UNIVERSITY OF CALIFORNIA,
IRVINE

# Understanding the molecular and functional consequences of epigenome dynamics in cell fate, aging, and disease

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Biomedical Engineering

By

Julien Laurent Pierre Morival

Dissertation Committee:
Assistant Professor Timothy L. Downing, Chair
Assistant Professor Elizabeth Read
Associate Professor Chang Liu

2021

## DEDICATION

To

My loving wife, Chloé,
and our growing family.
For giving me the motivation to give it my all,
And grounding me when I needed it most.

To

My parents, Pierre and Nathalie
My sister, Camille
and the rest of my family,
For giving me the help and confidence
To accomplish anything I set my heart out to

To

My grandfather, Georges,
For teaching me to reflect on life
And the value of hard work

## TABLE OF CONTENTS

**CONTENTS**.....Pages
**LIST OF FIGURES**.....v
**LIST OF TABLES**.....vi
**ACKNOWLEDGEMENTS**.....vii
**VITA**.....viii
**ABSTRACT OF THE DISSERTATION**.....ix
**INTRODUCTION**.....1
**SECTION 1: Genome replication programs both cell fate and aging**
Introduction.....3
1.1.1. DNA replication plays a role in cell fate transitions.....3
1.1.2. Multicellular life and aging are intrinsically linked.....3
1.1.3. DNA methylation is temporally dynamic.....3
1.2. Results and Discussion.....4
1.2.1. Post-replication DNA remethylation kinetics create a transient window of epigenetic entropy.....4
1.2.2. Coordinated temporal dynamics across the epigenome point to a regulatory function of the DNA replication-associated transient window of entropy.....10
1.2.3. Slow remethylation kinetics may provide a prolonged window of time for increased gene expression variability, allowing for cell fate transitions.....15
1.2.4. The transient window of regulatory heterogeneity leaves the genome vulnerable to age-related epigenetic drift over an organism's lifetime.....20
1.3. Materials and Methods.....26
1.3.1. Replication-associated bisulfite sequencing (Repli-BS) datasets.....26
1.3.2. Temporally dynamic tile generation and binning.....26
1.3.3. Methylation entropy and read-level calculations.....27
1.3.4. Stochastic modeling of post-replication remethylation kinetics.....28
1.3.5. Annotations and downloaded datasets.....28
1.3.6. Transcription factor binding site (TFBS) enrichment and gene ontology (GO) analyses.....30
1.3.7. Replication-associated assay for transposase-accessible chromatin sequencing (Repli-ATAC) and insert size entropy calculations.....30
1.3.8. Statistical analyses.....32

**SECTION 2: DNA methylation analysis reveals epimutation hotspots in patients with dilated cardiomyopathy-associated laminopathies**
2.1. Introduction.....34
2.1.1. Lamin A/C in the nuclear envelope.....34
2.1.2. Lamins interact with DNA.....34
2.1.3. Laminopathies and Dilated Cardiomyopathy.....34
2.1.4. DNA methylation in *LMNA*-mutated DCM samples.....35
2.2. Results and Discussion.....35
2.2.1. Genome-wide DNA methylation analysis within family-specific primary fibroblasts and iPSCs.....35
2.2.2. Family-specific epigenetic signatures dominate DMR landscape at distal regulatory features and transcriptionally repressed chromatin in fibroblasts.....41
2.2.3. Fibroblast DMRs associate with distal regulatory features and transcriptionally repressed chromatin.....47

2.2.4. Fibroblast DMR-associated genes enrich for family-specific disease ontologies.....49
2.2.5. Genes dysregulated in both fibroblast and DCM cardiac tissues associate with DMRs from both families and LADs.....52
2.2.6. Reprogramming reveals epigenetic hotspots for aberrant methylation during early development.....57
2.3. Materials and Methods.....68
2.3.1. Fibroblast and iPSC cell lines.....68
2.3.2. RNA-sequencing (RNA-seq) and differentially expressed gene (DEG) analysis.....73
2.3.3. Reduced representation bisulfite sequencing (RRBS) and differentially methylated region (DMR) analysis.....73
2.3.4. Genomic feature annotation.....76
2.3.5. Identification of gene network and ontologies from DMR-associated gene lists.....76
2.3.6. Determining differentially methylated transcription factor binding sites (TFBS).....77
2.3.7. Lamina-associated Domain (LAD) redistribution analyses.....77
2.3.8. Statistical Analyses.....78

**Section 3**
**CONCLUSION**.....81
**FUTURE WORK**.....83
**REFERENCES**.....84
**APPENDIX SECTION 1**.....98
**APPENDIX SECTION 2**.....109

## LIST OF FIGURES

Figures.....pages

**Section 1**

Figure 1.1: Entropy and read-level analyses of replication-associated bisulfite sequencing (Repli-BS) data reveal a temporal window of epigenetic entropy.....7
Figure 1.2: Temporal methylation differences vary widely across 1Kb tile fragmented genome.....9
Figure 1.3: Remethylation rates and temporally dynamic tiles inversely correlate.....10
Figure 1.4: Temporal dynamics in chromosomal architecture and post-translational histone modifications associate with DNA remethylation kinetics at regulatory elements of the genome.....13
Figure 1.5: Accessibility and methylation entropy exponentially decays with time.....14
Figure 1.6: Temporally dynamic tiles associate with various genomic features.....15
Figure 1.7: Post-replication remethylation kinetics associate with gene expression variability and developmental elements.....18
Figure 1.8: Remethylation kinetics inversely correlates with single cell RNA sequencing (scRNA-seq) gene expression variability.....20
Figure 1.9: CpG density and context combined with slow remethylation kinetics affect CpG susceptibility to age-related epigenetic drift.....23
Figure 1.10: Whole genome bisulfite sequencing (WGBS) methylation data for newborn and centenarian samples.....25

**Section 2**

Figure 2.1: Characterization of DNA methylation in *LMNA*-mutant fibroblasts and iPSCs..37
Figure 2.2: Quantification of captured CpGs and corresponding DNA methylation by family.....39
Figure 2.3: Computational workflow of DNA methylation analyses.....40
Figure 2.4: Hypermethylated and hypomethylated DMRs localize at distal regulatory features and transcriptionally repressed chromatin in fibroblasts.....43
Figure 2.5: Sex chromosomes had minimal impact on genome-wide and DMR results.....45
Figure 2.6: Clustering of differentially methylated regions (DMRs) by DNA methylation level.....46
Figure 2.7: DMRs associate to dysregulated and disease-relevant genes near redistributed LADs.....51
Figure 2.8: Inter-DMR distances overlap across both families in all chromosomes.....56
Figure 2.9: Chromosomal distribution of conserved DMR-associated differentially expressed genes (DEGs) near redistributed LADs.....57
Figure 2.10: DMRs in iPSCs reveal tissue-persistent epimutation hotspots at developmentally and laminopathy relevant genes.....61
Figure 2.11: DMR-associated genes overlap in iPSCs and fibroblasts.....63
Figure 2.12: Validation of induced pluripotent stem cell (iPSC) pluripotency.....72
Figure 2.13: Validation of normal chromosome constitution in each induced pluripotent stem cell clone.....73

## LIST OF TABLES

Tables..... pages

**Section 2**

Table 2.1: Fibroblast and iPSC line pairs with corresponding genotype, sex, and age when skin biopsies were performed (N = 10)..... 68

Table 2.2: Table of antibodies used for pluripotency characterization of iPSCs..... 70

## ACKNOWLEDGEMENTS

I would like to first of all thank my wife, Chloé, who has been there by my side since our early days in undergraduate. She has been understanding and supportive every step of the way during this PhD, giving me the confidence to be the best that I could. With that, I also want to say “un grand merci!” to my parents, sister, and my family in France, who has always given me their unconditional trust and support in all my endeavors, despite it keeping me away from them across the Atlantic. Along with them, I also thank the family here that I have gained through marriage, and who took me in without a second thought.

I would like to acknowledge my committee chair, Dr. Timothy L. Downing, for giving me the tools and guidance to become the independent researcher that I am today. Seeing first-hand what it takes to start a new lab and dream about its potential has been truly humbling. Along with that, a big thank you to all of the current and previous Downing lab members. Coming in to work to such a fantastic and knowledgeable group of people every day truly makes it all worth it. And a special shout-out to Navied Akhtar, Annie Trinh, and Nandor Laszik for their help with getting section 1 of this dissertation together. Thank you as well to all the members of the Edward’s Lifesciences Center for Advanced Cardiovascular Technology for creating such a collaborative and friendly environment to work in. I also want to thank the friends I’ve made during my time here at UCI, you provided a lot of support and relief during the most stressful times, and helped make southern California feel like home to me and Chloé.

I would like to acknowledge my defense committee, Drs. Elizabeth Read and Chang Liu, for taking the time to evaluate this dissertation and providing valuable feedback. I also want to thank the wonderful teams I have had a chance to collaborate with, these include Drs. Michael Zaragoza and Lily Widyastuti, Dr. Elizabeth Read, Honglei Ren, and Dr. Qing Nie. Their contributions were invaluable for putting together the work presented in this dissertation.

I would also like to thank the funding sources that have supported me during my time at the University of California, Irvine. These include The National Science Foundation and The Simons Foundation. As well as on-campus centers that provided the facilities and resources to perform the work presented, including the NSF-Simons Center for Multiscale Cell Fate Research, the Center for Complex Biological Systems, the Edward’s Lifesciences Center for Advanced Cardiovascular Technology, the Genomics High Throughput Facility, and the RCIC. In addition, the work in this dissertation was supported by an Opportunity Award from the UCI Center for Complex Biological Systems, and funded in part by UCI’s School of Medicine Systems Pathology Initiative (Fund 60242), UCI’s NSF-Simons Center for Multiscale Cell Fate Research grants NSF DMS1763272 and a Simons Foundation grant (594598, QN), and grant 1R01HL129008 from the NIH National Heart, Lung, and Blood Institute (MVZ).

## VITA

### Julien Laurent Pierre Morival

### EDUCATION

**Ph.D.** **University of California, Irvine**, Irvine, CA — August 2021
Biomedical Engineering
*Advisor:* Timothy L. Downing, Ph.D.

**M.S.** **University of California, Irvine**, Irvine, CA — March 2018
Biomedical Engineering

**B.S.** **The Johns Hopkins University**, Baltimore, MD — May 2015
Biomedical Engineering (cell and tissue specialization)
*General and Departmental Honors*

---

### PUBLICATIONS

#### Published

**Morival, J.L.P.**, Widyastuti, H.P., Nguyen, C.H.H. *et al.* (2021). DNA methylation analysis reveals epimutation hotspots in patients with dilated cardiomyopathy-associated laminopathies. *Clin Epigenet*, 13(139)

Busto-Moner, L., **Morival, J.**, Ren, H., Fahim, A., Reitz, Z., Downing, T.L., Read, E.L. (2020). Stochastic modeling reveals kinetic heterogeneity in post-replication DNA methylation. *PLOS Computational Biology,*16(4)

Chu, M., Nguyen, T.T., Lee, Eugene E.K., **Morival, J.L.**, Khine, M. (2017). Plasma free reversible and irreversible microfluidic bonding. *Lab on a Chip*, 17(2), 267-273

---

### AWARDS

Center Fellow Award at NSF-Simons Center for Multiscale Cell Fate Research (UC Irvine), 6-month stipend funding, 2020-2021

Opportunity Award Recipient at the Center for Complex Biological Systems (UC Irvine), *Targeted epigenetic tools for the control of variability in cardiac differentiation of patient-derived iPSCs*, in collaboration with Halida Widyastuti, Ph.D. at the Zaragoza Lab (UC Irvine), $10,000, 2019-2020

Center Fellow Award at NSF-Simons Center for Multiscale Cell Fate Research (UC Irvine), 6-month stipend funding, 2019-2020

Opportunity Award Recipient at the Center for Complex Biological Systems (UC Irvine), *Methylation pseudotime: a novel tool for observing protein-DNA interactions*, in collaboration with Adam McLean, Ph.D. at the Nie Lab (UC Irvine), $10,000, 2018-2019

Center Fellow Award at NSF-Simons Center for Multiscale Cell Fate Research (UC Irvine), 6-month stipend funding, 2018-2019

Best Poster Presentation Award at UC Systemwide Bioengineering Symposium, June 2018

## ABSTRACT

Understanding the molecular and functional consequences of epigenome dynamics in cell fate, aging, and disease

By

Julien Laurent Pierre Morival

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2021

Professor Timothy L. Downing, Chair

DNA replication plays an important part in allowing cells to proliferate and develop into complex tissues. The advent of multicellular organisms, however, has been theorized to be intertwined with the tradeoff of aging and disease. These events are highly associated with drastic changes in gene expression across a cell population, often regulated by the epigenome. The set of heritable modifications that make up the epigenetic landscape are known to be altered by cell fate, aging, and disease. However, the dynamic processes by which the changes in the epigenome, and subsequently transcriptome, lead to these modified cell states are not clearly understood. In this dissertation, we demonstrate that DNA replication leads to a transient window of epigenetic entropy, providing the first evidence of a molecular link between cell fate, aging, and disease. In order to elucidate this link, we made use of replication-associated bisulfite sequencing (Repli-BS) and replication-associated assay for transposase-accessible chromatin sequencing (Repli-ATAC) datasets in human embryonic stem cells (hESCs). Our results suggest that the temporality of this window for both the chromatin architecture and DNA methylation differs across the genome. Specifically, we identified that the regions with the most prolonged window of epigenetic entropy are located at regulatory features, associate with expression variability, and are susceptible to age- and disease-related epigenetic drift. Additionally, this dissertation explores the impact of individual *LMNA* mutations on the

epigenome that lead to unique disease outcomes of dilated cardiomyopathy (DCM) and brachydactyly using patient-derived fibroblasts and induced pluripotent stem cells (iPSCs). Analyses combining multiple epigenetic features and transcriptomic data suggest that differentially methylated regions (DMRs) are associated with the misregulation of regulatory elements, and that, in combination with chromatin remodeling, could lead to gene dysregulation ending in DCM. Ultimately, our results provide evidence that somatic and reprogrammed patient cells could serve as models to understand the mechanism behind which disease-related regulatory abnormalities lead to laminopathies like DCM and brachydactyly.

## INTRODUCTION

DNA replication allows for the faithful inheritance of genetic information from one cell generation to the next, giving way to proliferation. In certain eukaryotic organisms, proliferative events like asymmetric division allow for multicellular life to develop. This mechanism, by which stem cells simultaneously self-proliferate and give rise to a differentiated daughter cell, is accompanied by unique changes in gene expression which help define a new cell state[1]. The central regulating mechanism of gene expression in cells is the epigenome, a group of modifications that affect genes without modifying the genetic sequence. These modifications are inherited from parental to daughter strands through maintenance enzymes. The epigenome operates at the chromatin (chromatin architecture, hetero- vs. euchromatin), the nucleosome (histone post-translation modifications), and the DNA (DNA methylation). Due to their involvement in gene expression, it is not surprising that events like cell fate have been associated with a modified epigenome[1]. Over the course of multiple cycles and mutation events, however, a cell's ability to correctly perform cellular functions can degrade[2], resulting in an altered epigenome and ultimately replicative aging[3,4] and disease[5,6]. The underlying mechanism by which the epigenome dynamically changes and allows for these modified cell states to arise is not fully understood.

In order to explore this problem, we mainly focused on DNA methylation, as it is a highly characterized epigenetic modification in cell fate[1,7], aging[4,8], and disease[9,10]. In mammals, this modification, consisting of a methyl ($CH_3$), is added to cytosines at CpG dinucleotide locations by enzymes, which copy methylation from parental strands to daughter strands. Its presence has directly been linked to changes in protein binding to the DNA, as well as a direct correlation with gene inhibition. Using this known epigenetic modification, we aim to elucidate the mechanism by which DNA methylation, in conjunction with other epigenetic features, can give rise to unique cellular events like multicellular organisms, aging, and disease. Ultimately, understanding the underlying role that the epigenome plays in allowing for these

dynamic processes to take place could prove to be essential to understanding how to modulate or negate them.

# SECTION 1

# Genome replication programs both cell fate and aging

## 1.1 Introduction

### 1.1.1 DNA replication plays a role in cell fate transitions

All living systems utilize DNA replication as a means to proliferate and increase population size. During this process, the genetic and epigenetic codes are dismantled, copied, and faithfully re-established in both parental and daughter cells, as part of the reliable maintenance of cell identity. In certain eukaryotic organisms, replication can also bring about the rise of multicellular life, associated with drastic changes in the transcriptome and epigenome across the cell population[1]. Understanding how this change is initiated, during a cell's replication, has been a key point of interest for developmental research[11]. Recently, the rise of single-cell technology has revealed the presence of a previously unappreciated molecular variability across cell populations taking place at the proteomic[11,12], transcriptomic[13], and epigenetic level[14]. This intrinsic regulatory noise has been suggested to be a potential source for explaining how seemingly homogeneous cell populations can give rise to a multitude of cell types over the course of several cell divisions[11].

### 1.1.2 Multicellular life and aging are intrinsically linked

The advent of complex organisms, however, has been theorized to be evolutionarily intertwined with the tradeoff of aging and disease, as a means of regulating resource demand and therefore population size[15]. Similar to cell fate transitions, aging also leads to transcriptomic[16] and epigenetic changes[17]. The molecular and functional mechanisms that connect and allow for both cell fate transition and aging to take place on such different timescales (days vs a lifetime) still remain unclear.

### 1.1.3 DNA methylation is temporally dynamic

Cytosine methylation, a highly conserved epigenetic modification across DNA replication, has been shown to be variable in cell populations at regulatory domains[18], and also across multiple cell generations[15]. As such, DNA methylation has been implicated in stem cell differentiation[19–21], aging[4,22,23], and the emergence of age-related diseases [3,24,25]. Although originally attributed to cell-to-cell heterogeneity, we have previously shown that much of the observed DNA methylation heterogeneity is actually due to a global delay in post-replication maintenance of this epigenetic mark[26].

We hypothesize that the temporal re-establishment of epigenetic marks, initiated by replication, could have a role in both creating regulatory noise needed for cell fate transitions to take place and for age-related epigenetic drift to arise. To explore this, we investigated the post-replication landscape of epigenetic modifications and chromatin architecture in a human embryonic stem cell (hESC) line. We provide the first direct evidence of a molecular framework that describes the co-dependency of multicellular life with mechanisms of aging.

## 1.2 Results and Discussion

### 1.2.1 Post-replication DNA remethylation kinetics create a transient window of epigenetic entropy

Quantification of the modification's genome-wide stochasticity through normalized methylation entropy (NME) revealed a gradual decrease in NME across timepoints, eventually reaching bulk levels (Figure 1.1A). This indicated the presence of a previously unappreciated transient window of time during which methylation entropy is elevated. To explore this temporary heterogeneity, we focused our analyses on the two most extreme changes in methylation levels (0hr and 16hr timepoints). Genome-wide, differences in methylation between the two timepoints appeared to vary considerably based on the region of interest (Figure 1.1B). Breaking the genome up into 1Kb tiles further revealed that this temporal difference was inconsistent across the tiles (mean methylation difference: 33.25 ± 14.39) (Figure 1.2). In order to capture local regions where

substantial methylation differences were most prevalent, we further refined our analysis to custom region tiles. Tiles were generated by grouping only CpGs showing an increase in methylation over time, and were separated into decile bins (D1-D10) of increasing average methylation difference across each tile (Figure 1.1C). Tiles with the largest differences in methylation over time (D10) were referred to as replication-associated differentially methylated regions (Repli-DMRs). Although efficient to get general genomic trends of remethylation, Repli-DMRs are limited due to data scarcity and tile requirements, thus leaving some CpGs from being taken into account (Figure 1.1C gray shading). In order to achieve CpG-specific resolution, we made use of previously established kinetic rate parameters that numerically reflect the speed at which individual cytosines achieve steady-state methylation levels after replication[27]. Despite greater CpG coverage, rates appeared to confirm the temporal remethylation delay in our generated tiles, as the two were found to inversely correlate, with the slowest rates found in Repli-DMRs (Figure 1.3).

NME results informed of the presence of a temporal heterogeneity across the tiles, the implication that this may have across a cell population cannot be fully appreciated using averages across reads at individual CpGs (Figure 1.1D). We therefore decided to perform read-level analyses to resolve how this temporal heterogeneity presented on an inter-cellular level. We first calculated the proportion of discordant reads (PDR), interpreted as the fraction of cells with locally disordered methylation at each CpG[28]. In agreement with our NME data, mean PDR decreased over time (Figure 1.1E), suggesting that following the re-establishment of methylation, reads become more homogeneously methylated throughout. Interestingly, within dynamic tiles, PDR was found to be significantly lower (1-way ANOVA post hoc Tukey: $p < 2.2 \times 10^{-16}$) in temporally dynamic tiles from D8-D10, compared to other groups of tile bins. This points to the fact that reads had more consecutive methylation compared to other regions with less drastic changes in methylation over time.

Although insightful, PDR remains limited in its ability to identify the degree of disorder in methylation across a read (Figure 1.1D). Understanding the way in which the methylation pattern presents itself temporally across a cell population could have important functional consequences on gene expression regulation. We therefore made use of the transition score calculation[26], which determines the number of transitions in methylation state that take place along a read between neighboring CpGs. This measurement can more clearly distinguish between a “cell state” and “random” pattern of methylation along reads. To do so, we generated a set of synthetic transition score distributions, modeling a population of reads with either a “randomized” or “cell state” methylation pattern (Figure 1.1D), and compared it to each deciles’ distribution using Jensen-Shannon divergence (JSD). JSD was significantly lower (student t-test: $p \leq 0.033$) for the “cell state” comparison for all groupings of temporally dynamic tiles (Figure 1.1F), indicating that following replication, methylation transiently takes on this pattern across the cell population. This agrees with a stochastic state that allows for the development of multicellular tissues from seemingly homogenous cell populations[11].

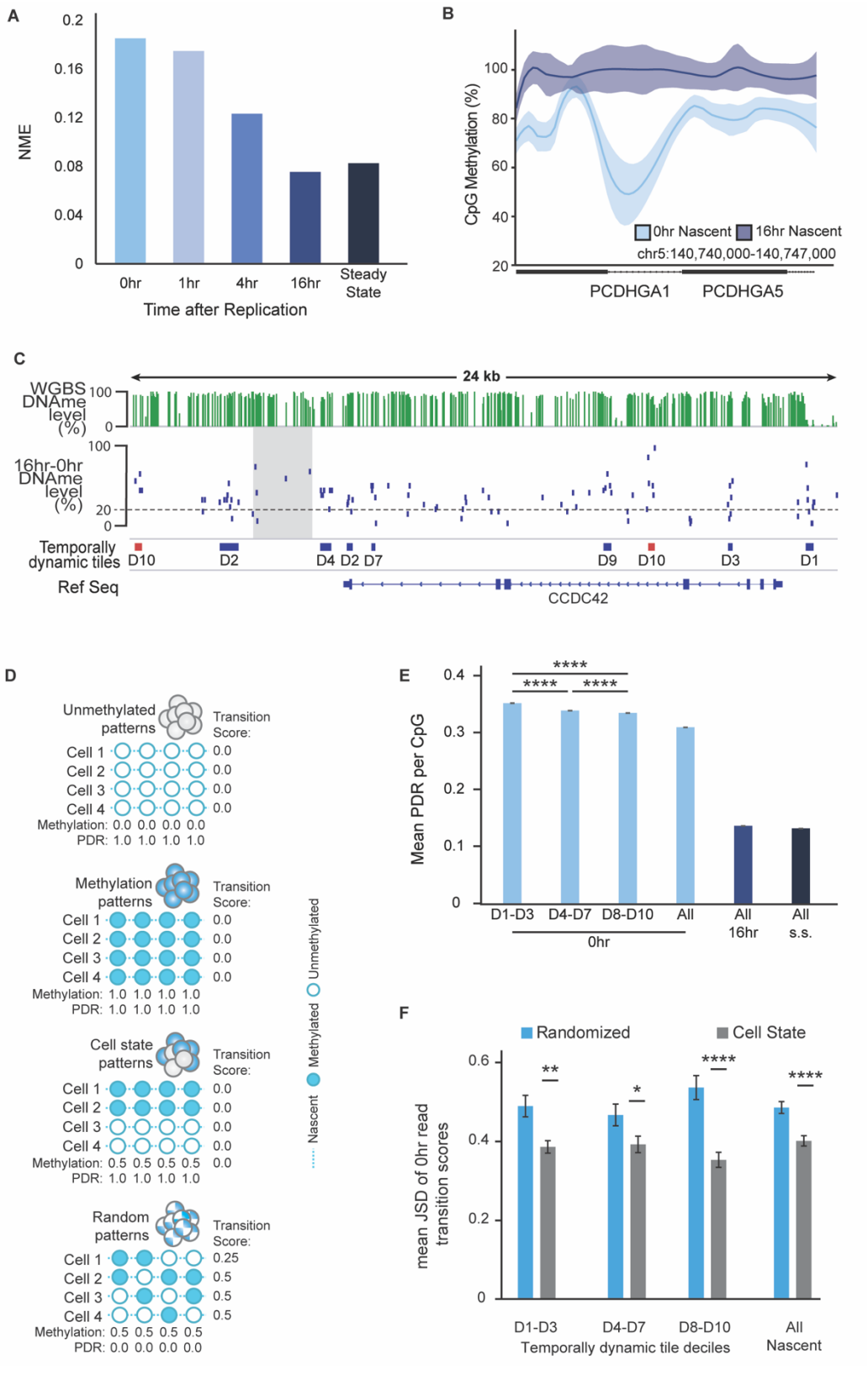
A
NME
0.2
0.16
0.12
0.08
0.04
0
0hr
1hr
4hr
16hr
Steady State
Time after Replication
B
CpG Methylation (%)
100
80
60
40
20
0hr Nascent
16hr Nascent
chr5:140,740,000-140,747,000
PCDHGA1
PCDHGA5
C
24 kb
WGBS DNAme level (%)
100
0
16hr-0hr DNAme level (%)
100
20
0
Temporally dynamic tiles
D10
D2
D4 D2 D7
D9
D10
D3
D1
Ref Seq
CCDC42
D
Unmethylated patterns
Transition Score:
Cell 1 0.0
Cell 2 0.0
Cell 3 0.0
Cell 4 0.0
Methylation: 0.0 0.0 0.0 0.0
PDR: 1.0 1.0 1.0 1.0
Methylation patterns
Transition Score:
Cell 1 0.0
Cell 2 0.0
Cell 3 0.0
Cell 4 0.0
Methylation: 1.0 1.0 1.0 1.0
PDR: 1.0 1.0 1.0 1.0
Cell state patterns
Transition Score:
Cell 1 0.0
Cell 2 0.0
Cell 3 0.0
Cell 4 0.0
Methylation: 0.5 0.5 0.5 0.5
PDR: 1.0 1.0 1.0 1.0
Random patterns
Transition Score:
Cell 1 0.25
Cell 2 0.5
Cell 3 0.5
Cell 4 0.5
Methylation: 0.5 0.5 0.5 0.5
PDR: 0.0 0.0 0.0 0.0
Methylated
Unmethylated
Nascent
E
Mean PDR per CpG
0.4
0.3
0.2
0.1
0
****
****
****
D1-D3
D4-D7
D8-D10
All
All 16hr
All s.s.
0hr
F
Randomized
Cell State
mean JSD of 0hr read transition scores
0.6
0.4
0.2
0
**
*
****
****
D1-D3
D4-D7
D8-D10
All Nascent
Temporally dynamic tile deciles

**Figure 1.1: Entropy and read-level analyses of replication-associated bisulfite sequencing (Repli-BS) data reveal a temporal window of epigenetic entropy. A.** Barplot showing the normalized methylation entropy (NME) for Repli-BS methylation data from timepoints collected following replication. **B.** Top, Genome browser track (chr5:140,740,000-140,7407,000) displaying a smooth average curve fitting for 0hr (light blue) and 16hr (dark blue) CpG methylation percentage from Repli-BS data. Bottom, Depiction of RefSeq gene annotation. **C.** Genome browser track (chr17:8,625,887-8,650,300) showing Top, Barplot of whole genome bisulfite sequencing (WGBS) methylation percentage, Middle Top, Scatter plot of methylation difference (16hr minus 0hr) for Repli-BS data. Dashed line indicates 20% methylation, the minimum methylation required for a methylation value to be considered to generate temporally dynamic tiles. Gray region contains CpGs that could not be captured in temporally dynamic tiles. Middle Bottom, Location of tiles generated using Repli-BS data, and separated into ten decile bins (D1-D10). Red tiles represent Repli-DMRs, the tiles with the largest difference in methylation over time. Bottom, Depiction of RefSeq gene annotation. **D.** Schematic depicting CpGs (blue circles) on nascent DNA (blue lines) either methylated (filled blue circles) or unmethylated (empty circles) in four theoretical models of methylation across a cell population. Below each CpG is the mean methylation and proportion of discordant reads (PDR) values per CpG. Each row indicates a read from a different cell in the population, along with its corresponding transition score. **E.** Barplot showing the mean PDR per CpG calculated across reads in the 0hr timepoint (light blue) either only at particular groups of temporally dynamic tile deciles or across all CpGs, 16hr timepoints (dark blue), or steady state (s.s., black). 1-way ANOVA post hoc Tukey test: **** $P \leq 0.0001$. **F.** Barplot of the mean Jensen-Shannon Distance (JSD) between the transition score from 0hr timepoint reads of samples and a synthetic dataset from the “randomized” or “cell state” methylation pattern models. This was performed at decile groups of temporally dynamic tile and all CpGs. Student t-test: * $P \leq 0.05$, ** $P \leq 0.01$, **** $P \leq 0.0001$.

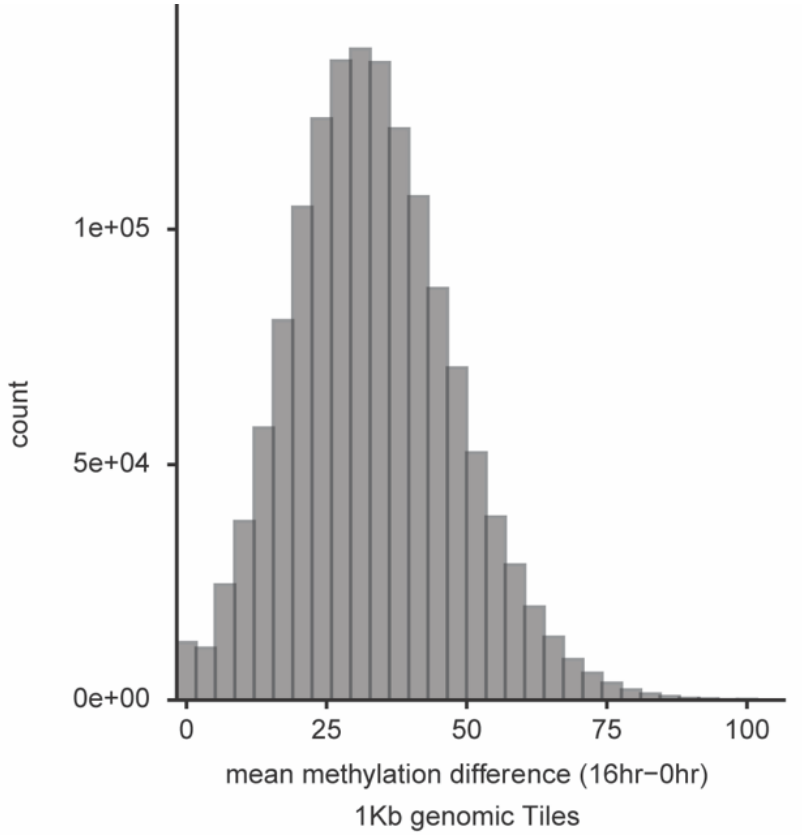


**Figure 1.2**: **Temporal methylation differences vary widely across 1Kb tile fragmented genome.** Histogram of mean methylation differences (16hr minus 0hr) from Repli-BS data at 1,000 bp (1Kb) tiles generated across the genome.

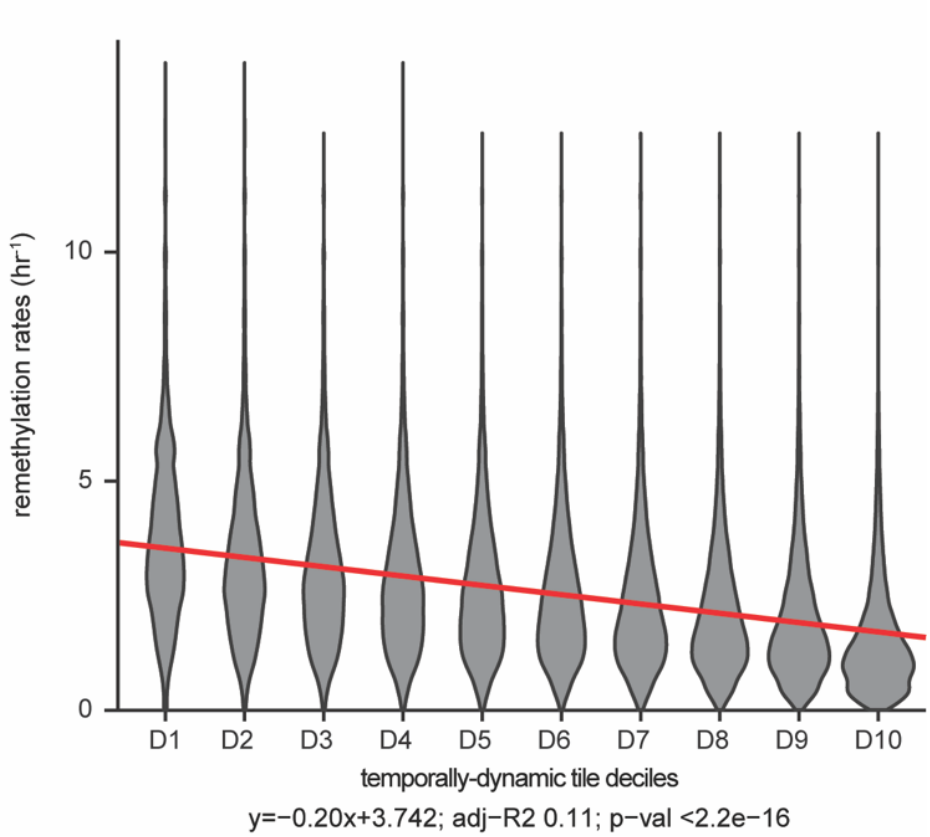


**Figure 1.3: Remethylation rates and temporally dynamic tiles inversely correlate.** Violin plots showing the distribution of remethylation rates captured in the different temporally dynamic tile deciles. Linear fit of the data is shown as a red line, with the corresponding formula and statistics shown at the bottom.

### 1.2.2 Coordinated temporal dynamics across the epigenome point to a regulatory function of the DNA replication-associated transient window of entropy

The presence of a transient window of inter-cellular methylation entropy can have important consequences in the context of regulatory function. However, it is also important to note that the epigenome operates on multiple levels, and that these are interconnected. Indeed, several studies have also demonstrated, through single cell or newly-developed sequencing techniques, that the chromatin is both disrupted by the replication fork[29–31], and that nucleosome occupancy is inversely correlated with DNA methylation[32,33]. In order to

investigate if post-replication chromatin accessibility operates on a similar time scale as our DNA methylation results, we performed an altered form of Repli-ATAC-seq[29], enabling us to capture reads representative of integer multiples of nucleosomes in hESCs over the same timecourse as our Repli-BS data (Figure 1.4A). Density plots of Repli-ATAC-seq insert size confirmed the expected nucleosome compaction periodicity observed from ATAC-seq[34]. Over time, though, a shift from less compact nucleosomes to a higher density of compacted chromatin emerged (median: 0h: 476.63, 1h: 490.78, 4h: 513.29, 16h: 604.18) (Figure 1.4B). Furthermore, accessibility entropy (replication-associated entropy minus sample background entropy) of insert size decreased over time (Figure 1.4C), mirroring our observations in DNA methylation (Figure 1.1A and 1.5). Repli-DMRs were additionally found to be strongly enriched for DNase I hypersensitivity sites (DHS) (Figure 1.4D), further confirming a possible coordination in post-replication remodeling of the chromatin architecture and DNA remethylation kinetics.

Interestingly, intersection of Repli-DMRs showed a fold enrichment for several epigenetic-modifying enzymes chromatin immunoprecipitation sequencing (ChIP-seq) peaks (Figure 1.4D). CTCF, a methylation-sensitive protein[35] involved in controlling chromatin architecture, was strongly associated with Repli-DMRs, indicative of a possible mechanism by which slow remethylation kinetics could account for dynamic changes in accessibility. Additionally, while EZH2, known for inhibitory H3K4me27 deposition, was only slightly enriched in Repli-DMRs, while P300, a histone acetyltransferase, had a 4-fold higher log odds ratio. This observation is particularly interesting as EZH2 is intrinsically linked to DNA methylation as part of the repression machinery[36], whereas histone acetylation deposition by P300 is typically associated with transcriptional activation[37]. We decided to elucidate the relationship of DNA remethylation kinetics with histone post-translational modifications, as these have also been found to undergo cyclic changes in levels[30] with different rates of recovery[31,38,39]. Intersection of our temporally dynamic tile deciles with ChIP-seq peaks for several histone marks revealed that some were represented across all tiles (H3K27Ac, H3K27me3, H3K4me3), while others had a clear

bias toward smaller (H3K36me3, H3K9me3) or larger (H3K4me1) temporal methylation differences (Figure 1.4E). To understand the significance of temporal methylation heterogeneity on histone marks, we decided to focus our analysis on regions of the genome where the window of inter-cellular methylation heterogeneity/entropy was most prolonged, namely Repli-DMRs. Intersection of Repli-DMRs with these histone marks revealed the highest enrichment for H3K4me1 (avg: 0.35), both uniquely and overlapped with either H3K4me3 or H3K27me3 (Figure 1.4F). Interestingly, H3K4me1 sites were previously found to have high heterogeneity and oscillations of DNA methylation in primed ESCs[40]. Consistent with our observations in histone modifications, Repli-DMRs enriched for enhancers, specifically hESC-specific non-super enhancers (Figure 1.4G). Interestingly, promoters were found to be enriched in both D1 and Repli-DMR (D10) tiles (Figure 1.6). This observation prompted us to question if the subgroups of promoters on either end of the epigenetic temporal spectrum had unique functions. Bivalent promoters, traditionally marked by both H3K27me3 and H3K4me3, were previously identified at developmental genes when the latter modification was cell cycle-regulated in hESCs[30]. Overlap of promoter regions with remethylation rates revealed that this subclass of cell cycle-regulated bivalent promoters had significantly slower kinetics (1-way ANOVA post hoc Tukey test: $p < 2.2 \times 10^{-16}$), in comparison to other forms of promoters (Figure 1.4H). Furthermore, bivalency has been theorized to be brought on by the co-occupancy of repressive H3K27me3 and P300[37], which could explain the large enrichment of the acetyltransferase in Repli-DMRs (Figure 1.4D). The presence of a prolonged window of epigenetic heterogeneity at regulatory elements known to be associated with development suggests that this transient state may play a part in regulating cell fate.

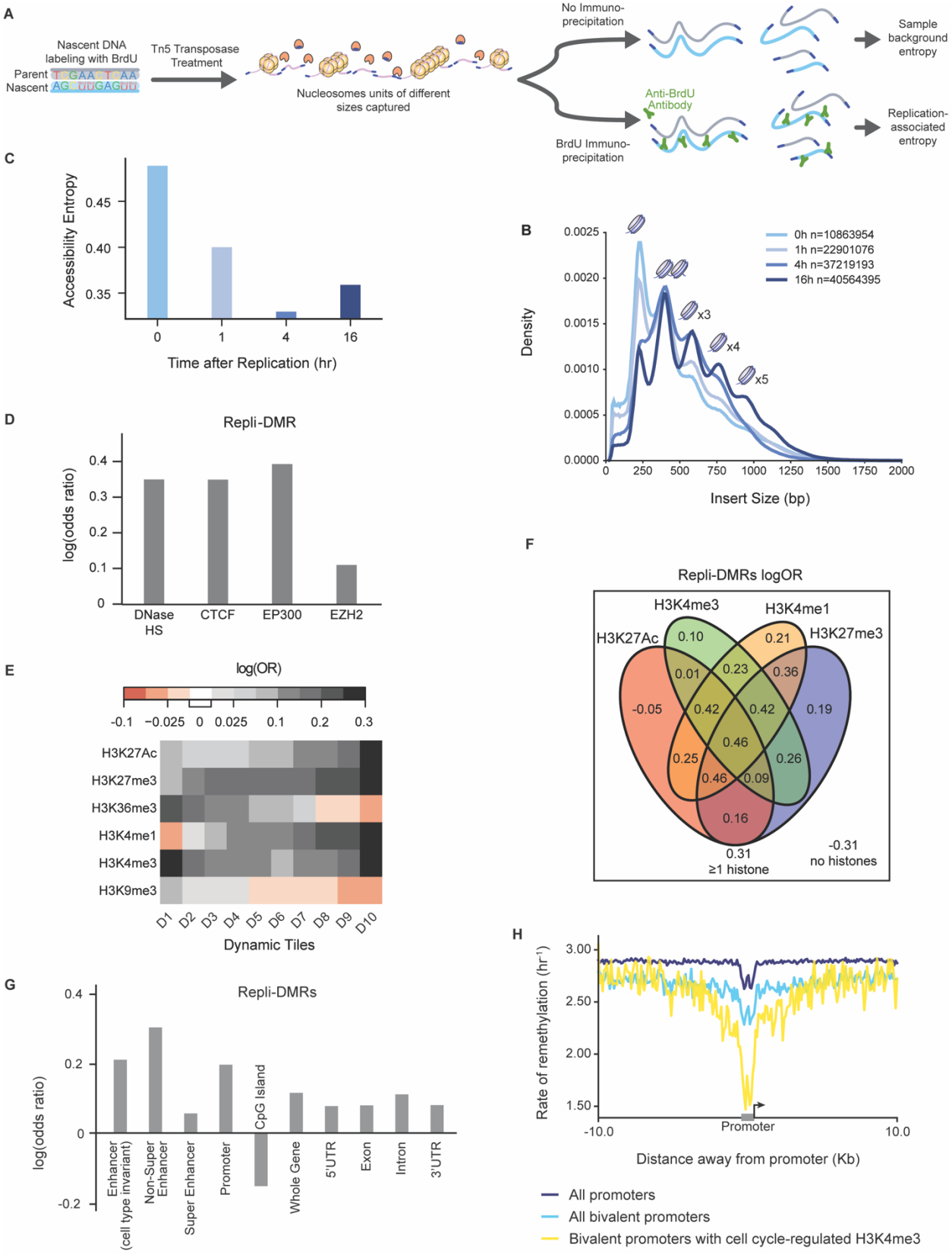
A
Nascent DNA labeling with BrdU
Parent
Nascent
Tn5 Transposase Treatment
Nucleosomes units of different sizes captured
No Immuno-precipitation
Sample background entropy
Anti-BrdU Antibody
BrdU Immuno-precipitation
Replication-associated entropy
B
Density
0h n=10863954
1h n=22901076
4h n=37219193
16h n=40564395
x3
x4
x5
Insert Size (bp)
C
Accessibility Entropy
Time after Replication (hr)
D
Repli-DMR
log(odds ratio)
DNase HS
CTCF
EP300
EZH2
E
log(OR)
H3K27Ac
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me3
Dynamic Tiles
F
Repli-DMRs logOR
H3K4me3
H3K4me1
H3K27Ac
H3K27me3
≥1 histone
no histones
G
Repli-DMRs
log(odds ratio)
Enhancer (cell type invariant)
Non-Super Enhancer
Super Enhancer
Promoter
CpG Island
Whole Gene
5'UTR
Exon
Intron
3'UTR
H
Rate of remethylation (hr⁻¹)
Promoter
Distance away from promoter (Kb)
All promoters
All bivalent promoters
Bivalent promoters with cell cycle-regulated H3K4me3

**Figure 1.4: Temporal dynamics in chromosomal architecture and post-translational histone modifications associate with DNA remethylation kinetics at regulatory elements of the genome A.** Schematic showing the methodology of replication-associated assay for transposase-accessible chromatin sequencing (Repli-ATAC-seq) and the expected entropy outputs. **B.** Density distributions of insert sizes captured in Repli-ATAC-seq for different timepoints after replication (0hr, 1hr, 4hr, and 16hr). The number of compacted nucleosomes is depicted above each of their corresponding peaks. **C.** Barplot showing accessibility entropy across each of the timepoints. **D.** Barplot showing the log odds ratio enrichment of DNase hypersensitivity sites (DNase HS) and epigenetic-modifying proteins in Repli-DMR tiles. **E.** Heatmap showing the log odds ratio enrichment of histone modifications in temporally dynamic tiles from each decile group. **F.** Venn diagram showing the log odds ratio (logOR) enrichment of unique, overlapping, no histone modifications in Repli-DMR tiles. **G.** Barplot showing the log odds ratio enrichment of genomic features in Repli-DMRs. **H.** Smooth median curve fitting for remethylation rates at and within ±10Kb of all promoters (dark blue), bivalent (light blue), and cell cycle-regulated H3K4me3 bivalent promoters (yellow).

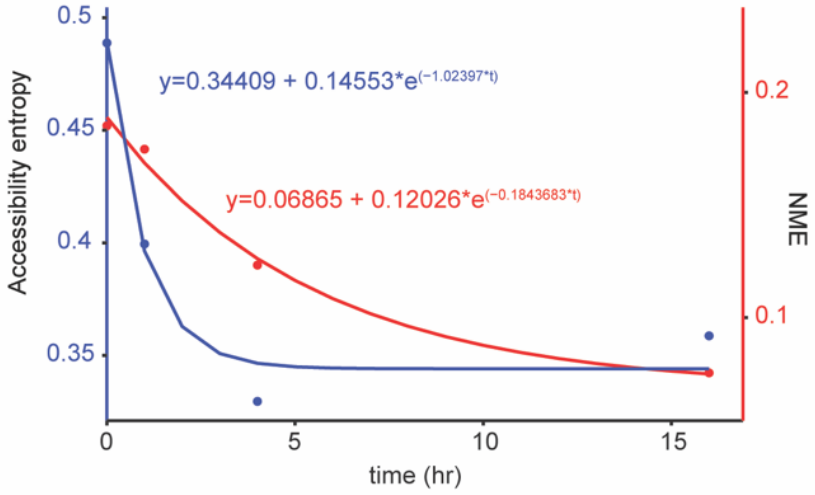

**Figure 1.5 Accessibility and methylation entropy exponentially decays with time.** Non-linear fitting for accessibility entropy (dark blue) and normalized methylation entropy (NME, red) across Repli-BS timepoints.

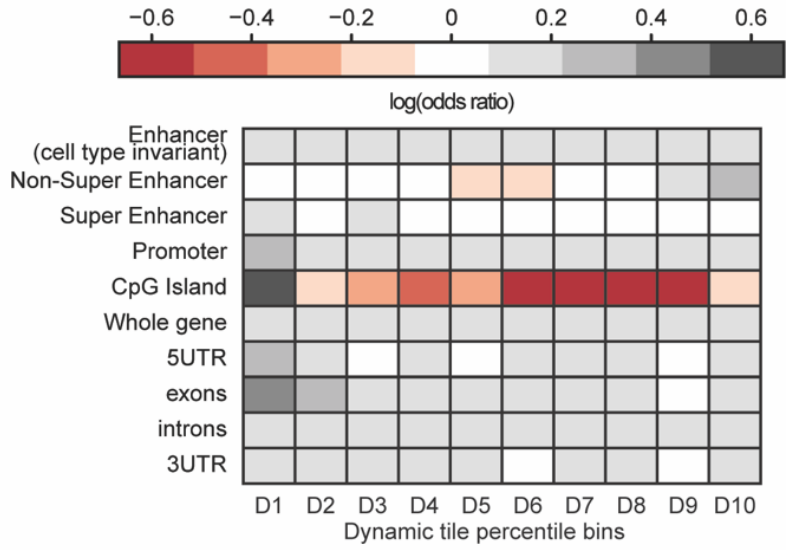


**Figure 1.6 Temporally dynamic tiles associate with various genomic features**. Heatmap showing the log odds ratio enrichment of genomic features in temporally dynamic tiles from each decile group.

### 1.2.3 Slow remethylation kinetics may provide a prolonged window of time for increased gene expression variability, allowing for cell fate transitions

Epigenetic memory, in other words, the faithful inheritance of epigenetic marks from parental to daughter cells, plays an important role in maintaining the transcriptional state of cells[41]. It is therefore unsurprising that previous studies have noted that the epigenetic memory of the transcriptional state also gets disrupted by replication[42]. Considering that the epigenome of silenced regions may be transiently heterogeneous at regulatory features, we theorized DNA may be left temporarily vulnerable to transcription factor (TF) binding, which could lead to further cellular changes. TF binding site analysis of Repli-DMRs revealed the presence of several

development-regulating transcription factor families, including POU, FOX, GATA, and HOX (Figure 1.7A, Appendix 1.1). Gene ontology (GO) term analysis of the top 20 most significant TFs confirmed their association with cell fate, development, and transcription regulation (Appendix 1.2). Seeing the enrichment of these TFs in our Repli-DMRs, regions with prolonged epigenetic heterogeneity that associate with regulatory features, we next decided to investigate if gene expression could be impacted. Transcriptional noise, fluctuations in gene expression in cells, has recently been identified as an important tool for stem cells to undergo specific cell fate specification[11,43]. In order to determine if the identified window of epigenetic heterogeneity could account for some of these increased fluctuations in gene expression, we measured gene expression variability across single cells, using previously published scRNA-seq data in hESC[44]. We observed that gene expression variability had an inverse relationship with remethylation rates at and around the gene's body, with lower variability genes having significantly higher remethylation kinetics (1-way ANOVA post hoc Tukey test: $p < 2.2 \times 10^{-16}$; Figure 1.7B and 1.8). We theorized that remethylation rates may dictate the duration of methylation heterogeneity at a particular gene's regulatory domain, and therefore impact the likelihood of cells yet to have promoter remethylation, resulting in temporary heterogeneous gene expression across the cell population (Figure 1.7C). With this in mind, we hypothesized that, in the event of a TF binding and leading to sustained transcription, regions with slow remethylation kinetics could remain hypomethylated in their new cell state. We therefore intersected Repli-DMRs with previously identified hypomethylated DMRs during the transition from hESCs to each of the three germ layers (ectoderm, mesoderm, and endoderm)[7]. We found that Repli-DMRs were enriched in hypomethylated DMRs for all three cell types, particularly in these found in ectoderm (Figure 1.7D). Separately, GO analysis of genes associated with Repli-DMRs also showed a strong enrichment for developmental genes (Figure 1.7E). The noticeable presence of neuro-related terms also reflected the ectoderm DMR results mentioned above. Seeing as a group of cell-cycle regulated genes has previously been shown to be made up of developmental regulators[45], we

calculated the average remethylation rate across their promoters. We found that rates were significantly lower in this cluster, as opposed to other cell-cycle regulated genes (Figure 1.7F). Overall, our results point to a potential mechanism by which this temporal window of epigenetic heterogeneity could play a role in development and cell fate transitions.

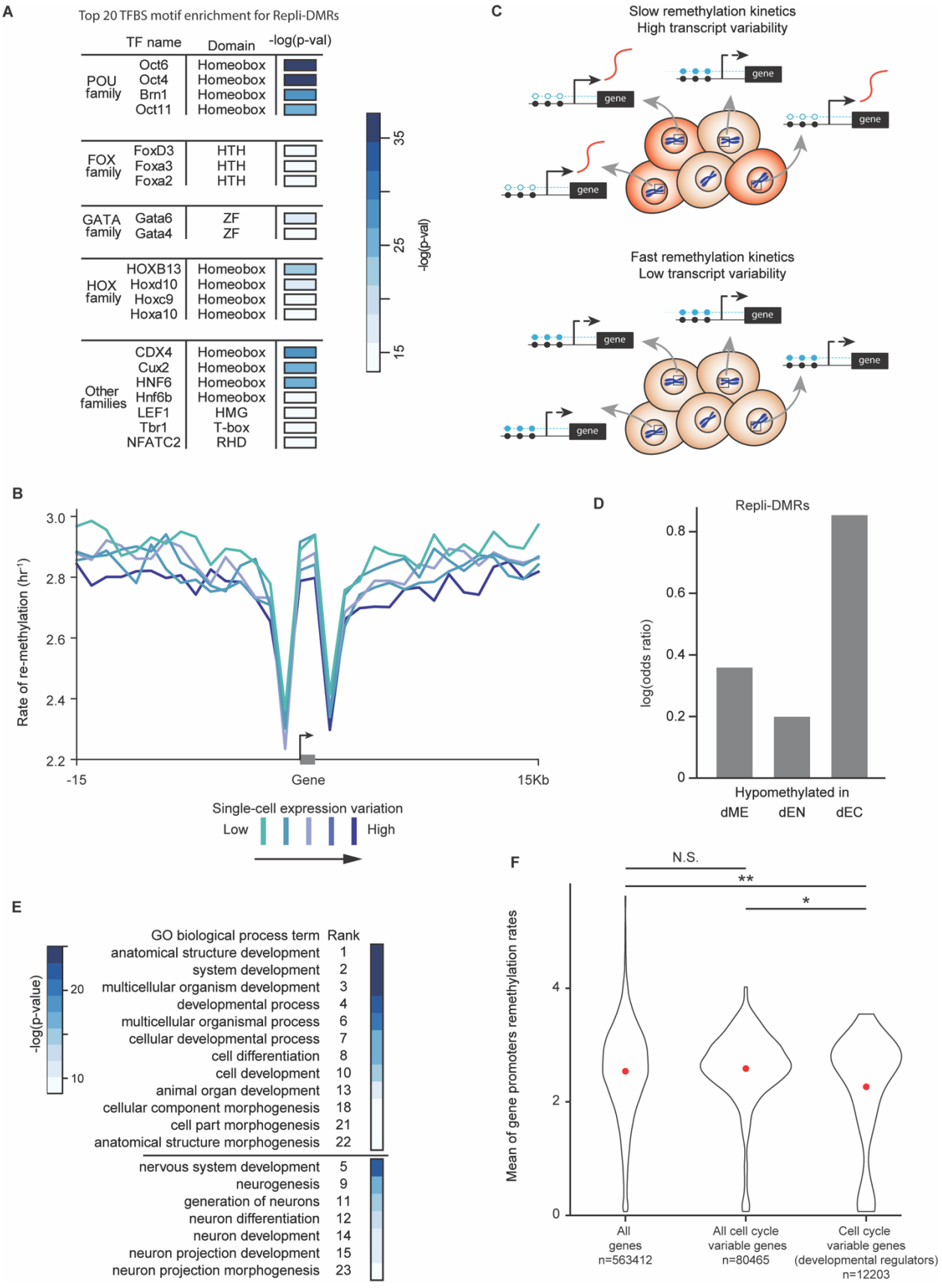
A
Top 20 TFBS motif enrichment for Repli-DMRs
TF name
Domain
-log(p-val)
POU family
Oct6
Oct4
Brn1
Oct11
Homeobox
FOX family
FoxD3
Foxa3
Foxa2
HTH
GATA family
Gata6
Gata4
ZF
HOX family
HOXB13
Hoxd10
Hoxc9
Hoxa10
Other families
CDX4
Cux2
HNF6
Hnf6b
LEF1
Tbr1
NFATC2
HMG
T-box
RHD
35
25
15
-log(p-val)
C
Slow remethylation kinetics
High transcript variability
gene
Fast remethylation kinetics
Low transcript variability
B
Rate of re-methylation (hr⁻¹)
3.0
2.8
2.6
2.4
2.2
-15
Gene
15Kb
Single-cell expression variation
Low
High
D
Repli-DMRs
log(odds ratio)
0.8
0.6
0.4
0.2
0
Hypomethylated in
dME
dEN
dEC
E
-log(p-value)
20
10
GO biological process term
Rank
anatomical structure development 1
system development 2
multicellular organism development 3
developmental process 4
multicellular organismal process 6
cellular developmental process 7
cell differentiation 8
cell development 10
animal organ development 13
cellular component morphogenesis 18
cell part morphogenesis 21
anatomical structure morphogenesis 22
nervous system development 5
neurogenesis 9
generation of neurons 11
neuron differentiation 12
neuron development 14
neuron projection development 15
neuron projection morphogenesis 23
F
N.S.
**
*
Mean of gene promoters remethylation rates
4
2
0
All genes n=563412
All cell cycle variable genes n=80465
Cell cycle variable genes (developmental regulators) n=12203

**Figure 1.7: Post-replication remethylation kinetics associate with gene expression variability and developmental elements A.** Table highlighting the top 20 most significant transcription factor binding site (TFBS) motifs enriched in Repli-DMRs. TFs are organized by protein family and heatmap reports the degree of statistical significance for TFBS motif enrichment. **B.** Smooth median curve fitting for remethylation rates at and within ±15Kb of genes. Each line depicts different bins of gene expression variability from low (light blue) to high (dark blue). **C.** Schematic depicting the theoretical mechanism by which slow (top) and fast (bottom) DNA remethylation rates could influence gene expression variability across a population. **D.** Barplot showing the log odds ratio enrichment of regions, hypomethylated in differentiated mesoderm (dME), endoderm (dEN), and ectoderm (dEC), in Repli-DMRs. **E.** Top 23 most significant gene ontology biological process terms enriched in genes associated with Repli-DMRs, related to development (top) and ectoderm development (bottom). **F.** Violin plots showing the distribution of mean promoter remethylation rates for all genes, cell-cycle variable genes, and cell-cycle variable genes involved in regulating development. The number of genes included is displayed below each category. 1-way ANOVA post hoc Tukey test: * $P \leq 0.05$, ** $P \leq 0.01$, N.S. = non-significant.

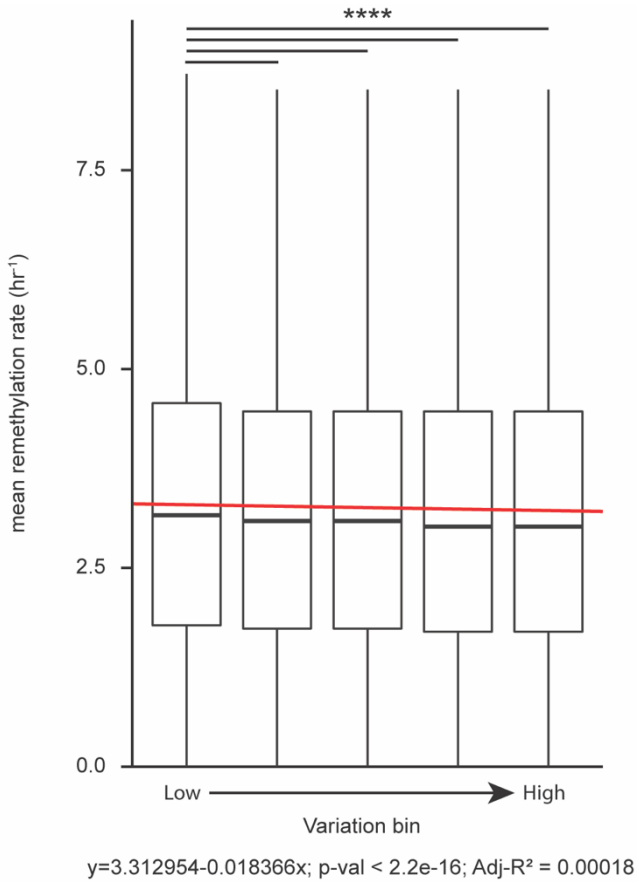


**Figure 1.8: Remethylation kinetics inversely correlates with single cell RNA sequencing (scRNA-seq) gene expression variability**. Boxplot showing the distribution of mean remethylation rates within ±10Kb of the transcription start site (TSS) of genes binned according to expression variability in the cell population. Dark line indicates the median and edges of the box show the 25$^{th}$ and 75$^{th}$ percentile values. Linear fit of the data is shown as a red line, with the corresponding formula and statistics shown at the bottom. 1-way ANOVA post hoc Tukey test: **** $P \leq 0.0001$.

### 1.2.4 The transient window of regulatory heterogeneity leaves the genome vulnerable to age-related epigenetic drift over an organism's lifetime

Our results suggest that delays in remethylation are associated with transient regulatory heterogeneity from cell-to-cell, which can be essential for allowing important developmental

changes to take place. Over the course of multiple cycles, however, a cell's ability to perform cellular functions can degrade, leading to observed replicative aging[2,3]. In order to elucidate if post-replication DNA methylation maintenance kinetics could be the common molecular link between cell fate and aging, we compared the methylation level of newborn and nonagenarians/centenarian DNA[4,46] across the temporally dynamic decile tiles. We observed an increase in age-related methylation difference in tiles of increasing temporal methylation difference (Figure 1.9A and 1.10A), suggesting that increases in the duration of the window of heterogeneity could account for increased loss of methylation with age. These results prompted us to question whether certain regions of the genome may be more susceptible to epigenetic drift, while others remain resilient with age. Previously, CpG density was found to be an important factor in susceptibility to age-related epigenetic drift[15], attributed to the methylation enzymes' processivity[5]. As expected, we observed a larger loss in methylation at CpGs with fewer neighbors (Figure 1.9B). However, we also found a positive correlation between remethylation rate and CpG density (Figure 1.9C), suggesting that CpG-poor regions may be more vulnerable to age-related epigenetic drift due to lack of maintenance exacerbated by slower remethylation kinetics (Figure 1.9D). This theory is in line with others that highlight a deregulation of maintenance machinery with age[15]. Local loss of methylation accumulated across multiple mitotic divisions has also been reported in the context of diseases, like cancer, where CpG context can be predictive of susceptibility. Specifically, CpGs in the WCGW context, where W stands for A or T, have been shown to be more prone to loss in methylation in cancer, unlike those in a SCGS context, where S strands for C or G[5]. CpGs in the SCGS context were found to have significantly faster remethylation kinetics (1-way ANOVA post hoc Tukey test: $p < 2.2x\ 10^{-16}$) than genome-wide CpGs, unlike WCGW CpGs which had significantly slower rates (1-way ANOVA post hoc Tukey test: $p < 2.2x\ 10^{-16}$; Figure 1.9E). Breaking WCGW CpGs according to the number of neighboring CpGs, within ±35bp window, we observed two different types of behaviors related to age-related loss and remethylation rate (Figure 1.9F and 1.10B). Notably, WCGW CpGs with

0 to 1 neighbors seemed to be most susceptible to methylation loss with slower remethylation rates, consistent with our previous findings. However, more surprisingly, CpGs with 2 or more neighbors seemed uncorrelated, regardless of the remethylation rate, suggesting that CpG density may overcome a susceptibility factor like CpG context. Overall, our results suggest that while a transient window of epigenetic entropy at regulatory regions provides context for multicellular development, it may also be the source of both age and disease-related methylation loss in situations where maintenance is not reliably copied to the newly divided cells (Figure 1.9D).

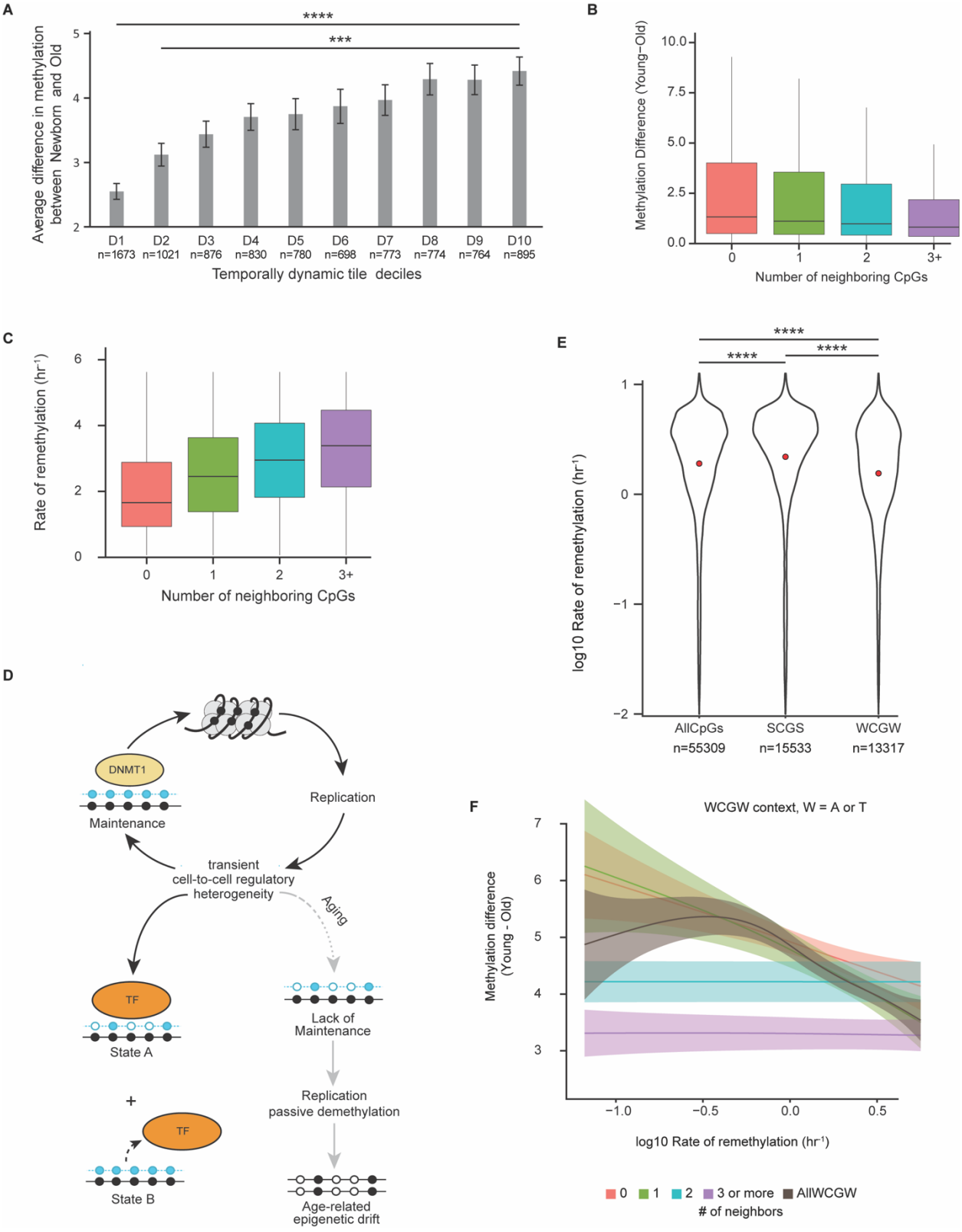
A
****
***
Average difference in methylation between Newborn and Old
5
4
3
2
D1 n=1673
D2 n=1021
D3 n=876
D4 n=830
D5 n=780
D6 n=698
D7 n=773
D8 n=774
D9 n=764
D10 n=895
Temporally dynamic tile deciles
B
Methylation Difference (Young−Old)
10.0
7.5
5.0
2.5
0.0
0
1
2
3+
Number of neighboring CpGs
C
Rate of remethylation (hr⁻¹)
6
4
2
0
0
1
2
3+
Number of neighboring CpGs
E
****
****
****
log10 Rate of remethylation (hr⁻¹)
1
0
−1
−2
AllCpGs n=55309
SCGS n=15533
WCGW n=13317
D
DNMT1
Maintenance
Replication
transient cell-to-cell regulatory heterogeneity
Aging
TF
State A
Lack of Maintenance
+
TF
State B
Replication passive demethylation
Age-related epigenetic drift
F
WCGW context, W = A or T
Methylation difference (Young - Old)
7
6
5
4
3
−1.0
−0.5
0.0
0.5
log10 Rate of remethylation (hr⁻¹)
0
1
2
3 or more
AllWCGW
# of neighbors

**Figure 1.9: CpG density and context combined with slow remethylation kinetics affect CpG susceptibility to age-related epigenetic drift. A.** Barplot showing the average difference in methylation (young - old) from methylation array data across each of the temporally dynamic tile decile groups. The number of CpGs captured in each group is shown below each category. 1-way ANOVA post hoc Tukey test: *** $P \leq 0.001$, *** $P \leq 0.0001$. **B.** Boxplot showing the distribution of methylation difference (young – old) from methylation array data according to the number of neighboring CpGs present within ±35bp. Dark line indicates the median and edges of the box show the $25^{th}$ and $75^{th}$ percentile values. **C.** Boxplot showing the distribution of remethylation rates according to the number of neighboring CpGs present within ±35bp. **D.** Schematic showing the theoretical by which transient epigenetic heterogeneity acts as a double-edged sword, able to bring about multicellular life, while being susceptible to age-related epigenetic drift over time. Here, CpGs are depicted as circles on nascent reads, with methylation represented with filled blue circles. **E.** Violin plots showing the distribution of remethylation rates (in log form) at CpGs in different contexts (W = A or T, S = C or G). Red circles indicate the mean, and the number of CpGs is indicated under each condition. 1-way ANOVA post hoc Tukey test: **** $P \leq 0.0001$. **F.** Smooth average curve fitting of methylation difference (young - old) from methylation array data versus remethylation rates (in log form) at CpGs in the WCGW context. Line colors indicate a different number of neighboring CpGs within ±35bp and all CpGs in WCGW context.

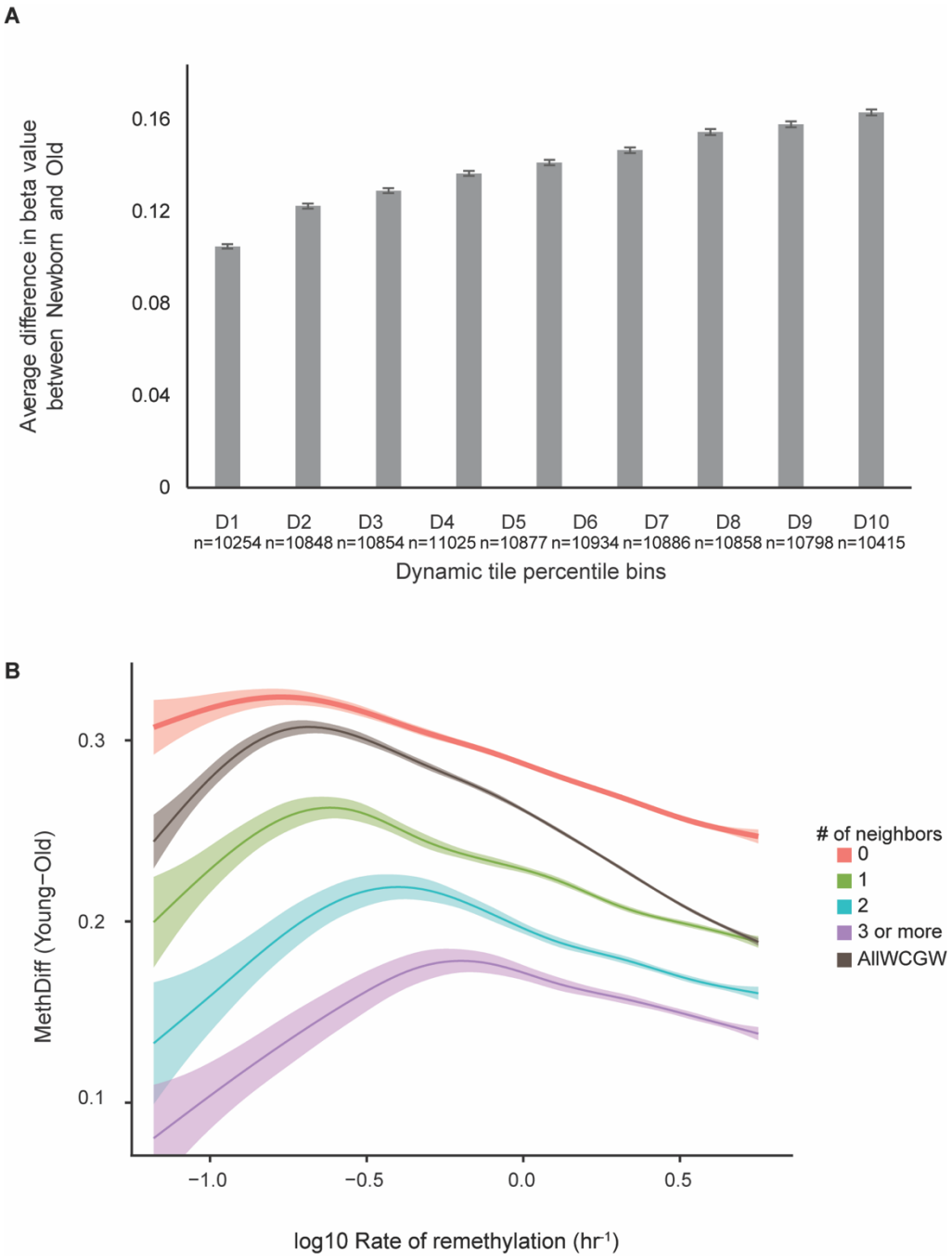


**Figure 1.10: Whole genome bisulfite sequencing (WGBS) methylation data for newborn and centenarian samples. A.** Barplot showing the average difference in methylation (young - old) from WGBS data across each of the temporally dynamic tile decile groups. The number of CpGs captured in each group is shown below each category. **B.** Smooth average curve fitting of methylation difference (young - old) from WGBS data versus remethylation rates (in log

form) at CpGs in the WCGW context. Line colors indicate a different number of neighboring CpGs within ±35bp and all CpGs in WCGW context.

## 1.3 Materials and Methods

### 1.3.1 Replication-associated bisulfite sequencing (Repli-BS) datasets

Repli-BS datasets for 0hr, 1hr, 4hr, 16hr, and arrested HUES64 human embryonic stem cell (hESC) samples were accessed from GSE82045[26]. Raw fastq files for the 0hr (S1-S6 fractions), 16hr, and arrested timepoints were downloaded and had adapters trimmed using TrimGalore (Version 0.4.4)[47]. Trimmed reads were then aligned to hg19/GRCh37 using Bowtie2 [48] as part of Bismark (Version 0.20.1) [49]. Paired-end read mapping efficiency varied between 70.4-87.9%, with an average of 81.13% (Appendix 1.3). Aligned BAM files from each of the 6 S fractions of the 0hr time point were merged. Methylation calls were finally generated through Bismark, with values from neighboring CpGs on opposite sides of the strand merged. Finally, the methylation ratios generated were filtered to keep only CpGs with a minimum read coverage of ≥ 5x, for increased confidence in CpG methylation ratios.

### 1.3.2 Temporally dynamic tile generation and binning

0hr and 16hr Repli-BS BED files, containing methylation score values across captured CpGs, were downloaded from GSE82045[26]. Files were then filtered for CpGs with a minimum read coverage of ≥ 5x and overlapping both timepoints. CpGs were further filtered so as to keep only those with a methylation difference (16hr minus 0hr) ≥ 20%. The remaining CpGs were then either tiled every 1000bp or through a custom method. For the custom method, CpGs within ±250bp were merged into a single tile using BEDTools' (Version 2.25.0) *merge* function[50]. To ensure that captured tiles were rich in CpGs, only the top 10% of tiles with the highest number of CpGs (220877 tiles) were kept. Generate tiles (both 1Kb and custom) were then intersected with

files containing CpGs with at least ≥ 5x coverage and a methylation difference (16hr minus 0hr) > 0 for both 16hr and 0hr. Tiles were then sorted by mean methylation difference (16hr minus 0hr) and binned into deciles. Tiles in the bin with the highest mean methylation difference were termed Repli-DMRs.

### 1.3.3 Methylation entropy and read-level calculations

Normalized methylation entropy (NME) was calculated by normalizing Shannon entropy (H) using the previously derived formula $NME = -\frac{H}{log_2(N+1)}$[51], where N represents the number of CpGs used in the calculation. Shannon entropy was calculated using the "entropy" function from python's SciPy.stats package[52] (Version 1.5.2), and histogram distributions of CpG methylation ratios from Repli-BS BED files for the 0hr, 4hr, 16hr, and arrested timepoints, with CpGs filtered for ≥ 5x coverage, as inputs.

Read-level methylation calculations were performed on reads from Repli-BS BAM files for the 0hr (all reads and filtered by dynamic tile decile overlap), 16hr, and arrested. To do so, methylation calls along reads, generated by Bismark[49], were extracted and filtered for CpG methylation information only. 1) Proportion of discordant reads (PDR) and 2) transition scores were then calculated as follows: 1) Each read's methylation calls were first analyzed to assign concordance or discordance to each read, using a custom python script. Reads were then filtered to retain only those with methylation for ≥ 2 CpGs per read. The remaining were intersected with a BED file of CpGs captured in 0hr Repli-BS data, and the number of discordant reads overlapping each CpG was determine using BEDTools'[50] *intersect* and *merge* functions. PDR was finally calculated at each CpG based using the previously described formula $\frac{\#\ of\ discordant\ reads}{Total\ \#\ of\ reads}$ [28]. A mean PDR was then calculated for each timepoint. The same analysis was performed using only reads and CpGs present in dynamic tile deciles. 2) From read-level methylation, consecutive CpG

methylation status was determined to calculate the number of transitions taking place along the read. Transition score calculations were then calculated as $\frac{\#\ of\ transitions\ along\ the\ read}{\#\ of\ CpGs\ used\ in\ the\ read}$.

In order to create both the "random" and the "cell state" distribution models, reads were split according to the total number of CpGs present in each read, and the total number of methylated CpGs and total number of CpGs captured were determined. Total methylation was then reassigned either randomly or in a consecutive fashion along the reads until no methylated CpGs were left. Transition scores were calculated, as described above, for the synthetically methylated reads in each model. Jensen-Shannon distance (JSD) was then calculated between the histogram distributions of the transition score for either the "random" or "cell state" models and the distribution of the actual data. This was performed using the *distance.jensenshannon* function from the SciPy.spatial[52] (version 1.1.0) python package. This calculation was done only between reads with the same number of CpGs so as to ensure a fair comparison between distributions. A mean JSD was then calculated using values from every instance of number of CpGs per read.

### 1.3.4 Stochastic modeling of post-replication remethylation kinetics

Post-replication rates of methylation re-establishment were generated from HUES64 Repli-BS data[26] using a previously established stochastic model[27]. Briefly, a maximum likelihood estimation was used to infer a per-CpG remethylation rate (k) and steady-state methylation fraction (f). So as to investigate the consequences of remethylation kinetics at methylation-rich CpGs, only CpGs assigned with a value $f \geq 0.8$.

### 1.3.5 Annotations and downloaded datasets

In order to determine the impact of our generated temporally dynamic tiles and remethylation rates, files were intersected with genomic features, histone modification and protein

(EP300, EZH2, CTCF) chromatin immunoprecipitation sequencing (ChIP-seq) peak files, and DNase hypersensitivity peaks for H1 hESCs from the UCSC genome table browser[53], using BEDtools' *intersection* function[50]. Additionally, Repli-DMR tiles were intersected with hypomethylated differentially methylated regions (DMRs) between the hESC cell line HUES64 and each of the three germ layers (mesoderm, ectoderm, endoderm), downloaded from the roadmap epigenomics project database[7], and with a track of super-enhancer locations in H1 hESCs[54].

For the bivalent promoter analysis, a list of regions for bivalent promoters (defined as 1Kb upstream of the transcription start site (TSS) and 1.5Kb downstream of the TSS) with cell-cycle regulated H3K4me3[30], and a list of genes with bivalent promoters[55] were downloaded. In order to identify promoters from the gene list, promoter regions were generated using TSS of genes acquired from the hg19 biomart database[56]. Promoter regions were generated using the definition above, namely 1Kb upstream of the TSS and 1.5Kb downstream of the TSS. Finally, a median rate of remethylation was calculated in 100bp windows within a ±10Kb region around promoters using deeptools'[57] *computeMatrix* function (Version 3.5.0).

For our gene expression analyses, scRNA-seq data from H1-hESCs was downloaded from GSE36552[44]. Variation was calculated at each gene using the coefficient of variation equation (standard deviation/mean) on gene expression RPKM values for each cell at that particular gene. Remethylation rates, described above, within ±15Kb from each gene's TSS were then identified. Genes were then filtered to retain only those with RPKM contributions from at least 3 cells, and having at least 20 remethylation rates within the ±15Kb region. The remaining genes were then separated into 5 bins of equal size. A median rate of remethylation was then calculated in 1Kb windows within a ±15Kb region around each gene using deeptools'[57] *computeMatrix* function (Version 3.5.0). Additionally, a list of cell-cycle regulated genes was downloaded[45] and promoter regions were generated by extending 2Kb upstream and 500bp downstream of gene

TSS. These were then intersected with remethylation rates using BEDtools' *intersection* function[50].

Age-related epigenetic analyses were performed using both DNA methylation information from a WGBS study of newborn and centenarian blood samples (GSE31263)[4], for fair CpG comparability to our Repli-BS dataset, and a methylation microarray dataset of 19 newborn and 19 nonagenarians (GSE30870)[46], for increased sample size. For disease-related loses in methylation at different CpG contexts, WCGW and SCGS locations were identified in the genome using the Hypergeometric Optimization of Motif EnRichment (HOMER)[58] (Version 4.7) software's *seq2profile.pl* function to create .motif files for each that were then scanned across the hg19 genome using *scanMotifGenomeWide.pl* function. Finally, the number of neighboring CpGs was calculated within a window of ±35bp around each CpG, as previously defined[10], using a combination of BEDTool's[50] *getfasta* function (Version 2.25.0) and UCSC's[59] faCount (Version 327).

### 1.3.6 Transcription factor binding site (TFBS) enrichment and gene ontology (GO) analyses

In order to determine TFBS present within the Repli-DMR tiles, we made use of the HOMER software[58] (Version 4.7). Using the tiles as inputs, HOMER was performed using the hg19 genome as background, along with a specified motif size parameter based on average tile size. TFBS motif results were finally filtered for p-value $\leq 0.01$. As part of HOMER, a gene ontology term enrichment analysis was also performed using the parameters above. GO term analysis was performed on the top 20 most significant TF results from HOMER through the Gene Ontology Resource's PantherDB[60].

### 1.3.7 Replication-associated assay for transposase-accessible chromatin sequencing (Repli-ATAC) and insert size entropy calculations

A Repli-ATAC-seq protocol was derived from an established ATAC-seq protocol[61], with modifications for nascent read pulldown. Human embryonic stem cells (HUES64) were grown on in feeder-free conditions using Geltrex (Thermo Fisher Scientific). Once ready, cells were given fresh mTesr1 media (STEMCELL Technologies) and were treated with 50 mM BrdU (BD Pharmingen, BD Biosciences) for 1 hour. Following treatment, media, containing BrdU, was aspirated and cells were washed twice with mTesr1 media. Cells were then collected at timepoints (0, 1, 4, 16 hour) post-BrdU treatment, through Accutase (Innovative Cell Technology Inc.) treatment and subsequent wash and spin steps. Following collection, cells were counted to ensure retrieval of at least 100,000 cells, and were immediately assayed using the ATAC-seq protocol described previously [61] up to the PCR Amplification step 2. Thermal cycling was performed for 1 cycle at 72˚C to allow for extension of both ends of primer after transposition. At this point, the DNA was purified using a Qiagen MinElute PCR Purification Kit and half of the sample were immunoprecipitated with anti-BrdU antibody (BD Pharmagen). For immunoprecipitation, DNA was first denatured through incubation 95˚C for 5 mins, then cooled for 2 mins on ice-water and added to a tube with IP buffer (1 mM sodium phosphate, 140 mM NaCl and 0.02% TritonX-100). 0.5 mg anti-BrdU antibody (BD Pharmagen) was added to the sample and incubated for 20 min at room temperature with constant rotation in the dark. 20 µg of rabbit anti-mouse IgG (BD Biosciences) was added for 20 mins at room temperature with constant rotation before centrifugation at 17000xg for 5 min at 4˚C. The supernatant was entirely removed and ice-cold IP buffer was added, followed by a centrifugation step at 17000xg for 5 mins at 4˚C. Following removal of the supernatant, the pellet was resuspended in 200 µl of digestion buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, and 0.5% SDS) with 0.25 mg/ml proteinase K before incubating the samples overnight at 37˚C. A further 100 µl of fresh digestion buffer with 0.25 mg/ml proteinase K was added to samples before incubating for another 60 mins at 56˚C. DNA purification was performed using AMPure XP beads (Beckman Coulter). The ATAC-seq protocol described in [61] was then continued from step 1 of the PCR Amplification

step until library completion. Repli-ATAC-seq libraries were sent to the UCI Genomics High-Throughput Facility and sequenced on an Illumina NovaSeq6000 sequencer. We performed paired-end sequencing runs for a total of 200 cycles.

Reads were analyzed using the nfcore/atacseq package[62] (Version 1.2.1), available on Github. The pipeline performs the following steps: 1) Raw read QC on FASTQ files with FastQC, 2) Adapter trimming with Trim Galore!, 3) Alignment with BWA to generate BAM files, 4) Alignment quality control removing mitochondrial DNA, blacklisted regions, duplicates, unmapped reads, reads mapping to multiple locations, reads with mismatches, reads with insert size >2kb, and reads that map to different chromosomes with a combination of SAMtools, BAMTools, and Pysam, 5) Creation of bigWig files scaled to 1 million mapped reads with BEDTools and bedGraphToBigWig, and finally 6) peak-calling with MACS2.

Insert length distributions were then generated from BAM files using custom code in Python (Version 3.8.4). Insert length entropy at a particular timepoint was calculated using the following formula:

$$avg_{rep}\left(\sum_{i=1}^{2000} P(x_i)_{ctrl} * \ln P(x_i)_{ctrl} - \sum_{i=1}^{2000} P(x_i)_{ip} * \ln P(x_i)_{ip}\right)$$

where $P(x_i)$ is the probability of occurrence of an insert of a particular length i.

The central calculation is the Shannon's entropy of insert length in both the IP and control samples. The IP sample can be considered a subsampling of the control sample; thus a differential entropy is taken between the two to ascertain the difference in entropy by taking the subsampling. The final value is the mean of differential entropies across replicates. The time-dependent decrease in entropy is present in both replicates, irrespective of the mean. This calculation was performed in Python 3.8.4 using the SciPy package[52].

### 1.3.8 Statistical analyses

All statistical tests were performed through R (Version 3.6.2) [63]. Student t-test and 1-

way ANOVA, followed by a subsequent post-hoc Tukey Honest Significant Differences, were performed on samples to determine significance.

Odds ratio (OR) analyses were performed to determine the significance of Repli-DMR association to particular features (for example, histone modifications and genomic features). OR were calculated as follows: $\frac{a/c}{b/d}$, where *a* = the number of basepairs that fall within a Repli-DMR and within the context of interest, *b* = the number of basepairs that fall within Repli-DMRs and outside of the context of interest, *c* = the number of basepairs that fall outside of Repli-DMRs and within the context of interest, *d* = the number of basepairs that fall outside of Repli-DMRs and outside of the context of interest. The logarithmic OR value (logOR) was then reported for each context of interest. Fisher's exact test was used to determine significance of odds ratios.

Finally, non-linear least squares (NLS) curve fitting was performed on accessibility entropy and NME using values for each timepoint using self-starter parameters generated by the *SSAsymp* from the stats R package.

# SECTION 2

# DNA methylation analysis reveals epimutation hotspots in patients with dilated cardiomyopathy-associated laminopathies

## 2.1 Introduction

### 2.1.1 Lamin A/C in the nuclear envelope

The gene *LMNA* gives rise to both Lamin A and C through alternative splicing. These two intermediate filaments line the inner membrane of the nuclear envelope, and are essential in providing structure to the nucleus, while simultaneously linking the chromatin to the cytoskeleton [64].

### 2.1.2 Lamins interact with DNA

DNA regions associated to lamins at the periphery of the nucleus, termed lamina-associated domains (LADs), have previously been shown to be part of heterochromatin, the condensed region of chromatin where gene expression is silenced [65]. These structural associations, however, are disrupted in cases of mutated *LMNA*, leading to nuclear blebbing and subsequently nuclear envelope rupture [66]. Together, these events lead to DNA damage [67], as well as altered gene expression and chromatin organization [68].

### 2.1.3 Laminopathies and Dilated Cardiomyopathy

Mutations in the *LMNA* gene cause a variety of diseases, called laminopathies, including premature aging, muscular dystrophy, lipodystrophy, and bone abnormities. Cardiac disease such as dilated cardiomyopathy (DCM) remains the most common type among the *LMNA*-related diseases. Patients with DCM typically present with enlargement of the ventricles, resulting in systolic dysfunction, eventually leading to heart failure [69]. While cardiac symptoms typically present in adulthood, other laminopathy-associated phenotypes, such as facial and digital bone abnormalities (ex. brachydactyly), are congenital and indicative of disease mechanisms occurring early in development.

### 2.1.4 DNA methylation in *LMNA*-mutated DCM samples

The role of DNA methylation, which works in conjunction with the chromatin to control gene expression, has not been thoroughly investigated in the context of *LMNA* mutations. A recent study examined the impact of DNA methylation in heart tissue from patients with DCM [6]. This study concluded that altered CpG methylation, in combination with LAD redistribution and dysregulated gene expression, plays a key role in DCM pathogenesis. Although this study further solidifies the potential role of DNA methylation in the context of DCM, the individual impact of each family-specific *LMNA* mutation was not considered. Taking into account the specific mutation remains important since laminopathies arise in a large variety of tissue types and tissue abnormalities often appear in a mutation-specific fashion [69–71]. Furthermore, it was previously shown that methylation levels varied at the promoter of laminopathy-related genes in cells with two distinct *LMNA* mutations [72].

## 2.2 Results

### 2.2.1 Genome-wide DNA methylation analysis within family-specific primary fibroblasts and iPSCs

To investigate the effect of *LMNA* mutations on the DNA methylation landscape, RRBS was performed on primary skin fibroblasts (and their iPSC derivatives) obtained from two families harboring unique *LMNA* mutations, and an additional unaffected (and unrelated) donor control cell line (Figure 2.1A). After filtering, we captured an average of 2.2 million CpGs per sample in both cell types (Appendix 2.1), of which 1539576 (62.2-73.2% of total CpGs) and 1418269 (58.2-62.9% of total CpGs) overlapped all samples in fibroblasts and iPSCs, respectively (Figure 2.2A). Filtered CpGs represented a large portion of CpGs found in exons (13.7-20.0%) and promoters (12.1-20.5%) in fibroblasts, and in iPSCs (12.0-19.2% and 12.8-19.9%, respectively) (Figure 2.1B). This represented a coverage of approximately half of all promoters in both fibroblasts and iPSCs (Appendix 2.2). The relative distribution of CpGs captured in exon, intergenic, intron, and promoter was similar within each sample, in both cell types. These results agree with previous reports that RRBS captures about 2.8 million CpGs, within 60% of promoters [73,74].

Globally, average methylation levels of controls (60.6 ± 0.6 in fibroblast and 69.7 ± 0.3% in iPSC) and patients (61.42 ± 0.9% in fibroblast and 70.9 ± 0.6% in iPSC) did not vary between the two groups (Figure 2.1C). This observation was consistent when separated by family. At the single CpG level, however, we observed differences between patient and control sample methylation levels in both fibroblast and iPSCs (Figure 2.2B), with the largest differences observed at CpGs with intermediate methylation (30-60%) in controls. To obtain a regional view of how methylation patterns change in patient samples compared to unaffected controls, we focused on differences in methylation levels over sections of the genome rather than individual CpGs. Interestingly, some differences in methylation, in the fibroblast genome for example, appeared to be shared across both families (Figure 2.1D). In contrast, other methylation differences were unique to one family, with little differences seen across samples in the other family. Due to the presence of distinct regional methylation difference between patient samples and unaffected controls, we focused our analysis on DMR tiles (Figure 2.3), classified as "Shared" (Figures 2.1D & E, orange shaded region), "Family A-specific" (green shaded region), or "Family C-specific" (purple shaded region). Methylation differences of Family A and Family C samples were confirmed to significantly correlate genome-wide at shared DMRs (Figure 2.1E, left panel, Pearson correlation: $R = 0.49$, $p < 2.2 \times 10^{-16}$), while no positive correlation was observed at Family-specific DMRs (Pearson correlation: $R = -0.017$, $p = 0.026$ for Family A; $R = -0.026$, $p = 0.0019$ for Family C).

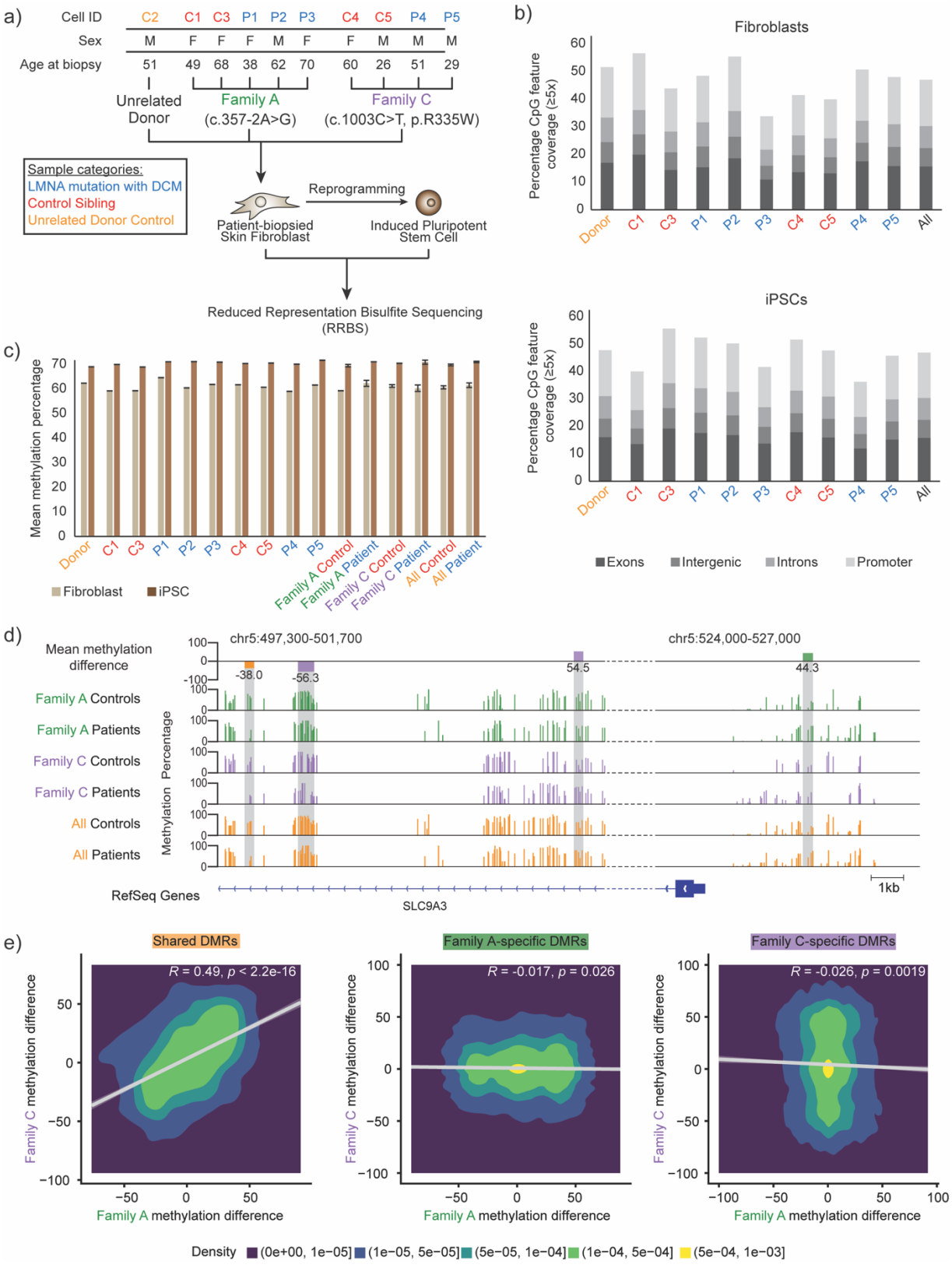


**Figure 2.1: Characterization of DNA methylation in *LMNA*-mutant fibroblasts and iPSCs. A,**

Schematic representation of the experimental setup. Cluster branches indicate groups of samples by family. **B**, Stacked bar plot showing the percentage of CpGs (≥ 5x depth) in a particular feature (Exon, Intergenic, Introns, Promoter from bottom to top) for all samples individually and merged in fibroblast (top) and iPSC (bottom). **C**, Bar plot displaying mean genome-wide DNA methylation percentage using CpGs (≥ 5x depth) across all samples individually and merged by groups in fibroblasts (tan) and iPSCs (brown). **D**, Example of regions with CpG methylation differences between patient and control fibroblasts. Top, Genome browser track (chr5:497,300-501,700 and chr5:524,000-527,000) displaying DMRs based on mean methylation differences (patient minus control) by group (Family A-specific – green, Family C-specific – purple, Shared – orange). Middle, Methylation levels for patient and control samples by group. Gray regions reflect the location of DMRs from the top track. Bottom, Depiction of RefSeq gene annotation. **E**, 2D density plots of CpG methylation difference (patient minus control) in fibroblasts from Family C (y-axis) or Family A (x-axis) at Shared, Family A-specific, and Family C-specific DMRs.

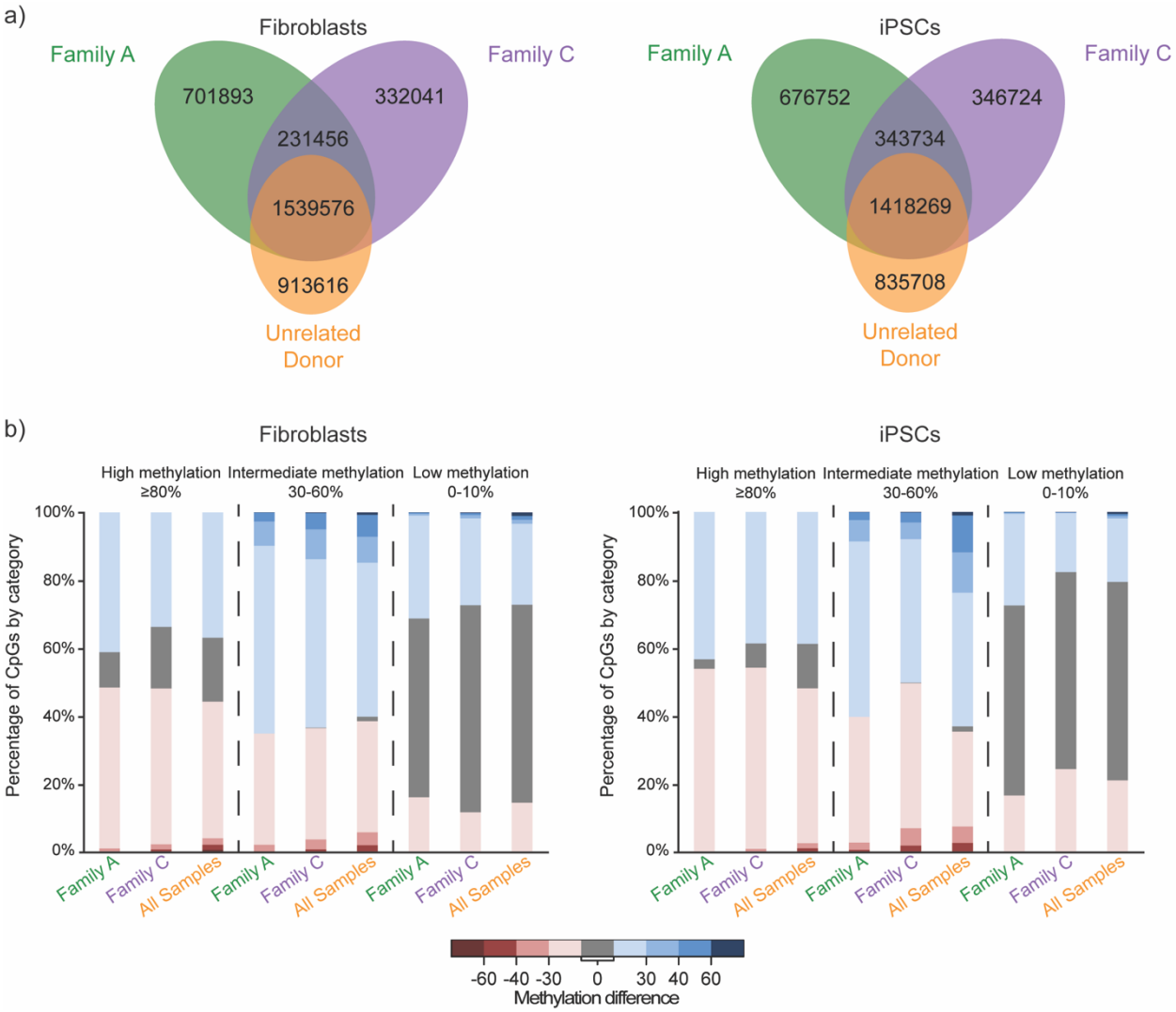


**Figure 2.2: Quantification of captured CpGs and corresponding DNA methylation by family.** **A**, Venn diagrams of the number of overlapping CpGs, captured in RRBS and filtered for ≥ 5x depth across grouped samples for fibroblast (top) and induced pluripotent stem cell (iPSC) (bottom). **B**, Top, Classification of CpGs based on methylation percentage of input control samples (high – left, intermediate – center, or low – right). Middle, Percentage stacked bar plot of CpGs based on the degree of methylation difference (patient-control), as indicated by heatmap legend. Bottom, Group of samples used for percentage calculation.

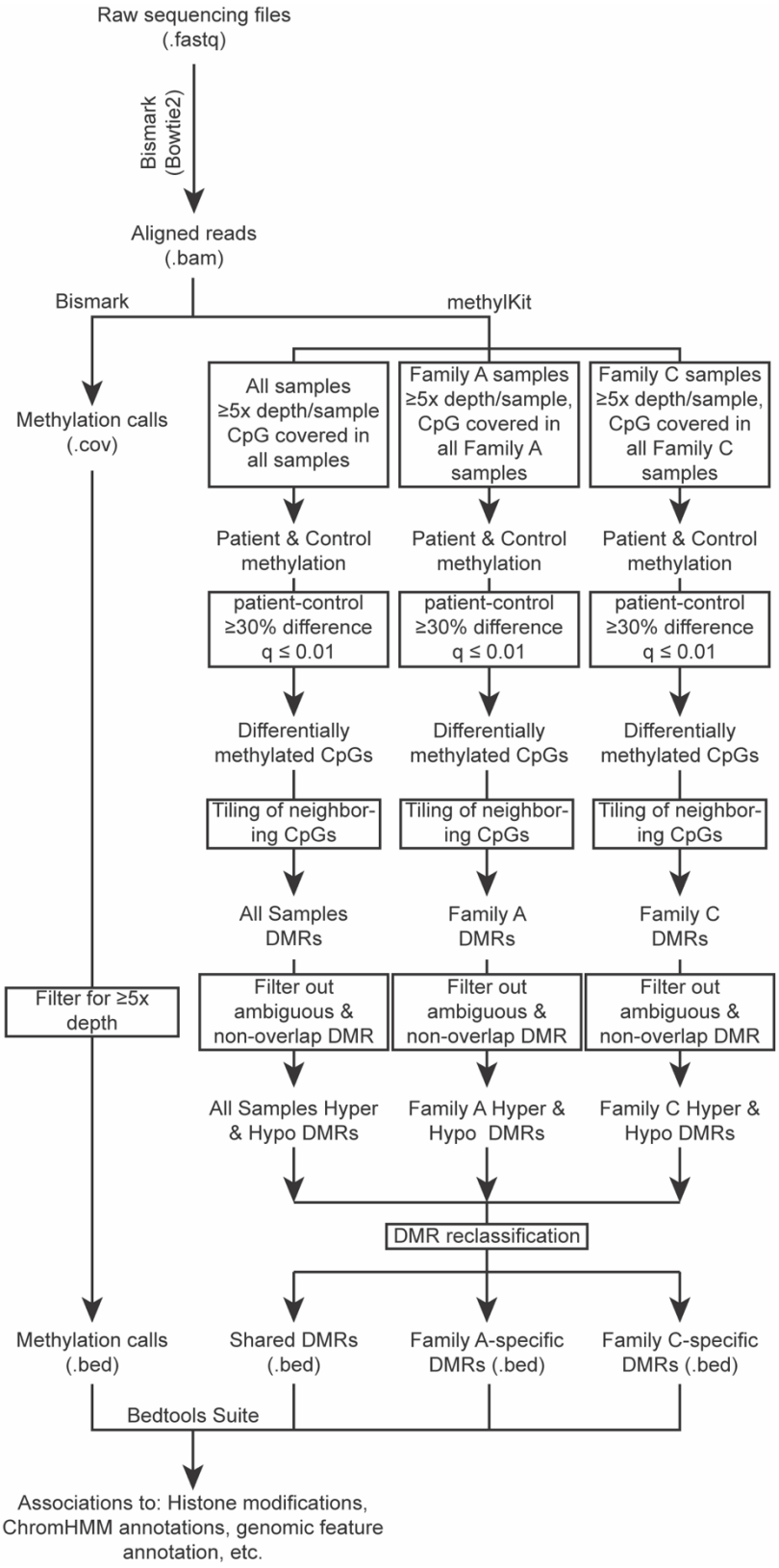


**Figure 2.3: Computational workflow of DNA methylation analyses.** Flowchart showing the

computational workflow for producing methylation call and DMR BED files from raw fastq files, and their subsequent analyses.

### 2.2.2 Family-specific epigenetic signatures dominate DMR landscape in fibroblasts

To characterize family-specific and shared DNA methylation differences between patient and control samples, we first focused on data from patient-biopsied fibroblasts only. Despite no differences in global methylation levels between families (Figure 2.1C), hierarchical clustering of samples based on all DNA methylation data showed that samples tended to group according to family (Figure 2.4A). Clustering was also performed on samples following removal of sex chromosomes X and Y, in order to identify possible sex biases. Despite clusters no longer segregating by family (Figure 2.5A), the average Pearson correlation coefficient of genome-wide methylation data between samples was higher when compared between samples of the same family than when compared across families (Figure 2.5B), indicating that genome-wide methylation signatures were more dependent on family than sex. Furthermore, DMRs in sex chromosomes made up only 0.76-2.63% of total DMRs generated for each category (Figure 2.4B, 2.5C).

By performing methylation comparisons between patient and control samples within the same family, we posited that disease-specific patterns of differential methylation would more strongly emerge from our analyses, while normalizing for family-specific methylation pattern biases. We therefore focused the rest of our analyses on “shared” (Figure 2.4B, orange shaded region), “Family A-specific” (green shaded region), or “Family C-specific” (purple shaded region) DMRs. These three groupings were replicated through hierarchical clustering based on methylation at DMR locations (Figure 2.6). While clustering based on shared DMR methylation showed a clear separation between patient and control samples, family-specific clusters still emerged from within each patient and control sub-cluster (Figure 2.4C). This evidence, together

with the identification of a relatively low number of shared DMRs overall (Figure 2.4B), show that family-specific changes dominated our DMR analysis. Furthermore, we noted that the absolute median methylation difference across DMR tiles was significantly higher across family-specific comparisons (41.60 for Family A, and 52.63 for Family C) relative to DMRs obtained from our shared comparison (34.10) (Figure 2.4D, Kruskal-Wallis test: p-value $<2.2x10^{-16}$). These findings indicated that epimutations that arise in DCM patients occur largely in a family-specific manner.

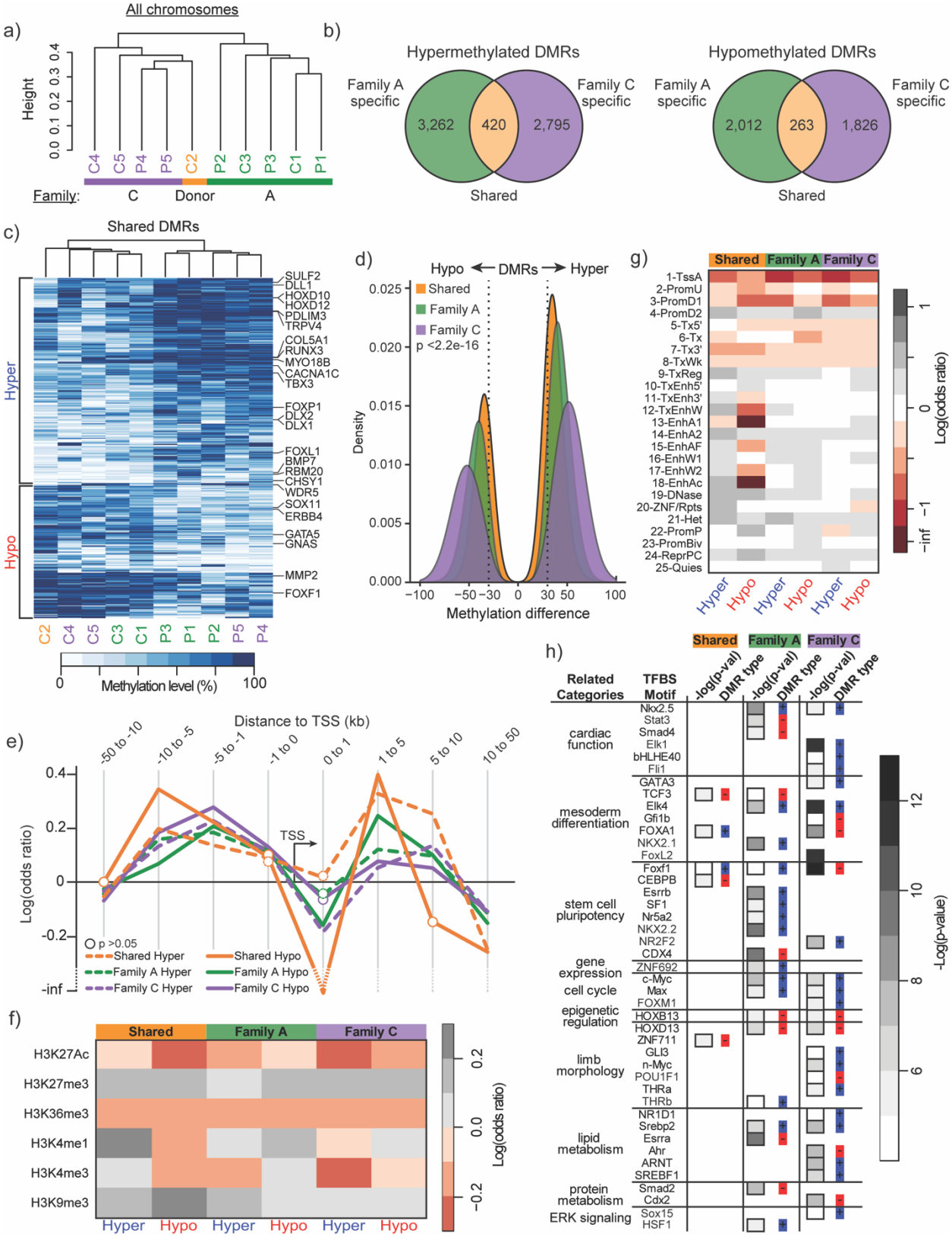


**Figure 2.4: Hypermethylated and hypomethylated DMRs localize at distal regulatory**

**features and transcriptionally repressed chromatin in fibroblasts. A**, Hierarchical clustering of all fibroblast samples by genome-wide DNA methylation. Colors represent family groupings. **B**, Venn diagrams showing the number of DMRs captured by group for both hypermethylated and hypomethylated DMRs. Orange regions denote “Shared DMRs”, green regions denote “Family A-specific DMRs”, and purple regions denote “Family C-specific DMRs”. **C**, Top, Hierarchical clustering of all samples by shared DMR methylation. Bottom, Heatmap of average CpG (≥ 5x depth) methylation percentage across shared DMRs for each individual sample. Genes associated to heart and skeletal system development are shown next to the associated DMR. **D**, Density plot of mean methylation difference (patient minus control) within DMRs by group. Overall Kruskal-Wallis test p-value is displayed. **E**, Line plot of log odds ratio of the likelihood of CpGs to fall within a hypermethylated (“Hyper”) or hypomethylated (“Hypo”) DMR and a given range of genomic distance away from a gene’s TSS. Open circles designate log odd ratios that were non-significant (p-value > 0.05) by Fisher’s exact test. **F**, Heatmap showing the log odds ratio of a CpG falling within both a DMR group and a given histone modification. **G**, Heatmap showing the log odds ratio of a CpG falling within both a DMR group and one of 25 ChromHMM annotated genomic regions. **H**, Table highlighting TFBS motifs enriched in shared, Family A, and Family C DMRs, grouped by TF-related categories. Heatmap reports the degree of statistical significance for TFBS motif enrichment. Results were categorized as hypomethylated (red) or hypermethylated (blue) according to the type of DMR associated to a particular TFBS motif.

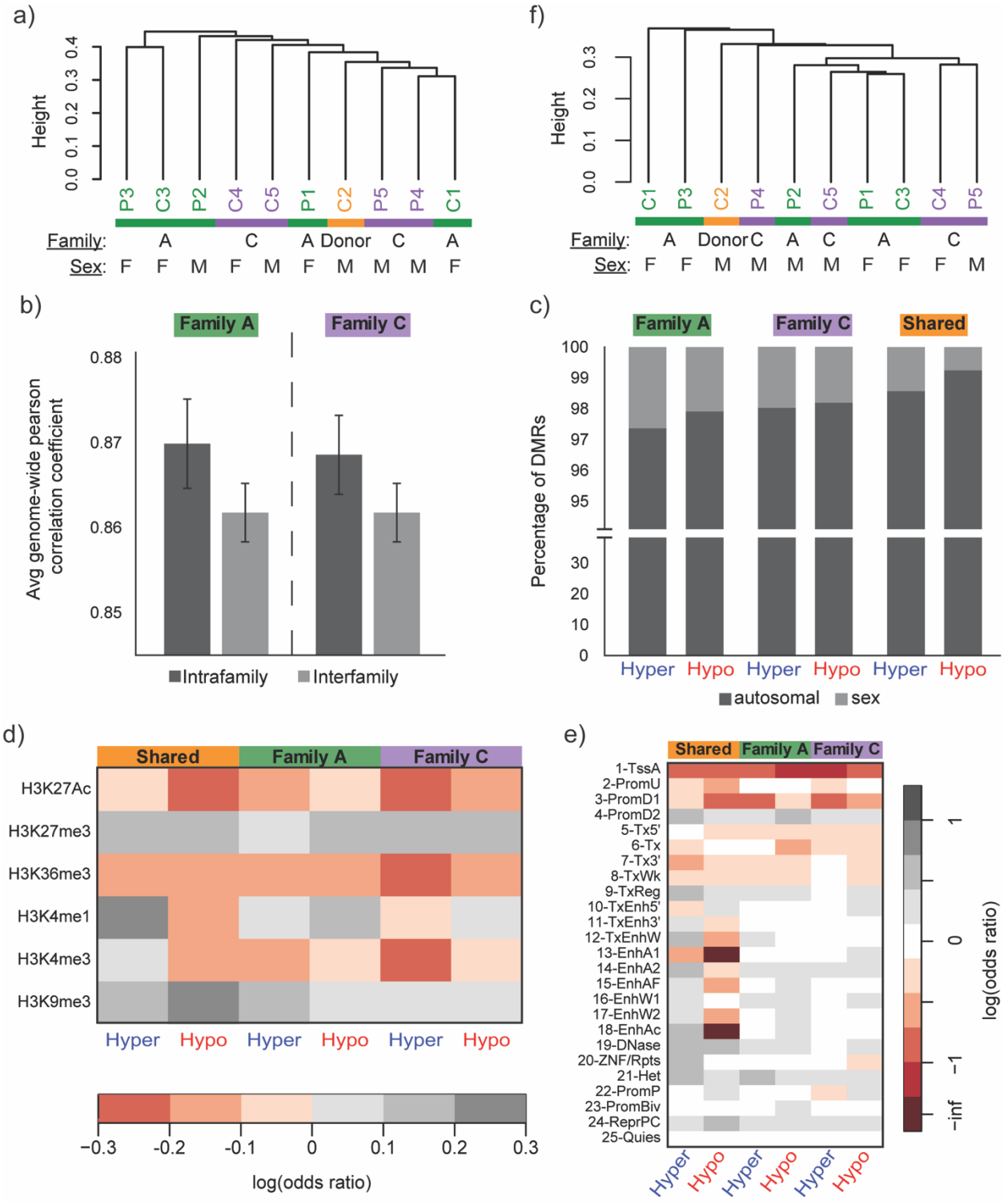


**Figure 2.5: Sex chromosomes had minimal impact on genome-wide and DMR results. A**, Hierarchical clustering of all fibroblast samples by genome-wide DNA methylation of autosomal chromosomes. Colors represent family groupings. **B**, Bar plot of the average Pearson correlation

coefficient of genome-wide DNA methylation between fibroblast samples belonging either to the same family ("Intrafamily") or to the other family ("Interfamily") in Family A (left) or Family C (right). **C**, Stacked barplot showing the percentage of fibroblast DMRs present either in autosomal or sex chromosomes for hypermethylated ("Hyper") or hypomethylated ("Hypo") differentially methylated regions (DMRs) in Family A-specific (left), Family C-specific (center), or Shared (right) groupings. **D**, Heatmap showing the log odds ratio of a CpG falling within both a given histone modification and a fibroblast DMR generated with only autosomal chromosomes. **E**, Heatmap showing the log odds ratio of a CpG falling within both one of 25 ChromHMM annotated genomic regions and a fibroblast DMR generated with only autosomal chromosomes. **F**, Hierarchical clustering of all induced pluripotent stem cell samples by genome-wide DNA methylation of autosomal chromosomes. Colors represent family groupings.

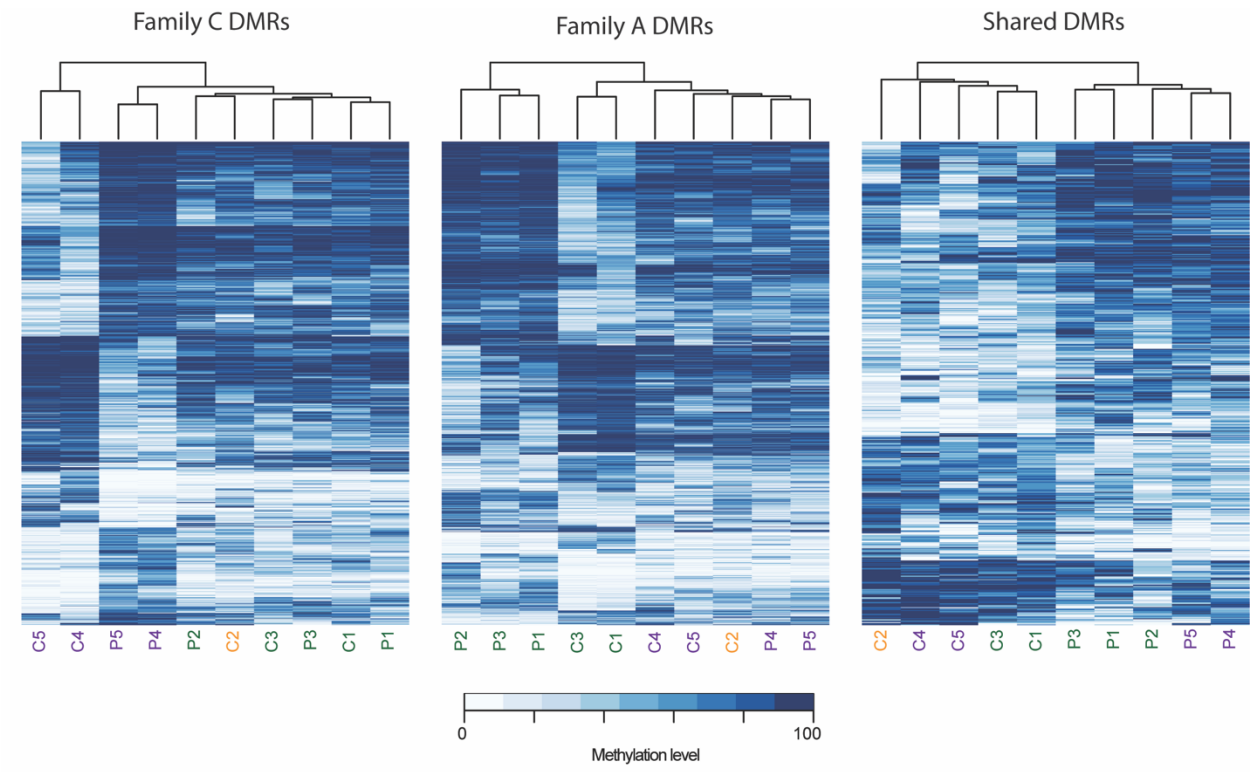


**Figure 2.6: Clustering of differentially methylated regions (DMRs) by DNA methylation level.** Top, Hierarchical clustering by methylation level in fibroblast samples. Bottom, Heatmap of

average CpG (≥ 5x depth) methylation percentage across Family C, Family A, and shared DMRs. Sample IDs are colored based on family of origin: Family C – purple, Family A – green, Unrelated Donor – orange.

### 2.2.3 Fibroblast DMRs associate with distal regulatory features and transcriptionally repressed chromatin

To investigate the potential regulatory impact of the DMRs identified, we used the Genomic Regions Enrichment of Annotations Tool (GREAT) [75] to identify genes that our DMRs may be regulating, both proximally and distally. Shared DMRs, despite their low frequency, revealed an association to 62 genes included in heart (eg. *GATA5*, *FOXL1*, *TBX3*, *MYO18B*, *CACNA1C*, *BMP7*) and skeletal system (eg. *HOXD10*, *HOXD12, RUNX3*) development GO terms (Figure 2.4C, full list shown in Appendix 2.3). To examine the potential regulatory impact of methylation on these DMR-associated genes, we performed an odds ratio (OR) analysis to determine the likelihood of CpGs falling within each of the three DMR groups and within a given genomic distance of a gene's transcriptional start site (TSS). This analysis revealed that CpGs within DMRs were generally more significantly likely to fall within genomic locations 1 to 10Kb upstream of a given gene's TSS and, more proximally, between 1 to 5kb downstream of the TSS (Figure 2.4E, Fisher's exact test: p-value ≤0.05, unless specified as non-significant).

The tendency of DMR-overlapping CpGs to fall distally to TSSs, beyond ±1kb, suggested that disease-associated changes in methylation could exist within diverse chromatin context that lie largely outside of promoters (which generally showed an odds ratio close to 1) and potentially within distal gene regulatory elements. To explore this, we performed a similar odds ratio analysis across a broader chromatin context (Figure 2.4F,G, Fisher's exact test: p-value ≤0.05, unless specified as non-significant in Appendix 2.4 & 2.5), to infer any potential role that aberrant methylation patterns might have on gene regulation in patient cells. Interestingly, an analysis

based on CpG overlap within fibroblast-specific histone modification landscapes (rather than distance from TSS) revealed that CpGs within hypermethylated DMRs obtained from our shared category showed a strong association (logOR = 0.24; p-value = $2.08 \times 10^{-20}$) with regions marked by histone 3 lysine 4 mono-methylation (H3K4me1), a histone mark traditionally enriched at enhancers [76,77] (Figure 2.4F). This was in stark contrast to CpGs within hypermethylated Family C DMRs, which displayed a protective effect with respect to H3K4me1 marks (logOR = -0.05; p-value = $3.2 \times 10^{-5}$). Conversely, Family A hypermethylated (logOR = 0.04; p-value = $2.71 \times 10^{-5}$) and hypomethylated DMRs (logOR = 0.12; p-value = $4.61 \times 10^{-24}$) both showed a slightly stronger association with this histone modification. A similar analysis which included the removal of sex chromosomes showed similar histone modification enrichment (Figure 2.5D) to those previously mentioned.

We next took a more focused approach towards understanding the relationship between the occurrence of CpGs in DMRs and functionally annotated genomic regions, as assigned (computationally) by ChromHMM [78,79]. These results revealed that all of our DMR categories showed a significant increased association with at least one subtype of enhancer annotation, including those functionally characterized as weak (annotation 16-18), strong (annotation 13-15) or transcribed (annotation 10-12). (Figure 2.4G, Fisher's exact test: p-value $\leq 0.05$, unless specified as non-significant in Appendix 2.5). Additionally, we saw a general negative association with promoter annotations (annotation 2-3, Fisher's exact test: p-value $\leq 0.02$), however we did observe strong associations with "downstream promoter elements" (annotation 4, Fisher's exact test: p-value $\leq 0.04$), which likely coincide with the increased association of DMRs at genomic distances 1-5kb downstream of gene TSSs that we observed previously (Figure 2.4E). Removal of sex chromosomes did not affect the results above for our ChromHMM analysis (Figure 2.5E).

We also observed that DMRs showed a strong likelihood to fall within histone modifications – H3K27me3 and H3K9me3 (Figure 2.4F, Fisher's exact test: p-value $\leq 2.2 \times 10^{-5}$) –

and functional genomic annotations – heterochromatin (annotation 21) and polycomb repression (annotation 24) ChromHMM annotations (Figure 2.4G, Fisher's exact test: p-value $\leq$0.03) – associated with gene repression. This was particularly interesting given that LADs, which are disrupted due to numerous *LMNA* mutations [6,80,81], typically co-localize to the nuclear periphery along with heterochromatic regions of DNA and also marked by H3K9m3 and H3K27me3 [65].

We next wanted to investigate whether DMR locations co-localized with certain classes of regulatory factor binding sites (TFBS). This could reveal important molecular targets within key signaling pathways that might be impacted by family-specific epimutations. We performed TFBS motif enrichment analysis in our DMRs using HOMER [58], focusing on TFBS motifs enriched only in either hypo- or hypermethylated DMRs. Few TFBS motifs were enriched within shared DMRs, however, these motifs were involved in mesoderm differentiation (e.g., *TCF3*, *FOXA1*) and stem cell pluripotency (e.g. *Foxf1* and *CEBPB*) (Figure 2.4H, full list shown in Appendix 2.6). Conversely, family-specific DMRs enriched for TFBS motifs of transcription factors (TFs) previously shown to be implicated in multiple categories relevant to laminopathies (cardiac function, limb morphology, lipid metabolism, mesoderm differentiation). The tendency of Family C DMRs to enrich for several TFBS motifs associated with limb morphology was particularly interesting given this family's presentation of a brachydactyly phenotype. In general, DMRs related to the enriched motifs were largely hypermethylated, though this could be due to the larger amount of hypermethylated DMRs present in fibroblasts.

### 2.2.4 Fibroblast DMR-associated genes enrich for family-specific disease ontologies

Due to the enrichment of TFBS motifs associated with pathways critical for tissue functions commonly disrupted in laminopathy diseases, we decided to investigate if shared and/or family-specific DMRs enriched for certain disease ontologies (Appendix 2.7). We performed disease ontology enrichment on genes associated with either hypo- or hypermethylated DMR contexts. The large presence of disease ontology terms represented by genes associated to

hypermethylated DMRs (Figure 2.7A) further demonstrated the bias towards this type of DMR. We also found that Family A and C DMRs showed enriched association with several laminopathy disease categories, while shared DMRs showed no enrichment within these categories (Figure 2.7A). This observation corroborated the low number of TFBS motifs that were associated with categories related to laminopathy-impacted tissues ("cardiac development", "limb development", "lipid metabolism") that we noted previously (Figure 2.4H). Both families equally enriched for a variety of cardiovascular diseases, including both cardiac remodeling and hypertensive diseases, which supported the DCM phenotype observed in both families. Despite patients not exhibiting hypertensive disease, both sets of family-specific DMRs enriched for this phenotype, which has been shown to lead to excessive remodeling of the myocardium, resulting in the development of DCM [82]. Similar to our motif enrichment, we also observed a strong enrichment for diseases associated with skeletal malformations in Family C DMRs. Indeed, brachydactyly, which Family C patients exhibit, was the most enriched laminopathy-related ontology associated with our Family C DMR dataset. Family A DMRs instead favored diseases related to neuro-muscular phenotypes. Surprisingly, we also observed the presence of kidney-related disease terms in genes associated to Family A DMRs. Although not widely recognized as a form of laminopathy, several studies have documented the occurrence of kidney-related diseases in patients with *LMNA* mutation-induced lipodystrophy or DCM [83,84]. A large majority of the remaining disease ontologies (Appendix 2.7) were found to be involved in either cancer (21%) or nervous system disorders/abnormalities (50%). The documented low levels of lamins in several types of cancers [85] and the known involvement of neurodegeneration [86] and neuropathies [87] in laminopathies could account for some of these observations.

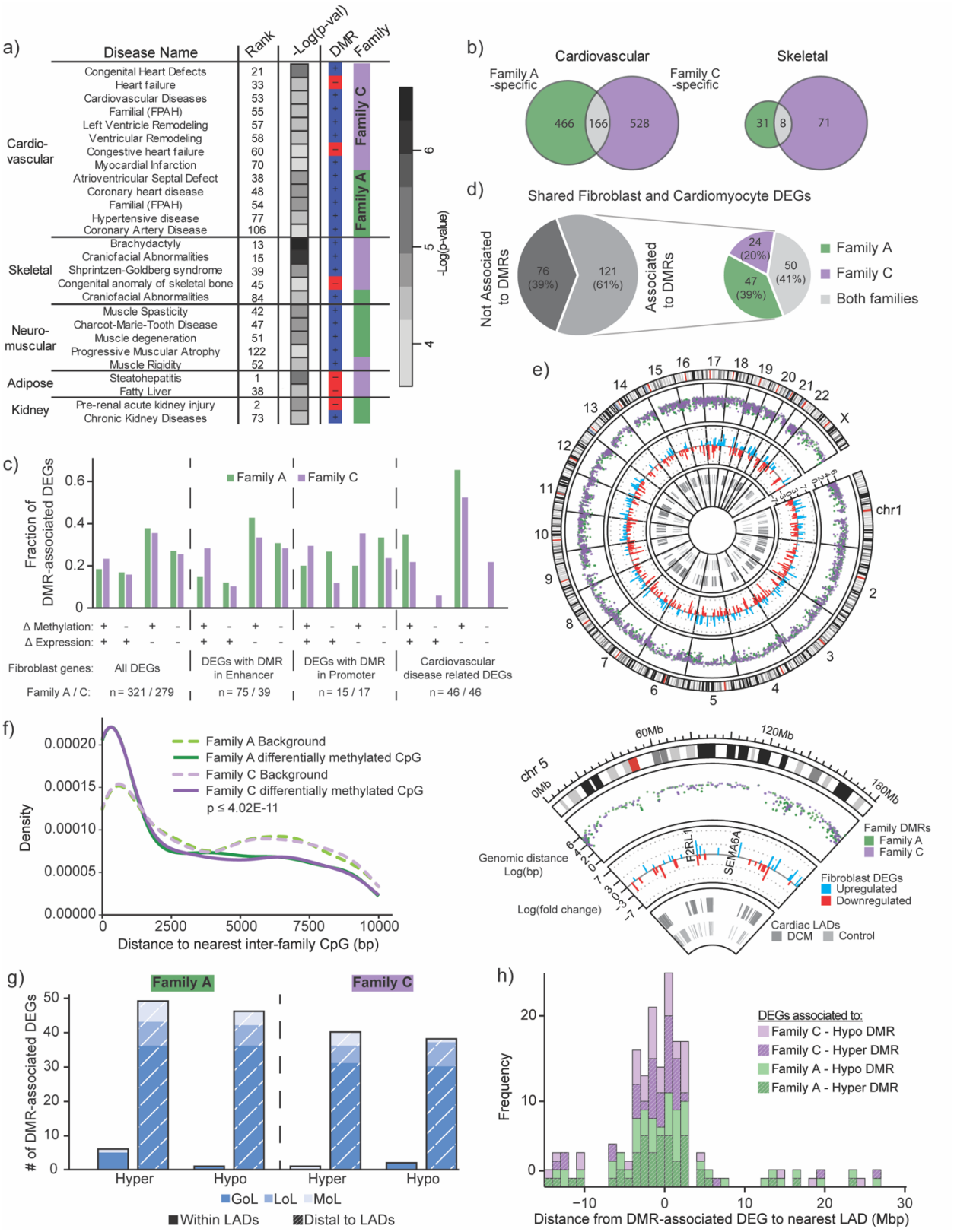


**Figure 2.7: DMRs associate to dysregulated and disease-relevant genes near redistributed**

**LADs. A**, Disease ontology terms enriched in DMRs, grouped by disease type. Heatmap reports the degree of statistical significance for enrichment. Results were categorized as hypomethylated (red) or hypermethylated (blue) by type of DMR associated to a particular disease. **B**, Number of genes in cardiovascular and skeletal disease associated to Family A-specific and Family C-specific DMRs. **C**, Top, Fraction of DMR-associated fibroblast DEGs present in one of four combinatorial groups of differential methylation (Δ Methylation) and differential gene expression (Δ Expression). Middle, (+) indicate patient > control, while (-) indicate patient < control for both differential methylation and gene expression. Bottom, Category of fibroblast DEGs and number of DEGs by family (Family A / Family C). **D**, Number and percentage of DEGs shared between fibroblast and cardiac tissue associated with DMRs in Family A only, Family C only, or both. **E**, Circos map of the genome (Top) and zoomed in chromosome 5 (Bottom). Outer to inner rings represent the following: Track I - genomic distance (log 10) between DMRs within Family A or Family C. Track II - fold change (log 2) of fibroblast DEGs, highlighting two genes found within the top 10 most differentially expressed. Track III - location of LADs in cardiomyocytes from either *LMNA*-related DCM or control samples from prior study [6]. **F**, Density of genomic distance to the nearest inter-family CpG for differentially methylated CpGs and a random sample of CpGs. Wilcoxon rank sum test p-value is displayed. **G**, Number of DEGs shared between fibroblast and cardiac tissue associated to DMRs in Family A or Family C falling within or distal to redistributed LADs (Gain of LAD (GoL), Loss of LAD (LoL), or Maintenance of LAD (MoL)). **H**, Stacked histogram of the distance between DMR-associated DEGs, shared between fibroblast and cardiac tissue, and the nearest redistributed LAD.

### 2.2.5 Genes dysregulated in both fibroblast and DCM cardiac tissues associate with DMRs from both families and LADs

The lack of enrichment for diseases related to tissues affect by laminopathies in genes associated with shared DMRs led us to focus on family-specific DMRs only. Given that both family-specific DMR sets were enriched for cardiovascular and skeletal disease ontology categories, we evaluated for inter-family gene overlap within each of the corresponding disease-associated gene sets. Unexpectedly, we found no overlap for the majority of these genes including those in the cardiovascular category despite both families exhibiting DCM (Figure 2.7B).

To examine this more thoroughly, we first compared our DMR data with the list of differentially expressed genes (DEGs) between patient and their unaffected controls previously obtained from transcriptome-wide expression data from Family A fibroblasts [88] (Figure 2.7C). To ascertain the potential role of aberrant DNA methylation on differential expression, we compared the direction of methylation change within DMRs to the direction of expression change for associated DEGs. Genome-wide DMRs associated to both families were weakly inversely correlated (54.5%, Quadrant Count Ratio (QCR) = -0.09 for Family A and 51.2%, QCR = -0.03 for Family C), with expression changes (i.e. higher methylation level in patients compared to controls (+) was associated with lower gene expression (-)) (Figure 2.7C). Notably, analysis of DEGs present within our previously identified cardiovascular disease-related gene list (Figure 2.7B), also showed an inverse correlation between methylation and gene expression though more pronounced in both families (65.2%, QCR = -0.30 for Family A and 56.5%, QCR = -0.13 for Family C). Most of these correlations were the result of hypermethylation association to decreased expression. This observation corroborates our previous observations of hypermethylation also being associated with disease-related genes (Figure 2.7A).

When broken down into DEGs associated to DMRs located within gene enhancers, we noted that this bias was also present. However, we did observe more DEGs were inversely correlated with DMR methylation changes in Family A (QCR = -0.09), unlike in Family C (QCR = 0.13). In the promoter context, however, DMRs associated to DEGs did not show any negative trends with expression changes (QCR = 0.07 for Family A and QCR = 0.06 for Family C). These findings are

consistent with a more important regulatory function for enhancer-located DMRs in Family A compared to Family C, and a lack of association in both families for DMRs in upstream promoters (Figure 2.4G), also observed in a prior study on *LMNA*-related DCM cardiac tissues [6].

To relate our DMR data to DEGs observed within a more physiologically relevant context, we identified DEGs found in both our patient fibroblasts and within DCM patient cardiac tissues from a prior study [6]. Interestingly, 61% of the 197 conserved DEGs were associated to a DMR from at least one of the families (Figure 2.7D). Remarkably, despite the lack of inter-family overlap seen for disease-related genes (Figure 2.7B), 41% of DEGs in this category were found to associate with at least one DMR from both families. Given this overlap, we wondered if inter-family DMRs occurred in close genomic proximity more broadly. To explore this, we compared the density distributions of CpG proximity in DMRs for each family and random background (Figure 2.7F, Wilcoxon rank sum test: p-value $\leq 4.02 \times 10^{-11}$ for both families). We found that differentially methylated CpGs (DMCpGs) indeed showed a greater density bias towards smaller inter-family distances (median for Family A: 2192.5bp, Family C: 2036.5bp) compared to the random background (median for Family A: 3640bp, Family C: 3645bp), up until about 1450bp. This proximity between Family A and C DMRs was also observed in our circos and rainfall plot analysis (Figures 2.7E, 2.8).

Given these results along with our previous observations that DMRs, in general, tended to associate with epigenomic features that co-localize to the nuclear periphery (Figure 2.4F,G), we next analyzed the proximity of DMRs associated with conserved DEGs (between fibroblasts and cardiac tissues) to LADs known to be dynamic (or redistributed) in *LMNA*-related DCM [6] (Figure 2.7E). In addition to two previously defined domain redistribution categories [6], Gain of LAD (GoL) and Loss of LAD (LoL), genomic regions were also assigned to Maintenance of LAD (MoL). Of the DMR-associated DEGs found in both fibroblasts and cardiac tissues, we found that only a small fraction fell directly within a redistributed LAD (0 to 6.2% for GoLs, and 0% for LoLs) or MoLs (0 to 2.1%), comparable to those previously observed in DCM tissues [6]. The remainder

of the DMR-associated DEGs were mostly distal to GoLs (73.5 to 78.9%) (Figure 2.7G). Moreover, identified DMR-associated DEGs were found be significantly more likely to fall within 2Mbp of their closest redistributed LAD (Figure 2.7H) than outside of that range (logOR = 0.50, $p=1.31x10^{-7}$). Interestingly, chromosome 19 did not contain any conserved DEGs distal to redistributed LADs (Figure 2.9).

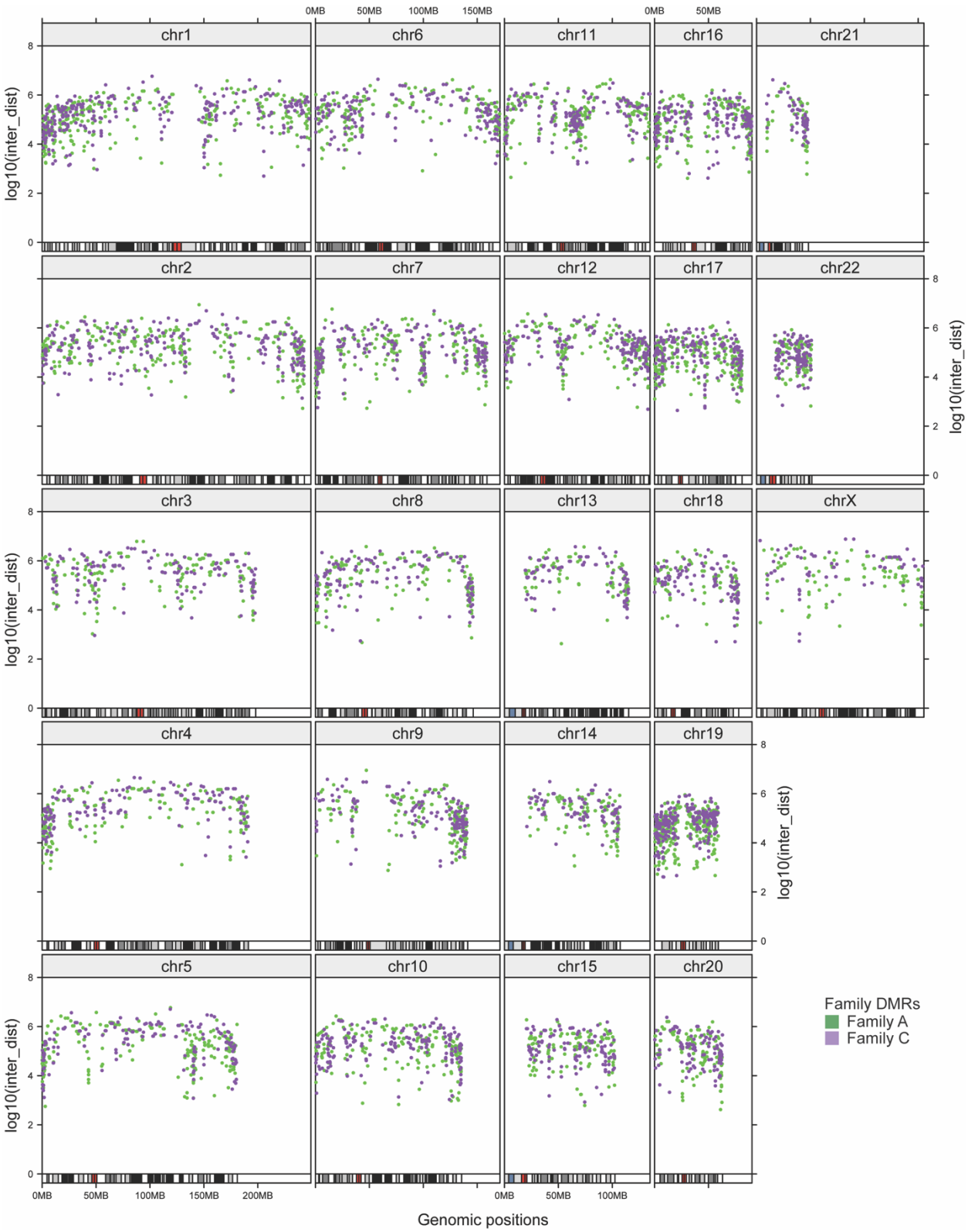


**Figure 2.8: Inter-DMR distances overlap across both families in all chromosomes.** Top, Trellis rainfall plot showing genomic distance (log10 bp) between DMRs within Family A (green)

or Family C (purple) for chromosomes 1 through 22 and X.

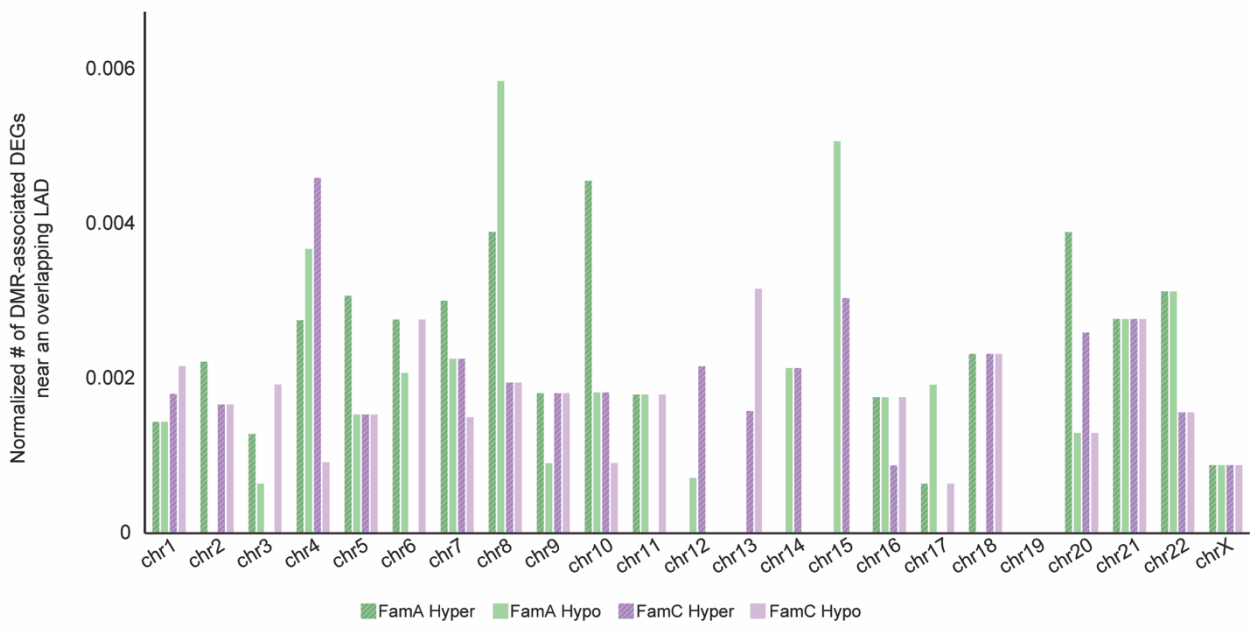


**Figure 2.9: Chromosomal distribution of conserved DMR-associated differentially expressed genes (DEGs) near redistributed LADs.** Bar plot showing the number of DMR-associated DEGs conserved between fibroblast and cardiomyocyte samples that fall near a redistributed LAD at a particular chromosome. The value is normalized to the number of genes present within the specified chromosome.

### 2.2.6 Reprogramming reveals epigenetic hotspots for aberrant methylation during early development

Given that patients in Family C presented with developmental abnormalities in bone formation (brachydactyly), we wanted to see if our *in vitro* cell system could be used to better understand the influence of DNA methylation epimutations in the early stages of development. We therefore performed similar studies in iPSC, as an early developmental model of LMNA mutations. Unlike in fibroblasts, hierarchical clustering of iPSC samples based on DNA methylation from all chromosomes (Figure 2.10A) or autosomal chromosomes only (Figure 2.5F)

did not cluster according to family. This confirmed our expectation that reprograming would lead to massive epigenetic remodeling and resetting (at least partially) of somatic methylation patterns that might have arose due to family-specific conditions [89–91]. Despite this global change in DNA methylation levels (Figure 2.1D), we still identified DMRs in patient iPSCs across each category (Figure 2.10B). However, the number of DMRs found in iPSCs (2674 DMRs) were still only ~¼ of the number found in fibroblast (10578 DMRs). Direct overlap between fibroblast and iPSC DMRs was greatest in Family C by almost 3-fold (19.6% compared to 6.7% for Family A and 1.9% for our shared category) (Figure 2.10C). In addition to the greatest amount of intercell-type DMR overlap, Family C had the largest fraction (0.97 versus 0.52 for Family A) of overlapped DMRs with conserved directionality (hyper or hypomethylated).

We also found that iPSC DMRs varied in their association to histone modifications compared to their fibroblast counterparts (Figure 2.10D, Fisher's exact test: p-value $\leq 0.05$ unless specified as non-significant in Appendix 2.8). Particularly, we saw an increased presence of iPSC DMRs in H3K9me3 and H3K27me3, further highlighting the presence of aberrant methylation in the compacted and silenced regions of chromatin. We also observed an overall increase in odds ratio at H3K4me3, a histone mark enriched at active promoters[77].

Although direct overlap of DMRs across cell types was low (Figure 2.10C), we observed genomic regions where iPSC DMRs were in close proximity to fibroblast DMRs (Figure 2.10E), which made us wonder if regions highly susceptible to epimutations were conserved between fibroblast and iPSC states. We therefore compared the distance between CpGs in iPSCs and their closest neighboring CpG in fibroblast for both a randomized set of CpGs and our DMCpGs (Figure 2.10F, 1-tailed Fisher's exact test: p-values $\leq 0.05$). Interestingly, compared to our randomized background, 3.3 and 6.5 times more CpGs fell within 1kb of each other between the two cell types in Family A and Family C, respectively. This fold difference decreased in both families for bins of larger inter-CpG distances. Moreover, when we focused on genes associated

to DMRs in iPSCs and fibroblasts, we found a large amount of overlap between the two gene sets (Figure 2.10G). Specifically, 59.8% and 61.6% of genes that were associated with an iPSC DMR were also associated to a fibroblast DMR in Family A and Family C, respectively (Figure 2.11). We also saw a comparable number of DMR-associated genes that switched in the methylation change direction between fibroblasts and iPSCs (e.g, hyper → hypo, or hypo → hyper) for both families.

Analysis of these DMR-associated genes showed enrichment for laminopathy-related disease ontologies (Figure 2.10H, full list shown in Appendix 2.9). Family A showed enrichment only in genes associated with DMRs hypermethylated in fibroblast and hypomethylated in iPSCs. In contrast, Family C enrichment in all categories of DMRs except those that were uniquely found in iPSCs. Most notably, genes associated with Family C DMRs hypermethylated in fibroblast but hypomethylated in iPSCs showed specific enrichment for brachydactyly, abnormality of the skeletal system, and congenital abnormality. Genes associated to Family C DMRs hypomethylated in fibroblast but hypermethylated in iPSCs enriched for *LMNA*-related DCM. All four diseases were ranked in the top 10 diseases, and, interestingly, both the skeletal disease-associated DMRs and brachydactyly phenotype were unique to Family C [92].

To gain further insight into disease mechanism in our early development model, we performed protein-protein interaction network analysis, using STRING. The list of 519 genes for Family C DMRs hypermethylated in fibroblast but hypomethylated in iPSC (Figure 2.10G) was filtered for association to *LMNA*-related DCM (Concept ID: C1449563) and Congenital Abnormality (Concept ID: C0000768), both of which are phenotypes that Family C patients exhibited. The resulting STRING output included a large interaction network that included 28 genes with high confidence interactions (Figure 2.10I). Of the genes associated to congenital abnormality, four genes (*HDAC4*, *PTCH1*, *EHMT1*, *SYK*) were associated to brachydactyly, according to the DisGeNET [93] database. Interestingly, *LMNB1*, which codes for one of the two types of B-type lamins and is associated to DCM [94], was present within this network. Within this

network, *CCND1*, the most connected node (8 associations), was involved in three pathways (Wnt signaling, Hedgehog (Hh) signaling, and the cell cycle) found to be enriched in this gene set (Appendix 2.10). Another 60.7% of the genes in this network were previously identified as DEGs in hearts from *LMNA*-related DCM patients [6], substantiating that our analysis was able to reveal a highly-networked set of disease-associated genes that may be dysregulated due to methylation changes linked to *LMNA* mutations.

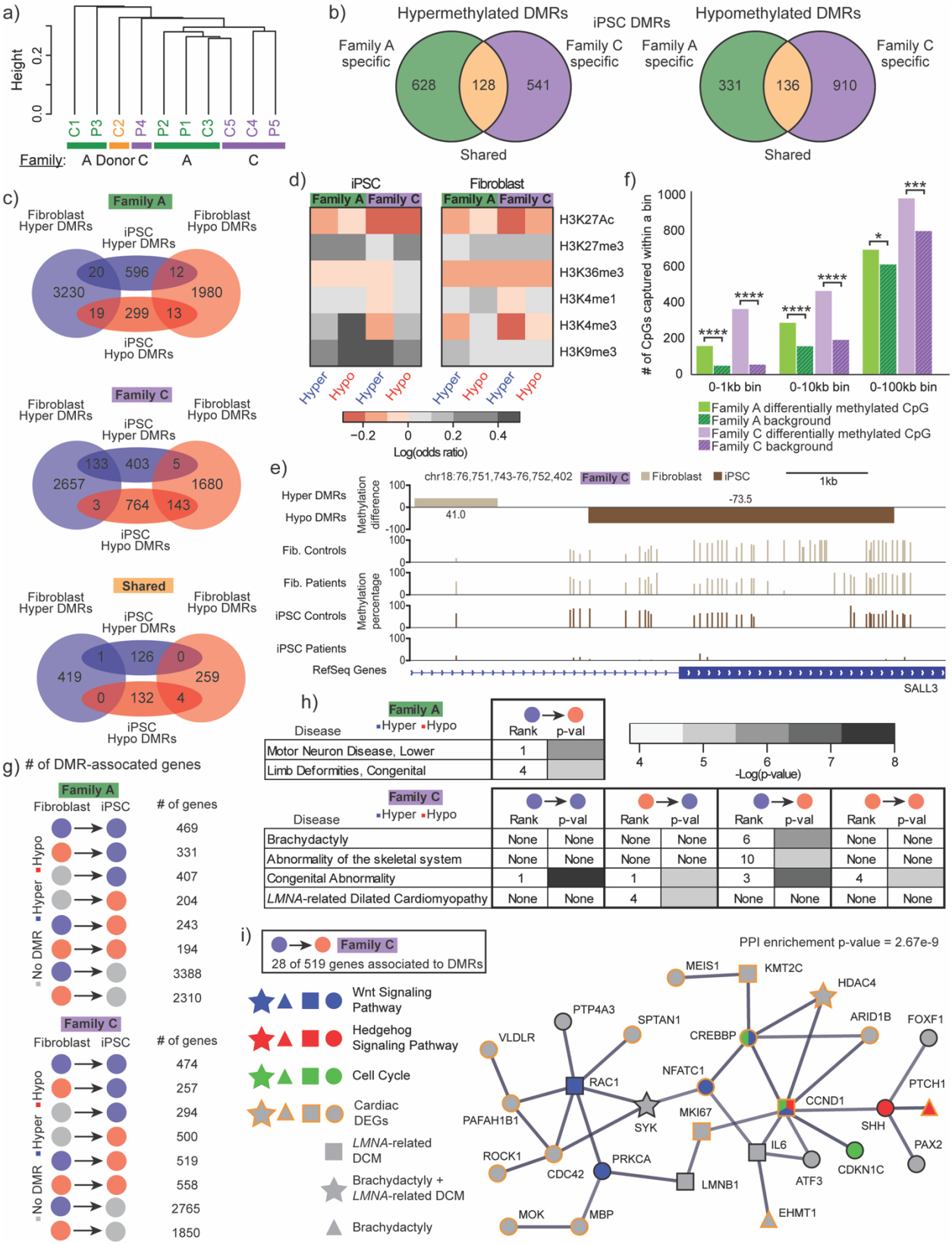


**Figure 2.10: DMRs in iPSCs reveal tissue-persistent epimutation hotspots at**

**developmentally and laminopathy relevant genes. A**, Hierarchical clustering of iPSC samples by genome-wide DNA methylation. Colors represent family groups. **B,** Venn diagram showing the number of DMRs captured by group for both hypermethylated and hypomethylated DMRs. Orange regions denote “Shared DMRs”, green regions denote “Family A-specific DMRs”, and purple regions denote “Family C-specific DMRs”. **C**, Number of DMRs captured within fibroblast and iPSC samples for each grouping for hypermethylated and hypomethylated DMRs. **D**, Log odds ratio of a CpG falling within both a DMR group and a given histone modification in iPSC and fibroblast. **E**, Example of Family C DMR proximity in both cell types. Top, Genome browser track displaying DMRs based on mean methylation differences (patient minus control) in fibroblasts and iPSCs. Middle, Methylation levels for patient and control samples for each cell type. Bottom, Depiction of RefSeq gene annotation. **F**, Number of either differentially methylated CpGs or randomly sampled CpGs in iPSC that fell within a range of genomic distances from their closest neighboring fibroblast CpG in the same family; Fisher’s exact test: * $P \leq 0.05$, *** $P \leq 0.001$, **** $P \leq 0.0001$ **G**, Diagram depicting the number of genes associated with DMRs falling within one of eight categories of DMR methylation patterns in fibroblast and iPSCs. **H**, Table highlighting laminopathy-related disease ontologies enriched in DMRs grouped by fibroblast and iPSC DMR state (hyper- or hypomethylated). Heatmap reports the degree of statistical significance for disease enrichment. **I**, Protein-protein interaction (PPI) network of 28 genes associated to Family C-specific DMRs (hypermethylated in fibroblasts and hypomethylated in iPSCs) and either *LMNA*-related dilated cardiomyopathy (DCM), congenital abnormality, or both. Pathway enrichment and disease association are denoted by color and shape, respectively. Orange node borders indicate that the gene is differentially expressed in cardiac tissue (cardiac DEG).

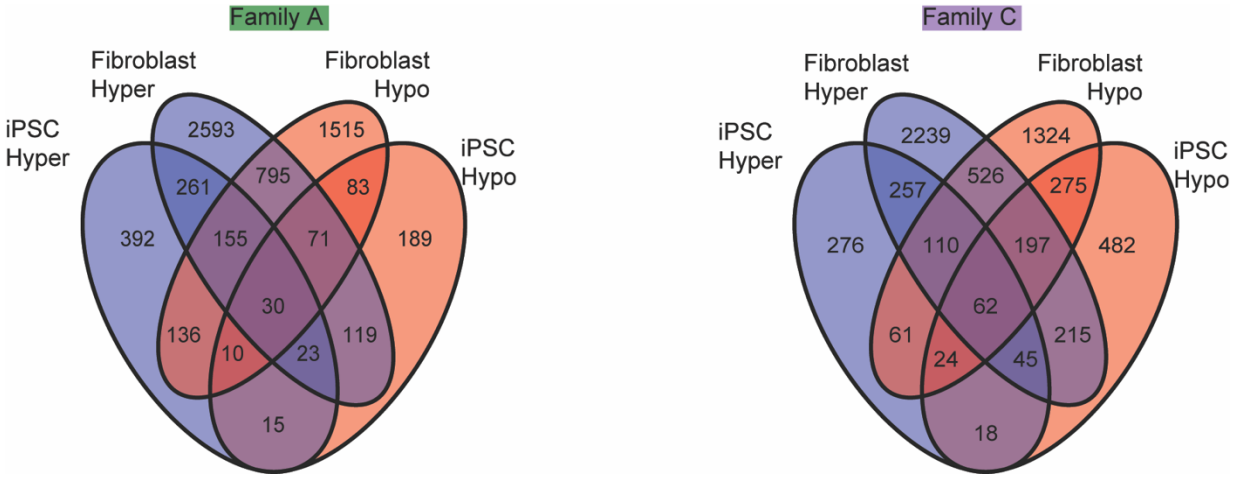


**Figure 2.11: DMRs in iPSCs reveal tissue-persistent epimutation hotspots at developmentally and laminopathy relevant genes.** Venn diagrams showing the number of genes associated to hyper methylated (blue) and hypomethylated (red) DMRs groups overlapped between both fibroblasts and iPSCs for Family A (left) and Family C (right).

We performed a comprehensive analysis of differential DNA methylation for ten matched pairs of fibroblasts and iPSC from DCM patients in two families with distinct *LMNA* mutations and their unaffected sibling controls. Our results provide new insight into mutation-specific mechanisms that influence both common and unique aspects of phenotypic expression of laminopathies.

First, our observations suggest that aberrant DNA methylation in *LMNA*-mutated cells affect not only normally silenced regions of the genome but also previously unappreciated regulatory features such as enhancers and downstream promoters. Although large differences in methylation level were not observed from genome-wide averages in either cell type, closer inspection of the RRBS data at a regional level, revealed DMRs in *LMNA*-mutant samples compared to controls. In fibroblasts, we observed an increased likelihood of finding CpGs in DMRs falling 1-5Kb downstream of TSS and distally upstream of the gene promoter. Along with

DMR association to relevant histone marks such as H3K4me1 [76,77], this suggests that Family A DMRs serve a more important regulatory function as enhancers relative to Family C DMRs, and that neither Family DMRs had much association to upstream promoters as previously shown [6]. In contrast, the association of iPSC DMRs to H3K4me3 suggested that the regulatory mechanism most impacted by differential methylation in this cell type is at promoters. In addition, the association of fibroblast and iPSC DMRs to histone modifications related to both heterochromatin and LADs suggests that, despite each of our families showing largely unique DMR landscapes, both families experience epimutations within these normally silenced regions of the genome, which could contribute to (or be associated with) the dysregulated of genes. This concept adds to the previous observation that altered CpG methylation was associated to redistributed LADs and gene dysregulation in DCM hearts [6].

Second, our results for DMRs identified multiple epimutation hotspots in the genome across all samples that may play an important role in the expression of DCM, a common laminopathy phenotype. Several shared DMRs were notably associated to genes in close genomic proximity to one another (ex. *HOXD10* and *HOXD12*), and fibroblast DEGs associated to family-specific DMRs showed a substantial amount of inter-family overlap. These inter-family epimutation hotspots were supported with observations in fibroblasts that the distance between inter-family DMCpGs had a higher density bias at short genomic distances than a random background. Furthermore, despite shared DMRs having little to no association to TFBS motif pathways and disease ontologies related to laminopathies, a relatively larger number of DMR-associated genes related to cardiovascular disease were present in both Family A and C. Thus, the identification of these epimutation hotspots across samples from families with distinct *LMNA* mutation suggests that family-specific aberrances in DNA methylation might lead to common functional consequences in DCM.

Our findings for a common subset of laminopathy epimutations in family-specific DMRs, in conjunction with LAD redistribution, also suggest a significant role of Lamin A/C in epigenetic

regulating mechanisms of laminopathy-related pathways in multiple affected tissues but insufficient to express disease phenotype. The close proximity of family-specific DMRs at epimutation hotspots and silenced chromatin could explain our observation that both sets of family-specific DMRs had overlapping DEGs, shared between fibroblast and cardiac tissue DCM samples. This commonality between the two families further extended to DMR-associated DEG localization outside of redistributed LADs. Interestingly, family-specific DMRs also both showed enrichment for disease in laminopathy-related tissues outside of those affected in patients (e.g., neuromuscular, adipose, and kidney). Family A DMRs, for example, enriched for "Charcot-Marie-Tooth disease", known to be caused by a *LMNA* mutation [87], despite neither family having muscular dystrophy. Furthermore, a previous study of patients with DCM revealed a GO term enrichment for "lipid metabolism" in genes with transcript level correlated to their associated methylation status and LAD localization [6]. Another study on Emery-Dreifuss muscular dystrophy (EDMD) similarly suggested that nuclear envelope disorders could account for a unifying molecular model responsible for the wide range of laminopathy phenotypes [72].

In addition to a possible common laminopathic mechanism, our study identified family-specific epimutations with unique regulatory functions in chromatin remodeling, disease mechanism, and phenotypic expression. Until now, DNA methylation studies using samples from DCM patients did not consider the role for specific *LMNA* mutation in affected families [6,95]. The individual impact of specific mutations is further highlighted by the previous observation that expression of the *LMNA* mutation responsible for familial partial lipodystrophy did not induce epigenetic alterations of myogenic loci in a human myogenic cell line unlike the *LMNA* mutation involved in EDMD [72]. In our study, the presence of divergent mutation-specific epimutations is apparent in the limited overlap of disease-related genes associated to Family A and C DMRs. Family C DMRs were particularly interesting due to the strikingly significant enrichment for disease ontology of brachydactyly, a unique phenotype in patients from Family C [92]. *De novo* enhancer-promoter interactions from the disruption of topology associated domains (TADs)

previously was demonstrated to result in ectopic gene expression and subsequently brachydactyly [96]. The significant presence of many redistributed LADs, mostly GoLs, within 2Mb, the maximum distance for enhancer-promoter interacting pairs [97], of DEGs associated to DMRs in Family C further supports the involvement of TAD restructuring. It is therefore conceivable that the aberrant methylation observed at enhancers is a signature remnant of disease-induced chromatin remodeling.

In iPSC samples, the presence of mutation-specific epimutations also supports a disease mechanism during early development. Despite little direct overlap between iPSC and fibroblast DMRs, Family C hypermethylated and hypomethylated DMRs were more conserved from fibroblast to iPSC than DMRs in Family A. The presence of retained epimutations further supported Family C's involvement in the iPSC's primed pluripotent state. Paradoxically, the subset of Family C DMRs, which reversed methylation directionality from being hypermethylated in fibroblasts to hypomethylated in iPSCs, was associated to developmental genes implicated in skeletal malformations, echoing the family's unique brachydactyly phenotype. This suggests aberrant increases and decreases in DNA methylation in regions more susceptible to epimutations is important in disease pathogenesis.

Finally, the set of genes associated to these reversed Family C DMRs, when filtered, provided us with a particularly interesting network of protein-protein interaction that provides further involvement of the Wnt signaling pathway and cell cycle regulation in the disease mechanism of laminopathies for DCM. Despite Family C patients having skeletal involvement, our network showed a specific association also to cardiac disease in several ways. Foremost, over half of the genes identified within our network was previously identified as DEGs in hearts from DCM patients [6]. Additionally, the Wnt signaling pathway, enriched in our network, is known to be involved in heart development and disease [98,99] and dysregulated in *LMNA*-mutated mouse models of DCM [100]. In parallel, Wnt proteins regulate the cell cycle, itself involved in cardiac development and disease [101]. Specifically, cell cycle-related GO terms previously were

observed in genes associated to redistributed LADs with altered CpG methylation and differential expression in cardiac tissue from *LMNA*-related DCM patients [6,102]. Furthermore, cell cycle progression is tightly regulated during cardiac development, with the exit of G1 phase mediated through E2F transcription of its target genes [101]. Despite not being associated to cell cycle, the expression of *LMNB1*, encoding for Lamin B1, previously was shown to be regulated by E2F as part of cell cycle progression [103]. The presence of Lamin B1 is especially significant in the context of iPSCs since this isoform is expressed in early embryo and differentiating cells, unlike Lamin A/C which is expressed primarily in differentiated somatic cells [69]. E2F TF target genes previously were shown to be dysregulated in *LMNA*-mutated cardiomyocytes with DCM [102]. Of the dysregulated E2F target genes [102], three (*CCND1*, *CDKN1C*, *MKI67*) were identified in our network. *CCND1*'s involvement in cardiac disease is supported by its presence in both the cell cycle and Wnt [104] signaling and previous observations of upregulation in DCM [102,105]. Interestingly, a previous study of EDMD also implicated E2F and cell cycle dysregulation as a key feature of the disease mechanism [72].

In addition to DCM, our protein-protein interaction network provides further involvement of the Hedgehog (Hh) signaling pathway and cell cycle regulation in the disease mechanism for brachydactyly. *CCND1*, as mentioned above, encodes for Cyclin D1 that also is involved in Hh [106] signaling, an important regulating pathway in limb development [107]. *SHH*, one of the three Hh proteins, has specifically been shown to be tightly regulated by a long-range enhancer region, whose disruption can lead to *SHH* dysregulation and subsequent finger malformation [107]. The relevance of our network in finger malformation was further highlighted by the presence of genes involved in brachydactyly (*HDAC4*, *PTCH1*, *EHMT1*, *SYK*). Of particular note, *HDAC4* is considered highly associated to brachydactyly (2$^{nd}$ highest gene-disease association according to the disease database DisGeNET [93]), due in part because of its direct involvement in inducing brachydactyly mental retardation syndrome (BDMR) [108,109]. Additionally, *PTCH1* has also been previously involved in brachydactyly as part of Hh signaling [110]. Together, these results

suggest that epimutations at important cell cycle genes such as *CCND1* could provide a molecular link for how both cardiovascular disease and limb malformation may be present in patients.

## 2.3 Materials and Methods

### 2.3.1 Fibroblast and iPSC lines

Ten matched pairs of PATIENT and CONTROL fibroblasts and iPSC lines were used in this study (Figure 2.1A and Table 2.1). For the PATIENT group, dermal fibroblasts were cultured from skin biopsies obtained from five affected individuals of two *LMNA* study families (A & C) as previously reported [92,111]. Family A includes three patients (P1, P2, and P3) heterozygous for *LMNA* splice-site (c.357-2A>G) that exhibit sick sinus syndrome and DCM leading to heart failure [111]. Family C includes two patients (P4 and P5) heterozygous for *LMNA* missense (p.Arg335Trp) mutation displaying conduction disease, DCM, and brachydactyly, similar to HHS IV [92]. For the CONTROL group, dermal fibroblasts were cultured from skin biopsies obtained from four unaffected siblings (C1, C3, C4, and C5) and from a purchased sample obtained from one healthy, unrelated "Donor" individual (C2) (CC-2511, Lonza, Basel, Switzerland). Fibroblast culture and genomic DNA (gDNA) extraction were performed as described previously [92,111]. By Sanger sequencing of all 12 *LMNA* exons in fibroblast DNA, presence or absence of the LMNA mutation was confirmed in all PATIENT and CONTROL lines, respectively.

**Table 2.1**

**Fibroblast and iPSC line pairs with corresponding genotype, sex, and age when skin biopsies were performed (N = 10)**

| Cell ID | Family | Genotype* | Sex | Age at Skin Biopsy (years)† |
|---|---|---|---|---|
| P1 | A | +/- | F | 38 |
| P2 | A | +/- | M | 62 |
| P3 | A | +/- | F | 70 |
| P4 | C | +/- | M | 51 |
| P5 | C | +/- | M | 29 |
| C1 | A | +/+ | F | 49 |
| C2 | Donor | +/+ | M | 51 |

| | | | | |
|---|---|---|---|---|
| C3 | A | +/+ | F | 68 |
| C4 | C | +/+ | F | 60 |
| C5 | C | +/+ | M | 26 |

(*) Genotype: +/+ homozygous normal, +/- heterozygous *LMNA* mutation. (†) Age: average age ± SD of Control (50.8 ± 16) vs. Patient (50 ± 17) is not significantly different p> 0.05, (t-test).

To generate matched iPSC lines, the PATIENT and CONTROL fibroblasts were reprogrammed using the CytoTune-iPS 2.0 Sendai Reprogramming Kit (Life Technologies, Carlsbad, CA) that uses a replication-defective Sendai virus as vectors to introduce reprogramming factors (OCT3/4, SOX2, KLF4, c-MYC) into the host cell [112,113]. Cryopreserved fibroblasts at passage 5 were revived for culture in 20% FBS (Sigma-Aldrich, St. Louis, MO) and DMEM (Life Technologies) at 37C and 5% CO2. At passage 7, fibroblasts were confirmed free of mycoplasma infection using MycoAlert Mycoplasma Detection Kit and Assay Control (Lonza) and plated at the appropriate density on 6-well plates two days prior to Sendai viral transduction to achieve 50-80% confluency. The cells were transduced (Day 0) using the calculated volumes of each virus to reach the target MOI. Twenty-four hours after transduction (Day 1), media was changed, and cells were cultured for six days with fibroblast media changes every other day. Seven days after transduction (Day 7), transduced fibroblasts were replated onto 60-mm tissue culture dishes pre-coated with recombinant Vitronectin (Life Technologies) in fibroblast medium. After twenty-four hours, medium was replaced with Essential 8 Media (Life Technologies), and cells were cultured with iPSC media changes every day. Eight days after transduction (Day 8), the cells were checked under the microscope for the emergence of cell clumps indicative of transformed cells. Three to four weeks post-transduction after sufficient growth, individual undifferentiated colonies were selected by iPSC morphology, manually picked (passage 0), and transferred to plates pre-coated with Corning Matrigel Matrix (Thermo Fisher Scientific, Waltham, MA) in TeSR-E8 media (STEMCELL Technologies) for culture at 37C and

5% CO2 with daily media changes. The iPSC clones first were passaged manually (passage 1-5) and thereafter passaged using ReLeSR (STEMCELL Technologies). For each iPSC line, independent clones were created, serially passaged, expanded, and cryopreserved in Bambanker media (Thermo Fisher Scientific) for long-term storage in liquid nitrogen.

For each iPSC line at passage 10 or above, independent clones were validated for normal pluripotency (Figure 2.12). iPSC clones were tested for positive staining by immunocytochemistry (ICC) of established pluripotency makers. For ICC, iPSCs clones for each line were grown, processed, and analyzed directly on Matrigel-coated, Nunc Lab-Tek 4-well Chamber Slides (Thermo Fisher Scientific) for pluripotent stem cell markers (OCT4, SOX2, SSEA4, and TRA-1-60) using the Pluripotent Stem Cell 4-Marker Immunocytochemistry Kit (A24881, Life Technologies). Cells were fixed, permeabilized, and incubated with blocking solution and antibodies (Table 2.2). Cells were nuclear counterstained using Fluoroshield with DAPI (Sigma-Aldrich) and visualized using a Nikon Ti-E Inverted Fluorescent Microscope.

**Table 2.2**

**Table of antibodies used for pluripotency characterization of iPSCs**

| **Primary antibodies** | | | **Secondary antibodies** | | |
|---|---|---|---|---|---|
| **Antigen (host)** | **Company, Catalog no.** | **Dilution** | **Antigen (host)** | **Company, Catalog no.** | **Dilution** |
| OCT4 (rabbit) | TFS, A24867* | 1:100 | Anti-rabbit (donkey) AF-594 | TFS, A24870* | 1:250 |
| OCT4 (rabbit) | Abcam, ab181557 | 1:500 | Anti-rabbit (donkey) AF-594 | TFS, A24870* | 1:250 |
| SSEA4 (mouse) | TFS, A24866* | 1:100 | Anti-mouse (goat) AF-488 | TFS, A24877* | 1:250 |
| SSEA4 (mouse) | TFS, 414000 | 1:500 | Anti-mouse (goat) AF-488 | TFS, A24877* | 1:250 |
| SOX2 (rat) | TFS, A24759* | 1:100 | Anti-rat (donkey) AF-488 | TFS, A24876* | 1:250 |
| TRA-1-60 (mouse) | TFS, A24868* | 1:100 | Anti-mouse (goat) AF-594 | TFS, A24872* | 1:250 |
| TRA-1-60 (mouse) | TFS, MAB4360 | 1:500 | Anti-mouse (goat) AF-594 | TFS, A24872* | 1:250 |

(*) Pluripotent Stem Cell 4-Marker Immunocytochemistry Kit (A24881, Life Technologies).

For each iPSC line at passage 9 or above, independent clones were validated for normal chromosome constitution by karyotype (Figure 2.13). iPSC cultures in Matrigel-coated T25 flasks with TeSR-E8 media were sent to WiCell Genetics (Madison, WI) for routine study of G-banded chromosomes by counting 20 cells and analyzing eight cells. Karyotype results were classified as either normal (46,XX or 46,XY) or abnormal with clonal or nonclonal findings. Clonal findings were defined as chromosome gain or structural rearrangement in at least two cells or chromosome loss in at least three cells. Nonclonal findings were defined as chromosome gain and structural rearrangements in a single cell consistent with technical artifact, developing clonal abnormality, or low-level mosaicism. If the result of the first clone was abnormal (clonal or nonclonal), a second independent clone isolated from the iPSC line was analyzed by ICC and then karyotyped. This process was repeated until at least one chromosomally normal clone was identified with validation of pluripotency.

After fibroblast reprogramming and characterization for normal pluripotency and karyotype, iPSCs were cultured from cryopreserved vials and maintained on Matrigel-coated 6-well plates with mTESR1 for gDNA extraction. At 90-100% confluency, iPSCs were harvested using ReLeSR, and gDNA was isolated using MasterPure Complete DNA Purification Kit (Lucigen, Middleton, WI). Total gDNA was then quantified using Nanodrop Spectrophotometer (Thermo Fisher Scientific).

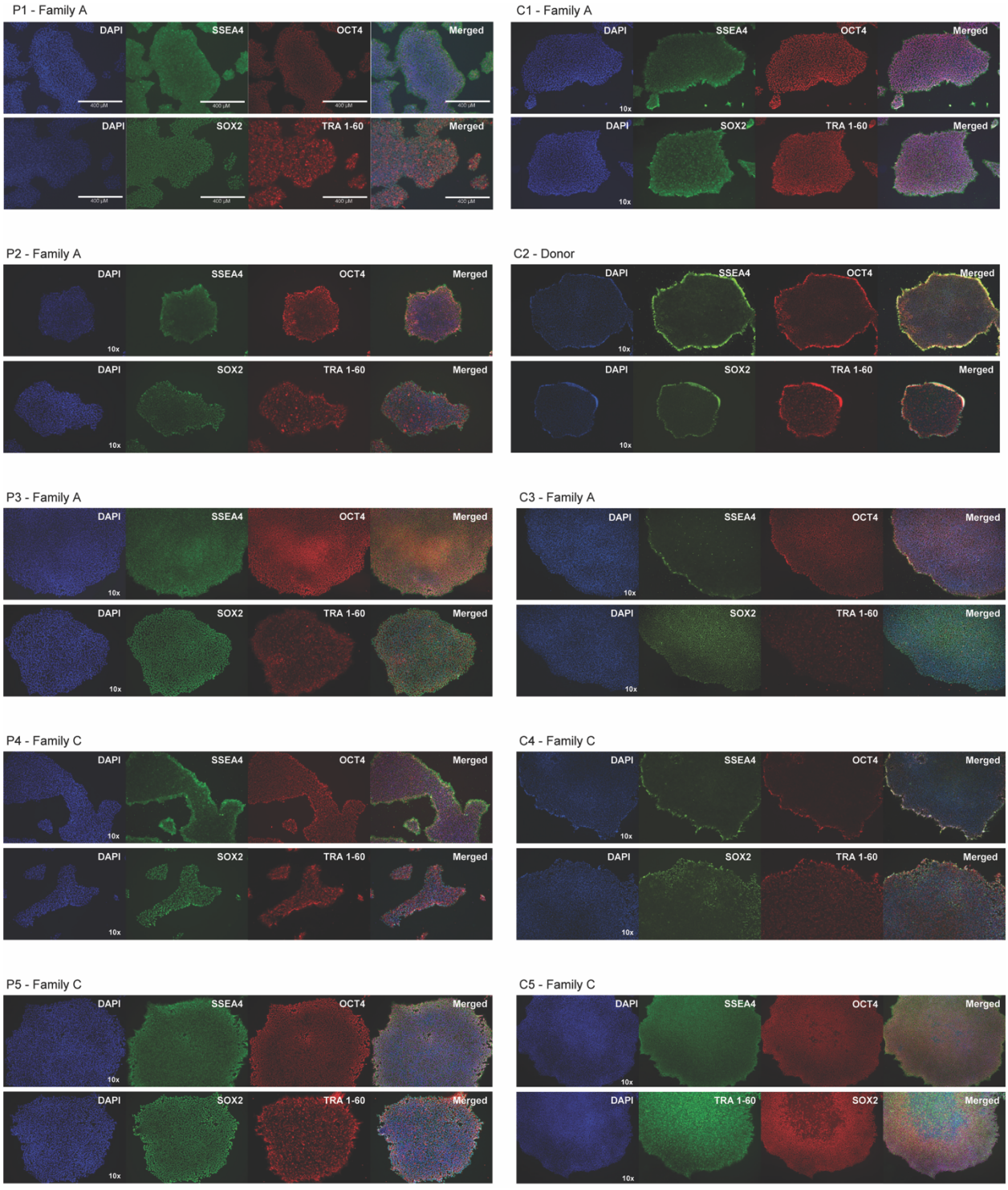


**Figure 2.12: Validation of induced pluripotent stem cell (iPSC) pluripotency.** Immunocytochemistry staining of all 10 iPSC lines for pluripotent stem cell markers: top, SSEA4 (green) and OCT4 (red) and bottom, SOX2 (green) and TRA 1-60 (red). Nuclei were visualized

with DAPI (blue). All images were taken at 10x magnification.

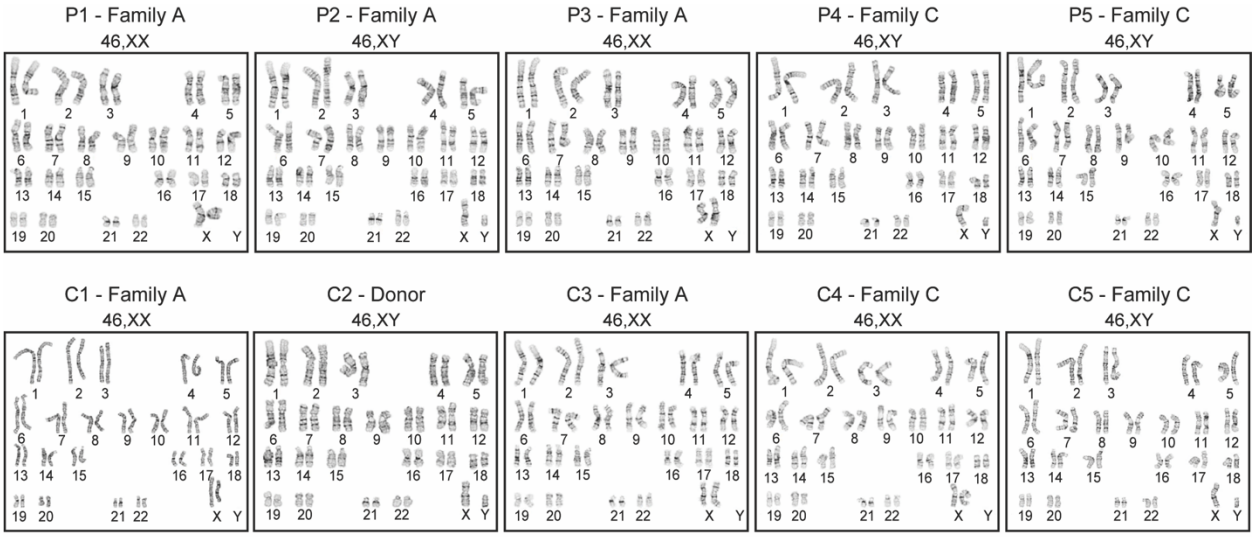


**Figure 2.13: Validation of normal chromosome constitution in each induced pluripotent stem cell clone.** G-banding metaphase karyotype of all 10 iPSC lines derived from dermal fibroblasts.

### 2.3.2 RNA-sequencing (RNA-seq) and differentially expressed gene (DEG) analysis

Bulk RNA-seq was previously performed on Family A fibroblasts (3 unaffected mutation-negative family members and 3 patients heterozygous for *LMNA* splice-site (c.357-2A>G)) and 3 healthy, unrelated individuals (Donors 2, 3, and 4). A list of DEGs between patients, control siblings, and the unrelated controls was attained from GSE125990 [88]. DEGs were filtered for FDR-adjusted p-value $\leq 0.05$. RNA-seq data for control and DCM heart tissue was accessed from GSE120836 [6]. Provided $\log_2$ fold change values of DCM over Control, filtered for genes with p-values $\leq 0.05$, were then intersected with fibroblast DEGs for analyses.

### 2.3.3 Reduced representation bisulfite sequencing (RRBS) and Differentially methylation region (DMR) analysis

Extracted gDNA from fibroblasts and iPSCs were subjected to RRBS for DNA methylation analysis. For all twenty samples, 4.5μg of DNA was first mixed with 4μL MspI (20,000 U/mL, New England BioLabs) and 1x CutSmart, and incubated at 37˚C for 24 hrs. 0.5x Agencourt Ampure XP beads (Beckman Coulter) were then used to keep fragments ≤300bp, which were then concentrated using Zymo Clean and Concentrator kit's protocol. Zymo DNA Methylation-Gold Kit was used according to manufacturer's protocol to perform bisulfite conversion on all samples, with a final volume of 15μL in elution buffer. The eluted DNA was then processed through the Accel-NGS Methyl-seq DNA library kit (Swift Biosciences), following the manufacturer's protocol, for adapter ligation. Post-ligation DNA was subjected to 10 PCR cycles for indexing. PCR products were then eluted in 21μL of low EDTA elution buffer, of which 1μL was run in a 2200 TapeStation (Agilent) to ensure correct band sizes of approximately 300bp. Pooled multiplex RRBS libraries were sent to the UCI Genomics High-Throughput Facility and sequenced on an Illumina HiSeq4000 sequencer. We performed paired-end sequencing runs for a total of 100 cycles.

Raw fastq files were trimmed by 11bp on both 5' and 3' ends of both reads 1 and 2 using Trim Galore (Version 0.4.4) [47]. Trimmed reads were then aligned to hg19/GRCh37 using Bowtie2 [48] as part of Bismark (Version 0.20.1) [49]. Paired-end read mapping efficiency varied between 68.0-82.3%, with an average of 77.4% across all twenty samples (Appendix 2.1). Bismark was used to make methylation calls, which were then merged for neighboring CpGs on opposite sides of the strand. Finally, the methylation ratios generated were filtered to keep only CpGs with a minimum read coverage of ≥ 5x, thus ensuring fair comparisons across samples.

DNA methylation data and DMRs were visualized across the hg19 genome using the Broad Institute's Integrative Genome Viewer (IGV) [114], Circos and Trellis plots generated with R packages circlize (Version 0.4.5) [115] and gtrellis (Version 1.16.1) [116]. Hierarchical clustering of samples based on genome-wide DNA methylation was performed using the ward

method as part of methylKit. Additional heatmaps of DNA methylation levels in DMRs was generated through heatmap.2 from R package gplots (Version 2.11.0) [117] was used to generate heatmap and corresponding dendrograms for DMRs.

To obtain DMCpGs, methylation call BAM files were inputted into the R package methylKit (Version 1.16.0) [118], with a specified minimum read coverage of 5 (≥5x) per sample and assembly hg19. The unite() function was then applied to compare methylation calls of ≥5x CpGs, overlapped across all input samples, generated after destranding to merge methylation calls on both sides of DNA strand at CpG dinucleotides. A filter of minimum q-value of ≤0.01 and a ±30% CpG methylation difference cutoff between CONTROL and PATIENT samples were used to ensure reliable differential methylation results. This generated a set of DMCpGs, where negative DNA methylation differences indicated scenarios where patient samples were hypomethylated relative to controls and positive differences indicated where patient samples were hypermethylated. DMRs were generated by merging neighboring DMCpGs within ±500bp of one another into a single tile. Tiles with a size <100bp were extended equally on each side until a size of 100bp was attained, similar to previously described methods [119]. Tiles containing DMCpGs with methylation differences with opposite directionality (hyper- or hypomethylation) were considered ambiguous and were removed from further analyses (0.13-0.8% of total DMRs generated) (Appendix 2.11). Methylation difference of DMCpGs falling within the same tile was averaged in the remaining DMRs. This methodology was applied with three different inputs (1) all samples, (2) Family A samples (C1, C3, P1, P2, P3), (3) Family C samples (C4, C5, P4, P5), thus yielding three categories of DMR tiles. To compare across all three categories, DMRs were filtered to keep only those with CpG methylation data overlapped in both Family A and Family C. DMR tiles from the three groups were reclassified as follows: “Family-Specific” tiles were defined as DMRs only found in one of two family DMR categories (2) or (3), described above, or found in one of the two family DMR categories (2) or (3) and in the all samples category (1). “Shared” tiles

were defined as DMRs found in both categories (2) and (3), or found only in all samples (1) and not in family categories (2) or (3), or found in all three categories (1), (2), and (3). This DMR methodology and grouping was applied to both fibroblast and iPSC samples separately. When comparing iPSC DMRs to their fibroblast counterparts, tiles were filtered to keep only those that had CpG methylation in both cell types. A detailed workflow of the computational methods used for DNA methylation analyses in this study is available at Figure 2.3.

### 2.3.4 Genomic feature annotation

To determine DMR association to inferred and experimentally derived genomic features, DMR files were annotated against ChromHMM's 25-state chromatin model [120] for normal human dermal fibroblasts (NHDFs), acquired from NIH Epigenome Roadmap, and RefSeq genomic features and histone modifications for NHDFs and a human embryonic stem cell line (HUES64), acquired from UCSC genome table browser, using BEDTools' *intersection* function [50]. Genomic promoter features were defined as 2Kb upstream of gene transcription start sites (TSS) acquired from UCSC genome table browser. Intergenic features were acquired by finding regions outside of gene bodies, against acquired from UCSC genome table browser, using BEDTools' *subtract* function. A list of double elite enhancer locations, including their associated genes, used for annotation were acquired from the GeneHancer database [121] available on the UCSC genome table browser.

### 2.3.5 Identification of gene network and ontologies from DMR-associated gene lists

Stanford's Genomic Regions Enrichment of Annotations Tool (GREAT) software [75] (Version 4.0.4) was used with default parameters (basal plus extension/proximal 5Kb upstream, 1Kb downstream, plus distal up to 1000Kb) to find hg19 UCSC genes associated to input DMR files. From there, (1) disease ontology, (2) gene ontology, (3) protein-protein interaction networks, and (4) pathway enrichment analysis were performed as follows: (1) Disease ontology was

performed on acquired gene lists using ToppFun, a part of the ToppGene suite [122], using default correction and p-value cutoff parameters (FDR correction with p-value ≤ 0.05) and “Gene Limits” increased to include the number of genes inputted. Additionally, gene lists related to diseases of interest were acquired from DisGeNET database [93] (Version 7.0). (2) Gene lists for GO terms heart development (GO:0007507) and skeletal system development (GO:0001501) were acquired from the AmiGO database [123,124]. (3) Gene lists were submitted to STRING [125] (Version 11.0b) to identify protein-protein interaction (PPI) networks. The minimum required interaction score for all PPI was set at 0.700 (considered “high confidence”) for the network. (4) The Kyoto Encyclopedia of Genes and Genomes (KEGG) [126] database was used, as part of STRING [125], to identify enriched of pathways within a PPI network. Strength scores are calculated as $\log_{10}$(observed/expected) by STRING. Enriched pathways are filtered for a false discovery rate (calculated according to the Benjamini & Hochberg method [127]) ≤ 0.05 by STRING. PPI enrichment p-value for the generated network was provided by STRING.

### 2.3.6 Determining differentially methylated transcription factor binding sites (TFBS)

DMR files, in BED format, were inputted into Hypergeometric Optimization of Motif EnRichment (HOMER) software [58] (Version 4.7) to identify enrichment of known TFBS motifs, reposited within the software’s vertebrae database. Analyses were performed with hg19 genome as background, along with a specified motif size parameter based on average DMR tile size. TFBS motif results were finally filtered for p-value ≤ 0.01. Known related categories for each transcription factor (TF) were determined using GeneCards’ Human Phenotype Ontology (HPO) and SuperPathways databases[128].

### 2.3.7 Lamina-associated Domain (LAD) redistribution analyses

*LMNA* peaks, generated by anti-Lamin A/C ChIP-seq, from cardiomyocytes derived from DCM patients and control individuals were acquired from GSE120837 [6]. In order to determine the location of redistributed LAD, BEDtools' *subtract* function [50] was used to compare DCM and control LAD locations. Gain of LAD (GoL) regions demarcated LAD locations that were present in diseased tissues but absent in unaffected donors. Loss of LAD (LoL) regions demarcated LAD locations that were present in unaffected donors but absent in diseased tissues. Regions where LADs were present in both control and diseased tissues were termed MoL (maintenance of LAD) regions.

LADs from normal human primary dermal fibroblast (AD04) were acquired from GSM1313399 [129] and compared to the aforementioned cardiomyocyte redistributed LADs to identify LADs conserved across both cell types. Fibroblast LADs locations were compared to those of the three LAD categories (GoL, LoL, and MoL) generated in the cardiomyocyte samples. Genomic regions identified as cardiomyocyte GoLs that did not overlap with a fibroblast LAD were kept for downstream analyses. Similarly, genomic regions annotated as LoLs and MoLs in cardiomyocytes that overlapped with a fibroblast LAD were retained for further analyses. Distance between DEGs and closest redistributed LADs were determined using BEDtools' *closest* function [50].

### 2.3.8 Statistical Analyses

All statistical tests were performed through R (Version 2.15.2) [63]. Data distributions were first tested for normality using the Shapiro-Wilks test. The Kruskal-Wallis and Wilcoxon rank sum tests were performed for datasets with non-normal distribution.

Quadrant count ratio (QCR) was calculated as $\frac{n(Quadrant\ I)+n(Quadrant\ III)-n(Quadrant\ II)-n(Quadrant\ IV)}{N_{total}}$, where n(Quadrant) is the number of observations present within a given quadrant, and $N_{total}$ is the total number of observations across

all four quadrants.

Odds ratio (OR) analyses were performed to determine the significance of DMR association to particular chromatin contexts (for example, distance from a gene's transcriptional start site (TSS), histone modifications, and ChromHMM annotations). CpGs (filtered for ≥5x depth) captured in our RRBS study for each sample were merged according to the three categories previously described (all samples, Family A samples, Family C samples), thus creating three categories of background CpGs. The resulting background CpG files were then intersected with one of the six DMR files previously generated (Hyper and hypomethylated DMRs for shared, Family A, and Family C). Subsequently, the number of DMR-filtered CpGs and background CpGs that intersected with a particular context of interest, were compared. For distance from a gene's TSS, CpGs were intersected with bins of distance (from 0-1Kb up to 10-50Kb) in both up and downstream directions relative to each gene's genomic orientation. For histone modifications and ChromHMM annotations, CpGs were simply intersected with the Chip-seq peak tiles or annotated tiles. OR was then calculated as follows: $\frac{a/c}{b/d}$, where *a* = the number of CpGs that fall within a DMR and within the context of interest, *b* = the number of CpGs that fall within DMRs and outside of the context of interest, *c* = the number of CpGs that fall outside of DMRs and within the context of interest, *d* = the number of CpGs that fall outside of DMRs and outside of the context of interest. The logarithmic OR value (logOR) was then reported for each context of interest. Fisher's exact test was used to determine significance of odds ratios.

To determine the significance of proximity between DMRs in different contexts of interest (across families or cell types), we randomly sampled our set of captured CpGs to match the number of differentially methylated CpGs found within each DMR category. We then calculated the distance between CpGs from one category to the nearest sampled CpG from the category of comparison (e.g. Family A CpGs vs. Family C CpGs, or iPSC CpGs vs. fibroblast CpGs). This comparison served as our background distribution for CpG distance in the context of interest. The

same analysis was performed for differentially methylated CpGs. These distributions were plotted as a density distribution for interfamily CpG distance or using histogram bins for inter-cell type CpG distance. Significance was determine using Wilcoxon rank sum test and 1-tailed Fisher's exact test for interfamily and inter-cell type analyses, respectively.

# SECTION 3

# Conclusion

## 3.1 Summary and Conclusion:

***Genome replication programs both cell fate and aging***

This study demonstrated the temporal dynamics of post-replication DNA remethylation and nucleosomal occupancy using replication-associated sequencing techniques. We showed that these kinetics vary widely across the genome, leading to a prolonged window of time during which epigenetic entropy is present across the cell population. Moreover, the regions with the largest temporal delay, termed Repli-DMRs, were found to be at important regulatory features of the genome, associated with high gene expression variability and other elements highly linked to cell fate. Finally, our data suggest that these same Repli-DMRs are made up of CpGs with the most susceptibility to age-related epigenetic drift. More precisely, we confirmed previous observations that CpG context and CpG density are important factors that impact drift susceptibility, both of which were directly shown to be significantly associated with Repli-DMRs.

Ultimately, we provide the first evidence that the temporal dynamics of post-replication re-establishment of the epigenome may be the link between cell fate, aging, and disease. More precisely, we theorize that the same window of epigenetic heterogeneity that brings about multicellular life may also be its downfall, as a deterioration of the molecular epigenetic maintenance machinery, brought on by age and mutations, could lead to previously observed age- and disease-related epigenetic drift. This hypothesis adds sustenance to previous theories that multicellular life, age, and disease have all arisen in conjunction with evolutionary needs.

***DNA methylation analysis reveals epimutation hotspots in patients with dilated cardiomyopathy-associated laminopathies***

The laminopathy research presented in this dissertation[130] describes a framework for how DMR analysis of *in vitro* systems can be utilized to understand how regulatory elements become misregulated in laminopathy-associated diseases. Our results add to the previous studies substantiating that DNA methylation and chromatin remodeling of LADs/TADs have a combinatorial impact on the dysregulation of genes responsible for the development of DCM. Additionally, the family-specific DMR gene associations suggest the presence of both a laminopathy-shared and a mutation-unique set of epimutations. This type of analysis may prove to be highly beneficial for identifying networks of disease-relevant genes for rare diseases such as Family C's HHS IV, which have a limited disease-gene association database.

Still, certain limitations of this study must be considered. First, our study only had a limited number of patients and sibling controls per mutation and were not sex-diverse. This limits our ability to attain high statistical power and entirely rule out any sex bias, respectively. Additionally, our observations were made in patient skin fibroblasts and their iPSCs derivatives, neither of which are directly involved in the observed disease phenotypes. The study was performed, however, under the assumption that these more easily obtainable cell types could maintain a disease-specific epigenetic signature, and thus provide us with a powerful model to use as a foundation for future works.

Ultimately, our study highlights the potential for DNA methylation to provide new perspective on the etiology of mutation-specific laminopathies, as well as an alternative therapeutic substrate.

### 3.2 Future works:

The studies in this dissertation both deliver initial findings in their respective topics, without providing concrete mechanistic pathways. Future research on the first project will therefore focus on trying to validate the relationship between epigenetic heterogeneity and gene expression variability through single cell sequencing studies. The recent technological advances in replication-associated single cell sequencing[131] suggest that questions regarding temporal cell-to-cell heterogeneity will be able to be answered in the near future. Furthermore, studies related to epigenetic drift in long-term cell cultures[132] may provide more controllable ways to investigate how DNA methylation is lost over time. In addition, long-term cultures provide a way to understand if and how replication stress, previously suggested to be involved in age- and disease-related epigenetic alterations[42], is linked to the temporal component of the epigenome. Future studies in the second work will focus on validating the misregulation of identified genes and performing similar analyses on iPSCs-derived cardiomyocytes and osteoblasts from the two *LMNA*-mutated families to confirm our findings and to identify further gene networks associated to epimutations.

It is this author's hope that the research presented in this dissertation demonstrates two important, and sometimes unappreciated, notions with regards to the epigenome. The first is that the epigenome is a dynamic entity, shifting constantly over multiple timescales, from a cell's lifetime to hours, in response to its environment. Furthermore, it is also imperative to consider that the epigenome is regulated by numerous proteins working in conjunction and operating on different layers of the same DNA architecture. Moving forward, I hope to make use of these two principles to provide a further understanding of the epigenome's role and impact in disease, and in doing inform the development of new therapeutics better suited for pathologies with known epigenetic alterations, like laminopathies.

## REFERENCES


1. Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell [Internet]. 2013;153(5):1149–63. http://dx.doi.org/10.1016/j.cell.2013.04.037
2. Spivey EC, Jones SK, Rybarski JR, Saifuddin FA, Finkelstein IJ. An aging-independent replicative lifespan in a symmetrically dividing eukaryote. Elife. 2017;6:1–25.
3. Sen P, Shah PP, Nativio R, Berger SL. Epigenetic Mechanisms of Longevity and Aging. Cell. 2016;166(4):822–39.
4. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A. 2012;109(26):10522–7.
5. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet [Internet]. 2018;50(4):591–602. http://dx.doi.org/10.1038/s41588-018-0073-4
6. Cheedipudi SM, Matkovich SJ, Coarfa C, Hu X, Robertson MJ, Sweet M, et al. Genomic Reorganization of Lamin-Associated Domains in Cardiac Myocytes Is Associated With Differential Gene Expression and DNA Methylation in Human Dilated Cardiomyopathy. Circ Res. 2019;124(8):1198–213.
7. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–29.
8. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet. 2012;8(4):1–12.
9. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005;6(8):597–

610.

10. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet [Internet]. 2018;50(April). http://dx.doi.org/10.1038/s41588-018-0073-4
11. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. 2010;467(7312):167–73.
12. Mahdessian D, Cesnik AJ, Gnann C, Danielsson F, Stenström L, Arif M, et al. Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. Nature. 2021;590(7847):649–54.
13. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, et al. Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells. Mol Cell. 2014;55(2):319–31.
14. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015;523(7561):486–90.
15. Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: Implications, and a case of epigenetic thrift? Hum Mol Genet. 2013;22(R1):7–15.
16. Harris SE, Riggio V, Evenden L, Gilchrist T, McCafferty S, Murphy L, et al. Age-related gene expression changes and transcriptome wide association study of physical and cognitive aging traits in the Lothian Birth Cohort 1936. Aging (Albany NY). 2017;9(12):2489–503.
17. Pal S, Tyler JK. Epigenetics and aging. Sci Adv [Internet]. 2016 Jul 29;2(7):1–19. https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.1600584
18. Song Y, van den Berg PR, Markoulaki S, Soldner F, Dall'Agnese A, Henninger JE, et al. Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs. Mol Cell [Internet]. 2019;75(5):905–20.

https://doi.org/10.1016/j.molcel.2019.06.045

19. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, et al. Editing DNA Methylation in the Mammalian Genome. Cell [Internet]. 2016;167(1):233-247.e17. http://dx.doi.org/10.1016/j.cell.2016.08.056

20. Yoon BS, Yoo SJ, Lee JE, You S, Lee HT, Yoon HS. Enhanced differentiation of human embryonic stem cells into cardiomyocytes by combining hanging drop culture and 5-azacytidine treatment. Differentiation. 2006;(74):149–59.

21. Izzo F, Lee SC, Poran A, Chaligne R, Gaiti F, Gross B, et al. DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nat Genet [Internet]. 2020;52(4):378–87. http://dx.doi.org/10.1038/s41588-020-0595-4

22. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MPM, van Eijk K, et al. Aging effects on DNA methylation modules in human brain and blood tissue. Genome Biol [Internet]. 2012;13(10):1–18. http://genomebiology.com/2012/13/10/R97

23. Bork S, Pfister S, Witt H, Horn P, Korn B, Ho AD, et al. DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. Aging Cell. 2010;9(1):54–63.

24. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: Early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci. 2014;17(9):1156–63.

25. Aavik E, Babu M, Ylä-Herttuala S. DNA methylation processes in atheosclerotic plaque. Atherosclerosis [Internet]. 2019;281(December 2018):168–79. https://doi.org/10.1016/j.atherosclerosis.2018.12.006

26. Charlton J, Downing TL, Smith ZD, Gu H, Clement K, Pop R, et al. Global delay in nascent strand DNA methylation. Nat Struct Mol Biol. 2018;25(4):327–32.

27. Busto-Moner L, Morival J, Ren H, Fahim A, Reitz Z, Downing TL, et al. Stochastic modeling reveals kinetic heterogeneity in post-replication DNA methylation. PLoS

Comput Biol [Internet]. 2020;16(4):1–23. http://dx.doi.org/10.1371/journal.pcbi.1007195

28. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, et al. Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. Cancer Cell [Internet]. 2014 Dec;26(6):813–25. https://linkinghub.elsevier.com/retrieve/pii/S1535610814004164

29. Stewart-Morgan KR, Reverón-Gómez N, Groth A. Transcription Restart Establishes Chromatin Accessibility after DNA Replication. Mol Cell. 2019;75(2):284-297.e6.

30. Singh AM, Sun Y, Li L, Zhang W, Wu T, Zhao S, et al. Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. Stem Cell Reports. 2015;5(3):323–36.

31. Reverón-Gómez N, González-Aguilera C, Stewart-Morgan KR, Petryk N, Flury V, Graziano S, et al. Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication. Mol Cell. 2018;72(2):239–49.

32. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. 2012;22(12):2497–506.

33. Clark SJ, Argelaguet R, Kapourani CA, Stubbs TM, Lee HJ, Alda-Catalinas C, et al. ScNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun [Internet]. 2018;9(781):1–9. http://dx.doi.org/10.1038/s41467-018-03149-4

34. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213–8.

35. Zuo Z, Roy B, Chang YK, Granas D, Stormo GD. Measuring quantitative effects of methylation on transcription factor – DNA binding affinity. 2017;3(11):1–11.

36. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, et al. The Polycomb group

protein EZH2 directly controls DNA methylation. Nature [Internet]. 2006;439(7078):871–4. http://www.ncbi.nlm.nih.gov/pubmed/16357870

37. Blanco E, González-Ramírez M, Alcaine-Colet A, Aranda S, Di Croce L. The Bivalent Genome: Characterization, Structure, and Regulation. Trends Genet. 2020;36(2):118–31.

38. Bar-Ziv R, Voichek Y, Barkai N. Chromatin dynamics during DNA replication. Genome Res. 2016;26(9):1245–56.

39. Alabert C, Barth TK, Reverón-Gómez N, Sidoli S, Schmidt A, Jensen O, et al. Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. 2015;29(6):585–90.

40. Rulands S, Lee HJ, Clark SJ, Angermueller C, Smallwood SA, Krueger F, et al. Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. Cell Syst. 2018;7(1):63–76.

41. Kim M, Costello J. DNA methylation: An epigenetic mark of cellular memory. Exp Mol Med. 2017;49(4):1–8.

42. Alabert C, Groth A. Chromatin replication and epigenome maintenance. Nat Rev Mol Cell Biol. 2012;13(3):153–67.

43. Desai R V., Chen X, Martin B, Chaturvedi S, Hwang DW, Li W, et al. A DNA-repair pathway can affect transcriptional noise to promote cell fate transitions. Science (80- ) [Internet]. 2021 Jul 22;6506:1–17. https://www.sciencemag.org/lookup/doi/10.1126/science.abc6506

44. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol. 2013;20(9):1131–9.

45. Singh AM, Chappell J, Trost R, Lin L, Wang T, Tang J, et al. Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. Stem Cell Reports [Internet]. 2013;1(6):532–44.

http://dx.doi.org/10.1016/j.stemcr.2013.10.009

46. Simo-Riudalbas L, Diaz-Lagares A, Gatto S, Gagliardi M, Crujeiras AB, Matarazzo MR, et al. Genome-Wide DNA methylation analysis identifies novel hypomethylated non-Pericentromeric genes with potential clinical implications in ICF syndrome. PLoS One. 2015;10(7):1–20.

47. Krueger F. Trim Galore [Internet]. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

49. Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571–2.

50. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

51. Jenkinson G, Abante J, Feinberg AP, Goutsias J. An information-theoretic approach to the modeling and analysis of whole-genome bisulfite sequencing data. BMC Bioinformatics. 2018;19(87):1–23.

52. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72.

53. Karolchik D, Hinricks AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC table browser data retrieval tool. Nucleic Acids Res. 2004;32(Database Iss.):493–6.

54. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934–47.

55. Sharov AA, Ko MSH. Human ES Cell Profiling Broadens the Reach of Bivalent Domains. Cell Stem Cell. 2007;1(3):237–8.

56. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Ridwan Amode M, et al. Ensembl 2021. Nucleic Acids Res. 2021;49(Database Iss.):D884–91.

57. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44:W160–5.

58. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010;38(4):576–89.

59. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002;12(6):996–1006.

60. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019;47(D1):D419–26.

61. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015 Jan;109:21.29.1-21.29.9.

62. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol [Internet]. 2020 Mar 13;38(3):276–8. http://www.nature.com/articles/s41587-020-0435-1

63. R Core Team. R: A language and environment for statistical computing. [Internet]. R Foundation for Statistical Computing, Vienna, Austria. 2014. http://www.r-project.org/

64. Dechat T, Pfleghaar K, Sengupta K, Shimi T, Shumaker DK, Solimando L, et al. Nuclear lamins: Major factors in the structural organization and function of the nucleus and chromatin. Genes Dev. 2008;22(7):832–53.

65. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.

Nature. 2008;453:948–51.

66. Shah P, Wolf K, Lammerding J. Bursting the bubble – nuclear envelope rupture as a path to genomic instability? Trends Cell Biol. 2017;27(8):546–55.

67. Earle AJ, Kirby TJ, Fedorchak GR, Isermann P, Patel J, Iruvanti S, et al. Mutant lamins cause nuclear envelope rupture and DNA damage in skeletal muscle cells. Nat Mater. 2020;19(4):464–73.

68. Puckelwartz MJ, Depreux FFS, McNally EM. Gene expression, chromosome position and lamin A/C mutations. Nucleus. 2011;2(3):162–7.

69. Lu JT, Muchir A, Nagy PL, Worman HJ. LMNA cardiomyopathy: Cell biology and genetics meet clinical medicine. DMM Dis Model Mech. 2011;4(5):562–8.

70. Fatkin D, MacRae C, Sasaki T, Wolff MR, Porcu M, Frenneaux M, et al. Missense Mutations in the Rod Domain of the Lamin A/C Gene as Causes of Dilated Cardiomyopathy and Conduction-System Disease. N Engl J Med. 1999 Dec 2;341(23):1715–24.

71. Bonne G, Di Barletta MR, Varnous S, Bécane HM, Hammouda EH, Merlini L, et al. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery- Dreifuss muscular dystrophy. Nat Genet. 1999;21(3):285–8.

72. Perovanovic J, Dell'Orso S, Gnochi VF, Jaiswal JK, Sartorelli V, Vigouroux C, et al. Laminopathies disrupt epigenomic developmental programs and cell fate. Sci Transl Med. 2016;8(335):335ra58.

73. Bock C, Tomazou EM, Brinkman AB, Müller F, Simmer F, Gu H, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol. 2010;28(10):1106–14.

74. Gu H, Bock C, Mikkelsen TS, Jäger N, Smith ZD, Tomazou E, et al. Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat Methods. 2010;7(2):133–6.

75. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28(5):495–501.

76. Rada-Iglesias A. Is H3K4me1 at enhancers correlative or causative? Nat Genet. 2018;50(1):4–5.

77. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39(3):311–8.

78. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.

79. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc. 2017;12(12):2478–92.

80. Briand N, Collas P. Laminopathy-causing lamin A mutations reconfigure lamina-associated domains and local spatial chromatin conformation. Nucleus. 2018;9(1):216–26.

81. Köhler F, Bormann F, Raddatz G, Gutekunst J, Corless S, Musch T, et al. Epigenetic deregulation of lamina-associated domains in Hutchinson-Gilford progeria syndrome. Genome Med. 2020;12(1):1–16.

82. Kehat I, Molkentin JD. Molecular Pathways Underlying Cardiac Remodeling During Pathophysiological Stimulation. Circulation. 2010 Dec 21;122(25):2727–35.

83. Fountas A, Giotaki Z, Dounousi E, Liapis G, Bargiota A, Tsatsoulis A, et al. Familial partial lipodystrophy and proteinuric renal disease due to a missense c.1045C > T LMNA mutation. Endocrinol Diabetes Metab Case Reports. 2017;2017(June).

84. Fujita K, Hatta K. Membranous glomerulonephritis with an LMNA mutation. CEN Case Reports. 2018;7(1):98–100.

85. Irianto J, Pfeifer CR, Ivanovska IL, Swift J, Discher DE. Nuclear Lamins in Cancer. Cell Mol Bioeng. 2016;9(2):258–67.

86. Frost B. Alzheimer's disease: An acquired neurodegenerative laminopathy. Nucleus [Internet]. 2016;7(3):275–83. http://dx.doi.org/10.1080/19491034.2016.1183859

87. De Sandre-Giovannoli A, Chaouch M, Kozlov S, Vallat JM, Tazir M, Kassouri N, et al. Homozygous defects in LMNA, encoding lamin A/C nuclear-envelope proteins, cause autosomal recessive axonal neuropathy in human (Charcot-Marie-Tooth disorder type 2) and mouse. Am J Hum Genet. 2002;70(3):726–36.

88. Widyastuti HP, Norden-Krichmar TM, Grosberg A, Zaragoza M V. Gene expression profiling of fibroblasts in a family with LMNA-related cardiomyopathy reveals molecular pathways implicated in disease pathogenesis. BMC Med Genet. 2020;21(1):1–12.

89. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature. 2007;448(7151):318–24.

90. Alegría-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. Epigenomics. 2011;3(3):267–77.

91. Breton C V., Byun HM, Wenten M, Pan F, Yang A, Gilliland FD. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. Am J Respir Crit Care Med. 2009;180(5):462–7.

92. Zaragoza M V., Hakim SA, Hoang V, Elliott AM. Heart-hand syndrome IV: a second family with LMNA-related cardiomyopathy and brachydactyly. Clin Genet. 2017;91(3):499–500.

93. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48:D845–55.

94. Bhattacharjee P, Dasgupta D, Sengupta K. DCM associated LMNA mutations cause distortions in lamina structure and assembly. Biochim Biophys Acta - Gen Subj. 2017 Nov;1861(11):2598–608.

95. Haas J, Frese KS, Park YJ, Keller A, Vogel B, Lindroth AM, et al. Alterations in cardiac DNA methylation in human dilated cardiomyopathy. EMBO Mol Med. 2013;5(3):413–29.

96. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015;161(5):1012–25.

97. Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat Genet. 2016;48(5):488–96.

98. Gay A, Towler DA. Wnt Signaling in Cardiovascular Disease: Opportunities and Challenges. Curr Opin Lipidol. 2017;28(5):387–96.

99. Foulquier S, Daskalopoulos EP, Lluri G, Hermans KCM, Deb A, Blankesteijn WM. WNT signaling in cardiac and vascular disease. Pharmacol Rev. 2018;70(1):68–141.

100. Le Dour C, Macquart C, Sera F, Homma S, Bonne G, Morrow JP, et al. Decreased WNT/β-catenin signalling contributes to the pathogenesis of dilated cardiomyopathy caused by mutations in the lamin a/C gene. Hum Mol Genet. 2017;26(2):333–43.

101. Ahuja P, Sdek P, Maclellan RW. Cardiac Myocyte Cell Cycle Control in Development, Disease and Regeneration. Physiol Rev. 2007;87(2):521–44.

102. Chen SN, Lombardi R, Karmouch J, Tsai JY, Czernuszewicz G, Taylor MRG, et al. DNA Damage Response/TP53 Pathway Is Activated and Contributes to the Pathogenesis of Dilated Cardiomyopathy Associated with LMNA (Lamin A/C) Mutations. Circ Res. 2019;124(6):856–73.

103. Shimi T, Butin-Israeli V, Adam SA, Hamanaka RB, Goldman AE, Lucas CA, et al. The role of nuclear lamin B1 in cell proliferation and senescence. Genes Dev. 2011;25(24):2579–93.

104. Röhrs S, Kutzner N, Vlad A, Grunwald T, Ziegler S, Müller O. Chronological expression of Wnt target genes Ccnd1, Myc, Cdkn1a, Tfrc, Plf1 and Ramp3. Cell Biol Int. 2009;33(4):501–8.

105. Tatman PD, Woulfe KC, Karimpour-Fard A, Jeffrey DA, Jaggers J, Cleveland JC, et al. Pediatric dilated cardiomyopathy hearts display a unique gene expression profile. JCI insight. 2017;2(14).

106. Briscoe J, Thérond P. Hedgehog signaling: From the drosophila cuticle to anti-cancer drugs. Dev Cell. 2005;8(2):143–51.

107. Tickle C, Towers M. Sonic hedgehog signaling in limb development. Front Cell Dev Biol. 2017;5(14):1–19.

108. Villavicencio-Lorini P, Klopocki E, Trimborn M, Koll R, Mundlos S, Horn D. Phenotypic variant of Brachydactyly-mental retardation syndrome in a family with an inherited interstitial 2q37.3 microdeletion including HDAC4. Eur J Hum Genet. 2013;21(7):743–8.

109. Williams SR, Aldred MA, Der Kaloustian VM, Halal F, Gowans G, McLeod DR, et al. Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. Am J Hum Genet. 2010;87(2):219–28.

110. Gao B, Hu J, Stricker S, Cheung M, Ma G, Law KF, et al. A mutation in Ihh that causes digit abnormalities alters its signalling capacity and range. Nature. 2009;458(7242):1196–200.

111. Zaragoza M V., Fung L, Jensen E, Oh F, Cung K, McCarthy LA, et al. Exome sequencing identifies a novel LMNA splice-site mutation and multigenic heterozygosity of potential modifiers in a family with sick sinus syndrome, dilated cardiomyopathy, and sudden cardiac death. PLoS One. 2016;11(5):1–19.

112. Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Japan Acad Ser B Phys Biol Sci. 2009;85(8):348–62.

113. Lieu PT, Fontes A, Vemuri MC, MacArthur CC. Generation of Induced Pluripotent Stem

Cells with CytoTune, a Non-Integrating Sendai Virus. In: Methods Mol Biol. 2013. p. 45–56.

114. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.

115. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811–2.

116. Gu Z, Eils R, Schlesner M. Gtrellis: An R/Bioconductor package for making genome-level Trellis graphics. BMC Bioinformatics [Internet]. 2016;17(169):1–7. http://dx.doi.org/10.1186/s12859-016-1051-4

117. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Wolfgang H, Liaw A, et al. gplots: Various R programming tools for plotting data. R package version 2.17.0. 2015.

118. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012;13(10):R87.

119. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LTY, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500(7463):477–81.

120. Earnst J, Kellis M. ChromHMM: automating chromatin state discovery and characterization. Nat Methods. 2012;9(3):215–6.

121. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford). 2017;2017:1–17.

122. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009;37:305–11.

123. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000 May;25(1):25–9.

124. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: Online access to ontology and annotation data. Bioinformatics. 2009;25(2):288–9.

125. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–13.

126. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27–30.

127. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995 Jan;57(1):289–300.

128. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: From gene data mining to disease genome sequence analyses. Curr Protoc Bioinforma. 2016;54:1.30.1-1.30.33.

129. Lund E, Oldenburg AR, Collas P. Enriched domain detector: A program for detection of wide genomic enrichment domains robust against local variations. Nucleic Acids Res. 2014;42(11):e92.

130. Morival JLP, Widyastuti HP, Nguyen CHH, Zaragoza M V., Downing TL. DNA methylation analysis reveals epimutation hotspots in patients with dilated cardiomyopathy-associated laminopathies. Clin Epigenetics [Internet]. 2021 Dec 10;13(139):1–20. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164365

131. Miura H, Takahashi S, Shibata T, Nagao K, Obuse C, Okumura K, et al. Mapping replication timing domains genome wide in single mammalian cells with single-cell DNA replication sequencing. Nat Protoc [Internet]. 2020;15(12):4058–100. http://dx.doi.org/10.1038/s41596-020-0378-5

132. Franzen J, Georgomanolis T, Selich A, Kuo CC, Stöger R, Brant L, et al. DNA methylation changes during long-term in vitro cell culture are caused by epigenetic drift.

Commun Biol. 2021;4(598):1–12.

**Appendix Section 1**

## Appendix 1.1

### List of transcription factor binding site motifs enriched in Repli-DMRs from HOMER

| Rank | Motif Name | Log P-value | q-value (Benjamini) |
|---|---|---|---|
| 1 | Oct6 | -3.73E+01 | 0 |
| 2 | Oct4 | -3.51E+01 | 0 |
| 3 | Brn1 | -2.79E+01 | 0 |
| 4 | CDX4 | -2.75E+01 | 0 |
| 5 | Cux2 | -2.58E+01 | 0 |
| 6 | HNF6 | -2.55E+01 | 0 |
| 7 | Oct11 | -2.40E+01 | 0 |
| 8 | HOXB13 | -2.35E+01 | 0 |
| 9 | Hoxd10 | -1.79E+01 | 0 |
| 10 | Gata6 | -1.60E+01 | 0 |
| 11 | Hnf6b | -1.50E+01 | 0 |
| 12 | LEF1 | -1.47E+01 | 0 |
| 13 | Gata4 | -1.46E+01 | 0 |
| 14 | Hoxc9 | -1.44E+01 | 0 |
| 15 | Hoxa10 | -1.39E+01 | 0 |
| 16 | FoxD3 | -1.36E+01 | 0 |
| 17 | Tbr1 | -1.36E+01 | 0 |
| 18 | Foxa3 | -1.33E+01 | 0 |
| 19 | NFATC2 | -1.32E+01 | 0 |
| 20 | Foxa2 | -1.32E+01 | 0 |
| 21 | Prop1 | -1.29E+01 | 0 |
| 22 | BMYB | -1.28E+01 | 0 |
| 23 | Oct2 | -1.27E+01 | 0 |
| 24 | DLX5 | -1.24E+01 | 0.0001 |
| 25 | Zic1/2 | -1.23E+01 | 0.0001 |
| 26 | Bcl6 | -1.22E+01 | 0.0001 |
| 27 | Foxh1 | -1.21E+01 | 0.0001 |
| 28 | STAT6 | -1.14E+01 | 0.0001 |
| 29 | FOXP1 | -1.14E+01 | 0.0001 |
| 30 | Foxa2 | -1.13E+01 | 0.0002 |
| 31 | Atoh1 | -1.07E+01 | 0.0003 |
| 32 | Phox2a | -1.06E+01 | 0.0003 |
| 33 | Pdx1 | -1.05E+01 | 0.0003 |
| 34 | Rbpjl | -1.04E+01 | 0.0003 |

| | | | |
|---|---|---|---|
| 35 | CUX1 | -1.03E+01 | 0.0004 |
| 36 | Dlx3 | -1.03E+01 | 0.0004 |
| 37 | Cdx2 | -9.92E+00 | 0.0005 |
| 38 | Lhx1 | -9.63E+00 | 0.0006 |
| 39 | Otx2 | -9.62E+00 | 0.0006 |
| 40 | Zic3 | -9.54E+00 | 0.0007 |
| 41 | GATA3 | -9.53E+00 | 0.0007 |
| 42 | Six2 | -9.42E+00 | 0.0007 |
| 43 | Eomes | -9.14E+00 | 0.0009 |
| 44 | STAT6 | -9.13E+00 | 0.0009 |
| 45 | PBX2 | -9.11E+00 | 0.0009 |
| 46 | Six4 | -9.08E+00 | 0.0009 |
| 47 | Pax7 | -8.99E+00 | 0.001 |
| 48 | Brn2 | -8.76E+00 | 0.0012 |
| 49 | Mef2d | -8.72E+00 | 0.0013 |
| 50 | MafA | -8.70E+00 | 0.0013 |
| 51 | HOXA2 | -8.46E+00 | 0.0016 |
| 52 | Atf3 | -8.39E+00 | 0.0017 |
| 53 | NFY | -8.33E+00 | 0.0018 |
| 54 | CEBP | -8.14E+00 | 0.0021 |
| 55 | STAT1 | -8.14E+00 | 0.0021 |
| 56 | STAT4 | -8.07E+00 | 0.0022 |
| 57 | Sox21 | -7.89E+00 | 0.0026 |
| 58 | Sox9 | -7.75E+00 | 0.0029 |
| 59 | Sox10 | -7.72E+00 | 0.003 |
| 60 | Olig2 | -7.61E+00 | 0.0033 |
| 61 | HOXA1 | -7.56E+00 | 0.0034 |
| 62 | HRE | -7.49E+00 | 0.0036 |
| 63 | Fra1 | -7.43E+00 | 0.0037 |
| 64 | Bach1 | -7.40E+00 | 0.0038 |
| 65 | Fos | -7.32E+00 | 0.0041 |
| 66 | EWS | -7.24E+00 | 0.0043 |
| 67 | GATA3 | -7.23E+00 | 0.0043 |
| 68 | NFE2L2 | -7.23E+00 | 0.0043 |
| 69 | PAX3 | -7.14E+00 | 0.0046 |
| 70 | HNF1b | -6.97E+00 | 0.0054 |
| 71 | Oct7 | -6.95E+00 | 0.0054 |
| 72 | Gata2 | -6.90E+00 | 0.0056 |
| 73 | Sox7 | -6.89E+00 | 0.0056 |
| 74 | MYNN | -6.84E+00 | 0.0058 |
| 75 | NR1H2 | -6.71E+00 | 0.0065 |

| | | | |
|---|---|---|---|
| 76 | Pit1 | -6.68E+00 | 0.0067 |
| 77 | STAT5 | -6.66E+00 | 0.0067 |
| 78 | TRPS1 | -6.60E+00 | 0.007 |
| 79 | IRF4 | -6.58E+00 | 0.0071 |
| 80 | Pit1 | -6.40E+00 | 0.0084 |
| 81 | NeuroD1 | -6.40E+00 | 0.0084 |
| 82 | MafB | -6.39E+00 | 0.0084 |
| 83 | En1 | -6.38E+00 | 0.0084 |
| 84 | Zic2 | -6.36E+00 | 0.0084 |
| 85 | Hoxb4 | -6.20E+00 | 0.0097 |
| 86 | Duxbl | -6.19E+00 | 0.0097 |
| 87 | FOXK1 | -6.08E+00 | 0.0107 |
| 88 | ZNF7 | -6.05E+00 | 0.0109 |
| 89 | LHX9 | -6.01E+00 | 0.0113 |
| 90 | RFX | -5.95E+00 | 0.0119 |
| 91 | DLX2 | -5.92E+00 | 0.0121 |
| 92 | Lhx2 | -5.89E+00 | 0.0123 |
| 93 | Hoxd13 | -5.88E+00 | 0.0123 |
| 94 | PRDM15 | -5.78E+00 | 0.0134 |
| 95 | CEBP | -5.75E+00 | 0.0137 |
| 96 | Gata1 | -5.54E+00 | 0.0168 |
| 97 | Stat3 | -5.49E+00 | 0.0175 |
| 98 | THRa | -5.38E+00 | 0.0193 |
| 99 | Lhx3 | -5.32E+00 | 0.0202 |
| 100 | NeuroG2 | -5.23E+00 | 0.022 |
| 101 | SPI1 | -5.13E+00 | 0.0241 |
| 102 | GRHL2 | -5.12E+00 | 0.0241 |
| 103 | Fosl2 | -5.07E+00 | 0.0251 |
| 104 | JunB | -5.02E+00 | 0.0261 |
| 105 | IRF3 | -4.98E+00 | 0.027 |
| 106 | DUX4 | -4.94E+00 | 0.0279 |
| 107 | Sox17 | -4.88E+00 | 0.0293 |
| 108 | Pitx1 | -4.88E+00 | 0.0293 |
| 109 | Six1 | -4.87E+00 | 0.0293 |
| 110 | Phox2b | -4.87E+00 | 0.0293 |
| 111 | BATF | -4.87E+00 | 0.0293 |
| 112 | Fra2 | -4.87E+00 | 0.0293 |
| 113 | FOXM1 | -4.83E+00 | 0.0294 |
| 114 | Foxo3 | -4.78E+00 | 0.0305 |
| 115 | NRSF | -4.75E+00 | 0.0313 |
| 116 | Bcl11a | -4.72E+00 | 0.032 |

| | | | |
|---|---|---|---|
| 117 | RBPJ | -4.71E+00 | 0.032 |
| 118 | Rfx1 | -4.68E+00 | 0.0326 |

## Appendix 1.2

### List of gene ontology terms enriched from top 20 transcription factor binding sites in Repli-DMRs

| GO biological process complete | fold Enrichment | p-value | FDR |
|---|---|---|---|
| endocrine pancreas development (GO:0031018) | > 100 | 8.53E-08 | 2.81E-05 |
| positive regulation of cardioblast differentiation (GO:0051891) | > 100 | 2.50E-05 | 4.87E-03 |
| atrioventricular node development (GO:0003162) | > 100 | 3.21E-05 | 6.11E-03 |
| regulation of cardioblast differentiation (GO:0051890) | > 100 | 4.01E-05 | 7.19E-03 |
| peripheral nervous system neuron differentiation (GO:0048934) | > 100 | 8.08E-05 | 1.36E-02 |
| peripheral nervous system neuron development (GO:0048935) | > 100 | 8.08E-05 | 1.37E-02 |
| endodermal cell fate commitment (GO:0001711) | > 100 | 8.08E-05 | 1.39E-02 |
| atrioventricular canal development (GO:0036302) | > 100 | 9.32E-05 | 1.52E-02 |
| positive regulation of cardiocyte differentiation (GO:1905209) | > 100 | 1.51E-04 | 2.30E-02 |
| positive regulation of stem cell differentiation (GO:2000738) | > 100 | 1.86E-04 | 2.72E-02 |
| proximal/distal pattern formation (GO:0009954) | 93.61 | 5.46E-06 | 1.22E-03 |
| intestinal epithelial cell differentiation (GO:0060575) | 93.61 | 2.44E-04 | 3.44E-02 |
| cell fate commitment involved in formation of primary germ layer (GO:0060795) | 79.21 | 3.33E-04 | 4.58E-02 |
| regulation of cardiocyte differentiation (GO:1905207) | 73.55 | 3.83E-04 | 5.00E-02 |
| pancreas development (GO:0031016) | 59.7 | 6.72E-07 | 1.83E-04 |
| endoderm development (GO:0007492) | 54.92 | 9.24E-07 | 2.39E-04 |
| cell fate specification (GO:0001708) | 50.85 | 1.24E-06 | 3.02E-04 |
| peripheral nervous system development (GO:0007422) | 39.1 | 6.59E-05 | 1.16E-02 |
| anterior/posterior pattern specification (GO:0009952) | 35.51 | 7.13E-10 | 2.97E-07 |
| cell fate commitment (GO:0045165) | 34.61 | 4.08E-11 | 2.08E-08 |
| endocrine system development (GO:0035270) | 33.22 | 6.36E-06 | 1.40E-03 |
| response to BMP (GO:0071772) | 32.86 | 1.09E-04 | 1.72E-02 |
| cellular response to BMP stimulus (GO:0071773) | 32.86 | 1.09E-04 | 1.73E-02 |

| | | | |
|---|---|---|---|
| cardiocyte differentiation (GO:0035051) | 28.87 | 1.58E-04 | 2.37E-02 |
| formation of primary germ layer (GO:0001704) | 27.58 | 1.80E-04 | 2.66E-02 |
| gastrulation (GO:0007369) | 26.24 | 1.58E-05 | 3.23E-03 |
| regionalization (GO:0003002) | 25.99 | 3.74E-10 | 1.64E-07 |
| embryonic appendage morphogenesis (GO:0035113) | 25.96 | 2.14E-04 | 3.08E-02 |
| embryonic limb morphogenesis (GO:0030326) | 25.96 | 2.14E-04 | 3.11E-02 |
| male gonad development (GO:0008584) | 22.39 | 3.28E-04 | 4.55E-02 |
| liver development (GO:0001889) | 22.39 | 3.28E-04 | 4.59E-02 |
| development of primary male sexual characteristics (GO:0046546) | 22.22 | 3.35E-04 | 4.56E-02 |
| hepaticobiliary system development (GO:0061008) | 21.91 | 3.49E-04 | 4.59E-02 |
| appendage morphogenesis (GO:0035107) | 21.91 | 3.49E-04 | 4.63E-02 |
| limb morphogenesis (GO:0035108) | 21.91 | 3.49E-04 | 4.67E-02 |
| pattern specification process (GO:0007389) | 19.34 | 3.65E-09 | 1.41E-06 |
| regulation of transmembrane receptor protein serine/threonine kinase signaling pathway (GO:0090092) | 16.28 | 9.77E-05 | 1.58E-02 |
| embryonic morphogenesis (GO:0048598) | 16.26 | 1.30E-09 | 5.15E-07 |
| reproductive structure development (GO:0048608) | 14.85 | 2.16E-06 | 5.09E-04 |
| reproductive system development (GO:0061458) | 14.75 | 2.25E-06 | 5.23E-04 |
| embryonic organ development (GO:0048568) | 14.2 | 2.78E-06 | 6.38E-04 |
| in utero embryonic development (GO:0001701) | 14.07 | 2.28E-05 | 4.56E-03 |
| embryonic organ morphogenesis (GO:0048562) | 13.96 | 1.75E-04 | 2.61E-02 |
| positive regulation of transcription by RNA polymerase II (GO:0045944) | 13.59 | 8.84E-17 | 6.99E-13 |
| gland development (GO:0048732) | 12.44 | 4.08E-05 | 7.25E-03 |
| chordate embryonic development (GO:0043009) | 11.76 | 1.19E-06 | 2.93E-04 |
| embryo development (GO:0009790) | 11.43 | 3.80E-10 | 1.62E-07 |
| embryo development ending in birth or egg hatching (GO:0009792) | 11.39 | 1.47E-06 | 3.51E-04 |
| positive regulation of RNA biosynthetic process (GO:1902680) | 10.75 | 1.82E-16 | 5.76E-13 |
| positive regulation of transcription, DNA-templated (GO:0045893) | 10.75 | 1.80E-16 | 7.13E-13 |
| positive regulation of nucleic acid-templated transcription (GO:1903508) | 10.75 | 1.80E-16 | 9.51E-13 |
| positive regulation of RNA metabolic process (GO:0051254) | 9.87 | 7.48E-16 | 1.31E-12 |

| | | | |
|---|---|---|---|
| animal organ morphogenesis (GO:0009887) | 9.69 | 1.10E-07 | 3.33E-05 |
| positive regulation of macromolecule biosynthetic process (GO:0010557) | 9.4 | 1.70E-15 | 2.69E-12 |
| negative regulation of transcription by RNA polymerase II (GO:0000122) | 9.09 | 1.13E-06 | 2.84E-04 |
| positive regulation of nucleobase-containing compound metabolic process (GO:0045935) | 9.02 | 3.37E-15 | 4.44E-12 |
| epithelial cell differentiation (GO:0030855) | 9.01 | 3.63E-05 | 6.74E-03 |
| positive regulation of cellular biosynthetic process (GO:0031328) | 8.89 | 4.28E-15 | 5.20E-12 |
| positive regulation of biosynthetic process (GO:0009891) | 8.74 | 5.73E-15 | 6.04E-12 |
| epithelium development (GO:0060429) | 8.33 | 3.95E-07 | 1.16E-04 |
| circulatory system development (GO:0072359) | 8.22 | 1.24E-05 | 2.60E-03 |
| anatomical structure formation involved in morphogenesis (GO:0048646) | 8.16 | 1.29E-05 | 2.69E-03 |
| regulation of transcription by RNA polymerase II (GO:0006357) | 7.94 | 1.09E-18 | 1.73E-14 |
| negative regulation of transcription, DNA-templated (GO:0045892) | 7.94 | 1.04E-07 | 3.36E-05 |
| negative regulation of nucleic acid-templated transcription (GO:1903507) | 7.93 | 1.06E-07 | 3.34E-05 |
| negative regulation of RNA biosynthetic process (GO:1902679) | 7.92 | 1.07E-07 | 3.33E-05 |
| tube morphogenesis (GO:0035239) | 7.87 | 3.42E-04 | 4.62E-02 |
| developmental process involved in reproduction (GO:0003006) | 7.4 | 2.43E-05 | 4.80E-03 |
| negative regulation of RNA metabolic process (GO:0051253) | 7.31 | 2.24E-07 | 6.69E-05 |
| tube development (GO:0035295) | 7.22 | 1.23E-04 | 1.92E-02 |
| anatomical structure morphogenesis (GO:0009653) | 7.16 | 1.95E-11 | 1.10E-08 |
| tissue development (GO:0009888) | 7.15 | 8.41E-09 | 3.09E-06 |
| positive regulation of gene expression (GO:0010628) | 7.05 | 7.49E-06 | 1.62E-03 |
| negative regulation of nucleobase-containing compound metabolic process (GO:0045934) | 6.74 | 4.79E-07 | 1.38E-04 |
| negative regulation of cellular macromolecule biosynthetic process (GO:2000113) | 6.69 | 5.15E-07 | 1.45E-04 |
| negative regulation of macromolecule biosynthetic process (GO:0010558) | 6.64 | 5.50E-07 | 1.52E-04 |
| negative regulation of cellular biosynthetic process (GO:0031327) | 6.4 | 7.76E-07 | 2.08E-04 |
| negative regulation of biosynthetic process (GO:0009890) | 6.27 | 9.35E-07 | 2.38E-04 |
| regulation of nucleic acid-templated transcription (GO:1903506) | 5.93 | 3.68E-16 | 8.31E-13 |
| regulation of transcription, DNA-templated (GO:0006355) | 5.93 | 3.66E-16 | 9.63E-13 |
| regulation of RNA biosynthetic process (GO:2001141) | 5.92 | 3.79E-16 | 7.48E-13 |
| positive regulation of developmental process (GO:0051094) | 5.67 | 1.31E-04 | 2.03E-02 |

| | | | |
|---|---|---|---|
| reproductive process (GO:0022414) | 5.65 | 3.74E-05 | 6.88E-03 |
| reproduction (GO:0000003) | 5.63 | 3.80E-05 | 6.90E-03 |
| positive regulation of nitrogen compound metabolic process (GO:0051173) | 5.61 | 8.84E-12 | 5.82E-09 |
| regulation of RNA metabolic process (GO:0051252) | 5.45 | 1.99E-15 | 2.86E-12 |
| positive regulation of cellular metabolic process (GO:0031325) | 5.28 | 2.34E-11 | 1.28E-08 |
| positive regulation of macromolecule metabolic process (GO:0010604) | 5.26 | 2.17E-12 | 1.56E-09 |
| regulation of cellular macromolecule biosynthetic process (GO:2000112) | 5.19 | 5.18E-15 | 5.85E-12 |
| neurogenesis (GO:0022008) | 5.18 | 2.30E-04 | 3.27E-02 |
| regulation of macromolecule biosynthetic process (GO:0010556) | 5.15 | 6.06E-15 | 5.98E-12 |
| regulation of nucleobase-containing compound metabolic process (GO:0019219) | 5.09 | 7.76E-15 | 7.21E-12 |
| cell development (GO:0048468) | 5 | 8.84E-05 | 1.46E-02 |
| regulation of cellular biosynthetic process (GO:0031326) | 4.95 | 1.35E-14 | 1.19E-11 |
| cell differentiation (GO:0030154) | 4.9 | 8.12E-11 | 3.89E-08 |
| regulation of biosynthetic process (GO:0009889) | 4.87 | 1.82E-14 | 1.52E-11 |
| positive regulation of metabolic process (GO:0009893) | 4.85 | 9.02E-12 | 5.71E-09 |
| cellular developmental process (GO:0048869) | 4.82 | 1.06E-10 | 4.94E-08 |
| animal organ development (GO:0048513) | 4.8 | 5.89E-09 | 2.22E-06 |
| negative regulation of cellular metabolic process (GO:0031324) | 4.73 | 8.29E-07 | 2.18E-04 |
| negative regulation of nitrogen compound metabolic process (GO:0051172) | 4.2 | 3.45E-05 | 6.49E-03 |
| regulation of gene expression (GO:0010468) | 4.18 | 3.82E-13 | 3.02E-10 |
| negative regulation of metabolic process (GO:0009892) | 3.99 | 5.17E-06 | 1.17E-03 |
| anatomical structure development (GO:0048856) | 3.86 | 1.87E-12 | 1.41E-09 |
| system development (GO:0048731) | 3.81 | 2.99E-08 | 1.03E-05 |
| multicellular organism development (GO:0007275) | 3.75 | 8.06E-10 | 3.27E-07 |
| regulation of nitrogen compound metabolic process (GO:0051171) | 3.59 | 8.21E-12 | 5.64E-09 |
| negative regulation of macromolecule metabolic process (GO:0010605) | 3.58 | 1.38E-04 | 2.12E-02 |
| developmental process (GO:0032502) | 3.55 | 1.00E-11 | 6.08E-09 |
| regulation of primary metabolic process (GO:0080090) | 3.48 | 1.55E-11 | 9.06E-09 |
| regulation of cellular metabolic process (GO:0031323) | 3.36 | 3.13E-11 | 1.65E-08 |
| regulation of macromolecule metabolic process (GO:0060255) | 3.25 | 5.98E-11 | 2.95E-08 |

| | | | |
|---|---|---|---|
| positive regulation of cellular process (GO:0048522) | 3.23 | 1.08E-08 | 3.90E-06 |
| regulation of metabolic process (GO:0019222) | 3 | 3.00E-10 | 1.36E-07 |
| negative regulation of cellular process (GO:0048523) | 2.94 | 1.64E-05 | 3.32E-03 |
| positive regulation of biological process (GO:0048518) | 2.93 | 5.84E-08 | 1.96E-05 |
| multicellular organismal process (GO:0032501) | 2.82 | 1.41E-08 | 4.95E-06 |
| negative regulation of biological process (GO:0048519) | 2.62 | 6.65E-05 | 1.15E-02 |
| regulation of cellular process (GO:0050794) | 1.81 | 9.49E-06 | 2.03E-03 |
| regulation of biological process (GO:0050789) | 1.73 | 2.95E-05 | 5.68E-03 |
| biological regulation (GO:0065007) | 1.64 | 8.28E-05 | 1.38E-02 |

## Appendix 1.3

## SRA codes and Bismark mapping efficiencies for downloaded Repli-BS samples

| Sample | SRA | Efficiency |
|---|---|---|
| S1 - 0hr | SRR3609267 | 80.80% |
| S1 - 0hr | SRR3609268 | 80.70% |
| S1 - 0hr | SRR3609269 | 81.30% |
| S1 - 0hr | SRR3609270 | 81.30% |
| S2 - 0hr | SRR3609271 | 85.30% |
| S2 - 0hr | SRR3609272 | 85.10% |
| S2 - 0hr | SRR3609273 | 83.70% |
| S2 - 0hr | SRR3609274 | 83.50% |
| S2 - 0hr | SRR3609275 | 86.00% |
| S2 - 0hr | SRR3609276 | 86.00% |
| S3 - 0hr | SRR3609277 | 86.70% |
| S3 - 0hr | SRR3609278 | 86.60% |
| S3 - 0hr | SRR3609279 | 77.10% |
| S3 - 0hr | SRR3609280 | 77.00% |
| S3 - 0hr | SRR3609281 | 87.30% |
| S3 - 0hr | SRR3609282 | 87.30% |
| S4 - 0hr | SRR3609283 | 86.00% |
| S4 - 0hr | SRR3609285 | 85.90% |
| S4 - 0hr | SRR3609286 | 84.00% |
| S4 - 0hr | SRR3609287 | 83.80% |
| S4 - 0hr | SRR3609288 | 86.50% |
| S4 - 0hr | SRR3609289 | 86.50% |
| S5 - 0hr | SRR3609290 | 76.50% |
| S5 - 0hr | SRR3609291 | 76.40% |
| S5 - 0hr | SRR3609292 | 83.70% |
| S5 - 0hr | SRR3609293 | 83.50% |
| S5 - 0hr | SRR3609294 | 76.80% |
| S5 - 0hr | SRR3609295 | 76.80% |
| S6 - 0hr | SRR3609296 | 70.80% |
| S6 - 0hr | SRR3609297 | 70.70% |
| S6 - 0hr | SRR3609298 | 79.30% |
| S6 - 0hr | SRR3609299 | 79.10% |
| S6 - 0hr | SRR3609300 | 71.10% |
| S6 - 0hr | SRR3609301 | 71.00% |

| | | |
|---|---|---|
| 16hr Nasc | SRR3609323 | 73.70% |
| 16hr Nasc | SRR3609324 | 73.60% |
| 16hr Nasc | SRR3609325 | 70.60% |
| 16hr Nasc | SRR3609326 | 70.40% |
| 16hr Nasc | SRR5621968 | 86.40% |
| Arrested | SRR3609311 | 87.10% |
| Arrested | SRR3609312 | 87.20% |
| Arrested | SRR3609313 | 87.80% |
| Arrested | SRR3609314 | 87.90% |

## Appendix Section 2

### Appendix 2.1

**A.** RRBS read and methylation call data before and after depth filtering for fibroblast lines

| Cell ID | Family | Mapping efficiency (%) | # of mapped reads | Mean read depth | # of CpGs | # of CpGs with ≥5x depth | % of total CpGs with ≥5x depth | Mean methylation level (%) ≥5x |
|---|---|---|---|---|---|---|---|---|
| C2 | Donor | 80.10 | 11167514 | 5.93 | 5765443 | 2453192 | 42.55 | 62.27 |
| C1 | A | 78.90 | 12955944 | 7.62 | 5732669 | 2478080 | 43.23 | 59.15 |
| C3 | A | 76.40 | 9405102 | 5.07 | 5566829 | 2082358 | 37.41 | 59.21 |
| P1 | A | 79.70 | 11919395 | 4.96 | 6687065 | 2418012 | 36.16 | 64.54 |
| P2 | A | 74.70 | 11726435 | 5.55 | 6434663 | 2574977 | 40.02 | 60.37 |
| P3 | A | 68.00 | 8205263 | 4.06 | 5658271 | 1617429 | 28.59 | 61.75 |
| C4 | C | 71.60 | 9253758 | 4.79 | 5722213 | 1988035 | 34.74 | 61.66 |
| C5 | C | 74.70 | 8597498 | 4.77 | 5368763 | 1897892 | 35.35 | 60.61 |
| P4 | C | 71.30 | 9723382 | 5.27 | 5842361 | 2240017 | 38.34 | 58.93 |
| P5 | C | 79.20 | 9976610 | 4.96 | 6087224 | 2273148 | 37.34 | 61.52 |
| Control | Avg | 76.34 | 10275963.2 | 5.63 | 5631183.4 | 2179911.4 | 38.66 | 60.58 |
| Patient | Avg | 74.58 | 10310217 | 4.96 | 6141916.8 | 2224716.6 | 36.09 | 61.42 |
| All | Avg | 75.46 | 10293090.1 | 5.30 | 5886550.1 | 2202314 | 37.37 | 61.00 |

**B.** RRBS read and methylation call data before and after depth filtering for iPSC lines

| Cell ID | Family | Mapping efficency (%) | # of mapped reads | Mean read depth | # of CpGs | # of CpGs with ≥5x depth | % of total CpGs with ≥5x depth | Mean methylation level (%) ≥5x |
|---|---|---|---|---|---|---|---|---|
| C2 | Donor | 81.8 | 15412513 | 3.76 | 9679992 | 2253977 | 23.28 | 68.94 |
| C1 | A | 82.3 | 14005546 | 3.37 | 9272820 | 1863154 | 20.09 | 69.89 |
| C3 | A | 79.9 | 12818488 | 5.22 | 7269552 | 2524314 | 34.72 | 68.85 |
| P1 | A | 80 | 13239664 | 5.01 | 7303946 | 2456742 | 33.64 | 70.97 |
| P2 | A | 74.4 | 13889429 | 4.16 | 8538282 | 2369711 | 27.75 | 71.06 |
| P3 | A | 77.3 | 11897293 | 4.01 | 7459790 | 1994609 | 26.74 | 70.85 |
| C4 | C | 74.7 | 12541395 | 4.16 | 8285819 | 2304838 | 27.82 | 70.27 |
| C5 | C | 80.9 | 11418775 | 4.74 | 6786321 | 2245024 | 33.08 | 70.43 |
| P4 | C | 81.1 | 12307525 | 3.33 | 8391565 | 1729506 | 20.61 | 69.99 |
| P5 | C | 81.2 | 11057593 | 4.44 | 6912886 | 2201672 | 31.85 | 71.62 |
| Control | Avg | 79.92 | 13239343.4 | 4.25 | 8258900.8 | 2238261.4 | 27.80 | 69.68 |
| Patient | Avg | 78.8 | 12478300.8 | 4.19 | 7721293.8 | 2150448 | 28.12 | 70.90 |
| All | Avg | 79.36 | 12858822.1 | 4.22 | 7990097.3 | 2194354.7 | 27.96 | 70.29 |

## Appendix 2.2

**A.** Number of genomic features captured in RRBS by each sample in fibroblast

| Feature | C1 | C2 | C3 | C4 | C5 | P1 | P2 | P3 | P4 | P5 | *All samples | total # of features | % of features in all samples |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| promoters | 16100 | 16389 | 15799 | 15652 | 15446 | 16491 | 16769 | 14901 | 16230 | 16384 | 18686 | 28180 | 66.30 |
| exons | 43866 | 45265 | 40795 | 39704 | 38893 | 44346 | 47258 | 34130 | 43881 | 44157 | 62713 | 242221 | 25.89 |
| introns | 71387 | 74197 | 70271 | 69760 | 68389 | 75080 | 75997 | 64738 | 71261 | 73333 | 90438 | 188793 | 47.90 |
| intergenic | 17447 | 17576 | 17249 | 17124 | 17110 | 17574 | 17730 | 16817 | 17431 | 17522 | 18700 | 21508 | 86.94 |

**B.** Number of genomic features captured in RRBS by each sample in iPSC

| Feature | C1 | C2 | C3 | C4 | C5 | P1 | P2 | P3 | P4 | P5 | *All samples | total # of features | % of features in all samples |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| promoters | 15467 | 16401 | 16411 | 16524 | 16025 | 16494 | 16577 | 15869 | 15248 | 16091 | 19458 | 28180 | 69.05 |
| exons | 38711 | 44089 | 46231 | 45644 | 43324 | 45603 | 45212 | 40139 | 36913 | 43356 | 65888 | 242221 | 27.20 |
| introns | 70186 | 75164 | 74271 | 74233 | 72394 | 75095 | 75528 | 71255 | 68008 | 73273 | 98064 | 188793 | 51.94 |
| intergenic | 17210 | 17604 | 17554 | 17591 | 17414 | 17564 | 17626 | 17291 | 17026 | 17425 | 19061 | 21508 | 88.62 |

(*) Features found in all samples were merged together, without any duplicates

## Appendix 2.3

**A.** Full list of genes associated to Shared DMRs included in GO term heart development (GO:0007507) (n=34)

| | | | |
|---|---|---|---|
| TBX3 | SOX11 | ZFPM2 | MYO18B |
| CACNA1C | RPS6KA2 | RBM20 | GLI2 |
| COL5A1 | FOXL1 | MSX1 | ZMIZ1 |
| ERBB4 | DLL1 | SMYD2 | FOXN4 |
| JMJD6 | DNAH5 | BMP7 | |
| PKD1 | SMG9 | MIXL1 | |
| FOLR1 | PDLIM3 | SORBS2 | |
| FOXF1 | RXRA | GATA5 | |
| EYA1 | ZFPM1 | SIX1 | |
| ZFP36L1 | TAB1 | ZBTB14 | |

**B.** Full list of genes associated to Shared DMRs included in GO term skeletal system development (GO: 0001501) (n=37)

| | | | |
|---|---|---|---|
| ALPL | FAM20C | MSX1 | SULF2 |
| ALX3 | FOXP1 | PBX1 | TBX3 |
| BMP7 | GLI2 | PKD1 | TPO |
| CHSY1 | GNAS | RASSF2 | TRPV4 |
| CYTL1 | HMGA2 | RPL13 | WDR5 |
| DLX1 | HOXD10 | RUNX3 | XYLT1 |
| DLX2 | HOXD12 | SIX1 | ZFPM1 |
| DSCAML1 | LHX1 | SNX19 | |
| EYA1 | LRRK1 | SOX11 | |
| FAM101A | MMP2 | SP5 | |

## Appendix 2.4

## OR statistics for fibroblast DMR and histone modifications

| DMR Group | DMR Type | Histone Mark | a* | b* | c* | d* | OR† | log(OR) | p-value‡ |
|---|---|---|---|---|---|---|---|---|---|
| Shared | Hyper | H3K27Ac | 326 | 1015793 | 1126 | 3408263 | 0.97 | -0.01 | 6.62E-01 |
| Shared | Hyper | H3K27me3 | 532 | 1238261 | 920 | 3185589 | 1.49 | 0.17 | 1.02E-12 |
| Shared | Hyper | H3K36me3 | 366 | 1355818 | 1086 | 3068198 | 0.76 | -0.12 | 5.12E-06 |
| Shared | Hyper | H3K4me1 | 445 | 899637 | 1007 | 3524300 | 1.73 | 0.24 | 2.08E-20 |
| Shared | Hyper | H3K4me3 | 328 | 951104 | 1124 | 3472950 | 1.07 | 0.03 | 3.07E-01 |
| Shared | Hyper | H3K9me3 | 308 | 747719 | 1144 | 3676355 | 1.32 | 0.12 | 2.19E-05 |
| Shared | Hypo | H3K27Ac | 116 | 1016003 | 890 | 3408709 | 0.44 | -0.36 | 1.95E-20 |
| Shared | Hypo | H3K27me3 | 361 | 1238432 | 645 | 3186035 | 1.44 | 0.16 | 5.85E-08 |
| Shared | Hypo | H3K36me3 | 235 | 1355949 | 771 | 3068644 | 0.69 | -0.16 | 3.27E-07 |
| Shared | Hypo | H3K4me1 | 169 | 899913 | 837 | 3524746 | 0.79 | -0.10 | 4.78E-03 |
| Shared | Hypo | H3K4me3 | 160 | 951272 | 846 | 3473396 | 0.69 | -0.16 | 9.73E-06 |
| Shared | Hypo | H3K9me3 | 249 | 747778 | 757 | 3676801 | 1.62 | 0.21 | 2.96E-10 |
| Family A | Hyper | H3K27Ac | 1946 | 912786 | 9406 | 3009559 | 0.68 | -0.17 | 2.13E-57 |
| Family A | Hyper | H3K27me3 | 3613 | 1099621 | 7739 | 2821057 | 1.20 | 0.08 | 1.01E-18 |
| Family A | Hyper | H3K36me3 | 2645 | 1211411 | 8707 | 2710235 | 0.68 | -0.17 | 6.64E-72 |
| Family A | Hyper | H3K4me1 | 2502 | 801355 | 8850 | 3120434 | 1.10 | 0.04 | 2.71E-05 |
| Family A | Hyper | H3K4me3 | 1931 | 860760 | 9421 | 3061600 | 0.73 | -0.14 | 5.58E-39 |
| Family A | Hyper | H3K9me3 | 2626 | 661875 | 8726 | 3259790 | 1.48 | 0.17 | 6.65E-65 |
| Family A | Hypo | H3K27Ac | 1540 | 913192 | 5963 | 3013408 | 0.85 | -0.07 | 1.47E-08 |
| Family A | Hypo | H3K27me3 | 2781 | 1100453 | 4722 | 2824906 | 1.51 | 0.18 | 3.08E-64 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Family A | Hypo | H3K36me3 | 1871 | 1212185 | 5632 | 2714084 | 0.74 | -0.13 | 8.57E-30 |
| Family A | Hypo | H3K4me1 | 1896 | 801961 | 5607 | 3124283 | 1.32 | 0.12 | 4.61E-24 |
| Family A | Hypo | H3K4me3 | 1654 | 861037 | 5849 | 3065449 | 1.01 | 0.00 | 8.12E-01 |
| Family A | Hypo | H3K9me3 | 1475 | 663026 | 6028 | 3263639 | 1.20 | 0.08 | 3.37E-10 |
| Family C | Hyper | H3K27Ac | 1405 | 805630 | 7602 | 2623702 | 0.60 | -0.22 | 4.48E-76 |
| Family C | Hyper | H3K27me3 | 3151 | 963853 | 5856 | 2463733 | 1.38 | 0.14 | 1.75E-45 |
| Family C | Hyper | H3K36me3 | 2013 | 1058107 | 6994 | 2370617 | 0.64 | -0.19 | 2.21E-72 |
| Family C | Hyper | H3K4me1 | 1701 | 707970 | 7306 | 2721066 | 0.89 | -0.05 | 3.20E-05 |
| Family C | Hyper | H3K4me3 | 1362 | 763549 | 7645 | 2665826 | 0.62 | -0.21 | 7.55E-65 |
| Family C | Hyper | H3K9me3 | 1696 | 574902 | 7311 | 2854139 | 1.15 | 0.06 | 2.53E-07 |
| Family C | Hypo | H3K27Ac | 1231 | 805804 | 5425 | 2626053 | 0.74 | -0.13 | 8.45E-23 |
| Family C | Hypo | H3K27me3 | 2421 | 964583 | 4235 | 2466084 | 1.46 | 0.16 | 3.34E-48 |
| Family C | Hypo | H3K36me3 | 1586 | 1058534 | 5070 | 2372968 | 0.70 | -0.15 | 8.19E-37 |
| Family C | Hypo | H3K4me1 | 1391 | 708280 | 5265 | 2723417 | 1.02 | 0.01 | 6.06E-01 |
| Family C | Hypo | H3K4me3 | 1352 | 763559 | 5304 | 2668177 | 0.89 | -0.05 | 1.33E-04 |
| Family C | Hypo | H3K9me3 | 1249 | 575349 | 5407 | 2856490 | 1.15 | 0.06 | 1.68E-05 |

(*) *a*, *b*, *c*, and *d* values are the contingency parameters used to calculate OR. (†) OR was calculated as described in the Methods section. (‡) P-values were calculated by Fisher's exact test

## Appendix 2.5

## Odds ratio statistics for fibroblast DMRs and ChromHMM annotations

| DMR Group | DMR Type | Annotation | a* | b* | c* | d* | OR† | log(OR) | p-value‡ |
|---|---|---|---|---|---|---|---|---|---|
| Shared | Hyper | 1_TssA | 11 | 187497 | 1441 | 4236874 | 0.17 | -0.76 | 2.75E-15 |
| Shared | Hyper | 2_PromU | 41 | 176466 | 1411 | 4247875 | 0.70 | -0.16 | 2.23E-02 |
| Shared | Hyper | 3_PromD1 | 33 | 186423 | 1419 | 4237926 | 0.53 | -0.28 | 8.33E-05 |
| Shared | Hyper | 4_PromD2 | 33 | 34490 | 1419 | 4389859 | 2.96 | 0.47 | 1.07E-07 |
| Shared | Hyper | 5_Tx5' | 51 | 126216 | 1401 | 4298115 | 1.24 | 0.09 | 1.34E-01 |
| Shared | Hyper | 6_Tx | 8 | 58069 | 1444 | 4366305 | 0.42 | -0.38 | 7.50E-03 |
| Shared | Hyper | 7_Tx3' | 39 | 313455 | 1413 | 4110888 | 0.36 | -0.44 | 2.24E-13 |
| Shared | Hyper | 8_TxWk | 76 | 357266 | 1376 | 4067040 | 0.63 | -0.20 | 3.26E-05 |
| Shared | Hyper | 9_TxReg | 28 | 32655 | 1424 | 4391699 | 2.64 | 0.42 | 7.20E-06 |
| Shared | Hyper | 10_TxEnh5' | 7 | 20788 | 1445 | 4403587 | 1.03 | 0.01 | 8.47E-01 |
| Shared | Hyper | 11_TxEnh3' | 9 | 17011 | 1443 | 4407362 | 1.62 | 0.21 | 1.37E-01 |
| Shared | Hyper | 12_TxEnhW | 29 | 22785 | 1423 | 4401568 | 3.94 | 0.60 | 1.60E-09 |
| Shared | Hyper | 13_EnhA1 | 4 | 22488 | 1448 | 4401890 | 0.54 | -0.27 | 2.67E-01 |
| Shared | Hyper | 14_EnhA2 | 20 | 17341 | 1432 | 4407021 | 3.55 | 0.55 | 2.26E-06 |
| Shared | Hyper | 15_EnhAF | 23 | 37791 | 1429 | 4386568 | 1.87 | 0.27 | 5.94E-03 |
| Shared | Hyper | 16_EnhW1 | 26 | 39436 | 1426 | 4384920 | 2.03 | 0.31 | 1.08E-03 |
| Shared | Hyper | 17_EnhW2 | 30 | 56505 | 1422 | 4367847 | 1.63 | 0.21 | 1.32E-02 |
| Shared | Hyper | 18_EnhAc | 15 | 12699 | 1437 | 4411668 | 3.63 | 0.56 | 3.06E-05 |
| Shared | Hyper | 19_DNase | 33 | 31149 | 1419 | 4393200 | 3.28 | 0.52 | 1.04E-08 |
| Shared | Hyper | 20_ZNF/Rpts | 4 | 4576 | 1448 | 4419802 | 2.67 | 0.43 | 6.58E-02 |
| Shared | Hyper | 21_Het | 27 | 24691 | 1425 | 4399664 | 3.38 | 0.53 | 1.19E-07 |
| Shared | Hyper | 22_PromP | 24 | 61721 | 1428 | 4362637 | 1.19 | 0.07 | 3.70E-01 |
| Shared | Hyper | 23_PromBiv | 52 | 130099 | 1400 | 4294231 | 1.23 | 0.09 | 1.61E-01 |
| Shared | Hyper | 24_ReprPC | 197 | 367844 | 1255 | 4056341 | 1.73 | 0.24 | 2.06E-11 |
| Shared | Hyper | 25_Quies | 635 | 2080494 | 817 | 2343253 | 0.88 | -0.06 | 1.25E-02 |
| Shared | Hypo | 1_TssA | 14 | 187494 | 992 | 4237320 | 0.32 | -0.50 | 4.20E-07 |
| Shared | Hypo | 2_PromU | 10 | 176497 | 996 | 4248321 | 0.24 | -0.62 | 1.75E-08 |
| Shared | Hypo | 3_PromD1 | 6 | 186450 | 1000 | 4238372 | 0.14 | -0.87 | 3.28E-12 |
| Shared | Hypo | 4_PromD2 | 14 | 34509 | 992 | 4390305 | 1.80 | 0.25 | 4.45E-02 |
| Shared | Hypo | 5_Tx5' | 14 | 126253 | 992 | 4298561 | 0.48 | -0.32 | 3.19E-03 |
| Shared | Hypo | 6_Tx | 14 | 58063 | 992 | 4366751 | 1.06 | 0.03 | 7.81E-01 |
| Shared | Hypo | 7_Tx3' | 29 | 313465 | 977 | 4111334 | 0.39 | -0.41 | 8.33E-09 |
| Shared | Hypo | 8_TxWk | 40 | 357302 | 966 | 4067486 | 0.47 | -0.33 | 2.24E-07 |
| Shared | Hypo | 9_TxReg | 12 | 32671 | 994 | 4392145 | 1.62 | 0.21 | 9.54E-02 |
| Shared | Hypo | 10_TxEnh5' | 11 | 20784 | 995 | 4404033 | 2.34 | 0.37 | 9.20E-03 |
| Shared | Hypo | 11_TxEnh3' | 2 | 17018 | 1004 | 4407808 | 0.52 | -0.29 | 6.03E-01 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Shared | Hypo | 12_TxEnhW | 1 | 22813 | 1005 | 4402014 | 0.19 | -0.72 | 7.29E-02 |
| Shared | Hypo | 13_EnhA1 | 0 | 22492 | 1006 | 4402336 | 0.00 | -Inf | 1.22E-02 |
| Shared | Hypo | 14_EnhA2 | 6 | 17355 | 1000 | 4407467 | 1.52 | 0.18 | 3.01E-01 |
| Shared | Hypo | 15_EnhAF | 3 | 37811 | 1003 | 4387014 | 0.35 | -0.46 | 5.69E-02 |
| Shared | Hypo | 16_EnhW1 | 7 | 39455 | 999 | 4385366 | 0.78 | -0.11 | 6.16E-01 |
| Shared | Hypo | 17_EnhW2 | 3 | 56532 | 1003 | 4368293 | 0.23 | -0.64 | 2.72E-03 |
| Shared | Hypo | 18_EnhAc | 0 | 12714 | 1006 | 4412114 | 0.00 | -Inf | 1.28E-01 |
| Shared | Hypo | 19_DNase | 19 | 31163 | 987 | 4393646 | 2.71 | 0.43 | 1.41E-04 |
| Shared | Hypo | 20_ZNF/Rpts | 1 | 4579 | 1005 | 4420248 | 0.96 | -0.02 | 1.00E+00 |
| Shared | Hypo | 21_Het | 11 | 24707 | 995 | 4400110 | 1.97 | 0.29 | 3.20E-02 |
| Shared | Hypo | 22_PromP | 36 | 61709 | 970 | 4363083 | 2.62 | 0.42 | 5.38E-07 |
| Shared | Hypo | 23_PromBiv | 26 | 130125 | 980 | 4294677 | 0.88 | -0.06 | 5.75E-01 |
| Shared | Hypo | 24_ReprPC | 253 | 367788 | 753 | 4056787 | 3.71 | 0.57 | 2.11E-57 |
| Shared | Hypo | 25_Quies | 475 | 2080654 | 531 | 2343699 | 1.01 | 0.00 | 9.25E-01 |
| Family A | Hyper | 1_TssA | 58 | 171117 | 11294 | 3753116 | 0.11 | -0.95 | 6.38E-141 |
| Family A | Hyper | 2_PromU | 336 | 160571 | 11016 | 3763384 | 0.71 | -0.15 | 2.21E-10 |
| Family A | Hyper | 3_PromD1 | 93 | 170717 | 11259 | 3753481 | 0.18 | -0.74 | 2.19E-111 |
| Family A | Hyper | 4_PromD2 | 204 | 30961 | 11148 | 3893126 | 2.30 | 0.36 | 2.57E-25 |
| Family A | Hyper | 5_Tx5' | 196 | 112115 | 11156 | 3811980 | 0.60 | -0.22 | 1.08E-14 |
| Family A | Hyper | 6_Tx | 126 | 51250 | 11226 | 3872915 | 0.85 | -0.07 | 6.83E-02 |
| Family A | Hyper | 7_Tx3' | 446 | 278242 | 10906 | 3645603 | 0.54 | -0.27 | 8.39E-46 |
| Family A | Hyper | 8_TxWk | 511 | 317826 | 10841 | 3605954 | 0.53 | -0.27 | 3.50E-52 |
| Family A | Hyper | 9_TxReg | 203 | 28851 | 11149 | 3895237 | 2.46 | 0.39 | 1.26E-28 |
| Family A | Hyper | 10_TxEnh5' | 46 | 18226 | 11306 | 3906019 | 0.87 | -0.06 | 4.06E-01 |
| Family A | Hyper | 11_TxEnh3' | 56 | 14927 | 11296 | 3909308 | 1.30 | 0.11 | 5.60E-02 |
| Family A | Hyper | 12_TxEnhW | 82 | 20017 | 11270 | 3904192 | 1.42 | 0.15 | 2.91E-03 |
| Family A | Hyper | 13_EnhA1 | 73 | 19476 | 11279 | 3904742 | 1.30 | 0.11 | 3.20E-02 |
| Family A | Hyper | 14_EnhA2 | 75 | 15074 | 11277 | 3909142 | 1.72 | 0.24 | 1.33E-05 |
| Family A | Hyper | 15_EnhAF | 132 | 32884 | 11220 | 3891275 | 1.39 | 0.14 | 3.01E-04 |
| Family A | Hyper | 16_EnhW1 | 167 | 35231 | 11185 | 3888893 | 1.65 | 0.22 | 2.63E-09 |
| Family A | Hyper | 17_EnhW2 | 189 | 49086 | 11163 | 3875016 | 1.34 | 0.13 | 1.64E-04 |
| Family A | Hyper | 18_EnhAc | 28 | 11248 | 11324 | 3913015 | 0.86 | -0.07 | 4.82E-01 |
| Family A | Hyper | 19_DNase | 125 | 27672 | 11227 | 3896494 | 1.57 | 0.20 | 2.89E-06 |
| Family A | Hyper | 20_ZNF/Rpts | 13 | 4079 | 11339 | 3920199 | 1.10 | 0.04 | 6.61E-01 |
| Family A | Hyper | 21_Het | 183 | 22073 | 11169 | 3902035 | 2.90 | 0.46 | 3.88E-34 |
| Family A | Hyper | 22_PromP | 201 | 54673 | 11151 | 3869417 | 1.28 | 0.11 | 9.97E-04 |
| Family A | Hyper | 23_PromBiv | 380 | 119007 | 10972 | 3804904 | 1.11 | 0.04 | 5.50E-02 |
| Family A | Hyper | 24_ReprPC | 1460 | 325138 | 9892 | 3597693 | 1.63 | 0.21 | 1.03E-60 |
| Family A | Hyper | 25_Quies | 5955 | 1831369 | 5397 | 2086967 | 1.26 | 0.10 | 4.32E-34 |
| Family A | Hypo | 1_TssA | 59 | 171116 | 7444 | 3756965 | 0.17 | -0.76 | 3.71E-76 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Family A | Hypo | 2_PromU | 345 | 160562 | 7158 | 3767233 | 1.13 | 0.05 | 2.86E-02 |
| Family A | Hypo | 3_PromD1 | 178 | 170632 | 7325 | 3757330 | 0.54 | -0.27 | 1.05E-19 |
| Family A | Hypo | 4_PromD2 | 165 | 31000 | 7338 | 3896975 | 2.83 | 0.45 | 8.79E-30 |
| Family A | Hypo | 5_Tx5' | 105 | 112206 | 7398 | 3815829 | 0.48 | -0.32 | 7.70E-17 |
| Family A | Hypo | 6_Tx | 30 | 51346 | 7473 | 3876764 | 0.30 | -0.52 | 1.06E-15 |
| Family A | Hypo | 7_Tx3' | 324 | 278364 | 7179 | 3649452 | 0.59 | -0.23 | 1.72E-23 |
| Family A | Hypo | 8_TxWk | 382 | 317955 | 7121 | 3609803 | 0.61 | -0.22 | 3.72E-24 |
| Family A | Hypo | 9_TxReg | 109 | 28945 | 7394 | 3899086 | 1.99 | 0.30 | 1.33E-10 |
| Family A | Hypo | 10_TxEnh5' | 44 | 18228 | 7459 | 3909868 | 1.27 | 0.10 | 1.25E-01 |
| Family A | Hypo | 11_TxEnh3' | 27 | 14956 | 7476 | 3913157 | 0.94 | -0.02 | 8.51E-01 |
| Family A | Hypo | 12_TxEnhW | 42 | 20057 | 7461 | 3908041 | 1.10 | 0.04 | 5.17E-01 |
| Family A | Hypo | 13_EnhA1 | 48 | 19501 | 7455 | 3908591 | 1.29 | 0.11 | 8.36E-02 |
| Family A | Hypo | 14_EnhA2 | 55 | 15094 | 7448 | 3912991 | 1.91 | 0.28 | 1.37E-05 |
| Family A | Hypo | 15_EnhAF | 106 | 32910 | 7397 | 3895124 | 1.70 | 0.23 | 6.46E-07 |
| Family A | Hypo | 16_EnhW1 | 133 | 35265 | 7370 | 3892742 | 1.99 | 0.30 | 1.08E-12 |
| Family A | Hypo | 17_EnhW2 | 150 | 49125 | 7353 | 3878865 | 1.61 | 0.21 | 7.42E-08 |
| Family A | Hypo | 18_EnhAc | 31 | 11245 | 7472 | 3916864 | 1.45 | 0.16 | 5.03E-02 |
| Family A | Hypo | 19_DNase | 86 | 27711 | 7417 | 3900343 | 1.63 | 0.21 | 2.36E-05 |
| Family A | Hypo | 20_ZNF/Rpts | 8 | 4084 | 7495 | 3924048 | 1.03 | 0.01 | 8.57E-01 |
| Family A | Hypo | 21_Het | 65 | 22191 | 7438 | 3905884 | 1.54 | 0.19 | 1.14E-03 |
| Family A | Hypo | 22_PromP | 154 | 54720 | 7349 | 3873266 | 1.48 | 0.17 | 5.36E-06 |
| Family A | Hypo | 23_PromBiv | 376 | 119011 | 7127 | 3808753 | 1.69 | 0.23 | 4.23E-20 |
| Family A | Hypo | 24_ReprPC | 1147 | 325451 | 6356 | 3601542 | 2.00 | 0.30 | 9.69E-88 |
| Family A | Hypo | 25_Quies | 3335 | 1833989 | 4168 | 2090816 | 0.91 | -0.04 | 7.83E-05 |
| Family C | Hyper | 1_TssA | 28 | 150782 | 8979 | 3279927 | 0.07 | -1.17 | 2.61E-132 |
| Family C | Hyper | 2_PromU | 177 | 144059 | 8830 | 3286501 | 0.46 | -0.34 | 1.23E-31 |
| Family C | Hyper | 3_PromD1 | 69 | 153340 | 8938 | 3277328 | 0.16 | -0.78 | 2.94E-96 |
| Family C | Hyper | 4_PromD2 | 141 | 27367 | 8866 | 3403229 | 1.98 | 0.30 | 4.09E-13 |
| Family C | Hyper | 5_Tx5' | 126 | 97654 | 8881 | 3332957 | 0.48 | -0.31 | 9.69E-20 |
| Family C | Hyper | 6_Tx | 79 | 45364 | 8928 | 3385294 | 0.66 | -0.18 | 1.21E-04 |
| Family C | Hyper | 7_Tx3' | 476 | 244075 | 8531 | 3186186 | 0.73 | -0.14 | 2.08E-12 |
| Family C | Hyper | 8_TxWk | 550 | 273945 | 8457 | 3156242 | 0.75 | -0.13 | 1.03E-11 |
| Family C | Hyper | 9_TxReg | 69 | 26335 | 8938 | 3404333 | 1.00 | 0.00 | 1.00E+00 |
| Family C | Hyper | 10_TxEnh5' | 32 | 15901 | 8975 | 3414804 | 0.77 | -0.12 | 1.40E-01 |
| Family C | Hyper | 11_TxEnh3' | 44 | 13370 | 8963 | 3417323 | 1.25 | 0.10 | 1.49E-01 |
| Family C | Hyper | 12_TxEnhW | 41 | 17594 | 8966 | 3413102 | 0.89 | -0.05 | 5.06E-01 |
| Family C | Hyper | 13_EnhA1 | 38 | 17315 | 8969 | 3413384 | 0.84 | -0.08 | 2.97E-01 |
| Family C | Hyper | 14_EnhA2 | 52 | 13301 | 8955 | 3417384 | 1.49 | 0.17 | 6.37E-03 |
| Family C | Hyper | 15_EnhAF | 99 | 29015 | 8908 | 3401623 | 1.30 | 0.11 | 1.12E-02 |
| Family C | Hyper | 16_EnhW1 | 110 | 31534 | 8897 | 3399093 | 1.33 | 0.12 | 3.99E-03 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Family C | Hyper | 17_EnhW2 | 115 | 43366 | 8892 | 3387256 | 1.01 | 0.00 | 8.87E-01 |
| Family C | Hyper | 18_EnhAc | 23 | 9793 | 8984 | 3420921 | 0.89 | -0.05 | 6.92E-01 |
| Family C | Hyper | 19_DNase | 82 | 24391 | 8925 | 3406264 | 1.28 | 0.11 | 2.79E-02 |
| Family C | Hyper | 20_ZNF/Rpts | 12 | 3572 | 8995 | 3427153 | 1.28 | 0.11 | 4.09E-01 |
| Family C | Hyper | 21_Het | 114 | 19776 | 8893 | 3410847 | 2.21 | 0.34 | 9.60E-14 |
| Family C | Hyper | 22_PromP | 89 | 47610 | 8918 | 3383038 | 0.71 | -0.15 | 8.27E-04 |
| Family C | Hyper | 23_PromBiv | 361 | 107137 | 8646 | 3323239 | 1.30 | 0.11 | 3.42E-06 |
| Family C | Hyper | 24_ReprPC | 1172 | 290148 | 7835 | 3139417 | 1.62 | 0.21 | 1.89E-47 |
| Family C | Hyper | 25_Quies | 4896 | 1580512 | 4111 | 1845329 | 1.39 | 0.14 | 7.21E-55 |
| Family C | Hypo | 1_TssA | 65 | 150745 | 6591 | 3282278 | 0.21 | -0.67 | 1.31E-59 |
| Family C | Hypo | 2_PromU | 232 | 144004 | 6424 | 3288852 | 0.82 | -0.08 | 3.62E-03 |
| Family C | Hypo | 3_PromD1 | 123 | 153286 | 6533 | 3279679 | 0.40 | -0.39 | 3.41E-31 |
| Family C | Hypo | 4_PromD2 | 88 | 27420 | 6568 | 3405580 | 1.66 | 0.22 | 9.46E-06 |
| Family C | Hypo | 5_Tx5' | 94 | 97686 | 6562 | 3335308 | 0.49 | -0.31 | 1.20E-14 |
| Family C | Hypo | 6_Tx | 50 | 45393 | 6606 | 3387645 | 0.56 | -0.25 | 1.27E-05 |
| Family C | Hypo | 7_Tx3' | 311 | 244240 | 6345 | 3188537 | 0.64 | -0.19 | 2.73E-16 |
| Family C | Hypo | 8_TxWk | 311 | 274184 | 6345 | 3158593 | 0.56 | -0.25 | 9.77E-27 |
| Family C | Hypo | 9_TxReg | 78 | 26326 | 6578 | 3406684 | 1.53 | 0.19 | 4.13E-04 |
| Family C | Hypo | 10_TxEnh5' | 57 | 15876 | 6599 | 3417155 | 1.86 | 0.27 | 1.83E-05 |
| Family C | Hypo | 11_TxEnh3' | 22 | 13392 | 6634 | 3419674 | 0.85 | -0.07 | 4.91E-01 |
| Family C | Hypo | 12_TxEnhW | 28 | 17607 | 6628 | 3415453 | 0.82 | -0.09 | 3.44E-01 |
| Family C | Hypo | 13_EnhA1 | 78 | 17275 | 6578 | 3415735 | 2.34 | 0.37 | 5.14E-11 |
| Family C | Hypo | 14_EnhA2 | 40 | 13313 | 6616 | 3419735 | 1.55 | 0.19 | 9.80E-03 |
| Family C | Hypo | 15_EnhAF | 47 | 29067 | 6609 | 3403974 | 0.83 | -0.08 | 2.28E-01 |
| Family C | Hypo | 16_EnhW1 | 96 | 31548 | 6560 | 3401444 | 1.58 | 0.20 | 3.55E-05 |
| Family C | Hypo | 17_EnhW2 | 83 | 43398 | 6573 | 3389607 | 0.99 | -0.01 | 9.56E-01 |
| Family C | Hypo | 18_EnhAc | 21 | 9795 | 6635 | 3423272 | 1.11 | 0.04 | 6.44E-01 |
| Family C | Hypo | 19_DNase | 78 | 24395 | 6578 | 3408615 | 1.66 | 0.22 | 3.87E-05 |
| Family C | Hypo | 20_ZNF/Rpts | 3 | 3581 | 6653 | 3429504 | 0.43 | -0.36 | 1.79E-01 |
| Family C | Hypo | 21_Het | 92 | 19798 | 6564 | 3413198 | 2.42 | 0.38 | 1.53E-13 |
| Family C | Hypo | 22_PromP | 177 | 47522 | 6479 | 3385389 | 1.95 | 0.29 | 2.64E-15 |
| Family C | Hypo | 23_PromBiv | 231 | 107267 | 6425 | 3325590 | 1.11 | 0.05 | 1.05E-01 |
| Family C | Hypo | 24_ReprPC | 1062 | 290258 | 5594 | 3141768 | 2.05 | 0.31 | 4.78E-87 |
| Family C | Hypo | 25_Quies | 3189 | 1582219 | 3467 | 1847680 | 1.07 | 0.03 | 3.68E-03 |

(*) *a*, *b*, *c*, and *d* values are the contingency parameters used to calculate OR. (†) OR was calculated as described in the Methods section. (‡) P-values were calculated by Fisher's exact test.

## Appendix 2.6

## Complete list of TFBS motif enrichment for fibroblast DMRs acquired from HOMER

| DMR group | DMR type | Rank | Motif | Related Gene Name | HOMER TF Name | -Log(p-value)* |
|---|---|---|---|---|---|---|
| Family A | Hypo | 1 | TCAGACGTAGTCTCGAAGTCTCAGGC ATCTAGCTAGAGCT | USF2 | Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer | 19.61 |
| Family A | Hypo | 2 | TCGATCAGTCGAACTGCATGACGTAG TCCTGA | NR2F2 | COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer | 12.41 |
| Family A | Hypo | 3 | AGTCTGCATCGACTGAACTGCATGAC GTATGCGTCATACG | ESRRA | Erra(NR)/HepG2-Erra-ChIP-Seq(GSE31477)/Homer | 9.41 |
| Family A | Hypo | 4 | TACGTCAGGATCGTACTCGAGACTGC TAGCTAGCTACGTAGATCGTCA | CDX4 | CDX4(Homeobox)/ZebrafishEmbryos-Cdx4.Myc-ChIP-Seq(GSE48254)/Homer | 9.19 |
| Family A | Hypo | 5 | ATGCGACTACTGCAGTGATCACGTTA CGTACG | SMAD2 | Smad2(MAD)/ES-SMAD2-ChIP-Seq(GSE29422)/Homer | 8.00 |
| Family A | Hypo | 6 | AGTCGACTCAGTGTACAGTCATCGTC AGACTGGTCACGTA | STAT3 | Stat3(Stat)/mES-Stat3-ChIP-Seq(GSE11431)/Homer | 6.48 |
| Family A | Hypo | 7 | AGCTGACTCTAGCGTACATGCGATCT AGATCGGACTCAGT | NKX3-2 | Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer | 6.37 |
| Family A | Hypo | 8 | CGATTAGCGACTGTCACGTAACGTCG TACGTACGTAGCTA | HOXD13 | HOXD13(Homeobox)/Chicken-Hoxd13-ChIP-Seq(GSE38910)/Homer | 6.22 |
| Family A | Hypo | 9 | TACGATCGTAGCGATCACTGACGTAG TCACGTCTAGATCG | SMAD4 | Smad4(MAD)/ESC-SMAD4-ChIP-Seq(GSE29422)/Homer | 5.91 |
| Family A | Hypo | 10 | CATGGTACGACTGCTACGTACGTACG TAGCTAGACTCTGATCAGGTAC | MEF2C | Mef2c(MADS)/GM12878-Mef2c-ChIP-Seq(GSE32465)/Homer | 5.86 |
| Family A | Hypo | 11 | GTACGACTCGTACTGATCGACGTAGC TACAGTCTGATACG | MEF2A | Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer | 5.62 |
| Family A | Hypo | 12 | CATGAGTCGACTCGTACGATGCATGA CTGCATCGATCTAGCATGTGAC | MEF2B | Mef2b(MADS)/HEK293-Mef2b.V5-ChIP-Seq(GSE67450)/Homer | 5.49 |
| Family A | Hypo | 13 | CGATACGTACGTACGTCGTAAGCTCA GTCTAGATCGACTG | HOXB13 | HOXB13(Homeobox)/Prostate Tumor-HOXB13-ChIP-Seq(GSE56288)/Homer | 5.37 |
| Family A | Hypo | 14 | CTGATCGACGTAATGCCGTACGTACG ATCTAGTCAGGATC | SOX15 | Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer | 5.17 |
| Family A | Hypo | 15 | CGTATGACTCGAAGTCCGTAATCGAT GCACGTACTGAGTC | TCF3 | E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer | 4.69 |
| Family A | Hypo | 16 | ATCGAGCTCTGACTAGACTGACGTGT ACGCTAATGCACGTCTAGCATGTACG CGATATGCCGTA | NR1D1 | Reverb(NR),DR2/RAW-Reverba.biotin-ChIP-Seq(GSE45914)/Homer | 4.67 |
| Family A | Hyper | 1 | CATGGTACGACTGCTACGTACGTACG TAGCTAGACTCTGATCAGGTAC | MEF2C | Mef2c(MADS)/GM12878-Mef2c-ChIP-Seq(GSE32465)/Homer | 26.42 |
| Family A | Hyper | 2 | GTACGACTCGTACTGATCGACGTAGC TACAGTCTGATACG | MEF2A | Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer | 25.80 |
| Family A | Hyper | 3 | CATGAGTCGACTCGTACGATGCATGA CTGCATCGATCTAGCATGTGAC | MEF2B | Mef2b(MADS)/HEK293-Mef2b.V5-ChIP-Seq(GSE67450)/Homer | 21.14 |
| Family A | Hyper | 4 | AGCTGACTCTAGCGTACATGCGATCT AGATCGGACTCAGT | NKX3-2 | Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer | 18.29 |
| Family A | Hyper | 5 | TCAGACGTAGTCTCGAAGTCTCAGGC ATCTAGCTAGAGCT | USF2 | Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer | 14.54 |
| Family A | Hyper | 6 | ATGCGACTAGCTCTAGCGTACTAGCG ATCTAGATCGGATC | NKX2-2 | Nkx2.2(Homeobox)/NPC-Nkx2.2-ChIP-Seq(GSE61673)/Homer | 9.68 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Family A | Hyper | 7 | CTAGTACGAGTCCGTAAGTCACGTAG TCTCGACGTATACG | NKX2-1 | Nkx2.1(Homeobox)/LungAC-Nkx2.1-ChIP-Seq(GSE43252)/Homer | 8.58 |
| Family A | Hyper | 8 | CATGAGCTTACGGTCAGTACTAGCAG CTGACTATCGTCGA | ESRRB | Esrrb(NR)/mES-Esrrb-ChIP-Seq(GSE11431)/Homer | 8.33 |
| Family A | Hyper | 9 | CTGACTGATAGCGATCGCTAGTACAC GTGATCTGCACGTA | NKX2-5 | Nkx2.5(Homeobox)/HL1-Nkx2.5.biotin-ChIP-Seq(GSE21529)/Homer | 8.19 |
| Family A | Hyper | 10 | GATCCTGAAGTCCGATCGATGATCAG TCACTGATCGAGCT | ELK4 | Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer | 7.10 |
| Family A | Hyper | 11 | TCAGTCAGTAGCAGTCCTGAAGTCCT AGACGTACTGATCG | MYC | c-Myc(bHLH)/mES-cMyc-ChIP-Seq(GSE11431)/Homer | 7.02 |
| Family A | Hyper | 12 | ACGTCTAGAGCTACGTACGTCTGAAG TCGACTAGCTCGTA | FOXM1 | FOXM1(Forkhead)/MCF7-FOXM1-ChIP-Seq(GSE72977)/Homer | 6.88 |
| Family A | Hyper | 13 | TACGGACTACTGACTGCTAGATGCAG TCAGTCAGTCCTGA | ZNF692 | ZNF692(Zf)/HEK293-ZNF692.GFP-ChIP-Seq(GSE58341)/Homer | 6.84 |
| Family A | Hyper | 14 | TCGATCAGTCGAACTGCATGACGTAG TCCTGA | NR2F2 | COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer | 6.46 |
| Family A | Hyper | 15 | ATCGAGCTCTGACTAGACTGACGTGT ACGCTAATGCACGTCTAGCATGTACG CGATATGCCGTA | NR1D1 | Reverb(NR),DR2/RAW-Reverba.biotin-ChIP-Seq(GSE45914)/Homer | 6.35 |
| Family A | Hyper | 16 | GTACCTAGTCAGAGCTTAGCCGTAAT GCTACGAGTCGTACGTCAAGTC | SREBF2 | Srebp2(bHLH)/HepG2-Srebp2-ChIP-Seq(GSE31477)/Homer | 6.14 |
| Family A | Hyper | 17 | TCGATGACAGTCCGTAAGTCCTAGAC GTACTGACTGAGCTAGTCGCAT | MAX | Max(bHLH)/K562-Max-ChIP-Seq(GSE31477)/Homer | 5.96 |
| Family A | Hyper | 18 | TGACCGTACTGAACTGACTGGACTGA TCTGCAGTACTACG | SF1 | SF1(NR)/H295R-Nr5a1-ChIP-Seq(GSE44220)/Homer | 5.54 |
| Family A | Hyper | 19 | ACGTGACTTAGCCGTACTGACATGCT AGGACTGATCCGTA | NR5A2 | Nr5a2(NR)/Pancreas-LRH1-ChIP-Seq(GSE34295)/Homer | 5.41 |
| Family A | Hyper | 20 | ATGCTAGCAGCTAGCTTGACGACTTC AGTACGGTCACTGAATCGTAGCGACT CAGTAGTCAGCTTCGAATCGTGCATG CA | HSF1 | HRE(HSF)/HepG2-HSF1-ChIP-Seq(GSE31477)/Homer | 5.40 |
| Family A | Hyper | 21 | CTGATCGACGTAATGCCGTACGTACG ATCTAGTCAGGATC | SOX15 | Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer | 4.97 |
| Family A | Hyper | 22 | GCATGCATCTGAACGTCTGAACGTCG TACGTACGTAAGTCGTCAGTCA | FOXF1 | Foxf1(Forkhead)/Lung-Foxf1-ChIP-Seq(GSE77951)/Homer | 4.68 |
| Family A | Hyper | 23 | GCATTCAGCTGAATCGACTGCGATGA TCCTGA | THRB | THRb(NR)/Liver-NR1A2-ChIP-Seq(GSE52613)/Homer | 4.66 |
| Family C | Hypo | 1 | GCATGCATCTGAACGTCTGAACGTCG TACGTACGTAAGTCGTCAGTCA | FOXF1 | Foxf1(Forkhead)/Lung-Foxf1-ChIP-Seq(GSE77951)/Homer | 12.57 |
| Family C | Hypo | 2 | CGTAGCTACGATCTAGACGTGTCACG TACGTAAGTCCGTATGCATACG | FOXL2 | FoxL2(Forkhead)/Ovary-FoxL2-ChIP-Seq(GSE60858)/Homer | 11.90 |
| Family C | Hypo | 3 | ACGTCTAGAGCTACGTACGTCTGAAG TCGACTAGCTCGTA | FOXM1 | FOXM1(Forkhead)/MCF7-FOXM1-ChIP-Seq(GSE72977)/Homer | 10.01 |
| Family C | Hypo | 4 | CATGGTACGACTGCTACGTACGTACG TAGCTAGACTCTGATCAGGTAC | MEF2C | Mef2c(MADS)/GM12878-Mef2c-ChIP-Seq(GSE32465)/Homer | 9.77 |
| Family C | Hypo | 5 | AGCTGACTCTAGCGTACATGCGATCT AGATCGGACTCAGT | NKX3-2 | Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer | 9.62 |
| Family C | Hypo | 6 | TCAGACGTAGTCTCGAAGTCTCAGGC ATCTAGCTAGAGCT | USF2 | Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer | 9.57 |
| Family C | Hypo | 7 | GCTATCGACGTACTAGAGCTGTCAGT CACGTAAGTCCGTA | FOXA1 | FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq(GSE26831)/Homer | 8.61 |
| Family C | Hypo | 8 | GCTATCGACGTACTAGAGCTGTCAGT CACGTAAGTCCGTA | FOXA1 | FOXA1(Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer | 7.53 |
| Family C | Hypo | 9 | GCATATCGCATGGTACGCTAAGTCTC AGTGACGTCATGCA | ARNT, AHR | Arnt:Ahr(bHLH)/MCF7-Arnt-ChIP-Seq(Lo_et_al.)/Homer | 7.51 |
| Family C | Hypo | 10 | TCAGAGCTGTACCGTAACGTCGTACG TACGTAGCTAGACT | CDX2 | Cdx2(Homeobox)/mES-Cdx2-ChIP-Seq(GSE14586)/Homer | 7.35 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Family C | Hypo | 11 | ATGCGACTAGCTCTAGCGTACTAGCG ATCTAGATCGGATC | NKX2-2 | Nkx2.2(Homeobox)/NPC-Nkx2.2-ChIP-Seq(GSE61673)/Homer | 6.83 |
| Family C | Hypo | 12 | CGATTAGCGACTGTCACGTAACGTCG TACGTACGTAGCTA | HOXD13 | HOXD13(Homeobox)/Chicken-Hoxd13-ChIP-Seq(GSE38910)/Homer | 6.53 |
| Family C | Hypo | 13 | GTACGACTCGTACTGATCGACGTAGC TACAGTCTGATACG | MEF2A | Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer | 6.29 |
| Family C | Hypo | 14 | CTAGTACGAGTCCGTAAGTCACGTAG TCTCGACGTATACG | NKX2-1 | Nkx2.1(Homeobox)/LungAC-Nkx2.1-ChIP-Seq(GSE43252)/Homer | 6.02 |
| Family C | Hypo | 15 | CGATACGTACGTACGTCGTAAGCTCA GTCTAGATCGACTG | HOXB13 | HOXB13(Homeobox)/Prostate Tumor-HOXB13-ChIP-Seq(GSE56288)/Homer | 5.55 |
| Family C | Hypo | 16 | GCTAACGTCTAGGTACGCTAGACTCT GAGCATCATGGATC | POU1F1 | Pit1(Homeobox)/GCrat-Pit1-ChIP-Seq(GSE58009)/Homer | 4.99 |
| Family C | Hypo | 17 | CATGAGTCGACTCGTACGATGCATGA CTGCATCGATCTAGCATGTGAC | MEF2B | Mef2b(MADS)/HEK293-Mef2b.V5-ChIP-Seq(GSE67450)/Homer | 4.72 |
| Family C | Hypo | 18 | TGCACGTAGTCAAGCTAGTCGCTATA GCCGATCTAGGATC | GFI1B | Gfi1b(Zf)/HPC7-Gfi1b-ChIP-Seq(GSE22178)/Homer | 4.70 |
| Family C | Hyper | 1 | CATGGTACGACTGCTACGTACGTACG TAGCTAGACTCTGATCAGGTAC | MEF2C | Mef2c(MADS)/GM12878-Mef2c-ChIP-Seq(GSE32465)/Homer | 22.11 |
| Family C | Hyper | 2 | AGCTGACTCTAGCGTACATGCGATCT AGATCGGACTCAGT | NKX3-2 | Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer | 19.74 |
| Family C | Hyper | 3 | TCAGACGTAGTCTCGAAGTCTCAGGC ATCTAGCTAGAGCT | USF2 | Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer | 16.78 |
| Family C | Hyper | 4 | GTACGACTCGTACTGATCGACGTAGC TACAGTCTGATACG | MEF2A | Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer | 16.38 |
| Family C | Hyper | 5 | CATGAGTCGACTCGTACGATGCATGA CTGCATCGATCTAGCATGTGAC | MEF2B | Mef2b(MADS)/HEK293-Mef2b.V5-ChIP-Seq(GSE67450)/Homer | 13.09 |
| Family C | Hyper | 6 | GATCTCGAAGTCCGATCGATAGTCAT GCACTGATCGGACT | ELK1 | Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer | 11.99 |
| Family C | Hyper | 7 | GATCCTGAAGTCCGATCGATGATCAG TCACTGATCGAGCT | ELK4 | Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer | 11.53 |
| Family C | Hyper | 8 | ATGCGACTAGCTCTAGCGTACTAGCG ATCTAGATCGGATC | NKX2-2 | Nkx2.2(Homeobox)/NPC-Nkx2.2-ChIP-Seq(GSE61673)/Homer | 8.19 |
| Family C | Hyper | 9 | CTGACAGTCTGAAGTCCTAGGACTAT CGGTAC | ARNT | HIF-1b(HLH)/T47D-HIF1b-ChIP-Seq(GSE59937)/Homer | 7.79 |
| Family C | Hyper | 10 | TCGATCAGTCGAACTGCATGACGTAG TCCTGA | NR2F2 | COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer | 7.47 |
| Family C | Hyper | 11 | GTACCTAGTCAGAGCTTAGCCGTAAT GCTACGAGTCGTACGTCAAGTC | SREBF2 | Srebp2(bHLH)/HepG2-Srebp2-ChIP-Seq(GSE31477)/Homer | 7.24 |
| Family C | Hyper | 12 | TCGAGCATATGCCTGAATGCTAGCAG TCGTACTCGAAGCT | SREBF1 | Srebp1a(bHLH)/HepG2-Srebp1a-ChIP-Seq(GSE31477)/Homer | 6.65 |
| Family C | Hyper | 13 | TCAGTCAGTAGCAGTCCTGAAGTCCT AGACGTACTGATCG | MYC | c-Myc(bHLH)/mES-cMyc-ChIP-Seq(GSE11431)/Homer | 6.46 |
| Family C | Hyper | 14 | CGTACTAGGACTGTCAGTCACGTAAG TCCGTATCGATCGATCGACGTACTGA CTAGGCTACGTATAGCCGTACGATCG TA | FOXA1 | FOXA1:AR(Forkhead,NR)/LNCAP-AR-ChIP-Seq(GSE27824)/Homer | 6.26 |
| Family C | Hyper | 15 | TACGTCGATAGCAGTCCGTAAGTCCT AGGCATACTGATCG | MYCN | n-Myc(bHLH)/mES-nMyc-ChIP-Seq(GSE11431)/Homer | 6.25 |
| Family C | Hyper | 16 | AGTCCTGAAGTCCGATCAGTGATCAT GCACTGATCGGACT | FLI1 | Fli1(ETS)/CD8-FLI-ChIP-Seq(GSE20898)/Homer | 5.95 |
| Family C | Hyper | 17 | CTGACTGATAGCGATCGCTAGTACAC GTGATCTGCACGTA | NKX2-5 | Nkx2.5(Homeobox)/HL1-Nkx2.5.biotin-ChIP-Seq(GSE21529)/Homer | 5.92 |
| Family C | Hyper | 18 | CGTAACTGGTCAACGTATCGCAGTCT AGTCAGCGTAACTGCGTAACGTCGTA CTGATACG | GATA3 | GATA3(Zf),DR4/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer | 5.62 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Family C | Hyper | 19 | CTAGCATGACGTAGTCGCTAAGCTAGTCAGCTTCAGCTGAACTGCATGGCATATGCCGTA | THRA | THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer | 5.48 |
| Family C | Hyper | 20 | TCGATGACAGTCCGTAAGTCCTAGACGTACTGACTGAGCTAGTCGCAT | MAX | Max(bHLH)/K562-Max-ChIP-Seq(GSE31477)/Homer | 5.30 |
| Family C | Hyper | 21 | ACGTCTAGAGCTACGTACGTCTGAAGTCGACTAGCTCGTA | FOXM1 | FOXM1(Forkhead)/MCF7-FOXM1-ChIP-Seq(GSE72977)/Homer | 5.25 |
| Family C | Hyper | 22 | CTAGTACGAGTCCGTAAGTCACGTAGTCTCGACGTATACG | NKX2-1 | Nkx2.1(Homeobox)/LungAC-Nkx2.1-ChIP-Seq(GSE43252)/Homer | 4.91 |
| Family C | Hyper | 23 | ATCGAGCTCTGACTAGACTGACGTGTACGCTAATGCACGTCTAGCATGTACGCGATATGCCGTA | NR1D1 | Reverb(NR),DR2/RAW-Reverba.biotin-ChIP-Seq(GSE45914)/Homer | 4.91 |
| Family C | Hyper | 24 | CTGATCGACGTAATGCCGTACGTACGATCTAGTCAGGATC | SOX15 | Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer | 4.79 |
| Family C | Hyper | 25 | AGTCATCGGCATCTAGACTGTACGCGATTCAGCATGAGCTTAGCGATC | GLI3 | GLI3(Zf)/Limb-GLI3-ChIP-Chip(GSE11077)/Homer | 4.62 |
| Family C | Hyper | 26 | CATGGTACCGTAAGTCCTAGACGTACTGGTACAGTCAGCT | BHLH40E | bHLHE40(bHLH)/HepG2-BHLHE40-ChIP-Seq(GSE31477)/Homer | 4.61 |
| Shared | Hypo | 1 | TGCAAGCTACGTCTAGGATCCTAGGATCGTCACTGAAGTC | CEBPB | CEBP(bZIP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer | 5.68 |
| Shared | Hypo | 2 | TGCAACTGTACGATGCAGTCGACTTCGAATCG | ZNF711 | ZNF711(Zf)/SHSY5Y-ZNF711-ChIP-Seq(GSE20673)/Homer | 5.41 |
| Shared | Hypo | 3 | CGTATGACTCGAAGTCCGTAATCGATGCACGTACTGAGTC | TCF3 | E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer | 5.19 |
| Shared | Hyper | 1 | GCTATCGACGTACTAGAGCTGTCAGTCACGTAAGTCCGTA | FOXA1 | FOXA1(Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer | 5.85 |
| Shared | Hyper | 2 | GCTATCGACGTACTAGAGCTGTCAGTCACGTAAGTCCGTA | FOXA1 | FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq(GSE26831)/Homer | 4.93 |
| Shared | Hyper | 3 | GCATGCATCTGAACGTCTGAACGTCGTACGTACGTAAGTCGTCAGTCA | FOXF1 | Foxf1(Forkhead)/Lung-Foxf1-ChIP-Seq(GSE77951)/Homer | 4.87 |

(*) Significance of the motif is displayed in the last column as -log(p-value), calculated using the hypergeometric test through HOMER[58].

## Appendinx 2.7

## Complete list of disease ontology terms from ToppGene for gene lists associated with either hypo or hypermethylated fibroblast DMR contexts

| DMR group | DMR type | Rank | ID* | Name | Source | p-value† | FDR B&H‡ |
|---|---|---|---|---|---|---|---|
| Shared | Hyper | 1 | C0014544 | Epilepsy | DisGeNET BeFree | 1.43E-06 | 1.04E-02 |
| Shared | Hyper | 2 | C1535926 | Neurodevelopmental Disorders | DisGeNET Curated | 3.30E-06 | 1.20E-02 |
| Shared | Hyper | 3 | 20090507: Lasky-Su | Hyperactive-impulsive symptoms | GWAS | 8.74E-06 | 2.12E-02 |
| Shared | Hyper | 4 | C0086743 | Osteoarthrosis Deformans | DisGeNET Curated | 1.81E-05 | 2.25E-02 |
| Shared | Hyper | 5 | C0029408 | Degenerative polyarthritis | DisGeNET Curated | 1.81E-05 | 2.25E-02 |
| Shared | Hyper | 6 | C3714756 | Intellectual Disability | DisGeNET BeFree | 1.86E-05 | 2.25E-02 |
| Shared | Hypo | - | - | None | - | - | - |
| Family A | Hyper | 1 | C0028754 | Obesity | DisGeNET BeFree | 1.50E-09 | 1.65E-05 |
| Family A | Hyper | 2 | C0014544 | Epilepsy | DisGeNET BeFree | 3.12E-09 | 1.65E-05 |
| Family A | Hyper | 3 | C0036341 | Schizophrenia | DisGeNET BeFree | 3.86E-09 | 1.65E-05 |
| Family A | Hyper | 4 | C0001418 | Adenocarcinoma | DisGeNET BeFree | 6.22E-09 | 1.65E-05 |
| Family A | Hyper | 5 | C0278878 | Adult Glioblastoma | DisGeNET BeFree | 6.57E-09 | 1.65E-05 |
| Family A | Hyper | 6 | C0280474 | Childhood Glioblastoma | DisGeNET BeFree | 6.57E-09 | 1.65E-05 |
| Family A | Hyper | 7 | C0027765 | nervous system disorder | DisGeNET BeFree | 6.76E-09 | 1.65E-05 |
| Family A | Hyper | 8 | C0338656 | Impaired cognition | DisGeNET BeFree | 1.69E-08 | 3.62E-05 |
| Family A | Hyper | 9 | C0007758 | Cerebellar Ataxia | DisGeNET BeFree | 3.57E-08 | 6.78E-05 |
| Family A | Hyper | 10 | C0699790 | Colon Carcinoma | DisGeNET BeFree | 4.71E-08 | 8.05E-05 |
| Family A | Hyper | 11 | C3714756 | Intellectual Disability | DisGeNET BeFree | 5.29E-08 | 8.21E-05 |
| Family A | Hyper | 12 | C1535926 | Neurodevelopmental Disorders | DisGeNET BeFree | 8.99E-08 | 1.28E-04 |
| Family A | Hyper | 13 | C0001973 | Alcoholic Intoxication, Chronic | DisGeNET Curated | 9.80E-08 | 1.29E-04 |
| Family A | Hyper | 14 | C0007102 | Malignant tumor of colon | DisGeNET BeFree | 2.44E-07 | 2.98E-04 |
| Family A | Hyper | 15 | C1510586 | Autism Spectrum Disorders | DisGeNET BeFree | 3.50E-07 | 3.98E-04 |
| Family A | Hyper | 16 | C0344315 | Depressed mood | DisGeNET BeFree | 4.49E-07 | 4.80E-04 |
| Family A | Hyper | 17 | C0009319 | Colitis | DisGeNET BeFree | 5.35E-07 | 5.38E-04 |
| Family A | Hyper | 18 | C0011570 | Mental Depression | DisGeNET BeFree | 7.66E-07 | 7.27E-04 |
| Family A | Hyper | 19 | C0007097 | Carcinoma | DisGeNET BeFree | 9.95E-07 | 8.95E-04 |
| Family A | Hyper | 20 | C0025286 | Meningioma | DisGeNET BeFree | 1.09E-06 | 9.32E-04 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family A | Hyper | 21 | C0011581 | Depressive disorder | DisGeNET BeFree | 1.17E-06 | 9.48E-04 |
| Family A | Hyper | 22 | C0013384 | Dyskinetic syndrome | DisGeNET BeFree | 1.22E-06 | 9.48E-04 |
| Family A | Hyper | 23 | C0011849 | Diabetes Mellitus | DisGeNET BeFree | 1.95E-06 | 1.45E-03 |
| Family A | Hyper | 24 | C0085281 | Addictive Behavior | DisGeNET BeFree | 2.33E-06 | 1.66E-03 |
| Family A | Hyper | 25 | C0002736 | Amyotrophic Lateral Sclerosis | DisGeNET BeFree | 2.92E-06 | 2.00E-03 |
| Family A | Hyper | 26 | C0557874 | Global developmental delay | DisGeNET BeFree | 4.01E-06 | 2.61E-03 |
| Family A | Hyper | 27 | C0007785 | Cerebral Infarction | DisGeNET BeFree | 4.12E-06 | 2.61E-03 |
| Family A | Hyper | 28 | C0011847 | Diabetes | DisGeNET BeFree | 4.39E-06 | 2.68E-03 |
| Family A | Hyper | 29 | C0042769 | Virus Diseases | DisGeNET BeFree | 4.78E-06 | 2.82E-03 |
| Family A | Hyper | 30 | C0524851 | Neurodegenerative Disorders | DisGeNET BeFree | 6.01E-06 | 3.42E-03 |
| Family A | Hyper | 31 | C0005586 | Bipolar Disorder | DisGeNET Curated | 8.40E-06 | 4.47E-03 |
| Family A | Hyper | 32 | C0006142 | Malignant neoplasm of breast | DisGeNET Curated | 8.73E-06 | 4.47E-03 |
| Family A | Hyper | 33 | C0026764 | Multiple Myeloma | DisGeNET BeFree | 8.76E-06 | 4.47E-03 |
| Family A | Hyper | 34 | C1458155 | Mammary Neoplasms | DisGeNET BeFree | 8.89E-06 | 4.47E-03 |
| Family A | Hyper | 35 | C0424605 | Developmental delay (disorder) | DisGeNET BeFree | 9.50E-06 | 4.64E-03 |
| Family A | Hyper | 36 | C0000768 | Congenital Abnormality | DisGeNET BeFree | 1.11E-05 | 5.25E-03 |
| Family A | Hyper | 37 | C0010054 | Coronary Arteriosclerosis | DisGeNET BeFree | 1.18E-05 | 5.45E-03 |
| Family A | Hyper | 38 | C1389018 | Atrioventricular Septal Defect | DisGeNET BeFree | 1.27E-05 | 5.69E-03 |
| Family A | Hyper | 39 | C0271650 | Impaired glucose tolerance | DisGeNET BeFree | 1.31E-05 | 5.75E-03 |
| Family A | Hyper | 40 | C0278877 | Adult Meningioma | DisGeNET BeFree | 1.43E-05 | 5.99E-03 |
| Family A | Hyper | 41 | C0036341 | Schizophrenia | DisGeNET Curated | 1.48E-05 | 5.99E-03 |
| Family A | Hyper | 42 | C0026838 | Muscle Spasticity | DisGeNET BeFree | 1.52E-05 | 5.99E-03 |
| Family A | Hyper | 43 | C0006118 | Brain Neoplasms | DisGeNET BeFree | 1.54E-05 | 5.99E-03 |
| Family A | Hyper | 44 | C0025202 | melanoma | DisGeNET BeFree | 1.54E-05 | 5.99E-03 |
| Family A | Hyper | 45 | C0030193 | Pain | DisGeNET BeFree | 1.64E-05 | 6.21E-03 |
| Family A | Hyper | 46 | C1328504 | Hormone refractory prostate cancer | DisGeNET BeFree | 1.78E-05 | 6.63E-03 |
| Family A | Hyper | 47 | C0007959 | Charcot-Marie-Tooth Disease | DisGeNET BeFree | 1.96E-05 | 7.12E-03 |
| Family A | Hyper | 48 | C0010068 | Coronary heart disease | DisGeNET BeFree | 2.11E-05 | 7.53E-03 |
| Family A | Hyper | 49 | C1762616 | Meningioma, benign, no ICD-O subtype | DisGeNET BeFree | 2.30E-05 | 8.01E-03 |
| Family A | Hyper | 50 | C0004936 | Mental disorders | DisGeNET BeFree | 2.41E-05 | 8.11E-03 |
| Family A | Hyper | 51 | C0234958 | Muscle degeneration | DisGeNET BeFree | 2.42E-05 | 8.11E-03 |
| Family A | Hyper | 52 | C0025958 | Microcephaly | DisGeNET BeFree | 2.65E-05 | 8.57E-03 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family A | Hyper | 53 | C0019569 | Hirschsprung Disease | DisGeNET BeFree | 2.77E-05 | 8.57E-03 |
| Family A | Hyper | 54 | C1611743 | Familial (FPAH) | DisGeNET BeFree | 2.85E-05 | 8.57E-03 |
| Family A | Hyper | 55 | C3539878 | Triple Negative Breast Neoplasms | DisGeNET BeFree | 2.89E-05 | 8.57E-03 |
| Family A | Hyper | 56 | C0153690 | Secondary malignant neoplasm of bone | DisGeNET BeFree | 3.15E-05 | 8.57E-03 |
| Family A | Hyper | 57 | C0700095 | Central neuroblastoma | DisGeNET BeFree | 3.24E-05 | 8.57E-03 |
| Family A | Hyper | 58 | C4316881 | Prescription Drug Abuse | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 59 | C0013170 | Drug habituation | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 60 | C0013146 | Drug abuse | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 61 | C0013222 | Drug Use Disorders | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 62 | C0038580 | Substance Dependence | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 63 | C0038586 | Substance Use Disorders | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 64 | C0236969 | Substance-Related Disorders | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 65 | C1510472 | Drug Dependence | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 66 | C0029231 | Organic Mental Disorders, Substance-Induced | DisGeNET Curated | 3.34E-05 | 8.57E-03 |
| Family A | Hyper | 67 | C4086165 | Childhood Neuroblastoma | DisGeNET BeFree | 3.36E-05 | 8.57E-03 |
| Family A | Hyper | 68 | C0027819 | Neuroblastoma | DisGeNET BeFree | 3.45E-05 | 8.67E-03 |
| Family A | Hyper | 69 | C0019348 | Herpes Simplex Infections | DisGeNET BeFree | 3.66E-05 | 9.06E-03 |
| Family A | Hyper | 70 | C0023418 | leukemia | DisGeNET BeFree | 3.75E-05 | 9.15E-03 |
| Family A | Hyper | 71 | C0740858 | Substance abuse problem | DisGeNET Curated | 4.02E-05 | 9.67E-03 |
| Family A | Hyper | 72 | C0030567 | Parkinson Disease | DisGeNET BeFree | 4.08E-05 | 9.67E-03 |
| Family A | Hyper | 73 | C1561643 | Chronic Kidney Diseases | DisGeNET BeFree | 4.79E-05 | 1.12E-02 |
| Family A | Hyper | 74 | C0004352 | Autistic Disorder | DisGeNET BeFree | 4.99E-05 | 1.15E-02 |
| Family A | Hyper | 75 | C0153676 | Secondary malignant neoplasm of lung | DisGeNET BeFree | 5.04E-05 | 1.15E-02 |
| Family A | Hyper | 76 | C0020429 | Hyperalgesia | DisGeNET BeFree | 5.15E-05 | 1.16E-02 |
| Family A | Hyper | 77 | C0020538 | Hypertensive disease | DisGeNET BeFree | 5.49E-05 | 1.22E-02 |
| Family A | Hyper | 78 | C0598766 | Leukemogenesis | DisGeNET BeFree | 5.68E-05 | 1.24E-02 |
| Family A | Hyper | 79 | C0017638 | Glioma | DisGeNET BeFree | 5.77E-05 | 1.25E-02 |
| Family A | Hyper | 80 | C0008073 | Developmental Disabilities | DisGeNET BeFree | 6.13E-05 | 1.30E-02 |
| Family A | Hyper | 81 | C0003873 | Rheumatoid Arthritis | DisGeNET BeFree | 6.14E-05 | 1.30E-02 |
| Family A | Hyper | 82 | C0011860 | Diabetes Mellitus, Non-Insulin-Dependent | DisGeNET BeFree | 6.35E-05 | 1.32E-02 |
| Family A | Hyper | 83 | C0036572 | Seizures | DisGeNET BeFree | 6.93E-05 | 1.42E-02 |
| Family A | Hyper | 84 | C0376634 | Craniofacial Abnormalities | DisGeNET Curated | 7.00E-05 | 1.42E-02 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family A | Hyper | 85 | C0019340 | Herpes NOS | DisGeNET BeFree | 7.09E-05 | 1.43E-02 |
| Family A | Hyper | 86 | C0020179 | Huntington Disease | DisGeNET BeFree | 8.73E-05 | 1.73E-02 |
| Family A | Hyper | 87 | C1332977 | Childhood Leukemia | DisGeNET BeFree | 8.99E-05 | 1.77E-02 |
| Family A | Hyper | 88 | C4722518 | Triple-Negative Breast Carcinoma | DisGeNET BeFree | 9.11E-05 | 1.77E-02 |
| Family A | Hyper | 89 | C0040822 | Tremor | DisGeNET BeFree | 9.26E-05 | 1.78E-02 |
| Family A | Hyper | 90 | C0004134 | Ataxia | DisGeNET BeFree | 1.10E-04 | 2.09E-02 |
| Family A | Hyper | 91 | C0233514 | Abnormal behavior | DisGeNET BeFree | 1.16E-04 | 2.16E-02 |
| Family A | Hyper | 92 | C0598589 | Inherited neuropathies | DisGeNET BeFree | 1.17E-04 | 2.16E-02 |
| Family A | Hyper | 93 | C0009241 | Cognition Disorders | DisGeNET BeFree | 1.19E-04 | 2.19E-02 |
| Family A | Hyper | 94 | C0037763 | Spasm | DisGeNET BeFree | 1.27E-04 | 2.28E-02 |
| Family A | Hyper | 95 | C0235974 | Pancreatic carcinoma | DisGeNET BeFree | 1.27E-04 | 2.28E-02 |
| Family A | Hyper | 96 | C0014175 | Endometriosis | DisGeNET BeFree | 1.30E-04 | 2.28E-02 |
| Family A | Hyper | 97 | C0278595 | Adult Fibrosarcoma | DisGeNET BeFree | 1.30E-04 | 2.28E-02 |
| Family A | Hyper | 98 | C0025149 | Medulloblastoma | DisGeNET BeFree | 1.38E-04 | 2.41E-02 |
| Family A | Hyper | 99 | C3266262 | Multiple Chronic Conditions | DisGeNET BeFree | 1.52E-04 | 2.62E-02 |
| Family A | Hyper | 100 | C0023467 | Leukemia, Myelocytic, Acute | DisGeNET BeFree | 1.53E-04 | 2.62E-02 |
| Family A | Hyper | 101 | C2677180 | Congenital microcephaly | DisGeNET BeFree | 1.59E-04 | 2.68E-02 |
| Family A | Hyper | 102 | C0858600 | Taste sweet | DisGeNET BeFree | 1.60E-04 | 2.68E-02 |
| Family A | Hyper | 103 | C0026650 | Movement Disorders | DisGeNET BeFree | 1.65E-04 | 2.69E-02 |
| Family A | Hyper | 104 | C0027868 | Neuromuscular Diseases | DisGeNET BeFree | 1.66E-04 | 2.69E-02 |
| Family A | Hyper | 105 | C0029408 | Degenerative polyarthritis | DisGeNET BeFree | 1.67E-04 | 2.69E-02 |
| Family A | Hyper | 106 | C1956346 | Coronary Artery Disease | DisGeNET BeFree | 1.67E-04 | 2.69E-02 |
| Family A | Hyper | 107 | C0023449 | Acute lymphocytic leukemia | DisGeNET BeFree | 1.69E-04 | 2.70E-02 |
| Family A | Hyper | 108 | C0346647 | Malignant neoplasm of pancreas | DisGeNET BeFree | 1.71E-04 | 2.70E-02 |
| Family A | Hyper | 109 | C0042063 | Urogenital Abnormalities | DisGeNET BeFree | 1.80E-04 | 2.82E-02 |
| Family A | Hyper | 110 | C0014544 | Epilepsy | DisGeNET Curated | 1.81E-04 | 2.82E-02 |
| Family A | Hyper | 111 | C0023434 | Chronic Lymphocytic Leukemia | DisGeNET BeFree | 1.84E-04 | 2.84E-02 |
| Family A | Hyper | 112 | C0016057 | Fibrosarcoma | DisGeNET BeFree | 1.95E-04 | 2.97E-02 |
| Family A | Hyper | 113 | C0085220 | Cerebral Amyloid Angiopathy | DisGeNET BeFree | 2.09E-04 | 3.16E-02 |
| Family A | Hyper | 114 | C0280222 | stage, pancreatic cancer | DisGeNET BeFree | 2.15E-04 | 3.17E-02 |
| Family A | Hyper | 115 | C1842937 | AURAL ATRESIA, CONGENITAL | DisGeNET BeFree | 2.15E-04 | 3.17E-02 |
| Family A | Hyper | 116 | C0004352 | Autistic Disorder | DisGeNET Curated | 2.15E-04 | 3.17E-02 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family A | Hyper | 117 | C0086438 | Hypogammaglobulinemia | DisGeNET BeFree | 2.22E-04 | 3.25E-02 |
| Family A | Hyper | 118 | C0027794 | Neural Tube Defects | DisGeNET BeFree | 2.27E-04 | 3.28E-02 |
| Family A | Hyper | 119 | C0162809 | Kallmann Syndrome | DisGeNET BeFree | 2.32E-04 | 3.30E-02 |
| Family A | Hyper | 120 | cv: | Progressive myoclonus epilepsy | Clinical Variations | 2.32E-04 | 3.30E-02 |
| Family A | Hyper | 121 | C0027819 | Neuroblastoma | DisGeNET Curated | 2.57E-04 | 3.63E-02 |
| Family A | Hyper | 122 | C0917981 | Progressive Muscular Atrophy | DisGeNET BeFree | 2.77E-04 | 3.88E-02 |
| Family A | Hyper | 123 | C0001973 | Alcoholic Intoxication, Chronic | DisGeNET BeFree | 2.90E-04 | 4.03E-02 |
| Family A | Hyper | 124 | C0021841 | Intestinal Neoplasms | DisGeNET BeFree | 3.00E-04 | 4.13E-02 |
| Family A | Hyper | 125 | C0021390 | Inflammatory Bowel Diseases | DisGeNET BeFree | 3.02E-04 | 4.13E-02 |
| Family A | Hyper | 126 | C0021141 | Inappropriate ADH Syndrome | DisGeNET BeFree | 3.31E-04 | 4.35E-02 |
| Family A | Hyper | 127 | C0017178 | Gastrointestinal Diseases | DisGeNET Curated | 3.31E-04 | 4.35E-02 |
| Family A | Hyper | 128 | C0559031 | Functional Gastrointestinal Disorders | DisGeNET Curated | 3.31E-04 | 4.35E-02 |
| Family A | Hyper | 129 | C1565321 | Cholera Infantum | DisGeNET Curated | 3.31E-04 | 4.35E-02 |
| Family A | Hyper | 130 | C0023440 | Acute Erythroblastic Leukemia | DisGeNET BeFree | 3.31E-04 | 4.35E-02 |
| Family A | Hyper | 131 | C0023467 | Leukemia, Myelocytic, Acute | DisGeNET Curated | 3.40E-04 | 4.44E-02 |
| Family A | Hypo | 1 | C0004352 | Autistic Disorder | DisGeNET BeFree | 6.27E-06 | 3.62E-02 |
| Family A | Hypo | 2 | C3854173 | Pre-renal acute kidney injury | DisGeNET BeFree | 1.30E-05 | 3.62E-02 |
| Family A | Hypo | 3 | C0011581 | Depressive disorder | DisGeNET BeFree | 1.31E-05 | 3.62E-02 |
| Family A | Hypo | 4 | C1510586 | Autism Spectrum Disorders | DisGeNET BeFree | 1.57E-05 | 3.62E-02 |
| Family A | Hypo | 5 | C0344315 | Depressed mood | DisGeNET BeFree | 1.72E-05 | 3.62E-02 |
| Family A | Hypo | 6 | C0011570 | Mental Depression | DisGeNET BeFree | 1.75E-05 | 3.62E-02 |
| Family C | Hyper | 1 | C0036341 | Schizophrenia | DisGeNET BeFree | 1.16E-11 | 1.89E-07 |
| Family C | Hyper | 2 | C1510586 | Autism Spectrum Disorders | DisGeNET BeFree | 3.05E-09 | 1.86E-05 |
| Family C | Hyper | 3 | C0000768 | Congenital Abnormality | DisGeNET BeFree | 3.42E-09 | 1.86E-05 |
| Family C | Hyper | 4 | C0011581 | Depressive disorder | DisGeNET BeFree | 1.39E-08 | 5.67E-05 |
| Family C | Hyper | 5 | C0557874 | Global developmental delay | DisGeNET BeFree | 2.01E-08 | 6.54E-05 |
| Family C | Hyper | 6 | C0036341 | Schizophrenia | DisGeNET Curated | 2.62E-08 | 7.12E-05 |
| Family C | Hyper | 7 | C0424295 | Hyperactive behavior | DisGeNET BeFree | 5.03E-08 | 1.14E-04 |
| Family C | Hyper | 8 | C0344315 | Depressed mood | DisGeNET BeFree | 5.59E-08 | 1.14E-04 |
| Family C | Hyper | 9 | C0011570 | Mental Depression | DisGeNET BeFree | 1.12E-07 | 1.83E-04 |
| Family C | Hyper | 10 | C0424605 | Developmental delay (disorder) | DisGeNET BeFree | 1.12E-07 | 1.83E-04 |
| Family C | Hyper | 11 | C0338656 | Impaired cognition | DisGeNET BeFree | 1.65E-07 | 2.45E-04 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family C | Hyper | 12 | C0004352 | Autistic Disorder | DisGeNET BeFree | 2.08E-07 | 2.80E-04 |
| Family C | Hyper | 13 | C0221357 | Brachydactyly | DisGeNET BeFree | 2.23E-07 | 2.80E-04 |
| Family C | Hyper | 14 | C0020456 | Hyperglycemia | DisGeNET BeFree | 2.64E-07 | 3.08E-04 |
| Family C | Hyper | 15 | C0376634 | Craniofacial Abnormalities | DisGeNET Curated | 4.63E-07 | 5.03E-04 |
| Family C | Hyper | 16 | C0028754 | Obesity | DisGeNET BeFree | 1.14E-06 | 1.16E-03 |
| Family C | Hyper | 17 | C0524528 | Pervasive Development Disorder | DisGeNET BeFree | 1.77E-06 | 1.70E-03 |
| Family C | Hyper | 18 | C0030193 | Pain | DisGeNET BeFree | 2.74E-06 | 2.47E-03 |
| Family C | Hyper | 19 | C1269683 | Major Depressive Disorder | DisGeNET BeFree | 2.88E-06 | 2.47E-03 |
| Family C | Hyper | 20 | C0030567 | Parkinson Disease | DisGeNET BeFree | 3.99E-06 | 3.25E-03 |
| Family C | Hyper | 21 | C0018798 | Congenital Heart Defects | DisGeNET BeFree | 4.93E-06 | 3.83E-03 |
| Family C | Hyper | 22 | C0036572 | Seizures | DisGeNET BeFree | 7.94E-06 | 5.84E-03 |
| Family C | Hyper | 23 | C0041696 | Unipolar Depression | DisGeNET BeFree | 8.24E-06 | 5.84E-03 |
| Family C | Hyper | 24 | C0003467 | Anxiety | DisGeNET BeFree | 9.23E-06 | 6.27E-03 |
| Family C | Hyper | 25 | C0302142 | Deformity | DisGeNET BeFree | 1.03E-05 | 6.72E-03 |
| Family C | Hyper | 26 | C0020676 | Hypothyroidism | DisGeNET BeFree | 1.08E-05 | 6.75E-03 |
| Family C | Hyper | 27 | C0011269 | Dementia, Vascular | DisGeNET BeFree | 1.12E-05 | 6.75E-03 |
| Family C | Hyper | 28 | C0033975 | Psychotic Disorders | DisGeNET BeFree | 1.16E-05 | 6.75E-03 |
| Family C | Hyper | 29 | C0025286 | Meningioma | DisGeNET BeFree | 1.23E-05 | 6.90E-03 |
| Family C | Hyper | 30 | C1535926 | Neurodevelopmental Disorders | DisGeNET BeFree | 1.47E-05 | 8.01E-03 |
| Family C | Hyper | 31 | C0349204 | Nonorganic psychosis | DisGeNET BeFree | 1.53E-05 | 8.03E-03 |
| Family C | Hyper | 32 | C0027765 | nervous system disorder | DisGeNET BeFree | 1.64E-05 | 8.37E-03 |
| Family C | Hyper | 33 | C0008073 | Developmental Disabilities | DisGeNET BeFree | 1.81E-05 | 8.96E-03 |
| Family C | Hyper | 34 | C0233514 | Abnormal behavior | DisGeNET BeFree | 2.42E-05 | 1.16E-02 |
| Family C | Hyper | 35 | C0003469 | Anxiety Disorders | DisGeNET BeFree | 2.48E-05 | 1.16E-02 |
| Family C | Hyper | 36 | C0001973 | Alcoholic Intoxication, Chronic | DisGeNET Curated | 2.67E-05 | 1.17E-02 |
| Family C | Hyper | 37 | C3714756 | Intellectual Disability | DisGeNET BeFree | 2.72E-05 | 1.17E-02 |
| Family C | Hyper | 38 | C0011860 | Diabetes Mellitus, Non-Insulin-Dependent | DisGeNET BeFree | 2.73E-05 | 1.17E-02 |
| Family C | Hyper | 39 | C1321551 | Shprintzen-Goldberg syndrome | DisGeNET BeFree | 2.82E-05 | 1.18E-02 |
| Family C | Hyper | 40 | C0236733 | Amphetamine-Related Disorders | DisGeNET Curated | 3.35E-05 | 1.30E-02 |
| Family C | Hyper | 41 | C0236807 | Amphetamine Abuse | DisGeNET Curated | 3.35E-05 | 1.30E-02 |
| Family C | Hyper | 42 | C0236804 | Amphetamine Addiction | DisGeNET Curated | 3.35E-05 | 1.30E-02 |
| Family C | Hyper | 43 | C0278878 | Adult Glioblastoma | DisGeNET BeFree | 3.66E-05 | 1.36E-02 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family C | Hyper | 44 | C0280474 | Childhood Glioblastoma | DisGeNET BeFree | 3.66E-05 | 1.36E-02 |
| Family C | Hyper | 45 | C0877015 | Pelvic Organ Prolapse | DisGeNET BeFree | 3.78E-05 | 1.37E-02 |
| Family C | Hyper | 46 | C3714796 | Isolated somatotropin deficiency | DisGeNET BeFree | 4.08E-05 | 1.45E-02 |
| Family C | Hyper | 47 | C0751265 | Learning Disabilities | DisGeNET BeFree | 4.23E-05 | 1.47E-02 |
| Family C | Hyper | 48 | C1535926 | Neurodevelopmental Disorders | DisGeNET Curated | 6.28E-05 | 2.13E-02 |
| Family C | Hyper | 49 | C0014544 | Epilepsy | DisGeNET Curated | 6.60E-05 | 2.15E-02 |
| Family C | Hyper | 50 | C0345967 | Malignant mesothelioma | DisGeNET Curated | 6.60E-05 | 2.15E-02 |
| Family C | Hyper | 51 | C0004936 | Mental disorders | DisGeNET BeFree | 6.96E-05 | 2.22E-02 |
| Family C | Hyper | 52 | C0026837 | Muscle Rigidity | DisGeNET BeFree | 7.47E-05 | 2.34E-02 |
| Family C | Hyper | 53 | C0007222 | Cardiovascular Diseases | DisGeNET BeFree | 7.74E-05 | 2.38E-02 |
| Family C | Hyper | 54 | C0027819 | Neuroblastoma | DisGeNET BeFree | 7.96E-05 | 2.40E-02 |
| Family C | Hyper | 55 | C1611743 | Familial (FPAH) | DisGeNET BeFree | 8.56E-05 | 2.51E-02 |
| Family C | Hyper | 56 | C0020429 | Hyperalgesia | DisGeNET BeFree | 8.64E-05 | 2.51E-02 |
| Family C | Hyper | 57 | C0600520 | Left Ventricle Remodeling | DisGeNET Curated | 9.81E-05 | 2.76E-02 |
| Family C | Hyper | 58 | C0600519 | Ventricular Remodeling | DisGeNET Curated | 9.81E-05 | 2.76E-02 |
| Family C | Hyper | 59 | C0424296 | Social disinhibition | DisGeNET BeFree | 1.03E-04 | 2.81E-02 |
| Family C | Hyper | 60 | C0005586 | Bipolar Disorder | DisGeNET Curated | 1.04E-04 | 2.81E-02 |
| Family C | Hyper | 61 | C0700095 | Central neuroblastoma | DisGeNET BeFree | 1.06E-04 | 2.84E-02 |
| Family C | Hyper | 62 | C4086165 | Childhood Neuroblastoma | DisGeNET BeFree | 1.10E-04 | 2.89E-02 |
| Family C | Hyper | 63 | C0009241 | Cognition Disorders | DisGeNET BeFree | 1.19E-04 | 3.08E-02 |
| Family C | Hyper | 64 | C0014544 | Epilepsy | DisGeNET BeFree | 1.23E-04 | 3.12E-02 |
| Family C | Hyper | 65 | C0524620 | Metabolic Syndrome X | DisGeNET BeFree | 1.28E-04 | 3.13E-02 |
| Family C | Hyper | 66 | C0020538 | Hypertensive disease | DisGeNET BeFree | 1.28E-04 | 3.13E-02 |
| Family C | Hyper | 67 | C0221271 | Elastosis perforans serpiginosa | DisGeNET BeFree | 1.29E-04 | 3.13E-02 |
| Family C | Hyper | 68 | C1565489 | Renal Insufficiency | DisGeNET BeFree | 1.46E-04 | 3.49E-02 |
| Family C | Hyper | 69 | C0233794 | Memory impairment | DisGeNET BeFree | 1.54E-04 | 3.60E-02 |
| Family C | Hyper | 70 | C0027051 | Myocardial Infarction | DisGeNET BeFree | 1.55E-04 | 3.60E-02 |
| Family C | Hyper | 71 | C0001973 | Alcoholic Intoxication, Chronic | DisGeNET BeFree | 1.76E-04 | 4.03E-02 |
| Family C | Hyper | 72 | C0006012 | Borderline Personality Disorder | DisGeNET BeFree | 2.04E-04 | 4.61E-02 |
| Family C | Hyper | 73 | C0026650 | Movement Disorders | DisGeNET BeFree | 2.17E-04 | 4.85E-02 |
| Family C | Hypo | 1 | C2711227 | Steatohepatitis | DisGeNET BeFree | 3.96E-06 | 2.84E-02 |
| Family C | Hypo | 2 | C0334583 | Pilocytic Astrocytoma | DisGeNET BeFree | 4.68E-06 | 2.84E-02 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family C | Hypo | 3 | C0027765 | nervous system disorder | DisGeNET BeFree | 8.90E-06 | 2.84E-02 |
| Family C | Hypo | 4 | C0023448 | Lymphoid leukemia | DisGeNET BeFree | 1.82E-05 | 2.84E-02 |
| Family C | Hypo | 5 | C0149931 | Migraine Disorders | DisGeNET BeFree | 1.98E-05 | 2.84E-02 |
| Family C | Hypo | 6 | C0013384 | Dyskinetic syndrome | DisGeNET BeFree | 2.03E-05 | 2.84E-02 |
| Family C | Hypo | 7 | C1332977 | Childhood Leukemia | DisGeNET BeFree | 2.15E-05 | 2.84E-02 |
| Family C | Hypo | 8 | C0740858 | Substance abuse problem | DisGeNET BeFree | 2.30E-05 | 2.84E-02 |
| Family C | Hypo | 9 | C0023418 | leukemia | DisGeNET BeFree | 3.37E-05 | 2.84E-02 |
| Family C | Hypo | 10 | C0271650 | Impaired glucose tolerance | DisGeNET BeFree | 4.57E-05 | 2.84E-02 |
| Family C | Hypo | 11 | C0338656 | Impaired cognition | DisGeNET BeFree | 4.91E-05 | 2.84E-02 |
| Family C | Hypo | 12 | C2267227 | Bulimia Nervosa | DisGeNET BeFree | 5.14E-05 | 2.84E-02 |
| Family C | Hypo | 13 | C3642347 | Basal-Like Breast Carcinoma | DisGeNET BeFree | 5.59E-05 | 2.84E-02 |
| Family C | Hypo | 14 | C0270824 | Visual seizure | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 15 | C0270846 | Epileptic drop attack | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 16 | C0234533 | Generalized seizures | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 17 | C0234535 | Clonic Seizures | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 18 | C0751056 | Non-epileptic convulsion | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 19 | C0751123 | Atonic Absence Seizures | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 20 | C0751110 | Single Seizure | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 21 | C0751494 | Convulsive Seizures | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 22 | C0751496 | Seizures, Sensory | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 23 | C0149958 | Complex partial seizures | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 24 | C3495874 | Nonepileptic Seizures | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 25 | C4505436 | Generalized Absence Seizures | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 26 | C0422855 | Vertiginous seizure | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 27 | C0422854 | Gustatory seizure | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 28 | C0422850 | Seizures, Somatosensory | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 29 | C0422853 | Olfactory seizure | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 30 | C0422852 | Seizures, Auditory | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 31 | C0022333 | Jacksonian Seizure | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 32 | C4317109 | Epileptic Seizures | DisGeNET Curated | 7.13E-05 | 2.84E-02 |
| Family C | Hypo | 33 | C0018801 | Heart failure | DisGeNET BeFree | 7.39E-05 | 2.85E-02 |
| Family C | Hypo | 34 | C4316903 | Absence Seizures | DisGeNET Curated | 8.24E-05 | 2.92E-02 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family C | Hypo | 35 | C4048158 | Convulsions | DisGeNET Curated | 8.24E-05 | 2.92E-02 |
| Family C | Hypo | 36 | C0270844 | Tonic Seizures | DisGeNET Curated | 8.24E-05 | 2.92E-02 |
| Family C | Hypo | 37 | C0085207 | Gestational Diabetes | DisGeNET BeFree | 8.62E-05 | 2.97E-02 |
| Family C | Hypo | 38 | C0015695 | Fatty Liver | DisGeNET BeFree | 9.11E-05 | 3.06E-02 |
| Family C | Hypo | 39 | C0279583 | Childhood T Acute Lymphoblastic Leukemia | DisGeNET BeFree | 9.37E-05 | 3.06E-02 |
| Family C | Hypo | 40 | C0005586 | Bipolar Disorder | DisGeNET Curated | 1.01E-04 | 3.16E-02 |
| Family C | Hypo | 41 | C0279565 | Invasive Lobular Breast Carcinoma | DisGeNET BeFree | 1.04E-04 | 3.16E-02 |
| Family C | Hypo | 42 | C0751495 | Seizures, Focal | DisGeNET Curated | 1.09E-04 | 3.16E-02 |
| Family C | Hypo | 43 | C0494475 | Tonic - clonic seizures | DisGeNET Curated | 1.09E-04 | 3.16E-02 |
| Family C | Hypo | 44 | C4317123 | Myoclonic Seizures | DisGeNET Curated | 1.09E-04 | 3.16E-02 |
| Family C | Hypo | 45 | C0265509 | Congenital anomaly of skeletal bone | DisGeNET BeFree | 1.21E-04 | 3.20E-02 |
| Family C | Hypo | 46 | C0853892 | Catabolic state | DisGeNET BeFree | 1.21E-04 | 3.20E-02 |
| Family C | Hypo | 47 | C0677886 | Epithelial ovarian cancer | DisGeNET BeFree | 1.22E-04 | 3.20E-02 |
| Family C | Hypo | 48 | C0038443 | Stress, Psychological | DisGeNET BeFree | 1.22E-04 | 3.20E-02 |
| Family C | Hypo | 49 | C0206658 | Smooth Muscle Tumor | DisGeNET BeFree | 1.23E-04 | 3.20E-02 |
| Family C | Hypo | 50 | C4288891 | Infant T Acute Lymphoblastic Leukemia | DisGeNET BeFree | 1.48E-04 | 3.78E-02 |
| Family C | Hypo | 51 | C0005699 | Blast Phase | DisGeNET BeFree | 1.53E-04 | 3.81E-02 |
| Family C | Hypo | 52 | C0019569 | Hirschsprung Disease | DisGeNET BeFree | 1.66E-04 | 4.00E-02 |
| Family C | Hypo | 53 | C0007134 | Renal Cell Carcinoma | DisGeNET BeFree | 1.67E-04 | 4.00E-02 |
| Family C | Hypo | 54 | C0011581 | Depressive disorder | DisGeNET BeFree | 1.84E-04 | 4.35E-02 |
| Family C | Hypo | 55 | C0030567 | Parkinson Disease | DisGeNET BeFree | 1.95E-04 | 4.51E-02 |
| Family C | Hypo | 56 | C0024301 | Lymphoma, Follicular | DisGeNET BeFree | 2.09E-04 | 4.71E-02 |
| Family C | Hypo | 57 | C0035344 | Retinopathy of Prematurity | DisGeNET BeFree | 2.14E-04 | 4.71E-02 |
| Family C | Hypo | 58 | C2062441 | Influenza A | DisGeNET BeFree | 2.14E-04 | 4.71E-02 |
| Family C | Hypo | 59 | C0001973 | Alcoholic Intoxication, Chronic | DisGeNET BeFree | 2.31E-04 | 4.92E-02 |
| Family C | Hypo | 60 | C0018802 | Congestive heart failure | DisGeNET BeFree | 2.32E-04 | 4.92E-02 |

(*) IDs are unique to the associated database. (†) P-values were calculated using the hypergeometric test. (‡) FDR B&H: False discovery rates were calculated by the Benjamini and Hochberg method[127].

## Appendix 2.8

## OR statistics for iPSC DMR and histone modifications

| DMR Group | DMR Type | Histone Mark | a* | b* | c* | d* | OR† | log(OR) | p-value‡ |
|---|---|---|---|---|---|---|---|---|---|
| Shared | Hyper | H3K27Ac | 326 | 1015793 | 1126 | 3408263 | 9.71E-01 | -1.26E-02 | 6.62E-01 |
| Shared | Hyper | H3K27me3 | 532 | 1238261 | 920 | 3185589 | 1.49E+00 | 1.73E-01 | 1.02E-12 |
| Shared | Hyper | H3K36me3 | 366 | 1355818 | 1086 | 3068198 | 7.63E-01 | -1.18E-01 | 5.12E-06 |
| Shared | Hyper | H3K4me1 | 445 | 899637 | 1007 | 3524300 | 1.73E+00 | 2.38E-01 | 2.08E-20 |
| Shared | Hyper | H3K4me3 | 328 | 951104 | 1124 | 3472950 | 1.07E+00 | 2.76E-02 | 3.07E-01 |
| Shared | Hyper | H3K9me3 | 308 | 747719 | 1144 | 3676355 | 1.32E+00 | 1.22E-01 | 2.19E-05 |
| Shared | Hypo | H3K27Ac | 116 | 1016003 | 890 | 3408709 | 4.37E-01 | -3.59E-01 | 1.95E-20 |
| Shared | Hypo | H3K27me3 | 361 | 1238432 | 645 | 3186035 | 1.44E+00 | 1.58E-01 | 5.85E-08 |
| Shared | Hypo | H3K36me3 | 235 | 1355949 | 771 | 3068644 | 6.90E-01 | -1.61E-01 | 3.27E-07 |
| Shared | Hypo | H3K4me1 | 169 | 899913 | 837 | 3524746 | 7.91E-01 | -1.02E-01 | 4.78E-03 |
| Shared | Hypo | H3K4me3 | 160 | 951272 | 846 | 3473396 | 6.91E-01 | -1.61E-01 | 9.73E-06 |
| Shared | Hypo | H3K9me3 | 249 | 747778 | 757 | 3676801 | 1.62E+00 | 2.09E-01 | 2.96E-10 |
| Family A | Hyper | H3K27Ac | 1946 | 912786 | 9406 | 3009559 | 6.82E-01 | -1.66E-01 | 2.13E-57 |
| Family A | Hyper | H3K27me3 | 3613 | 1099621 | 7739 | 2821057 | 1.20E+00 | 7.84E-02 | 1.01E-18 |
| Family A | Hyper | H3K36me3 | 2645 | 1211411 | 8707 | 2710235 | 6.80E-01 | -1.68E-01 | 6.64E-72 |
| Family A | Hyper | H3K4me1 | 2502 | 801355 | 8850 | 3120434 | 1.10E+00 | 4.17E-02 | 2.71E-05 |
| Family A | Hyper | H3K4me3 | 1931 | 860760 | 9421 | 3061600 | 7.29E-01 | -1.37E-01 | 5.58E-39 |
| Family A | Hyper | H3K9me3 | 2626 | 661875 | 8726 | 3259790 | 1.48E+00 | 1.71E-01 | 6.65E-65 |
| Family A | Hypo | H3K27Ac | 1540 | 913192 | 5963 | 3013408 | 8.52E-01 | -6.94E-02 | 1.47E-08 |
| Family A | Hypo | H3K27me3 | 2781 | 1100453 | 4722 | 2824906 | 1.51E+00 | 1.80E-01 | 3.08E-64 |
| Family A | Hypo | H3K36me3 | 1871 | 1212185 | 5632 | 2714084 | 7.44E-01 | -1.29E-01 | 8.57E-30 |
| Family A | Hypo | H3K4me1 | 1896 | 801961 | 5607 | 3124283 | 1.32E+00 | 1.20E-01 | 4.61E-24 |
| Family A | Hypo | H3K4me3 | 1654 | 861037 | 5849 | 3065449 | 1.01E+00 | 2.93E-03 | 8.12E-01 |
| Family A | Hypo | H3K9me3 | 1475 | 663026 | 6028 | 3263639 | 1.20E+00 | 8.08E-02 | 3.37E-10 |
| Family C | Hyper | H3K27Ac | 1405 | 805630 | 7602 | 2623702 | 6.02E-01 | -2.20E-01 | 4.48E-76 |
| Family C | Hyper | H3K27me3 | 3151 | 963853 | 5856 | 2463733 | 1.38E+00 | 1.38E-01 | 1.75E-45 |
| Family C | Hyper | H3K36me3 | 2013 | 1058107 | 6994 | 2370617 | 6.45E-01 | -1.91E-01 | 2.21E-72 |
| Family C | Hyper | H3K4me1 | 1701 | 707970 | 7306 | 2721066 | 8.95E-01 | -4.83E-02 | 3.20E-05 |
| Family C | Hyper | H3K4me3 | 1362 | 763549 | 7645 | 2665826 | 6.22E-01 | -2.06E-01 | 7.55E-65 |
| Family C | Hyper | H3K9me3 | 1696 | 574902 | 7311 | 2854139 | 1.15E+00 | 6.13E-02 | 2.53E-07 |
| Family C | Hypo | H3K27Ac | 1231 | 805804 | 5425 | 2626053 | 7.39E-01 | -1.31E-01 | 8.45E-23 |
| Family C | Hypo | H3K27me3 | 2421 | 964583 | 4235 | 2466084 | 1.46E+00 | 1.65E-01 | 3.34E-48 |
| Family C | Hypo | H3K36me3 | 1586 | 1058534 | 5070 | 2372968 | 7.01E-01 | -1.54E-01 | 8.19E-37 |
| Family C | Hypo | H3K4me1 | 1391 | 708280 | 5265 | 2723417 | 1.02E+00 | 6.84E-03 | 6.06E-01 |
| Family C | Hypo | H3K4me3 | 1352 | 763559 | 5304 | 2668177 | 8.91E-01 | -5.03E-02 | 1.33E-04 |
| Family C | Hypo | H3K9me3 | 1249 | 575349 | 5407 | 2856490 | 1.15E+00 | 5.95E-02 | 1.68E-05 |

(*) *a*, *b*, *c*, and *d* values are the contingency parameters used to calculate OR. (†) OR was calculated as described in the Methods section. (‡) P-values were calculated by Fisher's exact test.

## Appendix 2.9

## Complete list of disease ontology terms from ToppGene for gene lists associated with fibroblast and iPSC DMRs

| DMR group | DMR type (Fibroblast to iPSC) | Rank | ID* | Name | Source | p-value† | FDR B&H‡ |
|---|---|---|---|---|---|---|---|
| Family C | Hyper to Hypo | 1 | C0014544 | Epilepsy | DisGeNET BeFree | 2.22E-07 | 9.66E-04 |
| Family C | Hyper to Hypo | 2 | C0424605 | Developmental delay (disorder) | DisGeNET BeFree | 2.75E-07 | 9.66E-04 |
| Family C | Hyper to Hypo | 3 | C0000768 | Congenital Abnormality | DisGeNET BeFree | 4.71E-07 | 1.10E-03 |
| Family C | Hyper to Hypo | 4 | C0008073 | Developmental Disabilities | DisGeNET BeFree | 1.23E-06 | 2.17E-03 |
| Family C | Hyper to Hypo | 5 | C0557874 | Global developmental delay | DisGeNET BeFree | 3.22E-06 | 4.53E-03 |
| Family C | Hyper to Hypo | 6 | C0221357 | Brachydactyly | DisGeNET BeFree | 4.3E-06 | 5.04E-03 |
| Family C | Hyper to Hypo | 7 | C0023418 | leukemia | DisGeNET BeFree | 7.58E-06 | 7.61E-03 |
| Family C | Hyper to Hypo | 8 | C0598766 | Leukemogenesis | DisGeNET BeFree | 2.29E-05 | 1.89E-02 |
| Family C | Hyper to Hypo | 9 | C3714756 | Intellectual Disability | DisGeNET BeFree | 2.47E-05 | 1.89E-02 |
| Family C | Hyper to Hypo | 10 | C4021790 | Abnormality of the skeletal system | DisGeNET BeFree | 2.69E-05 | 1.89E-02 |
| Family C | Hyper to Hypo | 11 | C0524528 | Pervasive Development Disorder | DisGeNET BeFree | 3.7E-05 | 2.25E-02 |
| Family C | Hyper to Hypo | 12 | C0036572 | Seizures | DisGeNET BeFree | 4.13E-05 | 2.25E-02 |
| Family C | Hyper to Hypo | 13 | C1332977 | Childhood Leukemia | DisGeNET BeFree | 4.15E-05 | 2.25E-02 |
| Family C | Hyper to Hypo | 14 | C0221356 | Brachycephaly | DisGeNET BeFree | 5.11E-05 | 2.25E-02 |
| Family C | Hyper to Hypo | 15 | C0410179 | Ullrich congenital muscular dystrophy 1 | DisGeNET Curated | 5.11E-05 | 2.25E-02 |
| Family C | Hyper to Hypo | 16 | C1834674 | BETHLEM MYOPATHY 1 | DisGeNET Curated | 5.11E-05 | 2.25E-02 |
| Family C | Hyper to Hypo | 17 | C0240340 | Microdontia (disorder) | DisGeNET BeFree | 8.81E-05 | 3.65E-02 |
| Family C | Hyper to Hypo | 18 | C0079218 | Fibromatosis, Aggressive | DisGeNET BeFree | 9.9E-05 | 3.87E-02 |
| Family C | Hyper to Hypo | 19 | C0008029 | Cherubism | DisGeNET BeFree | 0.000106 | 3.92E-02 |
| Family C | Hyper to Hypo | 20 | C0025958 | Microcephaly | DisGeNET BeFree | 0.000129 | 4.53E-02 |
| Family C | Hyper to Hypo | 21 | C0265354 | CHARGE Syndrome | DisGeNET BeFree | 0.000141 | 4.72E-02 |
| Family C | Hyper to Hyper | 1 | C0000768 | Congenital Abnormality | DisGeNET BeFree | 8.69E-09 | 5.12E-05 |
| Family C | Hyper to Hyper | 2 | C0000846 | Agenesis | DisGeNET BeFree | 2.94E-06 | 8.68E-03 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family C | Hypo to Hyper | 1 | C0000768 | Congenital Abnormality | DisGeNET BeFree | 1.02E-05 | 3.06E-02 |
| Family C | Hypo to Hyper | 2 | C0424605 | Developmental delay (disorder) | DisGeNET BeFree | 1.28E-05 | 3.06E-02 |
| Family C | Hypo to Hyper | 3 | C0013080 | Down Syndrome | DisGeNET BeFree | 2.67E-05 | 4.26E-02 |
| Family C | Hypo to Hyper | 4 | C1449563 | Cardiomyopathy, Familial Idiopathic | DisGeNET BeFree | 4.51E-05 | 4.77E-02 |
| Family C | Hypo to Hyper | 5 | C0003873 | Rheumatoid Arthritis | DisGeNET Curated | 5.67E-05 | 4.77E-02 |
| Family C | Hypo to Hyper | 6 | C0557874 | Global developmental delay | DisGeNET BeFree | 5.98E-05 | 4.77E-02 |
| Family C | Non-DMR to Hyper | none | none | none | none | none | none |
| Family C | Hypo to Hypo | 1 | C0014544 | Epilepsy | DisGeNET BeFree | 3.66E-07 | 2.71E-03 |
| Family C | Hypo to Hypo | 2 | C0008925 | Cleft Palate | DisGeNET BeFree | 1.69E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 3 | C1535926 | Neurodevelopmental Disorders | DisGeNET BeFree | 2.64E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 4 | C0000768 | Congenital Abnormality | DisGeNET BeFree | 2.97E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 5 | C0270824 | Visual seizure | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 6 | C0270846 | Epileptic drop attack | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 7 | C0234533 | Generalized seizures | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 8 | C0234535 | Clonic Seizures | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 9 | C0751056 | Non-epileptic convulsion | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 10 | C0751123 | Atonic Absence Seizures | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 11 | C0751110 | Single Seizure | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 12 | C0422855 | Vertiginous seizure | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 13 | C0422854 | Gustatory seizure | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 14 | C0422850 | Seizures, Somatosensory | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 15 | C0422853 | Olfactory seizure | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 16 | C0422852 | Seizures, Auditory | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 17 | C0751494 | Convulsive Seizures | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 18 | C0751496 | Seizures, Sensory | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 19 | C0149958 | Complex partial seizures | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 20 | C3495874 | Nonepileptic Seizures | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 21 | C4505436 | Generalized Absence Seizures | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 22 | C0022333 | Jacksonian Seizure | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 23 | C4317109 | Epileptic Seizures | DisGeNET Curated | 5.9E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 24 | C4707243 | Familial thoracic aortic aneurysm and aortic dissection | DisGeNET Curated | 5.91E-05 | 1.68E-02 |
| Family C | Hypo to Hypo | 25 | C4316903 | Absence Seizures | DisGeNET Curated | 6.46E-05 | 1.68E-02 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Family C | Non-DMR to Hypo | none | none | none | none | none | none |
| Family A | Hyper to Hypo | 1 | C0270764 | Motor Neuron Disease, Lower | DisGeNET BeFree | 2.29E-06 | 4.70E-03 |
| Family A | Hyper to Hypo | 2 | C0524730 | Odontome | DisGeNET Curated | 3.07E-06 | 4.70E-03 |
| Family A | Hyper to Hypo | 3 | C0040427 | Tooth Abnormalities | DisGeNET Curated | 3.07E-06 | 4.70E-03 |
| Family A | Hyper to Hypo | 4 | C0206762 | Limb Deformities, Congenital | DisGeNET BeFree | 1.05E-05 | 1.21E-02 |
| Family A | Hyper to Hypo | 5 | C1839839 | MAJOR AFFECTIVE DISORDER 2 | DisGeNET Curated | 3.27E-05 | 3.00E-02 |
| Family A | Hyper to Hypo | 6 | C0850639 | premalignant lesion | DisGeNET BeFree | 6.48E-05 | 4.96E-02 |
| Family A | Hyper to Hyper | none | none | none | none | none | none |
| Family A | Hypo to Hyper | none | none | none | none | none | none |
| Family A | Non-DMR to Hyper | 1 | C0917796 | Optic Atrophy, Hereditary, Leber | DisGeNET Curated | 5.01E-06 | 2.36E-02 |
| Family A | Non-DMR to Hyper | 2 | 535000 | LEBER OPTIC ATROPHY | OMIM MedGen | 1.26E-05 | 2.36E-02 |
| Family A | Non-DMR to Hyper | 3 | cv:C09177 96 | Leber's optic atrophy | Clinical Variations | 1.26E-05 | 2.36E-02 |
| Family A | Hypo to Hypo | 1 | C0266544 | Microcornea | DisGeNET BeFree | 4.34E-06 | 1.76E-02 |
| Family A | Hypo to Hypo | 2 | C1839839 | MAJOR AFFECTIVE DISORDER 2 | DisGeNET Curated | 1.4E-05 | 2.85E-02 |
| Family A | Hypo to Hypo | 3 | C0007124 | Noninfiltrating Intraductal Carcinoma | DisGeNET BeFree | 2.53E-05 | 3.43E-02 |
| Family A | Non-DMR to Hypo | none | none | none | none | none | none |

(*) IDs are unique to the associated database. (†) P-values were calculated using the hypergeometric test. (‡) FDR B&H: False discovery rates were calculated by the Benjamini and Hochberg method[127].

## Appendix 2.10

## KEGG pathway enrichment for the set of 28 genes associated to DMRs whose methylation change is hypermethylated in fibroblast and hypomethylated in iPSC, acquired from STRING

| #term ID | term description | Strength* | FDR† | matching proteins in the network (labels)‡ |
|---|---|---|---|---|
| hsa04340 | Hedgehog signaling pathway | 1.66 | 4.10E-04 | CCND1,SHH,PTCH1 |
| hsa04933 | AGE-RAGE signaling pathway in diabetic complications | 1.63 | 1.01E-06 | CCND1,RAC1,CDC42,IL6,NFATC1,PRKCA |
| hsa05143 | African trypanosomiasis | 1.61 | 4.90E-03 | IL6,PRKCA |
| hsa05130 | Pathogenic Escherichia coli infection | 1.6 | 5.50E-04 | ROCK1,CDC42,PRKCA |
| hsa04370 | VEGF signaling pathway | 1.55 | 7.10E-04 | RAC1,CDC42,PRKCA |
| hsa05132 | Salmonella infection | 1.52 | 1.10E-04 | RAC1,ROCK1,CDC42,IL6 |
| hsa05131 | Shigellosis | 1.52 | 7.80E-04 | RAC1,ROCK1,CDC42 |
| hsa04666 | Fc gamma R-mediated phagocytosis | 1.5 | 1.20E-04 | RAC1,SYK,CDC42,PRKCA |
| hsa04664 | Fc epsilon RI signaling pathway | 1.5 | 8.50E-04 | RAC1,SYK,PRKCA |
| hsa05211 | Renal cell carcinoma | 1.49 | 8.50E-04 | CREBBP,RAC1,CDC42 |
| hsa04520 | Adherens junction | 1.47 | 9.30E-04 | CREBBP,RAC1,CDC42 |
| hsa04662 | B cell receptor signaling pathway | 1.47 | 9.30E-04 | RAC1,SYK,NFATC1 |
| hsa05212 | Pancreatic cancer | 1.45 | 9.70E-04 | CCND1,RAC1,CDC42 |
| hsa05205 | Proteoglycans in cancer | 1.4 | 1.01E-06 | CCND1,SHH,PTCH1,RAC1, ROCK1,CDC42,PRKCA |
| hsa04670 | Leukocyte transendothelial migration | 1.4 | 2.30E-04 | RAC1,ROCK1,CDC42,PRKCA |
| hsa05416 | Viral myocarditis | 1.4 | 1.02E-02 | CCND1,RAC1 |
| hsa04310 | Wnt signaling pathway | 1.39 | 4.12E-05 | CCND1,CREBBP,RAC1,NFATC1,PRKCA |
| hsa05161 | Hepatitis B | 1.39 | 4.12E-05 | CCND1,CREBBP,IL6,NFATC1,PRKCA |
| hsa04360 | Axon guidance | 1.38 | 6.48E-06 | SHH,PTCH1,RAC1,ROCK1,CDC42,PRKCA |
| hsa05206 | MicroRNAs in cancer | 1.37 | 4.12E-05 | CCND1,CREBBP,HDAC4,ROCK1,PRKCA |
| hsa00310 | Lysine degradation | 1.37 | 1.06E-02 | KMT2C,EHMT1 |
| hsa05167 | Kaposi's sarcoma-associated herpesvirus infection | 1.36 | 7.14E-06 | CCND1,CREBBP,RAC1,SYK,IL6,NFATC1 |
| hsa05203 | Viral carcinogenesis | 1.36 | 7.14E-06 | CCND1,CREBBP,HDAC4,RAC1,SYK,CDC42 |
| hsa04650 | Natural killer cell mediated cytotoxicity | 1.35 | 3.00E-04 | RAC1,SYK,NFATC1,PRKCA |
| hsa05321 | Inflammatory bowel disease (IBD) | 1.35 | 1.14E-02 | IL6,NFATC1 |
| hsa05217 | Basal cell carcinoma | 1.35 | 1.15E-02 | SHH,PTCH1 |
| hsa04720 | Long-term potentiation | 1.34 | 1.17E-02 | CREBBP,PRKCA |
| hsa04066 | HIF-1 signaling pathway | 1.33 | 2.10E-03 | CREBBP,IL6,PRKCA |
| hsa05120 | Epithelial cell signaling in Helicobacter pylori infection | 1.33 | 1.21E-02 | RAC1,CDC42 |
| hsa05223 | Non-small cell lung cancer | 1.33 | 1.21E-02 | CCND1,PRKCA |
| hsa05214 | Glioma | 1.31 | 1.22E-02 | CCND1,PRKCA |
| hsa05100 | Bacterial invasion of epithelial cells | 1.29 | 1.33E-02 | RAC1,CDC42 |
| hsa04921 | Oxytocin signaling pathway | 1.27 | 5.30E-04 | CCND1,ROCK1,NFATC1,PRKCA |
| hsa04071 | Sphingolipid signaling pathway | 1.26 | 3.20E-03 | RAC1,ROCK1,PRKCA |

| | | | | |
|---|---|---|---|---|
| hsa04919 | Thyroid hormone signaling pathway | 1.26 | 3.20E-03 | CCND1,CREBBP,PRKCA |
| hsa04024 | cAMP signaling pathway | 1.25 | 1.20E-04 | CREBBP,PTCH1,RAC1,ROCK1,NFATC1 |
| hsa04510 | Focal adhesion | 1.25 | 1.20E-04 | CCND1,RAC1,ROCK1,CDC42,PRKCA |
| hsa01521 | EGFR tyrosine kinase inhibitor resistance | 1.25 | 1.52E-02 | IL6,PRKCA |
| hsa04110 | Cell cycle | 1.23 | 3.60E-03 | CCND1,CREBBP,CDKN1C |
| hsa04380 | Osteoclast differentiation | 1.23 | 3.60E-03 | RAC1,SYK,NFATC1 |
| hsa04350 | TGF-beta signaling pathway | 1.23 | 1.68E-02 | CREBBP,ROCK1 |
| hsa04530 | Tight junction | 1.22 | 7.10E-04 | CCND1,RAC1,ROCK1,CDC42 |
| hsa05210 | Colorectal cancer | 1.22 | 1.73E-02 | CCND1,RAC1 |
| hsa04068 | FoxO signaling pathway | 1.21 | 3.80E-03 | CCND1,CREBBP,IL6 |
| hsa04912 | GnRH signaling pathway | 1.2 | 1.82E-02 | CDC42,PRKCA |
| hsa04972 | Pancreatic secretion | 1.17 | 2.03E-02 | RAC1,PRKCA |
| hsa05146 | Amoebiasis | 1.17 | 2.03E-02 | IL6,PRKCA |
| hsa05215 | Prostate cancer | 1.16 | 2.08E-02 | CCND1,CREBBP |
| hsa05166 | HTLV-I infection | 1.15 | 2.80E-04 | CCND1,CREBBP,ATF3,IL6,NFATC1 |
| hsa04932 | Non-alcoholic fatty liver disease (NAFLD) | 1.15 | 5.20E-03 | RAC1,CDC42,IL6 |
| hsa04660 | T cell receptor signaling pathway | 1.15 | 2.09E-02 | CDC42,NFATC1 |
| hsa04916 | Melanogenesis | 1.15 | 2.09E-02 | CREBBP,PRKCA |
| hsa05231 | Choline metabolism in cancer | 1.15 | 2.09E-02 | RAC1,PRKCA |
| hsa04620 | Toll-like receptor signaling pathway | 1.14 | 2.15E-02 | RAC1,IL6 |
| hsa04659 | Th17 cell differentiation | 1.14 | 2.15E-02 | IL6,NFATC1 |
| hsa04218 | Cellular senescence | 1.13 | 5.80E-03 | CCND1,IL6,NFATC1 |
| hsa04630 | Jak-STAT signaling pathway | 1.12 | 6.00E-03 | CCND1,CREBBP,IL6 |
| hsa05225 | Hepatocellular carcinoma | 1.11 | 6.20E-03 | CCND1,ARID1B,PRKCA |
| hsa05164 | Influenza A | 1.1 | 6.60E-03 | CREBBP,IL6,PRKCA |
| hsa05200 | Pathways in cancer | 1.09 | 1.45E-06 | CCND1,CREBBP,SHH,PTCH1,RAC1, ROCK1,CDC42,IL6,PRKCA |
| hsa05152 | Tuberculosis | 1.09 | 6.90E-03 | CREBBP,SYK,IL6 |
| hsa04722 | Neurotrophin signaling pathway | 1.08 | 2.66E-02 | RAC1,CDC42 |
| hsa04270 | Vascular smooth muscle contraction | 1.07 | 2.76E-02 | ROCK1,PRKCA |
| hsa04062 | Chemokine signaling pathway | 1.06 | 7.80E-03 | RAC1,ROCK1,CDC42 |
| hsa04611 | Platelet activation | 1.06 | 2.89E-02 | SYK,ROCK1 |
| hsa05169 | Epstein-Barr virus infection | 1.03 | 9.20E-03 | CREBBP,HDAC4,SYK |
| hsa04210 | Apoptosis | 1.02 | 3.30E-02 | LMNB1,SPTAN1 |
| hsa04371 | Apelin signaling pathway | 1.02 | 3.30E-02 | CCND1,HDAC4 |
| hsa05162 | Measles | 1.02 | 3.30E-02 | CCND1,IL6 |
| hsa04015 | Rap1 signaling pathway | 1.01 | 1.02E-02 | RAC1,CDC42,PRKCA |
| hsa04810 | Regulation of actin cytoskeleton | 1.01 | 1.02E-02 | RAC1,ROCK1,CDC42 |
| hsa04151 | PI3K-Akt signaling pathway | 1 | 8.50E-04 | CCND1,RAC1,SYK,IL6,PRKCA |
| hsa04010 | MAPK signaling pathway | 0.98 | 3.60E-03 | RAC1,CDC42,NFATC1,PRKCA |
| hsa04072 | Phospholipase D signaling pathway | 0.98 | 3.72E-02 | SYK,PRKCA |
| hsa05226 | Gastric cancer | 0.98 | 3.77E-02 | CCND1,SHH |
| hsa04014 | Ras signaling pathway | 0.96 | 1.21E-02 | RAC1,CDC42,PRKCA |
| hsa05165 | Human papillomavirus infection | 0.95 | 4.50E-03 | CCND1,CREBBP,HDAC4,CDC42 |
| hsa04022 | cGMP-PKG signaling pathway | 0.94 | 4.35E-02 | ROCK1,NFATC1 |

| | | | | |
|---|---|---|---|---|
| hsa05202 | Transcriptional misregulation in cancer | 0.92 | 4.75E-02 | MEIS1,IL6 |

(*) Strength is calculated by STRING as $\log_{10}$(observed/expected). (†) False discovery rates (FDR) were calculated by the Benjamini and Hochberg method[127] as part of STRING. (‡) Specific genes related to each pathway.

## Appendix 2.11

**A.** Hyper and hypomethylated DMR statistics in fibroblasts for all samples and by family

| | All Samples | Family A | Family C |
|---|---|---|---|
| **Total DMR** | 1485 | 5713 | 4924 |
| **Ambiguous*** | 5 | 46 | 25 |
| **Unambiguous†** | 1480 | 5667 | 4899 |
| **Overlap filter‡** | 1479 | 5378 | 4725 |
| **Hyper DMR** | 885 | 3339 | 2872 |
| **Hypo DMR** | 594 | 2039 | 1853 |

**B.** Hyper and hypomethylated DMR statistics in iPSCs for all samples and by family

| | All Samples | Family A | Family C |
|---|---|---|---|
| **Total DMR** | 511 | 1083 | 1547 |
| **Ambiguous*** | 1 | 8 | 2 |
| **Unambiguous†** | 510 | 1075 | 1545 |
| **Overlap filter‡** | 506 | 1004 | 1496 |
| **Hyper DMR** | 238 | 646 | 559 |
| **Hypo DMR** | 268 | 358 | 937 |

(*) Tiles containing differentially methylated CpGs with methylation differences with opposite directionality (hyper- or hypomethylation) were considered ambiguous. (†) Tiles containing differentially methylated CpGs with methylation differences with the same directionality. (‡) DMRs were filtered to keep only those with CpG methylation data found in both Family A and Family C.