The role of aberrant transcription factor expression and loss of epigenetic control in activating long-terminal-repeats in Hodgkin's lymphoma.

Benjamin Edginton-White

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences

College of Medical and Dental Sciences

University of Birmingham

September 2017

ABSTRACT

Approximately 8% of the human genome is composed of long terminal repeat (LTR) elements which originate from ancient retroviral germline infections. LTRs have the potential to act as alternative promoters and enhancers meaning that their expression is usually under tight epigenetic control to prevent aberrant gene expression. LTR activation has been reported in a number of diseases including Hodgkin's Lymphoma (HL) where an LTR acts as a promoter for the growth factor receptor gene *CSF1R* which is required for HL cell survival.

To investigate the genome-wide activation of LTRs in HL and their impact on gene expression we developed a novel targeted next generation sequencing approach (RACE-Seq) and integrated this data with global gene expression RNA-Sequencing as well as chromatin profiling data from HL and non-HL cell lines.

We discovered a unique pattern of LTR activation in each cell line, which correlated with changes in gene expression, including a number of HL-associated genes. We also showed that LTR expression can be induced by activation of inflammatory signalling pathways. Together these results show that LTR activation presents an additional level of gene expression deregulation in HL and highlight the potential for the impact of genome-wide LTR activation in other inflammatory diseases.

ACKNOWLEDGEMENTS

Foremost I would like to express my gratitude to Professor Constanze Bonifer for her guidance, support and commitment to both me and my project over the last 4 years. I also thank my co-supervisor Professor Peter Cockerill for his input throughout my project. I also would like to thank Arthur Riggs from the Beckman Research Center of City of Hope who funded my PhD through a generous Research Contract and the Kay Kendal Leukaemia fund for funding this research project.

Further to this I am very grateful to all the members of the Bonifer and Cockerill Labs, past and present for making me feel so welcome and for always being helpful, friendly and supportive. Particular thanks goes to Dr. Pierre Cauchy for his extensive involvement in the bioinformatic analysis for this project, his scientific input and for sharing his knowledge to allow me to carry out my own bioinformatic analysis. Additional thanks also goes to Dr Maria Rosaria Imperato for her support and guidance particularly in performing DNaseI-Seq.

I would like to extend a special thanks to two people. Firstly Leigh O'Connor-Jones for all of our scientific and not so scientific conversations and for making me a better scientist by always encouraging me to think critically. Secondly, Panchali Kanvatirth for being a wonderfully supportive and understanding friend during the ups and downs of PhD life.

Finally thank you to my family, friends and partner Jen who have offered support and encouragement throughout my PhD.

CONTENTS

1. INTRODUCTION 1

1.1. Chromatin and Nuclear Organisation 1

1.1.1. Chromatin 1

1.1.2. Nucleosome Structure 1

1.1.3. Histone Variants 2

1.1.4. Histone exchange 3

1.1.5. Post-translational modifications (PTMs) of Histones 4

1.2. Eukaryotic Transcription 9

1.2.1. Promoters 9

1.2.2. Enhancers 9

1.2.3. Transcriptional Initiation 10

1.2.4. Transcriptional Elongation 12

1.2.5. Termination, cleavage and polyadenylation 14

1.2.6. Transcription Factors 15

1.2.7. Silencers and Insulators 15

1.2.8. Chromatin Conformation 16

1.3. Repeat Elements in the Human Genome 18

1.3.1. Tandem Repeats 19

1.3.2. Interspersed Repeats 21

1.3.3. DNA Transposons 22

1.3.4. Retrotransposons 23

1.3.5. Non-LTR Retrotransposons 23

1.3.6. Non-LTR Retrotransposons function in the Human genome 24

1.3.7. LTR Retrotransposons – Human Endogenous Retroviruses (HERVs) 25

1.3.8. Classification of Human Endogenous Retroviruses 26

1.3.9. ERVL 27

1.3.10. MaLR 27

1.3.11. ERV1 28

1.3.12. ERVK 28

1.3.13. LTRs as Promoter and Enhancers 29
1.3.14. HERV role in Healthy Human Tissue 31
1.3.15. HERV role in Human Evolution 34
1.3.16. HERVs in Human disease 35
1.3.17. Regulation of HERVs 39
1.3.18. Methods for Identifying Active HERVs 42
1.4. B-Cell Development 44
1.5. Hodgkin's Lymphoma 49
1.5.1. HRS cells 50
1.5.2. Origin of HRS cells 51
1.5.3. HRS Cell Development 52
1.5.4. NF-κB Expression in HRS Cells 53
1.5.5. HRS Cell Expression Patterns 54
1.5.6. EBV 55
1.5.7. CSF1R 55
1.6. Aims and Objectives 57
2. MATERIALS AND METHODS 60
2.1. Cell Culture 60
2.2. Gene Expression Analysis 61
2.2.1. RNA Extraction 61
2.2.2. cDNA Synthesis 61
2.2.3. qPCR 61
2.2.4. RNA-Sequencing 62
2.3. Chromatin Accessibility Assays 64
2.3.1. DNaseI Hypersensitive Site Mapping 65
2.3.2. Assay for Transposase Accessible Chromatin 68
2.4. Rapid Amplification of cDNA Ends followed by Sequencing (RACE-Seq) 70
2.4.1. 5' RACE 70
2.4.2. RACE Validation by Cloning 72
2.4.3. RACE-Seq 73
2.5. ChIP-Seq 75
2.6. siRNA Knockdown *TNFRSF11A* 78

2.7. Nuclear Protein extraction and Western Blotting 78
2.8. iIkKβ Cloning 79
2.8.1. Cloning iIkKβ into PCW57.1 79
2.8.2. Lentivirus Production and Transduction of Reh cells 80
2.9. Data Analysis 81
2.9.1. RNA-Seq 81
2.9.2. DNaseI-Seq, ATAC-Seq and ChIP-Seq 82
2.9.3. Digital Genomic Foot-printing 83
2.9.4. Digital genomic footprinting motif co-occurrence clustering 83
2.9.5. RACE-Seq 84
2.9.6. LTR presence by gene expression fold change 85
3. RESULTS 87
3.1. The Hodgkin's Lymphoma transcriptional network has a global impact on chromatin structure and gene expression. 87
3.1.1. Hodgkin's Lymphoma displays a unique gene expression pattern 87
3.1.2. B-Cell specific gene expression is downregulated and pro-inflammatory genes are upregulated in HL. 91
3.1.3. THE1B LTRs are more active in HL cell lines than control cell lines. 105
3.2. HL cells display a global activation of long terminal repeat elements. 107
3.2.1. CSF1R is expressed from an active THE1B LTR in HL cell lines 107
3.2.2. 5' RACE-Seq can be used to identify active THE1B LTRs. 108
3.2.3. THE1B RACE-Seq identifies other associated MalR sub-families of repeats. 111
3.2.4. HL cell lines have a unique activation pattern of LTRs 116
3.3. LTR activation contributes to global deregulation of gene expression in HL cell lines. 120
3.3.1. Active LTRs are mainly located in intergenic and intronic regions 120
3.3.2. Active LTRs in HL act as alternative promoters 123
3.3.3. LTR activation in HL contributes to global deregulation of gene expression 124
3.3.4. TNFRSF11A is expressed from an active LTR in the L1236 cell line and is involved in the up-regulation of NF-κB activation 136
3.4. Long Terminal Repeats can be activated by inflammatory signalling 141
3.4.1. Treatment of the Reh cell line with PMA induces *CSF1R-LTR* expression 141
3.4.2. Treatment of the Reh cell line with PMA induces global THE1B LTR activation 143
3.4.3. LTRs activated by PMA treatment have a genome-wide impact on gene expression 146

3.4.4. PMA treatment of Reh cells induces global changes in chromatin accessibility and pushes the cells towards the HL chromatin accessibility pattern 150

3.4.5. Constitutive NF-κB activation can be triggered in Reh cells using doxycycline inducible IκKβ expression 153

3.4.6. LTR activity can be induced by aberrant NF-κB activation 157

4. DISCUSSION 164

4.1. The Hodgkin's Lymphoma transcriptional network has a global impact on chromatin structure and gene expression. 164

4.1.1. Generating a novel RNA-Seq dataset 164

4.1.2. Generating a high-resolution DNAseI-Seq dataset for digital foot-printing 169

4.2. RACE-Seq is an effective technique for the genome-wide identification of specific types of active LTRs 171

4.3. HL cells display wide-spread activation of long terminal repeat elements 174

4.3.1. LTR activation contributes to global deregulation of gene expression in HL cell lines. 176

4.3.2. HL specific genes transcribed from activated LTRs may be part of HL pathology 178

4.4. Long Terminal Repeats can be activated by inflammatory signalling 182

4.5. Conclusion 187

4.6. Future Work 188

4.6.1. Further examination of the cis-regulatory elements driving HL 188

4.6.2. Genome-wide mapping of LTRs 188

4.6.3. LTRs as Alternative Promoters and Enhancers 189

4.6.4. Direct Impact of LTR activation 191

4.6.5. Epigenetic control mechanisms for LTR repression 192

4.6.6. Primary Cells 193

5. SUPPLEMENTARY TABLES 194

6. REFERENCES 259

LIST OF FIGURES

Figure 2.1. DNaseI digestion validation gel.. 66

Figure 2.2. qPCR validation of DNaseI digestion levels 67

Figure 2.3 THE1B primer design 70

Figure 2.4 Schematic representation of RACE-Seq protocol 74

Figure 3.1 High level of correlation between biological replicates of RNA-Seq 88

Figure 3.2. Correlation clustering of gene expression patterns of the HL and non-HL cell lines 89

Figure 3.3 Gene expression level of a panel of 13 genes with a known involvement in HL.....90

Figure 3.4. Gene expression of a panel of 13 genes with a known involvement in HL measured by qPCR. 91

Figure 3.5. Gene Ontology analysis shows a down-regulation of genes associated with B cell processes in HL cell lines and an up-regulation of inflammatory and immune response genes.. 95

Figure 3.6. KEGG pathway analysis of up- and down-regulated genes in HL. 96

Figure 3.7. Network of up- and down-regulated genes in HL and their involvement in KEGG pathways 98

Figure 3.8 DNaseI-Seq data UCSC Genome Browser screenshot. 99

Figure 3.9. Correlation of HL and control cell lines based on chromatin accessibility 100

Figure 3.10. Transcription factor binding motif enrichment in DNase-I footprints in L428, KM-H2 and Reh 102

Figure 3.11. The transcriptional programme in HL cell lines is driven by the transcription factors NF-κB and AP-1. 104
Figure 3.12. PAX5 expression is lost in HL cell lines 105
Figure 3.13. Intergenic THE1B LTRs are more active in HL cell lines than control cell lines 106
Figure 3.14. *CSF1R* is expressed from an upstream THE1B LTR in HL cell lines. . 108
Figure 3.15. Scheme depicting *CSF1R* transcription start sites and 5' RACE strategy 109
Figure 3.16. The THE1B LTR which acts as a promoter for *CSF1R* in HL can be detected by RACE-Seq 111
Figure 3.17. Overlap of biological replicates of LTRs detected by RACE-Seq. 113
Figure 3.18. The THE1B primer anneals with multiple mismatching bases during the RACE-Seq PCR 113
Figure 3.19. RACE-Seq using a THE1B primer identifies a wide range of MalR repeat elements 115
Figure 3.20. HL and control cell lines have a unique pattern of LTR activation 116
Figure 3.21. Comparison of active LTRs in three HL cell lines 117
Figure 3.22. HL cell lines correlate based on LTR activation determined by RACE-Seq 119
Figure 3.23 Active LTRs are primarily located in intergenic and intronic regions in HL and control cell lines 122
Figure 3.24. Active LTRs produce downstream RNA transcripts 124
Figure 3.25. Active LTRs cluster with up-regulated genes in HL cell lines. 125

Figure 3.26. HL cell lines cluster based on expression of genes down-stream of active LTRs. 127
Figure 3.27. Intergenic LTRs act as alternative promoters in HL..... 129
Figure 3.28. *NLRP1* is expressed from a THE1B LTR promoter in HL..... 130
Figure 3.29 Intragenic LTRs produce shorter isoforms of some genes in HL..... 131
Figure 3.30. Active LTRs can produce anti-sense RNA transcripts correlating with reduced gene expression 133
Figure 3.31 CHD1L is down-regulated in HL cell lines compared to control cell lines 134
Figure 3.32. Active LTRs produce unannotated lncRNA transcripts..... 135
Figure 3.33. *TNFRSF11A* is expressed in the L1236 HL cell line from an active upstream THE1B LTR 137
Figure 3.34 High levels of TNFRSF11A are expressed from a THE1B LTR in the L1236 HL cell line. 138
Figure 3.35. TNFRSF11A can be knocked down by siRNA 139
Figure 3.36. Knockdown of *TNFRSF11A* in L1236 cells results in reduced NF-κB activation 140
Figure 3.37. THE1B consensus sequence contains an NF-κB binding motif. 141
Figure 3.38. Treatment of Reh cells with PMA induces CSF1R LTR expression. ... 142
Figure 3.39. ETO2 expression reduces following PMA treatment of Reh cells 143
Figure 3.40. Overlap of biological replicates of LTRs detected by RACE-Seq in PMA treated Reh cells 144
Figure 3.41. Activation of transcription from the *CSF1R LTR* can be detected by RACE-Seq in Reh cell line following PMA treatment..... 145

Figure 3.42. Following PMA treatment LTRs are transcriptionally activated 145
Figure 3.43. Correlation of RNA-Seq biological replicates from PMA treated Reh cells 146
Figure 3.44. PMA treatment of Reh cells results in an increased correlation of its gene expression pattern with that of HL cells 147
Figure 3.45 Reh cell growth is stopped following PMA treatment.....148
Figure 3.46. Transcriptionally active LTRs in PMA treated Reh cells are associated with up-regulated genes 149
Figure 3.47. LTRs activated by PMA treatment of Reh cells associate with genes which are highly expressed in HL cell lines 150
Figure 3.48. Following PMA treatment chromatin accessibility patterns of Reh cells shift towards those of HL cell lines 152
Figure 3.49. Inducible NF-κB activation scheme 153
Figure 3.50. Scheme depicting the mechanism of NF-κB activation 154
Figure 3.51. Flow cytometry analysis showed induction of IκKβ and IRES GFP construct following doxycycline treatment 155
Figure 3.52. IκKβ(EE) is induced by dox treatment 156
Figure 3.53. Nuclear NF-κB protein level increases following IκKβ induction.....157
Figure 3.54. Overlap of RACE-Seq biological replicates 158
Figure 3.55. High levels of NF-κB activation induces activation of LTRs.....158
Figure 3.56. A proportion of LTRs active in HL is activated by constitutive NF-κB activation in Reh cells.....160
Figure 3.57 Supervised clustering of LTR activation across cell lines 161

Figure 3.58 Clustering of gene expression data from cells with induced NF-κB activation showed little change in overall gene expression 162
Figure 3.59 Up-regulation of HL gene expression following constitutive NF-κB activation 163

LIST OF TABLES

Table 2.1. Culture densities of cell lines. 60

Table 2.2. Expression primers were designed based on DNA sequences from RefSeq or from PrimerBank (*).....62

Table 2.3 RACE Primers 72

Table 2.4 ChIP-Seq Primers.....77

Table 2.5 Western blotting antibodies 79

Table 3.1 Multiple active MalR family LTRs can be identified using 5' RACE and a THE1B primer.....110

ABBREVIATIONS

AP1 – Activator Protein 1

ATAC – assay for transposase accessible chromatin

ATP – adenosine triphosphate

bp – base pair

BSA – bovine serum albumin

ChIP – chromatin immunoprecipitation

cDNA – complementary DNA

DMEM – Dulbecco's Modified Eagle's Medium

DNMT – DNA methyltransferase

Dox – Doxycycline

DSG – disuccinimidyl glutarate

ERV - Endogenous Retrovirus

FACS – fluorescence activated cell sorting

FCS – foetal calf serum

FPKM – Fragments per kilobase of transcript per million mapped reads

GO – Gene Ontology

HAT – histone acetyltransferase

HDAC – histone deacetylase

HERV – Human Endogenous Retrovirus

HL – Hodgkin's Lymphoma

HRS – Hodgkin/Reed Strernberg

IMDM – Iscove Modified Dulbecco Medium

JAK-STAT – Janus Kinase-Signal Transducer and Activator of Transcription

Kb – kilobase

KARB – Krüppel Associated Box

lncRNA – Long Non-coding RNA

LTR – Long Terminal Repeat

MaLR – Mammalian apparent LTR Retrotransposon

MAPK - Mitogen-activated protein kinase

NF-κB – Nuclear Factor kappa B

nt – nucleotides

PBS - phosphate buffered saline

PCR – polymerase chain reaction

PIC – protease inhibitor cocktail

PKC – Protein Kinase C

PMA - phorbol 12-myristate 13-acetate

qPCR – quantitative PCR

RACE – Rapid Amplification of cDNA Ends

RIPA – radioimmunoprecipitation assay

RNAPII – RNA polymerase II

Seq – sequencing

TF – transcription factor

TNF – Tumour Necrosis Factor

TSS – Transcription Start Site

TTS – Transcription Termination Site

ZFP – Zinc Finger Protein

ZNF – zinc finger

1. INTRODUCTION

1.1. Chromatin and Nuclear Organisation

1.1.1. Chromatin

The presence of chromosomes within the nuclei of cells was first discovered in the mid-19th century long before Franklin provided the data for Watson and Crick to elucidate the double helix structure of DNA (Watson and Crick 1953). Flemming demonstrated that during cell replication the material within the nucleus formed thread like structures and he went on to identify the different stages of mitosis (Flemming 1882, Paweletz 2001). We now know that the nucleus of human cells contains around 3 billion base pairs of DNA equalling around 2 metres in length which require intricate packaging to fit within the nucleus (Bloom and Joglekar 2010). To achieve this level of compaction, the double helix of DNA is coiled around nucleosomes which fold to produce fibres and are then tightly coiled into the chromatids of chromosomes. Although it is important to neatly package DNA within the nucleus it is also necessary to maintain accessibility at specific regions for replication and transcription which is mostly achieved by modifications at the nucleosome level.

1.1.2. Nucleosome Structure

A nucleosome is formed of 147 bp of DNA coiled 1.65x around an octamer of histones. Each nucleosome contains 2 of each of 4 different histones, H2A, H2B, H3 and H4. Each histone has a structural domain with 3-alpha-helices separated by 2 loops which are essential for DNA interaction (Kornberg 1974, Kornberg and Lorch 1999). H2A heterodimerizes with H2B and H3 with H4, and the two H3/H4 dimers then form a central tetrameric axis. The two H2A/H2B dimers join the tetramer to form the complete nucleosome octamer. Histone H1 attaches to the outside of the nucleosome and serves to both lock the DNA in place and as a linker to other nucleosomes to form higher order chromatin. It is essential to maintain accessibility to regions of DNA for transcription. This is achieved by a combination of post-

translational modifications (PTMs) to the protruding amino-terminal tails of the histones, through histone variants and nucleosome remodelling complexes (Fig. 1.1).

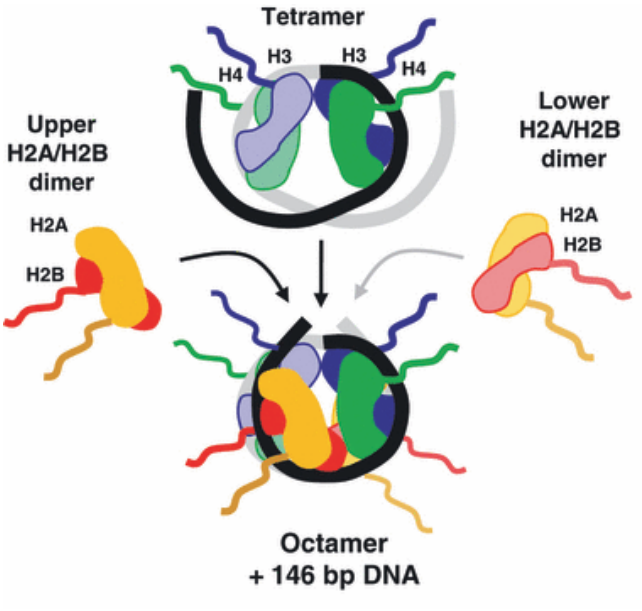


Figure 1.1. Nucleosome Structure – Two H3-H4 dimers form a central tetramer which Two H2A-H2B dimers join to form the final octamer structure of the nucleosome. Histone tails can be seen protruding from the nucleosome. Adapted from: (Cockerill 2011).

1.1.3. Histone Variants

The canonical histones that make up the nucleosome can be replaced with variants that have anywhere from a few amino acid substitutions to entire domains being gained or lost. The integration of these histones into nucleosomes can change the properties of the nucleosome impacting on chromatin structure and DNA – protein interactions (Kornberg 1974).

The canonical histone H3.1 is upregulated during mitosis and substitutes for variant H3.3 at replication origins to destabilise nucleosomes allowing access to the DNA for replication (Tagami, Ray-Gallet et al. 2004). H3.3, unlike H3.1 is a replication-independent variant. Replication-independent variants have a continuous level of expression throughout cell cycle.

H3.3 differs from H3.1 only by 5 amino acids and shows no other major structural differences. However, it plays an important role in transcriptional regulation as H3.3 is highly dynamic and is able to contribute to a more rapid activation of genes silenced by histone modifications independent of replication (Ahmad and Henikoff 2002, Tagami, Ray-Gallet et al. 2004).

H2A.Z is another important histone variant which is expressed in all organisms along with H2A (Talbert and Henikoff 2010). The incorporation of H2A.Z in a nucleosome only produces subtle structural changes, however, this is enough to destabilise the nucleosome allowing for transcription and DNA repair to take place (Meneghini, Wu et al. 2003, Venkatesh and Workman 2015). Both H3.1 and H2A.Z are also associated with transcription start sites (TSS's). H3.1 is often associated with silenced genes whereas the H3.3 variant is often located in proximity to TSS's of transcribed genes (Tagami, Ray-Gallet et al. 2004). H2A.Z is incorporated into the +1 phased nucleosome that sits upstream of the nucleosome free region at TSS's and is often associated with genes which have a moderate expression level (Soboleva, Nekrasov et al. 2014).

1.1.4. Histone exchange

Histone exchange is a critical process for the integration of histone variants and also to allow chromatin to have a fluid structure for replication, repair and transcription to take place. A number of factors contribute to histone exchange including chromatin re-modellers, histone chaperones and post translational histone modifications (PTM), which each play roles in either weakening the histone-histone or histone-DNA interactions (Venkatesh and Workman 2015).

Chromatin remodellers act in an ATP dependent manner driving a DNA translocase and are able to slide or remove nucleosomes, opening DNA for interaction with histone chaperones and for transcription (Tsukiyama, Becker et al. 1994). In the context of transcription the

INO80 and SWR chaperones (yeast orthologue to human SRCAP) are needed to exchange H2A histones for the H2A.Z variant which reduces the stability of the nucleosome allowing transcription to take place (Ruhl, Jin et al. 2006). SWR functions by carrying out ATP hydrolysis in the presence of the H2A-H2B dimer meaning that it will only be active at sites where the H2A.Z variant is not present so it can add the variant but cannot remove it. It is thought that the INO80 complex functions in opposition to SWR and removes non-acetylated H2A.Z dimers avoiding miss-targeting of the histone variant to areas of chromatin which should remain closed and avoiding aberrant transcription (Papamichos-Chronakis, Watanabe et al. 2011, Watanabe, Radman-Livaja et al. 2013).

Unlike chromatin remodellers, histone chaperones are not usually ATP dependent and rely on spontaneous DNA movement (Hondele, Stuwe et al. 2013). Chaperones perform a number of functions including assisting in maintaining the stability and organisation of nucleosomes. They are able to interact with remodellers to act as histone sinks and also to regulate histone post translational modifications (PTMs). For example the repressive H3K9me3 histone modification has been shown to be maintained at heterochromatin by deposition of H3.3 by the ATRX/DAXX chaperone complex (Voon and Wong 2016) Chaperones are also required for the assembly of newly synthesised nucleosomes following replication as part of the FACT complex (discussed further in 1.2.4). Finally following histone modifications of other domains within the nucleosome octamer chaperones are needed to provide access to the PTM enzymes (Youdell, Kizer et al. 2008).

1.1.5. Post-translational modifications (PTMs) of Histones

The presence of an amino-terminal tail on each of the histones which protrudes from the nucleosome allows for many potential post-translational modifications to be easily added to histones (Luger, Mader et al. 1997, Margueron and Reinberg 2010). It has been shown in numerous studies that these histone modifications play an important role in the control of

chromatin accessibility and therefore gene expression (Brownell, Zhou et al. 1996, Vaquero, Loyola et al. 2003, Kouzarides 2007). Many different histone modifications have been identified and there are no doubt still more to be discovered. Although many modifications occur on amino acids in the protruding tails of the histones there are also a number which occur at other amino acids including those of the globular domain.

Post-translational histone modifications have a number of important functions in biological processes including transcriptional initiation and repression, elongation, mitosis and maintenance of both euchromatin and heterochromatin (Table 1.1)(Lawrence, Daujat et al. 2016). Histone modifications can act alone or in the case of double modifications can influence each other (Cheung, Tanner et al. 2000, Lee, Smith et al. 2010). This can be achieved by 3 main routes: (1) altering the histone-DNA interaction (increasing or decreasing the binding affinity of the nucleosome), (2) acting as an interaction partner to recruit proteins and regulate transcription and (3) to prevent other proteins from binding. Heterochromatin displays a generally repressed transcriptional state which is predominantly maintained through methylation of H3. Although generally transcriptionally inactive it is now recognised that it can be split into 2 groups, facultative and constitutive (Bannister and Kouzarides 2011). Facultative heterochromatin contains genes which are differentially regulated often during cell differentiation where gene activation is only required at certain stages, this chromatin often has H3K27me3 and H2AK119ub modifications. Constitutive heterochromatin, often with marked H3K9me2/3, is found at the repeat regions of telomeres and centromeres where genes are permanently silenced. H3k9me3 is also often found at transposable elements throughout the genome to maintain silencing preventing the elements from acting as transposons, promoters or enhancers (Hurst and Magiorkinis 2017).

Euchromatin represents the less tightly packed chromatin incorporating H2A.Z histone variants and contains many transcriptionally active genes. Expressed protein coding genes have a fairly well defined set of histone modifications spread throughout the gene. H3K4me3

is present at the promoters of actively transcribed genes (Liang, Lin et al. 2004). These marks are added to H3 by the SETD1/MLL complex which has a preference for methylating histones in vicinity to CpG islands, a common feature of eukaryotic promoters (Deaton and Bird 2011). The presence of H3K4me3 has been shown to both have involvement in the recruitment of Pol II and specific transcription factors to the promoter and also as part of a positive feedback loop to reinforce its own deposition (Yokoyama, Wang et al. 2004, Zhang, Cooper et al. 2015).

Another notable histone modification on nucleosomes of actively transcribed genes is H3K36me3 which accumulates towards the 3' end of the gene (Pokholok, Harbison et al. 2005). In mammals this modification is carried out by the SETD2 enzyme (Wagner and Carpenter 2012). It has been shown that the presence of H3K36me3 recruits HDACs which deacetylate histones within the gene body. This activity is the opposite to that of H3K4me3 which causes an increase in acetylation at promoters and an increase in transcription initiation complex recruitment. By promoting deacetylation of gene bodies H3K36me3 helps to prevent the formation of spurious transcription start sites within genes (Zhang, Cooper et al. 2015).

In addition to the role of histone modifications at promoters they can also play a role at other transcriptional regulatory elements such as enhancers. Studies have shown that many elements in the genome which act as enhancers are flanked by H3K4me1, H3K4me2 and H3K27ac. Further studies attempting to classify subgroups of enhancers have also shown presence of H3K27me3 and H3K36me3 (Zentner, Tesar et al. 2011). It is thought that H3K4me1 facilitates the binding of readers to enhancers which then leads to their interaction with promoters (Calo and Wysocka 2013). However the exact function of enhancer histone marks is yet to be elucidated (Heinz, Romanoski et al. 2015).

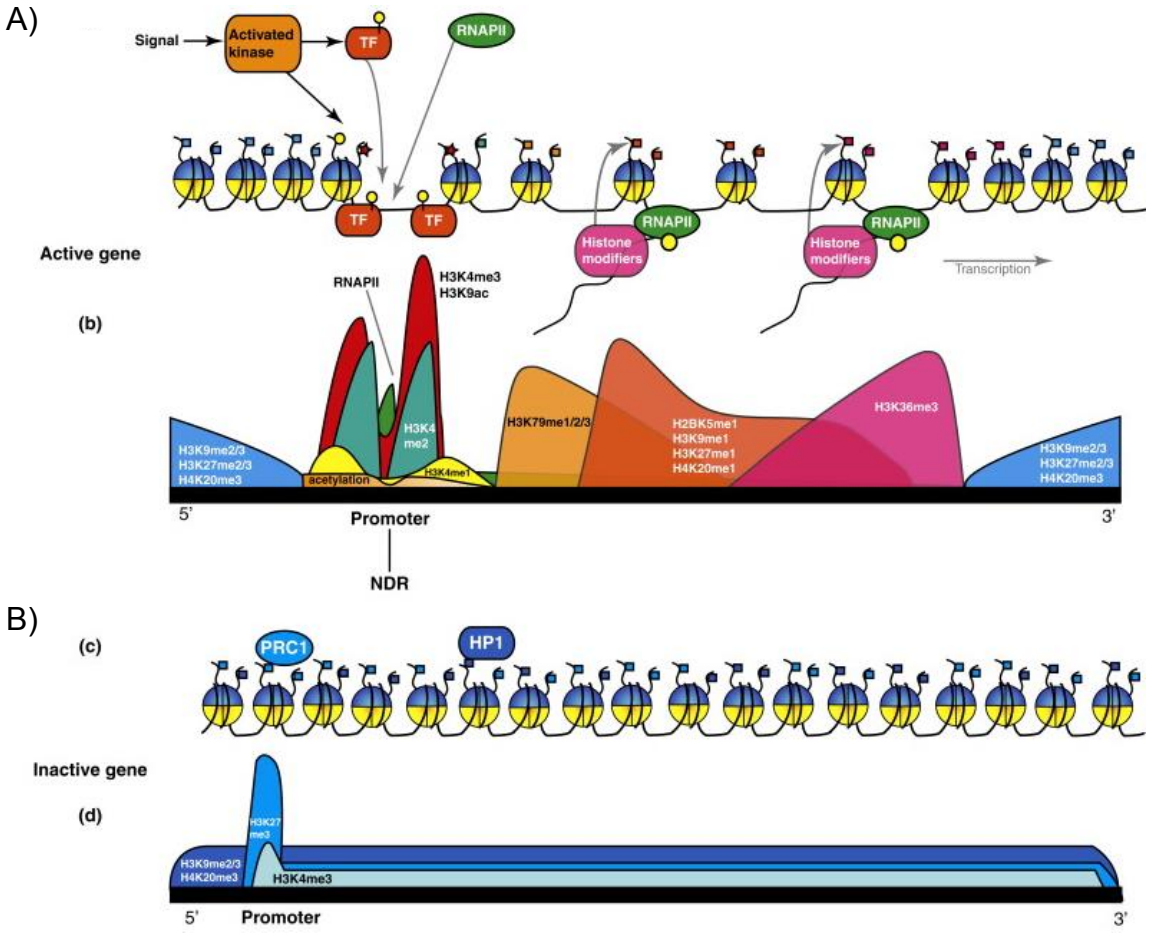


Figure 1.2. Distribution of histone modifications at active (A) and inactive (B) protein coding genes. Adapted from (Barth and Imhof 2010)

Table 1.1. Post translational histone modifications (human) and their role in chromatin. Adapted from(Lawrence, Daujat et al. 2016)

Histone	Modifications	Role
H2A	K4/5ac,K7ac	Transcriptional Activation
	S1P	Mitosis, Chromosome Assembly
	K119P	Spermatogenesis
	119uq	Transcriptional Repression
H2B	S14P	Apoptosis
	S33P, K5ac, K11/12ac, K15/16ac, K20ac, K123uq	Transcriptional Activation
	K120uq	Spermatogenesis/meiosis
H3	K4me2	Permissive euchromatin
	K4me3	Transcriptional elongation, active euchromatin
	K9me3	Transcriptional repression; imprinting; DNA methylation
	R17me, K4ac, K27ac	Transcriptional Activation
	K27me3	Transcriptional silencing; X-inactivation; bivalent genes/gene poising
	K36me3	Transcriptional Elongation
	K9ac	Histone deposition; transcriptional activation
	K14ac, K23ac	Transcriptional activation; DNA repair
	K18ac	Transcriptional activation; DNA repair; DNA replication
	T3P, T11/S28P	Mitosis
	S10P S28P	Mitosis; meiosis; transcriptional activation
H4	R3me	Transcriptional Activation
	K20me1	Transcriptional silencing
	K20me3	Heterochromatin
	K5ac	Histone deposition; transcriptional activation; DNA repair
	K8ac	Transcriptional activation; DNA repair; transcriptional elongation
	K12ac	Histone deposition; telomeric silencing; transcriptional activation; DNA repair
	K16ac	Transcriptional activation; DNA repair
	S1P (serine 1 phosphorylation)	Mitosis

1.2. Eukaryotic Transcription

1.2.1. Promoters

The core promoter region of a gene lies upstream of the transcription start site (TSS) and defines the location of the TSS. Promoters are traditionally classified into 2 groups, TATA box and CpG island promoters. The TATA box is recognised by TBP or a TAF mediating the formation of the RNA Pol II initiation complex. CpG island promoters are sequences containing an atypically high proportion of CpG sites and are spread throughout the human genome. They are able to bind transcription factors which can then recruit TBP and the Pol II initiation complex (PIC) (Illingworth and Bird 2009).

Although the TATA-box promoters were the first to be discovered a far more diverse range of promoters have since been identified. These can include combinations of a number of elements including Inr (initiator motif), DPE (downstream core promoter element), MTE (motif ten element) which act as TFIID binding sites. A number of other elements can also be involved including BRE a TFIIB recognition element and DCE (downstream core element) (Juven-Gershon, Hsu et al. 2008). These elements work in combination with tissue specific transcription factor motifs to initiate the formation of the PIC.

It should be noted that many CpG island have been identified at sites which are not annotated gene promoters (Illingworth and Bird 2009). It is thought that these sequences may still act as promoters for other elements and many of these sites may be long terminal repeat elements from endogenous retroviruses (discussed in 1.3.7).

1.2.2. Enhancers

Enhancers are transcriptional regulatory elements defined by their ability to activate transcription regardless of their distance from the promoter of a gene as shown in

transfection assays of reporter genes such as luciferase (Bulger and Groudine 2010). Enhancers are often around 200-400bp in size and can reside within a gene or be located many hundreds of kilo-bases from the promoter with which they interact (de Villiers and Schaffner 1981, He, Meyer et al. 2010). Enhancers contain motifs to attract binding of specific transcription factors which in turn recruit cofactors. Through looping of the DNA these complexes interact with the PIC at a promoter increasing the transcriptional activity and the expression level of the gene. The looping activity of enhancers was first shown in vivo at the locus control region (LCR) of the β-globin locus which is located 40-60 Kb away from the gene promoter but is in close spatial proximity (Carter, Chakalova et al. 2002, Tolhuis, Palstra et al. 2002). Individual enhancers are often tissue specific and not necessarily required for the transcription of a gene *per se* but mediate tissue specific expression. Enhancer promoter interactions can be highly specific and often do not involve the promoter of the nearest gene and some promoters have interactions with more than one enhancer in different tissues (Sanyal, Lajoie et al. 2012, Mifsud, Tavares-Cadete et al. 2015). A recent study also demonstrated that many promoters can also act as distal enhancers for unrelated genes (Dao, Galindo-Albarran et al. 2017). This means that the enhancer effect of a gene promoter becoming active should also be considered along with expression of its transcribed gene.

1.2.3. Transcriptional Initiation

Almost every cell in a eukaryotic organism carries the same DNA within its nucleus, providing the code for every protein that is required by the organism. As every cell contains the same set of genes it is necessary to regulate the expression of genes on a cell type specific basis to ensure correct functioning of a cell. For tissue-specific genes, the largest part of gene expression control occurs at the stage of transcription initiation, where DNA is processed by RNA Polymerase II (Pol II) to form mRNA which is later translated to amino acid sequences and functional proteins.

The initiation of transcription of a protein coding gene starts with the formation of the Pol II initiation complex (PIC) which is formed by a group of proteins termed general transcription factors (GTF) (Liu, Bushnell et al. 2013). The process begins with the binding of TFIID, TATA box binding protein (TBP) and TBP associated factors (TAFS) to double stranded DNA at the promoter of a gene. This forms a complex with TFIIB which assists with TBP binding forming the core initiation complex(Kostrewa, Zeller et al. 2009).

The TBP and TFIIB complex has been shown to be sufficient for the recruitment of Pol II and for transcription (Tyree, George et al. 1993). TBP recognises and binds to TATA box sequences (TATAWAWR) which are often located around 30 bp upstream of the transcription start site, however, only 10-20% of promoters have a TATA box (Basehoar, Zanton et al. 2004). As TBP is an essential member of the initiation complex, it is thought that in TATA-less promoters binding is assisted by the TAFs recognising elements around the promoters. Once bound to the minor groove of double stranded DNA at a promoter, TBP bends the DNA at almost 90° which effectively functions to wrap the DNA around Pol II and this feature may be the reason that TBP is required even at TATA-less promoters (Sainsbury, Bernecky et al. 2015).

The binding of TBP and its bending function is further assisted by TFIIB which binds to upstream and downstream recognition elements. Through the interaction with recognition elements TFIIB also acts to orientate the complex for transcription towards the direction of the gene. It should be noted that many Pol II promoters also act in a bidirectional manner (Fong, Saldi et al. 2017). TFIIB also serves to recruit Pol II through the binding of the B-ribbon of TFIIB to the dock domain of Pol II (Chen and Hahn 2003). Once the core initiation complex is formed Pol II and TFIIF are recruited to the complex and are joined by TFIIE and TFIIH forming the complete PIC.

The TFIIF component of the complex forms a heterodimer binding to the Pol II near the downstream DNA. The winged helix domain of TFIIF interacts with TFIIB and its downstream recognition elements thereby helping to prevent non-specific interaction of Pol II with DNA (Cabart, Ujvari et al. 2011). The final two components of the complex, TFIIE and TFIIH are required for the opening of the double stranded DNA to form a transcription bubble. TFIIE binds to Pol II and acts as a bridge for the binding of TFIIH. TFIIH functions as an ATP dependent molecular 'wrench' which creates torque, melting the double stranded DNA to form a single stranded DNA bubble which the Pol II can use as a template to produce an mRNA (Grunberg, Warfield et al. 2012).

Once the PIC has formed at a promoter the process of elongation can begin. Transcription usually begins around 30 bp downstream of the TATA box for TATA box containing promoters.

1.2.4. Transcriptional Elongation

The process of elongation can be split into two stages, early elongation and productive elongation. Early elongation begins once around 12 nucleotides of nascent RNA have been produced, at this point RNA is capped and the GTFs are no longer needed(Tang, Roy et al. 2009). Although not required for elongation TFIID, TFIIA and TFIIH can remain bound at the promoter of some genes acting as a scaffold for the re-initiation of transcription (Yudkovsky, Ranish et al. 2000). TFIIF can also remain bound to Pol II as an elongation factor (Cabart, Ujvari et al. 2011). In the initial models of transcription it was thought that once Pol II escaped the initiation complex it would then proceed uninterrupted to produce the full RNA transcript. Based on observation of accumulation of Pol II in close proximity to promoters of many genes it has been shown that Pol II pauses after transcribing 20-60 nucleotides at many genes (Kwak and Lis 2013).

Pol II pausing is an important level of regulation in gene expression and the strongest rate limiting step for Pol II (Kwak and Lis 2013). Pausing can be explained by three different mechanisms. The kinetic model suggests that pausing is dependent on the energy state of the DNA-RNA hybrid and is influenced by factors such as the initial rate of elongation. An example for this notion is that in a higher energy state Pol II can backtrack and requires TFIIS interaction to restart elongation from a paused state (Adelman, Marr et al. 2005). The second model is based on a physical barrier being present which pauses elongation. This barrier can be a nucleosome which is supported by the observation that pausing often occurs between the promoter and first nucleosome (Izban and Luse 1991). These 2 models alone do not account for all promoters with proximal Pol II pausing. The third model demonstrates pausing as a result of specific binding factors which interact with Pol II and the RNA transcript. In particular NELF and DSIF bind the nascent RNA and are associated with Pol II pausing (Missra and Gilmour 2010). For transcription to proceed it is necessary for Pol II to escape into the productive-elongation phase.

The majority of genes which show promoter proximal pausing are expressed at least some of the time, showing that Pol II can escape from the paused state. Two processes are required for Pol II to escape from pausing. Firstly 5' capping of the nascent RNA transcript which is has been shown to occur as the Pol II escapes from its paused state (Rasmussen and Lis 1993). The second process is the phosphorylation of Pol II and pausing factors DSIF and NELF. This phosphorylation is carried out by P-TEFb which can be recruited to Pol II by activators such as RelA which recruits P-TEFb to TNF-α genes (Barboric, Nissen et al. 2001). It can also be recruited by co-activator Brd4 which recognises acetylated histone tails and by the Mediator complex which also assists Pol II in passing the +1 nucleosome barrier (Yang, Yik et al. 2005, Takahashi, Parmely et al. 2011, Nock, Ascano et al. 2012).

Following escape from pausing the remaining barrier to transcription is the presence of nucleosomes along the gene body. Passing these barriers is achieved using a number of

different factors including FACT, Spt6, PARP and PAF complex. FACT assists in progression of elongation by breaking the H2A-H2B dimer from a nucleosome which allows Pol II to continue through the remaining histones. FACT has also been shown to reassemble the H2A-H2B dimer with the other histones to restore the nucleosome (Orphanides, LeRoy et al. 1998, Xin, Takahata et al. 2009). Spt6 has also been shown to play a similar role of disassembling and reassembling nucleosomes through interaction with histones H3 and H4 (Bortvin and Winston 1996). After successful elongation transcription then needs to undergo termination.

1.2.5. Termination, cleavage and polyadenylation

The final stages of transcription are termination, cleavage of the RNA transcript and also polyadenylation of the transcript. Termination occurs anywhere between a few bases and several kilobases downstream of the 3' end of a transcript (Proudfoot 1989, Richard and Manley 2009). Termination is directed by the 3'-end processing of the mRNA transcript which involves cleavage and polyadnylation specific factor (CPSF) binding to an AAUAAA sequence along with cleavage stimulation factor (CstF) which binds at a GU rich sequence. These complexes include a polymerase which is able to polyadenylate the mRNA which signals for termination of Pol II elongation. There were 2 models proposed in the 1980's for the termination of Pol II elongation (Richard and Manley 2009). The allosteric model suggests that transcription through the poly A site leads to a conformational change in the elongation complex(Logan, Falck-Pedersen et al. 1987). The Torpedo model suggests that the rapid degradation of 3' RNA allows the entry of an exonuclease which releases the Pol II(Connelly and Manley 1988). More recent studies would suggest a combination of these theories is more likely to be true with Xrn2 degradation of the RNA transcript downstream of the poly A cleavage site either prior or post cleavage resulting in RNAPII release(West, Proudfoot et al. 2008).

1.2.6. Transcription Factors

Transcription factors bind specific DNA motifs and through protein-protein interaction domains are able to activate or in some cases repress transcription (Kadonaga 2004). The first specific transcription factors were identified and purified in the 1980's. Specificity protein 1 (Sp1) was one of the first transcription factors to be purified, it was shown to bind to the GC box of the SV40 enhancer (Dynan and Tjian 1983). Many transcription factors have since been discovered most of which bind to defined unique DNA binding motifs and are expressed in a tissue specific manner (von Strandmann, Nastos et al. 1997). Transcription factors can interact with each other and also recruit cofactors and bridging factors to maintain specific chromatin states as well as to recruit other proteins to activate or repress transcription. Through these interactions transcription factors play a central role in defining the gene expression programme of cells and therefore the specification of cell types from embryonic development through to adult tissues. Transcription factors can also act in a signalling dependent manor to trigger gene expression in response to certain stimuli. For example the transcription factors AP-1 (Activator protein 1) and NF-κB are stimulated by cytokines, growth factors and infection to produce an inflammatory response in cells (Hess, Angel et al. 2004, Brasier 2006).

1.2.7. Silencers and Insulators

There are a number of other elements which can influence control of gene expression. Silencers share many of the properties of enhancers, however effectively have the opposite function resulting in repression of gene expression (Baniahmad, Steiner et al. 1990). Silencers can function in a similar way to enhancers, by a looping mechanism interacting with a specific promoter via cofactors and preventing transcription. Silencers can also act through the generation of double stranded (duplex) RNA which can also interact with promoters to downregulated gene expression (Kolovos, Knoch et al. 2012). Through

interaction with binding factors such as the zinc-finger protein NRSF/REST (neuron-restrictive silencer factor) silencers are also able to act in a tissue specific manner repressing transcription in all but specific tissues (Schoenherr and Anderson 1995, Jones and Meech 1999).

Insulators act as boundaries preventing unwanted interaction between enhancers or silencers and promoters of genes which should not be targeted by the regulatory element (Kolovos, Knoch et al. 2012). The importance of insulators has been clearly demonstrated by the mutation or deletion of these regions which results in developmental defects (Gaszner and Felsenfeld 2006). Insulators, however, play a far larger role than just blocking unwanted chromatin interaction (Yang and Corces 2011). Studies have shown that through the binding of the transcription factor CTCF insulators are responsible for the looping of chromatin, its localisation within the nucleus and therefore potential interactions of regulatory elements (Reviewed by (Deng, Patel et al. 2015).

1.2.8. Chromatin Conformation

As previously discussed interaction of enhancers with promoters allows a physical proximity between promoter and enhancer elements. In recent years the importance of the conformational chromatin structure as a regulator of gene expression has been highlighted by the use of new techniques such as chromatin conformation capture (3C)(Dekker, Rippe et al. 2002) and genome wide approaches including Hi-C which enable the physical interactions between regions of DNA to be observed(Lieberman-Aiden, van Berkum et al. 2009). These studies have enabled the identification of topologically associated domains (TADS), regions which exhibit a high level of topological interactions internally with very few if any interactions passing the boundaries into neighbouring TADs (Dixon, Selvaraj et al. 2012). This observation has been further confirmed by mouse studies where the boundary sequences have been mutated, resulting in developmental defects (Lupianez, Kraft et al. 2015). The

binding of the transcription factor CTCF plays a crucial role in demarcating the boundaries of TADS and has been shown to be required at insulators to block enhancers (Bell, West et al. 1999). CTCF is further responsible for genome-wide chromatin looping, however it has been recently shown that in the absence of CTCF transcription is disrupted but not genomic compartmentalisation (Nora, Goloborodko et al. 2017). This suggests that other factors play a role in the overall topological arrangement of the genome.

1.3. Repeat Elements in the Human Genome

The disparity between genome size and organism complexity puzzled many early researchers, however, in the 1970s the discovery of functional non-coding DNA helped to resolve this puzzle (Gregory 2005). Many theories were proposed for the existence of what was assumed to be 'junk DNA'. The three main theories were that it consisted of functionless copies of genes or 'pseudogenes', that it consisted entirely of introns or that it served a structural purpose forming a nuclear skeleton (Comings 1972, Cavalier-Smith 1978, Gilbert 1978).

A number of studies followed the discovery of non-coding DNA, however, the biggest step forward in this field was no doubt the sequencing of the human genome. We now know that only ~1.5% of DNA codes for proteins and a further ~26% resides within introns of these genes which leaves ~73% of DNA outside of protein coding genes (Lander, Linton et al. 2001, Gregory 2005). A large proportion of the non-coding DNA is made up of repeat elements. A conservative estimate based on the RepeatMasker approach is that ~50% of the entire genome is composed of repeat elements (Figure 1.3) (Smit 2013-2015). This estimate is based on alignment of genomic DNA to repeat family consensus sequences from Repbase (Jurka 2000). Alternative estimates using a de novo approach would suggest that up to 69% of the genome is repeat derived (de Koning, Gu et al. 2011).

Repeat elements can be split into two distinct groups based on sequence structure, tandem repeats and interspersed repeats.

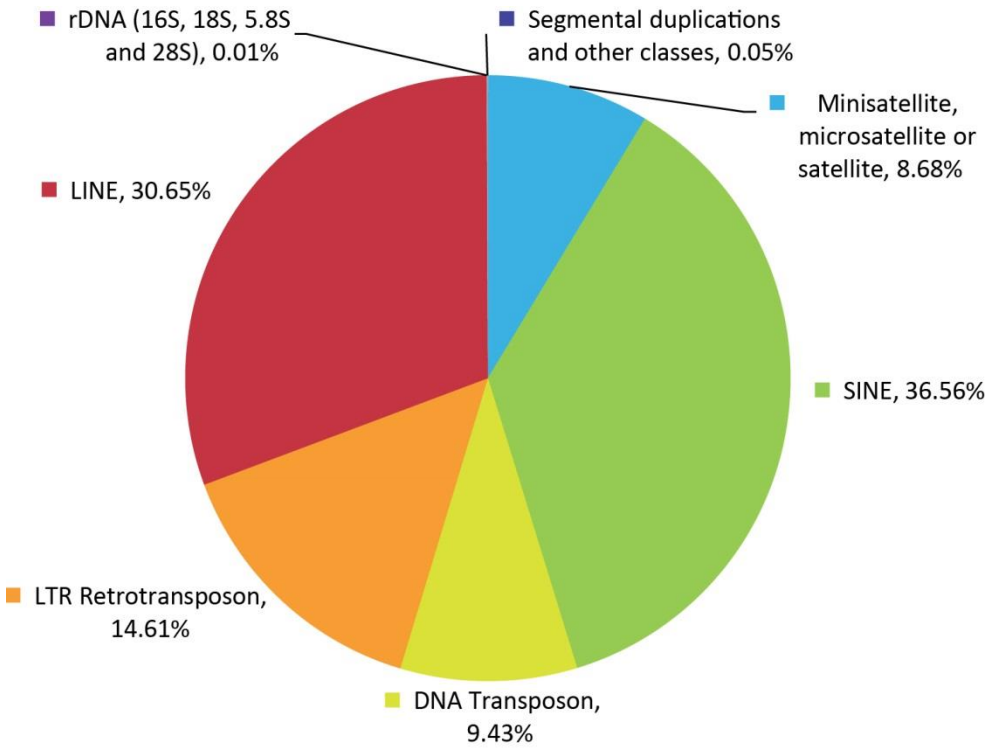


Figure 1.3. Percentage of repeat genome occupied by types of repeat element in humans (adapted from (Treangen and Salzberg 2011)).

1.3.1. Tandem Repeats

Tandem repeats can range anywhere from single bases to mega-bases in length and are divided into 5 groups; satellites, mini-satellites, micro-satellites, centromeric satellites and telomeric repeats. Mini-satellite repeats have a repeat length of 30-35 bp with a conserved 10-15 bp core sequence and can total 1-15 kb in length. Microsatellites are composed of dinucleotide to pentanucleotide repeats with a total length of up to 100's of base pairs (Fig 1.4, Table 1.2) (Padeken, Zeller et al. 2015). The reason for the formation of the microsatellite repeats is most likely slipped strand mispairing (Levinson and Gutman 1987). Slipped strand mispairing occurs when the polymerase briefly dissociates from the template strand and re-associates slightly up- or downstream which can result in repeated sequence. This occurs most often at short tandem repeats and in germ line cells which means that the length of the tandem repeats can vary greatly even between individuals of the same species.

There is also a reasonable amount of variability in the length of mini-satellite repeats, however, the reason for this is much less clear as the length of the repeats means it is much less likely to occur by chance (Ahmed and Liang 2012). The most likely cause for this variation is during gene conversion in meiosis (Richard and Paques 2000) although some are also thought to be derived from transposable elements (Ahmed and Liang 2012). The final class of tandem repeats are the centromeric and telomeric repeats. Telomeric repeats are simple arrays of tandem repeats (TTAGG in humans) located at the ends of chromosomes and through interaction with telomerase binding proteins serve a number of roles including protecting the chromosomes from exonuclease activity (Wai 2004). The centromeric repeats are also arrays of tandem repeats and are located at the centromere of chromosomes. They are bound by the histone variant CENH3 which plays an important role in mitosis (McKinley and Cheeseman 2016).

The general function of the tandem repeats in the human genome is not clear. However, in specific cases tandem repeat elements can act as origins of replication(Liu, Bissler et al. 2007), promoter components (Alakurtti, Virtaneva et al. 2000), enhancers (Tassone, Hagerman et al. 2000, Tassone, Beilina et al. 2007), gene silencers (van Overveld, Lemmers et al. 2003), transcriptional elongation blockers and translational regulators (Krol, Fiszer et al. 2007). These different functions are often influenced by the expansion or contraction of a tandem repeat element and in a number of cases have been linked to diseases, often resulting from the dysregulation of expression of a particular gene linked to such a repeat (Usdin 2008). Examples of this have been shown in triplet repeat neurological diseases. Regions of tandem repeats often form heterochromatin and through position effect variegation (PEV) nearby genes can also be silenced in a proportion of cells. It has been shown that in a number of neurological diseases such as myotonic dystrophy and Friedreich's ataxia small expansions of triplet-repeat regions results in PEV which contributes

to deregulation of gene expression (Saveliev, Everett et al. 2003, Nageshwaran and Festenstein 2015).

1.3.2. Interspersed Repeats

Interspersed repeats, also known as transposable elements (TEs), make up a far larger proportion of the human genome than tandem repeats and have the potential to play significant functional roles. The interspersed repeat elements can be classified into 2 general groups: DNA transposons making up around 9% of repeat elements in the human genome and retrotransposons which account for 82% of repeat elements. Retrotransposons are further classified into the Long-terminal-repeat (LTR) retrotransposons and the non-LTR retrotransposons.

Transposons were first identified by McClintock in the 1960's who observed 'jumping genes', showing for the first time that the genome was not stationary and fixed but that some sections of DNA could move around within the genome (McClintock 1968, Ravindran 2012). Transposons are basically defined as sections of DNA which are able to move around within the genome and can be found in all organisms (Munoz-Lopez and Garcia-Perez 2010).

Table 1.2. Length of different repeat element classes within the human genome Adapted from (Padeken, Zeller et al. 2015)

Repeat Class	Repeat Type	Length (bp)
Minisatellite, microsatellite or satellite	Tandem	2-100
SINE	Interspersed	100-300
DNA Transposon	Interspersed	200-2,000
LTR Retrotransposon	Interspersed	200-5,000
LINE	Interspersed	500-8,000
rDNA (16S, 18S, 5.8S and 28S)	Tandem	2,000-43,000
Segmental duplications and other classes	Tandem or Interspersed	1,000-100,000

1.3.3. DNA Transposons

A DNA transposon is formed of a pair of TIR (Terminal Interspersed Repeat) elements which sit either side of a transposase gene (Fig. 1.4). It is thought that DNA transposons became integrated into the genome ~50 million years ago following infection with dsDNA viruses termed virophages (Fischer and Suttle 2011). DNA transposons work in a 'cut and paste' manner where the transposase recognises the TIR elements, excises the DNA and reinserts it elsewhere in the genome. The reintegration sites vary based on the family of DNA transposon but in the most common family, Tc1/*mariner* (originally discovered in *Drosophila*) this can be any TA dinucleotide. On reintegration into the genome the target site DNA is duplicated producing sequences known as Target Site Duplications (TSD) which are a hallmark of transposed DNA (Munoz-Lopez and Garcia-Perez 2010).

Current evidence suggests that the DNA transposons within the human genome became inactive around 37 million years ago although the reason for this is currently unknown (Pace and Feschotte 2007). The inactivation of DNA transposons is essential to maintain genome stability as theoretically the expression of only one transposase is required to activate many DNA transposon elements (Padeken, Zeller et al. 2015). The random insertion of transposable elements within promoters, enhancers and gene bodies could easily dysregulate gene expression resulting in any number of developmental defects and diseases. There are exceptions to this where DNA transposons have been repurposed within the genome for essential functions. The best example of this is the RAG proteins which are involved in V(d)J recombination in B-Cell development which are most likely to have originated from DNA transposons (Agrawal, Eastman et al. 1998). The RAG proteins, however, are also a significant source of cancer causing mutations in lymphomas (Lieber 2016).

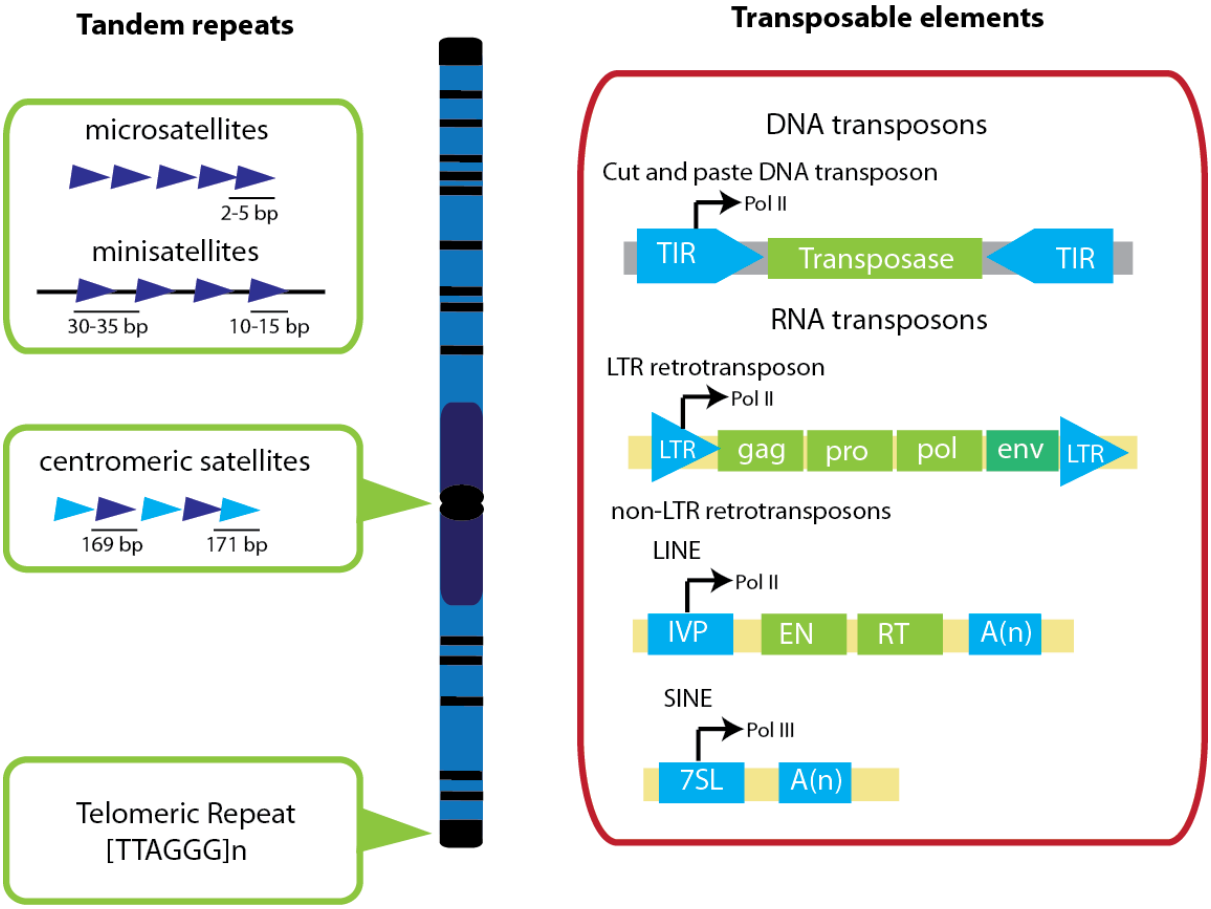


Figure 1.4. Structure of repeat elements of different classes within the human genome (adapted from (Padeken, Zeller et al. 2015)).

1.3.4. Retrotransposons

Unlike the DNA transposons the retrotransposons or RNA transposons carry out transposition through an RNA intermediary. A retrotransposon must first be transcribed and then uses a reverse transcriptase to convert the RNA back into DNA before insertion back into the genome. The endogenous retrotransposons within the human genome are classified based on the presence or absence of long terminal repeat (LTR) elements in their sequence.

1.3.5. Non-LTR Retrotransposons

In humans there are 2 main families of non-LTR retrotransposons: Long Interspersed Elements (LINE) and Short Interspersed Elements (SINE). LINE retrotransposons have an

internal 5' promoter which allows for initiation of transcription by RNAPII expressing a reverse-transcriptase and an endonuclease. The reverse-transcriptase complex recognises the 5' poly-A tail of the element following transcription which allows for transposition to take place (Fig. 1.4). The most abundant LINE in humans is L1 which has ~500,000 copies in the genome and is overall the most abundant transposable element based on genome coverage. The L1 elements are fully autonomous expressing all required enzymes for transposase activity and are the only TE's which have been shown to still have transposon activity within the Human genome. However, only ~100 still retain their complete retrotransposon sequence out of the potential 500,000 (Cordaux and Batzer 2009).

SINE retrotransposons are non-autonomous and only have a 7SL promoter and a poly-A tail (Fig. 1.4). For these elements to become active a reverse-transcriptase is required from an active LINE transposon. In the same way as LINEs, SINEs are recognised by the transposition machinery based on the poly-A tail. The most abundant family of SINEs in humans is the Alu elements totalling ~1 million copies. The Alu family is derived from 7SL RNA and possess the 7SL RNA Pol III promoter sequence (Ullu and Tschudi 1984, Kriegs, Churakov et al. 2007).

1.3.6. Non-LTR Retrotransposons function in the Human genome

Although mostly no longer active L1 and Alu elements played a critical role in inducing genomic variation in early primate evolution mainly through insertional mutagenesis. A small proportion of Alu elements are still active particularly within the germ line, with an estimated 1 in 20 human births having an Alu insertion (Xing, Zhang et al. 2009). Alu insertions can occur within coding regions and close to splice junctions giving them the potential to disrupt gene expression (Deininger 2011). As a result of this Alu retrotransposons are responsible for 1 in every 1,000 new genetic diseases in Humans (Deininger and Batzer 1999). L1 elements are also active within the genome and are required for the Alu elements to function. A number of

studies have shown a significant role for L1 elements within neuronal and brain development and it is suggested that L1 elements contribute to variation between neurons (Muotri, Marchetto et al. 2010, Vogel 2011, Upton, Gerhardt et al. 2015). L1 elements also contribute to genetic disease and this was first discovered in patients with haemophilia A who have an L1 insertion into the *Factor VIII* gene which resulted in the disease in the affected individuals (Kazazian, Wong et al. 1988).

1.3.7. LTR Retrotransposons – Human Endogenous Retroviruses (HERVs)

LTR retrotransposons also often referred to as Human Endogenous Retroviruses (HERVs) account for around 8% of the Human genome (Lander, Linton et al. 2001). Unlike the other transposable elements which often originate from early evolutionary mutations LTR retrotransposons originate from retroviral germ line infections. These infections primarily occurred over 35 million years ago and those which became integrated into the germ line have persisted throughout evolution (Bannert and Kurth 2006). At least a proportion of HERVs have also undergone 'exaptation', through evolution, becoming functional units within the genome (Gould and Vrba 1982, Brosius and Gould 1992).

The basic structure of an LTR retrotransposon is a pair of LTRs flanking 4 ORFs which code for the viral proteins Gag, Pol, Pro and Env (Fig 1.4). The Gag protein forms the central structural core of the virus, Pol codes for a retrotranscriptase and integrase, Pro for a protease and finally Env is a viral envelope protein. This means that a LTR retrotransposon has the ability to form complete viral particles (Bannert and Kurth 2006).

Upon retroviral infection a virion would usually bind to the cell membrane and be transported into the cell. Following entry into the cytoplasm, the viral RNA is then reverse-transcribed by the viral reverse transcriptase enzyme to produce double stranded DNA (dsDNA). This dsDNA is assembled into a pre-integration complex which is imported into the nucleus and is integrated into the host genome by the viral integrase enzyme. The presence of the flanking

LTRs which act as RNAP II promoters allow the viral components to be transcribed by the host cell machinery. Following transcription, some of the RNA is spliced and exported from the nucleus to produce the viral proteins and the remainder is exported to be encapsulated into the virus. In most cases the virion is assembled at the plasma membrane and expelled from the cell in preparation to infect a new host cell. The viral DNA remains integrated within the genome and will be replicated if the cell undergoes mitosis. The production of virions will continue in cells with the integrated viral DNA unless the cell is targeted as part of an immune response or the viral DNA silenced by epigenetic mechanisms (Deininger and Batzer 2002).

1.3.8. Classification of Human Endogenous Retroviruses

Exogenous retroviruses are classified into the family *Retroviridae* which is then further subdivided between 7 genera; alpha-, beta-, gamma-, delta-, epsilon-, lenti- and spuma-retrovirus (ICTV 2017). The taxonomy of endogenous viruses however, was carried out separately to this because many years of evolution mean that many endogenous retroviruses have only a distant relationship to current exogenous viruses. Endogenous viruses are classified into 3 groups, class I which has a resemblance to gamma- and epsilon-retroviruses, class II which is similar to beta-retroviruses and has a distant relationship to delta- and lenti-retroviruses and finally class III which has a similarity to spuma-retroviruses (Bannert and Kurth 2006). From this point the classification becomes much more complicated due to lack of a unified system.

Traditionally the classes are split into families which are named based on the amino acid code of the tRNA which would have originally bound to the primer binding site (PBS) for reverse transcription. Therefore the HERV-K family would denote those with a lysine tRNA PBS (Lavie, Medstrand et al. 2004). The HERVs are then further sub-classified under these groups based on phylogenetic analysis (Polavarapu, Bowen et al. 2006). The downside of

this classification system, other than the incorrect use of the term 'family', is that many of the tRNA annotations are wrong. For example it is now thought that many of the HERV-K family would have associated with methionine tRNA rather than lysine tRNA (Lavie, Medstrand et al. 2004). Regardless of these shortcomings, the most complete classification system is that used by RepBase, a fairly comprehensive database of human repeat elements. The repeats are grouped into 4 classes; ERVL, MaLR, ERV1 and ERVK, each of which is split into numerous families totalling over 500 members (Jurka 2000).

1.3.9. ERVL

The ERVL family was first identified in 1995 in human placenta tissue and was shown to have an integrase domain and also a pol domain which closely resemble that of foamy retroviruses (Cordonnier, Casella et al. 1995). It is now known this family accounts for ~22% of HERVs and is the oldest detectable family at 100-150 Myr (million years) old. It has been shown that ERVL retrotransposons had a significant burst of transposon activity at 45 – 65 Myr ago, however, this has since ceased and no recent activity has been reported (Benit, Lallemand et al. 1999).

1.3.10. MaLR

The MaLR or Mammalian Like Retrotransposons are the largest HERV family (48%) and are very similar to the ERVL family both of which fit within class III (Bannert and Kurth 2006). They were actually described prior to the ERVL family in 1993, when the similarity between a number of human and rodent repeat elements was noted and these were combined to form the new family MaLR (Smit 1993). The family was named based on the absence of any retroviral protein sequence between the LTRs which was likely lost as a result of homologous recombination leaving many solitary LTRs spread throughout the genome (Lander, Linton et al. 2001). The family has a further 10 subfamilies classified based on phylogenetic analysis

(Fig 1.5). Although MaLR elements no longer have coding regions for viral proteins the LTRs still play significant roles in both healthy and diseased tissue (discussed further in 1.3.13).

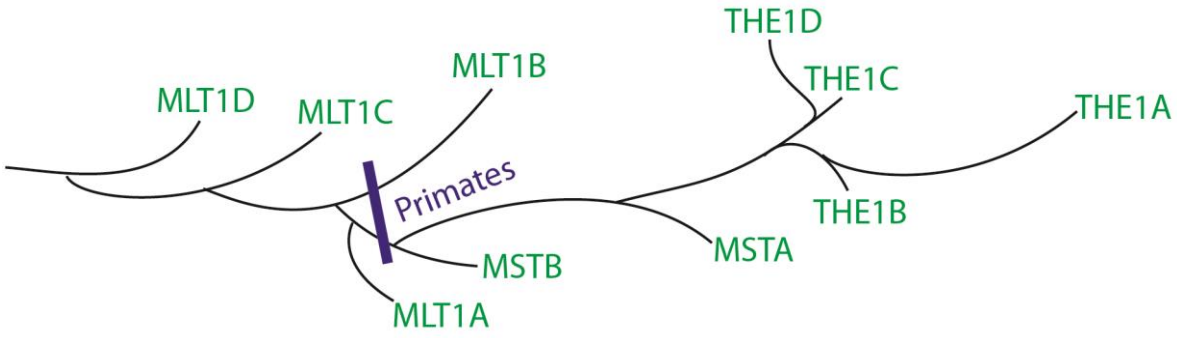


Figure 1.5. Evolutionary development of the MaLR super family of repeats (adapted from (Smit 1993)).

1.3.11. ERV1

ERV1 are class I HERVs and encompass the HERV-H and HERV-F subfamilies. The first HERV-H element was discovered in the β-globin gene cluster where the presence of a pair of LTRs was noted in the sequence and homologous recombination was occurring between the LTRs during cloning experiments (Mager and Henthorn 1984). The ERV1 family has much more recent history than the previously discussed families and can be dated to a common ancestor between the new and old world monkeys around 35 Myr ago (Bannert and Kurth 2006). Although no recent transposition events were reported, an increase in transcription of ERV1 elements has been described in both healthy and diseased tissue. This includes the placenta and lung tissue as well as cancer tissue of the colon, squamous cells and a number of other tumours (discussed further in 1.3.16) (Hirose, Takamatsu et al. 1993).

1.3.12. ERVK

ERVK are class II HERVs and are a family recently integrated into the Human genome with the earliest example dating to ~40 Myr ago. HERVK10 is thought to be the most recently active germ line transposable element of the HERVs with evidence of transposition since the

chimpanzee divergence 7 Myr ago (Medstrand and Mager 1998). The family includes a number of full length copies of retrotransposons spread throughout the human genome which theoretically have the potential to both retrotranspose and produce viral particles (Lower, Lower et al. 1996). It has been shown that at least 6 of the ERVK elements have retrotransposon activity *in vitro* when cloned and also that the Env proteins show activity against the antiviral protein tetherin (Berkhout, Jebbink et al. 1999, Lemaitre, Harper et al. 2014). Although there is no reported viral activity from ERVK elements in healthy tissue a recent study has shown viral-like particles from some teratocarcinoma and breast cancer cell lines are able to infect and become integrated into healthy cells *in vitro* (Contreras-Galindo, Kaplan et al. 2015).

1.3.13. LTRs as Promoter and Enhancers

Most HERVs have lost the ability to code viral proteins through mutation and homologous recombination over millions of years of evolution. This means that many LTRs are now solitary elements within the genome. Because LTRs act as Pol II promoters during their viral function solitary endogenous LTRs have potential to play significant roles as promoters and enhancers.

It was shown using luciferase reporter assays in a number of human cancer cell lines that ERVK LTRs had promoter activity ranging from undetectable in some cell lines to very strong in Tera-1 (a testicular carcinoma cell line). In the same study the enhancer activity of ERVK LTRs was also measured with activity only detectable in the Tera-1 cell line (Ruda, Akopov et al. 2004).

LTRs are split into 3 sections 3’ Unique (U3), Repeat (R) and 5’ Unique (U5) and have a number of shared sequence characteristics. Most LTRs have a 5’ TG and a 3’ CA dinucleotide along with a TATA box often with the sequence TAATAAA and may also include a polyadenylation signal, an initiator motif, a splice site and transcription factor motifs

(Benachenhou, Jern et al. 2009, Benachenhou, Sperber et al. 2013). The transcription start site usual resides at the boundary between the R and U5 regions (Kovalskaya, Buzdin et al. 2006).

LTRs have been shown to have promoter activity and produce transcripts through the binding of TBP to the TATA box present in many LTRs. There are also a number of LTRs, particularly those of the HERVK family which have been shown to not have a TATA box sequence but instead rely on Sp1 and Sp3 proteins binding to GC rich elements to initiate transcription (Fuchs, Kraft et al. 2011). Along with regulation by the main transcriptional machinery, LTRs have been shown to have transcription factor (TF) motifs for many tissue specific transcription factors. This means that LTRs can have tissue specific promoter activity both in healthy and diseased tissue.

The particular transcription factor motifs present within LTRs vary by family. The HERVK family is the best studied of the HERV families and 39 potential TF binding motifs including NF-κB, AP1, POU5F1-SOX2 and GATA have been identified (Bourque, Leong et al. 2008, Manghera and Douville 2013). Of these factors only a small number of factors have been experimentally validated as binding to the HERVK LTRs including YY1, NF-κB, NFAT-1, MITF-M, PR, AR and ER (Manghera and Douville 2013). One study has also identified the tumour suppressor P53 to be bound to many members of the HERVK family and to play a significant role in the P53 regulatory network (Wang, Zeng et al. 2007). An enrichment of P53 binding sites is also seen in members of the ERV1 family (Bourque, Leong et al. 2008). The transcription factor Ying Yang 1 (YY1) is ubiquitously expressed and frequently overexpressed in inflammatory diseases such as cancers (Nicholson, Whitehouse et al. 2011). The binding of YY1 to LTRs has been shown to increase their activation by up to 50% implicating at least HERVK LTRs in inflammatory diseases (Knossl, Lower et al. 1999).

It has been shown that members of the MaLR family of LTRs also have Sp1, AP1, GATA and NF-κB motifs and in the inflammatory environment of Hodgkin lymphoma NF-κB is involved in the activation of LTRs (Lamprecht, Walter et al. 2010). NF-κB is also involved in many other diseases with an inflammatory signature. Along with the regulation of LTRs by various transcription factors there is evidence to suggest that HERVs actually played a critical role in the distribution of transcription factor motifs in the genome throughout evolution (See 1.3.15).

LTR promoters have been shown to act in sense, antisense and as bidirectional promoters (Dunn, Romanish et al. 2006, Huh, Kim et al. 2008, Faulkner, Kimura et al. 2009). As alternative promoters LTRs not only have the potential to simply induce gene expression in a tissue specific manner but also to change splicing of a gene including or excluding exons and producing different isoforms. This process can potentially change the conformational structure of the protein and its function or cellular localisation (Xin, Hu et al. 2008). Most LTRs however, have been shown to generate only subtle deregulatory effects in healthy tissue (Cohen, Lock et al. 2009). This is supported by the theory that LTR promoters have the potential to induce subtle phenotypic effects between individuals due to a population wide mosaic effect in the silencing of LTRs (Whitelaw and Martin 2001). An exception to this is seen in tissue of the placenta where LTRs act as the main tissue specific promoter for certain genes.

1.3.14. HERV role in Healthy Human Tissue

The placenta is one of the best-studied healthy tissues where LTRs act as regulatory elements. HERVs are highly transcribed in the placenta, testis and germ line cells meaning that LTR promoters are more active in these tissues (Cohen, Lock et al. 2009). There are 2 potential theories to explain this high level of activity. It is generally accepted that cells in the early developmental stages have less methylation so this could allow activity of LTRs, which would be supressed in somatic tissue (Fuke, Shimabukuro et al. 2004). However, studies

have suggested that this does not exclusively account for the level of activity (Reiss, Zhang et al. 2007). The other potential reason is that because retroviral infection levels were generally higher in reproductive tissue it is a side effect of the LTRs becoming endogenous parts of the genome (Cohen, Lock et al. 2009).

A number of HERVs are involved in gene expression in the placenta; the most significant of these is located in the *ERVWE1* locus which harbours a complete ERV1 provirus including 2 LTRS, gag, pol and env coding regions. The envelope protein coded for by this HERV has been repurposed to be essential for trophoblast cell fusion and differentiation in the syncytiotrophoblast layer of the placenta (Mi, Lee et al. 2000, Prudhomme, Oriol et al. 2004). There are also a number of other genes within the placenta which use an LTR as a tissue specific promoter. The *P450 Aromatase* gene contains a MER21A LTR from the ERVL family which acts as a placenta specific promoter (Conley and Hinshelwood 2001). The *EDNRB, Pleotorphin* and *Mid1* genes also display placenta specific transcripts from LTR promoters, however these genes also contain active native promoters meaning that the LTRs are just acting to increase expression (Cohen, Lock et al. 2009).

There are a number of other genes which use LTRs as alternate promoters. *Apolipoprotein C-I* is expressed from both its native and an ERV1 LTR promoter in most tissues but is significantly up-regulated from its LTR promoter in the liver (Medstrand, Landry et al. 2001). *AMY1C*, a human salivary amylase gene shows parotid tissue specific expression from an LTR promoter (Ting, Rosenberg et al. 1992). Other tissue specific expression from LTR promoters includes *NAIP* in the testes, *β1,3-galactosyltransferase* in the colon and *Mid1* in the foetal kidney (Landry, Rouhi et al. 2002, Dunn, Medstrand et al. 2003, Romanish, Lock et al. 2007).

Overall it would appear that the most active family of LTRs in healthy tissue is the ERV-9 sub-family which is part of the ERV1 family. The ERV-9 LTRs are unique within the HERV

families as they have TF binding motifs for NF-Y, MZF1 and GATA (Yu, Zhu et al. 2005, Liu and Eiden 2011, Hu, Pi et al. 2017). An ERV-9 LTR forms a novel transcript within the *P63* gene in the testes and germ cell precursors which has been shown to be lost in testicular cancer cells (Liu and Eiden 2011). It has also been shown that solitary ERV-9 LTRs in many human cell types produce both sense and anti-sense transcripts whereby the anti-sense transcript is expressed at a higher level than the sense transcript. The level of the antisense transcript has been shown to drop in many cancer cell lines and when over-expressed in vitro it slows the rate of growth in many cancer cell lines. It is thought that this occurs because the anti-sense transcript acts as a decoy or trap for NF-Y a key factor in cell proliferation (Xu, Elkahloun et al. 2013). Finally, the *β-globin* locus contains an upstream control region, which incorporates an ERV-9 LTR. The ERV-9 LTR acts as a tissue specific enhancer in erythroid progenitor cells priming the expression of globin genes in erythroid cells (Pi, Yang et al. 2004).

As in the case of the β-globin locus LTRs can also act as tissue specific enhancers. As previously mentioned a mild enhancer activity from ERVK LTRs was shown by luciferase reporter assays in the TERA-1 cell line (Ruda, Akopov et al. 2004). The *LEP* gene which codes for leptin in adipose tissue is also known to have an upstream LTR enhancer element which increase expression in the placenta (Bi, Gavrilova et al. 1997). There are few other examples however the identification of enhancer elements tends to be more challenging than promoters as it is much harder to directly associate them with genes which may be many killobases away (Cohen, Lock et al. 2009). With new techniques such as ‘chromatin conformation capture’ it may be easier to build up a picture of HERV enhancer activity which has not yet been identified.

An ERV-9 LTR is also responsible for the upregulation of the GSDML (gasdermin-like) protein coding gene in the alimentary tract, oesophagus, stomach and skin. Rather than acting as a direct promoter for the *GSDML* gene the LTR produces an antisense transcript

which interacts with the native promoter to positively regulate transcriptional activity (Huh, Kim et al. 2008). Many long non-coding RNAs (lncRNA) also originate from ERV-9 LTRs. It has been shown that knockdown of ERV-9 lncRNAs in erythroid cells results in significantly reduced transcription of a number of genes, many of which are erythroid specific. Depletion of these RNAs also resulted in the inhibition of ex vivo erythropoiesis (Hu, Pi et al. 2017).

1.3.15. HERV role in Human Evolution

HERVs have clearly had an effect on the evolution of the human genome as shown by their use as tissue specific promoters for a number of genes. To have remained through millions of years of evolution suggests that the presence of HERVs confers a survival advantage. Currently only a small number of such elements have been linked directly to being required for regulation of gene expression and many are kept in an inactive methylated state (discussed in 1.3.17). This would suggest that many may have been simply retained because of the difficulty in removing them (Nelson, Carnegie et al. 2003).

Using comparative genomics approaches between mouse and human on all transposable elements including ERVs it has been shown that there is strong selective pressure to retain TEs in the genome (Silva, Shabalina et al. 2003, Lowe, Bejerano et al. 2007). It has also been suggested that ERVs are an important source of regulatory elements within the genome as they have significantly increased the proportion of TF motifs to allow for tissue specific regulation (Feschotte 2008). This notion combined with the fact that the majority of examples of LTR-driven expression in healthy tissue only result in subtle changes in expression of already active genes would suggest that ERVs are acting to fine tune gene expression levels (Sverdlov 2000, Wray 2007, Bohne, Brunet et al. 2008). Although HERVs have been retained through evolution and have roles in the genome they have also been linked to number of diseases.

1.3.16. HERVs in Human disease

There are 2 main routes by which HERVs can play a role in disease, through the production of viral proteins or as regulatory elements disrupting gene expression either through activation or repression of a gene. The presence of viral proteins from HERVs has been shown in several inflammatory diseases. In multiple sclerosis it has been shown that HERVs from the ERV1 and ERVK families are active and produce virus like particles which bud from the cell surface (Perron, Garson et al. 1997, Christensen, Dissing Sorensen et al. 1998, Christensen 2005). Although it is not clear whether these are able to infect other cells they do create a pro-inflammatory response which has potential to exacerbate the condition (Clerici, Fusi et al. 1999). Antibodies towards the viral proteins can also be detected in patient samples(Jolivet-Reynaud, Perron et al. 1999). Nucleic acid binding of gag proteins which originate from the HERVs has been shown which could also increase the immune response(Christensen, Dissing Sorensen et al. 2000). Antibodies towards antigens from HERVs have also been detected in the synovial compartment in Rheumatoid Arthritis, another pro-inflammatory disease (Nakagawa, Brusic et al. 1997, Nelson, Lever et al. 1999). There is currently a lack of evidence for roles of HERV proteins in disease development and it is hypothesised that HERV activation may result from immune system activity and via inflammation rather than be causative of the disease (Johnston, Silva et al. 2001). This theory is based on evidence that treatment of cell lines with pro-inflammatory cytokines such as TNF can induce activity in ERVK and ERV1 families (Katsumata, Ikeda et al. 1999, Johnston, Silva et al. 2001).

HERVs have also been linked to many cancers, which also supports their activation by an immune response as many cancers have an inflammatory component. Patients with seminoma, a form of testicular cancer, have been shown to express antibodies against the HERV-K9 gag protein (Sauter, Schommer et al. 1995) and 85% of patients with germ cell tumours have been reported to exhibit antibodies against HERV Env proteins (Sauter,

Roemer et al. 1996). Proteins from the HERVK family have also been identified in melanoma, breast cancer and ovarian cancer (Wang-Johanning, Liu et al. 2007, Golan, Hizi et al. 2008, Wang-Johanning, Radvanyi et al. 2008, Schmitt, Reichrath et al. 2013). In cancers it would seem that the protein coding role of HERVs is a minor player in cancer progression and survival, however, the regulatory role of LTRs has been associated with cell survival in several cancers.

There are a number of routes by which an LTR can act to dysregulate gene expression and lead to cancer. The most straightforward of these is the ectopic or over expression of a gene, which occurs when the LTR is upstream of the genes native promoter and induces expression without modifying the open reading frames (ORFs) of the gene. The first identified example of this which was shown to be required for cancer cell survival was in Hodgkin's Lymphoma (HL). It was shown that the *Colony Stimulating Factor 1 Receptor (CSF1R*) gene was essential for survival of HL cell lines and that it was driven by an upstream THE1B (MaLR) LTR (Lamprecht, Walter et al. 2010). *CSF1R* is normally expressed in myeloid and trophoblast cells from 2 different promoter elements. In myeloid cells expression is strictly regulated by an upstream purine-rich promoter within exon 2 and an intronic enhancer element known as FIRE (Fms-Intronic Regulatory Element)(Himes, Tagoh et al. 2001). The FIRE enhancer contains transcription factor binding sites for PU.1, RUNX1, SP1 and AP1, of which PU.1 is most essential for *CSF1R* expression (Ross, Yue et al. 1998, Bonifer and Hume 2008, Sauter, Bouhlel et al. 2013). PU.1 is expressed in myeloid cells but expression is lost in Hodgkin/Reed-Sternberg (HRS) cells meaning that *CSF1R* cannot be activated by the same regulatory elements as in myeloid cells (Jundt, Kley et al. 2002).

In human placental trophoblasts a second promoter 25 kb upstream has been shown to be active (Visvader and Verma 1989). Based on this Lamprecht, *et al.* (2010) also looked for CSF1R activation from this promoter, however, they showed that in HRS cells *CSF1R* was

expressed from a promoter approximately 6.2 kb upstream. Further study showed that this site corresponded to the sequence of a long terminal repeat (LTR) from the MaLR THE1B family. The *THE1B-CS*F1R transcript was shown to be essential for cell growth and survival in cell lines and to be present in 39-48% of patient samples (Lamprecht, Walter et al. 2010).This region which is normally methylated to prevent activity was shown to have a DNaseI Hypersensitive site in HRS cells which was in contrast to control cells confirming that this was an active promoter which could be initiating the transcription of *CSF1R*.

A follow up study by Babaian, *et.al*. showed that the *IRF5* gene is also expressed in HL from a solitary LOR1a LTR. This finding differs slightly from that of *CSF1R* as *IRF5* also has a low level expression from its native promoter. *IRF5* is known to be upregulated in HL and to be a central regulator of the HL transcriptome (Kreher, Bouhlel et al. 2014). As well as acting as a promoter the LTR insertion upstream of *IRF5* also creates an interferon regulatory factor binding element (IRFE) which factors including IRF5 can interact with. This means that the LTR creates a positive feedback loop strengthening *IRF5* expression (Babaian, Romanish et al. 2016).

LTRs can also be implicated in cancer through expression of truncated proteins, often occurring when the LTR is located within an intron downstream of the canonical promoter. This can result in the loss of regulatory elements within the protein, therefore giving it oncogenic potential. The best example of this is the receptor tyrosine kinase gene, *Anaplastic Lymphoma Kinase* (*ALK*) which carries an alternative LTR16B2 promoter in intron 19 (Wiesner, Lee et al. 2015). This promoter results in expression of 3 different isoforms of *ALK* which lack the extracellular domain but retain the catalytic intracellular tyrosine kinase domain. The LTR promoter is active in 11% of skin cutaneous melanomas and LTR driven isoforms have been shown to increase oncogenic signalling, cell proliferation and also tumour formation in mice. There is also evidence for activation of this promoter in monocyte-derived macrophages, however the function of this transcript is currently unknown.

ALK-negative anaplastic large-cell lymphoma (ALCL) was investigated by Scarfo *et al.* (2016) and it was shown that in 24% of cases cells displayed high expression of *ERBB4* and *COL29A1*. Intriguingly 2 isoforms of *ERBB4* were identified, neither of which originated from the native promoter (Scarfo, Pellegrino et al. 2016). Both transcripts were shown to originate from MaLR family LTRs and it was noted that two thirds of the samples had a 'Hodgkin-like' morphology which is only usually seen in 3% of ALCL cases. Both isoforms are thought to be potentially oncogenic and mutations in the *ERBB4* gene have previously been associated with other cancers (Scarfo, Pellegrino et al. 2016).

Finally, an additional direct way in which an LTR promoter has been shown to drive cancer is through the formation of chimeric proteins where non-coding DNA becomes fused to downstream exons. This has the potential to form proteins with significantly altered structure and function. In 5% of Diffuse Large B-Cell Lymphomas (DLBCL), *Fatty acid binding protein 7* is expressed as a fusion to an anti-sense LTR2 element. The chimeric protein which is produced has an altered N-terminus and has been shown to be required for optimal cell growth and cellular localisation in DLBCL cell lines (Lock, Rebollo et al. 2014).

Many active solitary LTRs also produce long non-coding RNA (lncRNA) transcripts which can interact with other genes (Babaian and Mager 2016). A study in hepatocellular carcinoma identified a subset of ncRNAs expressed by LTRs which were 10-fold upregulated in most tumour samples suggesting that they may play a significant role in liver cancer pathogenesis (Hashimoto, Suzuki et al. 2015). The *SchLAP1* lncRNA is known to be overexpressed in 25% of prostate cancer cases and originates from an ERV-9 LTR *SchLAP1* has been shown to inhibit the function of the SWI/SNF complex which can act as a tumour repressor. This supports the observation that *SchLAP1* overexpression is associated with cell survival and proliferation in prostate cancer (Prensner, Iyer et al. 2013, Masliah-Planchon, Bieche et al. 2015).

Non-coding RNAs can also produce antisense transcripts of genes, for example a THE1A LTR promoter produces an anti-sense transcript of the *AFAP1* gene. Changes in expression of AFAP1 have been associated with several types of cancer including HL. The antisense transcript of *AFAP1* results in a loss of gene expression (Babaian and Mager 2016). It is thought that down-regulation of AFAP1 results in up-regulation of RhoA/Rac2 signalling which increases cell proliferation (Zhang, Weng et al. 2016).

Finally LTRs have been shown to play a role in cancer through enhancer activity which is probably the least studied LTR function due to the challenges of linking enhancers to genes. In prostate cancer hypomethylation of many HERVs is seen although only a specific subset appear to produce transcripts (Goering, Ribarska et al. 2011). It is known that the *KLK3* gene encoding a prostate specific antigen is highly androgen sensitive and is located in a cluster of androgen responsive genes. It has been shown that the cluster is regulated by an upstream region which is formed of an LTR40a element. The LTR40a acts as an androgen responsive element, particularly when it has a specific duplication in the LTR sequence, and this upregulates expression of the surrounding cluster of androgen responsive genes (Lawrence, Stephens et al. 2012).

1.3.17. Regulation of HERVs

It is now clear from multiple studies that the presence of active HERVs in the genome has the potential to contribute to multiple diseases. For HERVs to have remained throughout evolution, cells must have developed strict control mechanisms to avoid negative selection and for the overall maintenance of genome stability (Glinsky 2015). It has long been known that CpG methylation is largely directed towards transposable elements within the genome (Yoder, Walsh et al. 1997). Methylation is mostly removed from the genome during sexual reproduction but is replaced early in embryonic development and was originally thought to permanently silence transposable elements (Altun, Loring et al. 2010).

We now know that methylation of retrotransposons within the human embryo employs more complicated developmental mechanisms with only a fraction of LTRs being hypermethylated at the preimplantation stage. A transition was observed following fertilisation with MaLR family LTRs becoming methylated and down-regulated and ERV1 and ERVK elements becoming demethylated and having up-regulated expression (Smith, Chan et al. 2014). The same study showed that following implantation a stable LTR methylation programme is established which remains into adult somatic cells. It has also been suggested that the loss and rewriting of methylation during the early developmental stages is not perfect and produces a mosaic pattern which varies between individuals of the same species therefore contributing to subtle variation in phenotype (Whitelaw and Martin 2001). It is also possible that in some cases the methylation pattern of certain ERVs could cross the germ line being inherited across generations (Chong, Vickaryous et al. 2007).

Loss of CpG methylation has been shown in multiple studies to result in HERV activation often contributing to disease states. It has been shown that expression of HERK LTRs in the human teratocarcinoma cell line, TERA-1, is differentially regulated based on the level of methylation (Lavie, Kitova et al. 2005). This same finding was reproduced in melanoma cell lines (Stengel, Fiebig et al. 2010). The MaLR family of LTRs are also regulated by CpG methylation in Hodgkin's lymphoma (HL). Both the THE1B LTR found upstream of the *CSF1R* gene and the LOR1A LTR upstream of the *IRF5* gene show a loss of methylation at CpG element when compared to human B-cell lines (Lamprecht, Walter et al. 2010, Babaian, Romanish et al. 2016).

There are very few studies which investigate the causes of demethylation to induce LTR activation in humans, however, there is evidence that a hypomorphic allele of DNA methyltransferase-1 (*Dnmt1*) results in activation of ERVs and leads to T-cell lymphomas in mice (Eden, Gaudet et al. 2003, Howard, Eiges et al. 2008). A mechanism has been proposed for the demethylation seen in HL. Surprisingly in HL there is no loss of the DNA

methylating enzymes, however it was shown that there is a loss of the transcriptional repressor ETO2. ETO2 (also called CBFA2T3) acts via interaction with histone deacetylases (HDACs) (Hug and Lazar 2004). When ETO2 is not expressed demethylation is seen at the CpG element within the THE1B LTR which causes activation of the *THE1B-CSF1R* transcript in HL. It was further shown that when combined with NF-κB constitutive activation, a central feature of HL, a strong activation of the LTR was achieved (Lamprecht, Walter et al. 2010). This also highlights the role that tissue specific transcription factors play in activation of ERVs.

An alternative repressive mechanism that has been proposed for HERV silencing is Krüppel-associated box zinc finger proteins (KRAB-ZFP) (Wolf, Greenberg et al. 2015). For reverse transcription of a retroviral genome to take place a tRNA primer sequence is required to prime the minus strand synthesis. In mice the KRAB-ZFP protein ZFP809 has been shown to bind to the tRNA primer binding site (a sequence required for the priming of the minus strand during reverse transcription) of murine leukaemia virus in stem cell and recruit corepressors KAP1 and SETDB1 which, in turn, induces H3K9me3 (Wolf, Yang et al. 2015). It was further shown that ZFP809 and KAP1 are not required to maintain repression of ERVs in adult somatic tissues, however, SETDB1 is required in some differentiated cell types (Wiznerowicz, Jakobsson et al. 2007, Fasching, Kapopoulou et al. 2015). There is also evidence to suggest that in humans ERVs of the ERVK and ERV1 families are repressed by KRAB-ZFPs, although follow up validations are yet to be performed (Turelli, Castro-Diaz et al. 2014, Wolf, Greenberg et al. 2015). A recent study has also suggested a requirement for Histone H3.3 in ERV silencing in embryonic stem cells, however the significance of the role of H3.3 is debated (Elsasser, Noh et al. 2015, Elsasser, Noh et al. 2017, Wolf, Rebollo et al. 2017).

1.3.18. Methods for Identifying Active HERVs

The study of active HERVs and particularly the solitary LTRs poses a number of challenges, particularly for genome wide bioinformatics analysis. The nature of repeat elements having little or no unique sequence makes them intrinsically hard to align meaning that they often get overlooked in genome wide screens such as RNA-Seq. A number of lab based and bioinformatics approaches have been developed to overcome this issue.

The use of 5' Rapid Identification of cDNA ends (RACE) has long been used to validate individual transcripts originating from LTRs. It uses a known tagging sequence ligated to the 5' end of the transcript and a primer within the transcript to allow for amplification. RACE however does not allow for the full complement of active LTRs to be determined. Capped analysis of gene expression (CAGE) has been used in a number of studies and is a technique which makes use of the 5' cap (modified nucleotide) which is present on mature mRNA transcripts. The mRNA transcripts are captured on biotinylated beads using the 5' cap and digested to form short fragments representing the transcription start site. These fragments can then be sequenced and mapped back to the genome (Hashimoto, Suzuki et al. 2015, Babaian, Romanish et al. 2016). This method produces a map of all transcription start sites across the genome which would therefore identify LTRs which were acting as transcription start sites. Although CAGE only produces 27 bp reads the identification of LTR promoters is likely to be better than RNA-Seq as the level of background should be far lower.

A significant challenge is the ability to identify active LTR transcripts in sequencing data which has not been produced specifically for this purpose. The lack of unique mapping to LTR sequences poses a significant problem in the accurate estimation of enrichment of a particular element. This is because current alignment approaches either discard multiple aligning reads or assign them randomly to one of the locations, meaning that many reads at a particular element may be lost. This is particularly a problem with short sequencing read

data from experiments such as ChIP-Seq and DNaseI-Seq. It has been proposed that the estimation of enrichment can be improved by mapping the multiple aligning reads to repeat sub-families defined by RepBase and combining that enrichment information with the uniquely mapping tags on the LTRs on that family (Day, Luquette et al. 2010).

Paired end RNA-seq data helps to alleviate the problem of lack of unique LTR sequence to align by overlapping splice junctions and retrieving sequence from a gene which an LTR may be spliced to. As most LTR promoters will produce unannotated isoforms the challenge with this method is the requirement to identify unique splice junctions when aligning the RNA-Seq data. A technique has been developed which helps to overcome this problem by assigning the LTR sequences to a virtual chromosome composed of genome specific LTR sequences (Sokol, Jessen et al. 2016). Following alignment, paired reads are filtered for those where one read of the pair maps to an LTR within the artificial chromosome. These reads can then be mapped back to the reference genome using the BLAT algorithm (Kent 2002). A similar technique has also been used successfully to identify active LTRs in other studies (Lock, Rebollo et al. 2014).

1.4. B-Cell Development

The B cell lineage originates from haematopoietic stem cells (HSC) located within the bone marrow, which are able to self-replicate and differentiate into all blood cell types (Fig. 1.6). In the initial stage of differentiation the HSC first develops into a multipotent progenitor (MPP) still with the potential of differentiating into all blood cell types but losing its self-renewal abilities. From this population of cells develop cells which differentiate towards the lymphoid lineage known as Lymphoid-Primed multipotential progenitors (LMPP). At this point cells have lost the ability to become megakaryocytes or follow the erythroid lineage but still have potential to become either myeloid or lymphoid cells. The main deciding factor in how differentiation of these cells proceeds is the level of transcription factor PU.1. It has been shown that a high level of PU1 at the LMPP stage results in cells committing to the myeloid lineage whereas a low level results in commitment to the lymphoid lineage (DeKoter and Singh 2000, Leddin, Perrod et al. 2011). This being said, the presence PU.1 is also required for the lymphoid lineage since it has been demonstrated that knock-outs result in a differentiation block at the MPP stage (McKercher, Torbett et al. 1996) (Fig. 1.6).

In the following stage the cells become committed lymphoid progenitors (CLP), which lose the ability to become myeloid cells and have the ability to differentiate into either T cells or B cells. A number of transcription factors regulate this transition, including E2A (Herblot, Aplan et al. 2002). The most essential of these transcription factors for B cell development is PAX5, which is regulated by E2A, Ebf1 and FOXO1. PAX5 is thought to be responsible for the activation of the main B cell factors and repression of non-B cell genes (Nutt, Heavey et al. 1999). Once fully committed to the B cell lineage (pro-B cells), cells begin to express B220 (an isoform of CD45) but at this stage still do not express cell surface immunoglobulin (Ig) (Bauer, Rodiger et al. 2005). The main functional component of B cells, which is required for their survival, is the B Cell Receptor (BCR). The BCR is a complex including the Immunoglobulin and CD79 subunits (Duncan, Webster et al. 2010), which is responsible for

binding to antigens as part of the humoral immune system. To achieve adaptive immunity to a wide range of different antigens the immunoglobulin genes undergo a large amount of genetic rearrangement to produce many different variable sites. The basic immunoglobulin protein is made up of 2 identical heavy chains and 2 identical light chains with the 2 heavy chains linked together and one light chain attached to each heavy chain.

To achieve a diverse range of antigen receptors a process known as V(D)J recombination occurs, resulting in the rearrangement of the variable (V), diversity (D) and joining (J) regions (Dudley, Chaudhuri et al. 2005). During this process the D and J segments are first recombined and are then attached to a V segment. Once this has occurred, a complex is formed with the light chains which can be one of two isotypes, either kappa or lambda. The light chains also undergo recombination of the V and J segments, once combined with the heavy chains this forms an immunoglobulin known as IgM or IgD (Geisberger, Lamers et al. 2006, Kirkham 2014) (Fig. 1.6).

Once V(D)J recombination is complete and the BCR is presented on the surface of the B cell, it then leaves the bone marrow and moves into the blood stream where it travels to the spleen to undergo further maturation (Cariappa, Chase et al. 2007). Within the spleen further selection is carried out and B cells can either develop into marginal zone B cells or follicular B cells (Reviewed by (Pillai and Cariappa 2009)). Follicular B cells recirculate through the secondary lymphoid organs awaiting activation by a specific antigen. Marginal zone B cells are selected based on the expression of less specific BCRs, which will react to many different antigens. The marginal zone B cells then inhabit regions of the spleen, which interface with the blood and are able to rapidly produce plasma cells on activation (Mackay and Browning 2002, Cerutti, Cols et al. 2013). The follicular B cells then remain in the spleen and lymph nodes until presented with an antigen by a T-helper cell. At this point the cell migrates into the dark zone of the germinal centre where it is known as a centroblast and undergoes rapid proliferation (Bannard, Horton et al. 2013). Activation also results in

upregulation of activation induced cytidine deaminase (AID) by PAX5 and E2A (Kirkham 2014). AID results in point mutation in the variable region of the immunoglobulin known as somatic hypermutation. Following this the cells migrate to the light region of the germinal centre where they are known as centrocytes. In the light region the cells are selected based on whether the mutation is advantageous or not. Within the light region there are T-helper cells, which present the antigen to the newly mutated immunoglobulin, based on this binding affinity the cells then undergo apoptosis or class switch recombination swapping IgM or IgD for IgA or IgG. More recent imaging data suggest that the processing of B cells within the GC is slightly more complex with cells cycling between the light and dark zones and possibly with higher affinity Ig removing antigens from other lower affinity Ig (Allen, Okada et al. 2007, Kuppers 2009). Following this selection process, the highest affinity cells will either form long-lived plasma cells which produce large quantities of antibodies, which are secreted into the blood or memory B cells which display high affinity binding to the specific antigen so that a fast immune response is possible in the future if the immune system is again exposed to the same antigen (Fig. 1.6).

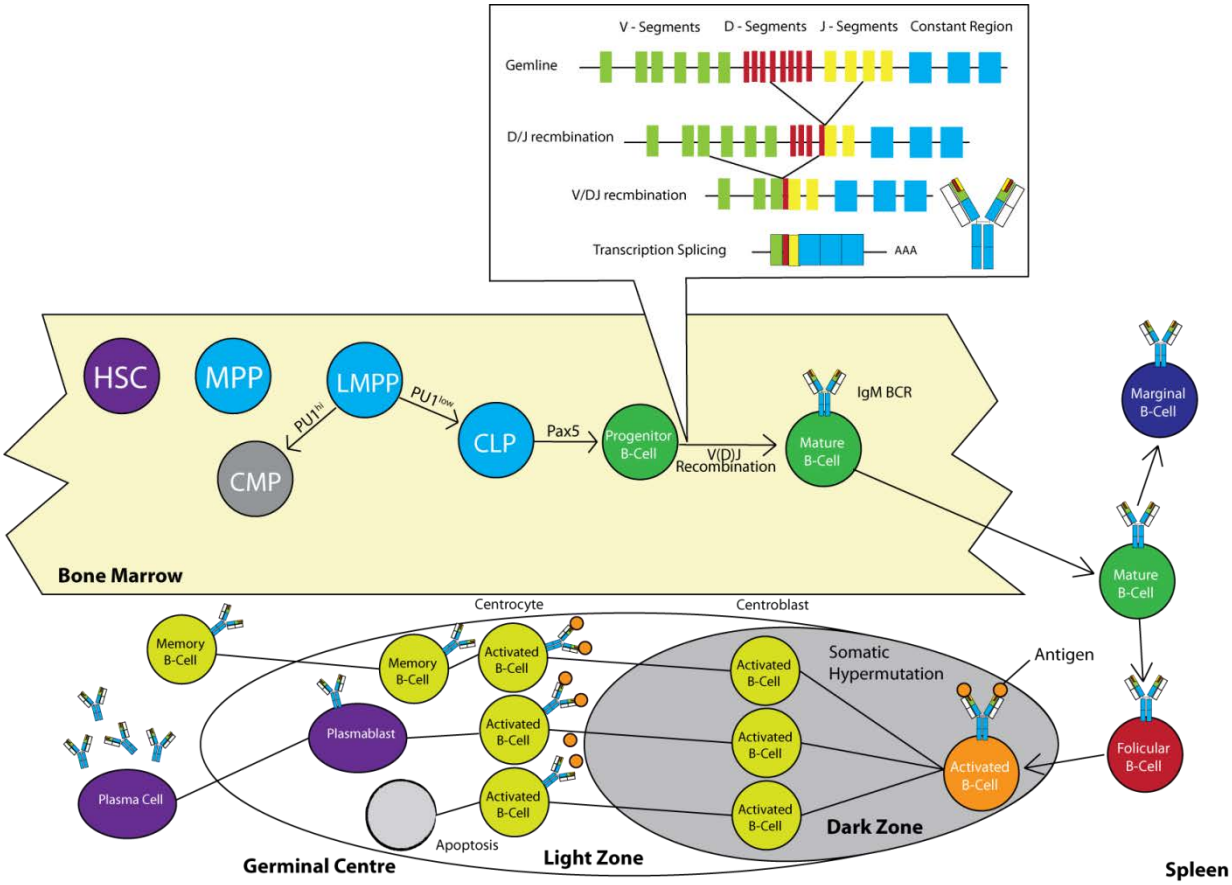


Figure 1.6. Simplified B Cell development pathway from undifferentiated HSCs in bone marrow through V(D)J recombination of Immunoglobulin and to activation by an antigen in the spleen, somatic hypermutation and production of antibodies as part of the humeral immune response. Adapted from: (Mackay and Browning 2002, Klein and Dalla-Favera 2008, Kirkham 2014).

The V(D)J recombination and somatic hyper-mutation processes have to be strictly controlled because there is a high risk of developing unfavourable mutations in the B cell genome resulting from the formation of double strand breaks. Once developed, a homeostatic balance of follicular B cells has to be maintained in preparation for antigen recognition. Cell survival is maintained by a combination of BCR and BAFF signalling. BAFF is a ligand of the tumour necrosis factor family (TNF) has also been shown to play an important role in B cell survival (Mackay and Browning 2002, Mackay, Figgett et al. 2010). It has been shown that mature B cells are not present in BAFF knockout mice and it is thought that BAFF interacts with the BCR signalling.

Within the germinal centre apoptosis of cells with lower antigen binding affinity occurs through the FAS pathway. At the GC stages of B cell development the cells are destined for apoptosis throughout the process and are saved by positive selection by high affinity binding to an antigen and signalling involving CD40 and CD154 (Guzman-Rojas, Sims-Mourtada et al. 2002).

1.5. Hodgkin's Lymphoma

Hodgkin's Lymphoma (HL) is split into two main sub-types, Classical Hodgkin's Lymphoma and Nodular Lymphocyte-Predominant Hodgkin's Lymphoma (NLPHL)(Kuppers 2009). Classical Hodgkin's Lymphoma affects 2.8 per 100,000 people in the US and 3.0 per 100,000 people in the UK. It is predicted that in the UK the overall likelihood of developing HL at some point during a person's lifetime is 0.2%. The age at which classical HL has the highest rate of occurrence has two peaks at 20-24 years old and again at 70-79 years old, particularly in men(CRUK 2013). A survival rate of 83.25% at 5 years and 77.65% at 10 years is reported in the UK and this figure has risen steadily over the past 30 years (CRUK 2012). Classical Hodgkin's Lymphoma usually presents as a tumour mass on lymph nodes but can also occasionally be found in the spleen, bone marrow, liver, and lungs (Connors 2009). NLPHL is much less common only accounting for around 5% of Hodgkin's lymphoma cases. Current treatment for early stage Hodgkin's lymphoma is four to six rounds of ABVD chemotherapy (Adriamycin, bleomycin, vinblastine and imidazole carboxamide). In cases with predicted poorer clinical outcome this is complimented with radiation therapy (Derenzini and Younes 2011).

Hodgkin's Lymphoma was first described by Thomas Hodgkin in 1832 who reported cases of lesions on the 'absorbent glands' (now known as the lymph nodes) and spleen, later named Hodgkin's Disease (Hodgkin 1832). At a cellular level the main identifying features of Classical HL are the mononucleated Hodgkin and multinucleated Reed-Sternberg (HRS) cells identified by Dorothy Reed (1902) and Carl Sternberg (1898) (Kuppers and Hansmann 2005)(Fig. 1.7). It has since been shown (see 1.5.2) that these cells are most likely derived from B cells therefore the disease was renamed Hodgkin's Lymphoma. The predominant cells in NLPHL are known as lymphocytic and histiocytic (L & H) cells and have very different expression patterns to HRS cells (Nogova, Rudiger et al. 2006). The characterisation of the cells in both classical HL and NLPHL is problematic as in most cases they only account for

around 1% of the tumour bulk with the majority of the tumour comprised of an infiltrate of other immune cells (Kuppers 2009).

Based on variation in the cellular composition of classical HL it has now been split into 4 sub-types. These are nodular sclerosis, mixed cellularity, lymphocyte depletion and lymphocyte rich (Kuppers and Hansmann 2005).

Figure 1.7. Bi-Nulceated RS cell surrounded by lymphocytes (NCI 2008).

1.5.1. HRS cells

HRS cells are the main identifying feature of classical HL. Phenotypically they have a very visible nucleolus in each nucleus and can be over 50µm in diameter (Kuppers and Hansmann 2005). The most commonly used expression marker to identify HRS cells is CD30 a member of the tumour necrosis factor (TNF) receptor family. CD30 is also the hallmark of a number of other lymphomas including Anaplastic Large Cell lymphoma (de

Leval and Gaulard 2010). The gene expression patterns of HRS cells do not resemble any other hematopoietic cell, expressing genes from a wide range of lineages making identifying the origin of these cells challenging.

1.5.2. Origin of HRS cells

The combination of the unusual expression pattern and the scarcity of the cells within tumour tissue mean that the B cell origin of HRS cells was long debated. In 1994, Kuppers, *et al.* demonstrated by micromanipulation of cells from histological sections that HRS cells were likely to have originated from a range of differentiation stages in the B cell lineage.

The B cell origin of most HRS cells was eventually confirmed by the analysis of immunoglobulin genes which showed rearrangement consistent with V(D)J recombination (Kuppers and Hansmann 2005). It was also shown that cells in each case were of monoclonal origin (a common feature of tumour cells), with identical Ig variable region rearrangements (Kuppers and Hansmann 2005). There was still debate as to whether this confirmed the B cell origin of HRS cells as T-cell receptors (TCR) also undergo similar recombination. However further study by (Kanzler, Kuppers et al. 1996) showed that the rearranged V regions had high levels of somatic mutation consistent with those occurring in the Germinal Centre (GC) confirming the B cell origin. For this reason it is also thought that in the majority of cases HRS cells originate from GC B cells or Post-GC B cells. It was also shown that in 25% of cases the somatic mutation resulted in non-functional immunoglobulins through the introduction of stop codons (Kuppers 2009). This suggests that HRS cells originate from pro-apoptotic GC B cells, which for some reason have avoided apoptosis. It is further suggested that as mutations resulting in non-functional immunoglobulins are not detectable in HRS cells from many patients that in the majority of cases HRS cells are probably cells which were selected for apoptosis. It is still thought that in rare cases HRS cells could originate from T-cells, as they have been shown to express the T-cell Receptor

and lack Ig mutations (Kuppers 2009). It should however be noted that these cases may not be Hodgkin's Lymphoma and instead T-cell lymphomas which are expressing the markers usually used to identify HL, CD30 and CD15 (Barry, Jaffe et al. 2003).

1.5.3. HRS Cell Development

If HRS cells develop from B cells marked for apoptosis this must mean that they somehow manage to escape the apoptosis pathway. To achieve this escape anti-apoptotic or pro-survival signals are necessary along with suppression of differentiation. A number of transcription factors which are important for the differentiation of B-cells are downregulated including OCT-2, SPI1 and EBF1 (Stein, Marafioti et al. 2001, McCune, Syrbu et al. 2006, Bohle, Doring et al. 2013) resulting in the downregulation of many B-cell specific genes and therefore creating a block in differentiation (See 1.5.5 HRS cell expression patterns). To survive within the germinal centre HRS cells have to avoid their predisposition to apoptosis through the FAS pathway. Through the expression of the FAS cell surface receptor germinal centre B-cells are predisposed to apoptosis on interaction with the Fas ligand which is present on CD4+ Th1 cells (Rothstein 2000, Mizuno, Zhong et al. 2003). This predisposition should prevent cells with unfavourable mutations such as auto-reactivity from developing further. The most obvious route for this to be overcome in HRS cells is through *FAS* gene mutations; however studies have shown that these mutations are very rare therefore the cells must survive by other routes (Maggio, Van Den Berg et al. 2003). HRS cells express c-FLIP, an anti-apoptotic gene, which is usually only expressed in B cells with high affinity antigen binding and is a potent inhibitor of FAS activity (Thomas, Kallenborn et al. 2002). It is therefore possible that this may assist to prevent apoptosis of HRS cells, however the exact signalling route in the absence of the BCR is unclear (Thomas, Re et al. 2004). It has also been suggested that the significant deviation of HRS cells from the B cell gene expression program may also play a role. Mimicking other cell types such as granulocytes with CD15 expression and T-cells with CD30 expression may prevent the cells from being recognized

as GC B cells (Thomas, Re et al. 2004). This helps the cells to evade immunological surveillance within the germinal centre.

To survive, HRS cells also require growth signals, which would usually originate from the BCR. The full complement of factors involved is not fully understood, however, NF-κB and the JAK/STAT pathway have been suggested to play a major role (Liu, Sattarzadeh et al. 2014). Expression of STAT3, STAT6 and STAT5a has been demonstrated in HRS cells which all have involvement in maintaining cell survival signals (Thomas, Re et al. 2004). Colony Stimulating Factor 1 Receptor (CSF1R) has also been shown to play a vital role in HRS cell survival (Lamprecht, Walter et al. 2010). It was shown that in the absence of CSF1R HRS cells were unable to survive making this another critical factor for survival without the BCR.

1.5.4. NF-κB Expression in HRS Cells

In 1997 Bargou, *et al.* showed that constitutively active NF-κB is essential for survival and proliferation of HRS cells. NF-κB is primarily a transcription factor activated as part of an inflammatory response. NFkB can become activated through 2 pathways the canonical pathway and the alternative pathway. The canonical pathway is activated by microbial products or inflammatory cytokines and results in the activation of RelA and RelC NF-κB complexes (Karin and Ben-Neriah 2000). Activation of this pathway results in the phosphorylation and ubiquitination of IkB by IkKβ and IkKγ which allows NF-κB to translocate to the nucleus and become active. The alternative pathway is activated by TNF-family cytokines and signalling. The alternative pathway functions through the phosphorylation of p100 which is a precursor to p52 (Lawrence 2009). Once activated p52 and RelB localise to the nucleus allowing NF-κB to activate expression of target genes.

There have been many studies attempting to determine the cause of NF-κB activation and there appears to be a lot of variation between HL cases (Kuppers, Engert et al. 2012). In a

study 44% of HL cases were shown to have a mutation of the *TNFAIP3* gene, which codes for protein A20 (an inhibitor of NF-κB) (Schmitz, Hansmann et al. 2009). This protein was also shown to function as a tumour suppressor as reactivation in HL cell lines resulted in impaired survival (Brauninger, Schmitz et al. 2006). In 30% of HL cases an increased level of REL expression has been detected resulting in an increased activation of NF-κB (Martin-Subero, Gesk et al. 2002, Kuppers, Engert et al. 2012). Other factors increasing NF-κB activation include activation of the alternative NF-κB pathway through NIK upregulation and mutations in the NκB inhibitors IκBα and IκBε (Kuppers, Engert et al. 2012). In most cases several of these pathways are active which suggests that strong NF-κB activity is needed for HRS cell survival (Kuppers, Engert et al. 2012).

1.5.5. HRS Cell Expression Patterns

As well as the factors required for HRS cell survival, these cells have a unique gene expression pattern unlike any other immune or blood cell. Firstly, there is an overall down regulation of expression of B cell specific transcription factors including OCT2, PU.1 and BOB1 which leads to down-regulation of their target genes (Stein, Marafioti et al. 2001, Torlakovic, Tierens et al. 2001). There is also increased expression of ectopic factors such as NotchI (a T-cell factor) and ID2 (usually found in NK cells) (Thomas, Re et al. 2004). Upregulation of non-B cell factors further down-regulates B cell factor genes. The ID2 protein acts as an inhibitor of E2A which contains E12 and E47 both of which are essential transcription factors in B cell development (Mathas, Janz et al. 2006). It is also suggested that ID2 may play a role in inhibition of PAX5 which has a significant role maintaining the B cell phenotype (Renne, Martin-Subero et al. 2006). As well as the expression of genes related to survival and development, HRS cells are also known to express high levels of chemokines which are responsible for attracting the infiltrate of immune cells which makes up the majority of the tumour mass. The most significant chemokine expressed is CCL17 (TRAC) which attracts Th2 cells making up the main infiltrate (Liu, Sattarzadeh et al. 2014).

Added to this are also extra chemokines produced by the cells in the microenvironment around the HRS cells, which further increases immune cell infiltrate. Finally, the HRS cells also produce other cytokines, which contribute to the tumour microenvironment, particularly migration inhibitory factor (MIF), which attracts M2 macrophages to the tumour (Liu, Sattarzadeh et al. 2014).

1.5.6. EBV

Epstein Barr Virus infection is seen in cells in around 40% of HL cases and has been shown to be a major initiating factor (Kapatai and Murray 2007). It was initially thought that maybe an earlier but undetectable EBV infection may have resulted in non-EBV HL cases, however, further studies suggest that there is no evidence of this (Staratschek-Jox, Kotkowski et al. 2000). The presence of EBV within HRS cells results in the expression of the viral proteins EBNA1, LMP1 and LMP2a (Kuppers 2009). LMP1 mimics CD40 which is able to induce NF-κB expression, aiding the development and survival of HRS cells (Kilger, Kieser et al. 1998). It is also suggested that LMP2a can replace the signalling of the BCR which also promotes survival allowing cells the escape apoptosis and become HRS cells (Caldwell, Wilson et al. 1998).

1.5.7. CSF1R

Expression of Colony Stimulating Factor 1 Receptor (CSF1R) has been shown to have an important role in the survival of HRS cells (previously discussed in 1.5.3) (Lamprecht, Walter et al. 2010). CSF1R is encoded by the *FMS* gene and its expression is usually restricted to myeloid cells but has also been demonstrated in cells of the female reproductive tract and neuronal cells (MacDonald, Rowe et al. 2005, Bonifer and Hume 2008, Droin and Solary 2010). It is a tyrosine kinase receptor for the ligand CSF-1 but has also been shown to bind to a new ligand IL-34 (Droin and Solary 2010). CSF-1 is responsible for regulating survival, proliferation and differentiation in mononuclear phagocytes and osteoclasts and also has a

role in fertility (Droin and Solary 2010). CSF1R and CSF-1 have been implicated in the pathology of a number of cancers, in particular, prostate cancer, breast cancer and leukemias (Ide, Seligson et al. 2002, Aikawa, Katsumoto et al. 2010, Morandi, Barbetti et al. 2011). In most cases it was shown that upregulation of CSF-1 results in increased proliferation within cancer cells. In several cases it is also suggested that an autocrine loop is involved whereby the cells express both CSF1R and CSF-1 resulting in a rapid proliferation of cells (Patsialou, Wyckoff et al. 2009). This autocrine loop has also been suggested to play a major role in the proliferation and survival of HRS cells. Lamprecht, et al. (2010) demonstrated that the expression of CSF1R was essential for the survival of HRS cells. They also went on to show expression of CSF-1 in HRS cells supporting the theory of an autocrine loop. Interestingly (Ingram, Valeaux et al. 2011) showed that in B cells the expression of CSF1R is further supressed by PAX5 which binds to the *CSF1R* promoter and an enhancer (Tagoh, Ingram et al. 2006). This is however less relevant in HRS cells as PAX5 is rarely expressed and Lamprecht et al. (2010) confirmed that *CSF1R* is activated by an upstream THE1B LTR not the canonical promoter (discussed in 1.3.16).

1.6. Aims and Objectives

Lamprecht *et al.* (2010) demonstrated that the activation of a THE1B LTR in HL acted as a promoter for the *CSF1R* gene which is required for the survival of HL cells. They also showed that LTR activation in HL resulted from a loss of epigenetic control due to down-regulation of *CBFA2T3* and the constitutive activation of NF-κB. Finally they presented preliminary results indicating that LTR activation may be a genome-wide phenomenon. Based on these data the aims of this thesis were to:

1) **Investigate the impact of the Hodgkin's Lymphoma specific transcriptional network on chromatin structure and gene expression.**

   a) Microarray data have previously been published for the L428, L1236 and KM-H2 HL cell lines. However, microarray data lack important information such as the expression of splicing variants and low abundance transcripts, as well as alternative transcription start sites. Therefore we plan to generate high-quality RNA-Seq data to enable us to study HL-specific gene expression patterns and alternative transcription start sites, in particular those residing within LTRs.

   b) Previous work mapping DNaseI HS sites at low resolution has shown the involvement of a number of inducible transcription factors driving the HL gene expression program, including IRF, NF-κB and AP-1. Here we aim to extend this work by producing high resolution DNaseI data which can be used for digital foot-printing to determine the occupied DNA binding motifs present within regions showing transcription factor binding. This will allow us to determine the driving factors of the HL gene expression program.

2) **Determine the global activation pattern of Long Terminal Repeat elements within HL cell lines.**

   We aim to develop a next generation sequencing technique (RACE-Seq) based on 5' Rapid Amplification of cDNA Ends (RACE) to map the global activation of THE1B LTRs within HL and non-HL cell lines. The LTR activation patterns of HL and non-HL cell lines will be compared to determine a HL specific pattern of LTR activation.

3) **Elucidate the impact on gene expression of long terminal repeat activation in Hodgkin's Lymphoma.**

   By integrating the RACE-Seq and RNA-Seq data we will investigate the genome-wide impact of LTRs as alternative promoters and enhancers in HL. We also aim to investigate whether other genes besides *CSF1R* are LTR-driven. We aim to identify specific target genes which are up-regulated as a result of LTR activation and assess the impact of these on the HL phenotype.

4) **Explore the impact of inflammation driven LTR activation on the control cell line, Reh.**

   The HL phenotype displays a significant inflammatory gene expression signature which is caused by the constitutive activation of NF-κB, MAPK-signalling via AP-1 factor family members and a number of other inflammatory pathways. Constitutive NF-κB and AP-1 activity has also been shown to drive LTR promoter activity.

Firstly, we will evaluate the impact of inflammatory stimuli by the treatment of Reh cells with phorbol 12-myristate 13-acetate (PMA) which is known to activate NF-κB, MAPK and a number of other pathways through PKC signalling. We will then investigate LTR activation by RACE-Seq and examine its impact on gene expression using RNA-Seq. Finally we aim to study the impact of NF-κB activation alone using an inducible activation system.

These studies will allow us to determine the direct effects of LTR activation on gene expression and highlight the first steps of transforming a normal B cell gene expression pattern into that of the highly deregulated HL-specific pattern.

2. MATERIALS AND METHODS

2.1. Cell Culture

Three Hodgkin's lymphoma cell lines (L428, L1236 and KM-H2) were cultured in IMDM (Sigma) with 10% heat-inactivated foetal calf serum (FCS), 2mM Glutamine, 100u penicillin and 100u streptomycin (GIBCO). The two control cell lines Reh and Namalwa were cultured in RPMI 1640 (Sigma) with 10% heat-inactivated FCS, 2mM glutamine, 100u penicillin and 100u streptomycin (GIBCO). All cultures were incubated at 37°C in a humidified incubator with 5% CO2. Sub-culturing was carried out based on the cell densities in Table 2.1.

The 293T cell line used for the production of viral particles was cultured in DMEM (Sigma) with 10% FCS, 2mM Glutamine, 100u penicillin, 100u streptomycin (GIBCO). Sub-culturing of cells was carried out by removing media from the plate, washing with PBS and treating with 3ml trypsin (GIBCO) in PBS (Sigma) solution and re-suspending the cells in full growth media (DMEM, as above).

Table 2.1. Culture densities of cell lines.

Cell Line	Maximum Culture Density	Sub-Culture Density
Reh	$5x10^6$/ml	$0.5x10^6$/ml
Namalwa	$2x10^6$/ml	$0.5x10^6$/ml
L428	$1x10^6$/ml	$0.3x10^6$/ml
L1236	$0.6x10^6$/ml	$0.3x10^6$/ml
KM-H2	$1x10^6$/ml	$0.3x10^6$/ml
293T	80% confluent	15% confluent

2.2. Gene Expression Analysis

2.2.1. RNA Extraction

RNA was extracted from between $0.5x10^6$ and $5x10^6$ cells which were first pelleted by centrifugation at 300 x g for 5 minutes. The cells were lysed using 350 µl RA1 buffer (Macherey-Nagel) and 3.5 µl β- mercaptorthanol (Sigma). RNA extraction was carried out using the NucleoSpin® RNA kit (Macherey-Nagel) in accordance to the manufacturer's instructions. On column DNase digestion was also carried out to remove any genomic DNA from the sample according to the manufacturer's protocol (Macherey-Nagel). The final RNA was eluted in between 30 and 50 µl H_2O and quantified using NanoDrop™ 2000 (Thermo Scientific). For RNA-Seq the quality of the RNA extracts was validated by running on a Bioanalyzer 2100 (Agilent) and only RNA with a RIN value of 9 or more was used for library preparation.

2.2.2. cDNA Synthesis

cDNA was produced from RNA extracts using Superscript III reverse transcriptase (Thermo Fisher) as follows: 2 µg of RNA were added to 1 µl of oligo (dT) 500 µg/ml and the mix was made up to 11 µl with H_2O. It was then incubated at 65°C for 5 minutes using a thermo-cycler and held at 4°C. In the next stage 10 µl of 5x RT buffer, 5 µl 0.1M DTT, 5 µl 10mM dNTP, 1 µl RNase out, 1 µl Superscript III RT Enzyme and 17 µl of H_2O were added. This was then incubated for 1 hour at 50 °C, 15 minutes at 70 °C and stored at -20°C ready for use in qPCR.

2.2.3. qPCR

Gene expression was measured by qPCR using Sybr ® green master mix (Sigma) and normalised to *GAPDH* expression. Each reaction contained 0.5 µl of 10 µM primer mix (Table 2.2), 5 µl Sybr® Green Master mix, 2.5 µg DNA and was made up to 10 µl with dd H_2O.

Quantitative real time PCR was run using an Applied Biosystems StepOne Plus RT PCR system and the default PCR program (95°C – 10', 40 x 95 °C - 15" and 50 °C – 60" followed by a melting curve from 60°C to 95°C in 0.3°C steps). Quantitation was carried out using a standard curve with mixed cDNA samples and dilutions of 25ng, 5ng, 1ng, 0.2ng, 0.04ng. For DNaseI, ChIP and ATAC validation a genomic DNA standard curve was used instead.

Table 2.2. Expression primers were designed based on DNA sequences from RefSeq or from PrimerBank (*).

Primer	Forward	Reverse
CSF1R EX13/14	AGCACGAGAACATCGTCAACC	TTCGCAGAAAGTTGAGCAGGT
CSF1R EX2/3	CACCTGCCTGCCACTTCC	CCACACATCGCAAGGTCAC
CSF1R-LTR	TTGGATGTGATTCTGCTCCTC	CCACACATCGCAAGGTCAC
PAX5	CCATGTTTGCCTGGGAGATC	GGTTGGTTGGGTGGCTGCTG
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
LTA2	CATCTACTTCGTCTACTCCCAGG	CCCCGTGGTACATCGAGTG
RANTES (CCL5)	TACACCAGTGGCAAGTGCTC	TGTACTCCCGAACCCATTTC
CSF2 (*371502128c1)*	TCCTGAACCTGAGTAGAGACAC	TGCTGCTTGTAGTGGCTGG
CXCL10 (*323422857c1)*	GTGGCATTCAAGGAGTACCTC	TGATGGCCTTCGATTCTGGATT
CXCL11 (*307611978c1)*	GACGCTGTCTTTGCATAGGC	GGATTTAGGCATCGTTGTCCTTT
JUNB (*44921611c2)*	ACAAACTCCTGAAACCGAGCC	CGAGCCCTGACCAGAAAAGTA
CCR7 (*299473754c1)*	TGAGGTCACGGACGATTACAT	GTAGGCCCACGAAACAAATGAT
LITAF (*210147490c1)*	ATGTCGGTTCCAGGACCTTAC	TACGAAGGAGGATTCATGCCC
IL13	CATCGAGAAGACCCAGAGGA	TTTACAAACTGGGCCACCTC
IL6	CGAGCCCACCGGGAACGAAAG	GTGGCTGTCTGTGTGGGGCG
IRF4 (*305410879c1)*	GCTGATCGACCAGATCGACAG	CGGTTGTAGTCCTGCTTGC
IRF5	CAGGGGAGCTATCTTGGTCA	GATGGAGCTCCTTGAATTGC
TNF (*25952110c2)*	GAGGCCAAGCCCTGGTATG	CGGGCCGATTGATCTCAGC
CBFA2T3	CAGTTTGGCAGCGACATCTC	GCCTCCTGAAGCTTGGAATG
TNFRSF11A (*22547111c1)*	AGATCGCTCCTCCATGTACCA	GCCTTGCCTGTATCACAAACTTT
TNFRSF11A LTR	AGCCACATGGAACTGTAAGTC	ACTGGTACATGGAGGAGCGA
Chr18	ACTCCCCTTTCATGCTTCTG	AGGTCCCAGGACATATCCATT
TBP Promoter	CTGGCGGAAGTGACATTATCAA	GCCAGCGGAAGCGAAGTTA

2.2.4. RNA-Sequencing

RNA Sequencing libraries were produced in duplicate using the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human/Mouse/Rat (Illumina) according to manufacturer's

protocol. RNA was extracted using the NucleoSpin® RNA kit (Macherey-Nagel) with on column DNase digestion as previously described (2.2.1). RNA quality was assessed using a Bioanalyzer 2100 with a Eukaryote Total RNA Pico chip (Agilent). Ribosomal RNA was removed by diluting between 100ng and 300ng of total RNA to 10 µl with H_2O and adding rRNA Binding Buffer (5 µl) and rRNA Removal Mix (5 µl) and incubating at 68°C for 5 minutes. Magnetic rRNA removal beads were warmed to room temperature (35 µl) and the RNA mix added to them and incubated at room temperature for 1 minute. Ribsomal RNA bound to the beads was then removed by magnetic separation.

RNA clean-up was carried out using RNAClean XP beads by mixing 100 µl with of beads each sample, incubating at room temperature for 15 minutes and magnetic separation prior to the removal of the supernatant. The beads were then washed with 70% EtOH and dried for 15 minutes before eluting in 11 µl of elution buffer for 2 minutes. Elute, Prime and Fragment mix (8.5 µl) was added to the eluted RNA (8.5 µl) and incubated at 94°C for 8 minutes. To synthesise the first strand DNA 1 µl Superscript II reverse transcriptase (Invitrogen) was added to 9 µl First Strand Synthesis Act D mix and 8 µl transferred to each sample prior to incubation (25°C 10 minutes, 42°C 15 minutes, 70°C 15 minutes). For second strand synthesis 5 µl of resuspension buffer and 20 µl of Second Strand Mix (TruSeq kit) was added to each sample and incubated at 16°C for 1 hour. DNA from the reaction was purified using AMPure XP bead (Beckman) by addition of 90 µl of beads to each sample. DNA was bound to the beads by 15 minute incubation at room temperature and the supernatant removed following magnetic separation. The beads were washed twice with 80% EtOH and dried for 15 minutes at room temperature before eluting in 17.5 µl resuspension buffer.

The 3' ends of the DNA fragments were adenylated by addition of 12.5 µl A-Tailing mix (TruSeq kit) and 2.5 µl resuspension buffer to 15 µl elute and incubation at 37°C for 30 minutes and 70°C for 5 minutes. Indexed Illumina Adaptors were ligated to the DNA

fragments to allow for multiplexed sequencing by addition of 2.5 µl resuspension buffer, 2.5 µl Ligation Mix and 2.5 µl RNA Adaptor Index (diluted 1:4). The ligation was incubated for 10 minutes at 30°C and the reaction stopped by addition of Stop Ligation Buffer (5 µl). The adaptor ligated DNA was purified using 42 µl AMPure XP beads as previously described and eluted in 52.5 µl resuspension buffer. The purification was repeated with 50 µl supernatant and 50 µl beads with the final elution in 22.5 µl. Finally a 15 cycle PCR amplification of the library was carried out using 20 µl of eluted DNA, 5µl PCR Primer Cocktail and 25 µl PCR Master Mix (98°C 30 seconds, 15 x 98°C 10 seconds, 60°C 30 seconds, 72°C 30 seconds and final extension 72°C 5 minutes). A final AMPure purification with 50 µl of beads was carried out and the final library eluted in 30 µl.

The libraries were run on a Bioanalyzer 2100 with a High Sensitivity DNA Assay chip (Agilent) to determine the average fragment size and quantified by PCR using the Kappa Illumina Library Quantification Kit on an Applied Biosystems StepOne Plus RT PCR system. Sequencing was carried out using an Illuimina NextSeq 500 with each library run as 1/12 of a 150 cycle flow cell.

2.3. Chromatin Accessibility Assays

Two assays were used for the genome-wide assessment of open chromatin. DNaseI hypersensitive site mapping works by carrying out a short digestion with the DNaseI enzyme on permeabalised cells. The DNaseI initially cuts the DNA preferentially in open regions of chromatin where the DNA is not bound by nucleosomes. By limiting the length of digestion this technique produces a library of fragments representing these regions. The second method, Assay for Transposase Accessible Chromatin (ATAC-Seq) uses a transposase enzyme. In the same way as DNaseI the TN5 transposase will function preferentially in regions of open chromatin. This results in the DNA in these regions being cut into fragments

and tagged with a known sequence that can be amplified to produce a genome-wide library for sequencing.

2.3.1. DNaseI Hypersensitive Site Mapping

DNaseI Hypersensitive Site (DHS) mapping was carried out using the protocol from Bert, *et al.* (2007). Cell pellets of $4.5x10^6$ cells were washed in PBS and then suspended in 150 µl of DNaseI sucrose buffer (60mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris pH 7.4, 300 mM sucrose). DNaseI dilutions ranging between 90u and 200u were produced in DNaseI dilution buffer (60mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris pH 7.4, 0.4% NP40 and 2mM $CaCl_2$). Each sample was incubated at 22°C for 3 minutes and then 150ul of diluted DNaseI stocks were added to each reaction and incubated for a further 3 minutes. To stop the digestion after 3 minutes 300ul of Cell lysis buffer (3oo mM NaAcetate, 10 mM EDTA pH 7.4, 1% SDS and 1 mg/ml proteinase K) was added. The samples were then incubated overnight at 45 °C to allow the proteinase K to digest the protein. RNaseA (6 µl, 10 mg/ml) was added and incubated for 30 minutes prior to running 6 µl of the DNA on a 0.7% TAE agarose gel to visualize the digestions (Figure 2.1). The samples with optimal digestion were chosen for down-stream validation.

Figure 2.1. DNaseI digestion validation gel
DNaseI digestions from the DHS assay were run on a 0.7% agarose gel to determine the degree of digestion. The samples digested with 30u, 40u and 50u of DNaseI (highlighted in red) were chosen as the optimal level of digestion.

Further validation was performed by qPCR to assess the best digestion points for optimal enrichment of open chromatin in the cells. Primers were used to amplify a known highly DNaseI hypersensitive region (*TBP* promoter) a region of low sensitivity (*IVL*) and a region of minimal or no sensitivity (*Chr18*). The ideal digestion was chosen based on high *TBP/Chr18* and a low *IVL/Chr18* ratio.

The best DNaseI concentrations giving the highest open chromatin enrichment based on the initial qPCR results were chosen and size selection of fragments between 50 and 250 bp was carried out. Briefly, 12 µg of each digested DNA sample was run on a 0.8% gel and a fragment of between 50 and 250 bp cut out from each lane. DNA was extracted from the gel using a Qiagen mini-elute gel extraction kit in accordance to the manufacturer's instructions

and the extracted DNA was eluted in 30 ul of H_2O. The size-selected samples were again analyzed by qPCR using primers for low hypersensitivity regions *ACTB* (actin, beta) and Chr18 (a gene desert on chromosome 18) and highly hypersensitive region, *TBP* promoter (Table 2.2) (Figure 2.2). Optimal samples were then chosen to progress to library preparation.

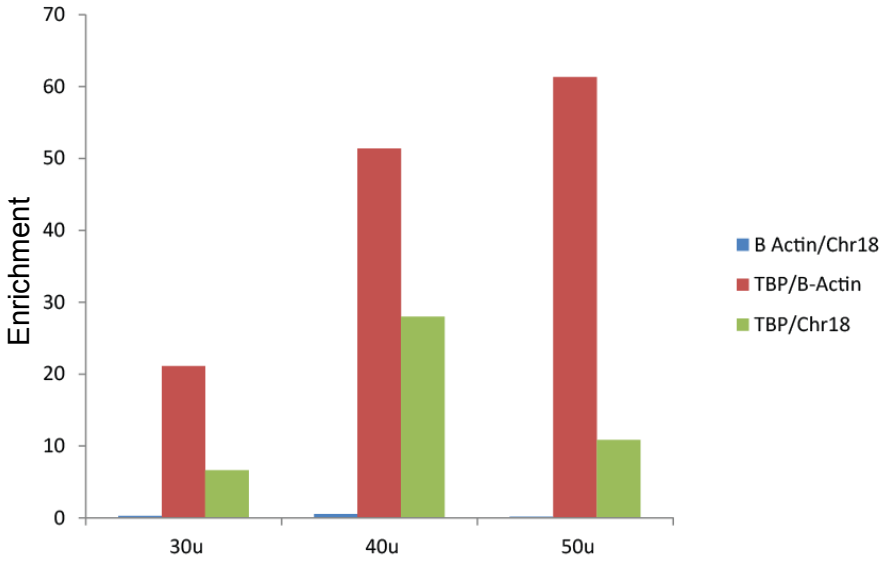


Figure 2.2. qPCR validation of DNaseI digestion levels

qPCR validation was carried out on size selected DNaseI samples to determine the optimal digestion level to use for library preparation. In this example 40 units was chosen as the optimal DNaseI concentration due to a high ratio of TBP/B-Actin and TBP/Chr18.

Indexed Illumina libraries were prepared using the MicroPlex Library Preparation Kit v2 (Diagenode) following manufacture's protocol. Briefly, 10 µl of DNA from the DNaseI assay was prepared with Template Preparation Buffer (2 µl) and Template Preparation Enzyme (1 µl) and incubated at 22 °C for 25 minutes and 55 °C for 20 minutes. The samples were then treated with Library Synthesis Buffer and Enzyme (1 µl of each) and incubated for a further 40 minutes at 22 °C. The library then underwent a step to amplify the library and add barcoded adaptors for Illumina sequencing. The samples were mixed with Library

Amplification Buffer and Enzyme (25 µl and 1 µl), H_2O (5 µl) and Indexing Barcode reagent (5 µl). The libraries were then incubated as follows; 72 °C - 3 minutes, 85 °C – 2 minutes, 98 °C – 2 minutes, 4 cycles of 98 °C – 20 seconds, 67 °C – 20 seconds, 72 °C – 40 seconds, followed by 12 cycles of 98 °C – 20 seconds and 72 °C – 50 seconds. The libraries were then run on a 2% agarose gel and size selected between 190-300 bp. The libraries were finally quantified by qPCR using the Kappa library quantification kit (Kappa Biosystems), pooled and run on an Illumina Hi-Seq 2000 75bp paired end.

2.3.2. Assay for Transposase Accessible Chromatin

ATAC-Seq was performed using the protocol originally developed by Buenrostro, et al. 2015 and modified by Corces, *et al.* 2016. The cells were first pelleted by centrifugation at 300 x g from 5 minutes, re-suspended in a volume of cold PBS to $1x10^6$ cells/ml and 50 µl (50,000 cells) were then pelleted by centrifugation at 300 x g for 5 minutes. The PBS was then removed and 50 µl transposition mix (25 µl 2 x TD Buffer (Illumina), 2.5 µl TN5 Transposase (Illumina), 5 µl digitonin (0.1% stock) and 17.5 µl H2O) was added. The cells were gently pipetted to mix and re-suspend and then incubated at 37 °C in a 300 rpm shaking heat block for 30 minutes. The resulting product was purified using the Qiagen MiniElute Reaction Cleanup kit following manufacturer's protocol and eluted in 20 µl H2O.

The transposase treated DNA was then amplified in 2 PCR amplification steps. Firstly by the addition of 2.5 µl 25 µM Nextera PCR Primer 1, 2.5 µl 25 µM Nextera Barcoded PCR Primer and 25 µl NEBNext Master Mix followed by incubation at 72 °C – 5 minutes, 5 cycles 98 °C – 30 seconds, 63 °C – 30 seconds and 72 °C 1 minute. A qPCR step was then carried out using 5 µl of the amplified library, H_2O (4.4 µl), primers as before (0.25 µl of each), 10 x SYBR Green I (0.6 µl) and NEBNext Master Mix (25 µl) and run for 40 cycles using the program as before. This step allowed for the optimal number of cycles to be calculated to achieve 25% amplification of the library, therefore minimising clonal duplication of sequences

in the library. The remaining material was further amplified based on the cycle number determined from the qPCR. The libraries were finally purified using AMPure XP beads following manufacturer's protocol and eluting in 22.5 µl elution buffer. The purified libraries were quantified by Kappa Library Quantification kit (Kappa Biosystems) and Bioanalyzer 2100 (Agilent) and sequenced using Illumina NextSeq 500.

2.4. Rapid Amplification of cDNA Ends followed by Sequencing (RACE-Seq)

We developed the RACE-Seq protocol to allow for the identification of active LTRs related to the THE1B LTR family. Briefly this technique works by the addition of a known adaptor sequence to the 5' end of all transcripts which then allows for the amplification of fragments between the 5' adaptor and a primer degenerate primer designed to target the THE1B elements. This allows for the production of a library of active LTRs which can then be barcoded and sequenced using the Illumina platform to produce a genome-wide dataset of LTR activation.

```
                    Predicted Transcription Initation Site
                    ↓
THE1B   TCCCCTGCACAWGCTCTCTTGCCTGCCGCCATGTAAGACGTGMCTTTGCTCCTCCTTCGC
        |||||||||  |  |||| |||||||||   |||||  || |||  | |||||||||||| ||
CSF1R   TCCCCTGCATATACTCTTTTGCCTGCCATCATGTTGGATGTGATTCTGCTCCTCCTTTGC

                GATTGTGAGGCCTCCCCAGCCATG  THE1B Primer Sequence
THE1B   CTTCCGCCATGATTGTGAGGCCTCCCCAGCCATGTGGAACT/gtgagtCCATTAAA
        ||||| | ||||||| ||||||||||  ||||||||  ||||| ||||||| ||||||
CSF1R   CTTCCACTATGATTCTGAGGCCTCCTCAGCCATGCTGAACT/gtgagtCAATTAAA
                                                 ↖
                                                 Predicted Splice Site
```

Figure 2.3 THE1B primer design

Alignment of the THE1B consensus sequence and the THE1B primer upstream of CSF1R. The THE1B primer was designed based on the region of the consensus sequence with greatest homology to the CSF1R THE1B LTR.

2.4.1. 5' RACE

RACE was carried out based on the ExactSTART™ Eukaryotic mRNA 5'-& 3'-RACE Kit (Epicentre) and supplied protocol, however due to the discontinuation of the Tobacco Acid Pyrophosphatase enzyme and therefore the above kit a number of modifications were made (Figure 2.4). RNA was treated with Alkaline Phosphatase for 15 minutes at 37°C (10 µl Apex Buffer, 5 µl Apex Heat-liable Alkaline Phosphatase (Epicentre), 1 µg RNA, made up to 100 µl with H_2O). An Ampure RNA-clean bead purification then was carried out prior to treatment with RNA 5' Pyrophosphohydrolase (RppH) for 1 hour at 37 °C (Purified RNA (40 µl), 10 x

Thermopol Buffer (NEB)(5 µl), RppH Enzyme (NEB)(5 µl)). The RppH reaction was stopped by addition of 1 µl of 500 mM EDTA solution and then purified with a further Ampure RNA-clean bead purification and eluted in 16 µl H_20. The 5' RACE acceptor oligonucleotide was ligated at 37 ° for 30 minutes (4 µl H_2O, 2 µl RNA ligase buffer, 1 µl 5' acceptor oligonucleotide, 1 µl 2mM ATP solution and 1 µl T4 RNA ligase (NEB)). First strand cDNA synthesis was carried out by addition of 14 µl H_2O, 1 µl cDNA synthesis primer, 2 µl dNTP PreMix, 2 µl MMLVV RT Buffer and 1 µl MMLV RT, the reaction was incubated at 37°C for 1 hour and 85°C for 10 minutes. The reaction then had 1 µl of RNase added and was incubated at 55°C for 5 minutes The second strand cDNA synthesis reaction components were added at 55°C. A degenerate biotinylated THE1B LTR primer (20 µM, 2 µl) was used along with a primer for the 5' RACE linker (5 µl), 21 µl H_2O, 30 µl PCR mix and 1 µl (2.5 U) PfuUltra II polymerase and incubated for 21 cycles (95°C 30 seconds, 21 x 95°C 20 seconds, 60 °C 20 seconds, 72°C 3 minutes).

Following second strand synthesis the fragments incorporating a biotinylated primer were selected using T1 Dynabeads (Thermo Fisher Scientific) as follows. T1 dynabeads (20 µl) were washed in B&W buffer (10 mM Tris-HCL, 1 mM EDTA, 2 M NaCl) using magnetic separation and re-suspended in 50 µl of 2 x buffer B&W buffer. An equal volume of RACE product was added and samples were mixed on a slow rotating wheel at room temperature for 1 hour. The beads were then captured using a magnetic separator and following washes, with B&W buffer and TE, were re-suspended in 33.75 µl 1 x TE.

The selected DNA was amplified off the beads by a 3 cycle PCR ((95°C 2 minutes, 3 x 95°C 20 seconds, 56 °C 20 seconds, 72°C 30 seconds and 72°C 3 minutes) using a non-biotinylated THE1B LTR primer (1.25 µl, 20µM), the 5' RACE adaptor primer (4 µl), 5 µl dNTPs (10mM), 5 µl PFU Ultra Buffer and 1 µ PFU Ultra polymerase. Finally purification was carried out using the Qiagen MiniElute PCR Cleanup Kit and eluted in 11 µl elution buffer.

2.4.2. RACE Validation by Cloning

To validate the genome wide LTR libraries produced by RACE, fragments were blunt-end cloned into the pBlueScript vector which allowed for Blue/White selection of positive clones. The individual clones could then be picked and the inserts were Sanger sequenced to determine the sequences of a number of fragments within the library.

The pBlueScript vector was digested with SmaI (NEB) using a standard protocol and the RACE fragments were ligated into the vector using T4 DNA ligase (NEB). The ligation reaction was then incubated for 5 hours at room temperature. Transformation of the ligated vectors into DH5α competent E.coli was carried out by heat shock at 42 °C and the culture was plated and incubated overnight on 1.5% agar LB plates containing ampicillin (100ng/ml) with added X-Gal to a concentration of 200 µg/ml. Clones containing fragments were picked based on blue/white selection and grown in culture (+ ampicillin 100ng/ml) at 37 °C overnight prior to mini-prep using the Qiagen Mini-Prep kit. The resulting DNA was digested using BamHI and EcoRI to ascertain inserted fragment sizes. The digested samples were separated on a 1% TAE agarose gel stained with ethidium bromide. All positive clones were then sequenced using the pBlueScriptKS primer Table 2.3. Sequenced clones were mapped to the genome individually using UCSC Blat to establish whether they aligned uniquely and which regions they aligned to.

Table 2.3 RACE Primers

Name	Sequence
THE1B	CATGGCTGGGGAGGCCTC
pBlueScriptKS	TCTAGAACTAGTGGATC
RACE 5' Primer	TCATACACATACGATTTAGGTGACACTATAGAGCGGCCGCC TGCAGGAAA

2.4.3. RACE-Seq

Libraries for genome wide RACE-Seq were produced using the MicroPlex Library Preparation Kit v2 (Diagenode). Purified RACE material (10 µl) was added to 2 µl of template preparation buffer and 1 µl of Template Preparation Enzyme and incubated at 22°C for 25 minutes and 55°C for 20 minutes. The prepared template was then mixed with 1 µl Library synthesis buffer according to the manufacturer's instructions and 1 µl Library synthesis enzyme and incubated at 22°C for 40 minutes. The libraries were then mixed with amplification buffer (25 µl), amplification enzyme (1 µl), H_2O (4 µl) and a Barcoded Indexing Reagent (5 µl) to allow for the samples to be multiplexed for sequencing. The mix was split into 2 reactions of 25 µl which were amplified at 12 and 14 cycles to allow for selection of enough material to sequence without introducing clonal amplification (Extension and Cleavage: 72°C 3 minutes, 85°C 2 minutes, Denaturation: 98°C 2 minutes, Addition of Indexed oligonucleotides: 4 x 98°C 20 seconds, 67°C 20 seconds, 72°C 40 seconds, Library Amplification: 12 or 14 x 98°C 20 seconds, 72°C 50 seconds).

Finally size selection and purification of the libraries was carried out by running products on a 1.2% TAE agarose gel (with 0.05% ethidium bromide) and excising fragments between 190 and 300 bp. These were then extracted using the Qiagen mini-elute gel extraction kit and eluted twice in 12 µl H_2O. The libraries were run on a Bioanalyzer 2100 with a High Sensitivity DNA Assay chip (Agilent) to determine the average fragment size and quantified by PCR using the Kappa Illumina Library Quantification Kit on an Applied Biosystems StepOne Plus RT PCR system.

The indexed libraries were pooled and sequenced on Illumina MiSeq using the 150-Cycle paired end kit or a NextSeq 500 as a fraction of a 150 cycle flow cell.

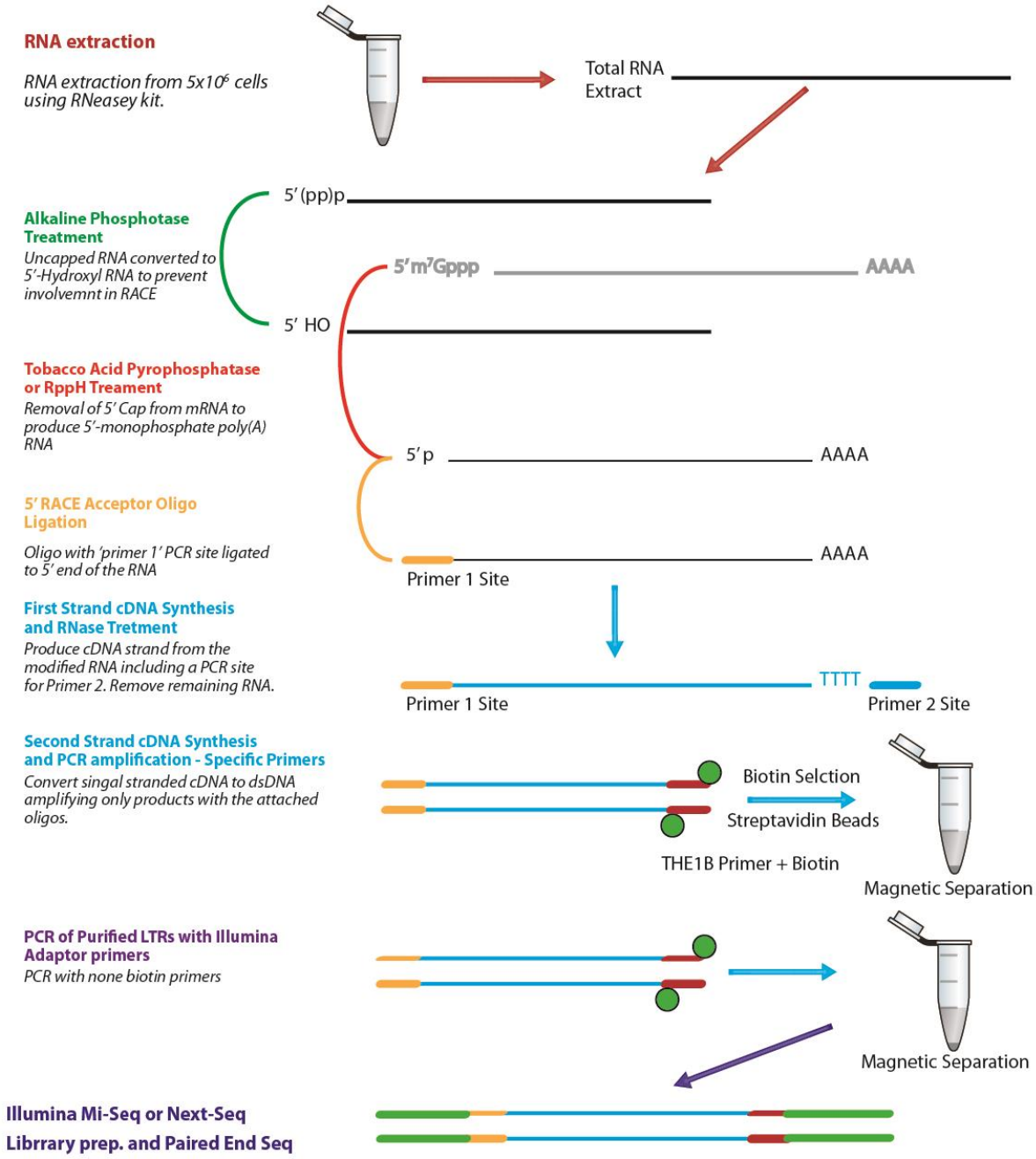


Figure 2.4 Schematic representation of RACE-Seq protocol

2.5. ChIP-Seq

To study the distribution of H3K4me3 at transcribed LTRs chromatin immune-precipitation followed by sequencing (ChIP-Seq) was used. ChIP-Seq was carried out using double cross-linked chromatin to ensure that we were able to capture all protein-DNA interactions.

To produce double cross-linked chromatin cultured cells were washed three times in PBS (Sigma) with centrifugation for 5 minutes at 300xg between washes and finally re-suspended in PBS at $1x10^7$ cells per ml. The first crosslinking step was performed by adding 8.3 µl of 100mg/ml disuccinimidyl glutarate (DSG) (Sigma Aldrich) per ml of cells and incubating for 45 minutes on a roller at room temperature. The cells were then washed a further 4 times in PBS using the previous conditions. The second cross-linking step was performed by re-suspending the cells in 1 ml of 1% formaldehyde (Thermo Scientific) diluted in PBS and incubated at room temperature on a roller for 10 minutes. The reaction was then stopped by addition of $1/10^{th}$ volume of 2M Glycine and mixed by pipetting. The cross-linked cells then pelleted by centrifugation at 400xg for 5 minutes and washed twice in ice-cold PBS.

The chromatin was prepared by firstly re-suspending the cross-linked cells at $1x10^7$ cells/ml in ice cold Buffer A (10 mM HEPES pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, 1:1000 PIC (Protease Inhibitor Cocktail)(Sigma) and 0.1 mM PMSF) and incubated at 4 °C on a rotating wheel for 10 minutes. The nuclei were then pelleted by centrifugation for 5 minutes at 400 x g and resuspended at $1x10^7$ cells/ml in ice-cold Buffer B (10 mM HEPES pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton X-100, 1:1000 PIC and 0.1 mM PMSF). The resuspended pellet was incubated on a rotating wheel at 4°C for 10 minutes and pelleted by centrifugation at 500 x g for 5 minutes. Finally the nuclei were resuspended in ice-cold IP Buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.25% SDS, 1:1000 PIC and 0.1 mM PMSF) at $3.3x10^7$ cells/ml.

The chromatin was fragmented by sonication in aliquots of 300 µl using a Picorupter (Diagenode) for 8 cycles (30 seconds on, 30 seconds off). The sonicated chromatin was centrifuged at 16,000 x g for 10 minutes and the supernatant was diluted with 2 volumes of IP Buffer II (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2mM EDTA, 1% Triton X-100, 7.5% glycerol, 1:1,000 PIC and 0.1 mM PMSF).

Prior to ChIP the antibody was bound to Protein G Dynabeads as follows; 10 µl of beads were washed by addition of 500 µl 0.1M citrate-phosphate buffer pH 5.0. The beads were then immobilised using a magnetic rack and resuspended in 15 µl 0.1 M citrate-phosphate buffer. The beads were then mixed with 2 µg H3K4me3 antibody (Millipore 07-473) and 5% BSA (final concentration 0.5%), this was then incubated on a rotating wheel at 4°C overnight. At this point 10% of the cross-linked chromatin was taken as an input control and mixed with 4 µl 5M NaCl and 0.5 µl 50 mg/ml proteinase K (Sigma) The input control was then reverse cross-linked overnight at 65 °C.

An additional pre-clearing step was performed to reduce the level of background in the ChIP. Firstly 10 µl Protein G Dynabeads were washed with 500 µl, 0.1 M citrate-phosphate buffer pH 5.0 prior to immobilisation using a magnetic rack and resuspention in 440 µl of prepared chromatin. The pre-clearing mix was then incubated on a rotating wheel at 4°C for 1 hour before immobilisation of the beads using a magnetic stand.

The Protein G Dynabeads with the pre-bound H3K4me3 antibody were washed in 500 µl 0.1 M citrate-phosphate buffer. The beads were then separated from the buffer by magnetic separation and resuspended in 15 µl citrate-phosphate buffer with 5% BSA. The pre-bound beads were then added to the supernantant from the pre-clearing step and incubated overnight a 4 °C on a rotating wheel.

Following overnight incubation the beads with bound chromatin were separated from the supernatant by magnetic separation and the following washes were performed each with 500

µl wash buffer; one wash with Wash Buffer 1 (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 10% Triton-X 100, 0.1% SDS), two washes with Wash Buffer 2 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA pH 8.0, 10% Triton-X 100, 0.1% SDS), one wash with LiCL Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA pH 8.0, 0.5% NP-40 and 0.5% Na-Deoycholate) and finally two washes with TE/NaCl Buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl and 1 mM EDTA pH 8.0). The beads were then resuspended in 50 µl elution buffer and incubated at 65 °C for 15 minutes. The beads were then removed by magnetic separation and the supernatant mixed with 5 µl of 5M NaCl and 0.5 µl proteinase K (50 mg/ml) and incubated at 65 °C overnight.

The final ChIP step was to purify the DNA using AMPure XP beads (BD) following manufactures protocol and eluting in 50 µl 0.1 x TE pH 8.0. The ChIP experiment was then validated by qPCR using primers designed to a known region of H3K4me3 (TBP promoter) and a known region with no H3K4me3 (IVL). This data was normalised to input DNA and measured against a standard curve of genomic DNA. Following validation libraries were produced using the Kappa Hyperprep Kit (Kappa Biosystems) following manufacturers protocol with an amplification step of 12 cycles. The library was size selected between 250 – 500 bp by gel electrophoresis on a 1% TAE gel followed by extraction using the Qiagen MiniElute Gel Extraction kit, following manufacturer's instructions and eluting in 15 µl. A final qPCR validation step was carried out using the previous primers to determine the library enrichment. The library was quantified by Bioanalyser (Agilent) and Kappa Library Quantification Kit and sequenced on a NextSeq 500 (Illumina).

Table 2.4 ChIP-Seq Primers

Name	Sequence - Fwd	Sequence - Rev
TBP-promoter	CTGGCGGAAGTGACATTATCAA	GCCAGCGGAAGCGAAGTTA
IVL - promoter	GCCGTGCTTTGGAGTTCTTA	CCTCTGCTGCTGCCACTT

2.6. siRNA Knockdown *TNFRSF11A*

Small interfering RNA knockdown of *TNFRSF11A* was performed in the L1236 cell line by electroporation. Briefly, $1x10^7$ cells were pelleted by centrifugation at 300 x g for 5 minutes and resuspended in 700 µl Optimem medium (Sigma). This was mixed with siRNA (SMARTpool: siGENOME TNFRSF11A, Dharmacon) to a final concentration of 200 Nm in a 0.4mm electroporation cuvette and electroporated using a Gene-Pulser X-cell (Bio-Rad) with 960 µF and 0.18 kV. The cells were then returned to culture for down-stream analysis.

2.7. Nuclear Protein extraction and Western Blotting

Nuclear protein extraction was carried out using the Active Motif Nuclear Extraction Kit following manufacturer's protocol and the extracted protein was quantified by using the Pierce BCA Protein Assay Kit. The quantified protein was mixed with 6x loading buffer (0.3125M Tris-Cl pH6.8, 10% SDS, 25% glycerol, 10% β-2 Mercaptoethanol and 0.05% Bromophenol blue) and was heated to 95 °C for 5 minutes. The denatured protein was then run on a Mini-PROTEAN TGX precast gel (Bio-Rad) in SDS-PAGE running buffer (0.025M Tris Base pH8.3, 0.192M Glycine and 0.1% SDS). The protein was transferred onto a nitrocellulose membrane using Trans-Blot Turbo transfer system with the default turbo blot program.

The membrane was then blocked with 5% milk in TBST (0.2% Tween™ 20, 0.075 M NaCl and 0.01 M Tris pH 7.5) for 30 minutes. The membrane was then probed with primary antibody (Table 2.5) diluted in 5% milk TBST by rocking at 4 °C overnight. Following this the membrane was washed 3 times in TBST and re-probed with the appropriate HRP secondary antibody in 5% milk TBST (Table 2.5). The membrane was then developed using ECL detection (GE Healthcare) and imaged on a Bio-Rad ChemiDoc XRS+.

Table 2.5 Western blotting antibodies

Antibody	Dilution
NF-κB p65 Mouse (6956s - Cell Signalling)	1:1,000
β-Actin (A1978 - Sigma)	1:5,000
Anti-Mouse HRP (Jackson ImunoResearch)	1:5,000

2.8. iIkKβ Cloning

To investigate the impact of constitutive NF-κB activation on LTR activation in the Reh cell line, a doxycycline inducible lentiviral vector was produced containing cDNA for *IκBKβ* and IRES GFP. This vector was transduced into Reh cells, making a stable cell line with doxycycline inducible NF-κB constitutive activation.

2.8.1. Cloning iIkKβ into PCW57.1

THE PCW57.1 (Adgene) vector uses the Gateway cloning system meaning that all cDNA was first cloned into the pENTR vector (Addgene). IRES GFP was excised from the pSIEW vector by restriction enzyme digest with BamHI (20U), SalI (20U) and 5 µl SmartCut Buffer (NEB) in a 50 µl reaction and incubated at 37°C for 1 hour. The pENTR vector was digested with BamHI and XhoI using the same protocol and both digests were run on a 1% agarose, TAE electrophoresis gel. The digested fragments were excised from the gel and extracted using the Qiagen gel extraction kit in accordance to the manufacturer's instructions and the extracted DNA was eluted in 30 ul of H_2O. The IRES GFP fragment was then ligated into the pENTR vector by addition of T4 DNA Liagse (1 µl, NEB), T4 DNA Ligase Buffer (1 µl, NEB) with an insert to vector ratio of 3:1 and incubation at room temperature for 2 hours.

The ligation product was used to transform chemically competent Sub-cloning Efficiency™ DH5α™ *E.Coli* (Invitrogen) by heat-shock at 42°C. The bacteria were grown on LB plates with Kanamycin (50 µg/ml) selection and colonies picked to grow up for mini-prep. Mini-Preps were carried out using the Qiaprep® Spin Miniprep Kit (Qiagen) following the supplied protocol and eluting in 50 µl H_2O. The resulting clones were sequenced by sanger sequencing using primers designed to either side of the insert site (pENTR-Fwd (CTACAAACTCTTCCTGTTAGTTAG) and pENTR-Rev (ATGGCTCATAACACCCCTTG)). Clones matching the correct sequence were chosen and Maxi-Prepped using the NucleoBond ® Xtra Midi Plus EF kit following manufacturer's instructions.

The cDNA for *IκBKβ* was excised from a vector obtained from Stephan Mattas, Charité–Universitätsmedizin, Berlin, by digestion with 20 U SmaI and 20 U NotI (plus 5 µl CutSmart Buffer in 50 µl reaction) with a sequential incubation of 25°C 1 hour and 37°C for 1 hour. The digested fragments were separated on an agarose electrophoresis gel and the correct size fragment extracted as before. The pENTR IRES GFP construct was digested with BamHI following the previous protocol and both the vector at *IκBKβ* were blunt-ended using 3.75 µl T4 DNA polymerase (Promega), 10 µl T4 Buffer and 4 µl dNTPs (10 mM). The reaction was incubated at 37°C for 5 minutes and stopped by addition of 4 µl 0.5M EDTA and the DNA purified using the Qiaquick ® PCR Purification kit and eluted in 30 µl. The *IκBKβ* fragment was ligated into the vector at a 3:1 ratio with 1 µl T4 DNA Ligase (NEB), 2 µl T4 DNA Ligase Buffer (NEB) in a 20 µl reaction and incubated at room temperature for 2 hours.

Sub-cloning Efficiency™ DH5α™ chemically competent *E.Coli* (Invitrogen) were transformed as previously described and the resulting DNA preps sequenced.

2.8.2. Lentivirus Production and Transduction of Reh cells

Lentivirus was produced in the 293T cell line, concentrated and used to infect Reh cells as described below.

The 293T cells were transfected using TransIT293 (Mirus) which was firstly warmed to room temperature. The plasmid DNA was mixed; 12 µg PCW57.1, 0.6 µg Tat, 06 µg Rev, 0.6 µg gag/pol and 1.2 µg vsv-g. The DNA was then mixed with 1.5 ml of Optimem media and 45 µl TransIT293 was added and mixed by pipetting then incubated for 30 minutes at room temperature. The transfection mix was then added dropwise to the 293T cells in 15.5 ml of DMEM and incubated for 24 hours. Virus containing media was collected at 12 hour intervals for 3 days and refrigerated until day 3.

The virus was concentrated prior to infection of the Reh cells as follows. The collected virus containing media was centrifuged at 500 x g for 15 minutes at 4 °C and the supernatant filtered using a 0.45 µm acrodisk syringe filter (PVDF low protein binding)(Sigma). Concentration was then carried out using protein purification columns with 100-kD cut-off (Centricon Plus concentrators, Millipore) with centrifugation at 2,000 x g for 20 minutes.

This virus was then used to infect Reh cells by spin infection. Reh cells were plated at $1x10^6$ cells/ml and the concentrated virus and 8 µg/ml polybrene was added. The plate was then centrifuged for 2 hours at 1,500 x g at 32 °C. Following spin infection the plate was returned to the incubator and the media changed after ~12 hours. The cells were then selected by treatment with puromycin and used for downstream experiments.

2.9. Data Analysis

2.9.1. RNA-Seq

RNA-Seq reads were mapped to the hg19 human reference genome using Tophat2 (Kim, Pertea et al. 2013). Reads mapping to the sense and anti-sense strands were split into separate files and histogram density plots were created from the mapped reads using Bedtools genomecov with the '–d –split' option and uploaded to UCSC genome browser (Quinlan and Hall 2010). To obtain normalised FPKM (fragments per kilobase of transcript per million mapped reads) values for gene expression CuffNorm was used (Trapnell, Roberts

et al. 2012). Further analysis was carried out using LOG2 FPKM values in R and Microsoft Excel (R-Development-Core-Team 2008). Expressed genes were defined as any gene with a LOG2 FPKM value of 0 or above and differential expression between cell lines was defined based on at least a 2-fold change in expression.

To perform clustering of the RNA-seq data from the cell lines pairwise Pearson correlation of gene expression was used to produce a correlation matrix. Clustering of the RNA-Seq data by Pearson correlation was performed using R with the heatmap.2 function in the gplots package using hierarchical clustering with Euclidean distance and average linkage.

Gene Ontology analysis was performed using DAVID on lists of up and down regulated genes as previously defined. KEGG Pathway analysis was carried out using the ClueGo and CluePedia packages in Cytoscape with lists of up and down regulated genes combined from each HL cell line compared to each control cell line (Shannon, Markiel et al. 2003, Bindea, Mlecnik et al. 2009, Bindea, Galon et al. 2013). The network was produced based on KEGG terms with a pV < 0.05 and the layout was manually adjusted to enable all interactions to be visualised.

2.9.2. DNaseI-Seq, ATAC-Seq and ChIP-Seq

Quality control data was obtained for DNaseI-Seq, ATAC-Seq and ChIP-Seq reads using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were the mapped to the hg19 version of the human reference genome using Bowtie2 with the '–very-sensitive' parameter (Langmead and Salzberg 2012). Histogram alignment density plots were produced using the genomecov function of Bedtools and the plots uploaded to UCSC genome browser. Regions of enrichment (peaks) in DNaseI-Seq, ATAC-Seq and ChIP-Seq data were identified using Macs 1.4 with default parameters (Zhang, Liu et al. 2008). Overlaps of ATAC-Seq, ChIP-seq and DNaseI-Seq datasets with other datasets were performed using the intersect function of bedtools (Quinlan and Hall 2010).

Clustering of the DNaseI-Seq and ATAC-Seq datasets from each cell line was performed based on the Pearson correlation of read (tag) counts at the summit of each peak. To define a set of hypersensitive sites (peaks) for correlation to be performed the summits of peaks identified in all datasets were concatenated and any summits occurring within 400 bp of each other were merged using Bedtools merge. The count of reads overlapping with the summit of these newly defined peaks was then obtained using the annotatePeaks function of Homer (Heinz, Benner et al. 2010). Pairwise correlation of this data was then performed in R and hierarchical clustering and heatmap production carried out using the heatmap.2 function of the gplots package in R as previously described.

2.9.3. Digital Genomic Foot-printing

Digital genomic footprinting was performed using the Wellington algorithm (Piper, Elze et al. 2014), using standard parameters. Input files were peaks identified via MACS (Zhang, Liu et al. 2008) for high-depth sequencing DNaseI-Seq runs performed in resting Reh, KM-H2 and L428 cells. Retained footprints were those identified with an FDR ≤ 0.01, as per standard parameters of the algorithm.

2.9.4. Digital genomic footprinting motif co-occurrence clustering

Digital genomic footprinting motif co-occurrence clustering was performed as previously described (Cauchy, James et al. 2015, Obier, Cauchy et al. 2016). Briefly, motif discovery was first performed in footprints via the Homer findMotifsGenome algorithm (Heinz, Benner et al. 2010), using -size given as a parameter. 16 motifs representing motifs found in HL and NHL cell lines were selected for computing constraints. Motif mapping to footprints was performed via Homer annotatePeaks on specific populations. Motif co-occurrences within 50bp were computed via pyBedTools intersection_matrix (Dale, Pedersen et al. 2011). To assess for significance of co-occurrence enrichments, background co-occurrences of footprinted motifs were computed using whole footprint populations of the reference cell type,

using 1000 random subsamplings of sizes equal to the specific population in focus. An enrichment score was then derived using co-occurrences of the specific population versus the background as $Z=(x-\mu)/\sigma$, where x is the co-occurrence in the specific population for a given motif, μ and σ are the average co-occurrence and standard deviation of co-occurrence as computed from the background.

2.9.5. RACE-Seq

RACE-Seq reads were first trimmed using Cutadapt to remove the RACE adaptor sequence (TCATACACATACGATTTAGGTGACACTATAGAGCGGCCGCCTGCAG
GAAA) and the THE1B primer sequence (CATGGCTGGGGAGGCCTC). The trimmed reads were then mapped to the hg19 version of the human reference genome using Bowtie2 with the '—very-sensitive' parameter (Langmead and Salzberg 2012). Histogram density plots were produced for each biological replicate and for the merged replicates using bedtools genomecov and the resulting plots uploaded to UCSC genome browser. Regions of enrichment (peaks) were identified using Macs1.4 with the '–keep-dup=all' parameter. The resulting peaks from each biological replicate were overlapped to identify the shared peaks using the ChipPeakAnno package in R and venn diagrams produced. High-confidence RACE-Seq peaks were defined by the presence of a peak in at least 2 out of 3 biological replicates. High confidence peaks were selected using an in house bedtools script and used for all further analysis. All further venn diagrams comparing the RACE-Seq peaks between cell lines were performed using the ChipPeakAnno package in R and lists of overlapping and specific peaks were created using the intersect function in bedtools (Zhu, Gazin et al. 2010). Annotation of repeat elements also made use of the bedtools intersect function overlapping the datasets with the Repeat Masker annotation track obtained from UCSC genome browser.

The clustering of RACE data between cell lines was carried out by creating a matrix of the number of peaks shared between each pair of cell lines using the ChipPeakAnno package in

R. To compare these binary datasets the Dice index coefficient was calculated for each pairwise comparison in the context of the entire population and clustering was carried out as previously described using the heatmap.2 function in the gplots package of R.

Annotation of the genomic regions in which the active LTRs (RACE peaks) resided was performed using the annotatePeaks function of Homer. This function was also used to create the average profile plots by using the –hist parameter to plot average RNA-Seq tag counts within 20 bp bins 250 bp around the RACE peaks.

Closest genes to RACE peaks were identified using bedtools closest and a hg19 gene annotation reference. To determine closest genes in the same orientation and on the same strand the expressed LTR strand was first determined using bedtools intersect with the '–wao' parameter to obtain strand annotation form the repeat masker annotation. The closest genes which shared the same strand as the active LTR were then annotated using bedtools closest with the '-s -t first -iu -D a' parameters.

Finally supervised clustering of the individual LTRs was performed by concatenating high confidence RACE peaks identified in all cell lines and merging any which overlapped using bedtools merge. The newly defined set of peaks were annotated for their presence or absence in each of the sets of high confidence peaks from the cell lines and these were then sorted based on which cell lines the peaks were shared in. The resulting clusters were plotted as a heatmap using the heatmap.2 in the gplots package of R and carrying out hierarchical clustering of the overall LTR activation pattern in each cell line.

2.9.6. LTR presence by gene expression fold change

Pairwise RNA-Seq datasets were ranked by $\log_2$ FPKM fold change, as defined as FC=($\log_2$ sample A FPKM+1)/($\log_2$ sample B FPMK+1) to avoid dividing by 0, with all genes ranked accordingly. Separately, LTRs identified via RACE-Seq were annotated to the closest gene using bedtools closest –a <LTR peak file.bed> -b hg19_refGene.bed -t first as parameters

(Quinlan and Hall 2010). LTR presence for all genes was thus computed by performing a left outer join of all genes and genes annotated as closest to LTR peaks via the merge function of R, using merge(<all genes sorted by log_2 FPKM fold change>, <gene list of annotated LTR peak file>, all.x=T), then replacing all matches with the value 1 and NULL values by 0 in the column originating from the gene list of the annotated LTR peak file. Resulting files were subsequently written as text files via write.table in R, then visualised and saved as heatmap images using Java TreeView (Saldanha 2004).

3. RESULTS

3.1. The Hodgkin's Lymphoma transcriptional network has a global impact on chromatin structure and gene expression.

3.1.1. Hodgkin's Lymphoma displays a unique gene expression pattern

It is known from many studies that the HRS cells of Hodgkin's Lymphoma (HL) have a very different gene expression pattern to cells of the B-Cell lineage with loss of B cell receptor expression and gain of a number of other factors required for cell survival (Schwering, Brauninger et al. 2003). The majority of published genome-wide HL expression data is based on micro-array technology which has limited utility. To enable us to investigate the use of alternative promoters (including LTRs), splicing patterns, quantitative gene expression patterns in HL we used RNA-Seq technology. RNA-Seq was performed in duplicate, on 3 HL cell lines (L428, L1236 and KM-H2) and 2 control cell lines, Reh; a B-cell leukaemia (Rosenfeld, Goutner et al. 1977) and Namalwa; a Burkitt's lymphoma (Klein, Dombos et al. 1972). The RNA-Seq duplicates were validated by calculation of linear regression producing R^2 values between 0.97 and 0.87 (Figure 3.1). The slight level of variation seen between replicates of the HL cell lines compared to the control cell lines is likely caused by a number of gene with unstable expression due to the global deregulation of gene regulation. This probably also accounts for the variation in cell size, adherent properties and other phenotypic variability observed between cells of HL cell lines.

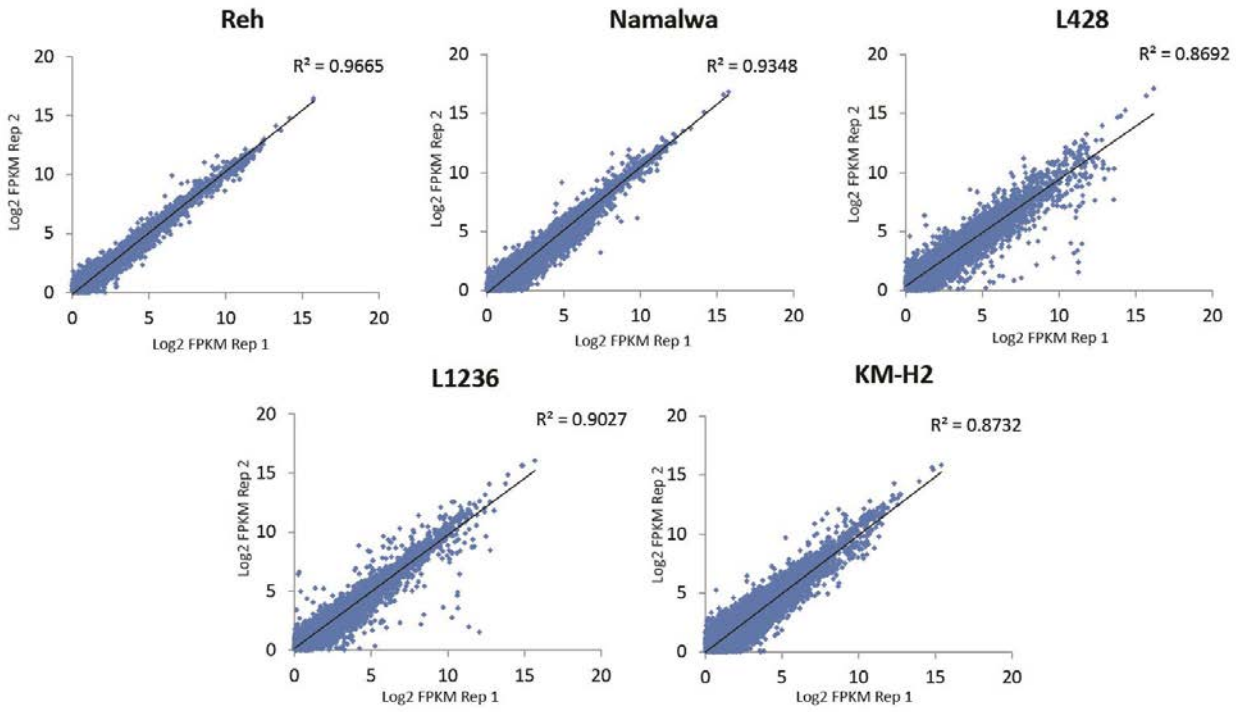


Figure 3.1 High level of correlation between biological replicates of RNA-Seq
Comparison of RNA-Seq biological replicates from 2 control cell lines (Reh and Namalwa) and 3 HL cell lines (L428, L1236 and KM-H2). Log2 FPKM values for each gene with an FPKM of at least 1 were plotted and linear regression calculated to showing a high similarity between replicates.

To illustrate the variation in gene expression between the cell lines a pairwise Pearson correlation was calculated between the FPKM values. Hierarchical unsupervised clustering of these correlations showed that the 3 HL cell lines cluster together and the 2 control lines cluster also cluster together but away from the HL cell lines (Figure 3.2). The clustering also showed heterogeneity between the HL cell lines, with L428 clustering further away from L1236 and KM-H2. The L428 cell line also has the least correlation with the control cell lines suggesting it is the furthest from a B-cell in terms of gene expression pattern.

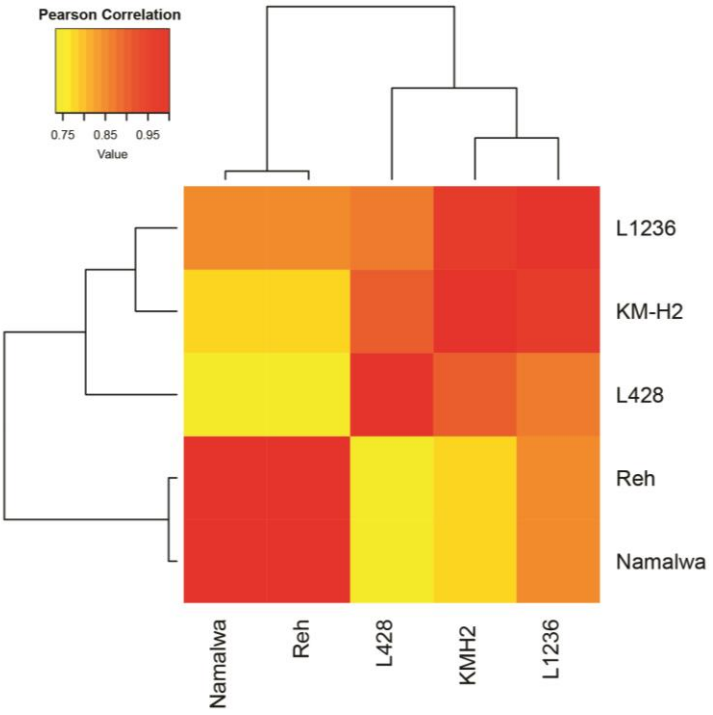


Figure 3.2. Correlation clustering of gene expression patterns of the HL and non-HL cell lines
Pearson correlation of gene expression patterns determined by RNA-Seq and clustered by hierarchical unsupervised clustering. The HL cell lines (L428, L1236 and KM-H2) cluster independently to the control cell lines (Reh and Namalwa) and L428 clusters further away from the other HL cell lines indicating heterogeneity between the HL cell lines.

To further validate the global gene expression data of the different cell lines we assessed gene expression of a panel of genes known to be upregulated in HL compared to normal B-cells (Kreher, Bouhlel et al. 2014) (Figure 3.3). The panel was made up of 10 pro-inflammatory cytokines and 3 transcription factor genes known to be involved in the survival and proliferation of HL. The analysis showed expression of the entire panel of genes in the HL cell lines and HL specific expression of *IL13, IL6* and *CSF2*. The Reh and Namalwa cell lines also showed expression of *CCL5, IRF5, JUNB, CCR7, LITAF, LTA, IRF4* and *TNF*. Additionally a low level of expression of *CXCL10* and *CXCL11* was also seen in Namalwa. The expression of many of these genes in the control cell lines, many of which are cytokines involved in immune and inflammatory response, is not surprising as both cell lines are derived from cancers, most of which have some degree of inflammatory signature. The expression pattern of these genes shows a large amount of variation between the 3 HL cell

lines. This could be partially down to experimental variation although the biological replicates should help to avoid this. Variation between the cell lines is also to be expected as they originate from patients with different genetic backgrounds.

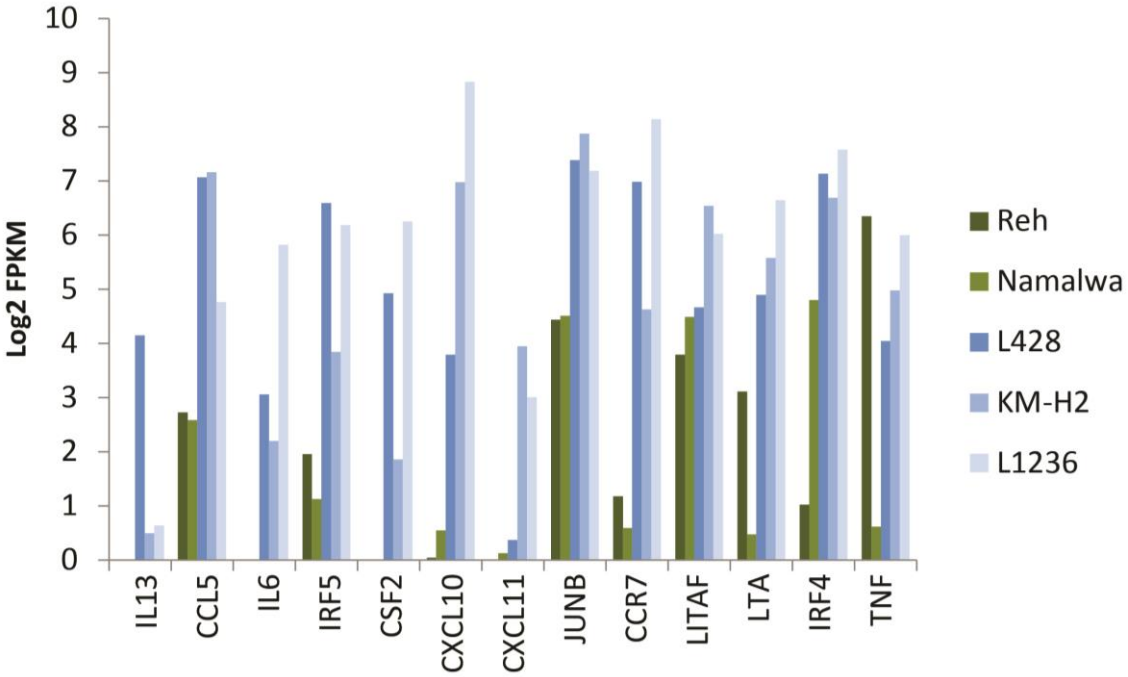


Figure 3.3 Gene expression level of a panel of 13 genes with a known involvement in HL

Mean Log2 FPKM values obtained from RNA-Seq replicates in HL and control cell lines were plotted for each of a panel of 13 HL associated genes (Kreher, Bouhlel et al. 2014) . Expression of all of the genes is observed in the HL cell lines, 3 of which show no expression in the control cell lines (*IL13*, *IL6* and *CSF2*). The other genes all have expression in the control cell lines but in most cases at a lower level to that seen in the HL cell lines.

To validate the RNA-Seq data, qPCR primers were designed against the same panel of genes using RNA from independent experiments. Overall the gene expression measured by qPCR showed a similar pattern with the entire panel being expressed in the HL cell lines (Figure 3.4). The qPCR results showed little expression of all but *LITAF, LTA* and *TNF i*n the control cell lines. The relative levels of expression between the cell lines vary from what is seen in the RNA-Seq data. It is likely this is the result of a combination of variable primer efficiency, normalisation between the cell lines and the small region of the genes which are measured by qPCR compared to RNA-Seq which measures the entire transcript.

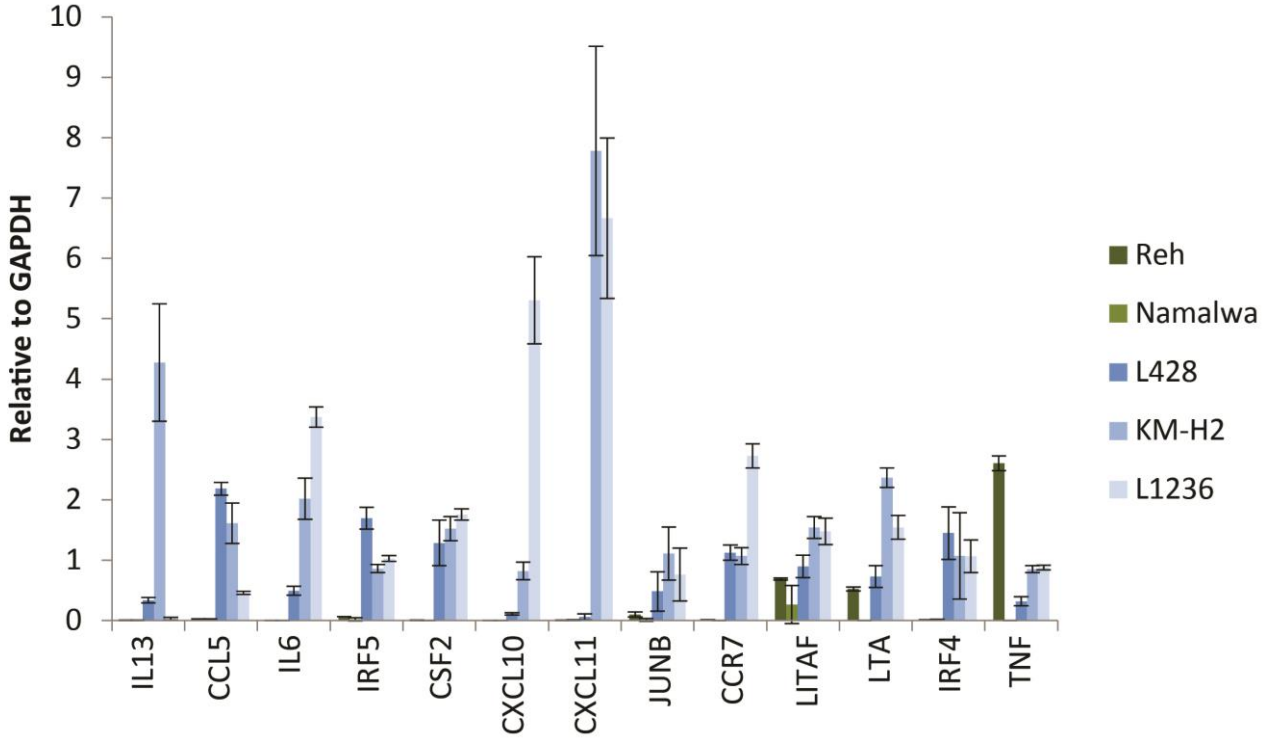


Figure 3.4. Gene expression of a panel of 13 genes with a known involvement in HL measured by qPCR
Gene expression relative to housekeeping gene *GAPDH* of a panel of 13 genes associated with HL in 3 HL cell line (L428, L1236 and KM-H2) and 2 control cell lines (Reh and Namalwa). Expression of all genes is observed in the HL cell lines although some are only present at low levels. Expression of *LITAF*, *LTA* and *IRF4* is also observed in the control cell lines.

3.1.2. B-Cell specific gene expression is downregulated and pro-inflammatory genes are upregulated in HL.

We next wanted to assess how the global de-regulation of gene expression seen in HL impacted on the pathways and processes within HL cells. To achieve this any genes which were at least 2-fold up- or down-regulated compared to the control cell lines underwent Gene Ontology (GO) analysis (Figure 3.5). The resulting GO terms showed some variation between the HL cell lines which is to be expected from the variation in gene expression between the cell lines.

All 3 HL cell lines show a down-regulation of genes involved in transcription and transcriptional regulation. Many genes which are involved in maintaining the B cell gene expression programme are down-regulated such as *BCL6, EBF1, PRMT5, MYC* and *PAX5*. *CBFA2T3* and *FOXO1* (Lamprecht, Walter et al. 2010, Xie, Ushmorov et al. 2012). A large number of the ZNF family of transcription factors are also down-regulated, many of which are known to be associated with the transcriptional regulation of repeat elements (see 1.3.17). The down-regulation of these genes involved in transcriptional regulation contributes to the overall deregulation of gene expression seen in HL through the loss of B cell specific transcriptional regulation.

The GO terms also highlight genes associated with chromatin regulation including, nucleosome assembly and chromatin silencing as being down-regulated. The down-regulation of these genes which are mainly histone coding may result from the slower cell cycle of HL cells compared to the control cell lines.

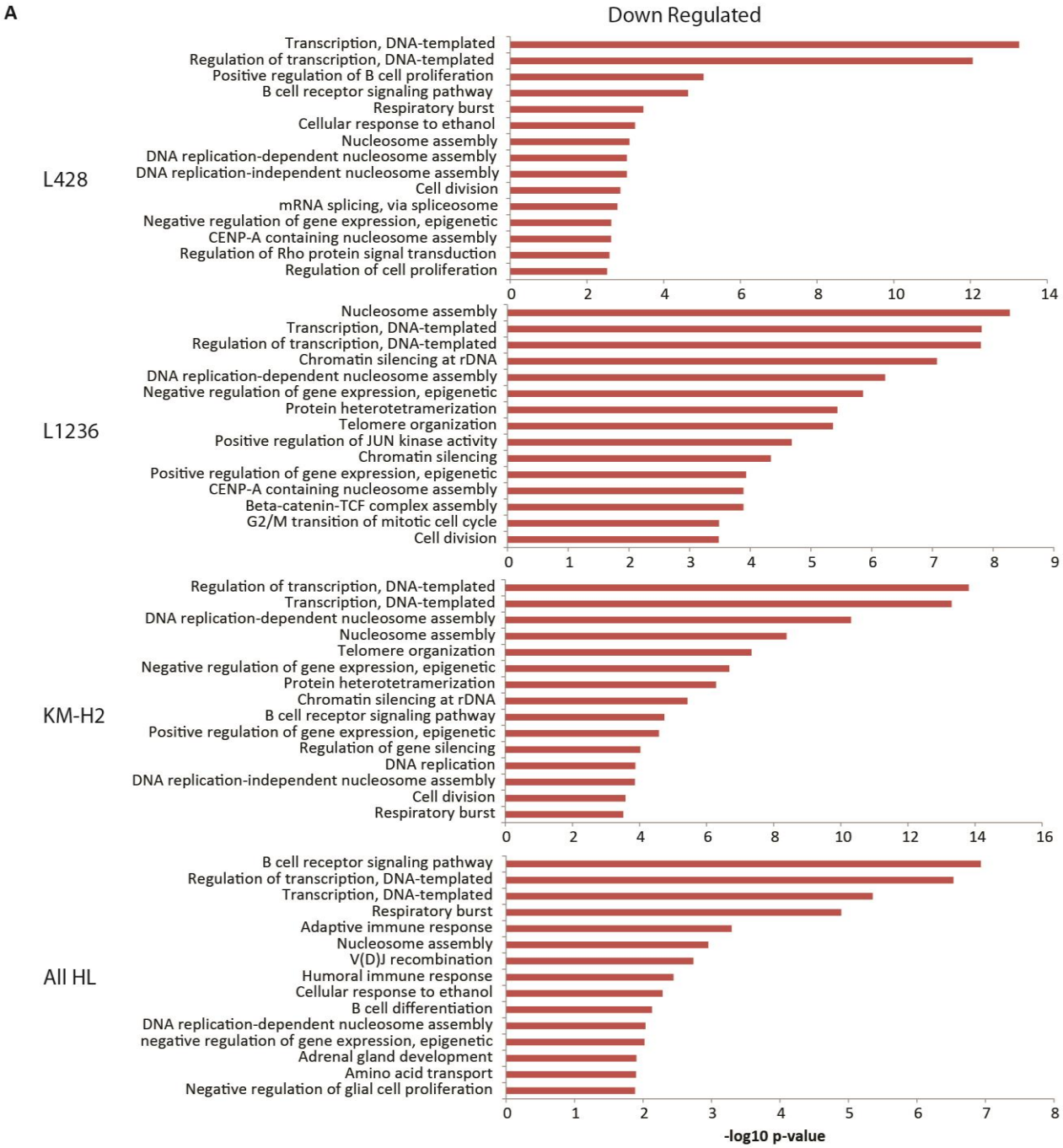
There was also a significant down-regulation of B-cell specific genes including those involved in B cell receptor signalling, V(D)J recombination and B-cell differentiation. These become particularly prominent when combining the downregulated genes from all HL cell lines. The B cell signalling genes included known cell surface markers of the B cell lineage including *CD19*, *CD38*, *CD79A* and *CD79B*. A down-regulation of the *RAG1* and *RAG2* genes both of which are important for V(D)J recombination in B cells was also observed. The decrease in expression of all of these genes shows the overall move from B cell lineage gene expression to a HL specific expression pattern.

The GO analysis also showed an upregulation of genes associated with cell adhesion and inflammatory and immune responses including response to cytokines and signal transduction. These ontology groups incorporated 172 up-regulated genes including all of the HL associated cytokines used in the expression validation panel (Figure 3.3). Based on the

known phenotype of HL cells, producing high levels of inflammatory cytokines and chemokines and producing an inflammatory microenvironment the upregulation of immune and inflammation genes is not entirely surprising.

There was also some up-regulation of genes associated with positive regulation of transcription and cell proliferation. Many of these genes were shared with the inflammatory response groups including *CXCL10*, *TNFRSF11A*, *CSF2* and *TNFSF13B*. There were also many transcription factors up-regulated including *SOX9*, *RUNX2*, *GATA3*, *GATA4*, *JUN* and *JUND*. The up-regulation of these factors represents a change in the regulation of gene expression from the B cell lineage with a HL specific regulatory network.

There was also evidence for an up-regulation of genes associated with negative regulation of apoptosis. The combination of this with up-regulation of transcription factors and growth factors may be at least in part responsible for the survival of HL cells after escaping apoptosis in the germinal centre. The up-regulated features appeared to remain common across all of the HL cell lines.


A
Down Regulated
L428
Transcription, DNA-templated
Regulation of transcription, DNA-templated
Positive regulation of B cell proliferation
B cell receptor signaling pathway
Respiratory burst
Cellular response to ethanol
Nucleosome assembly
DNA replication-dependent nucleosome assembly
DNA replication-independent nucleosome assembly
Cell division
mRNA splicing, via spliceosome
Negative regulation of gene expression, epigenetic
CENP-A containing nucleosome assembly
Regulation of Rho protein signal transduction
Regulation of cell proliferation
0 2 4 6 8 10 12 14
L1236
Nucleosome assembly
Transcription, DNA-templated
Regulation of transcription, DNA-templated
Chromatin silencing at rDNA
DNA replication-dependent nucleosome assembly
Negative regulation of gene expression, epigenetic
Protein heterotetramerization
Telomere organization
Positive regulation of JUN kinase activity
Chromatin silencing
Positive regulation of gene expression, epigenetic
CENP-A containing nucleosome assembly
Beta-catenin-TCF complex assembly
G2/M transition of mitotic cell cycle
Cell division
0 1 2 3 4 5 6 7 8 9
KM-H2
Regulation of transcription, DNA-templated
Transcription, DNA-templated
DNA replication-dependent nucleosome assembly
Nucleosome assembly
Telomere organization
Negative regulation of gene expression, epigenetic
Protein heterotetramerization
Chromatin silencing at rDNA
B cell receptor signaling pathway
Positive regulation of gene expression, epigenetic
Regulation of gene silencing
DNA replication
DNA replication-independent nucleosome assembly
Cell division
Respiratory burst
0 2 4 6 8 10 12 14 16
All HL
B cell receptor signaling pathway
Regulation of transcription, DNA-templated
Transcription, DNA-templated
Respiratory burst
Adaptive immune response
Nucleosome assembly
V(D)J recombination
Humoral immune response
Cellular response to ethanol
B cell differentiation
DNA replication-dependent nucleosome assembly
negative regulation of gene expression, epigenetic
Adrenal gland development
Amino acid transport
Negative regulation of glial cell proliferation
0 1 2 3 4 5 6 7 8
-log10 p-value

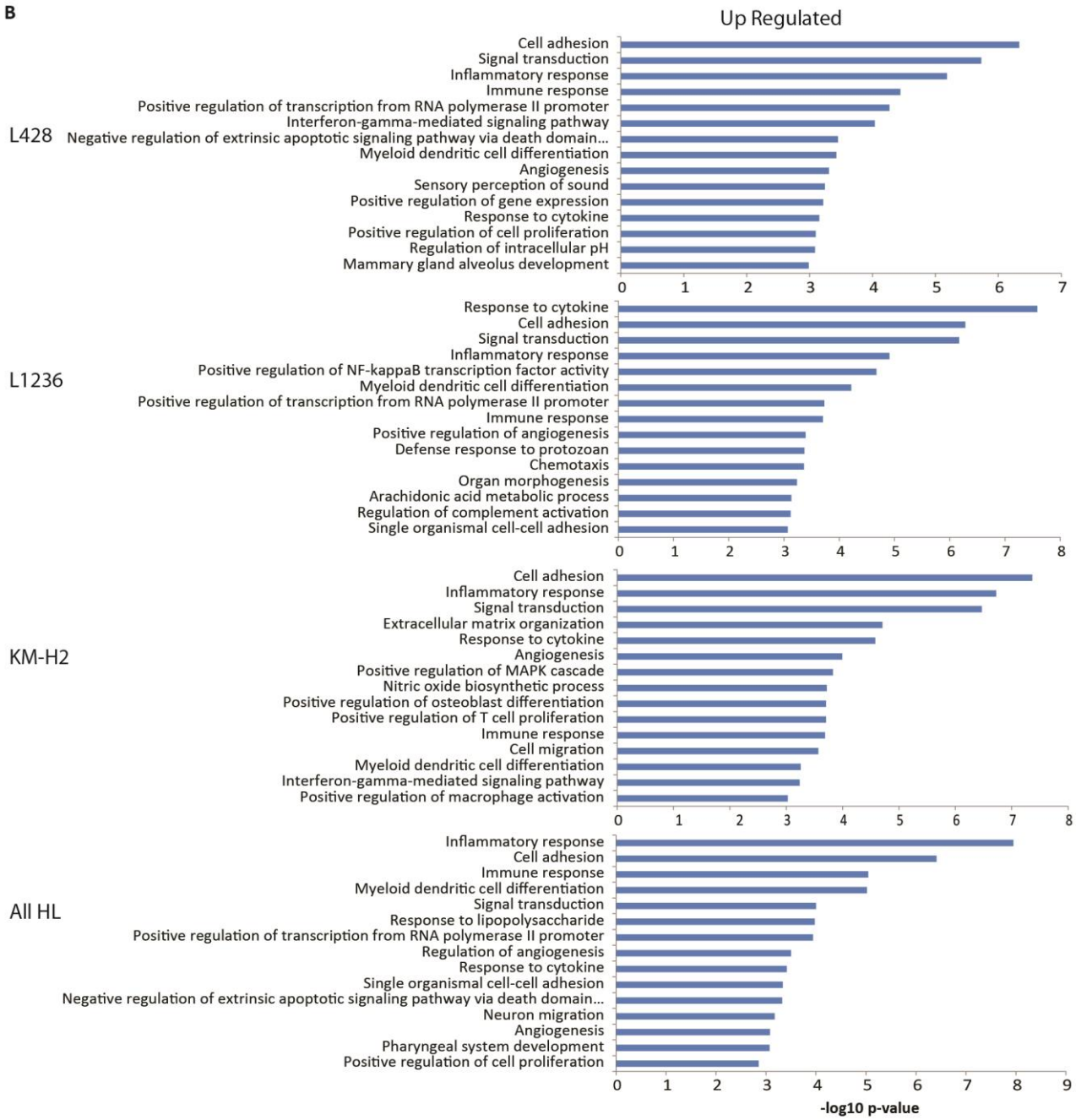


Figure 3.5. Gene Ontology analysis shows a down-regulation of genes associated with B cell processes in HL cell lines and an up-regulation of inflammatory and immune response genes.

Gene ontology analysis was carried out on genes which were at least 2-fold up- or down-regulated in HL cell lines (L428, L1236 and KM-H2) compared to 2 control cell lines (Reh and Namalwa). The gene ontology groups show the top 15 GO terms in each cell line for either up- or down-regulated genes. (A) HL cell lines show a downregulation of genes associated with transcription and chromatin regulation and B cell signalling and (B) an up-regulation of cytokine signalling and inflammatory response genes. When the list of down-regulated genes from all 3 HL cell lines was combined, B cell receptor signalling became the most down regulated group of genes (A - All HL).

To further examine the de-regulated pathways within the HL gene expression programme, KEGG pathway analysis was carried out to determine how the overall changes in gene expression impact on the functions of the HL cells (Figure 3.6). This analysis again showed a down-regulation of genes involved in B cell receptor signalling and primary immunodeficiency and a move of gene expression away from the B cell programme. The upregulated pathways give a very clear picture of the HL cellular environment and processes. These include cytokine receptor interaction, TNF signalling, NF-κB signalling and several other pathways involved in an inflammatory immune response. Again this supports the findings from the GO analysis and the contribution of gene expression to the cells inflammatory phenotype.

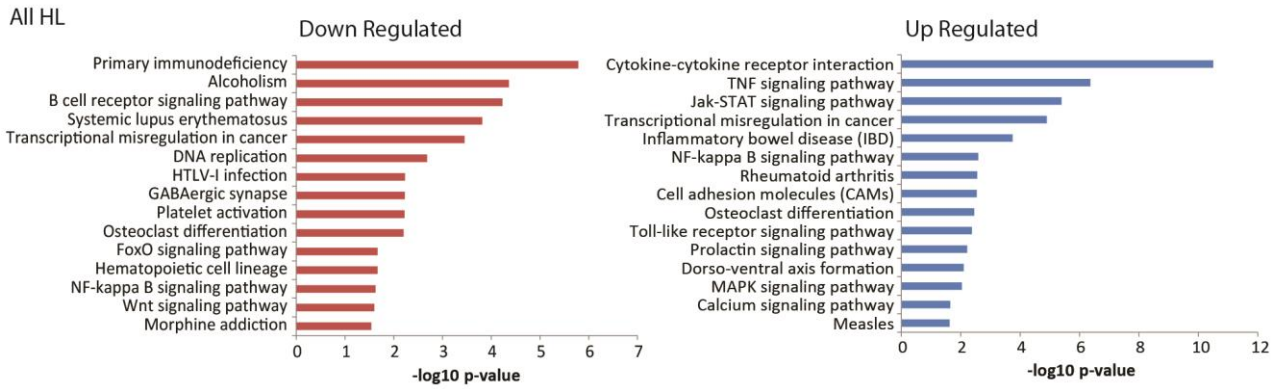


Figure 3.6. KEGG pathway analysis of up- and down-regulated genes in HL
The genes in all 3 HL cell lines (L428, L1236 and KM-H2) which are up and down regulated at least 2-fold compared to control cell lines (Reh and Namalwa) were subjected to KEGG pathway analysis. This analysis showed a down-regulation of genes in the B cell signalling pathways and an upregulation of inflammatory and immune signalling genes.

To understand how the identified KEGG pathways may be interacting, a gene expression network of KEGG pathways was produced showing the shared genes between pathways. This revealed that the gene expression programme was very much centred on cytokine-cytokine receptor signalling upregulation and interaction with other inflammatory signalling pathways including TNF, JAK-STAT and NF-κB (Figure 3.7).

It also showed the upregulation of many genes involved in transcriptional deregulation in cancer including *PLAU, SIX1, SIX4*, *CCR7* and *CSF2* which have all been previously associated with HL (Mathas, Hinz et al. 2002, Kuppers 2009, Nagel, Meyer et al. 2015, de Oliveira, Kaergel et al. 2016). We also observed the down-regulation of genes related to B cell signalling and primary immunodeficiency. However, the majority of genes related to primary immunodeficiency are also linked to B cell signalling and those that aren't such as RAG1 and RAG2 are involved in B-cell development. This analysis identifies a number of other genes which are deregulated in HL and may pose previously undiscovered pathological genes in HL (see 4.1).

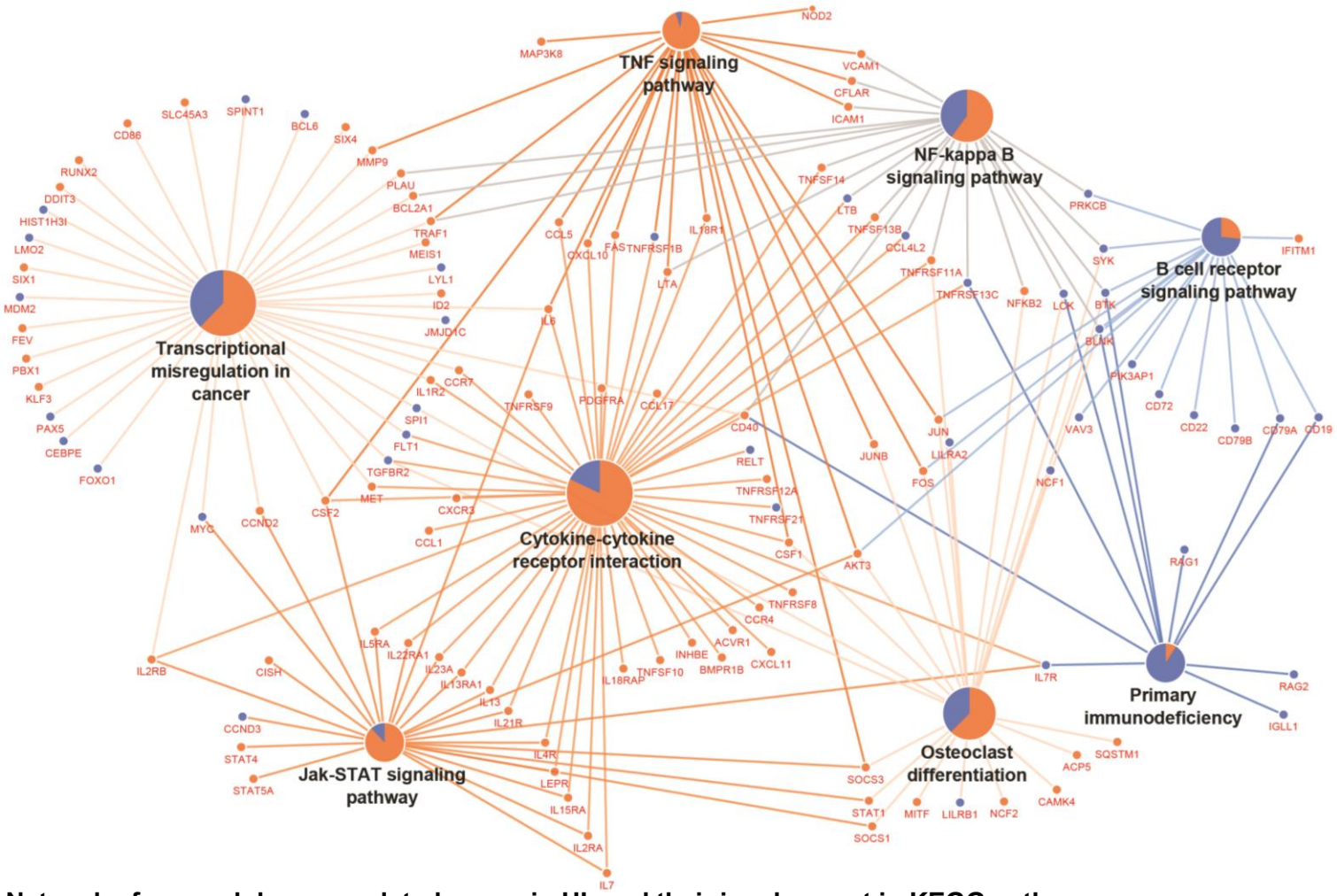


Figure 3.7. Network of up- and down-regulated genes in HL and their involvement in KEGG pathways

The results of KEGG pathway analysis linking 2-fold dysregulated genes in HL cell lines (L428, L1236 and KM-H2) to control cell lines (Reh and Namalwa were plotted as a network to show genes which are shared between multiple pathways. Orange represents genes which are up-regulated in at least one HL cell line compared to the control cell lines (Reh and Namalwa) and blue represents genes which are down-regulated. The pie charts and colour of the lines for each pathway indicate the proportion of the dysregulated genes which are up- (orange) or down-regulated (blue) in that pathway.

Characterisation of cis-regulatory elements driving Hodgkin's Lymphoma–specific gene expression - the HL transcriptional programme is driven by the transcription factors NF-κB and AP1.

Kreher et al. 2014 published an analysis using low resolution DNase-Seq to identify the cis-regulatory regions driving the unique gene expression profile of HL and comparing it to non-HL cells. This study identified enriched motifs within HL-specific DHS and used functional assays to identify IRF5 as a major regulator of the HL phenotype. However, this study was unable to identify the full complement of HL-specific transcription factors binding to HL-specific DHS. We therefore carried out DNaseI-Seq to identify regions of open chromatin, highlighting the presence of regulatory regions bound by transcription factors in HL cell lines, L428 and KM-H2 and control cell lines Reh and Namalwa (Figure 3.8). To test whether these sequences were actually occupied, we ran our DNase-Seq libraries at high sequencing depths to be able to conduct high resolution digital foot-printing experiments which identify transcription factor binding motifs.

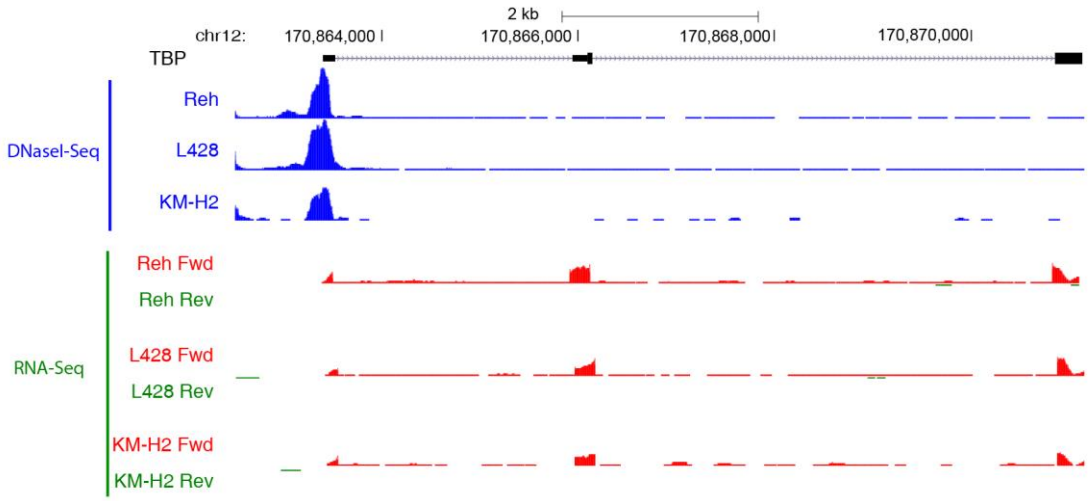


Figure 3.8 DNaseI-Seq data UCSC Genome Browser screenshot
Screenshot of UCSC genome browser showing DNaseI-Seq and RNA-Seq data at the TBP promoter in the Reh, L428 and KM-H2 cell lines demonstrating the low background of the high quality DNaseI-Seq data.

The sensitivity of open chromatin to nucleases was first demonstrated by Weintraub and Groudine, (1976) which lead to the development of genome wide DHS mapping (Weintraub and

Groudine 1976, Boyle, Davis et al. 2008). DNaseI hypersensitive sites have been shown to represent genomic elements including enhancers, promoters, locus control regions and insulators (Cockerill, Bert et al. 1999, Cockerill 2011).

The sequences existing as regions of open chromatin in each cell line were clustered and showed a high correlation between L428 and KM-H2 and also a high correlation between Reh and Namalwa (Figure 3.9). The HL cell lines again clustered separately to the control cell lines and showed little correlation with them, most of which is likely to be associated with shared housekeeping genes. This data confirms Kreher et al and demonstrates that HL-specific gene expression is driven by a specific set of cis-regulatory elements.

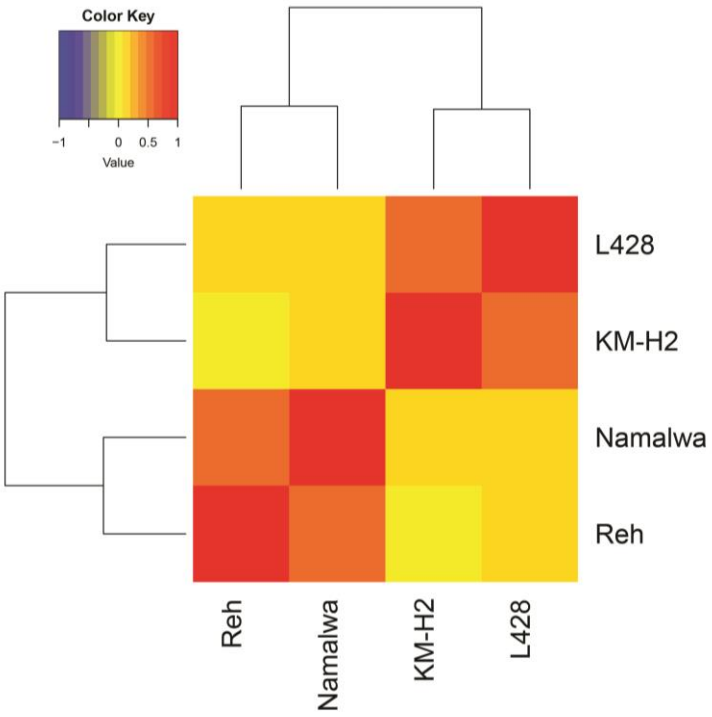


Figure 3.9. Correlation of HL and control cell lines based on chromatin accessibility. Regions hypersensitive to DNaseI, indicating regions of open chromatin, were determined by DNaseI-Seq. The sequences found in each of the cell lines were combined and the Pearson correlation of the tag counts for each site in each cell line was calculated. The correlation was then clustered by hierarchical unsupervised clustering showing that the HL cell lines (L428 and KM-H2) cluster together separately from the control cell lines (Reh and Namalwa).

As of today, only very limited information exists regarding the genome-wide binding activity of transcription factors involved in driving the unique expression profile in HL, most of which centred on NF-κB (de Oliveira, Kaergel et al. 2016). Most importantly, it was largely unclear how such factors would interact with each other. To understand the transcriptional regulation network in HL it was therefore important to establish which transcription factors were likely to be bound in HL-specific cis-regulatory regions. This information was determined by analysing our high-read depth DNAse-Seq data from Reh, L428 and KM-H2 cells using the Wellington algorithm to identify regions which were likely to be bound by transcription factors at the point of DNaseI digestion.

The binding of transcription factors to DHS's can be identified due to their presence protecting the DNA from cleavage. Piper *et al*. (2013) observed that in high read-depth DNaseI-Seq data a strand imbalance of aligned reads is present when a DHS is bound by a transcription factor. The Wellington algorithm is able to identify DHS's with bound transcription factors by analysing testing for this strand imbalance and has been shown to accurately predict foot-printed DHS's (Piper, Elze et al. 2013).

Using Homer (See 2.9.4) analysis of motif enrichment within foot-printed regions was performed to determine the transcription factors most likely to be producing the footprint. This analysis showed an enrichment of CTCF, AP1, NF-κB, NFY, SP1 and IRF in HL specific footprints. Analysis of the Reh specific footprints showed enrichment of NFY, SP1 and CTCF motifs which highlights AP1, NF-κB and IRF as the HL specific motifs (Figure 3.10). This result suggests that AP1, NF-κB and IRF have a significant involvement in the deregulated gene expression programme in HL and also links to the inflammatory signatures observed in the GO and KEGG pathway analysis (Figure 3.7).

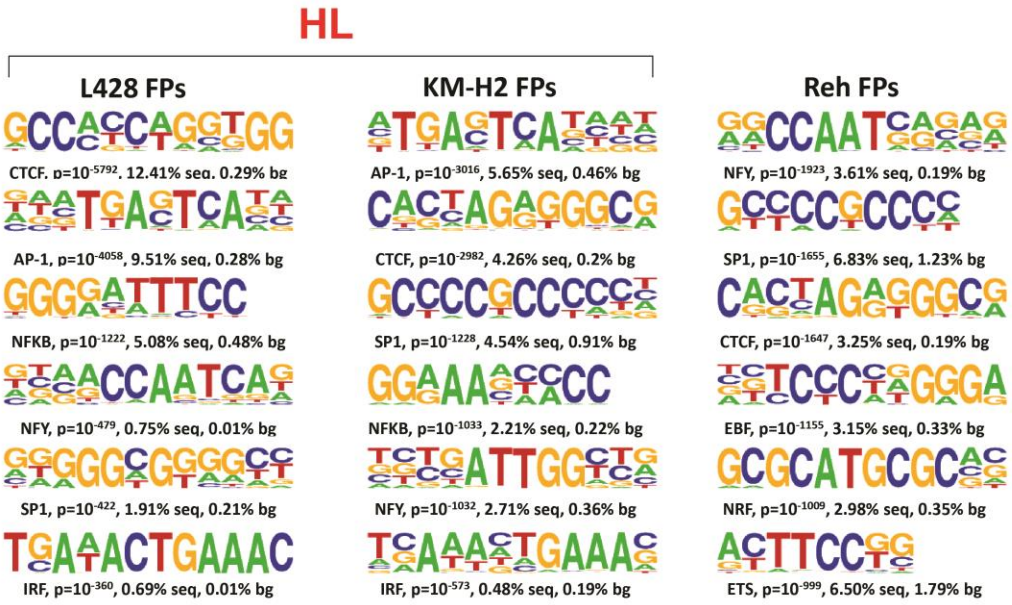


Figure 3.10. Transcription factor binding motif enrichment in DNase-I footprints in L428, KM-H2 and Reh
Digital foot-printing analysis was carried out on high read-depth DNaseI-Seq data from L428, KM-H2 and Reh cell lines using the Wellington algorithm (Piper, Elze et al. 2013). This method identifies regions with likely transcription factor binding. A motif enrichment analysis was then carried out on cell line specific foot-printed regions which showed AP-1, NF-κB and IRF were enriched at distal foot-printed DHS's in the HL cell lines. This indicates that these transcription factors likely play a role in HL gene expression.

To assess the co-localisation of motifs within foot-prints are therefore potentially the co-localisation of transcription factors which may interact a bootstrapping analysis was carried out. This analysis determines the co-localisation of motifs, selected from the motif enrichment analysis, within 50 bp of each other by comparing it to what would be expected by random based on the background of the sample. This showed that the control of gene expression in HL is dependent on AP-1 and NF-κB which have motifs co-localising with each other in addition to many other transcription factors known to be drivers of gene expression in HL (Figure 3.11).

It was also noted that motifs for the B cell specific transcription factor Pax5 were not enriched in the boot strapping analysis indicating its lack of involvement in HL. This is to be expected as Pax5 is an important transcription factor in B cell signalling and expression is lost in HL (Figure 3.12). This analysis was also repeated for KM-H2 specific footprint motifs and showed a similar result (Figure 3.11).

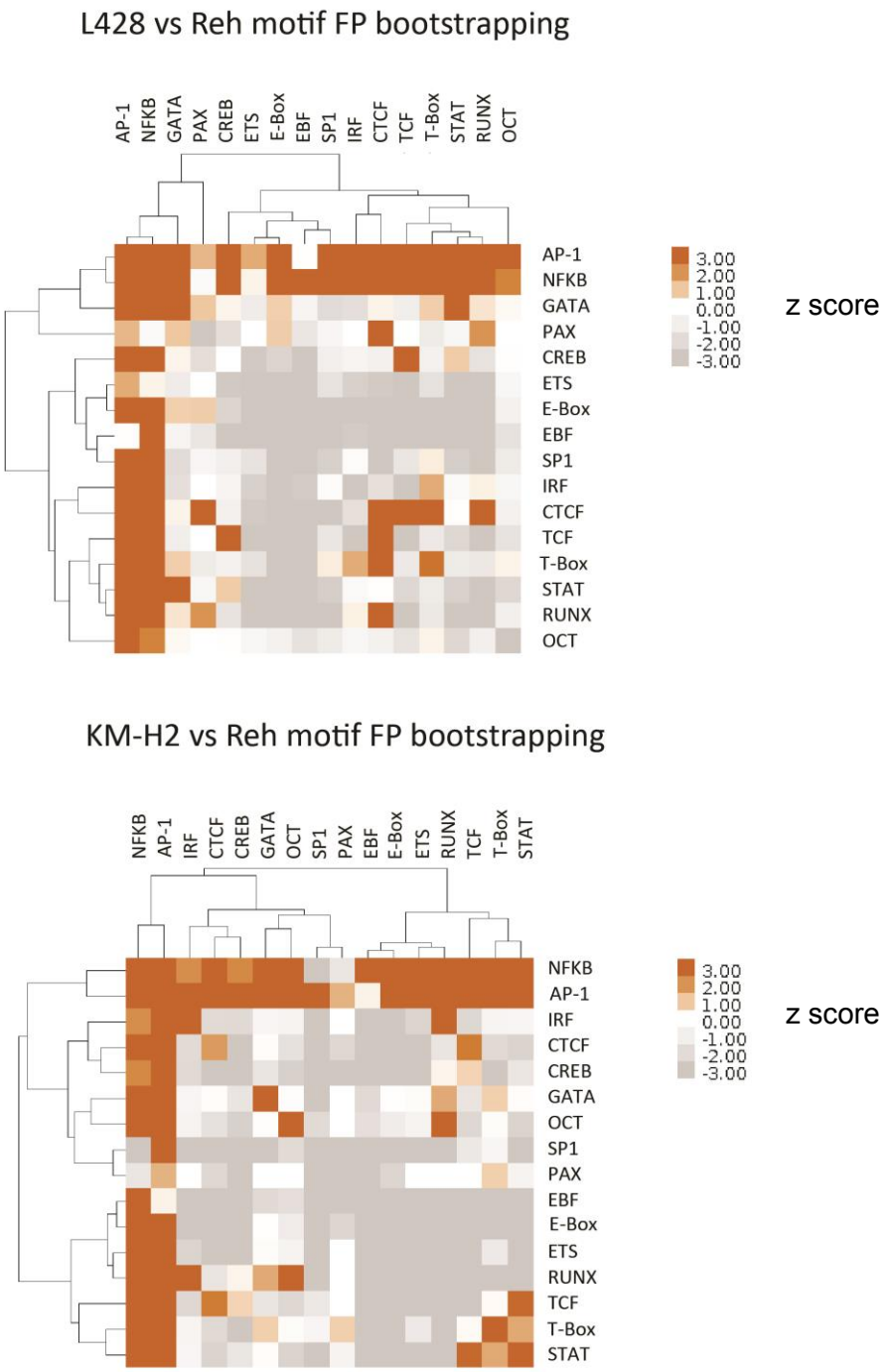


Figure 3.11. The transcriptional programme in HL cell lines is driven by the transcription factors NF-κB and AP-1

Bootstrapping analysis was performed on L428 and KM-H2 specific footprints in comparison to all Reh footprints to determine the co-localisation of transcription factor binding motifs. This showed a statistically significant co-localisation of the AP-1 and NF-κB motifs with many motifs of transcription factors known to be involved in HL. The implication of this was that the main transcription factors involved in the gene expression regulation in both HL cell line (L428 and KM-H2) was centred on NF-κB and AP-1.

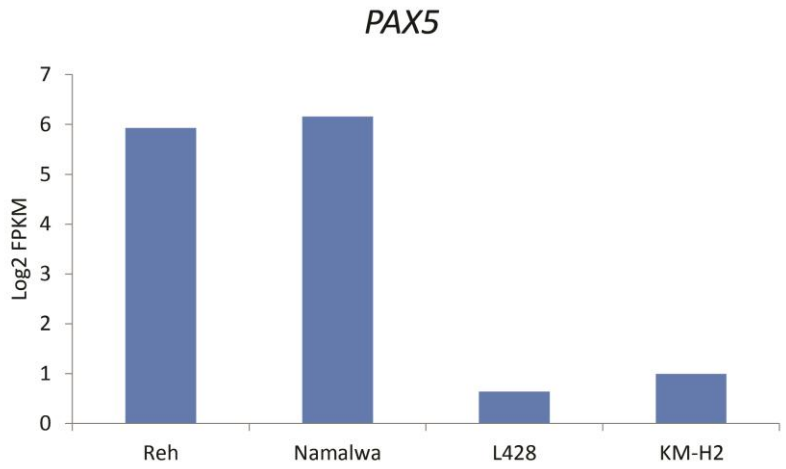


Figure 3.12. PAX5 expression is lost in HL cell lines
RNA-Seq Log2 FPKM values show a downregulation of *PAX5* expression in the HL cell lines (L428 and KM-H2), compared to the control B cell lines (Reh and Namalwa).

3.1.3. THE1B LTRs are more active in HL cell lines than control cell lines.

Lamprecht, *et al*. (2010) showed that an active LTR of the THE1B sub-family was acting as a promoter for *CSF1R* in HL, demonstrating that the activation of an LTR can impact on gene expression in HL and be required for cell survival. If LTR activation was a genome wide phenomenon it could contribute to and account for some of the gene expression patterns seen in HL. To assess whether this may be the case, the RNA-Seq reads around annotated THE1B LTRs were plotted. This analysis showed an increase in reads originating from THE1B LTRs in all 3 HL cell lines when compared to the control cell lines (Figure 3.13). There was also some variation in the level of LTR activation within HL cells which could suggest there is more THE1B LTR activation in KM-H2 and L428 than in L1236. The overall background level of reads around the annotated LTRs was also higher the HL cell lines. This would suggest that at least a proportion of the active THE1B LTRs also have up and downstream transcripts meaning that they may be acting as bidirectional promoters. It is also possible that the upregulated LTRs could act as enhancers for other genes even in the absence of downstream reads. This

increase in LTR activation could account for some of the changes in chromatin accessibility and gene expression seen in HL cells when compared to control cell lines.

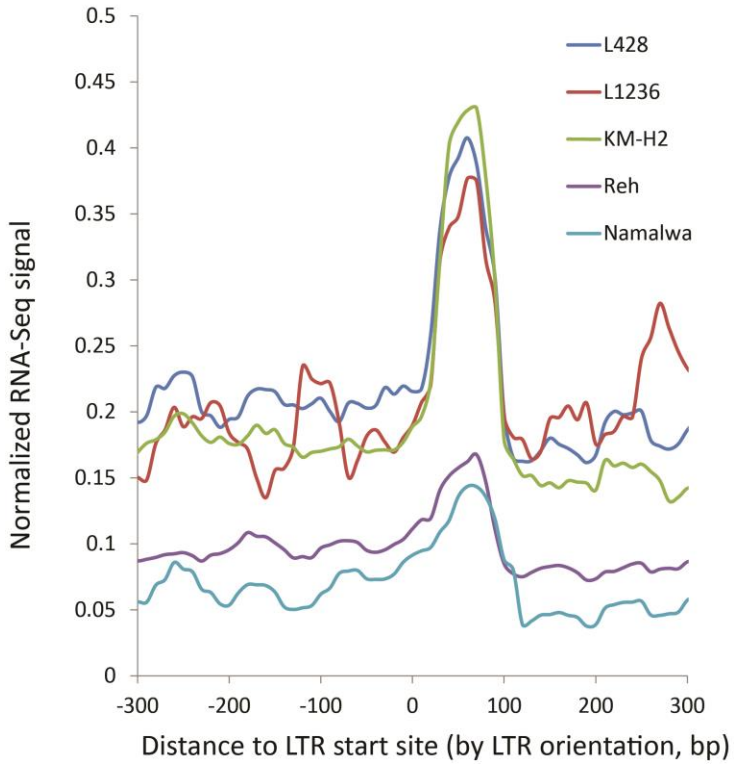


Figure 3.13. Intergenic THE1B LTRs are more active in HL cell lines than control cell lines Normalized RNA-Seq signals were plotted against intergenic THE1B LTRs annotated by repeat masker with reads centred on the LTR transcription start site and orientated by the annotated LTR direction. This analysis shows an increase in the overall level of transcription at LTRs in the HL cell lines (L428, L1236 and KM-H2) and also an increase in up- and down-stream transcription around the THE1B LTRs.

3.2. HL cells display a global activation of long terminal repeat elements.

3.2.1. CSF1R is expressed from an active THE1B LTR in HL cell lines

Following the identification of an overall increase in THE1B LTR activity based on the RNA-Seq data from the HL cell lines we wanted to further investigate which THE1B LTR elements were active in HL compared to the control cell lines. Lamprecht, *et al*. (2010) showed expression of *CSF1R* from an upstream THE1B LTR in HL. To confirm whether this finding could also be seen in our assays, qPCR was carried out using primers designed in exon 2 and 3 of *CSF1R* to assess the level of gene expression and also in the THE1B LTR and exon 2 of *CSF1R* to assess expression originating from the LTR (Figure 3.15). The qPCR results using the exonic primers showed a significant increase in expression of the *CSF1R* gene in the HL cell lines compared to the control cell lines (Figure 3.14). This expression originated from the LTR as there was also a significant increase in expression occurring from the THE1B LTR. The LTR transcript appears lower than the exon 2/3 transcript however this is likely to be an artefact of qPCR primer efficiency.

This experiment confirmed the presence of an active LTR upstream of *CSF1R* which has a transcript running into the second exon. As well as confirming the finding by Lamprecht *et al*., 2010 this established the use of qPCR with primers designed for transcripts between the LTR and gene body as a viable method for experiments assaying the molecular mechanisms of activation of THEIB LTR promoters.

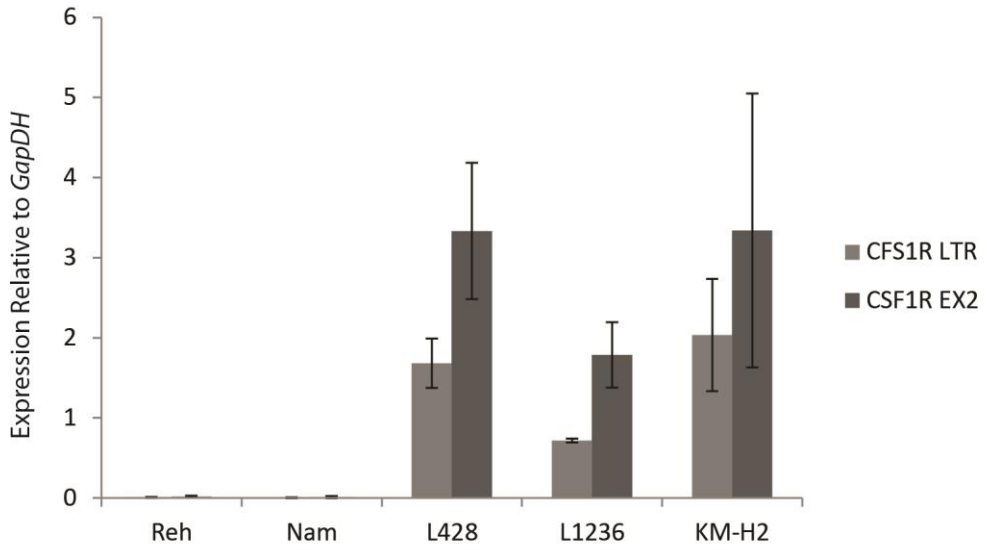


Figure 3.14. *CSF1R* is expressed from an upstream THE1B LTR in HL cell lines qPCR gene expression analysis was carried out using primers designed in exon 2 and 3 of *CSF1R* and in the upstream THE1B LTR and exon 2, the resulting values were normalised to *GapDH* expression. The primers within exon 2 and 3 and the primers within the LTR and exon 2 both showed a significant up-regulation (P<0.002) of expression in all 3 HL cell lines (L428, L1236 and KM-H2) when compared to control cell lines (Reh and Namalwa). N=3, Error bars show standard deviation.

3.2.2. 5' RACE-Seq can be used to identify active THE1B LTRs.

The analysis of RNA-Seq reads around annotated THE1B LTRs showed a genome-wide activation of LTRs in HL cell lines. To identify which particular LTRs were active genome-wide I developed a novel approach based on 5' Rapid Amplification of cDNA Ends (5' RACE) (See 2.4.3). The 5' RACE technique works by the addition of a known RNA adaptor sequence to the 5' end of all transcripts, effectively tagging the transcription start site (TSS). Following conversion to cDNA by reverse transcriptase a primer complementary to the adaptor sequence can be used along with a second primer which is specific to a region downstream of the TSS in a PCR reaction to amplify this fragment. In the traditional 5' RACE protocol this could then be

used to determine the size of the fragment and therefore the distance between the TSS and downstream region. The fragment could also be sequenced using Sanger sequencing to determine the sequence of the region between the TSS and downstream primer.

To enable all THE1B LTRs to be captured by this method a primer was used which was designed against the most homogenous region of the LTR sequence which is shared between many of the THE1B LTRs (Figure 3.15 & Figure 2.3). This technique allowed many active THE1B LTRs to be identified in a single assay effectively producing a library of active THE1B LTRs and was initially performed in the Reh, L428 and KM-H2 cell lines.

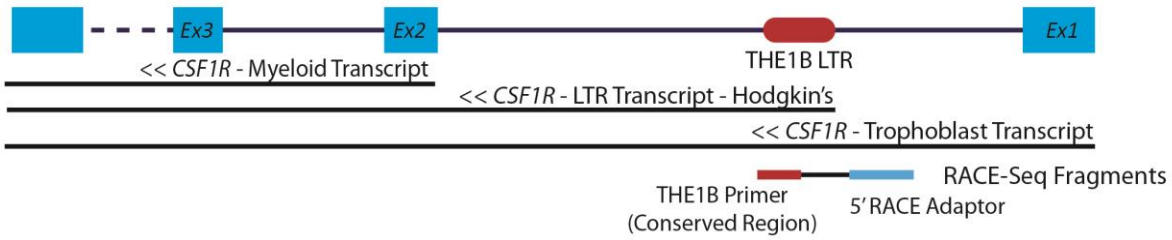


Figure 3.15. Scheme depicting *CSF1R* transcription start sites and 5' RACE strategy

Schematic showing the 3 potential transcription start sites of *CSF1R* in different cellular contexts. The RACE-Seq strategy is also shown, amplifying fragments between the TSS and a promoter designed within a conserved region of the THE1B LTR.

To initially examine the validity of the assay and to deconvolute this pool of active LTR DNA sequences we blunt end ligated isolated fragments into the pBluescript vector allowing individual clonal populations of bacteria to be obtained each containing a single LTR DNA sequence (see 2.4.2) This DNA was extracted from a number of clones and Sanger sequenced. Alignment of these sequences to the human genome showed that 6 out of 8 positive clones uniquely aligned to LTR elements. These included three THE1B elements and one MALT1A element from the L1236 cell line, one THE1B element from the L428 cell line and surprisingly one THE1D element from the Reh cell line (Table 3.1).

These cloning experiments not only validated the technique for identifying active LTRs but also showed evidence of active THE1B LTRs in HL other than the previously identified LTR. The clone from the Reh cell line showed that there was also clearly a level of LTR activation present in other B cell lines as well as the HL cell lines. The identification of a THE1D element in the Reh cell line and a MALT1A element in the L1236 cell line also demonstrated that the THE1B primer was able to detect other related members of the MalR family of LTRs.

Table 3.1 Multiple active MalR family LTRs can be identified using 5' RACE and a THE1B primer

Sanger sequencing and alignment to the human genome using BLAT shows that 5' RACE fragments produced by amplification using a primer designed to the THE1B LTR can be uniquely aligned to different regions of the genome.

Cell Line	Total Number of Positive Clones	Repeat Element	Genomic Position
Reh	2	THE1D	chr8:16314495-16314551
L428	1	THE1B	chr8:78849291-78849333
L1236	5	THE1B	chr2:155062150-155062203
		THE1B	chr4:38552326 -38552386
		MLT1A	chr5:106774949-106774968

We next wanted to make the 5' RACE assay truly genome wide to identify the full complement of active TH1B related LTRs in the HL cells. To achieve this we used Ilumina's short read next generation sequencing platform allowing for all of the LTR fragments in the pool of amplified RACE DNA to be sequenced simultaneously. The sequences were then uniquely aligned to the human genome and were uploaded onto the UCSC genome browser as well as being subjected to numerous downstream analyses. To initially validate that the aligned reads were identifying active THE1B LTRs the presence of an active LTR at *CSF1R* was confirmed which showed a peak in all 3 HL cell lines and no peak in the control cell lines (Figure 3.16). The identification of a known active LTR confirmed that in principle the RACE-Seq technique was successful.

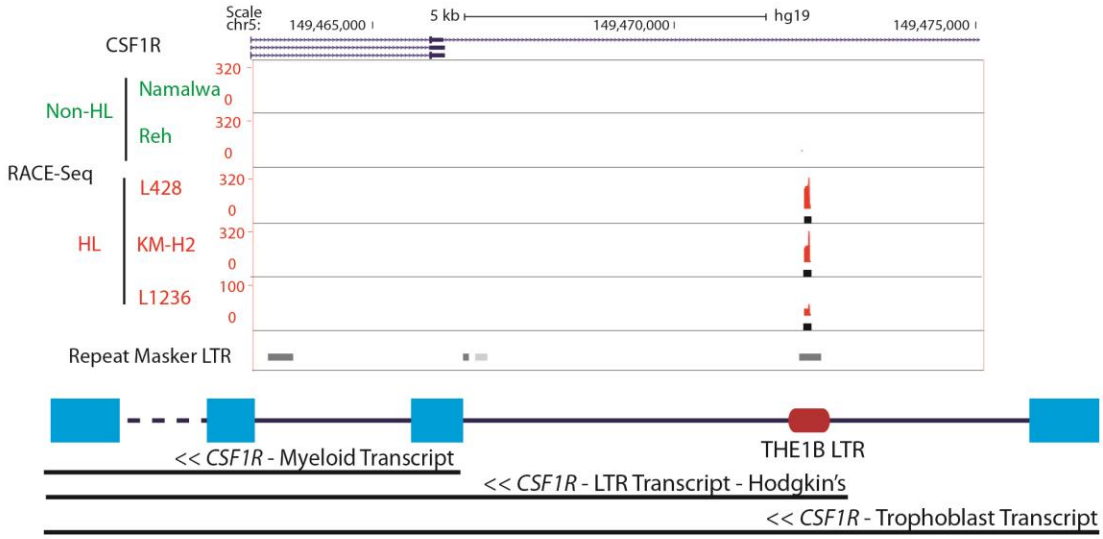


Figure 3.16. The THE1B LTR which acts as a promoter for *CSF1R* in HL can be detected by RACE-Seq
Alignment of RACE-Seq fragments to the human genome displayed on the UCSC genome browser. The alignments produce a peak at the THE1B LTR (annotated by the Repeat Masker track) upstream of the CSF1R gene. This peak is present in all 3 HL cell lines (L428, L1236 and KM-H2) and absent in the 2 control cell lines (Reh and Namalwa). The black bars below the aligned reads represent peaks called by MACs (See 2.9.5).

3.2.3. THE1B RACE-Seq identifies other associated MalR sub-families of repeats.

Before commencing further analyses and experiments we wanted to ensure that the RACE-Seq technique identified active LTRs reproducibly in each cell line. To this end 3 independent replicates were performed in each of the 5 cell lines. Overlap of the replicates showed that only between 7% and 30% of identified peaks were shared between all 3 replicates (Figure 3.17). We hypothesised that the reason for this low reproducibility, besides technical variation in the processing of samples, could have been the fact that we were using a partly degenerate consensus primer for the first amplification. Variation in sequence of the related THE1B family members would display weaker binding of the THE1B primer and may not have always been detected. However, selecting only LTRs which were associated with THE1B elements for

analysis did not improve the replicate overlap. Another possibility was that the less active LTRs could be harder to detect and therefore may have been missed in replicates, however, again selecting the peaks with highest read counts again did not improve the overlap. We believe that the reason for replicate variation was most likely down to the binding nature of the THE1B primer. The experimental procedure was optimised to allow annealing of the primer with mismatches to capture the widest array of MalR active LTRs possible. The sequencing data showed that overall only 10% of peaks had no mismatches in the primer sequence, half of the sequences had up to 5 mismatches and the remainder had up to 8 (Figure 3.18). This means that because of the clonal nature of the PCR reaction used during RACE the variation of annealing in the first PCR cycle dictates the LTRs which are detectable at the final stage, introducing a stochastic element into the assay. The LTRs shared by all replicates are most likely those which the primer anneals to with fewest mismatches.

Although the proposed theory for replicate variation would suggest that all identified peaks are genuinely active LTRs, we used a conservative approach of only carrying out further analysis on those active LTRs which were identified in at least 2 out of the 3 replicates (Figure 3.17).

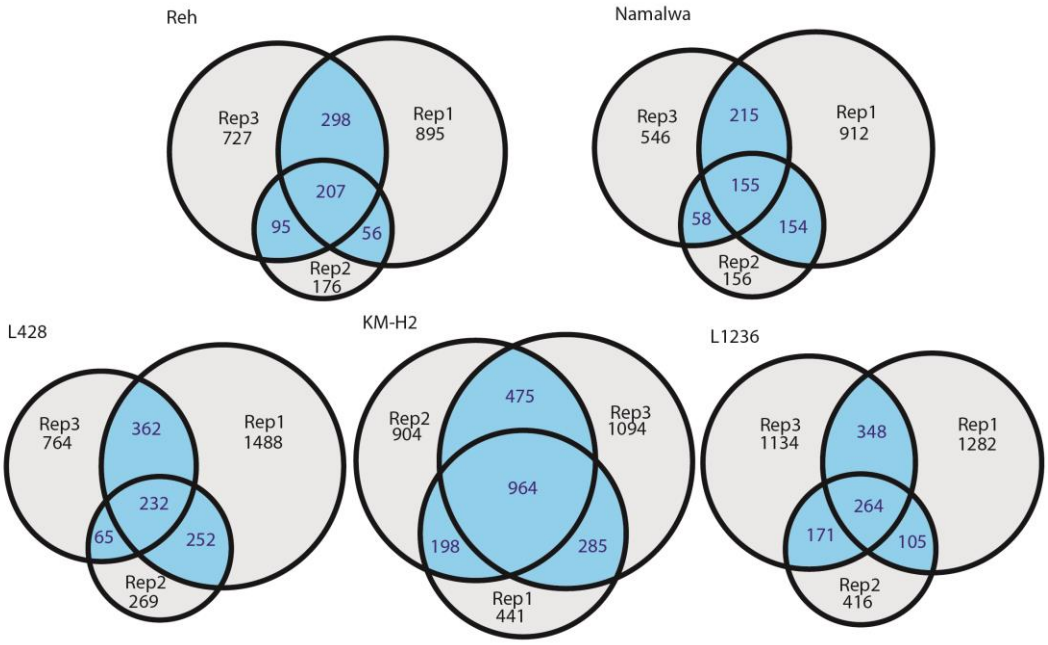


Figure 3.17. Overlap of biological replicates of LTRs detected by RACE-Seq Peak regions representing active LTRs were determined from RACE-Seq fragment alignments and overlaps of the peaks between 3 biological replicates for each of the 5 cell lines were performed. The areas highlighted in blue represent those peaks which are shared by at least 2 replicates in a cell line ($p<0.01$) and are the peaks which were selected for further downstream analysis.

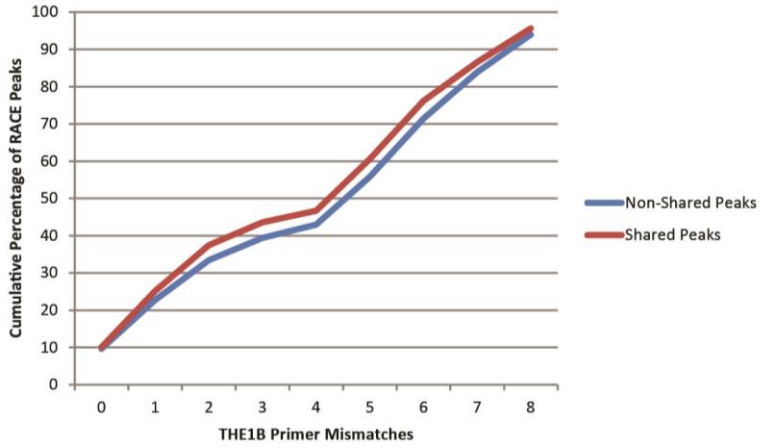


Figure 3.18. The THE1B primer anneals with multiple mismatching bases during the RACE-Seq PCR

The number of mismatches between the primer sequence and the identified active LTRs was determined by comparison of the primer sequence to the genome transcript aligning to the primer sequence in the vicinity of the aligned fragments. The overall cumulative percentage of RACE peaks based on the number of mismatches to the primer sequence was plotted showing that up to 5 mismatches were present in 50% of identified RACE peaks and over 5 mismatches were present in the remainder.

From the original cloning experiments we knew that the THE1B RACE primer was able to recognize and cause amplification of other associated members of the MalR family of LTRs. To determine the array of repeat elements in the RACE-Seq data, the aligned sequences were annotated to repeat elements using the hg19 RepeatMasker data. This annotation showed an array of MalR LTRs including, THE1B, THE1D, THE1C, THE1A, MSTA and MSTB. A predominance of THE1B elements was seen in all cell lines, however with a lower overall percentage in the control cell lines (33%) as compared to the HL cell lines (58%). As the primer is based on the THE1B sequence it is expected that THE1B LTRs would primarily be observed and it is likely that the full array of other active MalR LTRs is not captured. The lower proportion of THE1B activity in the control cell lines could be related to a different activation mechanism resulting in both a smaller number and different types of repeat elements being active (Figure 3.19). This analysis confirmed the activation of many MalR LTRs throughout both the HL and control cell genomes.

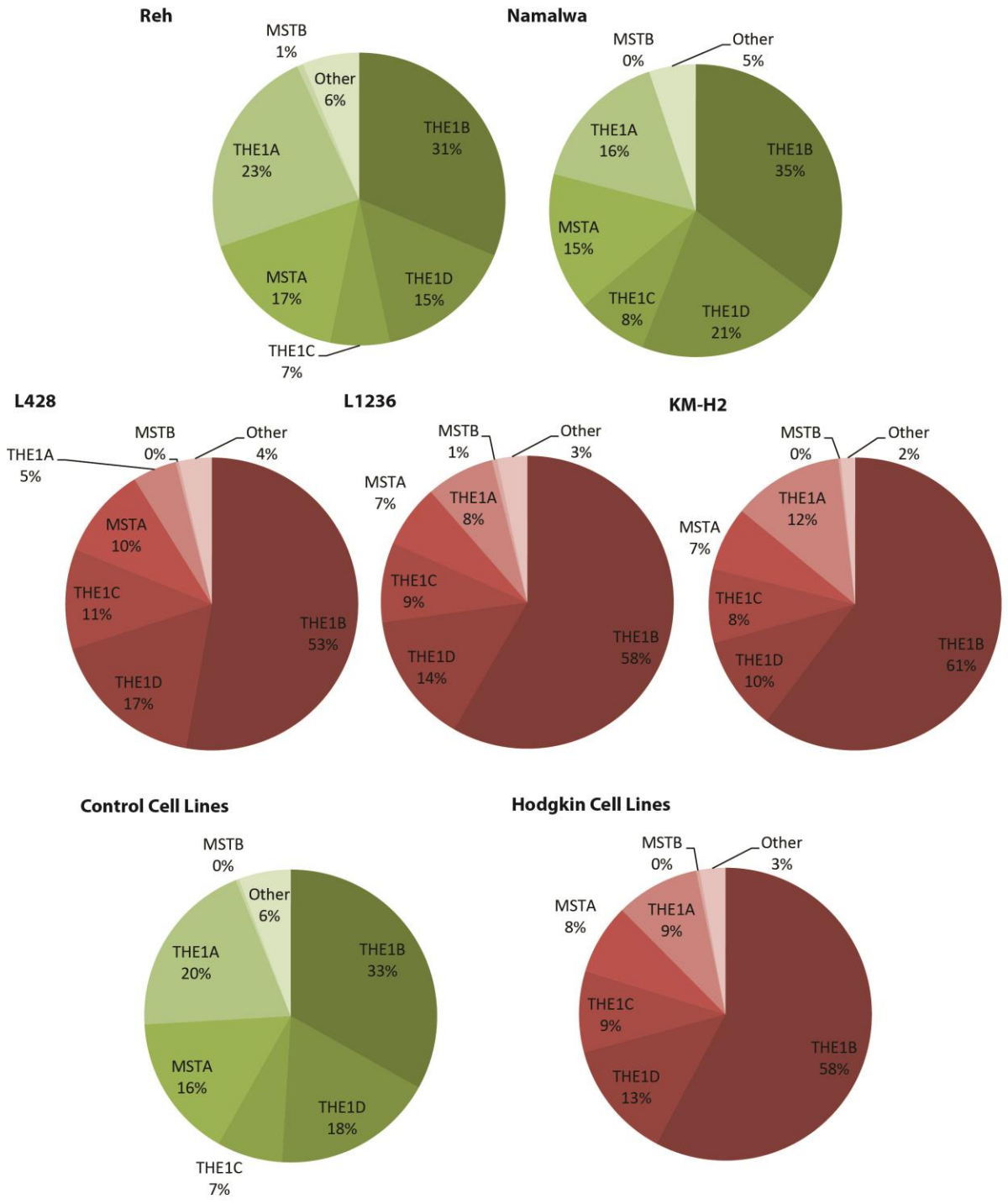


Figure 3.19. RACE-Seq using a THE1B primer identifies a wide range of MalR repeat elements

Overlap of THE1B RACE-Seq peaks in HL and control cell lines with the Repeat Masker hg19 annotation shows a prevalence of THE1B elements but also a number of other repeat element types of the MalR family in each of the HL and control cell lines. Averages across all HL and all control cell lines also show a similar pattern.

3.2.4. HL cell lines have a unique activation pattern of LTRs

The alignment data from the RACE-Seq experiments showed an activation of LTRs in both the HL and control cell lines. To determine whether a HL specific pattern of LTR activation existed we overlapped the peaks from all 3 HL cell lines with the peaks from the 2 control cell lines. This analysis showed the presence of 2822 HL specific active LTRs and 340 active LTRs specific to the control cell lines. The presence of a HL unique pattern of LTR activation suggests LTRs may play a significant role in the regulation of the HL genome. The control cell lines unique LTR activation also suggest that LTRs may play a role in these cell lines, however, at a lower level than in HL.

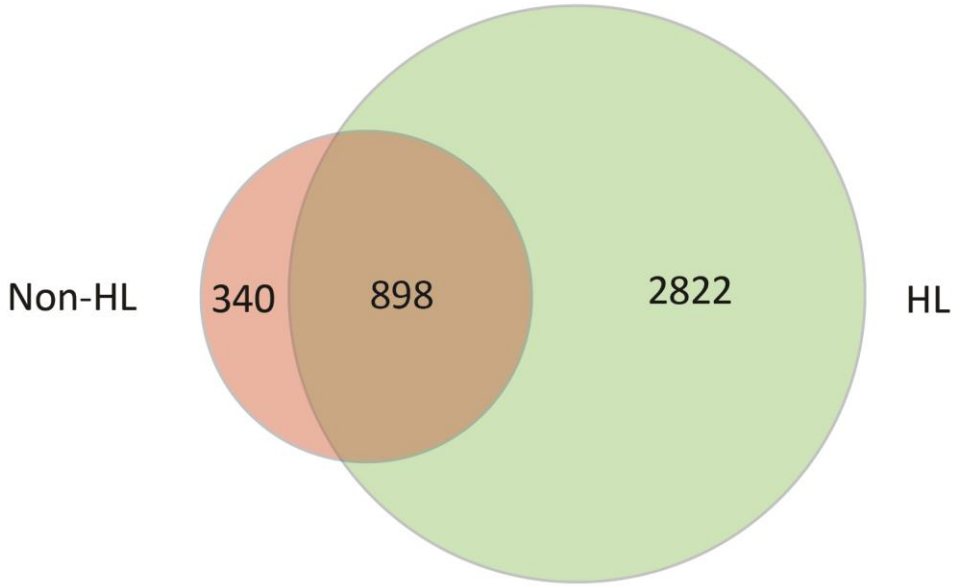


Figure 3.20. HL and control cell lines have a unique pattern of LTR activation Overlap of active LTR sequences detected by RACE-Seq in 3 HL cell lines (L428, L1236 and KM-H2) (green) and 2 control cell lines (Reh and Namalwa) (Red). A unique LTR activation pattern is observed in both the HL cell lines are the control cell lines.

To further investigate the pattern of LTR activation in HL, the RACE-Seq peaks from active LTRs in each of the HL cell lines were compared. Although there was overlap, this particular analysis showed a number of unique active LTRs in each of the HL cell lines with 411 active LTRs shared between all lines. Given the partly stochastic nature of the assay the extent of the overlap is likely to vary, nevertheless based on hypergeometric analysis there was significant variation between the cell lines ($p<0.01$). This feature may have developed as a result of them being immortalised cell lines that have acquired mutations through many rounds of replication. However, it is also possible that all 3 cell lines represent different subtypes of HL which may have different survival and proliferation mechanisms promoted by the many active LTRs.

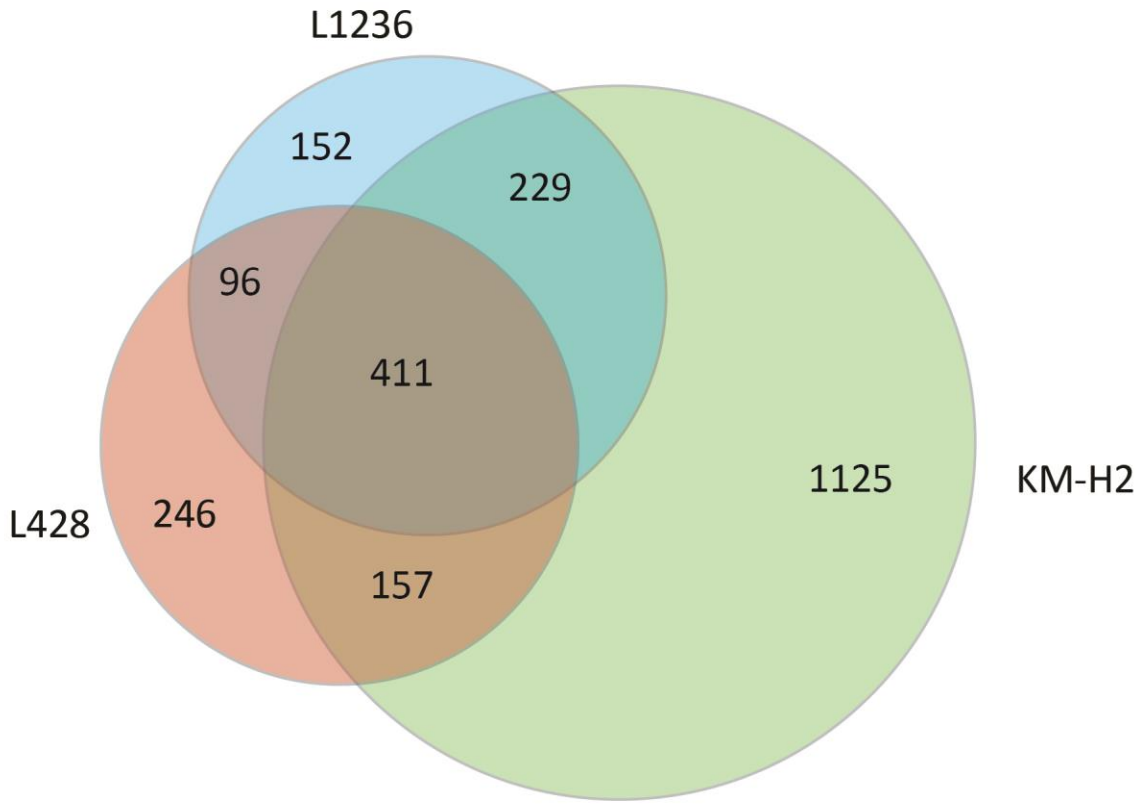


Figure 3.21. Comparison of active LTRs in three HL cell lines

Overlap of the active LTRs detected by RACE-Seq in each of the 3 HL cell lines shows a shared group of 411 active LTRs between all HL cell lines and also a large group of unique peaks for each cell lines representing a unique LTR activation pattern.

To visualise the overlap of active LTRs between each of the 5 cell lines a DICE index score was calculated for each pair showing the similarity in LTR activation between each pair. This method allowed for the correlation between the binary data sets (presence or absence of active LTRs) to be clustered to determine the relationship between the cell lines based on LTR activation (Figure 3.22). The clustering showed clearly that the control and HL cell lines clustered separately and that the HL cell lines clustered together, confirming a specific LTR pattern in HL cells as compared to non-HL cells. Based on LTR activation the L1236 and L428 cell lines cluster more closely together than with KM-H2. It is possible this could be an artefact of sequencing read depth with a higher read depth in KM-H2 resulting in the discovery of more active LTRs which may have been missed in the L428 and L1236 cell lines. Although the Reh and Namalwa cell lines cluster together the similarity based on DICE index is not as high as between the HL cell lines. As both of the control cell lines are derived from patients with different diseases, B cell leukaemia (Reh) and Burkitt's Lymphoma (Namalwa), this may point towards a specific LTR activation pattern being present in many different cancers.

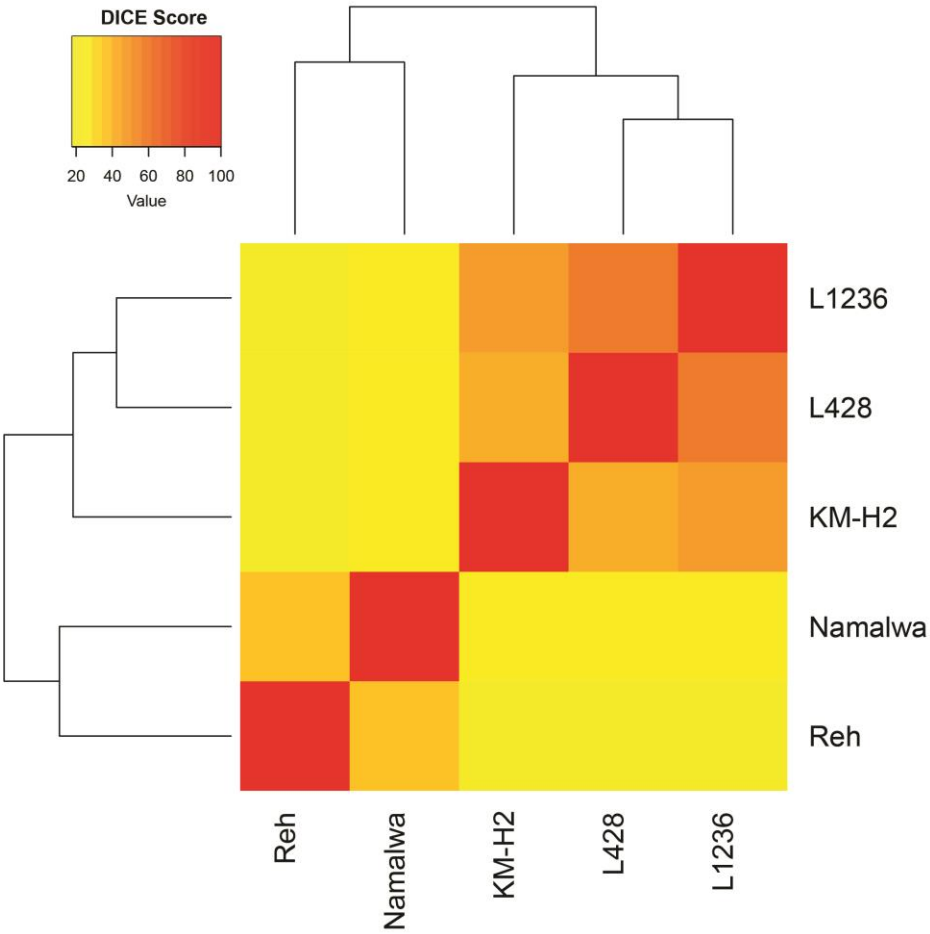


Figure 3.22. HL cell lines correlate based on LTR activation determined by RACE-Seq DICE index scores were calculated for each pair of cell lines based on active LTR presence identified by RACE-Seq allowing for correlation of LTR activation between the cell lines to be determined. Clustering analysis shows that the HL cell lines (L428, L1236 and KM-H2) cluster together and the control cell lines (Reh and Namalwa) also cluster together but separately from the HL cell lines representing the disease specific LTR activation pattern.

3.3. LTR activation contributes to global deregulation of gene expression in HL cell lines.

We know from previous studies and our own RNA-Seq data that HL cells have a vastly different gene expression pattern to cells from the B cell lineage and that they cluster separately based on gene expression (Stein, Marafioti et al. 2001, Torlakovic, Tierens et al. 2001) (Figure 3.2). It has been shown in many studies that LTRs can act as promoter and enhancer elements (See 1.3.13), therefore the genome-wide activation of LTRs observed in HL could be at least in part responsible for the HL gene expression pattern. Lamprecht, *et al*. 2010, showed that *CSF1R* in HL is expressed from an active LTR promoter, so we wanted to determine whether other genes in HL were also being expressed from active LTRs.

3.3.1. Active LTRs are mainly located in intergenic and intronic regions

To elucidate the overall impact of active LTRs we firstly determined the genomic regions in which active LTRs could be detected. This allowed us to establish the proportions of active LTRs which were already annotated as promoters and exonic elements and whether there was a bias of LTR activation to specific regions.

All cell lines showed a predominance of active intergenic LTRs, with an average of 63% in the HL cell lines and 56% in the control cell lines (Figure 2.22). This finding indicated that the active LTRs were mainly located in regions, which have potential to act as upstream promoters for genes. A large proportion of the active LTRs were also intronic, an average of 29% in the HL cell lines and 24% in the control cell lines. As in the case of *CSF1R,* these LTRs may also be acting as alternative promoters for shorter isoforms of genes either expressing an additional isoform of an already expressed gene or by being the sole promoter inducing expression of a gene.

The percentages of intergenic and intronic active LTRs are comparable to the overall distribution of all MaLR family LTRs (Figure 3.23). Notably the promoter and exon annotated LTRs were observed at a higher percentage in the active LTRs in HL cell lines and to an even greater extent in the control cell lines. The overall bias towards LTRs being located in intergenic regions is most likely due to insertional bias. It has been reported in a number of studies that viral integrations into the genome are more prevalent in open chromatin and at regions with active histone marks (reviewed by (de Jong, Wessels et al. 2014)), therefore a bias towards LTR integration within intergenic regions would be expected. The active LTRs in intergenic regions identified in HL which reside in active genes may also be examples of LTRs which have been repurposed for regulation of endogenous gene expression.

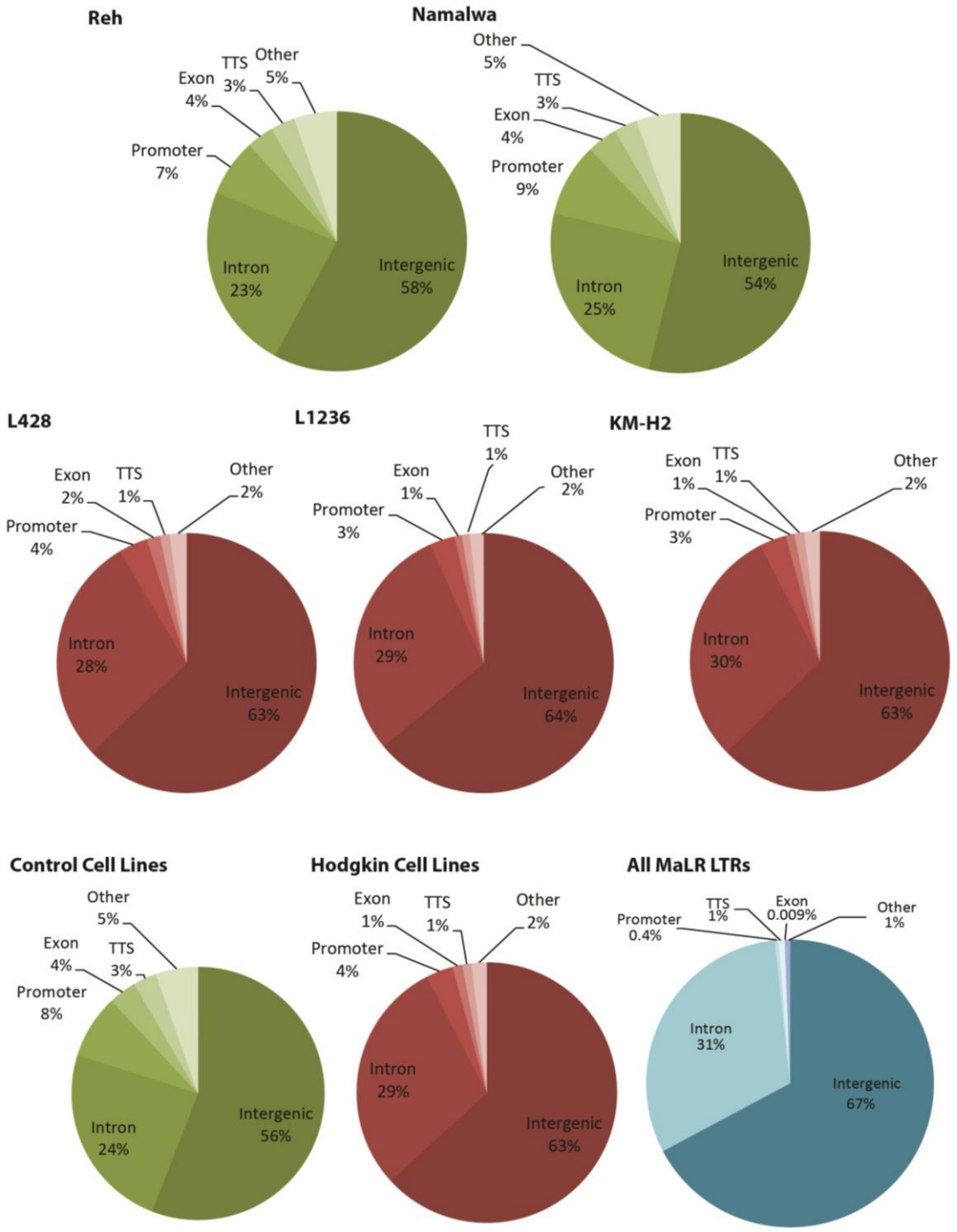


Figure 3.23 Active LTRs are primarily located in intergenic and intronic regions in HL and control cell lines

Active LTRs identified by THE1B RACE-Seq in the HL and control cell lines were annotated to genomic regions, this showed that active LTRs are mainly intergenic and intronic. Annotation of all MaLR family LTRs annotated by Repeat Masker showed a similar distribution of LTRs in the intergenic and intronic regions and far less in the promoter and exonic regions than was seen in the HL and control cell lines.

3.3.2. Active LTRs in HL act as alternative promoters

We next wanted to establish whether the active LTRs were acting as promoters based on histone marks. It is known that most actively transcribed promoters are marked with H3K4me3 (Liang, Lin et al. 2004). To study this in our genome-wide LTR activation datasets we performed chromatin immunoprecipitation with sequencing (ChIP-Seq) for H3K4me3, in the L428 and Reh cell lines. The ChIP signal around active LTRs was overall very low, however, overlaps of H3K4me3 peaks with the RACE-Seq datasets for the corresponding cell lines showed that 5.03% of Reh active LTRs and 11.86% of L428 active LTRs corresponded to genomic regions containing H3K4me3. This finding would suggest that only a small proportion of the active LTRs are active promoters and the remainder may be acting as enhancers which may have different histones marks. Although the presence of histone marks other than H3k4me3 is a possibility for some active LTRs the overlaps performed may be an underestimate due to the difficulty of aligning ChIP-Seq data uniquely to repeat elements with shared sequence. The difference observed between the Reh and L428 cell lines may also indicate that a higher proportion of the active LTRs are acting as promoters in HL when compared to the control cell line.

Finally to further confirm the activity of LTRs as alternative promoters we integrated the RACE-Seq data with the RNA-Seq data to study the downstream transcripts originating from active LTRs. The nature of the RACE assay means that strand information is lost during library preparation. The disadvantage of this being that if an LTR is acting as a promoter the strand information would help to infer which nearby gene may be affected. It has however been shown in a number of studies that LTRs can act as bi-directional promoters.

In the same way as ChIP-seq the alignment of RNA-Seq to repeat sequences is challenging, however as shown in figure 3.13 it is possible to detect active LTRs on a genome-wide basis .

To investigate the transcripts originating from the active LTRs strand information was inferred from the Repeat Masker annotation.

RNA-Seq reads were centred on the active LTRs and a peak corresponding to the active LTR region was seen in the average profiles for all cell lines Figure 3.24. Further to this there was a notably higher level of downstream transcription in the HL cell lines suggesting that the LTRs are functioning as promoters. There was little evidence of transcripts in the anti-sense direction, however if only a small number of promoters were acting bi-directionally these transcripts could be masked in the profiles.

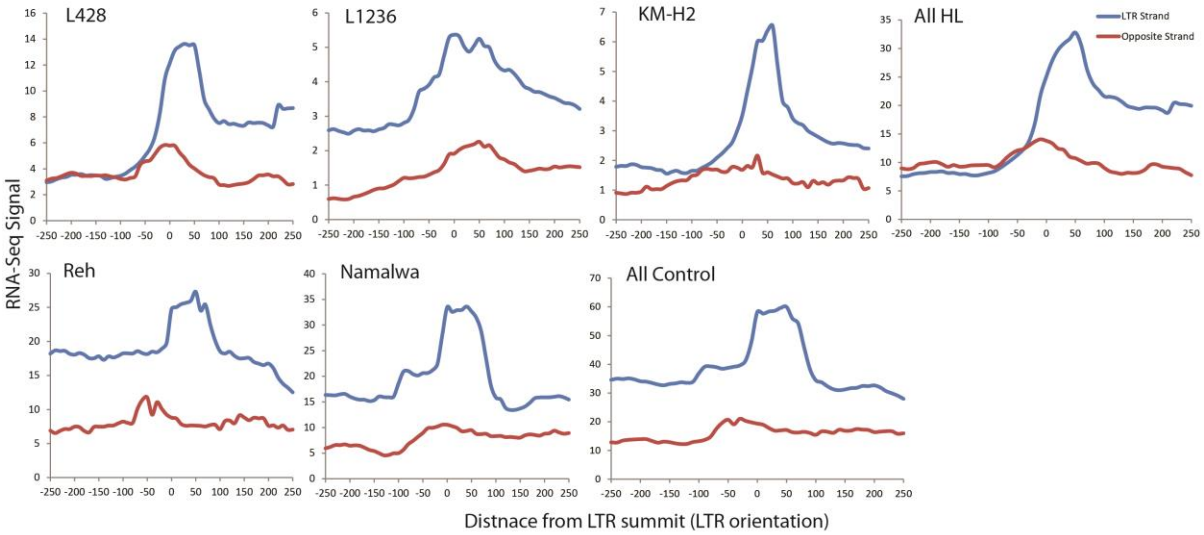


Figure 3.24. Active LTRs produce downstream RNA transcripts

Average profiles produced using RNA-Seq data centred on active LTRs identified by THE1B RACE-Seq and orientated based on LTR strand annotated by Repeat Masker show an increased level of transcription around active LTRs in both HL and control cell lines indicating the potential activity of LTRs as promoters.

3.3.3. LTR activation in HL contributes to global deregulation of gene expression

Following the observation that a proportion of LTRs were marked as promoters by H3K4me3 and that on average active LTRs have downstream transcripts, in at least the HL cell lines, we wanted to determine the impact on expression of genes in the vicinity of active LTRs. To

achieve this we mapped the active LTRs in each cell line to their closest gene and plotted the expression of the genes in relation to the control cell lines (Figure 3.25) (See supplementary tables 1-6 for full list of closest genes). This analysis showed an increase of active LTRs close to the genes which are up-regulated in each of the HL cell lines compared to the control cell lines. The same pattern was not generally observed for the active LTRs specific to the control cell lines. There was however a slight increase in active LTRs close to the Reh and Namalwa specific genes when compared to the L428 cell line.

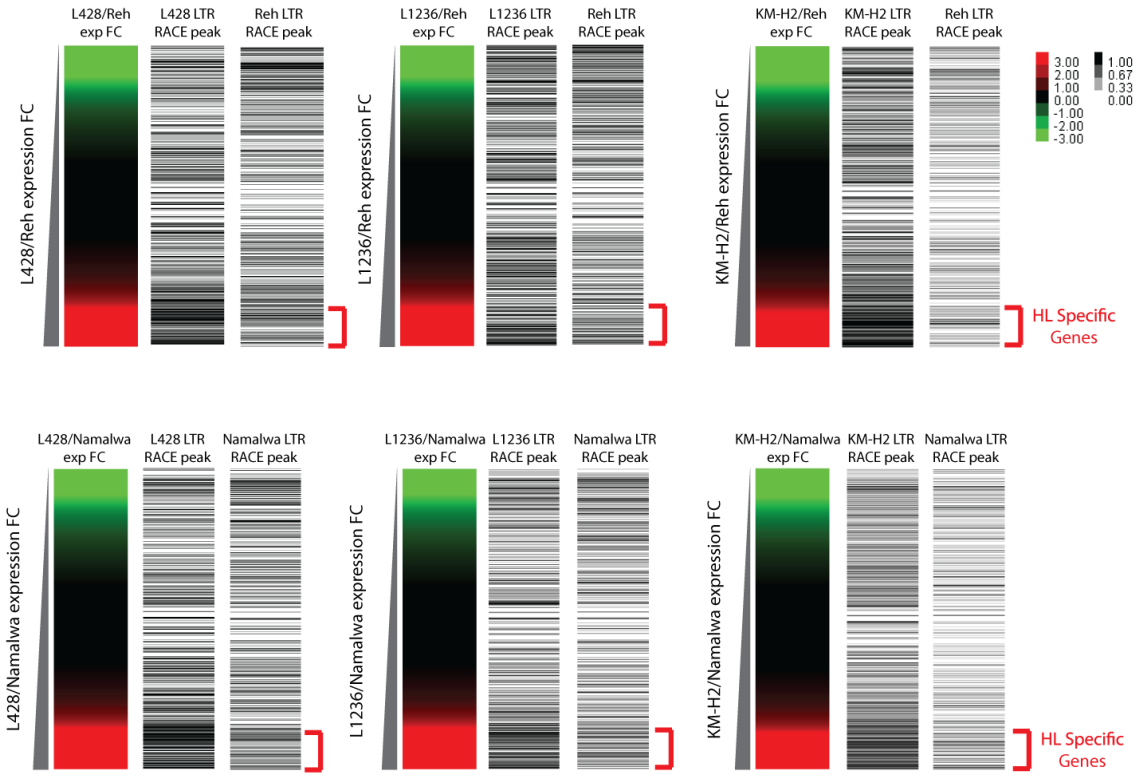


Figure 3.25. Active LTRs cluster with up-regulated genes in HL cell lines Fold-change of gene expression obtained by RNA-Seq for each of the HL cell lines over each of the control cell lines was plotted. The presence of active LTRs closest to these genes was plotted based on the gene expression fold-change axis. This showed an increased proportion of active LTRs close to the genes up-regulated in HL compared to the control cell lines. The control cell line active LTRs plotted on the same axis show very little relationship to gene expression.

The presence of active LTRs close to genes specifically up-regulated in HL confirmed a genome-wide relationship between HL gene expression and LTR activation. The LTRs may be acting as promoters or enhancers for the up-regulated genes, however the enhancer effect is much harder to determine as many enhancers do not interact with the promoters of their nearest genes (Sanyal, Lajoie et al. 2012, Mifsud, Tavares-Cadete et al. 2015).

Because the function of LTRs as promoters was easier to validate we focused on LTRs which were most likely to be acting as promoters assuming that the majority of LTRs acting as promoters would produce downstream transcripts on the same strand based on the average profiles (Figure 3.24). To study this on a gene specific basis we reassigned the closest genes to the LTRs based on the strand annotation from repeat masker. Correlation clustering of the expression of these genes close to active LTRs showed that the HL cell lines and control cell lines formed 2 separate clusters. The L428 cell line clustered further away from the L1236 and KM-H2 cell lines the same as when clustering the cell lines using expression of all genes (Figure 3.26). This finding implies that the action of LTRs as promoters of their nearest downstream gene is sufficient to differentiate between the cell lines and plays a significant part in the gene expression program of the cells.

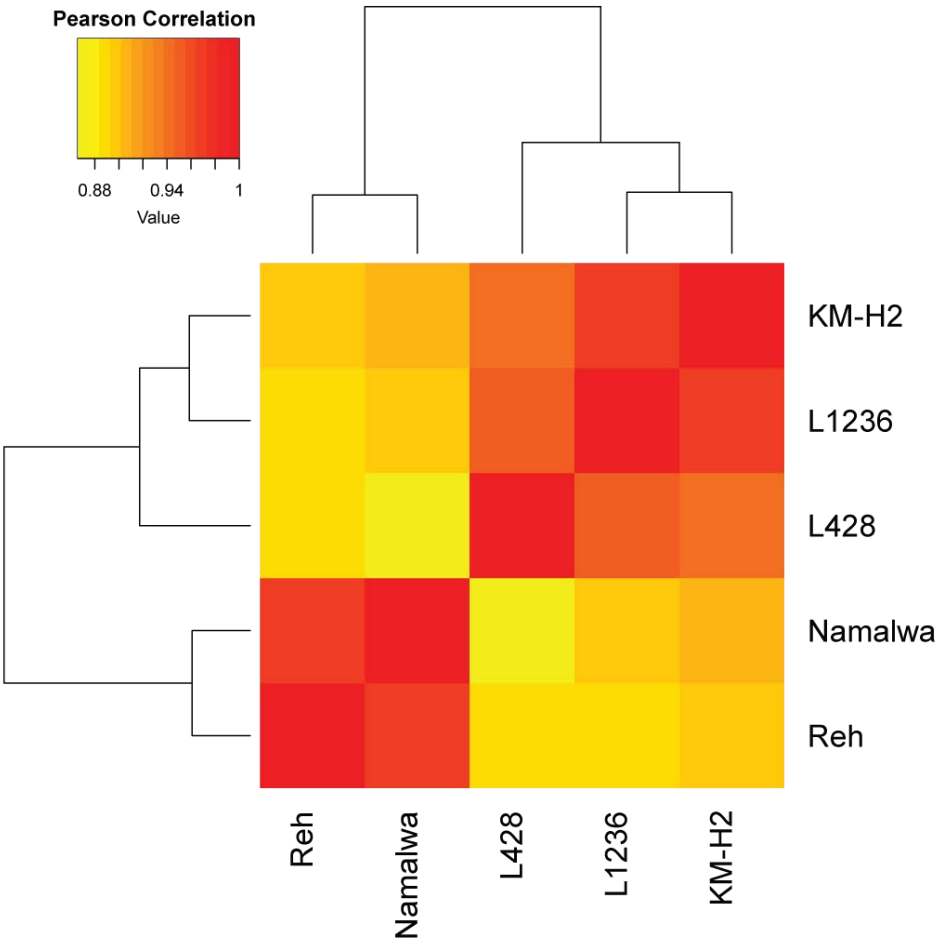


Figure 3.26. HL cell lines cluster based on expression of genes down-stream of active LTRs
The orientation of active LTRs identified by RACE-Seq was inferred from annotation of LTR orientation by repeat masker. The closest genes downstream of active LTRs were annotated and the expression values obtained from RNA-Seq data. The expression of these genes was correlated by Pearson correlation and the result clustered. The clustering showed 2 separate groups for the HL and control cell lines based on expression of these genes.

To understand how active LTRs influenced expression of individual genes and how this may contribute to the HL phenotype we manually curated the active LTRs using UCSC genome browser to observe changes in RNA expression around the active LTRs. Although this method is far from definitive as many active LTRs have very few reads associated with them in the RNA-Seq data it proved to be useful for identifying at least a proportion of the LTR driven transcripts without having to individually validate every active LTR by qPCR. The overall finding

of this analysis was that LTRs produced 4 different types of transcript: those originating from an upstream promoter or from an intragenic promoter as well as generating an RNA anti-sense to a protein-coding gene or as an un-annotated, intergenic long non-coding RNA.

As an upstream promoter an LTR can either be solely responsible for the expression of the gene or increase the level of expression of a gene with an active native promoter. The *NLRP1* gene displayed an increase in expression from an upstream LTR in all HL cell lines. The THE1C LTR located ~35 Kb upstream of the native *NLRP1* promoter was active in all HL cell lines and not active in the control cell lines. The LTR was in a DNaseI hypersensitive site and carried the H3K4me3 mark in the L428 cell line but not in the Reh cell line. The RNA-Seq data showed a clear read-through transcript linking the LTR to the first exon of *NLRP1* (Figure 3.27). RNA-Seq data showed that *NLRP1* was active in all cell lines, however, there was at least a 4-fold increase in expression observed between the Reh and HL cell lines and at least an 8-fold in the between the Namalwa and HL cell lines (Figure 3.28 A). To confirm that the transcript from the LTR shown by the RNA-Seq data was linked to the up-regulation of *NLRP1* we performed a qPCR using primers designed in the LTR and second exon of *NLRP1*. The results confirmed expression of this transcript in the HL cell lines with the highest level in L428 and an absence of the transcript in the control cell lines (Figure 3.28 B).

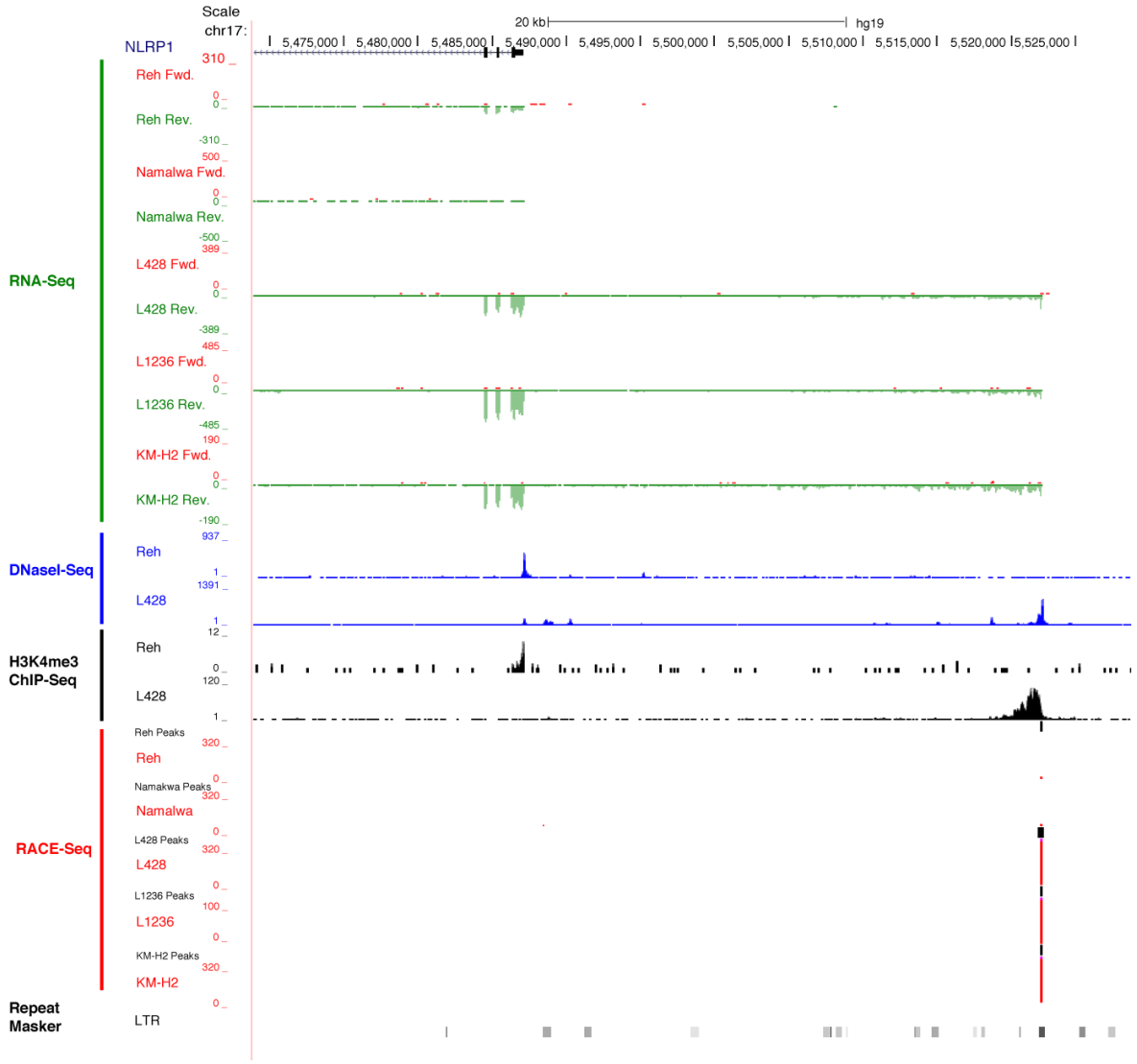


Figure 3.27. Intergenic LTRs act as alternative promoters in HL

UCSC genome browser screen shot showing an active THE1C LTR only active in HL cell lines and acting as a promoter for *NLRP1*. The LTR also lies within a region of open chromatin and is marked with H3K4me3 in the L428 cell line. RACE Peaks represent the regions identified as peaks by Macs and were used to define active LTRs in RACE-Seq data. A RNA-Seq transcript was observed originating from the LTR and proceeding downstream to the first exon of *NLRP1*.

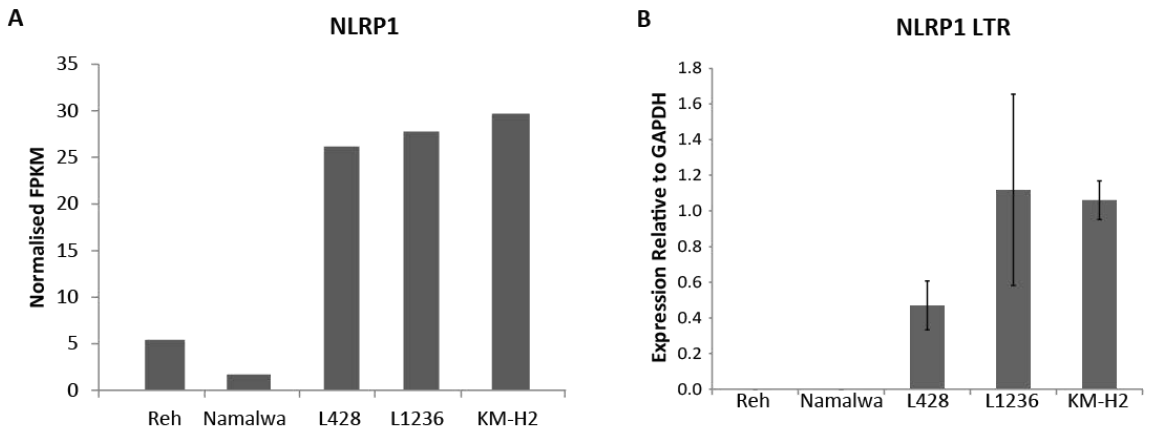


Figure 3.28. *NLRP1* is expressed from a THE1B LTR promoter in HL

A. *NLRP1* was more highly expressed in HL than control cell lines based on normalised RNA-Seq expression data. B. qPCR data relative to *GapDH*, using primers designed in the upstream LTR and exon 2 of *NLRP1*. Expression of the LTR transcript is present in all HL cell lines and absent from the control cell lines indictating the role of the THE1C LTR as a HL specific promoter for this gene. qPCR n=3, error bars show standard deviation.

An intragenic LTR promoter has the potential to produce a longer isoform of a gene whereas an intragenic promoter will always produce a shorter isoform. Depending upon the gene in question this may produce a non-functional protein or a protein with modified function. The activity of intragenic LTRs is more difficult to determine by visual inspection of genome browser tracks as many genes with intragenic LTRs also have some degree of native promoter activity meaning that an alternate isoform can be hidden by the full length isoform. An example of an intragenic LTR where the activity could be determined was seen in the *CACNA2D1* gene in the KM-H2 cell line (Figure 3.29). This gene carried an active intronic THE1B LTR located between exons 3 and 4. A low level of expression was present in exons 2 and 3 prior to the LTR and the exon expression level following the LTR was dramatically increased, producing an isoform 3 exons shorter than the native promoter.

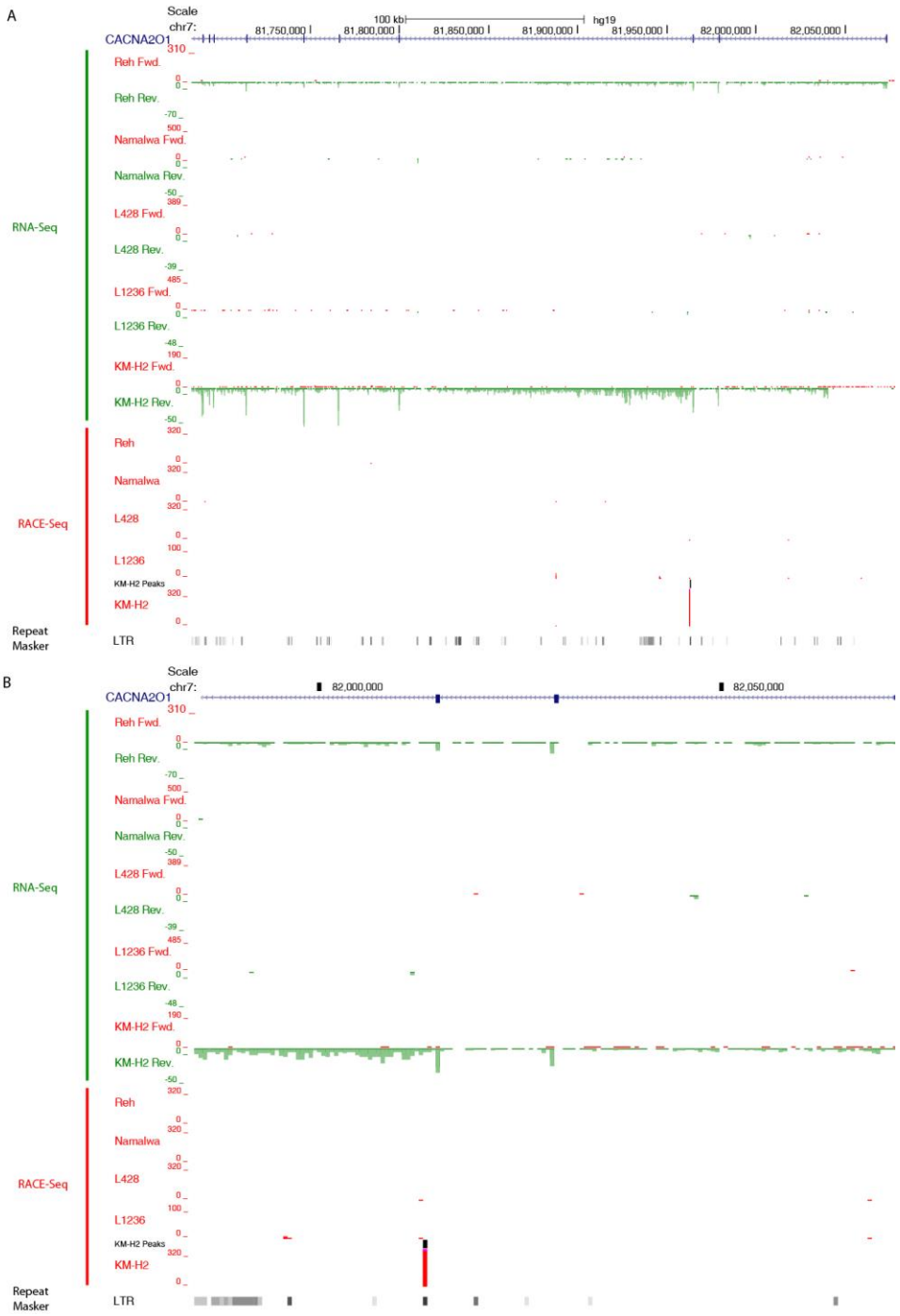


Figure 3.29 Intragenic LTRs produce shorter isoforms of some genes in HL UCSC genome browser screen shot showing a THE1B LTR only active in the KM-H2 cell line and acting as a promoter for *CACNA2D1*. As an alternative intragenic promoter it produces a shorter isoform of the gene without exons 1-3. RACE Peaks represent the regions identified as peaks by Macs (See Methods 2.9.5) and were used to define active LTRs in RACE-Seq data. A) Showing transcript running into downstream exons. B) Magnified version showing exact position of THE1B LTR.

An anti-sense RNA promoter can be either intergenic or intragenic and produces an anti-sense transcript of a gene often resulting in reduced gene expression through disruption at the initiation, transcriptional processing or the post-transcriptional processing stages of the genes sense transcript (Pelechano and Steinmetz 2013). Our RNA-Seq data showed that the gene expression of CHD1L was reduced in the HL cell lines L428 and KM-H2. A downstream THE1B LTR, 20Kb from the termination site of the CHD1L transcript, was observed producing an anti-sense RNA. The active LTR was also within a DNaseI hypersensitive site and carried H3K4me3 in L428 cells, both of which were absent in Reh cells (Figure 3.30). The expression data obtained from the RNA-Seq showed an average ~1.5 fold decrease in CHD1L expression in both the L428 and KM-H2 cell lines when compared to the cell lines without the active LTR (Figure 3.31). This demonstrates that when transcribed in the antisense direction, active LTRs have the potential to reduce the level of gene expression. Therefore, highlighting another potential route by which LTR activation could influence the overall gene expression program in HL cells.

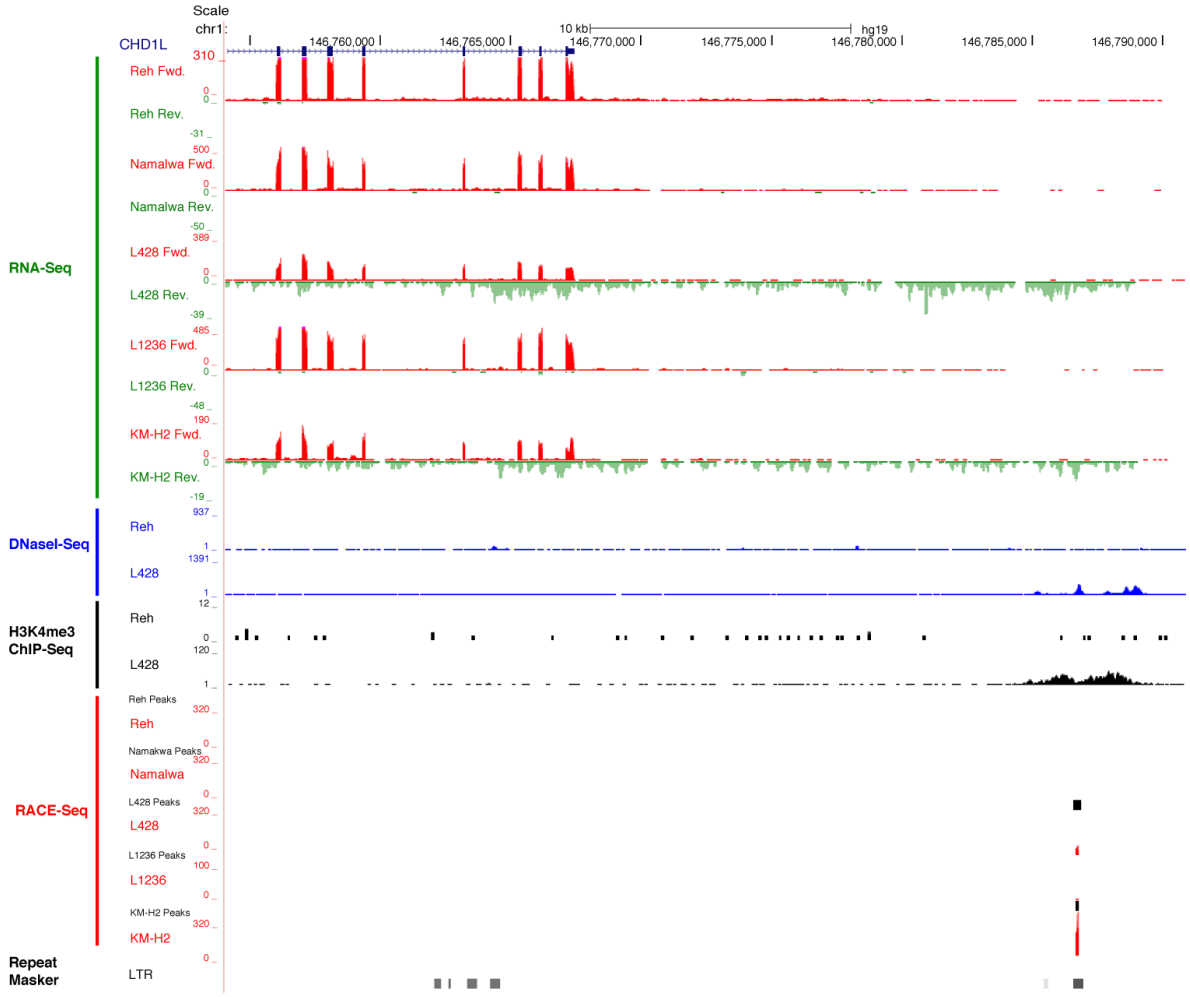


Figure 3.30. Active LTRs can produce anti-sense RNA transcripts correlating with reduced gene expression

UCSC genome browser screenshot showing a THE1B LTR only active in the L428 and KM-H2 HL cell lines and acting as a promoter for and anti-sense RNA as seen in the RNA-Seq data. The LTR also lies within a region of open chromatin and is marked with H3K4me3 in the L428 cell line. The anti-sense RNA overlaps with the *CHD1L* gene and correlates with a reduction in gene expression. RACE Peaks represent the regions identified as peaks by Macs and were used to define active LTRs in RACE-Seq data.

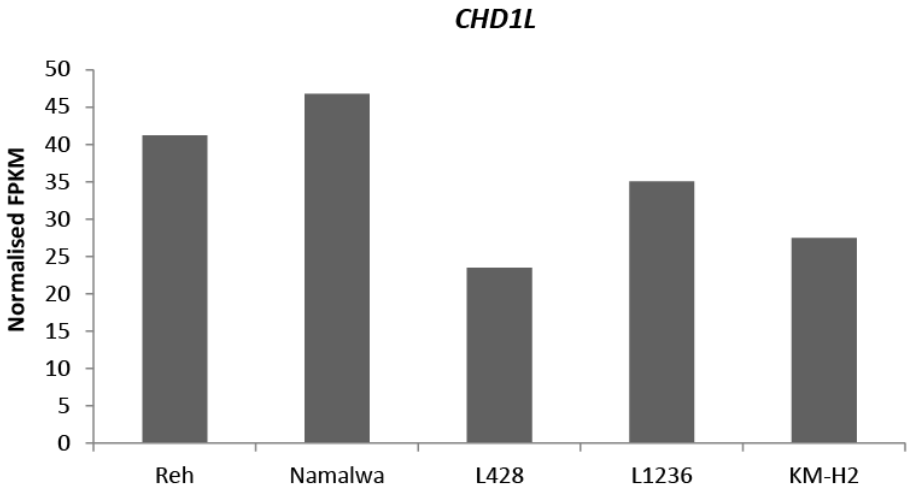


Figure 3.31 CHD1L is down-regulated in HL cell lines compared to control cell lines

Normalised FPKM expression values obtained from RNA-Seq data shows a down-regulation of CHD1L in HL cell lines when compared to the control cell lines.

Finally, our data show that many of the active LTRs produce un-annotated long non-coding RNAs (lncRNAs) which do not overlap with protein-coding genes. These transcripts were often 10's of kilobases in length and located in gene deserts with an RNA transcript that gradually reduced in expression level with length (Figure 3.32). The function of these elements is far harder to determine as they may just be a benign side effect of an LTR being active and inducing transcription of RNA with no function, which is simply degraded. It is also possible, however, that some of these lncRNAs physically interact with the promoters of other genes to inhibit or enhance transcription. A proportion of these un-annotated transcripts also showed indications of splicing events in the RNA-Seq data. An example of this phenomenon was seen on chromosome 12 where a THE1B LTR was acting as a promoter in all HL cell lines for an un-annotated RNA transcript. As with the other active LTR promoters both a DNaseI hypersensitive site and H3K4me3 were present at the LTR in L428 and not in Reh. The transcript lay ~1 Mb from the nearest annotated coding gene and appeared to have a number of peaks in RNA-Seq

data which could represent spliced exons. It is possible that this could be a previously unidentified protein coding gene although it is more likely to be a non-coding transcript which may influence with the expression of another gene.

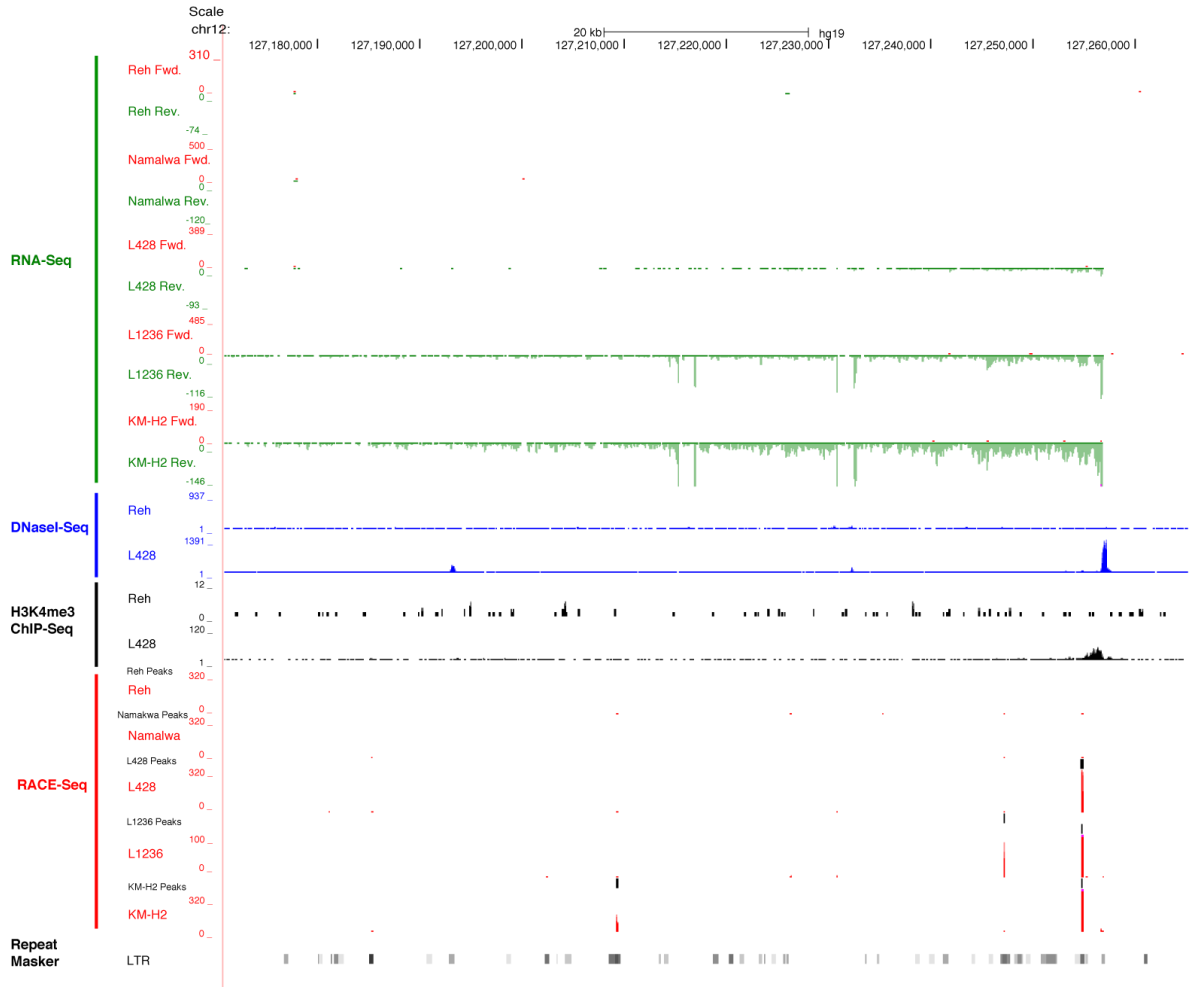


Figure 3.32. Active LTRs produce unannotated lncRNA transcripts
UCSC genome browser screenshot showing a THE1B LTR only active in the HL cell lines and acting as a promoter for an unannotated long non-coding RNA. The LTR also lies within a region of open chromatin and is marked with H3K4me3 in the L428 cell line. RACE Peaks represent the regions identified as peaks by Macs and were used to define active LTRs in RACE-Seq data.

Overall it is clear that active LTRs in HL have both a direct impact through their activity as alternative promoters and the potential for an indirect impact through non-coding RNAs on the expression of genes. Active LTRs also have the potential to act as enhancers, which may be

seen to some extent in the link between LTRs and the up-regulated HL genes (Figure 3.25). It is much more difficult, however, to identify which enhancers may be acting on which genes.

3.3.4. TNFRSF11A is expressed from an active LTR in the L1236 cell line and is involved in the up-regulation of NF-κB activation

After showing that LTR activation had a global impact on gene expression in HL we next wanted to determine whether any of the up-regulated genes were critical regulators of the HL phenotype. An active THE1B LTR was identified 1.2 Kb upstream of the *TNFRSF11A* (tumour necrosis factor receptor superfamily member 11a) gene, producing a transcript of *TNFRSF11A* in the L1236 cell line and also a weak transcript in the L428 cell line (Figure 3.33). A DNaseI hypersensitive site and H3K4me3 were not observed in the L428 or Reh datasets which would suggest in the case of L428 the histone mark and DHS may be too weak to detect or may just be present in a small subset of L428 cells. The expression of *TNFRSF11A* was clearly up-regulated in L1236 and this was confirmed by qPCR (Figure 3.34). To validate the existence of the LTR driven transcript, qPCR was also performed using primers in the LTR and in exon 2 of *TNFRSF11A* demonstrating expression of the LTR driven transcript at a level at least as high as the transcript from primers designed within exons 2 and 3 of the gene.

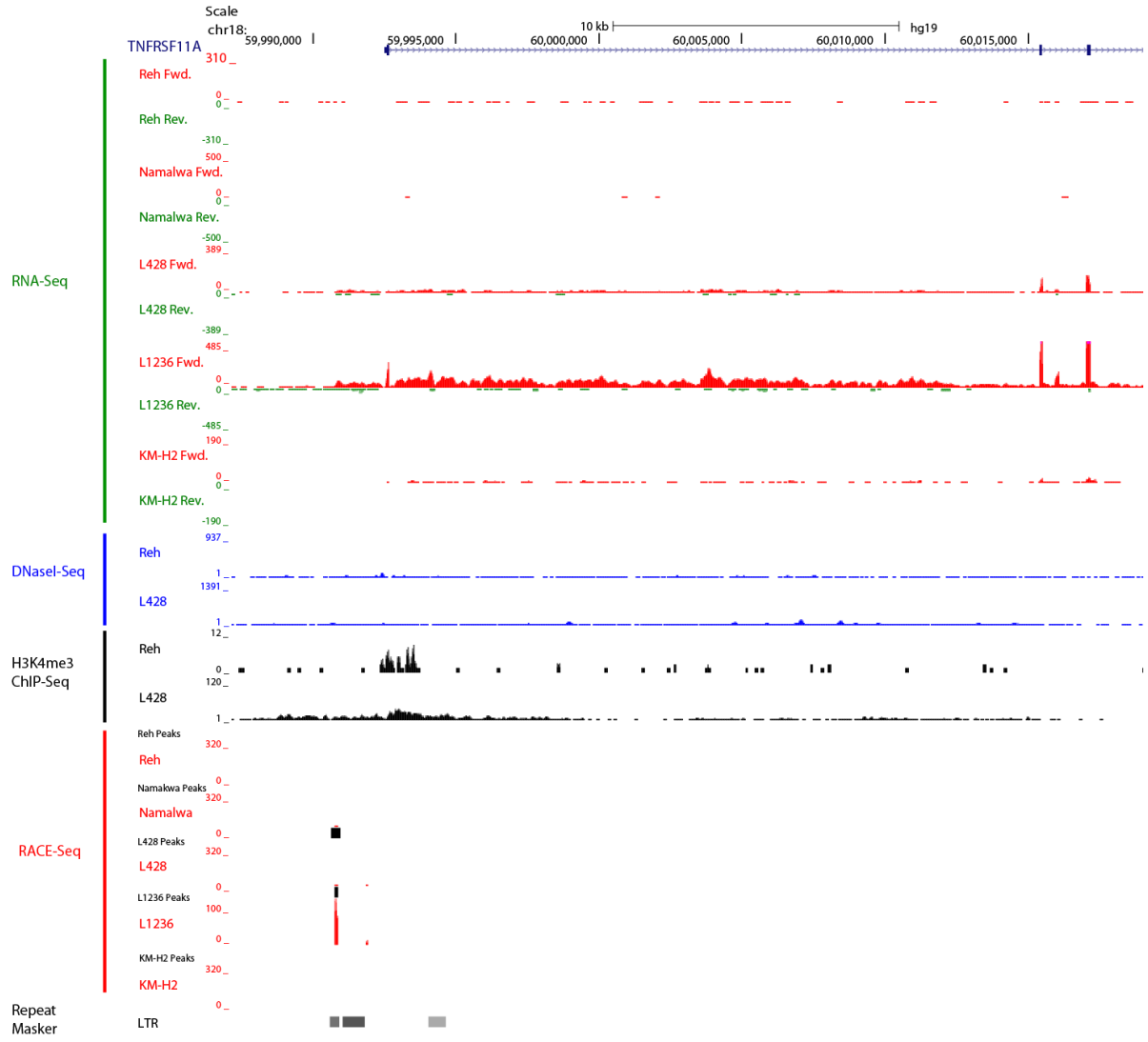


Figure 3.33. *TNFRSF11A* is expressed in the L1236 HL cell line from an active upstream THE1B LTR

UCSC genome browser screen shot showing a THE1B LTR only active in the L1236 and L428 HL cell lines and acting as a promoter for the *TNFRSF11A* gene. A low level of H3K4me3 in the L428 cell line around the LTR can be observed. RACE Peaks represent the regions identified as peaks by Macs and were used to define active LTRs in RACE-Seq data.

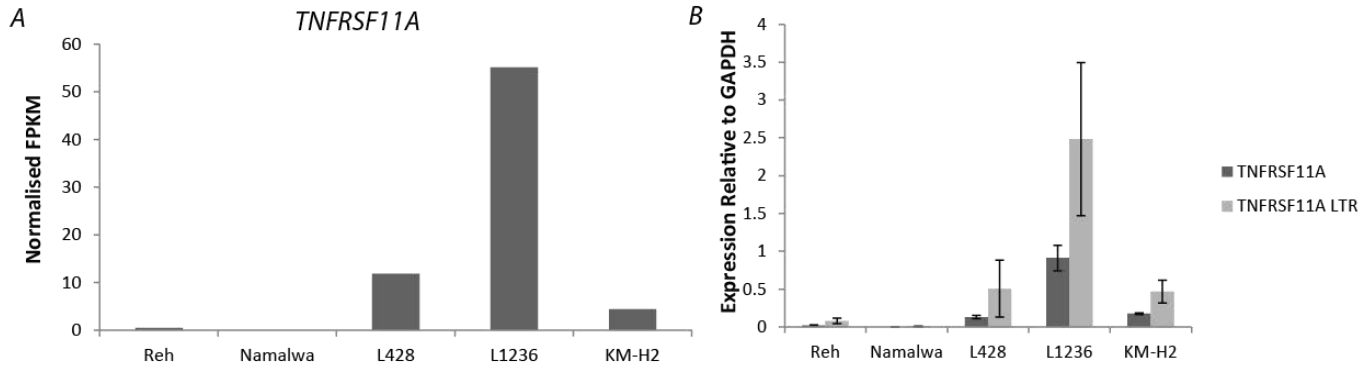


Figure 3.34 High levels of TNFRSF11A are expressed from a THE1B LTR in the L1236 HL cell line.
A. Normalised FPKM values from RNA-Seq data showed a high level of *TNFRSF11A* expression in the L1236 cell line. B. The high level of expression seen in RNA-Seq data was confirmed by qPCR with primers design in exon 2 and 3 of *TNFRSF11A* and the transcript originating from the LTR was confirmed using primers in the LTR and exon 2 of *TNFRSF11A*. Average of 3 biological replicates, error bars show standard deviation.

The discovery of another LTR driven gene with specificity to the L1236 cell line was an interesting finding as it showed that an LTR with cell line specific expression could have a cell specific impact on a cellular phenotype and may indicate that a sub-type specific LTR activation in HL. The other reason that *TNFRSF11A* was of particular interest is because it has been shown in other studies to be up-regulated in HL patient samples and it is involved in the regulation of NF-κB signalling through its interaction with its ligand TNFSF11 (Steidl, Diepstra et al. 2012). TNFRSF11A interacts with tumour necrosis factor receptor-associated factors (TRAFs), which feed in to the NF-κB and JNK pathways resulting in increased activity (Fiumara, Snell et al. 2001). This is of particular relevance to HL as we known that the central control mechanisms for gene expression in HL are through NF-κB and AP-1. NF-κB also potentially plays a very important role in the activation of LTRs as the THE1B consensus sequence has an NF-κB motif and it was shown that the *CSF1R* LTR required aberrant NF-κB activation to become fully active (Lamprecht, Walter et al. 2010).

TNFRSF11A is primarily activated through interaction with its ligand TNFSF11, however, our RNA-Seq data showed no expression of TNFSF11 in any of the HL or control cell lines.

Although TNFSF11 is the major activating partner of TNFRSF11A it was shown by Anderson et al. (1997) that overexpression of the receptor alone could induce NF-κB activation. To determine whether TNFRSF11A was a driving factor in HL pathology we performed an siRNA knockdown in the L1236 cell line. The knockdown successfully reduced the level of *TNFRSF11A* mRNA (Figure 3.35). However, there was no observable difference in the growth rate or cell morphology following the knock down.

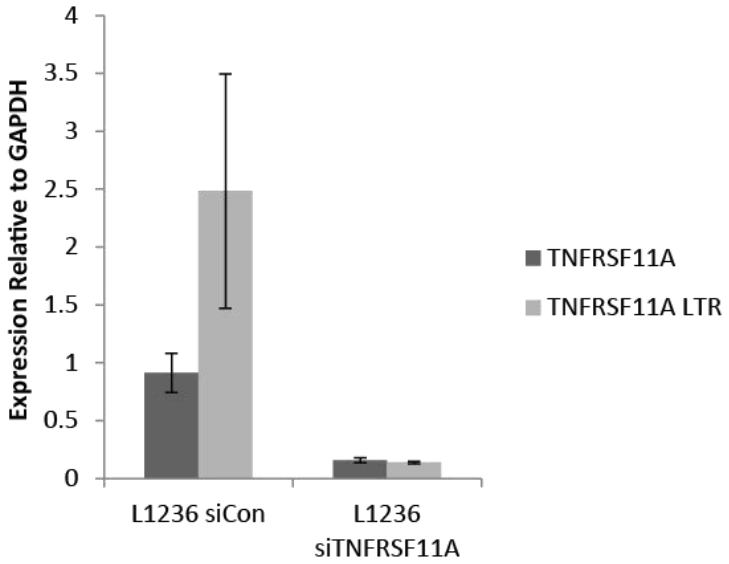


Figure 3.35. TNFRSF11A can be knocked down by siRNA

RNA interference was used to knockdown of *TNFRSF11A* in theL1236 cell line. Expression of *TNFRSF11A* using primers designed for exon 3 and the upstream LTR and exon 2 was measured by qPCR comparing cells treated with siTNFRSF11A to cells treated with a non-targeting control (siCon). Treatment with siTNFRSF11A showed a good level of knockdown of both transcripts. Error bars show standard deviation, n=3.

We next wanted to establish whether TNFRSF11A expression had an impact on NF-κB activation. When NF-κB becomes active it is translocated to the nucleus (Trask 2004). We therefore determined activation by analysing nuclear NF-κB levels in L1236 cells with an siRNA knockdown of *TNFRSF11A* and compared to non-targeting control siRNA treated cells.

Following knock down of TNFRSF11A the level of NF-κB p65 present within the nucleus decreased showing a role for TNFRSF11A in NF-κB activation in the L1236 cell line (Figure 3.36). This finding shows the exciting potential of a positive feedback loop where the LTR activation supports and enhances its self via the increase of NF-κB signalling.

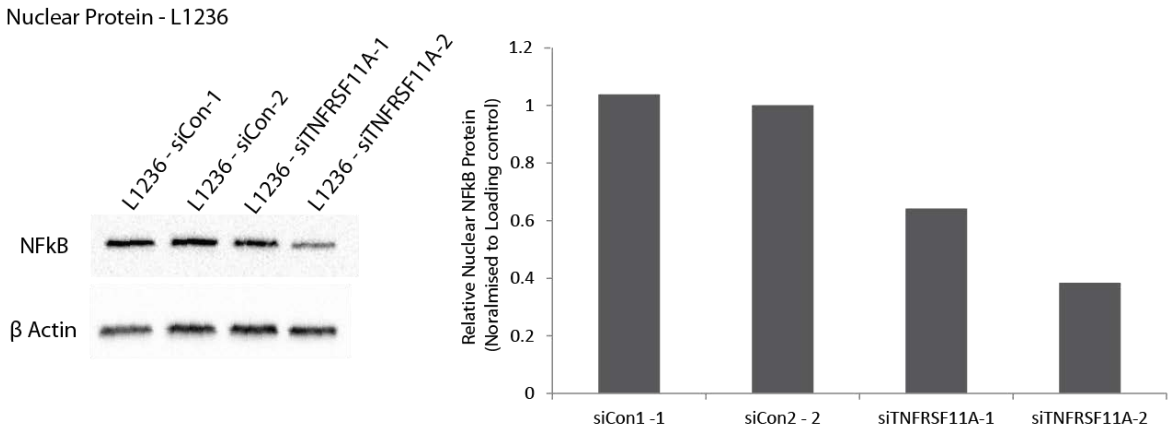


Figure 3.36. Knockdown of *TNFRSF11A* in L1236 cells results in reduced NF-κB activation.
Western blot showing nuclear protein extract from L1236 cells treated with either a control siRNA (siCon) or a siRNA against *TNFRSF11A* probed for NF-κB p65. Two biological replicates are shown and beta-Actin was also probed as a loading control. A nuclear specific protein loading control was not performed, although would have been useful to confirm the purity of the nuclear protein fraction. The right panel shows the relative expression values normalised to the loading control demonstrating a drop in NF-κB protein in the nuclear fraction resulting from knockdown of *TNFRSF11A*.

3.4. Long Terminal Repeats can be activated by inflammatory signalling

Lamprecht et al. 2010, demonstrated that activation of the CSF1R LTR could be recapitulated in the Reh cell line by the knockdown of transcriptional repressor ETO2 and constitutive activation of NF-κB, both of which are features of HL. We have shown in our foot-printing data that NF-κB and AP-1 are the main driving factors of the HL phenotype and know that the overall HL gene expression pattern comprises many genes involved in inflammation. To study the factors involved in transcriptional activation of the THE1B LTR we analysed its consensus sequence and observed an NF-κB binding motif up-stream of the transcription start site (Figure 3.37), which had also been previously identified (Lamprecht et al. 2010). Based on this result we wanted to determine whether inflammatory signalling particularly in relation to NF-κB activation could be driving the genome-wide activation of LTRs in HL.

THE1B Consensus Sequence

```
 GATA                    SP1
TGATATGGTTTGGCTGTGTCCCCACCCAAATCTCATCTTGAATTGTAGCTCCCATAATT
    E-Box        SP1                        AP-1        SP1
CCCACGTGTCGTGGGAGGGACCCGGTGGGAGGTAATTGAATCATGGGGGCGGGTC
                     GATA                                  TATA
TTTCCCGTGCTGTTCTCGTGATAGTGAATAAGTCTCACGAGATCTGATGGTTTTATAAA
      NFκB                Start
GGGGAGTTCCCCTGCACAWG
```

Figure 3.37. THE1B consensus sequence contains an NF-κB binding motif. THE1B LTR consensus sequence obtained from RepBase and annotated with transcription factor binding motifs. The consensus sequence includes AP-1 and NF-κB motifs both of which have the potential to bind these transcription factors which are critical for driving the HL gene expression programme.

3.4.1. Treatment of the Reh cell line with PMA induces *CSF1R-LTR* expression

The compound phorbol 12-myristate 13-acetate (PMA) is a potent activator of the protein kinase C (PKC) enzyme due to its similarity to diacylglycerol, a natural activator of PKC (Slater, Kelly et al. 1994). Activation of PKC drives many downstream cellular pathways which have been linked

to inflammation and tumour growth including MAPK and NF-κB pathways, both of which are up regulated in HL (Zheng, Fiumara et al. 2003). To establish whether the inflammatory signalling, particularly NF-κB activation, was able to induce LTR activation we treated Reh cells with 2ng/ml PMA for a period of 8 or 16 hours.

At both time points transcription of the *CSF1R LTR* was induced as measured by qPCR (Figure 3.38). The induction was slightly reduced at the 16 hour time point, therefore further experiments were carried out following 8 hours of PMA treatment. It was previously shown by Lamprecht et al. (2010) that loss of the transcriptional repressor CBFA2T3 (ETO2) was also required to activate the CSF1R LTR. To establish whether PMA had an effect on *CBFAT23* expression, qPCR was performed. This showed a downregulation of ETO2 upon treatment with PMA (Figure 3.39). Although the route by which this downregulation occurs is not clear, it shows that the LTRs may be activated by a combination of ETO2 loss and NF-κB activation as previously published.

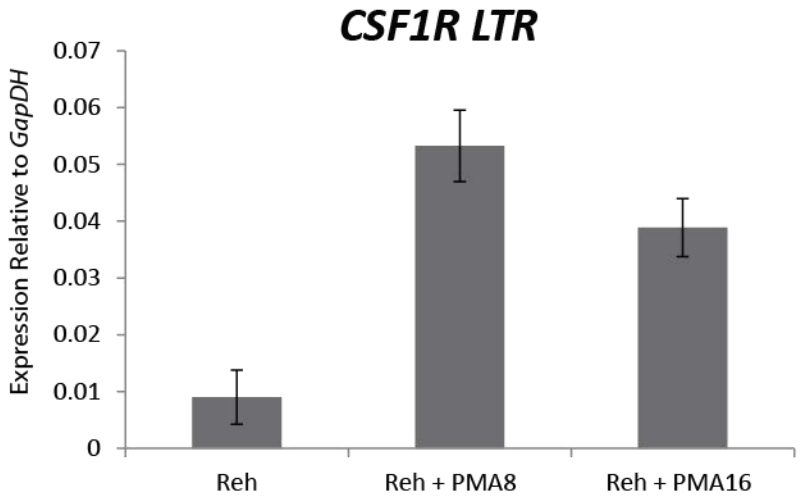


Figure 3.38. Treatment of Reh cells with PMA induces CSF1R LTR expression Treatment of Reh cells with 2ng/ml PMA for 8 or 16 hours followed by measurement of CSF1R LTR expression by qPCR. Induction of *CFS1R* expression can be observed following PMA treatment resulting from activation of the LTR promoter. N=3 error bars show standard deviation.

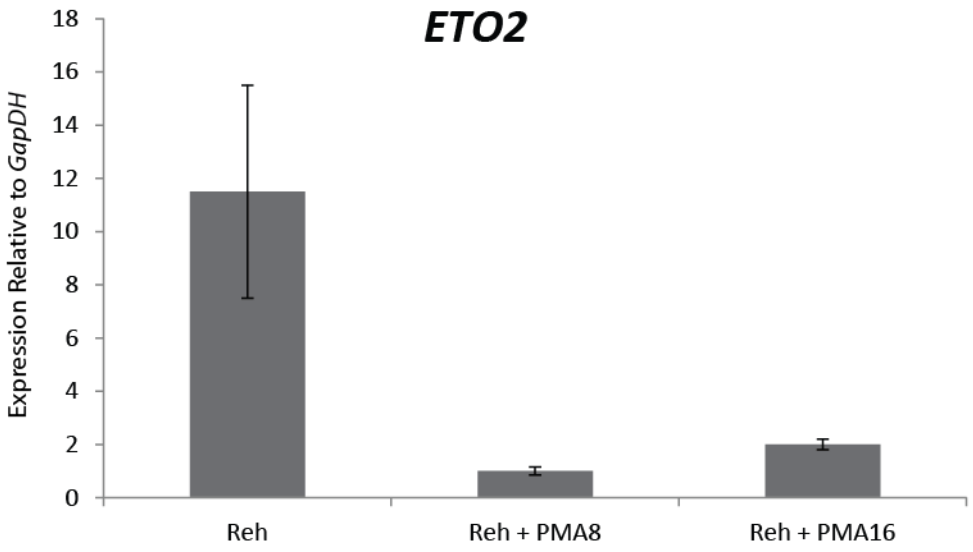


Figure 3.39. ETO2 expression reduces following PMA treatment of Reh cells Gene expression measured by qPCR showing a drop in *ETO2* expression following treatment of Reh cells with 2ng/ml PMA for 8 or 16 hours. N=3 error bars show standard deviation.

3.4.2. Treatment of the Reh cell line with PMA induces global THE1B LTR activation

Following the discovery that PMA treatment induced activation of the *CSF1R LTR* in Reh cells, we wanted to use our RACE-Seq technique to determine the impact of PMA treatment on THE1B related LTRs genome-wide. RACE-Seq was carried out in triplicate on cells treated with PMA for 8 hours. The RACE-Seq data was validated by the observation of reads aligned to the *CSF1R LTR* using UCSC Genome Browser (Figure 3.41). The replicates had a similar level of overlap to the previous RACE experiments and the same strategy of using RACE peaks present in at least 2 replicates for further analysis was applied (Figure 3.40). Previous analysis of the RACE data obtained from the Reh cell line showed that a proportion of LTRs were already active. Therefore we next wanted to find out how activation changed following PMA treatment and whether the LTR activation pattern became more HL like. Overlap of the RACE-Seq peak regions from the untreated Reh cells, the union of HL cell lines and PMA treated Reh cells showed a loss of 153 Reh specific sites and a gain of 1,295 sites resulting from PMA treatment

(Figure 3.42). Following treatment with PMA the Reh cells gain activation of 295 HL specific LTRs indicating that the PMA induces a partial move towards the HL LTR activation pattern.

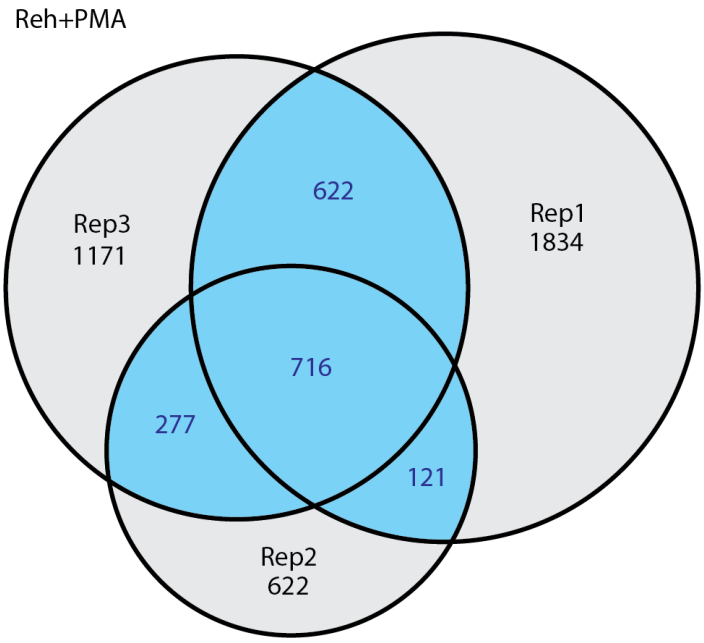


Figure 3.40. Overlap of biological replicates of LTRs detected by RACE-Seq in PMA treated Reh cells

Overlap of active LTR peaks identified by THE1B RACE-Seq in 3 biological replicates following PMA treatment of Reh cells. Further analysis was carried out on peaks shared in at least 2 replicates as indicated by the blue regions.

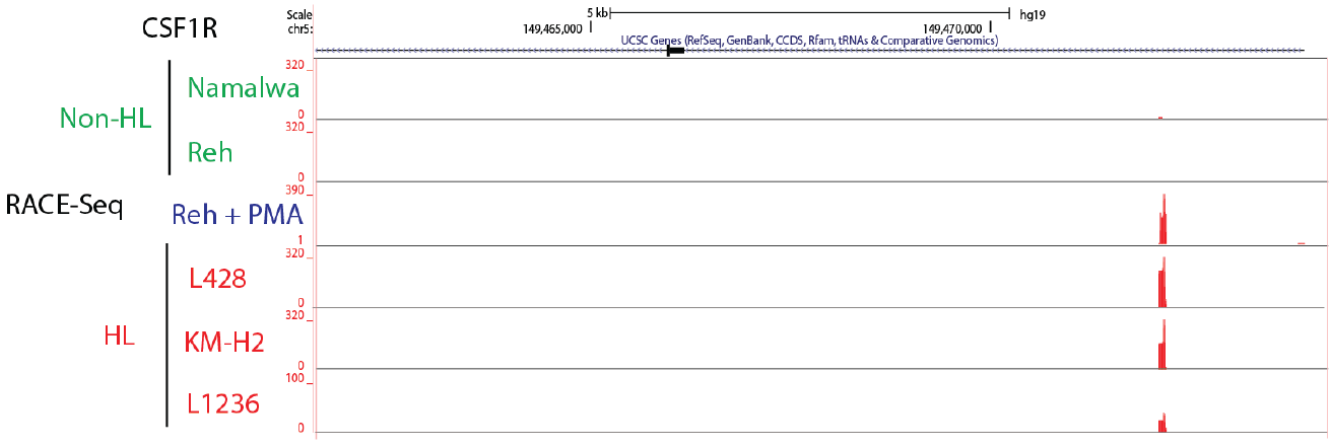


Figure 3.41. Activation of transcription from the *CSF1R LTR* can be detected by RACE-Seq in Reh cell line following PMA treatment

UCSC genome browser screenshot showing aligned RACE-Seq reads at the THE1B LTR upstream of *CSF1R*. Following treatment with PMA the *CSF1R LTR* became active in Reh cells.

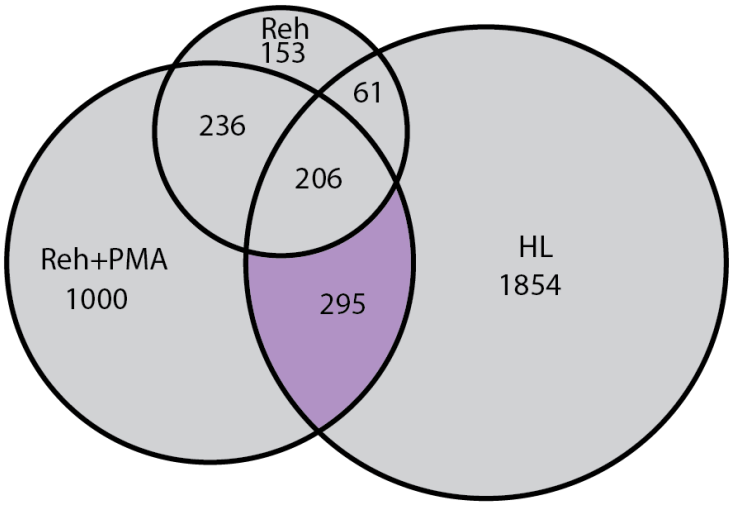


Figure 3.42. Following PMA treatment LTRs are transcriptionally activated

Peaks representing active LTRs determined by RACE-Seq were overlapped to compare the LTR activation in Reh cells following 8 hours of PMA treatment to untreated Reh cells and the union of all HL active LTRs. This showed that 295 LTRs activated by PMA treatment were also active in HL cell lines.

3.4.3. LTRs activated by PMA treatment have a genome-wide impact on gene expression

Having shown that LTR activation has an impact on the gene expression pattern in HL but to a much lesser extent in Reh cells it was important to examine whether PMA activated LTRs also influenced the expression of their associated genes. We first carried out RNA-Seq on the PMA treated Reh cells, with two biological replicates which showed very good correlation (Figure 3.43). Pearson correlation of the LOG2 FPKM values obtained from RNA-Seq between the PMA treated Reh cells and the control and HL cell lines, followed by clustering, showed the PMA treatment moved the overall gene expression pattern of the cells away from the control cell lines (Figure 3.44). A slight increase in correlation of the PMA treated Reh cells to the HL cell lines was also observed, which suggests the PMA treatment changes the gene expression pattern to be more like that of the HL cells.

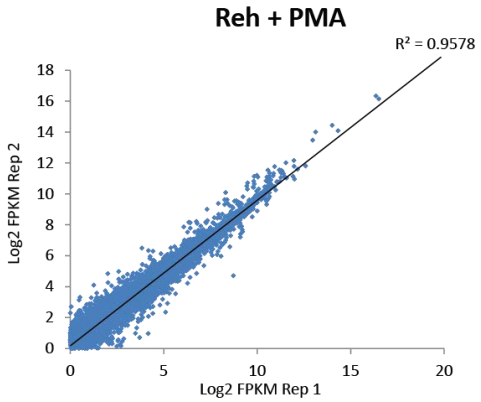


Figure 3.43. Correlation of RNA-Seq biological replicates from PMA treated Reh cells
Correlation of Log2 FPKM values from 2 biological replicates of RNA-Seq data obtained from Reh cells treated with 2ng/ml PMA for 8 hours. R^2 value shows a good correlation between the replicates.

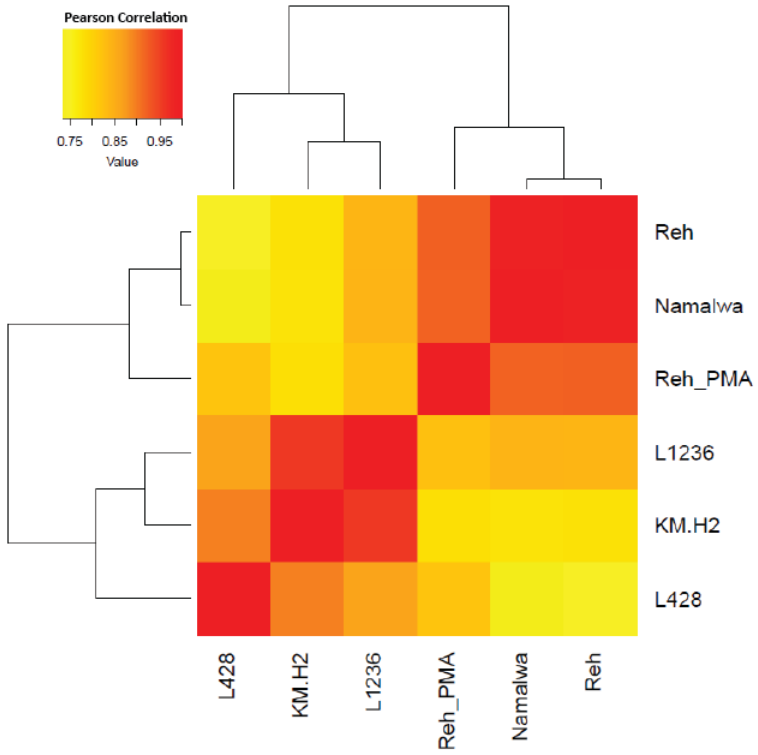


Figure 3.44. PMA treatment of Reh cells results in an increased correlation of its gene expression pattern with that of HL cells
Pearson correlation between the FPKM values from RNA-Seq in each of the HL and control cell lines including PMA treated Reh cells was carried out followed by clustering of the gene expression correlations for the cell lines. This showed that PMA treatment of Reh cells made their gene expression correlate less with untreated cells and begin to correlate slightly more with HL cell lines.

In addition to the gene expression pattern shifting towards that of HL, a number of genes dysregulated following PMA treatment are known to be involved in HL. B-cell specific genes including *Pax5*, *CD79B, RAG1* and *Rag2* were down-regulated by PMA treatment and genes which contribute to the HL phenotype such as *CSF1R, JUNB, LTA* and *TNF* were up-regulated. As shown by the correlation of the gene expression data, PMA did not induce a complete move to the HL gene expression programme. Many important genes were not up-regulated including *IL13, CCL5, IL6, IRF5* and *CCR7*. We were unable to determine whether expression of these genes was upregulated at a later time point following PMA treatment as long term PMA treatment had a major impact on cell survival (Figure 3.45).

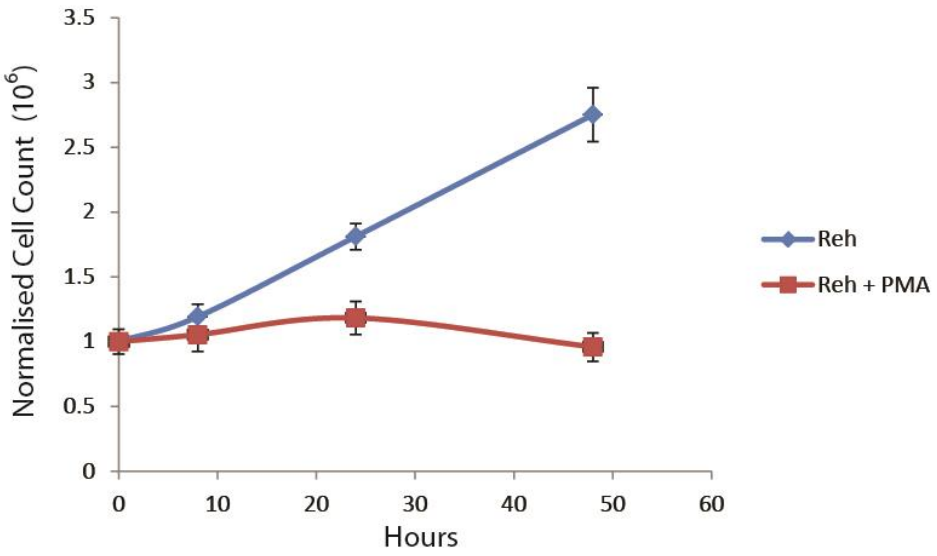


Figure 3.45 Reh cell growth is stopped following PMA treatment

Following treatment of Reh cells with 2ng/ml PMA a growth curve was created over a 48 hour period showing that proliferation of Reh cells is down-regulated following treatment. Cell counts normalised to $1x10^6$ at 0 hours. Error bars show standard deviation across 3 biological replicates.

As PMA impacts on the cellular environment through many different pathways we wanted to establish which part of the gene expression pattern resulted from LTR activation. We carried out the same analysis previously conducted on the HL cell lines by first determining the closest genes to the active LTRs in the Reh and Reh+PMA RACE-Seq data. The change in expression levels of such genes based on RNA-Seq data was then plotted against the presence or absence of active LTRs in PMA treated and untreated Reh cells (Figure 3.46). Genes up-regulated in Reh+PMA cells showed an increased presence of active LTRs in their vicinity which was not seen for active LTRs in the untreated Reh cells. This finding suggests that the LTRs which are activated by PMA treatment have an impact on the regulation of gene expression in the treated Reh cells and result in the up-regulation of some genes.

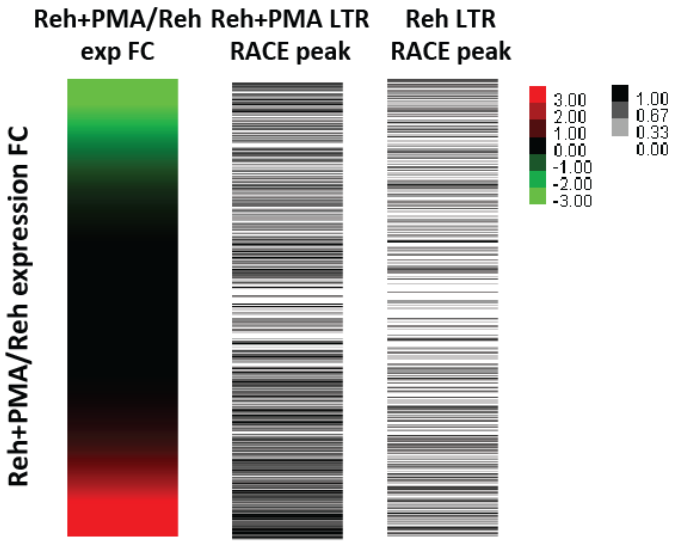


Figure 3.46. Transcriptionally active LTRs in PMA treated Reh cells are associated with up-regulated genes

Gene expression Log2 fold change for PMA treated Reh cells over untreated Reh cells was plotted (Red; up-regulated genes, Green; down-regulated genes). The presence of active LTRs, observed by RACE-Seq, close to the genes was then plotted on the same axis for treated and untreated cells. This showed a larger proportion of active LTRs associated with up-regulated genes in the treated cells than the untreated.

PMA treatment most likely results in the activation of LTRs in Reh cells via a similar route to that in HL as the induced signalling pathways mimic those seen in HL. To compare LTRs activated by PMA treatment with gene expression in HL cells, we ranked the gene expression fold change of several HL cell lines over Reh and aligned ranked genes with the presence of neighbouring LTRs (Figure 3.47). This analysis showed that PMA treatment of Reh cells induced LTRs which were associated with up-regulated genes in the L428 and KM-H2 HL cell lines and to a lesser extent with genes from the L1236 cell line. When compared to the L428 cell line we also found a group of LTRs which were associated with down-regulated genes. The reason for this could be LTRs which produce anti-sense transcripts and decrease the expression of nearby genes. However, as most THE1B related LTRs contain NF-κB binding sites, it is likely that a number of these LTRs are already active in Reh cells and are further activated following PMA treatment.

It should be noted that although PMA activates a proportion of active LTRs in Reh cells which are also active in HL many remain inactive. These include the LTRs involved in the up-regulation of *NLRP1* and the expression of *TNFRSF11A* in the L1236 cell line. This finding may point towards additional control mechanisms for LTR expression other than the loss of ETO2 and constitutive activation of NF-κB which was sufficient for the activation of the *CSF1R* LTR (Lamprecht, Walter et al. 2010).

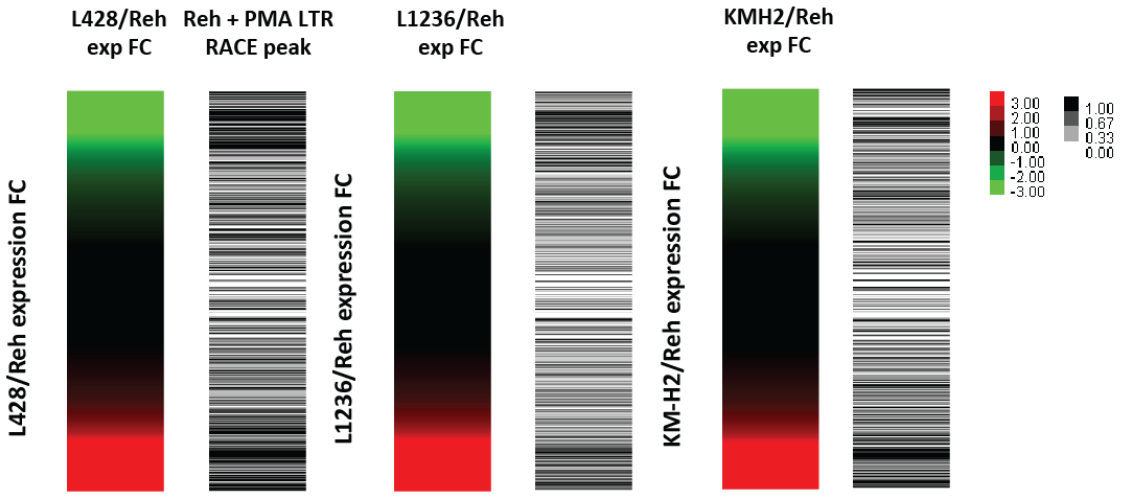


Figure 3.47. LTRs activated by PMA treatment of Reh cells associate with genes which are highly expressed in HL cell lines

Gene expression Log2 fold difference for HL cell lines over the Reh cell line was plotted (Red; up-regulated genes, Green; down-regulated genes). The presence of active LTRs, observed by RACE-Seq, close to the genes was then plotted on the same axis for PMA treated Reh cells showing PMA driven LTR activity close to genes upregulated in HL.

3.4.4. PMA treatment of Reh cells induces global changes in chromatin accessibility and pushes the cells towards the HL chromatin accessibility pattern

Following our finding that PMA treatment of Reh cells shifted gene expression patterns towards those seen in HL cells, we also wanted to assess whether these changes were associated with alterations in chromatin accessibility. Global changes in chromatin accessibility could play a role in genome-wide LTR activation and may lead to the up-regulation of genes not associated with LTRs. To this end, we carried out ATAC-Seq on the PMA treated Reh cells. This assay provides

the same information as the DNaseI-seq assay but has the advantage of needing a smaller amount of starting material. We also performed ATAC-Seq on the L428 cell line for comparison to the data produced by the DNaseI-Seq assay. The regions of open chromatin in each cell line, either determined by ATAC-Seq or DNaseI-Seq were correlated using Pearson correlation and clustered based on the correlations (Figure 3.48). The result showed that the ATAC-Seq and DNaseI-Seq data from the L428 cell line correlated very well showing that the assays are very comparable for the identification of open chromatin regions. As seen in the gene expression clustering the Reh cells treated with PMA still clustered with the control cell lines but have an increased correlation with the HL cell lines. This shows that the changes chromatin accessibility following PMA treatment of Reh cells follows a similar pattern to the changes in gene expression, moving towards a HL pattern.

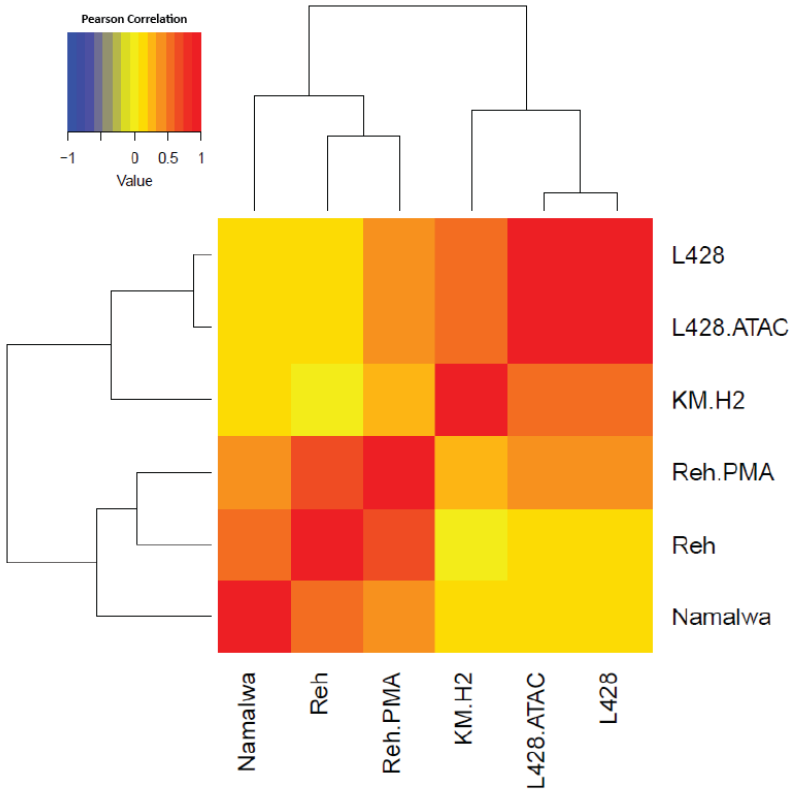


Figure 3.48. Following PMA treatment chromatin accessibility patterns of Reh cells shift towards those of HL cell lines

Regions of open chromatin identified by DNaseI-Seq or ATAC-Seq were determined. The sequences found in each of the cell lines were combined and the Pearson correlation of the tag counts for each site in each cell line was calculated. The correlation values were then clustered by hierarchical unsupervised clustering showing that the DNaseI-Seq data for L428 clusters with the ATAC-Seq data for L428 and also that the Reh cells treated with PMA have a higher correlation with the HL cell lines than the untreated cells have with the HL cell lines.

The induction of LTRs by PMA treatment is an exciting finding as it points towards inflammatory signalling as a potential pathway for LTR activation which could then lead to gene deregulation and disease.

3.4.5. Constitutive NF-κB activation can be triggered in Reh cells using doxycycline inducible IκKβ expression

The use of PMA as a trigger of inflammatory signalling leading to LTR activation worked effectively, however due to PMA affecting many cellular pathways by acting through PKC it was not possible to link LTR activation to a specific regulator. The presence of an NF-κB binding motif within the consensus sequence of THE1B LTRs and the findings by Lamprecht et al. 2010 suggest that NF-κB activation is the most likely reason for the LTR activation induced by PMA. To test this idea, we constructed a system to allow NF-κB activation to be induced in Reh cells by treatment with doxycycline (Figure 3.49). The advantage of this inducible system was that we were able to grow a stably transduced cell line and induce NF-κB activation at a short time prior to any downstream assays. This meant that we were more likely to observe the direct effects of NF-κB activation on LTRs rather than the indirect effects resulting from down-stream changes in gene expression produced by NF-κB.

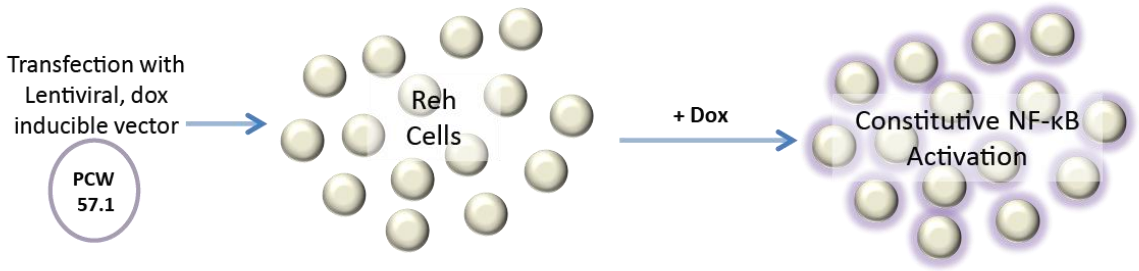


Figure 3.49. Inducible NF-κB activation scheme
Reh cells with doxycycline inducible NF-κB activation were produced by the transduction of cells with an inducible lentiviral expression vector (PCW 57.1) containing the *IκKβ* gene and IRES GFP. This produced a stable cell line in which NF-κB activation could be induced by treatment with doxycycline.

The doxycycline (dox) inducible system worked by expressing an *IκKβ* with the activating mutation S117E, S181E (IκKβ(EE)) upon dox treatment which is unable to become phosphorylated. Wild type IκKβ mediates phosphorylation of IκBα resulting in its proteasomal

degradation which, in turn, allows NF-κB to localise to the nucleus and become active (Figure 3.50) (Delhase, Hayakawa et al. 1999). IκKβ(EE) was cloned into the PCW 57.1 inducible lentiviral vector to achieve NF-kB constitutive activation (Delhase, *et al.* 1999) (See 2.8 for cloning strategy).

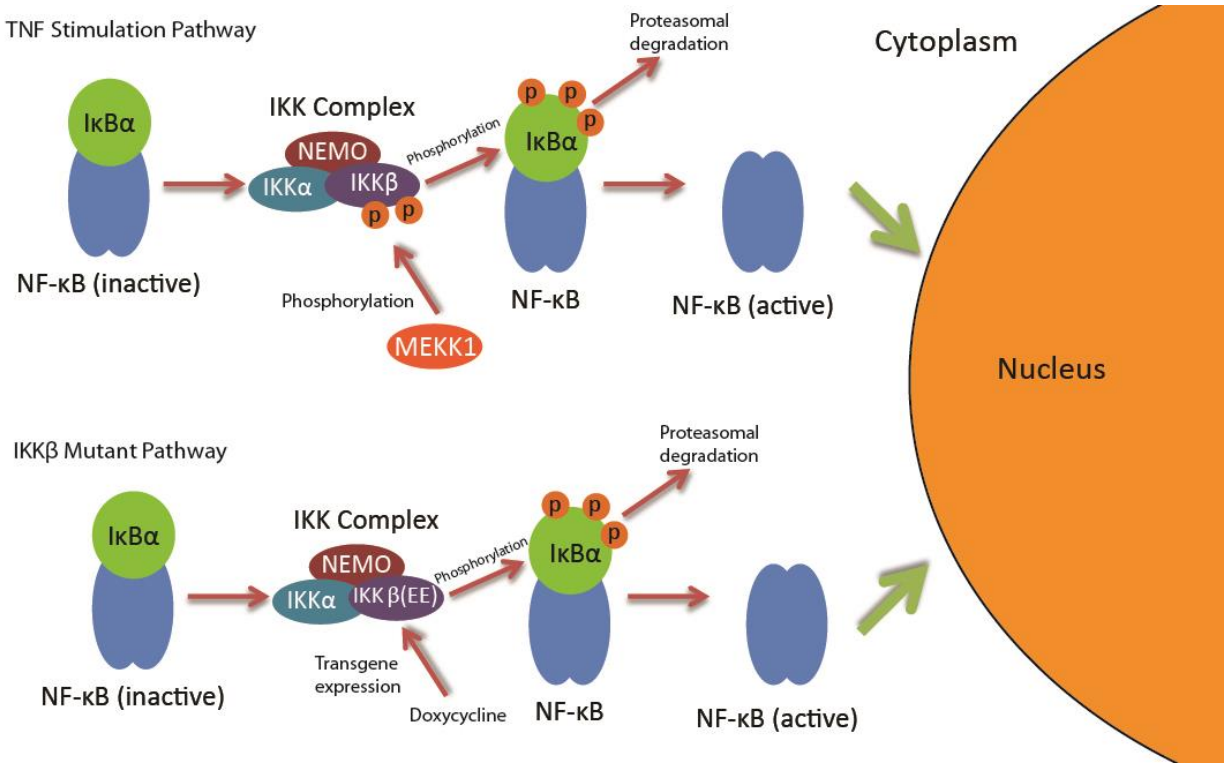


Figure 3.50. Scheme depicting the mechanism of NF-κB activation

Activation of NF-κB is achieved through over-expression of mutant IκKβ which does not require phosphorylation by MEKK1 and leads to the phosphorylation and degradation of IκBα allowing NF-κB to localise to the nucleus and become active.

An IRES GFP sequence was also added to the mutant *IκKβ* sequence to allow for easy monitoring of induction efficiency. The efficiency of induction was tested by flow cytometry analysis comparing the percentage of GFP positive cells with and without treatment with 2 µg/ml dox for 48 hours (Figure 3.51). Cells containing the *IκKβ* and *IRES GFP* (i*IκKβ*) construct showed an induction of GFP expression in ~33% of cells compared to only ~0.3% without dox

treatment. The i*IκKβ* cells were also compared to a control cell line with only an inducible IRES GFP (iGFP). The control cell line showed ~78% of cells induced following dox treatment. The reason for the reduced induction in the cells containing the *IκKβ* is likely due to the increased size of the inducible cassette making activation less stable. To observe the effect of *IκKβ* induction on the cells without the data being diluted by the cells which had not induced all further experiments were conducted on GFP positive cells sorted from the induced population by FACS.

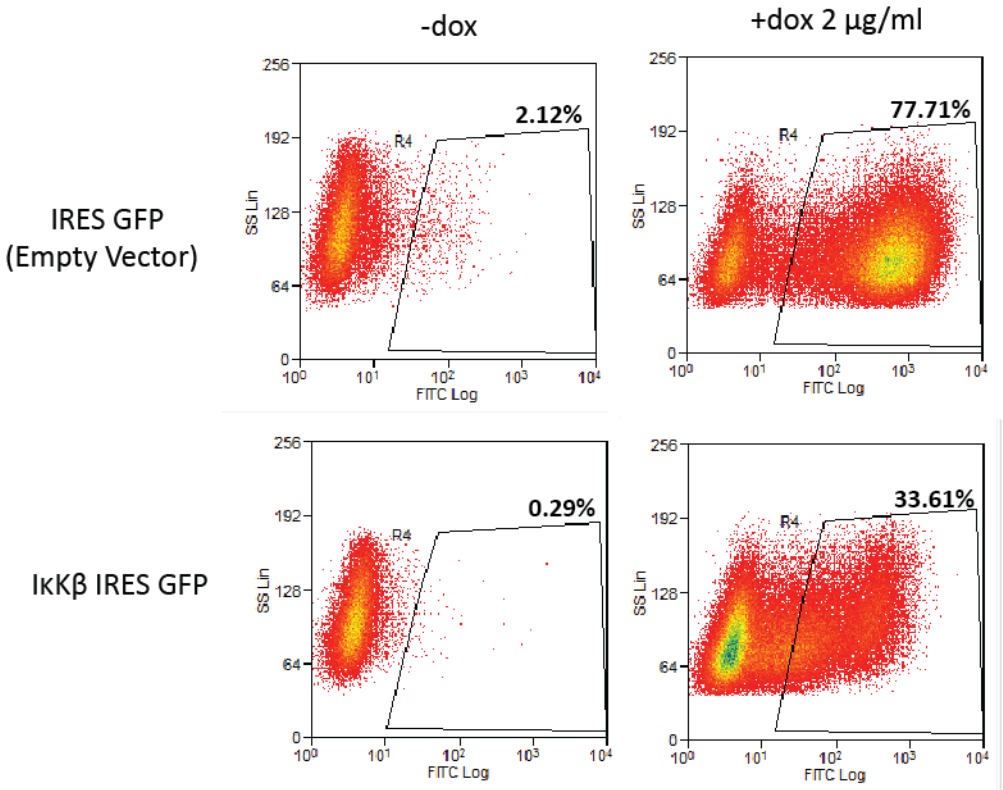


Figure 3.51. Flow cytometry analysis showed induction of IκKβ and IRES GFP construct following doxycycline treatment
Reh cells containing a dox inducible *IκKβ* and *IRES GFP* construct were treated with 2 µg/ml doxycycline for 48 hours and the GFP levels of cells measured by flow cytometry. This result showed a successful induction of the *IκKβ* and *IRES GFP* construct and the *IRES GFP* alone.

To ensure that expression of the *IκKβ* was being induced we performed qPCR to measure the expression of the endogenous *IκKβ* and the inducible mutant *IκKβ (*Figure 3.52*).* This analysis showed a fairly consistent level of endogenous *IκKβ* expression and little of the mutant *IκKβ* prior to dox treatment. Following dox treatment for 48 hours a large increase of mutant *IκKβ* expression was seen. We next wanted to establish the level of NF-κB activation in the cells by measuring the level of NF-κB protein localising to the nucleus. We measured the level of NF-κB protein within the nuclear fraction in dox induced iGFP cells, induced i*IκKβ* cells and in the L428 cell line. This experiment showed a ~3.5 fold up-regulation of nuclear NF-κB protein after induction which was higher than the constitutive level observed in the L428 cell line. By achieving a level of NF-κB activation at least as high as in the L428 cell line this shows that our inducible system should be fairly representative of the activation state in HL.

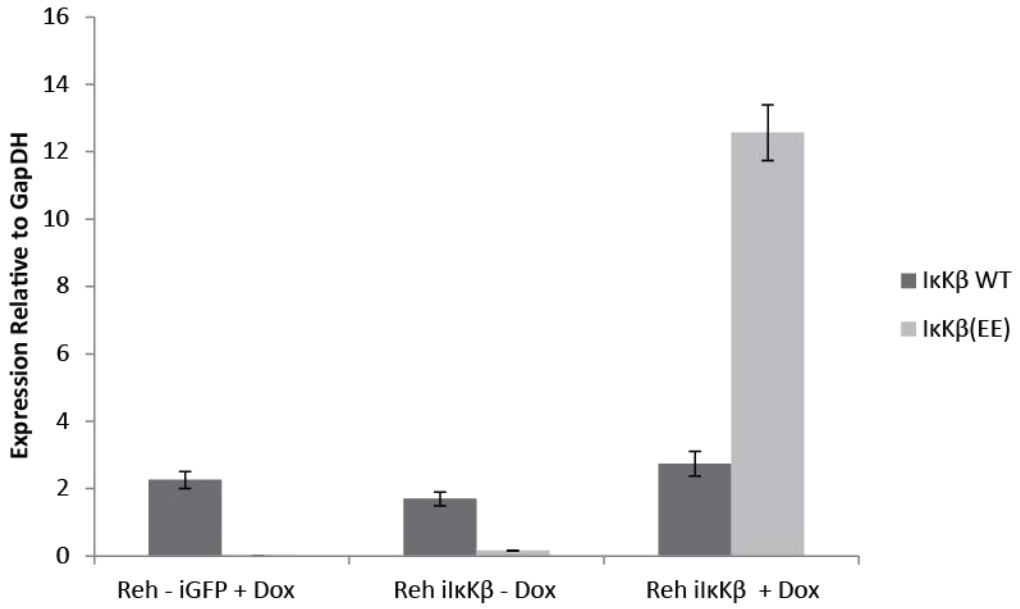


Figure 3.52. IκKβ(EE) is induced by dox treatment
iIκKβ cells were treated with2 µg/ml doxycycline for 48 hours and induction was measured by qPCR using primers deigned for either the endogenous IκKβ of the cloned mutant (IκKβ(EE)). This was compared to the iIκKβ with no dox treatment and the iGFP cell with dox treatment.

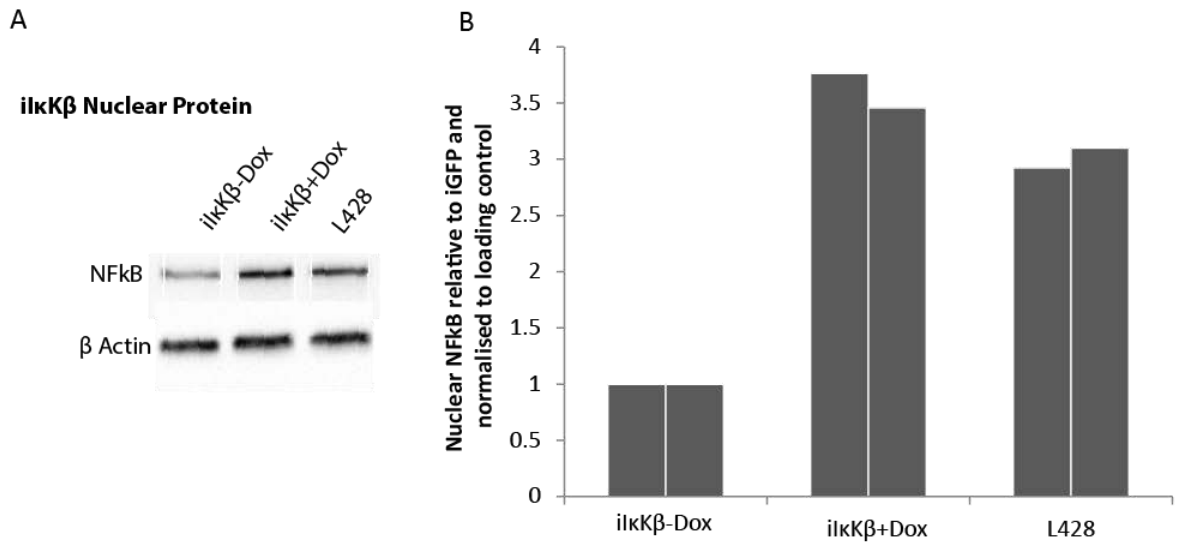


Figure 3.53. Nuclear NF-κB protein level increases following IκKβ induction A) The level of NF-κB protein in the nuclear fraction was measured by western blotting B) the level of protein calculated by densitometry relative to the level in iGFP cells and normalised to β-Actin loading control n=2.

3.4.6. LTR activity can be induced by aberrant NF-κB activation

After establishing that we were able to induce high levels of active NF-κB activation in Reh cells, we wanted to investigate the impact of NF-κB on genome-wide activation of LTRs. To achieve this aim, we carried out THE1B RACE-Seq in triplicate on the iGFP cell line with dox treatment and the iIκKβ cell line with (iIκKβ+Dox) and without (iIκKβ –Dox) dox treatment. The overlap of the data from the biological replicates of RACE-Seq data was determined and as in the previous approach, peaks appearing in at least 2 replicates were selected (Figure 3.54). We then compared the 2 control data sets (iGFP and iIκKβ-Dox) and the NF-κB induced dataset (iIκKβ+Dox) (Figure 3.55). This analysis showed a small number of peaks unique to each of the control datasets, most likely resulting from experimental variation. The remainder of the peaks within the control datasets were also present in the iIκKβ+Dox data which would be expected as all of these cells have the same Reh cell background. Finally the iIκKβ+Dox showed a large

number (660) of uniquely active LTRs, this confirmed that activation of NF-κB at a high level induces the activation of many THE1B related LTRs.

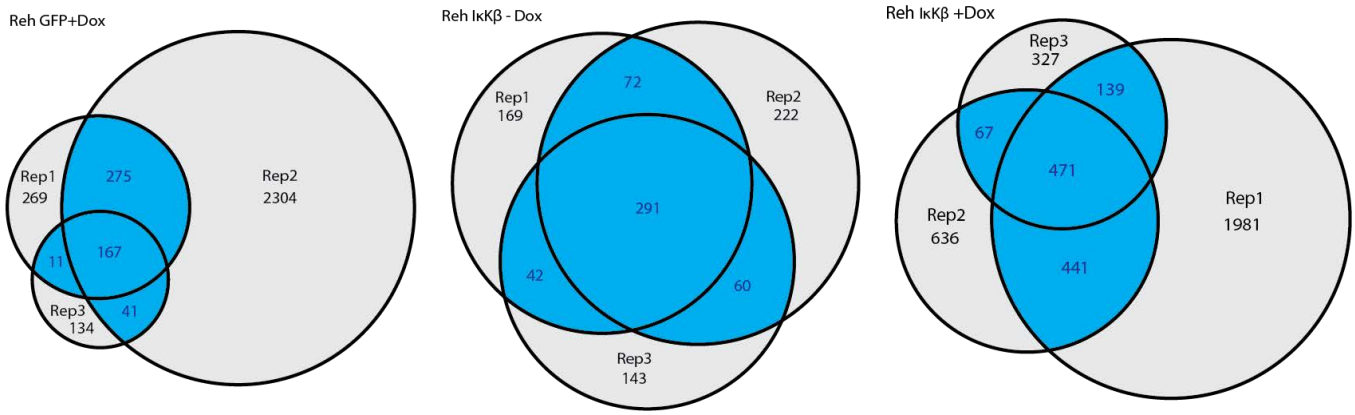


Figure 3.54. Overlap of RACE-Seq biological replicates
The peaks identified by Macs in the biological replicate THE1B RACE-Seq data sets for iGFP, iIκKβ-Dox and iIκKβ+Dox were overlapped and the peaks shared in at least to replicates (shown in blue) were selected for downstream analysis.

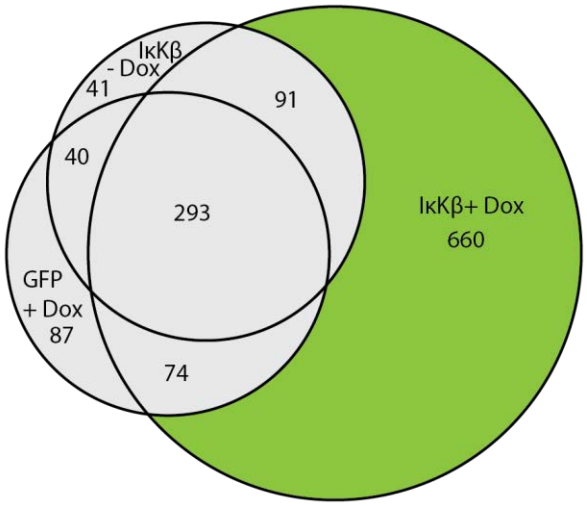


Figure 3.55. High levels of NF-κB activation induces activation of LTRs

Overlap of RACE-Seq data for active THE1B related LTRs with and without constitutive NF-κB activation induced by Dox treatment of cells. A large set (660) of uniquely active LTRs were identified in cell following NF-κB activation.

To study whether the set of active LTRs induced by NF-κB in Reh cells resembles that of HL cells, we compared the RACE-Seq data from the iIκKβ-Dox and iIκKβ+dox cells with all active 'high-confidence' LTRs identified in the HL cell lines(Figure 3.56). A proportion of the peaks (179) shared between the iIκKβ+Dox and iIκKβ-Dox datasets were also seen in the HL cell lines. This was to be expected as Reh cells share a number of active LTRs with the HL (see 3.2.4). A further 174 which were activated following Dox treatment were also shared with the HL cell lines. This indicates that NF-κB activation alone is able to induce activation of some HL specific LTRs. However, we found also 939 active LTRs identified in the IκKβ+Dox cells which were not shared with the HL cell lines. We also visualised this data in a supervised clustering analysis showing the shared and unique peaks between each cell line and the PMA treated and IκKβ+Dox cells (Figure 3.57). This analysis showed a large number of LTRs in each cell line and Reh cell treatment which were unique and only a small number shared between all of the HL and IκKβ+Dox cells.

It is possible that some LTRs may have been included or excluded through experimental variation, however due to the large number it is likely that this finding points towards additional control mechanisms preventing the activation of many LTRs purely through high levels of active NF-κB. There was also an absence of activation of the CSF1R LTR which is active in HL cells and was activated by PMA treatment. This finding points to the requirement of a different epigenome driven by ETO2 downregulation for the activation of at least a subset of the LTRs active in HL.

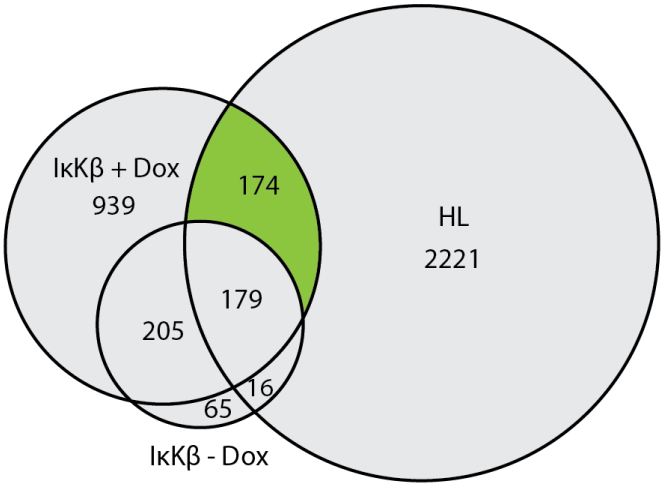


Figure 3.56. A proportion of LTRs active in HL is activated by constitutive NF-κB activation in Reh cells

The RACE-Seq data identifying active LTRs in control cells (IκKβ-Dox), cells with induced NF-κB activation (IκKβ+Dox) and in HL cell lines (L428, L1236 and KM-H2) was overlap to examine the LTR activation shared between the datasets. A proportion of the LTRs activated by NF-κB constitutive activation are also active in HL cell lines (shown in green).

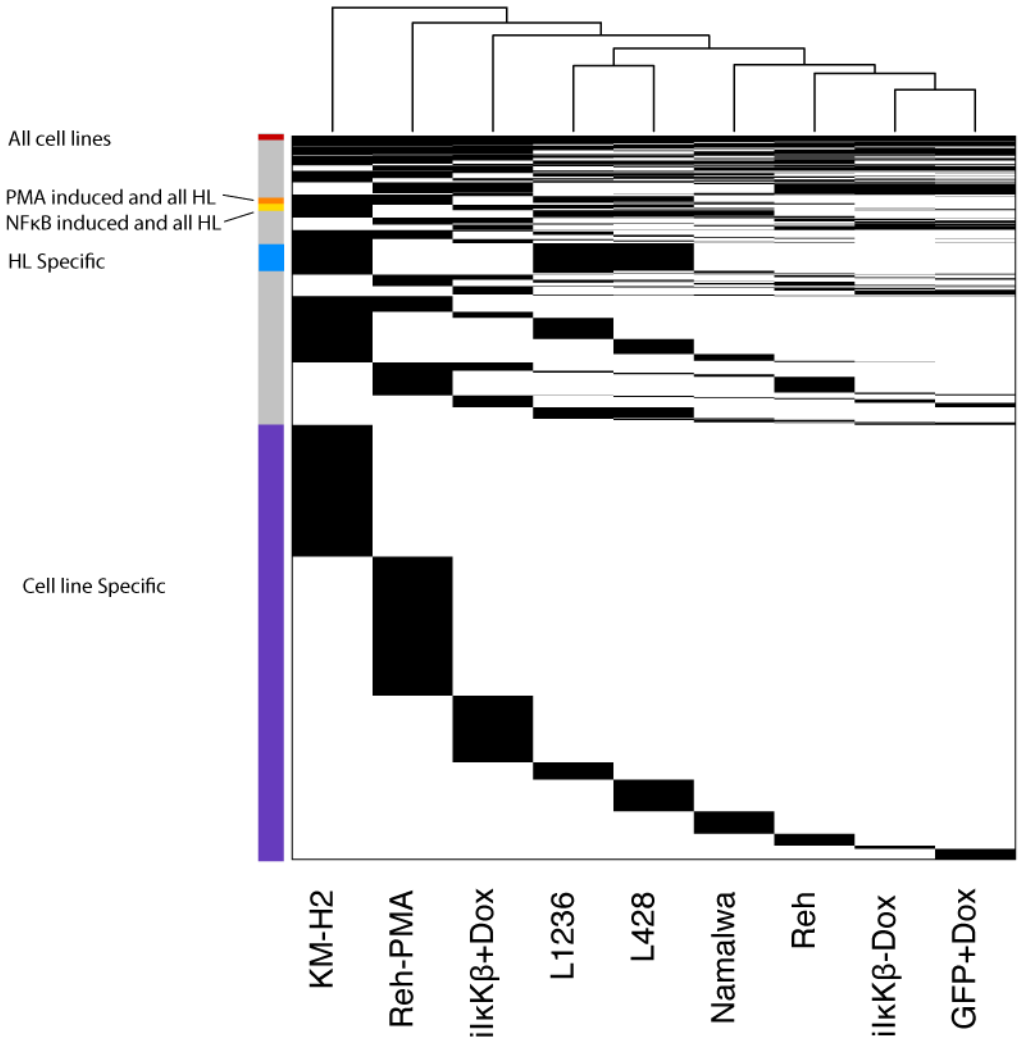


Figure 3.57 Supervised clustering of LTR activation across cell lines

Supervised clustering of LTR activation across all cell lines and IκKβ+Dox cells shows a unique pattern of LTR activation in each condition and also subsets of LTRs which are active in the HL cell lines and following PMA treatment of NF-κB activation.

Finally, to study the impact that constitutive NF-κB activation had on gene expression we performed RNA-Seq on the iIκKβ cells before and after doxycycline treatment. The overall gene expression pattern showed very little change in correlation with the NF-κB induced cells clustering with the un-induced cells. We did however observe an up-regulation of IκKβ which confirmed the functionality of the inducible system and we also saw the up-regulation of a number of the known HL markers analysed previously.

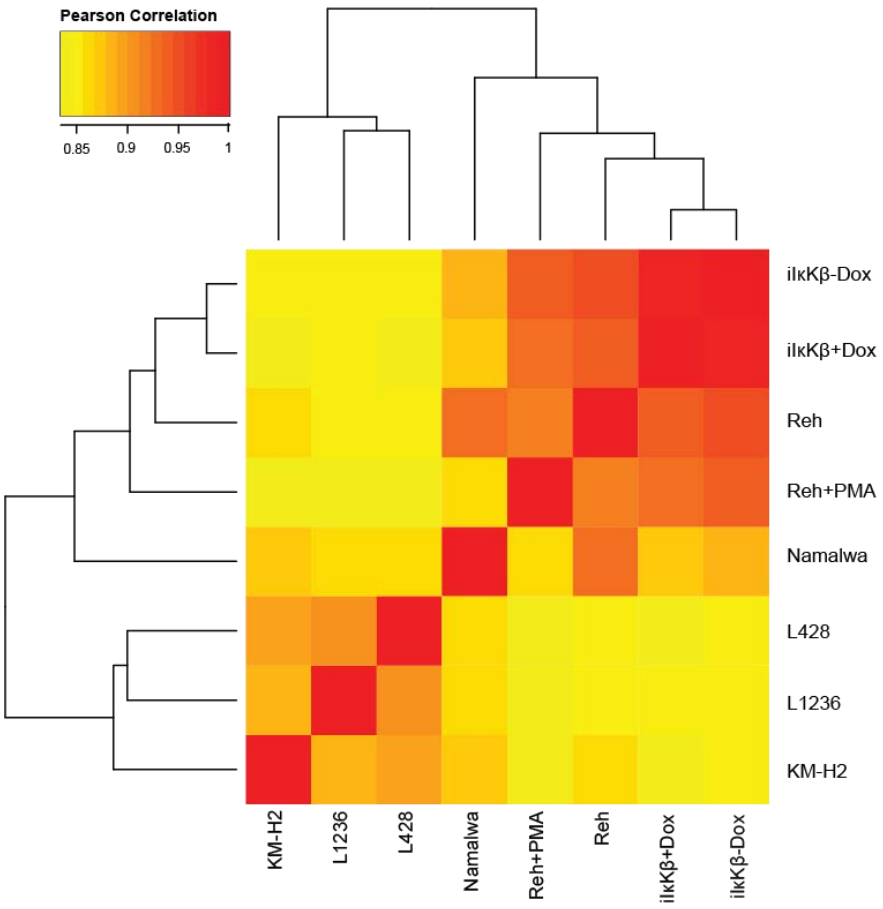


Figure 3.58 Clustering of gene expression data from cells with induced NF-κB activation showed little change in overall gene expression

Clustering of RNA-Seq data for HL cell lines, control cell lines and Reh cells with untreated, with PMA treatment and constitutive NF-κB activation (IκKβ+Dox). This analysis shows that induced constitutive NF-κB activation does not change the overall gene expression pattern in Reh cells at a 48 hour time point.

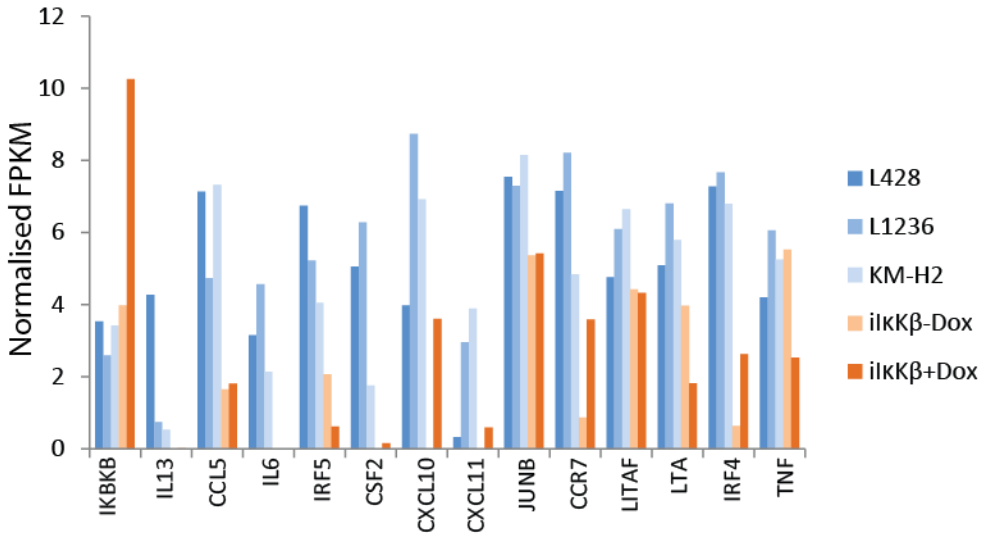


Figure 3.59 Up-regulation of HL gene expression following constitutive NF-κB activation

Following constitutive NF-κB activation by induction of IκKβ an up-regulation in gene expression of a number of genes known to be associated with HL was observed by RNA-Seq.

4. DISCUSSION

4.1. The Hodgkin's Lymphoma transcriptional network has a global impact on chromatin structure and gene expression.

4.1.1. Generating a novel RNA-Seq dataset

The HL gene expression network has been investigated in many studies and is known to differ greatly from the gene expression network of B-cells. B-cell lineage genes such as *BCR*, *CD19* and *CD20* are down-regulated and genes from many other cell lineages, particularly those with pro-inflammatory properties are up-regulated (Kuppers, Klein et al. 2003, Kuppers 2009). The majority of published HL gene expression studies use expression data obtained from qPCR and micro-array experiments. The disadvantage of this is that crucial data such as different isoforms of genes, low-abundance transcripts and alternative transcription start sites cannot be detected (Zhao, Fung-Leung et al. 2014). To overcome these disadvantages, particularly for the study of LTRs as alternative promoters, we performed RNA-Seq as no such data-set was available. RNA-Seq data allowed us to both study alternative isoforms of genes and also to further study the overall gene expression patterns in a set of HL cell lines (L428, L1236 and KM-H2). The gene expression patterns were then compared with those of control cell lines which more closely resemble the B-cell lineage (Reh and Namalwa).

The RNA-Seq data we produced confirmed the large difference in gene expression between the HL and control cell lines, and also showed a variation in gene expression between the HL cell lines. The variation in gene expression between the HL cell lines originates from 2 sources. Firstly the clonal expansion of the cell lines over many passages may have resulted in the introduction of random mutations or epigenetic drift in genes which have conferred further survival advantage and therefore been propagated through the cell line. The second and likely

predominant source of variation between the cell lines is the pathology of the patient samples from which they were derived.

All 3 HL cell lines were originally defined based on the expression of CD15 and CD30 surface markers and lack B-cell lineage antigens CD20 and CD19 and T-cell lineage antigens CD45 and CD3, which are the standard clinical markers for identification of HRS cells (Schaadt, Diehl et al. 1980, Kamesaki, Fukuhara et al. 1986, Wolf, Kapp et al. 1996, Piccaluga, Agostinelli et al. 2011). The L1236 and KM-H2 cell lines have highly complex karyotypes with at least 10% of each population displaying polyploidy and a proportion of cells in each population forming large multinucleated cells (Kanzler, Hansmann et al. 1996, Uehira, Amakawa et al. 2001). The L428 cell line also contains a proportion of multinucleated cells, however the karyotype is more stable than the other cell lines with all L428 cells lacking one copy of chromosome 14 and having an additional copy of chromosome 12 (Schaadt, Diehl et al. 1980).

Variation in the treatment received by the patients from whom the cell lines were derived is also likely to contribute to their gene expression patterns. The L428 and KM-H2 cell lines were both obtained from pleural effusions of relapse patients who had originally received chemotherapy and also radiotherapy in the case of L428 (Schaadt, Diehl et al. 1980, Kamesaki, Fukuhara et al. 1986). The L1236 cell line has a more complex background. L1236 cells originate from a patient who presented with advanced HL following treatment with radiotherapy of initial presentation. Chemo-therapy was then administered on a second relapse, surgery but no chemo- or radio-therapy on a third relapse and finally unsuccessful anti-CD25 immunotoxin therapy at the stage of a fourth relapse. Following the immunotoxin therapy the cell line was derived from separated lymphocytes obtained from peripheral blood (Wolf, Kapp et al. 1996). Although the overall phenotype of the cell lines matches that of primary HRS cells the treatment received by the patients prior to extraction of cells had the potential to impact significantly on the overall gene expression pattern, contributing to the gene expression differences between the

cell lines. It is also possible that the cell lines each represent a different subset of gene expression either specific to a particular case of HL or representing a particular subset of the disease.

Studies comparing the gene expression patterns of micro-dissected primary HRS cells and HL cell lines have shown that the expression of genes involved in cell proliferation and growth are significantly up-regulated in HL cell lines (Tiacci, Doring et al. 2012). However, we have shown that the main HL gene expression signature such as that dependent on constitutive NF-κB activation is present in the cell lines. This has also been confirmed in other studies suggesting that cell lines are still a good model for the study of HL (Steidl, Diepstra et al. 2012). This finding highlights the importance of comparing the results obtained in the cell lines to data from primary samples (See future work 4.6.5).

When analysing the overall gene expression patterns we observed a down-regulation of B-cell lineage genes in the HL cell lines when compared to the control cell lines. The loss of B-cell lineage genes in HL is well known; in particular the markers *CD19*, *CD38*, *CD79A* and *CD79B* are lost in the majority of patient samples (Tzankov, Zimpfer et al. 2003). We also observed a down-regulation of *EBF1* which plays a major role in down-regulating the B-cell lineage genes in HL (Bohle, Doring et al. 2013). *RAG1* and *RAG2* which are involved in recombination during B-cell development are also down-regulated in the HL cell lines in our gene expression data which has also been previously reported in patient samples (Meru, Jung et al. 2001). We also observed the down-regulation of BCL6 which acts as a transcriptional repressor in B-cells and is also expressed in the majority of lymphomas with the exception of classical Hodgkin's. It has been demonstrated that *BCL6* and *CD138* can be used to determine the stage in B-cell development which the malignant cells originated. The L1236 and KM-H2 cell lines express a very low level of *BCL6* and no *CD138* indicating that they originate from late germinal centre/early post germinal centre cells. These represent the most common type of HRS cells

found in ~59% of patients. The L428 cell line still lacks expression of CD138 but has a higher level of BCL6 indicating a far less common indeterminate phenotype seen in only ~18% of patients (Bai, Panoulas et al. 2006).

Our RNA-Seq data also confirmed an up-regulation in expression of genes associated with inflammatory signalling and pathways. HRS cells reside within an infiltrate of immune cells attracted by the HRS cells highly inflammatory signature, composed of a wide range of chemokines and cytokines. We observed up-regulation of a number of genes known to be expressed in HL including *CCL5* (*RANTES*) which is involved in the recruitment of T_H2 cells, T_{Reg} cells, eosinophils and mast cells and the inflammatory cytokines *IL-6* and *IL-13* which are involved in both B-Cell maturation and stimulation of growth in HL (Kurzrock, Redman et al. 1993, Kapp, Yeh et al. 1999, Skinnider and Mak 2002, Fischer, Juremalm et al. 2003, Aldinucci, Lorenzon et al. 2008). Many other genes known to be involved in the recruitment of immune response cells in HL were also up-regulated including *CCL17*, *CCL22*, *IL-5*, *IL-8* and *GM-CSF* (Skinnider and Mak 2002, Kuppers 2009).

We additionally found many up-regulated genes related to cytokine interactions which were shared with the JAK-STAT, TNF and NF-κB signalling pathways. These pathways form a complex network of interactions which jointly drive the survival and proliferation of HL cells. The JAK-STAT (Janus kinase/signal transducers and activators of transcription) pathway is central to many processes within the cell and is the signalling mechanism for cytokines and growth factors to induce cell proliferation, differentiation and apoptosis (Rawlings, Rosler et al. 2004). Signalling through the Jak-STAT pathway is required for HL cell survival which is driven in part through the cells own production of cytokines. It has been shown that in HL STAT5a can be activated through interaction with NF-κB signalling which is also upregulated in HL (Hinz, Lemke et al. 2002). This particular interaction is not seen in our KEGG pathway analysis for deregulated genes in HL. The absence of a number of interactions specific to HL is due to the

pathways being based around signalling in healthy tissue and not specific to the deregulated signalling in HL.

There is also evidence from a number of studies that members of the TNF pathway can interact directly to activate the JAK-STAT pathway, particularly as part of an inflammatory response (Guo, Dunbar et al. 1998, Ahmad, Ansari et al. 2015). The TNF pathway can also drive activation of NF-κB signalling through both the canonical and non-canonical pathways (Wajant, Pfizenmaier et al. 2003, Sun 2011). Reciprocally NF-κB can also drive TNF signalling which can trigger further inflammation (Dong, Jimi et al. 2010).

Besides the activation of STAT, constitutive NF-κB signalling is also central to driving the HL phenotype and inhibition of NF-κB results in rapid apoptosis (Izban, Ergin et al. 2001). Members of the NF-κB transcription factor family play an important role in the activation of inflammatory and immune responses. In healthy lymphocytes strictly controlled NF-κB activation promotes expression of caspase inhibitors to prevent apoptosis, cytokines involved in proliferation and cell survival including IL2 and IL6 and also a number of cell cycle regulators (Jost and Ruland 2007). Due to the loss of BCR signalling in HL, aberrant NF-κB signalling is required for cell survival. Up-regulated NF-κB signalling results from interaction of a number of factors seen in our pathway analysis. CD40, which we show to be shared between the NF-κB signalling, cytokine-cytokine receptor interactions and the transcriptional misregulation in cancer pathways, is an important activator of NF-κB signalling in HL (Gruss, Hirschstein et al. 1994, O'Grady, Stewart et al. 1994). The ligand of CD40 is expressed by CD4+ T-cells which accumulate around HRS cells leading to activation of NF-κB through TRAF proteins. Internal mechanisms of NF-κB activation also include TNFRSF8 (CD30) up-regulation which undergoes self-oligomerization to recruit TRAF proteins to activate NF-κB. Up-regulation of TNFRSF8 in particular has been demonstrated as a requirement for HL cell survival (Thakar, Ovchinnikov et

al. 2015). We have also shown that upregulation of TNFRSF11A also plays an important role in NF-κB activation in HL (Discussed in 4.3.2).

The pathways which we have identified form a complex inflammatory response network which has been shown across many studies to drive the survival and proliferation of HL cells (Kuppers 2009). To achieve the diverse and inflammatory gene expression pattern in HL cells significant changes to gene expression regulation are required. So far no RNA-Seq data from HL cell lines have been published and our new data will provide an important resource to the field. Using the RNA-Seq data we have been able to show the presence of previously unannotated non-coding transcripts and anti-sense transcripts that impact gene expression, for example in the case of the *CHD1L* gene. We have also been able to integrate this data with RACE-Seq data to show unannotated alternative transcription start sites.

4.1.2. Generating a high-resolution DNAseI-Seq dataset for digital foot-printing

Kreher, *et al.* 2014 published an analysis using low resolution DNaseI-Seq data identifying the cis-regulatory regions driving the HL-specific gene expression pattern. From an analysis of enriched motifs within DHS peaks they identified IRF5, which we also see upregulated in all HL cell lines, as a major regulator of the HL phenotype. Due to the low resolution of their DNase-I data they were unable to identify the full complement of HL-specific transcription factor binding sites. Moreover, from this data it is unclear which binding motifs are actually occupied. We therefore performed a high read-depth DNaseI-Seq experiment to generate better data and to enable us to carry out high-resolution digital foot-printing to determine the full complement of bound transcription factors. Our data show an enrichment of occupied IRF, NF-κB and AP-1 motifs in distal, HL specific, foot-printed regions when compared to Reh cells. This showed that

expression of distal regulatory regions in HL is driven by IRF, NF-κB and AP-1 which also confirms the findings of Kreher, et al. 2014.

The up-regulation of both NF-κB and AP-1 are hallmarks of the regulatory network of HL gene expression as shown by a number of studies (Mathas, Hinz et al. 2002, Kreher, Bouhlel et al. 2014). The AP-1 family transcription factors function by the formation of a DNA binding heterodimer between c-Jun or JunB and c-Fos, all of which were up-regulated in our HL expression data (Shaulian and Karin 2002). AP-1 activation usually occurs as a result of a number of stimuli including cytokines, growth factors and infections which means that in HL activation most likely occurs as a result of a number of inflammatory mechanisms (Hess, Angel et al. 2004). AP-1 activation pathways include signalling through JNK, MAPK and NF-κB phenotype (Chang and Karin 2001) all of which are central to the HL. It has been shown that in HL, AP-1 DNA binding activity is driven by IRF5 which induces JUN, JUNB and ATF3 forming a HL-specific AP-1 complex.

To determine which transcription factors were likely to be interacting in the regulation of HL gene expression we performed bootstrapping analysis which tests the probability of the co-occurrence of transcription factor motifs within foot-printed regions, indicative of a interacting factor complex. Our data showed AP-1 and NF-κB as the main drivers of gene expression in HL cells both by co-localising with each other and with a number of other transcription factors including GATA, IRF, STAT and RUNX. Interestingly GATA is not expressed in B cells but is aberrantly expressed in HL, the finding that GATA motifs co-localise with the AP-1 and NF-κB motifs shows that GATA participates in the regulation of HL gene expression. The likely interaction between these factors in the activation of the unique HL gene expression pattern is an exciting finding. AP-1 and NF-κB have previously been shown to be involved in the regulation of CD44 a pro-oncogene in breast cancer (Smith and Cai 2012). This study went on to show that AP-1 and NF-κB have been implicated in a number of other inflammation driven

diseases including keratitis, rheumatoid arthritis and hepatocellular carcinomas (Han, Boyle et al. 1998, Liu, Kimmoun et al. 2002, Dong, Jimi et al. 2010). Interestingly hepatocellular carcinoma and rheumatoid arthritis also both exhibit a genome-wide pattern of ERV expression which may implicate AP-1 and NF-κB in LTR activation in inflammatory diseases. It is also known that THE1B LTRs have binding sites for both NF-κB and AP-1 (Kreher, Bouhlel et al. 2014). To our knowledge there is currently no published data to show that AP-1 and NF-κB physically interact however their interaction with the same cis-element has been reported in the expression of CD44 in breast cancer (Smith and Cai 2012).

4.2. RACE-Seq is an effective technique for the genome-wide identification of specific types of active LTRs

Lamprecht et al. (2010) showed for the first time that an active LTR could induce pathological gene expression required for cell survival in a disease state. The finding that an active THE1B LTR in HL was acting as the sole promoter for *CSF1R* and that the expression of this gene was required for cell survival opened up the possibility that LTRs activation could be playing an important role in disease. To investigate the genome-wide activation of LTRs related to the THE1B family in HL we developed a targeted next-generation sequencing approach based on the 5' RACE technique. The 5' RACE technique was pioneered in the 1980's, originally developed as a method for cloning low-abundance mRNA transcripts (Frohman, Dush et al. 1988). The technique has since been repurposed and is now used mainly to identify the transcription start site for alternative isoforms of genes, such as genes which have an LTR acting as an alternative promoter (Lamprecht, Walter et al. 2010, Babaian, Romanish et al. 2016). In its simplest form 5' RACE works by the ligation of a known adaptor sequence to the 5'

end of transcripts which can then be used in conjunction with a single primer located within the transcript in a PCR reaction to amplify the region between the 5' end and the primer.

To harness RACE technology for genome-wide screening of LTR activation we used a degenerate primer based on the most conserved region of the THE1B LTR consensus sequence. By carrying out PCR using a low annealing temperature to promoter the binding of the primer with mismatches we were able to successfully produce genome-wide libraries of THE1B LTR activation. When combined with Illumina next-generation sequencing this allowed for datasets to be produced showing the global activation of LTRs in cell lines. The RACE-Seq technique was also able to identify a number of other related LTR families. It is however not likely that the full complement of activated LTRs in families other than THE1B is seen in our data as it is known that the efficiency of PCR reactions drops rapidly with the introduction of mismatching nucleotides between the primer and template sequence (Bru, Martin-Laurent et al. 2008). The use of a degenerate primer also introduced an issue with regard to the reproducibility of the assay. Due to the primer annealing with 8 or more mismatches 50% of the time when compared to the genomic sequence the amplification of LTRs with greater sequence variation produced variable levels of reads in each replicate. It may be possible to overcome this issue by higher read-depth sequencing of the libraries to allow for detection of the less amplified transcripts. The advantage of higher read-depth was seen in the 3rd replicate of RACE-Seq on the KM-H2 cell line which showed higher overlaps with replicates 1 and 2 due to an overall higher read-depth. To our knowledge there are currently no published studies which assess the influence of read-depth on the identification of regions selected by PCR however in terms of peak size, ChIP-Seq for factors binding to narrow regions (point-source factors) is the nearest comparison. In ChIP-Seq for point-source factors such as sequence-specific transcription factors a read-depth of 20 million reads is recommended by the ENCODE project and it has been shown that saturation for identification of all peaks can require up to 100 million reads

(Sims, Sudbery et al. 2014). The number of peaks expected in a ChIP-Seq experiment may be greater than what is expected for the number of active LTRs in a particular cell type and also the specificity of the reads should be higher in RACE-Seq due to the targeted amplification. However it would suggest that an increase from the average 5 million reads per replicate currently obtained in our RACE-Seq experiments may improve the reproducibility.

There are very few other reports of RACE being used in conjunction with next-generation sequencing (NGS) approaches. 5' RACE has recently been used with NGS to determine the multiple transcription start sites of 2 genes which have a large amount of variation in their start site (Leenen, Vernocchi et al. 2016). 'RACE-Seq' has also been used for the investigation of long non-coding RNAs for which the transcription start site is unknown. 'RACE-Seq' was carried out on 398 lncRNAs as a proof of principle and fragments amplified between the 5' or 3' end and a known exon within the lncRNA followed by long-read high-throughput sequencing (Lagarde, Uszczynska-Ratajczak et al. 2016). As far as we are aware our RACE-Seq approach is novel, particularly in terms of the identification of active LTRs.

Other techniques have been developed for the purpose of identifying genome-wide LTR activation. Both CAGE (Capped Analysis of Gene Ends) and RNA-Seq have been used in other studies to identify LTR activation. The RNA-seq techniques are based on the development of bioinformatics pipelines which align reads specifically to ERV sequences prior to mapping these back to the genome (Sokol, Jessen et al. 2016). The downside of this technique is that due to the limited coverage of active LTRs in an entire RNA-Seq library only a small number of reads are likely to be able to uniquely map in the vicinity of active LTRs. It is also not possible with this technique to conclusively determine whether the LTR is acting as a transcription start site or is expressed as part of another transcript. The use of CAGE technology overcomes this problem as it specifically identifies transcription start sites. CAGE has been successfully used to identify the expression of ncRNAs driven by LTR promoters in virus-induced hepatocellular carcinoma

(Hashimoto, Suzuki et al. 2015). The expression of families of ERVs in a range of healthy tissue including developmental tissue has also been shown by CAGE (Faulkner, Kimura et al. 2009, Fort, Hashimoto et al. 2014). These studies show that CAGE data can be successfully used for the identification of active ERVs. However, the unique alignment of CAGE reads to repeated regions of the genome is potentially more challenging than with RACE-Seq due to the very short read length lacking unique binding data. To further investigate the efficacy of the 2 techniques in identifying the full complement of active LTRs CAGE data would be required for the HL cell lines.

Due to the challenges of aligning uniquely to repeat elements which have a high sequence homology we believe that our targeted approach is likely to be able to identify a greater proportion of the active THE1B LTRs in a cell line (see future work 4.6.1).

4.3. HL cells display wide-spread activation of long terminal repeat elements

The potential for genome-wide activation of LTRs in HL was demonstrated by Lamprecht at al. (2010), who identified an overall activation of LTRs in HL by using both 3' and 5' RACE followed by cloning. The study showed that a number of members of the THE1 family were active in HL including THE1A, THE1B and THE1C, but did not attempt to map all of the expressed LTRs genome-wide. To follow up on the findings of Lamprecht at al. (2010) we aimed to identify the full complement of LTR activation in HL cell lines. By using RACE-Seq we were able to identify the activation of a novel set of LTRs specific to HL when compared to our control cell lines. A number of studies have identified active LTRs in both diseased and normal tissue however there are few published datasets of genome-wide LTR activation. Genome-wide disease specific LTR activation has been shown in hepatocellular carcinoma by the use of CAGE (Capped Analysis of Gene Ends) to identify novel transcription start sites (Hashimoto, Suzuki et al. 2015). This analysis identified 935 active distal LTRs which were specific to hepatocellular

carcinoma (HCC) tumour cells and not transcribed in healthy hepatocytes. Mining of RNA-Seq data from diffuse large B-cell lymphoma patient samples has also been used to identify ERVs acting as alternative promoters (Lock, Rebollo et al. 2014). It is clear based on these studies that the genome-wide activation of LTRs in cancerous cells is not an exclusive phenomenon to HL cells and this is supported by our observation that our control cell lines also exhibit LTR activation.

A unique set of activated LTRs in each cell line, both HL and control was also displayed in our data. It is possible that some of the variation between cell lines is due to activation of a proportion of LTRs being missed due to low sequencing depth. However, we are able to show by statistical comparison of the cell line specific transcribed LTRs to the total LTR population that the cell line specific LTR activation is significant. Variation in LTR expression between the cell lines most likely results from the different epigenetic background of the both the HL and control cell lines (discussed in 4.1.1). Variation in LTR activation between different but closely related cell types has been shown during early embryonic development (See 1.3.14). In this study it was shown that ~2% of RNA-Seq reads at the 2 and 8 cell stage of human embryonic development map to ERVs including some of the THE1 family. They also observed that the proportion of active ERVs reduced as embryonic development progressed and also that the expressed ERVs changed throughout development marking specific cell populations. The epigenetic landscape during early embryonic development is known to change rapidly so the variation in LTR activation in not surprising. However, it does highlight the possibility that changes in LTR activation can to some degree define a cell type. The large variability in active LTRs which we observed in HL is an exciting finding as the variability in LTR activation may define different subsets of the disease and also may contribute to both the gene expression and chromatin accessibility variation seen between the HL cell lines.

4.3.1. LTR activation contributes to global deregulation of gene expression in HL cell lines.

A number of studies have shown that expressed LTRs can act as promoters and enhancers. The exaptation of LTRs as promoters in HL has been shown at both the CSF1R gene and the IRF5 gene (Lamprecht, Walter et al. 2010, Babaian, Romanish et al. 2016). Based on these findings we wanted to determine whether other expressed LTRs in HL were also acting as promoters. To study the promoter activity of LTRs in HL we performed ChIP-Seq for the active promoter histone mark H3K4me3.

We observed that 11.86% of LTRs corresponded to genomic regions containing H3K4me3 in the L428 cell line compared to only 5.03% in the Reh cell line. H3K4me3 is known to mark active promoters and has been previously shown at transcribed LTRs acting as promoters in HCC tumour cells (Liang, Lin et al. 2004, Hashimoto, Suzuki et al. 2015). The presence of H3K4me3 at only a small percentage of expressed LTRs does not necessarily accurately reflect that only these LTRs are active promoters. The alignment of short fragments produced by the ChIP-Seq experiments is intrinsically challenging due to the shared nature of sequence at repeat elements which means that reads either have to be assigned to all matching sequences or at random to one of the sites with corresponding sequence (Trapnell and Salzberg 2009). The potential alignment issues mean that many of the ChIP-Seq reads located around LTRs may be lost. It is however also possible that active LTRs are not active in all cells simultaneously and may not be captured by ChIP which is less sensitive than RACE due to its reliance on antibody binding. It is also possible that some LTRs may harbour alternative histone marks such as the enhancer mark H3K4me1. H3K4me1 has been observed at ERV-9 LTRs which act as promoters for the transcription of lncRNAs in human erythroblasts and at ERV driven enhancers in many cell types (Xie, Hong et al. 2013, Hu, Pi et al. 2017). An array of activating histone marks has also been described at LTRs in $CD4^{+}$T cells including H3K37me1,

H3K36me1 and H3K27me1 (Huda, Marino-Ramirez et al. 2010). This implies that the absence of H3K4me3 on active LTRs in HL does not mean that these LTRs have no impact on gene expression in the HL cells but may indicate that a proportion are acting as enhancers.

We also screened active LTRs for DNaseI hypersensitive sites (DHS) primarily with the aim of developing computational methods of identifying active LTRs in DHS data. Following high read-depth DNaseI-Seq only a proportion of active LTRs were shown to coincide with a DHS. This finding was difficult to understand as to be transcriptionally active, chromatin needs to be in an accessible state. However, another study reported a similar finding that active repeat elements could not be identified in DNaseI-Seq or ATAC-Seq datasets but could be identified by Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-Seq). They also went on to show, by MNase, that the sites of active repeat elements were located between phased nucleosomes. This led to the hypothesis that the repeat regions have unstable nucleosome binding so are less susceptible to digestion by DNaseI but are susceptible to the biochemical extraction process of FAIRE-Seq (Gomez, Hepperla et al. 2016).

To determine the influence of active LTRs on gene expression in HL we integrated our RACE-Seq and RNA-Seq data which showed clustering of active LTRs with up-regulated genes in each of the HL cell lines. Clustering of the cell lines based on correlation of gene expression for genes in the vicinity of active LTRs also showed the same clusters as correlation based on total gene expression. These findings would suggest that active LTRs in HL influence the expression of their nearest genes and in part define the expression pattern of each cell line. There are few other studies which compare actively transcribed LTRs to the expression of coding genes on a genome-wide basis, however there are a number of studies which identify individual genes regulated by LTRs (See 1.3.13). In HL cell lines we identified a number of cases where an active LTR could be linked directly to the expression of a down-stream transcript. Based on the transcripts observed we defined 4 different types of transcript; expression of a coding gene from

an alternative up-stream promoter, a shorter isoform of a coding gene from an intronic promoter, an anti-sense transcript or a lncRNA. The observation of these different transcripts originating from an LTR is not exclusive to HL cell lines and is seen across published studies of LTR activation (reviewed by (Babaian and Mager 2016)).

4.3.2. HL specific genes transcribed from activated LTRs may be part of HL pathology

We identified 3 new protein-coding genes which were directly up-regulated as a result of LTR promoters in HL. These genes were initially defined based on the presence of a read-through transcript, displayed in the RNA-Seq data, originating from the LTR to the downstream gene. To be able to find genes based on this technique requires manual curation therefore there are likely to be other LTR driven genes in HL yet to be identified. The first gene we identified was *NLRP1* which is up-regulated from an up-stream THE1C LTR.

NLRP1 (NLR family pyrin domain-containing 1) is a cytosolic pattern recognition receptor which triggers an immune response to a microbial infection (Chavarria-Smith and Vance 2015). Following the recognition of a pathogen NLRP1 activates CASPASE-1 through interaction with its CARD (caspase activating and recruitment) domain. In turn CASPASE-1 cleaves precursors to the inflammatory cytokines Il-1β and Il-18 enabling these cytokines to be secreted triggering an immune response (Cerretti, Kozlosky et al. 1992, Ghayur, Banerjee et al. 1997). CASPASE-1 is also able to induce cell death via pyropotosis in the absence of Il-1β and Il-18 (Masters, Gerlic et al. 2012). A recent study has shown that the up-regulation of *NLRP1* expression can promote tumour growth through increasing inflammatory signalling in metastatic melanoma (Zhai, Liu et al. 2017). The findings in melanoma point towards a possible pathological function of NLRP1 which could also apply following the upregulation which we observed in HL. However, the inflammatory function that NLRP1 plays in melanoma is through the activation of Il-1β which based on our RNA-Seq data is not expressed in HL. A number of other potential targets for

caspase-1 cleavage have been identified but further study is required to determine whether these can be linked to an inflammatory response (Denes, Lopez-Castejon et al. 2012). Further work would be required to determine whether the up-regulation of *NLRP1* resulting from LTR activation plays a significant role in the HL phenotype or whether it is simply a non-functional by-product of LTR activation.

We have also shown that *WNT5A* is expressed in the KM-H2 cell line from an active THE1D LTR. WNT5A is a member of the WNT family of ligands which are involved in development and many cellular processes (Logan and Nusse 2004). WNT proteins interact with the Frizzled (Fzd) family of receptors and in particular WNT5A interacts with Fzd3, Fzd4, Fzd5 and Fz8 (Kikuchi, Yamamoto et al. 2012, Linke, Zaunig et al. 2017). WNT5A activates the β-catenin independent pathway, which is best known for the modulation of cell movement at the stage of gastrulation during embryogenesis (Veeman, Axelrod et al. 2003). The involvement of WNT5A in cell motility has been identified as an important factor in the metastasis of a number of cancers including gastric, brain, colon and breast cancer (Kurayoshi, Oue et al. 2006, Klemm, Bleckmann et al. 2011, Bakker, Das et al. 2013). A number of studies have noted the up-regulation in gene expression of *WNT5A* in HL, however none have identified that this originates from a promoter in an upstream LTR (Klemm, Bleckmann et al. 2011, Tiacci, Doring et al. 2012, Linke, Zaunig et al. 2017). A recent study implicated WNT5A in cell migration in HL cell lines and the potential for this activity in HL tumours. The study showed that WNT5A interacted with the Fzd5 receptor which activates DVL3 leading to activation of RHOA which regulates the actin cytoskeleton and promotes motility in the cells (Tkach, Bock et al. 2005). When combined with our findings with regard to the role of LTR activation in the expression of WNT5A, this may suggest that LTR activation promotes metastasis of HL tumours.

The WNT signalling pathways also play a role in a number of other cellular processes which may be implicated in HL. The interaction of WNT5A with Fzd5 does not only activate DVL3 but

also upregulates PKC signalling through release of CA^{2+} which has been shown to contribute to increase invasion of melanoma cells (Weeraratna, Jiang et al. 2002). In HL PKC signalling also has the potential to increase activation NF-κB which in turn is essential for the maintenance of the HL gene expression pattern, particularly in relation to the activation of LTRs. There is also evidence that PKC activity plays a role in the DNA binding capacity of both NF-κB and AP-1 (Li, Ping et al. 2000). The interaction of WNT5A with PKC and the potential for involvement in NF-κB and AP-1 activity may point towards a positive feedback loop in the regulation of the WNT5A LTR.

Our study also found that the expression of *TNFRSF11A* (*RANK*) is driven from a promoter within a THE1B LTR. The *TNFRSF11A* gene has been identified in a number of studies to be up-regulated in HL cell lines and in an average of 75% of HRS cells from primary tumour samples (Fiumara, Snell et al. 2001, Kuppers 2009). However, the expression of an isoform originating from an LTR promoter had previously not been reported. TNFRSF11A is a member of the TNF receptor family and acts as the receptor for TNFSF11A (RANKL) (Anderson, Maraskovsky et al. 1997). Activated TNFRSF11A interacts with TNF receptor associated factors (TRAFs) to up-regulate signalling in a number of pathways including NF-κB and JNK (Anderson, Maraskovsky et al. 1997, Darnay, Ni et al. 1999). A study into the impact of TNFRSF11A signalling on HL cell lines L428 and KM-H2 has shown that following incubation of cells with TNFSF11A the level of NF-κB activation increases (Fiumara, Snell et al. 2001). Because expression of the TNFSF11A ligand has also been shown in HL this would suggest TNFRSF11A expression may play a significant role in the up-regulation of NF-κB activation in HL cells (Fiumara, Snell et al. 2001). Further to this it has also been shown that the overexpression of TNFRSF11A even in the absence of its ligand can increase the activation of NF-κB (Anderson, Maraskovsky et al. 1997). To investigate the influence of *TNFRSF11A* expression on NF-κB activation in HL cell lines we performed a siRNA knockdown of

TNFRSF11A in the L1236 cell line, which our data showed to have the highest gene expression. The knockdown resulted in reduced NF-κB activation in the cell line although no obvious change in phenotype. This finding demonstrates another potential positive feedback loop for the activation of LTRs in HL.

Our study also highlighted the ability of transcribed LTRs to produce anti-sense RNA transcripts and potentially reduce the expression of a gene by RNA interference. In particular we observed the reduction in expression of the *CHD1L* gene which contains an LTR driven anti-sense transcript. CHD1L has been shown to carry out chromatin remodelling functions in early development and may acts as a transcription factor (Chen, Chan et al. 2010, Snider, Leong et al. 2013). The overexpression of CHD1L is linked to hepatocellular carcinoma and overexpression in mice promotes spontaneous tumour formation (Chen, Huang et al. 2009). There is no published data with regard to the impact of reduced expression of *CHD1L*, however its role in DNA damage repair may suggest that the genome would be more susceptible to damage (Ahel, Horejsi et al. 2009). In HL it is likely that the knock-down in gene expression resulting from the anti-sense RNA has no overall impact on the cell phenotype, however it does highlight the potential for LTR driven down-regulation of gene expression.

Our final observation with regards to LTR driven transcripts concerned a high abundance of unannotated lncRNAs resulting from active LTRs. These may be non-functional transcripts which are degraded soon after transcription and are formed simply as a result of an LTR promoter being actively transcribed. Long non-coding RNAs, however, have been shown to play a regulatory role in gene expression through interaction with chromatin remodelling, transcription and post-transcriptional processing (Cao 2014). The production of lncRNAs and vlincRNAs (very long non-coding RNAs) from LTR promoters has also been observed in other studies and linked to driving the progression of cancer. Knock-down of 10 vlncRNAs specific to the K562 myeloid leukaemia cell line and expressed from ERV promoters resulted in an

increase in apoptosis within the cells showing that ERV driven vlncRNAs can promote cell survival and potentially cancer cell growth (St Laurent, Shtokalo et al. 2013).

The remaining transcribed LTRs identified in our RACE data show no associated RNA-seq transcripts. The lack of RNA-seq transcripts at the LTRs is most likely due to the inability to generate unique alignments when mapping the RNA-Seq reads, however the lack of downstream transcript would suggest that these LTRs are not acting as promoters. The lack of transcript does not necessarily mean that the LTRs have no function though as they may be acting as enhancers for topologically close promoters. A study of the DNA methylation of a range of cell types including both foetal tissue and adult epithelial and haematopoietic cells identified a large number of TEs with the enhancer related histone mark H3K4me1 and showed enhancer activity in gene reporter assays (Xie, Hong et al. 2013). The fact that many TEs are able to act as enhancers opens up an additional level of gene regulation which may be occurring as a result of LTR activation in HL.

4.4. Long Terminal Repeats can be activated by inflammatory signalling

Our data has shown a clear genome-wide activation pattern of the THE1B family of LTRs in HL and additionally cell line specific LTR activation. Transcribed LTRs pose a significant threat to regulation of gene expression and therefore can contribute to the survival and growth of malignant cells. This means that strict control mechanisms are be required to ensure that ERVs are not expressed and these control mechanisms are lost in HL and other diseases which display LTR activation. The expression of subsets of LTRs in each HL cell line also implies that the expression of different members of the THE1B family may be regulated by different mechanisms.

Loss of DNA methylation was the first genome-wide epigenetic change to be reported in cancer and overall genomic hypomethylation has since been reported in many types of cancer (Gama-Sosa, Slagel et al. 1983, Ehrlich 2009). A large proportion of DNA hypomethylation in cancer is targeted towards repeat elements including Alu, LINE1, Satellite and LTRs (Qu, Dubeau et al. 1999, Florl, Steinhoff et al. 2004, Rodriguez, Vives et al. 2008, Benesova, Trejbalova et al. 2017). The THE1B LTR upstream of *CSF1R* was shown to be methylated at 2 CpG elements in mononuclear cells from healthy donors and this methylation is lost in HL (Lamprecht, Walter et al. 2010). To further investigate the epigenetic mechanisms controlling the expression of the *CSF1R* LTR Lamprecht, et al. (2010) screened for the expression of a number of corepressors, histone deacetylases and DNA methyltransferases in HL. No changes in the expression of any of these proteins were shown, however, a loss of the expression of the transcriptional repressor CBFA2T3 (ETO2) was discovered. ETO2 acts as a repressor through interaction with HDACs, a number of corepressors and possibly also EZH2 which is a component of the polycomb repressor complex and H3K27 methyltransferase complex (Hug and Lazar 2004, Wael, Fujiwara et al. 2011). It was further demonstrated that loss of ETO2 expression in Reh and Namalwa cell lines could induce weak expression of the CSF1R LTR and when this was combined with constitutive NF-κB activation a strong expression could be observed (Lamprecht, Walter et al. 2010).

Knowing that constitutive NF-κB activation is a part of the inflammatory signalling network in HL and that THE1B LTRs harbour NF-κB, AP-1 and IRF motifs we wanted to establish whether inflammatory signalling alone could induce LTR expression. To achieve this we treated the cells with PMA which activates NF-κB and a number of other inflammatory pathways through PKC signalling. We showed that PMA treatment of Reh cells was able to induce the expression of the CSF1R LTR and a number of other LTRs. The PMA treatment recapitulated the previous findings as it not only resulted in NF-κB activation but also the down-regulation of ETO2 by an

unknown mechanism. Our data also showed a correlation of LTR expression with changes in the gene expression and chromatin accessibility patterns indicating a move towards the HL phenotype. This finding demonstrated that inflammatory signalling alone through PKC was able to induce the expression of many LTRs and a partial HL phenotype.

The presence of inflammatory transcription factor motifs has been previously noted in ERVs of the HERV-K and THE1B families (Lamprecht, Walter et al. 2010, Manghera and Douville 2013). Further to this HERV-K activation has been shown in a number of inflammatory diseases including rheumatoid arthritis, schizophrenia and several types of cancer and THE1B activation in HL (Frank, Giehl et al. 2005, Ruprecht, Mayer et al. 2008, Freimanis, Hooley et al. 2010, Lamprecht, Walter et al. 2010). The induction of ERV expression by treatment with PMA has also been shown previously, however not on a genome-wide scale. Induction of monocytic differentiation of the U-937 cell line by PMA treatment results in expression of an ERV3 element of the HERV-R family which is usually expressed in placental tissue (Larsson, Venables et al. 1996). An increase in expression from an already active LTR can also be induced by PMA treatment, as described in a study of a HERV-K ERV in MeWo melanoma cells (Katoh, Mirova et al. 2011). Additionally several studies have shown that the activity of LTRs driving HIV infections can also be induced by PMA treatment (West, Lowe et al. 2001).

The move we observed towards the HL phenotype following PMA treatment of Reh cells is not likely to be solely driven by LTR activation. Through interaction with PKC, PMA induces many pathways which are up-regulated in HL including NF-κB, MAPK, JNK and AP-1 (Schultz, Engel et al. 1997, Sharma and Richards 2000, Bagowski, Besser et al. 2003). Due to the interaction between these pathways and the transcription factor activity of AP-1 and NF-κB PMA has the potential to induce wide spread gene expression changes which are additional to those resulting from LTR expression.

To study the impact of LTR expression in a more tightly regulated system we produced an inducible NF-κB activation system through the overexpression of IkKβ. The induction of NF-κB activity resulted in the expression of a large number of LTRs shown in our RACE-Seq data. However, only a small proportion of these LTRs were also expressed in the HL cell lines and the *CSF1R* LTR was not expressed. This finding suggests that the transcriptional repression of at least the *CSF1R* LTR by ETO2 is sufficient to prevent expression even in the presence of constitutive NF-κB activation. Very little overall change in gene expression was observed, however, we did see the up-regulation of a number of inflammatory response genes. The induction of inflammatory gene expression is not surprising as constitutive NF-κB activation has long been known to play a significant role in the inflammatory signature of HL cells (Izban, Ergin et al. 2001).

The LTR expression which we observe both following PMA treatment and induction of constitutive NF-κB activation only displayed a partial overlap with the LTR expression in HL cells. This finding implies that additional epigenetic mechanisms for transcriptional control are inhibiting the activation of some LTRs including those responsible for the expression of *CSF1R*, *WNT5A*, *NLRP1* and *TNFRSF11A*. We already know from the work of Lamprecht *et al.* (2010) that the transcription of the *CSF1R* LTR is repressed by ETO2, however, this does not account for many other LTRs which are not expressed following PMA treatment. A number of other mechanisms for controlling the expression of ERVs have been shown in other studies, including histone acetylation, histone methylation and KRAB-ZFP binding (Hurst and Magiorkinis 2017). However a recent study has also suggested that, in the case of the L1HS-Ta subfamily of LINE1 elements, activation only occurs when the element is in a region of chromatin which is already primed for expression (Philippe, Vargas-Landin et al. 2016).

Acetylation of lysine residues in histones has been linked to the activation of HERV-Fc1 when combined with a loss of CpG methylation (Laska, Brudek et al. 2012). Acetylation of lysine

residues by histone acetyltransferases (HATs) blocks positive charges and therefore destabilises nucleosome binding allowing for transcription to take place. To maintain stable chromatin at these regions histone deacetylases (HDACs) remove acetylation from the lysine residues (Bannister and Kouzarides 2011). In a disease state the loss of HDACs could potentially lead to chromatin destabilisation and the activation of ERVs, however, it has been shown that inhibition of HDACs alone is not sufficient to upregulate ERV expression (Laska, Brudek et al. 2012). Our gene expression data does not show any significant change in the levels of HDAC proteins compared to the control cell lines with the exception of HDAC9. HDAC9 overexpression has been linked to other B-cell malignancies but there is a lack of research into the impact of HDAC9 down-regulation (Gil, Bhagat et al. 2016).

Histone methylation also has potential to play a role in the epigenetic regulation of ERV expression. We show the presence of the activating lysine methylation mark H3K4me3 at a proportion of active LTRs in HL, however, we did not study potential repressive marks residing at inactive LTRs in both the control and HL cell lines. A strong association has previously been shown between repressed HERV-K elements and the repressive histone mark H3K9me3 (Campos-Sanchez, Cremona et al. 2016). H3K9me3 at ERVs is driven largely by Krüppel associated box zinc finger proteins (KRAB-ZFP) which exhibit sequence specific DNA binding and interact with KAP1 (TRIM28) to recruit histone modification complexes. In particular KARB-ZFPs have been shown to recruit SETDB1 to ERVs resulting in the repressive H3K9me3 modification (Imbeault, Helleboid et al. 2017). There is also evidence to suggest that the wide array of KRAB-ZFP proteins within the human genome developed as a result of new ERV integrations (Thomas and Schneider 2011). Although the action of KRAB-ZFPs is yet to be investigated at MalR LTRs, this finding highlights an interesting potential regulatory mechanism as the binding of different KRAB-ZFPs could account for the different activation patterns which we observed in each cell line.

The final factor which is known to play a role in control of ERV expression and may be contributing to the LTR expression patterns in HL is DNA methylation. As discussed previously (1.3.17) CG-hypomethylation is a major phenotype of many cancers and may contribute to increased LTR activation. A recent study has highlighted this notion by showing that the treatment of lung cancer cells with histone deacetylase inhibitors (HDACi) and DNA methyltransferase inhibitors (DNMTi) resulted in the genome wide activation of LTRs, particularly of the ERV9 family (Brocks, Schmidt et al. 2017). A number of other studies have also shown that although global hypomethylation occurs in cancer that some tumour suppressor genes are hypermethylated which suggests a more complex role for DNMTs (Mager and Lorincz 2017). We know from previous studies (discussed above) that a combination of activation and loss of repression is required for activation of other LTR families and these findings with regard to DNA methylation add an additional level of complexity.

4.5. Conclusion

In conclusion, we have developed a new genome-wide technique for the identification of active THE1B LTRs and used this technique to investigate the activation of LTRs in HL. We have shown that HL cell lines exhibit a specific activation pattern of LTRs both in comparison to control cell lines and to each other. Moreover, the expressed LTRs have also been shown to play a role in the up-regulation of potentially pathological gene expression.

The inflammatory signature and the involvement of NF-κB and AP-1 in HL had been previously described, however for the first time we have shown co-localisation of AP-1 and NF-κB motifs with many other transcription factors motifs in HL suggesting that these factors may co-localise. This demonstrates the central role of AP-1 and NF-κB in contributing to the deregulated HL gene expression program. We have also shown that the inflammatory signature of HL may be

the driving force in the genome-wide activation of LTRs, which highlights the potential involvement of LTR expression in other inflammatory disease.

These exciting new insights into the activation and function of LTRs within the genome of HL cells highlight the significance of LTR activation and the role that it may play in the pathogenesis of HL. We have also shown LTR activation as a significant gene regulatory mechanism which may have implications for the pathogenesis of other diseases.

4.6. Future Work

4.6.1. Further examination of the cis-regulatory elements driving HL

We have shown that NF-κB and AP1 motifs co-localise in regions of chromatin which show transcription factor binding, based on our footprinting data. This finding implies that there may be an interaction between the NF-κB and AP-1 transcription factors. To further study this idea we could obtain ChIP-Seq data for NF-κB and AP-1 to determine their actual binding pattern rather than just that inferred from the presence of their motifs. The co-localisation of these factors could also be studied using immune florescence and confocal microscopy. We also now have an IRF4 ChIP-Seq dataset which can be used to analyse the binding at the IRF motifs which we also identified in foot-printed regions.

4.6.2. Genome-wide mapping of LTRs

We have successfully demonstrated the utility of our THE1B LTR RACE-Seq technique in the identification of expressed LTRs in both HL and control cell lines. Although successful our RACE-Seq technique currently has a significant variability between biological replicates. We hypothesised that this was due to the low sequencing depth capturing only a proportion of the RACE-Seq library in each run. To overcome this we will re-sequence the libraries at higher read-depth.

We believe that the RACE-Seq technique should have significant advantage over CAGE for the identification of expressed LTRs due to the longer read length producing more accurate unique mapping. To determine whether there is a significant advantage to the RACE-Seq technique it would be necessary to perform CAGE on the HL cell lines. The RACE-Seq data could also be compared to RNA-Seq bioinformatics based approaches for the identification of active LTRs (discussed in 1.3.18), however, it is expected that the lower abundance LTR transcripts would be missed in the RNA-Seq data.

Currently the RACE-Seq technique has only been targeted towards the THE1B LTR family and through sequence homology we have shown that other members of the THE1 family are active and can influence gene expression. To further investigate the other active ERVs in HL an array of primers could be designed allowing a far broader spectrum of expressed ERVs to be studied.

The RACE-Seq technique could also prove valuable for the genome-wide identification of expressed LTRs in other diseases. LTR activation has been reported in many diseases, particularly those with an inflammatory aspect and it would be exciting to build up datasets for the LTR activation in these diseases to both analyse the impact on gene expression and to compare activation in different disease contexts. RACE-Seq could also be used to identify cells with chronic inflammation based on the identification of chronic LTR activation.

4.6.3. LTRs as Alternative Promoters and Enhancers

We successfully identified a number of protein coding genes which exhibit expression from an upstream expressed LTR promoter in HL. The WNT5A gene has shown involvement in the HL phenotype in a number of studies and we have proposed a number of other regulatory mechanisms which may form a positive feedback loop for the regulation of gene expression from the LTR. It may be of interest to further investigate the role of this gene in the overall HL

phenotype by performing siRNA knockdown. The NLRP1 gene which we showed to be up-regulated from an LTR has not before been associated with HL before and may prove to be an interesting target particularly if it is functioning through a novel pathway. Again in the first instance an siRNA knockdown of NLRP1 would be require to assess the influence of cell phenotype and survival.

We also identified many unannotated lncRNAs originating from expressed LTRs. The function of lncRNAs is more complicated to determine than protein coding genes as they are most likely to act as regulatory elements. Other studies have shown a significant impact on the survival and proliferation of cells resulting from the knockdown of LTR driven lncRNA transcripts. To study this would require a large scale screening method, a number of which have been developed in recent years. A particularly successful method has been the use of CRISPR interference (CRISPRi) which recruits a non-functional Cas9 enzyme to the transcription start site of lncRNAs preventing their transcription (Liu, Horlbeck et al. 2017). The efficacy of this technique when targeting lncRNAs with LTR transcription start sites would have to be assessed as the repeated nature of the sequence through the genome would make them more difficult to uniquely target.

Many active LTRs did not display any associated transcript in our RNA-Seq data. We have suggested that these LTRs may be acting as enhancer for promoters of coding genes which are topologically associated. The enhancer function of active LTRs could play an important role in HL and potentially other disease. The function of LTRs as enhancers could be investigated by 3 routes. Firstly performing genome-wide ChIP-Seq for histone modifications often associated with enhancer elements such as H3K4me1 (Calo and Wysocka 2013). Secondly promoter capture-Hi-C technology could be used to identify the topological association of active LTRs with promoters (Mifsud, Tavares-Cadete et al. 2015). Finally there are a number of luciferase reported based assay which have been developed for enhancer screening and luciferase

assays would be required to validate the function of any enhancers identified by the other methods (Dailey 2015).

Both the RACE-Seq and CAGE techniques suffer the significant problem of identification of down-stream transcripts originating from active LTRs because both techniques only identify the transcription start sites. We have overcome this to some degree by the manual screening of read-through transcripts in RNA-Seq data followed by the use of qPCR to confirm the presence of a transcript originating from an expressed LTR. This technique is not ideal as it is not possible to rapidly identify entire LTR driven transcripts. A potential way to overcome this would be the development of a capture RNA-Seq technique whereby entire transcripts containing an LTR could be captured and sequenced on a genome-wide basis. To achieve this biotinylated baits targeted towards the homologous LTR sequences could be designed and used to pull down RNA containing an LTR. For the THE1B elements this may be possible with a single bait and an annealing step allowing binding with mismatches. The selected RNA could then be used in an RNA-Seq library preparation protocol which would produce a genome-wide library of transcripts which incorporate an LTR. This technique could have a number of advantages including the opportunity to create a capture array for many families of ERV and potentially a simplified mapping strategy as alignment to the LTR its self may not be necessary. This capture technique combined with either RACE-Seq or CAGE to identify transcription start sites could prove to be a valuable and rapid technique to determine the true impact of LTRs as alternative promoters in HL and other healthy and diseased tissue.

4.6.4. Direct Impact of LTR activation

In the current study we began to assess the potential direct impact of LTR activation in the Reh cell line by the constitutive activation of NF-κB. We showed the activation of a large number of LTRs, however, only a small proportion of these were shared within the HL datasets. Based on

previous work it is clear that for at least the expression of the *CSF1R* LTR the knockdown of ETO2 is also required (Lamprecht, Walter et al. 2010). To establish the direct impact that genome-wide LTR activation has on gene expression we have already begun to carry out siRNA knockdown of ETO2 in our iIkKβ cells with the aim to carry out RNA-Seq and RACE-Seq, however time constraints prohibited it's inclusion in this thesis. This will allow us to both establish the genome-wide LTR activation pattern generated by combined ETO2 knockdown and constitutive NF-κB activation and the impact on gene expression at an early time point prior to downstream changes in the cell phenotype. Hopefully this will provide us with exciting insights into how LTR expression can deregulate gene expression.

4.6.5. Epigenetic control mechanisms for LTR repression

Although we know from the work of Lamprecht et al. (2010) that loss of ETO2 expression is involved in the up-regulation of LTRs the way in which ETO2 interacts with the LTR sequence to inhibit transcription is not known. It is known that ETO2 does not directly bind DNA and functions as part of a complex. To determine the way in which ETO2 contributes to LTR activation in HL ChIP-Seq could be performed to identify the ETO2 complex binding site. It is known, however, from previous work in our lab that ETO2 is notoriously difficult to ChIP (Bonifer, personal communication). Therefore, to overcome this it may be necessary to produce a tagged version of the ETO2 protein which could be expressed in Reh cells and the ChIPed based on the tag. This method is not ideal as an overexpression of ETO2 may result in aberrant binding, however, it may elucidate the binding motif of the complex in the vicinity of repressed LTRs. It may also be possible to carry out pull-down mass-spectrometry assays on the ETO2 repression complex at the repressed LTRs to determine the interacting partners which bind the DNA.

It is clear from the unique activation pattern of THE1B LTRs observed in each cell line that different epigenetic repression mechanisms act on different LTRs. In the first instance assessing the genome-wide DNA methylation state at repressed LTRs in the HL cell lines would be beneficial to determine whether this is the main controlling factor of LTR expression even in the generally hypomethylated state of HL cells. Other ERVs have been shown to harbour the H3K9me3 repressive chromatin mark which is likely to be deposited by SETDB1 as a result of specific KRAB-ZFP binding. To investigate whether H3K9me3 plays a role in repression of the THE1B LTRs ChIP-Seq could be carried out in a control cell line to look for its presence in the vicinity of inactive TH1B LTRs. The role of KRAB-ZFP proteins could also be investigated through ChIP-Seq of KAP1 the binding partner of many of the KRAB-ZFP proteins.

4.6.6. Primary Cells

Finally, it is important to note that all assays within the current study were carried out in HL and control cell lines which, although well characterised, are not a completely accurate representation of primary HL cells. To determine the relevance of our findings in the clinical context it is important to repeat aspects of our work in primary samples. This is a challenging prospect as obtaining primary HL cells is a notoriously arduous task due to their low percentage within tumour samples. It may be possible to optimise the RACE-Seq technique for very low cell numbers, as has been done for many genome-wide sequencing techniques in recent years. A simpler approach may be to use RNA-Seq data obtained from primary HL cells and carry out a targeted bioinformatics approach to identify active LTRs. Although this approach may not be able to identify the full complement of LTR expression shown by RACE-Seq, it may prove useful in validating the activity of particular LTRs in primary tissue.

5. SUPPLEMENTARY TABLES

Supplementary Data Table 1. Closest genes to active LTRs detected by RACE-Seq in **Reh** cells. Gene expression values (FPKM) from RNA-Seq are also shown for each gene.

RACE Peak	Gene (FPKM)
chr1 2123459-2123568	PRKCZ (0.37)
chr1 11467916-11468065	PTCHD2 (0.33)
chr1 11723228-11723324	FBXO6 (1.99)
chr1 31945309-31945449	LOC149086 (0)
chr1 38959152-38959284	LOC339442 (0)
chr1 57080279-57080415	PRKAA2 (0)
chr1 57271439-57271571	C1orf168 (0)
chr1 58185274-58185370	DAB1 (0.99)
chr1 60251163-60251289	FGGY (2.76)
chr1 72181715-72181857	NEGR1 (3.55)
chr1 79645542-79645678	ELTD1 (0)
chr1 83498791-83498918	TTLL7 (0)
chr1 91328368-91328538	ZNF644 (5.45)
chr1 101455533-101455850	DPH5 (4.19)
chr1 105955070-105955205	LOC100129138 (0)
chr1 108092966-108093310	VAV3 (6.39)
chr1 108436106-108436235	VAV3 (6.39)
chr1 112218475-112218571	RAP1A (2.52)
chr1 115735413-115735555	NGF (0)
chr1 116082660-116082801	VANGL1 (0)
chr1 150280559-150280660	PRPF3 (5.83)
chr1 153633929-153634046	ILF2 (8.32)
chr1 157866049-157866184	CD5L (0.77)
chr1 157867974-157868110	CD5L (0.77)
chr1 177701417-177701605	SEC16B (0)
chr1 193832613-193832748	CDC73 (2.7)
chr1 194317005-194317148	CDC73 (2.7)
chr1 194699318-194699446	CDC73 (2.7)
chr1 194780246-194780388	KCNT2 (0)
chr1 195611798-195611930	KCNT2 (0)
chr1 198999693-198999916	LOC100131234 (0)
chr1 209874557-209874702	HSD11B1 (0)
chr1 211640084-211640229	RD3 (0)
chr1 215159433-215159559	KCNK2 (0)
chr1 221760504-221760637	DUSP10 (0.98)
chr1 234730404-234730527	IRF2BP2 (6.06)
chr1 236326933-236327067	GPR137B (2.65)
chr1 245120978-245121151	EFCAB2 (4.25)
chr10 1496247-1496385	ADARB2-AS1 (0.56)
chr10 1801338-1801475	ADARB2 (0.06)
chr10 6278973-6279103	PFKFB3 (2.57)
chr10 13478502-13478631	BEND7 (0.13)
chr10 15080561-15080692	OLAH (0.52)
chr10 23834957-23835098	OTUD1 (1.6)
chr10 24139694-24139822	KIAA1217 (0.13)
chr10 26717788-26717915	APBB1IP (5.68)
chr10 30536793-30536927	MTPAP (3.58)
chr10 44180164-44180299	ZNF32-AS3 (0)
chr10 50791274-50791402	CHAT (0.12)
chr10 52555746-52555887	A1CF (0)
chr10 53022429-53022552	PRKG1 (0.01)
chr10 67021523-67021655	ANXA2P3 (0)
chr10 78271865-78272006	KCNMA1 (0.12)
chr10 79664370-79664512	DLG5 (2.2)
chr10 81114899-81115016	ZCCHC24 (0.93)
chr10 82388098-82388227	SH2D4B (1.65)
chr10 92353932-92354066	HTR7 (2.72)
chr10 92426108-92426228	HTR7 (2.72)
chr10 93656807-93656948	FGFBP3 (1.08)
chr10 95441248-95441350	PDE6C (0.03)
chr10 98497237-98497398	PIK3AP1 (7.24)
chr10 101711795-101711966	DNMBP (2.13)
chr10 108053831-108053956	SORCS1 (0)
chr10 110936974-110937109	XPNPEP1 (5.6)
chr10 115780559-115780702	ADRB1 (0.57)
chr10 125171640-125171762	BUB3 (8.8)
chr10 125305960-125306102	GPR26 (0)
chr11 1285714-1286156	MUC5B (0)
chr11 3206211-3206351	OSBPL5 (2.69)
chr11 10830368-10830491	EIF4G2 (8.04)
chr11 11110831-11110974	GALNTL4 (0)
chr11 11845428-11845570	USP47 (4.19)
chr11 13010984-13011140	RASSF10 (0.43)
chr11 14541835-14541970	PSMA1 (7.59)
chr11 23658261-23658386	LOC100500938 (0)
chr11 27911881-27912109	KIF18A (4.63)
chr11 35071547-35071706	PDHX (4.39)
chr11 36602021-36602155	RAG1 (8.75)
chr11 36607823-36608045	RAG2 (7.37)
chr11 46450246-46450424	AMBRA1 (4.61)
chr11 57357410-57357530	SERPING1 (3.46)
chr11 63595036-63595151	MARK2 (3.04)
chr11 64185765-64185899	RPS6KA4 (4.59)
chr11 67450457-67450559	ALDH3B2 (0)
chr11 73685720-73685942	DNAJB13 (0.13)
chr11 89984349-89984473	CHORDC1 (4.11)
chr11 98765707-98765834	CNTN5 (0)
chr11 103987549-103987690	PDGFD (0)
chr11 109031422-109031563	DDX10 (4.54)
chr11 111776418-111776533	CRYAB (0.1)
chr11 113749276-113749609	USP28 (4.66)
chr11 119773496-119773635	PVRL1 (0.67)
chr11 126517995-126518114	KIRREL3 (0.05)

chr11 127792055-127792177	ETS1 (4.32)	**chr14 35779976-35780108**	KIAA0391 (5.58)
chr12 6576145-6576288	VAMP1 (2.98)	**chr14 38973137-38973271**	LOC283547 (0)
chr12 9804052-9804177	CLEC2D (3.8)	**chr14 39011191-39011317**	LOC283547 (0)
chr12 13420857-13420978	EMP1 (0.17)	**chr14 41208380-41208513**	LRFN5 (0)
chr12 21810677-21810807	LDHB (10.18)	**chr14 50872977-50873211**	CDKL1 (1.96)
chr12 25108470-25108603	BCAT1 (0.01)	**chr14 52731145-52731291**	PTGDR (2.44)
chr12 26087145-26087268	LOC100506451 (0)	**chr14 70753641-70753782**	SYNJ2BP-COX16
chr12 38625342-38625442	ALG10B (1.4)	**chr14 81675607-81675744**	GTF2A1 (3.2)
chr12 46380565-46380699	ARID2 (4.2)	**chr14 81723116-81723231**	STON2 (0.7)
chr12 46544822-46544947	SLC38A1 (4.68)	**chr14 82889319-82889454**	SEL1L (3.87)
chr12 50505950-50506075	C12orf62 (0)	**chr14 85271390-85271518**	FLRT2 (0)
chr12 56506765-56506876	RPL41 (12.23)	**chr14 88746234-88746375**	KCNK10 (0.31)
chr12 59880360-59880489	SLC16A7 (0.08)	**chr14 92630270-92630521**	CPSF2 (5.83)
chr12 78714442-78714679	NAV3 (1.55)	**chr14 93336859-93336998**	GOLGA5 (3.59)
chr12 84255504-84255641	TMTC2 (0)	**chr14 93338642-93338775**	GOLGA5 (3.59)
chr12 88535708-88535816	CEP290 (2.64)	**chr14 93339790-93339929**	GOLGA5 (3.59)
chr12 89445744-89445875	LOC728084 (0.22)	**chr14 96245201-96245325**	TCL1A (0.18)
chr12 93335088-93335239	EEA1 (2.82)	**chr14 97375237-97375365**	VRK1 (6.55)
chr12 94027778-94027968	CRADD (3.44)	**chr14 97903762-97903895**	LOC100129345 (0)
chr12 97477865-97477991	NEDD1 (4.86)	**chr14 100800165-100800276**	SLC25A47 (0)
chr12 104864046-104864177	CHST11 (4.28)	**chr15 33502030-33502169**	RYR3 (0.08)
chr12 111298758-111298887	MYL2 (0)	**chr15 34880538-34880676**	GOLGA8B (2.41)
chr12 111848608-111848767	ATXN2 (2.64)	**chr15 38318073-38318214**	TMCO5A (0)
chr12 112847270-112847376	PTPN11 (5.34)	**chr15 38974421-38974556**	C15orf53 (0)
chr12 115803801-115803938	MED13L (4.99)	**chr15 43490121-43490263**	EPB42 (0.03)
chr12 117561629-117561763	FBXO21 (4.55)	**chr15 46166854-46166984**	SQRDL (1.02)
chr12 125597277-125597408	AACS (1.82)	**chr15 49907716-49907835**	C15orf33 (0)
chr12 126814995-126815112	LOC100128554 (0)	**chr15 53315164-53315305**	ONECUT1 (0.26)
chr12 129594177-129594319	TMEM132D (0)	**chr15 62574055-62574200**	C2CD4B (0)
chr13 21751159-21751291	MRP63 (0)	**chr15 69745120-69745258**	KIF23 (5.19)
chr13 21776034-21776159	MRP63 (0)	**chr15 70885582-70885712**	UACA (1.96)
chr13 27766731-27766873	USP12 (2.75)	**chr15 74756744-74756874**	LOC440288 (0)
chr13 28841731-28841873	FLT1 (2.45)	**chr15 75230242-75230468**	RPP25 (1.37)
chr13 31037342-31037464	LINC00426 (4.92)	**chr15 78953409-78953517**	CHRNB4 (0.21)
chr13 31122395-31122516	USPL1 (3.52)	**chr15 96603966-96604108**	NR2F2 (0)
chr13 38410175-38410353	TRPC4 (2.49)	**chr15 101835319-101835461**	PCSK6 (0.3)
chr13 38558527-38558685	TRPC4 (2.49)	**chr16 2011183-2011279**	RPS2 (11.27)
chr13 38624776-38625015	TRPC4 (2.49)	**chr16 2011606-2011931**	RPS2 (11.27)
chr13 54539416-54539551	MIR1297 (0)	**chr16 11931461-11931561**	RSL1D1 (6.04)
chr13 56357020-56357162	PRR20C (0)	**chr16 13515180-13515321**	SHISA9 (2.83)
chr13 59562742-59562860	DIAPH3 (4.07)	**chr16 20035576-20035705**	GPR139 (0)
chr13 60603706-60603831	DIAPH3 (4.07)	**chr16 21127729-21127888**	DNAH3 (0.08)
chr13 60982103-60982346	TDRD3 (2.96)	**chr16 27258614-27258734**	NSMCE1 (4.28)
chr13 63631316-63631432	OR7E156P (0)	**chr16 31123555-31123837**	KAT8 (5.29)
chr13 68917837-68917972	LOC338862 (0)	**chr16 54907974-54908106**	CRNDE (4.88)
chr13 89571375-89571608	SLITRK5 (1.18)	**chr16 60118530-60118661**	LOC644649 (0)
chr13 89681350-89681456	MIR622 (0)	**chr16 67978238-67978366**	SLC12A4 (2.43)
chr13 104846787-104846928	SLC10A2 (0)	**chr16 69355949-69356117**	COG8 (4.43)
chr13 109306691-109306832	MYO16 (0.01)	**chr16 76268947-76269137**	CNTNAP4 (0)
chr13 110354394-110354508	IRS2 (1.98)	**chr16 80926391-80926522**	C16orf61 (0)
chr13 110617233-110617368	IRS2 (1.98)	**chr16 88876087-88876225**	APRT (6.96)
chr13 112262578-112262701	C13orf16 (0)	**chr16 89727676-89727892**	C16orf55 (0)
chr13 114518538-114518683	FAM70B (0)	**chr16 90029710-90029826**	DEF8 (4.88)
chr14 21134849-21134992	ANG (0.89)	**chr17 1554038-1554161**	PRPF8 (5.54)
chr14 24037062-24037185	JPH4 (0.07)	**chr17 4843930-4844048**	RNF167 (4.88)
chr14 30348573-30348708	PRKD1 (0)	**chr17 7529654-7529770**	SHBG (0)

chr17 7760777-7760896	CYB5D1 (2.95)	**chr19 50363811-50363976**	PNKP (3.57)
chr17 10770809-10770941	PIRT (0)	**chr19 56824676-56824772**	ZNF542 (0)
chr17 27717442-27717571	MIR4523 (0)	**chr2 11029080-11029221**	KCNF1 (0)
chr17 30771505-30771629	PSMD11 (3.32)	**chr2 11720088-11720329**	MIR4429 (0)
chr17 33296817-33296948	ZNF830 (3.9)	**chr2 11821533-11821673**	NTSR2 (0)
chr17 35075637-35075770	MRM1 (3.56)	**chr2 14131243-14131387**	FAM84A (0)
chr17 41395560-41395722	TMEM106A (2.42)	**chr2 14375957-14376079**	FAM84A (0)
chr17 46752964-46753088	MIR196A1 (0)	**chr2 22065522-22065647**	LOC645949 (0)
chr17 65700583-65700727	PITPNC1 (3.5)	**chr2 22445716-22445845**	LOC645949 (0)
chr17 67559490-67559625	MAP2K6 (2.98)	**chr2 27907978-27908097**	SLC4A1AP (4.24)
chr17 67744804-67744929	MAP2K6 (2.98)	**chr2 30490026-30490154**	LBH (3.92)
chr17 69617064-69617194	SOX9 (0)	**chr2 33931893-33932037**	MYADML (0)
chr17 74560556-74560724	ST6GALNAC2 (0.81)	**chr2 36275754-36275883**	LOC100288911 (0)
chr17 77158335-77158486	RBFOX3 (0.13)	**chr2 40144672-40144804**	LOC100128590 (0)
chr17 79336000-79336133	TMEM105 (0)	**chr2 40808066-40808198**	SLC8A1 (3.07)
chr18 5270335-5270462	ZFP161 (0)	**chr2 54501233-54501370**	ACYP2 (1.76)
chr18 8445098-8445240	PTPRM (2.3)	**chr2 54621620-54621758**	C2orf73 (0.04)
chr18 8938989-8939113	CCDC165 (0)	**chr2 60343686-60343828**	MIR4432 (0)
chr18 13149934-13150054	CEP192 (4.77)	**chr2 65275253-65275376**	CEP68 (3.33)
chr18 20857268-20857393	CABLES1 (0.42)	**chr2 70645593-70645723**	TGFA (0.26)
chr18 29545587-29545718	TRAPPC8 (4.56)	**chr2 71349617-71350063**	MPHOSPH10 (3.6)
chr18 32485040-32485192	DTNA (0.02)	**chr2 74154009-74154157**	DGUOK (5.54)
chr18 36807047-36807147	LOC647946 (0)	**chr2 76532724-76532875**	LRRTM4 (0)
chr18 39015049-39015187	KC6 (0)	**chr2 89375689-89375824**	MIR4436A (0)
chr18 39580615-39580752	PIK3C3 (6.35)	**chr2 90163095-90163228**	MIR4436A (0)
chr18 44722435-44722566	IER3IP1 (5.25)	**chr2 100544450-100544579**	AFF3 (0.77)
chr18 44812047-44812173	IER3IP1 (5.25)	**chr2 111881294-111881411**	ACOXL (0)
chr18 48043839-48043965	MAPK4 (0.02)	**chr2 136007260-136007395**	ZRANB3 (2.66)
chr18 53264253-53264382	TCF4 (5.34)	**chr2 136743591-136743689**	DARS (6.72)
chr18 57212996-57213137	CCBE1 (0.05)	**chr2 137033983-137034097**	CXCR4 (7.16)
chr18 61222174-61222295	SERPINB12 (0.19)	**chr2 155062114-155062249**	GALNT13 (0)
chr18 62453073-62453198	LOC284294 (0)	**chr2 162419058-162419181**	SLC4A10 (0.01)
chr18 62658709-62658848	LOC284294 (0)	**chr2 177293855-177293995**	MTX2 (0)
chr18 71197398-71197532	LOC100505817 (0)	**chr2 188616708-188616845**	TFPI (1.52)
chr18 71452132-71452262	FBXO15 (0.76)	**chr2 191573349-191573480**	NAB1 (3.76)
chr18 71825364-71825498	C18orf55 (0)	**chr2 196344281-196344407**	SLC39A10 (5.44)
chr18 72057488-72057601	C18orf63 (0.09)	**chr2 211777511-211777654**	CPS1 (0.64)
chr18 73341580-73341717	C18orf62 (0)	**chr2 217363562-217363662**	SMARCAL1 (2.95)
chr18 73447887-73448014	C18orf62 (0)	**chr2 217363946-217364044**	SMARCAL1 (2.95)
chr18 75705583-75705711	GALR1 (0)	**chr2 222529817-222529946**	EPHA4 (0.07)
chr18 77352749-77352873	NFATC1 (1.19)	**chr2 224887630-224887764**	SERPINE2 (0.45)
chr19 2328397-2328580	SPPL2B (3.52)	**chr2 225291544-225291667**	FAM124B (0)
chr19 2477769-2477891	GADD45B (5)	**chr2 239683152-239683293**	TWIST2 (0)
chr19 3178371-3178504	S1PR4 (6.54)	**chr2 242670236-242670362**	ING5 (3.84)
chr19 5676822-5676933	C19orf70 (4.75)	**chr20 807139-807267**	FAM110A (2.73)
chr19 5978382-5978548	RANBP3 (4.51)	**chr20 4557233-4557365**	PRNP (3.66)
chr19 6381159-6381428	GTF2F1 (5.27)	**chr20 12638321-12638455**	SPTLC3 (0.9)
chr19 7320952-7321093	INSR (5.47)	**chr20 18109287-18109434**	PET117 (3.72)
chr19 7696336-7696479	PCP2 (0.71)	**chr20 22940392-22940534**	SSTR4 (0)
chr19 11488541-11488674	EPOR (3.06)	**chr20 35617257-35617387**	RBL1 (4.85)
chr19 34046695-34046823	PEPD (4.61)	**chr20 39570752-39570878**	TOP1 (4.68)
chr19 41302697-41302839	MIA-RAB4B (0.45)	**chr20 40256511-40256776**	CHD6 (3.69)
chr19 42891510-42891637	MEGF8 (2.5)	**chr20 43095404-43095524**	TTPAL (1.46)
chr19 47029653-47029793	LOC100506012 (0)	**chr20 43588896-43589039**	LOC100505826 (0)
chr19 47776906-47777022	PRR24 (0)	**chr20 55453491-55453624**	TFAP2C (0)
chr19 50363565-50363667	PNKP (3.57)	**chr20 55813770-55813907**	BMP7 (0)

chr20 56058374-56058650	CTCFL (0.95)	**chr3 109282195-109282321**	FLJ25363 (0)
chr20 60715845-60715987	SS18L1 (2.95)	**chr3 112151745-112151879**	BTLA (0.03)
chr21 16052027-16052161	SAMSN1 (4.49)	**chr3 113806447-113806669**	KIAA1407 (0.98)
chr21 16061745-16061879	SAMSN1 (4.49)	**chr3 116911876-116912011**	LSAMP-AS3 (0)
chr21 16891388-16891516	USP25 (4.66)	**chr3 120026814-120026935**	LRRC58 (3.29)
chr21 17005818-17005943	USP25 (4.66)	**chr3 120315155-120315293**	HGD (0.24)
chr21 17248781-17248902	LINC00478 (0)	**chr3 121753591-121753726**	ILDR1 (0)
chr21 19858103-19858241	TMPRSS15 (0)	**chr3 122393307-122393424**	PARP14 (5.73)
chr21 20834127-20834253	TMPRSS15 (0)	**chr3 129037052-129037187**	H1FX-AS1 (2.5)
chr21 23406648-23406783	LINC00308 (0)	**chr3 129037336-129037432**	H1FX-AS1 (2.5)
chr21 24043565-24043701	LINC00308 (0)	**chr3 129143253-129143387**	MBD4 (4.68)
chr21 28195424-28195567	ADAMTS1 (0)	**chr3 131606067-131606199**	CPNE4 (0.05)
chr21 29420731-29420864	LINC00314 (0)	**chr3 135882660-135882776**	PPP2R3A (0.97)
chr21 29570152-29570294	LINC00314 (0)	**chr3 140480162-140480296**	TRIM42 (0)
chr21 35287996-35288136	LOC100506334 (0)	**chr3 150321084-150321242**	SELT (5.16)
chr21 37177004-37177138	MIR802 (0)	**chr3 151269914-151270050**	MIR548H2 (0)
chr21 37470446-37470550	LOC100133286 (0.21)	**chr3 151305596-151305728**	MIR548H2 (0)
chr21 38431524-38431642	PIGP (4.23)	**chr3 157328836-157328958**	C3orf55 (0)
chr21 40110402-40110537	LINC00114 (1.53)	**chr3 166641916-166642049**	ZBBX (0)
chr21 40150491-40150618	LINC00114 (1.53)	**chr3 169530341-169530476**	LRRIQ4 (0.26)
chr21 40917295-40917437	SH3BGR (1.48)	**chr3 173559808-173559945**	NLGN1 (0)
chr21 42312106-42312229	DSCAM (0)	**chr3 174991882-174992024**	NAALADL2 (0)
chr21 45191561-45191690	CSTB (3.95)	**chr3 177345985-177346114**	TBL1XR1 (5.41)
chr21 45276170-45276282	AGPAT3 (4.06)	**chr3 182796631-182796830**	LAMP3 (0.02)
chr21 46224459-46224579	SUMO3 (6.79)	**chr3 183079277-183079409**	MCF2L2 (1.97)
chr21 47013482-47013767	SLC19A1 (4.24)	**chr3 184443714-184443849**	MAGEF1 (4.36)
chr21 47556854-47556988	FTCD (0.13)	**chr3 186314755-186314984**	DNAJB11 (4.65)
chr21 47673893-47674021	MCM3AP (4.87)	**chr3 190707170-190707294**	SNAR-I (0)
chr22 19441977-19442073	UFD1L (5.54)	**chr3 195913907-195914038**	ZDHHC19 (0.41)
chr22 27766484-27766618	MN1 (2.51)	**chr4 1250082-1250221**	C4orf42 (0)
chr22 39710277-39710414	RPL3 (10.68)	**chr4 3915912-3916046**	FAM86EP (1.67)
chr22 46016635-46016760	FBLN1 (0)	**chr4 4552164-4552305**	LOC100507266 (0)
chr22 50051054-50051195	BRD1 (4.09)	**chr4 8289185-8289311**	HTRA3 (1.05)
chr22 50133196-50133337	BRD1 (4.09)	**chr4 11687577-11687710**	HS3ST1 (0.81)
chr3 574690-574821	CHL1 (0.11)	**chr4 17425407-17425531**	QDPR (2.43)
chr3 12882886-12883025	RPL32 (9.69)	**chr4 20785021-20785163**	KCNIP4 (1.16)
chr3 20905789-20905925	VENTXP7 (0)	**chr4 22789819-22789945**	GBA3 (3.19)
chr3 22423377-22423625	ZNF385D (0)	**chr4 23109698-23109839**	MIR548AJ2 (0)
chr3 25824790-25824935	NGLY1 (6.41)	**chr4 35507928-35508059**	ARAP2 (3.09)
chr3 26160799-26160962	LOC285326 (0)	**chr4 38552290-38552425**	FLJ13197 (0)
chr3 26188422-26188545	LOC285326 (0)	**chr4 40300678-40300802**	CHRNA9 (0.07)
chr3 28065186-28065327	CMC1 (4.55)	**chr4 46725842-46726033**	COX7B2 (0)
chr3 31247508-31247619	GADL1 (0.61)	**chr4 56683246-56683367**	LOC644145 (0.32)
chr3 45833137-45833265	LZTFL1 (2.59)	**chr4 67823463-67823605**	CENPC1 (0)
chr3 51074297-51074429	DOCK3 (0.19)	**chr4 77446319-77446479**	SHROOM3 (4.59)
chr3 58464625-58464766	KCTD6 (2.44)	**chr4 77870071-77870202**	40787 (5.45)
chr3 60781681-60781823	PTPRG (3.04)	**chr4 80246570-80246705**	LOC100505875 (0)
chr3 73261431-73261579	PPP4R2 (6.62)	**chr4 81797529-81797672**	C4orf22 (0)
chr3 77778479-77778614	ROBO2 (0.01)	**chr4 95983104-95983234**	UNC5C (0)
chr3 78354645-78354767	ROBO1 (0.03)	**chr4 101087854-101088024**	DDIT4L (6.33)
chr3 79693570-79693698	ROBO1 (0.03)	**chr4 104337436-104337625**	TACR3 (0.1)
chr3 83666176-83666309	LOC440970 (0)	**chr4 107615545-107615664**	DKK2 (0.02)
chr3 89720881-89720981	EPHA3 (0)	**chr4 111140363-111140505**	ELOVL6 (0.81)
chr3 99563666-99563794	MIR548G (0)	**chr4 113033330-113033465**	C4orf32 (2.91)
chr3 104539745-104539845	ALCAM (2.82)	**chr4 113759339-113759475**	ANK2 (2.32)
chr3 107843794-107844001	CD47 (5.01)	**chr4 120221492-120221740**	USP53 (0.74)

chr4 133184010-133184403	PCDH10 (1.37)	**chr6 6649291-6649425**	LY86-AS1 (0)
chr4 142579314-142579441	IL15 (1.26)	**chr6 8672194-8672312**	LOC100506207 (0.91)
chr4 144658193-144658318	FREM3 (0)	**chr6 13472198-13472333**	GFOD1 (3.52)
chr4 145124254-145124375	GYPA (0.09)	**chr6 14911561-14911697**	JARID2 (4.42)
chr4 147887375-147887517	TTC29 (0)	**chr6 18522946-18523072**	MIR548A1 (0)
chr4 153116437-153116579	FBXW7 (6.55)	**chr6 22402081-22402201**	PRL (0.25)
chr4 156167736-156167879	NPY2R (0)	**chr6 31021183-31021327**	HCG22 (1.06)
chr4 162585950-162586088	FSTL5 (3.5)	**chr6 33048598-33048695**	HLA-DPB1 (4.65)
chr4 166771207-166771333	TLL1 (0)	**chr6 33218245-33218345**	RPS18 (12.38)
chr4 185296924-185297043	IRF2 (5.07)	**chr6 39938955-39939097**	MOCS1 (0)
chr4 185451425-185451558	IRF2 (5.07)	**chr6 40741020-40741158**	LRFN2 (3.54)
chr5 1318061-1318193	CLPTM1L (6.48)	**chr6 41873596-41873760**	USP49 (4.04)
chr5 5006905-5007027	LOC340094 (0)	**chr6 45993588-45993720**	CLIC5 (4.09)
chr5 8952112-8952250	SEMA5A (0.01)	**chr6 52129274-52129393**	MCM3 (8.08)
chr5 11057528-11057662	CTNND2 (0)	**chr6 63320765-63320896**	KHDRBS2 (0)
chr5 20922217-20922340	GUSBP1 (5.85)	**chr6 65553916-65554016**	EYS (0.34)
chr5 21918618-21918789	CDH12 (0.32)	**chr6 67778155-67778387**	MCART3P (0)
chr5 35729920-35730057	SPEF2 (2.44)	**chr6 67780069-67780198**	MCART3P (0)
chr5 38868618-38868760	RICTOR (4.03)	**chr6 82804914-82805056**	IBTK (4.62)
chr5 40604014-40604157	PTGER4 (2.21)	**chr6 85506210-85506347**	TBX18 (0)
chr5 43549556-43549701	PAIP1 (5.84)	**chr6 86417333-86417475**	SNHG5 (7.04)
chr5 53356296-53356429	ARL15 (1.81)	**chr6 100070081-100070216**	PRDM13 (0)
chr5 53358354-53358485	ARL15 (1.81)	**chr6 100412309-100412434**	MCHR2 (0)
chr5 53686642-53686758	HSPB3 (0)	**chr6 107627530-107627667**	SOBP (1.2)
chr5 66840974-66841094	CD180 (0.12)	**chr6 111599793-111599902**	KIAA1919 (3.09)
chr5 67794748-67794884	PIK3R1 (4.89)	**chr6 116238597-116238732**	FRK (1.42)
chr5 79696220-79696362	ZFYVE16 (3.25)	**chr6 116579912-116580018**	TSPYL4 (3.69)
chr5 87343402-87343526	TMEM161B (3.66)	**chr6 125351121-125351263**	RNF217 (0.75)
chr5 100963552-100963680	SLCO4C1 (0)	**chr6 128696428-128696557**	PTPRK (6.16)
chr5 115411013-115411136	COMMD10 (4.82)	**chr6 131429444-131429544**	AKAP7 (2.26)
chr5 115481477-115481601	COMMD10 (4.82)	**chr6 134719740-134719871**	LOC154092 (0)
chr5 115518618-115518742	COMMD10 (4.82)	**chr6 138200121-138200221**	PERP (2.72)
chr5 120500450-120500594	PRR16 (0)	**chr6 139329327-139329469**	REPS1 (5.17)
chr5 123823907-123824048	ZNF608 (6.58)	**chr6 145790211-145790337**	EPM2A (2.36)
chr5 127541722-127541858	SLC12A2 (2.53)	**chr6 156598966-156599132**	ARID1B (4.14)
chr5 130204729-130204864	HINT1 (8.67)	**chr6 156614765-156614886**	ARID1B (4.14)
chr5 131753988-131754112	SLC22A5 (2.16)	**chr6 164469271-164469406**	QKI (3.13)
chr5 138282775-138282908	SIL1 (3.53)	**chr7 4806876-4806986**	KIAA0415 (0)
chr5 139624808-139625018	C5orf32 (0)	**chr7 6441885-6442106**	RAC1 (4.82)
chr5 141019033-141019157	RELL2 (1.74)	**chr7 12971266-12971401**	ARL4A (0.23)
chr5 157249352-157249494	CLINT1 (4.55)	**chr7 14087829-14087949**	ETV1 (0)
chr5 157465408-157465537	CLINT1 (4.55)	**chr7 16891568-16891680**	AGR3 (0)
chr5 162864559-162864671	NUDCD2 (5.52)	**chr7 44053208-44053332**	POLR2J4 (1.75)
chr5 163681917-163682074	MAT2B (5.31)	**chr7 44545458-44545572**	NPC1L1 (0)
chr5 166959408-166959537	WWC1 (0)	**chr7 45412787-45412916**	RAMP3 (0)
chr5 168043828-168043909	PANK3 (5.37)	**chr7 66071823-66071971**	KCTD7 (2.2)
chr5 169492575-169492708	DOCK2 (5.65)	**chr7 68782073-68782213**	AUTS2 (4.14)
chr5 176884003-176884211	DBN1 (6.86)	**chr7 69740740-69740862**	AUTS2 (4.14)
chr5 179264005-179264179	C5orf45 (3.33)	**chr7 71222097-71222231**	CALN1 (1.95)
chr6 3195908-3196040	TUBB2B (0.34)	**chr7 75137117-75137243**	PMS2P3 (3.2)
chr6 4422166-4422295	CDYL (4.14)	**chr7 76157028-76157154**	UPK3B (1.64)
chr6 5548212-5548354	LYRM4 (5.34)	**chr7 82523594-82523877**	CACNA2D1 (1.02)
chr6 5756640-5756774	FARS2 (4.21)	**chr7 92218074-92218232**	CDK6 (7.47)
chr6 5795597-5795739	FARS2 (4.21)	**chr7 92244317-92244552**	CDK6 (7.47)
chr6 5805585-5805756	FARS2 (4.21)	**chr7 92561996-92562119**	CDK6 (7.47)
chr6 5864969-5865107	FARS2 (4.21)	**chr7 93002680-93002815**	CCDC132 (4.03)

chr7 93586828-93586967	GNG11 (5.2)	**chr9 2303015-2303158**	SMARCA2 (5.59)
chr7 93652093-93652224	BET1 (3.3)	**chr9 7989015-7989135**	C9orf123 (0)
chr7 99532790-99532898	GJC3 (0.37)	**chr9 16469380-16469515**	BNC2 (0.01)
chr7 101313579-101313718	MYL10 (0)	**chr9 19037512-19037647**	FAM154A (0)
chr7 101982472-101982696	SPDYE6 (1.16)	**chr9 26752854-26752979**	C9orf82 (0)
chr7 102207456-102207684	SPDYE2 (0)	**chr9 30139771-30139888**	MIR873 (0)
chr7 102212913-102213046	RASA4 (1.87)	**chr9 33917775-33918000**	UBAP2 (3.85)
chr7 102306538-102306757	SPDYE2L (0)	**chr9 38261886-38262022**	ALDH1B1 (3.63)
chr7 102311989-102312159	SPDYE2L (0)	**chr9 83145423-83145564**	TLE4 (3.1)
chr7 105646449-105646576	CDHR3 (0.04)	**chr9 83537533-83537668**	TLE1 (5.31)
chr7 105648378-105648507	CDHR3 (0.04)	**chr9 83714049-83714183**	TLE1 (5.31)
chr7 117636066-117636208	CTTNBP2 (0.01)	**chr9 92470386-92470521**	UNQ6494 (0)
chr7 119681687-119681822	KCND2 (0)	**chr9 93725465-93725584**	SYK (5.56)
chr7 121080886-121081022	FAM3C (2.9)	**chr9 93759281-93759418**	LOC100129316 (0)
chr7 123193294-123193428	NDUFA5 (3.82)	**chr9 96892984-96893109**	PTPDC1 (2.89)
chr7 125141052-125141148	POT1 (3.66)	**chr9 100828108-100828242**	TRIM14 (3.03)
chr7 127172205-127172324	GCC1 (3.15)	**chr9 102665041-102665182**	STX17 (3.12)
chr7 127769436-127769577	SND1 (6.92)	**chr9 103308510-103308643**	C9orf30-TMEFF1 (0)
chr7 128303374-128303503	FLJ45340 (0)	**chr9 103801901-103802038**	LPPR1 (0)
chr7 139502615-139502717	HIPK2 (2.52)	**chr9 106225386-106225520**	CYLC2 (0)
chr7 139677024-139677159	TBXAS1 (0.7)	**chr9 110550881-110551022**	KLF4 (0)
chr7 140895540-140895712	LOC100507421 (0)	**chr9 112407435-112407564**	PALM2 (0)
chr7 140980248-140980380	LOC100507421 (0)	**chr9 116460433-116460567**	RGS3 (1.98)
chr7 148336218-148336389	C7orf33 (0)	**chr9 116466404-116466526**	RGS3 (1.98)
chr7 151462282-151462405	PRKAG2 (1.21)	**chr9 117099091-117099188**	AKNA (3.93)
chr7 154998350-154998485	INSIG1 (7.1)	**chr9 125196265-125196406**	PTGS1 (0.73)
chr7 158172527-158172657	PTPRN2 (0.13)	**chr9 125395177-125395313**	OR1B1 (0)
chr8 2980164-2980305	CSMD1 (0.83)	**chr9 129547267-129547409**	ZBTB43 (2.04)
chr8 3137210-3137322	CSMD1 (0.83)	**chr9 137764383-137764515**	FCN2 (0)
chr8 4632314-4632425	CSMD1 (0.83)	**chr9 139548292-139548426**	EGFL7 (6.59)
chr8 15199098-15199220	SGCZ (0.11)	**chrM 5124-5257**	-1 (0)
chr8 15274672-15274813	TUSC3 (3.66)	**chrUn_gl000220 109887-110022**	RN5-8S1 (0)
chr8 16314354-16314677	MSR1 (4.46)	**chrUn_gl000220 114671-114919**	RN5-8S1 (0)
chr8 16661016-16661116	FGF20 (0)	**chrUn_gl000220 153859-153994**	RN5-8S1 (0)
chr8 18820518-18820659	PSD3 (4.21)	**chrUn_gl000220 158650-158891**	RN5-8S1 (0)
chr8 20917364-20917505	LOC286114 (0)	**chrX 1863342-1863480**	ASMT (0)
chr8 22351713-22351847	SORBS3 (3.11)	**chrX 7029122-7029270**	MIR4767 (0)
chr8 49430533-49430684	EFCAB1 (0)	**chrX 7285584-7285713**	STS (1.82)
chr8 57736361-57736490	IMPAD1 (4.44)	**chrX 7819833-7819975**	VCX (0.09)
chr8 58890922-58891057	FAM110B (0)	**chrX 12102371-12102513**	FRMPD4 (0)
chr8 60558667-60558794	TOX (0.02)	**chrX 26928857-26928991**	VENTXP1 (0)
chr8 65415723-65415854	LOC401463 (0)	**chrX 33733174-33733343**	DMD (0.75)
chr8 68421177-68421300	ARFGEF1 (4.1)	**chrX 33739868-33740003**	DMD (0.75)
chr8 71452837-71452956	TRAM1 (6.16)	**chrX 34097574-34097715**	FAM47A (0.83)
chr8 72804831-72804972	LOC100132891 (0)	**chrX 48760226-48760344**	PQBP1 (3.94)
chr8 82963493-82963603	SNX16 (2.26)	**chrX 62192812-62192947**	SPIN4 (2.41)
chr8 85828753-85828885	RALYL (0)	**chrX 68019175-68019316**	EFNB1 (4.03)
chr8 93281929-93282033	RUNX1T1 (0)	**chrX 78887098-78887228**	ITM2A (0.75)
chr8 100249258-100249385	VPS13B (3.37)	**chrX 85270289-85270391**	DACH2 (0)
chr8 111580965-111581106	KCNV1 (0)	**chrX 93946269-93946405**	FAM133A (0)
chr8 117816628-117816728	UTP23 (4.14)	**chrX 93990317-93990446**	FAM133A (0)
chr8 127019377-127019550	LOC100130231 (0)	**chrX 105349131-105349251**	MUM1L1 (0)
chr8 130034971-130035107	LOC728724 (0)	**chrX 111782272-111782407**	ZCCHC16 (0)
chr8 134392062-134392181	ST3GAL1 (3.41)	**chrX 119138349-119138491**	NKAP (3.76)
chr8 140026425-140026564	COL22A1 (0.01)	**chrX 135056048-135056172**	SLC9A6 (2.8)
chr8 144272817-144272943	GPIHBP1 (0)	**chrX 153060058-153060247**	SSR4 (7.79)

chrX 153195426-153195654	NAA10 (6.13)	**chrY 15398778-15399017**	UTY (0)
chrY 1813342-1813480	ASMT (0)		

Supplementary Data Table 2. Closest genes to active LTRs detected by RACE-Seq in **Namalwa** cells. Gene expression values (FPKM) from RNA-Seq are also shown for each gene.

RACE Peak	Gene (FPKM)
chr1 565687-565793	OR4F16 (0)
chr1 2873191-2873294	ACTRT2 (0)
chr1 3492137-3492246	MEGF6 (0)
chr1 11467915-11468064	PTCHD2 (0.08)
chr1 11640142-11640274	PTCHD2 (0.08)
chr1 16160830-16161078	SPEN (3.89)
chr1 24814729-24814843	RCAN3 (1.7)
chr1 25589898-25590066	RHD (0.88)
chr1 25660224-25660391	RHD (0.88)
chr1 31945290-31945453	LOC149086 (0)
chr1 38959146-38959288	LOC339442 (0)
chr1 55404744-55404851	TMEM61 (0)
chr1 57271450-57271561	C1orf168 (0.02)
chr1 66594411-66594525	PDE4B (3.21)
chr1 68019995-68020148	SERBP1 (6.87)
chr1 85839200-85839328	DDAH1 (0.07)
chr1 87911300-87911404	LOC100505768 (0)
chr1 91328391-91328524	ZNF644 (4.97)
chr1 97721043-97721154	DPYD (4.44)
chr1 101455646-101455978	DPH5 (4.33)
chr1 101491228-101491359	DPH5 (4.33)
chr1 101853796-101853916	S1PR1 (1.62)
chr1 108093064-108093214	VAV3 (0.39)
chr1 148213989-148214116	PPIAL4D (0)
chr1 148632427-148632529	PPIAL4E (0)
chr1 148794459-148794586	PPIAL4D (0)
chr1 153478547-153478695	S100A6 (0)
chr1 154281795-154281904	AQP10 (0)
chr1 159687104-159687209	CRP (0)
chr1 160682581-160682707	CD48 (7.1)
chr1 164446863-164446974	PBX1 (0.04)
chr1 168731055-168731165	MGC4473 (0)
chr1 171238114-171238225	FMO4 (1.89)
chr1 177701423-177701557	SEC16B (0)
chr1 182202214-182202319	GLUL (4.3)
chr1 193978557-193978671	CDC73 (2.5)
chr1 197964819-197964935	LHX9 (0)
chr1 207269193-207269296	C4BPA (0)
chr1 209874574-209874688	HSD11B1 (0)
chr1 215159433-215159549	KCNK2 (0)
chr1 219056302-219056439	LOC643723 (0)
chr1 224773977-224774092	CNIH3 (0.1)
chr1 226414505-226414605	MIXL1 (0.36)
chr1 241575474-241575588	RGS7 (0)
chr1 243269465-243269608	LOC731275 (0)
chr1 245121001-245121152	EFCAB2 (3.57)
chr10 15225856-15225970	NMT2 (0.67)
chr10 30536812-30536917	MTPAP (3.98)
chr10 31328037-31328147	ZNF438 (1.55)
chr10 85909878-85910002	C10orf99 (0)
chr10 88405003-88405110	OPN4 (0)
chr10 89370432-89370534	PAPSS2 (0.15)
chr10 98497235-98497402	PIK3AP1 (6.64)
chr10 103086554-103086703	BTRC (2.64)
chr10 107729823-107729931	SORCS1 (0.02)
chr10 125171650-125171754	BUB3 (7.41)
chr10 132469163-132469278	MIR378C (0)
chr10 134073068-134073173	STK32C (0.6)
chr11 214864-215084	RIC8A (4.98)
chr11 1285719-1285824	MUC5B (0.02)
chr11 5526794-5526909	OR51B5 (0)
chr11 7486962-7487112	SYT9 (0.02)
chr11 10830384-10830488	EIF4G2 (8.22)
chr11 13010984-13011131	RASSF10 (0)
chr11 14541845-14541960	PSMA1 (7.76)
chr11 21736666-21736782	NELL1 (0)
chr11 27911830-27912125	KIF18A (4.6)
chr11 36607827-36608045	RAG2 (5.31)
chr11 43702273-43702376	HSD17B12 (4.9)
chr11 46450217-46450387	AMBRA1 (4)
chr11 73685681-73685984	DNAJB13 (0)
chr11 76647125-76647268	ACER3 (2.32)
chr11 89984314-89984482	CHORDC1 (4.22)
chr11 102217944-102218085	BIRC2 (4.44)
chr11 113749374-113749523	USP28 (3.92)
chr11 115797114-115797215	LOC283143 (0)
chr11 116088771-116088880	LOC283143 (0)
chr11 119773491-119773639	PVRL1 (0.9)
chr11 126517987-126518125	KIRREL3 (0.02)
chr11 128240076-128240191	ETS1 (6.81)
chr12 1635020-1635188	LOC100292680 (0)
chr12 1756248-1756377	MIR3649 (0)
chr12 9797863-9797991	LOC374443 (5.13)
chr12 9804044-9804181	CLEC2D (5.18)
chr12 21810684-21810797	LDHB (10.32)
chr12 22421445-22421575	KIAA0528 (0)
chr12 25108462-25108607	BCAT1 (6.86)
chr12 43345713-43345828	PRICKLE1 (0.4)
chr12 46380557-46380706	ARID2 (3.84)
chr12 47066210-47066324	SLC38A4 (0)
chr12 50505955-50506066	C12orf62 (0)
chr12 56043446-56043549	OR10P1 (0)
chr12 56506761-56506865	RPL41 (12.42)
chr12 57081803-57081912	NACA (9.69)
chr12 65598794-65598905	MSRB3 (0.02)
chr12 66414345-66414459	HMGA2 (0)
chr12 76924859-76924965	OSBPL8 (5.19)
chr12 83014550-83014654	TMTC2 (1.26)
chr12 85244592-85244748	SLC6A15 (0)

chr12 87715690-87715794	MKRN9P (0)
chr12 91058850-91058961	C12orf37 (0)
chr12 93335084-93335242	EEA1 (2.9)
chr12 110995605-110995706	PPTC7 (2.61)
chr12 111848628-111848749	ATXN2 (3.02)
chr12 112847271-112847384	PTPN11 (5.71)
chr12 113471008-113471119	OAS2 (4.76)
chr12 115345785-115345883	TBX3 (0)
chr12 116356598-116356713	MED13L (3.13)
chr12 124328605-124328706	DNAH10 (0.04)
chr12 125597287-125597398	AACS (2.84)
chr12 129263089-129263203	SLC15A4 (3.72)
chr12 129594172-129594323	TMEM132D (0)
chr13 21751173-21751268	MRP63 (0)
chr13 28651435-28651586	PAN3-AS1 (1.49)
chr13 31037352-31037463	LINC00426 (2.45)
chr13 38410191-38410335	TRPC4 (0)
chr13 45647999-45648098	KIAA1704 (0)
chr13 59883237-59883342	DIAPH3 (4.59)
chr13 63631326-63631429	OR7E156P (0)
chr13 78003747-78003898	MYCBP2 (5.38)
chr13 78500948-78501141	EDNRB (0)
chr13 87778434-87778548	MIR4500HG (0)
chr13 91723885-91723999	LINC00410 (0)
chr13 96308340-96308454	DZIP1 (0.51)
chr13 100104206-100104323	MIR548AN (0)
chr13 101360611-101360719	TMTC4 (3.41)
chr13 103752584-103752684	SLC10A2 (1.27)
chr13 112025171-112025313	C13orf16 (0)
chr13 114518528-114518686	FAM70B (0)
chr14 19680939-19681050	POTEG (0)
chr14 19894334-19894445	POTEM (0)
chr14 21134864-21134978	ANG (0.29)
chr14 30348582-30348698	PRKD1 (0)
chr14 50872776-50873227	CDKL1 (3.04)
chr14 51651857-51652002	TMX1 (4.07)
chr14 51960191-51960301	FRMD6 (0.34)
chr14 56780067-56780181	PELI2 (2.43)
chr14 60677148-60677263	PPM1A (3.5)
chr14 71713069-71713167	PCNX (2.51)
chr14 80186459-80186569	NRXN3 (0.02)
chr14 81723108-81723221	STON2 (0.85)
chr14 88746274-88746362	KCNK10 (0)
chr14 89744494-89744609	FOXN3 (2.63)
chr14 92630286-92630528	CPSF2 (6)
chr14 92868969-92869077	RIN3 (0.05)
chr14 93336869-93336984	GOLGA5 (4.05)
chr14 93338650-93338765	GOLGA5 (4.05)
chr14 93339800-93339915	GOLGA5 (4.05)
chr14 94734966-94735081	SERPINA10 (0.05)
chr14 96288070-96288173	LOC100507043 (0)
chr14 97903754-97903899	LOC100129345 (0)
chr14 100800122-100800292	SLC25A47 (0)
chr15 23188121-23188273	WHAMMP3 (1.49)
chr15 33502030-33502182	RYR3 (0.08)
chr15 37745919-37746021	MEIS2 (1.96)

chr15 40826581-40826683	MRPL42P5 (0.96)
chr15 45923317-45923434	SQRDL (1.74)
chr15 56135259-56135370	PRTG (0.03)
chr15 62574071-62574186	C2CD4B (0)
chr15 69745134-69745246	KIF23 (5.18)
chr15 70885592-70885701	UACA (0.88)
chr15 74756740-74756878	LOC440288 (0)
chr15 75230268-75230447	RPP25 (4.2)
chr15 87581916-87582057	AGBL1 (0)
chr15 89198635-89198730	ISG20 (4.7)
chr15 89499696-89499795	MFGE8 (2.62)
chr15 93946479-93946628	RGMA (0)
chr15 101835334-101835447	PCSK6 (0.14)
chr16 8708565-8708679	METTL22 (3.59)
chr16 15657464-15657574	KIAA0430 (3.1)
chr16 21127725-21127866	DNAH3 (0.05)
chr16 23592362-23592466	NDUFAB1 (7.84)
chr16 27050426-27050537	C16orf82 (0)
chr16 27147946-27148051	JMJD5 (0)
chr16 31123571-31123821	KAT8 (5.11)
chr16 54907968-54908119	CRNDE (5.09)
chr16 55889901-55890039	CES5A (0)
chr16 58455103-58455339	GINS3 (5.28)
chr16 60118540-60118651	LOC644649 (0)
chr16 69669475-69669625	NFAT5 (2.84)
chr16 70002236-70002337	CLEC18A (0.04)
chr16 70225150-70225251	CLEC18C (0)
chr16 74438090-74438193	CLEC18B (0.13)
chr16 76268959-76269073	CNTNAP4 (0.07)
chr16 80926390-80926530	C16orf61 (0)
chr16 83806987-83807088	CDH13 (0.02)
chr16 83879075-83879169	HSBP1 (5.92)
chr16 85360119-85360277	LOC727710 (0)
chr16 88876100-88876217	APRT (7.52)
chr17 7529673-7529769	SHBG (0.17)
chr17 27717452-27717561	MIR4523 (0)
chr17 30771497-30771628	PSMD11 (5)
chr17 35075647-35075757	MRM1 (3.7)
chr17 58140237-58140385	HEATR6 (2.05)
chr17 64899520-64899627	CACNG5 (0.2)
chr17 69617074-69617168	SOX9 (0)
chr17 70016624-70016776	SOX9 (0)
chr17 73262119-73262266	MIF4GD (3.91)
chr17 74560542-74560709	ST6GALNAC2 (0.82)
chr17 79336009-79336125	TMEM105 (0)
chr17 79891163-79891269	PYCR1 (6.68)
chr18 18770173-18770322	GREB1L (0.04)
chr18 26055023-26055136	CDH2 (0)
chr18 29545582-29545726	TRAPPC8 (4.85)
chr18 33077777-33077892	MIR3975 (0)
chr18 42251873-42251981	SETBP1 (1.8)
chr18 44812039-44812178	IER3IP1 (5.2)
chr18 45953740-45953842	CTIF (1.09)
chr18 48043831-48043971	MAPK4 (0)
chr18 48668633-48668760	MEX3C (5.94)
chr18 55636112-55636213	NEDD4L (2.21)

chr18 56447358-56447482	MALT1 (5.1)
chr18 57855825-57855937	MC4R (0)
chr18 75705590-75705701	GALR1 (0)
chr18 76141404-76141504	SALL3 (0)
chr18 76717551-76717751	SALL3 (0)
chr18 77352739-77352881	NFATC1 (3.67)
chr19 5978361-5978555	RANBP3 (4.98)
chr19 6413865-6414019	KHSRP (5.98)
chr19 7696354-7696449	PCP2 (0.82)
chr19 14281838-14281992	LPHN1 (0)
chr19 34046691-34046836	PEPD (4.19)
chr19 34273547-34273648	CHST8 (0)
chr19 40157661-40157810	LGALS16 (0)
chr19 42891514-42891627	MEGF8 (2.33)
chr19 47776921-47777021	PRR24 (0)
chr19 49993711-49993864	SNORD33 (0)
chr19 50363571-50363657	PNKP (4.08)
chr19 50363833-50363953	PNKP (4.08)
chr19 54846585-54846693	LAIR1 (0.33)
chr19 56824647-56824764	ZNF542 (0)
chr2 668003-668228	LOC339822 (0)
chr2 6102010-6102121	LOC400940 (0)
chr2 10266050-10266225	C2orf48 (1.6)
chr2 11720083-11720233	MIR4429 (0)
chr2 14131239-14131391	FAM84A (0)
chr2 27907986-27908097	SLC4A1AP (4.33)
chr2 30554544-30554658	LBH (3.31)
chr2 30906180-30906320	LCLAT1 (3.35)
chr2 33306924-33307022	LTBP1 (2.61)
chr2 33931890-33932042	MYADML (0)
chr2 36275764-36275873	LOC100288911 (0)
chr2 38127199-38127309	FAM82A1 (0)
chr2 40144686-40144774	LOC100128590 (0)
chr2 49398277-49398388	FSHR (0)
chr2 53156003-53156112	ASB3 (4.4)
chr2 54342720-54342847	ACYP2 (2.97)
chr2 54501248-54501356	ACYP2 (2.97)
chr2 60343682-60343832	MIR4432 (0)
chr2 62003030-62003147	FAM161A (2.57)
chr2 62024772-62024880	FAM161A (2.57)
chr2 65275035-65275387	CEP68 (2.81)
chr2 68934399-68934502	ARHGAP25 (1.5)
chr2 69383725-69383836	ANTXR1 (0)
chr2 71349610-71350073	MPHOSPH10 (4.98)
chr2 74154005-74154141	DGUOK (5.99)
chr2 76532732-76532874	LRRTM4 (0)
chr2 76571471-76571573	LRRTM4 (0)
chr2 79213601-79213754	REG3G (0)
chr2 100544366-100544573	AFF3 (4.05)
chr2 104629686-104629811	LOC100287010 (0.1)
chr2 111881302-111881398	ACOXL (0.04)
chr2 118480734-118480875	DDX18 (6.02)
chr2 118781948-118782097	CCDC93 (3.2)
chr2 124350029-124350169	CNTNAP5 (0)
chr2 128586982-128587130	POLR2D (5.04)
chr2 132693173-132693323	C2orf27B (0)

chr2 152470238-152470394	NEB (0.11)
chr2 155062106-155062253	GALNT13 (0)
chr2 164518258-164518406	KCNH7 (0.02)
chr2 170543424-170543576	C2orf77 (0)
chr2 177293849-177293999	MTX2 (4.43)
chr2 180167047-180167193	SESTD1 (1.88)
chr2 191573359-191573470	NAB1 (3.54)
chr2 192364093-192364242	MYO1B (0)
chr2 196313207-196313355	SLC39A10 (4)
chr2 216848784-216848895	PECR (0)
chr2 217363565-217363652	SMARCAL1 (3.17)
chr2 217363958-217364045	SMARCAL1 (3.17)
chr2 224651841-224651946	WDFY1 (3.94)
chr2 225232401-225232500	FAM124B (0)
chr2 238378158-238378263	MLPH (0)
chr2 239683163-239683279	TWIST2 (0)
chr20 1772415-1772560	LOC100289473 (0)
chr20 7391918-7392068	HAO1 (0)
chr20 16859130-16859269	OTOR (0)
chr20 21797663-21797752	PAX1 (0)
chr20 35617265-35617434	RBL1 (4.4)
chr20 39570744-39570900	TOP1 (5.13)
chr20 40256542-40256675	CHD6 (3.27)
chr20 43588934-43588999	LOC100505826 (0)
chr20 56058404-56058553	CTCFL (0.03)
chr20 57395709-57395818	GNAS-AS1 (2.01)
chr20 58487911-58488017	SYCP2 (0.34)
chr21 14599916-14600066	ANKRD30BP2 (0.03)
chr21 19858095-19858245	TMPRSS15 (1.81)
chr21 20184882-20184986	TMPRSS15 (1.81)
chr21 20834136-20834243	TMPRSS15 (1.81)
chr21 26734081-26734196	LINC00158 (0.24)
chr21 28372021-28372136	ADAMTS5 (0)
chr21 29420727-29420870	LINC00314 (0)
chr21 30638206-30638350	BACH1 (4.22)
chr21 35284584-35284736	LOC100506334 (0)
chr21 43881482-43881632	RSPH1 (0.25)
chr21 46038162-46038264	TSPEAR (0.11)
chr21 47013485-47013624	SLC19A1 (3.96)
chr21 47556864-47556978	FTCD (0.09)
chr21 47673885-47674011	MCM3AP (4.8)
chr22 16162031-16162142	POTEH (0)
chr22 27766481-27766617	MN1 (0)
chr22 28382177-28382337	TTC28-AS1 (2.92)
chr22 32014338-32014554	PRR14L (3.55)
chr22 35626932-35627092	HMGXB4 (3.19)
chr22 41347926-41348012	XPNPEP3 (3.07)
chr22 46016631-46016768	FBLN1 (0)
chr22 50051057-50051197	BRD1 (3.79)
chr3 12882913-12883029	RPL32 (9.77)
chr3 25824782-25824922	NGLY1 (5.5)
chr3 26160840-26160944	LOC285326 (0)
chr3 51074308-51074419	DOCK3 (0.16)
chr3 51820908-51821023	IQCF6 (0)
chr3 53231284-53231397	PRKCD (2.03)
chr3 59819602-59819704	FHIT (2.39)

chr3 62936082-62936193	CADPS (2.87)
chr3 70928314-70928452	FOXP1 (4.04)
chr3 80577733-80577876	ROBO1 (0)
chr3 88354685-88354822	C3orf38 (3.2)
chr3 99563680-99563784	MIR548G (0)
chr3 107843761-107844005	CD47 (5.04)
chr3 110599699-110599785	PVRL3-AS1 (0)
chr3 116911886-116912001	LSAMP-AS3 (0)
chr3 118528784-118528899	IGSF11 (0)
chr3 118529555-118529663	IGSF11 (0)
chr3 120315147-120315375	HGD (0.09)
chr3 129037346-129037437	H1FX-AS1 (1.27)
chr3 130584513-130584629	ATP2C1 (3.9)
chr3 139891519-139891624	CLSTN2 (0)
chr3 140480170-140480286	TRIM42 (0)
chr3 144087395-144087506	C3orf58 (3.35)
chr3 148677815-148677962	GYG1 (4.51)
chr3 150321087-150321228	SELT (5.69)
chr3 151792400-151792514	LOC401093 (0)
chr3 158770769-158770885	IQCJ (0.12)
chr3 169530337-169530480	LRRIQ4 (0.36)
chr3 174991777-174992418	NAALADL2 (0.5)
chr3 182796648-182796830	LAMP3 (0.08)
chr3 183670436-183670538	ABCC5 (3.26)
chr3 185743973-185744122	ETV5 (0.1)
chr3 188057286-188057387	LPP (3.88)
chr3 191018563-191018664	CCDC50 (3.59)
chr3 193563170-193563275	LOC647323 (0)
chr3 194968496-194968611	ACAP2 (4.65)
chr3 195913918-195914024	ZDHHC19 (0)
chr4 53192-53307	ZNF876P (0.31)
chr4 309644-309748	ZNF732 (0.04)
chr4 467830-467962	PIGG (2.76)
chr4 1250135-1250221	C4orf42 (0)
chr4 2384530-2384679	ZFYVE28 (0.52)
chr4 4822987-4823141	MSX1 (0.09)
chr4 16242612-16242717	FLJ39653 (0)
chr4 34898769-34898883	ARAP2 (2.72)
chr4 36722659-36722761	DTHD1 (0)
chr4 38552300-38552415	FLJ13197 (0)
chr4 42463443-42463589	ATP8A1 (4.39)
chr4 46725796-46726037	COX7B2 (0)
chr4 61911261-61911373	LPHN3 (0)
chr4 63788175-63788263	LPHN3 (0)
chr4 65213256-65213358	LOC401134 (0)
chr4 77446326-77446462	SHROOM3 (0.04)
chr4 77870081-77870192	40787 (5.46)
chr4 80101015-80101117	LOC100505875 (0)
chr4 81797509-81797675	C4orf22 (1.8)
chr4 81952971-81953167	PRKG2 (0)
chr4 101087876-101088028	DDIT4L (0)
chr4 104337471-104337614	TACR3 (0)
chr4 113203286-113203402	TIFA (5.41)
chr4 113759350-113759455	ANK2 (2.61)
chr4 120221477-120221773	USP53 (4.58)
chr4 123001147-123001283	KIAA1109 (2.91)

chr4 130000058-130000176	C4orf33 (3.17)
chr4 133184024-133184213	PCDH10 (0)
chr4 142579328-142579432	IL15 (0.04)
chr4 168971037-168971153	ANXA10 (0)
chr4 175285024-175285166	CEP44 (3.67)
chr4 178653522-178653635	LOC285501 (0)
chr5 1318002-1318196	CLPTM1L (6.15)
chr5 1507743-1507854	SLC6A3 (0)
chr5 2268714-2268834	IRX4 (0)
chr5 6195529-6195635	FLJ33360 (0)
chr5 8952108-8952255	SEMA5A (0)
chr5 11057543-11057653	CTNND2 (0)
chr5 17123884-17123983	LOC285696 (0)
chr5 22618992-22619104	CDH12 (0.25)
chr5 23828371-23828481	PRDM9 (0)
chr5 38868614-38868764	RICTOR (3.13)
chr5 40604012-40604197	PTGER4 (1.3)
chr5 49981571-49981680	PARP8 (3.21)
chr5 51202158-51202272	ISL1 (0)
chr5 53686633-53686774	HSPB3 (0)
chr5 68575714-68575875	CCDC125 (3)
chr5 77262627-77262729	AP3B1 (5.6)
chr5 91206664-91206813	LOC100129716 (0)
chr5 92840276-92840381	FLJ42709 (0)
chr5 99708329-99708441	LOC100133050
chr5 100963543-100963693	SLCO4C1 (0)
chr5 103717017-103717126	RAB9BP1 (0)
chr5 105502151-105502247	RAB9BP1 (0)
chr5 113462037-113462136	KCNN2 (0.06)
chr5 117818363-117818477	DTWD2 (0.84)
chr5 120500320-120500617	PRR16 (0)
chr5 121663268-121663369	SNCAIP (0)
chr5 123510312-123510415	ZNF608 (3.93)
chr5 127541731-127541848	SLC12A2 (3.81)
chr5 131162422-131162527	FNIP1 (3.34)
chr5 134831318-134831467	NEUROG1 (0)
chr5 141019043-141019147	RELL2 (4.29)
chr5 149776168-149776385	CD74 (9.56)
chr6 3195904-3196044	TUBB2B (0.04)
chr6 6649301-6649415	LY86-AS1 (1.22)
chr6 13472208-13472323	GFOD1 (3.86)
chr6 16990395-16990506	FLJ23152 (0)
chr6 31021197-31021312	HCG22 (0.23)
chr6 33029816-33029920	HLA-DPA1 (6.81)
chr6 35907226-35907331	SLC26A8 (0.03)
chr6 37717355-37717488	MDGA1 (0)
chr6 71085430-71085580	FAM135A (3.92)
chr6 72163935-72164036	LINC00472 (0)
chr6 74555612-74555761	CD109 (0.01)
chr6 85506206-85506349	TBX18 (0)
chr6 86417346-86417461	SNHG5 (5.84)
chr6 90775000-90775216	BACH2 (6.48)
chr6 91195836-91195947	MAP3K7 (4.93)
chr6 98376249-98376363	MIR2113 (0)
chr6 111599770-111599892	KIAA1919 (3.22)
chr6 115348269-115348377	FRK (0.54)

chr6 119765569-119765673	MAN1A1 (6.33)
chr6 123225590-123225736	CLVS2 (0)
chr6 131083093-131083198	LOC100507203 (0)
chr6 136546672-136546786	FAM54A (0)
chr6 137314307-137314421	IL20RA (0)
chr6 138367531-138367645	PERP (0)
chr6 142287284-142287380	NMBR (0)
chr6 143981411-143981522	PHACTR2 (0.4)
chr6 167463692-167463788	FGFR1OP (3.59)
chr7 1899641-1900049	MAD1L1 (5.21)
chr7 6441881-6442190	RAC1 (5.99)
chr7 7558756-7558870	COL28A1 (0.02)
chr7 12971276-12971391	ARL4A (3.42)
chr7 14087839-14087941	ETV1 (0)
chr7 20941907-20942053	RPL23P8 (0)
chr7 26129297-26129410	NFE2L3 (4.34)
chr7 38435928-38436037	AMPH (0)
chr7 44053205-44053336	POLR2J4 (1.79)
chr7 44054181-44054284	POLR2J4 (1.79)
chr7 66071810-66071971	KCTD7 (3.41)
chr7 70419588-70419703	AUTS2 (4.46)
chr7 70674746-70674895	WBSCR17 (0)
chr7 72486819-72486933	PMS2L2 (0)
chr7 72514881-72514995	PMS2L2 (0)
chr7 74707626-74707740	LOC100093631
chr7 74921899-74922072	SPDYE8P (0.39)
chr7 74949915-74950074	SPDYE8P (0.39)
chr7 75137116-75137247	PMS2P3 (3.79)
chr7 76156935-76157156	UPK3B (2.61)
chr7 82523708-82523823	CACNA2D1 (0.03)
chr7 87130546-87130647	ABCB1 (0)
chr7 91884314-91884423	ANKIB1 (3.65)
chr7 93652087-93652224	BET1 (3.98)
chr7 99229640-99229810	ZNF498 (0)
chr7 99532776-99532898	GJC3 (0.08)
chr7 99923308-99923478	PMS2P1 (5.13)
chr7 99933458-99933617	PILRB (3.46)
chr7 101982568-101982699	SPDYE6 (2.4)
chr7 102207457-102207684	SPDYE2 (0)
chr7 102208433-102208536	SPDYE2 (0)
chr7 102212913-102213033	RASA4 (2.93)
chr7 102306537-102306764	SPDYE2L (0)
chr7 102307513-102307616	SPDYE2L (0)
chr7 102311966-102312192	SPDYE2L (0)
chr7 105646455-105646566	CDHR3 (0.57)
chr7 106145151-106145265	C7orf74 (0)
chr7 110148678-110148795	IMMP2L (3.43)
chr7 117636080-117636194	CTTNBP2 (0)
chr7 123193286-123193436	NDUFA5 (4.78)
chr7 124996202-124996342	POT1 (4.2)
chr7 125141036-125141138	POT1 (4.2)
chr7 127769423-127769584	SND1 (7.91)
chr7 128303360-128303475	FLJ45340 (0)
chr7 130016485-130016596	CPA1 (0)
chr7 131007219-131007329	MKLN1 (4.05)
chr7 133550390-133550494	EXOC4 (5.33)

chr7 134847227-134847375	C7orf49 (5.7)
chr7 140895581-140895697	LOC100507421 (0)
chr7 148565395-148565529	EZH2 (5.59)
chr7 151578399-151578546	LOC100505483 (0)
chr7 158172537-158172647	PTPRN2 (0.03)
chr8 1710764-1711040	CLN8 (1.79)
chr8 15199108-15199214	SGCZ (0)
chr8 15274672-15274800	TUSC3 (0)
chr8 16314453-16314604	MSR1 (0.16)
chr8 16694396-16694511	FGF20 (0)
chr8 19115703-19115818	LOC100128993 (0)
chr8 29605536-29605675	LOC286135 (0)
chr8 31064945-31065058	WRN (2.97)
chr8 31683928-31684042	NRG1 (0)
chr8 39462111-39462224	ADAM18 (0)
chr8 49430503-49430784	EFCAB1 (0.11)
chr8 66893305-66893454	DNAJC5B (0.13)
chr8 68684387-68684563	CPA6 (0)
chr8 71452838-71452959	TRAM1 (6.97)
chr8 74288472-74288585	LOC100128126 (0)
chr8 96135058-96135207	PLEKHF2 (4.83)
chr8 99417949-99418058	KCNS2 (0.1)
chr8 111494282-111494381	KCNV1 (0)
chr8 114769767-114769916	CSMD3 (0)
chr8 117816611-117816765	UTP23 (4.07)
chr8 117864193-117864296	RAD21-AS1 (0.49)
chr8 117912409-117912499	RAD21-AS1 (0.49)
chr8 118275725-118275836	SLC30A8 (0)
chr8 120955552-120955704	DEPTOR (4.92)
chr8 121925287-121925392	SNTB1 (4.06)
chr8 123583734-123583834	ZHX2 (4.14)
chr8 126545237-126545345	TRIB1 (3.83)
chr8 134392073-134392171	ST3GAL1 (0.72)
chr9 2252773-2252886	SMARCA2 (4.07)
chr9 6313198-6313301	TPD52L3 (0.07)
chr9 16469390-16469505	BNC2 (0)
chr9 16917485-16917595	BNC2 (0)
chr9 22058811-22058924	CDKN2B-AS (0)
chr9 22197110-22197281	CDKN2B-AS (0)
chr9 26752846-26752989	C9orf82 (0)
chr9 30139762-30139893	MIR873 (0)
chr9 31578578-31578692	ACO1 (3.63)
chr9 41902191-41902327	MGC21881 (0)
chr9 44454549-44454685	CNTNAP3B (0)
chr9 46896404-46896540	KGFLP1 (0.59)
chr9 66533924-66534061	LOC442421 (0)
chr9 80056215-80056360	GNA14 (0.07)
chr9 83714060-83714175	TLE1 (0.36)
chr9 86720349-86720435	RMI1 (3.06)
chr9 92470396-92470511	UNQ6494 (0)
chr9 93759291-93759408	LOC100129316 (0)
chr9 100153136-100153285	BDAG1 (0)
chr9 101998630-101998824	SEC61B (8.09)
chr9 103308502-103308629	C9orf30-TMEFF1 (0)
chr9 103801913-103802024	LPPR1 (0)
chr9 116466413-116466516	RGS3 (0.56)

chr9 117099076-117099187	AKNA (4.2)
chr9 124879595-124879703	MIR4478 (0)
chr9 138273294-138273409	LOC100506599 (0)
chrUn_gl000220 114674-114821	RN5-8S1 (0)
chrUn_gl000220 158646-158844	RN5-8S1 (0)
chrX 4218120-4218225	LOC389906 (3.85)
chrX 7029124-7029275	MIR4767 (0)
chrX 11815266-11815371	MSL3 (3.85)
chrX 13321435-13321549	ATXN3L (0)
chrX 15498258-15498400	PIR-FIGF (0)
chrX 18889465-18889615	LOC100132163 (0)
chrX 26928849-26928995	VENTXP1 (0)
chrX 34486610-34486710	TMEM47 (0)
chrX 34587108-34587249	TMEM47 (0)
chrX 41973802-41973907	CASK (2.28)
chrX 47003999-47004121	RBM10 (5.15)
chrX 48760209-48760352	PQBP1 (4.71)
chrX 68019189-68019304	EFNB1 (1.78)

chrX 78887112-78887219	ITM2A (0)
chrX 79478694-79478832	FAM46D (0)
chrX 82925572-82925687	POU3F4 (0)
chrX 91733094-91733196	PCDH11X (0)
chrX 92094170-92094310	PCDH11X (0)
chrX 105349141-105349244	MUM1L1 (0)
chrX 113028105-113028258	HTR2C (0)
chrX 118425457-118425565	PGRMC1 (5.69)
chrX 124347004-124347090	LOC100129520 (0)
chrX 131759358-131759456	HS6ST2 (0)
chrX 135056058-135056174	SLC9A6 (2.87)
chrX 141171314-141171413	MAGEC2 (0)
chrX 145722669-145722826	CXorf51A (0)
chrX 150023844-150023980	CD99L2 (2.2)
chrX 152370302-152370449	MAGEA1 (0)
chrX 153060016-153060227	SSR4 (7.72)
chrX 153195418-153195663	NAA10 (6.33)

Supplementary Data Table 3. Closest genes to active LTRs detected by RACE-Seq in **L428** cells. Gene expression values (FPKM) from RNA-Seq are also shown for each gene.

RACE Peak	Gene (FPKM)
chr1 2123459-2123567	PRKCZ (2.15)
chr1 6981705-6981820	CAMTA1 (5.78)
chr1 8225875-8225988	ERRFI1 (0.1)
chr1 16160878-16161019	SPEN (3.89)
chr1 21719060-21719234	ECE1 (2.55)
chr1 24526768-24526885	IL28RA (0)
chr1 24814727-24814845	RCAN3 (1.7)
chr1 30874495-30874611	MATN1 (0.05)
chr1 40371047-40371152	MYCL1 (0)
chr1 55404738-55404853	TMEM61 (0)
chr1 56337634-56337767	PPAP2B (0.03)
chr1 58368619-58368736	DAB1 (0)
chr1 62386028-62386169	INADL (0.94)
chr1 64634984-64635099	ROR1 (4.11)
chr1 64957070-64957169	CACHD1 (0.17)
chr1 66594409-66594527	PDE4B (3.21)
chr1 78588148-78588249	GIPC2 (0.06)
chr1 81428611-81428730	LPHN2 (0)
chr1 82218886-82219001	LPHN2 (0)
chr1 85191690-85191840	SSX2IP (3.88)
chr1 85429341-85429472	MCOLN2 (4.58)
chr1 88966312-88966418	PKN2 (3.72)
chr1 97721041-97721156	DPYD (4.44)
chr1 97722055-97722174	DPYD (4.44)
chr1 101455619-101455876	DPH5 (4.33)
chr1 102479013-102479131	OLFM3 (0)
chr1 109978215-109978330	PSMA5 (5.97)
chr1 160597763-160597881	SLAMF1 (0.89)
chr1 164446861-164446976	PBX1 (0.04)
chr1 168189759-168189895	SFT2D2 (4.93)
chr1 168731038-168731179	MGC4473 (0)
chr1 170803684-170803801	PRRX1 (0)
chr1 171409389-171409529	PRRC2C (5.82)
chr1 174173947-174174072	RABGAP1L (6.96)
chr1 177701431-177701553	SEC16B (0)
chr1 180528078-180528190	ACBD6 (6.52)
chr1 181419165-181419282	CACNA1E (0.01)
chr1 181511712-181511831	CACNA1E (0.01)
chr1 182202214-182202321	GLUL (4.3)
chr1 183281078-183281209	NMNAT2 (0)
chr1 186213435-186213547	MIR548F1 (0)
chr1 188701637-188701776	FAM5C (0)
chr1 200247978-200248120	C1orf98 (0)
chr1 204011683-204011809	LINC00303 (0)
chr1 216442638-216442753	ESRRG (0.02)
chr1 219114803-219114942	LOC643723 (0)
chr1 222360694-222360813	HHIPL2 (0)
chr1 222401271-222401389	HHIPL2 (0)
chr1 223461596-223461706	SUSD4 (0)
chr1 224847884-224848002	CNIH3 (0.1)
chr1 234081473-234081613	SLC35F3 (0)
chr1 234831807-234831926	LOC100506810 (0)
chr1 237780090-237780208	LOC100130331 (0)
chr1 239712269-239712400	CHRM3 (1.04)
chr1 241450704-241450823	RGS7 (0)
chr1 243269481-243269591	LOC731275 (0)
chr1 244472679-244472797	C1orf100 (0)
chr1 244830097-244830267	PPPDE1 (0)
chr1 245353128-245353246	KIF26B (0.02)
chr10 4915374-4915511	tAKR (0)
chr10 7144565-7144674	SFMBT2 (3.62)
chr10 8383285-8383391	GATA3 (0.03)
chr10 9724534-9724669	SFTA1P (0)
chr10 9966871-9966989	SFTA1P (0)
chr10 10066135-10066250	SFTA1P (0)
chr10 13478510-13478616	BEND7 (0)
chr10 15548643-15548758	ITGA8 (0)
chr10 15981514-15981619	FAM188A (3.67)
chr10 17568234-17568376	PTPLA (0)
chr10 25302007-25302101	ENKUR (0)
chr10 25481137-25481239	GPR158 (0)
chr10 26683181-26683284	APBB1IP (5.36)
chr10 27660519-27660638	PTCHD3 (0.04)
chr10 28691711-28691823	LOC220906 (0)
chr10 34102187-34102325	LOC100505583 (0)
chr10 47109739-47109879	LOC643650 (0)
chr10 49498137-49498283	FRMPD2 (0.02)
chr10 56710195-56710333	PCDH15 (0)
chr10 56965585-56965750	MTRNR2L5 (0)
chr10 60256212-60256328	BICC1 (0)
chr10 62416791-62416948	ANK3 (0.04)
chr10 68275394-68275501	CTNNA3 (0.01)
chr10 72672059-72672165	PCBD1 (2.67)
chr10 75006859-75007005	C10orf103 (0)
chr10 79468093-79468206	KCNMA1 (0)
chr10 81629600-81629740	LOC100288974 (0)
chr10 81955486-81955599	ANXA11 (4.47)
chr10 85909886-85909897	C10orf99 (0)
chr10 90305149-90305257	RNLS (2.12)
chr10 90573504-90573667	LIPM (0)
chr10 90597573-90597692	LIPM (0)
chr10 92995502-92995615	LOC100188947 (0)
chr10 95579267-95579383	LGI1 (0)
chr10 107729824-107729933	SORCS1 (0.02)
chr10 108042658-108042795	SORCS1 (0.02)
chr10 108792838-108792954	SORCS3 (0)
chr10 111568710-111568820	XPNPEP1 (6.03)
chr10 112373134-112373249	SMC3 (5.73)
chr10 118178152-118178271	PNLIPRP3 (0)

chr10 125305963-125306095	GPR26 (0)	**chr12 65915292-65915398**	MSRB3 (0.02)
chr10 125820855-125820996	CHST15 (4.31)	**chr12 66414343-66414461**	HMGA2 (0)
chr10 129754048-129754157	PTPRE (2.28)	**chr12 68066305-68066421**	DYRK2 (3.99)
chr11 1285717-1285826	MUC5B (0.02)	**chr12 68835694-68836256**	MDM1 (4.16)
chr11 3206233-3206339	OSBPL5 (0)	**chr12 73345011-73345124**	TRHDE (0.01)
chr11 5526781-5526923	OR51B5 (0)	**chr12 88543918-88544036**	CEP290 (2.76)
chr11 9189859-9189972	DENND5A (2.92)	**chr12 88721976-88722114**	TMTC3 (2.4)
chr11 10458289-10458407	AMPD3 (0.2)	**chr12 89445751-89445867**	LOC728084 (0)
chr11 11259654-11259766	GALNTL4 (0)	**chr12 91058848-91058964**	C12orf37 (0)
chr11 13010988-13011127	RASSF10 (0)	**chr12 91351941-91352058**	C12orf12 (0)
chr11 17890250-17890367	SERGEF (4.2)	**chr12 94202036-94202153**	CRADD (3.46)
chr11 19548781-19548901	NAV2 (0.14)	**chr12 97286221-97286340**	NEDD1 (4.33)
chr11 26961601-26961716	FIBIN (0)	**chr12 99386279-99386388**	ANKS1B (1.22)
chr11 27911882-27912101	KIF18A (4.6)	**chr12 102685554-102685664**	PMCH (2.52)
chr11 33753564-33753683	FBXO3 (3.99)	**chr12 104864054-104864169**	CHST11 (0.54)
chr11 39491593-39491729	LRRC4C (0)	**chr12 114438145-114438258**	RBM19 (4.8)
chr11 44098364-44098482	ACCS (0)	**chr12 115345771-115345896**	TBX3 (0)
chr11 59263853-59263971	OR4D11 (0)	**chr12 118419389-118419530**	KSR2 (0.01)
chr11 63059596-63059714	SLC22A10 (0)	**chr12 118419874-118420023**	KSR2 (0.01)
chr11 70160986-70161106	PPFIA1 (3.79)	**chr12 127254724-127254865**	LOC100507206 (0)
chr11 71881679-71881794	FOLR1 (0)	**chr12 127468378-127468493**	LOC440117 (0)
chr11 73685709-73686074	DNAJB13 (0)	**chr12 128452964-128453052**	FLJ37505 (0)
chr11 74168900-74169038	LIPT2 (3.27)	**chr12 128891783-128891883**	TMEM132C (0)
chr11 76647129-76647264	ACER3 (2.32)	**chr12 129110028-129110143**	TMEM132C (0)
chr11 77554858-77554995	RSF1 (3.75)	**chr12 129594176-129594318**	TMEM132D (0)
chr11 78905144-78905264	ODZ4 (0)	**chr13 19128876-19128992**	ANKRD20A9P (0.39)
chr11 85494206-85494312	SYTL2 (3.76)	**chr13 21665974-21666095**	LATS2 (0)
chr11 88268080-88268207	GRM5 (0)	**chr13 27766743-27766860**	USP12 (2.69)
chr11 92332386-92332500	FAT3 (0)	**chr13 35314331-35314450**	NBEA (1.03)
chr11 97252984-97253095	JRKL (1.84)	**chr13 46234667-46234783**	SPERT (0)
chr11 98490121-98490255	CNTN5 (0)	**chr13 49259270-49259384**	CYSLTR2 (0)
chr11 102304547-102304688	BIRC2 (4.44)	**chr13 54539414-54539553**	MIR1297 (0)
chr11 110214728-110214855	RDX (4.15)	**chr13 56917553-56917689**	PRR20C (0)
chr11 112367295-112367397	C11orf34 (0)	**chr13 63631285-63631445**	OR7E156P (0)
chr11 115797100-115797226	LOC283143 (0)	**chr13 64882038-64882141**	OR7E156P (0)
chr11 119773504-119773623	PVRL1 (0.9)	**chr13 65606632-65606783**	PCDH9 (4.79)
chr11 126517954-126518150	KIRREL3 (0.02)	**chr13 69229703-69229875**	LOC338862 (0)
chr11 127264558-127264695	KIRREL3-AS3 (0)	**chr13 69251436-69251586**	LOC338862 (0)
chr11 128239961-128240254	ETS1 (6.81)	**chr13 71010946-71011072**	ATXN8OS (0)
chr12 9804058-9804165	CLEC2D (5.18)	**chr13 72893893-72893996**	MZT1 (4.9)
chr12 13420858-13420970	EMP1 (0.06)	**chr13 72916013-72916138**	MZT1 (4.9)
chr12 14432832-14432962	ATF7IP (4.78)	**chr13 72976580-72976794**	MZT1 (4.9)
chr12 15659435-15659544	PTPRO (0.17)	**chr13 78003763-78003882**	MYCBP2 (5.38)
chr12 21810682-21810799	LDHB (10.32)	**chr13 78498975-78499087**	EDNRB (0)
chr12 22567635-22567747	KIAA0528 (0)	**chr13 87346177-87346310**	MIR4500HG (0)
chr12 25108459-25108609	BCAT1 (6.86)	**chr13 87778357-87778562**	MIR4500HG (0)
chr12 25445983-25446288	KRAS (3.35)	**chr13 89896538-89896645**	MIR622 (0)
chr12 30020945-30021053	TMTC1 (0)	**chr13 96132271-96132389**	CLDN10 (0)
chr12 39279665-39279770	KIF21A (0)	**chr13 98395866-98395993**	IPO5 (7.68)
chr12 41706344-41706462	PDZRN4 (0)	**chr13 99772200-99772319**	DOCK9 (2.85)
chr12 50505958-50506067	C12orf62 (0)	**chr13 100104204-100104324**	MIR548AN (0)
chr12 59427482-59427600	LRIG3 (0)	**chr13 100113234-100113379**	TM9SF2 (6.15)
chr12 59880356-59880492	SLC16A7 (2.4)	**chr13 102161444-102161583**	NALCN (0)
chr12 60826109-60826215	SLC16A7 (2.4)	**chr13 103752582-103752686**	SLC10A2 (1.27)
chr12 63402435-63402554	PPM1H (0.77)	**chr13 104440020-104440162**	SLC10A2 (1.27)
chr12 64428040-64428146	SRGAP1 (0.05)	**chr14 19680927-19681064**	POTEG (0)

chr14 19894320-19894456	POTEM (0)	**chr16 21314303-21314411**	RUNDC2B (0)
chr14 21134732-21135029	ANG (0.29)	**chr16 26644119-26644238**	C16orf82 (0)
chr14 27075087-27075192	NOVA1 (0)	**chr16 27258580-27258734**	NSMCE1 (4.78)
chr14 29558644-29558776	C14orf23 (0)	**chr16 31123626-31123821**	KAT8 (5.11)
chr14 29593676-29593791	MIR548AI (0)	**chr16 54907984-54908094**	CRNDE (5.09)
chr14 30348581-30348700	PRKD1 (0)	**chr16 58455194-58455341**	GINS3 (5.28)
chr14 35829097-35829203	NFKBIA (4.48)	**chr16 60118538-60118653**	LOC644649 (0)
chr14 37469986-37470112	SLC25A21 (0)	**chr16 67403545-67403663**	TPPP3 (0)
chr14 37594005-37594253	SLC25A21 (0)	**chr16 76268957-76269075**	CNTNAP4 (0.07)
chr14 41907146-41907256	LRFN5 (0)	**chr16 76269447-76269554**	CNTNAP4 (0.07)
chr14 43687715-43687829	FSCB (0)	**chr16 80926290-80926626**	C16orf61 (0)
chr14 48203570-48203676	MIR548Y (0)	**chr16 86380942-86381060**	LOC732275 (0)
chr14 50053303-50053593	RPS29 (10.33)	**chr16 88876084-88876217**	APRT (7.52)
chr14 50873090-50873200	CDKL1 (3.04)	**chr17 5522640-5522784**	NLRP1 (1.44)
chr14 51651869-51651986	TMX1 (4.07)	**chr17 25597719-25597860**	MIR4522 (0)
chr14 51960189-51960305	FRMD6 (0.34)	**chr17 27717450-27717563**	MIR4523 (0)
chr14 57445260-57445415	OTX2OS1 (0)	**chr17 30771513-30771618**	PSMD11 (5)
chr14 59502600-59502715	DAAM1 (2.68)	**chr17 31279715-31279834**	TMEM98 (0)
chr14 60677143-60677281	PPM1A (3.5)	**chr17 38946998-38947105**	KRT28 (0)
chr14 69377536-69377674	ACTN1 (0)	**chr17 42022285-42022405**	PPY (0)
chr14 71411543-71411685	PCNX (2.51)	**chr17 49498006-49498128**	UTP18 (6.18)
chr14 76012924-76013197	FLVCR2 (0.18)	**chr17 50681008-50681121**	LOC100506650 (0)
chr14 89744371-89744627	FOXN3 (2.63)	**chr17 52433895-52434010**	KIF2B (0)
chr14 90410505-90410619	TDP1 (4.47)	**chr17 64410429-64410541**	PRKCA (2.28)
chr14 92630310-92630493	CPSF2 (6)	**chr17 64934372-64934491**	CACNG4 (0.02)
chr14 93336858-93336998	GOLGA5 (4.05)	**chr17 66823157-66823270**	ABCA8 (0)
chr14 93338639-93338779	GOLGA5 (4.05)	**chr17 69227567-69227700**	SOX9 (0)
chr14 93339789-93339929	GOLGA5 (4.05)	**chr17 69617072-69617207**	SOX9 (0)
chr14 94114160-94114286	PRIMA1 (0)	**chr17 70016632-70016770**	SOX9 (0)
chr14 97903730-97903934	LOC100129345 (0)	**chr17 71146860-71147015**	SSTR2 (0.29)
chr14 100800129-100800299	SLC25A47 (0)	**chr17 71999238-71999356**	LINC00469 (0)
chr15 23188003-23188295	WHAMMP3 (1.49)	**chr17 73754260-73754605**	GALK1 (4.9)
chr15 34880551-34880664	GOLGA8B (3.08)	**chr17 74560620-74560714**	ST6GALNAC2 (0.82)
chr15 38339671-38339786	TMCO5A (0)	**chr17 79335996-79336139**	TMEM105 (0)
chr15 39527801-39527919	C15orf54 (0)	**chr18 10890477-10890590**	PIEZO2 (0.52)
chr15 42788980-42789116	ZFP106 (0)	**chr18 19886298-19886390**	GATA6 (0)
chr15 45103608-45103713	TRIM69 (0.27)	**chr18 20452935-20453060**	RBBP8 (6.03)
chr15 47835590-47835708	SEMA6D (0)	**chr18 22537531-22537649**	ZNF521 (0.02)
chr15 49907724-49907827	C15orf33 (0)	**chr18 30136366-30136483**	WBP11P1 (0)
chr15 53016821-53016939	ONECUT1 (0)	**chr18 32485069-32485203**	DTNA (0)
chr15 58113629-58113748	GCOM1 (0)	**chr18 32486999-32487105**	DTNA (0)
chr15 66721530-66721655	MAP2K1 (4.28)	**chr18 33069290-33069426**	MIR3975 (0)
chr15 69745132-69745250	KIF23 (5.18)	**chr18 42251832-42252031**	SETBP1 (1.8)
chr15 74756735-74756874	LOC440288 (0)	**chr18 44812020-44812208**	IER3IP1 (5.2)
chr15 75230266-75230387	RPP25 (4.2)	**chr18 45721780-45721922**	ZBTB7C (0.02)
chr15 81096283-81096421	KIAA1199 (0)	**chr18 47982707-47982825**	SKA1 (5.6)
chr15 87581914-87582054	AGBL1 (0)	**chr18 50287463-50287600**	DCC (0)
chr15 91284017-91284135	BLM (5.39)	**chr18 55636098-55636225**	NEDD4L (2.21)
chr15 95965601-95965714	LOC145820 (0)	**chr18 55664170-55664261**	NEDD4L (2.21)
chr15 96537229-96537371	NR2F2 (0)	**chr18 56447356-56447466**	MALT1 (5.1)
chr16 6147114-6147232	RBFOX1 (0)	**chr18 57161150-57161265**	CCBE1 (1.79)
chr16 13354736-13354856	SHISA9 (2.65)	**chr18 58210297-58210409**	MC4R (0)
chr16 13431183-13431302	SHISA9 (2.65)	**chr18 59990762-59990875**	TNFRSF11A (0)
chr16 17119508-17119657	XYLT1 (2.68)	**chr18 62267760-62267873**	LOC284294 (0)
chr16 20035572-20035707	GPR139 (0)	**chr18 62453081-62453192**	LOC284294 (0)
chr16 21127733-21127858	DNAH3 (0.05)	**chr18 62658732-62658836**	LOC284294 (0)

chr18 63016212-63016315	CDH7 (0)	**chr2 84790162-84790280**	DNAH6 (0.03)
chr18 65051482-65051594	DSEL (0)	**chr2 99229885-99230002**	COA5 (3.74)
chr18 68603722-68603833	LOC100505776 (0)	**chr2 104629690-104629825**	LOC100287010 (0.1)
chr18 68919816-68919920	LOC100505776 (0)	**chr2 110294435-110294550**	40422 (0)
chr18 68975425-68975567	LOC100505776 (0)	**chr2 110451130-110451245**	ANKRD57 (0)
chr18 69552667-69552809	LOC100505776 (0)	**chr2 110971196-110971314**	NPHP1 (0.82)
chr18 69776943-69777067	CBLN2 (0)	**chr2 111881290-111881403**	ACOXL (0.04)
chr18 70147093-70147206	CBLN2 (0)	**chr2 114448583-114448781**	SLC35F5 (3.98)
chr18 73447895-73448004	C18orf62 (0)	**chr2 116514940-116515055**	DPP10 (0)
chr18 75705580-75705715	GALR1 (0)	**chr2 118480741-118480859**	DDX18 (6.02)
chr19 2477781-2477881	GADD45B (4.32)	**chr2 121013791-121014247**	TMEM185B (2.2)
chr19 6381159-6381451	GTF2F1 (5.57)	**chr2 123811158-123811262**	CNTNAP5 (0)
chr19 7696347-7696468	PCP2 (0.82)	**chr2 128586999-128587114**	POLR2D (5.04)
chr19 13864465-13864577	MRI1 (4.06)	**chr2 132737354-132737467**	ANKRD30BL (0)
chr19 16308721-16308840	AP1M1 (4.86)	**chr2 148549202-148549312**	ACVR2A (1.46)
chr19 23887328-23887468	ZNF675 (3.88)	**chr2 152470238-152470386**	NEB (0.11)
chr19 28620111-28620219	LOC148189 (0)	**chr2 155061992-155062392**	GALNT13 (0)
chr19 34046697-34046823	PEPD (4.19)	**chr2 155342895-155343013**	GALNT13 (0)
chr19 35090318-35090437	SCGBL (0)	**chr2 157881105-157881243**	GALNT5 (0)
chr19 46743645-46743762	IGFL1 (0)	**chr2 160266972-160267101**	BAZ2B (0.31)
chr19 50363849-50363937	PNKP (4.08)	**chr2 162419054-162419191**	SLC4A10 (0.44)
chr2 1318520-1318638	SNTG2 (0.05)	**chr2 163969119-163969230**	KCNH7 (0.02)
chr2 3119038-3119156	TSSC1 (5.21)	**chr2 164518262-164518402**	KCNH7 (0.02)
chr2 4091294-4091413	LOC100505964 (0)	**chr2 166178046-166178179**	SCN2A (0.02)
chr2 6101997-6102123	LOC400940 (0)	**chr2 169280671-169280783**	CERS6 (0)
chr2 7860336-7860496	LOC339788 (0)	**chr2 176249775-176249893**	ATP5G3 (6.29)
chr2 9255032-9255146	ASAP2 (0.2)	**chr2 176251669-176251788**	ATP5G3 (6.29)
chr2 10266072-10266268	C2orf48 (1.6)	**chr2 177293847-177293995**	MTX2 (4.43)
chr2 11037116-11037234	KCNF1 (0)	**chr2 182030652-182030770**	UBE2E3 (5.17)
chr2 11719997-11720229	MIR4429 (0)	**chr2 187079749-187079858**	ZC3H15 (6.66)
chr2 12007290-12007409	MIR4262 (0)	**chr2 188969733-188969852**	GULP1 (0)
chr2 14131243-14131388	FAM84A (0)	**chr2 191573345-191573534**	NAB1 (3.54)
chr2 15444372-15444476	NBAS (4.31)	**chr2 191860096-191860214**	GLS (5.47)
chr2 15974652-15974767	MYCNOS (0)	**chr2 196972875-196973015**	STK17B (5.62)
chr2 16048806-16048921	MYCNOS (0)	**chr2 198220811-198220929**	SF3B1 (7.92)
chr2 16909369-16909486	FAM49A (1.22)	**chr2 204991755-204991874**	ICOS (0)
chr2 16995366-16995481	FAM49A (1.22)	**chr2 215674336-215674434**	ABCA12 (0.03)
chr2 17253370-17253484	FAM49A (1.22)	**chr2 220553182-220553299**	SLC4A3 (0)
chr2 20263076-20263355	LAPTM4A (2.92)	**chr2 226802918-226803037**	KIAA1486 (0)
chr2 21385745-21385883	APOB (0)	**chr2 229227273-229227381**	SPHKAP (0)
chr2 27383113-27383218	TCF23 (0)	**chr2 235095748-235095863**	SPP2 (0)
chr2 30554542-30554660	LBH (3.31)	**chr2 238125828-238125945**	COL6A3 (0.04)
chr2 33931865-33932035	MYADML (0)	**chr2 241647948-241648061**	KIF1A (0)
chr2 36275760-36275875	LOC100288911 (0)	**chr20 1772427-1772544**	LOC100289473 (0)
chr2 38035738-38035851	FAM82A1 (0)	**chr20 6093809-6093926**	LRRN4 (0.25)
chr2 38391430-38391548	CYP1B1-AS1 (0)	**chr20 8354899-8355004**	PLCB1 (0)
chr2 44523998-44524149	PREPL (4.44)	**chr20 15932709-15932844**	MACROD2 (0)
chr2 45337143-45337255	UNQ6975 (0)	**chr20 15952051-15952169**	MACROD2 (0)
chr2 49305302-49305417	FSHR (0)	**chr20 23478253-23478371**	CST8 (0)
chr2 49398162-49398432	FSHR (0)	**chr20 33650945-33651063**	EDEM2 (3.64)
chr2 53155995-53156176	ASB3 (4.4)	**chr20 34547338-34547447**	SCAND1 (4.84)
chr2 56503657-56503797	CCDC85A (1.56)	**chr20 34620302-34620412**	C20orf152 (0)
chr2 64261720-64261823	VPS54 (3.49)	**chr20 39485856-39485960**	MAFB (0)
chr2 65275099-65275295	CEP68 (2.81)	**chr20 42457706-42457821**	TOX2 (0)
chr2 67517083-67517227	LOC644838 (0)	**chr20 43588896-43589039**	LOC100505826 (0)
chr2 79213605-79213749	REG3G (0)	**chr20 48944827-48944926**	LOC284751 (0)

chr20 52716951-52717088	BCAS1 (0.1)	**chr3 77707849-77707963**	ROBO2 (0)
chr20 55453499-55453615	TFAP2C (0)	**chr3 79363572-79363708**	ROBO1 (0)
chr20 56058420-56058538	CTCFL (0.03)	**chr3 80577749-80577860**	ROBO1 (0)
chr20 56343862-56343991	PMEPA1 (1.71)	**chr3 83433624-83433733**	LOC440970 (0)
chr21 14555344-14555460	ANKRD30BP2 (0.03)	**chr3 87306525-87306644**	CHMP2B (3.03)
chr21 15971645-15971782	SAMSN1 (2.33)	**chr3 88082264-88082398**	CGGBP1 (3.95)
chr21 16042596-16042711	SAMSN1 (2.33)	**chr3 88593131-88593238**	C3orf38 (3.2)
chr21 16052023-16052208	SAMSN1 (2.33)	**chr3 98829913-98830055**	DCBLD2 (2.28)
chr21 16118794-16118905	SAMSN1 (2.33)	**chr3 99563633-99563830**	MIR548G (0)
chr21 17296940-17297092	USP25 (4.25)	**chr3 99976565-99976679**	TBC1D23 (3.52)
chr21 17832232-17832327	LINC00478 (0)	**chr3 102001158-102001285**	ZPLD1 (0.02)
chr21 19858084-19858262	TMPRSS15 (1.81)	**chr3 102577929-102578064**	ZPLD1 (0.02)
chr21 20834124-20834257	TMPRSS15 (1.81)	**chr3 103960075-103960190**	MIR548A3 (0)
chr21 20852686-20852805	TMPRSS15 (1.81)	**chr3 106921644-106921873**	LOC100302640 (0)
chr21 21066614-21066754	LINC00320 (0)	**chr3 107843796-107843997**	CD47 (5.04)
chr21 23058563-23058681	LINC00317 (0)	**chr3 112032952-112033107**	CD200 (0.04)
chr21 24750219-24750338	D21S2088E (0)	**chr3 112101271-112101413**	CD200 (0.04)
chr21 25309325-25309445	D21S2088E (0)	**chr3 113806465-113806583**	KIAA1407 (1.2)
chr21 26839466-26839573	LINC00158 (0.24)	**chr3 114500665-114500781**	ZBTB20 (0.74)
chr21 27082030-27082146	ATP5J (6.23)	**chr3 116407860-116408002**	LSAMP-AS3 (0)
chr21 28372021-28372138	ADAMTS5 (0)	**chr3 116688247-116688541**	LSAMP-AS3 (0)
chr21 30447265-30447383	CCT8 (8.09)	**chr3 116911873-116912015**	LSAMP-AS3 (0)
chr21 30638210-30638351	BACH1 (4.22)	**chr3 118528782-118528901**	IGSF11 (0)
chr21 31330501-31330639	GRIK1 (0.05)	**chr3 121698403-121698518**	ILDR1 (0.45)
chr21 31799625-31799734	KRTAP13-3 (0)	**chr3 121734415-121734524**	CD86 (2.35)
chr21 35624520-35624647	LINC00310 (0.56)	**chr3 121753599-121753714**	ILDR1 (0.45)
chr21 36066843-36066971	CLIC6 (0)	**chr3 131606062-131606198**	CPNE4 (0.03)
chr21 47013481-47013766	SLC19A1 (3.96)	**chr3 141754570-141754718**	TFDP2 (5.21)
chr21 47556862-47556980	FTCD (0.09)	**chr3 148677831-148677946**	GYG1 (4.51)
chr21 47673888-47674013	MCM3AP (4.8)	**chr3 149146318-149146425**	TM4SF4 (0)
chr22 16162017-16162153	POTEH (0)	**chr3 150321107-150321230**	SELT (5.69)
chr22 29656693-29656823	EMID1 (0.13)	**chr3 160479537-160479655**	PPM1L (0.38)
chr22 34186149-34186291	LARGE (3.51)	**chr3 163199738-163199856**	LOC647107 (0)
chr22 35458729-35458854	ISX (0.16)	**chr3 166801087-166801205**	ZBBX (0)
chr22 35626942-35627088	HMGXB4 (3.19)	**chr3 166909343-166909455**	ZBBX (0)
chr22 40269979-40270091	ENTHD1 (0)	**chr3 172671982-172672090**	SPATA16 (0)
chr22 45493403-45493516	LOC100506714 (0)	**chr3 174991754-174992117**	NAALADL2 (0.5)
chr22 45513703-45513791	LOC100506714 (0)	**chr3 175815778-175815890**	NAALADL2 (0.5)
chr22 46016641-46016752	FBLN1 (0)	**chr3 177345983-177346118**	TBL1XR1 (5.56)
chr22 46067721-46067847	ATXN10 (5.12)	**chr3 183079268-183079409**	MCF2L2 (0.65)
chr22 47702793-47702898	LOC339685 (0)	**chr3 184175713-184175852**	CHRD (0)
chr3 278939-279042	CNTN6 (0)	**chr3 185743989-185744106**	ETV5 (0.1)
chr3 3168676-3168824	IL5RA (0)	**chr3 188057283-188057411**	LPP (3.88)
chr3 6917115-6917266	LOC100288428 (0)	**chr3 189121164-189121298**	TPRG1 (0.03)
chr3 8226505-8226622	LOC100288428 (0)	**chr3 193563168-193563276**	LOC647323 (0)
chr3 13965278-13965383	LOC100132526 (0)	**chr3 195913909-195914038**	ZDHHC19 (0)
chr3 29254183-29254293	RBMS3 (0.21)	**chr3 196669747-196669837**	LOC152217 (0)
chr3 32972910-32973047	CCR4 (0.05)	**chr4 3915920-3916038**	FAM86EP (2.17)
chr3 43665965-43666101	ANO10 (1.65)	**chr4 5068800-5068919**	STK32B (0)
chr3 52740136-52740275	SPCS1 (5.41)	**chr4 12763410-12763516**	HSP90AB2P (0)
chr3 53231270-53231411	PRKCD (2.03)	**chr4 14809489-14809628**	LOC441009 (0)
chr3 61504851-61505007	PTPRG (0.05)	**chr4 29023038-29023141**	MIR4275 (0)
chr3 62300695-62300814	LOC100506994 (0)	**chr4 33796983-33797100**	ARAP2 (2.72)
chr3 66471313-66471450	LRIG1 (3.69)	**chr4 34272599-34272705**	ARAP2 (2.72)
chr3 72529245-72529486	RYBP (1.48)	**chr4 34540850-34540968**	ARAP2 (2.72)
chr3 72529487-72529575	RYBP (1.48)	**chr4 38552259-38552428**	FLJ13197 (0)

chr4 41914704-41914826	TMEM33 (4.62)	**chr5 41281455-41281568**	C6 (0)
chr4 42465377-42465492	ATP8A1 (4.39)	**chr5 41478921-41479009**	PLCXD3 (0)
chr4 42980951-42981076	GRXCR1 (0)	**chr5 42101089-42101218**	FBXO4 (1.84)
chr4 61911247-61911387	LPHN3 (0)	**chr5 43068939-43069143**	LOC100132356 (2.25)
chr4 63136311-63136458	LPHN3 (0)	**chr5 51202097-51202386**	ISL1 (0)
chr4 73079586-73079726	NPFFR2 (0)	**chr5 55240473-55240730**	IL31RA (0)
chr4 77446327-77446452	SHROOM3 (0.04)	**chr5 59995862-59995954**	DEPDC1B (3.19)
chr4 77791142-77791259	ANKRD56 (0)	**chr5 66118929-66119031**	MAST4 (0.3)
chr4 88654236-88654350	IBSP (0)	**chr5 66786403-66786520**	CD180 (4.65)
chr4 89947689-89947790	FAM13A (0.58)	**chr5 67794749-67794886**	PIK3R1 (3.04)
chr4 94224152-94224259	GRID2 (0.02)	**chr5 68604295-68604424**	CDK7 (4.46)
chr4 95983086-95983264	UNC5C (0)	**chr5 79114595-79114749**	CMYA5 (0.1)
chr4 112723789-112723919	C4orf32 (3.88)	**chr5 83032193-83032311**	HAPLN1 (0)
chr4 117150691-117150799	MIR1973 (0)	**chr5 86938164-86938277**	CCNH (5.95)
chr4 120221493-120221733	USP53 (4.58)	**chr5 91206689-91206797**	LOC100129716 (0)
chr4 122127028-122127156	TNIP3 (0)	**chr5 95382454-95382594**	MIR583 (0)
chr4 123001164-123001273	KIAA1109 (2.91)	**chr5 99708314-99708455**	LOC100133050 (0.14)
chr4 133184029-133184156	PCDH10 (0)	**chr5 100208301-100208443**	ST8SIA4 (3.28)
chr4 134283845-134283963	PCDH10 (0)	**chr5 100217485-100217646**	ST8SIA4 (3.28)
chr4 138978550-138978668	LOC641364 (0)	**chr5 115481489-115481595**	COMMD10 (4.62)
chr4 142579319-142579445	IL15 (0.04)	**chr5 116556682-116556814**	SEMA6A (0.22)
chr4 150430085-150430227	DCLK2 (0.06)	**chr5 119410254-119410366**	PRR16 (0)
chr4 155657151-155657302	LRAT (0)	**chr5 120500390-120500593**	PRR16 (0)
chr4 160900894-160901003	RAPGEF2 (4.8)	**chr5 122037086-122037195**	SNX2 (4.6)
chr4 166771215-166771321	TLL1 (0)	**chr5 129630694-129630813**	CHSY3 (0)
chr4 176499390-176499510	GPM6A (2.49)	**chr5 130802299-130802418**	RAPGEF6 (4.08)
chr4 179189048-179189158	LOC285501 (0)	**chr5 132839582-132839693**	FSTL4 (0)
chr4 185451433-185451663	IRF2 (3.59)	**chr5 139624820-139625001**	C5orf32 (0)
chr4 188077177-188077292	LOC339975 (0)	**chr5 141019041-141019149**	RELL2 (4.29)
chr4 188154413-188154531	LOC339975 (0)	**chr5 142589739-142589852**	ARHGAP26 (1.67)
chr4 189441858-189441973	LOC401164 (0)	**chr5 142970524-142970648**	NR3C1 (4.29)
chr4 189444112-189444228	LOC401164 (0)	**chr5 144779312-144779427**	PRELID2 (0)
chr5 2058890-2059010	IRX4 (0)	**chr5 149472094-149472212**	PDGFRB (0.06)
chr5 2115723-2115842	IRX4 (0)	**chr5 149776150-149776390**	CD74 (9.56)
chr5 2268698-2268846	IRX4 (0)	**chr5 151338389-151338530**	GLRA1 (0.3)
chr5 5006913-5007025	LOC340094 (0)	**chr5 151482694-151482800**	GLRA1 (0.3)
chr5 5039964-5040109	ADAMTS16 (0.05)	**chr5 161353888-161354007**	GABRA1 (0)
chr5 10311294-10311410	CMBL (0.08)	**chr5 162864555-162864671**	NUDCD2 (5.42)
chr5 12079986-12080121	CTNND2 (0)	**chr5 163798356-163798491**	MAT2B (4.63)
chr5 13372760-13372878	DNAH5 (0.04)	**chr5 174478455-174478573**	MIR4634 (0)
chr5 14202804-14202930	TRIO (4.21)	**chr5 176830518-176830635**	F12 (0.42)
chr5 17259222-17259328	LOC285696 (0)	**chr6 3195908-3196040**	TUBB2B (0.04)
chr5 17884805-17884914	LOC401177 (0)	**chr6 6649189-6649447**	LY86-AS1 (1.22)
chr5 18697900-18698125	CDH18 (0.03)	**chr6 9766829-9766988**	TFAP2A (0)
chr5 20922217-20922370	GUSBP1 (4.91)	**chr6 11295205-11295297**	NEDD9 (2.46)
chr5 22618978-22619118	CDH12 (0.25)	**chr6 12677103-12677218**	PHACTR1 (1.67)
chr5 23443620-23443727	PRDM9 (0)	**chr6 16402134-16402269**	ATXN1 (0.35)
chr5 30855646-30855744	CDH6 (0)	**chr6 18522954-18523066**	MIR548A1 (0)
chr5 34280086-34280201	C1QTNF3-AMACR	**chr6 27686879-27686989**	LOC100507173 (0)
chr5 34349766-34349883	C1QTNF3-AMACR	**chr6 31021182-31021328**	HCG22 (0.23)
chr5 35560178-35560293	SPEF2 (0)	**chr6 32430708-32430827**	HLA-DRA (8.53)
chr5 36277925-36278073	RANBP3L (0.07)	**chr6 36321474-36321706**	ETV7 (0)
chr5 38415872-38415978	EGFLAM (0)	**chr6 37717359-37717484**	MDGA1 (0)
chr5 38868631-38868748	RICTOR (3.13)	**chr6 40741031-40741154**	LRFN2 (0)
chr5 40225322-40225466	PTGER4 (1.3)	**chr6 43099194-43099326**	PTK7 (2.47)
chr5 40323412-40323516	PTGER4 (1.3)	**chr6 48855979-48856067**	MUT (3.12)

chr6 54604850-54604968	FAM83B (0)	**chr7 37743963-37744077**	GPR141 (0)
chr6 55495535-55495650	HMGCLL1 (0)	**chr7 45412787-45412920**	RAMP3 (0)
chr6 67778166-67778279	MCART3P (0)	**chr7 46727534-46727632**	TNS3 (3.96)
chr6 71339465-71339584	SMAP1 (4.21)	**chr7 48169599-48169712**	UPP1 (0)
chr6 72372797-72372932	RIMS1 (0)	**chr7 49774986-49775100**	VWC2 (0)
chr6 77054334-77054469	IMPG1 (0)	**chr7 50330068-50330200**	IKZF1 (6.94)
chr6 79307602-79307715	IRAK1BP1 (2.91)	**chr7 53535984-53536102**	FLJ45974 (0)
chr6 81662236-81662480	BCKDHB (4.44)	**chr7 63512747-63512859**	ZNF727 (0)
chr6 81926824-81926942	FAM46A (0.1)	**chr7 68782085-68782205**	AUTS2 (4.46)
chr6 82247253-82247395	FAM46A (0.1)	**chr7 70713087-70713190**	WBSCR17 (0)
chr6 82296403-82296542	FAM46A (0.1)	**chr7 71222105-71222224**	CALN1 (0.01)
chr6 82804931-82805039	IBTK (4.33)	**chr7 81315457-81315565**	HGF (0.55)
chr6 83909290-83909408	RWDD2A (2.54)	**chr7 82523706-82523825**	CACNA2D1 (0.03)
chr6 84815043-84815158	MRAP2 (0)	**chr7 87492680-87492811**	DBF4 (6.02)
chr6 85506212-85506347	TBX18 (0)	**chr7 89020756-89020873**	ZNF804B (0.05)
chr6 85718574-85718681	TBX18 (0)	**chr7 92218103-92218227**	CDK6 (5.13)
chr6 89244212-89244322	RNGTT (4.82)	**chr7 92572171-92572289**	CDK6 (5.13)
chr6 91602251-91602391	MAP3K7 (4.93)	**chr7 93652080-93652244**	BET1 (3.98)
chr6 92685430-92685545	MIR4643 (0)	**chr7 93878939-93879049**	COL1A2 (0)
chr6 94989559-94989674	TSG1 (0)	**chr7 96380361-96380465**	SHFM1 (7.65)
chr6 98376247-98376365	MIR2113 (0)	**chr7 97430391-97430496**	ASNS (7.04)
chr6 101149203-101149338	ASCC3 (5.01)	**chr7 106131454-106131568**	C7orf74 (0)
chr6 102996032-102996144	GRIK2 (0)	**chr7 106145137-106145284**	C7orf74 (0)
chr6 111612239-111612370	REV3L (4.31)	**chr7 106224344-106224456**	C7orf74 (0)
chr6 113777038-113777146	MARCKS (1.09)	**chr7 112709785-112709896**	GPR85 (0.17)
chr6 115034063-115034231	HS3ST5 (0)	**chr7 112947887-112948003**	LOC401397 (0)
chr6 115267892-115267987	HS3ST5 (0)	**chr7 115936062-115936180**	TES (0.19)
chr6 115409554-115409703	FRK (0.54)	**chr7 116907999-116908119**	WNT2 (0)
chr6 117763596-117763709	ROS1 (0)	**chr7 117636023-117636260**	CTTNBP2 (0)
chr6 120326837-120326968	MAN1A1 (6.33)	**chr7 120702879-120703019**	C7orf58 (0)
chr6 125169243-125169359	NKAIN2 (0)	**chr7 121403786-121403901**	PTPRZ1 (0)
chr6 128299543-128299650	PTPRK (0)	**chr7 122851385-122851500**	SLC13A1 (0.13)
chr6 131430514-131430633	AKAP7 (1.04)	**chr7 123193290-123193432**	NDUFA5 (4.78)
chr6 134433538-134433667	HMGA1P7 (0)	**chr7 124996207-124996337**	POT1 (4.2)
chr6 134719719-134719914	LOC154092 (0)	**chr7 125121557-125121684**	POT1 (4.2)
chr6 136546662-136546804	FAM54A (0)	**chr7 125140951-125141158**	POT1 (4.2)
chr6 138128308-138128426	TNFAIP3 (1.46)	**chr7 128761668-128761807**	LOC407835 (0)
chr6 138440545-138440660	PERP (0)	**chr7 130708011-130708131**	MKLN1 (4.05)
chr6 139370341-139370522	C6orf115 (0)	**chr7 144859699-144859858**	TPK1 (1.35)
chr6 140908984-140909113	MIR4465 (0)	**chr7 147082062-147082175**	MIR548I4 (0)
chr6 145790218-145790325	EPM2A (3.12)	**chr7 147084001-147084116**	MIR548I4 (0)
chr6 147173630-147173734	LOC729178 (0)	**chr7 147607689-147607806**	MIR548T (0)
chr6 150580005-150580120	PPP1R14C (0.19)	**chr7 151578411-151578530**	LOC100505483 (0)
chr6 153038695-153038810	VIP (0)	**chr7 154998358-154998485**	INSIG1 (6.36)
chr6 155951973-155952486	NOX3 (0)	**chr7 156162743-156162878**	LOC285889 (0)
chr6 162163307-162163426	PARK2 (0.02)	**chr7 158172523-158172672**	PTPRN2 (0.03)
chr6 164942237-164942364	C6orf118 (0)	**chr8 15199095-15199290**	SGCZ (0)
chr6 166913063-166913180	RPS6KA2 (1.03)	**chr8 15274672-15274804**	TUSC3 (0)
chr7 1561148-1561266	MAFK (4.36)	**chr8 16314447-16314600**	MSR1 (0.16)
chr7 2889073-2889189	GNA12 (3.77)	**chr8 19075653-19075770**	SH2D4A (0)
chr7 12574886-12574976	SCIN (0)	**chr8 20368449-20368564**	LZTS1 (2.67)
chr7 12971259-12971405	ARL4A (3.42)	**chr8 23864217-23864329**	STC1 (0)
chr7 14277920-14278029	DGKB (0)	**chr8 27787781-27787925**	SCARA5 (0)
chr7 30872880-30872995	FAM188B (0.94)	**chr8 37340572-37340732**	ZNF703 (0.03)
chr7 34329154-34329273	AAA1 (0)	**chr8 61402648-61402756**	RAB2A (4.17)
chr7 36031055-36031170	39326 (6.76)	**chr8 63655403-63655521**	NKAIN3 (0)

chr8 67001335-67001448	DNAJC5B (0.13)	**chr9 26752818-26752992**	C9orf82 (0)
chr8 68684429-68684545	CPA6 (0)	**chr9 29724573-29724688**	MIR873 (0)
chr8 69152000-69152115	PREX2 (0)	**chr9 30139758-30139902**	MIR873 (0)
chr8 70216755-70216874	SULF1 (0.01)	**chr9 30601396-30601508**	MIR873 (0)
chr8 71413673-71413788	TRAM1 (6.97)	**chr9 30925182-30925294**	ACO1 (3.63)
chr8 74288470-74288588	LOC100128126 (0)	**chr9 31194211-31194319**	ACO1 (3.63)
chr8 76041799-76041908	CRISPLD1 (0)	**chr9 31578562-31578706**	ACO1 (3.63)
chr8 78434057-78434175	PEX2 (3.58)	**chr9 31724422-31724541**	ACO1 (3.63)
chr8 80659316-80659434	HEY1 (0.34)	**chr9 39630751-39630860**	LOC653501 (0)
chr8 86851851-86852020	REXO1L2P (0)	**chr9 41775783-41775887**	LOC653501 (0)
chr8 90358680-90358861	RIPK2 (3.05)	**chr9 47022872-47022975**	KGFLP1 (0.59)
chr8 90757029-90757140	RIPK2 (3.05)	**chr9 75488945-75489060**	ALDH1A1 (0)
chr8 92453599-92453708	SLC26A7 (0)	**chr9 80438503-80438622**	GNAQ (0)
chr8 94351961-94352080	LOC642924 (0)	**chr9 80847492-80847635**	CEP78 (4.14)
chr8 101344583-101344689	RNF19A (4.02)	**chr9 92470357-92470525**	UNQ6494 (0)
chr8 101977446-101977551	YWHAZ (7.09)	**chr9 93759289-93759408**	LOC100129316 (0)
chr8 105714553-105714672	LRP12 (0)	**chr9 100153163-100153269**	BDAG1 (0)
chr8 105734582-105734694	LRP12 (0)	**chr9 101928114-101928230**	TGFBR1 (3.91)
chr8 107552919-107553034	OXR1 (4.03)	**chr9 101998600-101998832**	SEC61B (8.09)
chr8 112386802-112386941	CSMD3 (0)	**chr9 102000159-102000428**	SEC61B (8.09)
chr8 113767676-113767779	CSMD3 (0)	**chr9 102015355-102015485**	SEC61B (8.09)
chr8 118532934-118533073	EXT1 (2.94)	**chr9 103308510-103308643**	C9orf30-TMEFF1 (0)
chr8 118580736-118580849	MED30 (5.13)	**chr9 103801899-103802040**	LPPR1 (0)
chr8 118776761-118776879	EXT1 (2.94)	**chr9 106120928-106121044**	CYLC2 (0)
chr8 119919115-119919258	TNFRSF11B (0.04)	**chr9 106503290-106503430**	SMC2 (5.15)
chr8 120835697-120835807	DSCC1 (4.64)	**chr9 110420567-110420676**	KLF4 (0)
chr8 123279495-123279605	ZHX2 (4.14)	**chr9 113082632-113082750**	SVEP1 (0)
chr8 123583731-123583836	ZHX2 (4.14)	**chr9 130883425-130883557**	PTGES2 (5.11)
chr8 124685073-124685200	ANXA13 (0)	**chr9 130887602-130887692**	LOC389791 (0)
chr8 124724434-124724546	FAM91A1 (3.56)	**chr9 133272510-133272649**	ASS1 (2.46)
chr8 124840804-124840911	FAM91A1 (3.56)	**chr9 137764379-137764519**	FCN2 (0)
chr8 125755053-125755171	MTSS1 (1.58)	**chr9 138273293-138273412**	LOC100506599 (0)
chr8 126545235-126545345	TRIB1 (3.83)	**chrUn_gl000220 118422-118476**	RN5-8S1 (0)
chr8 127019454-127019594	LOC100130231 (0)	**chrX 3935857-3935966**	LOC389906 (3.85)
chr8 134392054-134392185	ST3GAL1 (0.72)	**chrX 4346910-4347028**	LOC389906 (3.85)
chr8 135340956-135341163	ZFAT (2.56)	**chrX 4352535-4352646**	LOC389906 (3.85)
chr8 135368506-135368621	ZFAT (2.56)	**chrX 5461067-5461174**	NLGN4X (0)
chr8 135784079-135784192	MIR30B (0)	**chrX 5613347-5613489**	NLGN4X (0)
chr8 136876322-136876465	KHDRBS3 (1.19)	**chrX 6458606-6458726**	VCX3A (0)
chr8 138931603-138931722	FAM135B (0)	**chrX 7285579-7285715**	STS (3.93)
chr8 144899823-144899968	PUF60 (6.87)	**chrX 8427959-8428099**	VCX3B (0)
chr9 2252773-2252888	SMARCA2 (4.07)	**chrX 12250420-12250526**	FRMPD4 (0)
chr9 2388633-2388744	FLJ35024 (0)	**chrX 18889481-18889600**	LOC100132163 (0)
chr9 3307049-3307162	RFX3 (2.61)	**chrX 22821880-22821985**	DDX53 (0)
chr9 5580609-5580744	PDCD1LG2 (0.21)	**chrX 24526235-24526344**	PCYT1B (0.16)
chr9 11393643-11393751	PTPRD (0.01)	**chrX 26432707-26432825**	VENTXP1 (0)
chr9 16917483-16917600	BNC2 (0)	**chrX 26928847-26929043**	VENTXP1 (0)
chr9 16983808-16983924	BNC2 (0)	**chrX 31831584-31831702**	DMD (1.07)
chr9 18838059-18838198	ADAMTSL1 (0)	**chrX 34421962-34422104**	TMEM47 (0)
chr9 19037520-19037638	FAM154A (0)	**chrX 34587110-34587245**	TMEM47 (0)
chr9 19252110-19252245	DENND4C (3.58)	**chrX 36466302-36466419**	CXorf30 (0)
chr9 22058798-22058938	CDKN2B-AS (0)	**chrX 36694826-36694969**	CXorf30 (0)
chr9 24481116-24481218	ELAVL2 (0)	**chrX 39293033-39293152**	LOC286442 (0)
chr9 25462603-25462710	TUSC1 (0)	**chrX 45149477-45149608**	CXorf36 (0)
chr9 26157484-26157608	LOC100506422 (0)	**chrX 46182419-46182534**	ZNF673 (0)
chr9 26748174-26748325	C9orf82 (0)	**chrX 69216691-69216832**	EDA (0.89)

chrX 74566456-74566592	ZDHHC15 (0.07)
chrX 78154791-78154906	MIR4328 (0)
chrX 78885074-78885189	ITM2A (0)
chrX 78887097-78887242	ITM2A (0)
chrX 89349733-89349834	TGIF2LX (0)
chrX 92660015-92660123	NAP1L3 (0)
chrX 92797978-92798085	NAP1L3 (0)
chrX 93500187-93500298	FAM133A (0)
chrX 93990317-93990450	FAM133A (0)
chrX 95720766-95720872	LOC643486 (0)
chrX 96883273-96883503	DIAPH2 (4.29)
chrX 97926999-97927109	LOC442459 (0)
chrX 105349136-105349250	MUM1L1 (0)
chrX 111587631-111587770	TRPC5 (0.04)
chrX 113028122-113028240	HTR2C (0)
chrX 118425449-118425567	PGRMC1 (5.69)
chrX 119195832-119195956	RHOXF2B (0.01)
chrX 119308217-119308341	RHOXF2B (0.01)
chrX 119371632-119371742	ZBTB33 (3.72)
chrX 124346974-124347092	LOC100129520 (0)
chrX 135056076-135056164	SLC9A6 (2.87)
chrX 138072661-138072794	FGF13 (0)
chrX 145724218-145724336	CXorf51A (0)
chrX 151963457-151963569	MAGEA3 (0)
chrX 152170887-152171010	PNMA5 (0)
chrX 153060010-153060235	SSR4 (7.72)
chrX 153195440-153195652	NAA10 (6.33)

Supplementary Data Table 4. Closest genes to active LTRs detected by RACE-Seq in **L1236** cells. Gene expression values (FPKM) from RNA-Seq are also shown for each gene.

RACE Peak	Gene (FPKM)
chr1 2123459-2123591	PRKCZ (1.73)
chr1 4036165-4036289	LOC728716 (0)
chr1 6981707-6981818	CAMTA1 (3.87)
chr1 8225877-8225978	ERRFI1 (0.04)
chr1 8245003-8245104	SLC45A1 (0.35)
chr1 18313499-18313645	IGSF21 (0)
chr1 21719062-21719233	ECE1 (3.16)
chr1 24814729-24814843	RCAN3 (2.3)
chr1 55404740-55404851	TMEM61 (0)
chr1 56337652-56337774	PPAP2B (1.79)
chr1 57271444-57271564	C1orf168 (0.02)
chr1 58368618-58368740	DAB1 (0.07)
chr1 58606898-58607023	DAB1 (0.07)
chr1 62386050-62386173	INADL (2.88)
chr1 64634986-64635097	ROR1 (0)
chr1 81428611-81428728	LPHN2 (0)
chr1 82218882-82219062	LPHN2 (0)
chr1 85429340-85429464	MCOLN2 (7.2)
chr1 97721043-97721154	DPYD (4.65)
chr1 97722062-97722172	DPYD (4.65)
chr1 101455649-101455761	DPH5 (5.54)
chr1 103338945-103339054	COL11A1 (0)
chr1 118996484-118996595	SPAG17 (0.29)
chr1 148213989-148214106	PPIAL4D (0)
chr1 148632437-148632554	PPIAL4E (0)
chr1 148794469-148794586	PPIAL4D (0)
chr1 154281789-154281908	AQP10 (0.07)
chr1 157867982-157868106	CD5L (0.11)
chr1 158194472-158194583	CD1A (0)
chr1 160597759-160597885	SLAMF1 (2.1)
chr1 164446863-164446974	PBX1 (1.11)
chr1 168731046-168731171	MGC4473 (0)
chr1 170803680-170803801	PRRX1 (0.07)
chr1 171170877-171170991	FMO2 (0)
chr1 173439099-173439212	LOC100506023
chr1 174173961-174174073	RABGAP1L (6)
chr1 174569167-174569282	RABGAP1L (6)
chr1 179370863-179370988	SOAT1 (2.7)
chr1 182202212-182202325	GLUL (0.15)
chr1 182261512-182261619	GLUL (0.15)
chr1 182505793-182505904	RGSL1 (0)
chr1 183281086-183281207	NMNAT2 (0)
chr1 186213426-186213551	MIR548F1 (0)
chr1 188701651-188701762	FAM5C (0)
chr1 191367830-191367935	LOC440704 (0)
chr1 193446094-193446207	CDC73 (3.09)
chr1 200247878-200248112	C1orf98 (0)
chr1 210249989-210250101	SYT14 (0.05)
chr1 219114811-219114930	LOC643723 (0)
chr1 219180057-219180170	LOC643723 (0)
chr1 220976461-220976585	MOSC1 (0)
chr1 222238026-222238137	DUSP10 (3.15)
chr1 222401273-222401393	HHIPL2 (0.07)
chr1 234081473-234081607	SLC35F3 (0.14)
chr1 234831802-234831924	LOC100506810 (0)
chr1 236326943-236327057	GPR137B (1.35)
chr1 241575474-241575588	RGS7 (0.43)
chr1 243269481-243269593	LOC731275 (0)
chr1 244472681-244472795	C1orf100 (0.14)
chr1 244830105-244830229	PPPDE1 (0)
chr1 245353128-245353244	KIF26B (0.04)
chr10 2941449-2941551	PFKP (6.57)
chr10 3780202-3780343	KLF6 (4.38)
chr10 4915382-4915500	tAKR (0)
chr10 7144562-7144708	SFMBT2 (1.47)
chr10 8383281-8383423	GATA3 (4.91)
chr10 9407412-9407522	GATA3 (4.91)
chr10 9724547-9724659	SFTA1P (0)
chr10 12774465-12774593	CAMK1D (4.54)
chr10 13478503-13478616	BEND7 (0.33)
chr10 14256010-14256124	FRMD4A (0.59)
chr10 15225850-15225976	NMT2 (3.24)
chr10 15548645-15548756	ITGA8 (0)
chr10 17568240-17568368	PTPLA (0)
chr10 25301979-25302099	ENKUR (0)
chr10 28691710-28691821	LOC220906 (0)
chr10 46734506-46734625	BMS1P5 (1.11)
chr10 47109748-47109871	LOC643650 (0)
chr10 48924280-48924395	BMS1P5 (1.11)
chr10 49498151-49498250	FRMPD2 (3.61)
chr10 52464613-52464743	ASAH2B (0)
chr10 56965559-56965710	MTRNR2L5 (0)
chr10 60256209-60256328	BICC1 (0.02)
chr10 62416788-62416929	ANK3 (4.87)
chr10 68275397-68275499	CTNNA3 (1.55)
chr10 77592459-77592570	C10orf11 (0)
chr10 78271873-78271998	KCNMA1 (4.84)
chr10 79468100-79468235	KCNMA1 (4.84)
chr10 81629606-81629734	LOC100288974 (0)
chr10 81955500-81955597	ANXA11 (5.46)
chr10 82388108-82388217	SH2D4B (0.07)
chr10 89145915-89146038	FAM22D (0)
chr10 89404413-89404536	PAPSS2 (0.26)
chr10 89408781-89408892	PAPSS2 (0.26)
chr10 90573498-90573670	LIPM (0.28)
chr10 92765106-92765255	ANKRD1 (0.12)
chr10 103086566-103086693	BTRC (2.52)
chr10 106162985-106163111	CCDC147 (0)
chr10 107729822-107729931	SORCS1 (0.02)
chr10 108042672-108042781	SORCS1 (0.02)
chr10 110437471-110437590	XPNPEP1 (4.99)

chr10 110936986-110937101	XPNPEP1 (4.99)
chr11 1285723-1285824	MUC5B (0.01)
chr11 5526796-5526909	OR51B5 (0)
chr11 9189854-9189976	DENND5A (5.28)
chr11 10458285-10458416	AMPD3 (0.67)
chr11 10830387-10830483	EIF4G2 (7.55)
chr11 13010998-13011113	RASSF10 (0.02)
chr11 17890254-17890377	SERGEF (4.16)
chr11 25613603-25613721	LUZP2 (2.29)
chr11 25613928-25614020	LUZP2 (2.29)
chr11 29181704-29181805	METTL15 (3.43)
chr11 29999917-30000038	KCNA4 (0)
chr11 39491602-39491721	LRRC4C (0)
chr11 44098361-44098486	ACCS (0)
chr11 63059599-63059713	SLC22A10 (0)
chr11 64185778-64185891	RPS6KA4 (3.3)
chr11 73685839-73685925	DNAJB13 (0.25)
chr11 74168914-74169013	LIPT2 (2.05)
chr11 77726216-77726319	KCTD14 (0)
chr11 78905146-78905262	ODZ4 (0)
chr11 84202351-84202468	DLG2 (0.15)
chr11 98458625-98458748	CNTN5 (0.04)
chr11 102217944-102218056	BIRC2 (5.19)
chr11 109187966-109188074	C11orf87 (0)
chr11 110214736-110214853	RDX (4.47)
chr11 113749386-113749511	USP28 (3.67)
chr11 115797093-115797230	LOC283143 (0)
chr11 123538403-123538512	SCN3B (0.02)
chr11 126517976-126518143	KIRREL3 (0)
chr11 128240069-128240263	ETS1 (3.16)
chr12 9804062-9804163	CLEC2D (5.3)
chr12 23286637-23286750	SOX5 (1.54)
chr12 25108425-25108700	BCAT1 (6.22)
chr12 25446058-25446173	KRAS (6.67)
chr12 25890980-25891090	IFLTD1 (0)
chr12 28298986-28299112	CCDC91 (3.26)
chr12 30020947-30021049	TMTC1 (3.96)
chr12 41706302-41706471	PDZRN4 (0)
chr12 56506773-56506865	RPL41 (12.63)
chr12 58713396-58713536	XRCC6BP1 (5.39)
chr12 58748926-58749035	XRCC6BP1 (5.39)
chr12 59427478-59427604	LRIG3 (3.87)
chr12 59880364-59880481	SLC16A7 (0.06)
chr12 60826111-60826213	SLC16A7 (0.06)
chr12 61195153-61195273	FAM19A2 (0.28)
chr12 62603102-62603199	FAM19A2 (0.28)
chr12 63402431-63402553	PPM1H (4.09)
chr12 66414341-66414465	HMGA2 (0.04)
chr12 68835708-68836226	MDM1 (4.04)
chr12 72622999-72623120	LOC283392 (0)
chr12 73345008-73345128	TRHDE (0)
chr12 81171245-81171364	LIN7A (0.05)
chr12 88543924-88544049	CEP290 (2.75)
chr12 88721988-88722100	TMTC3 (3.24)
chr12 88991667-88991781	KITLG (0)
chr12 89445754-89445865	LOC728084 (1.18)

chr12 91058843-91058963	C12orf37 (0)
chr12 92955015-92955121	CLLU1 (0)
chr12 93335096-93335215	EEA1 (3.92)
chr12 94202038-94202151	CRADD (2.35)
chr12 97286223-97286338	NEDD1 (4.69)
chr12 99978457-99978571	ANKS1B (0.7)
chr12 108294566-108294717	LOC728739 (3.24)
chr12 110494591-110494706	C12orf76 (2.82)
chr12 115345770-115345895	TBX3 (1.48)
chr12 116356598-116356719	MED13L (3.72)
chr12 117561639-117561753	FBXO21 (3)
chr12 118419888-118420029	KSR2 (1.9)
chr12 118421849-118422030	KSR2 (1.9)
chr12 125597285-125597404	AACS (4.1)
chr12 125905871-125905985	TMEM132B (0.29)
chr12 127254733-127254855	LOC100507206 (0)
chr12 129594156-129594311	TMEM132D (0.1)
chr12 132087450-132087565	SFSWAP (4.65)
chr13 19128872-19128993	ANKRD20A9P (0.38)
chr13 46234669-46234781	SPERT (0)
chr13 53400049-53400171	MIR759 (0)
chr13 54539430-54539539	MIR1297 (0)
chr13 56917567-56917681	PRR20C (0)
chr13 59883238-59883348	DIAPH3 (3.32)
chr13 63631258-63631454	OR7E156P (0)
chr13 65606636-65606751	PCDH9 (1.79)
chr13 69229719-69229813	LOC338862 (0)
chr13 69251448-69251589	LOC338862 (0)
chr13 71010960-71011058	ATXN8OS (0)
chr13 72916017-72916137	MZT1 (5.09)
chr13 72959912-72960027	MZT1 (5.09)
chr13 77307128-77307242	KCTD12 (1.05)
chr13 78003759-78003886	MYCBP2 (4.15)
chr13 85672105-85672206	SLITRK6 (0.26)
chr13 87346186-87346302	MIR4500HG (0)
chr13 99772196-99772321	DOCK9 (2.28)
chr13 100104200-100104324	MIR548AN (0)
chr13 100113268-100113363	TM9SF2 (5.94)
chr13 102161456-102161577	NALCN (0.02)
chr13 102351896-102351997	FGF14 (0.09)
chr13 107224998-107225098	ARGLU1 (5.34)
chr14 19680937-19681050	POTEG (0)
chr14 19894334-19894445	POTEM (0.07)
chr14 21134748-21135027	ANG (4.45)
chr14 24037066-24037176	JPH4 (0.31)
chr14 27075089-27075194	NOVA1 (0.05)
chr14 30348576-30348704	PRKD1 (0.3)
chr14 37469962-37470105	SLC25A21 (5.6)
chr14 37593913-37594182	SLC25A21 (5.6)
chr14 45260471-45260593	C14orf28 (2.21)
chr14 50053306-50053595	RPS29 (11.22)
chr14 50329289-50329485	NEMF (4.12)
chr14 50403197-50403314	ARF6 (4.98)
chr14 50872978-50873201	CDKL1 (0.96)
chr14 51651869-51651985	TMX1 (4.75)
chr14 51960191-51960303	FRMD6 (2.41)

chr14 56263318-56263431	C14orf34 (0)
chr14 60677146-60677269	PPM1A (4.87)
chr14 67273777-67273886	GPHN (2.51)
chr14 69377551-69377660	ACTN1 (3.05)
chr14 76012920-76013216	FLVCR2 (1.31)
chr14 81723126-81723220	STON2 (1.63)
chr14 82079718-82079804	SEL1L (6.39)
chr14 84620644-84620759	FLRT2 (0.01)
chr14 86559143-86559257	FLRT2 (0.01)
chr14 88317134-88317246	GALC (0)
chr14 89744488-89744634	FOXN3 (5.06)
chr14 90164122-90164238	FOXN3 (5.06)
chr14 90410503-90410617	TDP1 (3.65)
chr14 92630326-92630506	CPSF2 (4.84)
chr14 93336830-93336990	GOLGA5 (4.28)
chr14 93338653-93338765	GOLGA5 (4.28)
chr14 93339800-93339915	GOLGA5 (4.28)
chr14 94114156-94114281	PRIMA1 (0)
chr14 97903734-97903995	LOC100129345
chr14 101879706-101879820	DIO3OS (0)
chr14 103269237-103269341	TRAF3 (3.09)
chr14 105222933-105223083	SIVA1 (5.88)
chr15 23188096-23188261	WHAMMP3 (2.1)
chr15 29002972-29003163	LOC100289656 (0)
chr15 31425037-31425152	TRPM1 (0)
chr15 33502048-33502155	RYR3 (0.04)
chr15 37490008-37490113	MEIS2 (3.03)
chr15 38339672-38339774	TMCO5A (0)
chr15 39527788-39527922	C15orf54 (0)
chr15 39945911-39946028	FSIP1 (3.39)
chr15 42788974-42789114	ZFP106 (0)
chr15 47835592-47835706	SEMA6D (0)
chr15 50024005-50024131	DTWD1 (2.92)
chr15 53016823-53016937	ONECUT1 (0.45)
chr15 56135253-56135372	PRTG (0)
chr15 58113625-58113752	GCOM1 (0)
chr15 58552280-58552406	AQP9 (0.4)
chr15 69745133-69745248	KIF23 (5.61)
chr15 72537233-72537352	PARP6 (4.19)
chr15 75230259-75230449	RPP25 (4.01)
chr15 76721427-76721546	SCAPER (2.68)
chr15 81096287-81096415	KIAA1199 (0)
chr15 87581908-87582049	AGBL1 (0)
chr15 91284013-91284136	BLM (4.14)
chr15 93173103-93173214	FAM174B (0.05)
chr15 94198465-94198575	RGMA (0)
chr15 96537143-96537363	NR2F2 (3.19)
chr16 6494500-6494624	RBFOX1 (0.06)
chr16 16573477-16573596	PKD1P1 (3.02)
chr16 17119472-17119752	XYLT1 (1.62)
chr16 17992390-17992516	XYLT1 (1.62)
chr16 18027413-18027522	XYLT1 (1.62)
chr16 20035580-20035697	GPR139 (0)
chr16 21127741-21127850	DNAH3 (0.08)
chr16 21314298-21314409	RUNDC2B (0)
chr16 27401044-27401167	IL21R (5.69)

chr16 31123626-31123821	KAT8 (4.64)
chr16 54907982-54908107	CRNDE (5.11)
chr16 55106598-55106709	IRX5 (1.68)
chr16 58455214-58455320	GINS3 (3.65)
chr16 60118534-60118657	LOC644649 (0)
chr16 67978248-67978356	SLC12A4 (2.45)
chr16 70912081-70912207	HYDIN (0.04)
chr16 76268955-76269120	CNTNAP4 (0.25)
chr16 80926276-80926554	C16orf61 (0)
chr16 82142909-82143033	HSD17B2 (0.09)
chr16 86632960-86633056	FOXL1 (0)
chr16 88876097-88876202	APRT (7.54)
chr17 5522642-5522776	NLRP1 (4.92)
chr17 7529674-7529768	SHBG (0.14)
chr17 11111585-11111678	SHISA6 (0)
chr17 27717452-27717561	MIR4523 (0)
chr17 38946994-38947101	KRT28 (0)
chr17 42022293-42022391	PPY (0.36)
chr17 50544008-50544134	CA10 (0)
chr17 52433897-52434010	KIF2B (0)
chr17 59322439-59322558	BCAS3 (2.44)
chr17 63698843-63698956	CEP112 (0.51)
chr17 65700593-65700719	PITPNC1 (2.32)
chr17 66823159-66823270	ABCA8 (0.05)
chr17 69227575-69227691	SOX9 (2.1)
chr17 69617068-69617228	SOX9 (2.1)
chr17 70016636-70016762	SOX9 (2.1)
chr17 71146867-71146994	SSTR2 (1.42)
chr17 73754311-73754585	GALK1 (3.9)
chr17 79336003-79336131	TMEM105 (0.04)
chr18 308294-308406	COLEC12 (0.09)
chr18 10890481-10890588	PIEZO2 (0.17)
chr18 22537540-22537653	ZNF521 (1.1)
chr18 29545600-29545708	TRAPPC8 (4.3)
chr18 30136377-30136484	WBP11P1 (0.53)
chr18 42251873-42251999	SETBP1 (1.94)
chr18 44812025-44812303	IER3IP1 (5.43)
chr18 47982679-47982829	SKA1 (4.54)
chr18 50287475-50287586	DCC (0)
chr18 54833285-54833402	BOD1P (0)
chr18 55558331-55558453	ATP8B1 (0.05)
chr18 55636106-55636213	NEDD4L (1.74)
chr18 56447355-56447479	MALT1 (3.14)
chr18 57855825-57855936	MC4R (0)
chr18 59990764-59990875	TNFRSF11A (3.82)
chr18 68975439-68975554	LOC100505776 (0)
chr18 69552680-69552786	LOC100505776 (0)
chr18 69969229-69969344	CBLN2 (0)
chr18 70147095-70147202	CBLN2 (0)
chr18 75705590-75705707	GALR1 (0)
chr19 6381159-6381431	GTF2F1 (5.86)
chr19 7696348-7696449	PCP2 (0.73)
chr19 23887336-23887460	ZNF675 (4.26)
chr19 34046707-34046809	PEPD (4.18)
chr19 35090320-35090435	SCGBL (0)
chr19 46743641-46743764	IGFL1 (0)

chr19 51876984-51877108	NKG7 (1.2)
chr2 184858-184960	SH3YL1 (1.26)
chr2 1318516-1318642	SNTG2 (0)
chr2 3351202-3351309	TTC15 (0)
chr2 6101996-6102129	LOC400940 (0.03)
chr2 6924926-6925041	LINC00487 (0)
chr2 7472125-7472227	LOC100506274
chr2 7860347-7860498	LOC339788 (0)
chr2 8072087-8072203	LINC00299 (0.12)
chr2 9255027-9255148	ASAP2 (0.16)
chr2 11037118-11037229	KCNF1 (0)
chr2 11720095-11720241	MIR4429 (0)
chr2 12007286-12007412	MIR4262 (0)
chr2 12756709-12756824	TRIB2 (0.67)
chr2 14131251-14131427	FAM84A (0)
chr2 16048802-16048921	MYCNOS (0)
chr2 16909365-16909481	FAM49A (1.13)
chr2 16995368-16995479	FAM49A (1.13)
chr2 20062809-20062917	FLJ12334 (0)
chr2 20263062-20263254	LAPTM4A (4.57)
chr2 30554538-30554660	LBH (0.62)
chr2 33931903-33932027	MYADML (0)
chr2 44524007-44524143	PREPL (5.27)
chr2 49305302-49305421	FSHR (0)
chr2 53155997-53156175	ASB3 (4.66)
chr2 54501248-54501356	ACYP2 (2.17)
chr2 54621638-54621750	C2orf73 (0)
chr2 54908781-54908892	SPTBN1 (2.46)
chr2 56503666-56503792	CCDC85A (0.28)
chr2 57858313-57858435	VRK2 (5.75)
chr2 58083384-58083497	VRK2 (5.75)
chr2 64261716-64261825	VPS54 (4.62)
chr2 67517091-67517312	LOC644838 (0)
chr2 68054682-68054785	C1D (4.44)
chr2 72126688-72126802	DYSF (0.08)
chr2 75769217-75769334	FAM176A (0)
chr2 78508399-78508511	SNAR-H (0)
chr2 79213617-79213761	REG3G (0)
chr2 99349273-99349384	MGAT4A (2.64)
chr2 104629696-104629824	LOC100287010
chr2 111969397-111969514	BCL2L11 (1.67)
chr2 114448626-114448738	SLC35F5 (4.73)
chr2 118480742-118480899	DDX18 (6.25)
chr2 121014049-121014233	TMEM185B (4.24)
chr2 123811160-123811257	CNTNAP5 (0)
chr2 128586992-128587118	POLR2D (5.04)
chr2 143874853-143874981	ARHGAP15 (3.76)
chr2 152470255-152470378	NEB (0.2)
chr2 155061994-155062379	GALNT13 (2.13)
chr2 155342889-155343017	GALNT13 (2.13)
chr2 157881120-157881229	GALNT5 (0)
chr2 159145373-159145497	CCDC148 (1.34)
chr2 160266975-160267099	BAZ2B (0.42)
chr2 160740620-160740727	PLA2R1 (0)
chr2 165373612-165373739	COBLL1 (1.08)
chr2 166178052-166178171	SCN2A (1.41)

chr2 169280673-169280781	CERS6 (3.59)
chr2 170543442-170543558	C2orf77 (0)
chr2 176249754-176249912	ATP5G3 (6.54)
chr2 177293855-177293982	MTX2 (6.02)
chr2 181802410-181802508	UBE2E3 (4.44)
chr2 191573353-191573535	NAB1 (4.71)
chr2 192364112-192364224	MYO1B (0.02)
chr2 194838673-194838784	PCGEM1 (0)
chr2 196972871-196973014	STK17B (7.6)
chr2 204991750-204991877	ICOS (0.02)
chr2 205278783-205278892	PARD3B (0.35)
chr2 215986511-215986621	ABCA12 (0.13)
chr2 220855577-220855691	MIR4268 (0)
chr2 224923576-224923691	SERPINE2 (2.86)
chr2 229227270-229227385	SPHKAP (0)
chr2 238363279-238363394	MLPH (0.04)
chr2 241647950-241648059	KIF1A (0.03)
chr20 15932717-15932836	MACROD2 (0.03)
chr20 22529162-22529277	LINC00261 (0)
chr20 23478254-23478369	CST8 (0)
chr20 24876120-24876221	CST7 (1.11)
chr20 33650940-33651063	EDEM2 (4.41)
chr20 34620298-34620411	C20orf152 (0)
chr20 34642431-34642557	LOC647979 (0)
chr20 37260027-37260164	ARHGAP40 (0)
chr20 39485858-39485958	MAFB (0.1)
chr20 42457708-42457819	TOX2 (1.84)
chr20 55350163-55350266	TFAP2C (0)
chr20 55453501-55453612	TFAP2C (0)
chr20 56058422-56058536	CTCFL (0)
chr20 56343871-56343983	PMEPA1 (0.01)
chr20 57395705-57395824	GNAS-AS1 (0.82)
chr20 59732883-59733004	CDH4 (0)
chr21 14555340-14555461	ANKRD30BP2 (0.02)
chr21 15971611-15971779	SAMSN1 (9.19)
chr21 16042592-16042733	SAMSN1 (9.19)
chr21 16052031-16052209	SAMSN1 (9.19)
chr21 17296948-17297063	USP25 (4.82)
chr21 19858107-19858233	TMPRSS15 (0.03)
chr21 20834123-20834249	TMPRSS15 (0.03)
chr21 21066625-21066740	LINC00320 (0)
chr21 23058563-23058679	LINC00317 (0)
chr21 23119246-23119371	LINC00317 (0)
chr21 23406652-23406779	LINC00308 (0)
chr21 25309300-25309447	D21S2088E (3.62)
chr21 28372029-28372135	ADAMTS5 (0)
chr21 30638218-30638377	BACH1 (3.71)
chr21 31330474-31330627	GRIK1 (0)
chr21 31799627-31799731	KRTAP13-3 (0)
chr21 33200816-33200925	HUNK (0)
chr21 35624530-35624645	LINC00310 (0.2)
chr21 47556864-47556978	FTCD (0.1)
chr21 47673874-47674014	MCM3AP (4.81)
chr22 16162031-16162142	POTEH (0)
chr22 34186158-34186282	LARGE (3.76)
chr22 35458727-35458875	ISX (3.83)

chr22 35626948-35627081	HMGXB4 (3.71)
chr22 40269976-40270095	ENTHD1 (0.71)
chr22 45493399-45493520	LOC100506714 (0)
chr22 46016643-46016750	FBLN1 (0.1)
chr22 47702795-47702896	LOC339685 (0)
chr22 50051075-50051178	BRD1 (3.08)
chr3 278941-279041	CNTN6 (0)
chr3 3168709-3168801	IL5RA (5.6)
chr3 6210153-6210251	GRM7 (0.01)
chr3 6917127-6917255	LOC100288428 (0)
chr3 15644425-15644536	BTD (2.59)
chr3 16154022-16154136	GALNTL2 (0)
chr3 16946319-16946417	PLCL2 (3.03)
chr3 32972924-32973034	CCR4 (7.45)
chr3 53231284-53231397	PRKCD (1.99)
chr3 60781690-60781815	PTPRG (3.1)
chr3 61504753-61504995	PTPRG (3.1)
chr3 64311657-64311762	PRICKLE2 (0.1)
chr3 66471325-66471435	LRIG1 (0.2)
chr3 71987292-71987403	PROK2 (0)
chr3 72529259-72529381	RYBP (4.61)
chr3 72529489-72529575	RYBP (4.61)
chr3 72534111-72534225	RYBP (4.61)
chr3 76967441-76967553	ROBO2 (1.33)
chr3 77707841-77707967	ROBO2 (1.33)
chr3 79531584-79531709	ROBO1 (0.16)
chr3 88082274-88082384	CGGBP1 (6.69)
chr3 97777470-97777580	GABRR3 (0)
chr3 98829925-98830047	DCBLD2 (1.85)
chr3 99563646-99563809	MIR548G (0)
chr3 103960071-103960190	MIR548A3 (0)
chr3 106921750-106921879	LOC100302640 (0)
chr3 107843764-107843993	CD47 (6.32)
chr3 112032970-112033067	CD200 (2.74)
chr3 113806566-113806581	KIAA1407 (1.72)
chr3 116911862-116912051	LSAMP-AS3 (0)
chr3 118528780-118528905	IGSF11 (0.15)
chr3 121698367-121698522	ILDR1 (2.13)
chr3 121753598-121753718	ILDR1 (2.13)
chr3 131606070-131606192	CPNE4 (2.09)
chr3 134314094-134314194	KY (0)
chr3 140480167-140480292	TRIM42 (0)
chr3 141754580-141754708	TFDP2 (4.63)
chr3 141760310-141760424	TFDP2 (4.63)
chr3 148677827-148677950	GYG1 (3.77)
chr3 150321108-150321232	SELT (6.24)
chr3 158770763-158770887	IQCJ (0)
chr3 163199739-163199854	LOC647107 (0)
chr3 166330668-166330793	ZBBX (0.48)
chr3 166641930-166642041	ZBBX (0.48)
chr3 166801089-166801203	ZBBX (0.48)
chr3 172296185-172296296	NCEH1 (2.77)
chr3 174991780-174992233	NAALADL2 (0.48)
chr3 175815802-175815927	NAALADL2 (0.48)
chr3 177345991-177346110	TBL1XR1 (5.83)
chr3 180448260-180448377	CCDC39 (0.85)

chr3 183079282-183079397	MCF2L2 (1.58)
chr3 185743984-185744110	ETV5 (0.4)
chr3 188057286-188057387	LPP (2.28)
chr3 189121168-189121295	TPRG1 (0.74)
chr3 195913917-195914024	ZDHHC19 (0)
chr3 196507222-196507331	PAK2 (4.41)
chr4 1250129-1250254	C4orf42 (0)
chr4 5068800-5068923	STK32B (0.1)
chr4 5129635-5129749	STK32B (0.1)
chr4 11687585-11687702	HS3ST1 (3.11)
chr4 14809486-14809611	LOC441009 (0)
chr4 16242612-16242723	FLJ39653 (0)
chr4 19987710-19987833	SLIT2 (0)
chr4 26566767-26566876	TBC1D19 (1.74)
chr4 34540852-34540966	ARAP2 (3.67)
chr4 34898769-34898883	ARAP2 (3.67)
chr4 38552239-38552437	FLJ13197 (0)
chr4 42465352-42465529	ATP8A1 (4.82)
chr4 42980950-42981073	GRXCR1 (0)
chr4 44017714-44017819	KCTD8 (0)
chr4 47206706-47206825	GABRB1 (0)
chr4 57694758-57694873	SPINK2 (1.61)
chr4 59024219-59024332	LOC255130 (0)
chr4 59812611-59812723	LOC255130 (0)
chr4 61911261-61911373	LPHN3 (0)
chr4 65366099-65366209	TECRL (0)
chr4 73079597-73079712	NPFFR2 (0)
chr4 74586700-74586811	IL8 (0)
chr4 77791137-77791263	ANKRD56 (0)
chr4 77870075-77870202	40787 (4.41)
chr4 87261512-87261616	MAPK10 (0)
chr4 88963360-88963470	PKD2 (1.68)
chr4 89947687-89947788	FAM13A (2.74)
chr4 94224142-94224263	GRID2 (0.02)
chr4 94260926-94261028	GRID2 (0.02)
chr4 95983112-95983224	UNC5C (0.81)
chr4 104337489-104337600	TACR3 (0)
chr4 106750617-106750764	NPNT (0)
chr4 112723785-112723921	C4orf32 (2.35)
chr4 113203287-113203406	TIFA (4.56)
chr4 115335685-115335774	UGT8 (1.74)
chr4 118926275-118926374	NDST3 (0)
chr4 120221574-120221725	USP53 (1.96)
chr4 120584048-120584152	PDE5A (2.35)
chr4 123001166-123001265	KIAA1109 (3.57)
chr4 123440513-123440628	IL2 (0)
chr4 133184038-133184148	PCDH10 (0.03)
chr4 134283847-134283961	PCDH10 (0.03)
chr4 141227266-141227388	SCOC (5.5)
chr4 142579321-142579436	IL15 (2.79)
chr4 149680947-149681055	NR3C2 (2.03)
chr4 150430093-150430219	DCLK2 (0.04)
chr4 155657179-155657289	LRAT (0)
chr4 166771211-166771345	TLL1 (0.01)
chr4 168581999-168582116	SPOCK3 (0.15)
chr4 176499363-176499512	GPM6A (0)

chr4 178583094-178583275	LOC285501 (0)
chr4 182732734-182732856	MGC45800 (0)
chr4 184724159-184724276	C4orf41 (0)
chr4 185451429-185451555	IRF2 (4.67)
chr4 188077179-188077290	LOC339975 (0)
chr4 189441860-189441969	LOC401164 (0)
chr5 2058886-2059014	IRX4 (0)
chr5 2115726-2115840	IRX4 (0)
chr5 2268679-2268945	IRX4 (0)
chr5 2332278-2332398	IRX2 (0)
chr5 5006908-5007019	LOC340094 (0)
chr5 5039976-5040077	ADAMTS16 (2.74)
chr5 6123067-6123177	FLJ33360 (0.08)
chr5 10311297-10311408	CMBL (0.09)
chr5 12079994-12080113	CTNND2 (0.01)
chr5 13596631-13596748	DNAH5 (0.12)
chr5 17259224-17259326	LOC285696 (0.09)
chr5 17884801-17884914	LOC401177 (0)
chr5 22618985-22619111	CDH12 (0.05)
chr5 31080627-31080737	CDH6 (0)
chr5 31975319-31975433	PDZD2 (0.33)
chr5 36277939-36278059	RANBP3L (0.62)
chr5 38415868-38415982	EGFLAM (0)
chr5 40027631-40027756	DAB2 (0.01)
chr5 40225285-40225490	PTGER4 (7)
chr5 40532030-40532142	PTGER4 (7)
chr5 43068909-43069135	LOC100132356
chr5 51202146-51202386	ISL1 (2.39)
chr5 53686645-53686761	HSPB3 (0)
chr5 66118930-66119035	MAST4 (3.22)
chr5 68604297-68604426	CDK7 (5.08)
chr5 76805968-76806077	WDR41 (4.5)
chr5 80923976-80924089	SSBP2 (4.36)
chr5 83032195-83032309	HAPLN1 (0)
chr5 91206680-91206795	LOC100129716 (0)
chr5 92840294-92840522	FLJ42709 (0)
chr5 95382458-95382588	MIR583 (0)
chr5 99708323-99708448	LOC100133050
chr5 100208259-100208435	ST8SIA4 (6.21)
chr5 100217498-100217742	ST8SIA4 (6.21)
chr5 116556696-116556808	SEMA6A (3.45)
chr5 120183541-120183627	PRR16 (0.22)
chr5 120355474-120355584	PRR16 (0.22)
chr5 120500406-120500580	PRR16 (0.22)
chr5 127541726-127541854	SLC12A2 (4.13)
chr5 130802301-130802416	RAPGEF6 (4.77)
chr5 133512341-133512430	PPP2CA (6.68)
chr5 141019043-141019147	RELL2 (3.09)
chr5 142970515-142970691	NR3C1 (6.66)
chr5 144779314-144779431	PRELID2 (0.04)
chr5 146717830-146717923	DPYSL3 (0.24)
chr5 147076722-147076846	JAKMIP2 (2.63)
chr5 149472094-149472216	PDGFRB (0.02)
chr5 151338398-151338625	GLRA1 (0.21)
chr5 159929909-159930033	MIR146A (0)
chr5 163798359-163798492	MAT2B (6.69)

chr5 166959414-166959533	WWC1 (2.01)
chr5 167396323-167396439	WWC1 (2.01)
chr5 174478457-174478569	MIR4634 (0)
chr5 179216403-179216520	LTC4S (0)
chr6 3195918-3196026	TUBB2B (0.64)
chr6 6562598-6562690	LY86-AS1 (0)
chr6 6649190-6649450	LY86-AS1 (0)
chr6 9766831-9766960	TFAP2A (2.5)
chr6 12677099-12677217	PHACTR1 (0.17)
chr6 19356206-19356312	ID4 (0)
chr6 27686873-27686994	LOC100507173 (0)
chr6 27714722-27714836	LOC100131289 (0)
chr6 37717364-37717471	MDGA1 (0.93)
chr6 40741029-40741148	LRFN2 (0)
chr6 43099206-43099312	PTK7 (0.08)
chr6 45528051-45528172	RUNX2 (6.34)
chr6 52860430-52860732	GSTA4 (1.9)
chr6 54604845-54604971	FAM83B (0.05)
chr6 55495531-55495654	HMGCLL1 (0)
chr6 79307606-79307713	IRAK1BP1 (2.86)
chr6 81662243-81662401	BCKDHB (6.31)
chr6 82247266-82247381	FAM46A (3.64)
chr6 82296408-82296534	FAM46A (3.64)
chr6 82804933-82805037	IBTK (5.3)
chr6 83909286-83909412	RWDD2A (3.3)
chr6 85506220-85506337	TBX18 (0)
chr6 89244214-89244320	RNGTT (4.83)
chr6 91602259-91602422	MAP3K7 (4.57)
chr6 97203089-97203215	GPR63 (2.81)
chr6 98376241-98376369	MIR2113 (0)
chr6 101149216-101149328	ASCC3 (5.66)
chr6 102996034-102996142	GRIK2 (0.01)
chr6 111382668-111382781	GSTM2P1 (0.17)
chr6 111612257-111612368	REV3L (4.38)
chr6 115034077-115034192	HS3ST5 (0.04)
chr6 115267892-115267987	HS3ST5 (0.04)
chr6 115409557-115409671	FRK (0.22)
chr6 118184273-118184383	SLC35F1 (0.14)
chr6 120326818-120327007	MAN1A1 (4.03)
chr6 123987298-123987412	TRDN (0.03)
chr6 128299545-128299650	PTPRK (2.54)
chr6 129085148-129085268	LAMA2 (0.43)
chr6 131430510-131430631	AKAP7 (3.05)
chr6 131797809-131797936	ARG1 (0.61)
chr6 132852311-132852429	TAAR9 (0.24)
chr6 134433560-134433659	HMGA1P7 (0.33)
chr6 134719721-134719979	LOC154092 (0)
chr6 134929124-134929217	LOC154092 (0)
chr6 136546666-136546792	FAM54A (0)
chr6 138128310-138128413	TNFAIP3 (8.35)
chr6 138177459-138177568	TNFAIP3 (8.35)
chr6 138440546-138440658	PERP (7.27)
chr6 139036350-139036447	LOC100507462
chr6 139370337-139370463	C6orf115 (0)
chr6 140908992-140909117	MIR4465 (0)
chr6 140985139-140985241	MIR4465 (0)

chr6 150580005-150580124	PPP1R14C (1.41)
chr6 150740825-150740940	IYD (0)
chr6 155952304-155952575	NOX3 (0)
chr6 155953769-155953893	NOX3 (0)
chr6 158215249-158215363	SNX9 (1.65)
chr6 164280067-164280186	QKI (5.25)
chr6 164942252-164942361	C6orf118 (0)
chr6 166913065-166913178	RPS6KA2 (2.03)
chr7 1561144-1561279	MAFK (3.34)
chr7 4283121-4283236	SDK1 (1.51)
chr7 8385305-8385414	ICA1 (1.3)
chr7 11606833-11606935	THSD7A (0.15)
chr7 12971270-12971538	ARL4A (0.1)
chr7 16319547-16319646	LOC100506025 (0)
chr7 18749638-18749753	HDAC9 (0.27)
chr7 25115084-25115197	CYCS (5.07)
chr7 30872880-30872999	FAM188B (0.77)
chr7 32029916-32030029	PDE1C (0.02)
chr7 34329150-34329329	AAA1 (0)
chr7 36031051-36031174	39326 (5.15)
chr7 37592996-37593108	ELMO1 (5.48)
chr7 37904359-37904468	SFRP4 (0.04)
chr7 38122165-38122280	STARD3NL (4.8)
chr7 38539474-38539600	AMPH (4.98)
chr7 44233862-44233962	GCK (0)
chr7 45412789-45412912	RAMP3 (0)
chr7 47833044-47833130	C7orf69 (0)
chr7 50330074-50330213	IKZF1 (6.21)
chr7 70419582-70419708	AUTS2 (2.13)
chr7 70713102-70713188	WBSCR17 (0)
chr7 71222101-71222226	CALN1 (0.01)
chr7 81315455-81315563	HGF (0.05)
chr7 87492694-87492798	DBF4 (5.53)
chr7 88768105-88768215	ZNF804B (0)
chr7 89020751-89020870	ZNF804B (0)
chr7 92218048-92218229	CDK6 (5.53)
chr7 93652003-93652261	BET1 (3.71)
chr7 93878939-93879061	COL1A2 (0)
chr7 96380355-96380506	SHFM1 (6.91)
chr7 97089535-97089661	TAC1 (0)
chr7 99532776-99532894	GJC3 (0.63)
chr7 103710101-103710271	ORC5 (5.16)
chr7 106145145-106145276	C7orf74 (0)
chr7 106224351-106224454	C7orf74 (0)
chr7 112709785-112709894	GPR85 (0.01)
chr7 112947889-112948002	LOC401397 (0)
chr7 117636073-117636203	CTTNBP2 (0.03)
chr7 118561345-118561470	ANKRD7 (1.72)
chr7 120702887-120703011	C7orf58 (0)
chr7 121403784-121403905	PTPRZ1 (0.02)
chr7 123193298-123193430	NDUFA5 (4.05)
chr7 124996210-124996335	POT1 (5.6)
chr7 125121567-125121682	POT1 (5.6)
chr7 125141019-125141162	POT1 (5.6)
chr7 125329632-125329753	GRM8 (0.24)
chr7 128761678-128761807	LOC407835 (2.65)

chr7 130708007-130708155	MKLN1 (3.79)
chr7 131007219-131007333	MKLN1 (3.79)
chr7 131636720-131636844	PLXNA4 (0.02)
chr7 131641284-131641392	PLXNA4 (0.02)
chr7 132183099-132183206	CHCHD3 (6.7)
chr7 139677042-139677163	TBXAS1 (0.29)
chr7 140980252-140980376	LOC100507421 (0)
chr7 143715948-143716065	OR6B1 (0)
chr7 144859727-144859846	TPK1 (1.99)
chr7 145916108-145916209	CNTNAP2 (0.06)
chr7 147084001-147084120	MIR548I4 (0)
chr7 148677002-148677091	PDIA4 (5.46)
chr7 151578407-151578534	LOC100505483 (0)
chr7 151661580-151661693	GALNT11 (0.34)
chr7 154998368-154998477	INSIG1 (6.73)
chr7 158172531-158172653	PTPRN2 (3.04)
chr8 3888825-3888940	CSMD1 (0.01)
chr8 15199070-15199324	SGCZ (0.06)
chr8 15274687-15274794	TUSC3 (0.03)
chr8 15364049-15364163	TUSC3 (0.03)
chr8 15823500-15823598	MSR1 (0.43)
chr8 16314428-16314592	MSR1 (0.43)
chr8 16999619-16999745	ZDHHC2 (5.14)
chr8 17456345-17456457	PDGFRL (3.1)
chr8 19075655-19075768	SH2D4A (4.18)
chr8 20917371-20917497	LOC286114 (0)
chr8 25215711-25215838	DOCK5 (0.39)
chr8 27787789-27787915	SCARA5 (0)
chr8 37340621-37340722	ZNF703 (0.67)
chr8 40153211-40153326	C8orf4 (0.07)
chr8 49430556-49430670	EFCAB1 (0)
chr8 52353707-52353818	PXDNL (0.04)
chr8 55074608-55074731	MRPL15 (3.92)
chr8 63655409-63655519	NKAIN3 (0)
chr8 68684431-68684546	CPA6 (0)
chr8 69152000-69152119	PREX2 (0)
chr8 74288465-74288592	LOC100128126 (0)
chr8 77114204-77114363	LOC100192378 (0)
chr8 77484424-77484515	LOC100192378 (0)
chr8 78434063-78434164	PEX2 (5.59)
chr8 80062445-80062564	IL7 (1.46)
chr8 90757032-90757135	RIPK2 (4.56)
chr8 94351963-94352101	LOC642924 (0)
chr8 95395020-95395123	RAD54B (3.55)
chr8 96135076-96135191	PLEKHF2 (3.08)
chr8 99417954-99418058	KCNS2 (0)
chr8 103897842-103897946	AZIN1 (7.92)
chr8 104576009-104576120	RIMS2 (0.35)
chr8 105291431-105291517	RIMS2 (0.35)
chr8 105714555-105714670	LRP12 (0)
chr8 112386816-112386930	CSMD3 (1.05)
chr8 114710579-114710696	CSMD3 (1.05)
chr8 117688590-117688725	EIF3H (8.59)
chr8 117816613-117816745	UTP23 (4.82)
chr8 117828019-117828116	RAD21 (9.05)
chr8 117864193-117864298	RAD21-AS1 (1.55)

chr8 118275719-118275838	SLC30A8 (0.34)
chr8 118670550-118670662	MED30 (6.28)
chr8 118776757-118776879	EXT1 (4.37)
chr8 119919131-119919245	TNFRSF11B (0)
chr8 124685069-124685211	ANXA13 (0)
chr8 125755056-125755169	MTSS1 (0.82)
chr8 126362800-126362917	NSMCE2 (5.42)
chr8 127019461-127019586	LOC100130231 (0)
chr8 131084636-131084759	ASAP1 (2.19)
chr8 135341052-135341156	ZFAT (3)
chr8 135368508-135368619	ZFAT (3)
chr8 135784073-135784196	MIR30B (0)
chr8 136876329-136876457	KHDRBS3 (4.25)
chr8 140026433-140026598	COL22A1 (6.22)
chr8 143218660-143218765	MIR4472-1 (0)
chr8 144274111-144274216	GPIHBP1 (0.03)
chr9 1930284-1930410	SMARCA2 (7.12)
chr9 2388628-2388748	FLJ35024 (0)
chr9 3742808-3742933	GLIS3 (0.72)
chr9 5580617-5580736	PDCD1LG2 (3.23)
chr9 9442061-9442288	PTPRD (0)
chr9 16917479-16917601	BNC2 (0.57)
chr9 18838067-18838190	ADAMTSL1 (0.01)
chr9 19252122-19252231	DENND4C (4.34)
chr9 26157481-26157606	LOC100506422
chr9 26748184-26748317	C9orf82 (0)
chr9 26752810-26753095	C9orf82 (0)
chr9 29724574-29724687	MIR873 (0)
chr9 30139741-30139992	MIR873 (0)
chr9 31578572-31578696	ACO1 (4.23)
chr9 39630744-39630857	LOC653501 (0)
chr9 41775775-41775890	LOC653501 (0)
chr9 41902210-41902324	MGC21881 (0)
chr9 47022840-47023012	KGFLP1 (0.06)
chr9 66083508-66083648	LOC442421 (0)
chr9 72456607-72456721	C9orf135 (0)
chr9 80847500-80847633	CEP78 (4.25)
chr9 83714053-83714178	TLE1 (2.63)
chr9 92470390-92470513	UNQ6494 (0.03)
chr9 93759291-93759405	LOC100129316 (0.9)
chr9 100153151-100153273	BDAG1 (0)
chr9 101928116-101928341	TGFBR1 (2.94)
chr9 101998558-101998836	SEC61B (6.83)
chr9 102000166-102000480	SEC61B (6.83)
chr9 103801907-103802030	LPPR1 (0)
chr9 106503302-106503416	SMC2 (4.79)
chr9 107092611-107092735	OR13F1 (0)
chr9 110559661-110559786	KLF4 (0)
chr9 113854454-113854578	LPAR1 (0.02)
chr9 117446585-117446696	LOC100505478 (0)
chr9 119787205-119787329	ASTN2 (1.46)
chr9 124879597-124879710	MIR4478 (0)
chr9 137764387-137764511	FCN2 (0)
chr9 138273289-138273417	LOC100506599 (0)
chrM 2791-3172	-1 (0)
chrUn_gl000220 118319-118497	RN5-8S1 (0)

chrX 5330736-5330843	NLGN4X (0)
chrX 5373177-5373288	NLGN4X (0)
chrX 5613355-5613481	NLGN4X (0)
chrX 6140887-6140998	NLGN4X (0)
chrX 6458602-6458730	VCX3A (0)
chrX 7285588-7285711	STS (1.09)
chrX 8427966-8428091	VCX3B (0)
chrX 10657284-10657384	MID1 (1.57)
chrX 13321433-13321555	ATXN3L (0.54)
chrX 13469093-13469201	EGFL6 (0.07)
chrX 15216882-15217009	ASB9 (2.18)
chrX 18889475-18889603	LOC100132163 (0)
chrX 22375020-22375135	ZNF645 (0)
chrX 24526239-24526348	PCYT1B (0.72)
chrX 26432707-26432823	VENTXP1 (0)
chrX 26928846-26929104	VENTXP1 (0)
chrX 28025373-28025478	DCAF8L1 (0.07)
chrX 29678083-29678184	IL1RAPL1 (0.05)
chrX 29789980-29790075	IL1RAPL1 (0.05)
chrX 30398888-30398987	NR0B1 (0)
chrX 34421976-34422090	TMEM47 (0)
chrX 34587120-34587237	TMEM47 (0)
chrX 36466282-36466427	CXorf30 (0)
chrX 36694824-36694961	CXorf30 (0)
chrX 39393821-39393926	LOC286442 (0)
chrX 42824391-42824513	PPP1R2P9 (0)
chrX 74566470-74566578	ZDHHC15 (0.02)
chrX 78154790-78154910	MIR4328 (0)
chrX 78885076-78885187	ITM2A (0.14)
chrX 78886993-78887239	ITM2A (0.14)
chrX 82083735-82083832	POU3F4 (0)
chrX 85270256-85270382	DACH2 (0)
chrX 86638139-86638258	KLHL4 (3.87)
chrX 88861723-88861837	TGIF2LX (0)
chrX 92094184-92094292	PCDH11X (0)
chrX 93502113-93502222	FAM133A (0.55)
chrX 93633542-93633657	FAM133A (0.55)
chrX 93990323-93990442	FAM133A (0.55)
chrX 95720773-95720870	LOC643486 (0)
chrX 96883268-96883421	DIAPH2 (4.55)
chrX 111587628-111587753	TRPC5 (0.23)
chrX 111782276-111782403	ZCCHC16 (0)
chrX 113028118-113028244	HTR2C (0)
chrX 117030753-117030867	KLHL13 (0.13)
chrX 118425457-118425565	PGRMC1 (6.49)
chrX 119138358-119138483	NKAP (5.5)
chrX 119195846-119195943	RHOXF2B (0)
chrX 119308224-119308327	RHOXF2B (0)
chrX 124346976-124347090	LOC100129520
chrX 131759358-131759455	HS6ST2 (0)
chrX 140455351-140455489	SPANXC (0)
chrX 140722648-140722761	SPANXA2-OT1 (0)
chrX 141161893-141162004	MAGEC2 (0)
chrX 141168314-141168428	MAGEC2 (0)
chrX 146956821-146956919	FMR1-AS1 (0.17)
chrX 150044247-150044371	CD99L2 (4.32)

chrX 152370320-152370432	MAGEA1 (0)
chrX 153060062-153060217	SSR4 (7.77)

chrX 153195436-153195668	NAA10 (6.34)

Supplementary Data Table 5. Closest genes to active LTRs detected by RACE-Seq in **KM-H2** cells. Gene expression values (FPKM) from RNA-Seq are also shown for each gene.

RACE Peak	Gene (FPKM)
chr1 135124-135227	LOC729737 (0.47)
chr1 3492135-3492248	MEGF6 (0.02)
chr1 3536145-3536238	TPRG1L (3.83)
chr1 3835054-3835173	LOC100133612 (0)
chr1 4036166-4036290	LOC728716 (0)
chr1 6784220-6784312	DNAJC11 (4.35)
chr1 7036985-7037100	CAMTA1 (4.55)
chr1 7040763-7040852	CAMTA1 (4.55)
chr1 7968801-7968917	TNFRSF9 (3.31)
chr1 8225875-8225984	ERRFI1 (0.16)
chr1 8245000-8245106	SLC45A1 (0)
chr1 8900213-8900318	ENO1 (9.6)
chr1 11467937-11468043	PTCHD2 (0.55)
chr1 14328921-14329042	PRDM2 (4.48)
chr1 18905856-18905949	PAX7 (0.01)
chr1 19839395-19839514	CAPZB (5.49)
chr1 21719072-21719178	ECE1 (4.02)
chr1 24325553-24325666	SRSF10 (0.02)
chr1 24526768-24526884	IL28RA (0)
chr1 24814726-24814865	RCAN3 (2)
chr1 35198466-35198549	GJB5 (0.08)
chr1 38959164-38959272	LOC339442 (0)
chr1 42259806-42259910	HIVEP3 (2.64)
chr1 43094328-43094433	PPIH (6.66)
chr1 43217321-43217407	LEPRE1 (0)
chr1 53806983-53807084	LRP8 (5.26)
chr1 55203474-55203574	PARS2 (2.6)
chr1 55404744-55404853	TMEM61 (0)
chr1 56854440-56854553	PPAP2B (3.53)
chr1 57271454-57271557	C1orf168 (0.13)
chr1 58368620-58368736	DAB1 (0)
chr1 58606900-58607019	DAB1 (0)
chr1 59980384-59980502	FGGY (2.29)
chr1 60201768-60201865	FGGY (2.29)
chr1 60255942-60256048	HOOK1 (0)
chr1 62386060-62386163	INADL (1.64)
chr1 65464423-65464526	JAK1 (5.85)
chr1 65553383-65553483	MIR101-1 (0)
chr1 66594408-66594527	PDE4B (1.5)
chr1 67673268-67673386	IL23R (0.01)
chr1 71406208-71406326	PTGER3 (0.65)
chr1 72944797-72944912	NEGR1 (0.98)
chr1 73938542-73938647	LRRIQ3 (0.79)
chr1 78588137-78588249	GIPC2 (0.02)
chr1 78919473-78919576	PTGFR (0)
chr1 79017335-79017440	PTGFR (0)
chr1 80201070-80201176	ELTD1 (0)
chr1 81428611-81428730	LPHN2 (0)
chr1 81721811-81721926	LPHN2 (0)
chr1 82218894-82219001	LPHN2 (0)
chr1 83498805-83498908	TTLL7 (1.78)
chr1 84498158-84498276	TTLL7 (1.78)
chr1 85191720-85191827	SSX2IP (2.41)
chr1 93727503-93727601	CCDC18 (3.53)
chr1 93866909-93866998	DR1 (5.06)
chr1 95152048-95152168	SLC44A3 (0.03)
chr1 97721047-97721146	DPYD (3.04)
chr1 97722060-97722174	DPYD (3.04)
chr1 98070998-98071092	DPYD (3.04)
chr1 101455655-101455880	DPH5 (4.02)
chr1 101491248-101491341	DPH5 (4.02)
chr1 101808582-101808695	S1PR1 (2)
chr1 102403107-102403214	OLFM3 (0)
chr1 102479013-102479131	OLFM3 (0)
chr1 108093076-108093198	VAV3 (0.02)
chr1 111053783-111053902	KCNA10 (0)
chr1 111543829-111543931	LRIF1 (4.36)
chr1 118258820-118258923	FAM46C (0.55)
chr1 118996482-118996657	SPAG17 (0.08)
chr1 146669168-146669287	FMO5 (0.03)
chr1 146786689-146786794	CHD1L (5.19)
chr1 148213993-148214114	PPIAL4D (0)
chr1 148632429-148632550	PPIAL4E (0)
chr1 148794461-148794582	PPIAL4D (0)
chr1 152401709-152401826	CRNN (0.02)
chr1 154281793-154281908	AQP10 (0.19)
chr1 157866064-157866170	CD5L (0.44)
chr1 157867990-157868096	CD5L (0.44)
chr1 158194470-158194585	CD1A (0)
chr1 160682597-160682685	CD48 (2.12)
chr1 163511954-163512059	NUF2 (5.26)
chr1 164446858-164446979	PBX1 (0.33)
chr1 164519526-164519608	PBX1 (0.33)
chr1 164633935-164634045	PBX1 (0.33)
chr1 165429234-165429335	RXRG (0.19)
chr1 166747089-166747200	POGK (5.82)
chr1 167160219-167160324	POU2F1 (3.3)
chr1 168731052-168731167	MGC4473 (0)
chr1 169842378-169842465	SCYL3 (3.87)
chr1 170585242-170585411	PRRX1 (0.14)
chr1 170803684-170803801	PRRX1 (0.14)
chr1 171409397-171409517	PRRC2C (5.62)
chr1 172215096-172215201	DNM3 (0.59)
chr1 173439088-173439214	LOC100506023 (0.61)
chr1 174173949-174174066	RABGAP1L (4.4)
chr1 174569165-174569284	RABGAP1L (4.4)
chr1 174571374-174571490	RABGAP1L (4.4)
chr1 177701423-177701556	SEC16B (0.14)
chr1 178636257-178636376	MIR4424 (0)

chr1 178913875-178913994	RALGPS2 (0.02)	**chr10 7316477-7316571**	SFMBT2 (0.46)
chr1 179370863-179370986	SOAT1 (3.23)	**chr10 8383281-8383391**	GATA3 (4.33)
chr1 179445123-179445221	AXDND1 (0.1)	**chr10 9407410-9407528**	GATA3 (4.33)
chr1 180528084-180528184	ACBD6 (5.78)	**chr10 9724552-9724652**	SFTA1P (0)
chr1 181511716-181511814	CACNA1E (5.12)	**chr10 10862041-10862159**	SFTA1P (0)
chr1 182202215-182202321	GLUL (0)	**chr10 11389670-11389770**	CELF2 (2.45)
chr1 182247077-182247175	GLUL (0)	**chr10 12774469-12774618**	CAMK1D (4.41)
chr1 182505791-182505906	RGSL1 (0)	**chr10 13478504-13478618**	BEND7 (0)
chr1 183281079-183281204	NMNAT2 (0)	**chr10 14256008-14256126**	FRMD4A (0.59)
chr1 183418541-183418705	DKFZP564C196 (0)	**chr10 15080558-15080682**	OLAH (0.24)
chr1 186213428-186213566	MIR548F1 (0)	**chr10 15225854-15225972**	NMT2 (2.13)
chr1 186715860-186715963	PTGS2 (0)	**chr10 15548643-15548758**	ITGA8 (0)
chr1 188701645-188701764	FAM5C (0)	**chr10 15981505-15981631**	FAM188A (4.27)
chr1 191367827-191367937	LOC440704 (0.18)	**chr10 17277260-17277382**	VIM (8.97)
chr1 193446098-193446205	CDC73 (1.88)	**chr10 17568246-17568364**	PTPLA (0)
chr1 195754706-195754824	KCNT2 (0)	**chr10 19520530-19520632**	ARL5B (2.57)
chr1 197197938-197198058	ZBTB41 (2.83)	**chr10 20647714-20647816**	PLXDC2 (0)
chr1 197964817-197964936	LHX9 (3.66)	**chr10 21068198-21068298**	NEBL (0.19)
chr1 200247989-200248108	C1orf98 (0)	**chr10 24566570-24566658**	KIAA1217 (0.18)
chr1 204011679-204011815	LINC00303 (0)	**chr10 24821478-24821579**	ARHGAP21 (4.35)
chr1 209874572-209874690	HSD11B1 (0)	**chr10 25176404-25176495**	PRTFDC1 (0)
chr1 210249985-210250118	SYT14 (1.22)	**chr10 25301982-25302101**	ENKUR (0.05)
chr1 211714288-211714389	SLC30A1 (4.36)	**chr10 25481137-25481238**	GPR158 (0.06)
chr1 213070988-213071126	FLVCR1 (4.53)	**chr10 30351545-30351651**	KIAA1462 (2.5)
chr1 215405122-215405240	KCTD3 (3.31)	**chr10 34732403-34732502**	PARD3 (0.05)
chr1 216310651-216310771	USH2A (0.21)	**chr10 36588897-36588981**	FZD8 (0)
chr1 216757102-216757205	ESRRG (0.17)	**chr10 38193739-38193840**	ZNF25 (2.33)
chr1 219114821-219114924	LOC643723 (0)	**chr10 45397168-45397283**	LOC220980 (0)
chr1 220394100-220394183	RAB3GAP2 (4.41)	**chr10 47109752-47109870**	LOC643650 (0)
chr1 220532669-220532787	RAB3GAP2 (4.41)	**chr10 49394570-49394678**	FRMPD2 (2.15)
chr1 220976469-220976575	MOSC1 (0)	**chr10 49498149-49498252**	FRMPD2 (2.15)
chr1 222238030-222238133	DUSP10 (2.87)	**chr10 50791282-50791400**	CHAT (0)
chr1 222360694-222360812	HHIPL2 (0.16)	**chr10 52464617-52464735**	ASAH2B (0.3)
chr1 222401271-222401389	HHIPL2 (0.16)	**chr10 52820931-52821035**	PRKG1 (1.17)
chr1 223312317-223312422	SUSD4 (2.91)	**chr10 57338466-57338584**	MTRNR2L5 (0)
chr1 224847891-224847996	CNIH3 (0.1)	**chr10 59821017-59821137**	IPMK (2)
chr1 234081483-234081603	SLC35F3 (4.64)	**chr10 59933169-59933279**	IPMK (2)
chr1 234312901-234313004	SLC35F3 (4.64)	**chr10 62312661-62312770**	ANK3 (0.98)
chr1 234730410-234730519	IRF2BP2 (6.15)	**chr10 62416804-62416915**	ANK3 (0.98)
chr1 234831806-234831917	LOC100506810 (0)	**chr10 64018897-64018994**	RTKN2 (0.52)
chr1 236326941-236327059	GPR137B (3.63)	**chr10 67021529-67021647**	ANXA2P3 (0.13)
chr1 237144761-237144898	RYR2 (0.14)	**chr10 69291679-69291794**	CTNNA3 (0.5)
chr1 238533010-238533127	LOC339535 (0)	**chr10 72878637-72878736**	UNC5B (0)
chr1 241450705-241450821	RGS7 (0.01)	**chr10 75006850-75007045**	C10orf103 (0)
chr1 241575471-241575590	RGS7 (0.01)	**chr10 77592457-77592572**	C10orf11 (0)
chr1 242218073-242218174	PLD5 (0.04)	**chr10 77870431-77870538**	C10orf11 (0)
chr1 243269479-243269597	LOC731275 (0)	**chr10 78271864-78271998**	KCNMA1 (3.55)
chr1 244472685-244472791	C1orf100 (0)	**chr10 79468100-79468235**	KCNMA1 (3.55)
chr1 244528720-244528826	ADSS (5.3)	**chr10 81114911-81115006**	ZCCHC24 (2.24)
chr1 244830105-244830225	PPPDE1 (0)	**chr10 81629612-81629730**	LOC100288974 (0)
chr1 245121026-245121138	EFCAB2 (0.76)	**chr10 82104458-82104572**	DYDC2 (0.34)
chr1 245353128-245353258	KIF26B (0.02)	**chr10 82388111-82388213**	SH2D4B (0.07)
chr10 2941453-2941547	PFKP (5.85)	**chr10 84608586-84608692**	NRG3 (0)
chr10 3780218-3780318	KLF6 (2.54)	**chr10 85909896-85910014**	C10orf99 (0.05)
chr10 4915274-4915501	tAKR (0)	**chr10 86904150-86904251**	LOC100507470 (0)
chr10 7144562-7144674	SFMBT2 (0.46)	**chr10 88405001-88405112**	OPN4 (0)

chr10 89145919-89146037	FAM22D (0)	**chr11 37827793-37827898**	C11orf74 (0.04)
chr10 89404417-89404536	PAPSS2 (0.16)	**chr11 39284098-39284201**	LRRC4C (0)
chr10 89408785-89408888	PAPSS2 (0.16)	**chr11 41206403-41206506**	LRRC4C (0)
chr10 92426123-92426214	HTR7 (0.01)	**chr11 45437648-45437741**	SYT13 (0.01)
chr10 92763169-92763286	ANKRD1 (0.07)	**chr11 46193468-46193551**	PHF21A (2.92)
chr10 92765110-92765236	ANKRD1 (0.07)	**chr11 46450246-46450405**	AMBRA1 (4.36)
chr10 93095169-93095273	LOC100188947 (0)	**chr11 60677262-60677365**	PRPF19 (4.95)
chr10 95943481-95943605	PLCE1 (0.28)	**chr11 61235289-61235395**	C11orf66 (0)
chr10 106162989-106163107	CCDC147 (0)	**chr11 63059597-63059714**	SLC22A10 (0.29)
chr10 107729820-107729936	SORCS1 (3.83)	**chr11 63595051-63595135**	MARK2 (2.65)
chr10 108046161-108046266	SORCS1 (3.83)	**chr11 64185770-64185887**	RPS6KA4 (3.33)
chr10 109115335-109115458	SORCS1 (3.83)	**chr11 67450464-67450556**	ALDH3B2 (0)
chr10 109262913-109263028	SORCS1 (3.83)	**chr11 68903626-68903757**	TPCN2 (1.35)
chr10 109482761-109482876	SORCS1 (3.83)	**chr11 70160995-70161085**	PPFIA1 (3.58)
chr10 110936990-110937091	XPNPEP1 (5.01)	**chr11 71881685-71881788**	FOLR1 (0)
chr10 111357871-111357972	XPNPEP1 (5.01)	**chr11 73685721-73685921**	DNAJB13 (0)
chr10 112373134-112373249	SMC3 (4.76)	**chr11 74145705-74145814**	KCNE3 (0.01)
chr10 116081825-116081943	AFAP1L2 (0.71)	**chr11 77554876-77554977**	RSF1 (3.48)
chr10 117803536-117803629	GFRA1 (0.04)	**chr11 78787418-78787524**	ODZ4 (0)
chr10 118178158-118178261	PNLIPRP3 (0.02)	**chr11 78905144-78905263**	ODZ4 (0)
chr10 118455941-118456045	HSPA12A (0)	**chr11 82918080-82918191**	ANKRD42 (2.36)
chr10 118588707-118588837	ENO4 (0.1)	**chr11 84202355-84202470**	DLG2 (0.08)
chr10 119062313-119062430	SLC18A2 (1.26)	**chr11 85494213-85494306**	SYTL2 (1.7)
chr10 123123688-123123807	FGFR2 (0)	**chr11 85631078-85631196**	CCDC83 (0.09)
chr10 123479796-123479902	ATE1 (4.55)	**chr11 88419877-88419996**	GRM5 (1.74)
chr10 123751370-123751468	NSMCE4A (5.9)	**chr11 92332399-92332500**	FAT3 (0)
chr10 125171648-125171754	BUB3 (6.4)	**chr11 94969583-94969695**	SESN3 (3.41)
chr10 125305963-125306085	GPR26 (0)	**chr11 95682839-95682936**	MTMR2 (2.8)
chr10 125820867-125820986	CHST15 (0.09)	**chr11 97997125-97997227**	CNTN5 (1.82)
chr10 128565687-128565799	DOCK1 (0.12)	**chr11 98490141-98490244**	CNTN5 (1.82)
chr10 131050865-131050983	MGMT (3.8)	**chr11 102217944-102218085**	BIRC2 (5.29)
chr10 132469161-132469280	MIR378C (0)	**chr11 102759211-102759329**	MMP12 (0)
chr11 214884-215035	RIC8A (4.38)	**chr11 110183027-110183139**	RDX (4.02)
chr11 2497040-2497143	KCNQ1 (0)	**chr11 110214740-110214855**	RDX (4.02)
chr11 2644349-2644461	KCNQ1 (0)	**chr11 113749387-113749507**	USP28 (3.82)
chr11 3090769-3090879	CARS (5.28)	**chr11 114997252-114997378**	CADM1 (2.78)
chr11 3206241-3206333	OSBPL5 (0)	**chr11 115453099-115453183**	CADM1 (2.78)
chr11 3402161-3402280	LOC650368 (0)	**chr11 115797096-115797220**	LOC283143 (0)
chr11 5526791-5526927	OR51B5 (0.1)	**chr11 115965792-115965903**	LOC283143 (0)
chr11 6756086-6756205	GVINP1 (0.51)	**chr11 116088773-116088876**	LOC283143 (0)
chr11 7698385-7698522	CYB5R2 (4.96)	**chr11 119483054-119483159**	PVRL1 (2.9)
chr11 9189817-9189970	DENND5A (4.66)	**chr11 119773507-119773623**	PVRL1 (2.9)
chr11 10458289-10458406	AMPD3 (1.08)	**chr11 123538407-123538510**	SCN3B (0.18)
chr11 13010992-13011127	RASSF10 (0)	**chr11 126517974-126518115**	KIRREL3 (0.09)
chr11 14541845-14541974	PSMA1 (7.33)	**chr11 127264576-127264677**	KIRREL3-AS3 (0)
chr11 19548787-19548893	NAV2 (0.17)	**chr11 128240074-128240193**	ETS1 (5.28)
chr11 25389367-25389480	LUZP2 (0)	**chr11 129009537-129009649**	ARHGAP32 (1.79)
chr11 26500986-26501092	ANO3 (1.59)	**chr12 1189820-1189940**	RAD52 (2.96)
chr11 26755471-26755592	SLC5A12 (0.08)	**chr12 1635084-1635168**	LOC100292680 (0)
chr11 26963481-26963594	FIBIN (0)	**chr12 4514536-4514651**	FGF23 (0)
chr11 27363533-27363634	CCDC34 (3)	**chr12 6852659-6852777**	MLF2 (6.52)
chr11 27911940-27912094	KIF18A (3.37)	**chr12 9797863-9797993**	LOC374443 (0.42)
chr11 32154175-32154293	RCN1 (4.1)	**chr12 9804066-9804159**	CLEC2D (3.38)
chr11 33753564-33753676	FBXO3 (3.49)	**chr12 10800431-10800541**	STYK1 (0.32)
chr11 36607845-36607947	RAG2 (0)	**chr12 12261926-12262034**	MIR1244-1 (0)
chr11 37150918-37151027	C11orf74 (0.04)	**chr12 12971653-12971773**	DDX47 (5.74)

chr12 14296899-14297014	GRIN2B (0.32)	**chr12 95100203-95100310**	TMCC3 (0.07)
chr12 14432849-14432944	ATF7IP (4.61)	**chr12 97225910-97226024**	NEDD1 (3.84)
chr12 21810688-21810799	LDHB (10.4)	**chr12 97286224-97286309**	NEDD1 (3.84)
chr12 22280557-22280656	CMAS (5)	**chr12 97477865-97477982**	NEDD1 (3.84)
chr12 22421450-22421584	KIAA0528 (0)	**chr12 102054514-102054618**	CHPT1 (1.73)
chr12 22983609-22983728	ETNK1 (5.72)	**chr12 102685554-102685665**	PMCH (0.72)
chr12 23286635-23286753	SOX5 (0.7)	**chr12 102739289-102739377**	IGF1 (0.13)
chr12 25108451-25108596	BCAT1 (5.63)	**chr12 103421726-103421845**	ASCL1 (0)
chr12 25446062-25446167	KRAS (4.63)	**chr12 104093186-104093286**	NT5DC3 (1.25)
chr12 25890981-25891086	IFLTD1 (0)	**chr12 108077731-108077836**	PWP1 (5.2)
chr12 28298990-28299108	CCDC91 (4.41)	**chr12 108294605-108294714**	LOC728739 (2.95)
chr12 30020945-30021046	TMTC1 (2.15)	**chr12 108479105-108479211**	WSCD2 (0)
chr12 33171264-33171366	PKP2 (0)	**chr12 109775293-109775380**	FOXN4 (0.06)
chr12 40506886-40506992	SLC2A13 (2.48)	**chr12 111565570-111565676**	CUX2 (0.3)
chr12 40597566-40597675	LRRK2 (0.05)	**chr12 111848625-111848749**	ATXN2 (2.68)
chr12 41473402-41473508	CNTN1 (0.34)	**chr12 112847271-112847384**	PTPN11 (4.95)
chr12 41706344-41706462	PDZRN4 (0)	**chr12 114244952-114245040**	RBM19 (5.14)
chr12 42008364-42008465	PDZRN4 (0)	**chr12 114438141-114438258**	RBM19 (5.14)
chr12 44675857-44675975	NELL2 (0.12)	**chr12 115061119-115061208**	TBX3 (1.55)
chr12 46544812-46544941	SLC38A1 (0.58)	**chr12 115345763-115345890**	TBX3 (1.55)
chr12 50505957-50506075	C12orf62 (0)	**chr12 115803815-115803923**	MED13L (3.63)
chr12 56043450-56043540	OR10P1 (0)	**chr12 115826213-115826320**	MED13L (3.63)
chr12 58564583-58564683	XRCC6BP1 (3.8)	**chr12 115965956-115966075**	MED13L (3.63)
chr12 58713402-58713505	XRCC6BP1 (3.8)	**chr12 116356601-116356706**	MED13L (3.63)
chr12 59427482-59427598	LRIG3 (2.26)	**chr12 117561636-117561755**	FBXO21 (3.25)
chr12 59517578-59517681	LRIG3 (2.26)	**chr12 118792525-118792633**	TAOK3 (3.59)
chr12 59880368-59880483	SLC16A7 (1.5)	**chr12 120398589-120398693**	CCDC64 (3.66)
chr12 60826106-60826215	SLC16A7 (1.5)	**chr12 121675802-121675981**	CAMKK2 (2.84)
chr12 61195157-61195273	FAM19A2 (0.19)	**chr12 125597291-125597394**	AACS (3.95)
chr12 61923504-61923608	FAM19A2 (0.19)	**chr12 125905869-125905987**	TMEM132B (0.02)
chr12 62608126-62608229	FAM19A2 (0.19)	**chr12 125973092-125973195**	TMEM132B (0.02)
chr12 62841257-62841360	MON2 (2.72)	**chr12 126533999-126534113**	LOC400084 (0)
chr12 63402435-63402553	PPM1H (3.63)	**chr12 127128423-127128541**	LOC387895 (0)
chr12 66414342-66414479	HMGA2 (0)	**chr12 127254736-127254866**	LOC100507206 (0)
chr12 68835835-68836256	MDM1 (2.99)	**chr12 127468384-127468487**	LOC440117 (0)
chr12 71633155-71633274	TSPAN8 (0)	**chr12 128045151-128045257**	FLJ37505 (0.29)
chr12 72623002-72623116	LOC283392 (0)	**chr12 128194281-128194387**	FLJ37505 (0.29)
chr12 74595298-74595401	LOC100507377 (0.51	**chr12 129263088-129263205**	SLC15A4 (3.42)
chr12 76365243-76365358	PHLDA1 (0.6)	**chr12 129594079-129594310**	TMEM132D (0)
chr12 77730026-77730145	E2F7 (2.33)	**chr12 130769019-130769134**	PIWIL1 (0.14)
chr12 81171248-81171364	LIN7A (0)	**chr12 131698628-131698767**	LOC116437 (0)
chr12 82124159-82124241	PPFIA2 (0.02)	**chr13 19128875-19128991**	ANKRD20A9P (1.16)
chr12 83839056-83839174	TMTC2 (1.57)	**chr13 21665981-21666077**	LATS2 (0.39)
chr12 85872024-85872126	ALX1 (0)	**chr13 27766745-27766857**	USP12 (3.93)
chr12 86547473-86547576	MGAT4C (0.43)	**chr13 28651452-28651570**	PAN3-AS1 (2.24)
chr12 88535708-88535816	CEP290 (2.62)	**chr13 29118653-29118759**	FLT1 (0)
chr12 88543925-88544036	CEP290 (2.62)	**chr13 31501576-31501681**	LOC100507064 (0)
chr12 88721992-88722102	TMTC3 (2.17)	**chr13 35205298-35205402**	NBEA (1.28)
chr12 88991665-88991784	KITLG (4.15)	**chr13 36506660-36506759**	MIR548F5 (0)
chr12 89445752-89445867	LOC728084 (0)	**chr13 36946963-36947067**	SPG20 (0)
chr12 89541372-89541475	LOC728084 (0)	**chr13 38410207-38410323**	TRPC4 (0.01)
chr12 91058848-91058963	C12orf37 (0)	**chr13 45647997-45648100**	KIAA1704 (0)
chr12 92739751-92739856	CLLU1OS (0)	**chr13 46234665-46234783**	SPERT (0.02)
chr12 92955007-92955125	CLLU1 (0)	**chr13 47730203-47730317**	HTR2A (0)
chr12 93335100-93335215	EEA1 (2.61)	**chr13 49259276-49259384**	CYSLTR2 (4.58)
chr12 94202042-94202147	CRADD (2.02)	**chr13 52997141-52997244**	THSD1 (0.85)

chr13 54539428-54539541	MIR1297 (0)	**chr14 49188861-49188979**	RPS29 (10.97)
chr13 56917565-56917678	PRR20C (0)	**chr14 50053389-50053591**	RPS29 (10.97)
chr13 58716873-58716989	PCDH17 (0)	**chr14 50872998-50873202**	CDKL1 (1.41)
chr13 59134882-59134997	PCDH17 (0)	**chr14 51651877-51651980**	TMX1 (4.38)
chr13 59883236-59883344	DIAPH3 (2.77)	**chr14 51869186-51869290**	LOC283553 (0)
chr13 60982144-60982347	TDRD3 (3.89)	**chr14 51960189-51960307**	FRMD6 (1.47)
chr13 62840882-62840985	PCDH20 (0)	**chr14 52372553-52372658**	GNG2 (5.29)
chr13 63631303-63631433	OR7E156P (0)	**chr14 54453587-54453699**	BMP4 (0)
chr13 68267457-68267577	PCDH9 (3.81)	**chr14 56263321-56263427**	C14orf34 (0)
chr13 69251468-69251583	LOC338862 (0)	**chr14 56780071-56780177**	PELI2 (0.08)
chr13 70322448-70322562	KLHL1 (1.38)	**chr14 57484550-57484657**	OTX2OS1 (0)
chr13 71231369-71231486	ATXN8OS (0.05)	**chr14 57871385-57871494**	C14orf105 (0.07)
chr13 72664298-72664416	DACH1 (0.44)	**chr14 60677146-60677321**	PPM1A (3.79)
chr13 72916027-72916129	MZT1 (3.37)	**chr14 66447217-66447316**	FUT8 (5.85)
chr13 76798442-76798540	LMO7 (2.18)	**chr14 66521580-66521695**	FUT8 (5.85)
chr13 77307132-77307238	KCTD12 (0.04)	**chr14 67271815-67271930**	GPHN (3.11)
chr13 78003763-78003882	MYCBP2 (3.09)	**chr14 67273779-67273886**	GPHN (3.11)
chr13 78500970-78501078	EDNRB (0.1)	**chr14 69377549-69377657**	ACTN1 (1.23)
chr13 79320204-79320294	RNF219 (3.51)	**chr14 70202928-70203010**	KIAA0247 (0)
chr13 84882890-84882995	SLITRK1 (0)	**chr14 70729399-70729519**	ADAM21P1 (0.12)
chr13 87346181-87346298	MIR4500HG (0.56)	**chr14 71673290-71673387**	PCNX (3.09)
chr13 88848754-88848860	SLITRK5 (0.52)	**chr14 71713067-71713163**	PCNX (3.09)
chr13 89964572-89964686	MIR622 (0)	**chr14 76012909-76013263**	FLVCR2 (0.74)
chr13 92540381-92540465	GPC5 (1.33)	**chr14 78414926-78415029**	ADCK1 (1.74)
chr13 93591201-93591311	GPC5 (1.33)	**chr14 80785911-80786026**	LOC100628307 (0)
chr13 93611467-93611586	GPC5 (1.33)	**chr14 80945720-80945838**	CEP128 (1.7)
chr13 99439638-99439733	DOCK9 (1.92)	**chr14 81675618-81675703**	GTF2A1 (3)
chr13 99772193-99772313	DOCK9 (1.92)	**chr14 81723112-81723230**	STON2 (0)
chr13 100104204-100104321	MIR548AN (0)	**chr14 83317256-83317362**	SEL1L (4.82)
chr13 100113263-100113369	TM9SF2 (4.69)	**chr14 86559141-86559257**	FLRT2 (0.01)
chr13 101360615-101360708	TMTC4 (2.18)	**chr14 86674032-86674144**	FLRT2 (0.01)
chr13 102161456-102161573	NALCN (0.02)	**chr14 87240614-87240741**	LOC283585 (0)
chr13 103752583-103752686	SLC10A2 (0.05)	**chr14 88073541-88073650**	GALC (0)
chr13 103919748-103919861	SLC10A2 (0.05)	**chr14 88317138-88317244**	GALC (0)
chr13 103997782-103997883	SLC10A2 (0.05)	**chr14 88388292-88388422**	GALC (0)
chr13 108682808-108682913	FAM155A (0.04)	**chr14 89471489-89471608**	TTC8 (3.1)
chr13 108843832-108843931	LIG4 (2.65)	**chr14 89522339-89522457**	FOXN3 (1.66)
chr13 110617242-110617358	IRS2 (1.96)	**chr14 89744491-89744611**	FOXN3 (1.66)
chr13 110974057-110974166	COL4A2 (0)	**chr14 90164132-90164232**	FOXN3 (1.66)
chr13 111594774-111594893	ANKRD10 (5.16)	**chr14 90339470-90339577**	TDP1 (4.72)
chr13 112025193-112025287	C13orf16 (0)	**chr14 90514137-90514255**	TDP1 (4.72)
chr13 113799969-113800074	PROZ (0.18)	**chr14 92630319-92630504**	CPSF2 (5.18)
chr14 19680937-19681052	POTEG (0)	**chr14 92753561-92753666**	SLC24A4 (0)
chr14 19894332-19894447	POTEM (0.11)	**chr14 92868973-92869073**	RIN3 (0.04)
chr14 20925343-20925436	TMEM55B (3.8)	**chr14 93336867-93337020**	GOLGA5 (4.44)
chr14 21134860-21135028	ANG (4.09)	**chr14 93338648-93338782**	GOLGA5 (4.44)
chr14 24104085-24104196	DHRS2 (8.9)	**chr14 93339798-93339934**	GOLGA5 (4.44)
chr14 28645719-28645820	LOC100505967 (0)	**chr14 93366476-93366595**	CHGA (0.29)
chr14 30348580-30348701	PRKD1 (1.39)	**chr14 94114160-94114277**	PRIMA1 (0)
chr14 30515075-30515206	PRKD1 (1.39)	**chr14 94143688-94143805**	PRIMA1 (0)
chr14 37469995-37470101	SLC25A21 (1.98)	**chr14 94288209-94288327**	PRIMA1 (0)
chr14 37594019-37594141	SLC25A21 (1.98)	**chr14 94762672-94762788**	SERPINA10 (0.06)
chr14 38825620-38825735	CLEC14A (0)	**chr14 94884443-94884549**	SERPINA11 (0)
chr14 41208395-41208501	LRFN5 (0)	**chr14 94886389-94886519**	SERPINA11 (0)
chr14 44143114-44143217	FSCB (0.12)	**chr14 97903770-97903885**	LOC100129345 (0.41)
chr14 45260477-45260583	C14orf28 (1.48)	**chr14 98498455-98498550**	C14orf64 (0)

chr14 98748846-98748948	C14orf64 (0)	**chr15 100438376-100438491**	ADAMTS17 (0.04)
chr14 100800131-100800265	SLC25A47 (0)	**chr15 100947071-100947164**	CERS3 (0.08)
chr14 105222904-105223078	SIVA1 (6.53)	**chr15 101835338-101835443**	PCSK6 (0.34)
chr14 105771045-105771160	BRF1 (3.97)	**chr15 102182661-102182796**	TARSL2 (3.54)
chr15 23188081-23188257	WHAMMP3 (2.27)	**chr15 102347613-102347745**	OR4F6 (0)
chr15 26619162-26619268	GABRB3 (0)	**chr16 5120972-5121073**	ALG1 (1.97)
chr15 29002980-29003086	LOC100289656 (0)	**chr16 6147114-6147232**	RBFOX1 (0.92)
chr15 32495824-32495931	CHRNA7 (0.02)	**chr16 7826358-7826459**	RBFOX1 (0.92)
chr15 33054496-33054615	FMN1 (0.02)	**chr16 8036749-8036855**	RBFOX1 (0.92)
chr15 33502046-33502157	RYR3 (0.15)	**chr16 8149819-8149935**	RBFOX1 (0.92)
chr15 37490006-37490118	MEIS2 (3.28)	**chr16 8708563-8708681**	METTL22 (2.61)
chr15 38339671-38339771	TMCO5A (0)	**chr16 10002939-10003039**	GRIN2A (0.01)
chr15 38582796-38582916	FAM98B (2.64)	**chr16 11060464-11060574**	CLEC16A (4.66)
chr15 39527796-39527913	C15orf54 (0.16)	**chr16 12543915-12544015**	SNX29 (2.6)
chr15 39945916-39946031	FSIP1 (2.27)	**chr16 12753698-12753802**	CPPED1 (2.86)
chr15 42788990-42789100	ZFP106 (0)	**chr16 13145039-13145151**	SHISA9 (2.49)
chr15 44530090-44530200	FRMD5 (1.14)	**chr16 13354736-13354856**	SHISA9 (2.49)
chr15 45103618-45103707	TRIM69 (0.58)	**chr16 15027011-15027113**	NPIP (0)
chr15 47835588-47835708	SEMA6D (0)	**chr16 15342891-15342973**	MIR3180-4 (0)
chr15 48312659-48312750	SLC24A5 (0.67)	**chr16 15370192-15370293**	MPV17L (0)
chr15 48362570-48362688	SLC24A5 (0.67)	**chr16 15666758-15666858**	KIAA0430 (2.5)
chr15 49907724-49907827	C15orf33 (0)	**chr16 15667941-15668044**	KIAA0430 (2.5)
chr15 50024008-50024132	DTWD1 (2.67)	**chr16 16593363-16593481**	PKD1P1 (3.62)
chr15 53016821-53016939	ONECUT1 (0.44)	**chr16 16757508-16757610**	PKD1P1 (3.62)
chr15 53597013-53597105	WDR72 (0.06)	**chr16 17119539-17119648**	XYLT1 (0.62)
chr15 54071963-54072066	WDR72 (0.06)	**chr16 17992400-17992506**	XYLT1 (0.62)
chr15 54902018-54902113	UNC13C (0.01)	**chr16 18298360-18298473**	MIR3180-3 (0)
chr15 56135263-56135366	PRTG (0)	**chr16 18306408-18306526**	MIR3180-3 (0)
chr15 56689988-56690087	MNS1 (0.17)	**chr16 18326114-18326236**	MIR3180-3 (0)
chr15 57832483-57832599	CGNL1 (0)	**chr16 20035584-20035699**	GPR139 (0)
chr15 58113629-58113748	GCOM1 (0)	**chr16 21127745-21127850**	DNAH3 (0.07)
chr15 58552284-58552403	AQP9 (0.29)	**chr16 21314303-21314411**	RUNDC2B (0)
chr15 60181631-60181734	FOXB1 (0.35)	**chr16 22871503-22871621**	HS3ST2 (0)
chr15 62574069-62574188	C2CD4B (0)	**chr16 22910970-22911074**	HS3ST2 (0)
chr15 65536702-65536820	PARP16 (3.7)	**chr16 23715753-23715866**	ERN2 (0.02)
chr15 69745127-69745247	KIF23 (5.65)	**chr16 26644127-26644232**	C16orf82 (0)
chr15 70656463-70656581	TLE3 (2.88)	**chr16 27258628-27258717**	NSMCE1 (5.3)
chr15 70885590-70885702	UACA (0.01)	**chr16 27401056-27401157**	IL21R (5.7)
chr15 71983002-71983105	THSD4 (0.02)	**chr16 29525594-29525702**	LOC440354 (0)
chr15 75230267-75230432	RPP25 (0.56)	**chr16 30265579-30265687**	LOC613037 (0.68)
chr15 76721437-76721540	SCAPER (3.02)	**chr16 31123645-31123810**	KAT8 (4.76)
chr15 80976586-80976704	FAM108C1 (0)	**chr16 54907984-54908103**	CRNDE (4.44)
chr15 81096291-81096409	KIAA1199 (0)	**chr16 55106596-55106711**	IRX5 (1.37)
chr15 87581908-87582042	AGBL1 (0)	**chr16 55331542-55331660**	IRX6 (0)
chr15 88076999-88077114	LINC00052 (0)	**chr16 55935090-55935208**	CES5A (0)
chr15 88186360-88186457	LINC00052 (0)	**chr16 56762981-56763084**	NUP93 (5.51)
chr15 88246997-88247116	LINC00052 (0)	**chr16 58455184-58455322**	GINS3 (3.79)
chr15 89198642-89198730	ISG20 (4.56)	**chr16 59721352-59721445**	LOC644649 (0)
chr15 91284016-91284134	BLM (4.4)	**chr16 60118537-60118658**	LOC644649 (0)
chr15 91803964-91804085	SV2B (0.05)	**chr16 66365666-66365772**	CDH5 (0.01)
chr15 93173107-93173208	FAM174B (1.84)	**chr16 69355935-69356037**	COG8 (4.51)
chr15 95232723-95232813	MCTP2 (0.01)	**chr16 70912086-70912205**	HYDIN (0.02)
chr15 95749213-95749314	LOC400456 (0)	**chr16 75918806-75918909**	TERF2IP (5.84)
chr15 95751457-95751570	LOC400456 (0)	**chr16 76268959-76269087**	CNTNAP4 (0.04)
chr15 96537241-96537376	NR2F2 (2.44)	**chr16 78208089-78208206**	WWOX (4.9)
chr15 97795260-97795360	SPATA8 (0)	**chr16 80926395-80926514**	C16orf61 (0)

chr16 82142915-82143029	HSD17B2 (0)	**chr18 10341456-10341568**	APCDD1 (0)
chr16 82721058-82721164	CDH13 (0)	**chr18 13281972-13282054**	C18orf1 (0)
chr16 86380948-86381054	LOC732275 (0)	**chr18 13998316-13998424**	ZNF519 (2.48)
chr16 86632908-86633046	FOXL1 (0.07)	**chr18 15135167-15135282**	LOC644669 (0.02)
chr16 87294579-87294677	C16orf95 (2.55)	**chr18 20452939-20453087**	RBBP8 (5.16)
chr16 88876084-88876229	APRT (7.15)	**chr18 22537538-22537657**	ZNF521 (0.05)
chr17 5522646-5522773	NLRP1 (4.87)	**chr18 28720938-28721057**	DSC1 (0)
chr17 7529674-7529768	SHBG (0.04)	**chr18 28794449-28794566**	DSC1 (0)
chr17 8892074-8892171	PIK3R5 (1.49)	**chr18 29545595-29545710**	TRAPPC8 (4.37)
chr17 10170290-10170409	MYH13 (0)	**chr18 32485083-32485174**	DTNA (1.94)
chr17 10195103-10195230	MYH13 (0)	**chr18 33069296-33069402**	MIR3975 (0)
chr17 20515269-20515414	CDRT15L2 (0)	**chr18 33325342-33325446**	GALNT1 (5.29)
chr17 25702008-25702111	LOC440419 (0)	**chr18 36053200-36053300**	LOC647946 (0)
chr17 27717445-27717563	MIR4523 (0)	**chr18 36093082-36093194**	LOC647946 (0)
chr17 30771499-30771623	PSMD11 (5.2)	**chr18 43454452-43454558**	EPG5 (2.37)
chr17 31279721-31279828	TMEM98 (0)	**chr18 44812046-44812167**	IER3IP1 (5.42)
chr17 35704142-35704243	ACACA (4.15)	**chr18 47982707-47982825**	SKA1 (4.18)
chr17 38946998-38947101	KRT28 (0)	**chr18 50287481-50287567**	DCC (0)
chr17 42022291-42022387	PPY (0)	**chr18 51188649-51188764**	DCC (0)
chr17 44488571-44488701	ARL17A (0.39)	**chr18 53849326-53849432**	LOC100505474 (0)
chr17 44706151-44706269	NSF (3.34)	**chr18 54833283-54833402**	BOD1P (0)
chr17 46736655-46736765	MIR196A1 (0)	**chr18 55558335-55558453**	ATP8B1 (0.71)
chr17 49498016-49498121	UTP18 (5.55)	**chr18 55636098-55636214**	NEDD4L (3.2)
chr17 50375185-50375290	CA10 (0.66)	**chr18 55664166-55664276**	NEDD4L (3.2)
chr17 53451535-53451756	MMD (4.08)	**chr18 55879585-55879690**	NEDD4L (3.2)
chr17 55229433-55229539	AKAP1 (4.24)	**chr18 56447362-56447466**	MALT1 (6.3)
chr17 59322437-59322596	BCAS3 (3.21)	**chr18 57333311-57333414**	CCBE1 (0.02)
chr17 63958878-63958991	CEP112 (0.74)	**chr18 57730289-57730380**	PMAIP1 (5.72)
chr17 64410444-64410534	PRKCA (4.93)	**chr18 58124559-58124668**	MC4R (0.53)
chr17 64934372-64934491	CACNG4 (0.02)	**chr18 58210303-58210401**	MC4R (0.53)
chr17 66823157-66823272	ABCA8 (0.33)	**chr18 59702329-59702430**	PIGN (3.22)
chr17 68979387-68979502	KCNJ2 (3.44)	**chr18 61428946-61429054**	SERPINB7 (0)
chr17 69227585-69227682	SOX9 (1.45)	**chr18 68975443-68975546**	LOC100505776 (0)
chr17 69617077-69617177	SOX9 (1.45)	**chr18 69552676-69552800**	LOC100505776 (0)
chr17 70002228-70002337	SOX9 (1.45)	**chr18 69776943-69777067**	CBLN2 (0.02)
chr17 70016639-70016760	SOX9 (1.45)	**chr18 69941156-69941270**	CBLN2 (0.02)
chr17 70032649-70032764	SOX9 (1.45)	**chr18 70147098-70147198**	CBLN2 (0.02)
chr17 70063246-70063359	SOX9 (1.45)	**chr18 71197412-71197518**	LOC100505817 (0)
chr17 70340935-70341050	LOC100499467 (0)	**chr18 72234153-72234274**	CNDP1 (0)
chr17 70421503-70421612	LOC100499467 (0)	**chr18 75207172-75207301**	GALR1 (0)
chr17 70435831-70435937	LOC100499467 (0)	**chr18 75437137-75437252**	GALR1 (0)
chr17 70493443-70493540	LOC100499467 (0)	**chr18 75705586-75705708**	GALR1 (0)
chr17 70510422-70510528	LOC100499467 (0)	**chr19 1085162-1085250**	POLR2E (4.89)
chr17 70989027-70989153	SLC39A11 (3.68)	**chr19 2041859-2042015**	MKNK2 (6.66)
chr17 71146870-71146990	SSTR2 (0.03)	**chr19 6323451-6323576**	CLPP (5.57)
chr17 72514915-72515024	CD300LB (0)	**chr19 6381159-6381241**	GTF2F1 (5.16)
chr17 73262119-73262280	MIF4GD (4.13)	**chr19 6381340-6381431**	GTF2F1 (5.16)
chr17 74560626-74560724	ST6GALNAC2 (0.62)	**chr19 6413849-6414014**	KHSRP (5.17)
chr17 76095495-76095602	TNRC6C (1.43)	**chr19 14751624-14751752**	EMR3 (0)
chr17 76663088-76663183	CYTH1 (6.01)	**chr19 23887340-23887460**	ZNF675 (3.87)
chr17 79336004-79336127	TMEM105 (0)	**chr19 28620111-28620219**	LOC148189 (0)
chr18 308291-308409	COLEC12 (0)	**chr19 34046707-34046814**	PEPD (4.69)
chr18 1275070-1275178	LINC00470 (0)	**chr19 35090318-35090437**	SCGBL (0)
chr18 1928434-1928541	LINC00470 (0)	**chr19 35284902-35285014**	ZNF599 (1.42)
chr18 4216086-4216252	DLGAP1 (0)	**chr19 46743644-46743762**	IGFL1 (0)
chr18 5270339-5270454	ZFP161 (0)	**chr19 49792758-49792876**	SLC6A16 (0.33)

chr19 49993711-49993864	SNORD33 (12.13)	**chr2 62502615-62502715**	B3GNT2 (4.75)
chr19 50363849-50363968	PNKP (3.31)	**chr2 64261720-64261823**	VPS54 (5.11)
chr2 184856-184963	SH3YL1 (2.72)	**chr2 65275117-65275277**	CEP68 (3.51)
chr2 316125-316229	FAM150B (0.66)	**chr2 65943129-65943232**	SPRED2 (2.41)
chr2 668027-668116	LOC339822 (0)	**chr2 67517096-67517215**	LOC644838 (0)
chr2 1318526-1318632	SNTG2 (0)	**chr2 69383723-69383824**	ANTXR1 (0.03)
chr2 1445809-1445922	SNTG2 (0)	**chr2 70645610-70645727**	TGFA (0.24)
chr2 4091301-4091386	LOC100505964 (0)	**chr2 72126693-72126800**	DYSF (0.59)
chr2 6101976-6102124	LOC400940 (0)	**chr2 74154009-74154137**	DGUOK (5.98)
chr2 7860370-7860470	LOC339788 (0)	**chr2 75769209-75769330**	FAM176A (0)
chr2 8072077-8072221	LINC00299 (0)	**chr2 76215892-76216001**	C2orf3 (0)
chr2 9255038-9255140	ASAP2 (0.97)	**chr2 77564717-77564814**	LRRTM4 (0.4)
chr2 9818806-9818916	YWHAQ (8.77)	**chr2 78274128-78274232**	SNAR-H (0)
chr2 11037122-11037223	KCNF1 (0)	**chr2 79000887-79001004**	REG3G (0)
chr2 11720099-11720217	MIR4429 (0)	**chr2 79213618-79213737**	REG3G (0)
chr2 12007268-12007410	MIR4262 (0)	**chr2 80258492-80258604**	CTNNA2 (0.22)
chr2 14131255-14131387	FAM84A (0.01)	**chr2 84402081-84402187**	FUNDC2P2 (0.14)
chr2 16048806-16048921	MYCNOS (0)	**chr2 85590817-85590899**	ELMOD3 (3.19)
chr2 16995366-16995481	FAM49A (1.55)	**chr2 86082383-86082486**	ST3GAL5 (5.06)
chr2 17093091-17093203	FAM49A (1.55)	**chr2 87855332-87855441**	LINC00152 (5.04)
chr2 18110579-18110674	KCNS3 (0)	**chr2 89113051-89113167**	MIR4436A (0)
chr2 20263110-20263232	LAPTM4A (5.27)	**chr2 99229898-99230002**	COA5 (4.36)
chr2 21385753-21385871	APOB (0)	**chr2 104629698-104629823**	LOC100287010 (0)
chr2 21657429-21657530	LOC645949 (0.14)	**chr2 108004251-108004335**	LOC729121 (0)
chr2 23312478-23312579	KLHL29 (0.31)	**chr2 110294435-110294550**	40422 (0)
chr2 24674827-24674938	ITSN2 (4.66)	**chr2 110971202-110971308**	NPHP1 (1.78)
chr2 24773241-24773358	NCOA1 (4.31)	**chr2 111881302-111881390**	ACOXL (1.43)
chr2 25067850-25067970	CENPO (4.11)	**chr2 111969407-111969508**	BCL2L11 (3.96)
chr2 27907986-27908097	SLC4A1AP (5.32)	**chr2 112424662-112424776**	ANAPC1 (4.36)
chr2 30554532-30554660	LBH (3.21)	**chr2 114448623-114448728**	SLC35F5 (4.11)
chr2 33931891-33932027	MYADML (0)	**chr2 116514941-116515055**	DPP10 (0.01)
chr2 36275760-36275875	LOC100288911 (0)	**chr2 117359539-117359671**	DPP10 (0.01)
chr2 40010065-40010177	THUMPD2 (3.94)	**chr2 118781970-118782077**	CCDC93 (4.19)
chr2 40808088-40808184	SLC8A1 (3.55)	**chr2 121014122-121014225**	TMEM185B (3.93)
chr2 41517312-41517415	LOC388942 (0)	**chr2 123442080-123442199**	TSN (6.24)
chr2 43317979-43318094	ZFP36L2 (2.44)	**chr2 123545441-123545542**	TSN (6.24)
chr2 44275097-44275215	LRPPRC (6.81)	**chr2 123811164-123811253**	CNTNAP5 (0)
chr2 44524018-44524137	PREPL (5.17)	**chr2 124350042-124350153**	CNTNAP5 (0)
chr2 47081263-47081366	LOC100134259 (0)	**chr2 126822370-126822490**	GYPC (6.01)
chr2 49305304-49305417	FSHR (0.19)	**chr2 128586998-128587114**	POLR2D (4.6)
chr2 50243647-50243748	NRXN1 (0.26)	**chr2 128721337-128721441**	SAP130 (3.32)
chr2 52334372-52334481	NRXN1 (0.26)	**chr2 132737349-132737467**	ANKRD30BL (0)
chr2 53155999-53156176	ASB3 (4.3)	**chr2 140700407-140700514**	LRP1B (0.08)
chr2 53641860-53641969	ASB3 (4.3)	**chr2 143874853-143874977**	ARHGAP15 (0.29)
chr2 53885710-53885810	ASB3 (4.3)	**chr2 144852044-144852147**	ZEB2 (4.63)
chr2 54342736-54342833	ACYP2 (1.97)	**chr2 146923044-146923152**	PABPC1P2 (1.92)
chr2 54501246-54501358	ACYP2 (1.97)	**chr2 148233321-148233437**	ACVR2A (1.08)
chr2 54621638-54621758	C2orf73 (0.06)	**chr2 148549202-148549311**	ACVR2A (1.08)
chr2 54908745-54908896	SPTBN1 (1.41)	**chr2 151986540-151986656**	RBM43 (1.64)
chr2 55757335-55757443	SMEK2 (5.36)	**chr2 152470256-152470378**	NEB (0.11)
chr2 56503657-56503789	CCDC85A (1.18)	**chr2 153573867-153573968**	ARL6IP6 (3.24)
chr2 57189213-57189320	CCDC85A (1.18)	**chr2 155062121-155062248**	GALNT13 (0.17)
chr2 57858301-57858431	VRK2 (4.96)	**chr2 155342895-155343014**	GALNT13 (0.17)
chr2 58083391-58083493	VRK2 (4.96)	**chr2 155361898-155362019**	GALNT13 (0.17)
chr2 58972566-58972675	FLJ30838 (0)	**chr2 157881118-157881233**	GALNT5 (0.03)
chr2 60343702-60343810	MIR4432 (0)	**chr2 158033328-158033427**	GALNT5 (0.03)

chr2 158535723-158535832	ACVR1C (0.88)	**chr2 234505178-234505280**	UGT1A8 (0)
chr2 159145379-159145487	CCDC148 (0.41)	**chr2 234718914-234719033**	HJURP (4.44)
chr2 160266991-160267095	BAZ2B (0.53)	**chr2 234896491-234896587**	TRPM8 (0)
chr2 160916227-160916341	PLA2R1 (2.99)	**chr2 235095754-235095857**	SPP2 (0)
chr2 161986910-161986998	TANK (5.78)	**chr2 235781887-235781996**	SH3BP4 (0.04)
chr2 162419066-162419179	SLC4A10 (0.09)	**chr2 237409720-237409817**	IQCA1 (0)
chr2 164518273-164518394	KCNH7 (0.05)	**chr2 237799438-237799532**	COPS8 (4.89)
chr2 165856169-165856265	SLC38A11 (0.04)	**chr2 238125828-238125942**	COL6A3 (0.01)
chr2 166629797-166629903	CSRNP3 (0)	**chr2 238363276-238363390**	MLPH (0.03)
chr2 166994033-166994144	SCN9A (0)	**chr2 238378153-238378266**	MLPH (0.03)
chr2 168797605-168797731	STK39 (5.54)	**chr2 238960152-238960267**	UBE2F (3.56)
chr2 169272637-169272730	CERS6 (0)	**chr20 807147-807258**	FAM110A (1.94)
chr2 169280674-169280783	CERS6 (0)	**chr20 1772431-1772544**	LOC100289473 (0)
chr2 170543440-170543560	C2orf77 (0)	**chr20 5333002-5333122**	PROKR2 (0)
chr2 173750244-173750363	RAPGEF4 (0.12)	**chr20 10337881-10337994**	MKKS (2.88)
chr2 176249781-176249868	ATP5G3 (5.97)	**chr20 10773554-10773669**	JAG1 (0.9)
chr2 177293859-177293983	MTX2 (4.68)	**chr20 10799147-10799253**	JAG1 (0.9)
chr2 178507650-178507765	PDE11A (0)	**chr20 15569695-15569814**	MACROD2 (0.05)
chr2 179090373-179090489	OSBPL6 (1.62)	**chr20 15932721-15932836**	MACROD2 (0.05)
chr2 180795645-180795762	CWC22 (4.18)	**chr20 16447185-16447291**	KIF16B (2.95)
chr2 184637903-184638004	NUP35 (4.77)	**chr20 19169699-19169802**	SLC24A3 (0)
chr2 186272255-186272421	FSIP2 (0.59)	**chr20 20223715-20223830**	C20orf26 (0)
chr2 187638435-187638537	FAM171B (3.71)	**chr20 22529160-22529279**	LINC00261 (0)
chr2 190375513-190375612	WDR75 (6.63)	**chr20 23478252-23478392**	CST8 (0)
chr2 191573357-191573532	NAB1 (4.55)	**chr20 24876117-24876226**	CST7 (0.08)
chr2 192364110-192364226	MYO1B (0)	**chr20 33650951-33651063**	EDEM2 (5.18)
chr2 192610059-192610166	OBFC2A (0)	**chr20 34547342-34547441**	SCAND1 (5.31)
chr2 193582278-193582394	PCGEM1 (0)	**chr20 34620301-34620409**	C20orf152 (0)
chr2 195427214-195427313	SLC39A10 (2.16)	**chr20 37915034-37915153**	LOC339568 (0)
chr2 196972884-196973003	STK17B (3.69)	**chr20 39447152-39447258**	MAFB (1.5)
chr2 197475976-197476083	HECW2 (0)	**chr20 40256554-40256653**	CHD6 (3.93)
chr2 198351754-198351881	HSPD1 (8)	**chr20 42386546-42386656**	GTSF1L (0)
chr2 198453984-198454092	RFTN2 (0.18)	**chr20 42457712-42457815**	TOX2 (0)
chr2 203683398-203683498	ICA1L (2.36)	**chr20 42670726-42670844**	TOX2 (0)
chr2 205278781-205278896	PARD3B (0.12)	**chr20 43493102-43493205**	YWHAB (6.61)
chr2 206489307-206489426	PARD3B (0.12)	**chr20 45313032-45313147**	SLC2A10 (0)
chr2 206843616-206843727	INO80D (2.22)	**chr20 47373375-47373493**	ARFGEF2 (4.21)
chr2 207101364-207101503	GPR1 (0.27)	**chr20 50613837-50613955**	ZFP64 (2.67)
chr2 210661398-210661515	UNC80 (0)	**chr20 50650372-50650530**	ZFP64 (2.67)
chr2 211143700-211143818	MYL1 (0)	**chr20 52716969-52717070**	BCAS1 (5.77)
chr2 214195618-214195727	SPAG16 (3.69)	**chr20 55453505-55453610**	TFAP2C (0.01)
chr2 215986515-215986626	ABCA12 (0.12)	**chr20 55474561-55474658**	TFAP2C (0.01)
chr2 216848784-216848897	PECR (5.17)	**chr20 56058420-56058539**	CTCFL (0)
chr2 217211722-217211815	SMARCAL1 (3.27)	**chr20 56343875-56343980**	PMEPA1 (0)
chr2 223623786-223623882	MOGAT1 (0)	**chr20 57395711-57395814**	GNAS-AS1 (0.77)
chr2 223852857-223852972	ACSL3 (4.94)	**chr20 58487915-58488010**	SYCP2 (4.25)
chr2 224532898-224533011	SCG2 (0.05)	**chr20 60028955-60029075**	CDH4 (0.02)
chr2 224604011-224604129	AP1S3 (2.23)	**chr21 14555344-14555459**	ANKRD30BP2 (0)
chr2 224923580-224923693	SERPINE2 (0.05)	**chr21 16042596-16042712**	SAMSN1 (5.92)
chr2 226307618-226307724	KIAA1486 (0)	**chr21 16052035-16052153**	SAMSN1 (5.92)
chr2 227408293-227408413	IRS1 (1.41)	**chr21 17005826-17005935**	USP25 (4.45)
chr2 227965887-227965986	COL4A3 (0)	**chr21 17071913-17072007**	USP25 (4.45)
chr2 228512348-228512464	C2orf83 (0)	**chr21 17296952-17297059**	USP25 (4.45)
chr2 228970442-228970562	SPHKAP (0)	**chr21 17832236-17832339**	LINC00478 (0)
chr2 233227045-233227149	ALPP (0)	**chr21 19371847-19371966**	CHODL (0.69)
chr2 234503043-234503162	UGT1A8 (0)	**chr21 19858111-19858229**	TMPRSS15 (4.3)

chr21 20599549-20599650	TMPRSS15 (4.3)	**chr3 4467495-4467600**	ITPR1 (4.93)
chr21 20834126-20834303	TMPRSS15 (4.3)	**chr3 6210151-6210263**	GRM7 (0)
chr21 20873380-20873542	TMPRSS15 (4.3)	**chr3 6277500-6277603**	GRM7 (0)
chr21 23374333-23374437	LINC00308 (0)	**chr3 6917133-6917251**	LOC100288428 (0)
chr21 23406656-23406775	LINC00308 (0)	**chr3 12882891-12883013**	RPL32 (9.04)
chr21 24131048-24131158	D21S2088E (0.27)	**chr3 16154026-16154132**	GALNTL2 (0)
chr21 25309325-25309445	D21S2088E (0.27)	**chr3 17065000-17065082**	PLCL2 (3.74)
chr21 25358760-25358868	D21S2088E (0.27)	**chr3 17245159-17245280**	TBC1D5 (4.32)
chr21 26839466-26839572	LINC00158 (1.5)	**chr3 18343483-18343595**	SATB1 (3.36)
chr21 26874728-26874851	MIR155HG (6.77)	**chr3 20903851-20903974**	VENTXP7 (0)
chr21 27727431-27727534	CYYR1 (1.33)	**chr3 21377362-21377468**	VENTXP7 (0)
chr21 27842836-27842938	APP (0.35)	**chr3 22628939-22629029**	UBE2E2 (1.84)
chr21 28372027-28372132	ADAMTS5 (0.26)	**chr3 26584188-26584278**	LRRC3B (0)
chr21 29244367-29244470	LINC00113 (0)	**chr3 30108004-30108107**	RBMS3 (0.34)
chr21 30577507-30577607	C21orf7 (0)	**chr3 32972923-32973036**	CCR4 (4.57)
chr21 30638221-30638343	BACH1 (4.45)	**chr3 43665994-43666099**	ANO10 (2.53)
chr21 30806576-30806695	BACH1 (4.45)	**chr3 48402343-48402460**	FBXW12 (0)
chr21 31330520-31330627	GRIK1 (0.19)	**chr3 51074312-51074415**	DOCK3 (1.4)
chr21 31799631-31799728	KRTAP13-3 (0)	**chr3 51792252-51792371**	IQCF6 (1.31)
chr21 32652652-32652767	TIAM1 (1.05)	**chr3 53231288-53231394**	PRKCD (2.85)
chr21 32745693-32745817	TIAM1 (1.05)	**chr3 54546899-54547000**	CACNA2D3 (0)
chr21 32777872-32777971	TIAM1 (1.05)	**chr3 55330875-55330967**	WNT5A (1.1)
chr21 33139478-33139583	SCAF4 (3.3)	**chr3 55539200-55539305**	ERC2 (0.24)
chr21 33903282-33903397	C21orf63 (0)	**chr3 55875381-55875487**	ERC2 (0.24)
chr21 35284586-35284699	LOC100506334 (0)	**chr3 60031038-60031159**	FHIT (0.79)
chr21 42077432-42077547	DSCAM (0)	**chr3 61504864-61504993**	PTPRG (3.88)
chr21 42274876-42274979	DSCAM (0)	**chr3 65329582-65329687**	MAGI1 (3.24)
chr21 42789589-42789704	MX1 (6.63)	**chr3 67561918-67562024**	SUCLG2 (4.72)
chr21 43017196-43017296	LINC00111 (0)	**chr3 72529272-72529368**	RYBP (3.84)
chr21 43881506-43881598	RSPH1 (0.18)	**chr3 72529491-72529575**	RYBP (3.84)
chr21 45191563-45191679	CSTB (5.04)	**chr3 74181980-74182091**	CNTN3 (0.09)
chr21 46224477-46224565	SUMO3 (5.51)	**chr3 76949168-76949325**	ROBO2 (0.87)
chr21 47013482-47013648	SLC19A1 (2.25)	**chr3 76967439-76967558**	ROBO2 (0.87)
chr21 47278810-47278896	PCBP3 (0.66)	**chr3 77033079-77033189**	ROBO2 (0.87)
chr21 47460876-47460983	COL6A1 (0)	**chr3 77707853-77707963**	ROBO2 (0.87)
chr21 47556862-47556980	FTCD (0.02)	**chr3 79363575-79363691**	ROBO1 (0.88)
chr21 47673880-47674013	MCM3AP (4.46)	**chr3 79531543-79531711**	ROBO1 (0.88)
chr22 16162029-16162144	POTEH (0.13)	**chr3 80577750-80577854**	ROBO1 (0.88)
chr22 28760240-28760343	CHEK2 (3.17)	**chr3 81808146-81808255**	GBE1 (3.95)
chr22 33355197-33355311	SYN3 (0.01)	**chr3 86473101-86473201**	CADM2 (0)
chr22 34186161-34186281	LARGE (1.87)	**chr3 88082278-88082380**	CGGBP1 (5.27)
chr22 34618938-34619064	LARGE (1.87)	**chr3 95835908-95836027**	EPHA6 (0.3)
chr22 35458737-35458857	ISX (0.42)	**chr3 95837839-95837956**	EPHA6 (0.3)
chr22 35626959-35627070	HMGXB4 (3.21)	**chr3 96156540-96156655**	EPHA6 (0.3)
chr22 36075763-36075876	APOL6 (4.09)	**chr3 99563660-99563786**	MIR548G (0)
chr22 40269980-40270103	ENTHD1 (0)	**chr3 100304825-100304907**	TMEM45A (0.51)
chr22 41347924-41348038	XPNPEP3 (2.29)	**chr3 102339952-102340064**	ZPLD1 (0)
chr22 45493403-45493518	LOC100506714 (0)	**chr3 102377762-102377877**	ZPLD1 (0)
chr22 46016641-46016752	FBLN1 (0)	**chr3 102444416-102444534**	ZPLD1 (0)
chr22 46067721-46067847	ATXN10 (4.06)	**chr3 106540146-106540252**	LOC100302640 (0)
chr22 47702799-47702892	LOC339685 (0)	**chr3 106921754-106921873**	LOC100302640 (0)
chr22 48054488-48054597	FLJ46257 (0)	**chr3 107843866-107843989**	CD47 (4)
chr22 49839316-49839424	C22orf34 (0.54)	**chr3 111071740-111071843**	PVRL3 (0)
chr3 278945-279038	CNTN6 (0.04)	**chr3 112032973-112033069**	CD200 (1.39)
chr3 3168701-3168802	IL5RA (4.86)	**chr3 112238004-112238119**	ATG3 (6.82)
chr3 3827416-3827529	LRRN1 (0.18)	**chr3 116858045-116858160**	LSAMP-AS3 (0)

chr3 116911884-116912003	LSAMP-AS3 (0)	**chr4 10679265-10679347**	CLNK (3.35)
chr3 118528782-118528905	IGSF11 (0.18)	**chr4 11569917-11570006**	HS3ST1 (1.63)
chr3 121734421-121734521	CD86 (6.63)	**chr4 11687592-11687692**	HS3ST1 (1.63)
chr3 125960813-125960916	ALDH1L1 (0.07)	**chr4 12763419-12763510**	HSP90AB2P (0)
chr3 131606075-131606190	CPNE4 (0.77)	**chr4 14809490-14809619**	LOC441009 (0)
chr3 134314082-134314197	KY (0.18)	**chr4 15519425-15519538**	CC2D2A (0.26)
chr3 141754585-141754710	TFDP2 (5.07)	**chr4 15521242-15521345**	CC2D2A (0.26)
chr3 144087399-144087513	C3orf58 (4.39)	**chr4 16242610-16242725**	FLJ39653 (0)
chr3 145759514-145759629	PLOD2 (3.54)	**chr4 16489657-16489773**	LDB2 (0.05)
chr3 148677831-148677946	GYG1 (4.55)	**chr4 17454168-17454272**	QDPR (2.71)
chr3 149175976-149176077	TM4SF4 (0.02)	**chr4 17480257-17480377**	QDPR (2.71)
chr3 150321093-150321230	SELT (5.31)	**chr4 20338836-20338939**	SLIT2 (0.01)
chr3 153709387-153709494	ARHGEF26-AS1 (0.13)	**chr4 24102853-24102953**	PPARGC1A (0.03)
chr3 156716844-156716945	LEKR1 (1.23)	**chr4 26566771-26566872**	TBC1D19 (1.61)
chr3 158347110-158347227	GFM1 (5.45)	**chr4 29173579-29173682**	MIR4275 (0)
chr3 158611965-158612080	MFSD1 (4.54)	**chr4 30554013-30554130**	PCDH7 (0.04)
chr3 158770767-158770885	IQCJ (0.02)	**chr4 31217610-31217716**	PCDH7 (0.04)
chr3 159922615-159922709	C3orf80 (0.09)	**chr4 34898763-34898885**	ARAP2 (2.49)
chr3 160479536-160479655	PPM1L (3.6)	**chr4 35258081-35258200**	ARAP2 (2.49)
chr3 162366365-162366467	LOC647107 (0)	**chr4 35503512-35503630**	ARAP2 (2.49)
chr3 163289161-163289280	LOC647107 (0)	**chr4 35596342-35596445**	ARAP2 (2.49)
chr3 166330671-166330789	ZBBX (0.37)	**chr4 38552282-38552417**	FLJ13197 (0)
chr3 166801087-166801205	ZBBX (0.37)	**chr4 38824545-38824670**	TLR6 (0.33)
chr3 168494265-168494374	EGFEM1P (0.14)	**chr4 40345338-40345424**	RBM47 (0.04)
chr3 169898619-169898774	PHC3 (3.17)	**chr4 40970986-40971105**	APBB2 (0.16)
chr3 172295179-172295307	NCEH1 (3.02)	**chr4 41848119-41848214**	TMEM33 (3.26)
chr3 172296159-172296300	NCEH1 (3.02)	**chr4 42463412-42463595**	ATP8A1 (3.68)
chr3 172298047-172298162	NCEH1 (3.02)	**chr4 42465377-42465492**	ATP8A1 (3.68)
chr3 172340113-172340207	NCEH1 (3.02)	**chr4 42980969-42981058**	GRXCR1 (0.04)
chr3 172671978-172672102	SPATA16 (0.02)	**chr4 45735752-45735865**	GABRG1 (0.17)
chr3 173559827-173559927	NLGN1 (1.26)	**chr4 47206716-47206819**	GABRB1 (0.11)
chr3 173745786-173745904	NAALADL2 (0.69)	**chr4 57694756-57694875**	SPINK2 (0.07)
chr3 174991893-174992029	NAALADL2 (0.69)	**chr4 58142274-58142397**	LOC255130 (0)
chr3 175453316-175453406	NAALADL2 (0.69)	**chr4 58729403-58729518**	LOC255130 (0)
chr3 176175057-176175170	TBL1XR1 (5.41)	**chr4 59024205-59024330**	LOC255130 (0)
chr3 177345991-177346106	TBL1XR1 (5.41)	**chr4 59812607-59812725**	LOC255130 (0)
chr3 179344882-179344989	NDUFB5 (6.54)	**chr4 61911259-61911375**	LPHN3 (0)
chr3 179803419-179803525	PEX5L (3.57)	**chr4 63360970-63361077**	LPHN3 (0)
chr3 180448260-180448373	CCDC39 (1.37)	**chr4 63391558-63391678**	LPHN3 (0)
chr3 182779789-182779915	LAMP3 (2.43)	**chr4 73079595-73079714**	NPFFR2 (0)
chr3 182796648-182796750	LAMP3 (2.43)	**chr4 73780243-73780350**	COX18 (2.74)
chr3 183079289-183079399	MCF2L2 (1.83)	**chr4 76350013-76350114**	RCHY1 (1.92)
chr3 183743156-183743268	HTR3D (0)	**chr4 76456879-76456992**	THAP6 (2.56)
chr3 184175733-184175834	CHRD (0)	**chr4 77446337-77446446**	SHROOM3 (0.76)
chr3 185743988-185744104	ETV5 (1.2)	**chr4 77791140-77791271**	ANKRD56 (0)
chr3 189121157-189121294	TPRG1 (3.88)	**chr4 77870079-77870194**	40787 (4.33)
chr3 190934883-190934971	UTS2D (0)	**chr4 80050034-80050147**	LOC100505875 (0)
chr3 191018561-191018667	CCDC50 (3.83)	**chr4 80655579-80655674**	GDEP (0)
chr3 194241781-194241883	FLJ34208 (0)	**chr4 84209739-84209858**	COQ2 (3.06)
chr3 195913921-195914038	ZDHHC19 (0.03)	**chr4 86886307-86886411**	ARHGAP24 (2.14)
chr3 196507224-196507327	PAK2 (4.46)	**chr4 87261516-87261612**	MAPK10 (0.69)
chr4 53196-53297	ZNF876P (0)	**chr4 88654238-88654344**	IBSP (0)
chr4 3915920-3916038	FAM86EP (1.14)	**chr4 93082479-93082597**	GRID2 (1.05)
chr4 4605314-4605426	LOC100507266 (0)	**chr4 95983104-95983224**	UNC5C (0.23)
chr4 5115538-5115651	STK32B (0.02)	**chr4 104337487-104337602**	TACR3 (0.08)
chr4 7755773-7755893	AFAP1 (2.94)	**chr4 104417100-104417218**	TACR3 (0.08)

chr4 106750617-106750737	NPNT (0.2)	**chr4 188154413-188154533**	LOC339975 (0.08)
chr4 107701379-107701485	DKK2 (0.09)	**chr4 189441863-189441971**	LOC401164 (0)
chr4 108446708-108446820	PAPSS1 (5.5)	**chr4 189444112-189444226**	LOC401164 (0)
chr4 108693606-108693719	SGMS2 (1.3)	**chr4 189893225-189893329**	LOC401164 (0)
chr4 109395605-109395705	LOC285456 (0)	**chr5 1318021-1318174**	CLPTM1L (6.27)
chr4 112378833-112378927	C4orf32 (2.2)	**chr5 1507741-1507856**	SLC6A3 (0)
chr4 112723789-112723911	C4orf32 (2.2)	**chr5 2058889-2059010**	IRX4 (0)
chr4 112993734-112993836	C4orf32 (2.2)	**chr5 2069744-2069864**	IRX4 (0)
chr4 113040984-113041083	C4orf32 (2.2)	**chr5 2115723-2115842**	IRX4 (0)
chr4 113203297-113203400	TIFA (5.53)	**chr5 2268717-2268836**	IRX4 (0)
chr4 115335689-115335793	UGT8 (0.96)	**chr5 2332283-2332394**	IRX2 (0)
chr4 116889234-116889339	MIR1973 (0)	**chr5 4526910-4527015**	LOC340094 (0)
chr4 117048159-117048278	MIR1973 (0)	**chr5 5006913-5007021**	LOC340094 (0)
chr4 118060028-118060127	TRAM1L1 (0)	**chr5 5039972-5040073**	ADAMTS16 (0.66)
chr4 118400844-118400943	TRAM1L1 (0)	**chr5 5789402-5789506**	KIAA0947 (0)
chr4 119495435-119495544	CEP170P1 (1.74)	**chr5 6056596-6056709**	FLJ33360 (0.19)
chr4 120221587-120221715	USP53 (3.29)	**chr5 6082280-6082380**	FLJ33360 (0.19)
chr4 120584052-120584148	PDE5A (5.11)	**chr5 6517903-6518049**	UBE2QL1 (0)
chr4 121756727-121756840	PRDM5 (0.08)	**chr5 6548437-6548534**	LOC255167 (0)
chr4 123001164-123001267	KIAA1109 (3.1)	**chr5 9949864-9949979**	LOC285692 (0.04)
chr4 123440504-123440632	IL2 (0)	**chr5 10311295-10311410**	CMBL (0.01)
chr4 125297655-125297775	ANKRD50 (1.9)	**chr5 11057547-11057649**	CTNND2 (0)
chr4 132879653-132879768	PCDH10 (0.62)	**chr5 12079996-12080109**	CTNND2 (0)
chr4 132978324-132978474	PCDH10 (0.62)	**chr5 13596636-13596734**	DNAH5 (0.28)
chr4 133184042-133184154	PCDH10 (0.62)	**chr5 14202806-14202920**	TRIO (5.41)
chr4 134283851-134283957	PCDH10 (0.62)	**chr5 15639558-15639657**	FBXL7 (0)
chr4 135286868-135286978	PABPC4L (0)	**chr5 16197516-16197619**	40603 (0)
chr4 138978550-138978669	LOC641364 (0)	**chr5 17259222-17259328**	LOC285696 (0.27)
chr4 141227266-141227384	SCOC (4.17)	**chr5 17860476-17860582**	LOC401177 (0)
chr4 141901827-141901923	TBC1D9 (0.02)	**chr5 20855999-20856116**	GUSBP1 (4.38)
chr4 142579326-142579432	IL15 (1.65)	**chr5 20922232-20922326**	GUSBP1 (4.38)
chr4 142807079-142807169	INPP4B (1.8)	**chr5 21659369-21659487**	GUSBP1 (4.38)
chr4 146543400-146543501	C4orf51 (0)	**chr5 22464242-22464356**	CDH12 (0.16)
chr4 149511611-149511729	NR3C2 (0.64)	**chr5 22618990-22619107**	CDH12 (0.16)
chr4 149680951-149681051	NR3C2 (0.64)	**chr5 22776047-22776165**	CDH12 (0.16)
chr4 153067696-153067818	FBXW7 (4.72)	**chr5 22922695-22922808**	CDH12 (0.16)
chr4 154815726-154815821	SFRP2 (0)	**chr5 25010801-25010919**	LOC340107 (0.02)
chr4 155657177-155657292	LRAT (0.01)	**chr5 26541851-26541954**	CDH9 (0.28)
chr4 156572039-156572150	GUCY1A3 (0.14)	**chr5 31975323-31975429**	PDZD2 (0.21)
chr4 156672824-156672939	GUCY1B3 (0.05)	**chr5 33162264-33162391**	LOC340113 (0)
chr4 159126872-159126969	TMEM144 (0)	**chr5 34280092-34280186**	C1QTNF3-AMACR
chr4 164336760-164336880	TKTL2 (0)	**chr5 34349766-34349883**	C1QTNF3-AMACR
chr4 166771215-166771324	TLL1 (0.64)	**chr5 35331315-35331471**	PRLR (0)
chr4 166813154-166813259	TLL1 (0.64)	**chr5 35560184-35560287**	SPEF2 (0.06)
chr4 167848505-167848625	SPOCK3 (0.46)	**chr5 35729928-35730047**	SPEF2 (0.06)
chr4 168580068-168580157	SPOCK3 (0.46)	**chr5 36277942-36278061**	RANBP3L (0.48)
chr4 168582003-168582106	SPOCK3 (0.46)	**chr5 37869232-37869337**	GDNF (0)
chr4 168971035-168971153	ANXA10 (0.03)	**chr5 38415878-38415978**	EGFLAM (0)
chr4 169457679-169457775	PALLD (0.28)	**chr5 38851984-38852103**	RICTOR (3.6)
chr4 170719791-170719897	C4orf27 (3.92)	**chr5 38868631-38868748**	RICTOR (3.6)
chr4 178653520-178653622	LOC285501 (0)	**chr5 40027659-40027746**	DAB2 (0.01)
chr4 179212629-179212737	LOC285501 (0)	**chr5 40225334-40225453**	PTGER4 (6.44)
chr4 184724163-184724266	C4orf41 (0)	**chr5 40532028-40532144**	PTGER4 (6.44)
chr4 185451439-185451544	IRF2 (4.52)	**chr5 40561036-40561154**	PTGER4 (6.44)
chr4 187680188-187680318	FAT1 (0.09)	**chr5 41690642-41690760**	OXCT1 (5.09)
chr4 188077180-188077282	LOC339975 (0.08)	**chr5 42101101-42101210**	FBXO4 (3.67)

chr5 43069013-43069134	LOC100132356 (2.74)	**chr5 151482691-151482806**	GLRA1 (0.11)
chr5 44397007-44397106	FGF10 (0)	**chr5 151695971-151696065**	NMUR2 (0)
chr5 44852994-44853112	MRPS30 (5.1)	**chr5 152317687-152317782**	NMUR2 (0)
chr5 49981573-49981676	PARP8 (2.38)	**chr5 155332779-155332879**	SGCD (0)
chr5 50205883-50205985	PARP8 (2.38)	**chr5 156816194-156816292**	CYFIP2 (4.78)
chr5 51202156-51202274	ISL1 (0.22)	**chr5 156825163-156825281**	CYFIP2 (4.78)
chr5 53651045-53651148	ARL15 (1.59)	**chr5 159863822-159863914**	PTTG1 (6.89)
chr5 53686589-53686765	HSPB3 (0)	**chr5 159929912-159930029**	MIR146A (0)
chr5 57808750-57808869	GAPT (0)	**chr5 166959420-166959523**	WWC1 (0.99)
chr5 66245259-66245364	MAST4 (0.17)	**chr5 167396327-167396442**	WWC1 (0.99)
chr5 66786403-66786520	CD180 (0.01)	**chr5 173014101-173014188**	LOC285593 (0.45)
chr5 66930448-66930551	CD180 (0.01)	**chr5 174452918-174453027**	MIR4634 (0)
chr5 67794767-67794870	PIK3R1 (4.36)	**chr5 174496162-174496271**	MIR4634 (0)
chr5 68604295-68604413	CDK7 (5.05)	**chr5 179216396-179216522**	LTC4S (0.95)
chr5 73129539-73129658	RGNEF (0)	**chr5 179264005-179264217**	C5orf45 (3.55)
chr5 75833894-75834012	IQGAP2 (0.25)	**chr6 3195914-3196037**	TUBB2B (1.62)
chr5 78160095-78160194	ARSB (2.69)	**chr6 5555330-5555497**	LYRM4 (3.34)
chr5 79114571-79114753	CMYA5 (0.16)	**chr6 9766834-9766947**	TFAP2A (0.45)
chr5 80924003-80924103	SSBP2 (3.18)	**chr6 10281393-10281510**	TFAP2A (0.45)
chr5 82113761-82113876	MIR3977 (0)	**chr6 10293780-10293875**	TFAP2A (0.45)
chr5 82307545-82307634	TMEM167A (4.35)	**chr6 12677095-12677215**	PHACTR1 (0.2)
chr5 89187883-89188002	MIR3660 (0)	**chr6 16990393-16990508**	FLJ23152 (0)
chr5 90581012-90581109	ARRDC3 (3.35)	**chr6 20184061-20184179**	E2F3 (2.76)
chr5 91206688-91206797	LOC100129716 (0)	**chr6 23961997-23962098**	NRSN1 (0.36)
chr5 95382469-95382582	MIR583 (0)	**chr6 25946123-25946232**	SLC17A2 (0)
chr5 99434278-99434381	LOC100133050 (0.22	**chr6 27686871-27686989**	LOC100507173 (0)
chr5 99708327-99708443	LOC100133050 (0.22)	**chr6 27714720-27714850**	LOC100131289 (0)
chr5 100208313-100208432	ST8SIA4 (4.66)	**chr6 30015025-30015121**	ZNRD1-AS1 (1.83)
chr5 100217513-100217625	ST8SIA4 (4.66)	**chr6 31021195-31021317**	HCG22 (0.25)
chr5 107090431-107090533	EFNA5 (0.21)	**chr6 32430708-32430827**	HLA-DRA (8.69)
chr5 111555013-111555116	EPB41L4A (1.07)	**chr6 32686694-32686798**	HLA-DQA2 (0.21)
chr5 115411010-115411162	COMMD10 (4.57)	**chr6 33048597-33048720**	HLA-DPB1 (4.54)
chr5 115664219-115664325	COMMD10 (4.57)	**chr6 37717371-37717473**	MDGA1 (0.08)
chr5 120183520-120183648	PRR16 (2.65)	**chr6 40741033-40741144**	LRFN2 (0)
chr5 120288457-120288574	PRR16 (2.65)	**chr6 50388982-50389100**	TFAP2D (0)
chr5 120500416-120500582	PRR16 (2.65)	**chr6 50774205-50774324**	TFAP2B (0.01)
chr5 121689337-121689454	SNCAIP (0.05)	**chr6 52033951-52034071**	IL17A (0)
chr5 123491350-123491459	ZNF608 (0.14)	**chr6 52129232-52129381**	MCM3 (7.88)
chr5 123748532-123748651	ZNF608 (0.14)	**chr6 52860576-52860715**	GSTA4 (1.4)
chr5 127541728-127541850	SLC12A2 (4.02)	**chr6 54604856-54604960**	FAM83B (0)
chr5 129630700-129630797	CHSY3 (0)	**chr6 55495538-55495647**	HMGCLL1 (0)
chr5 130802299-130802412	RAPGEF6 (4.19)	**chr6 55999567-55999668**	COL21A1 (0.12)
chr5 134860307-134860403	NEUROG1 (0)	**chr6 58364721-58364829**	GUSBP4 (1.96)
chr5 137381825-137381927	FAM13B (4.01)	**chr6 71085452-71085556**	FAM135A (2.17)
chr5 142589738-142589865	ARHGAP26 (1.74)	**chr6 71339471-71339574**	SMAP1 (2.92)
chr5 142970524-142970655	NR3C1 (4.93)	**chr6 71700263-71700362**	B3GAT2 (0.49)
chr5 144779312-144779427	PRELID2 (0.01)	**chr6 74677013-74677116**	CD109 (1.72)
chr5 144790723-144790809	PRELID2 (0.01)	**chr6 76941246-76941350**	IMPG1 (0)
chr5 145170670-145170772	GRXCR2 (0)	**chr6 77054352-77054453**	IMPG1 (0)
chr5 146717834-146717940	DPYSL3 (0.08)	**chr6 77440355-77440470**	IMPG1 (0)
chr5 147076722-147076842	JAKMIP2 (1.99)	**chr6 78271665-78271774**	HTR1B (0)
chr5 147218677-147218767	SPINK1 (0)	**chr6 78900393-78900499**	IRAK1BP1 (1.54)
chr5 148778488-148778594	MIR143HG (0)	**chr6 80096708-80096825**	LCA5 (0.47)
chr5 149472097-149472212	PDGFRB (0.1)	**chr6 81005039-81005127**	TTK (3.84)
chr5 149776186-149776341	CD74 (9.61)	**chr6 81662247-81662473**	BCKDHB (3.21)
chr5 151338401-151338520	GLRA1 (0.11)	**chr6 82247264-82247400**	FAM46A (1.54)

chr6 82804938-82805039	IBTK (3.07)	**chr6 150821865-150821992**	IYD (0)
chr6 83909289-83909408	RWDD2A (2.02)	**chr6 152492097-152492216**	SYNE1 (0.66)
chr6 85506230-85506333	TBX18 (0)	**chr6 155952338-155952462**	NOX3 (0)
chr6 91195840-91195943	MAP3K7 (2.83)	**chr6 158215253-158215359**	SNX9 (0)
chr6 91602269-91602384	MAP3K7 (2.83)	**chr6 160723841-160723946**	SLC22A2 (0)
chr6 91964490-91964593	MIR4643 (0)	**chr6 162163307-162163426**	PARK2 (0)
chr6 92685430-92685545	MIR4643 (0)	**chr6 164280077-164280178**	QKI (1.84)
chr6 94989565-94989662	TSG1 (0)	**chr6 164933423-164933539**	C6orf118 (0)
chr6 97203110-97203205	GPR63 (0.99)	**chr6 164942238-164942363**	C6orf118 (0)
chr6 98376247-98376365	MIR2113 (0)	**chr6 166913045-166913180**	RPS6KA2 (0.41)
chr6 101149221-101149330	ASCC3 (3.43)	**chr6 168193542-168193662**	C6orf124 (0)
chr6 102996032-102996142	GRIK2 (0.19)	**chr6 168493645-168493763**	FRMD1 (0)
chr6 107227590-107227689	LOC100422737 (0)	**chr7 1561148-1561266**	MAFK (2.99)
chr6 107229517-107229620	LOC100422737 (0)	**chr7 3188635-3188744**	CARD11 (0.03)
chr6 109681293-109681392	CD164 (5.33)	**chr7 3717022-3717132**	SDK1 (0)
chr6 111382666-111382783	GSTM2P1 (0)	**chr7 4008538-4008640**	SDK1 (0)
chr6 111612254-111612384	REV3L (1.94)	**chr7 4283120-4283238**	SDK1 (0)
chr6 111932660-111932763	TRAF3IP2 (0.11)	**chr7 7558770-7558866**	COL28A1 (0.04)
chr6 115034075-115034193	HS3ST5 (0.14)	**chr7 7865057-7865167**	LOC729852 (0)
chr6 115267885-115267990	HS3ST5 (0.14)	**chr7 8564376-8564477**	NXPH1 (0.37)
chr6 115409565-115409661	FRK (0.51)	**chr7 11606831-11606931**	THSD7A (0.21)
chr6 118184271-118184387	SLC35F1 (0.03)	**chr7 12082302-12082387**	TMEM106B (2.54)
chr6 119419348-119419454	FAM184A (0.49)	**chr7 12971249-12971396**	ARL4A (4.2)
chr6 120326828-120326957	MAN1A1 (2.08)	**chr7 16319545-16319648**	LOC100506025 (0)
chr6 124340116-124340222	NKAIN2 (0.04)	**chr7 20941927-20942041**	RPL23P8 (0)
chr6 125806803-125806895	HDDC2 (3.53)	**chr7 25115084-25115199**	CYCS (4.23)
chr6 127883832-127883934	C6orf58 (0.12)	**chr7 26540970-26541059**	LOC441204 (0)
chr6 128299549-128299646	PTPRK (0.14)	**chr7 26677956-26678064**	SKAP2 (2.95)
chr6 129085151-129085264	LAMA2 (0)	**chr7 27401188-27401313**	EVX1 (0)
chr6 131083089-131083200	LOC100507203 (0)	**chr7 32029920-32030025**	PDE1C (0.05)
chr6 131797818-131797926	ARG1 (0.18)	**chr7 34329154-34329316**	AAA1 (0)
chr6 132057480-132057586	ENPP3 (3.47)	**chr7 35341568-35341661**	LOC401324 (0)
chr6 132852309-132852427	TAAR9 (0)	**chr7 36031055-36031182**	39326 (4.38)
chr6 134433549-134433661	HMGA1P7 (0)	**chr7 36362214-36362304**	KIAA0895 (1.53)
chr6 134719760-134719853	LOC154092 (0)	**chr7 36683419-36683528**	AOAH (3.58)
chr6 135388195-135388292	HBS1L (3.23)	**chr7 37904355-37904470**	SFRP4 (0)
chr6 135972502-135972605	LINC00271 (0)	**chr7 38122171-38122276**	STARD3NL (4.2)
chr6 136048260-136048366	LINC00271 (0)	**chr7 38539479-38539597**	AMPH (0.36)
chr6 136546670-136546788	FAM54A (0)	**chr7 42422379-42422490**	GLI3 (0)
chr6 137314305-137314423	IL20RA (0)	**chr7 45412793-45412908**	RAMP3 (0)
chr6 137697522-137697629	OLIG3 (0)	**chr7 47833047-47833130**	C7orf69 (0)
chr6 139370342-139370460	C6orf115 (0)	**chr7 48169599-48169712**	UPP1 (0.25)
chr6 140908994-140909107	MIR4465 (0)	**chr7 50329994-50330211**	IKZF1 (5.99)
chr6 140916426-140916522	MIR4465 (0)	**chr7 53535984-53536102**	FLJ45974 (0)
chr6 141027317-141027415	MIR4465 (0)	**chr7 53837812-53837905**	FLJ45974 (0)
chr6 141386468-141386574	MIR4465 (0)	**chr7 56279908-56280009**	LOC389493 (0)
chr6 141752351-141752471	NMBR (0)	**chr7 63512753-63512852**	ZNF727 (0.08)
chr6 143292120-143292237	HIVEP2 (0)	**chr7 66071824-66071971**	KCTD7 (2.35)
chr6 143981409-143981524	PHACTR2 (1.04)	**chr7 67902596-67902704**	STAG3L4 (1.74)
chr6 144702811-144702915	UTRN (1.73)	**chr7 68782091-68782199**	AUTS2 (1.63)
chr6 144951533-144951638	UTRN (1.73)	**chr7 70419585-70419705**	AUTS2 (1.63)
chr6 147981815-147981931	SAMD5 (0)	**chr7 70713085-70713190**	WBSCR17 (0)
chr6 148337801-148337903	SASH1 (0.02)	**chr7 71222111-71222217**	CALN1 (0.01)
chr6 149188757-149188876	UST (1.02)	**chr7 80851353-80851454**	SEMA3C (0)
chr6 150580005-150580151	PPP1R14C (0)	**chr7 81140645-81140765**	HGF (1.39)
chr6 150814918-150815031	IYD (0)	**chr7 81229765-81229865**	HGF (1.39)

chr7 81315457-81315601	HGF (1.39)	**chr7 144859696-144859846**	TPK1 (1.37)
chr7 81962725-81962843	CACNA2D1 (0)	**chr7 145721365-145721471**	CNTNAP2 (0.09)
chr7 82203115-82203230	CACNA2D1 (0)	**chr7 145800390-145800496**	CNTNAP2 (0.09)
chr7 82225263-82225426	CACNA2D1 (0)	**chr7 145916108-145916211**	CNTNAP2 (0.09)
chr7 86883011-86883123	C7orf23 (0)	**chr7 147082066-147082169**	MIR548I4 (0)
chr7 89020756-89020868	ZNF804B (0)	**chr7 148086457-148086543**	MIR548T (0)
chr7 90068972-90069089	CLDN12 (0.9)	**chr7 148565409-148565525**	EZH2 (4.04)
chr7 91884312-91884425	ANKIB1 (3.63)	**chr7 151462297-151462391**	PRKAG2 (0.82)
chr7 92218106-92218226	CDK6 (4)	**chr7 151578411-151578535**	LOC100505483 (0)
chr7 92535484-92535618	CDK6 (4)	**chr7 151661584-151661691**	GALNT11 (0.16)
chr7 92572177-92572283	CDK6 (4)	**chr7 153193188-153193285**	DPP6 (0)
chr7 93652097-93652223	BET1 (2.14)	**chr7 154907505-154907611**	HTR5A (0)
chr7 93878939-93879051	COL1A2 (0)	**chr7 154998368-154998485**	INSIG1 (6.38)
chr7 96380375-96380465	SHFM1 (5.84)	**chr7 155037501-155037598**	INSIG1 (6.38)
chr7 96796585-96796682	ACN9 (0)	**chr7 156162761-156162862**	LOC285889 (0)
chr7 97088990-97089091	TAC1 (0)	**chr7 156330202-156330310**	LINC00244 (0)
chr7 97089539-97089658	TAC1 (0)	**chr7 158172533-158172705**	PTPRN2 (3.38)
chr7 99006234-99006348	BUD31 (4.49)	**chr8 1785043-1785163**	ARHGEF10 (0)
chr7 99229641-99229787	ZNF498 (0)	**chr8 2220367-2220470**	MYOM2 (0.2)
chr7 99532774-99532891	GJC3 (0.13)	**chr8 2573162-2573279**	CSMD1 (0.01)
chr7 99933463-99933584	PILRB (2.19)	**chr8 3888821-3888936**	CSMD1 (0.01)
chr7 105241233-105241323	ATXN7L1 (0.09)	**chr8 4632322-4632422**	CSMD1 (0.01)
chr7 105646453-105646568	CDHR3 (0)	**chr8 4768519-4768613**	CSMD1 (0.01)
chr7 106145149-106145272	C7orf74 (0)	**chr8 4774618-4774708**	CSMD1 (0.01)
chr7 112947879-112947998	LOC401397 (0)	**chr8 9794691-9794795**	MIR124-1 (0)
chr7 114423250-114423356	FOXP2 (1.41)	**chr8 11080485-11080578**	XKR6 (0)
chr7 114490185-114490314	MDFIC (2.3)	**chr8 14838052-14838152**	SGCZ (0.97)
chr7 116885885-116886002	ST7 (4.71)	**chr8 15199105-15199214**	SGCZ (0.97)
chr7 116908005-116908113	WNT2 (0)	**chr8 15274685-15274796**	TUSC3 (1.72)
chr7 117636035-117636198	CTTNBP2 (0)	**chr8 15364047-15364165**	TUSC3 (1.72)
chr7 118519275-118519374	ANKRD7 (0.46)	**chr8 16314468-16314593**	MSR1 (0.33)
chr7 118561350-118561466	ANKRD7 (0.46)	**chr8 16892587-16892692**	EFHA2 (0)
chr7 121403786-121403901	PTPRZ1 (0)	**chr8 16999622-16999741**	ZDHHC2 (3.06)
chr7 123193302-123193420	NDUFA5 (2.71)	**chr8 17456355-17456462**	PDGFRL (0.54)
chr7 123972571-123972718	TMEM229A (0)	**chr8 19075659-19075764**	SH2D4A (0.61)
chr7 124996219-124996327	POT1 (2.9)	**chr8 19973498-19973626**	SLC18A1 (0)
chr7 125121571-125121676	POT1 (2.9)	**chr8 21321247-21321350**	GFRA2 (0)
chr7 125141034-125141144	POT1 (2.9)	**chr8 22826596-22826720**	RHOBTB2 (1.67)
chr7 128761680-128761798	LOC407835 (0.23)	**chr8 24092605-24092723**	ADAM28 (0)
chr7 130016489-130016592	CPA1 (0)	**chr8 24958606-24958725**	DOCK5 (0.03)
chr7 130708012-130708131	MKLN1 (2.49)	**chr8 25215721-25215819**	DOCK5 (0.03)
chr7 131007219-131007301	MKLN1 (2.49)	**chr8 28314984-28315088**	FBXO16 (1.04)
chr7 131636733-131636843	PLXNA4 (0)	**chr8 28530464-28530583**	EXTL3 (2.08)
chr7 133550394-133550496	EXOC4 (3.68)	**chr8 29874155-29874274**	MIR548O2 (0)
chr7 134539713-134539826	CALD1 (0.07)	**chr8 30076043-30076149**	DCTN6 (3.64)
chr7 134847271-134847359	C7orf49 (3.82)	**chr8 30742891-30742990**	TEX15 (0.01)
chr7 135015914-135016024	CNOT4 (2.33)	**chr8 31408816-31408924**	NRG1 (0)
chr7 135843385-135843485	LUZP6 (5.11)	**chr8 31683926-31684044**	NRG1 (0)
chr7 139677036-139677165	TBXAS1 (1.38)	**chr8 33524627-33524733**	DUSP26 (0)
chr7 140746063-140746183	LOC100507421 (0)	**chr8 34662893-34663008**	UNC5D (0)
chr7 140895579-140895699	LOC100507421 (0)	**chr8 34730865-34730969**	UNC5D (0)
chr7 140980256-140980375	LOC100507421 (0)	**chr8 37340625-37340719**	ZNF703 (0.12)
chr7 141208400-141208523	LOC100507421 (0)	**chr8 40153209-40153328**	C8orf4 (0)
chr7 143715952-143716055	OR6B1 (0)	**chr8 49430553-49430672**	EFCAB1 (0)
chr7 143872842-143872945	CTAGE4 (0)	**chr8 52532155-52532273**	PXDNL (0.13)
chr7 144831255-144831361	TPK1 (1.37)	**chr8 58890930-58891049**	FAM110B (0)

chr8 63655416-63655515	NKAIN3 (0)	**chr8 144274105-144274218**	GPIHBP1 (0)
chr8 63686241-63686347	NKAIN3 (0)	**chr9 1930288-1930406**	SMARCA2 (6.3)
chr8 64079032-64079150	YTHDF3 (3.17)	**chr9 2388636-2388738**	FLJ35024 (0)
chr8 69961266-69961367	LOC100505718 (0)	**chr9 2676163-2676280**	VLDLR (0.04)
chr8 74288476-74288582	LOC100128126 (0)	**chr9 3742812-3742930**	GLIS3 (0.04)
chr8 75482540-75482658	FLJ39080 (0)	**chr9 5580630-5580726**	PDCD1LG2 (4.25)
chr8 76774200-76774318	HNF4G (0.96)	**chr9 9442123-9442283**	PTPRD (0.53)
chr8 80062451-80062554	IL7 (0.91)	**chr9 11393649-11393751**	PTPRD (0.53)
chr8 88823894-88824013	DCAF4L2 (0)	**chr9 16917482-16917596**	BNC2 (0.01)
chr8 90757036-90757131	RIPK2 (3.72)	**chr9 16983807-16983927**	BNC2 (0.01)
chr8 94079268-94079369	LOC389676 (0)	**chr9 17192909-17193029**	BNC2 (0.01)
chr8 94351961-94352087	LOC642924 (0)	**chr9 17889077-17889181**	SH3GL2 (0.57)
chr8 95103850-95103955	CDH17 (0)	**chr9 18333349-18333453**	ADAMTSL1 (0.05)
chr8 95481559-95481666	KIAA1429 (3.44)	**chr9 18340710-18340827**	ADAMTSL1 (0.05)
chr8 97439495-97439601	SDC2 (0.04)	**chr9 18693980-18694080**	ADAMTSL1 (0.05)
chr8 99362835-99362954	NIPAL2 (2.69)	**chr9 18701978-18702090**	ADAMTSL1 (0.05)
chr8 99417953-99418054	KCNS2 (0.08)	**chr9 18838071-18838191**	ADAMTSL1 (0.05)
chr8 103211647-103211753	RRM2B (2.32)	**chr9 19037526-19037632**	FAM154A (0)
chr8 104576013-104576116	RIMS2 (1.1)	**chr9 19252124-19252227**	DENND4C (3.79)
chr8 112386814-112386931	CSMD3 (0.11)	**chr9 19956526-19956639**	SLC24A2 (0.11)
chr8 114261342-114261443	CSMD3 (0.11)	**chr9 20584524-20584695**	KIAA1797 (0)
chr8 114589565-114589668	CSMD3 (0.11)	**chr9 22058806-22058926**	CDKN2B-AS (0)
chr8 116294158-116294246	TRPS1 (0.18)	**chr9 22197149-22197265**	CDKN2B-AS (0)
chr8 117688599-117688698	EIF3H (6.32)	**chr9 25462604-25462709**	TUSC1 (0.05)
chr8 117864193-117864298	RAD21-AS1 (0.13)	**chr9 26157484-26157604**	LOC100506422 (0)
chr8 117910451-117910570	RAD21-AS1 (0.13)	**chr9 26634190-26634293**	C9orf82 (0)
chr8 118275729-118275832	SLC30A8 (0)	**chr9 26748188-26748313**	C9orf82 (0)
chr8 118532938-118533058	EXT1 (2.52)	**chr9 26752854-26752977**	C9orf82 (0)
chr8 118776761-118776879	EXT1 (2.52)	**chr9 29724570-29724688**	MIR873 (0)
chr8 119919135-119919241	TNFRSF11B (0)	**chr9 30139770-30139887**	MIR873 (0)
chr8 119997049-119997166	TNFRSF11B (0)	**chr9 31578570-31578695**	ACO1 (3.85)
chr8 120835700-120835806	DSCC1 (2.24)	**chr9 32075067-32075170**	ACO1 (3.85)
chr8 123277325-123277443	ZHX2 (2.83)	**chr9 32316509-32316640**	ACO1 (3.85)
chr8 123579819-123579933	ZHX2 (2.83)	**chr9 32455358-32455440**	DDX58 (2.69)
chr8 123583738-123583830	ZHX2 (2.83)	**chr9 34367028-34367140**	KIAA1161 (0.24)
chr8 123973802-123973902	ZHX2 (2.83)	**chr9 39630751-39630859**	LOC653501 (0)
chr8 124685073-124685199	ANXA13 (0)	**chr9 41775780-41775890**	LOC653501 (0)
chr8 124724440-124724538	FAM91A1 (3.35)	**chr9 41902199-41902311**	MGC21881 (0)
chr8 124840804-124840911	FAM91A1 (3.35)	**chr9 44454571-44454674**	CNTNAP3B (0)
chr8 126362804-126362907	NSMCE2 (3.78)	**chr9 46896420-46896532**	KGFLP1 (0.7)
chr8 126545238-126545338	TRIB1 (0.97)	**chr9 47022868-47022978**	KGFLP1 (0.7)
chr8 127019462-127019582	LOC100130231 (0)	**chr9 66533933-66534045**	LOC442421 (0)
chr8 131482412-131482508	ASAP1 (1.61)	**chr9 74272377-74272483**	TMEM2 (1.49)
chr8 131892277-131892379	ASAP1 (1.61)	**chr9 80759387-80759489**	CEP78 (3.96)
chr8 132168967-132169069	ADCY8 (0)	**chr9 80847504-80847623**	CEP78 (3.96)
chr8 134392077-134392167	ST3GAL1 (2.76)	**chr9 81318037-81318156**	PSAT1 (5.1)
chr8 135341048-135341157	ZFAT (1.96)	**chr9 81341587-81341703**	PSAT1 (5.1)
chr8 135345747-135345865	ZFAT (1.96)	**chr9 81704897-81705016**	TLE4 (1.42)
chr8 135784077-135784192	MIR30B (0)	**chr9 83259872-83259976**	TLE4 (1.42)
chr8 136324165-136324269	LOC286094 (0)	**chr9 83714057-83714178**	TLE1 (0.55)
chr8 136706843-136706951	KHDRBS3 (0.62)	**chr9 84176118-84176221**	TLE1 (0.55)
chr8 137459499-137459602	KHDRBS3 (0.62)	**chr9 88530868-88530985**	NAA35 (3.4)
chr8 138316697-138316803	FAM135B (0.18)	**chr9 89152579-89152667**	ZCCHC6 (2.76)
chr8 138523188-138523294	FAM135B (0.18)	**chr9 90562273-90562360**	CDK20 (0)
chr8 140471359-140471465	KCNK9 (0.15)	**chr9 92470358-92470513**	UNQ6494 (0.06)
chr8 143218658-143218767	MIR4472-1 (0)	**chr9 93759289-93759405**	LOC100129316 (0.02)

chr9 94504359-94504478	ROR2 (0)	**chrX 18889481-18889600**	LOC100132163 (0)
chr9 95909126-95909232	NINJ1 (2.7)	**chrX 24315443-24315561**	FAM48B2 (0)
chr9 96160505-96160622	FAM120AOS (1.83)	**chrX 24526240-24526344**	PCYT1B (0.05)
chr9 96893001-96893095	PTPDC1 (2.79)	**chrX 24701328-24701434**	PCYT1B (0.05)
chr9 97479997-97480111	C9orf3 (2.21)	**chrX 26928813-26928984**	VENTXP1 (0)
chr9 99317543-99317633	HABP4 (2.21)	**chrX 27257510-27257610**	SMEK3P (0)
chr9 100153153-100153273	BDAG1 (0)	**chrX 29789964-29790071**	IL1RAPL1 (0)
chr9 101928111-101928233	TGFBR1 (1.38)	**chrX 30291458-30291576**	MAGEB1 (0.98)
chr9 101998622-101998809	SEC61B (6.24)	**chrX 34421974-34422092**	TMEM47 (0)
chr9 102000216-102000395	SEC61B (6.24)	**chrX 34587124-34587227**	TMEM47 (0)
chr9 103308522-103308625	C9orf30-TMEFF1 (0)	**chrX 34760321-34760434**	TMEM47 (0)
chr9 103801909-103802086	LPPR1 (0)	**chrX 35687970-35688090**	MAGEB16 (0)
chr9 106120936-106121044	CYLC2 (0.06)	**chrX 36466302-36466419**	CXorf30 (0)
chr9 106503308-106503412	SMC2 (3.69)	**chrX 39293036-39293152**	LOC286442 (0)
chr9 107040222-107040324	SMC2 (3.69)	**chrX 39294007-39294115**	LOC286442 (0)
chr9 107092615-107092732	OR13F1 (0)	**chrX 40291668-40291786**	ATP6AP2 (4.06)
chr9 110420565-110420676	KLF4 (0)	**chrX 42294702-42294811**	PPP1R2P9 (0)
chr9 110559690-110559776	KLF4 (0)	**chrX 44523851-44523957**	FUNDC1 (2.53)
chr9 112989705-112989808	TXN (6.74)	**chrX 45149486-45149596**	CXorf36 (0)
chr9 113082632-113082750	SVEP1 (0)	**chrX 46182425-46182528**	ZNF673 (0)
chr9 113251644-113251750	SVEP1 (0)	**chrX 47004001-47004120**	RBM10 (4.28)
chr9 113854456-113854574	LPAR1 (0)	**chrX 47685031-47685133**	ZNF81 (2.24)
chr9 115127701-115127804	HSDL2 (3.21)	**chrX 50948842-50948960**	NUDT10 (0)
chr9 116586999-116587105	ZNF618 (0.19)	**chrX 52726199-52726486**	SSX2 (0)
chr9 117446574-117446697	LOC100505478 (0.21)	**chrX 52790072-52790233**	SPANXN5 (0)
chr9 118380982-118381099	37226 (0)	**chrX 55885949-55886070**	RRAGB (2.18)
chr9 118595096-118595193	LINC00474 (0)	**chrX 63613921-63614024**	MTMR8 (0.58)
chr9 119680313-119680420	ASTN2 (0.4)	**chrX 64738215-64738315**	LAS1L (4)
chr9 119787209-119787327	ASTN2 (0.4)	**chrX 68884205-68884289**	EDA (0.07)
chr9 124879593-124879705	MIR4478 (0)	**chrX 69000552-69000659**	EDA (0.07)
chr9 126698983-126699083	DENND1A (1.88)	**chrX 69429630-69429733**	DGAT2L6 (0.08)
chr9 133272524-133272637	ASS1 (0.08)	**chrX 77795559-77795674**	ZCCHC5 (0.05)
chr9 137337042-137337138	RXRA (0.14)	**chrX 77922158-77922265**	ZCCHC5 (0.05)
chr9 137764389-137764507	FCN2 (0)	**chrX 78885080-78885183**	ITM2A (0)
chr9 138273293-138273411	LOC100506599 (0)	**chrX 78887054-78887223**	ITM2A (0)
chr9 138324366-138324480	LOC100506599 (0)	**chrX 78899251-78899363**	ITM2A (0)
chr9_gl000200_random 2743-2844	-1 (0)	**chrX 80772009-80772122**	SH3BGRL (5.02)
chrM 2824-3146	-1 (0)	**chrX 84594122-84594235**	POF1B (2.4)
chrUn_gl000220 114686-114809	RN5-8S1 (0)	**chrX 86301704-86301811**	DACH2 (0.08)
chrUn_gl000220 117521-118081	RN5-8S1 (0)	**chrX 86638149-86638252**	KLHL4 (2.81)
chrUn_gl000220 158692-158846	RN5-8S1 (0)	**chrX 88861727-88861833**	TGIF2LX (0)
chrUn_gl000220 159709-159964	RN5-8S1 (0)	**chrX 90388782-90388897**	PABPC5 (0)
chrX 1798394-1798495	ASMT (0)	**chrX 93500193-93500292**	FAM133A (0)
chrX 3347882-3347996	MXRA5 (0)	**chrX 93633542-93633659**	FAM133A (0)
chrX 5613358-5613477	NLGN4X (0)	**chrX 93990323-93990438**	FAM133A (0)
chrX 6140883-6141002	NLGN4X (0)	**chrX 94748385-94748491**	LOC643486 (0)
chrX 6547031-6547137	VCX3A (0)	**chrX 95720766-95720878**	LOC643486 (0)
chrX 7285592-7285707	STS (0.54)	**chrX 97927003-97927109**	LOC442459 (0)
chrX 8340180-8340275	VCX3B (0)	**chrX 99121883-99121973**	PCDH19 (0.04)
chrX 12102382-12102502	FRMPD4 (0)	**chrX 102074127-102074247**	LOC100287765 (0)
chrX 13079454-13079557	FAM9C (0)	**chrX 103472562-103472665**	ESX1 (0)
chrX 13321433-13321551	ATXN3L (0)	**chrX 105349145-105349246**	MUM1L1 (0)
chrX 13469091-13469201	EGFL6 (0)	**chrX 106449536-106449663**	NUP62CL (1.54)
chrX 15216905-15216999	ASB9 (1.35)	**chrX 111587637-111587743**	TRPC5 (0.13)
chrX 15498274-15498385	PIR-FIGF (0)	**chrX 111782280-111782399**	ZCCHC16 (0)
chrX 15920118-15920219	AP1S2 (1.92)	**chrX 113028122-113028240**	HTR2C (0)

chrX 114435601-114435704	RBMXL3 (0)	**chrX 141161889-141162006**	MAGEC2 (0)
chrX 114594404-114594511	LUZP4 (0)	**chrX 141168311-141168430**	MAGEC2 (0)
chrX 115822027-115822134	CXorf61 (0)	**chrX 141831144-141831247**	SPANXN4 (0)
chrX 116331751-116331869	KLHL13 (0.08)	**chrX 142132418-142132524**	SPANXN4 (0)
chrX 116735089-116735198	KLHL13 (0.08)	**chrX 142216047-142216164**	SPANXN4 (0)
chrX 117030757-117030863	KLHL13 (0.08)	**chrX 144262021-144262118**	SPANXN1 (0.26)
chrX 118425454-118425567	PGRMC1 (4.53)	**chrX 145724218-145724334**	CXorf51A (0)
chrX 122893877-122893974	THOC2 (4.28)	**chrX 146956819-146956925**	FMR1-AS1 (0.11)
chrX 124346974-124347092	LOC100129520 (0)	**chrX 148774536-148774638**	MAGEA11 (0)
chrX 131296408-131296521	FRMD7 (0.04)	**chrX 151047187-151047309**	MAGEA4 (0)
chrX 133733450-133733532	LOC100506757 (0)	**chrX 151057503-151057591**	MAGEA4 (0)
chrX 135056062-135056165	SLC9A6 (2.83)	**chrX 151963457-151963569**	MAGEA3 (0)
chrX 135287948-135288041	MAP7D3 (3.16)	**chrX 152370324-152370427**	MAGEA1 (0)
chrX 137256678-137256784	LOC158696 (0)	**chrX 153060084-153060217**	SSR4 (7.14)
chrX 138072656-138072802	FGF13 (2.78)	**chrX 153195446-153195612**	NAA10 (5.46)
chrX 140455357-140455472	SPANXC (0)		

Supplementary Data Table 6. Closest genes to active LTRs detected by RACE-Seq in **Reh** cells treated with **PMA**. Gene expression values (FPKM) from RNA-Seq are also shown for each gene.

RACE Peak	Gene (FPKM)
chr1 565650-565804	OR4F16 (0)
chr1 2123363-2123568	PRKCZ (1.89)
chr1 2208881-2209013	SKI (2.43)
chr1 11467932-11468049	PTCHD2 (1.71)
chr1 14803711-14803826	KAZN (3.45)
chr1 16998071-16998192	MIR3675 (0)
chr1 18215517-18215648	ACTL8 (0.22)
chr1 18905848-18905957	PAX7 (0)
chr1 19236099-19236221	IFFO2 (3.77)
chr1 31945312-31945443	LOC149086 (0)
chr1 36164901-36164997	C1orf216 (2.39)
chr1 38959158-38959277	LOC339442 (0)
chr1 41601131-41601246	SCMH1 (4.37)
chr1 42166571-42166684	HIVEP3 (2.38)
chr1 42187896-42188000	HIVEP3 (2.38)
chr1 42233846-42234018	HIVEP3 (2.38)
chr1 43094319-43094441	PPIH (6.97)
chr1 43952641-43952759	SZT2 (2.51)
chr1 51397087-51397215	FAF1 (5.7)
chr1 54632345-54632474	CYB5RL (1.86)
chr1 55203462-55203586	PARS2 (2.42)
chr1 57080282-57080409	PRKAA2 (0)
chr1 58185280-58185394	DAB1 (0)
chr1 58368622-58368738	DAB1 (0)
chr1 60201756-60201876	FGGY (4.46)
chr1 60898919-60899019	C1orf87 (0)
chr1 68020006-68020135	SERBP1 (6.32)
chr1 68791579-68791709	WLS (3.67)
chr1 72181725-72181847	NEGR1 (0.23)
chr1 75349869-75349985	TYW3 (4.4)
chr1 76917907-76918022	ST6GALNAC3 (0)
chr1 79007296-79007421	PTGFR (0)
chr1 79645553-79645647	ELTD1 (0)
chr1 79862755-79862866	ELTD1 (0)
chr1 83498797-83498909	TTLL7 (3.08)
chr1 88166158-88166276	LOC100505768 (0)
chr1 99503267-99503382	LOC100129620 (0)
chr1 101418637-101418745	DPH5 (4.95)
chr1 101455664-101455890	DPH5 (4.95)
chr1 102834806-102834920	OLFM3 (0)
chr1 105158459-105158578	LOC100129138 (0)
chr1 107779355-107779475	NTNG1 (0)
chr1 107781279-107781396	NTNG1 (0)
chr1 108093073-108093214	VAV3 (0)
chr1 108436114-108436288	VAV3 (0)
chr1 108547892-108548000	VAV3 (0)
chr1 112068984-112069124	ADORA3 (0.08)
chr1 115735419-115735545	NGF (0)
chr1 116082664-116082795	VANGL1 (0)
chr1 117994214-117994388	MAN1A2 (3.95)
chr1 118258812-118258927	FAM46C (1.07)
chr1 118996484-118996599	SPAG17 (0.29)
chr1 119198990-119199105	TBX15 (0.04)
chr1 119906782-119906900	HAO2 (0.1)
chr1 145277583-145277706	NBPF10 (1.59)
chr1 148632439-148632534	PPIAL4E (0)
chr1 148794462-148794579	PPIAL4D (0)
chr1 153478561-153478681	S100A6 (4.13)
chr1 153633935-153634053	ILF2 (7.77)
chr1 157866056-157866178	CD5L (0.06)
chr1 157867977-157868108	CD5L (0.06)
chr1 163078387-163078558	RGS4 (0)
chr1 163448435-163448557	NUF2 (5.69)
chr1 163511946-163512067	NUF2 (5.69)
chr1 163678154-163678249	NUF2 (5.69)
chr1 164633931-164634051	PBX1 (3.59)
chr1 166747083-166747204	POGK (5.58)
chr1 168731051-168731169	MGC4473 (0)
chr1 169842279-169842492	SCYL3 (3.93)
chr1 170803682-170803797	PRRX1 (1.61)
chr1 171409395-171409523	PRRC2C (5.65)
chr1 174173928-174174062	RABGAP1L (5.6)
chr1 174571371-174571502	RABGAP1L (5.6)
chr1 177429205-177429318	FAM5B (0)
chr1 177701423-177701596	SEC16B (0.28)
chr1 186217855-186217970	MIR548F1 (0)
chr1 186715852-186715969	PTGS2 (0.24)
chr1 188005721-188005847	PLA2G4A (2.15)
chr1 188564197-188564315	FAM5C (0)
chr1 188701644-188701766	FAM5C (0)
chr1 191457317-191457457	RGS18 (0.1)
chr1 194317011-194317142	CDC73 (2.73)
chr1 194699325-194699436	CDC73 (2.73)
chr1 194780252-194780382	KCNT2 (0.04)
chr1 194835292-194835387	KCNT2 (0.04)
chr1 198999779-198999910	LOC100131234 (0)
chr1 199647390-199647502	NR5A2 (1.22)
chr1 200247987-200248114	C1orf98 (0)
chr1 201798358-201798459	IPO9 (5.08)
chr1 204011709-204011799	LINC00303 (0.66)
chr1 204026517-204026644	SOX13 (1.6)
chr1 204598621-204598783	LRRN2 (0.15)
chr1 209874568-209874696	HSD11B1 (0.28)
chr1 210730901-210731030	HHAT (3.48)
chr1 211640091-211640221	RD3 (0.03)
chr1 215159425-215159553	KCNK2 (0)
chr1 216757098-216757213	ESRRG (0.1)
chr1 222238018-222238155	DUSP10 (3.72)
chr1 226532797-226532918	PARP1 (6.85)

chr1 228000689-228000801	PRSS38 (0.38)	**chr10 81075202-81075298**	ZMIZ1 (0.48)
chr1 234312893-234313008	SLC35F3 (0)	**chr10 81114905-81114999**	ZCCHC24 (0.59)
chr1 234347775-234347898	SLC35F3 (0)	**chr10 81955474-81955605**	ANXA11 (4.61)
chr1 234831811-234831919	LOC100506810 (0)	**chr10 82388100-82388224**	SH2D4B (0)
chr1 236326912-236327066	GPR137B (0.1)	**chr10 85909892-85910011**	C10orf99 (0)
chr1 236558790-236558898	EDARADD (0.92)	**chr10 88404999-88405089**	OPN4 (0.06)
chr1 237144761-237144891	RYR2 (0.74)	**chr10 91931347-91931479**	LOC643529 (0)
chr1 238221057-238221171	LOC100130331 (0)	**chr10 93095167-93095274**	LOC100188947 (0)
chr1 240800370-240800474	GREM2 (0)	**chr10 93656817-93656938**	FGFBP3 (1.59)
chr1 245121020-245121144	EFCAB2 (0.35)	**chr10 98497245-98497374**	PIK3AP1 (3.83)
chr1 247361696-247361919	MIR3916 (0)	**chr10 98509988-98510246**	PIK3AP1 (3.83)
chr10 528281-528393	DIP2C (1.54)	**chr10 101709917-101710045**	DNMBP (2.22)
chr10 1112737-1112868	WDR37 (3.04)	**chr10 108053838-108053944**	SORCS1 (0)
chr10 1189512-1189647	WDR37 (3.04)	**chr10 109262915-109263030**	SORCS1 (0)
chr10 1496248-1496383	ADARB2-AS1 (0)	**chr10 109732090-109732226**	SORCS1 (0)
chr10 1801346-1801467	ADARB2 (0.78)	**chr10 110936984-110937099**	XPNPEP1 (5.57)
chr10 2051633-2051768	ADARB2 (0.78)	**chr10 112302087-112302245**	SMC3 (4.84)
chr10 2105561-2105749	LOC282980 (0)	**chr10 114166710-114166832**	ACSL5 (4.1)
chr10 2107495-2107632	LOC282980 (0)	**chr10 115780565-115780696**	ADRB1 (0)
chr10 2213216-2213355	LOC399708 (0)	**chr10 120248598-120248735**	PRLHR (0)
chr10 3556036-3556158	KLF6 (2.93)	**chr10 120432082-120432205**	C10orf46 (0)
chr10 4263550-4263669	LOC100216001 (0)	**chr10 121376858-121376979**	TIAL1 (3.79)
chr10 4442466-4442584	LOC100216001 (0)	**chr10 125305958-125306094**	GPR26 (0)
chr10 6278979-6279097	PFKFB3 (2.36)	**chr10 132469158-132469286**	MIR378C (0)
chr10 7144559-7144680	SFMBT2 (1.69)	**chr10 132517772-132517895**	MIR378C (0)
chr10 8541543-8541638	GATA3 (2.99)	**chr10 134947503-134947608**	GPR123 (0)
chr10 15080567-15080682	OLAH (0.24)	**chr11 1286011-1286150**	MUC5B (0)
chr10 17277243-17277399	VIM (9.93)	**chr11 3206227-3206354**	OSBPL5 (0)
chr10 17568244-17568362	PTPLA (0)	**chr11 3402155-3402287**	LOC650368 (0)
chr10 17953654-17953744	MRC1 (0)	**chr11 4999292-4999410**	MMP26 (0)
chr10 18200537-18200657	MRC1 (0)	**chr11 6756080-6756205**	GVINP1 (1.52)
chr10 23198724-23198837	ARMC3 (0.12)	**chr11 8489642-8489760**	TRIM66 (2.18)
chr10 23834963-23835091	OTUD1 (2.37)	**chr11 9189857-9189970**	DENND5A (5.68)
chr10 24139708-24139816	KIAA1217 (2.88)	**chr11 9212018-9212137**	TMEM41B (4.01)
chr10 24335406-24335514	KIAA1217 (2.88)	**chr11 10830371-10830495**	EIF4G2 (7.22)
chr10 26015668-26015789	GPR158 (2.77)	**chr11 11110837-11110969**	GALNTL4 (0)
chr10 26041401-26041525	GPR158 (2.77)	**chr11 11845433-11845560**	USP47 (4.16)
chr10 30536804-30536921	MTPAP (3.36)	**chr11 13010990-13011140**	RASSF10 (0)
chr10 31991149-31991243	ARHGAP12 (2.98)	**chr11 14541842-14541964**	PSMA1 (7.42)
chr10 37937174-37937303	MTRNR2L7 (0)	**chr11 23658269-23658384**	LOC100500938 (0)
chr10 43389485-43389614	BMS1 (3.79)	**chr11 26268340-26268474**	ANO3 (0.38)
chr10 49394558-49394686	FRMPD2 (0.25)	**chr11 27911861-27912104**	KIF18A (3.71)
chr10 52555752-52555883	A1CF (0)	**chr11 34460431-34460541**	CAT (5.03)
chr10 56965591-56965718	MTRNR2L5 (0)	**chr11 35071553-35071684**	PDHX (4.67)
chr10 62312659-62312774	ANK3 (2.15)	**chr11 36607831-36607969**	RAG2 (0.05)
chr10 64018895-64019017	RTKN2 (0.27)	**chr11 37566533-37566656**	C11orf74 (4.14)
chr10 64188129-64188320	ZNF365 (0.46)	**chr11 39491597-39491731**	LRRC4C (1.91)
chr10 65005133-65005241	JMJD1C (2.75)	**chr11 40352112-40352232**	LRRC4C (1.91)
chr10 65720470-65720599	REEP3 (2.47)	**chr11 42275923-42276040**	LOC100507205 (0)
chr10 67021529-67021649	ANXA2P3 (0)	**chr11 46450159-46450402**	AMBRA1 (3.79)
chr10 72866084-72866206	UNC5B (0.04)	**chr11 57357416-57357524**	SERPING1 (0)
chr10 75006858-75006997	C10orf103 (0)	**chr11 59900683-59900813**	MS4A2 (0)
chr10 76340346-76340436	KAT6B (2.39)	**chr11 63595042-63595155**	MARK2 (2.69)
chr10 78271871-78272000	KCNMA1 (4.64)	**chr11 64185766-64185893**	RPS6KA4 (3.4)
chr10 78913897-78914004	C10orf11 (1.76)	**chr11 73685720-73685946**	DNAJB13 (0)
chr10 79664376-79664505	DLG5 (0.05)	**chr11 74168910-74169032**	LIPT2 (1.26)

chr11 74371087-74371207	POLD3 (3.99)	**chr12 76056310-76056441**	KRR1 (4.96)
chr11 80926103-80926228	MIR4300 (0)	**chr12 76220673-76220793**	PHLDA1 (3.3)
chr11 84040489-84040644	DLG2 (0.53)	**chr12 76365241-76365363**	PHLDA1 (3.3)
chr11 84202353-84202466	DLG2 (0.53)	**chr12 78714548-78714667**	NAV3 (1.43)
chr11 88417944-88418066	GRM5 (0.07)	**chr12 84255508-84255639**	TMTC2 (2.18)
chr11 88419874-88419998	GRM5 (0.07)	**chr12 85244603-85244703**	SLC6A15 (0)
chr11 94658597-94658704	CWC15 (5.3)	**chr12 86547472-86547580**	MGAT4C (0)
chr11 96732865-96732981	JRKL (2.61)	**chr12 87108273-87108443**	MGAT4C (0)
chr11 97252978-97253101	JRKL (2.61)	**chr12 87391528-87391643**	MGAT4C (0)
chr11 98765707-98765828	CNTN5 (0.03)	**chr12 88535708-88535866**	CEP290 (3.18)
chr11 102217943-102218088	BIRC2 (5.06)	**chr12 89445746-89445925**	LOC728084 (0.16)
chr11 102718551-102718664	MMP3 (0)	**chr12 89447703-89447829**	LOC728084 (0.16)
chr11 103987555-103987685	PDGFD (1.54)	**chr12 90631320-90631440**	LOC338758 (0)
chr11 108338014-108338165	C11orf65 (1.35)	**chr12 90674641-90674766**	LOC338758 (0)
chr11 109031417-109031557	DDX10 (3.86)	**chr12 90676542-90676684**	LOC338758 (0)
chr11 109479295-109479416	C11orf87 (0)	**chr12 91041254-91041362**	C12orf37 (0)
chr11 109645774-109645891	ZC3H12C (2.51)	**chr12 93335094-93335221**	EEA1 (3.46)
chr11 109674654-109674776	ZC3H12C (2.51)	**chr12 93551924-93552052**	LOC643339 (0.28)
chr11 110183030-110183143	RDX (5.77)	**chr12 93559065-93559186**	LOC643339 (0.28)
chr11 110201565-110201671	RDX (5.77)	**chr12 94027833-94027960**	CRADD (4.66)
chr11 110214738-110214857	RDX (5.77)	**chr12 98488747-98488877**	MIR4303 (0)
chr11 112367293-112367397	C11orf34 (0)	**chr12 104864053-104864171**	CHST11 (2.44)
chr11 113749381-113749523	USP28 (3.16)	**chr12 105927820-105927933**	C12orf75 (5.49)
chr11 126518001-126518109	KIRREL3 (0.35)	**chr12 111298763-111298881**	MYL2 (0)
chr11 127963611-127963701	ETS1 (3.3)	**chr12 111374368-111374487**	LOC100131138 (0)
chr11 128240075-128240195	ETS1 (3.3)	**chr12 111568929-111569050**	CUX2 (0)
chr12 1635061-1635174	LOC100292680 (0)	**chr12 111571167-111571343**	CUX2 (0)
chr12 6576151-6576282	VAMP1 (2.39)	**chr12 111848596-111848767**	ATXN2 (3.58)
chr12 9392490-9392597	LOC100499405 (0)	**chr12 112847260-112847398**	PTPN11 (5.76)
chr12 10691290-10691405	KLRAP1 (1.91)	**chr12 113470990-113471113**	OAS2 (3.33)
chr12 13420864-13420972	EMP1 (0.05)	**chr12 113784690-113784808**	PLBD2 (2.34)
chr12 15916997-15917126	EPS8 (1.58)	**chr12 113975423-113975541**	LHX5 (0)
chr12 19737519-19737656	AEBP2 (2.68)	**chr12 115803793-115803936**	MED13L (6.03)
chr12 21810682-21810801	LDHB (9.76)	**chr12 118792519-118792643**	TAOK3 (4.39)
chr12 24875773-24875912	BCAT1 (4.95)	**chr12 121629159-121629275**	P2RX7 (1.73)
chr12 24877922-24878044	BCAT1 (4.95)	**chr12 124961824-124961950**	ZNF664-FAM101A
chr12 24928810-24928931	BCAT1 (4.95)	**chr12 126796474-126796595**	LOC100128554 (0)
chr12 25005009-25005135	BCAT1 (4.95)	**chr12 126815001-126815110**	LOC100128554 (0)
chr12 25108470-25108602	BCAT1 (4.95)	**chr12 126872770-126872888**	LOC100128554 (0)
chr12 31836534-31836666	AMN1 (2.5)	**chr12 128045126-128045264**	FLJ37505 (0.11)
chr12 38625317-38625436	ALG10B (2.82)	**chr12 128203053-128203161**	FLJ37505 (0.11)
chr12 39872574-39872698	KIF21A (4.31)	**chr12 129594182-129594314**	TMEM132D (0)
chr12 41725501-41725620	PDZRN4 (0)	**chr12 131697998-131698106**	LOC116437 (0)
chr12 43345709-43345831	PRICKLE1 (0.52)	**chr12 132102692-132102790**	SFSWAP (4.69)
chr12 50505942-50506078	C12orf62 (0)	**chr13 19128858-19129000**	ANKRD20A9P (1.85)
chr12 52042299-52042428	SCN8A (2.04)	**chr13 19156190-19156371**	ANKRD20A9P (1.85)
chr12 57032941-57033055	BAZ2A (4.28)	**chr13 21776041-21776149**	MRP63 (0)
chr12 57081796-57081916	NACA (10.24)	**chr13 28841735-28841866**	FLT1 (0.64)
chr12 59880362-59880485	SLC16A7 (1.17)	**chr13 30861501-30861623**	LINC00426 (0)
chr12 63821157-63821286	DPY19L2 (1.07)	**chr13 31037352-31037447**	LINC00426 (0)
chr12 66032534-66032647	RPSAP52 (0.24)	**chr13 31501568-31501689**	LOC100507064 (0)
chr12 66316652-66316773	HMGA2 (0.72)	**chr13 32559756-32559877**	EEF1DP3 (0.28)
chr12 66414345-66414463	HMGA2 (0.72)	**chr13 35302383-35302496**	NBEA (1.18)
chr12 66451357-66451465	LLPH (5.17)	**chr13 36946955-36947074**	SPG20 (3.01)
chr12 68381098-68381225	IFNG (0.18)	**chr13 38410190-38410329**	TRPC4 (0.12)
chr12 68837765-68837886	MDM1 (2.92)	**chr13 38558533-38558664**	TRPC4 (0.12)

chr13 38624893-38624987	TRPC4 (0.12)	**chr13 101444308-101444430**	TMTC4 (2.39)
chr13 38704304-38704421	UFM1 (5.98)	**chr13 102161454-102161577**	NALCN (0)
chr13 39160372-39160494	FREM2 (1.21)	**chr13 103985021-103985152**	SLC10A2 (0.04)
chr13 43331542-43331648	C13orf30 (0)	**chr13 109109861-109109977**	MYO16 (1.19)
chr13 43393870-43393993	C13orf30 (0)	**chr13 109306697-109306826**	MYO16 (1.19)
chr13 47497650-47497806	HTR2A (0)	**chr13 110354375-110354516**	IRS2 (0)
chr13 47553270-47553378	HTR2A (0)	**chr13 110617240-110617358**	IRS2 (0)
chr13 48620633-48620774	MED4 (3.59)	**chr13 111594768-111594899**	ANKRD10 (4.8)
chr13 49458136-49458266	FNDC3A (2.04)	**chr13 111602188-111602328**	ANKRD10 (4.8)
chr13 56357026-56357153	PRR20C (0)	**chr13 111627519-111627644**	ANKRD10 (4.8)
chr13 56917306-56917425	PRR20C (0)	**chr13 114518544-114518676**	FAM70B (0)
chr13 58720242-58720373	PCDH17 (0)	**chr13 114571050-114571161**	LOC100506394 (0)
chr13 59883235-59883346	DIAPH3 (3.53)	**chr14 19680940-19681054**	POTEG (1.96)
chr13 60525397-60525519	TDRD3 (3)	**chr14 19894330-19894449**	POTEM (3.72)
chr13 60529470-60529595	TDRD3 (3)	**chr14 20925343-20925447**	TMEM55B (3.72)
chr13 60603708-60603823	DIAPH3 (3.53)	**chr14 21134861-21134982**	ANG (2.1)
chr13 60982109-60982346	TDRD3 (3)	**chr14 21458360-21458479**	SLC39A2 (0)
chr13 63306782-63306948	OR7E156P (0)	**chr14 23019278-23019403**	DAD1 (7.3)
chr13 63450393-63450514	OR7E156P (0)	**chr14 24037056-24037186**	JPH4 (3.05)
chr13 63452319-63452450	OR7E156P (0)	**chr14 25018572-25018687**	CTSG (0)
chr13 63457663-63457769	OR7E156P (0)	**chr14 26714702-26714820**	NOVA1 (0)
chr13 63631308-63631430	OR7E156P (0)	**chr14 32523895-32524012**	C14orf128 (0)
chr13 64636567-64636758	OR7E156P (0)	**chr14 35343884-35344005**	BAZ1A (4.35)
chr13 64665486-64665626	OR7E156P (0)	**chr14 35409138-35409359**	SRP54 (6.09)
chr13 68917841-68917966	LOC338862 (0)	**chr14 37158307-37158419**	SLC25A21 (0.49)
chr13 69251469-69251578	LOC338862 (0)	**chr14 37469988-37470107**	SLC25A21 (0.49)
chr13 71293196-71293334	ATXN8OS (0)	**chr14 38800238-38800358**	CLEC14A (0)
chr13 72664296-72664417	DACH1 (0.04)	**chr14 38973140-38973269**	LOC283547 (0)
chr13 73801912-73802036	KLF5 (2.1)	**chr14 39011198-39011311**	LOC283547 (0)
chr13 74103606-74103721	KLF12 (2.33)	**chr14 39967036-39967159**	FBXO33 (2.52)
chr13 77845769-77845894	FBXL3 (3.47)	**chr14 42457948-42458078**	LRFN5 (0)
chr13 80607661-80607791	SPRY2 (0.36)	**chr14 46341498-46341618**	MIS18BP1 (4.33)
chr13 80618388-80618496	SPRY2 (0.36)	**chr14 47426112-47426213**	MDGA2 (0)
chr13 80721414-80721513	SPRY2 (0.36)	**chr14 50053365-50053593**	RPS29 (11.35)
chr13 83966970-83967093	SLITRK1 (0.99)	**chr14 50403195-50403316**	ARF6 (4.87)
chr13 84202953-84203101	SLITRK1 (0.99)	**chr14 50872979-50873207**	CDKL1 (3.15)
chr13 85625707-85625827	SLITRK6 (0)	**chr14 51960187-51960303**	FRMD6 (2.44)
chr13 86270246-86270362	SLITRK6 (0)	**chr14 52372551-52372659**	GNG2 (5.09)
chr13 87087729-87087845	SLITRK6 (0)	**chr14 52646918-52647043**	PTGDR (0)
chr13 87127778-87127905	SLITRK6 (0)	**chr14 54382593-54382706**	BMP4 (0)
chr13 87778354-87778556	MIR4500HG (0.25)	**chr14 54453584-54453699**	BMP4 (0)
chr13 89571479-89571603	SLITRK5 (0)	**chr14 54613693-54613854**	BMP4 (0)
chr13 89681323-89681465	MIR622 (0)	**chr14 54779660-54779768**	CDKN3 (6.04)
chr13 89964556-89964696	MIR622 (0)	**chr14 58274585-58274700**	C14orf37 (0.5)
chr13 91013207-91013319	MIR622 (0)	**chr14 61566646-61566762**	SLC38A6 (2.57)
chr13 91297129-91297257	LINC00410 (0)	**chr14 62697537-62697651**	FLJ43390 (0)
chr13 91782665-91782796	LINC00410 (0)	**chr14 64713074-64713164**	ESR2 (0.69)
chr13 92528240-92528421	GPC5 (1.71)	**chr14 65846854-65846974**	FUT8 (4.17)
chr13 92530202-92530334	GPC5 (1.71)	**chr14 66117119-66117288**	FUT8 (4.17)
chr13 92570314-92570441	GPC5 (1.71)	**chr14 66521578-66521693**	FUT8 (4.17)
chr13 93498337-93498525	GPC5 (1.71)	**chr14 66969784-66969899**	GPHN (0.13)
chr13 95054946-95055050	GPC6 (0.42)	**chr14 69192829-69192949**	ZFP36L1 (5.15)
chr13 95073232-95073374	GPC6 (0.42)	**chr14 70753651-70753772**	SYNJ2BP-COX16 (0)
chr13 100104202-100104322	MIR548AN (0)	**chr14 71825449-71825574**	SNORD56B (0)
chr13 100113264-100113372	TM9SF2 (5.09)	**chr14 81413278-81413419**	CEP128 (2.6)
chr13 101031876-101031984	PCCA (2.41)	**chr14 81675598-81675738**	GTF2A1 (3.23)

chr14 81723121-81723230	STON2 (0.29)	**chr15 62574063-62574210**	C2CD4B (0)
chr14 82495018-82495287	SEL1L (5.13)	**chr15 66721537-66721645**	MAP2K1 (4.79)
chr14 82661029-82661151	SEL1L (5.13)	**chr15 69745130-69745249**	KIF23 (5.36)
chr14 82889326-82889448	SEL1L (5.13)	**chr15 70656459-70656597**	TLE3 (3.84)
chr14 83317248-83317369	SEL1L (5.13)	**chr15 70815599-70815704**	UACA (0)
chr14 83764442-83764570	SEL1L (5.13)	**chr15 70885588-70885705**	UACA (0)
chr14 84620635-84620767	FLRT2 (0)	**chr15 71982996-71983117**	THSD4 (0.02)
chr14 85271396-85271512	FLRT2 (0)	**chr15 75230265-75230442**	RPP25 (3.43)
chr14 85654494-85654612	FLRT2 (0)	**chr15 75918192-75918304**	SNUPN (4.77)
chr14 86029811-86029914	LOC283585 (0)	**chr15 77910854-77910958**	LOC253044 (0)
chr14 86633038-86633158	FLRT2 (0)	**chr15 80478579-80478673**	FAH (5.2)
chr14 87430415-87430542	LOC283585 (0)	**chr15 81579196-81579299**	IL16 (0.77)
chr14 87700394-87700516	LOC283585 (0)	**chr15 91284011-91284137**	BLM (4.23)
chr14 87766493-87766614	LOC283585 (0)	**chr15 91803961-91804086**	SV2B (0.96)
chr14 88338606-88338741	GALC (2.52)	**chr15 95751455-95751569**	LOC400456 (0)
chr14 88351594-88351704	GALC (2.52)	**chr15 96603973-96604103**	NR2F2 (0.9)
chr14 88746240-88746372	KCNK10 (0.2)	**chr15 97592377-97592507**	SPATA8 (0.27)
chr14 90186522-90186644	EFCAB11 (2.99)	**chr15 99941440-99941602**	LRRC28 (3.27)
chr14 90480179-90480287	KCNK13 (0.58)	**chr15 101835329-101835465**	PCSK6 (0.74)
chr14 92630281-92630508	CPSF2 (5.59)	**chr15 102182678-102182802**	TARSL2 (3.86)
chr14 93328732-93328847	GOLGA5 (4.56)	**chr16 7839029-7839155**	RBFOX1 (1.97)
chr14 93338644-93338774	GOLGA5 (4.56)	**chr16 10251375-10251489**	GRIN2A (0.51)
chr14 93339794-93339923	GOLGA5 (4.56)	**chr16 13515186-13515315**	SHISA9 (3.53)
chr14 93366486-93366597	CHGA (0.1)	**chr16 15629660-15629787**	KIAA0430 (3.09)
chr14 95920000-95920122	C14orf49 (0)	**chr16 15657459-15657582**	KIAA0430 (3.09)
chr14 96245203-96245318	TCL1A (3.14)	**chr16 15666746-15666876**	KIAA0430 (3.09)
chr14 97375242-97375368	VRK1 (6.4)	**chr16 15667927-15668050**	KIAA0430 (3.09)
chr14 97644017-97644114	VRK1 (6.4)	**chr16 19103046-19103172**	COQ7 (3.41)
chr14 97646057-97646218	VRK1 (6.4)	**chr16 20035578-20035701**	GPR139 (0)
chr14 97903764-97903886	LOC100129345 (0)	**chr16 21127739-21127866**	DNAH3 (0.05)
chr14 99005428-99005561	C14orf177 (0)	**chr16 27258606-27258735**	NSMCE1 (6.52)
chr14 99523030-99523159	BCL11B (0.44)	**chr16 31123645-31123821**	KAT8 (5.29)
chr14 100545306-100545430	EVL (1.16)	**chr16 31769860-31769974**	ZNF720 (1.23)
chr14 100800131-100800272	SLC25A47 (0)	**chr16 51457136-51457257**	SALL1 (0)
chr14 105140391-105140504	MIR4710 (0)	**chr16 51660959-51661093**	LOC388276 (0)
chr15 23188131-23188263	WHAMMP3 (1.88)	**chr16 51666347-51666455**	LOC388276 (0)
chr15 27622604-27622726	OCA2 (0)	**chr16 51890865-51890984**	LOC388276 (0)
chr15 33502044-33502168	RYR3 (0.06)	**chr16 53109515-53109646**	CHD9 (2.31)
chr15 34880553-34880662	GOLGA8B (2.9)	**chr16 54510840-54510975**	IRX3 (0)
chr15 38318066-38318210	TMCO5A (0)	**chr16 54907978-54908100**	CRNDE (0.56)
chr15 38332250-38332531	TMCO5A (0)	**chr16 58455205-58455323**	GINS3 (3.7)
chr15 38339666-38339819	TMCO5A (0)	**chr16 60118540-60118655**	LOC644649 (0)
chr15 38428136-38428285	SPRED1 (2.25)	**chr16 60128450-60128577**	LOC644649 (0)
chr15 38531423-38531538	SPRED1 (2.25)	**chr16 61287015-61287144**	CDH8 (0.67)
chr15 38562723-38562830	FAM98B (1.87)	**chr16 67293724-67293823**	SLC9A5 (2.49)
chr15 38974430-38974554	C15orf53 (0.63)	**chr16 67978246-67978360**	SLC12A4 (3.35)
chr15 43054197-43054317	TTBK2 (3.16)	**chr16 76268953-76269081**	CNTNAP4 (3.2)
chr15 43490131-43490253	EPB42 (0.06)	**chr16 77446490-77446618**	ADAMTS18 (0)
chr15 45103610-45103715	TRIM69 (0.53)	**chr16 77530867-77530985**	ADAMTS18 (0)
chr15 48365095-48365207	SLC24A5 (0.09)	**chr16 78128713-78128842**	WWOX (4.75)
chr15 49907722-49907826	C15orf33 (0)	**chr16 80704255-80704371**	CDYL2 (3.31)
chr15 53315174-53315295	ONECUT1 (0.12)	**chr16 81755280-81755384**	CMIP (2.57)
chr15 54902011-54902121	UNC13C (2.18)	**chr16 82399045-82399154**	MPHOSPH6 (4.72)
chr15 56651045-56651177	TEX9 (1.2)	**chr16 82965238-82965360**	CDH13 (0.04)
chr15 56689930-56690100	MNS1 (0)	**chr16 82984069-82984244**	CDH13 (0.04)
chr15 61881207-61881349	VPS13C (4.36)	**chr16 83252490-83252618**	CDH13 (0.04)

chr16 88876083-88876202	APRT (7.99)	**chr18 33077773-33077886**	MIR3975 (0)
chr16 89571780-89571802	SPG7 (4.72)	**chr18 36807053-36807182**	LOC647946 (0)
chr17 7529674-7529769	SHBG (0)	**chr18 39394632-39394741**	PIK3C3 (3.68)
chr17 7760698-7760886	CYB5D1 (1.32)	**chr18 39438755-39438862**	PIK3C3 (3.68)
chr17 12070265-12070405	MAP2K4 (4.09)	**chr18 39455678-39455837**	PIK3C3 (3.68)
chr17 13684382-13684513	HS3ST3A1 (0.27)	**chr18 39580619-39580748**	PIK3C3 (3.68)
chr17 15683356-15683485	MEIS3P1 (0)	**chr18 39695462-39695595**	PIK3C3 (3.68)
chr17 18154163-18154266	FLII (5.62)	**chr18 42251871-42251993**	SETBP1 (0.11)
chr17 21274032-21274147	KCNJ12 (0)	**chr18 43569317-43569446**	PSTPIP2 (0)
chr17 27717448-27717563	MIR4523 (0)	**chr18 44722454-44722556**	IER3IP1 (5.73)
chr17 30771479-30771628	PSMD11 (4.73)	**chr18 44812053-44812162**	IER3IP1 (5.73)
chr17 34465546-34465666	TBC1D3B (0)	**chr18 55879577-55879698**	NEDD4L (2.99)
chr17 39878891-39879017	HAP1 (0)	**chr18 56447358-56447472**	MALT1 (3.64)
chr17 46752970-46753095	MIR196A1 (0)	**chr18 57213002-57213132**	CCBE1 (0.08)
chr17 48224527-48224688	PPP1R9B (5.27)	**chr18 59702321-59702440**	PIGN (3.03)
chr17 48940277-48940446	LOC400604 (0)	**chr18 61222171-61222295**	SERPINB12 (0)
chr17 50544006-50544136	CA10 (0)	**chr18 61427010-61427131**	SERPINB7 (0)
chr17 59322447-59322556	BCAS3 (4.34)	**chr18 61428938-61429064**	SERPINB7 (0)
chr17 62011806-62011903	CD79B (0.22)	**chr18 62453065-62453169**	LOC284294 (0)
chr17 62060132-62060247	SCN4A (0.03)	**chr18 62658730-62658842**	LOC284294 (0)
chr17 65700590-65700721	PITPNC1 (3.32)	**chr18 63882058-63882186**	CDH19 (0.02)
chr17 66823148-66823274	ABCA8 (0)	**chr18 65728267-65728397**	LOC643542 (0)
chr17 67521671-67521817	MAP2K6 (0.58)	**chr18 65906195-65906316**	LOC643542 (0)
chr17 67550270-67550438	MAP2K6 (0.58)	**chr18 65936807-65936924**	LOC643542 (0)
chr17 67559491-67559624	MAP2K6 (0.58)	**chr18 66096421-66096566**	TMX3 (3.7)
chr17 67595773-67595903	MAP2K6 (0.58)	**chr18 66231147-66231330**	TMX3 (3.7)
chr17 67696430-67696553	MAP2K6 (0.58)	**chr18 71197391-71197530**	LOC100505817 (0)
chr17 67744806-67744934	MAP2K6 (0.58)	**chr18 71452136-71452256**	FBXO15 (0.24)
chr17 67781899-67782021	MAP2K6 (0.58)	**chr18 71657465-71657567**	FBXO15 (0.24)
chr17 67906893-67906997	KCNJ16 (0)	**chr18 72731699-72731841**	ZNF407 (2.24)
chr17 68028148-68028270	KCNJ16 (0)	**chr18 72785986-72786133**	ZNF407 (2.24)
chr17 68338976-68339097	KCNJ2 (2.38)	**chr18 72816048-72816170**	ZNF407 (2.24)
chr17 68394927-68395049	KCNJ2 (2.38)	**chr18 73341586-73341711**	C18orf62 (0)
chr17 68572859-68572980	KCNJ2 (2.38)	**chr18 73447893-73448005**	C18orf62 (0)
chr17 69617070-69617184	SOX9 (0.12)	**chr18 75422023-75422146**	GALR1 (0)
chr17 70387123-70387301	LOC100499467 (0)	**chr19 5676828-5676925**	C19orf70 (6.17)
chr17 72514913-72515028	CD300LB (0)	**chr19 5916284-5916394**	RANBP3 (4.5)
chr17 72931464-72931579	OTOP3 (0.06)	**chr19 6000617-6000735**	RFX2 (2.59)
chr17 73262112-73262280	MIF4GD (4.77)	**chr19 6323445-6323586**	CLPP (6.09)
chr17 73754326-73754585	GALK1 (3.49)	**chr19 6381159-6381431**	GTF2F1 (5.99)
chr17 74560618-74560718	ST6GALNAC2 (0.39)	**chr19 6413804-6414014**	KHSRP (5.84)
chr17 77158349-77158502	RBFOX3 (0.51)	**chr19 7320964-7321085**	INSR (0.01)
chr17 78622681-78622794	RPTOR (3.05)	**chr19 7696346-7696458**	PCP2 (0.52)
chr17 79336006-79336127	TMEM105 (0)	**chr19 8431208-8431320**	LOC100507567 (0)
chr18 548595-548715	CETN1 (0)	**chr19 11488473-11488668**	EPOR (2.01)
chr18 6123474-6123564	L3MBTL4 (0.15)	**chr19 13864455-13864593**	MRI1 (3.65)
chr18 8445102-8445234	PTPRM (1.33)	**chr19 16308721-16308846**	AP1M1 (5.45)
chr18 13035334-13035463	CEP192 (4.18)	**chr19 17615076-17615215**	PGLS (5.17)
chr18 15031453-15031611	ANKRD30B (3.44)	**chr19 23887334-23887465**	ZNF675 (2.88)
chr18 15109912-15110004	LOC644669 (0)	**chr19 28819641-28819766**	LOC148189 (0)
chr18 15135158-15135289	LOC644669 (0)	**chr19 29087923-29088016**	LOC148145 (0)
chr18 20857274-20857381	CABLES1 (0.05)	**chr19 30054649-30054764**	LOC284395 (0)
chr18 22276324-22276450	LOC729950 (0)	**chr19 30534170-30534299**	C19orf2 (0)
chr18 26055019-26055140	CDH2 (0)	**chr19 31735176-31735302**	TSHZ3 (0)
chr18 29545591-29545716	TRAPPC8 (3.56)	**chr19 34046701-34046817**	PEPD (4.18)
chr18 32485066-32485191	DTNA (2.03)	**chr19 38874210-38874328**	GGN (1.5)

chr19 42891509-42891631	MEGF8 (2.67)	**chr2 40808075-40808252**	SLC8A1 (2.01)
chr19 47029663-47029783	LOC100506012 (0)	**chr2 40838102-40838291**	SLC8A1 (2.01)
chr19 47568234-47568332	ZC3H4 (2.6)	**chr2 42205655-42205776**	LOC400950 (0)
chr19 47776903-47777024	PRR24 (0)	**chr2 44406267-44406380**	PPM1B (5.2)
chr19 50363569-50363659	PNKP (3.28)	**chr2 44524017-44524129**	PREPL (4.61)
chr19 50363843-50363942	PNKP (3.28)	**chr2 45819216-45819342**	PRKCE (1.83)
chr19 56824639-56824777	ZNF542 (0)	**chr2 47684532-47684647**	MSH2 (5.75)
chr19 57999415-57999547	ZNF419 (1.78)	**chr2 54342708-54342842**	ACYP2 (2.92)
chr19 58011545-58011677	ZNF773 (2.71)	**chr2 54501240-54501368**	ACYP2 (2.92)
chr2 682715-682829	TMEM18 (5.1)	**chr2 54621635-54621762**	C2orf73 (0)
chr2 1635917-1636054	PXDN (0)	**chr2 58485795-58485913**	FANCL (5.41)
chr2 1803670-1803763	MYT1L (0)	**chr2 60343696-60343818**	MIR4432 (0)
chr2 2431270-2431388	MYT1L (0)	**chr2 65275130-65275286**	CEP68 (3.79)
chr2 2535700-2535790	MYT1L (0)	**chr2 66582856-66582981**	MIR4778 (0)
chr2 3955082-3955204	LOC100505964 (0)	**chr2 66972395-66972517**	MEIS1 (3.65)
chr2 4295740-4295862	LOC100505964 (0)	**chr2 67042192-67042316**	MEIS1 (3.65)
chr2 4980726-4980847	LOC727982 (0)	**chr2 67342569-67342698**	LOC644838 (0)
chr2 5566446-5566585	SOX11 (0)	**chr2 67735498-67735618**	ETAA1 (4.68)
chr2 5683386-5683494	SOX11 (0)	**chr2 68054669-68054856**	C1D (4.88)
chr2 6101996-6102129	LOC400940 (0.44)	**chr2 68623157-68623291**	FBXO48 (1.76)
chr2 6334660-6334790	LOC400940 (0.44)	**chr2 70645597-70645722**	TGFA (1.41)
chr2 6924918-6925047	LINC00487 (0)	**chr2 74154006-74154156**	DGUOK (6.69)
chr2 7239993-7240096	RNF144A (0)	**chr2 75769219-75769335**	FAM176A (0)
chr2 7780411-7780534	LOC100506274 (0.16)	**chr2 76532754-76532860**	LRRTM4 (3.2)
chr2 8034492-8034623	LOC339788 (0)	**chr2 78007895-78008025**	SNAR-H (0)
chr2 8856192-8856312	KIDINS220 (4.58)	**chr2 80719333-80719440**	CTNNA2 (2.35)
chr2 11029086-11029215	KCNF1 (0)	**chr2 89375691-89375814**	MIR4436A (0)
chr2 11402709-11402832	ROCK2 (4.02)	**chr2 90163103-90163226**	MIR4436A (0)
chr2 11720093-11720223	MIR4429 (0)	**chr2 92195966-92196069**	ACTR3BP2 (0)
chr2 11821535-11821647	NTSR2 (0)	**chr2 99285791-99285906**	MGAT4A (2.16)
chr2 12603365-12603488	TRIB2 (1.85)	**chr2 99349273-99349388**	MGAT4A (2.16)
chr2 12756710-12756828	TRIB2 (1.85)	**chr2 99920751-99920882**	LYG1 (1.71)
chr2 13095355-13095478	LOC100506474 (0)	**chr2 100544442-100544573**	AFF3 (1.35)
chr2 14131250-14131381	FAM84A (0)	**chr2 101786016-101786141**	TBC1D8 (1.35)
chr2 14375963-14376071	FAM84A (0)	**chr2 104629707-104629819**	LOC100287010 (0.21)
chr2 15038793-15038900	FAM84A (0)	**chr2 110451127-110451247**	ANKRD57 (0)
chr2 15054104-15054222	NBAS (4.28)	**chr2 110905815-110905951**	NPHP1 (0.88)
chr2 15633228-15633336	NBAS (4.28)	**chr2 110911705-110911863**	NPHP1 (0.88)
chr2 16165995-16166118	MYCN (0)	**chr2 110913820-110913991**	NPHP1 (0.88)
chr2 17884009-17884131	VSNL1 (0.14)	**chr2 111881300-111881401**	ACOXL (1.19)
chr2 18005042-18005193	MSGN1 (0)	**chr2 114448621-114448731**	SLC35F5 (3.61)
chr2 18351344-18351467	KCNS3 (0)	**chr2 114603519-114603639**	ACTR3 (5.6)
chr2 18730633-18730751	RDH14 (3.67)	**chr2 132619286-132619442**	C2orf27B (0)
chr2 20678251-20678385	RHOB (0.44)	**chr2 132693184-132693313**	C2orf27B (0)
chr2 22065539-22065652	LOC645949 (0.33)	**chr2 132704195-132704314**	C2orf27B (0)
chr2 24674824-24674941	ITSN2 (4.41)	**chr2 132737343-132737473**	ANKRD30BL (0)
chr2 25109961-25110076	DNAJC27 (1.31)	**chr2 136007265-136007387**	ZRANB3 (1.98)
chr2 25876155-25876286	ASXL2 (3.13)	**chr2 141815155-141815295**	LRP1B (0.04)
chr2 27907983-27908115	SLC4A1AP (4.36)	**chr2 143874854-143874984**	ARHGAP15 (6.12)
chr2 30490026-30490160	LBH (0.07)	**chr2 147581166-147581279**	PABPC1P2 (0.17)
chr2 30554540-30554662	LBH (0.07)	**chr2 148077919-148078031**	ACVR2A (2.74)
chr2 30594330-30594466	LCLAT1 (3.12)	**chr2 148177658-148177781**	ACVR2A (2.74)
chr2 32951900-32952026	TTC27 (4.08)	**chr2 150985221-150985342**	RND3 (0.35)
chr2 33931900-33932032	MYADML (0)	**chr2 151919017-151919147**	RBM43 (1.56)
chr2 36275756-36275881	LOC100288911 (0)	**chr2 153572505-153572662**	ARL6IP6 (3.62)
chr2 36384369-36384477	LOC100288911 (0)	**chr2 153573782-153573970**	ARL6IP6 (3.62)

chr2 155062116-155062243	GALNT13 (0.03)	**chr20 22529161-22529281**	LINC00261 (0.21)
chr2 157879192-157879307	GALNT5 (6)	**chr20 22531141-22531264**	LINC00261 (0.21)
chr2 158033320-158033436	GALNT5 (6)	**chr20 22940397-22940530**	SSTR4 (0)
chr2 159145375-159145495	CCDC148 (0.67)	**chr20 25782621-25782742**	FAM182B (0.02)
chr2 162419060-162419183	SLC4A10 (0.09)	**chr20 31522523-31522644**	SUN5 (0)
chr2 163954253-163954381	KCNH7 (0.91)	**chr20 32482224-32482344**	CHMP4B (3.77)
chr2 164518268-164518396	KCNH7 (0.91)	**chr20 40256542-40256675**	CHD6 (3.49)
chr2 166629785-166629915	CSRNP3 (0.05)	**chr20 42455728-42455841**	TOX2 (0.11)
chr2 173750243-173750365	RAPGEF4 (0.09)	**chr20 42735256-42735378**	JPH2 (0.12)
chr2 179946504-179946618	SESTD1 (2.45)	**chr20 43588896-43589040**	LOC100505826 (0)
chr2 182958841-182958952	PDE1A (0.14)	**chr20 49678720-49678847**	KCNG1 (0)
chr2 184132749-184132879	NUP35 (4.89)	**chr20 50879398-50879514**	ZFP64 (2.16)
chr2 189248967-189249086	GULP1 (0)	**chr20 52671871-52672000**	MIR4756 (0)
chr2 190851947-190852067	MSTN (0.3)	**chr20 56058214-56058546**	CTCFL (0)
chr2 191573355-191573478	NAB1 (3.8)	**chr20 56617023-56617140**	C20orf85 (0)
chr2 192086028-192086137	MYO1B (0.04)	**chr20 57102170-57102284**	LOC149773 (0)
chr2 196229586-196229715	SLC39A10 (3.03)	**chr20 57111923-57112044**	LOC149773 (0)
chr2 196344291-196344398	SLC39A10 (3.03)	**chr20 58867317-58867438**	MIR4533 (0)
chr2 207101094-207101268	GPR1 (0.07)	**chr20 61368262-61368385**	NTSR1 (0)
chr2 207101284-207101503	GPR1 (0.07)	**chr21 14555335-14555469**	ANKRD30BP2 (0.06)
chr2 212856951-212857072	ERBB4 (0.06)	**chr21 14582642-14582839**	ANKRD30BP2 (0.06)
chr2 222529817-222529944	EPHA4 (0)	**chr21 14599928-14600056**	ANKRD30BP2 (0.06)
chr2 224532895-224533013	SCG2 (0.43)	**chr21 14668572-14668701**	MIR3156-3 (0)
chr2 224604006-224604135	AP1S3 (1.87)	**chr21 14702650-14702772**	MIR3156-3 (0)
chr2 224887625-224887758	SERPINE2 (4.41)	**chr21 14709803-14709960**	MIR3156-3 (0)
chr2 224980281-224980385	SERPINE2 (4.41)	**chr21 16042590-16042713**	SAMSN1 (4.52)
chr2 224990233-224990355	SERPINE2 (4.41)	**chr21 16052037-16052155**	SAMSN1 (4.52)
chr2 225291551-225291657	FAM124B (2.07)	**chr21 16215973-16216102**	NRIP1 (3.23)
chr2 226884344-226884442	KIAA1486 (0)	**chr21 16234082-16234203**	NRIP1 (3.23)
chr2 227408287-227408416	IRS1 (1.04)	**chr21 17005828-17005937**	USP25 (3.88)
chr2 227642412-227642525	RHBDD1 (2.97)	**chr21 17248786-17248900**	LINC00478 (0)
chr2 229227267-229227396	SPHKAP (0)	**chr21 18054603-18054693**	LINC00478 (0)
chr2 231137220-231137343	SP140 (4.53)	**chr21 18483327-18483449**	C21orf37 (0)
chr2 234485992-234486106	USP40 (0.61)	**chr21 19378370-19378482**	CHODL (0.07)
chr2 235166328-235166451	SPP2 (0)	**chr21 19858105-19858235**	TMPRSS15 (0)
chr2 235192096-235192296	SPP2 (0)	**chr21 20834129-20834251**	TMPRSS15 (0)
chr2 235218086-235218217	ARL4C (0.09)	**chr21 23058566-23058683**	LINC00317 (0)
chr2 238363284-238363392	MLPH (1.27)	**chr21 23406649-23406791**	LINC00308 (0)
chr2 239491515-239491621	LOC151171 (0)	**chr21 23966238-23966368**	LINC00308 (0)
chr2 239527007-239527141	LOC151171 (0)	**chr21 25166166-25166283**	D21S2088E (0.44)
chr2 239683158-239683297	TWIST2 (2.25)	**chr21 26637135-26637248**	LINC00158 (3.93)
chr2 239869549-239869674	FLJ43879 (0)	**chr21 26874721-26874855**	MIR155HG (6.23)
chr2 241647942-241648067	KIF1A (0.36)	**chr21 27727427-27727542**	CYYR1 (0.19)
chr2 242670240-242670355	ING5 (2.82)	**chr21 27842834-27842939**	APP (0.48)
chr20 1145731-1145886	C20orf46 (0)	**chr21 28195430-28195560**	ADAMTS1 (0.04)
chr20 1828125-1828265	SIRPA (0)	**chr21 29277510-29277643**	LINC00314 (0)
chr20 4004646-4004775	RNF24 (1.25)	**chr21 29288683-29288824**	LINC00314 (0)
chr20 4557225-4557363	PRNP (4.41)	**chr21 29376724-29376865**	LINC00314 (0)
chr20 5332999-5333128	PROKR2 (0)	**chr21 29413631-29413753**	LINC00314 (0)
chr20 7121440-7121553	BMP2 (0)	**chr21 29420711-29420867**	LINC00314 (0)
chr20 9134091-9134212	PLCB4 (0)	**chr21 29430126-29430257**	LINC00314 (0)
chr20 11161916-11162058	LOC339593 (0)	**chr21 29570158-29570288**	LINC00314 (0)
chr20 12638327-12638448	SPTLC3 (0)	**chr21 29715685-29715819**	LINC00161 (0)
chr20 15903870-15903991	MACROD2 (0.4)	**chr21 30908676-30908804**	GRIK1 (0.23)
chr20 18109316-18109424	PET117 (4.08)	**chr21 32722912-32723022**	TIAM1 (2.03)
chr20 19709731-19709898	SLC24A3 (0)	**chr21 33903275-33903399**	C21orf63 (0)

chr21 34359756-34359867	OLIG2 (0)	**chr3 52897261-52897355**	MUSTN1 (2.81)
chr21 35284586-35284714	LOC100506334 (0)	**chr3 58464635-58464762**	KCTD6 (2.68)
chr21 35902210-35902331	RCAN1 (3.06)	**chr3 60031034-60031131**	FHIT (0.21)
chr21 36529950-36530064	RUNX1 (2.35)	**chr3 60781688-60781817**	PTPRG (0.33)
chr21 36669210-36669331	RUNX1 (2.35)	**chr3 67561910-67562028**	SUCLG2 (5.83)
chr21 36940982-36941108	MIR802 (0)	**chr3 69040966-69041087**	TMF1 (4.25)
chr21 37177006-37177136	MIR802 (0)	**chr3 72316188-72316310**	LOC201617 (0)
chr21 37214314-37214435	MIR802 (0)	**chr3 72529282-72529575**	RYBP (2.79)
chr21 37470446-37470556	LOC100133286 (0)	**chr3 73261434-73261562**	PPP4R2 (6.37)
chr21 37838671-37838779	CLDN14 (0)	**chr3 77778485-77778608**	ROBO2 (2.13)
chr21 40150497-40150608	LINC00114 (0)	**chr3 78611045-78611147**	ROBO1 (5.29)
chr21 40917301-40917431	SH3BGR (1.54)	**chr3 79984673-79984791**	ROBO1 (5.29)
chr21 41229345-41229475	PCP4 (1.72)	**chr3 80577743-80577866**	ROBO1 (5.29)
chr21 41621290-41621413	DSCAM (1.84)	**chr3 83666188-83666278**	LOC440970 (0)
chr21 42312114-42312223	DSCAM (1.84)	**chr3 84557474-84557582**	LOC440970 (0)
chr21 42524190-42524307	LINC00323 (0)	**chr3 84869318-84869440**	CADM2 (0.02)
chr21 43039773-43039900	LINC00111 (0)	**chr3 87645421-87645542**	POU1F1 (0)
chr21 45191562-45191701	CSTB (4.47)	**chr3 89819579-89819698**	EPHA3 (1.78)
chr21 46579076-46579187	ADARB1 (1.74)	**chr3 95145153-95145281**	LOC255025 (0)
chr21 47013493-47013662	SLC19A1 (3.15)	**chr3 96154582-96154713**	EPHA6 (0)
chr21 47556864-47556982	FTCD (0.21)	**chr3 99563661-99563786**	MIR548G (0)
chr22 16162027-16162146	POTEH (2.28)	**chr3 102377756-102377883**	ZPLD1 (0)
chr22 17232428-17232549	XKR3 (0)	**chr3 102574912-102575052**	ZPLD1 (0)
chr22 17358578-17358707	HSFY1P1 (0)	**chr3 103436519-103436609**	MIR548A3 (0)
chr22 27766481-27766608	MN1 (0.06)	**chr3 106953495-106953610**	LOC100302640 (0)
chr22 39130848-39130966	SUN2 (3.75)	**chr3 107843859-107843990**	CD47 (3.69)
chr22 39635878-39635987	RPL3 (10.38)	**chr3 108150274-108150400**	MYH15 (0.23)
chr22 42907921-42908026	SERHL2 (0.8)	**chr3 109282201-109282311**	FLJ25363 (0)
chr22 44573875-44573996	PARVG (0)	**chr3 109437796-109437908**	FLJ25363 (0)
chr22 46067730-46067848	ATXN10 (4.49)	**chr3 109449782-109449918**	FLJ25363 (0)
chr22 46409276-46409399	LOC730668 (0.22)	**chr3 109906643-109906750**	FLJ25363 (0)
chr22 47512301-47512418	TBC1D22A (2.38)	**chr3 110041166-110041331**	PVRL3-AS1 (0)
chr22 49583686-49583814	LOC100128946 (0)	**chr3 110050519-110050654**	PVRL3-AS1 (0)
chr22 49788681-49788798	C22orf34 (3.53)	**chr3 110052452-110052582**	PVRL3-AS1 (0)
chr22 50051070-50051186	BRD1 (2.59)	**chr3 110223933-110224068**	PVRL3-AS1 (0)
chr22 50133202-50133341	BRD1 (2.59)	**chr3 112475278-112475398**	CD200R1L (0)
chr3 3168700-3168801	IL5RA (0.65)	**chr3 113806556-113806849**	KIAA1407 (0.57)
chr3 6210149-6210254	GRM7 (0.1)	**chr3 116911883-116912005**	LSAMP-AS3 (0)
chr3 7382004-7382128	LOC100288428 (0)	**chr3 118528784-118528903**	IGSF11 (1.24)
chr3 11448120-11448237	ATG7 (2.6)	**chr3 119823286-119823423**	GSK3B (2.7)
chr3 12817628-12817750	TMEM40 (1.47)	**chr3 125960805-125960924**	ALDH1L1 (2.3)
chr3 12882893-12883039	RPL32 (9.83)	**chr3 131477669-131477797**	CPNE4 (2.63)
chr3 17064939-17065118	PLCL2 (4.34)	**chr3 132618659-132618780**	NPHP3-AS1 (0)
chr3 19036703-19036817	KCNH8 (0.04)	**chr3 140480163-140480294**	TRIM42 (0)
chr3 21998569-21998690	ZNF385D (0)	**chr3 143739918-143740038**	C3orf58 (3.12)
chr3 25824794-25824936	NGLY1 (4.92)	**chr3 147672662-147672792**	ZIC1 (0)
chr3 26160821-26160949	LOC285326 (0)	**chr3 150321069-150321232**	SELT (4.58)
chr3 26188417-26188540	LOC285326 (0)	**chr3 151269928-151270038**	MIR548H2 (0)
chr3 28065192-28065322	CMC1 (5.2)	**chr3 151305605-151305721**	MIR548H2 (0)
chr3 28067396-28067494	CMC1 (5.2)	**chr3 156287838-156287952**	SSR3 (5.73)
chr3 31113969-31114155	GADL1 (0.06)	**chr3 157282505-157282627**	C3orf55 (0)
chr3 36764175-36764289	TRANK1 (0.72)	**chr3 160296297-160296421**	KPNA4 (4.36)
chr3 37184373-37184598	LRRFIP2 (4.23)	**chr3 161319990-161320101**	OTOL1 (0)
chr3 39183424-39183541	CSRNP1 (3.16)	**chr3 161494019-161494147**	OTOL1 (0)
chr3 40836781-40836894	ZNF621 (1.74)	**chr3 161668128-161668232**	OTOL1 (0)
chr3 45833145-45833259	LZTFL1 (0.09)	**chr3 161934237-161934383**	OTOL1 (0)

chr3 162366357-162366469	LOC647107 (0)	**chr4 23107124-23107236**	MIR548AJ2 (0)
chr3 162904439-162904558	LOC647107 (0)	**chr4 23109704-23109834**	MIR548AJ2 (0)
chr3 163199722-163199858	LOC647107 (0)	**chr4 23128754-23128873**	MIR548AJ2 (0)
chr3 164366444-164366571	SI (0)	**chr4 26019686-26019811**	C4orf52 (0)
chr3 166641920-166642045	ZBBX (0.34)	**chr4 26042091-26042205**	C4orf52 (0)
chr3 167888066-167888183	GOLIM4 (4.81)	**chr4 26824726-26824841**	STIM2 (3.65)
chr3 168236523-168236632	EGFEM1P (1.4)	**chr4 27740178-27740315**	STIM2 (3.65)
chr3 169530347-169530470	LRRIQ4 (0)	**chr4 30106407-30106529**	PCDH7 (0)
chr3 171076088-171076209	PLD1 (0.93)	**chr4 30615702-30615938**	PCDH7 (0)
chr3 172295191-172295315	NCEH1 (2.55)	**chr4 30648777-30648916**	PCDH7 (0)
chr3 172296181-172296294	NCEH1 (2.55)	**chr4 31194645-31194776**	PCDH7 (0)
chr3 173559818-173559935	NLGN1 (0.04)	**chr4 35454465-35454575**	ARAP2 (3.34)
chr3 174991887-174992020	NAALADL2 (0.08)	**chr4 35503508-35503636**	ARAP2 (3.34)
chr3 177345987-177346120	TBL1XR1 (4.81)	**chr4 36617046-36617154**	DTHD1 (0)
chr3 177852488-177852596	KCNMB2 (0.88)	**chr4 36733685-36733826**	DTHD1 (0)
chr3 178692133-178692248	ZMAT3 (4.27)	**chr4 36934956-36935070**	MIR4801 (0)
chr3 181754497-181754618	SOX2-OT (2.28)	**chr4 38420063-38420176**	FLJ13197 (0)
chr3 183079282-183079403	MCF2L2 (2.71)	**chr4 38552292-38552421**	FLJ13197 (0)
chr3 183985148-183985283	CAMK2N2 (3.18)	**chr4 38824545-38824663**	TLR6 (1.88)
chr3 184443720-184443842	MAGEF1 (5.83)	**chr4 40300684-40300796**	CHRNA9 (0.44)
chr3 186314754-186314984	DNAJB11 (5.85)	**chr4 40922228-40922349**	APBB2 (0.72)
chr3 189330145-189330269	TP63 (5)	**chr4 41848113-41848220**	TMEM33 (4.37)
chr3 190895729-190895834	OSTN (0.69)	**chr4 42465374-42465490**	ATP8A1 (3.89)
chr3 190934865-190934975	UTS2D (0)	**chr4 42567010-42567126**	ATP8A1 (3.89)
chr3 191201933-191202073	PYDC2 (0)	**chr4 45735750-45735867**	GABRG1 (0)
chr3 192492143-192492233	MB21D2 (2.67)	**chr4 46725841-46726028**	COX7B2 (0)
chr3 194361588-194361698	LSG1 (5.8)	**chr4 46913998-46914113**	COX7B2 (0)
chr3 194624360-194624476	LOC100507391 (0.25)	**chr4 47200833-47200934**	GABRB1 (0)
chr3 195913913-195914039	ZDHHC19 (0)	**chr4 47638642-47638763**	CORIN (0.43)
chr3 197820107-197820243	LOC348840 (0)	**chr4 57489850-57489981**	HOPX (0.24)
chr4 467844-467966	PIGG (3.99)	**chr4 58322341-58322470**	LOC255130 (0)
chr4 1042619-1042741	FGFRL1 (2.38)	**chr4 63711162-63711279**	LPHN3 (0)
chr4 1250127-1250258	C4orf42 (0)	**chr4 67823473-67823594**	CENPC1 (0)
chr4 1967429-1967635	WHSC1 (5.52)	**chr4 71993877-71993992**	SLC4A4 (0.07)
chr4 2384540-2384669	ZFYVE28 (1.6)	**chr4 72628445-72628572**	SLC4A4 (0.07)
chr4 4369974-4370095	D4S234E (0)	**chr4 72975921-72976043**	ADAMTS3 (0.18)
chr4 4552170-4552300	LOC100507266 (0)	**chr4 73079594-73079716**	NPFFR2 (0)
chr4 4925536-4925648	MSX1 (2.49)	**chr4 74682247-74682366**	CXCL6 (0)
chr4 5068800-5068921	STK32B (0)	**chr4 74992724-74992846**	CXCL2 (0.12)
chr4 5129631-5129722	STK32B (0)	**chr4 76072999-76073174**	PARM1 (5.62)
chr4 5551655-5551774	EVC2 (0.05)	**chr4 76458840-76458953**	THAP6 (3.25)
chr4 7642986-7643092	SORCS2 (0.08)	**chr4 77446335-77446468**	SHROOM3 (0.39)
chr4 7755772-7755895	AFAP1 (1.49)	**chr4 77870073-77870200**	40787 (5.92)
chr4 8289191-8289305	HTRA3 (1.21)	**chr4 79102043-79102148**	FRAS1 (0)
chr4 11687579-11687704	HS3ST1 (0.14)	**chr4 79327531-79327661**	FRAS1 (0)
chr4 14346848-14346940	LOC152742 (0)	**chr4 81797535-81797661**	C4orf22 (0)
chr4 15910643-15910769	FGFBP1 (0)	**chr4 85577970-85578093**	CDS1 (1.48)
chr4 19218037-19218164	SLIT2 (1.93)	**chr4 86478326-86478424**	ARHGAP24 (2.74)
chr4 19309252-19309371	SLIT2 (1.93)	**chr4 86696554-86696674**	ARHGAP24 (2.74)
chr4 19987712-19987827	SLIT2 (1.93)	**chr4 86886304-86886419**	ARHGAP24 (2.74)
chr4 20785027-20785157	KCNIP4 (0.93)	**chr4 93716479-93716600**	GRID2 (0)
chr4 22304594-22304703	LOC100505912 (0)	**chr4 95983106-95983234**	UNC5C (0.18)
chr4 22789802-22789942	GBA3 (1.26)	**chr4 96664529-96664646**	PDHA2 (0)
chr4 22999083-22999247	GBA3 (1.26)	**chr4 98356729-98356842**	C4orf37 (0)
chr4 23103070-23103207	MIR548AJ2 (0)	**chr4 101074258-101074391**	DDIT4L (0.36)
chr4 23105174-23105306	MIR548AJ2 (0)	**chr4 101087879-101088018**	DDIT4L (0.36)

chr4 104337479-104337612	TACR3 (0)	**chr5 12079994-12080115**	CTNND2 (0)
chr4 104417098-104417215	TACR3 (0)	**chr5 12873507-12873622**	TAG (0)
chr4 104422746-104422853	TACR3 (0)	**chr5 14034559-14034685**	DNAH5 (0)
chr4 107615537-107615671	DKK2 (0.07)	**chr5 17209315-17209439**	LOC285696 (0)
chr4 111140368-111140499	ELOVL6 (3.99)	**chr5 17259216-17259334**	LOC285696 (0)
chr4 111520230-111520340	PITX2 (0.76)	**chr5 17288606-17288724**	BASP1 (5.46)
chr4 113033337-113033459	C4orf32 (3.52)	**chr5 21218508-21218598**	GUSBP1 (4.19)
chr4 113759346-113759456	ANK2 (2.64)	**chr5 21800943-21801062**	CDH12 (2.93)
chr4 114320974-114321096	ANK2 (2.64)	**chr5 21918618-21918769**	CDH12 (2.93)
chr4 115335677-115335804	UGT8 (1.67)	**chr5 22618984-22619110**	CDH12 (2.93)
chr4 116791424-116791537	MIR1973 (0)	**chr5 23828374-23828485**	PRDM9 (0)
chr4 118060012-118060146	TRAM1L1 (0)	**chr5 25124445-25124567**	LOC340107 (0)
chr4 118179206-118179319	TRAM1L1 (0)	**chr5 25536644-25536764**	LOC340107 (0)
chr4 118400832-118400955	TRAM1L1 (0)	**chr5 25674171-25674278**	LOC340107 (0)
chr4 119003984-119004133	NDST3 (0.02)	**chr5 28016477-28016597**	LOC643401 (0)
chr4 120221489-120221773	USP53 (1.32)	**chr5 30476257-30476378**	CDH6 (0)
chr4 121220558-121220686	MAD2L1 (6.66)	**chr5 38868624-38868756**	RICTOR (3.53)
chr4 122761359-122761481	BBS7 (3.71)	**chr5 40604022-40604201**	PTGER4 (5.73)
chr4 127780559-127780678	INTU (0)	**chr5 42101095-42101215**	FBXO4 (2.74)
chr4 130755766-130755888	C4orf33 (4.03)	**chr5 42103042-42103151**	FBXO4 (2.74)
chr4 132405716-132405829	PCDH10 (0)	**chr5 43069011-43069133**	LOC100132356 (2.23)
chr4 133184021-133184155	PCDH10 (0)	**chr5 43549536-43549694**	PAIP1 (5.4)
chr4 133604114-133604218	PCDH10 (0)	**chr5 51138566-51138687**	ISL1 (0.32)
chr4 139589035-139589152	CCRN4L (4.59)	**chr5 53356302-53356425**	ARL15 (2)
chr4 139653397-139653509	CCRN4L (4.59)	**chr5 53609465-53609578**	ARL15 (2)
chr4 139673476-139673580	CCRN4L (4.59)	**chr5 56597922-56598031**	GPBP1 (5.09)
chr4 144792053-144792181	GYPE (0)	**chr5 58733708-58733839**	PDE4D (2.77)
chr4 144912848-144912991	GYPB (0)	**chr5 64897187-64897307**	PPWD1 (4.02)
chr4 145016102-145016239	GYPA (0)	**chr5 65167650-65167782**	NLN (2.96)
chr4 147887382-147887512	TTC29 (0)	**chr5 65375436-65375526**	ERBB2IP (4.42)
chr4 153116443-153116573	FBXW7 (5.43)	**chr5 66786402-66786522**	CD180 (0)
chr4 154994458-154994568	DCHS2 (0)	**chr5 66966030-66966145**	CD180 (0)
chr4 156167742-156167873	NPY2R (0)	**chr5 67794754-67794880**	PIK3R1 (3.64)
chr4 162021482-162021598	FSTL5 (0.96)	**chr5 68473336-68473561**	CENPH (6.04)
chr4 162585960-162586082	FSTL5 (0.96)	**chr5 78857431-78857542**	HOMER1 (3.84)
chr4 164336758-164336849	TKTL2 (0.07)	**chr5 79409388-79409509**	SERINC5 (2.67)
chr4 168581999-168582114	SPOCK3 (0)	**chr5 79696226-79696356**	ZFYVE16 (2.45)
chr4 172396393-172396507	GALNTL6 (0)	**chr5 82028369-82028494**	MIR3977 (0)
chr4 174782105-174782220	NBLA00301 (0)	**chr5 82307545-82307636**	TMEM167A (4.12)
chr4 175285029-175285255	CEP44 (3.86)	**chr5 82623867-82623988**	VCAN (0.03)
chr4 175319470-175319593	MIR4276 (0)	**chr5 83032177-83032313**	HAPLN1 (0)
chr4 181040090-181040221	LINC00290 (0)	**chr5 83670614-83670722**	EDIL3 (0)
chr4 183734096-183734218	ODZ3 (0)	**chr5 87343408-87343520**	TMEM161B (3.65)
chr4 185296924-185297046	IRF2 (3.48)	**chr5 88714524-88714730**	MEF2C (1.64)
chr4 185451416-185451557	IRF2 (3.48)	**chr5 89589329-89589450**	CETN3 (4.19)
chr4 188216606-188216735	LOC339975 (0)	**chr5 91206673-91206808**	LOC100129716 (0)
chr4 189441213-189441338	LOC401164 (0)	**chr5 92840287-92840416**	FLJ42709 (0)
chr4 189444110-189444226	LOC401164 (0)	**chr5 95382456-95382588**	MIR583 (0)
chr5 1318016-1318190	CLPTM1L (5.41)	**chr5 95865465-95865591**	PCSK1 (0.17)
chr5 1507739-1507858	SLC6A3 (0)	**chr5 97403795-97403956**	RGMB (1.62)
chr5 2071651-2071772	IRX4 (0.12)	**chr5 97816271-97816385**	RGMB (1.62)
chr5 5006902-5007019	LOC340094 (0)	**chr5 98092686-98092824**	RGMB (1.62)
chr5 6271931-6272062	FLJ33360 (0.25)	**chr5 98343946-98344074**	LOC100289230 (1.17)
chr5 9974495-9974614	LOC285692 (0)	**chr5 99445400-99445533**	LOC100133050 (0)
chr5 11057535-11057666	CTNND2 (0)	**chr5 100208311-100208433**	ST8SIA4 (3.13)
chr5 11253584-11253704	CTNND2 (0)	**chr5 100219454-100219558**	ST8SIA4 (3.13)

chr5 100362636-100362758	ST8SIA4 (3.13)	**chr6 11295129-11295321**	NEDD9 (3.89)
chr5 100963562-100963674	SLCO4C1 (0)	**chr6 13472206-13472327**	GFOD1 (2.27)
chr5 108113416-108113513	FER (2.87)	**chr6 15123088-15123220**	JARID2 (2.43)
chr5 111257558-111257697	C5orf13 (0)	**chr6 20431841-20431964**	CDKAL1 (4.12)
chr5 111494298-111494466	EPB41L4A-AS1 (4.46)	**chr6 21169973-21170118**	CDKAL1 (4.12)
chr5 115518620-115518735	COMMD10 (4.07)	**chr6 22734334-22734452**	HDGFL1 (0)
chr5 116556692-116556807	SEMA6A (1.4)	**chr6 23996964-23997102**	NRSN1 (0.32)
chr5 117818356-117818485	DTWD2 (1.89)	**chr6 24014151-24014281**	NRSN1 (0.32)
chr5 119713260-119713375	PRR16 (0)	**chr6 24024069-24024187**	NRSN1 (0.32)
chr5 119741799-119741935	PRR16 (0)	**chr6 25946121-25946232**	SLC17A2 (0)
chr5 120500455-120500584	PRR16 (0)	**chr6 25985933-25986160**	TRIM38 (3.29)
chr5 121935243-121935354	SNCAIP (2.54)	**chr6 31021188-31021318**	HCG22 (0.76)
chr5 123491336-123491461	ZNF608 (1.96)	**chr6 34231234-34231344**	C6orf1 (4.83)
chr5 123748528-123748657	ZNF608 (1.96)	**chr6 36321585-36321703**	ETV7 (0.1)
chr5 123823912-123824042	ZNF608 (1.96)	**chr6 39938962-39939092**	MOCS1 (0.03)
chr5 129630687-129630815	CHSY3 (0)	**chr6 40741027-40741160**	LRFN2 (0)
chr5 133619054-133619185	CDKL3 (1.25)	**chr6 45993575-45993732**	CLIC5 (0.17)
chr5 139624818-139624927	C5orf32 (0)	**chr6 52860445-52860716**	GSTA4 (2.05)
chr5 141019039-141019151	RELL2 (1.85)	**chr6 53855796-53855918**	MLIP (0)
chr5 145374036-145374175	SH3RF2 (0.91)	**chr6 54889433-54889548**	FAM83B (0)
chr5 149028343-149028473	ARHGEF37 (0.17)	**chr6 63320779-63320886**	KHDRBS2 (0)
chr5 149472092-149472218	PDGFRB (0)	**chr6 63355169-63355285**	KHDRBS2 (0)
chr5 149776171-149776415	CD74 (9.46)	**chr6 64450184-64450311**	EYS (0.26)
chr5 150199307-150199435	C5orf62 (0)	**chr6 65553886-65554007**	EYS (0.26)
chr5 151257268-151257390	GLRA1 (0.4)	**chr6 67104453-67104575**	MCART3P (0)
chr5 154423824-154423946	KIF4B (0)	**chr6 67778160-67778281**	MCART3P (0)
chr5 154450696-154450818	KIF4B (0)	**chr6 67780075-67780190**	MCART3P (0)
chr5 154452619-154452740	KIF4B (0)	**chr6 67869770-67869884**	MCART3P (0)
chr5 156640117-156640235	ITK (0.52)	**chr6 72119979-72120096**	LINC00472 (0)
chr5 157249362-157249484	CLINT1 (4.87)	**chr6 72166018-72166145**	LINC00472 (0)
chr5 157465418-157465525	CLINT1 (4.87)	**chr6 77223860-77223981**	IMPG1 (0)
chr5 157832016-157832140	EBF1 (1.89)	**chr6 81662242-81662388**	BCKDHB (3.13)
chr5 163213711-163213832	MAT2B (4.37)	**chr6 82486881-82486988**	FAM46A (0.82)
chr5 163681927-163682042	MAT2B (4.37)	**chr6 82804929-82805054**	IBTK (3.86)
chr5 166959410-166959545	WWC1 (2.33)	**chr6 84544298-84544412**	RIPPLY2 (0)
chr5 169492585-169492698	DOCK2 (6.23)	**chr6 84673644-84673773**	CYB5R4 (3.79)
chr5 171404636-171404758	STK10 (3.67)	**chr6 85506218-85506339**	TBX18 (0)
chr5 172107987-172108092	NEURL1B (0.02)	**chr6 86417339-86417469**	SNHG5 (7.23)
chr5 172712384-172712515	STC2 (0)	**chr6 87772755-87772877**	CGA (0.31)
chr5 174452916-174453031	MIR4634 (0)	**chr6 89207315-89207423**	RNGTT (3.85)
chr5 175192952-175193062	CPLX2 (0.04)	**chr6 89244214-89244328**	RNGTT (3.85)
chr5 176883916-176884139	DBN1 (4.72)	**chr6 89270862-89270981**	RNGTT (3.85)
chr5 177625153-177625260	HNRNPAB (6.87)	**chr6 90681362-90681486**	BACH2 (0.06)
chr5 178824706-178824828	ADAMTS2 (0.13)	**chr6 91964482-91964598**	MIR4643 (0)
chr5 179216405-179216521	LTC4S (1.11)	**chr6 92373717-92373824**	MIR4643 (0)
chr5 179264031-179264214	C5orf45 (4.78)	**chr6 96333985-96334134**	FUT9 (0.61)
chr6 3195918-3196030	TUBB2B (5.57)	**chr6 100070087-100070210**	PRDM13 (1.93)
chr6 4422170-4422289	CDYL (2.94)	**chr6 103230605-103230727**	GRIK2 (0.04)
chr6 5548217-5548434	LYRM4 (5.56)	**chr6 107229505-107229624**	LOC100422737 (2.05)
chr6 5555330-5555499	LYRM4 (5.56)	**chr6 110565753-110565859**	C6orf186 (0)
chr6 5795607-5795729	FARS2 (5.38)	**chr6 110630750-110630866**	C6orf186 (0)
chr6 5805608-5805737	FARS2 (5.38)	**chr6 110783166-110783296**	SLC22A16 (0)
chr6 5864969-5865102	FARS2 (5.38)	**chr6 111599798-111599892**	KIAA1919 (2.66)
chr6 6580563-6580678	LY86-AS1 (0)	**chr6 112801158-112801288**	RFPL4B (0.11)
chr6 7313289-7313433	CAGE1 (0.35)	**chr6 113055478-113055596**	RFPL4B (0.11)
chr6 8064558-8064663	EEF1E1 (6.1)	**chr6 113777052-113777147**	MARCKS (5.21)

chr6 113832310-113832414	MARCKS (5.21)	**chr7 7194485-7194576**	C1GALT1 (2.62)
chr6 116910843-116910975	RSPH4A (0.31)	**chr7 7865054-7865169**	LOC729852 (0)
chr6 119419339-119419460	FAM184A (2.53)	**chr7 11776593-11776699**	THSD7A (0)
chr6 123225588-123225733	CLVS2 (0)	**chr7 12082291-12082414**	TMEM106B (2.76)
chr6 123445554-123445679	CLVS2 (0)	**chr7 12971267-12971401**	ARL4A (0.07)
chr6 125169241-125169362	NKAIN2 (0)	**chr7 14087835-14087941**	ETV1 (0.02)
chr6 125171188-125171309	NKAIN2 (0)	**chr7 15898080-15898211**	MEOX2 (0)
chr6 125256174-125256297	RNF217 (0)	**chr7 16761403-16761503**	BZW2 (6.07)
chr6 125271779-125271885	RNF217 (0)	**chr7 17014271-17014397**	AGR3 (0)
chr6 125351126-125351257	RNF217 (0)	**chr7 17307952-17308069**	AHR (3.76)
chr6 127969611-127969728	C6orf58 (3.11)	**chr7 19298407-19298525**	FERD3L (0)
chr6 129908666-129908787	ARHGAP18 (2.43)	**chr7 26129289-26129418**	NFE2L3 (5.02)
chr6 131083087-131083202	LOC100507203 (0)	**chr7 32029898-32030036**	PDE1C (0.08)
chr6 131429412-131429534	AKAP7 (1.55)	**chr7 34207882-34208011**	BMPER (0.22)
chr6 134031887-134032002	MGC34034 (0)	**chr7 34329153-34329275**	AAA1 (0)
chr6 134551540-134551656	SGK1 (1.91)	**chr7 37904353-37904476**	SFRP4 (0.07)
chr6 134719740-134719865	LOC154092 (0)	**chr7 44053216-44053326**	POLR2J4 (1.25)
chr6 136048251-136048371	LINC00271 (0.09)	**chr7 44545460-44545574**	NPC1L1 (0)
chr6 136546672-136546790	FAM54A (0)	**chr7 45412789-45412914**	RAMP3 (0)
chr6 137600512-137600634	IFNGR1 (4.17)	**chr7 48512547-48512678**	ABCA13 (1.64)
chr6 137613988-137614110	IFNGR1 (4.17)	**chr7 49336977-49337106**	CDC14C (0)
chr6 138155096-138155227	TNFAIP3 (3.64)	**chr7 49444789-49444985**	VWC2 (0)
chr6 138177451-138177575	TNFAIP3 (3.64)	**chr7 49453993-49454121**	VWC2 (0)
chr6 139329331-139329463	REPS1 (4.28)	**chr7 49851016-49851132**	ZPBP (0)
chr6 139370325-139370467	C6orf115 (0)	**chr7 66071809-66071988**	KCTD7 (2.2)
chr6 140710464-140710589	MIR3668 (0)	**chr7 67592715-67592832**	STAG3L4 (2.64)
chr6 141825587-141825709	NMBR (0)	**chr7 68113239-68113357**	AUTS2 (1.02)
chr6 143378672-143378790	AIG1 (4)	**chr7 68368643-68368764**	AUTS2 (1.02)
chr6 144238307-144238411	PLAGL1 (0.27)	**chr7 68509858-68509952**	AUTS2 (1.02)
chr6 145790207-145790350	EPM2A (1.5)	**chr7 68782079-68782210**	AUTS2 (1.02)
chr6 146028718-146028850	LOC100507557 (0.31)	**chr7 69740687-69740852**	AUTS2 (1.02)
chr6 146030568-146030703	LOC100507557 (0.31)	**chr7 70741573-70741663**	WBSCR17 (2.32)
chr6 146051146-146051264	LOC100507557 (0.31)	**chr7 71222099-71222225**	CALN1 (0.26)
chr6 147165424-147165533	LOC729178 (0)	**chr7 71401871-71401992**	CALN1 (0.26)
chr6 147943947-147944041	SAMD5 (0.04)	**chr7 71508868-71508977**	CALN1 (0.26)
chr6 150740807-150740948	IYD (0)	**chr7 75137123-75137283**	PMS2P3 (3.38)
chr6 150814912-150815033	IYD (0)	**chr7 76157033-76157146**	UPK3B (1.46)
chr6 152397391-152397513	SYNE1 (0.52)	**chr7 78809667-78809789**	MAGI2 (0.32)
chr6 154681665-154681778	IPCEF1 (0)	**chr7 78811584-78811719**	MAGI2 (0.32)
chr6 156179417-156179533	NOX3 (0)	**chr7 79922983-79923108**	GNAI1 (0)
chr6 156614775-156614880	ARID1B (2.46)	**chr7 80173990-80174111**	GNAT3 (0)
chr6 157125637-157125879	ARID1B (2.46)	**chr7 81140639-81140768**	HGF (6.76)
chr6 158215240-158215371	SNX9 (0.03)	**chr7 81315453-81315571**	HGF (6.76)
chr6 159017717-159017848	TMEM181 (3.52)	**chr7 82225303-82225427**	CACNA2D1 (3.05)
chr6 160085754-160085855	SOD2 (5.12)	**chr7 82523700-82523849**	CACNA2D1 (3.05)
chr6 162490273-162490394	PACRG (0)	**chr7 83261812-83261935**	SEMA3A (0.83)
chr6 164280065-164280190	QKI (4.02)	**chr7 83290416-83290555**	SEMA3E (0.34)
chr6 164469281-164469397	QKI (4.02)	**chr7 83383708-83383883**	SEMA3E (0.34)
chr6 165448876-165448999	C6orf118 (0.12)	**chr7 83449448-83449584**	SEMA3A (0.83)
chr6 168477218-168477336	FRMD1 (0.07)	**chr7 83472782-83472912**	SEMA3A (0.83)
chr7 1899529-1900067	MAD1L1 (3.94)	**chr7 83510567-83510689**	SEMA3A (0.83)
chr7 2157229-2157319	MAD1L1 (3.94)	**chr7 84045574-84045703**	SEMA3A (0.83)
chr7 3202569-3202698	CARD11 (0.75)	**chr7 88642219-88642349**	ZNF804B (0)
chr7 4259016-4259138	SDK1 (1.11)	**chr7 88768077-88768225**	ZNF804B (0)
chr7 4444647-4444782	SDK1 (1.11)	**chr7 91096078-91096186**	FZD1 (2.45)
chr7 4806877-4806972	KIAA0415 (0)	**chr7 92244323-92244568**	CDK6 (4.68)

chr7 92373549-92373663	CDK6 (4.68)	**chr7 155154320-155154414**	INSIG1 (7.7)
chr7 92561993-92562117	CDK6 (4.68)	**chr7 156558030-156558162**	LMBR1 (5.09)
chr7 93586838-93586959	GNG11 (2.01)	**chr7 156652902-156653027**	NOM1 (5.17)
chr7 93652075-93652228	BET1 (3.59)	**chr7 157677286-157677409**	PTPRN2 (5.06)
chr7 95542827-95542953	DYNC1I1 (1.07)	**chr7 158172529-158172655**	PTPRN2 (5.06)
chr7 96380352-96380471	SHFM1 (6.65)	**chr7 158546785-158546903**	ESYT2 (5.85)
chr7 99006226-99006344	BUD31 (6.42)	**chr7 158692389-158692508**	WDR60 (2.17)
chr7 99229631-99229867	ZNF498 (0)	**chr8 1710850-1710964**	CLN8 (0.43)
chr7 101313587-101313708	MYL10 (0)	**chr8 2220366-2220478**	MYOM2 (1.86)
chr7 101982578-101982689	SPDYE6 (1.06)	**chr8 2347312-2347434**	MYOM2 (1.86)
chr7 102207468-102207583	SPDYE2 (0)	**chr8 2735368-2735545**	CSMD1 (2.87)
chr7 102212937-102213045	RASA4 (1.74)	**chr8 2980170-2980300**	CSMD1 (2.87)
chr7 102306548-102306658	SPDYE2L (0)	**chr8 4632320-4632426**	CSMD1 (2.87)
chr7 102311991-102312125	SPDYE2L (0)	**chr8 8779595-8779707**	MFHAS1 (4.12)
chr7 105648378-105648505	CDHR3 (0.02)	**chr8 9915200-9915328**	MSRA (4.76)
chr7 106145143-106145286	C7orf74 (0)	**chr8 10143888-10144012**	MSRA (4.76)
chr7 114780613-114780707	MDFIC (5.66)	**chr8 13488629-13488749**	C8orf48 (0)
chr7 116103685-116103810	CAV2 (3.28)	**chr8 15199100-15199214**	SGCZ (0.35)
chr7 116227049-116227179	CAV1 (6.19)	**chr8 15274678-15274806**	TUSC3 (0.19)
chr7 116633337-116633458	ST7 (4.06)	**chr8 16314431-16314621**	MSR1 (4.21)
chr7 116907998-116908121	WNT2 (0.07)	**chr8 16607140-16607258**	FGF20 (0)
chr7 117081159-117081256	ASZ1 (0)	**chr8 18820522-18820653**	PSD3 (3.96)
chr7 117634141-117634262	CTTNBP2 (0.03)	**chr8 19075647-19075776**	SH2D4A (0.31)
chr7 117636072-117636203	CTTNBP2 (0.03)	**chr8 20917367-20917499**	LOC286114 (0)
chr7 118561343-118561472	ANKRD7 (0.43)	**chr8 22478352-22478546**	BIN3 (3.56)
chr7 119568480-119568576	KCND2 (0.65)	**chr8 23618276-23618392**	NKX2-6 (0)
chr7 121158265-121158366	FAM3C (3.21)	**chr8 25957412-25957531**	EBF2 (1.61)
chr7 121403786-121403903	PTPRZ1 (0)	**chr8 27787787-27787918**	SCARA5 (0.33)
chr7 121590110-121590231	PTPRZ1 (0)	**chr8 28162858-28162988**	PNOC (0)
chr7 123193302-123193422	NDUFA5 (3.54)	**chr8 32293613-32293736**	NRG1 (0.63)
chr7 123746459-123746572	TMEM229A (0)	**chr8 47360155-47360253**	LINC00293 (0)
chr7 124228251-124228369	GPR37 (0)	**chr8 49430546-49430678**	EFCAB1 (0)
chr7 124996208-124996332	POT1 (3.76)	**chr8 51097806-51097928**	SNTG1 (0)
chr7 125469005-125469115	GRM8 (2.22)	**chr8 52213619-52213756**	PXDNL (1.8)
chr7 125757184-125757300	GRM8 (2.22)	**chr8 52215990-52216109**	PXDNL (1.8)
chr7 125865644-125865744	GRM8 (2.22)	**chr8 52643299-52643410**	PXDNL (1.8)
chr7 125916766-125916882	GRM8 (2.22)	**chr8 55074557-55074747**	MRPL15 (3.64)
chr7 127172211-127172319	GCC1 (2.51)	**chr8 55444230-55444347**	SOX17 (0)
chr7 133403913-133404044	EXOC4 (5.64)	**chr8 58890928-58891050**	FAM110B (0.06)
chr7 133519714-133519839	EXOC4 (5.64)	**chr8 59220739-59220861**	UBXN2B (3.31)
chr7 135015912-135016024	CNOT4 (3.6)	**chr8 60650322-60650442**	CA8 (2.06)
chr7 135777668-135777761	LUZP6 (6.46)	**chr8 60850776-60850891**	CA8 (2.06)
chr7 139677026-139677157	TBXAS1 (1.15)	**chr8 63655404-63655547**	NKAIN3 (0)
chr7 140895572-140895705	LOC100507421 (0)	**chr8 64579000-64579160**	LOC286184 (0)
chr7 141657244-141657375	CLEC5A (0.17)	**chr8 66353812-66353917**	LOC286186 (0)
chr7 143203011-143203116	LOC285965 (0)	**chr8 66792827-66792955**	PDE7A (3)
chr7 143715940-143716067	OR6B1 (0)	**chr8 71452839-71452956**	TRAM1 (5.71)
chr7 143912886-143913013	OR2A1 (0)	**chr8 72804836-72804979**	LOC100132891 (0)
chr7 144032163-144032290	OR2A42 (0)	**chr8 74498135-74498268**	STAU2 (3.57)
chr7 147229759-147229888	MIR548I4 (0)	**chr8 75012269-75012389**	LY96 (2.01)
chr7 148336253-148336383	C7orf33 (4.98)	**chr8 82963475-82963631**	SNX16 (1.4)
chr7 151462290-151462399	PRKAG2 (1.16)	**chr8 83116810-83116949**	SNX16 (1.4)
chr7 153161865-153161986	DPP6 (1.14)	**chr8 83161079-83161191**	SNX16 (1.4)
chr7 154199594-154199717	DPP6 (1.14)	**chr8 85839152-85839259**	RALYL (0)
chr7 154544316-154544430	DPP6 (1.14)	**chr8 90366501-90366616**	RIPK2 (3.82)
chr7 154998362-154998489	INSIG1 (7.7)	**chr8 96135068-96135197**	PLEKHF2 (2.06)

chr8 97400855-97400973	PTDSS1 (5.24)	**chr9 38261895-38262012**	ALDH1B1 (3.91)
chr8 99918572-99918681	OSR2 (1.4)	**chr9 41902205-41902313**	MGC21881 (0)
chr8 101141485-101141575	FBXO43 (1.8)	**chr9 42234197-42234325**	ANKRD20A2 (0.1)
chr8 102002266-102002389	YWHAZ (6.43)	**chr9 42778394-42778549**	FOXD4L2 (0)
chr8 104284589-104284695	FZD6 (0.8)	**chr9 42842794-42842970**	AQP7P3 (0.71)
chr8 104576009-104576124	RIMS2 (0.43)	**chr9 44454563-44454671**	CNTNAP3B (0)
chr8 106056315-106056445	ZFPM2 (0)	**chr9 46896418-46896526**	KGFLP1 (0.93)
chr8 109353477-109353605	EIF3E (7.95)	**chr9 66533939-66534047**	LOC442421 (0)
chr8 111580975-111581096	KCNV1 (0)	**chr9 67015830-67015977**	LOC100133920 (0)
chr8 114769781-114769902	CSMD3 (0.04)	**chr9 67304408-67304555**	AQP7P1 (0.25)
chr8 117864193-117864298	RAD21-AS1 (1.71)	**chr9 69666379-69666496**	LOC100133920 (0)
chr8 118532943-118533061	EXT1 (4.89)	**chr9 69750941-69751103**	LOC100133920 (0)
chr8 121838784-121838919	SNTB1 (0.5)	**chr9 69801825-69802017**	LOC100133920 (0)
chr8 121875299-121875411	SNTB1 (0.5)	**chr9 70117634-70117796**	FOXD4L5 (0.46)
chr8 121947688-121947800	SNTB1 (0.5)	**chr9 70368075-70368241**	FOXD4L4 (0)
chr8 121951793-121951902	SNTB1 (0.5)	**chr9 70629019-70629153**	CBWD3 (1.64)
chr8 124250507-124250631	ZHX1 (3.82)	**chr9 72951817-72951942**	KLF9 (1.75)
chr8 124503997-124504118	FBXO32 (1.27)	**chr9 74272369-74272487**	TMEM2 (2.69)
chr8 124685059-124685190	ANXA13 (0)	**chr9 75887760-75887868**	ANXA1 (5.56)
chr8 124692264-124692385	ANXA13 (0)	**chr9 81325675-81325812**	PSAT1 (6.06)
chr8 125740159-125740318	MTSS1 (2.31)	**chr9 81341581-81341707**	PSAT1 (6.06)
chr8 125755048-125755206	MTSS1 (2.31)	**chr9 83145433-83145538**	TLE4 (1.11)
chr8 125773036-125773195	MTSS1 (2.31)	**chr9 83147352-83147481**	TLE4 (1.11)
chr8 126333127-126333279	NSMCE2 (4.45)	**chr9 83714051-83714180**	TLE1 (0.99)
chr8 126360901-126361025	NSMCE2 (4.45)	**chr9 83956434-83956555**	TLE1 (0.99)
chr8 126545229-126545352	TRIB1 (2.35)	**chr9 86181034-86181164**	FRMD3 (0.36)
chr8 127019457-127019588	LOC100130231 (0)	**chr9 89152562-89152673**	ZCCHC6 (3.72)
chr8 129255857-129255986	MIR1208 (0)	**chr9 90965448-90965569**	SPIN1 (3.19)
chr8 130034999-130035094	LOC728724 (0)	**chr9 90971262-90971383**	SPIN1 (3.19)
chr8 134249541-134249659	WISP1 (0.07)	**chr9 92470388-92470519**	UNQ6494 (3.28)
chr8 134392068-134392175	ST3GAL1 (1)	**chr9 92979115-92979231**	LOC286370 (0)
chr8 135340956-135341142	ZFAT (2.04)	**chr9 93273812-93273938**	LOC340515 (0)
chr8 135630238-135630359	ZFAT (2.04)	**chr9 93581721-93581843**	SYK (1.82)
chr8 140261510-140261641	COL22A1 (3.81)	**chr9 93583644-93583765**	SYK (1.82)
chr8 141965424-141965543	PTK2 (3.39)	**chr9 93759287-93759409**	LOC100129316 (0)
chr8 144272815-144272951	GPIHBP1 (0)	**chr9 95872954-95873083**	NINJ1 (5.87)
chr8 144274107-144274220	GPIHBP1 (0)	**chr9 100828114-100828235**	TRIM14 (3.62)
chr8 144672141-144672251	TIGD5 (1.93)	**chr9 102665046-102665176**	STX17 (3.25)
chr9 2303022-2303152	SMARCA2 (5.44)	**chr9 103308506-103308652**	C9orf30-TMEFF1 (0)
chr9 3304641-3304768	RFX3 (2.3)	**chr9 103801905-103802032**	LPPR1 (0)
chr9 9190406-9190527	PTPRD (0.04)	**chr9 105180427-105180550**	CYLC2 (0)
chr9 9442149-9442266	PTPRD (0.04)	**chr9 106225392-106225514**	CYLC2 (0)
chr9 12475847-12475968	TYRP1 (0)	**chr9 110550887-110551016**	KLF4 (0.09)
chr9 13384169-13384278	FLJ41200 (0)	**chr9 116448131-116448263**	RGS3 (1.87)
chr9 19037513-19037644	FAM154A (0)	**chr9 116460439-116460561**	RGS3 (1.87)
chr9 20012215-20012356	SLC24A2 (0.08)	**chr9 116520605-116520721**	ZNF618 (0)
chr9 22058803-22058932	CDKN2B-AS (0)	**chr9 116529195-116529323**	ZNF618 (0)
chr9 25435260-25435365	TUSC1 (0)	**chr9 117099092-117099187**	AKNA (3.28)
chr9 26157467-26157599	LOC100506422 (0)	**chr9 117297625-117297761**	DFNB31 (2.62)
chr9 26752856-26752975	C9orf82 (0)	**chr9 117446529-117446705**	LOC100505478 (0)
chr9 27075433-27075563	IFT74 (4.72)	**chr9 117490904-117491028**	LOC100505478 (0)
chr9 30559198-30559299	MIR873 (0)	**chr9 118284246-118284369**	37226 (0)
chr9 31578569-31578700	ACO1 (5.06)	**chr9 120280747-120280860**	ASTN2 (2.7)
chr9 32357080-32357188	ACO1 (5.06)	**chr9 124430168-124430297**	DAB2IP (3.26)
chr9 32796955-32797086	TMEM215 (0)	**chr9 125196271-125196401**	PTGS1 (1.51)
chr9 33917876-33917997	UBAP2 (4.35)	**chr9 129547273-129547403**	ZBTB43 (2.96)

chr9 130232519-130232641	LRSAM1 (3.46)
chr9 134967412-134967553	MED27 (4.51)
chrM 2802-3137	-1 (0)
chrM 5118-5251	-1 (0)
chrUn_gl000220 114688-115211	RN5-8S1 (0)
chrUn_gl000220 153851-154000	RN5-8S1 (0)
chrUn_gl000220 158643-159185	RN5-8S1 (0)
chrX 5831942-5832070	NLGN4X (1.6)
chrX 6140886-6141001	NLGN4X (1.6)
chrX 7029127-7029265	MIR4767 (0)
chrX 7285586-7285711	STS (0.76)
chrX 7819838-7819969	VCX (0.98)
chrX 8276428-8276549	VCX2 (1.06)
chrX 11142876-11143004	HCCS (4.14)
chrX 13079453-13079565	FAM9C (1)
chrX 13321463-13321556	ATXN3L (0)
chrX 18048947-18049048	BEND2 (0)
chrX 18889475-18889602	LOC100132163 (0)
chrX 26928857-26928989	VENTXP1 (0)
chrX 29678080-29678188	IL1RAPL1 (0.08)
chrX 33733184-33733306	DMD (3.58)
chrX 33739875-33739997	DMD (3.58)
chrX 34097584-34097705	FAM47A (0.55)
chrX 34353273-34353395	FAM47A (0.55)
chrX 34587118-34587239	TMEM47 (0.18)
chrX 37287143-37287250	LANCL3 (0.18)
chrX 39293027-39293158	LOC286442 (0)
chrX 39293998-39294127	LOC286442 (0)
chrX 42824389-42824515	PPP1R2P9 (0)
chrX 45418774-45418900	MIR221 (0)
chrX 45899146-45899274	MIR222 (0)
chrX 46031914-46032047	ZNF673 (0)
chrX 46182417-46182540	ZNF673 (0)
chrX 47003999-47004120	RBM10 (4.93)
chrX 68019181-68019310	EFNB1 (4.08)
chrX 69216696-69216826	EDA (0.18)
chrX 74128724-74128853	ABCB7 (3.63)
chrX 77795563-77795676	ZCCHC5 (0)
chrX 78899252-78899364	ITM2A (0.61)
chrX 82925564-82925694	POU3F4 (3.95)
chrX 85270259-85270386	DACH2 (0)
chrX 87041373-87041492	KLHL4 (1.96)
chrX 90994410-90994540	PCDH11X (0.41)
chrX 93990319-93990444	FAM133A (0)
chrX 94539489-94539610	LOC643486 (0)
chrX 95481421-95481543	LOC643486 (0)
chrX 105349133-105349264	MUM1L1 (0.38)
chrX 109756901-109757012	TDGF1P3 (0)
chrX 110100670-110100801	CHRDL1 (0.1)
chrX 110720529-110720646	DKFZp686D0853 (0)
chrX 111587632-111587742	TRPC5 (0.4)
chrX 111782275-111782405	ZCCHC16 (0)
chrX 114659135-114659258	LUZP4 (0.24)
chrX 118875618-118875749	ANKRD58 (0)
chrX 119138356-119138485	NKAP (5.04)
chrX 122891941-122892064	THOC2 (4.9)
chrX 126505785-126505916	CXorf64 (0)
chrX 131887524-131887643	HS6ST2 (0.16)
chrX 133170447-133170569	GPC3 (0)
chrX 135056044-135056180	SLC9A6 (3.15)
chrX 135287948-135288044	MAP7D3 (3.93)
chrX 135530724-135530852	GPR112 (0)
chrX 144625835-144625944	SLITRK2 (0)
chrX 148732463-148732582	TMEM185A (3.8)
chrX 153060097-153060229	SSR4 (7.51)
chrX 153195444-153195638	NAA10 (6.16)
chrX 153969397-153969519	DKC1 (5.67)
chrY 16405321-16405445	NLGN4Y (0)
chrY 17293477-17293595	NLGN4Y (0)

6. REFERENCES

Adelman, K., M. T. Marr, J. Werner, A. Saunders, Z. Ni, E. D. Andrulis and J. T. Lis (2005). "Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS." Mol Cell **17**(1): 103-112.

Agrawal, A., Q. M. Eastman and D. G. Schatz (1998). "Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system." Nature **394**(6695): 744-751.

Ahel, D., Z. Horejsi, N. Wiechens, S. E. Polo, E. Garcia-Wilson, I. Ahel, H. Flynn, M. Skehel, S. C. West, S. P. Jackson, T. Owen-Hughes and S. J. Boulton (2009). "Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1." Science **325**(5945): 1240-1243.

Ahmad, K. and S. Henikoff (2002). "The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly." Mol Cell **9**(6): 1191-1200.

Ahmad, S. F., M. A. Ansari, K. M. Zoheir, S. A. Bakheet, H. M. Korashy, A. Nadeem, A. E. Ashour and S. M. Attia (2015). "Regulation of TNF-alpha and NF-kappaB activation through the JAK/STAT signaling pathway downstream of histamine 4 receptor in a rat model of LPS-induced joint inflammation." Immunobiology **220**(7): 889-898.

Ahmed, M. and P. Liang (2012). "Transposable elements are a significant contributor to tandem repeats in the human genome." Comp Funct Genomics **2012**: 947089.

Aikawa, Y., T. Katsumoto, P. Zhang, H. Shima, M. Shino, K. Terui, E. Ito, H. Ohno, E. R. Stanley, H. Singh, D. G. Tenen and I. Kitabayashi (2010). "PU.1-mediated upregulation of CSF1R is crucial for leukemia stem cell potential induced by MOZ-TIF2." Nat Med **16**(5): 580-585, 581p following 585.

Alakurtti, K., K. Virtaneva, T. Joensuu, J. J. Palvimo and A. E. Lehesjoki (2000). "Characterization of the cystatin B gene promoter harboring the dodecamer repeat expanded in progressive myoclonus epilepsy, EPM1." Gene **242**(1-2): 65-73.

Aldinucci, D., D. Lorenzon, L. Cattaruzza, A. Pinto, A. Gloghini, A. Carbone and A. Colombatti (2008). "Expression of CCR5 receptors on Reed-Sternberg cells and Hodgkin lymphoma cell lines: involvement of CCL5/Rantes in tumor cell growth and microenvironmental interactions." Int J Cancer **122**(4): 769-776.

Allen, C. D., T. Okada and J. G. Cyster (2007). "Germinal-center organization and cellular dynamics." Immunity **27**(2): 190-202.

Altun, G., J. F. Loring and L. C. Laurent (2010). "DNA methylation in embryonic stem cells." J Cell Biochem **109**(1): 1-6.

Anderson, D. M., E. Maraskovsky, W. L. Billingsley, W. C. Dougall, M. E. Tometsko, E. R. Roux, M. C. Teepe, R. F. DuBose, D. Cosman and L. Galibert (1997). "A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function." Nature **390**(6656): 175-179.

Babaian, A. and D. L. Mager (2016). "Endogenous retroviral promoter exaptation in human cancer." Mob DNA **7**: 24.

Babaian, A., M. T. Romanish, L. Gagnier, L. Y. Kuo, M. M. Karimi, C. Steidl and D. L. Mager (2016). "Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma." Oncogene **35**(19): 2542-2546.

Bagowski, C. P., J. Besser, C. R. Frey and J. E. Ferrell, Jr. (2003). "The JNK cascade as a biochemical switch in mammalian cells: ultrasensitive and all-or-none responses." Curr Biol **13**(4): 315-320.

Bai, M., V. Panoulas, A. Papoudou-Bai, N. Horianopoulos, P. Kitsoulis, K. Stefanaki, D. Rontogianni, N. J. Agnantis and P. Kanavaros (2006). "B-cell differentiation

immunophenotypes in classical Hodgkin lymphomas." Leuk Lymphoma **47**(3): 495-501.
Bakker, E. R., A. M. Das, W. Helvensteijn, P. F. Franken, S. Swagemakers, M. A. van der Valk, T. L. ten Hagen, E. J. Kuipers, W. van Veelen and R. Smits (2013). "Wnt5a promotes human colon cancer cell migration and invasion but does not augment intestinal tumorigenesis in Apc1638N mice." Carcinogenesis **34**(11): 2629-2638.
Baniahmad, A., C. Steiner, A. C. Kohne and R. Renkawitz (1990). "Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site." Cell **61**(3): 505-514.
Bannard, O., R. M. Horton, C. D. Allen, J. An, T. Nagasawa and J. G. Cyster (2013). "Germinal center centroblasts transition to a centrocyte phenotype according to a timed program and depend on the dark zone for effective selection." Immunity **39**(5): 912-924.
Bannert, N. and R. Kurth (2006). "The evolutionary dynamics of human endogenous retroviral families." Annu Rev Genomics Hum Genet **7**: 149-173.
Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." Cell Res **21**(3): 381-395.
Barboric, M., R. M. Nissen, S. Kanazawa, N. Jabrane-Ferrat and B. M. Peterlin (2001). "NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II." Mol Cell **8**(2): 327-337.
Barry, T. S., E. S. Jaffe, L. Sorbara, M. Raffeld and S. Pittaluga (2003). "Peripheral T-cell lymphomas expressing CD30 and CD15." Am J Surg Pathol **27**(12): 1513-1522.
Barth, T. K. and A. Imhof (2010). "Fast signals and slow marks: the dynamics of histone modifications." Trends Biochem Sci **35**(11): 618-626.
Basehoar, A. D., S. J. Zanton and B. F. Pugh (2004). "Identification and distinct regulation of yeast TATA box-containing genes." Cell **116**(5): 699-709.
Bauer, A., C. Rodiger, C. Greif, M. Kaatz and P. Elsner (2005). "Vulvar dermatoses--irritant and allergic contact dermatitis of the vulva." Dermatology **210**(2): 143-149.
Bell, A. C., A. G. West and G. Felsenfeld (1999). "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators." Cell **98**(3): 387-396.
Benachenhou, F., P. Jern, M. Oja, G. Sperber, V. Blikstad, P. Somervuo, S. Kaski and J. Blomberg (2009). "Evolutionary conservation of orthoretroviral long terminal repeats (LTRs) and ab initio detection of single LTRs in genomic data." PLoS One **4**(4): e5179.
Benachenhou, F., G. O. Sperber, E. Bongcam-Rudloff, G. Andersson, J. D. Boeke and J. Blomberg (2013). "Conserved structure and inferred evolutionary history of long terminal repeats (LTRs)." Mob DNA **4**(1): 5.
Benesova, M., K. Trejbalova, D. Kovarova, Z. Vernerova, T. Hron, D. Kucerova and J. Hejnar (2017). "DNA hypomethylation and aberrant expression of the human endogenous retrovirus ERVWE1/syncytin-1 in seminomas." Retrovirology **14**(1): 20.
Benit, L., J. B. Lallemand, J. F. Casella, H. Philippe and T. Heidmann (1999). "ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals." J Virol **73**(4): 3301-3308.
Berkhout, B., M. Jebbink and J. Zsiros (1999). "Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus." J Virol **73**(3): 2365-2375.

Bi, S., O. Gavrilova, D. W. Gong, M. M. Mason and M. Reitman (1997). "Identification of a placental enhancer for the human leptin gene." J Biol Chem **272**(48): 30583-30588.
Bindea, G., J. Galon and B. Mlecnik (2013). "CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data." Bioinformatics **29**(5): 661-663.
Bindea, G., B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W. H. Fridman, F. Pages, Z. Trajanoski and J. Galon (2009). "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks." Bioinformatics **25**(8): 1091-1093.
Bloom, K. and A. Joglekar (2010). "Towards building a chromosome segregation machine." Nature **463**(7280): 446-456.
Bohle, V., C. Doring, M. L. Hansmann and R. Kuppers (2013). "Role of early B-cell factor 1 (EBF1) in Hodgkin lymphoma." Leukemia **27**(3): 671-679.
Bohne, A., F. Brunet, D. Galiana-Arnoux, C. Schultheis and J. N. Volff (2008). "Transposable elements as drivers of genomic and biological diversity in vertebrates." Chromosome Res **16**(1): 203-215.
Bonifer, C. and D. A. Hume (2008). "The transcriptional regulation of the Colony-Stimulating Factor 1 Receptor (csf1r) gene during hematopoiesis." Front Biosci **13**: 549-560.
Bortvin, A. and F. Winston (1996). "Evidence that Spt6p controls chromatin structure by a direct interaction with histones." Science **272**(5267): 1473-1476.
Bourque, G., B. Leong, V. B. Vega, X. Chen, Y. L. Lee, K. G. Srinivasan, J. L. Chew, Y. Ruan, C. L. Wei, H. H. Ng and E. T. Liu (2008). "Evolution of the mammalian transcription factor binding repertoire via transposable elements." Genome Res **18**(11): 1752-1762.
Boyle, A. P., S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey and G. E. Crawford (2008). "High-resolution mapping and characterization of open chromatin across the genome." Cell **132**(2): 311-322.
Brasier, A. R. (2006). "The NF-kappaB regulatory network." Cardiovasc Toxicol **6**(2): 111-130.
Brauninger, A., R. Schmitz, D. Bechtel, C. Renne, M. L. Hansmann and R. Kuppers (2006). "Molecular biology of Hodgkin's and Reed/Sternberg cells in Hodgkin's lymphoma." Int J Cancer **118**(8): 1853-1861.
Brocks, D., C. R. Schmidt, M. Daskalakis, H. S. Jang, N. M. Shah, D. Li, J. Li, B. Zhang, Y. Hou, S. Laudato, D. B. Lipka, J. Schott, H. Bierhoff, Y. Assenov, M. Helf, A. Ressnerova, M. S. Islam, A. M. Lindroth, S. Haas, M. Essers, C. D. Imbusch, B. Brors, I. Oehme, O. Witt, M. Lubbert, J. P. Mallm, K. Rippe, R. Will, D. Weichenhan, G. Stoecklin, C. Gerhauser, C. C. Oakes, T. Wang and C. Plass (2017). "DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats." Nat Genet **49**(7): 1052-1060.
Brosius, J. and S. J. Gould (1992). "On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA"." Proc Natl Acad Sci U S A **89**(22): 10706-10710.
Brownell, J. E., J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth and C. D. Allis (1996). "Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation." Cell **84**(6): 843-851.
Bru, D., F. Martin-Laurent and L. Philippot (2008). "Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example." Appl Environ Microbiol **74**(5): 1660-1663.

Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide." Curr Protoc Mol Biol **109**: 21 29 21-29.
Bulger, M. and M. Groudine (2010). "Enhancers: the abundance and function of regulatory sequences beyond promoters." Dev Biol **339**(2): 250-257.
Cabart, P., A. Ujvari, M. Pal and D. S. Luse (2011). "Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes." Proc Natl Acad Sci U S A **108**(38): 15786-15791.
Caldwell, R. G., J. B. Wilson, S. J. Anderson and R. Longnecker (1998). "Epstein-Barr virus LMP2A drives B cell development and survival in the absence of normal B cell receptor signals." Immunity **9**(3): 405-411.
Calo, E. and J. Wysocka (2013). "Modification of enhancer chromatin: what, how, and why?" Mol Cell **49**(5): 825-837.
Campos-Sanchez, R., M. A. Cremona, A. Pini, F. Chiaromonte and K. D. Makova (2016). "Integration and Fixation Preferences of Human and Mouse Endogenous Retroviruses Uncovered with Functional Data Analysis." PLoS Comput Biol **12**(6): e1004956.
Cao, J. (2014). "The functional role of long non-coding RNAs and epigenetics." Biol Proced Online **16**: 11.
Cariappa, A., C. Chase, H. Liu, P. Russell and S. Pillai (2007). "Naive recirculating B cells mature simultaneously in the spleen and bone marrow." Blood **109**(6): 2339-2345.
Carter, D., L. Chakalova, C. S. Osborne, Y. F. Dai and P. Fraser (2002). "Long-range chromatin regulatory interactions in vivo." Nat Genet **32**(4): 623-626.
Cauchy, P., S. R. James, J. Zacarias-Cabeza, A. Ptasinska, M. R. Imperato, S. A. Assi, J. Piper, M. Canestraro, M. Hoogenkamp, M. Raghavan, J. Loke, S. Akiki, S. J. Clokie, S. J. Richards, D. R. Westhead, M. J. Griffiths, S. Ott, C. Bonifer and P. N. Cockerill (2015). "Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature." Cell Rep **12**(5): 821-836.
Cavalier-Smith, T. (1978). "Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox." J Cell Sci **34**: 247-278.
Cerretti, D. P., C. J. Kozlosky, B. Mosley, N. Nelson, K. Van Ness, T. A. Greenstreet, C. J. March, S. R. Kronheim, T. Druck, L. A. Cannizzaro and et al. (1992). "Molecular cloning of the interleukin-1 beta converting enzyme." Science **256**(5053): 97-100.
Cerutti, A., M. Cols and I. Puga (2013). "Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes." Nat Rev Immunol **13**(2): 118-132.
Chang, L. and M. Karin (2001). "Mammalian MAP kinase signalling cascades." Nature **410**(6824): 37-40.
Chavarria-Smith, J. and R. E. Vance (2015). "The NLRP1 inflammasomes." Immunol Rev **265**(1): 22-34.
Chen, H. T. and S. Hahn (2003). "Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex." Mol Cell **12**(2): 437-447.
Chen, L., T. H. Chan, Y. F. Yuan, L. Hu, J. Huang, S. Ma, J. Wang, S. S. Dong, K. H. Tang, D. Xie, Y. Li and X. Y. Guan (2010). "CHD1L promotes hepatocellular carcinoma progression and metastasis in mice and is associated with these processes in human patients." J Clin Invest **120**(4): 1178-1191.

Chen, M., J. D. Huang, L. Hu, B. J. Zheng, L. Chen, S. L. Tsang and X. Y. Guan (2009). "Transgenic CHD1L expression in mouse induces spontaneous tumors." PLoS One **4**(8): e6727.
Cheung, P., K. G. Tanner, W. L. Cheung, P. Sassone-Corsi, J. M. Denu and C. D. Allis (2000). "Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation." Mol Cell **5**(6): 905-915.
Chong, S., N. Vickaryous, A. Ashe, N. Zamudio, N. Youngson, S. Hemley, T. Stopka, A. Skoultchi, J. Matthews, H. S. Scott, D. de Kretser, M. O'Bryan, M. Blewitt and E. Whitelaw (2007). "Modifiers of epigenetic reprogramming show paternal effects in the mouse." Nat Genet **39**(5): 614-622.
Christensen, T. (2005). "Association of human endogenous retroviruses with multiple sclerosis and possible interactions with herpes viruses." Rev Med Virol **15**(3): 179-211.
Christensen, T., P. Dissing Sorensen, H. Riemann, H. J. Hansen and A. Moller-Larsen (1998). "Expression of sequence variants of endogenous retrovirus RGH in particle form in multiple sclerosis." Lancet **352**(9133): 1033.
Christensen, T., P. Dissing Sorensen, H. Riemann, H. J. Hansen, M. Munch, S. Haahr and A. Moller-Larsen (2000). "Molecular characterization of HERV-H variants associated with multiple sclerosis." Acta Neurol Scand **101**(4): 229-238.
Clerici, M., M. L. Fusi, D. Caputo, F. R. Guerini, D. Trabattoni, A. Salvaggio, C. L. Cazzullo, D. Arienti, M. L. Villa, H. B. Urnovitz and P. Ferrante (1999). "Immune responses to antigens of human endogenous retroviruses in patients with acute or stable multiple sclerosis." J Neuroimmunol **99**(2): 173-182.
Cockerill, P. N. (2011). "Structure and function of active chromatin and DNase I hypersensitive sites." FEBS J **278**(13): 2182-2210.
Cockerill, P. N., A. G. Bert, D. Roberts and M. A. Vadas (1999). "The human granulocyte-macrophage colony-stimulating factor gene is autonomously regulated in vivo by an inducible tissue-specific enhancer." Proc Natl Acad Sci U S A **96**(26): 15097-15102.
Cohen, C. J., W. M. Lock and D. L. Mager (2009). "Endogenous retroviral LTRs as promoters for human genes: a critical assessment." Gene **448**(2): 105-114.
Comings, D. E. (1972). "The structure and function of chromatin." Adv Hum Genet **3**: 237-431.
Conley, A. and M. Hinshelwood (2001). "Mammalian aromatases." Reproduction **121**(5): 685-695.
Connelly, S. and J. L. Manley (1988). "A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II." Genes Dev **2**(4): 440-452.
Connors, J. M. (2009). "Clinical manifestations and natural history of Hodgkin's lymphoma." Cancer J **15**(2): 124-128.
Contreras-Galindo, R., M. H. Kaplan, D. Dube, M. J. Gonzalez-Hernandez, S. Chan, F. Meng, M. Dai, G. S. Omenn, S. D. Gitlin and D. M. Markovitz (2015). "Human Endogenous Retrovirus Type K (HERV-K) Particles Package and Transmit HERV-K-Related Sequences." J Virol **89**(14): 7187-7201.
Corces, M. R., J. D. Buenrostro, B. Wu, P. G. Greenside, S. M. Chan, J. L. Koenig, M. P. Snyder, J. K. Pritchard, A. Kundaje, W. J. Greenleaf, R. Majeti and H. Y. Chang (2016). "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution." Nat Genet **48**(10): 1193-1203.
Cordaux, R. and M. A. Batzer (2009). "The impact of retrotransposons on human genome evolution." Nat Rev Genet **10**(10): 691-703.

Cordonnier, A., J. F. Casella and T. Heidmann (1995). "Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence." J Virol **69**(9): 5890-5897.
CRUK. (2012). "Hodgkin Lymphoma Survival Statistics." from http://www.cancerresearchuk.org/cancer-info/cancerstats/types/hodgkinslymphoma/survival/.
CRUK. (2013). "Hodgkin's Lymphoma Incidence Statistics." from http://www.cancerresearchuk.org/cancer-info/cancerstats/types/hodgkinslymphoma/incidence/.
Dailey, L. (2015). "High throughput technologies for the functional discovery of mammalian enhancers: new approaches for understanding transcriptional regulatory network dynamics." Genomics **106**(3): 151-158.
Dale, R. K., B. S. Pedersen and A. R. Quinlan (2011). "Pybedtools: a flexible Python library for manipulating genomic datasets and annotations." Bioinformatics **27**(24): 3423-3424.
Dao, L. T. M., A. O. Galindo-Albarran, J. A. Castro-Mondragon, C. Andrieu-Soler, A. Medina-Rivera, C. Souaid, G. Charbonnier, A. Griffon, L. Vanhille, T. Stephen, J. Alomairi, D. Martin, M. Torres, N. Fernandez, E. Soler, J. van Helden, D. Puthier and S. Spicuglia (2017). "Genome-wide characterization of mammalian promoters with distal enhancer functions." Nat Genet **49**(7): 1073-1081.
Darnay, B. G., J. Ni, P. A. Moore and B. B. Aggarwal (1999). "Activation of NF-kappaB by RANK requires tumor necrosis factor receptor-associated factor (TRAF) 6 and NF-kappaB-inducing kinase. Identification of a novel TRAF6 interaction motif." J Biol Chem **274**(12): 7724-7731.
Day, D. S., L. J. Luquette, P. J. Park and P. V. Kharchenko (2010). "Estimating enrichment of repetitive elements from high-throughput sequence data." Genome Biol **11**(6): R69.
de Jong, J., L. F. Wessels, M. van Lohuizen, J. de Ridder and W. Akhtar (2014). "Applications of DNA integrating elements: Facing the bias bully." Mob Genet Elements **4**(6): 1-6.
de Koning, A. P., W. Gu, T. A. Castoe, M. A. Batzer and D. D. Pollock (2011). "Repetitive elements may comprise over two-thirds of the human genome." PLoS Genet **7**(12): e1002384.
de Leval, L. and P. Gaulard (2010). "CD30+ lymphoproliferative disorders." Haematologica **95**(10): 1627-1630.
de Oliveira, K. A., E. Kaergel, M. Heinig, J. F. Fontaine, G. Patone, E. M. Muro, S. Mathas, M. Hummel, M. A. Andrade-Navarro, N. Hubner and C. Scheidereit (2016). "A roadmap of constitutive NF-kappaB activity in Hodgkin lymphoma: Dominant roles of p50 and p52 revealed by genome-wide analyses." Genome Med **8**(1): 28.
de Villiers, J. and W. Schaffner (1981). "A small segment of polyoma virus DNA enhances the expression of a cloned beta-globin gene over a distance of 1400 base pairs." Nucleic Acids Res **9**(23): 6251-6264.
Deaton, A. M. and A. Bird (2011). "CpG islands and the regulation of transcription." Genes Dev **25**(10): 1010-1022.
Deininger, P. (2011). "Alu elements: know the SINEs." Genome Biol **12**(12): 236.
Deininger, P. L. and M. A. Batzer (1999). "Alu repeats and human disease." Mol Genet Metab **67**(3): 183-193.
Deininger, P. L. and M. A. Batzer (2002). "Mammalian retroelements." Genome Res **12**(10): 1455-1465.

Dekker, J., K. Rippe, M. Dekker and N. Kleckner (2002). "Capturing chromosome conformation." Science **295**(5558): 1306-1311.
DeKoter, R. P. and H. Singh (2000). "Regulation of B lymphocyte and macrophage development by graded expression of PU.1." Science **288**(5470): 1439-1441.
Delhase, M., M. Hayakawa, Y. Chen and M. Karin (1999). "Positive and negative regulation of IkappaB kinase activity through IKKbeta subunit phosphorylation." Science **284**(5412): 309-313.
Denes, A., G. Lopez-Castejon and D. Brough (2012). "Caspase-1: is IL-1 just the tip of the ICEberg?" Cell Death Dis **3**: e338.
Deng, C., B. Patel, X. Lin, Y. Li and S. Huang (2015). Chapter 5 - Chromatin dynamics and genome organization in development and disease. Epigenetic Gene Expression and Regulation. Oxford, Academic Press**:** 95-115.
Derenzini, E. and A. Younes (2011). "Predicting treatment outcome in classical Hodgkin lymphoma: genomic advances." Genome Med **3**(4): 26.
Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu and B. Ren (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions." Nature **485**(7398): 376-380.
Dong, J., E. Jimi, C. Zeiss, M. S. Hayden and S. Ghosh (2010). "Constitutively active NF-kappaB triggers systemic TNFalpha-dependent inflammation and localized TNFalpha-independent inflammatory disease." Genes Dev **24**(16): 1709-1717.
Droin, N. and E. Solary (2010). "Editorial: CSF1R, CSF-1, and IL-34, a "menage a trois" conserved across vertebrates." J Leukoc Biol **87**(5): 745-747.
Dudley, D. D., J. Chaudhuri, C. H. Bassing and F. W. Alt (2005). "Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences." Adv Immunol **86**: 43-112.
Duncan, L., K. Webster, V. Gupta, S. Nair and E. Deane (2010). "Molecular characterisation of the CD79a and CD79b subunits of the B cell receptor complex in the gray short-tailed opossum (Monodelphis domestica) and tammar wallaby (Macropus eugenii): Delayed B cell immunocompetence in marsupial neonates." Vet Immunol Immunopathol **136**(3-4): 235-247.
Dunn, C. A., P. Medstrand and D. L. Mager (2003). "An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon." Proc Natl Acad Sci U S A **100**(22): 12841-12846.
Dunn, C. A., M. T. Romanish, L. E. Gutierrez, L. N. van de Lagemaat and D. L. Mager (2006). "Transcription of two human genes from a bidirectional endogenous retrovirus promoter." Gene **366**(2): 335-342.
Dynan, W. S. and R. Tjian (1983). "The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter." Cell **35**(1): 79-87.
Eden, A., F. Gaudet, A. Waghmare and R. Jaenisch (2003). "Chromosomal instability and tumors promoted by DNA hypomethylation." Science **300**(5618): 455.
Ehrlich, M. (2009). "DNA hypomethylation in cancer cells." Epigenomics **1**(2): 239-259.
Elsasser, S. J., K. M. Noh, N. Diaz, C. D. Allis and L. A. Banaszynski (2015). "Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells." Nature **522**(7555): 240-244.
Elsasser, S. J., K. M. Noh, N. Diaz, C. D. Allis and L. A. Banaszynski (2017). "Elsasser et al. reply." Nature **548**(7665): E7-E9.
Fasching, L., A. Kapopoulou, R. Sachdeva, R. Petri, M. E. Jonsson, C. Manne, P. Turelli, P. Jern, F. Cammas, D. Trono and J. Jakobsson (2015). "TRIM28 represses

transcription of endogenous retroviruses in neural progenitor cells." Cell Rep **10**(1): 20-28.
Faulkner, G. J., Y. Kimura, C. O. Daub, S. Wani, C. Plessy, K. M. Irvine, K. Schroder, N. Cloonan, A. L. Steptoe, T. Lassmann, K. Waki, N. Hornig, T. Arakawa, H. Takahashi, J. Kawai, A. R. Forrest, H. Suzuki, Y. Hayashizaki, D. A. Hume, V. Orlando, S. M. Grimmond and P. Carninci (2009). "The regulated retrotransposon transcriptome of mammalian cells." Nat Genet **41**(5): 563-571.
Feschotte, C. (2008). "Transposable elements and the evolution of regulatory networks." Nat Rev Genet **9**(5): 397-405.
Fischer, M., M. Juremalm, N. Olsson, C. Backlin, C. Sundstrom, K. Nilsson, G. Enblad and G. Nilsson (2003). "Expression of CCL5/RANTES by Hodgkin and Reed-Sternberg cells and its possible role in the recruitment of mast cells into lymphomatous tissue." Int J Cancer **107**(2): 197-201.
Fischer, M. G. and C. A. Suttle (2011). "A virophage at the origin of large DNA transposons." Science **332**(6026): 231-234.
Fiumara, P., V. Snell, Y. Li, A. Mukhopadhyay, M. Younes, A. M. Gillenwater, F. Cabanillas, B. B. Aggarwal and A. Younes (2001). "Functional expression of receptor activator of nuclear factor kappaB in Hodgkin disease cell lines." Blood **98**(9): 2784-2790.
Flemming, W. (1882). Zellsubstanz, Kern und Zelltheilung ... Mit ... Tafeln, Leipzig.
Florl, A. R., C. Steinhoff, M. Muller, H. H. Seifert, C. Hader, R. Engers, R. Ackermann and W. A. Schulz (2004). "Coordinate hypermethylation at specific genes in prostate carcinoma precedes LINE-1 hypomethylation." Br J Cancer **91**(5): 985-994.
Fong, N., T. Saldi, R. M. Sheridan, M. A. Cortazar and D. L. Bentley (2017). "RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction." Mol Cell **66**(4): 546-557 e543.
Fort, A., K. Hashimoto, D. Yamada, M. Salimullah, C. A. Keya, A. Saxena, A. Bonetti, I. Voineagu, N. Bertin, A. Kratz, Y. Noro, C. H. Wong, M. de Hoon, R. Andersson, A. Sandelin, H. Suzuki, C. L. Wei, H. Koseki, F. Consortium, Y. Hasegawa, A. R. Forrest and P. Carninci (2014). "Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance." Nat Genet **46**(6): 558-566.
Frank, O., M. Giehl, C. Zheng, R. Hehlmann, C. Leib-Mosch and W. Seifarth (2005). "Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders." J Virol **79**(17): 10890-10901.
Freimanis, G., P. Hooley, H. D. Ejtehadi, H. A. Ali, A. Veitch, P. B. Rylance, A. Alawi, J. Axford, A. Nevill, P. G. Murray and P. N. Nelson (2010). "A role for human endogenous retrovirus-K (HML-2) in rheumatoid arthritis: investigating mechanisms of pathogenesis." Clin Exp Immunol **160**(3): 340-347.
Frohman, M. A., M. K. Dush and G. R. Martin (1988). "Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer." Proc Natl Acad Sci U S A **85**(23): 8998-9002.
Fuchs, N. V., M. Kraft, C. Tondera, K. M. Hanschmann, J. Lower and R. Lower (2011). "Expression of the human endogenous retrovirus (HERV) group HML-2/HERV-K does not depend on canonical promoter elements but is regulated by transcription factors Sp1 and Sp3." J Virol **85**(7): 3436-3448.
Fuke, C., M. Shimabukuro, A. Petronis, J. Sugimoto, T. Oda, K. Miura, T. Miyazaki, C. Ogura, Y. Okazaki and Y. Jinno (2004). "Age related changes in 5-methylcytosine

content in human peripheral leukocytes and placentas: an HPLC-based study." Ann Hum Genet **68**(Pt 3): 196-204.
Gama-Sosa, M. A., V. A. Slagel, R. W. Trewyn, R. Oxenhandler, K. C. Kuo, C. W. Gehrke and M. Ehrlich (1983). "The 5-methylcytosine content of DNA from human tumors." Nucleic Acids Res **11**(19): 6883-6894.
Gaszner, M. and G. Felsenfeld (2006). "Insulators: exploiting transcriptional and epigenetic mechanisms." Nat Rev Genet **7**(9): 703-713.
Geisberger, R., M. Lamers and G. Achatz (2006). "The riddle of the dual expression of IgM and IgD." Immunology **118**(4): 429-437.
Ghayur, T., S. Banerjee, M. Hugunin, D. Butler, L. Herzog, A. Carter, L. Quintal, L. Sekut, R. Talanian, M. Paskind, W. Wong, R. Kamen, D. Tracey and H. Allen (1997). "Caspase-1 processes IFN-gamma-inducing factor and regulates LPS-induced IFN-gamma production." Nature **386**(6625): 619-623.
Gil, V. S., G. Bhagat, L. Howell, J. Zhang, C. H. Kim, S. Stengel, F. Vega, A. Zelent and K. Petrie (2016). "Deregulated expression of HDAC9 in B cells promotes development of lymphoproliferative disease and lymphoma in mice." Dis Model Mech **9**(12): 1483-1495.
Gilbert, W. (1978). "Why genes in pieces?" Nature **271**(5645): 501.
Glinsky, G. V. (2015). "Transposable Elements and DNA Methylation Create in Embryonic Stem Cells Human-Specific Regulatory Sequences Associated with Distal Enhancers and Noncoding RNAs." Genome Biol Evol **7**(6): 1432-1454.
Goering, W., T. Ribarska and W. A. Schulz (2011). "Selective changes of retroelement expression in human prostate cancer." Carcinogenesis **32**(10): 1484-1492.
Golan, M., A. Hizi, J. H. Resau, N. Yaal-Hahoshen, H. Reichman, I. Keydar and I. Tsarfaty (2008). "Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker." Neoplasia **10**(6): 521-533.
Gomez, N. C., A. J. Hepperla, R. Dumitru, J. M. Simon, F. Fang and I. J. Davis (2016). "Widespread Chromatin Accessibility at Repetitive Elements Links Stem Cells with Human Cancer." Cell Rep **17**(6): 1607-1620.
Gould, S. J. and E. S. Vrba (1982). "Exaptation-A Missing Term in the Science of Form." Paleobiology **8**(1): 4-15.
Gregory, T. R. (2005). "Synergy between sequence and size in large-scale genomics." Nat Rev Genet **6**(9): 699-708.
Grunberg, S., L. Warfield and S. Hahn (2012). "Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening." Nat Struct Mol Biol **19**(8): 788-796.
Gruss, H. J., D. Hirschstein, B. Wright, D. Ulrich, M. A. Caligiuri, M. Barcos, L. Strockbine, R. J. Armitage and S. K. Dower (1994). "Expression and function of CD40 on Hodgkin and Reed-Sternberg cells and the possible relevance for Hodgkin's disease." Blood **84**(7): 2305-2314.
Guo, D., J. D. Dunbar, C. H. Yang, L. M. Pfeffer and D. B. Donner (1998). "Induction of Jak/STAT signaling by activation of the type 1 TNF receptor." J Immunol **160**(6): 2742-2750.
Guzman-Rojas, L., J. C. Sims-Mourtada, R. Rangel and H. Martinez-Valdez (2002). "Life and death within germinal centres: a double-edged sword." Immunology **107**(2): 167-175.
Han, Z., D. L. Boyle, A. M. Manning and G. S. Firestein (1998). "AP-1 and NF-kappaB regulation in rheumatoid arthritis and murine collagen-induced arthritis." Autoimmunity **28**(4): 197-208.

Hashimoto, K., A. M. Suzuki, A. Dos Santos, C. Desterke, A. Collino, S. Ghisletti, E. Braun, A. Bonetti, A. Fort, X. Y. Qin, E. Radaelli, B. Kaczkowski, A. R. Forrest, S. Kojima, D. Samuel, G. Natoli, M. A. Buendia, J. Faivre and P. Carninci (2015). "CAGE profiling of ncRNAs in hepatocellular carcinoma reveals widespread activation of retroviral LTR promoters in virus-induced tumors." Genome Res **25**(12): 1812-1824.
He, H. H., C. A. Meyer, H. Shin, S. T. Bailey, G. Wei, Q. Wang, Y. Zhang, K. Xu, M. Ni, M. Lupien, P. Mieczkowski, J. D. Lieb, K. Zhao, M. Brown and X. S. Liu (2010). "Nucleosome dynamics define transcriptional enhancers." Nat Genet **42**(4): 343-347.
Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.
Heinz, S., C. E. Romanoski, C. Benner and C. K. Glass (2015). "The selection and function of cell type-specific enhancers." Nat Rev Mol Cell Biol **16**(3): 144-154.
Herblot, S., P. D. Aplan and T. Hoang (2002). "Gradient of E2A activity in B-cell development." Mol Cell Biol **22**(3): 886-900.
Hess, J., P. Angel and M. Schorpp-Kistner (2004). "AP-1 subunits: quarrel and harmony among siblings." J Cell Sci **117**(Pt 25): 5965-5973.
Himes, S. R., H. Tagoh, N. Goonetilleke, T. Sasmono, D. Oceandy, R. Clark, C. Bonifer and D. A. Hume (2001). "A highly conserved c-fms gene intronic element controls macrophage-specific and regulated expression." J Leukoc Biol **70**(5): 812-820.
Hinz, M., P. Lemke, I. Anagnostopoulos, C. Hacker, D. Krappmann, S. Mathas, B. Dorken, M. Zenke, H. Stein and C. Scheidereit (2002). "Nuclear factor kappaB-dependent gene expression profiling of Hodgkin's disease tumor cells, pathogenetic significance, and link to constitutive signal transducer and activator of transcription 5a activity." J Exp Med **196**(5): 605-617.
Hirose, Y., M. Takamatsu and F. Harada (1993). "Presence of env genes in members of the RTVL-H family of human endogenous retrovirus-like elements." Virology **192**(1): 52-61.
Hodgkin (1832). "On some Morbid Appearances of the Absorbent Glands and Spleen." Med Chir Trans **17**: 68-114.
Hondele, M., T. Stuwe, M. Hassler, F. Halbach, A. Bowman, E. T. Zhang, B. Nijmeijer, C. Kotthoff, V. Rybin, S. Amlacher, E. Hurt and A. G. Ladurner (2013). "Structural basis of histone H2A-H2B recognition by the essential chaperone FACT." Nature **499**(7456): 111-114.
Howard, G., R. Eiges, F. Gaudet, R. Jaenisch and A. Eden (2008). "Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice." Oncogene **27**(3): 404-408.
Hu, T., W. Pi, X. Zhu, M. Yu, H. Ha, H. Shi, J. H. Choi and D. Tuan (2017). "Long non-coding RNAs transcribed by ERV-9 LTR retrotransposon act in cis to modulate long-range LTR enhancer function." Nucleic Acids Res **45**(8): 4479-4492.
Huda, A., L. Marino-Ramirez and I. K. Jordan (2010). "Epigenetic histone modifications of human transposable elements: genome defense versus exaptation." Mob DNA **1**(1): 2.
Hug, B. A. and M. A. Lazar (2004). "ETO interacting proteins." Oncogene **23**(24): 4270-4274.

Huh, J. W., D. S. Kim, D. W. Kang, H. S. Ha, K. Ahn, Y. N. Noh, D. S. Min, K. T. Chang and H. S. Kim (2008). "Transcriptional regulation of GSDML gene by antisense-oriented HERV-H LTR element." Arch Virol **153**(6): 1201-1205.
Hurst, T. P. and G. Magiorkinis (2017). "Epigenetic Control of Human Endogenous Retrovirus Expression: Focus on Regulation of Long-Terminal Repeats (LTRs)." Viruses **9**(6).
ICTV, I. C. o. T. o. V. (2017). Virus taxonomy: classification and nomenclature of viruses: 10th report of the International Committee on Taxonomy of Viruses. San Diego.
Ide, H., D. B. Seligson, S. Memarzadeh, L. Xin, S. Horvath, P. Dubey, M. B. Flick, B. M. Kacinski, A. Palotie and O. N. Witte (2002). "Expression of colony-stimulating factor 1 receptor during prostate development and prostate cancer progression." Proc Natl Acad Sci U S A **99**(22): 14404-14409.
Illingworth, R. S. and A. P. Bird (2009). "CpG islands--'a rough guide'." FEBS Lett **583**(11): 1713-1720.
Imbeault, M., P. Y. Helleboid and D. Trono (2017). "KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks." Nature **543**(7646): 550-554.
Ingram, R. M., S. Valeaux, N. Wilson, M. A. Bouhlel, D. Clarke, I. Kruger, D. Kulu, G. Suske, S. Philipsen, H. Tagoh and C. Bonifer (2011). "Differential regulation of sense and antisense promoter activity at the Csf1R locus in B cells by the transcription factor PAX5." Exp Hematol **39**(7): 730-740 e731-732.
Izban, K. F., M. Ergin, Q. Huang, J. Z. Qin, R. L. Martinez, B. Schnitzer, H. Ni, B. J. Nickoloff and S. Alkan (2001). "Characterization of NF-kappaB expression in Hodgkin's disease: inhibition of constitutively expressed NF-kappaB results in spontaneous caspase-independent apoptosis in Hodgkin and Reed-Sternberg cells." Mod Pathol **14**(4): 297-310.
Izban, M. G. and D. S. Luse (1991). "Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing." Genes Dev **5**(4): 683-696.
Johnston, J. B., C. Silva, J. Holden, K. G. Warren, A. W. Clark and C. Power (2001). "Monocyte activation and differentiation augment human endogenous retrovirus expression: implications for inflammatory brain diseases." Ann Neurol **50**(4): 434-442.
Jolivet-Reynaud, C., H. Perron, P. Ferrante, L. Becquart, P. Dalbon and B. Mandrand (1999). "Specificities of multiple sclerosis cerebrospinal fluid and serum antibodies against mimotopes." Clin Immunol **93**(3): 283-293.
Jones, F. S. and R. Meech (1999). "Knockout of REST/NRSF shows that the protein is a potent repressor of neuronally expressed genes in non-neural tissues." Bioessays **21**(5): 372-376.
Jost, P. J. and J. Ruland (2007). "Aberrant NF-kappaB signaling in lymphoma: mechanisms, consequences, and therapeutic implications." Blood **109**(7): 2700-2707.
Jundt, F., K. Kley, I. Anagnostopoulos, K. Schulze Probsting, A. Greiner, S. Mathas, C. Scheidereit, T. Wirth, H. Stein and B. Dorken (2002). "Loss of PU.1 expression is associated with defective immunoglobulin transcription in Hodgkin and Reed-Sternberg cells of classical Hodgkin disease." Blood **99**(8): 3060-3062.
Jurka, J. (2000). "Repbase update: a database and an electronic journal of repetitive elements." Trends Genet **16**(9): 418-420.

Juven-Gershon, T., J. Y. Hsu, J. W. Theisen and J. T. Kadonaga (2008). "The RNA polymerase II core promoter - the gateway to transcription." Curr Opin Cell Biol **20**(3): 253-259.
Kadonaga, J. T. (2004). "Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors." Cell **116**(2): 247-257.
Kamesaki, H., S. Fukuhara, E. Tatsumi, H. Uchino, H. Yamabe, H. Miwa, S. Shirakawa, M. Hatanaka and T. Honjo (1986). "Cytochemical, immunologic, chromosomal, and molecular genetic analysis of a novel cell line derived from Hodgkin's disease." Blood **68**(1): 285-292.
Kanzler, H., M. L. Hansmann, U. Kapp, J. Wolf, V. Diehl, K. Rajewsky and R. Kuppers (1996). "Molecular single cell analysis demonstrates the derivation of a peripheral blood-derived cell line (L1236) from the Hodgkin/Reed-Sternberg cells of a Hodgkin's lymphoma patient." Blood **87**(8): 3429-3436.
Kanzler, H., R. Kuppers, M. L. Hansmann and K. Rajewsky (1996). "Hodgkin and Reed-Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells." J Exp Med **184**(4): 1495-1505.
Kapatai, G. and P. Murray (2007). "Contribution of the Epstein Barr virus to the molecular pathogenesis of Hodgkin lymphoma." J Clin Pathol **60**(12): 1342-1349.
Kapp, U., W. C. Yeh, B. Patterson, A. J. Elia, D. Kagi, A. Ho, A. Hessel, M. Tipsword, A. Williams, C. Mirtsos, A. Itie, M. Moyle and T. W. Mak (1999). "Interleukin 13 is secreted by and stimulates the growth of Hodgkin and Reed-Sternberg cells." J Exp Med **189**(12): 1939-1946.
Karin, M. and Y. Ben-Neriah (2000). "Phosphorylation meets ubiquitination: the control of NF-[kappa]B activity." Annu Rev Immunol **18**: 621-663.
Katoh, I., A. Mirova, S. Kurata, Y. Murakami, K. Horikawa, N. Nakakuki, T. Sakai, K. Hashimoto, A. Maruyama, T. Yonaga, N. Fukunishi, K. Moriishi and H. Hirai (2011). "Activation of the long terminal repeat of human endogenous retrovirus K by melanoma-specific transcription factor MITF-M." Neoplasia **13**(11): 1081-1092.
Katsumata, K., H. Ikeda, M. Sato, A. Ishizu, Y. Kawarada, H. Kato, A. Wakisaka, T. Koike and T. Yoshiki (1999). "Cytokine regulation of env gene expression of human endogenous retrovirus-R in human vascular endothelial cells." Clin Immunol **93**(1): 75-80.
Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips and S. E. Antonarakis (1988). "Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man." Nature **332**(6160): 164-166.
Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." Genome Res **12**(4): 656-664.
Kikuchi, A., H. Yamamoto, A. Sato and S. Matsumoto (2012). "Wnt5a: its signalling, functions and implication in diseases." Acta Physiol (Oxf) **204**(1): 17-33.
Kilger, E., A. Kieser, M. Baumann and W. Hammerschmidt (1998). "Epstein-Barr virus-mediated B-cell proliferation is dependent upon latent membrane protein 1, which simulates an activated CD40 receptor." EMBO J **17**(6): 1700-1709.
Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley and S. L. Salzberg (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biol **14**(4): R36.
Kirkham, C. M., Scott, J.N., Boyes, J., Bevington, S. (2014). The Molecular Basis of B Cell Development and the Role of Deregulated Transcription and Epigenetics in Leukaemia and Lymphoma. Transcriptional and Epigenetic Mechanisms Regulating

Normal and Aberrant Blood Cell Development. C. C. Bonifer, P.N. Berlin, Springer: 313-366.
Klein, G., L. Dombos and B. Gothoskar (1972). "Sensitivity of Epstein-Barr virus (EBV) producer and non-producer human lymphoblastoid cell lines to superinfection with EB-virus." Int J Cancer **10**(1): 44-57.
Klein, U. and R. Dalla-Favera (2008). "Germinal centres: role in B-cell physiology and malignancy." Nat Rev Immunol **8**(1): 22-33.
Klemm, F., A. Bleckmann, L. Siam, H. N. Chuang, E. Rietkotter, D. Behme, M. Schulz, M. Schaffrinski, S. Schindler, L. Trumper, F. Kramer, T. Beissbarth, C. Stadelmann, C. Binder and T. Pukrop (2011). "beta-catenin-independent WNT signaling in basal-like breast cancer and brain metastasis." Carcinogenesis **32**(3): 434-442.
Knossl, M., R. Lower and J. Lower (1999). "Expression of the human endogenous retrovirus HTDV/HERV-K is enhanced by cellular transcription factor YY1." J Virol **73**(2): 1254-1261.
Kolovos, P., T. A. Knoch, F. G. Grosveld, P. R. Cook and A. Papantonis (2012). "Enhancers and silencers: an integrated and simple model for their function." Epigenetics Chromatin **5**(1): 1.
Kornberg, R. D. (1974). "Chromatin structure: a repeating unit of histones and DNA." Science **184**(4139): 868-871.
Kornberg, R. D. and Y. Lorch (1999). "Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome." Cell **98**(3): 285-294.
Kostrewa, D., M. E. Zeller, K. J. Armache, M. Seizl, K. Leike, M. Thomm and P. Cramer (2009). "RNA polymerase II-TFIIB structure and mechanism of transcription initiation." Nature **462**(7271): 323-330.
Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.
Kovalskaya, E., A. Buzdin, E. Gogvadze, T. Vinogradova and E. Sverdlov (2006). "Functional human endogenous retroviral LTR transcription start sites are located between the R and U5 regions." Virology **346**(2): 373-378.
Kreher, S., M. A. Bouhlel, P. Cauchy, B. Lamprecht, S. Li, M. Grau, F. Hummel, K. Kochert, I. Anagnostopoulos, K. Johrens, M. Hummel, J. Hiscott, S. S. Wenzel, P. Lenz, M. Schneider, R. Kuppers, C. Scheidereit, M. Giefing, R. Siebert, K. Rajewsky, G. Lenz, P. N. Cockerill, M. Janz, B. Dorken, C. Bonifer and S. Mathas (2014). "Mapping of transcription factor motifs in active chromatin identifies IRF5 as key regulator in classical Hodgkin lymphoma." Proc Natl Acad Sci U S A **111**(42): E4513-4522.
Kriegs, J. O., G. Churakov, J. Jurka, J. Brosius and J. Schmitz (2007). "Evolutionary history of 7SL RNA-derived SINEs in Supraprimates." Trends Genet **23**(4): 158-161.
Krol, J., A. Fiszer, A. Mykowska, K. Sobczak, M. de Mezer and W. J. Krzyzosiak (2007). "Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets." Mol Cell **25**(4): 575-586.
Kuppers, R. (2009). "The biology of Hodgkin's lymphoma." Nat Rev Cancer **9**(1): 15-27.
Kuppers, R. (2009). "Molecular biology of Hodgkin lymphoma." Hematology Am Soc Hematol Educ Program: 491-496.
Kuppers, R., A. Engert and M. L. Hansmann (2012). "Hodgkin lymphoma." J Clin Invest **122**(10): 3439-3447.
Kuppers, R. and M. L. Hansmann (2005). "The Hodgkin and Reed/Sternberg cell." Int J Biochem Cell Biol **37**(3): 511-517.

Kuppers, R., U. Klein, I. Schwering, V. Distler, A. Brauninger, G. Cattoretti, Y. Tu, G. A. Stolovitzky, A. Califano, M. L. Hansmann and R. Dalla-Favera (2003). "Identification of Hodgkin and Reed-Sternberg cell-specific genes by gene expression profiling." J Clin Invest **111**(4): 529-537.
Kurayoshi, M., N. Oue, H. Yamamoto, M. Kishida, A. Inoue, T. Asahara, W. Yasui and A. Kikuchi (2006). "Expression of Wnt-5a is correlated with aggressiveness of gastric cancer by stimulating cell migration and invasion." Cancer Res **66**(21): 10439-10448.
Kurzrock, R., J. Redman, F. Cabanillas, D. Jones, J. Rothberg and M. Talpaz (1993). "Serum interleukin 6 levels are elevated in lymphoma patients and correlate with survival in advanced Hodgkin's disease and with B symptoms." Cancer Res **53**(9): 2118-2122.
Kwak, H. and J. T. Lis (2013). "Control of transcriptional elongation." Annu Rev Genet **47**: 483-508.
Lagarde, J., B. Uszczynska-Ratajczak, J. Santoyo-Lopez, J. M. Gonzalez, E. Tapanari, J. M. Mudge, C. A. Steward, L. Wilming, A. Tanzer, C. Howald, J. Chrast, A. Vela-Boza, A. Rueda, F. J. Lopez-Domingo, J. Dopazo, A. Reymond, R. Guigo and J. Harrow (2016). "Extension of human lncRNA transcripts by RACE coupled with long-read high-throughput sequencing (RACE-Seq)." Nat Commun **7**: 12339.
Lamprecht, B., K. Walter, S. Kreher, R. Kumar, M. Hummel, D. Lenze, K. Kochert, M. A. Bouhlel, J. Richter, E. Soler, R. Stadhouders, K. Johrens, K. D. Wurster, D. F. Callen, M. F. Harte, M. Giefing, R. Barlow, H. Stein, I. Anagnostopoulos, M. Janz, P. N. Cockerill, R. Siebert, B. Dorken, C. Bonifer and S. Mathas (2010). "Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma." Nat Med **16**(5): 571-579, 571p following 579.
Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki,

S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki and C. International Human Genome Sequencing (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Landry, J. R., A. Rouhi, P. Medstrand and D. L. Mager (2002). "The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter." Mol Biol Evol **19**(11): 1934-1942.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

Larsson, E., P. J. Venables, A. C. Andersson, W. Fan, S. Rigby, J. Botling, F. Oberg, M. Cohen and K. Nilsson (1996). "Expression of the endogenous retrovirus ERV3 (HERV-R) during induced monocytic differentiation in the U-937 cell line." Int J Cancer **67**(3): 451-456.

Laska, M. J., T. Brudek, K. K. Nissen, T. Christensen, A. Moller-Larsen, T. Petersen and B. A. Nexo (2012). "Expression of HERV-Fc1, a human endogenous retrovirus, is increased in patients with active multiple sclerosis." J Virol **86**(7): 3713-3722.

Lavie, L., M. Kitova, E. Maldener, E. Meese and J. Mayer (2005). "CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2)." J Virol **79**(2): 876-883.

Lavie, L., P. Medstrand, W. Schempp, E. Meese and J. Mayer (2004). "Human endogenous retrovirus family HERV-K(HML-5): status, evolution, and reconstruction of an ancient betaretrovirus in the human genome." J Virol **78**(16): 8788-8798.

Lawrence, M., S. Daujat and R. Schneider (2016). "Lateral Thinking: How Histone Modifications Regulate Gene Expression." Trends Genet **32**(1): 42-56.

Lawrence, M. G., C. R. Stephens, E. F. Need, J. Lai, G. Buchanan and J. A. Clements (2012). "Long terminal repeats act as androgen-responsive enhancers for the PSA-kallikrein locus." Endocrinology **153**(7): 3199-3210.

Lawrence, T. (2009). "The nuclear factor NF-kappaB pathway in inflammation." Cold Spring Harb Perspect Biol **1**(6): a001651.

Leddin, M., C. Perrod, M. Hoogenkamp, S. Ghani, S. Assi, S. Heinz, N. K. Wilson, G. Follows, J. Schonheit, L. Vockentanz, A. M. Mosammam, W. Chen, D. G. Tenen, D. R. Westhead, B. Gottgens, C. Bonifer and F. Rosenbauer (2011). "Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells." Blood **117**(10): 2827-2838.

Lee, J. S., E. Smith and A. Shilatifard (2010). "The language of histone crosstalk." Cell **142**(5): 682-685.

Leenen, F. A., S. Vernocchi, O. E. Hunewald, S. Schmitz, A. M. Molitor, C. P. Muller and J. D. Turner (2016). "Where does transcription start? 5'-RACE adapted to next-generation sequencing." Nucleic Acids Res **44**(6): 2628-2645.
Lemaitre, C., F. Harper, G. Pierron, T. Heidmann and M. Dewannieux (2014). "The HERV-K human endogenous retrovirus envelope protein antagonizes Tetherin antiviral activity." J Virol **88**(23): 13626-13637.
Levinson, G. and G. A. Gutman (1987). "Slipped-strand mispairing: a major mechanism for DNA sequence evolution." Mol Biol Evol **4**(3): 203-221.
Li, R. C., P. Ping, J. Zhang, W. B. Wead, X. Cao, J. Gao, Y. Zheng, S. Huang, J. Han and R. Bolli (2000). "PKCepsilon modulates NF-kappaB and AP-1 via mitogen-activated protein kinases in adult rabbit cardiomyocytes." Am J Physiol Heart Circ Physiol **279**(4): H1679-1689.
Liang, G., J. C. Lin, V. Wei, C. Yoo, J. C. Cheng, C. T. Nguyen, D. J. Weisenberger, G. Egger, D. Takai, F. A. Gonzales and P. A. Jones (2004). "Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome." Proc Natl Acad Sci U S A **101**(19): 7357-7362.
Lieber, M. R. (2016). "Mechanisms of human lymphoid chromosomal translocations." Nat Rev Cancer **16**(6): 387-398.
Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander and J. Dekker (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." Science **326**(5950): 289-293.
Linke, F., S. Zaunig, M. M. Nietert, F. von Bonin, S. Lutz, C. Dullin, P. Janovska, T. Beissbarth, F. Alves, W. Klapper, V. Bryja, T. Pukrop, L. Trumper, J. Wilting and D. Kube (2017). "WNT5A: a motility-promoting factor in Hodgkin lymphoma." Oncogene **36**(1): 13-23.
Liu, G., J. J. Bissler, R. R. Sinden and M. Leffak (2007). "Unstable spinocerebellar ataxia type 10 (ATTCT*(AGAAT) repeats are associated with aberrant replication at the ATX10 locus and replication origin-dependent expansion at an ectopic site in human cells." Mol Cell Biol **27**(22): 7828-7838.
Liu, M. and M. V. Eiden (2011). "Role of human endogenous retroviral long terminal repeats (LTRs) in maintaining the integrity of the human germ line." Viruses **3**(6): 901-905.
Liu, P., E. Kimmoun, A. Legrand, A. Sauvanet, C. Degott, B. Lardeux and D. Bernuau (2002). "Activation of NF-kappa B, AP-1 and STAT transcription factors is a frequent and early event in human hepatocellular carcinomas." J Hepatol **37**(1): 63-71.
Liu, S. J., M. A. Horlbeck, S. W. Cho, H. S. Birk, M. Malatesta, D. He, F. J. Attenello, J. E. Villalta, M. Y. Cho, Y. Chen, M. A. Mandegar, M. P. Olvera, L. A. Gilbert, B. R. Conklin, H. Y. Chang, J. S. Weissman and D. A. Lim (2017). "CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells." Science **355**(6320).
Liu, X., D. A. Bushnell and R. D. Kornberg (2013). "RNA polymerase II transcription: structure and mechanism." Biochim Biophys Acta **1829**(1): 2-8.
Liu, Y., A. Sattarzadeh, A. Diepstra, L. Visser and A. van den Berg (2014). "The microenvironment in classical Hodgkin lymphoma: An actively shaped and essential tumor component." Semin Cancer Biol **24**: 15-22.

Lock, F. E., R. Rebollo, K. Miceli-Royer, L. Gagnier, S. Kuah, A. Babaian, M. Sistiaga-Poveda, C. B. Lai, O. Nemirovsky, I. Serrano, C. Steidl, M. M. Karimi and D. L. Mager (2014). "Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma." Proc Natl Acad Sci U S A **111**(34): E3534-3543.
Logan, C. Y. and R. Nusse (2004). "The Wnt signaling pathway in development and disease." Annu Rev Cell Dev Biol **20**: 781-810.
Logan, J., E. Falck-Pedersen, J. E. Darnell, Jr. and T. Shenk (1987). "A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene." Proc Natl Acad Sci U S A **84**(23): 8306-8310.
Lowe, C. B., G. Bejerano and D. Haussler (2007). "Thousands of human mobile element fragments undergo strong purifying selection near developmental genes." Proc Natl Acad Sci U S A **104**(19): 8005-8010.
Lower, R., J. Lower and R. Kurth (1996). "The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences." Proc Natl Acad Sci U S A **93**(11): 5177-5184.
Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.
Lupianez, D. G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel and S. Mundlos (2015). "Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions." Cell **161**(5): 1012-1025.
MacDonald, K. P., V. Rowe, H. M. Bofinger, R. Thomas, T. Sasmono, D. A. Hume and G. R. Hill (2005). "The colony-stimulating factor 1 receptor is expressed on dendritic cells during differentiation and regulates their expansion." J Immunol **175**(3): 1399-1405.
Mackay, F. and J. L. Browning (2002). "BAFF: a fundamental survival factor for B cells." Nat Rev Immunol **2**(7): 465-475.
Mackay, F., W. A. Figgett, D. Saulep, M. Lepage and M. L. Hibbs (2010). "B-cell stage and context-dependent requirements for survival signals from BAFF and the B-cell receptor." Immunol Rev **237**(1): 205-225.
Mager, D. L. and P. S. Henthorn (1984). "Identification of a retrovirus-like repetitive element in human DNA." Proc Natl Acad Sci U S A **81**(23): 7510-7514.
Mager, D. L. and M. C. Lorincz (2017). "Epigenetic modifier drugs trigger widespread transcription of endogenous retroviruses." Nat Genet **49**(7): 974-975.
Maggio, E. M., A. Van Den Berg, D. de Jong, A. Diepstra and S. Poppema (2003). "Low frequency of FAS mutations in Reed-Sternberg cells of Hodgkin's lymphoma." Am J Pathol **162**(1): 29-35.
Manghera, M. and R. N. Douville (2013). "Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors?" Retrovirology **10**: 16.
Margueron, R. and D. Reinberg (2010). "Chromatin structure and the inheritance of epigenetic information." Nat Rev Genet **11**(4): 285-296.
Martin-Subero, J. I., S. Gesk, L. Harder, T. Sonoki, P. W. Tucker, B. Schlegelberger, W. Grote, F. J. Novo, M. J. Calasanz, M. L. Hansmann, M. J. Dyer and R. Siebert (2002). "Recurrent involvement of the REL and BCL11A loci in classical Hodgkin lymphoma." Blood **99**(4): 1474-1477.

Masliah-Planchon, J., I. Bieche, J. M. Guinebretiere, F. Bourdeaut and O. Delattre (2015). "SWI/SNF chromatin remodeling and human malignancies." Annu Rev Pathol **10**: 145-171.
Masters, S. L., M. Gerlic, D. Metcalf, S. Preston, M. Pellegrini, J. A. O'Donnell, K. McArthur, T. M. Baldwin, S. Chevrier, C. J. Nowell, L. H. Cengia, K. J. Henley, J. E. Collinge, D. L. Kastner, L. Feigenbaum, D. J. Hilton, W. S. Alexander, B. T. Kile and B. A. Croker (2012). "NLRP1 inflammasome activation induces pyroptosis of hematopoietic progenitor cells." Immunity **37**(6): 1009-1023.
Mathas, S., M. Hinz, I. Anagnostopoulos, D. Krappmann, A. Lietz, F. Jundt, K. Bommert, F. Mechta-Grigoriou, H. Stein, B. Dorken and C. Scheidereit (2002). "Aberrantly expressed c-Jun and JunB are a hallmark of Hodgkin lymphoma cells, stimulate proliferation and synergize with NF-kappa B." EMBO J **21**(15): 4104-4113.
Mathas, S., M. Janz, F. Hummel, M. Hummel, B. Wollert-Wulf, S. Lusatis, I. Anagnostopoulos, A. Lietz, M. Sigvardsson, F. Jundt, K. Johrens, K. Bommert, H. Stein and B. Dorken (2006). "Intrinsic inhibition of transcription factor E2A by HLH proteins ABF-1 and Id2 mediates reprogramming of neoplastic B cells in Hodgkin lymphoma." Nat Immunol **7**(2): 207-215.
McClintock, B. (1968). "The states of a gene locus in maize." Carnegie Institute of Washington Yearbook **66**(20-28).
McCune, R. C., S. I. Syrbu and M. A. Vasef (2006). "Expression profiling of transcription factors Pax-5, Oct-1, Oct-2, BOB.1, and PU.1 in Hodgkin's and non-Hodgkin's lymphomas: a comparative study using high throughput tissue microarrays." Mod Pathol **19**(7): 1010-1018.
McKercher, S. R., B. E. Torbett, K. L. Anderson, G. W. Henkel, D. J. Vestal, H. Baribault, M. Klemsz, A. J. Feeney, G. E. Wu, C. J. Paige and R. A. Maki (1996). "Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities." EMBO J **15**(20): 5647-5658.
McKinley, K. L. and I. M. Cheeseman (2016). "The molecular basis for centromere identity and function." Nat Rev Mol Cell Biol **17**(1): 16-29.
Medstrand, P., J. R. Landry and D. L. Mager (2001). "Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans." J Biol Chem **276**(3): 1896-1903.
Medstrand, P. and D. L. Mager (1998). "Human-specific integrations of the HERV-K endogenous retrovirus family." J Virol **72**(12): 9782-9787.
Meneghini, M. D., M. Wu and H. D. Madhani (2003). "Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin." Cell **112**(5): 725-736.
Meru, N., A. Jung, R. Lisner and G. Niedobitek (2001). "Expression of the recombination activating genes (RAG1 and RAG2) is not detectable in Epstein-Barr virus-associated human lymphomas." Int J Cancer **92**(1): 75-78.
Mi, S., X. Lee, X. Li, G. M. Veldman, H. Finnerty, L. Racie, E. LaVallie, X. Y. Tang, P. Edouard, S. Howes, J. C. Keith, Jr. and J. M. McCoy (2000). "Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis." Nature **403**(6771): 785-789.
Mifsud, B., F. Tavares-Cadete, A. N. Young, R. Sugar, S. Schoenfelder, L. Ferreira, S. W. Wingett, S. Andrews, W. Grey, P. A. Ewels, B. Herman, S. Happe, A. Higgs, E. LeProust, G. A. Follows, P. Fraser, N. M. Luscombe and C. S. Osborne (2015). "Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C." Nat Genet **47**(6): 598-606.

Missra, A. and D. S. Gilmour (2010). "Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex." Proc Natl Acad Sci U S A **107**(25): 11301-11306.
Mizuno, T., X. Zhong and T. L. Rothstein (2003). "Fas-induced apoptosis in B cells." Apoptosis **8**(5): 451-460.
Morandi, A., V. Barbetti, M. Riverso, P. Dello Sbarba and E. Rovida (2011). "The colony-stimulating factor-1 (CSF-1) receptor sustains ERK1/2 activation and proliferation in breast cancer cell lines." PLoS One **6**(11): e27450.
Munoz-Lopez, M. and J. L. Garcia-Perez (2010). "DNA transposons: nature and applications in genomics." Curr Genomics **11**(2): 115-128.
Muotri, A. R., M. C. Marchetto, N. G. Coufal, R. Oefner, G. Yeo, K. Nakashima and F. H. Gage (2010). "L1 retrotransposition in neurons is modulated by MeCP2." Nature **468**(7322): 443-446.
Nagel, S., C. Meyer, M. Kaufmann, H. G. Drexler and R. A. MacLeod (2015). "Aberrant expression of homeobox gene SIX1 in Hodgkin lymphoma." Oncotarget **6**(37): 40112-40126.
Nageshwaran, S. and R. Festenstein (2015). "Epigenetics and Triplet-Repeat Neurological Diseases." Front Neurol **6**: 262.
Nakagawa, K., V. Brusic, G. McColl and L. C. Harrison (1997). "Direct evidence for the expression of multiple endogenous retroviruses in the synovial compartment in rheumatoid arthritis." Arthritis Rheum **40**(4): 627-638.
NCI, N. C. I. (2008, 2008). "Reed-Sternberg Cell." Retrieved 18th March, 2018, from https://visuals.nci.nih.gov/details.cfm?imageid=7172.
Nelson, P. N., P. R. Carnegie, J. Martin, H. Davari Ejtehadi, P. Hooley, D. Roden, S. Rowland-Jones, P. Warren, J. Astley and P. G. Murray (2003). "Demystified. Human endogenous retroviruses." Mol Pathol **56**(1): 11-18.
Nelson, P. N., A. M. Lever, S. Smith, R. Pitman, P. Murray, S. A. Perera, O. M. Westwood, F. C. Hay, H. D. Ejtehadi and J. C. Booth (1999). "Molecular investigations implicate human endogenous retroviruses as mediators of anti-retroviral antibodies in autoimmune rheumatic disease." Immunol Invest **28**(4): 277-289.
Nicholson, S., H. Whitehouse, K. Naidoo and R. J. Byers (2011). "Yin Yang 1 in human cancer." Crit Rev Oncog **16**(3-4): 245-260.
Nock, A., J. M. Ascano, M. J. Barrero and S. Malik (2012). "Mediator-regulated transcription through the +1 nucleosome." Mol Cell **48**(6): 837-848.
Nogova, L., T. Rudiger and A. Engert (2006). "Biology, clinical course and management of nodular lymphocyte-predominant hodgkin lymphoma." Hematology Am Soc Hematol Educ Program: 266-272.
Nora, E. P., A. Goloborodko, A. L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny and B. G. Bruneau (2017). "Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization." Cell **169**(5): 930-944 e922.
Nutt, S. L., B. Heavey, A. G. Rolink and M. Busslinger (1999). "Commitment to the B-lymphoid lineage depends on the transcription factor Pax5." Nature **401**(6753): 556-562.
O'Grady, J. T., S. Stewart, J. Lowrey, S. E. Howie and A. S. Krajewski (1994). "CD40 expression in Hodgkin's disease." Am J Pathol **144**(1): 21-26.
Obier, N., P. Cauchy, S. A. Assi, J. Gilmour, A. L. M. Lie, M. Lichtinger, M. Hoogenkamp, L. Noailles, P. N. Cockerill, G. Lacaud, V. Kouskoff and C. Bonifer

(2016). "Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate." Development **143**(23): 4324-4340.
Orphanides, G., G. LeRoy, C. H. Chang, D. S. Luse and D. Reinberg (1998). "FACT, a factor that facilitates transcript elongation through nucleosomes." Cell **92**(1): 105-116.
Pace, J. K., 2nd and C. Feschotte (2007). "The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage." Genome Res **17**(4): 422-432.
Padeken, J., P. Zeller and S. M. Gasser (2015). "Repeat DNA in genome organization and stability." Curr Opin Genet Dev **31**: 12-19.
Papamichos-Chronakis, M., S. Watanabe, O. J. Rando and C. L. Peterson (2011). "Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity." Cell **144**(2): 200-213.
Patsialou, A., J. Wyckoff, Y. Wang, S. Goswami, E. R. Stanley and J. S. Condeelis (2009). "Invasion of human breast cancer cells in vivo requires both paracrine and autocrine loops involving the colony-stimulating factor-1 receptor." Cancer Res **69**(24): 9498-9506.
Paweletz, N. (2001). "Walther Flemming: pioneer of mitosis research." Nat Rev Mol Cell Biol **2**(1): 72-75.
Pelechano, V. and L. M. Steinmetz (2013). "Gene regulation by antisense transcription." Nat Rev Genet **14**(12): 880-893.
Perron, H., J. A. Garson, F. Bedin, F. Beseme, G. Paranhos-Baccala, F. Komurian-Pradel, F. Mallet, P. W. Tuke, C. Voisset, J. L. Blond, B. Lalande, J. M. Seigneurin and B. Mandrand (1997). "Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. The Collaborative Research Group on Multiple Sclerosis." Proc Natl Acad Sci U S A **94**(14): 7583-7588.
Philippe, C., D. B. Vargas-Landin, A. J. Doucet, D. van Essen, J. Vera-Otarola, M. Kuciak, A. Corbin, P. Nigumann and G. Cristofari (2016). "Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci." Elife **5**.
Pi, W., Z. Yang, J. Wang, L. Ruan, X. Yu, J. Ling, S. Krantz, C. Isales, S. J. Conway, S. Lin and D. Tuan (2004). "The LTR enhancer of ERV-9 human endogenous retrovirus is active in oocytes and progenitor cells in transgenic zebrafish and humans." Proc Natl Acad Sci U S A **101**(3): 805-810.
Piccaluga, P. P., C. Agostinelli, A. Gazzola, C. Tripodo, F. Bacci, E. Sabattini, M. T. Sista, C. Mannu, M. R. Sapienza, M. Rossi, M. A. Laginestra, C. A. Sagramoso-Sacchetti, S. Righi and S. A. Pileri (2011). "Pathobiology of hodgkin lymphoma." Adv Hematol **2011**: 920898.
Pillai, S. and A. Cariappa (2009). "The follicular versus marginal zone B lymphocyte cell fate decision." Nat Rev Immunol **9**(11): 767-777.
Piper, J., M. C. Elze, P. Cauchy, P. N. Cockerill, C. Bonifer and S. Ott (2013). "Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data." Nucleic Acids Res **41**(21): e201.
Piper, J., M. C. Elze, P. Cauchy, P. N. Cockerill, C. Bonifer and S. Ott (2013). "Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data." Nucleic Acids Research **41**(21).
Piper, J., M. C. Elze, P. Cauchy, P. N. Cockerill, C. Bonifer and S. Ott (2014). "Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data." Nucleic Acids Research **42**(17): 11272.

Pokholok, D. K., C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford and R. A. Young (2005). "Genome-wide map of nucleosome acetylation and methylation in yeast." Cell **122**(4): 517-527.
Polavarapu, N., N. J. Bowen and J. F. McDonald (2006). "Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses." Genome Biol **7**(6): R51.
Prensner, J. R., M. K. Iyer, A. Sahu, I. A. Asangani, Q. Cao, L. Patel, I. A. Vergara, E. Davicioni, N. Erho, M. Ghadessi, R. B. Jenkins, T. J. Triche, R. Malik, R. Bedenis, N. McGregor, T. Ma, W. Chen, S. Han, X. Jing, X. Cao, X. Wang, B. Chandler, W. Yan, J. Siddiqui, L. P. Kunju, S. M. Dhanasekaran, K. J. Pienta, F. Y. Feng and A. M. Chinnaiyan (2013). "The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex." Nat Genet **45**(11): 1392-1398.
Proudfoot, N. J. (1989). "How RNA polymerase II terminates transcription in higher eukaryotes." Trends Biochem Sci **14**(3): 105-110.
Prudhomme, S., G. Oriol and F. Mallet (2004). "A retroviral promoter and a cellular enhancer define a bipartite element which controls env ERVWE1 placental expression." J Virol **78**(22): 12157-12168.
Qu, G., L. Dubeau, A. Narayan, M. C. Yu and M. Ehrlich (1999). "Satellite DNA hypomethylation vs. overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential." Mutat Res **423**(1-2): 91-101.
Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.
R-Development-Core-Team. (2008). "R: A language and environment for statistical computing.", from http://www.R-project.org.
Rasmussen, E. B. and J. T. Lis (1993). "In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes." Proc Natl Acad Sci U S A **90**(17): 7923-7927.
Ravindran, S. (2012). "Barbara McClintock and the discovery of jumping genes." Proc Natl Acad Sci U S A **109**(50): 20198-20199.
Rawlings, J. S., K. M. Rosler and D. A. Harrison (2004). "The JAK/STAT signaling pathway." J Cell Sci **117**(Pt 8): 1281-1283.
Reiss, D., Y. Zhang and D. L. Mager (2007). "Widely variable endogenous retroviral methylation levels in human placenta." Nucleic Acids Res **35**(14): 4743-4754.
Renne, C., J. I. Martin-Subero, M. Eickernjager, M. L. Hansmann, R. Kuppers, R. Siebert and A. Brauninger (2006). "Aberrant expression of ID2, a suppressor of B-cell-specific gene expression, in Hodgkin's lymphoma." Am J Pathol **169**(2): 655-664.
Richard, G. F. and F. Paques (2000). "Mini- and microsatellite expansions: the recombination connection." EMBO Rep **1**(2): 122-126.
Richard, P. and J. L. Manley (2009). "Transcription termination by nuclear RNA polymerases." Genes Dev **23**(11): 1247-1269.
Rodriguez, J., L. Vives, M. Jorda, C. Morales, M. Munoz, E. Vendrell and M. A. Peinado (2008). "Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells." Nucleic Acids Res **36**(3): 770-784.
Romanish, M. T., W. M. Lock, L. N. van de Lagemaat, C. A. Dunn and D. L. Mager (2007). "Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution." PLoS Genet **3**(1): e10.

Rosenfeld, C., A. Goutner, C. Choquet, A. M. Venuat, B. Kayibanda, J. L. Pico and M. F. Greaves (1977). "Phenotypic characterisation of a unique non-T, non-B acute lymphoblastic leukaemia cell line." Nature **267**(5614): 841-843.
Ross, I. L., X. Yue, M. C. Ostrowski and D. A. Hume (1998). "Interaction between PU.1 and another Ets family transcription factor promotes macrophage-specific Basal transcription initiation." J Biol Chem **273**(12): 6662-6669.
Rothstein, T. L. (2000). "Inducible resistance to Fas-mediated apoptosis in B cells." Cell Res **10**(4): 245-266.
Ruda, V. M., S. B. Akopov, D. O. Trubetskoy, N. L. Manuylov, A. S. Vetchinova, L. L. Zavalova, L. G. Nikolaev and E. D. Sverdlov (2004). "Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR." Virus Res **104**(1): 11-16.
Ruhl, D. D., J. Jin, Y. Cai, S. Swanson, L. Florens, M. P. Washburn, R. C. Conaway, J. W. Conaway and J. C. Chrivia (2006). "Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes." Biochemistry **45**(17): 5671-5677.
Ruprecht, K., J. Mayer, M. Sauter, K. Roemer and N. Mueller-Lantzsch (2008). "Endogenous retroviruses and cancer." Cell Mol Life Sci **65**(21): 3366-3382.
Sainsbury, S., C. Bernecky and P. Cramer (2015). "Structural basis of transcription initiation by RNA polymerase II." Nat Rev Mol Cell Biol **16**(3): 129-143.
Saldanha, A. J. (2004). "Java Treeview--extensible visualization of microarray data." Bioinformatics **20**(17): 3246-3248.
Sanyal, A., B. R. Lajoie, G. Jain and J. Dekker (2012). "The long-range interaction landscape of gene promoters." Nature **489**(7414): 109-113.
Sauter, K. A., M. A. Bouhlel, J. O'Neal, D. P. Sester, H. Tagoh, R. M. Ingram, C. Pridans, C. Bonifer and D. A. Hume (2013). "The function of the conserved regulatory element within the second intron of the mammalian Csf1r locus." PLoS One **8**(1): e54935.
Sauter, M., K. Roemer, B. Best, M. Afting, S. Schommer, G. Seitz, M. Hartmann and N. Mueller-Lantzsch (1996). "Specificity of antibodies directed against Env protein of human endogenous retroviruses in patients with germ cell tumors." Cancer Res **56**(19): 4362-4365.
Sauter, M., S. Schommer, E. Kremmer, K. Remberger, G. Dolken, I. Lemm, M. Buck, B. Best, D. Neumann-Haefelin and N. Mueller-Lantzsch (1995). "Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas." J Virol **69**(1): 414-421.
Saveliev, A., C. Everett, T. Sharpe, Z. Webster and R. Festenstein (2003). "DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing." Nature **422**(6934): 909-913.
Scarfo, I., E. Pellegrino, E. Mereu, I. Kwee, L. Agnelli, E. Bergaggio, G. Garaffo, N. Vitale, M. Caputo, R. Machiorlatti, P. Circosta, F. Abate, A. Barreca, D. Novero, S. Mathew, A. Rinaldi, E. Tiacci, S. Serra, S. Deaglio, A. Neri, B. Falini, R. Rabadan, F. Bertoni, G. Inghirami, R. Piva and T. C. L. S. G. European (2016). "Identification of a new subclass of ALK-negative ALCL expressing aberrant levels of ERBB4 transcripts." Blood **127**(2): 221-232.
Schaadt, M., V. Diehl, H. Stein, C. Fonatsch and H. H. Kirchner (1980). "Two neoplastic cell lines with unique features derived from Hodgkin's disease." Int J Cancer **26**(6): 723-731.
Schmitt, K., J. Reichrath, A. Roesch, E. Meese and J. Mayer (2013). "Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma." Genome Biol Evol **5**(2): 307-328.

Schmitz, R., M. L. Hansmann, V. Bohle, J. I. Martin-Subero, S. Hartmann, G. Mechtersheimer, W. Klapper, I. Vater, M. Giefing, S. Gesk, J. Stanelle, R. Siebert and R. Kuppers (2009). "TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma." J Exp Med **206**(5): 981-989.
Schoenherr, C. J. and D. J. Anderson (1995). "Silencing is golden: negative regulation in the control of neuronal gene transcription." Curr Opin Neurobiol **5**(5): 566-571.
Schultz, H., K. Engel and M. Gaestel (1997). "PMA-induced activation of the p42/44ERK- and p38RK-MAP kinase cascades in HL-60 cells is PKC dependent but not essential for differentiation to the macrophage-like phenotype." J Cell Physiol **173**(3): 310-318.
Schwering, I., A. Brauninger, U. Klein, B. Jungnickel, M. Tinguely, V. Diehl, M. L. Hansmann, R. Dalla-Favera, K. Rajewsky and R. Kuppers (2003). "Loss of the B-lineage-specific gene expression program in Hodgkin and Reed-Sternberg cells of Hodgkin lymphoma." Blood **101**(4): 1505-1512.
Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Res **13**(11): 2498-2504.
Sharma, S. C. and J. S. Richards (2000). "Regulation of AP1 (Jun/Fos) factor expression and activation in ovarian granulosa cells. Relation of JunD and Fra2 to terminal differentiation." J Biol Chem **275**(43): 33718-33728.
Shaulian, E. and M. Karin (2002). "AP-1 as a regulator of cell life and death." Nat Cell Biol **4**(5): E131-136.
Silva, J. C., S. A. Shabalina, D. G. Harris, J. L. Spouge and A. S. Kondrashovi (2003). "Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes." Genet Res **82**(1): 1-18.
Sims, D., I. Sudbery, N. E. Ilott, A. Heger and C. P. Ponting (2014). "Sequencing depth and coverage: key considerations in genomic analyses." Nat Rev Genet **15**(2): 121-132.
Skinnider, B. F. and T. W. Mak (2002). "The role of cytokines in classical Hodgkin lymphoma." Blood **99**(12): 4283-4297.
Slater, S. J., M. B. Kelly, F. J. Taddeo, E. Rubin and C. D. Stubbs (1994). "Evidence for discrete diacylglycerol and phorbol ester activator sites on protein kinase C. Differences in effects of 1-alkanol inhibition, activation by phosphatidylethanolamine and calcium chelation." J Biol Chem **269**(25): 17160-17165.
Smit, A. F. (1993). "Identification of a new, abundant superfamily of mammalian LTR-transposons." Nucleic Acids Res **21**(8): 1863-1872.
Smit, A. F. A. H., R.; Green, P. (2013-2015). "RepeatMasker Open-4.0." from http://www.repeatmasker.org.
Smith, S. M. and L. Cai (2012). "Cell specific CD44 expression in breast cancer requires the interaction of AP-1 and NFkappaB with a novel cis-element." PLoS One **7**(11): e50867.
Smith, Z. D., M. M. Chan, K. C. Humm, R. Karnik, S. Mekhoubad, A. Regev, K. Eggan and A. Meissner (2014). "DNA methylation dynamics of the human preimplantation embryo." Nature **511**(7511): 611-615.
Snider, A. C., D. Leong, Q. T. Wang, J. Wysocka, M. W. Yao and M. P. Scott (2013). "The chromatin remodeling factor Chd1l is required in the preimplantation embryo." Biol Open **2**(2): 121-131.

Soboleva, T. A., M. Nekrasov, D. P. Ryan and D. J. Tremethick (2014). "Histone variants at the transcription start-site." Trends Genet **30**(5): 199-209.
Sokol, M., K. M. Jessen and F. S. Pedersen (2016). "Utility of next-generation RNA-sequencing in identifying chimeric transcription involving human endogenous retroviruses." APMIS **124**(1-2): 127-139.
St Laurent, G., D. Shtokalo, B. Dong, M. R. Tackett, X. Fan, S. Lazorthes, E. Nicolas, N. Sang, T. J. Triche, T. A. McCaffrey, W. Xiao and P. Kapranov (2013). "VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer." Genome Biol **14**(7): R73.
Staratschek-Jox, A., S. Kotkowski, G. Belge, T. Rudiger, J. Bullerdiek, V. Diehl and J. Wolf (2000). "Detection of Epstein-Barr virus in Hodgkin-Reed-Sternberg cells : no evidence for the persistence of integrated viral fragments inLatent membrane protein-1 (LMP-1)-negative classical Hodgkin's disease." Am J Pathol **156**(1): 209-216.
Steidl, C., A. Diepstra, T. Lee, F. C. Chan, P. Farinha, K. Tan, A. Telenius, L. Barclay, S. P. Shah, J. M. Connors, A. van den Berg and R. D. Gascoyne (2012). "Gene expression profiling of microdissected Hodgkin Reed-Sternberg cells correlates with treatment outcome in classical Hodgkin lymphoma." Blood **120**(17): 3530-3540.
Stein, H., T. Marafioti, H. D. Foss, H. Laumen, M. Hummel, I. Anagnostopoulos, T. Wirth, G. Demel and B. Falini (2001). "Down-regulation of BOB.1/OBF.1 and Oct2 in classical Hodgkin disease but not in lymphocyte predominant Hodgkin disease correlates with immunoglobulin transcription." Blood **97**(2): 496-501.
Stengel, S., U. Fiebig, R. Kurth and J. Denner (2010). "Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation." Genes Chromosomes Cancer **49**(5): 401-411.
Sun, S. C. (2011). "Non-canonical NF-kappaB signaling pathway." Cell Res **21**(1): 71-85.
Sverdlov, E. D. (2000). "Retroviruses and primate evolution." Bioessays **22**(2): 161-171.
Tagami, H., D. Ray-Gallet, G. Almouzni and Y. Nakatani (2004). "Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis." Cell **116**(1): 51-61.
Tagoh, H., R. Ingram, N. Wilson, G. Salvagiotto, A. J. Warren, D. Clarke, M. Busslinger and C. Bonifer (2006). "The mechanism of repression of the myeloid-specific c-fms gene by Pax5 during B lineage restriction." EMBO J **25**(5): 1070-1080.
Takahashi, H., T. J. Parmely, S. Sato, C. Tomomori-Sato, C. A. Banks, S. E. Kong, H. Szutorisz, S. K. Swanson, S. Martin-Brown, M. P. Washburn, L. Florens, C. W. Seidel, C. Lin, E. R. Smith, A. Shilatifard, R. C. Conaway and J. W. Conaway (2011). "Human mediator subunit MED26 functions as a docking site for transcription elongation factors." Cell **146**(1): 92-104.
Talbert, P. B. and S. Henikoff (2010). "Histone variants--ancient wrap artists of the epigenome." Nat Rev Mol Cell Biol **11**(4): 264-275.
Tang, G. Q., R. Roy, R. P. Bandwar, T. Ha and S. S. Patel (2009). "Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase." Proc Natl Acad Sci U S A **106**(52): 22175-22180.
Tassone, F., A. Beilina, C. Carosi, S. Albertosi, C. Bagni, L. Li, K. Glover, D. Bentley and P. J. Hagerman (2007). "Elevated FMR1 mRNA in premutation carriers is due to increased transcription." RNA **13**(4): 555-562.

Tassone, F., R. J. Hagerman, A. K. Taylor, L. W. Gane, T. E. Godfrey and P. J. Hagerman (2000). "Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome." Am J Hum Genet **66**(1): 6-15.
Thakar, N. Y., D. A. Ovchinnikov, M. L. Hastie, B. Kobe, J. J. Gorman and E. J. Wolvetang (2015). "TRAF2 recruitment via T61 in CD30 drives NFkappaB activation and enhances hESC survival and proliferation." Mol Biol Cell **26**(5): 993-1006.
Thomas, J. H. and S. Schneider (2011). "Coevolution of retroelements and tandem zinc finger genes." Genome Res **21**(11): 1800-1812.
Thomas, R. K., A. Kallenborn, C. Wickenhauser, J. L. Schultze, A. Draube, M. Vockerodt, D. Re, V. Diehl and J. Wolf (2002). "Constitutive expression of c-FLIP in Hodgkin and Reed-Sternberg cells." Am J Pathol **160**(4): 1521-1528.
Thomas, R. K., D. Re, J. Wolf and V. Diehl (2004). "Part I: Hodgkin's lymphoma—molecular biology of Hodgkin and Reed-Sternberg cells." The lancet oncology **5**(1): 11-18.
Tiacci, E., C. Doring, V. Brune, C. J. van Noesel, W. Klapper, G. Mechtersheimer, B. Falini, R. Kuppers and M. L. Hansmann (2012). "Analyzing primary Hodgkin and Reed-Sternberg cells to capture the molecular and cellular pathogenesis of classical Hodgkin lymphoma." Blood **120**(23): 4609-4620.
Ting, C. N., M. P. Rosenberg, C. M. Snow, L. C. Samuelson and M. H. Meisler (1992). "Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene." Genes Dev **6**(8): 1457-1465.
Tkach, V., E. Bock and V. Berezin (2005). "The role of RhoA in the regulation of cell morphology and motility." Cell Motil Cytoskeleton **61**(1): 21-33.
Tolhuis, B., R. J. Palstra, E. Splinter, F. Grosveld and W. de Laat (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." Mol Cell **10**(6): 1453-1465.
Torlakovic, E., A. Tierens, H. D. Dang and J. Delabie (2001). "The transcription factor PU.1, necessary for B-cell development is expressed in lymphocyte predominance, but not classical Hodgkin's disease." Am J Pathol **159**(5): 1807-1814.
Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn and L. Pachter (2012). "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks." Nat Protoc **7**(3): 562-578.
Trapnell, C. and S. L. Salzberg (2009). "How to map billions of short reads onto genomes." Nat Biotechnol **27**(5): 455-457.
Trask, O. J., Jr. (2004). Nuclear Factor Kappa B (NF-kappaB) Translocation Assay Development and Validation for High Content Screening. Assay Guidance Manual. G. S. Sittampalam, N. P. Coussens, K. Brimacombe et al. Bethesda (MD).
Treangen, T. J. and S. L. Salzberg (2011). "Repetitive DNA and next-generation sequencing: computational challenges and solutions." Nat Rev Genet **13**(1): 36-46.
Tsukiyama, T., P. B. Becker and C. Wu (1994). "ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor." Nature **367**(6463): 525-532.
Turelli, P., N. Castro-Diaz, F. Marzetta, A. Kapopoulou, C. Raclot, J. Duc, V. Tieng, S. Quenneville and D. Trono (2014). "Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements." Genome Res **24**(8): 1260-1270.
Tyree, C. M., C. P. George, L. M. Lira-DeVito, S. L. Wampler, M. E. Dahmus, L. Zawel and J. T. Kadonaga (1993). "Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II." Genes Dev **7**(7A): 1254-1265.

Tzankov, A., A. Zimpfer, A. C. Pehrs, A. Lugli, P. Went, R. Maurer, S. Pileri and S. Dirnhofer (2003). "Expression of B-cell markers in classical hodgkin lymphoma: a tissue microarray analysis of 330 cases." Mod Pathol **16**(11): 1141-1147.
Uehira, K., R. Amakawa, T. Ito, T. Uehira, Y. Ozaki, T. Shimizu, M. Fujimoto, M. Inaba and S. Fukuhara (2001). "A Hodgkin's disease cell line, KM-H2, shows biphenotypic features of dendritic cells and B cells." Int J Hematol **73**(2): 236-244.
Ullu, E. and C. Tschudi (1984). "Alu sequences are processed 7SL RNA genes." Nature **312**(5990): 171-172.
Upton, K. R., D. J. Gerhardt, J. S. Jesuadian, S. R. Richardson, F. J. Sanchez-Luque, G. O. Bodea, A. D. Ewing, C. Salvador-Palomeque, M. S. van der Knaap, P. M. Brennan, A. Vanderver and G. J. Faulkner (2015). "Ubiquitous L1 mosaicism in hippocampal neurons." Cell **161**(2): 228-239.
Usdin, K. (2008). "The biological effects of simple tandem repeats: lessons from the repeat expansion diseases." Genome Res **18**(7): 1011-1019.
van Overveld, P. G., R. J. Lemmers, L. A. Sandkuijl, L. Enthoven, S. T. Winokur, F. Bakels, G. W. Padberg, G. J. van Ommen, R. R. Frants and S. M. van der Maarel (2003). "Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy." Nat Genet **35**(4): 315-317.
Vaquero, A., A. Loyola and D. Reinberg (2003). "The constantly changing face of chromatin." Sci Aging Knowledge Environ **2003**(14): RE4.
Veeman, M. T., J. D. Axelrod and R. T. Moon (2003). "A second canon. Functions and mechanisms of beta-catenin-independent Wnt signaling." Dev Cell **5**(3): 367-377.
Venkatesh, S. and J. L. Workman (2015). "Histone exchange, chromatin structure and the regulation of transcription." Nat Rev Mol Cell Biol **16**(3): 178-189.
Visvader, J. and I. M. Verma (1989). "Differential transcription of exon 1 of the human c-fms gene in placental trophoblasts and monocytes." Mol Cell Biol **9**(3): 1336-1341.
Vogel, G. (2011). "Retrotransposons. Do jumping genes spawn diversity?" Science **332**(6027): 300-301.
von Strandmann, E. P., A. Nastos, B. Holewa, S. Senkel, H. Weber and G. U. Ryffel (1997). "Patterning the expression of a tissue-specific transcription factor in embryogenesis: HNF1 alpha gene activation during Xenopus development." Mech Dev **64**(1-2): 7-17.
Voon, H. P. and L. H. Wong (2016). "New players in heterochromatin silencing: histone variant H3.3 and the ATRX/DAXX chaperone." Nucleic Acids Res **44**(4): 1496-1501.
Wael, A. Y., T. Fujiwara, Y. Okitsu, Y. Onishi, K. Ishizawa and H. Harigae (2011). "Exploring the Mechanism of ETO2-Dependent Transcriptional Regulation in Erythroid Cells." Blood **118**(21): 1307-1307.
Wagner, E. J. and P. B. Carpenter (2012). "Understanding the language of Lys36 methylation at histone H3." Nat Rev Mol Cell Biol **13**(2): 115-126.
Wai, L. K. (2004). "Telomeres, telomerase, and tumorigenesis--a review." MedGenMed **6**(3): 19.
Wajant, H., K. Pfizenmaier and P. Scheurich (2003). "Tumor necrosis factor signaling." Cell Death Differ **10**(1): 45-65.
Wang-Johanning, F., J. Liu, K. Rycaj, M. Huang, K. Tsai, D. G. Rosen, D. T. Chen, D. W. Lu, K. F. Barnhart and G. L. Johanning (2007). "Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer." Int J Cancer **120**(1): 81-90.

Wang-Johanning, F., L. Radvanyi, K. Rycaj, J. B. Plummer, P. Yan, K. J. Sastry, C. J. Piyathilake, K. K. Hunt and G. L. Johanning (2008). "Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients." Cancer Res **68**(14): 5869-5877.
Wang, T., J. Zeng, C. B. Lowe, R. G. Sellers, S. R. Salama, M. Yang, S. M. Burgess, R. K. Brachmann and D. Haussler (2007). "Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53." Proc Natl Acad Sci U S A **104**(47): 18613-18618.
Watanabe, S., M. Radman-Livaja, O. J. Rando and C. L. Peterson (2013). "A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme." Science **340**(6129): 195-199.
Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." Nature **171**(4356): 737-738.
Weeraratna, A. T., Y. Jiang, G. Hostetter, K. Rosenblatt, P. Duray, M. Bittner and J. M. Trent (2002). "Wnt5a signaling directly affects cell motility and invasion of metastatic melanoma." Cancer Cell **1**(3): 279-288.
Weintraub, H. and M. Groudine (1976). "Chromosomal subunits in active genes have an altered conformation." Science **193**(4256): 848-856.
West, M. J., A. D. Lowe and J. Karn (2001). "Activation of human immunodeficiency virus transcription in T cells revisited: NF-kappaB p65 stimulates transcriptional elongation." J Virol **75**(18): 8524-8537.
West, S., N. J. Proudfoot and M. J. Dye (2008). "Molecular dissection of mammalian RNA polymerase II transcriptional termination." Mol Cell **29**(5): 600-610.
Whitelaw, E. and D. I. Martin (2001). "Retrotransposons as epigenetic mediators of phenotypic variation in mammals." Nat Genet **27**(4): 361-365.
Wiesner, T., W. Lee, A. C. Obenauf, L. Ran, R. Murali, Q. F. Zhang, E. W. Wong, W. Hu, S. N. Scott, R. H. Shah, I. Landa, J. Button, N. Lailler, A. Sboner, D. Gao, D. A. Murphy, Z. Cao, S. Shukla, T. J. Hollmann, L. Wang, L. Borsu, T. Merghoub, G. K. Schwartz, M. A. Postow, C. E. Ariyan, J. A. Fagin, D. Zheng, M. Ladanyi, K. J. Busam, M. F. Berger, Y. Chen and P. Chi (2015). "Alternative transcription initiation leads to expression of a novel ALK isoform in cancer." Nature **526**(7573): 453-457.
Wiznerowicz, M., J. Jakobsson, J. Szulc, S. Liao, A. Quazzola, F. Beermann, P. Aebischer and D. Trono (2007). "The Kruppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis." J Biol Chem **282**(47): 34535-34541.
Wolf, G., D. Greenberg and T. S. Macfarlan (2015). "Spotting the enemy within: Targeted silencing of foreign DNA in mammalian genomes by the Kruppel-associated box zinc finger protein family." Mob DNA **6**: 17.
Wolf, G., R. Rebollo, M. M. Karimi, A. D. Ewing, R. Kamada, W. Wu, B. Wu, M. Bachu, K. Ozato, G. J. Faulkner, D. L. Mager, M. C. Lorincz and T. S. Macfarlan (2017). "On the role of H3.3 in retroviral silencing." Nature **548**(7665): E1-E3.
Wolf, G., P. Yang, A. C. Fuchtbauer, E. M. Fuchtbauer, A. M. Silva, C. Park, W. Wu, A. L. Nielsen, F. S. Pedersen and T. S. Macfarlan (2015). "The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses." Genes Dev **29**(5): 538-554.
Wolf, J., U. Kapp, H. Bohlen, M. Kornacker, C. Schoch, B. Stahl, S. Mucke, C. von Kalle, C. Fonatsch, H. E. Schaefer, M. L. Hansmann and V. Diehl (1996). "Peripheral blood mononuclear cells of a patient with advanced Hodgkin's lymphoma give rise to permanently growing Hodgkin-Reed Sternberg cells." Blood **87**(8): 3418-3428.

Wray, G. A. (2007). "The evolutionary significance of cis-regulatory mutations." Nat Rev Genet **8**(3): 206-216.
Xie, L., A. Ushmorov, F. Leithauser, H. Guan, C. Steidl, J. Farbinger, C. Pelzer, M. J. Vogel, H. J. Maier, R. D. Gascoyne, P. Moller and T. Wirth (2012). "FOXO1 is a tumor suppressor in classical Hodgkin lymphoma." Blood **119**(15): 3503-3511.
Xie, M., C. Hong, B. Zhang, R. F. Lowdon, X. Xing, D. Li, X. Zhou, H. J. Lee, C. L. Maire, K. L. Ligon, P. Gascard, M. Sigaroudinia, T. D. Tlsty, T. Kadlecek, A. Weiss, H. O'Geen, P. J. Farnham, P. A. Madden, A. J. Mungall, A. Tam, B. Kamoh, S. Cho, R. Moore, M. Hirst, M. A. Marra, J. F. Costello and T. Wang (2013). "DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape." Nat Genet **45**(7): 836-841.
Xin, D., L. Hu and X. Kong (2008). "Alternative promoters influence alternative splicing at the genomic level." PLoS One **3**(6): e2377.
Xin, H., S. Takahata, M. Blanksma, L. McCullough, D. J. Stillman and T. Formosa (2009). "yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement." Mol Cell **35**(3): 365-376.
Xing, J., Y. Zhang, K. Han, A. H. Salem, S. K. Sen, C. D. Huff, Q. Zhou, E. F. Kirkness, S. Levy, M. A. Batzer and L. B. Jorde (2009). "Mobile elements create structural variation: analysis of a complete human genome." Genome Res **19**(9): 1516-1526.
Xu, L., A. G. Elkahloun, F. Candotti, A. Grajkowski, S. L. Beaucage, E. F. Petricoin, V. Calvert, H. Juhl, F. Mills, K. Mason, N. Shastri, J. Chik, C. Xu and A. S. Rosenberg (2013). "A novel function of RNAs arising from the long terminal repeat of human endogenous retrovirus 9 in cell cycle arrest." J Virol **87**(1): 25-36.
Yang, J. and V. G. Corces (2011). "Chromatin insulators: a role in nuclear organization and gene expression." Adv Cancer Res **110**: 43-76.
Yang, Z., J. H. Yik, R. Chen, N. He, M. K. Jang, K. Ozato and Q. Zhou (2005). "Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4." Mol Cell **19**(4): 535-545.
Yoder, J. A., C. P. Walsh and T. H. Bestor (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet **13**(8): 335-340.
Yokoyama, A., Z. Wang, J. Wysocka, M. Sanyal, D. J. Aufiero, I. Kitabayashi, W. Herr and M. L. Cleary (2004). "Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression." Mol Cell Biol **24**(13): 5639-5649.
Youdell, M. L., K. O. Kizer, E. Kisseleva-Romanova, S. M. Fuchs, E. Duro, B. D. Strahl and J. Mellor (2008). "Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36." Mol Cell Biol **28**(16): 4915-4926.
Yu, X., X. Zhu, W. Pi, J. Ling, L. Ko, Y. Takeda and D. Tuan (2005). "The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active LTR enhancer complex NF-Y/MZF1/GATA-2." J Biol Chem **280**(42): 35184-35194.
Yudkovsky, N., J. A. Ranish and S. Hahn (2000). "A transcription reinitiation intermediate that is stabilized by activator." Nature **408**(6809): 225-229.
Zentner, G. E., P. J. Tesar and P. C. Scacheri (2011). "Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions." Genome Res **21**(8): 1273-1283.
Zhai, Z., W. Liu, M. Kaur, Y. Luo, J. Domenico, J. M. Samson, Y. G. Shellman, D. A. Norris, C. A. Dinarello, R. A. Spritz and M. Fujita (2017). "NLRP1 promotes tumor

growth by enhancing inflammasome activation and suppressing apoptosis in metastatic melanoma." Oncogene **36**(27): 3820-3830.
Zhang, J. Y., M. Z. Weng, F. B. Song, Y. G. Xu, Q. Liu, J. Y. Wu, J. Qin, T. Jin and J. M. Xu (2016). "Long noncoding RNA AFAP1-AS1 indicates a poor prognosis of hepatocellular carcinoma and promotes cell proliferation and invasion via upregulation of the RhoA/Rac2 signaling." Int J Oncol **48**(4): 1590-1598.
Zhang, T., S. Cooper and N. Brockdorff (2015). "The interplay of histone modifications - writers that read." EMBO Rep **16**(11): 1467-1481.
Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.
Zhao, S., W. P. Fung-Leung, A. Bittner, K. Ngo and X. Liu (2014). "Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells." PLoS One **9**(1): e78644.
Zheng, B., P. Fiumara, Y. V. Li, G. Georgakis, V. Snell, M. Younes, J. N. Vauthey, A. Carbone and A. Younes (2003). "MEK/ERK pathway is aberrantly active in Hodgkin disease: a signaling pathway shared by CD30, CD40, and RANK that regulates cell proliferation and survival." Blood **102**(3): 1019-1027.
Zhu, L. J., C. Gazin, N. D. Lawson, H. Pages, S. M. Lin, D. S. Lapointe and M. R. Green (2010). "ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data." BMC Bioinformatics **11**: 237.