MOLECULAR INVESTIGATION OF BECKWITH-WIEDEMANN SYNDROME AND SILVER-RUSSELL SYNDROME

By

BENOÎT LAN-LEUNG

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

DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences
University of Birmingham
January 2017
The investigation of human imprinting disorders has provided important insights into the role of genomic imprinting in normal health and development. Beckwith-Wiedemann syndrome (BWS) is a congenital overgrowth disorder associated with abnormal function of 11p15.5 imprinted genes that’s result, most commonly, from the epimutation (loss of maternal allele methylation) at the imprinting centre KCNQ1OT1:TSS-DMR (BWS_IC2). In contrast, Silver-Russell syndrome (SRS) is characterised by pre- and postnatal growth retardation and, most commonly, epimutations (loss of paternal allele methylation) at H19/IGF2:IG-DMR. Using Infinium 450K methylation array, I performed methylation profiling at 46 imprinted differentially methylated regions in 90 BWS and 21 SRS patients. I report epimutations at other imprinting centres outside of chromosome 11p15.5 in 40% of BWS_IC2 but not in SRS_IC1. The investigation of the potential underlying causes of this multilocus methylation disturbances (MLID) epigenotype in BWS_IC2 individuals indicated that several factors might contribute to the BWS phenotype and MLID epigenotype. Although not a universal finding, the use of assisted reproductive technology was significantly associated with MLID in my cohort of BWS_IC2 patients. Furthermore, using whole-exome sequencing strategy, I describe new potential candidate genes for trans-acting factors regulating methylation at imprinting DMRs.
DEDICATION

To my family for supporting me with this endeavour.

To Hannah, the love of my life, for being you, for everything, for giving me the most beautiful gift in the world.

To Emilie, my daughter, for your beautiful smile that brighten my days.
AKNOWLEDGEMENTS

I would like to thank my supervisor Professor Eamonn Maher for his advice and support throughout this project. Thanks also to Professor Farida Latif for her supervision of this project.

Thanks to my funders, Marie Skłodowska-Curie, without who this project won’t be possible.

Special thanks to Scottie for the immeasurable help and advices you gave me during my PhD. Special mention to the invaluable coffee time we had, the English you tried to teach me and your (failed) attempt to making me like your Scottish accent.

Thank you Frenchie for all the hard work you have done for me. I won’t forget your kindness and dedication any time soon.

I also wish to thanks my colleagues for your warm welcome, assistance and daily support.

Big thank you ‘Copains malins’, my old-time friends. You have made my life, to some extent, more enjoyable since I’ve known you.

My blondie Hannah for your limitless patience, support and love. You were there with me the entire time. I don’t know what I would do without you.

And finally my family, Mom and Dad, for their love and support.
# TABLE OF CONTENTS

Chapter 1 - General Introduction .......................................................................................... 1

1.1 Epigenetics ..................................................................................................................... 2

1.1.1 DNA methylation ....................................................................................................... 3

1.1.1.1 Life cycle ............................................................................................................. 4

1.1.1.2 The machinery ..................................................................................................... 6

1.1.1.2.1 De novo DNA methylation .................................................................................. 6

1.1.1.2.2 DNA methylation maintenance ........................................................................ 7

1.1.1.2.3 DNA demethylation .......................................................................................... 8

1.1.1.2.3.1 Passive DNA demethylation .......................................................................... 8

1.1.1.2.3.2 Active DNA demethylation ........................................................................... 9

1.1.2 DNA methylation and transcriptional regulation ......................................................... 10

1.1.2.1 Direct regulation .................................................................................................... 11

1.1.2.2 Indirect regulation ................................................................................................ 11

1.2 Genomic imprinting ........................................................................................................ 14

1.2.1 Historical overview .................................................................................................... 15

1.2.2 Life cycle of genomic imprinting ................................................................................ 19

1.2.2.1 Imprinting erasure ............................................................................................. 20

1.2.2.2 Imprinting establishment .................................................................................... 22

1.2.2.3 Imprinting maintenance ....................................................................................... 23

1.2.3 Consequences of aberrant genomic imprinting ......................................................... 24

1.3 Beckwith-Wiedemann Syndrome .................................................................................. 27

1.3.1 Molecular defects in BWS aetiology .......................................................................... 28

1.3.1.1 Disturbances at H19/IGF2:IG-DMR region ......................................................... 32

1.3.1.2 Disturbances at KCNQ1OT1:TSS-DMR region .................................................. 35

1.3.1.3 Disturbances at both H19/IGF2:IG-DMR and KCNQ1OT1:TSS-DMR regions ......... 37

1.3.1.4 Abnormality detection rate in BWS ..................................................................... 37

1.4 Silver-Russell Syndrome ................................................................................................. 38
1.4.1 Molecular defects in SRS aetiology ............................................................ 39
  1.4.1.1 Disturbances at chromosome 11p15.5 ............................................. 40
  1.4.1.2 Disturbances at chromosome 7 ...................................................... 41
  1.4.1.3 Disturbances at other chromosome ............................................... 43
  1.4.1.4 Abnormality detection rate in SRS ................................................ 44

1.5 Imprinting disorders and multilocus methylation disturbances ............. 45

1.6 General summary of this project ............................................................... 47

1.7 Aims of this project .................................................................................. 49

Chapter 2 - General Methods .......................................................................... 51
  2.1 Buffers recipe .......................................................................................... 52
    2.1.1 1X Transfer buffer ......................................................................... 52
    2.1.2 1X PBS-Tween 20 (PBST) ............................................................ 52
    2.1.3 Blocking buffer ............................................................................ 52
    2.1.4 4X Laemmli sample buffer ........................................................... 52
    2.1.5 RIPA ............................................................................................. 52
    2.1.6 2X SSC ........................................................................................ 53
    2.1.7 1X TAE ....................................................................................... 53

  2.2 DNA purification .................................................................................... 53

  2.3 DNA quantification ............................................................................... 54

  2.4 Sodium bisulfite treatment .................................................................... 55

  2.5 Infinium HumanMethylation450 BeadChip ........................................... 56

  2.6 Analysis software ................................................................................ 57

  2.7 DNA fragments analysis of DNA libraries ............................................ 57
    2.7.1 Setting up the chip priming station ............................................... 57
    2.7.2 Loading the gel dye mix ............................................................... 58
    2.7.3 Loading the marker .................................................................... 58
    2.7.4 Loading the ladder and the samples ............................................ 58
    2.7.5 Starting the chip run ................................................................... 59
    2.7.6 Data analysis ............................................................................... 59
Chapter 3 - Development of a novel next generation sequencing strategy for the genomic imprinting methylation profiling of individual with BWS and SRS

3.1 Introduction .................................................................................................................. 61
  3.1.1 What is NGS? ........................................................................................................ 61
  3.1.2 Sequencing-by-synthesis; an Illumina technology ............................................. 62
  3.1.3 What is bisulfite sequencing? ............................................................................. 66
  3.1.4 NGS based methods for DNA methylation profiling ...................................... 67
  3.1.5 DNA methylation profiling using the bisulfite padlock probes .................... 70
  3.1.6 Aim ..................................................................................................................... 74

3.2 Materials and methods ............................................................................................... 76
  3.2.1 DNA purification ............................................................................................... 76
  3.2.2 DNA quantification .......................................................................................... 76
  3.2.3 Sodium bisulfite treatment ............................................................................... 76
  3.2.4 Amplicon size selection ................................................................................... 76
  3.2.5 Preparation of sequencing libraries using bisulfite padlock probes .......... 77
    3.2.5.1 Bisulfite padlock probes design and synthesis ........................................... 77
    3.2.5.2 Bisulfite padlock probes production (for LCSciences only) .................... 78
      3.2.5.2.1 Amplification (for LCSciences only) .................................................. 78
      3.2.5.2.2 Nicking digestion (for LCSciences only) ....................................... 79
      3.2.5.2.3 Denaturing PAGE purification (for LCSciences only) ................. 80
    3.2.5.3 Multiplex capture of bisulfite treated target regions ............................. 82
    3.2.5.4 Circularisation of captured sequences ...................................................... 83
    3.2.5.5 Enzymatic digestion .................................................................................. 83
    3.2.5.6 Capture circle amplification ..................................................................... 83
    3.2.5.7 Sequencing ............................................................................................... 85

3.3 Results .......................................................................................................................... 85
  3.3.1 First experimental design ................................................................................. 85
    3.3.1.1 Design requirements ............................................................................... 85
    3.3.1.2 Padlock probes to synthetise ................................................................... 87
3.3.1.3 Library preparation .............................................................. 87
3.3.1.4 Sequencing ........................................................................... 89
3.3.2 Second experimental design ....................................................... 90
3.3.2.1 Design requirements .......................................................... 90
3.3.2.2 Library preparation .............................................................. 92
3.3.2.3 Sequencing ........................................................................... 94
3.4 Discussion ................................................................................. 95

Chapter 4 - Methylation Profiling in Beckwith-Wiedemann Syndrome and Silver-Russell Syndrome ........................................ 99
4.1 Introduction .................................................................................. 100
4.1.1 DNA methylation profiling using the Infinium HumanMethylation 450K BeadChips .............................................................. 100
4.1.2 Aim .......................................................................................... 103
4.2 Materials and methods ................................................................. 103
4.2.1 Patient blood DNA .................................................................... 103
4.2.2 DNA quantification ................................................................. 104
4.2.3 Sodium bisulfite treatment ....................................................... 104
4.2.4 DNA methylation profiling using HM450K ................................ 104
4.2.4.1 Bioinformatic pipeline to analyse HM450K methylation dataset 105
4.2.4.1.1 HM450K data preprocessing ............................................. 105
4.2.4.1.2 HM450K data batch correction .......................................... 106
4.2.4.1.3 Epimutation methodology ............................................... 106
4.2.4.1.4 DMR hunting .................................................................. 108
4.3 Results ......................................................................................... 109
4.3.1 Data pre-processing ................................................................. 109
4.3.1.1 Normalisation ....................................................................... 109
4.3.1.2 Filtering ............................................................................... 110
4.3.1.3 Batch effect and biological artefacts correction ................... 111
4.3.2 Targeted methylation profiling ........................................................ 114
  4.3.2.1 Methodology validation ............................................................ 115
  4.3.2.2 Polymorphic imprinted DMRs .................................................. 116
  4.3.2.3 Targeted methylation profiling of SRS patients at imprinted DMRs 119
    4.3.2.3.1 SRS with mUPD7 cohort ....................................................... 119
    4.3.2.3.2 SRS with IC1 LOM cohort ...................................................... 119
  4.3.2.4 Targeted methylation profiling of BWS patients at imprinted DMRs 120
    4.3.2.4.1 BWS with pUPD11 cohort ...................................................... 120
    4.3.2.4.2 BWS with IC1 GOM cohort .................................................... 121
    4.3.2.4.3 BWS with IC2 LOM cohort ..................................................... 121
  4.3.3 DMR hunting ................................................................................... 131
4.4 Discussion .............................................................................................. 134
  4.4.1 Preprocessing ................................................................................. 134
  4.4.2 Methodology ................................................................................... 135
  4.4.3 Polymorphic epimutation ................................................................. 136
  4.4.4 Genomic imprinting methylation in SRS_UPD7 ................................ 136
  4.4.5 Genomic imprinting methylation in SRS_IC1 .................................. 136
  4.4.6 Genomic imprinting methylation in BWS_UPD11 ......................... 137
  4.4.7 Genomic imprinting methylation in BWS_IC1 ............................... 137
  4.4.8 Genomic imprinting methylation in BWS_IC2 ............................... 138
  4.4.9 DMR hunting ................................................................................... 140

Chapter 5 - Investigations of the underlying causes of MLID in BWS_IC2... 142
  5.1  Introduction ......................................................................................... 143
    5.1.1 Assisted reproductive technologies and MLID ........................... 143
      5.1.1.1 Ovarian stimulation ............................................................... 143
      5.1.1.2 In vitro fertilisation ................................................................. 145
      5.1.1.3 Intra-cytoplasmic sperm injection ......................................... 145
5.1.1.4 Evidence of ART disrupting genomic imprints ......................... 146
5.1.1.5 Evidence of association between ART and imprinting disorders 147
5.1.1.6 Investigation of potential ART association with MLID in BWS. 148

5.1.2 Trans-imprinting defect and MLID ............................................... 149
5.1.2.1 Maternal-effect and maternal-zygotic effect genes .............. 150
  5.1.2.1.1 Zygotic effect .................................................................. 151
  5.1.2.1.2 Maternal-effect .............................................................. 151
  5.1.2.1.3 Maternal-zygotic effect .................................................. 152
5.1.2.2 NLRP7 .............................................................................. 155
5.1.2.3 ZFP57 ............................................................................... 157
5.1.2.4 NLRP2 .............................................................................. 159
5.1.2.5 KHDC3L ........................................................................... 161
5.1.2.6 NLRP5 .............................................................................. 162
5.1.2.7 Other candidate genes ...................................................... 163

5.1.3 Aim .......................................................................................... 164

5.2 Materials and methods ................................................................. 165
5.2.1 Patient DNA ............................................................................. 165
5.2.2 DNA quantification .................................................................. 165
5.2.3 DNA purification ..................................................................... 165
5.2.4 Sodium bisulfite treatment .................................................... 166
5.2.5 DNA methylation profiling using HM450K ......................... 166
5.2.6 Whole-exome sequencing ....................................................... 166
  5.2.6.1 Library preparation and sequencing ................................ 166
  5.2.6.2 Alignment and variant calling ....................................... 167
  5.2.6.3 Filtering and annotation .................................................. 167
  5.2.6.4 Variant call analysis ........................................................ 168
5.2.7 Sanger sequencing ................................................................. 175
  5.2.7.1 Polymerase chain reaction ............................................. 175
  5.2.7.2 PCR product purification ................................................. 178
5.2.7.3  Sequencing reaction ................................................................. 178
5.2.7.4  Sequencing clean-up ............................................................... 179
5.2.7.5  Sequence electrophoresis ....................................................... 179
5.2.7.6  Sequence data analysis .......................................................... 179

5.3  Results ............................................................................................. 179

5.3.1  Methylation profiling of BWS_IC2 patients conceived with ART .... 179
5.3.2  Genetic investigation of BWS_IC2 patients ................................. 184
  5.3.2.1  Analysis of methylation associated genes ............................. 186
  5.3.2.2  Analysis of maternal effect genes ......................................... 189
  5.3.2.3  Analysis of imprinted genes ................................................ 193

5.4  Discussion ...................................................................................... 196

5.4.1  ART and MLID ............................................................................ 196
5.4.2  Trans-mechanism involving a gene associated with methylation pathway 198
5.4.3  Trans-mechanisms involving maternal-effect genes .................... 199
5.4.4  Trans-mechanism involving imprinting gene network disruption .... 202

Chapter 6 - Investigations of a novel candidate gene for BWS ............ 205

6.1  Introduction .................................................................................... 206

6.1.1  UHRF1, a critical epigenetic modifier ........................................ 206
6.1.2  Case Report .............................................................................. 209
6.1.3  Aim ............................................................................................. 210

6.2  Materials and methods ................................................................. 211

6.2.1  Cells .......................................................................................... 211
6.2.2  Plasmids ................................................................................... 212
6.2.3  Maintenance of adherent cells ................................................. 213
6.2.4  Transient transfection .............................................................. 214
6.2.5  Quantitative PCR ........................................................................ 215
  6.2.5.1  RNA extraction from adherent cells ................................. 215
  6.2.5.2  DNase treatment ................................................................. 216
6.2.5.3  Bioanalyzer RNA 6000 Nano Kit quality check and quantification

6.2.5.3.1  Setting up the chip priming station .......................................................... 217
6.2.5.3.2  Loading the gel dye mix ........................................................................ 217
6.2.5.3.3  Loading the marker ............................................................................. 217
6.2.5.3.4  Loading the ladder and the samples ...................................................... 217
6.2.5.3.5  Starting the chip run ............................................................................. 218
6.2.5.3.6  Data analysis ....................................................................................... 218

6.2.5.4  cDNA synthesis ....................................................................................... 218

6.2.5.5  Quantitative PCR .................................................................................... 219
6.2.5.5.1  Samples preparation ............................................................................. 219
6.2.5.5.2  Serial dilution preparation for standard curve ..................................... 219
6.2.5.5.3  Reaction set-up and PCR amplification .............................................. 220
6.2.5.5.4  Data analysis using the relative standard curve .................................. 221

6.2.6  Protein assay .............................................................................................. 222
6.2.6.1  Protein extraction from adherent cells ................................................... 222
6.2.6.2  Protein quantification ............................................................................. 222
6.2.6.3  Western Blot ........................................................................................... 223
6.2.6.3.1  SDS-PAGE ......................................................................................... 223
6.2.6.3.2  Transfer ............................................................................................... 224
6.2.6.3.3  Immunoblotting ................................................................................... 224
6.2.6.3.4  Membrane revelation ........................................................................... 225
6.2.6.3.5  Gel analysis ........................................................................................ 225

6.2.7  Methylation assay ...................................................................................... 225
6.2.7.1  DNA extraction from adherent cells ...................................................... 225
6.2.7.2  Methylation-sensitive restriction digestion assay ................................... 226
6.2.7.3  Manual DNA dot-blot ........................................................................... 227
6.2.7.3.1  DNA spotting ...................................................................................... 227
6.2.7.3.2  5-mC visualisation .............................................................................. 228
6.2.7.3.3  DNA quantity assessment .................................................................. 228
6.2.7.3.4 Membrane analysis................................................................. 229
6.2.7.4 Targeted bisulfite sequencing assay ....................................... 229
6.2.7.4.1 Library preparation and sequencing....................................... 229
  6.2.7.4.1.1 First round PCR: target capture ........................................... 229
  6.2.7.4.1.2 Purification of PCR products............................................... 230
  6.2.7.4.1.3 Second round PCR: adaptor tagging and indexing .......... 231
  6.2.7.4.1.4 Purification of PCR products............................................... 231
  6.2.7.4.1.5 Sequencing libraries quality assessment................................. 232
  6.2.7.4.1.6 Sequencing........................................................................ 232
6.2.7.4.2 Processing and analysis of NGS data .................................... 232
6.2.7.4.3 5-mC quantification................................................................. 233
6.2.7.4.4 PCR bias correction for 5-mC methylation level ..................... 233
6.3 Results ............................................................................................... 233
  6.3.1 Transfection optimisation................................................................. 233
  6.3.2 Low level of UHRF1 mRNA and protein in HeLa shUHRF1 ......... 234
  6.3.3 Transfection of UHRF1 does not rescue global methylation ....... 236
  6.3.4 Transfection of UHRF1 does not rescue methylation at IGS-rDNA 239
6.4 Discussion ............................................................................................ 243

Chapter 7 - General discussion ........................................................................ 247
  7.1 Multilocus imprinting disturbances in imprinting disorders, a need for
  standardisation ........................................................................................... 248
  7.1.1 Where to look? ................................................................................ 248
  7.1.2 Too many platforms might lead to confusion .................................. 249
  7.1.3 A common informatic approach to detect epimutations ............... 249
  7.1.4 Need to find a better alternative for diagnostic services ............... 250
  7.2 The possible underlying causes of BWS and MLID ......................... 252
    7.2.1 Environmental pressure ................................................................. 252
    7.2.2 Genetic component ...................................................................... 253
    7.2.3 Methylation abnormalities outside of imprinted DMRs ............... 255
7.2.4 5-hydroxymethylation, a hot topic for the future? ........................................ 256
7.3 Final conclusions ........................................................................................................ 258

Chapter 8 - Appendix ........................................................................................................ 259
Appendix 1 - R code used for methylation data preprocessing. .............................. 260
Appendix 2 - R code used for batch effect correction. ............................................. 262
Appendix 3 - R code used for genomic imprinting methylation profiling. .......... 264
Appendix 4 - R code used for bump hunting in BWS_IC1. Similar code was used for other disease group. .......................................................... 276
Appendix 5 - Shell script used for the trimming, alignment and methylation calling of bisulfite converted reads generated from MiSeq sequencer. ............... 278
Appendix 6 - R code used for the quantification of 5-mC at H19/IGF2:IG-DMR; KCNQ1OT1:TSS-DMR and IGS-rDNA in each sample................................. 280
Appendix 7 - R code used for the PCR amplification bias correction. ............... 286

Chapter 9 - List of References ....................................................................................... 289
Figure 1.1. Interaction between RNA, histone modification and DNA methylation in heritable gene silencing.............................................................................................................. 3
Figure 1.2. Dynamic changes in DNA methylation during PGCs maturation, fertilisation and embryo development................................................................. 6
Figure 1.3. De novo methylation mechanism. .......................................................... 7
Figure 1.4. DNA methylation maintenance mechanism. ....................................... 8
Figure 1.5. DNA demethylation mechanism......................................................... 10
Figure 1.6. DNA methylation-mediated gene silencing. ....................................... 13
Figure 1.7. Genomic imprinting. ......................................................................... 19
Figure 1.8. Life-cycle of genomic imprinting....................................................... 20
Figure 1.9. Molecular mechanisms leading to imprinting disorders. ................. 27
Figure 1.10. Localisation of the three breakpoints associated to BWS at chromosome 11p15............................................................... 30
Figure 1.11. Linear map of the BWS locus on 11p15.5........................................ 31
Figure 1.12. Proposed model for the monoallelic expression of IGF2 and H19.... 34
Figure 1.13. Enrichment of the type of abnormalities describe in the molecular aetiology of BWS................................................................. 38
Figure 1.14. Enrichment of the type of abnormalities describe in the molecular aetiology of SRS.................................................................................... 45
Figure 3.1. Sequencing-by-synthesis technology used in Illumina....................... 65
Figure 3.2. Principle of sodium bisulfite treatment. ........................................... 67
Figure 3.3. Bisulfite padlock probes................................................................. 73
Figure 3.4. Preliminary steps used for the probes syntthesised by LCSiences. ... 82
Figure 3.5. BSPP sequencing library preparation (first design). ......................... 88
Figure 3.6. BSPP sequencing library preparation (second design)..................... 93
Figure 4.1: Overview of the Infinium I and Infinium II bead types. .................... 102
Figure 4.2: Overview of the genome-wide methylation profiling workflow........ 105
Figure 4.3: Illustration of epimutation methodology ............................................. 107
Figure 4.4: Effect of the correction on methylation value. ................................. 110
Figure 4.5: Proportion of retained and removed probes................................. 111
Figure 4.6: Effects of technical and biological artefacts. .................................. 114
Figure 4.7: Methodology validation. ................................................................. 116
Figure 4.8: Multilocus imprinting disturbances and SRS. .............................. 120
Figure 4.9: Multilocus imprinting disturbances and BWS. ............................. 122
Figure 4.10: Comparison of the frequency of epimutated DMRs in BWS_IC2 with MLID patients. ................................................................. 123
Figure 4.11: Meta-analysis of reported MLID frequency in BWS_IC2 cohort ...... 126
Figure 4.12: Imprinting disturbances in BWS_IC2 patients ......................... 128
Figure 4.13: Epimutations correlation calculated at epimutated imprinted DMRs in BWS_IC2 MLID patients. ......................................................... 130
Figure 4.14: DMR hunting in BWS. ................................................................. 132
Figure 4.15: KCNAB3 is also aberrantly methylated in BWS_IC2. .......... 134
Figure 5.1: Maternal to zygotic gene transition in the mouse ........................... 151
Figure 5.2: Overview of the whole-exome sequencing workflow. ..................... 169
Figure 5.3: ART and genomic imprinting methylation in BWS_IC2 ............... 181
Figure 5.4: Significance of ART of MLID epigenotype in BWS_IC2 patients ... 182
Figure 5.5: ART is not associated with a specific epimutation at imprinted DMRs in BWS_IC2 cohort ................................................................. 183
Figure 5.6: Distribution of the sequencing depth of read across the 42 BWS_IC2 individuals ................................................................. 184
Figure 5.7: Proportion of the coverage of the exons of gene associated with the methylation pathway ................................................................. 188
Figure 5.8: Validation of the variants found in gene associated with the methylation pathway ................................................................. 189
Figure 5.9: Proportion of the coverage of the exons of maternal-effect genes ... 192
Figure 5.10: Validation of the variants found in maternal-effect genes .......... 193
Figure 5.11: Proportion of the coverage of the exons of imprinted genes ....... 195
Figure 5.12: Validation of the variants found in imprinted genes.................... 196
Figure 6.1: Domain structure of UHRF1............................................................. 208
Figure 6.2: Case report of two siblings with BWS_IC2 and a homozygous mutation in the UHRF1 gene................................................................. 210
Figure 6.3: HeLa cells transfection optimisation using pRP plasmid................. 234
Figure 6.4: Efficient knockdown of UHRF1 in HeLa cells.............................. 235
Figure 6.5: Transfection of pRP UHRF1 WT does not rescue global methylation. ........................................................................................................... 238
Figure 6.6: Transfection of pTag-2C UHRF1 WT does not rescue global methylation. ........................................................................................................... 239
Figure 6.7: Degree of bias introduced by PCR amplification............................ 241
Figure 6.8: Transfection of pTag-2C UHRF1 WT does not rescue methylation at KCNQ1OT1:TSS-DMR and IGS-rDNA.................................................. 242
### LIST OF TABLES

Table 3.1: List of imprinted DMRs. ................................................................. 75  
Table 3.2: List of primers used for the amplification of the padlock probes. ....... 79  
Table 3.3: List of primers used for sequencing libraries enrichment. .................. 84  
Table 3.4: List of primers used for sequencing using MiSeq sequencer. .............. 85  
Table 3.5: Number of padlock probes to synthesise. ....................................... 87  
Table 4.6: An example of statistical test at the imprinted locus PLAGL1_alt-TSS-DMR. ......................................................................................... 108  
Table 4.7: Age distribution in disease and normal individual groups. ................. 112  
Table 4.8: Imprinted DMRs with polymorphic epimutations. ............................ 118  
Table 4.9: Frequency of disturbances reported in the current and seven previous reports. .............................................................................................................. 125  
Table 4.10: Candidate differentially methylated regions in BWS and SRS. .......... 133  
Table 5.12: List of methylation associated genes. ........................................... 170  
Table 5.13: List of imprinted genes. ............................................................... 174  
Table 5.14: Primers used for Sanger sequencing validation. ............................. 177  
Table 5.15: Summary of interesting variants found by whole-exome sequencing. ............................................................................................................................ 185  
Table 6.16: shRNA clones used in HeLa cells. ................................................. 212  
Table 6.17: List of primers used to quantify relative UHRF1 mRNA level........... 221  
Table 6.18: List of primers used for targeted bisulfite sequencing. .................... 230
LIST OF ABBREVIATIONS

5-caC  5-carboxylcytosine
5-fC   5-formylcytosine
5-hmC  5-hydroxymethylcytosine
5-mC   5-methylcytosine
ART    assisted reproductive technology
AS     Angelman syndrome
bisDNA bisulfite treated deoxyribonucleic acid
bp     base pair
BSPP   bisulfite padlock probes
BWS    Beckwith-Wiedemann syndrome
BWS_IC1 Beckwith-Wiedemann syndrome with gain-of-methylation at
         H19/IGF2:IG-DMR
BWS_IC2 Beckwith-Wiedemann syndrome with loss-of-methylation at
         KCNQ1OT1:TSS-DMR
BWS_UPD11 Beckwith-Wiedemann syndrome with paternal uniparental disomy 11
COH    controlled ovarian hyperstimulation
CpG    cytosine-phosphate-guanine
dH₂O   distilled water
DMR    differentially methylated region
DNA    deoxyribonucleic acid
ESP    NHLBI Exome Sequencing Project
ExACT  Exome Aggregation Consortium
FBHM   Familial biparental hydatidiform moles
FDR    False discovery rate
FWER   Family-wise error rate
gDNA   genomic deoxyribonucleic acid
GOM    gain-of-methylation
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM450K</td>
<td>Infinium HumanMethylation 450K BeadChips</td>
</tr>
<tr>
<td>ICSI</td>
<td>intra-cytoplasmic sperm injection</td>
</tr>
<tr>
<td>ID</td>
<td>imprinting disorder</td>
</tr>
<tr>
<td>IMAGe</td>
<td>Intrauterine Growth Restriction, Metaphyseal Dysplasia, Adrenal Hypoplasia Congenita, and Genital Anomalies</td>
</tr>
<tr>
<td>indel</td>
<td>insertion deletion</td>
</tr>
<tr>
<td>IVF</td>
<td><em>in vitro</em> fertilisation</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>KOS</td>
<td>Kagami-Ogata syndrome</td>
</tr>
<tr>
<td>LOM</td>
<td>loss-of-methylation</td>
</tr>
<tr>
<td>Mb</td>
<td>megabase</td>
</tr>
<tr>
<td>MeDIP</td>
<td>Methylated DNA immunoprecipitation</td>
</tr>
<tr>
<td>MI</td>
<td>methylation index</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>MLID</td>
<td>multilocus imprinting disturbances</td>
</tr>
<tr>
<td>NGS</td>
<td>next-generation sequencing</td>
</tr>
<tr>
<td>OS</td>
<td>ovulatory stimulation</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PGC</td>
<td>primordial germ cell</td>
</tr>
<tr>
<td>PHD</td>
<td>plant homeodomain</td>
</tr>
<tr>
<td>PHP1B</td>
<td>pseudohypoparathyroidism type Ib</td>
</tr>
<tr>
<td>PWS</td>
<td>Prader-Willi syndrome</td>
</tr>
<tr>
<td>RING</td>
<td>really interesting new gene</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RRBS</td>
<td>reduced representation bisulfite sequencing</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>sec</td>
<td>second</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SRA</td>
<td>SET and RING-associated</td>
</tr>
<tr>
<td>SRS</td>
<td>Silver-Russell syndrome</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SRS_IC1</td>
<td>Silver-Russell syndrome with loss-of-methylation at H19/IGF2:IG-DMR</td>
</tr>
<tr>
<td>SRS_UPD7</td>
<td>Beckwith-Wiedemann syndrome with maternal uniparental disomy 7</td>
</tr>
<tr>
<td>TBS-seq</td>
<td>targeted bisulfite sequencing</td>
</tr>
<tr>
<td>TNDM1</td>
<td>Transient neonatal diabetes mellitus type 1</td>
</tr>
<tr>
<td>TRC</td>
<td>The RNAi Consortium</td>
</tr>
<tr>
<td>TS</td>
<td>Temple syndrome</td>
</tr>
<tr>
<td>TTD</td>
<td>tandem tudor domain</td>
</tr>
<tr>
<td>WES</td>
<td>whole-exome sequencing</td>
</tr>
<tr>
<td>WGBS</td>
<td>whole genome bisulfite sequencing</td>
</tr>
<tr>
<td>WGS</td>
<td>whole-genome sequencing</td>
</tr>
</tbody>
</table>
Chapter 1

General Introduction
1.1 Epigenetics

The term epigenetics was first coined by Waddington in 1942 (Waddington 1942). Originally, it referred to the complex mechanisms at work that link genotype to phenotype. Since then, the term has changed in meaning and the most widely accepted definition designates every mechanism that results in reversible heritable (via meiosis or mitosis) changes altering gene expression without alterations to the DNA sequence itself (Probst et al. 2009).

The epigenetic mechanisms capable of switching genes from an ON to OFF state are initiated and sustained by at least three systems. This includes not only DNA methylation, which is the most broadly studied and well-characterised epigenetic modification, but also the post-translational modification of histone proteins and RNA-mediated gene silencing. These systems are highly interactive with each other in order to maintain efficient gene expression regulation (Figure 1.1). Any unscheduled compromise at any of these levels can have dramatic effects on gene expression and in fine leads to disease.
Figure 1.1. Interaction between RNA, histone modification and DNA methylation in heritable gene silencing.

Chromatin condensation resulting from histone deacetylation or the methylation of the lysine 9 within histone H3 residues can lead to transcriptional repression. The post-translational modification of histones can also attract and trigger the DNA methylation machinery (DNA methyltransferases) which in turn can reinforce the transcriptional silencing initiated by the histone modification patterns. RNA interference could also trigger chromatin remodelling through the modification of histones or the methylation of DNA and therefore results in gene silencing. (Adapted from Egger et al. 2004).

1.1.1 DNA methylation

The methylation of DNA is a heritable epigenetic mark referring to the covalent transfer of a methyl group from S-adenosyl-L-methionine to the fifth carbon of a cytosine. Consequently, this modification results in a 5-methylcytosine (5-mC) (Messerschmidt et al. 2014). In mammals, DNA methylation is most frequently found
in symmetrical cytosine-phosphate-guanine (CpG) context, although rare non-CpG
(CpH, where H is adenine, thymine or cytosine) methylation occurs (Ramsahoye et

In somatic cells, DNA methylation generally affects 70-80% of all CpG sites. Highly
methylated sequences are found at repetitive genomic regions, including satellite
dNA, long interspersed transposable elements (LINEs) and short interspersed
transposable elements (SINEs), at non-repetitive intergenic DNA and at exons of
genes. In contrast, CpG islands (CGI), which are on average 1000 base pairs long
GC-rich regions with high CpG density, are generally unmethylated. CGI are found
predominantly at sites of transcription initiation including the gene promoters of
approximatively 60% of human genes (Bird 2002; Deaton and Bird 2011; E. Li and
Y. Zhang 2014).

Historically, DNA methylation has been linked to transcriptional repression but other
critical roles in the control of genomic imprinting (Bartolomei and Ferguson-Smith
2011; Ferguson-Smith 2011) (see section 1.2 - Genomic imprinting) and the
inactivation of the X chromosome in females (Sharp et al. 2011) or the maintenance
of the genome integrity (Eden et al. 2003) have been described.

1.1.1.1 Life cycle

In the life cycle of an individual, the genome undergoes several dynamic changes in
DNA methylation (Figure 1.2). These modifications take place at different times and
involve different pathways. During the maturation of the primordial germ cells (PGCs) of an embryo, DNA methylation is erased through genome-wide demethylation event. Following sex determination, the male (blue line) and female (red line) germ cells acquire new gametic methylation profile via global de novo DNA methylation. This process takes place at different times between the male and female germ cells. In the male embryo, the de novo DNA methylation is established during prospermatogonia and is completed before birth. In the female embryo, the de novo DNA methylation is established after birth during oocytes growth. In the zygote, a second wave of epigenetic reprogramming occurs. During the pre-implantation period, genome-wide DNA methylation marks from the paternal and maternal genome are erased. This mechanism takes place at different times for both parental genomes. After implantation, the embryo re-acquires a new epigenetic profile, which is associated with cell differentiation, through genome-wide de novo methylation. The DNA methylation in genomic imprinted regions escapes this second global epigenetic event (Smallwood and Kelsey 2012) (Figure 1.2).
Figure 1.2. Dynamic changes in DNA methylation during PGCs maturation, fertilisation and embryo development. (From Smallwood and Kelsey 2012).

1.1.1.2 The machinery

1.1.1.2.1 De novo DNA methylation

In mammals, DNA methylation is catalysed by DNA methyltransferase (DNMT) enzymes. The methylation patterns are established during embryonic development by the de novo methylating enzymes DNMT3A and DNMT3B (Okano et al. 1998; Okano et al. 1999) and the non-catalytic activating co-factor DNMT3L (Bourch'his et al. 2001; Bourch'his and Bestor 2004) (Figure 1.3). The deletion of mouse Dnmt3b results in substantial global loss-of-methylation and embryonic lethality (E. Li et al. 1992; Okano et al. 1999; Bostick et al. 2007). Similarly, the Dnmt3a knockout mice lack methylation but are, for a few weeks, partially viable (Okano et al. 1999). Finally, Dnmt3L knockout mice are viable, but lack de novo methylation in the germline.
associated with male sterility and embryonic lethality of maternal null-derived embryos (Bourc'his et al. 2001; Hata et al. 2002; Bourc'his and Bestor 2004).

![Image](image.png)

**Figure 1.3. De novo methylation mechanism.**
De novo methylation is catalysed by the DNA methyltransferases DNMT3A, DNMT3B and supported by the non-catalytic DNMT3L.

1.1.1.2.2 DNA methylation maintenance
During replication, the maintenance methylating enzyme DNMT1 localises at the replication fork on the parental methylated strand and restores methylation to the newly synthesised unmethylated strand (Leonhardt et al. 1992). For the faithful DNMT1-mediated reproduction of the methylation profile that is present in the parent cell, the maintenance mechanism also requires the protein UHRF1 (ubiquitin like with PHD and ring finger domains 1; also known as NP95 or ICBP90) which recognises and targets DNMT1 to hemi-methylated CpG residues (Figure 1.4). The deletion of mouse Dnmt1 or Uhrf1 results in substantial global loss-of-methylation and embryonic lethality (E. Li et al. 1992; Okano et al. 1999; Bostick et al. 2007).
Figure 1.4. DNA methylation maintenance mechanism.
The DNA methyltransferase DNMT1, supported by UHRF1, maintains the DNA methylation at hemi-methylated CpG during DNA replication

1.1.1.2.3 DNA demethylation
DNA demethylation can be achieved by one of two mechanisms; a passive replication-dependant process or an active enzyme-catalysed and replication-independent process.

1.1.1.2.3.1 Passive DNA demethylation
Global loss of 5-mC in mitotic cells can be achieved through the gradual passive dilution of the methylation mark over subsequent cleavage division due to DNA replication. Additionally, the gradual loss of 5-mC is performed in the absence or exclusion of both de novo and maintenance DNA methylase activity, including DNMT1, DNMT3A, DNMT3B and DNMT3L, and the key cofactor UHRF1 (P.A. Jones and Taylor 1980; Seisenberger et al. 2012; Kagiwada et al. 2013).
1.1.1.2.3.2 **Active DNA demethylation**

Loss of 5-mC can also be achieved via indirect and active mechanisms. Active genome-wide DNA demethylation is mostly achieved by the iterative oxidation of 5-mC to 5-hydroxymethylcytosine (5-hmC) and further to 5-formylcytosine (5-fmC) and 5-carboxycytosine (5-caC) by the ten-eleven translocation enzymes, TET1 and TET2 (He et al. 2011; Inoue et al. 2011; Ito et al. 2011; Hackett et al. 2013). 5-hmC, 5-fmC and 5-caC marks are then lost during replication since it is inefficiently recognised by DNMT1 (Inoue et al. 2011; Hashimoto et al. 2012) (Figure 1.5-bottom panel, red arrow). Alternatively, an enzyme-catalysed DNA demethylation mechanism involves the deamination of 5-mC to thymine by the DNA deaminases AID (activation-induced cytidine deaminase) and APOBEC1 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1). The thymine base resulting from the resulting T:G mismatch is recognised and removed by the thymine-DNA glycosylase (TDG) or the methyl-CpG-binding domain protein 4 (MBD4). The base excision repair (BER) machinery recognises the abasic site and reinstates an unmodified cytosine, effectively resulting in removal of the methyl mark (Morgan et al. 2004; Maiti and Drohat 2011) (Figure 1.5-top panel, red arrow). Finally, it is also suggested that TDG, MBD4 or a yet unidentified glycosylase can directly recognise and remove the 5-hmC, 5-fmC and 5-caC marks. The BER pathway will then repair the resulting abasic site and reinstate an unmodified cytosine (He et al. 2011; Inoue et al. 2011; Maiti and Drohat 2011; Shen et al. 2013; Messerschmidt et al. 2014) (Figure 1.5-bottom panel, red arrow).
Figure 1.5. DNA demethylation mechanism.

Passive DNA methylation is achieved via absence of DNMT and UHRF1 during subsequent round of cell division (middle panel, blue arrow). Possible active DNA demethylation mechanisms: AID/APOBEC1 can deaminate 5-mC to thymine, which in return can be excised by TDG/MBD4 and repaired by BER (top panel, red arrows). Alternatively, 5-mC can be lost via the TET-mediated oxidation of 5-mC to 5-hmC/5-fC/5-caC. The iterative oxidised 5-mC is then lost during replication or removed by TDG/MBD4 and repaired by BER (bottom panel, red arrows).

1.1.2 DNA methylation and transcriptional regulation

Several lines of evidence demonstrated that DNA methylation can regulate gene expression via either a direct or indirect processes.
1.1.2.1 Direct regulation

DNA methylation can mediate gene silencing through the direct inhibition of the binding of specific transcription factors to their targets (Figure 1.6). This was first demonstrated in HeLa cells in which the binding of the transcription factor MLTF (major late transcription factor) to DNA was affected by DNA methylation, hence leading in the silencing of the adenovirus major late promoter (Watt and Molloy 1988). The investigation of additional transcription factors, including CREB (cAMP response element-binding protein), c-Myc and E2F (E2 factor), further validated the model (Iguchi-Ariga and Schaffner 1989; Prendergast and Ziff 1991; Campanero et al. 2000).

1.1.2.2 Indirect regulation

DNA methylation mediated gene silencing can also be achieved indirectly through the recruitment of methyl-CpG-binding domain (MBD) proteins, including MeCP2 (methyl-CpG-binding proteins 2), MBD1, MBD2, MBD3 and MBD4, at methylated promoter regions. In return, MBD proteins recruit at methylated sites additional repressor complexes associated with the remodelling of chromatin and the covalent modification of histones. Finally, this mechanism results in the compaction of the chromatin (i.e. heterochromatin state) leading to transcriptional repression (E. Li and Y. Zhang 2014) (Figure 1.6-a). Several lines of evidence have validated this model. The protein MeCP2 recruits the co-repressor molecule SIN3A (transcriptional regulator, SIN3A) and the histone deacetylases (HDAC) HDAC1 and HDAC2 at CpG methylated sites. The consequent transcriptional repression is then achieved
via HDAC-mediated chromatin remodelling (P.L. Jones et al. 1998; Nan et al. 1998). Furthermore, the protein MBD2 was shown to recruit the nucleosome remodelling and deacetylase (NuRD) co-repressor complex, which comprises HDAC enzymes and large ATP-dependent chromatin remodelling proteins CHD3 (chromodomain helicase DNA binding protein 3) and CHD4 (chromodomain helicase DNA binding protein 4), at CpG methylated sites. This results in the deacetylation of histone proteins, the compaction of chromatin and consequently to gene silencing (Feng and Y. Zhang 2001). Finally, the protein MBD1 was shown to recruit the histone lysine methyltransferase SETDB1 (SET domain bifurcated 1) during the S phase of the cell cycle when DNA replication occurs. Subsequently, SETDB1 methylates the lysine 9 of the histone H3 (H3K9me) hence resulting in transcriptional repression (Sarraf and Stancheva 2004).

DNA methylation can also indirectly mediate gene silencing through the DNA methylation-dependant binding of the insulator protein CTCF (CCCTC-binding factor) (Bell et al. 1999; Bell and Felsenfeld 2000; Hark et al. 2000; Murrel et al. 2004; Kurukuti et al. 2006). The absence of methylation allows the protein CTCF to bind to the DNA. The binding of CTCF can lead to the formation of chromatin loops and the establishment of an active chromatin domain that could include genes, gene promoters and enhancers and an inactive chromatin domain that could include genes and gene promoters, away from enhancers. The interactions between gene promoters located in the inactive chromatin domain and their enhancers located in the active chromatin domain are blocked by the CTCF insulating activity. This
mechanism results in genes silencing (Figure 1.6-b, left panel). In contrast, DNA methylation prevents the protein CTCF to bind to the DNA and the creation of the associated chromatin boundaries. Gene promoters previously located within the inactive chromatin domain are now free to interact with their enhancers, hence leading to gene expression (Figure 1.6-b, right panel).

Figure 1.6. DNA methylation-mediated gene silencing.
(a) DNA methylation inhibits binding of transcription factors, hence leading to gene silencing. Alternatively, DNA methylation can recruit MBD proteins which in turn associate with histone modification proteins, hence inducing chromatin remodelling and gene silencing. (b) The absence of DNA methylation allows the binding of the insulator protein CTCF which in turn blocks interaction between gene promoters and enhancers, hence leading to genes silencing. DNA methylation inhibits the binding of CTCF which in turn allows interaction between gene promoters and enhancers, hence leading to genes expression. Chromatin looping is the proposed mechanism by which CTCF boundary elements separate active and silent and active chromatin domains (grey area).
1.2 Genomic imprinting

Mammals are diploid organisms whose cells possess two matched sets of chromosomes, one inherited from the mother and one from the father. Thus, mammals have two copies of every gene. In usual physiological state, most genes are biallelically expressed which means that both of the maternal and paternal allele of the gene is active. A minority of genes, probably numbering 100 in humans and in mice, are subject to genomic imprinting and show differences in expression according to the parental origin of the allele (Catalogue of Parent of Origin Effects, http://igc.otago.ac.nz/home.html (Morison et al. 2001)).

The machinery regulating genomic imprinting must fulfil different properties. It must influence transcription via the deliberate silencing of one allele of specific genes according to whether the allele comes from the father or the mother. The imprinting marks need to be placed onto parental inherited chromosomes at a time they are not in the same nucleus (i.e. during gametogenesis). Subsequently, following fertilisation the marks needs to be transmitted to the somatic lineages of the male and female offspring. Finally, an erasure mechanism must ensure that the imprinting marks are removed and replaced in the new female or male germline (Bartolomei and Ferguson-Smith 2011; Ferguson-Smith 2011). Considering these properties, DNA methylation fulfils these criteria and is described as one of the main mechanisms regulating genomic imprinting. However, other mechanisms including allele-specific RNA transcription, antisense transcripts, histone modifications and differences in
replication timing are also thought to be involved in the genomic imprinting regulatory
machinery (Barlow and Bartolomei 2014).

1.2.1 Historical overview

The notion of genomic imprinting was first described in a series of experiments
aiming to understand the inheritance of the maize kernel coloration (Kermicle 1970). The author noted that the maize kernels were solidly coloured when the R allele (i.e. the allele responsible for the coloration) was carried by the female gametes but, surprisingly, were mottled when carried by the male gametes. The author suggested that the R allele might function differently accordingly to the parental origin of the carrier (Kermicle 1970).

In 1974, genetic studies on mice demonstrated that the maternal-inheritance of a deletion at the Tme (T-associated maternal effect) locus located on chromosome 17 resulted in death at birth or in utero. In contrast, the paternal-inheritance of the deletion resulted in less abnormal and viable embryos (Johnson 1974). This suggested a differential contribution between the maternal and paternal allele.

Around that time, the researchers studying familial cases with Beckwith-Wiedemann Syndrome (BWS; see 1.3 - Beckwith-Wiedemann Syndrome) noted that BWS affected male and female offspring were born from female carriers only. These observations supported the idea that a parent-of-origin effect was associated with
the familial BWS phenotype (Lubinsky et al. 1974).

In the 1980's, pronuclear transplantation in newly fertilised mouse eggs aiming to remove and replace the paternal pronucleus with a second maternal one (i.e. generating a parthenogenic conceptus) and vice versa (i.e. generating an androgenic conceptus), resulted in the failure of embryos to develop normally. The parthenogenic mouse embryos developed tissues predominantly of embryonic origin but failed to develop the extra-embryonic lineages. In contrast, the androgenic mouse embryos developed predominantly extra-embryonic lineages but lacked, or had very underdeveloped, embryonic components. Consequently, it was suggested that mammalian genomes possess some genes that may be imprinted in different ways on the two parental genomes, making the presence of both maternal and paternal genome essential for normal embryonic development (McGrath and Solter 1983; Surani and Barton 1983; McGrath and Solter 1984; Surani et al. 1984).

In 1985, mouse genome engineering aiming to create regions of uniparental disomy by the introduction of reciprocal translocations resulted in mice with anomalous phenotypes which depart from normal in opposite directions. The discrete regions involved in the anomalous phenotypes were suggested to be subject to parent-of-origin effects (Cattanach and Kirk 1985). Interestingly, the study of non-Mendelian human disorder showed very similar inheritance to phenotypes seen in the disomic mice. As an example, paternally-inherited deletion of a region encompassing chromosome 15q11q13 results in Prader-Willi syndrome whereas the maternal-
inheritance of the deletion results in a very different clinical disorder, Angelman syndrome. This suggested that the paternal and maternal contribution of a gene or genes in region 15q11q13 were required for normal human growth and development (Nicholls et al. 1989).

In the early 1990’s, several research groups identified three endogenous imprinted genes \((Igf2r, Igf2, \text{and } H19)\) in mice. The \(Igf2r\) gene encodes for the insulin-like growth factor 2 receptor and is located within the \(Tme\) locus previously described to be subject to a parental-origin effect (see above). \(Igf2r\) was found to be maternally expressed and paternally silenced (Barlow et al. 1991). The \(Igf2\) gene encodes for the insulin growth factor 2 and was identified as a paternally expressed imprinted gene and maternally silenced (DeChiara et al. 1991; Ferguson-Smith et al. 1991). Finally, the \(H19\) gene (imprinted maternally expressed transcript) encodes for a long non-coding RNA that may have a tumour suppressor activity (Hao et al. 1993; Yoshimizu et al. 2008). \(H19\) was identified as maternally expressed and paternally silenced (Bartolomei et al. 1991). Additionally, \(H19\) is located 90 kb downstream of the \(Igf2\) imprinted gene, hence supporting the idea that imprinted genes can be clustered together. Finally, it was shown that the imprinting status of a gene may also be tissue-specific. The biallelic expression of \(Igf2\) has been described in the choroid plexus and leptomeninges of the brain whilst, as mentioned above, it is exclusively expressed from the paternally-inherited chromosome in the embryo (DeChiara et al. 1991). Similarly, the imprinted gene \(Dlk1\) (Delta-like homologue 1), normally expressed from the paternally-inherited chromosome, has been described to be
biallelically expressed in the postnatal neurogenic niche (Ferrón et al. 2011).

Finally, not long after the identification of the first endogenous imprinted genes it was shown that both parental genomes were differentially marked by DNA methylation at imprinting regions (Bartolomei et al. 1993; Ferguson-Smith et al. 1993; Stöger et al. 1993). In these reports, the imprinted differentially methylated regions (DMRs), mainly known as imprinting centres (ICs) or imprinting control regions (ICRs), were shown to be cis-acting elements controlling the expression of nearby or distant cluster of imprinted genes (Figure 1.7). Imprinted DMRs can be classified in germline imprinted DMRs and somatic (or secondary) imprinted DMRs. Germline imprinted DMRs exhibit differences in methylation states between the sperm and the egg. These differences are maintained post-fertilisation. In contrast, at somatic imprinted DMRs the DNA methylation is acquired after fertilisation but is still parent-of-origin specific (Woodfine et al. 2011). The parental-specific DNA methylation pattern at ICs is essential to maintain genomic imprinting and the disruption of the normal methylation pattern leads to aberrant imprinted gene expression (E. Li et al. 1993) and to diseases (Robertson 2005).
Figure 1.7. Genomic imprinting.

Gene X, Y and Z are imprinted genes and their gene expression is regulated by a cis-acting imprinting centre (blue oval). Methylation at the maternal imprinting centre leads to the expression of Gene Y and silencing of Gene X and Gene Z. On the paternal chromosome, the imprinting centre is unmethylated which leads to expression of Gene X and Gene Z but silencing of Gene Y.

1.2.2 Life cycle of genomic imprinting

The life cycle of genomic imprinting in mammals consists of three major steps: erasure, establishment and maintenance (Figure 1.8).
Figure 1.8. Life-cycle of genomic imprinting.

In the primordial germ cells, global methylation, including at imprinted regions, is erased via active and passive demethylation. Once devoid of methylation, the DNA methyltransferases DNMT3A, DNMT3B, further enhanced by DNTM3L, ensure re-methylation of the genome, including imprinted regions, via de novo methylation activity. After fertilisation, ZFP57, DPPA3 and DNMT1 protect the imprinting methylation marks against the epigenetic reprogramming events that take place during the pre-implantation (genome-wide demethylation) and post-implantation (genome-wide de novo methylation) period. (From Sanchez-Delgado et al. 2016).

1.2.2.1 Imprinting erasure

Erasure of imprinting methylation occurs in primordial germ cells (PGCs) at an early stage during differentiation of the female and male germlines. This global demethylation ensures the somatic cell epigenetic profile is completely erased. This epigenetic reprogramming is also required for the pluripotency of future germ cells,
imprinting switching and for the prevention of the inheritance of epigenetic defects (Morgan et al. 2005). The methylation erasure is achieved in a two-step process involving both passive (see section 1.1.1.2.3.1 - Passive DNA demethylation) and active (see section 1.1.1.2.3.2 - Active DNA demethylation) demethylation. Initially, PGCs proliferate actively, ensuring replication-dependent global passive demethylation (Kagiwada et al. 2013). This mechanism is further enhanced by the transcriptional silencing of the DNA methyltransferases $Dnmt1$, $Dnmt3a$, $Dnmt3b$ and $Dnmt3L$ and the key cofactor $Uhrf1$ (P.A. Jones and Taylor 1980; Kurimoto et al. 2008; Seisenberger et al. 2012; Kagiwada et al. 2013). A second wave of global active demethylation takes place during the migration of the PGCs to the genital ridge which are the precursors of the gonads (Hajkova et al. 2002; J. Lee et al. 2002). This active second epigenetic event involves the deamination of 5-mC to thymine by AID/APOBEC1, followed by the thymine excision by TDG or MBD4 and finally the repair of the resulting abasic site by the BER pathway (Morgan et al. 2004; Hajkova et al. 2010; Maiti and Drohat 2011). Alternatively, the TET enzymes can catalyse the oxidation of 5-mC to 5-hmC/5-fC/5-caC (He et al. 2011; Inoue et al. 2011; Ito et al. 2011; Hackett et al. 2013). The resulting oxidised products can be lost and replaced by unmodified cytosine through DNA replication (Inoue et al. 2011; Hashimoto et al. 2012) or via excision by TDG, MBD4 or an un-identified glycosylase followed by repair by the BER pathway (He et al. 2011; Inoue et al. 2011; Maiti and Drohat 2011; Shen et al. 2013).
1.2.2.2 **Imprinting establishment**

Once devoid of methylation, the PGCs undergo global genome re-methylation to acquire the gametes new epigenetic states. This includes the establishment of germline sex-specific methylation at imprinted loci. The timing of acquisition of imprinting methylation marks is different between the male and female germline. In the male germline, the DNA methylation imprint is established prenatally in prospermatogonia (Davis et al. 1999; Ueda et al. 2000; J.-Y. Li et al. 2004). In the female germline, the DNA methylation imprint is established asynchronously at different loci but in all cases completed postnatally during the oocyte growth (by the metaphase II) (Lucifero et al. 2002; Lucifero et al. 2004). The *de novo* DNA methyltransferase DNMT3A, helped by the enzymatically inactive DNA methyltransferase DNMT3L to stimulate its activity and stabilise its binding to the unmodified lysine 4 of histone 3 (H3K4me0), gives rise to the methylation profile of the gametes including the sex-specific germline imprints (Bourc'his et al. 2001; Kaneda et al. 2004; Suetake et al. 2004; Ooi et al. 2007). Additional mechanisms also take part to this *de novo* DNA methylation establishment. In the oocytes, the acquisition of methylation at the maternal imprinted DMR Gnas (GNAS (guanine nucleotide binding protein, alpha stimulating) complex locus) requires the active transcription of the *Nesp* transcript, which initiates furthest upstream in this imprinted domain (Chotalia et al. 2009). This relationship between transcription and establishment of DNA methylation in the oocytes is not exclusive to the imprinted *Gnas* locus. Intragenic methylated CpG islands are enriched at active transcription units compared to intragenic unmethylated CpG islands. Additionally, methylated
promoter CpG islands are more frequently found at overlapping transcript regions compared to unmethylated promoter CpG islands (Smallwood et al. 2011). Besides an active transcription, the presence of the histone H3K4 demethylase KDM1B (lysine demethylase 1B) is also required during oogenesis for the acquisition of DNA methylation at a subset of imprinted loci (Ciccone et al. 2009). The establishment of DNA methylation failed at the maternal imprinted DMRs Mest (mesoderm specific transcript), Grb10 (growth factor receptor bound protein 10), Plagl1 (pleiomorphic adenoma gene-like 1) and Impact (impact, RWD domain protein) in KDM1B-deficient oocyte whilst the acquisition of DNA methylation at the maternal imprinted DMRs Kcnq1ot1 (KCNQ1 overlapping transcript 1), Snrpn (small nuclear ribonucleoprotein N) and Igf2r are not affected by the absence of KDM1B. In the sperm, the PIWI-interacting RNA machinery is required for the establishment of DNA methylation at the paternal imprinted DMR Rasgrf1 (RAS protein-specific guanine nucleotide-releasing factor 1) but not at other imprinted paternal loci (Watanabe et al. 2011).

1.2.2.3 Imprinting maintenance

The paternal and maternal epigenetic imprints established in the germline are transmitted to the zygote through fertilisation and maintained faithfully throughout development and adulthood. Furthermore, the maintenance mechanism also protects the imprint against global embryonic epigenetic reprogramming events that take place before and after implantation. During pre-implantation, global epigenetic remodelling takes place and consists of erasing the epigenetic information presents on parental genomes. Global methylation is lost via the fast-active demethylation of
the paternal genome (Oswald et al. 2000) and the subsequent slow passive demethylation of both paternal and maternal genomes (Rougier et al. 1998; Santos et al. 2002). However, the newly established methylation marks at imprinted regions is strikingly fully resistant to the reprogramming. Several proteins are involved in this protection and maintenance mechanism. This includes the maternal and zygotic DNA methyltransferase DNMT1 (Howell et al. 2001; Hirasawa et al. 2008), the developmental pluripotency-associated protein 3 DPPA3 (also known as STELLA or PGC7) (T. Nakamura et al. 2007) and the Krüppel-associated box (KRAB) domain and 7 zinc fingers protein ZFP57 (X. Li et al. 2008). After implantation, the differential methylation at genomic imprinting regions is maintained by both DNMT1 and UHRF1 (Bostick et al. 2007; Sharif et al. 2007). In parallel, the unmethylated imprinted alleles escape another global epigenetic change that takes place at that time. This global epigenetic event consists of a wave of global de novo DNA methylation resulting in the hypermethylation of many genes including the totipotency, pluripotency and germ-cell specific genes (Messerschmidt et al. 2014). The insulator protein CTCF, the pluripotency transcription factor POU5F1 (POU class 5 homeobox 1; also known as OCT4) and the permissive histone modification H3K4me2/3 were shown to be involved in this protection pathway (N. Engel et al. 2006; Demars et al. 2010).

1.2.3 Consequences of aberrant genomic imprinting

It is widely believed that imprinted genes that are paternally expressed are associated with promoting the foetal growth, which would increase the chance of the
offspring to survive and reproduce. In contrast, it seems that imprinted gene that are
maternally expressed are more likely to be associated with restricting the foetal
growth by limiting access to the mother’s nutrients and resources, which would
increase her chance to survive and to have future offspring (Moore and Haig 1991).
To support this theory, the knockout in mice of paternally expressed genes, such as
\textit{Igf2}, mesoderm-specific transcript (\textit{Mest}) and paternally expressed gene 3 (\textit{Peg3}),
leads to foetal growth restriction (DeChiara et al. 1990; Lefebvre et al. 1998; L. LI et
al. 1999). In contrast, the knockout in mice of maternally expressed genes, such as
\textit{Igf2r}, \textit{H19} and \textit{Grb10}, is associated with foetal overgrowth (Lau et al. 1994; Leigthon
et al. 1995; Charalambous et al. 2003). Consistently, several evidences from the
mouse and rare human imprinting disorders suggest that imprinting predominantly
occur in genes influencing foetal growth, brain function and neurological behavioural

As suggested above, genomic imprinting failures can have dramatic consequences
on the normal growth and development of the embryo. Several epigenetic and/or
genomic alterations in imprinted gene clusters and in ICs have been associated with
a number of human diseases including cancer and imprinting disorders (IDs). The
over expression of the growth promoting gene \textit{IGF2} and the low-level expression of
the candidate tumour suppressor \textit{H19} have been described in the aetiology of
various tumour types including colon, liver, lung, ovarian cancer and Wilms tumour.
Loss-of-imprinting is also associated with IDs which are a group of ten rare but
probably underdiagnosed congenital diseases mainly characterised by growth,
metabolic and neurological abnormalities. These include Angelman syndrome (AS; OMIM n°105830), Prader-Willi syndrome (PWS; OMIM n°176270), transient neonatal diabetes mellitus type 1 (TNDM1; OMIM n°601410), pseudohypoparathyroidism type Ib (PHP1B; OMIM n°603233), Silver-Russell syndrome (SRS; OMIM n° 180860), Beckwith-Wiedemann syndrome (BWS; OMIM n°130650), Temple syndrome (TS; OMIM n°616222), Kagami-Ogata syndrome (KOS; OMIM n°608149), maternal uniparental disomy of chromosome 20 syndrome and precocious puberty syndrome (European Network of Human Congenital Imprinting Disorders, http://www.imprinting-disorders.eu). In the molecular aetiology of IDs, the deregulation of the genomic imprinting machinery, and consequently the aberrant expression of imprinted genes, is achieved by four different mechanisms: paternal or maternal uniparental disomy (UPD), genomic imbalances (duplications/deletions), epimutations (disturbed methylation) or point mutations in an imprinted gene (Figure 1.9).
1

2

3 **Figure 1.9. Molecular mechanisms leading to imprinting disorders.**

4 The four different mechanisms leading to the deregulation of imprinted gene expression and ultimately to imprinting disorders (adapted from Eggermann et al. 2015).

5

6

7

8 **1.3 Beckwith-Wiedemann Syndrome**

9 Beckwith-Wiedemann syndrome is a paediatric overgrowth condition firstly described in 1963-64 by Dr. J. Bruce Beckwith (Beckwith 1963) and independently by Dr. H.E. Wiedemann (Wiedemann 1964). BWS occurs once in approximately
every 15,000 births (Eggermann, Perez de Nanclares, et al. 2015), and is associated with an increased risk of various tumour types during childhood (DeBaun et al. 1998).

The clinical phenotype of BWS is highly variable and while some children are relatively mildly affected, others have a wider range of physical problems. The syndrome is most commonly associated with pre- and/or postnatal overgrowth, macroglossia and abdominal wall defects. Additional features include visceromegaly (particularly kidneys, liver and pancreas), neonatal hypoglycaemia, hemihyperplasia, genitourinary abnormalities and placental mesenchymal dysplasia. Embryonal tumours such as Wilms tumour (the most frequent), hepatoblastoma, neuroblastoma, adrenal carcinoma and rhabdomyosarcoma, are observed in approximately 7.5% of BWS patients (J.R. Engel 2000; Bliek et al. 2001; Weksberg et al. 2010). The vast majority of patients, with adequate treatment following diagnosis, survive infancy and develop normally.

1.3.1 Molecular defects in BWS aetiology

Analyses of rare chromosome 11p15 rearrangements have identified three regions, BWSCR1, BWSCR2 and BWSCR3, believed to play a role in BWS (Figure 1.10). BWSCR1 is the most frequent and most documented breakpoint and maps to a region of several imprinted genes including IGF2, H19, CDKN1C (cyclin dependent kinase inhibitor 1C) and KCNQ1OT1 (KCNQ1 opposite strand/antisense transcript 1) (Hoovers et al. 1995). Genetic and/or epigenetic abnormalities within BWSCR1
can disrupt the imprinted gene expression and lead to BWS. The less frequent breakpoints, BWSCR2 and BWSCR3, are located 5 and 7 Mb centromeric to BWSCR1 respectively (Redeker et al. 1995). BWSCR2 is defined by two breakpoints and may be associated with BWS (Alders et al. 2000). Two zinc finger genes, ZNF214 and ZNF215, have been identified within BWSCR2. The ZNF215 gene is imprinted in a tissue-specific manner and is expressed preferentially from the maternal allele, whereas ZNF214 is not imprinted (Alders et al. 2000; Sofos et al. 2011). Both BWSCR2 breakpoints disrupt two of the five alternatively spliced ZNF215 transcripts and parts of the 3-prime end of these splice forms are transcribed from the antisense strands of ZNF214. These data supported a role for ZNF215, and possibly for ZNF214, in the aetiology of BWS.

Through a number of different mechanisms, the primary epigenetic alterations or the genetic alterations resulting in the disruption of the imprinting of several genes located on chromosome 11p15.5 are described in the molecular aetiology of BWS (Figure 1.11).
Figure 1.10. Localisation of the three breakpoints associated to BWS at chromosome 11p15.

The more telomeric breakpoint, BWSCR1, includes IGF2, H19, CDKN1C and KCNQ1OT1 imprinted genes. The more centromeric breakpoints, BWSCR2 and BWSCR3, are respectively located 5 and 7 Mb from BWSCR1. BWSCR2 includes the tissue-specific imprinted gene ZNF215 and not imprinted gene ZNF214. (Adapted from Maher and Reik 2000).
Figure 1.11. Linear map of the BWS locus on 11p15.5.
(a) Linear schematic representation of the normal parent of origin-specific imprinted allelic expression. (b) H19/IGF2:IG-DMR gain-of-methylation, (c) KCNQ1OT1:TSS-DMR loss-of-methylation, (d) CDKN1C mutations and (e) H19/IGF2:IG-DMR and KCNQ1OT1:TSS-DMR paternal UPD. Mat, maternal; pat, paternal; OT1, refers to KCNQ1 antisense transcript, KCNQ1OT1. (Adapted from Choufani et al. 2010).
1.3.1.1 Disturbances at H19/IGF2:IG-DMR region

The telomeric imprinting centre H19/IGF2:IG-DMR (also known as IC1, ICR1 or H19-DMR) controls the IGF2 and H19 imprinted gene expression. H19/IGF2:IG-DMR is located within 2kb upstream of H19. Both IGF2 and H19 imprinted genes interact with the same enhancer located downstream of H19 in an enhancer-competition manner. The IGF2 gene encodes an embryonic foetal growth factor essential for normal development. IGF2 is normally paternally expressed and maternally silenced. H19 is a maternally expressed gene that encodes for a non-coding RNA that may function as a tumour suppressor (Hao et al. 1993) and is also implicated in growth restriction (Guo et al. 2008).

H19/IGF2:IG-DMR controls the allele-specific imprinted gene expression of IGF2 and H19 in two distinct ways. IGF2 gene expression is regulated by the methylation-sensitive insulator activity of the protein CTCF that can bind to several CTCF binding sites within the H19-IGF2 imprinted domain. That includes at the H19/IGF2:IG-DMR imprinted centre, at a site upstream to the IGF2 gene, at a centrally conserved region between IGF2 and H19 and at a site downstream of the enhancer (Figure 1.11-a). Once bind to the DNA, CTCF recruits the protein complex cohesin and each of the resulting CTCF/cohesin complexes interact with each other, hence forming chromatin loops and generate aboth active and inactive chromatin domain (Figure 1.11-a and Figure 1.12). The H19 gene expression is regulated by a secondary DMRs located within the H19 promoter. The methylation of the secondary DMR is acquired somatically and is mediated by H19/IGF2:IG-DMR (Hark et al. 2000; Bell
and Felsenfeld 2000; Srivastava et al. 2000; Murrel et al. 2004; Nativio et al. 2011) (Figure 1.11-a and Figure 1.12).

In normal conditions, $H19/IGF2$-IG-DMR is unmethylated on the maternal allele and methylated on the paternal allele. The absence of methylation on the maternal $H19/IGF2$:IG-DMR allows the insulator protein CTCF to bind to its target within $H19/IGF2$:IG-DMR, which in turn recruits the cohesin protein complex. The newly formed CTCF/cohesin complex at $H19/IGF2$:IG-DMR can then interact with the other CTCF/cohesin sites located upstream to the $IGF2$ gene and downstream to the enhancer. This leads to the formation of chromatin loops that isolate $IGF2$ and its promoter to an inactive chromatin domain away from the enhancer. As a result, the $IGF2$ gene promoter can’t interact with the enhancer and $IGF2$ is silenced. $H19$ and the enhancer are located within the same active chromatin loop and they are free to interact. This interaction and the absence of methylation at the $H19$ secondary DMR lead to $H19$ gene expression (Figure 1.11-a and Figure 1.12). In contrast, the presence of DNA methylation on the paternal $H19/IGF2$:IG-DMR prevents the insulator protein CTCF to bind to CTCF binding sites within $H19/IGF2$:IG-DMR. This results in the exclusion of $H19/IGF2$:IG-DMR from the CTCF/cohesin interacting region. The exclusion of $H19/IGF2$:IG-DMR results in $IGF2$ and its promoter to locate in an active chromatin domain near to the enhancer. This allows the interaction between $IGF2$ promoter with its enhancer and consequently leads to $IGF2$ gene expression. The methylation at $H19$ promoter inhibits $H19$ gene expression and this gene silencing is reinforced by the loss of interaction between $H19$ promoter and its
enhancer (Hark et al. 2000; Bell and Felsenfeld 2000; Srivastava et al. 2000; Murrel et al. 2004; Nativio et al. 2011) (Figure 1.11-a and Figure 1.12).

**Figure 1.12. Proposed model for the monoallelic expression of IGF2 and H19.** (Adapted from Nativio et al. 2011).

In BWS, the biallelic expression and the upregulation of IGF2 gene play a critical role in the molecular aetiology of BWS (Maher and Reik 2000) and sporadic tumours (Tycko 2000; Schofield et al. 2001). The primary isolated epigenetic alteration or in association with an underlying genomic alteration resulting in the gain-of-methylation
(GOM) of $H19/IGF2$:IG-DMR (i.e. both paternal and maternal $H19/IGF2$:IG-DMR are methylated) is responsible for the disruption of $IGF2$ and $H19$ imprinted gene expression (Figure 1.11-b). The main genomic alterations associated with the GOM of $H19/IGF2$:IG-DMR are variable length maternally inherited microdeletions removing a subset of the CTCF binding sites within the IC (Sparago et al. 2004; Prawitt, Enklaar, Gärtner-Rupprecht, et al. 2005; De Crescenzo et al. 2011; Beygo, Citro, et al. 2013). In addition, $OCT4/SOX2$ binding sites located within $H19/IGF2$:IG-DMR have been found to be important to maintain the unmethylated profile of the maternal allele (Poole et al. 2012). Accordingly, maternally-inherited point mutations disrupting conserved $OCT4/SOX2$-binding motif located in $H19/IGF2$:IG-DMR have been described in familial cases with BWS.

1.3.1.2 Disturbances at $KCNQ1OT1$:TSS-DMR region

The centromeric imprinting centre $KCNQ1OT1$:TSS-DMR (also known as IC2, ICR2 or KvDMR1) controls the $CDKN1C$ and $KCNQ1OT1$ gene expression. The $CDKN1C$ (cyclin-dependent kinase inhibitor 1C; also known as p57Kip2) gene encodes for a negative regulator of cell proliferation (M.H. Lee et al. 1995; Tsugu et al. 2000). In human, $CDKN1C$ is imprinted and primarily expressed from the maternal allele, though some expression (5 to 30%) is observed from the paternal chromosome (Chung 1996; Hatada et al. 1996). $KCNQ1OT1$ is a paternally expressed non-coding RNA with antisense transcription to $KCNQ1$ (potassium voltage-gated channel subfamily Q member 1). The 5' end of $KCNQ1OT1$ transcript overlaps with the differentially methylated imprinting centre $KCNQ1OT1$:TSS-DMR (M.P. Lee et al.
The expression of the imprinted genes located at the \textit{KCNQ1OT1}:TSS-DMR locus are controlled by the \textit{KCNQ1OT1} non-coding RNA negative regulator of gene expression activity and the CTCF enhancer blocking activity. On the paternal allele, \textit{KCNQ1OT1}:TSS-DMR is unmethylated and results in the expression of \textit{KCNQ1OT1}. Consequently, a nuclear compartment is established in the vicinity of the \textit{KCNQ1OT1} locus and a \textit{KCNQ1OT1} non-coding RNA-mediated bidirectional (i.e. downstream and upstream) gene silencing occurs. Several imprinted genes and enhancers are affected by this negative regulation, including \textit{CDKN1C} and its respective enhancer (Pandey et al. 2008). This repressor activity is further facilitated by the binding of the insulator protein CTCF to the unmethylated paternal \textit{KCNQ1OT1}:TSS-DMR. Similarly to the \textit{H19}/\textit{IGF2}:IG-DMR, the binding of CTCF to the unmethylated IC represses the interactions between \textit{CDKN1C} and its enhancer located downstream of \textit{KCNQ1OT1}:TSS-DMR. This results in the \textit{CDKN1C} gene silencing. In contrast, the maternal \textit{KCNQ1OT1}:TSS-DMR is methylated and results in the silencing of \textit{KCNQ1OT1}. Subsequently, the absence of the \textit{KCNQ1OT1} transcript leads to \textit{CDKN1C} gene expression. \textit{CDKN1C} gene expression is further enhanced by the lack of the enhancer blocking activity associated with the absence of CTCF binding (Algar et al. 2011) (Figure 1.11-a).

\textit{KCNQ1OT1} biallelic expression and \textit{CDKN1C} biallelic silencing play critical role in BWS. The primary loss-of-methylation (LOM) of \textit{KCNQ1OT1}:TSS-DMR (i.e. both
paternal and maternal \textit{KCNQ1OT1}:TSS-DMR are unmethylated) results in the disruption of \textit{KCNQ1OT1} and \textit{CDKN1C} imprinting. Consequently, this epimutation leading to the biallelic expression of \textit{KCNQ1OT1} and the biallelic silencing of \textit{CDKN1C} results in BWS phenotype (Figure 1.11-c). Chromosome deletion/duplication affecting the \textit{KCNQ1OT1}:TSS-DMR is rarely associated with BWS. However, the rare maternal inherited deletions occurring just in \textit{KCNQ1OT1}:TSS-DMR (Niemitz et al. 2004; Algar et al. 2011) or encompassing a larger locus including \textit{CDKN1C} (Zollino et al. 2010) have been reported. Finally, loss-of-function mutations of the maternally-inherited \textit{CDKN1C} allele are also documented in the aetiology of BWS (Hatada et al. 1997; Lee et al. 1997; Milani et al. 2014) (Figure 1.11-d).

1.3.1.3 Disturbances at both \textit{H19/IGF2}:IG-DMR and \textit{KCNQ1OT1}:TSS-DMR regions

Segmental paternal uniparental disomy (pUPD) affecting chromosome 11p15.5 is also described in BWS aetiology. The reported BWS cases with pUPD show somatic mosaicism with over expression of \textit{IGF2} due to the GOM at \textit{H19/IGF2}:IG-DMR and biallelic silencing of \textit{CDKN1C} due to the LOM at \textit{KCNQ1OT1}:TSS-DMR (Henry et al. 1993; Slatter et al. 1994; Cooper et al. 2007) (Figure 1.11-d).

1.3.1.4 Abnormality detection rate in BWS

The LOM at \textit{KCNQ1OT1}:TSS-DMR contributes to approximatively 50\% of BWS cases. The paternal UPD of chromosome 11p15 is described in approximatively 20\%
of cases. The GOM at *H19/IGF2*:IG-DMR contributes for approximatively 5 to 10% of BWS cases. Genetic mutations altering the maternal *CDKN1C* function accounts for approximatively 10% of cases (40 to 50% of familial BWS cases). Chromosomal aberrations (i.e. duplications, translocations, inversions) affecting chromosome 11p15 are described in a minority of BWS cases (<2-4%). Finally, in 13-15% of BWS cases the molecular defects are yet to be elucidated (Weksberg et al. 2010; Eggermann, Perez de Nanclares, et al. 2015) (Figure 1.13).

![Figure 1.13. Enrichment of the type of abnormalities describe in the molecular aetiology of BWS.](chart)

1.4 Silver-Russell Syndrome

Silver-Russell syndrome is a paediatric growth retardation condition first described in 1953 by Dr. Silver (Silver et al. 1953) and independently by Dr. Russell in 1954 (Russell 1954). SRS occurs once in approximatively every 75,000 to 100,000 births.
(Eggermann, Netchine, et al. 2015). Most cases are sporadic although a few familial
cases have been reported (Duncan et al. 1990).

The clinical presentation of SRS is variable but the disease is most commonly
associated with pre- and/or post-natal growth failure, relative macrocephaly at birth,
a frontal bossing in early life, body asymmetry, problems feeding and/or a low body
mass index at 2 years. Additional features include clinodactyly of the fifth finger,
syndactyly of the second and third toes and shoulder dimples. Finally, congenital
malformations are also associated with the disease. These include genital
abnormalities (cryptorchidism, hypospadias, Müllerian agenesis), renal and cardiac
defects (Netchine et al. 2007; Eggermann, Netchine, et al. 2015).

1.4.1 Molecular defects in SRS aetiology

SRS and BWS are very similar diseases in the sense that both syndromes are the
clinical mirror to the other. Unlike BWS the exact genes causing SRS are currently
unknown. However, several lines of evidence indicate that (epi)genetic alterations
leading to the deregulation of the imprinted genes \textit{IGF2}, \textit{CDKN1C}, \textit{MEST}
(mesoderm expressed transcript; also known as \textit{PEG1}) located on chromosome
7q32 and \textit{GRB10} (growth factor receptor bound protein 10) located on chromosome
7p11.2-p13 may be responsible for the disorder. Finally, a few SRS cases were also
reported with abnormalities at chromosomes other than chromosome 7 and 11,
suggesting that other genes might be involved in SRS molecular aetiology. In some
of these patients, additional and atypical features are also described (Abu-Amero et al. 2008).

1.4.1.1 Disturbances at chromosome 11p15.5

As discussed above (see section 1.3.1.1 - Disturbances at H19/IGF2:IG-DMR region), the expression of IGF2 is controlled by the imprinting centre H19/IGF2:IG-DMR. IGF2 is preferentially expressed from the paternally-inherited allele and is maternally silenced. In mirror to BWS, the biallelic silencing of IGF2 is described in SRS aetiology. The LOM at the paternal H19/IGF2:IG-DMR resulting in a concomitant decrease in IGF2 expression was identified in SRS individuals (Gicquel et al. 2005). Other mechanisms leading to the IGF2 biallelic silencing have also been described. These very rare genetic defects include the mosaic maternal UPD of chromosome 11 (Bullman et al. 2008), the maternal duplication of chromosome 11p15 encompassing both H19/IGF2:IG-DMR and KCNQ1OT1:TSS-DMR domains (Eggermann et al. 2005; Eggermann et al. 2010) or KCNQ1OT1:TSS-DMR domain only (Schönherr et al. 2007), the paternal deletions of the H19/IGF2 enhancer region (Grønskov et al. 2011) and the paternally-inherited IGF2 loss-of-function mutation (Begemann et al. 2015).

Another imprinted gene located at chromosome 11p15.5 has been described in SRS aetiology. Gain-of-function mutations in the imprinted gene CDKN1C are normally associated with the IMAGe (Intrauterine Growth Restriction, Metaphyseal Dysplasia, Adrenal Hypoplasia Congenita, and Genital Anomalies) syndrome (OMIM n°614732).
which is a severe growth retardation condition overlapping with SRS. Interestingly, a rare maternally-inherited gain-of-function mutation in the \textit{CDKN1C} gene has been identified in familial cases with SRS. It was subsequently suggested that the mutation was responsible for the SRS phenotype (Brioude et al. 2013).

1.4.1.2 Disturbances at chromosome 7

The identification of several SRS individuals with maternal UPD of chromosome 7 (Kotzot et al. 1995; Eggermann et al. 1997; Preece et al. 1997) raised the hypothesis that maternally transcribed imprinted genes with predicted growth suppression activity or paternally transcribed imprinting gene with predicted growth promoting activity were located on chromosome 7 and consequently were likely to be responsible for the observed SRS phenotype. Key candidate genes including the MEST and \textit{GRB10} imprinted gene have since been proposed as causative genes.

The \textit{MEST} gene encodes for a member of the alpha/beta hydrolase superfamily with unknown function. \textit{MEST} is an imprinted gene preferentially expressed from the paternally-inherited allele. \textit{Mest} knockout mice show pre- and postnatal growth restriction when the mutant gene is transmitted from the father (Lefebvre et al. 1998). Due to the high phenotypic similarity between \textit{Mest} knockout mice and maternal UPD 7 SRS cases, the \textit{MEST} imprinted gene was suggested to be responsible for the disease. In one SRS patient, the identification of a 35 Mb segmental maternal UPD at chromosome 7q31-qter, including \textit{MEST}, further supported the hypothesis that \textit{MEST} was involved in SRS aetiology (Hannula et al. 2001). However, despite...
this evidence, no conclusive reports indicating a clear pathogenic role of *MEST* has been found and, consequently, a *MEST* relevant role in SRS aetiology is questionable. Genetic mutations (e.g. paternal loss-of-function) leading to the deregulation of *MEST* expression have yet to be reported in SRS individuals (Riesewijk et al. 1998). Furthermore, the methylation analysis at the IC controlling *MEST* imprinting, *MEST*:alt-TSS-DMR, didn’t reveal epimutations that could deregulate *MEST* expression in SRS (S. Kobayashi et al. 2001; Schöherr et al. 2008).

The *GRB10* gene encodes for a growth factor receptor-binding protein. The SH2 (Src Homology 2) domain of human GRB10 was shown to inhibit the insulin receptor and insulin-like growth-factor receptor mediated mitogenesis activity. Therefore, *GRB10* has growth suppressing activity (F. Liu and Roth 1995; O'Neill et al. 1996). The *GRB10* gene is imprinted in an isoform and tissue specific manner. In human foetal brain, the parental specific gene expression is exclusively derived from the paternally-inherited allele. In contrast, the imprinted *GRB10* isoform γ1 is maternally expressed in skeletal muscle. However, biallelic expression of all *GRB10* isoforms was found in several other foetal tissues including intestine, kidney and liver (Blagitko et al. 2000; Hitchins et al. 2001). Several pieces of evidence suggested that *GRB10* is a strong candidate in SRS molecular aetiology. Mice with maternal disomy encompassing the mouse orthologous *Grb10* (located on chromosome 11) showed prenatal growth failure whilst the reciprocal paternal disomy resulted in the opposite phenotype, consisting of prenatal overgrowth (Cattanach and Kirk 1985).
The disruption of Grb10 imprinting via maternally inherited Grb10 mutation results in a disproportionate mice overgrowth phenotype with an Igf2-independent mechanism (Charalambous et al. 2003). Finally, the genetic investigation of SRS individuals revealed maternally inherited duplication at chromosome 7p11.2-p13, encompassing the GRB10 region, hence suggesting that GRB10 over-expression might be involved in SRS aetiology (Joyce et al. 1999; Monk et al. 2000; Monk et al. 2002). However, despite this evidence, conclusive reports involving GRB10 have yet to be reported in SRS. The methylation analysis of the IC regulating GRB10 expression, GRB10:alt-TSS-DMR, showed no significant changes that would deregulate GRB10 imprinted gene expression (Arnaud 2003; Monk et al. 2003). Furthermore, genetic alterations leading to GRB10 over-expression (e.g. maternal duplication) or biallelic expression (e.g. paternal gain-of-function) have not been reported in SRS individuals (Yoshihashi et al. 2000; Hitchins et al. 2001).

1.4.1.3 Disturbances at other chromosome

Copy number variations and UPD affecting chromosomes other than chromosome 7 and 11 have been identified in cohorts of clinically diagnosed SRS patients with, in some individuals, additional and atypical clinical features. The majority of these rare defects were identified in single cases and the associated prevalence is yet to be determined (Abu-Amero et al. 2008).

More recently, the hypomethylation of ICs located at chromosome 14, MEG3/DLK1:IG-DMR and MEG3:TSS-DMR, has been identified in three SRS
compatible individuals with no known defects at chromosome 7 or 11 (Azzi et al. 2015; Kagami et al. 2015). These ICs control the imprinted gene expression of \textit{DLK1} (delta like non-canonical Notch ligand 1) and \textit{MEG3} (Maternally expressed gene 3).

In mice, the low-level expression of \textit{Dlk1} has been implicated in growth retardation and accelerated adiposity (Moon et al. 2002; Cleaton et al. 2016). The imprinted gene \textit{MEG3} is suggested to be a tumour suppressor and the loss of \textit{MEG3} expression has been found in various types of human tumours and tumour cell lines (Zhou et al. 2012). These new findings at chromosome 14 imply that the maternal UPD of chromosome 14 and the epigenetic abnormalities affecting \textit{MEG3/DLK1}:IG-DMR and \textit{MEG3}:TSS-DMR constitute a rare, but potentially important, underlying factors leading to SRS. Furthermore, it was interesting to note that the LOM of both ICs is mostly associated with Temple syndrome (TS), which is an imprinting disorder sharing a high degree of phenotypic overlap with SRS. Consequently, this raises the question if TS individuals that fit the SRS description should be classed as SRS.

1.4.1.4 Abnormality detection rate in SRS

In the general population, the LOM at H19/\textit{IGF2}:IG-DMR contributes for more than 38\% of SRS cases and the maternal UPD of chromosome 7 is described in approximatively 10\% of cases. The other mechanisms leading to SRS are very rare. The copy number variations (deletion/duplication) affecting H19/\textit{IGF2}:IG-DMR and \textit{KCNQ1OT1}:TSS-DMR imprinted loci, the maternal UPD of chromosome 11 and the genomic abnormalities affecting chromosomes other than chromosome 7 and 11 are responsible for less than 1-2\% of SRS cases. An \textit{IGF2} loss-of-function mutation has
been identified in one case only. Similarly, a CDKN1C gain-of-function mutation has been reported in a single family with SRS. Finally, the molecular defects are yet to be identified for the remaining SRS cases (approximately 40-50%) (Abu-Amero et al. 2008; Eggermann, Perez de Nanclares, et al. 2015) (Figure 1.14).

Figure 1.14. Enrichment of the type of abnormalities describe in the molecular aetiology of SRS.

1.5 Imprinting disorders and multilocus methylation disturbances.

Genomic imprinting disorders are individually rare and at a clinical level they are usually considered distinct disorders. However, clinical features such as disordered
growth are common to several disorders. In some cases, a similar clinical phenotype may result from different types of disordered imprinting. Thus, SRS has been associated with epimutations at $H19/IGF2$:IG-DMR (SRS_IC1) on chromosome 11p15.5 but also with maternal uniparental disomy of chromosome 7 (SRS_UPD7) (see section 1.4 - Silver-Russell Syndrome). Occasionally a patient may present with a clinical diagnosis of a specific ID but molecular investigation reveals an epigenetic abnormality that is associated with a clinically distinct disorder. Thus, several studies have reported rare patients referred with a clinical diagnosis of BWS who have LOM of the paternal $H19/IGF2$:IG-DMR allele at 11p15.5, which is characteristically associated with SRS (Tee et al. 2013). By opposition, rare patients were referred with a clinical diagnosis of SRS but have LOM of the maternal $KCNQ1OT1$:TSS-DMR allele, which is characteristically associated with BWS (Azzi et al. 2009). Such epigenotype-phenotype inconsistencies may result from the observation that a subset of patients with an ID have epimutations at multiple ICs (Rossignol et al. 2006; Bliek et al. 2008; Mackay et al. 2008; Azzi et al. 2009; Lim et al. 2009; Turner et al. 2010; Court et al. 2013; Poole et al. 2013; Tee et al. 2013; Maeda et al. 2014; Mackay et al. 2015). This phenomenon is also known as multilocus imprinting disturbances (MLIDs).

The first evidence of MLID was reported in individuals clinically diagnosed with TNDM1 and carrying not only a LOM at the $PLAGL1$:alt-TSS-DMR, epimutation accounting for approximatively 20% of TNDM1 cases, but also hypomethylation at $KCNQ1OT1$:TSS-DMR, which is the main epimutation associated with BWS (Arima...
Since then, MLID has been identified in approximately 50% of TNDM1 individuals with LOM at *PLAGL1*:alt-TSS-DMR (Mackay et al. 2008), in approximately 25% of BWS with LOM at *KCNQ1OT1*:TSS-DMR (Rossignol et al. 2006; Bliék et al. 2008; Azzi et al. 2009; Court et al. 2013; Poole et al. 2013; Tee et al. 2013; Maeda et al. 2014), in approximately 10% of SRS individuals with LOM at *H19/IGF2*:IG-DMR (Azzi et al. 2009; Turner et al. 2010) and in approximately 12.5% of PHP1B individuals carrying isolated epimutation at the *GNAS* imprinting region (Perez-Nanclares et al. 2012; Court et al. 2013; Maupetit-Méhouas et al. 2013). Though the MLID phenomenon has now been documented in several IDs, the causes and clinical significance of MLID are, in most cases, not well defined (Eggermann, Perez de Nanclares, et al. 2015; Mackay et al. 2015).

1.6 General summary of this project

DNA methylation is a heritable epigenetic mark that is, for historical reasons, mainly associated with transcriptional repression. However, it is now well established that DNA methylation also play critical roles in other mechanisms such as the maintenance of the genome integrity, the inactivation of the X chromosome in female or in the regulation of genomic imprinting.

Genomic imprinting is an epigenetic phenomenon that results in the deliberate silencing of one copy (allele) of specific genes, according to whether the allele comes
from the father or the mother. Those genes, known as imprinted genes, are currently believed to be mainly associated with foetal growth, brain development, and neurological behavioural traits.

At imprinted domains, both parental genomes are differentially marked by DNA methylation (i.e. methylated on the paternal allele but unmethylated on the maternal allele). These imprinted differentially methylated regions, known as imprinting centres or imprinted DMRs, are cis-acting elements controlling the expression of nearby or distant cluster of imprinted genes. The parent-specific DNA methylation marks found at imprinted centres can be established during the development of parental germ cells into egg or sperm. The methylation marks at these germline imprinted centres are maintained post-fertilisation but are erased and then replaced by a new one during the new generation of germ cells. Alternatively, secondary imprinted DMRs are imprinted loci that acquire their parent-specific methylation marks post-zygotically.

Genomic imprinting is essential for normal growth and development and its deregulation, for example through aberrant methylation at imprinting centres, is associated in with number of human diseases such as the imprinting disorders BWS and SRS. BWS is a paediatric overgrowth condition caused by epigenetic or genetic alterations at the imprinted gene cluster located on chromosome 11p15.5. In contrast, SRS is a paediatric growth restriction condition caused by epigenetic or genetic alterations at imprinted gene clusters located on chromosome 7 and 11.
Interestingly, in a subset of BWS patients associated with a LOM at \textit{KCNQ1OT1}:TSS-DMR or SRS patients associated with a LOM at \textit{H19/IGF2}:IG-DMR, multiple epimutations (MLIDs) occurring at imprinted loci other than \textit{KCNQ1OT1}:TSS-DMR (for BWS) and \textit{H19/IGF2}:IG-DMR (for SRS) are also described. These aberrant MLID epigenotypes are not exclusively found in BWS and SRS individuals. MLIDs are also described in patients diagnosed with other imprinting disorders, including TNDM1 and PHP1B. However, in most cases, the causes, nature, frequency and clinical significance of MLID is currently not well defined yet.

1.7 Aims of this project

The principal aim of my thesis was to perform the methylation profiling of imprinted loci in BWS and SRS individuals. The resulting comprehensive analysis would have helped to define the nature, frequency and range of MLID in these cohorts. It also would have provided valuable informations on the possible clinical significance of these widespread imprinting failures in BWS and SRS. Additionally, I also aimed to decipher the possible underlying causes that could be responsible for the abnormal MLID epigenotype.

To that aim, I first sought to develop an affordable and highly sensitive high-throughput method to assess the methylation at imprinted loci in both cohorts. Following the unsuccessful attempt to develop the assay, I carried out the comprehensive methylation profiling using methylation array and my own
bioinformatic methodology. I studied the significance of the use of assisted reproductive technologies in BWS_IC2 individuals with MLID and, using whole-exome sequencing, I investigated the possibility that genetic alterations may be responsible for the widespread imprinting methylation disturbances. Finally, by using functional assays, I sought to know if a novel candidate gene variant found in siblings with BWS and LOM at KCNQ1OT1:TSS-DMR could be responsible for MLID.

Taken together, the work undertaken during my thesis could improve not only our knowledge about the molecular aetiology of BWS and SRS, but also our understanding of genomic imprinting and its disturbances.
Chapter 2

General Methods
2.1 Buffers recipe

2.1.1 1X Transfer buffer

192 mM glycine, 25 mM Tris, 10% pure methanol, adjust volume with dH₂O if needed.

2.1.2 1X PBS-Tween 20 (PBST)

137 mM NaCl, 12 mM Phosphate, 2.7 mM KCl, 0.1% Tween 20, pH 7.4, adjust volume with dH₂O if needed.

2.1.3 Blocking buffer

5 g non-fat milk, 100 ml PBST

2.1.4 4X Laemmli sample buffer

240 mM Tris/HCl pH 6.8, 8% SDS, 40% glycerol, 0.04% bromophenol blue, 5% beta-mercaptoethanol, adjust volume with dH₂O if needed.

2.1.5 RIPA

50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% Igepal, 0.5% sodium deoxycholate, 0.1% SDS, 1X protease inhibitors, adjust volume with dH₂O if needed.
2.1.6 2X SSC

300 mM NaCl, 30 mM sodium citrate, adjust volume with dH₂O if needed.

2.1.7 1X TAE

40 mM Tris, 20 mM acetic acid, 1 mM EDTA, adjust volume with dH₂O if needed.

2.2 DNA purification

DNA was purified using the QIAquick PCR Purification Kit (QIAGEN). 5 volumes of Buffer PB was added to 1 volume of sample and mixed by pipetting. To bind DNA to the membrane, the sample was loaded into a QIAquick spin column placed in a 2 ml collection tube and centrifuged a 10,000 x g for 30-60 seconds. The flow-through was discarded and the collection tube was reused. To wash the DNA, 750 µl of Buffer PE was added to the QIAquick spin column and the column was centrifuged at 10,000 x g for 30-60 seconds. The flow-through was discarded and the collection tube was reused. The column was centrifuged again for 1 minute to remove residual ethanol from Buffer PE. To elute the DNA, the column was transferred in a 1.5 ml tube, 50 µl of Buffer EB (or water) was added to the centre of the Qiaquick membrane, let to incubate for 1 minute and the column was centrifuged at 10,000 x g for 1 minute. The purified DNA was used immediately or stored at -20 °C for later use.

Alternatively, DNA purification was done by ethanol precipitation. In a 15 ml tube, 2.5
volumes of 100% ethanol and 0.1 volume of 3M NaOAc (pH 5.2 or 5.5) were added to 1 volume of sample. The mixture was mixed well and left at -80 °C for at least 30 minutes. To precipitate the DNA, the tube was centrifuged at 3,000 rpm (or maximum speed) for 30 minutes at 4 °C. After centrifugation, the supernatant was carefully discarded and 800 µl of cold 75% ethanol was added to the tube. Using a 1 ml tip (with tip pre-cut) the DNA pellet was transferred to a 1.5 ml tube and the tube was centrifuged at 14,000 rpm for 5 minutes. The supernatant was carefully discarded and the pellet was let to air dry for 3-5 minutes. To resuspend the DNA, 10-150 µl of double distilled water (pre-incubated at 37 °C) was added to the pellet and mixed well by pipetting up and down several times. The purified DNA was used immediately or stored at -20 °C for later use.

2.3 DNA quantification

2 µl of DNA was loaded on NanoDrop® ND-1000 (Thermo Scientific) and the measured absorbance at 260 nm was used to calculate the DNA concentration. The presence of proteins in DNA samples was evaluated with the measurement of absorbance at 280 nm. A ratio of sample absorbance at 260 nm and 280 nm (A260/A280) greater than 1.8 suggested DNA samples exempt from proteins. Similarly, salt and organic compounds contaminations were evaluated with the measurement of absorbance at 230 nm. A ratio of sample absorbance at 260 nm and 230 nm (A260/A230) greater than 1.5, ideally close to 1.8, suggested minimal or non existant contaminants carryover.
DNA concentration was determined by the measurement of absorbance at 260 nm and 280 nm. 2 µl of DNA was loaded on NanoDrop® ND-1000 (Thermo Scientific) and DNA concentration was obtained with the ratio of sample absorbance at 260 nm and 280 nm (A260/A280).

Alternatively, DNA concentration was determined by fluorescent dyes method using Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific). For samples with expected DNA concentration below 100 ng/µl, 2 µl of sample DNA was mixed with 198 µl of Qubit® dsDNA HS Reagent diluted in 1:200 in Qubit® dsDNA HS Buffer. Samples fluorescence intensities were measured with Qubit® 2.0 Fluorometer and DNA concentration was calculated from the fluorescence intensities of provided low and high standard (10 µl of low or high Qubit® standard was mixed with 190 µl of Qubit® dsDNA HS Reagent diluted in 1:200 in Qubit® dsDNA HS Buffer). For samples above 100 ng/µl, DNA was first diluted in distilled water to bring concentration below 100 ng/µl and then processed as above.

### 2.4 Sodium bisulfite treatment

Genomic DNA (gDNA) was quantified with NanoDrop® ND-1000 (see section 2.3 - DNA quantification) and 500 ng of DNA was bisulfite treated using EZ DNA Methylation-Lightning™ Kit (Zymo Research). Bisulfite treatment was done as follows: first gDNA was mixed with 130 µl of Lightning Conversion Reagent and then incubated in a thermal cycle with the following programme: 98 °C for 8 minutes, 54
°C for 60 minutes, 4 °C for up to 20 hours. Sample was then mixed with 600 µl of M-
Binding Buffer, loaded into a Zymo-Spin™ IC Column placed into a provided
Collection Tube and centrifuged at full speed for 30 seconds. To remove conversion
reagent 100 µl of M-Wash Buffer was added to the column and centrifuge at full
speed for 30 seconds. Complete methylated cytosine to uracil conversion was
achieved by adding 200 µl of L-Desulphonation Buffer to the column and let stand at
room temperature (20-30 °C) for 15-20 minutes. To stop the desulphonation reaction
the tubes were centrifuged at full speed for 30 seconds and then washed twice in
200 µl of M-Wash Buffer with centrifugation at full speed for 30 seconds after each
wash steps. Elution step was done by adding 15 µl of M-Elution Buffer to the column
and centrifuged at full speed for 30 seconds. Recovered bisulfite treated DNA
(bisDNA) was used immediately or stored at -20 °C for later use.

2.5 Infinium HumanMethylation450 BeadChip

Genomic DNA was quantified with Qubit® dsDNA HS Assay Kits (see section 2.3 -
DNA quantification) and 500 ng of gDNA was processed on the Infinium
HumanMethylation 450K BeadChips (HM450K) (Illumina). This robust platform can
interrogate the methylation status of more than 485,000 CpG across the genome.
Processing of the samples on the HM450K array were performed at the Cambridge
Genomic Services (Cambridge, UK), according to the manufacturer’s protocol
(Infinium HD Assay Methylation Protocol Guide). In brief, gDNA was sodium bisulfite
converted, denatured, neutralised and whole-genome amplified. Amplified bisulfite
treated gDNA was then enzymatically fragmented, purified by isopropanol precipitation and the resulting fragmented purified DNA was hybridized to BeadChip. The chip was then washed to remove all unhybridized and nonspecifically hybridized DNA. Using the capture DNA as template, specific fluorophore labels were added at CpG sites query by single-base extension. Staining detection was done by recording high resolution images of the light emitted from the fluorophore of the single-base extension products after laser excitation. GenomeStudio software (Illumina) was used to extract signal intensities for each probe and generate raw ‘idat’ files. The resulting raw data files were imported and analysed in R.

2.6 Analysis software

Unless stated otherwise, data handling, processing and statistical analysis were performed in R v3.3.0 and in Microsoft Excel v15.24.

2.7 DNA fragments analysis of DNA libraries

Bioanalyzer High Sensitivity DNA Kit (Agilent) was used to analyse DNA fragments.

2.7.1 Setting up the chip priming station

A new syringe was placed in the chip priming station, the plate base was inserted in position C and the syringe clip was positioned at the lowest position.
2.7.2 Loading the gel dye mix

Protected from light, the gel-dye mix was put at room temperature for 30 minutes before use. After equilibration at room temperature, 9 µl of gel-dye mix was pipetted at the bottom of the well located on the fourth column and third row of the DNA chip. The DNA chip was then placed into the chip priming station. The chip priming station was closed and the plunger of the syringe that was previously positioned at 1 ml was pressed down until held by the clip. The plunger was kept down for exactly 60 seconds and then released. The plunger was inspected that it moved back to the 0.3 ml mark and after 5 seconds it was slowly pull back to the 1 ml position. Finally, the chip priming station was opened and 9 µl of gel-dye mix was loaded in the three wells located on the fourth column and first, second and fourth row.

2.7.3 Loading the marker

5 µl of High Sensitivity DNA marker was pipetted into the wells with a ladder symbol and into each of the 11 sample wells.

2.7.4 Loading the ladder and the samples

1 µl of High sensitivity DNA ladder was pipetted in the well marked with the ladder symbol. 1 µl of sample was pipetted in each of the 11 sample wells. The chip was placed horizontally in the adapter of the IKA vortex mixer and vortex for 60 seconds at 2400 rpm. Finally, the DNA chip was placed into the Agilent 2100 Bioanalyzer (with chip selector was set to position (1)).
2.7.5 **Starting the chip run**

Using the 2100 Expert software, the appropriate assay was selected from the Assay menu located in the instrument context tab. Then the sample information’s were inserted into the sample name table. Finally, the button start was pressed to start the run.

2.7.6 **Data analysis**

Results were analysed directly using the 2100 Expert software built-in analysis module. The detected peaks were labelled in base-pair to indicate the products size.
Chapter 3

Development of a novel next generation sequencing strategy for the genomic imprinting methylation profiling of individual with BWS and SRS
3.1 Introduction

3.1.1 What is NGS?

Next-generation sequencing (NGS), also known as high-throughput sequencing or deep sequencing, is a generic term used to describe a number of different modern sequencing technologies. Examples include (i) sequencing-by-synthesis technology used by Illumina in its commercially available sequencers MiSeq, HiSeq, NextSeq and NovaSeq (see section 3.1.2 - Sequencing-by-synthesis; an Illumina technology); (ii) pyrosequencing used by Roche 454 Life Sciences in the now discontinued Genome Sequencer FLX and GS Junior sequencers; (iii) sequencing-by-ligation used by ThermoFisher Scientific in the ABI SOLID sequencers; and (iv) detection of pH changes that follow the release of a hydrogen ion used by ThermoFisher Scientific in the Ion Proton, Ion PGM and Ion S5 sequencers.

NGS platforms are reliable methods to sequence DNA and have many significant advantages when compared to classical Sanger sequencing. (i) It is quicker. NGS allows the sequencing of 300 Gb of DNA in a single run whilst a single read of 1 kb can be sequenced in a single reaction with Sanger sequencing. (ii) It is cheaper. Nowadays the sequencing of whole human genome with NGS may cost £3,000 or less whilst the predicted cost using Sanger sequencing could be approximatively £6M. (iii) It is more accurate and reliable. In a single NGS reaction, each portion of DNA can be covered by many short overlapping DNA fragments that are sequenced multiple times, hence resulting in reliable base calling. In contrast many time
consuming Sanger sequencing reactions will be required to achieve the same level of coverage. (iv) It needs less DNA. Approximately 5 µg of genomic DNA would be required for the sequencing of the whole human genome using NGS. In contrast, Sanger sequencing of a small genomic region of 1 kb will required approximately 20 ng of genomic DNA.

3.1.2 Sequencing-by-synthesis; an Illumina technology

The sequencing-by-synthesis technology used by Illumina relies on several steps. Using the appropriate method, sequencing libraries are prepared from genomic DNA (Figure 3.1- step 1). In the 5' to 3' order, the libraries consist of a P5 adaptor, possibly index 1 (i5) barcode, sequencing primer for read 1, DNA insert, sequencing primer for read 2, possibly index 2 (i7) barcode and P7 adaptor sequences. Once prepared and ready for sequencing, the double-stranded libraries are denatured and loaded onto the flow cell. The DNA molecules attach through the attachment of either the P5 or P7 adapter of the libraries to the surface of the flow cell (Figure 3.1- step 2). In parallel, the other end that did not attach to the flow cell anneals to a dense lawn of complimentary oligonucleotides that coat the surface of the flow cell, hence forming a kind of bridge (Figure 3.1- step 3). The single-stranded DNA molecules that are now in a bridge shape are double-stranded through the addition of unlabelled nucleotides, primers, buffer and polymerase enzyme (Figure 3.1- step 4). The double-stranded molecules are denatured and the original strands are washed away, leaving behind the strands that had been synthesised and covalently bonded
to the flow cell surface in a mixture of orientations (Figure 3.1- step 5). Step 3 to 5 are repeated multiple times to create clusters of DNA molecules that contains thousands of identical copies of the same sequence (in a mixture of orientations) (Figure 3.1- step 6). Prior to sequencing cycles, the P5 region is then cleaved, resulting in clusters containing only fragments which are attached by the P7 adaptor region. This is to ensure that all copies are in the same direction. Sequencing primers, polymerase, fluorescently labelled nucleotides with reversible terminator (one fluorescent colour for each base) and buffer are then added to the flow cell. The primers anneal to the P5 end of the fragments, which leads to the incorporation of the first base and therefore begins the sequencing-by-synthesis procedure (Figure 3.1- step 7). After the incorporation of the first locked base, the flow cell is washed, hence resulting in removing unincorporated nucleotides and excess of primers. The sequencer starts the acquisition of the clear fluorescent signals and records the identity of the first base that was added to each clusters (Figure 3.1- step 8). The terminator dye and the fluorescent dye of the first base are removed, ultimately resulting in an unblocked and non-fluorescent nucleotide. Subsequently, a new sequencing chemistry cycle is initiated to determine the second base. Sequencing primers, polymerase, fluorescently labelled nucleotides with reversible terminator and buffer are added to the flow cell, which result in the incorporation of another fluorescently labelled nucleotide (Figure 3.1- step 9). The flow cell is then washed and the sequencer acquires the fluorescent signals and records the identity of the second base that was incorporated (Figure 3.1- step 10). The terminator dye and the fluorescent dye are removed which allows the start of a new sequencing cycle. This
sequencing procedure is repeated multiple times which leads to the identification of each base of a cluster (Figure 3.1- step 11). Finally, all the acquired sequencing reads are preprocessed and aligned for further analysis (Figure 3.1- step 12).
Figure 3.1. Sequencing-by-synthesis technology used in Illumina.

3.1.3 **What is bisulfite sequencing?**

DNA methylation plays a critical role in normal growth and development and aberrant DNA methylation changes can cause a number of human disorders, including cancer and imprinting disorders. Consequently, it is crucial to characterise methylome states during normal development and under pathological conditions.

Bisulfite-sequencing (Frommer et al. 1992) is one of the most reliable methods to characterise DNA methylation within a genomic region. It relies on the treatment of a single stranded DNA with a chemical compound, sodium bisulfite, and results in the deamination of unmodified cytosine residues to uracil. In contrast, 5-mC (and the iterative oxidised form 5-hmC/5-fC/5-caC), is protected against the bisulfite treatment. Consequently, cytosine residues that survive the bisulfite treatment are indirectly identified as methylated, whilst cytosines that are converted to thymines after polymerase chain reaction are identified as unmethylated (Figure 3.2). Used in combination with NGS, it constitutes nowadays the method of choice for the high-resolution methylation profiling of genomic regions.
Unmethylated cytosine residues are deaminated to uracil whereas methylated cytosine residues are unaffected. After polymerase chain reaction and sequencing, all cytosines that survive the treatment were methylated cytosine.

3.1.4 NGS based methods for DNA methylation profiling

The status of DNA methylation of a genomic region can be analysed by many different methods that rely on three basic principles: (i) the enzymatic digestion of methylated and unmethylated DNA with a restriction enzyme; (ii) the enrichment of methylated DNA using anti-methylcytosine antibodies or methyl-binding domain proteins; (iii) the sodium bisulfite conversion of genomic DNA which results in the conversion of unmodified cytosine to thymine whereas methylated cytosines remain unchanged. These principles have been integrated into high-throughput analytical applications such as next-generation sequencing (NGS) platforms.

Whole-genome bisulfite sequencing (WGBS) is the gold standard method for the single base-pair resolution complete mapping of cytosine methylation across the genome. It has made significant contributions to determine the methylome of individuals and especially in cancer cells (Hansen et al. 2011; Hon et al. 2012). However, despite undeniable advantages, it is an expensive approach especially when applied to a large set of samples.

RRBS is a large-scale approach that uses CpG-specific restriction enzyme (e.g. MspI, which is methylation-insensitive) digestion to randomly enrich for regions of high CpG content, such as CpG islands and promoters in any genome (Meissner et al. 2008; Z.D. Smith et al. 2009; Gu et al. 2011; Boyle et al. 2012; Junwen Wang et al. 2013). The method allows the profiling of approximatively 10-20% of the human
CpG content and requires a very low amount of DNA (10-100 ng). However, due to the intrinsic principle of using restriction enzymatic digestion, many CpG-sparse regions could be left uncharacterised. Additionally, the use of a high output sequencer is recommended in order to obtain a good depth of sequencing coverage.

MeDIP-seq is a large-scale approach that uses monoclonal antibodies directed against 5-mC to precipitate and enrich methylated DNA fragment (MeDIP; (Weber et al. 2005)) before sequencing. It was used to complete the first high resolution methylome of a mammalian (i.e. human) genome (Down et al. 2008). The method has now been refined and requires a low input of DNA (i.e. 50 ng) (Taiwo et al. 2012) for the characterisation of 60-90% of the human CpG content. Similarly to RRBS, sequencing following MeDIP will require a high output sequencer to obtain a good depth of sequencing coverage.

Targeted capture methods may be designed to interrogate any region of the genome. First developed to conduct whole-exome sequencing at a low cost compared to whole-genome sequencing (Albert et al. 2007; Hodges et al. 2007; Okou et al. 2007; Porreca et al. 2007), these methods were gradually integrated with bisulfite sequencing. Several targeted bisulfite sequencing (TBS-seq) methods have been developed and rely on different capture technics, including bisulphite padlock probes (Deng et al. 2009; Diep et al. 2012) (see section 3.1.5 - DNA methylation profiling using the bisulfite padlock probes below), microdroplet-based PCR amplification coupled with NGS protocols (Herrmann et al. 2011; Komori et al. 2011; Guilhamon
et al. 2013; Paul et al. 2014), on array hybridisation (Hodges et al. 2009) or in solution hybridisation (E.-J. Lee et al. 2011; Ivanov et al. 2013; Allum et al. 2015). TBS-seq offers several advantages over WGBS, RRBS or MeDIP-seq: (i) they are scalable approaches that can adapt to the amount of targeted regions; (ii) they are designed in answer to the hypothesis in question; (iii) depending on the size of the project they can result in high depth of sequencing coverage (>1000x) even when sequenced with a benchtop sequencer (e.g. MiSeq); (iv) they are developed with the scope of processing many samples hence offering a high degree of throughput for sample processing and multiplexing and (v) they don’t require, for the majority, a large amount of input DNA (200-500 ng). However, TBS-seq has the disadvantage of (i) limiting findings to the captured targeted regions which consequently excludes possible new discoveries that are located out of these boundaries; (ii) for some methods, the design of capturing probes may be complicated by reduced sequence complexity (resulting from bisulfite treatment) and high GC-rich regions; (iii) they have a starting high cost (e.g. for probe synthesis) but they become much more affordable when used for the analysis of large set of samples.

3.1.5 DNA methylation profiling using the bisulfite padlock probes

Bisulfite padlock probes (BSPPs) have been suggested as a practical, flexible and cost-effective approach to interrogate the methylation profile of multiple regions of interest across a large set of samples (Ball et al. 2009; Deng et al. 2009; Diep et al.
The method requires a very low amount of input DNA (e.g. 200 ng of bisulfite-treated DNA) and is highly specific and reproducible.

Historically, BSPP was firstly used to assess the methylation status of 66,000 CpG sites in 2,020 CGI on human chromosomes 12 and 20 (Deng et al. 2009). In the same year, another research group used BSPP to profile 7,000 CpG sites within the ENCODE pilot project regions (Ball et al. 2009). Both studies used approximately 30,000 padlock probes to capture the targeted regions and both showed that the method is highly specific and reproducible. More recently, a third group released an upgraded version of the protocol and demonstrated the high specificity of the technique by characterising the methylation profile of approximately 500,000 CpG sites across 34 Mb of sequences. For this large-scale project, 330,000 padlock probes were designed, synthesised and used to capture the sequence of interest (Diep et al. 2012). Although the initial cost of BSPP is high, the cost per sample dramatically decreases with increased sample sizes (Diep et al. 2012).

BSPP are long single stranded DNA oligonucleotides (100-150 mers) designed to hybridise to bisulfite converted genomic DNA targets in a horseshoe manner (Figure 3.3-a). The targeted regions for sequence capture are the gap between the two hybridised and locus-specific arms of a padlock probe and their capture are done by filling the gap with a polymerase (Figure 3.3-b). Uncaptured linear DNA is digested (Figure 3.3-c) and the resultant circular DNA is enriched and amplified using the common linker sequence between the two ligation arms with primers containing
sequencer's adaptors and barcode (sequencing library preparation) (Figure 3.3-d).

BSPP sequencing libraries are mainly designed to be sequenced on Illumina systems. However, the BSPP approach is flexible and can be adapted to another sequencing platform.
Figure 3.3. Bisulfite padlock probes.
(a) Padlock probes bind to bisulfite converted genomic DNA. (b) The sequence of interest is captured through gap filling step. (c) Linear DNA is digested with nuclease and (d) circles DNA, which contain sequence of interest, are amplified by PCR with primers containing P5 and P7 Illumina sequencing adaptors and index sequences. Both primers are specific to a different portion of the BSPP common linker. (Adapted from Diep et al. 2012).
3.1.6 **Aim**

As discussed previously (see section 1.5 - Imprinting disorders and multilocus methylation disturbances), a subset of BWS and SRS individuals have been identified with MLID but the nature, frequency and clinical significance of such aberrant epigenotype has not been clearly defined yet. Additionally, the comparison of different studies of MLID in BWS and SRS is complicated by different strategies for detecting epimutations and different choices of ICs analysed. In order to undertake a comprehensive survey of the frequency and nature of MLID in patients with BWS and SRS I attempted to develop a NGS assay for a highly sensitive methylation profiling at 46 imprinting DMRs (Table 3.1) (Court et al. 2014). The assay was intended to fulfil a list of criteria that includes (i) a streamlined workflow for the processing of a large number of clinical samples; (ii) being cost-efficient; (iii) the characterisation of CpG methylation within the targets at a single base-pair resolution; (iv) a very high depth of sequencing coverage (e.g. > 500x); and (v) requires a low amount of input DNA (e.g. <500 ng). At the time of the project, this was one of the first attempt of genome-wide screening for alteration at multiple imprinted loci using NGS method and applied to a large cohort of BWS and SRS individuals (more than 100). Additionally, although the primary focus was to characterise DNA methylation at imprinting regions in BWS and SRS individuals, the NGS assay could also be used for the genomic imprinting methylation profiling of other diseases, including other IDs and cancer.
### Table 3.1: List of imprinted DMRs.

(Court et al. 2014)

<table>
<thead>
<tr>
<th>Imprinted DMR</th>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Number CpGs</th>
<th>Methylation origin</th>
<th>Germline derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPIE: Ex1-DMR</td>
<td>1</td>
<td>4002426</td>
<td>4002540</td>
<td>39</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>DIRAS3: TSS-DMR</td>
<td>1</td>
<td>6851543</td>
<td>6851754</td>
<td>88</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>DIRAS3: Ex2-DMR</td>
<td>1</td>
<td>6851250</td>
<td>6851386</td>
<td>39</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>GPR1-AS: TSS-DMR</td>
<td>2</td>
<td>2070696</td>
<td>2070944</td>
<td>86</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>ZDBF2-GPR1: 1G-DMR</td>
<td>2</td>
<td>2071145</td>
<td>2071364</td>
<td>439</td>
<td>Paternal</td>
<td>Sperm gDMR-secondary DMR</td>
</tr>
<tr>
<td>NAP1L5: TSS-DMR</td>
<td>4</td>
<td>8961818</td>
<td>8961237</td>
<td>57</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>VTRNA2-1: DMR</td>
<td>5</td>
<td>1354140</td>
<td>1354166</td>
<td>76</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>FAM90B: TSS-DMR</td>
<td>6</td>
<td>3849082</td>
<td>3850359</td>
<td>90</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>PLAG1: alt-TSS-DMR</td>
<td>6</td>
<td>1443280</td>
<td>1443288</td>
<td>143</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>IGFB2: In2-DMR</td>
<td>6</td>
<td>1604285</td>
<td>1604756</td>
<td>74</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>WDR27: Int13-DMR</td>
<td>6</td>
<td>1700545</td>
<td>1700561</td>
<td>58</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>GRB10: TSS-DMR</td>
<td>7</td>
<td>5084872</td>
<td>5085132</td>
<td>171</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>PEG10: TSS-DMR</td>
<td>7</td>
<td>4206537</td>
<td>4206560</td>
<td>110</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>MEST: alt-TSS-DMR</td>
<td>7</td>
<td>1301301</td>
<td>1301348</td>
<td>226</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>HTR5A: TSS-DMR</td>
<td>7</td>
<td>1548627</td>
<td>1548633</td>
<td>55</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>ERLN2: Int6-DMR</td>
<td>8</td>
<td>3760492</td>
<td>3760608</td>
<td>37</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>PEG13: TSS-DMR</td>
<td>8</td>
<td>1211814</td>
<td>1211854</td>
<td>193</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>FANCC: Int1-DMR</td>
<td>9</td>
<td>9807540</td>
<td>9807674</td>
<td>26</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>INPP5F: In2-DMR</td>
<td>10</td>
<td>1215780</td>
<td>1215787</td>
<td>52</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>H19/IGF2: IG-DMR</td>
<td>11</td>
<td>2018812</td>
<td>2024740</td>
<td>250</td>
<td>Paternal</td>
<td>Sperm gDMR</td>
</tr>
<tr>
<td>IGF2: Ex8-DMR</td>
<td>11</td>
<td>2153991</td>
<td>2155112</td>
<td>63</td>
<td>Paternal</td>
<td>No-secondary DMR</td>
</tr>
<tr>
<td>IGF2: alt-TSS-DMR</td>
<td>11</td>
<td>2168333</td>
<td>2168768</td>
<td>33</td>
<td>Paternal</td>
<td>Sperm gDMR</td>
</tr>
<tr>
<td>KCNQ1OT1: TSS-DMR</td>
<td>11</td>
<td>2171948</td>
<td>2172269</td>
<td>192</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>RB1: Int2-DMR</td>
<td>13</td>
<td>4889234</td>
<td>4889763</td>
<td>195</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>MEG3/DLK1: IG-DMR</td>
<td>14</td>
<td>1012754</td>
<td>1012705</td>
<td>64</td>
<td>Paternal</td>
<td>Sperm gDMR</td>
</tr>
<tr>
<td>MEG3: TSS-DMR</td>
<td>14</td>
<td>1012905</td>
<td>1012978</td>
<td>188</td>
<td>Paternal</td>
<td>No-secondary DMR</td>
</tr>
<tr>
<td>MEG6: Int2-DMR</td>
<td>14</td>
<td>1013707</td>
<td>1013714</td>
<td>19</td>
<td>Maternal</td>
<td>No-secondary DMR</td>
</tr>
<tr>
<td>MIR163: TSS-DMR</td>
<td>15</td>
<td>2350706</td>
<td>2351240</td>
<td>109</td>
<td>Maternal</td>
<td>Oocyte gDMR/secondary DMR</td>
</tr>
<tr>
<td>MAGEL2: TSS-DMR</td>
<td>15</td>
<td>2389242</td>
<td>2389429</td>
<td>51</td>
<td>No secondary gDMR</td>
<td></td>
</tr>
<tr>
<td>NDN: TSS-DMR</td>
<td>15</td>
<td>2393341</td>
<td>2393279</td>
<td>108</td>
<td>No secondary gDMR</td>
<td></td>
</tr>
<tr>
<td>SNRPN: alt-TSS-DMR</td>
<td>15</td>
<td>2500656</td>
<td>2500640</td>
<td>19</td>
<td>No secondary DMR</td>
<td></td>
</tr>
<tr>
<td>SNURF: TSS-DMR</td>
<td>15</td>
<td>2520004</td>
<td>2520197</td>
<td>113</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>IGF1R: In2-DMR</td>
<td>15</td>
<td>9940496</td>
<td>9940560</td>
<td>55</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>ZNF597: 3' DMR</td>
<td>16</td>
<td>3481801</td>
<td>3482688</td>
<td>29</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>ZNF597: TSS-DMR</td>
<td>16</td>
<td>3492928</td>
<td>3494663</td>
<td>76</td>
<td>Paternal</td>
<td>No-secondary DMR</td>
</tr>
<tr>
<td>ZNF331: alt-TSS-DMR</td>
<td>19</td>
<td>5404050</td>
<td>5404212</td>
<td>125</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>ZNF331: alt-TSS-DMR</td>
<td>19</td>
<td>5405706</td>
<td>5405886</td>
<td>102</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>PEG3: TSS-DMR</td>
<td>19</td>
<td>5734849</td>
<td>5735271</td>
<td>221</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>MCTS2P: TSS-DMR</td>
<td>20</td>
<td>3013463</td>
<td>3015303</td>
<td>47</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>N NAT: TSS-DMR</td>
<td>20</td>
<td>3814880</td>
<td>3815258</td>
<td>135</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>L3MBTL1: alt-TSS-DMR</td>
<td>20</td>
<td>4214236</td>
<td>4214400</td>
<td>84</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>GNAS-NESP: TSS-DMR</td>
<td>20</td>
<td>5741403</td>
<td>5741625</td>
<td>257</td>
<td>Paternal</td>
<td>No-secondary DMR</td>
</tr>
<tr>
<td>GNAS-AST1: TSS-DMR</td>
<td>20</td>
<td>5742854</td>
<td>5742803</td>
<td>128</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>GNAS-XL: Ex1-DMR</td>
<td>20</td>
<td>5742895</td>
<td>5743148</td>
<td>200</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>GNAS-A8: TSS-DMR</td>
<td>20</td>
<td>5746325</td>
<td>5746520</td>
<td>198</td>
<td>No secondary gDMR</td>
<td></td>
</tr>
<tr>
<td>WRB: alt-TSS-DMR</td>
<td>21</td>
<td>4075751</td>
<td>4075826</td>
<td>43</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
<tr>
<td>SNU13: alt-TSS-DMR</td>
<td>22</td>
<td>4207774</td>
<td>4207873</td>
<td>63</td>
<td>Maternal</td>
<td>Oocyte gDMR</td>
</tr>
</tbody>
</table>

Genomic location map the Genome Reference Consortium Human Build 37 (GRCh37/hg19)
3.2 Materials and methods

3.2.1 DNA purification

See section 2.2 - DNA purification.

3.2.2 DNA quantification

See section 2.3 - DNA quantification.

3.2.3 Sodium bisulfite treatment

See section 2.4 - Sodium bisulfite treatment.

3.2.4 Amplicon size selection

Size selection of a pool of amplicons was done using E-Gel® SizeSelect™ 2% pre-cast agarose gels (Thermo Fisher Scientific). 20-25 µl of sample was loaded in the top wells (lane 1 to 8) of the gel. 5-10 µl of diluted molecular weight markers was loaded into the small middle well (lane M) at the top of the gel. 25 µl of deionised water was loaded into all of the wells (lane 1 to 8) in the lower row and 5-10 µl of deionised water was loaded into the small well of the lower row (lane M). Following sample loading, the gel was inserted into the E-Gel® iBase™ Power System and the amber filter was put over the iBase™ device. For the electrophoresis, the program called ‘Run SizeSelect 2%’ was selected and an appropriate time (according to ‘Run Time Estimation Table’) was set to ‘Run Time to Reference’ (Run Time Estimation Table). Once set, the electrophoresis was run until the band of interest reached the...
reference line. At that point the electrophoresis was paused, the collection wells were refilled to 25 µl of deionised water, the appropriate time listed under ‘Run Time from Reference Line to Collection Well’ from the ‘Run Time Estimation Table’ was entered and the electrophoresis was run. The migration of the band of interest was carefully monitored until it migrates into the Collection Well. At that point the run was stopped and the DNA of the appropriate size was collected from the Collection Well. The collected DNA was used immediately or stored at -20 °C for later use.

3.2.5 Preparation of sequencing libraries using bisulfite padlock probes

3.2.5.1 Bisulfite padlock probes design and synthesis

ppDesigner software (Diep et al. 2012) was used to design bisDNA specific padlock probes within regions of interest (Table 3.1). In design 1 (see section 3.3.1 – First experimental design), overlapping probes were designed to cover as much as possible of the regions of interest and CpG inclusion within the probes was allowed. The probes were synthesised by LCSiences. In design 2 (see section 3.3.2 – Second experimental design), one probe was design for each of PPIEL:Ex1-DMR, ZDBF2/GPR1:IG-DMR, FAM50B:TSS-DMR, H19/IGF2:IG-DMR, KCNQ1OT1:TSS-DMR, L3MBTL1:alt-TSS-DMR and SNU13:alt-TSS-DMR. CpG inclusion within the probes was not allowed and the probes were synthesised by Integrated DNA Technologies.
3.2.5.2 Bisulfite padlock probes production (for LCSciences only)

The following amplification, nicking digestion and purification steps of the padlock probes were done for the one synthesised by LCSiences only (Figure 3.4). The quantity of probes synthesised by LCSiences was very low and consequently it was necessary to amplify them prior target capture step. On the contrary, the quantity of probes synthesised by Integrated DNA Technologies was high and therefore it was not necessary to amplify them prior target capture.

3.2.5.2.1 Amplification (for LCSciences only)

To overcome the low concentration of the padlock probes received from LCSciences, amplification primer site 1 (AP1) and amplification primer site 2 (AP2) were added at each end of all probes. The amplification of the single stranded padlock probes was done as previously described (Diep et al. 2012) (Figure 3.4-a). In brief, in 200 μl reaction volume, 1 nM of template oligonucleotides was mixed with 400 nM each of forward primer AP1_F and reverse primer AP2_R (Table 3.2), and 100 μL of KAPA SYBR fast Bio-Rad qPCR Master Mix (Kapa Biosystems). The mixture was placed in a thermocycler for PCR amplification and incubated at 95 °C for 30 sec, 5 cycles of 95 °C for 5 sec; 52 °C for 1 min; and 72 °C for 30 sec, 10-12 cycles of 95 °C for 5 sec; 60 °C for 30 sec; and 72 °C for 30 sec, and 72 °C for 2 min. The resulting amplicons were purified using QIAquick PCR purification columns (QIAGEN) (see section 2.2 - DNA purification) and re-amplified by PCR in 32 reactions (100 μL each) with 0.02 nM first round amplicons, 400 nM each of AP1_F primer and AP2_R primer, and 50 μL of KAPA SYBR fast Bio-Rad qPCR Master Mix. For PCR amplification,
the mixture was incubated in a thermocycler at 95 °C for 30 sec, 13-15 cycles of 95 °C for 5 sec; 60 °C for 30 sec; and 72 °C for 30 sec, and 72 °C for 2 min. The resulting double stranded amplicons were purified by ethanol precipitation and re-purified with QIAquick PCR purification columns (see section 2.2 - DNA purification). The purified amplicons were quantified by Nanodrop (see section 2.3 - DNA quantification) and immediately processed with next step or stored at -20 °C for later use.

Table 3.2: List of primers used for the amplification of the padlock probes.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5'&gt;3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1_F</td>
<td>TGCCTAGGACCGGTACACT</td>
</tr>
<tr>
<td>AP2_R</td>
<td>GAGCTTCGGTACGCAATG</td>
</tr>
</tbody>
</table>

3.2.5.2.2 Nicking digestion (for LCSciences only)

Nicking digestion was used to remove the amplification primer sites AP_1 and AP_2 from the double stranded amplicons (Figure 3.4-b). Approximately 4 μg of the purified amplicons were digested with 100 units of Nt.Alwl (100 U/μl, New England Biolabs, cleaves at 5'-GGATCNNNN/N-3' only on one strand of a double stranded DNA) at 37°C for 1 h in NEBuffer 2 (New England Biolabs). The enzyme was heat inactivated at 80 °C for 20 min. The digested amplicons were then incubated with 100 units of Nb.BrsDI (10 U/μl, New England Biolabs, cleaves at 3'-CGTTAC/NN-5' only on one strand of a double stranded DNA) at 65 °C for 1 h. The resulting double stranded nicked DNA was purified by QIAquick PCR purification column (see section 2.2 - DNA purification). The nicking digestion of the amplicons with Nt.Alwl and Nb.BrsDI resulted in a double stranded DNA with a 70 bp long top strand (i.e.
equivalent to the padlock probes) and a 100 bp bottom strand (i.e. equivalent to the unwanted complementary strand).

3.2.5.2.3 Denaturing PAGE purification (for LCSciences only)

To discard the 100 bp complementary strands and to keep the 70 bp padlock probes, the DNA molecules were purified using denaturing Novex® TBE-Urea Gels 6% (Thermo Fisher Scientific) (Figure 3.4-c). In 100 µl final volume, 2 µg of nicked DNA was mixed with 2X Novex® TBE-Urea Sample Buffer in a 1:1 ratio volume. For the DNA ladder, in a 10 µl final volume 1 µg of 10 bp DNA ladder (Thermo Fisher Scientific) was mixed with 1X Novex® TBE-Urea Sample Buffer. The Novex® TBE-Urea gel was put into XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Fisher Scientific), the upper and lower chamber of the electrophoresis system was filled with 0.5X Novex® TBE Running Buffer (Thermo Fisher Scientific) and the gel was pre-run at 200 V for 10-15 minutes. At the same time, the DNA and DNA ladder-loading dye mixtures were incubated at 75 °C for 8-10 minutes and then quickly cooled on ice for at least 1 minute. The wells of the Novex® TBE-Urea gel were flushed with a 1 ml pipette to remove residual urea. The DNA and DNA ladder-loading dye mixtures were quickly loaded into the gel and the gel was run at 200 V for 30 minutes. Once the run finished, the gel was removed from the cassette and stained in 25-50 ml of 0.5X Novex® TBE Running Buffer and 1X SYBR Gold (Thermo Fisher Scientific) for 5 minutes and gentle shaking. The gel was then placed on a clean Saran wrap on a UV transilluminator (UVP). The area of the gel containing the band of approximately 70 bp was cut as close to the band using a clean scalpel (with
minimising exposing the gel to UV light). The cut gel was moved to a clean area and chop into small pieces with a scalpel. Two 0.5 ml tubes were pierced at their bottom with a 22G needle and then put into two 1.5 ml DNA LoBind tubes (Eppendorf). The chopped gel was transferred into the two 0.5 ml tubes and then centrifuged at 15,000 rpm for 3 minutes at room temperature. The gel remaining in the 0.5 ml tube was transferred to the 1.5 ml tube below with a sterile pipette tip. 400 µl of 1X TAE buffer (see section 2.1.7 - 1X TAE) was then added to the tube, and the mixture was shake vigorously on vortex for at least 45 minutes at 37 °C. Once finished, the clear supernatant was transferred to a Nanosep column (Pall Corporation) and centrifuged at 15,000 rpm for 3 minutes. The lower layer of the flow-through was carefully (i.e. avoiding the large chunk of gel) transferred again to the Nanosep column and then centrifuged for an additional 3 minutes at 15,000 rpm. The supernatant was transferred to a fresh 1.5 ml tube and precipitated with ethanol (see section 2.2 - DNA purification). After ethanol precipitation, the DNA pellet was dissolved with 10-15 µl of distilled water. All probes were pooled and the size of the probes and DNA concentration was estimated using 6% TBE-urea gel. The resulting bisulfite padlock probes were used immediately or stored at -20 °C for later use.
**Figure 3.4. Preliminary steps used for the probes syntthesised by LCSiences.**

(a) PCR amplification was used to increase bisulfite padlock probes quantity. (b) AP1 and AP2 sites were removed from double stranded amplicons using nicking digestion. The enzymatic digestion resulted in a DNA molecules with 70 bp top strand and 100 bp bottom strand. (c) The unwanted bottom strand (i.e. complementary product) was removed using denaturing urea PAGE.

3.2.5.3 Multiplex capture of bisulfite treated target regions

500 ng of sample gDNA was quantified with Nanodrop (see section 2.3 - DNA quantification) and bisulfite treated (see section 2.4 - Sodium bisulfite treatment). In 25 µl volume reaction, 50 nM of a pool of padlock probes, approximatively 500 ng of bisDNA and 1X Ampligase buffer (Epicentre) were mixed and incubated in a thermocycler using the following reaction conditions: 95°C for 10 min, gradual cooling at 70 °C (0.1 °C/sec), 70 °C for 30 min, gradual cooling at 65 °C (0.1 °C/sec), 65 °C for 60 min, gradual cooling at 60 °C (0.1 °C/sec), 60 °C for 120 min, gradual cooling at 55 °C (0.1 °C/sec) and 55 °C for 20 h. The capture reaction was immediately processed with next step.
3.2.5.4 Circularisation of captured sequences

2.5 µl gap-filling mix consisting of 100 µM dNTP, 10 U/µl Titanium Taq (Takara Bio USA), 0.5 U/µl Ampligase and 1X Ampligase buffer (Epicentre) was incubated at 95 °C for 1 min, cooled at 55 °C and added to the previous capture reaction. Gap-filling and circularisation reactions were then performed by incubating the reaction at 55 °C for 20 h. The gap-filled reaction was then cooled on ice and immediately processed with next step.

3.2.5.5 Enzymatic digestion

Linear DNA was digested by adding 4 µl of exonuclease mix (20 U exonuclease I and 200 U exonuclease III, 1X ampligase buffer; New England Biolabs) to the target capture reaction. The reaction was then incubated at 37 °C for 2 h followed by exonuclease inactivation at 95 °C for 2 min and hold at 4 °C indefinitely. Reaction was immediately processed with next step or stored at 4 °C or -20 °C for later use.

3.2.5.6 Capture circle amplification

Sequencing library was generated by amplifying the exonuclease treated circle DNA with barcoded primers. In 100 µl volume reaction, all of the circularised DNA was mixed with 400 nM of each forward AmpF_P5 and reverse AmpR_P7_Ind_001 primer (for design 1) or forward SLXA_PE_MIPBC_FOR and SLXA_PE_MIPBC_REV_001 primer (for design 2) (Table 3.3), 50 µl of KAPA HiFi 2x HotStart ReadyMix Taq (Kapa Biosystems) and PCR-grade water. The mixture was incubated in a thermocycler and PCR amplified using one of condition 1,
condition 2 or condition 3 programme (see below). The resulting barcoded sequencing library was processed immediately after with next steps consisting of fragment analysis using bioanalyzer (see section 2.7 - Fragment analysis of DNA libraries), size selection and purification using E-Gel® SizeSelect™ gels (see section Amplicon size selection), followed by DNA quantification using Qubit® dsDNA HS (see section 2.3 - DNA quantification). Finally, the libraries were again checked for correct size using bioanalyzer (see section 2.7 - Fragment analysis of DNA libraries). The libraries were used immediately after or stored at -20 °C for later use.

Table 3.3: List of primers used for sequencing libraries enrichment.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’&gt;3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOR FIRST DESIGN</td>
<td></td>
</tr>
<tr>
<td>AmpF_P5</td>
<td>AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAAGATGT GTATAAGAGACAG</td>
</tr>
<tr>
<td>AmpR_P7_ind_001</td>
<td>CAAGCAGAAGACGGCATACGAGATGCCTGATCTGGACTGAGTTCTTGGC ACCCGAGAATTCCA</td>
</tr>
<tr>
<td>FOR SECOND DESIGN</td>
<td></td>
</tr>
<tr>
<td>SLXA_PE_MIPBC_FOR</td>
<td>AATGATACGGCGACCACCGAGATCTACACTCTCAGATTTATCGAG GTCCGCAC</td>
</tr>
<tr>
<td>SLXA_PE_MIPBC_REV_001</td>
<td>CAAGCAGAAGACGGCATACGAGATGTTAAGACGCTAGGAACGATGAGCC TCAAC</td>
</tr>
</tbody>
</table>

Underlined sequence are index barcode

PCR programme:

Condition 1: 98 °C for 30 sec, 10 cycles of 98 °C for 15 sec; 50 °C for 20 sec; and 72 °C for 20 sec, 20 cycles of 98 °C for 15 sec; 72 °C for 20 sec, and 72 °C for 3 min, and hold at 4 °C.

Condition 2: 98 °C for 30 sec, 25 cycles of 98 °C for 15 sec; 58 °C for 20 sec, and 72 °C for 20 sec, and 72 °C for 3 min, and hold at 4 °C.
Condition 3: 98 °C for 30 sec, 25 cycles of 98 °C for 15 sec; 50 °C for 20 sec; and 72 °C for 20 sec, 72 °C for 3 min, and hold at 4 °C.

3.2.5.7 Sequencing

The sequencing of the libraries was performed by the Stratified Medicine Core Laboratory (SMCL) Next Generation Sequencing facility. A MiSeq sequencer was used for the sequencing. The sequencing primer used are listed in Table 3.4.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’&gt;3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIPBS_SEQ_NEXT (from Illumina Nextera)</td>
<td>TCGTCGCGAAGCTGAGGATTATCGTAAGAGACAG</td>
</tr>
<tr>
<td>MIPBS_SEQ_TRUSEQ (from Illumina TruSeq Small RNA)</td>
<td>GTGACTGGAGTTCCTTGGCACCAGAGATTTCA</td>
</tr>
<tr>
<td>MIPBS_SEQ_INDEX (from TruSeq Small RNA)</td>
<td>TGGAATTCTCGGGTGCCAGGAAACTCCAGTCAC</td>
</tr>
</tbody>
</table>

**Table 3.4: List of primers used for sequencing using MiSeq sequencer.**

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’&gt;3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIPBC_SEQ_FOR</td>
<td>TACACCATCTCAGATTTATCGAGGAGTCCGAC</td>
</tr>
<tr>
<td>MIPBC_SEQ_REV</td>
<td>GCTAGGAACGATGAGGCTCCAAC</td>
</tr>
<tr>
<td>MIPBC_SEQ_INDEX</td>
<td>GTTGGAGGATCATCGTTCTAGGC</td>
</tr>
</tbody>
</table>

3.3 Results

3.3.1 First experimental design

3.3.1.1 Design requirements

I designed BSPP with the following specific requirements: the capture insert size would be of 120-125 nucleotides; the maximum length of the two ligation arms would
be of 25 nucleotides; the maximum combined length of the two ligation arms would be of 40 nucleotides. The amplification primer site 1 (AP1) required for padlock probes amplification contains the nicking endonuclease Nt.AlwI sequence. Similarly, the amplification primer site 2 (AP2) contains the nicking endonuclease Nb.BrsDI sequence. The nucleotide composition of the common linker was dictated by the sequencing primer sequences. However, the choice of the sequencing primers to include in the common linker was complicated by several factors. The sequencing libraries were predicted to be sequenced using a MiSeq sequencer. One of the particularities of this sequencer is that it runs hotter (65 °C) than other Illumina sequencers, such as the HiSeq sequencer (55 °C), during the deblocking and extension stages. Consequently, custom sequencing primers have to have a melting temperature above 65 °C to prevent them from dissociating from the target. Additionally, the custom sequencing primers need to meet Illumina recommendations which are 33 bp long, 51.5% GC, with a Tm of 65.5 °C for sequencing primer for read 1 and 37 bp long, 59.5% GC with a Tm of 70.1 °C for sequencing primer for read 2. However, despite fulfilling these requirements, it is not guaranteed the custom sequencing primers will work properly. Finally, Illumina doesn’t provide technical support when custom sequencing primers are used. Consequently, I decided to incorporate a mixture of Illumina proprietary sequencing primers, Illumina Nextera sequencing primer for read 1 and Illumina TruSeq Small RNA sequencing primer for read 2, into the common linker. They are both natively recognised by the MiSeq sequencer and they have a limited impact on introducing additional Nt.AlwI and Nb.BrsDI restriction sites.
3.3.1.2 Padlock probes to synthetise

Following the design criteria, 516 overlapping padlock probes were generated *in silico* using ppDesigner to capture 61.6 kb of DNA sequence, including the targeted regions. Analysis of the probes sequence revealed CpG sites inclusion within the ligation arms of 412 (80.5%) padlock probes. Therefore, to avoid binding bias towards the methylated allele, a second copy of each of the 412 probes was computed to be specific to the unmethylated allele, which resulted in the design of a total of 928 padlock probes. The pool of 928 oligonucleotides were checked to be free from additional Nt.Alwl and Nb.BrsDI recognition site sequences to avoid unwanted digestion. Out of 928, the procedure identified 19 (~ 2%) probes with additional Nt.Alwl and Nb.BrsDI recognition site sequences, which consequently were removed from the pool of padlock probes. This resulted in 909 padlock probes to be synthesised (Table 3.5).

![Table 3.5: Number of padlock probes to synthesise.](image)

<table>
<thead>
<tr>
<th>Methylated + unmethylated padlocks</th>
<th>Padlocks containing extra restriction sites for Nt.Alwl and Nb.BrsDI</th>
<th>Final total amount of padlocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>516</td>
<td>928</td>
<td>909</td>
</tr>
</tbody>
</table>

3.3.1.3 Library preparation

Quality control of the sequencing libraries before size selection by Bioanalyser identified mainly two amplicons of different sizes. The amplicons with a length of approximatively 286 bp were predicted to be the expected sequencing libraries. The
amplicons with a length of approximately 237 bp were predicted to be a truncated form of the sequencing libraries resulting from inefficient amplification (i.e. only one primer anneals to the DNA template) (Figure 3.5).

The libraries were then size selected (retained product at approximately 280-290 bp) and cleaned-up. qPCR analysis indicated that the resulting libraries were of the correct size and were at an appropriate concentration for the sequencing reaction (data not show).

Figure 3.5. BSPP sequencing library preparation (first design).

Bioanalyzer analysis identified two kinds of amplicons in the sequencing libraries. One is predicted to be the full-length library construct whilst the other is predicted to be a truncated form of the expected constructs.
The resulting sequencing libraries were sequenced on Illumina MiSeq sequencer. However, no sequencing reads were generated during the sequencing run. The quality control performed during the sequencing run, which mainly checks the cluster generation output, did not indicate any failures with the instrument. After investigation with the core sequencing facility, I formulated several hypothesis that may explain the unsuccessful sequencing. Firstly, balance A/C/G/T bases composition is required for the optimal differentiation of the different clonally amplified DNA molecules spatially separated in the flow-cell. My libraries were generated from bisulfite converted DNA, which meant that the A/C/G/T bases proportion was un-balanced and consequently could have been an issue for the sequencer. However, to avoid this problem, PhiX sequencing control was spiked into my libraries in sufficient quantity, hence resulting in a pool of libraries with a balanced A/C/G/T composition. Secondly, my custom paired-end, single-indexed libraries were pooled with paired-end, dual-indexed libraries, which were provided by other customers. It was conceivable that the mixture of libraries prepared by different methods could interfere with each other (e.g. cross-hybridization) and consequently lead to failure in sequencing. However, I was not able to verify this assumption. Another possible issue associated with the mix of sequencing libraries could originate from the different index 1 (i7) size length used in the different libraries. To achieve sample multiplexing I used a custom six nucleotides barcode for index 1 (i7) (barcode for index 2 (i5) has 8 nucleotides) whilst in the other libraries an eight nucleotides index 1 (i7) barcode was used. This could have impacted the whole sequencing process.
via interference with the index sequencing read stage. It is possible that during the sequencing of index 1 (i7) the sequencer completed the first six sequencing cycles without problems but encountered an un-identified issue during the remaining two cycles. As a result, the sequencer could have interpreted the issue as a major malfunction and consequently discarded the already acquired sequencing data. I was not able to confirm this hypothesis. Alternatively, it is also possible that an issue arose from the demultiplexing procedure that takes place once the whole sequencing run is finished. The software used for the demultiplexing could have struggled to identify my six nucleotides-indexed samples. If it was the case, the acquired sequencing data consisting of sequencing artefacts or sequences that were not assign to a sample, mine included, would have been saved but placed into a ‘garbage’ file. However, no sequences in relation to my target regions were present in that file. Finally, another source of failure may come from the failure of the sequencing primers to anneal to the DNA molecules.

3.3.2 Second experimental design

3.3.2.1 Design requirements

The original BSPP design was modified to take into account and minimise the previously discussed possible sources of errors.

i. Only seven oligonucleotides were designed to reduce the cost of the experiment. The chosen target regions were \textit{PPIEL:Ex1-DMR}, \textit{ZDBF2/GPR1:IG-DMR}, \textit{FAM50B:TSS-DMR}, \textit{H19/IGF2:IG-DMR},
ii. The padlock probes were ordered from Integrated DNA Technologies company, which allowed more freedom in the design. This included the synthesis of a higher oligonucleotides concentration, which meant avoiding the initial round of padlock probes amplification (no need for AP1 and AP2 primer sites), digestion and purification; the design of a longer padlock probes (101-111 bp; 70 bp previously) which meant incorporating longer ligation arms (up to 70 bp combined; up to 40 bp previously) for more specificity during the target capture step, and longer common linker for more primers padding sequence during the capture circle amplification step.

iii. The size of the target inserts was changed to be between 120 to 185 bp (120 to 125 bp previously) which gave more flexibility to ppDesigner software to find regions that fit the design requirements.

iv. Custom sequencing primers (Diep et al. 2012) were used in place of the proprietary ones from Illumina. Although this approach is not supported by the company, the custom oligonucleotides were previously reported to perform well with both the library preparation protocol and the sequencing with MiSeq (Diep et al. 2012).

v. Different PCR amplification conditions were tested during the capture circle amplification step (see section 3.2.5.6 - Capture circle amplification for condition 1, 2 and 3 details). This may prevent the formation of non-specific, and truncated, products and increase the yield of enrichment. Therefore, this
may result in libraries with better quality, hence increasing the chance of optimal sequencing.

3.3.2.2 Library preparation

Condition 1, 2 and 3 were used to generate and enrich for the libraries. According to Bioanalyzer analysis performed before size selection and clean-up step, the condition 1 used to amplify the captured targets failed to generate amplicons, which suggested that the PCR condition was inadequate. Condition 2 and condition 3 had similar outcome. Both generated amplicons but the majority of them were likely to be non-specific product. Additionally, Bioanalyzer analysis also indicated that library concentration was within the pg/µl range, hence indicated poor yield during amplification (Figure 3.6). This could be explained by non-optimal PCR conditions during the capture circle amplification stage. Alternatively, this could also be explained by the fact that only seven regions were targeted and captured, as oppose of the hundreds in the previous design.

The libraries were then size selected and cleaned-up. Bioanalyzer analysis failed to detect any product within the samples (data not shown). This suggested that size selection and clean-up procedures led to the complete or partial loss of samples. DNA quantification with Qubit High Sensitivity kit detected very low amount of DNA, hence confirming the size selection and clean-up procedures were, in part, responsible for the loss of samples (data not shown). However, whilst the concentration was too low for the Bioanalyzer detection threshold, it was still
concentrated enough for MiSeq sequencing (only 2-10 pM is needed for sequencing).

Figure 3.6. BSPP sequencing library preparation (second design).

Sequencing library preparation with second BSPP design and reviewed method. (a) The captured circle amplified with PCR condition 1 did not generate amplicons whilst (b) using PCR condition 2 and (c) condition 3 resulted in a pool of specific and non-specific products.
### 3.3.2.3 Sequencing

The two libraries that were generated using condition 2 and condition 3 were sequenced using MiSeq sequencer. Unfortunately, the sequencing of the samples did not produce a successful outcome. Among the few reads that were acquired during the sequencing run, only a very few (<1%) aligned to my target regions whilst the vast majority (>99%) were either artefacts or garbage sequences that aligned to the common linker of the padlock probes. This was unexpected as the original single stranded free form of the oligonucleotides is expected to be digested during the exonuclease I and III enzymatic treatment. Furthermore, the design of the experiment was done in such way that even in the presence of carried over free form padlock probes, they should not be PCR amplified during the capture circle amplification step. The long primers used for library enrichment during the capture circle amplification step were designed to amplify specifically circle DNA that contains both the common linker of the padlock probes and the target regions but not single stranded linear padlock probes only. Finally, in the absence of amplification, the free form padlock probes should not contain any P5 and P7 adaptor sequences, which are both required to bind to the flow-cell. Alternatively, the failed sequencing may have also originated from failure of the custom sequencing primers to bind to the libraries during read 1 or read 2 sequencing stages. However, according to previous reports that made use of them with success (Deng et al. 2009; Diep et al. 2012) it is unlikely that they were responsible for the failed sequencing.

Taken together, the lack of reads aligning to target regions but instead aligning to
3.4 **Discussion**

Padlock probes based assay have been used successfully in the past to complete large projects that include exome sequencing, RNA allelotyping, copy number variation analysis, and SNP identification (Porreca et al. 2007; J.B. Li, Gao, et al. 2009; J.B. Li, Levanon, et al. 2009; K. Zhang et al. 2009; H Wang et al. 2010; Gore et al. 2011; Noggle et al. 2011; O’Roak et al. 2012; Carvill et al. 2013; Hiatt et al. 2013; Nuttle et al. 2013; Yoon et al. 2015). The method has since been adapted for DNA methylation analysis and BSPPs were found to be highly specific and consistent in the characterisation of DNA methylation of a large number of CpG. Additionally, the method requires a low amount of DNA and can easily be scaled to adapt to the size of the genomic regions of interest and to the sample size of the project (Deng et al. 2009; Diep et al. 2012). Based on these reports, I decided to use BSPPs in the novel high-throughput genomic imprinting methylation assay I aimed to develop.

Two types of padlock probes design (i.e. first and second design) were tested and in both cases several sequencing libraries were generated using different experimental
conditions. Quality control steps using bioanalyzer was performed before and after final size selection and the best candidate libraries (i.e. bioanalyzer profile closer to the theoretical end product, higher yield) for both padlock probes design was sent for sequencing. On overall, the method encountered several issues which led to unsuccessful target capture, enrichment of the targets, sequencing and methylation profiling. The absence of additional quality control steps throughout the method, such as after the target capture (Figure 3.3-a) and after the enzymatic digestion of circulised target sequences (Figure 3.3-c), has made troubleshooting difficult. However, based on several pieces of evidence I hypothesised that the failure of the assay could have originated from different factors:

(i) In the first design, the size of the BSPPs may have been too short. This was a supplier constraint for which at the time of the project was not able to synthesise a pool of hundreds of oligonucleotides of more than 100 bp. The short size of BSPPs implied having two short locus-specific target recognition and hybridisation arms, hence resulting in a loss of specificity to recognise the bisulfite treated targeted regions. Additionally, the short BSPPs also meant that the common linker was also short (30 bp) and consequently it left little room to design annealing sequences for the two long (60 bp) primers used during the multiplex PCR amplification of capture circle DNA step (15 bp annealing sequence for each primer). The BSPPs size issue was addressed with the second experimental design in which longer oligonucleotides (i.e 100 bp versus 70 bp) were synthesised. However, I was not able to assess how that benefited to the second design as it also failed.
(ii) The sequencing primers may have been an issue during the sequencing library preparation and/or during sequencing. In the first design I opted for the use of a mixture of Illumina proprietary sequencing primers, hence avoiding possible incompatibility with the MiSeq sequencer. However, the downside of these primers was that their last 15 bp (i.e. the ones that anneal to the common linker) have a melting temperature difference of 5 °C, which meant possible loss of PCR efficiency during the capture circle amplification step and therefore a low efficiency to generate final sequencing libraries. This may have explained the presence of a non-specific amplicon in the pool of sequencing libraries. Alternatively, the mixture of Illumina sequencing primers may have interfered with the sequencing itself as in both sequencing primers are not meant to work in combination on the same sequencing library. To address this possible problem, I opted for the use of the published custom sequencing primers that were originally designed with BSPPs method and were compatible, to some extent, with the MiSeq sequencer (Diep et al. 2012). However, I was not able to assess how that benefited to the second design as it also failed.

(iii) Finally, although originally not an issue, the 6 nucleotides index used in the first experimental design may have interfered with sequencing once my libraries were pooled with 8 nucleotides indexed sequencing libraries. The differences of index length may have led to sequencer malfunction or issue during demultiplexing. To avoid further complication, I included 8 nucleotides index in the second experimental design. However, I was not able to assess how that benefited to the second design
as it also failed.

In conclusion, various factors that were hypothesised to be a possible source of problems in the first place were taken into account during a second experimental design. However, despite the taken measures, this resulted in an unsuccessful attempt. Whilst the exact causes of failures are still unknown, it is likely that a combination of mentioned and unknown parameters may have influenced the assay and led to the failure in the target capture, enrichment of the targets and sequencing.
Chapter 4

Methylation Profiling in Beckwith-Wiedemann Syndrome

and Silver-Russell Syndrome
4.1 Introduction

4.1.1 DNA methylation profiling using the Infinium HumanMethylation 450K BeadChips

The Infinium HumanMethylation 450K BeadChips (HM450K) is an array-based method developed for the genome-wide profiling of DNA methylation. Whilst the platform is limited to the characterisation of a very small proportion of the human CpG content (~ 2%), it remains a popular choice in the analysis of large sample cohorts for which it represents an affordable approach to generate DNA methylome data (Rakyan et al. 2011; Y. Liu et al. 2013; Michels et al. 2013). Furthermore, the method requires a low amount of DNA input (500 ng) which make it suitable for studies in which DNA is a limiting factor.

The HM450K platform contains more than 485,000 probes distributed throughout the genome. Of the pool of probes, 482,421 interrogate CpG sites (~ 2% of human CpG content), 3091 non-CpG sites and 65 interrogate random single-nucleotide polymorphism sites. They were selected from WGBS data and input from DNA methylation experts and they cover 99% of RefSeq genes with an average of 17 probes per gene, 96% of CpG islands from the UCSC database, CpG island shores, DNAse I hypersensitive sites and regulatory elements such as enhancers (Bibikova et al. 2011).

HM450K is constituted of two types of beads, Infinium I and Infinium II. Infinium I
beads were designed for the Infinium HumanMethylation27 BeadChip, which is the precursor of HM450K. The design of Infinium I beads consists of two probes per locus, one for the methylated allele and one for the unmethylated allele, and the base extension is the same for the methylated or unmethylated alleles (i.e. the fluorescent signal does not carry any information on the methylation status). Infinium II beads are specific for HM450K. The design of Infinium II beads consists of a single type of probe per locus, and the methylation information is obtained through dual channel single-nucleotide primer extension with labelled dideoxynucleotides on the methylation variable position of a CpG (Figure 4.1) (Bibikova et al. 2011; Dedeurwaerder et al. 2014; Morris and Beck 2015).

Infinium I and Infinium II also differ in the reported derived methylation values, which for both is reported as beta-value ($\beta = \frac{\text{intensity of the Methylated allele (M)}}{\text{intensity of the Unmethylated allele (U) + intensity of the Methylated allele (M) + 100}}$) and where $\beta = 0$ is equivalent to unmethylated and $\beta = 1$ is equivalent to 100% methylated. Infinium II probes were found to exhibit a much lower dynamic range of methylation values compared to Infinium I probes (0 to 0.922 for Infinium II versus 0 to 0.971 for Infinium I) (Dedeurwaerder et al. 2011). Infinium II probes were also found to be less sensitive for the detection of extreme methylation values (i.e. 0 and 1) and less reproducible than Infinium I (standard deviation = 0.029 for Infinium II versus 0.008 for Infinium I) (Dedeurwaerder et al. 2011). Therefore, it is important to normalise the data to correct for the Infinium I and Infinium II probe differences. Several methods have been developed to address these issues (Cazaly et al. 2016).
These include Subset-quantile Within Array Normalisation (SWAN) (Maksimovic et al. 2012), Dasen (Pidsley et al. 2013), Beta-Mixture Quantile dilation (BMIQ) (Teschendorff et al. 2013) and Functional normalisation (FunNorm) (Fortin et al. 2014).

Figure 4.1: Overview of the Infinium I and Infinium II bead types.
(a) Infinium I and (b) Infinium II present on the HM450K. M, Methylated; U, Unmethylated. (From Dedeurwaerder et al.:2011).
4.1.2 Aim

Following the unsuccessful design of a novel NGS strategy for the methylation profiling of imprinting regions in BWS and SRS, I undertook the characterisation of the nature and frequency of MLID in a large cohort of BWS and SRS individuals using the robust HM450K platform. To that end, I developed a bioinformatic pipeline that can preprocess and handle large cohort methylation dataset and detect with accuracy epimutations at targeted regions, including imprinting DMRs (Table 3.1), in BWS and SRS individuals. Finally, I screened the patients methylomes to identify genomic regions that appeared to have an aberrant methylation profile and could possibly be involved in the molecular aetiology of the diseases.

4.2 Materials and methods

4.2.1 Patient blood DNA

According to the criteria described in the literature (http://www.geneclinics.org), BWS was diagnosed in 90 screened patients and SRS was diagnosed in 21 screened patients. For BWS, the molecular diagnostic investigations undertaken in NHS laboratories identified GOM at *H19/IGF2*:IG-DMR (BWS_IC1) in 4 individuals, paternal UPD11 (BWS_UPD11) in 8 individuals and LOM at *KCNQ1OT1*:TSS-DMR (BWS_IC2) in 78 individuals. For SRS, the molecular diagnostic identified LOM at *H19/IGF2*:IG-DMR (SRS_IC1) in 15 individuals and maternal UPD7 (SRS_UPD7) in 6 individuals. Written informed consent was obtained and ethical approval was
obtained from South Birmingham Research Ethics committee. For the 1472 healthy individuals that were included in my control group, 30 originated in-house whilst the remaining 1442 originated from collaborators.

4.2.2 DNA quantification
See section 2.3 - DNA quantification.

4.2.3 Sodium bisulfite treatment
See section 2.4 - Sodium bisulfite treatment.

4.2.4 DNA methylation profiling using HM450K
In order to obtain the genome wide methylation profiles, genomic DNA from patients with BWS (n=90) or SRS (n=21) and normal control individuals (n=1472; 30 individuals in-house and 1442 individuals from collaborators) was bisulfite treated (see section 2.4 - Sodium bisulfite treatment), processed on the HM450K (see section 2.5 - Infinium HumanMethylation450 BeadChip) and analysed as described below (Figure 4.2).

The sample preparation (i.e. DNA quantification and dilution prior HM450K) of the 111 BWS and SRS patients and of 30 out of 1472 normal individuals was done by me. The methylation profiling of these individuals was then done at the Cambridge Genomic Services. The sample preparation of the remaining 1442 normal individuals
was done by collaborators and, of those, the methylation profiling of 1402 was also
done at the Cambridge Genomic Services. The methylation profiling of the remaining
40 normal individuals was done at a different platform in Spain.

---

**Figure 4.2: Overview of the genome-wide methylation profiling workflow.**

4.2.4.1 Bioinformatic pipeline to analyse HM450K methylation dataset

4.2.4.1.1 HM450K data preprocessing

Data filtering and normalisation were performed in R with RnBeads (Assenov et al.
2014). Samples with detection P-value > 0.01 were filtered out. Infinium probes
containing SNPs and sites with excessive missing values were filtered out. In addition, probes with detection P-value > 0.01 and probes that lack signal values in one or more of the DNA samples analysed were also filtered out. Probe intensities were then normalised with beta-mixture quantile (BMIQ) normalisation (Teschendorff et al. 2013). Finally, probes located on sex chromosomes and in a non-CpG context were removed to discard any bias within samples. β values (i.e. equivalent to the absolute DNA methylation levels) and M-value (i.e. logistically transformed β values) were then generated for each probes. The R script used to process the data is available in Appendix 1.

4.2.4.1.2 HM450K data batch correction

Methylation data were corrected for technical (e.g. array slides and array positions) and biological (e.g. age) effects. I used ComBat method (Johnson et al. 2007) included in the sva (Leek et al. 2012) R package to perform the correction. The R script used to process the data is available in Appendix 2.

4.2.4.1.3 Epimutation methodology

To define LOM and GOM, I extracted the methylation value (β-value) of each HM450K CpG probes within target regions for all patients and normal individuals. I then computed the mean methylation index (MI) at each target regions for all individuals (i.e. patients and controls) and calculated MI standard errors (SE) for all patients. I also calculated the grand mean MI and associated standard deviation
(SD) for normal individuals. At each target region, samples were considered demonstrating GOM (above) or LOM (below) if both conditions were filled:

i. their MI and SE were above or below 3 SDs calculated from the grand mean of the normal individuals group. 99.73% of values lie within 3 SDs of the mean. Consequently, using such stringent threshold should significantly limit the discovery of false positives.

ii. at least 60% of the patient interrogated probes were above or below the 3 SDs threshold (Figure 4.3).

![Graph showing epimutation methodology](image)

**Figure 4.3: Illustration of epimutation methodology.**

Additionally, polymorphic methylation change at each target regions were assessed by Fisher’s exact test followed by false discovery rate (FDR) correction. Statistical tests were performed in R. Aberrant methylation, either GOM or LOM, with FDR ≥
0.05 were considered polymorphic (Table 4.6).

Table 4.6: An example of statistical test at the imprinted locus PLAGL1_alt-TSS-DMR.

<table>
<thead>
<tr>
<th></th>
<th>LOM is not polymorphic</th>
<th>GOM is polymorphic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PLAGL1_alt-TSS-DMR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOM</td>
<td>Other</td>
<td>GOM</td>
</tr>
<tr>
<td>Cases</td>
<td>5</td>
<td>106</td>
</tr>
<tr>
<td>Controls</td>
<td>7</td>
<td>1432</td>
</tr>
<tr>
<td>FDR</td>
<td>0.003</td>
<td>FDR</td>
</tr>
</tbody>
</table>

LOM, loss-of-methylation; GOM, gain-of-methylation; Other, either GOM + no methylation change or LOM + no methylation change

Establishing methylation profiles for all BWS and SRS patients across the targeted regions (e.g. imprinted DMRs) was performed in R. Providing the manual input of the required variables, the R script will automatically load methylation data, subset and save methylation data associated with CpG probes overlapping the region of interest, compute GOM and LOM in patient and in control individuals and save a corrected (i.e. all reported loci that are not polymorphic) and uncorrected (i.e. some reported loci may be polymorphic) epimutation table for all patients. The R script used to process the data is available in Appendix 3.

4.2.4.1.4 DMR hunting

Bumphunter (Jaffe et al. 2012; Aryee et al. 2014) function was used to identify genomic regions that were differentially methylated between two conditions (e.g.
BWS_IC1 patients versus control individuals). The function was run in R and the script used to process the data is available in Appendix 4. It first computed a t-statistic with smoothing at each genomic location. Then, it defined a candidate region to be a cluster of probes for which all the t-statistics exceeded a predefined threshold (delta $\beta \geq 0.15$ between the two groups) and significance of the candidate regions was calculated using permutations test ($n=1000$). Candidate regions with FDR > 0.05, FWER (family-wise error rate) > 0.3 and encompassing less than three probes were filtered out.

4.3 Results

4.3.1 Data pre-processing

4.3.1.1 Normalisation

In normal conditions, the majority of CpG sites can be either methylated or unmethylated. On the HM450K, this assumption can be defined with a density methylation peak at $\beta \sim 0$ (i.e. 0% methylated) and $\beta \sim 1$ (i.e. 100% methylated). In the uncorrected HM450K dataset (i.e. 21 BWS, 91 SRS and 1471 normal control individuals), I observed a peak of methylation distribution at $\beta \sim 0.2$ (i.e. 20% methylated) and another peak at $\beta \sim 0.7$ (i.e. 70% methylated) (Figure 4.4, red line). This could be explained by the chemistry differences between the two types of probes within HM450K array (i.e. Infinium I and Infinium II). Indeed, Infinium II probes are known to be less accurate than Infinium I probes and therefore, data needed to be corrected prior to analysis (Dedeurwaerder et al. 2011). Hence, methylation
values were normalised with Beta Mixture Quantile (BMHQ) normalisation algorithm (Teschendorff et al. 2013) included in RnBeads R package (Assenov et al. 2014). After correction, peaks of density methylation shifted from $\beta \sim 0.2$ to $\beta \sim 0.05$ and from $\beta \sim 0.7$ to $\beta \sim 0.9$ (Figure 4.4, blue line), illustrating the effectiveness of the data normalisation.

Figure 4.4: Effect of the correction on methylation value. Infinium I and Infinium II $\beta$ value (i.e. methylation value) were corrected with BMHQ normalisation algorithm included in RnBeads.

4.3.1.2 Filtering

After normalisation, problematic samples and probes were filtered out. Out of the 1583 samples, 33 control individuals were removed due to bad sample P-value (higher than 0.01). Out of the 485577 probes present on the HM450K array, 413997
were retained for analysis. The remaining 71580 probes were removed because 12416 probes overlapped with SNPs in their last 5 bases of their targets, 29675 probes contained non-specific sequences and had a high likelihood of cross-hybridisation, 19071 probes had low quality (P-value > 0.01), 1074 probes were not in CpG contexts and 9344 probes were on sex chromosomes Figure 4.5).

**Figure 4.5: Proportion of retained and removed probes.**
Probes overlapping SNPs, non-specific, with P-value above 0.01, on sex chromosomes or in a non-CpG context were removed from analysis.

**4.3.1.3 Batch effect and biological artefacts correction**
Batch effects (Leek et al. 2010) are systematic biases in the data that are unrelated to the research question but that arise from undesirable differences in sample handling (i.e. chip or instrument used, date of experiment, technician running samples, etc.). The control individuals methylation dataset has been processed on
the same platform (Cambridge Genomic Services) but on different date, different
HM450K array slides, and by different technicians, resulting in potential batch
effects. Although not all technical variabilities were known, the array slides and array
positions were accounted for batch effects correction.

Biological artefacts are unwanted variations (i.e. gender, sex, tissue, etc.) that could
also lead to biases in the data. As an example, loss of DNA methylation is observed
in ageing (Wilson et al. 1987; Fraga et al. 2007). BWS and SRS are both paediatric
disorders. Hence, it was important to use an age-match control individual population
for comparison. Age of patient and control individuals were first estimated ((Horvath
2013), https://labs.genetics.ucla.edu/horvath/dnamage/). An age difference was
observed between the disease patients group (mean: 6.9 years old) and normal
individual controls group (mean: 56 years old) (Table 4.7). Age of all individuals were
then categorised into 9 groups (A, 0 to <10 years old; B, 10 to <20 years old; C, 20
to <30 years old; D, 30 to <40 years old; E, 40 to <50 years old; F, 50 to <60 years
old; G, 60 to <70 years old; H, 70 to <80 years old; I, 80 to <90 years old) and the
age variable was accounted for biological artefacts correction.

Table 4.7: Age distribution in disease and normal individual groups.

<table>
<thead>
<tr>
<th></th>
<th>Mean age (in years)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease patient group</td>
<td>6.9</td>
<td>9.6</td>
</tr>
<tr>
<td>(n=111)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal individual group</td>
<td>56</td>
<td>13.2</td>
</tr>
<tr>
<td>(n=1439)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To remove both technical (i.e. array slides and array positions) and biological (i.e. age) effects, I used the ComBat method (Johnson et al. 2007) included in the sva R package (Leek et al. 2012). Principal component analysis before applying ComBat showed variable data (multiple clusters) most likely related to technical and biological effects as further confirmed by principal component regression analysis (Figure 4.6-a and b). After applying ComBat on the normalised methylation data, the variability observed among the 1550 individuals is now mostly associated with the disease status of individuals (i.e. Sample_Group) and most of the bias associated with both technical batch effects (i.e. Sentrix_Position and Sentrix_ID) was efficiently removed from data. Bias associated with age was also removed but in a lesser extent (Figure 4.6-c and d). The age of individuals was confounded with their disease status (i.e. young BWS and SRS patients, and old normal control individuals), thus resulting in ComBat being less efficient to address this issue.
Figure 4.6: Effects of technical and biological artefacts.

On normalised data, the 10000 most variable probes were analysed by principal component analysis and by principal component regression analysis (a-b) before and (c-d) after ComBat correction. Principal component regression analysis was performed in R using ENmix (Xu et al. 2016).

4.3.2 Targeted methylation profiling

I first investigated and compared the methylation index (MI) at 46 targeted imprinted DMRs in BWS and SRS patients with the MI in a group of normal individuals. I defined a MI in patients above or below 3 standard deviations calculated from the grand mean of the normal individuals as hyper- or hypomethylated respectively.
4.3.2.1 Methodology validation

My method based on three standard deviations epimutation threshold was validated by investigating MI at $H19/IGF2$:IG-DMR and $KCNQ1OT1$:TSS-DMR in five groups of patients (i.e. BWS_IC1, BWS_IC2, BWS_UPD11, SRS_IC1 and SRS_UPD7). In our cohort, MI at both loci have been previously tested in a clinical setting for diagnostic purposes and epimutations, if any, were already known. Using HM450K array and my methodology, the $H19/IGF2$:IG-DMR locus was hypermethylated in all (n=4) BWS_IC1 patients and in 7 out of 8 patients with BWS_UPD11, hypomethylated in all (n=15) SRS_IC1 patients, and a MI within the normal range was observed in all BWS_IC2 (n=78) and SRS_UPD7 (n=6) patients and in 1 out of 8 patient with BWS_UPD11 (Figure 4.7-a). The $KCNQ1OT1$:TSS-DMR locus was hypomethylated in all BWS_IC2 and BWS_UPD11 patients, and normal methylation was observed in all BWS_IC1, SRS_IC1 and SRS_UPD7 patients (Figure 4.7-b). Although this methodology was very stringent and conservative, the results were mostly consistent with the results from clinical molecular testing using methylation-specific multiplex ligation-dependent probe amplification. I found complete correlation between the results from clinical testing and HM450K analysis of the methylation patterns at $H19/IGF2$:IG-DMR and $KCNQ1OT1$:TSS-DMR in 110 of 111 patients. The exception was one individual with BWS diagnosed with pUPD11 in whom the expected methylation profile was LOM at $KCNQ1OT1$:TSS-DMR and GOM at $H19/IGF2$:IG-DMR.
Figure 4.7: Methodology validation.

Methylation index (β-value) of all patients was measured at (a) H19/IGF2:IG-DMR and (b) KCNQ1OT1:TSS-DMR. The black box represents normal individuals range: top of the box is 3 standard deviations above control individuals grand mean, the bottom of the box is 3 standard deviations below control individuals grand mean and the middle line is grand mean of control individuals. BWS_IC1, BWS patient with IC1 GOM; BWS_IC2, BWS patients with IC2 LOM; BWS_UPD11, BWS patients with pUPD11; SRS_IC1, SRS patients with IC1 LOM; SRS_UPD7, SRS patients with mUPD7.

4.3.2.2 Polymorphic imprinted DMRs

Following method validation, methylation profiling was performed at 46 imprinted DMRs (Table 3.1) across all individuals. Whilst multiple epimutations were observed in our disease patient group it was necessary to discriminate between methylation
abnormalities that were (i) probably linked to the disorder (i.e. outcome of interest),
(ii) random epigenetic events or (iii) common within general population (Tee et al.
2013). Although discriminating between the random occurrence of epigenetic events
and linkage to the disorder was complicated, it was possible to assess whether or
not the observation was common in the general population (i.e. polymorphic). In this
regard, the number of GOM and LOM findings were calculated for all loci in both
populations and compared using Fisher’s exact test followed by false discovery rate
(FDR) adjustment using the Benjamini-Hochberg procedure (Benjamini and
Hochberg 1995). For GOM, out of the 46 investigated imprinted DMRs 12 had MI
within normal individual range and 34 loci were hypermethylated, of which 8 were
significant (FDR <0.05) and 26 were polymorphic. For LOM, 6 imprinted DMRs had
MI within normal individual range and 40 loci were found hypomethylated, of which
22 were significant (FDR <0.05) and 18 were polymorphic (Table 4.8).
## Table 4.8: Imprinted DMRs with polymorphic epimutations.

<table>
<thead>
<tr>
<th>Imprinted DMR</th>
<th>Gain-of-methylation</th>
<th>Loss-of-methylation</th>
<th>Number of 450K Illumina probes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient with MLD/no MLID</td>
<td>Normal control individuals with MLD/no MLID</td>
<td>Outcome</td>
</tr>
<tr>
<td>PPIE: Ex1-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>DIRAS3: TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>DIRAS3: Ex2-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>GPR1A: TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>ZDBF2: GPR1A:IG-DMR</td>
<td>6/105</td>
<td>0/1439</td>
<td>significant</td>
</tr>
<tr>
<td>NAP1L5: TSS-DMR</td>
<td>1/110</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>VDRN2: TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>FAM108B: TSS-DMR</td>
<td>0/111</td>
<td>3/1436</td>
<td>polymorphic</td>
</tr>
<tr>
<td>PLAGL1: alt-TSS-DMR</td>
<td>1/110</td>
<td>3/1436</td>
<td>polymorphic</td>
</tr>
<tr>
<td>IGFB2: Ex9-DMR</td>
<td>2/109</td>
<td>29/1430</td>
<td>polymorphic</td>
</tr>
<tr>
<td>WDR27: Int3-DMR</td>
<td>0/111</td>
<td>2/1437</td>
<td>polymorphic</td>
</tr>
<tr>
<td>NRR3: TSS-DMR</td>
<td>7/110</td>
<td>7/1438</td>
<td>significant</td>
</tr>
<tr>
<td>PEG10: TSS-DMR</td>
<td>6/103</td>
<td>5/1434</td>
<td>significant</td>
</tr>
<tr>
<td>MEST: alt-TSS-DMR</td>
<td>7/104</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>HTRSA: TSS-DMR</td>
<td>7/104</td>
<td>4/1435</td>
<td>polymorphic</td>
</tr>
<tr>
<td>ERLN2: Int9-DMR</td>
<td>0/111</td>
<td>2/1437</td>
<td>polymorphic</td>
</tr>
<tr>
<td>PEG13: TSS-DMR</td>
<td>1/110</td>
<td>4/1435</td>
<td>polymorphic</td>
</tr>
<tr>
<td>FANCC: Int1-DMR</td>
<td>0/111</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>INPP5F: Int2-DMR</td>
<td>0/111</td>
<td>2/1437</td>
<td>polymorphic</td>
</tr>
<tr>
<td>HTGFR2: IG-DMR</td>
<td>11/100</td>
<td>0/1439</td>
<td>polymorphic</td>
</tr>
<tr>
<td>IGF2: Ex9-DMR</td>
<td>4/107</td>
<td>7/1432</td>
<td>polymorphic</td>
</tr>
<tr>
<td>IGF2: alt-TSS-DMR</td>
<td>11/100</td>
<td>1/1436</td>
<td>significant</td>
</tr>
<tr>
<td>KCNQ7: TSS-DMR</td>
<td>0/111</td>
<td>4/1435</td>
<td>polymorphic</td>
</tr>
<tr>
<td>RB1: Int2-DMR</td>
<td>0/111</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>MEG3: TSS-DMR</td>
<td>0/111</td>
<td>4/1435</td>
<td>polymorphic</td>
</tr>
<tr>
<td>MEG8: Int2-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>MKRN3: TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>MAGEL2: TSS-DMR</td>
<td>0/111</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>NDN: TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>SNRPW: alt-TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>SNURF: TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>IGF1: Int2-DMR</td>
<td>0/111</td>
<td>2/1431</td>
<td>polymorphic</td>
</tr>
<tr>
<td>ZNF977:3 DMR</td>
<td>1/110</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>ZNF587: TSS-DMR</td>
<td>2/109</td>
<td>2/1437</td>
<td>polymorphic</td>
</tr>
<tr>
<td>ZNF337: alt-TSS-DMR</td>
<td>0/111</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>ZNF337: alt-TSS-DMR</td>
<td>0/111</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>PEG2: TSS-DMR</td>
<td>0/111</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>MCT28: TSS-DMR</td>
<td>1/110</td>
<td>11/1428</td>
<td>polymorphic</td>
</tr>
<tr>
<td>ANAT: TSS-DMR</td>
<td>0/111</td>
<td>4/1435</td>
<td>polymorphic</td>
</tr>
<tr>
<td>L3MBTL1: alt-TSS-DMR</td>
<td>1/110</td>
<td>2/1437</td>
<td>polymorphic</td>
</tr>
<tr>
<td>GNAS: NE5P: TSS-DMR</td>
<td>2/109</td>
<td>1/1438</td>
<td>polymorphic</td>
</tr>
<tr>
<td>GNAS: xl: Ex1-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
<tr>
<td>GNAS: A8: TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>polymorphic</td>
</tr>
<tr>
<td>WDR4: alt-TSS-DMR</td>
<td>0/111</td>
<td>7/1432</td>
<td>polymorphic</td>
</tr>
<tr>
<td>SNUR3: alt-TSS-DMR</td>
<td>0/111</td>
<td>0/1439</td>
<td>NA</td>
</tr>
</tbody>
</table>

polymorphic, FDR<0.05; significant, FDR<0.05, NA, no methylation change detected in both patient and control individual groups.
4.3.2.3 Targeted methylation profiling of SRS patients at imprinted DMRs

4.3.2.3.1 SRS with mUPD7 cohort

Methylation profiling of 6 individuals in the SRS_UPD7 group showed hypermethylation at \textit{GRB10}:alt-TSS-DMR, \textit{PEG10}:TSS-DMR, \textit{MEST}:alt-TSS-DMR and \textit{HTR5A}:TSS-DMR in all patients. This was expected as these DMRs are located on chromosome 7 and are methylated on the maternal allele. In addition, 1 patient had GOM at \textit{ZDBF2/GPR1}:IG-DMR and LOM at \textit{IGF1R}:Int2-DMR. MLID was detected in only 1 patient out of 6 (17%) (Figure 4.8).

4.3.2.3.2 SRS with IC1 LOM cohort

Methylation profiling of 15 individuals in the SRS_IC1 group showed, as expected, hypomethylation at \textit{H19/IGF2}:IG-DMR in all patients. In addition, hypomethylation at \textit{IGF2}:alt-TSS-DMR was also detected in 11 patients out of 15, but this was not considered as MLID as both \textit{H19/IGF2}:IG-DMR and \textit{IGF2}:alt-TSS-DMR belong to the same imprinted cluster. Finally, one SRS_IC1 patient had an additional LOM at \textit{L3MBTL1}:alt-TSS-DMR and, interestingly, one other patient had GOM at imprinted DMRs (\textit{GRB10}:alt-TSS-DMR, \textit{PEG10}:TSS-DMR, \textit{MEST}:alt-TSS-DMR and \textit{HTR5A}:TSS-DMR) located on chromosome 7 and normally associated with SRS_UPD7 aetiology (Figure 4.8). Taken together, these results suggested that MLID occurred in approximatively 13% (2/15) of SRS_IC1 patients.
Figure 4.8: Multilocus imprinting disturbances and SRS.

Epimutations at imprinted DMRs in SRS_UPD7 (n=6) and SRS_IC1 (n=15). Epimutations distribution in the different SRS group of patients (x-axis) across the 46 interrogated imprinted DMRs (y-axis). Red, loss-of-methylation; green, gain-of-methylation; black: no epimutations.

4.3.2.4 Targeted methylation profiling of BWS patients at imprinted DMRs

4.3.2.4.1 BWS with pUPD11 cohort

As previously mentioned (see section 4.3.2.1 - Methodology validation) methylation profiling of 8 individuals in the BWS_UPD11 group showed GOM at H19/IGF2:IG-DMR and IGF2:alt-TSS-DMR in 7 out of 8 patients, GOM at IGF2:Ex9-DMR in 1 out of 8 patients and LOM at KCNQ1OT1:TSS-DMR in all patients. For 7 out of 8 patients, these results are consistent with clinical diagnosis. As expected, H19/IGF2:IG-DMR is hypermethylated and KCNQ1OT1:TSS-DMR is hypomethylated. For one patient, we failed to detect hypermethylation at
Finally, 1 out of 8 (12.5%) patient has MLID. The patient has hypermethylation at $PEG10$:TSS-DMR (Figure 4.9).

### 4.3.2.4.2 BWS with IC1 GOM cohort

Methylation profiling of 4 individuals in the BWS_IC1 group showed, as expected, hypermethylation at $H19/IGF2$:IG-DMR and $IGF2$:alt-TSS-DMR in all patients and two patients had hypermethylation at $IGF2$:Ex9-DMR. MLID was not detected in this group of patient (Figure 4.9).

### 4.3.2.4.3 BWS with IC2 LOM cohort

Methylation profiling of 78 individuals in the BWS_IC2 group showed, as expected, LOM at $KCNQ1OT1$:TSS-DMR in all patients. Additionally, MLID was detected in 31 out of 78 (40%) of patients of this cohort. Of the 31 BWS_IC2 with MLID, 5 (16%) patients had GOM at $ZDBF2/GPR1$:IG-DMR and LOM was observed in 8 (26%) at $DIRAS3$:TSS-DMR, 5 (16%) at $DIRAS3$:Ex2-DMR, 8 (26%) at $FAM50B$:TSS-DMR, 5 (16%) at $MEST$:alt-TSS-DMR, 5 (16%) at $IFG2R$:Int2-DMR, 4 (13%) at $MEST$:alt-TSS-DMR, 14 (45%) at $FANCC$:Int1-DMR, 2 (6%) at $INPP5F$:Int2-DMR, 2 (6%) at $SNURF$:TSS-DMR, 5 (16%) at $IGF1R$:Int2-DMR, 3 (10%) at $ZNF331$:alt-TSS-DMR1, 3 (10%) at $ZNF331$:alt-TSS-DMR2, 3 (10%) at $NNAT$:TSS-DMR, 10 (32%) at $L3MBTL1$:alt-TSS-DMR, 3 (10%) at $GNAS-AS1$:TSS-DMR, 2 (6%) at $GNAS-XL$:Ex1-DMR, 3 (10%) at $GNAS-AB$:TSS-DMR, 3 (10%) at $WRB$:alt-TSS-DMR and 15 (48%) at $SNU13$:alt-TSS-DMR (Figure 4.9 and Figure 4.10).
Figure 4.9: Multilocus imprinting disturbances and BWS.

Epimutations at imprinted DMRs in BWS_UPD11 (n=8), BWS_IC1 (n=4) and BWS_IC2 (n=78). Epimutations distribution in the different BWS group of patients (x-axis) across the 46 interrogated imprinted DMRs (y-axis). Red, loss-of-methylation; green, gain-of-methylation; black: no epimutations.
Figure 4.10: Comparison of the frequency of epimutated DMRs in BWS_IC2 with MLID patients.
MLID frequency calculated in the current study was significantly higher (p=0.0051) than the frequency obtained from seven combined previous reports. However, when compared to two previous studies that also used HM450K for methylation profiling (Court et al. 2013; Maeda et al. 2014), the MLID detection rates were similar (p=0.421) (Table 4.9).

Compared to the imprinted DMRs that were investigated in at least 3 other studies, the frequency of disturbances calculated in my cohort was concordant with previous reports at PEG10:TSS-DMR, H19/IGF2:IG-DMR, MEG3:TSS-DMR, SNRPN:alt-TSS-DMR, PEG3:TSS-DMR. Those loci were found mostly unaffected in the majority of studies, including mine. However, the frequency of disturbances observed at PLAGL1:alt-TSS-DMR, IGF2R:Int2-DMR, GRB10:alt-TSS-DMR, MEST:alt-TSS-DMR, GNAS-NESP:TSS-DMR, GNAS-AS1:TSS-DMR, GNAS-XL:Ex1-DMR was below the 25th percentile calculated from the 7 previous studies at those DMRs. Nonetheless, it was noted that the observed frequency of epimutations reported previously was very variable between studies. As an example, PLAGL1:alt-TSS-DMR was epimutated in 16% in our cohort and in 39% (median) with a minimum of 7%, a maximum of 63% and a standard deviation of 18.6 in previous studies. Similarly, MEST:alt-TSS-DMR was epimutated in 13% in our cohort and in 35% (median) with a of minimum 7%, a maximum of 61% and a standard deviation of 17.8 in previous studies.
Table 4.9: Frequency of disturbances reported in the current and seven previous reports.

<table>
<thead>
<tr>
<th></th>
<th>MLID</th>
<th>with no MLID</th>
<th>MLID frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current study</td>
<td>31</td>
<td>47</td>
<td>40%</td>
</tr>
<tr>
<td>Combined previous reports</td>
<td>113</td>
<td>360</td>
<td>24%</td>
</tr>
<tr>
<td>Rossignol et al. 2006</td>
<td>10</td>
<td>30</td>
<td>25%</td>
</tr>
<tr>
<td>Bliek et al. 2008</td>
<td>17</td>
<td>64</td>
<td>21%</td>
</tr>
<tr>
<td>Azzi et al. 2009</td>
<td>16</td>
<td>52</td>
<td>24%</td>
</tr>
<tr>
<td>Poole et al. 2013</td>
<td>8</td>
<td>13</td>
<td>38%</td>
</tr>
<tr>
<td>Tee et al. 2013</td>
<td>33</td>
<td>143</td>
<td>18%</td>
</tr>
<tr>
<td>Court et al. 2013</td>
<td>14</td>
<td>29</td>
<td>33%</td>
</tr>
<tr>
<td>Maeda et al. 2014</td>
<td>15</td>
<td>29</td>
<td>34%</td>
</tr>
</tbody>
</table>

**Fisher's exact test (p-value calculated from current study against the 7 previous reports)**

0.0051

**Fisher's exact test (p-value calculated from current study against Court et al. and Maeda et al. reports)**

0.421
Figure 4.11: Meta-analysis of reported MLID frequency in BWS_IC2 cohort.

Comparison of the frequency of disturbances across imprinted DMRs in BWS_IC2 MLID observed in the current study (red triangle) and in 7 previous studies (colour dots). Box-and-whisker plots were calculated from the 7 previous reports indicated in plot key.
The number of additional epimutated DMRs per patient ranged from 1 to 12. The majority of patients had only 1 (10/31) or 2 (8/31) additional epimutated DMRs and the remaining 13 patients had between 3 and 12 additional epimutated DMRs, indicating that the more additional epimutated DMRs were observed the fewer patients were affected (Figure 4.12-a). Within these 31 patients, I counted a total of 139 epimutations affecting 21 imprinted DMRs. Out of the 139 observed epimutations, 134 were LOM affecting 20 maternal imprinted DMRs (DIRAS3:TSS-DMR, DIRAS3:Ex2-DMR, FAM50B:TSS-DMR, PLAGL1:alt-TSS-DMR, IGF2R:Int2-DMR, MEST:alt-TSS-DMR, FANCC:Int1-DMR, INPP5F:Int2-DMR, KCNQ1OT1:TSS-DMR, SNURF:TSS-DMR, IGF1R:Int2-DMR, ZNF331:alt-TSS-DMR, ZNF331:alt-TSS-DMR, NNAT:TSS-DMR, L3MBTL1:alt-TSS-DMR, GNAS-AS1:TSS-DMR, GNAS-XL:Ex1-DMR, GNAS-AB:TSS-DMR, WRB:alt-TSS-DMR, SNU13:alt-TSS-DMR). The remaining 5 disturbances were GOM affecting the paternal DMRs ZDBF2/GPR1:IG-DMR (Figure 4.12-b). This suggested that LOM was more common than GOM in BWS_IC2, and that LOM occurred at maternal DMRs only whilst GOM was more associated with paternal DMRs. No epimutation preference toward maternal DMRs was found as statistical test failed to reach significance (p=0.106, Fisher’s exact test) (Figure 4.12-c).
Figure 4.12: Imprinting disturbances in BWS_IC2 patients.
(a) Comparison of the number of additional epimutated loci in BWS_IC2 with MLID patients. (b) Comparison of the total GOM and LOM in BWS_IC2 with MLID patients. (c) Comparison of the number of aberrantly methylated DMRs between matDMRs and patDMRs. Fisher’s exact test was used for statistical analysis. GOM, gain-of-methylation; LOM, loss-of-methylation; matDMR, maternal DMR; patDMR, paternal DMR.

Epimutation correlation and associated significance level were calculated for all imprinted DMRs detected to be epimutated in BWS_IC2 MLID patients. Using an absolute correlation threshold of at least 0.6 and significance level threshold of 0.01, IGF1R:Int2-DMR, GNAS-AS1:TSS-DMR, GNAS-XL:Ex1-DMR and GNAS A/B:TSS-
DMR had significant positive correlation. The two imprinted DMRs located at *DIRAS3* locus (*DIRAS3*:TSS-DMR and *DIRAS3*:Ex2-DMR) had significant positive correlation. The two imprinted DMRSs located at *ZNF331* locus (*ZNF331*:alt-TSS-DMR1 and *ZNF331*:alt-TSS-DMR2) had significant positive correlation. Finally, *NNAT*:TSS-DMR on chromosome 20 had significant positive correlation with *ZNF331*:alt-TSS-DMR1, *ZNF331*:alt-TSS-DMR2 on chromosome 19 and *PLAGL1*:alt-TSS-DMR on chromosome 6 (Figure 4.13).
Figure 4.13: Epimutations correlation calculated at epimutated imprinted DMRs in BWS_IC2 MLID patients.

Pearson’s product-moment correlation was used to compute correlation and associated p-value. Numeric values, r correlation; red, toward negative correlation; blue, toward positive correlation; black square, correlation with |r| ≥ 0.6 and p ≤ 0.01).
4.3.3 DMR hunting

Bumphunter (Jaffe et al. 2012; Aryee et al. 2014) R package was used to identify genomic regions that were differentially methylated between two conditions (e.g. BWS_IC1 patients versus control individuals). Resulting candidate regions were then processed with our epimutation methodology (see section 4.2.4.1.3 - Epimutation methodology) to retain the most interesting ones. Following this procedure, I detected DMRs at most of the well-described loci involved in the aetiology of each group of patients (H19/IGF2:IG-DMR, KCNQ1OT1:TSS-DMR, PEG10:TSS-DMR, MEST:alt-TSS-DMR and GRB10:alt-TSS-DMR). In addition, I found two new significant DMRs overlapping the promoter of the lymphotoxin alpha (LTA) gene (Figure 4.14-a) and in the exon 4 of the solute carrier family 12 member 9 (SLC12A9) gene (Figure 4.14-b) in BWS_IC2. LTA promoter was found hypomethylated in 7 out of the 78 (9%) BWS_IC2 patient and SLC12A9 exon 4 was found hypermethylated in 34 out of 78 (44%) patients (Figure 4.14-c). I also found two new significant DMRs overlapping SLC12A9 exon 4 (Figure 4.14-d) and in the promoter of the potassium voltage-gated channel subfamily A regulatory beta subunit 3 (KCNAB3) gene (Figure 4.14-e) in BWS_UPD11. To be noted that two probes within SLC12A9 DMR detected in BWS_UPD11 were not found significant in BWS_IC2, resulting in a smaller cluster in the latter group. SLC12A9 exon 4 was found hypermethylated in 6 out of 8 (75%) BWS_UPD11 patients and KCNAB3 promoter was found hypomethylated in 4 out of 8 (50%) patients (Figure 4.14-f). No additional significant DMRs were found in BWS_IC1, SRS_IC1 and SRS_UPD7.
Finally, although it was not detected by bumphunter, I also found that \textit{KCNAB3} promoter was hypomethylated in 14 out of 78 (18\%) BWS\_IC2 patients (Figure 4.15-a, b).

\textbf{Figure 4.14: DMR hunting in BWS.} (a,b,d,e) Methylation index (y-axis) of HM450K CpG probes (x-axis) of patients (orange) and normal individuals (blue) at (a) LTA (b) SLC12A9 in BWS\_IC2 and at (d) SLC12A9 and (e) KCNAB3 in BWS\_UPD11. Vertical black lines represent DMR boundaries defined by bumphunter. (c,f) Frequency of (c) BWS\_IC2 and (f) BWS\_UPD11 patients with hypo- or hypermethylation at the indicated loci
<table>
<thead>
<tr>
<th>name</th>
<th>region</th>
<th>description</th>
<th>Entrez</th>
<th>chromosome</th>
<th>start</th>
<th>end</th>
<th>delta β</th>
<th>number of probes</th>
<th>FDR</th>
<th>fwer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BWS_IC1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H19</td>
<td>promoter</td>
<td>H19/GF2:G-DMR, imprinted</td>
<td>283120</td>
<td>11</td>
<td>2019079</td>
<td>2020560</td>
<td>-0.264</td>
<td>40</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td><strong>BWS_IC2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCNQ1</td>
<td>inside intron</td>
<td>KCNQ1T1:TSS-DMR, imprinted</td>
<td>3784</td>
<td>11</td>
<td>2721207</td>
<td>2722440</td>
<td>0.331</td>
<td>36</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td>SLC12A9</td>
<td>inside exon 4</td>
<td>solute carrier family 12 member</td>
<td>56996</td>
<td>7</td>
<td>100463563</td>
<td>100464145</td>
<td>-0.158</td>
<td>4</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td>LTA</td>
<td>promoter</td>
<td>lymphotoxin alpha</td>
<td>4049</td>
<td>6</td>
<td>31539539</td>
<td>31539735</td>
<td>0.159</td>
<td>3</td>
<td>0.009</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>BWS_UPD11</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCNQ1</td>
<td>inside intron</td>
<td>KCNQ1T1:TSS-DMR, imprinted</td>
<td>3784</td>
<td>11</td>
<td>2721207</td>
<td>2722062</td>
<td>0.161</td>
<td>22</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td>H19</td>
<td>promoter</td>
<td>H19/GF2:G-DMR, imprinted</td>
<td>283120</td>
<td>11</td>
<td>2020279</td>
<td>2020560</td>
<td>-0.165</td>
<td>10</td>
<td>0.006</td>
<td>0.058</td>
</tr>
<tr>
<td>KCNAB1</td>
<td>promoter</td>
<td>potassium voltage-gated channel</td>
<td>9196</td>
<td>17</td>
<td>7832769</td>
<td>7833237</td>
<td>0.205</td>
<td>7</td>
<td>0.025</td>
<td>0.217</td>
</tr>
<tr>
<td>subfamily A regulatory beta subunit 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC12A9</td>
<td>inside exon 4</td>
<td>solute carrier family 12 member</td>
<td>56996</td>
<td>7</td>
<td>100463416</td>
<td>100464145</td>
<td>-0.207</td>
<td>6</td>
<td>0.029</td>
<td>0.252</td>
</tr>
<tr>
<td><strong>SRS_IC1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H19</td>
<td>promoter</td>
<td>H19/GF2:G-DMR, imprinted</td>
<td>283120</td>
<td>11</td>
<td>2019079</td>
<td>2020560</td>
<td>0.302</td>
<td>40</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td><strong>SRS_UPD7</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGCE</td>
<td>overlaps 5UTR - exon 1</td>
<td>PEG10:TSS-DMR, imprinted</td>
<td>8910</td>
<td>7</td>
<td>94285327</td>
<td>94287242</td>
<td>-0.350</td>
<td>59</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td>MEST</td>
<td>overlaps 5UTR - exon 1</td>
<td>MEST:alt-TSS-DMR, imprinted</td>
<td>4232</td>
<td>7</td>
<td>130129946</td>
<td>130132453</td>
<td>-0.350</td>
<td>56</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td>GRB10</td>
<td>inside intron - 5' UTR</td>
<td>GRB10:alt-TSS-DMR, imprinted</td>
<td>2887</td>
<td>7</td>
<td>50849639</td>
<td>50849931</td>
<td>-0.343</td>
<td>4</td>
<td>0.006</td>
<td>0.082</td>
</tr>
</tbody>
</table>

region: genomic features where significant clusters are located; delta β: mean methylation difference between normal and patient individuals; number of probes: number of 450K CpG probes within cluster, FDR: false discovery rate; fwer: family-wise error rate
Figure 4.15: KCNAB3 is also aberrantly methylated in BWS_IC2.
(a) Methylation index (y-axis) of HM450K CpG probes (x-axis) of BWS_IC2 patients (orange) and normal individuals (blue) at KCNAB3 promoter. Vertical black lines represent DMR boundaries detected in BWS_UPD11. (b) Frequency of BWS_IC2 patients with hypomethylation at KCNAB3 promoter.

4.4 Discussion

4.4.1 Preprocessing

Data preprocessing was performed with the R package RnBeads, a specialist bioinformatic pipeline designed to handle HM450K data. Technical and biological biases were efficiently removed from data using the R package ComBat although the algorithm started to show its limitations whilst dealing with confounding covariates such as the age in our dataset. The use of younger normal control individuals and a better random allocation of the samples to HM450K slides and HM450K positions (i.e. Sentrix_ID and Sentrix_Position respectively) to minimise confounding covariates would have greatly helped to reduce the bias associated with
those variables. Additionally, the ComBat algorithm would have performed better for bias correction. However, the methylation data used in this study was generated by multiple groups (i.e. our group and collaborators) and therefore it was hardly possible to elaborate a better experimental design that took into account all of those variables.

4.4.2 Methodology

My epimutation methodology was validated by assessing the methylation level at \textit{H19/IGF2}:IG-DMR and at \textit{KCNQ1OT1}:TSS-DMR in all patients. By comparison with clinical diagnostic testing results, my method detected the correct epimutation pattern in all individuals except for one BWS_UPD11 patient for which it failed to detect \textit{H19/IGF2}:IG-DMR GOM. We hypothesised that mosaicism could explain the epimutation detection failure observed in that patient. Indeed, mosaicism is commonly associated with BWS_UPD11 patients (Slatter et al. 1994; Weksberg et al. 2010) and, hence, our method may not be sensitive enough to detect lower-level disomy in that group of patient. Selecting a two standard deviations threshold may have improved epimutation sensitivity detection for low-level disomy but this may have been at the cost of increasing false positive epimutation rate. In conclusion, our method performed well for the methylation profiling of BWS and SRS individuals at \textit{H19/IGF2}:IG-DMR and at \textit{KCNQ1OT1}:TSS-DMR and, therefore, it can be applied to other target regions such as other imprinted DMRs. However, additional care need to be taken when dealing with patients in which low-level mosaicism is expected.
4.4.3 Polymorphic epimutation

The epimutation rate at each targeted imprinted DMR was evaluated in both patients and normal individuals. Comparison of both groups using Fisher's exact test followed by FDR adjustment showed that more than half of imprinted DMRs with abnormal methylation findings were in fact polymorphic for GOM or LOM only or for both. This confirmed previous published data where the imprinted DMRs \textit{IGF2R}:Int2-DMR was found polymorphic for GOM but not LOM (Tee et al. 2013).

4.4.4 Genomic imprinting methylation in SRS_UPD7

In SRS_UPD7, MLID was detected in 1 out of 6 (17%) patient. The individual had GOM at \textit{ZDBF2/GPR1}:IG-DMR and LOM at \textit{IGF1R}:Int2-DMR. Previous published data has shown hypermethylation at \textit{RB1}, \textit{ANKRD11} and \textit{MCTS2} (Prickett et al. 2015) in SRS_UPD7 patients, but none of these loci were found to be epimutated in our cohort. In our study, \textit{RB1} and \textit{MCTS2} were found polymorphic for both GOM and LOM (i.e. high frequency of GOM and LOM is observed in the normal control population) and \textit{ANKRD11} has not been evaluated.

4.4.5 Genomic imprinting methylation in SRS_IC1

In SRS_IC1, MLID was detected in 13% (2 out of 15). This is slightly above the published common consensus of 7-10% (9 out of 97 in combined previous reports; \(p=0.64\)) (Azzi et al. 2009; Turner et al. 2010). Of the two individuals with MLID, one had additional LOM at \textit{L3MBTL1}:alt-TSS-DMR and, interestingly, the other one had
additional GOM at \textit{GRB10}:alt-TSS-DMR, \textit{PEG10}:TSS-DMR, \textit{MEST}:alt-TSS-DMR and \textit{HTR5A}:TSS-DMR. This observation was very intriguing as GOM at chromosome 7 is normally associated with SRS\_UPD7 and not SRS\_IC1. Additionally, methylation disturbances at these loci, if any, are generally LOM according to previous published data (Azzi et al. 2009; Turner et al. 2010; Court et al. 2013). Accordingly, this patient sample was suspected to have been mislabelled as SRS\_IC1 instead of SRS\_UPD7. Taking the latter patient out of that group, MLID was detected in approximatively 7% (1 of 14) of SRS\_IC1.

4.4.6 Genomic imprinting methylation in BWS\_UPD11

Of the 8 BWS\_UPD11 patient, 1 (12.5\%) had MLID. The patient has GOM at \textit{PEG10}:TSS-DMR. To the best of my knowledge, this is the first time that an additional epimutation affecting another chromosome than chromosome 11 has been identified in BWS\_UPD11 cohort. However, the clinical relevance of the findings is yet to be described.

4.4.7 Genomic imprinting methylation in BWS\_IC1

MLID was not detected in any individuals although a recent report suggested that MLID can occurred in up to 30\% (3 in 10) of BWS\_IC1 patient (Maeda et al. 2014). In the latter, the germline DMR \textit{INPP5Fv2} was found hypomethylated in two patients and the somatic DMR \textit{NESP} was found hypermethylated in a third patient. In our cohort, only a few patients (n=4) were methylation profiled and, hence, could
explained why no MLID was detected.

4.4.8 Genomic imprinting methylation in BWS_IC2

MLID was detected in 31 out of 78 (40%) of patients, which was significantly higher \((p=0.0051)\) than the common consensus of approximatively 24\% (all previous studies combined: 113 of 473) (Rossignol et al. 2006; Bliek et al. 2008; Azzi et al. 2009; Court et al. 2013; Poole et al. 2013; Tee et al. 2013; Maeda et al. 2014). 

However, with the increasing use of modern technics such as Illumina methylation array, imprinted DMRs that were not assessed previously are also shown to be epimutated in BWS. Consequently, the previously reported consensus is likely to underestimate the frequency of MLID in BWS with IC2 LOM as some patients may have additional epimutations at previously unscreened imprinted DMRs but no abnormalities at previously screened imprinted DMRs. To support this hypothesis, the authors of two recent studies used high-throughput methods for the methylation profiling of more than 20 imprinted DMRs and both authors reported MLID in 14 out of 43 (33\%) and in 15 out of 44 (34.1\%) (Court et al. 2013; Maeda et al. 2014). Taken together with my study, it seems that the methylation profiling of more imprinted DMRs is associated with and increased frequency of MLID discovery through the identification of epimutated DMRs that were not screened in the past. Additionally, it confirmed that MLID is likely to be more frequent than originally thought.

There was considerable variance in the number of loci affected, ranging from one to
12 additional imprinted DMRs and the majority of individuals had less than two disturbed loci. This showed that BWS_IC2 cases with extreme abnormal epigenotype are rare but still exist. It remained to evaluate the consequences of such extreme epigenotype and see if those rare patients have a more severe form of BWS or an unusual phenotype.

Both paternal and maternal DMRs were affected with GOM associated with paternal DMRs and LOM associated with maternal DMRs. No epimutation preference toward maternal or paternal DMRs was found (p=0.106, Fisher’s exact test) which contrasts with a previous report where maternal DMRs have been found significantly more epimutated than paternal DMRs (p=0.042, Fisher’s exact test) (Maeda et al. 2014).

The somatic locus ZDBF2/GPR1:IG-DMR was shown to gain methylation in 5 (16%) of our patient and it was the only interrogated imprinted DMRs that was hypermethylated. This acquisition was believed to being due to a concomitant LOM in the nearby maternally methylated GPR1-AS:TSS-DM, which is known to regulate the methylation of the somatic DMR ZDBF2/GPR1:IG-DMR in a hierarchical fashion (H. Kobayashi et al. 2013). In my study, methylation profiling at GPR1-AS:TSS-DMR did not show LOM but this was expected as GPR1-AS:TSS-DMR acquires immediate biparental methylation following implantation (H. Kobayashi et al. 2013).

Finally, no distinct epimutation correlations were found between ‘unrelated DMRs’. Imprinted DMRs belonging to the same imprinting cluster tend to be epimutated
simultaneously but this was expected.

4.4.9 DMR hunting

DMR hunting revealed 3 new potential candidate regions (SLC12A9, LTA and KCNAB3) that may be involved in BWS_IC2 and BWS_UPD11 aetiology.

SLC12A9 belongs to the solute carrier 12 (SLC12) gene family. This gene family encodes electroneutral inorganic cation-chloride cotransporters (CCCs) that are plasma membrane proteins mediating the movement of inorganic sodium (Na⁺) and/or potassium (K⁺) cations, tightly coupled to the movement of chloride (Cl⁻) anions. Mutations and or dysfunctions in members of this gene family have been associated with pathophysiological disorders such as Bartter syndrome, Gitleman syndrome and Andermann syndromes (Gagnon and Delpire 2013). However, little is known about SLC12A9 for which no protein function has yet been described and relevance to human disease is not known.

LTA gene encodes a cytokine produced by lymphocytes. The protein, belonging to the tumour necrosis factor family, is highly inducible, secreted, and forms heterotrimers with lymphotoxin-beta which anchor lymphotoxin-alpha to the cell surface. This protein also mediates a large variety of inflammatory, immunostimulatory, and antiviral responses, is involved in the formation of secondary lymphoid organs during development and plays a role in apoptosis.
Genetic variations in this gene are associated with susceptibility to leprosy type 4, myocardial infarction, non-Hodgkin's lymphoma, and psoriatic arthritis. However, LTA clinical relevance in BWS is currently not known.

*KCNAB3* gene encodes a member of the potassium channel, voltage-gated, shaker-related subfamily. The encoded protein is one of the beta subunits, which are auxiliary proteins associating with functional Kv-alpha subunits. The encoded protein forms a heterodimer with the potassium voltage-gated channel, shaker-related subfamily, member 5 (*KCNA5*) gene product and regulates the activity of the alpha subunit (Leicher et al. 1998). Interestingly, recent published papers reported complete or partial LOM at *KCNAB3* in MLID patients with, in some cases, mutation in NLR family pyrin domain containing 7 (*NLRP7*) gene and in ZFP57 zinc finger protein (*ZFP57*) gene (Beygo, Ammerpohl, et al. 2013; Rezwan et al. 2015; Bak et al. 2016).
Chapter 5

Investigations of the underlying causes of MLID in BWS_IC2
5.1 Introduction

5.1.1 Assisted reproductive technologies and MLID

Assisted reproductive technology (ART) is a generic term given to describe a number of different treatments, including ovarian stimulation, in vitro fertilisation (IVF) and intra-cytoplasmic sperm injection (ICSI), that can be performed to help achieve pregnancy. ART is used primarily for infertility treatment but can also be used for fertile couples to reduce the risk of potential diseases (e.g. when combined with preimplantation diagnosis). The potentially increased frequency of deregulated genomic imprinting in children conceived with the help of ART has raised concerns about how the procedures may predispose the embryos to acquire imprinting disturbances and in fine to diseases.

5.1.1.1 Ovarian stimulation

Ovarian stimulation (OS) is the induction of ovulation by the injection of a fertility hormone (e.g. clomiphene, gonadotropin). It is usually used to stimulate the development of ovarian follicles to reverse anovulation (where the ovaries do not release an oocyte) or oligoovulation (where ovulation is infrequent or irregular). To achieve ovulation and pregnancy, anovulatory and oligoovulatory patients are given enough hormones to reach a threshold sufficient to initiate growth and development of a number of follicles but ideally only maintain the growth of one follicle or certainly not more than three. When more than one egg is produced, the risk of multiple pregnancy increases.
OS can also be used for controlled ovarian hyperstimulation (COH). In contrast to OS in anovulatory and oligoovulatory patients in which the aim is to produce a monofollicular ovulation, COH aims to produce multiple follicular development in order to harvest a suitable number of oocytes. Patients that undergo COH are given a larger dose of hormones than anovulatory and oligoovulatory patients to treat their conditions. COH is used in conjunction with ART procedures such as IVF or ICSI.

In both OS and COH procedures, a trigger shot of hormones (e.g. human chorionic gonadotropin) can also be administered to patients to induce final maturation of oocytes and/or trigger oocytes release. This would allow scheduling sexual intercourse or intrauterine insemination and thus increasing chances of pregnancy in anovulatory and oligoovulatory patients. In the case of IVF or ICSI, this would avail the retrieval of fully mature eggs.

Ovarian stimulation is considered a safe procedure but complication may arise in some cases. One of the main risk associated with OS and COH is that patients may develop an ovarian hyperstimulation syndrome (OHSS), which can occur when too much hormone medication is administered. Symptoms of OHSS range from mild abdominal distention due to swollen and painful ovaries alone or with an accompanying fluid shift into the abdomen, to renal failure and death as a result of haemoconcentration and reduced perfusion of organs such as the kidneys, heart and brain (Fiedler and Ezcurra 2012). There is also a risk of multiple pregnancy when an excess of follicles are produced.
5.1.1.2 In vitro fertilisation

IVF treatment involves the fertilisation of an egg (or eggs) outside the body. The treatment can be performed using an individual's own eggs and sperm, or using either donated sperm or donated eggs, or both. IVF is an ART procedure that is recommended in cases where there is underlying unexplained infertility problems, the fallopian tubes are blocked, the male partner has non severe fertility problems, or the embryos need to be tested to avoid passing on a genetic condition. The protocol used for IVF is not fully consistent between places, they can differ between birth centres and countries for example, but the concept remains the same. The IVF procedure starts with COH (see section 5.1.1.1 - Ovarian stimulation). The stimulated and matured eggs are then collected from ovaries and prepared for fertilisation. In parallel, sperm is washed to remove inactive cells and seminal fluid. Prepared eggs and sperm are incubated together in a culture media to achieve fertilisation. The embryos (fertilised eggs) are passed into growth media and cultured until cleavage stage or blastocyst stage. In some cases, preimplantation genetic diagnosis can be performed and one or two cells are removed from the developing embryo and tested for a specific genetic disease. Finally, selected embryos are transferred into the uterus. Non transferred embryos are stored by cryopreservation and may be used in subsequent cycles if pregnancy is not achieved after the first attempt.

5.1.1.3 Intra-cytoplasmic sperm injection

ICSI is most commonly used to overcome male fertility problems. The procedure is
very similar to IVF and only differs in the techniques used to fertilise the eggs. With ICSI, a matured and healthy single sperm is first selected and then directly injected into an egg. The embryos are cultured and transferred into the uterus in a similar manner to IVF.

5.1.1.4 Evidence of ART disrupting genomic imprints

Evidence to suggest that ART may disrupt normal growth and development through epigenetic alterations of genomic imprinting was first seen in animal studies. In large domestic animals, such as cattle and sheep, a considerable number of abnormally large animals were born subsequent to various embryo manipulations before the blastocyst stage. The resulting aberrant overgrowth phenotype seen in the animals was very similar to the human disorder BWS and it was referred to as the large offspring syndrome (LOS) (Young et al. 1998). At the time, it was hypothesised that LOS may be caused by an epigenetic deregulation that would impair gene expression, especially imprinted gene expression, and this could be a consequence of the embryo manipulations. The investigation of sheep foetuses recovered at day 125 of gestation (term=147 days) and conceived by ART revealed that 12 out of 48 foetuses had LOS (compared to none in naturally conceived foetuses, n=22). Interestingly, the aberrant phenotype observed in the 12 foetuses was associated with a significant reduction in the expression of the imprinted gene IGF2R. In 9 of the 12 foetuses, the reduced IGF2R gene expression was further explained by a loss-of-methylation of IGF2R-DMR2, the imprinting control region regulating IGF2R expression (Young et al. 2001). In mice, IGF2R was shown to degrade the excess
IGF2 (Ludwig et al. 1995) and mouse embryos lacking Igf2r were shown to be larger than normal at day 18.5 of gestation and at birth (Ludwig et al. 1996). More recently, the investigators of another study showed that in addition to the increased of bodyweight, LOS hybrid bovine foetuses conceived with ART have also macrosomia, macroglossia, and umbilical hernia, which are primary characteristics of BWS. More importantly, besides the phenotypic similarities the foetuses with LOS and conceived with ART also had epigenetic similarities with BWS. Indeed the investigators found that in these foetuses KCNQ1OT1, which is the main locus involved in BWS molecular aetiology (~ 50%), was biallelically expressed and this was due to a loss-of-methylation on the maternal allele of KCNQ1OT1:TSS-DMR (Z. Chen et al. 2013).

5.1.1.5 Evidence of association between ART and imprinting disorders

In humans, an association between ICSI and Angelman syndrome (AS) due to an imprinting defect was suggested following the report of three children who were conceived in such manner and subsequently developed the syndrome. Methylation profiling revealed complete or partial LOM at the imprinted DMR SNRPN:alt-TSS-DMR (Cox et al. 2002; Ørstavik et al. 2003).

A similar association between ART and BWS has also been reported. In the general population, the prevalence of ART is estimated between 0.8 and 1.3% whilst the prevalence of ART in BWS population is estimated between 4 and 4.6% (DeBaun et al. 2003; Gicquel et al. 2003; Maher et al. 2003). The investigators of a case-control study performed in Australia reported a prevalence of ART in BWS group as high as
10.81% whilst in the general population the ART prevalence was of 0.67% (Halliday et al. 2004). Interestingly, across these studies 95% of BWS patients conceived with ART were due to the LOM at $KCNQ1OT1$:TSS-DMR whilst in BWS patients conceived naturally the imprinting disturbance accounts for approximatively 50% of cases (DeBaun et al. 2003; Gicquel et al. 2003; Maher et al. 2003; Halliday et al. 2004; Weksberg et al. 2010). Taken together, these reports showed clear evidence that ART is associated with BWS due to imprinting defects at $KCNQ1OT1$:TSS-DMR.

5.1.1.6 Investigation of potential ART association with MLID in BWS

The study of several BWS with LOM at $KCNQ1OT1$:TSS-DMR (BWS_IC2) cohorts have identified a significant associations between MLID in BWS_IC2 and the use of ART. A first study reported that 3 out of 8 BWS_IC2 patients conceived with ART had MLID compared to 3 out of 47 for naturally conceived patients ($p=0.034$) (Lim et al. 2009). In another report, MLID was detected in one BWS_IC2 patient conceived with the help of ART (ICSI) and in one out of five BWS_IC2 patients conceived naturally. Although only one ART patient was available, the authors suggested a possible association between BWS_IC2 and the widespread methylation disturbances (Hiura et al. 2012). Finally, the association was further demonstrated in a third report in which MLID was reported in 7 out of 14 BWS_IC2 ART patients compared to 26 out of 173 of patients conceived naturally ($p=0.0033$) (Tee et al. 2013). However, such associations are not consistent between publications and other research studies have reported there was no significant association between the two. A first study suggesting there was no association reported MLID in 3 out of 11 BWS_IC2 patients.
1 conceived with ART and in 7 out of 29 patients conceived normally (p=1) (Rossignol et al. 2006). Furthermore, the investigators of another study reported MLID in BWS_IC2 in 4 out of 12 patients conceived with ART and in 12 out of 55 patients conceived naturally and they concluded that ART was not responsible for MLID (p=0.46) (Azzi et al. 2009). It remains unclear what is the impact of ART on genomic imprinting methylation but the study of more extensive patient cohort will help to determine whether or not ART predispose embryos to acquire imprinting errors and diseases.

5.1.2 Trans-imprinting defect and MLID

The genetic alterations of genes required for establishing and maintaining methylation, or shown to be essential for embryonic development may lead to the disruption of genomic imprinting methylation through trans-mechanism defects. Furthermore, several pieces of evidence have demonstrated that a subset of imprinted genes belongs to a co-ordinately regulated network. Within the same imprinted gene network, imprinted genes can modify in trans the expression of other imprinted genes (Varrault et al. 2006; Patten et al. 2016; Soellner et al. 2016). As an example, experiments in mouse brain neuroblastoma cell lines and meta-analysis of mouse microarray data sets have shown that the imprinted transcription factor Plagl1 promotes the expression of several genes, including the imprinted genes Igf2, H19, Kcnq1ot1, Cdkn1c and Dlk1 (Varrault et al. 2006). Therefore, the existence of such network may suggest that the genetic insult of a member of the network may first
result in a cis-defect, and subsequently to further downstream trans-imprinting failures. However, no evidences reporting that this mechanism may acts at the imprinting methylation level were found yet.

To date, a few pathological mutations affecting maternal-effect genes (\textit{NLRP7}, \textit{KHDC3L} (previously known as \textit{C6orf221}), \textit{NLRP2} and \textit{NLRP5}) and maternal-zygotic effect gene (\textit{ZFP57}) were identified in syndromes with MLID.

### 5.1.2.1 Maternal-effect and maternal-zygotic effect genes

Maternal-effect genes are a special class of genes that are required for normal embryonic development. The RNA or protein products of those genes are supplied by the mother and are produced or deposited in the oocyte or are present in the fertilised egg or embryo before the expression of zygotic genes is initiated. In contrast, genes whose RNA and protein products are only produced after zygotic genome activation, at or after the maternal to zygotic transition stage, are zygotic genes (Marlow 2010). The maternal to zygotic transition starts during the early two-cell stages in the mouse (L. Li et al. 2010) (Figure 5.1) and by the 4- to 8-cell stages in the human (Schultz 2002).

To date, the gene knockout and gene knockdown in mammalian animal models (mice) and the study of rare human disorders have helped to identify approximatively 60 maternal-effect genes for which the zygotic transcription has not rescued the embryonic lethality due to the loss of maternal gene function (Table 5.11) (Condic...
Figure 5.1: Maternal to zygotic gene transition in the mouse. Maternal RNA and proteins accumulate during oocytes maturation and will slowly degrade starting from the ovulation stage. In contrast, the activation of the embryonic genome will start from the 1/2-cell stage (E1.5) and will remain active thereafter. (Adapted from L. Li et al. 2010).

5.1.2.1.1 Zygotic effect

Recessive mutations disrupting the zygotic gene function will be visible in embryos homozygous for the mutant genotype. The abnormal embryo will inherit one copy of the mutated gene from its father and one copy of the mutated gene from its mother. Both parents will be heterozygous for the pathogenic mutation. Embryos with a heterozygous mutant genotype will appear normal.

5.1.2.1.2 Maternal-effect

Recessive mutations disrupting the maternal gene function will be visible in embryos for which the mother is homozygous for the mutant genotype (X. Li 2010; Marlow
The embryo genotype, either homozygous or heterozygous for the mutant allele, has no effect on the phenotype.

5.1.2.1.3 Maternal-zygotic effect

Recessive mutations disrupting maternal-zygotic gene function will be visible in embryos for which its genotype and its mother genotype are both homozygous for the mutant allele (X. Li 2010; Marlow 2010). If one of the embryo or the mother is heterozygous for the mutant genotype, the embryo may display a reduced or no signs of the abnormal phenotype.
Table 5.11: List of mammalian maternal-effect genes.

(Maternal-effect was mainly described in mice (Condic 2016)).

<table>
<thead>
<tr>
<th>GENE</th>
<th>DESCRIPTION</th>
<th>ROLE IN DEVELOPMENT</th>
<th>LETHALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell cycle regulation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fmn2</td>
<td>Formin-2</td>
<td>Control of mitotic spindle</td>
<td>embryo</td>
</tr>
<tr>
<td>Kpna6</td>
<td>Importin alpha7; karyopherin alpha 6</td>
<td>Nuclear import, spindle assembly</td>
<td>2-cell</td>
</tr>
<tr>
<td>MiR181A1</td>
<td>microRNA 181a-1 (human)</td>
<td>Negative regulator of NPM2</td>
<td>early embryo</td>
</tr>
<tr>
<td>Npm2</td>
<td>Nucleoplaemin2</td>
<td>Nuclear end nucleolar organization, chromatin remodelling</td>
<td>2-cell</td>
</tr>
<tr>
<td>Tcd1</td>
<td>T cell lymphoma breakpoint 1</td>
<td>Cell division, transcriptional regulation</td>
<td>morula</td>
</tr>
<tr>
<td>TP73</td>
<td>TAp73; Tumor protein 73</td>
<td>Cell division and ploidy</td>
<td>embryo</td>
</tr>
<tr>
<td>Zar1</td>
<td>zygote arrest 1</td>
<td>RNA processing, pronuclear fusion</td>
<td>2-cell</td>
</tr>
<tr>
<td></td>
<td><strong>Epigenetic reprogramming/pluripotency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ago2</td>
<td>argonaute 2, RISC catalytic component</td>
<td>RNA stability, degradation of maternal factors, zygotic gene activation</td>
<td>2-cell</td>
</tr>
<tr>
<td>Atg5</td>
<td>autophagy related 5</td>
<td>Degradation of maternal factors</td>
<td>4-8 cell</td>
</tr>
<tr>
<td>Dicer1</td>
<td>dicer 1, ribonuclease type III</td>
<td>Degradation of maternal factors</td>
<td>embryo</td>
</tr>
<tr>
<td>Dnmt1</td>
<td>DNA methyltransferase 1 associated protein 1</td>
<td>DNA methylation, epigenetic modification</td>
<td>embryo</td>
</tr>
<tr>
<td>H1fh2a</td>
<td>DNA methyltransferase (cytosine-5) 1</td>
<td>DNA methylation, epigenetic modification</td>
<td>embryo</td>
</tr>
<tr>
<td>H1fh2b</td>
<td>TH2A; histone cluster 1, H2aa</td>
<td>Genomic reprogramming, paternal genome activation, pluripotency</td>
<td>embryo</td>
</tr>
<tr>
<td>H1fh3f</td>
<td>TH2B; histone cluster 1, H2ba</td>
<td>Genomic reprogramming, paternal genome activation, pluripotency</td>
<td>embryo</td>
</tr>
<tr>
<td>H1fh4f</td>
<td>Histone H3; histone cluster 1, H3f</td>
<td>Chromatin structure, epigenetic modification</td>
<td>2-cell</td>
</tr>
<tr>
<td>H1fh4f</td>
<td>Histone H4; histone cluster 1, H4f</td>
<td>Chromatin structure, epigenetic modification</td>
<td>2-cell</td>
</tr>
<tr>
<td>Kdm1B</td>
<td>lysine (K)-specific demethylase 1B</td>
<td>DNA methylation, epigenetic modification</td>
<td>mid gestation</td>
</tr>
<tr>
<td>Pou5F1</td>
<td>Oct4; POU class 5 homeobox 1</td>
<td>Pluripotency determinant, transcription factor, regulation of PGC7</td>
<td>2-cell</td>
</tr>
<tr>
<td>Ring1</td>
<td>ring finger protein 1</td>
<td>Chromatin structure, silencing differentiation genes</td>
<td>2-cell</td>
</tr>
<tr>
<td>Rnf2</td>
<td>ring finger protein 2</td>
<td>Chromatin structure, silencing differentiation genes</td>
<td>2-cell</td>
</tr>
<tr>
<td>Slbp</td>
<td>stem-loop binding protein</td>
<td>Histone gene regulation</td>
<td>2-cell</td>
</tr>
<tr>
<td>Sox2</td>
<td>SRY (sex determining region Y)-box 2</td>
<td>Pluripotency determinant, transcription factor</td>
<td>early embryo</td>
</tr>
<tr>
<td>Tet3</td>
<td>tet methylcytosine dioxygenase 3</td>
<td>Paternal genome demethylation</td>
<td>early embryo</td>
</tr>
<tr>
<td>Trim24</td>
<td>Tif1A; tripartite motif-containing 24</td>
<td>Chromatin structure, epigenetic modification</td>
<td>2-4 cell</td>
</tr>
<tr>
<td>Trim28</td>
<td>Tif1b; tripartite motif-containing 28</td>
<td>Epigenetic modification, interacts with Zfp57</td>
<td>early embryo</td>
</tr>
<tr>
<td>Xist</td>
<td>inactive X specific transcripts</td>
<td>Epigenetic modification, X-inactivation</td>
<td>embryo</td>
</tr>
<tr>
<td>Dnmt3a</td>
<td>DNA methyltransferase 3A</td>
<td>DNA methylation, genomic imprinting</td>
<td>embryo</td>
</tr>
<tr>
<td>Dnmt3l</td>
<td>DNA (cytosine-5)-methyltransferase 3-like</td>
<td>DNA methylation, genomic imprinting</td>
<td>mid gestation</td>
</tr>
<tr>
<td>Dppa3</td>
<td>Pgc7/Stella; developmental pluripotency-associated 3</td>
<td>DNA methylation, genomic imprinting</td>
<td>2-cell</td>
</tr>
</tbody>
</table>
Table 5.11: List of mammalian maternal-effect genes (continued).

(Maternal-effect was mainly described in mice (Condic 2016)).

<table>
<thead>
<tr>
<th>GENE</th>
<th>DESCRIPTION</th>
<th>ROLE IN DEVELOPMENT</th>
<th>LETHALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Esr2</td>
<td>estrogen receptor 2 (beta)</td>
<td>Genomic imprinting, endocrine response</td>
<td>embryo</td>
</tr>
<tr>
<td>Gja4</td>
<td>gap junction protein, alpha 4</td>
<td>Genomic imprinting, gap-junctional communication</td>
<td>embryo</td>
</tr>
<tr>
<td>KHDC3L</td>
<td>ECAT1/C8orf221; KH domain containing 3-like, subcortical maternal complex member (human)</td>
<td>Genomic imprinting, associates with NLRP7</td>
<td>zygote</td>
</tr>
<tr>
<td>Nlrp2</td>
<td>NLR family, pyrin domain containing 2</td>
<td>DNA methylation, genomic imprinting</td>
<td>2-cell</td>
</tr>
<tr>
<td>NLRP7</td>
<td>NLR family, pyrin domain containing</td>
<td>DNA methylation, genomic imprinting</td>
<td>zygote</td>
</tr>
<tr>
<td></td>
<td><strong>Oocyte cytoplasmic lattice</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khdc3</td>
<td>FILIA; KH domain containing 3, subcortical maternal complex member</td>
<td>Chromosome stability, binds NLRP5, oocyte</td>
<td>embryo</td>
</tr>
<tr>
<td>Nlrp5</td>
<td>MATER; NLR family, pyrin domain containing 5</td>
<td>Oocyte cytoplasmic lattice assembly, mitochondrial activation</td>
<td>2-cell</td>
</tr>
<tr>
<td>Ooep</td>
<td>Flapped; oocyte expressed protein</td>
<td>Oocyte cytoplasmic lattice assembly</td>
<td>2-cell</td>
</tr>
<tr>
<td>Padi6</td>
<td>peptidyl arginine deiminase, type VI</td>
<td>Oocyte cytoplasmic lattice assembly</td>
<td>2-cell</td>
</tr>
<tr>
<td>Tle6</td>
<td>transducin-like enhancer of split 6</td>
<td>Oocyte cytoplasmic lattice assembly</td>
<td>2-cell</td>
</tr>
<tr>
<td></td>
<td><strong>Transcriptional regulation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bnc1</td>
<td>basonulin 1</td>
<td>Transcription factor</td>
<td>2-cell</td>
</tr>
<tr>
<td>Brwd1</td>
<td>bromodomain and WD repeat domain containing 1</td>
<td>Zygotic gene activation, chromatin remodeling</td>
<td>2-cell</td>
</tr>
<tr>
<td>Ctfc</td>
<td>CCTC-binding factor</td>
<td>Transcription factor; Epigenetic modification</td>
<td>early embryo</td>
</tr>
<tr>
<td>Figla</td>
<td>folliculogenesis specific basic helix-loop-helix</td>
<td>Transcription factor; Genomic imprinting</td>
<td>embryo</td>
</tr>
<tr>
<td>Hira</td>
<td>histone cell cycle regulator</td>
<td>DNA transcription and replication</td>
<td>zygote</td>
</tr>
<tr>
<td>Hsf1</td>
<td>heat shock factor 1</td>
<td>Transcription factor</td>
<td>zygote</td>
</tr>
<tr>
<td>Kmt2d</td>
<td>MI22: lysine (K)-specific methyltransferase 2D</td>
<td>Zygotic gene activation, DNA methylation</td>
<td>2-cell</td>
</tr>
<tr>
<td>MIR196a-1</td>
<td>microRNA 196-1 (human)</td>
<td>Negative regulator of NOBOX</td>
<td>embryo</td>
</tr>
<tr>
<td>MIR212</td>
<td>microRNA 212 (human)</td>
<td>Regulates FIGLA stability</td>
<td>embryo</td>
</tr>
<tr>
<td>Nobox</td>
<td>NOBOX oogenesis homeobox</td>
<td>Transcription factor</td>
<td>embryo</td>
</tr>
<tr>
<td>Pdk1</td>
<td>pyruvate dehydrogenase kinase, isoenzyme 1</td>
<td>Zygotic gene activation</td>
<td>2-cell</td>
</tr>
<tr>
<td>Pik3r1</td>
<td>phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 1 (p85 alpha)</td>
<td>Zygotic gene activation</td>
<td>2-cell</td>
</tr>
<tr>
<td>Sebox</td>
<td>SEBOX homeobox</td>
<td>Transcription factor</td>
<td>2-cell</td>
</tr>
<tr>
<td>Smarca4</td>
<td>Brg1; SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4</td>
<td>Zygotic gene activation</td>
<td>2-cell</td>
</tr>
<tr>
<td>Zfp36i2</td>
<td>zinc finger protein 36, C3H type-like</td>
<td>Zinc-finger protein; Degradation of</td>
<td>2-cell</td>
</tr>
<tr>
<td>Zfp57</td>
<td>zinc finger protein 57</td>
<td>Zinc-finger protein; Genomic imprinting</td>
<td>embryo</td>
</tr>
</tbody>
</table>
### Table 5.11: List of mammalian maternal-effect genes (continued).

(Maternal-effect was mainly described in mice (Condic 2016)).

<table>
<thead>
<tr>
<th>GENE</th>
<th>DESCRIPTION</th>
<th>ROLE IN DEVELOPMENT</th>
<th>LETHALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bcas2</td>
<td>breast cancer amplified sequence 2</td>
<td>DNA repair/RPA complex</td>
<td>2–4 cell</td>
</tr>
<tr>
<td>Pms2</td>
<td>postmeiotic segregation increased 2</td>
<td>Ubiquitin ligase; X-inactivation</td>
<td>early embryo</td>
</tr>
<tr>
<td>Rilim</td>
<td>Rnf12; ring finger protein, LIM domain interacting</td>
<td>Ubiquitin-conjugating enzyme; Deubiquitinating enzyme</td>
<td>embryo</td>
</tr>
<tr>
<td>Ube2a</td>
<td>Hr6a; ubiquitin-conjugating enzyme</td>
<td>Ubiquitin-conjugating enzyme; Deubiquitinating enzyme</td>
<td>2-cell</td>
</tr>
<tr>
<td>Uchl1</td>
<td>ubiquitin carboxy-terminal hydrolase L1</td>
<td></td>
<td>morula</td>
</tr>
</tbody>
</table>

**Undetermined embryonic functions**

<table>
<thead>
<tr>
<th>GENE</th>
<th>DESCRIPTION</th>
<th>ROLE IN DEVELOPMENT</th>
<th>LETHALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCAR4</td>
<td>breast cancer anti-estrogen resistance 4 (human)</td>
<td>Breast cancer anti-estrogen resistance factor</td>
<td>morula</td>
</tr>
<tr>
<td>Gas6</td>
<td>growth arrest specific 6</td>
<td>Cytokine, Axl ligand</td>
<td>zygote</td>
</tr>
<tr>
<td>Gclm</td>
<td>glutamate-cysteine ligase, modifier subunit</td>
<td>Glutathione synthesis</td>
<td>blastocyst</td>
</tr>
</tbody>
</table>

### 5.1.2.2 NLRP7

Familial biparental hydatidiform moles (FBHM) (OMIM 231090) are aberrant human pregnancies in which there is nonexistent or abnormal embryonic development, excessive trophoblastic proliferation, and cystic degeneration of chorionic villi into grape-like structures (Rezaei et al. 2016). Unlike complete hydatidiform moles which are mostly sporadic and wholly androgenic, women with FBHM suffer from recurrent molar pregnancies and the molar tissues are not androgenetic but show a normal pattern of bi-parental diploid inheritance (Van den Veyver and Al-Hussaini 2006). FBHM is associated with global failure in the establishment and or the maintenance of methylation marks at maternal imprinting centres during oocyte growth or post-zygotic development, respectively. The genome-wide methylation analysis of molar
tissues showed widespread hypomethylation at maternal DMRs, including placental-
specific DMRs (Court et al. 2014), whilst the paternal DMRs (e.g. \textit{H19/IGF2}:IG-DMR
and \textit{MEG3/DLK1}:IG-DMR) and non-imprinted regions are unaffected (Judson et al.
2002; El-Maarri 2003; Djuric et al. 2006; Kou et al. 2008; Hayward et al. 2009;
Sanchez-Delgado et al. 2015). The genetic investigations of FBHM identified
recessive mutations in the maternal-effect gene \textit{NLRP7} in approximatively 70% of
cases. Consequently, it was suggested that the \textit{NLRP7} maternal loss of function was
responsible for FBHM phenotype (Murdoch et al. 2006; Kou et al. 2008; Hayward et
al. 2009). To date, 59 pathogenic recessive mutations in \textit{NLRP7} have been
described in the aetiology of FBHM (Reddy et al. 2016).

\textit{NLRP7} (NLR family pyrin domain containing 7) is a maternal-effect gene for which
the transcript product has been documented in a large number of human tissues
including liver, lung, placenta, spleen, thymus, peripheral blood leukocytes, testis
and ovaries (Slim and Wallace 2013). \textit{NLRP7} transcripts are found in all oocyte and
preimplantation embryo stages (Murdoch et al. 2006). Its transcript concentration
decreases during oocyte maturation until the morula stage where it reaches its
lowest level. The transcript level then increases rapidly from day 3 to day 5, which
corresponds to the blastocyst stage and the activation of the embryonic genome
(Murdoch et al. 2006; P. Zhang et al. 2008; Slim and Wallace 2013). \textit{NLRP7} gene
product was also found to be involved in cell proliferation, although its precise role
has not been fully described yet (Okada et al. 2004; Khare et al. 2012). Some studies
have suggested that \textit{NLRP7} silencing induces growth inhibition (Okada et al. 2004)
whilst others have suggested the opposite (Khare et al. 2012). NLRP7 is also known to be a negative feedback regulator of the interleukin 1 beta (IL-1B) cytokine and consequently plays a role in the inflammatory pathway (Kinoshita et al. 2005; Messaed et al. 2011). The inflammasome is an important mechanism during pre-implantation of the embryo. It facilitates the implantation of the blastocyst, regulates the protease network and controls the extent to which the trophoblast may invade the maternal endometrium (Karmakar and Das 2002; Strakova et al. 2002; Murdoch et al. 2006). Finally, pathogenic mutations in NLRP7 may also be associated with the global maternal genomic imprinting alteration in FBHM, hence suggesting that NLRP7 may have a role in regulating genomic imprinting methylation in an unrecognised manner.

5.1.2.3 ZFP57

Transient neonatal diabetes mellitus type 1 (TNDM1) is a form of diabetes mellitus. TNDM1 is caused by the over-expression of the PLAGL1 imprinted gene (paternally expressed) located at chromosome 6q24 (Gardner et al. 2000; Kamiya et al. 2000). The loss-of-methylation of the imprinting centre PLAGL1:alt-TSS-DMR, which is normally maternally methylated and regulates PLAGL1 gene expression, accounts for approximatively 20% of TNDM1 cases. Among these patients, approximatively half of them also have mosaic pattern of hypomethylation affecting other maternally methylated imprinted DMRs (Mackay et al. 2006). Recessive homozygous and compound heterozygous mutations in the maternal-zygotic gene ZFP57 have been identified in approximatively 50% of TNDM1 with MLID patients, consequently
suggesting that MLID may be the result of a *trans*-mechanism defects involving ZFP57 (Mackay et al. 2008).

ZFP57 (ZFP57 zinc finger protein) is a maternal-zygotic gene that encodes for a protein containing a *Krüppel*-associated box (KRAB) domain and 7 zinc fingers. The maternal and zygotic products are both required for normal embryonic development. In mice, the single loss of zygotic Zfp57 results in partial neonatal lethality whilst the loss of both the maternal and the zygotic function results in a highly penetrant embryonic lethality (X. Li et al. 2008). Zfp57 was shown to be required for the establishment of DNA methylation at some maternally methylated DMRs but overall that it was not an essential factor for general maternal germline imprinting. Indeed, the loss of Zfp57 maternal function in oocytes results in the failure to establish methylation at the maternally methylated *Snrpn* DMR but not at *Peg1*, *Peg3*, and *Peg5/Nnat* DMRs (X. Li et al. 2008). ZFP57 is also essential for faithful maintenance of DNA methylation at paternal and maternal imprinted DMRs. The maintenance of DNA methylation at the paternally methylated IG-DMR and at the maternally methylated *Snrpn, Peg1, Peg3*, and *Peg5/Nnat* DMRs failed in embryos lacking both maternal and zygotic Zfp57 functions. In embryos lacking Zfp57 maternal function only, the loss of DNA methylation is rescued by the zygotic Zfp57, whilst the maintenance of DNA methylation is sometimes compromised in embryos lacking the zygotic Zfp57 only (X. Li et al. 2008). In humans, the study of TNDM1 patients showed that pathogenic mutations in *ZFP57* was associated with mosaic pattern of MLID, hence demonstrating the critical role of *ZFP57* in the maintenance of genomic
imprinting methylation (Mackay et al. 2008). Finally, the investigation of mouse embryonic stem cells revealed that ZFP57 recognises and binds specifically to the methylated TGCC[met]GC hexanucleotide motif and subsequently associates with TRIM28 (also known as KAP1), the H3K9 methyltransferase SETDB1 and the heterochromatin HP1 (Quenneville et al. 2011; Zuo et al. 2012). The protein complex is particularly enriched at all known imprinting centres, protect the DNA methylation from the genome-wide epigenetic reprogramming occurring after fertilisation and maintain the DNA methylation status during development, consequently preserving the parent of origin-specific gene expression (Quenneville et al. 2011). The knockout of Zfp57 in mouse embryonic stem cells demonstrated that the loss of Zfp57 results in hypomethylation at multiple imprinted loci (Quenneville et al. 2011; Zuo et al. 2012) and that the loss of the epigenetic memory at these DMRs was irreversible (i.e. not re-established) even upon re-introduction of exogenous Zfp57 (Zuo et al. 2012).

5.1.2.4 NLRP2

A trans-mechanism involving the maternal-effect gene NLRP2 (NLR family pyrin domain containing 2) has been described in BWS with MLID. NLRP2 is a member of the NACHT, LRR and PYD domains-containing protein family and is highly homologous to NLRP7. A recessive homozygous frameshift mutation affecting NLRP2 and resulting in a truncated protein has been found in the unaffected mother of three children, of which two were diagnosed with BWS. Complex consanguinity was described in the family and the mother was also suspected to have a hydatidiform mole pregnancy. One child had LOM at KCNQ1OT1:TSS-DMR and
was homozygous for the mutation. The second child with BWS had LOM at both
KCNQ1OT1:TSS-DMR and at MEST:alt-TSS-DMR and was heterozygous for the
mutation. Consequently, it was suggested that the BWS phenotype was likely due to
NLRP2 maternal-effect mutation rather than the loss of NLRP2 zygotic function. To
date, it is the only report linking imprinting disorders to mutations in NLRP2, which
suggests that NLRP2-induced imprinting disorders is very rare and appears to affect
a small minority of cases.

NLRP2 has been shown to negatively regulate the activation of the transcription
factor NFkB induced by various pro-inflammatory stimuli, including tumour necrosis
factor alpha (TNFa) and interleukin 1 beta (IL1B) (Bruey et al. 2004). The interaction
of NLRP2 with PYCARD, a member of the caspase-associated recruitment domains-
containing adaptor protein family, was shown to enhance caspase 1 activation and
IL1B secretion (Bruey et al. 2004). Multiple evidences have suggested that NLRP2
is a maternal-effect gene required for early embryonic development. Similarly to
NLRP7, NLRP2 transcripts are found in all oocyte and pre-implantation embryo
stages (Murdoch et al. 2006) and NLRP2 transcript level is at the highest in the
oocyte, then slowly decrease from the germinal vesicle stage to day 3 embryos and
finally increase again on day 5 (P. Zhang et al. 2008). The knockdown of Nlrp2 in
mouse oocytes results in normal maturation of the oocytes but the development of
parthenogenetic embryos derived from Nlrp2 knockdown oocytes mainly arrest at
the 2-cell stage. In contrast, 60% of the control parthenogenetic embryos reached
the blastocyst stage (Peng et al. 2012). Similarly, the depletion of maternal Nlrp2 in
zygotic embryos also showed an early embryonic development arrest between the 2- to 8-cells stage (Peng et al. 2012). Finally, as mentioned above, a maternal-effect mutation in \textit{NLRP2} was linked to MLID and BWS phenotype. Therefore, it was suggested that \textit{NLRP2} might play a role in the establishment and or maintenance of the DNA methylation at imprinted regions, but the exact mechanism remains unknown (Meyer et al. 2009).

5.1.2.5 \textbf{KHDC3L}


\textit{KHDC3L} (KH domain containing 3 like, subcortical maternal complex member), is a maternal-effect gene that has been found responsible for approximately 5% of \textit{NLRP7}-negative familial biparental hydatidiform moles cases (Parry et al. 2011). To date, 6 maternal-effect recessive mutations in \textit{KHDC3L} gene have been described in the aetiology of FBHM (Parry et al. 2011; Reddy et al. 2013; Rezaei et al. 2016).

KHDC3L precise protein function remains unknown but the protein belongs to the KHDC1 (KH homology domain containing 1) protein family which are known to bind RNA. Several pieces of evidence suggest that \textit{KHDC3L} is a maternal-effect gene and may have similar or overlapping functions to \textit{NLRP7} in the oocyte and in early embryonic development. As mentioned above, pathogenic maternal recessive mutations in \textit{KHDC3L} are associated with abnormal embryonic development in FBHM (Parry et al. 2011). \textit{KHBC3L} has a very similar expression pattern than \textit{NLRP7}; \textit{KHDC3L} transcript levels appear to be at their highest in germinal vesicle oocytes and then decrease during preimplantation development and become
undetectable at the blastocyst stage (Parry et al. 2011). KHDC3L colocalises with NLRP7 to the microtubule organising centre and the Golgi apparatus in human hematopoietic cells (Messaed et al. 2011; Reddy et al. 2013), and at the cytoskeleton of the oocytes where they are both particularly abundant at the cortical region (Akoury et al. 2014). Finally, KHDC3L forms with NLRP5 (NLR family pyrin domain containing 5), OOEP (oocyte expressed protein) and TLE6 (transducin like enhancer of split 6) the human subcortical maternal complex (SMC) protein (Zhu et al. 2015). The human SMC may have a very similar role to the SMC counterpart in mice, which was found to be essential for developmental progression beyond the first zygotic cell divisions (L. Li et al. 2008; Yu et al. 2014). The MLID at maternal imprinted DMRs associated with KHDC3L-positive FBHM cases also suggests that KHDC3L may play a role in the establishment or maintenance of methylation at maternal imprinted DMRs during oogenesis and early embryogenesis respectively. However, and similar to NLRP7, this mechanism currently remains unknown.

5.1.2.6 NLRP5

Seven individuals with MLID were recently reported in five families. The clinical features of the affected individuals were heterogeneous, with three children diagnosed with BWS and MLID, two with SRS and MLID and two with non-specific phenotype and MLID. Pregnancy losses and infertility issues were also reported in some families. Genetic investigations of proband’s family pedigree have identified pathogenic homozygous and compound heterozygous mutations in the maternal-effect gene NLRP5 in the proband’s mother. It was consequently suggested that the
maternal NLRP5 depletion was responsible for the diseases phenotype and the pregnancy losses and was likely involved in the genomic imprinting disturbances (Docherty et al. 2015).

Nlrp5, also known as Mater, is exclusively expressed in the oocyte and was first described as an antigen associated with ovarian autoimmunity in mice (Tong and Nelson 1999). Nlrp5 was later found to be a maternal-effect gene required for normal embryonic development (Tong et al. 2000). The Nlrp5 null female mice have normal oocyte maturation and ovulation, but they are sterile. The Nlrp5 maternal loss of function results in embryonic developmental arrest at the 2-cell stage. Similar phenotypes were also observed in rhesus macaque monkeys, in which the maternal depletion of NLRP5 resulted in embryonic developmental arrest between the 8-cell and the 16-cell stage (Wu 2009). Some data also showed that the depletion of NLRP5 in mice disrupts the assembly of the subcortical maternal complex, which was found to be essential for developmental progression beyond the first zygotic cell divisions (L. Li et al. 2008; Yu et al. 2014). Finally, both paternal and maternal imprinted DMRs are disturbed in MLID patients with a maternal depletion of NLRP5 (Docherty et al. 2015), hence suggesting an unrecognised role in the maintenance of genomic imprinting methylation.

5.1.2.7 Other candidate genes

In addition to the few known causative genes described in the aetiology of MLID, other candidates have been proposed to be involved in the disruption of genomic
imprinting via *trans*-mechanism defects. This includes the maintenance DNA methyltransferase DNMT1, the active *de novo* DNA methyltransferases DNMT3A, DNMT3B and the enzymatically inactive DNMT3L, the ZFP57 co-factor TRIM28, the DNA-binding protein CTCF, the methyl-CpG-binding proteins MBD3 and the developmental pluripotency-associated protein DPPA3 (Begemann et al. 2011; Azzi et al. 2014; Caliebe et al. 2014). However, to date no additional causative gene mutations have been described in MLID pathway.

### 5.1.3 Aim

The underlying causes resulting in MLID are still not clearly understood but a few hypotheses have been formulated to explain the widespread methylation defects observed in imprinting disorders. As seen previously, environmental insults such as ART could directly disrupt the genomic imprinting. Additionally, mutations occurring in maternal-effect genes have been found in patients with MLID, hence suggesting those genes might play a critical role in MLID aetiology. Although it has not been reported yet, I also hypothesised that mutations in genes associated to the methylation pathway could cause methylation disturbances at imprinted loci via a *trans*-mechanism. Finally, many imprinted genes belong to a complex imprinted gene network. I hypothesised that pathogenic mutations occurring in one of the member of the network may result to *cis*-defects and then to possible subsequent *trans*-methylation disturbances at other members of the network. In that regards, in order to gain additional insights into the underlying causes responsible for MLID in
BWS I performed methylation profiling using the HM450K platform and my bioinformatic pipeline (see section 4.2.4 - DNA methylation profiling using HM450K) of BWS_IC2 patients conceived with the help of ART and studied the effects of the ART procedures on genomic imprinting methylation. Additionally, using whole-exome sequencing (WES) I compared the genetics of BWS_IC2 individuals with and without MLID to identify potential genetic components responsible for the widespread genomic imprinting disturbances.

5.2 Materials and methods

5.2.1 Patient DNA

According to the criteria described in the literature (http://www.geneclinics.org), BWS was diagnosed in 78 screened patients. The molecular diagnostic investigations undertaken in NHS laboratories identified LOM at \( KCNQ1OT1: \) TSS-DMR (BWS_IC2) as the cause of the BWS phenotype. 13 of these patients were reported as conceived with the help of ART (2 following OS, 3 following IVF and 8 following ICSI). The use of ART was not reported in the remaining 65 BWS_IC2 individuals.

5.2.2 DNA quantification

See section 2.3 - DNA quantification.

5.2.3 DNA purification

See section 2.2 - DNA purification.
5.2.4 Sodium bisulfite treatment

See section 2.4 – Sodium bisulfite treatment.

5.2.5 DNA methylation profiling using HM450K

See section 4.2.4 - DNA methylation profiling using HM450K.

5.2.6 Whole-exome sequencing

In order to investigate the potential genetic causes that could lead to MLID, 25 BWS_IC2 patients with MLID and 17 BWS_IC2 patients with no MLID were processed for whole-exome sequencing and analysed as described below (Figure 5.2).

5.2.6.1 Library preparation and sequencing

gDNA was quantified with Qubit® dsDNA HS Assay Kits (see section 2.3 - DNA quantification) and 50 ng was used to perform whole-exome sequencing. Library preparation (Nextera Rapid Capture Exomes (Dual Indexed), Illumina) and sequencing (NextSeq500, High Output 300 cycle (2x150bp Paired End), Illumina) was performed at the Stratified Medicine Core Laboratory (SMCL) Next Generation Sequencing facility hosted at the Department of Medical Genetics, University of Cambridge.
5.2.6.2 Alignment and variant calling

Read alignments and variant calling were kindly done by Dr. Ezequiel Martín Rodríguez. The bioinformatic pipeline relies on freely available bioinformatic tools such as Burrows–Wheeler transformation (BWA) v0.7.5a-r405 for sequences alignment to the Genome Reference Consortium Human Build 38 patch release 6 (GRCh38.p6), samtools (H. Li et al. 2009) v0.1.19 for manipulating files and removing reads duplicates and Genome Analysis Toolkit (GATK) (McKenna et al. 2010; DePristo et al. 2011) v3.3-0-g37228af for realigning complicated regions such as insertion-deletion (indel), recalibration of base quality scores (i.e. Q scores) and variant calling.

5.2.6.3 Filtering and annotation

vcftools (Danecek et al. 2011) v0.1.12b was used to filter out low quality score (Q <30), low read depth coverage (DP <20x) and homozygous variants for the reference allele. Filtered files were then annotated with Annovar (Kai Wang et al. 2010) using available database. This includes, but not limited to, gene, transcript and protein names (e.g. NCBI Reference Sequence (RefSeq)), single nucleotide polymorphism (e.g. avSNP147), allele frequency (e.g. 1000 Genome Project, Exome Aggregation Consortium (ExAC), NHLBI GO Exome Sequencing Project (ESP)), genomic variation and its relationship to human health (e.g. ClinVar, Catalogue of Somatic Mutations in Cancer (COSMIC)), effects of amino acid substitution or indels on the structure and function of a human protein (e.g. SIFT (Henikoff and M.M. Smith 2015),
PROVEAN (Choi et al. 2012), PolyPhen-2 (Adzhubei et al. 2010)) and effects of
genetic variants on human splicing (e.g. ada-boost and random forest dbscSNV
scores (Jian et al. 2014), SPIDEX™ (Xiong et al. 2015)).

5.2.6.4 Variant call analysis

Filtered variant call files were analysed in R. Variants occurring in exonic or splicing
regions with minor allele frequency below 1% in the frequency databases 1000
Genome Project, ExAC and ESP and affecting protein function or splicing were
retained for analysis. Based on literature research, I elaborated a list of candidate
genes for which I believed genetic mutation altering their functions may have
dramatic consequences in the establishment, maintenance or erasure of genomic
imprinting. Consequently, rare variants occurring in gene shown to have a maternal-
effect in mammalian species (Table 5.11) or associated with methylation pathways
(Table 5.12) were flagged as potential disease causing candidates. Furthermore, due
to the existence of complex interactions between imprinted genes (i.e. imprinted
gene network (Varrault et al. 2006; Patten et al. 2016; Soellner et al. 2016)), rare
variants occurring in imprinted genes (Table 5.13) were also flagged as potential
disease causing candidates. Finally, web-tools such as ToppGene (J. Chen et al.
2009) were also used to prioritise candidate genes.
25 BWS_IC2 with MLID  
17 BWS_IC2 with no MLID

42 individuals

Whole-exome sequencing

BWA
Read alignments to reference genome (GRCh38.p6)

samtools
PCR duplicates removal

GATK
Re-alignment of complicated regions such as insertion-deletion and re-calibration of base quality scores

GATK
Variant calling

vcftools
Variant filtering
(low quality score (Q < 30), low read depth coverage (DP < 20x) and homozygous variants for the reference allele)

Annovar
Variant annotation
(e.g. dbSNP, allele frequency, effect of the variant on the structure and function of the protein and the splicing)

R
Variant filtering
(exclusion of variants with minor allele frequency in ExACT, ESP, 1000 Genome project data > 1%, in non exonic or splicing regions, with no predicted damaging effects on protein or splicing)

Identification of possible causal mutations that could lead to MLID epigenotype in BWS_IC2

- Prioritisation of genes based on candidate genes list (methylation associated genes, maternal-effect genes, imprinted genes)
- Prioritisation of candidate genes based on functional similarity to training genes list (known genes in BWS aetiology)

Microsoft Excel
ToppGene

Sanger sequencing validation

Figure 5.2: Overview of the whole-exome sequencing workflow.
## Table 5.12: List of methylation associated genes.

<table>
<thead>
<tr>
<th>GENE</th>
<th>LOCATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAZ2A</td>
<td>12q13.3</td>
<td>bromodomain adjacent to zinc finger domain 2A</td>
</tr>
<tr>
<td>BAZ2B</td>
<td>2q24.2</td>
<td>bromodomain adjacent to zinc finger domain 2B</td>
</tr>
<tr>
<td>MBD1</td>
<td>18q21</td>
<td>methyl-CpG binding domain protein 1</td>
</tr>
<tr>
<td>MBD2</td>
<td>18q21</td>
<td>methyl-CpG binding domain protein 2</td>
</tr>
<tr>
<td>MBD3</td>
<td>19p13</td>
<td>methyl-CpG binding domain protein 3</td>
</tr>
<tr>
<td>MBD4</td>
<td>3q21.3</td>
<td>methyl-CpG binding domain 4, DNA glycosylase</td>
</tr>
<tr>
<td>MBD5</td>
<td>2q23.2</td>
<td>methyl-CpG binding domain protein 5</td>
</tr>
<tr>
<td>MBD6</td>
<td>12q13.2</td>
<td>methyl-CpG binding domain protein 6</td>
</tr>
<tr>
<td>MECP2</td>
<td>Xq28</td>
<td>methyl-CpG binding protein 2</td>
</tr>
<tr>
<td>SETDB1</td>
<td>1q21</td>
<td>SET domain bifurcated 1</td>
</tr>
<tr>
<td>SETDB2</td>
<td>13q14</td>
<td>SET domain bifurcated 2</td>
</tr>
<tr>
<td>SUV39H1</td>
<td>Xp11.23</td>
<td>suppressor of variegation 3-9 homolog 1</td>
</tr>
<tr>
<td>SUV39H2</td>
<td>10p13</td>
<td>suppressor of variegation 3-9 homolog 2</td>
</tr>
<tr>
<td>EHMT2</td>
<td>6p21.3</td>
<td>euchromatic histone lysine methyltransferase 2</td>
</tr>
<tr>
<td>FHMT1</td>
<td>9q34.3</td>
<td>euchromatic histone lysine methyltransferase 1</td>
</tr>
<tr>
<td>KMT2A</td>
<td>11q23</td>
<td>lysine methyltransferase 2A</td>
</tr>
<tr>
<td>KMT2B</td>
<td>19q13.12</td>
<td>lysine methyltransferase 2B</td>
</tr>
<tr>
<td>KMT2C</td>
<td>7q36</td>
<td>lysine methyltransferase 2C</td>
</tr>
<tr>
<td>KMT2D</td>
<td>12q13.12</td>
<td>lysine methyltransferase 2D</td>
</tr>
<tr>
<td>KMT2E</td>
<td>7q22.1</td>
<td>lysine methyltransferase 2E</td>
</tr>
<tr>
<td>SETD1A</td>
<td>16p11.2</td>
<td>SET domain containing 1A</td>
</tr>
<tr>
<td>SETD1B</td>
<td>12q24.31</td>
<td>SET domain containing 1B</td>
</tr>
<tr>
<td>ASH1L</td>
<td>1q22</td>
<td>ASH1 like histone lysine methyltransferase</td>
</tr>
<tr>
<td>SETD2</td>
<td>3p21.31</td>
<td>SET domain containing 2</td>
</tr>
<tr>
<td>NSD1</td>
<td>5q35</td>
<td>nuclear receptor binding SET domain protein 1</td>
</tr>
<tr>
<td>SMYD2</td>
<td>1q32.3</td>
<td>SET and MYND domain containing 2</td>
</tr>
<tr>
<td>SMYD1</td>
<td>2p11.1</td>
<td>SET and MYND domain containing 1</td>
</tr>
<tr>
<td>SMYD3</td>
<td>1q44</td>
<td>SET and MYND domain containing 3</td>
</tr>
<tr>
<td>DOT1L</td>
<td>19p13.3</td>
<td>DOT1 like histone lysine methyltransferase</td>
</tr>
<tr>
<td>KMT5A</td>
<td>12q24.31</td>
<td>lysine methyltransferase 5A</td>
</tr>
<tr>
<td>KMT5B</td>
<td>11q13.2</td>
<td>lysine methyltransferase 5B</td>
</tr>
<tr>
<td>KMT5C</td>
<td>19q13.42</td>
<td>lysine methyltransferase 5C</td>
</tr>
<tr>
<td>SETD7</td>
<td>4q31.1</td>
<td>SET domain containing lysine methyltransferase 7</td>
</tr>
<tr>
<td>PRDM2</td>
<td>1p36</td>
<td>PR/SET domain 2</td>
</tr>
<tr>
<td>KDM1A</td>
<td>1p36.12</td>
<td>lysine demethylase 1A</td>
</tr>
<tr>
<td>KDM1B</td>
<td>6p22.3</td>
<td>lysine demethylase 1B</td>
</tr>
<tr>
<td>KDM2A</td>
<td>11q13.1</td>
<td>lysine demethylase 2A</td>
</tr>
<tr>
<td>KDM2B</td>
<td>12q24.31</td>
<td>lysine demethylase 2B</td>
</tr>
<tr>
<td>KDM3A</td>
<td>2p11.2</td>
<td>lysine demethylase 3A</td>
</tr>
<tr>
<td>KDM3B</td>
<td>5q31</td>
<td>lysine demethylase 3B</td>
</tr>
<tr>
<td>KDM4A</td>
<td>1p34.1</td>
<td>lysine demethylase 4A</td>
</tr>
<tr>
<td>KDM4B</td>
<td>19p13.3</td>
<td>lysine demethylase 4B</td>
</tr>
<tr>
<td>KDM4C</td>
<td>9p24-p23</td>
<td>lysine demethylase 4C</td>
</tr>
<tr>
<td>KDM4D</td>
<td>11q21</td>
<td>lysine demethylase 4D</td>
</tr>
</tbody>
</table>
Table 5.12: List of methylation associated genes (continued).

<table>
<thead>
<tr>
<th>GENE</th>
<th>LOCATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDM4E</td>
<td>11q21</td>
<td>lysine demethylase 4E</td>
</tr>
<tr>
<td>KDM5A</td>
<td>12p13.33</td>
<td>lysine demethylase 5A</td>
</tr>
<tr>
<td>KDM5B</td>
<td>1q32.1</td>
<td>lysine demethylase 5B</td>
</tr>
<tr>
<td>KDM5C</td>
<td>Xp11.22-p11.21</td>
<td>lysine demethylase 5C</td>
</tr>
<tr>
<td>KDM5D</td>
<td>Yq11</td>
<td>lysine demethylase 5D</td>
</tr>
<tr>
<td>KDM6A</td>
<td>Xp11.2</td>
<td>lysine demethylase 6A</td>
</tr>
<tr>
<td>KDM6B</td>
<td>17p13.1</td>
<td>lysine demethylase 6B</td>
</tr>
<tr>
<td>KDM7A</td>
<td>7q34</td>
<td>lysine demethylase 7A</td>
</tr>
<tr>
<td>PHF8</td>
<td>Xp11.22</td>
<td>PHD finger protein 8</td>
</tr>
<tr>
<td>PHF2</td>
<td>9q22</td>
<td>PHD finger protein 2</td>
</tr>
<tr>
<td>KDM8</td>
<td>16p12.1</td>
<td>lysine demethylase 8</td>
</tr>
<tr>
<td>HDAC1</td>
<td>1p34</td>
<td>histone deacetylase 1</td>
</tr>
<tr>
<td>HDAC2</td>
<td>6q21</td>
<td>histone deacetylase 2</td>
</tr>
<tr>
<td>HDAC3</td>
<td>5q31.1-q31.2</td>
<td>histone deacetylase 3</td>
</tr>
<tr>
<td>HDAC8</td>
<td>Xq13</td>
<td>histone deacetylase 8</td>
</tr>
<tr>
<td>HDAC4</td>
<td>2q37.3</td>
<td>histone deacetylase 4</td>
</tr>
<tr>
<td>HDAC5</td>
<td>17q21</td>
<td>histone deacetylase 5</td>
</tr>
<tr>
<td>HDAC7</td>
<td>12q13.1</td>
<td>histone deacetylase 7</td>
</tr>
<tr>
<td>HDAC9</td>
<td>7p21.1</td>
<td>histone deacetylase 9</td>
</tr>
<tr>
<td>HDAC6</td>
<td>Xp11.23</td>
<td>histone deacetylase 6</td>
</tr>
<tr>
<td>HDAC10</td>
<td>22q13.31</td>
<td>histone deacetylase 10</td>
</tr>
<tr>
<td>HDAC11</td>
<td>3p25.1</td>
<td>histone deacetylase 11</td>
</tr>
<tr>
<td>SIRT1</td>
<td>10q21</td>
<td>sirtuin 1</td>
</tr>
<tr>
<td>SIRT2</td>
<td>19q13</td>
<td>sirtuin 2</td>
</tr>
<tr>
<td>SIRT3</td>
<td>11p15.5</td>
<td>sirtuin 3</td>
</tr>
<tr>
<td>SIRT4</td>
<td>12q24.31</td>
<td>sirtuin 4</td>
</tr>
<tr>
<td>SIRT5</td>
<td>6p23</td>
<td>sirtuin 5</td>
</tr>
<tr>
<td>SIRT6</td>
<td>19p13.3</td>
<td>sirtuin 6</td>
</tr>
<tr>
<td>SIRT7</td>
<td>17q25.1</td>
<td>sirtuin 7</td>
</tr>
<tr>
<td>PHC1</td>
<td>12p13</td>
<td>polyhomeotic homolog 1</td>
</tr>
<tr>
<td>PHC2</td>
<td>1p34.3</td>
<td>polyhomeotic homolog 2</td>
</tr>
<tr>
<td>PHC3</td>
<td>3q26.32</td>
<td>polyhomeotic homolog 3</td>
</tr>
<tr>
<td>CBX6</td>
<td>22q13.1</td>
<td>chromobox 6</td>
</tr>
<tr>
<td>CBX2</td>
<td>17q25.3</td>
<td>chromobox 2</td>
</tr>
<tr>
<td>CBX4</td>
<td>17q25.3</td>
<td>chromobox 4</td>
</tr>
<tr>
<td>CBX7</td>
<td>22q13.1</td>
<td>chromobox 7</td>
</tr>
<tr>
<td>CBX8</td>
<td>17q25.3</td>
<td>chromobox 8</td>
</tr>
<tr>
<td>PCGF1</td>
<td>2p13.1</td>
<td>polycomb group ring finger 1</td>
</tr>
<tr>
<td>PCGF2</td>
<td>17q12</td>
<td>polycomb group ring finger 2</td>
</tr>
<tr>
<td>PCGF3</td>
<td>4p16.3</td>
<td>polycomb group ring finger 3</td>
</tr>
<tr>
<td>BMI1</td>
<td>10p13</td>
<td>BMI1 proto-oncogene, polycomb ring finger</td>
</tr>
<tr>
<td>PCGF5</td>
<td>10q23.33</td>
<td>polycomb group ring finger 5</td>
</tr>
<tr>
<td>PCGF6</td>
<td>10q24.33</td>
<td>polycomb group ring finger 6</td>
</tr>
<tr>
<td>RING1</td>
<td>6p21.3</td>
<td>ring finger protein 1</td>
</tr>
</tbody>
</table>
1 Table 5.12: List of methylation associated genes (continued).

<table>
<thead>
<tr>
<th>GENE</th>
<th>LOCATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNF2</td>
<td>1q25.3</td>
<td>ring finger protein 2</td>
</tr>
<tr>
<td>EZH1</td>
<td>17q21.1-q21.3</td>
<td>enhancer of zeste 1 polycomb repressive complex 2 subunit</td>
</tr>
<tr>
<td>EZH2</td>
<td>7q35-q36</td>
<td>enhancer of zeste 2 polycomb repressive complex 2 subunit</td>
</tr>
<tr>
<td>SUZ12</td>
<td>17q21</td>
<td>SUZ12 polycomb repressive complex 2 subunit</td>
</tr>
<tr>
<td>EED</td>
<td>11q14.2-q22.3</td>
<td>embryonic ectoderm development</td>
</tr>
<tr>
<td>ACTL6A</td>
<td>3q26.33</td>
<td>actin like 6A</td>
</tr>
<tr>
<td>ACTL6B</td>
<td>7q22</td>
<td>actin like 6B</td>
</tr>
<tr>
<td>ARID1A</td>
<td>1p36.1-p35</td>
<td>AT-rich interaction domain 1A</td>
</tr>
<tr>
<td>ARID1B</td>
<td>6q25.3</td>
<td>AT-rich interaction domain 1B</td>
</tr>
<tr>
<td>ARID2</td>
<td>12q13.11</td>
<td>AT-rich interaction domain 2</td>
</tr>
<tr>
<td>PBRM1</td>
<td>3p21</td>
<td>polybromo 1</td>
</tr>
<tr>
<td>SMARCA2</td>
<td>9p24.3</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily a, member 2</td>
</tr>
<tr>
<td>SMARCA4</td>
<td>19p13.3</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily a, member 4</td>
</tr>
<tr>
<td>SMARCA5</td>
<td>4q31.1-q31.2</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily a, member 5</td>
</tr>
<tr>
<td>SMARCBI</td>
<td>22q11.23</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily b, member 1</td>
</tr>
<tr>
<td>SMARCC1</td>
<td>3p21.31</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily c member 1</td>
</tr>
<tr>
<td>SMARCC2</td>
<td>12q13.2</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily c member 2</td>
</tr>
<tr>
<td>SMARCD1</td>
<td>12q13-q14</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily d, member 1</td>
</tr>
<tr>
<td>SMARCD2</td>
<td>17q23.3</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily d, member 2</td>
</tr>
<tr>
<td>SMARCD3</td>
<td>7q35-q36</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily d, member 3</td>
</tr>
<tr>
<td>SMARCE1</td>
<td>17q21.2</td>
<td>SWI/SNF related, matrix associated, actin dependent regulator of chromatin,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subfamily e, member 1</td>
</tr>
<tr>
<td>BCL11A</td>
<td>2p16.1</td>
<td>B-cell CLL/lymphoma 11A</td>
</tr>
<tr>
<td>BCL11B</td>
<td>14q32</td>
<td>B-cell CLL/lymphoma 11B</td>
</tr>
<tr>
<td>BCL7A</td>
<td>12q24.1</td>
<td>BCL tumor suppressor 7A</td>
</tr>
<tr>
<td>BCL7B</td>
<td>7q11.23</td>
<td>BCL tumor suppressor 7B</td>
</tr>
<tr>
<td>BCL7C</td>
<td>16p11</td>
<td>BCL tumor suppressor 7C</td>
</tr>
<tr>
<td>BRD7</td>
<td>16q12.1</td>
<td>bromodomain containing 7</td>
</tr>
<tr>
<td>BRD9</td>
<td>5p15.33</td>
<td>bromodomain containing 9</td>
</tr>
<tr>
<td>SS18</td>
<td>18q11.2</td>
<td>SS18, nBAF chromatin remodeling complex subunit</td>
</tr>
<tr>
<td>DPFI</td>
<td>19q13.12</td>
<td>double PHD fingers 1</td>
</tr>
<tr>
<td>DPF2</td>
<td>11q13.1</td>
<td>double PHD fingers 2</td>
</tr>
<tr>
<td>DPF3</td>
<td>14q24.2</td>
<td>double PHD fingers 3</td>
</tr>
<tr>
<td>DNMT1</td>
<td>19p13.2</td>
<td>DNA methyltransferase 1</td>
</tr>
<tr>
<td>DNMT3A</td>
<td>2p23</td>
<td>DNA methyltransferase 3 alpha</td>
</tr>
</tbody>
</table>
**Table 5.12: List of methylation associated genes (continued).**

<table>
<thead>
<tr>
<th>GENE</th>
<th>LOCATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNMT3B</td>
<td>20q11.2</td>
<td>DNA methyltransferase 3 beta</td>
</tr>
<tr>
<td>DNMT3L</td>
<td>21q22.3</td>
<td>DNA methyltransferase 3 like</td>
</tr>
<tr>
<td>TET1</td>
<td>10q21</td>
<td>tet methylcytosine dioxygenase 1</td>
</tr>
<tr>
<td>TET2</td>
<td>4q24</td>
<td>tet methylcytosine dioxygenase 2</td>
</tr>
<tr>
<td>TET3</td>
<td>2p13.1</td>
<td>tet methylcytosine dioxygenase 3</td>
</tr>
<tr>
<td>TDG</td>
<td>12q24.1</td>
<td>thymine DNA glycosylase</td>
</tr>
<tr>
<td>TRIM28</td>
<td>19q13.4</td>
<td>tripartite motif containing 2B</td>
</tr>
<tr>
<td>CBX1</td>
<td>17q21.32</td>
<td>chromobox 1</td>
</tr>
<tr>
<td>CBX3</td>
<td>7p15.2</td>
<td>chromobox 3</td>
</tr>
<tr>
<td>CBX5</td>
<td>12q13.13</td>
<td>chromobox 5</td>
</tr>
<tr>
<td>CREBBP</td>
<td>16p13.3</td>
<td>CREB binding protein</td>
</tr>
<tr>
<td>EP300</td>
<td>22q13.2</td>
<td>E1A binding protein p3C0</td>
</tr>
<tr>
<td>PRAM1</td>
<td>19p13.2</td>
<td>PML-RARA regulated adaptor molecule 1</td>
</tr>
<tr>
<td>PCNA</td>
<td>20p13-p12.3</td>
<td>proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>MLH1</td>
<td>3p22.3</td>
<td>mutL homolog 1</td>
</tr>
<tr>
<td>UHRF1</td>
<td>19p13.3</td>
<td>ubiquitin like with PHD and ring finger domains 1</td>
</tr>
</tbody>
</table>
1 Table 5.13: List of imprinted genes.

<table>
<thead>
<tr>
<th>GENE</th>
<th>LOCATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIRAS3</td>
<td>1p31</td>
<td>DIRAS family GTPase 3</td>
</tr>
<tr>
<td>TP73</td>
<td>1p36.3</td>
<td>tumor protein p73</td>
</tr>
<tr>
<td>LRRTM1</td>
<td>2p12</td>
<td>leucine rich repeat transmembrane neuronal 1</td>
</tr>
<tr>
<td>ZDBF2</td>
<td>2q33.3</td>
<td>zinc finger DBF-type containing 2</td>
</tr>
<tr>
<td>GPR1</td>
<td>2q33.3</td>
<td>G protein-coupled receptor 1</td>
</tr>
<tr>
<td>GPR1-AS</td>
<td>2q33.3</td>
<td>GPR1 antisense RNA</td>
</tr>
<tr>
<td>NAP1L5</td>
<td>4q21-q22</td>
<td>nucleosome assembly protein 1 like 5</td>
</tr>
<tr>
<td>RHOBTB3</td>
<td>5q15</td>
<td>Rho related BTB domain containing 3</td>
</tr>
<tr>
<td>FAM50B</td>
<td>6p25.2</td>
<td>family with sequence similarity 50 member B</td>
</tr>
<tr>
<td>LIN28B</td>
<td>6q21</td>
<td>lin-28 homolog B</td>
</tr>
<tr>
<td>AIM1</td>
<td>6q21</td>
<td>absent in melanoma 1</td>
</tr>
<tr>
<td>PLAGL1</td>
<td>6q24-q25</td>
<td>PLAG1 like zinc finger 1</td>
</tr>
<tr>
<td>HYMAI</td>
<td>6q24.2</td>
<td>hydatidiform mole associated and imprinted (non-protein coding)</td>
</tr>
<tr>
<td>PHACTR2</td>
<td>6q24.1</td>
<td>phosphatase and actin regulator 2</td>
</tr>
<tr>
<td>SLC22A2</td>
<td>6q25.3</td>
<td>solute carrier family 22 member 2</td>
</tr>
<tr>
<td>SLC22A3</td>
<td>6q25.3</td>
<td>solute carrier family 22 member 3</td>
</tr>
<tr>
<td>GRB10</td>
<td>7p12.2</td>
<td>growth factor receptor bound protein 10</td>
</tr>
<tr>
<td>CALC1</td>
<td>7q21.3</td>
<td>calcitonin receptor</td>
</tr>
<tr>
<td>TFFP2</td>
<td>7q</td>
<td>tissue factor pathway inhibitor 2</td>
</tr>
<tr>
<td>SGCE</td>
<td>7q21.3</td>
<td>sarcoglycan epsilon</td>
</tr>
<tr>
<td>PEG10</td>
<td>7q21</td>
<td>paternally expressed 10</td>
</tr>
<tr>
<td>CPA4</td>
<td>7q32</td>
<td>carboxypeptidase A4</td>
</tr>
<tr>
<td>MEST</td>
<td>7q32</td>
<td>mesoderm specific transcript</td>
</tr>
<tr>
<td>MESTIT1</td>
<td>7q32.2</td>
<td>MEST intronic transcript 1, antisense RNA</td>
</tr>
<tr>
<td>COPG2IT1</td>
<td>7q32</td>
<td>COPG2 imprinted transcript 1 (non-protein coding)</td>
</tr>
<tr>
<td>COPG2</td>
<td>7q32</td>
<td>coatomer protein complex subunit gamma 2</td>
</tr>
<tr>
<td>KLF14</td>
<td>7q32.3</td>
<td>Kruppel like factor 14</td>
</tr>
<tr>
<td>DLGAP2</td>
<td>8p23</td>
<td>DLG associated protein 2</td>
</tr>
<tr>
<td>ZFAT-AS1</td>
<td>8q24.22</td>
<td>ZFAT antisense RNA 1</td>
</tr>
<tr>
<td>KCNK9</td>
<td>8q24.3</td>
<td>potassium two pore domain channel subfamily K member 9</td>
</tr>
<tr>
<td>INPP5F</td>
<td>10q26.13</td>
<td>inositol polyphosphate-5-phosphatase F</td>
</tr>
<tr>
<td>WT1</td>
<td>11p13</td>
<td>Wilms tumor 1</td>
</tr>
<tr>
<td>WT1-AS</td>
<td>11p13</td>
<td>WT1 antisense RNA</td>
</tr>
<tr>
<td>ZNF215</td>
<td>11p15.4</td>
<td>zinc finger protein 215</td>
</tr>
<tr>
<td>H19</td>
<td>11p15.5</td>
<td>H19, imprinted maternally expressed transcript (non-protein coding)</td>
</tr>
<tr>
<td>IGF2</td>
<td>11p15.5</td>
<td>insulin like growth factor 2</td>
</tr>
<tr>
<td>MIR483</td>
<td>11p15.5</td>
<td>microRNA 483</td>
</tr>
<tr>
<td>IGF2-AS</td>
<td>11p15.5</td>
<td>IGF2 antisense RNA</td>
</tr>
<tr>
<td>INS</td>
<td>11p15.5</td>
<td>insulin</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>11p15.5</td>
<td>potassium voltage-gated channel subfamily Q member 1</td>
</tr>
<tr>
<td>KCNQ1OT1</td>
<td>11p15.5</td>
<td>KCNQ1 opposite strand/antisense transcript 1 (non-protein coding)</td>
</tr>
<tr>
<td>KCNQ1DN</td>
<td>11p15.5</td>
<td>KCNQ1 downstream neighbor (non-protein coding)</td>
</tr>
</tbody>
</table>
Table 5.13: List of imprinted genes (continued).

<table>
<thead>
<tr>
<th>GENE</th>
<th>LOCATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDKN1C</td>
<td>11p15.5</td>
<td>cyclin dependent kinase inhibitor 1C</td>
</tr>
<tr>
<td>SLC22A18</td>
<td>11p15.5</td>
<td>solute carrier family 22 member 18 antisense</td>
</tr>
<tr>
<td>SLC22A18</td>
<td>11p15.5</td>
<td>solute carrier family 22 member 18</td>
</tr>
<tr>
<td>PHLDA2</td>
<td>11p15.4</td>
<td>pleckstrin homology like domain family A member 2</td>
</tr>
<tr>
<td>ANO1</td>
<td>11q13.2</td>
<td>anoctamin 1</td>
</tr>
<tr>
<td>WIF1</td>
<td>12q14.2</td>
<td>WNT inhibitory factor 1</td>
</tr>
<tr>
<td>RB1</td>
<td>13q14.2</td>
<td>RB transcriptional corepressor 1</td>
</tr>
<tr>
<td>RTL1</td>
<td>14q32.2</td>
<td>retrotransposon-like 1</td>
</tr>
<tr>
<td>DLK1</td>
<td>14q32.2</td>
<td>delta like non-canonical Notch ligand 1</td>
</tr>
<tr>
<td>MEG3</td>
<td>14q32.2</td>
<td>maternally expressed 3 (non-protein coding)</td>
</tr>
<tr>
<td>MEG8</td>
<td>14q32.31</td>
<td>maternally expressed 8 (non-protein coding)</td>
</tr>
<tr>
<td>MKRN3</td>
<td>15q11-q13</td>
<td>makorin ring finger protein 3</td>
</tr>
<tr>
<td>MAGEL2</td>
<td>15q11.2</td>
<td>MAGE family member L2</td>
</tr>
<tr>
<td>NDN</td>
<td>15q11.2</td>
<td>neclin, MAGE family member</td>
</tr>
<tr>
<td>PWRN1</td>
<td>15q11.2</td>
<td>Prader-Willi region non-protein coding RNA 1</td>
</tr>
<tr>
<td>NPAP1</td>
<td>15q11.2</td>
<td>nuclear pore associated protein 1</td>
</tr>
<tr>
<td>SNURF</td>
<td>15q11.2</td>
<td>SNRPN upstream reading frame</td>
</tr>
<tr>
<td>UBE3A</td>
<td>15q11.2</td>
<td>ubiquitin protein liçase E3A</td>
</tr>
<tr>
<td>NAA60</td>
<td>16p13.3</td>
<td>N(alpha)-acetyltransferase 60, NatF catalytic subunit</td>
</tr>
<tr>
<td>ZNF597</td>
<td>16p13.3</td>
<td>zinc finger protein 597</td>
</tr>
<tr>
<td>TCEB3C</td>
<td>18q21.1</td>
<td>transcription elongation factor B subunit 3C</td>
</tr>
<tr>
<td>DNM1</td>
<td>19p13.2</td>
<td>DNA methyltransferase 1</td>
</tr>
<tr>
<td>AXL</td>
<td>19q13.1</td>
<td>AXL receptor tyrosine kinase</td>
</tr>
<tr>
<td>ZNF331</td>
<td>19q13</td>
<td>zinc finger protein 331</td>
</tr>
<tr>
<td>MIMT1</td>
<td>19q13.43</td>
<td>MER1 repeat containing imprinted transcript 1 (non-protein coding)</td>
</tr>
<tr>
<td>PEG3</td>
<td>19q13.4</td>
<td>paternally expressed 3</td>
</tr>
<tr>
<td>ZIM2</td>
<td>19q13.4</td>
<td>zinc finger imprinted 2</td>
</tr>
<tr>
<td>BLCAP</td>
<td>20q11.23</td>
<td>bladder cancer associated protein</td>
</tr>
<tr>
<td>NNT</td>
<td>20q11.2-q12</td>
<td>neuronatin</td>
</tr>
<tr>
<td>MCTS2P</td>
<td>20q11.21</td>
<td>malignant T-cell amplified sequence 2, pseudogene</td>
</tr>
<tr>
<td>L3MBTL1</td>
<td>20q13.12</td>
<td>l(3)mbt-like 1 (Drosophila)</td>
</tr>
<tr>
<td>GNAS</td>
<td>20q13.2-q13</td>
<td>GNAS complex locus</td>
</tr>
<tr>
<td>SGK2</td>
<td>20q13.2</td>
<td>SGK2, serine/threonine kinase</td>
</tr>
<tr>
<td>SNU13</td>
<td>22q13</td>
<td>SNU13 homolog, small nuclear ribonucleoprotein (U4/U6.U5)</td>
</tr>
</tbody>
</table>

5.2.7 Sanger sequencing

5.2.7.1 Polymerase chain reaction

For polymerase chain reaction (PCR) amplification, patient gDNA was quantified with
NanoDrop® ND-1000 (see section 2.3 - DNA quantification) and 30 ng of gDNA was mixed with 0.125 μl of 5 U/μl AmpliTaq Gold DNA Polymerase (Thermo Fisher Scientific), 2.5 μl of 10X PCR Buffer I, 0.5 μl of 10 mM dNTP mix, 0.5 μl of each 10 μM forward and reverse primers (see Table 5.14) and distilled water for a final reaction volume of 25 μl. The mixture was mixed by pipetting up and down several times and then placed in a thermocycler. The reaction was then incubated at 95 °C for 10 minutes, 30 cycles of 95 °C for 1 minute; 65 °C for 45 seconds and 72 °C for 45 minutes, 72 °C for 10 minutes and then hold at 4 °C. PCR products and 1 kb DNA ladder (Bioline) were loaded on a 1.5% agarose gel electrophoresis (1.5 g agarose dissolved in 100 ml 1X TAE (see section 2.1.7 - 1X TAE), 2 μl of SYBR Safe) and the gel was run for 40 minutes at 100 V. Once finished, the gel was placed into Gel Doc XR+ System (Bio-Rad) for UV exposition. The size of amplicons was compared to the 1 kb DNA ladder and successful amplified PCR products (i.e. the one with correct size) were used immediately after or stored at -20 °C for later use.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Name</th>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Ref</th>
<th>Alt</th>
<th>Function</th>
<th>Exonic function</th>
<th>dbSNP147</th>
<th>Forward primer (5'→3')</th>
<th>Reverse primer (5'→3')</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDG</td>
<td>variant 1</td>
<td>12</td>
<td>103984920</td>
<td>103984920</td>
<td>A</td>
<td></td>
<td>exonic</td>
<td>frameshift insertion</td>
<td>rs764159857</td>
<td>AGCTCTGCTATGTTATGCTCACATC</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>534</td>
</tr>
<tr>
<td>TDG</td>
<td>variant 2</td>
<td>12</td>
<td>103984921</td>
<td>103984921</td>
<td>G</td>
<td>A</td>
<td>splicing</td>
<td></td>
<td>rs760400700</td>
<td>AGCTCTGCTATGTTATGCTCACATC</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>534</td>
</tr>
<tr>
<td>TDG</td>
<td>variant 3</td>
<td>12</td>
<td>103984922</td>
<td>103984922</td>
<td>T</td>
<td>G</td>
<td>splicing</td>
<td></td>
<td>rs760554309</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>534</td>
</tr>
<tr>
<td>TDG</td>
<td>variant 4</td>
<td>12</td>
<td>103980073</td>
<td>103980073</td>
<td>A</td>
<td>G</td>
<td>splicing</td>
<td></td>
<td>rs760554309</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>534</td>
</tr>
<tr>
<td>TDG</td>
<td>variant 5</td>
<td>12</td>
<td>103984747</td>
<td>103984747</td>
<td>A</td>
<td>C</td>
<td>splicing</td>
<td></td>
<td>rs762057949</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>534</td>
</tr>
<tr>
<td>TDG</td>
<td>variant 6</td>
<td>12</td>
<td>103984748</td>
<td>103984748</td>
<td>C</td>
<td>T</td>
<td>splicing</td>
<td></td>
<td>rs762057949</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>534</td>
</tr>
<tr>
<td>TDG</td>
<td>variant 7</td>
<td>12</td>
<td>103985729</td>
<td>103985729</td>
<td>G</td>
<td>T</td>
<td>splicing</td>
<td></td>
<td>rs762057949</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>AGAGCAGCTCTCCTGCTGGAATA</td>
<td>534</td>
</tr>
<tr>
<td>PHF10</td>
<td>variant 8</td>
<td>6</td>
<td>169715806</td>
<td>169715806</td>
<td>A</td>
<td>C</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs562092150</td>
<td>AACACACACACATAGTGTAGAA</td>
<td>TGGGTCAAAATAGCTTACACGAT</td>
<td>505</td>
</tr>
<tr>
<td>PHF10</td>
<td>variant 9</td>
<td>6</td>
<td>169715755</td>
<td>169715755</td>
<td>A</td>
<td>G</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs77919800</td>
<td>AACACACACACATAGTGTAGAA</td>
<td>TGGGTCAAAATAGCTTACACGAT</td>
<td>505</td>
</tr>
<tr>
<td>PHF10</td>
<td>variant 10</td>
<td>6</td>
<td>169715761</td>
<td>169715761</td>
<td>A</td>
<td>C</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs144595699</td>
<td>AACACACACACATAGTGTAGAA</td>
<td>TGGGTCAAAATAGCTTACACGAT</td>
<td>505</td>
</tr>
<tr>
<td>BNC1</td>
<td>variant 11</td>
<td>15</td>
<td>83264026</td>
<td>83264026</td>
<td>G</td>
<td>T</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs145472704</td>
<td>AGATATCCAGAGCTTGTTGCTTA</td>
<td>GCAAAAGTCTAAGCAGGAA</td>
<td>503</td>
</tr>
<tr>
<td>CTCF</td>
<td>variant 12</td>
<td>16</td>
<td>67363779</td>
<td>67363779</td>
<td>C</td>
<td>T</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs145727034</td>
<td>TTTTCTTTCATGCTACTGCAACCACTTCTT</td>
<td>TATCCAAATACCCGCCAATTACAC</td>
<td>504</td>
</tr>
<tr>
<td>NLRP2</td>
<td>variant 13</td>
<td>19</td>
<td>54982915</td>
<td>54982915</td>
<td>G</td>
<td>C</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs139903547</td>
<td>TTTTCTTTCATGCTACTGCAACCACTTCTT</td>
<td>TATCCAAATACCCGCCAATTACAC</td>
<td>504</td>
</tr>
<tr>
<td>NLRP2</td>
<td>variant 14</td>
<td>19</td>
<td>54986222</td>
<td>54986222</td>
<td>C</td>
<td>T</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs200375320</td>
<td>TTTTCTTTCATGCTACTGCAACCACTTCTT</td>
<td>TATCCAAATACCCGCCAATTACAC</td>
<td>504</td>
</tr>
<tr>
<td>NLRP2</td>
<td>variant 15</td>
<td>19</td>
<td>54990056</td>
<td>54990056</td>
<td>G</td>
<td>A</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs117066658</td>
<td>TATTTTCTTTCATGCTACTGCAACCACTTCTT</td>
<td>TATCCAAATACCCGCCAATTACAC</td>
<td>504</td>
</tr>
<tr>
<td>ZDBF2</td>
<td>variant 16</td>
<td>2</td>
<td>206304878</td>
<td>206304878</td>
<td>T</td>
<td>G</td>
<td>exonic</td>
<td>stop-gain</td>
<td>rs575604527</td>
<td>TATTTTCTTTCATGCTACTGCAACCACTTCTT</td>
<td>TATCCAAATACCCGCCAATTACAC</td>
<td>508</td>
</tr>
<tr>
<td>ZDBF2</td>
<td>variant 17</td>
<td>2</td>
<td>206310752</td>
<td>206310752</td>
<td>G</td>
<td>T</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs36095066</td>
<td>TATTTTCTTTCATGCTACTGCAACCACTTCTT</td>
<td>TATCCAAATACCCGCCAATTACAC</td>
<td>504</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>variant 18</td>
<td>11</td>
<td>2585225</td>
<td>2585225</td>
<td>G</td>
<td>C</td>
<td>exonic</td>
<td>non-synonymous SNV</td>
<td>rs36095066</td>
<td>TATTTTCTTTCATGCTACTGCAACCACTTCTT</td>
<td>TATCCAAATACCCGCCAATTACAC</td>
<td>504</td>
</tr>
</tbody>
</table>

*Primers were selected from Primer Designer™ Tool database (Thermo Fisher Scientific)*
5.2.7.2 PCR product purification

Excess of primers and other types of single stranded DNA contaminants were removed from the successfully amplified PCR products using MultiScreen-PCR96 Filter Plate (Millipore). All the PCR product were pipetted into the filter plate. The plate was placed on a vacuum manifold and pressed down until a vacuum is created. The plate was let on the vacuum manifold until the filter was dry (approximatively 3 minutes). The plate was then removed from the vacuum manifold, 20-25 µl of distilled water was added to the purified product and the plate was left at room temperature for 20 minutes. To dissolve completely the DNA, the water was mixed by pipetting up and down several times and the resulting purified DNA was transferred into a new plate. The purified DNA was used immediately after or stored at -20 °C for later use.

5.2.7.3 Sequencing reaction

The sequencing reaction was set up in PCR plates (Starlab) and sealed with Thermo Seal Film (Starlab). In a PCR plate, 2 µl of purified PCR product was mixed with 1 µl of BigDye Terminator 3.1 (Applied Biosystems), 4 µl of BigDye Terminator sequencing buffer, 1 µl of 10 µM forward (or reverse) primer (see Table 5.14) and 11 µl of distilled water. The plate was placed in a thermocycler and incubated at 96 °C for 5 minutes and 25 cycles of 96 °C for 10 seconds; 50 °C for 5 seconds and 60 °C for 4 minutes, and hold at 4 °C. Products of BigDye termination sequencing reactions were used immediately after or stored at -20 °C for later use.
5.2.7.4 Sequencing clean-up

Prior sequence data collection, the sequencing reaction was cleaned-up using ethanol precipitation (see section 2.2 - DNA purification). However, the purified pellet was kept dry and stored at -20 °C until ready to put the reaction on the sequencer.

5.2.7.5 Sequence electrophoresis

The dry pellet was resuspended in 10 µl of distilled water and 10 of µl HiDi™ formamide (Applied Biosystems). The mixture was denatured at 96 °C for 2 minutes, and then cooled down on ice protected from lights. The mixture was then ready to be placed in an ABI 3730 capillary sequencer (Applied Biosystems) for sequencing electrophoresis.

5.2.7.6 Sequence data analysis

Sequencing data were analysed using Sequencher v5.4.5 (Gene Codes). In Sequencher, the sequencing reads were first aligned against an appropriate gene specific reference file (GenBank file) downloaded from NCBI and the discrepancies were detected and highlighted by the software.

5.3 Results

5.3.1 Methylation profiling of BWS_IC2 patients conceived with ART

Among the 78 BWS_IC2 patients, 13 were conceived with the help of ART: 2 were
following ovarian stimulation (OS) only, 3 were following in vitro fertilisation (IVF) and 8 were following intra-cytoplasmic sperm injection (ICSI). ART was not reported in the 65 other patients. Additionally to LOM at \textit{KCNQ1OT1}:TSS-DMR, methylation profiling of ART patients revealed MLID in 1 out of 2 (50%) patient conceived with the help of OS, 2 out of 3 (67%) patient conceived with the help of IVF and 6 out of 8 (75%) patients conceived with the help of ICSI. Although 1 patient conceived with OS had MLID, I didn’t draw a conclusion regarding the prevalence of MLID in this group as a limited number of cases were studied. IVF and ICSI follow similar clinical procedures and therefore were grouped. Taken together, MLID was detected in 8 out of 11 (73%) patients conceived with the help of IVF/ICSI. Out of the 65 naturally conceived patients 22 had MLID (34%) (Figure 5.3). MLID frequency in the IVF/ICSI group was compared to the MLID frequency in the no ART reported group. Statistical analysis using Fisher’s exact test showed that ART procedures (i.e. IVF and ICSI) were significantly associated with MLID in BWS_IC2 (p=0.021) (Figure 5.4). Finally, the frequency of epimutation at imprinted DMRs in IVF/ICSI MLID group was compared to the frequency of epimutation at imprinted DMRs in the no ART reported MLID group. Whilst some DMRs appeared to be more frequently epimutated in one group than in the other (e.g. \textit{MEST}:alt-TSS-DMR: 25% in ART, 9% in non-ART; \textit{FAM50B}:TSS-DMR: 12.5% in ART, 32% in non-ART) no significant association between the IVF/ICSI procedures and epimutations at imprinted DMRs was found (p-value threshold of 0.05 using Fisher’s exact test) (Figure 5.5).
**Figure 5.3:** ART and genomic imprinting methylation in BWS_IC2.
Multilocus imprinting disturbances and ART. Epimutations distribution at 46 imprinted DMRs (y-axis) in BWS_IC2 patients conceived with or without the help of assisted reproductive technology (x-axis). Red, loss-of-methylation; green, gain-of-methylation; black: no epimutations. OS, n=2; IVF, n=3; ICSI, n=8, ART not reported, n=65.
**Figure 5.4: Significance of ART of MLID epigenotype in BWS.IC2 patients.**

Comparison of the number of BWS.IC2 patients conceived with or without the help of ART (IVF/ICSI) and with or without MLID. ART, assisted reproductive technology; MLID, multilocus imprinting disturbances; statistical analysis was done using Fisher’s exact test.
Figure 5.5: ART is not associated with a specific epimutation at imprinted DMRs in BWS_IC2 cohort

Comparison of imprinted DMRs epimutation frequency in 8 BWS_IC2 patients with MLID conceived with ART (IVF/ICSI) and in 22 BWS_IC2 patients with MLID conceived naturally. ART, assisted reproductive technology.
5.3.2 Genetic investigation of BWS_IC2 patients

WES was performed on 42 BWS_IC2 patients (25 with MLID and 17 with no MLID). The sequencing read depth was assessed in all samples and at least 87% of bases in 41 samples and at least 80% of bases in 1 sample were sequenced at least 20 times (i.e. 20x) (Figure 5.6). This meant a reliable bases calling, hence a good certainty for further variants analysis. In total, WES analysis revealed 18 rare genetic alterations that were predicted to be damaging for the protein function or splicing and therefore might potentially be related to the presence of MLID in the BWS patients (Table 5.15).

Figure 5.6: Distribution of the sequencing depth of read across the 42 BWS_IC2 individuals.
Table 5.15: Summary of interesting variants found by whole-exome sequencing.
5.3.2.1 Analysis of methylation associated genes

I analysed WES data for the 138 genes identified as important in or associated with methylation pathways (see Table 5.12 in section 5.2.6.4 - Variant call analysis) and interesting variants were validated by Sanger sequencing wherever possible. An average of the proportion of exons successfully sequenced for each gene for all patients was calculated. Out of 138, 6 genes had a very poor to average (<60%) proportion of exons sequenced and 132 had a high to complete coverage (>60%) (Figure 5.7).

No stop-gain affecting a gene associated with methylation pathways were found. A single frameshift insertion in the thymine DNA glycosylase (TDG) gene (variant 1: NM_003211, c.965dupA, p.E322fs, rs764159587) was found in 33 patients; all were heterozygotes for the variant. Additionally, 6 heterozygotes splicing variants in TDG (variant 2: NM_003211, c.964+1G>A; variant 3: NM_003211, c.964+2T>G, rs760400700; variant 4: NM_003211, c.408+1G>A, rs780554309; variant 5: NM_003211, c.793-2A>C; variant 6: NM_003211, c.793-1G>T; variant 7: NM_003211, c.1090+1G>T, rs762057949) affecting in total 28, 32, 4, 13, 14 and 11 individuals respectively were found. Finally, three non-synonymous variants occurring in the PHD finger protein 10 (PHF10) gene were found; variant 8 (NM_018288 c.T595G, p.Y199D, rs562092150) was present in two individuals. These two individuals, as well as one additional patient, also carried variant 9 (NM_018288, c.C646T, p.R216W, rs77919800) and variant 10 (NM_018288, c.T640G, p.L214V, rs144595699). All variants in each patient were in a heterozygous
state. Sanger sequencing of an independent PCR product for variant 1 in 31 of 33
individuals, variant 2 in 28 of 28 individuals, variant 3 in 31 of 32 individuals, variant
4 in 4 of 4 individuals, variant 5 in 13 of 13 individuals and variant 6 in 14 of 14
individuals in the TDG gene and for variant 8 in 2 of 2 individuals, variant 9 in 3 of 3
individuals and variant 10 in 3 of 3 individuals in the PHF10 gene did not successfully
confirmed the results obtained from WES for those patients (Figure 5.8 and Table
5.15). Further investigation with the sequencing facility revealed that those variants
were also commonly detected in other sequencing project, hence suggesting that
they were sequencer specific sequencing artefacts. Sanger sequencing of variant 7
was not attended
Figure 5.7: Proportion of the coverage of the exons of gene associated with the methylation pathway.

Proportion of exons of 138 gene associated with methylation pathway successfully sequenced. Toward red, low coverage; toward green, high coverage.
Figure 5.8: Validation of the variants found in gene associated with the methylation pathway.

Sanger sequencing of an independent PCR product of variant 1, 2, 3, 4, 5 and 6 (TDG) and variant 8, 9 and 10 (PHF10). Top, patient sequencing trace; bottom, control individual sequencing trace; blue inverted triangle, variant position.

5.3.2.2 Analysis of maternal effect genes

I analysed WES data for the 63 genes identified as maternal-effect genes in mammals (see Table 5.11 in section 5.1.2.1 - Maternal and maternal-zygotic genes). An average of the proportion of exons successfully sequenced for each gene for all patients was calculated. Out of 63, 7 genes had a very poor to average (<60%)
proportion of exons sequenced and 56 had high to complete coverage (>60%) 
(Figure 5.9).

No stop-gain, frameshifting indel or splicing mutations were found in maternal-effect 
gene. A non-synonymous variant in the basonuclin 1 (BNC1) gene (variant 11: 
NM_001301206, c.C1204A, p.P402T) was found in a heterozygous state in one 
patient with GOM at ZDBF2/GPR1:IG-DMR and LOM at IGF2R:Int2-DMR, 
MEST:alt-TSS-DMR, FANCC:Int1-DMR, KCNQ1OT1:TSS-DMR, SNU13:alt-TSS-
DMR. A non-synonymous variant in the CCCTC-binding factor (CTCF) gene (variant 
12: NM_001191022, c.C943T, p.P315S, rs145727304) was found in a heterozygous 
state in one patient with LOM at DIRAS3:TSS-DMR, DIRAS3:Ex2-DMR, 
FAM50B:TSS-DMR, FANCC:Int1-DMR, KCNQ1OT1:TSS-DMR, IGF1R:Int2-DMR, 
L3MBTL1:alt-TSS-DMR, GNAS-AS1:TSS-DMR and SNU13:alt-TSS-DMR. Finally, 
three non-synonymous variants in the NLRP2 gene (variant 13: NM_001174082, 
rs200375320; variant 15: NM_001174082, c.G2335A, p.A779T, rs117066658) were 
found. One patient was a homozygous for variant 13 and had LOM at FAM50B:TSS-
DMR and KCNQ1OT1:TSS-DMR. One patient was heterozygous for both variant 14 
and variant 15 and had LOM at KCNQ1OT1:TSS-DMR, L3MBTL1:alt-TSS-DMR and 
SNU13:alt-TSS-DMR. Finally, one patient was heterozygous for variant 15 only and 
had LOM at DIRAS3:TSS-DMR, DIRAS3:Ex2-DMR, FAM50B:TSS-DMR, 
KCNQ1OT1:TSS-DMR, ZNF331:alt-TSS-DMR1, ZNF331:alt-TSS-DMR2,
NNAT: TSS-DMR, SNU13:alt-TSS-DMR. For which DNA was available, Sanger sequencing of an independent PCR product for variant 11 in 1 individual (BNC1), variant 12 in 1 individual (CTCF), and variant 15 in 1 of 2 individual (NLRP2) successfully confirmed the results obtained from WES (Figure 5.10 and Table 5.15). The Sanger sequencing of one individual with variant 13, one individual with variant 14 and one of two individual with variant 15 was not attended due to the lack of DNA for those patients.
**Figure 5.9:** Proportion of the coverage of the exons of maternal-effect genes.

Proportion of exons of 63 maternal-effect genes successfully sequenced. Toward red, low coverage; toward green, high coverage.
**Figure 5.10: Validation of the variants found in maternal-effect genes.**

Sanger sequencing of an independent PCR product of variant 11 (BNC1), 12 (CTCF) and 15 (NLRP2). Top, patient sequencing trace; bottom, control individual sequencing trace; blue inverted triangle, variant position.

**5.3.2.3 Analysis of imprinted genes**

I analysed WES data for the 76 genes identified as imprinted or potentially imprinted (see Table 5.13 in section 5.2.6.4 - Variant call analysis). An average of the proportion of exons successfully sequenced for each gene for all patients was calculated. Out of 76, 8 genes were not sequenced, 29 genes had a very poor to average (<60%) proportion of exons sequenced and 39 had high to complete coverage (>60%) (Figure 5.11).

A stop-gain variant (variant 16: NM_020923, c.T350G, p.L117X, rs755604527) in a individual in heterozygous state with LOM at *FAM50B:TSS-DMR* and *KCNQ1OT1:TSS-DMR* and a non-synonymous variant (variant 17: NM_020923,
c.G6224T, p.R2075M, rs36095066) in another individual in heterozygous state with LOM at FANCC:Int1-DMR, INPP5F:Int2-DMR, KCNQ1OT1:TSS-DMR, IGF1R:Int2-DMR, GNAS-AS1:TSS-DMR, GNAS-XL:Ex1-DMR, GNAS-AB:TSS-DMR were found in the zinc finger DBF-type containing 2 (ZDBF2) imprinted gene. Finally, a non-synonymous variant (variant 18: NM_000218, c.C1046T, p.S349L) in the potassium voltage-gated channel subfamily Q member 1 (KCNQ1) imprinted gene was found in a heterozygous state in one patient with LOM at FANCC:Int1-DMR, INPP5F:Int2-DMR, KCNQ1OT1:TSS-DMR, IGF1R:Int2-DMR, GNAS-NESP:TSS-DMR, GNAS-AS1:TSS-DMR and GNAS-XL:Ex1-DMR. The Sanger sequencing of an independent PCR product for variant 16 and 17 (ZDBF2) and variant 18 (KCNQ1) successfully confirmed the results obtained from the WES data (Figure 5.12 and Table 5.15)
**Figure 5.11:** Proportion of the coverage of the exons of imprinted genes.

Proportion of exons of 76 imprinted or potentially imprinted gene successfully sequenced. Toward red, low coverage; toward green, high coverage.
**Figure 5.12:** Validation of the variants found in imprinted genes.

Sanger sequencing of an independent PCR product of variant 16 and 17 (ZDBF2) and variant 18 (KCNQ1). Top, patient sequencing trace; bottom, control individual sequencing trace; blue inverted triangle, variant position.

### 5.4 Discussion

#### 5.4.1 ART and MLID

Among the 11 patients conceived with the help of ART (IVF/ICSI), 8 had MLID (73%). This was significantly higher (p=0.021, Fisher’s exact test) than the frequency observed in the no ART reported group in which 22 out 65 (34%) patients had MLID. This result corroborates with previous studies in which a significant association between ART and MLID in BWS_IC2 were found (Lim et al. 2009; Hiura et al. 2012; Tee et al. 2013) (it should be noted that the cohort I studied may overlap with those in the Lim et al. and Tee et al. reports). However, my observation also contrasts with other studies in which ART was not reported to be associated with MLID in BWS_IC2 (Rossignol et al. 2006; Azzi et al. 2009).
The number of disturbances reported at a given imprinted DMRs were variable between ART and no ART reported BWS patients. Some DMRs appeared to be more frequently epimutated in one group than in the other but statistical analysis did not reveal a significant association between the use of ART and the disturbances at specific imprinted DMRs. In conclusion, with the exception of \textit{KCNQ1OT1}:TSS-DMR, ART seems to disrupt genomic imprinting in a random manner.

Taken together, my data suggests that ART is associated with MLID in BWS IC2. However, it remains to be understood if the imprinting disturbances are the direct consequences of ART alone, the products of other factors such as the parental infertility or the age of the individuals, or the results of a combination of all. If ART is involved, how does it disrupt the genomic imprint and which step of the procedure is involved in this mechanism? The ovarian stimulation and the use of an immature oocyte may affect the establishment of DNA methylation of genomic imprinting whilst the fertilisation, the embryo culture and the uterus implantation steps may disrupt the maintenance of this epigenetic mechanism (Hiura et al. 2014). The genome-wide methylation analysis using next-generation sequencing and microarrays of animal models and extensive patient cohorts will hopefully provide a better picture of the extent and implications of ART-induced imprinting changes in human disorders.
5.4.2 Trans-mechanism involving a gene associated with methylation pathway

WES analysis revealed rare variants that have the potential to alter protein function or splicing of the TDG and PHF10 genes, both described to be associated with the methylation pathway. Though the heterozygous state of the BWS with MLID individuals indicated that the TDG and PHF10 potential pathogenic variants raised the possibility of being responsible for genomic imprinting disturbances following an autosomal dominant mode of inheritance, the identified candidate genetic alterations were not validated by Sanger sequencing and were found to be sequencer specific sequencing artefacts and so were not considered further.

In summary, I did not identify candidate mutations in a methylation associated gene that is likely to be responsible for BWS phenotype and MLID in our cohort. To date, no additional candidate mutations in one of these genes have been associated with MLID and BWS. Therefore, with the exception of ZFP57 in a subset of TNDM1 patients, causative pathogenic mutations of such genes are very rare events in the aetiology of imprinting disorders with MLID. However, it should be noted that the identification of pathogenic mutations is complicated by (i) the increasing burden that represents the analysis and interpretation of WES and (ii) some type of mutation, such as internal exon deletion, might not be detected by WES. It has been established that a sequenced exome of an individual can harbour between 20,000 to 50,000 variants. Of these, several steps of filtering criteria to remove false-positive
calls (e.g. reads quality, out of coding sequence, synonymous coding variants, common in dbSNP) and prioritisation strategy (e.g. non-synonymous variants, splicing variants) will reduce the list of potential candidates and retain approximatively 150 to 500 variants. Of these, further analysis and interpretation in the context of the phenotype will be undertaken to hopefully identify the potential true pathogenic variants (Gilissen et al. 2012; Robinson et al. 2014). It should be noted that this would be true on the assumption that the pathogenic mutations was not filtered out during upstream filtering and prioritisation processes, such as a patient with heterozygous missense substitution might be undetected compound heterozygotes.

5.4.3 Trans-mechanisms involving maternal-effect genes

I identified five rare variants that may be responsible for BWS phenotype and MLID in the maternal-effect genes BNC1, CTCF and NLRP2. These variants were found in four distinct individuals with MLID and all variants were predicted to be protein damaging. One of the five variants (NLRP2, variant 13) was found in a homozygous state whilst the four others (BNC1, variants 11; CTCF, variant 12; NLRP2, variants 14 and 15) were found in a heterozygous state. Finally, all these variants were reported as rare in ESP and ExAC database (variants 11, not reported in ESP and ExAC; variant 12, ESP: T=57/C=12939, ExAC: T=258/C=120954; variant 13: ESP: G=87/C=12917, ExAC: G=690/C=118704; variant 14: ESP: T=1/C=13005, ExAC: T=7/C=121367; and variant 15: ESP: A=138/G=12868, ExAC: A=1126/G=120172).
Taking into consideration the rarity of the variants and their potential damaging effects on the protein, I hypothesised that they may have been pathogenic and associated to BWS and MLID.

To further investigate the potential significance of these findings it would be necessary to sequence the mothers of these cases to see if they were homozygous for the variants (or were compound heterozygotes). If this were to be found then it could be hypothesised that the genetic variants found in $BNC1$, $CTCF$ and $NLRP2$ might predispose to BWS and MLID in a maternal-effect inheritance manner. Additionally, the absence of the variants in the fathers would further support this hypothesis.

$BNC1$ encodes for a zinc-finger protein found in abundance in oocytes and regulates the rRNA transcription. $BNC1$ deficiency in mouse oocytes perturbs both RNA polymerase I- and II- mediated transcription and it affects the morphology and biochemistry of oocytes. The fertilised $Bnc1$-deficient eggs failed to develop beyond the two-cell stage (Ma et al. 2006; Kim and K.-A. Lee 2014). If the maternal-effect inheritance in this patient could be confirmed, this would be the first description of $BNC1$ in BWS aetiology. Moreover, the additional epimutations associated with the patient may also indicate that the variant is likely to cause MLID and consequently could have an unrecognised role in genomic imprinting.

$CTCF$ encodes for a highly conserved transcription factor with 11 zinc-finger DNA binding domains. Using different combinations of zinc fingers domains to bind DNA,
CTCF can function as a transcriptional insulator, repressor, or activator, depending on the context of the binding site (Ohlsson et al. 2001). Depletion of Ctcf in fertilised Ctcf-deficient eggs arrest at the morula stage or at various stages prior to morula compaction and only a very small proportion of the embryos are able to develop to the blastocyst stage (Fedoriw et al. 2004; Wan et al. 2008). CTCF is critical for regulating genomic imprinting at multiple loci including the Igf2/H19 locus. The protein binds the unmethylated maternal allele of Igf2/H19 DMR and inhibits the maternal Igf2 gene expression via enhancer blocking activity (Bell and Felsenfeld 2000; Hark et al. 2000). The depletion of Ctcf in mouse oocytes also results in the hypermethylation of Igf2/H19 DMR during oocyte growth, indicating that CTCF is required to maintain the unmethylated status of the locus (Fedoriw et al. 2004). Similarly, the maternal transmission of microdeletion removing CTCF binding sites within H19/IGF2:IG-DMR is associated with hypermethylation of the locus, biallelic silencing of H19, biallelic expression of IGF2 and, in fine, to BWS phenotype (Sparago et al. 2004; Prawitt, Enklaar, Gartner-Rupprecht, et al. 2005; Sparago et al. 2006).

As seen previously (see section 5.1.2.4 - NLRP2), maternal-effect mutation in NLRP2 was previously reported to be associated with the BWS phenotype and MLID (Meyer et al. 2009). Hence sequencing of parental samples would be indicated to determine if the variants were of paternal or maternal inheritance and if the mother was homozygous or compound heterozygous.
Further studies, including sequencing of the parents of the probands, are necessary to determine a more precise role of these variants in the molecular aetiology of MLID and BWS.

5.4.4 Trans-mechanism involving imprinting gene network disruption

The complex co-regulation of multiple imprinted genes and the existence of imprinted gene networks (Varrault et al. 2006; Patten et al. 2016; Soellner et al. 2016) suggested that the genetic insults of a member of a network might have dramatic consequences on not only the imprinted gene itself but also on the other imprinted gene members of the network. I hypothesised this may act at the methylation level and may be involved in the abnormal imprinted epigenotype seen in BWS patients with MLID.

The exons of 37 out of the 76 imprinting genes were not sequenced or had a poor coverage. This can be explained by the fact that a majority of these genes are non-protein coding genes. The library preparation method used for WES was not designed to capture these non-protein coding regions. However, despite the lack of sequencing coverage I identified one stop-gain (not reported in ESP and ExAC) and one non-synonymous (ESP: T=77/G=11847; ExAC: T=756/G=118228) rare variants in the ZDBF2 imprinted gene and one rare non-synonymous (not reported in ESP and ExAC) variant in the KCNQ1 imprinted gene (KCNQ1 variant did not overlap with KCNQ1OT1 genomic region). Mutations in KCNQ1 have been associated with
hereditary long QT syndrome 1 (Q Wang et al. 1996), short QT syndrome 2 (Bellocq et al. 2004) and familial atrial fibrillation (heart conditions) (Y.-H. Chen et al. 2003), and Jervell and Lange-Nielsen syndrome (a hearing loss condition) (Neyroud et al. 1997) but not with an imprinting disorder. The three variants were reported as rare and were predicted to be damaging for the protein by several scoring databases, hence suggesting they might be pathogenic.

In placental tissue, the imprinted expression of \textit{ZDBF2} is expressed from the paternal allele, although stochastic maternal expression of \textit{ZDBF2} is detected. In somatic tissue, \textit{ZDBF2} paternal expression is maintained whilst maternal \textit{ZDBF2} is silenced. I hypothesised that the stop-gain and the non-synonymous dominant variant could affect the \textit{ZDBF2} paternal gene expression, which in return would (i) induce aberrant methylation at other imprinting regions through \textit{trans}-imprinting mechanism and (ii) result in aberrant embryonic development. To support my hypothesis, the author of a recent study has found that the knockout of \textit{Liz} (Long isoform of \textit{Zdbf2}) in mice results in the absence of \textit{ZDBF2} expression and in an abnormal growth phenotype (methylation profiling at other imprinting DMRs was not performed) (Greenberg et al. 2017). However, the authors found that the lack of \textit{ZDBF2} activation was associated with growth restriction rather than growth promotion, which \textit{in fine} was not compatible with my hypothesis.

\textit{KCNQ1} is located within \textit{KCNQ1OT1}:TSS-DMR region, which is the main locus involved in BWS aetiology. I hypothesised that the non-synonymous variant could
be a dominant maternally-inherited mutation affecting a regulatory sequence
required for the cis methylation of the maternal KCNQ1OT1:TSS-DMR IC.
Consequently, this would primarily result in maternal KCNQ1OT1:TSS-DMR
hypomethylation, biallelic expression of KCNQ1OT1, bisilencing of CDKN1C and, in
fine, to BWS phenotype. Furthermore, through a mechanism that would by-pass the
KCNQ1OT1-mediated transcriptional repression activity, the non-synonymous
mutation could result in the translation of an altered KCNQ1 protein that would
disrupt genomic imprinting methylation at other loci. Although it was an interesting
hypothesis, several lines of evidence contradicted this hypothesis. Firstly, it seems
that the variant does not affect a regulatory element required for the methylation of
the IC. Secondly, the mechanisms that would explain the by-pass of KCNQ1OT1-
mediated gene silencing are yet to be identified. Finally, there is a lack of evidence
suggesting that KCNQ1 (as opposed to KCNQ1OT1) has a role in imprinting.

In conclusion, further investigations are required to further characterise these
variants. However, the current line of evidence suggests that the ZDBF2 and KCNQ1
variants are non-pathogenic and consequently are not involved in the molecular
aetiology of BWS and MLID.
Chapter 6

Investigations of a novel candidate gene for BWS
6.1 Introduction

6.1.1 UHRF1, a critical epigenetic modifier

Ubiquitin-like with PHD and RING finger domains 1 (UHRF1), also known as inverted CCAAT box protein of 90 kDa (ICBP90) in human and NP95 in mouse, is a nuclear multi-structural domain and functional protein. UHRF1 has been shown to promote cell proliferation in various cancers, including breast, colon, prostate, and lung cancer (Bronner et al. 2013). In these, it was demonstrated that the over expression of UHRF1 contributed to the silencing of tumour suppressor genes, the inhibition of the DNA repair pathway, the tumour growth and metastasis (Bronner et al. 2013). The down-regulation of UHRF1 in human colon colorectal carcinoma cells was shown to induce cell cycle to arrest at the G1/S transition, hence suggesting that UHRF1 is required for progression in the cell cycle (Y. Arima et al. 2004). UHRF1 is also known to be essential for the maintenance of DNA methylation through recruiting DNMT1 to the replication fork in S phase of the cell cycle (Bostick et al. 2007; Sharif et al. 2007). The depletion of UHRF1 in human cells, mouse embryonic stem cells and in zebrafish results in the reduction of global DNA methylation and at locus specific DNA methylation, including IGS-rDNA (intergenic spacer of ribosomal DNA) and imprinted loci such as H19, Kcnq1ot1, Nnat, Igf2r, Dlk1 (Bostick et al. 2007; Sharif et al. 2007; Tittle et al. 2011; Rothbart et al. 2012; Qi et al. 2015). However, the transient expression of wild-type UHRF1 in knockdown UHRF1 human cells, or in Uhrf1 null embryonic stem cells, was reported to restore the loss-of-methylation associated with UHRF1 knockdown (Rothbart et al. 2012; X. Liu et al.)
2013). This also includes methylation rescue in only a small subset of mouse imprinted loci, \textit{H19, Nnat, Dlk1} (Qi et al. 2015). Consequently, this line of evidences demonstrated that UHRF1 is a key epigenetic modifier.

5

UHRF1 harbours at least five functional domains: a N-terminal ubiquitin-like domain (UBL), followed by a tandem tudor domain (TTD), a plant homeodomain (PHD), a SET and RING-associated (SRA) domain, and a C-terminal really interesting new gene (RING) domain (Figure 6.1). The SRA domain preferentially binds to hemi-methylated CpG, which during semiconservative replication of DNA recruits DNMT1 at the replication fork to copy the methylation pattern onto the daughter strand (Sharif et al. 2007). This process may also involve the RING domain through the ubiquitylation of H3K23 and H3K18, which creates ubiquitylated docking sites for DNMT1 (Nishiyama et al. 2013; Qin et al. 2015). The isolated TTD domain was shown to recognise and bind preferentially to H3K9me3 (Nady et al. 2011) whilst the isolated PHD domain binds to the unmodified extreme H3 N terminus, and more particularly to unmodified histone H3 arginine residue 2 (Hu et al. 2011; Rajakumara et al. 2011). Through the existence of functional cooperation in reading histone modifications, both TTD and PHD domains were suggested to play a role in the DNA methylation maintenance mechanism. Interestingly, whilst some studies suggested that these domains were essential for the maintenance of DNA methylation (Rothbart et al. 2012; Rothbart et al. 2013), others suggested that both domains promoted DNA methylation maintenance but were not essential to it (X. Liu et al. 2013; Zhao et al. 2016). The crosstalk between both epigenetic marks (i.e. DNA methylation and
histone modifications) is to date not well understood. However, a few recent studies had started to decipher possible mechanisms that might link both epigenetic marks. In a recent report, the existence of a poly basic region (PBR) between the SRA and RING domains was suggested. The PBR was shown to facilitate the recognition of semi-methylated DNA by the SRA (Fang et al. 2016) and to mediate the interaction of the TTD domain with H3K9me3 and the interaction of the PHD domain with H3K4me0 (Gelato et al. 2014; Fang et al. 2016). This mechanism was dependant on the presence or absence of phosphatidylinositol 5-phosphate (PIP5). The binding of PIP5 to the PBR was shown help to stabilise UHRF1 and then DNMT1 to genomic regions containing H3K9me3 and undergoing replication. In the absence of PIP5, UHRF1 and then DNMT1 stabilise at genomic regions undergoing replication that contain H3K4me0 (Gelato et al. 2014).

**Figure 6.1: Domain structure of UHRF1.**

**UBL, ubiquitin-like domain; TTD, tandem tudor domain; PHD, plant homeodomain; SRA, SET and RING-associated; RING, really interesting new gene; PBR, polybasic region. (From Gelato et al. 2014).**
6.1.2 Case Report

Two siblings born to consanguineous parents were diagnosed with BWS. Previous molecular genetic analysis had demonstrated that both siblings had loss-of-methylation at KCNQ1OT1:TSS-DMR. The methylation profiling at imprinted loci other than KCNQ1OT1:TSS-DMR was not performed and is currently under investigation. Whole-exome sequencing was performed prior to my study and the data was inspected to identify candidate mutations that might cause a trans-imprinting defect. No candidate mutations were identified in NLRP2, NLRP5 (a heterozygous candidate mutation would be present in the siblings if the mother was a homozygote or compound heterozygote for mutations in these genes) or in ZFP57. In view of the parental consanguinity a candidate mutation would be predicted to be homozygous in both siblings and therefore rare homozygous variants were identified in the exome data and evaluated as potential candidate mutations.

It was noted that both siblings had a novel homozygous missense variant in UHRF1 (Figure 6.2-a) (the variant was not present in ESP nor in ExAC databases). The guanine to cytosine substitution at position 4954693 of chromosome 19 (Build GRCh38) resulted in an amino acid substitution of lysine to asparagine at position 667 (NM_001048201.2, c.2237G>C, p.K667N). The rare non synonymous missense variant was predicted to be protein damaging by Mutation Taster (score: 0.999) and SIFT (score: 0.015) and probably damaging by PolyPhen-2 (score: 0.991) (Figure 6.2-b). Additionally, the affected amino acid was found to be conserved across
species including mice, zebrafish and *Xenopus tropicalis* (Figure 6.2-c). Finally, the variant is located near to the polybasic region (PBR) within UHRF1. As discussed previously, this region may also be involved in the faithful DNMT1-mediated DNA methylation maintenance mechanism during replication.

![Image](image.png)

**Figure 6.2: Case report of two siblings with BWS_IC2 and a homozygous mutation in the UHRF1 gene.**

(a) Family pedigree of the two affected siblings. Both siblings have a novel homozygous missense variant affecting UHRF1 (g.49544693G>C, GRCh38). (b) Deleterious scores obtained with Mutation taster, PolyPhen-2, SIFT and PROVEAN. (c) The affected amino acid (lysine 667 in human) is conserved across species.

### 6.1.3 Aim

To date several trans-acting factors have been described in the molecular aetiology of familial MLID disorders including *KHDC3L* (also known as *C6ORF221*), *NLRP2*, *...*
NLRP5, NLRP7 and ZFP57 (see section 5.1.2 - Trans-imprinting defect and MLID). However, an underlying genetic defect responsible for MLID in BWS is very rare and only two of these genes have been linked to MLID in BWS (NLRP2 and NLRP5) (Meyer et al. 2009; Docherty et al. 2015). In this chapter I describe investigations into the potential pathogenicity of a homozygous variant in the UHRF1 gene found in two siblings with BWS_IC2. It was suggested that the missense substitution in UHRF1 p.K667N might partially or fully disturb the function of the protein and compromise normal DNA methylation maintenance mechanisms at imprinted DMRs. If this hypothesis was correct then UHRF1 would be the third described trans-acting factor responsible for methylation disturbances associated with BWS. However in the absence of additional families with a similar phenotype the candidacy of UHRF1 remained unproven. Therefore it was decided to investigate, in vitro, whether the missense substitution could be shown to compromise UHRF1 function. I therefore proceeded to assess whether wild-type and mutant UHRF1 could rescue methylation abnormalities in human HeLa cells in which UHRF1 gene expression had been knocked down by shRNA lentiviral transduction.

6.2 Materials and methods

6.2.1 Cells

HeLa cells knockdown for UHRF1 (HeLa shUHRF1) and control HeLa cells (HeLa shLuc) were a generous gift from Dr. Scott Rothbart (Rothbart et al. 2012). The shRNA were designed to target the 3' UTR region of the UHRF1 and Luciferase
gene, hence allowing subsequent rescue experiments. The shRNA used to
knockdown targeted genes were obtained from the RNAi Consortium and were used
following standard TRC lentivirus production and infection protocols (Table 6.16).

Table 6.16: shRNA clones used in HeLa cells.
(From Rothbart et al. 2012).

<table>
<thead>
<tr>
<th>shRNA</th>
<th>TRC clone name</th>
<th>TRC clone ID</th>
<th>Target sequence</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>control</td>
<td>promegaLuc_221s1c1</td>
<td>TRCN0000072246</td>
<td>CAAATCACAGAATCGTGAT</td>
<td>3' UTR</td>
</tr>
<tr>
<td>UHRF1</td>
<td>NM_013282.3-3485s21c1</td>
<td>TRCN0000273256</td>
<td>GCCCTTGATTCGTTCTTTCTT</td>
<td>3' UTR</td>
</tr>
</tbody>
</table>

6.2.2 Plasmids

The plasmids pRP+EV (i.e. empty vector) and pRP+WT (i.e. full-length human cDNA
of wild-type UHRF1) were designed and ordered from Vector Builder. The pRP
expression vector contains a cytomegalovirus (CMV) promoter, a cDNA sequence
coding for the enhanced green fluorescent protein (EGFP) marker, and the C-
terminus of the protein of interest (UHRF1) has a dual FLAG and HA tags.

The plasmids pTag-2C+EV (i.e. empty vector), pTag-2C+WT (i.e full-length human
cDNA of wild-type UHRF1) and pTag-2C+Y188A (i.e full-length human cDNA of
UHRF1 Y188A mutant) were a generous gift from Dr. Scott Rothbart (Rothbart et al.
2012). The pTag-2C expression vector contains a CMV promoter and the N-terminus
of the protein of interest (UHRF1) has a FLAG tag.
## 6.2.3 Maintenance of adherent cells

Complete growth medium, PBS 1X and Trypsin 1X solutions were warmed up at 37°C in water bath for approximately 30 minutes prior starting procedure. In a class II cabinet growth medium was removed from cells container (i.e. Flasks (StarLab)/Petri Dishes (Corning)). In order to completely remove growth medium the cells were rapidly washed by adding a small quantity of sterile PBS 1X and by gently rocking the container. Supernatant was then discard from container. To detach adherent cells from the bottom of the container a small amount of Trypsin 1X - EDTA 0.02% solution was added to the container. Cells were then placed into CO₂ incubator at 37 °C until they start to detach. Once the cells are detached the container was transferred back to class II cabinet and complete growth medium was added to the cells to stop the reaction of the Trypsin. To remove dead cells and Trypsin the cell suspension was transferred in a falcon tube and then centrifuged for 5 minutes at 2000 rpm. Supernatant was removed and cell pellet was resuspended carefully but thoroughly in an appropriate volume of complete growth medium. Cell counting was performed on LUNA™ Automated Cell Counter (logos biosystems). 10 µl of sample was mixed with 10 µl of 0.4% trypan blue stain. 10~12 µl of the mixture was loaded into the chamber port on the LUNA™ counting slide, slide was inserted into slide port of the counter, focus was done if necessary and live cells (i.e. bright centres and dark edges while dead cells have blue colour with no bright centres) were counted. Once the cell concentration was determined, the required volume of cell suspension was transferred into a new container and an appropriate volume of fresh complete growth
medium was added to the container. The new container was then placed into CO₂ incubator at 37 °C.

6.2.4 Transient transfection

One day prior to transfection HeLa shLuc and HeLa shUHRF1 cells were seeded in sterile 6 cm Petri Dish (Corning) (see section 6.2.3 - Maintenance of adherent cells) in normal growth medium for an expected cells at confluence of 70-80% the next day. At the day of transfection, growth medium was removed from Petri Dish and 9 ml of fresh antibiotic free growth medium was gently added to the plate. For 1 Petri Dish plate, transfection reagents were prepared as followed: TransIT-LT1 Reagent (Mirus) was put at room temperature and vortexed gently before using. In a 1.5 ml sterile tube (Eppendorf), 250 µl of Opti-MEM I Reduced-Serum Medium (Thermo Fisher Scientific) was mixed gently with 1.5 µl of 1 µg/µl plasmid DNA (one of pRP+EV; pRP+WT; pTag-2C+EV; pTag-2C+WT; pTag-2C+Y188A). 3-9 µl of TransIT-LT1 Reagent was added to the tube and mixed completely by pipetting. The diluted TransIT-LT1 Reagent was incubated at room temperature for 15-30 minutes. The resulting TransIT-LT1 Reagent:DNA complexes were added drop-wise at different area of the plate. The Petri Dish was rocked gently to distribute the TransIT-LT1 Reagent:DNA complexes evenly and then put into CO₂ incubator at 37 °C. Visualisation and analysis of transfected cells were performed 48 hours after transfection.
6.2.5 Quantitative PCR

6.2.5.1 RNA extraction from adherent cells

Total RNA of adherent cells was extracted using RNeasy Mini Kit (Qiagen). Cells were trypsinized (see section 6.2.3 - Maintenance of adherent cells) and pelleted in falcon tube by centrifugation at 500 x g for 5 minutes. The cells were washed with 1.5 ml of PBS 1X and centrifuged at 500 x g for 5 minutes at 4 °C. To lyse the cells, supernatant was removed and 350 µl of Buffer RLT was added directly to the cells pellet. Pellet was resuspended by vortexing 4-6 seconds and the lysate was directly pipet into a QIAshredder spin column in a 2 ml collection tube and centrifuge at 10,000 x g during 2 minutes for homogenisation. 350 µl of freshly prepared 70% ethanol was added to the homogenised lysate and the mixture was mixed well by pipetting. Up to 700 µl of sample, including any precipitate that may have formed, to an RNeasy spin column placed in a 2 ml collection tube. The lid was close gently and the tube was centrifuged at 10,000 x g for 15 seconds. The flow-through was discarded from collection tube. 700 µl of Buffer RW1 was added to the RNeasy spin column in a 2 ml collection tube. The lid was closed gently and the tube was centrifuged at 10,000 g for 15 seconds. The flow-through was discarded and the collection tube was reuse. To wash the spin column membrane, 500 µl of Buffer RPE was added to the RNeasy spin column in a 2 ml collection tube. The lid was closed gently and the tube was centrifuged at 10,000 g for 15 seconds. The flow-through was discarded and the collection tube was reuse. To wash the spin column membrane, 500 µl of Buffer RPE was added to the RNeasy spin column in a 2 ml collection tube.
collection tube. The lid was closed gently and the tube was centrifuged at 10,000 g for 2 minutes. To eliminate any carryover of buffers, the RNeasy spin column (with lid close) was then placed in a new 2 ml collection tube and was centrifuged at 10,000 x g for 1 minute. To elute the RNA, the RNeasy spin column was put in a new 1.5 ml collection tube and 30-50 µl of RNase-free water was directly added to the spin column membrane and the tube was centrifuged at 10,000 x g for 1 minute. The eluted RNA was used immediately or stored at -20 °C for future use.

6.2.5.2 DNase treatment

DNase treatment of freshly extracted RNA was done using the TURBO DNA-free™ Kit (Thermo Fisher Scientific). 0.1 volume of 10X TURBO DNase Buffer and 1 µl of TURBO DNase were added to the RNA and mixed gently. The mixture was incubated for 20-30 minutes at 37 °C. 0.1 volume of resuspended DNase Inactivation Reagent was added to the sample and mixed well. With occasional mixing, the mixture was incubated for 5 minutes at room temperature. Finally, the tube was centrifuged at 10,000 x g for 1.5 minutes and the treated RNA was transferred to a fresh tube. The DNase treated RNA was used immediately or stored at -20 °C for future use.

6.2.5.3 Bioanalyzer RNA 6000 Nano Kit quality check and quantification

RNA integrity and quantity were assessed using Bioanalyzer RNA 6000 nano kit (Agilent).
6.2.5.3.1 Setting up the chip priming station

A new syringe was placed in the chip priming station, the plate base was inserted in position C and the syringe clip was positioned at the top position.

6.2.5.3.2 Loading the gel dye mix

Protected from light, the gel-dye mix was put at room temperature for 30 minutes before use. After equilibration at room temperature, 9 µl of gel-dye mix was pipetted at the bottom of the well located on the fourth column and third row of the RNA chip. The RNA chip was then placed into the chip priming station. The chip priming station was closed and the plunger of the syringe that was previously positioned at 1 ml was pressed down until held by the clip. The plunger was kept down for exactly 30 seconds and then released. Finally, the plunger was inspected that it moved back to the 0.3 ml mark and after 5 seconds it was slowly pull back to the 1 ml position. After waiting 5 seconds, the plunger was gently pull back and 9 µl of gel-dye mix was pipetted in the wells located at the fourth column and first and second row.

6.2.5.3.3 Loading the marker

5 µl of RNA 6000 Nano marker was pipetted into the wells with a ladder symbol and into each of the 12 sample wells.

6.2.5.3.4 Loading the ladder and the samples

Before use, the ladder aliquots were thaw and kept on ice. Before loading, the samples were heat denatured at 70 °C for 2 minutes, hence minimising the formation
of secondary structure. 1 µl of the RNA ladder was pipetted in the well marked with
the ladder symbol. 1 µl of sample was pipetted in each of the 21 sample wells. The
chip was then placed horizontally in the adapter of the IKA vortex mixer and vortex
for 60 seconds at 2400 rpm. Finally, the RNA chip was placed into the Agilent 2100
Bioanalyzer (with chip selector was set to position (1)).

6.2.5.3.5 Starting the chip run

Using the 2100 Expert software, the appropriate assay was selected from the Assay
menu located in the instrument context tab. Then the sample informations were
inserted into the sample name table. Finally, the button start was pressed to start the
run.

6.2.5.3.6 Data analysis

RNA integrity and concentration were assessed directly using the 2100 Expert
software built-in analysis module.

6.2.5.4 cDNA synthesis

Complementary DNA (cDNA) was synthesised using the High Capacity cDNA
Reverse Transcription kit (Thermo Fisher Scientific). In a PCR tube on ice, 10 µl of
RNA (equivalent to 1µg; see section 6.2.5.3 - Bioanalyzer RNA 6000 Nano Kit quality
check and quantification) was mixed with 10 µl of 2X RT Master mix (2 µl of 10X RT
Buffer; 0.8 µl of 25X dNTP mix (100 mM); 2 µl of10X RT random primers; 1 µl of
MultiScribe Reverse Transcriptase; 1 µl of RNase inhibitor; and 3.2 µl of nuclease-
free water). The mixture was mixed by pipetting up and down several times and kept on ice. The mixture was then placed in a thermocycler and incubated at 25 °C for 10 minutes, 37 °C for 120 minutes, 85 °C for 5 minutes and then hold at 4 °C. The synthesised cDNA was used immediately or stored at -20 °C for future use.

6.2.5.5 Quantitative PCR

6.2.5.5.1 Samples preparation

10 µl of sample cDNA (i.e. from untransfected HeLa shUHRF1, untransfected HeLa shLuc) was mixed with 90 µl of RNase free water, hence resulting in a diluted sample cDNA at 5 ng/µl.

6.2.5.5.2 Serial dilution preparation for standard curve

A mixture of both untransfected HeLa shLuc and untransfected HeLa shUHRF1 cDNA was used to generate the standard curve.

5 µl of 50 ng/µl of untransfected HeLa shLuc cDNA was mixed with 5 µl of 50 ng/µl of untransfected HeLa shUHRF1 cDNA and 90 µl of RNase free water, hence resulting in a diluted mixture of cDNA 'D1' at 5 ng/µl. 4 µl of D1 was mixed with 16 µl of RNase free water hence resulting in a diluted cDNA 'D2' at 1 ng/µl. 4 µl of D2 was mixed with 16 µl of RNase free water hence resulting in a diluted cDNA 'D3' at 200 pg/µl. 4 µl of D3 was mixed with 16 µl of RNase free water hence resulting in a diluted cDNA 'D4' at 40 pg/µl.
6.2.5.5.3 Reaction set-up and PCR amplification

All the following reactions were done in triplicate to ensure accuracy. In a MicroAmp™ Fast Optical 96-Well Reaction Plate (Thermo Fisher Scientific),

• 4 µl of diluted samples was mixed with 10 µl of SYBR Select 2X Mix (Thermo Fisher Scientific), 0.4 µl of each 10 µM forward and reverse primers (Table 6.17) and 5.2 µl of water.

• 4 µl of each of D1, D2, D3 and D4 was mixed with 10 µl of SYBR Select 2X Mix, 0.4 µl of each 10 µM forward and reverse primers (Table 6.17) and 5.2 µl of water.

• 4 µl of negative controls (i.e. water) was mixed with 10 µl of SYBR Select 2X Mix, 0.4 µl of each 10 µM forward and reverse primers (Table 6.17) and 5.2 µl of water.

The plate was sealed using MicroAmp™ Optical Adhesive Film (Thermo Fisher Scientific) and centrifuged at 1,500 rpm for 30 seconds. The plate was then placed in StepOne Plus system (Thermo Fisher Scientific) and incubated at 50 °C for 2 minutes, 95 °C for 5 minutes and 40 cycles of 95 °C for 15 seconds; 60 °C for 1 minute; fluorescent signal acquisition.
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’&gt;3’)</th>
<th>Amplicon size</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHRF1_FW</td>
<td>GCCATACCCTCTTCGACTACG</td>
<td>237 bp</td>
</tr>
<tr>
<td>UHRF1_RV</td>
<td>GCCCAATTCCGTCTCATCC</td>
<td></td>
</tr>
<tr>
<td>GAPDH_FW</td>
<td>GGAGCGAGATCCCTCCAAAAT</td>
<td>197 bp</td>
</tr>
<tr>
<td>GAPDH_RV</td>
<td>GGCTGGTGTCACTTCTCATGG</td>
<td></td>
</tr>
</tbody>
</table>

*Primers were chosen from PrimerBank open database
Wang et al. 2003; Wang et al. 2012*

6.2.5.5.4 Data analysis using the relative standard curve

Relative standard curve method was used to determine the relative target quantity in samples. The StepOne software (Thermo Fisher Scientific) measures amplification of the target (*UHRF1*) and of the endogenous control (*GAPDH*) in sample (i.e. HeLa shUHRF1), in a reference sample (i.e. HeLa shLuc), and in a standard dilution series (i.e. D1, D2, D3 and D4). Measurements are normalised using the endogenous control *GAPDH*. Data from the standard dilution series are used to generate the standard curve. Using the standard curve, the software interpolates *UHRF1* quantity in the sample HeLa shUHRF1 and in the reference sample HeLa shLuc. The software determines the relative quantity of *UHRF1* in HeLa shUHRF1 by comparison to target quantity in the reference sample HeLa shLuc.
6.2.6 Protein assay

6.2.6.1 Protein extraction from adherent cells

Cells were trypsinized (see section 6.2.3 - Maintenance of adherent cells) and pelleted in falcon tube by centrifugation at 500 x g for 5 minutes. The cells were washed with 1.5 ml of PBS 1X and centrifuged at 500 x g for 5 minutes at 4 °C. To lyse the cells and extract proteins, 150 µl of ice-cold RIPA buffer (see section 2.1.5 - RIPA) was added to the pellet and the mixture was mixed thoroughly by vortexing for 10 seconds. The mixture was placed on ice for 15 minutes with vortexing every 5 minutes for 10 seconds. Tubes were then centrifuged at 16,000 x g for 15 minutes at 4 °C. The supernatant was collected and transferred to ice-cold 1.5 ml tube (Eppendorf). Protein lysates were used immediately or stored at -20 °C for later use.

6.2.6.2 Protein quantification

Proteins were quantified using Quick Start™ Bradford Protein Assay (Bio-Rad). Before starting, the 1x dye reagent were removed from 4 °C storage, let to warm at ambient temperature and inverted few times before use. Bovine serum albumin was diluted into RIPA buffer to make a series of 8 protein standards ranging from 0 to 2000 µg/ml. In triplicate, 5 µl of each standards and samples were added to FLUOTRAC 200 microplate (Greiner Bio-One) and mixed with 250 µl of 1x dye reagent. The microplate was incubated for 5 min at room temperature and then placed into a microplate reader (PHERAsar® FS, BMG LABTECH). The absorbance at 595 nm was measured for each standards and samples. To determine
samples concentration, absorbance of each triplicate was averaged and the blank (i.e. 0 µg/mL of protein) value was subtracted from all standards and samples. Linear equation was generated by finding the line of best fit of the absorbance of each standard and their respective protein concentration. Samples protein concentration were estimated by solving $x$ in the linear equation $y = ax + b$, where $y$ is the absorbance, $x$ is the protein concentration, $a$ is the slope of the line and $b$ is the $y$-intercept.

**6.2.6.3 Western Blot**

**6.2.6.3.1 SDS-PAGE**

20 µg of protein lysate (see section 6.2.6.1 - Protein extraction from adherent cells and 6.2.6.2 - Protein quantification) was mixed with 5 µl of 4X Laemmli sample buffer (see section 2.1.4 - 4X Laemmli sample buffer) and, if needed, deionized water for a total volume of 20 µl. Samples were then heated at 100 °C for 10 minutes, vortexed and quickly centrifuged. NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels (Thermo Fisher Scientific) was put into XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Fisher Scientific). Upper chamber of the electrophoresis system was filled with 200 ml of 1X NuPAGE® MOPS SDS Running Buffer (Thermo Fisher Scientific) with 500 µl NuPAGE® Antioxidant (Thermo Fisher Scientific) and the lower chamber was filled with 600 ml of 1X NuPAGE® MOPS SDS Running Buffer only. Samples were then loaded on gel and run at 200 V constant for 50 minutes.
6.2.6.3.2 Transfer

Proteins were transferred onto a membrane of nitrocellulose (GE Healthcare Life Sciences). Sponges, Whatman papers (Thermo Fisher Scientific), nitrocellulose membrane and protein gel were incubated for 5 minutes in pre-cooled 1X transfer buffer (see section 2.1.1 - 1X Transfer buffer). They were then assembled into a “sandwich” making sure no bubbles were present between layers in the following order: black side transfer - sponge - Whatman paper - gel - membrane - Whatman paper - sponge - red side transfer. The ‘sandwich’ was then loaded into a transfer apparatus filled with pre-cooled 1X transfer buffer. Transfer was performed for 90 minutes at 100 V constant and at 4 °C. After transfer, the nitrocellulose membrane was incubated in blocking buffer (see section 2.1.3 - Blocking buffer) for 1 hour at room temperature and on a rocking platform for gentle agitation.

6.2.6.3.3 Immunoblotting

The blocked nitrocellulose membrane was incubated with primary antibody against UHRF1 (mouse monoclonal; 1:1,000 in blocking buffer; BD Biosciences) or beta-actin (goat polyclonal; 1,2000 in blocking buffer, Santa Cruz Biotechnology) overnight at 4 °C and on rocking platform. Membrane was then washed with 1X PBST (see section 2.1.2 - 1X PBS-Tween 20 (PBST)) 3 times for 10 minutes and incubated in HRP-conjugated secondary antibody for 1 hour at room temperature and on rocking platform. Membrane was washed again 3 times for 10 minutes with 1X PBST on rocking platform.
6.2.6.3.4 Membrane revelation

The chemiluminescence reaction was done by incubating for 1 minute and gentle agitation the nitrocellulose membrane with an equal mixed volumes of the Enhanced Luminol Reagent and the Oxidizing Reagent included in the Western Lightning Plus ECL kit (Perkin Elmer). Excess of chemiluminescence reagent was removed by draining and the membrane was placed in plastic sheet protector. Protein visualisation was done by chemiluminescence imaging using an automatic image capture GeneGnome XRQ (Syngene).

6.2.6.3.5 Gel analysis

ImageJ (http://imagej.nih.gov) was used to compare protein band densities. Density of protein of interest (i.e. UHRF1) was measured using ImageJ gel analysis module (top menu: Analyze>Gels), normalised to the measured loading control density (i.e. beta-actin) and scaled between 0 and 1.

6.2.7 Methylation assay

6.2.7.1 DNA extraction from adherent cells

Genomic DNA of adherent cells was extracted using Quick-DNA™ Miniprep Kit (Zymo Research). Cells were trypsinized (see section 6.2.3 - Maintenance of adherent cells) and pelleted in a falcon tube by centrifugation at 500 x g for 5 minutes. To lyse the cells, supernatant was removed and 500 µl of Genomic Lysis Buffer was added directly to the cells pellet. Pellet was resuspended by vortexing 4-
6 seconds and let stand for 5-10 minutes at room temperature. After lysis, the mixture was transferred to a Zymo-Spin™ Column in a Collection Tube and centrifuged at 10,000 x g for 1 minute. Zymo-Spin™ Column was transferred to a new Collection Tube, 200 µl of DNA Pre-Wash Buffer was added to the spin column and centrifuged at 10,000 x g for 1 minute. 500 µl of g-DNA Wash Buffer was added to the spin column and centrifuged at 10,000 x g for 1 minute. For elution, the spin column was transferred to a clean micro centrifuge tube, 50 µl of DNA Elution Buffer was added to the spin column and let incubated for 2-5 minutes at room temperature. The spin column was centrifuged at 10,000 x g for 30 seconds. The eluted DNA was quantified by Nanodrop (see section 2.3 - DNA quantification) and immediately used or stored at -20 °C for future use.

6.2.7.2 Methylation-sensitive restriction digestion assay

The genomic DNA was extracted from the different cell lines (see section DNA extraction from adherent cells) and digested with Msp I and Hpa II (New England Biolabs). Restriction digestion reaction was prepared as followed: in 2 different tubes 2 units of Msp I or 2 units of Hpa II were mixed with 1 µg of gDNA, 1X CutSmart® Buffer (New England Biolabs) and deionized water if needed for a total final volume of 50 µl. Both tubes reactions were incubated at 37 °C for 2 hours. The digested DNA samples were resolved on a 1% (w/v) agarose gel (1 g agarose dissolved in 100 ml 1X TAE buffer (see section 2.1.7 - 1X TAE) containing SYBR® Safe DNA gel staining (Thermo Fisher Scientific) for 30 min at 100 V constant. The gel was then placed into Gel Doc XR+ System (Bio-Rad) for UV exposition and image capture. ImageJ was
used to determine the top gel band density and the associated level of enzymatic
digestion. HpaII cuts at unmethylated sites (bright band) whilst methylated sites
remain protected to the enzyme (darker band).

6.2.7.3 Manual DNA dot-blot

6.2.7.3.1 DNA spotting

In 6 µl final volume reaction, 100 ng of sample genomic DNA (see section 2.3 - DNA
quantification) or a set of calibrated human methylated DNA (Zymo Research;
standard curve at 0%, 25%, 50%, 75%, 100% methylated DNA) was mixed with 0.6
µl of 0.1 M EDTA, 0.48 µl of 5 M NaOH and, if necessary, dH$_2$O. Samples were
denatured at 99 ºC for 10 minutes and let cool down at room temperature without
snap-cooling. Charged nylon membrane (Zeta-Probe Membranes, Bio-Rad) was wet
in autoclaved dH$_2$O, and placed face-up on Whatman paper (Thermo Fisher
Scientific) to remove the excess of water. Denatured DNA samples were quickly but
carefully manually spotted on the membrane whilst the latter was still wet. Nylon
membrane was let air-dried for at least 10 minutes and rinsed face down in a clean
box containing 100-200 ml of 2x SSC buffer (see section 2.1.6 - 2X SSC) for 1–2
minutes at room temperature and gentle agitation. The membrane was oven baked
at 80 ºC for 1 hour to immobilise the DNA onto it and then blocked in blocking buffer
(see section 2.1.3 - Blocking buffer) for 1 hour at room temperature and gentle
agitation.
6.2.7.3.2 5-mC visualisation

The blocked nylon membrane was incubated with primary monoclonal antibody against 5-mC (Eurogentec) diluted at 1:1,000 in blocking buffer (see section 2.1.3 - Blocking buffer) overnight at 4 °C and on rocking platform. Membrane was then washed with 1X PBST (see section 2.1.2 - 1X PBS-Tween 20 (PBST)) 3 times for 10 minutes and incubated in HRP-conjugated secondary antibody (1:2,000 in 1X PBS-T) for 1 hour at room temperature and on rocking platform. Membrane was washed again 3 times for 10 minutes with 1X PBST on rocking platform. The chemiluminescence reaction was done by incubating for 1 minute and gentle agitation the nylon membrane with an equal mixed volume of the Enhanced Luminol Reagent and the Oxidizing Reagent included in the Western Lightning Plus ECL kit (Perkin Elmer). Excess of chemiluminescence reagent was removed by draining and the membrane was placed in plastic sheet protector. Protein visualisation was done by chemiluminescence imaging using an automatic image capture GeneGnome XRQ (Syngene).

6.2.7.3.3 DNA quantity assessment

Following 5-mC detection, the nylon membrane was washed 2 times 1 minute in 1x PBST. DNA staining was done by incubating the membrane in 1x SYBR® Gold (Thermo Fisher Scientific) in 1x TAE (see section 2.1.7 - 1X TAE) for 5-10 minutes and gentle agitation. Membrane was quickly washed 2-3 times in 1X PBST to remove the excess of SYBR® Gold and then placed into Gel Doc XR+ System (Bio-Rad) for UV exposition and image capture.
6.2.7.3.4 Membrane analysis

ImageJ (http://imagej.nih.gov) and the MicroArray Profile plugin (Bob Dougherty and Wayne Rasband, http://www.optinav.info/MicroArray_Profile.htm) were used to analyse the membrane. The measured 5-mC density of the standard curve and of the samples was normalised to their corresponding measured DNA loading control density. The methylation level in the samples was calculated by resolving $x$ in the linear equation $y = ax + b$ derived from the methylated standard points regression analysis.

6.2.7.4 Targeted bisulfite sequencing assay

6.2.7.4.1 Library preparation and sequencing

The following library preparation method was developed by Dr. Eguzkine Ochoa. The data were generated by Dr France Docquier. The data processing and analysis were performed by me.

6.2.7.4.1.1 First round PCR: target capture

500 ng (see section 2.3 - DNA quantification) of gDNA or a set of calibrated human methylated DNA (Zymo Research; standard curve at 0%, 25%, 50%, 75%, 100% methylated DNA) was bisulfite treated (see section 2.4 - Sodium bisulfite treatment). 1 µl of bisulfite converted DNA was mixed with 8.5 µl of ZymoTaq™ PreMix (Zymo Research), 0.85 µl of each 10 µM forward and reverse primers (Table 6.18) and 5.8
μl of distilled water. The mixture was placed in a thermocycler and incubated at 95
° C for 10 min, 40 cycles of 95 °C for 30 sec; 58 °C for 40 sec and 72 °C for 1 min,
72 °C for 7 min and then hold at 4 °C. PCR reaction was loaded on a 1.5% agarose
gel electrophoresis (1.5 g agarose dissolved in 100 ml 1X TAE (see section 2.1.7 -
1X TAE), 1X SYBR Safe) and the gel was run for 40 minutes at 100 V. Once finished,
the gel was placed into Gel Doc XR+ System (Bio-Rad) for UV exposition. The size
of amplicons was compared to the 1 kb DNA ladder. The PCR products were used
immediately after or stored at -20 °C for later use.

Table 6.18: List of primers used for targeted bisulfite sequencing.

KvDMR1 primers amplify a region within the imprinting centre KCNQ1OT1:TSS-
DMR. H19 primers amplify a region within the imprinting centre H19/IGF2:IG-DMR.
KvDMR1 and H19 primers were designed by Dr. Eguzkine Ochoa. IGS-rDNA
primers amplify a region within the intergenic spacer of ribosomal DNA (Rothbart et
al. 2012).

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5'→3')</th>
<th>Amplicon size</th>
</tr>
</thead>
<tbody>
<tr>
<td>KvDMR1_Fw</td>
<td>TCGTCGGACGAGGCTGAGATTAGTATATTAGTTAG</td>
<td>163 bp</td>
</tr>
<tr>
<td>KvDMR1_Rv</td>
<td>GTCCTGTGGGGTGGAAAGAAGGCCATGAGGATTAGTATATTAGTTAG</td>
<td></td>
</tr>
<tr>
<td>IGS-rDNA_Fw</td>
<td>TCGTCGGACGAGGCTGAGATTAGTATATTAGTTAG</td>
<td>274 bp</td>
</tr>
<tr>
<td>IGS-rDNA_Rv</td>
<td>GTCCTGTGGGGTGGAAAGAAGGCCATGAGGATTAGTATATTAGTTAG</td>
<td></td>
</tr>
</tbody>
</table>

In red, common sequence used for multiplex 2nd round PCR amplification (adaptor tagging and indexing).
In blue, sequence specific to targeted region.

6.2.7.4.1.2 Purification of PCR products

The PCR products were purified using QIAquick PCR purification kit and eluted in
50 μl of double distilled water (see section 2.2 - DNA purification). The purified PCR
products were used immediately after or stored at -20 °C for later use.
6.2.7.4.1.3 Second round PCR: adaptor tagging and indexing

5 µl of purified PCR products was mixed with 25 µl of 2X KAPA HiFi HotStart ReadyMix (Kapa Biosystems), 5 µl of each Nextera XT Index forward and reverse primers (Nextera XT Index Kit v2 Set A, Illumina) and 10 µl of distilled water. The mixture was placed in a thermocycler and incubated at 95 °C for 3 min, 8 cycles of 95 °C for 30 sec; 55 °C for 30 sec and 75 °C for 30 sec, 72 °C for 5 min and then hold at 4 °C. The PCR products were used immediately after or stored at -20 °C for later use.

6.2.7.4.1.4 Purification of PCR products

The resulting adaptor-tagged and indexed PCR product was purified and size selected using AGENCOURT® AMPURE® XP (Beckman Coulter). The PCR product was transferred to a 300 µl round bottom plate (Thermo Fisher Scientific). The AMPure XP bottle was gently mixed to resuspend the magnetic particles that may have settled. 90 µl of AMPure XP was added to the PCR reaction and then mixed thoroughly by pipetting up and down 10 times. The mixed sample was incubated for 5 minutes at room temperature. To separate the beads from the solution, the plate was placed onto an Agencourt SPRIPlate 96 Super Magnet Plate (Beckman Coulter) for 2 minutes. The cleared solution was aspirated from the reaction plate and discarded. With the plate still on the magnetic stand, the beads were washed by adding carefully (i.e. without disturbing the separated magnetic beads) 200 µl of freshly prepared 70% ethanol to the well of the reaction plate and let to incubate for 30 seconds at room temperature. The ethanol was carefully removed and the wash
step was repeated for a total of two washes. Once the ethanol removed, the plate was let to air dry for 3 minutes. The plate was then removed from the magnetic stand and the sequencing libraries were eluted by adding 56 µl of distilled water to it and mixing by pipetting up and down 10 times. The eluted DNA was used immediately after or stored at -20 °C for later use.

6.2.7.4.1.5 Sequencing libraries quality assessment

The resulting purified sequencing libraries were quantified using Qubit® dsDNA HS Assay Kit (see section 2.3 - DNA quantification) and their size were checked using Bioanalyzer High Sensitivity DNA Kit (see section 2.7 - Fragment analysis of DNA libraries). The sequencing libraries were then diluted to 4 nM, pooled if necessary (i.e. in case of sample multiplexing) and stored at -20 °C until sequencing.

6.2.7.4.1.6 Sequencing

The sequencing of the libraries was performed by the Stratified Medicine Core Laboratory (SMCL) Next Generation Sequencing facility. A MiSeq sequencer was used for the sequencing.

6.2.7.4.2 Processing and analysis of NGS data

Trim Galore v0.4.0 was used to remove and trim low-quality sequences and sequencer adaptors from bisulfite converted reads. Bismark v0.14.4 (Krueger and Andrews 2011) was used to align the trimmed sequencing reads to an in silico bisulfite converted human genome (GRCh38.p6). Bismark Methylation Extractor
v0.14.4 was used for cytosine methylation calling. The shell code used to trim, align
and extract 5-mC level from the sequencing reads is available in Appendix 5.

6.2.7.4.3 5-mC quantification
The resulting methylation called file was used to assess the methylation level for
each amplicon (i.e. H19/IGF2:IG-DMR; KCNQ1OT1:TSS-DMR and IGS-rDNA). The level of 5-mC in each locus was estimated by calculating the average of 5-mC level reported at each CpG sites within each locus. The R script used for 5-mC quantification is available in Appendix 6.

6.2.7.4.4 PCR bias correction for 5-mC methylation level
To account for the PCR amplification bias, the degree of experimental methylation measured in each samples at each loci was corrected by resolving $x$ in the cubic polynomial equation $y = ax + cx^2 + dx^3 + c$ derived from the methylated standard points regression analysis (Moskalev et al. 2011). The R script used to adjust the methylation data is available in Appendix 7.

6.3 Results
6.3.1 Transfection optimisation
The following data were generated by Dr France Docquier and further analysed by myself. Optimal transfection efficiency was evaluated on parental non-knocked down
HeLa cells. Different concentrations of transfecting reagents were mixed with a constant amount of plasmid DNA (plasmid pRP, 1.5 μg) and transfection efficiency was evaluated by visual estimation of GFP positive cells using fluorescent microscope. The best condition of transfection was achieved using 1.5 μg of plasmid DNA and 9 μl of transfecting reagents (Figure 6.3). This condition was used to transfect wild-type and mutant UHRF1 in HeLa cells in which UHRF1 was knocked down.

![Image showing transfection optimization using pRP plasmid](image.png)

**Figure 6.3: HeLa cells transfection optimisation using pRP plasmid.**

For the transfection of the pRP plasmid into HeLa cells, different concentration of transfection reagent was used as indicated and transfection efficiency (level of GFP positive cells) was visually evaluated by fluorescent cell imager. Top, brightfield; bottom, green channel. (Data were generated by Dr France Docquier and analysed by myself).

### 6.3.2 Low level of UHRF1 mRNA and protein in HeLa shUHRF1

The following data were generated by Dr France Docquier and further analysed by myself. Quantitative PCR measuring UHRF1 mRNA level showed efficient
knockdown of more than 80% of UHRF1 between control HeLa cells (i.e. HeLa shLuc) and HeLa UHRF1 knockdown cells (i.e. HeLa shUHRF1) (Figure 6.4-a). This result was further confirmed by Western-blot. Compared to HeLa shLuc, UHRF1 protein level in HeLa shUHRF1 decreased by more than 80%, thus indicating efficient gene silencing of UHRF1 by shRNA (Figure 6.4-b).

Figure 6.4: Efficient knockdown of UHRF1 in HeLa cells.

UHRF1 shRNA knockdown efficiency was compared to a control luciferase shRNA in HeLa cells (a) by qPCR and (b) by Western blot (top: Western blot, bottom: UHRF1 gel band relative density obtained with ImageJ). HeLa shLuc, HeLa cells transduced with shRNA targeting luciferase; HeLa shUHRF1, HeLa cells transduced with shRNA targeting UHRF1. (Data were generated by Dr France Docquier and analysed by myself).
6.3.3 Transfection of UHRF1 does not rescue global methylation

The following data were generated by Dr France Docquier and further analysed by myself. In comparison to the control cells HeLa shLuc, the digestion of HeLa shUHRF1 cells by the methylation-sensitive restriction enzyme HpaII suggested that the knockdown of *UHRF1* induced a global reduction of DNA methylation. However, the transient expression of wild-type *UHRF1* seemed to fail to restore the methylation to a similar level than observed in the control cells (Figure 6.5-a). These qualitative observations provided unexpected suggestions as a methylation rescue was expected following the transient expression of wild-type *UHRF1*. DNA dot-blots assays were carried out to confirm this trend. In comparison to the control cell HeLa shLuc, in which the methylation level was 60.3%, the *UHRF1* knockdown in HeLa shUHRF1 induced the global methylation to drop to 23%. However, although the transient expression of wild-type *UHRF1* in HeLa shUHRF1 cells seemed to induce the methylation to increase slightly to 27.4%, it seemed to fail to restore the methylation to a similar level than observed in the control cells (Figure 6.5-b). UHRF1 protein level was assessed by Western-blot and significant UHRF1 over expression was observed after transient transfection of wild-type *UHRF1* in both HeLa shLuc and HeLa shUHRF1 (Figure 6.5-c). Altogether, these results suggested that although it was expressed in cells, the exogenous UHRF1 might have been non-functional. I hypothesised that this could be linked to the pRP vector backbone or the FLAG and HA protein tags. To address these issues, a plasmid construct identical to the one used in previous studies (Rothbart et al. 2012; Rothbart et al. 2013) was tested on
control and UHRF1 knockdown cells for DNA methylation rescue. In comparison to the control cells HeLa shLuc in which the methylation level was 70.7%, the knockdown of UHRF1 in HeLa shUHRF1 seemed to reduce the global methylation level to 37.9%. However, similar to experiments using the plasmid pRP, the transient expression of wild-type UHRF1 expressed from the new vector backbone pTag-2C seemed to fail to rescue the methylation to the level seen in the control cells. Interestingly, the transient expression of wild-type UHRF1 seemed to induce a small loss-of-methylation (drop to 60.8%) in the control cells HeLa shLuc. As expected, the transient expression of the mutant UHRF1 Y188A, used here as a positive control for loss of protein function (Rothbart et al. 2012), seemed to have no effect on DNA methylation in HeLa shUHRF1 cells (Figure 6.6-a). UHRF1 wild-type and UHRF1 Y188A protein levels were assessed by Western-blot in HeLa shUHRF1 cells and significant over expression was observed for both constructs (Figure 6.6-b). It has to be noted that in this set of experiment, the level of repeatability between my biological replicates was very low, hence the necessity to draw careful interpretations in regard to the obtained results.
Figure 6.5: Transfection of pRP UHRF1 WT does not rescue global methylation.

The global methylation level in HeLa shLuc and HeLa shUHRF1 transfected with different pRP plasmid (as indicated) was assessed by (a) digestion using methylation-sensitive restriction enzymes (n=1) and by (b) DNA dot-blot (n=2). (c) UHRF1 protein level in HeLa shLuc and HeLa shUHRF1 transfected with indicated plasmid was assessed by Western-blot. +NT, not transfected; +EV, empty vector; +WT, UHRF1-WT. (Data were generated by Dr France Docquier and analysed by myself).
Figure 6.6: Transfection of pTag-2C UHRF1 WT does not rescue global methylation.

(a) The global methylation level in HeLa shLuc and HeLa shUHRF1 transfected with different pTag-2C plasmids (as indicated) was assessed by DNA dot-blot (n=2). (b) UHRF1 protein level in HeLa shLuc and HeLa shUHRF1 transfected with indicated plasmids was assessed by Western-blot. +NT, not transfected; +EV, empty vector; +WT, UHRF1-WT; +Y188A, UHRF1-Y188A. (Data were generated by Dr France Docquier and analysed by myself).

6.3.4 Transfection of UHRF1 does not rescue methylation at IGS-rDNA

The following data were generated by Dr France Docquier and further analysed by myself. Targeted bisulfite sequencing was used to measure methylation levels at a CpG island within the intergenic spacer of ribosomal DNA (IGS-rDNA), a known target of UHRF1 (Bostick et al. 2007; Rothbart et al. 2012; Rothbart et al. 2013), and
at the imprinted locus \textit{KCNQ1OT1}:TSS-DMR. Both targets were amplified by bisulfite-PCR followed by next-generation sequencing on an Illumina sequencing platform (MiSeq). Using calibrated DNA standards, preferential amplification of the methylated alleles (4.41 fold at IGS-rDNA and 3.02 fold at \textit{KCNQ1OT1}:TSS-DMR) was observed at both loci (Figure 6.7). Consequently, the degree of experimental methylation measured in each samples was corrected for PCR amplification bias (Moskalev et al. 2011). The interpretation of the adjusted values showed that in agreement with previous studies (Bostick et al. 2007; Rothbart et al. 2012; Rothbart et al. 2013), \textit{UHRF1} knockdown seemed to induce a loss-of-methylation (~ 57-59%) at IGS-rDNA. However, and similar to our findings on global methylation (Figure 6.5 and Figure 6.6), the transient expression of wild-type \textit{UHRF1} expressed from the \textit{pRP} and \textit{pTag-2C} expression vector seemed to fail to restore the DNA methylation at IGS-rDNA (~ 57-58% LOM compare to control cells) (Figure 6.8-a). Similarly, DNA methylation at the imprinted locus \textit{KCNQ1OT1}:TSS-DMR seemed to decrease by approximatively 48-50% in \textit{UHRF1} knockdown cells but the transient expression of wild-type \textit{UHRF1} seemed to fail to restore the DNA methylation (~ 46-53% LOM compare to control cells) to a similar level than observed in control cells (Figure 6.8-b).
Figure 6.7: Degree of bias introduced by PCR amplification.

IGS-rDNA and KCNQ1OT1:TSS-DMR amplification bias was calculated from calibrated methylated DNA standards (0%, 25%, 50%, 75%, 100% methylated). The experimental degree of methylation observed after amplification (y-axis) was plotted as a function of the actual methylation percentage (x-axis). The value of b, equivalent to the PCR bias, was calculated by averaging individual b values calculated for each curve point with the following equation: $b = [y \times (100 - x)]/[x \times (100 - y)]$ where y is the uncorrected experimental and x is the real value. Red line, cubic polynomial regression line derived from calibrated DNA standards; black dotted line, derived theoretical regression line derived from un-bias amplification. (Data were generated by Dr France Docquier and analysed by myself).
Figure 6.8: Transfection of pTag-2C UHRF1 WT does not rescue methylation at KCNQ1OT1:TSS-DMR and IGS-rDNA. Methylation at IGS-rDNA and at KCNQ1OT1:TSS-DMR loci were obtained by bisulfite-PCR followed by next-generation sequencing. Corrected methylation estimated in HeLa shLuc and HeLa shUHRF1 cells transfected with the indicated construct at (a) IGS-rDNA and (b) KCNQ1OT1:TSS-DMR. +NT, not transfected; +EV, empty vector; +WT, wild-type UHRF1; +Y188A, UHRF1 Y188A mutant. (Data were generated by Dr France Docquier and analysed by myself.)
6.4 Discussion

I hypothesised that the novel homozygous missense variant in the *UHRF1* gene (NM_001048201.2, c.2237G>C, p.K667N) found in the two siblings with BWS_IC2 would affect the maintenance of DNA methylation at imprinted DMRs such as *KCNQ1OT1*:TSS-DMR without disrupting global DNA methylation. To test this hypothesis I undertook DNA methylation rescue assays in HeLa cells knockdown for *UHRF1* (HeLa shUHRF1). Those cells were a generous gift from Dr. Scott Rothbart and they were previously shown to have decreased DNA methylation genome-wide and at IGS-rDNA compared to control HeLa cells (i.e. HeLa shLuc). Additionally, it was shown that transient expression of wild-type UHRF1 in those cells should rescue the loss of DNA methylation resulting from the knockdown of UHRF1 (Rothbart et al. 2012).

Before evaluating the effects that the *UHRF1* variant may have on DNA methylation, experiments were undertaken to assess the effects of the transient expression of wild-type *UHRF1* using pRB expression vector in HeLa shUHRF1 cells. However, the choice of assessing the global level of methylation by DNA digestion using methylation-sensitive restriction enzymes was not indicated for my objectives. This method only provided the qualitative, and not quantitative, assessment of the global level of DNA methylation. The use of manual DNA dot-blot assays revealed to be more quantitative and appropriate for my objectives. However, the second set of experiments using the pTag-2C plasmid achieved a very low level of repeatability,
hence indicating the lack of both biological and technical replicates. Therefore, additional biological and technical replicates were required to gain more accuracy and confidence in the results. Consequently, the conclusions I drawn from these experiments are limited to the suggestion of possible changes in DNA methylation.

Global and targeted DNA loss-of-methylation associated with \textit{UHRF1} knockdown seemed not to be rescued following the transient expression of wild-type \textit{UHRF1}. This result was unexpected as it appeared to contradict the previous published results reported by Dr. Scott Rothbart and colleagues (Rothbart et al. 2012). I hypothesised that the presence of a dual FLAG and HA protein tags in C-terminus may prevent the exogenous UHRF1 from rescuing the DNA methylation defect. To test this hypothesis, Dr. France Docquier performed experiments using wild-type \textit{UHRF1} expressed from pTag-2C expression vectors. This plasmid construct, a generous gift from Dr. Scott Rothbart, displayed a single FLAG tag in N-terminus and was previously shown to be efficient in rescuing DNA methylation abnormality (Rothbart et al. 2012). Similar to the results obtained using pRB, the transient expression of wild-type \textit{UHRF1} using pTag-2C seemed to fail to rescue the global loss-of-methylation associated with \textit{UHRF1} knockdown. Additionally, no DNA methylation rescue was observed at IGS-rDNA and \textit{KCNQ1OT1}:TSS-DMR whilst it was expected that both loci would re-gain DNA methylation after transient expression of wild-type \textit{UHRF1}.

Following these failed attempts to rescue the DNA methylation associated with
245

1. **UHRF1** knockdown by expressing wild-type **UHRF1** with both the pRB and pTag-2C expression vector, I did not continue with the analysis of the UHRF1 missense variant found in the siblings with BWS. After carefully reviewing the experimental procedures and discussion with Dr. Scott Rothbarts’ collaborators, I hypothesised that in my experiments **UHRF1** was too overexpressed and this may have deregulated UHRF1 function, ultimately leading to a loss of DNA methylation. To support this hypothesis, a small loss-of-methylation was observed in HeLa controls cells following the transient expression of wild-type **UHRF1** whilst no DNA methylation change was expected. Additionally, UHRF1 overexpression had been associated with DNA hypomethylation in cancer such as hepatocellular carcinoma and esophageal squamous cell carcinoma (Mudbhary et al. 2014; K. Nakamura et al. 2016). However, although these observations would be compatible with my hypothesis, further studies would be required to confirm these.

The potential causal link between the mutant UHRF1_K667N and the DNA methylation maintenance failure at **KCNQ1OT1**:TSS-DMR is still under investigation. Further DNA methylation rescue experiments involving the use of an expression vector including a tag free protein and a weak promoter, such as the human ubiquitin C promoter, for gene expression is considered. Another approach using gene knockin in mice is currently being discussed. The CRISPR/Cas9 system could be used to generate a NP95 p.K662N (NP_035061) mouse line (p.K662N in mice is equivalent to p.K667N in human). Using this model would have the advantages of being able to investigate in more detail the effects of the mutant in the
maintenance of DNA methylation. Furthermore, although UHRF1 has mainly been
described as a DNA methylation maintenance protein, the genome-edited mouse
line may provide potential insights on the role of the protein (and the mutant) in the
establishment of DNA methylation.
Chapter 7

General discussion
7.1 Multilocus imprinting disturbances in imprinting disorders, a need for standardisation

Aberrant DNA methylation at imprinting control regions plays a critical role in the molecular aetiology of imprinting disorders (IDs) such as BWS and SRS. In the current report, I combined the use of a reliable and robust genome-wide methylation array and an accurate and efficient bioinformatic pipeline to identify widespread methylation disturbances across multiple imprinting control regions in a subset of BWS and SRS cohorts. Though my findings were consistent with previously published reports (Rossignol et al. 2006; Bliek et al. 2008; Azzi et al. 2009; Court et al. 2013; Poole et al. 2013; Tee et al. 2013; Maeda et al. 2014), they also highlighted a current need to harmonise the way in which aberrant DNA methylation in ID individuals is defined, analysed and reported. Indeed, the variability in (i) the loci investigated, (ii) the platform used to perform the methylation profiling and (iii) the methodology used to identify methylation change complicated comparisons with other studies (discussed below). To address these issues, the research community has begun adopting different measures that would help standardisation of results and consequently leading to better comparison between studies (Monk et al. 2016).

7.1.1 Where to look?

To define the frequency of multilocus imprinting disturbances (MLIDs) in ID cohort researchers focused on a subset of arbitrary loci. In the current study I analysed DNA methylation at 46 imprinted differentially methylated regions (DMRs), which
contrasts with the limited number of loci that were assessed in others (e.g. 7 in Azzi et al. 2009; 4 in Tee et al. 2013). As discussed in Chapter 4, it seems that the more loci being investigated, the higher the reported MLID frequency. Therefore, whilst assessing all imprinted DMRs might not be possible in all cases (e.g. technical constrains associated with the used platform), it will be beneficial to define a core set of imprinted DMRs that would need to be included in all reports.

### 7.1.2 Too many platforms might lead to confusion

The use of different platforms to assess methylation may also hamper comparison between studies. Whilst some studies, including mine, used high-throughput genome-wide methylation array, others relied on combined bisulphite restriction analysis and methylation-specific PCR (Bliek et al. 2008), allele-specific methylated multiplex real time quantitative PCR (Azzi et al. 2009), or pyrosequencing (Poole et al. 2013). The sensitivity to detect subtle changes in DNA methylation and low-level mosaicism may differ between platforms, hence leading to the identification, or exclusion, of findings that may be relevant to the research question and to diagnostic testing.

### 7.1.3 A common informatic approach to detect epimutations

A standardised method to define a loss- or a gain-of-methylation should be defined. Should the method rely on a cut-off threshold calculated from normal control individuals? How many normal control individuals should be used for the
In the current study, I developed a method relying on a cut-off threshold of ± three standard deviations calculated from the mean of a cohort of more than 1400 normal control individuals. Whilst this is an impressive number that should increase the robustness of statistical analysis, it is unlikely that such large amounts of data would be available to all laboratories. To overcome this issue, alternative informatics methodologies have been developed, such as a single case-control comparison test that requires 20 normal control individuals only and which is based on Crawford-Howell t-test (Rezwan et al. 2015).

### 7.1.4 Need to find a better alternative for diagnostic services

With the scope to address these issues, I originally undertook the development of a novel NGS strategy for the genomic imprinting methylation profiling of individual with BWS and SRS, and more globally with a clinically diagnosed ID. The methodology was based on bisulfite padlock probes, a potentially reliable and sensitive method that was previously used to characterise 66,000 CpG sites within 2,020 CpG islands on human chromosome 12, chromosome 20, and 34 selected regions (Deng et al. 2009), and 500,000 CpG sites within genomic regions known to contain differentially methylated regions or sites, CTCF binding sites, DNase I hypersensitive regions, all microRNA genes and all promoters for human NCBI Reference Sequence genes (Diep et al. 2012). Although the development of the method was not successful, it could have offered a highly sensitive method to characterise at a single base pair resolution the majority of CpGs within all the currently known imprinted DMRs (Court
et al. 2014) in a single multiplex reaction. The method was intended to allow the sequencing of a large number of samples using the bench top MiSeq sequencer whilst maintaining great depth of sequencing coverage (>150 samples at >500x). Consequently, this could have resulted in very precise detection of DNA methylation change at a given region and CpG site level, hence improving accuracy in epimutation identification rate even in samples with low level mosaicism. Finally, the method could have been used for the genomic imprinting methylation profiling in several tissues, hence leading to the characterisation of the nature and frequency of MLID in these.

In conclusion another NGS-based method, if successfully developed and applied, might be part of the answer to standardise findings in the scope of better comparison of findings between studies. In the meantime, until such NGS-based method is successfully developed, I would recommend the use of HM450K for the clinical diagnostic of individuals with suspected ID. Despite some limiting factors, such as low probe coverage at some loci or presence of technical and biological batch effects, it provided 100% sensitivity and >99% specificity in identifying all of my BWS and SRS patients and similar efficiency has been seen in other studies focussing on other ID (Court et al. 2013; Docherty et al. 2014; Maeda et al. 2014; Prickett et al. 2015; Rochtus et al. 2016). Finally, the HM450K allows in a single analytic procedure the comprehensive and simultaneous assessment of DNA methylation at not only the disease-causing loci but also at other regions not known to be involved in the molecular aetiology of the disease which allows potential new discovery.
7.2 The possible underlying causes of BWS and MLID

The underlying causes of MLID in BWS, and more globally in most ID, remain to date unknown. However, several hypotheses that could explain the aberrant epigenotype have been formulated.

7.2.1 Environmental pressure

It is well established that assisted-reproductive technology (ART) is associated with the LOM of KCNQ1OT1:TSS-DMR in BWS individuals (DeBaun et al. 2003; Gicquel et al. 2003; Maher et al. 2003; Halliday et al. 2004; Weksberg et al. 2010). Consequently, it was natural to suggest that ART might also be implicated in the MLID epigenotype. To that end, several studies, including the current one, undertook the characterisation of MLID in BWS_IC2 individuals conceived with ART. Whilst I and others found an association (Lim et al. 2009; Hiura et al. 2012; Tee et al. 2013), others reported the opposite (Rossignol et al. 2006; Azzi et al. 2009). The discrepancy between reports might be explained by several factors that might include (i) the platform/methodology used to identify the epimutation (which could relate to the discussion above), (ii) the overlap between cohort used for the different studies (the current one may overlap with Lim et al. 2009 and Tee et al. 2013) and therefore the possible bias in the findings and (iii) differences between the ART procedures themselves. For the latter, it is conceivable that the techniques used to perform the ART may differ to some degree between hospitals and countries. Consequently, some aspects of a procedure may interfere with DNA methylation
differently, hence resulting in different epigenotype.

Overall, although it is not a universal finding, cohort studies suggested that ART might be associated with MLID epigenotype. However, it is yet to be determined if the methylation disturbances result directly from the ART procedures themselves, other factors such as the infertility or the age of the parents, or a combination of all. Furthermore, there is a need to identify which step of the ART procedures is involved in this mechanism. Therefore, the study of more extensive cohort will bring some clarity in the characterisation of ART in MLID epigenotype.

7.2.2 Genetic component

Recently, the identification of MLID in a subset of individuals with IDs has triggered a particular interest in performing WES screening to determine if there might be potential underlying genetic causes that are responsible for MLID. However, despite much effort, only a few causative trans-acting mutations have been found. These include the maternal-zygotic gene \textit{ZFP57} and the maternal genes \textit{NLRP7}, \textit{KHDC3L}, \textit{NLRP2} and \textit{NLRP5} (Murdoch et al. 2006; Mackay et al. 2008; Meyer et al. 2009; Parry et al. 2011; Docherty et al. 2015). In the current study I undertook WES with the scope to identify new pathogenic mutations that might explain BWS and MLID phenotype. In addition to a homozygous missense variant in the \textit{UHRF1} gene for which the potential pathogenicity is currently being assessed through functional studies within our group, I report additional potential candidate gene mutations that
affect maternal-effect genes (*BNC1, CTCF, NLRP2*) and could be implicated in BWS phenotype and MLID epigenotype. However, the potential pathogenicity of these still require further validation. Overall, taking into consideration the current study and previous reports (Begemann et al. 2011; Azzi et al. 2014; Caliebe et al. 2014) it appears that the number of disease-causing mutations found by targeted exome and WES is very rare in individuals with an ID. Alternatively, it is possible that the potential pathological variants may have been missed due to the complexity that poses the analysis and interpretation of WES. The filtering and prioritisation criteria are critical in WES analysis such as using wrong parameters might result in either too many variants left for interpretation (resulting in false positives) or in losing the pathogenic disease-causing ones (resulting in false negatives).

On the other hand, it is possible that the underlying genetic causes reside elsewhere in the genome (e.g. intronic or other genomic material such as non-coding RNA) or is not well detected by WES (e.g. internal exon deletion, copy number variation, chromosomal rearrangement, repeats variations). Consequently, it is worth considering performing whole-genome sequencing (WGS) to obtain the most comprehensive view of the genomic informations. However, using such approach also comes with some limitations. *(i)* Whilst the WGS of a very small subset of carefully selected individuals is financially possible, the processing of large sample size cohort is unlikely. Hence the notion of carefully selected individuals. Under which criteria should a patient have his genome sequenced? Should the ones with more severe phenotype, a family history or a negative WES be prioritised? *(ii)* If WES is
challenging to interpret WGS will be prove even more so since the process is expected to yield millions of variants. Consequently, appropriate filtering and prioritisation criteria will be the key to remove false-positive variants without filtering-out the pathogenic ones.

7.2.3 Methylation abnormalities outside of imprinted DMRs

Finally, it is possible that the underlying causes of BWS and MLID are not of a genetic nature but an epigenetic origin. Besides methylation disturbances at imprinted DMRs, it is conceivable that other genomic regions might also present a pathogenic methylation profile that could participate to the BWS phenotype and MLID epigenotype. In the current study, using a genome-wide approach, I identified three additional regions outside of imprinted DMRs, *SLC12A9*, *LTA*, and *KCNAB3*, that display aberrant DNA methylation. However, the clinical significance of these findings is currently unknown and further investigations are needed to define their role in BWS and MLID aetiology. It should be noted that my analysis was restricted to the captured sequences on the arrays. Indeed, the Infinium HumanMethylation 450K BeadChips allows the characterisation of more than 485,000 CpG sites, which is equivalent to approximatively 2% of the human CpG content. Consequently, it is possible that pathologic methylation changes at an undocumented important locus might be associated with BWS and MLID but remain unidentified due to the lack of coverage by the array. The use of a more comprehensive method such as whole-genome bisulfite sequencing (WGBS), reduced-representation-bisulphite-
sequencing (RRBS) (Meissner et al. 2008; Z.D. Smith et al. 2009; Gu et al. 2011; Boyle et al. 2012) or methylated DNA immunoprecipitation sequencing (MeDIP-seq) could be an answer to alleviate this issue. Although they are more costly, especially to process large sample size cohorts, they will undoubtedly cover greater amount of CpG sites (WGBS, >90% of human CpG content; RRBS, 10-20%; MeDIP-seq, 60-90% (Plongthongkum et al. 2014)), hence allowing new discoveries that were not possible with methylation array alone.

7.2.4 5-hydroxymethylation, a hot topic for the future?

5-hydroxymethylcytosine (5-hmC) is a cytosine modification recently described to be an intermediate in the DNA demethylation pathway (see section 1.1.1.2.3 - DNA demethylation). The modification results from the oxidation of 5-methylcytosine (5-mC) by the ten-eleven translocation enzymes, TET1 and TET2 (He et al. 2011; Inoue et al. 2011; Ito et al. 2011; Hackett et al. 2013). The levels of 5-hmC varied in different organs with highest content found in the brain, followed by the rectum, liver and colon. In contrast, the level of 5-hmC is very low in the heart, breast and placenta (W. Li and M. Liu 2011). In mouse embryonic stem cells, 5-hmC was found to be mostly associated with euchromatin and, unlike 5-mC, was particularly enriched at CpG islands, gene promoters, particularly the ones marked with H3K4me3 and H3K27me3, and enhancers, particularly the ones marked with H3K4me1 and H3K27ac (Ficz et al. 2011; Pastor et al. 2011). Consequently, it was suggested that 5-hmC has functional significance in gene expression regulation. Since the recent
development of methods that allow the discrimination of 5-mC to 5-hmC (oxidative bisulfite sequencing), several studies undertook the characterisation of the impact of 5-hmC in diseases. Subsequently, multiple suggestive data indicated that 5-hmC might potentially be involved in the pathophysiology of diseases (Jingyu Wang et al. 2014) such as Rett syndrome (Szulwach et al. 2011), Huntington diseases (Fengli Wang et al. 2013), melanoma (Lian et al. 2012; Gambichler et al. 2013), gastrointestinal stromal tumours (Mason and Hornick 2013) and hepatocellular carcinoma (M.-L. Chen et al. 2013).

As an epigenetic regulator, it was natural to think that aberrant change in 5-hmC level might be implicated in the molecular aetiology of imprinting disorders via either directly deregulating imprinted DMRs or indirectly through the silencing or activation of a trans-acting factor. Accordingly, a report exploring the relationship between 5-hmC and Kagami-Ogata syndrome has recently been published (Matsubara et al. 2015). The authors concluded that the hypermethylation of the two imprinting control regions MEG3/DLK1:IG-DMR and MEG3:TSS-DMR was not due to an increased level of 5-hmC. Additionally, the analysis of blood taken from the patients did not revealed conclusive evidence of global changes of the levels of 5-hmC that could explain the disease phenotype. However, although inconclusive this report was the very first one to explore the relationship between 5-hmC and ID. The analysis of more individuals with an ID, including BWS, will surely help deciphering the potential implication of this epigenetic modification in the molecular aetiology of a such disorder.
7.3 Final conclusions

Genomic imprinting is a fascinating but incompletely understood process. Whilst breakthrough discoveries have been made via in-depth molecular investigation of model organisms, particularly mice, the study of imprinting disorders has provided additional critical insights into both the pathogenesis and associated phenotypic consequences of aberrant imprinting. Application of high-throughput NGS-based techniques to large cohorts of ID patients has the potential to provide further insights and this would be further facilitated by data-sharing and international collaborative efforts such as the INGENIUM training network.
Chapter 8

Appendix
Appendix 1

*R code used for methylation data preprocessing.*

```
> library("RnBeads")
# Set and create result directory folder
> resultDir <- "/path/to/result/directory/"
> dir.create(resultDir, showWarnings = FALSE, recursive = TRUE)
# Set data directory (idat and sample sheet files need to be both located in that folder)
> data.source <- "/path/to/data/directory/"
# Set directory where the output should be written to
> analysis.dir <- file.path(resultDir)
# Set directory where the report files should be written to
> report.dir <- file.path(analysis.dir, "reports")
# Initialising reports
> rnb.initialize.reports(report.dir)
# Enable multicore
> logger.start(fname = NA)
> num.cores <- 20
> parallel.setup(num.cores)
# Loading raw methylation data (idat file format) in R
> result.raw <- rnb.run.import(data.source = data.source, data.type = "infinium.idat.dir", dir.reports = report.dir)
> rnb.set.raw <- result.raw$rnb.set
# Generate quality check report
> rnb.run.qc(rnb.set.raw, report.dir)
# Set options for filtering and normalisation
> rnb.options(disk.dump.big.matrices = FALSE,
            identifiers.column = "Sample_Name",
            filtering.context.removal = c("CC", "CAG", "CAH", "CTG", "CTH", "Other"),
            filtering.snp = "5", filtering.greedycut = TRUE,
```
filtering.greedycut.pvalue.threshold = 0.01,
filtering.sex.chromosomes.removal = TRUE,
filtering.cross.reactive = TRUE,
filtering.coverage.threshold = 0,
filtering.low.coverage.masking = FALSE,
filtering.missing.value.quantile = 0,
normalization = TRUE,
normalization.method = "bmiq",
normalization.background.method = "methylumi.noob",
normalization.plot.shifts = TRUE,
qc.snp.boxplot = TRUE)

# Perform normalisation with BMIQ method
> result.norm.filt <- rnb.run.preprocessing(rnb.set.raw,
dir.reports = report.dir)
> rnb.set.norm.filt <- result.norm.filt$rnb.set
# Extract normalised methylation data (M-value)
> datMval <- as.data.frame(mval(rnb.set.norm.filt, row.names = TRUE, epsilon = 1e-05))
# Remove CpG probes with missing data in all samples
> datMval <- datMval[complete.cases(datMval), ]
# Save normalised methylation data
> write.csv(datMval, paste(resultDir, "mval_norm.csv", sep=""""), row.names = TRUE, quote = FALSE)
Appendix 2

R code used for batch effect correction.

```r
# Import data in CSV file format
> library("sqldf")
# Set and create result directory folder
> resultDir <- "/path/to/result/directory/
> dir.create(resultDir, showWarnings = FALSE, recursive = TRUE)
# Import data (in .csv format)
> Data = read.csv.sql("/path/to/data.csv", header = T)
> rownames(Data) <- gsub("\\", ",", Data$X)
> Data[1] <- NULL
# Batch effect correction with ComBat
> library("sva")
# Import phenotype file (tab separated file format) Phenotype file is in general the same as the samplesheet provided for the 450K
> Pheno <- read.table("/path/to/phenotype.txt", header = T, sep = "\t")
# Batch correction step 1 (Sentrix_ID correction)
> DataBat.1 <- ComBat(dat = Data, batch = Pheno$Sentrix_ID,
> mod = model.matrix(~as.factor(Sample_Group) +
> as.factor(Sentrix_Position) + as.factor(Age), data = Pheno))
# Batch correction step 2 (Sentrix_Position correction)
> DataBat.2 <- ComBat(dat = DataBat.1, batch =
> Pheno$Sentrix_Position, mod =
> model.matrix(~as.factor(Sample_Group) + as.factor(Age), data = Pheno))
# Batch correction step 3 (age correction)
> DataBat.3 <- ComBat(dat = DataBat.2, batch = Pheno$Age,
> mod = model.matrix(~as.factor(Sample_Group), data = Pheno))
# Save normalised and batch effect corrected methylation data
```

262
> write.csv(DataBat.3, paste(resultDir, "mval_norm_combat.csv", sep = ""), row.names = TRUE, quote = FALSE)
Appendix 3

R code used for genomic imprinting methylation profiling.

```r
#### Set variables below ####
# Need to be edited Character vector indicated group of interest. It has to match with the "Disease" column in phenotype file (i.e. Pheno$Disease)
GroupOfInterest <- c("Patient")
# Character vector indicated from what standard deviation will be calculated from. It has to match with the "Disease" column in phenotype file (i.e. Pheno$Disease)
GroupOfControl <- "Control"
# Indicate standard deviation threshold
nbSd <- 3
# Fill with target names to remove from analysis. Can be empty
TargetToRemove <- c("MEG3-DLK1_IG-DMR")
# Indicate where results should be saved
resultDir <- "/path/to/result/directory/"
# Indicate where methylation data is located (CSV file format, row: CpG probes, column: samples)
dataFile <- "/path/to/data.csv"
# Indicate where phenotype file is located (tab separated file)
phenoFile <- "/path/to/phenotype.txt"
# Indicate where target file (i.e. region of interests) is located (bed file format with .txt file extension). Column names need to include 'chr', 'start', 'end'
targetFile <- "/path/to/target.file.txt"
#### Set variables above ####

#### Import files Create result directory folder

```
dir.create(resultDir, showWarnings = FALSE, recursive = TRUE)
# Import phenotype file
Pheno <- read.table(phenoFile, header = T, sep = "\t")
# Import target file
target <- read.table(targetFile, header = TRUE, sep = "\t")
target$chr <- paste("chr", target$chr, sep = "")
library("GenomicRanges")
grTarget <- makeGRangesFromDataFrame(target, seqnames.field = "chr", start.field = "start", end.field = "end",
keep.extra.columns = TRUE, ignore.strand = TRUE, seqinfo = NULL)
# Set final target specify
FinalTarget <- as.character(grTarget$Name[!grTarget$Name %in% TargetToRemove])
FinalTarget <- factor(FinalTarget, levels = unique(FinalTarget))
# Import data file
library("sqldf")
Data = read.csv.sql(dataFile, header = T)
rownames(Data) <- gsub(""", "", Data$X)
Data[1] <- NULL
# Convert data in Beta value if necessary
library("lumi")
Data <- as.data.frame(m2beta(Data))

##### Convert methylation data into GenomicRanges object
library("FDb.InfiniumMethylation.hg19")
library("GenomicRanges")
hm450 <- as.data.frame(get450k())
hm450 <- hm450[rownames(hm450) %in% rownames(Data), ]
hm450 <- hm450[order(hm450$seqnames, hm450$probeTarget), ]
Data.2 <- as.data.frame(Data[match(rownames(Data),
rownames(hm450)), ])
Data.2$chr <- hm450$seqnames
# Df.2$chr <- gsub('chr', '', Data.2$chr) #Need to un-comment
if chromosome is indicated as 1,2,3,... and not chr1, chr2, chr3 in target file
Data.2$start <- hm450$probeTarget
gpData.2 <- makeGRangesFromDataFrame(df = Data.2,
seqnames.field = "chr", start.field = "start", end.field = "start", keep.extra.columns = TRUE, ignore.strand = TRUE, seqinfo = NULL)

##### Subset data regarding target files
dir.create(paste(resultDir,
"Methylation.analysis/meth.score/", sep = ""), showWarnings = FALSE, recursive = TRUE)
setwd(paste(resultDir, "Methylation.analysis/meth.score/", sep = ""))
# Extract and save CpG beta value for all samples within
regions of interest
meth.score <- list()
for (i in seq(grTarget)) {
  x <- as.data.frame(subsetByOverlaps(gpData.2, grTarget[i]))
  x <- x[, -(1:5)]
  write.table(x, paste(resultDir, "Methylation.analysis/meth.score/", grTarget[i]$Name, sep = ""), sep = "\t", row.names = FALSE, quote = FALSE)
  meth.score[[i]] <- x
}
names(meth.score) <- grTarget$Name

setwd(resultDir)

##### Compute methylation index at target loci in all
individuals (patients and controls).
##### Compute 3 standard deviations from grand mean of control
individuals
library("plotrix")

# For GroupOfInterest (i.e. Patient) Put a list in a list
GroupPatient <- list()
for (j in seq(GroupOfInterest)) {
  y <- list()
  for (i in seq(grTarget)) {
    x <- meth.score[[i]][Pheno$Disease %in%
                        GroupOfInterest[j]]
    y[[i]] <- x
  }
  names(y) <- grTarget$Name
  y[TargetToRemove] <- NULL
  GroupPatient[[j]] <- y
}

names(GroupPatient) <- GroupOfInterest

# Calculate mean for all patients
GroupPatientMean <- list()
for (j in seq(GroupOfInterest)) {
  x <- as.data.frame(lapply(GroupPatient[[j]], colMeans))
  GroupPatientMean[[j]] <- x
}

# Calculate standard error for all patients
GroupPatientStdError <- list()
for (j in seq(GroupOfInterest)) {
  for (i in 1:nrow(Pheno[Pheno$Disease %in%
                        GroupOfInterest, ])) x <- lapply(GroupPatient[[j]],
                                                   std.error)
  x <- do.call(cbind, x)
  x[is.na(x)] <- 0
  GroupPatientStdError[[j]] <- x
}

names(GroupPatientStdError) <- GroupOfInterest
# For GroupOfControl (i.e. Control)
GroupControl <- list()
for (i in seq(grTarget)) {
  x <- meth.score[[i]][Pheno$Disease %in% GroupOfControl]
  GroupControl[[i]] <- x
}
names(GroupControl) <- grTarget$Name
GroupControl[TargetToRemove] <- NULL

# Compute mean of control group, grand mean of control group and standard deviation
GroupControlMean <- lapply(GroupControl, colMeans)
GroupControlMean <- do.call(cbind, GroupControlMean)
GroupControlGrandMean <-
as.data.frame(colMeans(GroupControlMean))
rownames(GroupControlGrandMean) <- FinalTarget

library("matrixStats")
GroupControlSd <- as.data.frame(colSds(GroupControlMean))
rownames(GroupControlSd) <- FinalTarget

# Create data.list table with 'patient.Mean', 'patient.StdError', 'control.GrandMean' and 'control.Stdeviation'
DataMI.P <- list()
for (j in seq(GroupOfInterest)) {
  for (i in 1:length(FinalTarget)) {
    y <- cbind(GroupPatientMean[[j]][, i],
               GroupPatientStdError[[j]][, i],
               GroupControlGrandMean[i, ],
               GroupControlSd[i, ])
    colnames(y) <- c("patient.Mean", "patient.StdError", "control.GrandMean", "control.Stdeviation")
    DataMI.P[[i]] <- as.data.frame(y)
  }
}

268
names(DataMI.P) <- FinalTarget

#### Compute hypomethylation (i.e. LOM) and hypermethylation (i.e. GOM) in patient group.
#### Standard deviation threshold: cf. nbSd variable above.
#### 60% of interrogated probes need to be over or below Sdev threshold
dir.create(paste(resultDir, "Methylation.analysis/epimutation/", sep = ""), showWarnings = FALSE, recursive = TRUE)
setwd(paste(resultDir, "Methylation.analysis/epimutation/", sep = ""))

PatientEpimut <- list()
y <- data.frame(matrix(NA, nrow = nrow(Pheno[Pheno$Disease %in% GroupOfInterest, ]), ncol = 1))
for (i in 1:length(FinalTarget)) {
    for (j in 1:nrow(Pheno[Pheno$Disease %in% GroupOfInterest, ])) {
        for (k in 1:length(GroupOfInterest)) y[j, 1] <-
            ifelse((DataMI.P[[i]][j, 1] > DataMI.P[[i]][j, 3] +
                DataMI.P[[i]][j, 4] * nbSd & sum(GroupPatient[[k]][[i]][, j] >
                DataMI.P[[i]][j, 3] + DataMI.P[[i]][j, 4] * nbSd &
                sum(Pheno$Disease %in% GroupOfInterest, ])
                * 100/length(GroupPatient[[k]][[i]][, j]) > 60), 1,
                ifelse((DataMI.P[[i]][j, 1] < DataMI.P[[i]][j, 3] - DataMI.P[[i]][j, 4] * nbSd &
                    sum(Pheno$Disease %in% GroupOfInterest, ])
                    * 100/length(GroupPatient[[k]][[i]][, j]) > 60), -1, 0))
            PatientEpimut[[i]] <- as.data.frame(y)
    }
}

names(PatientEpimut) <- FinalTarget
PatientEpimut <- do.call(cbind, PatientEpimut)
rownames(PatientEpimut) <- rownames(DataMI.P[[1]])
colnames(PatientEpimut) <- FinalTarget
tmpPatientMap <- PatientEpimut
tmpPatientMap[tmpPatientMap == -1] <- "LOM"
tmpPatientMap[tmpPatientMap == 1] <- "GOM"
tmpPatientMap[tmpPatientMap == 0] <- ".
write.table(tmpPatientMap, paste(resultDir, 
"Methylation.analysis/epimutation/", "Patient_epimut.txt", 
sep = ""), sep = "\t", quote = FALSE)
rm(tmpPatientMap)

# Subset control group and define hypo/hypermethylated loci in control group For GroupOfControl (ie. Control)
GroupControl <- list()
for (i in seq(grTarget)) {
    x <- meth.score[[i]][Pheno$Disease %in% GroupOfControl]
    GroupControl[[i]] <- x
}
names(GroupControl) <- grTarget$Name
GroupControl[TargetToRemove] <- NULL

# Compute mean of control group, grand mean of control group and standard deviation
GroupControlMean <- lapply(GroupControl, colMeans)
GroupControlMean <- do.call(cbind, GroupControlMean)
GroupControlMean <- as.data.frame(GroupControlMean)

GroupControlGrandMean <- as.data.frame(colMeans(GroupControlMean))

GroupControlSd <- as.data.frame(colSds(as.matrix(GroupControlMean)))
rownames(GroupControlSd) <- FinalTarget
GroupControlStdError <- lapply(GroupControl, std.error)
GroupControlStdError <- do.call(cbind, GroupControlStdError)
GroupControlStdError <- as.data.frame(GroupControlStdError)
GroupControlStdError[is.na(GroupControlStdError)] <- 0

# Create data.list table to compute 3 standard deviations
DataMI.C <- list()
for (j in seq(GroupOfControl)) {
  for (i in 1:length(FinalTarget)) {
    y <- cbind(GroupControlMean[, i],
               GroupControlStdError[, i], GroupControlGrandMean[i, ],
               GroupControlSd[i, ])
    colnames(y) <- c("control.Mean", "control.StdError",
                    "control.GrandMean", "control.Stdeviation")
    DataMI.C[[i]] <- as.data.frame(y)
  }
}
names(DataMI.C) <- FinalTarget

### Create list to store data
### Determine if Control individual are epimutated
ControlEpimut <- list()
y <- data.frame(matrix(NA, nrow = nrow(Pheno[Pheno$Disease %in% GroupOfControl, ]), ncol = 1))
for (i in 1:length(FinalTarget)) {
  for (j in 1:nrow(Pheno[Pheno$Disease %in% GroupOfControl, ])) {
                       sum(GroupControl[[i]][, j] > DataMI.C[[i]][j, 3] +
                       DataMI.C[[i]][j, 4] * nbSd) * 100/length(GroupControl[[i]][, j]) > 60), 1, ifelse((DataMI.C[[i]][j, 1] < DataMI.C[[i]][j, 3] - DataMI.C[[i]][j, 4] * nbSd &
                        DataMI.C[[i]][j, 4] * nbSd) * 100/length(GroupControl[[i]][, j]) > 60), 0)
  }
}
sum(\text{GroupControl}[[i]][, j] < \text{DataMI.C}[[i]][j, 3] - \text{DataMI.C}[[i]][j, 4] * \text{nbSd}) * 100/\text{length(\text{GroupControl}[[i]][, j]) > 60}, -1, 0)) \text{ControlEpimut}[[i]] \leftarrow y 
\}
\}
names(\text{ControlEpimut}) \leftarrow \text{FinalTarget}
\text{ControlEpimut} \leftarrow \text{do.call(cbind, ControlEpimut)}
\text{rownames(ControlEpimut)} \leftarrow \text{rownames(DataMI.C[[1]])}
\text{colnames(ControlEpimut)} \leftarrow \text{FinalTarget}
\text{tmpControltMap} \leftarrow \text{ControlEpimut}
\text{tmpControltMap[tmpControltMap == -1]} \leftarrow "LOM"
\text{tmpControltMap[tmpControltMap == 1]} \leftarrow "GOM"
\text{tmpControltMap[tmpControltMap == 0]} \leftarrow "."
\text{write.table(tmpControltMap, paste(resultDir, "Methylation.analysis/epimutation/", "Control_epimut.txt", sep = ""), sep = "\t", quote = FALSE)}
\text{rm(tmpControltMap)}

#### Create contingency table (2x2) for hypomethylated and hypermethylated loci in cases versus controls
#### Calculate Fisher's exact test P-value and FDR P-value

# Create hypomethylation contingency table
\text{hypo} \leftarrow \text{list()}
\text{for (i in 1:length(PatientEpimut)) \{}
  \text{case.hypo} \leftarrow \text{data.frame(Hypo = sum(PatientEpimut[, i] == -1), Other = sum(PatientEpimut[, i] != -1))}
  \text{control.hypo} \leftarrow \text{data.frame(Hypo = sum(ControlEpimut[, i] == -1), Other = sum(ControlEpimut[, i] != -1))}
  \text{o} \leftarrow \text{rbind(case.hypo, control.hypo)}
  \text{rownames(o)} \leftarrow \text{c("Cases", "Controls")}
  \text{colnames(o)} \leftarrow \text{c("Hypo", "Other")}
\text{\}}
hypo[[i]] <- o
}
names(hypo) <- FinalTarget
rm(case.hypo, control.hypo, o)

# Perform Fisher's exact test and FDR correction for hypomethylation observations
hypoFisherTest <- data.frame(matrix(NA, nrow = length(FinalTarget), ncol = 1))
for (i in 1:length(hypo)) {
  x <- fisher.test(hypo[[i]])
  hypoFisherTest[i, ] <- x
}

rownames(hypoFisherTest) <- names(hypo)

hypoFisherTest <- cbind(hypoFisherTest, p.adjust(hypoFisherTest$p.value, method = "BH"))

rownames(hypoFisherTest) <- c("p.value", "FDR (Benjamini & Hochberg)"
write.table(hypoFisherTest, paste(resultDir, "Methylation.analysis/epimutation/", "hypoFisherTest.txt", sep = ""), sep = "\t", quote = FALSE)

# Create hypermethylation contingency table
hyper <- list()

for (i in 1:length(PatientEpimut)) {
  case.hyper <- data.frame(Hypo = sum(PatientEpimut[, i] == 1), Other = sum(PatientEpimut[, i] != 1))
  control.hyper <- data.frame(Hypo = sum(ControlEpimut[, i] == 1), Other = sum(ControlEpimut[, i] != 1))
  o <- rbind(case.hyper, control.hyper)
  rownames(o) <- c("Cases", "Controls")
  colnames(o) <- c("Hyper", "Other")
hyper[[i]] <- o
}
names(hyper) <- FinalTarget
rm(case.hyper, control.hyper, o)

# Perform Fisher's exact test and FDR correction for
# hypermethylation observations
hyperFisherTest <- data.frame(matrix(NA, nrow = length(FinalTarget), ncol = 1))
for (i in 1:length(hyper)) {
  x <- fisher.test(hyper[[i]])
  hyperFisherTest[i, ] <- x
}
colnames(hyperFisherTest) <- "p.value"
rownames(hyperFisherTest) <- names(hyper)

hyperFisherTest <- cbind(hyperFisherTest, p.adjust(hyperFisherTest$p.value, method = "BH"))
colnames(hyperFisherTest) <- c("p.value", "FDR (Benjamini & Hochberg)")
write.table(hyperFisherTest, paste(resultDir, "Methylation.analysis/epimutation/", "hyperFisherTest.txt", sep = ""), sep = "\t", quote = FALSE)

##### Correct epimutation following hypo/hyper FDR p.value
PatientEpimutCorrected <- list()
for (i in 1:length(PatientEpimut)) {
  h <- ifelse((PatientEpimut[i] == -1 & hypoFisherTest[i, 2] < 0.05), -1, ifelse((PatientEpimut[i] == 1 & hyperFisherTest[i, 2] < 0.05), 1, 0))
  PatientEpimutCorrected[[i]] <- h
}
PatientEpimutCorrected <- as.data.frame(do.call(cbind, 
  PatientEpimutCorrected))
colnames(PatientEpimutCorrected) <- FinalTarget
rownames(PatientEpimutCorrected) <- rownames(PatientEpimut)
Appendix 4

*R code used for bump hunting in BWS_IC1. Similar code was used for other disease group.*

```r
# Import data in CSV file format
> library("sqldf")
> Data = read.csv.sql("/path/to/data.csv", header = T)
> rownames(Data) <- gsub("\"", ",", Data$X)
> Data[1] <- NULL

# Convert data in Beta value if necessary
> library("lumi")
> Data <- as.data.frame(m2beta(Data))  #Comment this line if data is b-value and not m-value

# Import phenotype file in tab separated file format
> Pheno <- read.table("/path/to/phenotype.txt", header = T, sep="\t")

# Set and create result directory folder
> resultDir <- "/path/to/result/directory/
> dir.create(resultDir, showWarnings = FALSE, recursive = TRUE)

# Import 450K annotation file and reorder methylation data
> library("FDb.InfiniumMethylation.hg19")
> hm450 <- as.data.frame(get450k())
> hm450 <- hm450[rownames(hm450) %in% rownames(Data), ]
> hm450 <- hm450[order(hm450$seqnames, hm450$probeTarget), ]
> Data.2 <- as.data.frame(Data[match(rownames(Data), rownames(hm450)), ])

# Subset data for bump hunting: cases (e.g. BWS_IC1 patients) versus controls (e.g. normal indivuals)
> DataSubset <- Data.2[colnames(Data.2) %in% Pheno$Sample_Name[Pheno$Sample_Group %in% c("BWS_IC1", "Control")]]
```
# Subset phenotype for bump hunting: cases (e.g. BWS_IC1 patients) versus controls (e.g. normal individuals)

```r
> PhenoSubset <- Pheno[Pheno$Sample_Group %in% c("BWS_IC1", "Control"), ]
> PhenoSubset <- PhenoSubset[, -c(6, 5, 4, 2)]
> PhenoSubset$Sample_Name <- factor(PhenoSubset$Sample_Name, levels = unique(PhenoSubset$Sample_Name))
> PhenoSubset$Sample_Group <- factor(PhenoSubset$Sample_Group, levels = unique(PhenoSubset$Sample_Group))

# DMR hunting

```r
> library("bumphunter")
> cl <- clusterMaker(hm450$seqnames, hm450$probeTarget, maxGap = 300)
> library("doParallel")
> registerDoParallel(cores = 10)
> designMatrix <- model.matrix(~as.factor(Sample_Group), data = PhenoSubset)
```r
> bumps <- bumphunter(object = as.matrix(DataSubset), design = designMatrix, chr = hm450$seqnames, pos = hm450$probeTarget, cluster = cl, coef = 2, cutoff = 0.15, B = 1000, maxGap = 300, permutations = TRUE, smooth = TRUE, smoothFunction = loessByCluster, useWeights = TRUE)

# Save bumps

```r
> bumps <- bumps$table
> write.table(bumps, paste(resultDir, "Bumps.txt", sep=""), sep= "\t", quote = FALSE)
```
Appendix 5

Shell script used for the trimming, alignment and methylation calling of bisulfite converted reads generated from MiSeq sequencer.

#!/bin/bash

#############################################################
# Script for trimming, alignment (hg38, hg38, rDNA) and methylation calling #
# Need to be executed in RAW folder #
# Create samples.txt file
#############################################################

# Create samples.txt file
> echo create samples name file
> ls *_L001_R1_001.fastq.gz > samples.txt
> vim -c "%s/_L001_R1_001.fastq.gz//g|wq" samples.txt

# Trimming, alignment and non-directional BS-seq library
> echo trimming
> mkdir ../trimmed &> /dev/null
> for i in `cat samples.txt`; do trim_galore --paired --clip_R2 3 ${i}_L001_R1_001.fastq.gz ${i}_L001_R2_001.fastq.gz --output_dir ../trimmed &> /dev/null
> done

# Align to hg38
> echo align to hg38
> mkdir -p ../hg38/hg38_aligned &> /dev/null
> for i in `cat samples.txt`; do bismark --bowtie2 --non_directional --bam --output_dir ../hg38/hg38_aligned --temp_dir ../temp --multicore 10 /media/Sasha/Resources/hg38/
-1 ../trimmed/${i}_L001_R1_001_val_1.fq.gz -2
../trimmed/${i}_L001_R2_001_val_2.fq.gz &> /dev/null
> done

> echo methylation call
> mkdir -p ../hg38/hg38_methCall &> /dev/null
> for i in `cat samples.txt`; do
bismark_methylation_extractor --bedGraph --p --merge_non_CpG
--report --output ../hg38/hg38_methCall --multicore 10 --gzip
../hg38/hg38_aligned/${i}_L001_R1_001_val_1.fq.gz_bismark_bt2
_pe.bam &> /dev/null
> done
Appendix 6

*R code used for the quantification of 5-mC at H19/IGF2:IG-DMR; KCNQ1OT1:TSS-DMR and IGS-rDNA in each sample.*

```r
## Script to quantify methylation in target regions (ie. DMRs or IGS-rDNA)
##
## Execute R script in parent folder of RAW, hg19, hg38 and rDNA folders
library(GenomicRanges)
library(ggplot2)
library(matrixStats)
library(plotrix)
################################
################################
## Result for hg38
################################
################################
setwd("../hg38/hg38_methCall")
temp = list.files(pattern="*bismark.cov")
# get sample name
sample_name <- sapply(strsplit(temp, "-_"), "]", 1)
sample_name <- as.factor(sample_name)
# read all table in the folder and store them in a list
methCall = lapply(temp, read.table, sep = "\t", header = FALSE)
setwd("../")

# Convert methylation call file into GenomicRanges
for (i in 1:length(methCall)) {
    methCall[[i]] <- GRanges(methCall[[i]]$V1,
    IRanges(methCall[[i]]$V2, methCall[[i]]$V3),
```
meth=methCall[[i]]$V4, countMeth=methCall[[i]]$V5, countUnmeth=methCall[[i]]$V6)
}
names(methCall) <- sample_name

#Import target file and convert to GenomicRanges
target <- read.table("../hg38_target", sep="\t", header=F)
target <- GRanges(target$V1, IRanges(target$V2, target$V3, names=target$V4))

#Find overlap between methylation call file and target file for all samples and convert into data frame object
methSubset <- list()
for (i in 1:length(methCall)) {
methSubset[[i]] <- subsetByOverlaps(methCall[[i]], target)
#Add DMR names in overlap object
overlaps_tmp <- findOverlaps(methCall[[i]], target)
match_hit <-
data.frame(names(target)[subjectHits(overlaps_tmp)], stringsAsFactors=F)
names(match_hit) <- "DMR"
mcols(methSubset[[i]]) <- c(mcols(methSubset[[i]]), match_hit)
methSubset[[i]] <- as.data.frame(methSubset[[i]])
methSubset[[i]] <-
methSubset[[i]][!rowSums(data.frame(methSubset[[i]]$countMeth, methSubset[[i]]$countUnmeth)) < 1000,]
}
names(methSubset) <- sample_name

#Calculate mean and std.error for samples at DMRs
x <- data.frame(matrix(ncol = 4, nrow = length(target)))
y <- list()
for (i in 1:length(methSubset)) {
  for (j in 1:length(target)) {

x[j,] <- c(sample=names(methSubset[i]),
locus=names(target[j]),
mean=mean(methSubset[[i]]$meth[methSubset[[i]]$DMR %in%
names(target[j]))),
std.error=std.error(methSubset[[i]]$meth[methSubset[[i]]$DMR
%in% names(target[j])))
y[[i]] <- x
}

methAnalysis <- do.call(rbind,y)
methAnalysis <- data.frame(as.factor(methAnalysis[,1]),
as.factor(methAnalysis[,2]), as.numeric(methAnalysis[,3]),
as.numeric(methAnalysis[,4]))
colnames(methAnalysis) <- c("sample", "locus", "mean",
"std.error")

sample.order <- c("0meth", "25meth", "50meth", "75meth",
"100meth", "sample01", " sample02", " sample03", "sample04")

#Barplot methylation level
dir.create("result", showWarnings = FALSE, recursive = TRUE)
setwd("result")
for (i in 1:length(target)) {
q <- ggplot(methAnalysis[methAnalysis$locus ==
names(target[i]),], aes(x=sample, y=mean)) +
  geom_bar(position=position_dodge(), stat="identity",
  fill="#FF9999", colour="black", size=.3) +
  geom_errorbar(aes(ymin=mean-std.error,
  ymax=mean+std.error), size=1, width=.2,
  position=position_dodge(.9)) +
  ggtitle(paste("Methylation index at",
  names(target[i]),\n  "Average nb CpG:",round(sum(do.call(rbind,
  methSubset)$locus==
  names(target[i]))/length(sample_name),0))) +
  ylab('Methylation level (%)') +

```r
theme(panel.grid.minor.y = element_blank()) +
theme(plot.title=element_text(face="bold", hjust=0.5)) +
theme(axis.text.x = element_text(angle = 45, hjust=1, face = "bold", colour = "black")) +
theme(axis.text.y = element_text(face = "bold", colour = "black")) +
theme(axis.title.x = element_blank()) +
scale_y_continuous(limits=c(0,110), expand = c(0, 0)) +
scale_x_discrete(limits = sample.order)
ggsave(filename=paste("MI at_", names(target[[i]]), ".tiff", sep=""", plot=q, dpi=600, compression="lzw")
}

#############################
## FOR IGS-rDNA anaylsis ONLY
#############################
#import file folder
setwd("rDNA/rDNA_methCall")
temp = list.files(pattern="*bismark.cov")
#get sample name
sample_name <- sapply(strsplit(temp, "_"), ";","[", 1)
sample_name <- as.factor(sample_name)
#read all table in the folder and store them in a list
methCall = lapply(temp, read.table,sep = "\t", header = FALSE)
setwd("./")

#Re-name column name in methCall object and remove CpG with low count reads (ie. total < 1000)
for (i in 1:length(methCall)) {
  colnames(methCall[[i]]) <- c("chr", "start", "end", "meth", "countMeth", "countUnmeth")
  methCall[[i]] <- methCall[[i]][!rowSums(data.frame(methCall[[i]]$countMeth, methCall[[i]]$countUnmeth)) < 1000,]
}
```
names(methCall) <- sample_name

#Calculate mean and std.error for samples at DMRs
x <- data.frame(matrix(ncol = 4, nrow = 1))
y <- list()
for (i in 1:length(methCall)) {
  x <- c(sample=names(methCall[[i]]), locus="IGS-rDNA",
         mean=mean(methCall[[i]]$meth),
         std.error=std.error(methCall[[i]]$meth))
  y[[i]] <- x
}

methAnalysis <- do.call(rbind,y)
methAnalysis <- data.frame(as.factor(methAnalysis[,1]),
                           as.factor(methAnalysis[,2]), as.numeric(methAnalysis[,3]),
                           as.numeric(methAnalysis[,4]))
colnames(methAnalysis) <- c("sample", "locus", "mean",
                           "std.error")

sample.order <- c("0meth", "25meth", "50meth", "75meth",
                   "100meth", "sample01", "sample02", "sample03", "sample04")

#Barplot methylation level
dir.create("result", showWarnings = FALSE, recursive = TRUE)
setwd("result")
q <- ggplot(methAnalysis, aes(x=sample, y=mean)) +
  geom_bar(position=position_dodge(), stat="identity",
           fill="#FF9999", colour="black", size=.3) +
  geom_errorbar(aes(ymin=mean-std.error,
                    ymax=mean+std.error), size=1, width=.2,
                position=position_dodge(.9)) +
  ggtitle(paste("Methylation index at IGS-rDNA", 
                "\nAverage nb CpG: ", round(nrow(do.call(rbind,
                methCall))/length(sample_name),0))) +
  ylab('Methylation level (%)') +
theme(panel.grid.minor.y = element_blank()) +
theme(plot.title=element_text(face="bold", hjust=0.5)) +
theme(axis.text.x = element_text(angle = 45, hjust=1, face = "bold", colour = "black")) +
theme(axis.text.y = element_text(face = "bold", colour = "black")) +
theme(axis.title.x = element_blank()) +
scale_y_continuous(limits=c(0,110), expand = c(0, 0)) +
scale_x_discrete(limits = sample.order)
ggsave(filename=paste("MI at IGS-rDNA", ".tiff", sep=""),
plot=q, dpi=600, compression="lzw")

write.table(methAnalysis, "methTable_IGS.txt", sep="\t",
quote=F, row.names=F)
Appendix 7

*R code used for the PCR amplification bias correction.*

#Import data
Data <- read.table("/Users/NoName/Desktop/test.txt", sep='\t', header=T, row.names = 1)

#Create empty DataCorrected object
DataCorrected <- data.frame(matrix(ncol = length(Data), nrow=nrow(Data)))
rownames(DataCorrected) <- rownames(Data)
colnames(DataCorrected) <- colnames(Data)

#Theoric values
ytheo <- c(0, 25, 50, 75, 100)
xtheo <- c(0, 25, 50, 75, 100)

for (i in 1:nrow(Data)){
  #Standard Curve
  yobserved <- as.numeric(Data[i,1:5])
xexpected <- c(0, 25, 50, 75, 100)

  ### Methylation correction with cubic polynomiale regression
  #Model
  fit1 <- lm(xexpected ~ poly(yobserved,3, raw = TRUE))
  #Correct standard curve
  SCcorrected <- predict(fit1,data.frame(yobserved))
  #Correct Samples
  Samples <- as.numeric(Data[i,6:length(Data)])
  SamplesCorrected <- predict(fit1, data.frame(yobserved=Samples))
  SamplesCorrected[SamplesCorrected<0] <- 0
  #Store corrected values
DataCorrected[i,] <- c(SCcorrected, SamplesCorrected)

### Plot standard curve before and after correction

#### BEFORE correction
#bias
bias <- (yobserved * (100 - xexpected))/(xexpected * (100 - yobserved))
bias[bias==Inf] <- 0
bias[bias==-Inf] <- 0
bias[is.na(bias)] <- 0
bias <- bias[!bias==0]
MeanBias1 <- mean(bias)

#predicted curve for uncorrected standard curve
q <- seq(from=0, to=100, by=0.1)
fit2 <- lm(yobserved ~ poly(xexpected, 3, raw = TRUE))
R.SQUARE1 <- summary(fit2)$r.squared
pred1 <- predict(fit2, data.frame(xexpected=q))

#### AFTER correction
#bias
bias <- (SCcorrected * (100 - xexpected))/(xexpected * (100 - SCcorrected))
bias[bias==Inf] <- 0
bias[bias==-Inf] <- 0
bias[is.na(bias)] <- 0
bias <- bias[!bias==0]
MeanBias2 <- mean(bias)

#predicted curve for uncorrected standard curve
q <- seq(from=0, to=100, by=0.1)
fit3 <- lm(SCcorrected ~ poly(xexpected, 3, raw = TRUE))
R.SQUARE2 <- summary(fit3)$r.squared
pred2 <- predict(fit3, data.frame(xexpected=q))

#plot
pdf(paste(rownames(Data)[i],"_bias.pdf", sep=""), height = 5, width = 10)
par(mfrow=c(1,2))
plot(yobserved~xexpected, col="red", ylim=c(0,100), xlim=c(0,100), xlab="Actual methylation (%)", ylab="Uncorrected estimated methylation (%)", main="Standard curve before correction")
lines(x=q, y=pred1, col="red")
lines(ytheo~xtheo, lty=2)
text(x=10,y=98,labels = paste("b=", round(MeanBias1,digits=2), sep=""), col="red")
text(x=80,y=0, labels= bquote(R^2 == .(round(R.SQUARE1,digits=3))))

plot(SCcorrected~xexpected, col="red", ylim=c(0,100), xlim=c(0,100), xlab="Actual methylation (%)", ylab="Corrected estimated methylation (%)", main="Standard curve after correction")
lines(x=q, y=pred2, col="red")
lines(ytheo~xtheo, lty=2)
text(x=10,y=98,labels = paste("b=", round(MeanBias2,digits=2), sep=""), col="red")
text(x=80,y=0, labels= bquote(R^2 == .(round(R.SQUARE2,digits=3))))
dev.off()
###
}

write.table(DataCorrected,"DataCorrected.txt", quote=FALSE, sep='\t')
Chapter 9

List of References


next-generation capture sequencing identifies novel disease-associated variants.

Arima T. 2005. ZAC, LIT1 (KCNQ1OT1) and p57KIP2 (CDKN1C) are in an imprinted gene network that may play a role in Beckwith-Wiedemann syndrome. Nucleic Acids Research 33:2650–2660.


Barlow DP, Stöger R, Herrmann BG, Saito K, Schweifer N. 1991. The mouse insulin-
like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature 349:84–87.


Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. 2001. Dnmt3L and the


Chen J, Bardes EE, Aronow BJ, Jegga AG. 2009. ToppGene Suite for gene list


51:229–238.


a network to decipher their aetiology and to improve the diagnostic and clinical care. Clin Epigenetics 7:23.


23:218–220.


Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2


defined by Beckwith-Wiedemann syndrome rearrangement breakpoints and subchromosomal transferable fragments. Proc Natl Acad Sci USA 92:12456–12460.


proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333:1300–1303.


Joyce CA, Sharp A, Walker JM, Bullman H, Temple IK. 1999. Duplication of 7p12. 1-
p13, including GRB10 and IGFBP1, in a mother and daughter with features of Silver-


Karmakar S, Das C. 2002. Regulation of Trophoblast Invasion by IL-1beta and TGF-


Komori HK, LaMere SA, Torkamani A, Hart GT, Kotsopoulos S, Warner J, Samuels


antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. Proc Natl Acad Sci USA 96:5203–5208.


Li L, Baibakov B, Dean J. 2008. A subcortical maternal complex essential for


Mackay DJG, Hahnemann JMD, Boonen SE, Poerksen S, Bunyan DJ, White HE, Durston VJ, Thomas NS, Robinson DO, Shield JPH, et al. 2006. Epimutation of the


Marlow FL. 2010. Maternal Control of Development in Vertebrates: My Mother Made Me Do It!


Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a


Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K,


Russell A. 1954. A syndrome of intra-uterine dwarfism recognizable at birth with


Journal of Medical Genetics 44:59–63.


specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. Cell 73:61–71.


methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. Epigenetics Chromatin 4:1.


