UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Functional analysis of epigenetic differentially methylated regions in the context of rare diseases

Catarina Zorro Nobre Mesquita Macedo

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Professor Matthieu Defrance
Professora Margarida Henriques da Gama Carvalho

2024

Acknowledgements

I would like to thank first Professor Matthieu Defrance for accepting me in his lab at the Interuniversity Institute of Bioinformatics in Brussels (IBsquare), for introducing me to a very interesting research project, as well as for his availability to answer and guide my questions.

I would also like to thank my colleagues at IBsquare for making me feel welcome in their group since the first day, inside and outside of the lab. A special thanks to Robin Grolaux and Edoardo Giuili who always helped me with questions on this work.

I also want to thank my internal supervisor, Professor Margarida Carvalho, for reading my thesis and giving very detailed suggestions on how to improve it.

Finally, I want to give a huge thanks to my parents, sister, grandma, aunt and all of my friends, for all the emotional support they gave me through the development of this thesis.

Resumo

A epigenética consiste no estudo de todas as mudanças que afetam a expressão génica (e o transcritoma) sem alterar a sequência do DNA, transmitidas por mitose e meiose. Ao conjunto de todas as modificações epigenéticas ocurrentes ao longo do tempo num organismo, dá-se o nome de epigenoma. Este é dinâmico e influenciado por fatores genéticos e ambientais. Os mecanismos de modificação epigenética podem dividir-se em três, que colaboram entre si, contribuindo para um desenvolvimento normal. Dois afetam a transcrição controlando a acessibilidade dos fatores de transcrição (TFs) ao DNA, sendo estes a metilação do DNA (DNAm) e modificações nas histonas. As modificações das histonas alteram a estrutura da cromatina para estruturas mais relaxadas – eucromatina – ou mais constritas – heterocromatina – as quais facilitam ou dificultam o acesso dos TFs ao DNA, respetivamente. A DNAm consiste na transferência covalente de um grupo metilo do S-adenosil metionina ao carbono-5 de uma citosina, catalisada por DNA metiltransferases. A metilcitosina (5-mC) está presente em 4% do genoma humano, maioritariamente em contextos CG (CpGs) nos animais. Estes tendem a aglomerar-se formando regiões chamadas ilhas CpGs (CGIs), que os protege da metilação. Cerca de metade dos CpGs localiza-se nos promotores génicos, perto dos locais de iniciação de transcrição (TSS), e o resto distribui-se por regiões intergénicas (IGRs) ou no corpo génico. No genoma humano, mais de 80% dos CpGs estão metilados, excluindo os promotores e enchancers de genes ativamente expressos. Promotores ou enhancers densamente metilados associam-se à heterocromatina e repressão da expressão génica. Contrariamente, a metilação dos CpGs no corpo génico favorece a transcrição. As tecnologias de deteção da DNAm englobam tecnologias baseadas no pré-tratamento do DNA com bisulfito e tecnologias de sequenciamento de terceira geração como o nanopore. Na deteção com bisulfito, é aplicado bisulfito ao DNA desnaturado, convertendo as citosinas não metiladas em uracilo, e deixando as citosinas metiladas intactas. Depois, para análises envolvendo todo o genoma, são utilizados normalmente microarrays genómicos, como o Illumina Infinium Human Methylation 450 BeadChip (HM450) e o Illumina Infinium Methylation EPIC BeadChip (EPIC). Estes contêm milhares de micropérolas, cada uma contendo centenas de cópias de sondas oligonucleotídicas complementares a citosinas metiladas e não metiladas em CpGs específicos. Depois da conversão bisulfito e amplificação por PCR, os fragmentos de DNA hibridizam com as sondas, e o sinal de intensidade medido para cada CpG reflete o seu nível de metilação. O HM450 cobra um total de 485 577 CpGs distribuídos nos promotores, 5UTRs, corpo génico e 3UTRs, cobrando 96% das CGIs. O EPIC é um microarray mais atualizado, cobrando 850 000 CpGs. Mutações em genes codificantes de proteínas envolvidas nas modificações epigenéticas são causadores de Distúrbios Mendelianos da Máquina Epigenética (MDEMs). MDEMs incluem distúrbios do neurodesenvolvimento (NDDs), caraterizados por alta comorbidade e problemas cognitivos, comportamentais e/ ou motores desde a infância até ao fim da vida. Também podem incluir malformações dos membros, retardamento do crescimento ou crescimento excessivo, e problemas no sistema imunitário. O diagnóstico começa com uma avaliação clínica dos sintomas, seguida de testes genéticos. Estes podem ser direcionados a um gene/ conjunto de genes de interesse, a uma avaliação de desequilíbrios cromossómicos ou um sequenciamento total do exoma ou do genoma. No entanto, um diagnóstico pode demorar 1-8 anos, e 50-66% dos indivíduos permanece não diagnosticado. Assim, uma análise do metiloma surgiu como um teste especialmente útil no diagnóstico destes casos. As mutações génicas relacionadas com MDEMs provocam padrões de DNAm ao longo do genoma caraterísticos da mutação subjacente, denominadas assinaturas epigenéticas (esigs). Estas têm vindo a ser utilizadas como biomarcadores no diagnóstico de várias MDEMs, e no estudo dos mecanismos da doença afetados por estes padrões de metilação. O uso de esigs depende de modelos de aprendizado de máquina (ML). O metiloma analisado provém normalmente de amostras de

sangue periférico de indivíduos com diagnóstico clínico validado e portadores da variante patogénica respetiva. O resultante conjunto de citosinas diferentemente metiladas (DMCs) é dividido em grupos de treino e de teste, que são aplicados a modelos de ML supervisionados. O grupo de treino é treinado para dar a probabilidade da amostra pertencer a um indivíduo afetado. O grupo de teste avalia o desempenho do modelo treinado, incluindo a precisão, sensibilidade e especificidade. O modelo também é normalmente validado utilizando um conjunto de amostras independente. Nesta tese, foram comparados os conjuntos de DMCs pertencentes a dois modelos de esigs de três NDDs: CHARGE, Kabuki e Sotos. Estes são maioritariamente causados por mutações nos genes CHD7, KMT2D e KDM6A (Kabuki I e II), e NSD1, respetivamente. Estes modelos foram desenvolvidos por dois grupos de investigação liderados por Sadikovic e Weksberg. A comparação consistiu em analisar a distribuição genómica dos DMCs e os processos biológicos (BPs) em que os genes diferentemente metilados (DMGs) estavam envolvidos e qual a sua relação com as manifestações clínicas. As CHARGE-esigs consistiram em DMCs maioritariamente distribuídas no corpo génico e IGRs. Os BPs consistiram no desenvolvimento embriónico dos sistemas esqueléticos e circulatórios, cérebro, especificação do eixo anteroposterior, regulação transcricional, proliferação celular, sinalização do receptor do tipo Toll e adesão celular. Estes relacionam-se com alguns sintomas do CHARGE, a exemplo, escoliose, defeitos no septo atrioventricular, excesso de fluido cerebrospinal, e mudanças na adesão de células da crista neural (NCCs). Os genes com maiores valores diferenciais de metilação – valores $\Delta\beta$ – foram o HOXA5 e HOXA-AS3 (hipermetilados) e HOTAIRM1, HOXA1, e SLITRK5 (hipometilados) à semelhança de estudos prévios, também como LINC02914 (hipometilado). Aqueles envolvidos em vias moleculares do desenvolvimento foram o HOXA1 (especificação das NCCs e desenvolvimento do coração e ouvido), SLITRK5 (sinapses e neurogénese), HOXA5 (inervação do diafragma), APP (neurogénese, sinaptogénese e adesão celular), COL4A2 (vascularização cerebral), GFI1 (diferenciação ciliar do ouvido interior), NOX4 (protetor do coração aquando elevada pressão sanguínea), KIRREL3 (mediador sináptico, com variantes associadas a hipoplasia cerebelar), DAB1 (processo cognitivo), ARHGEF15 (vascularização neoretinal), FOXP2 (desenvolvimento craniofacial) e PARVA (desenvolvimento do septo cardíaco). As Kabuki-esigs consistiram em DMCs maioritariamente distribuídos no corpo génico. No grupo de Sadikovic, os DMGs estavam envolvidos no metabolismo e transporte transmembranar de lípidos, regulação da transcrição, fagocitose, morte e proliferação celular, e desenvolvimento do nervo cranial V. Estes relacionam-se com o Kabuki. O KMT2D participa no metabolismo lipídico, e a sua perda de função rompe o ciclo de proliferação celular e promove a atrofia do tronco cerebral (que engloba o nervo cranial V). No grupo Weksberg, os BPs estavam envolvidos na apoptose, desenvolvimento esquelético e glândula mamária, especificação anteroposterior, angiogénese, transporte da leucina e prolina, adesão e senescência celular e regulação do movimento ciliar. No Kabuki, há aumento dos processos apópticos, desenvolvimento mamário precoce, defeitos na angiogénese (ex: defeito do arco aórtico), ciliopatias (ex: hidrocefalia) e escoliose. A prolina é abudante nas sinapses e a leucina ativa a secreção de insulina, sendo hipoglicémia uma manifestação rara do Kabuki. Os genes mais diferentemente metilados consistiram no ZMIZ1 (hipermetilado), e TNS1, AGAP2 e AGAP2-AS1 (hipometilados). Aqueles envolvidos no desenvolvimento foram o ZMIZ1 (gene candidato do Kabuki e participante da diferenciação neuronal e da glia), LAMA1 (desenvolvimento vascular retinal e adesão da célula de copo óptico, podendo relacionar-se com a falta de vista no Kabuki), ARHGEF7 (sinaptogénese, com haploinsuficiência em casos de deficiência intelectual ou ID), RPS6 (envolvida em retardamento do crescimento fetal), MXRA8 (maturação da barreira hematoencefálica), ESR1 (ligante do KMT2D; relacionado com hipospadias e diferenciação glandular mamária), CNTN5 (sinapses, relacionado com autismo, raro no Kabuki) e KIRREL3 (variantes envolvidas no autismo e ID). Finalmente, na Sotos-esig, os DMCs distribuíram-se maioritariamente no corpo génico, seguido por

IGRs e promotores, e perfil totalmente hipometilado. Os BPs estiveram enriquecidos em vias biológicas de migração e crescimento dos fibroblastos, morfogénese de estruturas ramificadas, proliferação celular, assemblamento dos nucleossomas e regulação da sinalização MAPK no grupo Sadikovic. No Sotos, há relatórios de atividade diminuída da cascada MAPK relacionada com o crescimento excessivo e proliferação celular desregulada. Na esig de Weksberg, diversos processos de regulação, desenvolvimento, transporte, metabolismo, vias de sinalização e adesão celular estiveram enriquecidos. Destacam-se as sinaléticas Wnt (desregulação envolvida no crescimento excessivo) e BMP (essencial em etapas iniciais do desenvolvimento), desenvolvimento esquelético e da pele, e desenvolvimento do sistema nervoso e transporte de neurotransmissores, relacionados com ID. Os DMGs de topo foram os PCAT6 e KDM5B. Aqueles envolvidos no desenvolvimento foram o KDM5 (diferenciação de células progenitoras neuronais), ZPR1, MIR200B, H4C1 (microcefalias e deficiência intelectual ou ID), ITGA2B (sépsis, por vezes presente em recém-nascidos com Sotos), TP73, TPPP3 (desenvolvimento musculoesquelético e regeneração axonal), TET1 (reprogramação epigenética), LDB3 (miopatias relacionadas com dismorfismo facial do Sotos), CERS2 (defeitos mielínicos), SFN, HLX, HYAL2 (variantes em dismorfismo facial), CXCL12 (neurogénese; variantes em NDDs), TDGF1 (defeitos no coração) , VAX2, DDR1 (formação mielínica), NOTCH4 (cardiomiopatias), ACSL6 (produção do ácido docosahexaenóico, protetor neuronal), DLX3 (desenvolvimento craniofacial; sobreexpressão na preclampsia presente no Sotos), CREB3L1 (osteogénese imperfeita relacionada com idade óssea avançada no Sotos), GAB1, FOXL1 (desenvolvimento cerebral), RD3, GIT1 (sinapses), RASIP1, BBS4, COLI2A1 (causativo de um distúrbio do tecido conetivo e muscular), NCS1 (diferentemente expresso em NDDs), e o ZNF423 (neurogénese). Esta análise funcional pode ser utilizada no desenvolvimento de modelos de esigs mais robustos e mais relacionados com as caraterísticas do síndrome.

Palavras-chave

CHARGE; Kabuki; Sotos; assinaturas epigenéticas

Abstract

Epigenetics consists of modifications affecting gene expression (and transcriptome) without affecting the DNA sequence. Cytosine methylation is the most prevalent modification. Mutations of genes involved in the epigenetic machinery create DNA methylation (DNAm) patterns - episignatures (esigs) – that lead to differential expression of several genes. These patterns form in early embryo development and are thought to cause Mendelian neurodevelopmental disorders (NDDs). This thesis focused on CHARGE, Kabuki and Sotos rare NDDs, caused by mutations in CHD7, KMT2D or KDM6A (Kabuki subtypes) and NSD1, respectively, which create specific patterns of differentially methylated cytosines (DMCs). A functional analysis of a pair of esigs (DMCs set) from Sadikovic and Weksberg research groups for each syndrome was performed to compare genic distribution, see differentially methylated genes (DMGs), main affected biological processes (BPs) and link to clinical manifestations. The least differences were of CHARGE-esigs. In Kabuki and Sotos, however, the Weksberg results were more aligned with the clinical manifestations in the BP enrichment analysis. In CHARGE, the top DMGs were homeobox and related genes (HOXA5, HOXA-AS3, HOTAIRM1 and HOXA1), LINC02914 and SLITRK5. Those involved in developmental pathways were HOXA1, SLITRK5, HOXA5, APP, COL4A2, GFI1, NOX4, KIRREL3, DAB1, ARHGEF15, FOXP2 and PARVA. In Kabuki, top DMGs were ZMIZ1, TNS1, AGAP2 and AGAP2-AS1. Those of developmental pathways were ZMIZ1, LAMA1, ARHGEF7, RPS6, MXRA8, ESR1, CNTN5 and KIRREL3. In Sotos, the top DMGs were PCAT6 and KDM5B. DMGs in developmental pathways were KDM5, ZPR1, MIR200B, H4C1, ITGA2B, TP73, TPPP3, TET1, LDB3, CERS2, SFN, HLX, HYAL2, CXCL12, TDGF1, VAX2, DDR1, NOTCH4, ACSL6, DLX3, CREB3L1, GAB1, FOXL1, RD3, GIT1, RASIP1, BBS4, COLI2A1, NCS1, and ZNF423. Together, this analysis paves the way for improvement on the classification models for studying DNAm patterns by selecting the most relevant CpGs, and provides a better understanding of the disorder mechanisms.

Keywords

CHARGE; Kabuki; Sotos; episignatures

Abbreviation of terms

5-caC - 5-carboxycytosine
5-fC - 5-formylcytosine
5-hmC - 5-hydroxymethylcytosine
5-mC - methylcytosine
6-mA - 6-methyladenine
CGI - CpG island
AD - Alzheimer's disease
ADHD - Attention-deficit/hyperactivity disorder
AP - Anteroposterior pattern
ASD - Autism spectrum disorder
BER - Base-excision repair pathway
BP - Biological process
chr - Chromosome
CMA - Chromosomal microarray analysis
CNS - Central nervous system
CNV - Copy number variant
DD - Developmental delay
DMC - Differentially methylated cytosine
DMG - Differentially methylated gene
DNAm - DNA methylation
DNMT - DNA methyltransferase
dsDNA - Double-strand DNA
EA - Enrichment analysis
EPIC - Illumina Infinium Methylation EPIC BeadChip
esig - Episignature
HM450 - Human Methylation 450 BeadChip
hmTOP-seq - 5hmC-specific tethered oligonucleotide-primed sequencing
ID - Intellectual disability
LASSO - Least absolute shrinkage and selection operator regression
lncRNA - Long non-coding RNA
MAPK - Mitogen-activated protein kinase
MBD - Methyl-CpG-binding domain
MeDIP - Methylated DNA immunoprecipitation
MDEM - Mendelian Disorders of the Epigenetic Machinery
ML - Machine learning
MSP - Methylation-specific PCR
MRE - Methylation-sensitive restriction enzyme
NDD - Neurodevelopmental disorders
NCC - Neural crest cells
pCGI - Promoter in CpG island
PGCs - Primordial germ cell
RRBS - Reduced representation bisulfite sequencing
SAM - S-adenosyl methionine

SNV - Single nucleotide variant
TETs - Ten-eleven translocation enzymes
TDG - Thymine DNA glycosylase
TF - Transcription factor
TLR - Toll-like receptor
TSS - Transcription starting site
UTR - Untranslated region
VUS - Variant of uncertain clinical significance
WES - Whole exome sequencing
WGS - Whole genome sequencing
WGBS - Whole genome bisulfite sequencing

Contents

List of Figures

List of Tables

I. Introduction

1.1 Epigenetics definition and mechanisms

The term ”epigenetics” (”over the gene”) refers to the study of all stable, heritable changes affecting gene expression during mitosis/meiosis without altering the DNA sequence [1]. It has been added to this definition all the molecular mechanisms that merely alter the transcriptional profile of a cell [2]. Related terms include ”epigenome”, capturing all epigenetic modifications over time. The epigenome is dynamic, heritably transmitted [3], and is affected by both genetic factors - gene mutations - and environmental factors, namely diet, physical exercise, stress, smoking, toxins, and drugs. Epigenetic mechanisms can be divided into three main processes: DNA methylation, histone modifications and noncoding RNA activity. The first two influence transcription processes, by changing the accessibility of genomic loci to the transcription enzymes [4]. Histone modifications include acetylation, methylation and phosphorylation, which either loosen (euchromatin) or tighten (heterochromatin) the chromatin, affecting gene expression positively or negatively, respectively [5]. At the RNA level, on the other hand, epigenetic marks affect translation processes; for example, microRNAs target and degrade some transcripts, thereby preventing their turning into proteins [6]. These mechanisms collaborate for normal cell development and differentiation across all body cell types [7]. DNA methylation (DNAm) is discussed further as it has a central focus in this thesis.

1.2 DNA epigenetic modifications

DNA modifications occur in adenine or cytosine residues, and cytosine methylation is the more extensively investigated. Modified DNA bases found in the genome are 5-mC (5-methylcytosine), 5-hmC (5-hydroxymethylcytosine), 5-fC (5-formylcytosine), 5-caC (5-carboxycytosine), and 6-mA (6-methyladenine) [8].

Methylcytosine (5-mC) is the predominant modified base, constituting over 4% of the human genome [9]. 5-mC can occur symmetrically in CG and CHG contexts or asymmetrically in CHH regions (H=A, T, or C) (Figure 1.1). In animals, 5-mC mainly occurs in CG contexts, where a cytosine is followed by a guanine, abbreviated as CpG sites (CpGs) [10]. CpGs tend to cluster together forming CpG high-density regions called CpG islands (CGIs) [3] [11], which protects them from DNA methylation. Thus, the majority of CGIs are usually unmethylated in the genome [12]. CGIs are over 200 base pairs (bps) long, with more than 50% CG content and CpG density over 60% compared to the whole genome [13] [14]. Half are in gene promoters near transcription start sites (TSS), while the rest are unannotated ”orphans”, found in intergenic or intragenic regions [15] [16]. Regions near CGIs also have been designated, namely CpG shores (0-2kb from CGIs) and CpG shelves (3-4 kB from CGIs). Areas outside these regions are termed ”open sea” or inter-CGIs [17]. In the human genome, over 80% of CpGs are methylated, with site-specific gaps in enhancer and promoter regions of actively

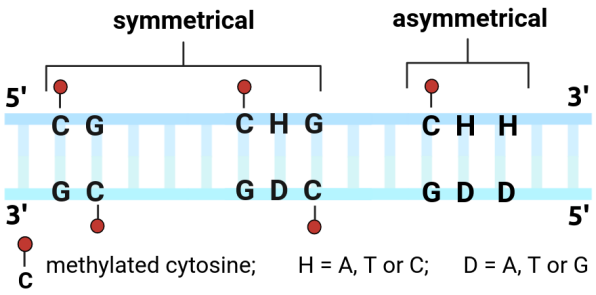


Figure 1.1. Cytosine methylation contexts. Cytosine methylation can occur in symmetrical contexts, that is in opposite cytosines in the double DNA strand, or asymmetrical contexts, that is in cytosines in just one of the DNA strands. Symmetrical methylation happens in either CG or CHG regions, and asymmetrical methylation happens in CHH regions, in which H can be A, T or C. This Figure was created with BioRender.com.

expressed genes, which are regulatory regions of gene transcription [18]. Figure 1.2 depicts the potential genomic locations of CpGs.

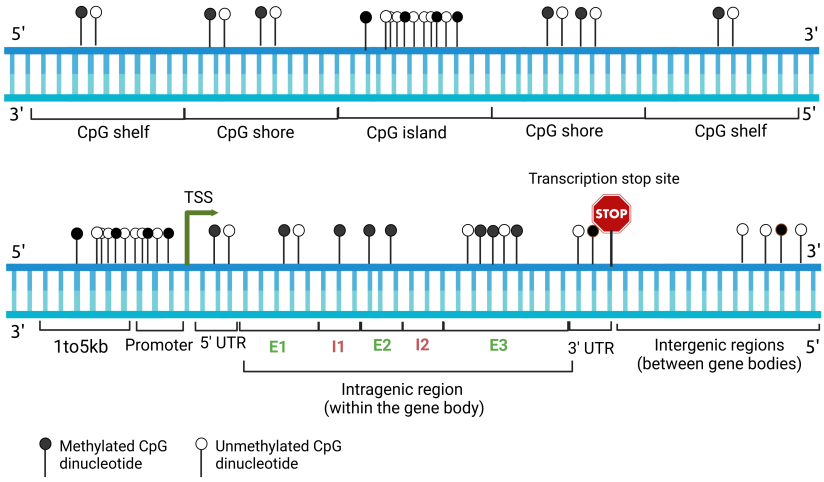


Figure 1.2. Schematic representation of the distribution of methylated and unmethylated CpG sites across the genome. The upper scheme translates the position of CpG shores and shelves relative to CpG islands. The down scheme translates the possible position of CpG sites across the region 1-5kb upstream of the promoter, the promoter region, transcription start site (TSS), 5'UTR (untranslated region), intragenic/gene body region, 3'UTR, transcription stop site, and intergenic regions. E1, E2 and E3 represent exons, and I1 and I2 represent introns. Figure created with BioRender.com.

Regulation of gene expression by CpG methylation

CpG methylation's impact on gene expression varies according to CpG genomic location. High DNAm in promoter CGIs (pCGIs) or enhancers relate to repressive heterochromatin structures, hindering transcription activators from assessing their cis-regulatory elements, and additionally recruiting repressive proteins [19] [20] [21]. Low pCGI methylation allows euchromatin states, promoting active transcription and proper transcription factor (TF) binding [19] [22] [23]. On the other hand, the methylation of CpGs within gene bodies (coding region excluding the TSS and transcription stop site) favours both transcription and post-transcriptional events, such as splicing, which is explained by the observation that exons are usually more heavily methylated than introns [24] [25] [26]. Additionally, methylation in centromeres and repeat regions enhances chromosome stability, also favouring gene expression [27].

Chemistry of cytosine methylation reaction

Cytosine methylation consists of the covalent transfer of a methyl group (CH_3) from the donor S-adenosyl methionine (SAM) to the C-5 position of cytosine rings on a DNA strand and is catalyzed by DNA methyltransferases (DNMTs). There are four human DNMT types - DNMT1, DNMT2, DNMT3a, and DNMT3b - crucial for normal development. DNMT1 methylates DNA already hemimethylated during DNA replication, thus being crucial for maintaining the DNAm patterns in the daughter strands. DNMT3a and DNMT3b prefer unmethylated CpG dinucleotides, participating in *de novo* methylation during development. Therefore, these three enzymes (DNMT1, DNMT3a and DNMT3b) cooperate in both the initiation and maintenance of DNA methylation. DNMT2, however, methylates cytosine residues in transfer-RNA (tRNA), having no part in DNA methylation [28] [29] [30][31].

1.3 Technologies available to detect DNA methylation

DNAm, particularly 5-mC, is the most studied and widely distributed DNAm type in eukaryotes [32]. For detecting DNAm, DNA strands are usually first submitted to a pretreatment which discriminates between methylated and unmethylated cytosines [33] [34] [35]. Depending on the pretreatment, DNAm detection techniques can be grouped into three main techniques, those being based on restriction enzymes, bisulfite conversion, or affinity enrichment [36]. Nonetheless, recently third-generation sequencing technologies, such as Oxford nanopore, have also been applied to study DNAm, having the advantage of not requiring DNA pre-treatment. Nanopore can detect the bulk presence of unmethylated cytosines, 5-mC, 5-hmC, adenine (A) and 6-mA forms, in double-strand native DNA (dsDNA). Whilst the previous NGS detected only short-range methylation patterns, the nanopore can detect ultra-long reads at a genome-wide level, and in significantly less time than bisulfite sequencing methylation analysis [37]. Furthermore, nanopore has more genomic coverage and lower GC bias, greater reproducibility, and faster data analysis than bisulfite sequencing [38]. Nanopore detects DNA modifications by detecting differences in the electric current intensity produced when a modified (or unmodified) nucleotide passes by the nanopore [39]. This difference is determined after nanopore read base calling and alignment, either by comparing the electric current pattern to a reference sample [40], or by using pre-trained supervised machine learning (ML) models [41]. Limitations of nanopore include that it assumes that all CpGs within a 10-bp region have the same methylation status, thus it has trouble in detecting close CpGs with different methylation statuses [42]. Furthermore, while optimal for bulk samples, it is unsuitable for small DNA amounts as it does not rely on PCR, contrarly to bisulfite techniques. Plus, it has lower read accuracy [43] [39]. Bisulfite treatment-based technologies are further explained due to their relevance in the present thesis.

Bisulfite conversion treatment-based techniques

Bisulfite conversion is the gold standard technique for detecting DNAm status [44]. It starts on denaturation of the DNA strands and appliance of bisulfite, which converts the unmethylated cytosines to uracil (C→U), thus distinguishing methylated from unmethylated cytosines [25]. Post-treatment considerations include the study scale (low, medium, or high-throughput), initial DNA amount, cost, time, sensitivity, accuracy, and whether it is a qualitative or quantitative analysis [45] [46] [47]. For **locus-specific analysis** (specific target regions/genes of interest), the pre-treated DNA strands can undergo direct PCR followed by Sanger sequencing or pyrosequencing, or methylation-specific PCR (MSP). In simple PCR, uracil residues (from unmethylated cytosines) become thymines, while methylated cytosines remain unchanged. MSP employs methylated (M primer) and unmethylated (U primer) primers. The results are visualized by gel electrophoresis. Simple PCR allows detection of all CpGs within the target region, whilst MSP specifically detects CpGs with the desired methylation status. Pyrosequencing uses a sequencing-by-synthesis approach, with light production using a luciferin-luciferase system indicating nucleotide incorporation. The methylation percentage of each cytosine is calculated from the ratio of heights of the cytosine peak (methylated signal) and the sum of cytosine and thymine peaks (methylated and unmethylated signals) [36] [48]. For **genome-wide analysis**, DNA methylation status is assessed using genomic microarrays or high-throughput NGS [49] [50]. Commonly used microarrays include Illumina Infinium Human Methylation 450 BeadChip (HM450) and Illumina Infinium Methylation EPIC BeadChip (EPIC). BeadChip arrays have thousands of silica microbeads containing oligonucleotide probes binding to methylated or non-methylated cytosines at specific CpGs. After bisulfite conversion and PCR amplification, DNA fragments hybridize with the probes, and array scanning measures signal intensity at each CpG, which reflects the level of methylation at that site. There are two types of Infinium designs.

Infinium I assays two bead types per CpG locus, one for methylated and the other for non-methylated cytosines. Infinium II assays use only one bead type, and the methylated state is determined by single base extension after hybridization [51]. HM450 covers a total of 485,577 CpGs. Those are distributed in promoters, 5UTRs, first exons, gene body, 3UTRs, and 96% of all CGI and near-CGI regions, as well as other additional categories outside of those [51]. However, it lacks coverage in certain regulatory regions. EPIC is an improvement of HM450 that covers an additional 413,743 CpGs encompassing enhancer regions, totalling over 850,000 CpGs [52]. On the other hand, high-throughput NGS commonly employs whole genome bisulfite sequencing (WGBS) for comprehensive CpG information. However, it is high-cost and generates vast data that requires sophisticated bioinformatics analysis. Since only a small fraction of the genome is found differentially methylated, cheaper alternative methods such as reduced representation bisulfite sequencing (RRBS) have been used instead of WGBS. RRBS integrates restriction enzyme digestion, bisulfite conversion and NGS, covering 85% of CGIs [53] [54].

Despite their popularity, bisulfite conversion methods have some drawbacks. The converted DNA is unstable and prone to degradation, which affects the quality and quantity of DNA available for downstream analyses, and complicates PCR. Techniques using restriction enzymes or affinity enrichment address this issue but offer lower genome resolution and coverage compared to bisulfite conversion, which provides single nucleotide resolution [55] [56].

1.4 Technologies to detect oxidised forms of methylation

Similar to 5-mC, its oxidized form, 5-hmC, is safeguarded from bisulfite-induced deamination. Consequently, bisulfite sequencing cannot differentiate between these two forms. To address this, modified bisulfite sequencing methods, such as oxidative bisulfite sequencing (OxBS-seq), were developed. In this method, 5-hmC is first oxidised to 5-fC and then 5-fC is decarbonylated and deaminated to uracyl under bisulfite conditions. 5-mC remains unchanged, thus it is possible to distinguish 5-mC from 5-hmC sites. The amount of 5-hmC at a particular position is deduced by subtracting the OxBS-seq readout from the unoxidised bisulfite sequencing (BS-seq) [57]. hmTOP-seq surged recently as another method for 5-hmC detection that does not rely on bisulfite pretreatment, and with single-base resolution and higher performance [58]. On the other hand, both 5-fC and 5-caC are deaminated by bisulfite and read as unmethylated cytosines (that is thymines after PCR). Thus, 5-fC and 5-caC positions are determined by protecting them from the bisulfite-induced deamination and then subtracting the cumulative levels of 5mC and 5hmC obtained in ordinary BS-seq (fCAB-seq and CAB-seq, respectively) [59] [60]. 5-fC can also be detected by selectively reducing it to 5-hmC (redBS-Seq) [61].

1.5 Epigenetics and relationship with neurodevelopmental disorders

The acquiring of epigenetic information starts in early development, such that multiple waves of epigenetic reprogramming were already identified. Mutations in genes coding for proteins that read, write or erase epigenetic modifications (e.g. DNMTs, histone-modifying enzymes or chromatin regulators) are the cause of several developmental disorders, which are also named ”Mendelian Disorders of the Epigenetic Machinery” (MDEMs) [62]. These mutations ultimately lead to differential TF accessibility to chromatin binding sites, thus disrupting gene expression that causes developmental errors. MDEMs include neurodevelopmental disorders (NDDs), which consist of a group of conditions that disturb the development of the nervous system, resulting in cognitive, behavioural and/or motor impairments that start manifesting in childhood and persist throughout the individual’s life [63]. NDDs comprise patients with intellectual disabilities (ID), autism spectrum disorder (ASD), communication disorders, attention-deficit/hyperactivity

disorder (ADHD), neurodevelopmental motor disorders (e.g. Tic disorder like Tourette's), epilepsy and other specific learning disorders [64]. NDDs have high rates of comorbidity, and they often have overlapping clinical manifestations resulting from the improper regulation (both temporal and spatial) of common molecular pathways during embryonic development, which makes their diagnosis difficult. Besides neurological and intellectual dysfunctions, NDD features can also include limb malformations, growth anomalies (growth retardation or overgrowth), and dysfunction of the immune system [62]. Genetic mutations causing NDDs can range from single nucleotide variants (SNVs) to whole chromosome aneuploidies [65].

NDD diagnosis usually starts with a clinical assessment, where a particular disorder or set of disorders might be suggested. In plenty of cases, however, the individual presents a set of atypical features that together do not suggest any particular disorder [66] [67] [68]. Moreover, not all individuals with clinical manifestations have necessarily pathogenic (or likely pathogenic) variants of the gene-associated syndrome and, in some cases, they have a variant of this gene but in which there is no validation yet of its significance (that is, a variant of uncertain clinical significance or VUS) [67] [68]. Based on this first assessment, genetic testing follows, which can include targeted gene sequencing (in case of suspecting of one SNV), a multigene panel approach (multiple genes of interest), global genomic screening (such as chromosomal microarray analysis, shortened as CMA), chromosome karyotyping (much lower resolution compared to CMA), whole exome sequencing (WES) or whole genome sequencing (WGS). CMA is considered a first-tier test in NDD diagnosis, often performed early in clinical evaluations for individuals with developmental delays (DD), sometimes before even considering a specific disorder. CMA detects copy number variants (CNVs) [69][70] [71], using oligonucleotides or single nucleotide polymorphism (SNP) arrays [72]. Most NDD-related CNVs involve dosage-sensitive genes, which are highly dependent on the normal number of alleles for proper product expression. Those include haploinsufficient, triplosensitive, or imprinted genes [69]. WES and WGS are more recent molecular techniques that offer higher diagnostic yield than CMA, being particularly useful when no specific disorder was suspected [66] [67] [68]. Those advances also led to discovery of new VUS [73]. Despite these advances, around 50-66% of patients remain undiagnosed with these methods [74] [75] [76]. Adding to this, a proper diagnosis may take years and many clinical evaluations and tests, with an average time in diagnosing individuals referred for genetic testing ranging from 1 to 8 years [69]. This is not suitable for either individual's family and health care systems, therefore any testing technology that can reduce this time is valuable [77].

In this context, mutations in epigenetic regulatory genes involved in DDs were shown to create genome-wide DNAm patterns specific to the underlying mutation. These disorder-specific DNAm patterns are called episignatures (esigs). There is increasing evidence that esigs can be used as highly sensitive and specific epigenetic biomarkers for the diagnosis of several developmental disorders, which is particularly useful for solving unresolved cases and reclassification of VUS [69] [73]. Adding to this, they are also used to study the disease mechanisms affected by these widespread alterations in DNAm [73] [78]. So far, 30 genetic syndromes with mutations in 50+ genes have been identified to have specific DNAm esigs. Notably, esigs from different disorders or disorder subtypes often overlap [79]. For instance, disorders caused by the disruption of common protein complexes can have highly overlapping esigs, as it is with the case of BAFopathies, caused by the disruption of the chromatin remodelling complex BAF [80]. Conversely, distinct esigs can result from mutations in the same gene affecting different protein domains, as seen in ADNP syndrome with two genome-wide DNAm profiles from gene mutations affecting separate protein domains [81]. Differentially methylated genes (DMGs) can go from hundreds to tens of thousands of CpGs in the methylation arrays, and the related mutated gene most commonly does

not have disrupted DNAm [82].

1.6 Use of machine learning methods in building episignatures

Esig mapping relies on genomic microarrays, such as the HM450 and EPIC arrays [83]. The ability to use esigs as viable diagnostic methods is due to large-scale reference DNAm databases and ML algorithms that make sense of the inputted DNAm data to classify disease states. A simplified pipeline for building ML esig models goes as follows. First, a genome-wide analysis of (usually) peripheral blood samples is performed using the microarrays, which generates a set of differentially methylated cytosines (DMCs) of patients with known disorders and presenting respective pathogenic variants *vs.* unaffected controls. This set of DMCs is then split into two sub-data sets: the training and testing sets. Those are combined within supervised ML models (classifiers), which rely on the input of labelled features, which consist on the DNAm profiles labelled with having the disorder (or different disorder subtypes) *vs.* being healthy (controls) [84]. Limitations of supervised ML are its dependence on both data quality and correct labelling, and the fact that it is more prone to have bias towards the training set, which means that it can work very well for the training set while having limited performance for other datasets [85]. Common supervised ML models include linear/logit regression, support vector machine (most common approach), random forest algorithms and LASSO [86] [87]. In the training phase, the model is trained to give the probability of having the disorder based on the labeled dataset. The testing set is smaller in size and is used to evaluate the performance of the trained model. Unlike the training phase, the testing phase does not train the model to best fit the data. If the previously trained model effectively predicts the outcome in the testing set with good performance, then it is a model that can be further used in unseen samples. In balanced datasets, performance evaluation includes accuracy, sensitivity (testing for false negatives), specificity (testing for false positives), and precision of the model [88] [89]. The specificity is evaluated by comparing the patient's DNAm data against controls, and can be improved by a multiclass approach where other MDEMs sharing clinical manifestations are included in the control set [90] [91]. For imbalanced datasets (few cases compared to controls), alternative metrics like F1-score, area under the curve (AUC), and Cohen's Kappa are more appropriate to measure performance [88] [89] [92]. Finally, there is also a validation set, that must come from other sources than the original data and it is used to validate the model. This step is not always performed due to data unavailability. In such cases, an alternative is incorporating a validation step during training, like k-fold cross-validation, where the training set data is randomly split into k training and test sets [93].

Unsupervised ML algorithms can also be used in epigenomics studies, for clustering and association tasks [86], for example, to identify DMCs with significant variability across samples, indicative of specific conditions or subgroups within the dataset [94] [95]. Or, for example, to ensure the esig can differentiate affected patients from controls [91]. Unsupervised ML models include k-means clustering, hierarchical clustering, principal component analysis (PCA), and partial least squares discriminant analysis [96] [97]. The main limitation of unsupervised ML is that whilst it can connect points of data, it cannot assign any meaning to them. Hence, it relies on the user to analyse if such relationships are plausible and have relevance within the context of the data [86] [98]. Furthermore, the correct pre-processing of data is of the utmost importance, since if there is a lot of noise in the data, the algorithm can cluster points together that make no sense [82].

Limitations of episignatures

One of the challenges that esig classifiers (as well as molecular genetic testing) face is the detection of mosaicism. Around 5 to 10% of pathogenic variants in Mendelian disorders are mosaics [99]. If the mutation occurs in the early developmental stages, it is more likely to find DNAm differences in the peripheral blood, whilst, if happening later, they may only affect specific tissues and not be detected in the blood [90]. To address that issue, several studies compared esigs based on peripheral blood samples and esigs based on other tissues such as the brain or fibroblasts. One study in Sotos syndrome reported that the set of DMCs could also differentiate fibroblast DNA of patients *vs.* controls, but with milder differences in methylation [100], and in a study of Down syndrome, both brain and blood esigs were distinct from controls but the DMCs were found in different loci respective of the tissue [101]. Moreover, even if esigs can help to solve the unresolved cases, and reclassify VUS, they cannot by themselves validate the VUS as a pathogenic variant causative of the specific disorder. For VUS validation, further studies are necessary such as gene expression studies and/or experiments that study directly particular DNAm changes [102] [103]. Notably, while blood tests are the most common for DNAm profile testing, it can also be conducted in prenatal testing, buccal swabs, biopsies, or amniotic fluid, particularly when no known DNAm signature was detected in the blood [102]. Moreover, diagnosing rare disorders, such as MDEMs, poses an additional challenge in DNAm mapping due to the limited number of affected individuals compared to studies on more prevalent conditions like cancer. Esig mapping heavily relies on the effect size - that is the difference in methylation levels also named $\Delta\beta$ values - and on the extent of DMCs. Conditions with a large effect size and numerous DMCs are easier to map with a smaller sample size. For rare disorders, because the number of samples is smaller, it is essential to take into account those effects, which can highly vary between syndromes [82]. Disorders with a milder DNAm pattern (lower $\Delta\beta$ values) typically require a larger sample size, around 10-20 samples [104]. Of note, esig models may have different cutoffs for $\Delta\beta$ and p-values, commonly set at $\sim$5-20% for $\Delta\beta$ values, with varying p-values based on sample size and confounding factors [82].

In this thesis, the focus is going to be on the functional analysis of episignatures of three NDDs, those being CHARGE, Sotos and Kabuki syndromes, which are all rare disorders with overlapping clinical features. A representation of the manifestations is represented in Figure 1.3.

1.7 CHARGE syndrome

CHARGE syndrome (CS) is a rare genetic disorder affecting around 1 per 10,000 births [66][105]. Its name is an acronym for Coloboma, Heart defects, Atresia choanae, Retarded growth and development, Genital hypoplasia, and Ear anomalies including deafness [106]. CS is considered in anyone with multiple anomalies and at least one of the major features. Major features include eye coloboma, choanal atresia or stenosis, cranial nerve abnormalities, swallowing and feeding difficulties, and CHARGE ear (short, wide outer ear with little or no lobe, malformed bones of the middle ear related to conductive hearing loss, and small/absent semicircular canals causing hearing loss and balance problems). Minor less specific features include heart defects, cleft lip or palate, tracheoesophageal fistula, small/missing/misplaced kidney with reflux, genital hypoplasia, short stature, typical CHARGE face (broad prominent forehead, arched eyebrows, large eyes, droopy eyelids, broad nasal bridge, thick nostrils, prominent nasal columella between the nostrils, flat midface, and small mouth) and ”hockey-stick” palmar crease. Other features include ear infections, sloping shoulders, absent or extra thumbs, vertebral abnormalities, scoliosis, hydrocephalus (extra CSF in brain ventricles), seizures, small or missing thymus (rare), omphalocele or umbilical hernias, extra/missing/misplaced nipples and hypotonia. Over 90% of CS cases are caused by heterozygous

LOF mutations in CHD7 (chromodomain helicase DNA-binding protein 7) located in 8q12.2 (OMIM: 608892), that occur mostly *de novo*. In rare cases, they can also be transmitted by autosomal dominant inheritance [66]. CHD7 encodes a chromatin-remodelling factor (CRF) that alters nucleosomes' position, thus modifying TF accessibility [107]. There is also evidence that CS can be caused by a mutation in the semaphorin-3E gene (SEMA3E) on chromosome 7q21 [108], which may provide a genetic link between CS and Kallman syndromes, which is an allelic disorder with less severe manifestations, and main features consisting of hypogonadism and impaired sense of smell [109].

1.8 Kabuki Syndrome

Kabuki syndrome (KS) is caused by either heterozygous variants of the gene KMT2D (lysine (K)-specific methyltransferase 2D) located on 12q13.12 (OMIM:602113), or hemizygous variants of KDM6A (Lysine Demethylase 6A) located on Xp11.3 (OMIM: 300128), and the syndrome subtype is named as Kabuki syndrome type 1 (KS-1) or type 2 (KS-2), respectively. It has an estimated incidence of 1 in 32,000 newborns [67][110]. KMT2D mutations occur mostly *de novo*, but in some rare cases, they can be transmitted by autosomal dominant inheritance. KMT2D is part of the SET-domain-containing family of histone methyltransferases, and it tri-methylates histone H3 at lysine 4 (H3K4me3), a histone marker involved in DNA methylation and transcriptional activation of genes [111] [112]. KDM6A pathogenic variants can also be *de novo* or inherited in an X-linked manner [67], and KDM6A is involved in the removal of H3K27me3 mark related to gene silencing [113] [114]. KS patients have psychomotor developmental delays, hypotonia, hypertrichosis (excessive hair growth anywhere in the body), and distinctive facial dysmorphism, particularly in the eyes, nose, lips, and ears. Skeletal anomalies include abnormalities in the spine (e.g. sagittal cleft or butterfly vertebrae, and scoliosis), hand anomalies (cone-shaped epiphyses, brachydactyly and clinodactyly), hip dislocation, pseudarthrosis of the clavicles, and others [115] [116] [117] [118] [67]. Besides, KS patients can present postnatal growth deficiency and anomalies in the cardiac, gastrointestinal and renal systems. Sometimes they also have behavioural issues such as ASD [119]. Another manifestation although very rare is hypoglycemia where patients have persistent hyperinsulinism which might be explained by KDM6A mutation and subsequent loss of H3K27me3 demethylation leading to deregulation of pancreatic beta cell development [120].

1.9 Sotos Syndrome

Sotos syndrome (SS) is caused by NDS1 (Nuclear Receptor Binding SET Domain Protein 1) haploinsufficiency, located in the 5q35 region. This is caused by heterozygous mutations or partial gene deletion or 5q35 microdeletions [121]. NSD1 is a histone methyltransferase (HMTase) that acts on H4K20 and H3K36 [68]. It has an incidence of 1 in 10,000 to 14,000 newborns. However, many of its symptoms relate to other disorders, thus it is likely that its incidence may be much higher, rounding 1 in 5,000 [122]. Sotos patients have facial dysmorphisms, such as broad and prominent foreheads, sparse frontotemporal hair, down slanted palpebral fissures (”sad” eyes), malar flushing (rash on the nose and cheeks), and long and narrow faces with small and pointed chins. Besides, they have mild-to-severe intellectual impairments, and overgrowth in their height and head circumference, which is more than 2 standard deviations (SD) above the mean, a condition known as macrocephaly. These are the cardinal features of SS, but other common features include behavioural issues such as ASD, advanced bone age, cardiac and renal anomalies, cranial abnormalities, joint hyperlaxity (very flexible joints) sometimes accompanied by pes planus (flat feet), scoliosis, seizures, and neonatal complications such as jaundice (abnormal heme metabolism causing high bilirubin levels, which causes yellow discolouration on the newborn's skin and

eyes), hypotonia and poor feeding. Besides, mothers of SS infants sometimes suffer maternal preeclampsia, characterized by hypertension and protein excess in their urine, forcing premature births [68] [123]. Other SS subtypes are SOTOS 2, caused by a mutation in NFIX on 19p13, and SOTOS 3, caused by a mutation in APC2 on 19p13 [108].

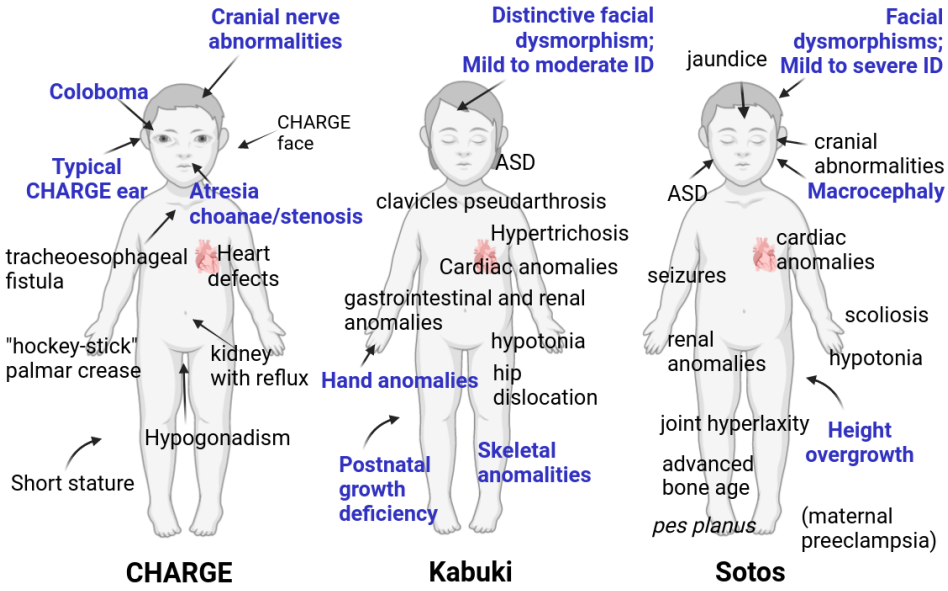


Figure 1.3. Clinical manifestations of CHARGE, Kabuki and Sotos syndromes. The main characteristic manifestations are highlighted in blue. "CHARGE face" consists of a broad prominent forehead, arched eyebrows, large eyes, droopy eyelids, broad nasal bridge, thick nostrils, prominent nasal columella between the nostrils, flat midface, and small mouth. Other less characteristic CHARGE features include ear infections, sloping shoulders, absent or extra thumbs, vertebral abnormalities, scoliosis, hydrocephalus, seizures, small or missing thymus (rare), omphalocele or umbilical hernias, extra/missing/misplaced nipples and hypotonia. Kabuki hand anomalies consist of cone-shaped epiphyses, brachydactyly and clinodactyly, and skeletal anomalies include abnormalities in the spine such as sagittal cleft or butterfly vertebrae, and scoliosis. Sotos facial dysmorphism consists of broad and prominent foreheads, sparse frontotemporal hair, down-slanted palpebral fissures, malar flushing, and long and narrow faces with small and pointed chins. ASD = Autism spectrum disorder. ID = Intellectual disability. Figure created with BioRender.com.

1.10 Objectives

Esig models are being more and more considered for clinical diagnosis on the classification of unresolved NDD cases, as well as possibly its implementation as a first-tier diagnosis. However, sometimes these models do not provide satisfactory results regarding functional annotation and relation with syndrome manifestations. In this context, it is important to validate these models by a doing functional analysis of their DMCs sets. This holds especially true for rare NDDs due to the reduced datasets available for building the models and the overlap sometimes extensive between NDD methylation patterns. The objectives of this thesis were to compare same-syndrome esig models of CHARGE, Kabuki and Sotos rare NDDs through functional analysis. This consisted on comparison of DMCs genic distribution, CGI context, and on the set of DMGs and their involvment in biological pathways, mainly those related to early development.

II. Material and Methods

2.1 Initial episignatures data

Two distinct esigs for syndromes CHARGE, Kabuki, and Sotos were obtained from two research groups, the Sadikovic's (S.) group [82] and Weksberg's (W.) group [124] [100]. Information about cohorts/sample sizes, final number of selected probes (DMCs), and the respective range of absolute $\Delta\beta$ values are described in Table 2.1. Further info can be found on the respective articles. The two groups had different cut-offs for what they considered to be differential methylation between the CpGs of discovery cohorts *vs.* controls. In all cases, probes with p-values >0.01, those located on sexual chromosomes, those which contained SNPs at CpG interrogation or single-nucleotide extension, and probes known to cross-react with other chromosomal locations other than target regions were removed. The genomic microarrays used for constructing the esigs were the HM450 array in Weksberg's group and both HM450 and EPIC arrays in Sadikovic's group, but in this last case, only HM450 was selected for downstream analyses.

Table 2.1. Cohorts sample sizes used in the development of episignatures by Sadikovic's (S.) and Weksberg's (W.) research groups for CHARGE, Kabuki and Sotos syndromes. The train and test sets consisted of patients with confirmed diagnosis of each syndrome and carriers of the associated gene variants. Blank values (-) mean the authors didn't supply information about the size of those cohorts/sets for the building of the episignature. Regarding **Sadikovic's group**, Kabuki subtypes 1 (KMT2D) and 2 (KDM6A) were treated as just one disorder for the establishment of the episignature.

Author	Syndrome	Gene variant	Train Set	Test Set	Validation Set	Number of selected DMCs	Absolute $\Delta\beta$ values
Sadikovic [82]	CHARGE	CHD7	45	15	-	148	[0.04,0.21]
	Kabuki	KMT2D, KDM6A	66	21	-	153	[0.04,0.24]
	Sotos	NSD1	47	15	-	112	[0.22,0.65]
Weksberg	CHARGE [124]	CHD7	19	-	40	75	[0.10,0.20]
	Kabuki [124]	KMT2D	11	-	17	113	[0.10,0.30]
	Sotos [100]	NSD1	57	19	1056	7,085	[0.20,0.67]

2.2 Annotation of differentially methylated cytosines

The DMCs for each syndrome-associated esig were annotated to genes and relation to CGIs on genome assembly GRCh37 (hg19) using the R (v. 4.3.1) package "annotatr" (v. 1.26.0). Genic annotations included 1-5Kb upstream of the TSS, gene promoters (defined as up to 1kb from the TSS), intergenic regions (IGRs), 5'UTRs and 3'UTRs, exons and introns by using the following built-in annotations: 'hg19_basicgenes' and 'hg19_genes_intergenic'. The exons and introns were considered part of the "gene body" category. CGI annotations ('hg19_cpgs') included both CpG islands and the regions upstream or downstream from the CGIs, that is the CpG shores (up to 2kb), CpG shelves (2-4kb) and inter CGIs (more than 4kb). Those were built using the built-in annotation 'hg19_cpgs'.

2.3 Enrichment analysis of biological processes

Enrichment of gene ontology biological processes (GO BPs) was performed using the DAVID (Database for Annotation, Visualization and Integrated Discovery) tool [125] [126] with the option "`GOTERM_BP_DIRECT`". The terms were further filtered for P-value ≤ 0.05. Then, they were reduced and summarized by REVIGO (REduce and VIsualize Gene Ontology) which removed redundant GO terms using the Homo

sapiens model (9606), and a semantic similarity cutoff of C = 0.70 for all esigs, except for Sotos-esig by Weksberg group, in which it was used a similarity cutoff of C = 0.50. Interactive graphs were also obtained from REVIGO [127].

2.4 Functional profile of gene sets involved in developmental pathways

For the analysis of genes involved in developmental pathways, only the common set of genes between the same syndrome esigs was selected. Then, they were mapped to a GO profile at a GO level 5 using the ”getGroupGO” function of ”CeTF” (v. 1.12.0). Those sets of genes were further filtered for the ones involved in GO BPs related to development processes (Description=”development”). Other more general terms were excluded, those being ”cell development”, ”embryo development”, ”regulation of multicellular organismal development”, ”embryonic organ development” and ”system development”.

2.5 Graphical representations

All plots using Rstudio were made with ”ggplot2” (v. 3.4.3). Representation of the distribution of CpG sites across the genome was represented using karyoploteR (v. 1.26.0).

III. Results

3.1 Comparison of differentially methylated cytosine distribution in the genome for same syndrome episignatures.

To first get a picture of the DMCs sets by the same syndrome esig models, the genomic context was accessed. Promoter DMCs are often related to gene silencing whilst methylation in the body is associated with enhanced gene expression [128] [129] [130]. Probes were annotated in the context of both CGIs (CpG shores, CpG shelves, CGIs, Inter CGIs) and genes (the 1-5Kb region upstream of the TSS, promoters, 5'UTRs, exons, introns, 3'UTRs, and intergenic regions/IGRs which excludes all the others). Exons and introns were grouped as a single category of gene body. The annotations were built using the annotatr R package, which uses TxDb.* and AnnotationHub resources. In this package, the promoter is defined as the region <1kb upstream of the TSS [131]. The number of DMCs per chr and their relative loci within chrs were also checked. This first analysis aimed to see if the same syndrome esigs models had similar DMCs distribution in the genome, despite having distinct DMCs sets.

In **CHARGE**, the DMCs methylation status ratio (hyper/hypo ratio) was drastically distinct between esigs, however those in common (n=19) showed the same methylation status. From the total of 148 probes selected by the S. group, most were hypermethylated (n=95, 64.19% hyper *vs.* n=53, 35.81% hypo). On the contrary, in the W. group, from the total of 75 probes, most were hypomethylated (n=21, 28% hyper *vs.* n=54, 72% hypo). The genic distribution was similar. In both groups, the majority of DMCs were present in the body (∼44–51%), following IGRs (∼16–25%), promoters (∼11–13%), 1to5kb (∼11–12%), and minimal amount in UTRs (∼4–6% in 5UTRs and ∼2–6% in 3UTRs) (Tables S1 and S2). Body and promoters were mostly hypermethylated in S. and mostly hypomethylated in W. Regarding CGI context, most DMCs were present in inter CGIs (∼49–59%), followed by shores (∼21–28%). In S., the rest of the DMCs were mostly found in CGIs (16.22%) and then shelves (4.73%), whilst in W. the rest of the DMCs were found in similar proportions in CGIs and shelves (n=8, 10.67%, and n=9, 12.00%, respectively) (Tables S3 and S4). DMCs could be found in all chrs in S. esig, but not in chrs 15, 19, 20 and 22 in W. esig. Chr 7 had the majority of DMCs in S. esig and chr 12 in W. esig. Other chrs with many DMCs were chrs 6 and 7 in W. esig. The higher differences in distribution were in chr12 (higher peak in W.), 15 and 16 (higher peaks in S.), and 21 (higher peak in W.) (Figures S1a and S1b). Inside chrs, most of them were present in distant loci, with few sharing proximal loci (Figure S2).

In **Kabuki**, the DMC methylation status ratio was similar for both groups, being around fifty-fifty (hyper/hypo ratio), and the probes in common (n=23) had the same methylation status. For the total of 153 probes in S. group, there was a ratio of 55.55% hypo (n=85) *vs.* 44% hyper (n=68). For the 113 probes in W. group, there was a ratio of 51.33% hypo (n=58) *vs.* 48.67% hyper (n=55). The genic distribution, while not exactly the same, was also similar. Most DMCs were present in the body (∼52–56%). In the case of S. group, the rest were mostly present in IGRs and 1to5kb regions (n=39, 16.96%), then in promoters (n=23, 10.00%) and less amount in UTRs (n=11, 4.78% in 5UTRs and n=8, 3.48% in 3UTRs). In the case of W., the rest of DMCs were mostly present in IGRs (n=37, 15.89%), followed by 1to5kb and promoters in similar amounts (n=19, 12.58% and n=18, 11.92%, respectively), and then in less amount in UTRs, with similar number in 3UTRs and 5UTRs (n=6,3.97%, n=5, 3.31%, respectively). In the body, most DMCs were hypomethylated. In promoters, most were hypomethylated in S., and the same amount was hypo or hypermethylated in W. (Tables S5 and S6). Concerning CGI annotation, the DMCs distribution was similar. Most DMCs were present in shores (∼41–46%), followed by inter CGIs (∼24–28%),

CGIs (∼22%) and shelves (∼8–9%) (Tables S7 and S8). DMCs could be found in all chrs in both esigs, but the percentage of DMCs per chr in the two esigs was different. Higher differences were found in chr12, 17 and 19 (higher peaks in W.) (Figures S1c and S1d). Inside the chrs, some of the DMCs were present in distant loci, while others were found in proximal loci (Figure S3).

In **Sotos**, DMC methylation status was overall hypomethylated. From the total of 112 probes selected by the S. group, 100% were hypomethylated, and from the total of 7085 probes selected by the W. group, almost all were hypomethylated (n=7038, 99.33 %) with a minimal amount hypermethylated (n=47, 0.67%). 108 probes were in common with the same methylation status (hypomethylated). The DMCs genic distribution was similar. Most DMCs were found in the body (∼30–36%), followed by IGRs (∼24–26%), promoters (21–32%), and 1to5kb regions (∼14–17%), and minimal amounts in UTRs (∼3–4% in 5UTRs and ∼2–3% in 3UTRs) (Tables S9 and S10). About CGI context, most DMCs were found in shores (∼36–46%). In the case of S., the rest were mostly found in CGIs (n=34, 30.86%), inter CGIs (n=16, 14.29%) and shelves (n=11, 9.82%). In the W. group, the rest were mostly found in inter CGIs and CGIs in similar proportions (n=2071, 29.23% and n=2030, 28.65%, respectively), and in less amount in shelves (n=431, 6.08%) (Tables S11 and S12). All chrs had DMCs in W. esig, but they weren't found in chrs 18 and 21 in S. esig. They were more prevalent in chrs 1 and 17 in S. esig and in chr 1 in W. esig. The higher difference was found in chr 9 (higher peak in S.) and 11 (higher peak in W.) (Figures S1e and S1f). In W. esig, DMCs covered many loci in each chr. S. esig had less amount of DMCs but they were in close loci with the DMCs from W. esig (Figures S4 and S5).

From this first analysis, it was possible to conclude that the distribution of DMCs per genomic context was similar between the same syndrome esigs in all cases, with a bit of difference in Kabuki esigs, in regards to 1to5kb and promoter regions. In all cases, the gene body had the major fraction of DMCs. That corresponded to close to half of DMCs in CHARGE and Kabuki and ∼30–36% in Sotos. That was followed by IGRs. Promoter DMCs corresponded to ∼10–13% for CHARGE and Kabuki, and a bit more in Sotos (∼21–32%). In all cases, minimal amounts of DMCs were distributed in UTRs. This concludes that the gene body was the most affected by methylation, which in Sotos was related to overall enhanced expression since they were all hypomethylated. In CGI context, DMCs distribution was similar for Kabuki, whilst in CHARGE and Sotos, it was a bit distinct. In Kabuki and Sotos, DMCs were more heavily distributed in shores (∼41–46% in Kabuki and ∼36–46 in Sotos), and in CHARGE, they were heavier in inter CGIs (∼49–59%), followed by shores (∼21–28%). Shelves presented the less amount of DMCs in all cases. This concludes that CGIs were less affected by differential methylation, however their closest regions (shores) were largely affected. In CHARGE, the farthest CGI regions (inter CGIs) were the most affected. The comparison of hyper/hypo methylated CpGs ratio was similar in all same syndromes esigs, except for CHARGE, where there was a clear difference. The distribution per chr and within the chr was distinct in all cases. Finally, in all cases, there was a small fraction of DMCs in common, except for Sotos, where almost all DMCs in S. were in common with W. (108/112).

3.2 Enrichment analysis of biological processes of differentially methylated genes between same syndrome episignatures.

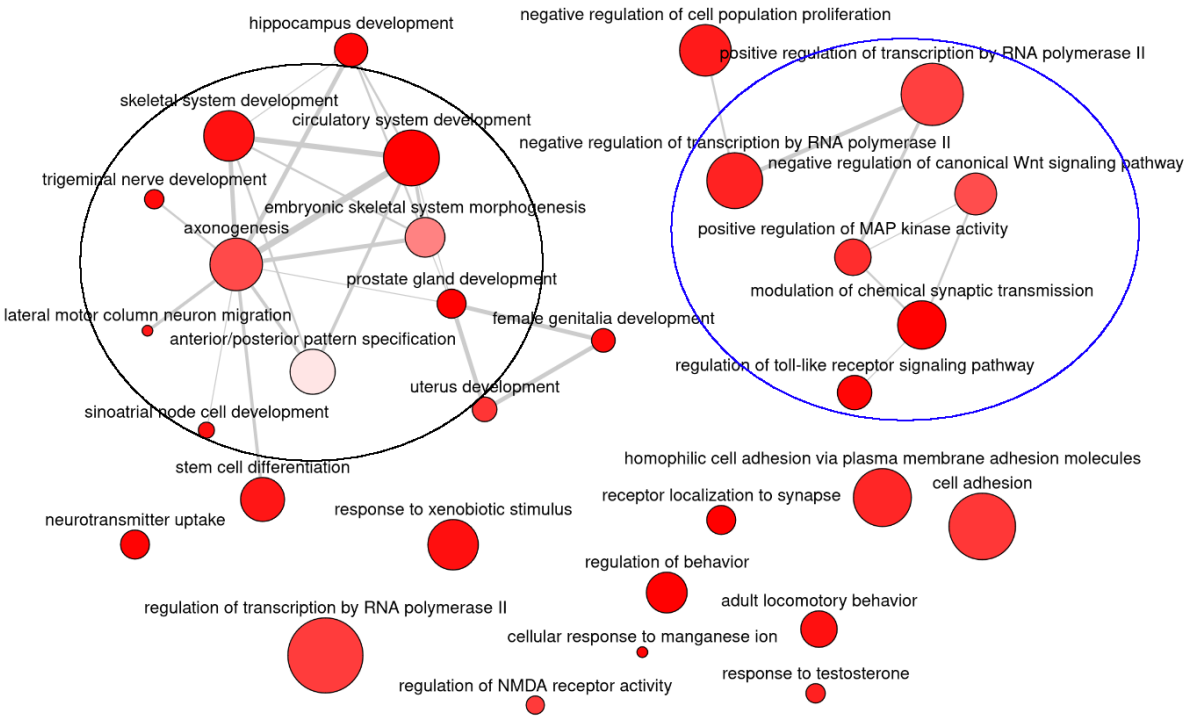
For both CHARGE and Kabuki esigs, there was a small set of DMCs in common. In Sotos, all of DMCs of S. model were present in W. model as well, however, W. model presented a much higher number of DMCs (112 *vs.* 7,805). To better understand the functional consequence of the methylation patterns, DMCs were annotated to genes. The total number of unique genes covered by DMCs in each esig is

present in Table S13. To question which biological processes (BPs) were the differentially methylated genes (DMGs) mostly implicated, an enrichment analysis (EA) was performed using DAVID, and then redundant terms were removed using REVIGO with similarity cutoffs. The cutoff provided removed the maximum amount of semantically similar GO terms by calculating the degree of frequency, uniqueness and dispensability between each pair of terms. Frequency corresponded to the proportion of a single GO term in the EBI GOA database for the selected species (Homo Sapiens 9606) thus higher frequencies corresponded to more general terms and minor frequencies corresponded to more specific terms. Dispensability corresponded to the semantic threshold that REVIGO considered to remove or not the GO term and assign it to a cluster representative. In cluster representatives, this value was always less than the cutoff chosen and was always less than the dispensability values of the terms it was representing (less dispensability meant the term was less dispensable and should be considered as a non-redundant term). REVIGO also provided graphical representations such as interactive graphs. The interactive graph displayed the connection between the set of GO terms based on the GO hierarchy. The colour of the bubbles corresponded to the p-values provided along with the GO terms in the DAVID output. Less red values had minor p-values. The size of the bubbles corresponded to the log10 (number of annotations for the GO Term in the selected species in the EBI GOA database). The bubble representing a non-redundant term was connected by a line to another term based on the semantic similarity of the two. The width of the line corresponded to the degree of similarity (more thickness meant more similarity). BPs involved in similar BPs were clustered together.

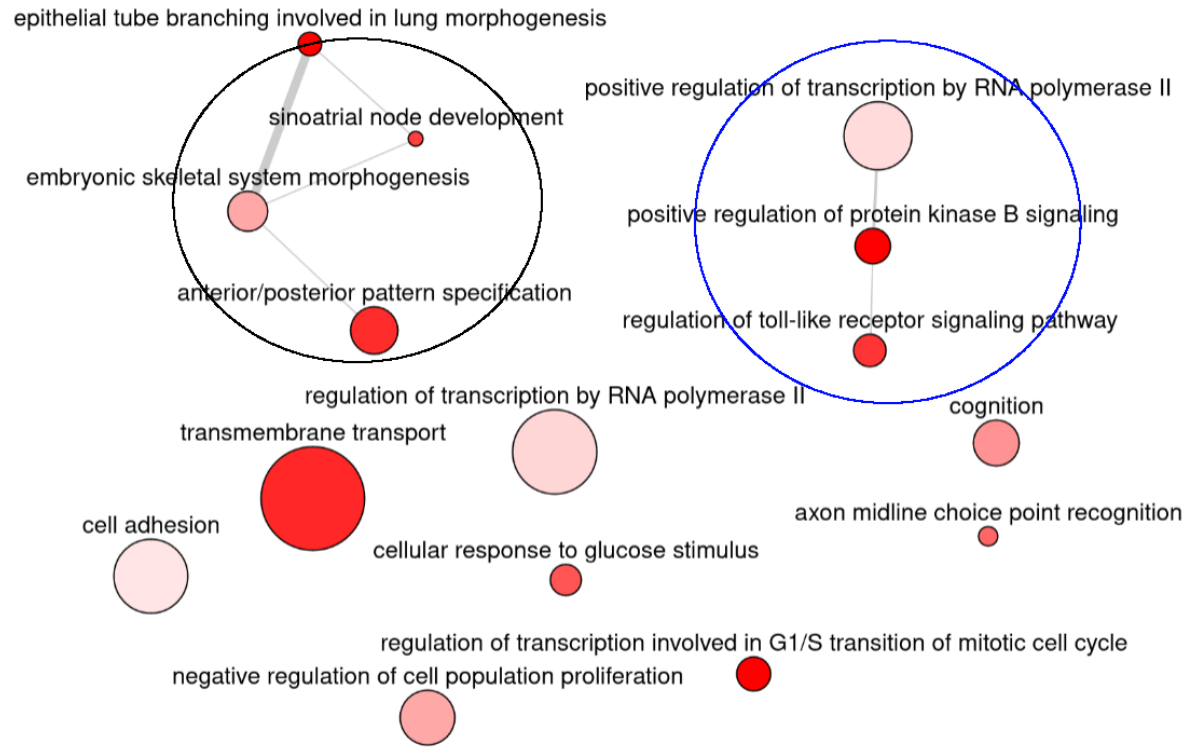
In **CHARGE**, the BP EA using DAVID had 44 enriched BPs for the S. group and 18 for the W. group. REVIGO reduction with cutoff=0.7 gave 42 enriched BPs for S. esig (Table S14) and kept the same 18 for W. esig (Table S15). In both cases, the interactive graph showed 2 different clusters that could be identified as developmental processes (black) which could all relate to CS - development of the embryonic skeletal system, heart related/ circulatory system and anteroposterior (AP) pattern specification - and regulation processes (dark blue), including regulation of transcription, cell proliferation and toll-like receptor pathway. Others included cell adhesion (Figure 3.1). S. group provided more detailed results since it also included genitalia and cranial nerve (trigeminal nerve) development, which are related to CS.

In **Kabuki**, the BP EA using DAVID had 11 enriched BPs for the S. group and 14 for the W. group, and none was in common between the two esigs. REVIGO reduction with cutoff=0.7 identified all of the BPs as non-redundant (Tables S16 and S17 for S. and W. groups, respectively). In S., BPs clustered into a cluster composed of lipid metabolism/ transport and regulation of transcription-related BPs (orange), and another of phagocytosis-related BPs (green) (Figure 3.2a). In W., BPs clustered into developmental (black) and signalling processes (dark blue) (Figure 3.2b). W. group seemed to provide better results since it included development processes of AP specification, angiogenesis and embryonic skeletal system morphogenesis, which can relate to KS.

In **Sotos**, the BP EA using DAVID had 9 enriched BPs for the S. group and 235 for the W. group. In the S. group, BPs were further provided to REVIGO with a cutoff=0.7 which classified all 9 enriched BPs as non-redundant (Table S18). In the W. group, the cutoff chosen was 0.5 due to the high number of enriched GO BPs provided by DAVID. REVIGO classified a total of 107 non-redundant GO terms (Table S19). In S., BPs clustered into fibroblast growth and migration (pink) and morphogenesis of branching structures-related processes (grey). Others included cell proliferation, nucleosome assembly and regulation of MAPK cascade (Figure 3.3a and Table S18). In W., BPs clustered into regulation processes (black circle), development processes (dark blue), transport-related processes (orange), metabolism pro-

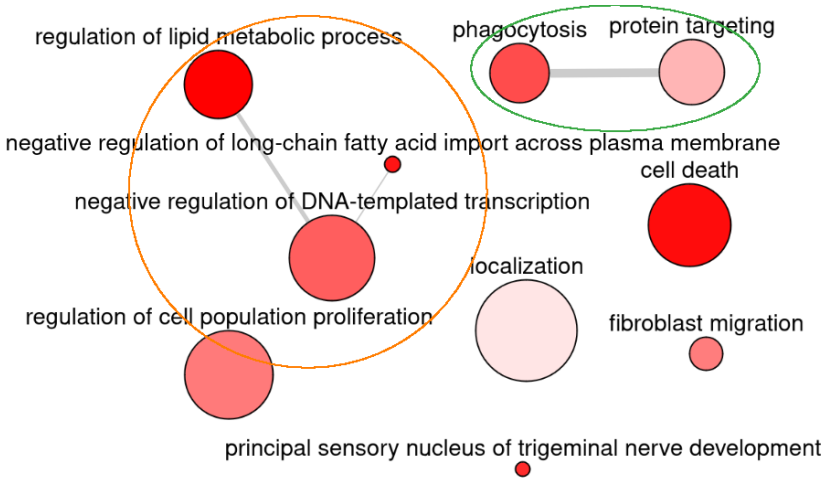


(a) Sadikovic episignature

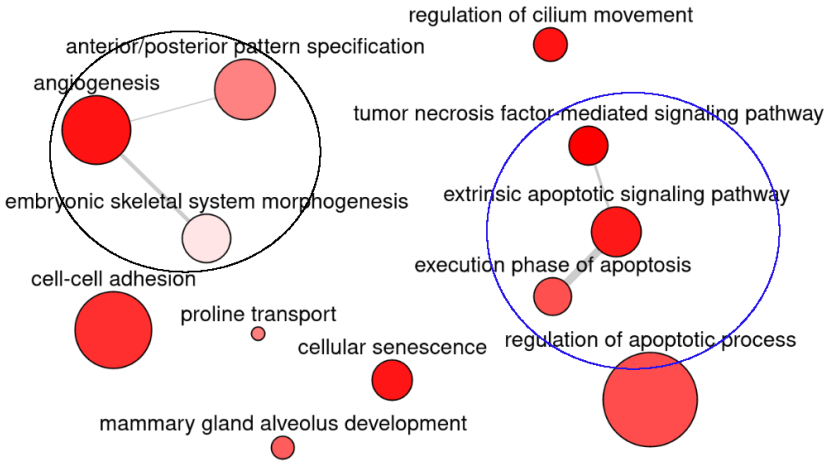


(b) Weksberg episignature

Figure 3.1. Interactive graph using REVIGO which plots all non-redundant enriched biological processes (BPs) for the differentially methylated genes in the episignature for **CHARGE** syndrome by (a) Sadikovic, (b) Weksberg (c) common genes in the two groups. Highly similar BPs are linked by edges and line width corresponds to the degree of similarity. Bubbles with lighter red colour have lower p-values. Bubbles with bigger sizes have a higher number of annotations in the EBI GOA database. The size corresponds to the Log10 values. The black circle highlights enriched developmental processes. The dark blue circle highlights enriched regulation processes.



(a) Sadikovic episignature



(b) Weksberg episignature

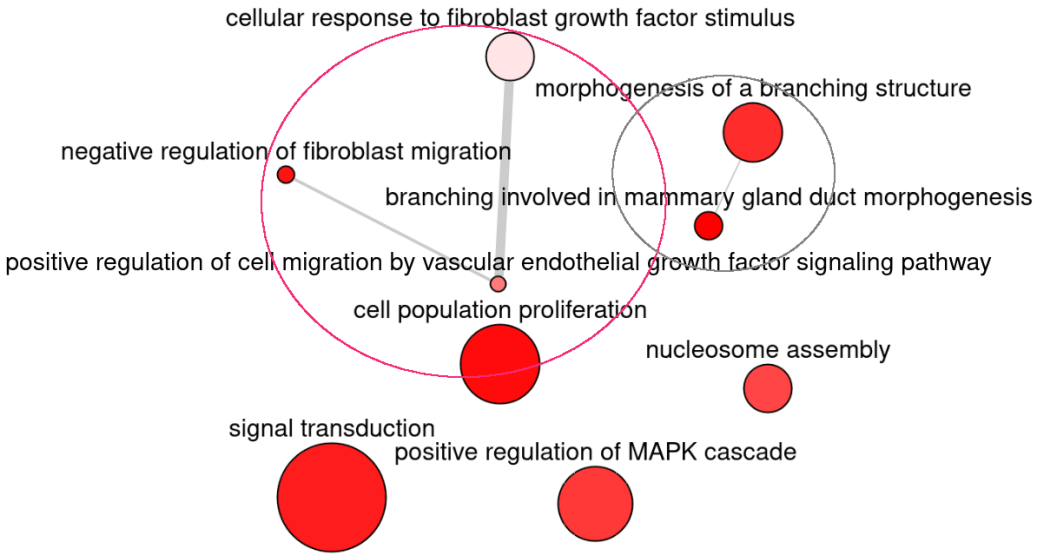
Figure 3.2. Interactive graph using REVIGO which plots all non-redundant enriched biological processes (BPs) for the differentially methylated genes in the episignature for **Kabuki** syndrome by (a) Sadikovic, (b) Weksberg (c) common genes in the two groups. Highly similar BPs are linked by edges and line width corresponds to the degree of similarity. Bubbles with lighter red colour have lower p-values. Bubbles with bigger sizes have a higher number of annotations in the EBI GOA database. The size corresponds to the Log10 values. The orange circle highlights the regulation of lipid metabolism/transport processes and the regulation of transcription. The green circle highlights phagocytosis-related processes. The black circle highlights enriched developmental processes. The dark blue circle highlights enriched signalling pathways related to apoptosis.

cesses (light blue) and signalling pathways (light green) (Figure 3.3b and Table S19). W. esig seemed to provide the best results since it had several important development processes related to SS.

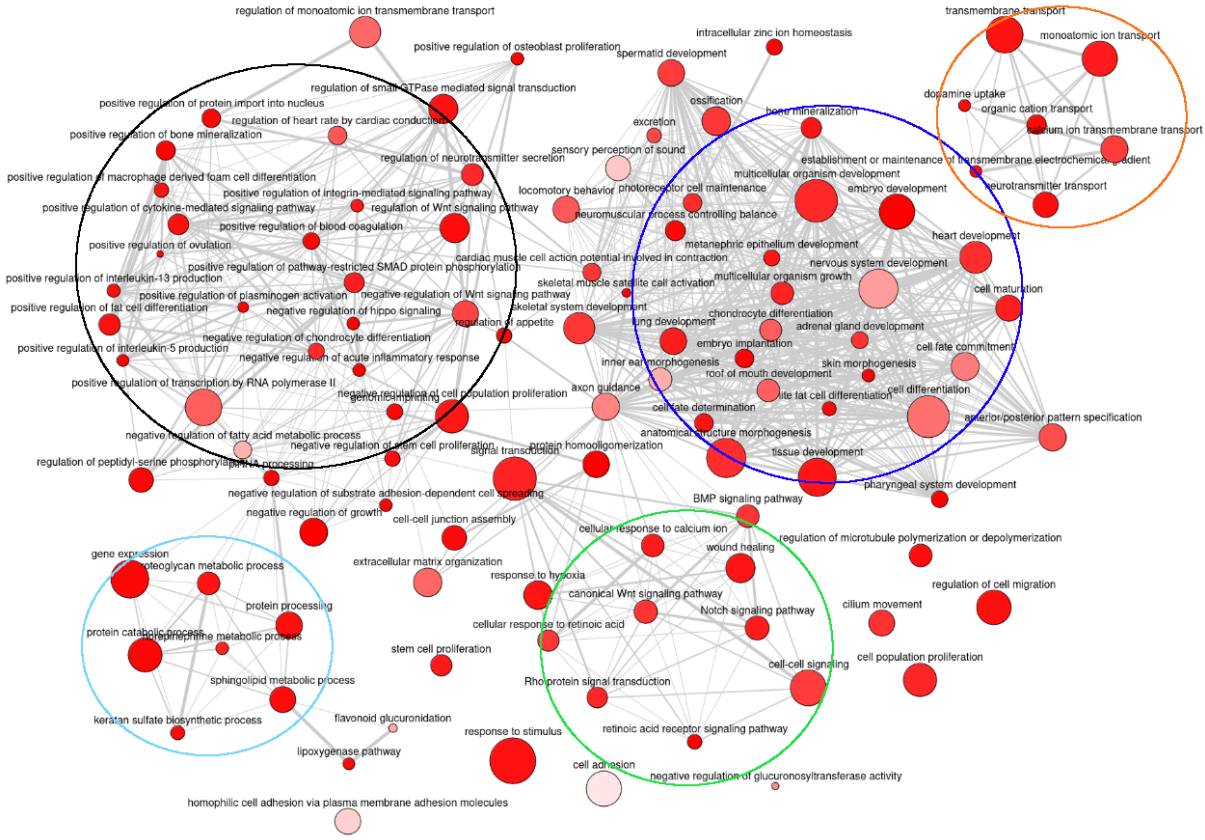
3.3 Comparison of the top differentially methylated probes between same syndrome episignatures

The EA was performed to assess which BPs were significantly overrepresented by comparing the set of DMGs to the total number of genes in the database and the expected distribution of genes involved in a BP that is expected by chance. For this reason, EA can only go so far since it doesn't take into consideration the genes with higher differences in methylation compared to the rest. Other BPs involving lesser genes can also be relevant, especially if those genes had higher differences in methylation (higher absolute $\Delta\beta$ values). In this context, the top DMCs with higher $\Delta\beta$ values and gene correspondence were considered.

In **CHARGE**, from the total of 148 probes in S. esig, 98 (66.22%) of them were assigned to 96 unique genes, and from the total of 75 probes selected in W. esig, 56 (74.67%) of them were given to 58 unique



(a) Sadikovic episignature



(b) Weksberg episignature

Figure 3.3. Interactive graph using REVIGO which plots all non-redundant enriched biological processes (BPs) for the differentially methylated genes in the episignature for **Sotos** syndrome by (a) Sadikovic, (b) Weksberg (c) common genes in the two groups. Highly similar BPs are linked by edges and line width corresponds to the degree of similarity. Bubbles with lighter red colour have lower p-values. Bubbles with bigger sizes have a higher number of annotations in the EBI GOA database. The size corresponds to the Log10 values. The pink circle (upper image) highlights enriched processes involved in fibroblast growth/migration and cell proliferation. The grey circle (upper image) highlights branching structures morphogenesis processes. The black circle highlights enriched regulation processes. The dark blue circle highlights enriched developmental processes. The orange circle highlights enriched transport processes. The light blue circle highlights metabolism processes. The light green highlights signalling pathways.

genes. The methylation status of those genes is found in Tables S20 and S21 for S., and Tables S22 and S23 for W. Within the top 5 hyper-methylated CpGs, there were the in-common genes HOXA5, HOXA-AS3 and VWF, and within the top 5 hypo-methylated CpGs there were HOTAIRM1, HOXA1, LINC02914 and SLITRK5 (Table 3.1). Full tables are in the Appendix (Tables S31, S32, S33, and S34).

Table 3.1. Top 5 hypermethylated and hypomethylated probes with gene correspondence and higher absolute $\Delta\beta$ values, for both Sadikovic and Weksberg episignatures of CHARGE syndrome.

Probes	**Genes**	**Chr**	$\Delta\beta$	**genomic_features**	**CGI annot.**
Top 5 hypermethylated probes in Sadikovic episignature					
cg04863892	HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3	chr7	0.208	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg04053108	VWF	chr12	0.181	Body	CGIs
cg09549073	HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5	chr7	0.175	1to5kb, 1to5kb, 5UTRs, 5UTRs, Body, Body	CGIs
cg09319828	TTC24	chr1	0.147	Body	Inter
cg26229043	KIAA1217	chr10	0.136	Body	Inter
Top 5 hypermethylated probes in Weksberg episignature					
cg04053108	VWF	chr12	0.203	Body	CGIs
cg18274664	APP	chr21	0.150	Body	Inter
cg04392082	SORBS2, SORBS2, SORBS2	chr4	0.136	Promoters, 5UTRs, Body	Inter
cg26023912	HOXA-AS3, HOXA5, HOXA-AS3, HOXA5	chr7	0.124	1to5kb, 1to5kb, Body, Body	CGIs
cg16923485	SLCO1A2	chr12	0.111	Body	Inter
Top 5 hypomethylated probes in Sadikovic episignature					
cg08941355	HOTAIRM1, HOXA1, HOTAIRM1, HOXA1, HOTAIRM1, HOXA1	chr7	-0.163	1to5kb, 1to5kb, Body, Body, 3UTRs, 3UTRs	shores
cg06906435	LINC02914	chr14	-0.130	Promoters	Inter
cg06602723	HOXB7, HOXB8	chr17	-0.120	1to5kb, 1to5kb	shores
cg16787483	SLITRK5	chr13	-0.117	Body	shores
cg16915863	LINC02381	chr12	-0.111	Body	shelves
Top 5 hypomethylated probes in Weksberg episignature					
cg07318204	HHIP, HHIP-AS1, HHIP, HHIP-AS1	chr4	-0.187	Promoters, Promoters, Body, Body	CGIs
cg21090457	ROBO2	chr3	-0.157	Body	Inter
cg08941355	HOTAIRM1, HOXA1, HOTAIRM1, HOXA1, HOTAIRM1, HOXA1	chr7	-0.154	1to5kb, 1to5kb, Body, Body, 3UTRs, 3UTRs	shores

Table 3.1 continued from previous page

cg06906435	LINC02914	chr14	-0.150	Promoters	Inter
cg16787483	SLITRK5	chr13	-0.150	Body	shores

In **Kabuki**, of the total of 153 probes in S., 132 (86.27%) were assigned to 168 unique genes, and from the total of 113 probes in W., 88 (77.88%) of them were given to 106 unique genes. Genes methylation status is in Tables S24, S25, S26 and S27. Within the top 5 hyper-methylated CpGs, there was the in-common gene ZMIZ1, and within the top 5 hypo-methylated CpGs there were AGAP2-AS1, AGAP2 and TNS1 (Table 3.2). Full tables are in the Appendix (Tables S35, S36, S37, and S38).

Table 3.2. Top 5 hypermethylated and hypomethylated probes with gene correspondence and higher absolute $\Delta\beta$ values, for both Sadikovic and Weksberg episignatures of Kabuki syndrome.

Probes	**Genes**	**Chr**	$\Delta\beta$	**genomic_features**	**CGI annot.**
Top 5 hypermethylated probes in Sadikovic episignature					
cg20744163	ZMIZ1, ZMIZ1	chr10	0.194	1to5kb, Body	shelves
cg18322589	TACC2	chr10	0.150	Body	Inter
cg21721340	SOX18	chr20	0.149	1to5kb	shores
cg21637392	RNF216, RNF216	chr7	0.145	1to5kb, Body	Inter
cg10213542	ADAMTS2	chr5	0.139	Body	Inter
Top 5 hypermethylated probes in Weksberg episignature					
cg01246520	RAI1	chr17	0.298	Body	Inter
cg08818610	RIPOR2	chr6	0.262	Body	CGIs
cg20744163	ZMIZ1, ZMIZ1	chr10	0.238	1to5kb, Body	shelves
cg05463589	IL17C	chr16	0.232	Body	CGIs
cg23061725	CASP8, CASP8	chr2	0.208	1to5kb, Body	Inter
Top 5 hypomethylated probes in Sadikovic episignature					
cg01178624	KCNK7, KCNK7	chr11	-0.236	Body, 3UTRs	CGIs
cg20225999	TNS1	chr2	-0.213	Body	shores
cg14845962	AGAP2, AGAP2-AS1, AGAP2, AGAP2-AS1	chr12	-0.212	Body, Body, 3UTRs, 3UTRs	CGIs
cg24869272	TSPAN4, TSPAN4	chr11	-0.155	5UTRs, Body	CGIs
cg18587137	TNFAIP2, TNFAIP2	chr14	-0.151	5UTRs, Body	CGIs
Top 5 hypomethylated probes in Weksberg episignature					
cg16370398	FLJ12825, HOXC4, FLJ12825, HOXC4	chr12	-0.214	1to5kb, 1to5kb, Body, Body	shores
cg13541527	SNORD52, SNHG32, SNORD52, SNHG32	chr6	-0.207	Promoters, Promoters, Body, Body	shores
cg20225999	TNS1	chr2	-0.202	Body	shores
cg23387569	AGAP2-AS1, AGAP2, AGAP2-AS1, AGAP2, AGAP2-AS1, AGAP2	chr12	-0.200	Promoters, Promoters, Body, Body, 3UTRs, 3UTRs	CGIs

Table 3.2 continued from previous page

cg12474798	ADO, ADO	chr10	-0.195	Body, 3UTRs	CGIs

In **Sotos**, of the total of 112 probes in S., 73 (65.18%) were assigned to 91 unique genes, and from the total of 7085 probes selected by W., 4684 (66.11%) were assigned to 2456 unique genes. Genes methylation status is in Tables S28, S29 and S30. Within the top 5 hypo-methylated CpGs, there were the in-common genes PCAT6 and KDM5B (Table 3.3). Full tables are in the Appendix (Tables S39 and S40).

Table 3.3. Top 5 hypermethylated and hypomethylated probes with gene correspondence and higher absolute $\Delta\beta$ values, for both Sadikovic and Weksberg episignatures of Sotos syndrome.

Probes	**Genes**	**Chr**	$\Delta\beta$	**genomic_features**	**CGI annot.**
Top 5 hypomethylated probes in Sadikovic episignature					
cg21105875	PCAT6, KDM5B	chr1	-0.650	1to5kb, 1to5kb	shores
cg07816556	H3C1, H4C1, H1-1, H3C1, H4C1, H1-1, H3C1, H4C1, H1-1	chr6	-0.559	1to5kb, 1to5kb, 1to5kb, Body, Body, Body, 3UTRs, 3UTRs, 3UTRs	shelves
cg15477144	FAM118A	chr22	-0.502	Promoters	shores
cg06154311	RBBP8NL	chr20	-0.494	Promoters	CGIs
cg25968569	KCTD12	chr13	-0.490	Promoters	shores
Top 5 hypomethylated probes in Weksberg episignature					
cg21105875	PCAT6, KDM5B	chr1	-0.668	1to5kb, 1to5kb	shores
cg07600533	KLHDC7B	chr22	-0.661	Promoters	CGIs
cg03133378	TMEM225B, TMEM225B	chr7	-0.650	5UTRs, Body	CGIs
cg25807487	MRGPRF, MRGPRF	chr11	-0.642	1to5kb, Body	shelves
cg09684846	TVP23A	chr16	-0.636	Promoters	CGIs

3.4 Differentially methylated genes involved in developmental pathways

Finally, common DMGs involved in developmental pathways were considered for bibliography analyses, since the three syndromes discussed in this thesis are due to problems in developmental stages. It was assumed that the $\Delta\beta$ value of a DMC in a gene was a measure of the differential methylation of the gene. For genes having more than one DMC, the mean values of all DMCs were selected as the correspondent $\Delta\beta$ measure of the gene differential methylation. Only the common genes between the same syndrome esig were considered. Common genes with mean $\Delta\beta$ are in Tables S41, S42 and S43 for CHARGE, Kabuki and Sotos syndromes, respectively. The developmental pathways and respective genes are represented in Figures 3.4, 3.5 and 3.6. The $\Delta\beta$ values shown are a mean of the $\Delta\beta$ values assigned to each gene in respective esig. In **CHARGE**, all of the genes presented in Figure 3.4 seemed to be involved in important developmental pathways for CS. In **Kabuki**, all of the genes seemed to be involved as well in KS manifestations except for RPS6 since it was involved only in placenta development (Figure 3.5). In **Sotos** as well, all of the genes seemed to be related to SS, except for RD3 which was only involved in sensory organ and system development (Figure 3.6).

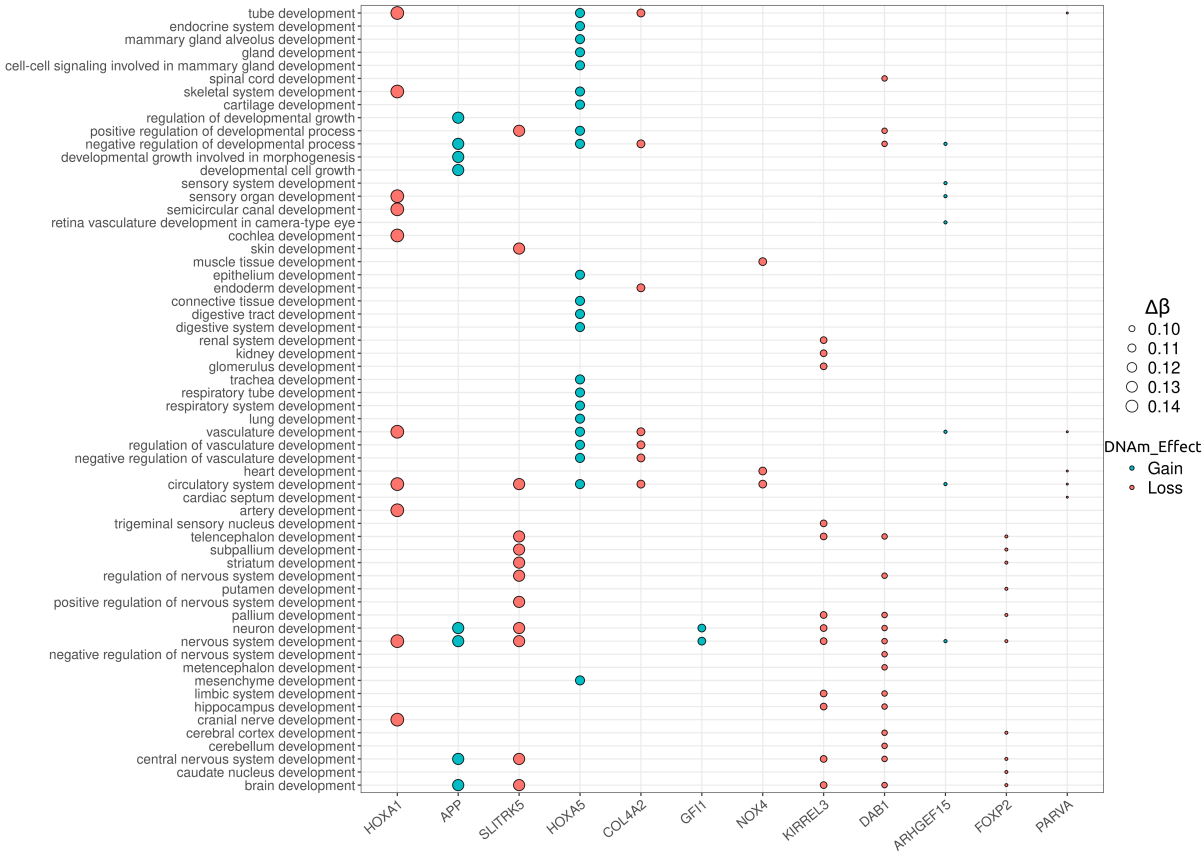


Figure 3.4. Genes with the highest differential methylated CpGs and involved in developmental processes for each CHARGE-signature. The 'DNAm_Effect' corresponds to whether the CpG was hypermethylated (Gain) or hypomethylated (Loss). The $\Delta\beta$ values presented are the mean values from the $\Delta\beta$ values of those probes in Sadikovic's and Weksberg's episignatures.

IV. Discussion

Single-gene NDDs aborded in this thesis were shown to have distinctive and specific methylation patterns caused by the mutation in specific causative genes. Two esigs were analysed for each of the syndromes CHARGE, Kabuki and Sotos, built by S. and W. research groups, and the focus was to compare them in their final set of selected probes with respect to genomic location and gene correspondence. The DMGs set was further investigated in their play in biological pathways, mainly those of developmental stages, and it was investigated their implicance in the disorder manifestations. The number of DMCs selected in the building of esigs can vary a lot for different models, which was the case in this comparative study, with the most notable difference in the Sotos esig, in which S. selected 112 CpGs, whilst W. selected 7,085 DMCs. Besides, $\Delta\beta$ values had different cutoffs for each research group, thus one of the aims was also to see if the choice of those values led to distinct conclusions regarding the relatability with the mechanisms of disease. Sotos esigs were the only ones with similar absolute $\Delta\beta$ ranges between groups.

In respect to **CHARGE**, the esigs had only 21 common DMGs, but despite that, the outcome of BPs EA produced similar results. Those BPs could be generally grouped into processes related to development, regulation or cell adhesion. Developmental processes included the **embryonic development of skeletal and circulatory systems**, the **brain development** and **AP pattern specification**. The improper functioning of all of those can relate to CS. Scoliosis is the main CS skeletal problem mainly due to hypotonia, while less common anomalies include vertebral fusions, hemivertebrae or abnormal number of ribs [132] [133]. Heart defects are part of the less specific features of CS with $\sim$ 75% frequency, including defects in the outflow tract, atrioventricular septum, and aortic arch [134] [135]. CS patients can also exhibit

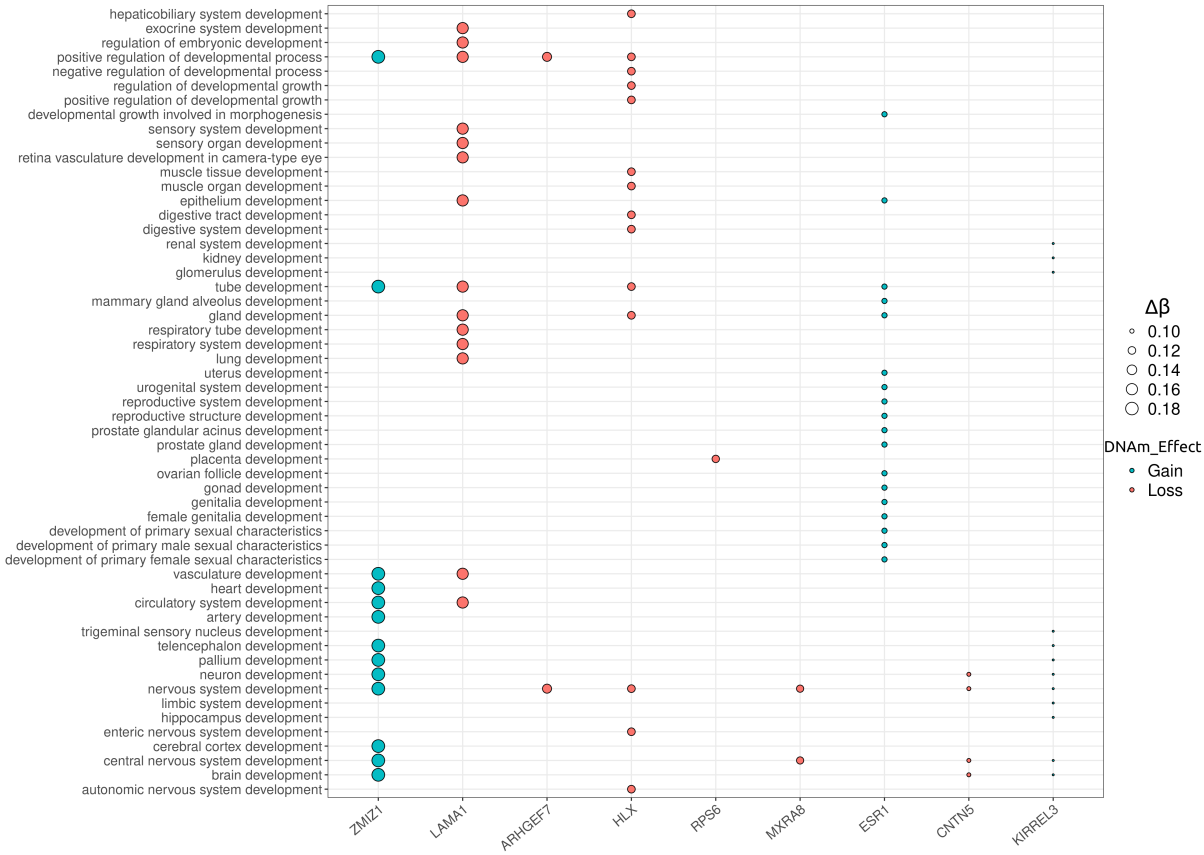


Figure 3.5. Genes with the highest differential methylated CpGs and involved in developmental processes for each Kabuki episignature. The 'DNAm_Effect' corresponds to whether the CpG was hypermethylated (Gain) or hypomethylated (Loss). The $\Delta\beta$ values presented are the mean values from the $\Delta\beta$ values of those probes in Sadikovic's and Weksberg's episignatures.

persistent ductus arteriosus, where the oxygenated blood flows back into the lungs [134]. CS brain is underdeveloped, showing several intracranial anomalies, including hypoplasia of the semicircular canal [136], cranial nerves [136] [137], basal skull [136][138], cerebellum [136][139] [140], brainstem [136] and frontal lobe [141]. The CS brain ventricles were also found enlarged by CSF excessive buildup (hydrocephalus) [136]. Moreover, another study reported that CHD7 depletion causes aberrant brain folding in mice, similar to what is described in CS [142]. AP axis consists of the axis going along the spinal cord. Problems in its establishment might explain the CS scoliosis. Enriched regulation BPs included **regulation of transcription**, **cell proliferation** and **toll-like receptor (TLR) signalling**. CHD7 is a chromatin remodeler that regulates nucleosome assembly via ATP hydrolysis thus controlling transcription [143]. CHD7 haploinsufficiency increases neural stem cell (NSC) proliferation [144]. Concerning TLRs, those play in the innate immune response. However, CS immunodeficiency is mainly in the adaptive immune response [145]. Finally, **cell adhesion** enriched BPs relate to defects in neural crest cells (NCCs) migration and adhesion seen in CS [146]. Those stem cells start developing in the border of the neural plate induced by signals of neighbouring tissues [147] [148]. Changes in their morphology and adhesion allow them to detach from the neural tube [149] [150] and migrate to embryonic tissues where they differentiate [148]. A 2023 study also reported a significant difference in the expression of cell-adhesion genes [151]. Although BP EA gave similar results for the different models, it is fair to say that the S. gave more detailed results including **genitalia** and **cranial nerve development**. Hypogonadism is a CS feature, and cranial nerve development abnormalities make part of the main CS features [106].

Concerning the genes with CpGs with higher $\Delta\beta$ values, **HOXA5** and **HOXA-AS3** were hypermethylated. Homeobox (HOX) genes are TFs that regulate embryo axial patterning [152], and HOX methy-

Figure 3.6. Genes with the highest differential methylated CpGs and involved in developmental processes for each Sotos episignature. The 'DNAm_Effect' corresponds to whether the CpG was hypermethylated (Gain) or hypomethylated (Loss). The $\Delta\beta$ values presented are the mean values from the $\Delta\beta$ values of those probes in Sadikovic's and Weksberg's episignatures.

lation is crucial in regulating HOX activity post-natal. HOX genes are part of a large family organized into 4 chromosomal clusters, each containing up to 11 genes [153] [154]. Their transcription during embryonic development follows a temporal and spatial colinearity from the 3' to the 5'end along the cluster [155]. Each cluster has unique HOX combinations, which determines the cell's placement along the AP axis and within the hindbrain [156] [157] [158], limbs [159] [160] and genitalia [161] during embryo development. Hypermethylation of promoter, 1to5kb and body regions of HOXA5 was found in previous studies in both CS and KS, and may justify part of their clinical overlap [162]. Furthermore, abnormal HOXA5 expression is related to congenital diaphragmatic hernia (CDH) [163], a very rare condition seen in CS [164], associated with pulmonary hypoplasia. HOXA5 is necessary for axonal growth, synaptogenesis, establishment of neural circuits and synaptic plasticity [165]. HOXA-AS3 is a long non-coding RNA (lncRNA) produced in HOXA cluster, and highly expressed in the embryo. As part of the lncRNAs of the HOX family, it is responsible for the colinear expression of HOX clusters [166]. In another CS study, HOXA-AS3 was found hypermethylated in CGIs in the body, along with HOXA5 in CGIs in the body and 1to5kb regions [167]. **HOTAIRM1**, **HOXA1**, **LINC02914** and **SLITRK5** were distinctly hypomethylated. HOTAIRM1 is a lncRNA that interacts with chromatin-modifying complexes [168], thus helping the temporal collinear activation of genes in HOXA cluster [169]. HOTAIRM1 also regulates integrin signalling, necessary for cerebral cortex development via expansion of the neural stem cells [170] [171]. This gene was differentially methylated in previous CS studies [124]. HOXA1 plays in neural crest specification, hindbrain patterning, and heart and ear development [172] [173] [174]. SLITRK5 is a transmembrane protein involved in CNS synapses, where it can make cis-interactions with the TRKB receptors upon BDNF stimulation [175], critical in neurogenesis in CNS development [176]. Both HOXA1

and SLITRK5 were found previously hypomethylated in CS studies [124]. LINC02914 is a lncRNA associated with the rare syndrome Bardet-Biedl (BBS) which has some features overlapping CS namely hypogonadism, renal and cardiovascular anomalies, and learning disabilities [177] [178].

Of the 21 common genes, 12 were involved in developmental pathways. Those included **HOXA1**, **SLITRK5** and **HOXA5**, highly differentially methylated. Other developmental genes are worth mentioning. **APP** plays in neurogenesis, neuron differentiation, growth and migration, synaptogenesis and cell adhesion [179]. APP was found hypermethylated in body in previous CS studies [167]. **COL4A2** pathogenic variants relate to vasculature defects in the brain [180]. **GFI1** is a TF with a mostly repressing effect. It enters several developmental processes, of interest, differentiation of hair cells and survival of cochlear ganglion neurons in the inner ear, relating to CHD7 phenotype. Defects in those cause deafness in CS [181]. As well, it plays in commitment of B and T cells in the adaptive immune system, sometimes defective in CS [182] [183] [184] [185]. **NOX4** is constitutively active in cardiomyocytes, endothelial cells, fibroblasts, and vascular smooth muscle cells. It is also expressed highly in kidney proximal tubular cells [186] [187]. **KIRREL3** is highly expressed in CNS neurons, including the limbic system, and the cerebellum, mediating synapses. Pathogenic variants relate to ID, cerebellar hypoplasia, and behavioural anomalies [188]. **DAB1** is a cytoplasmic adaptor protein involved in cerebral cortex development, crucial for cognitive processes [189]. **ARHGEF15** enhances eye retina neovascularization [190] which can relate to CS coloboma. **FOXP2** plays in craniofacial development including facial muscles and striatal brain. FOXP2 mutations relate to speech and language deficits, also present in CS [191] [192]. Moreover, CHD7 is known to bind to FOXP2 promoter in embryonic stem cells (ESCs) and NPCs [193]. **PARVA** enters in sarcomere organization and contraction of smooth muscle cells, being crucial for vascular development and the cardiac septum [194].

Concerning **Kabuki**, 44 DMGs were in common. BPs EA gave quite different results between esigs, being that W. provided better results. In S., BPs could be divided into processes related to **lipid metabolism and transport across the membrane**, **regulation of transcription** and **phagocytosis**-related BPs. Moreover, it also included **cell death**, **proliferation** and the **development of the principal sensory nucleus of the trigeminal nerve**. Studies of hepatic steatosis and pancreatic cancer proposed that KMT2D targets gene sequences involved in lipid metabolism, by PPARγ receptor [195].[196]. KMT2D tri methylates H3K4, associated with active gene transcription [111] [112] so the dysregulation of transcription processes was expected. Moreover, KMT2D LOF was reported to disrupt the cell cycle, proliferation and survival in both mice and human KS1 models, with higher proportions of cell apoptosis in induced pluripotent stem cells (iPSCs) and NPCs in the KS1 model *vs.* controls [197]. KS patients can have microcephaly and several structural brain anomalies such as atrophy of the cerebellum and the brainstem [198] [199], which englobes the sensory and motor nuclei of the trigeminal nerve (cranial nerve V). On the other hand, in W., enriched BPs consisted mainly of **apoptosis**-related signalling pathways and several developmental processes, such as those related to the **embryonic skeletal system**, **AP pattern specification**, **development of the mammary gland** and **angiogenesis**. It also included **proline and leucine transport**, **cell adhesion**, **cellular senescence** and **regulation of cilium movement**, which is a signalling-related process. KMT2D mutation in KS was suggested to increase apoptotic processes, slow growth and retard development [200]. Furthermore, several skeletal anomalies have been described in KS [67]. Altered AP specification may justify KS scoliosis [119]. KMT2D is critical in the regulation of several HOX genes, namely the HOXC6 which is essential in mammary gland development. KS patients can present premature thelarche and precocious puberty caused by low hypothalamic sensitivity to the suppressive effects sent by gonadotropins [201] [202] [203] [204]. A 2019 study using zebrafish with

KMT2D null mutants (KS model) revealed the critical role of KMT2D mutation in causing cardiac defects by impacting vasculogenesis and angiogenesis. Angiogenesis issues included abnormal aortic arch development, hyperactive ectopic blood vessel sprouting, and aberrant brain vascular plexus patterning [205]. L-proline is heavily found in CNS where it acts in glutamatergic transmissions at hippocampal synapses [206]. Abnormally L-proline concentrations in humans relate to impairment of cognition [207] [208]. Mice with L-proline transporter PROT defects exhibited altered expression of glutamate transmission-related proteins in cortex and thalamus synapses. This led to reduced locomotion, motivation, and memory extinction issues [209]. Leucine (Leu) uptake through the LAT1 transporter in the mammalian brain activates the TORC1 target in a pathway impacting protein synthesis, brain excitability, and neuron plasticity [210]. Leu also influences gene transcription and protein synthesis through mTOR-dependent and independent pathways. Additionally, it triggers insulin secretion in pancreatic β cells by activating glutamate dehydrogenase [211], thus, problems in Leu transport may correlate with the rare hypoglycemia observed in KS [120]. KMT2D generally activates transcription being that KMT2D KO downregulates genes involved in cell adhesion [212]. Senescent cells are present in the hindbrain, midbrain, pronephros, mesonephros, neural tube, endolymphatic sac, pharyngeal arches, endoderm of the gut, olfactory epithelium, and nerve fascicles [213] [214]. Motile cilia localise on the surface of specialised epithelial cells in CNS. They aid CSF circulation in the brain [215]. Motile ciliopathies include hydrocephalus which is seen in KS patients with KDM6A mutations [216]. Ciliopathies englobing both motile and non-motile cilia include CHD which is found in 70% of KS patients [217] [218]. EA results of W. esig were coherent with another KS study, that showed enrichment of genes involved in embryonic organ and skeletal development/morphogenesis [219]. However, S. esig EA results did not relate much.

Considering the genes with CpGs with higher $\Delta\beta$ values, **ZMIZ1** was highly hypermethylated. ZMIZ1 is a TF that doesn't bind directly to DNA [220] and directly interacts with NOTCH1 receptor in brain formation and differentiation of neuronal and glial cells [221] [222]. ZMIZ1 is another candidate gene for which mutations can cause KS [223]. Individuals with ZMIZ1 variants and ID/DD syndromes were reported, with features including growth failure, feeding difficulties, microcephaly and facial dysmorphism. Furthermore, overexpression of ZMIZ1 variants in developing mice brains resulted in abnormal pyramidal neuron morphology, polarization and positioning, which supported the idea of those variants causing rare NDDs [223]. **TNS1** was highly hypomethylated in shores in the body, and **AGAP2** and **AGAP2-AS1** were highly hypomethylated in CGIs in promoters, body and 3UTRs. TNS1 codes for a protein localized in focal adhesions, a transmembrane linkage between the extracellular matrix and the cytoskeleton. TNS1 expression is high in the heart, kidneys, lungs, colon, prostate and testicles, but is low/unexisting in the brain thymus and in circulating leukocytes. Its abnormal expression causes numerous diseases, including cardiac diseases by disrupting downstream signalling pathways [224]. TNS1 hypomethylation was also present in patients with KMT2D LOF in another study, (also by Sadikovic), but in this case, it was present in CGIs [225]. In another study of WGBS of KS patients targetting CGIs, TNS1 was found to differentially methylated [226]. AGAP2 is widely expressed across several tissues and cellular compartments, and it participates in cellular organisation in the actin remodelling system, and focal adhesion [227] [228]. AGAP2 is overexpressed in several cancers, promoting cell proliferation and migration and inhibiting cancer cells apoptosis [229]. AGAP2-AS1 is an antisense lncRNA implied in several cancer types where it promotes cell proliferation and migration [230]. Both AGAP2 and AGAP-AS1 have been reported before to have several hypo-DMCs in KS patients in previous studies of both Sadikovic and Weksberg. Those were located in CGIs in the body and 1to5kb regions [225] [167].

Of the 44 common genes, only 9 were involved in developmental pathways, including **ZMIZ1**. **LAMA1**

is necessary for eye retina vasculature development [231] and maintenance of optic cup cell polarity and adhesion. Defects in this lead to disorganized retina [232]. This could relate to the loss of vision sometimes seen in KS [233]. LAMA1 was a DMG in a previous KS study also by Sadikovic [225]. **ARHGEF7** is fundamental in synaptogenesis and axons and dendrites development along the cortex and hippocampus. Its haploinsufficiency was reported in epilepsy and ID patients [234] [235]. **HLX** is a homeobox gene crucial for early enteric nervous system development that regulates gastrointestinal behaviour [236]. HLX polymorphisms are connected to fetal growth restriction (FGR) which highers the risk of contracting diseases and losing their life [237]. Short stature is a main KS feature [67] [238]. **RPS6** codes for a cytoplasmic ribosomal protein part of the 40S ribosomal subunit playing in cell growth and proliferation regulation. It was reported underexpressed in FGR placentas [239], thus it may also relate to KS postnatal growth deficiency. **MXRA8** codes for an immunoglobulin highly present in CNS, where it plays a role in the maturation of the blood-brain barrier [240]. It is also involved in cell adhesion and migration, and in hypoxia and TGF-β pathways, which promote angiogenesis [241] [242]. **ESR1** plays a crucial role in fetal human genital development. Disruptions in ESR1 are implicated in hypospadias, where the urethral opening is incorrectly located on the underside rather than the tip of the penis, which is found in some male KS patients [243] [244]. ESR1 binds to KMT2D [245]. In a previous KS study, ESR1 had six hyper-DMCs and one hypo-DMC [246]. **CNTN5** codes for a neuronal membrane participating in axonal targeting, and synaptic formation and plasticity. CNTN5 pathogenic variants are thought to increase the risk for ASD [247], very rarely reported in KS [248]. **KIRREL3** is an immunoglobulin (Ig) hypermethylated in inter CGIs in the body. It is expressed in both fetal and adult brains, and also in kidney glomeruli. KIRREL3 CNV or missense variants are associated with ASD and ID [249] [250].

Regarding **Sotos**, the BPs EA gave different outcomes between groups. In S., BPs were mainly connected to **fibroblast growth and migration**, **morphogenesis of branching structures**, and also **cell proliferation**, **nucleosome assembly** and **regulation of MAPK signalling cascade**. SS patients were found to have diminished activity of the MAPK/ERK signalling. This pathway is strongly affected by fibroblast growth factors (FGFs) differential expression [251]. Dysregulation of the pathway in SS alters the hypertrophic differentiation of NSD1-expressing chondrocytes, leading to stature overgrowth and faster skeletal maturation [252] [251]. It also dysregulates cell proliferation [253][254]. NSD1 methylates H4K20 and H3K36 [121] which alters the structure of the nucleosome and influences the DNA access to the transcriptional machinery. Branching morphogenesis are controlled by epigenetic mechanisms and cell adhesion processes [255]. In W., BPs could be clustered into processes involved in regulation, development, transport, metabolism and signalling. **Regulation** processes included regulation of interleukins production (13 and 5), protein import to the nucleus, canonical Wnt signalling, pathway-restricted SMAD protein phosphorylation, blood coagulation and plasminogen activation, hippo signalling, acute inflammatory response, transcription, cell proliferation (englobing stem cells), fatty acid metabolism, peptidyl-serine phosphorylation, and regulation of processes related with skeletal system (bone mineralization), neurotransmission and ovulation. Some of those are worth mentioning. Wnt signalling plays in skeletal development and bone homeostasis, and its dysregulation relates to several skeletal dysplasias, including SS overgrowth [256]. Abnormal expression of TGF-β and activin receptors in cardiac muscle can lead to cardiomyopathies [257]. Activin and BMP mediate bone formation [258]. The plasminogen activation system (PAS) catalyzes the cleavage of fibrinogen by thrombin, to dissolve blood clots. The Hippo pathway dysregulation is implied in cardiac [259] and renal diseases [260]. **Developmental** processes englobed development of the skeletal system of the skin, cardiovascular system , renal system , AP pattern specification, nervous system , lungs, inner ear, and spermatid. All these BPs relate to SS. In

SS, characteristics such as overgrowth make part of its cardinal manifestations [121]. This can be due to altered bone mineralization and chondrocyte differentiation. Moreover, SS patients can have cardiac and renal abnormalities, although those do not constitute cardinal features [68]. Perturbations in nervous system development relate to the mild-to-severe intellectual disabilities seen commonly in SS [68]. Problems in the AP pattern axis may justify SS scoliosis cases [68]. All of the Wnt, BMP, FGFs and RA signalling pathways play in AP patterning by inhibition/activation of their signals [261] [262]. Problems with the inner ear haven't been reported in SS so far, only chronic infections of the middle ear (chronic otitis media) that result in hearing loss [263] [264]. **Transport** processes included synapses-related processes, which might relate to ID in SS. **Signalling** processes englobed Wnt pathway, Rho protein signal transduction, RA receptor pathway, Notch signalling, and BMP signalling. Wnt canonical pathway is crucial for cell proliferation, apoptosis, cell fate and migration [265]. Wnt is a key factor in skeletal development and homeostasis, found perturbed in several skeletal dysplasias, which affect bone and cartilage growth [256]. Rho GTPases are crucial for all main eukaryotic cell processes, including gene expression and neuronal plasticity. They play in cognitive, immunological and cardiovascular disorders[266][267][268]. The Notch pathway regulates cell fate and differentiation in developmental stages [269]. Problems in this pathway are related to different developmental phenotypes involving the skeleton, heart, vasculature, face, eye, kidney and liver [270]. BMP signalling is essential in early development stages. BMP LOF mutations were shown to be implicated in several diseases englobing the neurological, ophthalmic, cardiovascular, pulmonary, urinary, gastrointestinal, and musculoskeletal systems [271]. **Cell adhesion** consisted of the most significantly enriched BP. Problems in cell-cell adhesion lead to disrupted homeostasis and embryonic development. Moreover, the reduction of cell adhesion or cytoskeleton tension in early phases of genetic reprogramming leads to more open chromatin structures, a decrease in DNAm and heterochromatin marks, and an increase of euchromatin marks at the promoter of neuronal genes, which helps TF accessibility to DNA [272]. This may explain the overall hypomethylation of genes since hypomethylation is generally related to enhanced transcription. Cell contacts can happen in diverse contexts, namely focal adhesion[273], and immune [274] and neuronal synapses [275]. These EA results of W. esig were coherent with another Sotos study with DMGs mostly enriched in nervous system development, system development, developmental process, and cell–cell adhesion. However, EA results of S. esig did not relate so much with those results [219].

The most highly hypomethylated genes were PCAT6 and KDM5B, both hypomethylated in shores in 1to5kb regions. **PCAT6** is a lncRNA found aberrantly expressed in several cancer tissues, including prostate cancer, hence the gene name [276]. It is thought that its role in tumour progression is by promoting the Wnt/catenin signalling pathway. **KDM5B** is a histone demethylase that specifically catalyzes the removal of di- and tri-methylation of H3K4 (H3K4me2 and H3K4me3) [277]. H3K4me3 is a chromatin mark associated with promoters of actively transcribed genes, thus KDM5B acts as a transcriptional repressor by removing this mark [278]. KDM5B is highly expressed in ESCs and trophoblast stem cells, and in NPCs, where it modulates their differentiation and fate [279] [280]. In healthy adults, however, KMD5B is only significantly expressed in the brain, eye, spleen, thymus testes and ovaries, and its inhibition has been shown to promote neurogenesis and decrease the proliferation of adult neural stem cells [281][282]. KDM5B may also act sometimes as a transcriptional repressor[283].

Of the 91 common genes, only a third of them were involved in developmental pathways, including **KDM5B** which was one of the most hypomethylated genes. Other development genes considered relevant for SS are worth mentioning. Regarding **H4C1**, missense mutations in lysine 91 in H4 were reported in individuals with growth delays, microcephalies and ID, due to increased genomic instability

[284]. **ITGA2B** is crucial for normal platelet hemostatic function by complexing with integrin 3, and it is found highly expressed in circulating platelets in sepsis conditions in humans, where it prevents platelet death [285]. Some newborn patients with SS had sepsis not long after birth [286]. **TP73** makes part of the p53 TF family which activates genes involved in cell cycle arrest, apoptosis and cell differentiation. It is related to neurological development, ciliogenesis, fertility and metabolism [287]. **TPPP3** plays a role in musculoskeletal system development and axon regeneration [288]. It is connected with cardiovascular and metabolic-associated vascular complications by promoting VDAC activity, promoting apoptosis, ROS production and tube formation [289]. **TET1** typically binds to CpG-dense regions, where it hydroxylates 5mc to 5hmc, thus starting the DNA demethylation, very important for the epigenetic reprogramming of germ cells and cell differentiation including the neurons [290]. **LDB3** recessive LOF variants induce severe early-onset cardiomyopathies and myopathies [291]. **CERS2** is highly expressed in liver, kidney and white brain matter, being that CERS2 LOF is related to myelin sheath defects and cerebellum degeneration, resulting in problems in motor function [292]. Furthermore, reduced levels of CERS2 were proposed to induce progressive myoclonic epilepsy (PME), englobing several CNS-related disorders and the presence of muscle contractions and seizures [293]. **HLX** is a homeobox gene that regulates mesenchymal-epithelial interaction in early enteric nervous system development [236]. HLX overexpression inhibits the migration of endothelial cells, decreases sprouting and reduces vasculogenesis [294]. **HYAL2** is essential for degrading extracellular hyaluronan (HA), which is a main component of the extracellular matrix in various tissues such as skin, lungs, brain and kidneys [295]. Missense HYAL2 variants were found in individuals with craniofacial dysmorphism that included orofacial clefting, congenital heart anomalies, cataracts, hearing loss and single palmar creases [296]. **CXCL12** codes for a small cytokine molecule widely expressed in CNS, where it aids normal neurogenesis and migration of neural precursor cells and guides axon growth. Decreased expression of CXCL12 interferes with neuronal transmission causing problems in memory processing. Moreover, CXCL12 KO was reported to damage hippocampal neurons which resulted in behavioral and learning defects. On the contrary, overexpression of both CXCL12 and its receptor CSCR4 reduced long-term epileptic seizures and cognitive impairment. Abnormalities in CXCL12 expression were found in several neurodegenerative disorders [297] [298]. Besides the brain, CXCL12 is also expressed in the thymus, heart, lungs, kidney (where it aids renal vascular development [299]), spleen and bone marrow. It aids hematopoietic cell migration from the liver to the bone marrow of the fetus, and vasculogenesis [300] [301]. Adding to those, CXCL12 is significantly overexpressed in individuals with androgenic alopecia [300]. **TDGF1** participates in cell proliferation, migration and survival, in epithelial-to-mesenchymal transition, angiogenesis and tissue homeostasis. Upregulated hypomethylation in TDGF1 promoter was seen in patients with outflow tract (TOF) defects, due to its crucial role in early cardiac progenitors differentiation and maintenance [302]. **DDR1** is a tyrosine kinase receptor that aids myelin formation and differentiation in oligodendrocytes and axons [303], which is very important for rapid action potentials in neurons. Aberrant myelination was observed in children with chronic epilepsy and NDDs [304]. **NOTCH4** is a cell surface receptor that regulates cell fate and promotes epithelial-mesenchymal transition in endocardium development and vascular smooth muscle cell development [305]. SNPs of NOTCH4 were suggested to increase the risk of dilated cardiomyopathy (DCM), which consists of enlarged ventricles and systolic dysfunction [306]. **ACSL6** has a major role in fatty acid metabolism in the brain, by catalysing the production of acyl-CoA from fatty acids, ATP and CoA. ACSL6 KO mice had reduced amounts of docosahexaenoic acid (DHA) in the brain, which is a very abundant lipid in the brain and reduces neurodegenerative disease risk [307]. **DLX3** plays in craniofacial and ventral forebrain development [308], and hair follicle differentiation and regeneration [309]. It also aids normal placental growth, being found overexpressed

in preeclampsia-affected placentae [310]. **CREB3L1** promotes bone morphogenesis, and neurogenesis, modulates the neuroendocrine system, and plays in angiogenesis and vasculature integrity maintenance. CREB3L1 mutations were reported in osteogenesis imperfecta, which is a heritable bone disease consisting of decreased bone mass and density, which increases bone fractures. This relates to the advanced bone age seen in SS [311]. **GAB1**, an insulin receptor, is expressed in cardiomyocytes, and its deletion induces dilated cardiomyopathies. It mediates PI3K/Akt and PTPN11/MAPK signalling that play in cell proliferation and growth, cell survival and differentiation [312]. **FOXL1** is a TF that regulates collagen gene expression and is a causative gene for autosomal dominant otosclerosis (abnormal bone growth in the middle ear causing hearing loss) and a candidate gene for osteoporosis. A zebrafish FOXL1 mutant had an atypical craniofacial skeleton due to aberrant collagen expression [313]. FOXL1 was found overexpressed in the midbrain, hindbrain and otic vesicles in embryonic stages in zebrafish, and FOXL1 LOF caused midbrain and eye defects [314]. **GIT1** deficient mice had profound defects in learning and memory and reduction of synaptic plasticity, which compose the ID phenotype [315]. **RASIP1** is critical in regulating GTPase signalling, morphogenesis of endothelial cells and blood vessel tubulogenesis [316]. **BBS4** is implied in an autosomal recessive disorder consisting of severe retinopathy, polydactyly, renal malformation cognitive disability and obesity [317]. **COL12A1** LOF homozygous recessive mutations and *de novo* mutations in COL12A1 were proposed to cause a syndrome with defects in both muscle and connective tissue [318]. **NCS1** is a neuronal calcium sensor involved in brain development, and it was found differentially expressed in neurodevelopmental and neurodegenerative diseases [319]. Finally, **ZNF423** is a critical partner in various signalling pathways, namely Notch and Smad signalling, implied in neurogenesis [320].

4.1 Concluding Remarks and Future Perspectives

For the syndromes engaged in this thesis, CHARGE-esig gave the most consistent results between research groups, whilst the other two revealed significant differences in the set of DMGs and respective BPs. This enhanced the notion that esig models are dependent on the type of disease and dataset used, for example, each patient's own characteristics. As well, the difference in the range of $\Delta\beta$ values for each group could be a reason for the difference in the Kabuki-esigs, since esig mapping heavily relies on effect size, especially in rare disorders [82]. Nonetheless, that did not have great impact on CHARGE cases. BP EA of W. model gave the best results in both Kabuki and Sotos. That can be explained by the reason that S. built the classification models to distinguish a variety of conditions contrarly to W. group, which only considered one or two. The Sadikovic approach, while improving specifity, gave results less aligned with the clinical manifestations. Surprinsigly, only a small fraction of the common DMGs were implied in development in Kabuki and Sotos esigs, even though they are both essentially development disorders. This motivates further investigation of DMGs involved in other type of pathways. Nevertheless, many of the common genes involved in developmental pathways could relate to syndrome symptoms. The identification of DMGs can help to explore novel therapeutics to attenuate or reverse some of the neurodevelopmental deficits (as well as others e.g. cardiac deficits). Previous attempts revealed reversion of postnatal neurological deficits in mice models for Rett syndrome [321] and of memory deficits in Kabuki [322]. It is also suggested to do a single-patient approach (single-patient *vs.* controls) instead of a single model to discriminate all patients from controls could lead to models more adapted to the patient characteristics (e.g. age, sex, type of manifestations). Furthermore, since the models were built based on blood samples, future studies in development model systems such as human induced pluripotent stem cells derived cortical neurons are suggested.

Bibliography

[1] S. L. Berger, T. Kouzarides, R. Shiekhattar, and A. Shilatifard, "An operational definition of epigenetics," *Genes and development*, vol. 23, no. 7, p. 781–3, Apr 2009.

[2] B. Griffiths and R. Hunter, "Neuroepigenetics of stress," *Neuroscience*, vol. 275, p. 420–435, Sep 2014.

[3] K. R. v. Eijk, "Quantitative studies of dna methylation and gene expression in neuropsychiatric traits," PhD Thesis, University Medical Center Utrecht, Jun 2014.

[4] I. S. Kiselev, O. G. Kulakova, A. N. Boyko, and O. O. Favorova, "Dna methylation as an epigenetic mechanism in the development of multiple sclerosis," *Acta Naturae*, vol. 13, no. 2, p. 45–57, Jul 2021.

[5] G. Felsenfeld, "A brief history of epigenetics," *Cold Spring Harbor Perspectives in Biology*, vol. 6, no. 1, p. a018200–a018200, Jan 2014.

[6] P. Peixoto, P.-F. Cartron, A. A. Serandour, and E. Hervouet, "From 1957 to nowadays: A brief history of epigenetics," *International Journal of Molecular Sciences*, vol. 21, no. 20, p. 7571, Oct 2020.

[7] A. Quach, M. E. Levine, T. Tanaka, A. T. Lu, B. H. Chen, L. Ferrucci, B. Ritz, S. Bandinelli, M. L. Neuhouser, J. M. Beasley, L. Snetselaar, R. B. Wallace, P. S. Tsao, D. Absher, T. L. Assimes, J. D. Stewart, Y. Li, L. Hou, A. A. Baccarelli, and E. A. Whitsel, "Epigenetic clock analysis of diet, exercise, education, and lifestyle factors," *Aging*, vol. 9, no. 2, p. 419–446, Feb 2017.

[8] S. Kumar, "Epigenomics of plant responses to environmental stress," *Epigenomes*, vol. 2, no. 1, p. 6, Mar 2018.

[9] A. Breiling and F. Lyko, "Epigenetic regulatory functions of dna modifications: 5-methylcytosine and beyond," *Epigenetics and Chromatin*, vol. 8, no. 1, Jul 2015.

[10] X. Wang, Q. Li, W. Yuan, Z. Cao, B. Qi, S. Kumar, Y. Li, and W. Qian, "The cytosolic fe-s cluster assembly component met18 is required for the full enzymatic activity of ros1 in active dna demethylation," *Scientific Reports*, vol. 6, no. 1, May 2016.

[11] J. L. Miller and P. A. Grant, "The role of dna methylation and histone modifications in transcriptional regulation in humans," *Subcellular Biochemistry*, vol. 61, p. 289–317, Jun 2012.

[12] D. N. Cooper, M. Taggart, and A. Bird, "Unmethlated domains in vertebrate dna," *Nucleic Acids Research*, vol. 11, no. 3, p. 647–658, Jan 1983.

[13] J. W. Cain, B. Montibus, and R. J. Oakey, "Intragenic cpg islands and their impact on gene regulation," *Frontiers in Cell and Developmental Biology*, vol. 10, Feb 2022.

[14] S. Sarda and S. Hannenhalli, "Orphan cpg islands as alternative promoters," *Transcription*, vol. 9, no. 3, p. 171–176, Nov 2017.

[15] R. Illingworth, A. Kerr, D. DeSousa, H. Jørgensen, P. Ellis, J. Stalker, D. Jackson, C. Clee, R. Plumb, J. Rogers, S. Humphray, T. Cox, C. Langford, and A. Bird, "A novel cpg island set

identifies tissue-specific methylation at developmental gene loci," *PLoS Biology*, vol. 6, no. 1, p. e22, Jan 2008.

[16] N. P. Blackledge, H. K. Long, J. Zhou, S. Kriaucionis, R. Patient, and R. J. Klose, "Bio-cap: a versatile and highly sensitive technique to purify and characterise regions of non-methylated dna," *Nucleic Acids Research*, vol. 40, no. 4, p. e32–e32, Dec 2011.

[17] R. Visone, M. G. Bacalini, S. D. Franco, M. Ferracin, M. L. Colorito, S. Pagotto, N. Laprovitera, D. Licastro, M. D. Marco, E. Scavo, C. Bassi, E. Saccenti, A. Nicotra, M. Grzes, P. Garagnani, V. D. Laurenzi, N. Valeri, R. Mariani-Costantini, M. Negrini, and G. Stassi, "Dna methylation of shelf, shore and open sea cpg positions distinguish high microsatellite instability from low or stable microsatellite status colon cancer stem cells," *Epigenomics*, vol. 11, no. 6, p. 587–604, May 2019.

[18] S. Kumar, V. Chinnusamy, and T. Mohapatra, "Epigenetics of modified dna bases: 5-methylcytosine and beyond," *Frontiers in Genetics*, vol. 9, p. 640, Dec 2018.

[19] W.-Y. Choi, J.-H. Hwang, A.-N. Cho, A. J. Lee, J. Lee, I. Jung, S.-W. Cho, L. K. Kim, and Y.-J. Kim, "Dna methylation of intragenic cpg islands are required for differentiation from ipsc to npc," *Stem Cell Reviews and Reports*, vol. 16, no. 6, p. 1316–1327, Sep 2020.

[20] P. A. Jones and S. B. Baylin, "The epigenomics of cancer," *Cell*, vol. 128, no. 4, p. 683–692, Feb 2007.

[21] L. D. Moore, T. Le, and G. Fan, "Dna methylation and its basic function," *Neuropsychopharmacology*, vol. 38, no. 1, p. 23–38, Jul 2012.

[22] E. Li and Y. Zhang, "Dna methylation in mammals," *Cold Spring Harbor Perspectives in Biology*, vol. 6, no. 5, p. a019133–a019133, May 2014.

[23] Y. A. Medvedeva, A. M. Khamis, I. V. Kulakovskiy, W. Ba-Alawi, M. S. I. Bhuyan, H. Kawaji, T. Lassmann, M. Harbers, A. R. Forrest, and V. B. Bajic, "Effects of cytosine methylation on transcription factor binding sites," *BMC Genomics*, vol. 15, no. 1, p. 119, Mar 2014.

[24] X. Yang, H. Han, D. D. De Carvalho, F. D. Lay, P. A. Jones, and G. Liang, "Gene body methylation can alter gene expression and is a therapeutic target in cancer," *Cancer cell*, vol. 26, no. 4, p. 577–590, Oct 2014.

[25] P. A. Jones, "Functions of dna methylation: islands, start sites, gene bodies and beyond," *Nature Reviews Genetics*, vol. 13, no. 7, p. 484–492, May 2012.

[26] Q. Tan, B. T. Heijmans, J. v. B. Hjelmborg, M. Soerensen, K. Christensen, and L. Christiansen, "Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort," *International Journal of Epidemiology*, p. dyw132, Aug 2016.

[27] A. H. Moarefi and F. Chédin, "Icf syndrome mutations cause a broad spectrum of biochemical defects in dnmt3b-mediated de novo dna methylation," *Journal of Molecular Biology*, vol. 409, no. 5, p. 758–772, Jun 2011.

[28] B. Jin, Y. Li, and K. D. Robertson, "Dna methylation: Superior or subordinate in the epigenetic hierarchy?" *Genes and Cancer*, vol. 2, no. 6, p. 607–617, Jan 2011.

[29] F. Lyko, "The dna methyltransferase family: a versatile toolkit for epigenetic regulation," *Nature Reviews Genetics*, vol. 19, no. 2, p. 81–92, Oct 2017.

[30] M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C.-L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, and T. H. Bestor, "Methylation of trna asp by the dna methyltransferase homolog dnmt2," *Science*, vol. 311, no. 5759, p. 395–398, Jan 2006.

[31] B. Jin and K. D. Robertson, "Dna methyltransferases, dna damage repair, and cancer," *Advances in Experimental Medicine and Biology*, vol. 754, p. 3–29, Jul 2012.

[32] P. A. Jones, "The role of dna methylation in mammalian epigenetics," *Science*, vol. 293, no. 5532, p. 1068–1070, Aug 2001.

[33] A. Sulewska, W. Niklinska, M. Kozlowski, L. Minarowski, W. Naumnik, J. Niklinski, K. Dabrowska, and L. Chyczewski, "Detection of dna methylation in eucaryotic cells," *Folia Histochemica Et Cytobiologica*, vol. 45, no. 4, p. 315–324, 2007.

[34] M. J. Pajares, C. Palanca-Ballester, R. Urtasun, E. Alemany-Cosme, A. Lahoz, and J. Sandoval, "Methods for analysis of specific dna methylation status," *Methods*, vol. 187, p. 3–12, Mar 2021.

[35] H. Wu, J. Tao, and Y. E. Sun, "Regulation and function of mammalian dna methylation patterns: a genomic perspective," *Briefings in Functional Genomics*, vol. 11, no. 3, p. 240–250, Mar 2012.

[36] R. Halabian, V. Arshad, A. Ahmadi, P. Saeedi, S. Azimzadeh Jamalkandi, and M. R. Alivand, "Laboratory methods to decipher epigenetic signatures: a comparative review," *Cellular and Molecular Biology Letters*, vol. 26, no. 1, Nov 2021.

[37] S. E. Levy and R. M. Myers, "Advancements in next-generation sequencing," *Annual Review of Genomics and Human Genetics*, vol. 17, no. 1, p. 95–115, Aug 2016.

[38] O. N. Technologies, "Benchmarking nanopore methylation analysis by comparison to publicly available bisulphite datasets," in *Oxford Nanopore Technologies*, Dec 2022.

[39] A. H. Laszlo, I. M. Derrington, H. Brinkerhoff, K. W. Langford, I. C. Nova, J. M. Samson, J. J. Bartlett, M. Pavlenok, and J. H. Gundlach, "Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore mspa," *Proceedings of the National Academy of Sciences*, vol. 110, no. 47, p. 18904–18909, Oct 2013.

[40] Q. Liu, D. C. Georgieva, D. Egli, and K. Wang, "Nanomod: a computational tool to detect dna modifications using nanopore long-read sequencing data," *BMC Genomics*, vol. 20, no. S1, Feb 2019.

[41] Q. Liu, L. Fang, G. Yu, D. Wang, C.-L. Xiao, and K. Wang, "Detection of dna base modifications by deep recurrent neural network on oxford nanopore sequencing data," *Nature Communications*, vol. 10, no. 1, Jun 2019.

[42] J. T. Simpson, R. E. Workman, P. C. Zuzarte, M. David, L. J. Dursi, and W. Timp, "Detecting dna cytosine methylation using nanopore sequencing," *Nature Methods*, vol. 14, no. 4, p. 407–410, Feb 2017.

[43] S. Kingan, H. Heaton, J. Cudini, C. Lambert, P. Baybayan, B. Galvin, R. Durbin, J. Korlach, and M. Lawniczak, "A high-quality de novo genome assembly from a single mosquito using pacbio sequencing," *Genes*, vol. 10, no. 1, p. 62, Jan 2019.

[44] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, and C. L. Paul, "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual dna strands." *Proceedings of the National Academy of Sciences*, vol. 89, no. 5, p. 1827–1831, Mar 1992.

[45] S. Tierling, M. Schuster, R. Tetzner, and J. Walter, "A combined hm-pcr/snupe method for high sensitive detection of rare dna methylation," *Epigenetics and Chromatin*, vol. 3, no. 1, Jun 2010.

[46] I. Wong, "Qualitative and quantitative polymerase chain reaction-based methods for dna methylation analyses," *Humana Press eBooks*, vol. 336, p. 33–44, Sep 2006.

[47] D. Diep, N. Plongthongkum, A. Gore, H.-L. Fung, R. Shoemaker, and K. Zhang, "Library-free methylation sequencing with bisulfite padlock probes," *Nature methods*, vol. 9, no. 3, p. 270–272, Feb 2012.

[48] J. Tost and I. G. Gut, "Dna methylation analysis by pyrosequencing," *Nature Protocols*, vol. 2, no. 9, p. 2265–2275, Sep 2007.

[49] W.-S. Yong, F.-M. Hsu, and P.-Y. Chen, "Profiling genome-wide dna methylation," *Epigenetics and Chromatin*, vol. 9, no. 1, Jun 2016.

[50] Q. Gouil and A. Keniry, "Latest techniques to study dna methylation," *Essays in Biochemistry*, vol. 63, no. 6, p. 639–648, Dec 2019.

[51] M. Bibikova, B. Barnes, C. Tsan, V. Ho, B. Klotzle, J. M. Le, D. Delano, L. Zhang, G. P. Schroth, K. L. Gunderson, J.-B. Fan, and R. Shen, "High density dna methylation array with single cpg site resolution," *Genomics*, vol. 98, no. 4, p. 288–295, Oct 2011.

[52] R. Pidsley, E. Zotenko, T. J. Peters, M. G. Lawrence, G. P. Risbridger, P. Molloy, S. Van Djik, B. Muhlhausler, C. Stirzaker, and S. J. Clark, "Critical evaluation of the illumina methylationepic beadchip microarray for whole-genome dna methylation profiling," *Genome Biology*, vol. 17, no. 1, Oct 2016.

[53] D. R. Masser, A. S. Berg, and W. M. Freeman, "Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing," *Epigenetics and Chromatin*, vol. 6, no. 1, p. 33, Oct 2013.

[54] A. Meissner, "Reduced representation bisulfite sequencing for comparative high-resolution dna methylation analysis," *Nucleic Acids Research*, vol. 33, no. 18, p. 5868–5877, Oct 2005.

[55] Y. Li, "Modern epigenetics methods in biological research," *Methods*, Jul 2020.

[56] D. P. Genereux, W. C. Johnson, A. F. Burden, R. Stoger, and C. D. Laird, "Errors in the bisulfite conversion of dna: modulating inappropriate- and failed-conversion frequencies," *Nucleic Acids Research*, vol. 36, no. 22, p. e150–e150, Nov 2008.

[57] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, and S. Balasubramanian, "Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution," *Science*, vol. 336, no. 6083, p. 934–937, Apr 2012.

[58] P. Gibas, M. Narmontė, Z. Staševskij, J. Gordevičius, S. Klimašauskas, and E. Kriukienė, "Precise genomic mapping of 5-hydroxymethylcytosine via covalent tether-directed sequencing," *PLOS Biology*, vol. 18, no. 4, p. e3000684, Apr 2020.

[59] C.-X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, and C. He, "Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming," *Cell*, vol. 153, no. 3, p. 678–691, Apr 2013.

[60] X. Lu, C.-X. Song, K. E. Szulwach, Z. Wang, P. A. Weidenbacher, P. Jin, and C. He, "Chemical modification-assisted bisulfite sequencing (cab-seq) for 5-carboxylcytosine detection in dna," *Journal of the American Chemical Society*, vol. 135, no. 25, p. 9315–9317, Jun 2013.

[61] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, and S. Balasubramanian, "Quantitative sequencing of 5-formylcytosine in dna at single-base resolution," *Nature Chemistry*, vol. 6, no. 5, p. 435–440, Mar 2014.

[62] H. T. Bjornsson, "The mendelian disorders of the epigenetic machinery," *Genome Research*, vol. 25, no. 10, p. 1473–1481, Oct 2015.

[63] J. A. López-Rivera, E. Pérez-Palma, J. Symonds, A. S. Lindy, D. A. McKnight, C. Leu, S. Zuberi, A. Brunklaus, R. S. Møller, and D. Lal, "A catalogue of new incidence estimates of monogenic neurodevelopmental disorders caused by de novo variants," *Brain*, vol. 143, no. 4, p. 1099–1105, Mar 2020.

[64] D. J. Morris-Rosendahl and M.-A. Crocq, "Neurodevelopmental disorders—the history and future of a diagnostic concept," *Dialogues in Clinical Neuroscience*, vol. 22, no. 1, p. 65–72, Mar 2020.

[65] D. Pinto, E. Delaby, D. Merico, M. Barbosa, A. Merikangas, L. Klei, B. Thiruvahindrapuram, X. Xu, R. Ziman, Z. Wang, J. A. Vorstman, A. Thompson, R. Regan, M. Pilorge, G. Pellecchia, A. T. Pagnamenta, B. Oliveira, C. R. Marshall, T. R. Magalhaes, and J. K. Lowe, "Convergence of genes and cellular pathways dysregulated in autism spectrum disorders," *The American Journal of Human Genetics*, vol. 94, no. 5, p. 677–694, May 2014.

[66] N. Usman and M. Sur, "Statpearls [internet] - charge syndrome," Treasure Island (FL), Mar 2023.

[67] M. P. Adam, L. Hudgins, and M. Hannibal, "Genereviews [internet] - kabuki syndrome," Seattle (WA), Sep 2022.

[68] K. Tatton-Brown, T. R. Cole, and N. Rahman, "Genereviews [internet] - sotos syndrome," Seattle (WA), Dec 2022.

[69] K. Rooney and B. Sadikovic, "Dna methylation episignatures in neurodevelopmental disorders associated with large structural copy number variants: Clinical implications," *International Journal of Molecular Sciences*, vol. 23, no. 14, p. 7862, Jul 2022.

[70] D. T. Miller, M. P. Adam, S. Aradhya, L. G. Biesecker, A. R. Brothman, N. P. Carter, D. M. Church, J. A. Crolla, E. E. Eichler, C. J. Epstein, W. A. Faucett, L. Feuk, J. M. Friedman, A. Hamosh, L. Jackson, E. B. Kaminsky, K. Kok, I. D. Krantz, R. M. Kuhn, and C. Lee, "Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies," *The American Journal of Human Genetics*, vol. 86, no. 5, p. 749–764, May 2010.

[71] M. Silva, N. de Leeuw, K. Mann, H. Schuring-Blom, S. Morgan, D. Giardino, K. Rack, and R. Hastings, "European guidelines for constitutional cytogenomic analysis," *European Journal of Human Genetics*, vol. 27, no. 1, p. 1–16, Jan 2019.

[72] G. D'Amours, M. Langlois, G. Mathonnet, R. Fetni, S. Nizard, M. Srour, F. Tihy, M. S Phillips, J. L Michaud, and E. Lemyre, "Snp arrays: comparing diagnostic yields for four platforms in children with developmental delay," *BMC Medical Genomics*, vol. 7, no. 1, Dec 2014.

[73] B. Sadikovic, M. A. Levy, J. Kerkhof, E. Aref-Eshghi, L. Schenkel, A. Stuart, H. McConkey, P. Henneman, A. Venema, C. E. Schwartz, R. E. Stevenson, S. A. Skinner, B. R. DuPont, R. S. Fletcher, T. B. Balci, V. M. Siu, J. L. Granadillo, J. Masters, M. Kadour, and M. J. Friez, "Clinical epigenomics: genome-wide dna methylation analysis for the diagnosis of mendelian disorders," *Genetics in Medicine*, vol. 23, no. 6, p. 1065–1074, Jun 2021.

[74] K. Schwarze, J. Buchanan, J. C. Taylor, and S. Wordsworth, "Are whole-exome and whole-genome sequencing approaches cost-effective? a systematic review of the literature," *Genetics in Medicine*, vol. 20, no. 10, p. 1122–1130, Feb 2018.

[75] B. Sikkema-Raddatz, L. F. Johansson, E. N. de Boer, R. Almomani, L. G. Boven, M. P. van den Berg, K. Y. van Spaendonck-Zwarts, J. P. van Tintelen, R. H. Sijmons, J. D. H. Jongbloed, and R. J. Sinke, "Targeted next-generation sequencing can replace sanger sequencing in clinical diagnostics," *Human Mutation*, vol. 34, no. 7, p. 1035–1042, Apr 2013.

[76] Y. S. Fraiman and M. H. Wojcik, "The influence of social determinants of health on the genetic diagnostic odyssey: who remains undiagnosed, why, and to what effect?" *Pediatric Research*, vol. 89, no. 2, p. 295–300, Jan 2021.

[77] J. M. Savatt and S. M. Myers, "Genetic testing in neurodevelopmental disorders," *Frontiers in Pediatrics*, vol. 9, Feb 2021.

[78] L. C. Schenkel, E. Aref-Eshghi, K. Rooney, J. Kerkhof, M. A. Levy, H. McConkey, R. C. Rogers, K. Phelan, S. M. Sarasua, L. Jain, R. Pauly, L. Boccuto, B. DuPont, G. Cappuccio, N. Brunetti-Pierri, C. E. Schwartz, and B. Sadikovic, "Dna methylation epi-signature is associated with two molecularly and phenotypically distinct clinical subtypes of phelan-mcdermid syndrome," *Clinical Epigenetics*, vol. 13, no. 1, Jan 2021.

[79] S. Haghshenas, M. A. Levy, J. Kerkhof, E. Aref-Eshghi, H. McConkey, T. Balci, V. M. Siu, C. D. Skinner, R. E. Stevenson, B. Sadikovic, and C. Schwartz, "Detection of a dna methylation signature for the intellectual developmental disorder, x-linked, syndromic, armfield type," *International Journal of Molecular Sciences*, vol. 22, no. 3, p. 1111, Jan 2021.

[80] E. Aref-Eshghi, E. G. Bend, R. L. Hood, L. C. Schenkel, D. A. Carere, R. Chakrabarti, S. C. S. Nagamani, S. W. Cheung, P. M. Campeau, C. Prasad, V. M. Siu, L. Brady, M. A. Tarnopolsky,

D. J. Callen, A. M. Innes, S. M. White, W. S. Meschino, A. Y. Shuen, G. Paré, and D. E. Bulman, "Bafopathies' dna methylation epi-signatures demonstrate diagnostic utility and functional continuum of coffin–siris and nicolaides–baraitser syndromes," *Nature Communications*, vol. 9, no. 1, Nov 2018.

[81] E. G. Bend, E. Aref-Eshghi, D. B. Everman, R. C. Rogers, S. S. Cathey, E. J. Prijoles, M. J. Lyons, H. Davis, K. Clarkson, K. W. Gripp, D. Li, E. Bhoj, E. Zackai, P. Mark, H. Hakonarson, L. A. Demmer, M. A. Levy, J. Kerkhof, A. Stuart, and D. Rodenhiser, "Gene domain-specific dna methylation episignatures highlight distinct molecular entities of adnp syndrome," *Clinical Epigenetics*, vol. 11, no. 1, Apr 2019.

[82] E. Aref-Eshghi, J. Kerkhof, V. P. Pedro, G. D. France, M. Barat-Houari, N. Ruiz-Pallares, J.-C. Andrau, D. Lacombe, J. Van-Gils, P. Fergelot, C. Dubourg, V. Cormier-Daire, S. Rondeau, F. Lecoquierre, P. Saugier-Veber, G. Nicolas, G. Lesca, N. Chatron, D. Sanlaville, and A. Vitobello, "Evaluation of dna methylation episignatures for diagnosis and phenotype correlations in 42 mendelian neurodevelopmental disorders," *American Journal of Human Genetics*, vol. 106, no. 3, p. 356–370, Mar 2020.

[83] S. Kurdyukov and M. Bullock, "Dna methylation analysis: Choosing the right method," *Biology*, vol. 5, no. 1, p. 3, Jan 2016.

[84] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, p. 159–190, Nov 2006.

[85] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent Data Analysis*, vol. 6, no. 5, p. 429–449, Nov 2002.

[86] C. Krittanawong, H. Zhang, Z. Wang, M. Aydar, and T. Kitai, "Artificial intelligence in precision cardiovascular medicine," *Journal of the American College of Cardiology*, vol. 69, no. 21, p. 2657–2664, May 2017.

[87] T. Husson, F. Lecoquierre, G. Nicolas, A.-C. Richard, A. Afenjar, S. Audebert-Bellanger, C. Badens, F. Bilan, V. Bizaoui, A. Boland, M.-N. Bonnet-Dupeyron, E. Brischoux-Boucher, C. Bonnet, M. Bournez, O. Boute, P. Brunelle, R. Caumes, P. Charles, N. Chassaing, and N. Chatron, "Episignatures in practice: independent evaluation of published episignatures for the molecular diagnostics of ten neurodevelopmental disorders," *European Journal of Human Genetics*, p. 1–10, Oct 2023.

[88] A. BENDAVID, "Comparison of classification accuracy using cohen's weighted kappa," *Expert Systems with Applications*, vol. 34, no. 2, p. 825–832, Feb 2008.

[89] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, no. 4, p. 427–437, Jul 2009.

[90] M. A. Levy, H. McConkey, J. Kerkhof, M. Barat-Houari, S. Bargiacchi, E. Biamino, M. P. Bralo, G. Cappuccio, A. Ciolfi, A. Clarke, B. R. DuPont, M. W. Elting, L. Faivre, T. Fee, R. S. Fletcher, F. Cherik, A. Foroutan, M. J. Friez, C. Gervasini, and S. Haghshenas, "Novel diagnostic dna methylation episignatures expand and refine the epigenetic landscapes of mendelian disorders," *Human Genetics and Genomics Advances*, vol. 3, no. 1, Jan 2022.

[91] E. Aref-Eshghi, D. I. Rodenhiser, L. C. Schenkel, H. Lin, C. Skinner, P. Ainsworth, G. Paré, R. L. Hood, D. E. Bulman, K. D. Kernohan, K. M. Boycott, P. M. Campeau, C. Schwartz, and B. Sadikovic, "Genomic dna methylation signatures enable concurrent diagnosis and clinical genetic variant classification in neurodevelopmental syndromes," *The American Journal of Human Genetics*, vol. 102, no. 1, p. 156–174, Jan 2018.

[92] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems with Applications*, vol. 73, p. 220–239, May 2017.

[93] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, p. 2839–2846, Sep 2015.

[94] S. Brasil, C. J. Neves, T. Rijoff, M. Falcão, G. Valadão, P. A. Videira, and V. Ferreira, "Artificial intelligence in epigenetic studies: Shedding light on rare diseases," *Frontiers in Molecular Biosciences*, vol. 8, May 2021.

[95] S. M. Waszak, P. A. Northcott, I. Buchhalter, G. W. Robinson, C. Sutter, S. Groebner, K. B. Grund, L. Brugières, D. T. W. Jones, K. W. Pajtler, A. S. Morrissy, M. Kool, D. Sturm, L. Chavez, A. Ernst, S. Brabetz, M. Hain, T. Zichner, M. Segura-Wang, and J. Weischenfeldt, "Spectrum and prevalence of genetic predisposition in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial cohort," *The Lancet Oncology*, vol. 19, no. 6, p. 785–798, Jun 2018.

[96] A. L. Tarca, V. J. Carey, X.-w. Chen, R. Romero, and S. Drăghici, "Machine learning and its applications to biology," *PLoS Computational Biology*, vol. 3, no. 6, p. e116, 2007.

[97] A.-L. Boulesteix and K. Strimmer, "Partial least squares: a versatile tool for the analysis of high-dimensional genomic data," *Briefings in Bioinformatics*, vol. 8, no. 1, p. 32–44, May 2006.

[98] H. O. Alanazi, A. H. Abdullah, and K. N. Qureshi, "A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care," *Journal of Medical Systems*, vol. 41, no. 4, Mar 2017.

[99] A. M. D'Gama and C. A. Walsh, "Somatic mosaicism and neurodevelopmental disease," *Nature Neuroscience*, vol. 21, no. 11, p. 1504–1514, Nov 2018.

[100] S. Choufani, C. Cytrynbaum, B. H. Y. Chung, A. L. Turinsky, D. Grafodatskaya, Y. A. Chen, A. S. A. Cohen, L. Dupuis, D. T. Butcher, M. T. Siu, H. M. Luk, I. F. M. Lo, S. T. S. Lam, O. Caluseriu, D. J. Stavropoulos, W. Reardon, R. Mendoza-Londono, M. Brudno, W. T. Gibson, and D. Chitayat, "Nsd1 mutations generate a genome-wide dna methylation signature," *Nature Communications*, vol. 6, no. 1, Dec 2015.

[101] M. Mendioroz, C. Do, X. Jiang, C.-H. Liu, H. Darbary, C. H. Lang, J. Lin, A. Thomas, S. Abu-Amero, P. Stanier, A. Temkin, A. Yale, M.-M. Liu, Y. Li, M. Salas, K. Kerkel, G. T. Capone, W. Silverman, Y. Yu, and G. E. Moore, "Trans effects of chromosome aneuploidies on dna methylation patterns in human down syndrome and mouse models," *Genome Biology*, vol. 16, no. 1, Nov 2015.

[102] B. Sadikovic, E. Aref-Eshghi, M. A. Levy, and D. Rodenhiser, "Dna methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype," *Epigenomics*, vol. 11, no. 5, p. 563–575, Apr 2019.

[103] X. S. Liu, H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R. A. Young, and R. Jaenisch, "Editing dna methylation in the mammalian genome," *Cell*, vol. 167, no. 1, pp. 233–247.e17, 2016.

[104] B. Sadikovic, M. A. Levy, and E. Aref-Eshghi, "Functional annotation of genomic variation: Dna methylation episignatures in neurodevelopmental mendelian disorders," *Human Molecular Genetics*, vol. 29, no. R1, p. R27–R32, Jul 2020.

[105] M. C. J. Jongmans, "Charge syndrome: the phenotypic spectrum of mutations in the chd7 gene," *Journal of Medical Genetics*, vol. 43, no. 4, p. 306–314, Sep 2005.

[106] R. A. Pagon, J. M. Graham, J. Zonana, and S.-L. Yong, "Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: Charge association," *The Journal of Pediatrics*, vol. 99, no. 2, p. 223–227, Aug 1981.

[107] K. Klement, M. S. Luijsterburg, J. Pinder, C. S. Cena, V. D. Nero, C. M. Wintersinger, G. Dellaire, H. van Attikum, and A. A. Goodarzi, "Opposing iswi- and chd-class chromatin remodeling activities orchestrate heterochromatic dna repair," *The Journal of cell biology*, vol. 207, no. 6, p. 717–733, Dec 2014.

[108] R. Cacabelos, "Chapter 22 - epigenetics and pharmacoepigenetics of neurodevelopmental and neuropsychiatric disorders," p. 609–709, Jan 2019.

[109] R. Ufartes, J. Schwenty-Lara, L. Freese, C. Neuhofer, J. Möller, P. Wehner, C. M. A. van Ravenswaaij-Arts, M. T. Y. Wong, I. Schanze, A. Tzschach, O. Bartsch, A. Borchers, and S. Pauli, "Sema3a plays a role in the pathogenesis of charge syndrome," *Human Molecular Genetics*, vol. 27, no. 8, p. 1343–1352, Feb 2018.

[110] N. Niikawa, Y. Kuroki, T. Kajii, N. Matsuura, S. Ishikiriyama, H. Tonoki, N. Ishikawa, Y. Yamada, M. Fujita, U. H, Y. Iwama, I. Kondoh, Y. Fukushima, Y. Nako, I. Matsui, T. Urakami, S. Aritaki, M. Hara, Y. Suzuki, and H. Chyo, "Kabuki make-up (niikawa-kuroki) syndrome: A study of 62 patients," *American Journal of Medical Genetics*, vol. 31, no. 3, p. 565–589, Nov 1988.

[111] E. Smith, C. Lin, and A. Shilatifard, "The super elongation complex (sec) and mll in development and disease," *Genes and Development*, vol. 25, no. 7, p. 661–672, Apr 2011.

[112] C. Martin and Y. Zhang, "The diverse functions of histone lysine methylation," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 11, p. 838–849, Nov 2005.

[113] M. R. Hubner and D. L. Spector, "Role of h3k27 demethylases jmjd3 and utx in transcriptional regulation," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 75, no. 0, p. 43–49, Jan 2010.

[114] R. Margueron and D. Reinberg, "The polycomb complex prc2 and its mark in life," *Nature*, vol. 469, no. 7330, p. 343–349, Jan 2011.

[115] R. M. Shawky, R. Gamal, and N. Mostafa, "Kabuki make-up syndrome with genitourinary anomalies, ophthalmologic features and hyperpigmentation in an egyptian child," *Egyptian Journal of Medical Human Genetics*, vol. 18, no. 1, p. 87–92, Jan 2017.

[116] N. Niikawa, N. Matsuura, Y. Fukushima, T. Ohsawa, and T. Kajii, "Kabuki make-up syndrome: A syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency," *The Journal of Pediatrics*, vol. 99, no. 4, p. 565–569, Oct 1981.

[117] S. Phillips, S. Hemmady, P. Thomas, and D. ODoherty, "Kabuki syndrome presenting with congenital talipes equinovarus," *Journal of Pediatric Orthopaedics B*, vol. 14, no. 4, p. 285, Jul 2005.

[118] N. Philip, P. Meinecke, A. David, J. Dean, S. Ayme, R. Clark, E. Gross-Kieselstein, D. Hosenfeld, A. Moncla, and D. Muller, "Kabuki make-up (niikawa-kuroki) syndrome: a study of 16 non-japanese cases," *Clinical Dysmorphology*, vol. 1, no. 2, p. 63–77, Apr 1992.

[119] S. Boniel, K. Szymańska, R. Śmigiel, and K. Szczałuba, "Kabuki syndrome—clinical review with molecular aspects," *Genes*, vol. 12, no. 4, p. 468, Mar 2021.

[120] H. Gole, R. Chuk, and D. Coman, "Persistent hyperinsulinism in kabuki syndrome 2: Case report and literature review," *Clinics and Practice*, vol. 6, no. 3, p. 68–70, Aug 2016.

[121] K. Tatton-Brown, J. Douglas, K. L. Coleman, G. Baujat, T. Cole, S. Das, D. Horn, H. M. Hughes, I. K. Temple, F. Faravelli, D. Waggoner, S. Türkmen, V. Cormier-Daire, A. Irrthum, and N. Rahman, "Genotype-phenotype associations in sotos syndrome: An analysis of 266 individuals with nsd1 aberrations," *American Journal of Human Genetics*, vol. 77, no. 2, p. 193–204, Aug 2005.

[122] B. M. N. L. of Medicine (US), "Sotos syndrome," accessed on 10.11.2023. [Online]. Available: https://medlineplus.gov/genetics/condition/sotos-syndrome/

[123] C. H. Backes, K. Markham, P. Moorehead, L. Cordero, C. A. Nankervis, and P. J. Giannone, "Maternal preeclampsia and neonatal outcomes," *Journal of Pregnancy*, vol. 2011, 2011.

[124] D. T. Butcher, C. Cytrynbaum, A. L. Turinsky, M. T. Siu, M. Inbar-Feigenberg, R. Mendoza-Londono, D. Chitayat, S. Walker, J. Machado, O. Caluseriu, L. Dupuis, D. Grafodatskaya, W. Reardon, B. Gilbert-Dussardier, A. Verloes, F. Bilan, J. M. Milunsky, R. Basran, B. Papsin, and T. L. Stockley, "Charge and kabuki syndromes: Gene-specific dna methylation signatures identify epigenetic mechanisms linking these clinically overlapping conditions," *The American Journal of Human Genetics*, vol. 100, no. 5, p. 773–788, May 2017.

[125] B. T. Sherman, M. Hao, J. Qiu, X. Jiao, M. W. Baseler, H. C. Lane, T. Imamichi, and W. Chang, "David: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)," *Nucleic Acids Research*, vol. 50, no. W1, Mar 2022.

[126] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using david bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, p. 44–57, Dec 2008.

[127] F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc, "Revigo summarizes and visualizes long lists of gene ontology terms," *PLoS ONE*, vol. 6, no. 7, p. e21800, Jul 2011.

[128] M. Ejarque, V. Ceperuelo-Mallafré, C. Serena, E. Maymo-Masip, X. Duran, A. Díaz-Ramos, M. Millan-Scheiding, Y. Núñez-Álvarez, C. Núñez-Roa, P. Gama, P. M. Garcia-Roves, M. A. Peinado, J. M. Gimble, A. Zorzano, J. Vendrell, and S. Fernández-Veledo, "Adipose tissue mitochondrial dysfunction in human obesity is linked to a specific dna methylation signature in adipose-derived stem cells," *International Journal of Obesity*, vol. 43, no. 6, p. 1256–1268, Sep 2018.

[129] N. T. Ventham, N. A. Kennedy, A. T. Adams, R. Kalla, S. Heath, K. R. O'Leary, H. Drummond, D. C. Wilson, I. G. Gut, E. R. Nimmo, and J. Satsangi, "Integrative epigenome-wide analysis demonstrates that dna methylation may mediate genetic risk in inflammatory bowel disease," *Nature Communications*, vol. 7, no. 1, Nov 2016.

[130] M. J. Ball, J. Li, Y. Gao, J.-H. Lee, E. M. LeProust, I.-H. Park, B. Xie, G. Q. Daley, and G. M. Church, "Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells," *Nature Biotechnology*, vol. 27, no. 4, p. 361–368, Mar 2009.

[131] R. G. Cavalcante and M. A. Sartor, "annotatr: genomic regions in context," *Bioinformatics*, vol. 33, no. 15, p. 2381–2383, Mar 2017.

[132] D. Sanlaville and A. Verloes, "Charge syndrome: an update," *European Journal of Human Genetics*, vol. 15, no. 4, p. 389–399, Feb 2007.

[133] C. Doyle and K. Blake, "Scoliosis in charge: A prospective survey and two case reports," *American Journal of Medical Genetics - Part A*, vol. 133A, no. 3, p. 340–343, Feb 2005.

[134] N. Corsten-Janssen, W. S. Kerstjens-Frederikse, G. J. du Marchie Sarvaas, M. E. Baardman, M. K. Bakker, J. E. Bergman, H. D. Hove, K. R. Heimdal, C. F. Rustad, R. C. Hennekam, R. M. Hofstra, L. H. Hoefsloot, C. M. Van Ravenswaaij-Arts, and L. Kapusta, "The cardiac phenotype in patients with a chd7 mutation," *Circulation: Cardiovascular Genetics*, vol. 6, no. 3, p. 248–254, Jun 2013.

[135] N. Corsten-Janssen and P. J. Scambler, "Clinical and molecular effects ofchd7in the heart," *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 175, no. 4, p. 487–495, Oct 2017.

[136] M. Hoch, S. Patel, D. Jethanamest, W. Win, G. Fatterpekar, J. Roland, and M. Hagiwara, "Head and neck mri findings in charge syndrome," *AJNR: American Journal of Neuroradiology*, vol. 38, no. 12, p. 2357–2363, Dec 2017.

[137] A. E. Lin, J. R. Siebert, and J. M. Graham, "Central nervous system malformations in the charge association," *American journal of medical genetics*, vol. 37, no. 3, p. 304–310, Nov 1990.

[138] K. Fujita, N. Aida, Y. Asakura, K. Kurosawa, T. Niwa, K. Muroya, M. Adachi, G. Nishimura, and T. Inoue, "Abnormal basiocciput development in charge syndrome," *American Journal of Neuroradiology*, vol. 30, no. 3, p. 629–634, Dec 2008.

[139] T. Yu, L. C. Meiners, K. Danielsen, M. T. Wong, T. Bowler, D. Reinberg, P. J. Scambler, C. M. van Ravenswaaij-Arts, and M. A. Basson, "Deregulated fgf and homeotic gene expression underlies cerebellar vermis hypoplasia in charge syndrome," *eLife*, vol. 2, Dec 2013.

[140] Y. B. Sohn, J. M. Ko, C. H. Shin, S. W. Yang, J. H. Chae, and K. A. Lee, "Cerebellar vermis hypoplasia in charge syndrome: clinical and molecular characterization of 18 unrelated korean patients," *Journal of Human Genetics*, vol. 61, no. 3, p. 235–239, Nov 2015.

[141] L. Gregory, E. Gevers, J. S. Baker, T. Kasia, K. Chong, D. Josifova, M. Caimari, F. Bilan, M. J. McCabe, and M. Dattani, "Structural pituitary abnormalities associated with charge syndrome," *The Journal of Clinical Endocrinology and Metabolism*, vol. 98, no. 4, p. E737–E743, Apr 2013.

[142] N. C. Reddy, S. P. Majidi, L. Kong, M. Nemera, C. J. Ferguson, M. Moore, T. M. Goncalves, H.-K. Liu, J. A. J. Fitzpatrick, G. Zhao, T. Yamada, A. Bonni, and H. W. Gabel, "Charge syndrome protein chd7 regulates epigenomic activation of enhancers in granule cell precursors and gyrification of the cerebellum," *Nature Communications*, vol. 12, no. 1, Sep 2021.

[143] C. R. Clapier, J. Iwasa, B. R. Cairns, and C. L. Peterson, "Mechanisms of action and regulation of atp-dependent chromatin-remodelling complexes," *Nature Reviews Molecular Cell Biology*, vol. 18, no. 7, p. 407–422, Jul 2017.

[144] K. M. Jones, N. Sarić, J. P. Russell, C. L. Andoniadou, P. J. Scambler, and M. A. Basson, "Chd7 maintains neural stem cell quiescence and prevents premature stem cell depletion in the adult hippocampus," *STEM CELLS*, vol. 33, no. 1, p. 196–210, Dec 2014.

[145] S. Mehr, P. Hsu, and D. E. Campbell, "Immunodeficiency in charge syndrome," *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 175, no. 4, p. 516–523, Nov 2017.

[146] H. Okuno, F. R. Mihara, S. Ohta, K. Fukuda, K. Kurosawa, W. Akamatsu, T. Sanosaka, J. Kohyama, K. Hayashi, K. Nakajima, T. Takahashi, J. Wysocka, K. Kosaki, and H. Okano, "Charge syndrome modeling using patient-ipscs reveals defective migration of neural crest cells harboring chd7 mutations," *eLife*, p. e21114, Nov 2017.

[147] A. K Knecht and M. Bronner-Fraser, "R e v i e w s," *Nature reviews*, Jun 2002.

[148] L. S. Gammill and M. Bronner-Fraser, "Neural crest specification: migrating into genomics," *Nature Reviews Neuroscience*, vol. 4, no. 10, p. 795–805, Oct 2003.

[149] P. D. Lampe and A. F. Lau, "Regulation of gap junctions by phosphorylation of connexins," *Archives of Biochemistry and Biophysics*, vol. 384, no. 2, p. 205–215, Dec 2000.

[150] S. Kuriyama and R. Mayor, "Molecular analysis of neural crest migration," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1495, p. 1349–1362, Apr 2008.

[151] L.-Y. Shi, Z. Wang, Y. Li, S. Zheng, Y. Wu, and B. Hu, "Deletion of the chd7 hinders oligodendrocyte progenitor cell development and myelination in zebrafish," *International Journal of Molecular Sciences*, vol. 24, no. 17, p. 13535–13535, Aug 2023.

[152] M. Mallo, D. M. Wellik, and J. Deschamps, "Hox genes and regional patterning of the vertebrate body plan," *Developmental Biology*, vol. 344, no. 1, p. 7–15, Aug 2010.

[153] W. McGinnis and R. Krumlauf, "Homeobox genes and axial patterning," *Cell*, vol. 68, no. 2, p. 283–302, Jan 1992.

[154] R. Krumlauf, "Hox genes in vertebrate development," *Cell*, vol. 78, no. 2, p. 191–201, Jul 1994.

[155] C. Nolte and R. Krumlauf, "Expression of hox genes in the nervous system of vertebrates," 2013.

[156] D. G. Wilkinson, S. Bhatt, M. Cook, E. Boncinelli, and R. Krumlauf, "Segmental expression of hox-2 homoeobox-containing genes in the developing mouse hindbrain," *Nature*, vol. 341, no. 6241, p. 405–409, Oct 1989.

[157] P. Hunt, D. Wilkinson, and R. Krumlauf, "Patterning the vertebrate head: murine hox 2 genes mark distinct subpopulations of premigratory and migrating cranial neural crest," *Development*, vol. 112, no. 1, p. 43–50, May 1991.

[158] P. Hunt, M. Gulisano, M. Cook, M.-H. Sham, A. Faiella, D. Wilkinson, E. Boncinelli, and R. Krumlauf, "A distinct hox code for the branchial region of the vertebrate head," *Nature*, vol. 353, no. 6347, p. 861–864, Oct 1991.

[159] P. Dollé, J.-C. Izpisúa-Belmonte, H. Falkenstein, A. Renucci, and D. Duboule, "Coordinate expression of the murine hox-5 complex homoeobox-containing genes during limb pattern formation," *Nature*, vol. 342, no. 6251, p. 767–772, Dec 1989.

[160] J. Izpisúa-Belmonte, H. Falkenstein, P. Dollé, A. Renucci, and D. Duboule, "Murine genes related to the drosophila abdb homeotic genes are sequentially expressed during development of the posterior part of the body." *The EMBO Journal*, vol. 10, no. 8, p. 2279–2289, Aug 1991.

[161] P. Dolle, J. C. Izpisua-Belmonte, J. M. Brown, C. Tickle, and D. Duboule, "Hox-4 genes and the morphogenesis of mammalian genitalia." *Genes and Development*, vol. 5, no. 10, p. 1767–1776, Oct 1991.

[162] S. Sakata, S. Okada, K. Aoyama, K. Hara, C. Tani, R. Kagawa, A. Utsunomiya-Nakamura, S. Miyagawa, T. Ogata, H. Mizuno, and M. Kobayashi, "Individual clinically diagnosed with charge syndrome but with a mutation in kmt2d, a gene associated with kabuki syndrome: A case report," *Frontiers in Genetics*, vol. 8, Dec 2017.

[163] M. V. Volpe, K. Ting, H. C. Nielsen, and M. R. Chinoy, "Unique spatial and cellular expression patterns of hoxa5, hoxb4, and hoxb6 proteins in normal developing murine lung are modified in pulmonary hypoplasia," *Birth Defects Research*, vol. 82, no. 8, p. 571–584, Jun 2008.

[164] A. Barkat, L. Essabar, and F. Aglili, "An unusual case of charge syndrome presenting with intrathoracic kidney and right-sided diaphragmatic hernia," *Pediatric Oncall*, vol. 15, no. 3, Jan 2018.

[165] B. Lizen, B. Hutlet, D. Bissen, D. Sauvegarde, M. Hermant, M.-T. Ahn, and F. Gofflot, "Hoxa5 localization in postnatal and adult mouse brain is suggestive of regulatory roles in postmitotic neurons," *Journal of Comparative Neurology*, vol. 525, no. 5, p. 1155–1175, Apr 2017.

[166] N. Degani, Y. Lubelsky, R. B.-T. Perry, E. Ainbinder, and I. Ulitsky, "Highly conserved and cis-acting lncrnas produced from paralogous regions in the center of hoxa and hoxb clusters in the endoderm lineage," *PLOS Genetics*, vol. 17, no. 7, p. e1009681–e1009681, Jul 2021.

[167] R. Weksberg, S. Choufani, D. Grafodatskaya, and D. BUTCHER, "Dna methylation markers for neurodevelopmental syndromes," 2016.

[168] M.-C. Tsai, O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal, and H. Y. Chang, "Long noncoding rna as modular scaffold of histone modification complexes," *Science*, vol. 329, no. 5992, p. 689–693, Jul 2010.

[169] X. Q. Wang and J. Dostie, "Reciprocal regulation of chromatin state and architecture byhotairm1contributes to temporal collinearhoxagene activation," *Nucleic Acids Research*, p. gkw966–gkw966, Oct 2016.

[170] K. Long, L. Moss, L. Laursen, L. Boulter, and C. ffrench Constant, "Integrin signalling regulates the expansion of neuroepithelial progenitors and neurogenesis via wnt7a and decorin," *Nature Communications*, vol. 7, no. 1, Feb 2016.

[171] K. Loulier, J. D. Lathia, V. Marthiens, J. Relucio, M. R. Mughal, S.-C. Tang, T. Coksaygan, P. E. Hall, S. Chigurupati, B. Patton, H. Colognato, M. S. Rao, M. P. Mattson, T. F. Haydar, and C. ffrench Constant, "β1 integrin maintains integrity of the embryonic neocortical stem cell niche," *PLoS Biology*, vol. 7, no. 8, p. e1000176, Aug 2009.

[172] T. M. Bosley, I. A. Alorainy, M. A. Salih, H. Aldhalaan, K. K. Abu-Amero, D. T. Oystreck, M. A. Tischfield, E. C. Engle, and R. P. Erickson, "The clinical spectrum of homozygoushoxa1 mutations," *American Journal of Medical Genetics - Part A*, vol. 146A, no. 10, p. 1235–1240, Jan 2008.

[173] N. Makki and M. R. Capecchi, "Cardiovascular defects in a mouse model of hoxa1 syndrome," *Human Molecular Genetics*, vol. 21, no. 1, p. 26–31, Sep 2011.

[174] ——, "Identification of novel hoxa1 downstream targets regulating hindbrain, neural crest and inner ear development," *Developmental Biology*, vol. 357, no. 2, p. 295–304, Sep 2011.

[175] M. Song, J. Giza, C. C. Proenca, D. Jing, M. Elliott, I. Dincheva, S. V. Shmelkov, J. Kim, R. Schreiner, S.-H. Huang, E. Castrén, R. Prekeris, B. L. Hempstead, M. V. Chao, J. B. Dictenberg, S. Rafii, Z.-Y. Chen, E. Rodriguez-Boulan, and F. S. Lee, "Slitrk5 mediates bdnf-dependent trkb receptor trafficking and signaling," *Developmental cell*, vol. 33, no. 6, p. 690–702, Jun 2015.

[176] E. J. Huang and L. F. Reichardt, "Neurotrophins: Roles in neuronal development and function," *Annual Review of Neuroscience*, vol. 24, no. 1, p. 677–736, Mar 2001.

[177] R. Barshir, S. Fishilevich, T. Iny-Stein, O. Zelig, Y. Mazor, Y. Guan-Golan, M. Safran, and D. Lancet, "Genecarna: A comprehensive gene-centric database of human non-coding rnas in the genecards suite," *Journal of Molecular Biology*, vol. 433, no. 11, p. 166913, Mar 2021.

[178] A. Melluso, F. Secondulfo, G. Capolongo, G. Capasso, and M. Zacchia, "Bardet-biedl syndrome: Current perspectives and clinical outlook," *Therapeutics and Clinical Risk Management*, vol. Volume 19, p. 115–132, Jan 2023.

[179] T. L. Young-Pearse, J. Bai, R. Chang, J. B. Zheng, J. J. LoTurco, and D. J. Selkoe, "A critical function for -amyloid precursor protein in neuronal migration revealed by in utero rna interference," *Journal of Neuroscience*, vol. 27, no. 52, p. 14459–14469, Dec 2007.

[180] M. Jeanne, C. Labelle-Dumais, J. Jorgensen, W. B. Kauffman, G. M. Mancini, J. Favor, V. Valant, S. M. Greenberg, J. Rosand, and D. B. Gould, "Col4a2 mutations impair col4a1 and col4a2 secretion and cause hemorrhagic stroke," *The American Journal of Human Genetics*, vol. 90, no. 1, p. 91–101, Jan 2012.

[181] J. Nie, Y. Ueda, A. J. Solivais, and E. Hashino, "Chd7 regulates otic lineage specification and hair cell differentiation in human inner ear organoids," *Nature Communications*, vol. 13, no. 1, Nov 2022.

[182] D. Wallis, "The zinc finger transcription factor gfi1, implicated in lymphomagenesis, is required for inner ear hair cell differentiation and survival," *Development*, vol. 130, no. 1, p. 221–232, Jan 2003.

[183] H. Tsuda, H. Jafar-Nejad, A. J. Patel, Y. Sun, H.-K. Chen, M. F. Rose, K. J. Venken, J. Botas, H. T. Orr, H. J. Bellen, and H. Y. Zoghbi, "The axh domain of ataxin-1 mediates neurodegeneration through its interaction with gfi-1/senseless proteins," *Cell*, vol. 122, no. 4, p. 633–644, Aug 2005.

[184] A. Kazanjian, D. Wallis, N. Au, R. Nigam, Koen, P. T. Cagle, B. F. Dickey, H. J. Bellen, C. B. Gilks, and H. L. Grimes, "Growth factor independence-1 is expressed in primary human neuroendocrine lung carcinomas and mediates the differentiation of murine pulmonary neuroendocrine cells," *Cancer Research*, vol. 64, no. 19, p. 6874–6882, Oct 2004.

[185] M. Bjerknes and H. Cheng, "Cell lineage metastability in gfi1-deficient mouse intestinal epithelium," *Developmental Biology*, vol. 345, no. 1, p. 49–63, Sep 2010.

[186] S. P. Gray, A. M. Shah, and I. Smyrnias, "Nadph oxidase 4 and its role in the cardiovascular system," *Vascular Biology*, vol. 1, no. 1, p. H59–H66, Aug 2019.

[187] M. Zhang, A. C. Brewer, K. Schröder, C. X. C. Santos, D. J. Grieve, M. Wang, N. Anilkumar, B. Yu, X. Dong, S. J. Walker, R. P. Brandes, and A. M. Shah, "Nadph oxidase-4 mediates protection against chronic load-induced stress in mouse hearts by enhancing angiogenesis," *Proceedings of the National Academy of Sciences*, vol. 107, no. 42, p. 18121–18126, Oct 2010.

[188] T. Hisaoka, T. Komori, K. Fujimoto, T. Kitamura, and Y. Morikawa, "Comprehensive expression pattern of kin of irregular chiasm-like 3 in the adult mouse brain," *Biochemical and Biophysical Research Communications*, vol. 563, p. 66–72, Jul 2021.

[189] Y. Matsunaga, M. Noda, H. Murakawa, H. Kanehiro, A. Nagasaka, S. Inoue, T. Miyata, T. Miura, K. I. Kubo, and K. Nakajima, "Reelin transiently promotes n-cadherin–dependent neuronal adhesion during mouse cortical development," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 8, p. 2048–2053, Feb 2017.

[190] S. Kusuhara, Y. Fukushima, S. Fukuhara, L. M. Jakt, M. Okada, Y. Shimizu, M. Hata, K. Nishida, A. Negi, M. Hirashima, N. Mochizuki, S.-I. Nishikawa, and A. Uemura, "Arhgef15 promotes retinal angiogenesis by mediating vegf-induced cdc42 activation and potentiating rhoj inactivation in endothelial cells," *PLoS ONE*, vol. 7, no. 9, p. e45858, Sep 2012.

[191] J. M. Cesario, A. A. Almaidhan, and J. Jeong, "Expression of forkhead box transcription factor genes foxp1 and foxp2 during jaw development," *Gene Expression Patterns*, vol. 20, no. 2, p. 111–119, Mar 2016.

[192] J.-R. van Rhijn and S. C. Vernes, "Retinoic acid signaling: A new piece in the spoken language puzzle," *Frontiers in Psychology*, vol. 6, Nov 2015.

[193] M. P. Schnetz, C. F. Bartels, K. Shastri, D. Balasubramanian, G. E. Zentner, R. Balaji, X. Zhang, L. Song, B. Wang, T. LaFramboise, G. P. Crawford, and P. C. Scacheri, "Genomic distribution of chd7 on chromatin tracks h3k4 methylation patterns," *Genome Research*, vol. 19, no. 4, p. 590–601, Apr 2009.

[194] E. Montanez, S. A. Wickström, J. Altstätter, H. Chu, and R. Fässler, "α-parvin controls vascular mural cell recruitment to vessel wall by regulating rhoa/rock signalling," *The EMBO Journal*, vol. 28, no. 20, p. 3132–3144, Oct 2009.

[195] T. Okumura, "Role of lipid droplet proteins in liver steatosis," *Journal of Physiology and Biochemistry*, vol. 67, no. 4, p. 629–636, Dec 2011.

[196] Q. Zhai, M. Luo, Y. Zhang, W. Zhang, C. Wu, S. Lv, and Q. Wei, "Histone methyltransferase kmt2d mediated lipid metabolism via peroxisome proliferator-activated receptor gamma in prostate cancer," *Translational Cancer Research*, vol. 11, no. 8, p. 2607–2621, Aug 2022.

[197] G. A. Carosso, L. Boukas, J. J. Augustin, H. N. Nguyen, B. L. Winer, G. H. Cannon, J. D. Robertson, L. Zhang, K. D. Hansen, L. A. Goff, and H. T. Bjornsson, "Precocious neuronal differentiation and disrupted oxygen responses in kabuki syndrome," *JCI Insight*, vol. 4, no. 20, Oct 2019.

[198] K. L. Ciprero, J. Clayton-Smith, D. Donnai, R. A. Zimmerman, E. H. Zackai, and J. E. Ming, "Symptomatic chiari i malformation in kabuki syndrome," *American Journal of Medical Genetics Part A*, vol. 132A, no. 3, p. 273–275, Nov 2004.

[199] S. Yano, T. Matsuishi, M. Yoshino, H. Kato, and K. Kojima, "Cerebellar and brainstem "atrophy" in a patient with kabuki make-up syndrome," *American Journal of Medical Genetics*, vol. 71, no. 4, p. 486–487, Sep 1997.

[200] S. Glaser, J. Schaft, S. Lubitz, K. Vintersten, F. van der Hoeven, K. R. Tufteland, R. Aasland, K. Anastassiadis, S.-L. Ang, and A. F. Stewart, "Multiple epigenetic maintenance factors implicated by the loss of mll2 in mouse development," *Development*, vol. 133, no. 8, p. 1423–1432, Apr 2006.

[201] A. Garcia-Gasca and D. D. Spyropoulos, "Differential mammary morphogenesis along the anteroposterior axis in hoxc6 gene targeted mice," *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, vol. 219, no. 2, p. 261–276, Oct 2000.

[202] J.-E. Moon, S.-J. Lee, and C. W. Ko, "A de novo kmt2d mutation in a girl with kabuki syndrome associated with endocrine symptoms: a case report," *BMC Medical Genetics*, vol. 19, no. 1, Jun 2018.

[203] S. Banka, R. Veeramachaneni, W. Reardon, E. Howard, S. Bunstone, N. Ragge, M. J. Parker, Y. J. Crow, B. Kerr, H. Kingston, K. Metcalfe, K. Chandler, A. Magee, F. Stewart, V. P. M. McConnell, D. E. Donnelly, S. Berland, G. Houge, J. E. Morton, and C. Oley, "How genetically heterogeneous is kabuki syndrome?: Mll2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum," *European Journal of Human Genetics*, vol. 20, no. 4, p. 381–388, Nov 2011.

[204] Y. Kuroki, N. Katsumata, T. Eguchi, Y. Fukushima, S. Suwa, and T. Kajii, "Precocious puberty in kabuki makeup syndrome," *The Journal of Pediatrics*, vol. 110, no. 5, p. 750–752, May 1987.

[205] M. d. l. A. Serrano, B. L. Demarest, T. Tone-Pah-Hote, M. Tristani-Firouzi, and H. J. Yost, "Inhibition of notch signaling rescues cardiovascular development in kabuki syndrome," *PLOS Biology*, vol. 17, no. 9, p. e3000087, Sep 2019.

[206] A. T. S. Wyse and C. A. Netto, "Behavioral and neurochemical effects of proline," *Metabolic Brain Disease*, vol. 26, no. 3, p. 159–172, Jun 2011.

[207] P. Roussos, S. G. Giakoumaki, and P. Bitsios, "A risk prodh haplotype affects sensorimotor gating, memory, schizotypy, and anxiety in healthy male subjects," *Biological Psychiatry*, vol. 65, no. 12, p. 1063–1070, Jun 2009.

[208] C. L. Clelland, L. L. Read, A. N. Baraldi, C. P. Bart, C. A. Pappas, L. J. Panek, R. H. Nadrich, and J. D. Clelland, "Evidence for association of hyperprolinemia with schizophrenia and a measure of clinical outcome," *Schizophrenia Research*, vol. 131, no. 1-3, p. 139–145, Sep 2011.

[209] D. Schulz, J. Morschel, S. Schuster, V. Eulenburg, and J. Gomeza, "Inactivation of the mouse l-proline transporter prot alters glutamatergic synapse biochemistry and perturbs behaviors required to respond to environmental changes," *Frontiers in Molecular Neuroscience*, vol. 11, Aug 2018.

[210] C. A. Hoeffer and E. Klann, "mtor signaling: At the crossroads of plasticity, memory, and disease," *Trends in neurosciences*, vol. 33, no. 2, p. 67, Feb 2010.

[211] C. A. Stanley, "Hyperinsulinism/hyperammonemia syndrome: insights into the regulatory role of glutamate dehydrogenase in ammonia metabolism," *Molecular Genetics and Metabolism*, vol. 81, p. 45–51, Apr 2004.

[212] I. Issaeva, Y. Zonis, T. Rozovskaia, K. Orlovsky, C. M. Croce, T. Nakamura, A. Mazo, L. Eisenbach, and E. Canaani, "Knockdown of alr (mll2) reveals alr target genes and leads to alterations in cell adhesion and growth," *Molecular and Cellular Biology*, vol. 27, no. 5, p. 1889–1903, Mar 2007.

[213] D. Muñoz-Espín, M. Cañamero, A. Maraver, G. Gómez-López, J. Contreras, S. Murillo-Cuesta, A. Rodríguez-Baeza, I. Varela-Nieto, J. Ruberte, M. Collado, and M. Serrano, "Programmed cell senescence during mammalian embryonic development," *Cell*, vol. 155, no. 5, p. 1104–18, 2013.

[214] M. Storer, A. Mas, A. Robert-Moreno, M. Pecoraro, M. C. Ortells, V. Di Giacomo, R. Yosef, N. Pilpel, V. Krizhanovsky, J. Sharpe, and W. M. Keyes, "Senescence is a developmental mechanism that contributes to embryonic growth and patterning," *Cell*, vol. 155, no. 5, p. 1119–1130, Nov 2013.

[215] N. Spassky and A. Meunier, "The development and functions of multiciliated epithelia," *Nature Reviews Molecular Cell Biology*, vol. 18, no. 7, p. 423–436, Apr 2017.

[216] Z. Guo, F. Liu, and H. J. Li, "Novel kdm6a splice-site mutation in kabuki syndrome with congenital hydrocephalus: a case report," *BMC Medical Genetics*, vol. 19, no. 1, Dec 2018.

[217] J. F. Reiter and M. R. Leroux, "Genes and molecular pathways underpinning ciliopathies," *Nature Reviews Molecular Cell Biology*, vol. 18, no. 9, p. 533–547, Jul 2017.

[218] M. C. Digilio, M. Gnazzo, F. Lepri, M. L. Dentici, E. Pisaneschi, A. Baban, C. Passarelli, R. Capolino, A. Angioni, A. Novelli, B. Marino, and B. Dallapiccola, "Congenital heart defects in molecularly proven kabuki syndrome patients," *American Journal of Medical Genetics Part A*, vol. 173, no. 11, p. 2912–2922, Sep 2017.

[219] E. Giuili, R. Grolaux, C. Macedo, L. Desmyter, B. Pichon, S. Neuens, C. Vilain, C. Olsen, S. Van Dooren, G. Smits, and M. Defrance, "Comprehensive evaluation of the implementation of episignatures for diagnosis of neurodevelopmental disorders (ndds)," *Human Genetics*, vol. 142, no. 12, p. 1721–1735, Oct 2023.

[220] K. Shuai and B. Liu, "Regulation of gene-activation pathways by pias proteins in the immune system," *Nature Reviews Immunology*, vol. 5, no. 8, p. 593–605, Aug 2005.

[221] N. Gaiano, J. S. Nye, and G. Fishell, "Radial glial identity is promoted by notch1 signaling in the murine forebrain," *Neuron*, vol. 26, no. 2, p. 395–404, May 2000.

[222] C. Angulo-Rojo, R. Manning□Cela, A. Aguirre, A. Ortega, and E. López-Bayghen, "Involvement of the notch pathway in terminal astrocytic differentiation: Role of pka," *Asn Neuro*, vol. 5, no. 5, p. AN20130023–AN20130023, Nov 2013.

[223] R. Carapito, E. Ivanova, A. Morlon, L. Meng, A. Molitor, □. Erdmann, B. Kieffer, A. Pichot, L. Naegely, A. Kolmer, N. Paul, A. Hanauer, F. T. Mau-Them, N. Jean-Marçais, S. M. Hiatt, G. M. Cooper, T. Tvrdik, A. M. Muir, C. Dimartino, and M. Chopra, "Zmiz1 variants cause a syndromic neurodevelopmental disorder," *American Journal of Human Genetics*, vol. 106, no. 1, p. 137–137, Jan 2020.

[224] S.-B. Liu, X. Meng, Y. Li, J.-M. Wang, H. Guo, C. Wang, and B.-M. Zhu, "Histone methyltransferase kmt2d contributes to the protection of myocardial ischemic injury," *Frontiers in Cell and Developmental Biology*, vol. 10, Jul 2022.

[225] E. Aref-Eshghi, L. C. Schenkel, H. Lin, C. Skinner, P. Ainsworth, G. Paré, D. Rodenhiser, C. Schwartz, and B. Sadikovic, "The defining dna methylation signature of kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance," *Epigenetics*, vol. 12, no. 11, p. 923–933, Nov 2017.

[226] Y. Hamaguchi, H. Mishima, T. Kawai, S. Saitoh, K. Hata, A. Kinoshita, and K.-I. Yoshiura, "Identification of unique dna methylation sites in kabuki syndrome using whole genome bisulfite sequencing and targeted hybridization capture followed by enzymatic methylation sequencing," *Journal of Human Genetics*, vol. 67, no. 12, p. 711–720, Dec 2022.

[227] T. Vitali, S. Girald-Berlingeri, P. A. Randazzo, and P.-W. Chen, "Arf gaps: A family of proteins with disparate functions that converge on a common structure, the integrin adhesion complex," *Small GTPases*, p. 1–9, Mar 2017.

[228] Y. Zhu, Y. Wu, J. I. Kim, Z. Wang, Y. Daaka, and Z. Nie, "Arf gtpase-activating protein agap2 regulates focal adhesion kinase activity and focal adhesion remodeling," *Journal of Biological Chemistry*, vol. 284, no. 20, p. 13489–13496, May 2009.

[229] Q. Qi, S. H. Kang, S. Zhang, C. D. Pham, H. Fu, D. J. Brat, and K. Ye, "Co-amplification of phosphoinositide 3-kinase enhancer a and cyclin-dependent kinase 4 triggers glioblastoma progression," *Oncogene*, vol. 36, no. 32, p. 4562–4572, Apr 2017.

[230] S. Nakken, Øystein Eikrem, H. Marti, C. Beisland, L. Bostad, A. Scherer, A. Flatberg, V. Beisvåg, E. Skandalou, J. Furriol, and P. Strauss, "Agap2-as1 as a prognostic biomarker in low-risk clear cell renal cell carcinoma patients with progressing disease," *Cancer Cell International*, vol. 21, no. 1, Dec 2021.

[231] M. M. Edwards, D. S. McLeod, R. Grebe, C. Heng, O. Lefèbvre, and G. A. Lutty, "Lama1 mutations lead to vitreoretinal blood vessel formation, persistence of fetal vasculature, and epiretinal membrane formation in mice," *BMC Developmental Biology*, vol. 11, no. 1, Oct 2011.

[232] C. D. Bryan, C.-B. Chien, and K. M. Kwan, "Loss of laminin alpha 1 results in multiple structural defects and divergent effects on adhesion during vertebrate optic cup morphogenesis," *Developmental Biology*, vol. 416, no. 2, p. 324–337, Aug 2016.

[233] F. M. C. Medina, P. Rodriguez, R. T. d. S. Cabral, M. B. d. Castro, J. C. Llerena Junior, and R. M. Japiassú, "Kabuki syndrome: a case report with severe ocular abnormalities," *Revista Brasileira de Oftalmologia*, vol. 72, no. 5, p. 341–343, Oct 2013.

[234] Y. Kwon, S. J. Lee, E. Lee, D. Kim, and D. Park, "β pix heterozygous mice have defects in neuronal morphology and social interaction," *Biochemical and Biophysical Research Communications*, vol. 516, no. 4, p. 1204–1210, Sep 2019.

[235] A. Orsini, A. Bonuccelli, P. Striano, A. Azzarà, G. Costagliola, R. Consolini, D. Peroni, A. Valetto, and V. Bertini, "Generalized epilepsy and mild intellectual disability associated with 13q34 deletion: A potential role for sox1 and arhgef7," *Seizure: European Journal of Epilepsy*, vol. 59, p. 38–40, Jul 2018.

[236] M. D. Bates, D. T. Dunagan, L. C. Welch, A. Kaul, and R. P. Harvey, "The hlx homeobox transcription factor is required early in enteric nervous system development," *BMC Developmental Biology*, vol. 6, no. 1, p. 33, 2006.

[237] W. I. Wujcicka, M. Kacerovsky, M. Krekora, P. Kaczmarek, B. Leśniczak, and M. Grzesiak, "Rs868058 in the homeobox gene hlx contributes to early-onset fetal growth restriction," *Biology*, vol. 11, no. 3, p. 447, Mar 2022.

[238] D. A. Schott, M. J. Blok, W. J. M. Gerver, K. Devriendt, L. J. I. Zimmermann, and C. T. R. M. Stumpel, "Growth pattern in kabuki syndrome with akmt2dmutation," *American Journal of Medical Genetics Part A*, vol. 170, no. 12, p. 3172–3179, Aug 2016.

[239] K. M. Hirschi, K. Y. F. Tsai, T. Davis, J. Clark, M. Knowlton, B. T. Bikman, P. Reynolds, and J. A. Arroyo, "Growth arrest-specific protein-6/axl signaling induces preeclampsia in rats†," *Biology of Reproduction*, vol. 21, no. 2, Jul 2019.

[240] N. Shiba, R. A. Daza, L. G. Shaffer, A. J. Barkovich, W. B. Dobyns, and R. F. Hevner, "Neuropathology of brain and spinal malformations in a case of monosomy 1p36," *Acta Neuropathologica Communications*, vol. 1, no. 1, Aug 2013.

[241] L. Tan, D. Fu, F. Liu, J. Liu, Y. Zhang, X. Li, J. Gao, K. Tao, G. Wang, L. Wang, and Z. Wang, "Mxra8 is an immune-relative prognostic biomarker associated with metastasis and cd8+ t cell infiltration in colorectal cancer," *Frontiers in Oncology*, vol. 12, Jan 2023.

[242] Y.-T. Lin and K.-J. Wu, "Epigenetic regulation of epithelial-mesenchymal transition: focusing on hypoxia and tgf-β signaling," *Journal of Biomedical Science*, vol. 27, no. 1, Mar 2020.

[243] L. Baskin, A. Sinclair, A. Derpinghaus, M. Cao, Y. Li, M. Overland, S. Aksel, and G. R. Cunha, "Estrogens and development of the mouse and human external genitalia," *Differentiation*, vol. 118, p. 82–106, Mar 2021.

[244] L. Baskin, M. Cao, A. Sinclair, Y. Li, M. Overland, D. Isaacson, and G. R. Cunha, "Androgen and estrogen receptor expression in the developing human penis and clitoris," *Differentiation*, vol. 111, p. 41–59, Jan 2020.

[245] R. Mo, S. Rao, and Y.-J. Zhu, "Identification of the mll2 complex as a coactivator for estrogen receptor α," *Journal of Biological Chemistry*, vol. 281, no. 23, p. 15714–15720, Jun 2006.

[246] N. Sobreira, M. Brucato, L. Zhang, C. Ladd-Acosta, C. M. Ongaco, J. Romm, K. F. Doheny, R. C. Mingroni-Netto, D. R. Bertola, C. J. Kim, A. Beatriz, M. I. Melaragno, D. Valle, V. A. Meloni, and H. T. Bjornsson, "Patients with a kabuki syndrome phenotype demonstrate dna methylation abnormalities," *European Journal of Human Genetics*, vol. 25, no. 12, p. 1335–1344, Nov 2017.

[247] M. Dauar, C. Picard, P. Rosa-Neto, S. Villeneuve, and J. Poirier, "Cntn5 is associated with disease risk and pathology throughout the alzheimer's disease continuum." *Alzheimer's and Dementia*, vol. 17 Suppl 3, p. e052359–e052359, Dec 2021.

[248] M. Sertcelik, C. Ugur, A. Sahin Akozel, and C. K. Gurkan, "A child with kabuki syndrome and autism spectrum disorder," *Noro Psikiyatri Arsivi*, vol. 53, no. 3, p. 280–282, Aug 2016.

[249] M. R. Taylor, E. A. Martin, B. Sinnen, R. Trilokekar, E. Ranza, S. E. Antonarakis, and M. E. Williams, "Kirrel3-mediated synapse formation is attenuated by disease-associated missense variants," *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 40, no. 28, p. 5376–5388, Jul 2020.

[250] Janet, A. Brignall, T. Cutforth, K. Shen, and J.-F. Cloutier, "Kirrel3 is required for the coalescence of vomeronasal sensory neuron axons into glomeruli and for male-male aggression," *Development*, vol. 140, no. 11, p. 2398–2408, Jun 2013.

[251] R. Visser, Ellie, J. J. Goeman, J. M. Wit, and M. Karperien, "Sotos syndrome is associated with deregulation of the mapk/erk-signaling pathway," *PLoS ONE*, vol. 7, no. 11, p. e49229–e49229, Nov 2012.

[252] S. Murakami, "Constitutive activation of mek1 in chondrocytes causes stat1-independent achondroplasia-like dwarfism and rescues the fgfr3-deficient mouse phenotype," *Genes and Development*, vol. 18, no. 3, p. 290–305, Feb 2004.

[253] M. Rio, "Spectrum of nsd1 mutations in sotos and weaver syndromes," *Journal of Medical Genetics*, vol. 40, no. 6, p. 436–440, Jun 2003.

[254] E. K. Kim and E.-J. Choi, "Pathological roles of mapk signaling pathways in human diseases," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1802, no. 4, p. 396–405, Apr 2010.

[255] K. Goodwin and C. M. Nelson, "Branching morphogenesis," *Development*, vol. 147, no. 10, May 2020.

[256] Y. Huybrechts, G. Mortier, E. Boudin, and W. V. Hul, "Wnt signaling and bone: Lessons from skeletal dysplasias and disorders," *Frontiers in Endocrinology*, vol. 11, Apr 2020.

[257] M. Mahmoudabady, M. Mathieu, L. Dewachter, I. Hadad, L. Ray, P. Jespers, S. Brimioulle, R. Naeije, and K. McEntee, "Activin-a, transforming growth factor-β, and myostatin signaling pathway in experimental dilated cardiomyopathy," *Journal of Cardiac Failure*, vol. 14, no. 8, p. 703–709, Oct 2008.

[258] K. Tsuchida, M. Nakatani, K. Hitachi, A. Uezumi, Y. Sunada, H. Ageta, and K. Inokuchi, "Activin signaling as an emerging target for therapeutic interventions," *Cell Communication and Signaling*, vol. 7, no. 1, Jun 2009.

[259] J. Wang, S. Liu, T. Heallen, and J. F. Martin, "The hippo pathway in the heart: pivotal roles in development, disease, and regeneration," *Nature Reviews Cardiology*, vol. 15, no. 11, p. 672–684, Aug 2018.

[260] S. Ma and K. Guan, "Polycystic kidney disease: a hippo connection," *Genes and Development*, vol. 32, no. 11-12, p. 737–739, Jun 2018.

[261] A. Mongera, A. Michaut, C. Guillot, F. Xiong, and O. Pourquié, "Mechanics of anteroposterior axis formation in vertebrates," *Annual Review of Cell and Developmental Biology*, vol. 35, no. 1, p. 259–283, Oct 2019.

[262] J. Gerhart, M. Danilchik, T. Doniach, S. Roberts, B. Rowning, and R. Stewart, "Cortical rotation of the xenopus egg: consequences for the anteroposterior pattern of embryonic dorsal development," *Development*, vol. 107, no. Supplement, p. 37–51, Apr 1989.

[263] V.-H. Lourdes, S.-C. Mario, C.-A. Didac, B. Mercè, M. J. Di, P. Leticia, F. Lucía, A. Martinez□Monseny, and S. Mercedes, "Beyond the known phenotype of sotos syndrome: a 31-individuals cohort study," *Frontiers in Pediatrics*, vol. 11, Jun 2023.

[264] C. Gerald Leonard, S. Ranguis, S. Lynn Cushing, S. Blaser, and A. James, "Cholesteatoma in children with sotos syndrome," *The Journal of International Advanced Otology*, vol. 18, no. 2, p. 139–144, Apr 2022.

[265] P. Kaur, H. Jin, J. Lusk, and N. Tolwinski, "Modeling the role of wnt signaling in human and drosophila stem cells," *Genes*, vol. 9, no. 2, p. 101, Feb 2018.

[266] N. Mosaddeghzadeh and M. R. Ahmadian, "The rho family gtpases: Mechanisms of regulation and signaling," *Cells*, vol. 10, no. 7, p. 1831, Jul 2021.

[267] A. B. Jaffe and A. Hall, "Rho gtpases: Biochemistry and biology," *Annual Review of Cell and Developmental Biology*, vol. 21, no. 1, p. 247–269, Nov 2005.

[268] S. I. J. Ellenbroek and J. G. Collard, "Rho gtpases: functions and association with cancer," *Clinical and Experimental Metastasis*, vol. 24, no. 8, p. 657–672, Nov 2007.

[269] E. C. Lai, "Notch signaling: control of cell communication and cell fate," *Development*, vol. 131, no. 5, p. 965–973, Mar 2004.

[270] A. L. Penton, L. D. Leonard, and N. B. Spinner, "Notch signaling in human development and disease," *Seminars in cell and developmental biology*, vol. 23, no. 4, p. 450–457, Jun 2012.

[271] R. N. Wang, J. Green, Z. Wang, Y. Deng, M. Qiao, M. Peabody, Q. Zhang, J. Ye, Z. Yan, S. Denduluri, O. Idowu, M. Li, C. Shen, A. Hu, R. C. Haydon, R. Kang, J. Mok, M. J. Lee, H. L. Luu, and L. L. Shi, "Bone morphogenetic protein (bmp) signaling in development and human diseases," *Genes and Diseases*, vol. 1, no. 1, p. 87–105, Sep 2014.

[272] J. Soto, Y. Song, Y. Wu, C. Bin-ru, H. Park, N. Akhtar, P. Wang, T. Hoffman, C. Ly, J. Sia, S. Y. Wong, D. Kelkhoff, J. Chu, M. Poo, T. L. Downing, A. C. Rowat, and S. Li, "Reduction of intracellular tension and cell adhesion promotes open chromatin structure and enhances cell reprogramming," *Advanced Science*, vol. 10, no. 24, Jun 2023.

[273] O. Rossier and G. Giannone, "The journey of integrins and partners in a complex interactions landscape studied by super-resolution microscopy and single protein tracking," *Experimental Cell Research*, vol. 343, no. 1, p. 28–34, Apr 2016.

[274] J. Rossy, S. V. Pageon, D. M. Davis, and K. Gaus, "Super-resolution microscopy of the immunological synapse," *Current Opinion in Immunology*, vol. 25, no. 3, p. 307–312, Jun 2013.

[275] M. Maglione and S. J. Sigrist, "Seeing the forest tree by tree: super-resolution light microscopy meets the neurosciences," *Nature Neuroscience*, vol. 16, no. 7, p. 790–797, Jun 2013.

[276] W. Zhuang, X. Ge, S. Yang, M. Huang, W. Zhuang, P. Chen, X.-H. Zhang, J. Fu, J. Qu, and B.-Z. Li, "Upregulation of lncrna meg3 promotes osteogenic differentiation of mesenchymal stem cells from multiple myeloma patients by targeting bmp4 transcription," *Stem Cells*, vol. 33, no. 6, p. 1985–1997, Jun 2015.

[277] K. Yamane, K. Tateishi, R. J. Klose, J.-Y. Fang, L. A. Fabrizio, H. Erdjument-Bromage, J. Taylor-Papadimitriou, P. Tempst, and Y. Zhang, "Plu-1 is an h3k4 demethylase involved in transcriptional repression and breast cancer cell proliferation," *Molecular Cell*, vol. 25, no. 6, p. 801–812, Mar 2007.

[278] L. P. Blair, J. Cao, M. R. Zou, J. Sayegh, and Q. Yan, "Epigenetic regulation by lysine demethylase 5 (kdm5) enzymes in cancer," *Cancers*, vol. 3, no. 1, p. 1383–1404, Mar 2011.

[279] B. L. Kidder, G. Hu, and K. Zhao, "Kdm5b focuses h3k4 methylation near promoters and enhancers during embryonic stem cell self-renewal and differentiation," *Genome Biology*, vol. 15, no. 2, p. R32, 2014.

[280] B. K. Dey, L. Stalker, A. Schnerch, M. Bhatia, J. Taylor-Papidimitriou, and C. Wynder, "The histone demethylase kdm5b/jarid1b plays a role in cell fate decisions by blocking terminal differentiation," *Molecular and Cellular Biology*, vol. 28, no. 17, p. 5312–5327, Sep 2008.

[281] S. U. Schmitz, M. Albert, M. Malatesta, L. Morey, J. V. Johansen, M. Bak, N. Tommerup, I. Abarrategui, and K. Helin, "Jarid1b targets genes regulating development and is involved in neural differentiation," *The EMBO Journal*, vol. 30, no. 22, p. 4586–4600, Oct 2011.

[282] Q. Zhou, E. A. Obana, K. L. Radomski, G. Sukumar, C. Wynder, C. L. Dalgard, and M. L. Doughty, "Inhibition of the histone demethylase kdm5b promotes neurogenesis and derepresses reln (reelin)

in neural stem cells from the adult subventricular zone of mice," *Molecular Biology of the Cell*, vol. 27, no. 4, p. 627–639, Feb 2016.

[283] L. Xie, C. Pelz, W. Wang, A. Bashar, O. Varlamova, S. Shadle, and S. Impey, "Kdm5b regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription," *The EMBO Journal*, vol. 30, no. 8, p. 1473–1484, Mar 2011.

[284] F. Tessadori, K. Duran, K. Knapp, M. Fellner, S. Smithson, A. Beleza-Meireles, M. W. Elting, Q. Waisfisz, A. O'Donnell-Luria, C. Nowak, J. Douglas, A. Ronan, T. Brunet, U. Kotzaeridou, S. Svihovec, M. Saenz, I. Thiffault, F. D. Viso, P. Devine, and S. Rego, "Recurrent de novo missense variants across multiple histone h4 genes underlie a neurodevelopmental syndrome," *American Journal of Human Genetics*, vol. 109, no. 4, p. 750–758, Apr 2022.

[285] E. A. Middleton, J. W. Rowley, R. A. Campbell, C. K. Grissom, S. M. Brown, S. J. Beesley, H. Schwertz, Y. Kosaka, B. K. Manne, K. Krauel, N. D. Tolley, A. S. Eustes, L. Guo, R. Paine, E. S. Harris, G. A. Zimmerman, A. S. Weyrich, and M. T. Rondina, "Sepsis alters the transcriptional and translational landscape of human and murine platelets," *Blood*, vol. 134, no. 12, p. 911–923, Sep 2019.

[286] R. Corrado, A. F. Wilson, C. Tello, M. Noel, E. Galaretto, and E. Bersusky, "Sotos syndrome and scoliosis surgical treatment: a 10-year follow-up," *European Spine Journal*, vol. 20, no. S2, p. 271–277, Jan 2011.

[287] D. Jm, S. B. Gibson, S. Tsetsou, K. Russell, Keefe, F. Kp, B. Mb, M. Lc, B. Jl, P. Sm, and L. B. Jorde, "Loss oftp73function contributes to amyotrophic lateral sclerosis pathogenesis," *bioRxiv (Cold Spring Harbor Laboratory)*, Oct 2018.

[288] X. Zhang, M. Shi, X. Zhao, E. Bin, Y. Hu, N. Tang, H. Dai, and C. Wang, "Telomere shortening impairs alveolar regeneration," *Cell Proliferation*, vol. 55, no. 4, Mar 2022.

[289] N. Liu, Y. Li, N. Wu, W. Zhou, J. Huang, R. Li, L. Zhou, and R. Hu, "Interaction of tppp3 with vdac1 promotes endothelial injury through activation of reactive oxygen species," *Oxidative Medicine and Cellular Longevity*, vol. 2020, p. 1–13, Oct 2020.

[290] O. Bogdanović, A. H. Smits, Elisa, J. J. Tena, E. Ford, R. M. Williams, U. Senanayake, M. D. Schultz, S. Hontelez, I. van Kruijsbergen, T. Rayón, F. Gnerlich, T. Carell, J. C. Veenstra, M. Manzanares, T. Sauka□Spengler, J. R. Ecker, M. Vermeulen, J. L. Gómez□Skarmeta, and R. Lister, "Active dna demethylation at enhancers during the vertebrate phylotypic period," *Nature Genetics*, vol. 48, no. 4, p. 417–426, Feb 2016.

[291] T. T. Koopmann, Y. Jamshidi, M. Naghibi-Sistani, H. M. van der Klift, H. Birjandi, Z. Al-Hassnan, A. Alwadai, G. Zifarelli, E. G. Karimiani, S. Sedighzadeh, A. Bahreini, N. Nouri, M. Peter, K. Watanabe, H. A. van Duyvenvoorde, C. A. L. Ruivenkamp, A. K. K. Teunissen, A. D. J. Ten Harkel, S. G. van Duinen, and M. C. Haak, "Biallelic loss of ldb3 leads to a lethal pediatric dilated cardiomyopathy," *European Journal of Human Genetics*, vol. 31, no. 1, p. 97–104, Jan 2023.

[292] S. Imgrund, D. Hartmann, H. Farwanah, M. Eckhardt, R. Sandhoff, J. Degen, V. Gieselmann, K. Sandhoff, and K. Willecke, "Adult ceramide synthase 2 (cers2)-deficient mice exhibit myelin

sheath defects, cerebellar degeneration, and hepatocarcinomas," *Journal of Biological Chemistry*, vol. 284, no. 48, p. 33549–33560, Nov 2009.

[293] M. Mosbech, A. S. B. Olsen, D. Neess, O. Ben□David, L. L. Klitten, J. Larsen, A. Sabers, J. Vissing, J. E. Nielsen, L. Hasholt, A. D. Klein, M. M. Tsoory, H. Hjalgrim, N. Tommerup, A. H. Futerman, R. S. Møller, and N. J. Færgeman, "Reduced ceramide synthase 2 activity causes progressive myoclonic epilepsy," *Annals of Clinical and Translational Neurology*, vol. 1, no. 2, p. 88–98, Jan 2014.

[294] J. Testori, B. Schweighofer, I. Helfrich, C. Sturtzel, K. Lipnik, S. Gesierich, P. Nasarre, R. Hofer-Warbinek, M. Bilban, H. G. Augustin, and E. Hofer, "The vegf-regulated transcription factor hlx controls the expression of guidance cues and negatively regulates sprouting of endothelial cells," *Blood*, vol. 117, no. 9, p. 2735–2744, Mar 2011.

[295] V. Colombaro, I. Jadot, A.-E. Declèves, V. Voisin, L. Giordano, I. Habsch, B. Flamion, and N. Caron, "Hyaluronidase 1 and hyaluronidase 2 are required for renal hyaluronan turnover," *Acta Histochemica*, vol. 117, no. 1, p. 83–91, Jan 2015.

[296] M. M. A. Muggenthaler, B. Chowdhury, S. N. Hasan, H. E. Cross, B. Mark, G. V. Harlalka, M. A. Patton, M. Ishida, E. R. Behr, S. Sharma, K. Zahka, E. Faqeih, B. Blakley, M. Jackson, M. Lees, V. Dolinsky, L. Cross, P. Stanier, C. Salter, and E. L. Baple, "Mutations in hyal2, encoding hyaluronidase 2, cause a syndrome of orofacial clefting and cor triatriatum sinister in humans and mice," *PLoS Genetics*, vol. 13, no. 1, Jan 2017.

[297] F. Trousse, A. Jemli, M. Silhol, E. Garrido, L. Crouzier, G. Naert, T. Maurice, and M. Rossel, "Knockdown of the cxcl12/cxcr7 chemokine pathway results in learning deficits and neural progenitor maturation impairment in mice," *Brain Behavior and Immunity*, vol. 80, p. 697–710, Aug 2019.

[298] Y. Yan, J. Su, and Z. Zhang, "The cxcl12/cxcr4/ackr3 response axis in chronic neurodegenerative disorders of the central nervous system: Therapeutic target and biomarker," *Cellular and Molecular Neurobiology*, vol. 42, no. 7, p. 2147–2156, Jun 2021.

[299] Y. Takabatake, T. Sugiyama, H. Kohara, T. Matsusaka, H. Kurihara, P. A. Koni, Y. Nagasawa, T. Hamano, I. Matsui, N. Kawada, E. Imai, T. Nagasawa, H. Rakugi, and Y. Isaka, "The cxcl12 (sdf-1)/cxcr4 axis is essential for the development of renal vasculature," *Journal of the American Society of Nephrology*, vol. 20, no. 8, p. 1714–1723, May 2009.

[300] M. Zheng, S. H. Oh, N. Choi, Y. J. Choi, J. Kim, and J.-H. Sung, "Cxcl12 inhibits hair growth through cxcr4," *Biomedicine and Pharmacotherapy*, vol. 150, p. 112996, Jun 2022.

[301] K. Golan, O. Kollet, and T. Lapidot, "Dynamic cross talk between s1p and cxcl12 regulates hematopoietic stem cells migration, development and bone remodeling," *Pharmaceuticals*, vol. 6, no. 9, p. 1145–1169, Sep 2013.

[302] M. Grunert, C. Dorn, H. Cui, I. Dunkel, K. Schulz, S. Schoenhals, W. Sun, F. Berger, W. Chen, and S. R. Sperling, "Comparative dna methylation and gene expression analysis identifies novel genes for structural congenital heart diseases," *Cardiovascular Research*, vol. 112, no. 1, p. 464–477, Aug 2016.

[303] R. Mei, W. Qiu, Y. Yang, S. Xu, Y. Rao, Q. Li, Y. Luo, H. Huang, A. Yang, H. Tao, M. Qiu, and X. Zhao, "Evidence that ddr1 promotes oligodendrocyte differentiation during development and myelin repair after injury," *International Journal of Molecular Sciences*, vol. 24, no. 12, p. 10318–10318, Jun 2023.

[304] K. Schoch, L. Meng, S. Szelinger, D. R. Bearden, A. Stray-Pedersen, O. L. Busk, N. Stong, E. Liston, R. D. Cohn, F. Scaglia, J. A. Rosenfeld, J. Tarpinian, C. M. Skraban, M. A. Deardorff, J. N. Friedman, Z. C. Akdemir, N. Walley, M. A. Mikati, P. G. Kranz, and J. Jasien, "A recurrent de novo variant in nacc1 causes a syndrome characterized by infantile epilepsy, cataracts, and profound developmental delay," *The American Journal of Human Genetics*, vol. 100, no. 2, p. 343–351, Feb 2017.

[305] M. Noseda, G. McLean, K. Niessen, L. Chang, I. Pollet, R. Montpetit, R. Shahidi, K. Dorovini-Zis, L. Li, B. Beckstead, R. E. Durand, P. A. Hoodless, and A. Karsan, "Notch activation results in phenotypic and functional changes consistent with endothelial-to-mesenchymal transformation," *Circulation Research*, vol. 94, no. 7, p. 910–917, Apr 2004.

[306] B. P. Halliday, J. G. Cleland, J. J. Goldberger, and S. K. Prasad, "Personalizing risk stratification for sudden death in dilated cardiomyopathy," *Circulation*, vol. 136, no. 2, p. 215–231, Jul 2017.

[307] R. F. Fernandez, S. Q. Kim, Y. Zhao, R. Foguth, M. M. Weera, J. L. Counihan, D. K. Nomura, J. A. Chester, J. R. Cannon, and J. M. Ellis, "Acyl-coa synthetase 6 enriches the neuroprotective omega-3 fatty acid dha in the brain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 49, p. 12525–12530, Nov 2018.

[308] H. Zhu and A. J. Bendall, "Dlx3 is expressed in the ventral forebrain of chicken embryos: implications for the evolution of the dlx gene family," *The International Journal of Developmental Biology*, vol. 50, no. 1, p. 71–75, Jan 2006.

[309] J. Hwang, T. Mehrani, S. E. Millar, and M. I. Morasso, "Dlx3 is a crucial regulator of hair follicle differentiation and cycling," *Development*, vol. 135, no. 18, p. 3149–3159, Sep 2008.

[310] F. Hou, H. Jiang, L. Cao, X. Jiao, B. Wang, H. Liu, and B. Cui, "The imbalance expression of dlx3 may perform critical function in the occurrence and progression of preeclampsia," *Computational and Mathematical Methods in Medicine*, vol. 2022, p. 1–10, Jan 2022.

[311] K. Andersson, B. Malmgren, E. Åström, A. Nordgren, F. Taylan, and G. Dahllöf, "Mutations in col1a1/a2 and creb3l1 are associated with oligodontia in osteogenesis imperfecta," *Orphanet Journal of Rare Diseases*, vol. 15, no. 1, Mar 2020.

[312] J. Zhao, M. Yin, H. Deng, F. Q. Jin, S. Xu, Y. Lu, M. A. Mastrangelo, H. Luo, and Z. G. Jin, "Cardiac gab1 deletion leads to dilated cardiomyopathy associated with mitochondrial damage and cardiomyocyte apoptosis," *Cell Death and Differentiation*, vol. 23, no. 4, p. 695–706, Oct 2015.

[313] A. Hawkey-Noble, J. A. Pater, R. Kollipara, M. Fitzgerald, A. S. Maekawa, C. S. Kovacs, T.-L. Young, and C. R. French, "Mutation of foxl1 results in reduced cartilage markers in a zebrafish model of otosclerosis," *Genes*, vol. 13, no. 7, p. 1107, Jun 2022.

[314] C. Nakada, S. Satoh, Y. Tabata, K. Arai, and S. Watanabe, "Transcriptional repressor foxl1 regulates central nervous system development by suppressing shh expression in zebra fish," *Molecular and Cellular Biology*, vol. 26, no. 19, p. 7246–7257, Oct 2006.

[315] A. C. Martyn, K. Tóth, R. Schmalzigaug, N. G. Hedrick, R. M. Rodriguiz, R. Yasuda, W. C. Wetsel, and R. T. Premont, "Git1 regulates synaptic structural plasticity underlying learning," *PLOS ONE*, vol. 13, no. 3, p. e0194350–e0194350, Mar 2018.

[316] K. Xu, A. Sacharidou, S. Fu, D. C. Chong, B. Skaug, Z. J. Chen, G. E. Davis, and O. Cleaver, "Blood vessel tubulogenesis requires rasip1 regulation of gtpase signaling," *Developmental Cell*, vol. 20, no. 4, p. 526–539, Apr 2011.

[317] N. Haq, C. Schmidt-Hieber, F. J. Sialana, L. Ciani, J. P. Heller, M. Stewart, L. Bentley, S. Wells, R. J. Rodenburg, P. M. Nolan, E. Forsythe, M. C. Wu, G. Lubec, P. Salinas, M. Häusser, P. L. Beales, and S. Christou-Savina, "Loss of bardet-biedl syndrome proteins causes synaptic aberrations in principal neurons," *PLoS biology*, vol. 17, no. 9, p. e3000414, Sep 2019.

[318] Y. Zou, D. Zwolanek, Y. Izu, S. Gandhy, G. Schreiber, K. Brockmann, M. Devoto, Z. Tian, Y. Hu, G. Veit, M. K. Meier, J. Stetefeld, D. Hicks, V. Straub, N. C. Voermans, D. E. Birk, E. R. Barton, M. Koch, and C. G. Bönnemann, "Recessive and dominant mutations in col12a1 cause a novel eds/myopathy overlap syndrome in humans and mice," *Human molecular genetics*, vol. 23, no. 9, p. 2339–2352, May 2014.

[319] T. T. Fischer, L. D. Nguyen, and B. E. Ehrlich, "Neuronal calcium sensor 1 (ncs1) dependent modulation of neuronal morphology and development," *The FASEB Journal*, vol. 35, no. 10, Sep 2021.

[320] L. Harder, A.-C. Puller, and M. A. Horstmann, "Znf423: Transcriptional modulation in development and cancer," *Molecular and Cellular Oncology*, vol. 1, no. 3, p. e969655, Sep 2014.

[321] J. Guy, J. Gan, J. Selfridge, S. Cobb, and A. Bird, "Reversal of neurological defects in a mouse model of rett syndrome," *Science*, vol. 315, no. 5815, p. 1143–1147, Feb 2007.

[322] H. T. Bjornsson, J. S. Benjamin, L. Zhang, J. Weissman, E. E. Gerber, Y.-C. Chen, R. G. Vaurio, M. C. Potter, K. D. Hansen, and H. C. Dietz, "Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of kabuki syndrome," *Science translational medicine*, vol. 6, no. 256, p. 256ra135, Oct 2014.

V. Appendix

5.1 Figures

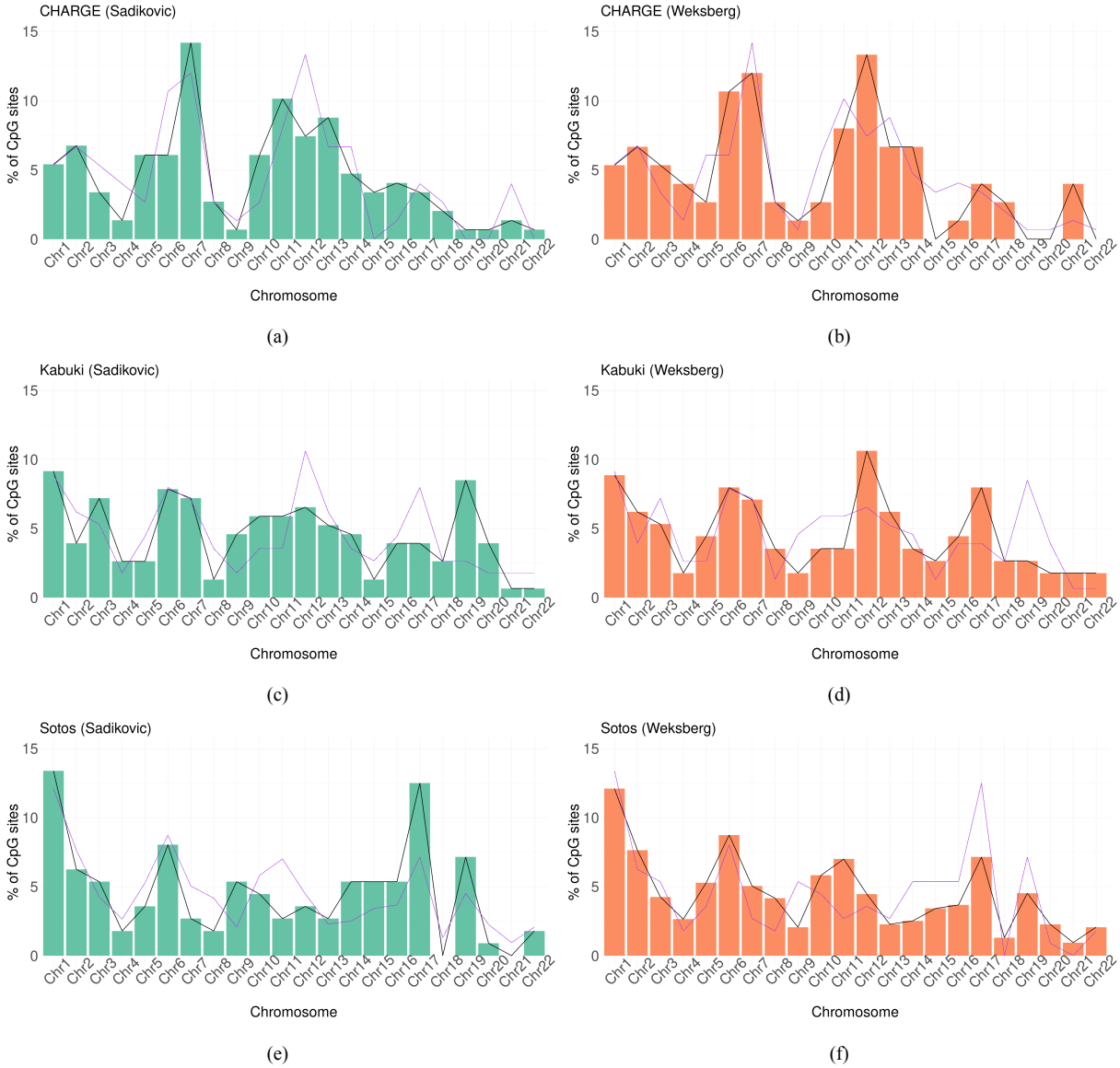


Figure S1. Distribution of differentially methylated selected CpG sites per chromosome in each syndrome. (a) (c) (e) Green bars correspond to episignatures developed by Sadikovic's group. (b) (d) (f) Orange bars correspond to episignatures developed by Weksberg's group. The two plots in the first row correspond to CHARGE syndrome, the second row corresponds to Kabuki syndrome, and the third row corresponds to Sotos syndrome. The black line is a tendency line to facilitate the reading.

Figure S2. Karyoplot with the distribution of differentially methylated CpG sites over chromosomes in **CHARGE** syndrome. Green - Selected DMCs in Sadikovic episignature. Orange - Selected DMCs in Weksberg episignature.

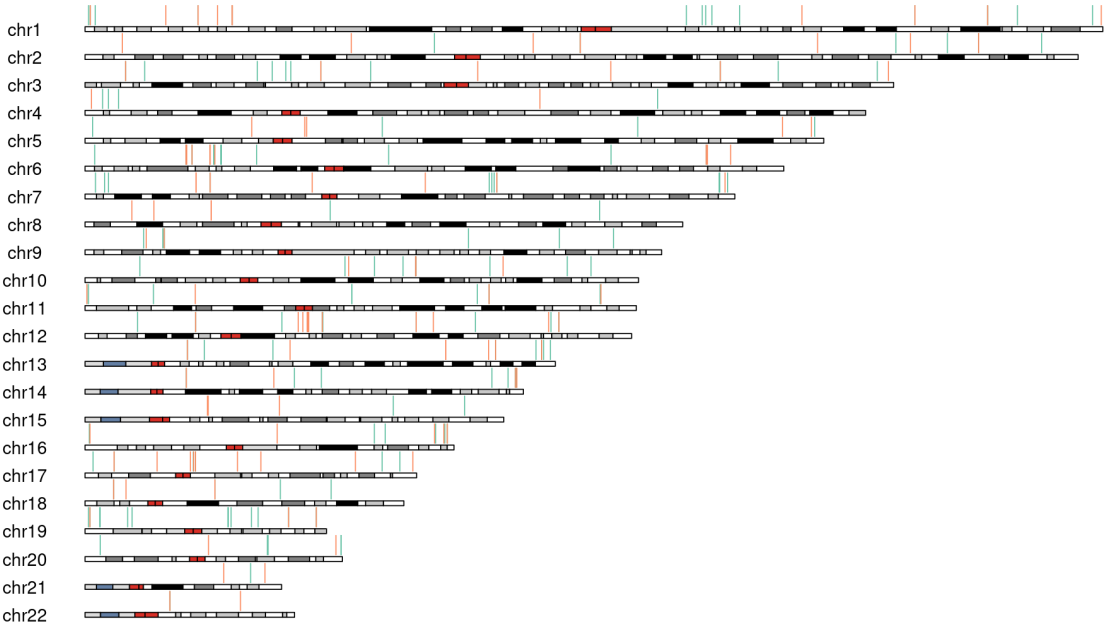


Figure S3. Karyoplot with the distribution of differentially methylated CpG sites over chromosomes in **Kabuki** syndrome. Green - Selected DMCs in Sadikovic episignature. Orange - Selected DMCs in Weksberg episignature.

Figure S4. Karyoplot with the distribution of differentially methylated CpG sites over chromosomes in **Sotos** syndrome for Sadikovic episignature

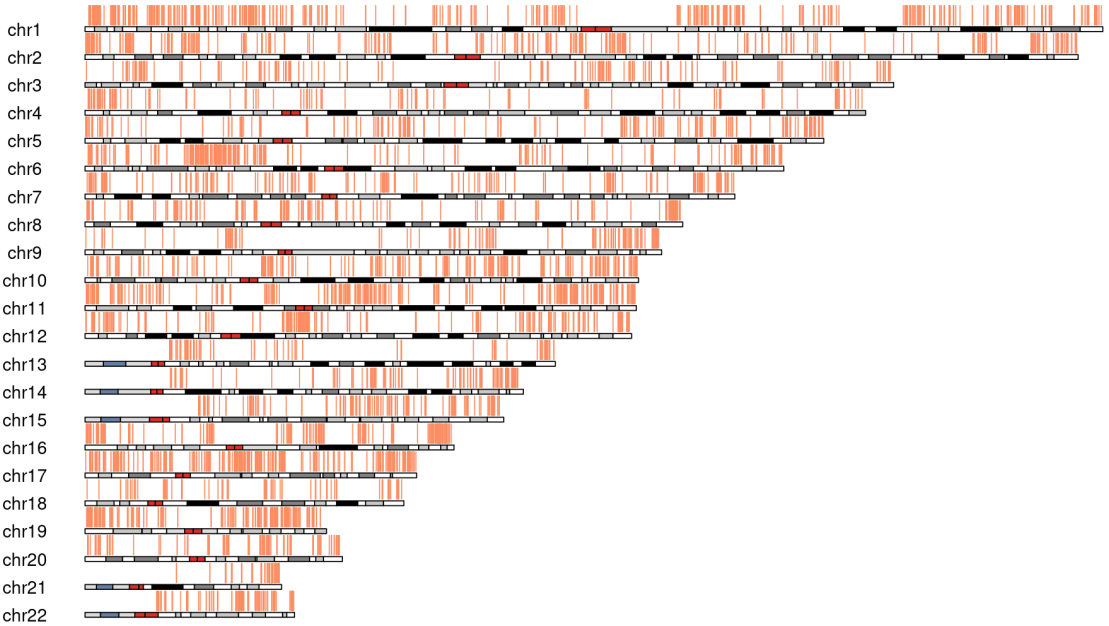


Figure S5. Karyoplot with the distribution of differentially methylated CpG sites over chromosomes in **Sotos** syndrome for Weksberg episignature

5.2 Tables

Table S1. Summary of the number (#) and percentage of CpGs genomic position relative to genes, for Sadikovic CHARGE episignature.

Genic annot.	#	%	DNAm Effect	#	%
1to5kb	23	11.56	Gain	14	7.04
			Loss	9	4.52
Promoters	25	12.56	Gain	18	9.05
			Loss	7	3.52
5UTRs	11	5.53	Gain	10	5.03
			Loss	1	0.50
Gene body	88	44.22	Gain	60	30.15
			Loss	28	14.07
3UTRs	3	1.51	Gain	2	1.01
			Loss	1	0.50
Intergenic	49	24.62	Gain	29	14.57
			Loss	20	10.05

Table S2. Summary of the number (#) and percentage of CpGs genomic position relative to genes, for Weksberg CHARGE episignature.

Genic annot.	#	%	DNAm Effect	#	%
1to5kb	13	11.21	Gain	4	3.45
			Loss	9	7.76
Promoters	13	11.21	Gain	4	3.45
			Loss	9	7.76
5UTRs	5	4.31	Gain	3	2.59
			Loss	2	1.72
Gene body	59	50.86	Gain	20	17.24
			Loss	39	33.62
3UTRs	7	6.03	Gain	1	0.86
			Loss	6	5.17
Intergenic	24	16.38	Gain	4	3.45
			Loss	15	12.93

Table S3. Summary of the number (#) and percentage of CpGs genomic position relative to CpG islands (CGIs), for Sadikovics CHARGE episignature.

CGI annot.	#	%	DNAm Effect	#	%
shores	31	20.95	Gain	12	8.11
			Loss	19	12.84
shelves	7	4.73	Gain	2	1.35
			Loss	5	3.38
CGIs	24	16.22	Gain	19	12.84
			Loss	5	3.38
Inter	86	58.11	Gain	62	41.89
			Loss	24	16.22

Table S4. Summary of the number (#) and percentage of CpGs genomic position relative to CpG islands (CGIs), for Weksberg CHARGE episignature.

CGI annot.	#	%	DNAm Effect	#	%
shores	21	28	Gain	3	4
			Loss	18	24
shelves	9	12.00	Gain	1	1.33
			Loss	8	10.67
CGIs	8	10.67	Gain	4	5.33
			Loss	4	5.33
Inter	37	49.33	Gain	13	17.33
			Loss	24	32

Table S5. Summary of the number (#) and percentage of CpGs genomic position relative to genes, for Sadikovic Kabuki episignature.

Genic annot.	#	%	DNAm Effect	#	%
1to5kb	39	16.96	Gain	10	4.35
			Loss	29	12.61
Promoters	23	10.00	Gain	9	3.91
			Loss	14	6.09
5UTRs	11	4.78	Gain	4	1.74
			Loss	7	3.04
Gene body	128	55.65	Gain	44	19.13
			Loss	84	36.52
3UTRs	8	3.48	Gain	2	0.87
			Loss	6	2.61
Intergenic	39	9.13	Gain	19	8.26
			Loss	2	0.87

Table S6. Summary of the number (#) and percentage of CpGs genomic position relative to genes, for Weksberg Kabuki episignature.

Genic annot.	#	%	DNAm Effect	#	%
1to5kb	19	12.58	Gain	3	1.99
			Loss	16	10.60
Promoters	18	11.92	Gain	9	5.96
			Loss	9	5.96
5UTRs	5	3.31	Gain	3	1.99
			Loss	2	1.32
Gene body	79	52.32	Gain	36	23.84
			Loss	43	28.48
3UTRs	6	3.97	Gain	1	0.66
			Loss	5	3.31
Intergenic	37	15.89	Gain	17	11.26
			Loss	7	4.64

Table S7. Summary of the number (#) and percentage of CpGs genomic position relative to CpG islands (CGIs), for Sadikovics Kabuki episignature.

CGI annot.	#	%	DNAm Effect	#	%
shores	71	46.41	Gain	16	10.46
			Loss	55	35.95
shelves	12	7.84	Gain	8	5.23
			Loss	4	2.61
CGIs	33	21.57	Gain	15	9.80
			Loss	18	11.76
Inter	37	24.18	Gain	29	18.95
			Loss	8	5.23

Table S8. Summary of the number (#) and percentage of CpGs genomic position relative to CpG islands (CGIs), for Weksberg Kabuki episignature.

CGI annot.	#	%	DNAm Effect	#	%
shores	46	40.71	Gain	19	16.81
			Loss	27	23.89
shelves	10	8.85	Gain	7	6.19
			Loss	3	2.65
CGIs	25	22.12	Gain	11	9.73
			Loss	14	12.39
Inter	32	28.32	Gain	18	15.93
			Loss	14	12.39

Table S9. Summary of the number (#) and percentage of CpGs genomic position relative to genes, for Sadikovic Sotos episignature.

Genic annot.	DNAm Effect	#	%
1to5kb	Loss	24	16.55
Promoters	Loss	32	22.07
5UTRs	Loss	4	2.76
Gene body	Loss	43	29.65
3UTRs	Loss	4	2.76
Intergenic	Loss	38	26.21

Table S10. Summary of the number (#) and percentage of CpGs genomic position relative to genes, for Weksberg Sotos episignature.

Genic annot.	#	%	DNAm Effect	#	%
1to5kb	1308	14.02	Gain	5	0.05
			Loss	1303	13.97
Promoters	1925	20.64	Gain	11	0.12
			Loss	1914	20.52
5UTRs	350	3.75	Gain	1	0.01
			Loss	349	3.74
Gene body	3347	35.89	Gain	37	0.40
			Loss	3310	35.49
3UTRs	173	1.85	Gain	3	0.03
			Loss	170	1.82
Intergenic	31	23.84	Gain	11	0.12
			Loss	2213	23.73

Table S11. Summary of the number (#) and percentage of CpGs genomic position relative to CpG islands (CGIs), for Sadikovics Sotos episignature.

CGI annot.	DNAm Effect	#	%
shores	Loss	51	45.54
shelves	Loss	11	9.82
CGIs	Loss	34	30.36
Inter	Loss	16	14.29

Table S12. Summary of the number (#) and percentage of CpGs genomic position relative to CpG islands (CGIs), for Weksberg Sotos episignature.

CGI annotation	#	%	DNAm Effect	#	%
shores	2553	36.03	Gain	17	0.24
			Loss	2536	35.79
shelves	431	6.08	Gain	1	0.01
			Loss	430	6.07
CGIs	2030	28.65	Gain	11	0.16
			Loss	2019	28.50
Inter	2071	29.23	Gain	18	0.25
			Loss	2053	28.98

Table S13. Number of common genes covered by the selected differentially methylated CpGs between episignatures for CHARGE, Kabuki and Sotos syndromes built by Sadikovic's and Weksberg's research groups. SC = Sadikovic CHARGE, SK = Sadikovic Kabuki, SS = Sadikovic Sotos, WC = Weksberg CHARGE, WK = Weksberg Kabuki, WS = Weksberg Sotos.

	SC	SK	SS	WC	WK	WS
SC	96	1	0	21	6	30
SK	-	168	2	1	44	34
SS	-	-	91	0	2	91
WC	-	-	-	58	4	14
WK	-	-	-	-	106	30
WS	-	-	-	-	-	2456

Table S14. Enriched non-redundant biological processes (BPs) retrieved by REVIGO for CHARGE episignature in Sadikovic group. Semantic similarity cutoff = 0.7. Higher frequency (Freq.) values correspond to more general terms. Uniqueness (Uniq.) semantically compares the GO terms to the whole list, and uniqueness closer to 1 means more unique. Less dispersibility (Disp.) means the term is more unique and thus less dispensable (non-redundant). # is the number of genes involved in the enriched BP.

GO BP	#	Freq.	Disp.	PValue
anterior/posterior pattern specification	8	1.149	0.000	0.00000
embryonic skeletal system morphogenesis	5	0.535	0.244	0.00004
negative regulation of canonical Wnt signaling pathway	6	0.754	0.000	0.00067
axonogenesis	5	1.942	0.399	0.00078
positive regulation of transcription by RNA pol. II	14	6.959	0.574	0.00139
regulation of transcription by RNA pol. II	17	13.631	0.053	0.00166
regulation of NMDA receptor activity	3	0.124	0.258	0.00194
cell adhesion	9	5.360	0.006	0.00221
uterus development	3	0.129	0.256	0.00246
positive regulation of MAP kinase activity	4	0.664	0.301	0.00396
homophilic cell adhesion via plasma membrane adhesion molecules	5	0.940	0.000	0.00581
positive regulation of protein tyrosine kinase activity	3	0.253	0.692	0.00677
negative regulation of transcription by RNA pol. II	11	5.445	0.617	0.00724
response to testosterone	3	0.231	0.000	0.00768
embryonic forelimb morphogenesis	3	0.180	0.573	0.00768
embryonic organ development	3	2.551	0.603	0.00815
negative regulation of gene expression	6	5.276	0.381	0.00900
negative regulation of cell population proliferation	7	3.913	0.244	0.01143
lateral motor column neuron migration	2	0.017	0.411	0.01231
stem cell differentiation	3	1.025	0.173	0.01418
skeletal system development	4	2.753	0.295	0.01796
outflow tract morphogenesis	3	0.439	0.515	0.01877
adult locomotory behavior	3	0.456	0.125	0.01946
sinoatrial node cell development	2	0.028	0.203	0.02044
response to xenobiotic stimulus	5	2.399	0.244	0.02059
central nervous system development	4	5.917	0.541	0.02328
trigeminal nerve development	2	0.062	0.200	0.02850
female genitalia development	2	0.107	0.597	0.02850
hippocampus development	3	0.512	0.290	0.03293
regulation of toll-like receptor signaling pathway	2	0.175	0.265	0.03649
cellular response to manganese ion	2	0.068	0.179	0.03649
neurotransmitter uptake	2	0.169	0.021	0.04046
ventral spinal cord development	2	0.259	0.474	0.04046
secondary heart field specification	2	0.062	0.654	0.04442
regulation of behavior	2	0.428	0.027	0.04442
cell-cell adhesion	4	3.023	0.678	0.04670
modulation of chemical synaptic transmission	3	2.663	0.317	0.04704
prostate gland development	2	0.253	0.638	0.04836

Table S14 continued from previous page

atrial septum morphogenesis	2	0.084	0.670	0.04836
circulatory system development	2	5.101	0.397	0.04836
receptor localization to synapse	2	0.236	0.165	0.04836

Table S15. Enriched non-redundant biological processes (BPs) retrieved by REVIGO for CHARGE episignature in Weksberg group. Semantic similarity cutoff = 0.7. Higher frequency (Freq.) values correspond to more general terms. Uniqueness (Uniq.) semantically compares the GO terms to the whole list, and uniqueness closer to 1 means more unique. Less dispersibility (Disp.) means the term is more unique and thus less dispensable (non-redundant). # is the number of genes involved in the enriched BP.

Name	#	Freq.	Disp.	PValue
cell adhesion	7	0.706	0.000	0.00281
positive regulation of transcription by RNA pol. II	10	0.369	0.434	0.00315
regulation of transcription by RNA pol. II	12	1.918	0.000	0.00335
negative regulation of cell population proliferation	6	0.093	0.231	0.00576
embryonic skeletal system morphogenesis	3	0.020	0.000	0.00577
cognition	3	0.035	0.443	0.00735
axon midline choice point recognition	2	0.002	0.499	0.01256
cellular response to glucose stimulus	3	0.008	0.007	0.01540
sinoatrial node development	2	0.001	0.512	0.02002
regulation of toll-like receptor signaling pathway	2	0.009	0.191	0.02249
anterior/posterior pattern specification	3	0.039	0.531	0.02495
transmembrane transport	4	13.535	0.014	0.02648
cardiac septum morphogenesis	2	3.402	0.537	0.03233
positive regulation of protein kinase B signaling	3	0.012	0.555	0.04206
regulation of transcription involved in G1/S transition of mitotic cell cycle	2	0.011	0.317	0.04207
epithelial tube branching involved in lung morphogenesis	2	0.004	0.676	0.04207
embryonic brain development	2	2.818	0.590	0.04449

Table S16. Enriched non-redundant biological processes (BPs) retrieved by REVIGO for Kabuki episignature in Sadikovic group. Semantic similarity cutoff = 0.7. Higher frequency (Freq.) values correspond to more general terms. Uniqueness (Uniq.) semantically compares the GO terms to the whole list, and uniqueness closer to 1 means more unique. Less dispersibility (Disp.) means the term is more unique and thus less dispensable (non-redundant). # is the number of genes involved in the enriched BP.

GO BP	#	Freq.	Disp.	PValue
localization	4	25.297	0.000	0.00258
protein targeting	4	1.436	0.000	0.00452
fibroblast migration	3	0.107	0.004	0.00860
regulation of cell population proliferation	6	9.329	0.044	0.00899
negative regulation of DNA-templated transcription	11	7.303	0.229	0.01238
phagocytosis	4	0.890	0.272	0.01523
principal sensory nucleus of trigeminal nerve development	2	0.017	0.000	0.02209
negative regulation of long-chain fatty acid import across plasma membrane	2	0.023	0.000	0.02934

Table S16 continued from previous page

cell death	3	6.199	0.006	0.03227
regulation of lipid metabolic process	3	1.914	0.022	0.03710

Table S17. Enriched non-redundant biological processes (BPs) retrieved by REVIGO for Kabuki episignature in Weksberg group. Semantic similarity cutoff = 0.7. Higher frequency (Freq.) values correspond to more general terms. Uniqueness (Uniq.) semantically compares the GO terms to the whole list, and uniqueness closer to 1 means more unique. Less dispersibility (Disp.) means the term is more unique and thus less dispensable (non-redundant). # is the number of genes involved in the enriched BP.

GO BP	#	Freq.	Disp.	PValue
embryonic skeletal system morphogenesis	5	0.535	0.000	0.00005
anterior/posterior pattern specification	5	1.149	0.244	0.00095
proline transport	3	0.062	0.000	0.00106
mammary gland alveolus development	3	0.113	0.253	0.00290
execution phase of apoptosis	3	0.287	0.000	0.00433
regulation of apoptotic process	6	8.158	0.042	0.00474
cell-cell adhesion	5	3.023	0.005	0.01150
extrinsic apoptotic signaling pathway	3	0.625	0.489	0.02277
cellular senescence	3	0.343	0.005	0.02528
regulation of cilium movement	2	0.225	0.000	0.02659
angiogenesis	5	1.881	0.397	0.02772
leucine transport	2	1.041	0.874	0.04393
tumor necrosis factor-mediated signaling pathway	3	0.321	0.115	0.04787

Table S18. Enriched non-redundant biological processes (BPs) retrieved by REVIGO for Sotos episignature in Sadikovic group. Semantic similarity cutoff = 0.7. Higher frequency (Freq.) values correspond to more general terms. Uniqueness (Uniq.) semantically compares the GO terms to the whole list, and uniqueness closer to 1 means more unique. Less dispersibility (Disp.) means the term is more unique and thus less dispensable (non-redundant). # is the number of genes involved in the enriched BP.

GO BP	#	Freq.	Disp.	PValue
cellular response to fibroblast growth factor stimulus	3	0.019	0.000	0.00633
positive regulation of cell migration by vascular endothelial growth factor signalling pathway	2	0.001	0.696	0.01638
nucleosome assembly	4	0.091	0.000	0.02622
positive regulation of MAPK cascade	4	0.077	0.000	0.02936
morphogenesis of a branching structure	2	0.030	0.000	0.03250
signal transduction	11	8.136	0.295	0.03755
negative regulation of fibroblast migration	2	0.001	0.135	0.04046
cell population proliferation	4	0.111	0.007	0.04381
branching involved in mammary gland duct morphogenesis	2	0.002	0.514	0.04836

Table S19. Enriched non-redundant biological processes (BPs) retrieved by REVIGO for Sotos episignature in Weksberg group. Semantic similarity cutoff = 0.5. Higher frequency (Freq.) values correspond to more general terms. Uniqueness (Uniq.) semantically compares the GO terms to the whole list, and uniqueness closer to 1 means more unique. Less dispersibility (Disp.) means the term is more unique and thus less dispensable (non-redundant). # is the number of genes involved in the enriched BP.

GO BP	#	Freq.	Disp.	PValue
cell adhesion	107	5.360	0.007	0.00000
homophilic cell adhesion via plasma membrane adhesion molecules	46	0.940	0.000	0.00000
sensory perception of sound	43	0.867	0.000	0.00000
negative regulation of fatty acid metabolic process	10	0.197	0.000	0.00000
inner ear morphogenesis	22	0.574	0.124	0.00000
flavonoid glucuronidation	9	0.028	0.003	0.00000
nervous system development	79	12.359	0.468	0.00000
negative regulation of glucuronosyltransferase activity	8	0.063	0.017	0.00000
axon guidance	43	1.177	0.381	0.00000
cell fate commitment	20	1.447	0.166	0.00000
cell differentiation	110	19.903	0.440	0.00001
regulation of monoatomic ion transmembrane transport	32	2.618	0.037	0.00002
extracellular matrix organization	40	1.543	0.005	0.00002
roof of mouth development	20	0.512	0.168	0.00003
positive regulation of transcription by RNA polymerase II	173	6.959	0.351	0.00003
chondrocyte differentiation	18	0.467	0.289	0.00004
locomotory behavior	23	1.098	0.133	0.00006
regulation of heart rate by cardiac conduction	15	0.236	0.023	0.00006
anterior/posterior pattern specification	24	1.149	0.246	0.00014
excretion	12	0.107	0.327	0.00019
negative regulation of Wnt signaling pathway	16	0.946	0.182	0.00026
cell-cell signaling	42	6.092	0.226	0.00051
spermatid development	23	1.115	0.416	0.00051
calcium ion transmembrane transport	25	1.222	0.005	0.00052
skeletal system development	28	2.753	0.294	0.00067
cardiac muscle cell action potential involved in contraction	8	0.214	0.348	0.00068
ossification	20	1.571	0.142	0.00070
cellular response to retinoic acid	18	0.366	0.004	0.00075
canonical Wnt signaling pathway	23	0.586	0.122	0.00079
BMP signaling pathway	22	0.512	0.363	0.00085
adrenal gland development	10	0.158	0.262	0.00091
cilium movement	14	0.952	0.005	0.00104
heart development	38	3.102	0.369	0.00138
anatomical structure morphogenesis	28	12.409	0.309	0.00149
photoreceptor cell maintenance	13	0.242	0.114	0.00150
negative regulation of chondrocyte differentiation	9	0.135	0.237	0.00169
multicellular organism development	28	21.924	0.499	0.00204
regulation of neurotransmitter secretion	9	0.462	0.464	0.00230

Table S19 continued from previous page

cell population proliferation	34	4.031	0.006	0.00263
Rho protein signal transduction	15	0.343	0.178	0.00268
multicellular organism growth	20	0.479	0.149	0.00295
signal transduction	166	26.812	0.436	0.00316
norepinephrine metabolic process	5	0.073	0.073	0.00328
positive regulation of pathway-restricted SMAD protein phosphorylation	13	0.282	0.283	0.00477
cell maturation	11	0.946	0.408	0.00481
lung development	18	1.019	0.378	0.00520
cellular response to calcium ion	19	0.490	0.472	0.00534
Notch signaling pathway	23	0.647	0.372	0.00541
monoatomic ion transport	26	5.642	0.335	0.00566
stem cell proliferation	12	0.394	0.004	0.00695
negative regulation of cell population proliferation	66	3.913	0.251	0.00720
positive regulation of integrin-mediated signaling pathway	6	0.073	0.342	0.00796
wound healing	19	1.892	0.088	0.00848
tissue development	10	9.752	0.398	0.00862
bone mineralization	13	0.321	0.378	0.00910
response to hypoxia	28	1.509	0.330	0.00971
metanephric epithelium development	4	0.141	0.352	0.00984
proteoglycan metabolic process	5	0.507	0.152	0.00992
transmembrane transport	40	7.370	0.444	0.01069
cell fate determination	7	0.242	0.252	0.01085
positive regulation of macrophage derived foam cell differentiation	7	0.101	0.331	0.01085
response to stimulus	17	45.853	0.000	0.01185
regulation of small GTPase mediated signal transduction	22	1.695	0.341	0.01207
positive regulation of fat cell differentiation	13	0.377	0.465	0.01218
protein processing	17	1.098	0.249	0.01477
regulation of cell migration	21	5.169	0.055	0.01495
positive regulation of cytokine-mediated signaling pathway	5	0.327	0.387	0.01517
regulation of microtubule polymerization or depolymerization	6	0.495	0.041	0.01554
positive regulation of interleukin-13 production	6	0.079	0.227	0.01554
neurotransmitter transport	11	0.788	0.263	0.01653
regulation of Wnt signaling pathway	7	1.847	0.370	0.01809
negative regulation of stem cell proliferation	7	0.124	0.433	0.01809
dopamine uptake	4	0.062	0.310	0.01815
sphingolipid metabolic process	8	0.890	0.190	0.01881
cell-cell junction assembly	11	0.732	0.178	0.01919
keratan sulfate biosynthetic process	6	0.090	0.446	0.02069
establishment or maintenance of transmembrane electrochemical gradient	5	0.062	0.491	0.02189
organic cation transport	5	0.259	0.353	0.02189
lipoxygenase pathway	5	0.062	0.466	0.02189

Table S19 continued from previous page

neuromuscular process controlling balance	11	0.287	0.432	0.02542
protein catabolic process	15	4.324	0.318	0.02576
gene expression	22	9.633	0.078	0.02591
positive regulation of blood coagulation	6	0.158	0.320	0.02683
white fat cell differentiation	6	0.096	0.231	0.02683
genomic imprinting	7	0.129	0.150	0.02807
regulation of appetite	4	0.113	0.237	0.02930
regulation of peptidyl-serine phosphorylation	4	0.766	0.493	0.02930
positive regulation of interleukin-5 production	5	0.062	0.483	0.03016
skeletal muscle satellite cell activation	3	0.028	0.095	0.03079
positive regulation of ovulation	3	0.017	0.274	0.03079
embryo implantation	10	0.225	0.438	0.03218
positive regulation of protein import into nucleus	10	0.225	0.430	0.03218
intracellular zinc ion homeostasis	9	0.135	0.371	0.03542
positive regulation of bone mineralization	10	0.248	0.376	0.03681
positive regulation of osteoblast proliferation	5	0.073	0.411	0.04002
negative regulation of hippo signaling	5	0.073	0.416	0.04002
skin morphogenesis	5	0.073	0.444	0.04002
negative regulation of substrate adhesion-dependent cell spreading	5	0.073	0.148	0.04002
piRNA processing	7	0.118	0.376	0.04113
embryo development	6	5.799	0.480	0.04229
negative regulation of growth	6	1.379	0.192	0.04229
retinoic acid receptor signaling pathway	6	0.107	0.161	0.04229
pharyngeal system development	6	0.163	0.463	0.04229
protein homooligomerization	21	1.064	0.364	0.04266
negative regulation of acute inflammatory response	4	0.068	0.387	0.04328
positive regulation of plasminogen activation	4	0.045	0.345	0.04328

Table S20. Number of differentially methylated cytosines (#DMCs) present in hypermethylated genes of Sadikovic CHARGE episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
HOXA-AS3	8	0.119	VSIG2	1	0.078
HOXA5	8	0.119	SLC1A3	1	0.076
FMN2	2	0.093	LRRC4C	1	0.076
TEAD1	2	0.068	CDH11	1	0.075
VWF	1	0.181	KLHL29	1	0.074
TTC24	1	0.147	CCDC175	1	0.071
KIAA1217	1	0.136	OCA2	1	0.069
CLDN9	1	0.128	PTPRR	1	0.069
DNAL4	1	0.123	PCCA	1	0.069

Table S20 continued from previous page

SUN2	1	0.123	CTAGE1	1	0.069
APP	1	0.117	RFTN2	1	0.068
GFI1	1	0.116	SLC1A7	1	0.068
DLG2	1	0.107	CYP19A1	1	0.067
SLCO1A2	1	0.106	ACACB	1	0.066
MYO1F	1	0.105	TBX3	1	0.066
C10orf90	1	0.104	KCNMB2	1	0.066
SLC9C1	1	0.102	PRKN	1	0.065
RAB3C	1	0.101	C5orf34	1	0.065
LOC283194	1	0.100	HECW1	1	0.063
N4BP2L2	1	0.100	RIN2	1	0.063
RCCD1	1	0.099	HTR2A	1	0.063
TRUB1	1	0.091	OPRM1	1	0.063
TMEM132C	1	0.090	MIR30A	1	0.062
RELN	1	0.088	RBM28	1	0.062
TMEM131L	1	0.084	FBXO41	1	0.060
ARHGEF15	1	0.081	DLC1	1	0.058
TAFA1	1	0.080	KIF25	1	0.055
HOXA7	1	0.079	HEY2	1	0.054
THBS1	1	0.079	CHST13	1	0.053
GPM6A	1	0.078	PCDH17	1	0.048
ESAM	1	0.078			

Table S21. Number of differentially methylated cytosines (#DMCs) present in hypomethylated genes of Sadikovic CHARGE episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
FMN2	2	-0.093	PARVA	1	-0.074
COL4A2	2	-0.081	HOXA11-AS	1	-0.073
TEAD1	2	-0.068	PCDH20	1	-0.071
HOXA11	2	-0.068	GPR151	1	-0.070
HOTAIRM1	1	-0.163	DIP2C	1	-0.069
HOXA1	1	-0.163	KCNJ6	1	-0.067
LINC02914	1	-0.130	HOXD4	1	-0.062
HOXB7	1	-0.120	MIR10B	1	-0.062
HOXB8	1	-0.120	KLHL14	1	-0.057
SLITRK5	1	-0.117	AXIN2	1	-0.057
LINC02381	1	-0.111	SGPP2	1	-0.056
NOX4	1	-0.099	HOXA10	1	-0.055
DAB1	1	-0.096	HOXA10-HOXA9	1	-0.055
TFAP2A	1	-0.088	FLJ12825	1	-0.054
COLEC12	1	-0.084	NALF1	1	-0.051

Table S21 continued from previous page

NKX2-8	1	-0.082	MIR708	1	-0.048
KIRREL3	1	-0.081	TENM4	1	-0.048
FOXP2	1	-0.078	SGK1	1	-0.045
ISL1	1	-0.076			

Table S22. Number of differentially methylated cytosines (#DMCs) present in hypermethylated genes of Weksberg CHARGE episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
HOXA-AS3	3	0.111	PSMB9	1	0.106
HOXA5	2	0.115	TAP1	1	0.106
HOXA6	2	0.104	ARHGEF15	1	0.105
VWF	1	0.203	PPP2R5C	1	0.104
APP	1	0.150	LINC01550	1	0.103
SORBS2	1	0.136	LMO3	1	0.102
SLCO1A2	1	0.111	FAM210A	1	0.102
HLA-E	1	0.111	RNMT	1	0.102
LMO7	1	0.109	GFI1	1	0.101
ZDHHC22	1	0.108			

Table S23. Number of differentially methylated cytosines (#DMCs) present in hypomethylated genes of Weksberg CHARGE episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
HOTAIRM1	3	-0.134	NOX4	1	-0.118
HOXA1	3	-0.134	MYORG	1	-0.115
NR4A2	3	-0.109	SARM1	1	-0.113
HHIP	1	-0.187	CHD7	1	-0.112
HHIP-AS1	1	-0.187	PARVA	1	-0.110
ROBO2	1	-0.157	MKS1	1	-0.109
LINC02914	1	-0.150	C6orf89	1	-0.108
SLITRK5	1	-0.150	BACH1	1	-0.108
LINC02381	1	-0.150	FOXP2	1	-0.108
COL4A2	1	-0.137	HOXC9	1	-0.107
TBX5	1	-0.132	GJB6	1	-0.106
NAA25	1	-0.129	PCDH20	1	-0.105
CDC42BPB	1	-0.127	EIF4E3	1	-0.104
KIRREL3	1	-0.125	FMN2	1	-0.104
BACH2	1	-0.125	CD9	1	-0.103
TRA2B	1	-0.124	ANAPC13	1	-0.103
WIPI2	1	-0.122	CEP63	1	-0.103

Table S23 continued from previous page

TEAD1	1	-0.121	DAB1	1	-0.100
AFDN	1	-0.118	NCALD	1	-0.100
AFDN-DT	1	-0.118			

Table S24. Number of differentially methylated cytosines (#DMCs) present in hypermethylated genes of Sadikovic Kabuki episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
ZMIZ1	2	0.166	WTIP	1	0.090
TCEA2	2	0.109	TRAK1	1	0.086
TACC2	1	0.150	S100A11	1	0.085
SOX18	1	0.149	CCDC172	1	0.083
RNF216	1	0.145	S100A4	1	0.083
ADAMTS2	1	0.139	S100A5	1	0.083
LAMB2	1	0.139	LMX1B	1	0.082
CD79A	1	0.134	PIP5K1C	1	0.081
CD37	1	0.124	KIRREL3	1	0.080
MAP3K6	1	0.123	ARID5B	1	0.078
CLYBL	1	0.123	NFIB	1	0.076
CDH22	1	0.123	NHERF1	1	0.073
RBFOX3	1	0.121	CACTIN-AS1	1	0.071
IL17C	1	0.115	TBXA2R	1	0.071
LFNG	1	0.112	ESR1	1	0.071
CDKN1B	1	0.107	CLDN15	1	0.067
PTDSS2	1	0.105	EVC	1	0.065
FOXK1	1	0.103	EVC2	1	0.065
ST8SIA2	1	0.098	LRMDA	1	0.061
BCL11B	1	0.095	CTDSP2	1	0.059
FBXO24	1	0.095	NIN	1	0.059
LRCH4	1	0.095	MITF	1	0.059
PCOLCE-AS1	1	0.095	ZNF787	1	0.057
CIDEC	1	0.094	EHBP1L1	1	0.049
IL17RE	1	0.094	NLK	1	0.046
CHIT1	1	0.093	HDAC7	1	0.045
SH3TC1	1	0.091	FGD6	1	0.044

Table S25. Number of differentially methylated cytosines (#DMCs) present in hypomethylated genes of Sadikovic Kabuki episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
RPS6	2	-0.125	SRGN	1	-0.093

Table S25 continued from previous page

TCEA2	2	-0.109	PSMG1	1	-0.091
KCNK7	1	-0.236	GPR160	1	-0.090
TNS1	1	-0.213	SLC17A5	1	-0.087
AGAP2	1	-0.212	PTPRK	1	-0.087
AGAP2-AS1	1	-0.212	ZKSCAN1	1	-0.086
TSPAN4	1	-0.155	GMDS-DT	1	-0.085
TNFAIP2	1	-0.151	MAML2	1	-0.084
SAMD11	1	-0.151	MXRA8	1	-0.084
RAB34	1	-0.151	FGFR1OP2	1	-0.083
RPL23A	1	-0.151	INTS13	1	-0.083
SNORD42A	1	-0.151	VAC14	1	-0.083
SNORD4A	1	-0.151	SLC33A1	1	-0.083
SNORD4B	1	-0.151	GABARAP	1	-0.083
SLC6A20	1	-0.151	PHF23	1	-0.083
LAMA1	1	-0.142	LAMP1	1	-0.082
SLMAP	1	-0.138	IRS2	1	-0.081
CIDEB	1	-0.137	PSMB2	1	-0.080
LTB4R	1	-0.137	TFAP2E	1	-0.080
GARS1	1	-0.137	C9orf43	1	-0.080
GARS1-DT	1	-0.137	POLE3	1	-0.080
RGL3	1	-0.136	MYL1	1	-0.080
SQLE	1	-0.135	CNTN5	1	-0.077
CFD	1	-0.131	TGFBI	1	-0.077
ELANE	1	-0.131	PTPRN2	1	-0.076
TOX	1	-0.125	PARP4	1	-0.074
ARHGEF7	1	-0.123	CSNK2B	1	-0.072
RRP12	1	-0.120	GPANK1	1	-0.072
C6orf62	1	-0.120	LY6G5B	1	-0.072
LYSMD1	1	-0.117	POMP	1	-0.072
SCNM1	1	-0.117	COPA	1	-0.072
TNFAIP8L2-SCNM1	1	-0.117	NCSTN	1	-0.072
PLD4	1	-0.113	SCARF2	1	-0.071
H2BC7	1	-0.113	NAA30	1	-0.070
CBX5	1	-0.109	GUK1	1	-0.070
HNRNPA1	1	-0.109	TMEM115	1	-0.069
CNST	1	-0.107	SLC12A5	1	-0.067
TFB2M	1	-0.107	ATP6V1G2-DDX39B	1	-0.067
CCND3	1	-0.107	DDX39B	1	-0.067
TAF8	1	-0.107	SNORD117	1	-0.067
B3GALT4	1	-0.106	NDUFC1	1	-0.067
RPS18	1	-0.106	CUTA	1	-0.066
VPS52	1	-0.106	SYNGAP1	1	-0.066
MARS2	1	-0.104	ICAM3	1	-0.065
LYAR	1	-0.104	RAVER1	1	-0.065

Table S25 continued from previous page

ZBTB49	1	-0.104	RBM33	1	-0.063
LMF1	1	-0.103	SCGB1B2P	1	-0.061
SOX8	1	-0.103	DGKD	1	-0.060
CXXC1	1	-0.101	TGOLN2	1	-0.059
NOL4	1	-0.101	C11orf58	1	-0.058
HSPA12B	1	-0.100	SOX6	1	-0.058
ARPC1B	1	-0.097	RRAGA	1	-0.055
HIC1	1	-0.096	BTF3	1	-0.055
HLX	1	-0.096	MRPS18B	1	-0.054
AKT2	1	-0.096	PPP1R10	1	-0.054
CCNP	1	-0.096	FOXRED1	1	-0.050
LSR	1	-0.095	SRPRA	1	-0.050
USF2	1	-0.095			

Table S26. Number of differentially methylated cytosines (#DMCs) present in hypermethylated genes of Weksberg Kabuki episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
ZMIZ1	2	0.201	PTDSS2	1	0.130
SLITRK5	2	0.109	CD37	1	0.124
RAI1	1	0.298	KSR1	1	0.120
RIPOR2	1	0.262	ENPEP	1	0.120
IL17C	1	0.232	CHIT1	1	0.114
CASP8	1	0.208	CAPZB	1	0.112
HOXA-AS3	1	0.179	KIRREL3	1	0.112
HOXA5	1	0.179	ZDHHC14	1	0.111
MELTF	1	0.179	BRF1	1	0.111
MIR548N	1	0.179	CIDEC	1	0.110
TTN-AS1	1	0.179	IL17RE	1	0.110
CLYBL	1	0.175	MAP3K6	1	0.106
ARL5C	1	0.164	ZNF787	1	0.105
KRT18	1	0.156	MOGAT3	1	0.104
KRT8	1	0.156	HOXA7	1	0.104
VOPP1	1	0.149	SEMA6D	1	0.103
LGALS1	1	0.145	PPP1R18	1	0.103
CACNA1H	1	0.144	COX4I2	1	0.103
ESR1	1	0.138	SLC6A15	1	0.101
RASAL1	1	0.135			

Table S27. Number of differentially methylated cytosines (#DMCs) present in hypomethylated genes of Weksberg Kabuki episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
HOXC4	2	-0.158	SCARF2	1	-0.143
PSMB2	2	-0.110	FIGNL2	1	-0.141
TFAP2E	2	-0.110	SH3RF3	1	-0.137
PURG	2	-0.110	SH3RF3-AS1	1	-0.137
WRN	2	-0.110	C6orf62	1	-0.130
FLJ12825	1	-0.214	PLD4	1	-0.130
SNHG32	1	-0.207	MIR3912	1	-0.129
SNORD52	1	-0.207	NPM1	1	-0.129
TNS1	1	-0.202	COL9A3	1	-0.126
AGAP2	1	-0.200	FGFR1OP2	1	-0.126
AGAP2-AS1	1	-0.200	INTS13	1	-0.126
ADO	1	-0.195	CNTN5	1	-0.122
HOXA3	1	-0.180	RPL37	1	-0.117
HOXA4	1	-0.180	SNORD72	1	-0.117
LAMA1	1	-0.179	CFAP298	1	-0.117
SLC1A4	1	-0.176	TCP10L	1	-0.117
RAB34	1	-0.169	EFCC1	1	-0.116
RPL23A	1	-0.169	H2BC7	1	-0.114
SNORD42A	1	-0.169	DLG4	1	-0.112
SNORD4A	1	-0.169	RPS6	1	-0.112
SNORD4B	1	-0.169	ZNF385A	1	-0.111
GARS1	1	-0.165	FARP1	1	-0.109
GARS1-DT	1	-0.165	TNR	1	-0.107
CIDEB	1	-0.163	PARP4	1	-0.106
LTB4R	1	-0.163	ESM1	1	-0.104
MBOAT2	1	-0.159	MRPS18B	1	-0.103
SLC7A5	1	-0.156	PPP1R10	1	-0.103
NOL4	1	-0.155	ARHGAP27P2	1	-0.103
SLMAP	1	-0.151	SNX18	1	-0.103
MXRA8	1	-0.150	COL23A1	1	-0.101
LOC389705	1	-0.146	CD7	1	-0.101
HLX	1	-0.146	SECTM1	1	-0.101
ARHGEF7	1	-0.144	FGF20	1	-0.101
SLC33A1	1	-0.143			

Table S28. Number of differentially methylated cytosines (#DMCs) present in hypomethylated genes of Sadikovic Sotos episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
MIR200B	2	-0.447	C2orf81	1	-0.336
H1-1	2	-0.441	GCAT	1	-0.335
H3C1	2	-0.441	H1-0	1	-0.335
H4C1	2	-0.441	GALNT6	1	-0.333
TMEM190	2	-0.297	SLC4A8	1	-0.333
KDM5B	1	-0.650	CXCL12	1	-0.330
PCAT6	1	-0.650	F10	1	-0.329
FAM118A	1	-0.502	ACSL6	1	-0.322
RBBP8NL	1	-0.494	MEIKIN	1	-0.322
KCTD12	1	-0.490	AMN	1	-0.320
C17orf98	1	-0.484	TMEM105	1	-0.317
FAM83A	1	-0.469	GAB1	1	-0.313
FAM83A-AS1	1	-0.469	CDC14B	1	-0.311
KRBA1	1	-0.467	FOXL1	1	-0.311
APOA5	1	-0.457	CREB3L1	1	-0.309
ZPR1	1	-0.457	NUP210	1	-0.307
PRPSAP1	1	-0.444	NABP2	1	-0.306
TP73	1	-0.441	RNF41	1	-0.306
SDHAF1	1	-0.433	RASIP1	1	-0.304
SSC5D	1	-0.416	UBALD1	1	-0.302
TPPP3	1	-0.412	CHIT1	1	-0.299
ZDHHC1	1	-0.412	CERS2	1	-0.297
PPP1R1B	1	-0.412	LRRC2	1	-0.295
RPLP1	1	-0.411	TDGF1	1	-0.295
TET1	1	-0.410	CYGB	1	-0.289
LINC01623	1	-0.409	PRCD	1	-0.289
MIR200A	1	-0.408	COL12A1	1	-0.286
MIR429	1	-0.408	BBS4	1	-0.285
MYO1C	1	-0.398	HIGD2B	1	-0.285
SFN	1	-0.376	ANKRD13B	1	-0.282
DPM2	1	-0.369	GIT1	1	-0.282
RTKN	1	-0.363	DOK2	1	-0.279
TMEM72-AS1	1	-0.361	XPO7	1	-0.279
VAX2	1	-0.355	RD3	1	-0.279
HLX	1	-0.354	GPD1	1	-0.277
GPSM3	1	-0.353	ASIC4	1	-0.275
NOTCH4	1	-0.353	BTBD17	1	-0.267
DDR1	1	-0.352	KCNH2	1	-0.264
RWDD3	1	-0.348	CERS4	1	-0.255
TLCD4-RWDD3	1	-0.348	CKM	1	-0.253

Table S28 continued from previous page

CARD9	1	-0.344	IL11	1	-0.252
DNLZ	1	-0.344	ZNF423	1	-0.249
HYAL2	1	-0.344	NCS1	1	-0.247
LDB3	1	-0.341	COCH	1	-0.242
DLX3	1	-0.340	ZNF710	1	-0.239
ITGA2B	1	-0.339			

Table S29. Number of differentially methylated cytosines (#DMCs) present in hypermethylated genes of Weksberg Sotos episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
ZNF833P	13	0.281	PCDHGA9	1	0.283
KIRREL3	10	0.284	PCDHGB1	1	0.283
SHANK1	7	0.274	PCDHGB2	1	0.283
TCEAL7	6	0.290	PCDHGB3	1	0.283
SORCS2	4	0.263	PCDHGB4	1	0.283
HSPA12B	3	0.234	PCDHGB5	1	0.283
LCE3A	2	0.256	PCDHGB6	1	0.283
LOC93429	2	0.230	PCDHGB7	1	0.283
CIRBP-AS1	2	0.229	CRCT1	1	0.262
JAKMIP1	2	0.202	PCDHB7	1	0.224
TCHH	1	0.318	PCDHB8	1	0.224
PCDHGA1	1	0.283	FAM174C	1	0.217
PCDHGA10	1	0.283	MAT2B	1	0.213
PCDHGA11	1	0.283	DOCK6	1	0.211
PCDHGA2	1	0.283	DNAAF3	1	0.203
PCDHGA3	1	0.283	TNNI3	1	0.203
PCDHGA4	1	0.283	B3GALT5	1	0.203
PCDHGA5	1	0.283	PCDHB3	1	0.200
PCDHGA6	1	0.283	RPL27A	1	0.200
PCDHGA7	1	0.283	SNORA3A	1	0.200
PCDHGA8	1	0.283	SNORA3B	1	0.200

Table S30. Number of differentially methylated cytosines (#DMCs) present in hypomethylated genes of Weksberg Sotos episignature, with respective mean $\Delta\beta$ values.

Genes	#DMCs	mean $\Delta\beta$	Genes	#DMCs	mean $\Delta\beta$
PPT2	34	-0.334	PHAF1	1	-0.524
PPT2-EGFL8	34	-0.334	GUCY2D	1	-0.519
PRRT1	34	-0.334	SHANK2	1	-0.508
DDR1	30	-0.300	FAM238C	1	-0.505

Table S30 continued from previous page

LOC100507547	23	-0.341	RPL23	1	-0.493
LINC01623	22	-0.373	SNORA21	1	-0.493
NBPF8	22	-0.324	PRPSAP1	1	-0.491
TP73	17	-0.293	RPRD2	1	-0.474
TNXB	16	-0.309	KRBA1	1	-0.473
CASZ1	16	-0.282	B3GNT2	1	-0.456
NRXN2	13	-0.281	CLN6	1	-0.451
ZNF833P	13	-0.281	BARX1	1	-0.446
TMEM190	12	-0.291	TMEM14C	1	-0.446
SPIDR	11	-0.339	RECK	1	-0.437
SFN	11	-0.318	LEFTY2	1	-0.433
HOXC4	11	-0.270	ATP5MC2	1	-0.432
AIRE	10	-0.421	USP2	1	-0.431
PLEKHB1	10	-0.340	HCG25	1	-0.431
CMYA5	10	-0.334	TMEM132E	1	-0.431
KIRREL3	10	-0.284	LINC02709	1	-0.426
APOB	10	-0.281	SBF2-AS1	1	-0.426
BOLA2	10	-0.272	AZGP1P1	1	-0.426
VWA7	10	-0.266	DPH7	1	-0.421
WFIKKN2	9	-0.414	KRTAP2-3	1	-0.418
RTKN	9	-0.359	ESX1	1	-0.417
MIR4640	9	-0.358	ITPKA	1	-0.410
LRRC2	9	-0.354	SEMA3F	1	-0.409
MIR200B	9	-0.348	MIR572	1	-0.409
METTL24	9	-0.337	CLUH	1	-0.408
EGFLAM	9	-0.295	SLC25A51	1	-0.408
GRM4	9	-0.287	LY6E	1	-0.407
DMTN	9	-0.270	LY6E-DT	1	-0.407
CRISP2	9	-0.264	ERGIC3	1	-0.407
DLEU7	9	-0.264	CT62	1	-0.400
DLEU7-AS1	9	-0.264	PNLIPRP1	1	-0.398
MOB2	9	-0.253	ATP6V0CP3	1	-0.395
CLDN9	8	-0.436	PRPH2	1	-0.395
PITX1-AS1	8	-0.422	TEAD1	1	-0.395
KLHDC7B	8	-0.396	MIS18BP1	1	-0.394
TDGF1	8	-0.371	MDM2	1	-0.391
TEX26-AS1	8	-0.358	SZRD1P1	1	-0.391
HCAR1	8	-0.357	CYP46A1	1	-0.391
ZNF678	8	-0.356	KNL1	1	-0.391
NBPF20	8	-0.340	SEMA6C	1	-0.389
LINC03034	8	-0.329	MAMDC2	1	-0.388
SLC51A	8	-0.318	NTNG2	1	-0.388
NGFR	8	-0.317	FBXO3	1	-0.387
PIWIL2	8	-0.314	ADAMTSL2	1	-0.383

Table S30 continued from previous page

DPPA5	8	-0.311	RCN1	1	-0.382
FAM83A	8	-0.300	FGF18	1	-0.382
FAM83A-AS1	8	-0.300	CALML6	1	-0.382
TRIM41	8	-0.298	GLTPD2	1	-0.381
ADAMTS2	8	-0.297	VMO1	1	-0.381
SLC35F3	8	-0.295	LINC02396	1	-0.379
TBX1	8	-0.279	KCNF1	1	-0.378
PRSS42P	8	-0.277	LOC100287846	1	-0.378
TUB	8	-0.273	PYCR2	1	-0.378
LY6G5C	8	-0.265	ELL3	1	-0.378
CAMTA1	8	-0.261	SERF2	1	-0.378
LRRC4B	8	-0.257	EFNA1	1	-0.376
CRTAC1	8	-0.251	CHCHD6	1	-0.374
SOD3	7	-0.451	TMEM184B	1	-0.372
LYNX1	7	-0.428	GLIS1	1	-0.370
COX4I2	7	-0.404	IL1RL2	1	-0.370
MRGPRF	7	-0.395	QSOX2	1	-0.368
C17orf98	7	-0.392	DAZAP1	1	-0.368
LINC01101	7	-0.382	COX7A2	1	-0.366
TEX26	7	-0.379	CREBBP	1	-0.365
SSC5D	7	-0.368	PXT1	1	-0.361
CDX1	7	-0.364	MYO15A	1	-0.361
GPR45	7	-0.338	ADAMTS12	1	-0.361
MIR200A	7	-0.337	PDCD6IPP2	1	-0.360
H3C1	7	-0.329	ADCY5	1	-0.359
SPEG	7	-0.303	CERKL	1	-0.359
UCN3	7	-0.299	IL4I1	1	-0.358
CACNA2D4	7	-0.298	LOC100129175	1	-0.355
CHKB-CPT1B	7	-0.292	HNMT	1	-0.355
CHKB-DT	7	-0.292	SLC30A2	1	-0.355
CPT1B	7	-0.292	LMNTD2	1	-0.354
GCSAML	7	-0.291	RASSF7	1	-0.354
PACRG	7	-0.281	TMEM105	1	-0.353
RHOD	7	-0.278	SPRY4	1	-0.352
CCKBR	7	-0.276	HIBADH	1	-0.351
SHANK1	7	-0.274	AK7	1	-0.351
NTN1	7	-0.273	WLS	1	-0.351
TRIM2	7	-0.273	UNC5A	1	-0.349
ZDHHC19	7	-0.272	EYA2	1	-0.348
MYT1L	7	-0.272	KCTD14	1	-0.348
SRRM3	7	-0.260	NDUFC2-KCTD14	1	-0.348
FANCD2OS	7	-0.258	IGSF9B	1	-0.347
LINC00654	7	-0.224	ERF	1	-0.346
PTPRN2	7	-0.223	CCNO	1	-0.345

Table S30 continued from previous page

DDO	6	-0.385	TUBB6	1	-0.345
CHCT1	6	-0.377	AIFM3	1	-0.345
RBBP8NL	6	-0.377	STPG2	1	-0.345
ABCA13	6	-0.369	WTIP	1	-0.343
TRIM7	6	-0.356	PLIN5	1	-0.343
MIR429	6	-0.349	TIMM50	1	-0.342
USP44	6	-0.347	TNNC2	1	-0.341
KRT86	6	-0.345	FBXO24	1	-0.340
F10	6	-0.337	LRCH4	1	-0.340
PLBD1	6	-0.329	PCOLCE-AS1	1	-0.340
HSPB9	6	-0.329	CPN1	1	-0.340
KAT2A	6	-0.329	ENOX1	1	-0.339
WT1	6	-0.329	CXCL12	1	-0.339
SALL4	6	-0.323	PTPRQ	1	-0.338
KCNH2	6	-0.322	FTCD	1	-0.337
CRYL1	6	-0.318	MTMR7	1	-0.337
MYO7A	6	-0.315	H2AC21	1	-0.336
LINC00482	6	-0.314	H2BC21	1	-0.336
RASGEF1C	6	-0.314	GALNT17	1	-0.336
RNF112	6	-0.313	DYNC1I1	1	-0.335
MIRLET7BHG	6	-0.310	SNRPD1	1	-0.334
GRID1	6	-0.308	WRNIP1	1	-0.334
NTRK1	6	-0.308	SRARP	1	-0.332
LRMDA	6	-0.307	LAG3	1	-0.332
PIWIL3	6	-0.305	GRK2	1	-0.332
TOP1P2	6	-0.305	MIR1253	1	-0.331
LOC284798	6	-0.305	NUP210	1	-0.331
CHRM1	6	-0.304	ARHGEF28	1	-0.330
RBM24	6	-0.304	NPPC	1	-0.329
MIR3917	6	-0.302	MKRN2	1	-0.329
STMN1	6	-0.302	RNF175	1	-0.329
GPSM3	6	-0.298	CRYAA	1	-0.328
NOTCH4	6	-0.298	LMO1	1	-0.327
GDF2	6	-0.297	FBLIM1	1	-0.326
KHDC3L	6	-0.293	FAM98B	1	-0.326
ADARB2	6	-0.293	GTPBP10	1	-0.326
DPF3	6	-0.291	GDF6	1	-0.326
TCEAL7	6	-0.290	IFNLR1	1	-0.326
FSCN2	6	-0.287	TSLP	1	-0.325
RAP1GAP	6	-0.283	MN1	1	-0.325
LAMB2	6	-0.282	CSDC2	1	-0.325
CDH22	6	-0.281	ADGRB1	1	-0.324
DENND2B	6	-0.279	PPP1R12C	1	-0.324
CACNA1G	6	-0.279	CLMP	1	-0.324

Table S30 continued from previous page

PEX10	6	-0.278	BAIAP2L2	1	-0.323
MRGPRE	6	-0.275	BTF3	1	-0.323
DOC2GP	6	-0.275	HNF1B	1	-0.323
TRPM4	6	-0.271	LINC01168	1	-0.323
LRCOL1	6	-0.267	NKD2	1	-0.323
PACSIN1	6	-0.266	OR6S1	1	-0.322
NTM	6	-0.261	POLR3H	1	-0.322
IFITM10	6	-0.261	PLAAT5	1	-0.322
LMTK3	6	-0.260	AICDA	1	-0.321
SPON1	6	-0.260	MYADM	1	-0.321
PSMA8	6	-0.260	PTGDS	1	-0.320
GSDME	6	-0.258	PRMT8	1	-0.320
GATA4	6	-0.257	ACLY	1	-0.320
PLA2G4D	6	-0.254	GALR2	1	-0.319
HORMAD2	6	-0.251	SRP68	1	-0.319
MIR548H4	6	-0.251	RAET1K	1	-0.319
NOX5	6	-0.251	EPOP	1	-0.319
ADGRA1	6	-0.247	NHLH1	1	-0.318
MIR9-3HG	6	-0.246	SOX6	1	-0.318
KIAA1210	6	-0.242	SLC5A7	1	-0.318
LDHAL6A	6	-0.237	FOXA2	1	-0.318
DTX1	6	-0.237	KLF8	1	-0.318
LY6G6C	6	-0.211	CACNA1D	1	-0.317
MPIG6B	6	-0.211	PITX1	1	-0.316
ANXA11	5	-0.456	FCSK	1	-0.316
LINC00857	5	-0.456	NARS1	1	-0.315
DNAJB13	5	-0.439	GMNN	1	-0.314
CCDC17	5	-0.431	SLC11A1	1	-0.314
SDHAF1	5	-0.427	FAM110D	1	-0.314
ANKRD65	5	-0.398	PDE11A	1	-0.314
RWDD3	5	-0.396	SH3PXD2B	1	-0.313
TLCD4-RWDD3	5	-0.396	ALKBH8	1	-0.312
PRSS36	5	-0.396	TMEM37	1	-0.312
NBPF10	5	-0.390	LINC02881	1	-0.312
FYB2	5	-0.388	GAB1	1	-0.312
DMKN	5	-0.383	ANKRD13B	1	-0.311
RAD51B	5	-0.375	GIT1	1	-0.311
TVP23A	5	-0.359	NXPH3	1	-0.311
KRT81	5	-0.346	RBP4	1	-0.311
SCAND3	5	-0.338	RTEL1	1	-0.311
TMEM72-AS1	5	-0.331	RTEL1-TNFRSF6B	1	-0.311
GRIK4	5	-0.329	STMN3	1	-0.311
USP43	5	-0.327	XRCC6P5	1	-0.311
ACSL6	5	-0.327	KBTBD12	1	-0.310

Table S30 continued from previous page

MEIKIN	5	-0.327	EBF4	1	-0.310
PDZD7	5	-0.325	KLHL14	1	-0.309
MACF1	5	-0.324	GLI2	1	-0.309
S100A1	5	-0.324	PER2	1	-0.308
S100A13	5	-0.324	ARHGAP22	1	-0.308
MUC4	5	-0.321	TTPA	1	-0.307
ZNF710	5	-0.321	SCN1B	1	-0.307
RCOR2	5	-0.320	BTBD3	1	-0.306
ESPNL	5	-0.317	CRLF1	1	-0.306
MON1A	5	-0.312	ZNF670	1	-0.306
GCSAML-AS1	5	-0.304	ZNF670-ZNF695	1	-0.306
OR2C3	5	-0.304	ACSF2	1	-0.305
FBXO17	5	-0.303	CHAD	1	-0.305
ZFYVE28	5	-0.303	SYT13	1	-0.305
MYO1C	5	-0.302	ESRRB	1	-0.305
RPLP1	5	-0.299	IMPG1	1	-0.304
BHMT2	5	-0.298	SIX5	1	-0.304
DMGDH	5	-0.298	SLFN13	1	-0.303
TACC2	5	-0.292	LIMA1	1	-0.303
CCDC152	5	-0.292	TMEM63A	1	-0.302
PDLIM4	5	-0.292	TBXT	1	-0.302
KRT12	5	-0.291	GAL	1	-0.302
INSRR	5	-0.291	PTGES	1	-0.301
IZUMO1	5	-0.288	MEGF10	1	-0.301
HS3ST3B1	5	-0.287	GFPT2	1	-0.300
CCDC144NL	5	-0.286	ZNF281	1	-0.300
GPR85	5	-0.286	HSPB8	1	-0.300
NDRG4	5	-0.282	ANKRD63	1	-0.300
MATN4	5	-0.280	ESM1	1	-0.299
GJB6	5	-0.280	WDR19	1	-0.299
LINC00880	5	-0.280	CDC14B	1	-0.299
PKNOX2	5	-0.279	MLXIPL	1	-0.298
CYP1A1	5	-0.278	TCEA3	1	-0.298
LOXHD1	5	-0.277	ESAM	1	-0.298
NEURL3	5	-0.277	HSPA12A	1	-0.297
BEGAIN	5	-0.276	KAZN	1	-0.297
MT1A	5	-0.276	MIR137HG	1	-0.296
TACSTD2	5	-0.274	MT1JP	1	-0.295
PI16	5	-0.272	FAM245A	1	-0.295
MMRN2	5	-0.272	DOK2	1	-0.295
SNCG	5	-0.272	XPO7	1	-0.295
SYN2	5	-0.270	LINC00592	1	-0.295
TIMP4	5	-0.270	LINC02874	1	-0.295
SEPTIN8	5	-0.266	CECR2	1	-0.295

Table S30 continued from previous page

GJD3	5	-0.264	IP6K3	1	-0.295
RARA	5	-0.264	SRP9	1	-0.294
STRA6	5	-0.264	ABCA4	1	-0.294
RBM20	5	-0.263	MIR1275	1	-0.293
SLC38A11	5	-0.262	CTAGE1	1	-0.293
GRB7	5	-0.261	GPT2	1	-0.293
ANK1	5	-0.260	USP2-AS1	1	-0.292
HOXC5	5	-0.260	RIGI	1	-0.292
HOXC6	5	-0.260	MAD2L2	1	-0.292
M1AP	5	-0.258	B4GALT3	1	-0.292
CYGB	5	-0.256	PPOX	1	-0.292
PRCD	5	-0.256	ICMT	1	-0.292
ZBTB38	5	-0.256	CLEC4F	1	-0.292
MELTF	5	-0.254	LINC02731	1	-0.292
DNAH2	5	-0.252	ADAMTSL5	1	-0.292
POMC	5	-0.251	LRRC15	1	-0.292
PPP2R2A	5	-0.249	MCAM	1	-0.291
ASCL2	5	-0.249	NUP43	1	-0.291
FBXO47	5	-0.249	PCMT1	1	-0.291
NKX6-3	5	-0.247	TEF	1	-0.291
EMILIN1	5	-0.238	CRB3	1	-0.291
SECTM1	5	-0.238	SLC25A23	1	-0.291
TDRD1	5	-0.238	DUSP6	1	-0.290
CHRND	5	-0.237	CSHL1	1	-0.290
KCNH3	5	-0.235	IFITM2	1	-0.290
LINC00301	5	-0.235	ZEB2	1	-0.290
SHROOM1	5	-0.235	TLE6	1	-0.289
TAPBP	5	-0.234	SAXO1	1	-0.289
ZBTB22	5	-0.234	ST3GAL4	1	-0.289
SLC9C1	5	-0.232	CES1	1	-0.288
NOS3	5	-0.232	SDC3	1	-0.288
ARHGAP19-SLIT1	5	-0.228	BBS4	1	-0.288
SLIT1	5	-0.228	HIGD2B	1	-0.288
RNF39	5	-0.224	GAS2L3	1	-0.287
FOXL2NB	5	-0.223	SEC23B	1	-0.287
COL11A2	5	-0.216	GABRQ	1	-0.286
KDM5B	4	-0.476	CHL1	1	-0.286
PCAT6	4	-0.476	CAPSL	1	-0.286
SMAD3	4	-0.463	SLURP1	1	-0.286
TMEM88B	4	-0.435	LINC00606	1	-0.285
FAM118A	4	-0.424	DNAAF11	1	-0.285
KCTD12	4	-0.417	TPGS1	1	-0.284
ACKR2	4	-0.396	HSPA1A	1	-0.284
HIGD1A	4	-0.396	HSPA1L	1	-0.284

Table S30 continued from previous page

AHNAK	4	-0.390	PSKH2	1	-0.283
LOC284950	4	-0.390	CD164L2	1	-0.283
C1QTNF9	4	-0.390	GRHL1	1	-0.283
SPATA13	4	-0.390	SLC17A9	1	-0.283
SLC39A14	4	-0.387	LRRC47	1	-0.283
APOA5	4	-0.386	ARHGEF16	1	-0.283
ZPR1	4	-0.386	SMPD3	1	-0.283
LOC100131289	4	-0.383	CACNA1G-AS1	1	-0.282
ZMYND15	4	-0.382	EMP3	1	-0.281
LRCH1	4	-0.371	ODAD1	1	-0.281
PLA2G4E	4	-0.370	CRX	1	-0.281
GCAT	4	-0.369	ADAD1	1	-0.281
H1-0	4	-0.369	WDR59	1	-0.281
RIPK4	4	-0.360	PNLIPRP2	1	-0.281
COL13A1	4	-0.357	CRTC1	1	-0.280
ATP5MF	4	-0.348	GJC1	1	-0.280
ATP5MF-PTCD1	4	-0.348	COLEC12	1	-0.280
ZNF789	4	-0.348	PPFIA3	1	-0.280
UBALD1	4	-0.344	MAN1C1	1	-0.279
HYAL2	4	-0.341	SLC4A3	1	-0.279
CX3CL1	4	-0.340	PLEKHD1	1	-0.279
MIR346	4	-0.339	ARHGAP39	1	-0.279
EFCAB9	4	-0.339	C8orf82	1	-0.279
STK10	4	-0.339	LRRC24	1	-0.279
RASGEF1A	4	-0.336	AQP1	1	-0.279
FADD	4	-0.331	NRARP	1	-0.278
BMP7	4	-0.329	COL12A1	1	-0.278
ADARB1	4	-0.323	TNN	1	-0.278
SSR4P1	4	-0.323	DGKI	1	-0.277
MFSD1	4	-0.322	EXOC3L4	1	-0.277
RD3	4	-0.321	RPL24	1	-0.277
CHRD	4	-0.320	LINC01140	1	-0.277
H1-1	4	-0.318	FRMD4A	1	-0.276
H4C1	4	-0.318	PPEF2	1	-0.276
TH	4	-0.318	TACC1	1	-0.276
PEBP4	4	-0.315	EPHB1	1	-0.276
ALX4	4	-0.314	MAP3K5	1	-0.276
LINC01135	4	-0.312	FBXO44	1	-0.276
S100A10	4	-0.312	REST	1	-0.275
RGN	4	-0.310	ABCC12	1	-0.275
PLA2G3	4	-0.310	GGN	1	-0.275
KLHDC8A	4	-0.309	SPRED3	1	-0.275
KLF15	4	-0.309	RPS6KC1	1	-0.275
FAM228A	4	-0.307	IRF2BPL	1	-0.275

Table S30 continued from previous page

HAPLN1	4	-0.307	PLEKHM3	1	-0.275
FEZ1	4	-0.305	SORBS3	1	-0.274
LRRN4	4	-0.305	EFS	1	-0.274
COLEC11	4	-0.305	ASPG	1	-0.274
ADD3	4	-0.305	C9orf131	1	-0.273
ADD3-AS1	4	-0.305	TF	1	-0.273
TET1	4	-0.304	DLGAP4	1	-0.273
FBXO27	4	-0.304	MDGA1	1	-0.273
CPEB1	4	-0.303	SLC2A2	1	-0.273
TLX1NB	4	-0.302	ARF5	1	-0.273
ADIRF	4	-0.302	FSCN3	1	-0.273
AGAP11	4	-0.302	GCC1	1	-0.273
GPD1	4	-0.301	ISYNA1	1	-0.273
PRKCZ	4	-0.298	CFAP299	1	-0.273
CACNG6	4	-0.298	METTL21C	1	-0.272
IGFBP2	4	-0.297	NAALADL1	1	-0.272
PRPH	4	-0.296	PHOX2A	1	-0.272
PRSS41	4	-0.295	MTCL1	1	-0.272
PKD2L1	4	-0.295	MYBPC2	1	-0.271
PDGFRA	4	-0.295	SRRM4	1	-0.271
ALX3	4	-0.293	POLR2F	1	-0.271
MT1M	4	-0.289	SOX10	1	-0.271
GFRA3	4	-0.288	DVL3	1	-0.271
POU2AF1	4	-0.286	IFFO2	1	-0.271
CCDC85C	4	-0.286	RTN4RL2	1	-0.271
RBPJL	4	-0.285	MT1IP	1	-0.271
KIFC3	4	-0.283	IKZF3	1	-0.271
KIF12	4	-0.282	ZPBP2	1	-0.271
HOXB6	4	-0.281	APBB2	1	-0.270
HOXB7	4	-0.281	LIN54	1	-0.270
CDH16	4	-0.281	G3BP1	1	-0.270
RRAD	4	-0.281	LOC100652758	1	-0.270
NR5A1	4	-0.277	C2orf73	1	-0.270
RHBDL2	4	-0.277	PCOLCE2	1	-0.270
TMEM176A	4	-0.277	TBC1D31	1	-0.270
TMEM176B	4	-0.277	ST14	1	-0.269
KLHDC7A	4	-0.277	VSX2	1	-0.269
GFRA4	4	-0.276	LLCFC1	1	-0.269
PLXNB2	4	-0.275	PITPNC1	1	-0.269
DUSP5	4	-0.275	TMEM97	1	-0.269
BTBD17	4	-0.274	TNNT2	1	-0.269
LGALS8	4	-0.273	SPATA3	1	-0.269
LGALS8-AS1	4	-0.273	SPATA3-AS1	1	-0.269
ZNF541	4	-0.272	LINC01257	1	-0.268

Table S30 continued from previous page

RASIP1	4	-0.271	HELZ	1	-0.268
ARHGEF4	4	-0.266	NACAD	1	-0.268
SNRPF	4	-0.266	HPCAL4	1	-0.268
SH3RF3	4	-0.266	MAPRE2	1	-0.268
SH3RF3-AS1	4	-0.266	KRT28	1	-0.268
NANOS2	4	-0.264	MIR5700	1	-0.268
TBPL2	4	-0.264	CYS1	1	-0.268
SORCS2	4	-0.263	PRELID3B	1	-0.267
CARD14	4	-0.263	SLMO2-ATP5E	1	-0.267
C2CD6	4	-0.263	JPH2	1	-0.267
NECTIN4	4	-0.262	ARSI	1	-0.267
WIPF3	4	-0.261	SPAG17	1	-0.267
KCNG1	4	-0.260	CENPV	1	-0.267
PGPEP1L	4	-0.258	RAE1	1	-0.267
OTOP3	4	-0.257	COCH	1	-0.266
SDK1	4	-0.255	IGFN1	1	-0.266
CFAP77	4	-0.255	LINC02898	1	-0.266
GTF2A1L	4	-0.255	CHST1	1	-0.266
STON1-GTF2A1L	4	-0.255	CLIC6	1	-0.266
NBPF9	4	-0.254	KNCN	1	-0.266
DPPA4	4	-0.254	MKNK1-AS1	1	-0.266
EBF1	4	-0.254	EPAS1	1	-0.266
MACROD1	4	-0.253	ZBTB12	1	-0.266
TMPRSS12	4	-0.253	PCSK5	1	-0.266
ABHD11	4	-0.252	CXCL14	1	-0.266
EFNA3	4	-0.251	HROB	1	-0.266
ALOX5	4	-0.251	ANGPTL5	1	-0.265
WNT3	4	-0.250	CEP126	1	-0.265
TRIM61	4	-0.250	CCDC81	1	-0.265
GNG7	4	-0.249	RHOB	1	-0.265
MAMSTR	4	-0.249	MRPL9	1	-0.265
SLC22A18	4	-0.247	OAZ3	1	-0.265
OCA2	4	-0.246	MIR141	1	-0.264
CCDC186	4	-0.245	MIR200C	1	-0.264
SMIM24	4	-0.245	KCNA2	1	-0.264
DAGLA	4	-0.245	CCDC33	1	-0.264
RIMBP2	4	-0.244	MTAP	1	-0.264
DNALI1	4	-0.241	ZNF423	1	-0.264
SNIP1	4	-0.241	MYL10	1	-0.264
PRDM9	4	-0.241	STC2	1	-0.264
MGAT5B	4	-0.241	SMCO4	1	-0.264
RGMA	4	-0.240	C2orf50	1	-0.263
FAM131C	4	-0.239	FLJ33534	1	-0.263
PRKCH	4	-0.239	CLPSL1	1	-0.263

Table S30 continued from previous page

TMEM30B	4	-0.239	CHD6	1	-0.263
CDH20	4	-0.238	TNK1	1	-0.263
ASGR1	4	-0.238	BAHD1	1	-0.263
JUP	4	-0.237	NFKBIE	1	-0.263
CAMK2A	4	-0.235	ZNF287	1	-0.263
MFSD2B	4	-0.233	PNPT1	1	-0.263
PAX8	4	-0.233	SMPDL3A	1	-0.263
FGF20	4	-0.233	SYT15	1	-0.263
PRSS23	4	-0.232	RBL2	1	-0.263
TEX14	4	-0.230	UGT1A1	1	-0.263
CPLX1	4	-0.228	DISP1	1	-0.262
SLC9B2	4	-0.226	LRRC59	1	-0.262
LINC00111	4	-0.226	SCUBE3	1	-0.262
CCN1	4	-0.223	CADM3	1	-0.262
DDAH1	4	-0.223	CADPS	1	-0.262
ANXA2R-AS1	4	-0.222	LTK	1	-0.262
PLAT	4	-0.222	ODAD3	1	-0.262
CLSTN1	4	-0.222	RGL3	1	-0.262
PHYHIP	4	-0.219	CYB5B	1	-0.262
TTBK1	4	-0.218	KCNE1	1	-0.261
MAP3K6	4	-0.216	TMEM175	1	-0.261
TMEM225B	3	-0.544	KCNK5	1	-0.261
LINC02774	3	-0.438	MEFV	1	-0.261
ASIC1	3	-0.423	ADORA1	1	-0.261
SMARCD1	3	-0.423	CTTNBP2NL	1	-0.261
ANKRD20A19P	3	-0.423	LHX6	1	-0.261
NSD1	3	-0.420	LRIT1	1	-0.261
CERS2	3	-0.401	ENTPD4	1	-0.261
ATP1A2	3	-0.396	LOXL2	1	-0.261
SUGT1P3	3	-0.390	EFCC1	1	-0.260
TPTE2P5	3	-0.390	SYT7	1	-0.260
FAM163A	3	-0.385	IRX1	1	-0.260
COL2A1	3	-0.385	RNF111	1	-0.260
CIDEA	3	-0.384	SF3B5	1	-0.260
ZBTB16	3	-0.374	HLA-B	1	-0.259
HS6ST1	3	-0.374	CPE	1	-0.259
CATSPER4	3	-0.373	PRDM6	1	-0.259
SLC25A18	3	-0.372	RUBCNL	1	-0.259
CACNA1C	3	-0.369	N4BP1	1	-0.259
GALNT6	3	-0.366	SKP1	1	-0.259
SLC4A8	3	-0.366	KCNQ1	1	-0.259
TMC2	3	-0.361	PSMC1	1	-0.259
ESRP2	3	-0.354	KRT42P	1	-0.258
DPM2	3	-0.353	SBSPON	1	-0.258

Table S30 continued from previous page

CYP7B1	3	-0.353	DKK3	1	-0.258
ORMDL3	3	-0.351	MEGF8	1	-0.258
CPNE5	3	-0.350	TMEM145	1	-0.258
CERS4	3	-0.347	CCDC149	1	-0.257
SLC22A8	3	-0.344	RFC5	1	-0.257
TTYH1	3	-0.343	OR10J5	1	-0.257
LOC286177	3	-0.341	SYNDIG1L	1	-0.257
SINHCAF	3	-0.339	KCTD15	1	-0.257
ENTPD2	3	-0.338	HMSD	1	-0.257
BSND	3	-0.336	SCD5	1	-0.257
GPR25	3	-0.334	MIR4686	1	-0.257
TDRP	3	-0.334	ADGRD2	1	-0.257
TBC1D16	3	-0.333	FGF6	1	-0.257
SLC16A12	3	-0.331	GLI1	1	-0.256
SHISA8	3	-0.329	INHBE	1	-0.256
NOM1	3	-0.329	C1QTNF1	1	-0.256
PPP1R1B	3	-0.328	C1QTNF1-AS1	1	-0.256
MYOM2	3	-0.326	ELF3	1	-0.256
LBX2	3	-0.325	RGS20	1	-0.256
LBX2-AS1	3	-0.325	ADGRD1	1	-0.256
CACNG2	3	-0.323	TBX4	1	-0.256
ZNF536	3	-0.323	ARHGEF33	1	-0.256
ARID5B	3	-0.322	LOC375196	1	-0.256
CREB3L1	3	-0.321	METAP1D	1	-0.256
CFAP73	3	-0.321	SP6	1	-0.256
KLK1	3	-0.318	YIPF4	1	-0.256
KLK15	3	-0.318	MCM10	1	-0.256
DHRS3	3	-0.318	LRP5	1	-0.255
FAM217A	3	-0.316	DNAH5	1	-0.255
TEX56P	3	-0.316	LCN10	1	-0.255
C10orf82	3	-0.316	LOC100128593	1	-0.255
ETNK1	3	-0.316	SYN1	1	-0.255
CKM	3	-0.312	NID2	1	-0.255
ARL4A	3	-0.311	ZC3HAV1	1	-0.255
RBPMS2	3	-0.310	PKP1	1	-0.255
GCK	3	-0.310	PDXK	1	-0.255
AKAP10	3	-0.310	PEDS1	1	-0.255
KCNC4	3	-0.310	PEDS1-UBE2V1	1	-0.255
VAX2	3	-0.309	RNF207	1	-0.255
ESRRG	3	-0.308	GAS2L2	1	-0.254
GARIN5B	3	-0.307	USP3	1	-0.254
FLJ12825	3	-0.307	KCNQ2	1	-0.254
COL23A1	3	-0.306	KRTAP5-11	1	-0.254
PITRM1	3	-0.306	NR2F1	1	-0.254

Table S30 continued from previous page

MYL9	3	-0.305	ESRP1	1	-0.254
VRTN	3	-0.302	LINC02894	1	-0.254
ASPHD1	3	-0.301	TPBGL	1	-0.254
SEZ6L2	3	-0.301	IL1A	1	-0.253
TMEM263	3	-0.300	SLC15A4	1	-0.253
FAAP20	3	-0.300	TBC1D14	1	-0.253
HILPDA	3	-0.299	SOST	1	-0.253
MIR9-1	3	-0.298	KRTAP5-5	1	-0.253
MIR9-1HG	3	-0.298	IDE	1	-0.253
DES	3	-0.296	HAPLN2	1	-0.253
MIR4763	3	-0.294	FAM107A	1	-0.253
MIRLET7A3	3	-0.294	MYL7	1	-0.253
MIRLET7B	3	-0.294	FSTL4	1	-0.252
APBA2	3	-0.294	COBL	1	-0.252
AGBL4	3	-0.292	KRTAP10-6	1	-0.252
ELN	3	-0.292	TSPEAR	1	-0.252
PLCD3	3	-0.290	GPR143	1	-0.252
NOVA2	3	-0.289	SGPP2	1	-0.251
GIPC2	3	-0.288	CSNK1E	1	-0.251
RADIL	3	-0.287	TPTEP2	1	-0.251
BCL7C	3	-0.286	IL17D	1	-0.251
CTF1	3	-0.286	COL7A1	1	-0.251
MIR148A	3	-0.285	FANCG	1	-0.251
RAI1	3	-0.283	NFATC2	1	-0.251
FERMT1	3	-0.283	PNCK	1	-0.251
CRISPLD2	3	-0.283	SLC6A8	1	-0.251
ANKRD2	3	-0.282	PAQR5	1	-0.251
DLX4	3	-0.282	EXOC7	1	-0.251
NUAK1	3	-0.281	RAN	1	-0.251
CHST2	3	-0.281	ADAM30	1	-0.250
C9orf50	3	-0.281	NBPF14	1	-0.250
PTPRF	3	-0.281	UBE2E2	1	-0.250
SPATS1	3	-0.281	MIR4472-1	1	-0.250
TMEM151B	3	-0.281	GABARAPL1	1	-0.250
ABCG4	3	-0.281	EIF2AK1	1	-0.250
CACNG8	3	-0.280	FGF17	1	-0.250
ATP2B2	3	-0.280	KAAG1	1	-0.249
ITGA3	3	-0.280	VWA5B1	1	-0.249
LINC00908	3	-0.280	ELAVL3	1	-0.249
WWC2	3	-0.279	ZNF32-AS3	1	-0.249
APCDD1L	3	-0.279	SIAH3	1	-0.249
HHIP	3	-0.278	ECHDC2	1	-0.249
HHIP-AS1	3	-0.278	BICDL2	1	-0.249
MT1B	3	-0.277	LOC100128770	1	-0.249

Table S30 continued from previous page

ATP6V1B1	3	-0.277	KCNA5	1	-0.249
PP12613	3	-0.277	ISOC2	1	-0.249
SMIM43	3	-0.277	OLIG1	1	-0.249
CDH4	3	-0.276	WDR35	1	-0.249
STEAP2	3	-0.276	SULT2B1	1	-0.249
AMN	3	-0.276	PAX5	1	-0.249
GPR37L1	3	-0.274	ITGB3	1	-0.249
LYSMD4	3	-0.274	SPOCD1	1	-0.249
CRB2	3	-0.273	DNASE2	1	-0.249
OPCML	3	-0.273	KLF1	1	-0.249
DOC2A	3	-0.273	SEMA3C	1	-0.248
C1QL4	3	-0.271	KATNAL2	1	-0.248
CRHR1	3	-0.270	PIAS2	1	-0.248
NUP58	3	-0.270	DRGX	1	-0.248
OBSCN	3	-0.269	TP53BP1	1	-0.248
OBSCN-AS1	3	-0.269	LAMP3	1	-0.248
TMPRSS6	3	-0.269	GLIPR2	1	-0.248
EBF2	3	-0.268	DNPEP	1	-0.248
ZBTB44	3	-0.268	CDKN1C	1	-0.248
NTN5	3	-0.268	AKNAD1	1	-0.247
SEC1P	3	-0.268	GPSM2	1	-0.247
ABCA7	3	-0.267	TTLL10	1	-0.247
ARHGAP45	3	-0.267	TULP1	1	-0.247
KCNA3	3	-0.267	ETV7	1	-0.247
ELAVL4	3	-0.266	CFAP221	1	-0.247
MOS	3	-0.266	GDPD1	1	-0.246
PLA2G5	3	-0.266	MIR615	1	-0.246
SP7	3	-0.266	BEND5	1	-0.246
CNGB1	3	-0.265	SMOC1	1	-0.246
ZNF518B	3	-0.265	SLC45A1	1	-0.246
GSTA4	3	-0.264	STX12	1	-0.246
PIWIL1	3	-0.264	CCDC105	1	-0.246
POU2AF2	3	-0.263	BTN2A3P	1	-0.246
TMEM266	3	-0.263	ITGB4	1	-0.246
SNAI2	3	-0.262	CTSD	1	-0.246
KRT18P55	3	-0.262	CDC42EP4	1	-0.246
LINC00473	3	-0.262	ADM2	1	-0.246
ZSCAN10	3	-0.262	SBF1	1	-0.246
ALDH3A1	3	-0.261	DCDC1	1	-0.245
ESYT3	3	-0.261	KRT7	1	-0.245
MEOX1	3	-0.260	CD207	1	-0.245
ASB16	3	-0.260	CNKSR1	1	-0.245
IL17REL	3	-0.259	RORA	1	-0.245
BTBD19	3	-0.259	SYBU	1	-0.245

Table S30 continued from previous page

LUC7L	3	-0.258	TCEA1	1	-0.245
LRRC14B	3	-0.258	MFSD4A	1	-0.244
FAM83C	3	-0.258	SLC4A9	1	-0.244
GAPDHS	3	-0.257	CD1D	1	-0.244
CELF4	3	-0.257	CAP2	1	-0.244
LGR6	3	-0.256	EMX1	1	-0.244
OTOP2	3	-0.256	TACR2	1	-0.244
USH1G	3	-0.256	LINC00620	1	-0.244
LOC653513	3	-0.256	RNF8	1	-0.244
PDE4DIP	3	-0.256	EVI5L	1	-0.244
FABP7	3	-0.256	METRN	1	-0.244
COX16	3	-0.256	MCMBP	1	-0.243
SYNJ2BP-COX16	3	-0.256	IHO1	1	-0.243
H2AJ	3	-0.255	FN3KRP	1	-0.243
H4C16	3	-0.255	COLGALT2	1	-0.243
SRCIN1	3	-0.254	CPNE4	1	-0.243
DAAM2	3	-0.254	DUOX2	1	-0.243
SRPRB	3	-0.253	DUOXA2	1	-0.243
MEGF11	3	-0.253	VSTM5	1	-0.242
IQSEC3	3	-0.253	LINC00261	1	-0.242
LOC574538	3	-0.253	AQP5	1	-0.242
PTCRA	3	-0.253	ISLR2	1	-0.242
AHRR	3	-0.253	LOC283731	1	-0.242
GRIP2	3	-0.252	ZIC4	1	-0.242
TTC22	3	-0.252	KREMEN2	1	-0.242
GAL3ST2	3	-0.252	CLDN11	1	-0.242
ERRFI1	3	-0.251	POLR2D	1	-0.242
KNDC1	3	-0.251	PAMR1	1	-0.242
CRABP1	3	-0.250	TMEM216	1	-0.242
DIRAS1	3	-0.250	LOC283194	1	-0.241
SLCO6A1	3	-0.250	PF4V1	1	-0.241
SLC24A4	3	-0.248	CIRBP	1	-0.241
RBM46	3	-0.248	SV2C	1	-0.241
ZBTB47	3	-0.247	WDR27	1	-0.241
IER5	3	-0.247	UGT3A1	1	-0.241
IFITM5	3	-0.247	LARP1B	1	-0.241
UCMA	3	-0.247	FSTL1	1	-0.241
RTL4	3	-0.247	WWC1	1	-0.241
TRPC5	3	-0.247	RASGRF1	1	-0.241
RGS22	3	-0.247	ARHGAP12	1	-0.241
GJA3	3	-0.246	CALCA	1	-0.241
BORCS7-ASMT	3	-0.246	LINC02418	1	-0.241
FGFBP1	3	-0.246	SLC36A2	1	-0.240
ARL5C	3	-0.246	SYT8	1	-0.240

Table S30 continued from previous page

HLA-DQB2	3	-0.245	MYO3A	1	-0.240
HEXIM1	3	-0.244	ASCL5	1	-0.240
LINC01342	3	-0.243	TSNARE1	1	-0.240
FBLN7	3	-0.243	WNT9B	1	-0.240
LPIN3	3	-0.243	LINC01750	1	-0.240
ILDR1	3	-0.242	ALDH1L1	1	-0.240
ALPI	3	-0.242	ALDH1L1-AS2	1	-0.240
IQCJ-SCHIP1	3	-0.241	HES5	1	-0.240
SCHIP1	3	-0.241	CCDC172	1	-0.239
KIF26B	3	-0.241	LEP	1	-0.239
GTF2B	3	-0.240	PSD2	1	-0.239
BMP2	3	-0.240	ADAP1	1	-0.239
TRIM15	3	-0.239	COX19	1	-0.239
LHPP	3	-0.238	TIAM2	1	-0.239
CSPG4	3	-0.238	ARID3C	1	-0.239
TRPV3	3	-0.238	AHDC1	1	-0.239
GLB1L2	3	-0.238	ATP12A	1	-0.239
AMDHD1	3	-0.238	ADAM33	1	-0.239
MARCHF10	3	-0.238	APBA1	1	-0.238
ADORA2A-AS1	3	-0.237	CPXM2	1	-0.238
UPB1	3	-0.237	ULBP1	1	-0.238
PCBP3	3	-0.237	LINC03040	1	-0.238
CCDC162P	3	-0.236	SAAL1	1	-0.238
MTMR11	3	-0.236	SLC15A3	1	-0.238
USH1C	3	-0.235	PTHLH	1	-0.237
ATP8A2	3	-0.235	RHBDL1	1	-0.237
SHTN1	3	-0.235	STUB1	1	-0.237
LMO7DN	3	-0.234	NPFFR1	1	-0.237
VGLL4	3	-0.234	MLPH	1	-0.237
HSPA12B	3	-0.234	FAM184B	1	-0.237
CFB	3	-0.234	HSPB2	1	-0.237
LYPD3	3	-0.233	HSPB2-C11orf52	1	-0.237
ADAMTS7	3	-0.233	CSMD1	1	-0.237
PFN3	3	-0.233	MT3	1	-0.237
KLHL29	3	-0.233	PRR15	1	-0.237
EDN2	3	-0.232	VENTX	1	-0.237
RIMS3	3	-0.232	GRHL2	1	-0.237
DUSP29	3	-0.231	MIR3529	1	-0.237
TCP11	3	-0.231	MYMK	1	-0.237
MIR193A	3	-0.231	CHADL	1	-0.236
EXT1	3	-0.230	LACTB	1	-0.236
SBK2	3	-0.230	AKIP1	1	-0.236
PTPRU	3	-0.230	RNU5F-1	1	-0.236
TRPC3	3	-0.230	SLC30A10	1	-0.236

Table S30 continued from previous page

CNMD	3	-0.227	PSEN1	1	-0.236
SEMA4B	3	-0.226	CDYL	1	-0.236
ZNF664-RFLNA	3	-0.226	LINC01548	1	-0.236
KLK4	3	-0.226	WSCD2	1	-0.236
KCP	3	-0.226	PANX2	1	-0.236
MALL	3	-0.225	ATXN10	1	-0.235
CD300LG	3	-0.225	TPO	1	-0.235
FOXL2	3	-0.225	MAP3K9	1	-0.235
HR	3	-0.224	CDC20B	1	-0.235
DSCAML1	3	-0.224	GPX8	1	-0.235
LDHC	3	-0.223	CCDC113	1	-0.235
RAB5IF	3	-0.221	ACTL8	1	-0.235
TGIF2-RAB5IF	3	-0.221	GAD1	1	-0.235
NHSL1	3	-0.220	GRIN3B	1	-0.235
IL22RA1	3	-0.220	GSTO2	1	-0.235
OTOG	3	-0.220	HID1	1	-0.234
FAM166C	3	-0.220	PHLDA3	1	-0.234
OTOF	3	-0.220	PLEK2	1	-0.234
GORASP2	3	-0.220	SLC25A25-AS1	1	-0.234
KDM2B	3	-0.219	FBP2	1	-0.234
ECT2L	3	-0.218	EIF2D	1	-0.234
MYOM3	3	-0.217	APCS	1	-0.234
LINC01749	3	-0.217	KLHL33	1	-0.234
AQP11	3	-0.217	STUM	1	-0.234
MAFA	3	-0.216	KIT	1	-0.234
SPSB1	3	-0.214	HOTAIRM1	1	-0.234
COLCA1	3	-0.212	HOXA1	1	-0.234
POU2AF3	3	-0.212	PXDNL	1	-0.234
SLC4A11	3	-0.212	FOXI2	1	-0.234
DPEP3	3	-0.212	ICE1	1	-0.234
USP29	3	-0.212	DUSP4	1	-0.234
LGI3	3	-0.211	DMBX1	1	-0.234
SFTPC	3	-0.211	NAGA	1	-0.234
GALP	2	-0.465	PHETA2	1	-0.234
MANBA	2	-0.455	CTTN	1	-0.234
DUSP15	2	-0.454	IQCE	1	-0.233
FOXS1	2	-0.454	VPS13A	1	-0.233
APCDD1	2	-0.448	ATP1A1	1	-0.233
CABP1	2	-0.448	SLC29A4	1	-0.233
SUGT1P4-STRA6LP	2	-0.445	HAGLR	1	-0.233
SUGT1P4-STRA6LP-CCDC180	2	-0.445	S1PR3	1	-0.233
ELFN2	2	-0.441	PAX7	1	-0.233
PRSS56	2	-0.436	ADAM10	1	-0.233
FHL1P1	2	-0.424	GPAM	1	-0.232

Table S30 continued from previous page

SEC62	2	-0.424	OR10C1	1	-0.232
RGPD8	2	-0.415	FUOM	1	-0.232
LINC02381	2	-0.414	SLCO5A1	1	-0.232
ITGA2B	2	-0.407	CDH3	1	-0.232
ADAM5	2	-0.400	SCG5	1	-0.232
SLC22A11	2	-0.395	ZNF292	1	-0.232
NGEF	2	-0.395	ST6GALNAC1	1	-0.232
EXOSC2	2	-0.393	CHRDL2	1	-0.232
OCSTAMP	2	-0.390	FAM167A	1	-0.232
ALOXE3	2	-0.386	PTCH1	1	-0.232
LNX1	2	-0.379	MFSD14CP	1	-0.232
LNX1-AS2	2	-0.379	HTR4	1	-0.232
MFN2	2	-0.378	LINC01088	1	-0.232
IL17RA	2	-0.373	MRPL58	1	-0.232
GPX5	2	-0.372	FAM234A	1	-0.232
NCAM2	2	-0.370	MMP19	1	-0.232
KALRN	2	-0.369	GATA3	1	-0.232
CLMN	2	-0.368	GATA3-AS1	1	-0.232
LDB3	2	-0.367	ASB18	1	-0.231
ATG5	2	-0.365	CFAP100	1	-0.231
GDF5	2	-0.363	HLF	1	-0.231
CARD9	2	-0.363	UQCC4	1	-0.231
DNLZ	2	-0.363	STAC2	1	-0.231
RASAL1	2	-0.363	PVALB	1	-0.231
C2orf81	2	-0.361	SILC1	1	-0.231
TNPO3	2	-0.361	IL18	1	-0.231
TPI1P2	2	-0.361	TEX12	1	-0.231
SGMS2	2	-0.359	DLX6	1	-0.231
NKAPL	2	-0.356	DLX6-AS1	1	-0.231
ZKSCAN4	2	-0.356	ITIH5	1	-0.230
CPEB1-AS1	2	-0.355	CNTFR	1	-0.230
KIF3C	2	-0.355	CNTFR-AS1	1	-0.230
SFXN3	2	-0.354	OSMR	1	-0.230
FHAD1	2	-0.354	PRRX2	1	-0.230
SCRT2	2	-0.353	NTNG1	1	-0.230
SRXN1	2	-0.353	ANOS1	1	-0.230
SBK3	2	-0.352	KRT8P41	1	-0.230
TMEM114	2	-0.352	MIR5691	1	-0.230
RAB11FIP4	2	-0.352	SCUBE2	1	-0.230
DOCK9	2	-0.351	LRRN2	1	-0.230
IGLON5	2	-0.350	MDM4	1	-0.230
USF3	2	-0.348	KLHL30	1	-0.230
MIR4499	2	-0.346	ROPN1	1	-0.229
KCNQ1DN	2	-0.343	PTPN21	1	-0.229

Table S30 continued from previous page

IL17C	2	-0.343	IGFBPL1	1	-0.229
FRAS1	2	-0.342	LINC00922	1	-0.229
FBXL5	2	-0.340	MOG	1	-0.229
LEFTY1	2	-0.340	KBTBD11	1	-0.229
COL9A2	2	-0.340	CRIP2	1	-0.229
CHM	2	-0.339	LRRC28	1	-0.228
HLX	2	-0.337	TTC23	1	-0.228
MAPK4	2	-0.336	DMRTB1	1	-0.228
NFIX	2	-0.336	ZNF564	1	-0.228
REM2	2	-0.336	ZNF709	1	-0.228
MORF4L1	2	-0.334	RASSF4	1	-0.228
VWA1	2	-0.334	IGDCC4	1	-0.228
HEYL	2	-0.334	PLTP	1	-0.228
TEAD3	2	-0.333	SHISAL1	1	-0.228
HLCS	2	-0.331	MAD1L1	1	-0.228
SNTG2	2	-0.329	ANTXR2	1	-0.228
ARHGEF15	2	-0.328	CLCNKB	1	-0.228
RBAK-RBAKDN	2	-0.328	NKPD1	1	-0.228
RBAKDN	2	-0.328	PBX4	1	-0.227
RCC2	2	-0.325	TBX5	1	-0.227
NES	2	-0.324	CDH15	1	-0.227
C11orf97	2	-0.322	MAP6D1	1	-0.227
WNT7B	2	-0.322	POLR2L	1	-0.227
PRELID2	2	-0.321	TSPAN4	1	-0.227
LINC00189	2	-0.321	RARRES1	1	-0.227
CAVIN3	2	-0.321	FKBP6	1	-0.227
KRTAP2-2	2	-0.320	LOC541473	1	-0.227
DAB2IP	2	-0.318	NSUN5P2	1	-0.227
PRICKLE1	2	-0.317	STAG3L3	1	-0.227
ALDH1A3	2	-0.316	ARFGEF3	1	-0.227
PLD6	2	-0.316	C1QTNF12	1	-0.227
NAT8L	2	-0.316	CTRC	1	-0.227
ANGPTL2	2	-0.314	ALKAL1	1	-0.227
RALGPS1	2	-0.314	REM1	1	-0.227
KRTAP2-1	2	-0.314	PKLR	1	-0.227
IBTK	2	-0.314	SGSM1	1	-0.226
FOXL1	2	-0.313	BCL6B	1	-0.226
DNAJB5	2	-0.313	MIR497	1	-0.226
CHST5	2	-0.311	MIR497HG	1	-0.226
CFAP69	2	-0.311	FIGNL1	1	-0.226
CD160	2	-0.311	DOCK8	1	-0.226
PLEKHB2	2	-0.311	DOCK8-AS1	1	-0.226
RASD2	2	-0.310	RAPGEFL1	1	-0.226
MT1L	2	-0.309	FKBP10	1	-0.226

Table S30 continued from previous page

TPPP3	2	-0.309	P3H4	1	-0.226
ZDHHC1	2	-0.309	RHOBTB1	1	-0.226
DLX3	2	-0.309	PPARG	1	-0.226
SNX31	2	-0.309	GALNT8	1	-0.226
ZBTB7C	2	-0.309	CCNJ	1	-0.225
THSD4	2	-0.308	ENTPD1-AS1	1	-0.225
RPS6KA2	2	-0.308	ST18	1	-0.225
SPICE1	2	-0.307	TRPM8	1	-0.225
SLC7A4	2	-0.306	INHBA	1	-0.225
C10orf90	2	-0.306	INHBA-AS1	1	-0.225
TRIM31	2	-0.305	COL1A2	1	-0.225
ASPDH	2	-0.304	TAS1R1	1	-0.225
RIIAD1	2	-0.303	ZBTB48	1	-0.225
SETDB1	2	-0.303	LINC00112	1	-0.225
VIPR2	2	-0.302	LINC00479	1	-0.225
AGPAT2	2	-0.301	PSMD1	1	-0.225
OVCH1-AS1	2	-0.301	ARHGAP21	1	-0.224
CHIT1	2	-0.300	MIR4256	1	-0.224
SYTL3	2	-0.299	WNT2B	1	-0.224
BNIP5	2	-0.299	SHCBP1L	1	-0.224
C1QL1	2	-0.298	ANKRD33B	1	-0.224
MIR183	2	-0.297	CLEC3B	1	-0.224
MIR96	2	-0.297	CCDC197	1	-0.224
KLHL21	2	-0.297	MBP	1	-0.224
MGAT4C	2	-0.296	TOM1L1	1	-0.224
MAP3K15	2	-0.295	TOB2P1	1	-0.224
WTAPP1	2	-0.295	SCMH1	1	-0.224
LINC00881	2	-0.295	MT1DP	1	-0.223
ETV1	2	-0.294	LINC02092	1	-0.223
FANCI	2	-0.294	AKAP7	1	-0.223
CNRIP1	2	-0.293	CCDC63	1	-0.223
CTIF	2	-0.293	CCNI2	1	-0.223
LGI1	2	-0.292	CXCL16	1	-0.223
MIR935	2	-0.292	KIF11	1	-0.223
SLC22A3	2	-0.291	NPHS1	1	-0.223
B4GALNT1	2	-0.290	SLC44A3	1	-0.223
B3GNT3	2	-0.289	TESC	1	-0.223
RLF	2	-0.289	FAM124A	1	-0.223
TLX1	2	-0.288	RGS12	1	-0.223
PKHD1L1	2	-0.288	POMGNT2	1	-0.222
CCDC8	2	-0.287	LINC03042	1	-0.222
NABP2	2	-0.287	ANXA13	1	-0.222
RNF41	2	-0.287	ANKDD1A	1	-0.222
PRRT4	2	-0.286	LINC00518	1	-0.222

Table S30 continued from previous page

OPRD1	2	-0.286	GPR68	1	-0.222
ERCC1	2	-0.286	PTP4A1	1	-0.222
FOSB	2	-0.286	NHERF2	1	-0.222
WFDC5	2	-0.285	SPTBN4	1	-0.222
ZDHHC13	2	-0.285	TNR	1	-0.222
SLC37A2	2	-0.285	USP47	1	-0.222
ROBO3	2	-0.285	PPP1R32	1	-0.222
ESPNP	2	-0.285	LYPD1	1	-0.222
HPN	2	-0.285	LAMA2	1	-0.222
COL6A1	2	-0.285	ATP8B3	1	-0.222
FOXE3	2	-0.284	PDE1B	1	-0.222
DCTN2	2	-0.284	KCNJ1	1	-0.221
KIF5A	2	-0.284	SIX4	1	-0.221
ATOH8	2	-0.284	ALDH1L2	1	-0.221
TRARG1	2	-0.284	LTBR	1	-0.221
LRP4	2	-0.284	NCAM1	1	-0.221
GPR108	2	-0.283	NCAM1-AS1	1	-0.221
TRIP10	2	-0.283	SUCLA2	1	-0.221
RGR	2	-0.283	KIF13A	1	-0.221
ARL4C	2	-0.283	PDE4A	1	-0.221
CNTN1	2	-0.283	CLCNKA	1	-0.221
GPRC5A	2	-0.283	HSPB7	1	-0.221
ZNF831	2	-0.283	RCN3	1	-0.221
LANCL2	2	-0.282	SLC22A1	1	-0.221
LASP1	2	-0.281	LPL	1	-0.221
P2RY6	2	-0.281	SCUBE1	1	-0.220
C16orf74	2	-0.281	DIXDC1	1	-0.220
SLC5A1	2	-0.281	CEP128	1	-0.220
TRIM29	2	-0.281	TSHR	1	-0.220
SLC7A9	2	-0.281	COG2	1	-0.220
ZC2HC1C	2	-0.280	ST8SIA2	1	-0.220
VPS18	2	-0.280	TDRD9	1	-0.219
TFAP2A	2	-0.279	ANKRD29	1	-0.219
LHFPL2	2	-0.279	IGFBP1	1	-0.219
NCS1	2	-0.279	SH2D4B	1	-0.219
NTMT1	2	-0.279	LINC01558	1	-0.219
ONECUT1	2	-0.279	CELF1	1	-0.219
HDHD5	2	-0.279	DNAJC2	1	-0.219
HDHD5-AS1	2	-0.279	PSMC2	1	-0.219
PPP1R3G	2	-0.279	RGS17	1	-0.219
ATOH7	2	-0.277	KCTD13	1	-0.219
LIMS2	2	-0.277	RPL37	1	-0.219
SLC9A3	2	-0.277	SNORD72	1	-0.219
WNT6	2	-0.277	CLDN5	1	-0.219

Table S30 continued from previous page

DNAI3	2	-0.277	FABP6	1	-0.218
SLC25A42	2	-0.277	FGF9	1	-0.218
TNFSF9	2	-0.276	ATF6B	1	-0.218
TRIM44	2	-0.276	MACROH2A1	1	-0.218
IGFBP7-AS1	2	-0.276	PNMA8B	1	-0.218
ANG	2	-0.275	PPP5D1P	1	-0.218
RNASE4	2	-0.275	AHCTF1P1	1	-0.218
KAZALD1	2	-0.274	SMKR1	1	-0.218
LINC02897	2	-0.274	BFSP2	1	-0.218
FNDC5	2	-0.274	CELSR1	1	-0.218
FBXO2	2	-0.273	GJB4	1	-0.218
BHMT	2	-0.273	ATP1A3	1	-0.217
ZNF503	2	-0.273	KDELR1	1	-0.217
ZNF503-AS2	2	-0.273	RFLNA	1	-0.217
WWC2-AS2	2	-0.273	NID1	1	-0.217
PDE2A	2	-0.273	CHRNA5	1	-0.217
CACNA1S	2	-0.272	ZNF48	1	-0.217
RTP5	2	-0.272	DYSF	1	-0.217
KRT83	2	-0.272	FDXR	1	-0.217
ATP8B5P	2	-0.272	ZNF649	1	-0.217
ARHGDIG	2	-0.272	NCOA4	1	-0.217
PDIA2	2	-0.272	PARG	1	-0.217
TPM1	2	-0.272	HELT	1	-0.217
DACT3	2	-0.271	SALL2	1	-0.217
ECEL1	2	-0.271	CHURC1-FNTB	1	-0.217
PLEKHG5	2	-0.271	RAB15	1	-0.217
RUFY4	2	-0.271	FKBP1A-SDCBP2	1	-0.217
CD177	2	-0.271	SDCBP2	1	-0.217
PSMB2	2	-0.270	ATP9B	1	-0.217
TFAP2E	2	-0.270	SLC35C1	1	-0.217
CCSAP	2	-0.270	THBS1	1	-0.216
LEXM	2	-0.270	RPH3AL	1	-0.216
SPTB	2	-0.270	PDE9A	1	-0.216
ZNF385C	2	-0.269	SLFN12	1	-0.216
NODAL	2	-0.269	NPR2	1	-0.216
COL11A1	2	-0.269	GATA2	1	-0.216
EMC4	2	-0.268	LRRC71	1	-0.216
PAK6	2	-0.268	MPO	1	-0.216
PAK6-AS1	2	-0.268	FLVCR2	1	-0.216
AQP6	2	-0.267	LINC01512	1	-0.215
ANKRD39	2	-0.266	ARHGAP23	1	-0.215
LINC00588	2	-0.266	GUCA1A	1	-0.215
FAIM2	2	-0.265	FNDC4	1	-0.215
H2AC14	2	-0.264	GCKR	1	-0.215

Table S30 continued from previous page

CERS3	2	-0.264	PRR35	1	-0.215
AS3MT	2	-0.264	LOC100652768	1	-0.215
C17orf75	2	-0.263	TAGLN	1	-0.215
NRG2	2	-0.263	ADAM12	1	-0.215
MAPK15	2	-0.263	AMER3	1	-0.215
CD9	2	-0.263	ULK2	1	-0.215
CYP2W1	2	-0.263	SCN5A	1	-0.215
STMN4	2	-0.262	MYL2	1	-0.215
OLFML3	2	-0.261	PRSS45P	1	-0.215
KCNN3	2	-0.261	PRSS50	1	-0.215
GPRC5C	2	-0.260	PPP1CC	1	-0.215
SSPOP	2	-0.260	MIR99B	1	-0.215
EEF1A2	2	-0.260	MIRLET7E	1	-0.215
MEAK7	2	-0.259	CPNE2	1	-0.215
TUBGCP3	2	-0.259	ALOX15B	1	-0.215
SEZ6	2	-0.259	SOBP	1	-0.215
AXL	2	-0.259	SYNPO2L	1	-0.215
PGAM2	2	-0.259	SHISA3	1	-0.215
LINC02800	2	-0.259	NKD1	1	-0.215
LACTB2-AS1	2	-0.259	EFNA2	1	-0.214
TRAM1	2	-0.259	CPLX3	1	-0.214
CACNA1A	2	-0.259	MAGOH	1	-0.214
AP2A2	2	-0.259	MAGOH-DT	1	-0.214
KLHL8	2	-0.258	KLC3	1	-0.214
NLRP14	2	-0.257	ANKRD53	1	-0.214
BCAR1	2	-0.257	MEDAG	1	-0.214
CNPY3	2	-0.257	GEMIN6	1	-0.214
LINC00205	2	-0.257	SOX2-OT	1	-0.214
CHRNB4	2	-0.257	TEPP	1	-0.214
MAPK8IP1	2	-0.257	FOXR1	1	-0.214
DAB1	2	-0.256	FAM131A	1	-0.213
SCGB1B2P	2	-0.256	NECAB2	1	-0.213
FAM76A	2	-0.256	RBM17	1	-0.213
ACTN3	2	-0.256	GPR137B	1	-0.213
ZDHHC24	2	-0.256	FXYD1	1	-0.213
CDC42EP1	2	-0.256	FXYD7	1	-0.213
NFATC4	2	-0.256	LGI4	1	-0.213
TLR5	2	-0.255	DNAAF8	1	-0.213
PLLP	2	-0.255	TRIM33	1	-0.213
COPZ2	2	-0.255	PYY	1	-0.213
CFAP45	2	-0.255	CELSR2	1	-0.213
CLDN4	2	-0.254	TAFA3	1	-0.213
NR2E1	2	-0.254	CRACDL	1	-0.213
TRIM10	2	-0.253	ALKBH7	1	-0.213

Table S30 continued from previous page

OSCP1	2	-0.253	INMT	1	-0.213
SVEP1	2	-0.253	INMT-MINDY4	1	-0.213
H2AC2P	2	-0.253	CACTIN-AS1	1	-0.212
ZNF890P	2	-0.252	TBXA2R	1	-0.212
THBS4	2	-0.252	RXRG	1	-0.212
SLC35E4	2	-0.252	CSF3	1	-0.212
NTF3	2	-0.252	PNPLA5	1	-0.212
KCNMA1	2	-0.252	LRRC37A6P	1	-0.212
EXD3	2	-0.252	MXI1	1	-0.212
CCDC184	2	-0.252	ACBD4	1	-0.212
ADD2	2	-0.251	METRNL	1	-0.212
EPS8L2	2	-0.251	DENND2C	1	-0.212
USP18	2	-0.251	PROKR1	1	-0.212
SLC45A4	2	-0.251	CCNH	1	-0.212
FLYWCH1	2	-0.251	C1QTNF8	1	-0.212
SPEN-AS1	2	-0.250	PNOC	1	-0.212
KBTBD13	2	-0.250	DNAH17	1	-0.212
RASL12	2	-0.250	DNAH17-AS1	1	-0.212
APOC1	2	-0.250	GPRIN2	1	-0.211
LMX1B	2	-0.250	ANKRD44	1	-0.211
DRAIC	2	-0.249	CNFN	1	-0.211
C8orf74	2	-0.249	LCA5L	1	-0.211
ICAM5	2	-0.249	SH3BGR	1	-0.211
CEP170B	2	-0.249	OSCAR	1	-0.211
LRRC66	2	-0.249	ISLR	1	-0.211
MAST4	2	-0.249	SLC12A5	1	-0.211
RARG	2	-0.249	MOCOS	1	-0.211
GRIN2D	2	-0.248	LOC440311	1	-0.211
POLD2	2	-0.248	GNAZ	1	-0.211
FAM234B	2	-0.248	RSPH14	1	-0.211
RHBDL3	2	-0.248	NOS1	1	-0.211
LAMA4	2	-0.247	KCNK9	1	-0.211
MGAT4D	2	-0.247	TFEB	1	-0.211
PDIA6	2	-0.247	BORCS7	1	-0.211
MARCKSL1	2	-0.247	INHBB	1	-0.211
KIF19	2	-0.247	IGSF9	1	-0.210
DPRXP4	2	-0.247	TMEM121	1	-0.210
RNF135	2	-0.247	MIGA2	1	-0.210
RIMS4	2	-0.247	WDR88	1	-0.210
ARHGEF10L	2	-0.246	NEUROG1	1	-0.210
ASZ1	2	-0.246	TRIM63	1	-0.210
SOX30	2	-0.245	C22orf42	1	-0.210
TSPAN11	2	-0.245	LY6K	1	-0.210
KDELR3	2	-0.245	LINC00602	1	-0.210

Table S30 continued from previous page

KCND3	2	-0.245	LINC00605	1	-0.210
ABT1	2	-0.244	KCNK10	1	-0.210
P2RX5	2	-0.244	CDC42EP5	1	-0.210
P2RX5-TAX1BP3	2	-0.244	PELI3	1	-0.210
MTERF1	2	-0.244	ANO2	1	-0.210
ERICH1	2	-0.244	RNF6	1	-0.210
C12orf56	2	-0.244	RP1	1	-0.210
CDH23	2	-0.243	SYCN	1	-0.210
SMTNL2	2	-0.243	CLYBL	1	-0.210
IL11	2	-0.243	GALNT9	1	-0.209
CLEC11A	2	-0.243	ZNF222	1	-0.209
LHX8	2	-0.243	ZNF284	1	-0.209
GPR84	2	-0.243	PLEKHO2	1	-0.209
ZNF385A	2	-0.243	RCL1	1	-0.209
GTF2IRD1	2	-0.243	LINC00673	1	-0.209
EPB41L1	2	-0.242	HRC	1	-0.209
LOC100286922	2	-0.242	SHC3	1	-0.209
UGT1A10	2	-0.242	L1TD1	1	-0.209
UGT1A3	2	-0.242	RIPPLY3	1	-0.209
UGT1A4	2	-0.242	HTRA1	1	-0.209
UGT1A5	2	-0.242	KDM4C	1	-0.209
UGT1A6	2	-0.242	ASB10	1	-0.208
UGT1A7	2	-0.242	ITM2C	1	-0.208
UGT1A8	2	-0.242	GJD4	1	-0.208
UGT1A9	2	-0.242	COPS7A	1	-0.208
CFAP46	2	-0.242	SPESP1	1	-0.208
TTLL11	2	-0.242	CD7	1	-0.208
IGDCC3	2	-0.242	DCHS2	1	-0.208
SERPINB5	2	-0.242	PPP1R14A	1	-0.208
PRLR	2	-0.242	GLI3	1	-0.208
ASIC4	2	-0.242	XYLT2	1	-0.208
DLK1	2	-0.242	TNS1	1	-0.208
C2	2	-0.242	PLEKHA2	1	-0.208
SYNPO	2	-0.242	DHX9	1	-0.208
TCF7L1	2	-0.241	ZNF707	1	-0.208
EGFR	2	-0.241	NOTUM	1	-0.208
CR1L	2	-0.241	LONRF2	1	-0.208
PIP5K1A	2	-0.241	SERINC5	1	-0.208
DNAH1	2	-0.240	HOXB8	1	-0.208
CRYGN	2	-0.240	CMC1	1	-0.207
TMEM200B	2	-0.240	LOC644189	1	-0.207
CEP104	2	-0.239	ZFP82	1	-0.207
DFFB	2	-0.239	TNFRSF18	1	-0.207
LRPAP1	2	-0.239	PIEZO2	1	-0.207

Table S30 continued from previous page

NRAP	2	-0.239	ALOX15	1	-0.207
SSMEM1	2	-0.239	BDKRB2	1	-0.207
TMEM209	2	-0.239	BTNL9	1	-0.207
ZNF764	2	-0.239	LYPD5	1	-0.207
ABHD10	2	-0.238	GRM7	1	-0.207
VAX1	2	-0.238	C4orf36	1	-0.207
KCNIP3	2	-0.238	TMEM119	1	-0.207
ANKRD33	2	-0.238	RTN4RL1	1	-0.207
ATP6V0A4	2	-0.238	AANAT	1	-0.207
TMEM213	2	-0.238	SMIM45	1	-0.207
MPPED1	2	-0.237	MIR3150B	1	-0.207
CTBP2	2	-0.237	NDUFAF6	1	-0.207
LINC00469	2	-0.237	MMP9	1	-0.207
NT5C1B	2	-0.237	AGPAT1	1	-0.207
NT5C1B-RDH14	2	-0.237	RNF5	1	-0.207
KRT19	2	-0.237	MIPEP	1	-0.207
GLIS3	2	-0.236	PCOTH	1	-0.207
C11orf91	2	-0.236	C1orf115	1	-0.206
NCOR1	2	-0.236	USP22	1	-0.206
PIGL	2	-0.236	RPRM	1	-0.206
ARK2C	2	-0.236	FXYD6	1	-0.206
DPY19L1	2	-0.236	FXYD6-FXYD2	1	-0.206
CACNA1I	2	-0.235	CMTM5	1	-0.206
WNT10B	2	-0.234	DLEU1	1	-0.206
FSD2	2	-0.234	MGAT3	1	-0.206
WHAMM	2	-0.234	RAD51C	1	-0.206
EIF5A2	2	-0.234	ETHE1	1	-0.206
TRABD2B	2	-0.234	F11-AS1	1	-0.206
GFAP	2	-0.233	LBX1	1	-0.206
MMP21	2	-0.233	LBX1-AS1	1	-0.206
TMEM61	2	-0.233	CLPB	1	-0.206
MIR100HG	2	-0.233	WDR33	1	-0.206
WNT7A	2	-0.233	NEBL	1	-0.206
PPP1R13L	2	-0.233	LRRC10B	1	-0.206
AXIN2	2	-0.233	MIR4488	1	-0.206
PRICKLE2	2	-0.233	H2BC14	1	-0.206
EPHB2	2	-0.232	TMEM63C	1	-0.206
ATG4C	2	-0.231	GRAMD2A	1	-0.205
PATL1	2	-0.231	LHB	1	-0.205
RNPEP	2	-0.231	LOC101059948	1	-0.205
LINC01115	2	-0.231	ZNF469	1	-0.205
ZNF273	2	-0.230	TEX2	1	-0.205
XPNPEP1	2	-0.230	OSR1	1	-0.205
GTF3A	2	-0.230	TRIM71	1	-0.205

Table S30 continued from previous page

LMBR1L	2	-0.229	OLFML2A	1	-0.205
ZNF648	2	-0.229	ARHGEF19	1	-0.205
MROH5	2	-0.229	TTL	1	-0.205
CAMK2B	2	-0.229	HSF2BP	1	-0.205
CIRBP-AS1	2	-0.229	RRP1B	1	-0.205
HHIPL1	2	-0.229	GARIN1B	1	-0.205
PRSS16	2	-0.229	HOXA10	1	-0.205
NOX4	2	-0.228	HOXA10-HOXA9	1	-0.205
SOX7	2	-0.228	FBXL2	1	-0.205
EXTL1	2	-0.228	ERN2	1	-0.205
HCN2	2	-0.228	TMTC1	1	-0.205
ECE2	2	-0.228	CFD	1	-0.205
LITAF	2	-0.228	GLB1L3	1	-0.205
INSYN1	2	-0.227	C5orf52	1	-0.204
TMEM220	2	-0.227	FAM66C	1	-0.204
TMEM220-AS1	2	-0.227	RPN1	1	-0.204
CACNB4	2	-0.227	PTH1R	1	-0.204
EFCAB10	2	-0.227	PET117	1	-0.204
ZNF575	2	-0.227	KLHL26	1	-0.204
SMARCD3	2	-0.227	LINC00511	1	-0.204
F7	2	-0.226	DRG2	1	-0.204
DCDC2	2	-0.226	MELTF-AS1	1	-0.204
LMNB2	2	-0.226	SPON2	1	-0.204
TIMM13	2	-0.226	IL17RD	1	-0.204
RFX2	2	-0.226	H2BC17	1	-0.204
TLL2	2	-0.225	H3C12	1	-0.204
TBX2	2	-0.225	MOGAT3	1	-0.204
SLC6A4	2	-0.225	ISG15	1	-0.204
LINC02875	2	-0.225	ABCB8	1	-0.204
TMEM229A	2	-0.225	ATG9B	1	-0.204
SPATC1	2	-0.225	TANC1	1	-0.204
KIAA1614	2	-0.224	WNT5A	1	-0.204
EMP2	2	-0.224	MIR129-1	1	-0.204
SRGAP3	2	-0.224	IFT70A	1	-0.204
ELFN1	2	-0.224	DCHS1	1	-0.203
ARHGAP29	2	-0.223	FNDC1	1	-0.203
HNF4A	2	-0.223	CDH1	1	-0.203
USP6	2	-0.222	PPP2R1A	1	-0.203
ZNF232	2	-0.222	EGR2	1	-0.203
SLC25A51P4	2	-0.222	LINC03048	1	-0.203
STAU2	2	-0.221	DDIT3	1	-0.203
SERTAD4	2	-0.221	MBD6	1	-0.203
SERTAD4-AS1	2	-0.221	GLP1R	1	-0.203
RARRES2	2	-0.221	DIP2C	1	-0.203

Table S30 continued from previous page

CKMT2	2	-0.221	DANCR	1	-0.203
RNU5E-1	2	-0.221	MIR4449	1	-0.203
CHP2	2	-0.220	SNORA26	1	-0.203
SHD	2	-0.220	VWA3A	1	-0.203
CDIPT	2	-0.219	LMX1A	1	-0.203
CDIPTOSP	2	-0.219	COL20A1	1	-0.203
COL9A1	2	-0.219	LOC100287944	1	-0.203
MOBP	2	-0.219	RFX4	1	-0.203
VPS51	2	-0.219	TGIF1	1	-0.203
PITPNM3	2	-0.218	LINC01107	1	-0.203
C1orf210	2	-0.218	FKBP7	1	-0.203
GAS2L1	2	-0.218	MIR548N	1	-0.203
CUX2	2	-0.218	PLEKHA3	1	-0.203
CDHR1	2	-0.218	SLC25A10	1	-0.203
DNAI1	2	-0.218	HLA-J	1	-0.203
FAM219A	2	-0.218	NIPAL3	1	-0.203
PLOD2	2	-0.217	STPG1	1	-0.203
SCN4B	2	-0.217	POU2F3	1	-0.202
MROH6	2	-0.216	SLC6A5	1	-0.202
SLC7A8	2	-0.216	CCDC42	1	-0.202
RNF180	2	-0.216	FIP1L1	1	-0.202
TRIL	2	-0.216	COQ5	1	-0.202
ALLC	2	-0.215	RNF10	1	-0.202
GNG8	2	-0.214	SNRPE	1	-0.202
ALPL	2	-0.214	MCF2	1	-0.202
EIF4E3	2	-0.214	TDH	1	-0.202
GPR27	2	-0.214	FAM174B	1	-0.202
ACTR3C	2	-0.214	KDF1	1	-0.202
LRRC61	2	-0.214	ECHS1	1	-0.202
CACYBP	2	-0.214	PAOX	1	-0.202
LINC00028	2	-0.214	TGFBI	1	-0.202
SYT2	2	-0.214	GABRG1	1	-0.202
KRT72	2	-0.214	CER1	1	-0.202
GRB10	2	-0.213	FIBCD1	1	-0.202
LRRN4CL	2	-0.213	SLC12A8	1	-0.202
TMEM232	2	-0.213	BCL7A	1	-0.201
LINC01060	2	-0.213	PDSS2	1	-0.201
SLC52A3	2	-0.212	SULT1C2	1	-0.201
KLB	2	-0.211	CUEDC1	1	-0.201
BCOR	2	-0.211	CFAP251	1	-0.201
RASA3	2	-0.210	CD72	1	-0.201
ZFR2	2	-0.210	CEACAM5	1	-0.201
BMP1	2	-0.210	SYCP1	1	-0.201
PRKN	2	-0.209	FMOD	1	-0.201

Table S30 continued from previous page

LMAN1L	2	-0.209	LPIN1	1	-0.201
INS	2	-0.208	PRDM16	1	-0.201
INS-IGF2	2	-0.208	SYNDIG1	1	-0.201
TLE2	2	-0.208	PLCL2	1	-0.201
PKP3	2	-0.208	CLN8	1	-0.201
TMEM128	2	-0.207	INSL3	1	-0.201
KIF17	2	-0.207	GATA6	1	-0.201
DPP10	2	-0.207	HECTD2-AS1	1	-0.201
SRMS	2	-0.207	COL16A1	1	-0.201
WNT10A	2	-0.206	CFAP53	1	-0.201
HLA-G	2	-0.206	ENTREP2	1	-0.201
HLA-H	2	-0.206	DIPK1C	1	-0.201
MLNR	2	-0.206	NKX2-8	1	-0.201
KCNJ5	2	-0.204	ZNF18	1	-0.201
KCNJ5-AS1	2	-0.204	ETNK2	1	-0.200
ANO1	1	-0.625	MESP2	1	-0.200
FOXN1	1	-0.624	PAX3	1	-0.200
MRPS25	1	-0.589	ITSN2	1	-0.200
ADGRG7	1	-0.567			

Table S31. Hypermethylated probes with gene correspondence in Sadikovic CHARGE episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Genes	Chr	$\Delta\beta$	genomic_features	CGI annot.
cg04863892	HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3	chr7	0.208	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg04053108	VWF	chr12	0.181	Body	CGIs
cg09549073	HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5	chr7	0.175	1to5kb, 1to5kb, 5UTRs, 5UTRs, Body, Body	CGIs
cg09319828	TTC24	chr1	0.147	Body	Inter
cg26229043	KIAA1217	chr10	0.136	Body	Inter
cg02646423	HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3	chr7	0.132	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg09674170	CLDN9	chr16	0.128	Promoters	shores
cg19543968	SUN2, DNAL4	chr22	0.123	Promoters, Promoters	shores
cg23204968	HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3	chr7	0.120	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg11321156	APP	chr21	0.117	Body	Inter

Table S31 continued from previous page

cg14179389	GFI1	chr1	0.116	Body	CGIs
cg07520269	DLG2, DLG2	chr11	0.107	5UTRs, Body	Inter
cg16923485	SLCO1A2	chr12	0.106	Body	Inter
cg22987448	MYO1F	chr19	0.105	Body	CGIs
cg04507071	C10orf90	chr10	0.104	Body	Inter
cg03352173	SLC9C1	chr3	0.102	Promoters	Inter
cg07593523	RAB3C	chr5	0.101	Body	Inter
cg25801742	LOC283194	chr11	0.100	Body	Inter
cg11630632	N4BP2L2, N4BP2L2	chr13	0.100	Promoters, 1to5kb	Inter
cg14696334	RCCD1, RCCD1	chr15	0.099	1to5kb, Body	CGIs
cg03934681	TRUB1	chr10	0.091	Promoters	shores
cg25506432	HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3	chr7	0.091	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg06130322	TMEM132C	chr12	0.090	Body	Inter
cg17713912	RELN	chr7	0.088	Body	Inter
cg03814093	TMEM131L	chr4	0.084	Body	Inter
cg22226821	FMN2	chr1	0.083	Body	Inter
cg14058329	HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3	chr7	0.081	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg14853974	ARHGEF15	chr17	0.081	Body	Inter
cg01709518	TAFA1, TAFA1, TAFA1	chr3	0.080	1to5kb, 5UTRs, Body	shelves
cg02000318	HOXA7	chr7	0.079	1to5kb	shores
cg23849826	THBS1	chr15	0.079	1to5kb	shores
cg27079740	GPM6A, GPM6A	chr4	0.078	5UTRs, Body	Inter
cg21929875	VSIG2, ESAM, VSIG2, ESAM, VSIG2, ESAM	chr11	0.078	5UTRs, 5UTRs, Body, Body, 3UTRs, 3UTRs	Inter
cg15355952	SLC1A3	chr5	0.076	Body	Inter
cg00350158	LRRC4C	chr11	0.076	Body	Inter
cg08070327	HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3	chr7	0.076	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg02973961	CDH11	chr16	0.075	Body	shores
cg20860188	KLHL29	chr2	0.074	Body	shelves
cg23930311	CCDC175	chr14	0.071	Promoters	shores
cg14013695	HOXA5, HOXA-AS3, HOXA5, HOXA-AS3, HOXA5, HOXA-AS3	chr7	0.071	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg10273821	OCA2	chr15	0.069	Body	shores
cg13178766	PTPRR	chr12	0.069	Body	Inter
cg08985968	PCCA	chr13	0.069	Body	Inter
cg12291059	CTAGE1	chr18	0.069	Promoters	Inter
cg09096031	RFTN2, RFTN2	chr2	0.068	5UTRs, Body	Inter

Table S31 continued from previous page

cg14212966	SLC1A7, SLC1A7	chr1	0.068	1to5kb, Body	CGIs
cg14424631	CYP19A1	chr15	0.067	Body	Inter
cg08335389	ACACB	chr12	0.066	Promoters	Inter
cg18161956	TBX3	chr12	0.066	Body	CGIs
cg08207154	KCNMB2	chr3	0.066	Body	Inter
cg21980394	PRKN	chr6	0.065	Body	Inter
cg12399536	TEAD1	chr11	0.065	Body	Inter
cg05465935	C5orf34	chr5	0.065	Promoters	shores
cg01663825	HECW1	chr7	0.063	Body	Inter
cg26130053	RIN2, RIN2	chr20	0.063	Body, 3UTRs	Inter
cg09361691	HTR2A, HTR2A	chr13	0.063	5UTRs, Body	Inter
cg13245264	OPRM1, OPRM1	chr6	0.063	5UTRs, Body	Inter
cg20815778	MIR30A	chr6	0.062	Promoters	Inter
cg26924747	RBM28	chr7	0.062	Promoters	shores
cg01383287	FBXO41	chr2	0.060	Body	shores
cg05226008	DLC1, DLC1	chr8	0.058	5UTRs, Body	Inter
cg21906947	KIF25	chr6	0.055	Body	Inter
cg10308749	HEY2	chr6	0.054	Body	CGIs
cg18664965	CHST13	chr3	0.053	Body	shores
cg07629125	PCDH17	chr13	0.048	Body	Inter

Table S32. Hypermethylated probes with gene correspondence in Weksberg CHARGE episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Gene	Chr	$\Delta\beta$	Genic annot.	CGI annot.
cg04053108	VWF	chr12	0.203	Body	CGIs
cg18274664	APP	chr21	0.150	Body	Inter
cg04392082	SORBS2, SORBS2, SORBS2	chr4	0.136	Promoters, 5UTRs, Body	Inter
cg26023912	HOXA-AS3, HOXA5, HOXA-AS3, HOXA5	chr7	0.124	1to5kb, 1to5kb, Body, Body	CGIs
cg16923485	SLCO1A2	chr12	0.111	Body	Inter
cg17615629	HLA-E	chr6	0.111	Body	shores
cg12166476	LMO7, LMO7	chr13	0.109	5UTRs, Body	Inter
cg23998119	ZDHHC22	chr14	0.108	Body	Inter
cg26033526	PSMB9, TAP1, PSMB9, TAP1	chr6	0.106	1to5kb, 1to5kb, Body, Body	shores
cg22469274	HOXA6, HOXA5, HOXA-AS3, HOXA6, HOXA5, HOXA-AS3, HOXA6, HOXA5, HOXA-AS3	chr7	0.105	Promoters, Promoters, Promoters, 1to5kb, 1to5kb, 1to5kb, Body, Body, Body	CGIs
cg14853974	ARHGEF15	chr17	0.105	Body	Inter

Table S32 continued from previous page

cg17904575	PPP2R5C	chr14	0.104	Body	Inter
cg16278496	LINC01550	chr14	0.103	Promoters	Inter
cg19816811	HOXA6, HOXA-AS3, HOXA6, HOXA-AS3	chr7	0.103	Promoters, Promoters, Body, Body	shores
cg25285743	LMO3, LMO3	chr12	0.102	Body, 3UTRs	Inter
cg23679819	FAM210A, RNMT, FAM210A, RNMT, FAM210A, RNMT	chr18	0.102	1to5kb, 1to5kb, 5UTRs, 5UTRs, Body, Body	shelves
cg14179389	GFI1	chr1	0.101	Body	CGIs

Table S33. Hypomethylated probes with gene correspondence in Sadikovic CHARGE episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Genes	Chr	$\Delta\beta$	Genic annot.	CGI annot.
cg08941355	HOTAIRM1, HOXA1, HOTAIRM1, HOXA1, HOTAIRM1, HOXA1	chr7	-0.163	1to5kb, 1to5kb, Body, Body, 3UTRs, 3UTRs	shores
cg06906435	LINC02914	chr14	-0.130	Promoters	Inter
cg06602723	HOXB7, HOXB8	chr17	-0.120	1to5kb, 1to5kb	shores
cg16787483	SLITRK5	chr13	-0.117	Body	shores
cg16915863	LINC02381	chr12	-0.111	Body	shelves
cg12806882	FMN2	chr1	-0.104	Body	shelves
cg24750308	NOX4, NOX4, NOX4	chr11	-0.099	Promoters, 1to5kb, Body	shores
cg11096515	COL4A2	chr13	-0.098	Body	Inter
cg17461600	DAB1	chr1	-0.096	Body	Inter
cg08401657	TFAP2A	chr6	-0.088	Body	shores
cg13569424	COLEC12	chr18	-0.084	Body	shores
cg15841205	NKX2-8	chr14	-0.082	1to5kb	shelves
cg11857140	KIRREL3	chr11	-0.081	Body	Inter
cg02211646	FOXP2, FOXP2	chr7	-0.078	5UTRs, Body	Inter
cg17686487	ISL1	chr5	-0.076	Body	shores
cg11758841	PARVA	chr11	-0.074	Body	Inter
cg00705992	HOXA11, HOXA11-AS, HOXA11, HOXA11-AS	chr7	-0.073	1to5kb, 1to5kb, Body, Body	shores
cg25037165	TEAD1	chr11	-0.071	Body	Inter
cg20706134	PCDH20, PCDH20	chr13	-0.071	Promoters, Body	Inter
cg08503681	GPR151	chr5	-0.070	Promoters	Inter
cg13775636	DIP2C	chr10	-0.069	Body	CGIs
cg06728970	KCNJ6	chr21	-0.067	Body	shelves
cg05167916	COL4A2	chr13	-0.065	Body	Inter
cg09600715	HOXA11	chr7	-0.063	1to5kb	shores

Table S33 continued from previous page

cg17104824	MIR10B, HOXD4, MIR10B, HOXD4	chr2	-0.062	Promoters, Promoters, 1to5kb, 1to5kb	CGIs
cg18774642	KLHL14	chr18	-0.057	Promoters	shores
cg06875131	AXIN2	chr17	-0.057	Body	shores
cg16171484	SGPP2	chr2	-0.056	Body	shores
cg05210501	HOXA10-HOXA9, HOXA10	chr7	-0.055	Body, Body	CGIs
cg04024827	FLJ12825	chr12	-0.054	Body	shores
cg22512830	NALF1	chr13	-0.051	Body	Inter
cg08115371	MIR708, TENM4, MIR708, TENM4	chr11	-0.048	1to5kb, 1to5kb, Body, Body	Inter
cg09404376	SGK1, SGK1, SGK1	chr6	-0.045	Promoters, 1to5kb, Body	shores

Table S34. Hypomethylated probes with gene correspondence in Weksberg CHARGE episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Genes	Chr	$\Delta\beta$	Genic annot.	CGI annot.
cg07318204	HHIP, HHIP-AS1, HHIP, HHIP-AS1	chr4	-0.187	Promoters, Promoters, Body, Body	CGIs
cg21090457	ROBO2	chr3	-0.157	Body	Inter
cg08941355	HOTAIRM1, HOXA1, HOTAIRM1, HOXA1, HOTAIRM1, HOXA1	chr7	-0.154	1to5kb, 1to5kb, Body, Body, 3UTRs, 3UTRs	shores
cg06906435	LINC02914	chr14	-0.150	Promoters	Inter
cg16787483	SLITRK5	chr13	-0.150	Body	shores
cg16915863	LINC02381	chr12	-0.150	Body	shelves
cg08657654	HOXA1, HOTAIRM1, HOXA1, HOTAIRM1	chr7	-0.147	1to5kb, 1to5kb, Body, Body	shelves
cg08959039	COL4A2	chr13	-0.137	Body	Inter
cg25556579	TBX5	chr12	-0.132	Body	Inter
cg17301248	NAA25, NAA25	chr12	-0.129	Body, 3UTRs	Inter
cg14191466	CDC42BPB	chr14	-0.127	Body	Inter
cg11857140	KIRREL3	chr11	-0.125	Body	Inter
cg18477569	BACH2	chr6	-0.125	Body	Inter
cg11341144	TRA2B, TRA2B	chr3	-0.124	Promoters, 1to5kb	shores
cg20071624	WIPI2, WIPI2	chr7	-0.122	Body, 3UTRs	Inter
cg25037165	TEAD1	chr11	-0.121	Body	Inter
cg22321572	AFDN, AFDN-DT, AFDN, AFDN-DT	chr6	-0.118	1to5kb, 1to5kb, Body, Body	shores
cg19981409	NOX4, NOX4, NOX4	chr11	-0.118	Promoters, 1to5kb, Body	shores
cg01746241	MYORG	chr9	-0.115	Body	CGIs

Table S34 continued from previous page

cg25356504	SARM1, SARM1	chr17	-0.113	1to5kb, Body	shores
cg16058600	NR4A2	chr2	-0.112	Body	shores
cg05575639	CHD7	chr8	-0.112	Body	shores
cg11758841	PARVA	chr11	-0.110	Body	Inter
cg15648345	MKS1	chr17	-0.109	Promoters	shores
cg17654050	NR4A2, NR4A2	chr2	-0.109	Promoters, Body	shores
cg22011526	C6orf89	chr6	-0.108	Body	shelves
cg22076474	BACH1, BACH1	chr21	-0.108	Promoters, Body	Inter
cg18546840	FOXP2, FOXP2	chr7	-0.108	5UTRs, Body	Inter
cg22679316	HOXC9, HOXC9	chr12	-0.107	Body, 3UTRs	shelves
cg09203312	GJB6, GJB6	chr13	-0.106	5UTRs, Body	shores
cg21758126	NR4A2	chr2	-0.105	Body	shores
cg20706134	PCDH20, PCDH20	chr13	-0.105	Promoters, Body	Inter
cg23333146	EIF4E3, EIF4E3	chr3	-0.104	Body, 3UTRs	Inter
cg12806882	FMN2	chr1	-0.104	Body	shelves
cg07659054	HOTAIRM1, HOXA1, HOTAIRM1, HOXA1, HOTAIRM1, HOXA1	chr7	-0.103	1to5kb, 1to5kb, Body, Body, 3UTRs, 3UTRs	CGIs
cg22222281	CD9	chr12	-0.103	Promoters	shores
cg00511167	ANAPC13, CEP63, ANAPC13, CEP63	chr3	-0.103	1to5kb, 1to5kb, Body, Body	shores
cg17461600	DAB1	chr1	-0.100	Body	Inter
cg05976283	NCALD	chr8	-0.100	1to5kb	shelves

Table S35. Hypermethylated probes with gene correspondence in Sadikovic Kabuki episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Genes	Chr	$\Delta\beta$	Genic annot.	CGI annot.
cg20744163	ZMIZ1, ZMIZ1	chr10	0.194	1to5kb, Body	shelves
cg18322589	TACC2	chr10	0.150	Body	Inter
cg21721340	SOX18	chr20	0.149	1to5kb	shores
cg21637392	RNF216, RNF216	chr7	0.145	1to5kb, Body	Inter
cg10213542	ADAMTS2	chr5	0.139	Body	Inter
cg05654765	LAMB2	chr3	0.139	Promoters	Inter
cg14371731	ZMIZ1, ZMIZ1	chr10	0.137	Promoters, Body	CGIs
cg09970481	CD79A	chr19	0.134	1to5kb	Inter
cg16423910	CD37, CD37, CD37	chr19	0.124	Promoters, Body, 3UTRs	CGIs
cg09230763	MAP3K6	chr1	0.123	Body	CGIs
cg27053299	CLYBL, CLYBL	chr13	0.123	Body, 3UTRs	CGIs
cg02661079	CDH22	chr20	0.123	Body	CGIs
cg16017144	RBFOX3	chr17	0.121	Body	CGIs

Table S35 continued from previous page

cg05463589	IL17C	chr16	0.115	Body	CGIs
cg10364115	LFNG	chr7	0.112	Body	shores
cg06826457	CDKN1B	chr12	0.107	1to5kb	shelves
cg24420089	PTDSS2	chr11	0.105	Body	shores
cg03461110	FOXK1	chr7	0.103	Body	Inter
cg23928920	ST8SIA2	chr15	0.098	Body	Inter
cg06341731	BCL11B	chr14	0.095	Body	CGIs
cg05003599	LRCH4, FBXO24, PCOLCE-AS1, LRCH4, FBXO24, PCOLCE-AS1	chr7	0.095	1to5kb, 1to5kb, 1to5kb, Body, Body, Body	shelves
cg15095327	IL17RE, CIDEC, IL17RE, CIDEC	chr3	0.094	5UTRs, 5UTRs, Body, Body	Inter
cg03890167	CHIT1	chr1	0.093	Promoters	CGIs
cg07816074	SH3TC1	chr4	0.091	Body	Inter
cg05620261	WTIP	chr19	0.090	Body	shelves
cg09990613	TCEA2, TCEA2	chr20	0.087	5UTRs, Body	shores
cg03168947	TRAK1, TRAK1	chr3	0.086	Promoters, Body	Inter
cg12280317	S100A11	chr1	0.085	Body	shores
cg03472000	CCDC172	chr10	0.083	Body	Inter
cg13091627	S100A4, S100A5, S100A4, S100A5, S100A4, S100A5	chr1	0.083	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	Inter
cg14916924	LMX1B	chr9	0.082	Body	CGIs
cg19423978	PIP5K1C	chr19	0.081	Body	shores
cg12689529	KIRREL3	chr11	0.080	Body	Inter
cg12845268	ARID5B	chr10	0.078	1to5kb	Inter
cg21212881	NFIB	chr9	0.076	Body	shores
cg09978353	NHERF1	chr17	0.073	Body	Inter
cg01209635	TBXA2R, CACTIN-AS1, TBXA2R, CACTIN-AS1, TBXA2R, CACTIN-AS1	chr19	0.071	Promoters, Promoters, 5UTRs, 5UTRs, Body, Body	CGIs
cg00601836	ESR1	chr6	0.071	Body	shores
cg22188918	CLDN15, CLDN15	chr7	0.067	Promoters, 1to5kb	Inter
cg17242664	EVC2, EVC, EVC2, EVC, EVC2, EVC, EVC2, EVC	chr4	0.065	Promoters, Promoters, 1to5kb, 1to5kb, 5UTRs, 5UTRs, Body, Body	shores
cg10658779	LRMDA	chr10	0.061	Body	shores
cg01644731	CTDSP2	chr12	0.059	Body	shelves
cg11190278	NIN	chr14	0.059	Body	Inter
cg06789445	MITF	chr3	0.059	Body	Inter
cg20956548	ZNF787	chr19	0.057	Body	shelves
cg03033176	EHBP1L1	chr11	0.049	Body	shores
cg11414202	NLK	chr17	0.046	Body	Inter

Table S35 continued from previous page

cg03888592	HDAC7	chr12	0.045	Body	shelves
cg16526732	FGD6	chr12	0.044	Body	Inter

Table S36. Hypermethylated probes with gene correspondence in Weksberg Kabuki episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Genes	Chr	$\Delta\beta$	Genic annot.	CGI annot.
cg01246520	RAI1	chr17	0.298	Body	Inter
cg08818610	RIPOR2	chr6	0.262	Body	CGIs
cg20744163	ZMIZ1, ZMIZ1	chr10	0.238	1to5kb, Body	shelves
cg05463589	IL17C	chr16	0.232	Body	CGIs
cg23061725	CASP8, CASP8	chr2	0.208	1to5kb, Body	Inter
cg11724970	HOXA-AS3, HOXA5, HOXA-AS3, HOXA5	chr7	0.179	1to5kb, 1to5kb, Body, Body	shores
cg18025886	MELTF	chr3	0.179	Body	shelves
cg17740434	TTN-AS1, MIR548N	chr2	0.179	Body, Body	Inter
cg27053299	CLYBL, CLYBL	chr13	0.175	Body, 3UTRs	CGIs
cg14371731	ZMIZ1, ZMIZ1	chr10	0.164	Promoters, Body	CGIs
cg07330481	ARL5C, ARL5C	chr17	0.164	5UTRs, Body	shores
cg04799958	KRT8, KRT18, KRT8, KRT18	chr12	0.156	Promoters, Promoters, Body, Body	shores
cg08352439	VOPP1	chr7	0.149	Body	shelves
cg27619353	LGALS1, LGALS1	chr22	0.145	5UTRs, Body	shores
cg10044575	CACNA1H	chr16	0.144	Body	shores
cg04063345	ESR1	chr6	0.138	Body	shores
cg19721065	RASAL1	chr12	0.135	Body	CGIs
cg24420089	PTDSS2	chr11	0.130	Body	shores
cg17393635	CD37, CD37	chr19	0.124	Promoters, Body	CGIs
cg05246522	KSR1	chr17	0.120	Promoters	Inter
cg07587796	ENPEP, ENPEP	chr4	0.120	5UTRs, Body	shores
cg00525681	SLITRK5	chr13	0.117	Body	shores
cg17220161	CHIT1	chr1	0.114	Promoters	CGIs
cg01832549	CAPZB	chr1	0.112	Body	Inter
cg12689529	KIRREL3	chr11	0.112	Body	Inter
cg04654167	ZDHHC14	chr6	0.111	Body	shelves
cg22811647	BRF1	chr14	0.111	Body	Inter
cg18738581	IL17RE, CIDEC, IL17RE, CIDEC	chr3	0.110	Promoters, Promoters, Body, Body	Inter
cg23989110	MAP3K6	chr1	0.106	Body	shores
cg20956548	ZNF787	chr19	0.105	Body	shelves
cg16620608	MOGAT3	chr7	0.104	Body	shelves
cg15372603	HOXA7	chr7	0.104	Promoters	shores

Table S36 continued from previous page

cg06683510	SEMA6D	chr15	0.103	Body	Inter
cg18335326	PPP1R18	chr6	0.103	Body	shores
cg27413508	COX4I2	chr20	0.103	Promoters	shores
cg09823859	SLITRK5	chr13	0.101	Body	shores
cg09892273	SLC6A15	chr12	0.101	Promoters	shores

Table S37. Hypomethylated probes with gene correspondence in Sadikovic Kabuki episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Gene	Chr	$\Delta\beta$	Genic annot.	CGI annot.
cg01178624	KCNK7, KCNK7	chr11	-0.236	Body, 3UTRs	CGIs
cg20225999	TNS1	chr2	-0.213	Body	shores
cg14845962	AGAP2, AGAP2-AS1, AGAP2, AGAP2-AS1	chr12	-0.212	Body, Body, 3UTRs, 3UTRs	CGIs
cg24869272	TSPAN4, TSPAN4	chr11	-0.155	5UTRs, Body	CGIs
cg18587137	TNFAIP2, TNFAIP2	chr14	-0.151	5UTRs, Body	CGIs
cg10146935	SAMD11, SAMD11	chr1	-0.151	1to5kb, Body	CGIs
cg16565409	SNORD4A, SNORD42A, SNORD4B, RAB34, RPL23A, SNORD4A, SNORD42A, SNORD4B, RAB34, RPL23A	chr17	-0.151	1to5kb, 1to5kb, 1to5kb, 1to5kb, 1to5kb, Body, Body, Body, Body, Body	shores
cg24940967	SLC6A20	chr3	-0.151	Body	shores
cg22220710	LAMA1	chr18	-0.142	Body	shores
cg01331992	RPS6	chr9	-0.142	Body	shores
cg03534375	SLMAP	chr3	-0.138	Promoters	shores
cg06768599	CIDEB, LTB4R, CIDEB, LTB4R	chr14	-0.137	1to5kb, 1to5kb, Body, Body	CGIs
cg16875104	GARS1-DT, GARS1, GARS1-DT, GARS1	chr7	-0.137	1to5kb, 1to5kb, Body, Body	shores
cg24937727	RGL3	chr19	-0.136	Body	CGIs
cg14660676	SQLE	chr8	-0.135	Body	shores
cg12176783	TCEA2, TCEA2, TCEA2	chr20	-0.132	Promoters, 1to5kb, Body	CGIs
cg11683663	CFD, ELANE, CFD, ELANE	chr19	-0.131	1to5kb, 1to5kb, Body, Body	CGIs
cg02892925	TOX	chr8	-0.125	1to5kb	shores
cg13127231	ARHGEF7	chr13	-0.123	Body	shores
cg06359132	RRP12	chr10	-0.120	Body	shores
cg12104246	C6orf62	chr6	-0.120	Body	shores
cg06902698	SCNM1, LYSMD1, TNFAIP8L2-SCNM1, SCNM1, LYSMD1, TNFAIP8L2-SCNM1	chr1	-0.117	Promoters, Promoters, Promoters, Body, Body, Body	Inter

Table S37 continued from previous page

cg20978937	PLD4, PLD4	chr14	-0.113	Body, 3UTRs	CGIs
cg08122386	H2BC7	chr6	-0.113	1to5kb	shores
cg14014731	RPS6	chr9	-0.109	Body	shores
cg02583484	CBX5, HNRNPA1, CBX5, HNRNPA1	chr12	-0.109	1to5kb, 1to5kb, Body, Body	shelves
cg12910806	CNST, TFB2M, CNST, TFB2M	chr1	-0.107	Promoters, Promoters, Body, Body	shores
cg14707053	CCND3, TAF8, CCND3, TAF8, CCND3, TAF8	chr6	-0.107	1to5kb, 1to5kb, 5UTRs, 5UTRs, Body, Body	shores
cg12086028	B3GALT4, VPS52, RPS18, B3GALT4, VPS52, RPS18	chr6	-0.106	1to5kb, 1to5kb, 1to5kb, Body, Body, Body	shores
cg07588439	MARS2	chr2	-0.104	Body	shores
cg02142461	LYAR, ZBTB49, LYAR, ZBTB49	chr4	-0.104	1to5kb, 1to5kb, Body, Body	shores
cg04124636	LMF1, SOX8, LMF1, SOX8	chr16	-0.103	1to5kb, 1to5kb, Body, Body	CGIs
cg19887750	CXXC1, CXXC1	chr18	-0.101	5UTRs, Body	shores
cg26125366	NOL4	chr18	-0.101	1to5kb	shores
cg07963563	HSPA12B	chr20	-0.100	Body	CGIs
cg05715492	ARPC1B	chr7	-0.097	Body	shores
cg13389502	HIC1	chr17	-0.096	Body	CGIs
cg21282663	HLX	chr1	-0.096	Body	shores
cg15153957	CCNP, AKT2, CCNP, AKT2, CCNP, AKT2	chr19	-0.096	1to5kb, 1to5kb, Body, Body, 3UTRs, 3UTRs	shelves
cg04731926	USF2, LSR, USF2, LSR	chr19	-0.095	1to5kb, 1to5kb, Body, Body	shores
cg14898243	SRGN	chr10	-0.093	Body	Inter
cg01103827	PSMG1	chr21	-0.091	Body	shores
cg12350863	GPR160	chr3	-0.090	Body	shores
cg11742202	SLC17A5	chr6	-0.087	Promoters	shores
cg16983110	PTPRK	chr6	-0.087	Body	Inter
cg04305808	ZKSCAN1	chr7	-0.086	Body	shores
cg05612654	GMDS-DT	chr6	-0.085	Body	Inter
cg26328687	MAML2	chr11	-0.084	Promoters	Inter
cg14270725	MXRA8	chr1	-0.084	Body	CGIs
cg21226442	FGFR1OP2, INTS13, FGFR1OP2, INTS13	chr12	-0.083	1to5kb, 1to5kb, Body, Body	shelves
cg08382220	VAC14	chr16	-0.083	Body	shores
cg15451248	SLC33A1	chr3	-0.083	Body	shores
cg19646165	PHF23, GABARAP, PHF23, GABARAP	chr17	-0.083	1to5kb, 1to5kb, Body, Body	shores
cg07642595	LAMP1	chr13	-0.082	Body	shores

Table S37 continued from previous page

cg24526103	IRS2	chr13	-0.081	Body	shores
cg01101380	TFAP2E, PSMB2, TFAP2E, PSMB2, TFAP2E, PSMB2	chr1	-0.080	5UTRs, 5UTRs, Body, Body, 3UTRs, 3UTRs	CGIs
cg00070052	C9orf43, POLE3, C9orf43, POLE3, C9orf43, POLE3	chr9	-0.080	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	shores
cg22904096	MYL1	chr2	-0.080	Promoters	Inter
cg03798743	CNTN5	chr11	-0.077	Body	Inter
cg16031250	TGFBI	chr5	-0.077	Body	CGIs
cg13912824	PTPRN2	chr7	-0.076	Body	shores
cg18582260	PARP4	chr13	-0.074	Body	shores
cg10586546	GPANK1, LY6G5B, CSNK2B, GPANK1, LY6G5B, CSNK2B, GPANK1, LY6G5B, CSNK2B, GPANK1, LY6G5B, CSNK2B	chr6	-0.072	Promoters, Promoters, Promoters, 1to5kb, 1to5kb, 1to5kb, 5UTRs, 5UTRs, 5UTRs, Body, Body, Body	shores
cg03437912	POMP	chr13	-0.072	Promoters	shores
cg10382814	NCSTN, COPA, NCSTN, COPA	chr1	-0.072	Promoters, Promoters, Body, Body	shores
cg04097078	SCARF2	chr22	-0.071	Body	CGIs
cg14271729	NAA30	chr14	-0.070	Body	shores
cg16000360	GUK1, GUK1, GUK1	chr1	-0.070	1to5kb, 5UTRs, Body	shores
cg10133500	TMEM115, TMEM115	chr3	-0.069	Promoters, Body	shores
cg23274660	SLC12A5, SLC12A5	chr20	-0.067	1to5kb, Body	shelves
cg24332685	SNORD117, DDX39B, ATP6V1G2-DDX39B, SNORD117, DDX39B, ATP6V1G2-DDX39B	chr6	-0.067	1to5kb, 1to5kb, 1to5kb, Body, Body, Body	shores
cg15392111	NDUFC1	chr4	-0.067	Body	shores
cg23565821	SYNGAP1, CUTA, SYNGAP1, CUTA	chr6	-0.066	1to5kb, 1to5kb, Body, Body	shores
cg07976603	RAVER1, ICAM3, RAVER1, ICAM3, RAVER1, ICAM3	chr19	-0.065	Promoters, Promoters, Body, Body, 3UTRs, 3UTRs	CGIs
cg13810727	RBM33, RBM33	chr7	-0.063	1to5kb, Body	shores
cg07457967	SCGB1B2P	chr19	-0.061	1to5kb	shores
cg06191495	DGKD	chr2	-0.060	Body	Inter
cg23404877	TGOLN2	chr2	-0.059	Body	shores
cg20474675	SOX6, C11orf58, SOX6, C11orf58	chr11	-0.058	1to5kb, 1to5kb, Body, Body	shores
cg21162977	RRAGA	chr9	-0.055	Body	shores
cg23035209	BTF3	chr5	-0.055	Body	shores

Table S37 continued from previous page

cg15961920	PPP1R10, MRPS18B, PPP1R10, MRPS18B	chr6	-0.054	1to5kb, 1to5kb, Body, Body	shores
cg09941876	FOXRED1, SRPRA, FOXRED1, SRPRA, FOXRED1, SRPRA	chr11	-0.050	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	shores

Table S38. Hypomethylated probes with gene correspondence in Weksberg Kabuki episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Gene	Chr	$\Delta\beta$	Genic annot.	CGI annot,
cg16370398	FLJ12825, HOXC4, FLJ12825, HOXC4	chr12	-0.214	1to5kb, 1to5kb, Body, Body	shores
cg13541527	SNORD52, SNHG32, SNORD52, SNHG32	chr6	-0.207	Promoters, Promoters, Body, Body	shores
cg20225999	TNS1	chr2	-0.202	Body	shores
cg23387569	AGAP2-AS1, AGAP2, AGAP2-AS1, AGAP2, AGAP2-AS1, AGAP2	chr12	-0.200	Promoters, Promoters, Body, Body, 3UTRs, 3UTRs	CGIs
cg12474798	ADO, ADO	chr10	-0.195	Body, 3UTRs	CGIs
cg04321618	HOXA4, HOXA3, HOXA4, HOXA3	chr7	-0.180	Promoters, Promoters, 1to5kb, 1to5kb	shores
cg03691722	LAMA1	chr18	-0.179	Body	CGIs
cg20698421	SLC1A4	chr2	-0.176	Body	shores
cg16565409	SNORD4A, SNORD42A, SNORD4B, RAB34, RPL23A, SNORD4A, SNORD42A, SNORD4B, RAB34, RPL23A	chr17	-0.169	1to5kb, 1to5kb, 1to5kb, 1to5kb, 1to5kb, Body, Body, Body, Body, Body	shores
cg16875104	GARS1-DT, GARS1, GARS1-DT, GARS1	chr7	-0.165	1to5kb, 1to5kb, Body, Body	shores
cg06768599	CIDEB, LTB4R, CIDEB, LTB4R	chr14	-0.163	1to5kb, 1to5kb, Body, Body	CGIs
cg24201793	MBOAT2	chr2	-0.159	Promoters	shores
cg07021906	SLC7A5	chr16	-0.156	Body	CGIs
cg26125366	NOL4	chr18	-0.155	1to5kb	shores
cg03534375	SLMAP	chr3	-0.151	Promoters	shores
cg14270725	MXRA8	chr1	-0.150	Body	CGIs
cg19338056	LOC389705, LOC389705	chr9	-0.146	1to5kb, Body	CGIs
cg21282663	HLX	chr1	-0.146	Body	shores
cg13127231	ARHGEF7	chr13	-0.144	Body	shores
cg03876340	SLC33A1	chr3	-0.143	Body	shores
cg22091609	SCARF2	chr22	-0.143	Body	CGIs
cg07529392	FIGNL2	chr12	-0.141	Body	CGIs

Table S38 continued from previous page

cg04286823	SH3RF3-AS1, SH3RF3, SH3RF3-AS1, SH3RF3	chr2	-0.137	Promoters, Promoters, Body, Body	shores
cg12104246	C6orf62	chr6	-0.130	Body	shores
cg20145009	PLD4	chr14	-0.130	Body	CGIs
cg11927233	MIR3912, NPM1, MIR3912, NPM1	chr5	-0.129	1to5kb, 1to5kb, Body, Body	shores
cg14880348	COL9A3	chr20	-0.126	1to5kb	shores
cg21226442	FGFR1OP2, INTS13, FGFR1OP2, INTS13	chr12	-0.126	1to5kb, 1to5kb, Body, Body	shelves
cg03798743	CNTN5	chr11	-0.122	Body	CGIs
cg15130459	SNORD72, RPL37, SNORD72, RPL37	chr5	-0.117	1to5kb, 1to5kb, Body, Body	shores
cg20857253	TCP10L, CFAP298	chr21	-0.117	Body, Body	CGIs
cg25197194	EFCC1, EFCC1	chr3	-0.116	Body, 3UTRs	CGIs
cg03785755	H2BC7	chr6	-0.114	1to5kb	shores
cg20786280	PURG, WRN, PURG, WRN	chr8	-0.113	1to5kb, 1to5kb, Body, Body	CGIs
cg26372517	TFAP2E, PSMB2, TFAP2E, PSMB2, TFAP2E, PSMB2	chr1	-0.112	5UTRs, 5UTRs, Body, Body, 3UTRs, 3UTRs	CGIs
cg13768953	DLG4	chr17	-0.112	Body	shelves
cg14014731	RPS6	chr9	-0.112	Body	shores
cg07684775	ZNF385A	chr12	-0.111	Body	CGIs
cg01997461	FARP1	chr13	-0.109	Body	shelves
cg22851880	TFAP2E, PSMB2, TFAP2E, PSMB2, TFAP2E, PSMB2	chr1	-0.108	Promoters, Promoters, Body, Body, 3UTRs, 3UTRs	shores
cg21577836	TNR	chr1	-0.107	Body	CGIs
cg08494221	PURG, WRN, PURG, WRN	chr8	-0.106	1to5kb, 1to5kb, Body, Body	CGIs
cg18582260	PARP4	chr13	-0.106	Body	shores
cg20673840	ESM1, ESM1	chr5	-0.104	5UTRs, Body	CGIs
cg15735240	PPP1R10, MRPS18B, PPP1R10, MRPS18B	chr6	-0.103	1to5kb, 1to5kb, Body, Body	shores
cg18117780	ARHGAP27P2	chr17	-0.103	Promoters	CGIs
cg21995347	SNX18	chr5	-0.103	Body	CGIs
cg15834355	HOXC4	chr12	-0.102	Body	shores
cg24527008	COL23A1	chr5	-0.101	Body	CGIs
cg14728380	CD7, SECTM1, CD7, SECTM1	chr17	-0.101	1to5kb, 1to5kb, Body, Body	shores
cg04451175	FGF20	chr8	-0.101	Promoters	shores

Table S39. Hypomethylated probes with gene correspondence in Sadikovic Sotos episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Gene	Chr	$\Delta\beta$	Genic annot.	CGI annot.
cg21105875	PCAT6, KDM5B	chr1	-0.650	1to5kb, 1to5kb	shores
cg07816556	H3C1, H4C1, H1-1, H3C1, H4C1, H1-1, H3C1, H4C1, H1-1	chr6	-0.559	1to5kb, 1to5kb, 1to5kb, Body, Body, Body, 3UTRs, 3UTRs, 3UTRs	shelves
cg15477144	FAM118A	chr22	-0.502	Promoters	shores
cg06154311	RBBP8NL	chr20	-0.494	Promoters	CGIs
cg25968569	KCTD12	chr13	-0.490	Promoters	shores
cg21778193	MIR200B	chr1	-0.485	1to5kb	CGIs
cg19817652	C17orf98	chr17	-0.484	Body	shores
cg21215576	FAM83A-AS1, FAM83A, FAM83A-AS1, FAM83A	chr8	-0.469	1to5kb, 1to5kb, Body, Body	shores
cg09089451	KRBA1	chr7	-0.467	1to5kb	CGIs
cg09044186	ZPR1, APOA5, ZPR1, APOA5	chr11	-0.457	1to5kb, 1to5kb, Body, Body	CGIs
cg00549574	PRPSAP1	chr17	-0.444	Promoters	shores
cg04159546	TP73, TP73	chr1	-0.441	1to5kb, Body	shelves
cg19389372	SDHAF1	chr19	-0.433	Promoters	CGIs
cg25240254	SSC5D	chr19	-0.416	Body	CGIs
cg05413199	TPPP3, ZDHHC1, TPPP3, ZDHHC1	chr16	-0.412	1to5kb, 1to5kb, Body, Body	shores
cg05897163	PPP1R1B, PPP1R1B	chr17	-0.412	Promoters, 1to5kb	shores
cg26218577	RPLP1	chr15	-0.411	Promoters	shores
cg19127638	TET1	chr10	-0.410	Body	shores
cg17774102	LINC01623	chr6	-0.409	Body	shelves
cg21475076	MIR200B, MIR200A, MIR429	chr1	-0.408	1to5kb, 1to5kb, 1to5kb	CGIs
cg08378788	MYO1C, MYO1C, MYO1C	chr17	-0.398	Promoters, 5UTRs, Body	shores
cg07786675	SFN	chr1	-0.376	Body	CGIs
cg23006204	DPM2	chr9	-0.369	Promoters	shores
cg00326131	RTKN, RTKN	chr2	-0.363	Promoters, Body	CGIs
cg20638675	TMEM72-AS1	chr10	-0.361	Body	CGIs
cg19185846	VAX2	chr2	-0.355	1to5kb	shelves
cg03329019	HLX	chr1	-0.354	1to5kb	shores
cg08801479	GPSM3, NOTCH4, GPSM3, NOTCH4	chr6	-0.353	1to5kb, 1to5kb, Body, Body	shores
cg24727290	DDR1, DDR1, DDR1	chr6	-0.352	Promoters, 1to5kb, Body	shores

Table S39 continued from previous page

cg03123541	TLCD4-RWDD3, RWDD3, TLCD4-RWDD3, RWDD3	chr1	-0.348	Promoters, Promoters, Body, Body	shores
cg18033770	DNLZ, CARD9, DNLZ, CARD9	chr9	-0.344	Promoters, Promoters, Body, Body	CGIs
cg12150256	HYAL2, HYAL2	chr3	-0.344	5UTRs, Body	CGIs
cg25921357	TMEM190	chr19	-0.343	Promoters	shores
cg19286437	LDB3, LDB3	chr10	-0.341	Promoters, Body	CGIs
cg12165772	DLX3	chr17	-0.340	Body	shores
cg14686645	ITGA2B, ITGA2B	chr17	-0.339	Body, 3UTRs	CGIs
cg03292743	C2orf81	chr2	-0.336	Body	CGIs
cg02477448	H1-0, GCAT	chr22	-0.335	1to5kb, 1to5kb	shores
cg01568522	GALNT6, SLC4A8, GALNT6, SLC4A8	chr12	-0.333	Promoters, Promoters, Body, Body	shores
cg19959917	CXCL12	chr10	-0.330	Body	shores
cg05092353	F10	chr13	-0.329	Promoters	CGIs
cg08511651	H3C1, H4C1, H1-1, H3C1, H4C1, H1-1	chr6	-0.324	1to5kb, 1to5kb, 1to5kb, Body, Body, Body	shelves
cg25547332	MEIKIN, ACSL6, MEIKIN, ACSL6	chr5	-0.322	Promoters, Promoters, Body, Body	CGIs
cg23897302	AMN	chr14	-0.320	Body	shores
cg02399371	TMEM105	chr17	-0.317	1to5kb	CGIs
cg01601573	GAB1	chr4	-0.313	1to5kb	shores
cg02024846	CDC14B	chr9	-0.311	Promoters	shores
cg06114987	FOXL1	chr16	-0.311	Promoters	shores
cg20068058	CREB3L1	chr11	-0.309	Body	shores
cg24642820	NUP210	chr3	-0.307	Promoters	shores
cg19893439	NABP2, RNF41, NABP2, RNF41	chr12	-0.306	Promoters, Promoters, 1to5kb, 1to5kb	shores
cg10347828	RASIP1	chr19	-0.304	Promoters	shores
cg07870378	UBALD1	chr16	-0.302	Promoters	CGIs
cg17220161	CHIT1	chr1	-0.299	Promoters	CGIs
cg22012583	CERS2	chr1	-0.297	Promoters	CGIs
cg20941258	TDGF1, LRRC2, TDGF1, LRRC2	chr3	-0.295	Promoters, Promoters, Body, Body	CGIs
cg08289350	CYGB, PRCD, CYGB, PRCD	chr17	-0.289	Body, Body, 3UTRs, 3UTRs	CGIs
cg03503642	COL12A1, COL12A1	chr6	-0.286	Promoters, Body	shores
cg25924688	BBS4, HIGD2B, BBS4, HIGD2B	chr15	-0.285	Promoters, Promoters, Body, Body	shores
cg08217285	ANKRD13B, GIT1, ANKRD13B, GIT1	chr17	-0.282	1to5kb, 1to5kb, Body, Body	shores
cg05998295	XPO7, DOK2	chr8	-0.279	1to5kb, 1to5kb	shores
cg14264194	RD3	chr1	-0.279	Body	CGIs

Table S39 continued from previous page

cg24210717	GPD1, GPD1	chr12	-0.277	5UTRs, Body	CGIs
cg02197228	ASIC4, ASIC4	chr2	-0.275	Body, 3UTRs	shores
cg11541486	BTBD17	chr17	-0.267	Promoters	shores
cg11857646	KCNH2	chr7	-0.264	Body	CGIs
cg24413781	CERS4	chr19	-0.255	Promoters	CGIs
cg24083746	CKM, CKM	chr19	-0.253	5UTRs, Body	shores
cg06462065	TMEM190, IL11	chr19	-0.252	1to5kb, 1to5kb	shelves
cg02136452	ZNF423	chr16	-0.249	Body	CGIs
cg13734535	NCS1	chr9	-0.247	Body	CGIs
cg26107597	COCH	chr14	-0.242	Promoters	shores
cg19925872	ZNF710	chr15	-0.239	1to5kb	shores

Table S40. Hypomethylated probes with gene correspondence in Weksberg Sotos episignature, and respective chromosome position, $\Delta\beta$ value, and annotation relative to genes and CpG islands. In blue colour, the top 5 probes are highlighted.

Probes	Gene	Chr	$\Delta\beta$	Genic annot.	CGI annot,
cg21105875	PCAT6, KDM5B	chr1	-0.668	1to5kb, 1to5kb	shores
cg07600533	KLHDC7B	chr22	-0.661	Promoters	CGIs
cg03133378	TMEM225B, TMEM225B	chr7	-0.650	5UTRs, Body	CGIs
cg25807487	MRGPRF, MRGPRF	chr11	-0.642	1to5kb, Body	shelves
cg09684846	TVP23A	chr16	-0.636	Promoters	CGIs
cg12363903	PITX1-AS1	chr5	-0.634	Body	shores
cg02956248	PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547	chr6	-0.632	Promoters, Promoters, Promoters, Promoters, 1to5kb, 1to5kb, 1to5kb, 1to5kb, Body, Body, Body, Body	shores
cg01996116	HIGD1A, ACKR2, HIGD1A, ACKR2, HIGD1A, ACKR2	chr3	-0.625	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	shores
cg24365896	ANO1	chr11	-0.625	Body	Inter
cg13568106	FOXN1	chr17	-0.624	Body	Inter
cg26863172	ANKRD65, TMEM88B, ANKRD65, TMEM88B	chr1	-0.620	1to5kb, 1to5kb, Body, Body	CGIs
cg00701051	NBPF20, NBPF10, NBPF20, NBPF10	chr1	-0.616	Promoters, Promoters, Body, Body	CGIs
cg06704455	ELFN2	chr22	-0.610	Body	shores
cg23689428	LINC01101	chr2	-0.608	Promoters	Inter

Table S40 continued from previous page

cg23359665	PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547	chr6	-0.601	Promoters, Promoters, Promoters, Promoters, 1to5kb, 1to5kb, 1to5kb, 1to5kb, Body, Body, Body, Body	shores
cg08199758	CLDN9	chr16	-0.597	Promoters	shores
cg23731272	SMAD3	chr15	-0.593	1to5kb	shores
cg25844921	MRPS25	chr3	-0.589	Promoters	shores
cg00931644	KCTD12	chr13	-0.585	Promoters	shores
cg12846837	SPIDR	chr8	-0.583	Promoters	CGIs
cg05820861	C1QTNF9, SPATA13, C1QTNF9, SPATA13	chr13	-0.580	1to5kb, 1to5kb, Body, Body	CGIs
cg16717549	AIRE, AIRE	chr21	-0.578	Promoters, 1to5kb	CGIs
cg06833564	UBALD1	chr16	-0.576	Promoters	CGIs
cg07816556	H3C1, H4C1, H1-1, H3C1, H4C1, H1-1, H3C1, H4C1, H1-1	chr6	-0.576	1to5kb, 1to5kb, 1to5kb, Body, Body, Body, 3UTRs, 3UTRs, 3UTRs	shelves
cg25449950	SOD3, SOD3, SOD3	chr4	-0.575	1to5kb, 5UTRs, Body	shelves
cg18689454	AIRE, AIRE	chr21	-0.574	Promoters, 1to5kb	CGIs
cg11291003	PITX1-AS1	chr5	-0.571	Body	shores
cg00186462	ADGRG7	chr3	-0.567	Body	Inter
cg22344745	ZNF678	chr1	-0.565	1to5kb	CGIs
cg03274456	AHNAK	chr11	-0.564	Body	Inter
cg02486855	SMAD3	chr15	-0.561	1to5kb	shores
cg13504434	CABP1, CABP1	chr12	-0.556	5UTRs, Body	Inter
cg08482837	LOC284950	chr2	-0.556	1to5kb	CGIs
cg25968569	KCTD12	chr13	-0.554	Promoters	shores
cg14220678	CCDC17	chr1	-0.553	Body	CGIs
cg19092981	TBX1	chr22	-0.553	Body	CGIs
cg16615454	SMARCD1, ASIC1, SMARCD1, ASIC1	chr12	-0.552	1to5kb, 1to5kb, Body, Body	CGIs
cg02781088	RGPD8, RGPD8	chr2	-0.551	Promoters, 1to5kb	shores
cg12290671	TMEM225B	chr7	-0.551	Promoters	CGIs
cg13502125	NBPF8	chr1	-0.546	Body	CGIs
cg10690919	WFIKKN2, WFIKKN2	chr17	-0.546	Promoters, Body	Inter
cg15477144	FAM118A	chr22	-0.546	Promoters	shores
cg11647108	ANXA11, LINC00857, ANXA11, LINC00857	chr10	-0.545	1to5kb, 1to5kb, Body, Body	CGIs
cg06686742	CERS4	chr19	-0.541	Promoters	shores
cg03962527	ADD3, ADD3-AS1, ADD3, ADD3-AS1, ADD3, ADD3-AS1	chr10	-0.541	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	shores

Table S40 continued from previous page

cg22249529	KLHDC7B	chr22	-0.540	Promoters	CGIs
cg19817652	C17orf98	chr17	-0.538	Body	shores
cg04044120	DNAJB13	chr11	-0.537	Body	Inter
cg26643967	SDHAF1	chr19	-0.536	Promoters	CGIs
cg10999992	H1-0, GCAT, H1-0, GCAT	chr22	-0.530	Promoters, Promoters, 1to5kb, 1to5kb	shores
cg26070540	COX4I2	chr20	-0.528	Promoters	CGIs
cg12131208	CHCT1	chr17	-0.528	Promoters	shores
cg18424208	CDX1	chr5	-0.527	Promoters	CGIs
cg21778193	MIR200B	chr1	-0.527	1to5kb	CGIs
cg08043592	GALP	chr19	-0.525	Promoters	Inter
cg04358214	PHAF1	chr16	-0.524	Promoters	CGIs
cg07187855	DDR1, MIR4640, DDR1, MIR4640	chr6	-0.523	1to5kb, 1to5kb, Body, Body	shores
cg10502244	COL2A1	chr12	-0.523	1to5kb	CGIs
cg09614479	ANXA11, LINC00857, ANXA11, LINC00857	chr10	-0.521	1to5kb, 1to5kb, Body, Body	shores
cg04157161	GUCY2D	chr17	-0.519	Body	CGIs
cg06154311	RBBP8NL	chr20	-0.517	Promoters	Inter
cg12233487	TMC2	chr20	-0.516	Promoters	Inter
cg04561804	SEC62, FHL1P1, SEC62, FHL1P1	chr3	-0.513	Promoters, Promoters, Body, Body	shores
cg26600753	TRIM7, TRIM7	chr5	-0.512	1to5kb, Body	CGIs
cg09593860	ZBTB16	chr11	-0.510	1to5kb	shores
cg13614409	TEX26, TEX26-AS1	chr13	-0.510	Promoters, Promoters	Inter
cg00990385	CFAP73	chr12	-0.510	Body	shelves
cg09044186	ZPR1, APOA5, ZPR1, APOA5	chr11	-0.509	1to5kb, 1to5kb, Body, Body	CGIs
cg25885280	SHANK2	chr11	-0.508	Body	Inter
cg17900689	ZMYND15	chr17	-0.507	Body	Inter
cg20137746	FAM238C	chr10	-0.505	1to5kb	Inter
cg17116120	COX4I2	chr20	-0.505	Promoters	CGIs
cg09674170	CLDN9	chr16	-0.504	Promoters	shores
cg00970752	FHAD1	chr1	-0.503	Body	Inter
cg11117637	CDX1, CDX1	chr5	-0.502	5UTRs, Body	CGIs
cg06842409	FOXS1, DUSP15, FOXS1, DUSP15	chr20	-0.502	1to5kb, 1to5kb, Body, Body	CGIs
cg05516295	PITX1-AS1	chr5	-0.502	Body	shores
cg10365562	TDGF1, LRRC2, TDGF1, LRRC2, TDGF1, LRRC2	chr3	-0.502	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	shelves
cg19196401	METTL24, DDO	chr6	-0.501	Body, Body	CGIs
cg02632490	MIR346, GRID1	chr10	-0.501	Body, Body	shores
cg25257677	SLC39A14, SLC39A14	chr8	-0.499	Promoters, 1to5kb	shores

Table S40 continued from previous page

cg12642717	CIDEA	chr18	-0.497	Body	shores
cg20737388	DNAJB13	chr11	-0.496	Body	Inter
cg04220541	SPIDR	chr8	-0.496	Promoters	CGIs
cg26695758	TNXB	chr6	-0.496	Body	CGIs
cg13939055	LINC02774	chr1	-0.495	Body	CGIs
cg00443981	CHCT1	chr17	-0.493	Promoters	shores
cg27500148	RPL23, SNORA21	chr17	-0.493	1to5kb, 1to5kb	shores
cg17976205	RBBP8NL	chr20	-0.493	Promoters	Inter
cg11372436	SOD3, SOD3	chr4	-0.491	Promoters, 1to5kb	Inter
cg07537443	NGFR	chr17	-0.491	Promoters	shores
cg00549574	PRPSAP1	chr17	-0.491	Promoters	shores
cg05308495	CACNG6	chr19	-0.489	Body	CGIs
cg06693983	TMEM190	chr19	-0.489	Body	CGIs
cg08379987	TEX26-AS1, TEX26, TEX26-AS1, TEX26	chr13	-0.488	Promoters, Promoters, Body, Body	Inter
cg09692695	RTKN, RTKN	chr2	-0.487	Promoters, Body	CGIs
cg10635895	CACNA1C	chr12	-0.486	Body	CGIs
cg10555383	LYNX1, LYNX1	chr8	-0.483	Promoters, 1to5kb	shores
cg18568930	IGLON5	chr19	-0.482	Body	shores
cg17724412	LINC01623	chr6	-0.482	Body	shelves
cg21406402	LINC01101, LINC01101	chr2	-0.482	5UTRs, Body	Inter
cg00643333	FAM118A	chr22	-0.481	Promoters	shores
cg00823843	LINC02774	chr1	-0.480	Body	CGIs
cg00944421	ESRP2, ESRP2	chr16	-0.480	1to5kb, Body	CGIs
cg24550212	OCSTAMP	chr20	-0.480	Body	shores
cg19940077	GPR25	chr1	-0.480	Body	CGIs
cg18409649	ZNF789, ATP5MF-PTCD1, ATP5MF, ZNF789, ATP5MF-PTCD1, ATP5MF	chr7	-0.479	Promoters, Promoters, Promoters, 1to5kb, 1to5kb, 1to5kb	shelves
cg18121224	NSD1, NSD1	chr5	-0.479	Promoters, 1to5kb	CGIs
cg00270878	SLC22A11	chr11	-0.476	Body	shores
cg21411366	SOD3, SOD3, SOD3	chr4	-0.476	1to5kb, 5UTRs, Body	shelves
cg19402939	CHRD	chr3	-0.476	Body	CGIs
cg00524708	PRSS36	chr16	-0.475	Body	shores
cg27362048	ZPR1, APOA5, ZPR1, APOA5	chr11	-0.475	1to5kb, 1to5kb, Body, Body	CGIs
cg13713218	SUGT1P4-STRA6LP, SUGT1P4-STRA6LP-CCDC180, SUGT1P4-STRA6LP, SUGT1P4-STRA6LP-CCDC180	chr9	-0.474	Promoters, Promoters, Body, Body	shores
cg25066665	RPRD2	chr1	-0.474	1to5kb	shores
cg14708990	MYO7A	chr11	-0.473	Promoters	Inter
cg09089451	KRBA1	chr7	-0.473	1to5kb	CGIs

Table S40 continued from previous page

cg11450541	ANKRD20A19P	chr13	-0.473	Body	Inter
cg23505823	HCAR1	chr12	-0.473	Promoters	Inter
cg19772114	LINC01623	chr6	-0.473	Body	shelves
cg02254407	PLEKHB1, PLEKHB1	chr11	-0.472	Promoters, 1to5kb	Inter
cg17453840	CPEB1, CPEB1-AS1, CPEB1, CPEB1-AS1, CPEB1, CPEB1-AS1	chr15	-0.471	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	CGIs
cg08772588	SSC5D	chr19	-0.471	Body	shores
cg04539705	PRSS56	chr2	-0.471	Promoters	shores
cg04942251	RCOR2	chr11	-0.471	1to5kb	shores
cg10365886	TNXB	chr6	-0.470	Body	CGIs
cg24524285	NRXN2	chr11	-0.470	Body	CGIs
cg21543589	COX4I2	chr20	-0.470	Promoters	CGIs
cg20084219	C17orf98	chr17	-0.468	Body	CGIs
cg12133451	ZNF678	chr1	-0.467	1to5kb	CGIs
cg11923631	AIRE, AIRE, AIRE	chr21	-0.466	1to5kb, 5UTRs, Body	CGIs
cg06303238	SALL4, SALL4	chr20	-0.466	5UTRs, Body	CGIs
cg03063309	SOD3, SOD3	chr4	-0.466	Promoters, 1to5kb	Inter
cg05133205	PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547	chr6	-0.466	Promoters, Promoters, Promoters, Promoters, 1to5kb, 1to5kb, 1to5kb, 1to5kb, 5UTRs, 5UTRs, 5UTRs, 5UTRs, Body, Body, Body, Body	shores
cg05630111	CERS2	chr1	-0.466	Promoters	shores
cg20625967	PLA2G4E	chr15	-0.465	Body	Inter
cg10854819	LINC03034	chr15	-0.465	Body	Inter
cg27251412	AIRE, AIRE, AIRE	chr21	-0.464	1to5kb, 5UTRs, Body	CGIs
cg23280506	HS3ST3B1	chr17	-0.464	1to5kb	CGIs
cg17261676	H3C1	chr6	-0.463	1to5kb	shelves
cg20701556	TLCD4-RWDD3, RWDD3, TLCD4-RWDD3, RWDD3	chr1	-0.463	Promoters, Promoters, Body, Body	shores
cg04159546	TP73, TP73	chr1	-0.463	1to5kb, Body	shelves
cg26010218	FRAS1	chr4	-0.462	1to5kb	shores
cg01235820	SCRT2, SRXN1, SCRT2, SRXN1	chr20	-0.462	Body, Body, 3UTRs, 3UTRs	CGIs
cg19049754	LINC00857, ANXA11, LINC00857, ANXA11	chr10	-0.461	Promoters, Promoters, 1to5kb, 1to5kb	CGIs
cg06825512	APCDD1	chr18	-0.461	Promoters	shores
cg11985480	LBX2-AS1, LBX2, LBX2-AS1, LBX2, LBX2-AS1, LBX2	chr2	-0.461	Promoters, Promoters, 1to5kb, 1to5kb, Body, Body	shores

Table S40 continued from previous page

cg01205267	LRCH1	chr13	-0.460	Promoters	shores
cg03541791	LINC01623	chr6	-0.460	Body	shelves
cg24024505	LINC01623	chr6	-0.460	Body	shelves
cg00003625	WFIKKN2, WFIKKN2	chr17	-0.460	Promoters, Body	Inter
cg01101782	PCAT6, KDM5B	chr1	-0.459	1to5kb, 1to5kb	shores
cg09510752	PITX1-AS1	chr5	-0.459	Body	shores
cg02497785	ABCA13, ABCA13	chr7	-0.459	Promoters, Body	CGIs
cg08247921	MANBA	chr4	-0.458	Promoters	shores
cg04638150	AHNAK	chr11	-0.458	Body	Inter
cg08378788	MYO1C, MYO1C, MYO1C	chr17	-0.458	Promoters, 5UTRs, Body	shores
cg12799314	ADARB2	chr10	-0.456	Body	CGIs
cg18206867	B3GNT2	chr2	-0.456	Promoters	CGIs
cg23327896	DNAJB13	chr11	-0.456	Body	Inter
cg13591848	SFXN3, PDZD7, SFXN3, PDZD7	chr10	-0.456	1to5kb, 1to5kb, Body, Body	CGIs
cg09368716	ANKRD65, TMEM88B, ANKRD65, TMEM88B	chr1	-0.456	1to5kb, 1to5kb, Body, Body	CGIs
cg13105599	SMAD3	chr15	-0.454	1to5kb	shores
cg26218577	RPLP1	chr15	-0.454	Promoters	shores
cg14372753	SLC25A18	chr22	-0.453	Body	CGIs
cg01721149	MIR9-1, MIR9-1HG, MIR9-1, MIR9-1HG	chr1	-0.452	1to5kb, 1to5kb, Body, Body	shelves
cg14802355	MANBA	chr4	-0.452	Promoters	shores
cg13934406	PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547	chr6	-0.452	Promoters, Promoters, Promoters, Promoters, 1to5kb, 1to5kb, 1to5kb, 1to5kb, Body, Body, Body, Body	shores
cg09965419	DDR1, MIR4640, DDR1, MIR4640	chr6	-0.452	1to5kb, 1to5kb, Body, Body	shores
cg12408990	RASGEF1C	chr5	-0.451	Body	CGIs
cg19760250	C17orf98	chr17	-0.451	Body	CGIs
cg22952142	CLN6, CLN6	chr15	-0.451	5UTRs, Body	Inter
cg16216407	SOD3, SOD3, SOD3	chr4	-0.450	1to5kb, 5UTRs, Body	shelves
cg11905611	PLBD1	chr12	-0.450	Promoters	CGIs
cg09117688	TPTE2P5, SUGT1P3	chr13	-0.450	Promoters, Promoters	CGIs
cg18192325	GPR45	chr2	-0.450	1to5kb	shores
cg20286956	MACF1, MACF1	chr1	-0.449	1to5kb, Body	CGIs
cg03102494	NBPF8	chr1	-0.449	Body	Inter
cg16915863	LINC02381	chr12	-0.449	Body	shelves
cg09428340	PACRG	chr6	-0.449	Body	CGIs
cg11557492	SSC5D	chr19	-0.449	Body	CGIs

Table S40 continued from previous page

cg05226558	TDGF1, LRRC2, TDGF1, LRRC2	chr3	-0.449	1to5kb, 1to5kb, Body, Body	shores
cg09510531	AIRE, AIRE, AIRE	chr21	-0.448	1to5kb, 5UTRs, Body	CGIs
cg17571919	LINC01623	chr6	-0.448	Body	shelves
cg22861548	LYNX1, LYNX1	chr8	-0.448	Promoters, 1to5kb	shores
cg11617964	PPT2, PPT2-EGFL8, PRRT1, PPT2, PPT2-EGFL8, PRRT1	chr6	-0.448	1to5kb, 1to5kb, 1to5kb, Body, Body, Body	CGIs
cg14759977	TPTE2P5, SUGT1P3	chr13	-0.447	Promoters, Promoters	CGIs
cg21182196	CLMN	chr14	-0.447	Promoters	shores
cg03188580	LINC01623	chr6	-0.447	Body	shelves
cg19389372	SDHAF1	chr19	-0.447	Promoters	CGIs
cg08123207	PPT2, PPT2-EGFL8, PRRT1, PPT2, PPT2-EGFL8, PRRT1	chr6	-0.447	1to5kb, 1to5kb, 1to5kb, Body, Body, Body	CGIs
cg15584071	SDHAF1	chr19	-0.447	1to5kb	CGIs
cg14459158	BARX1	chr9	-0.446	1to5kb	shores
cg19579217	TMEM14C	chr6	-0.446	1to5kb	shelves
cg08877188	SLC51A, SLC51A	chr3	-0.446	5UTRs, Body	Inter
cg19328828	HCAR1	chr12	-0.445	Promoters	Inter
cg04579398	SDHAF1	chr19	-0.445	Promoters	CGIs
cg00611789	CMYA5	chr5	-0.445	Promoters	Inter
cg25460807	DMTN	chr8	-0.445	1to5kb	shelves
cg07266144	PDLIM4	chr5	-0.445	Body	CGIs
cg17378342	CDX1	chr5	-0.444	Promoters	CGIs
cg19962075	MAPK4, MAPK4	chr18	-0.444	Body, 3UTRs	CGIs
cg02247178	TEX26, TEX26-AS1, TEX26, TEX26-AS1	chr13	-0.444	Promoters, Promoters, Body, Body	Inter
cg01206378	TLCD4-RWDD3, RWDD3, TLCD4-RWDD3, RWDD3	chr1	-0.444	Promoters, Promoters, Body, Body	shores
cg27438152	LINC03034	chr15	-0.443	Body	Inter
cg10244666	WT1	chr11	-0.443	Body	Inter
cg12914966	LINC01623	chr6	-0.443	Body	shores
cg06025456	PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547, PPT2, PPT2-EGFL8, PRRT1, LOC100507547	chr6	-0.442	Promoters, Promoters, Promoters, Promoters, 1to5kb, 1to5kb, 1to5kb, 1to5kb, Body, Body, Body, Body	shores
cg24010336	FBXO17	chr19	-0.441	Promoters	shores
cg25483741	LYNX1, LYNX1	chr8	-0.441	Promoters, 1to5kb	shores
cg08979191	SEPTIN8	chr5	-0.441	Promoters	shores
cg11669516	IGFBP2	chr2	-0.441	Body	shelves
cg09655403	CMYA5	chr5	-0.441	Promoters	Inter

Table S40 continued from previous page

cg19978312	RASGEF1C	chr5	-0.441	Body	CGIs
cg21475076	MIR200B, MIR200A, MIR429	chr1	-0.440	1to5kb, 1to5kb, 1to5kb	CGIs
cg00083206	METTL24, DDO	chr6	-0.439	Body, Body	CGIs
cg07429623	KLHDC8A, KLHDC8A, KLHDC8A	chr1	-0.439	Promoters, 5UTRs, Body	Inter
cg17330303	SFN, SFN	chr1	-0.438	5UTRs, Body	shores
cg16306978	APOB	chr2	-0.437	Promoters	CGIs
cg03679504	LRRC2, TDGF1, LRRC2, TDGF1	chr3	-0.437	5UTRs, 5UTRs, Body, Body	shelves
cg23180489	LYNX1, LYNX1	chr8	-0.437	Promoters, 1to5kb	shores
cg13392078	RECK	chr9	-0.437	Promoters	shores
cg17662083	DMKN, DMKN	chr19	-0.437	Promoters, 1to5kb	Inter
cg24808901	TLCD4-RWDD3, RWDD3, TLCD4-RWDD3, RWDD3	chr1	-0.437	Promoters, Promoters, Body, Body	shores
cg08089567	AIRE, AIRE	chr21	-0.436	1to5kb, Body	CGIs
cg07853425	PITX1-AS1	chr5	-0.435	Body	Inter
cg12231340	KRT81, KRT86	chr12	-0.435	Body, Body	CGIs
cg18361418	GPX5	chr6	-0.435	Body	Inter
cg17493885	NSD1, NSD1	chr5	-0.435	Promoters, 1to5kb	CGIs
cg19264571	APCDD1	chr18	-0.434	Promoters	CGIs
cg23279355	CMYA5	chr5	-0.434	Promoters	Inter
cg01791034	PLA2G4E	chr15	-0.434	Body	Inter
cg25665606	KLHDC7B	chr22	-0.434	Promoters	CGIs
cg19227031	PPT2, PPT2-EGFL8, PRRT1, PPT2, PPT2-EGFL8, PRRT1	chr6	-0.434	1to5kb, 1to5kb, 1to5kb, Body, Body, Body	CGIs
cg03453449	USP44, USP44, USP44	chr12	-0.434	1to5kb, 5UTRs, Body	shelves
cg20632887	ATP1A2, ATP1A2	chr1	-0.433	5UTRs, Body	Inter
cg00399059	LYNX1, LYNX1	chr8	-0.433	Promoters, 1to5kb	shores
cg07168556	WFIKKN2, WFIKKN2	chr17	-0.433	Promoters, Body	Inter
cg21277452	LINC01623	chr6	-0.433	Body	shelves
cg27179622	LEFTY2	chr1	-0.433	Body	CGIs
cg00070899	GRM4	chr6	-0.433	Body	shores
cg13058819	SLC35F3	chr1	-0.433	Body	shores
cg12407057	VWA1	chr1	-0.432	1to5kb	shelves
cg10859192	ATP5MC2	chr12	-0.432	1to5kb	shores
cg12716639	USP2	chr11	-0.431	Body	shores
cg24024660	TMEM225B	chr7	-0.431	Promoters	CGIs
cg16361867	GRIK4	chr11	-0.431	Body	Inter
cg12182937	HCG25	chr6	-0.431	1to5kb	shores
cg00213295	TMEM132E	chr17	-0.431	Body	Inter
cg06687305	CCDC17	chr1	-0.431	Body	CGIs
cg01119278	METTL24, DDO	chr6	-0.430	Body, Body	CGIs

Table S40 continued from previous page

cg25356715	CACNA2D4	chr12	-0.430	Body	shores
cg24874173	NBPF20, NBPF10, NBPF20, NBPF10	chr1	-0.430	Promoters, Promoters, Body, Body	shores
cg16677448	DMKN, DMKN	chr19	-0.430	Promoters, 1to5kb	Inter
cg12033622	COX4I2	chr20	-0.429	Promoters	CGIs
cg18700133	ALOXE3	chr17	-0.428	Body	CGIs
cg25642234	PLBD1	chr12	-0.428	Promoters	CGIs
cg13734833	MRGPRF, MRGPRF	chr11	-0.427	Promoters, Body	shores
cg20889395	WFIKKN2, WFIKKN2	chr17	-0.427	Promoters, Body	Inter
cg12127196	AGBL4	chr1	-0.427	Body	shores
cg21215576	FAM83A-AS1, FAM83A, FAM83A-AS1, FAM83A	chr8	-0.427	1to5kb, 1to5kb, Body, Body	shores
cg07191900	IL17RA	chr22	-0.427	Promoters	CGIs
cg07734975	SBF2-AS1, LINC02709	chr11	-0.426	Body, Body	shores
cg14552982	EXOSC2	chr9	-0.426	Promoters	shores
cg13202523	AZGP1P1	chr7	-0.426	Body	Inter
cg04715503	CLDN9, CLDN9	chr16	-0.426	5UTRs, Body	shores
cg08880849	HSPB9, KAT2A, HSPB9, KAT2A	chr17	-0.425	Promoters, Promoters, 1to5kb, 1to5kb	CGIs
cg13702536	HCAR1	chr12	-0.425	Promoters	Inter
cg22670033	TRIM41	chr5	-0.425	1to5kb	shores
cg03342928	SSC5D	chr19	-0.424	Body	CGIs
cg06489965	CLDN9, CLDN9	chr16	-0.424	5UTRs, Body	shores
cg27268279	COL13A1	chr10	-0.424	Body	Inter
cg27532360	ITGA2B	chr17	-0.424	Body	Inter
cg02723291	ETNK1	chr12	-0.424	Promoters	shores
cg07349815	KALRN	chr3	-0.423	1to5kb	shores
cg24460126	FAM163A	chr1	-0.423	Body	CGIs
cg14481208	RTKN, RTKN	chr2	-0.423	Promoters, 1to5kb	CGIs
cg12529273	LNX1-AS2, LNX1, PDGFRA, LNX1-AS2, LNX1, PDGFRA	chr4	-0.423	Promoters, Promoters, Promoters, Body, Body, Body	Inter
cg02179982	WNT7B	chr22	-0.423	Body	shores
cg26776957	PLEKHB1, PLEKHB1	chr11	-0.422	Promoters, 1to5kb	Inter
cg14947478	GDF5, GDF5, GDF5	chr20	-0.422	Promoters, 5UTRs, Body	shelves
cg21109038	ZNF710	chr15	-0.422	1to5kb	CGIs
cg01510388	CYP7B1	chr8	-0.422	Promoters	CGIs
cg13411784	DPH7	chr9	-0.421	Promoters	CGIs
cg03677069	MMRN2, SNCG, MMRN2, SNCG, MMRN2, SNCG	chr10	-0.421	Promoters, Promoters, 5UTRs, 5UTRs, Body, Body	Inter
cg11069430	USP44, USP44, USP44	chr12	-0.421	1to5kb, 5UTRs, Body	shelves
cg01343363	NDRG4, NDRG4	chr16	-0.421	Promoters, Body	shores

Table S40 continued from previous page

cg12229151	CAVIN3, CAVIN3	chr11	-0.421	Body, 3UTRs	shores
cg13266496	METTL24, DDO	chr6	-0.420	Body, Body	shores
cg08610426	IZUMO1, IZUMO1, IZUMO1	chr19	-0.419	Promoters, 5UTRs, Body	Inter
cg20595752	TEAD3	chr6	-0.418	Body	shelves
cg10541110	KRTAP2-2, KRTAP2-3, KRTAP2-2, KRTAP2-3, KRTAP2-2, KRTAP2-3	chr17	-0.418	1to5kb, 1to5kb, Body, Body, 3UTRs, 3UTRs	Inter
cg11210878	CHRM1	chr11	-0.418	Promoters	shores
cg01293485	ACKR2, HIGD1A, ACKR2, HIGD1A	chr3	-0.418	1to5kb, 1to5kb, Body, Body	shores
cg06672093	ESX1	chrX	-0.417	1to5kb	shelves
cg12733396	COL9A2	chr1	-0.417	Promoters	shores
cg25240254	SSC5D	chr19	-0.417	Body	CGIs
cg15602677	SNTG2	chr2	-0.417	Body	shelves
cg19127638	TET1	chr10	-0.416	Body	shores
cg07824824	CIDEA	chr18	-0.416	Body	CGIs
cg22031964	CHRM1	chr11	-0.416	Body	Inter
cg13647973	SBK3	chr19	-0.416	1to5kb	CGIs
cg08274637	DLEU7, DLEU7-AS1, DLEU7, DLEU7-AS1	chr13	-0.415	Promoters, Promoters, Body, Body	CGIs
cg01887374	HS6ST1	chr2	-0.415	1to5kb	CGIs
cg02299497	RAD51B	chr14	-0.415	Body	shores
cg14625636	SUGT1P4-STRA6LP, SUGT1P4-STRA6LP-CCDC180, SUGT1P4-STRA6LP, SUGT1P4-STRA6LP-CCDC180	chr9	-0.415	Promoters, Promoters, Body, Body	shores
cg03993109	ADAMTS2	chr5	-0.415	Body	CGIs
cg09221960	GRIK4, GRIK4	chr11	-0.415	Promoters, Body	shores
cg01043616	SMARCD1, ASIC1, SMARCD1, ASIC1	chr12	-0.414	1to5kb, 1to5kb, Body, Body	CGIs
cg05413199	TPPP3, ZDHHC1, TPPP3, ZDHHC1	chr16	-0.414	1to5kb, 1to5kb, Body, Body	shores
cg21123519	RAD51B	chr14	-0.414	Body	shores
cg11070193	CATSPER4	chr1	-0.414	Body	Inter
cg18947553	ZNF678	chr1	-0.414	1to5kb	CGIs
cg11237406	ZMYND15, ZMYND15	chr17	-0.414	Body, 3UTRs	Inter
cg17674042	PACSIN1, PACSIN1	chr6	-0.414	Promoters, Body	Inter
cg17112426	DMKN, DMKN	chr19	-0.414	Promoters, 1to5kb	Inter
cg05593775	PDZD7	chr10	-0.413	Body	CGIs
cg17512474	CDX1	chr5	-0.413	Promoters	CGIs
cg03577139	SOD3, SOD3	chr4	-0.413	Promoters, 1to5kb	Inter
cg21148531	ATG5	chr6	-0.413	Promoters	shores

Table S40 continued from previous page

cg08798685	LOC100131289	chr6	-0.413	Body	Inter
cg05374238	FYB2	chr1	-0.413	Promoters	Inter
cg01906695	COL2A1	chr12	-0.413	1to5kb	shores
cg17762073	GRM4	chr6	-0.413	Body	CGIs
cg00326131	RTKN, RTKN	chr2	-0.412	Promoters, Body	CGIs
cg06410849	CDH16, RRAD, CDH16, RRAD	chr16	-0.412	1to5kb, 1to5kb, Body, Body	CGIs
cg17356252	AIRE, AIRE	chr21	-0.412	Promoters, 1to5kb	CGIs
cg15006681	H1-0, GCAT	chr22	-0.412	1to5kb, 1to5kb	shores
cg25360385	GALNT6, SLC4A8, GALNT6, SLC4A8	chr12	-0.412	1to5kb, 1to5kb, Body, Body	shores
cg22623319	RBM20	chr10	-0.411	Body	shelves
cg10408284	WFIKKN2, WFIKKN2, WFIKKN2	chr17	-0.411	Promoters, 5UTRs, Body	Inter
cg14742937	ADAM5	chr8	-0.411	Promoters	Inter
cg02242734	SLC25A18	chr22	-0.411	Body	shores
cg04664126	SCAND3, SCAND3	chr6	-0.411	Promoters, Body	shores
cg06184825	CPNE5	chr6	-0.411	Body	Inter
cg05568941	GJD3, RARA, GJD3, RARA	chr17	-0.410	Body, Body, 3UTRs, 3UTRs	shores
cg07227033	ARL4A, ARL4A	chr7	-0.410	Promoters, 1to5kb	shores
cg04625615	ITPKA	chr15	-0.410	Body	CGIs
cg20955328	KLF15	chr3	-0.410	Body	CGIs
cg13341668	HYAL2, HYAL2	chr3	-0.410	5UTRs, Body	CGIs
cg06420834	MYL9, MYL9	chr20	-0.409	Promoters, Body	CGIs
cg19782271	STK10, EFCAB9, STK10, EFCAB9	chr5	-0.409	Promoters, Promoters, 1to5kb, 1to5kb	shores
cg00220225	SEMA3F, SEMA3F, SEMA3F	chr3	-0.409	Promoters, 1to5kb, Body	shores
cg18137414	SHISA8, SHISA8	chr22	-0.409	1to5kb, Body	shores
cg14190522	DAB2IP	chr9	-0.409	Body	Inter
cg06117341	CLDN9, CLDN9	chr16	-0.409	5UTRs, Body	shores
cg03609493	MIR572	chr4	-0.409	Promoters	CGIs
cg19873297	EFCAB9, STK10	chr5	-0.408	1to5kb, 1to5kb	shores
cg00668150	CLUH	chr17	-0.408	Promoters	shores
cg14478475	SLC25A51	chr9	-0.408	Promoters	shores
cg14920044	SLC16A12	chr10	-0.408	Promoters	shores
cg22275125	SLC22A8, SLC22A8	chr11	-0.407	Promoters, 1to5kb	Inter
cg00610692	RTKN, RTKN	chr2	-0.407	Promoters, Body	CGIs
cg17052170	LY6E, LY6E-DT, LY6E, LY6E-DT	chr8	-0.407	Promoters, Promoters, Body, Body	CGIs
cg17454857	ERGIC3	chr20	-0.407	1to5kb	shelves
cg24176731	PLEKHB1, PLEKHB1	chr11	-0.407	Promoters, 1to5kb	Inter

Table S40 continued from previous page

cg16667508	MATN4, RBPJL, MATN4, RBPJL, MATN4, RBPJL	chr20	-0.406	1to5kb, 1to5kb, 5UTRs, 5UTRs, Body, Body	shores
cg00079785	SPIDR	chr8	-0.406	Promoters	shores
cg15701111	KRT81, KRT86, KRT81, KRT86	chr12	-0.406	Promoters, Promoters, Body, Body	shores
cg18971434	LINC01623	chr6	-0.406	Body	shelves
cg25432738	MIR4499, CRYL1, MIR4499, CRYL1	chr13	-0.406	1to5kb, 1to5kb, Body, Body	CGIs
cg15175162	FBXL5	chr4	-0.406	Promoters	CGIs
cg05366561	FOXS1, DUSP15, FOXS1, DUSP15	chr20	-0.406	1to5kb, 1to5kb, Body, Body	CGIs
cg18604823	GALP	chr19	-0.405	Promoters	Inter
cg14644378	DNAJB5	chr9	-0.405	1to5kb	shores
cg24496745	FYB2, FYB2	chr1	-0.405	5UTRs, Body	Inter
cg00054210	CYP7B1	chr8	-0.405	Promoters	CGIs
cg25001930	NBPF8	chr1	-0.405	Body	Inter
cg02920396	MRGPRF, MRGPRF	chr11	-0.405	1to5kb, Body	shelves
cg11757040	DPF3	chr14	-0.404	Body	Inter
cg00689340	RTKN, RTKN	chr2	-0.404	Promoters, Body	CGIs
cg17774102	LINC01623	chr6	-0.404	Body	shelves
cg07298985	PIWIL2, PIWIL2	chr8	-0.404	Promoters, Body	CGIs
cg06599908	TTYH1	chr19	-0.404	Body	shelves
cg16034541	MUC4	chr3	-0.404	Body	shores
cg26644853	CX3CL1	chr16	-0.404	Promoters	Inter
cg05463589	IL17C	chr16	-0.403	Body	CGIs
cg17827650	NOM1	chr7	-0.403	Promoters	CGIs
cg02987388	CCDC17	chr1	-0.403	Body	CGIs
cg21160372	NCAM2	chr21	-0.403	1to5kb	shores
cg05254747	SLC39A14, SLC39A14	chr8	-0.403	Promoters, 1to5kb	shores
cg15116481	KLK1, KLK15, KLK1, KLK15	chr19	-0.403	1to5kb, 1to5kb, Body, Body	CGIs
cg13981310	ENTPD2, ENTPD2	chr9	-0.403	Promoters, 1to5kb	shores
cg04566512	MIRLET7BHG, MIRLET7BHG	chr22	-0.403	1to5kb, Body	CGIs
cg18675097	ZKSCAN4, NKAPL, ZKSCAN4, NKAPL, ZKSCAN4, NKAPL	chr6	-0.402	Promoters, Promoters, 5UTRs, 5UTRs, Body, Body	CGIs
cg13655803	WFIKKN2, WFIKKN2	chr17	-0.402	5UTRs, Body	Inter
cg17295878	TBC1D16, TBC1D16	chr17	-0.402	Promoters, Body	shores
cg23109344	TDRP, TDRP	chr8	-0.402	Promoters, 1to5kb	shores
cg18919720	FSCN2	chr17	-0.402	Promoters	CGIs
cg02002217	ANKRD20A19P	chr13	-0.402	Body	Inter
cg05485462	CLDN9	chr16	-0.402	Promoters	shores

Table S40 continued from previous page

cg09240693	LOC284950	chr2	-0.401	1to5kb	CGIs
cg03558837	ESPNL	chr2	-0.401	Body	Inter
cg25571880	PRSS56	chr2	-0.401	Promoters	shelves
cg06106599	FADD	chr11	-0.401	Promoters	shores
cg25556225	LINC03034	chr15	-0.401	Body	Inter
cg09336899	NGEF	chr2	-0.400	Body	CGIs
cg05415308	CT62	chr15	-0.400	Body	shores
cg12813394	RAD51B	chr14	-0.400	Body	CGIs

5.2.1 Common Genes

Table S41. Common differentially methylated genes of CHARGE episignature with respective mean $\Delta\beta$ values of Sadikovic and Weksberg episignatures.

Genes	DNAm Effect	mean $\Delta\beta$	Genes	DNAm Effect	mean $\Delta\beta$
VWF	Gain	0.192	GFI1	Gain	0.109
HOTAIRM1	Loss	0.149	NOX4	Loss	0.108
HOXA1	Loss	0.149	KIRREL3	Loss	0.103
LINC02914	Loss	0.140	FMN2	Loss	0.098
APP	Gain	0.134	DAB1	Loss	0.098
SLITRK5	Loss	0.133	TEAD1	Loss	0.094
LINC02381	Loss	0.130	ARHGEF15	Gain	0.093
HOXA5	Gain	0.117	FOXP2	Loss	0.093
HOXA-AS3	Gain	0.115	PARVA	Loss	0.092
COL4A2	Loss	0.109	PCDH20	Loss	0.088
SLCO1A2	Gain	0.109			

Table S42. Common differentially methylated genes of Kabuki episignature with respective mean $\Delta\beta$ values of Sadikovic and Weksberg episignatures.

Genes	DNAm Effect	mean $\Delta\beta$	Genes	DNAm Effect	mean $\Delta\beta$	Genes	DNAm Effect	mean $\Delta\beta$
TNS1	Loss	0.208	CLYBL	Gain	0.149	FGFR1OP2	Loss	0.104
AGAP2	Loss	0.206	SLMAP	Loss	0.145	INTS13	Loss	0.104
AGAP2-AS1	Loss	0.206	ARHGEF7	Loss	0.134	ESR1	Gain	0.104
ZMIZ1	Gain	0.183	NOL4	Loss	0.128	CHIT1	Gain	0.104
IL17C	Gain	0.173	C6orf62	Loss	0.125	IL17RE	Gain	0.102
LAMA1	Loss	0.160	CD37	Gain	0.124	CIDEC	Gain	0.102
SNORD4A	Loss	0.160	PLD4	Loss	0.122	CNTN5	Loss	0.099
SNORD42A	Loss	0.160	HLX	Loss	0.121	KIRREL3	Gain	0.096
SNORD4B	Loss	0.160	RPS6	Loss	0.118	TFAP2E	Loss	0.095
RAB34	Loss	0.160	PTDSS2	Gain	0.117	PSMB2	Loss	0.095
RPL23A	Loss	0.160	MXRA8	Loss	0.117	PARP4	Loss	0.090

Table S42 continued from previous page

GARS1-DT	Loss	0.151	MAP3K6	Gain	0.114	ZNF787	Gain	0.081
GARS1	Loss	0.151	H2BC7	Loss	0.113	PPP1R10	Loss	0.078
CIDEB	Loss	0.150	SLC33A1	Loss	0.113	MRPS18B	Loss	0.078
LTB4R	Loss	0.150	SCARF2	Loss	0.107			

Table S43. Common differentially methylated genes of Sotos episignature with respective mean $\Delta\beta$ values of Sadikovic and Weksberg episignatures.

Genes	DNAm Effect	mean $\Delta\beta$	Genes	DNAm Effect	mean $\Delta\beta$	Genes	DNAm Effect	mean $\Delta\beta$
PCAT6	Loss	0.563	RPLP1	Loss	0.355	FOXL1	Loss	0.312
KDM5B	Loss	0.563	LDB3	Loss	0.354	CDC14B	Loss	0.305
KRBA1	Loss	0.470	DNLZ	Loss	0.353	CERS4	Loss	0.301
PRPSAP1	Loss	0.468	CARD9	Loss	0.353	RD3	Loss	0.300
FAM118A	Loss	0.463	H1-0	Loss	0.352	CHIT1	Loss	0.299
KCTD12	Loss	0.453	GCAT	Loss	0.352	AMN	Loss	0.298
C17orf98	Loss	0.438	MYO1C	Loss	0.350	ANKRD13B	Loss	0.297
RBBP8NL	Loss	0.435	GALNT6	Loss	0.350	GIT1	Loss	0.297
SDHAF1	Loss	0.430	SLC4A8	Loss	0.350	NABP2	Loss	0.296
ZPR1	Loss	0.422	CERS2	Loss	0.349	RNF41	Loss	0.296
APOA5	Loss	0.422	C2orf81	Loss	0.349	TMEM190	Loss	0.294
MIR200B	Loss	0.397	SFN	Loss	0.347	KCNH2	Loss	0.293
SSC5D	Loss	0.392	TMEM72-AS1	Loss	0.346	GPD1	Loss	0.289
LINC01623	Loss	0.391	HLX	Loss	0.346	RASIP1	Loss	0.288
H3C1	Loss	0.385	HYAL2	Loss	0.342	XPO7	Loss	0.287
FAM83A-AS1	Loss	0.385	TMEM105	Loss	0.335	DOK2	Loss	0.287
FAM83A	Loss	0.385	CXCL12	Loss	0.334	BBS4	Loss	0.286
H4C1	Loss	0.380	F10	Loss	0.333	HIGD2B	Loss	0.286
H1-1	Loss	0.380	TDGF1	Loss	0.333	CKM	Loss	0.283
MIR429	Loss	0.379	VAX2	Loss	0.332	COL12A1	Loss	0.282
ITGA2B	Loss	0.373	DDR1	Loss	0.326	ZNF710	Loss	0.280
MIR200A	Loss	0.372	GPSM3	Loss	0.326	CYGB	Loss	0.272
TLCD4-RWDD3	Loss	0.372	NOTCH4	Loss	0.326	PRCD	Loss	0.272
RWDD3	Loss	0.372	MEIKIN	Loss	0.325	BTBD17	Loss	0.271
PPP1R1B	Loss	0.370	ACSL6	Loss	0.325	NCS1	Loss	0.263
TP73	Loss	0.367	LRRC2	Loss	0.324	ASIC4	Loss	0.258
DPM2	Loss	0.361	DLX3	Loss	0.324	ZNF423	Loss	0.256
RTKN	Loss	0.361	UBALD1	Loss	0.323	COCH	Loss	0.254
TPPP3	Loss	0.361	NUP210	Loss	0.319	IL11	Loss	0.248
ZDHHC1	Loss	0.361	CREB3L1	Loss	0.315			
TET1	Loss	0.357	GAB1	Loss	0.312			