THE REGULATION OF DNA METHYLATION BY A LONG NON-CODING RNA

by

RHIAN LOUISE JONES

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

DOCTOR OF PHILOSOPHY

School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
September 2019

ABSTRACT

Methylation of DNA is an epigenetic mechanism that is crucial in regulating gene expression. Aberrant changes in gene transcription observed in cancers and other diseases, is often attributed to changes in DNA methylation patterns. As a result, understanding the regulation of the DNA methyltransferase (DNMT) family of proteins, which catalyse addition of methyl groups to cytosine residues (5mC), is necessary. Recently, a number of long non-coding RNAs (lncRNAs) have been shown to play a role in regulating activity of DNA methyltransferase-1 or DNMT1. We show that our lncRNA of interest, which is frequently mutated in myeloid leukaemia, also interacts with DNMT1. In an exciting discovery, we have found that a knockout (KO) of this RNA results in DNMT1 re-localising to the cytoplasm of the cell. As a consequence, in KO cells, DNA is significantly hypomethylated and many genes show changes in expression. We also observe that KO cells grow significantly slower and are apoptotic. We have additionally seen changes in 3D nuclear architecture, although it is currently unclear how this occurs. These observations not only suggest a new role for our lncRNA in DNA methylation by regulating subcellular localisation of DNMT1, but also its potential in down-regulating cancer cell growth.

Dedicated to

Mam, Dad, Jay and Nathan

ACKNOWLEDGMENTS

Completion of this project and thesis would not have been possible without a number of important people. First and foremost, I would like to express thanks to my supervisor, Dr Aditi Kanhere, whose guidance, patience and expertise have helped me become the scientist I am today. I am hugely appreciative for the confidence you had in me, when I lacked it in myself. Thank you also to Dr Farhat Khanim for our invaluable discussions and your continual encouragement. I am hugely appreciative to the BALM facility, especially Dr Alessandro Di Maio, for his immense help and support with the confocal microscope and image analysis. Also, to Dr John Halsall for his assistance with histone extractions and other discussions.

The Kanhere lab group, Susanne, Diaa and Sanjana, were a valued part of my PhD journey, providing encouragement and help during the most difficult times. I will truly miss us working together. I must also thank the rest of the 8th floor members for making my experience so enjoyable, and for always being so willing to help or loan reagents. It was a privilege working alongside such a kind and intelligent group of people.

I must also thank my family and friends. To my parents who have always encouraged and supported me in everything I do, to Nathan who is always on the end of the phone during panicked rants of failed experiments, and a special thank you to Jay who has been my complete rock, not only throughout this PhD, but throughout the last 9 years. This thesis would not exist without you. Thank you also to my Tredegar girls for sticking with me, despite me being a rubbish friend and always being busy in the last four years!

A final thank you also to the BBSRC MIBTP for giving me this wonderful opportunity.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

3C	Chromosome conformation capture
5mC	5 methyl-cytosine
AgNOR	Argyrophilic nucleolar organiser region
AML	Acute myeloid leukemia
BAH	Bromo-adjacent homology
BSA	Bovine serum albumin
Cas9	CRISPR associated protein 9
CCDC26	Coiled-coil domain containing 26
cDNA	Complementary DNA
CHX	Cycloheximide
CML	Chronic myeloid leukemia
CpG	5'-Cytosine-phosphate-Guanine-3'
CRISPR	Clustered regularly interspaced short palindromic repeats
Ct	Cycle threshold
DAC	Decitabine (5-Aza-2'-deoxycytidine)
DACOR1	DNMT1-associated colon cancer repressed lncRNA 1
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA Pol	DNA Polymerase
DNase	Deoxyribonuclease
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3A
DSB	Double strand break
ENCODE	Encyclopaedia of DNA Elements
EZH2	Enhancer of Zeste 2
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
Firre	Functional intergenic repeating RNA element
FISH	Fluorescence *in situ* hybridisation
gDNA	Genomic DNA
GSDMC	Gasdermin C
HDAC	Histone deacetylase
HDM	Histone demethylase
hnRNPU	Heterogenous nuclear ribonucleoprotein U
HOTAIR	HOX transcript antisense RNA
HSAN1E	Hereditary sensory and autonomic neuropathy type 1E
IDP	Intrinsically disordered proteins
IF	Immunofluorescence
IP	Immunoprecipitation
KD	Knock down
KO	Knockout
LB	Lennox broth
LINEs	Long interspersed nuclear elements
lncRNA	Long non-coding RNA
LSD1	Lysine demethylase 1
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript
miRNA	MicroRNA
mRNA	Messenger RNA

NCBI	National Centre for Biotechnology Information
ncRNA	Non-coding RNA
NEAT1	Nuclear Enriched Abundant Transcript 1
NLS	Nuclear localisation signal
ORF	Open reading frame
O/x	Overexpression
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PI	Propidium iodide
piRNA	PIWI-interacting RNA
polyA	Polyadenylation
PRC1/2	Polycomb repressive complex 1 / 2
PTM	Post-translational modification
PVT1	Plasmacytoma Variant Translocation 1
qRT-PCR	Quantitative real time polymerase chain reaction
RAM	Retinoic Acid modulator
RefSeq	NCBI Reference Sequence Database
RFTS	Replication foci-targeting sequence
RIP	RNA immunoprecipitation
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNA Pol II	RNA Polymerase II
RNA-seq	RNA sequencing
RNase	Ribonuclease
RNP	Ribonucleoprotein
RRBS	Reduced representation bisulfite sequencing
rRNA	Ribosomal RNA
SAHA	Suberanilohydroxamic Acid
SAM	S-adenosylmethionine
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
siRNA	Small interfering RNA
sncRNA	Small non-coding RNA
snoRNA	Small nucleolar RNA
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoproteins
TAD	Topologically-associated domain
TBRS	Tatton-Brown-Rahman syndrome
TBS	Tris buffered saline
TBS-T	Tris buffered saline with Tween
tRNA	Transfer RNA
TSA	Trichostatin A
TSS	Transcription start site
TF	Transcription factor
UHRF1	Ubiquitin-like with PHD and RING finger domains 1
VPA	Valproic Acid
WT	Wild type
XCI	X-chromosome inactivation
Xi	Inactive X-chromosome
XIC	X inactivation centre
Xist	X-inactive specific transcript

CHAPTER 1

INTRODUCTION

1.1 Non-Coding RNAs

1.1.1 Introduction to Non-Coding RNAs

Throughout the last decade, the substantial development of high-throughput technology has helped reveal the sheer complexity of the mammalian genome. Less than 2% of the human genome is protein coding, however, transcription is far more extensive, with RNA-seq data demonstrating that >80% of the genome is transcribed (Hangauer, Vaughn, and McManus 2013). It was previously believed that most of these transcriptional events constituted irrelevant “transcriptional noise”, however, following the success in sequencing whole eukaryotic genomes, the naivety in the level of our understanding of the transcriptome is becoming clearer (Hangauer, Vaughn, and McManus 2013).

For years the understanding of RNA was primarily based on stably expressed protein-coding messenger RNAs (mRNAs) and housekeeping non-coding RNAs (ncRNAs) that are not translated to protein, but are largely involved in the process of translation, such as ribosomal RNA (rRNA), small nuclear and small nucleolar RNA (snRNA and snoRNA) and transfer RNA (tRNA). The early 2000’s however, led to the discovery that many non-protein-coding transcripts, previously disregarded, were stable enough to be cloned (Rinn et al. 2007). Evidence began to emerge of transcripts, formerly regarded as waste products, demonstrating

tissue-specific expression and localisation patterns, as well as unexpectedly long life spans (Clement et al. 1999; Clement, Maiti, and Wilkinson 2001; Lv et al. 2015). A key study in the discovery and appreciation of functional ncRNAs annotated >60,000 complementary DNA sequences (cDNA) in mice. Newly discovered ncRNAs comprised a far greater proportion of the murine transcriptome than expected, accounting for over a third of the analysed transcripts. This helped consolidate non-coding transcripts as one of the major components of the genome (Okazaki et al. 2002).

The theory that ncRNAs are mere transcriptional noise has dwindled somewhat in recent years, as the true intricacy of our transcriptome has been unveiled. It has become clear that ncRNAs, other than housekeeping RNAs are functional in their own right, without being directly connected to protein translation. Large numbers of functional transcripts appear to be transcribed from intergenic regions and DNA that was previously thought to be silent (Morozova and Marra 2008; Derrien et al. 2012). This change has been driven by the development of modern technologies including RNA deep sequencing, chromatin immunoprecipitation with next-generation sequencing (ChIP-seq) and high-resolution techniques.

1.1.2 Non-coding RNAs: Functional Components or Genetic Waste?

The suggestion that non-coding transcripts are functional has generated much debate. One of the reasons behind this is that a number of observations indicate that the generation of non-coding transcripts is coincidental (Struhl 2007). For instance, as compared to protein-coding genes, ncRNAs display poor evolutionary

conservation (Pollard et al. 2006), and are expressed at much lower levels. An expression analysis of 15 cell lines showed that just 25% of protein-coding RNAs are present at ≤1 copy/cell, whilst a staggering 80% of long ncRNAs (lncRNAs) are present at this frequency (Djebali et al. 2012). This argues against functionality of ncRNAs.

However, several observations suggest that ncRNAs are functional; for example, lncRNAs exhibit cellular specificity that is unexpected of mere noise. The same study that demonstrated low expression levels of ncRNAs, also showed that whilst only 7% of protein-coding gene expression is specific to a cell-line, almost 30% of lncRNAs demonstrate cell-specificity (Djebali et al. 2012). In addition, many non-coding transcripts also display disease-specific expression patterns. In 2018, an entire, comprehensive database was generated just for non-coding RNAs associated with heart disease (Wang, Wang, et al. 2018). There is also the sheer number of lncRNAs to consider; a 2015 study pooled >7000 RNA-seq libraries and of the 91,000 genes analysed, >58,000 (~70%) were catalogued as lncRNAs (Iyer et al. 2015). A more recent study generated a high-confidence atlas of ~30,000 lncRNA genes with accurately determined 5' ends. Of these, they argued that at least two thirds are potentially functional, based on the increased conservation of their exonic and transcription initiation regions, and their implications in genome-wide association studies (GWAS) traits and expression quantitative trait loci (eQTL) (Hon et al. 2017).

With regards to conservation, it is important to note that whilst ncRNA sequences are far less conserved than protein-coding genes, they still demonstrate a higher level of conservation than ancestral repeat regions (Orom et al. 2010). Evidence suggests that for ncRNAs, conservation of sequences outside the gene body, such as regulatory regions, is more important. Studies have shown that lncRNA promoter regions are highly conserved and generally demonstrate accumulation of relatively few mutations throughout evolution. Remarkably, the percentage of conserved sequences in lncRNA promoters approximately matches the percentage of conserved exonic sequences in protein-coding genes (Ponjavic, Ponting, and Lunter 2007). The fact that ncRNAs do not encode specific amino acid sequences might reflect the reason that they do not require such strict sequence conservation. Additionally, in some cases the act of transcription at certain regions might be more important than the end-product transcript. Initiation of a transcriptional event at a particular genomic region might be required to induce a desirable effect, such as altering the chromatin architecture or driving the accumulation of transcriptional machinery to that particular locus (Engreitz et al. 2016). For example, intergenic transcription of a ncRNA is required to drive the formation of an open chromatin environment at the β globin locus, thereby allowing access for transcription machinery and subsequently transcription of the gene (Gribnau et al. 2000).

It has been shown that several ncRNAs localise at their site of transcription. This means that they can exert their function immediately after being transcribed and thus perform their roles at highly effective and efficient concentrations (Zhang et al. 2013; Takayama et al. 2013). This might explain why many ncRNAs are expressed

at such low levels. For example, ncRNAs transcribed from genomic enhancer regions (eRNAs), can accumulate at their transcription site and behave as scaffolds onto which proteins such as Mediator can gather and subsequently drive changes in chromosomal architecture, organisation and gene expression (Lai et al. 2013).

Evidence is rapidly emerging of ncRNAs playing key roles in all aspects of cellular function and I will discuss this in detail later in this chapter. At present, there is a large discrepancy between the number of identified ncRNAs and the number that have been functionally characterised (Quek et al. 2015; Derrien et al. 2012). This is largely due to our current inability to interpret function solely from RNA sequence in the way that protein function can be determined, as well as the relative unavailability of methods to study them. Ultimately, it is likely that both functional and non-functional ncRNAs exist, and further work will be required to distinguish the functional from the non-functional.

1.1.3 Classification of Non-Coding RNAs

Non-coding RNAs are either regarded as housekeeping ncRNAs or regulatory ncRNAs. At present there is no definitive method of classifying the regulatory ncRNAs. Multiple categories have been proposed, however, ncRNAs can often be placed into several of these. A common overarching divider used however, is separation according to size. Long ncRNAs (lncRNAs) are those defined as having a length of >200bp, whilst small ncRNAs (sncRNAs) are those with a length of <200bp (Kung, Colognori, and Lee 2013). Numerous sub-categories then exist within these groups, examples of which can be found in Figure 1.1 and Table 1.1.

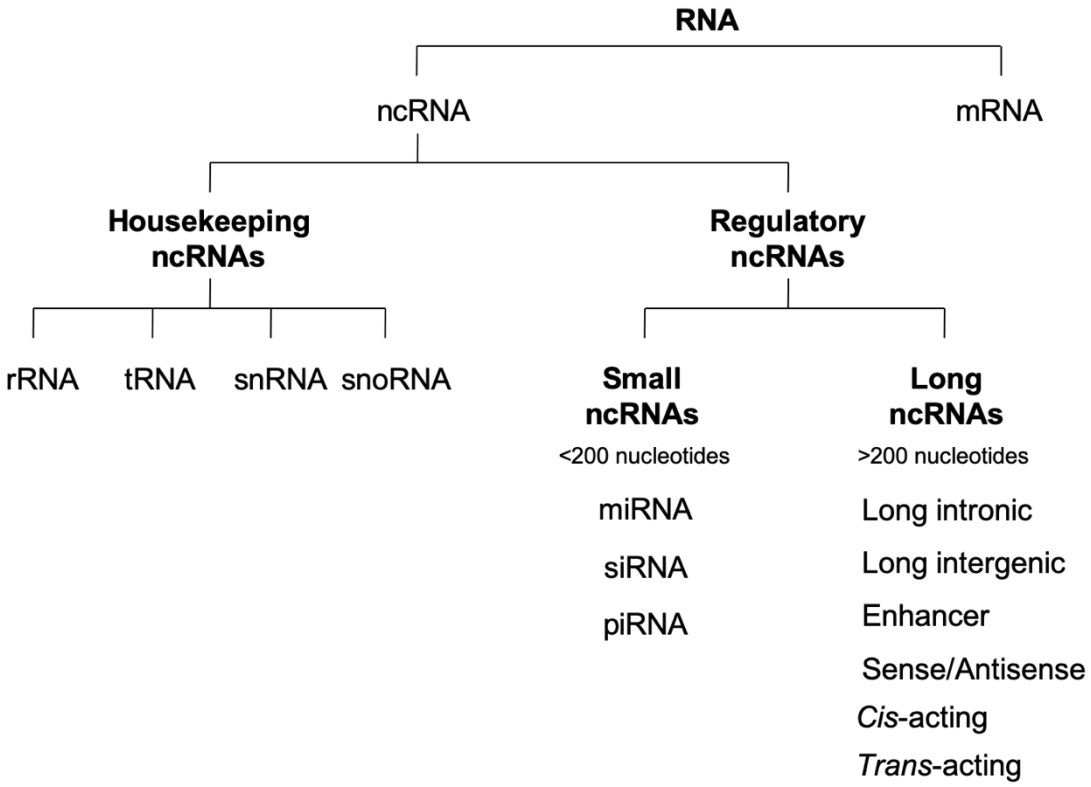


Figure 1.1. Classification of RNA. RNA is classified into two main categories: non-coding RNA (ncRNA) and messenger RNA (mRNA). ncRNA is comprised of regulatory ncRNAs and housekeeping ncRNAs which are generally required for cell viability; ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA). The former is further divided according to size into small ncRNAs and long ncRNAs. The three main types of small ncRNAs are microRNA (miRNA), small interfering RNA (siRNA) and PIWI-interacting RNA (piRNA). Long ncRNAs can be categorised in multiple ways, e.g. according to site of transcription (intronic, intergenic and enhancer RNA), direction of transcription (sense and anti-sense RNA) or mechanism of action (*cis*-acting or *trans*-acting RNA).

1.1.3.1 Housekeeping ncRNAs

Housekeeping ncRNAs differ from regulatory ncRNAs as they are required for normal cellular functionality and viability, which is reflected in their constitutive expression. This category includes transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) (Morey and Avner 2004).

rRNA is required to generate functional ribosomes by associating with ribosomal proteins. At the nucleolus, rRNA is transcribed as a large 45S precursor transcript, which is then cleaved to yield 5.8S, 18S and 28S RNA species (Henras et al. 2015). Small nucleolar RNAs are typically <300 nucleotides in size and their predominant function involves post-transcriptionally modifying and processing rRNA. snoRNAs can be categorised according to their conserved RNA sequences and the specific RNA modification they are associated with; H/ACA snoRNAs target pseudouridylation sites whilst box C/D snoRNAs aid rRNA 2'-O-methylation (Kiss 2001).

In addition to rRNA, tRNA is also a critical component for protein biosynthesis. During translation, amino acid chains are assembled using mRNA codons as templates. tRNAs act as adaptor molecules; they contain anti-codon sequences that can recognise and bind specific mRNA codons, and also carry amino acids associated with that specific codon. As the codons bind the anti-codons, the enzymatic component of the ribosome catalyses peptide bond formation between adjacent amino acids attached to tRNA molecules (Kirchner and Ignatova 2015).

snRNAs on the other hand are not associated with translation, but rather are primarily involved in processing precursor-mRNA (pre-mRNA). They typically form small nuclear ribonucleoproteins (snRNPs) by binding specific proteins, and help splice introns out of newly synthesized transcripts (Valadkhan 2005).

1.1.3.2 Regulatory ncRNAs

1.1.3.2.1 Small Non-Coding RNAs (sncRNAs)

Unlike housekeeping ncRNAs, regulatory ncRNAs generally tend to be expressed as a consequence of environmental stimuli, or at specific stages of the cell cycle and development (Morey and Avner 2004). Admittedly, it is sncRNAs, in both animals and plants, that have primarily governed the literature so far. One of the best-studied categories of sncRNAs are microRNAs (miRNAs). miRNAs were first discovered in *C. elegans* in the early 1990s (Lee, Feinbaum, and Ambros 1993), and since then almost 3000 human miRNAs have been recorded, with thousands more potential candidates predicted (Griffiths-Jones et al. 2008; Backes et al. 2018). The main role of miRNAs is in gene silencing, by targeting degradation of specific mRNA transcripts. Following transcription and processing into mature miRNAs, they join the RNA-induced silencing complex (RISC), where they associate with the Argonaute protein (ARGO2). They then act as guides for the complex and complementary base-pair with the 3'-untranslated regions (3'UTRs) of their complementary mRNAs. The target mRNA is subsequently cleaved by ARGO2 and degraded (Krol, Loedige, and Filipowicz 2010).

One of the major focuses of miRNA research has been how they are implicated in cancer. Since an association between the two was first reported in the early 2000's, the precise functions of miRNAs, as well as investigation into their potential use as prognostic biomarkers, has received significant attention (Tan et al. 2018). For example, *miR-15* and *miR-16* have been implicated in chronic lymphocytic leukemia (CLL). These miRNAs are down-regulated in CLL, which results in enhanced expression of the *miR-15/miR-16* target, B-cell lymphoma 2 (BCL2). BCL2 inhibits apoptosis, and consequently contributes to carcinogenesis (Calin et al. 2002).

In addition to miRNAs, other classes of sncRNAs include small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs), which possess similar roles in RNA silencing and examples of which are highlighted in Table 1.1 (Choudhuri 2010).

1.1.3.2.2 Long Non-coding RNAs (LncRNAs)

LncRNAs are generally subclassified to reflect the genomic region from which the RNA is transcribed, or the genomic region at which it functions. For example, intergenic, intronic and enhancer lncRNAs are transcribed from regions between genes, within introns or within enhancers respectively, sense and anti-sense lncRNAs are transcribed from sense and anti-sense DNA strands respectively, and *cis*-acting and *trans*-acting lncRNAs regulate either neighbouring genes *in cis*, or distant genes *in trans* respectively (Table 1.1).

Interestingly, lncRNAs share some characteristics with mRNAs. First, the genes of both possess some similar epigenetic chromatin marks. Early analyses detected

comparable patterns of histone modifications, including tri-methylation of Histone 3 at Lysine 4 (H3K4me3) at promoter sites, and tri-methylation of Histone 3 at Lysine 36 (H3K36me3) along the actively transcribed DNA region (Guttman et al. 2009). Similar to protein coding genes, lncRNA transcription start sites have also demonstrated increased levels of various active histone marks, including acetylation of Histone 3 at Lysines 9 and 27 (H3K9ac and H3K27ac) (Derrien et al. 2012). More recent studies however, have identified certain distinct differences between chromatin signatures. For example, ChIP-seq experiments have shown that long intergenic ncRNAs (lincRNAs) and mRNAs with similar expression levels, possess strikingly different levels of tri-methylated Histone 3 at Lysine 9 (H3K9me3). This mark is typically associated with transcriptional repression when deposited at mRNA genes; however, it was found to be enriched at actively transcribed lincRNA promoters (Mele et al. 2017). A subset of lncRNAs also demonstrate methylation of Histone 3 at Lysine 4 (H3K4me1), which is typically associated with active enhancers, indicating enhancer-lncRNA gene overlap (De Santa et al. 2010). Similar ChIP data has also demonstrated that despite the presence of many conserved transcription factor binding sites at lincRNA promoters, they are generally bound by fewer transcription factors than mRNAs. It is suggested that this is due to many lincRNAs only requiring activation in response to particular stimuli or in certain tissues (Mele et al. 2017).

At present, transcriptional regulation of lncRNAs is not well understood, thus, several studies have begun comparing mRNA and ncRNA synthesis. In some ways, the process appears similar; both tend to be transcribed by RNA Pol II (Derrien et

al. 2012), and similar to protein coding genes, RNA Pol II promoter-proximal pausing, regulated by transcription factors, TRIM28 and pTEFb, is prevalent in a subset of lncRNA genes (Bunch et al. 2016; Bunch et al. 2019). However, distinctive patterns of lincRNA nascent transcription have also been reported. Differentially phosphorylated forms of the C-terminal domain (CTD) of RNA Pol II were analysed at different stages of transcription and processing using mammalian native elongating transcript sequencing (mNET-seq). Interestingly, phosphorylated Threonine 4 of the CTD was localised across entire lincRNA genes. This was in contrast to protein coding genes where it was generally found at transcription termination regions. This suggests that many lincRNA transcripts may terminate prematurely at various different positions (Schlackow et al. 2017).

Another important part of RNA synthesis is processing. Typically, mRNAs are co-transcriptionally spliced and their 3' ends, cleaved and polyadenylated. Many lncRNAs also demonstrate similar splicing and polyadenylation, although this is not true for all lncRNAs (Guttman et al. 2009). This was demonstrated by mNET-seq profiles which showed that these two events coincide with the presence of RNA Pol II CTD phosphorylation at Serine 5 and Serine 2 respectively for mRNAs. Such CTD profiles however, were absent for transcription of multiple lincRNAs (Schlackow et al. 2017).

3' poly(A) tails are critical for RNA stability and functionality (Colgan and Manley 1997). For lncRNAs lacking poly(A) tails, various types of alternative, non-canonical 3' end processing can occur. In some cases, 3' triple helical structures have been

observed in mature transcripts. This has been reported for lncRNAs, *MALAT1* and *MENβ*. Immediately following transcription, their 3' ends adopt a structure similar to tRNAs. As a result, they are not processed by canonical cleavage machinery, but are instead cleaved by RNases P and Z. This generates the mature 3' end, which subsequently forms a stable, triple helical structure. This structure serves to protect the RNA from exonuclease digestion (Zong et al. 2016).

5'-capping is another transcriptional process that occurs during mRNA transcription. The 5' end of mRNAs are typically "capped" with a methylated guanosine nucleotide, commonly referred to as m^7G. Its' functions include protection against exonuclease digestion, and regulation of translation and nuclear export (Ramanathan, Robb, and Chan 2016). Many lncRNAs are also 5' capped and can consequently be detected by Cap Analysis of Gene Expression (CAGE), however this is not true for all (Shiraki et al. 2003; Guttman et al. 2009). In the last decade another new class of lncRNAs has been identified, and named sno-lncRNAs. These sno-lncRNAs lack both a 5' cap and 3' poly(A) tail, and are instead capped with sno-RNPs at their 5' and 3' ends. Such transcripts are generated when two intronic snoRNA genes are transcribed, along with the sequence between them, with studies observing both Box C/D and Box H/ACA snoRNAs being transcribed in this way (Yin et al. 2012). Similar 5' snoRNP 3' poly(A) transcripts, commonly referred to as SPAs, have also been reported. These lncRNAs possess a snoRNP at their 5' end and are polyadenylated at their 3' end, and arise as a result of polycistronic transcript processing (Wu et al. 2016).

In addition to typical linear structures, lncRNAs lacking 5' capping and 3' polyadenylation have also been shown to adopt closed, circular structures. These circRNAs are either generated from excised intron lariats or back-spliced exons, and are consequently resistant to digestion by exonuclease enzymes, as they effectively lose their 5' and 3' ends (Zhang et al. 2013; Ashwal-Fluss et al. 2014; Wu, Yang, and Chen 2017).

Similar to mRNAs, many lncRNAs are also spliced. It is believed that lncRNA splicing regulation is similar to mRNAs, due to the presence of canonical GT/AG splice site sequences. Mele et al. identified a large number of spliced lincRNAs possessing these conserved dinucleotides (Mele et al. 2017). Some lncRNAs have additionally demonstrated alternative splicing (Krchnakova et al. 2019; Soreq et al. 2014). *PCBP1-AS1*, an antisense lncRNA transcribed from the region which also encodes the poly(rC)- binding protein 1 (PCBP1), has 40 annotated splice-variants (Derrien et al. 2012; Ziegler and Kretz 2017). However, several studies have reported that lncRNAs are spliced less efficiently than mRNAs (Tilgner et al. 2012; Mele et al. 2017; Schlackow et al. 2017). Many demonstrate weaker binding to splice factor U2AF65, and alternative compositions of polypyrimidine tracts, resulting in reduced spliceosome assembly (Mele et al. 2017). It was recently proposed that splicing efficiency for lncRNAs is more dependent on the strength of their 5' splice sites, and the proportion of thymidine in the polypyrimidine tract than mRNAs, as they interact less with nuclear splicing factors (Krchnakova et al. 2019).

LncRNAs possess a number of additional features that are distinct from mRNAs. As previously mentioned, compared to protein-coding genes, lncRNA sequences are generally, less well conserved (Pollard et al. 2006). Also, as compared to mRNAs, lncRNAs are typically shorter in length and are preferentially two-exon transcripts (Derrien et al. 2012; Huang et al. 2018; Liu et al. 2018). Despite both exons and introns of lncRNAs being larger on average compared to their protein-coding counterparts, this “di-exonic” tendency results in shorter mature transcripts (Derrien et al. 2012). It is important to note however, that not all lncRNAs strictly possess the characteristics outlined above (Burd et al. 2010; Lyle et al. 2000).

Table 1.1 Characteristics of classes of non-coding RNAs (Choudhuri 2010; Ma, Bajic, and Zhang 2013)

sncRNAs (<200 nucleotides)	
Class	**Description**
Small interfering RNAs (siRNAs)	Length of ~21–24 nucleotides, generated from double-stranded RNA. Mediates post-transcriptional silencing of genes via RNA interference (RNAi) by directing the RNA-induced silencing complex (RISC).
microRNAs (miRNAs)	Length of ~22 nucleotides, derived from hairpin loop structures within RNA transcripts. The resulting mature miRNA forms a primary component of RISC to target gene silencing. E.g. Let7 miRNAs
PIWI-interacting RNAs (piRNAs)	Length of ~26-30 nucleotides. They bind proteins belonging to the Argonaute family, known as Piwi proteins. This family of proteins play key roles in RNA silencing; The RNAs and proteins form structures that predominantly act to silence transposable elements during development of the gametes to protect the germline cells. E.g. prenatal piRNAs – appear during spermatogenesis.
lncRNAs (>200 nucleotides)	
Long intergenic RNAs (lincRNAs)	Transcribed from genomic regions that are intergenic (i.e. between genes). E.g. SRA1
Long intronic RNAs	Transcribed from genomic sequences encompassing intronic regions of protein-coding genes. E.g. COLDAIR
Enhancer RNAs	Transcribed from enhancer regions of DNA E.g. HOTTIP
Sense RNAs	Transcribed from the sense strand of a protein-coding gene (i.e. the strand possessing the same sequence as the corresponding mRNA, running 5’ to 3’). E.g. AK011429
Antisense RNAs	Transcribed from the antisense strand of a protein-coding gene (i.e. the strand which is used as the template during transcription, running 3’ to 5’) E.g. ANRIL
Cis-acting RNAs	RNAs involved in regulation of neighbouring genes in the immediate vicinity. E.g. Xist
Trans-acting RNAs	RNAs that are able to influence the regulation of genomically distant genes. E.g. HOTAIR

1.1.4 Functions of lncRNAs: Transcriptional Regulation

1.1.4.1 Molecular Mechanisms of lncRNAs: Signals, Decoys, Guides and Scaffolds

Since their discovery, a vast array of functions related to gene expression regulation has been associated with lncRNAs, including transcriptional, post-transcriptional, translational, post-translational and epigenetic regulation. Investigations of individual lncRNAs have identified functional roles in almost all aspects of cell biology, including cell growth and differentiation, and extending to metabolism and cellular stress response pathways. At the transcriptional level, lncRNA-mediated gene regulation can involve chromatin and chromosome remodelling, as well as direct regulation of transcriptional machinery; these are discussed in more detail later in this chapter (He, Luo, and Mo 2019; Wang and Chang 2011). The molecular mechanisms by which lncRNAs perform their roles are broadly categorised into four types, i.e. signals, decoys, guides and/or scaffolds, and are discussed in greater detail below (Figure 1.2) (Wang and Chang 2011).

The controlled expression of lncRNAs at specific locations, time points, developmental phases and cell cycle stages means that they can both respond to various stimuli and subsequently behave as signals themselves (Figure 1.2A). In many instances, changes in the lncRNA expression or even the act of transcription itself, is sufficient to activate downstream effects (Wang and Chang 2011). An excellent example of a lncRNA that applies this kind of regulatory mechanism is *Air*, which is transcribed from a genomically imprinted gene (Nagano et al. 2008). Genomic imprinting is a process whereby genes are expressed mono-allelically

according to parental origin (McGrath and Solter 1984). Transcription of *Air* is required to induce repression of a number of *cis*-linked genes in an allele-specific manner. Following transcription in mice, *Air* assembles at the promoter region of the solute carrier family 22-member 3 (SLC22A3) gene, recruits a histone modifying enzyme which leads to silencing of the associated genes (Nagano et al. 2008).

In many instances lncRNAs can act as molecular decoys by binding and luring components away from a particular site or target (Figure 1.2B). Generally, lncRNAs involved in this kind of regulation do not employ additional functions other than that of masking regulatory factors, which can have both positive or negative effects on transcription. For example, *metastasis associated lung adenocarcinoma transcript 1* (*MALAT1*) can allure proteins required for mRNA splicing to nuclear speckles (Tripathi et al. 2010). Additionally, in a disease setting, increased *MALAT1* can bind the miRNA, *miR-217* and prevent it from being transported to the cytoplasm in pancreatic ductal adenocarcinoma. In doing so the mitogen-activated-protein-kinase (MAP-Kinase) signalling pathway is enhanced due to prevention of *miR-217* inhibiting the K-RAS protein, ultimately resulting in increased tumour growth and proliferation (Liu et al. 2017).

As the name suggests, guide lncRNAs bind and direct complexes to specific target sites, either *in cis* or *in trans* (Figure 1.2C). This often results in changes to the chromatin landscape and consequently gene expression (Wang and Chang 2011). A key epigenetic regulatory complex that has demonstrated interactions with over 200 lncRNAs is the Polycomb repressive complex 2 (PRC2) (Khalil et al. 2009). This complex functions by catalysing tri-methylation of Histone 3 at Lysine 27

(H3K27me3), a repressive histone modification, at the developmental genes (Cao et al. 2002). LncRNA *Kcnq1ot1* recruits and guides PRC2 as well as G9a, a Histone 3 Lysine 9 methyltransferase enzyme, to establish transcriptional silencing *in cis* (Pandey et al. 2008). *HOX transcript antisense RNA (HOTAIR*) has also been reported to guide PRC2 to the HOXD locus to repress its expression *in trans*. *HOTAIR* was the first ncRNA observed to impact transcription epigenetically on a chromosome different to the one from which it is transcribed (Rinn et al. 2007). Despite the ambiguity surrounding the precise mechanism by which lncRNAs can target complexes to specific genomic loci, the importance of lncRNAs as transcriptional regulators, is clear.

The final mechanism involves behaviour of lncRNAs as scaffolds (Figure 1.2D) onto which multiple molecular components are assembled. This can either allow components to collectively carry out a particular function, or to structurally stabilise complexes (Wang and Chang 2011). *Nuclear Enriched Abundant Transcript 1* (*NEAT1*) for example behaves as a scaffold onto which proteins such as PSP1 and p54 are assembled for the formation of paraspeckles (Fox, Bond, and Lamond 2005; Mao et al. 2011).

As further research continually emerges, it is clear that many lncRNAs fall into several of these categories. Here we will review a variety of different lncRNAs and their functional roles, all of which can be associated with one or more of the four functional mechanisms described above.

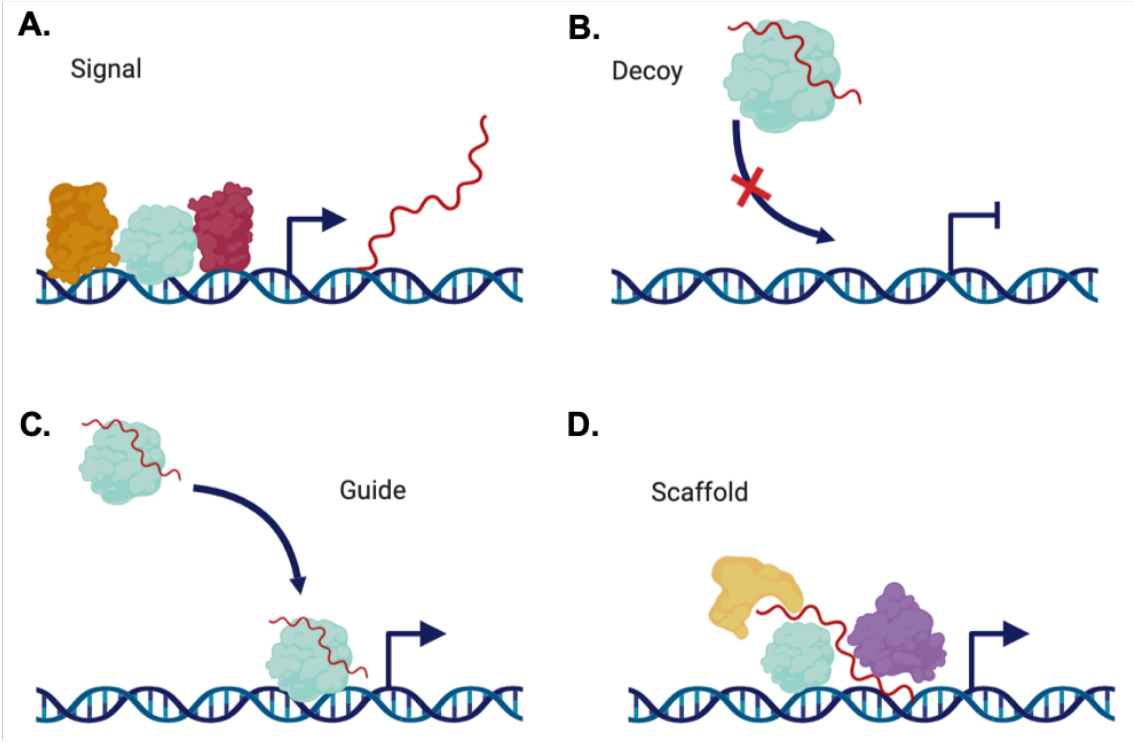


Figure 1.2. Molecular Mechanisms of lncRNAs. According to Wang and Chang, (2011), functional lncRNAs can adopt four different mechanisms of action, behaving as signals, decoys, guides or scaffolds. **A.** Signal lncRNAs can respond to a stimulus and elicit an appropriate response. **B.** Decoys can prevent components binding a particular target, **C.** whilst guides have an opposing effect and can direct components to a specific target. **D.** Scaffolds provide an anchor upon which multiple components can assemble (Wang and Chang 2011).

1.1.4.2 LncRNAs and Chromatin Remodelling

A key mechanism in regulation of gene expression at the transcriptional level, involves controlling the chromatin landscape. Heterochromatic and euchromatic states generally correlate with repressed and activated transcription respectively (Wang, Jia, and Jia 2016; Wang and Chang 2011). A number of lncRNAs regulate transcription by controlling this chromatin landscape.

Arguably, one of the most well-characterised lncRNAs that regulates transcription in this way, is *X-inactive specific transcript* (*Xist*), whose role is central to X-chromosome inactivation (XCI). This process aims to achieve dosage compensation, a phenomenon first hypothesised in the early 1960's (Lyon 1961). It involves suppression of one of the two X chromosomes in mammalian females (XX) to equalise gene expression with males (XY). Willard and colleagues were the first to identify the XIST gene in humans, an exciting finding that promised to provide more insight into X-inactivation (Brown et al. 1991). Observations that the gene lacked any credible open reading frame indicated that the 17kb *Xist* transcript itself could be important (Brown et al. 1992).

XCI can be considered in three key stages: Localisation of *Xist* to the inactive X chromosome (Xi), gene repression and finally preservation of an inactive state. Of course, the process requires a pathway whereby a single X chromosome should be specifically targeted among all other chromosomes found within the nucleus of a cell. Consequently, an *in cis* mechanism is suggested for *Xist*, which is transcribed from the Xi. The apparent stability of the RNA in combination with its distinctive

spatial organisation indicates a direct role in silencing (Figure 1.3) (Clemson et al. 1996; Pintacuda, Young, and Cerase 2017; Sahakyan, Yang, and Plath 2018)

Xist is transcribed from the XIST locus, which lies within a region known as the X inactivation centre (XIC). The RNA then localises to the Xi from which it is transcribed, by means which are understood only partially (Engreitz et al. 2013). It has been suggested that the presence of the SAF-A scaffold protein and the existence of several binding sequences within *Xist* help to mediate the coating of Xi by the lncRNA along its length (Wutz, Rasmussen, and Jaenisch 2002; Hasegawa et al. 2010; Pullirsch et al. 2010; Lyon 1998).

Localisation of *Xist* to Xi is closely followed by a significant reduction in transcriptional machinery surrounding the chromosome. Studies have demonstrated an absence of activating histone modifications and RNA around the Xi during this time (Chu et al. 2015). Next, polycomb repressive complexes (PRCs) accumulate and function in altering the chromatin state of the Xi; whilst PRC1 ubiquitylates histone H2A at Lysine 119 (H2AK119ub1), PRC2 tri-methylates H3K27 (H3K27me3) across the entirety of the chromosome (Plath et al. 2003; Schoeftner et al. 2006). Interestingly, a factor that is essential for this silencing initiation step and gene repression, is the repeated Adenine sequence at 5' *Xist*. A loss of this sequence disrupts silencing initiation and blocks repression of gene expression (Wutz, Rasmussen, and Jaenisch 2002; Patil et al. 2016). Interestingly, generation of the large majority of repressive Xi chromatin signatures are dependent on *Xist*, and do not arise in its absence. For example, without *Xist*, the PRC1/2

complexes are unable to ubiquitylate and methylate the chromatin, and several associated proteins are unable to bind the chromosome (Plath et al. 2003; Kohlmaier et al. 2004; Schoeftner et al. 2006). This emphasises the significance of the roles that lncRNAs can play with regards to transcriptional regulation.

The final stage of X-chromosome silencing requires gene repression to be maintained throughout the life of the female. This is achieved by the distinctive epigenetic landscape of the Xi in female somatic cells, which contrasts the active X chromosome. This characteristic epigenetic arrangement maintains a repressive state without the need for *Xist* to be constitutively expressed (Csankovszki, Nagy, and Jaenisch 2001; Boggs et al. 2002; McHugh et al. 2015). The lncRNA, antisense counterpart to *Xist,* known as *Tsix*, is also essential to the process. Upon X-inactivation, bi-allelic *Tsix* expression is altered to become mono-allelic from the active X chromosome. It functions as a *cis*- acting repressor of *Xist*, thus preventing silencing of both chromosomes (Stavropoulos, Lu, and Lee 2001)

Numerous other lncRNAs also function in chromatin remodelling. For example, *Antisense Non-coding RNA in the INK4 Locus* (*ANRIL*), is a lncRNA that regulates gene expression largely via regulation of histone modifications through interactions with PRC1/2. Functioning *in cis*, *ANRIL* recruits PRC1 to the CDKN2A gene, and PRC2 to the CDKN2B gene, where Histone 3, Lysine 27 is tri-methylated (H3K27me3), subsequently resulting in gene silencing (Yap et al. 2010; Kotake et al. 2011). Similar mechanisms of recruiting PRCs are also employed by *ANRIL* to target genes, such as CARD8 and E2F1, *in trans* (Sato et al. 2010; Bai et al. 2014).

The *HOX Transcript Antisense RNA* (*HOTAIR*) was one of the first lncRNAs to influence gene expression *in trans* and also to recruit PRC2 (Rinn et al. 2007). *HOTAIR* additionally interacts with Lysine Demethylase 1 (LSD1), which catalyses removal of methyl groups from histones (Tsai et al. 2010). These interactions mediate addition of the H3K27me3 mark and demethylation of activating methylated Histone 3 Lysine 4 marks, across the HOXD gene on chromosome 2, resulting in gene repression (Rinn et al. 2007; Tsai et al. 2010)

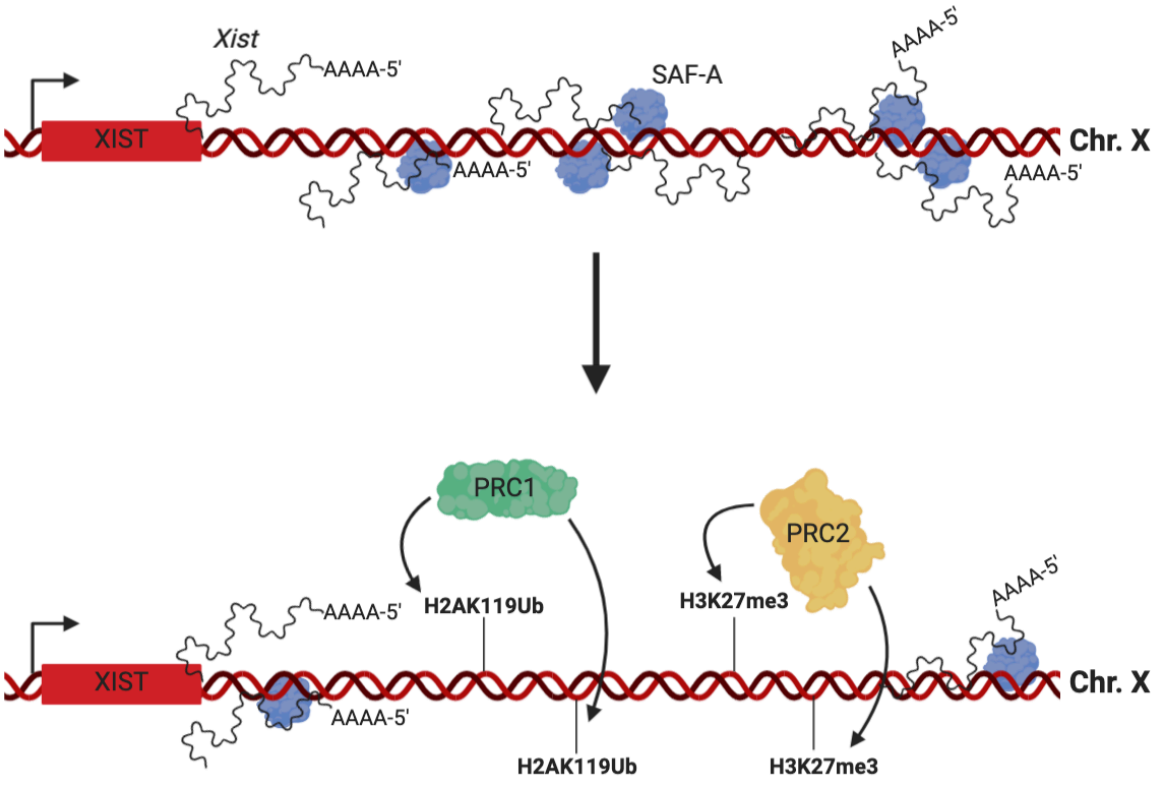


Figure 1.3. LncRNA *Xist* Regulates X-Chromosome Inactivation. LncRNA *Xist* is transcribed from and coats the inactive X chromosome (Xi) along its length, aided by the SAF-A scaffold protein. This leads to the accumulation of polycomb repressive complexes PRC1 and PRC2 which alter the chromatin state of the Xi. PRC1 and PRC2 catalyse addition of the H2AK119Ub and H3K27me3 repressive epigenetic modifications.

1.1.4.3 LncRNAs and Chromatin Looping

In addition to regulating local chromatin architecture, lncRNAs have also demonstrated the ability to influence chromatin organisation on a larger scale. Since the 1980's it has been recognized that RNA can play a structural role in the nucleus. Bouteille and colleagues were amongst the first to demonstrate that the nuclear matrix disintegrates upon nuclear RNA digestion (Bouvier et al. 1985). Since then, substantial evidence has emerged demonstrating a role for lncRNAs in regulating transcription by influencing the 3D organisation of chromosomes within the nucleus (Quinodoz and Guttman 2014)

One good example of this is *functional intergenic repeating RNA element* (*Firre*), a nuclear-localised lincRNA that escapes XCI and is one of very few transcripts expressed from the inactive X chromosome (Yang et al. 2010). Previous lncRNA screenings demonstrated that *Firre* is important in the regulation of adipogenesis (Sun et al. 2013). It was recently suggested that this occurs by *Firre*-mediated co-regulation of adipogenesis genes through organisation of chromosome positioning (Figure 1.4). A ~150bp repeat sequence within the transcript enables *Firre* to bind the heterogeneous nuclear ribonucleoprotein U (hnRNPU), a component of the nuclear matrix network. Both *Firre* and hnRNPU were proven essential for a clustered trans-chromosomal interaction to occur between the Firre locus and four *trans*-sites on chromosomes 2, 9, 15 and 17, containing genes associated with adipogenesis (Hacisuleyman et al. 2014; Maass et al. 2018). *Firre* has also demonstrated a role in XCI, by tethering the Xi chromosome to the nucleolus and preserving some of its repressive epigenetic marks (Yang et al. 2015).

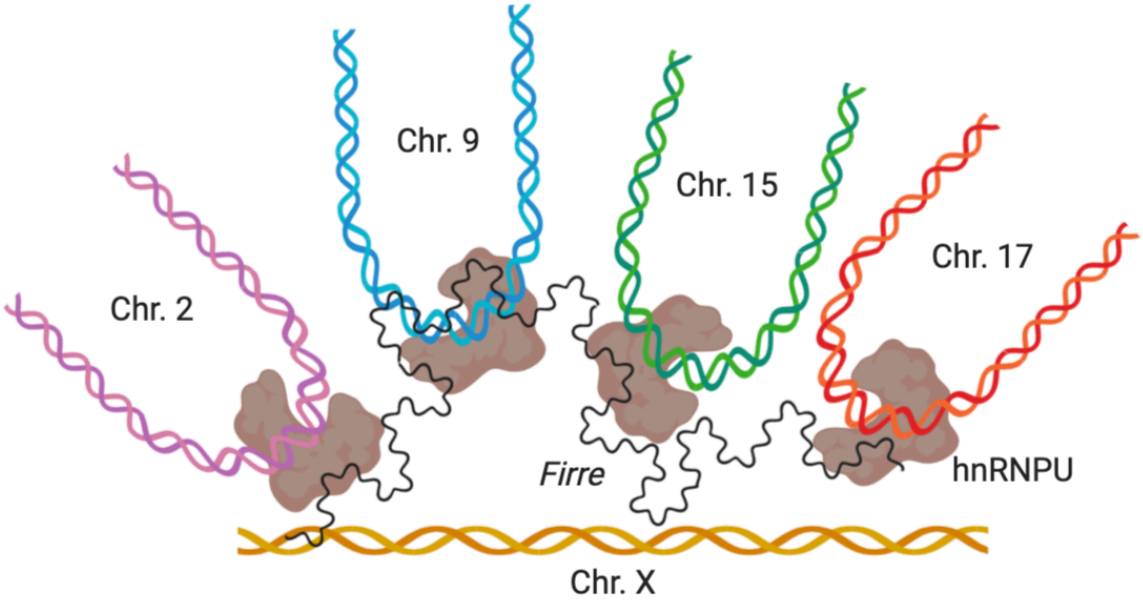


Figure 1.4 LncRNA *Firre* Mediates Formation of a *Trans*-Chromosomal Interaction. LncRNA *Firre* binds the hnRNPU protein and mediates a *trans*-chromosomal interaction between regions on chromosomes 2, 9, 15, 17 and the X chromosome. This provides a means of co-regulating genes at these loci associated with adipogenesis.

Another example is lncRNA-mediated topologically-associated domain (TAD) formation. Certain genomic loci interact with each other at far higher frequencies than other regions; these self-interacting, mega-base scale genomic units are known as TADs, and are conserved across a number of species and cell types (Dixon et al. 2012; Nora et al. 2012; Pope et al. 2014; Eser et al. 2017; Harmston et al. 2017). TADs provide an additional, structural layer of genomic regulation. The limitation on chromatin interactions that TADs pose is fundamental for proper gene regulation, with genes within a TAD often demonstrating correlated expression patterns (Nora et al. 2012).

A TAD boundary lies on the X-chromosome at the site of the gene encoding *Xist*; the *Xist* promoter lies in one TAD and the promoter for its antisense transcript, *Tsix*, lies in the adjacent TAD. Chromosomal looping has been observed within the latter between the *Tsix* promoter and an enhancer element named, Xite, which helps induce *Tsix* expression. The gene encoding lincRNA, *Linx*, can also be found within this region, and its expression is significantly, positively correlated with *Tsix*. It has been suggested that *Linx*, via an interaction with Xite, is important for promoting this chromosomal looping and thus *Tsix* expression (Nora et al. 2012).

1.1.4.4 LncRNAs and DNMT1

In recent years, there has been increasing evidence implicating lncRNAs in the regulation of gene expression via DNA methylation. Numerous studies have demonstrated significant interplay between DNA methylation machinery and

lncRNAs, with hundreds of DNA MethylTransferase 1 (DNMT1)-interacting lncRNAs having been identified (Merry et al. 2015; Di Ruscio et al. 2013).

One of the first DNMT1-lncRNA interactions was described in 2010, showing that *Kcnq1ot1* binds and recruits DNMT1 to a set of imprinted genes, whereby subsequent DNA methylation can mediate gene repression in an allele-specific manner (Mohammad et al. 2010). This was recently followed up by a team who demonstrated that preventing this methylation by knocking down *Kcnq1ot1,* can block progression of osteosarcoma and promote sensitivity to cisplatin treatment (Qi et al. 2019).

Since this initial finding, there has been an eruption of reports showing interactions between DNMT1 and a plethora of different lncRNAs, including *lincRNA-p21* (Bao et al. 2015), *HOXA11-AS* (Sun et al. 2016), *Dali* (Chalei et al. 2014), *Dum* (Wang et al. 2015), *PARTICLE* (O'Leary et al. 2017), *FOXD2-AS1* (Gao et al. 2019) and *HOXD-AS1* (Guo et al. 2019), to name a few. Generally, these interactions recruit DNMT1 to either mediate or prevent DNA methylation at specific genes.

In 2013, a non-polyadenylated lncRNA, *ecCEBP*, transcribed from the CEBPA gene in same-sense orientation, was identified in two myeloid cell lines (Di Ruscio et al. 2013). This lncRNA was shown to mediate expression of its protein-coding counterpart, CEBPA, by interacting with DNMT1 and preventing it from methylating the CEBPA promoter. Accordingly, *ecCEBP* knock down (KD) experiments showed increased CEBPA promoter methylation levels and subsequently reduced CEBPA

mRNA expression (Di Ruscio et al. 2013). This study went on to further investigate the nature of such interactions. Remarkably, electromobility shifts assays (EMSAs) indicated that via its C-terminal catalytic domain, DNMT1 binds RNA with a higher affinity than DNA, thus consolidating RNA as a critical regulatory component. Furthermore, this affinity is enhanced for RNAs that can form stem loops, indicating that RNA secondary structure is important (Di Ruscio et al. 2013).

One particularly interesting case involves lncRNA, *DNMT1-associated colon cancer repressed lncRNA 1* (*DACOR1*), in colon cancer cells. Its interaction with DNMT1, regulates the DNA methylome without alteration of DNMT1 protein levels. 450K DNA methylation arrays showed that induction of *DACOR1* results in increased DNA methylation at numerous loci, believed to occur by guiding DNMT1 to particular regions. Approximately 50 sites demonstrated differential methylation upon *DACOR1* induction, including the PHGDH and PSAT1 genes (Merry et al. 2015). Recent work however, suggests that *DACOR1* represents a more unique case, as reduced representation bisulfite sequencing (RRBS) has shown that this particular lncRNA can regulate DNA methylation on a genome wide scale. The Khalil group showed evidence of methylation restoration at more than 17,000 sites across the genome, that extended beyond just gene promoters, to intergenic regions and gene bodies, upon activation of *DACOR1* expression (Somasundaram et al. 2018). A possible reason for this could be because one of the genes silenced by *DACOR1*-mediated methylation encodes cystathionine β-synthase, whose repression leads to increased levels of the DNA-methylation methyl-donor, S-adenosylmethionine (SAM) (Merry et al. 2015).

1.1.5 Functions of lncRNAs: Post-Transcriptional

In addition to transcriptional gene regulation, emerging evidence also suggests that lncRNAs provide an additional layer of regulation at the post-transcriptional level. Following completion of transcription, gene expression continues to be regulated through control of processing, stability and subcellular localisation of both mRNA and protein. lncRNAs contribute to this regulation both directly and indirectly, often by formation of ribonucleoprotein (RNP) complexes (He, Luo, and Mo 2019; Lan et al. 2018; Diaz-Lagares et al. 2016).

1.1.5.1 LncRNAs and mRNA Stability

mRNA stability is largely regulated by RNA binding proteins, and consequently, lncRNAs found to play a role in mRNA stability, often interact with such proteins. These proteins can either act to stabilise or destabilise mRNAs, and whether lncRNAs help to recruit or prevent them from binding determines the stability of the mRNA (He, Luo, and Mo 2019). *LincRNA regulator of reprogramming* (*Linc-RoR*) is an example of a lncRNA that regulates stability of *c-Myc* mRNA. Interestingly, *Linc-RoR* prevents binding between *c-Myc* and the destabilising protein AUF1, as well as promotes an interaction between *c-Myc* and the stabilising protein, hnRNPI. Consequently, in the presence of *Linc-RoR*, *c-Myc* is stabilised and cell proliferation is enhanced (Huang et al. 2016).

1.1.5.2 LncRNAs and mRNA Splicing

Following transcription, mRNA is spliced, and in higher eukaryotes, alternative splicing provides a means of generating transcripts with different combinations of

exons from the same gene (Gilbert 1978). It has been suggested that lncRNAs play an important role in selecting specific splicing factors for specific genes (He, Luo, and Mo 2019). A well-known lncRNA associated with this process is *MALAT1*. *MALAT1* sequesters serine-arginine (SR) splicing factors in nuclear speckles and regulates their phosphorylation. It is not clear how this happens, but it is known that phosphorylation status can influence cycling of the factors between transcripts and nuclear speckles. Consequently, *MALAT1* regulates alternative splicing. In its absence, SR splicing factors are increasingly de-phosphorylated and alternative splicing patterns are altered (Tripathi et al. 2010).

1.1.5.3 LncRNAs and Protein Stability

In addition to mRNA, lncRNAs can also control protein stability. The RNA binding protein, ELAVL1, can be destabilised by binding the lncRNA, *overexpressed in colon carcinoma-1* (*OCC1*). *OCC1* promotes an interaction between ELAVL1 and a ubiquitin E3 ligase, which ultimately renders the protein vulnerable to degradation. Interestingly, ELAVL1 itself is involved in stabilising a large number of mRNA transcripts, meaning that as well as regulating protein stability, *OCC1* also indirectly regulates mRNA stability in this process (Lan et al. 2018).

Protein stability is largely determined by post-translational modifications (PTMs), so naturally, this is another feature utilised by lncRNAs. Phosphorylation of two specific threonine residues on the enhancer of zeste homolog 2 (EZH2) was found to be enhanced as a result of a lncRNA-mediated interaction between EZH2 and cyclin dependent kinase 1 (CDK1). This lncRNA was termed *anti-differentiation ncRNA*

(*ANCR*) and the resultant phosphorylation it induces leads to EZH2 degradation (Li, Hou, et al. 2017).

1.1.5.4 LncRNAs and Protein Localisation

Another means of regulation involves controlling where proteins reside within the cell. Prevention of proteins from reaching the subcellular compartment where their functions are required provides an additional layer of control. For example, DNA methyltransferase enzymes must be correctly localised to the nucleus in order to methylate DNA, and nucleolin requires localisation at nucleoli to aid ribosome assembly (Ma et al. 2007; Arzenani et al. 2011). Several lncRNAs have been shown to prevent nuclear localisation of proteins. These include *TP53TG1*, which directly sequesters YBX1 protein in the cytoplasm (Diaz-Lagares et al. 2016) and *CRYBG3* which prevents nuclear translocation of MAL protein via an interaction with G-actin (Pei et al. 2018). lncRNA *NEAT1* on the other hand regulates protein localisation not to the nucleus, but to nuclear bodies within the nucleus known as paraspeckles. As mentioned previously, *NEAT1* behaves as a scaffold onto which paraspeckle proteins such as NONO, PSPC1, SFPQ and p54 can assemble. In the absence of NEAT1, paraspeckles fail to generate (Clemson et al. 2009).

In an interesting instance, lncRNAs also play a role in the nucleolar stress response. If a cell is subjected to a stressor, such as heat shock or DNA damage, expression of a group of lncRNAs is induced from intergenic spacer (IGS) regions on ribosomal DNA. These IGS lncRNAs recognise a distinct amino acid sequence, commonly referred to as the Nucleolar Detention Sequence (NoDS), which allows them to bind

specific proteins and sequestrate them within a nucleolar sub-compartment called the Nucleolar Detention Centre (NoDC) (Audas, Jacob, and Lee 2012; Jacob et al. 2013). The aim is to gather the proteins necessary to overcome the cellular stressor. For example, in response to hypoxia, von Hippel-Lindau (VHL) is sequestered at the NoDC, thus preventing VHL-mediated hypoxia-inducible factor (HIF)-degradation. This allows HIF to stimulate transcription of a variety of genes involved in the maintenance of steady-state Oxygen levels (Mekhail et al. 2004).

1.1.6 LncRNAs and Phase Separation

Currently, the lncRNA field has turned its attention to roles in phase separation. Phase separation is a phenomenon whereby two discrete phases are generated within a homogeneous, liquid mix. This involves the stable separation of a saturated solution into a dense phase and a diluted phase, like oil droplets in water. In a cellular setting, this commonly refers to the generation of membrane-less compartments within the cytosol such as nucleoli or nuclear speckles, whose components remain distinct from the rest of the cytoplasm (Boeynaems et al. 2018).

It is believed that two key components are required for phase separation-mediated generation of membrane-less organelles. The first is a RNA component, that contains repetitive sequence elements that are capable of forming secondary structures, in order to sequester proteins (Langdon et al. 2018). It was recently shown that local RNA concentrations are a big driving force for phase separation, with lower concentrations driving aggregation and higher concentrations preventing it (Maharana et al. 2018). Second are intrinsically disordered proteins (IDPs), which

contain regions lacking stable 3D structures. These generally contain regions of hydrophobicity and hydrophilicity to enhance formation of aggregates as well as further protein interactions respectively (Tartaglia et al. 2008; Shin and Brangwynne 2017).

NEAT1 and the formation of paraspeckles is one example of phase separation-mediated organelle formation. One group argued that paraspeckle formation possesses many characteristics of phase-separated bodies. For example, the repeat nucleotide sequences within *NEAT1* generate hairpin loop structures that are required for binding paraspeckle proteins, and specific IDPs are critical for generation of paraspeckles *in vivo* (Yamazaki et al. 2018; Hennig et al. 2015).

Recently, phase-separation has been found to extend beyond formation of intracellular organelles, and also appears to occur to help regulate other biological processes such as gene expression (Cerase et al. 2019). An exciting hypothesis has been put forward, suggesting that phase separation is used by *Xist* in X-chromosome silencing. This process demonstrates very similar features to the formation of paraspeckles via phase separation (Cerase et al. 2019). Tartaglia and colleagues suggested that phase separation commences as *Xist* begins interacting with IDPs, following its coating of the Xi. *Xist*-interacting proteins accumulate, generating a more compact environment, resulting in the formation of phase separated foci (Cerase et al. 2019). Analysis of the *Xist* interactome greatly supports this theory. Proteins that directly interact with *Xist*, as well as the majority of proteins that interact indirectly, are IDPs and are more inclined to undergo phase separation

(Klus et al. 2014; Bolognesi et al. 2016; Cerase et al. 2019). Interestingly, a large portion of these proteins are also associated with paraspeckle formation (Van Nostrand et al. 2016; Cerase et al. 2019). In addition, *Xist* demonstrates structured repeat regions, which have been proven essential for interactions with XCI proteins (Pintacuda, Young, and Cerase 2017), and *Xist* assemblies are similar to paraspeckle assemblies in diameter and shape (Smeets et al. 2014; Cerase et al. 2014; Yamazaki et al. 2018).

Currently, studies focussed on this area appear to collectively agree that RNA secondary structure is fundamental to the foundation of such phase-separated bodies (Hutchinson et al. 2007; Yamazaki et al. 2018; Cerase et al. 2019). In the future, it will be interesting to further scrutinise the precise role that lncRNAs play in this process.

1.2 *Coiled-Coil Domain-Containing 26* (*CCDC26*)

1.2.1 The 8q24 Locus

The chr.8q24 locus on chromosome 8 is of particular interest in the ncRNA field, due to the abundance of lncRNAs expressed from this locus. This region, approximately 1.2Mb in length garnered initial interest due to its abundance of cancer-associated variants and the fact that the area was considered to be devoid of genes (Ghoussaini et al. 2008). Some of the first genome wide association studies conducted, focussed on 8q24 SNP variants and their links to prostate cancer (Amundadottir et al. 2006; Gudmundsson et al. 2007; Haiman et al. 2007; Thomas et al. 2008). Over the years, several additional SNPs, as well as other

mutation types within this region, were also found to be associated with various other types of cancer. In 2008, at least five discrete loci were identified within the 8q24 locus, variants of which were associated with increased risk of prostate, breast, colorectal and ovarian cancers (Ghoussaini et al. 2008). The following year, 111 cases of childhood AML were characterised. Although the number of genomic alterations found were few, of those identified, particular regions demonstrated frequent copy-number changes (Radtke et al. 2009). The most commonly amplified region was 8q24.21, with 14% of AML cases analysed demonstrating copy number gains, an observation replicated by both the Downing group and the Preudhomme group (Radtke et al. 2009; Kuhn et al. 2012; Duployez et al. 2018). Recently, these results were further confirmed when deletion of the 8q24 gene desert region in mice demonstrated a remarkable anti-tumorigenic effect (Homer-Bouthiette et al. 2018). A large-scale study concluded that the most recurrently amplified locus across almost all human cancers is the 8q24 region (Beroukhim et al. 2010). Furthermore, links with other disease types are also currently emerging, with one study having found more than 30 markers at chr.8q24.21 associated with lumbar disc herniation (Bjornsdottir et al. 2017).

It is clear that chr.8q24 is a mutational hotspot in cancers. Given that this locus is adjacent to the c-Myc oncogene, it is unsurprising that this has been the primary focus of most studies (Huppi et al. 2012). Multiple reports have demonstrated a correlation between amplification of c-Myc and increased cancer progression (Slovak et al. 1994; Fegan, White, and Sweeney 1995; O'Malley et al. 1999; Receveur et al. 2004). Almost half of the cases of myeloid disorders that Rayeroux

and Campbell analysed, demonstrated amplification of the c-Myc gene on double minute chromosomes (Rayeroux and Campbell 2009).

A common hypothesis is that the risk variant region comprises a c-Myc regulatory region. Epigenetic analysis identified several transcriptional enhancers at 8q24, whose genomic positions coincided with several cancer-associated SNPs (Jia et al. 2009). Furthermore, long-range chromatin looping has been observed between the c-Myc gene and predicted 8q24 enhancers (Ahmadiyeh et al. 2010).

However, there are conflicting results regarding 8q24 variants and levels of c-Myc. In many instances, the presence of 8q24 SNPs does not correlate with differential c-Myc levels. For example, comparison of approximately 200 cases of cancerous prostate tissue and normal tissue demonstrated no significant correlation between c-Myc expression and the presence of 6 different 8q24 cancer-associated SNPs (Pomerantz et al. 2009).

Following this, numerous pieces of evidence surfaced indicating that c-Myc should not necessarily be the sole focus for investigations of this region; it has emerged in recent years that this previously considered “gene desert” actually expresses numerous lncRNAs, including *PCAT1, CCAT1, POU5F1P1, PVT1* and *CCDC26* (Figure 1.5) (Huppi et al. 2012).

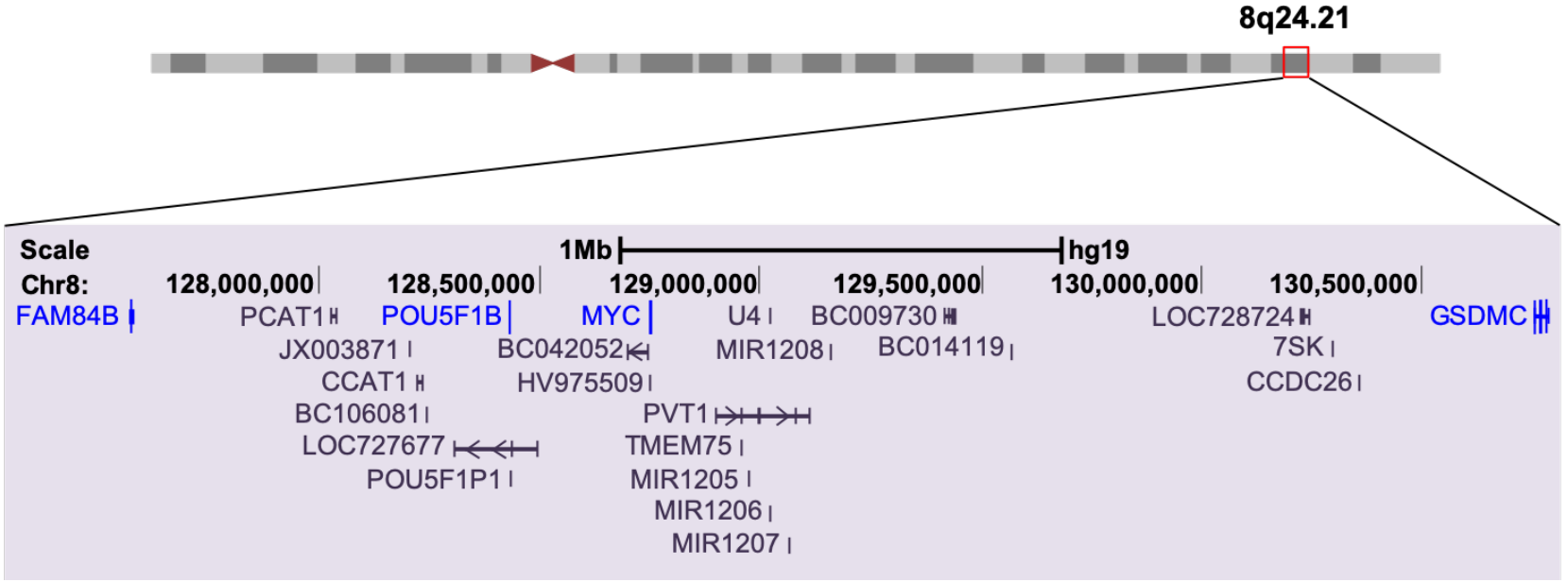


Figure 1.5. The 8q24 Locus is a Sea of LncRNA Genes. The chr.8q24 locus was previously thought to be a gene desert, however, numerous lncRNAs have been reported to be transcribed from this region in recent years, including *CCDC26*. Protein coding genes are indicated in blue, and genes encoding ncRNAs are indicated in purple. Figure is adapted from the UCSC Genome Browser view of the chr.8q24 locus (UCSC genes track, GRCh37/hg19 assembly).

It is extremely interesting to find that in various cancers, increased expression of c-Myc is very often positively correlated to increased expression of *PVT1* (Shtivelman and Bishop 1989). One murine study went further and demonstrated that overexpression of one of four genes (c-Myc, PVT1, CCDC26 or GSDMC) at the 8q24 locus alone, is insufficient to significantly promote cancer progression, however, amplification of all four genes prominently enhances tumorigenesis (Tseng et al. 2014). Furthermore, various lymphomas often exhibit translocation mutations targeting c-Myc, however, it has now been shown that in several cases, these mutations occur at the 8q24 locus, downstream of c-Myc (Huppi et al. 2012). Collectively, these observations suggest that mutations at the 8q24 locus are not necessarily always targeted to c-Myc. It seems that the somewhat overlooked genes encoding lncRNAs in this region might be more important than previously thought.

1.2.2 LncRNA *CCDC26*

One lncRNA encoded from within the 8q24 locus, is *coiled-coil domain-containing 26* (*CCDC26)*. *CCDC26* is of particular interest because it appears to be targeted by many of the mutations and variants discussed in 1.2.1. In AML, one of the most frequent lesions implicated in disease progression is focal copy-number gain of CCDC26 (Radtke et al. 2009; Kuhn et al. 2012; Duployez et al. 2018). It was additionally reported that a fusion gene between CCDC26 and NSMCE2, an E3 SUMO-protein ligase, identified in both patient-derived AML cells and the AML HL60 cell line, could also play a functional role in tumorigenesis (Chinen et al. 2014). A more recent study analysed expression levels of *CCDC26*, as well as *PVT1* and

CCAT1, in AML patients. Although no significant difference in *CCDC26* levels were observed between AML cases and controls when looking at the AML population as a whole, differential expression levels could be observed when analysing specific AML subtypes or stages of progression. Intermediate cytogenetic risk cases displayed increased levels of *CCDC26*, as well as patients with AML subtypes, AML-M2, AML-M4 and AML-M5 (Izadifard et al. 2018). Despite this, there have been very few publications investigating the biological significance of this lncRNA.

CCDC26 is a lincRNA encoded by a ~330kb gene located on chromosome 8. Its gene resides at the 8q24.21 locus, situated between the protein coding gene Gasdermin C (GSDMC) and the gene encoding lncRNA *Plasmacytoma Variant Translocation 1* (*PVT1*). At present, four *CCDC26* isoforms have been identified, that are transcribed from one of two transcription start sites (TSSs), separated by a distance of approximately 105kb (Figure 1.6). This is according to current Reference Sequence (RefSeq) gene predictions from the National Center for Biotechnology Information (NCBI). These isoforms comprise three short splice variants transcribed from TSS-2, and a single long variant transcribed from an upstream start site (TSS-1). Each consists of variations of exons 1-6, with exons 5 and 6 being common to all (Figure 1.6). Isoforms 1 and 2 are highly similar, differing only by the additional 17bp region at the 3' end of exon 4 in isoform 1. In addition to exon 4, these isoforms are made up of exons 2, 5 and 6. Isoform 3 is also transcribed from TSS-2, but lacks exon 4 completely, instead consisting of exons 2, 5 and 6 only. Isoform 4 on the other hand is comprised of exons 1,3, 5 and 6.

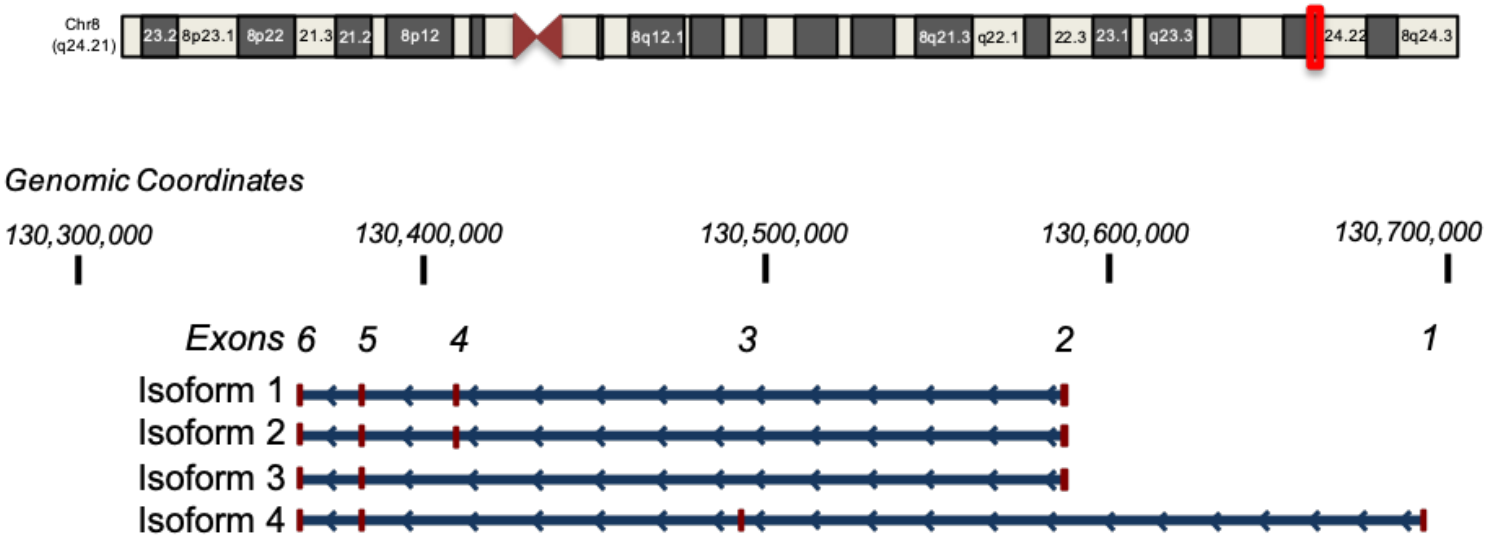


Figure 1.6 *CCDC26* Isoforms. The gene encoding CCDC26 is found at the 8q24 locus on chromosome 8 (red box). There are currently four known isoforms of *CCDC26* transcribed from one of two transcription start sites, and consisting of variations of exons 1-6. Three short isoforms are transcribed from TSS-1, and one long isoform is transcribed from TSS-2. Isoforms 1 and 2 consist of exons 2, 4, 5 and 6 (differing by just 17bp in exon 4), isoform 3 consists of exons 2, 5 and 6, and isoform 4 consists of exons 1, 3, 5 and 6. Figure is adapted from the UCSC Genome Browser view of the CCDC26 locus (NCBI RefSeq tracks, GRCh37/hg19 assembly).

CCDC26 was first discovered in 2006, by a group studying acute promyelocytic leukemia therapies. Specifically, they were investigating downstream mediators of differentiation and apoptosis, induced by retinoic acid treatment. By implementing a retroviral insertional mutagenesis technique, they generated a cell line resistant to this therapy, which upon further investigation demonstrated mutation of the *CCDC26* gene. This insertional mutation was localised to what we now know to be intron 1 of the gene (chr.8: 130,629,014). Interestingly, whilst wild type (WT) cells showed *CCDC26* silencing upon treatment with retinoic acid, the mutated cells, termed WY-1, presented persistent *CCDC26* expression following treatment, as well as impaired differentiation and apoptosis. Furthermore, lack of apoptosis and differentiation could be rescued in retinoic acid-treated WY-1 cells if *CCDC26* was knocked down (Yin et al. 2006).

Not only did this group establish *CCDC26* as a retinoic acid-dependent regulator of myeloid cell differentiation and death, but they were also the first to indicate that *CCDC26* is a ncRNA and is not translated into a protein. The largest hypothetical open reading frame (ORF) consists of 109 amino acids encoded from exon 6, however, no such protein was and has ever been detected endogenously. Furthermore, the only predictable motif within this “protein” would be a small 17-amino-acid signal region. Exogeneous overexpression of this ORF tagged with a B10 epitope in leukemia cells failed to yield a detectable protein construct, despite detection of the corresponding RNA, suggesting that either the CCDC26 protein is unstable or is not expressed at all. The B10 epitope tag was proven not to be the cause of instability, as a similar V5-epitope-tagged construct also failed to yield a

detectable protein. This, combined with the observed impact of siRNA-mediated *CCDC26* knock down, strongly supported the hypothesis that *CCDC26* RNA is functional (Yin et al. 2006). In further support of this, unlike protein-coding genes, the exonic regions of *CCDC26* are also seemingly not conserved between species. Consequently, proteins homologous to the hypothetical "CCDC26 protein" have also not been established in other species (Hirano et al. 2015).

1.2.3 *CCDC26* Function

Despite being identified more than 10 years ago, our current understanding of the biological role of *CCDC26* is still very limited. The initial *CCDC26* study by the Gronemeyer group, acknowledged a clear link with retinoid-induced differentiation and apoptosis in myeloid cells. Their results indicated a potential role for *CCDC26* in maintenance of an undifferentiated cell state (Yin et al. 2006). A subsequent study supported this by demonstrating that constitutive *CCDC26* expression was associated with inhibition of differentiation in the AML HL60 cell line. Yokoyama and colleagues demonstrated that late-passaged HL60 cells gained large extrachromosomal elements within their nuclei and were unable to differentiate. Large portions of these extrachromosomal elements encompassed the 8q24 region, where certain genes, including *CCDC26*, were particularly prominent. Consequently, these genes were found to be constitutively expressed, indicating a role in maintaining a pluripotent state (Hirano et al. 2008).

In the following years, a number of genome-wide studies emerged, linking *CCDC26* with glioma. Multiple analyses reported *CCDC26* as a susceptibility gene for glioma,

and numerous SNPs within this lncRNA were associated with increased risk of the disease (Lasho et al. 2012; Li et al. 2012; Wei et al. 2014; Adel Fahmideh et al. 2015; Cui 2015; Lu et al. 2015; Wang et al. 2016; Zeng et al. 2017; Gonzalez-Castro et al. 2019). The rs4295627 polymorphism has drawn particular attention, having been linked to an ethnicity-dependent, increased risk of glioma (Cui 2015; Lu et al. 2015; Wang et al. 2016; Zeng et al. 2017). One group went further to confirm the relationship between *CCDC26* and glioma progression; they demonstrated that the lncRNA is overexpressed in glioma cells, where it seemingly regulates glioma progression via direct targeting of *miR-203*. *CCDC26* siRNA knock down in glioma cells increased apoptosis and reduced both cell proliferation and motility, an effect observed both *in vitro* and *in vivo* (Wang, Hui, et al. 2018).

Another group also investigated the biological function of *CCDC26* in CML cells (Hirano et al. 2015). The Yamazaki group proposed that *CCDC26* functions at least in part by regulating expression of the tyrosine kinase receptor, KIT (Hirano et al. 2015). In high serum media (15% FBS), *CCDC26* shRNA knock down cells grew slowly, with a doubling time of almost twice that of WT CML K562 cells. In contrast, cell survival was significantly improved when grown in serum-depleted media (0.1% FBS). This suggested that *CCDC26* can behave oncogenically, enhancing cell proliferation. However, in harsher conditions where nutrient supply is low, the opposite appears true, with cells dividing faster and surviving longer with reduced *CCDC26* levels. Upon investigating differential gene expression using DNA microarrays, KIT was identified as being significantly up-regulated in the *CCDC26* knock down cells. When knock down cells were treated with a specific KIT inhibitor

under serum starvation, cell survival was reduced back to that of WT K562 cells (Hirano et al. 2015).

Similar to previous glioma and leukemia studies, Peng and Jiang too reported increased levels of *CCDC26* in pancreatic cancer tissues. Although they detected no significant correlation between *CCDC26* and tumour differentiation, siRNA knock down resulted in increased apoptosis and reduced proliferation. They hypothesised that this could be via regulation of PCNA and BCL2 (Peng and Jiang 2016).

Conversely, emerging evidence has linked *CCDC26* inhibition with drug resistance in gastrointestinal stromal tumours (GIST), suggesting that upregulation of this lncRNA is a more desirable effect in these particular tissues. *CCDC26* siRNA knock down made cells resistant to imatinib therapy and resulted in increased cell viability and proliferation and reduced apoptosis (Cao et al. 2018; Yan et al. 2019). One study has attributed this to *CCDC26*-mediated regulation of KIT. *CCDC26* knock down resulted in increased levels of KIT, whilst KIT siRNA knock down increased sensitivity of cells to Imatinib. Consequently, Cao et al. concluded that *CCDC26* inhibition potentially induces Imatinib resistance by activating increased expression of KIT (Cao et al. 2018). Similarly, an alternative study proposed that *CCDC26* inhibition potentially induces Imatinib resistance by activating increased expression of IGF-1R (Yan et al. 2019). Similar to KIT, IGF-1R is also increased following *CCDC26* knock down, with cells also showing increased Imatinib sensitivity upon IGF-1R knock down (Yan et al. 2019).

Taking into consideration all of the research and information gathered in 1.2, we decided to concentrate our attention on the Chr.8q24 region. Our research group is focussed on understanding the molecular mechanisms of lncRNAs, and Chr.8q24 has been shown to contain a sea of lncRNA-expressing genes. Furthermore, the abundance of cancer-associated SNPs and variants reported within this region make it particularly interesting (Ghoussaini et al. 2008). Most previous work on this area has focussed on the locally-expressed c-Myc gene, with studies suggesting that 8q24 contains c-Myc enhancers (Huppi et al. 2012; Jia et al. 2009; Ahmadiyeh et al. 2010). However, as discussed earlier, several conflicting results indicate that c-Myc-alternative functions of the region exist, thereby warranting further investigation (Pomerantz et al. 2009; Huppi et al. 2012). Multiple lncRNA genes reside at 8q24, however, we have chosen to study *CCDC26*. We noticed that many of the cancer-associated SNPs and variations reported in past publications can be mapped to the *CCDC26* gene (Radtke et al. 2009; Kuhn et al. 2012; Duployez et al. 2018; Tseng et al. 2014; Chinen et al. 2014). Furthermore, its expression has been associated with specific stages of leukemia progression (Izadifard et al. 2018). However, there are very few studies that have investigated *CCDC26* independently, and those that have, have done little beyond identifying an association with apoptosis and differentiation (Yin et al. 2006; Hirano et al. 2008; Hirano et al. 2015; Wang, Hui, et al. 2018; Peng and Jiang 2016). Taking all of this together, we identified a gap in the field. We decided it was important to functionally characterise this lncRNA in order to fully understand the implications of such mutations within its gene.

1.3 The Chromatin Landscape

The mammalian nucleus is the fundamental control centre of each cell, directly or indirectly responsible for coordination of all cell activity. Correct functioning of the nucleus and its processes requires strict organisation of the components within, including regulation of the spatial localisation of proteins, RNA and chromosomes. Organisation of DNA in particular is critical, as the hierarchical compaction of chromosomes is critically interwoven with the interpretation of genetic material (Felsenfeld and Groudine 2003). Walther Flemming first coined the term 'chromatin' when he observed a stainable, fibrous framework within the nuclei of dividing cells (Paweletz 2001). Today we use the term to describe the complex structure of DNA, RNA and proteins into which the eukaryotic genome is assembled, with the overarching purpose of efficiently folding, as well as protecting, DNA within a small space (Felsenfeld and Groudine 2003).

1.3.1 The Nucleosome

The basic repeating unit of chromatin is known as the nucleosome, which consists of a protein core around which approximately 147bp of DNA is wrapped (Figure 1.7). This gives rise to a ~10nm fibre, often referred to as the 'beads-on-a-string' structure, and comprises the first level of chromatin compaction (Kornberg 1974; Kornberg and Thomas 1974). The centre of each nucleosome contains 8 core histones, consisting of two copies each of histones H2A, H2B, H3 and H4, wound around by approximately 80bp of linker DNA (Richmond et al. 1984). Whereas DNA associated with this core is somewhat shielded, the linker DNA is accessible and susceptible to enzymatic digestion. (Luger et al. 1997; Davey et al. 2002). The

negatively charged DNA phosphate groups interact largely with positive Arginine (R) and Lysine (K) residues within N-terminal ends of histones (Luger et al. 1997). Eight highly conserved R residues are found on histones H2A, H3 and H4, and are critical for accurately positioning the wrapped DNA around the octamer (Wang, Ulyanov, and Zhurkin 2010).

The incorporation of DNA into nucleosomes is transient, thus, nucleosomes are highly dynamic structures. Approximately one quarter of the mass of each histone is found within an N-terminal tail domain. These tails project from the nucleosome structure such that they are exposed to enzymes involved in PTMs. These enzymes mediate the addition and removal of various chemical groups to and from histone tails, often at prevalent K and R residues, which is important for epigenetic signalling (detailed in section 1.4.1) (Davey et al. 2002; Allfrey, Faulkner, and Mirsky 1964). The final principal element of the nucleosome is the linker histone (H1). Unlike the core histones which make up the central octamer, H1 binds the linker DNA between nucleosomes. It's precise function is somewhat ambiguous, but studies have suggested roles including protection of the linker DNA from nuclease digestion, stabilisation of interacting DNA and organisation of higher order structures (Allan et al. 1986).

1.3.1.1 Histone H2A.X

Whilst the canonical histones mentioned above are rapidly generated and incorporated into nucleosomes closely behind replication forks, it is important to also note that non-canonical histone variants exist. Non-canonical substitution of

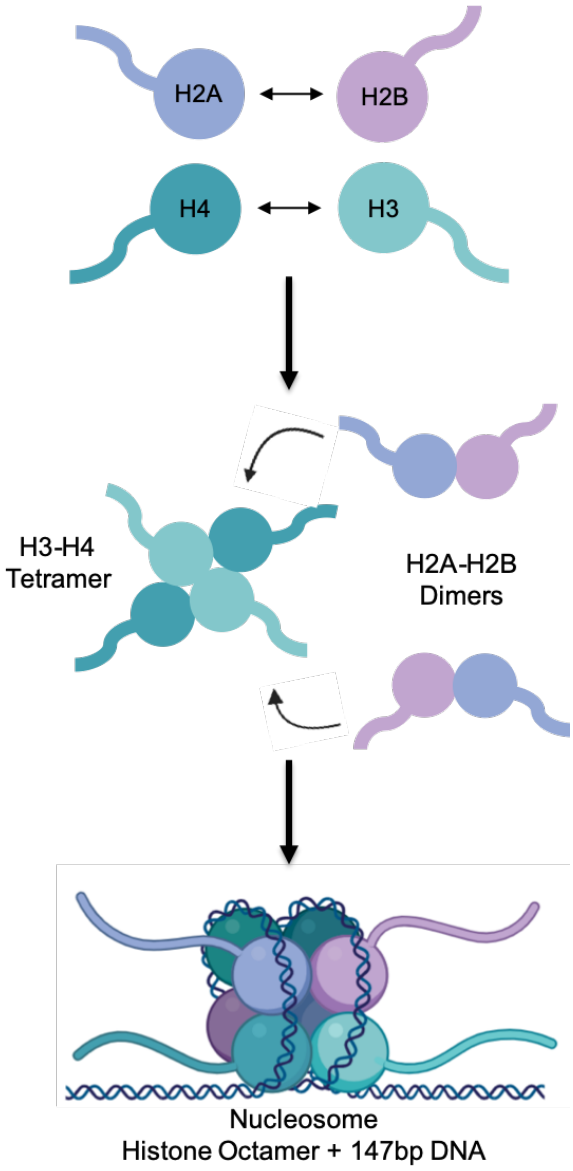


Figure 1.7. Nucleosome Structure. The nucleosome consists of two copies each of core histones H2A, H2B, H3 and H4, around which 147bp of DNA is wound. Core histones are first arranged as two sets of heterodimers: H2A-H2B and H3:H4. The H3:H4 heterodimers form a tetramer by associating at the H3-H3 interface. This tetramer binds the H2A:H2B dimers via an interaction between H4 and H2B.

histones occurs independently of DNA replication and can serve numerous purposes. A well-known histone H2A variant is H2A.X; this variant is phosphorylated to form **γ**-H2A.X in response to DNA damage, especially DNA double strand breaks, (DSBs) and is a key event in the DNA damage response pathway (Rogakou et al. 1999; Rogakou et al. 1998). H2A.X is incorporated randomly into chromatin throughout the genome (Rogakou et al. 1998). It has been suggested that DSBs are detected by DNA-dependent protein kinase, which ultimately leads to phosphorylation of H2A.X at the conserved Serine 139 residue, by either ataxia telangiectasia mutated (ATM) or ATM-RAD3-related (ATR) proteins (Burma et al. 2001; Durocher and Jackson 2001). This generates **γ**-H2A.X, which can radiate several mega-bases from the DSB site (Rogakou et al. 1999). This histone modification behaves as a signal onto which DNA repair machinery assembles, such as the NBS1/hMRE11/hRAD50 repair complex (Kobayashi 2004)

1.3.2 Higher-Order Chromatin Fibres

It was previously believed that the 10nm bead-like chromatin structure further folded into a structured 30nm fibre. This was once considered to be the standard chromatin unit from which higher order structures were arranged (Gilbert et al. 2004). However, this prediction was based on *in vitro* studies and has never been observed within the cell (Finch and Klug 1976). Since the late 1980's, techniques including cryogenic electron-microscopy (cryo-EM), small angle X-ray scattering and electron spectroscopic imaging have indicated that the 10nm chromatin fibres are folded in an irregular manner rather than forming a structured 30nm fibre (McDowall, Smith, and Dubochet 1986; Nishino et al. 2012; Fussner et al. 2012). Recently, live cell

imaging, super-resolution (STORM) imaging and ChromEMT, a combination of electron microscopy tomography and DNA labelling, showed evidence of these dynamic, irregular fibres (Ricci et al. 2015; Ou et al. 2017; Nozaki et al. 2017). These findings indicate that chromatin is far more dynamic than previous models suggested. It is likely that this ensures that DNA is easily accessible to machinery for vital, biological processes such as replication and transcription.

1.3.3 Chromosome Territories

At the largest scale of organisation, 46 chromosomes (humans) are localised in specific territories within the nucleus (Cremer and Cremer 2010). The organisation of the nucleus was first deemed to be non-random in the late 1800's, when the Rabl configuration was established in plants and yeast. This is the arrangement whereby chromosomes lie parallel across the nucleus, with centromeres and telomeres aligning at opposite poles during anaphase (Mizuguchi, Barrowman, and Grewal 2015). This led to the discovery of chromosome territories (CTs), distinct regions in the nucleus where chromosomes preferentially localise. Theodore Boveri concluded that chromosomes inhabit distinct spaces in the nucleus, and maintain a level of distinctiveness throughout interphase (Cremer and Cremer 2010). Although this phenomenon was largely ignored for many years (Wischnitzer 1973), in the late 1970s and 1980s, Boveri's concept re-emerged, when a group treated interphase root-tip cells and Chinese hamster ovary (CHO) cells with modified Giemsa staining treatments. They suggested that chromosomes do not decondense during interphase such that 3D organisation is lost, but rather they maintain their preferential nuclear positions (Stack, Brown, and Dewey 1977).

It is well-established that there is a correlation between territorial organisation and transcriptional activity; gene-rich, transcriptionally active chromatin regions generally demonstrate preferential localisation at the centre of the nucleus, whereas, transcriptionally repressed regions containing fewer genes are more often localised at the nuclear periphery (Boyle et al. 2001). Bickmore and colleagues were amongst the first to demonstrate this in the 1990's with 3D fluorescence *in situ* hybridisation (FISH), showing that the nuclear interior and periphery are occupied with the gene-rich chromosome 19 and the gene-poor chromosome 18 respectively (Croft et al. 1999).

1.3.4 Heterochromatin v Euchromatin

Upon staining with basophilic dyes, chromosomes appear as a series of differentially, intensely stained bands. Darker stained bands arise as a result of an increased concentration of DNA compacted within that particular region of the chromosome and is commonly referred to as heterochromatin, a term first coined by Emil Heitz in the 1920's. It was Heitz who noted that such regions failed to de-compact following mitosis. The decondensed, lighter stained regions into which chromatin is less tightly compacted is known as euchromatin. Heitz too recognised that gene density is generally greater in euchromatic regions, compared to heterochromatic regions. However, these patterns were found to change between cell types and throughout development and cell cycle. This led to the distinction between constitutive heterochromatin, which is consistently formed regardless of the cell-type or stage of cell cycle, and facultative heterochromatin, which can be locus and cell-type specific (Passarge 1979). Today, we acknowledge that

heterochromatin and euchromatin are functionally distinct, and are consequently more commonly used to refer to transcriptionally repressed or active regions respectively, rather than referring to how regions are stained (Wang, Jia, and Jia 2016).

Approximately 10 years ago, researchers concluded that the genome is divided overarchingly into two distinct, nuclear compartments, within which smaller domain structures such as TADs, can be observed. These compartments are known as the A and B compartments, and are associated with euchromatin and heterochromatin respectively. The A compartment generally occupies the centre of the nucleus, with the B compartment lying at the nuclear periphery (Lieberman-Aiden et al. 2009). Recent observations have demonstrated the highly dynamic nature of these compartments, with large portions of the genome demonstrating changes in compartment occupancy during mammalian development (Dixon et al. 2015).

The segregation of active and inactive chromatin serves several purposes. Since heterochromatin, unlike euchromatin, requires limited access to transcriptional machinery in the centre of the nucleus, these regions are localised at the nuclear periphery whereby they can interact with components of the nuclear lamina. Such lamina-interacting loci are known as lamina-associated domains (LADs) and provide anchoring points (Guelen et al. 2008). Past studies have suggested that tethering can influence the folding and organisation chromosomes (Kind et al. 2015) as well as potentially strengthen the nuclear lamina itself (Tajik et al. 2016) and even aid repair of DNA double strand breaks (Lemaitre et al. 2014).

1.4 Epigenetic Gene Regulation

Heterochromatin and euchromatin are fundamentally regulated by epigenetic mechanisms. Epigenetics refers to reversible changes to DNA that can influence gene expression, which do not involve changes to the underlying DNA sequence. The term was first applied as early as the 1940s, by C.H. Waddington, and was originally used to describe how phenotypes may be generated via complex interactions between genes and the environment (Waddington 2012). Today, there are considered to be to major types of epigenetic modifications: DNA methylation and histone modifications (Allis and Jenuwein 2016).

1.4.1 Histone Modifications

1.4.1.1 The Histone Code Hypothesis

Revolutionary research in the 1960s, led to the first reports of histones being post translationally modified (Allfrey, Faulkner, and Mirsky 1964). Today, over 130 histone PTMs have been reported (Zhao and Garcia 2015). As we have mentioned, histones form the nucleosome core, with their N-terminal tail domains projecting outwards. Modification of these tails can subsequently serve multiple purposes, including mediating interactions between nucleosomes which can impact chromatin compaction, as well as recruiting assorted proteins with various functions (Davey et al. 2002; Allfrey, Faulkner, and Mirsky 1964; Bannister and Kouzarides 2011).

The dynamic distribution and pattern of histone modifications throughout the genome comprises the histone code, and the histone code hypothesis states that gene expression, at least in part, is regulated by these modifications which occur

predominantly on the tail regions. The primary role of this code is in determining the transcriptional state of a given region. Histone PTM patterns in combination with DNA methylation patterns are capable of rendering genes transcriptionally active or silent (Strahl and Allis 2000).

Histones tails are most frequently modified at Lysine residues, but modifications of other amino acid residues, such as Arginine, have also been observed (Gary and Clarke 1998). Methylation and acetylation are arguably the most common modifications, however, numerous others, including phosphorylation, ubiquitylation and sumoylation to name a few, also exist (Strahl and Allis 2000). Table 1.2 provides a list of some of the most commonly observed histone modifications, along with their abbreviations and general roles. For example, tri-methylation of Histone 3 at Lysine 4 is commonly found at active promoter regions (Heintzman et al. 2009).

1.4.1.2 Writers, Readers and Erasers of Histone Modifications

Each specific histone modification needs to be deposited, interpreted and removed when necessary. This is carried out by specific proteins: “writers”, “readers” and “erasers” respectively. Two key “writer” enzymes include histone acetyltransferases (HATs) and histone methyltransferases (HMTs); the former catalyses addition of acetyl groups to histone tails from acetyl CoA molecules, and the latter catalyses addition of methyl groups, utilising S-adenosyl-methionine (SAM) as a donor (Chiang et al. 1996).

Opposing “eraser” proteins however, can remove these modifications, making the process highly dynamic. Histone deacetylases (HDACs) and histone demethylases (HDMs) respectively catalyse the removal of acetyl and methyl groups. LSD1 was the first HDM identified only ~20 years ago, largely due to the early, common belief that histone methylation was irreversible (Shi et al. 2004). HDAC activity on the other hand has been acknowledged since the 1960’s (Inoue and Fujimoto 1969), with HDAC1 being the first HDAC to be purified in 1996 (Taunton, Hassig, and Schreiber 1996). To date, 18 different HDAC enzymes have been identified in humans, which can be categorised into one of four classes based on sequence similarities (summarised in Table 1.3). Accurate identification of different HDAC substrates has proven challenging in recent years. This is largely due to a predicted functional redundancy within and between HDAC classes, as well as interactions with different complexes likely altering the preferential substrates of each HDAC (Seto and Yoshida 2014).

“Reader” proteins are required to interpret histone signatures. Such proteins can bind histone modifications via specific domains, thus targeting their function to that particular locus (Kim, Daniel, et al. 2006). For example, DNA methyltransferase enzymes can recognise and bind H3K36me3 signatures via PWWP domains, thus directing DNA methylation (Rondelet et al. 2016).

Table 1.2. Commonly Observed Histone Modifications

Histone Modification	Abbreviation	Function	Reference
Histone 3 Lysine 4 mono-methylation	H3K4me1	Enhancer activation	(Heintzman et al. 2009)
Histone 3 Lysine 4 tri-methylation	H3K4me3	Promoter activation	(Heintzman et al. 2009)
Histone 3 Lysine 36 tri-methylation	H3K36me3	Gene body activation	(Huang and Zhu 2018)
Histone 3 Lysine 79 di-methylation	H3K79me2	Gene body activation	(Farooq et al. 2016)
Histone 3 Lysine 9 acetylation	H3K9ac	Enhancer, promoter activation	(Gates et al. 2017)
Histone 3 Lysine 27 acetylation	H3K27ac	Enhancer, promoter activation	(Heintzman et al. 2009)
Histone 4 Lysine 5 acetylation	H4K5ac	Epigenetic bookmarking Allows rapid transcription of specific genes immediately following mitosis	(Zhao et al. 2011)
Histone 4 Lysine 16 acetylation	H4K16ac	Regulates higher order chromatin structure	(Shogren-Knaak et al. 2006)
Histone 3 Lysine 27 tri-methylation	H3K27me3	Promoter, gene-rich region repression	(Wiles and Selker 2017)
Histone 3 Lysine 9 mono-methylation	H3K9me1	Establishment of heterochromatin	(Rivera et al. 2015)
Histone 3 Lysine 9 tri-methylation	H3K9me3	Repression of satellite repeats, telomeres, pericentromeres	(Lehnertz et al. 2003)
Phosphorylated Histone	**γ**-H2A.X	DNA damage, DSBs	(Rogakou et al. 1998)
Ubiquitylated Histone H2A Lysine 119	H2AK119ub1	Repression of developmental genes	(Zhang, Cooper, and Brockdorff 2015)
Histone 4 Lysine 20 mono-methylation	H4K20me1	Promoter activation	(Wang et al. 2008)

Table 1.3 The HDAC Family of Proteins (Seto and Yoshida 2014)

Superfamily	Class	Member	Subcellular Localisation
Arginase/deacetylase superfamily	I	HDAC1	Nucleus
		HDAC2	Nucleus
		HDAC3	Nucleus/Cytoplasm
		HDAC8	Nucleus
	IIA	HDAC4	Nucleus/Cytoplasm
		HDAC5	Nucleus/Cytoplasm
		HDAC7	Nucleus/Cytoplasm
		HDAC9	Nucleus/Cytoplasm
	IIB	HDAC6	Nucleus/Cytoplasm
		HDAC10	Nucleus/Cytoplasm
deoxyhypusine synthase-like NAD/FAD-binding domain superfamily	III	Sirtuins 1, 2, 3,4, 5, 6, 7	Nucleus/Cytoplasm/ Mitochondria
Arginase/deacetylase superfamily	IV	HDAC11	Nucleus

1.4.2 DNA Methylation

DNA methylation is considered to be the principal mechanism of epigenetic regulation, alongside histone modifications (Jaenisch and Bird 2003). In eukaryotes, most DNA methylation occurs at the 5th Carbon atom of the cytosine ring to generate C5-methylcytosine, commonly shortened to 5mC (Figure 1.8A) (Hotchkiss 1948). Other residues can also be methylated, such as the 4th Nitrogen atom of cytosine (4mC) and the 6th Nitrogen atom of adenine (6mA), however these are more commonly observed in bacteria, archaea and protists (Ehrlich et al. 1985; Ehrlich et al. 1987). In eukaryotes, 5mC signatures predominantly exist at cytosine-phosphate-guanine (CpG) dinucleotides; within the human genome, approximately 80% of CpGs are ordinarily methylated (Schubeler 2015). The large majority of methylated CpGs are found in repetitive sequence elements such as long and short interspersed nuclear elements, as well as retrotransposons and centromeres (Zheng et al. 2017). This helps prevent these regions from being transcribed unnecessarily. Non-methylated CpGs generally tend to be found in clusters at gene promoters and transcription start sites, forming CpG islands. CpG islands are regions approximately 1kb in length, whose DNA sequences diverge from typical genomic configurations by their characteristic elevated CpG levels and frequent methylation deficiency (Bird et al. 1985; Cross et al. 1994; Deaton and Bird 2011). This allows transcription of these genes to occur and additionally provides a means of gene regulation; cell- and tissue-specific methylation at particular loci can give rise to differential gene expression, resulting in diverse phenotypes (Jaenisch and Bird 2003).

The most widely recognized role of DNA methylation is in transcriptional repression; a vast amount of evidence exists demonstrating that transcription cannot be initiated from transcription start sites containing highly methylated CpG islands (Kass, Landsberger, and Wolffe 1997; Venolia and Gartler 1983; Li, Beard, and Jaenisch 1993; Walsh, Chaillet, and Bestor 1998; Siegfried et al. 1999). Recent studies have employed dCas9 fusion proteins to study methylation. Both Liu et al. and Stepper et al. targeted a dCas9-methyltransferase fusion protein to un-methylated promoter regions, and both observed silencing of their respective targets (Stepper et al. 2017; Liu et al. 2016). Methylation-mediated repression is likely to occur through numerous means including exclusion of transcriptional machinery, attraction of DNA-binding transcriptional blocker proteins and chromatin-modifying complexes that promote a heterochromatic environment (Nan et al. 1998; Watt and Molloy 1988).

In contrast to most gene promoters, bodies of active genes are methylated extensively, regardless of whether they contain CpG islands (Wolf et al. 1984). Unlike promoter regions, gene body methylation is usually associated with active transcription and gene expression (Larsen, Solheim, and Prydz 1993; Hellman and Chess 2007). However, the role of methylation within gene bodies is not completely understood, and several studies have proposed a potential role in splicing, given the observed differential methylation levels between exons and introns (Laurent et al. 2010; Shukla et al. 2011). Current studies are beginning to reveal increased versatility and complexity of DNA methylation. The simplistic observation that promoter methylation is associated exclusively with gene repression may not be so

clear cut; instances have been reported associating the signature with activation of gene expression, with several developmental transcription factors having demonstrated preferential binding to 5mC signatures via hydrophobic interactions (Yin et al. 2017). This implies that genomic location and context of 5mC signatures determines gene expression, rather than its presence.

1.4.2.1 The DNA Methyltransferase Family

The enzymes responsible for methylating DNA comprise the DNA methyltransferase (DNMT) family of proteins, which transfer methyl groups from the methyl donor, S-Adenosyl-Methionine (SAM) to cytosine residues (Figure 1.8A) (Chiang et al. 1996). Humans possess five different DNMTs, three of which have the ability to catalyse this reaction; these are DNMT1, DNMT3A and DNMT3B (Lyko 2018). The additional DNMTs, DNMT2 and DNMT3L, are non-canonical. DNMT2 for example, whilst lacking the ability to robustly catalyse DNA methylation, has been found to play a more predominant role in transfer RNA (tRNA) methylation (Goll et al. 2006). DNMT3L on the other hand is a catalytically inactive version of the DNMT3 enzymes, which acts as a cofactor for the latter to increase binding affinity to DNA (Gowher et al. 2005).

1.4.2.2 The Structure of DNMTs

The general structure of the DNMTs consists of a catalytic C-terminal and a regulatory N-terminal (Figure 1.8B). DNMT1 is the largest methyltransferase with a size of approximately 190kDa, and consisting of around 1600 amino acids (Zhang et al. 2015). The N-terminal domain is responsible for binding multiple factors

including transcriptional repressor, DNMT1 associated protein 1 (DMAP1). The resulting complex also binds HDAC2 and the three proteins cooperate to maintain a repressive, heterochromatic state at loci following DNA replication (Rountree, Bachman, and Baylin 2000). Multiple other proteins involved in heterochromatin formation also bind DNMT1 at this region, including proliferating cell nuclear antigen (PCNA), the histone methyltransferase, G9a and heterochromatin-binding protein β (HP1 β), demonstrating the collaborative means by which heterochromatin is maintained (Iida et al. 2002; Esteve et al. 2006; Fuks et al. 2003). In addition, a nuclear localisation signal (NLS) is also located at N-terminus covering amino acids 177-205, and ensures DNMT1 is correctly localised to the nucleus where it can bind DNA and perform its function (Alvarez-Ponce et al. 2018).

Another important domain in the N-terminus is the replication foci-targeting sequence (RFTS) domain. This domain is responsible for targeting DNMT1 to DNA replication foci where it methylates the newly synthesised daughter DNA strand (Leonhardt et al. 1992). When DNMT1 is not bound to DNA, the RFTS domain occupies the catalytic pocket, bound by Hydrogen bonding (Syeda et al. 2011). Only when substrate DNA of a length around 30bp is available can DNMT1 methylate DNA at full catalytic activity, by triggering exclusion of the RFTS domain from the site (Berkyurek et al. 2014). It is likely that this autoinhibitory mechanism plays a role in ensuring DNMT1 is correctly localised to sites requiring DNA methylation (Misaki et al. 2016).

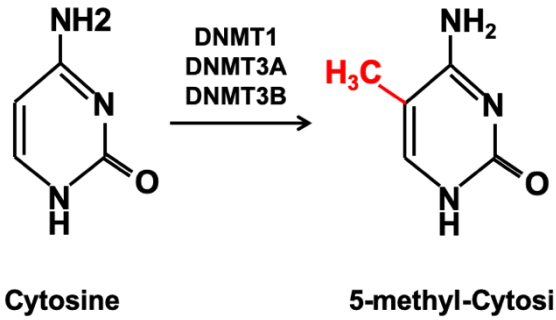
Adjacent to the RFTS domain is the CXXC domain. This domain consists of zinc finger motifs and specifically binds unmethylated CpG sites. It has been proposed that this specificity provides another DNMT1 autoinhibitory mechanism. Completely unmethylated CpGs are prevented from entering into the DNMT1 catalytic site by first binding the CXXC domain. Secondly, the subsequent positioning of the linker region between CXXC and its neighbouring BAH domain, blocks the catalytic pocket. This prevents DNMT1-mediated *de novo* methylation, and ensures that only hemi-methylated DNA can enter the catalytic site (Song et al. 2011). However, others have failed to see any impact on DNMT1 specificity upon loss of the CXXC domain, leaving this model unconfirmed (Bashtrykov et al. 2012).

Finally, prior to the catalytic C-terminal domain are dual bromo-adjacent homology (BAH) domains, which follow the CXXC domain. At present, the exact purpose of these domains is unclear. The first BAH domain potentially recognises and interacts with methylated histones (Takeshita et al. 2011), and the second appears to interact with both the catalytic site and substrate DNA (Song et al. 2012). Despite the ambiguity surrounding these regions, the preparation of enzymatically active DNMT1 has proven unsuccessful in their absence, indicating they are a critical component for this enzyme (Yarychkivska et al. 2018).

The DNMT3A and DNMT3B enzymes are smaller than DNMT1 with a size of approximately 100kDa each. They are structurally very similar, differing at a variable N-terminal region, which is ~60 amino acids smaller in DNMT3B (Okano, Xie, and Li 1998; Xie et al. 1999). In addition to a catalytic C-terminal, they possess two

additional, conserved domains, a PWWP domain and an ATRX-DNMT3-DNMT3L (ADD) domain. As the name suggests, the former contains a proline-tryptophan-tryptophan-proline motif at its centre, and is primarily involved in binding the enzymes to DNA (Chen, Tsujimoto, and Li 2004). The DNA binding affinity of DNMT3A however, is marginally lower than that of DNMT3B (Purdy, Holz-Schietinger, and Reich 2010), possibly contributing to the distinct biological functions of these proteins. The ADD domain on the other hand, primarily enables binding of these enzymes to unmethylated H3K4 residues, leading to methylation of DNA enriched with this epigenetic signature (Zhang et al. 2010). DNMT3L also possesses an ADD domain, however it lacks a functional catalytic domain, and consequently is incapable of methylating DNA itself. Rather, it behaves as a cofactor whose C-terminal interacts with the catalytic domains of DNMT3A and DNMT3B, resulting in increased catalytic activity of the latter (Suetake et al. 2004). Considerably different to the other DNMTs is DNMT2, which consists of approximately 400 amino acids, and is comprised of a catalytic domain and a smaller target-recognizing (TR) domain (Figure 1.8B) (Dong et al. 2001).

A.



Cytosine 5-methyl-Cytosine

B.

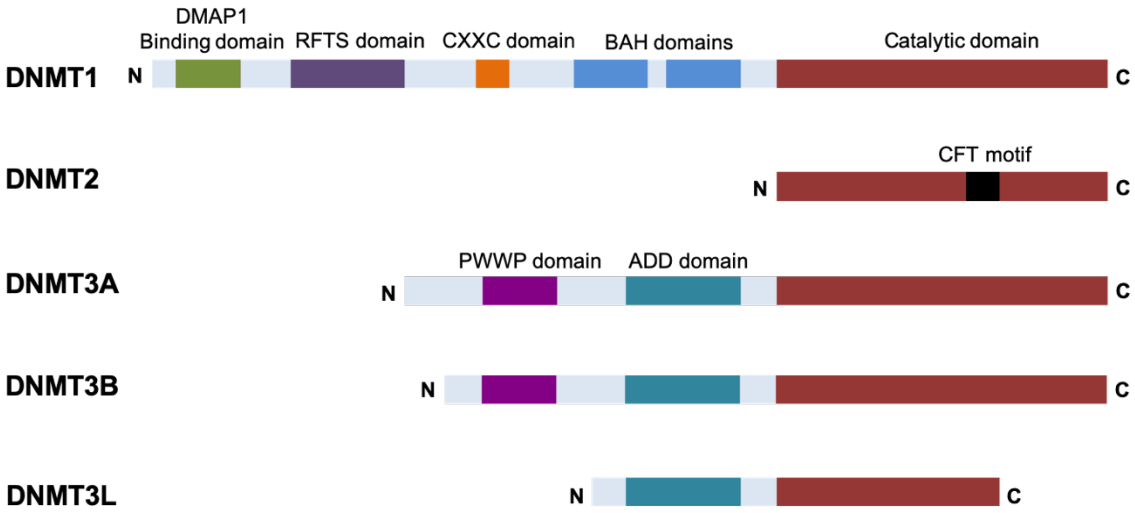


Figure 1.8. The DNA Methyltransferase Family. A. In eukaryotes, DNMT1/3A/3B have the ability to catalyse the addition of a methyl group to the fifth carbon atom of the cytosine in DNA, to form 5-methyl-cytosine (5mC). **B.** Five DNA methyltransferase family members are found in humans, which mostly consist of a catalytic C-terminal and a regulatory N-terminal. DNMT1 consists of a catalytic domain, dual bromo-adjacent homology (BAH) domains, a CXXC zinc finger domain, a replication foci targeting sequence (RFTS) domain, and a DNMT1-associated protein 1 (DMAP1) binding domain. DNMT2 consists only of a catalytic domain. The DNMT3 enzymes contain an ATRX-DNMT3-DNMT3L (ADD) domain and a Proline-Tryptophan (PWWP) domain. DNMT3L is catalytically inactive (Reproduced from Lyko 2018).

1.4.2.3 DNMT3A and DNMT3B

DNA methylation patterns across the genome are established by the *de novo* DNMTs, DNMT3A and DNMT3B. In the very early stages of development, DNA methylation is completely eradicated in primordial germ cells, resulting in an epigenetically “blank canvas”. This is first achieved by re-localisation of an oocyte-specific isoform of DNMT1 (DNMT1o) to the cytosol of cells where it is unable to bind and methylate DNA (Doherty, Bartolomei, and Schultz 2002). Second, a family of proteins, the ten-eleven translocation (TET) enzymes, catalyse DNA methylation reversal, converting 5-methyl-cytosine to 5-hydroxymethyl-cytosine (5hmC) (Rasmussen and Helin 2016). The DNMT3s then restore DNA methylation in a non-CpG specific and ubiquitous manner. They bind DNA at unmethylated H3K4 sites via their ADD domain and methylate surrounding DNA, whilst methylated H3K4 sites, in addition to other proteins, shield regions such as CpG islands and promoter sequences from being methylated (Otani et al. 2009).

It was previously accepted that DNMT3A and DNMT3B establish these initial methylation patterns, and that DNMT1 preserves them following DNA replication. Whilst that is largely true, in recent years it has become clear that this “black and white” concept is not as simple, and that there is crosstalk between maintenance and *de novo* methylation machineries. Researchers have observed that methylation of certain genomic regions is not maintained upon knock down of either of DNMT3A or 3B (Chen et al. 2003). It has been further suggested that these enzymes can anchor themselves to nucleosomes and potentially methylate CpGs missed by DNMT1 (Jeong et al. 2009).

Despite their similarities both structurally and functionally, DNMT3A and DNMT3B also possess some distinct roles. Deletion of each of these enzymes in mice resulted in different pathogenic phenotypes, with death ensuing at different developmental stages. Whilst pericentromeric heterochromatin is favourably methylated by DNMT3A, the methylation of minor centromeric satellite repeats is carried out exclusively by DNMT3B (Okano et al. 1999). Several studies have also suggested that gene body methylation may via targeting to H3K36me3 signatures may also be attributed to DNMT3B alone (Baubec et al. 2015; Duymich et al. 2016).

1.4.2.4 DNMT1

Following initial deposition of methylation marks by DNMT3A and 3B, DNMT1 primarily maintains post-replicative patterns in the long term, by preferentially binding hemi-methylated DNA, and methylating the newly synthesized daughter strand (Lyko 2018). At late S-phase of the cell cycle, DNMT1 is targeted to replication foci via its RFTS domain, a process dependent on an additional protein, ubiquitin-like with PHD and RING finger domains 1 (UHRF1) (Bostick et al. 2007). Immediately following DNA replication, UHRF1 specifically binds hemi-methylated CpG dinucleotides via its SET and RING-associated (SRA) domain, as well as the DNMT1 RFTS domain (Berkyurek et al. 2014). In addition, UHRF1 has also demonstrated an ability to simultaneously recognize both di- and tri-methylated H3K9 residues on DNA, via a tandem Tudor domain (Rothbart et al. 2012). This ultimately leads to inclusion of substrate DNA into the catalytic pocket of DNMT1 where it is methylated.

However, despite the knowledge that UHRF1 is absolutely required for maintenance methylation to occur, the precise mechanism by which UHRF1 and DNMT1 cooperate to achieve this remains elusive. In recent years, collective data has resulted in the proposal of two main possible models. The first model suggests that DNMT1 is targeted to hemi-methylated sites by ubiquitinated lysine residues present on histone H3. It was proposed that binding of UHRF1 to methylated H3K9 residues, triggers UHRF1-mediated mono-ubiquitination of three H3 lysine residues. DNMT1 is then recruited to DNA via its affinity for these ubiquitinated amino acids (Nishiyama et al. 2013; Ishiyama et al. 2017).

A second model argues that DNMT1 is recruited via several direct domain interactions between DNMT1 and UHRF1. A C-terminal linker region on UHRF1 has been reported to assist binding of the SRA domain with both the DNMT1 RFTS domain and hemi-methylated CpGs (Fang et al. 2016). This interaction appears important for the release of the RFTS domain from the catalytic site of DNMT1, consequently relieving the protein's auto-inhibition (Li et al. 2018). An interaction has also been reported between a ubiquitin-like (UBL) domain present within UHRF1 and the DNMT1 RFTS domain, which appears to be required for activation of DNMT1 catalytic activity (Foster et al. 2018). This model also acknowledges the ubiquitination activity of UHRF1, but rather, suggests that the protein ubiquitinates itself, generating an interface with which DNMT1 can bind (Vaughan et al. 2018). Ultimately, this model argues that UHRF1 binds DNA first, followed by DNMT1 in a more histone-independent manner.

Very recently, Mousli and colleagues put forward their perspective on this topic. They proposed that both models are likely correct and that a more complicated series of events occurs. They suggested that DNMT1 and UHRF1 work repetitively by permanently existing in close proximity as part of a larger macro-molecular complex. This complex travels along DNA until the UHRF1 SRA domain comes across hemi-methylation, which triggers a cascade of events, ultimately concluding with methylation of the new DNA strand. A structural change in UHRF1 is induced which activates dual-ubiquitination of itself as well as lysine residues on H3. Simultaneously, UHRF1 also binds H3K9me2/3. The DNMT1 RFTS subsequently interacts with one or both of these ubiquitinated residues, thus removing the RFTS from the catalytic pocket, and allowing DNMT1 to methylate the DNA. Until such observations can be confirmed *in vivo* however, this model cannot be confirmed (Figure 1.9) (Bronner et al. 2019).

1.4.2.5 DNA Methylation and Disease

Early experiments in mice demonstrated that DNMT1 is essential for cell viability. Despite DNMT1 knockout (KO) murine embryonic stem cells (mESCs) displaying no major growth abnormalities, when the enzyme was knocked out in embryos, severe aberrant development was observed, closely followed by death. (Li, Bestor, and Jaenisch 1992). Similar observations have been made in humans; a conditional DNMT1 KO cell line using colorectal carcinoma cells demonstrated that upon KO induction, cells initially arrest in the G2 phase of the cell cycle whilst displaying nuclear morphological defects, and eventually die ~7 days later (Chen et al. 2007). Conversely, cell death does not occur immediately upon KO of the DNMT3 proteins.

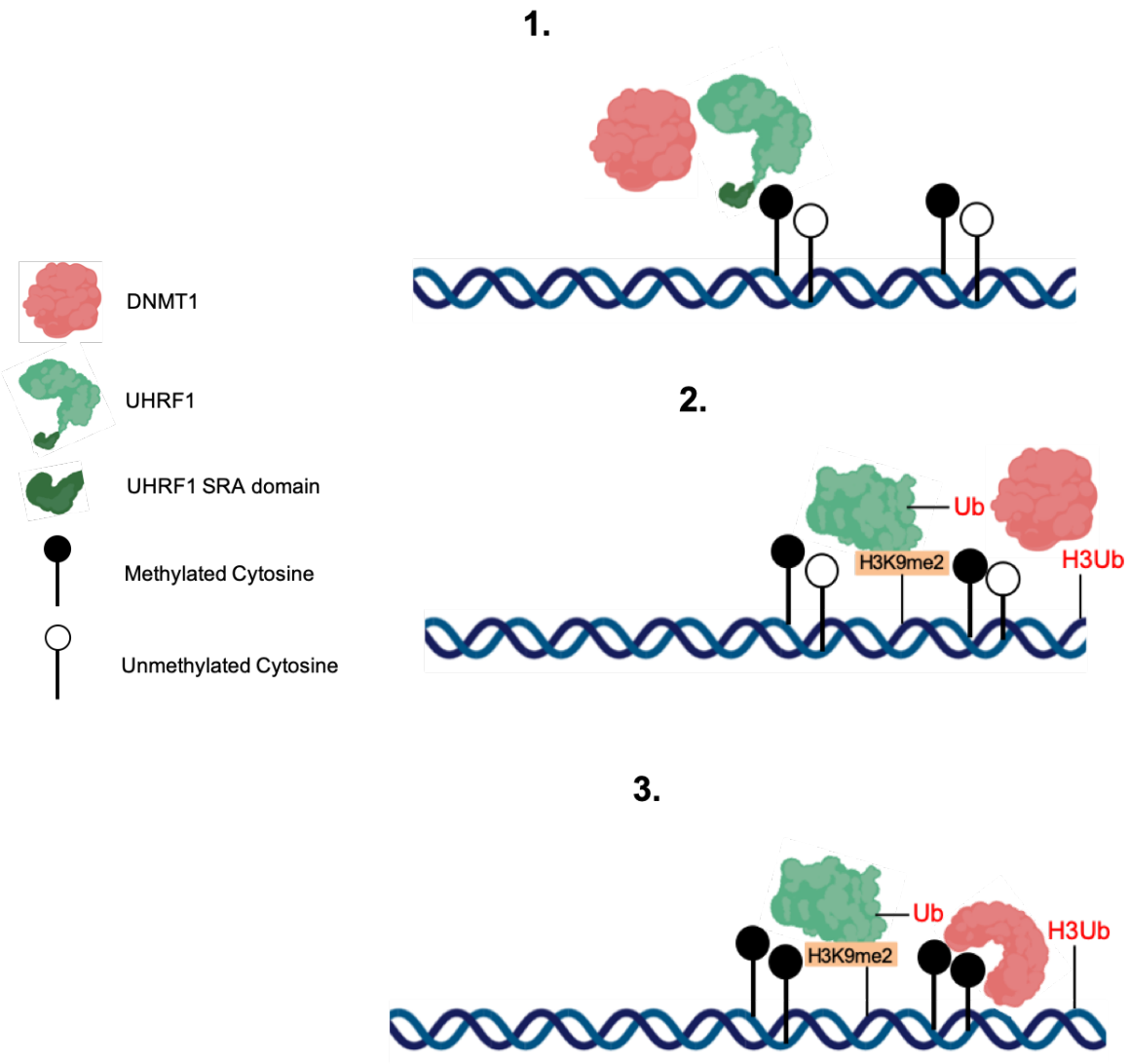


Figure 1.9. Proposed Model of UHRF1-Mediated DNMT1 Targeting to Hemi-Methylated CpGs. 1. UHRF1 and DNMT1 are in close proximity scanning post-replicative DNA until the UHRF1 SRA domain comes across a hemi-methylated CpG. **2.** This induces a structural change in UHRF1, which triggers UHRF1-mediated ubiquitination of itself and histone H3. UHRF1 also simultaneously binds di- and tri-methylated H3K9 residues. **3.** DNMT1 binds one or both of these ubiquitinated residues triggering a change in conformation. The RFTS domain is removed from the catalytic pocket, allowing DNMT1 to bind DNA and methylate the hemi-methylated CpG (Bronner et al. 2019).

Single and double KOs of DNMT3A and DNMT3B in human embryonic stem cells (hESCs) only mildly affected the cells, which retained >50% of normal methylation levels (Liao et al. 2015).

Abnormal DNA methylation is a common disease characteristic for pathologies involving genomic instability. For example, Immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome is an autosomal, recessive disorder characterised on a molecular level by genomic instability at several repetitive regions including pericentromeric heterochromatin. On a physiological level, ICF patients typically demonstrate delayed development, facial abnormalities and intellectual disability (Smeets et al. 1994). Studies have shown that in ~50% of cases, this disease is caused by mutations in the catalytic domain of DNMT3B that impact its methyltransferase activity (Xu et al. 1999). The remaining ~50% show mutations in additional components involved in DNMT3B-mediated methylation (Zhu et al. 2006). The loss of activity results in dysregulation of a number of genes, including several homeobox genes involved in craniofacial and neurological development (Jin et al. 2008). Mutations in DNMT3A are frequently associated with Tatton-Brown-Rahman syndrome (TBRS), a disorder that typically presents with abnormally fast growth, mental retardation and facial anomalies (Tatton-Brown et al. 2018).

DNMT1 on the other hand has been largely linked with hereditary sensory and autonomic neuropathy type 1E (HSAN1E). Patients with this neurological disorder display numerous symptoms including cognitive decline, seizures and hearing loss,

typically only presenting in late teens. Patient studies have identified various mutations in the RFTS domain of DNMT1, which cause this disorder (Klein et al. 2013; Baets et al. 2015). Interestingly, Klein and colleagues demonstrated that when such DNMT1 mutations are induced in cultured cells, DNMT1 is misfolded and subsequently re-localised to the cytoplasm, resulting in global hypomethylation. They speculate that this can cause aberrant gene activation which contributes to the pathogenicity of the disease (Baets et al. 2015).

Arguably, the disease most commonly associated with aberrant DNA methylation is cancer. Both hypo- and hyper-methylation of the genome have reportedly demonstrated contribution to tumorigenesis. Global hypomethylation can activate expression of numerous oncogenes, whilst hypermethylation of regulatory elements can result in silencing of tumour suppressor genes. At present, it is unclear precisely how these opposing effects on methylation can occur simultaneously within cancer cells (Morgan, Davies, and Mc Auley 2018).

Urokinase-type plasminogen activator (uPA) is an example of a protein commonly overexpressed in multiple cancers as a result of aberrant DNA hypomethylation. uPA expression is epigenetically regulated via methylation of its promoter; demethylation results in increased levels of uPA and often correlates with increased aggressiveness of cancer cells (Pakneshan, Tetu, and Rabbani 2004). As well as gene-specific methylation, hypomethylation of repeat sequences is also frequently observed in cancer cells. It is well-established that global hypomethylation is associated with genomic instability (Gaudet et al. 2003; Eden et al. 2003). De-

methylation of retro-transposable elements can contribute to tumorigenesis by inducing such instability (Daskalos et al. 2009). When retrotransposons such as long interspersed nuclear elements (LINEs) integrate at various positions in the genome, recombination events can lead to increased DNA strand breaks and disturbance of sequences can cause aberrant gene expression. Hypomethylation of repeat sequences such as these can trigger transcription, and consequently increase the chance of this kind of event happening (Daskalos et al. 2009).

On the other hand, hypermethylation of regions has also been widely observed in cancer cells, primarily at promoter regions of tumour suppressor genes. For example, the tumour suppressor protein, Src homology region 2 (SH2) domain–containing phosphatase 1 (SHP1), also known as Protein Tyrosine Phosphatase Non-Receptor Type 6 (PTPN6), is frequently downregulated in various cancer types. Inhibition of DNMT1 however, leads to hypomethylation of the corresponding gene and subsequent upregulation (Li, Liu, et al. 2017; Wang et al. 2017; Wen et al. 2018). Similar effects on other tumour suppressors such as Cyclin Dependent Kinase Inhibitor 2B (CDKN2B) (Herman et al. 1996; Yu et al. 2013), Retinoblastoma protein (RB1) (Stirzaker et al. 1997) and Cyclin Dependent Kinase Inhibitor 1A (CDKN1A) (Milutinovic et al. 2004; Schmelz et al. 2005) have also been observed. This has led to application of DNMT inhibitors such as 5-Azacytidine, 5-Aza-2'-deoxycytidine (Christman 2002) and Zebularine (Cheng et al. 2004) as cancer therapies. Moreover, a recent study analysed the epigenome of tumour cells in patients at different stages of chronic myeloid leukemia (CML). Interestingly, methylation and transcriptional patterns varied throughout disease progression,

meaning that analysis of the methylome could be used as a prognostic marker in the future (Heller et al. 2016).

1.5 Aims and Objectives

The Kanhere group is interested in disease-associated lncRNAs, particularly those associated with epigenetic regulation. Current data has led to the suggestion that *CCDC26* could act as a potential biomarker for diseases such as AML. *CCDC26* levels could be used as a means of risk stratification, and could one day potentially be used to monitor disease progression (Chen et al. 2019). However, despite these past studies collectively demonstrating clear correlations between cell viability and *CCDC26*, there has been no in depth, mechanistic analysis of this lncRNA. Potential targets of *CCDC26* have been suggested, including *mir-203*, KIT and IGF-1R, however, further exploration into the precise functional mechanism of *CCDC26* has not been conducted (Hirano et al. 2015; Cao et al. 2018; Wang, Hui, et al. 2018; Yan et al. 2019). Our group is interested in functionally analysing lncRNAs such as *CCDC26*. In recent years it has become clear that lncRNAs provide an additional, complex layer of genetic and cellular regulation. If we want to fully understand the complex molecular mechanisms of gene regulatory networks and other cellular systems, it is important to determine the functions and mechanisms of individual lncRNAs, such as *CCDC26*.

Aim:

To establish the functional significance of lncRNA, *CCDC26,* in myeloid leukemia cells, and mechanistically analyse its modes of action.

Objectives:

1. **To characterise *CCDC26* in a suitable cell line**

   We will adopt a dual data mining and experimental approach to confirm expression of *CCDC26*, establish which isoforms are present and where they are localised within the cell.

2. **Define and confirm the effects of a *CCDC26* knock out on cell viability**

   CCDC26 knock out cells will be functionally characterised and compared to WT cells to help elucidate its biological significance. Characteristics including cell growth, apoptosis and cell cycle will be analysed, as well as effects on global epigenetic modifications.

3. **Elucidate how *CCDC26* may be exerting such effects mechanistically**

   We will further investigate any changes to processes we observe in *CCDC26* KO cells compared to WT and attempt to determine the mechanistic role that *CCDC26* plays in these processes.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

Table 2.1 Buffer Compositions

SDS PAGE = sodium dodecyl sulphate polyacrylamide gel electrophoresis; WB = western blot; IF = immunofluorescence

Buffer	Application	Reagents
10X Running Buffer	SDS PAGE	60g Tris, 288g glycine, up to 2L dH_2O
1X Running Buffer	SDS PAGE	100ml 10X running buffer, 5ml 20% SDS, 895ml dH_2O
10% SDS-PAGE gel (X4)	SDS-PAGE	6.66ml 30% acrylamide, 5ml 1.5M Tris (pH8.8), 200ul 10% SDS, 200ul 10% APS, 20ul TEMED, 8ml H_2O
4% stacking for SDS-PAGE gel (X4)	SDS-PAGE	1.7ml 30% acrylamide, 2.5ml 0.5M Tris (pH6.8), 100ul 10% SDS, 100ul 10% APS, 20ul TEMED, 5.55ml H_2O
Blocking Solution	WB/IF	5% skimmed milk in TBS
TBS	WB	40ml 1M Tris (pH.8), 60ml 5M NaCl, 1900ml dH_2O
TBS-T	WB	40ml 1M Tris (pH.8), 60ml 5M NaCl, 2ml Tween 20, 1900ml dH_2O
1X Transfer Buffer	WB	200ml BIO-RAD TransBlot Turbo 5 x Transfer Buffer (Cat. 1704271), 200ml 100% Ethanol, 600ml nanopure H_2O
5X SDS Loading Dye	SDS	200mM Tris-HCL pH6.8, 40% glycerol, 4% SDS, 0.4% bromophenol blue, 200mM β-mercaptoethanol
Protein Lysis Buffer	Protein extraction	20mM Tris pH7.5, 150mM NaCl, 0.5% deoxycholic acid, 10mM EDTA, 0.5% Triton X-100
Buffer RLN	Cellular Fractionation	50mM Tris-HCl (pH.8), 1.5mM $MgCl_2$, 140mM NaCl, 0.5% NP-40
DNA FISH Wash Buffer	DNA FISH	50% formamide, 50% 2X SSC
FACS Cell Cycle Buffer	FACS Propidium Iodide staining	30µg propidium iodide (PI), 1% (w/v) sodium citrate, 0.1mM $NaCl_2$, 0.1% Triton X-100
RIP Buffer	DNMT1 RNA immunoprecipitation	25mM Tris, pH 7.4, 5mM EDTA, 150mM KCl, 0.5mM DTT, 0.5% NP40 Igepal, 100U/ml RNase inhibitor
Back Extraction Buffer	DNMT1 RNA immunoprecipitation	10mM Tris, pH8, 1mM EDTA, 100mM NaCl, 0.25% SDS
LB Broth	Bacterial culture	20g Lysogeny Buffer made up to 1L with H_2O, autoclaved

Table 2.2 Primary Antibodies

WB = western blot; IF = immunofluorescence; IP = immunoprecipitation; RIP = RNA immunoprecipitation; M = monoclonal; P = polyclonal

Antibody	Use	Working Dilution	Origin	Company	Catalog No.	Clonality
GAPDH	WB	1:1000	Mouse	ThermoFisher Scientific	MA5-15738	M
Nucleolin	WB	1:1000	Rabbit	Bethyl Laboratories. Inc.	A300-711A-M	P
NOL10	WB	1:1000	Rabbit	CUSABIO	CSB-PA015918 LA01HU	P
DNMT1	WB	1:1000	Mouse	NovusBio	NB100-56519	M
DNMT3A	WB	1:1000	Rabbit	NovusBio	NB100-265SS	P
DNMT3B	WB	1:1000	Rabbit	BioVision	3275	P
EZH2	WB	1:1000	Rabbit	Cell Signaling	#5246	M
HDAC2	WB	1:700	Rabbit	Santa Cruz	sc-7899	P
Phosphor-Tyr	WB	1:2000	Mouse	Cell Signaling	#9411	M
Acetyl-Lysine	WB	1:1000	Rabbit	Abcam	ab80178	P
Caspase 9	WB	1:1000	Rabbit	Cell Signaling	#9508	M
Caspase 3	WB	1:1000	Mouse	Cell Signaling	#9668	M
Caspase 8	WB	1:1000	Mouse	Cell Signaling	#9746	M
H3K27ac	WB	1:1000	Rabbit	Abcam	ab4729	P
H4K5ac	WB	Generated and kindly gifted by Dr John Halsall (Turner Group)				
H3K27me3	WB	1:1000	Mouse	Abcam	ab6002	M
H3K9ac	WB	Generated and kindly gifted by Dr John Halsall (Turner Group)				
H4K16ac	WB	1:1000	Rabbit	Abcam	ab109463	M
H3K9me3	WB	1:500	Rabbit	Millipore	07-442	P
Histone H3	WB	1:5000	Rabbit	Abcam	ab1791	P
c-JUN	WB	1:1000	Rabbit	Cell Signaling	#9165	M
DNMT1	IF	1:10	Mouse	NovusBio	NB100-56519	M
5mC	IF	1:250	Mouse	Epigentek	A-1014	M
Nucleolin	IF	1:30	Rabbit	Bethyl Laboratories. Inc.	A300-711A-M	P
γ-H2AX	IF	1:100	Mouse	Millipore	05-636-AF555	M
HDAC2	IF	1:50	Rabbit	Santa Cruz	sc-7899	P
Ki-67	IF	1:200	Mouse	Biolegend	35051	M
DNMT1	IP	/	Mouse	NovusBio	NB100-56519	M
DNMT1	RIP	/	Mouse	NovusBio	NB100-56519	M
IgG	RIP	/	Mouse	Sigma	18765-5MG	/

Table 2.3 Secondary Antibodies

WB = Western Blot; IF = Immunofluorescence

Antibody	Use	Working Dilution	Company	Catalog No.
Anti-rabbit IgG (H+L) (DyLight™ 800 4X PEG Conjugate)	WB	1:10,000	Cell Signalling	5151
Anti-mouse IgG (H+L) (DyLight™ 680 Conjugate)	WB	1:10,000	Cell Signalling	5470
Alexa Fluor 488 donkey anti-rabbit IgG (H+L)	IF	1:500	Invitrogen	A21206
Alexa Fluor 633 goat anti-mouse IgG (H+L)	IF	1:500	Invitrogen	1010093

Table 2.4 BACs for DNA FISH Probes (Source Bioscience)

BAC	Region of hybridisation	Length	Antibiotic	Incubation time at 15ºC for nick translation reaction
RP11-770K21	CCDC26	189,124bp	Chloramphenicol, 30ug/ml	60 minutes
RP4-735C1	Chr.1p13.3	133,451bp	Kanamycin, 50ug/ml	15 minutes
RP11-1072A3	Chr.16p11.2	136,630bp	Chloramphenicol, 30ug/ml	15 minutes

Table 2.5 List of Primers

All primers were purchased from Sigma and are shown 5'-3'

Primer	Forward	Reverse
CCDC26_All (all isoforms)	ATGGAAAGATTGTGCCTGCAG	CTCGATCTTTCCCAGTGTGG
CCDC26_Set1-2 (isoforms 1 and 2)	TCAGGCAACTGCAGAGTCTTAG	ACCCAGGCTTGTCTCATCTC
CCDC26_Set3 (isoform 3)	AGATCAGCTATGAAGGCCTGAG	CTCGATCTTTCCCAGTGTGG
CCDC26_Set4 (isoform 4)	TTCAAGAATGGCCTTTTAAAGGACC	CTCGATCTTTCCCAGTGTGG
GAPDH	GGAGCGAGATCCCTCCAAAAT	GGCTGTTGTCATACTTCTCATGG
U105	CCCTATCTCTCATGATGAAC	CCCATCTCTTCTTCAGAGCG
DNMT1	GCGGTATACCCACCATGACA	AGGCTTTGCCGGCTTCC
DNMT3A	GTTGTGAGAAGGAATGGGCG	TTGGCTTTCTTCTCAGCCGTAT
DNMT3B	CACTCTGTCCTGGGTGCTG	GTCTCCCTTCATGCTTTCCAAG
PTPN6	AACAGCCGTGTCATCGTCAT	ATCAGGTCTCCATTGTCCAGC
CDKN2B	TGGGGGCGGCAGCGATGAG	AGGTGGGTGGGGGTGGGAAAT
CDKN1A	AGCATGACAGATTTCTACCACTC	GATGTAGAGCGGGCCTTTGA
CD9	CATCTGTATCCAGCGCCAGG	CCGGCAAGCCAGAAGATGAA
VAV1	CTTACGGAGCTGGTGGAGTT	ACTTTGTGCTTCCCACTGCT
JUNB	CTGCTGGAAACAGACTCGATTC	CCACAGTACGGTGCAGAGAG
IGF1	TGCTCTCAACATCTCCCATCTC	TGGTGTGCATCTTCACCTTCA
Actin B	TACTCCTGCTTGCTGATCCA	GATCATTGCTCCTCCTGAGC
Genomic DNA control	TGAGCAGCCTGTCTCTCTGA	GGTGCCTCCACTTCAACATT

Table 2.6 siRNAs

siRNA	Company	Catalog No.	ID	Target sequence
MISSION siRNA Universal Negative Control	Sigma	SIC001	-	-
DNMT1 MISSION esiRNA	Sigma	EHU060601	HU-06060-1	GAGCCACAGATGCTGACAAATG AGAAGCTGTCCATCTTTGATGCC AACGAGTCTGGCTTTGAGAGTTA TGAGGCGCTTCCCCAGCACAAAC TGACCTGCTTCAGTGTGTACTGTA TGACCTGCTTCAGTGTGTACTGTA AGCACGGTCACCTGTGTCCCATC GACACCGGCCTCATCGAGAAGAA TATCGAACTCTTCTTTTCTGGTTCA GCAAAACCAATCTATGATGATGA CCCATCTCTTGAAGGTGGTGTTAA TGGCAAAAATCTTGGCCCCATAAA TGAATGGTGGATCACTGGCTTTGAT GGAGGTGAAAAGGCCCTCATCGG CTTCAGCACCTCATTTGCCGAATAC ATTCTGATGGATCCCAGTCCCGAGT ATGCGCCCATATTTGGGCTGATGCA GGAGAAGATCTACATCAGCAAGAT TGTGGTGGAGTTCCTGCAGAGCAAT TCCGACTCGACCTATGAGGACCTGA TCAACAAGATCGAGACCACGGTTCC TCCTTCT

2.2 Cell Culture

Table 2.7 Cell culture growth media for K562 cells

Growth Media for K562 cells	
RPMI 1640 (GIBCO, ThermoFisher Scientific)	500ml
Fetal Bovine Serum (GIBCO, ThermoFisher Scientific)	50ml (10%)
Antibiotic-Antimycotic (100X) Contains 10,000 units/ml penicillin, 10,000ug/ml streptomycin, 25ug/ml Amphotericin B (GIBCO, ThermoFisher Scientific)	5ml (1%)

Table 2.8 Cell culture freezing media

Freezing Media for K562 cells	
Fetal Bovine Serum (GIBCO, ThermoFisher Scientific)	25ml
Dimethyl sulfoxide (DMSO)	3ml

2.2.1 Culturing K562 Cells

K562 cells were maintained in an undifferentiated state in either 75cm^2 or 25cm^2 Greiner culture flasks (Sigma), at 37^{o}C, 5% CO_2. Cells were monitored daily, passaged (split) every ~48 hours and seeded at 2.5 x 10^5 cells in growth media. A hemocytometer was used to observe and count cells for seeding. Cell centrifugations were always performed at 1200rpm, for 5 minutes at room temperature unless stated otherwise.

2.2.2 Long-term Storage of K562 Cells in Liquid Nitrogen

K562 cells were stored long-term by centrifuging ~8-9 million cells, removing the media and resuspending in 3ml freezing media. The suspension was then aliquoted into 3 x 1ml cryotubes, which were frozen overnight at -80^{o}C, before being finally stored in liquid nitrogen.

2.2.3 Thawing K562 cells

Cells retrieved from liquid nitrogen were thawed rapidly at 37^{o}C and re-suspended in 2ml growth media, added drop-wise over ~5 minutes. Cells were centrifuged, and washed with media again. Following a second centrifugation, cells were re-suspended in 1ml media and added to 4ml media in a 25cm^{2} flask. Cell growth and maintenance was then followed as described in 2.2.1.

2.2.4 Cell Growth Curves

10^{5} cells/ml were seeded into a 6 well-plate and viable cells were counted after 24, 48 and 72 hours. Trypan Blue dye exclusion was used to distinguish viable cells. Before counting, 10ul cell suspension was mixed with 10ul 0.4% Trypan Blue solution (ThermoFisher, Cat: 15250061). 10ul of this mix was then applied to a hemocytometer and cells were counted, excluding those that appeared blue.

2.2.5 Inhibitor and Drug Treatments

Cells were seeded at a density of 2.5 x 10^{5} cells/ml and grown in the presence of various inhibitors and drugs. All inhibitors were prepared in a sterile tissue culture hood, and all solvents were filter sterilized before preparing stocks. Control cells were grown in the presence of the equivalent volume of solvent in which each inhibitor was dissolved, such as DMSO or H_2O (E.g. if 3ul of an inhibitor dissolved in DMSO was added to cells, then 3ul DMSO alone was added to control cells). Upon harvesting after treatment, cells were washed in PBS three times, centrifuging between washes.

Cells were grown in growth media containing a final concentration of either 0uM (control), 5uM or 10uM cisplatin (Millipore, Cat: 232120) for 24 hours, before harvesting. A 10mM cisplatin stock was freshly prepared in H_2O for each use.

Cells were grown in growth media containing a final concentration of either 0uM (control), 5uM or 10uM 5-Aza-2'-deoxycytidine (DAC) for 48 hours, before harvesting. A 220mM DAC stock was freshly prepared in DMSO for each use.

Cells were grown in growth media containing a final concentration of either 0uM (control) or 10uM MG132 for 24 hours before extraction of whole cell lysates (2.7.1). A 10mM MG132 stock was prepared in DMSO and stored at -20^{o}C.

HDAC inhibitor stocks of 10mM suberoylanilide hydroxamic acid (SAHA) (Cayman Chemicals, Cat: 10009929) in DMSO, and 100mM Valproic acid (VPA) (Sigma, Cat: P4543-256) in H_2O, were prepared fresh on the day of use. Cells were grown in growth media containing a final concentration of either 0uM (control) or 10uM SAHA, or 0uM (control) or 4mM VPA for 24 hours before harvesting.

Cells were grown in growth media containing a final concentration of 50ug/ml cyclohexamide (CHX) (Sigma, Cat: C7698) for 12 hours, before extraction of whole cell lysates (2.7.1). A 20mg/ml CHX stock was prepared in ethanol and stored at -20^{o}C.

2.2.6 Transient siRNA Transfection

All cell transfections were similarly performed via electroporation using the Cell Line Nucleofector Kit V (Lonza, Cat: VVCA-1003). For each transfection, 10^6 cells were centrifuged and washed once in 500ul PBS. For DNMT1 siRNA knock-down experiments, a nucleofector solution was prepared containing 100ul solution V and either 60pmol negative control siRNA or 60pmol DNMT1 siRNA (See Table 2.6 for siRNA details). The cell pellet was resuspended in the nucleofector solution and transferred into an electroporation cuvette. The cuvette was placed into a Nucleofector 2b Device (Lonza, Cat: AAB-1001) and the contents electroporated using the T-16 setting, specifically designed for K562 cells. The electroporated cells were then gently pipetted into 2.5ml growth media in a 6-well plate. DNMT1 siRNA KD cells were grown for 24 hours before harvesting.

2.2.7 Growing *CCDC26* Overexpression Cells

Stable *CCDC26* overexpression (O/x) cell lines were prepared by previous students, by ligating a spliced *CCDC26* transcript (isoform 1) into a modified pEF6 plasmid (ThermoFisher #V96220, myc epitope and His-tag were removed previously). Briefly, the resulting plasmid, as well as an empty control pEF6 plasmid lacking *CCDC26*, were used to transfect both WT and *CCDC26* KO cells via electroporation, which were subsequently selected with Blasticidin S HCl (TOKU-E B001). For use in this project, these cell lines were grown as described 2.2.1, in growth media supplemented with a final concentration of 4ug/ml Blasticidin S HCl. A 25mg/ml Blasticidin S HCl stock was prepared in H_2O and stored at -20^oC. *CCDC26* overexpression was confirmed via qRT-PCR, as described in 2.6.4.

2.2.8 Fluorescence Activated Cell Sorting (FACS) Cell Cycle Analysis

For cell cycle analysis, ~300,000 cells were centrifuged in FACS tubes and the growth media poured off. 300ul cold, cell cycle buffer (30µg propidium iodide (PI), 1% (w/v) sodium citrate, 0.1mM $NaCl_2$, 0.1% Triton X-100) was added to the pellet, and gently vortexed, before storing at 4°C for ~24 hours in the dark, to allow PI staining to occur. DNA content of cells was then analysed in triplicate by flow cytometry, to measure the percentage of cells at each stage of the cell cycle. This was performed using a BD FACS Calibur machine, and analysis conducted using the BD Cell Quest software (BD Biosciences).

2.3 DNA Fluorescence *in situ* Hybridisation (FISH)

Table 2.9 DNA FISH Hybridisation Mixture

Component	Quantity (1 x 10ul reaction)
Formamide (Sigma, Cat: F9037)	5ul
50% dextran sulphate	2ul
20 x SSC	1ul
Cot-1 DNA (Roche, Cat: 11581074001)	1ul
DNA FISH probe	2ng/ul (final concentration)

2.3.1 Generation of DNA FISH Probes

2.3.1.1 Bacterial Artificial Chromosome (BAC) Preparation

BACs spanning the desired genomic regions were purchased from Source Bioscience (further details in Table 2.4). These were prepared by first streaking onto agar plates containing the appropriate antibiotic for selection using sterile loops, and incubating overnight at 37°C to obtain single colonies. Single colonies were

selected and used to inoculate 3ml LB broth (20g Lysogeny Buffer in 1L H_2O, autoclaved) (containing appropriate antibiotic), and incubated overnight at 37°C, with agitation. The following day, 150ml LB broth (containing appropriate antibiotic) was inoculated with 500ul of the overnight cultures in a sterile, 500ml conical flask. This was again grown overnight at 37°C, with agitation and was followed by midi-prep BAC extraction as described below (2.3.1.2).

2.3.1.2 Midi-Prep

All midi-preps were performed using the NucleoBond Xtra Midi Plus Kit (Macherey-Nagel, Cat: 740412.10) following the manufacturers guidelines. Briefly,

the OD_{600} of the 150ml overnight culture was measured to determine the recommended culture volume to use (Volume (ml) = $400/OD_{600}$). This volume of cell culture was pelleted by centrifuging at 7000rpm for 10 minutes at 4°C and the supernatant was discarded. The pellet was re-suspended completely in 8ml Buffer RES before adding 8ml Buffer LYS and gently inverting to lyse the cells. The mix was incubated at room temperature for 5 minutes and then mixed with 8ml Buffer NEU and inverted to neutralize the suspension. This was then added to a NucleoBond Xtra Column, which had been previously equilibrated with 12ml Buffer EQU. The column and filter were washed with 5ml Buffer EQU before the column filter was removed. The column only was then washed with 8ml Buffer WASH, followed by eluting the DNA with 5ml Buffer ELU which was collected in a 15ml falcon. 3.5ml isopropanol was added and vortexed vigorously to precipitate the eluted DNA, and then centrifuged at 12,000rpm for 30 minutes at 4°C. The supernatant was discarded and 2ml 70% ethanol was added before centrifuging at

12,000rpm for 5 minutes at room temperature. The ethanol was carefully removed from the pellet, which was allowed to air dry completely at room temperature. Finally, the pellet was dissolved in 30ul dH_2O.

2.3.1.3 Synthesis of Amine-Modified DNA and Labelling with Fluorescent Dye

DNA FISH probes were generated by nick translation, using the Invitrogen FISH Tag DNA Multicolor Kit (Cat No. F32591) following the kit protocol. First, nick translation reactions were performed on ice to generate BAC DNA fragments of length 200-700bp (5ul 10X nick translation buffer, 5ul 0.1M DTT, 5ul 10X DNA nucleotide mix, 1ug BAC DNA, 1.7ul DNA polymerase I, 3ul DNase I, H_2O up to 50ul). Reactions were performed at 15^oC for varying incubation times that were optimised for different sized BACs (Table 2.4) (Appendix I, Figure 1). Following incubation 50ul H_2O was added to the mix and vortexed thoroughly to inactivate the DNase I.

The resulting amine-modified DNA was then column-purified by adding 400ul of binding buffer and mixing well. The entire volume was added to a spin column and centrifuged at maximum speed for 1 minute. The flow-through was discarded and the column was washed with 650ul of wash buffer. The column was centrifuged as before and the flow through discarded, followed by an additional dry centrifugation to remove any residual buffer. The column was placed into a fresh eppendorf and the DNA was eluted in 55ul elution buffer by centrifugation.

The column-purified DNA was then ethanol precipitated, by adding 10ul 3M sodium acetate (pH 5.2), 1ul glycogen, 39ul H_2O and 250ul 100% ethanol, and incubating at -20^oC overnight. The following morning the sample was centrifuged at maximum speed for 10 minutes and the supernatant removed. The resulting pellet was rinsed with 400ul 70% ethanol and centrifuged twice. All traces of ethanol were then removed and the sample was allowed to air dry. Finally, 5ul H_2O was added and incubated at 37^oC for 5 minutes to fully resuspend the DNA.

Next, the amine-modified DNA was labelled with fluorescent dye. First, the 5ul DNA suspension was denatured by heating at 96^oC for 5 minutes, followed by cooling on ice for 3 minutes and centrifuging at maximum speed for 3 minutes. 3ul sodium bicarbonate and 2ul reactive dye (in DMSO) was added and vortexed well. The labelling reaction was then incubated at room temperature for 1 hour in the dark, followed by addition of 90ul H_2O. The fluorescent dye-labelled DNA was then column purified and ethanol precipitated as described previously, except for being finally resuspended in 10ul H_2O. Fluorescent probes were stored in the dark at -80^oC.

The CCDC26, Chr.1p13.3 and Chr.16p11.2 probes were labelled with the AlexaFluor 594, 555 and 647 fluorescent dyes respectively.

2.3.2 *In Situ* Hybridisation

2.3.2.1 Cell Fixation

4 x 10^6 cells were harvested and centrifuged at 1200rpm for 5 minutes at room temperature. Growth media was removed from the pellet, which was then washed in 1ml PBS and centrifuged again as before. The PBS was removed and the pellet was re-suspended in 50ul PBS. 1ml of a 3:1 mix of Methanol:Acetic Acid fixative was added, gently mixed and incubated at room temperature for 10 minutes exactly. This was followed immediately with centrifugation, followed by three washes with 1ml PBS. Fixed cells were stored in 3ml 70% Ethanol at 4^{o}C at least overnight before use, and stored at 4^{o}C for no longer than 3 months.

2.3.2.2 Slide Preparation and Hybridisation

Fixed cells were inverted to resuspend any pellet that may have formed during storage. 400ul of cells were aliquoted per slide into sterile microcentrifuge tubes and centrifuged for 5 minutes at 7000 rpm at room temperature. The supernatant was removed and the pellet was re-suspended in 200ul PBS. Samples were then spun down onto microscope slides via cyto-centrifugation for 7 minutes at 350rpm. The slides were dehydrated through an ethanol series consisting of 70%, 90% and 100% ethanol for 4 minutes each. After air-drying, 10ul of Hybridisation mix (Table 2.9) was pipetted directly onto the slide and temporarily sealed with a 22mm x 22mm coverslip. The slides were optimally heated on a heat block for 3 minutes at 95^{o}C (Appendix I, Figure 2) and then incubated overnight at 37^{o}C.

2.3.2.3 Post-Hybridisation Washes

The following morning, coverslips were removed from slides gently using clean forceps, taking care not to shear cells. The slides were washed in DNA FISH Wash Buffer (50% formamide, 50% 2X SSC) at 45^{o}C for 20 minutes three times, followed by washing twice in 2X SSC at 45^{o}C for 20 minutes. The slides were then rinsed in 4X SSC at room temperature and allowed to dry. 10ul of SlowFade® Gold anti-fade reagent (Invitrogen DNA FISH kit) and 1ug/ml DAPI were added to slides which were then covered with a coverslip and sealed with nail polish.

2.3.2.4 Confocal Microscopy

Slides were imaged using a Leica TCS SP8 Confocal microscope, using either 20X, 40X or 63X objectives. For 3-dimensional analysis, the 12kHz Tandem scanner alongside the SuperZ Galvo stage allowed acquisition of accurate 3D stacks, collected at 1.4um intervals through the Z-axis.

2.3.2.5 Calculating the 3D Distance Between Signals

Calculations for the distance between two probes in three dimensions were performed by first manually segmenting the hybridisation signals. The X, Y and Z coordinates of the centroids of each were established using the “spot detector” function in the bioimage informatics software, *Icy*. Distance was then calculated using the equation below:

$$\text{Distance} = [(X_1-X_2)^2 + (Y_1-Y_2)^2 + (Z_1-Z_2)^2]$$

This was automated using a script written in R (Box 1). For example, if measuring the distances between CCDC26 (green signal) and Chr.1 (red signal), all X coordinates for CCDC26 signals in a Z-stack image were written into a vector in R (labeled ‘A’), and all X coordinates for Chr.1 signals for that same Z-stack image were written into a separate vector in R (labeled ‘B’). These two vectors were then put into the script detailed in Box. 1. The script outputs a matrix containing the distance between every CCDC26 X coordinate and every Chr.1 X coordinate. This was repeated for Y and Z coordinates, resulting in the generation of three matrices (X, Y and Z). Values in each matrix were then squared, and the values in the three resulting matrices were added together in R, to give a single matrix of values. Finally, the square root of the values in this matrix were calculated to give a final matrix containing the 3D distance between every CCDC26 signal and every Chr.1 signal. Each row represented the distance between a single CCDC26 signal and every other Chr.1 signal, whilst each column represented the distance between a single Chr.1 signal and every other CCDC26 signal. The smallest distance in either each row or each column indicated the minimum distance between each CCDC26 signal and a Chr.1 signal. All minimum distances for a matrix were returned as a vector using the ‘apply’ function in R.

Minimum distances between probes were calculated for multiple nuclei within a Z-stack, simultaneously. To ensure the nearest signal to a particular probe was not in an adjacent cell, images containing nuclei separated by large distances were manually selected for use. Minimum distances were calculated for at least 500 cells, per replicate, for each cell line.

Box 1. R script to calculate the distances between all Chr1 fluorescent signals and all CCDC26 fluorescent signals in each cell, following the input of Chr1 X coordinates (A) and CCDC26 X coordinates (B).

```
X_MATRIX<-function(A,B){
  rows<-length(A)
  cols<-length(B)
  M<-matrix(0,rows,cols)
  for(i in 1:rows){
    for(j in 1:cols){
      M[i,j]<-((A[i]-B[j])*(A[i]-B[j]))
    }
  }
 return(M)
}
```

2.4 Immunofluorescence

2.4.1 Slide Preparation

Cells were fixed in Methanol: Acetic Acid, as described in 2.3.2.1. 200ul cells per slide were aliquoted into sterile microcentrifuge tubes, which were centrifuged for 5 minutes at 7000 rpm, room temperature. The supernatant was removed and the pellet was re-suspended in 200ul PBS. Samples were then spun down onto microscope slides via cyto-centrifugation for 7 minutes at 350rpm.

2.4.2 Blocking and Primary Antibody

Cells were blocked by pipetting 50ul of 5% Bovine serum albumin (BSA) (Promega, Cat: W3841) directly onto slides, and incubating for 1 hour at room temperature, followed by a brief wash in PBS. 50ul of primary antibody diluted in 5% BSA (See Table 2.2 for a list of antibody working dilutions) was then pipetted onto each slide

and temporarily covered with a 22mm x 22mm coverslip. Slides were placed into a moist chamber and incubated overnight at 4°C. For 5mC immunofluorescence, slides were heated at 94°C for 3 minutes prior to loading primary antibody, in order to separate DNA strands to allow binding.

2.4.3 Secondary Antibody

The following morning, coverslips were removed and slides were carefully washed three times in PBS. Approximately 80ul of the appropriate secondary antibody, diluted in PBS (See Table 2.3 for list of antibody working dilutions) was then pipetted onto slides, sealed temporarily with a coverslip and incubated at room temperature for 1 hour in the dark. This was followed by three PBS washes and a final single wash in H_2O. After allowing slides to dry, 10ul of SlowFade® Gold anti-fade reagent (Invitrogen DNA FISH kit) and 1ug/ml DAPI were added to the slide which was then covered with a coverslip and sealed with nail polish.

2.4.4 Confocal Microscopy

Slides were imaged using a Leica TCS SP8 Confocal microscope as described in 2.3.2.4. The brightfield microscope setting was used to visualize cell outlines.

2.5 Nucleolar AgNOR Staining

Table 2.10 Solutions for AgNOR Staining

Solution	Components
Solution A (50% Silver Nitrate)	50g Silver Nitrate, 100ml dH_2O
Solution B (2% Gelatin)	2g Gelatin, 100ml dH_2O
Solution C (5% Sodium Thiosulfate)	5g Sodium thiosulfate, 100ml dH_2O

Approximately 500,000 cells were harvested per slide by centrifuging for 5 minutes at 1200rpm, washing in PBS followed by centrifuging again as before. The resulting pellet was re-suspended in 200ul PBS and cyto-centrifuged onto a microscope slide for 7 minutes at 350rpm. Slides were allowed to dry completely before proceeding.

10ul of formic acid was added to 1ml Solution B (1%) (Table 2.10) and then mixed with 2ml Solution A (2:1 ratio, Solution A: Solution B). Approximately 300ul of this working solution was pipetted onto slides and incubated at room temperature, in the dark for 30 minutes. Slides were washed three times in ddH_2O and allowed to air dry. 300ul of Solution C (Table 2.10) was pipetted onto slide and incubated at room temperature in the dark for 10 minutes, before washing three times as before and allowing to air dry. The slides were then dehydrated through an ethanol series (70%, 90% and 100% EtOH, for 2 minutes each). Finally, the slides were mounted with DEPEX and sealed with 22mm x 22mm coverslips and nail varnish. Slides were imaged using a Nikon Eclipse E600 Wide-field microscope.

2.6 RNA Manipulation

An RNase free-environment was maintained for all RNA work by treatment of equipment and work surfaces with RNase *Zap*, RNase decontamination solution (ThermoFisher Scientific, Cat. No. AM9780).

2.6.1 RNA Extraction

All RNA was extracted using the QIAGEN RNeasy Mini Kit (Cat: 74106), following manufacturers guidelines as described below (2.6.1.1). RNA concentrations were

determined using a NanoDrop ND-100 spectrophotometer (Thermo Scientific).

2.6.1.1 Total RNA Extraction

Briefly, ~5 x 10^6 cells were harvested for total RNA extraction. The cell pellet was re-suspended and lysed in 350ul Buffer RLT (containing 1% β-mercaptoethanol) and further homogenized using a 1ml syringe and needle (21G x 1.5" – Nr.2., 0.8mm x 40mm). The lysate was centrifuged for 3 minutes at maximum speed in order to remove cell debris, and the supernatant was removed and mixed with 1 x volume 70% ethanol. The sample was next transferred to a RNeasy mini spin column placed within a collection tube and centrifuged for 20 seconds at maximum speed. This was followed by column washes with 700ul Buffer RW1, 20 seconds centrifugation, 500ul Buffer RPE, 20 seconds centrifugation, 500ul Buffer RPE and a 2-minute centrifugation, discarding flow through between washes. A final dry centrifugation step was performed for 1 minute to dry the membrane, after which the RNA was eluted in 30ul RNase free water (warmed to ~50^{o}C). RNA quality was determined by running ~200-300ng on a 1% agarose gel.

2.6.1.2 Extraction of Nuclear and Cytosolic RNA Fractions

For extraction of nuclear and cytosolic RNA fractions, approximately 10^7 cells were harvested by centrifuging at 1200rpm at room temperature for 5 minutes, washing in 1ml PBS and centrifuging again. As much supernatant as possible was removed from the cell pellet, which was then re-suspended in 1ml of cold (4^{o}C) Buffer RLN (50mM Tris-HCl (pH.8), 1.5mM $MgCl_2$, 140mM NaCl, 0.5% NP-40, 100U/ml RNase inhibitor) and incubated on ice for 5 minutes. 250ul of the mix was then pipetted into

four eppendorfs labelled “nuclear”. These were centrifuged for 2 minutes at 3700rpm, 4°C. Approximately 500ul of supernatant collected from all four eppendorfs was transferred into a fresh eppendorf, labelled “Cytosolic”, taking extra care not to disturb the nuclear pellets. Each of the four nuclear pellets were washed twice with 100ul Buffer RLN, centrifuging for 2 minutes at 3200rpm, 4°C between washes. 200ul and 600ul Buffer RLT (containing 1% β-mercaptoethanol) was added to nuclear and cytosolic fractions respectively and vortexed vigorously. The nuclear pellets were further homogenized using a 1ml syringe and needle (21G x 1.5” – Nr.2., 0.8mm x 40mm). The fractions were then centrifuged for 3 minutes at maximum speed to remove cell debris, and the supernatants were removed and mixed with 1X volume 70% ethanol. The extraction then followed the RNA extraction protocol described above (2.6.1.1)

2.6.1.3 DNase I Treatment of RNA

All freshly extracted RNA was treated with DNase I using the Sigma, amplification grade DNase I kit (Cat: AMPD1), following the kit guidelines. 16ul of extracted RNA was added to 2ul of 10x reaction buffer and 2ul DNase I, and incubated for 15 minutes at room temperature. 2ul of the stop solution was then added to bind Calcium and Magnesium ions and to inactivate DNase I, before heating at 70°C for 10 minutes to denature the DNase I and the RNA.

2.6.2 Complementary DNA (cDNA) Synthesis by Reverse Transcription

RNA was converted into cDNA using the Bioline-Tetro cDNA Synthesis Kit (Cat. No. BIO-65042) following the manufacturers guidelines. 2-5ug of extracted RNA (eluted

in H_2O and treated with DNase I) was mixed with the following kit components: 1ul random hexamer primer, 1ul 10mM dNTP mix, 4ul 5X RT Buffer, 1ul RiboSafe RNase Inhibitor, 1ul Reverse Transcriptase and DEPC-treated water up to a total volume of 20ul. The components were mixed gently by pipetting and incubated at room temperature for 10 minutes. This was followed by incubation at 45°C for 30 minutes before terminating the reaction by heating at 85°C for 5 minutes. Samples were then chilled on ice briefly and stored at -20°C.

2.6.3 Real-Time Polymerase Chain Reaction (RT-PCR)

In order to check the quality of cDNA generated by reverse transcription, and to ensure no genomic DNA contamination, real-time polymerase chain reaction (RT-PCR) was performed on cDNA samples, using primers specific for actin B and genomic DNA (primer sequences are detailed in Table 2.5). Approximately 100ng cDNA was added to 25ul 2x MyTaq Red Mix (Bioline, Cat. BIO-25043), 2ul forward primer (10uM) and 2ul reverse primer (10uM) and made up to 50ul with dH_2O for each reaction. The PCR was then performed under the conditions outlined in Box. 2 using a bench-top Primus thermocycler.

Box. 2

1 x Cycle	-	94°C (1 min)
35 x Cycles	-	94°C (15 secs), ~50-60°C (30 secs), 72°C (10 secs)
1 x Extension	-	72°C (2 mins)

The PCR products were then analysed via agarose gel electrophoresis, by running on a 1% agarose gel stained with SYBR Safe DNA gel stain (Invitrogen, Cat.

S33102) to visualize cDNA bands. Bioline Hyperladders (100bp and 1kb) were used as size reference markers.

2.6.4 Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

Quantitative Real-Time Polymerase Chain Reactions (qRT-PCR) were performed in 96-well plates. 20ng of cDNA made up to 6.8ul with dH_2O was added to each well (in triplicate) with 1.6ul forward primer (5uM), 1.6ul reverse primer (5uM) and 10ul SensiFAST SYBR Hi-ROX mix (Bioline, Cat: BIO92020). The plate was then loaded into an Agilent AriaMax real time PCR machine (Agilent Technologies) and the qRT-PCR was performed using the conditions outlined in Box 3.

Box. 3

1 x Cycle	-	95.0°C (2 mins)
40 x Cycles	-	95.0°C (5 secs), 60.0°C (10 secs), 72.0°C (10 secs)

The average cycle threshold (Ct) values from each technical triplicate were calculated and used to determine expression levels of the gene of interest relative to the housekeeping gene, GAPDH using the formulae below:

$$\Delta Ct = Ct_{\text{(gene of interest)}} - Ct_{\text{(housekeeping gene)}}$$

$$\Delta\Delta Ct = \Delta Ct_{\text{(sample group)}} - \Delta Ct_{\text{(control group)}}$$

$$\text{Relative Expression} = 2^{-\Delta\Delta Ct}$$

After calculating gene expression relative to the housekeeping gene, GAPDH, unless stated otherwise, values for three biological replicates were averaged and

unpaired, two-tailed, parametric t-tests were performed to calculate the significance of any differences between control and treated samples. *P* values of <0.05 were considered as statistically significant.

2.7 Protein Work

2.7.1 Preparation of Whole Cell Lysates

Total protein was extracted from cells by first harvesting ~2 x 10^6 cells by centrifuging at 1200rpm for 5 minutes. Cell pellets were washed with 1ml PBS and centrifuged again as before. As much PBS as possible was removed from the pellets before re-suspending in ~80ul lysis buffer (20mM Tris, pH 7.5, 150mM NaCl, 10mM EDTA, 0.5% deoxycholic acid, 0.5% Triton X-100). 1 protease inhibitor tablet (Roche, Cat. 04693159001) was added to 10ml lysis buffer immediately before use. Samples were then homogenized using a 1ml syringe and needle (21G x 1.5" – Nr.2., 0.8mm x 40mm) and incubated on ice for 30 minutes before centrifuging at 14,000rpm for 20 minutes at 4^{o}C. The resulting supernatant was pipetted into a fresh, sterile eppendorf and stored at -20^{o}C.

2.7.2 Preparation of Nuclear and Cytosolic Protein Fractions

Approximately 1.5 x 10^7 cells were harvested by centrifuging at 1200rpm for 5 minutes at room temperature and washed in PBS. As much supernatant was removed from the pellet as possible, which was then re-suspended in 150ul cold Buffer RLN and incubated on ice for 10 minutes. One protease inhibitor tablet (Roche, Cat. 04693159001) was added to 10ml Buffer RLN immediately before use. The mix was centrifuged at 3700 rpm for 5 minutes at 4^{o}C and 100ul supernatant

was transferred to a new eppendorf (cytosolic fraction). The remainder of the supernatant was discarded and the pellet was washed twice with 100ul Buffer RLN, centrifuging between washes as before. The resulting pellet was re-suspended in ~80ul lysis buffer. One protease inhibitor tablet (Roche, Cat. 04693159001) was added to 10ml lysis buffer immediately before use. Samples were then homogenized using a 1ml syringe and needle (21G x 1.5" – Nr.2., 0.8mm x 40mm) to lyse nuclei, and incubated on ice for 30 minutes before centrifuging at 14,000rpm for 20 minutes at 4^{o}C. The resulting supernatant was pipetted into a fresh, sterile eppendorf (nuclear fraction) and stored at -20^{o}C.

2.7.3 Histone Protein Extraction

10^7 cells were harvested for histone extraction. Cells were first lysed in 500ul histone extraction buffer (PBS containing 0.5% Triton X-100, 2mM phenylmethylsulfonyl fluoride, 0.02% sodium azide) for 1 minute on ice, followed by centrifuging at 8000rpm for 10 minutes at 4^{o}C. The resulting pellet was washed in 250ul histone extraction buffer and centrifuged as before. The pellet was then resuspended in 0.4M HCl and incubated overnight at 4^{o}C for acid extraction of histones. The following morning, samples were centrifuged as before and the supernatant (containing histone protein) was saved. The HCl was neutralized by adding 0.1X volumes of 2M NaOH. The histone protein was then used in SDS-PAGE (2.7.6). All histone extractions were performed in collaboration with Dr John Halsall (Institute of Cancer and Genomic Sciences, University of Birmingham).

2.7.4 Bradford Assay Protein Quantification

To ensure equal loading of samples in SDS-PAGE, protein concentrations were determined via Bradford assay. Standard samples (5ul of known BSA protein concentration in serial dilution ranging from 0-2mg/ml) and 5ul of experimental protein samples were loaded in duplicate into a 96-well plate along with 250ul Bradford reagent (Sigma, Cat: B6916). Absorbance was measured at 570nm using a Tecan infinite 5200 pro plate reader and iControl™ Microplate Reader Software. A standard curve was plotted from the standard samples of known protein concentration (conc. vs absorbance). The protein concentration of the experimental samples was determined by subtracting the absorbance from a blank control absorbance value, and then dividing by the slope of the standard curve.

2.7.5 Preparation of SDS-PAGE Gels

For SDS-PAGE, 10% polyacrylamide SDS resolving gels were prepared with a 4% stack (Table 2.1) and stored in moist wrapping at 4°C for no more than 1 week before use.

2.7.6 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Before loading onto gels, extracted protein samples were mixed with 5X SDS loading dye (200mM Tris-HCL pH6.8, 40% glycerol, 4% SDS, 0.4% bromophenol blue, 200mM β-mercaptoethanol) and heated at 70°C for 10 minutes. In addition to protein samples, 3ul of protein ladder (PageRuler™ Plus Prestained Protein Ladder, ThermoFisher Scientific, Cat: 26619) was loaded onto gels as a molecular weight marker for use as a size standard reference. Gels were then run for approximately

90-120 minutes at 120V in cold 1X Running Buffer (100ml 10X running buffer, 5ml 20% SDS, 895ml dH_2O; See Table 2.1 for further detail).

2.7.7 Western Blotting

After running, a semi-dry transfer was performed onto nitrocellulose membranes using the Trans-Blot Turbo Transfer System (BIO-RAD, Cat. 1704271) and 1X Transfer Buffer (200ml BIO-RAD TransBlot Turbo 5 x Transfer Buffer (Cat. 1704271), 200ml 100% Ethanol, 600ml nanopure H_2O). Membranes were then blocked in 5% skimmed milk in TBS for 1 hour before incubating in primary antibody overnight at 4^oC, with agitation. All primary antibodies were prepared in 5% BSA at the dilutions indicated in Table 2.2. The following morning, the membranes were washed three times in TBS-T for 5 minutes and incubated in the appropriate secondary antibody, diluted in blocking solution (5% skimmed milk in TBS-T) in the dark for 1 hour, at room temperature, with agitation. The antibody was removed and the membranes were washed twice in TBS-T for 5 minutes, and once in TBS for 5 minutes for a final time, before storing in TBS at 4^oC.

2.7.8 SDS PAGE Image Analysis

Membranes were scanned and protein bands were detected using the Odyssey infrared detection system (LI-COR Biosciences). Images were then analysed and the protein bands were quantified using Image Studio Lite software.

2.7.9 Stripping Nitrocellulose Membranes

Occasionally, antibodies were stripped from nitrocellulose membranes by incubating in 0.5M NaOH for ~10-20 minutes, followed by washing three times in TBS for 5 minutes. Blots were then re-scanned to ensure complete removal of antibodies and re-incubated in a different primary antibody.

2.8 DNMT1 Immunoprecipitation (IP)

2 x 10^7 cells were harvested per IP. Whole cell lysates were prepared using 500ul lysis buffer, as described in 2.7.1, and protein was quantified via Bradford assay (2.7.3). 50ul (1.5mg) magnetic beads (Dynabeads Protein G Immunoprecipitation kit, Invitrogen, Cat: 10007D) per IP, were pipetted into an eppendorf, and the supernatant removed using a magnetic separator. Beads were then washed twice with 100ul lysis buffer (containing protease inhibitor). 5ul (5ug) DNMT1 antibody was diluted in 200ul Antibody binding and washing buffer (provided with kit), and added to the beads. These were mixed by rotating for 30 minutes at room temperature. The magnetic separator was used again to aid removal of the supernatant, and the bead-antibody complex were washed with 200ul Antibody binding and washing buffer. 2mg of the initially extracted protein was loaded onto the bead-antibody complex and mixed by rotating for 2 hours at room temperature. A 2% fraction of protein was left behind to use later as an “input” control. Following incubation, the eppendorf was placed on the magnetic separator and the supernatant was kept and stored as the “unbound fraction” of protein. This fraction represented protein not bound to the DNMT1 antibody, and should therefore contain very little DNMT1). The beads were then washed gently, three times with 200ul

washing buffer (provided by the kit). The beads-antibody-protein was resuspended in 100ul washing buffer and transferred to a new eppendorf, which was placed on the magnet for removal of the supernatant. Finally, the beads-antibody-protein were resuspended in 10ul elution buffer and 10ul loading buffer, gently vortexed and stored at -80^{o}C. Prior to loading the immunoprecipitated protein onto a polyacrylamide gel for SDS-PAGE, the mix was heated to 70^{o}C for 10 minutes to fully elute the protein from the beads. The eppendorf was placed on the magnet and the supernatant containing the pulled-down protein was loaded directly. SDS-PAGE and immunoblotting was performed according to 2.7.5 and 2.7.6.

2.9 DNMT1 RNA Immunoprecipitation (RIP)

2 x 10^{7} cells were harvested per RIP. Cell pellets were washed in PBS three times and resuspended in 500ul ice-cold Buffer RLN (containing 100U/ml RNase inhibitor and with 1 protease inhibitor tablet added to 10ml RLN). This was incubated on ice for 10 minutes before centrifuging at 3700rpm, for 5 minutes at 4^{o}C. Nuclear pellets were washed with 100ul Buffer RLN, followed by resuspending in 500ul freshly prepared RIP Buffer (25mM Tris, pH 7.4, 5mM EDTA, 150mM KCl, 0.5mM DTT, 0.5% NP40 Igepal, 100U/ml RNase inhibitor) with added protease inhibitor. This mix was incubated on ice for 3 hours with frequent, gentle agitation. Following the incubation, nuclei were homogenized using a 1ml syringe and needle (21G x 1.5" – Nr.2., 0.8mm x 40mm), then centrifuged at maximum speed for 10 minutes at 4^{o}C to pellet nuclear debris. The supernatant was kept and transferred to an RNase-free eppendorf, to which 6ug of either DNMT1 of IgG (control) antibody was added. This mix was incubated overnight at 4^{o}C with gentle rotation.

The following day, 50ul (1.5mg) magnetic beads (Dynabeads Protein G Immunoprecipitation kit, Invitrogen, Cat: 10007D) were prepared per RIP. Similar to 2.8, the beads were placed into an eppendorf on a magnetic separator and the supernatant was removed. This was followed by two washes with 100ul RIP buffer. The overnight antibody-protein-RNA suspension was added to the beads and incubated for 1 hour with gentle rotation at 4^{o}C. Following this incubation, the beads-antibody-protein-RNA were placed on the magnet, the supernatant was removed and the complex was washed three times with 500ul ice-cold RIP buffer and once with RNase-free PBS. At this point, 5% of the bead slurry was collected to be used in SDS-PAGE analysis to confirm DNMT1 pull-down. The remaining beads-antibody-protein-RNA was resuspended in 100ul RIP buffer with 50ug proteinase K and 0.1% SDS. This was incubated at 55^{o}C for 45 minutes to detach the protein-RNA complexes from the beads. The eppendorf was then placed onto the magnet and the supernatant was transferred to a new eppendorf. The beads were discarded.

To purify the RNA, 1X volume of phenol-chloroform-isoamyl alcohol was added to the supernatant and vortexed thoroughly. This was phase separated by centrifuging at 14,000rpm for 10 minutes at 4^{o}C. The aqueous phase (containing the RNA) was carefully collected and placed into a fresh eppendorf. Any remaining aqueous phase was further extracted by the addition of 150ul back extraction buffer (10mM Tris, pH8, 1mM EDTA, 100mM NaCl, 0.25% SDS), followed by vortexing and centrifugation as before. Any remaining aqueous phase was collected and added to the previous collection.

The RNA was further purified by ethanol precipitation. 0.1X volumes of 3M sodium acetate (pH 5.2), 2.2X volumes of 100% ice cold ethanol and 1ul glycogen was added to the RNA extract and incubated at -20^{o}C overnight. The following morning, the mix was centrifuged at 10,000rpm for 20 minutes at 4^{o}C and the supernatant carefully discarded. 500ul 70% ice cold ethanol was added and the mix centrifuged as before. The ethanol was carefully removed and the tubes were left open to allow any residual ethanol to evaporate. The purified RNA was then finally dissolved in 20ul RNase-free H_2O. This RNA was then DNase-treated and converted into cDNA as described in 2.6.1.3 and 2.6.2. qRT-PCRs were then performed. Results were analysed using the calculations below to determine DNMT1 enrichment for *CCDC26* compared to the IgG control.

$$\text{DNMT1 RIP } \Delta Ct = Ct_{\text{DNMT1\_RIP\_CCDC26}} - Ct_{\text{INPUT\_RNA\_CCDC26}}$$

$$\text{IgG RIP } \Delta Ct = Ct_{\text{IgG\_RIP\_CCDC26}} - Ct_{\text{INPUT\_RNA\_CCDC26}}$$

$$\Delta\Delta Ct = \text{DNMT1 RIP } \Delta Ct - \text{IgG RIP } \Delta Ct$$

$$\text{Relative Expression} = 2^{-\Delta\Delta Ct}$$

CHAPTER 3

CHARACTERISATION OF LONG NON-CODING RNA, *CCDC26*, IN CHRONIC MYELOID LEUKEMIA CELL LINE, K562

3.1 Introduction

Currently, our knowledge of lncRNAs comprises awareness of tens of thousands of transcripts, but with functional understanding of less than 10% of them (Iyer et al. 2015; Quek et al. 2015). In order to fully comprehend and appreciate the complexity of gene regulatory networks and other cellular processes, mechanistic analysis of independent lncRNAs is required. LncRNA, *CCDC26*, is a particularly interesting candidate which has undergone limited functional analysis (Yin et al. 2006). The gene encoding *CCDC26* is located at the chr.8q24 locus, a region renowned for being a mutational hotspot in numerous types of cancer (Ghoussaini et al. 2008; Amundadottir et al. 2006; Kuhn et al. 2012; Beroukhim et al. 2010). Mutations and SNPs associated with this region frequently encompass *CCDC26* (Radtke et al. 2009; Kuhn et al. 2012; Duployez et al. 2018), and several reports have shown correlations between increased *CCDC26* levels and cancer progression (Hirano et al. 2008; Wang, Hui, et al. 2018; Peng and Jiang 2016). However, since its discovery in 2006 (Yin et al. 2006), functional studies have done little beyond linking *CCDC26* depletion with reduced cell viability and identifying a few potential downstream targets (Hirano et al. 2015; Peng and Jiang 2016; Cao et al. 2018; Yan et al. 2019), with none of these publications having investigated the effects of a complete *CCDC26* CRISPR knock out (KO). As part of our first objective, here we

have begun initial characterisation of C*CDC26*, including determination of its localisation, expression and effects on epigenetic modifications in the CML K562 cell line.

3.2 Results

3.2.1 Expression and Localisation of *CCDC26*

Previous literature indicates that *CCDC26* expression is limited to myeloid cell lines of haematopoietic origin, and that the AML, HL60 cell line shows strongest expression (Hirano et al. 2015; Yin et al. 2006). To confirm this and establish a suitable cell line in which to proceed with this project, qRT-PCRs were performed by a former colleague in our group on a panel of cell lines, including those of both myeloid and lymphoid lineage amongst others (Figure 3.1). The results concurred with previous studies and showed significantly greater expression in myeloid cell lines; however, the CML K562 cell line demonstrated greatest *CCDC26* expression levels as opposed to HL60 cells (Wijesinghe et al. unpublished, manuscript in preparation). To further confirm *CCDC26* expression, I analysed publicly available CalTech RNA-seq data sets on multiple cell lines (ENCODE Project, GEO accession: GSM958729) (Djebali et al. 2012). Expression profiles showed detection of *CCDC26* in K562 cells, as well as three lymphoblastoid cell lines (GEO accession: GM12878, GM12891 and GM12892) although at significantly lower levels compared to the former. Expression was not detected however in the remaining 11 non-myeloid cell lines for which RNA-seq data was available (Figure

3.2) (Djebali et al. 2012). Following these results, the K562 cell line was chosen for subsequent analyses of *CCDC26* function.

According to NCBI RefSeq annotations, *CCDC26* has four isoforms (NR_130920.1, NR_130919.1, NR_130918.1 and NR_130917.1) that are transcribed from two discrete transcription start sites (TSSs), and are alternatively spliced to contain variations of six exons (Pruitt et al. 2014). Isoforms 1, 2 and 3 are transcribed from the same transcription start site (TSS2: Chr.8: 130,587,264) while isoform 4 is expressed from a distinct, upstream start site (TSS1: Chr.8: 130,692,485). Genome-wide histone modifications from K562 cells support the position of two TSSs (GEO accession: GSM733643) (Figure 3.3). Tracks for H3K4me3, H3K4me2 and H3K9ac show two distinct peaks at genomic regions corresponding to exons 1 and 2; these are epigenetic signatures for TSSs, transcription factor binding regions and active gene promoters respectively. Peaks corresponding to active enhancers (H3K27ac and H3K4me1) can also be seen within these regions, as well as RNA Pol II Chromatin immunoprecipitation (ChIP) peaks (Consortium 2012). However, these peaks are more predominant at the exon 2 region, potentially indicating increased activity at TSS2 over TSS1 (Figure 3.3).

Isoforms 1 (1666bp) and 2 (1649bp) are transcribed from TSS2, and differ only by an additional 17 nucleotide sequence at the 3' end of exon 4 in the former. Isoform 3 (1495bp) is also transcribed from TSS2, but lacks exon 4 completely. Isoform 4 (1718bp) is transcribed from TSS1 that is upstream of TSS2. This isoform lacks exon 2 and exon 4 (Figure 1.6). H3K36me3 peaks, a mark commonly observed

across transcribed regions of the gene body (Huang and Zhu 2018), can interestingly be observed between regions encoding exon 2 and 6, but less so across the region between exons 1 and 2. This potentially indicates that isoform 4 is not transcribed in K562 cells.

To functionally characterise *CCDC26* in CML and AML, it is important to understand which isoforms are expressed in K562 cells. This will also help in appropriately targeting *CCDC26* and manipulating its RNA levels in subsequent experiments. Primers were designed to amplify specific *CCDC26* isoforms and qRT-PCRs were performed to measure their expression levels. Given that isoforms 1 and 2 differ only by a 17bp difference, it was not possible to design primers that distinguish these two forms, and both were amplified by the same primer set (CCDC26_Set1-2). Isoforms 3 and 4 were amplified by primer sets CCDC26_Set3 and CCDC26_Set4 respectively. Primers were also designed to amplify all isoforms in K562 cells (CCDC26_all), by annealing to a common sequence within exon 6 (see Table 2.5 for primer details).

The qRT-PCR results show that the four isoforms are not uniformly expressed; all the three isoforms starting from transcription start site, TSS2, are expressed at different levels (Figure 3.4). Isoforms 1 and 2 are predominant, accounting for more than 80% of the total *CCDC26* transcripts in the cell. Isoform 3 is also detected but at considerably lower levels, accounting for approximately 15% of transcripts. Isoform 4 on the other hand, which is expressed from another transcription start

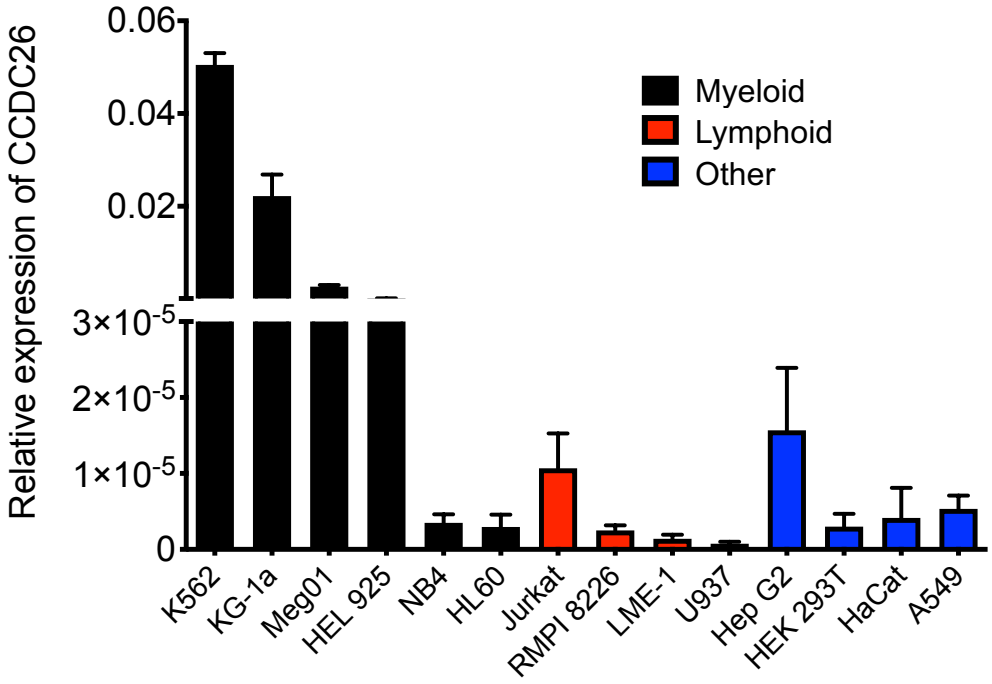


Figure 3.1 K562 cells demonstrate highest levels of *CCDC26* expression. *CCDC26* levels were measured in multiple cancer and non-cancer cell lines using qRT-PCR. Generally, myeloid cell lines showed highest expression levels, with the CML K562 cell line demonstrating greatest *CCDC26* expression of all. All qRT-PCRs and analyses were performed by former Kanhere laboratory member, Susanne Wijesinghe (Wijesinghe et al. unpublished, manuscript in preparation).

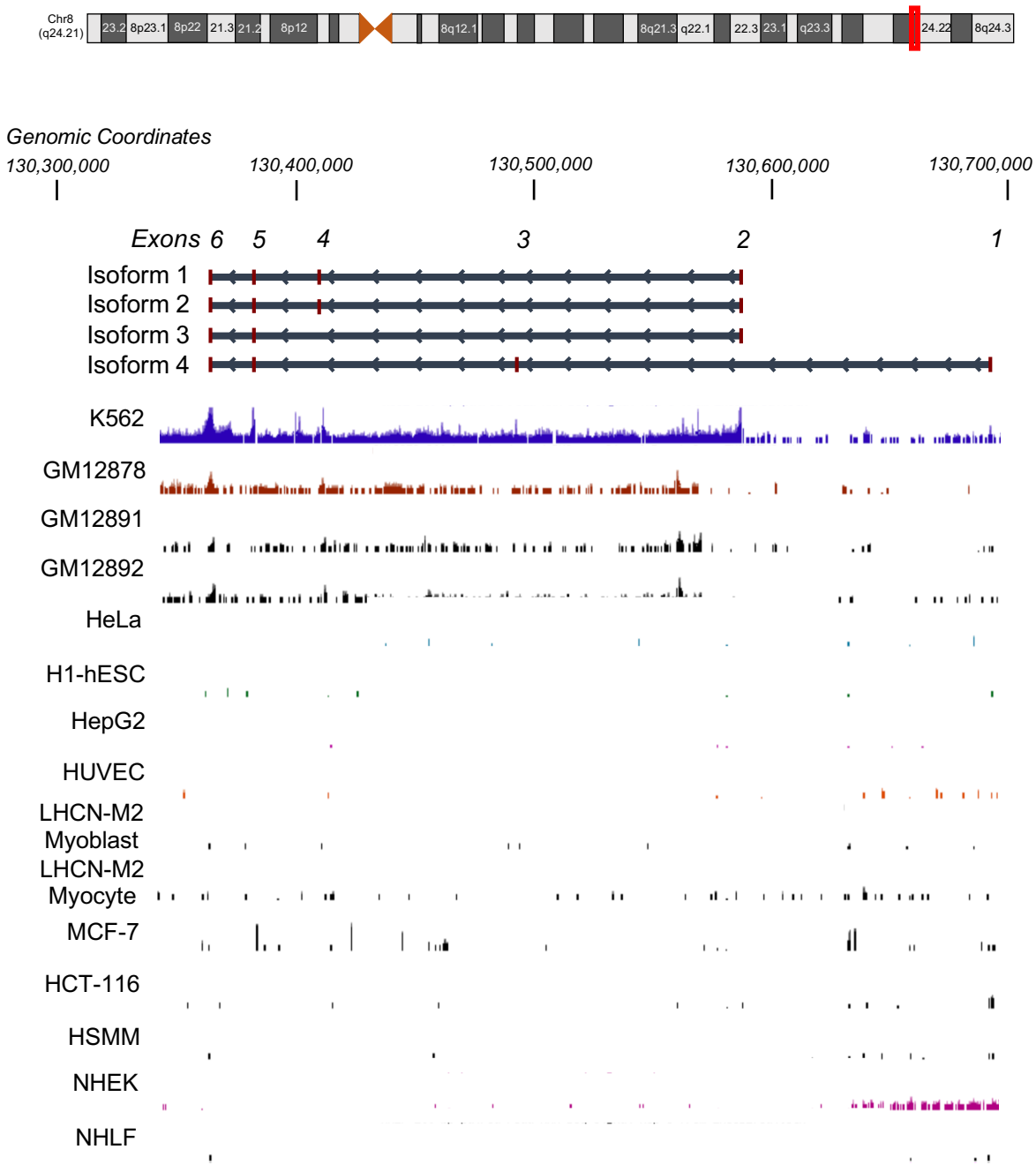


Figure 3.2 Expression profiles for *CCDC26*. Expression profile tracks produced from RNA-seq data as part of the ENCODE Project from CalTech (Djebali et al. 2012), show detection of *CCDC26* expression in the myeloid K562 cell line. *CCDC26* is also expressed at lower levels in GM12878, GM12891 and GM12892 lymphoblastoid cell lines, but does not appear to be expressed in the remaining non-myeloid cells. Data tracks were accessed and viewed on UCSC Genome Browser. Data points have been transformed by LOG (ln(1+x)). Tracks are within a vertical viewing range of 0-5 and have been smoothed by 5 pixels.

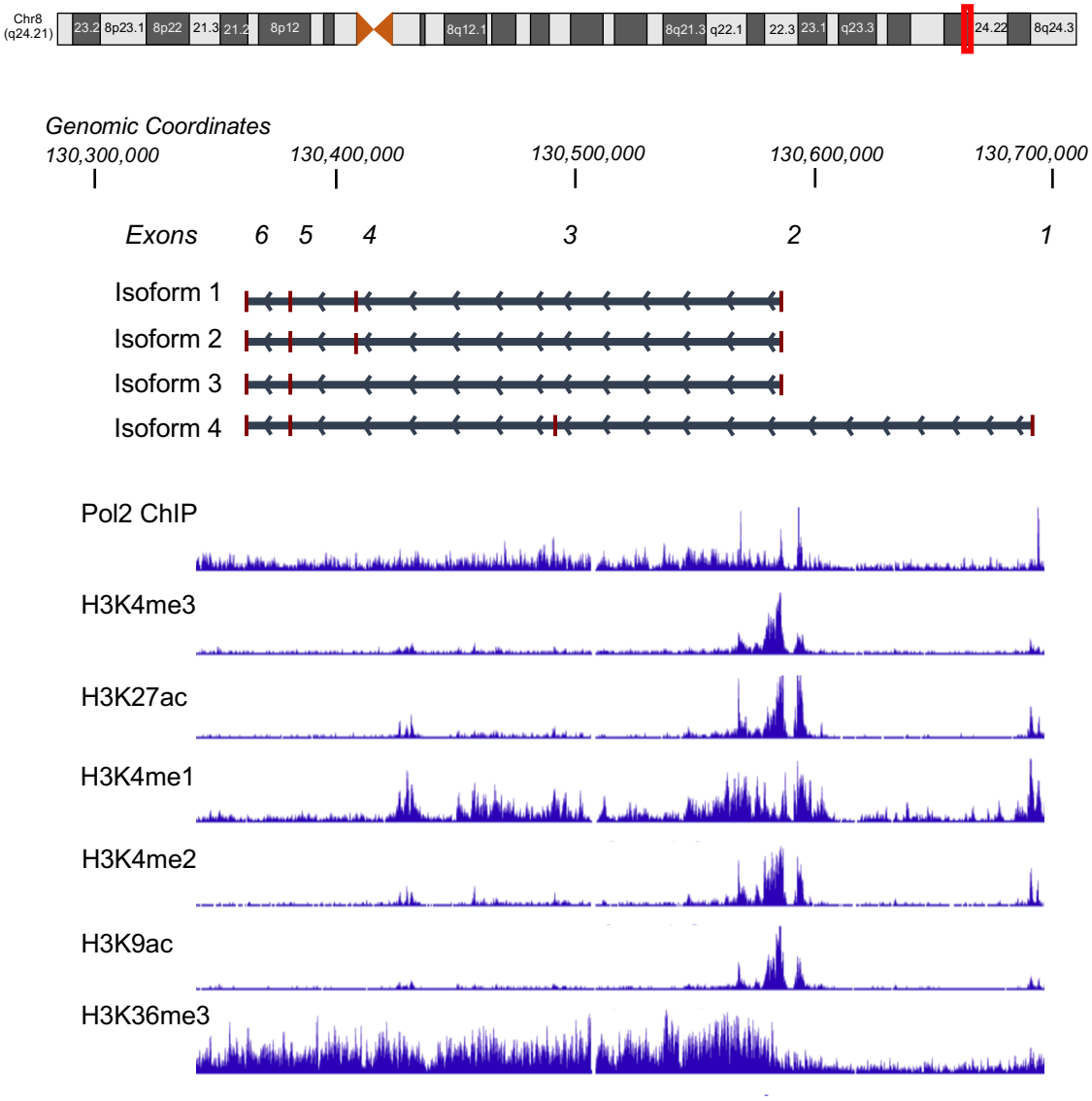
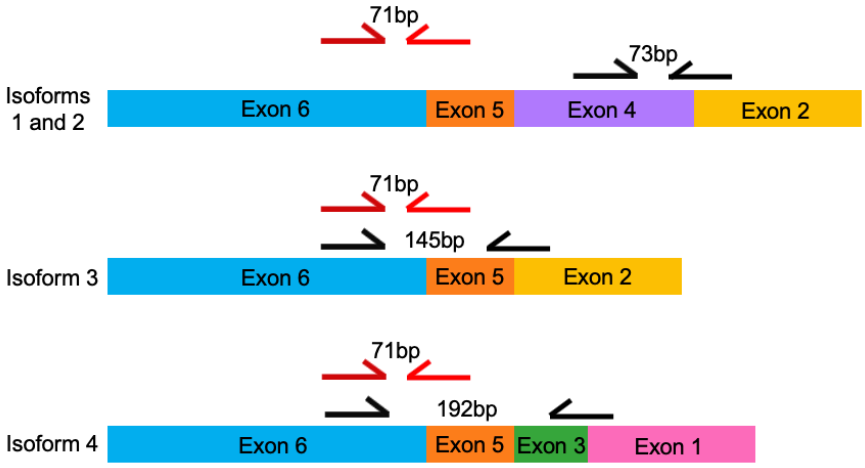


Figure 3.3 Mapping epigenetic marks and RNA Pol II binding to the CCDC26 locus. Histone modification tracks generated from ChIPseq data sets generated by the Broad/MGH ENCODE Project indicate the position of two TSSs in K562 cells. Two defined peaks can be observed for RNA Pol II binding as well as epigenetic signatures that correspond to TSSs, TF binding spots, active promoter and active enhancer regions (H3K4me3, H3K4me2, H3K9ac, H3K27ac and H3K4me1 respectively) at the position of exon 2 and to a lesser extent, exon 1. The signature for actively transcribed gene bodies (H3K36me3) also shows peaks between regions encoding exons 2-6, possibly indicating that isoform 4 is transcribed to a lesser extent than isoforms 1-3 in K562 cells (Consortium 2012).

A.



B.

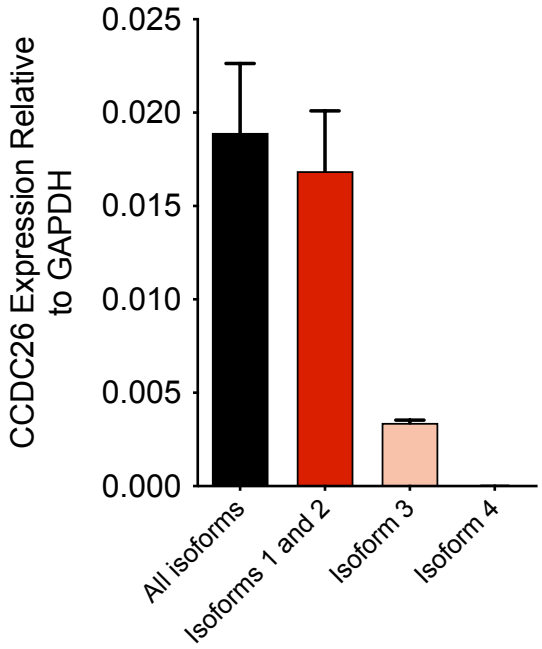


Figure 3.4 *CCDC26* isoforms 1 and 2 are predominant in K562 cells. A. Schematic portraying spliced forms of *CCDC26*. Positions of exon-spanning primers used to amplify specific isoforms (black) and size (bp) of amplification products are indicated. Position of primers used to amplify all isoforms simultaneously are also indicated (red). **B.** qRT-PCRs were performed using isoform-specific primers to determine which transcripts are present in K562 cells. Isoforms 1 and 2 account for ~80% of all *CCDC26* transcripts whilst isoform 3 accounts for ~15%. Isoform 4 accounts for <1% of total transcripts.

site, TSS1, is barely detectable and represents less than 1% of total *CCDC26*, as predicted by epigenetic signatures in Figure 3.3 (Figure 3.4).

3.2.2 Subcellular Localisation of *CCDC26*

A large proportion of lncRNAs have shown a predominantly nuclear localisation, either associated with chromatin or enriched in nuclear sub-compartments and organelles (Clemson et al. 2009; Hutchinson et al. 2007; Werner and Ruthenburg 2015). The subcellular localisation of any lncRNA can give an inkling to its functional mechanism. We began initially investigating the localisation of *CCDC26* by implementing a data-mining approach, and analysing RNA-seq data sets from various cellular compartments generated by the ENCODE Consortium (GEO sample accession: GSM758577). Both polyadenylated (polyA+) and non-polyadenylated (polyA-) RNAs were analysed in subcellular fractions as well in the whole cell in Figure 3.5. Polyadenylation is the process whereby a chain of adenine nucleotides is added to the 3' end of an RNA molecule. This makes the RNA more stable by protecting it from enzymatic degradation (Nevins and Darnell 1978). The data suggests that *CCDC26* is primarily nuclear, with enrichment for a non-polyadenylated form. Within the nucleus, the majority of *CCDC26* appears bound to chromatin, with a small amount also present in the nucleolus. Interestingly, there also appears to be a small amount of polyadenylated *CCDC26* present in the cytosol (Figure 3.5).

To further assess localisation, qRT-PCRs were performed on nuclear and cytosolic RNA fractions. snoRNA *U105* and *GAPDH* levels were used as a measure of the

purity of the nuclear and cytosolic fractions respectively. Actin was used as a housekeeping control gene against which *CCDC26* expression could be measured, given its similar levels in both the nucleus and cytosol. Consistent with the RNA-seq data, all *CCDC26* isoforms showed greater levels in the nucleus, although the cytosol was not devoid of *CCDC26*, and transcripts were still detected albeit at low levels in this compartment (Figure 3.6).

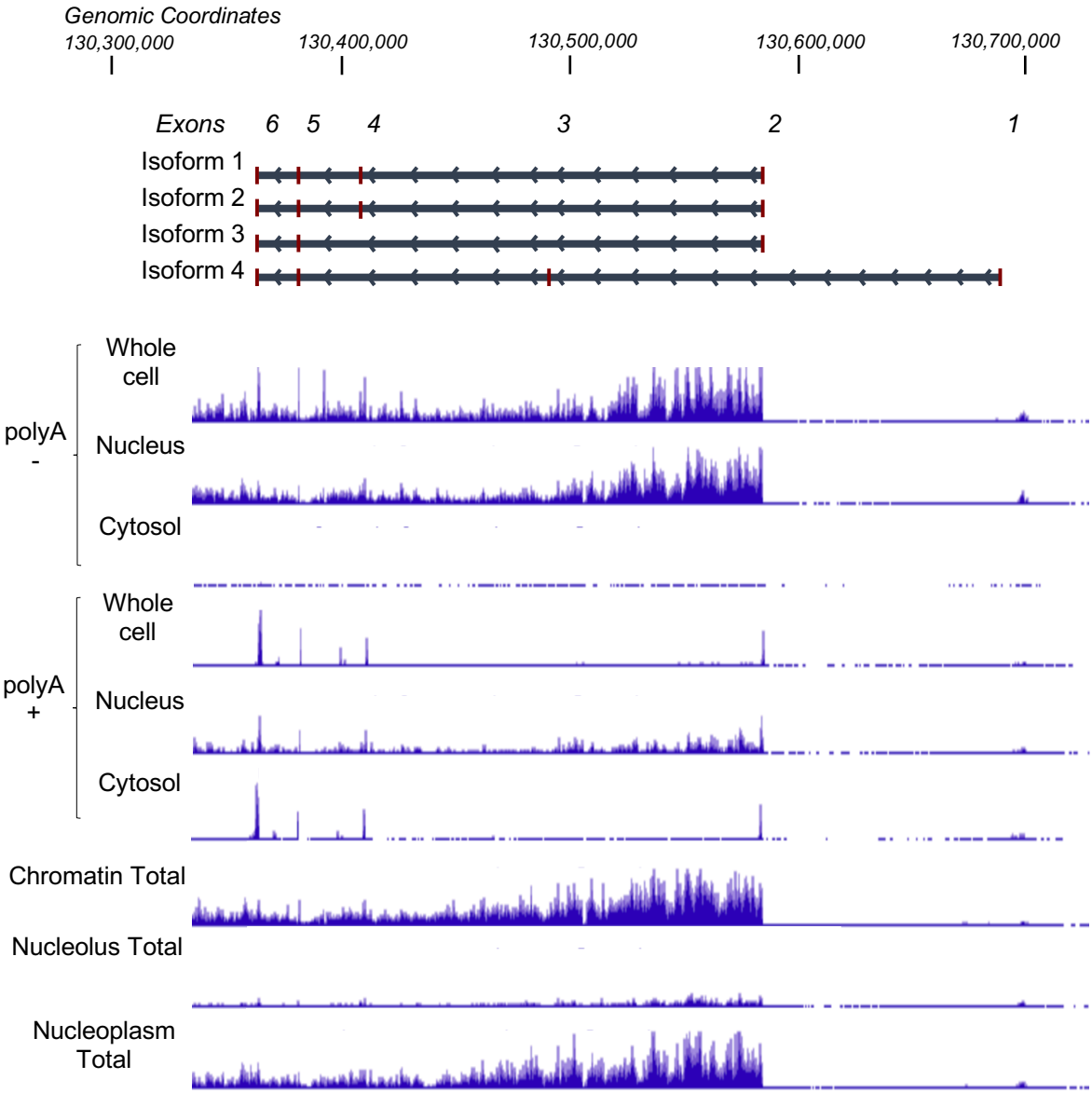


Figure 3.5 *CCDC26* localisation: a data mining approach. Long RNA-seq tracks generated by the ENCODE Consortium and viewed on UCSC Genome Browser show that greater levels of *CCDC26* are found in the nucleus of K562 cells. However, a small amount of polyA+ *CCDC26* also appears in the cytosol. Low levels of *CCDC26* are found in the nucleolus of the cell, but the majority of the lncRNA is predominantly found in the chromatin-bound fraction.

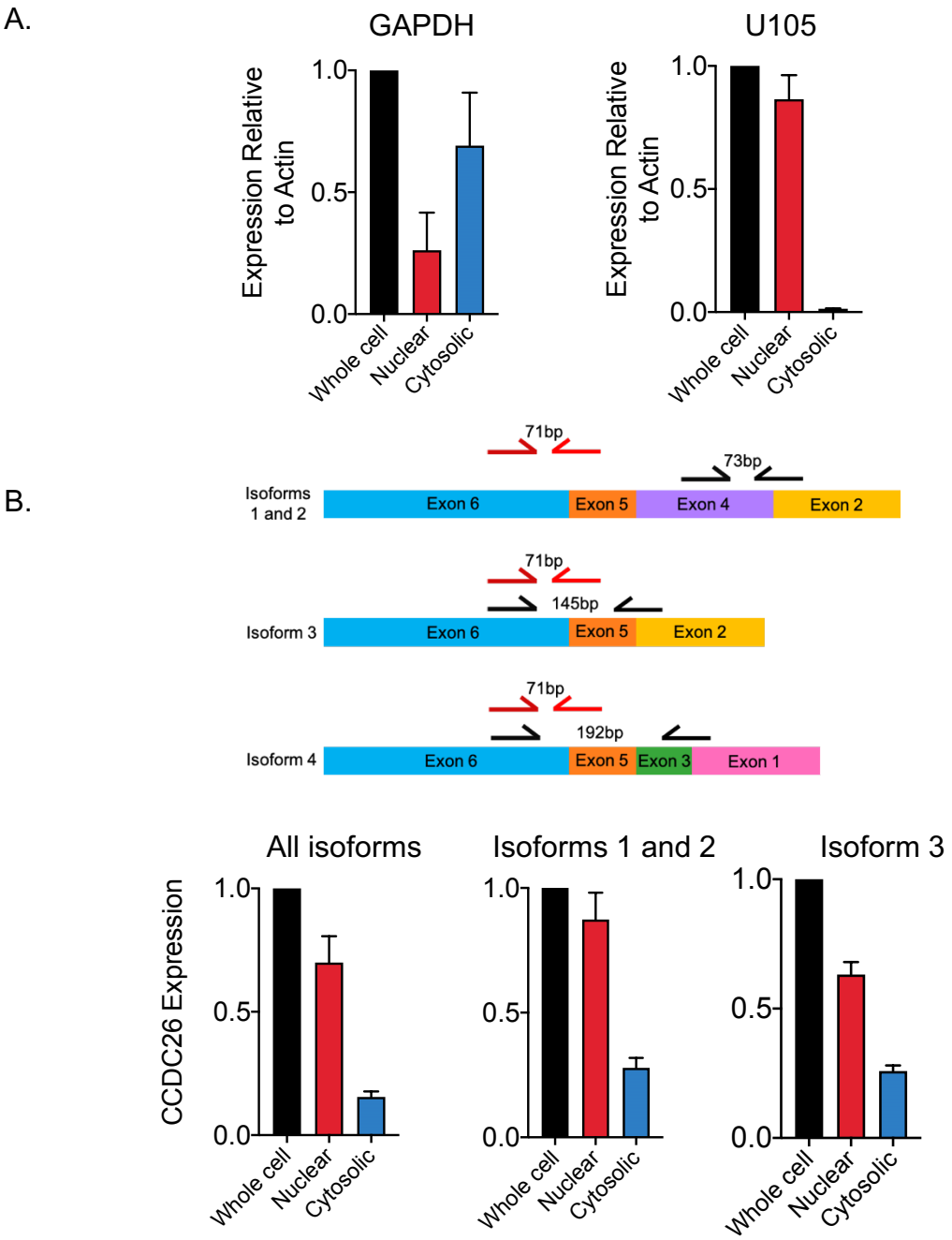


Figure 3.6 *CCDC26* is primarily localised in the nucleus of K562 cells. A. Nuclear and cytosolic RNA fractions were extracted and converted to cDNA. To assess the purity of fractions, qRT-PCRs were performed using *GAPDH* and *snoRNA U105* as cytosolic and nuclear controls respectively. **B.** qRT-PCRs were performed using *CCDC26* isoform-specific primers, measuring relative to Actin and total RNA levels. Greater expression of all *CCDC26* isoforms was found in the nuclear fraction, however, all isoforms were also present, to a lower degree, in the cytosolic fraction. Isoform 4 was not further analysed given its extremely low overall expression levels (Figure 3.4). Schematic portraying spliced forms of *CCDC26* is also displayed. Positions of exon-spanning primers used to amplify specific isoforms (black) and size (bp) of amplification products are indicated. Position of primers used to amplify all isoforms simultaneously are also indicated (red).

3.2.3 *CCDC26* KO cells grow slower and are more apoptotic

To understand the biological role of *CCDC26*, we began comparing K562 WT cells to *CCDC26* KO cells. Two *CCDC26* KO cell lines (KO.1 and KO.2) were generated by former colleague, Susanne Wijesinghe, using CRISPR Cas9 technology, and have been subsequently used throughout this project. Given the increased difficulty of achieving an RNA KO compared to protein (Yang et al. 2018), two small guide RNAs (sgRNAs) were used simultaneously (Figure 3.7A). These were designed to target TSS2, as previous qRT-PCRs indicated that almost all *CCDC26* transcripts arise from this site in K562 cells (Figure 3.4). Single cell clones were selected to obtain clean KO cell lines, which were subsequently sequenced. Sequencing (performed previously by Susanne Wijesinghe) showed that the 3 CCDC26 alleles in KO.1 cells (K562 cells possess 3 copies of chromosome 8), possess a single base insertion, a single base deletion and a 28-base deletion respectively. KO.2 cells possess a 34 base deletion and single base insertion in CCDC26 allele 1, and a 28-base deletion in CCDC26 allele 2. The third allele however, appears to be intact (Figure 3.7C). Before conducting any further experiments, I first checked the quality of the KOs in these cell lines by qRT-PCRs. Accordingly, KO.1 cells show a ~99.8% *CCDC26* depletion, whilst KO.2 cells demonstrate an almost 90% reduction (Figure 3.7B).

KO cell lines were grown in the same conditions as WT K562 cells (Chapter 2.2.1) and cell viability was measured by manually counting live cells using a hemocytometer via the trypan blue exclusion method. Cells were counted every 24hrs across a 72hr period and growth curves were subsequently plotted, showing

a significantly reduced rate of growth in both KO cell lines. Whilst WT cells double in number approximately every 24hrs as previously reported (Murray et al. 1993), the KO cells grow almost half as fast, increasing by two-fold approximately every 48hrs (Figure 3.8).

We next went on to investigate whether cell death or cell cycle stalling could be contributing to this reduction in growth rate. Apoptosis, also referred to as programmed cell death, is mediated by the caspase family of enzymes. Briefly, caspases are translated as inactive zymogens, which generate active cleavage products when apoptosis is triggered by the necessary signal. The initiator caspases (2, 8, 9 and 10) are activated first and are responsible for activating the downstream effector caspases (3, 6 and 7). Active effector caspases then function by proteolytically cleaving numerous proteins, such as PARP-1 and MEK Kinase 1, which eventually results in cell death. This process can be triggered either intrinsically, whereby the initial death signal arises from within the cell, e.g. DNA damage, or extrinsically, whereby the cell receives an external signal by the binding of a ligand to its extracellular death receptor (Riedl and Shi 2004).

Western blotting using anti-caspase antibodies showed increased levels of cleaved initiator caspases 8 and 9 in *CCDC26* KO cells, suggesting that these cells are more apoptotic (Figure 3.9). There is a particularly striking increase in the 35kDa caspase 9 cleavage product, and although levels of the 18kDa caspase 8 cleavage product also rise, the increase is not as statistically significant. It is possible that this is reflective of the apoptotic pathway that has been activated; caspase 9 cleavage is

associated with the intrinsic pathway, whilst caspase 8 is more involved in the extrinsic pathway (Riedl and Shi 2004). Caspase 3 is activated downstream of both intrinsic and extrinsic pathways (Shalini et al. 2015). Overall, the levels of the effector caspase 3 cleavage products however are not significantly increased (Figure 3.9).

Cell cycle was also analysed using propidium iodide (PI) staining to quantitate the DNA content of cells. FACS was then used to determine the percentage of cells in each stage of the cell cycle (Figure 3.10). The most interesting result obtained from this was the significant increase in the proportion of cells in the sub G0 phase in the KO cell lines compared to WT. Cells undergoing apoptosis are characterised by a number of morphological changes including nuclear disintegration and DNA fragmentation and degradation (Kerr, Wyllie, and Currie 1972; Nagata et al. 2003). The sub G1 portion represents such apoptotic cells with reduced DNA content. This compliments the anti-caspase immunoblot result, supporting that the *CCDC26* KO cell lines are more apoptotic. However, there is very little difference in the proportion of cells in each of the other stages of the cell cycle (G1, S, G2/M phase) and in the polynucleated fraction of cells in the KOs compared to the WT. It is possible that KO cells are moving through stages of the cell cycle at a normal rate, but appear to be growing more slowly due to a higher proportion of cell death, and therefore a greater loss of cells over a given period.

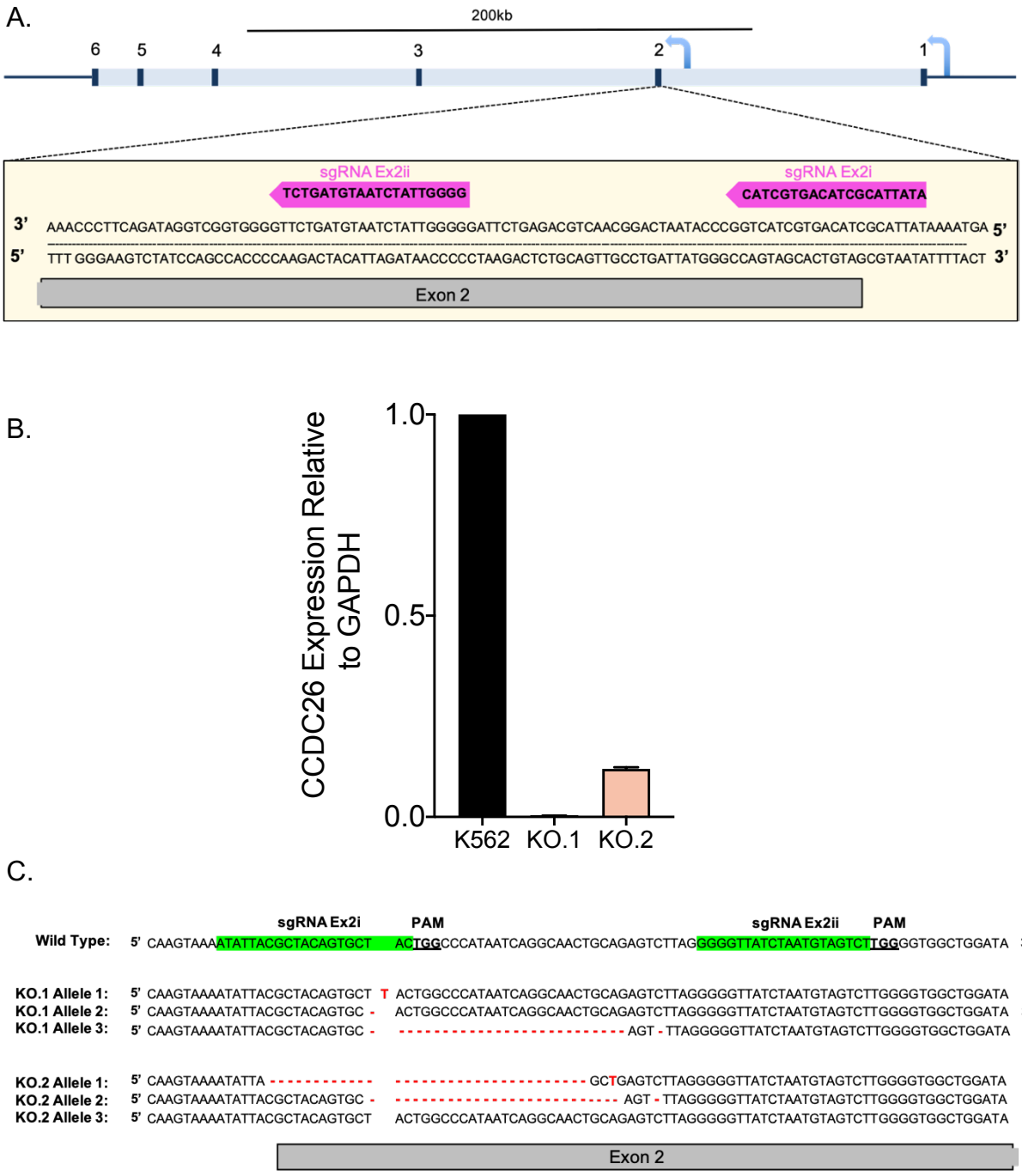


Figure 3.7 *CCDC26* CRISPR KO cell lines. A. Schematic of the CCDC26 gene showing that TSS2 was targeted with two sgRNAs simultaneously using CRISPR Cas9 technology to establish *CCDC26* KOs (initial CRISPR experiment performed by Susanne Wijesinghe). **B.** *CCDC26* KO's in two KO cell lines were confirmed by qRT-PCR with *CCDC26* primers. **C.** Schematic demonstrating the CCDC26 mutations present in each allele in the KO cell lines (sequencing performed by Susanne Wijesinghe). PAM sequences are indicated and green highlighted sequences represent the sgRNA targets.

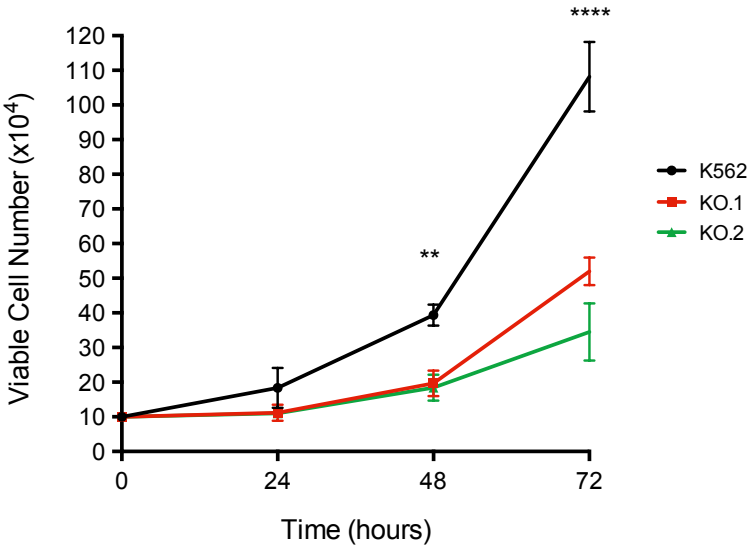


Figure 3.8 *CCDC26* KO cells grow slower than WT cells. WT and *CCDC26* KO cells were grown for 72hrs in 6 well-plates. Every 24hrs, cells were stained with Trypan blue, and viable cells were counted manually using a hemocytometer. KO cells grew approximately half as fast as WT cells. Values represent the mean ± standard deviation (n=3). * P<0.05; **P<0.005; ***P<0.001; ****P<0.0001 (unpaired, two-tailed *t* test).

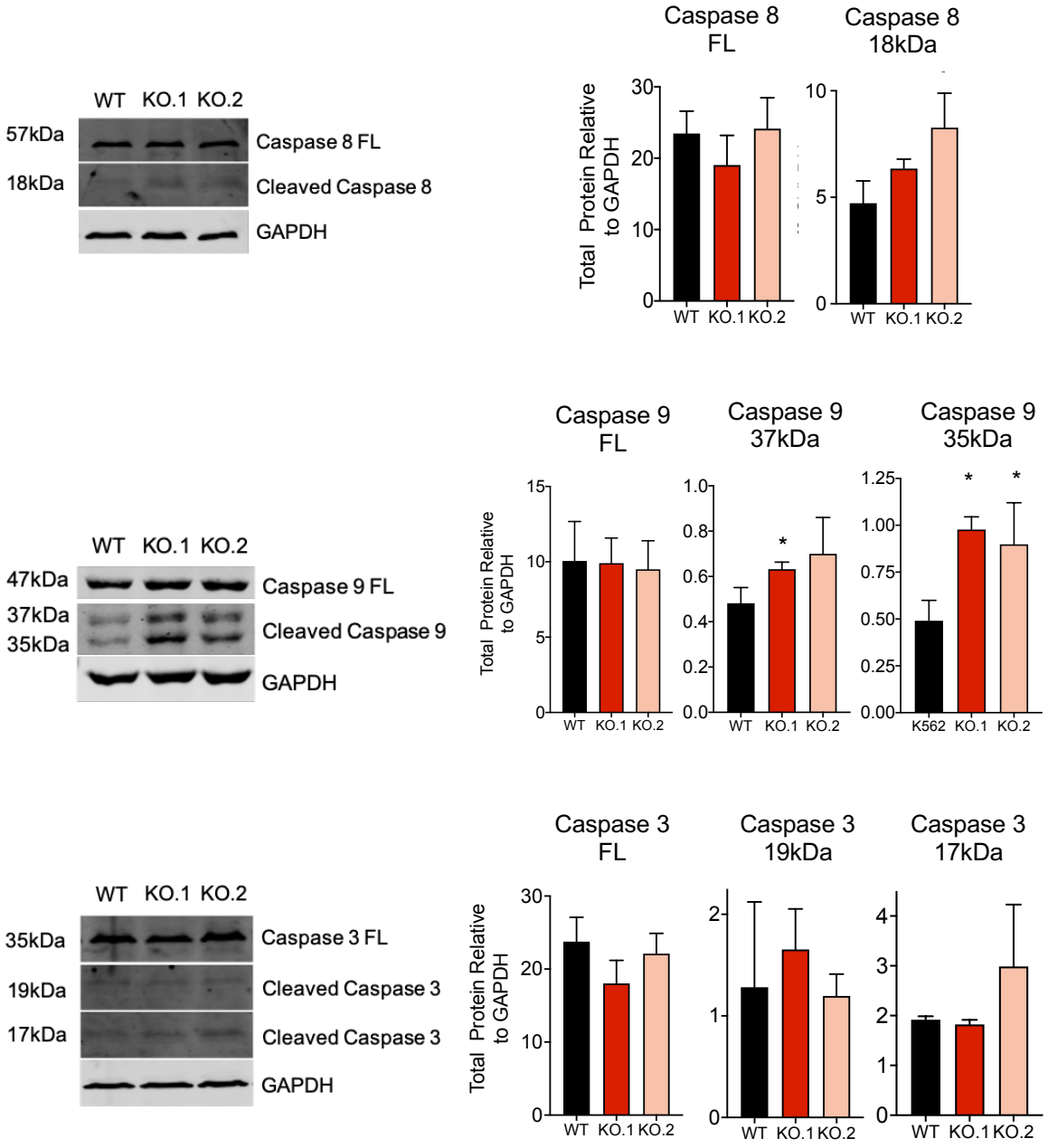


Figure 3.9 *CCDC26* KO cells are more apoptotic than WT cells. Cell lysates were collected from WT and KO cell lines and subjected to immunoblotting with a series of anti-caspase antibodies. Increased levels of the cleaved active forms of caspases 8, 9 and 3 can be observed in some cases in KO cells (FL = full length). Values represent the mean ± standard deviation (n=3). * P<0.05; **P<0.005 (unpaired, two-tailed *t* test).

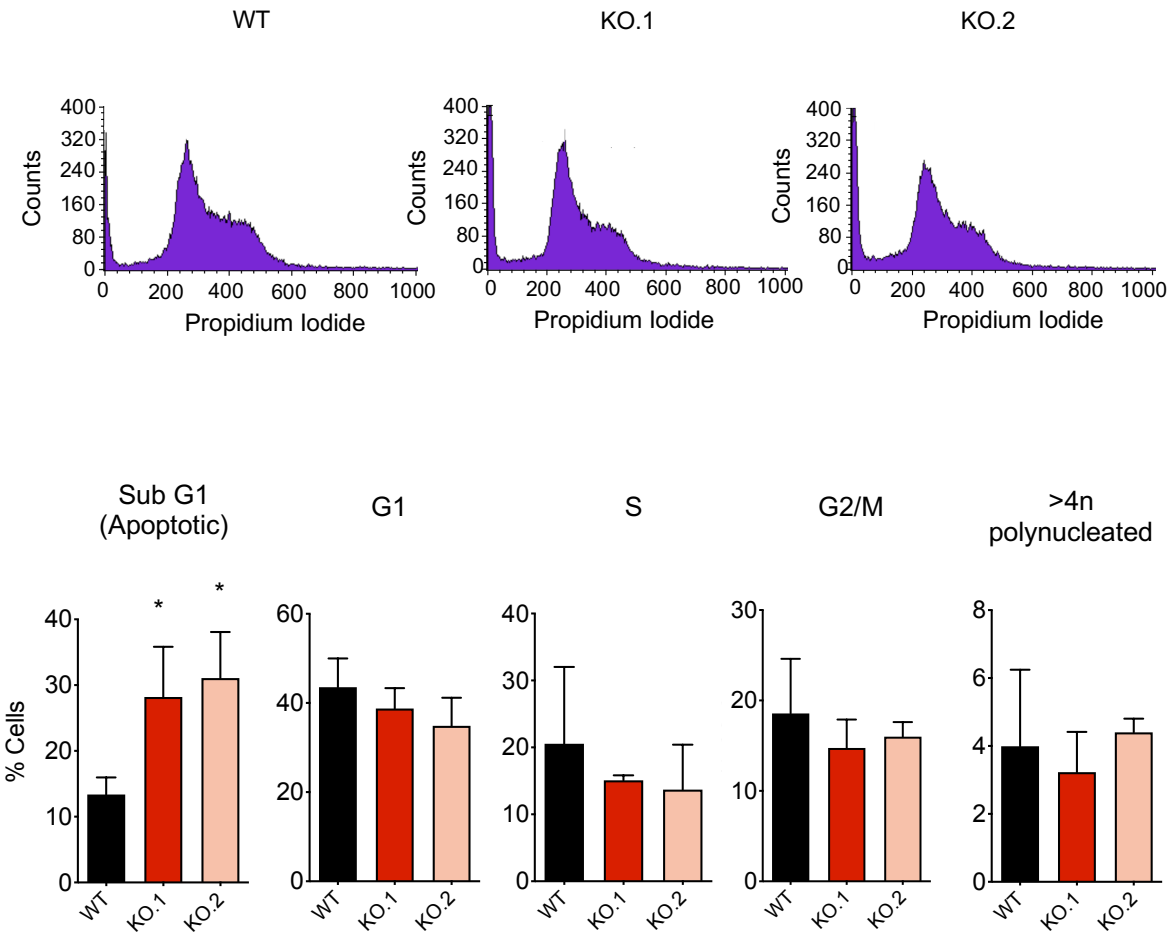


Figure 3.10 *CCDC26* KO cells show no changes in cell cycle. Propidium iodide staining followed by flow cytometry showed that there is a significantly greater proportion of apoptotic cells in the KO cell populations compared to WT (SubG1). However, there is no significant difference in the proportion of cells in each stage of the cell cycle between the WT and KO cells. Values represent the mean ± standard deviation (n=3). * P<0.05 (unpaired, two-tailed *t* test).

3.2.4 The Impact of *CCDC26* on Epigenetic Modifications

Previous analysis of caspase protein levels, as well as the subG1 proportion of cells, indicated that the *CCDC26* KO cells are more apoptotic (Figure 3.9 and 3.10). In relation to this, we investigated levels of DNA damage in cells given its link with cell death processes. Histone variant H2A.X is key in the cellular response to DNA damage, having its C-terminal tail rapidly phosphorylated at a unique Serine residue following the occurrence of a DNA double-strand break (DSB). The result is the formation of **γ**-H2A.X, which serves as an excellent DSB marker, and can be readily detected by its corresponding antibody (Rogakou et al. 1998). Immunofluorescence with anti-**γ**-H2A.X was performed to compare DNA damage in WT and KO cells (Figure 3.11). The WT cells showed very little, if any, fluorescence when cells were analysed by confocal microscopy, however, the majority of cells in both *CCDC26* KO lines showed several punctated **γ**-H2AX foci, indicating the location of DSBs (Figure 3.11). This provides further evidence that apoptosis has been induced in these cells.

Epigenetic modifications or chemical modifications of DNA and histones provide a broad and dynamic layering across the genome, allowing for multidimensional control of underlying gene sequences. Recently, many lncRNAs have been identified as key components in epigenetic regulation, frequently directing recruitment of chromatin-modifying proteins (Mercer and Mattick 2013; McHugh et al. 2015; Di Ruscio et al. 2013; Somasundaram et al. 2018). Arguably one of the best-known examples is lncRNA, *Xist*, known to recruit PRC2, amongst other proteins, to achieve X-chromosome inactivation, resulting in deposition of the

repressive H3K27me3 mark (McHugh et al. 2015). Following the observation that KO cells have increased levels of **γ**-H2A.X, we continued by investigating whether *CCDC26* has an effect on any other epigenetic modifications.

Generally, epigenetic modifications are considered as belonging to one of two categories, the first being histone post-translational modifications (PTMs). These reversible changes are applied to influence gene expression, typically by binding transcription factors, or aiding organisation of the chromatin structure and architecture. Histones can be modified by the addition of several types of chemical groups including acetyl and methyl groups (Davey et al. 2002; Allfrey, Faulkner, and Mirsky 1964) (Chapter 1.4.1). Global levels of other histone modifications were analysed by immunoblotting. Histones were extracted and blotted with a selection of anti-modified-histone antibodies, using total Histone H3 as a loading control (Figure 3.12A). Histone modifications related to gene activation as well as gene repression were analysed. H3K27ac and H3K9ac are markers of active enhancers and active gene promoters respectively (Creyghton et al. 2010; Gates et al. 2017), whilst H3K27me3 and H3K9me3 signify gene repression and heterochromatin (Nakayama, Rice, et al. 2001; Boyer et al. 2006). However, we observed no significant changes in the overall levels of these histone marks (Figure 3.12). We also analysed two other, less common histone modifications, H4K5ac and H4K16ac. H4K5ac mark is associated with a phenomenon known as “epigenetic bookmarking,” the means by which certain genes are not compacted during mitosis. Rather, an open chromatin state is maintained, allowing rapid gene transcription to resume, following mitosis, in a gene-specific manner (Park, Rehrauer, and Mansuy

2013). The latter is generally associated with transcriptional activation through regulation of chromatin structure (Zhang, Erler, and Langowski 2017). However, no significant changes in the global levels of these marks were observed either (Figure 3.12). Although there are countless other types of histone modifications, based on the selection that we have looked at, *CCDC26* does not appear to have a significant impact on how histones overall are post-translationally modified, other than H2A.X.

Further supporting this was our observation that total levels of histone methyltransferase enzymes, Enhancer of zeste homolog 2 (EZH2) and G9a, are also unchanged in KO cells (Figure 3.12B). EZH2 is a key component of the PRC2 complex, that alongside SUZ12 and EED, either mono-, di-, or tri-methylates Lysine 27 on Histone 3 (Margueron et al. 2008). Similarly, G9a either mono- or di-methylates Lysine 9 or 27 on Histone 3 (Shankar et al. 2013).

The second major class of epigenetic modifications is DNA methylation, the most widely recognised role of which is in gene repression (Curradi et al. 2002). In recent years, an increasing number of lncRNAs have been found to regulate gene expression by influencing DNA methylation, often via the DNMT enzymes (Chapter 1.1.4.4). To investigate global levels of DNA methylation, we performed immunofluorescence using anti-5-methyl-cytosine (anti-5mC) (Figure 3.13). The pattern of 5mC immunofluorescence was diffuse throughout the nuclei of both WT and *CCDC26* KO cells. However, there was a stark difference between the intensity of fluorescence between the cell lines. 5mC fluorescence intensity was measured for 200 individual nuclei (n=3) using FIJI imaging analysis software, and the values

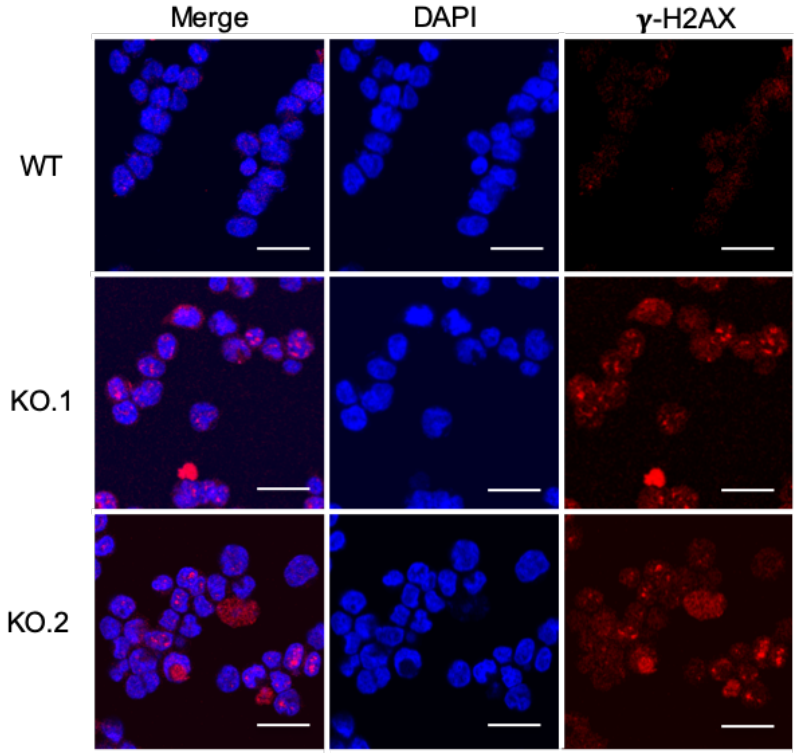


Figure 3.11 *CCDC26* KO cells have increased levels of DNA damage. Immunofluorescence shows increased γ-H2AX foci in *CCDC26* KO cells, indicating positions of DNA double-strand breaks. Cells were fixed with methanol and acetic acid and stained with DAPI nuclear stain (blue) and γ-H2AX (red) antibody. 2D confocal images were captured and compiled using FIJI image analysis software (scale bar = 25um).

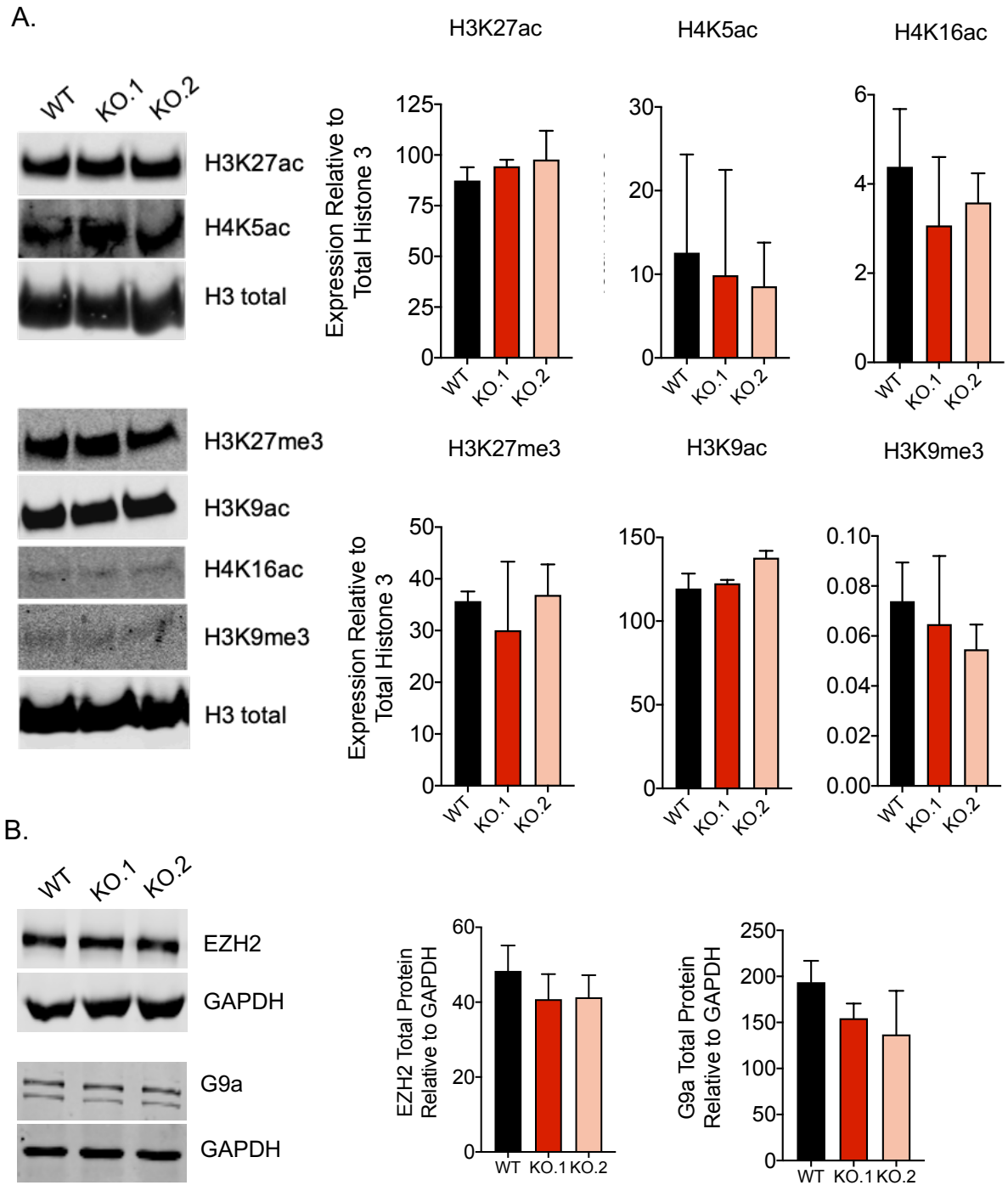


Figure 3.12 Global levels of histone modifications are unchanged in *CCDC26* KO cells. A. Histones were isolated from WT and KO cells and subjected to immunoblotting with antibodies against various histone modifications. No significant difference between global levels of these marks can be observed between WT and KO cells. **B.** Total cell lysates were extracted and subjected to immunoblotting with antibodies against histone modifying enzymes, EZH2 and G9a. No significant difference in total levels can be observed between WT and KO cells. Values represent the mean ± standard deviation (n=3) (unpaired, two-tailed *t* test).

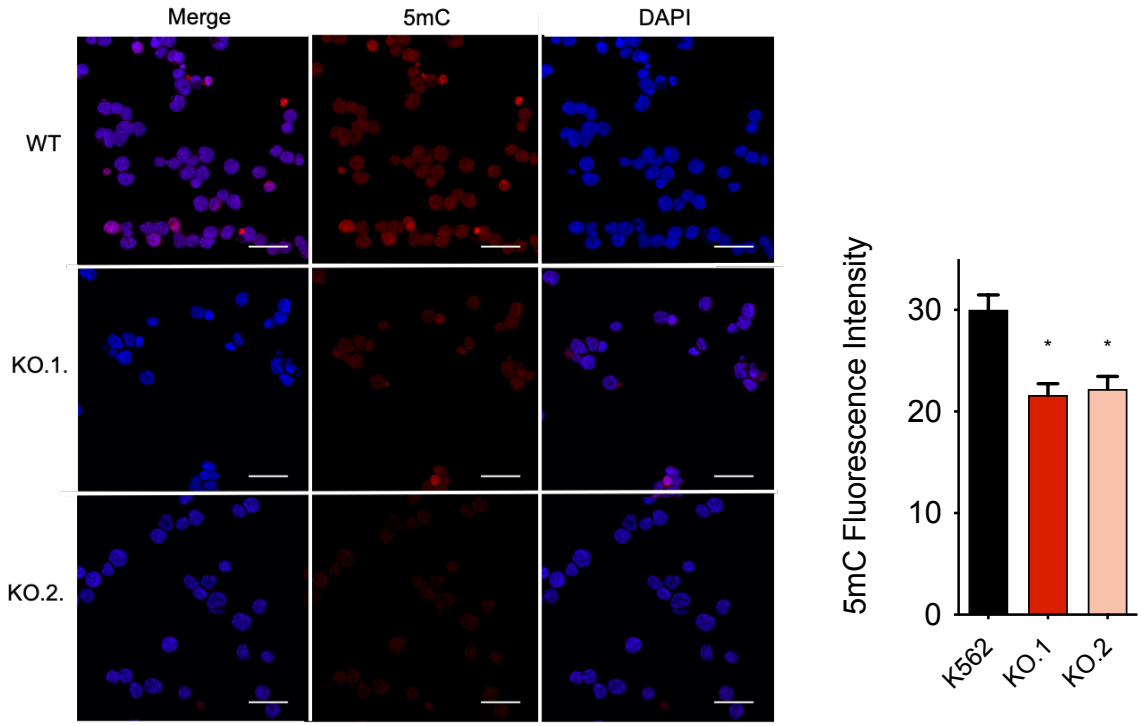


Figure 3.13 *CCDC26* KO cells are hypomethylated. 5mC Immunofluorescence showed a signficant decrease in 5mC fluorescence intensity in *CCDC26* KO cells compared to WT. Cells were fixed with methanol and acetic acid and stained with DAPI nuclear stain (Blue) and anti-5mC primary antibody (Red). 2D confocal images were taken and 5mC fluorescence intensity was measured for 200 individual nuclei, per replicate, using FIJI image analysis software (scale bar = 50um). Values represent the mean ± standard deviation (n=3). * P<0.05 (unpaired, two-tailed *t* test).

demonstrated a significant decrease in both KO cell lines (Figure 3.13). This suggests that globally, there is a reduction in the levels of DNA methylation in the absence of *CCDC26*.

3.3. Discussion

Our initial characterisation of *CCDC26* has involved investigating expression and localisation patterns of this lncRNA, as well as analysing its effects on cell viability and epigenetics. Through implementing both bioinformatical and hands-on techniques, we have expanded our basic knowledge of *CCDC26* and determined a *CCDC26* KO phenotype.

A few previous studies have demonstrated that *CCDC26* expression is cell line specific and is particularly abundant in myeloid cells of haematopoietic origin (Hirano et al. 2015; Yin et al. 2006). This has also been confirmed in our group by a colleague using qRT-PCRs (Wijesinghe et al. unpublished, manuscript in preparation) and followed by my analysis of RNA-seq and ChIP-seq data sets. I chose the immortalised K562 cell line, derived from a CML patient, as my model for this project, as these cells have consistently shown relatively high levels of *CCDC26* expression in ours as well as previous studies (Yin et al. 2006; Hirano et al. 2015). Although the AML cell line, HL60, has reportedly shown stronger expression than K562 (Hirano et al. 2015), the HL60 line has a complicated genomic structure. Double minute chromosomes, which are replaced by large extrachromosomal elements as cells are passaged, exist in these cells, and consist of regions spanning

the 8q24 locus, including CCDC26. The aberrant *CCDC26* transcripts, as well as fusions with other genes that have been observed, could influence results in unpredictable ways (Hirano et al. 2008). Given that K562 cells are not reported to possess such mutations encompassing *CCDC26,* yet expression levels are still relatively high, this cell line appeared suitable for subsequent analyses.

Analysis of expression of individual *CCDC26* isoforms is something that has not been investigated previously. At present, there are four known isoforms of *CCDC26,* annotated by NCBI RefSeq, which are transcribed from one of two TSSs. However, as of yet, no one has investigated whether or not all of these isoforms are present and functional in cells. Although Hirano et al. (2015) investigated other transcripts that arise from within intronic regions of the CCDC26 locus, they only consider two of the four *CCDC26* isoforms themselves (NR_130918.1 and NR_130917.1 respectively) (Hirano et al. 2015). Here, our analysis of ChIP-seq data sets for various histone modifications corresponding to active gene expression, showed stronger peaks for transcription arising from TSS2. Similarly, peaks corresponding to the actively transcribed gene body (H3K36me3) were stronger from exon 2 onwards, suggesting that expression of different isoforms is not uniform. For the first time, we have determined which *CCDC26* isoforms are present in K562 cells and assessed their relative levels. In accordance with ChIP-seq profiles, we found that expression of *CCDC26* occurred almost exclusively from TSS2, with isoforms 1 and 2 being predominant in K562 cells. It is possible that TSS1 is inactive in K562 cells. The small ChIP-seq peaks shown in Figure 3.3 indicate that there may be a small amount of active transcription from this region. It is possible that transcription

is initiated but then aborted, as indicated by the lack of H3K36me3 marks in the gene body region.

A critical first step in identifying the biological role of any cellular component, is to determine where it resides. Our approach combined bioinformatical analysis with qRT-PCRs on cellular fractions, with results agreeing with past reports that all *CCDC26* isoforms present in K562 cells are largely nuclear (Hirano et al. 2015). However, for the first time, here we show that *CCDC26* localisation is not limited to the nucleus, but rather it occupies several cellular compartments. Our analysis of long RNAseq data demonstrates that a large proportion of nuclear *CCDC26* is chromatin-bound, with low levels also found in the nucleolus. However, a small amount of the polyA+ *CCDC26* transcript is also present in the cytosol (Consortium 2012). Its chiefly nuclear localisation and apparent preference for binding chromatin points towards a potential role for *CCDC26* in gene regulation. However, its presence in multiple compartments potentially reflects that *CCDC26* is multifunctional.

Whilst several studies have investigated the effects of shRNA- and siRNA-mediated *CCDC26* KDs, there is currently no information on the effects of a *CCDC26* KO (Hirano et al. 2015; Wang, Hui, et al. 2018; Peng and Jiang 2016; Cao et al. 2018; Yan et al. 2019). Comparisons of KO and WT cells are very informative regarding a component's function; our group possesses two previously generated *CCDC26* CRISPR KO cell lines. We began characterising these cell lines with some initial experiments measuring cell viability. Previous reports have been inconsistent with

regards to the involvement of *CCDC26* in cell viability, death and apoptosis. In myeloid cell lines, down-regulation of *CCDC26* has mostly been associated with more pronounced apoptosis and slower growth (Hirano et al. 2015; Yin et al. 2006). Similarly, pancreatic cancer and glioma cells have too demonstrated growth arrest and increased cell death upon transfection with *CCDC26* siRNAs (Wang, Hui, et al. 2018; Peng and Jiang 2016). However, more recently, gastrointestinal stromal tumour cells have demonstrated conflicting results, with studies showing that *CCDC26* KDs in combination with Imatinib therapy results in improved cell viability, increased proliferation and reduced levels of apoptosis (Cao et al. 2018; Yan et al. 2019). Moreover, *CCDC26* KD CML cells have also shown similar results when grown in nutrient-poor, low serum media (Hirano et al. 2015). Here, our results appear to support the former studies; analysis of pro- and cleaved- caspase levels as well as propidium iodide staining and growth curves, indicate that *CCDC26* KO cells grow slower, and are more apoptotic than WT cells. It is possible that these inconsistent results reflect cell-specific functions of *CCDC26*. In the future, it would be interesting to see whether expression of different *CCDC26* isoforms determines this phenotype.

Although we see increases in levels of both initiator caspases 8 and 9 in KO cells, this difference is only statistically significant for caspase 9. Given that these caspases are activated in the extrinsic and intrinsic apoptotic pathways respectively, it is possible that this is indicative of activation of the latter pathway. This is perhaps also reflected by the observed increase in DNA damage in *CCDC26* KO cells, as shown by **γ**-H2AX immunofluorescence. This histone mark arises as a result of the

presence of DNA DSBs. DNA damage is known to trigger apoptosis if the damage in question cannot be repaired (Roos and Kaina 2006). However, apoptosis itself also leads to DNA fragmentation and hence increased levels of **γ**-H2AX (Rogakou et al. 2000). Consequently, it is not clear here whether the DNA damage acts as an intrinsic trigger, or whether apoptosis has already been induced and the DNA damage we observe is a consequence of that. Unexpectedly, we observed no significant change in caspase 3 levels. This is unexpected since activated caspases 8 and 9 both cleave caspase 3. The bands detected with this antibody were very weak, which subsequently gave rise to some relatively large error bars. Therefore, it is possible that there are technical issues with this result. Alternatively, a different effector caspase that we are yet to analyse (e.g. caspases 6 and 7), may be cleaved and activated in these cells (Riedl and Shi 2004).

If a *CCDC26* KO results in DNA damage subsequently triggering apoptosis, we'd perhaps expect to see some evidence of activation of DNA damage cell cycle checkpoints and hence cell cycle arrest (Elledge 1996). However, this is not something we have observed; FACS following PI staining showed no significant changes between WT and KO in the proportion of cells at each stage of the cell cycle. This might suggest that the DNA damage we observe is a consequence rather than a cause of apoptosis. To further confirm this, it would be interesting to perform a more sensitive analysis of the cell cycle. One future experiment might be to repeat this having synchronized the cells with a chemical such as hydroxyurea (Walker and Wanda 1987) rather than working with a mixed cell population. The

length of time it takes for cells to progress through each stage of the cell cycle could then be measured and compared.

In recent years, there has been a surge of studies reporting roles of lncRNAs in epigenetic gene regulation (Holoch and Moazed 2015). *HOTAIR, Xist, AIR* and *Kcnq1ot1* are just a few examples of nuclear, chromatin-binding lncRNAs that impact the epigenetic landscape (Kohlmaier et al. 2004; Pandey et al. 2008; Nagano et al. 2008; Rinn et al. 2007). As previously mentioned, *CCDC26* is primarily localised in the nucleus and a large proportion of this appears to be chromatin bound. This suggests that *CCDC26* might be important in chromatin regulation. All of this considered, we decided to assess the effect of *CCDC26* on global levels of various epigenetic marks. Numerous studies have reported that lncRNAs play a role in epigenetic gene regulation by interacting directly with the proteins that deposit, read and eliminate histone modifications (Kohlmaier et al. 2004; Pandey et al. 2008; Nagano et al. 2008). lncRNA *Xist* for example, binds PRC2, localising it to the inactive X chromosome where it can deposit H3K27me3 (Kohlmaier et al. 2004). Similarly, *Kcnq1ot1* and *AIR* lncRNAs guide the histone methyltransferase enzyme, G9a, to local genomic regions to mediate *in cis* repression of surrounding genes (Pandey et al. 2008; Nagano et al. 2008). However, aside from changes in levels of **γ**-H2AX, no major changes in other histone modifications were observed. It is important to note however, that this can only be concluded with regards to global levels of modifications. Small local changes in histone modifications have not been measured and may not be detectable at a global level. Therefore, we cannot categorically argue that *CCDC26* in no way impacts histone modifications.

However, we did observe significant changes in the levels of DNA methylation, with *CCDC26* KO cells exhibiting a significant reduction in 5mC immunofluorescence. Interestingly, this appears to be a global effect, as we see an overall reduction in 5mC levels within nuclei. It is possible that this change in the methylome leads to increased apoptosis. Previous studies have demonstrated that hypermethylation can result in repression of apoptotic and tumour suppressor genes, leading to increased tumorigenesis (Li, Liu, et al. 2017; Wang et al. 2017; Wen et al. 2018; Yu et al. 2013; Stirzaker et al. 1997; Schmelz et al. 2005). DNMT inhibitors have even been implemented as anti-cancer therapies (Christman 2002; Cheng et al. 2004). A more detailed analysis of differentially methylated regions in WT and KO cells, will help determine this.

Multiple lncRNAs have demonstrated roles in regulating DNA methylation, however this typically involves local effects at specific loci (Di Ruscio et al. 2013; Wang et al. 2015; Gao et al. 2019; Guo et al. 2019). For example, *ecCEBP*, is a DNMT1-binding lncRNA whose depletion results in increased methylation of the *CEBPA* promoter from which it is transcribed, resulting in silencing of the mRNA *CEBPA* (Di Ruscio et al. 2013). It is rare for a lncRNA to affect DNA methylation on a genome-wide scale, although a few instances have been reported. *DACOR1* expression has been shown to increase genome-wide DNA methylation in colon cancer cells, by indirectly increasing levels of the methyl donor SAM (Merry et al. 2015; Somasundaram et al. 2018). Similarly, lncRNA *PARTICLE* has been associated with increased DNMT1 activity and consequently genome-wide hypermethylation (O'Leary et al. 2017).

Overall so far, we have shown that although *CCDC26* KO does not condemn all K562 cells to immediate death, the effects of its absence are enough to slow growth and trigger increased levels of apoptosis. Furthermore, we have shown that its absence leads to notable changes in DNA methylation levels.

CHAPTER 4

CCDC26 REGULATES DNA METHYLATION BY CONTROLLING SUBCELLULAR LOCALISATION OF DNMT1

4.1 Introduction

The previous chapter provided evidence that *CCDC26* regulates DNA methylation levels, with KO cells demonstrating global DNA hypomethylation. One of the fundamental mechanisms by which gene expression is regulated, is the methylation of DNA at cytosine residues. These epigenetic marks are established and later maintained by members of the DNA methyltransferase (DNMT) family; DNMT3A/DNMT3B and DNMT1 respectively (Okano et al. 1999). DNA methylation is a dynamic signature whose role is context-dependent, commonly associated with gene repression when deposited at gene promoter regions, yet linked with processes such as splicing and transcriptional elongation when found within the gene body (Hellman and Chess 2007; Shukla et al. 2011; Jones 2012). Although a great deal of research has been performed on the diversity of this mark in different settings, our understanding of how it is targeted to specific sequences is limited.

The recent discovery and characterisation of lncRNAs and their functions has unveiled interplay between these molecules and DNA methylation (Zhao, Sun, and Wang 2016). This can largely be attributed to interactions with the DNMTs. DNMT1 has been found to interact with hundreds of different lncRNAs, which offer a means of direct targeting of the protein to methylate specific regions of the genome (Merry et al. 2015; Di Ruscio et al. 2013). LncRNAs have also shown interactions with other

DNMTs; the lncRNA, *Dum*, functions in regulating myogenesis by recruiting DNMT3A, DNMT3B, as well as DNMT1, and silencing adjacent genes (Wang et al. 2015). Other examples of lncRNAs have demonstrated that they can also employ indirect methods of regulating DNA methylation. *lincRNA-p21*, for example, recruits DNMT1 indirectly, through binding Heterogeneous nuclear ribonucleoprotein K (HNRNPK) and thus, by maintaining their heterochromatic state, represses genes encoding pluripotency factors to prevent cell reprogramming (Bao et al. 2015).

Our findings so far show that in the absence of *CCDC26*, overall levels of DNA methylation reduce significantly. At present, of the publications available that discuss *CCDC26*, none have investigated its involvement in regulating DNA methylation. In this chapter, we further elaborate on this result. Here we have investigated the mechanism of *CCDC26*-mediated DNA methylation regulation by examining DNMT levels and localisation in *CCDC26* KOs, as well as potential interactions between *CCDC26* and DNMT1.

4.2 Results

4.2.1 Total Levels of DNMT Enzymes are Unchanged in *CCDC26* KO Cells

The global decrease in 5mC in *CCDC26* KO cells (Chapter 3) could be a result of indirect changes in the levels of the DNA methyltransferase enzymes. Being the enzymes responsible for catalysing DNA methylation, any changes in expression of these proteins could affect genomic methylation levels. We began with DNMT1, the maintenance methyltransferase; qRT-PCRs and western blotting using anti-DNMT1

were performed to determine the levels of total DNMT1 mRNA (Figure 4.1A) and protein (Figure 4.1B) respectively. Protein levels were measured as described in 2.7.8, and normalised to the housekeeping protein, GAPDH, to account for differences in sample loading and transfer efficiency. Similarly, sample variations were corrected for in the qRT-PCRs, by normalising first to GAPDH mRNA expression, and then further normalising to the WT control, as described in 2.6.4. Although mRNA levels were slightly greater in the KO cells, no significant differences were observed between protein levels, indicating that hypomethylation is not due to a decrease in DNMT1 levels.

We also examined RNA and protein levels of the *de novo* methyltransferases, DNMT3A and DNMT3B. mRNA levels for these two genes generally appeared marginally higher in the KO cells, however, this did not appear to affect protein levels, which remained consistent between cell lines (Figure 4.2 and 4.3). Given the principal role of these DNMTs in *de novo* methylation, their protein levels were also considerably lower than that of the primary methyltransferase, DNMT1, as expected. DNMT3A and DNMT3B levels were approximately 4% and <1% (Figure 4.2 and 4.3) than that of housekeeping GAPDH levels, in comparison to DNMT1, which demonstrated levels of ~20% (Figure 4.1).

This further indicates that DNA hypomethylation is not caused by any significant changes in DNMT levels. We initially chose to focus on just these members of the DNMT family, as DNMT2 and DNMT3L are generally regarded as non-canonical. DNMT3L is a catalytically inactive protein that acts as a cofactor for DNMT3A

(Bourc'his et al. 2001; Jia et al. 2007), and to our knowledge has not shown evidence of any lncRNA-mediated regulation. The catalytic activity of DNMT2 on the other hand is not targeted for DNA methylation at all, but rather it functions in methylation of tRNA (Goll et al. 2006).

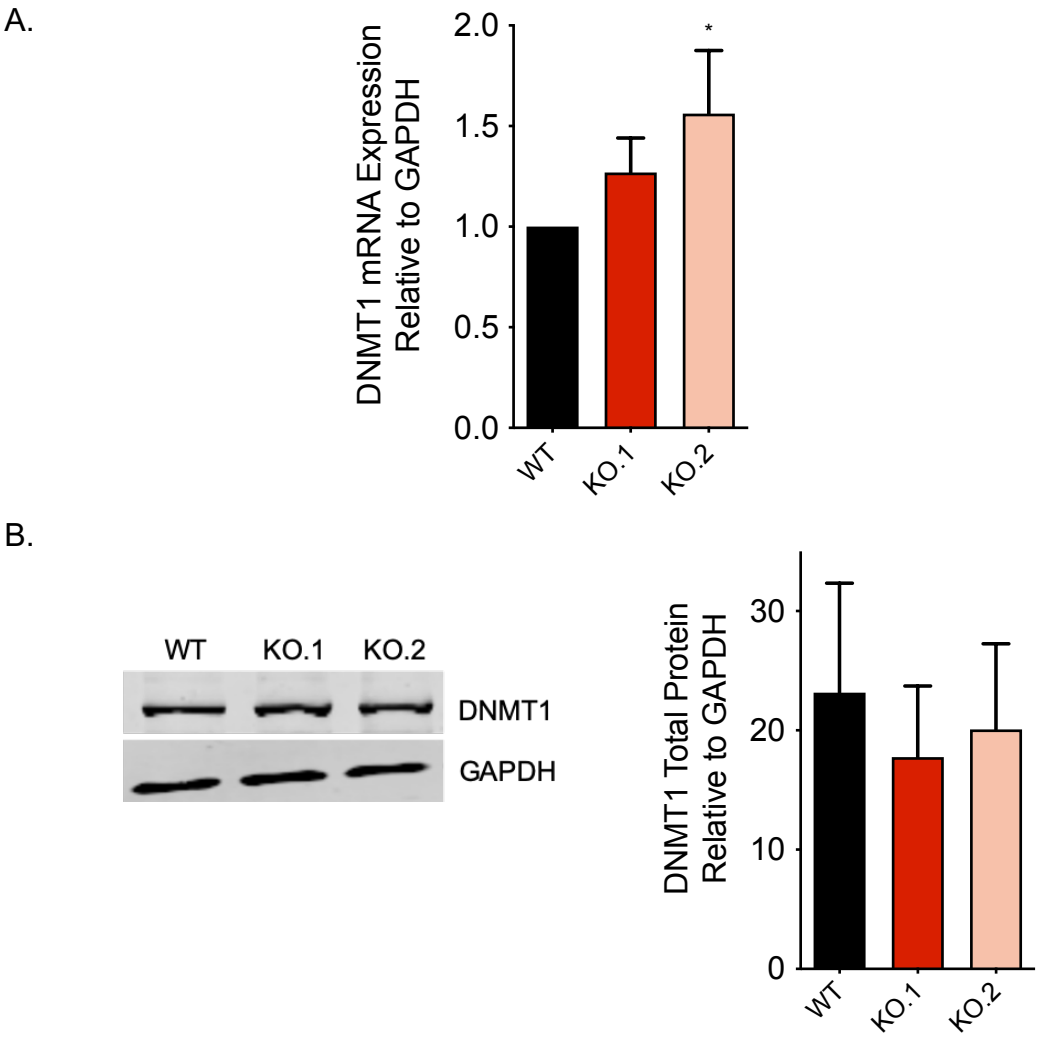
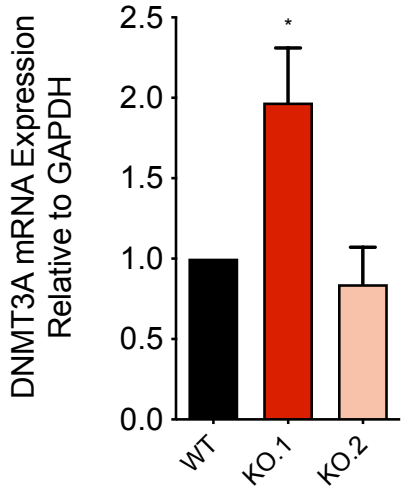
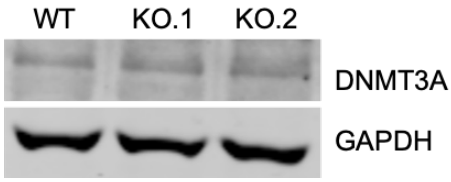


Figure 4.1. Total DNMT1 levels do not change upon *CCDC26* KO. A. qRT-PCRS demonstrate a small increase in DNMT1 mRNA expression levels in *CCDC26* KO cells compared to WT. **B.** However, total DNMT1 protein levels are not significantly different between WT and KO cells. DNMT1 mRNA and protein levels were measured relative to GAPDH. Values represent the mean ± standard deviation (n=3). * P<0.05 (unpaired, two-tailed *t* test).

A.



B.



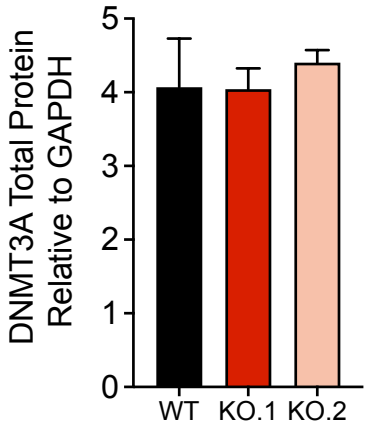
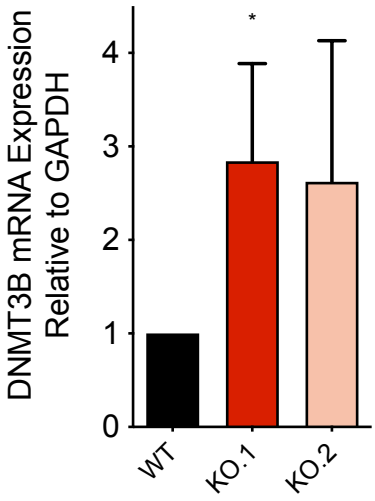
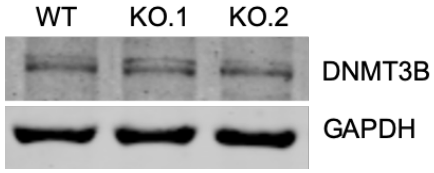


Figure 4.2. Total DNMT3A levels do not change upon *CCDC26* KO. A. qRT-PCRS demonstrate a small increase in DNMT3A mRNA expression levels in *CCDC26* KO.1 cells compared to WT. **B.** However, total DNMT3A protein levels are not significantly different between WT and KO cells. DNMT3A mRNA and protein levels were measured relative to GAPDH. Values represent the mean ± standard deviation (n=3). * P<0.05 (unpaired, two-tailed *t* test).

A.



B.



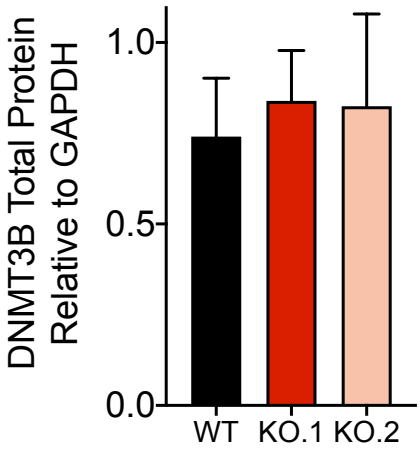


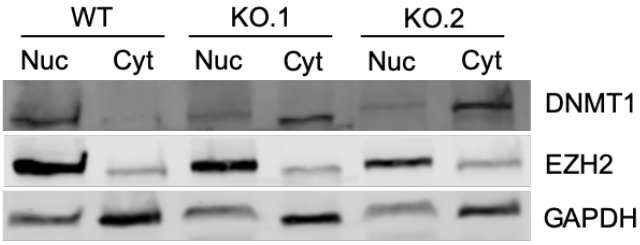
Figure 4.3. Total DNMT3B levels do not change upon *CCDC26* KO. A. qRT-PCRS demonstrate a small increase in DNMT3B mRNA expression levels in *CCDC26* KO cells compared to WT. **B.** However, total DNMT3B protein levels are not significantly different between WT and KO cells. DNMT3B mRNA and protein levels were measured relative to GAPDH. Values represent the mean ± standard deviation (n=3). * P<0.05 (unpaired, two-tailed *t* test).

4.2.2 *CCDC26* KO Implicates DNMT1 Subcellular Localisation

Our results show that DNMT levels in KO cells are unaffected and the expected level of DNMTs required to establish and maintain DNA methylation is present. In spite of the presence of DNMTs, KOs show overall genomic hypomethylation. A possible reason behind this could be either loss of enzymatic activity, or unavailability of these enzymes at genomic sites of action. Therefore, we next sought to determine whether the enzymes are actually available to perform their function, starting with DNMT1. This first involved checking it is correctly localised to the nucleus of the cell, where it can interact with genomic DNA. Nuclear and cytosolic cellular fractions were purified, and protein from each was extracted and subjected to western blotting with anti-DNMT1 (Figure 4.4A). This demonstrated a dramatic shift in the ratio of nuclear: cytosolic DNMT1 levels. In contrast to WT cells, where the majority of DNMT1 was found in the nuclear fraction, a significant increase in cytosolic DNMT1 was observed in both KO cell lines (Figure 4.4A). Other nuclear proteins such as EZH2 (Grzenda et al. 2013) did not show a similar shift to the cytosol in KO cells. In both WT and KOs, GAPDH, which is a marker of the cytosolic fraction (Mazzola and Sirover 2003), was also enriched in the cytoplasm. This indicates that nuclear and cytosolic fractions were pure and re-localisation was specific to DNMT1. To further confirm this result, we also performed anti-DNMT1 immunofluorescence, and the microscopic observation was in agreement with the immunoblot results. A large proportion of DNMT1 can be seen in the cytosol of KO cells (Figure 4.4B). The pattern of distribution of the remaining small amount of nuclear DNMT1 also changes from a diffuse arrangement in the WT, to a more punctated pattern in the KOs.

To determine whether these changes in protein localisation are also observed in other DNMTs, the western blot was repeated using anti-DNMT3A and anti-DNMT3B. The subcellular localisation of these enzymes however remained consistent between all three cell lines, suggesting the effect is specific to DNMT1 (Figure 4.5).

A.



B.

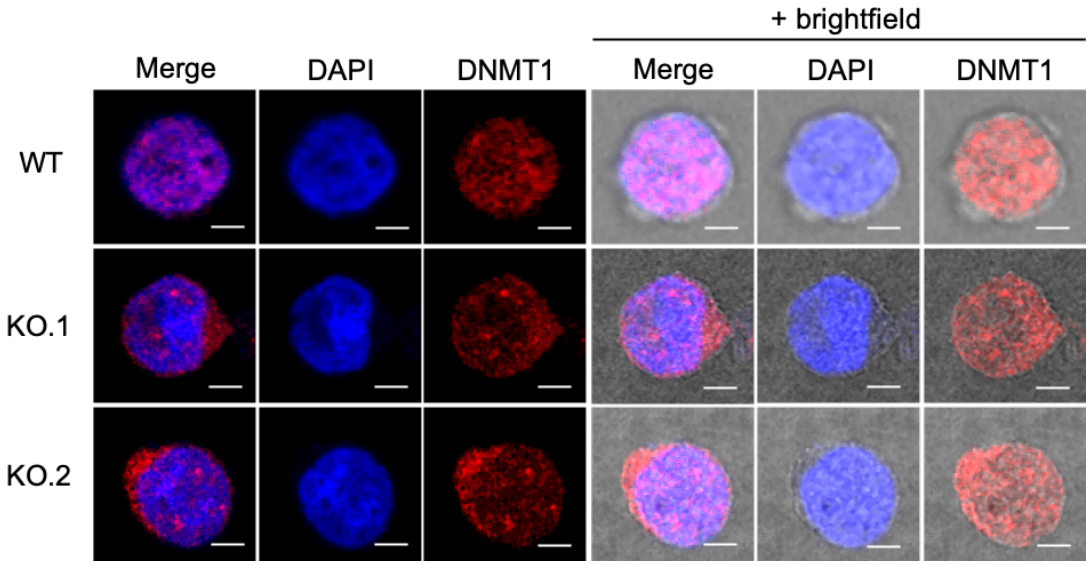


Figure 4.4 DNMT1 is primarily localised in the cytosol in *CCDC26* KO cells. A. Immunoblotting for DNMT1 on nuclear and cytosolic protein fractions shows a shift in the subcellular localisation of DNMT1. DNMT1 is almost exclusively nuclear in the WT cells, but appears both nuclear and cytosolic in *CCDC26* KO cells. EZH2 and GAPDH are used as nuclear and cytosolic markers respectively (nuc = nuclear protein fraction; cyt = cytosolic protein fraction). **B.** DNMT1 was also visualised in WT and *CCDC26* KO cells via anti-DNMT1 (Red) immunofluorescence. Nuclei were stained with DAPI (Blue). The outline of the cell membrane can be seen with the addition of the brightfield lens in the right-hand panels (scale bar = 5um).

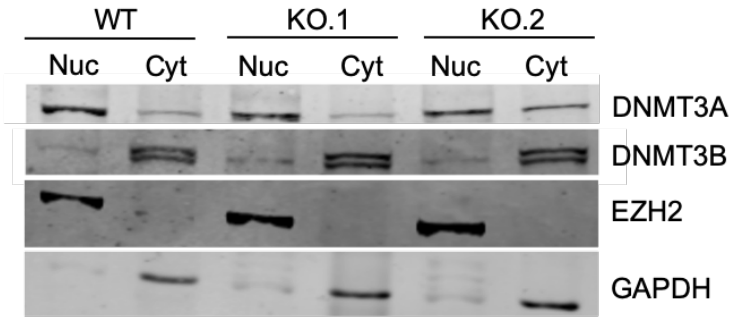


Figure 4.5 DNMT3A and DNMT3B localisation is unaffected by *CCDC26* KO. Immunoblotting for DNMT3A and DNMT3B on nuclear and cytosolic protein fractions show no significant difference in subcellular localisation between WT and *CCDC26* KO cells. EZH2 and GAPDH are used as nuclear and cytosolic markers respectively (nuc = nuclear protein fraction; cyt = cytosolic protein fraction).

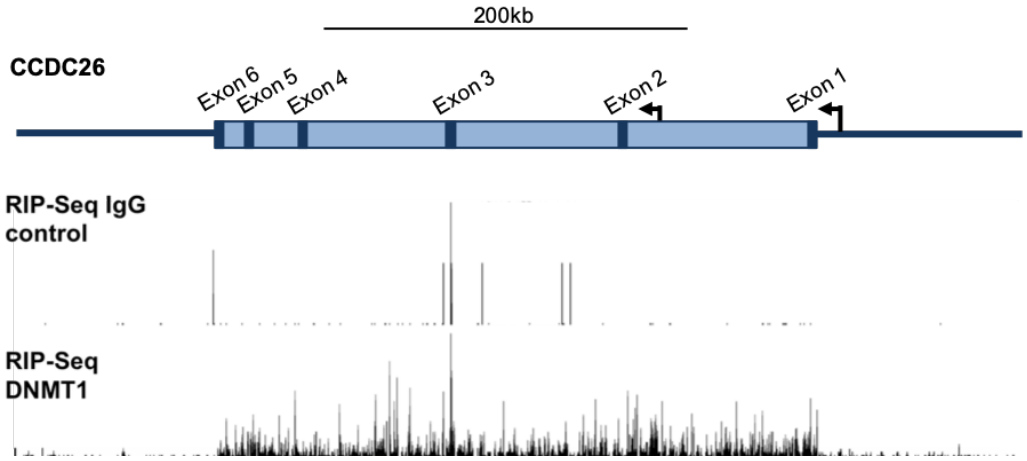
4.2.3 *CCDC26* may interact with DNMT1

Our results raise an interesting possibility that lncRNAs such as *CCDC26* might regulate the subcellular localisation of DNMT1. Therefore, it is important to establish whether this could be due to a direct interaction between *CCDC26* and DNMT1, or an indirect effect of *CCDC26* knockout. DNMT1 has previously been shown to bind and undergo regulation by numerous different lncRNAs (Merry et al. 2015; Di Ruscio et al. 2013). It has also been suggested that DNMT1 has higher affinity for RNA than DNA (Di Ruscio et al. 2013). A common method for analysing interactions between protein and RNA is via RNA immunoprecipitation (RIP) (Chapter 2.8). In recent years, a number of protein-RNA interactions have been studied using RIP or variations of this method (Di Ruscio et al. 2013; Hendrickson et al. 2016; Xu et al. 2019; Hui et al. 2019; Bai et al. 2019). At least two datasets exploring DNMT1-RNA interactions have been published recently in *CCDC26*-expressing cell lines (Di Ruscio et al. 2013; Hendrickson et al. 2016). I first adopted a bioinformatics approach and re-analysed DNMT1 RIP-seq data from Di Ruscio et al., remapping the data to the CCDC26 locus. The RIP binding profiles in HL60 cells show stronger binding of DNMT1 to *CCDC26* in comparison to an IgG RIP control (Figure 4.6A) (Di Ruscio et al. 2013). Furthermore, analysis of another dataset produced by a RIP method, formaldehyde-RIP-seq (fRIP-seq), also showed enrichment of *CCDC26* in DNMT1-bound RNAs in K562 cells (Hendrickson et al. 2016), suggesting an interaction may exist.

To test this further, I independently performed DNMT1 RIP using an anti-DNMT1 antibody. Anti-IgG was used to produce a control sample. The protein-bound-RNA

pulled down with each of the antibodies was used to perform qRT-PCRs with primers specific to *CCDC26*. The RNA levels were measured relative to the input (see Chapter 2.8). Anti-DNMT1 pull-down RNA showed approximately three times more enrichment of *CCDC26* as compared to the IgG control (Figure 4.7). In accordance with the previous RIP-seq data sets, this result also suggests *CCDC26* may interact with DNMT1. To confirm this, further controls will be required in the future. It will important to perform qRT-PCRs with primers specific to lncRNAs that have either been confirmed to interact with DNMT1 (positive control) or do not interact with DNMT1 (negative control). This will help confirm whether the significant increase in *CCDC26* pulled-down with DNMT1 is due to a specific interaction, rather than a general, non-specific 'stickiness' of the protein for RNA.

A.



B.

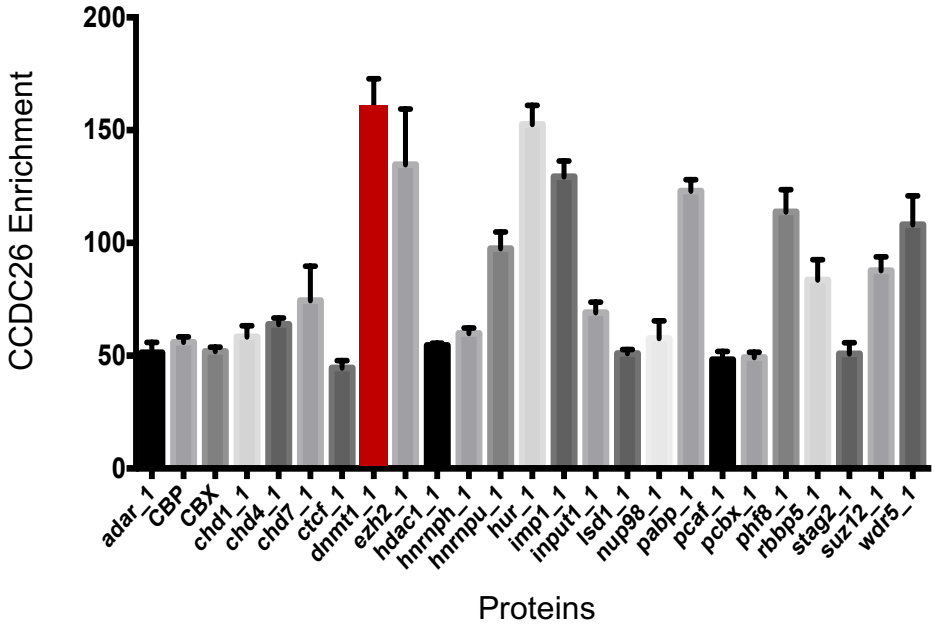


Figure 4.6 DNMT1 RIP-seq suggests *CCDC26* may interact with DNMT1. Analysis of previous DNMT1 **A.** RIP-seq (Di Ruscio et al. 2013) (GEO Accession: GSE32162) and **B.** fRIP-seq (Hendrickson et al. 2016) (GEO Accession: GSE67963) data sets both show enrichment for *CCDC26* in RNA pulled down with DNMT1, compared to IgG control. Data available at NCBI Gene Expression Omnibus (GEO).

A. B.

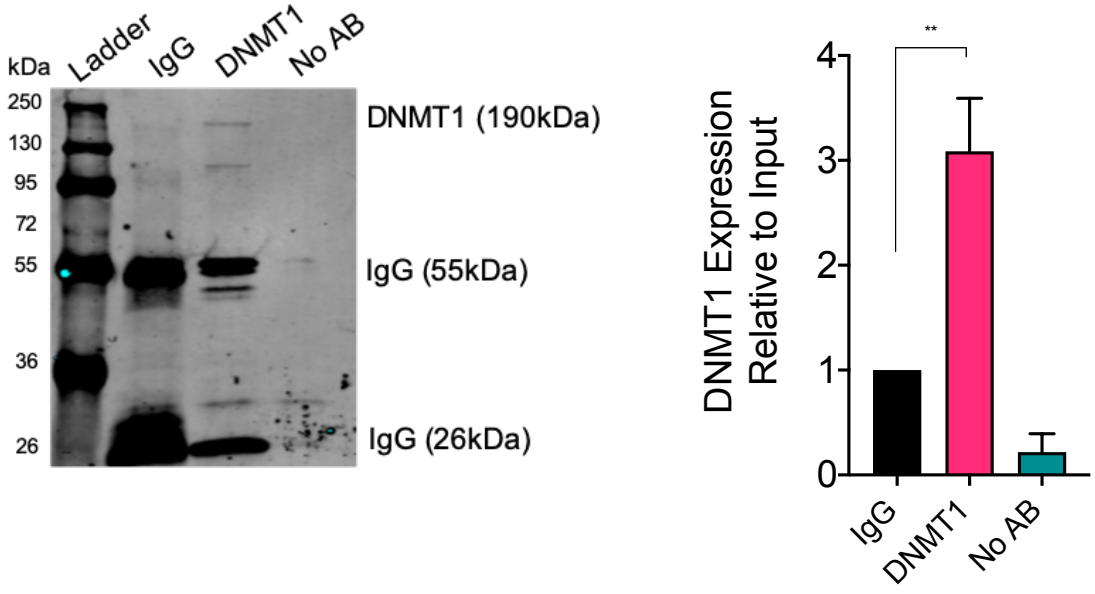


Figure 4.7 DNMT1 RIP shows that *CCDC26* may interact with DNMT1. A. Protein-RNA complexes pulled down with either anti-IgG, anti-DNMT1 or no antibody, were immunoblotted with anti-DNMT1, to ensure that the DNMT1 protein was correctly pulled down. **B.** RNA pulled-down in each IP was purified, converted to cDNA and subjected to qRT-PCR with *CCDC26* primers, to determine how much *CCDC26* was pulled down relative to the input in each instance. Approximately 3X more *CCDC26* was pulled down with anti-DNMT1, compared to anti-IgG. Values represent the mean ± standard deviation (n=3). * P<0.05; **P<0.01 (unpaired, two-tailed *t* test).

4.2.4 DNMT1 Genes are Up-Regulated in KO cells

We hypothesised that if a large proportion of DNMT1 protein is present in the cytosol of KO cells, then a sizeable portion of DNMT1 would be unable to carry out its primary function in the nucleus. Consequently, these cells should behave similarly to DNMT1 KD cells. As an additional confirmation of our previous results, we investigated expression levels of a selection of six genes, previously shown to be significantly impacted by DNMT1 and methylation levels in myeloid cells. These were Protein Tyrosine Phosphatase Non-Receptor Type 6 (PTPN6) (Li, Liu, et al. 2017; Wang et al. 2017), Cyclin Dependent Kinase Inhibitor 1A (CDKN1A) (Milutinovic et al. 2004; Schmelz et al. 2005), Cyclin Dependent Kinase Inhibitor 2B (CDKN2B) (Yu et al. 2013; Herman et al. 1996), CD9 (Kim 2012), VAV1 (Ilan and Katzav 2012; Fernandez-Zapico et al. 2005) and JUNB (Yang et al. 2003; Fiskus et al. 2009) all of which have previously demonstrated upregulation in response to DNMT1 down-regulation or DNA hypomethylation. We also measured levels of *IGF1*, which has previously been repressed in response to DNMT1 inhibition (Pastural et al. 2007). qRT-PCRs demonstrated that out of 7 genes we tested, five (PTPN6, CDKN1A, CDKN2B, CD9 and VAV1) were significantly upregulated in both KO cell lines as reported in past studies on DNMT1 KD or inhibition (Figure 4.8). JUNB, albeit not significantly, also demonstrated upregulation. The only exception was IGF1 which was significantly downregulated in both KO cells. This suggests that DNMT1-regulated genes are affected in *CCDC26* KO cells, presumably because of cytoplasmic mis-localisation of DNMT1, leading to DNA hypomethylation.

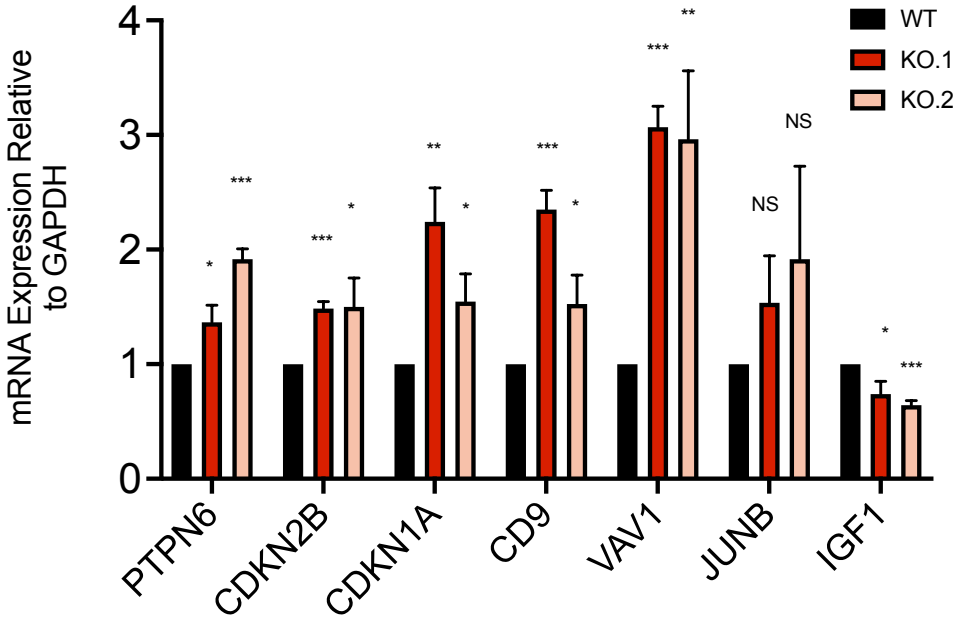


Figure 4.8 DNMT1-regulated genes are differentially expressed in *CCDC26* KO cells, similarly to DNMT1 KD cells. qRT-PCRs were used to measure levels of various genes whose expression have previously been shown to be impacted by DNMT1 depletion or DNA hypomethylation in myeloid leukemia. mRNA levels were measured relative to GAPDH. Values represent the mean ± standard deviation (n=3). * P<0.05; **P<0.01; ***P<0.001; NS = Not significant, *P*>0.05 (unpaired, two-tailed *t* test).

4.2.5 Exogenous *CCDC26* expression does not rescue DNMT1 localisation

To complement our results so far, we expressed *CCDC26* exogenously in *CCDC26* KO cell lines, with the aim of rescuing our observed phenotype. Previously, *CCDC26* overexpression cell lines (o/x) were generated by stably transfecting WT, KO.1 and KO.2 cells with a pEF6 plasmid into which the spliced isoform of *CCDC26* had been cloned. Simultaneously, control cell lines were also generated by transfecting cells with empty pEF6 plasmid, lacking *CCDC26*. After ensuring the empty control cells expressed *CCDC26* at levels comparable to original cells (Appendix II, Figure 1), *CCDC26* expression in the o/x cells was compared to empty controls by performing qRT-PCR analysis. WT, KO.1 and KO.2 o/x cells all similarly showed an approximately 25-fold increase in *CCDC26* expression, compared to WT control cells (Figure 4.9). However, despite *CCDC26* overexpression, DNMT1 localisation was not rescued in the KO o/x cells, and still appeared predominantly cytosolic, similar to the empty vector controls and non-transfected *CCDC26* KO cells (Figure 4.10). To see if *CCDC26* expression can rescue KOs of DNA damage, we also performed **γ**-H2A.X immunofluorescence on o/x cells. However, there was no significant change in the DNA damage levels of the o/x cells (Figure 4.11A). qRT-PCRs were also carried out to check if the expression of genes affected in KOs (4.2.4) can be rescued. However, the up-regulation of *PTPN6*, *VAV1* and *CD9* seen in *CCDC26* KOs was not rescued in o/x cells (Figure 4.11B).

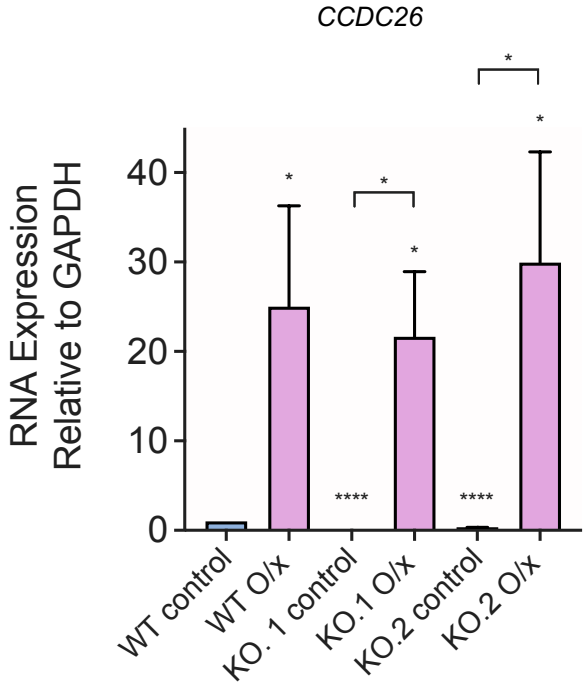


Figure 4.9 Exogenous *CCDC26* expression in WT and CCDC26 KO cells. qRT-PCRs show that *CCDC26* levels are significantly raised in cells stably transfected with spliced *CCDC26*-expressing plasmids, compared to empty plasmid controls. Compared to WT control cells, *CCDC26* expression is increased by approximately 25-fold in WT, KO.1 and KO.2 overexpression cells. Values represent the mean ± standard deviation (n=3). Statistical significance with respect to the WT control is indicated immediately above each bar. Statistical significance between KO controls and KO overexpression cells is also indicated; NS=P>0.05; *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 (unpaired, two-tailed *t* test).

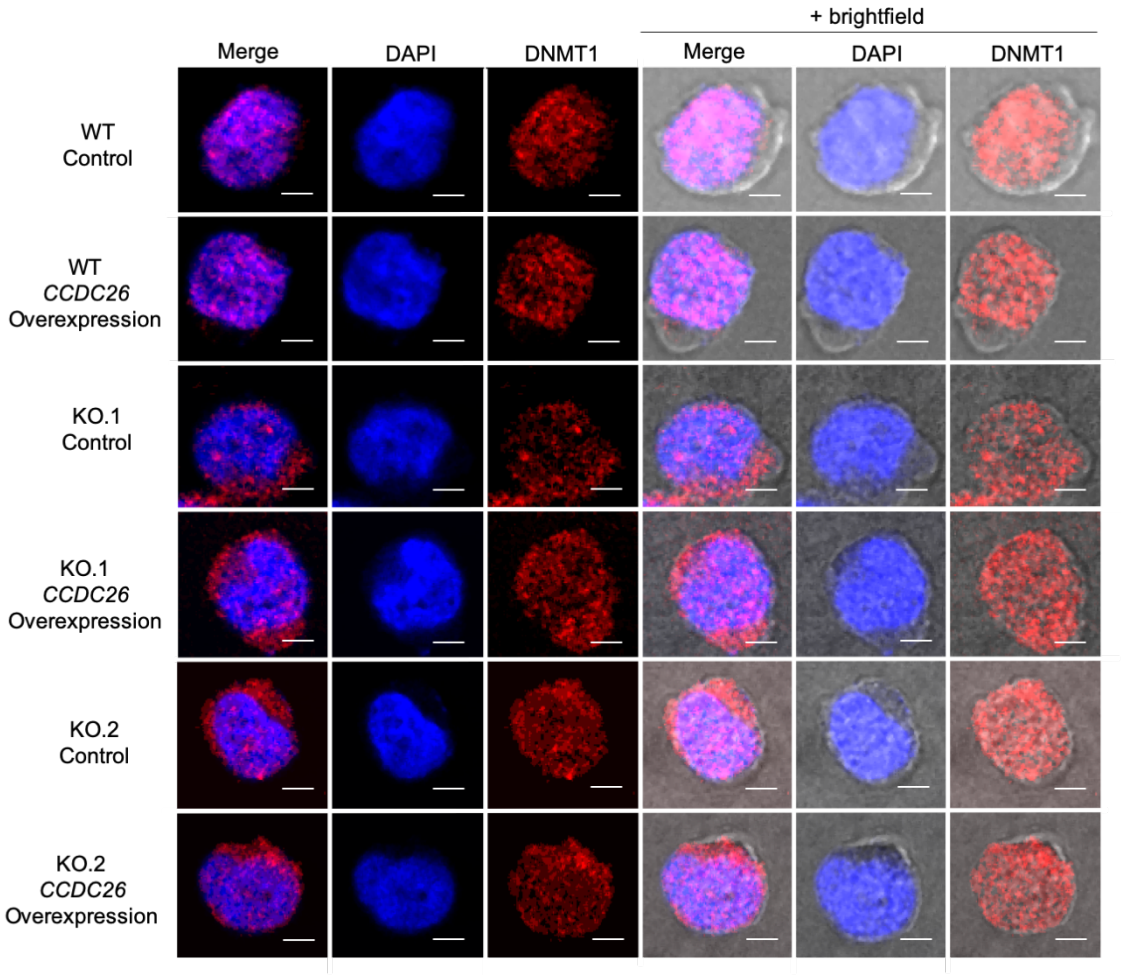


Figure 4.10 Exogenous *CCDC26* expression does not rescue DNMT1 localisation. Anti-DNMT1 (Red) immunofluorescence demonstrates that exogenous *CCDC26* expression in KO cell lines does not rescue DNMT1 localisation, which remains predominantly cytosolic in both *CCDC26* KO control and o/x cells. Nuclei were stained with DAPI (Blue). The outline of the cell membrane can be seen with the addition of the brightfield lens in the right-hand panels (scale bar = 5um)

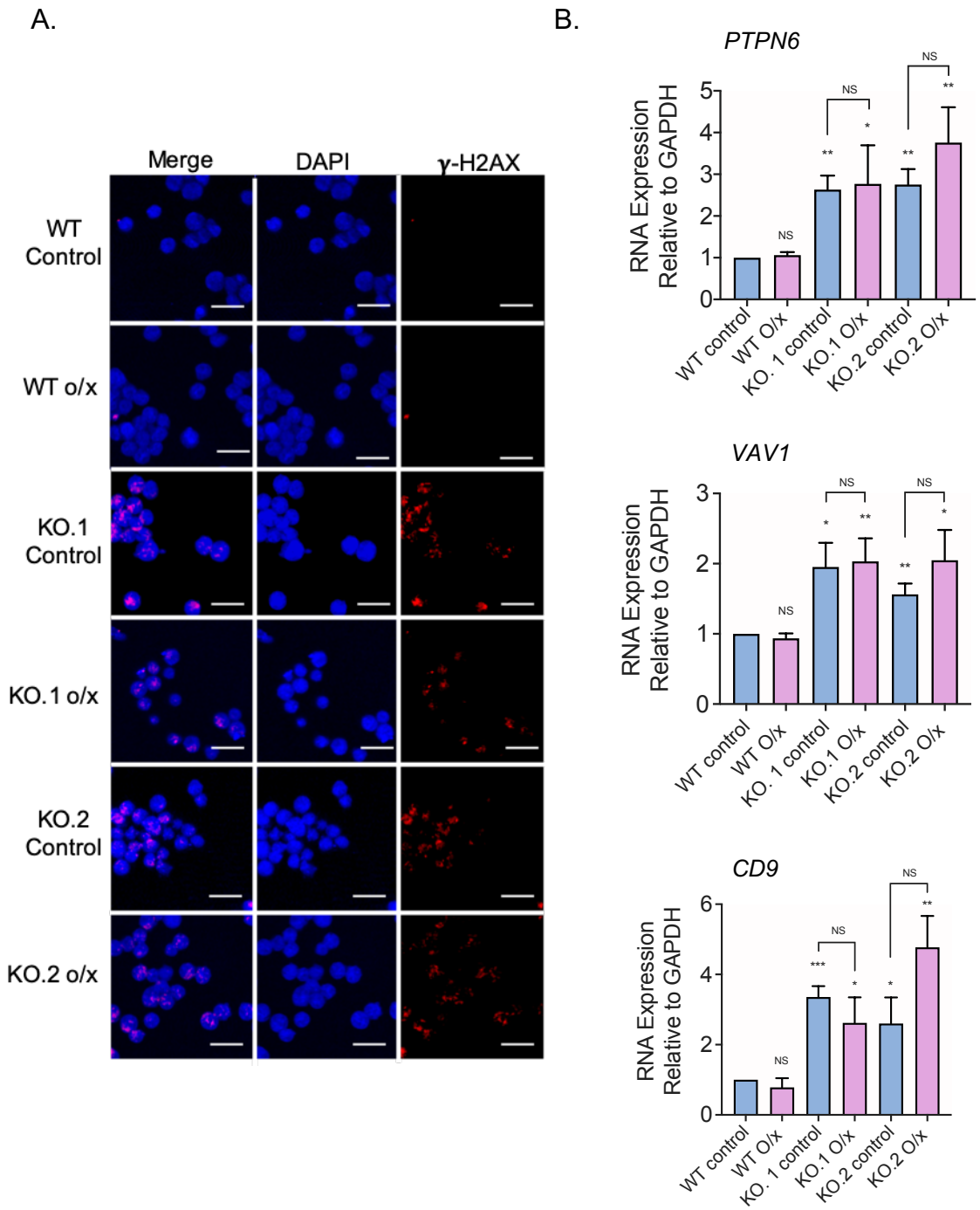


Figure 4.11 Exogenous *CCDC26* expression does not rescue *CCDC26* KO phenotypes A. Anti-γ-H2AX immunofluorescence (red) shows γ-H2AX foci in *CCDC26* KO o/x cells similar to KO control cells, indicating that DNA damage is not rescued by exogenously expressed *CCDC26* (scale bar = 25um) **B.** qRT-PCRs demonstrate that downregulation of *PTPN6*, *VAV1* and *CD9* gene expression is also not rescued with *CCDC26* overexpression. Values represent the mean ± standard deviation. Statistical significance with respect to the WT control is indicated immediately above each bar. Statistical significance between KO controls and KO overexpression cells is also indicated; NS=P>0.05; *P<0.05; **P<0.01; ***P<0.001 (unpaired, two-tailed *t* test).

4.3 Discussion

Here we have identified the likely cause of DNA hypomethylation in *CCDC26* KO cells. We have first shown that the total levels of DNMT3A, DNMT3B and DNMT1 are not affected by *CCDC26*, however, in its absence DNMT1 is mis-localised to the cytosol. Several lines of evidence, suggest that *CCDC26* directly interacts with DNMT1, alluding to a direct role for the lncRNA in regulating DNMT1 subcellular localisation. However, we were unable to rescue effects with exogenously expressed *CCDC26*.

In the past, non-coding RNAs have been shown to impact DNA methylation levels through transcriptional and post-transcriptional regulation of DNMT genes (Chen et al. 2015; Wang, Guo, Qian, et al. 2018; Cheng et al. 2018; Mohammad et al. 2010; Di Ruscio et al. 2013; Merry et al. 2015). *miR-124* and *miR-506* for example, have been shown to target DNMT3B and DNMT1 mRNAs respectively for degradation in breast and colorectal cancer. The overexpression of these miRNAs results in decreased DNMT protein levels and hence reduced DNA methylation (Chen et al. 2015; Wang, Guo, Qian, et al. 2018). Conversely, in hepatocellular carcinoma, lncRNA *HOTAIR* regulates DNMT1, 3A and 3B transcription by targeting polycomb repressor EZH2 to DNMT promoters (Cheng et al. 2018). However, unlike these previous publications, none of the DNMTs changed expression levels in *CCDC26* KOs, indicating that the mechanism behind DNA hypomethylation on removal of *CCDC26* is different.

LncRNA-mediated regulation of protein localisation within the cell is another well-established function. In some instances this can occur via a direct interaction, for example, lncRNA *TP53TG1* binds the transcription factor, YBX1, thereby preventing its nuclear trafficking (Diaz-Lagares et al. 2016). In many other instances, this can also occur via an indirect effect; an interaction between lncRNA *CRYBG3* and actin for example, is sufficient to prevent translocation of Myelin and Lymphocyte protein (MAL) into the nucleus (Pei et al. 2018). Similarly, *NFκB Interacting LncRNA* (*NKILA*) binds and prevents phosphorylation of the inhibitory IκB subunit. This blocks its degradation, which subsequently prevents the active p65 subunit of NFκB from re-localising from the cytosol to the nucleus (Liu et al. 2015). Protein trafficking by lncRNAs is also not restricted to just the nucleus and cytosol; a collection of lncRNAs in nerve cells, including *MEG3, MEG8, MEG9* and *RTL1-as*, have been shown to regulate translocation of AMPA glutamate receptors to the cell surface (Tan et al. 2017). Past publications show that DNMTs, especially DNMT1, interact with RNAs (Merry et al. 2015; Di Ruscio et al. 2013; Wang et al. 2015). We checked if DNMT1 directly or indirectly interacts with lincRNA *CCDC26*. Both approaches, data mining and experimental verification, show that *CCDC26* may interact with DNMT1. This has not been shown before, and could potentially provide a new mechanism by which *CCDC26* might function in regulating DNA methylation.

Remarkably, we also observed striking changes in DNMT1 subcellular localisation upon *CCDC26* KO. Both immunoblotting on cell fractions and immunofluorescence with anti-DNMT1, demonstrated a shift in the localisation of DNMT1, revealing a

prominent cytosolic DNMT1 fraction in KO cells. This is in contrast to the almost exclusively nuclear localisation of DNMT1 in WT cells. This, along with DNMT1 binding to *CCDC26* in RIP, indicates that *CCDC26* could play some role in direct sequestration of DNMT1 in the nucleus. DNMT3A and DNMT3B did not show any changes in localisation in KO cells. Furthermore, their reduced protein levels within cells, compared to DNMT1, indicate that the latter is responsible in large for DNA methylation in K562 cells. This suggests that the observed KO cell hypomethylation is likely attributed to DNMT1 specifically.

Despite the significant shift in DNMT1 localisation, DNMT1 is not re-localised to the cytosol in its entirety. Cells have previously been shown to grow for a long time with up to a 90% loss of DNMT1, whereas complete DNMT1 KOs are lethal; if the nuclei of our *CCDC26* KO cells were completely devoid of DNMT1, we would expect the cells to die (Chen et al. 2007; Takebayashi et al. 2007; Li, Bestor, and Jaenisch 1992). This can explain why *CCDC26* KO cells display apoptosis but still grow, albeit much slower compared to WT.

Together, these results indicate a novel role for *CCDC26*, as well as lncRNAs in general. Previous reports have shown that lncRNA-mediated regulation of DNMT1 primarily involves influencing methyltransferase activity at specific loci (Di Ruscio et al. 2013; Chalei et al. 2014; Gao et al. 2019; Guo et al. 2019; Mohammad et al. 2010), however, this is the first time a lncRNA has been shown to directly influence DNMT1 subcellular localisation and hence genome-wide methylation levels. It is yet to be identified, whether *CCDC26* directly sequesters DNMT1 in the nucleus similar

to lncRNA *TP53TG1* in the case of YBX1 (Diaz-Lagares et al. 2016), or whether its binding affects localisation-dependent modifications on DNMT1, similar to lncRNA *NKILA* and IκB (Liu et al. 2015).

Another observation of note, is the change in the nuclear pattern of DNMT1 remaining in the nuclei of KO cells. Nuclear distribution of DNMT1 can be described as somewhat diffuse in the WT cells, localising relatively evenly throughout the organelle. The DNMT1 that remains in the nuclei of the KO cells however, shows more punctated foci, indicating specific regions where DNMT1 has accumulated. At present it is not clear what these foci represent. We speculate that these could correspond to DNA repair sites. In Chapter 3 we showed increased γ-H2AX foci corresponding to increased numbers of DSBs and enhanced levels of DNA damage in the KO cells. Past studies have shown that proteins including DNMT1, accumulate at sites of DNA damage as part of the repair machinery. This allows DNMT1 to rapidly restore DNA methylation upon recovery of the DNA (Mortusewicz et al. 2005). Alternatively, these sites of DNMT1 accumulation could be due to aberrant DNMT1 activity. Previous studies have shown evidence of abnormal accumulations of DNMT1 as a result of disrupting the protein's 3D conformation (Misaki et al. 2016). It is also possible that DNMT1 is accumulating at some nuclear body, such as the nucleolus. Nevertheless, based on current data, it is not clear exactly what these foci correspond to.

As further confirmation of the results observed in this chapter, we performed a series of qRT-PCRs on WT and KO cells. We hypothesised that DNMT1 mis-

localisation should generate a phenotype similar to DNMT1 KD cells. Consequently, we selected several genes, as indicated by past literature, that have demonstrated increased expression in response to DNMT1 KD or inhibition; these included PTPN6, CDKN1A, CDKN2B, JUNB, CD9 and VAV1. PTPN6 encodes a tyrosine phosphatase protein known as SHP1, that is believed to act as a tumour suppressor. DNMT1 is known to interact with and methylate the PTPN6 promoter, and in particular has been shown to mediate aberrant repression of the gene in CML. Consequently, reductions in DNMT1 expression, correlate with rising PTPN6 expression (Li, Liu, et al. 2017; Wang et al. 2017). Cyclin-Dependent Kinase Inhibitors (CDKIs) are frequently repressed in cancer cells via hypermethylation of their promoter regions, resulting in aberrant cell cycle regulation. Consequently, expression of the CDKIs, CDKN1A and CKDN2B, is enhanced in response to DNMT1 depletion. CDKN2B appears to be selectively repressed in glioma and myeloid leukemia by aberrant methylation. Treatment of these cells with a DNMT1 inhibitor however, demonstrates re-expression of this gene (Herman et al. 1996). Furthermore, CDKN2B was found to be re-expressed following curcumin-mediated DNMT1 repression in acute myeloid leukemia cells (Yu et al. 2013). CDKN1A has similarly demonstrated activation following DNMT1 inhibition in both AML and CML cells amongst others (Milutinovic et al. 2004; Schmelz et al. 2005). JUNB is a transcription factor that typically demonstrates low expression levels in CML, but is upregulated upon demethylation by treatment with a DNMT inhibitor (Fiskus et al. 2009; Yang et al. 2003). CD9 is a member of the tetraspanin family of proteins involved in cell adhesion and migration, whose expression is activated in a selection of AML cell lines, including HL60 cells, upon DNMT inhibitor treatment (Kim 2012).

Similarly, methylation-mediated repression of the guanine nucleotide exchange factor, VAV1, has also been observed in a variety of cell lines (Ilan and Katzav 2012; Fernandez-Zapico et al. 2005).

To further argue our hypothesis, we also selected a gene whose expression is down-regulated upon DNA hypomethylation. IGF-1 is involved in an important signalling pathway, which involves binding a receptor tyrosine kinase (IGF1-receptor). This pathway is involved in regulation of cell proliferation and growth, as well as inhibition of apoptosis (Parrizas, Saltiel, and LeRoith 1997). Previous studies have shown that IGF-1 is down-regulated by the histone methyltransferase, Retinoblastoma-interacting zinc-finger protein 1 (RIZ1). Typically, in K562 cells RIZ1 is silenced by methylation of its promoter region, however, when DNA hypomethylation is induced, RIZ1 is expressed, and subsequently represses IGF-1 (Pastural et al. 2007).

In accordance with our hypothesis, expression patterns of the selected genes in the KO cells compared to WT, resembled the effects observed in past reports when DNMT1 is reduced. Of the 7 genes analysed, 6 were significantly differentially expressed in both KO cells as predicted. The exception was JUNB, whose expression was still upregulated in KO cells, despite not being statistically significant. These results give further support to our previous observations and hypothesis, that *CCDC26* KO cells behave similarly to DNMT1 KD cells, as a result of DNMT1 re-localisation. To distinguish this, it will be extremely interesting to

assess the methylation status of the promoter regions of these genes by including methylation specific PCR (MSP) or bisulphite sequencing in our future experiments. To close this chapter, attempts to rescue DNMT1 re-localisation as well as other effects observed in the KO cells were made. However, despite successfully overexpressing *CCDC26* exogenously in cell lines, we were unable to reverse the KO cell phenotypes. DNMT1 persisted in the cytosol of KO o/x cells, and DNA damage and gene expression patterns remained the same. A possible reason for this might be that the overexpression vector used was designed to express spliced *CCDC26*. Given that our DNMT1 RIP data indicates it is the spliced form of *CCDC26* that may interact with DNMT1, it is likely that the spliced form of *CCDC26* is involved in DNMT1 regulation. Nevertheless, we cannot completely rule out the possibility that the unspliced form, or even an as yet undiscovered isoform plays a role in this process. It will be interesting to perform further rescue experiments with alternative *CCDC26* isoforms, both spliced and unspliced. We must also consider the possibility that the rescue has not worked because the observed phenotype is due to some unexpected off-target effect. In the future, it will be important to rule this out by potentially sequencing the genome of the KO cells. In addition, it will also be interesting to try to rescue nuclear DNMT1 in the KO cells to determine whether exogenously expressed DNMT1 rescues the observed DNA damage and apoptosis.

The possibility that a single lncRNA can have such a profound effect on an essential enzyme such as DNMT1, is astounding. However, a considerable amount of further work is needed to fully comprehend this novel mechanism of regulation. So far, our

model suggests that in the presence of *CCDC26* in WT K562 cells, DNMT1 is primarily nuclear, where it is able to carry out its function in methylating DNA. In the absence of *CCDC26* however, DNMT1 is largely localised within the cytosol, resulting in global DNA hypomethylation (Figure 4.12). It is now vital to ascertain the mechanism by which this is occurring. Currently, we do not know in the absence of *CCDC26* if DNMT1 export to the cytosol is enhanced, if nuclear import of DNMT1 is affected or there is increased cytosolic accumulation DNMT1 due to its decreased proteosomal degradation. Understanding exactly how DNMT1 is being re-localised will help answer this question, as well as establishing the order of cellular events following loss of *CCDC26*.

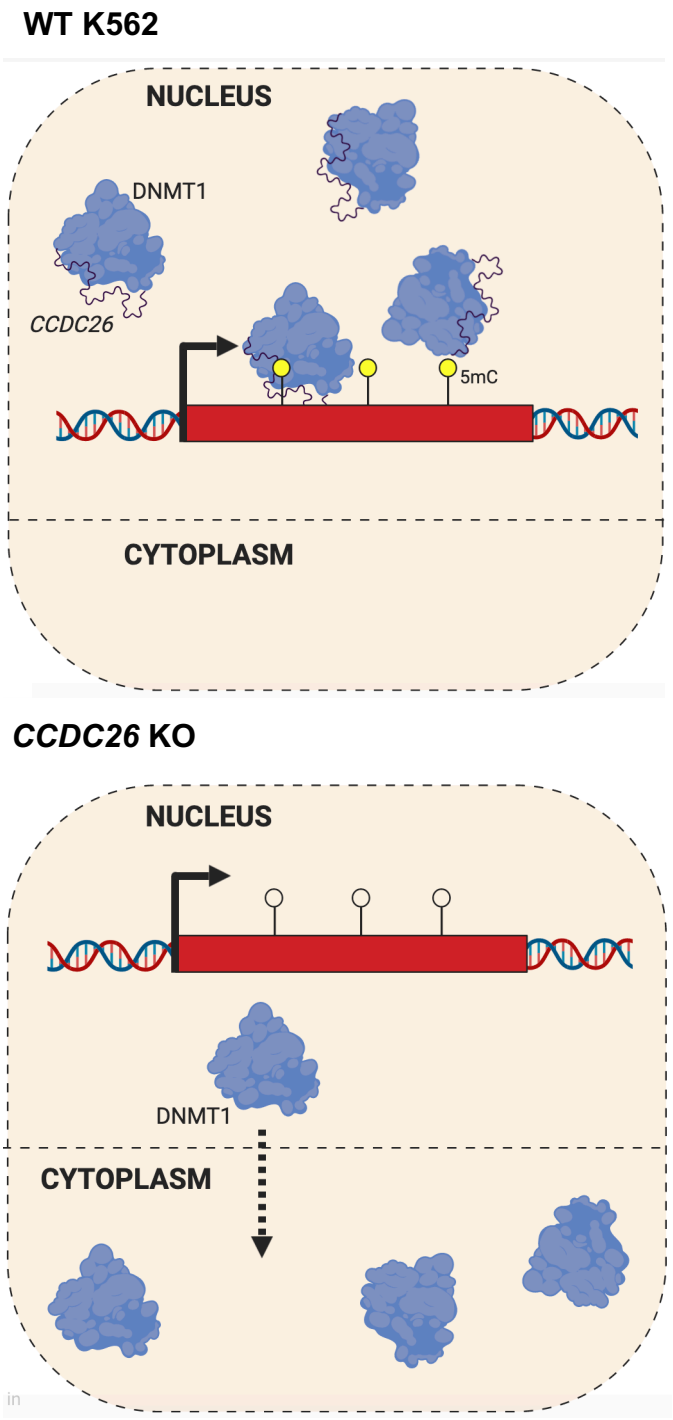


Figure 4.12 Model for *CCDC26*-mediated DNMT1 regulation. In WT K562 cells, DNMT1 interacts with *CCDC26*. DNMT1 is almost exclusively localised in the nucleus where it maintains DNA methylation patterns as cells replicate. In the absence of *CCDC26*, DNMT1 is re-localised to the cytoplasm and cells become hypomethylated.

CHAPTER 5

UNDERSTANDING HOW *CCDC26* REGULATES SUBCELLULAR LOCALISATION OF DNMT1

5.1 Introduction

In the previous chapter, we demonstrated that whilst DNMT levels are unchanged in *CCDC26* KO cells, the majority of DNMT1 is re-localised from the nucleus to the cytoplasm. This is a striking observation, as reports of cytosolic DNMT1 are rare. Arguably the best studied of these instances is in preimplantation during early development. Most developmental stages of the early embryo, up to and including the blastocyst stage, demonstrate cytosolic DNMT1 sequestration. This is required for demethylation of the embryonic genome, creating a 'blank canvas', upon which lineage specific methylation patterns are established. (Cardoso and Leonhardt 1999). However, in instances other than during embryonic development, cytosolic localisation of DNMT1 tends to be aberrant; for example, it has been associated with several neurological disorders including hereditary sensory and autonomic neuropathy type 1E (HSAN1E) (Baets et al. 2015), Alzheimer's disease (Mastroeni et al. 2013) and Parkinson's disease (Desplats et al. 2011), as well as cancer tumorigenesis (Hodge et al. 2007; Arzenani et al. 2011). Interestingly, these examples do not share a common mechanism for DNMT1 re-localisation; DNMT1 becomes cytosolic by a range of different means, including as a result of changes to post-translational modifications (Hodge et al. 2007), HDAC inhibition (Arzenani et al. 2011), mutations within the RFTS domain (Baets et al. 2015) and disruption to nucleo-cytoplasmic transport systems across the nuclear membrane (Mastroeni et al. 2013). In many of these instances however, DNMT1 appears to be affected at

the N-terminal domain, providing a common link (Hodge et al. 2007; Baets et al. 2015). The re-localisation of DNMT1 to the cytosol, observed in Chapter 4, likely explains the global DNA hypomethylation we observed in Chapter 3. We also showed that DNMT1 interacts with *CCDC26* in WT cells. It is now important to establish how *CCDC26* lncRNA could have such a profound effect on the localisation of such an important protein. Here, we address this question by studying DNMT1 displacement in KO cells.

5.2 Results

5.2.1 DNA Damage is a consequence and not a cause of DNMT1 re-localisation

We sought to establish the sequence of events that result in DNMT1 re-localising to the cytosol of *CCDC26* KO cells. In Chapters 3 we also observed that there are increased levels of DNA damage and apoptosis in the absence of *CCDC26*. In order to really understand the functional mechanism of *CCDC26*, it is important to examine whether cytosolic localisation of DNMT1 is a cause or consequence of DNA damage and apoptosis.

We began by determining whether the observed DNMT1 re-localisation and consequently DNA hypomethylation can trigger DNA damage. Since DNMT1-mediated DNA methylation occurs exclusively in the nucleus, we reasoned that the KO cells should behave similarly to DNMT1 KD cells. Consequently, to mimic the effects of DNMT1 cytosolic re-localisation, K562 cells were transiently transfected with 60pmol DNMT1 siRNA for 24hrs. The KD was confirmed by measuring DNMT1

mRNA and protein levels by qRT-PCR and western blotting, which showed a KD at both the RNA and protein level (Figure 5.1). 24hrs after cells were transfected, DNMT1 mRNA was reduced by approximately 50% when measured relative to the housekeeping gene, GAPDH (Figure 5.1A), and more importantly, DNMT1 protein levels showed a 70-80% KD (Figure 5.1B). RNA and protein were also extracted and subjected to the same analyses 48hrs post-transfection, however, at this point, DNMT1 protein levels had almost returned to normal and demonstrated levels of >90% than that of WT cells (Appendix III, Figure 1).

A fraction of DNMT1 KD cells were fixed in methanol and acetic acid 24hrs after transfection and subjected to **γ**-H2AX immunofluorescence to measure DNA damage (Figure 5.2A). There appeared to be no significant difference between DNMT1 KD cells and cells transfected with a negative control siRNA. No prominent **γ**-H2AX foci could be seen in either cell type, similar to WT K562 cells. This suggested that reduced levels of nuclear DNMT1 does not result in DNA damage.

However, when we also measured 5mC immunofluorescence, we found that levels were also not significantly different between control and KD cells, and resembled WT K562 cells (Figure 5.2B). Furthermore, genes previously shown to be impacted by DNMT1 and methylation (Chapter 4.2.4), and shown to be differentially expressed in *CCDC26* KO cells (Figure 4.8), showed no significant differences in their expression in DNMT1 KD cells (Figure 5.2C). It is possible that 24hrs is not long enough to see significant changes in global methylation levels and hence gene expression. We have seen that 24hrs after siRNA transfection, there is a significant

decrease in DNMT1 protein levels, however, after 48hrs, levels have almost returned to normal. It appears there is a very short window within which DNMT1 protein is significantly depleted, and given that K562 cells double every 24hrs, it is likely that this window is too short to see global DNA hypomethylation and subsequent downstream effects.

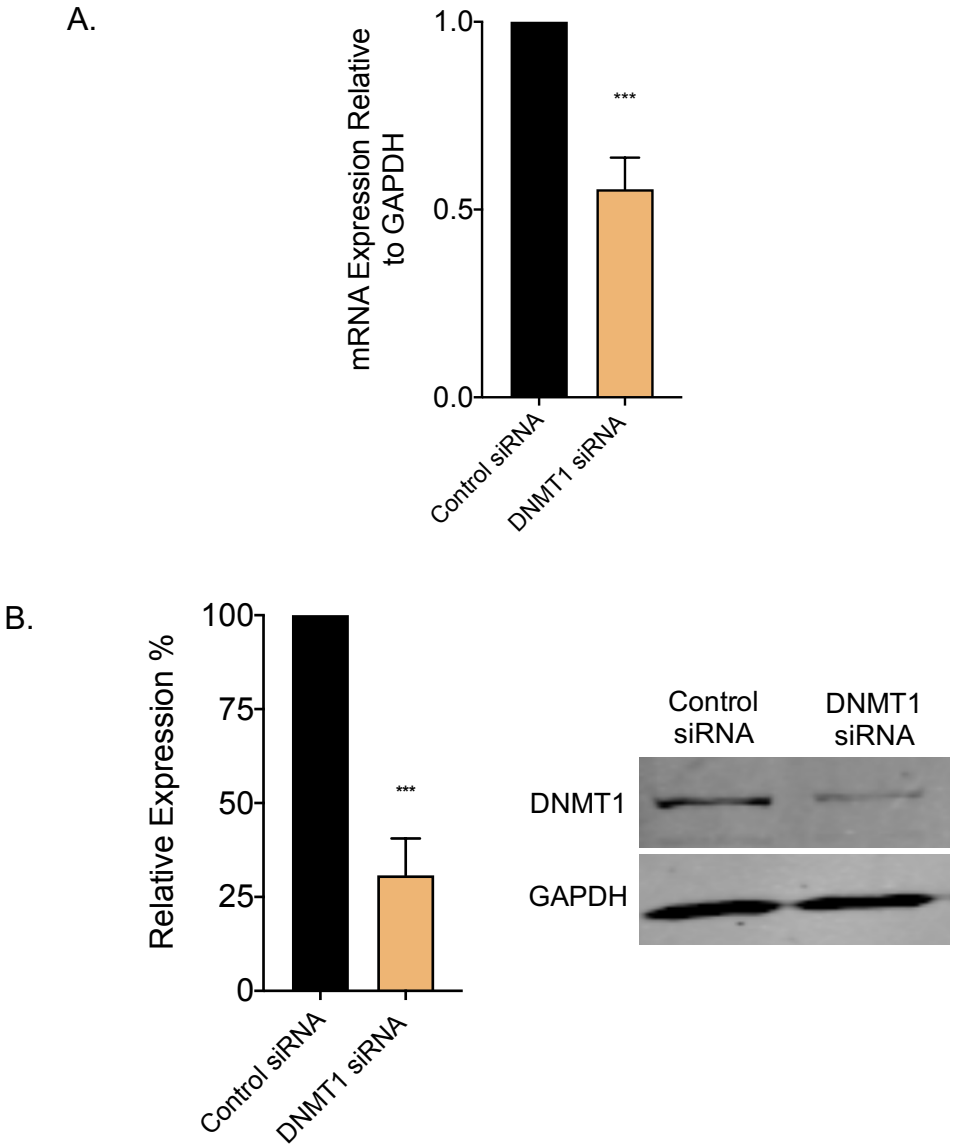


Figure 5.1 DNMT1 siRNA KD in K562 cells. WT K562 cells were transiently transfected with 60pmol DNMT1 siRNA for 24hrs, after which total RNA and protein were extracted. **A.** qRT-PCRs demonstrated an approximately 50% reduction in DNMT1 mRNA in cells transfected with DNMT1 siRNA compared to a control. **B.** Immunoblotting demonstrated an almost 80% reduction in DNMT1 protein in cells transfected with DNMT1 siRNA compared to control. mRNA and protein were measured relative to the housekeeping gene, GAPDH. Values represent the mean ± standard deviation (n=3). * P<0.05; **P<0.01; ***P<-0.001 (unpaired, two-tailed *t* test).

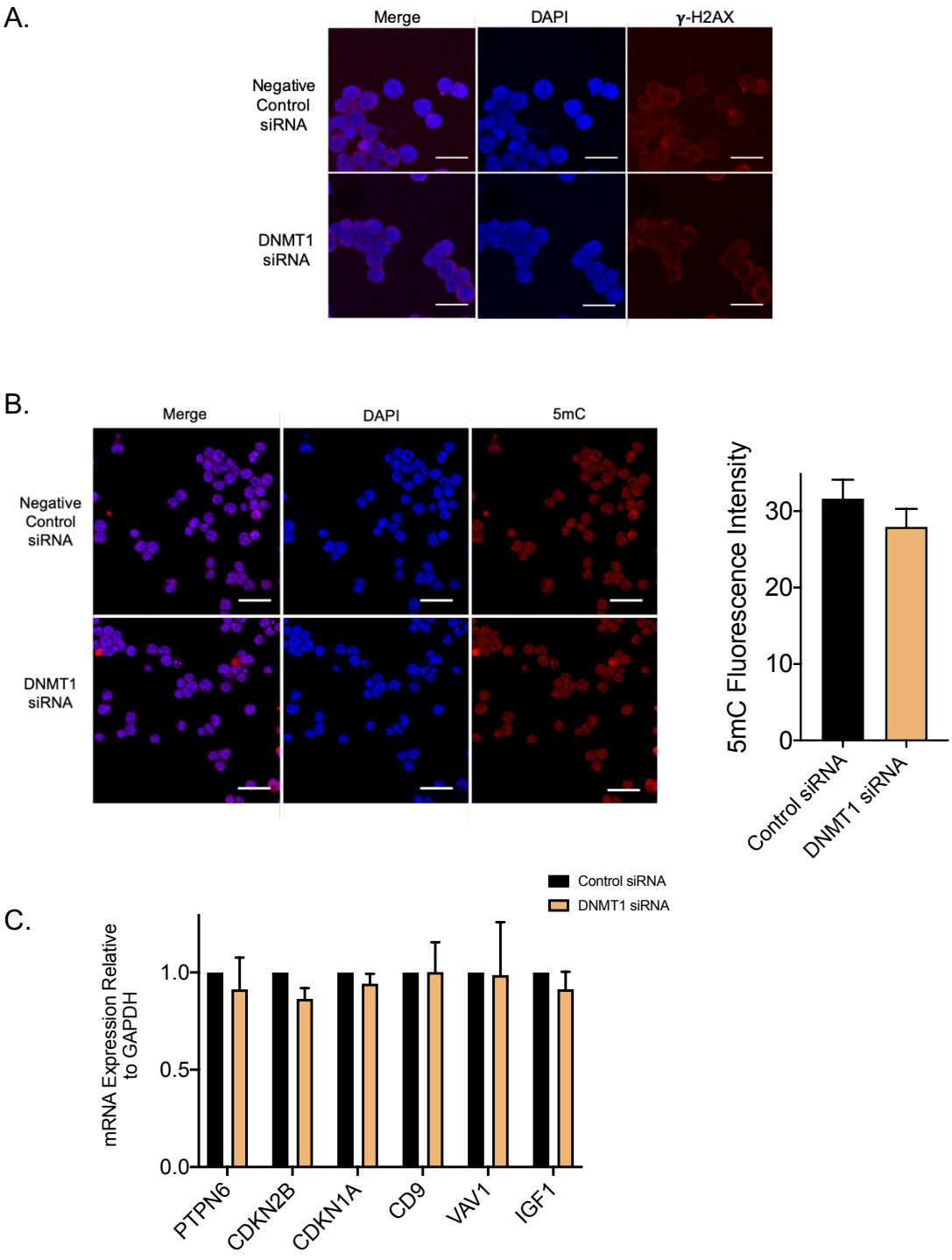


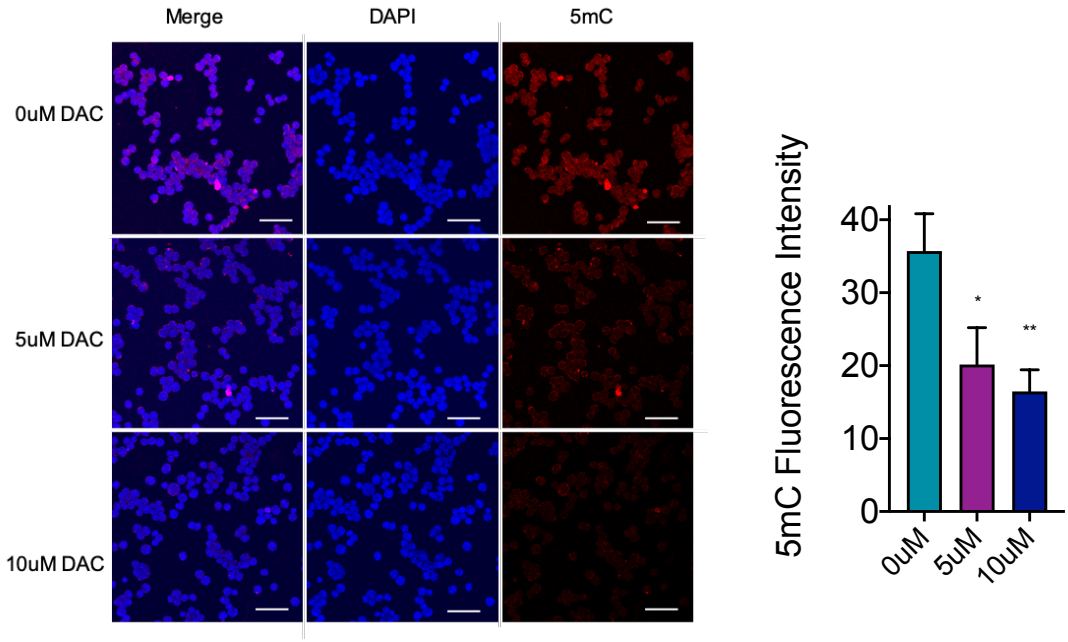
Figure 5.2 24-hour DNMT1 siRNA KD does not significantly affect phenotype. DNMT1 siRNA KD cells demonstrated **A.** no significant change in DNA damage, as indicated by **γ**-H2AX (red) immunofluorescence (scale bar = 25um), **B.** no significant change in global 5mC (red) immunofluorescence levels (scale bar = 50um) and **C.** no changes in expression levels of genes that are differentially expressed in *CCDC26* KO cells. Values represent the mean ± standard deviation (n=3) (unpaired, two-tailed *t* test). Nuclei were stained with DAPI (blue).

Since DNMT1 siRNA KD appeared insufficient to confirm the effect of DNMT1 depletion, we began contemplating other ways in which to deplete nuclear DNMT1, which would also allow time for a reduction in DNA methylation and induction of downstream effects. In other words, DNMT1 depletion needed to be maintained for a longer period of time. Consequently, we adopted a different approach and chose to treat cells with a DNMT1 inhibitor. 5-aza-2'-deoxycytidine, also known as, Decitabine (DAC), is a cytosine chemical analogue, and is incorporated into DNA as cells replicate. DNMT1 binds DAC as it would cytosine, but the Nitrogen atom in place of the usual Carbon-5 of the cytosine ring, causes DNMT1 to become covalently trapped to the DNA, thereby rendering it non-functional (Stresemann and Lyko 2008).

K562 WT cells were grown in the presence of either 0uM, 5uM or 10uM DAC for 48hrs. To ensure the inhibitor was working, treated cells were fixed and 5mC immunofluorescence was performed. Treatment with both 5uM and 10uM DAC concentrations significantly reduced global 5mC levels (Figure 5.3A). DAC has also been reported to cause reductions of DNMT1 levels, as it can be targeted for degradation after being mobilised (Ghoshal et al. 2018; Patel et al. 2010; Ghoshal et al. 2005). Western blotting for DNMT1 too demonstrated decreased protein levels, although not significantly so (Figure 5.3B). Combined with the 5mC immunofluorescence data however, we can conclude that the inhibitor treatment was successful. Measurements were also performed 24hrs after inhibitor treatment, however, 5mC levels were not significantly altered at this time point (Appendix III, Figure 2).

Following confirmation of global hypomethylation, we next examined whether this elicited similar effects to *CCDC26* KO. **γ**-H2AX immunofluorescence resembled *CCDC26* KO cells and indicated a significant increase in DNA damage in cells treated with both 5uM and 10uM DAC for 48hrs (Figure 5.4A), whilst qRT-PCRs for genes identified in Figure 4.8 showed similar patterns of expression to *CCDC26* KO cells. Significantly, the changes in gene expression were more pronounced in cells treated with 10uM DAC as compared 5uM DAC (Figure 5.4B). This confirmed that DNA hypomethylation leads to DNA damage.

A.



B.

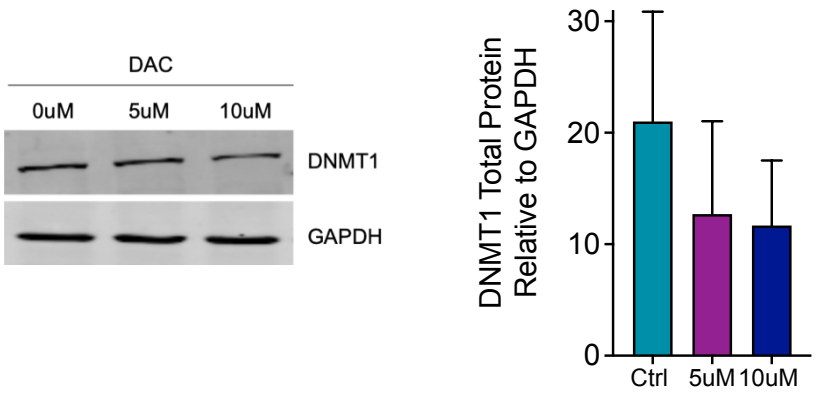
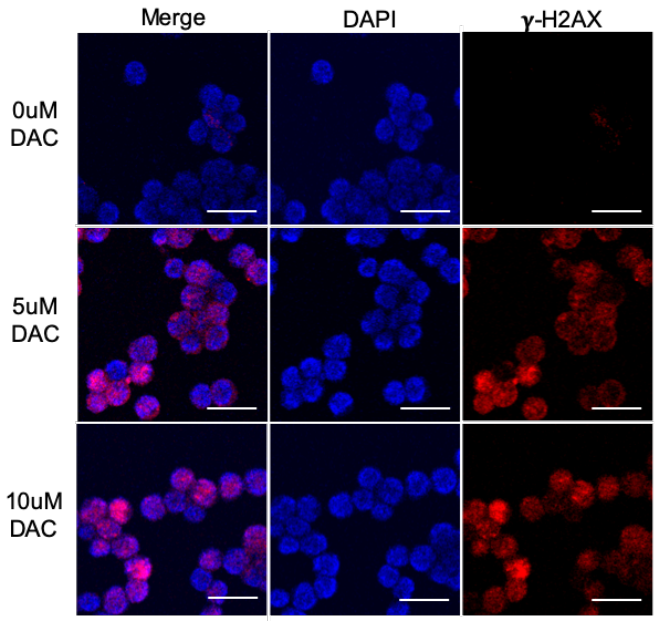


Figure 5.3. Confirmation of DAC-induced DNMT1 inhibition. A. Global 5mC immunofluorescence (red) levels are significantly reduced upon 48hrs treatment of K562 WT cells with both 5uM and 10uM DAC. Nuclei were stained with DAPI (blue) (Scale bar = 50um). **B.** DNMT1 total protein levels are also slightly reduced upon treatment with 5uM and 10uM DAC. Values represent the mean ± standard deviation (n=3) (unpaired, two-tailed *t* test).

A.



B.

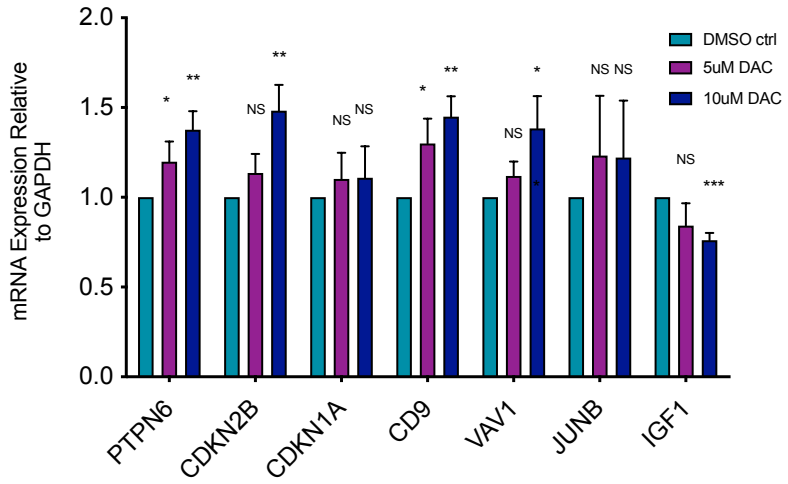


Figure 5.4. DAC treated cells demonstrate a phenotype similar to *CCDC26* KO cells. A. K562 cells show increased levels of DNA damage, as indicated by γ-H2AX immunofluorescence (red), when treated with 5uM or 10uM DAC for 48hrs. Nuclei were stained with DAPI (blue) (scale bar = 25um). **B** DAC-treated cells show similar changes in gene expression patterns to *CCDC26* KO cells. mRNA levels were measured relative to the housekeeping gene, GAPDH. Values represent the mean ± standard deviation (n=3). NS = Not significant, * P<0.05; **P<0.01; ***P<-0.001 (unpaired, two-tailed *t* test).

After confirming DNA damage is a consequence of DNA hypomethylation, we next wanted to confirm that DNMT1 mis-localisation is a result of *CCDC26* KO and not a consequence of DNA damage. It was critical to establish whether or not this type of DNMT1 movement is a general response to DNA damage and apoptosis. To investigate this, DNA damage was induced in WT cells using cisplatin. Cisplatin is a platinum-based drug that forms bonds with, and ultimately crosslinks, bases within and between DNA strands. This can distort the double helix, interfere with both DNA replication and transcription and consequently induce DNA damage and apoptosis (Goodsell 2006). To begin, the amount of cisplatin and treatment time required to induce DNA damage but prior to complete cell death, was optimised. Microscopic observations demonstrated that after 24hrs of cisplatin treatment, cells treated with 5uM or 10uM of the drug were still viable. However, cell viability significantly decreased when treated with 20uM cisplatin (Figure 5.5A). Therefore, further investigation was carried on in the cells treated with 5uM and 10uM cisplatin. DNA damage was also confirmed in the cells treated with 5uM and 10uM cisplatin by monitoring **γ**-H2AX foci using immunofluorescence (Figure 5.5B).

To assess DNMT1 localisation in DNA damage induced cells, DNMT1 immunofluorescence was performed on cisplatin-treated cells, and demonstrated no significant differences between 0uM control cells and the drug-treated cells. Similar to WT K562 cells, in cisplatin-treated cells, DNMT1 appeared primarily nuclear, demonstrating a diffused pattern of distribution throughout (Figure 5.6A). Accordingly, there was also no substantial change in the levels of 5mC immunofluorescence between control and cisplatin-treated cells (Figure 5.6B). This

suggests that DNMT1 re-localisation is not a general consequence of DNA damage. As further confirmation, the levels of genes that are up- or down-regulated in response to *CCDC26* KO and DNMT inhibition (Chapter 4.2.4, Figure 4.8) were also tested. Some genes, such as IGF1, demonstrated similar changes in expression patterns in response to cisplatin, as they did in response to DNMT inhibition and *CCDC26* KO. However, in the majority of instances, gene expression was either not significantly altered, or demonstrated an opposite change in expression. Ultimately, we did not see identical patterns of gene expression in cisplatin-treated cells (Figure 5.6C). This suggests that these responsive expression changes are specific to DNMT1 re-localisation and DNA hypomethylation.

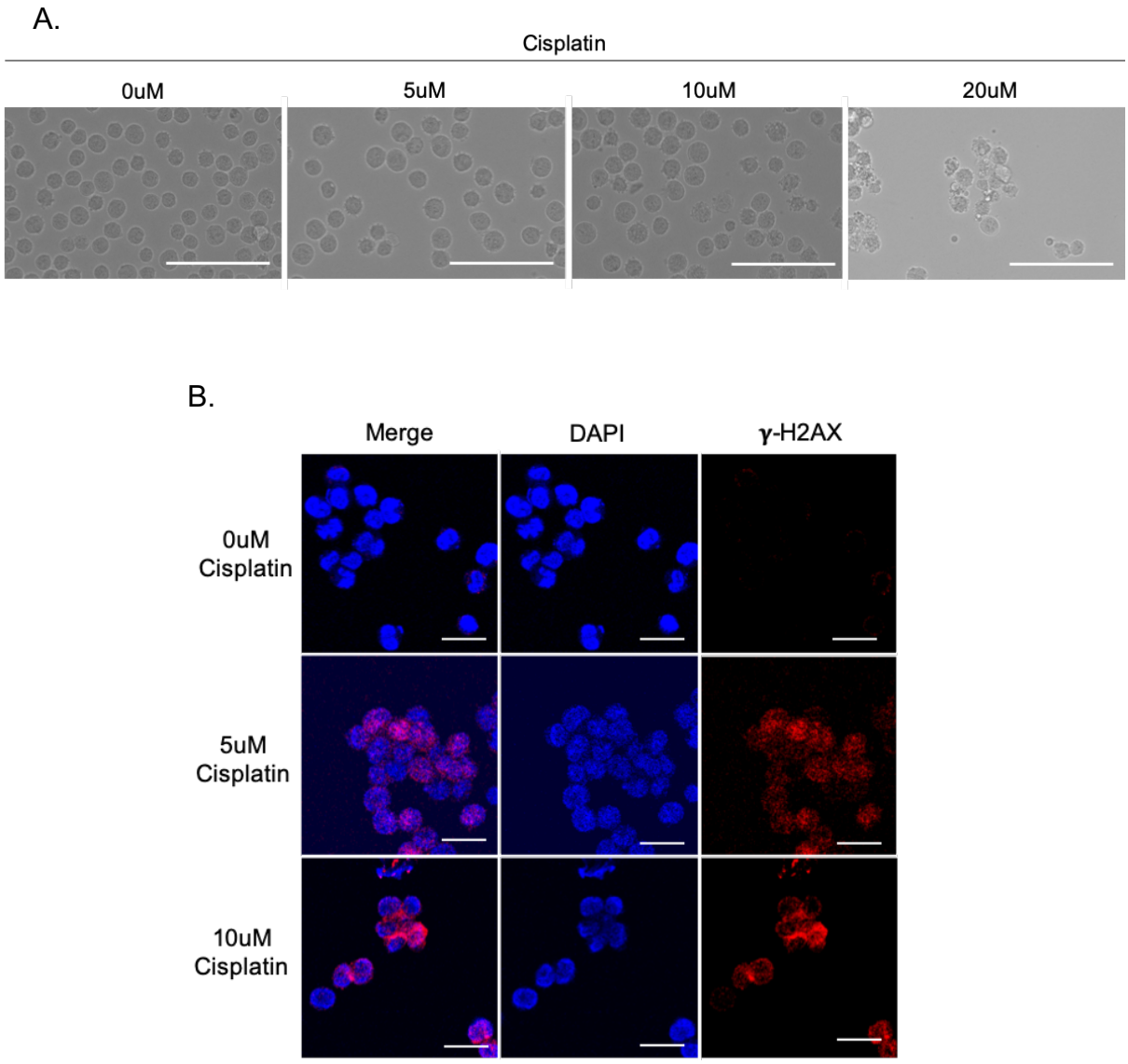
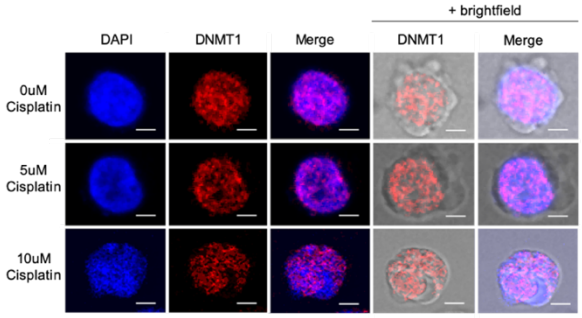
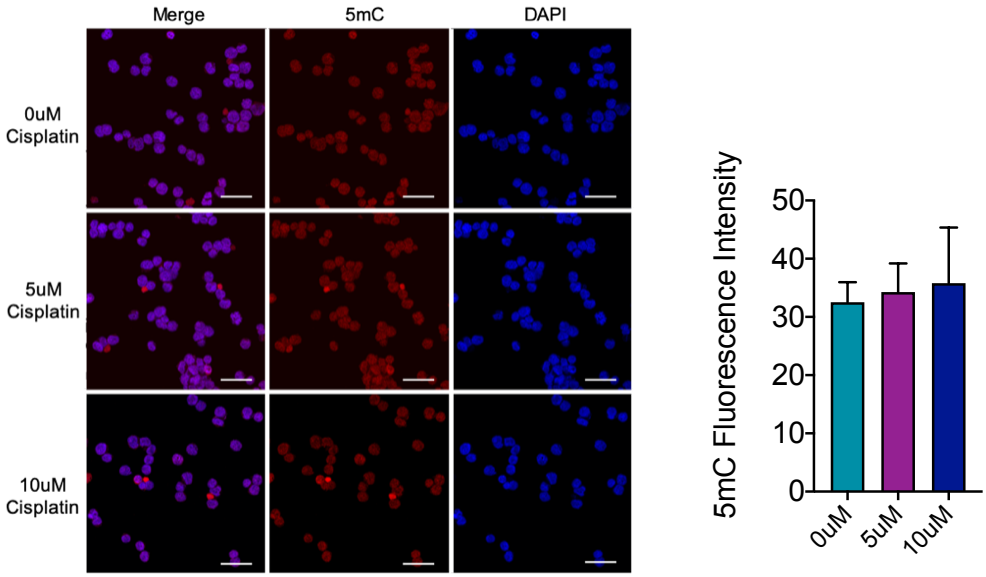


Figure 5.5 Confirmation of Cisplatin-induced DNA Damage A. Cells look increasingly distressed with increasing concentrations of Cisplatin as seen using brightfield microscopy (scale bar = 100um) **B.** γ-H2AX immunofluorescence (red) shows increased numbers of DNA damage foci. Nuclei were stained with DAPI (blue) (scale bar = 25um).

A.



B.



C.

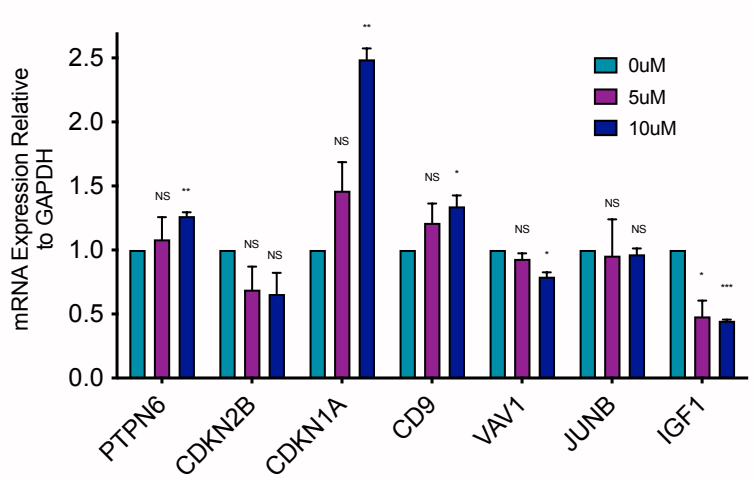


Figure 5.6 Cisplatin-induced DNA damage does not affect DNMT1 localisation **A.** DNMT1 immunofluorescence shows nuclear localisation in both control and cisplatin-treated cells (scale bar = 5um). **B.** 5mC immunofluorescence shows no significant change in global levels between control and cisplatin-treated cells (scale bar = 50um). **C.** Genes expression patterns are not identical to expression changes observed in *CCDC26* KO and DAC-treated cells. mRNA levels were measured relative to the housekeeping gene, GAPDH. Values represent the mean ± standard deviation (n=3). NS = Not significant, * P<0.05; **P<0.01; ***P<-0.001 (unpaired, two-tailed *t* test). Nuclei were stained with DAPI (blue).

5.2.2 Is DNMT1 re-localisation related to HDAC Inhibition?

HDACs are a family of histone deacetylase proteins, that primarily catalyse the removal of acetyl groups from histones in chromatin regulation (Seto and Yoshida 2014). Additionally, in recent years evidence has demonstrated that HDACs are also important regulators of other non-histone proteins involved in important biological processes, such as p53, and is subsequently implicated in tumorigenesis (Insinga et al. 2004). A previous publication (Arzenani et al. 2011) showed that DNMT1 localisation changes upon inhibition of HDACs with Trichostatin A (TSA). They found that upon treatment with TSA, Hep3B cells show increased cytosolic DNMT1 and reduced nuclear DNMT1 levels in a dose-dependent manner. Interestingly, TSA treated cells also showed increased levels of apoptosis (Arzenani et al. 2011). Given the similarities with our observations, we decided to investigate whether the DNMT1 re-localisation we detect in *CCDC26* KO cells is driven by a similar mechanism of HDAC reduction.

Arzenani et al. (2011) showed that TSA treatment had the most prominent effect on HDAC2, which demonstrated increased degradation with increasing inhibitor concentration. Consequently, we have focussed our attentions here on this particular HDAC family member. We first analysed HDAC2 protein levels and localisation by western blotting and immunofluorescence. Total HDAC2 protein levels showed no significant changes between WT and *CCDC26* KO cells (Figure 5.7A) and no significant changes in nuclear and cytosolic distribution (Figure 5.7B and C). Immunofluorescence experiments confirmed that there are no discernible differences in HDAC2 nuclear distribution in WT and *CCDC26* KO cells (Figure

5.7C). We also performed co-immunofluorescence of DNMT1 and HDAC2 in WT and KO cells, to check HDAC2 distribution in relation to DNMT1, and determine whether they are co-localised or in close proximity (Figure 5.8). Interestingly, the punctated, nuclear DNMT1 foci we see in KO cells, overlaps strongly with HDAC2 signals. HDAC2 is known to bind DNMT1 at replication foci (Rountree, Bachman, and Baylin 2000), and also accumulates at DNA damage sites along with DNMT1 as part of the DNA damage response (Miller et al. 2010). As mentioned in Chapter 4, it is possible that these foci represent more pronounced damage spots or replication sites. It was not possible to perform **γ**-H2AX and DNMT1 co-immunofluorescence in KO cells to confirm whether the punctated foci are damage spots however, as our antibodies were both derived from mouse. This will require further investigation in the future.

Since no major differences in HDAC2 levels could be observed between our WT and KO cells, before continuing examination of other HDACs, we decided to investigate whether the observations made by Arzenani et al. (2011) could be replicated in our cell line of interest. To do this, we induced HDAC inhibition in WT K562 cells to determine whether a similar change in DNMT1 localisation occurred. This could tell us whether the increase in cytosolic DNMT1 observed in Hep3B cells upon TSA treatment is a general consequence of HDAC inhibition, or whether this might be a cell-specific effect.

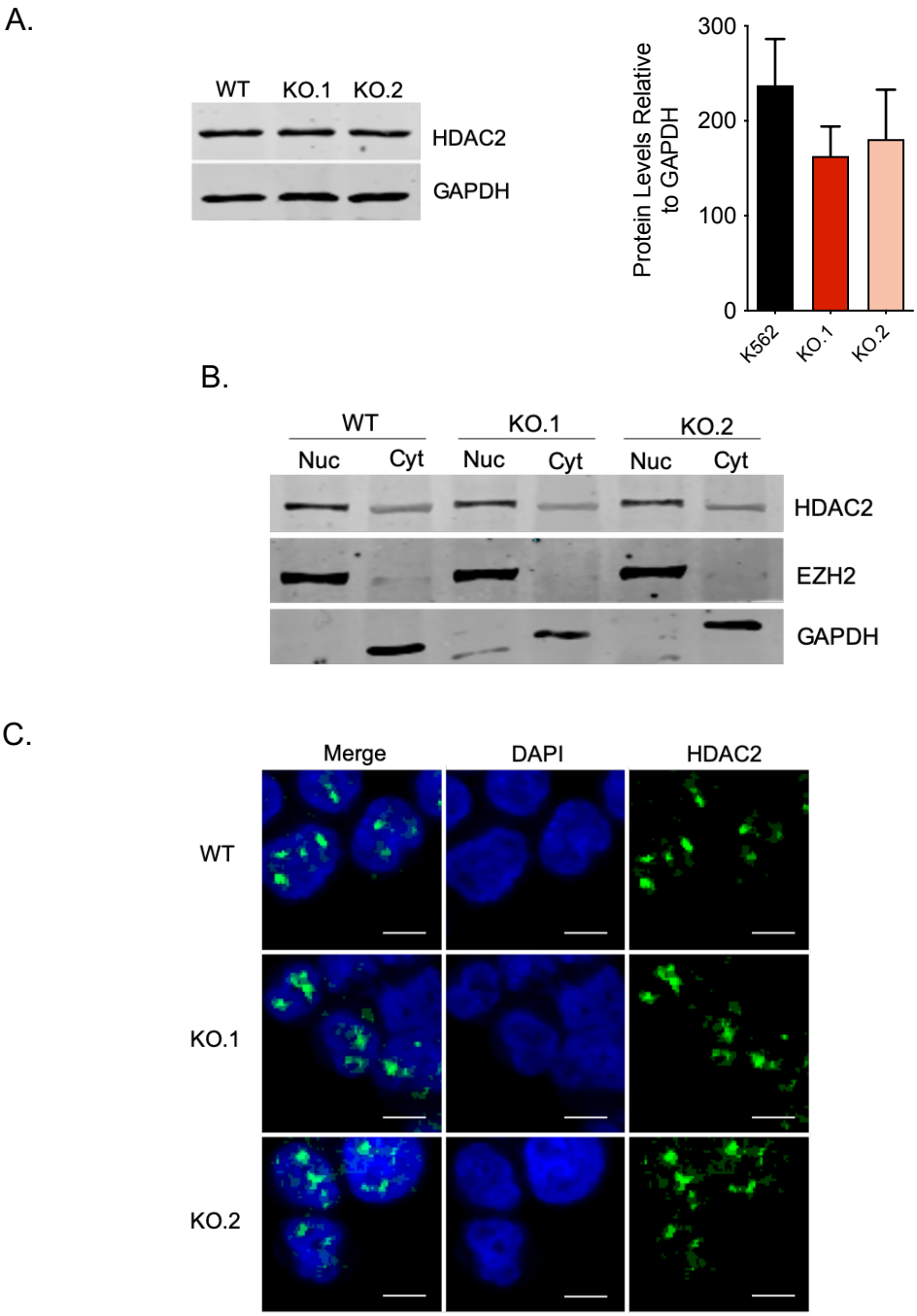


Figure 5.7 HDAC2 protein levels and localisation are not significantly different in *CCDC26* KO cells compared to WT. A. Western blotting indicates that there is no significant difference in total protein levels of HDAC2 and **B.** no change in subcellular localisation. Western blotting on nuclear and cytosolic fractions show that HDAC2 is mostly nuclear, with a small fraction present in the cytosol in both WT and CCDC26 KO cells. EZH2 and GAPDH were used as nuclear and cytosolic markers respectively. **C.** HDAC2 immunofluorescence (green) demonstrated a primarily nuclear localisation for HDAC2 in both WT and KO cells, with similar patterns of staining. Nuclei were stained with DAPI (blue) (scale bar = 10um).

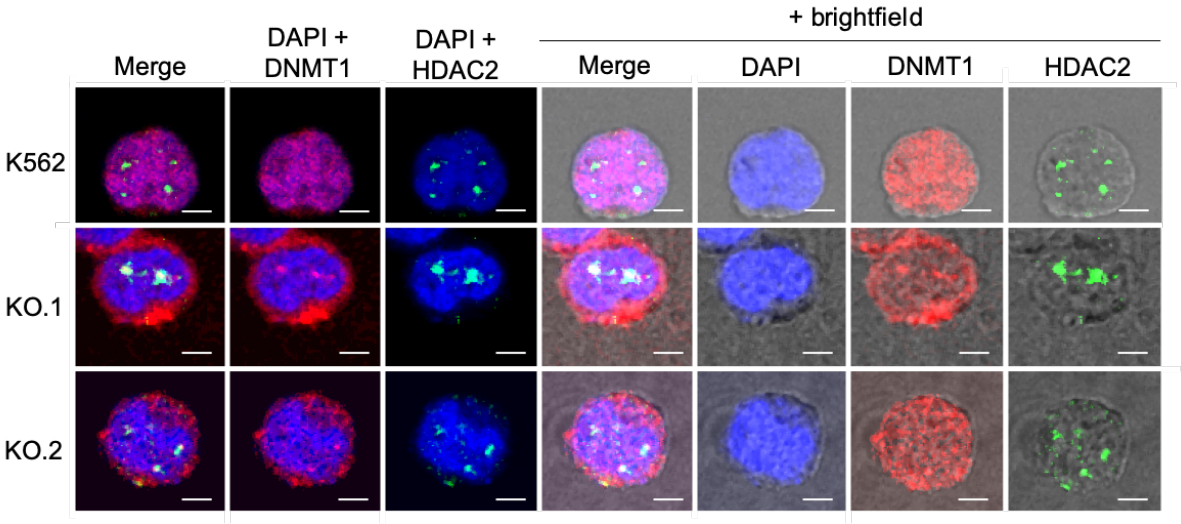


Figure 5.8 DNMT1-HDAC2 co-immunofluorescence. The punctate, nuclear DNMT1 (red) foci remaining in CCDC26 KO cells overlaps with HDAC2 (green) foci. Nuclei were stained with DAPI (blue). The outline of the cell membrane can be seen with the addition of the brightfield lens in the right-hand panels (scale bar = 5um).

Due to a lack of availability of TSA at the time of this project, cells were treated with similar, alternative HDAC inhibitors. WT cells were treated with either 10uM suberoylanilide hydroxamic acid (SAHA) or 4mM Valproic Acid (VPA) for 24hrs. SAHA is a pan-HDAC inhibitor, which functions by directly binding the HDAC catalytic site (Finnin et al. 1999), whilst VPA inhibits most class I and II HDACs, with the exception of HDACs 6, 8, 9 and 10 (Bradbury et al. 2005; Chateauvieux et al. 2010). TSA is particularly similar to SAHA, with both acting as pan-HDAC inhibitors and having hydroxamic acid-based structures. They also both function by binding a key zinc molecule in the active sites of HDACs (Kim and Bae 2011). First, to ensure the inhibitors were working, we tested cells for increased histone acetylation levels. We extracted and purified histones, which were then subjected to western blotting with antibodies against two different histone acetylation marks (H4K5ac and H3K27ac). Levels of these histone signatures were measured relative to total histone H3, and accordingly, were significantly increased in the cells treated with either SAHA or VPA, confirming that HDACs were being blocked (Figure 5.9). Inhibitor-treated cells were also tested for anti-DNMT1 immunofluorescence. However, DNMT1 localisation was not altered in the treated cells, and consistently appeared nuclear, similar to the control and WT K562 cells (Figure 5.10). It would seem that at least in this cell-type, HDAC inhibition does not lead to DNMT1 re-localisation. This experiment suggests that HDACs are not responsible for DNMT1 re-localisation upon *CCDC26* KO, however, it would be useful to repeat these experiments with TSA in the future in order to exactly replicate the experiment performed by Arzenani et al. (2011) to confirm these results.

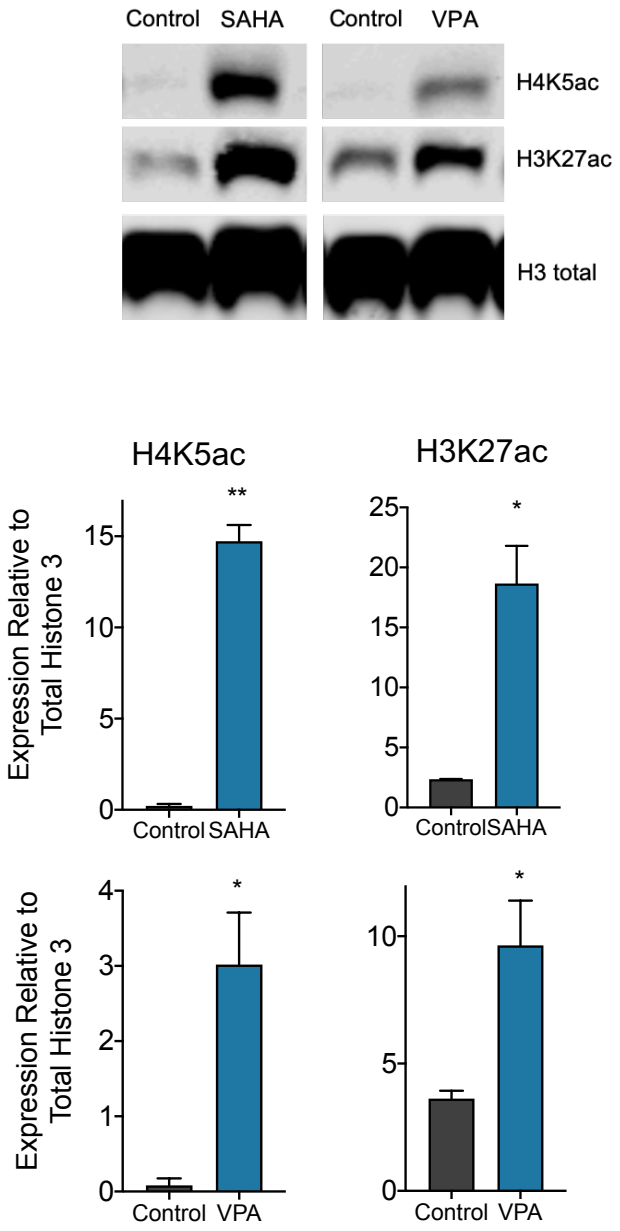


Figure 5.9. Confirmation of SAHA- and VPA-mediated HDAC inhibition. Immunoblotting using anti-H4K5ac and anti-H3K27ac antibodies demonstrates significant increases in these acetylation marks upon treatment of WT K562 cells with 10uM SAHA or 4mM VPA for 24 hours. Levels were measured relative to total levels of histone H3. Values represent the mean ± standard deviation. NS = Not significant, * P<0.05; **P<0.01 (unpaired, two-tailed *t* test).

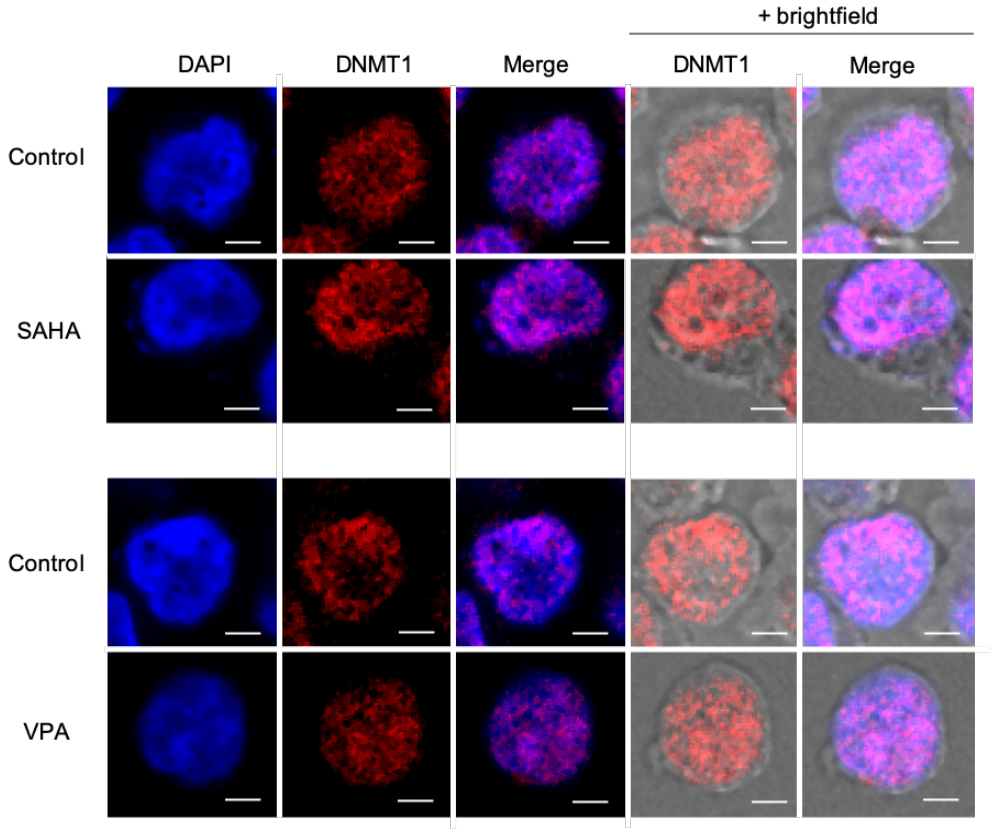


Figure 5.10 DNMT1 subcellular localisation does not change upon HDAC inhibition. DNMT1 (red) is primarily localised in the nucleus of both control cells and cells treated with HDAC inhibitors SAHA (10uM) and VPA (4mM) for 24 hours. Nuclei were stained with DAPI (Blue). The outline of the cell membrane can be seen with the addition of the brightfield lens in the right-hand panels (scale bar = 5um).

5.2.3 DNMT1 stability is not changed in *CCDC26* KO cells

We next focussed our analyses on DNMT1 itself, and any differences in the protein in WT compared to KO cells. We first chose to assess the stability of the protein. It is possible that DNMT1 is less stable in the absence of *CCDC26*, and subsequently more of the protein translocates to the cytoplasm for proteasomal degradation. Another possibility, is that DNMT1 is less prone to proteasomal-mediated degradation in *CCDC26* KO cells, and as a result, it is localised to the cytoplasm for degradation, but is not depleted as fast. Although we see no change in the total levels of DNMT1 in KO cells that might indicate these scenarios, we do see increased levels of DNMT1 transcription. We cannot rule out the possibility that a complex mix of changes could be occurring with regards to DNMT1 expression at transcriptional, translational and degradational levels, resulting in no clear changes to overall DNMT1 levels.

First, cells were treated with cycloheximide (CHX) for 12hrs and DNMT1 protein levels were monitored by immunoblotting. CHX is a translational inhibitor, which blocks synthesis of new proteins by binding the ribosome and halting translation elongation (Obrig et al. 1971). This inhibitor provides a means of measuring DNMT1 degradation. 12hrs after the addition of CHX to the cell media, DNMT1 levels had fallen, however, there appeared to be no major difference in the extent to which DNMT1 had been depleted between the WT and KO cells, indicating no difference in DNMT1 stability (Figure 5.11A).

DNMT1 is typically degraded by the ubiquitin-proteasomal pathway (Du et al. 2010). This first involves ubiquitination by ubiquitin ligase enzymes which allows targeting to the proteasome, a large protease complex, which unfolds and lyses the protein (Voges, Zwickl, and Baumeister 1999; Orlowski and Wilk 2000). MG132 is a drug which targets the catalytic subunit of the 26S proteasome and inhibits proteolytic activity (Lee and Goldberg 1998). We decided to also treat both our WT and KO cells with 10uM MG132 for 24hrs, followed by protein extraction and western blotting with anti-DNMT1, to further assess DNMT1 stability. The less stable the protein, the greater the impact of proteasomal inhibition, and consequently a greater increase in protein levels compared to control should be observed. However, we did not observe any significant differences in DNMT1 expression changes upon MG132 treatment between WT and KO cells, suggesting that DNMT1 is equally stable in all three cell lines (Figure 5.11B). Unusually, DNMT1 levels declined upon MG132 treatment. We were unable to treat cells for a longer period of time, or with higher concentrations of MG132 to see if DNMT1 levels eventually rose, as too many cells became non-viable and died (Appendix III, Figure 3). Immunoblotting for anti-c-JUN showed significant increases in protein levels upon treatment, confirming that the MG132 inhibitor had successfully blocked the proteasome (Nakayama, Furusu, et al. 2001).

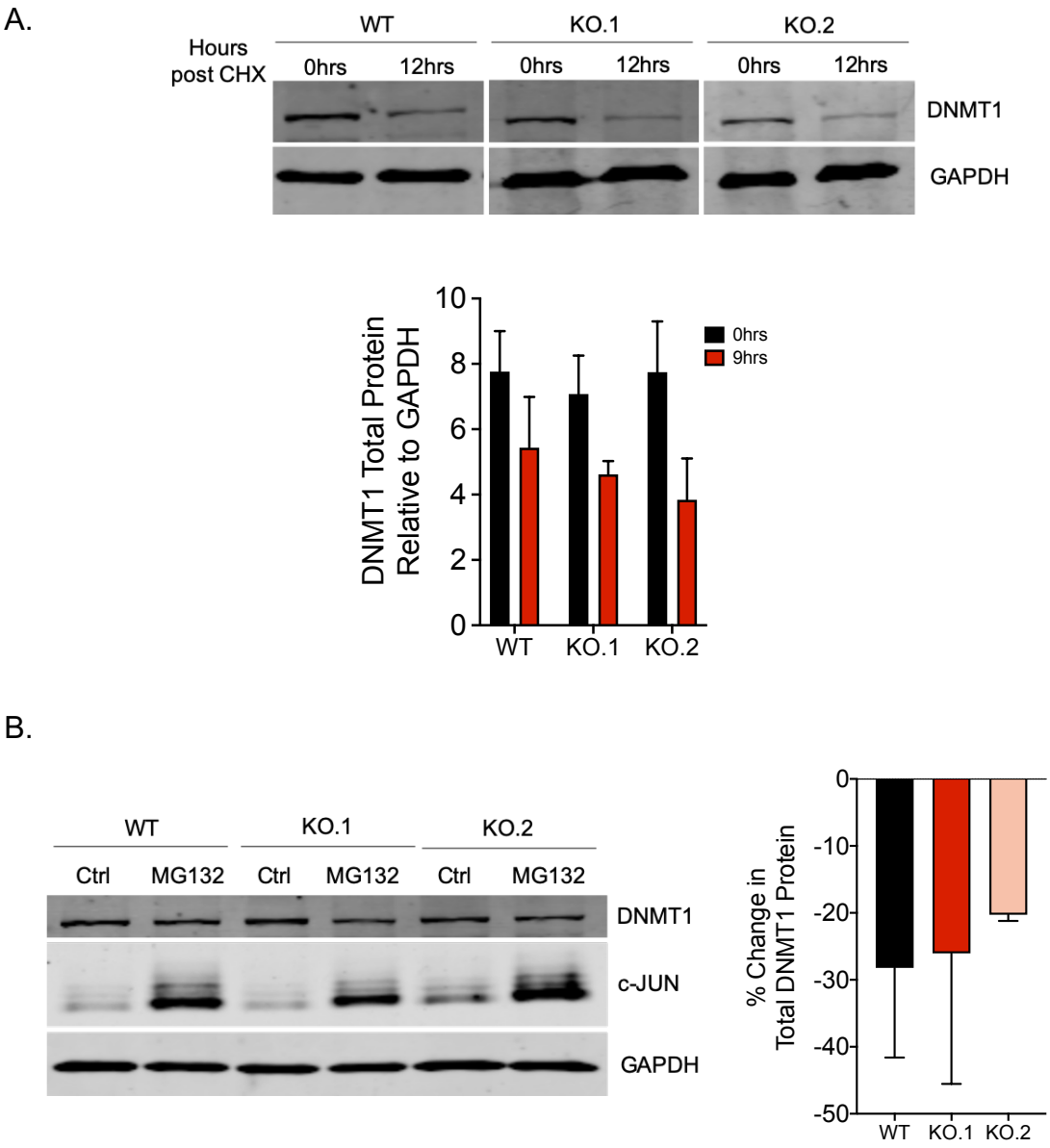


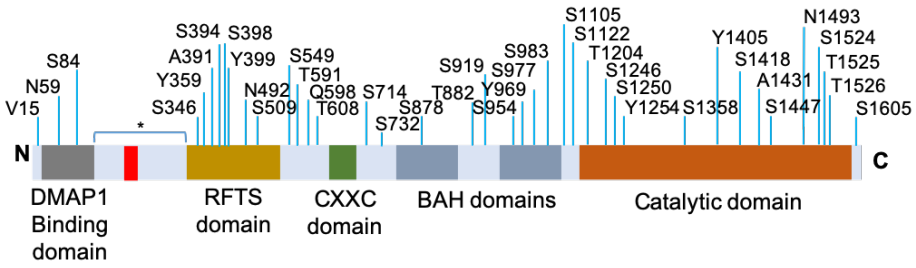
Figure 5.11. DNMT1 stability is not significantly affected in *CCDC26* KO cells. **A.** Cells were treated with CHX and DNMT1 immunoblotting was performed on protein extracted at 0hrs and 12hrs after treatment. DNMT1 levels decreased after 12hrs in all cell lines. The difference in the extent to which DNMT1 levels fell was not statistically significant between WT and KO cells. **B.** Upon treatment of cells with 10uM MG132, total DNMT1 protein levels fell in both WT and KO cells. The difference in the extent to which DNMT1 levels fell was not statistically significant between cell lines. Immunoblotting for c-JUN was also performed as a control to show that the MG132 inhibitor was working. C-JUN levels rose upon MG132 treatment in both WT and KO cells. Protein levels were measured relative to the housekeeping protein, GAPDH. Values represent the mean ± standard deviation (unpaired, two-tailed *t* test).

5.2.4 DNMT1 PTMs demonstrate no obvious changes in *CCDC26* KO cells

Post translational modifications (PTMs) have been shown to enhance the functional diversity of proteins, and influence numerous characteristics such as structure, enzymatic activity and localisation (Duan and Walther 2015). We therefore next began analysing how DNMT1 is post-translationally modified in our WT and KO cell lines. DNMT1 is a highly modified protein, possessing numerous ubiquitylation, phosphorylation, acetylation, sumoylation and mono- and di-methylation sites, many of which can be found within and surrounding the NLS (Figure 5.12). Nucleo-cytoplasmic trafficking has long been associated with protein PTMS (Nardozzi, Lott, and Cingolani 2010; Zhao et al. 2006; Rodriguez 2014). A previous study showed DNMT1 translocation is heavily impacted by phosphorylation of two key amino acids within the NLS (Hodge et al. 2007). They showed that AKT-mediated phosphorylation greatly enhances nuclear compartmentalisation, whilst mutations of these residues leads to a substantial increase in cytosolic DNMT1 (Hodge et al. 2007). In order to understand the mechanism of DNMT1 localisation in *CCDC26* KO, it will be interesting to profile DNMT1 PTM changes in KO cells.

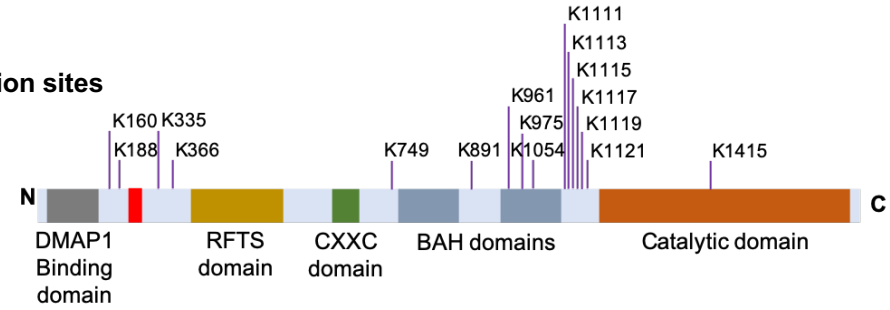
We began some preliminary analyses by purifying DNMT1 via immunoprecipitation followed by immunoblotting. Blotting with anti-phospho-tyrosine and anti-acetyl-lysine antibodies demonstrated no significant changes in total PTM levels for DNMT1 between WT and KO cells (Figure 5.13). This indicates that there are no major changes in these particular DNMT1 PTMs between WT and KO cells. However, significantly more research needs to be performed in this area in order to come to a conclusion regarding how DNMT1 might be differentially modified.

Phosphorylation sites

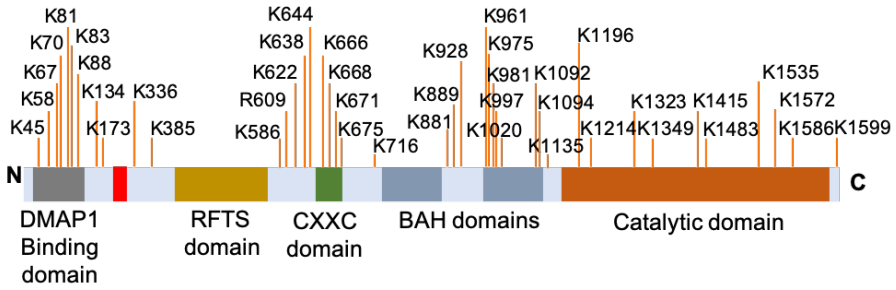


* = A125, S127, P128, P131, S133, T137, S141, S143, S152, S154, T165, T166, T168, S184, S189, S192, T208, S209, S227, S247, T249, T253, L258, S288, N308, S312, T328

Acetylation sites



Ubiquitylation sites



Sumoylation and methylation sites

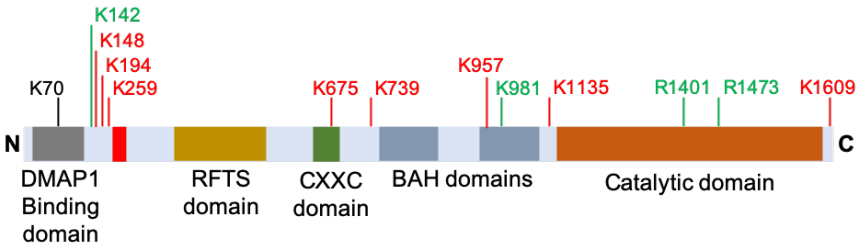


Figure 5.12 DNMT1 is a highly modified protein. DNMT1 sites of phosphorylation (blue; top panel), acetylation (purple; second panel), ubiquitylation (orange; third panel), sumoylation (red; fourth panel), mono-methylation (green; fourth panel) and di-methylation (black; fourth panel) are indicated, according to PhosphoSitePlus [Accessed:https://www.phosphosite.org/proteinAction?id=6047&showAllSites=true]. Asterisk (top panel) indicates the additional phosphorylation sites located between amino acids 103-331.

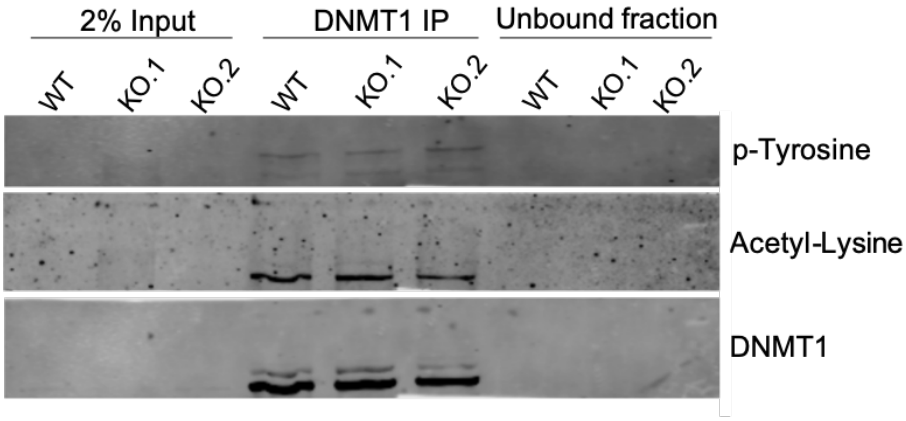


Figure 5.13 No obvious change in DNMT1 PTM levels between WT and *CCDC26* KO cells. DNMT1 was pulled down via immunoprecipitation and immunoblotted using anti-phospho-Tyrosine and anti-acetyl-lysine antibodies. No clear, significant changes in levels of these PTMs can be observed using this method, between WT and KO cells. 2% input represents 2% of the total protein extract prior to performing the DNMT1 pull-down. The unbound fraction represents the protein fraction that did not bind the DNMT1 antibody during the immunoprecipitation.

5.3 Discussion

Normal cellular function relies on the ability of the cell to transfer proteins between cellular compartments. Much of this involves protein translocation across the nuclear membrane. Since protein translation occurs in the cytosol, all nuclear proteins must be able to relocate back to the nucleus, and an inability to do this can be disastrous (Kim and Taylor 2017). Here, it is fascinating to observe such a dramatic localisation change in response to a lncRNA, especially considering that DNMT1 possesses an NLS (Alvarez-Ponce et al. 2018). In this chapter we have performed some initial experiments to determine the mechanism by which the DNMT1 NLS is somehow overridden and consequently re-localised to the cytosol.

Past studies that have reported cytosolic localisation of DNMT1 vary greatly, with each demonstrating different ways by which this occurs. During preimplantation development, an oocyte-specific form of DNMT1, referred to as DNMT1o, is generated. This protein demonstrates a preferential localisation within the cytosol, except at the 8-cell stage, at which it is transferred to the nucleus. Although DNMT1o is transcribed from an alternative promoter and consequently lacks 118 amino acids at the N-terminus compared to somatic DNMT1, this is not the reason for its cytosolic localisation. It remains unclear how this happens, but it has been postulated that an alternative, extended region of the N-terminus is critical, and complex folding of this area likely plays a large part in overriding the NLS (Cardoso and Leonhardt 1999).

In Alzheimer's disease, abnormal sequestration of DNMT1 as well as RNA Pol II has been reported in the cytosol (Mastroeni et al. 2013). This has been attributed to down-regulation of RAs-related Nuclear protein (RAN), a GTPase involved in both import and export of proteins between the nucleus and cytoplasm (Mastroeni et al. 2013). We cannot rule out the possibility that RAN or some similar nucleo-cytoplasmic trafficking protein could be implicated upon *CCDC26* KO. However, it is not something that we have immediately investigated as this is more of a general import/export mechanism. We would expect multiple proteins to be impacted if this were the case in *CCDC26* KO cells, however, DNMT1 is the only protein in our analyses that appears to be mis-localised in this way. Consequently, we suspect that the mechanism here is more specific to DNMT1.

Instead, we decided it was first important to establish the order of events following *CCDC26* KO, and to confirm whether DNMT1 re-localisation leads to DNA damage and apoptosis, or whether it is a downstream response of DNA damage and apoptosis. The latter would have meant the discovery of a novel, cellular distress pathway. Induction of DNA damage with oxidative stressor, hydrogen peroxide, has previously shown to cause re-localisation of DNMT1 to different regions of the genome, however, re-localisation from the nucleus entirely in response to damage, is not something that has ever been reported, to our knowledge (O'Hagan et al. 2011).

We first sought to demonstrate that DNMT1 mis-localisation can lead to DNA damage. Since DNMT1 functions in the nucleus, we hypothesised that a cytosolic

translocation should result in behaviour similar to a DNMT1 KD. Therefore, we attempted to mimic the DNMT1 re-localisation by generating a DNMT1 siRNA-mediated KD in WT cells. Unfortunately, this method was insufficient for our needs. 24hrs after siRNA transfection, DNMT1 protein levels were reduced by ~80%, however, we observed no significant changes in global DNA methylation levels nor DNA damage (**γ**-H2AX immunofluorescence). Taking in to consideration that K562 cells double approximately every 24hrs, we reasoned that this time period was not long enough to see the downstream effects of a DNMT1 KD such as reduced DNA methylation levels. However, given that the life span of siRNAs is short, later time points following transfection were also unsuitable, with DNMT1 levels being returned almost to normal by 48hrs (Appendix III, Figure 1). A more suitable alternative might have been to perform a stable transfection with DNMT1 shRNAs, and establishing a DNMT1 KD cell line. This would allow time for effects on DNA methylation to arise following depletion as well as subsequent downstream effects on gene expression and DNA damage. The alternative approach we applied however, was treatment of WT cells with a DNMT1 inhibitor.

DAC is a hypomethylating agent that is incorporated into DNA where it can inhibit methyltransferase activity (Stresemann and Lyko 2008). We have shown that 48hrs treatment of cells with this drug elicits a similar phenotype to *CCDC26* KO cells; global 5mC levels fall, levels of **γ**-H2AX rise and similar patterns of gene expression can be observed, indicating that a DNMT1 KD can lead to DNA damage. Although less time consuming, the disadvantage of this method for inducing a decrease in DNMT1 activity, is that the effects of DAC on DNMT3A and DNMT3B are

ambiguous. Whereas some reports insinuate that DAC solely affects DNMT1 (Chowdhury et al. 2015), others state that all three DNMTs are affected (Yu et al. 2018). Nevertheless, numerous other reports support our argument that DNMT1 KD leads to DNA damage and apoptosis (Milutinovic et al. 2003; Loughery et al. 2011; Unterberger et al. 2006).

After establishing that DNA damage can result from a DNMT1 KD or re-localisation, next we showed that DNMT1 re-localisation is not a result of DNA damage. Cisplatin-treated cells did not show any change in DNMT1 compartmentalisation, nor any changes in global DNA methylation. Furthermore, DNMT1-responsive genes showed varied expression patterns in cisplatin-treated cells compared to *CCDC26* KO and DAC-treated cells. Taken together, these findings suggest that DNMT1 re-localisation is a *CCDC26*-specific effect. Following *CCDC26* KO, DNMT1 is re-localised to the cytosol, generating a phenotype similar to DNMT1 KD cells, which ultimately leads to increased DNA damage and apoptosis.

Our next route of investigation concerned the possibility that DNMT1 re-localisation was occurring as a result of some sort of HDAC inhibition. A few years ago, a study reported that HDAC inhibitor, TSA, impacts DNMT1 nuclear kinetics. They used Fluorescence Correlation Spectroscopy (FCS) to demonstrate increased levels of DNMT1 movement in the nuclei of cells. TSA treatment subsequently led to increased cytosolic DNMT1, reduced global DNA methylation and increased apoptosis (Arzenani et al. 2011). Our research has produced some very similar observations. Another report that made this route of inquiry attractive, is that the

CDKN1A gene has been shown to be very responsive to HDAC inhibitors, with expression being rapidly induced upon treatment (Rocchi et al. 2005). In previous qRT-PCRs, we too have seen upregulation of this gene in *CCDC26* KO cells (Figure 4.8).

We have already analysed several histone acetylation marks (H3K27ac, H4K5ac, H3K9ac and H4K16ac) in Chapter 3 (Figure 3.12), but did not observe any significant differences in the global levels of these signatures between our WT and KO cells. This suggests that histone acetylation is not affected upon *CCDC26* KO. Nevertheless, this does not rule out the possibility that HDACs are in some way implicated in our observations. Substrate specificity of HDACs still remains somewhat ambiguous, with a great deal of target overlap existing between different members of the HDAC family (Seto and Yoshida 2014). Furthermore, it is not clear how HDAC inhibition causes DNMT1 re-localisation, and whether it can be attributed to one or several HDACs. Arzenani et al. (2011) show that TSA has the most prominent effect on HDAC2. If the DNMT1 re-localisation we see in our *CCDC26* KO cells is due to some inhibition or reduction of just one or two HDACs, due to the overlap in the targets of different HDACs, we are unlikely to see any effect on global histone acetylation levels. Considering all this, we decided to pursue this avenue further, first analysing levels and localisation of HDAC2, which showed no evidence of being affected by *CCDC26* KO. Following this, we attempted to replicate the results shown by Arzenani et al. (2011), by treating WT K562 cells with HDAC inhibitors, SAHA and VPA. We were unable to replicate their results, and DNMT1 did not appear to change its localisation in response to these drugs. All of

this evidence suggests that HDAC inhibition or reduction is not the means by which DNMT1 is re-localised in response to a *CCDC26* KO in K562 cells. Given that we were unable to replicate the results shown by Arzenani et al. (2011), it is possible that this is a cell-specific effect. To our knowledge, no other studies have reproduced this result in other cell lines.

We chose to move on and began examining DNMT1 itself and any differences in the protein upon *CCDC26* KO. Previous studies have shown increased cytosolic DNMT1 in response to mutations in the amino acid sequence as well as changes to PTMs. Single amino acid substitution mutations within the RFTS domain of DNMT1 have been reported as causal for the mis-localisation of DNMT1 to the cytosol in cases of HSAN1E. These mutations were believed to impact protein folding and consequently led to the formation of cytosolic aggresomes followed by induction of autophagy (Baets et al. 2015). These types of mutations are generally hereditary, what is interesting however, is the observation that misfolding of the DNMT1 protein can affect its localisation in this way.

Considering the possibility that DNMT1 could be mis-localised as a result of it being misfolded and unstable, and consequently needing to be degraded, we assessed the protein's stability in the presence or absence of *CCDC26*. We did this by measuring DNMT1 levels in WT and KO cells following treatment with MG132, with the expectation that less stable protein would be impacted more by proteasomal inhibition and consequently demonstrate increased levels. Unusually, we observed a decrease in DNMT1 levels upon MG132 treatment compared to control in both

WT and KO cells. The extent to which DNMT1 levels were affected were similar between WT and KO cells, indicating similar levels of stability, however, it is uncommon to observe a decrease in protein. In previous instances where MG132 has resulted in suppression of protein, evidence suggests that polyubiquitinated protein degradation can be triggered by MG132-mediated induction of autophagy (Wang et al. 2019). It is possible a similar process has occurred here.

Since unexpected variables seemed to be at play in our MG132 experiment, we also measured DNMT1 stability using CHX, an inhibitor of protein synthesis (Obrig et al. 1971). CHX treatment confirmed that DNMT1 stability was not significantly different in the KO. From these results, we can conclude that it is unlikely that DNMT1 re-localisation to the cytosol is occurring as a result of DNMT1 stability in the KO cells.

Another paper (Hodge et al. 2007) too demonstrates that mutations within the N-terminal of DNMT1 leads to cytosolic accumulation, however, this has been related to the phosphorylation status of residues within this region. There are several reports of DNMT1 regulation by post-translational modifications, which would explain why it is so heavily modified (Figure 5.12). For example, lysine methylation and demethylation by SET7 and LSD1 respectively, as well as phosphorylation of various amino acids, have been shown to impact DNMT1 degradation (Esteve et al. 2011; Esteve et al. 2009; Wang et al. 2009). Additionally, phosphorylation-mediated protein localisation has long been established (Nardozzi, Lott, and Cingolani 2010), however, there is very little literature which relates DNMT1 phosphorylation with

localisation. One study showed that removal of two AKT kinase target residues at the NLS abrogates nuclear localisation of DNMT1, probably by disturbing the affinity of DNMT1 for nuclear import proteins. Yet, replacement with phosphor-mimics allowed nuclear import, indicating that it is the phosphorylation of these residues that is important, rather than the specific residues themselves (Hodge et al. 2007). Similar mechanisms have been observed for other proteins; for example, phosphorylation of a Serine residue within the NLS of the Epstein-Barr virus nuclear antigen 1 (EBNA-1) protein enhances binding affinity to an importin protein, and subsequently increases translocation to the nucleus (Kitamura et al. 2006). Other PTMs have also been shown to affect protein localisation and dynamics; N-terminal acetylation has been shown to implicate protein targeting to both the golgi apparatus and endoplasmic reticulum (Behnia et al. 2004; Forte, Pool, and Stirling 2011) and blocking of Arginine methylation of splicing factor SF2/ASF leads to cytosolic accumulation (Sinha et al. 2010). More specifically, it has been reported that DNMT1 is acetylated within one of the BAH domains between the N-terminal regulatory and C-terminal catalytic domains, which appears to disrupt binding and increase the mobility of the protein (Kim, Sprung, et al. 2006).

We have performed some preliminary experiments in our investigation into *CCDC26*-induced changes to DNMT1 PTMs, by assessing phospho-tyrosine and acetyl-lysine signatures present on DNMT1 in WT and KO cells. Our initial application of immunoprecipitation and western blotting techniques does not indicate that there are any differences in the overall levels of these marks for DNMT1 between cell types. Of course, considerably more research needs to be

conducted before we can make any conclusions about this. So far, we have only analysed two PTMs, and we must also consider the extent to which DNMT1 is modified; there are so many PTM sites within this protein, that it is possible that different combinations of modifications can alter the protein in different ways. That being considered, it might be more useful in the future to apply a mass spectrometry-based approach.

Although we have not established the mechanism by which DNMT1 is re-localised upon *CCDC26* KO, we have made a considerable start to this investigation. Going forward, it will be important to complete the examination of differences in DNMT1 PTMs in the presence and absence of *CCDC26*, as well as begin structural analysis of DNMT1 folding.

CHAPTER 6

THE EFFECT OF *CCDC26* AND DNMT1 RE-LOCALISATION ON NUCLEAR ARCHITECTURE

6.1 Introduction

In addition to gaining mechanistic insight into the DNMT1 re-localisation observed in *CCDC26* KO cells, it is also important to understand the implications of these effects. In this chapter we have commenced investigation into alterations in other biological and cellular processes in the KO cells to further understand the impact of *CCDC26* and DNMT1 mis-localisation.

The most prominent organelle within the nucleus is the membrane-less nucleolus, which appears as multiple, irregularly stained bodies in humans when viewed under the microscope. These structures are assembled around heterochromatic DNA regions comprising operons encoding ribosomal RNA (rRNA), commonly referred to as the Nucleolar Organising Regions (NORs). They are found on the p- (short) arm of the five acrocentric chromosomes (13, 14, 15, 21 and 22) (Henderson, Warburton, and Atwood 1972; Sakai et al. 1995). Remarkably, nucleoli can generate wherever active rRNA genes lie, where they can then function in building ribosomes (Karpen, Schaefer, and Laird 1988). Earlier studies have proposed that DNMT1 plays a role in maintaining nucleolar integrity and structure. DNMT1 KD experiments have resulted in demethylation of genes encoding rRNA, as well as nucleolar fragmentation (Espada et al. 2007; Ianni et al. 2017). LncRNAs too have shown evidence of a role in nucleolar regulation, particularly in response to cellular stressors (Jacob et al. 2013; Bierhoff et al. 2014; Xing et al. 2017).

It is logical that similar factors regulate both nucleolar and chromosomal organisation, and accordingly, DNMT1 is also implicated in the organisation of larger chromatin domains within the nucleus. A past study reported that upon DNMT1 KD, the spatial localisation of heterochromatin was altered immensely, corresponding to changes in chromosome positioning and organisation (Espada et al. 2004). RNA has also long been known to also influence nuclear architecture. An excellent example is lncRNA, *FIRRE*, which has been proven essential for a clustered trans-chromosomal interaction to occur between the *FIRRE* locus and four trans-sites, for regulation of adipogenesis (Hacisuleyman et al. 2014).

Here I have analysed both nucleolar structure and chromosomal organisation to further comprehend the downstream effects of *CCDC26* KO, to help fully understand the biological role and impact of this lncRNA.

6.2 Results

6.2.1 Nucleolar structure and function is not significantly impacted by *CCDC26* KO

Following reports that both DNMT1 and lncRNAs can play roles in maintaining nucleolar integrity (Espada et al. 2004; Hacisuleyman et al. 2014; Rego et al. 2008; Zhang, Huynh, and Lee 2007), we compared the structure of nucleoli in WT and *CCDC26* KO cells. Hundreds of proteins have been shown to flux between the nucleolus and the nucleoplasm (Andersen et al. 2005), so we first selected three nucleolar proteins, Nucleolin, Nucleolar Protein 10 (NOL10) and Ki-67, to assess

the nucleolar proteome. Of the nucleolar proteins, Nucleolin has been the most extensively studied, functioning in all aspects of ribosome biogenesis (Ma et al. 2007; Ginisty et al. 1999). NOL10 has been studied considerably less, but is known to be required for synthesis of the 40S ribosomal subunit by complexing with additional proteins (Bammert et al. 2016). Ki-67 on the other hand has shown evidence of coordinating assembly and disassembly of nucleoli as well as ribosome assembly, in addition to acting as a proliferation marker (Schmidt et al. 2003; Rahmanzadeh et al. 2007).

Immunoblotting showed no significant change in the protein levels of Nucleolin and NOL10 (Figure 6.1A). Furthermore, immunostaining for nucleolin and Ki-67 showed no dramatic changes in the distribution of these proteins; staining for both demonstrated several nucleolar bodies in both WT and KO cells, with a debatably, slightly more diffuse pattern in the KO cells (Figure 6.1B). To assess nucleoli with regards to DNMT1, we also performed Nucleolin and DNMT1 co-immunofluorescence. We have previously observed punctate DNMT1 foci in the nuclei of KO cells (Figure 4.4B), that are arranged such that it is possible they are localised to some nuclear body. We wanted to assess whether these foci correspond to nucleoli, however, the immunostaining demonstrated that there is very little overlap between Nucleolin and DNMT1 (Figure 6.2).

The second approach I adopted to visualise nucleoli was AgNOR staining. A large number of the proteins associated with NORs are agyrophilic, a characteristic that can be taken advantage of by silver staining. The resulting black spots correspond

to nucleoli and are referred to as AgNORs (Trere 2000). Similar to the nucleolar protein immunostaining results, AgNOR staining demonstrated no obvious difference in nucleolar pattern between WT and KO cells (Figure 6.3).

To conclude our analysis of nucleolar organisation, we also examined rRNA levels to assess whether transcription at NORs had been impacted. rRNA is transcribed as a single 45S transcript, which is later cleaved to yield 5.8S, 18S and 28S RNAs. The cleaved transcripts are then assembled with other RNA molecules and ribosomal proteins for the construction of ribosomes (Pena, Hurt, and Panse 2017). qRT-PCRs using primers specific to each of these rRNA species was performed, but largely demonstrated no significant differences in levels between WT and KO cells. The only exception was KO.2 28S levels, which were slightly greater than the WT cells (Figure 6.4).

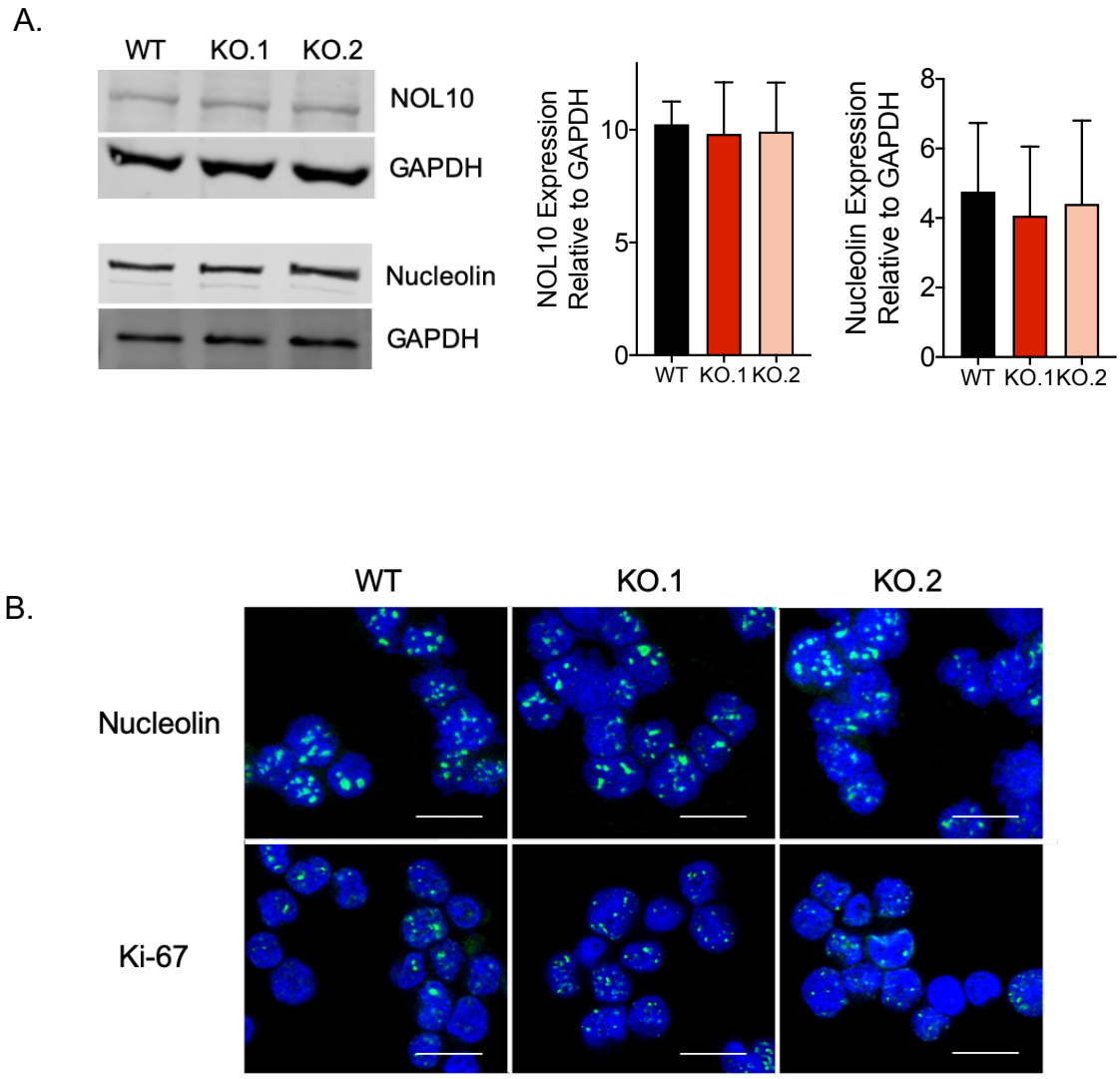


Figure 6.1 Nucleolar protein levels and localisation are not significantly altered upon *CCDC26* KO. A. Western blotting indicates no significant changes in the levels of nucleolar proteins, Nucleolin and NOL10, in *CCDC26* KO cells compared to WT cells. Values represent the mean ± standard deviation (n=3). (unpaired, two-tailed *t* test) **B.** Immunostaining for nucleolar proteins, Nucleolin and Ki-67 (green), demonstrates no significant changes in levels and patterns of the proteins between WT and KO cells. Nuclei were stained with DAPI (blue) (scale bar = 20um).

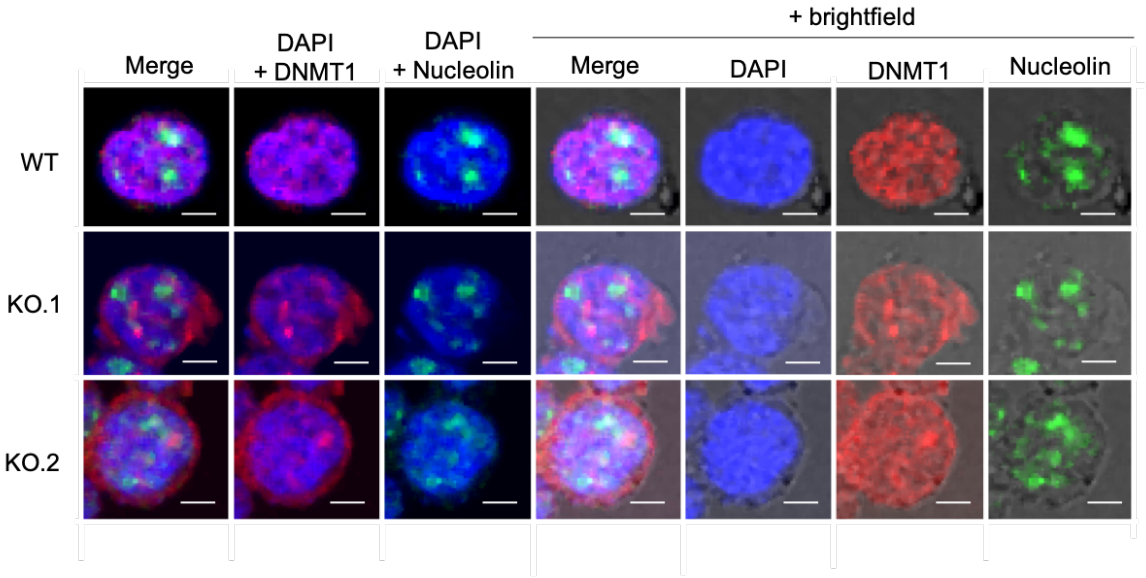


Figure 6.2 DNMT1 and Nucleolin co-immunofluorescence. The punctate, nuclear DNMT1 (red) foci remaining within CCDC26 KO cells, are not co-localised with Nucleolin (green). Nuclei were stained with DAPI (blue). The outline of the cell membrane can be seen with the addition of the brightfield lens in the right-hand panels (scale bar = 5um).

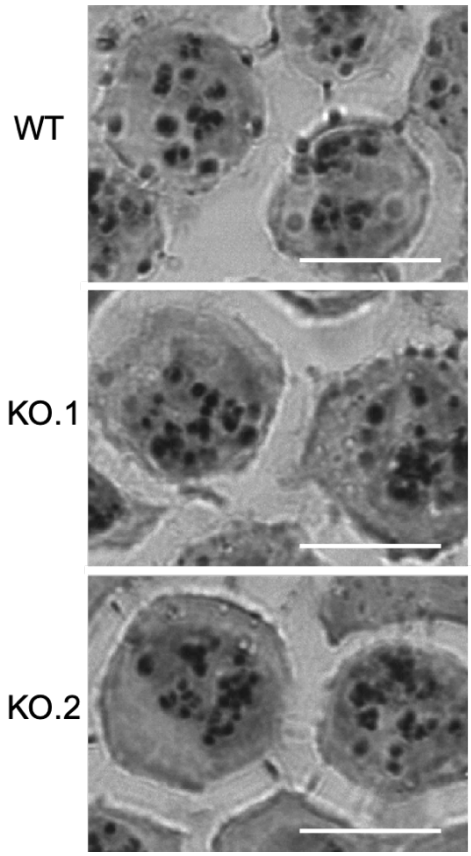


Figure 6.3 Nucleolar structure is similar in WT and *CCDC26* KO cells. Nucleoli were visualised by silver staining AgNORs. Both WT and KO cells demonstrate similar patterns and distributions of AgNORs. Cells were visualised by light microscopy using a Nikon Eclipse E600 Wide-field microscope (scale bar = 10um).

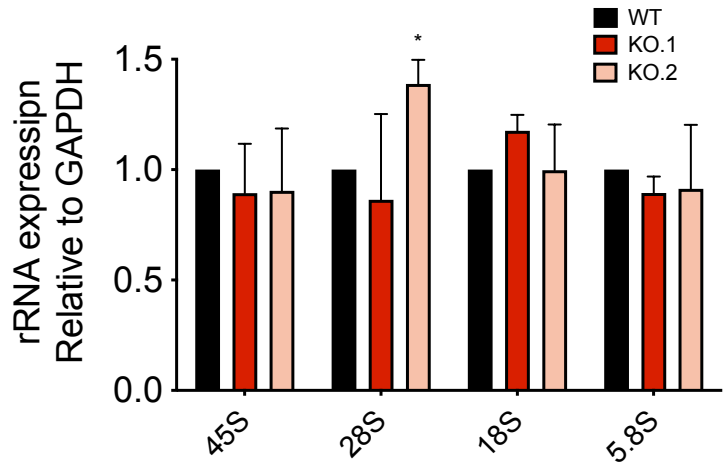


Figure 6.4 rRNA levels are not significantly affected by *CCDC26* KO. qRT-PCRs were performed using primers specific to 45S rRNA and its cleavage products, 28S rRNA, 18S rRNA and 5.8S rRNA. No significant difference can be observed in rRNA levels between WT and KO cells except for 28S rRNA in KO.2 cells. Values represent the mean ± standard deviation (n=3). * $P<0.05$ (unpaired, two-tailed *t* test).

6.2.2 *CCDC26* KO cells demonstrate changes in 3D nuclear localisation of chromosomal loci

Taking into consideration previous reports of both DNMT1 and lncRNAs being involved in the regulation of large-scale organisation of nuclear and chromatin domains (Espada et al. 2004; Hacisuleyman et al. 2014), we next considered whether *CCDC26* KO cells demonstrate any changes to chromosomal organisation and positioning.

In order to identify a starting point for this investigation, previously analysed RNA sequencing data for WT and KO cells was consulted (performed and analysed previously by Dr Aditi Kanhere). Here, we have taken significantly differentially-regulated genes (P<0.0025), and submitted them into the Database for Annotation, Visualisation and Integrated Discovery (DAVID). DAVID can be used for identifying any biological characteristics, functions and pathways that are overrepresented in a set of genes (Huang da, Sherman, and Lempicki 2009). Chromosome and cytoband annotations demonstrated that several loci contained stretches of neighbouring, down-regulated genes. Of these, the Chr.1p13.3 and Chr.16p11.2 loci showed overrepresentation of down-regulated genes at highest statistical significance (Figure 6.5). 3.8% of total down-regulated genes were located at Chr.1p13.3, and 4.9% were located at Chr.16p11.2. Consequently, these loci were used as a starting point for analysing any changes in chromosomal organisation. Lists of the down-regulated genes within these regions can be found in Tables 6.1 and 6.2, along with expression fold-change values and the statistical significance of the expression change according to RNA-seq data. We also used DAVID to assess

functional similarities, however, functional annotations identified no significant similarities in the functions and pathways associated with these genes.

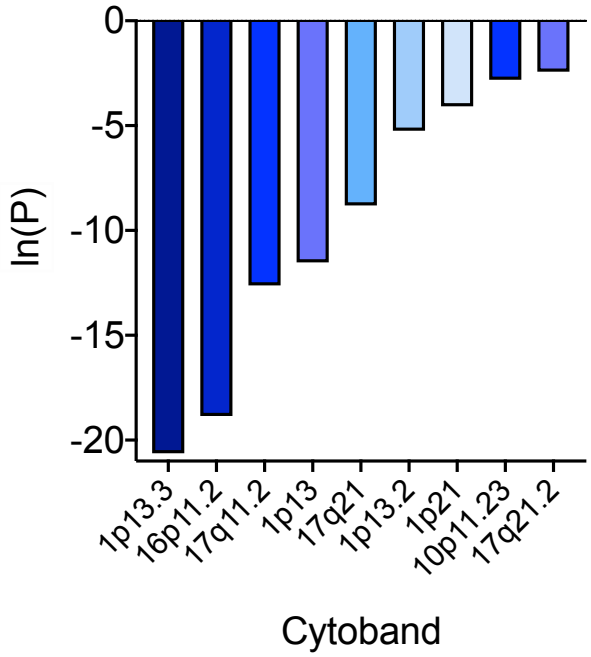


Figure 6.5 Locations of significantly down-regulated genes in *CCDC26* KO cells. Down-regulated genes identified by RNA-seq were analysed using DAVID. Several loci (indicated) demonstrated stretches of neighbouring, down-regulated genes. Chr.1p13.3 and Chr.16p11.2 loci showed overrepresentation of down-regulated genes at highest statistical significance

Table 6.1 *CCDC26* KO Chr.1p13.3 down-regulated genes
List of *CCDC26* KO down-regulated genes, according to RNA-seq data, located at Chr.1p13.3

Gene Symbol	Gene Name	Log2(fold-change)	*P*-value
GPSM2	G-protein signalling modulator 2 (GPSM2)	-1.2895	0.0004
RAP1A	RAP1A, member of RAS oncogene family (RAP1A)	-0.633606	0.0018
TAF13	TATA-box binding protein associated factor 13 (TAF13)	-0.99616	5.00E-05
CLCC1	chloride channel CLIC like 1 (CLCC1)	-1.14617	5.00E-05
CEPT1	choline/ethanolamine phosphotransferase 1 (CEPT1)	-1.02908	0.00015
FAM102B	family with sequence similarity 102 member B (FAM102B)	-1.06242	5.00E-05
LAMTOR5	late endosomal/lysosomal adaptor, MAPK and MTOR activator 5 (LAMTOR5)	-0.649508	0.0005
PSRC1	proline and serine rich coiled-coil 1 (PSRC1)	-1.00542	5.00E-05
PRMT6	protein arginine methyltransferase 6 (PRMT6)	-1.98625	5.00E-05
SLC25A24	solute carrier family 25 member 24 (SLC25A24)	-0.997384	5.00E-05
TMEM1676	transmembrane protein 167B (TMEM167B)	-1.23812	5.00E-05

Table 6.2 *CCDC26* KO Chr.16p11.2 down-regulated genes
List of *CCDC26* KO down-regulated genes, according to RNA-seq data, located at Chr.16p11.2

Gene Symbol	Gene Name	Log2(fold-change)	*P*-value
FBXL19	F-box and leucine rich repeat protein 19 (FBXL19)	-1.97115	5.00E-05
FBXL19-AS1	FBXL19 antisense RNA 1 (head to head) (FBXL19-AS1)	-1.70818	0.0008
ORAI3	ORAI calcium release-activated calcium modulator 3 (ORAI3)	-2.39615	5.00E-05
SETD1A	SET domain containing 1A (SETD1A)	-1.53015	5.00E-05
BCKDK	branched chain ketoacid dehydrogenase kinase (BCKDK)	-1.70685	5.00E-05
C16orf58	chromosome 16 open reading frame 58 (C16orf58)	-1.24626	5.00E-05
KAT8	lysine acetyltransferase 8 (KAT8)	-1.47784	5.00E-05
PRSS8	protease, serine 8 (PRSS8)	-1.35287	0.0008
STX1B	syntaxin 1B (STX1B)	-2.9954	0.00015
STX4	syntaxin 4 (STX4)	-1.61919	5.00E-05
VKORC1	vitamin K epoxide reductase complex subunit 1 (VKORC1)	-1.96572	5.00E-05
ZNF267	zinc finger protein 267 (ZNF267)	-1.20455	5.00E-05
ZNF668	zinc finger protein 668 (ZNF668)	-1.77992	0.0001
ZNF720	zinc finger protein 720 (ZNF710)	-1.45847	5.00E-05

DNA Fluorescence *In Situ* Hybridisation (FISH) was implemented in order to visualise these loci within the nucleus. Fluorescent probes specific to genomic loci were generated using the Invitrogen FISH Tag DNA Multicolour Kit (see Chapter 2.3.1.3 for more details). The length of time which nick translation reactions were allowed to proceed required optimisation to generate BAC DNA fragments of appropriate length (200-700bp). Running the fragmented DNA on agarose gels indicated that BACs of length ~190,000bp were best nick translated at 15^{o}C for 60 mins, whilst 15 mins was sufficient for smaller BACs of length 130,000-140,000bp (Appendix I, Figure 1).

Prior to hybridising probes to DNA overnight at 37^{o}C (Chapter 2.3.2.2), slides were heated to denature the double stranded DNA and probes, to allow annealing. Slides were heated to either 70^{o}C, 80^{o}C or 95^{o}C for 3-5 mins, however, the two lower temperatures failed to yield any fluorescent signal. Heating to 95^{o}C for 3 mins was required to sufficiently denature the probes and the DNA within fixed cells, in order for binding and subsequent visualisation of fluorescent signals (Appendix I, Figure 2).

Although increased temperatures enabled visualisation of fluorescent signals within nuclei, high levels of background fluorescence were present. This meant that the true signal corresponding to our locus of interest could not be distinguished. This was solved with the addition of Cot-1 DNA to the hybridisation mixture (Appendix I, Figure 2). Cot-1 is a DNA blocking agent, enriched for repetitive DNA sequences, and is applied to limit non-specific hybridisation (Kang et al. 2007).

We first investigated the nuclear localisation of CCDC26, Chr.1p13.3 and Chr.16.p11.2 genomic loci using the fluorescent probes specific to these genomic regions. As detailed in Chapter 2.3.2.2, probes were hybridised to DNA of fixed cells, which were also stained with the nuclear marker, DAPI. Cells were then analysed in three dimensions from Z-stack images taken using a Leica TCS SP8 confocal microscope, using a 63X objective lens and a Z-stack step size of 1.4um. DNA FISH with CCDC26, Chr.1p13.3 and Chr.16p11.2 probes recurrently demonstrated 3, 4 and 3 fluorescent signals respectively. This is consistent with the Catalogue of Somatic Mutations in Cancer (COSMIC) Cell Lines Project, which indicates that the average copy number of chromosomes 8, 1 and 16 are 3, 3.71 and 3.08 respectively in K562 cells. This indicated that the probes were specific to the loci of interest.

Observation of fluorescent signals showed that CCDC26 lies at the nuclear periphery in both WT and KO cells (Figure 6.6). The Chr.1p13.3 and Chr.16.p11.2 loci also showed preferential localisation at the nuclear periphery in WT cells, but demonstrated a more central nuclear positioning in the KO cells (Figure 6.7). To quantify this, we measured fluorescent signals at the nuclear periphery, as defined by the edge of the DAPI staining, and calculated the percentage of peripheral loci for each probe for both WT and KO cells. This was calculated for 500 individual nuclei per replicate (n=3). In WT cells, approximately 60% of Chr.1p13.3 signals measured lay at the nuclear periphery. This percentage however significantly

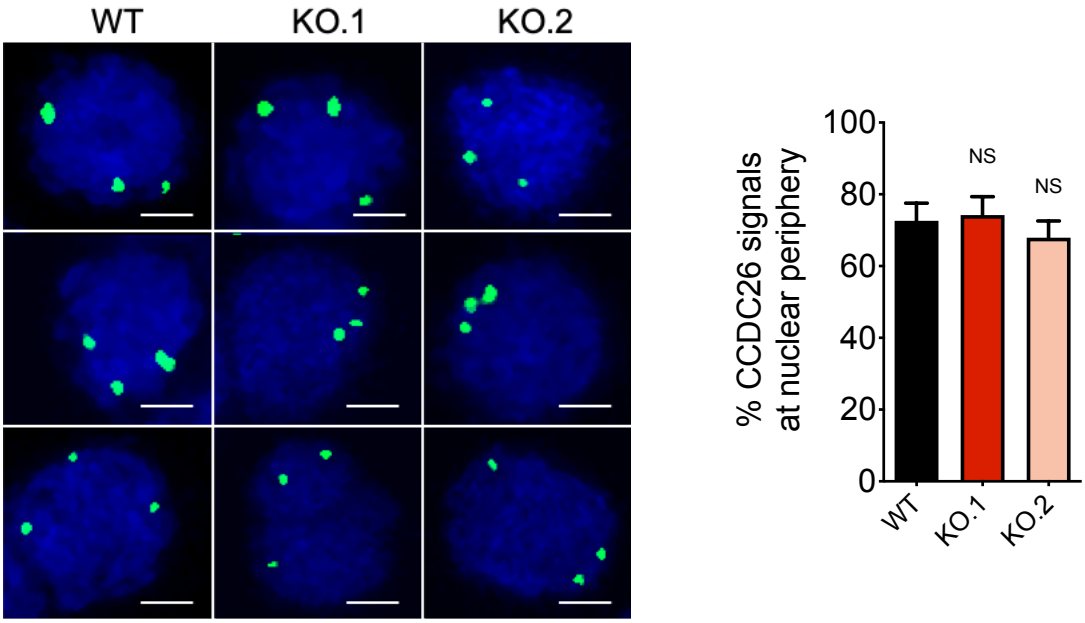


Figure 6.6 CCDC26 gene is located at the nuclear periphery. Maximum intensity projections of Z-stack images of both WT and *CCDC26* KO cell lines show that CCDC26 loci preferentially lie at the nuclear periphery. The percentage of CCDC26 signals (green) at the nuclear periphery was quantified and demonstrated no significant change in KO cells compared to WT. Nuclei were stained with DAPI (blue). Measurements were calculated for 500 cells per replicate. Values represent the mean ± standard deviation (n=3) (unpaired, two-tailed *t* test) (scale bar = 5um).

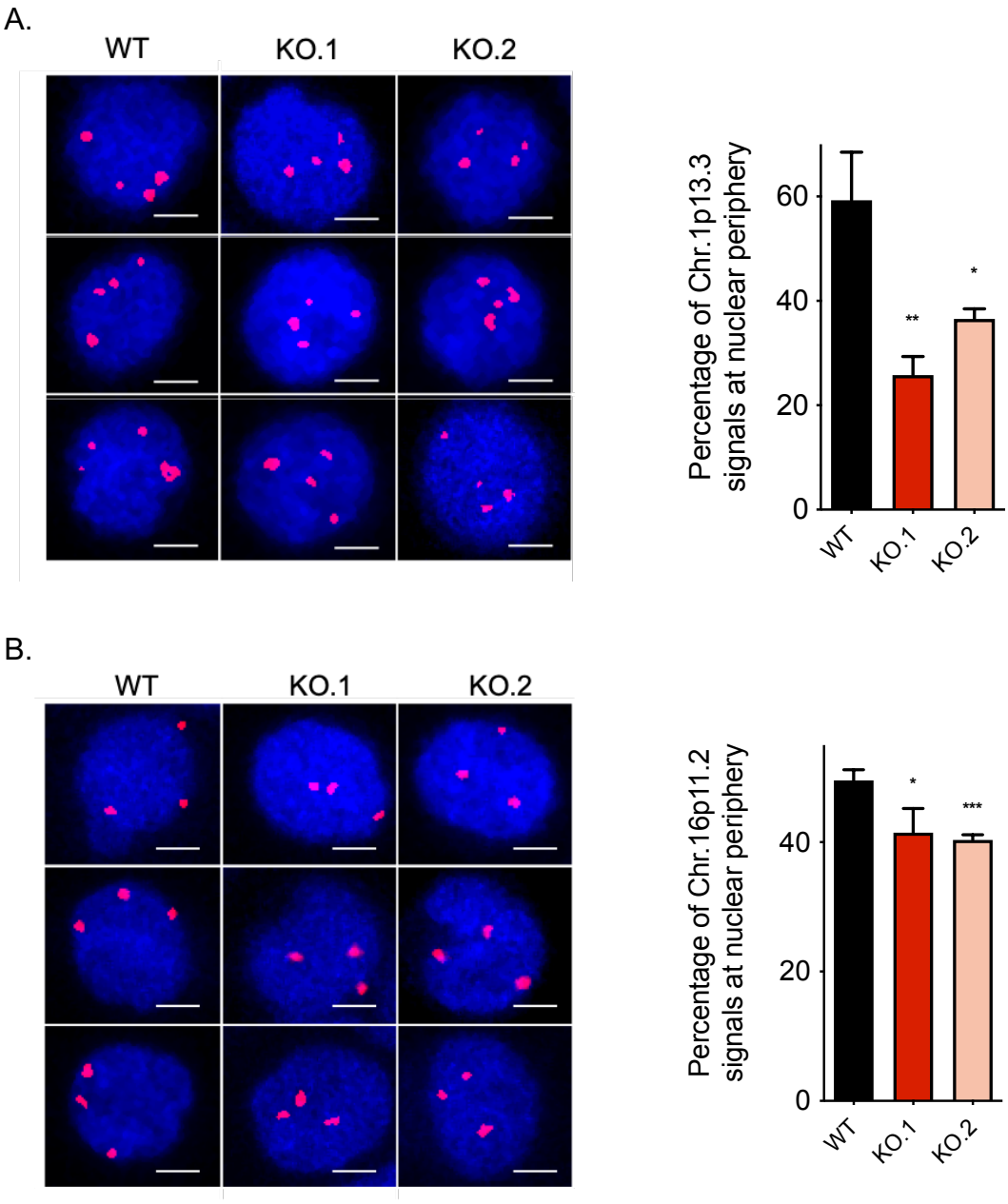


Figure 6.7 Chr.1p13.3 and Chr.16p11.2 loci are located at the nuclear periphery of WT but not *CCDC26* KO cells. Maximum intensity projections of Z-stack images of WT cells show that **A.** Chr.1p13.3 (red) and **B.** Chr.16p11.2 (red) loci preferentially lie at the nuclear periphery. However, they appear more centrally localised in *CCDC26* KO cells. The percentage of loci at the nuclear periphery were quantified and demonstrated no significant decreases in KO cells compared to WT. Nuclei were stained with DAPI (blue). Measurements were calculated for 500 cells per replicate. Values represent the mean ± standard deviation (n=3). * $P<0.05$; ** $P<0.005$; *** $P<0.001$ (unpaired, two-tailed *t* test) (scale bar = 5um).

dropped to approximately 25% and 30% in KO.1 and KO.2 cells respectively. A similar result was also observed for Chr.16p11.2 signals, although to a smaller extent, which demonstrated around 50% peripheral signals in WT compared to 40% in KO cells. Although the difference is smaller, it is significantly so for both KO cell lines. Approximately 75% of CCDC26 signals on the other hand were located at the nuclear periphery in both WT and KO cell lines.

Given that the genes within the Chr.1p13.3 and Chr.16p11.2 regions are similarly down-regulated in KO cells, we hypothesised that these genes are co-regulated in WT cells, that in some way requires *CCDC26*. Following the observations in Figures 6.6 and 6.7, we therefore questioned whether these loci lie in a close vicinity in the nuclei of WT cells. To verify this, I performed co-DNA FISH for Chr.1p13.3 and Chr.16p11.2 with CCDC26. Three-dimensional distances between loci were then calculated using the computational software, Icy, and an automated program written in R. The smallest distances between loci were then noted for 500 cells in each sample.

Interestingly, the WT cells showed that CCDC26 clustered with Chr.1p13.3 and Chr.16p11.2 loci (Figure 6.8). This pattern appeared to be lost however in the KO cells, which demonstrated an increase in spatial proximity between these loci. Quantification supported this observation, showing an increase of approximately 300 pixels in the distance between CCDC26 and these loci in KO cells.

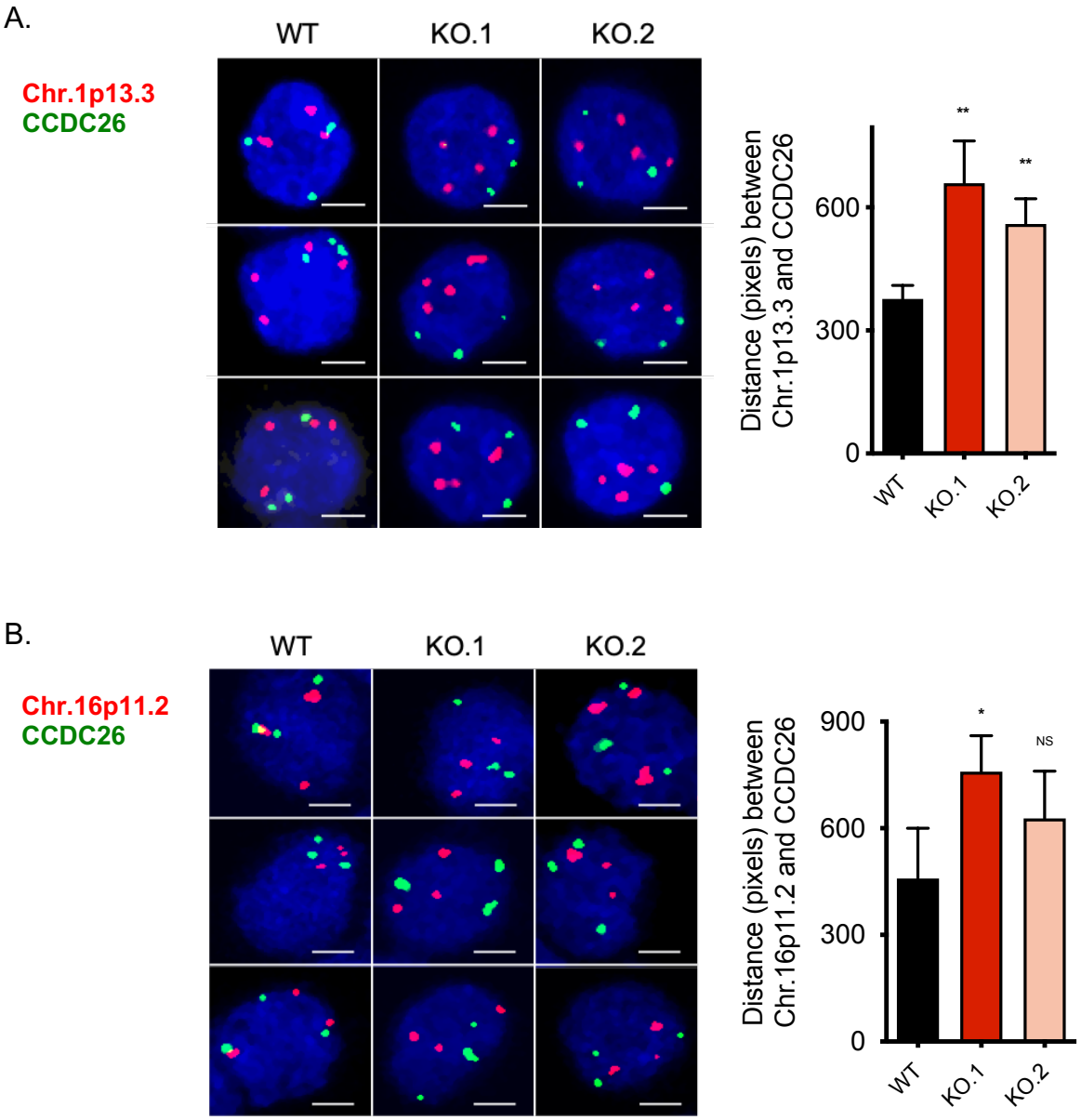


Figure 6.8 CCDC26, Chr.1p13.3 and Chr.16p11.2 loci are in close spatial proximity in WT cells but not CCDC26 KO cells. Maximum intensity projections of Z-stack images of WT and *CCDC26* KO cell lines show that CCDC26 (green) and either **A.** Chr.1p11.2 (red) or **B.** Chr.16p11.2 (red) loci lie in close proximity in WT cells. However, the loci move apart in the KO cells, with Chr.1 and Chr.16 loci moving toward the centre of the nucleus. 3D distances between CCDC26 and either **A.** Chr.1p13.3 or **B.** Chr.16p11.2 loci were measured and mostly show a significant increase in KO cells compared to WT cells. Nuclei were stained with DAPI (blue). Measurements were calculated for 500 cells per replicate. Values represent the mean ± standard deviation (n=3). * $P<0.05$; ** $P<0.005$; *** $P<0.001$ (unpaired, two-tailed *t* test) (scale bar = 5um).

6.3 Discussion

A 2007 study by the Esteller group was one of the first to report that DNMT1 KD can impair nucleolar structure (Espada et al. 2007). DNMT1 KD cells displayed fragmented nucleoli, with AgNOR staining clearly showing increased numbers and more dispersed patterns (Espada et al. 2007). Consistent with this, KO of histone deacetylase, Sirtuin 7 (Sirt7), has also shown evidence for nucleolar fragmentation. Sirt7 has been shown to recruit DNMT1 to rDNA promoters, a mechanism severely impaired upon Sirt7 KO, and which ultimately results in disintegration of nucleoli (Ianni et al. 2017). *CCDC26* KO cells behave similarly to DNMT1 KD cells, given the mass re-localisation of DNMT1 to the cytosol. However, our results do not correlate strongly in this instance. Immunofluorescence and AgNOR staining techniques have shown some slight differences in nucleolar patterns between WT and KO cells, but not as pronounced as those observed in DNMT1 KD cells by Espada et al (2007). Nucleoli appear somewhat more diffuse in the KO cells, but marginally so. We also assessed nucleolar function by measuring rRNA levels by qRT-PCR, and in support of there being minimal changes to the nucleolus upon *CCDC26* KO, overall, we observed no significant difference between WT and KO cells. The similarity in levels of both 45S rRNA as well as its cleaved products indicates no significant change to transcription, nor the cleavage process. Previously, there have been mixed results with regards to DNMT1 deficiency, rDNA methylation and rDNA transcription. Some have found that DNMT1 KD causes demethylation of rDNA promoter regions, but that this appears to have no significant effect on transcription (Espada et al. 2007). This however, is contradictory to early

reports that rDNA transcription rates can be influenced by the methylation status of a single CpG, upon DNMT inhibitor treatment (Santoro and Grummt 2001).

The discrepancy between our results and previous DNMT1 KD data could be explained by additional, as yet unknown, effects of *CCDC26* KO in cells. Although we have hypothesised that our KO cells should behave similarly to DNMT1 KDs, we are still in the process of characterising these cells. In recent years, many lncRNAs have been proven to be multifunctional (Ji et al. 2019; Lo, Wolfson, and Zhou 2016; Arzenani et al. 2011). As compared to DNMT1 KD cells, in *CCDC26* KOs other potential interacting and impacting factors might be present. It is interesting however, that we see some similar nucleolar effects, but to a much smaller extent. It is possible that the amount of DNMT1 remaining in the nucleus is sufficient for maintaining nucleoli.

Here, we have also shown that *CCDC26* KO leads to large scale chromosomal rearrangements. The organisation of chromatin domains has been implicated in response to both DNMT1 and lncRNAs (Rego et al. 2008; Zhang, Huynh, and Lee 2007; Rao et al. 2014; Espada et al. 2004; Hacisuleyman et al. 2014). Electron microscopy has shown severe impairment of nuclear architecture in cells upon DNMT1 KD, consisting of unusual, deformed folds of the nuclear envelope. Immunolocalisation of heterochromatin too showed abnormal distribution compared to WT cells, indicating large-scale changes to chromosomal positioning and organisation (Espada et al. 2004). Comparatively, it has been well-documented that lncRNAs also have roles in nuclear organisation. Since the 1980's it has been

recognised that RNA can play a structural role in the nucleus, with early experiments demonstrating disintegration of the nuclear matrix upon nuclear RNA digestion (Bouvier et al. 1985). This role is particularly well-demonstrated by lncRNA, *Xist*, and its influence on X-chromosome positioning. In females, one of the X-chromosomes (Xi) is silenced in a process largely dependent on *Xist*. A critical aspect of Xi silencing involves its localisation to the nuclear lamina and nucleolar periphery. In the absence of *Xist*, this localisation is impaired and genes on the chromosome can be reactivated (Rego et al. 2008; Zhang, Huynh, and Lee 2007). In addition to re-positioning the chromosome as a whole, *Xist* also mediates folding of the Xi into a specific structure consisting of two 'mega-domains' upon Xi silencing (Rao et al. 2014).

Three-dimensional DNA FISH analysis demonstrated clustering of CCDC26, Chr.1p13.3 and Chr.16p11.2 loci at the nuclear periphery in WT cells; this is an arrangement which is lost in KO cells, when Chr.1p13.3 and Chr.16p11.2 loci move towards the centre of the nucleus (Figure 6.9). Interestingly, long stretches of genes at these regions on chromosomes 1 and 16 are also similarly down-regulated in KO cells. It is unusual to observe a correlation between gene repression and localisation at the nuclear centre, although not completely unheard of (Solovei et al. 2009). A role for *CCDC26* in regulating this kind of nuclear architecture is novel. One possible reason for this arrangement could be to provide an efficient means of co-regulating multiple genes. A similar occurrence has been observed for loci on chromosomes 2, 9, 15, 17 and the X chromosome. lncRNA *Firre* drives the assembly of these loci to co-regulate associated genes for adipogenesis (Hacisuleyman et al. 2014; Sun

et al. 2013; Maass et al. 2018). However, in this instance, DAVID annotations indicate no functional similarity between the down-regulated genes, reducing the likelihood that the genes are co-regulated in this manner for a single, particular biological process. It is also currently unclear, which aspect of CCDC26 could be driving this organisation and regulating genes at these loci. At this point, it is possible that the RNA, the gene or even the act of transcription at this site is important. For example, it is possible that the CCDC26 locus behaves as a *trans*-acting enhancer region. Previous studies have already suggested that the 8q24 locus, including CCDC26, comprises several transcriptional enhancers of c-Myc, with which long-range chromatin looping has been observed (Jia et al. 2009; Ahmadiyeh et al. 2010). Alternatively, transcription of *CCDC26* could be the driving force for recruiting transcriptional machinery to the gene locus, whereby a transcription factory may be generated, thus allowing efficient transcriptional regulation of genes at multiple loci. A similar process occurs at the β globin locus, where transcription of an intergenic ncRNA allows transcription of the β globin gene, by generating an open chromatin environment and guiding the transcriptional machinery to the site (Gribnau et al. 2000). Combining DNA FISH with RNA FISH, using a probe specific to *CCDC26* will help us determine whether the RNA itself is important for mediating this localisation.

It is also unclear at this point whether the global hypomethylation observed in KO cells is in any way causative of this re-organisation, or whether we have identified an additional role for *CCDC26* completely separate from its function in regulating DNMT1 subcellular localisation. It will be interesting to assess the methylation

status of the Chr.1p13.3 and Chr.16p11.2 regions in the future to further investigate this. It will also be important to investigate nuclear architecture and chromosome positioning on a global level. We have confirmed a change in the expression and localisation of two loci, but we do not yet know whether this is a targeted effect, or whether a *CCDC26* KO impacts nuclear organisation on a whole.

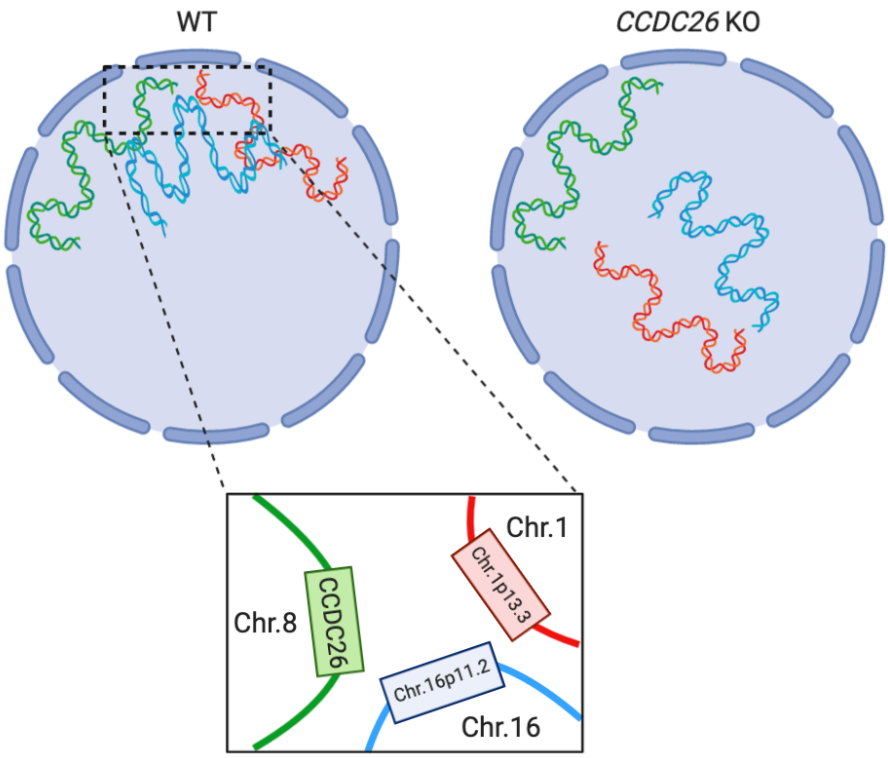


Figure 6.9 Model of 3D chromosomal rearrangements in WT and *CCDC26* KO cells. In WT K562 cells, CCDC26, Chr.1p13.3 and Chr.16p11.2 genomic loci cluster at the nuclear periphery. Upon *CCDC26* KO, genes located at Chr.1p13.3 and Chr.16p11.2 are down-regulated, and the loci move towards the centre of the nucleus, whilst CCDC26 remains at the nuclear periphery.

CHAPTER 7

GENERAL DISCUSSION

In the 60 years since the initial speculation that RNA molecules might function in gene regulation (Jacob and Monod 1961), lncRNAs have flourished from inconsequential by-products of transcription, to critical layers of regulation within gene regulatory networks and biological processes. Especially in the last three decades, following the development of whole-genome technologies, these molecules have gained importance (Hangauer, Vaughn, and McManus 2013). A number of pioneering transcriptomic studies emerged throughout the 2000's, demonstrating the scale of lncRNA abundance in mammalian cells, and consolidating them as a major component of the transcriptome. In 2002, almost 12,000 new ncRNAs were identified in mice (Okazaki et al. 2002). 10 years later, a comprehensive lncRNA database was established, having annotated a total of 14,880 human ncRNAs (Derrien et al. 2012). Demonstrating just how fast this field is moving, a 2015 study went further and identified approximately 58,000 lncRNA genes, almost 80% of which had never been annotated, via curation of thousands of previous RNA-seq datasets (Iyer et al. 2015). Of the lncRNAs identified, 13.5% were associated with cancer (Iyer et al. 2015). However, there is a huge inconsistency between the number of lncRNAs identified and the number that have been functionally characterised; a database for independently investigated, eukaryotic lncRNAs contained functional information for fewer than 300 transcripts (Quek et al. 2015). Research in this field must now extend beyond transcriptomic scale approaches, and focus on investigating the biological significance of these

components. Functional characterisation of lncRNAs has proven more difficult than proteins, largely due to our current inability to predict functionality from RNA sequence. Furthermore, for years efforts have been invested in methods of protein analysis, and consequently there is a dearth of biochemical and molecular methods for studying RNA (Leone and Santoro 2016). This means that investigation of independent lncRNAs, comprising detailed, mechanistic analysis is required. If we want to fully understand regulatory networks and the intricate dynamics of cellular processes, then better understanding of this critical layer of regulation is essential.

We identified *CCDC26* as an interesting lncRNA candidate for analysis from previous literature and genomic data. The chr.8q24 locus, from which *CCDC26* and several other lncRNAs are transcribed, has been well-established as a mutational hotspot, implicated in numerous forms of cancer (Ghoussaini et al. 2008; Thomas et al. 2008; Kuhn et al. 2012; Beroukhim et al. 2010). *CCDC26* is of particular interest given the frequency with which 8q24 mutations are located within or involve its gene (detailed in Chapter 1, 1.4.1) (Radtke et al. 2009; Kuhn et al. 2012; Chinen et al. 2014; Duployez et al. 2018; Izadifard et al. 2018). A few studies began functional analysis of *CCDC26*, however, none went beyond linking the lncRNA to cell viability and differentiation, and identifying a few potential downstream effector proteins (Yin et al. 2006; Hirano et al. 2008; Hirano et al. 2015; Wang, Hui, et al. 2018; Peng and Jiang 2016; Cao et al. 2018; Yan et al. 2019). In this work, I have carried out an in-depth, mechanistic analysis of *CCDC26* function in CML cells. In addition to confirming that *CCDC26* loss leads to reduced cell viability (Chapter 3), I have demonstrated a critical role for *CCDC26* in regulating global DNA methylation

(Chapters 4 and 5), as well as 3D chromosomal architecture (Chapter 6) (Figure 7.1).

7.1 *CCDC26* regulates global DNA methylation by controlling the subcellular localisation of DNMT1

Here, we have focussed our attention on the CML, K562 cell line, given the increased levels of *CCDC26* in these cells (Yin et al. 2006; Hirano et al. 2015) (Wijesinghe et al. unpublished, manuscript in preparation) (Figure 3.1, 3.2 and 3.3). Previously, most studies have reported that loss of *CCDC26* results in increased levels of apoptosis and reduced proliferation (Yin et al. 2006; Hirano et al. 2015; Peng and Jiang 2016; Wang, Hui, et al. 2018). In our initial characterisation of *CCDC26* in K562 cells, we confirmed this, demonstrating increased activation of caspases, and slower growth upon loss of *CCDC26*. This is the first time that such results have been confirmed in CRISPR KO cells (Chapter 3), as all prior analyses were conducted with KD cells (Wang, Hui, et al. 2018; Hirano et al. 2015; Peng and Jiang 2016; Cao et al. 2018; Yan et al. 2019). We also performed epigenetic characterisation. In a novel finding, the data presented here provides evidence that lncRNA, *CCDC26*, is associated/interacts with DNMT1 and regulates global DNA methylation by controlling the subcellular localisation of DNMT1. Expression of this transcript is important for localising DNMT1 to the nucleus, and in its absence, a significant proportion of the enzyme resides in the cytosol (Chapter 4). This result is coupled with a global loss of DNA methylation in KO cells (Chapter 3). Not only is this a novel finding for *CCDC26* specifically, but to our knowledge, this type of

functional mechanism has never been reported for any lncRNA with regards to DNMT1.

This is the first time that an interaction has been demonstrated between *CCDC26* and DNMT1. However, interactions between DNMT1 and other lncRNAs have been widely reported; 148 lncRNAs were found to associate with DNMT1 in colon cancer cells alone (Merry et al. 2015). These previously reported interactions however, have primarily demonstrated local effects on DNA methylation (Di Ruscio et al. 2013; Wang et al. 2015; Gao et al. 2019; Guo et al. 2019). For example, in HL60 cells, lncRNA, *ecCEBPA*, binds DNMT1 and prevents specific methylation of the CEBPA gene, resulting in increased expression of CEBPA mRNA (Di Ruscio et al. 2013). To our knowledge, there are very few lncRNAs that impact DNA methylation on a global scale. Two examples of such lncRNAs are *DACOR1* and *PARTICLE*. In colon cancer, reduced representation bisulfite sequencing (RRBS) showed that expression of *DACOR1* increases DNA methylation on a genome-wide scale (Somasundaram et al. 2018). The suspected mechanism in this instance however, involves increased available levels of methyl donor SAM. *DACOR1* has been shown to direct DNMT1-mediated methylation of the gene encoding cystathionine β synthase (CBS), resulting in gene repression. This leads to increased levels of methionine, and hence SAM (Merry et al. 2015; Somasundaram et al. 2018). Similarly, *PARTICLE* also enhances global DNA methylation levels, however, this occurs via a direct interaction with DNMT1, which subsequently increases overall methyltransferase activity (O'Leary et al. 2017). Importantly, none of these mechanisms involve re-localisation of DNMT1 similar to what we see with *CCDC26*.

Although DNA methylation has been studied for many years, new findings are adding complexity to our understanding of this epigenetic signature. Over the years, it has become clear that downstream effects of 5mC deposition are highly dependent on the genomic location. Whereas gene promoter methylation typically correlates with gene repression (Venolia and Gartler 1983; Stepper et al. 2017), methylation within the gene body is frequently associated with active transcription (Larsen, Solheim, and Prydz 1993; Hellman and Chess 2007). However, increasingly emerging studies are demonstrating that this is not so simplistic; for example, instances have been reported of transcription factors binding 5mC marks (Yin et al. 2017). Consequently, it is difficult to interpret the effects of global DNA hypomethylation in *CCDC26* KO cells without speculation. Going forward, it will be interesting to further analyse differential methylation patterns in the KO cells to help us understand the consequences.

This work has opened up a number of questions. From a clinical perspective, it would be interesting to compare DNA methylation levels in WT and KO cells with non-cancerous cells. Aberrant DNA methylation has previously been associated with K562 cells and CML progression. The Issa group analysed approximately 28,000 CpG sites, and found that 15% of these were hypermethylated in K562 cells (Jelinek et al. 2011). Given the trend of *CCDC26* being overexpressed in cancer cells (Tseng et al. 2014; Hirano et al. 2008; Wang, Hui, et al. 2018; Peng and Jiang 2016), it is possible that we have identified a novel cause of cancer-associated hypermethylation, via *CCDC26*-mediated prevention of DNMT1 exiting nuclei. It will be interesting to investigate whether reduced global methylation levels observed in

CCDC26 KO cells, are similar to levels in normal, non-cancerous cells. This means that targeting *CCDC26* could provide a new, alternative cancer therapy to DNMT inhibitor treatments such as Decitabine (Al-Ali, Jaekel, and Niederwieser 2014).

We are yet to establish the precise mechanism by which *CCDC26* regulates DNMT1 localisation, however, Chapter 5 rules out some possibilities including a HDAC-mediated mechanism and changes to DNMT1 stability. One possibility that requires investigation is that DNMT1 undergoes a structural change when bound to *CCDC26*, which impacts its localisation. Previous publications have demonstrated the importance of protein folding for both catalytic activity and localisation (Cummings et al. 1998; Almagro Armenteros et al. 2017; Scott, Thomas, and Hallett 2004). DNMT1 can be folded such that its activity is autoinhibited by its RFTS domain occupying the catalytic pocket (Syeda et al. 2011). Additionally, single amino acid substitutions near the NLS have resulted in DNMT1 misfolding and localisation to the cytosol where aggresomes are formed (Baets et al. 2015).

7.2 *CCDC26* expression is important in 3D nuclear architecture

For the first time, we also provide evidence that 3D nuclear architecture is implicated upon *CCDC26* KO, suggesting a potential structural role for this lncRNA. Both DNMT1 and lncRNAs, such as *Firre,* have demonstrated roles in chromosome positioning and organisation, making this an interesting feature to investigate (Espada et al. 2004; Hacisuleyman et al. 2014). We consulted previously analysed WT and *CCDC26* KO RNA-seq data to identify a region of interest, and identified long stretches of neighbouring, differentially expressed genes on chromosomes 1

and 16. Subsequent DNA FISH analyses using probes specific for these regions demonstrated large organisational changes. Whilst the CCDC26, chr.1p13.3 and chr.16p11.2 genetic loci appear clustered at the nuclear periphery in WT cells, they disperse in KO cells, with chr1.p13.3 and to some extent chr.16p11.2 loci moving in to the centre of the nucleus (Chapter 6). These data clearly suggest that some aspect of *CCDC26* expression is important in the structural organisation of chromosomes in the nucleus. Regulation of 3D chromosome organisation is a known function of lncRNAs. Previously, lncRNA *Linx*, has been shown to promote chromosomal looping by interacting with an enhancer, called Xite. The resulting chromosomal arrangement enables the lncRNA *Tsix* promoter to also interact with Xite, thus inducing *Tsix* expression (Nora et al. 2012). Furthermore, a chromosomal interaction between the CISTR-ACT and PTHLH genes on chromosome 12, and the SOX9 gene on chromosome 17 is implicated by lncRNA *CISTR-ACT* expression levels (Maass et al. 2012).

This observation has raised numerous new questions and a new investigational route for *CCDC26*. At present it is unclear whether this observation at all relates to the observed DNMT1 re-localisation. Establishing the methylation status of these chromosome 1 and 16 regions will help uncover this. Alternatively, we cannot rule out the possibility that this is an entirely new role for *CCDC26*, separate from its function in regulating DNA methylation, meaning that *CCDC26* is multifunctional. Other lncRNAs, such as *GAS5* for example, have shown evidence of multifunctionality. *GAS5* possesses roles in regulating histone modifications by increasing binding of EZH2 to transcription factors (Wang, Guo, Wang, et al. 2018),

regulating the immune response by preventing binding between glucocorticoid receptors and their substrates (Kino et al. 2010) as well as binding and preventing various miRNAs from reaching their targets (Guo et al. 2015; Ji et al. 2019).

Going forward it will also be interesting to establish the implications of this type of chromosomal rearrangement, and whether other chromosomes are also affected in a similar manner. *CCDC26* could be impacting chromosome arrangements on a large scale, or, the effect could be specific to chromosomes 1 and 16. If the latter is true, further work is required to investigate whether there is a purpose for co-regulation of these functionally unrelated, neighbouring genes (Chapter 6.2.2). Checking whether this chromosomal arrangement is present in other cells would determine its significance. One possibility is that these genes share regulatory elements, although this does not explain why these particular *trans* regions come together. There is a great deal of debate surrounding whether transcription levels drive chromosome positioning or vice versa (Osborne et al. 2004; Ragoczy et al. 2006). It is too early to say whether *CCDC26* is affecting transcription at these regions which subsequently results in co-localisation, or whether it is regulating the nuclear positioning of these loci, which then affects transcription.

7.3 Future Work and Final Conclusions

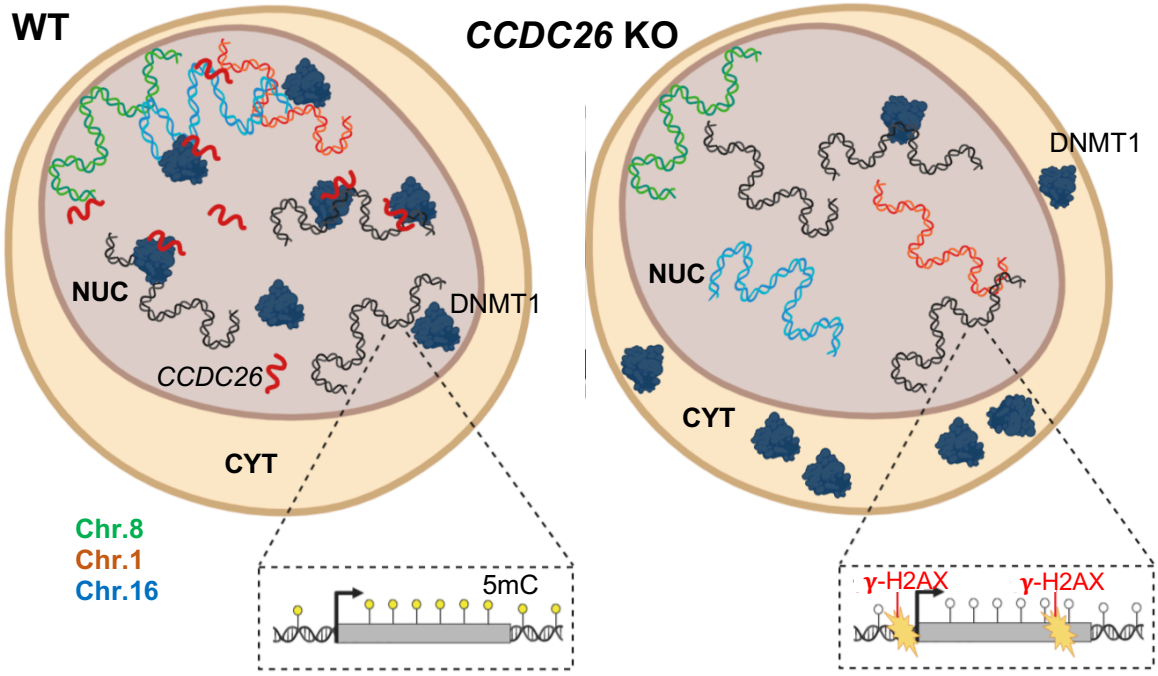
The work presented in this thesis has generated a number of as-yet-unanswered biological questions with regards to the function of *CCDC26*. It is next important to consolidate the findings presented here through further investigation. A critical, currently outstanding experiment is genome wide analysis of DNA methylation.

Techniques such as RRBS, a high-throughput, methylation profiling method, would demonstrate differentially methylated regions between WT and KO cells, subsequently allowing better interpretation of the effects of hypomethylation in the KOs. In addition, the methylation status of chr.1p13.3 and chr.16p11.2 might indicate whether DNMT1 re-localisation is involved in the observed change in nuclear architecture.

In order to establish the mechanism by which *CCDC26* regulates DNMT1 subcellular localisation, the next logical step is to assess the structure of DNMT1. As an initial approach, a technique such as circular dichroism could be applied to assess secondary structure. If it emerges that more detailed atomic information is required, X-ray crystallography could perhaps by applied. Additionally, it will be interesting to fully characterise the PTMs present on DNMT1 in WT and KO cells. We performed a preliminary DNMT1 IP to assess this (Figure 5.14), however, DNMT1 is so heavily modified (Figure 5.13) that a mass spectroscopy approach might be more applicable. This would allow detailed analysis of each PTM site individually.

Ultimately, the principal aims and objectives for this project have been achieved. In addition to characterising *CCDC26* KO cells, the data presented here also provides detailed, functional characterisation of this lncRNA for the first time. We have provided evidence of a striking, novel function for *CCDC26* and lncRNAs in general, in regulating global DNA methylation, as well as a potential, additional role in

coordinating nuclear architecture. Further work is required to fully comprehend the impact of *CCDC26* and determine its therapeutic potential.



- Increased apoptosis
- Slower growth
- Reduced global methylation levels

Figure 7.1. Model of the effects of *CCDC26* KO on K562 cells. Upon *CCDC26* KO (*CCDC26* RNA displayed in red) a number of changes occur within K562 cells. DNMT1 (dark blue) becomes primarily localised in the cytosol (CYT) rather than the nucleus (NUC), and consequently, global 5mC levels are significantly reduced. KO cells also demonstrate increased levels of DNA damage (γ-H2AX) and apoptosis, slower growth and nuclear, chromosomal rearrangements. Specific loci on chromosomes 1 (p13.3), 8 (*CCDC26* gene) and 16 (p11.2) reside in close proximity at the nuclear periphery in WT cells, however, upon *CCDC26* KO, the loci on chromosomes 1 and 16 relocate towards the nuclear center (Chr. 8 displayed in green; Chr. 1 displayed in orange; Chr. 16 displayed in blue). Further work will help determine whether these observations are all linked.

APPENDIX I

Figure 1. Optimisation of nick translation reactions for DNA FISH probe synthesis. DNA FISH probes were generated using the Invitrogen FISH Tag DNA Multicolour Kit (Section X). The protocol involved performing nick translation reactions on BACs to generate BAC DNA fragments. The length of time these reactions were allowed to proceed was optimised for different sized BACs in order to generate DNA fragments of length ~200-700bp. Time points varying from 15 minutes – 120 minutes were tested, and the resulting DNA was run on an agarose gel. For the RP11-770K21 BAC corresponding to the CCDC26 gene (length = 189,124bp), a 60-minute incubation was deemed optimal. For the RP4-735C1 BAC corresponding to the Chr.1p13.3 locus (length = 133,451bp), a 15-minute incubation time was deemed optimal (**1** = RP11-770K21 BAC undigested; **2** = RP4-735C1 BAC undigested).

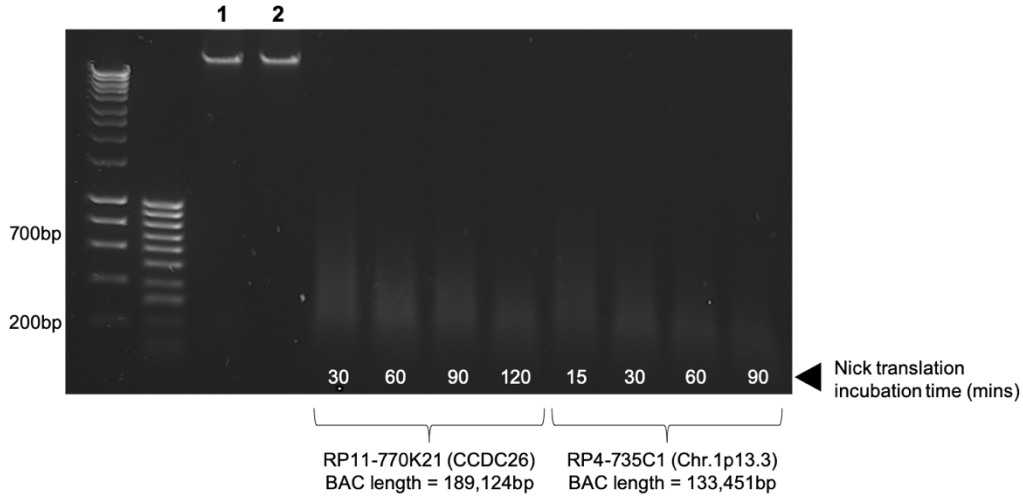
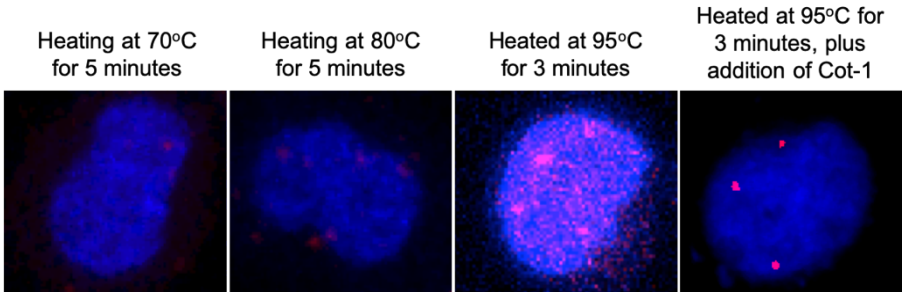
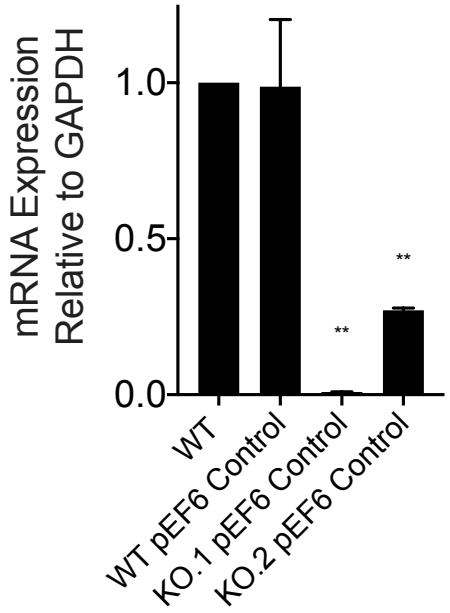


Figure 2. Optimisation of DNA FISH. Prior to overnight hybridisation, fixed cells and DNA FISH probes were heated at different temperatures. This was to denature DNA strands and probes to allow annealing to occur. Fluorescent signals could only be observed when cells and probes were heated to a temperature of 95°C. However, a large amount of non-specific fluorescent signal could be seen. This was improved by the addition of Cot-1 DNA to the hybridisation mixture. Maximum intensity projections of Z-stack images of K562 WT cells were captured using the CCDC26 probe (red). Nuclei were stained with DAPI (blue).



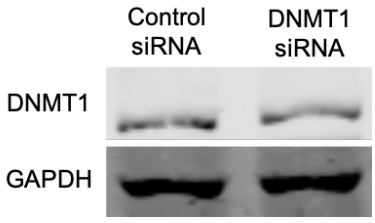
APPENDIX II

Figure 1. *CCDC26* expression in pEF6 control cells. *CCDC26* expression levels in cells transfected with empty pEF6 plasmids were measured. Levels were compared to non-transfected WT cells to ensure *CCDC26* remained knocked out. Values represent the mean ± standard deviation (n=3). * P<0.05; **P<0.01 (unpaired, two-tailed *t* test).



APPENDIX III

Figure 1. DNMT1 siRNA KD in K562 cells (48hrs). WT K562 cells were transiently transfected with 60pmol DNMT1 siRNA for 48hrs, after which total protein was extracted. Immunoblotting demonstrated that DNMT1 protein levels were not significantly different from the control, 48hrs post-transfection. Protein was measured relative to GAPDH. Values represent the mean ± standard deviation (n=3) (unpaired, two-tailed *t* test).



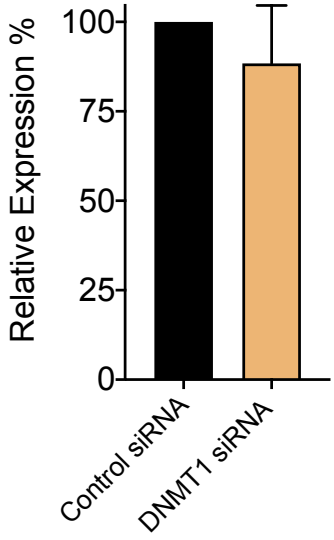


Figure 2. Cells treated with DAC for 24hrs show no significant change in global 5mC levels. Global 5mC immunofluorescence (red) levels are not significantly different 24hrs after treatment of K562 WT cells with both 5uM and 10uM DAC. Nuclei were stained with DAPI (blue) (scale bar = 50um).

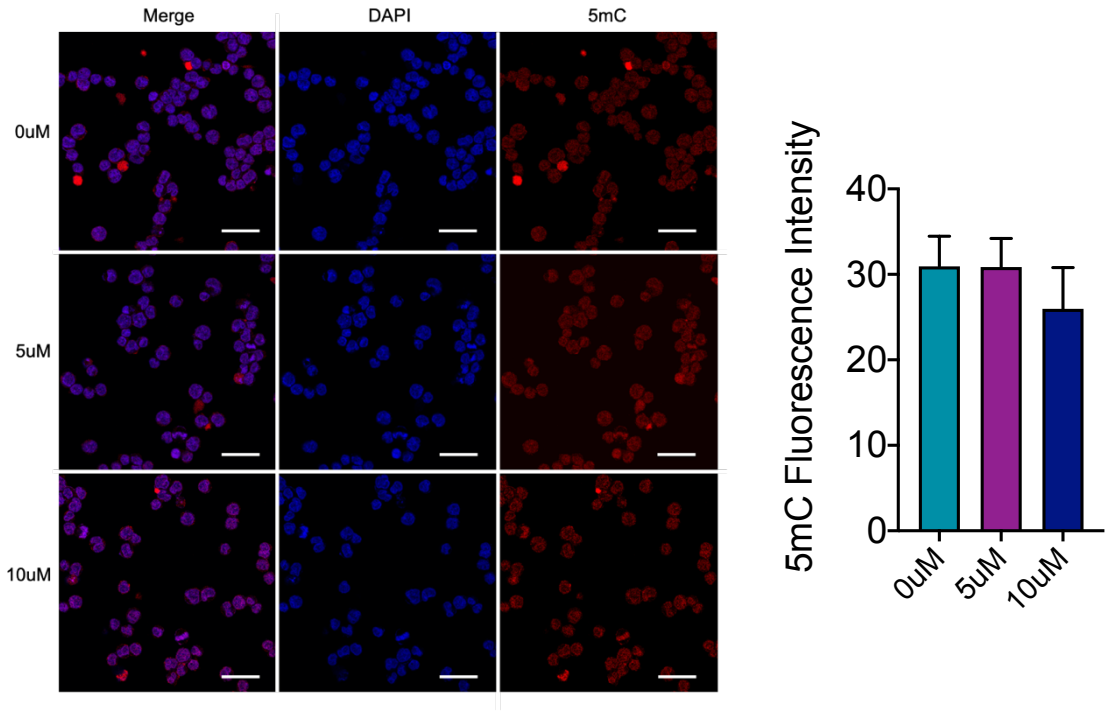
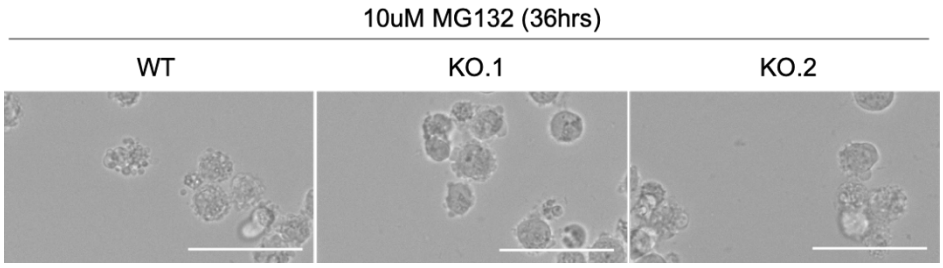


Figure 3. The majority of cells treated with 10uM MG132 for 36hrs were non-viable. Cells were imaged with brightfield microscopy (scale bar = 50um)



REFERENCES

Adel Fahmideh, M., C. Lavebratt, J. Schuz, M. Roosli, T. Tynes, M. A. Grotzer, C. Johansen, C. E. Kuehni, B. Lannering, M. Prochazka, L. S. Schmidt, and M. Feychting. 2015. 'CCDC26, CDKN2BAS, RTEL1 and TERT Polymorphisms in pediatric brain tumor susceptibility', *Carcinogenesis*, 36: 876-82.

Ahmadiyeh, N., M. M. Pomerantz, C. Grisanzio, P. Herman, L. Jia, V. Almendro, H. H. He, M. Brown, X. S. Liu, M. Davis, J. L. Caswell, C. A. Beckwith, A. Hills, L. Macconaill, G. A. Coetzee, M. M. Regan, and M. L. Freedman. 2010. '8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC', *Proc Natl Acad Sci U S A*, 107: 9742-6.

Al-Ali, H. K., N. Jaekel, and D. Niederwieser. 2014. 'The role of hypomethylating agents in the treatment of elderly patients with AML', *J Geriatr Oncol*, 5: 89-105.

Allan, J., T. Mitchell, N. Harborne, L. Bohm, and C. Crane-Robinson. 1986. 'Roles of H1 domains in determining higher order chromatin structure and H1 location', *J Mol Biol*, 187: 591-601.

Allfrey, V. G., R. Faulkner, and A. E. Mirsky. 1964. 'Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis', *Proc Natl Acad Sci U S A*, 51: 786-94.

Allis, C. D., and T. Jenuwein. 2016. 'The molecular hallmarks of epigenetic control', *Nat Rev Genet*, 17: 487-500.

Almagro Armenteros, J. J., C. K. Sonderby, S. K. Sonderby, H. Nielsen, and O. Winther. 2017. 'DeepLoc: prediction of protein subcellular localization using deep learning', *Bioinformatics*, 33: 3387-95.

Alvarez-Ponce, D., M. Torres-Sanchez, F. Feyertag, A. Kulkarni, and T. Nappi. 2018. 'Molecular evolution of DNMT1 in vertebrates: Duplications in marsupials followed by positive selection', *PLoS One*, 13: e0195162.

Amundadottir, L. T., P. Sulem, J. Gudmundsson, A. Helgason, A. Baker, B. A. Agnarsson, A. Sigurdsson, K. R. Benediktsdottir, J. B. Cazier, J. Sainz, M. Jakobsdottir, J. Kostic, D. N. Magnusdottir, S. Ghosh, K. Agnarsson, B. Birgisdottir, L. Le Roux, A. Olafsdottir, T. Blondal, M. Andresdottir, O. S. Gretarsdottir, J. T. Bergthorsson, D. Gudbjartsson, A. Gylfason, G. Thorleifsson, A. Manolescu, K. Kristjansson, G. Geirsson, H. Isaksson, J. Douglas, J. E. Johansson, K. Balter, F. Wiklund, J. E. Montie, X. Yu, B. K. Suarez, C. Ober, K. A. Cooney, H. Gronberg, W. J. Catalona, G. V. Einarsson, R. B. Barkardottir, J. R. Gulcher, A. Kong, U. Thorsteinsdottir, and K. Stefansson. 2006. 'A common variant associated with prostate cancer in European and African populations', *Nat Genet*, 38: 652-8.

Andersen, J. S., Y. W. Lam, A. K. Leung, S. E. Ong, C. E. Lyon, A. I. Lamond, and M. Mann. 2005. 'Nucleolar proteome dynamics', *Nature*, 433: 77-83.

Arzenani, M. K., A. E. Zade, Y. Ming, S. J. Vijverberg, Z. Zhang, Z. Khan, S. Sadique, L. Kallenbach, L. Hu, V. Vukojevic, and T. J. Ekstrom. 2011. 'Genomic DNA hypomethylation by histone deacetylase inhibition implicates DNMT1 nuclear dynamics', *Mol Cell Biol*, 31: 4119-28.

Ashwal-Fluss, R., M. Meyer, N. R. Pamudurti, A. Ivanov, O. Bartok, M. Hanan, N. Evantal, S. Memczak, N. Rajewsky, and S. Kadener. 2014. 'circRNA biogenesis competes with pre-mRNA splicing', *Mol Cell*, 56: 55-66.

Audas, T. E., M. D. Jacob, and S. Lee. 2012. 'Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA', *Mol Cell*, 45: 147-57.

Backes, C., T. Fehlmann, F. Kern, T. Kehl, H. P. Lenhof, E. Meese, and A. Keller. 2018. 'miRCarta: a central repository for collecting miRNA candidates', *Nucleic Acids Res*, 46: D160-D67.

Baets, J., X. Duan, Y. Wu, G. Smith, W. W. Seeley, I. Mademan, N. M. McGrath, N. C. Beadell, J. Khoury, M. V. Botuyan, G. Mer, G. A. Worrell, K. Hojo, J. DeLeon, M. Laura, Y. T. Liu, J. Senderek, J. Weis, P. Van den Bergh, S. L. Merrill, M. M. Reilly, H. Houlden, M. Grossman, S. S. Scherer, P. De Jonghe, P. J. Dyck, and C. J. Klein. 2015. 'Defects of mutant DNMT1 are linked to a spectrum of neurological disorders', *Brain*, 138: 845-61.

Bai, Y., S. Nie, G. Jiang, Y. Zhou, M. Zhou, Y. Zhao, S. Li, F. Wang, Q. Lv, Y. Huang, Q. Yang, Q. Li, Y. Li, Y. Xia, Y. Liu, J. Liu, J. Qian, B. Li, G. Wu, Y. Wu, B. Wang, X. Cheng, Y. Yang, T. Ke, H. Li, X. Ren, X. Ma, Y. Liao, C. Xu, X. Tu, and Q. K. Wang. 2014. 'Regulation of CARD8 expression by ANRIL and association of CARD8 single nucleotide polymorphism rs2043211 (p.C10X) with ischemic stroke', *Stroke*, 45: 383-8.

Bai, Y., W. Wang, Y. Zhang, F. Zhang, and H. Zhang. 2019. 'lncRNA MIAT suppression alleviates corneal angiogenesis through regulating miR-1246/ACE', *Cell Cycle*, 18: 661-69.

Bammert, L., S. Jonas, R. Ungricht, and U. Kutay. 2016. 'Human AATF/Che-1 forms a nucleolar protein complex with NGDN and NOL10 required for 40S ribosomal subunit synthesis', *Nucleic Acids Res*, 44: 9803-20.

Bannister, A. J., and T. Kouzarides. 2011. 'Regulation of chromatin by histone modifications', *Cell Res*, 21: 381-95.

Bao, X., H. Wu, X. Zhu, X. Guo, A. P. Hutchins, Z. Luo, H. Song, Y. Chen, K. Lai, M. Yin, L. Xu, L. Zhou, J. Chen, D. Wang, B. Qin, J. Frampton, H. F. Tse, D. Pei, H. Wang, B. Zhang, and M. A. Esteban. 2015. 'The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters', *Cell Res*, 25: 80-92.

Bashtrykov, P., G. Jankevicius, A. Smarandache, R. Z. Jurkowska, S. Ragozin, and A. Jeltsch. 2012. 'Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain', *Chem Biol*, 19: 572-8.

Baubec, T., D. F. Colombo, C. Wirbelauer, J. Schmidt, L. Burger, A. R. Krebs, A. Akalin, and D. Schubeler. 2015. 'Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation', *Nature*, 520: 243-7.

Behnia, R., B. Panic, J. R. Whyte, and S. Munro. 2004. 'Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p', *Nat Cell Biol*, 6: 405-13.

Berkyurek, A. C., I. Suetake, K. Arita, K. Takeshita, A. Nakagawa, M. Shirakawa, and S. Tajima. 2014. 'The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to

facilitate accession of the catalytic center to hemi-methylated DNA', *J Biol Chem*, 289: 379-86.

Beroukhim, R., C. H. Mermel, D. Porter, G. Wei, S. Raychaudhuri, J. Donovan, J. Barretina, J. S. Boehm, J. Dobson, M. Urashima, K. T. Mc Henry, R. M. Pinchback, A. H. Ligon, Y. J. Cho, L. Haery, H. Greulich, M. Reich, W. Winckler, M. S. Lawrence, B. A. Weir, K. E. Tanaka, D. Y. Chiang, A. J. Bass, A. Loo, C. Hoffman, J. Prensner, T. Liefeld, Q. Gao, D. Yecies, S. Signoretti, E. Maher, F. J. Kaye, H. Sasaki, J. E. Tepper, J. A. Fletcher, J. Tabernero, J. Baselga, M. S. Tsao, F. Demichelis, M. A. Rubin, P. A. Janne, M. J. Daly, C. Nucera, R. L. Levine, B. L. Ebert, S. Gabriel, A. K. Rustgi, C. R. Antonescu, M. Ladanyi, A. Letai, L. A. Garraway, M. Loda, D. G. Beer, L. D. True, A. Okamoto, S. L. Pomeroy, S. Singer, T. R. Golub, E. S. Lander, G. Getz, W. R. Sellers, and M. Meyerson. 2010. 'The landscape of somatic copy-number alteration across human cancers', *Nature*, 463: 899-905.

Bierhoff, H., M. A. Dammert, D. Brocks, S. Dambacher, G. Schotta, and I. Grummt. 2014. 'Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing', *Mol Cell*, 54: 675-82.

Bird, A., M. Taggart, M. Frommer, O. J. Miller, and D. Macleod. 1985. 'A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA', *Cell*, 40: 91-9.

Bjornsdottir, G., S. Benonisdottir, G. Sveinbjornsson, U. Styrkarsdottir, G. Thorleifsson, G. B. Walters, A. Bjornsson, I. H. Olafsson, E. Ulfarsson, A. Vikingsson, R. Hansdottir, K. O. Karlsson, T. Rafnar, I. Jonsdottir, M. L. Frigge, A. Kong, A. Oddsson, G. Masson, O. T. Magnusson, T. Gudbjartsson, H. Stefansson, P. Sulem, D. Gudbjartsson, U. Thorsteinsdottir, T. E. Thorgeirsson, and K. Stefansson. 2017. 'Sequence variant at 8q24.21 associates with sciatica caused by lumbar disc herniation', *Nat Commun*, 8: 14265.

Boeynaems, S., S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. Van Den Bosch, P. Tompa, and M. Fuxreiter. 2018. 'Protein Phase Separation: A New Phase in Cell Biology', *Trends Cell Biol*, 28: 420-35.

Boggs, B. A., P. Cheung, E. Heard, D. L. Spector, A. C. Chinault, and C. D. Allis. 2002. 'Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes', *Nat Genet*, 30: 73-6.

Bolognesi, B., N. Lorenzo Gotor, R. Dhar, D. Cirillo, M. Baldrighi, G. G. Tartaglia, and B. Lehner. 2016. 'A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression', *Cell Rep*, 16: 222-31.

Bostick, M., J. K. Kim, P. O. Esteve, A. Clark, S. Pradhan, and S. E. Jacobsen. 2007. 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science*, 317: 1760-4.

Bourc'his, D., G. L. Xu, C. S. Lin, B. Bollman, and T. H. Bestor. 2001. 'Dnmt3L and the establishment of maternal genomic imprints', *Science*, 294: 2536-9.

Bouvier, D., J. Hubert, A. P. Seve, and M. Bouteille. 1985. 'Nuclear RNA-associated proteins and their relationship to the nuclear matrix and related structures in HeLa cells', *Can J Biochem Cell Biol*, 63: 631-43.

Boyer, L. A., K. Plath, J. Zeitlinger, T. Brambrink, L. A. Medeiros, T. I. Lee, S. S. Levine, M. Wernig, A. Tajonar, M. K. Ray, G. W. Bell, A. P. Otte, M. Vidal, D. K. Gifford, R. A. Young, and R. Jaenisch. 2006. 'Polycomb complexes repress developmental regulators in murine embryonic stem cells', *Nature*, 441: 349-53.

Boyle, S., S. Gilchrist, J. M. Bridger, N. L. Mahy, J. A. Ellis, and W. A. Bickmore. 2001. 'The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells', *Hum Mol Genet*, 10: 211-9.

Bradbury, C. A., F. L. Khanim, R. Hayden, C. M. Bunce, D. A. White, M. T. Drayson, C. Craddock, and B. M. Turner. 2005. 'Histone deacetylases in acute myeloid leukaemia show a distinctive pattern of expression that changes selectively in response to deacetylase inhibitors', *Leukemia*, 19: 1751-9.

Bronner, C., M. Alhosin, A. Hamiche, and M. Mousli. 2019. 'Coordinated Dialogue between UHRF1 and DNMT1 to Ensure Faithful Inheritance of Methylated DNA Patterns', *Genes (Basel)*, 10.

Brown, C. J., A. Ballabio, J. L. Rupert, R. G. Lafreniere, M. Grompe, R. Tonlorenzi, and H. F. Willard. 1991. 'A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome', *Nature*, 349: 38-44.

Brown, C. J., B. D. Hendrich, J. L. Rupert, R. G. Lafreniere, Y. Xing, J. Lawrence, and H. F. Willard. 1992. 'The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus', *Cell*, 71: 527-42.

Bunch, H., H. Choe, J. Kim, D. S. Jo, S. Jeon, S. Lee, D. H. Cho, and K. Kang. 2019. 'P-TEFb Regulates Transcriptional Activation in Non-coding RNA Genes', *Front Genet*, 10: 342.

Bunch, H., B. P. Lawney, A. Burkholder, D. Ma, X. Zheng, S. Motola, D. C. Fargo, S. S. Levine, Y. E. Wang, and G. Hu. 2016. 'RNA polymerase II promoter-proximal pausing in mammalian long non-coding genes', *Genomics*, 108: 64-77.

Burd, C. E., W. R. Jeck, Y. Liu, H. K. Sanoff, Z. Wang, and N. E. Sharpless. 2010. 'Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk', *PLoS Genet*, 6: e1001233.

Burma, S., B. P. Chen, M. Murphy, A. Kurimasa, and D. J. Chen. 2001. 'ATM phosphorylates histone H2AX in response to DNA double-strand breaks', *J Biol Chem*, 276: 42462-7.

Calin, G. A., C. D. Dumitru, M. Shimizu, R. Bichi, S. Zupo, E. Noch, H. Aldler, S. Rattan, M. Keating, K. Rai, L. Rassenti, T. Kipps, M. Negrini, F. Bullrich, and C. M. Croce. 2002. 'Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia', *Proc Natl Acad Sci U S A*, 99: 15524-9.

Cao, K., M. Li, J. Miao, X. Lu, X. Kang, H. Zhu, S. Du, X. Li, Q. Zhang, W. Guan, Y. Dong, and X. Xia. 2018. 'CCDC26 knockdown enhances resistance of gastrointestinal stromal tumor cells to imatinib by interacting with c-KIT', *Am J Transl Res*, 10: 274-82.

Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang. 2002. 'Role of histone H3 lysine 27 methylation in Polycomb-group silencing', *Science*, 298: 1039-43.

Cardoso, M. C., and H. Leonhardt. 1999. 'DNA methyltransferase is actively retained in the cytoplasm during early development', *J Cell Biol*, 147: 25-32.

Cerase, A., A. Armaos, C. Neumayer, P. Avner, M. Guttman, and G. G. Tartaglia. 2019. 'Phase separation drives X-chromosome inactivation: a hypothesis', *Nat Struct Mol Biol*, 26: 331-34.

Cerase, A., D. Smeets, Y. A. Tang, M. Gdula, F. Kraus, M. Spivakov, B. Moindrot, M. Leleu, A. Tattermusch, J. Demmerle, T. B. Nesterova, C. Green, A. P. Otte, L. Schermelleh, and N. Brockdorff. 2014. 'Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy', *Proc Natl Acad Sci U S A*, 111: 2235-40.

Chalei, V., S. N. Sansom, L. Kong, S. Lee, J. F. Montiel, K. W. Vance, and C. P. Ponting. 2014. 'The long non-coding RNA Dali is an epigenetic regulator of neural differentiation', *Elife*, 3: e04530.

Chateauvieux, S., F. Morceau, M. Dicato, and M. Diederich. 2010. 'Molecular and therapeutic potential and toxicity of valproic acid', *J Biomed Biotechnol*, 2010.

Chen, C., P. Wang, W. Mo, Y. Zhang, W. Zhou, T. Deng, M. Zhou, X. Chen, S. Wang, and C. Wang. 2019. 'lncRNA-CCDC26, as a novel biomarker, predicts prognosis in acute myeloid leukemia', *Oncol Lett*, 18: 2203-11.

Chen, T., S. Hevi, F. Gay, N. Tsujimoto, T. He, B. Zhang, Y. Ueda, and E. Li. 2007. 'Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells', *Nat Genet*, 39: 391-6.

Chen, T., N. Tsujimoto, and E. Li. 2004. 'The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin', *Mol Cell Biol*, 24: 9048-58.

Chen, T., Y. Ueda, J. E. Dodge, Z. Wang, and E. Li. 2003. 'Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b', *Mol Cell Biol*, 23: 5594-605.

Chen, Z., S. Liu, L. Tian, M. Wu, F. Ai, W. Tang, L. Zhao, J. Ding, L. Zhang, and A. Tang. 2015. 'miR-124 and miR-506 inhibit colorectal cancer progression by targeting DNMT3B and DNMT1', *Oncotarget*, 6: 38139-50.

Cheng, D., J. Deng, B. Zhang, X. He, Z. Meng, G. Li, H. Ye, S. Zheng, L. Wei, X. Deng, R. Chen, and J. Zhou. 2018. 'LncRNA HOTAIR epigenetically suppresses miR-122 expression in hepatocellular carcinoma via DNA methylation', *EBioMedicine*, 36: 159-70.

Cheng, J. C., C. B. Yoo, D. J. Weisenberger, J. Chuang, C. Wozniak, G. Liang, V. E. Marquez, S. Greer, T. F. Orntoft, T. Thykjaer, and P. A. Jones. 2004. 'Preferential response of cancer cells to zebularine', *Cancer Cell*, 6: 151-8.

Chiang, P. K., R. K. Gordon, J. Tal, G. C. Zeng, B. P. Doctor, K. Pardhasaradhi, and P. P. McCann. 1996. 'S-Adenosylmethionine and methylation', *FASEB J*, 10: 471-80.

Chinen, Y., N. Sakamoto, H. Nagoshi, T. Taki, S. Maegawa, S. Tatekawa, T. Tsukamoto, S. Mizutani, Y. Shimura, M. Yamamoto-Sugitani, T. Kobayashi, Y. Matsumoto, S. Horiike, J. Kuroda, and M. Taniwaki. 2014. '8q24 amplified segments involve novel fusion genes between NSMCE2 and long noncoding RNAs in acute myelogenous leukemia', *J Hematol Oncol*, 7: 68.

Choudhuri, S. 2010. 'Small noncoding RNAs: biogenesis, function, and emerging significance in toxicology', *J Biochem Mol Toxicol*, 24: 195-216.

Chowdhury, B., A. McGovern, Y. Cui, S. R. Choudhury, I. H. Cho, B. Cooper, T. Chevassut, A. C. Lossie, and J. Irudayaraj. 2015. 'The hypomethylating agent Decitabine

causes a paradoxical increase in 5-hydroxymethylcytosine in human leukemia cells', *Sci Rep*, 5: 9281.
Christman, J. K. 2002. '5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy', *Oncogene*, 21: 5483-95.
Chu, C., Q. C. Zhang, S. T. da Rocha, R. A. Flynn, M. Bharadwaj, J. M. Calabrese, T. Magnuson, E. Heard, and H. Y. Chang. 2015. 'Systematic discovery of Xist RNA binding proteins', *Cell*, 161: 404-16.
Clement, J. Q., S. Maiti, and M. F. Wilkinson. 2001. 'Localization and stability of introns spliced from the Pem homeobox gene', *J Biol Chem*, 276: 16919-30.
Clement, J. Q., L. Qian, N. Kaplinsky, and M. F. Wilkinson. 1999. 'The stability and fate of a spliced intron from vertebrate cells', *RNA*, 5: 206-20.
Clemson, C. M., J. N. Hutchinson, S. A. Sara, A. W. Ensminger, A. H. Fox, A. Chess, and J. B. Lawrence. 2009. 'An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles', *Mol Cell*, 33: 717-26.
Clemson, C. M., J. A. McNeil, H. F. Willard, and J. B. Lawrence. 1996. 'XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure', *J Cell Biol*, 132: 259-75.
Colgan, D. F., and J. L. Manley. 1997. 'Mechanism and regulation of mRNA polyadenylation', *Genes Dev*, 11: 2755-66.
Consortium, Encode Project. 2012. 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489: 57-74.
Cremer, T., and M. Cremer. 2010. 'Chromosome territories', *Cold Spring Harb Perspect Biol*, 2: a003889.
Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. 2010. 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proc Natl Acad Sci U S A*, 107: 21931-6.
Croft, J. A., J. M. Bridger, S. Boyle, P. Perry, P. Teague, and W. A. Bickmore. 1999. 'Differences in the localization and morphology of chromosomes in the human nucleus', *J Cell Biol*, 145: 1119-31.
Cross, S. H., J. A. Charlton, X. Nan, and A. P. Bird. 1994. 'Purification of CpG islands using a methylated DNA binding column', *Nat Genet*, 6: 236-44.
Csankovszki, G., A. Nagy, and R. Jaenisch. 2001. 'Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation', *J Cell Biol*, 153: 773-84.
Cui, T. 2015. 'CCDC26 rs4295627 polymorphism and glioma risk: a meta-analysis', *Int J Clin Exp Med*, 8: 3862-8.
Cummings, C. J., M. A. Mancini, B. Antalffy, D. B. DeFranco, H. T. Orr, and H. Y. Zoghbi. 1998. 'Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1', *Nat Genet*, 19: 148-54.
Curradi, M., A. Izzo, G. Badaracco, and N. Landsberger. 2002. 'Molecular mechanisms of gene silencing mediated by DNA methylation', *Mol Cell Biol*, 22: 3157-73.
Daskalos, A., G. Nikolaidis, G. Xinarianos, P. Savvari, A. Cassidy, R. Zakopoulou, A. Kotsinas, V. Gorgoulis, J. K. Field, and T. Liloglou. 2009. 'Hypomethylation of

retrotransposable elements correlates with genomic instability in non-small cell lung cancer', *Int J Cancer*, 124: 81-7.

Davey, C. A., D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond. 2002. 'Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution', *J Mol Biol*, 319: 1097-113.

De Santa, F., I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B. K. Tusi, H. Muller, J. Ragoussis, C. L. Wei, and G. Natoli. 2010. 'A large fraction of extragenic RNA pol II transcription sites overlap enhancers', *PLoS Biol*, 8: e1000384.

Deaton, A. M., and A. Bird. 2011. 'CpG islands and the regulation of transcription', *Genes Dev*, 25: 1010-22.

Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow, and R. Guigo. 2012. 'The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression', *Genome Res*, 22: 1775-89.

Desplats, P., B. Spencer, E. Coffee, P. Patel, S. Michael, C. Patrick, A. Adame, E. Rockenstein, and E. Masliah. 2011. 'Alpha-synuclein sequesters Dnmt1 from the nucleus: a novel mechanism for epigenetic alterations in Lewy body diseases', *J Biol Chem*, 286: 9031-7.

Di Ruscio, A., A. K. Ebralidze, T. Benoukraf, G. Amabile, L. A. Goff, J. Terragni, M. E. Figueroa, L. L. De Figueiredo Pontes, M. Alberich-Jorda, P. Zhang, M. Wu, F. D'Alo, A. Melnick, G. Leone, K. K. Ebralidze, S. Pradhan, J. L. Rinn, and D. G. Tenen. 2013. 'DNMT1-interacting RNAs block gene-specific DNA methylation', *Nature*, 503: 371-6.

Diaz-Lagares, A., A. B. Crujeiras, P. Lopez-Serra, M. Soler, F. Setien, A. Goyal, J. Sandoval, Y. Hashimoto, A. Martinez-Cardus, A. Gomez, H. Heyn, C. Moutinho, J. Espada, A. Vidal, M. Paules, M. Galan, N. Sala, Y. Akiyama, M. Martinez-Iniesta, L. Farre, A. Villanueva, M. Gross, S. Diederichs, S. Guil, and M. Esteller. 2016. 'Epigenetic inactivation of the p53-induced long noncoding RNA TP53 target 1 in human cancer', *Proc Natl Acad Sci U S A*, 113: E7535-E44.

Dixon, J. R., I. Jung, S. Selvaraj, Y. Shen, J. E. Antosiewicz-Bourget, A. Y. Lee, Z. Ye, A. Kim, N. Rajagopal, W. Xie, Y. Diao, J. Liang, H. Zhao, V. V. Lobanenkov, J. R. Ecker, J. A. Thomson, and B. Ren. 2015. 'Chromatin architecture reorganization during stem cell differentiation', *Nature*, 518: 331-6.

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: 376-80.

Djebali, S., C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Roder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R.

Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigo, and T. R. Gingeras. 2012. 'Landscape of transcription in human cells', *Nature*, 489: 101-8.

Doherty, A. S., M. S. Bartolomei, and R. M. Schultz. 2002. 'Regulation of stage-specific nuclear translocation of Dnmt1o during preimplantation mouse development', *Dev Biol*, 242: 255-66.

Dong, A., J. A. Yoder, X. Zhang, L. Zhou, T. H. Bestor, and X. Cheng. 2001. 'Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA', *Nucleic Acids Res*, 29: 439-48.

Du, Z., J. Song, Y. Wang, Y. Zhao, K. Guda, S. Yang, H. Y. Kao, Y. Xu, J. Willis, S. D. Markowitz, D. Sedwick, R. M. Ewing, and Z. Wang. 2010. 'DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination', *Sci Signal*, 3: ra80.

Duan, G., and D. Walther. 2015. 'The roles of post-translational modifications in the context of protein interaction networks', *PLoS Comput Biol*, 11: e1004049.

Duployez, N., E. Boudry-Labis, C. Roumier, N. Boissel, A. Petit, S. Geffroy, N. Helevaut, K. Celli-Lebras, C. Terre, O. Fenneteau, W. Cuccuini, I. Luquet, H. Lapillonne, C. Lacombe, P. Cornillet, N. Ifrah, H. Dombret, G. Leverger, E. Jourdan, and C. Preudhomme. 2018. 'SNP-array lesions in core binding factor acute myeloid leukemia', *Oncotarget*, 9: 6478-89.

Durocher, D., and S. P. Jackson. 2001. 'DNA-PK, ATM and ATR as sensors of DNA damage: variations on a theme?', *Curr Opin Cell Biol*, 13: 225-31.

Duymich, C. E., J. Charlet, X. Yang, P. A. Jones, and G. Liang. 2016. 'DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells', *Nat Commun*, 7: 11453.

Eden, A., F. Gaudet, A. Waghmare, and R. Jaenisch. 2003. 'Chromosomal instability and tumors promoted by DNA hypomethylation', *Science*, 300: 455.

Ehrlich, M., M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, and C. W. Gehrke. 1985. 'DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine', *Nucleic Acids Res*, 13: 1399-412.

Ehrlich, M., G. G. Wilson, K. C. Kuo, and C. W. Gehrke. 1987. 'N4-methylcytosine as a minor base in bacterial DNA', *J Bacteriol*, 169: 939-43.

Elledge, S. J. 1996. 'Cell cycle checkpoints: preventing an identity crisis', *Science*, 274: 1664-72.

Engreitz, J. M., J. E. Haines, E. M. Perez, G. Munson, J. Chen, M. Kane, P. E. McDonel, M. Guttman, and E. S. Lander. 2016. 'Local regulation of gene expression by lncRNA promoters, transcription and splicing', *Nature*, 539: 452-55.

Engreitz, J. M., A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath, and M. Guttman. 2013. 'The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome', *Science*, 341: 1237973.

Eser, U., D. Chandler-Brown, F. Ay, A. F. Straight, Z. Duan, W. S. Noble, and J. M. Skotheim. 2017. 'Form and function of topologically associating genomic domains in budding yeast', *Proc Natl Acad Sci U S A*, 114: E3061-E70.

Espada, J., E. Ballestar, M. F. Fraga, A. Villar-Garea, A. Juarranz, J. C. Stockert, K. D. Robertson, F. Fuks, and M. Esteller. 2004. 'Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern', *J Biol Chem*, 279: 37175-84.

Espada, J., E. Ballestar, R. Santoro, M. F. Fraga, A. Villar-Garea, A. Nemeth, L. Lopez-Serra, S. Ropero, A. Aranda, H. Orozco, V. Moreno, A. Juarranz, J. C. Stockert, G. Langst, I. Grummt, W. Bickmore, and M. Esteller. 2007. 'Epigenetic disruption of ribosomal RNA genes and nucleolar architecture in DNA methyltransferase 1 (Dnmt1) deficient cells', *Nucleic Acids Res*, 35: 2191-8.

Esteve, P. O., Y. Chang, M. Samaranayake, A. K. Upadhyay, J. R. Horton, G. R. Feehery, X. Cheng, and S. Pradhan. 2011. 'A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability', *Nat Struct Mol Biol*, 18: 42-8.

Esteve, P. O., H. G. Chin, J. Benner, G. R. Feehery, M. Samaranayake, G. A. Horwitz, S. E. Jacobsen, and S. Pradhan. 2009. 'Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells', *Proc Natl Acad Sci U S A*, 106: 5076-81.

Esteve, P. O., H. G. Chin, A. Smallwood, G. R. Feehery, O. Gangisetty, A. R. Karpf, M. F. Carey, and S. Pradhan. 2006. 'Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication', *Genes Dev*, 20: 3089-103.

Fang, J., J. Cheng, J. Wang, Q. Zhang, M. Liu, R. Gong, P. Wang, X. Zhang, Y. Feng, W. Lan, Z. Gong, C. Tang, J. Wong, H. Yang, C. Cao, and Y. Xu. 2016. 'Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition', *Nat Commun*, 7: 11197.

Farooq, Z., S. Banday, T. K. Pandita, and M. Altaf. 2016. 'The many faces of histone H3K79 methylation', *Mutat Res Rev Mutat Res*, 768: 46-52.

Fegan, C. D., D. White, and M. Sweeney. 1995. 'C-myc amplification, double minutes and homogenous staining regions in a case of AML', *Br J Haematol*, 90: 486-8.

Felsenfeld, G., and M. Groudine. 2003. 'Controlling the double helix', *Nature*, 421: 448-53.

Fernandez-Zapico, M. E., N. C. Gonzalez-Paz, E. Weiss, D. N. Savoy, J. R. Molina, R. Fonseca, T. C. Smyrk, S. T. Chari, R. Urrutia, and D. D. Billadeau. 2005. 'Ectopic expression of VAV1 reveals an unexpected role in pancreatic cancer tumorigenesis', *Cancer Cell*, 7: 39-49.

Finch, J. T., and A. Klug. 1976. 'Solenoidal model for superstructure in chromatin', *Proc Natl Acad Sci U S A*, 73: 1897-901.

Finnin, M. S., J. R. Donigian, A. Cohen, V. M. Richon, R. A. Rifkind, P. A. Marks, R. Breslow, and N. P. Pavletich. 1999. 'Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors', *Nature*, 401: 188-93.

Fiskus, W., K. Buckley, R. Rao, A. Mandawat, Y. Yang, R. Joshi, Y. Wang, R. Balusu, J. Chen, S. Koul, A. Joshi, S. Upadhyay, P. Atadja, and K. N. Bhalla. 2009. 'Panobinostat treatment depletes EZH2 and DNMT1 levels and enhances decitabine mediated

de-repression of JunB and loss of survival of human acute leukemia cells', *Cancer Biol Ther*, 8: 939-50.
Forte, G. M., M. R. Pool, and C. J. Stirling. 2011. 'N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum', *PLoS Biol*, 9: e1001073.
Foster, B. M., P. Stolz, C. B. Mulholland, A. Montoya, H. Kramer, S. Bultmann, and T. Bartke. 2018. 'Critical Role of the UBL Domain in Stimulating the E3 Ubiquitin Ligase Activity of UHRF1 toward Chromatin', *Mol Cell*, 72: 739-52 e9.
Fox, A. H., C. S. Bond, and A. I. Lamond. 2005. 'P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner', *Mol Biol Cell*, 16: 5304-15.
Fuks, F., P. J. Hurd, R. Deplus, and T. Kouzarides. 2003. 'The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase', *Nucleic Acids Res*, 31: 2305-12.
Fussner, E., M. Strauss, U. Djuric, R. Li, K. Ahmed, M. Hart, J. Ellis, and D. P. Bazett-Jones. 2012. 'Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres', *EMBO Rep*, 13: 992-6.
Gao, J., C. Dai, X. Yu, X. B. Yin, W. J. Liao, Y. Huang, and F. Zhou. 2019. 'Silencing of long non-coding RNA FOXD2-AS1 inhibits the progression of gallbladder cancer by mediating methylation-dependent induction of MLH1', *Am J Physiol Gastrointest Liver Physiol*.
Gary, J. D., and S. Clarke. 1998. 'RNA and protein interactions modulated by protein arginine methylation', *Prog Nucleic Acid Res Mol Biol*, 61: 65-131.
Gates, L. A., J. Shi, A. D. Rohira, Q. Feng, B. Zhu, M. T. Bedford, C. A. Sagum, S. Y. Jung, J. Qin, M. J. Tsai, S. Y. Tsai, W. Li, C. E. Foulds, and B. W. O'Malley. 2017. 'Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation', *J Biol Chem*, 292: 14456-72.
Gaudet, F., J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt, and R. Jaenisch. 2003. 'Induction of tumors in mice by genomic hypomethylation', *Science*, 300: 489-92.
Ghoshal, K., J. Datta, S. Majumder, S. Bai, H. Kutay, T. Motiwala, and S. T. Jacob. 2005. '5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal', *Mol Cell Biol*, 25: 4727-41.
———. 2018. 'Correction for Ghoshal et al., "5-Aza-Deoxycytidine Induces Selective Degradation of DNA Methyltransferase 1 by a Proteasomal Pathway That Requires the KEN Box, Bromo-Adjacent Homology Domain, and Nuclear Localization Signal"', *Mol Cell Biol*, 38.
Ghoussaini, M., H. Song, T. Koessler, A. A. Al Olama, Z. Kote-Jarai, K. E. Driver, K. A. Pooley, S. J. Ramus, S. K. Kjaer, E. Hogdall, R. A. DiCioccio, A. S. Whittemore, S. A. Gayther, G. G. Giles, M. Guy, S. M. Edwards, J. Morrison, J. L. Donovan, F. C. Hamdy, D. P. Dearnaley, A. T. Ardern-Jones, A. L. Hall, L. T. O'Brien, B. N. Gehr-Swain, R. A. Wilkinson, P. M. Brown, J. L. Hopper, D. E. Neal, P. D. Pharoah, B. A. Ponder, R. A. Eeles, D. F. Easton, A. M. Dunning, U. K. Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, and U. K. ProtecT Study Collaborators. 2008. 'Multiple loci with

different cancer specificities within the 8q24 gene desert', *J Natl Cancer Inst*, 100: 962-6.
Gilbert, N., S. Boyle, H. Fiegler, K. Woodfine, N. P. Carter, and W. A. Bickmore. 2004. 'Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers', *Cell*, 118: 555-66.
Gilbert, W. 1978. 'Why genes in pieces?', *Nature*, 271: 501.
Ginisty, H., H. Sicard, B. Roger, and P. Bouvet. 1999. 'Structure and functions of nucleolin', *J Cell Sci*, 112 (Pt 6): 761-72.
Goll, M. G., F. Kirpekar, K. A. Maggert, J. A. Yoder, C. L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, and T. H. Bestor. 2006. 'Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2', *Science*, 311: 395-8.
Gonzalez-Castro, T. B., I. E. Juarez-Rojop, M. L. Lopez-Narvaez, C. A. Tovilla-Zarate, A. D. Genis-Mendoza, N. Perez-Hernandez, J. J. Martinez-Magana, and J. M. Rodriguez-Perez. 2019. 'Genetic Polymorphisms of CCDC26 rs891835, rs6470745, and rs55705857 in Glioma Risk: A Systematic Review and Meta-analysis', *Biochem Genet*, 57: 583-605.
Goodsell, D. S. 2006. 'The molecular perspective: Cisplatin', *Stem Cells*, 24: 514-5.
Gowher, H., K. Liebert, A. Hermann, G. Xu, and A. Jeltsch. 2005. 'Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L', *J Biol Chem*, 280: 13341-8.
Gribnau, J., K. Diderich, S. Pruzina, R. Calzolari, and P. Fraser. 2000. 'Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus', *Mol Cell*, 5: 377-86.
Griffiths-Jones, S., H. K. Saini, S. van Dongen, and A. J. Enright. 2008. 'miRBase: tools for microRNA genomics', *Nucleic Acids Res*, 36: D154-8.
Grzenda, A., G. Lomberk, P. Svingen, A. Mathison, E. Calvo, J. Iovanna, Y. Xiong, W. Faubion, and R. Urrutia. 2013. 'Functional characterization of EZH2beta reveals the increased complexity of EZH2 isoforms involved in the regulation of mammalian gene expression', *Epigenetics Chromatin*, 6: 3.
Gudmundsson, J., P. Sulem, A. Manolescu, L. T. Amundadottir, D. Gudbjartsson, A. Helgason, T. Rafnar, J. T. Bergthorsson, B. A. Agnarsson, A. Baker, A. Sigurdsson, K. R. Benediktsdottir, M. Jakobsdottir, J. Xu, T. Blondal, J. Kostic, J. Sun, S. Ghosh, S. N. Stacey, M. Mouy, J. Saemundsdottir, V. M. Backman, K. Kristjansson, A. Tres, A. W. Partin, M. T. Albers-Akkers, J. Godino-Ivan Marcos, P. C. Walsh, D. W. Swinkels, S. Navarrete, S. D. Isaacs, K. K. Aben, T. Graif, J. Cashy, M. Ruiz-Echarri, K. E. Wiley, B. K. Suarez, J. A. Witjes, M. Frigge, C. Ober, E. Jonsson, G. V. Einarsson, J. I. Mayordomo, L. A. Kiemeney, W. B. Isaacs, W. J. Catalona, R. B. Barkardottir, J. R. Gulcher, U. Thorsteinsdottir, A. Kong, and K. Stefansson. 2007. 'Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24', *Nat Genet*, 39: 631-7.
Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel. 2008. 'Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions', *Nature*, 453: 948-51.

Guo, C., W. Q. Song, P. Sun, L. Jin, and H. Y. Dai. 2015. 'LncRNA-GAS5 induces PTEN expression through inhibiting miR-103 in endometrial cancer cells', *J Biomed Sci*, 22: 100.

Guo, X., Z. Chen, L. Zhao, D. Cheng, W. Song, and X. Zhang. 2019. 'Long non-coding RNA-HAGLR suppressed tumor growth of lung adenocarcinoma through epigenetically silencing E2F1', *Exp Cell Res*, 382: 111461.

Guttman, M., I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, and E. S. Lander. 2009. 'Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals', *Nature*, 458: 223-7.

Hacisuleyman, E., L. A. Goff, C. Trapnell, A. Williams, J. Henao-Mejia, L. Sun, P. McClanahan, D. G. Hendrickson, M. Sauvageau, D. R. Kelley, M. Morse, J. Engreitz, E. S. Lander, M. Guttman, H. F. Lodish, R. Flavell, A. Raj, and J. L. Rinn. 2014. 'Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre', *Nat Struct Mol Biol*, 21: 198-206.

Haiman, C. A., N. Patterson, M. L. Freedman, S. R. Myers, M. C. Pike, A. Waliszewska, J. Neubauer, A. Tandon, C. Schirmer, G. J. McDonald, S. C. Greenway, D. O. Stram, L. Le Marchand, L. N. Kolonel, M. Frasco, D. Wong, L. C. Pooler, K. Ardlie, I. Oakley-Girvan, A. S. Whittemore, K. A. Cooney, E. M. John, S. A. Ingles, D. Altshuler, B. E. Henderson, and D. Reich. 2007. 'Multiple regions within 8q24 independently affect risk for prostate cancer', *Nat Genet*, 39: 638-44.

Hangauer, M. J., I. W. Vaughn, and M. T. McManus. 2013. 'Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs', *PLoS Genet*, 9: e1003569.

Harmston, N., E. Ing-Simmons, G. Tan, M. Perry, M. Merkenschlager, and B. Lenhard. 2017. 'Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation', *Nat Commun*, 8: 441.

Hasegawa, Y., N. Brockdorff, S. Kawano, K. Tsutui, K. Tsutui, and S. Nakagawa. 2010. 'The matrix protein hnRNP U is required for chromosomal localization of Xist RNA', *Dev Cell*, 19: 469-76.

He, R. Z., D. X. Luo, and Y. Y. Mo. 2019. 'Emerging roles of lncRNAs in the post-transcriptional regulation in cancer', *Genes Dis*, 6: 6-15.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren. 2009. 'Histone modifications at human enhancers reflect global cell-type-specific gene expression', *Nature*, 459: 108-12.

Heller, G., T. Topakian, C. Altenberger, S. Cerny-Reiterer, S. Herndlhofer, B. Ziegler, P. Datlinger, K. Byrgazov, C. Bock, C. Mannhalter, G. Hormann, W. R. Sperr, T. Lion, C. C. Zielinski, P. Valent, and S. Zochbauer-Muller. 2016. 'Next-generation sequencing identifies major DNA methylation changes during progression of Ph+ chronic myeloid leukemia', *Leukemia*, 30: 1861-8.

Hellman, A., and A. Chess. 2007. 'Gene body-specific methylation on the active X chromosome', *Science*, 315: 1141-3.

Henderson, A. S., D. Warburton, and K. C. Atwood. 1972. 'Location of ribosomal DNA in the human chromosome complement', *Proc Natl Acad Sci U S A*, 69: 3394-8.
Hendrickson, G., D. R. Kelley, D. Tenen, B. Bernstein, and J. L. Rinn. 2016. 'Widespread RNA binding by chromatin-associated proteins', *Genome Biol*, 17: 28.
Hennig, S., G. Kong, T. Mannen, A. Sadowska, S. Kobelke, A. Blythe, G. J. Knott, K. S. Iyer, D. Ho, E. A. Newcombe, K. Hosoki, N. Goshima, T. Kawaguchi, D. Hatters, L. Trinkle-Mulcahy, T. Hirose, C. S. Bond, and A. H. Fox. 2015. 'Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles', *J Cell Biol*, 210: 529-39.
Henras, A. K., C. Plisson-Chastang, M. F. O'Donohue, A. Chakraborty, and P. E. Gleizes. 2015. 'An overview of pre-ribosomal RNA processing in eukaryotes', *Wiley Interdiscip Rev RNA*, 6: 225-42.
Herman, J. G., J. Jen, A. Merlo, and S. B. Baylin. 1996. 'Hypermethylation-associated inactivation indicates a tumor suppressor role for p15INK4B', *Cancer Res*, 56: 722-7.
Hirano, T., F. Ike, T. Murata, Y. Obata, H. Utiyama, and K. K. Yokoyama. 2008. 'Genes encoded within 8q24 on the amplicon of a large extrachromosomal element are selectively repressed during the terminal differentiation of HL-60 cells', *Mutat Res*, 640: 97-106.
Hirano, T., R. Yoshikawa, H. Harada, Y. Harada, A. Ishida, and T. Yamazaki. 2015. 'Long noncoding RNA, CCDC26, controls myeloid leukemia cell growth through regulation of KIT expression', *Mol Cancer*, 14: 90.
Hodge, D. R., E. Cho, T. D. Copeland, T. Guszczynski, E. Yang, A. K. Seth, and W. L. Farrar. 2007. 'IL-6 enhances the nuclear translocation of DNA cytosine-5-methyltransferase 1 (DNMT1) via phosphorylation of the nuclear localization sequence by the AKT kinase', *Cancer Genomics Proteomics*, 4: 387-98.
Holoch, D., and D. Moazed. 2015. 'RNA-mediated epigenetic regulation of gene expression', *Nat Rev Genet*, 16: 71-84.
Homer-Bouthiette, C., Y. Zhao, L. B. Shunkwiler, B. Van Peel, E. Garrett-Mayer, R. C. Baird, A. I. Rissman, S. T. Guest, S. P. Ethier, M. C. John, P. A. Powers, J. D. Haag, M. N. Gould, and B. M. G. Smits. 2018. 'Deletion of the murine ortholog of the 8q24 gene desert has anti-cancer effects in transgenic mammary cancer models', *BMC Cancer*, 18: 1233.
Hon, C. C., J. A. Ramilowski, J. Harshbarger, N. Bertin, O. J. Rackham, J. Gough, E. Denisenko, S. Schmeier, T. M. Poulsen, J. Severin, M. Lizio, H. Kawaji, T. Kasukawa, M. Itoh, A. M. Burroughs, S. Noma, S. Djebali, T. Alam, Y. A. Medvedeva, A. C. Testa, L. Lipovich, C. W. Yip, I. Abugessaisa, M. Mendez, A. Hasegawa, D. Tang, T. Lassmann, P. Heutink, M. Babina, C. A. Wells, S. Kojima, Y. Nakamura, H. Suzuki, C. O. Daub, M. J. de Hoon, E. Arner, Y. Hayashizaki, P. Carninci, and A. R. Forrest. 2017. 'An atlas of human long non-coding RNAs with accurate 5' ends', *Nature*, 543: 199-204.
Hotchkiss, R. D. 1948. 'The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography', *J Biol Chem*, 175: 315-32.

Huang, C., and B. Zhu. 2018. 'Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity', *Biophys Rep*, 4: 170-77.

Huang da, W., B. T. Sherman, and R. A. Lempicki. 2009. 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nat Protoc*, 4: 44-57.

Huang, J., X. Luo, L. Zeng, Z. Huang, M. Huang, W. You, and C. Ke. 2018. 'Expression profiling of lncRNAs and mRNAs reveals regulation of muscle growth in the Pacific abalone, Haliotis discus hannai', *Sci Rep*, 8: 16839.

Huang, J., A. Zhang, T. T. Ho, Z. Zhang, N. Zhou, X. Ding, X. Zhang, M. Xu, and Y. Y. Mo. 2016. 'Linc-RoR promotes c-Myc expression through hnRNP I and AUF1', *Nucleic Acids Res*, 44: 3059-69.

Hui, B., H. Ji, Y. Xu, J. Wang, Z. Ma, C. Zhang, K. Wang, and Y. Zhou. 2019. 'RREB1-induced upregulation of the lncRNA AGAP2-AS1 regulates the proliferation and migration of pancreatic cancer partly through suppressing ANKRD1 and ANGPTL4', *Cell Death Dis*, 10: 207.

Huppi, K., J. J. Pitt, B. M. Wahlberg, and N. J. Caplen. 2012. 'The 8q24 gene desert: an oasis of non-coding transcriptional activity', *Front Genet*, 3: 69.

Hutchinson, J. N., A. W. Ensminger, C. M. Clemson, C. R. Lynch, J. B. Lawrence, and A. Chess. 2007. 'A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains', *BMC Genomics*, 8: 39.

Ianni, A., S. Hoelper, M. Krueger, T. Braun, and E. Bober. 2017. 'Sirt7 stabilizes rDNA heterochromatin through recruitment of DNMT1 and Sirt1', *Biochem Biophys Res Commun*, 492: 434-40.

Iida, T., I. Suetake, S. Tajima, H. Morioka, S. Ohta, C. Obuse, and T. Tsurimoto. 2002. 'PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA', *Genes Cells*, 7: 997-1007.

Ilan, L., and S. Katzav. 2012. 'Human Vav1 expression in hematopoietic and cancer cell lines is regulated by c-Myb and by CpG methylation', *PLoS One*, 7: e29939.

Inoue, A., and D. Fujimoto. 1969. 'Enzymatic deacetylation of histone', *Biochem Biophys Res Commun*, 36: 146-50.

Insinga, A., S. Monestiroli, S. Ronzoni, R. Carbone, M. Pearson, G. Pruneri, G. Viale, E. Appella, P. Pelicci, and S. Minucci. 2004. 'Impairment of p53 acetylation, stability and function by an oncogenic transcription factor', *EMBO J*, 23: 1144-54.

Ishiyama, S., A. Nishiyama, Y. Saeki, K. Moritsugu, D. Morimoto, L. Yamaguchi, N. Arai, R. Matsumura, T. Kawakami, Y. Mishima, H. Hojo, S. Shimamura, F. Ishikawa, S. Tajima, K. Tanaka, M. Ariyoshi, M. Shirakawa, M. Ikeguchi, A. Kidera, I. Suetake, K. Arita, and M. Nakanishi. 2017. 'Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance', *Mol Cell*, 68: 350-60 e7.

Iyer, M. K., Y. S. Niknafs, R. Malik, U. Singhal, A. Sahu, Y. Hosono, T. R. Barrette, J. R. Prensner, J. R. Evans, S. Zhao, A. Poliakov, X. Cao, S. M. Dhanasekaran, Y. M. Wu, D. R. Robinson, D. G. Beer, F. Y. Feng, H. K. Iyer, and A. M. Chinnaiyan. 2015. 'The landscape of long noncoding RNAs in the human transcriptome', *Nat Genet*, 47: 199-208.

Izadifard, M., H. Pashaiefar, M. Yaghmaie, M. Montazeri, M. Sadraie, M. Momeny, M. Jalili, M. Ahmadvand, S. H. Ghaffari, S. Mohammadi, K. Alimoghaddam, and A. Ghavamzadeh. 2018. 'Expression Analysis of PVT1, CCDC26, and CCAT1 Long Noncoding RNAs in Acute Myeloid Leukemia Patients', *Genet Test Mol Biomarkers*, 22: 593-98.

Jacob, F., and J. Monod. 1961. 'Genetic regulatory mechanisms in the synthesis of proteins', *J Mol Biol*, 3: 318-56.

Jacob, M. D., T. E. Audas, J. Uniacke, L. Trinkle-Mulcahy, and S. Lee. 2013. 'Environmental cues induce a long noncoding RNA-dependent remodeling of the nucleolus', *Mol Biol Cell*, 24: 2943-53.

Jaenisch, R., and A. Bird. 2003. 'Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals', *Nat Genet*, 33 Suppl: 245-54.

Jelinek, J., V. Gharibyan, M. R. Estecio, K. Kondo, R. He, W. Chung, Y. Lu, N. Zhang, S. Liang, H. M. Kantarjian, J. E. Cortes, and J. P. Issa. 2011. 'Aberrant DNA methylation is associated with disease progression, resistance to imatinib and shortened survival in chronic myelogenous leukemia', *PLoS One*, 6: e22110.

Jeong, S., G. Liang, S. Sharma, J. C. Lin, S. H. Choi, H. Han, C. B. Yoo, G. Egger, A. S. Yang, and P. A. Jones. 2009. 'Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA', *Mol Cell Biol*, 29: 5366-76.

Ji, J., X. Dai, S. J. Yeung, and X. He. 2019. 'The role of long non-coding RNA GAS5 in cancers', *Cancer Manag Res*, 11: 2729-37.

Jia, D., R. Z. Jurkowska, X. Zhang, A. Jeltsch, and X. Cheng. 2007. 'Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation', *Nature*, 449: 248-51.

Jia, L., G. Landan, M. Pomerantz, R. Jaschek, P. Herman, D. Reich, C. Yan, O. Khalid, P. Kantoff, W. Oh, J. R. Manak, B. P. Berman, B. E. Henderson, B. Frenkel, C. A. Haiman, M. Freedman, A. Tanay, and G. A. Coetzee. 2009. 'Functional enhancers at the gene-poor 8q24 cancer-linked locus', *PLoS Genet*, 5: e1000597.

Jin, B., Q. Tao, J. Peng, H. M. Soo, W. Wu, J. Ying, C. R. Fields, A. L. Delmas, X. Liu, J. Qiu, and K. D. Robertson. 2008. 'DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function', *Hum Mol Genet*, 17: 690-709.

Jones, P. A. 2012. 'Functions of DNA methylation: islands, start sites, gene bodies and beyond', *Nat Rev Genet*, 13: 484-92.

Kang, S. H., C. H. Park, H. C. Jeung, K. Y. Kim, S. Y. Rha, and H. C. Chung. 2007. 'Integrated in silico and biological validation of the blocking effect of Cot-1 DNA on Microarray-CGH', *Int J Mol Med*, 19: 901-8.

Karpen, G. H., J. E. Schaefer, and C. D. Laird. 1988. 'A Drosophila rRNA gene located in euchromatin is active in transcription and nucleolus formation', *Genes Dev*, 2: 1745-63.

Kass, S. U., N. Landsberger, and A. P. Wolffe. 1997. 'DNA methylation directs a time-dependent repression of transcription initiation', *Curr Biol*, 7: 157-65.

Kerr, J. F., A. H. Wyllie, and A. R. Currie. 1972. 'Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics', *Br J Cancer*, 26: 239-57.

Khalil, A. M., M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B. E. Bernstein, A. van Oudenaarden, A. Regev, E. S. Lander, and J. L. Rinn. 2009. 'Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression', *Proc Natl Acad Sci U S A*, 106: 11667-72.

Kim, H. J., and S. C. Bae. 2011. 'Histone deacetylase inhibitors: molecular mechanisms of action and clinical trials as anti-cancer drugs', *Am J Transl Res*, 3: 166-79.

Kim, H. J., and J. P. Taylor. 2017. 'Lost in Transportation: Nucleocytoplasmic Transport Defects in ALS and Other Neurodegenerative Diseases', *Neuron*, 96: 285-97.

Kim, J., J. Daniel, A. Espejo, A. Lake, M. Krishna, L. Xia, Y. Zhang, and M. T. Bedford. 2006. 'Tudor, MBT and chromo domains gauge the degree of lysine methylation', *EMBO Rep*, 7: 397-403.

Kim, K.; Mossman, D.; Small, D.; Scott, R. J. . 2012. 'Gene Expression Profiling of Human Myeloid Leukemic MV4-11 Cells Treated with 5-Aza-2'-deoxycytidine', *Journal of Cancer Therapy*, 3: 177-82.

Kim, S. C., R. Sprung, Y. Chen, Y. Xu, H. Ball, J. Pei, T. Cheng, Y. Kho, H. Xiao, L. Xiao, N. V. Grishin, M. White, X. J. Yang, and Y. Zhao. 2006. 'Substrate and functional diversity of lysine acetylation revealed by a proteomics survey', *Mol Cell*, 23: 607-18.

Kind, J., L. Pagie, S. S. de Vries, L. Nahidiazar, S. S. Dey, M. Bienko, Y. Zhan, B. Lajoie, C. A. de Graaf, M. Amendola, G. Fudenberg, M. Imakaev, L. A. Mirny, K. Jalink, J. Dekker, A. van Oudenaarden, and B. van Steensel. 2015. 'Genome-wide maps of nuclear lamina interactions in single human cells', *Cell*, 163: 134-47.

Kino, T., D. E. Hurt, T. Ichijo, N. Nader, and G. P. Chrousos. 2010. 'Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor', *Sci Signal*, 3: ra8.

Kirchner, S., and Z. Ignatova. 2015. 'Emerging roles of tRNA in adaptive translation, signalling dynamics and disease', *Nat Rev Genet*, 16: 98-112.

Kiss, T. 2001. 'Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs', *EMBO J*, 20: 3617-22.

Kitamura, R., T. Sekimoto, S. Ito, S. Harada, H. Yamagata, H. Masai, Y. Yoneda, and K. Yanagi. 2006. 'Nuclear import of Epstein-Barr virus nuclear antigen 1 mediated by NPI-1 (Importin alpha5) is up- and down-regulated by phosphorylation of the nuclear localization signal for which Lys379 and Arg380 are essential', *J Virol*, 80: 1979-91.

Klus, P., B. Bolognesi, F. Agostini, D. Marchese, A. Zanzoni, and G. G. Tartaglia. 2014. 'The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities', *Bioinformatics*, 30: 1601-8.

Kobayashi, J. 2004. 'Molecular mechanism of the recruitment of NBS1/hMRE11/hRAD50 complex to DNA double-strand breaks: NBS1 binds to gamma-H2AX through FHA/BRCT domain', *J Radiat Res*, 45: 473-8.

Kohlmaier, A., F. Savarese, M. Lachner, J. Martens, T. Jenuwein, and A. Wutz. 2004. 'A chromosomal memory triggered by Xist regulates histone methylation in X inactivation', *PLoS Biol*, 2: E171.

Kornberg, R. D. 1974. 'Chromatin structure: a repeating unit of histones and DNA', *Science*, 184: 868-71.

Kornberg, R. D., and J. O. Thomas. 1974. 'Chromatin structure; oligomers of the histones', *Science*, 184: 865-8.

Kotake, Y., T. Nakagawa, K. Kitagawa, S. Suzuki, N. Liu, M. Kitagawa, and Y. Xiong. 2011. 'Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene', *Oncogene*, 30: 1956-62.

Krchnakova, Z., P. K. Thakur, M. Krausova, N. Bieberstein, N. Haberman, M. Muller-McNicoll, and D. Stanek. 2019. 'Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5' splice-site sequences due to weak interactions with SR proteins', *Nucleic Acids Res*, 47: 911-28.

Krol, J., I. Loedige, and W. Filipowicz. 2010. 'The widespread regulation of microRNA biogenesis, function and decay', *Nat Rev Genet*, 11: 597-610.

Kuhn, M. W., I. Radtke, L. Bullinger, S. Goorha, J. Cheng, J. Edelmann, J. Gohlke, X. Su, P. Paschka, S. Pounds, J. Krauter, A. Ganser, A. Quessar, R. Ribeiro, V. I. Gaidzik, S. Shurtleff, J. Kronke, K. Holzmann, J. Ma, R. F. Schlenk, J. E. Rubnitz, K. Dohner, H. Dohner, and J. R. Downing. 2012. 'High-resolution genomic profiling of adult and pediatric core-binding factor acute myeloid leukemia reveals new recurrent genomic alterations', *Blood*, 119: e67-75.

Kung, J. T., D. Colognori, and J. T. Lee. 2013. 'Long noncoding RNAs: past, present, and future', *Genetics*, 193: 651-69.

Lai, F., U. A. Orom, M. Cesaroni, M. Beringer, D. J. Taatjes, G. A. Blobel, and R. Shiekhattar. 2013. 'Activating RNAs associate with Mediator to enhance chromatin architecture and transcription', *Nature*, 494: 497-501.

Lan, Y., X. Xiao, Z. He, Y. Luo, C. Wu, L. Li, and X. Song. 2018. 'Long noncoding RNA OCC-1 suppresses cell growth through destabilizing HuR protein in colorectal cancer', *Nucleic Acids Res*, 46: 5809-21.

Langdon, E. M., Y. Qiu, A. Ghanbari Niaki, G. A. McLaughlin, C. A. Weidmann, T. M. Gerbich, J. A. Smith, J. M. Crutchley, C. M. Termini, K. M. Weeks, S. Myong, and A. S. Gladfelter. 2018. 'mRNA structure determines specificity of a polyQ-driven phase separation', *Science*, 360: 922-27.

Larsen, F., J. Solheim, and H. Prydz. 1993. 'A methylated CpG island 3' in the apolipoprotein-E gene does not repress its transcription', *Hum Mol Genet*, 2: 775-80.

Lasho, T. L., A. Tefferi, A. Pardanani, C. M. Finke, S. R. Fink, A. A. Caron, P. A. Decker, and R. B. Jenkins. 2012. 'Differential distribution of CCDC26 glioma-risk alleles in myeloid malignancies with mutant IDH1 compared with their IDH2R140-mutated or IDH-unmutated counterparts', *Leukemia*, 26: 1406-7.

Laurent, L., E. Wong, G. Li, T. Huynh, A. Tsirigos, C. T. Ong, H. M. Low, K. W. Kin Sung, I. Rigoutsos, J. Loring, and C. L. Wei. 2010. 'Dynamic changes in the human methylome during differentiation', *Genome Res*, 20: 320-31.

Lee, D. H., and A. L. Goldberg. 1998. 'Proteasome inhibitors: valuable new tools for cell biologists', *Trends Cell Biol*, 8: 397-403.
Lee, R. C., R. L. Feinbaum, and V. Ambros. 1993. 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14', *Cell*, 75: 843-54.
Lehnertz, B., Y. Ueda, A. A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein, and A. H. Peters. 2003. 'Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin', *Curr Biol*, 13: 1192-200.
Lemaitre, C., A. Grabarz, K. Tsouroula, L. Andronov, A. Furst, T. Pankotai, V. Heyer, M. Rogier, K. M. Attwood, P. Kessler, G. Dellaire, B. Klaholz, B. Reina-San-Martin, and E. Soutoglou. 2014. 'Nuclear position dictates DNA repair pathway choice', *Genes Dev*, 28: 2450-63.
Leone, S., and R. Santoro. 2016. 'Challenges in the analysis of long noncoding RNA functionality', *FEBS Lett*, 590: 2342-53.
Leonhardt, H., A. W. Page, H. U. Weier, and T. H. Bestor. 1992. 'A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei', *Cell*, 71: 865-73.
Li, E., C. Beard, and R. Jaenisch. 1993. 'Role for DNA methylation in genomic imprinting', *Nature*, 366: 362-5.
Li, E., T. H. Bestor, and R. Jaenisch. 1992. 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality', *Cell*, 69: 915-26.
Li, S., T. Jin, J. Zhang, H. Lou, B. Yang, Y. Li, C. Chen, and Y. Zhang. 2012. 'Polymorphisms of TREH, IL4R and CCDC26 genes associated with risk of glioma', *Cancer Epidemiol*, 36: 283-7.
Li, T., L. Wang, Y. Du, S. Xie, X. Yang, F. Lian, Z. Zhou, and C. Qian. 2018. 'Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation', *Nucleic Acids Res*, 46: 3218-31.
Li, Y., X. Liu, X. Guo, X. Liu, and J. Luo. 2017. 'DNA methyltransferase 1 mediated aberrant methylation and silencing of SHP-1 gene in chronic myelogenous leukemia cells', *Leuk Res*, 58: 9-13.
Li, Z., P. Hou, D. Fan, M. Dong, M. Ma, H. Li, R. Yao, Y. Li, G. Wang, P. Geng, A. Mihretab, D. Liu, Y. Zhang, B. Huang, and J. Lu. 2017. 'The degradation of EZH2 mediated by lncRNA ANCR attenuated the invasion and metastasis of breast cancer', *Cell Death Differ*, 24: 59-71.
Liao, J., R. Karnik, H. Gu, M. J. Ziller, K. Clement, A. M. Tsankov, V. Akopian, C. A. Gifford, J. Donaghey, C. Galonska, R. Pop, D. Reyon, S. Q. Tsai, W. Mallard, J. K. Joung, J. L. Rinn, A. Gnirke, and A. Meissner. 2015. 'Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells', *Nat Genet*, 47: 469-78.
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: 289-93.
Liu, B., L. Sun, Q. Liu, C. Gong, Y. Yao, X. Lv, L. Lin, H. Yao, F. Su, D. Li, M. Zeng, and E. Song. 2015. 'A cytoplasmic NF-kappaB interacting long noncoding RNA blocks

IkappaB phosphorylation and suppresses breast cancer metastasis', *Cancer Cell*, 27: 370-81.
Liu, P., H. Yang, J. Zhang, X. Peng, Z. Lu, W. Tong, and J. Chen. 2017. 'The lncRNA MALAT1 acts as a competing endogenous RNA to regulate KRAS expression by sponging miR-217 in pancreatic ductal adenocarcinoma', *Sci Rep*, 7: 5186.
Liu, X. S., H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R. A. Young, and R. Jaenisch. 2016. 'Editing DNA Methylation in the Mammalian Genome', *Cell*, 167: 233-47 e17.
Liu, Y., M. Li, X. Bo, T. Li, L. Ma, T. Zhai, and T. Huang. 2018. 'Systematic Analysis of Long Non-Coding RNAs and mRNAs in the Ovaries of Duroc Pigs During Different Follicular Stages Using RNA Sequencing', *Int J Mol Sci*, 19.
Lo, P. K., B. Wolfson, and Q. Zhou. 2016. 'Cellular, physiological and pathological aspects of the long non-coding RNA NEAT1', *Front Biol (Beijing)*, 11: 413-26.
Loughery, J. E., P. D. Dunne, K. M. O'Neill, R. R. Meehan, J. R. McDaid, and C. P. Walsh. 2011. 'DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response', *Hum Mol Genet*, 20: 3241-55.
Lu, H. W., M. Huang, J. H. Wang, X. L. Sun, and Y. Q. Ke. 2015. 'CCDC26 rs4295627 polymorphism (8q24.21) and glioma risk: a meta-analysis', *Genet Mol Res*, 14: 12074-84.
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: 251-60.
Lv, J., H. Liu, S. Yu, H. Liu, W. Cui, Y. Gao, T. Zheng, G. Qin, J. Guo, T. Zeng, Z. Han, Y. Zhang, and Q. Wu. 2015. 'Identification of 4438 novel lincRNAs involved in mouse pre-implantation embryonic development', *Mol Genet Genomics*, 290: 685-97.
Lyko, F. 2018. 'The DNA methyltransferase family: a versatile toolkit for epigenetic regulation', *Nat Rev Genet*, 19: 81-92.
Lyle, R., D. Watanabe, D. te Vruchte, W. Lerchner, O. W. Smrzka, A. Wutz, J. Schageman, L. Hahner, C. Davies, and D. P. Barlow. 2000. 'The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1', *Nat Genet*, 25: 19-21.
Lyon, M. F. 1961. 'Gene action in the X-chromosome of the mouse (Mus musculus L.)', *Nature*, 190: 372-3.
———. 1998. 'X-chromosome inactivation: a repeat hypothesis', *Cytogenet Cell Genet*, 80: 133-7.
Ma, L., V. B. Bajic, and Z. Zhang. 2013. 'On the classification of long non-coding RNAs', *RNA Biol*, 10: 925-33.
Ma, N., S. Matsunaga, H. Takata, R. Ono-Maniwa, S. Uchiyama, and K. Fukui. 2007. 'Nucleolin functions in nucleolus formation and chromosome congression', *J Cell Sci*, 120: 2091-105.
Maass, P. G., A. R. Barutcu, C. L. Weiner, and J. L. Rinn. 2018. 'Inter-chromosomal Contact Properties in Live-Cell Imaging and in Hi-C', *Mol Cell*, 70: 188-89.
Maass, P. G., A. Rump, H. Schulz, S. Stricker, L. Schulze, K. Platzer, A. Aydin, S. Tinschert, M. B. Goldring, F. C. Luft, and S. Bahring. 2012. 'A misplaced lncRNA causes brachydactyly in humans', *J Clin Invest*, 122: 3990-4002.

Maharana, S., J. Wang, D. K. Papadopoulos, D. Richter, A. Pozniakovsky, I. Poser, M. Bickle, S. Rizk, J. Guillen-Boixet, T. M. Franzmann, M. Jahnel, L. Marrone, Y. T. Chang, J. Sterneckert, P. Tomancak, A. A. Hyman, and S. Alberti. 2018. 'RNA buffers the phase separation behavior of prion-like RNA binding proteins', *Science*, 360: 918-21.

Mao, Y. S., H. Sunwoo, B. Zhang, and D. L. Spector. 2011. 'Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs', *Nat Cell Biol*, 13: 95-101.

Margueron, R., G. Li, K. Sarma, A. Blais, J. Zavadil, C. L. Woodcock, B. D. Dynlacht, and D. Reinberg. 2008. 'Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms', *Mol Cell*, 32: 503-18.

Mastroeni, D., L. Chouliaras, A. Grover, W. S. Liang, K. Hauns, J. Rogers, and P. D. Coleman. 2013. 'Reduced RAN expression and disrupted transport between cytoplasm and nucleus; a key event in Alzheimer's disease pathophysiology', *PLoS One*, 8: e53349.

Mazzola, J. L., and M. A. Sirover. 2003. 'Subcellular localization of human glyceraldehyde-3-phosphate dehydrogenase is independent of its glycolytic function', *Biochim Biophys Acta*, 1622: 50-6.

McDowall, A. W., J. M. Smith, and J. Dubochet. 1986. 'Cryo-electron microscopy of vitrified chromosomes in situ', *EMBO J*, 5: 1395-402.

McGrath, J., and D. Solter. 1984. 'Completion of mouse embryogenesis requires both the maternal and paternal genomes', *Cell*, 37: 179-83.

McHugh, C. A., C. K. Chen, A. Chow, C. F. Surka, C. Tran, P. McDonel, A. Pandya-Jones, M. Blanco, C. Burghard, A. Moradian, M. J. Sweredoski, A. A. Shishkin, J. Su, E. S. Lander, S. Hess, K. Plath, and M. Guttman. 2015. 'The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3', *Nature*, 521: 232-6.

Mekhail, K., L. Gunaratnam, M. E. Bonicalzi, and S. Lee. 2004. 'HIF activation by pH-dependent nucleolar sequestration of VHL', *Nat Cell Biol*, 6: 642-7.

Mele, M., K. Mattioli, W. Mallard, D. M. Shechner, C. Gerhardinger, and J. L. Rinn. 2017. 'Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs', *Genome Res*, 27: 27-37.

Mercer, T. R., and J. S. Mattick. 2013. 'Structure and function of long noncoding RNAs in epigenetic regulation', *Nat Struct Mol Biol*, 20: 300-7.

Merry, C. R., M. E. Forrest, J. N. Sabers, L. Beard, X. H. Gao, M. Hatzoglou, M. W. Jackson, Z. Wang, S. D. Markowitz, and A. M. Khalil. 2015. 'DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer', *Hum Mol Genet*, 24: 6240-53.

Miller, K. M., J. V. Tjeertes, J. Coates, G. Legube, S. E. Polo, S. Britton, and S. P. Jackson. 2010. 'Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining', *Nat Struct Mol Biol*, 17: 1144-51.

Milutinovic, S., S. E. Brown, Q. Zhuang, and M. Szyf. 2004. 'DNA methyltransferase 1 knock down induces gene expression by a mechanism independent of DNA methylation and histone deacetylation', *J Biol Chem*, 279: 27915-27.

Milutinovic, S., Q. Zhuang, A. Niveleau, and M. Szyf. 2003. 'Epigenomic stress response. Knockdown of DNA methyltransferase 1 triggers an intra-S-phase arrest of DNA replication and induction of stress response genes', *J Biol Chem*, 278: 14985-95.

Misaki, T., L. Yamaguchi, J. Sun, M. Orii, A. Nishiyama, and M. Nakanishi. 2016. 'The replication foci targeting sequence (RFTS) of DNMT1 functions as a potent histone H3 binding domain regulated by autoinhibition', *Biochem Biophys Res Commun*, 470: 741-47.

Mizuguchi, T., J. Barrowman, and S. I. Grewal. 2015. 'Chromosome domain architecture and dynamic organization of the fission yeast genome', *FEBS Lett*, 589: 2975-86.

Mohammad, F., T. Mondal, N. Guseva, G. K. Pandey, and C. Kanduri. 2010. 'Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1', *Development*, 137: 2493-9.

Morey, C., and P. Avner. 2004. 'Employment opportunities for non-coding RNAs', *FEBS Lett*, 567: 27-34.

Morgan, A. E., T. J. Davies, and M. T. Mc Auley. 2018. 'The role of DNA methylation in ageing and cancer', *Proc Nutr Soc*, 77: 412-22.

Morozova, O., and M. A. Marra. 2008. 'Applications of next-generation sequencing technologies in functional genomics', *Genomics*, 92: 255-64.

Mortusewicz, O., L. Schermelleh, J. Walter, M. C. Cardoso, and H. Leonhardt. 2005. 'Recruitment of DNA methyltransferase I to DNA repair sites', *Proc Natl Acad Sci U S A*, 102: 8905-9.

Murray, N. R., G. P. Baumgardner, D. J. Burns, and A. P. Fields. 1993. 'Protein kinase C isotypes in human erythroleukemia (K562) cell proliferation and differentiation. Evidence that beta II protein kinase C is required for proliferation', *J Biol Chem*, 268: 15847-53.

Nagano, T., J. A. Mitchell, L. A. Sanz, F. M. Pauler, A. C. Ferguson-Smith, R. Feil, and P. Fraser. 2008. 'The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin', *Science*, 322: 1717-20.

Nagata, S., H. Nagase, K. Kawane, N. Mukae, and H. Fukuyama. 2003. 'Degradation of chromosomal DNA during apoptosis', *Cell Death Differ*, 10: 108-16.

Nakayama, J., J. C. Rice, B. D. Strahl, C. D. Allis, and S. I. Grewal. 2001. 'Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly', *Science*, 292: 110-3.

Nakayama, K., A. Furusu, Q. Xu, T. Konta, and M. Kitamura. 2001. 'Unexpected transcriptional induction of monocyte chemoattractant protein 1 by proteasome inhibition: involvement of the c-Jun N-terminal kinase-activator protein 1 pathway', *J Immunol*, 167: 1145-50.

Nan, X., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird. 1998. 'Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex', *Nature*, 393: 386-9.

Nardozzi, J. D., K. Lott, and G. Cingolani. 2010. 'Phosphorylation meets nuclear import: a review', *Cell Commun Signal*, 8: 32.

Nevins, J. R., and J. E. Darnell, Jr. 1978. 'Steps in the processing of Ad2 mRNA: poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing', *Cell*, 15: 1477-93.

Nishino, Y., M. Eltsov, Y. Joti, K. Ito, H. Takata, Y. Takahashi, S. Hihara, A. S. Frangakis, N. Imamoto, T. Ishikawa, and K. Maeshima. 2012. 'Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure', *EMBO J*, 31: 1644-53.

Nishiyama, A., L. Yamaguchi, J. Sharif, Y. Johmura, T. Kawamura, K. Nakanishi, S. Shimamura, K. Arita, T. Kodama, F. Ishikawa, H. Koseki, and M. Nakanishi. 2013. 'Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication', *Nature*, 502: 249-53.

Nora, E. P., B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker, and E. Heard. 2012. 'Spatial partitioning of the regulatory landscape of the X-inactivation centre', *Nature*, 485: 381-5.

Nozaki, T., R. Imai, M. Tanbo, R. Nagashima, S. Tamura, T. Tani, Y. Joti, M. Tomita, K. Hibino, M. T. Kanemaki, K. S. Wendt, Y. Okada, T. Nagai, and K. Maeshima. 2017. 'Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging', *Mol Cell*, 67: 282-93 e7.

O'Hagan, H. M., W. Wang, S. Sen, C. Destefano Shields, S. S. Lee, Y. W. Zhang, E. G. Clements, Y. Cai, L. Van Neste, H. Easwaran, R. A. Casero, C. L. Sears, and S. B. Baylin. 2011. 'Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands', *Cancer Cell*, 20: 606-19.

O'Leary, V. B., S. Hain, D. Maugg, J. Smida, O. Azimzadeh, S. Tapio, S. V. Ovsepian, and M. J. Atkinson. 2017. 'Long non-coding RNA PARTICLE bridges histone and DNA methylation', *Sci Rep*, 7: 1790.

O'Malley, F., K. Rayeroux, M. Cole-Sinclair, M. Tong, and L. J. Campbell. 1999. 'MYC amplification in two further cases of acute myeloid leukemia with trisomy 4 and double minute chromosomes', *Cancer Genet Cytogenet*, 109: 123-5.

Obrig, T. G., W. J. Culp, W. L. McKeehan, and B. Hardesty. 1971. 'The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes', *J Biol Chem*, 246: 174-81.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development', *Cell*, 99: 247-57.

Okano, M., S. Xie, and E. Li. 1998. 'Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases', *Nat Genet*, 19: 219-20.

Okazaki, Y., M. Furuno, T. Kasukawa, J. Adachi, H. Bono, S. Kondo, I. Nikaido, N. Osato, R. Saito, H. Suzuki, I. Yamanaka, H. Kiyosawa, K. Yagi, Y. Tomaru, Y. Hasegawa, A. Nogami, C. Schonbach, T. Gojobori, R. Baldarelli, D. P. Hill, C. Bult, D. A. Hume, J. Quackenbush, L. M. Schriml, A. Kanapin, H. Matsuda, S. Batalov, K. W. Beisel, J. A. Blake, D. Bradt, V. Brusic, C. Chothia, L. E. Corbani, S. Cousins, E. Dalla, T. A. Dragani, C. F. Fletcher, A. Forrest, K. S. Frazer, T. Gaasterland, M. Gariboldi, C. Gissi, A. Godzik, J. Gough, S. Grimmond, S. Gustincich, N. Hirokawa, I. J. Jackson, E. D. Jarvis, A. Kanai, H. Kawaji, Y. Kawasawa, R. M. Kedzierski, B. L. King, A. Konagaya, I. V. Kurochkin, Y. Lee, B. Lenhard, P. A. Lyons, D. R. Maglott, L. Maltais, L. Marchionni, L. McKenzie, H. Miki, T. Nagashima, K. Numata, T. Okido, W. J.

Pavan, G. Pertea, G. Pesole, N. Petrovsky, R. Pillai, J. U. Pontius, D. Qi, S. Ramachandran, T. Ravasi, J. C. Reed, D. J. Reed, J. Reid, B. Z. Ring, M. Ringwald, A. Sandelin, C. Schneider, C. A. Semple, M. Setou, K. Shimada, R. Sultana, Y. Takenaka, M. S. Taylor, R. D. Teasdale, M. Tomita, R. Verardo, L. Wagner, C. Wahlestedt, Y. Wang, Y. Watanabe, C. Wells, L. G. Wilming, A. Wynshaw-Boris, M. Yanagisawa, I. Yang, L. Yang, Z. Yuan, M. Zavolan, Y. Zhu, A. Zimmer, P. Carninci, N. Hayatsu, T. Hirozane-Kishikawa, H. Konno, M. Nakamura, N. Sakazume, K. Sato, T. Shiraki, K. Waki, J. Kawai, K. Aizawa, T. Arakawa, S. Fukuda, A. Hara, W. Hashizume, K. Imotani, Y. Ishii, M. Itoh, I. Kagawa, A. Miyazaki, K. Sakai, D. Sasaki, K. Shibata, A. Shinagawa, A. Yasunishi, M. Yoshino, R. Waterston, E. S. Lander, J. Rogers, E. Birney, Y. Hayashizaki, Fantom Consortium, Riken Genome Exploration Research Group Phase I, and I. I. Team. 2002. 'Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs', *Nature*, 420: 563-73.

Orlowski, M., and S. Wilk. 2000. 'Catalytic activities of the 20 S proteasome, a multicatalytic proteinase complex', *Arch Biochem Biophys*, 383: 1-16.

Orom, U. A., T. Derrien, M. Beringer, K. Gumireddy, A. Gardini, G. Bussotti, F. Lai, M. Zytnicki, C. Notredame, Q. Huang, R. Guigo, and R. Shiekhattar. 2010. 'Long noncoding RNAs with enhancer-like function in human cells', *Cell*, 143: 46-58.

Osborne, C. S., L. Chakalova, K. E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J. A. Mitchell, S. Lopes, W. Reik, and P. Fraser. 2004. 'Active genes dynamically colocalize to shared sites of ongoing transcription', *Nat Genet*, 36: 1065-71.

Otani, J., T. Nankumo, K. Arita, S. Inamoto, M. Ariyoshi, and M. Shirakawa. 2009. 'Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain', *EMBO Rep*, 10: 1235-41.

Ou, H. D., S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O'Shea. 2017. 'ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells', *Science*, 357.

Pakneshan, P., B. Tetu, and S. A. Rabbani. 2004. 'Demethylation of urokinase promoter as a prognostic marker in patients with breast carcinoma', *Clin Cancer Res*, 10: 3035-41.

Pandey, R. R., T. Mondal, F. Mohammad, S. Enroth, L. Redrup, J. Komorowski, T. Nagano, D. Mancini-Dinardo, and C. Kanduri. 2008. 'Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation', *Mol Cell*, 32: 232-46.

Park, C. S., H. Rehrauer, and I. M. Mansuy. 2013. 'Genome-wide analysis of H4K5 acetylation associated with fear memory in mice', *BMC Genomics*, 14: 539.

Parrizas, M., A. R. Saltiel, and D. LeRoith. 1997. 'Insulin-like growth factor 1 inhibits apoptosis using the phosphatidylinositol 3'-kinase and mitogen-activated protein kinase pathways', *J Biol Chem*, 272: 154-61.

Passarge, E. 1979. 'Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago', *Am J Hum Genet*, 31: 106-15.

Pastural, E., N. Takahashi, W. F. Dong, M. Bainbridge, A. Hull, D. Pearson, S. Huang, R. Lowsky, J. F. DeCoteau, and C. R. Geyer. 2007. 'RIZ1 repression is associated with insulin-like growth factor-1 signaling activation in chronic myeloid leukemia cell lines', *Oncogene*, 26: 1586-94.

Patel, K., J. Dickson, S. Din, K. Macleod, D. Jodrell, and B. Ramsahoye. 2010. 'Targeting of 5-aza-2'-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme', *Nucleic Acids Res*, 38: 4313-24.

Patil, D. P., C. K. Chen, B. F. Pickering, A. Chow, C. Jackson, M. Guttman, and S. R. Jaffrey. 2016. 'm(6)A RNA methylation promotes XIST-mediated transcriptional repression', *Nature*, 537: 369-73.

Paweletz, N. 2001. 'Walther Flemming: pioneer of mitosis research', *Nat Rev Mol Cell Biol*, 2: 72-5.

Pei, H., W. Hu, Z. Guo, H. Chen, J. Ma, W. Mao, B. Li, A. Wang, J. Wan, J. Zhang, J. Nie, G. Zhou, and T. K. Hei. 2018. 'Long Noncoding RNA CRYBG3 Blocks Cytokinesis by Directly Binding G-Actin', *Cancer Res*, 78: 4563-72.

Pena, C., E. Hurt, and V. G. Panse. 2017. 'Eukaryotic ribosome assembly, transport and quality control', *Nat Struct Mol Biol*, 24: 689-99.

Peng, W., and A. Jiang. 2016. 'Long noncoding RNA CCDC26 as a potential predictor biomarker contributes to tumorigenesis in pancreatic cancer', *Biomed Pharmacother*, 83: 712-17.

Pintacuda, G., A. N. Young, and A. Cerase. 2017. 'Function by Structure: Spotlights on Xist Long Non-coding RNA', *Front Mol Biosci*, 4: 90.

Plath, K., J. Fang, S. K. Mlynarczyk-Evans, R. Cao, K. A. Worringer, H. Wang, C. C. de la Cruz, A. P. Otte, B. Panning, and Y. Zhang. 2003. 'Role of histone H3 lysine 27 methylation in X inactivation', *Science*, 300: 131-5.

Pollard, K. S., S. R. Salama, N. Lambert, M. A. Lambot, S. Coppens, J. S. Pedersen, S. Katzman, B. King, C. Onodera, A. Siepel, A. D. Kern, C. Dehay, H. Igel, M. Ares, Jr., P. Vanderhaeghen, and D. Haussler. 2006. 'An RNA gene expressed during cortical development evolved rapidly in humans', *Nature*, 443: 167-72.

Pomerantz, M. M., C. A. Beckwith, M. M. Regan, S. K. Wyman, G. Petrovics, Y. Chen, D. J. Hawksworth, F. R. Schumacher, L. Mucci, K. L. Penney, M. J. Stampfer, J. A. Chan, K. G. Ardlie, B. R. Fritz, R. K. Parkin, D. W. Lin, M. Dyke, P. Herman, S. Lee, W. K. Oh, P. W. Kantoff, M. Tewari, D. G. McLeod, S. Srivastava, and M. L. Freedman. 2009. 'Evaluation of the 8q24 prostate cancer risk locus and MYC expression', *Cancer Res*, 69: 5568-74.

Ponjavic, J., C. P. Ponting, and G. Lunter. 2007. 'Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs', *Genome Res*, 17: 556-65.

Pope, B. D., T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, R. E. Thurman, Y. Cheng, G. Gulsoy, J. H. Dennis, M. P. Snyder, J. A. Stamatoyannopoulos, J. Taylor, R. C. Hardison, T. Kahveci, B. Ren, and D. M. Gilbert. 2014. 'Topologically associating domains are stable units of replication-timing regulation', *Nature*, 515: 402-5.

Pruitt, K. D., G. R. Brown, S. M. Hiatt, F. Thibaud-Nissen, A. Astashyn, O. Ermolaeva, C. M. Farrell, J. Hart, M. J. Landrum, K. M. McGarvey, M. R. Murphy, N. A. O'Leary, S. Pujar, B. Rajput, S. H. Rangwala, L. D. Riddick, A. Shkeda, H. Sun, P. Tamez, R. E.

Tully, C. Wallin, D. Webb, J. Weber, W. Wu, M. DiCuccio, P. Kitts, D. R. Maglott, T. D. Murphy, and J. M. Ostell. 2014. 'RefSeq: an update on mammalian reference sequences', *Nucleic Acids Res*, 42: D756-63.

Pullirsch, D., R. Hartel, H. Kishimoto, M. Leeb, G. Steiner, and A. Wutz. 2010. 'The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation', *Development*, 137: 935-43.

Purdy, M. M., C. Holz-Schietinger, and N. O. Reich. 2010. 'Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion', *Arch Biochem Biophys*, 498: 13-22.

Qi, X., X. J. Yu, X. M. Wang, T. N. Song, J. Zhang, X. Z. Guo, G. J. Li, and M. Shao. 2019. 'Knockdown of KCNQ1OT1 Suppresses Cell Invasion and Sensitizes Osteosarcoma Cells to CDDP by Upregulating DNMT1-Mediated Kcnq1 Expression', *Mol Ther Nucleic Acids*, 17: 804-18.

Quek, X. C., D. W. Thomson, J. L. Maag, N. Bartonicek, B. Signal, M. B. Clark, B. S. Gloss, and M. E. Dinger. 2015. 'lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs', *Nucleic Acids Res*, 43: D168-73.

Quinodoz, S., and M. Guttman. 2014. 'Long noncoding RNAs: an emerging link between gene regulation and nuclear organization', *Trends Cell Biol*, 24: 651-63.

Radtke, I., C. G. Mullighan, M. Ishii, X. Su, J. Cheng, J. Ma, R. Ganti, Z. Cai, S. Goorha, S. B. Pounds, X. Cao, C. Obert, J. Armstrong, J. Zhang, G. Song, R. C. Ribeiro, J. E. Rubnitz, S. C. Raimondi, S. A. Shurtleff, and J. R. Downing. 2009. 'Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia', *Proc Natl Acad Sci U S A*, 106: 12944-9.

Ragoczy, T., M. A. Bender, A. Telling, R. Byron, and M. Groudine. 2006. 'The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation', *Genes Dev*, 20: 1447-57.

Rahmanzadeh, R., G. Huttmann, J. Gerdes, and T. Scholzen. 2007. 'Chromophore-assisted light inactivation of pKi-67 leads to inhibition of ribosomal RNA synthesis', *Cell Prolif*, 40: 422-30.

Ramanathan, A., G. B. Robb, and S. H. Chan. 2016. 'mRNA capping: biological functions and applications', *Nucleic Acids Res*, 44: 7511-26.

Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. 2014. 'A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping', *Cell*, 159: 1665-80.

Rasmussen, K. D., and K. Helin. 2016. 'Role of TET enzymes in DNA methylation, development, and cancer', *Genes Dev*, 30: 733-50.

Rayeroux, K. C., and L. J. Campbell. 2009. 'Gene amplification in myeloid leukemias elucidated by fluorescence in situ hybridization', *Cancer Genet Cytogenet*, 193: 44-53.

Receveur, A., J. Ong, L. Merlin, Z. Azgui, H. Merle-Beral, R. Berger, and F. Nguyen-Khac. 2004. 'Trisomy 4 associated with double minute chromosomes and MYC amplification in acute myeloblastic leukemia', *Ann Genet*, 47: 423-7.

Rego, A., P. B. Sinclair, W. Tao, I. Kireev, and A. S. Belmont. 2008. 'The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure', *J Cell Sci*, 121: 1119-27.

Ricci, M. A., C. Manzo, M. F. Garcia-Parajo, M. Lakadamyali, and M. P. Cosma. 2015. 'Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo', *Cell*, 160: 1145-58.

Richmond, T. J., J. T. Finch, B. Rushton, D. Rhodes, and A. Klug. 1984. 'Structure of the nucleosome core particle at 7 A resolution', *Nature*, 311: 532-7.

Riedl, S. J., and Y. Shi. 2004. 'Molecular mechanisms of caspase regulation during apoptosis', *Nat Rev Mol Cell Biol*, 5: 897-907.

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, and H. Y. Chang. 2007. 'Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs', *Cell*, 129: 1311-23.

Rivera, C., F. Saavedra, F. Alvarez, C. Diaz-Celis, V. Ugalde, J. Li, I. Forne, Z. A. Gurard-Levin, G. Almouzni, A. Imhof, and A. Loyola. 2015. 'Methylation of histone H3 lysine 9 occurs during translation', *Nucleic Acids Res*, 43: 9097-106.

Rocchi, P., R. Tonelli, C. Camerin, S. Purgato, R. Fronza, F. Bianucci, F. Guerra, A. Pession, and A. M. Ferreri. 2005. 'p21Waf1/Cip1 is a common target induced by short-chain fatty acid HDAC inhibitors (valproic acid, tributyrin and sodium butyrate) in neuroblastoma cells', *Oncol Rep*, 13: 1139-44.

Rodriguez, J. A. 2014. 'Interplay between nuclear transport and ubiquitin/SUMO modifications in the regulation of cancer-related proteins', *Semin Cancer Biol*, 27: 11-9.

Rogakou, E. P., C. Boon, C. Redon, and W. M. Bonner. 1999. 'Megabase chromatin domains involved in DNA double-strand breaks in vivo', *J Cell Biol*, 146: 905-16.

Rogakou, E. P., W. Nieves-Neira, C. Boon, Y. Pommier, and W. M. Bonner. 2000. 'Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139', *J Biol Chem*, 275: 9390-5.

Rogakou, E. P., D. R. Pilch, A. H. Orr, V. S. Ivanova, and W. M. Bonner. 1998. 'DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139', *J Biol Chem*, 273: 5858-68.

Rondelet, G., T. Dal Maso, L. Willems, and J. Wouters. 2016. 'Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B', *J Struct Biol*, 194: 357-67.

Roos, W. P., and B. Kaina. 2006. 'DNA damage-induced cell death by apoptosis', *Trends Mol Med*, 12: 440-50.

Rothbart, S. B., K. Krajewski, N. Nady, W. Tempel, S. Xue, A. I. Badeaux, D. Barsyte-Lovejoy, J. Y. Martinez, M. T. Bedford, S. M. Fuchs, C. H. Arrowsmith, and B. D. Strahl. 2012. 'Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation', *Nat Struct Mol Biol*, 19: 1155-60.

Rountree, M. R., K. E. Bachman, and S. B. Baylin. 2000. 'DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci', *Nat Genet*, 25: 269-77.

Sahakyan, A., Y. Yang, and K. Plath. 2018. 'The Role of Xist in X-Chromosome Dosage Compensation', *Trends Cell Biol*, 28: 999-1013.
Sakai, K., T. Ohta, S. Minoshima, J. Kudoh, Y. Wang, P. J. de Jong, and N. Shimizu. 1995. 'Human ribosomal RNA gene cluster: identification of the proximal end containing a novel tandem repeat sequence', *Genomics*, 26: 521-6.
Santoro, R., and I. Grummt. 2001. 'Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription', *Mol Cell*, 8: 719-25.
Sato, K., H. Nakagawa, A. Tajima, K. Yoshida, and I. Inoue. 2010. 'ANRIL is implicated in the regulation of nucleus and potential transcriptional target of E2F1', *Oncol Rep*, 24: 701-7.
Schlackow, M., T. Nojima, T. Gomes, A. Dhir, M. Carmo-Fonseca, and N. J. Proudfoot. 2017. 'Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs', *Mol Cell*, 65: 25-38.
Schmelz, K., M. Wagner, B. Dorken, and I. Tamm. 2005. '5-Aza-2'-deoxycytidine induces p21WAF expression by demethylation of p73 leading to p53-independent apoptosis in myeloid leukemia', *Int J Cancer*, 114: 683-95.
Schmidt, M. H. H., R. Broll, H. P. Bruch, O. Bogler, and M. Duchrow. 2003. 'The proliferation marker pKi-67 organizes the nucleolus during the cell cycle depending on Ran and cyclin B', *J Pathol*, 199: 18-27.
Schoeftner, S., A. K. Sengupta, S. Kubicek, K. Mechtler, L. Spahn, H. Koseki, T. Jenuwein, and A. Wutz. 2006. 'Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing', *EMBO J*, 25: 3110-22.
Schubeler, D. 2015. 'Function and information content of DNA methylation', *Nature*, 517: 321-6.
Scott, M. S., D. Y. Thomas, and M. T. Hallett. 2004. 'Predicting subcellular localization via protein motif co-occurrence', *Genome Res*, 14: 1957-66.
Seto, E., and M. Yoshida. 2014. 'Erasers of histone acetylation: the histone deacetylase enzymes', *Cold Spring Harb Perspect Biol*, 6: a018713.
Shalini, S., L. Dorstyn, S. Dawar, and S. Kumar. 2015. 'Old, new and emerging functions of caspases', *Cell Death Differ*, 22: 526-39.
Shankar, S. R., A. G. Bahirvani, V. K. Rao, N. Bharathy, J. R. Ow, and R. Taneja. 2013. 'G9a, a multipotent regulator of gene expression', *Epigenetics*, 8: 16-22.
Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero, and Y. Shi. 2004. 'Histone demethylation mediated by the nuclear amine oxidase homolog LSD1', *Cell*, 119: 941-53.
Shin, Y., and C. P. Brangwynne. 2017. 'Liquid phase condensation in cell physiology and disease', *Science*, 357.
Shiraki, T., S. Kondo, S. Katayama, K. Waki, T. Kasukawa, H. Kawaji, R. Kodzius, A. Watahiki, M. Nakamura, T. Arakawa, S. Fukuda, D. Sasaki, A. Podhajska, M. Harbers, J. Kawai, P. Carninci, and Y. Hayashizaki. 2003. 'Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage', *Proc Natl Acad Sci U S A*, 100: 15776-81.
Shogren-Knaak, M., H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie, and C. L. Peterson. 2006. 'Histone H4-K16 acetylation controls chromatin structure and protein interactions', *Science*, 311: 844-7.

Shtivelman, E., and J. M. Bishop. 1989. 'The PVT gene frequently amplifies with MYC in tumor cells', *Mol Cell Biol*, 9: 1148-54.
Shukla, S., E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, and S. Oberdoerffer. 2011. 'CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing', *Nature*, 479: 74-9.
Siegfried, Z., S. Eden, M. Mendelsohn, X. Feng, B. Z. Tsuberi, and H. Cedar. 1999. 'DNA methylation represses transcription in vivo', *Nat Genet*, 22: 203-6.
Sinha, R., E. Allemand, Z. Zhang, R. Karni, M. P. Myers, and A. R. Krainer. 2010. 'Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF', *Mol Cell Biol*, 30: 2762-74.
Slovak, M. L., J. P. Ho, M. J. Pettenati, A. Khan, D. Douer, S. Lal, and S. T. Traweek. 1994. 'Localization of amplified MYC gene sequences to double minute chromosomes in acute myelogenous leukemia', *Genes Chromosomes Cancer*, 9: 62-7.
Smeets, D. F., U. Moog, C. M. Weemaes, G. Vaes-Peeters, G. F. Merkx, J. P. Niehof, and G. Hamers. 1994. 'ICF syndrome: a new case and review of the literature', *Hum Genet*, 94: 240-6.
Smeets, D., Y. Markaki, V. J. Schmid, F. Kraus, A. Tattermusch, A. Cerase, M. Sterr, S. Fiedler, J. Demmerle, J. Popken, H. Leonhardt, N. Brockdorff, T. Cremer, L. Schermelleh, and M. Cremer. 2014. 'Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci', *Epigenetics Chromatin*, 7: 8.
Solovei, I., M. Kreysing, C. Lanctot, S. Kosem, L. Peichl, T. Cremer, J. Guck, and B. Joffe. 2009. 'Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution', *Cell*, 137: 356-68.
Somasundaram, S., M. E. Forrest, H. Moinova, A. Cohen, V. Varadan, T. LaFramboise, S. Markowitz, and A. M. Khalil. 2018. 'The DNMT1-associated lincRNA DACOR1 reprograms genome-wide DNA methylation in colon cancer', *Clin Epigenetics*, 10: 127.
Song, J., O. Rechkoblit, T. H. Bestor, and D. J. Patel. 2011. 'Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation', *Science*, 331: 1036-40.
Song, J., M. Teplova, S. Ishibe-Murakami, and D. J. Patel. 2012. 'Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation', *Science*, 335: 709-12.
Soreq, L., A. Guffanti, N. Salomonis, A. Simchovitz, Z. Israel, H. Bergman, and H. Soreq. 2014. 'Long non-coding RNA and alternative splicing modulations in Parkinson's leukocytes identified by RNA sequencing', *PLoS Comput Biol*, 10: e1003517.
Stack, S. M., D. B. Brown, and W. C. Dewey. 1977. 'Visualization of interphase chromosomes', *J Cell Sci*, 26: 281-99.
Stavropoulos, N., N. Lu, and J. T. Lee. 2001. 'A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice', *Proc Natl Acad Sci U S A*, 98: 10232-7.
Stepper, P., G. Kungulovski, R. Z. Jurkowska, T. Chandra, F. Krueger, R. Reinhardt, W. Reik, A. Jeltsch, and T. P. Jurkowski. 2017. 'Efficient targeted DNA methylation with

chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase', *Nucleic Acids Res*, 45: 1703-13.
Stirzaker, C., D. S. Millar, C. L. Paul, P. M. Warnecke, J. Harrison, P. C. Vincent, M. Frommer, and S. J. Clark. 1997. 'Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors', *Cancer Res*, 57: 2229-37.
Strahl, B. D., and C. D. Allis. 2000. 'The language of covalent histone modifications', *Nature*, 403: 41-5.
Stresemann, C., and F. Lyko. 2008. 'Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine', *Int J Cancer*, 123: 8-13.
Struhl, K. 2007. 'Transcriptional noise and the fidelity of initiation by RNA polymerase II', *Nat Struct Mol Biol*, 14: 103-5.
Suetake, I., F. Shinozaki, J. Miyagawa, H. Takeshima, and S. Tajima. 2004. 'DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction', *J Biol Chem*, 279: 27816-23.
Sun, L., L. A. Goff, C. Trapnell, R. Alexander, K. A. Lo, E. Hacisuleyman, M. Sauvageau, B. Tazon-Vega, D. R. Kelley, D. G. Hendrickson, B. Yuan, M. Kellis, H. F. Lodish, and J. L. Rinn. 2013. 'Long noncoding RNAs regulate adipogenesis', *Proc Natl Acad Sci U S A*, 110: 3387-92.
Sun, M., F. Nie, Y. Wang, Z. Zhang, J. Hou, D. He, M. Xie, L. Xu, W. De, Z. Wang, and J. Wang. 2016. 'LncRNA HOXA11-AS Promotes Proliferation and Invasion of Gastric Cancer by Scaffolding the Chromatin Modification Factors PRC2, LSD1, and DNMT1', *Cancer Res*, 76: 6299-310.
Syeda, F., R. L. Fagan, M. Wean, G. V. Avvakumov, J. R. Walker, S. Xue, S. Dhe-Paganon, and C. Brenner. 2011. 'The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1', *J Biol Chem*, 286: 15344-51.
Tajik, A., Y. Zhang, F. Wei, J. Sun, Q. Jia, W. Zhou, R. Singh, N. Khanna, A. S. Belmont, and N. Wang. 2016. 'Transcription upregulation via force-induced direct stretching of chromatin', *Nat Mater*, 15: 1287-96.
Takayama, K., K. Horie-Inoue, S. Katayama, T. Suzuki, S. Tsutsumi, K. Ikeda, T. Urano, T. Fujimura, K. Takagi, S. Takahashi, Y. Homma, Y. Ouchi, H. Aburatani, Y. Hayashizaki, and S. Inoue. 2013. 'Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer', *EMBO J*, 32: 1665-80.
Takebayashi, S., T. Tamura, C. Matsuoka, and M. Okano. 2007. 'Major and essential role for the DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions', *Mol Cell Biol*, 27: 8243-58.
Takeshita, K., I. Suetake, E. Yamashita, M. Suga, H. Narita, A. Nakagawa, and S. Tajima. 2011. 'Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1)', *Proc Natl Acad Sci U S A*, 108: 9055-9.
Tan, M. C., J. Widagdo, Y. Q. Chau, T. Zhu, J. J. Wong, A. Cheung, and V. Anggono. 2017. 'The Activity-Induced Long Non-Coding RNA Meg3 Modulates AMPA Receptor Surface Expression in Primary Cortical Neurons', *Front Cell Neurosci*, 11: 124.
Tan, W., B. Liu, S. Qu, G. Liang, W. Luo, and C. Gong. 2018. 'MicroRNAs and cancer: Key paradigms in molecular therapy', *Oncol Lett*, 15: 2735-42.

Tartaglia, G. G., A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, and M. Vendruscolo. 2008. 'Prediction of aggregation-prone regions in structured proteins', *J Mol Biol*, 380: 425-36.

Taunton, J., C. A. Hassig, and S. L. Schreiber. 1996. 'A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p', *Science*, 272: 408-11.

Thomas, G., K. B. Jacobs, M. Yeager, P. Kraft, S. Wacholder, N. Orr, K. Yu, N. Chatterjee, R. Welch, A. Hutchinson, A. Crenshaw, G. Cancel-Tassin, B. J. Staats, Z. Wang, J. Gonzalez-Bosquet, J. Fang, X. Deng, S. I. Berndt, E. E. Calle, H. S. Feigelson, M. J. Thun, C. Rodriguez, D. Albanes, J. Virtamo, S. Weinstein, F. R. Schumacher, E. Giovannucci, W. C. Willett, O. Cussenot, A. Valeri, G. L. Andriole, E. D. Crawford, M. Tucker, D. S. Gerhard, J. F. Fraumeni, Jr., R. Hoover, R. B. Hayes, D. J. Hunter, and S. J. Chanock. 2008. 'Multiple loci identified in a genome-wide association study of prostate cancer', *Nat Genet*, 40: 310-5.

Tilgner, H., D. G. Knowles, R. Johnson, C. A. Davis, S. Chakrabortty, S. Djebali, J. Curado, M. Snyder, T. R. Gingeras, and R. Guigo. 2012. 'Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs', *Genome Res*, 22: 1616-25.

Trere, D. 2000. 'AgNOR staining and quantification', *Micron*, 31: 127-31.

Tripathi, V., J. D. Ellis, Z. Shen, D. Y. Song, Q. Pan, A. T. Watt, S. M. Freier, C. F. Bennett, A. Sharma, P. A. Bubulya, B. J. Blencowe, S. G. Prasanth, and K. V. Prasanth. 2010. 'The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation', *Mol Cell*, 39: 925-38.

Tsai, M. C., O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal, and H. Y. Chang. 2010. 'Long noncoding RNA as modular scaffold of histone modification complexes', *Science*, 329: 689-93.

Tseng, Y. Y., B. S. Moriarity, W. Gong, R. Akiyama, A. Tiwari, H. Kawakami, P. Ronning, B. Reuland, K. Guenther, T. C. Beadnell, J. Essig, G. M. Otto, M. G. O'Sullivan, D. A. Largaespada, K. L. Schwertfeger, Y. Marahrens, Y. Kawakami, and A. Bagchi. 2014. 'PVT1 dependence in cancer with MYC copy-number increase', *Nature*, 512: 82-6.

Unterberger, A., S. D. Andrews, I. C. Weaver, and M. Szyf. 2006. 'DNA methyltransferase 1 knockdown activates a replication stress checkpoint', *Mol Cell Biol*, 26: 7575-86.

Valadkhan, S. 2005. 'snRNAs as the catalysts of pre-mRNA splicing', *Curr Opin Chem Biol*, 9: 603-8.

Van Nostrand, E. L., G. A. Pratt, A. A. Shishkin, C. Gelboin-Burkhart, M. Y. Fang, B. Sundararaman, S. M. Blue, T. B. Nguyen, C. Surka, K. Elkins, R. Stanton, F. Rigo, M. Guttman, and G. W. Yeo. 2016. 'Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)', *Nat Methods*, 13: 508-14.

Vaughan, R. M., B. M. Dickson, M. F. Whelihan, A. L. Johnstone, E. M. Cornett, M. A. Cheek, C. A. Ausherman, M. W. Cowles, Z. W. Sun, and S. B. Rothbart. 2018. 'Chromatin structure and its chemical modifications regulate the ubiquitin ligase substrate selectivity of UHRF1', *Proc Natl Acad Sci U S A*, 115: 8775-80.

Venolia, L., and S. M. Gartler. 1983. 'Comparison of transformation efficiency of human active and inactive X-chromosomal DNA', *Nature*, 302: 82-3.

Voges, D., P. Zwickl, and W. Baumeister. 1999. 'The 26S proteasome: a molecular machine designed for controlled proteolysis', *Annu Rev Biochem*, 68: 1015-68.
Waddington, C. H. 2012. 'The epigenotype. 1942', *Int J Epidemiol*, 41: 10-3.
Walker, M. M., and P. E. Wanda. 1987. 'Immunochemical detection of cell cycle synchronization in a human erythroleukemia cell line, K562', *J Histochem Cytochem*, 35: 1143-8.
Walsh, C. P., J. R. Chaillet, and T. H. Bestor. 1998. 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation', *Nat Genet*, 20: 116-7.
Wang, D., N. B. Ulyanov, and V. B. Zhurkin. 2010. 'Sequence-dependent Kink-and-Slide deformations of nucleosomal DNA facilitated by histone arginines bound in the minor groove', *J Biomol Struct Dyn*, 27: 843-59.
Wang, D., Q. Xu, Q. Yuan, M. Jia, H. Niu, X. Liu, J. Zhang, C. Y. Young, and H. Yuan. 2019. 'Proteasome inhibition boosts autophagic degradation of ubiquitinated-AGR2 and enhances the antitumor efficiency of bevacizumab', *Oncogene*, 38: 3458-74.
Wang, J., S. Hevi, J. K. Kurash, H. Lei, F. Gay, J. Bajko, H. Su, W. Sun, H. Chang, G. Xu, F. Gaudet, E. Li, and T. Chen. 2009. 'The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation', *Nat Genet*, 41: 125-9.
Wang, J., L. Hua, M. Guo, L. Yang, X. Liu, Y. Li, X. Shang, and J. Luo. 2017. 'Notable roles of EZH2 and DNMT1 in epigenetic dormancy of the SHP1 gene during the progression of chronic myeloid leukaemia', *Oncol Lett*, 13: 4979-85.
Wang, J., S. T. Jia, and S. Jia. 2016. 'New Insights into the Regulation of Heterochromatin', *Trends Genet*, 32: 284-94.
Wang, K. C., and H. Y. Chang. 2011. 'Molecular mechanisms of long noncoding RNAs', *Mol Cell*, 43: 904-14.
Wang, L., Y. Zhao, X. Bao, X. Zhu, Y. K. Kwok, K. Sun, X. Chen, Y. Huang, R. Jauch, M. A. Esteban, H. Sun, and H. Wang. 2015. 'LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration', *Cell Res*, 25: 335-50.
Wang, M., C. Guo, L. Wang, G. Luo, C. Huang, Y. Li, D. Liu, F. Zeng, G. Jiang, and X. Xiao. 2018. 'Long noncoding RNA GAS5 promotes bladder cancer cells apoptosis through inhibiting EZH2 transcription', *Cell Death Dis*, 9: 238.
Wang, S., Y. Hui, X. Li, and Q. Jia. 2018. 'Silencing of lncRNA CCDC26 Restrains the Growth and Migration of Glioma Cells In Vitro and In Vivo via Targeting miR-203', *Oncol Res*, 26: 1143-54.
Wang, W. J., Y. M. Wang, Y. Hu, Q. Lin, R. Chen, H. Liu, W. Z. Cao, H. F. Zhu, C. Tong, L. Li, and L. Y. Peng. 2018. 'HDncRNA: a comprehensive database of non-coding RNAs associated with heart diseases', *Database (Oxford)*, 2018.
Wang, X., T. Luo, M. Ruan, P. Liu, S. Wang, and W. Zhu. 2016. 'Association of the CCDC26 rs4295627 polymorphism with the risk of glioma: Evidence from 7,290 cases and 11,630 controls', *Mol Clin Oncol*, 4: 878-82.
Wang, X. X., G. C. Guo, X. K. Qian, D. W. Dou, Z. Zhang, X. D. Xu, X. Duan, and X. H. Pei. 2018. 'miR-506 attenuates methylation of lncRNA MEG3 to inhibit migration and invasion of breast cancer cell lines via targeting SP1 and SP3', *Cancer Cell Int*, 18: 171.

Wang, Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang, and K. Zhao. 2008. 'Combinatorial patterns of histone acetylations and methylations in the human genome', *Nat Genet*, 40: 897-903.

Watt, F., and P. L. Molloy. 1988. 'Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter', *Genes Dev*, 2: 1136-43.

Wei, X. B., T. B. Jin, G. Li, T. T. Geng, J. Y. Zhang, C. P. Chen, G. D. Gao, C. Chen, and Y. K. Gong. 2014. 'CCDC26 gene polymorphism and glioblastoma risk in the Han Chinese population', *Asian Pac J Cancer Prev*, 15: 3629-33.

Wen, L. Z., K. Ding, Z. R. Wang, C. H. Ding, S. J. Lei, J. P. Liu, C. Yin, P. F. Hu, J. Ding, W. S. Chen, X. Zhang, and W. F. Xie. 2018. 'SHP-1 Acts as a Tumor Suppressor in Hepatocarcinogenesis and HCC Progression', *Cancer Res*, 78: 4680-91.

Werner, M. S., and A. J. Ruthenburg. 2015. 'Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes', *Cell Rep*, 12: 1089-98.

Wiles, E. T., and E. U. Selker. 2017. 'H3K27 methylation: a promiscuous repressive chromatin mark', *Curr Opin Genet Dev*, 43: 31-37.

Wischnitzer, S. 1973. 'The submicroscopic morphology of the interphase nucleus', *Int Rev Cytol*, 34: 1-48.

Wolf, S. F., D. J. Jolly, K. D. Lunnen, T. Friedmann, and B. R. Migeon. 1984. 'Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation', *Proc Natl Acad Sci U S A*, 81: 2806-10.

Wu, H., L. Yang, and L. L. Chen. 2017. 'The Diversity of Long Noncoding RNAs and Their Generation', *Trends Genet*, 33: 540-52.

Wu, H., Q. F. Yin, Z. Luo, R. W. Yao, C. C. Zheng, J. Zhang, J. F. Xiang, L. Yang, and L. L. Chen. 2016. 'Unusual Processing Generates SPA LncRNAs that Sequester Multiple RNA Binding Proteins', *Mol Cell*, 64: 534-48.

Wutz, A., T. P. Rasmussen, and R. Jaenisch. 2002. 'Chromosomal silencing and localization are mediated by different domains of Xist RNA', *Nat Genet*, 30: 167-74.

Xie, S., Z. Wang, M. Okano, M. Nogami, Y. Li, W. W. He, K. Okumura, and E. Li. 1999. 'Cloning, expression and chromosome locations of the human DNMT3 gene family', *Gene*, 236: 87-95.

Xing, Y. H., R. W. Yao, Y. Zhang, C. J. Guo, S. Jiang, G. Xu, R. Dong, L. Yang, and L. L. Chen. 2017. 'SLERT Regulates DDX21 Rings Associated with Pol I Transcription', *Cell*, 169: 664-78 e16.

Xu, G. L., T. H. Bestor, D. Bourc'his, C. L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo, and E. Viegas-Pequignot. 1999. 'Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene', *Nature*, 402: 187-91.

Xu, M., X. Xu, B. Pan, X. Chen, K. Lin, K. Zeng, X. Liu, T. Xu, L. Sun, J. Qin, B. He, Y. Pan, H. Sun, and S. Wang. 2019. 'LncRNA SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2', *Mol Cancer*, 18: 135.

Yamazaki, T., S. Souquere, T. Chujo, S. Kobelke, Y. S. Chong, A. H. Fox, C. S. Bond, S. Nakagawa, G. Pierron, and T. Hirose. 2018. 'Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation', *Mol Cell*, 70: 1038-53 e7.

Yan, J., D. Chen, X. Chen, X. Sun, Q. Dong, C. Hu, F. Zhou, and W. Chen. 2019. 'Downregulation of lncRNA CCDC26 contributes to imatinib resistance in human gastrointestinal stromal tumors through IGF-1R upregulation', *Braz J Med Biol Res*, 52: e8399.

Yang, F., T. Babak, J. Shendure, and C. M. Disteche. 2010. 'Global survey of escape from X inactivation by RNA-sequencing in mouse', *Genome Res*, 20: 614-22.

Yang, F., X. Deng, W. Ma, J. B. Berletch, N. Rabaia, G. Wei, J. M. Moore, G. N. Filippova, J. Xu, Y. Liu, W. S. Noble, J. Shendure, and C. M. Disteche. 2015. 'The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation', *Genome Biol*, 16: 52.

Yang, J., X. Meng, J. Pan, N. Jiang, C. Zhou, Z. Wu, and Z. Gong. 2018. 'CRISPR/Cas9-mediated noncoding RNA editing in human cancers', *RNA Biol*, 15: 35-43.

Yang, M. Y., T. C. Liu, J. G. Chang, P. M. Lin, and S. F. Lin. 2003. 'JunB gene expression is inactivated by methylation in chronic myeloid leukemia', *Blood*, 101: 3205-11.

Yap, K. L., S. Li, A. M. Munoz-Cabello, S. Raguz, L. Zeng, S. Mujtaba, J. Gil, M. J. Walsh, and M. M. Zhou. 2010. 'Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a', *Mol Cell*, 38: 662-74.

Yarychkivska, O., Z. Shahabuddin, N. Comfort, M. Boulard, and T. H. Bestor. 2018. 'BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo', *J Biol Chem*, 293: 19466-75.

Yin, Q. F., L. Yang, Y. Zhang, J. F. Xiang, Y. W. Wu, G. G. Carmichael, and L. L. Chen. 2012. 'Long noncoding RNAs with snoRNA ends', *Mol Cell*, 48: 219-30.

Yin, W., A. Rossin, J. L. Clifford, and H. Gronemeyer. 2006. 'Co-resistance to retinoic acid and TRAIL by insertion mutagenesis into RAM', *Oncogene*, 25: 3735-44.

Yin, Y., E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schubeler, C. Vinson, and J. Taipale. 2017. 'Impact of cytosine methylation on DNA binding specificities of human transcription factors', *Science*, 356.

Yu, J., Y. Peng, L. C. Wu, Z. Xie, Y. Deng, T. Hughes, S. He, X. Mo, M. Chiu, Q. E. Wang, X. He, S. Liu, M. R. Grever, K. K. Chan, and Z. Liu. 2013. 'Curcumin down-regulates DNA methyltransferase 1 and plays an anti-leukemic role in acute myeloid leukemia', *PLoS One*, 8: e55934.

Yu, J., B. Qin, A. M. Moyer, S. Nowsheen, T. Liu, S. Qin, Y. Zhuang, D. Liu, S. W. Lu, K. R. Kalari, D. W. Visscher, J. A. Copland, S. A. McLaughlin, A. Moreno-Aspitia, D. W. Northfelt, R. J. Gray, Z. Lou, V. J. Suman, R. Weinshilboum, J. C. Boughey, M. P. Goetz, and L. Wang. 2018. 'DNA methyltransferase expression in triple-negative breast cancer predicts sensitivity to decitabine', *J Clin Invest*, 128: 2376-88.

Zeng, J., Y. Luo, M. Yu, J. Li, and Z. Liu. 2017. 'CCDC26 rs4295627 polymorphisms associated with an increased risk of glioma: A meta-analysis', *Adv Clin Exp Med*, 26: 1275-81.

Zhang, L. F., K. D. Huynh, and J. T. Lee. 2007. 'Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing', *Cell*, 129: 693-706.

Zhang, R., J. Erler, and J. Langowski. 2017. 'Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction', *Biophys J*, 112: 450-59.

Zhang, T., S. Cooper, and N. Brockdorff. 2015. 'The interplay of histone modifications - writers that read', *EMBO Rep*, 16: 1467-81.

Zhang, Y., R. Jurkowska, S. Soeroes, A. Rajavelu, A. Dhayalan, I. Bock, P. Rathert, O. Brandt, R. Reinhardt, W. Fischle, and A. Jeltsch. 2010. 'Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail', *Nucleic Acids Res*, 38: 4246-53.

Zhang, Y., X. O. Zhang, T. Chen, J. F. Xiang, Q. F. Yin, Y. H. Xing, S. Zhu, L. Yang, and L. L. Chen. 2013. 'Circular intronic long noncoding RNAs', *Mol Cell*, 51: 792-806.

Zhang, Z. M., S. Liu, K. Lin, Y. Luo, J. J. Perry, Y. Wang, and J. Song. 2015. 'Crystal Structure of Human DNA Methyltransferase 1', *J Mol Biol*, 427: 2520-31.

Zhao, L. J., T. Subramanian, Y. Zhou, and G. Chinnadurai. 2006. 'Acetylation by p300 regulates nuclear localization and function of the transcriptional corepressor CtBP2', *J Biol Chem*, 281: 4183-9.

Zhao, R., T. Nakamura, Y. Fu, Z. Lazar, and D. L. Spector. 2011. 'Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation', *Nat Cell Biol*, 13: 1295-304.

Zhao, Y., and B. A. Garcia. 2015. 'Comprehensive Catalog of Currently Documented Histone Modifications', *Cold Spring Harb Perspect Biol*, 7: a025064.

Zhao, Y., H. Sun, and H. Wang. 2016. 'Long noncoding RNAs in DNA methylation: new players stepping into the old game', *Cell Biosci*, 6: 45.

Zheng, Y., B. T. Joyce, L. Liu, Z. Zhang, W. A. Kibbe, W. Zhang, and L. Hou. 2017. 'Prediction of genome-wide DNA methylation in repetitive elements', *Nucleic Acids Res*, 45: 8697-711.

Zhu, H., T. M. Geiman, S. Xi, Q. Jiang, A. Schmidtmann, T. Chen, E. Li, and K. Muegge. 2006. 'Lsh is involved in de novo methylation of DNA', *EMBO J*, 25: 335-45.

Ziegler, C., and M. Kretz. 2017. 'The More the Merrier-Complexity in Long Non-Coding RNA Loci', *Front Endocrinol (Lausanne)*, 8: 90.

Zong, X., S. Nakagawa, S. M. Freier, J. Fei, T. Ha, S. G. Prasanth, and K. V. Prasanth. 2016. 'Natural antisense RNA promotes 3' end processing and maturation of MALAT1 lncRNA', *Nucleic Acids Res*, 44: 2898-908.