Structural Maintenance of Chromosome Hinge Domain Containing 1 (SMCHD1) regulates gene expression

by
Shabnam Massah

B.Sc. (Microbiology and Immunology), Azad University, 2006

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

in the Doctor of Philosophy Program Faculty of Health Sciences

© Shabnam Massah 2017
SIMON FRASER UNIVERSITY Fall 2017

All rights reserved. However, in accordance with the Copyright Act of Canada, this work may be reproduced, without authorization, under the conditions for Fair Dealing. Therefore, limited reproduction of this work for the purposes of private study, research, education, satire, parody, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
Approval

Name: Shabnam Massah
Degree: Doctor of Philosophy (Health Sciences)
Title: Structural Maintenance of Chromosome Hinge Domain Containing 1 (SMCHD1) Regulates Gene Expression
Examine Committee: Chair: Tania Bubela
Gratien Prefontaine
Senior Supervisor
Associate Professor
Tim Beischlag
Co-Supervisor
Professor
Esther Verheyen
Supervisor
Professor
Department of Molecular Biology and Biochemistry
Michel Leroux
Supervisor
Professor
Department of Molecular Biology and Biochemistry
Colby Zaph
Supervisor
Professor
Department of Biochemistry and Molecular Biology
Monash University
Nadine Provencal
Internal Examiner
Assistant Professor
Carolyn Brown
External Examiner
Professor
Department of Medical Genetics
University of British Columbia
Date Defended/Approved: November 17, 2017
Abstract

Eukaryotic cells evolved by packaging genomic DNA into chromatin where DNA is wrapped around histones. This significantly reduces random transcriptional events by providing a barrier for gene expression. In addition, chemical modifications of histones and cytosine residues on DNA greatly impact regulation of gene expression. Structural maintenance of chromosome hinge domain containing 1 (SMCHD1) is a chromatin modifier. SMCHD1 was originally recognized as essential for X chromosome inactivation and survival in female mice where it plays a critical role in methylation of a subset of CpG islands. Structural studies suggest that SMCHD1 interaction with HP1 binding protein, HBiX1, mediates heterochromatin formation over the X chromosome by linking two chromatin domains enriched for repressive histone marks. In addition, loss of SMCHD1 is lethal in male mice in a mixed background, implying that SMCHD1 regulates genes on non-sex chromosomes. Thus, we identified a need to investigate the role of SMCHD1 in regulating expression of autosomal genes. In addition, I sought to determine SMCHD1 genome occupancy when global DNA methylation was greatly reduced and identify candidate binding partners. I used shRNA technology to knockdown SMCHD1 expression and identified genes that were up and down regulated in human embryonic kidney (HEK293) cells. A number of these genes are expressed in either a stochastic or parent-of-origin monoallelic fashion. Using chromatin immunoprecipitation-sequencing (ChIP), I identified genome-wide occupancy of SMCHD1 and showed that its genomic binding was sensitive to the DNA demethylating reagent, 5-azacytidine. SMCHD1 occupancy correlates with a number of genes belonging to the G protein-coupled receptor superfamily and loss of SMCHD1 in human neuroblastoma SH-SY5Y cells leads to increased levels of cellular cAMP. In addition, loss of SMCHD1 increases KCNQ1 expression, a subunit of the potassium voltage gated channel that plays a critical role in repolarization of the cardiac action potential. Moreover, using tandem tagged affinity purification, I investigated binding partners that potentially interact with SMCHD1 to regulate gene expression. Taken together, SMCHD1 might be involved in variety of diseases including Facioscapulohumeral Muscular Dystrophy (FSHD) and Bosma Arhinia Microphthalmia Syndrome (BAMS).
Keywords: SMCHD1; CpG methylation; monoallelic; cAMP; KCNQ1; Tap-tag
To my husband, Payam, my best friend and soul mate

and

To our son, Aryo, our bundle of joy, love and happiness
Acknowledgements

My journey at SFU has been exciting, encouraging and life changing. I would like to express my sincerest gratitude to my senior supervisor, Dr. Gratien Prefontaine for believing in me and providing the immense opportunity to learn and grow as a scientist. I am grateful for his guidance and mentorship throughout these years. Without his support and encouragement, none of this would have been possible. I am thankful to Dr. Tim Beischlag for his endless support, mentorship and guidance over the years. My sincere thanks also goes to Dr. Esther Verheyen, Dr. Michel Leroux and Dr. Colby Zaph for their guidance, valuable feedbacks and helping me to stay on track over the years. I am grateful to Dr. Frank Lee for his insightful comments, continuous support and encouragement. I would like to extend my thanks to my internal examiner, Dr. Nadine Provencal, and my external examiner Dr. Carolyn J. Brown for taking the time to read my thesis and for sitting on my examining committee.

I would like to thank all the past and present members of the Prefontaine lab, first and foremost, Rob Hollebakken, Dr. Addie Kolybaba and Dr. Gabrielle Rabu who were my first mentors in the lab and supported me over the years. I also want to express my deep gratitude to my research family and life-long friends, Mandeep Takhar, Nicole Bance, Shazina Khan, Samaneh Khakshour, Hadi Esmaeilsabzali, Mark Labrecque, Julienne Jagdeo, Kevin Tam, Rebecca Nason, Robert Payer, Alex Reers and Lillian Wong.

Finally, but most importantly I would like to thank my family. This journey would not have been possible without the support and encouragement of my husband, son, mom, dad, sister and grandma. Thank you for believing in me. Your unconditional love inspires me to work hard and follow my dreams.
# Table of Contents

Approval ................................................................. i  
Abstract ............................................................... iii  
Acknowledgements ..................................................... vi  
Table of Contents ..................................................... vii  
List of Tables .......................................................... xi  
List of Figures .......................................................... xii  
List of Acronyms ...................................................... xv  

## Chapter 1. General Introduction ........................................... 1

1.1. DNA packaging and chromatin structure ........................................ 1  
   1.1.1. Chromatin remodeling proteins ............................................. 3  
   1.1.2. Chromosome organizers ...................................................... 5  
      1.1.2.1. Structural maintenance of chromosome (SMC) complexes and their function ....................................................... 5  
      1.1.2.1.1. Cohesin ............................................................... 7  
      1.1.2.1.2. Condensin ......................................................... 9  
      1.1.2.1.3. The SMC5/6 complex .............................................. 10  
   1.1.2.2. Structural maintenance of chromosome hinge domain containing 1 (SMCHD1) ................................................................. 10  
   1.1.3. Epigenetic modifications and chromatin structure ....................... 11  
      1.1.3.1. Histone modifications ............................................... 12  
      1.1.3.2. Histone modification and transcriptional regulation ............... 13  
      1.1.3.3. DNA methylation ...................................................... 14  
      1.1.3.4. DNA demethylation .................................................... 16  
1.2. Genes regulated by SMCHD1 ............................................... 17  
   1.2.1. X chromosome inactivation ............................................... 17  
      1.2.1.1. Regulation of X chromosome inactivation ........................... 17  
      1.2.1.2. The X inactive specific transcript (Xist) ncRNA ................ 21  
      1.2.1.3. The Tsix ncRNA ..................................................... 22  
      1.2.1.4. Chromatin modifications associated with XCI ....................... 23  
      1.2.1.5. The role of DNA methylation in XCI ................................ 24  
      1.2.1.6. The role of regulatory proteins in XCI .............................. 25  
      1.2.1.7. The role of SMCHD1 in XCI ........................................ 26  
   1.2.2. Genomic imprinting and regulation of imprinted genes .................. 29  
   1.2.3. KCNQ1 imprinted locus, its epigenetic regulation and associated diseases ................................................................. 31  
      1.2.3.1. KCNQ1 role in mediating cellular cAMP production by adenylyl cyclases and associated diseases ........................................ 33  
      1.2.3.2. Adenosine 3',5'-cyclic monophosphate (cAMP) ...................... 33  
      1.2.3.3. The SNRPN imprinted locus is associated with PWS and AS ........ 34  
      1.2.3.4. DNA methylation and regulation of the SNRPN locus ................ 35  
      1.2.3.5. Histone tail modifications in regulation of the SNRPN locus ....... 36  
      1.2.3.6. The proteins involved in regulation of SNRPN imprinted locus .... 37  
      1.2.3.7. The role of SMCHD1 in regulating the expression of imprinted genes ................................................................. 38  
   1.2.4. Clustered protocadherin genes, their genomic organization and epigenetic regulation .................................................. 39  

vii
2.6.4. Pituitary imaging and Tissue immunofluorescence ............................................. 83
2.6.5. Manual bisulfite sequencing and CoBRA .......................................................... 83
2.6.6. Reverse transcription quantitative PCR .............................................................. 84
2.6.7. SDS-PAGE and immunoblot ............................................................................. 84
2.6.8. Chromatin Immunoprecipitation Assay ............................................................ 85
2.6.9. Preparation of nuclear extract and electrophoretic mobility shift assay (EMSA) .......................................................................................................................... 86
2.6.10. Large-scale protein purification ....................................................................... 87
2.6.11. Short hairpin knock-down of SMCHD1 in cells ............................................. 87
2.6.12. mRNA Microarray analysis ............................................................................ 88

2.7. Supplementary information .................................................................................. 90

Chapter 3. SMCHD1 is a regulator of forskolin-induced cellular cAMP levels

3.1. Abstract .................................................................................................................. 101
3.2. Introduction ........................................................................................................... 102
3.3. Results ................................................................................................................... 103
3.3.1. Identification of genome-wide occupancy of SMCHD1 that is sensitive to 5azaC .................................................................................................................. 104
3.3.2. SMCHD1 binds majorly over intron and intergenic regions and its occupancy overlaps with binding sites for specific classes of transcription factors .................................................................................................................. 107
3.3.3. SMCHD1 binds near genes acting in G-protein coupled receptor signalling pathways .................................................................................................................. 110
3.3.4. SMCHD1 knock-out upregulates KCNQ1 expression level ................................ 116
3.4. Discussion ............................................................................................................. 118
3.5. Materials and methods ....................................................................................... 121
3.5.1. Cells, Antibodies and Reagents ...................................................................... 121
3.5.2. Cell culture and 5azaC treatment ..................................................................... 122
3.5.3. SDS-PAGE and immunoblot .......................................................................... 122
3.5.4. ChIPSeq antiserum development ................................................................... 122
3.5.5. ChIP-seq assay ............................................................................................... 123
3.5.6. ChIP-seq data analysis .................................................................................... 124
3.5.7. CRISPR knockout of SMCHD1 in cells ......................................................... 125
3.5.8. Reverse transcription quantitative PCR ......................................................... 125
3.5.9. Cyclic AMP Immunoassay ............................................................................. 125
3.6. Supplementary information .................................................................................. 127

Chapter 4. Identification of SMCHD1 candidate binding partners ......................... 134

4.1. Introduction ........................................................................................................... 134
4.2. Creating a tool for identification of SMCHD1 binding partners ....................... 135
4.2.1. Tandem affinity purification (TAP) for isolation of SMCHD1 and its binding partners ........................................................................................................ 135
4.3. Results ................................................................................................................... 136
4.3.1. Sub-cloning. Strep II-FLAG (SF) tag to the N-terminus of SMCHD1 cDNA clone and generation of a modified HEK293 cell population that stably expresses SF-SMCHD1 ............................................................... 136
4.3.2. Identification of SMCHD1 potential binding partners .................................... 137
List of Tables

Table 1-1. The components of the SMC complexes in different species .................... 7
Table 1-2. Summary of repressors and activators of XCI ........................................... 18
Table 2-1. Oligonucleotides used for the EMSAs ....................................................... 96
Table 2-2. Genes identified with matched peptides from the DNA affinity purification assay using the methylated GH DMR and LC-MS. ................... 96
Table 2-3. Genes with monoallelic expression differentially regulated by SMCHD1 ........................................................................................................................... 97
Table 2-4. Primers used for ChIP/ChIP COBRA assay ................................................. 98
Table 2-5. Primers used for RT-PCR ............................................................................ 99
Table 2-6. Primers used for construction of the DNA affinity matrix ......................... 100
Table 2-7. Primers used for construction of shRNA KD ............................................... 100
Table 3-1. Oligonucleotides used for RT-qPCR and CRISPR KO .................................. 133
Table 4-1. Top candidates of SMCHD1 binding partners from the first mass spectrometry analysis ................................................................................................. 138
Table 4-2. Top candidates of SMCHD1 binding partners from the second mass spectrometry analysis ................................................................................................. 139
Table 4-3. Functional classification of SMCHD1 candidate interacting proteins ................................................................................................................................. 149
Table 4-4. First mass spectrometry analysis of SF-SMCHD1 immunoprecipitation .......................................................... 160
Table 4-5. Second mass spectrometry analysis of SF-SMCHD1 immunoprecipitation ................................................................................................................................. 162
Table 5-1. Summary of SMCHD1 mutations identified in FSHD patients ................. 167
# List of Figures

| Figure 1-1. | DNA packaging | 2 |
| Figure 1-2. | Architecture of SMC proteins | 6 |
| Figure 1-3. | A model for transcriptional regulation driven by cohesion | 8 |
| Figure 1-4. | DNA supercoiling by condensing | 9 |
| Figure 1-5. | The gene silencing function of condensin | 10 |
| Figure 1-6. | Structure of SMCHD1 protein | 11 |
| Figure 1-7. | DNA and histone modifications | 12 |
| Figure 1-8. | The X inactivation center is enriched in genes that encode for ncRNAs and are involved in X chromosome inactivation | 18 |
| Figure 1-9. | Pairing/counting determines the number of X chromosomes prior to inactivation | 20 |
| Figure 1-10. | Two distinct heterochromatin regions form a compact structure during XCI. Left panel is an illustration of a hypothetical model showing formation of two distinct domains | 28 |
| Figure 1-12. | ICR1 and KvDMR1; two imprinted control regions with distinct modes of regulation. A schematic illustration of mouse Kcnq1 and H19/Igf2 imprinted domains | 31 |
| Figure 1-13. | Imprinted control regions and antisense transcripts regulate the SNRPN imprinted locus | 37 |
| Figure 1-14. | Clustered protocadherin genes are arrayed in tandem on the same chromosome in the mammalian genome | 44 |
| Figure 2-1. | Identification of a GH promoter DMR using mouse BAC transgenes | 58 |
| Figure 2-2. | Comparative levels of GH promoter methylation in mice | 59 |
| Figure 2-3. | Inhibition of DNA methylation relieves transcriptional repression of GH and characterization of a methyl-DNA binding activity | 62 |
| Figure 2-4. | DNA affinity purification of SMCHD1 using a multimerized version of the methylated GH DMR | 67 |
| Figure 2-5. | Profiling gene expression from cells with knock-down levels of SMCHD1 | 70 |
| Figure 2-6. | The protocadherin β cluster genes were differentially regulated in SMCHD1 knock-down cells | 74 |
| Figure 2-7. | The H19/Igf2 imprinted locus was dis-regulated following SMCHD1 knock-down | 76 |
Figure 2-8. Comparison of the pyrosequencing results from WT-GH:RFP and ΔLCR-GH:RFP mouse pituitary. .................................................................90
Figure 2-9. SMCHD1 is bound to the GH promoter in cells in a DNA methylation dependent fashion. .................................................................91
Figure 2-10. SMCHD1 knockdown in HEK293 cells resulted in modified expression of genes when compared to control cells. .........................92
Figure 2-11. PCDHB10 up-regulation during SMCHD1 knockdown does not correlate with changes in promoter DNA methylation .........................94
Figure 2-12. The DNA used for affinity purification was resistant to digestion with Hgal, a DNA methylation sensitive enzyme. .................................95
Figure 3-1. Workflow for SMCHD1 ChIP-seq analysis in SH-SY5Y cells. ..........107
Figure 3-2. Genome-wide analysis of SMCHD1 binding sites in SH-SY5Y cells ........................................................................................................109
Figure 3-3. GO biological processes and functional clustering of SMCHD1 peaks and putative target genes ................................................................111
Figure 3-4. Forskolin stimulated cAMP levels and potentially G-protein coupled receptor signaling are altered in cells lacking SMCHD1. ..........114
Figure 3-5. Loss of SMCHD1 increases KCNQ1 expression level in SH-SY5Y cells ........................................................................................................116
Figure 3-6. Increased expression of KCNQ1 gene upon SMCHD1 KO is due to up-regulation of both alleles .................................................................117
Figure 3-7. 5azaC treatment induces loss of Dnmt1 level in SH-SY5Y cells. ......127
Figure 3-8. 5azaC treatment does not change SMCHD1 level in SH-SY5Y cells ........................................................................................................128
Figure 3-9. Anti-SMCHD1 antiserum efficiently immunoprecipitates SMCHD1. 129
Figure 3-10. Genome-wide analysis of SMCHD1 binding sites in cells treated with 5-azaC .............................................................................................130
Figure 3-11. Types of Transcription factor family binding sites found associated with SMCHD1 binding .........................................................................132
Figure 3-12. Loss of SMCHD1 did not change protein level of selected G protein-coupled receptor signaling proteins ........................................132
Figure 4-1. Outline of the SF-TAP procedure .................................................................................................................................136
Figure 4-2. Stable expression of SF tagged SMCHD1 in HEK293 cells. ..........137
Figure 4-3. Illustration of two co-immunoprecipitation strategies used for studying SMCHD1 and CNAP1 interaction ........................................141
Figure 4-4. Co-immunoprecipitation studies, examining SMCHD1 and CNAP1 protein interaction. .................................................................143
Figure 4-5. Co-immunoprecipitation studies, examining SMCHD1 and HBiX1 protein interaction. ................................................................. 144
Figure 4-6. Plasmid map of the SF-SMCHD1 pcDNA 3.0 clone. ......................... 152
Figure 4-7. Plasmid map of CNAP1 pcDNA 3.0 clone ................................ 155
Figure 4-8. Plasmid map of the SF-CNAP1 pcDNA 3.0 clone. ......................... 156
Figure 4-9. Plasmid map of the HA-HBiX1 pcDNA 3.0 clone ......................... 157
Figure 4-10. Plasmid map of the HBiX1–HA pcDNA 3.0 clone ....................... 158
Figure 4-11. Plasmid map of the HA-SMCHD1 pcDNA 3.0 clone .................... 159
Figure 5-1. The schematic illustration of SMCHD1 homodimer and mutations identified in FSHD and BAMS individuals ............................................. 170
Figure 5-2. SMCHD1 functional model ......................................................... 174
**List of Acronyms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-azaC</td>
<td>5-azacytidine</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine</td>
</tr>
<tr>
<td>AC</td>
<td>Adenylyl cyclase</td>
</tr>
<tr>
<td>ACF</td>
<td>Chromatin-assembly factor</td>
</tr>
<tr>
<td>ACSM2B</td>
<td>Acyl-CoA Synthetase Medium-Chain Family Member 2B</td>
</tr>
<tr>
<td>ADCY</td>
<td>adenylate cyclase</td>
</tr>
<tr>
<td>ADRA2C</td>
<td>Adrenoceptor Alpha 2C</td>
</tr>
<tr>
<td>AID</td>
<td>Activation-induced cytosine deaminase</td>
</tr>
<tr>
<td>AKAP</td>
<td>A-kinase anchor protein</td>
</tr>
<tr>
<td>ANAPC1</td>
<td>anaphase promoting complex subunit 1</td>
</tr>
<tr>
<td>ANT1</td>
<td>ADP/ATP translocase 1</td>
</tr>
<tr>
<td>APC/C</td>
<td>the anaphase-promoting complex or cyclosome</td>
</tr>
<tr>
<td>APOBEC4</td>
<td>Apolipoprotein B MRNA Editing Enzyme Catalytic Polypeptide Like 4</td>
</tr>
<tr>
<td>ARHGEF17</td>
<td>Rho Guanine Nucleotide Exchange Factor 17</td>
</tr>
<tr>
<td>ARID4A</td>
<td>AT-rich interaction domain 4A</td>
</tr>
<tr>
<td>ARID4B</td>
<td>AT-rich interaction domain 4B</td>
</tr>
<tr>
<td>AS</td>
<td>Angelman Syndrome</td>
</tr>
<tr>
<td>AS-SRO</td>
<td>AS-smallest region of deletion overlap</td>
</tr>
<tr>
<td>ASH1L</td>
<td>ASH1 Like Histone Lysine Methyltransferase</td>
</tr>
<tr>
<td>ATF5</td>
<td>Activating Transcription Factor 5</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ATPases</td>
<td>Adenosine triphosphatase, ATP hydrolase</td>
</tr>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
</tr>
<tr>
<td>BAF</td>
<td>BRG1-associated factor</td>
</tr>
<tr>
<td>BAF180</td>
<td>Polybromo BRG1-associated factor 180</td>
</tr>
<tr>
<td>BAF250</td>
<td>BRG1-associated factor 250 (AT-Rich Interaction Domain 1A)</td>
</tr>
<tr>
<td>BAMs</td>
<td>Bosma arhinia microphthalmia syndrome</td>
</tr>
<tr>
<td>BHC</td>
<td>BRAF-HDAC Complex</td>
</tr>
<tr>
<td>BHC80</td>
<td>BRF35/HDAC2 Complex (80 KDa)</td>
</tr>
<tr>
<td>BRF35</td>
<td>BRCA2-Associated Factor 35</td>
</tr>
<tr>
<td>BRG1(SMARCA4)</td>
<td>SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4</td>
</tr>
<tr>
<td>BRM(SMARCA2)</td>
<td>SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 2</td>
</tr>
<tr>
<td>BWS</td>
<td>Beckwith-Wiedemann syndrome</td>
</tr>
<tr>
<td>cAMP</td>
<td>Adenosine 3',5'-cyclic monophosphate</td>
</tr>
<tr>
<td>CAMTA1</td>
<td>Calmodulin Binding Transcription Activator 1</td>
</tr>
<tr>
<td>CARS</td>
<td>Cysteinyl-tRNA synthetase</td>
</tr>
<tr>
<td>CBP</td>
<td>Calmodulin binding affinity peptide</td>
</tr>
<tr>
<td>CCDC64</td>
<td>Coiled-Coil Domain Containing 64</td>
</tr>
<tr>
<td>CCDC87</td>
<td>Coiled-Coil Domain Containing 87</td>
</tr>
<tr>
<td>CD81</td>
<td>CD81 Molecule</td>
</tr>
<tr>
<td>CDK1</td>
<td>Cyclin-dependent kinase 1</td>
</tr>
<tr>
<td>CDKN1C</td>
<td>Cyclin-dependent kinase inhibitor 1C</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
</tbody>
</table>
CEP250  centrosome associated protein 250
CFB    Complement Factor B
CHD    Chromodomain helicase DNA binding
ChIP   chromatin immunoprecipitation
CHRAC  Chromatin accessibility complex
CHRM3  Cholinergic Receptor Muscarinic 3
CLIP-seq UV-crosslinking and immunoprecipitation deep sequencing
CNAP1  Centrosomal Nek2-Associated protein 1
CoBRA  combined bisulfite restriction analysis
Commd1 Copper Metabolism (Murr1) Domain Containing 1
CoREST REST Corepressor 1
CpG    C-phosphate-G dinucleotides
CRIP2/ESP1 cysteine rich protein 2
CRISPR Clustered regularly interspaced short palindromic repeats
CRKRS  CDK12 -Cyclin Dependent Kinase 12
CSE    conserved sequence element
CSMD1  CUB And Sushi Multiple Domains 1
CTCF   CCCTC-binding factor
CTCF   eleven-zinc finger protein
CTTNA1 Catenin Alpha 1
CTTNBP2 Cortactin Binding Protein 2
Cy3    cyanine 3
D4Z4   Macrosatellite repeats
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAVID</td>
<td>The Database for Annotation, Visualization and Integrated Discovery</td>
</tr>
<tr>
<td>DBE-T</td>
<td>D4Z4 Binding Element Transcript (Non-Protein Coding)</td>
</tr>
<tr>
<td>DIS3L2</td>
<td>DIS3 Like 3'-5' Exoribonuclease 2</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle’s Medium</td>
</tr>
<tr>
<td>DMR</td>
<td>differentially methylated region</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyrribonucleic acid</td>
</tr>
<tr>
<td>DNA-FISH</td>
<td>DNA fluorescence in situ hybridization</td>
</tr>
<tr>
<td>DNMT3B</td>
<td>DNA methyltransferase 3b</td>
</tr>
<tr>
<td>DOG</td>
<td>2-deoxy-galactose</td>
</tr>
<tr>
<td>DOK7</td>
<td>Docking Protein 7</td>
</tr>
<tr>
<td>DR4-TRE</td>
<td>direct repeat 4, thyroid hormone receptor element</td>
</tr>
<tr>
<td>DSB</td>
<td>DNA double-strand break</td>
</tr>
<tr>
<td>DSC1</td>
<td>Desmocollin 1</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>DUX4</td>
<td>double homeobox 4</td>
</tr>
<tr>
<td>Dw</td>
<td>dwarf mice</td>
</tr>
<tr>
<td>E-cad</td>
<td>E-cadherin</td>
</tr>
<tr>
<td>EC</td>
<td>Extra Cellular domain</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EEF1A2</td>
<td>Eukaryotic Translation Elongation Factor 1 Alpha 2</td>
</tr>
<tr>
<td>EGTA</td>
<td>ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid</td>
</tr>
<tr>
<td>EIF2a</td>
<td>Eukaryotic Translation Initiation Factor 2A</td>
</tr>
<tr>
<td>EIF3C</td>
<td>Eukaryotic Translation Initiation Factor 3 Subunit C</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>EMSA</td>
<td>electrophoresis mobility shift assay</td>
</tr>
<tr>
<td>EMT</td>
<td>epithelial to mesenchymal transition</td>
</tr>
<tr>
<td>EP300</td>
<td>E1A Binding Protein P300</td>
</tr>
<tr>
<td>EpiSC</td>
<td>differentiated ES cells, epiblast stem cell</td>
</tr>
<tr>
<td>EtOH</td>
<td>ethyl alcohol</td>
</tr>
<tr>
<td>EXOC4</td>
<td>Exocyst Complex Component 4</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit</td>
</tr>
<tr>
<td>FAM83G</td>
<td>Family With Sequence Similarity 83 Member G</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FBXL-14</td>
<td>F-box and leucine-rich repeat protein 14</td>
</tr>
<tr>
<td>FGF4</td>
<td>Fibroblast Growth Factor 4</td>
</tr>
<tr>
<td>FITC</td>
<td>fluorescein isothiocyanate</td>
</tr>
<tr>
<td>FRG1</td>
<td>FSHD region gene 1</td>
</tr>
<tr>
<td>FSHD</td>
<td>Facioscapulohumeral muscular dystrophy</td>
</tr>
<tr>
<td>FTX</td>
<td>Five prime to Xist</td>
</tr>
<tr>
<td>G9a</td>
<td>EHMT2- Euchromatic Histone Lysine Methyltransferase 2</td>
</tr>
<tr>
<td>GH</td>
<td>growth hormone</td>
</tr>
<tr>
<td>GHKL</td>
<td>gyrase, HSP90, histidine kinase, MutL</td>
</tr>
<tr>
<td>GHRH</td>
<td>GH releasing hormone</td>
</tr>
<tr>
<td>GLI3</td>
<td>GLI Family Zinc Finger 3</td>
</tr>
<tr>
<td>GNAl1</td>
<td>G Protein Subunit Alpha 1</td>
</tr>
<tr>
<td>GPCR</td>
<td>G-protein coupled receptor</td>
</tr>
<tr>
<td>GPR101</td>
<td>G-protein coupled receptor 101</td>
</tr>
<tr>
<td>GPR37</td>
<td>G-protein coupled receptor 37</td>
</tr>
</tbody>
</table>
GREAT  Genomic Regions of Enrichment Analysis Tool
GRID2  Glutamate Ionotropic Receptor Delta Type Subunit 2
GRM8  Glutamate Metabotropic Receptor 8
GWAS  genome wide association studies
H19  Imprinted Maternally Expressed Transcript (Non-Protein Coding)
H2A-Bdb  Histone H2A Barr body deficient
H2AK119ub  Histone H2A mono-ubiquitin lysine 119
H2AZ  Histone H2 variant AZ homolog to Htz1 in yeast
H3K14  Histone H3 Lysine 14
H3K20me3  Histone H3 trimethyl lysine 20
H3K27me3  Histone H3 trimethyl lysine 27
H3K4  histone H3 lysine 4
H3K4me3  tri-methylated at lysine 4 of histone H3
H3K9  Histone H3 lysine 9
H3K9me3  Histone H3 trimethyl lysine 9
H3S10  Histone H3 Serine 10
H4K20me3  Histone H4 trimethyl lysine 20
H4K5  Histone H4 lysine
H4K8  Histone H4 lysine 8
HAT  Histone acetyl transferases
HBiX1  HP1-binding protein
HDAC  Histone deacetylases
HEPES  4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid
HKMT  Histone lysine methyltransferases
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hmC</td>
<td>5-hydroxymethyl-2'-deoxycytidine</td>
</tr>
<tr>
<td>HNRNPK</td>
<td>Heterogeneous nuclear ribonucleoprotein K</td>
</tr>
<tr>
<td>HOMEZ</td>
<td>Homeobox And Leucine Zipper Encoding</td>
</tr>
<tr>
<td>HP1</td>
<td>heterochromatin protein 1</td>
</tr>
<tr>
<td>HR</td>
<td>homologous recombination</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish Peroxidase</td>
</tr>
<tr>
<td>HS</td>
<td>Hyper Sensitive</td>
</tr>
<tr>
<td>hTERT-RPE1</td>
<td>Homo Sapiens pigmented epithelium immortalized with hTERT</td>
</tr>
<tr>
<td>HYLS1</td>
<td>Centriolar And Ciliogenesis Associated</td>
</tr>
<tr>
<td>IBMX</td>
<td>3-isobutyl-1-methylxanthine</td>
</tr>
<tr>
<td>IC</td>
<td>imprinting center</td>
</tr>
<tr>
<td>ICK</td>
<td>Intestinal Cell Kinase</td>
</tr>
<tr>
<td>ICM</td>
<td>inner cell mass</td>
</tr>
<tr>
<td>ICR</td>
<td>imprinting control region</td>
</tr>
<tr>
<td>IGF-1</td>
<td>insulin-like growth factor -1</td>
</tr>
<tr>
<td>IGF2</td>
<td>Insulin like growth factor 2</td>
</tr>
<tr>
<td>IGF2R</td>
<td>Insulin-like growth factor 2 receptor</td>
</tr>
<tr>
<td>I_k_s</td>
<td>voltage gated K^+ channel</td>
</tr>
<tr>
<td>ING2</td>
<td>Inhibitor of Growth Family Member 2</td>
</tr>
<tr>
<td>INO80</td>
<td>Inositol-requiring 80</td>
</tr>
<tr>
<td>ISWI</td>
<td>Imitation switch</td>
</tr>
<tr>
<td>JPX</td>
<td>Just Proximal to Xist</td>
</tr>
<tr>
<td>KAT2B</td>
<td>Lysine Acetyltransferase 2B</td>
</tr>
<tr>
<td>KCl</td>
<td>Potassium chloride</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>KCNE1</td>
<td>Potassium Voltage-Gated Channel Subfamily E Regulatory Subunit 1</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>potassium voltage-gated channel KQT-like subfamily member1</td>
</tr>
<tr>
<td>KCNQ1OT1</td>
<td>Kcnq1 antisense transcript 1</td>
</tr>
<tr>
<td>KD</td>
<td>Knocked-down</td>
</tr>
<tr>
<td>KDM6A</td>
<td>Lysine Demethylase 6A</td>
</tr>
<tr>
<td>KO</td>
<td>knocked-out</td>
</tr>
<tr>
<td>KOH</td>
<td>Potassium hydroxide</td>
</tr>
<tr>
<td>KPRP</td>
<td>Keratinocyte Proline Rich Protein</td>
</tr>
<tr>
<td>LBR</td>
<td>Lamin B receptor</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>LCR</td>
<td>Locus Control Region</td>
</tr>
<tr>
<td>LENG1</td>
<td>Leukocyte Receptor Cluster Member 1</td>
</tr>
<tr>
<td>LINE</td>
<td>long interspersed nuclear element</td>
</tr>
<tr>
<td>LQTS</td>
<td>long QT syndrome</td>
</tr>
<tr>
<td>LRIF1</td>
<td>ligand dependent nuclear receptor interacting factor 1</td>
</tr>
<tr>
<td>LRRN4</td>
<td>Leucine Rich Repeat Neuronal 4</td>
</tr>
<tr>
<td>LSD1</td>
<td>lysine specific demethylase 1</td>
</tr>
<tr>
<td>LTCC</td>
<td>L-type Ca(^{2+}) channels</td>
</tr>
<tr>
<td>macroH2A</td>
<td>Histone H2A variant macro</td>
</tr>
<tr>
<td>MACS2</td>
<td>model-based analysis for ChIP-Seq peak calling with paired ends</td>
</tr>
<tr>
<td>MAGEL2</td>
<td>MAGE (melanoma antigen family)-like 2</td>
</tr>
<tr>
<td>MARCKS</td>
<td>Myristoylated Alanine Rich Protein Kinase C Substrate</td>
</tr>
<tr>
<td>MBD</td>
<td>Methyl-CpG binding domain</td>
</tr>
</tbody>
</table>

xxii
MeCP2  methyl CpG binding protein 2
MgCl2   Magnesium chloride
MKRN3  makorin ring finger protein 3
MPDZ   Multiple PDZ Domain Crumbs Cell Polarity Complex Component
mRNA   Messenger RNA
MTA    Metastasis-associated
NaCl   Sodium Chloride
Nap1l4 nucleosome assembly protein 1-like 4
NBEA   Neurobeachin
ncRNA  Non-coding RNA
NDN    Necdin
Nek2A  NIMA Related Kinase 2
NLRR4  Neuronal Leucine-Rich Repeat protein 4
NLS    nuclear localization signal
NP-40  detergent with CAS 9016-45-9
NR2F6  Nuclear Receptor Subfamily 2 Group F Member 6
NRF-1  nuclear respiratory factor 1
NRSF   neuron restrictive silencer factor
NURD   Nucleosome-remodeling and histone deacetylase
NURF   Nucleosome remodeling factor
ODZ3   TENM3-Teneurin Transmembrane Protein 3
P400   E1A Binding Protein P400
PANTHER Gene Ontology Consortium
PBAF   Polybromo BRG1-associated factor
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>PCDH</td>
<td>protocadherin</td>
</tr>
<tr>
<td>PCDH19</td>
<td>protocadherin 19</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PDE</td>
<td>phosphodiesterase</td>
</tr>
<tr>
<td>PDE4D3</td>
<td>phosphodiesterase 4D3</td>
</tr>
<tr>
<td>PDE8A</td>
<td>Phosphodiesterase 8A</td>
</tr>
<tr>
<td>PGC</td>
<td>primordial germ cell</td>
</tr>
<tr>
<td>PHF21A</td>
<td>PHD Finger Protein 21A</td>
</tr>
<tr>
<td>Phlda2</td>
<td>pleckstrin homology-like domain, family A, member 2</td>
</tr>
<tr>
<td>piRNAs</td>
<td>piwi-interacting small RNAs</td>
</tr>
<tr>
<td>PKA</td>
<td>Protein kinase A</td>
</tr>
<tr>
<td>poly dI-dC</td>
<td>of poly deoxyinosinic-deoxycytidylic</td>
</tr>
<tr>
<td>POU1-F1</td>
<td>Pit-1 protein</td>
</tr>
<tr>
<td>PP1</td>
<td>protein phosphatase 1</td>
</tr>
<tr>
<td>PP2A</td>
<td>Protein Phosphatase type 2A</td>
</tr>
<tr>
<td>pQCMIP</td>
<td>Retroviral vector</td>
</tr>
<tr>
<td>PRC2</td>
<td>Polycomb Repressive Complex 2</td>
</tr>
<tr>
<td>PREB</td>
<td>Prolactin Regulatory Element Binding</td>
</tr>
<tr>
<td>Prl</td>
<td>prolactin</td>
</tr>
<tr>
<td>PTTG1</td>
<td>pituitary tumor transforming 1</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene difluoride</td>
</tr>
<tr>
<td>PWS</td>
<td>Prader Willi Syndrome</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>PWS-SRO</td>
<td>PWS-smallest region of deletion overlap</td>
</tr>
<tr>
<td>QUMA</td>
<td>quantification tool for methylation analysis</td>
</tr>
<tr>
<td>RA</td>
<td>retinoic acid</td>
</tr>
<tr>
<td>RAD21</td>
<td>Double-Strand-Break Repair Protein 21</td>
</tr>
<tr>
<td>RbBP</td>
<td>Retinoblastoma-associated binding protein</td>
</tr>
<tr>
<td>RBBP1</td>
<td>retinoblastoma binding protein 1</td>
</tr>
<tr>
<td>RBBP1-L1</td>
<td>retinoblastoma binding protein 1-like 1</td>
</tr>
<tr>
<td>RepA</td>
<td>Repeat A</td>
</tr>
<tr>
<td>RFP</td>
<td>red fluorescent reporter</td>
</tr>
<tr>
<td>RGD</td>
<td>Arg-Gly-Asp</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>Rnf12</td>
<td>Ring Finger Protein 12</td>
</tr>
<tr>
<td>RNF2</td>
<td>ring finger protein 2</td>
</tr>
<tr>
<td>RORB</td>
<td>RAR Related Orphan Receptor B</td>
</tr>
<tr>
<td>RPE1</td>
<td>retinal pigmental epithelial</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Real Time-quantitative Polymerase Chain Reaction</td>
</tr>
<tr>
<td>RXR</td>
<td>retinoid X receptor</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosyl methionine</td>
</tr>
<tr>
<td>SAP180</td>
<td>Sin3A associated protein 180</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate-polyacrylamide</td>
</tr>
<tr>
<td>SET</td>
<td>Su(var)3-9, Enhancer-of-zester and Trithorax</td>
</tr>
<tr>
<td>SET7</td>
<td>SET Domain Containing Lysine Methyltransferase 7</td>
</tr>
<tr>
<td>SF</td>
<td>Strep II-FLAG</td>
</tr>
</tbody>
</table>
sgRNA  single guided RNA
shRNA  short hairpin RNA
Sin3A  SIN3 Transcription Regulator Family Member A
SLC12A7  Solute Carrier Family 12 Member 7
Slc22a2  Solute Carrier Family 22 Member 2
SLC25A39  Solute Carrier Family 25 Member 39
SLK  STE20 Like Kinase
SMC  structural maintenance of chromosome
SMCHD1  Structural maintenance of chromosome hinge domain containing-1
SNF  Sucrose Nonfermenting Protein 2 Homolog
SNP  single nucleotide polymorphism
SNRPN  Small Nuclear Ribonucleoprotein Polypeptide N
SP1  specificity protein 1
SRCAP  Snf2 Related CREBBP Activator Protein
sRNAs  small RNAs
SRRD  SRR1 Domain Containing
SRS  Silver-Russell syndrome
SSE  sequence specific element
SssI  CpG methyltransferase
STK36  Serine/Threonine Kinase 36
SUV39H  suppressor of variegation 3-9 homolog 1
SWI/SNF  Switching defective/sucrose nonfermenting
SWR1  Swi2/Snf2-related 1
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAP</td>
<td>tandem affinity purification</td>
</tr>
<tr>
<td>TARDBP</td>
<td>TAR DNA Binding Protein</td>
</tr>
<tr>
<td>TDRD9</td>
<td>Tudor domain containing 9</td>
</tr>
<tr>
<td>TE</td>
<td>Trophectoderm</td>
</tr>
<tr>
<td>TET</td>
<td>Tet Methylcytosine Dioxygenase</td>
</tr>
<tr>
<td>TEV</td>
<td>tobacco etch virus</td>
</tr>
<tr>
<td>TGIF1</td>
<td>TGFB Induced Factor Homeobox 1</td>
</tr>
<tr>
<td>Th</td>
<td>tyrosine hydroxylase</td>
</tr>
<tr>
<td>TIP60 (KAT5)</td>
<td>lysine acetyltransferase 5</td>
</tr>
<tr>
<td>TMEM202</td>
<td>Transmembrane Protein 202</td>
</tr>
<tr>
<td>TR</td>
<td>thyroid hormone receptor</td>
</tr>
<tr>
<td>TrxG</td>
<td>trithorax</td>
</tr>
<tr>
<td>TSC</td>
<td>trophoblast stem cell</td>
</tr>
<tr>
<td>TSHβ</td>
<td>thyroid stimulating hormone β</td>
</tr>
<tr>
<td>Tsix</td>
<td>Xist antisense RNA transcript</td>
</tr>
<tr>
<td>TTC28</td>
<td>Tetratricopeptide Repeat Domain 28</td>
</tr>
<tr>
<td>UBE3A</td>
<td>ubiquitin protein ligase E3A</td>
</tr>
<tr>
<td>UBE3A-ATS</td>
<td>antisense transcript of ubiquitin protein ligase E3A</td>
</tr>
<tr>
<td>ULK1</td>
<td>TUnc-51 Like Autophagy Activating Kinase 1</td>
</tr>
<tr>
<td>V(D)J</td>
<td>variable- (diversity)-joining DNA recombination</td>
</tr>
<tr>
<td>WCE</td>
<td>whole cell extract</td>
</tr>
<tr>
<td>WT</td>
<td>wild type</td>
</tr>
<tr>
<td>Xce</td>
<td>X-controlling element</td>
</tr>
<tr>
<td>XCI</td>
<td>X chromosome inactivation</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Xi</td>
<td>inactive X-chromosome</td>
</tr>
<tr>
<td>Xist</td>
<td>X-inactive specific transcript</td>
</tr>
<tr>
<td>Xite</td>
<td>X-inactivation intergenic element</td>
</tr>
<tr>
<td>Xite</td>
<td>X-inactivation intergenic transcription element</td>
</tr>
<tr>
<td>XKR4</td>
<td>XK Related 4</td>
</tr>
<tr>
<td>Xm</td>
<td>Maternal X chromosome</td>
</tr>
<tr>
<td>Xp</td>
<td>paternal X chromosome</td>
</tr>
<tr>
<td>Xpr</td>
<td>X pairing region</td>
</tr>
<tr>
<td>YBX1</td>
<td>Y-Box Binding Protein 1</td>
</tr>
<tr>
<td>YOTIAO</td>
<td>A-Kinase Anchoring Protein 9</td>
</tr>
<tr>
<td>YY1</td>
<td>Yin-Yang 1</td>
</tr>
<tr>
<td>ZCRB1</td>
<td>Zinc Finger CCHC-Type And RNA Binding Motif Containing 1</td>
</tr>
<tr>
<td>ZEB1</td>
<td>zinc finger E-box binding homeobox 1</td>
</tr>
<tr>
<td>ZFP57</td>
<td>Kruppel-associated box-containing zinc-finger proteins (KRAB-AFP)</td>
</tr>
</tbody>
</table>
Chapter 1.

General Introduction

A modified version of this chapter was published in Critical Reviews in Biochemistry and Molecular Biology: Epigenetic events regulating monoallelic gene expression, Massah S, Beischlag TV, Prefontaine GG. 2015;50(4):337-58

Contributions: Shabnam Massah was the main author of this paper.

1.1. DNA packaging and chromatin structure

Deoxyribonucleic acid (DNA) functions to store hereditary information in cells. This information is carried from generation to generation and encodes for all the proteins that make up an organism. In eukaryotes, each diploid cell contains about 2 meters of DNA that is fitted into the nucleus with a diameter about 10 to 20 microns. This is made possible by an important process in living cells, DNA packaging. At the basic level, DNA is wrapped around proteins called histones and form structures known as nucleosomes. The resulting DNA-Histone complex is called chromatin (Figure1-1).
DNA is wrapped around histone proteins and form structures called nucleosome. The overall DNA-histone complex is known is chromatin. DNA packaging to higher order structures enables DNA to fit in nucleus.

Histones are positively charged proteins and tightly bind to DNA due to the negative charge of the phosphate group in the DNA backbone. The most well-known histone proteins include histone H1, H2A, H2B, H3 and H4 [1]. Two of each histone H2A, H2B, H3 and H4 join and form an octamer that binds and wraps approximately 1.7 turns of DNA or about 146 base pairs [2]. The resulting nucleosome structures are joined by the DNA that runs between them, known as linker DNA. This long chain of DNA gives the appearance of a “string of beads” under an electron microscope [3, 4].
The packaging of DNA into a “string of beads” shortens the genome, however this structure is too long to fit into the nucleus. Chromatin is then further packaged by coiling into higher order structures, “30-nanometer fibre” and the chromosomes [5].

Packaging of DNA into higher order structures pose barriers to the enzymes involved in processes such as transcription and replication. These processes require both strands of DNA to unwind so that polymerases can access the DNA template. There are two major mechanisms that change chromatin accessibility. The first method involves displacement of histones by chromatin remodeling complexes [6] and the second mechanism is through addition/erasure of chemical modifications on DNA and histone proteins [7].

1.1.1. Chromatin remodeling proteins

Besides histone octamers that shape chromatin structure, other key factors that remodel chromatin and regulate gene expression are ATPase-dependent enzymes [8]. Four different families of chromatin remodeling complexes have been characterized; Switching defective/sucrose nonfermenting (SWI/SNF), Imitation switch (ISWI), Chromodomain helicase DNA binding (CHD), and inositol-requiring 80 (INO80) complexes.

The SWI/SNF complex is the major chromatin remodeling complex identified in mammals [9]. The core of this complex carries the catalytic activity of the complex and associates with variable multiple accessory subunits. Depending on the subunit compositions, they present different substrate specificity and could be separated into two groups: BRG1-associated factor (BAF) complex if BAF250 (also known as AT-Rich Interaction Domain 1A, ARID1A/B) exists and Polybromo BRG1-associated factor (PBAF) complex when Polybromo (also known as BAF180) is present within the complex [10]. SNF5 (Sucrose Nonfermenting Protein 2 Homolog) also known as BAF47 is the constant core subunit and is present in all variants [11-13]. BRM (also known as SMARCA2, SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 2) and BRG1 (also known as SMARCA4) are the two mutually exclusive subunits of the complex that account for the ATPase catalytic activity of the complex, thus play an essential role in chromatin organization [14].
The ISWI chromatin remodelers are separated into three groups: Nucleosome remodeling factor (NURF), Chromatin-assembly factor (ACF) and Chromatin accessibility complex (CHRAC) [15]. The ISWI ATPases (SNF2H or SNF2L in mammals) are the core subunits of all three groups. SNF2H and SNF2L present distinct functions and are found in different complexes [16]. For instance, SNF2L is present in NURF complex while SNF2H is the subunit of ACF and CHRAC chromatin remodeling complexes [17, 18]. The NURF complex is involved in transcriptional activation whereas ACF and CHRAC complexes regulate chromatin structure including nucleosome assembly and chromosome segregation during replication [19, 20].

The CHD family of proteins are also chromatin remodelers that regulate transcription. Among the nine members of the CHD family, CHD3 and CHD4 are subunits of the nucleosome-remodeling and histone deacetylase (NURD) complex [21]. The NURD complex contains histone deacetylases (HDACs), thus it is involved in transcriptional repression. Similar to the SWI/SNF complex, the NURD complex acquires substrate specificity by combinatorial assembly. CHD3 or CHD4 are the core subunits with catalytic activity. The accessory subunits are members of three gene families: Metastasis-associated (MTA), Methyl-CpG binding domain (MBD) and Retinoblastoma-associated binding protein (RbBP). Combinatorial assembly of subunits forms diversity of complexes with distinct functions [22].

The INO80 chromatin remodeler complex also contain ATPases that in mammals include INO80 and SWRI. They are large multiprotein complexes that contribute to nucleosome-remodeling and transcriptional regulation. In yeast, Swi2/Snf2-related 1 (Swr1) complex incorporates the Htz1 histone variant (known is H2AZ in mammals) into chromatin [23-25]. In mammals, E1A Binding Protein P400 (P400) and Snf2 Related CREBBP Activator Protein (SRCAP) are the homologs to the yeast SWR1 protein. SRCAP complexes are involved in deposition of the H2AZ histone variant [25]. The P400 containing complexes often consist of histone acetyltransferase, TIP60 (also known as lysine acetyltransferase 5, KAT5) and are involved in transcriptional regulation and DNA repair [26, 27].
Beside chromatin remodelers, there are protein complexes that are involved in chromosome dynamics and maintenance of chromosome structure. The next section reviews the role of Structural maintenance of chromosome (SMC) complexes in chromosome organization and focus on Structural maintenance of chromosome hinge domain containing 1 (SMCHD1) in regulating gene expression.

1.1.2. Chromosome organizers

1.1.2.1 Structural maintenance of chromosome (SMC) complexes and their function

SMC proteins in eukaryotes comprise a large family of ATPases that are conserved from bacteria to homo sapiens. In eukaryotes, each SMC protein consists of 1,000-1,500 amino acids and encompasses a central globular hinge domain that is flanked by two coiled-coils. The coiled coil domains end to nucleotide-binding motifs named Walker A and Walker B located at the N-and C-termini, respectively (Figure1-2) [28]. SMC proteins fold at the central hinge domain. This results in anti-parallel interaction of coiled-coil domains and the formation of a two-armed, symmetrical structure. This configuration brings the N- and C-terminal (Walker A and Walker B) together to build a globular structure with ATPase activity [29]. Generally, SMC proteins exist as dimers formed through interactions between hinge domains that allows opening and closing the two arms [30].
As subunits of multiprotein complexes, SMC proteins are involved in higher-order chromosome organization and dynamics. SMC complexes in eukaryotes include cohesion, condensin and SMC5/6 complexes. These complexes are involved in sister chromatid cohesion, chromosome condensation and DNA repair, respectively [31-34]. The SMC V-shaped dimer associates with non-SMC proteins. Usually, the kleisin family of proteins bridge the two head domains of the dimer to form a ring-like structure [35] (Figure 1-2 and Table 1-1).
Table 1-1. The components of the SMC complexes in different species

<table>
<thead>
<tr>
<th>Saccharomyces Cerevisiae</th>
<th>Drosophila melanogaster</th>
<th>Homo Sapiens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohesin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMC1</td>
<td>SMC1</td>
<td>SMC1a</td>
</tr>
<tr>
<td>SMC3</td>
<td>SMC3</td>
<td>SMC3</td>
</tr>
<tr>
<td>SCC1</td>
<td>RAD21</td>
<td>RAD21</td>
</tr>
<tr>
<td>SCC3</td>
<td>Stromalin</td>
<td>SA1 and SA2</td>
</tr>
<tr>
<td><strong>Condensin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMC2</td>
<td>SMC2</td>
<td>CAP-E</td>
</tr>
<tr>
<td>SMC4</td>
<td>SMC4</td>
<td>CAP-C</td>
</tr>
<tr>
<td>YCS4</td>
<td>CAP-D2 and CAP-D3</td>
<td>CAP-D2 and CAP-D3</td>
</tr>
<tr>
<td>YCS5</td>
<td>CAP-G</td>
<td>CAP-G and CAP-G2</td>
</tr>
<tr>
<td><strong>SMC5/6</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMC5</td>
<td>SMC5</td>
<td>SMC5</td>
</tr>
<tr>
<td>SMC6</td>
<td>SMC6</td>
<td>SMC6</td>
</tr>
<tr>
<td>NSE1</td>
<td>CG11329</td>
<td>NSE1</td>
</tr>
<tr>
<td>MMS21</td>
<td>CG13732</td>
<td>NSE2</td>
</tr>
<tr>
<td>NSE3</td>
<td>MAGE</td>
<td>MAGEG1</td>
</tr>
<tr>
<td>NSE4</td>
<td>CG13142</td>
<td>NSE4A and NSE4B</td>
</tr>
<tr>
<td>NSE5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSE6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1.1.2.1.1 Cohesin

The Cohesin complex holds the two sister chromatids together during replication and pulls them against the counter force of the mitotic spindle to regulate correct chromosome alignment and segregation [32, 34, 36, 37]. The Cohesin complex is comprised of SMC1 and SMC3 that are bridged by Kleisin subunits Scc1 and Scc3 in *Saccharomyces cerevisiae* and RAD21, SA1 and SA2 in *homo sapiens*. In *Saccharomyces cerevisiae* Pds5 and WAPL1 are other subunits of the cohesion complex [38-40].

In yeast, during G1 phase of mitosis, cohesion binds to chromosomes and establishes sister-chromatid cohesion in S phase [32, 34, 41]. This structure remains throughout G2
phase until the metaphase-anaphase transition. Dissociation of sister-chromatid cohesion involves cleavage of non-SMC subunits including Scc1 by the cysteine protease, Esp1 which also known as cysteine rich protein 2 (CRIP2) [42]. Initially, Esp1 is inactivated during mitosis up until metaphase through association with pituitary tumor transforming 1 (PTTG1) [42]. Then, through ubiquitin-dependent proteolysis pathway, PTTG1 is released that results in activation of Esp1.

Cohesin complex is also involved in transcriptional regulation where it acts as an intramolecular bridge and mediates long-range chromosomal interaction in cis [43-45]. Cohesin binds to the insulating CCCTC-binding factor (CTCF) protein and inhibits enhancer-promoter interactions to prevent transcription while it might cooperate with the transcriptional co-activator complexes to promote gene expression (Figure 1-3) [44, 46-48].

![Figure 1-3](image)

**Figure 1-3.  A model for transcriptional regulation driven by cohesion**

Cohesin acts as an intramolecular bridge and mediates long-range chromosomal interaction. Cohesin regulates promoter-enhancer-insulator interactions. It might mediate enhancer and promoter interaction to promote transcription (top) whereas it might insulate this interaction to prevent gene expression (bottom).
1.1.2.1.2 Condensin

The Condensin complex that contributes to chromosome condensation is comprised of the SMC2 and SMC4 heterodimer in *Saccharomyces cerevisiae* [49, 50]. SMC 2 and 4 homologs in *homo sapiens* are CAP-E and CAP-C proteins. Non-SMC proteins include CAP-D2/CNAP1, CAP-G and CAP-H in *homo sapiens* and Ycs4 and Ycg1 in *Saccharomyces cerevisiae* [49-51].

*In vitro* analysis suggests that condensin introduces supercoils into relaxed circular DNA [52, 53]. This reaction is ATP dependent and requires topoisomerase I. Electron-spectroscopic imaging reveals that the globular domains of SMC proteins wraps around 190 base pair of DNA, and thereby creates two gyres of DNA. This leads to formation of two compensatory supercoils at a distant site (Figure 1-4) [54].

![DNA supercoiling by condensing](image)

**Figure 1-4. DNA supercoiling by condensing**

Illustration of a model for DNA supercoiling by condensin complex. The head domains of condensin bind to about 190 base pairs of DNA and organize it into a compact structure that form two oriented gyres. This leads to formation of compensatory supercoils at a distant site.

Similar to cohesin, condensin has also been implicated in gene silencing [55]. Condensin acts as an intramolecular linker to form compacted chromosome structures (Figure 1-5). This reduces the chromatin accessibility to transcriptional activators, thereby promotes silencing.
The gene silencing function of condensin
Condensin silences gene expression by forming intramolecular loops that leads to a compacted chromosome structure.

1.1.2.1.3 The SMC5/6 complex
The SMC5/6 complex main function has been linked to DNA repair [31]. However, some evidence associates the SMC5/6 complex function with chromosome structure, genome stability and sister chromatid cohesion. This complex contains at least 4 non-SMC subunits including NSE1, MMS21, NSE2 and NSE4 [56-58].

Inactivation of the SMC5/6 complex in budding yeast, plants, chickens and human results in defects in DNA double-strand break (DSB) repair by homologous recombination (HR) [59-63]. Chromatin immunoprecipitation studies suggest recruitment of SMC5/6 complex to DSB sites in budding yeast [64]. The SMC5/6 also assists in DNA repair by repairing the collapsed replication fork [65]. In budding yeast, SMC6 localizes over the stalled replication fork suggesting its role in rescuing replication [64].

1.1.2.2 Structural maintenance of chromosome hinge domain containing 1 (SMCHD1)
SMCHD1 is a non-canonical member of the SMC family of proteins. The SMCHD1 ATPase domain is located at the N-terminus and unlike other SMC proteins, its ATPase domain is a signature of the GHKL (Gyrase, Hsp90, Histidine Kinase, MutL) ATPase family of proteins [66] (Figure 1-6). The SMCHD1 hinge domain is located at the C-terminus that is homologous to other SMC proteins [67]. The SMCHD1 hinge domain has DNA binding activity and it also contains the interface of SMCHD1
homodimerization. Electron microscopy analysis suggests that the SMCHD1 homodimer forms dumbbell-like rods [67].

**SmcHD1 homodimer**

![Diagram of SmcHD1 homodimer](image)

**Figure 1-6. Structure of SMCHD1 protein**

Illustration of SMCHD1 protein structure. SMCHD1 contains a GHKL-type ATPase domain. SMCHD1 hinge domain is homologous to SMC protein family. SMCHD1 homodimerize through their hinge domains and form a dumbbell-like rod structure.

The role of SMCHD1 has been associated with regulating expression of three distinct gene classes that are known to be monoallelically expressed; X-linked genes, imprinted genes and clustered protocadherin genes (section 1.2) [68-71]. In addition to these three classes, SMCHD1 also regulates a number of bi-allelically expressed genes [70]. More recently, SMCHD1 role has been implicated in the etiology of two developmental diseases that are discussed in the final chapter of this thesis [72, 73].

### 1.1.3. Epigenetic modifications and chromatin structure

Beside chromatin remodelers and chromosome organizers, chemical modifications on DNA and histones alter the electrostatic forces between DNA and histone proteins and ultimately change how tightly DNA is wrapped around histones (Figure 1-7). Epigenetics is the study of heritable changes that modify gene expression without changing individual DNA nucleotides. Eukaryotic genomes are divided into two forms; euchromatin and heterochromatin [74]. Euchromatin regions of the genome are relatively less compact and generally contain active genes while heterochromatin regions are compact structures and comprise inactive genes [74, 75]. Generally, heterochromatin and euchromatin regions encompass specific histone modifications. For instance, transcriptionally active regions tend to associate with histone H3 tri-methyl lysine 4
(H3K4me3) while histone H3 tri-methyl lysine 27 (H3K27me3) marks transcriptionally inactive regions of the genome (Figure 1-7) [76, 77].

**Figure 1-7. DNA and histone modifications**

Chemical modification of DNA and histone change electrostatic forces between DNA and histone, therefore alter chromatin structure and DNA accessibility. (A) Common chemical modifications on histone tails including phosphorylation, acetylation, methylation and ubiquitination. (B) DNA methylation (addition of methyl group to cytosine residue) and histone tail modifications (Ac: acetylation, Me: methylation, Ub: ubiquitination, P: phosphorylation). Pictures are adapted from [78].

### 1.1.3.1 Histone modifications

Histone modifications are principle component of chromatin and play a critical role in the manipulation and expression of DNA. There are a large number of histone post-translational modifications such as acetylation, phosphorylation and methylation.

Histone acetylation takes place on lysine residues and is dynamic. It is regulated by two groups of enzymes with opposing actions. Histone acetyl transferases (HATs) such as E1A Binding Protein P300 (EP300), and Lysine Acetyltransferase 2B (KAT2B) catalyze transfer of an acetyl group to lysine residues while histone deacetylases (HDACs) including HDAC1 and HDAC2 remove lysine acetylation [79-81]. Addition of acetyl groups neutralizes the lysine’s positive charge that weaken DNA and histone interaction [82]. Thus, it is generally accepted that HATs promote transcription while HDACs induce gene repression. Regulation of histone- DNA interaction often occurs through N-terminal
histone tails but there are sites over the globular histone core that are involved in this regulation [83].

Similar to histone acetylation, phosphorylation of histones greatly influences histone-DNA interaction. Addition and removal of phosphate groups is controlled by kinases such as Cyclin-dependent kinases 1 (CDK1) and phosphatases including Protein Phosphatase type 2A (PP2A), which add and remove these groups, respectively [84]. The majority of histone phosphorylation takes place on serine, threonine and tyrosine residues over the N-terminal histone tail; however, sites in the histone core also exist [85].

Histone methylation takes place on lysine and arginine residues. Interestingly, lysines may be mono-, di-, or tri-methylated whereas arginines may be mono- and di-methylated [86, 87]. Unlike acetylation and phosphorylation, methylation of histone tails does not change the charge. There are several methyltransferases. All histone lysine methyltransferases (HKMTs) that methylate N-terminal lysine chains have SET (Su(var)3-9, Enhancer-of-zester and Trithorax) domain that harbours the enzymatic activity [88]. HKMTs utilize S-adenosylmethionine (SAM) as a cofactor to add a methyl group to a lysine amino terminal [89]. There are several histone demethylases. The first lysine demethylase identified was Lysine-specific demethylase 1 (LSD1) that only demethylates mono- and di-methylated substrates [90]. Besides LSD1, there is another class of demethylases known as jumonji proteins that contain a so-called JmjC jumonji domain with enzymatic activity [91]. Overall, histone demethylases have a high level of substrate specificity and are sensitive to degree of methylation.

1.1.3.2 Histone modification and transcriptional regulation

Histone modifications alter gene expression through two main mechanisms. First, by directly changing the structure of chromatin and second, by regulating the binding of chromatin regulators.

Histone acetylation and phosphorylation decrease the positive charge of histones. This influences the electrostatic forces between DNA and histones and potentially alters chromatin structure. Numerous sites on histone tails are prone to acetylation such as
histone H3 Lysine 9 (H3K9), histone H3 Lysine 14 (H3K14), histone H4 lysine 5 (H4K5) and histone H4 lysine 8 (H4K8) [92-94]. Therefore, hyper-acetylated regions could have less compact structure. Enhancer elements and particularly gene promoters are often hyper-acetylated which facilitate transcription factor access and gene expression [95].

Histone phosphorylation is less common than acetylation but can still change chromatin structure. For instance, genome-wide phosphorylation of histone H3 Serine 10 (H3S10) is associated with chromatin condensation during mitosis [96].

Histone modifications influence binding of chromatin-associated factors to chromatin and there are numerous effector complexes that specifically interact with modified histones. These chromatin-associated complexes and associated factors often contain domains that allow for recognition and interaction with histone modifications. Some of the domains that recognize histone modification include PHD fingers, chromo-domains, Tudor, MBT domains and bromo-domain [97-99]. For instance, heterochromatin protein 1 (HP1) interacts with histone H3 tri-methyl lysine 9 (H3K9me3) through its chromo-domain and this interaction is associated with overall structure of compact and transcriptionally repressed regions of chromatin [100, 101]. Chromatin remodeling complexes harbouring bromo-domains frequently bind to acetylated histones. For example, Switching defective/sucrose nonfermenting (Swi2/Snf2) interacts with acetylated histones through its bromo-domain [102, 103]. Specific histone acetylations lead to recruitment of the SWI/SNF remodelling complex that in turn facilitates opening the chromatin structure.

Binding of chromatin factors that specifically interact with histone modifications recruits additional chromatin modifiers. For example, binding of Inhibitor of Growth Family Member 2 (ING2) to histone H3 tri-methylated lysine 4 (H3K4me3) recruits the mSin3a-HDAC1 complex to active proliferation-specific genes in response to DNA damage [104].

1.1.3.3 DNA methylation

DNA methylation is defined by the covalent addition of a methyl group to the cytosine residue of a CpG dinucleotide [105, 106]. DNA methylation together with other epigenetic modifications modulates chromatin structure, gene regulation and
chromosome stability. Therefore, DNA methylation is essential for regulation of gene expression during embryogenesis and development.

DNA methylation is regulated by DNA methyltransferases (DNMTs); DNMT1, DNMT2, DNMT3A and DNMT3B [107-111]. DNMT1 is known as maintenance methyltransferase and is responsible for replicating the pattern of DNA methylation to the newly synthesized daughter strands during DNA replication [112]. DNMT2 is responsible for methylation of cytosine-38 in the anticodon loop of aspartic acid transfer RNA [113]. DNMT3A and DNMT3B preferentially methylate unmethylated CpG dinucleotides and perform de novo methylation during development [114]. DNMTL has homology to DNMT3A and DNMT3B but lacks enzymatic activity. In addition, DNMTL assists de novo methylation by stimulating DNMT3A and DNMT3B and increasing their interaction with the methyl donor, SAM [115].

Chromatin structure is modulated by the interplay between DNA methylation and histone modifications. Histone modifications can direct the cellular machinery to promote DNA methylation and vice versa. Lehnertz and colleagues suggested that establishment of H3K9me3 by suppressor of variegation 3-9 homolog 1 (SUV39H), a histone methyltransferase and association of HP1 protein recruits DNA methyltransferase 3b (DNMT3B) and promotes DNA methylation at pericentric repeats in mouse ES cells [116]. Conversely, through the use of transgenic constructs containing methylated or unmethylated human beta-globin sequences in mice, Hashimshony and colleagues demonstrated that methylated transgenes can modify chromosome organization by inducing hypo-acetylation and hypo-methylation of H3K4 and hyper-methylation of H3K9 [117]. In addition, some evidence suggests that DNA methylation could direct di- and tri-methylation of histone H3K9 [118]. In Hela cells, DNMT1 mutation results in global loss of histone H3 di- and tri-methylated lysine 9 (H3K9me2 and H3K9me3) levels while re-expression of DNMT1 recovers this phenotype.

There are a number of examples that link DNA methylation and histone modifications and put DNA methyltransferases and histone modifiers in a single complex. For instance, histone H3K9 methyltransferase and SUV39H1 associate with DNMT1 and DNMT3A through the conserved PHD domain of DNMT3A [119]. Also, histone
methyltransferase SET7 (SET Domain Containing Lysine Methyltransferase 7) that methylates H3K4. SET7 regulates DNMT1 stability and degradation by co-localizing with DNMT1 and methylating at the K142 position [120]. Additional example is G9A histone methyltransferase that regulates mono- and di-methylation of histone H3K9. Genome-wide analysis suggests that loss of G9A in mice leads to reduction of DNA methylation [121-123]. Therefore, these results suggest crosstalk exists between DNA methylation and histone modifications which, greatly influences chromatin structure and gene expression.

1.1.3.4 DNA demethylation
DNA demethylation plays a critical role during development and tumorigenesis. During development, loss of DNA methylation is required for returning cells to the pluripotent state. Primordial germ cells (PGCs) lose DNA methylation over introns, intergenic regions and repeats and to a lesser degree over the promoters and exons of genes [124].

In plants, 5-methylcytosine (5mC) DNA glycosylases regulate active DNA demethylation [125, 126]. This form of DNA demethylation has not been identified in mammals yet however, other potential mechanisms have been proposed. One mechanism is through the influence of cytosine deaminases which converts 5mC to thymine. The T-G mismatch is then repaired by replacing thymine with cytosine. Current evidence suggests that activation-induced cytosine deaminase (AID) is associated with DNA demethylation [124, 127]. Loss of AID results in global DNA demethylation in murine PGCs. Furthermore, during reprogramming of somatic cells, AID is required for DNA demethylation of pluripotent-associated genes.

Another mechanism of DNA demethylation is through TET (Tet Methylcytosine Dioxygenase) family hydroxylases. This family of proteins converts 5mC to 5-hydroxymethyl-2'-deoxycytidine (hmC) via oxidation of the methyl group [124, 127]. HmC might be an intermediate product in a pathway that actively demethylates DNA. Also, hmC might associate with passive DNA demethylation as it is recognized poorly by maintenance DNA methyltransferase, DNMT1.
1.2. Genes regulated by SMCHD1

As mentioned above three distinct gene classes are regulated by SMCHD1; X-linked genes, imprinted genes and clustered protocadherin genes [68-71]. These genes are known to be monoallelically expressed. The following three sections summarize mechanisms involved in regulating expression of each gene class and the role SMCHD1 plays in this regard.

1.2.1. X chromosome inactivation

1.2.1.1 Regulation of X chromosome inactivation

X chromosome inactivation in mammals is a mechanism that regulates dosage compensation for most X-linked genes in females by shutting down expression of genes on the second X chromosome. Initially, Lyon introduced XCI as an explanation for variegated coat color observed in heterozygous female mice [128]. XCI is established in a stochastic or an imprinted manner.

In 1983, Rastan conducted one of the first studies that identified the X-inactivation center accountable for the establishment of XCI [129]. This region encompasses genes that mainly transcribe ncRNAs (Figure 1-8). The best-characterized genes in this region are X-inactive specific transcript (Xist) and Xist antisense RNA transcript (Tsix).

The Xist ncRNA is specifically expressed from the inactive X chromosome. This ncRNA coats the X chromosome and establishes an inactive state over the X chromosome while Tsix is a well-known regulator of the Xist gene and represses Xist gene expression.

Other genes located in the X-inactivation center work together to regulate expression of Xist and Tsix genes. The role of factors involved in XCI are listed in Table 1-2. Overall, the process of XCI can be categorized into four sections; counting, choice, inactivation of one X chromosome and maintenance of the active state of the second X chromosome.
Figure 1-8. The X inactivation center is enriched in genes that encode for ncRNAs and are involved in X chromosome inactivation.

A schematic illustration of the Xic region, which encompasses a number of positive regulators of Xist RNA such as Rnf12, Jpx and RepA genes and negative regulators including Xite and Tsix genes. The regions that are involved in the pairing and counting are indicated with blue line. The Xce region, which is responsible for the choice of inactivation, is in indicated in red. The dashed red line is the original region named as Xce and the solid red line indicates the minimal segment of Xce shown to be sufficient for the Xce effects.

Table 1-2. Summary of repressors and activators of XCI.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Mode of XCI regulation</th>
<th>Functions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xpr</td>
<td>Activator</td>
<td>Induces homologous pairing of Xite and Tsix, promotes initiation of XCI</td>
<td>Augui (2007)</td>
</tr>
<tr>
<td>Rnf12</td>
<td>Activator</td>
<td>Has dosage dependent activity, Induces initiation of XCI</td>
<td>Barakat et al. (2011), Shin et al. (2010), Jonkers et al. (2009), Gontan et al. (2012)</td>
</tr>
<tr>
<td>Jpx</td>
<td>Activator</td>
<td>Induces hyperacetylation of Xist promoter and Xist expression</td>
<td>Johnston et al. (2002), Tian et al. (2010), O’Neill et al. (1999)</td>
</tr>
<tr>
<td>RepA</td>
<td>Activator</td>
<td>Associates with polycomb protein complex, interacts with Xist encoding locus and induces its expression</td>
<td>Hoki et al. (2009), Zhao et al. (2008)</td>
</tr>
<tr>
<td>Ftx</td>
<td>Activator</td>
<td>Regulates expression of Jpx and Tsix</td>
<td>Chureau et al. (2011)</td>
</tr>
<tr>
<td>Xist</td>
<td>Activator</td>
<td>Interacts with polycomb protein complex, H3/K27me3 and coats the X chromosome targeting it for inactivation</td>
<td>Wutz et al. (2002), Beletskii et al. (2001), Brown et al. (1991), Borsani et al. (1991), Penny et al. (1996), Jeon &amp; Lee (2011)</td>
</tr>
<tr>
<td>Dspax34r</td>
<td>Activator</td>
<td>Reverse complement of Tsix and inhibits its expression</td>
<td>Courtier et al. (1995), Cohen et al. (2007), Vigneur et al. (2006)</td>
</tr>
<tr>
<td>Tsix</td>
<td>Repressor</td>
<td>Antisense to Xist. Inhibits binding of RepA-PRC2 complex to the Xist locus</td>
<td>Zhao et al. (2008), Leet al. (1999a,b), Sado et al. (2005), Morey et al. (2001), Ohbata et al. (2008), Ogawa et al. (2008)</td>
</tr>
<tr>
<td>Xite</td>
<td>Repressor</td>
<td>Drives expression of Tsix</td>
<td>Ogawa &amp; Lee (2003)</td>
</tr>
</tbody>
</table>

Note. Italic text represents the activator and the bold text the repressor activity of genes towards development of XCI

Counting is a process by which a diploid cell determines if more than one X chromosome is present in its genome. Although typical males contain a single copy (XY), and females contain two copies (XX), meiotic errors can produce individuals with
other combinations (e.g., XXX, XXY). When only one X should remain active, cells must know how many X chromosomes are present so that all additional X chromosomes are inactivated. Previously, the presence of only one active X chromosome in individuals harbouring XXX or XXY contributed to the identification of XCI in females [130-132]. Augui and colleagues, have suggested that in mammalian female cells, *trans*-interaction of the X pairing region (Xpr) from two X chromosomes induce pairing of the *Tsix* and X-inactivation intergenic element (*Xite*) genes and may act in chromosome counting and initiation of XCI (Figure 1-9, Model 1) [133]. In addition, recent data suggest that CCCTC-binding factor (CTCF) protein binds to Tsix ncRNA co-transcriptionally and mediates X-X pairing (Figure 1-9, Model 2) [134]. These models are not necessarily exclusive and it is not clear if homologous pairing of X chromosomes by Xpr occurs at the same time and through the same mechanism as chromosome pairing initiated by Xite and Tsix ncRNAs.
Determination of the number of X chromosomes

**Model 1**

- Pairing/counting determines the number of X chromosomes prior to inactivation.

**Model 2**

- A schematic illustration of two models proposed that explain the counting mechanism. Model 1 describes pairing of the Xpr region followed by Tsix and Xite pairing and initiation of inactivation. Model 2 explains pairing mediated by CTCF protein and depends on the transcriptional activation of Tsix. Tsix ncRNA guides CTCF to the inactivation center and interaction of CTCF proteins induces X–X pairing.

**Figure 1-9.** Pairing/counting determines the number of X chromosomes prior to inactivation.

A schematic illustration of two models proposed that explain the counting mechanism. Model 1 describes pairing of the Xpr region followed by Tsix and Xite pairing and initiation of inactivation. Model 2 explains pairing mediated by CTCF protein and depends on the transcriptional activation of Tsix. Tsix ncRNA guides CTCF to the inactivation center and interaction of CTCF proteins induces X–X pairing.
During XCI in embryonic tissues, cells choose and inactivate either the Xm or Xp. This is referred to as the choice phenomenon. Initially, it was thought that this process is completely random during stochastic XCI, but studies with heterozygous female mice determined that the genotype of a region called the X-controlling element (Xce) influences the choice between inactivation of paternal or maternal X chromosomes and skews the XCI ratio. So far, six strain specific Xce alleles (Xce^a to Xce^e) have been identified each with different susceptibility to inactivation (Figure 1-9) [135, 136].

Previously, Xite ncRNA that is located within the Xce region was thought to be the locus responsible for the allelic skewed ratio of XCI [137]. However, a recent study by Calaway and colleagues further narrowed the region within the Xce responsible for skewing inactivation of X chromosomes to a region that excludes Xite [135].

1.2.1.2 The X inactive specific transcript (Xist) ncRNA

Xist encodes for a 17 kilobase ncRNA that is only expressed from the inactive X chromosome [138-140]. Upon transcription, the spliced Xist RNA remains in the nucleus and spreads over the inactive X chromosome in cis. Xist RNAs that fail to bind to the X chromosome are efficiently degraded inhibiting trans-inactivation of the active X chromosome [141]. The sequences required for association of Xist RNA with the inactive X and coating of the X chromosome are scattered throughout the Xist coding sequence [141, 142].

Xist RNA stability has been linked to regulation of XCI. There are three Xist promoters (P0, P1, P2) that drive Xist expression and each isoform has different levels of RNA stability [143]. Prior to XCI, the Xist P0 promoter drives expression of the unstable Xist RNA from both chromosomes in ES cells. Upon initiation of XCI, transcription of Xist RNA initiates from the P1 and P2 promoters. These transcripts are more stable and occupy the inactive X chromosome. Therefore, alternative promoter usage alters Xist RNA stability [144, 145].

The Xist RNA works in collaboration with PRC2 to establish an inactive state over the X chromosome. PRC2 has a methyltransferase activity that places methylation marks on histone H3 lysine 27 (H3K27me3) [146]. The interaction between Xist RNA and PRC2
may bring about long term silencing of the inactive X chromosome. Along with the H3K27me3 histone mark, surrounding long interspersed nuclear elements (LINEs) facilitate the formation of nuclear compartmentalization [147]. LINEs are retrotransposons of about 6 kilobases in length that are more extensively enriched on X chromosome compared to autosomes [148]. X-linked genes that escape XCI have a lower number of flanking LINEs compared to those subjected to inactivation suggesting for a role of LINEs in inactivation of selected X-linked genes [149]. Recent data suggest interaction of Xist ncRNA with 81 proteins that are required at different phases during X inactivation [150]. Among the list of Xist interacting proteins, heterogeneous nuclear ribonucleoprotein K (HNRNPK) is required for gene silencing and incorporation of repressive chromatin modifications (H3K27me3, H2AK119ub) over the inactive X chromosome in differentiated ES cells, epiblast stem cells (EpiSC, undergo random XCI), trophoblast stem cells (TSC, undergo imprinted XCI) and not undifferentiated ESCs. Also, components of Sin3-HDAC1 complex including Drosophila Split ends homolog, SPEN bind to the Repeat A locus of Xist RNA and assist in gene silencing. This interaction does not occur in undifferentiated ESCs. This data suggest that Xist interaction with its binding partners is dynamic during XCI.

1.2.1.3 The Tsix ncRNA

Prior to XCI, Tsix expression is biallelic but becomes asymmetrically expressed only from the future active X chromosome [151]. Tsix ncRNA negatively regulates expression of Xist. At the early stages of XCI, Tsix RNA, which has overlapping antisense transcriptional direction to Xist, presumably antagonizes Xist expression.

Transcription of Tsix RNA induces Xist repression by modulating chromatin structure over the Xist locus in cis [152]. Deletion of a 65 kilobase region that encompasses the 3’ end of the Xist gene and the Tsix promoter in ES cells causes overexpression of Xist from that chromosome and diffusion of the transcript to sites other than the transcriptional site as observed by weakly detectable scattered pattern in some nuclei [153]. Upon restoration of this deleted region, Xist adopts normal transcript levels and Xist RNAs remain in close contact with the transcriptional start site. Furthermore, Tsix RNA interferes with Xist activation by blocking the binding of REPA with the PRC2
complex [154]. Together these results suggest that Tsix negatively regulates Xist gene expression.

The overlapping of the Tsix encoding region/transcriptional unit across the Xist promoter is essential for establishment of repressive chromatin structure over the Xist locus [155]. In line with this result, Ogawa and colleagues demonstrated that the double stranded RNAs formed by hybridization of the complementary strands of the Tsix and Xist overlap region is processed by dicer protein [156]. As a result, the small RNAs (sRNAs) likely act to repress the Xist gene over the active X chromosome. The XCI process is interrupted in ES cells when dicer protein levels are lowered by 5 percent. Also, in these cells, accumulation of Xist RNA over the X chromosome and association of repressive histone H3K27me3 marks is perturbed [156]. These results emphasize a regulatory role of antisense Tsix transcript for Xist gene repression and the importance of dicer in this process. Besides ncRNAs DNA methylation and histone modifications play critical roles in organizing chromatin structure over the active and inactive X chromosomes [116, 117, 157]. Also, various proteins have been linked to initiation and maintenance of XCI [68, 158, 159]. Here, we; 1) discuss the roles of DNA methylation, histone modifications and number of genes for organizing chromatin structure over the inactive X chromosome and; 2) summarize the roles of regulatory proteins in initiation and establishment of XCI.

1.2.1.4 Chromatin modifications associated with XCI

Combinations of different histone tail modifications and DNA methylation states regulate chromatin structure. These epigenetic modifications either provide accessible binding sites for transcription factors creating transcriptionally active euchromatin domains or block the access of transcription factors to chromatin by forming inactive heterochromatin regions.

During early stages of XCI and following the spread of Xist RNA over the X chromosome, the chromatin accumulates repressive histone marks including hyper-methylated histone H3 lysine 9 (H3K9), hypo-methylated histone H3 lysine 4 (H3K4) and hypo-acetylated histone H3 and H4 [160-164]. Incorporation of H3K27me3 on histones, mediated by PRC2, is essential at early stages of XCI [165, 166]. Moreover, isoforms of macroH2A histone including macroH2A1.1, macroH2A1.2 and macroH2A2 are involved
in establishment of XCI. Other H2A variants such as H2A Barr body deficient (H2A-Bdb) and H2AZ are absent from the inactive X chromosome [167, 168]. Together, these non-permissive histone modifications create a highly packed chromatin structure over the inactive X chromosome.

A study conducted in a retinal pigmented epithelial (RPE1) derived cell line, suggests the existence of two distinct heterochromatin domains on the inactive X chromosome. One involves association of heterochromatin protein 1 (HP1) with histone H3 trimethyl lysine 9 and lysine 20 (H3K9me3, H3K20me3) [169]. The other encompasses Xist RNA, MacroH2A and H3K27me3. Together folding of these two heterochromatin domains assist in the establishment of an inactive X chromosome [169].

1.2.1.5 The role of DNA methylation in XCI

DNA methylation in mammals is defined by the covalent addition of a methyl group to cytosine residue of a CpG dinucleotide. DNA methylation together with other epigenetic modifications modulates chromatin structure, gene regulation and chromosome stability. DNA methylation is essential for regulation of gene expression during embryogenesis and development. Studies in mice deficient in DNA methyltransferase 1, a protein that maintains DNA methylation, showed that maintenance of DNA methylation preserves the inactive state of X chromosome in embryonic (random inactivation) but not in extra-embryonic (imprinted inactivation) tissues, suggesting that other factors play key roles in imprinted XCI [157].

Chromatin structure is modulated by the interplay between DNA methylation and histone modifications. Histone modifications can direct the cellular machinery to promote DNA methylation and vice versa. Lehnertz and colleagues suggested that establishment of H3K9me3 by suppressor of variegation 3-9 homolog 1 (SUV39H), a histone methyltransferase and association of HP1 protein recruit DNA methyltransferase 3b (DNMT3B) and promotes DNA methylation at pericentric repeats in mouse ES cells [116]. Conversely, through the use of transgenic constructs containing methylated or unmethylated human beta globin sequences in mice, Hashimshony and colleagues demonstrated that methylated transgenes can modify chromosome organization by inducing hypo-acetylation and hypo-methylation of H3K4 and hyper-methylation of H3K9.
Therefore, these results suggest crosstalk between DNA methylation and histone modifications to establish chromatin structure and regulate gene expression.

### 1.2.1.6 The role of regulatory proteins in XCI

A number of studies have attempted to characterize the role of different regulatory proteins in the XCI process. For example, at the chromatin transitional zone between \textit{Xist} and \textit{Tsix} genes, a conserved element exists that has a strong binding activity for the \textit{Xist} repressor, CTCF protein. Deletion of this region perturbs XCI [159].

Another protein with a known regulatory role in XCI is the transcriptional repressor, Yin-Yang 1 (YY1). This protein has dual functions in that it binds to both \textit{Xist} RNA and chromatin, forming a physical link between mobile \textit{Xist} RNA and chromatin [146]. Upon activation of \textit{Xist} gene expression by the RepA-PRC2 complex, \textit{Xist} ncRNA binds to PRC2 co-transcriptionally. YY1 protein links the \textit{Xist}-PRC2 complex to the \textit{X} chromosome and promotes spreading of this complex throughout the \textit{X} chromosome [146]. Moreover, the DXPas34 locus located at 5' region of \textit{Tsix} encompasses binding sites for CTCF and its binding partner, YY1 [158]. Failure of CTCF and YY1 protein-protein association perturbs \textit{Tsix} gene regulation and disrupts both random and imprinted XCI [158]. CTCF binding to \textit{DXPas34} is sensitive to CpG methylation blocking CTCF binding to DNA [170]. Therefore, the chromatin modifications might also direct the cellular machinery and the choice of \textit{X} chromosome for inactivation.

Beside DXPas34, CTCF has binding sites within \textit{Xite}. CTCF binding to \textit{DXPas34} and \textit{Xite} and possibly their transcriptional activation is required for pairing of two \textit{X} chromosome [171]. Opposite to data shown by Donohoe and colleagues, YY1 protein was not essential for \textit{trans}-chromosomal pairing [134]. UV-crosslinking and immunoprecipitation using anti CTCF antibody followed by deep sequencing (CLIP-seq), revealed CTCF binding sites within RNA transcripts including \textit{Tsix} in mouse embryonic stem cells while CTCF mostly bound to noncoding intergenic region in genomic DNA. In line with previous observations CTCF interaction with \textit{Tsix} induces long-range chromosomal interactions and pairing.
1.2.1.7 The role of SMCHD1 in XCI

SMCHD1 protein has been implicated in X chromosome inactivation (XCI). Female mice carrying homozygous SMCHD1 mutations died in utero while half of the male mice with homogenous background were fertile [68]. Initially, hypo-methylation of CpG islands (CGIs) was thought to be responsible for the modified expression of X-linked genes [68]. However, later experiments showed that the modified DNA methylation pattern was not the sole factor in up-regulation of X-linked genes [69]. Mice deficient in DNMT3B, which lacked de novo methyltransferase activity failed to methylate CGIs and that loss was insufficient to de-repress the X-linked genes. This suggests that hypo-methylation of CGIs in X-linked genes is not the only contributor to modified expression in SMCHD1 mutant mice [69]. DNA methylation during XCI occurs in two phases; early and late. Gendrel and co-workers have suggested that methylation of CGIs over the X-linked genes is mediated through both SMCHD1 dependent or independent pathways [172]. The SMCHD1 dependent pathway was essential for late methylation of CGIs over the X chromosome, whereas the SMCHD1 independent pathway accounted for the early methylation at CGIs during initial stages of XCI [172].

Nozawa and colleagues have demonstrated that SMCHD1 is closely associated with X-inactive specific transcript (Xist) RNA and H3K27me3 histone marks [173]. SMCHD1 is recruited to the H3K27me3 domain presumably through interaction with Xist RNA. Whether SMCHD1 binds directly to Xist RNA requires further investigation. On the inactive X chromosome, HP1-binding protein (HBiX1) that is also called ligand dependent nuclear receptor interacting factor 1 (LRIF1) associates with H3K9me3 though interaction with HP1 protein. Interaction between SMCHD1 and HBiX1 brings Xist mediated H3K27me3 and HP1 mediated H3K9me3 heterochromatin regions together to form a compact chromatin structure [173]. SMCHD1 or HBiX1 knockdown in hTERT-RPE1 cells disrupts the compact structure of the inactive X chromosome and perturbs asynchronous replication of the inactive X chromosome [173]. This further supports the essential role of SMCHD1 and HBiX1 in establishment of XCI and links chromosome assembly and compaction with time of replication. We have summarized a hypothetical model, which explains existence of two distinct chromatin domains.
(H3K9me3, H3K27me3) while illustrating interactions between them resulting in precise folding (Figure 1-10) [169, 173].
Figure 1-10. Two distinct heterochromatin regions form a compact structure during XCI. Left panel is an illustration of a hypothetical model showing formation of two distinct domains. The hypothetical model explains two separate domains made of H3K9me3 and H3K27me3 heterochromatin regions on the inactive X chromosome while taking into consideration the folding of these domains mediated by SMCHD1 and HBiX1. Right panel is a schematic illustrating formation of a packed chromatin structure through linking two heterochromatin domains. SMCHD1 protein associates with Xist RNA and H3K27me3 histone mark while HBiX1 interacts with HP1 and H3K9me3 histone mark. Interaction of SMCHD1 and HBiX1 proteins brings two distinct heterochromatin domains in close vicinity likely for forming a highly compacted chromatin structure. The activities of SMCHD1 and HBiX1 are not likely to be dependent on the function of polycomb repressive complex 2 (PRC2) or its association with H3K27me3 because disruption of PRC2 function and loss of H3K27me3 did not change the highly ordered structure of the inactive X chromosome [173]. This also emphasizes that in the absence of H3K27me3 heterochromatin, SMCHD1 still interacts with Xist RNA and promotes inactivation of the X chromosome. Thus, it appears that SMCHD1 is involved in the establishment of highly ordered chromatin structure over the inactive X chromosome rather than directly regulating the expression of the inactive X-linked genes.
1.2.2. Genomic imprinting and regulation of imprinted genes

Initially, it was thought that the paternal and maternal genomes contribute equally to the zygotic cell. However, experiments with mammalian parthenotes, in which only a single parent contributes to the diploid genome, showed that the paternal and maternal genome contributions are not functionally equivalent and that both genomes are essential for viability of the zygote. In these studies, a single pronucleus was removed from a one-cell stage embryo followed by introducing the second pronucleus to construct a diploid mouse embryo with two female or two male pronuclei. This method demonstrated that both maternal and paternal genomes have distinct roles for completion of the embryogenesis [174]. Although, uni-parental disomy of chromosomes 1, 4, 5, 9, 13, 14 and 15 does not affect viability, having two copies of single parentally derived autosomes, such as chromosome 6, is lethal [175]. More importantly paternal and maternal disomy (such as chromosome 11) have different effects in littermates suggesting parental chromosome origin is very important and that these chromosomes are differentially regulated depending on their parent-of-origin [175]. For instance, in distinct human syndromes such as Prader Willi Syndrome (PWS) and Angelman Syndrome (AS), the identical 15q11-q13 chromosomal region is deleted from either the paternal or maternal chromosomes, respectively but with separate developmental outcomes. These imprinting disorders have distinct symptoms; PWS is associated with hypotonia and intellectual disability while individuals with AS suffer from speech and developmental impairment [176-178]. Later, in 1991, investigators identified Insulin-like growth factor 2 receptor (Igf2r) as another imprinted gene and in the same year, mouse H19 and Insulin like growth factor 2 (Igf2) genes were added to the list of imprinted genes with parent-of-origin specific expression [179-181]. Today we know that genomic imprinting plays significant roles in the establishment of proper growth [174]. Genome-wide analysis identified over 1300 protein coding genes and ncRNAs with parent-of-origin allelic exclusion in the mouse brain from which 604 genes have human orthologs. Maternal genome contributes more significantly in the mouse embryonic brain development whereas influence of paternal genome was more substantial in adult cortex and hypothalamus [182]. Using allele specific DNA methylation pattern by MethylC-seq in mouse frontal cortex, Xie and colleagues characterized 1952 parent-of-origin
dependent CG dinucleotides [183]. Methylation of numbers of these CG sites is
sequence specific and the polymorphism of the adjacent sequence influences the level
of methylation and subsequently expression.

There are differences in imprinted gene expression between human and mouse
placenta. Although many imprinted genes are common between mice and humans,
some genes have species-specific imprinting including Nap1l4, Cd81 and Slc22a2 that
are biallelically expressed in human placenta while imprinted in mouse placenta [184].
The majority of the known imprinted genes have been studied in the placenta. During the
prenatal stages of development, the placenta has the highest degree of imprinted gene
expression. This organ that facilitates nutrient exchange between the mother and the
fetus has an important role in balancing the parental resource allocation. Imprinted
genes show spatiotemporal dependent expression during development. This means that
depending on the tissue and stage of the development they may be imprinted or have
biallelic expression. Some of the genes that are imprinted in the placenta and important
for prenatal growth are biallelically expressed in other tissues during the postnatal
stages including IGF2 gene in humans that have biallelic expression in adult liver tissues
while it is imprinted in placenta [185].

The brain is another organ in which imprinted gene expression influences its function
and consequently neurodevelopmental processes. Some genes are specifically
imprinted in brain such as the ubiquitin protein ligase E3A (UBE3A) gene in humans
[186]. The spatiotemporal dependent expression of imprinted genes is also evident in
brain. Some genes that show biallelic expression in fetal brain, gain imprinted gene
expression in the adult brain. This includes the Copper Metabolism (Murr1) Domain
Containing 1 (Commd1) gene that is biallelically expressed in the embryonic mouse
brain and switches to monoallelic expression in the adult mouse [187]. Therefore,
depending on the tissue and stage of development, imprinted genes might acquire
different expression patterns.
1.2.3. **KCNQ1 imprinted locus, its epigenetic regulation and associated diseases**

Some imprint control regions (ICRs) overlap with the promoters of genes and transcribe ncRNAs, thus regulating the ncRNA expression. The ncRNAs govern the expression of imprinted genes *in cis*. An example of an ICR overlapping with a ncRNA is the KvDMR, which is located about 500 Kb from the \( H19/\text{Igf2} \) locus. This ICR, that is also called LIT1 (ICR2), contains the promoter of the Kcnq1 antisense transcript 1 (Kcnq1ot1) and regulates its expression [188-190]. In return, differential expression of this ncRNA regulates expression of genes *in cis* (Figure 1-12) [191].

---

**Figure 1-11.** **ICR1 and KvDMR1; two imprinted control regions with distinct modes of regulation.** A schematic illustration of mouse Kcnq1 and \( H19/\text{Igf2} \) imprinted domains.

The KvDMR1 region is methylated (closed circle) on the maternal (M) and unmethylated (open circle) on the paternal (P) chromosome. Kcnq1ot1 ncRNA is expressed only from the paternal allele, where its promoter, located within KvDMR1 is unmethylated. Expression of Kcnq1ot1 induces repression of other imprinted genes in the Kcnq1 domain. \( H19/\text{Igf2} \) is located upstream of the Kcnq1 domain, are regulated through differential methylation of the ICR that control expression of \( H19/\text{Igf2} \) imprinted genes. The non-imprinted, ubiquitously imprinted and placental specific imprinted genes are in light green, orange and blue, respectively. The genes expressed and repressed are depicted as green and red boxes, respectively.

In mice, the KvDMR1 region/ Kcnq1ot1 promoter on the maternal allele is methylated blocking the expression of the ncRNA KCNQ1OT1, which permits expression of genes
in close proximity, including *Kcnq1*, cyclin dependent kinase inhibitor 1C (*Cdkn1c*) and pleckstrin homology-like domain, family A, member 2 (*Phlda2*) [192]. Deletion of the KvDMR1 region results in loss of imprinting over most flanking genes on the paternal allele including the gene poor intervals between the ICR1 and KvDMR1[192, 193]. In humans, loss of imprinting over the KvDMR1 and biallelic expression of Kcnq1ot1 is observed in individuals with BWS [190]. The KvDMR1 has binding sites for CTCF [194]. This protein binds to the unmethylated KvDMR1 on the paternal allele and contributes to the expression of Kcnq1ot1 ncRNA.

Similar to Xist, the silencing role of ncRNA might be imparted by the recruitment of histone methyltransferases that induces repression of surrounding chromatin. Pandey and colleagues showed that in mice the 91 kilobase KCNQ1OT1 ncRNA that is mostly found in contact with the nucleolar compartment in placenta induces bidirectional repression of imprinted genes in the *Kcnq1* domain [195]. This is mediated through interaction of Kcnq1ot1 ncRNA with PRC2. In turn, G9a causes association of H3K27me3 and H3K9me1,2 silencing marks respectively.

In line with these observations, mice harbouring mutations in G9A or embryonic ectoderm development protein, a member of the PRC2 complex, lose silencing across some of the placental specific imprinted genes within the Kcnq1 locus [196]. Moreover, Terranova and colleagues demonstrated that the *Kcnq1ot1* is transcribed specifically from the paternal allele and localizes within the chromatin domain that is rich in transcriptionally non-permissive histone marks and free from permissive histone modifications and RNA polymerase II [197]. This region is highly compact on the paternal allele and is enriched for EZH2 and ring finger protein 2 (RNF2), members of the Polycomb repressive protein complex, which play roles in chromatin domain compaction. Together, these observations suggest that ncRNAs and histone modifying enzymes work in concert to make a higher order chromatin structure specifically over one allele, rendering this allele transcriptionally inactive.
1.2.3.1 KCNQ1 role in mediating cellular cAMP production by adenyl cyclases and associated diseases

The KCNQ1 gene encodes for the α-subunit of the voltage gated K⁺ channel (Iₓₛ) that together with other KCN family of proteins, is responsible for the slow delayed rectifying potassium current [198-201]. KCNQ1 was originally identified as a causative gene in specific cases of long QT syndrome (LQTS) [202]. LQTS is the result of prolongation of the QT interval of electrodiagram that reflects the delay in ventricular repolarization [203]. Currently KCNQ1 mutations may account for majority of LQTS cases. KCNQ1 loss of function can potentially lead to lethal cardiac arrhythmia and ultimately sudden cardiac death due to ventricular fibrillation [204].

Through KCNQ1 interaction with other effector molecules, the Iₓₛ responds to the sympathetic nervous system essential for physical and emotional adaptation. One of the characterized interactions, is the assembly of a macromolecular signaling hub around KCNQ1 mediated by the Yotiao protein [205-207]. The Yotiao protein (A-kinase anchoring protein 9; AKAP9), serves as a scaffold protein and brings a signaling complex to KCNQ1 [205-207]. This complex is comprised of Protein kinase A (PKA), protein phosphatase 1 (PP1), phosphodiesterase 4D3 (PDE4D3) and adenylyl cyclase 9 (ADCY 9). Activation of β-adrenergic receptors in response to sympathetic nervous system stimulation results in Adenosine 3′,5′-cyclic monophosphate (cAMP) production by ADCY 9 and consequently PKA activation. This leads to phosphorylation of KCNQ1 and activation of the K⁺ channel. Yotiao also gets phosphorylated by PKA. Perturbation of Yotiao phosphorylation impairs KCNQ1 activation suggesting that Yotiao possibly alters channel function.

1.2.3.2 Adenosine 3′,5′-cyclic monophosphate (cAMP)

cAMP acts as a second messenger by converting extracellular signals into intracellular ones, therefore playing a critical role in various cellular functions, including metabolism, cell growth, differentiation and gene expression [208-210]. Adenylate cyclases (ADCYs) are the proteins that convert triphosphate (ATP) into cAMP [211], while phosphodiesterases (PDEs) catalyze degradation of cAMP [212]. By binding to numerous transmembrane G-protein coupled receptors (GPCRs), extracellular stimuli modulate ADCY activity that leads to an increase or a decrease in cAMP levels. Cellular
cAMP level modulates activity of diverse proteins, including PKA. Phosphorylation of target proteins by PKA normally leads to their activation and results in various physiological outcomes. cAMP predominantly influences heart rate and contractility through β-adrenergic signaling pathway and it could be responsible for portions of heart failures. During instances of heart failure, overstimulation of β-adrenergic receptors by sympathetic nervous system leads to an increased cAMP levels that can potentially affect the heart function by promoting cell apoptosis [213, 214]. Voltage-dependent L-type Ca^{2+} channels (LTCC) is responsible for the Ca^{2+} influx essential for cardiac excitability and contraction of cardiac cells [215]. The Ca^{2+} influx initiates the release of intracellular supplies of Ca^{2+} from the sarcoplasmic reticulum that activates myofilaments contraction. Impairment in LTCC function associates with various cardiovascular diseases [216, 217]. cAMP can induce phosphorylation of LTCC to increase the intracellular level of Ca^{2+} required for heart contractibility. Thus, the regulation of cardiac myocyte contraction depends on fine-tuning of the cellular cAMP concentration.

Overall, as cellular level of cAMP regulates numerous cellular functions, therefore cAMP could be an indicator of the cell state and potentially could be used for prevention and treatment of various diseases.

1.2.3.3 The SNRPN imprinted locus is associated with PWS and AS

The mechanisms that govern SNRPN (Small Nuclear Ribonucleoprotein Polypeptide N) imprinted locus support the KCNQ1 imprinting model. The 2 megabase region on human chromosome 15q11-13 and mouse chromosome 7qC contains paternally imprinted genes including Snrpn, Necdin (NDN), makorin ring finger protein 3 (Mkrn3), MAGE (melanoma antigen family)-like 2 protein (Magel2), an antisense transcript of ubiquitin protein ligase E3A (Ube3a-ats) and one maternally expressed gene, Ube3a (Figure 1-12). This imprinted region is associated with Prader Willi Syndrome (PWS) and Angelman Syndrome (AS) in which the 15q11-q13 chromosomal regions are deleted/mutated from either the paternal or maternal chromosomes, respectively. Despite being caused by the similar genetic deletions on opposite parent-of-origin chromosomes, each syndrome has distinct symptoms [176, 177, 218-221].
In PWS patients, genes that are normally expressed from the paternal chromosome lose expression. Conversely in AS syndrome, the maternally expressed *UBE3A* gene, is most often repressed. The normal expression of these imprinted genes is coordinated by a bipartite imprinting center (IC). This center is differentially methylated and regulates chromatin structure and gene expression *in cis* [222]. Characterization of microdeletions in PWS and AS patients identified the regulatory regions in this locus responsible for establishment of correct imprinting [223, 224]. Mapping microdeletions from PWS patients identified a 4.3 kilobase region called the PWS-smallest region of deletion overlap (PWS-SRO) that includes the *SNRPN* promoter, exon 1 and the 5’ region of the first intron and plays a role in establishing the paternal epigenetic signature [178, 225]. By mapping microdeletions in AS patients, researchers have identified the AS-smallest region of deletion overlap (AS-SRO) region within IC that establishes the maternal epigenotype and is located 35 kilobases upstream of the first exon of *SNRPN* gene.

Imprinting of *UBE3A* has only been observed in the brain and in most other tissues this gene is biallelically expressed [186, 226]. The antisense transcript of UBE3A, UBE3A-ATS regulates the imprinted expression of *UBE3A* on the paternal allele [227]. Transcription of the UBE3A-ATS ncRNA starts from the PWS-SRO region and continues for another 460 kilobases [228]. The PWS-SRO region in humans contains a DMR. On the paternal allele, the unmethylated PWS-DMR is permissive for initiation of UBE3A-ATS ncRNA transcription, which drives through *SNRPN*, imprinted in Prader-Willi Syndrome (*IPW*) and *UBE3A*. As a result, *UBE3A* is repressed. On the maternal allele, the PWS-DMR region is methylated. Therefore, UBE3A-ATS ncRNA is not detectable but permissive for *UBE3A* expression.

### 1.2.3.4 DNA methylation and regulation of the SNRPN locus

The bipartite IC region encompasses both the PWS-SRO and AS-SRO domains. The IC region is deferentially methylated over the PWS-DMR region depending on the parent-of-origin. The differential methylation marks over this region are established upon fertilization based on the observation that the sperm DNA from individuals with PWS harbour a normal pattern of methylation. From this initial normal pattern of methylation, children lose methylation over the IC region, presumably contributing to the etiology of PWS [229]. Maintenance of imprinting marks at early stages of embryogenesis where
genome wide epigenetic reprogramming occurs is crucial. Kruppel-associated box-containing zinc-finger proteins (KRAB-AFP) ZFP57 and its cofactor KAP1 specifically binds to methylated IC region of SNRPN locus in mouse ES cells [230]. ZFP57 and KAP1 functions are necessary for maintenance of DNA methylation and histone modification of the IC region suggesting their essential role in regulating SNRPN locus early in development.

1.2.3.5 Histone tail modifications in regulation of the SNRPN locus

During spermatogenesis, protamines replace histones leading to removal of the normally chemically stable histone marks [231]. In contrast, the maternal allele encompasses repressive histone modification marks over the PWS-SRO region that is stable during gametogenesis. Following fertilization, histones with active marks including H3K4me1, 3 replace protamines on the paternal chromosome. Therefore, the maternal and paternal SNRPN locus is distinctly marked in gametes and this pattern is maintained in the zygotes.

Lymphoblastoid lines from PWS patients showed that the paternal chromosomes are marked with active chromatin markers such as histone hyper-acetylation over the PWS-SRO region while the same region was histone hypo-methylated over the maternal chromosomes [232]. In line with these observations, Perk and colleagues demonstrated that the AS-SRO region on the paternal allele has an inactive chromatin structure while the PWS-SRO has an open chromatin structure allowing activation of genes in close proximity [233]. Furthermore, the histone proteins associated with AS-SRO on the maternal allele are acetylated and tri-methylated at lysine 4 of histone H3 (H3K4me3) creating an open chromatin structure while inducing DNA methylation over the PWS-SRO region (Figure 1-13). These observations confirm that deletion of the PWS-SRO on the paternal allele causes repression of genes in cis in PWS patients but does not affect the maternal expression of the imprinted genes. It also underlines the role of histone marks, DNA accessibility and DNA methylation on the establishment and maintenance of closed and open chromatin structures that occur differentially over the maternal and paternal alleles.
Figure 1-12. Imprinted control regions and antisense transcripts regulate the SNRPN imprinted locus.

A schematic illustration of the SNRPN imprinted locus misregulated in individuals with PWS and AS. Normally, UBE3A and ATP10A are expressed from the maternal chromosome while the other genes in this locus are expressed only from the paternal chromosome. The PWS-SRO region is methylated on the maternal chromosome, which inhibits expression from the SNRPN promoter. On the paternal chromosome, the sequence encoded by the SNRPN promoter is a long non-coding transcript, which includes SNRPN, IPW, small nucleolar RNA (snoRNA) and UBE3A-ATS. Histone modifications have essential roles in establishment of imprinting over the SNRPN locus. The AS-SRO region associates with permissive histone marks (acetylated histone H3 (H3Ac) and H3K4me3) on the maternal allele (M). This induces DNA methylation (black oval) over the adjacent PWS-SRO region silencing the chromatin. On the paternal allele (P), the PWS-SRO region associates with active histone marks to activate the expression of adjacent genes. Genes expressed are colored in blue and those repressed in red.

1.2.3.6 The proteins involved in regulation of SNRPN imprinted locus

By studying two HS regions located at the promoter and first intron of SNRPN in PWS and AS patients, investigators identified the binding of a number of regulatory proteins including specificity protein 1 (SP1), nuclear respiratory factor 1 (NRF-1) and YY1 [234]. YY1 protein also interacts with regulatory regions of other imprinted genes [235, 236]. Using mice carrying deletion of the PWS-SRO region, Rodriguez-Jato and colleagues identified six DNase I HS regions over the Snrpn promoter, 5’ region of the
first exon and the first intron, spanning over the PWS-SRO region only on the paternal allele. These HS regions contain binding sites for YY1, SP1 and NRF-1 proteins as confirmed by chromatin immunoprecipitation (ChIP) experiments [237]. Although, a potential binding site for CTCF has been found over the Snrpn promoter, no research group has successfully demonstrated direct binding of CTCF to this region.

Another protein playing a regulatory role at the SNRPN locus is the methyl CpG binding protein 2 (MeCP2). As previously mentioned, the PWS-SRO region is unmethylated on the paternal allele. Therefore, it creates an open chromatin structure, which is permissive for expression of genes in cis including UBE3A-ATS. The UBE3A-ATS represses expression of UBE3A gene. On the other hand, the PWS-SRO region is methylated on the maternal allele. MeCP2 binds to this region on the maternal allele and possibly recruits histone-modifying enzymes to create a silenced allele. As a result, the genes in close proximity are silenced including UBE3A-ATS, which in turn allows expression of UBE3A. In individuals with Rett syndrome, in which MeCP2 fails to bind in this region, the silenced chromatin structure over the PWS-SRO region on maternal allele is lost [238]. This results in expression of some genes from maternal chromosomes, which are normally silent. Individuals with Rett syndrome share similar symptoms as AS patients where imprinting over many maternal alleles is lost and genes that are normally silent are expressed. Similarity between the expression pattern of AS and Rett individuals suggests a role for MeCP2 in regulation of imprinted genes on the PWS/AS locus.

1.2.3.7 The role of SMCHD1 in regulating the expression of imprinted genes

In SMCHD1 mutant mice, two imprinted loci showed modified expression patterns, Snrpn and Igf2r. The absence of functional SMCHD1 protein leads to loss of imprinting and biallelic expression of Igf2r genes and a few other imprinted genes in the Snrpn locus [172] [71]. Gain of biallelic expression over the imprinted genes was not due to the loss of DNA methylation or incorporation of the permissive histone marks of the ICR because their levels in SMCHD1 mutant mice were not significantly different from wild type mice. However, DNA methylation levels of the imprinted gene promoters remotely located from the ICR were affected.
Modified expression pattern of imprinted genes is evident in human neuroblastoma SH-SY5Y cells with reduced level of SMCHD1 protein [70]. Loss of SMCHD1 influenced two imprinted loci; H19/IGF2 and KCNQ1OT1 located on chromosome 15 that were not previously observed in mouse. The imprinted genes located in these loci acquired modified expression patterns mimicking the expression profile of the maternal chromosome. Presumably, genes that are normally expressed maternally became up-regulated while genes that are on the maternal chromosome became further repressed. Whether changes in gene expression upon loss of SMCHD1 caused the paternal chromosome to acquire maternal expression patterns remains to be determined. The expression of non-imprinted genes in this locus was also reduced upon SMCHD1 knockdown. The results infer the possibility that loss of SMCHD1 causes changes to the overall three dimensional structure of this locus since changes in gene expression was not limited to those governed by the ICR.

Overall, DNA methylation, histone modifications and transcription factors ensure establishment of imprinting temporally and in select tissues during development. It is still not understood how genomic imprinting is established in certain tissues while the same genes are biallelically expressed in others.

### 1.2.4. Clustered protocadherin genes, their genomic organization and epigenetic regulation

PCDH genes encode cell-surface adhesion molecules. They can be divided into two groups based on genomic organization: clustered PCDHs and non-clustered PCDHs [239]. The clustered PCDHs are further divided into three gene clusters of α (16 genes in human and mouse), β (19 genes in human, 22 genes in mouse) and γ (24 genes in human, 22 genes in mouse) (Figure 1-13). Clustered PCDH genes are interesting candidates for studying the regulation of monoallelic gene expression. First, clustered PCDH genes show stochastic patterns of allelic exclusion and are expressed combinatorially in Purkinje cells, which increases cellular diversity [240, 241]. Second, the long-range interactions and spatial organization of this cluster play an important role in establishment of allelic exclusion.
1.2.4.1.1 Stochastic or random monoallelic gene expression

Stochastic models of gene expression imply a single allele on one of the randomly selected chromosome is expressed while the other copy is silenced. Stochastic modes of monoallelic expression are observed in clustered PCDH, olfactory and immune cell receptor (antigen receptor) genes in select cells.

In 1965, Pernis and colleagues identified allelic exclusion of immune cell receptor genes [242]. This method of establishing monoallelic expression uses DNA recombination in immune cells to delete genomic regions in individual cells. This requires variable-(diversity)-joining [V(D)J] DNA recombination establishing the fate of the immune cell is established. Inefficient DNA recombination and subsequently a low probability of simultaneous DNA recombination might explain allelic exclusion of antigen receptor genes in immune cells [243]. Another model implicates timing or asynchronous replication for allelic exclusion [244]. A hallmark of monoallelically-expressed genes is asynchronous replication. This means that the active domains of the genome replicate during early S phase while the silenced genes, such as those on the inactive X chromosome, replicate during late S phase [245]. Following a successful rearrangement in an allele that replicates early, a feedback inhibition mechanism prevents DNA recombination over the other allele. Thus, recombination is prevented through DNA hyper-methylation and association of non-permissive or repressive histone marks over the inactive allele. In case of unsuccessful rearrangement, the second allele undergoes DNA recombination [246].

Another gene family with stochastic monoallelic expression is the olfactory receptor gene group that represents the largest gene family in vertebrates (2004 Nobel prize) [247]. In mice, this superfamily consists of about 1300 genes distributed among 27 clusters throughout the genome [248]. Each olfactory neuron expresses only a single olfactory receptor allele. The projections of olfactory neurons expressing the same allele extend to the same glomeruli in olfactory bulbs where they convey signals to mitral cells that ultimately transfer the signals to the brain [249]. Olfactory receptor gene expression provides guidance for neurons with similar sensitivity to form distinct glomeruli in olfactory bulb [250]. Misreading chemical signals occurs when olfactory neurons express multiple receptors [251-253].
Expression of one olfactory receptor gene from one allele in individual neurons elected from a large number of genes resembles the type of expression observed by immune cell receptor genes. In contrast to immune receptor genetic deletion strategy, olfactory receptors use epigenetic determinants. This was shown using cloned mice that were generated from nuclei of an olfactory neuron expressing a single receptor [254, 255]. The cloned mice contained neurons expressing one allele of a single gene from the whole olfactory receptor gene family, therefore ruling out DNA rearrangement as a mechanism for allelic exclusion of olfactory receptor genes.

Regulation of stochastic allelic exclusion of olfactory receptor genes occurs at multiple layers that are not necessarily exclusive; promoter choice, genomic organization and a feedback loop. They create a mechanism that explains the choice of a single olfactory receptor gene and stabilization of its transcription followed by inhibition of transcription from rest of the olfactory superfamily members.

The current findings suggest that the initial choice of an olfactory receptor gene is mediated by long-range interactions of distal regulatory elements with promoters. Previously two enhancers were identified, which mediated expression of few sensory genes in cis [256, 257]. However, recent data revealed existence of 35 enhancer elements that merge and make long-range interactions to regulate expression of olfactory receptor genes [258] Sequences in the olfactory receptor gene promoters might also influence promoter choice [259]. Differential affinity of factors bound to promoters and distal DNA regulatory elements might regulate promoter choice, therefore allelic exclusion. In addition, transient expression [260] and limited access to transcriptional de-repressors [261] might impact the promoter choice. DNA fluorescence in situ hybridization (DNA-FISH) revealed aggregation of inactive olfactory genes in heterochromatic regions enriched for transcriptional repressive marks (H3K9me3, H4K20me3), while active olfactory receptor genes located in euchromatic parts of the nucleus [261]. Lamin B receptor (LBR), a chromatin organizer that links chromatin to the nuclear membrane is essential for the nuclear organization of olfactory receptor genes [261]. Overall, spatial organization and compartmentalization of olfactory receptor genes restrict their access to transcriptional de-repressors. Current data suggest lysine specific demethylase 1 (LSD1) as a candidate protein that removes repressive marks from target
olfactory receptor genes [260]. Transient expression of LSD1 with an unknown demethylase that is able to demethylate H3K9me3 remove repressive marks from single olfactory receptor gene. The switching among olfactory receptor genes remains till one receptor triggers feedback signals. Translation of an olfactory receptor activates unfolded protein response in endoplasmic reticulum, which mediates activation of the Perk signalling pathway and consequently phosphorylation of translation initiation factor eIF2a [262]. eIF2a activates transient translation of Activating Transcription Factor 5 (ATF5) that in turn activates expression of Adenylyl Cyclase 3 (ADCY3). ADCY3 represses LSD1 that subsequently removes the unfolded protein signals and stabilizes the olfactory receptor gene expression leading to neuron maturation. Therefore, nuclear architecture, long-range chromatin interaction and limited expression of cellular de-repressors stabilizes individual gene expression and creates a framework for a feedback loop that inhibits expression of the other genes in the family.

Stochastic allelic exclusion of clustered PCDH genes is not well understood. Current data suggest that distally located regulatory elements are involved in modulating the monoallelic expression of clustered PCDH genes [263]. Clustered PCDH genes were identified in 1998 [239, 264] and their allelic exclusion was elucidated in 2005 [265]. This mechanism of activation ensures that one or a limited number of genes assembled in tandem will be expressed. However, it is not understood if interaction of the cis regulatory elements and gene promoters are required for regulation of allelic exclusion over the clustered PCDH genes and similar to olfactory receptor genes a feedback loop ensures expression of one gene from one allele.

1.2.4.2 Monoallelic and combinatorial expression of clustered protocadherin genes

Neuronal cell surface diversity is established in a similar fashion. Clustered PCDH genes are positioned in tandem along the same chromosome and each neuron expresses only a subset of these genes from one allele [265, 266]. This increases cell surface diversity providing a framework for self-recognition and non-self discrimination. Clustered PCDH receptors facilitate homophilic trans interactions between cultured K562 cells expressing distinct PCDH genes [267]. These interactions are highly specific and greatly restrict cell-cell recognition when non-identical sets of PCDH receptors are encountered. Cell
aggregation assays show specific homophilic interactions when the same five PCDH isoforms are expressed on the cell surface. Aggregation is blocked when only one isoform is mismatched indicating the cell-to-cell attachments are specific. However, further studies need to be conducted to investigate the role of clustered PCDH genes in cell-cell recognition in neuronal cells. Monoallelic expression of clustered PCDH genes provides a framework for dendritic self-avoidance [268]. Homophilic trans-interactions of PCDH proteins among the dendritic branches arising from a single neuron leads to repelling signals, which are essential for normal dendritic arborization. This is essential for self-recognition and non-self discrimination of neuronal cells in the neuronal wiring network.

1.2.4.3 Genomic organization of clustered protocadherin genes
In the mammalian genome, the clustered PCDH genes are assembled in tandem along the same chromosome, chromosome 5q31 in human and chromosome 18qB3 in mice. The α and γ gene clusters are arrays of variable exons that are placed in tandem (Figure 1-14). The variable exons are placed upstream of a set of three constant exons. Each PCDH α and γ protein is translated from a selected variable exon that is spliced to three constant exons. Each variable exon encodes for an extracellular domain, a transmembrane domain and a variable intracellular domain of a PCDH α or γ protein while three constant exons encode for a common distal intracellular domain. The variable exons in these clusters are divided into two groups: the C-type and non-C type exons. Analysis of PCDH α and γ clusters expression in Purkinje cells revealed that the C type variable exons show biallelic expression and are ubiquitously expressed. In contrast, the non-C type exons show combinatorial and monoallelic expression [240]. The β gene cluster (PCDH β) is an array of variable exons and lacks the constant exons in the α and γ family [269]. This means that only one exon encodes for the entire PCDH β protein opposite to PCDH α and γ proteins that require three exons (1 variable exon and three constant exons) (Figure 1-14). The PCDH β cluster exons are also expressed in a monoallelic and combinatorial fashion like the α and γ non-C type exons [241].
Clustered protocadherin genes are arrayed in tandem on the same chromosome in the mammalian genome.

A schematic model representing the organization of clustered PCDH genes in mouse and long-range chromatin interactions of PCDH α cluster forming a transcriptional active center in SK-N SH cells. Clustered PCDHs are divided into the α, β and γ clusters. The α and γ cluster are arrays of variable exons that are placed in tandem and upstream of three constant exons (red bars). Each PCDH α and γ protein is encoded by one variable exon and three constant exons. The variable exon encodes for the extracellular domain, the transmembrane domain and the variable intracellular domain while three constant exons encode for the cytoplasmic domain. The β cluster is an array of variable exons and lacks the constant exons. Therefore, each PCDH β protein is encoded by single exon and lacks the extended cytoplasmic domain characterized in PCDH α and γ proteins. For the PCDH α cluster, a schematic illustration of long-range chromatin interactions forming a transcriptionally active centre is shown. This chromatin organization places the HS7 and HS5-1 sites in vicinity of selected PCDH α genes. Inactive genes are excluded from the transcriptionally active site and form loops on the periphery. RAD21 (orange clamp) and RNA pol II (green ovals) drive the constitutive expression of alpha C2 gene. CTCF (pink oval) and RAD21 protein association bring the selected PCDH promoters (a4 and a10) as well as aC1 in close contact with enhancer HS5-1 and induce their transcription. The hypersensitive sites upstream of the γ cluster are shown to have regulatory roles for expression of the PCDH β cluster.
Each PCDH gene encodes a cell surface module consisting of six Extra Cellular domains (EC1-EC6), a transmembrane domain and the cytoplasmic domain [269]. The EC1 domain contains the Cys-X5-Cys motif that forms a loop due to disulfide bridges between the two cysteine residues [269]. This motif is conserved in clustered PCDH genes and is absent from the cadherin superfamily [269]. The EC1 domain also has an RGD (Arg-Gly-Asp) motif that can interact with beta-integrin [269]. In classical cadherin proteins, the EC1 region contains the hydrophobic pocket for homophilic adhesion (cell-cell contact) but this region is absent in the clustered PCDH α4 protein and possibly other clustered PCDH proteins [269]. Although, previous data suggested tetramerization of clustered PCDH genes [270], recent data favour a model in which clustered PCDH proteins form homo- and hetero-pentamers [267, 270, 271]. This may allow the formation of a homophilic binding pocket missing in the monomeric clustered PCDH proteins.

1.2.4.4 Long-range chromatin interactions and regulation of clustered protocadherin genes

For PCDH α and γ genes, the promoter is located just upstream of each distinct variable exon regulating expression of clustered PCDH genes [272]. These promoters are highly related and contain a conserved sequence element (CSE) with an essential role in driving transcription of PCHDs [272, 273]. The pre-mRNA includes the exon from which transcription started, plus all downstream exons including the three constant exons [272]. Upon post-translational modification, the most distally located exon adjacent to the 5’ cap is spliced together with the constant exons [273]. Therefore, the choice of a promoter from which transcription started indicates which variable exon is post-translationally cis-spliced together with the constant regions.

Studies have identified long-range chromatin interactions where DNase I HS regions regulate access to a second region over the clustered PCDH genes. The DNase I HS sites have open chromatin structure that renders them more sensitive to cleavage by DNase I. Transcription factors access the HS open chromatin structure correlating with active transcription. Two HS sites, HS5-1 and HS7 mapped to the PCDH α cluster [274, 275]. The HS5-1 region has an essential role for regulating monoallelic expression of clustered PCDH α1 to α12 and constitutive expression of α C1 but not α C2 [274, 275]. The HS5-1 region encompasses two CTCF binding sites referred to as HS5-1 α and
HS5-1 β [275]. Binding of CTCF protein to these regions is required for HS5-1 enhancer activity and neuronal-specific expression of PCDH α genes. In non-neuronal cells, association of neuron restrictive silencer factor (NRSF) within the HS5-1 region represses expression of the PCDH α cluster [275]. The HS7 region is involved in the regulation of most genes in the PCDH α cluster including α C2 [275]. In addition to the HS5-1 and HS7, a cluster control region is located 320 kilobases downstream of the PCDH γ cluster genomic region in mice [276]. This region encompasses 6 HSs (HS16, 17, 17’, 18, 19 and 20) that are known to exert regulatory effects on clustered PCDH β but not α or γ clusters, which further emphasizes the importance of long-range interactions in regulation of gene expression [276].

In addition, human genome wide analysis has identified CTCF protein binding sites located on the promoter of 12 clustered PCDH γ variable exons and the 3’ region of one constant exon suggesting a possible role of CTCF as an insulator in regulation of the promoter choice and monoallelic expression pattern of PCDHs [277]. In line with this finding, Monahan and colleagues demonstrated direct binding of CTCF and RAD21, a protein from the cohesion complex, to transcriptionally active PCDH α promoters over CSE, HS5-1 and active exonic regions [278]. Further research in mouse catecholaminergic neuronal tumour (CAD) cells showed that CTCF knockdown reduces transcript levels of certain PCDH α non-C type variable exons. While knockdown of RAD21 decreased transcript levels of the C type PCDH α variable exons [278]. These results indicate the essential and unique roles for both CTCF and RAD21 proteins in regulating expression of different types of exons located in the PCDH α cluster.

In agreement with the findings above, Guo and colleagues have proposed a model to explain the role of HS sites as well as structural and regulatory proteins in the promoter choice for regulation of gene expression [263]. In this model, CTCF/cohesion complex binding to the CSE defines active promoters. This complex binds to the HS5-1 and CSE and brings the active promoters within the vicinity of enhancers (Figure 1-13). This results in formation of chromosomal loop structures in which the inactive promoters are placed outside of the transcriptionally active “hub site”. In SK-N-SH neuroblastoma cells, RAD21 binding to the HS7 region and the α C2 promoter causes constitutive expression of the α C2 gene. RAD21 in the absence of CTCF binds to both the α C2 promoter and
HS7 region, suggesting the possibility that RAD21 places CTCF in close contact with the α C2 promoter [278].

Flanking the CSE found in all clustered PCDH gene promoters, a conserved sequence specific element (SSE) exists that is unique to each distinct variable exon [279]. The SSE is essential for expression of clustered PCDH genes. CTCF and YY1 proteins bind to both SSE and CSE within clustered PCDH α gene promoters and loss of CTCF reduces transcription from these promoters. Loss of CTCF also causes impaired dendritic arborisation and reduced expression level of 53 isoforms from clustered PCDH gene family in individual neurons from mouse cortex and hippocampus [280]. This emphasizes the role of CTCF in regulating the expression of PCDH that impacts neuronal diversity and dendritic arborization.

MeCP2 is another candidate protein with regulatory roles in expression of clustered PCDH genes. ChIP studies using an antibody specific to MeCP2 showed that MeCP2 binds to the promoter of clustered PCDH β1 and acts to repress gene expression [281]. Knockdown of MeCP2 in SH-SY5Y cells significantly increased the transcript level of the PCDH β1. Moreover, higher levels of PCDH β1 transcript could be detected in post-mortem brain cortical tissues of individuals with Rett syndrome in which MeCP2 is mutated and non-functional, providing further evidence for the critical role of this protein in expression of the PCDH β cluster.

Moreover, DNA methylation over the promoter and 5’ region of each exon influences transcriptional regulation of clustered PCDH genes. DNA hyper-methylation of these regions in mouse neuroblastoma cell lines, C1300 and M3 derived from brain tissues was sufficient to repress transcription of PCDH genes [282]. The promoter controlling the active allele of these monoallelically-expressed genes is DNA hypo-methylated while the repressed variable exons have hyper-methylated promoters [282]. The promoters of the C type exons that are biallelically expressed are hypo-methylated in both alleles [282]. Recent data suggest that differential methylation of clustered PCDH promoters by de novo DNA methyltransferase, Dnmt3b might explain the alternative promoter choice for expression during development [283]. This affects the number of PCDH genes expressed by each individual cell as well as dendritic arborization as individual purkinje
cells from DNMT3b knock out mouse expressed larger number of PCDH genes and presented impaired arborization.

1.2.4.5 The SMCHD1 role in regulating expression of clustered protocadherin genes

Besides CTCF, Rad21 and MeCP2, SMCHD1 is another candidate protein with regulatory roles in expression of clustered PCDH genes. Recently, two research groups showed that SMCHD1 mutation in mouse embryos results in up-regulation of clustered PCDH genes [69, 71]. The up-regulation of clustered PCDH genes is associated with hypo-methylation of their CpG islands (CGI). Moreover, SMCHD1 knockdown in human embryonic kidney (HEK293) cells results in elevated transcript levels of several PCDH β genes. In addition, up-regulation of clustered PCDH β genes occurs upon SMCHD1 knockdown in SH-SY5Y neuroblastoma cells [70]. However, opposite to previous results conducted in SMCHD1 mutant mice, up-regulated clustered PCDH β genes expression level does not associate with hypo-methylation of the promoter in human cells lines. The observed discrepancy might be due to the different regions that are assessed in these sets of experiments as well as different cell types or species of origin. Overall, observations from these experiments conducted in mice and human cell lines confirm a regulatory role for SMCHD1 in the expression of clustered PCDH genes.

1.3. Rationale, Hypothesis and Objectives of the thesis

1.3.1. Rationale

SMCHD1 is a non-canonical member of the SMC family of proteins [66, 67, 284]. Initially, it was identified as a modifier of the transgene variegation in mice and was recognized as essential for X chromosome inactivation in female mice [68]. Later studies suggested its role in the methylation of a subset of CpG islands in a late stage post-fertilization step [172]. In our preliminary studies, we purified SMCHD1 using a methylated DNA element that encompasses the growth hormone gene promoter in a DNA affinity purification assay suggesting that DNA methylation might impact SMCHD1 occupancy [70]. Chromatin modifiers often exist in high molecular weight complexes with
a series of other non-related proteins. SMCHD1 and other SMC family protein members are key components of multi-protein complexes with other effector molecules [31-34, 285]. As mentioned above, SMCHD1 forms a complex with HBiX1, a nuclear co-receptor that associates with heterochromatin protein 1 (HP1) to mediate formation of the X chromosome’s compact structure [173]. Yet, other proteins that interact with SMCHD1 and might be part of this complex or other multi-protein complexes remains unknown. The role of SMCHD1 in regulating X-linked genes and possibly the growth hormone gene may hold some promise. Furthermore, elucidation of the role of SMCHD1 in regulating expression of genes on autosomes will provide additional information regarding its function. In addition, SMCHD1 genome wide occupancy in relation to global levels of DNA methylation will delineate its role in regulating gene expression relative to epigenetic modification. Moreover, identification of SMCHD1 binding partners will provide insights into its role as a chromatin modifier that regulates gene expression.

1.3.2. **Hypothesis**

I propose that SMCHD1 regulates genes located on autosomes and its genome-wide occupancy is sensitive to the DNA demethylating reagent, 5 azacytidine. In addition, I propose that besides HBiX1, SMCHD1 regulates gene expression in complexes with other regulatory proteins involved in transcription regulation.

1.3.3. **Objectives**

1) Identify autosomal genes regulated by SMCHD1. By performing microarray analysis and comparing changes in gene expression upon reduced levels of SMCHD1, we investigated the role of SMCHD1 in genome-wide gene expression.

2) Determine genome-wide occupancy of SMCHD1 in relation to global levels of DNA methylation. By performing chromatin immunoprecipitation-sequencing (ChIP) in cells treated with DNA demethylating reagent, 5-azacytidine (5azaC), we identified genomic sites in which SMCHD1 recruitment was sensitive to 5azaC.
3) Identify SMCHD1 candidate binding partners. By tagging SMCHD1 with Strep II-FLAG (SF) and performing proteomic analysis in cells stably expressing SF-SMCHD1, we identified potential SMCHD1 interacting proteins.
Chapter 2.

SMCHD1 regulates genes on clusters with monoallelic expression pattern


Contributions: Shabnam Massah contributed to Figures: 2-3B, 2-3C, 2- 5, 2- 6, and 2-7. Rob Hollebakken contributed to Figures: 2-1, 2-2, 2-3A, 2-3D and 2-4A, 2-4B, 2-4C. Mark Labrecque contributed to Figure 2-4D. I was the main author of this paper. I was responsible for conceiving, designing and performing 80% of the experiments and figures.

In this paper I characterized the biological role of structural maintenance of chromosome hinge domain containing 1 (SMCHD1), a relatively uncharacterized protein. In this study, using chromatin immunoprecipitation and electrophoresis mobility shift assays I showed that SMCHD1 was recruited to growth hormone promoter in a methylation dependent fashion.

In addition, using shRNA knock-down of SMCHD1 in human embryonic kidney cells, I extensively profiled changes in gene expression and identified 385 genes in the human genome that are regulated by SMCHD1. I showed that majority of up-regulated genes were concentrated on the X-chromosome but some differentially regulated genes were also found on non-sex chromosomes. More significantly, I showed that SMCHD1 regulates genes that are expressed on non-sex chromosomes originating from one parent also known as genes with monoallelic expression. This includes the
protocadherin beta cluster and an imprinted cluster associated with the Beckwith-Wiedemann syndrome (BWS).

2.1. Abstract

Regulatory elements for the mouse growth hormone (GH) gene are located distally in a putative locus control region (LCR) in addition to key elements in the promoter proximal region. The role of promoter DNA methylation for GH gene regulation is not well understood. Pit-1 is a POU transcription factor required for normal pituitary development and obligatory for GH gene expression. In mammals, Pit-1 mutations eliminate GH production resulting in a dwarf phenotype. In this study, dwarf mice illustrated that Pit-1 function was obligatory for GH promoter hypo-methylation. By monitoring promoter methylation levels during developmental GH expression we found that the GH promoter became hypo-methylated coincident with gene expression. We identified a promoter differentially methylated region (DMR) that was used to characterize a methylation-dependent DNA binding activity. Upon DNA affinity purification using the DMR and nuclear extracts, we identified structural maintenance of chromosomes hinge domain containing -1 (SMCHD1). To better understand the role of SMCHD1 in genome-wide gene expression, we performed microarray analysis and compared changes in gene expression upon reduced levels of SMCHD1 in human cells. Knock-down of SMCHD1 in human embryonic kidney (HEK293) cells revealed a disproportionate number of up-regulated genes were located on the X-chromosome, but also suggested regulation of genes on non-sex chromosomes. Among those, we identified several genes located in the protocadherin β cluster. In addition, we found that imprinted genes in the H19/Igf2 cluster associated with Beckwith-Wiedemann and Silver-Russell syndromes (BWS & SRS) were dysregulated. For the first time using human cells, we showed that SMCHD1 is an important regulator of imprinted and clustered genes.
2.2. Introduction

Plasma growth hormone (GH) levels decline with age and contribute to decreased somatotropic axis signaling (GH releasing hormone [GHRH], GH and insulin-like growth factor -1 [IGF-1]) [286]. Understanding how GH is regulated will provide insight into events associated with declining levels of GH with age. It has been proposed that increasing GH levels in the elderly increases lean muscle tissue while decreasing adipose mass and may act to reverse some negative effects associated with aging [287].

The mammalian GH gene is expressed only in the pituitary and is dependent on the expression of a functional homeodomain containing transcription factor, the pituitary-specific Pit-1 protein (POU1-F1) [288]. During development in the absence of a functional Pit-1 protein, GH is not expressed resulting in a dwarf phenotype in mammals.

A distal locus control region (LCR) located ~14.5 kb upstream of the human GH-N gene is required for gene expression [289]. It is characterized by a series of pituitary-specific DNase I hypersensitive sites (HS) when expressed. The region representing the homologous LCR in rodent models is relatively uncharacterized, while the human LCR encompassing HSI and HSII represents an intergenic sequence that is homologous to mouse and rat genomic sequence.

The GH promoter is regulated by both positive and negative DNA elements through transcription factors and co-regulatory proteins [290]. Pit-1 binds to DNA elements in the promoter as well as the LCR [291]. In addition, promoter DNA methylation has been negatively correlated with gene transcription; loss of DNA methylation near the transcriptional start site is associated with GH gene expression [292-294].

Distally located DNA elements communicate with the promoter to regulate gene expression. Recombined bacterial artificial chromosome (BAC) transgenes in which distal elements are deleted, have proved useful for studying the influence of these elements on DNA methylation in cis [295]. Here we report the characterization of GH promoter DNA methylation in which the putative mouse LCR was deleted. The goal was to impair transcriptional expression via removal of the putative LCR to determine its influence on promoter CpGs methylation. We hypothesized that the hyper-methylation
would denote CpGs crucial for gene repression while the same CpGs when hypo-
methylated would be essential for gene expression. We defined this region as a
differentially methylated region (DMR). The direct role of promoter DNA methylation in
regulation of the GH gene is not understood. Factors responsible for mediating DNA
methylation dependent repression of the GH gene have not been identified. The ultimate
goal was to identify proteins that bind to the methylated GH DMR and promote
transcriptional repression. These studies should shed light on molecular mechanisms
directing long-term repression of the GH gene.

Our findings indicate that Structural Maintenance of Chromosomes hinge domain
containing-1 (SMCHD1) is a protein that can interact with the GH promoter and likely
regulates its expression. SMCHD1 is a non-canonical member of the structural
maintenance of chromosome (SMC) family. This family of proteins plays roles in
chromatin dynamics and condensation and DNA repair [57, 296]. These roles establish
DNA topology linking chromatin architecture with gene regulatory events. The SMC
family of proteins is characterized by a conserved ATPase globular domain consisting of
N- and C- terminal Walker A and B motifs characteristic of ABC-transporter ATPases
[296]. SMCHD1 lacks these discernable ATPase motifs found in authentic SMC proteins
and instead the ATPase domain of SMCHD1 resembles a GHKL (gyrase, HSP90, histidine kinase, MutL) ATPase domain found in the ATPase/kinase superfamily [297].
SMCHD1 homologues are found in vertebrates and in some plants [298].

The first part of this study focuses on the identification of a DMR, characterization of the
binding of a DNA methylation dependent protein and isolation of proteins that bind to the
DMR. For the first time, we show that SMCHD1 is likely a GH regulatory protein acting
directly through the promoter. In the second part, we used a genome wide approach to
identify autosomal genes regulated by SMCHD1. We confirm that SMCHD1 can repress
genes in the protocadherin β gene cluster and extends its targets to a gene cluster with
parent-of-origin imprinting associated with Beckwith-Wiedemann / Silver-Russell
syndromes (BWS/SRS, respectively).
2.3. Results

2.3.1. Identification of a GH promoter differential methylated region (DMR)

Several reports have identified DNA methylation as a regulatory component for expression of the pituitary-specific GH gene [292-294]. We found that genomic DNA from mouse pituitary was hypo-methylated over the GH promoter region. In order to study GH gene expression in real-time using mouse pituitary, we generated a series of BAC transgenic mice where the mouse GH gene was substituted with a homologous rat GH promoter and the coding sequence for red fluorescent reporter gene (wild-type, WT-GH:RFP) and another that contained an additional deletion in a putative upstream regulatory element or locus control region (LCR, ΔLCR-GH:RFP). The mouse LCR was identified by homology with the human locus [291, 299]. Transgenic mice carrying the recombined WT-GH:RFP BAC but not the ΔLCR-GH:RFP BAC expressed RFP only in the pituitary (Figure 2-1A).

Normally GH is expressed in somatotropes (GH+) and in some somatomamotropes (GH+ and prolactin, Prl+). To show that the RFP protein expression closely paralleled GH expression using the WT BAC, we labeled pituitary tissue sections from mice with anti-GH, anti-Prl and anti-thyroid stimulating hormone β (TSHβ) antibodies and visualized the presence of each cytoplasmic hormone with a fluorescein isothiocyanate (FITC)-conjugated secondary antibody (green, Figure 2-1B). These results demonstrated that pituitary of transgenic mice containing WT BAC transgenes is a suitable model system to study promoter DNA methylation levels compared to the ΔLCR version of the BAC transgene, which lacked RFP expression.

Next, we assessed the level of DNA methylation on the proximal promoter of the mouse transgenes. Genomic DNA was isolated from pituitaries and sequenced following bisulfite treatment. A number of clones were analyzed using pairwise statistics (displayed in a graphic plot in Figure 2-1C). Expression of GH:RFP paralleling GH expression coincided with significantly hypo-methylated CpGs at positions -8 through -6 as well as some in the RFP coding region. These same CpGs were fully methylated in
pituitary tissues from the ΔLCR transgenic mice where GH:RFP was silenced. In addition, pyrosequencing the targeted promoter CpGs demonstrated a similar trend (Supplementary Figure 2-8). In conclusion, GH:RFP BAC transgenic mice expressing RFP had significantly hypo-methylated CpGs at position -8 through -6 of the promoter region compared to transgenic mice, ΔLCR GH:RFP.
2.3.2. The GH promoter is heavily DNA methylated in dwarf mice pituitaries

Functional Pit-1 protein is required for expression of anterior pituitary secreted hormones including GH [288, 300]. In order to assess the level of promoter DNA methylation in pituitary cells lacking anterior pituitary hormones (GH, Prl and TSHβ), we compared DNA methylation levels in pituitaries from Snell dwarf mice (dw) with heterozygous wild type littermates (WT). Dw mice are characterized by an inactivating Pit-1 mutation [301]. To determine the level of methylation, we quantified bisulfite treated DNA using two methods. First, we directly sequenced a number of clones produced from PCR reactions following bisulfite treatment of genomic DNA (Figure 2-2A, upper panel). The GH promoter from dwarf mouse pituitaries was almost completely methylated immediately upstream of the Pit-1 binding sites (Figure 2-2A, CpGs -7 through -3, solid blue bars). Conversely, WT DNA samples were significantly hypo-methylated at CpG positions -7 through -3, consistent with transgenic mice carrying the rat GH promoter (Figure 2-2A, solid red bars). In a second experiment, following bisulfite treatment the PCR product was interrogated by DNA restriction digest targeting the CpG located at position -4 and showed that DNA hypo-methylation was only observed in the anterior pituitary from WT mice (WT, Figure 2-2A, lower panel). Together, these results demonstrate that the hypo-methylated promoter of the mouse GH gene was attributed to cells with functional Pit-1.
Figure 2-2. Comparative levels of GH promoter methylation in mice.
2.3.3. Postnatal hypo-methylation of the pituitary GH promoter

The developmental profile of GH gene expression in mice is well-established [301]. GH gene expression can be detected at around embryonic day (e) 15.5-e17.5 but not at e14.5. We were determined to characterize whether promoter demethylation correlated with the developmental expression of the GH gene: e14.5 (before GH expression), P0 (after GH expression) and P14.5 (postnatal GH expression). Bisulfite treated DNA and pyrosequencing revealed promoter hypo-methylation following GH gene expression including CpGs -8 to -5 (Figure 2B). Therefore, the GH promoter becomes hypo-methylated at a developmental time shortly after GH gene expression.

2.3.4. Treatment of cells with 5azaC relieves GH transcriptional repression

The pituitary is composed of many cells secreting a variety of hormones. We sought to investigate GH promoter methylation in clonally derived cells secreting a single hormone. Therefore, we chose cell types representing GH + (GC cells) and GH - (MMQ cells). Using manual bisulfite sequencing, we determined that the promoter DNA of GH + cells was largely unmethylated (Figure 2-3A top left panel, open circles) while that from GH - cells was hyper-methylated (Figure 2-3A top right panel, closed circles). Therefore, it is likely that DNA methylation in part restricts GH gene expression in MMQ cells.

In order to assess the role of DNA methylation on transcriptional repression in MMQ cells, we treated cells with a DNA demethylating agent, 5-azacytidine (5-azaC). Examination of mRNA levels revealed that 5-azaC relieved transcriptional silencing of
the GH gene (Figure 2-3A, lower panel) with little or no observable change in the transcription of Pit-1 and Prl. Together these results support that 5-azaC can relieve GH repression from a GH – cell line, characterized with a highly methylated DNA promoter.
Figure 2-3. Inhibition of DNA methylation relieves transcriptional repression of GH and characterization of a methyl-DNA binding activity.
A. GH gene repression can be relieved by treatment with 5-azaC. Above, schematic illustration of the CpG sites location in the rat GH promoter (the CpGs are numbered according to their relative position from the transcriptional start site). Middle, bisulfite sequencing of genomic DNA extracted from pituitary derived-GH+ (GC cells) or GH− cells (MMQ cells). The level of DNA methylation is displayed as unmethylated (open circles) or methylated (solid circles) from individual clones. Lower, GH− cells (MMQ) were treated with 5-azaC and the level of gene expression compared using RT-qPCR. The solid bars represent 5-azaC treated cells and the open bars, cells cultured under normal conditions. B. A methyl-specific binding protein binds to a region of DNA derived from the mouse GH promoter. Above, DNA sequence of the double-stranded oligonucleotides used in the EMSA. The position of the CpGs relative to the transcriptional start site are indicated and the position of a previously described binding site for the thyroid hormone receptor (TR) is boxed. Lower left, EMSA with MMQ nuclear extracts comparing the bound proteins from either methylated (lane 2) or unmethylated DNA probes (lane 5). A methylated competitor DNA (M) reveals a specific upper methyl-DNA protein-binding component (Lane 3, arrow). Lower right, an EMSA competition assay with nuclear extracts from pituitary-derived cell lines (MMQ, GC and GHFT) as indicated. Methylated or unmethylated (U) competitor oligonucleotides were used to identify the upper shifted component representing a DNA bound protein. All nuclear extracts contained a methyl-DNA binding protein. C. Methylation of CpGs at position −8 and −7 were required for recruiting a methyl DNA binding protein. Above, schematic of the result from the competition assay illustrated below. The shaded area indicates CpGs essential for the methyl DNA binding protein. Below, an EMSA with MMQ nuclear extracts (NE) and a methylated DNA probe with various competitor oligonucleotides as indicated. D. The methyl-DNA binding activity observed with the GH DMR appears to be unique. The sequence of the oligonucleotides used in the competition assay are listed in Table S1. An EMSA with increasing amounts of MMQ NE (6 and 12 µg) and a methylated GH DMR probe optionally in the presence of competitor oligonucleotides (100X) as indicated.

2.3.5. The methylated GH DMR recruits a methyl-DNA binding protein

DNA methylation can have profound effects on transcriptional gene regulation by either blocking or recruiting transcription factors [302]. In an attempt to determine how the DMR of GH influences the DNA binding activity of proteins, we used nuclear extracts and a GH DMR probe in electrophoresis mobility shift assays (EMSAs). MMQ cell nuclear extracts were tested for binding to either a methylated or unmethylated radiolabeled probe overlapping the DMR (Figure 2-3B, left panel). We obtained a strong upper band using nuclear extracts and the methylated probe but not with the unmethylated probe (see arrow, compare lanes 2 and 5). The specificity of binding was confirmed by efficient competition with methylated unlabeled competitor oligonucleotide (lanes 3,6). Next we assessed nuclear protein binding to a methylated radiolabeled probe by using nuclear extracts prepared from different pituitary derived rodent cell lines: MMQ, GC and GHFT. The mobility shift patterns were the same for GC and GHFT nuclear extracts (Figure 2-3B, right panel, lanes 5-10) while an additional lower band
was found using MMQ cells nuclear extracts (lanes 2-4). Furthermore, a methylated unlabeled competitor oligonucleotide competed for binding to the specific upper band (compare lanes 3 and 4, 6 and 7, 9 and 10). Taken together, this suggests that the GH DMR can recruit DNA binding proteins from number of cell lines in a methylation-dependent manner.

The DMR identified in this study contains four CpGs. To identify the methylated CpGs responsible for recruiting a methyl-specific binding protein, either the -8, -7 or -6, -5 were methylated and assessed in an EMSA competition assay (Figure 2-3C, compare lanes 2-6). Oligonucleotides methylated at positions -8 and -7 efficiently competed for binding while those methylated at positions -6 and -5, did not (compare lane 5 with lane 4). This suggests that the upstream methyl CpGs at positions -8 and -7 are likely sufficient to recruit a methyl-DNA binding protein, in vitro.

Some methyl-binding proteins require methylated cytosines on both DNA strands for binding [105]. Therefore, we tested if the protein binding was sensitive to upper and lower strand methylation or combinations thereof. Oligonucleotides with each strand individually methylated were not effective competitors (Figure 2-3C, lanes 8-11). Together, the methyl DNA binding activity described here likely requires symmetrically methylated CpG at positions -8,-7 of the rat GH promoter.

2.3.6. Characterization of GH DMR binding proteins with competitor oligonucleotides

A number of transcription factors regulating GH expression have been shown to bind within or near the DMR [303, 304]. For example, the thyroid hormone receptor site directly overlaps with the CpG at position -6. A series of competitor oligonucleotides were selected and tested for their ability to compete with the methyl-specific binding activity. Previously these oligonucleotides acted as efficient competitors presumably targeting the following transcription factors: the direct repeat 4, thyroid hormone receptor element (DR4-TRE) for the thyroid hormone receptor, the high affinity Mbox5 for zinc finger E-box binding homeobox 1 (Zeb1) [305], consensus Zeb 1 [306], the GC box for the specificity protein 1 (SP1) (Promega, Part # 9PIE639) family of transcription factors
and the high affinity MeCPG site for methyl CpG binding protein 2 (MeCP2) [307]. The result of the competitor EMSA can be found in the lower panel (Figure 2-3D) while a summary of the results is shown in the Table S1.

The DR4 TRE was an efficient competitor for the lower band (Figure 2-3D, compare lane 15 with 3). This suggests that the lower band likely represented the binding activity of a protein with similar sequence specificity and DNA affinity as the thyroid hormone receptor. The high affinity Zeb1 binding site, Mbox5, efficiently competed for binding to the band labeled “middle” (compare lane 18 with 3). Interestingly, a consensus Zeb1 binding site was an effective competitor for any of the bands (compare lane 21 with 3). Thus, the DMR oligonucleotide sequence likely contains a high affinity binding site for a protein(s) with similar sequence recognition and affinity as Zeb1. As well, the consensus GC box or SP1 family of protein binding site efficiently competed for binding to the middle band (compare lane 24 with 3). Together these results suggest that the probe used for these experiments contains multiple specific binding sites for proteins independent of DNA methylation.

In an attempt to further characterize the methyl binding activity, we used a high affinity MeCP2 competitor [307]. The results showed that the competitor was not able to displace the upper-methyl-specific binding activity (compare lane 27 with 3), likely ruling out MeCP2 as a candidate methyl DNA binding protein in this context.

2.3.7. **Enrichment of SMCHD1 using methylated GH DMR DNA affinity purification**

The EMSAs were unable to provide information on the identity of the DNA binding activity recruited to the methylated version of the DMR. Thus, to identify proteins recruited by the methylated DMR, we devised a protein purification strategy to isolate nuclear proteins using a methyl DNA affinity purification step based on a method described by Yaneva and Tempst [308] (Figure 2-4A and B upper). Individual fractions of nuclear extracts were collected by salt elution from a P11 phosphocellulose column and tested in an EMSA using a methylated DMR as a probe. The results showed that the middle, upper and lower specific binding activity in that order could be separated using
this strategy (Figure 2-4A). The middle band was eluted using 0.1M NaCl (fraction #14),
the upper methyl DNA-binding specific band was enriched by a 0.3M salt elution
(fraction #24) and the lower band was highly enriched by the 0.5M salt elution (fraction
#34). To provide support for the identity of proteins enriched in those fractions, we used
and immunoblot with an anti- thyroid hormone receptor (TR) antibody (Figure 2-4A,
upper right). The results showed that the thyroid hormone receptor was present in the
same fraction that was highly enriched with the lower shifted activity (fraction #34).
Together these and the competitor EMSA results show strong support that the lower
shifted component was likely due to the binding with thyroid hormone receptor.

A single high molecular weight band was the only unique protein band from the DNA
affinity column containing the methylated DMR (Figure 2-4B). Liquid chromatography-
mass spectroscopy analysis of the isolated band identified two peptides that matched
the mass predicted for amino acids sequences. Both of which were fragmentation
products of SMCHD1 polypeptide (Figure 2-4C and supplementary Table 2-2). Thus, our
data suggests that SMCHD1 can be a DMR regulatory protein, in vitro.
Figure 2-4. DNA affinity purification of SMCHD1 using a multimerized version of the methylated GH DMR.
A. Individual DNA binding proteins were separated by fractionation of MMQ nuclear extract. Upper left, schematic description of the step-wise elution of MMQ nuclear extract from a P11 phosphocellulose column. Below, EMSA analysis of individual samples collected from the fractionation. Fractions (#14, #24, and #34) isolated middle, upper and lower DNA binding proteins, respectively. Upper right, immunoblot of selected fractions to detect the presence of thyroid hormone receptor (TR). B. Isolation of a protein recruited by the methylated GH DMR. Top, a schematic illustrating the manipulation of the fraction (#24) containing the methyl-DNA specific binding activity. Three columns were generated with the multimerized GH DMR and the third and last one was methylated with SssI methyltransferase. Fraction #24 was applied sequentially to unmethylated columns and then to a third column to capture proteins with affinity to the methylated DNA. Lower, the proteins bound to each of the three columns were resolved on a gradient gel and visualized using a mass-spectroscopy compatible silver stain. A band retained by the DNA methylated affinity column is highlighted by an arrow. The protein was excised from the gel and analyzed by liquid chromatography and mass-spectroscopy. C. Identification of SMCHD1 as a molecular component recruited to the methylated DNA. Peptides (in red) matching the molecular mass of amino acid sequences predicted encoding SMCHD1. D. Binding of SMCHD1 to the GH promoter is sensitive to cells treated with 5-azaC. Upper, schematic illustration of the GH promoter and relative position of the primers used for the ChIP assays. MMQ cells were treated with 5-azaC or cultured under normal conditions and then processed in a conventional ChIP assays with anti-SMCHD1, anti-Pit1 or anti-E-cadherin antibodies. The samples were analyzed by quantitative PCR in triplicate and repeated a minimum of three times. The data is presented as percentage of the input. A Student's t-test was used to assess the statistical differences between untreated and 5-azaC treated cells. P-values are indicated above the sets (* indicates a P-value <0.05).

2.3.8. Interaction of SMCHD1 with the GH promoter in cells is sensitive to 5azaC

To investigate SMCHD1 binding to the GH promoter, we performed a chromatin immunoprecipitation (ChIP) assay with an anti-SMCHD1 antibody in MMQ cells with a preserved DNA methylation state or in cells treated with 5-azaC. The results showed that the anti-SMCHD1 antibody enriched the GH promoter only in untreated cells (Figure 2-4D and supplementary Figure 2-9). ChIP with anti-Pit1 antibody was used as a positive control and anti-E-cadherin (E-cad) antibody for estimating background. Together these data support the hypothesis that SMCHD1 can be recruited to the GH promoter in a DNA-methylation dependent manner in cultured cells. We attempted knock-down of SMCHD1 in MMQ cells to demonstrate a direct role for regulation of GH gene expression. However, all attempts to lower the level of SMCHD1 failed in these cells (data not shown). A stable cell line that expressed a Strep-FLAG-tagged SMCHD1 protein in human kidney derived HEK293 cells confirmed that SMCHD1 protein can bind to the GH promoter (Supplementary Figure S2-9) and enrichment was lost upon
treatment of cells with 5-azaC. These results are consistent with the results of the ChIP experiments in MMQ cells.

2.3.9. **Microarray analysis confirms that loss of SMCHD1 up-regulates X-linked genes**

In an attempt to further understand the role of SMCHD1 in gene expression, we generated a series of HEK293 cell populations with knock-down levels of SMCHD1 using a short hairpin (sh) retroviral delivery system and performed microarray analysis to determine the influence of reducing SMCHD1 protein level on genome-wide mRNA levels. The knock-down levels of SMCHD1 in these cells was visualized using an immunoblot with whole cell extracts (Figure 2-5A). shRNA3 and 4 were most effective in knocking-down SMCHD1 levels and thus were used to generate microarray gene expression data comparing the levels of shRNA 3 and 4 infected cells with those infected with a control shRNA, NC5. Triplicates of each data point were generated to produce a P-value and provided a level of confidence in the observed data. Using t-tests between pairs of data sets, those with the P-value <0.05 and an intensity difference greater than 1.8 fold were selected and defined as genes or loci IDs influenced by the loss of SMCHD1. A total of 385 gene “IDs” were identified. Of these, 115 were up-regulated while 270 were down-regulated (Appendix A. Supplementary Table S2-1 and S2-2, respectively). Hierarchical clustering revealed that the expression profile of KD3 (shRNA3) and KD4 (shRNA4) knock-down samples were more similar to each other than to the control samples, CTL (NC5) (Supplementary Figure 2-10, tree structure in the top left). Pie charts illustrate the chromosomal distribution of SMCHD1 differentially regulated genes (Figure 2-5B). As expected the majority of up-regulated genes (27 percent) were localized to the X-chromosome.
Figure 2-5. Profiling gene expression from cells with knock-down levels of SMCHD1.
A. Retroviral shRNA directed towards SMCHD1 efficiently down-regulated SMCHD1 protein levels in HEK293 cells. shRNAs directed towards SMCHD1, a control shRNA or empty plasmid (pQCXIP) was used for retroviral infection HEK293 cells. NEs were prepared from stably infected cells and analyzed by immunoblot with an anti-SMCHD1 antibody. An anti-LSD1 antibody was used as an internal control for loading. B. A disproportionate number of genes were up-regulated on the X-chromosome in SMCHD1 knock-down cells. A pie chart was used to illustrate the percentage of genes on each chromosome that were up- or down-regulated in SMCHD1 knock-down cells. C. Heat map and hierarchal clustering of selected up- and down-regulated genes in SMCHD1 knock-down cells or cells infected with control non-specific NC5 shRNA. Below, scaling of the fold differences of the genes from cells. Intense red indicates up-regulation and intense blue indicates down-regulation.

2.3.10. SMCHD1 regulates gene clusters on autosomes

Previous reports suggested that SMCHD1 may regulate gene clusters with monoallelic expression patterns including genes with parent-of-origin monoallelic expression [69, 71]. We identified differentially regulated genes upon loss of SMCHD1 in HEK293 cells that have been previously shown to exhibit monoallelic expression (Supplementary Table 2-3). We found two clusters of genes, including members of the protocadherin β gene cluster and imprinted genes associated with BWS/SRS including: potassium voltage-gated channel KQT-like subfamily member1 (Kcnq1); H19 and cyclin-dependent kinase inhibitor 1C (Cdkn1C). The expression of those gene sets are illustrated in a heat map using hierarchical clustering (Figure 2-5C). Upon knock-down of SMCHD1, all differentially regulated genes of the protocadherin β cluster were up-regulated (Figure 2-6A, red). We tested knock-down of SMCHD1 in a more relevant cell line SH-SY5Y cells, a neuroblastoma derived cell line. RT-qPCR confirmed that PCDHB 3, 8, 11 and 14 were significantly up-regulated following SMCHD1 knock-down (Figure 2-6B). To determine if DNA methylation was altered, we performed bisulfite pyrosequencing on the PCDHB 10 promoter but DNA methylation was not significantly changed (Supplementary Figure 2-11). Overall, these results suggest that through an unknown mechanism, SMCHD1 may repress gene expression of some members of the protocadherin β cluster. Conversely, three genes associated with BWS/SRS (Kcnq1, H19 and Cdkn1c) are positioned in a cluster at chromosome 11p15.5 and are down-regulated in SMCHD1 knock-down HEK293 cells (Figure 2-5C and 2-7A). Typically these genes are expected to be expressed from the maternal allele and are silenced on the paternal chromosome[309]. Genes in this locus are regulated through two differentially
methylated regions, KvDMR1 (imprinting control region 2 (ICR2)) and ICR1. The results showed that in SH-SY5Y cells genes typically expressed only on the maternal chromosome were up-regulated in knock-down cells (H19, Kcnq1, Cdkn1C) while those typically imprinted (silenced) on the same chromosome were further repressed including the non-coding Kcnq1ot1 (Figure 2-7B). The effects on tyrosine hydroxylase (Th) and insulin-like growth factor 2 (Igf2) expression are unclear. In addition, two non-imprinted genes, nucleosome assembly protein 1-like 4 (Nap1l4) and cysteinyl-tRNA synthetase (Cars) were down-regulated upon knock-down of SMCHD1. In summary, the imprinted cluster associated with BWS and SRS was dysregulated in SMCHD1 knock-down HEK293 and SH-SY5Y cells.
Figure 2-6. The protocadherin β cluster genes were differentially regulated in SMCHD1 knock-down cells.

A. A number of genes in the protocadherin β cluster were up-regulated in SMCHD1 knock-down SH-SY5Y cells. A graphical representation showing the position of differentially regulated genes on human chromosome 5. Below in red is the corresponding position of the upregulated genes in the protocadherin β cluster (5q31.3) upon loss of SMCHD1. B. mRNA quantitation of selected protocadherin β genes using RT-qPCR in SMCHD1 knock-down SH-SY5Y cells. The copy numbers are relative to and corrected using β-actin cDNA levels. * indicates P-values <0.05, ** P-values <0.01 and *** P-values <0.001 using an unpaired Students t-test.
Figure 2-7. The H19/lgf2 imprinted locus was dis-regulated following SMCHD1 knock-down.

A. A number of imprinted genes associated with BWS and SRS were dysregulated in SMCHD1 SH-SY5Y knock-down cells. A graphical representation of the H19/lgf2 locus on the human chromosome at position 11p15.5. Maternally imprinted genes are highlighted in blue, maternally expressed genes are in red, one of the placental-specific imprinted genes is colored in brown and not imprinted genes are in green. A non-coding RNA, Kcnq1ot1 is colored in blue and is typically expressed from the paternal chromosome presumably acting to silence genes normally expressed from the maternal chromosome (M) including Kcnq1 and Cdkn1c. Differentially DNA methylated regions (ICR1, KvDMR1 (ICR2)) are indicated by the trapezoids (solid indicates hyper-methylation and open hypo-methylation). B. mRNA quantitation of selected genes in the H19/lgf2 locus using RT-qPCR in SMCHD1 SH-SY5Y knock-down cells. The copy numbers are relative to and corrected using β-actin cDNA levels. * indicates $P$-values <0.05, ** $P$-values <0.01 and *** $P$-values <0.001 using an unpaired Students t-test.

2.4. Discussion

In this study, we demonstrated that SMCHD1 binds to the GH promoter and is likely a new regulatory protein for GH gene expression. Further, we identified a GH promoter DMR and characterized the binding of a methyl DNA binding protein to the methylated DMR. In addition, we show that SMCHD1 can repress the protocadherin β gene cluster and extends current findings to regulation of a second imprinted gene cluster associated with BWS/SRS.

2.4.1. SMCHD1 is involved in regulation of the GH gene

GH gene regulation in the pituitary provides an excellent model to study the effects of promoter DNA methylation in a non-CpG island context. GH plays a central role in modulating growth of some organisms from birth until puberty. In humans, GH secretion declines by 14% per decade in adulthood, with GH deficiency frequently occurring beyond 60 years of life [310-312]. GH hormone replacement therapy used in GH deficient individuals can increase muscle mass and strength, bone mass and quality of life while decreasing fat mass [287]. Somatotrope cells secrete GH, the predominant cell-type in the pituitary. The GH secreting lineages as well as two others, secreting Prl and TSH-β, lactotropes and thyrotropes, respectively, emerge from anterior portion of the pituitary. The identity of these cells is established by expression of a functional Pit-1
protein. Understanding mechanisms that govern GH regulation is useful to provide insight into growth and age related changes in GH production.

The GH gene is regulated by distal elements that are positioned over 13,000 bases upstream of the GH transcriptional start site. These include an evolutionarily conserved putative LCR that was originally characterized using human chromosome fragments and transgenic reporter genes in mice [299]. The distal LCR was defined as pituitary-specific DNaseI hypersensitive sites concomitant with gene expression. Interestingly, Lunyak et al. characterized a chromatin boundary element near the LCR using recombined human BAC transgenes in mice [313]. The proximal promoter has cis-regulatory elements located in the first 320 bp upstream of the transcriptional start site [314]. This minimal cassette, in the absence of the surrounding DNA sequences and chromatin is sufficient to drive expression of GH in somatotropes and repression or silencing in lactotropes using randomly inserted mouse transgenes [290]. However, in the presence of surrounding DNA and chromatin, the LCR is absolutely required for the expression of the GH gene.

Analysis of promoter CpGs using wild-type mice was not informative for identification of CpGs that were unmethylated and crucial for gene expression because most promoter CpGs were methylated to the same degree (Fig 2-2A and B). Here, we show that BAC recombined transgenes are useful for identification of CpG sites that can bind to regulatory proteins. Using this model, we identified three promoter CpGs that were hypo-methylated (-8,-7,-6) and was defined as the GH DMR (Figure 2-1).

The methylated version of the GH-DMR recruits a methyl-DNA binding protein (Figure 2-3B). We identified SMCHD1 as a DMR DNA binding protein in vitro (Figure 2-4C). In cells, we determined that SMCHD1 was bound to the methylated DNA preserved GH promoter and was dismissed upon pretreatment of cells with 5-azaC (Figure 2-4B). This is the first example demonstrating that SMCHD1 is a component of the machinery that could regulate GH gene expression.

SMCHD1 plays a role in the maintenance of DNA methylation on the inactive X-chromosome (Xi) and is essential for female viability [68]. During development, the Xi
acquires CpG island methylation in two phases: a rapid and late phase. The late phase is SMCHD1 dependent [172]. In cells, the Xi forms a compact structure referred to as barr body. SMCHD1 and HBiX1, a heterochromatic protein 1 (HP1) binding protein, together act to link the trimethylated histone H3 lys9 (Me3H3K9) and XIST- trimethylated histone H3 Lys27 (Me3H3K27) chromatin domains to organize the compact Xi structure [173], suggesting a role for SMCHD1 in organizing chromatin domains.

In humans, a deletion mutation (K274del) in SMCHD1 located at a conserved residue in the GHKL ATPase domain combined with a permissive 4q35 allele was suggested as a cause for Facioscapulohumeral muscular dystrophy type 2 (FSHD2) [315]. An autosomal dominant version of the disease mapped to 4q35 or more specifically a D4Z4 repeat. Whole exome sequencing of individuals from seven unrelated family members with FSHD2 identified heterozygous out-of-frame and missense mutations as well as splice variants of SMCHD1 (79% of those tested) [73]. In addition, ChIP analysis revealed that SMCHD1 bound directly to the D4Z4 metastable repeats with each repeat containing an open reading frame encoding double homeobox 4 (DUX4) [73]. Moreover, individuals with the autosomal-dominant FSHD type I form of the disease have at least one allele with 1-10 copies of a D4Z4 repeat. Typically, the general population has between 11-150 repeats further demonstrating the repeat is highly polymorphic. Families with 8 to 10 copies of the repeat were affected by SMCHD1 mutation, suggesting that it is a modifier of the type I form of disease [316].

2.4.2. **SMCHD1 is involved in regulation of genes on autosomes**

Homozygous SMCHD1 mutation was lethal in male mice with mixed backgrounds suggesting SMCHD1 is essential for regulation of genes on autosomes [71]. Our results support a major biological role for SMCHD1 in mediating silencing on the inactive X chromosome (Figure 2-5B). By mapping the position of differentially regulated genes to autosomes, we also found an unusual number of genes that have been characterized as imprinted or having monoallelic expression profiles (Figure 2-5C and supplementary Table 2-3).
2.4.3. The role of SMCHD1 in regulation of genes associated with BWS and SRS

BWS/SRS are reciprocal diseases, characterized by dysregulation of a cluster of genes with parent-of-origin imprinting. We identified imprinted genes belonging to the H19/ Igf2 and Kcnq1 locus located on chromosome 11p15 (Fig 2-5C and supplementary Table 2-3). This region has been shown to be associated with forms of SRS and BWS syndromes. BWS cases are sporadic and are characterized by different imprinting defects including defects in the ICR1 region that results in loss of imprinting of Igf2 and hyper-methylation of H19 and in most cases loss of methylation of the KvDMR1 that leads to gain of the non-coding Kcnq1ot1 transcript and loss of Kcnq1 and Cdkn1c expression [317]. Similarly, most SRS cases are also sporadic and are characterized by loss of methylation at the ICR1 locus and gain of expression of H19 at the expense of Igf2 expression [318]. In a more relevant cellular model, SH-SY5Y neuroblastoma cells, knock-down of SMCHD1 displayed dysregulation of BWS and SRS associated genes. All three genes (Kcnq1, H19 and Cdkn1c) found down-regulated in HEK293 cells were up-regulated in SH-SY5Y cells (Figure 2-7B). Differences in chromatin organization may account for distinct gene expression outcomes observed between the two cell lines. Th and Igf2 genes have mixed results, while non-imprinted genes in the locus, Nap1l4 and Cars, were down-regulated upon SMCHD1 knock-down. Interestingly, the non-coding transcript Kcnq1ot1 was down-regulated. This data supports a model in which the locus adopts imprinting patterns typical of the maternal chromosome as observed in patients with SRS. Determining gain or loss of nucleotide diversity in the mRNAs would conclusively demonstrate if biallelic expression of genes were encoded by the H19/Igf2 locus following knock-down of SMCHD1.

In this study, we present evidence that SMCHD1 is important for regulation of genes on the inactive X-chromosome and non-sex chromosomes in human cell lines. Identification of genes associated with BWS and SRS extends our understanding of the role of SMCHD1 in regulating imprinted clusters beyond that associated with Prader-Willi syndrome (PWS) and Angelman syndrome (AS) [69, 71]. BWS and SRS are congenital disorders with opposite outcomes on prenatal and postnatal growth: gigantism and dwarfism, respectively [319]. BWS is characterized by three major features: overgrowth,
macroglossia and anterior abdominal wall defects from diastasis recti to exomphalos [320]. The incidence of BWS is 1/13,000 [321] representing 300 births per year in the United States. SRS is characterized by intrauterine growth restriction and characteristic facial features [322]. Meta-analysis for tumor risk determined that 13.7% of individuals afflicted with BWS developed tumors [323]. The majority were Wilms tumors, hepatoblastomas, rhabdomyosarcomas and neuroblastoma. Thus, SMCHD1 is a key regulatory protein for correct expression of imprinted genes.

2.4.4. SMCHD1 represses the expression of genes from the protocadherin B cluster

Protocadherins are involved in mediating cell-to-cell contact/signaling, especially in human neuroblastoma cells [324]. An eleven-zinc finger protein (CTCF) in cooperation with the cohesion complex, which contains structural maintenance of chromosomes 3 (SMC3) protein play significant roles in monoallelic and combinatorial expression of protocadherin genes required for proper neuronal differentiation [263]. Previous studies showed that SMCHD1 regulates clustered protocadherin genes (α, β and γ) in mouse embryos [69, 71]. For the first time we showed in human cells that SMCHD1 also regulates expression of protocadherins β genes but this may be limited to the β cluster (Figure 2-5C and 2-6). Some studies demonstrated genes of the α and γ cluster were expressed from only one allele [240]. Moreover, the protocadherin β cluster has been shown to adopt monoallelic and combinatorial expression in Purkinje cells in mouse brain [241]. Only a subset of protocadherins are expressed in each neuron and only from one allele selected randomly. However, a number of protocadherin β genes are activated in SMCHD1 knock-down cells, therefore SMCHD1 may play important roles in defining molecular events governing neuronal networks.

2.4.5. Future Directions

In the future, identification of proteins associated with SMCHD1 will provide additional information on the role of SMCHD1 in chromatin-mediated events. In addition, it will be interesting to explore whether SMCHD1 is directly required for the establishment or
maintenance of monoallelic gene expression and whether the up-regulation of genes or non-coding RNA (i.e. H19) was due to a switch from mono- to bi-allelic expression.

2.5. Acknowledgements

We would like to thank Havilah Taylor from the Rosenfeld lab at UCSD for carrying out the husbandry for the transgenic mice used in this study and Dr. Frank Lee for helpful remarks and critical reading of the manuscript.

2.6. Materials and Methods

2.6.1. Ethics Statement

This study was carried out in strict accordance with the recommendations in the Guide for Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Institutional Animal Care and Utilization Committee of the University of California, San Diego.

2.6.2. Cells, Antibodies and Reagents

Cell lines used in this study were: MMQ (ATCC, CRL-10609), GC [325], GHFT [326], HEK293T (ATCC, CRL-11268) and SH-SY5Y (ATCC, CRL-2266). Antibodies used in this work included: anti-GH, anti-Prl (prolactin), anti-TSHβ, anti-Pit-1 (1769), anti-TRα/β (Santa Cruz, sc-772, fl-408), anti-FLAG (Sigma, Cat. # F3165), anti-E-cadherin (Santa Cruz, sc-7870, H-108) and anti-SMCHD1 antibody (Abcam, Cat. # ab31865). 5-Azacytidine was purchased from Invitrogen.

MMQ cells from rat pituitary were cultured in Ultraculture™ media (Biowhittaker, Cat. #12-725F) without L-glutamine and supplemented with 5% fetal bovine serum (FBS) and were maintained in a humidified atmosphere containing 5% CO₂ at 37°C. HEK293T, SH-SY5Y and GHFT cells were cultured in Dulbecco’s Modified Eagle’s Medium (DMEM; Gibco) containing 4.5g/L Glucose and L-Glutamine (Biowhittaker, Cat. # 12-604F) and
supplemented with 10% fetal bovine serum (FBS). GC cells were cultured in Dulbecco’s Modified Eagle’s Medium (DMEM; Gibco) containing 4.5g/L Glucose and Glutamax (Invitrogen, Cat. # 10566-016) supplemented with 12.5% horse serum and 2.5% FBS. Where indicated, cells were treated with 10uM 5-azaC every 24hrs for a minimum of 72hrs.

2.6.3. BAC recombination

WT GH:RFP BAC plasmids have been described elsewhere [313]. The mouse BAC 418O11 (Genbank accession number: AL604045.7) from the RPCI-23 library (http://bacpac.chori.org/) was used for recombination and contains 221.4 kb from chromosome 11 (nts:106,217,952-106,439,350, build Mouse Dec. 2011 (GRCm38/mm10) Assembly). This region includes 137.5 kb upstream of the major GH transcriptional start site to 82.3 kb downstream of the 3’UTR. The region deleted from the LCR BAC removed sequences -14,891 to -14,723 upstream of the GH transcriptional start site. The mouse GH promoter (starting at -316) and gene ending at nt +1,715 (downstream of the GH transcriptional start site) in the GH 3’ UTR were replaced by the homologous rat GH promoter and the coding sequence for DsRed derived from (Genbank accession number: AF506026). The sequence replaced included the rat promoter sequence -308 to +6 nts (chromosome 10: 95,694,111 -95,694,425, build Rat Nov. 2004 [Baylor 3.4/rn4]) fused to the DsRed through a short linker (127 bp) containing a sequence derived from the human GH-5’ UTR, 1st exon and the first intron (chromosome 17: 61,996,064-61,996,190, build Human Feb. 2009 (GRCh37/hg19) Assembly) beginning at position +8 of the human GH1 transcript. The splice between the human sequence with the DsRed encoding sequence was designed in a way that the initiating codon would begin with the DsRed coding sequence.

The LCR deletion was created in a similar way as described for the SineB2 deletion [313]. Briefly, 50 bp arms flanking the deleted region were used to amplify a galactose kinase (GalK) gene by PCR and the product electroporated into the rGH:RFP BAC containing SW102 E.coli strain. A recombination event was selected using minimal media plates containing galactose as the sole sugar source. Colonies positive for a
recombination event (GalK ΔLCR rGH:RFP) into the desired locus were verified by PCR and Southern analysis and then induced for a second recombination event. Synthetic oligonucleotides with the same arms but lacking the intervening GalK gene were electroporated into the GalK rGH:RFP containing BAC SW102 strain. The recombination event was negatively selected on 2-deoxy-galactose (DOG) containing minimal plates for elimination of the selectable GalK gene. The recombination event was verified using PCR and the integrity of the BAC confirmed using DNA fingerprinting. Finally, the LCR deletion was confirmed by sequencing the PCR product using oligonucleotides flanking the area targeted for deletion.

The BAC transgenes were amplified, purified and microinjected into 2-4 cell stage embryos in the University of California, San Diego transgenic core facility. Pup tail ends were genotyped with oligonucleotides recognizing the rGH:RFP component of the BAC. Those positive for the recombined BAC were bred and confirmed for germ line transmission and further propagated as individual mouse lineages.

2.6.4. Pituitary imaging and Tissue immunofluorescence

Fresh mouse pituitary and hindbrain tissue were carefully dissected and directly imaged using a rhodamine filter and then bright field microscopy. Fixed tissue sections were labeled with antibodies to GH, Prl or TSHβ and FITC and visualized with Cy2 conjugated secondary antibodies (Jackson Labs). Images were captured using a rhodamine and FITC filters.

2.6.5. Manual bisulfite sequencing and CoBRA

Following dissection of pituitaries or harvesting of cells, genomic DNA was prepared using the Qiagen Blood & Cell culture kit (Qiagen, Cat. # 13323) and stored at -20C. Genomic DNA was treated with bisulfite as previously described [327] or in other instances using the Imprint DNA Modification Kit (Sigma, Cat. # MOD50-1KT). The DNA was then amplified by PCR and products were cloned using the Stratagene pGEM T-easy kit (Promega, Cat. # A1360). A number of individual colonies were selected and prepared for sequencing (Eurofins MWG Operon). The sequencing results were
analyzed using the quantification tool for methylation analysis (QUMA, http://quma.cdb.riken.jp/) for statistical analysis.

For combined bisulfite restriction analysis (CoBRA), the PCR samples were digested with FokI to reveal the ratio of unmethylated vs methylated Cs. FokI cuts bisulfite modified DNA (C→T), unmethylated. The oligonucleotides used for amplification are listed in (Supplementary Table 2-4).

2.6.6. Reverse transcription quantitative PCR

RNA was prepared using Trizol (Life Technologies, Cat. # 15596018) according to the manufacturer’s protocol. Approximately, 100ng of RNA was reverse-transcribed using Superscript II (Life Technologies, Cat. # 18064-014) according to the manufacturer’s protocol. The cDNA was quantified using SYBR Advantage qPCR Premix (Clontech, Cat. # 638320), Rox as an internal standard and a StepOne Real Time PCR System (Life Technologies). For quality control, the melting curve for each primer set was verified and the PCR product ran on a agarose gel for detection of a single band of the expected length. The oligonucleotides are listed in Table S5. Following qPCR, the threshold levels were adjusted manually to logarithmic part of the curve to obtain a Ct value. The Ct values were normalized with those obtained by quantitation of the β-actin message and then the relative mRNA levels were determined. Where appropriate the data was analyzed using a Student’s t-test and the level of confidence displayed as P-values.

2.6.7. SDS-PAGE and immunoblot

Denatured proteins samples were separated on sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gels. DNA affinity purified proteins were run on a gradient gel (Invitrogen tris-tricine based buffer system) and were visualized using a mass-spectroscopy compatible silver stain kit (SilverQuest, Invitrogen, Cat. # LC6070) according to the manufacturer’s protocol.
For preparation of whole cell extracts (WCE), cells were pelleted and washed one time with PBS then lysed in one pellet volume of RIPA buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 1mM EDTA, 1% Sodium deoxycholate, 0.1% SDS) supplemented with a protease inhibitor cocktail (Bioshop, Cat. # PIC003). The re-suspended samples were placed on ice for 20 minutes, vortexed and then centrifuged for 5 min at maximum speed (14,000 xg). The protein supernatant was quantified and equal quantities diluted in sample buffer and boiled for 5 minutes. Proteins were separated on 6% SDS-PAGE acrylamide gels using a Tris-acetate buffering system [328] then transferred to nitrocellulose membranes (Millipore, Protran BA85). After transfer, membranes were blocked in 0.05% milk powder in PBS containing 0.01% Tween-20 then incubated overnight in 1:1000 primary antibody. After washing in PBS+0.01% Tween-20, the membranes were incubated with secondary HRP antibody (Jackson Labs, 1:50,000 dilution), developed using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific, Cat. # 37071) and visualized using a cooled CCD instrument (Dyversity, Syngene).

2.6.8. Chromatin Immunoprecipitation Assay

Chromatin immunoprecipitation (ChIP) was performed as previously described [329]. The cells or pituitary tissue were fixed using 1% formaldehyde in HEPES (pH 7.8) for 10 min at room temperature. Cells were collected, washed with PBS, re-suspended in lysis buffer (50 mM Tris-HCl, pH 8.1, 1% SDS and 10mM EDTA) and sonicated using a Branson Sonifier 450 with an output of 3.5 and constant duty cycle in pulses until the cross-linked DNA fragments were 300-500 bp in size. Five percent of the cross-linked chromatin was used as input and the rest was incubated with 5 µg of poly deoxyinosinic-deoxycytidylic (poly dI-dC) and either 40 µL of 50% slurry of anti FLAG affinity gel (Sigma) or a primary antibody as indicated overnight, at 4°C. For the primary antibody samples, Protein A agarose beads were added for an additional 20 minutes prior to washing. The beads were successively washed with RIPA (10mM Tris-HCl pH 8.0, 1mM EDTA, 0.5 mM EGTA, 140mM NaCl, 1% Triton-X 100, 0.1% Na-deoxycholate, 0.1% SDS and 1X protease inhibitor cocktail (Bioshop, Cat. # PIC003)), TSEII (20mM Tris-HCl pH 8.1, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100), TSE III (10mM Tris-HCl pH 8.1, 0.25 M LiCl, 1mM EDTA, 1%NP-40, 1% sodium deoxycholate) buffer and
then 3 washes with 0.1X TE. The DNA crosslinks were reversed overnight at 65°C in 0.1 M NaHCO₃. The DNA was then precipitated with 2 µL of pellet paint (Novagen), 1/10 volume 3M Na-acetate and 2 volumes of 100% EtOH followed by centrifugation for 10 min at 14,000 rpm. Pellets were washed with 70% EtOH, dried and re-suspended in 50 µL ddH₂O. For each PCR reaction 5 µL of the immunoprecipitated sample or five percent input samples were used to quantify enrichment.

### 2.6.9. Preparation of nuclear extract and electrophoretic mobility shift assay (EMSA)

Nuclear proteins from MMQ, GHFT and GC cells were prepared as previously described [330]. Briefly, the cells were washed twice with ice cold PBS followed by incubation on ice for 10 min in excess of buffer A (10 mM HEPES-KOH pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT, and protease inhibitor cocktail). Cells were lysed by vortexing and the nuclei were collected by centrifugation at 4400 rpm for 15 min. The pellets were re-suspended with 1 pellet volume of Buffer C (20 mM HEPES-KOH pH 7.9, 1.5 mM MgCl₂, 420 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT, 25% glycerol, and 1X protease inhibitor cocktail) and incubated on ice for 20 min and then centrifuged at >15000 rpm. The supernatants containing the nuclear proteins were collected and stored at -80 °C.

For radiolabelling oligonucleotides with α³²P-dCTP, 7 pmol of double stranded oligonucleotides were filled-in with T4 DNA polymerase (Klenow). The labeled probes were separated from the free nucleotides using Illustra Microspin G-50 columns (GE healthcare). The oligonucleotide sequences used in the EMSAs and competition assays are provided in the supplementary Table 2-1.

For the electrophoretic mobility shift assay (EMSA), the binding reactions were prepared using 12 µg of nuclear extracts in binding buffer (10 mM TRIS–HCl pH7.0, 1mM DTT, 5 mM MgCl₂, 50 ng/mL poly-dIdC, 2.5% glycerol, 0.05% Igepal, 0.05 M KCl and 50,000 cpm of labeled oligonucleotide). The reactions were incubated at room temperature for 20 min followed by separation using 4% non-denaturing polyacrylamide gels. For competition assays, cold/ unlabeled competitors were added at 100 fold molar excess.
Gels were dried and exposed to a phosphor storage cassette in the dark for 12hrs followed by scanning using Molecular Dynamics Storm 860 Phosphor imager.

### 2.6.10. Large-scale protein purification

Nuclear extracts were prepared from $4 \times 10^{10}$ MMQ cells and applied to a P11 column, pre-washed with PBS with no salt and then proteins were eluted in a step-wise fashion with PBS containing (0.1M NaCl, 0.3 M NaCl, 0.5 M NaCl and finally 0.85 M NaCl) and collected in 1 mL fractions. Methyl-DNA binding proteins were monitored using a standard EMSA with a methylated DMR oligonucleotide probe. Next, the methyl DNA binding activity concentrated in fraction #24 was applied to a DNA affinity column containing a mulitmerized unmethylated version of the DMR. The flow-through was added sequentially to a second unmethylated column and then finally to a third version that was DNA methylated. The bound proteins were washed with PBS, 0.01% NP-40 and eluted in SDS-sample buffer prior to gel electrophoresis. To generate the affinity matrix, an oligonucleotide with 2 copies of the DMR (RH75) was amplified in a PCR reaction with a second oligonucleotide containing a biotin group at the 5’ end (GP816) (supplementary Table 2-6). The PCR product was qualified on an acrylamide gel and the DNA methylated oligonucleotides prepared by treatment with SssI methyltransferase (NEB). The methylated version of the column was confirmed by digestion with the methylation sensitive Hgal restriction enzyme (supplementary Figure 2-12).

### 2.6.11. Short hairpin knock-down of SMCHD1 in cells

Double-stranded nucleotide sequences encoding shRNAs targeting SMCHD1 and a control NC5 sequence were cloned into the pQCXIP-GFP vector (Table S7) [331]. For production of shRNAs expressing virus particles, HEK293T cells were co-transfected with the individual shRNA expression vectors and the packaging plasmid (pCL10A1) using Polyfect (Qiagen) according to the manufacturer’s protocol. The media from transfected cells were collected after 48 and 72 hours and combined with 10 µL/mL Polybrene (2mg/mL in 1M HEPES pH 7.2). Then the virus stock was passed through a 0.2 µm syringe filter and stored at 4 ºC.
HEK293 or SH-SY5Y cells were infected two times in a sequential fashion. Briefly, 2mL of virus stock was added to cells seeded on 6 well plates and the plates were centrifuged for 1.5hrs at 2500 rpm and 30 °C and then the media was replaced with fresh DMEM complete media. The procedure was repeated the following day. Two days after the start of the infection process, cells were selected with 3 µg/mL of puromycin for 7 days. For the SH-SY5Y cells, the selection was initiated 1 day post-infection and harvested after 11 days post-infection. The infection was monitored using GFP expression.

2.6.12. mRNA Microarray analysis

The mRNA from stable cell lines was isolated using TRI reagent (Sigma, Cat. # T9424) according to the manufacturer’s protocol. The mRNA microarray was performed by the laboratory for Advanced Genome Analysis at the Vancouver Prostate Centre, Vancouver, Canada. Total RNA was qualified with the Agilent 2100 Bioanalyzer (RNA) and quantified with the NanoDrop ND-1000 UV-VIS spectrophotometer to measure A260/280 and A260/230 ratios. The RNA was converted to cDNA using T7 RNA polymerase in the presence of cyanine 3 (Cy3)-labeled CTP. Samples were prepared in biological triplicates following Agilent’s One-Color Microarray-Based Gene Expression Analysis Low Input Quick Amp Labeling v6.0. An input of 100ng of total RNA was used to generate Cyanine-3 labeled cRNA. Samples were hybridized on Agilent SurePrint G3 Human GE 8x60K Microarray (Design ID 028004).

Arrays were scanned with the Agilent DNA Microarray Scanner at a 3µm scan resolution and data was processed with Agilent Feature Extraction 10.10. Green processed signal was quintile normalized with Agilent GeneSpring 11.5.1. To find significantly regulated genes, fold changes between the SMCHD1 shRNAs and the NC5 shRNA control groups and P-values gained from t-test between the same groups were calculated with a Benjamini-Hochberg multiple testing correction. The t-tests were performed on log transformed normalized data and the variances were not assumed to be equal between sample groups. Up and down-regulated genes were selected if the P-value was < 0.05.
and fold difference greater or equal to 1.8 compared to the control. The raw data was submitted to the GEO repository, GSE52065.

Heat maps were created using the Hierarchical clustering program from the GenePattern website (http://genepattern.broadinstitute.org). To map the genes to chromosomal locations, we used the biomart program located at http://uswest.ensembl.org. The Ensemble Genes 70 and Homo sapiens genes (GRCh37.p10) were chosen as databases for analysis. Selected genes from the microarray analysis were mapped on chromosomes by filtering using the Agilent Sureprint G3 GE 8x60k probe’s IDs.
2.7. Supplementary information

![Pyrosequencing Graph](image)

**Primers used for pyrosequencing**

<table>
<thead>
<tr>
<th>Name of the primer</th>
<th>Species</th>
<th>Gene</th>
<th>Assay</th>
<th>Forward Sequence (5') to (3')</th>
<th>Reverse Sequence (5') to (3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp154</td>
<td>Mouse</td>
<td>GH</td>
<td>Nested primer for Pyrosequencing</td>
<td>TTGGATTTAAAAAT GGTGTTAGAGGATAT</td>
<td></td>
</tr>
<tr>
<td>gp155</td>
<td>Mouse</td>
<td>GH</td>
<td>Nested primer for Pyrosequencing</td>
<td></td>
<td>CCTCAAATCCATCTCTCA AAAACCCTAA</td>
</tr>
<tr>
<td>AK104</td>
<td>Mouse</td>
<td>GH</td>
<td>Nested primer for Pyrosequencing</td>
<td>GTGGGTTTTTTAG TGGGATTTTG</td>
<td></td>
</tr>
<tr>
<td>AK108</td>
<td>Mouse</td>
<td>GH</td>
<td>Pyrosequencing</td>
<td>TTACTATAACCATTAC CAAAACAATACT</td>
<td></td>
</tr>
<tr>
<td>AK105</td>
<td>Mouse</td>
<td>GH</td>
<td>Pyrosequencing</td>
<td>5Biosq/ACC CTT TTT ATA CCC TAC CTA TTT TC</td>
<td></td>
</tr>
<tr>
<td>AK106</td>
<td>Mouse</td>
<td>GH</td>
<td>Sequencing primer</td>
<td>ATTAATGTGTATA AGTTGGTAAAGA</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2-8. Comparison of the pyrosequencing results from WT-GH:RFP and ΔLCR-GH:RFP mouse pituitary.**

Percent methylation of the GH promoter was graphed for CpG positions -9 to -3. The PCR primers were designed using PuroMark assay Design software 2.0 (Qiagen). Three PCR reactions were performed: a primary with gp154 and gp155 and a final with AK104 and AL105. Ten \(\mu\)L of each sample was added to 70 \(\mu\)L of master mix containing GH Healthcare Strepavidin Sepharose High Performance beads (Cat. # 17-5113-01). Samples were loaded onto a PyroMark Q24 pyrosequencer and analyzed based on a sequencing primer (AK106). Note that the mouse genome has one less CpG in that region compared to the rat genome. Results comparing the intensity of signal for C versus T provides the percent methylation at cytosines.
Figure 2-9. SMCHD1 is bound to the GH promoter in cells in a DNA methylation dependent fashion.

Upper, schematic illustration of the GH promoter and relative position of the primers used for the ChIP assays. MMQ cells were treated with 5azaC or cultured under normal conditions and then processed in a conventional ChIP assays with anti-SMCHD1, anti-Pit1 or anti-E cadherin antibodies. Real time PCR results are illustrated in Figure 4D. Lower, stable HEK293 cells expressing strep/Flag-tagged SMCHD1 optionally with treatment of 5-AzaC were assayed in a conventional ChIP for enrichment of the GH promoter. Below, stable cell whole cell extracts were analyzed with anti-Flag or anti-SMCHD1 antibodies, as indicated, for determination of protein expression levels.
Figure 2-10. SMCHD1 knockdown in HEK293 cells resulted in modified expression of genes when compared to control cells.
The heat map and hierarchal clustering compare the transcriptional level of up-regulated and down-regulated genes in SMCHD1 shRNA KD cells (KD3 and 4) and control cells, CTL. The scale below represents the fold differences of the genes. Intense blue represents down-regulation and intense red indicates up-regulation.
Figure 2-11. PCDHB10 up-regulation during SMCHD1 knockdown does not correlate with changes in promoter DNA methylation.

The left panel represents PCDHB10 mRNA quantification in shRNA1 and shRNA NC5 sh-SY5Y cells. *represents P-value <0.001 using a two-way anova with Tukey’s post hoc analysis. The right panel represents the methylation level of CpGs -2 to -7 of PCDHB10 promoter in shRNA1 and shRNA NC5 sh-SY5Y cells measured by bisulfite pyrosequencing. Below, the table contains the oligonucleotides used for pyrosequencing. F and R-BIO were used for PCR and capture while S1 for the sequencing reaction.
Figure 2-12. The DNA used for affinity purification was resistant to digestion with Hgal, a DNA methylation sensitive enzyme.

Above, illustration of the process for generating methylated DNA. Biotin labeled DNA was amplified and multimerized using PCR. Then the DNA was methylated with SssI methyltransferase in the presence of S-adenosyl methionine (SAM). Below, to verify the DNA was methylated, its integrity was monitored by agarose gel electrophoresis following digestion with Hgal. The integrity of the methylated DNA (lane 4) was compared to unmethylated DNA (lane 2) or DNA not exposed to Hgal (lane 1 and 3).
### Table 2-1. Oligonucleotides used for the EMSAs

<table>
<thead>
<tr>
<th>Name</th>
<th>Oligo</th>
<th>Competed bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeProbe</td>
<td>gtaacgMeCggMeCgtgagaAGGTAAgataAGGGAMECgtgaeMeCgcaactag</td>
<td>Upper and lower</td>
</tr>
<tr>
<td>Me(-6,-5)Probe</td>
<td>gtaacgacggctggagaAGGTAAgataAGGGAMECgtaaMeCgcaactag</td>
<td>Lower only</td>
</tr>
<tr>
<td>unMeProbe</td>
<td>gtaacgacgctggagaAGGTAAgataAGGGACgtaaMeCgcaactag</td>
<td>Lower only</td>
</tr>
<tr>
<td>DR4, TRE</td>
<td>gacgttcAGGTCAcaggAGGGTCAGagagctc</td>
<td>Lower only</td>
</tr>
<tr>
<td>Mbox5, zeb1</td>
<td>ggcacggaggagctccagaggagctc</td>
<td>Middle only</td>
</tr>
<tr>
<td>Consensus, zeb1</td>
<td>ggtgtgcaaaagtgccaggtcc</td>
<td>None</td>
</tr>
<tr>
<td>GC box, Sp1 like</td>
<td>ggtgtgcggccccggcgccgccgc</td>
<td>Middle only</td>
</tr>
<tr>
<td>MeCpG, MeCP2 competitor</td>
<td>attacgttcctgMeCgtttttacttctgtctcMeCgcacccagacctggcagMeCgtgcctgtgcacactctcagctcag</td>
<td>None</td>
</tr>
</tbody>
</table>

### Table 2-2. Genes identified with matched peptides from the DNA affinity purification assay using the methylated GH DMR and LC-MS.

#### Rat IPI v3.4.7

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Gene</th>
<th>Mass</th>
<th>Score</th>
<th># of distinct peptides</th>
<th>emPAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP100212769</td>
<td>Anionic trypsin-2</td>
<td>26895</td>
<td>74</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>IP100212767</td>
<td>Anionic trypsin-1</td>
<td>26627</td>
<td>61</td>
<td>1</td>
<td>0.31</td>
</tr>
<tr>
<td>IP100362154</td>
<td>RGD13072354 Similar to SMC hinge domain containing 1</td>
<td>233507</td>
<td>44</td>
<td>2</td>
<td>0.03</td>
</tr>
</tbody>
</table>

#### Unipro-Swissprot

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Gene</th>
<th>Mass</th>
<th>Score</th>
<th># of distinct peptides</th>
<th>emPAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>P00761</td>
<td>Tryp</td>
<td>Pig</td>
<td>Trypsin</td>
<td>25078</td>
<td>280</td>
</tr>
<tr>
<td>P00763</td>
<td>Tryp2</td>
<td>Rat</td>
<td>Anionic trypsin-2</td>
<td>26895</td>
<td>75</td>
</tr>
<tr>
<td>P84733</td>
<td>PS17</td>
<td>Pinst</td>
<td>Putative cytochrome C oxidase subunit II</td>
<td>1707</td>
<td>71</td>
</tr>
<tr>
<td>A6NCN2</td>
<td>K781L</td>
<td>Human</td>
<td>Keratin-81-like protein</td>
<td>54972</td>
<td>63</td>
</tr>
<tr>
<td>P00762</td>
<td>Tryp1</td>
<td>Rat</td>
<td>Anionic trypsin-1</td>
<td>26627</td>
<td>60</td>
</tr>
<tr>
<td>O4FZU2</td>
<td>K268A</td>
<td>Rat</td>
<td>Keratin, type II cytoskeletal 6A</td>
<td>55555</td>
<td>59</td>
</tr>
<tr>
<td>Q7Z74</td>
<td>K2C18</td>
<td>Human</td>
<td>Keratin, type II cytoskeletal 1b</td>
<td>61935</td>
<td>59</td>
</tr>
<tr>
<td>Q0P5j</td>
<td>KRT35</td>
<td>Bovin</td>
<td>Keratin, type I cuticular Ha5</td>
<td>51217</td>
<td>52</td>
</tr>
</tbody>
</table>
Table 2-3. Genes with monoallelic expression differentially regulated by SMCHD1

<table>
<thead>
<tr>
<th>Location</th>
<th>Gene</th>
<th>Regulation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>2p16.3</td>
<td>NRXN1</td>
<td>up</td>
<td>[1]</td>
</tr>
<tr>
<td>2q36.3</td>
<td>COL4A3</td>
<td>down</td>
<td>[1]</td>
</tr>
<tr>
<td>3q22.2</td>
<td>EFNB1</td>
<td>down</td>
<td>[1]</td>
</tr>
<tr>
<td>5q31.3</td>
<td>PCDHB2</td>
<td>up</td>
<td>[2]</td>
</tr>
<tr>
<td>5q31.3</td>
<td>PCDHB8</td>
<td>up</td>
<td>[2]</td>
</tr>
<tr>
<td>5q31.3</td>
<td>PCDHB16</td>
<td>up</td>
<td>[2]</td>
</tr>
<tr>
<td>5q31.3</td>
<td>PCDHB10</td>
<td>up</td>
<td>[2]</td>
</tr>
<tr>
<td>5q31.3</td>
<td>PCDHB11</td>
<td>up</td>
<td>[2]</td>
</tr>
<tr>
<td>5q31.3</td>
<td>PCDHB14</td>
<td>up</td>
<td>[2]</td>
</tr>
<tr>
<td>5q34</td>
<td>SLIT3</td>
<td>down</td>
<td>[1]</td>
</tr>
<tr>
<td>7q21.11</td>
<td>MAGI2</td>
<td>down</td>
<td>[1]</td>
</tr>
<tr>
<td>7q31.33</td>
<td>GRM8</td>
<td>down</td>
<td>[1]</td>
</tr>
<tr>
<td>10q22.1</td>
<td>COL13A1</td>
<td>up</td>
<td>[1]</td>
</tr>
<tr>
<td>10q23.1</td>
<td>NRG3</td>
<td>down</td>
<td>[1]</td>
</tr>
<tr>
<td>11p13</td>
<td>CD44</td>
<td>up</td>
<td>[3]</td>
</tr>
<tr>
<td>11p14.3</td>
<td>GAS2</td>
<td>up</td>
<td>[1]</td>
</tr>
<tr>
<td>11p15.4</td>
<td>AMPD3</td>
<td>down</td>
<td>[3]</td>
</tr>
<tr>
<td>11p15.5</td>
<td>KCNC1</td>
<td>down</td>
<td>[3]</td>
</tr>
<tr>
<td>11p15.5</td>
<td>H19</td>
<td>down</td>
<td>[3]</td>
</tr>
<tr>
<td>11p15.5</td>
<td>CDKN1C</td>
<td>down</td>
<td>[3]</td>
</tr>
<tr>
<td>12q24.31</td>
<td>CAPB1</td>
<td>down</td>
<td>[1]</td>
</tr>
<tr>
<td>13q14.11</td>
<td>DNAJC15</td>
<td>down</td>
<td>[1]</td>
</tr>
<tr>
<td>14q32.2</td>
<td>BEGAIN</td>
<td>down</td>
<td>[3]</td>
</tr>
<tr>
<td>15q24.2</td>
<td>CSPG4</td>
<td>down</td>
<td>[4]</td>
</tr>
<tr>
<td>18p11.31</td>
<td>DLGAP1</td>
<td>up</td>
<td>[1]</td>
</tr>
<tr>
<td>20q13.2</td>
<td>SGK2</td>
<td>down</td>
<td>[3]</td>
</tr>
</tbody>
</table>

Supplemental references
<table>
<thead>
<tr>
<th>Name of the primer</th>
<th>Species</th>
<th>Gene</th>
<th>Assay</th>
<th>Forward Sequence (5’ to 3’)</th>
<th>Reverse Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp819</td>
<td>Rat</td>
<td>GH</td>
<td>ChIP</td>
<td>TCCCAGTGGACAAACGATGG</td>
<td></td>
</tr>
<tr>
<td>gp820</td>
<td>Rat</td>
<td>GH</td>
<td>ChIP</td>
<td>GAGGCCACTGACAGCTTGTG</td>
<td></td>
</tr>
<tr>
<td>SM84</td>
<td>Human</td>
<td>GH</td>
<td>ChIP</td>
<td>GGA GTC CTC AGC AGA AGT GGA ATTC</td>
<td></td>
</tr>
<tr>
<td>SM85</td>
<td>Human</td>
<td>GH</td>
<td>ChIP</td>
<td>CTTGTCCTCTTTAAGGTCACGTG</td>
<td></td>
</tr>
<tr>
<td>RH6</td>
<td>Mouse</td>
<td>GH</td>
<td>ChIP COBRA</td>
<td>CTCTAATATAAAAACATCCGACTTC</td>
<td></td>
</tr>
<tr>
<td>RH7</td>
<td>Mouse</td>
<td>GH</td>
<td>ChIP COBRA</td>
<td>GAATTTAGGATTTTGGAAGTAGTC</td>
<td></td>
</tr>
<tr>
<td>RH8</td>
<td>Mouse</td>
<td>GH</td>
<td>ChIP COBRA</td>
<td>TCCTATCTCCTATCCTCTTTC</td>
<td></td>
</tr>
<tr>
<td>RH9</td>
<td>Mouse</td>
<td>GH</td>
<td>ChIP COBRA</td>
<td>AGTTTTTTGTTATTATTGTATGGA</td>
<td></td>
</tr>
</tbody>
</table>
Table 2-5. Primers used for RT-PCR

<table>
<thead>
<tr>
<th>Name of the primer</th>
<th>Species</th>
<th>Gene</th>
<th>Assay</th>
<th>Forward Sequence (5' to 3')</th>
<th>Reverse Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM210</td>
<td>Human</td>
<td>PCDHB3</td>
<td>RT-PCR</td>
<td>CCCAGGCCTGAAAACCCAAGTA</td>
<td></td>
</tr>
<tr>
<td>SM211</td>
<td>Human</td>
<td>PCDHB3</td>
<td>RT-PCR</td>
<td></td>
<td>GGCAGCAGCGAGTGTTAC</td>
</tr>
<tr>
<td>SM214</td>
<td>Human</td>
<td>PCDHB8</td>
<td>RT-PCR</td>
<td>GCAAGCTCCTTGTCAACAA</td>
<td></td>
</tr>
<tr>
<td>SM215</td>
<td>Human</td>
<td>PCDHB8</td>
<td>RT-PCR</td>
<td></td>
<td>TCCCTACGGGTTCCATTTC</td>
</tr>
<tr>
<td>SM222</td>
<td>Human</td>
<td>PCDHB11</td>
<td>RT-PCR</td>
<td>CCCATCTTCTCGAAACAA</td>
<td></td>
</tr>
<tr>
<td>SM223</td>
<td>Human</td>
<td>PCDHB11</td>
<td>RT-PCR</td>
<td></td>
<td>AGCGCCTTGTCGAAACTAA</td>
</tr>
<tr>
<td>SM226</td>
<td>Human</td>
<td>PCDHB14</td>
<td>RT-PCR</td>
<td>AAACGACAACCACCCAGAAG</td>
<td></td>
</tr>
<tr>
<td>SM227</td>
<td>Human</td>
<td>PCDHB14</td>
<td>RT-PCR</td>
<td></td>
<td>TGACTCGGCTTGCTCTC</td>
</tr>
<tr>
<td>SM412</td>
<td>Human</td>
<td>KCNQ1ot1</td>
<td>RT-PCR</td>
<td>GTG AAG CCA GCT GTT</td>
<td></td>
</tr>
<tr>
<td>SM413</td>
<td>Human</td>
<td>KCNQ1ot1</td>
<td>RT-PCR</td>
<td>GGA GGG</td>
<td>ACA GGC CTG GTA AGG GAC</td>
</tr>
<tr>
<td>SM416</td>
<td>Human</td>
<td>CDKN1C</td>
<td>RT-PCR</td>
<td>CCTCAAGAGAGGCGGAGC</td>
<td>AACA GCAG CGA CGA CTC</td>
</tr>
<tr>
<td>SM417</td>
<td>Human</td>
<td>CDKN1C</td>
<td>RT-PCR</td>
<td></td>
<td>GAC ATC GCC CGA CGA CTT CTC</td>
</tr>
<tr>
<td>SM422</td>
<td>Human</td>
<td>KNCQ1</td>
<td>RT-PCR</td>
<td>GGC ATC CGC TTC CTG</td>
<td>CAG AT</td>
</tr>
<tr>
<td>SM423</td>
<td>Human</td>
<td>KNCQ1</td>
<td>RT-PCR</td>
<td></td>
<td>CCC CAC CAC AGC GCA TC</td>
</tr>
<tr>
<td>SM424</td>
<td>Human</td>
<td>H19</td>
<td>RT-PCR</td>
<td>GAT CCG TGC CTC AGC</td>
<td>GTT C</td>
</tr>
<tr>
<td>SM425</td>
<td>Human</td>
<td>H19</td>
<td>RT-PCR</td>
<td></td>
<td>GTT GTG GGT TCT GGG AGC</td>
</tr>
<tr>
<td>SM65</td>
<td>Human</td>
<td>Beta actin</td>
<td>RT-PCR</td>
<td>GCCACACACAGTGCTGCT</td>
<td></td>
</tr>
<tr>
<td>SM66</td>
<td>Human</td>
<td>Beta actin</td>
<td>RT-PCR</td>
<td></td>
<td>AGGAGCAATGATCTTGATCTT</td>
</tr>
<tr>
<td>VN3</td>
<td>Rat</td>
<td>Pit-1</td>
<td>RT-PCR</td>
<td>AAAACAAACCATCTGCCGAT</td>
<td></td>
</tr>
<tr>
<td>VN4</td>
<td>Rat</td>
<td>Pit-1</td>
<td>RT-PCR</td>
<td></td>
<td>TCGTTTCTCCTCCTTTTG</td>
</tr>
<tr>
<td>VN5</td>
<td>Rat</td>
<td>Prolactin</td>
<td>RT-PCR</td>
<td>TGGCCCTGGCAAATAACTC</td>
<td></td>
</tr>
<tr>
<td>VN6</td>
<td>Rat</td>
<td>Prolactin</td>
<td>RT-PCR</td>
<td></td>
<td>AGGGACACTCCTCCTGCTGAT</td>
</tr>
<tr>
<td>VN11</td>
<td>Rat</td>
<td>GH</td>
<td>RT-PCR</td>
<td>TGGCAACACAGACTGGAACAG</td>
<td></td>
</tr>
<tr>
<td>VN12</td>
<td>Rat</td>
<td>GH</td>
<td>RT-PCR</td>
<td></td>
<td>CGAGCCCAGATTCCAAACTA</td>
</tr>
</tbody>
</table>
### Table 2-6. Primers used for construction of the DNA affinity matrix

<table>
<thead>
<tr>
<th>Name of the primer</th>
<th>Species</th>
<th>Gene</th>
<th>Assay</th>
<th>Forward Sequence (5' to 3')</th>
<th>Reverse Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH175</td>
<td>Rat</td>
<td>GH DMR</td>
<td>DNA affinity</td>
<td>GAC GCT GCG GTC ACG TCC CTG ATC TTA CCT TTC CAC CGC CGC TTT TGA CGC TGC GGT CAC GTC CCT GAT CTT ACC TTT</td>
<td></td>
</tr>
<tr>
<td>gp816</td>
<td>Rat</td>
<td>GH DMR</td>
<td>DNA affinity</td>
<td>5' / 3'Biosp/nSpC3/sSpC3/GCG TCA AAG GCG GCG GTG GAA AGG TAA GAT C</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2-7. Primers used for construction of shRNA KD

<table>
<thead>
<tr>
<th>Name of the primer</th>
<th>Species</th>
<th>Gene</th>
<th>Assay</th>
<th>Forward Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM70</td>
<td>Human</td>
<td>SmoHD1/KD1</td>
<td>shRNA KD</td>
<td>TTTGGGAATAAGAGACAGAAGAGATCGTTCAAGAGACTGCTGATGCTATGGTATTTTCCTTTTTTTTT</td>
</tr>
<tr>
<td>SM72</td>
<td>Human</td>
<td>SmoHD1/KD3</td>
<td>shRNA KD</td>
<td>TTTGACTAGACAGTGTGAGGTATATGGAATATCATTCAAGAGAAGAAGATCTTTTCTACAGTCTAGTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>SM74</td>
<td>Human</td>
<td>SmoHD1/KD4</td>
<td>shRNA KD</td>
<td>TTTGCCAATTGGAATGCTACAATGGAAATTTCAAGAGAATATCCATTGTGATGCAATTCAATGTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>SM76</td>
<td>Human</td>
<td>ScrambledNC5</td>
<td>shRNA KD</td>
<td>TTTGCTATATTGCAGCGTATAGACAGGCTTTGTTCAAGAGACTAACGACTATACGGCAAATATGTCTTTTT</td>
</tr>
</tbody>
</table>
Chapter 3.

SMCHD1 is a regulator of forskolin-induced cellular cAMP levels

A modified version of this chapter is submitted to Molecular and Cellular Biology journal.

Contributions: Shabnam Massah contributed to all Figures. John Jubene contributed to antiserum development. I was the main author of this paper. I was responsible for conceiving, designing the experiments and figures and I performed 90% of the experiments. In the previous chapter, I showed that SMCHD1 binds to GH promoter in a DNA methylation sensitive manner and plays a role in regulating expression of X-linked genes as well as autosomal genes. In this manuscript, the question I was asking was how global DNA methylation affects binding of SMCHD1 and therefore, gene regulation.

I treated SH-SY5Y neuroblastoma cells with 5-azacytidine to degrade DNA methyltransferase 1 (DNMT1) and induce loss of global DNA methylation. Using a ChIP grade antiserum in chromatin immunoprecipitation (ChIP)-sequencing, I gathered information on genomic sites bound by SMCHD1 (structural maintenance of chromosome hinge domain containing-1) with high resolution. I showed that 5azaC treatment greatly reduced SMCHD1 binding. SMCHD1 occupancy coincides with binding sites for transcription factors from different families including Beta Beta Alpha-zinc fingers and Helix-loop-Helix transcription factors.

In addition, I also identified genes that were associated with SMCHD1 binding sites and their biological functions. I showed that SMCHD1 genome occupancy associates with genes involved in the G-protein coupled signaling pathway in a DNA methylation
sensitive manner. Loss of SMCHD1 changes expression of adenylate cyclases 7 and 9 (ADCY7 and ADCY9). More importantly, I demonstrated that loss of SMCHD1 influences the maximum available cellular levels of cAMP. I showed that loss of SMCHD1 induces expression level of KCNQ1, a gene normally shown to interact with ADCY9 to modulate cAMP production.

3.1. Abstract

Structural maintenance of chromosome hinge domain containing 1 (SMCHD1) plays an important role in regulating the expression of some imprinted genes [69-71]. In addition, SMCHD1 has an essential role for late stage CpG island methylation of the inactive X-chromosome [68, 172, 332]. A recent study suggests that SMCHD1 mutations in FSHD2 individuals results in DNA hypo-methylation of number of autosomal loci [333]. Therefore, DNA methylation may play a role in SMCHD1 genome wide occupancy, however understanding how global DNA methylation affects SMCHD1 binding and regulation of gene expression remains unclear. Here, we show genome wide occupancy of SMCHD1 in human neuroblastoma SH-SY5Y cells and identified binding sites sensitive to DNA demethylating reagent, 5-azacytidine (5azaC). SMCHD1 genomic binding associates with genes involved in G-protein couple receptor signaling. Moreover, this study reveals that cells lacking SMCHD1 have higher cellular cyclic adenosine monophosphate levels suggesting an altered second messenger signaling system. We propose loss of an epigenetic modifier may alter cell signaling and that up-regulation of two genes ADCY9 and KCNQ1 may explain increased cAMP signaling capabilities of these cells. Our studies suggest a role for SMCHD1 in the regulation of adenylate cyclase activity and second messenger signaling.
3.2. Introduction

Structural maintenance of chromosome hinge domain containing 1 (SMCHD1) is a chromatin modifier that regulates gene expression on the X-chromosome as well as autosomal genes [69-71]. Originally, SMCHD1 was identified in an N-ethyl-N nitrosourea mutagenesis screen as an epigenetic modifier and it was suggested to be essential for X-inactivation and survival in females [68]. Later studies confirmed the initial observation and showed that SMCHD1 is essential for methylation of subset of CpG islands at late stage of X-inactivation [172]. Loss of SMCHD1 is also lethal in male mice in a mixed background, suggesting an essential role for gene regulation on non-sex chromosomes [68, 332]. Indeed, we and others have shown that SMCHD1 is important for regulating monoallelically expressed genes including imprinted genes, and clustered protocadherin genes [69-71]. In human, SMCHD1 mutations associate with two distinct developmental diseases: facioscapulohumoral muscular dystrophy (FSHD) [73, 316] and bosma arhinia and microphthalmia (BAMS) [72, 334]. FSHD is a muscular dystrophy affecting upper arm, shoulder and face muscles and is characterized by chromatin relaxation of the D4Z4 microsatellite array on chromosome 4 [335]. SMCHD1 mutation in FSHD2 individuals results in DNA hypo-methylations of number of autosomal loci [333]. The most consistent signature of BAMS individuals is the complete absence of nose which might accompany with other malformations [336]. These findings suggest that an epigenetic modifier SMCHD1 impacts multiple genomic regions with different context, though its molecular function in regulating gene expression is unknown.

SMCHD1 is a non-canonical member of SMC protein family [68]. SMCHD1 contains a hinge domain homologous to members of SMC family. Unlike SMC proteins, SMCHD1 N-terminus encompasses a GHKL type ATPase domain [66]. SMCHD1 hinge domain was suggested to have DNA binding activity which is the interphase of SMCHD1 homodimerization [67]. Structural studies show that SMCHD1 is likely important for heterochromatin formation over the X chromosome by facilitating a link between two chromatin domains enriched for repressive histone marks (H3K9me3 and H3K27me3) [173]. Previously, we isolated SMCHD1 from cellular nuclear extracts using methylated DNA in an affinity purification column [70]. These findings suggest that DNA methylation
is an important determinant in recruitment of SMCHD1 to genomic loci and possibly its function in regulating gene expression. We hypothesize that SMCHD1 genomic binding is linked to DNA methylation level and treatment of cells with DNA demethylating reagent, 5azacytidine would influence SMCHD1 genomic occupancy.

The goal of this study was to better understand SMCHD1 function at the molecular level and investigate biological processes associated with SMCHD1 genome occupancy. We generated a ChIP grade antiserum and used it to gather information on genomic sites bound by SMCHD1 with high resolution and identify sites in which SMCHD1 recruitment was sensitive to 5azaC treatment. Here we show SMCHD1 occupancy is located mostly over the intron and intergenic regions. SMCHD1 occupancy coincides with binding sites for transcription factors including Beta Beta Alpha-zinc fingers, Helix-loop-Helix Families, hormone receptors and homeodomain proteins. In addition, our study reveals SMCHD1 contribution to second messenger signaling in cells through modulation of cAMP signaling. We showed that SMCHD1 binding sites associate with G protein-coupled receptor signaling that is sensitive to 5azaC treatment. Moreover, loss of SMCHD1 alters expression of some genes encoding adenylate cyclase (ADCY7 and ADCY9) and more importantly can influence the maximum available cellular levels of cAMP. Further, we demonstrated that loss of SMCHD1 induces expression level of KCNQ1 gene and current literature suggests it association with cAMP production through its interaction with ADCY9. This study proposed SMCHD1 functions in regulation of cAMP signaling in cells and suggests its contribution to biological processes that are greatly affected by second messenger signaling [337-339].

3.3. Results

3.3.1. Identification of genome-wide occupancy of SMCHD1 that is sensitive to 5azaC

Some observations showed SMCHD1 DNA binding and DNA methylation were intimately linked. Therefore, we sought to compare SMCHD1 genomic localization in SH-SY5Y cells cultured under normal conditions and those treated with 5azaC to induce
global loss of DNA methylation. 5azaC is an analog of cytidine and incorporates into DNA. The maintenance DNA methyltransferase, DNMT1 recognizes 5azaC. This results in a covalent bond between the enzyme and the 5azaC that could not be resolved, therefore leaving DNMT1 bound to DNA and blocking its activity. This triggers DNA damage signaling and proteasomal degradation of DNMT1 that leads to global loss of DNA methylation [340]. To examine the efficacy of 5azaC treatment, we measured DNMT1 protein levels. DNMT1 was almost completely lost in cells treated with 5azaC (Supplemental Figure 3-7), confirming that 5azaC treatment was effective in degrading DNMT1. We also measured SMCHD1 protein levels upon 5azaC treatment confirming that 5azaC treatment does not change SMCHD1 protein level (Supplemental Figure 3-8). We then used twenty replicates of 5azaC treated and control SH-SY5Y cells in chromatin immunoprecipitation (ChIP) assay using anti-SMCHD1 antiserum. Our guinea pig anti-SMCHD1 antiserum is generated specific to hydrophobic antigens to increase the efficiency of immunoprecipitation. The anti-SMCHD1 antiserum efficiently immunoprecipitated SMCHD1 in a comparable manner to the commercially available antibody (Supplemental Figure 3-9). Following chromatin immunoprecipitation, the replicate samples were pooled and used for generating DNA libraries and sequencing. Following ChIP-Seq, model-based analysis for ChIP-Seq peak calling with paired ends (MACS2) identified 5051 SMCHD1 binding sites across the genome in cells grown under normal conditions compared to 2100 binding site in cells treated with 5azaC, suggesting SMCHD1 genome binding declines with decreased DNA methylation. Figure 3-1A is a schematic illustrating the work-flow and analysis of the ChIP-Seq data. Figure 1B illustrates SMCHD1 binding peaks for the DUX4 repeats and shows that SMCHD1 occupies regions upstream and downstream of DUX4 close to FSHD region gene 1 (FRG1). SMCHD1 occupancy over this region is greatly affected by 5azaC treatment (Figure 3-1B). Previous studies implicated SMCHD1 role with methylation and chromatin condensation of this region that is lost in FSHD2 individuals. RT-PCR suggests that loss of SMCHD1 in SH-SH5Y cells increases expression of DUX4 gene (Figure 3-1C), emphasizing SMCHD1 role in regulating expression of DUX4 gene.
A

Raw ChIP-seq data

MACS2 Paired-end peak calling

SmcHD1 binding relative to TSSs

Motifs

Gene Ontology

B

No Treatment

Raw DATA

Peak calling

Control SmcHD1

5azaC Treatment

Raw DATA

Peak calling
Figure 3-1. Workflow for SMCHD1 ChIP-seq analysis in SH-SY5Y cells.

(A) Schematic illustration of the workflow for analysis of the ChIP-seq data. GREAT algorithm and SeqMonk software determined distribution of SMCHD1 binding sites relative to TSSs. SeqPos (Galaxy cistrome) identified potential SMCHD1-associated binding motifs. PAPST and Gene Ontology software identified biological processes associated with SMCHD1 occupancy relative to the nearest gene. (B) SMCHD1 occupancy was found in chromatin regions containing DUX4 and FRG1 genes as well as the intergenic regions between them on Chromosome 4. Illustration representing the conversion of SMCHD1 sequencing reads to peaks in SH-SY5Y cells treated with 5azaC (lower panel) and in cells without any treatment (upper panel). Red and blue lines represent forward and reverse sequencing reads, respectively. (C) mRNA quantification of DUX4 gene in SH-SY5Y (ctrl) and SMCHD1 KO cells (KO). The copy numbers are relative to beta-actin cDNA levels.

3.3.2. SMCHD1 binds majorly over intron and intergenic regions and its occupancy overlaps with binding sites for specific classes of transcription factors

To elucidate SMCHD1 binding across the genome, we determined how SMCHD1 called peaks positioned in relation to TSSs. SMCHD1 peaks were mostly found in intergenic and intron regions. When cells were treated with 5azaC there was a slight redistribution of SMCHD1 binding sites in intergenic regions (from 78.47% to 81.47%) at the expense of those located in introns (from 17.69% to 14.57%) (Supplemental Figure 3-10). The percentage of SMCHD1 binding sites in exon and promoter regions were unchanged with 5azaC treatment (Supplemental Figure 3-10). Most SMCHD1 binding sites were located distal from promoter regions (Figure 3-2B). These data suggest that SMCHD1 regulates gene expression through binding to regions distal to TSSs as well as more intergenic regions.
In an attempt to understand the nature of SMCHD1 DNA binding sites, we examined the composition of DNA sequences from called peaks where SMCHD1 were likely bound in cells grown under normal conditions as well as in cells treated with 5azaC. Under normal growth conditions, we found that SMCHD1 bound DNA sequences near binding sites for beta/beta/alpha-zinc fingers, helix-loop-helix families, hormone receptors and homeodomain proteins. A summary of the frequency of the binding sites for these and other transcription factors are illustrated in Supplemental Figure 3-11, including the consensus motifs and the associated transcription factor with the highest Z scores from the most frequently occurring transcription factor families. A total of 128 motif clusters were identified in peaks obtained from cells cultured under normal conditions and only 6 in cells treated with 5azaC. A complete list of individual transcription factors is provided in the supplemental information (Appendix B. Supplemental Table S3-1).
Figure 3-2. Genome-wide analysis of SMCHD1 binding sites in SH-SY5Y cells

(A) Distribution of SMCHD1 ChIP-seq peaks relative to promoters, gene bodies, exons, introns and intergenic regions. (B) SMCHD1 assigned peaks +/- 500kb relative to TSSs.
3.3.3. **SMCHD1 binds near genes acting in G-protein coupled receptor signalling pathways**

To further understand the role of SMCHD1 in the regulation of biological processes, we identified genes that were associated with SMCHD1 binding sites and their biological functions. We identified large number of SMCHD1 associated genes in control (5051) and 5azaC treated cells (2100), respectively (this is the same number as the total number of peaks) (Appendix B. Supplemental Table S3-2). Genome-wide SMCHD1 occupancy in control cells suggested G-protein coupled receptor signaling pathways and multicellular organism processes are important for SMCHD1 function (Figure 3-3A). 5azaC treatment significantly affects this association and no biological processes is assigned with SMCHD1 occupancy upon 5azaC treatment. Using the DAVID Bioinformatics resources, we generated twenty-one functional clusters that are associated with 33 genes identified in the G-protein coupled receptor signaling pathway (Appendix B. Supplemental Table S3-3). Figure 3-3B illustrates the top three clusters using genes vs functional categories. Next, we examined if SMCHD1 protein level changed the expression of genes involved in G-protein coupled receptor signaling. We generated a SH-SY5Y cell population in which we stably knocked-out (KO) SMCHD1 using CRISPR-Cas9 (Figure 3-4A). RT-qPCR suggested that loss of SMCHD1 in SH-SY5Y cells altered expression of genes involved in G-protein coupled signaling including GNAI1, ADRA2C, GPR37, GRM8, ADCY7 and 9 (Figure 3-4B and C). However, GPR37 protein levels did not appear to change in the absence of SMCHD1 (Supplemental Figure 3-12.). Also, treatment of cells with forskolin that activates the maximum levels of adenyl cyclase, resulted in elevated cyclic AMP levels in the SMCHD1 KO cells compared to wild type cells (Figure 3-4C).
A  Gene Ontology (GO) biological processes associated with SMCHD1

<table>
<thead>
<tr>
<th>GO Biological process</th>
<th>P value</th>
<th>Fold Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger</td>
<td>3.38E-02</td>
<td>2.5</td>
</tr>
<tr>
<td>multicellular organismal process</td>
<td>2.65E-02</td>
<td>1.18</td>
</tr>
</tbody>
</table>

B  Functional clustering of SMCHD1 peaks with putative target genes

Figure 3-3. GO biological processes and functional clustering of SMCHD1 peaks and putative target genes.
Biological processes associated with SMCHD1 occupancy in control cells. PAPST software identified SMCHD1 5051 peaks near genes in SH-SY5Y cells. Gene ontology consortium (Panther) found a number of genes that were overrepresented in the list associated with GO biological process. (B) The 33 genes identified in the G-protein coupled receptor signaling pathway were used to generate potential functional clusters. The top three are illustrated using genes vs functional categories. Above each cluster the gene names are given and to the right the associated GO terms are listed. The green squares indicate a positive gene-term association.
Figure 3-4. Forskolin stimulated cAMP levels and potentially G-protein coupled receptor signaling are altered in cells lacking SMCHD1.

(A) Protein level of SMCHD1 in SH-SY5Y and SMCHD1 sgRNA knockout cells. Beta-tubulin was used as an internal control for loading. (B) mRNA quantification of selected GPCR related genes (from the list of 33 genes) in SH-SY5Y (ctrl) and SMCHD1 KO cells (KO). The copy numbers are relative to beta-actin cDNA levels. (C) Left, mRNA quantification of ADCY7 and ADCY9 in SH-SY5Y (ctrl) and SMCHD1 KO cells (KO). The copy numbers are relative to beta-actin cDNA levels. Right, a box plot of cAMP concentration corrected for protein concentration in SH-SY5Y cells and SMCHD1 KO cells. Cyclic AMP levels were determined following forskolin activation as indicated. This experiment was repeated in four independent times in six replicates. Ctrl and SMCHD1 KO samples were compared using a Student’s T-test with the P-values indicated above the paired columns. ****P value < 0.0001 and ns= not significantly different.
Figure 3-5. Loss of SMCHD1 increases KCNQ1 expression level in SH-SY5Y cells.

(A) Schematic illustration representing the conversion of SMCHD1 ChIP sequencing reads to peaks in SH-SY5Y cells (ctrl) and cells treated with 5azaC. SMCHD1 occupancy was found in an intronic region of KCNQ1 gene. This occupancy was lost upon 5azaC treatment. (B) Left, mRNA quantification of KCNQ1 in SH-SY5Y (control) and SMCHD1 KO cells (KO). The copy numbers are relative to beta-actin cDNA levels. Right, SMCHD1 and KCNQ1 protein levels upon SMCHD1 KO in SY-SY5Y cells. LSD1 was used as an internal control.

3.3.4. SMCHD1 knock-out upregulates KCNQ1 expression level

It has been suggested that in response to activation of beta adrenergic receptors, ADCY9 and KCNQ1 interact and this interaction is mediated by a scaffolding A-kinase anchor protein (AKAP) Yotiao [205]. Yotiao brings about ADCY9, protein kinase A (PKA), protein phosphatase PP1, phosphodiesterase PDE4D3 and KCNQ1 protein to achieve regulation following Beta adrenergic stimulation. The ChIP-seq data suggests potential occupancy of SMCHD1 over the intronic region of KCNQ1 gene (Figure 3-5A). SMCHD1 occupancy was lost upon 5azaC treatment suggesting significant effects of 5azaC treatment on SMCHD1 occupancy over this region. We examined the role of SMCHD1 in regulating KCNQ1 gene expression level. KCNQ1 mRNA and protein levels were both elevated upon SMCHD1 KO in SH-SY5Y cells (Figure 3-5B and C). Since KCNQ1 is known to be imprinted and monoallelically expressed, we further examined if the upregulation of KCNQ1 is due to elevated expression level of one or both alleles. Using cDNA samples from SY5Y (ctrl) and SMCHD1 KO cells, we amplified KCNQ1 intronic and exonic regions encompassing single nucleotide polymorphism (SNP) (Figure 3-6). The amplified regions were then examined for the presence of SNPs by sanger sequencing. The sequencing results suggest that KCNQ1 is not imprinted in SH-SY5Y cells. In addition, it illustrates equal level of both alleles suggesting that upregulation of KCNQ1 gene upon SMCHD1 KO is due to elevated expression levels of both alleles.
Figure 3-6. Increased expression of KCNQ1 gene upon SMCHD1 KO is due to up-regulation of both alleles.

Sanger sequencing results of two regions containing SNPs over an intronic and exonic regions from SH-SY5Y (ctrl) and SMCHD1 KO cDNA samples.
3.4. Discussion

This is the first study that provides a high resolution of SMCHD1 occupied genomic sites that are sensitive to 5azaC treatment and provides support for its association with cellular signaling pathways.

Consistent with the previous study in murine NSC cells [341] which showed that SMCHD1 occupancy is not restricted to gene promoters, we found that SMCHD1 more actively binds to introns and intergenic regions in a DNA methylation sensitive manner. Our findings together with previous studies that showed SMCHD1 role in chromatin condensation raise the possibility that SMCHD1 acts as a chromatin organizer by mediating long-range chromatin folding. Indeed, the role of SMCHD1 in regulating gene clusters that are monoallelically expressed emphasizes its role in regulating regions with long-ranged interactions. We believe that it plays a role to organize chromatin to simultaneously regulate transcriptional activation and repression [342]. Chromatin capture could provide additional tools to understand its function. Compared to previous studies in murine NSCs, our ChIP-Seq data in human neuroblastoma SH-SY5Y cells found fewer binding sites near genes -5kb to +5kb relative to TSS (Figure 3-2B), although we identified more potential SMCHD1 binding sites using our custom antiserum. Others have used commercially available antibodies to explore the genome-wide association of SMCHD1. However, these antibodies are not designed for chromatin immunoprecipitation. We believe the antiserum we used may provide a better resolution of SMCHD1 genome-wide localization than other commercially available antibodies used to date. In this study, we examined the genome-wide occupancy of SMCHD1 relative to global level of DNA by treating cells with DNA demethylating reagent, 5azaC. 5azaC induces proteosomal degradation of DNMT1, therefore induces global loss of DNA methylation (Figure 3-7). Further analysis such as LUminometric Methylation Assay (LUMA) could estimate genome-wide DNA methylation level.

SMCHD1 binding sites overlap with specific transcription factor binding motifs. A previous study suggests opposing actions of SMCHD1 to CTCF in murine NSCs [341]. Our ChIP-seq data analysis and motif search suggest the CTCF binding motif as one of the SMCHD1 binding sites. However, this motif is not the most common in SMCHD1
ChIP-seq peaks (number 93 in the list of motifs ordered based on p-value). As it was previously shown, it would be interesting to determine if SMCHD1 has a role opposite to CTCF in human neuroblastoma SH-SY5Y cells and whether competition for the CTCF binding site plays a role in regulation of gene expression.

SMCHD1 occupancy overlaps with certain DNA binding motifs including beta/beta/alpha-zinc fingers, helix-loop-helix families, hormone receptors and homeodomain proteins, which are identified at high frequency in SMCHD1 ChIP-seq peaks. The majority of DNA binding motifs were sensitive to 5azaC treatment while a limited number were present in samples treated with 5azaC. This includes beta/beta/alpha-zinc-finger proteins, hormone nuclear receptor, Rel Homology region and CENP-B box binding family. The beta/beta/alpha-zinc-finger motif was the most common zinc finger motif that acts as a DNA binding domain and is found in various transcription factors including GLI3 [343]. GLI3 acts as both an activator and repressor of the Sonic hedgehog (Shh) signaling pathway and plays a critical role in embryonic development [343]. The GLI3 DNA binding domain has the highest p-value in SMCHD1 binding motifs. The next top two significant motifs are RXR-gamma and NR2F6 DNA binding motifs. These transcription factors belong to the hormone nuclear receptor family. RXR-gamma belongs to the retinoid X receptor (RXR) family of nuclear receptors that mediate the effects of retinoic acid (RA), and NR2F6 acts as a transcriptional repressor [344]. The significance of SMCHD1 binding to these DNA binding motifs and whether SMCHD1 has similar or opposing function requires investigation.

We used gene ontology and functional annotation bioinformatics to determine biological processes that are associated with SMCHD1 and if they are affected by 5azaC treatment. We identified G protein-coupled receptors as the top candidate biological process associated with SMCHD1 binding. Moreover, this association was affected by 5azaC treatment and no biological process was assigned to SMCHD1 functional clustering. The G protein-coupled receptor superfamily is comprised of about 1000 members that act as signal transducers [345, 346]. They trigger signaling molecules and act to signal environmental stimuli [347, 348]. Binding to signaling molecules causes conformational changes in these transmembrane proteins, leading to G protein binding. Activation of G proteins can affect thousands of intracellular second messengers.
including cyclic AMP (cAMP). cAMP production is involved in various biological processes including embryogenesis, olfaction, smooth muscle contraction, and learning and memory [349].

Adenylyl cyclase (AC) is an enzyme that converts ATP to cAMP and pyrophosphate. To investigate the association of SMCHD1 chromatin occupancy with the G protein-coupled signaling pathway using CRISPR Cas9, we knocked out SMCHD1 in SH-SY5Y cells. Loss of SMCHD1 led to upregulation of GPCRs including GPR37 and ADRA2C, upregulation of G protein GNAI1 and downregulation of the GPCR, GRM8. Loss of SMCHD1 also resulted in modified expression of at least two of the ten isoforms of adenyl cyclases in mammals; ADCY7 was downregulated while ADCY9 was upregulated. The modified expression of ADCY proteins did not lead to a change in the basal levels of cAMP. However, it significantly influenced the cell’s response to forskolin. Loss of SMCHD1 in SH-SY5Y cells resulted in more significant elevated level of cAMP in response to forskolin. It has been demonstrated that all ADCY family members except ADCY9 activate in response to forskolin treatment [350]. Thus, it is unclear what accounts for the high level of cAMP upon forskolin treatment in SMCHD1 KO cells. This also eliminates the possibility that SMCHD1 regulates cellular cAMP levels only through regulating ADCY9. There is not much known about the functional role of ADCY9, however recent data suggests that it associates with KCNQ1 protein through a scaffold protein, Yotiao in response to beta adrenergic receptor activation [205]. Our study shows upregulation of KCNQ1 at mRNA and protein level. KCNQ1 is the subunit of a voltage gated potassium channel (Iks) which mediates the slow delayed rectifying potassium current K⁺ crucial for the cardiac action potential repolarization [198, 199, 351]. Co-assembly of KCNQ1 and another member of KCN family of proteins like KCNE1 generates the Iks K⁺ current. Mutations in the KCNQ1 and KCNE1 interface cause long QT syndrome (LQT) and atrial fibrillation which results in prolongation of the QT interval of heart repolarization [206, 352]. Yotiao protein brings the Iks channel complex together with protein kinase PKA, protein phosphatase PP1, phosphodiesterase PDE4D3 and ADCY9 [205]. This leads to production of cAMP by ADCY9 that activates PKA and leads to phosphorylation of KCNQ1 at S27 residue. PDE4D3 in the complex ensures the basal Iks activity by cAMP degradation.
The KCNQ1 gene has been shown to be imprinted and monoallelically expressed. The Sanger sequencing suggests that elevated level of KCNQ1 gene expression is due to up-regulation of both alleles. The ChIP-seq data suggests potential occupancy of SMCHD1 in intronic region of KCNQ1 gene which is lost upon 5azaC treatment. Our data suggests no other SMCHD1 occupancy near KCNQ1 gene suggesting that SMCHD1 might function as a chromatin modifier by mediating long-ranged chromatin interactions, however further investigation is required. It should be mentioned that KCNQ1 gene is biallelically expressed in SH-SY5Y cells. To associate SMCHD1 function with regulation of imprinting and monoallelic gene expression, one would require a more suitable model such as human induced pluripotent stem cells that are known to maintain imprinting marks throughout the development [353]. Our study suggests regulation of KCNQ1 gene by SMCHD1 that could potentially contribute to potassium current and the LQT syndrome. Nevertheless, further investigation in cell model systems such as myoblast cells would be more informative.

In conclusion, this study reveals the influence of 5azaC treatment on SMCHD1 genome-wide occupancy and links it with second messenger signaling in cells. We showed that loss of SMCHD1 increases cellular cAMP levels in response to activation of adenylyl cyclases. cAMP as a second messenger plays roles in an array of functions in the human body [354, 355]. Therefore, further characterization of SMCHD1 molecular function and downstream actions have implications for drug development.

### 3.5. Materials and methods

#### 3.5.1. Cells, Antibodies and Reagents

Cell line used in this study was SH-SY5Y (ATCC, CRL-2266). 5-Azacytidine (5azaC) was purchased from Sigma (A1287).

Antibodies used in this study included: anti-beta-tubulin (Abcam, ab6046), commercial anti-SMCHD1 antibody (Bethyl NBP1-49969), anti-SMCHD1 antiserum (produced in house), anti-LSD1 antiserum (produced in house), anti-Dnmt1 (Abcam, ab19905), anti-
CHRM3 (Abcam, ab87199), anti-GPR37 (Abcam, ab166614), ADRA2C (Abcam, ab46536) and anti-GNAI1 (Sigma, SAB2100936).

3.5.2. Cell culture and 5azaC treatment

SH-SY5Y cells were cultured in Dulbecco’s Modified Eagle’s Medium (DMEM; Gibco) containing 4.5 g/L Glucose and L-Glutamine (Bio whittaker, Cat. # 12-604F) which was supplemented with 10% fetal bovine serum (FBS). The cells were maintained in a humidified atmosphere which contained 5% CO2 at 37°C. For 5azaC treatment, cells were treated with 10uM 5azaC every 24 hours for a period of 72 hours to induce global loss of DNA methylation.

3.5.3. SDS-PAGE and immunoblot

For preparation of whole cell lysate, cells were pelleted and washed once with PBS, then lysed in lysis buffer (PBS containing 1% triton X-100). Following resuspension of the pellet, the cells were sonicated briefly (Branson Sonifier 450. output 3.5 and constant duty cycle in pulses) and incubated on ice for 20 minutes, vortexed and then centrifuged for 5 minutes at 14,000 g. The supernatant was quantified, diluted and boiled in sample buffer for 5 minutes. Proteins were separated on 6% SDS-PAGE acrylamide gels using Tris-glycine buffering system [328]. After transferring of gels to PVDF membranes, the membranes were blocked in 0.05% milk powder in PBS containing 0.01% Tween-20 and then incubated with primary antibody (1:1000 dilution) overnight. Washes were done using PBS+0.01% Tween-20, and then the membranes were incubated with secondary HRP antibody (Jackson Labs,1: 50,000 dilution). The membranes were developed using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific, Cat. # 37071) and a cooled CCD instrument (Dyversity, Syngene) was used for detection.

3.5.4. ChIPSeq antiserum development

We designed immunogenic SMCHD1 peptides to generate anti-SMCHD1 antiserum from guinea pigs and used this antiserum for ChIPSeq. Nucleotides encoding amino acids 1620-1727 of human SMCHD1 were expressed in E. Coli BL21(pLsyS) using the
PET28a expression system (Novagen). The peptide antigen was isolated on a Nickel column using standard denaturing conditions (guanidium hydrochloride /Urea, Qiaexpressionist, Qiagen). After elution, the denatured peptide was renatured using a stepwise dilution protocol until the final buffer contained PBS. Next the peptide was mixed with alum for inoculation into a guinea pig. Following a standard 90-day inoculation protocol with a number of boosts, blood was collected and the serum tested for its efficiency in immunoprecipitation and used for ChIP seq.

3.5.5. ChIP-seq assay

We examined the ability of our ChIP-seq anti-SMCHD1 antiserum to immunoprecipitate SMCHD1 compared to a commercially available anti-SMCHD1 antibody. Chromatin immunoprecipitation with 5azaC treated and control SH-SY5Y cells was performed as previously described [70]. Briefly, twenty replicates (10cm plates) of 5azaC treated and control SH-SY5Y cells were fixed using 1% formaldehyde in HEPES (pH 7.8) for 8 min at room temperatures. Cells were then washed with PBS and re-suspended in lysis buffer (50mM Tris-HCl (pH 8.1), 1% SDS and 10mM EDTA) and sonicated using a Branon Sonifier 450 with an output of 3.5 and constant duty cycle in pulses to obtain 200 bp in size cross-linked DNA fragments. Five percent of the fragmented cross-linked chromatin was used as input and the rest was incubated with either 40 ul of anti-SMCHD1 antiserum or guinea pig serum overnight at 4 °C. Next, Protein A Sepharose beads were added for an 20 minutes prior to washing. The beads were then washed with RIPA (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 140 mM NaCl, 1% Triton-X 100, 0.1% Na-deoxycholate, 0.1% SDS and 1X protease inhibitor cocktail (Bioshop, Cat. # PIC003)), TSEII (20 mM Tris-HCl pH 8.1, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), TSE III (10 mM Tris-HCl pH 8.1, 0.25 M LiCl, 1 mM EDTA, 1%NP-40, 1% sodium deoxycholate) buffer and then 3 washes with 0.1X TE. Using 0.1 M NaHCO3 the DNA crosslinks were reversed overnight at 65°C. The replicate DNA samples were pooled and DNA was precipitated using 2 µL of pellet paint (Novagen), 1/10 volume 3M Na-acetate and 2 volumes of 100% EtOH by centrifugation for 10 min at 14,000 rpm. DNA pellets were washed with 70% EtOH, dried and re-suspended in 50 µL ddH2O. Michael Smith Genome Science Centre, Vancouver, Canada performed the
sequencing. The DNA libraries were prepared according to the Illumina’s (2000/2500) suggested protocol followed by paired-end sequencing. Details are available from their http://www.bcgsc.ca/platform/solexa/.

3.5.6. ChIP-seq data analysis

To identify genomic sites bound by SMCHD1 with high resolution, we first mapped the reads to the human genome (GRCh37, hg19) (bam file, performed by Canada’s Michael Smith Genome Science Centre), then used MACS2 paired peak calling for identification of SMCHD1 peaks [356]. The P-value and q-value cutoffs were both set as 0.05. The raw sequencing files as well as the MACS2 processed data are submitted to NCBI (GEO number GSE99227). A total of 5051 peaks were identified for the control SH-SY5Y cells and 2100 peaks for the 5azaC-treated SH-SY5Y cells (BED file). For evaluating SMCHD1 peak positions relative to the transcriptional start sites (TSSs), the Genomic Regions of Enrichment Analysis Tool (GREAT) software and SeqMonk program [357] (available at http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) were applied. In the GREAT software the default “basal plus extension” setting was used for associating genomic regions with genes (proximal 5.0kb upstream, 1.0 kb downstream, plus Distal up to 1000kb).

For motif analysis the SMCHD1 peaks were submitted to the SeqPos motif tool available in Galaxy cistrome [358]. Both cistrome and de novo motif search databases were used. The P-value cutoff was set as 0.001 (Appendix B. Supplemental Table S1).

For assigning SMCHD1 peaks to associated genes, PAPST software was used [359]. Using the SMCHD1 peaks (BED file) obtained from MACS2 paired, peaks were assigned +/- 1,750 kb relative to the TSSs of genes. The assigned genes were then submitted to the Gene Ontology Consortium (PANTHER) (http://amigo.geneontology.org/rte) for identification of biological processes associated with SMCHD1 in control and 5azaC-treated samples (Figure 3A).
3.5.7. CRISPR knockout of SMCHD1 in cells

We designed single guided RNA (sgRNA) targeting SMCHD1 and cloned them in CRISPR Cas9 PX459 plasmid following the Zhang Lab protocol [360]. In brief sgRNAs were designed using http://crispr.mit.edu/ software (Table S1). The SMCHD1 sgRNA was designed to target exon 18 of SMCHD1. The sgRNAs were then cloned into CRISPR Cas9 PX459 plasmid using BbsI as a cloning site according to the Zhang Lab protocol. Clones were sequenced (using oligonucleotide: gagggcctatttcccatgattcc) for confirmation of a positive clone. Transfection of SH-SY5Y cells was performed using jetPRIME transfection reagent (VWR Cat# CA89129-922) according to the manufacturer’s protocol. Stably transfected cells were selected using 3μg/ml puromycin 48 hours upon transfection.

3.5.8. Reverse transcription quantitative PCR

For RNA extraction, Trizol (Life Technologies, Cat. # 15596018) was used according to the manufacturer’s protocol. About 200ng of RNA was reverse-transcribed using Superscript II (Life Technologies, Cat. # 18964-014). StepOne Real Time PCR System (Life Technologies) and SYBR Advantage qPCR Premix (Clontech, Cat. # 638320) was used for quantification of cDNA. The Oligonucleotides used in this work are listed in Table S1. Following PCR, the PCR products were run on an agarose gel for confirmation of a single band amplification at the expected size. The threshold levels of each amplification were adjusted to the logarithmic part of the curve for determining a Ct value. Then the Ct values were normalized with those of the beta -actin to obtain the relative mRNA levels. The normalized data were analyzed using a Student’s t-test, and the confidence levels were displayed as P values.

3.5.9. Cyclic AMP Immunoassay

To evaluate cAMP expression, we pre-treated the SMCHD1-knockout and wild type SH-SY5Y cells with either 100nM or 1μM agonist, head activator (Cedarlane, Cat. # H-3790) for 30 min in DMEM supplemented with 0.5mM IBMX at 37°C. Then 10μM forskolin was added and the cells were incubated for another 15min to activate the maximum levels of
adenylyl cyclase. The cells were then washed with PBS and lysed using 1X cold lysis buffer (150mM NaCl, 1mM EDTA, 1% Triton X-100, 1 mM sodium orthovanadate) at 4°C for 30 min. Microplates coated with anti-mouse polyclonal antibody (Cedarlane, Cat. 400009) were incubated with 5 ul of 1/10,000 dilution of mAb cAMP antibody at 500 rpm for 1 hour at room temperature. The microplates were then washed four times with PBS-T (0.05% Tween). Next, samples, cAMP standards and lysis buffer were added to microplates. Following addition of 50 ul of 1/10,000 dilution of cAMP-HRP to all wells, the microplates were incubated at 500 rpm for 2 hours at room temperature. Following four washes with PBS-T, the TMB substrate (Cell Signalling, Cat. # 7004) was added to all wells and incubated for 15 minutes at room temperature. The reactions were halted by addition of Stop solution (Cell Signalling, Cat. # 7002) and the absorbance was measured at 450 nm using a Victor plate reader (PerkinElmer, Woodbridge, ON, Canada). Using the absorbance of the cAMP standards, the cAMP values were calculated for each sample.
3.6. Supplementary information

![Image of SH-SY5Y cells with 5azaC treatment](image)

**Figure 3-7.** 5azaC treatment induces loss of Dnmt1 level in SH-SY5Y cells.

Dnmt1 protein levels in SH-SY5Y cells upon 5azaC treatment. Beta-tubulin was used as an internal control for loading.
Figure 3-8. 5azaC treatment does not change SMCHD1 level in SH-SY5Y cells.

SMCHD1 protein levels in SH-SY5Y cells upon 5azaC treatment. LSD1 was used as an internal control for loading.
Figure 3-9. **Anti-SMCHD1 antiserum efficiently immunoprecipitates SMCHD1.**

mRNA quantification of DUX4 gene in SH-SY5Y (ctrl) and SMCHD1 KO cells (KO). The copy numbers are relative to beta-actin cDNA levels.
Figure 3-10. Genome-wide analysis of SMCHD1 binding sites in cells treated with 5-azaC.

(Top) Distribution of SMCHD1 ChIP-seq peaks relative to Intergenic, intron and gene promoters. (Below) SMCHD1 assigned peaks +/- 500kb relative to TSSs.
Figure 3-11. Types of Transcription factor family binding sites found associated SMCHD1 binding.

(Top) TF family binding sites associated with SMCHD1 bound sites in SH-SY5Y cells. The number of times each family was counted is indicated on the Y axis. * indicates the types of TF family motifs also found in cells treated with 5azaC. (Below) The highest score motifs (based on Z-scores) found for the TF family binding sites as indicated.

Figure 3-12. Loss of SMCHD1 did not change protein level of selected G protein-coupled receptor signaling proteins.

Protein level of CHRM3, GPR37 and HNAI1 upon loss of SMCHD1. Non-specific band was sued as a control for loading.
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM984</td>
<td>GGAGCGGAGTAAGATGATCG</td>
<td>GNAI1</td>
</tr>
<tr>
<td>SM985</td>
<td>CCAGCTAGCACAAGAGTTGG</td>
<td></td>
</tr>
<tr>
<td>SM986</td>
<td>CTGGCTTTCTGAGGGGAACTA</td>
<td>GRM8</td>
</tr>
<tr>
<td>SM987</td>
<td>GAAAATGCCACCTTTGGTTT</td>
<td></td>
</tr>
<tr>
<td>SM988</td>
<td>GTGCATCTGTGTGCCCATCA</td>
<td>ADRA2C</td>
</tr>
<tr>
<td>SM989</td>
<td>GATGCAGGAGGACAGGATGT</td>
<td></td>
</tr>
<tr>
<td>SM990</td>
<td>GGACTTCTCCTGCAAGATCG</td>
<td>GPR37</td>
</tr>
<tr>
<td>SM991</td>
<td>GCCACTAAACCCCCAAATCCT</td>
<td></td>
</tr>
<tr>
<td>SM1044</td>
<td>GTGCCGGACTTTCAAAGTGTT</td>
<td>ADCY7</td>
</tr>
<tr>
<td>SM1045</td>
<td>CGCCCGATGGTTGCTCTGTCT</td>
<td></td>
</tr>
<tr>
<td>SM1048</td>
<td>GGTGTCCCAGACCTACTCCA</td>
<td>ADCY9</td>
</tr>
<tr>
<td>SM1049</td>
<td>TGATCTTCTCGATGCTGTG</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CRISPR Oligos</th>
<th>Sequence</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM841</td>
<td>caccgTCGCATATTAGTCAACATGG</td>
<td>SMCHD1</td>
</tr>
<tr>
<td>SM842</td>
<td>aaacCCA TGT TGA CTA ATA TGC GAC</td>
<td>exon 18</td>
</tr>
</tbody>
</table>
Chapter 4.

Identification of SMCHD1 candidate binding partners

Contributions: Shabnam Massah contributed to all Figures. In the previous chapters (Chapter 2 and 3), I showed that SMCHD1 plays a role in regulating expression of X-linked and autosomal genes. In addition, I demonstrated that SMCHD1 genome-wide occupancy is sensitive to the global level of DNA methylation. In this chapter, I explored SMCHD1 candidate binding partners and examined SMCHD1’s interaction with candidate proteins.

4.1. Introduction

SMCHD1 is a non-canonical member of the SMC protein family [67]. Typically, the members of this protein family homo and hetero-dimerize [361]. For instance, eukaryotic SMC 1 and 3 constitute the core subunit of the cohesin complex and are involved in sister chromatid cohesion [32, 34]. SMC2 and 4 dimerize and act as the core of the condensin complex implicated in chromosome condensation [33, 285]. A dimer composed of SMC 5 and 6 functions in response to DNA repair [31]. It is well known that chromatin modifiers often exist in high molecular weight complexes with a series of other non-related proteins. SMCHD1 and other SMC family protein members are key components of multi-protein complexes with other effector molecules. Previously, it was shown that on the inactive X chromosome, SMCHD1 localized to heterochromatin regions enriched for histone H3 trimethyl Lysine at position 27 (H3K27me3) [68]. It interacts with nuclear receptor interacting factor 1 (LRIF1/HBiX1) which anchors SMCHD1 to regions enriched for histone H3 trimethyl lysine 9 (H3K9me3) to create a higher order folded sandwich chromatin structure [173]. However, it is unknown if SMCHD1 and HBiX1 interaction is essential for the role of SMCHD1 in regulating
expression of autosomal genes. In addition, identification of other SMCHD1 binding partners that participate in this or other regulatory complexes remains to be investigated.

4.2. Creating a tool for identification of SMCHD1 binding partners

4.2.1. Tandem affinity purification (TAP) for isolation of SMCHD1 and its binding partners

One of the key aspects of functional genomics is isolation of protein complexes and analysis of protein-protein interactions under native conditions. One of the most common approaches is tandem affinity purification (TAP) technology that was first developed by Rigaut, G et al. in 1999 for the analysis of protein-protein interactions in yeast [362]. This technique involves generating plasmids to create fusion of two affinity modules to the protein of interest that allows purification of the target protein along with its interacting partners in two consecutive steps. The original TAP tag consists of a calmodulin binding affinity peptide (CBP), two protein A IgG binding domains and a tobacco etch virus (TEV) protease cleavage site [363]. The dual affinity purification steps allow for protein isolation with marked reduction in nonspecific binding of proteins compared to the conventional single step immunoprecipitation. The original Prot A-CBP tag developed in yeast achieves lower protein recovery in higher eukaryotes. Specifically, the endogenous calmodulin protein in mammalian cells may block the target protein binding, therefore affecting protein recovery. In addition, the large size of the tag (21kDa) might impair protein function. The drawbacks of the original TAP tag technique lead to development of more than 30 alternative tags with improved efficiency in mammalian cells [364]. Among these modifications is 2x strep II- FLAG (SF) tag [365, 366]. This tag that is 5kDa in size consists of a tandem strep II and FLAG. The small size of this tag decreases the chance of impairing protein folding/function. During the first step of purification the Strep tag II of the bait protein is bound by Strep Tactin matrix and eluted by desthiobiotin. Next, the FLAG tag of the bait protein is bound to anti-FLAG agarose
and eluted by the FLAG peptide. Figure 4-1 illustrates the two-step purification of the SF-TAP tag.

**Figure 4-1. Outline of the SF-TAP procedure**

A) During the first phase of purification, the tandem Strep II tags are bound to the Strep Tactin Superflow. (B) Following washing steps, the bait protein and its interacting partners are eluted by desthiobiotin (blue diamonds). (C) During the second phase of purification the FLAG moiety tagged to the bait protein is bound to the anti-FLAG agarose. (D) Following washing steps the bait protein and its interacting partners are released by FLAG peptide.

### 4.3. Results

#### 4.3.1. Sub-cloning. Strep II-FLAG (SF) tag to the N-terminus of SMCHD1 cDNA clone and generation of a modified HEK293 cell population that stably expresses SF-SMCHD1

To identify potential SMCHD1 interacting proteins, I tagged SMCHD1 with Strep II-FLAG and carried out proteomic analysis in HEK293 cells stably expressing SF-SMCHD1. Among all the TAP-tag derivatives, we chose SF-tag for tagging SMCHD1 and
purification of its interacting partners because of its small size and high efficiency of isolation and elution. The SF-tag was inserted to the N-terminus of SMCHD1 cDNA clone.

HEK293 cells were transfected with a pcDNA3.0 plasmid encoding SF-SMCHD1. Following transfection, the cells were sparsely plated on 15cm plate to generate colonies from single cells. Upon formation of colonies, they were picked and expanded. Next, the colonies were examined for the expression of SF-SMCHD1. Figure 4-2 shows the results from an immunoblot of SF-SMCHD1 protein expression in HEK293 cells. One out of forty-eight clones were positive for the presence of SF-SMCHD1.

Figure 4-2. Stable expression of SF tagged SMCHD1 in HEK293 cells.
Immunoblot result illustrating the FLAG and SMCHD1 protein levels in HEK293 stably expressing SF-SMCHD1 and control HEK293 cells transfected with pcDNA3.0.

4.3.2. Identification of SMCHD1 potential binding partners

Since SMCHD1 is a nuclear protein, I used nuclear extracts from HEK293 cells stably expressing SF-SMCHD1 compared to control HEK293 cells expressing empty vector (pcDNA3.0) in tandem affinity purification. Figure 4-1 illustrates a flowchart of the SF-TAP purification procedure. After purification, the protein was concentrated by chloroform and methanol precipitation, then the samples were sent to the University of Victoria (Proteomics Core) for liquid chromatography-mass spectrometry (LC-MS/MS) analysis and identification of protein peptides. The SF-TAP purification protocol was
repeated twice. The first time at the final elution step, SMCHD1 and bound proteins were eluted using FLAG peptide. To increase the efficiency of the elution from FLAG agarose beads, in the second attempt, the bound proteins were boiled in SDS sample buffer for 10 min.

Based on peptides matched to genes/proteins, 18 and 33 unique candidate proteins from the first and second experiments were identified, respectively (Supplemental Table 4-4 and Table 4-5). In the SF-SMCHD1 sample, a number of predicted peptides masses (12 in the first and 94 in the second mass spectrometry) matched those of SMCHD1 protein indicating the tandem affinity purification was successful. Between the two SF-TAP purification trials, there were no matching candidate proteins. This suggests that addition of tag to SMCHD1 N-terminus possibly interrupted the proper folding, which could be accounted for non-reproducible data. We ranked the candidate proteins based on the number of peptides matches. The top 6 candidate proteins with the highest number of peptide matches from the first and second mass spectrometry are summarized in Table 4-1 and Table 4-2, respectively.

<table>
<thead>
<tr>
<th>Score</th>
<th>Mass</th>
<th>Matches</th>
<th>Sequences</th>
<th>emPAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMCHD1 Isoform 1</td>
<td>1</td>
<td>227942</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>CTTNBP2</td>
<td>0.75</td>
<td>183332</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>CEP250 Isoform 1</td>
<td>0.642857143</td>
<td>281880</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>ODZ3</td>
<td>0.571428571</td>
<td>305093</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td>Mass</td>
<td>Matches</td>
<td>Sequences</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>MPDZ Isoform 2</td>
<td>0.785714286</td>
<td>222792</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>DIS3L2</td>
<td>0.75</td>
<td>67954</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>ARID4B Isoform 2</td>
<td>0.642857143</td>
<td>138492</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 4-2.** Top candidates of SMCHD1 binding partners from the second mass spectrometry analysis.

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Mass</th>
<th>Matches</th>
<th>Sequences</th>
<th>emPAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMCHD1</td>
<td>1</td>
<td>227942</td>
<td>94</td>
<td>42</td>
<td>0.75</td>
</tr>
<tr>
<td>LRRN4</td>
<td>0.018592297</td>
<td>80162</td>
<td>85</td>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td>XKR4</td>
<td>0.013280212</td>
<td>72880</td>
<td>69</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>FBXL14</td>
<td>0.011952191</td>
<td>46769</td>
<td>61</td>
<td>2</td>
<td>0.08</td>
</tr>
<tr>
<td>TDRD9</td>
<td>0.011952191</td>
<td>148353</td>
<td>22</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>STK36 Isoform 1</td>
<td>0.015272244</td>
<td>145728</td>
<td>15</td>
<td>2</td>
<td>0.03</td>
</tr>
<tr>
<td>ACSM2B</td>
<td>0.016600266</td>
<td>65022</td>
<td>12</td>
<td>3</td>
<td>0.06</td>
</tr>
</tbody>
</table>
4.3.3. **Confirmation of the Mass spectrometry results by co-immunoprecipitation studies**

4.3.3.1  **Centrosome Associated Protein 250 (CEP250):**

One of the SMCHD1 candidate interacting protein suggested by the first mass spectrometry study was CEP250 that is also known as CNAP1 (Centrosomal Nek2-Associated protein 1). Some studies associate its role with cell cycle progression as it is involved in parental centrioles cohesion during interphase [367]. At the onset of mitosis, Nek2A kinase phosphorylates cohesion proteins including CNAP1, which leads to centrosome separation essential for spindle formation [367].

Other studies propose that CNAP1 is part of the condensin protein complex I [368]. This complex is composed of two SMC (SMC 2 and 4) and three non-SMC subunits. Condensin protein complexes are involved in sister chromatids condensation and resolution during mitosis. siRNA knockdown of CNAP1 in Hela cells impairs condensation and segregation of sister chromatids. The C-terminus of CNAP1 serves as a chromatin binding domain and nuclear localization signal (NLS) that co-precipitates with histone H3 in vivo suggesting that CNAP1 potentially targets itself and possibly the condensin complex to chromosomes [369]. It has been suggested that A kinase anchoring protein 95 recruits CNAP1 specifically to maternal chromosomes in zygotes to possibly mediate differential regulation of maternal and paternal chromosome condensation [370].

Since CNAP1 complexes with SMC proteins for its functional role, we sought to confirm the mass spectrometry data and determine if CNAP1 is a binding partner of SMCHD1. I performed co-immunoprecipitation studies to examine SMCHD1 binding with CNAP1 by two different strategies (Figure 4-3). A) I sub-cloned CNAP1 cDNA in a mammalian expression vector and transiently expressed CNAP1 in HEK293 cells stably expressing SF-SMCHD1. B) I sub-cloned HA and SF-tag to SMCHD1 and CNAP1 expressing vectors respectively and transiently expressed them in HEK293 cells for Co-immunoprecipitation studies.
Figure 4-3. Illustration of two co-immunoprecipitation strategies used for studying SMCHD1 and CNAP1 interaction.

(A) HEK293 cells stably expressing SF-SmHD1 were transfected with CNAP1-cDNA expression vector and used for immunoprecipitation using anti-FLAG antibody. (B) HEK293 cells were co-transfected with SF-CNAP1 and HA-SmHD1 cDNA clones for transient expression and immunoprecipitation with anti-HA antibody.

The results of both co-immunoprecipitation experiments did not suggest SMCHD1 and CNAP1 interaction (Figure 4-4). One possible factor that might have influenced the assay results is addition of affinity tags to SMCHD1 and CNAP1, that might change proper protein folding or interrupt the proteins interaction interphase. Also, transiently...
over expressed proteins might not achieve functional three-dimensional structure resulting in improper protein interaction and failure of the assay.

**A**

HEK293 cells expressing SF-SmcHD1 and CNAP1

300 KDa  
Immunoblot anti-SmcHD1

Beads  
IP-FLAG  
IP-CNAP1  
Input

250 KDa  
Immunoblot anti-CNAP1

Beads  
IP-FLAG  
IP-CNAP1  
Input

**B**

HEK293 cells expressing SF-CNAP1 and HA-SmcHD1

250 KDa  
Immunoblot anti-FLAG

IP-HA  
IP-FLAG  
Beads  
Input

300 KDa  
Immunoblot anti-HA

IP-HA  
IP-FLAG  
Beads  
Input
Figure 4-4. Co-immunoprecipitation studies, examining SMCHD1 and CNAP1 protein interaction.

(A) Co-immunoprecipitation studies were carried out with lysates prepared from HEK293 cells stably expressing SF-SMCHD1 and transiently expressing CNAP1. (Top) The immunoprecipitated samples were immunoblotted using anti-SMCHD1 antibody. (Bottom) The immunoprecipitated samples were immunoblotted using anti-CNAP1 antibody. (B) Co-immunoprecipitation studies were carried out using lysates from HEK293 cells transiently over-expressing SF-CNAP1 and HA-SMCHD1. (Top) The immunoprecipitated samples were immunoblotted using anti-FLAG antibody. (Bottom) The immunoprecipitated samples were immunoblotted using anti-HA antibody.

4.3.3.2 HP1 binding protein, HbiX1

As mentioned previously, over the inactive X chromosome, SMCHD1 interacts with HP1 binding protein, HbiX1. On the inactive X chromosome, SMCHD1 is localized over the heterochromatin regions enriched for trimethylation of histone H3 lysine 9 (H3K9me3) while HbiX1 is located over chromatin regions marked by trimethylation of histone H3 lysine 27 (H3K27me3). Interaction of these two proteins brings the two heterochromatin regions together to form a compact structure. Although the mass spectrometry data is not in line with the previous studies, we sought to examine SMCHD1 and HbiX1 interaction in cells transiently expressing HbiX1. To do so, I sub-cloned C-terminus region of HbiX1 containing the coiled coil domain that is shown to be the interface of SMCHD1 and HbiX1 interaction. In addition, I sub-cloned affinity HA-tag to either N- or C-terminus of HbiX1 cDNA clone. Next, I transiently over-expressed HbiX1-HA or HA-HbiX1 in HEK293 cells and used anti-HA antibody for immunoprecipitation of HbiX1 C-terminus domain and its interacting proteins. Co-immunoprecipitation using both cell lines failed to show interaction of endogenous SMCHD1 and over-expressed HbiX1 C-terminus domain (Figure 4-5). It is possible that the interactions interphase of these proteins are different in this cell system. In addition, the HA affinity tag might have interrupted the interaction.
Figure 4-5. Co-immunoprecipitation studies, examining SMCHD1 and HBiX1 protein interaction.

(A) Co-immunoprecipitation studies using anti-HA antibody were carried out with lysates prepared from HEK293 cells transiently over-expressing either HA-HBiX1 (Left panel) or HBiX1-HA (Right panel). The immunoprecipitated samples were immunoblotted using anti-SmcHD1 antibody. (B) Immunoprecipitated studies using anti-HA antibody were immunoblotted using anti-HA antibody.
4.4. Discussion

In this chapter, using tandem tag immunoprecipitation studies combined with mass spectrometry we explored SMCHD1 candidate binding partners and investigated SMCHD1 interaction with couple of candidates. We applied nuclear extraction of HEK293 cells stably expressing SF-SMCHD1 in tandem tag purification. As a control, nuclear extract of cells expressing empty vector (pcDNA3.0) was used. The SF-TAP purification was carried out twice, where SMCHD1 and interacting proteins were eluted either using FLAG peptide or by boiling in SDS sample buffer. Overall, 51 unique candidate proteins were identified, however, the TAP tag studies were non-reproducible that is possibly due to improper folding of tagged-SMCHD1. Further studies are required for examining the functional activity of tagged-SMCHD1. By introduction of tagged SMCHD1 into SMCHD1 KO cells and comparison of the expression profile, we could investigate the influence of adding a tag on SMCHD1 functional role.

Using co-immunoprecipitation studies, we investigated SMCHD1 interaction with couple of candidate proteins. One of the candidate proteins suggested by mass spectrometry was CNAP1 which has shown to be in the same complex with SMC proteins. Another candidate protein was HBiX1. Mass spectrometry did not suggest HBiX1 as a candidate binding partner, however previous studies showed interaction of SMCHD1 and HBiX1 [173]. Therefore, we investigated SMCHD1 interaction with HBiX1 by co-immunoprecipitation studies.

Co-immunoprecipitation studies failed to detect SMCHD1 interaction with either CNAP1 or HBiX1. One possible explanation is that addition of affinity tags might have influenced proper folding of SMCHD1 or have interrupted the interface of SMCHD1 interaction with binding partners. Lack of matching candidate proteins and inconsistency between the two attempts of purification is in line with this explanation. In addition, overexpression of proteins might result in improper protein folding, therefore affecting protein-protein interaction.

Despite limitations of this study, beside CNAP1 other candidate binding partners were suggested by mass spectrometry that might hold some promise and requires further
investigations. One of the SMCHD1 candidate genes suggested from the first mass spectrometry analysis is AT-rich interaction domain 4B (ARID4B) protein. ARID4B belongs to the ARID family of proteins with diverse function in chromatin remodeling and gene expression and is known to bind to AT-rich regions within chromatin [371]. ARID4B, also known as SAP180 (Sin3A associated protein 180) together with RBBP1-L1 (retinoblastoma binding protein 1-like 1) are components of the histone deacetylase dependent Sin3A transcriptional repressor in K562 leukemia cells [372]. It interacts with Sin3A-HDAC complex through a C-terminal repression domain. Previous studies suggested ARID4B and ARID4A (RBBP1) play roles as tumour suppressors and mutations in these genes contribute to malignant hematopoietic disorder and leukemia. Recent data links ARID4A and ARID4B with the control of gene expression at the imprinted domain associated with Prader-Willi Syndrome (PWS)/Angelman Syndrome (AS) [373]. Generally, it is known that the epigenetic imprinting of region 15q11-q13 is controlled in part by a bipartite imprinting center (PWS-IC/AS-IC) that is located within the SNRPN promoter. Deficiency of imprinted gene expression from the paternal and maternal chromosomes leads to the PWS and AS syndromes, respectively. Mouse ARID4A and ARID4B deletions resulted in reduced levels of H3K9me3, H4K20me3 and DNA methylation over the PWS-IC/AS-IC [373]. Moreover, recent data associates a role for SMCHD1 in regulating expression of the PWS/AS imprinted domain in mouse. It is possible that SMCHD1 regulates expression of the PWS/AS imprinted locus by partnering with effector molecules like ARID4B and ARID4A. Tudor domain containing 9 (TDRD9) besides ARID4B is another tudor domain containing protein identified in this screen. TDRD9 and its family member TDRD1 are involved in repression of retrotransposons in mammals [374]. About 46% and 39 % of the human and mouse genome is occupied by retrotransposable elements, respectively that are thought to be silenced by epigenetic mechanisms involving RNA interference (RNAi) [375, 376]. The RNAi guides effector complexes to the target mRNA to initiate degradation and suppression. There are different effector complexes that recognize and interact with RNAi including piwi proteins that interact with piwi-interacting small RNAs (piRNAs) [377, 378]. Two piwi proteins identified in mouse germline, MILI and MIWI2 that are key molecules in DNA methylation and repression of retrotrasnposons [379-381]. Some studies suggest that TDRD1 and TDRD9 are functional partners of MILI and MIWI2.
respectively and are essential for retrotransposon silencing in mice [374, 382, 383]. It is possible that SMCHD1 regulates gene expression through RNAi based mechanisms by partnering with effector complexes that recognizes and interact with RNAi. Some studies have investigated genome occupancy of SMCHD1 with the goal of informing on its functional role in regulating gene expression. There is a growing body of literature emphasizing the regulatory roles of RNA in epigenetic gene regulation. Thus, in the future it would be beneficial to investigate SMCHD1-RNA interactions as a mechanism in regulating gene expression.

Another potential SMCHD1 candidate binding protein is F-box and leucine-rich repeat protein 14 (FBXL-14) that is a member of F-box family of proteins. The members of this family are subunits of E3 ubiquitin ligase complexes, playing essential roles on proteasome degradation of specific substrates [384]. Several numbers of F-box proteins have been associated with epithelial to mesenchymal transition (EMT) that is defined by loss of epithelial phenotype and acquisition of mesenchymal and invasive characteristics including FBXL14 [385]. Generally, FBXL14 targets Snail1 and Twist, two EMT inducing transcription factors in mammalian cells [386, 387]. During low oxygen conditions or hypoxia that promotes EMT, FBXL14 gene expression is down regulated, which leads to stabilization of Snail1 and Twist. In colon cancer tumors, increased expression of hypoxia markers correlate with decreased transcriptional level of FBXL14 [386] and in neck squamous cell carcinoma, FBXL14 promotes proteasome degradation of Twist [387]. Interestingly, recent data suggest that SMCHD1 might function as a tumor suppressor. Transplantation of immune-deficient mice with SMCHD1-null mouse embryonic fibroblasts, enhanced tumor growth [388]. In Loss of SMCHD1 in Eμ-Myc transgenic mice that undergoes lymphomagenesis, accelerated development of B-cell lymphoma. Moreover, mutation or deletion of the region containing SMCHD1 has been identified in numerous cases of colorectal, breast and lung cancers. This suggests that SMCHD1 might act as a tumor suppressor in human malignancy; however, molecular mechanism of its function is unknown. It is possible that SMCHD1 partners with subunits of E3 ubiquitin ligase complexes like FBXL14 and mediate proteasome degradation of proteins that induce tumorigenesis, i.e. oncogenes.
Neuronal Leucine-Rich Repeat protein 4 (NLRR4), also known as LRRN4 is another SMCHD1 candidate binding partners. NLRR4 is a member of the LRR type I transmembrane family of proteins that are highly expressed in neuronal tissues and developing nervous system including NLRR1 and 2 [389, 390]. Number of LRR proteins are located in the synapse and contribute with the synapse formation in the developing central nervous system. NLRR4 is expressed in a subset of small-sized dorsal root ganglion; however, this pattern of expression is not consistent and is changed during development [391]. In addition, in the central nervous system NLRR4 is expressed in hippocampus and associates with hippocampus-dependent memory retention in mice [391]. The molecular mechanism of NLRR4 is yet to be identified. Moreover, whether SMCHD1 and NLRR4 interaction does exist and contribute to functional properties of nervous system requires investigations. It would be interesting to investigate if SMCHD1 knockout mice may also present some degree of hippocampus dependent memory loss.

4.5. Future directions

In this chapter, we explored SMCHD1 candidate binding partners and reported the top 6 candidate proteins with the highest number of the matched peptide from each mass spectrometry analysis. SMCHD1 interaction with CNAP1 and HBiX1 were investigated by co-immunoprecipitation studies, which failed to detect association of SMCHD1 and CNAP1 or HBiX1. Beside CNAP1 other potential binding partners were identified. It should be mentioned that the candidate partners with high number of peptide matches may not represent the actual interaction and might be simply present in the purified samples because of being abundant in cells. Nevertheless, due to the nature of the TAP-tag purification as well as the limitations of this study mentioned previously, the potential binding of SMCHD1 with candidate proteins requires further confirmation by co-immunoprecipitation and pull-down assays.

We have also categorized the candidate proteins according to their functions (Table 4-3).
Table 4-3. Functional classification of SMCHD1 candidate interacting proteins

<table>
<thead>
<tr>
<th>Functional Category</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulation Of Transcription</td>
<td>ZCRB1</td>
</tr>
<tr>
<td></td>
<td>PHF21A</td>
</tr>
<tr>
<td></td>
<td>PREB</td>
</tr>
<tr>
<td></td>
<td>TARDBP</td>
</tr>
<tr>
<td></td>
<td>TDRD9</td>
</tr>
<tr>
<td></td>
<td>CAMTA1</td>
</tr>
<tr>
<td></td>
<td>ARID4B Isoform 2</td>
</tr>
<tr>
<td></td>
<td>HOMEZ</td>
</tr>
<tr>
<td></td>
<td>RORB</td>
</tr>
<tr>
<td>Kinase / Kinase related</td>
<td>MARKS</td>
</tr>
<tr>
<td></td>
<td>CDK12</td>
</tr>
<tr>
<td></td>
<td>STK36</td>
</tr>
<tr>
<td></td>
<td>ULK1</td>
</tr>
<tr>
<td></td>
<td>NBEA</td>
</tr>
<tr>
<td>Cell Adhesion</td>
<td>PCDH19</td>
</tr>
<tr>
<td></td>
<td>CTNNA1</td>
</tr>
<tr>
<td></td>
<td>DSC1</td>
</tr>
<tr>
<td></td>
<td>ODZ3</td>
</tr>
<tr>
<td>Cetrosome, Centriolar, Cell Cycle</td>
<td>HYLS1</td>
</tr>
<tr>
<td></td>
<td>TTC28</td>
</tr>
<tr>
<td></td>
<td>ANAPC1</td>
</tr>
<tr>
<td></td>
<td>CEP250</td>
</tr>
<tr>
<td></td>
<td>C14orf145, CEP128</td>
</tr>
<tr>
<td>Translation/ Post translational modification</td>
<td>FBXL14</td>
</tr>
<tr>
<td></td>
<td>EEF1A2</td>
</tr>
<tr>
<td></td>
<td>EIF3CL, EIF3C</td>
</tr>
<tr>
<td>G-protein coupled receptor signaling</td>
<td>GPR101</td>
</tr>
<tr>
<td></td>
<td>PDE8A</td>
</tr>
<tr>
<td>Synaptic plasticity and Dendritic Spinogenesis</td>
<td>MPDZ</td>
</tr>
</tbody>
</table>
A functional category that is of interest contains proteins associated with the centrosome and centriolar and is involved in cell cycle. From this functional category, CNAP1 (CEP250) protein is discussed. Another candidate protein is anaphase promoting complex subunit 1 (ANAPC1) which is also named APC1. As its name suggests, this protein is a subunit of the anaphase-promoting complex or cyclosome (APC/C), a multi-subunit E3 ubiquitin ligase that plays an essential role in cell cycle progression and genome stability [392, 393]. Activation of this complex occurs through phosphorylation of APC1 and its partner APC3. Recent studies suggest that cyclin-dependent kinase 1 (CDK1) is essential for sequential phosphorylation of the APC3 and APC1 loop domains [394]. Interestingly, mass spectrometry data has also suggested candidate proteins with kinase activity including CDK12 that might involve in activation of APC/C.

Among 51 SMCHD1 associated candidate proteins, 9 of them are involved in transcriptional regulation. We are interested in transcription factors because it has been established that SMCHD1 is involved in regulation of gene expression. One candidate protein with low number of matched peptides that has transcriptional regulation function is PHF21A, also known as BHC80. BHC80 contains a PHD finger domain and is a key component of the BRAF-HDAC Complex (BHC) [395]. This complex consists of six subunits including CoREST, HDAC1, HDAC2, BRAF35, LSD1 and BHC80 and associates with neuron-specific gene repression [396]. BHC80 is required for LSD1 recruitment to and repression of target genes. In addition, BHC80 inhibits LSD1 histone demethylase activity in vitro while CoREST promotes LSD1 function [397]. BHC80 recognizes and binds to unmethylated H3K4 (H3K4) in vitro suggesting that BHC80 may act as a reader within the complex and adds unmethylated H3K4 as a histone code for regulation of chromatin state [397]. BHC80 binds to H3K4 through its PHD finger domain and this interaction is inhibited by methylation of H3K4.

The LC-MS/MS results suggest that SMCHD1 might be a subunit of transcriptional repressor complexes including Sin3A and CoREST. It is possible that SMCHD1 partner with different effector molecules to regulate gene expression depending on the target.
genes. Future research includes confirmation of the activity on these genes to regulate transcription. In addition, genome wide occupancy of SMCHD1 and binding partners in relation to histone modifications could be assessed.

4.6. Materials and methods

4.6.1. Sub-cloning SF-Tag to the N-terminus of SMCHD1 cDNA clone

The SF-TAP tag was amplified from N-SF-TAP (pcDNA3.0) obtained from Gloenchner et al. [366] using the following primers (Forward Primer: ATGGATCCATGGATTATAAAGATGATGATGATAAAGGGTCGG and Reverse Primer: ATGGATCCACCAGAACCACCTCCTTTCTCGAACTGCGGGGTG). The PCR product was then digested with BamHI and ligated into BamHI site of the SMCHD1 pcDNA cDNA clone (Figure 4-6).
Figure 4-6. Plasmid map of the SF-SMCHD1 pcDNA 3.0 clone.

(Top) representation of the circular map for SF- SMCHD1 pcDNA 3.0 clone. The SF-tag is inserted at N terminus of SMCHD1 cDNA clone using BamHI site. The relative position of restriction sites is indicated. (Bottom) illustrates the schematic of the SF-tag that consists one FLAG and two Strep II moiety.

4.6.2. Generation of HEK293 cells stably expressing SF- SMCHD1

HEK293 cells were seeded in 6 well tissue culture plates prior to transfection. The cells were transfected at 60% confluency with 4 µg DNA (control pcDNA3.0 or SF-SMCHD1 pcDNA3.0) in 400 µL serum free media containing 6 µL Turbofect (Fisher Scientific Cat # FERR0531) for 15 min at room temperature. The media was replaced after 6 hours. Twenty-four hours upon transfection, each well was transferred onto 15 cm plates and
selection with 500 µg/mL G418 for 3 weeks. Next, at least 48 clones were picked using hole punch sized whatman paper and expanded for SF-SMCHD1 immunoblot screening.

4.6.3. Nuclear Extract preparation

For nuclear extract preparation, the HEK293 cells were washed twice with ice cold PBS and incubated on ice in buffer A (10 mM HEPES-KOH pH 7.9, 1.5 mM MgCl2 10 mM KCl, 0.5 mM DTT and 1X protease inhibitor cocktail) for 10 min. Following cell lysis by vortexing, the nuclei were collected by centrifugation at 4400 rpm for 15 min. The pellet was re-suspended and incubated in 1 pellet volume of Buffer C (20 mM HEPES-KOH pH7.9, 1.5 mM MgCl2, 420 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT, 25% glycerol and 1X protease inhibitor cocktail) on ice for 20 min. Following centrifugation at 15,000 rpm, the supernatant was collected and applied in SF-TAP purification protocol.

4.6.4. SF-TAP purification protocol

The procedure was followed as Gloenchner et al [366]. In brief, the parental HEK293 cells and those stably expressing SF-SMCHD1 were plated in 15cm plates. Six plates were used per sample. Upon removal of media and a wash with warm PBS, the cells were scraped and lysed with one mL of lysis buffer (0.03 M Tris, 0.15M sodium chloride, 1X protease inhibitor, 1X phosphatase inhibitor (Bioshop Cat. # PIC002) and 0.5% NP-40) per 15 cm plate on ice for 20 min. Upon centrifugation, the lysates were cleared through 0.22 µm syringe filters. The lysates were then incubated with 50 µL per plate on Strep-Tactin superflow resin for one hour at 4°C. The unbound protein complexes were removed by centrifugation and washed three times with 500 µL wash buffer (0.03 M Tris, 1X phosphatase inhibitor I and II and 0.5% NP-40). The bound protein complexes were then eluted with 500 µL desthiobiotin elution buffer (2mM desthiobiotin) for 10 min by rocking at room temperature. For the second step of purification, the eluate from the first step were added to 25 µL per plate anti-FLAG-M2-agarose and incubated for one hour at 4°C with rocking. After incubation, the agarose beads were washed three times with wash buffer (0.03 M Tris, 1X phosphatase inhibitor (Bioshop Cat. # PIC002) and 0.5%
NP-40) and eluted either with 200 µL FLAG peptide elution buffer (200 µg FLAG peptide per mL) or boiled at 100°C in 6X SDS sample buffer for 10 min.

4.6.4.1 Precipitation of the purified protein complexes
The eluted protein was precipitated by mixing with 2 µL of pellet paint (Novagen Cat# 69049), 1/10 volume 3M Na-acetate and 2 volumes of 100% EtOH followed by centrifugation at 4°C for 10 min at max speed. Protein pellets were then washed with 70% EtOH, dried and re-suspended in 50 µL ddH2O for LC-MS/MS.

4.6.5. Sub-cloning SF-tag to CNAP1 cDNA clones
CNAP1 cDNA clone was purchased from Cedarlane (MHS1768-101377378) and sub-cloned into pcDNA 3.0 mammalian expression vector using BamHI and NotI restriction sites (Figure 4-7).
Figure 4-7. Plasmid map of CNAP1 pcDNA 3.0 clone

Representation of the circular map for CNAP1 pcDNA 3.0 clone. The relative position of restriction sites is indicated.

Similar to SF-SMCHD1, for sub-cloning SF-tag to CNAP1 cDNA clone, the SF-TAP tag was amplified from N-SF-TAP (pcDNA3.0) obtained from Gloenchner et al. [366] using the following primers (Forward Primer: ATGGATCCATGGATTATAAAGATGATGATGATAAAGGGTCGG and Reverse Primer: ATGGATCCACCAGAACCACCTCCTTTTCTCGAACTGCGGGTG). Following digestion of the PCR product with BamHI, the tag was ligated into BamHI site of the CNAP1 pcDNA cDNA clone (Figure 4-8).
Figure 4-8. Plasmid map of the SF-CNAP1 pcDNA 3.0 clone.
(Top) Illustration of the circular map for SF-CNAP1 pcDNA 3.0 clone. The SF-tag is inserted at the BamHI site located at the N terminus of CNAP1 cDNA clone. The relative position of restriction sites is indicated. (Bottom) Representation of the SF-tag that consists one FLAG and two Strep II moieties.

4.6.6. Sub-cloning HA-tag to HBiX1 cDNA clones

HBiX1 cDNA clone was purchased from Cedarlane (HG16564-U) and sub-cloned into pcDNA 3.0 using BamHI and XhoI restriction sites. The HA-tag was purchased as oligonucleotides with appropriate restriction site overhangs and following annealing of the forward and reverse strands, they were either ligated at KpnI/BamHI or XhoI/Apal.
sites of HBiX1 cDNA clone for addition of tag at N-and C- terminus respectively (Figure 4-9 and 4-10).

Figure 4-9. Plasmid map of the HA-HBiX1 pcDNA 3.0 clone.
(Top) Illustration of the circular map for HA-HBiX1 pcDNA 3.0 clone. The HA-tag is inserted at the KpnI/BamHI site located at the N-terminus of HBiX1 cDNA clone. The relative position of restriction sites is indicated. (Bottom) Representation of the HA-tag that consists of KpnI and BamHI restriction site overhangs.
Figure 4-10. Plasmid map of the HBiX1–HA pcDNA 3.0 clone.

(Top) Illustration of the circular map for HBiX1-HA pcDNA 3.0 clone. The HA-tag is inserted at the XhoI/Apal site located at the C-terminus of HBiX1 cDNA clone. The relative position of restriction sites is indicated. (Bottom) Representation of the HA-tag that consists of XhoI and Apal restriction site overhangs.

TCGAGTACCCATACGATGTTCCAGATTACGCTTAGGGGCC
CATGGGTATGTCTACGTTAAATAGGCTATCGGATCC
4.6.7. **Sub-cloning HA-tag to SMCHD1 cDNA clones**

The HA-tag was purchased as oligonucleotides with KpnI/BamHI restriction site overhangs. Following annealing of the forward and reverse strands, they were ligated at KpnI/BamHI site located at N-terminus of SMCHD1 cDNA clone (Figure 4-11).

![Diagram of SmcHD1 a.a. 1-2007 with HA tag and restriction sites](Figure 4-11)

**HA Tag with KpnI (N-terminus) and BamHI (C-terminus) restriction sites overhangs**

<table>
<thead>
<tr>
<th>CATGTACCCCATACGATGTTCCAGATTACGCTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATGATACATGGGATATGCTACAAGGTCTAATGCGACCTAG</td>
</tr>
</tbody>
</table>

**Figure 4-11.** Plasmid map of the HA-SMCHD1 pcDNA 3.0 clone.

(Top) Illustration of the circular map for HA-SMCHD1 pcDNA 3.0 clone. The HA-tag is inserted at the KpnI/BamHI site located at the N-terminus of SMCHD1 cDNA clone. The relative position of restriction sites is indicated. (Bottom) Illustration of the HA-tag consisting of KpnI and BamHI restriction site overhangs.
4.6.8. Co-Immunoprecipitation and SDS-PAGE

Whole cell lysate from HEK293 transfected cells were prepared by washing cell with PBS followed by cell lysis in one pellet volume lysis buffer (lysis buffer: 350mM NaCl, 0.5% NP-40, 5% Glyserol, 2 mM MgCl$_2$, 25mM Tris-HCl (pH 7.4), 1X protease inhibitor). The cells were lysed at 4°C for 20 min while rotating. Following cell lysis, the whole cell lysate was collected by centrifugation at 15000rpm for 5 min. For each immunoprecipitation experiment, 2 mg whole cell lysate was used in 2 mls total volume of PBS. The lysate was incubated with 1/300 primary antibody overnight at 4°C while rocking. Next, 25 ul of 50% A-sepharose slurry beads were added and the samples were incubated at room temperature for 20 min while rocking. The beads were collected by centrifugation at 2000rpm for 1 min. The beads were then wash 3 times with cold PBS for 2 min at room temperature. Following the last wash 20 ul of 6XSDS sample buffer was added to the beads. Proteins were separated on 6% SDS-PAGE acrylamide gels using a Tris-acetate buffering system [328]. Next the proteins were transferred to nitrocellulose membranes (Millipore, Protran BA85). Following transfer, membranes were blocked in 0.05% milk powder in PBS containing 0.01% Tween-20 overnight in 1:1000 primary antibody. Following washing in PBS+0.01% Tween-20, the membranes were incubated with secondary HRP antibody (Jackson Labs, 1: 50,000 dilution). The membranes were then developed using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific, Cat. # 37071) and visualized using a cooled CCD instrument (Dyversity, Syngene).

4.7. Supplementary information

Table 4-4. First mass spectrometry analysis of SF-SMCHD1 immunoprecipitation

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Mass</th>
<th>Matches</th>
<th>Sequences</th>
<th>emPAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCDC87</td>
<td>0.857142857</td>
<td>96812</td>
<td>151</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Gene</td>
<td>Value</td>
<td>ID</td>
<td>Repeat</td>
<td>Count</td>
<td>P-value</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>PCDH19 Isoform 2</td>
<td>1.035714286</td>
<td>121938</td>
<td>134</td>
<td>2</td>
<td>0.03</td>
</tr>
<tr>
<td>SLC12A7 Isoform 1</td>
<td>0.535714286</td>
<td>120283</td>
<td>14</td>
<td>4</td>
<td>0.03</td>
</tr>
<tr>
<td>SMCHD1 Isoform 1</td>
<td>1</td>
<td>227942</td>
<td>12</td>
<td>4</td>
<td>0.05</td>
</tr>
<tr>
<td>CTTNBP2</td>
<td>0.75</td>
<td>183332</td>
<td>12</td>
<td>2</td>
<td>0.02</td>
</tr>
<tr>
<td>CEP250 Isoform 1</td>
<td>0.642857143</td>
<td>281880</td>
<td>10</td>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>ODZ3</td>
<td>0.571428571</td>
<td>305093</td>
<td>10</td>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>MPDZ Isoform 2</td>
<td>0.785714286</td>
<td>222792</td>
<td>5</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>DIS3L2</td>
<td>0.75</td>
<td>67954</td>
<td>4</td>
<td>3</td>
<td>0.11</td>
</tr>
<tr>
<td>ARID4B Isoform 2</td>
<td>0.642857143</td>
<td>138492</td>
<td>4</td>
<td>2</td>
<td>0.03</td>
</tr>
<tr>
<td>EXOC4</td>
<td>0.607142857</td>
<td>111170</td>
<td>4</td>
<td>3</td>
<td>0.03</td>
</tr>
<tr>
<td>KPRP</td>
<td>0.714285714</td>
<td>67172</td>
<td>3</td>
<td>2</td>
<td>0.06</td>
</tr>
<tr>
<td>GRID2</td>
<td>0.607142857</td>
<td>114271</td>
<td>3</td>
<td>2</td>
<td>0.03</td>
</tr>
<tr>
<td>HOMEZ</td>
<td>0.5</td>
<td>58938</td>
<td>3</td>
<td>1</td>
<td>0.06</td>
</tr>
<tr>
<td>EEF1A2</td>
<td>1.107142857</td>
<td>50780</td>
<td>2</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>SRRD</td>
<td>0.821428571</td>
<td>39062</td>
<td>2</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>DSC1</td>
<td>0.714285714</td>
<td>94916</td>
<td>2</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Protein</td>
<td>Score</td>
<td>Mass</td>
<td>Matches</td>
<td>Sequences</td>
<td>emPAI</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>CSMD1</td>
<td>0.678571429</td>
<td>396718</td>
<td>2</td>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>EIF3CL EIF3C</td>
<td>0.642857143</td>
<td>105962</td>
<td>1</td>
<td>1</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 4-5. Second mass spectrometry analysis of SF-SMCHD1 immunoprecipitation
<table>
<thead>
<tr>
<th>Gene</th>
<th>Score</th>
<th>FPKM</th>
<th>Length</th>
<th>Hits</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLK Isoform 1</td>
<td>0.011952191</td>
<td>143234</td>
<td>7</td>
<td>3</td>
<td>0.03</td>
</tr>
<tr>
<td>TTC28</td>
<td>0.015272244</td>
<td>272653</td>
<td>6</td>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>NBEA</td>
<td>0.011952191</td>
<td>59563</td>
<td>6</td>
<td>2</td>
<td>0.06</td>
</tr>
<tr>
<td>YBX1</td>
<td>0.037848606</td>
<td>42333</td>
<td>5</td>
<td>3</td>
<td>0.09</td>
</tr>
<tr>
<td>LENG1</td>
<td>0.069721116</td>
<td>30511</td>
<td>4</td>
<td>2</td>
<td>0.12</td>
</tr>
<tr>
<td>RORB Isoform 2</td>
<td>0.015936255</td>
<td>54269</td>
<td>4</td>
<td>3</td>
<td>0.07</td>
</tr>
<tr>
<td>ANAPC1</td>
<td>0.013944223</td>
<td>218528</td>
<td>4</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>SLC25A39</td>
<td>0.013280212</td>
<td>39589</td>
<td>4</td>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>FAM83G</td>
<td>0.017264276</td>
<td>51850</td>
<td>3</td>
<td>2</td>
<td>0.07</td>
</tr>
<tr>
<td>ULK1</td>
<td>0.011952191</td>
<td>113557</td>
<td>3</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td>PDE8A</td>
<td>0.011288181</td>
<td>94385</td>
<td>3</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>CCDC64</td>
<td>0.01062417</td>
<td>20267</td>
<td>3</td>
<td>2</td>
<td>0.19</td>
</tr>
<tr>
<td>ZCRB1</td>
<td>0.02124834</td>
<td>24805</td>
<td>2</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>CRKRS</td>
<td>0.020584329</td>
<td>164738</td>
<td>2</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>HYLS1</td>
<td>0.017928287</td>
<td>34565</td>
<td>2</td>
<td>1</td>
<td>0.11</td>
</tr>
<tr>
<td>CFB Isoform 1</td>
<td>0.015936255</td>
<td>86847</td>
<td>2</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Gene</td>
<td>Value</td>
<td>Score</td>
<td>Level</td>
<td>Change</td>
<td>Value</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>PREB</td>
<td>0.014608234</td>
<td>11872</td>
<td>2</td>
<td>1</td>
<td>0.33</td>
</tr>
<tr>
<td>FGF4</td>
<td>0.013280212</td>
<td>22148</td>
<td>2</td>
<td>1</td>
<td>0.17</td>
</tr>
<tr>
<td>TARDBP</td>
<td>0.011952191</td>
<td>45305</td>
<td>2</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>TMEM202</td>
<td>0.011952191</td>
<td>31845</td>
<td>2</td>
<td>1</td>
<td>0.12</td>
</tr>
<tr>
<td>GPR101</td>
<td>0.011288181</td>
<td>57478</td>
<td>2</td>
<td>1</td>
<td>0.06</td>
</tr>
<tr>
<td>CAMTA1</td>
<td>0.011288181</td>
<td>18566</td>
<td>2</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>MARCKS</td>
<td>0.047808765</td>
<td>31707</td>
<td>1</td>
<td>1</td>
<td>0.12</td>
</tr>
<tr>
<td>PHF21A</td>
<td>0.019920319</td>
<td>70793</td>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>APOBEC4</td>
<td>0.013944223</td>
<td>41954</td>
<td>1</td>
<td>1</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Chapter 5.

SMCHD1 and Human Disease, Conclusions and Future Directions

Monoallelic gene expression or allelic exclusion mediate X chromosome inactivation in mammals, regulates stochastic cell specific expression of select clustered protocadherins and ensures parent-of-origin expression of imprinted genes. To date, it is understood that nearly 5 to 10 percent of all genes are expressed monoallelically in humans [398]. This would greatly impact genome wide association studies (GWAS) where single nucleotide polymorphisms (SNPs) are correlated with diseases in human populations. Most of these studies fail to incorporate single allelic expression in their algorithms and it is unknown if SNPs are located on the active or silent allele. Nonetheless, in most studies the individual is considered positive for the correlation between phenotype and a genetic determinant (i.e. SNP), even though the SNP might be located on the silent allele. Future studies interrogating SNPs or other genetic measures should consider incorporating monoallelic gene expression in their disease models.

Loss of allelic exclusion contributes to developmental diseases and the precise regulation of genome wide monoallelic gene expression is not known. Epigenetic mechanisms are fundamental players determining monoallelic expression in a homogenous genetic background. DNA methylation is one of the mediators of the epigenetic phenomena. DNA methylation is known to recruit and or block binding of different effector molecules or transcription factors. As discussed in this thesis, one epigenetic modulator is SMCHD1. Its role was first discovered during X chromosome inactivation where it is involved in methylation of a subset of CpG islands in a late stage post fertilization step [68, 172]. Later, its role in regulation of autosomal genes with monoallelic expression was proposed (chapter 2) [69-71].
In addition, a large proportion of literature associates SMCHD1 activity with facioscapulohumeral muscular dystrophy (FSHD) as an epigenetic silencer [73, 315, 316, 399]. FSHD is a muscular dystrophy characterized by progressive weakness of shoulder and face muscles extending to arms and legs [400-404]. Variations in severity, progression, gender bias and age suggest epigenetic mechanisms will play a central role in the etiology of FSHD [405]. The two forms of FSHD, FSHD1 and FSHD2, are both associated with the 4q35 sub-telomeric haplotype 4qA [73, 316, 402]. The more frequent form, FSHD1, is associated with deletion of a macrosatellite repeat designated as D4Z4 that is tandemly arrayed from head to tail [406]. The number of the repeats in healthy individuals varies between 11 to 100 and FSHD1 individuals contains less than 11 D4Z4 repeats [407]. Loss of D4Z4 repeats also correlates with a marked loss of DNA methylation and repressive histone marks [408-410]. On the other hand, in FSHD2 individuals that include about 5% of FSHD patients, the number of D4Z4 repeats is within the normal range. FSHD2 also associates with the 4qA haplotype and there is also a significant loss of DNA methylation at D4Z4 repeats [411]. FSHD2 is caused by missense, nonsense and/or deletion mutations in SMCHD1 [73, 335, 412]. The identified mutations encompass different regions of SMCHD1 including ATPase domain and hinge domain which all contribute to hypo-methylation of the D4Z4 array and FSHD2 (Figure 5-1 and Table 5-1). Each D4Z4 repeat (about 3.3kb) encompasses the coding sequence of a nuclear transcription factor DUX4. De-repression of the D4Z4 repeats leads to DUX4 gene expression. In addition, SMCHD1 is a modifier of disease severity in FSHD1 patients, as individuals carrying contracted D4Z4 repeats and SMCHD1 mutations display more severe symptoms than expected from their D4Z4 repeat copy number [316, 399].
Table 5-1. Summary of SMCHD1 mutations identified in FSHD patients.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Type of mutation</th>
<th>Outcome/SmcHD1 level of expression</th>
<th>Methylation of D4Z4 repeats</th>
<th>Author/ Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.1040+1G&gt;A, heterozygous</td>
<td>Exon 9</td>
<td>Error 5’ splice site</td>
<td>nonsense-mediated decay (NMD)</td>
<td>hypomethylated</td>
</tr>
<tr>
<td>c.2068C&gt;T</td>
<td>Exon 16</td>
<td>Missense</td>
<td></td>
<td>hypomethylated</td>
</tr>
<tr>
<td>c.3274_3276+2del</td>
<td>Intron 25</td>
<td>Missense</td>
<td>NO change at protein level</td>
<td></td>
</tr>
<tr>
<td>c.4661T&gt;C</td>
<td>Exon 37</td>
<td>Missense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.623_625 deletion of Lys275</td>
<td>Exon 38</td>
<td>Deletion</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Work from this thesis has demonstrated that in SH-SY5Y human neuroblastoma cells, SMCHD1 binds to the region on chromosome 4 that encompasses DUX4 and myopathic potential gene FRG1 (FSHD region gene 1). This occupancy was greatly affected by the global loss of DNA methylation. Therefore, it appears that DNA methylation is required for SMCHD1 binding over this region. In addition, loss of SMCHD1 results in up regulation of DUX4 gene expression. This supports a role for SMCHD1 in regulating DUX4 gene repression and potentially its contribution to FSHD.

Recently, a long non-coding RNA, DBE-T, is thought to contribute to the etiology of FSHD [413]. The DBE-T is located within each D4Z4 repeat. Normally, in healthy population the repressive epigenetic marks over the D4Z4 repeats are thought to silence DBE-T ncRNA expression. When the number of repeats reaches less than 11, the DNA methylation and repressive histone marks decreases, providing a permissive environment for DBE-T gene expression. This ncRNA recruits trithorax (TrxG) ASH1L
protein and promotes activation of DUX4 and the adjacent genes with myopathic potentials including ANT1 and FRG1. Interestingly, the ChIP-seq data from SH-SY5Y suggests that occupancy of SMCHD1 to DBE-T ncRNA is lost upon 5azaC treatment, thus global loss of DNA methylation. This suggests that SMCHD1 potentially represses DBE-T ncRNA expression and that SMCHD1 mutations lead to removal of repressive marks, resulting in DBE-T ncRNA expression and DUX4 gene activation. However, further investigation is required for characterization of SMCHD1-DBE-T ncRNA mediated gene silencing.

SMCHD1 knockdown in HEK293 cells (chapter 2) and knockout in SH-SY5Y cells (chapter 3) both resulted in elevated levels of KCNQ1 gene expression [70]. KCNQ1 role has been majorly investigated in cardiac cells. As mentioned in chapters 1 and 3, KCNQ1 is the alpha subunit of a voltage gated potassium channel (I\textsubscript{Ks}). KCNQ1 co-assembles with other members of the KCN family of proteins to form the ion channel. The best-known partner is KCNE1 [198, 199]. This potassium channel is responsible for slow delayed rectifying potassium current that is essential for cardiac cell repolarization upon cardiac action potential [351]. KCNQ1 interacts with KCNE1 distal C terminus through its helix C domain. This interface is crucial for \(I_{\text{Ks}}\) channel modulation and mutations interrupting this interaction cause long QT syndrome (LQT) [206, 352]. LQT syndrome is characterized by the prolongation of the QT intervals in the electrocardiogram (ECC/EKG) that represents delayed repolarization of the heart following the action potential [414]. SMCHD1 also regulates expression of ADCY9. Recent data suggests that the AKAP9 (Yotiao) protein functions as a scaffold and assembles with PP1, PDE4D3, PKA and ADCY9 through a C terminal leucine zipper motif on KCNQ1 [205]. This leads to cAMP production by ADCY9, which results in phosphorylation and activation of KCNQ1 by PKA.

Changes in expression of several AKAP genes associate with physiological dysfunction in heart including the gene Yotiao [415-421]. Mutations in the leucine zipper motif of Yotiao have been found in individuals with LQT syndrome. In particular, the S1570L has been found in 2% of individuals suffering from LQT syndrome [422]. Computational analysis suggests that this mutation reduces cAMP dependent phosphorylation of \(I_{\text{Ks}}\) and results in a prolonged action potential. This potentially links a role for SMCHD1 in
modulating heart function through regulating expression of KCNQ1 and ADCY9 and possibly Yotiao.

Association of FSHD with heart function alteration is controversial. While there are reports that documented cases of FSHD for cardiac involvement and suggested high susceptibility to induced atrial flutter or fibrillation during electrophysiologic studies [423-426], others did not find any cardiac dystrophy as a clinical feature of FSHD. A more recent study reports symptoms or signs of heart involvement in only 10 of 83 FSHD patients (12%) and were mainly represented by arrhythmic disturbances [427].

There is a growing body of literature that emphasizes miRNA mediated regulation of gene expression during myogenesis. The muscle specific miRNAs are known to be critical for proper skeletal and cardiac muscle development and function and have been associated with multiple myopathies such as hypertrophy, dystrophy and conduction defects [428]. There are miRNAs that are specifically expressed in FSHD individuals compared to the healthy control subjects. A recent study reported expression of miRNAs that potentially target about 40 genes [429]. Among which are miR-516b, miR-582 and miR-625 that target AKAP2, which is known to be involved in the cAMP signaling pathway in skeletal muscle. The ChIP-seq data analyzed in this thesis suggests a major association of SMCHD1 occupancy with miRNAs in general. There might be miRNAs that specifically target AKAP9 (Yotiao) through which SMCHD1 regulates cellular level of cAMP. In this regard, further research investigating SMCHD1 role in regulating expression of AKAP9 (Yotiao) would be informative.

Besides FSHD, two very recent publications have proposed key involvement of SMCHD1 in pathogenesis of a congenital disorder, Bosma arhinia and microphthalmia (BAMS) (Figure 5-1) [72, 334]. BAMS that is a rare condition and have only reported in 50 individuals thus far, involves absence of nose and reduced size of eyes and often accompanied with other abnormalities [336]. Using next generation sequencing, the two groups analyzed more than half of the known cases of BAMS and overall identified 24 missense mutations in SMCHD1 [72, 334]. Interestingly, unlike mutation identified in FSHD individuals, SMCHD1 mutations were localized over the ATPase domain and the adjacent region, presumably affecting ATPase activity. Gordon et al. applied Xenopus
Laevis as a model and showed that injection of human SMCHD1 mRNA which contains BAMS mutations induces abnormalities featuring arhinia [72]. In addition, biochemical analysis of ATPase activity in vitro suggested that SMCHD1 mutations would increase the ATPase activity. While Gordon et al. proposed gain of function mutation, studies by Shaw et al. suggested loss of function mutation in SMCHD1 [334]. Shaw et al. used Zebrafish as a model and showed that injection of wild type and not mutated human SMCHD1 rescue the arhinia features in zebrafish with knocked down levels of SMCHD1.

![SmcHD1 homodimer](image)

* Indicates similar mutation

**Figure 5-1. The schematic illustration of SMCHD1 homodimer and mutations identified in FSHD and BAMS individuals**

SMCHD1 mutation is FSHD individuals were distributed throughout SMCHD1 proteins while SMCHD1 mutations in BAMS individuals are localized in the ATPase domain.

The relationship between SMCHD1 mutations and the two very different developmental diseases is complex. FSHD2 individuals do not present facial abnormalities that are signatures of BAMS. On the other hand, BAMS individuals examined so far do not display muscular dystrophy. The latter might be due to the fact that these patients are examined at early stages in life and more time is required for FSHD progression. Among all the mutations identified in FSHD2 and BAMS so far, only one mutation (Gly137Glu) has been reported in both disorders, however individuals affected by this mutation do not display both symptoms. [334]. Only one individual has been reported thus far that display symptoms of both disorders and harbors a Asn139His mutation [430]. Nonetheless, DUX4 gene expression or loss of DNA methylation over 4q35 region has not been examined in this individual, raising the possibility that muscular dystrophy was misdiagnosed as FSHD.
Currently, it’s unknown how SMCHD1 mutations result in BAMS. Comparing the expression profile of BAMS individuals with SMCHD1 null mice neural stem cells (NSCs) suggests enrichment of numbers of downregulated genes [334]. Gene ontology pathway analysis suggested “depressed nasal tip” as the enriched phenotype which was driven by four genes: DOK7 (Docking protein 7), TGIF1 (TGFB induced factor Homeobox1), KDM6A (lysine demethylase 6A) and ICK (Intestinal cell kinase). These genes have been implicated in craniofacial development and mutations result in malformation of nasal bridges. In chapter 3 of this thesis, we investigated genome wide occupancy of SMCHD1 in SH-SY5Y cells. ChIP-seq data fails to suggest SMCHD1 occupancy over the differential expressed genes in BAMS. It is possible that SMCHD1 indirectly regulate expression of genes involved in craniofacial development. It would be interesting to investigate expression level of these gene in SH-SY5Y cells with knockout levels of SMCHD1.

Overall, the goal of this thesis was to delineate the role of SMCHD1 in regulating gene expression. To achieve this objective, my studies were separated into three sections. The first study involved the shRNA knockdown of SMCHD1 and measured global gene expression to identify genes regulated by SMCHD1. We found that SMCHD1 knockdown impacts 385 genes from which some are located in clusters on autosomes. These genes were characterized by monoallelic gene expression. The second study examined genome-wide occupancy of SMCHD1 in relation to global level of DNA methylation. We showed that SMCHD1 genomic binding was greatly affected by the loss of global DNA methylation. SMCHD1 genome occupancy associated with genes in the G-protein coupled receptor-signaling pathway and loss of SMCHD1 increased KCNQ1 protein expression. Interestingly, numbers of SMCHD1 binding sites identified in ChIP experiment correlated with differential expression pattern observed in the microarray including the imprinted genes located on the KCNQ1 locus. This might indicate direct roles of SMCHD1 on regulating gene expression over this locus. However, there were number of genes identified in the microarray experiment that lack direct binding of SMCHD1 such as the PCDH gene cluster. This might suggest the indirect regulatory roles of SMCHD1 in controlling gene expression. The third study explored SMCHD1
candidate protein binding partners. We suggested that SMCHD1 might be a subunit of transcriptional repressor complexes including Sin3A and CoREST.

Collectively, my PhD studies started by investigating how epigenetic modifications regulate expression of GH gene. SMCHD1 was shown to bind to methylated GH gene promoter. I followed our studies by examining the genes regulated by SMCHD1 as well as its genome wide occupancy. These studies suggest that SMCHD1 regulates genes in clusters and it majorly binds to intron and intergenic regions. In addition, genes were both up- and down regulated by loss of SMCHD1. These observations propose that SMCHD1 might regulates gene expression by mediating long-range chromatin interactions. It would be interesting to examine if SMCHD1 regulates monoallelic gene expression by mediating long-range chromatin interaction to simultaneously create transcriptionally active and repressed regions (Figure 5-2).
Figure 5-2. SMCHD1 functional model

A) Initially, reader and writer protein complexes add and recognize DNA and histone modifications. This results in recruitment of SMCHD1 (B). SMCHD1 mediates long-range chromatin interactions. (C) This creates two chromatin hubs, transcriptionally active and inactive regions. Transcriptionally repressed sites are compact chromatin regions that encompass hyper-methylated CpG sites while the transcriptionally active sites are loosely packed chromatin regions that are enriched in hype-methylated CpG sites.

Previous studies suggested that upon initial methylation of low number of CpG sites at early-stages of X chromosome inactivation, SMCHD1 mediates late-stage methylation of numbers of CpG sites. It is possible that writers and readers, the protein machinery that add and recognize DNA and histone modifications bind to and modify specific chromatin regions (Figure 5-2). This leads to recruitment of SMCHD1 and mediation of long-range chromatin interaction that generate transcriptionally active and repressed hubs. Over the transcriptionally inactive region, SMCHD1 induce hyper-methylation of CpG sites while the transcriptionally active chromatin, majorly encompasses hypo-methylated CpG sites.

Future research beyond this thesis may include assessing the relationship between allele specific binding of SMCHD1 and monoallelic gene expression. This would require using a cell system with a stable genome like human induced pluripotent stem cells (iPSCs) to determine if allele specific SMCHD1 occupancy correlates with the allelic exclusion. Instead of iPSCs one could use single cell analysis studies with mammalian cells. One of the major findings described in chapter 3 is the impact of SMCHD1 on KCNQ1 gene expression and the cellular level of cAMP. Moving forward, one could pursue further genetic and physiological studies to examine SMCHD1 role in regulating potassium channel activity and cAMP signaling in myoblast cells. In addition, SMCHD1 protein-protein interaction with potential binding partners and the activity of these protein complexes in regulating transcription could be assessed using traditional molecular biology techniques.
References


Appendix A.

**Supplementary Table S2-1**

**Description:**

The accompanying excel sheet provides the list of upregulated genes upon SMCHD1 shRNA knockdown in HEK293 cells.

**File name:**

Up-regulated genes Table S2-1

**Supplementary Table S2-2**

**Description:**

The accompanying excel sheet provides the list of downregulated genes upon SMCHD1 shRNA knockdown in HEK293 cells.

**File name:**

Down-regulated genes Table S2-2
Appendix B.

Supplementary Table S3-1

Description:

The accompanying excel sheet provides the list of motifs that associate with SMCHD1 genome wide occupancy.

File name:

SMCHD1 binding motifs Table S3-1

Supplementary Table S2

Description:

The accompanying excel sheet provides the list of genes that associate with SMCHD1 genome wide occupancy.

File name:

SMCHD1 binding associated genes Table S3-2

Supplementary Table S3-3

Description:

The accompanying excel sheet provides the list of twenty-one functional clusters generated by DAVID Bioinformatics resources which, associate with SMCHD1 genome occupancy. These clusters associate with 33 genes identified in the G-protein coupled receptor signaling pathway.

File name:

DAVID Annotation clusters Table S3-3