Electronic Thesis and Dissertation Repository

4-19-2022 1:00 PM

Identification of DNA Methylation Episignatures for Classification and Phenotype/Genotype Correlation in Mendelian Neurodevelopmental Disorders

John Reilly, *The University of Western Ontario*

Supervisor: Bekim Sadikovic, *London Health Sciences Center, Verspeeten Genome Center*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Pathology and Laboratory Medicine
© John Reilly 2022

ABSTRACT:

Diagnosis for neurodevelopmental disorders poses numerous challenges, related to the lack of specific findings and limited understanding of clinical impact of the majority of genetic variation. Epigenomics mechanisms involve chemical modifications in DNA that involve a range of cellular mechanisms. DNA methylation is an epigenetic mechanism involving addition and removal of methyl groups to cytosine residues. These methylation signals form episignatures; patterns of methylation that can be used as biomarkers capable of differentiating neurodevelopmental disorders. EpiSigns have enabled molecular diagnosis of a number of genetic conditions, classification of variants of unknown significance, and provided insights into the pathophysiology of neurodevelopmental disorders. I hypothesized DNA methylation can provide classifications of neurodevelopmental disorders, and identify epigenetic patterns that relate the phenotypic and genotypic variations seen in these patients. Main objectives of this work include 1) determination of syndrome specific episignatures, 2) analysis of domain specific variants and their effects on the methylation profiles and ensuing phenotypes 3) determine effectiveness of episignature assessment in classifying neurodevelopmental disorders in paralogous genes, 4) assessing phenotypic overlap between distinct neurodevelopmental disorders and correlation to their methylation profiles. My thesis demonstrates that these episignatures are robust biomarkers that inform effective methods to diagnose complex neurodevelopmental disorders and provide evidence of shared functional pathways highlighted by the various genomic and phenotypic contexts episignatures have been derived from.

LAY SUMMARY:

Within each cell, genetics acts as a blueprint to provide instructions for the creation and maintenance of various cellular structures and functions, however, given how this genetic blueprint is identical across all cells in an organism, additional methods of control must exist. While the blueprint remains the same, different levels of expression of genes within the genome allow for differentiation, resulting in the various different types of cells available in the human body. Additionally, cells can change over time, at different developmental periods, where particular genes are expressed, and eventually turned off when their function is complete. The study of this phenomenon, where the genome is not altered, but instead has the expression of different regions turned on and off, is referred to as epigenetics. One method for this type of epigenetic change is DNA methylation, a chemical mark that can be attached to parts of the genome that changes how genes at that region of the genome are expressed, similar to an on/off switch. When defects in the genome or epigenome occur, disorder of the cell's control systems is caused, and disease ensues, as switches supposed to remain off are switched on, or vice versa. This thesis works to observe how they differ from persons with disease compared to those without them, to create episignatures, chemical fingerprints of gene defects. In particular I have assessed DNA methylation in relation to neurodevelopmental disorders, syndromes that present with a complex set of characteristics related to intellectual development, and cognitive abilities. By cataloging and describing the genome of persons with neurodevelopmental disorders, we can identify which on/off switches are in disarray compared to their healthy counterparts, helping to better understand the ways in which these disorders present themselves, and provide ways to identify them in new persons.

TABLE OF CONTENTS

CO-AUTHORSHIP STATEMENT

Chapter 1: Introduction

Jack Reilly, Jennifer Kerkhof, Bekim Sadikovic

In the original book chapter published in Advances of Molecular Pathology entitled "DNA Methylation Episigns in Mendelian Neurodevelopmental Disorders as a Diagnostic Link Between a Genotype and Phenotype", Jack Reilly reviewed the literature, created figures in collaboration with Gavin Riddolls Graphic Design, and wrote significant portions of the text in collaboration with Co-authors Kerkhof and Sadikovic. Estimated contribution: 50%

Chapter 2: DNA methylation Episignature in Gabriele De Vries syndrome

Florian Cherik, Jack Reilly, Jennifer Kerkhof, Michael Levy, Haley McConkey, Mouna Barat-Houari, Kameryn M. Butler, Christine Coubes, Jennifer A. Lee, Gwenael Le Guyader, Raymond J. Louie, Wesley G. Patterson, Matthew L. Tedder, Mads Bak, Trine Bjørg Hammer, William Craigen, Florence Démurger, Christèle Dubourg, Mélanie Fradin, Rachel Franciskovich, Eirik Frengen, Jennifer Friedman, Nathalie Ruiz Palares, Maria Iascone, Doriana Misceo, Pauline Monin, Sylvie Odent, Christophe Philippe, Flavien Rouxel, Veronica Saletti, Petter Strømme, Perla Cassayre Thulin, Bekim Sadikovic*, David Genevieve*

Dr Cherik and Dr Genevieve coordinated the collection of peripheral blood samples from various clinical collaborators and provided detailed characteristics of each patient's presentation of the disorder. Jennifer Kerkhof processed these samples on microarrays at London Health Sciences Centre and provided the methylation data for the statistical analysis pipeline designed by Michael Levy. Analysis of methylation data was carried out by Jack Reilly, performing statistical analyses, creating the majority of paper figures, and providing detailed results summaries to the manuscript writing team. Dr Cherik and Jack Reilly then collaborated on the manuscript, with Dr Cherik focusing on description of clinical data, while Jack focused on the methods, and results of the methylation data analysis. Dr Sadikovic conceptualized and with Dr Genevieve co-supervised the project. Estimated contribution: 50%

Chapter 3: Delineation of a KDM2B-related neurodevelopmental disorder and its associated DNA methylation signature

Richard H. van Jaarsveld, Jack Reilly, Marie-Claire Cornips, Michael A. Hadders,

Emanuele Agolini, Priyanka Ahimaz,Kwame Anyane-Yeboa, Severine Audebert Bellanger, Ellen van Binsbergen, Marie-Jose van den Boogaard, Elise Brischoux-Boucher, Raymond

Caylor, Andrea Ciolfi, Ton AJ van Essen, Paolo Fontana, Saskia Hopman, Maria

Iascone, Margaret M Javier, Erik-Jan Kamsteeg, Jennifer Kerkhof, Jun Kido, HyungGoo Kim, Tjitske Kleefstra, Fortunato Lonardo, Abbe Lai, Dorit Lev, Michael A. Levy, Suzanne M.E. Lewis, Angie Lichty, Naomichi Matsumoto, Idit Maya, Haley McConkey, Andre Megarbane, Vincent Michaud, Evalina Miele, Marcello Niceta, Antonio Novelli, Roberta Onesimo, Rolph Pfundt, Bernt Popp, Eloise Prijoles, Raissa Relator, Sylvia Redon, Dmitrijs Rots, Karen Rouault, Ken Saida, Jolanda Schieving, Marco Tartaglia, Romano Tenconi, Kevin Uguen, Nienke Verbeek,

Christopher A Walsh, Keren Yosovich, Christopher J. Yuskaitis, Giuseppe Zampino

Bekim Sadikovic, Mariëlle Alders, Renske Oegema

Dr Oegema, and Dr Alders coordinated the collection of peripheral blood samples from various clinical collaborators, and provided detailed characteristics of each patient's presentation of the disorder. Jennifer Kerkhof processed these samples on microarrays at London Health Sciences Center, and provided the methylation data for the statistical analysis pipeline designed by Michael Levy. Analysis of methylation data was carried out by Jack Reilly, performing statistical analyses, creating the majority of paper figures, and providing detailed results summaries to the manuscript writing team. Dr Jaarsveld, Dr Oegema and Jack Reilly then collaborated on the manuscript, with Dr Jaarsveld and Oegema focusing on description of clinical data, while Jack focused on the methods, and results of the methylation data analysis. Dr Sadikovic conceptualized and with Drs Alders and Oegema co-supervised the project. Estimated contribution: 50%

Chapter 4: DNA methylation episignature are high sensitivity and specificity biomarkers for detection of patients with KAT6A and KAT6B mutations

Niels Vos MD, Jack Reilly, M.W. Elting MD,PhD, Philippe M Campeau MD, Coman D MD, Zornitza Stark MD, Tiong Yang Tan MD, David J Amor MD, Benjamin A Kamien MD, Chirag Patel MD, Matthew L Tedder PhD, Giuseppe Merla PhD, Paolo Prontera MD, Marco Castori MD, PhD, Kai Muru PhD, Felicity Collins MD, Janine Smith MD, Bruria Ben Zeev MD, Alessandra Murgia MD, PhD, Emanuela Leonardi PhD, Natacha Esber, Antonio Martinez-Monseny MD, Matthew Wallis MD, Marielle Alders MD,PhD, Bekim Sadikovic PhD

Dr Vos, and Dr Alders coordinated the collection of peripheral blood samples from various clinical collaborators, and provided detailed characteristics of each patient's presentation of each disorder. Jennifer Kerkhof processed these samples on microarrays at London Health Sciences Centre, and provided the methylation data for the statistical analysis pipeline designed by Michael Levy. Analysis of methylation data was carried out by Jack Reilly, performing statistical analyses, creating the majority of paper figures, and providing detailed results summaries to the manuscript writing team. Dr Vos and Jack Reilly then collaborated on the manuscript, with Dr Vos focusing on description of clinical data, while Jack focused on the methods, and results of the methylation data analysis. Dr Sadikovic conceptualized, and Dr Alders co-supervised the project. Estimated contribution: 50%

Chapter 5: Shared Blepharophimosis Phenotype Leads to Discovery of Common DNA Methylation Episignature in patients with SMARCA2 and ADNP variants

Nicola Brunetti, Jack Reilly, Gerarda Cappucio, Sadikovic B, Jennifer Kherkof

Dr Brunetti and Dr Cappucio coordinated the collection of peripheral blood samples from various clinical collaborators, and provided detailed characteristics of each patient's presentation of each disorder. Jennifer Kerkhof processed these samples on microarrays at London Health Sciences Centre, and provided the methylation data for the statistical analysis pipeline designed by Michael Levy. Analysis of methylation data was carried out by Jack Reilly, performing statistical analyses, creating the majority of paper figures, and providing detailed results summaries to the manuscript writing team. Dr Brunetti and Jack Reilly then collaborated on the manuscript, with Dr Brunetti focusing on description of clinical data, while Jack focused on the methods, and results of the methylation data analysis. Dr Sadikovic conceptualized and supervised the project. Estimated contribution: 50%

DEDICATION

I dedicate this thesis to my family, who always show me that the world is a bright place when you are curious and kind.

LIST OF FIGURES

Chapter 5: Shared Blepharophimosis Phenotype Leads to Discovery of Common DNA Methylation Episignature in patients with SMARCA2 and ADNP variants

LIST OF TABLES

LIST OF ABBREVIATIONS

5-caC   -   5-Carboxylcytosine

5-fC   -   5-Formylcytosine

5-hmC   -   5-hydroxymethylcytosine

5-mC   -   5-Methylcytosine

ACMG   -   American College of Medical Genetics

ADCADN   -   Adult-Onset Autosomal Dominant Cerebellar Ataxia, with Deafness and Narcolepsy

ADNP   -   Activity Dependent Neuroprotector Protein

aKG   -   Alpha Ketoglutarate

AMED   -   Agency for Medical Research and Development

AMP   -   Association for Molecular Pathology

APOBEC2   -   Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 2

Array-CGH   -   Array Comparative Genomic Hybridization

ASD   -   Autism Spectrum Disorder

ASD   -   Atrial Septal Defect

ATP   -   Adenosine Triphosphate

AUMC   -   Amsterdam University Medical Centers

BAF   -   BRG1/BRM Associated Factor

BER   -   Base Excision Repair

BH   -   Benjamini Hochberg

BIS   -   Blepharophimosis Impaired Intellectual Development Syndrome

BMP4   -   Bone Morphogenic Protein 4

BNL	-	Bipartite Nuclear Localization
BRG1	-	BOI
BRM	-	Brahma protein
BRPF1	-	Bromodomain and PHD Finger Containing Protein
cAMP	-	Cyclic Adenosine Monophosphate
CDLS	-	Cornelia De Lange Syndrome
CFCS	-	Communication Function Classification System
CHARGE	-	Choanal Atresia, Retardation, Genital and Ear Anomalies
CHD	-	Chromodomain
CLP	-	Cleft Lip/Palate
CMA	-	Chromosomal Microarray
CNS	-	Central Nervous System
CNV	-	Copy Number Variation
COPD	-	chronic obstructive pulmonary disease
CP	-	Cerebral Paresis
CpG	-	Cytosine
CREB	-	cAMP Response Element Binding Protein
CSS	-	Coffin Siris Syndrome
CUL1	-	Cullin 1
DD	-	Developmental Delay
DFG	-	Deutsche Forschungsgemeinschaft
DMG	-	Differentially Methylated Gene
DMR	-	Differentially Methylated Regions

DNA	-	Deoxyribonucleic Acid
DNMT	-	DNA Methyltransferase
DORV	-	Double Outlet Right Ventricle
DSM-5	-	Diagnostic and Statistical Manual of Mental Disorders 5
DYT28	-	Dystonia 28
EKD	-	Episign Knowledge Database
F	-	Female
FBXL10	-	F-Box and Leucine Rich Repeat Protein 10
FDR	-	False Discovery Rates
Fe2+	-	Iron Ion
GEO	-	Gene Expression Omnibus
GLI2	-	Gli-Kruppel Family Member 2
GNAT	-	Gcn5 Associated Acetyltransferases
gnomAD	-	Genome Aggregation Database
GTEx	-	Gene-Tissue Expression
GTPTS	-	Genitopatellar Syndrome
H3K36	-	Histone 3, Lysine 36
H3K4	-	Histone 3, Lysine 4
HAT	-	Histone Acetyltransferase
HDAC	-	Histone Deacetylase
HEY2	-	HES-Related Family bHLH Transcription Factor with YRPW Motif 2
HMA	-	Hunter McAlpine
HOTAIRM1	-	Hox Transcript Antisense RNA; Non-Coding

HOXA-AS3 - Hoxa Cluster Antisense RNA 3

HOXA1 - Homeobox A1

HOXA3 - Homeobox A3

HOXA5 - Homeobox A5

HOXA6 - Homeobox A6

HOXB3 - Homeobox B3

HOXB6 - Homeobox B6

HVDAS - Helsmoortel Van Der Aa Syndrome

ID - Intellectual Disability

IDDSELD - Intellectual Developmental Disorder with Seizures and Language Delay

ING5 - Inhibitor of Growth 5

INO80 - Inositol 80

ISWI - Imitation Switch

JHDM1B - Jumonji C Domain Containing Histone Demethylase Protein 1B

JmjC - Jumonji C

K2BNDD - KDM2B Related Neurodevelopmental Disorder

KAT6A - Lysine Acetyltransferase 6A

KAT6B - Lysine Acetyltransferase 6B

KDM2B - Lysine Demethylase 2B

KDM5C - Lysine Demethylase 5c

KMT2B - Lysine Methyltransferase 2B

KMT2D - Lysine Methyltransferase 2D

L - Left

LHSC	-	London Health Sciences Center
LOF	-	Loss of Function
LOOCV	-	Leave-One-Out-Cross-Validation
LRR	-	Leucine Rich Region
M	-	Male
MCD	-	Malformation of Cortical Development
MDB	-	Methyl CpG Binding Domain
MDEM	-	Mendelian Disorder of the Epigenetic Machinery
MDS	-	Multidimensional Scaling
MECP2	-	Methyl CpG Binding Protein 2
mESC	-	Murine Embryonic Stem Cells
MIM	-	Mendelian Inheritance in Man
MORF	-	Mortality Factor Like Protein
MOZ	-	Monocytic Leukemia Zinc Finger Protein
MR	-	Mitral Regurgitation
MVP	-	Methylation Variant Pathogenicity
MYST	-	Mys-1/TRR-1 Complex
NA	-	Not Assessed
NCBRS	-	Nicolaides Baraitser
ncRNA	-	Non- Coding Ribonucleic Acid
NDD	-	Neurodevelopmental Disorder
NGS	-	Next Generation Sequencing
NINDS	-	National Institute of Neurological Disorders and Stroke

NSD1	-	Nuclear Receptor- Binding SET Domain Protein 1
OMIM	-	Online Mendelian Inheritance In Man
PCA	-	Principal Component Analysis
PCR	-	Polymerase Chain Reaction
PDA	-	Persistent Ductus Arteriosus
PFO	-	Persistent Foramen Ovale
PHD	-	Plant Homeodomain
PHMDS	-	Phelan McDermid Syndrome
PVS	-	Pulmonary Valve Stenosis
RBBB	-	Right Bundle Branch Block
REB	-	Research Ethics Board
RING	-	Ring Finger Protein
ROC	-	Receiver Operating Characteristic
RTT	-	Rett Syndrome
SBBYSS	-	Say-Barber-Biesecker-Young-Simpson Syndrome
SD	-	Standard Deviation
SETD1B	-	SET-Domain Containing Protein 1B
SIFT	-	Sorting Intolerant From Tolerant
SMARCA2	-	SWI/SNF-Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily A, Member 2
SNHL	-	Sensorineural Hearing Loss
SNP	-	Single Nucleotide Polymorphism
SRCAP	-	SNF2-Related CBP Activator Protein

SVM	-	Support Vector Machine
SWI/SNF	-	Switch/Sucrose non-Fermentable
TBRS	-	Tatton-Brown- Rahman
TET	-	Ten-Eleven Translocases
Trio-ES	-	Trio Exome Sequencing
VSD	-	Ventricular Septal Defect
VUR	-	Vesicoureteral Reflux
VUS	-	Variant of Unknown Significance
WES	-	Whole Exome Sequencing
WGS	-	Whole Genome Sequencing
WT	-	Wild Type
YY1	-	Yin Yang 1
Zn2+	-	Zinc Ion

CONTRIBUTION TO ORIGINAL KNOWLEDGE

The work described in this thesis represents significant contributions to the field of genetics and neurodevelopmental disorders in several ways. First and foremost, I describe several novel episignatures, in chapter 2, an episignature for GADEVs, In chapter 3, an episignature for the newly describe K2BNDD, in Chapter 4, an episignature for KAT6A syndrome, and finally an episignature shared between ADNP and SMARCA2 cohorts. These signatures are effective biomarkers shown to have a high degree of sensitivity and specificity, and can be used in both clinical and research settings to better understand the disorders in question, and their molecular etiologies. Furthermore, I have integrated various sources of clinical information, including genetic phenotypic and epigenetic data, for assessment and refinement of the related episignatures. My work provides further evidence of the genotype and phenotype correlations with the consequent epigenetic patterns and provides further insights into the molecular pathophysiology of genetic neurodevelopmental disorders.

Chapter 1 : Introduction and Literature Review

Adapted from "DNA Methylation Episigns in Mendelian Neurodevelopmental Disorders as a Diagnostic Link Between a Genotype and Phenotype" Published in Advances in Molecular Pathology

1

1.1 LITERATURE REVIEW

1.1.1 DEFINING EPIGENETICS

Epigenetics refers to the study of cells ability to control gene activity without changes to the underlying genetic sequence [1,2]. The prefix "Epi-" from Greek, means above, or upon, and provides an effective insight into the mechanisms that epigenetics uses to modify the ensuing phenotypic expression of cells, namely, through modifications in the form of added molecular groups, such as methyl tags, and histone protein modifications. These chemical modifications involve DNA and the adjacent chromatin structures, allowing for regulation of gene expression through compaction of DNA elements, and restriction or easement of protein machinery access to the genetic sequence. Therefore, although no change is made to the composition or sequence of nucleotides, significant changes in expression of the various proteins and molecular signals that instruct the cells can be made.

1.1.2 ALTERNATE EPIGENETIC MECHANISMS: HISTONE MODIFICATIONS AND CHROMATIN REMODELERS

In this thesis, my research has focused primarily on the assessment of DNA methylation, which will be discussed in detail, however it is important to acknowledge the importance of other epigenetic modifiers, namely, histone modifications and chromatin remodeling complexes. Many of the previously established episignatures have been associated with the disruption of these machineries (See Figure 1-1), and they play an important part in the establishment of the epigenome. My primary assessment method throughout this project was based on DNA methylation biomarkers, however, the syndromes and cohorts of patients I will discuss further on concern disruption of genetic mechanisms related to several important histone modifying proteins, DNA methylation interactors, and chromatin remodeling complexes.
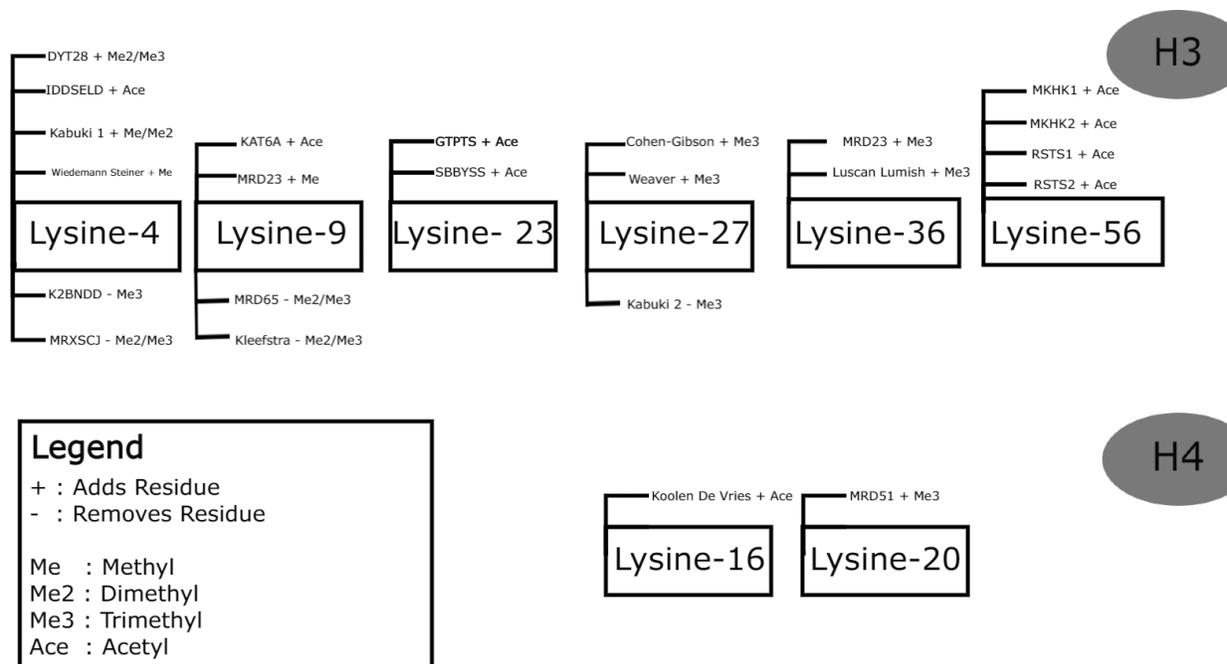
Figure 1-1. Episignatures associated with histone modifications. Outline of 23 disorders included in the Episign Version 3 paper [12] that involved direct modification of histone residues. Disorder names are paired with the histone modification for which the associated gene is responsible. 21 disorders were associated with histone 3 modifications, while an additional two disorders were associated with histone 4 modifications.

Histones are protein octamers that are organizational units that DNA is wrapped around, allowing for compact storage of the near 3 billion base pairs within the human genome [2]. This form of wrapped and compacted DNA is referred to as chromatin, comprising both the DNA sequence and its eight complexed organizational histone proteins. The compaction of chromatin into its transcriptionally inactive state; heterochromatin, or decompacted, transcriptionally active state; euchromatin is regulated by a plastic system of chemical modifiers. The addition or removal from the "tail" residues of the histone complexes results in changes to the histone's compaction, and therefore gene regulation activities. Histone tail modifications include but are not limited to, methylation, acetylation, phosphorylation, acylation, and ubiquitination, with each modification resulting in different chromatin states depending on the type and pattern of modification [3]. These changes, in concert with other epigenetic mechanisms, result in genetically identical cells within an organism being able to differentiate into the myriad of cell

lines and phenotypes, as well as providing a reversible plastic system of gene regulation, capable of interacting with and responding to environmental stimuli [4]. The interactions of histone modification and DNA methylation allow for an organism to differentiate both spatially, creating diverse and unique cell lines, as well as temporally, as epigenetic machinery can be recruited at various developmental timepoints to allow for preferential gene expression.

Chromatin remodelers are protein complexes capable of providing large scale changes to nucleosomes and their associated chromatin, promoting changes in nucleosome spacing and density, or facilitating histone variant exchange [5,6]. Disruption of these chromatin remodelers causes widespread changes to gene expression, which in the context of neural development, disrupts the tightly regulated network of factors which determine cell fates, resulting in the neurodevelopmental disorder (NDD) phenotypes observed. Proper reorganization of chromatin states, regulated by these complexes, allows for fine tuning of gene expression, making them essential [7,8]. Major actors include the Adenosine triphosphate (ATP) dependent chromatin remodeling enzyme family, which consists of four subgroups, the Switch/Sucrose non-fermentable (SWI/SNF), Imitation Switch family (ISWI) , Chromodomain-Helicase-DNA binding family (CHD), and the Inositol 80 family (INO80), which have been implicated in a number of neurodevelopmental disorders including Coffin Siris (CSS), Nicolaides Baraitser (NCBRS). The SWitch/Sucrose Non-Fermentable (SWI/SNF) complex, also known as the BRG1/BRM associated factor (BAF) is one such example of chromatin remodeler, which utilizes ATP dependent hydrolysis to alter the nucleosome, ultimately resulting in flexible changes between hetero-chromatin and euchromatin states. This SWI/SNF complex is involved in a number of epigenetic regulatory roles that impact developmental processes, particularly in the central nervous system (CNS). Lethality has been associated with complete knockout models of SWI/SNF, and even in the case of conditional knockout, severe neural agenesis was observed, indicating its essential status in proper development of neural structures[7,9,10]. Nevertheless, due to their complexity in terms of how they assemble their various subunits, and function in vivo, the exact role of each actor remains difficult to study, and future work remains to determine how these remodeler proteins can be accurately described and targeted for therapeutic interventions. [11]

1.1.3 DNA METHYLATION TECHNOLOGIES

DNA methylation can be assayed by several methods, including bisulfite conversion, restriction-enzyme digestion, DNA binding proteins with differential affinity for methylated sequences, pulldown antibody assays, and most recently, enzymatic methyl sequencing techniques [13, 14, 15, 16]. Microarray-based and sequencing methods have been built using these principles and allowed for extensive investigations into the epigenome. The current gold standard technique makes use of sodium bisulfite, which involves chemically modifying unmethylated cytosines. Bisulfite deaminates unmethylated cytosines to uracils, which are then converted to thymines following polymerase chain reaction (PCR) amplification, allowing for the identification of the remaining methylated cytosines through traditional sequencing or microarray methods. Whole genome bisulfite sequencing (WGBS), makes use of this technique, providing a genome wide view of methylation levels at single base resolutions [17, 18, 19]. This technology has shortcomings however, as the bisulfite conversion process can result in damage to the DNA sequence, as well as resulting in low levels of cytosine within the samples, making polymerase reactions required for sequencing difficult [20, 21]. This makes WGBS in tissues with limited genetic starting material difficult, as the margin of error for the amount of acceptable loss of DNA is quite small, and can create biases in the estimates of methylation levels.

Microarray technologies, such as the Illumina EPIC array used in my assessment of episignatures, make use of similar principals in the assessment of genome methylation levels. In DNA methylation microarrays, DNA is bisulfite converted to differentiate between methylated and unmethylated cytosines, as unmethylated cytosines are converted to uracil residues. The ensuing converted DNA is then provided to the microarray, wherein oligonucleotide probes, short sequences of DNA complementary to the processed DNA, are available for hybridization. For differentiation, probes specific to the methylated and unmethylated sequences are associated with different dyes, whose intensity can be measured to identify the distribution of methylated cytosines at each probe. Recent years have seen widespread adoption of microarray analysis in methylation research, due to its cost-effectiveness, versatility, high resolution at individual base pairs, and low amount of required DNA (~500ng) [22,23]. In comparison to sequencing

methods, microarrays only assay a selected portion of the genome, depending on the associated probeset, however it does so at a high level of coverage, providing a high level of precision [24]. Microarray methods suffer from the same shortfalls of WGBS, as the bisulfite conversion process can damage DNA, and requires significant amounts of starting material to proceed. Furthermore, the hybridization process is limited by bisulfite conversion, as bisulfite treated DNA only has 3 bases (A,T, and G) with which to create oligonucleotide probes for effective hybridization [25]. Nevertheless, microarray-based assessment of DNA methylation profiles remains a popular choice in research and clinical settings, representing a significantly more cost effective approach when compared to WGBS. Microarrays only assay a small proportion of the genome, they offer a high level of precision through high levels of coverage at these selected sites [24], furthermore, large increases in assayed regions have occurred over the past decade of technological advancements, with the latest version of the microarray platform from Illumina, the EPIC beadchip array covering over 850,000 CpG methylation sites [23].

Recently developed enzymatic methyl-sequencing techniques have been developed in response to the difficulties associated with the bisulfite conversion step of WGBS and microarray techniques, and validations of the process seem to indicate it to be an effective method of genomic methylation levels [15]. This technique makes use of epigenetic eraser proteins, known as ten-eleven translocases (TET), more specifically TET2 (OMIM# 612839), to oxidize methylated cytosines within the genome, which are then converted to uracils using an additional enzyme, apolipoprotein B mRNA editing enzyme catalytic subunit 2 (APOBEC2, OMIM# 604797) to convert them to thymines. This essentially mirrors the process of bisulfite conversion, but uses enzymatic modifiers, rather than chemical, and as a result has several improvements. These enzymatic modifiers do not have the same cross reaction with the DNA backbone, and do not cause DNA degradation, allowing this process to be used on smaller amounts of starting material. Furthermore, since the end products of these processes are the same, analytical tools based on bisulfite conversion techniques can be used [13,26].

1.1.4 DNA METHYLATION; MACHINERY, MECHANISMS AND MODULATION

The most well studied and understood epigenetic mechanism is DNA methylation, that is, the addition of a methyl group to nucleotide residues within the DNA sequence. First demonstrated in bacterial models in 1925, the presence of methyl (CH_3) chemical groups on

nucleotide bases remained poorly understood for several decades [27]. In 1975, researchers characterized the biological function of methyl groups connected to the 5' carbon of the cytosine residue skeleton, identifying these 5-methyl-cytosine (5mC) residues as important epigenetic modifiers that influence gene expression [28,29]. DNA methylation in humans occurs almost exclusively on cytosine residues, in cytosine-guanine dinucleotides (aka CpG). These CpG dinucleotides are interspersed throughout the human genome, and include randomly interspersed genomic CpGs, and gene promoter associated CpG islands, where large clusters of CpGs occur in high density [4]. Most of the interspersed CpGs in the human genome are methylated, however a subset of CpGs, termed CpG islands, are predominately unmethylated and associated with transcriptionally activated euchromatin configurations [30]. In these CpG islands, located near promoters, or promoter regulatory regions of genes, hypermethylation corresponds to a heterochromatin state, which lessens the ability of proteins to interact with the proximal DNA sequence, thereby reducing and/or silencing the expression of nearby genes [31]. However, epigenomic structures or profiles display regional differences across the human genome. For example, most CpGs in the epigenome, with the key exception of CpG islands of gene promoters for housekeeping genes [32], are methylated during development. Once cells are differentiated into separate cell lines, specific genes are differentially methylated in appropriate cells, allowing for expression of genes that are necessary for the unique functions of the given cell line, and subsequently maintain their function in these specialized tissues.

This plastic system requires several key functions for effective regulation of gene expression, necessitating several families of important proteins to carry out epigenetic modification. Components that interact with the epigenome are named "epigenetic machinery" and can be sorted into several classes according to their function, namely, readers, writers and erasers [33,34,35].

Writers, such as the DNA methyltransferase (DNMT) family, are responsible for the addition of methyl tags to cytosine bases. In the process of DNA methylation, methyl (CH_3) groups are added onto the C5 position of the carbon skeleton of the cytosine residue, and is carried out by a family of proteins known as the DNA methyltransferases (DNMTs). The DNMT family has 3 key proteins, *DNMT1, DNMT3A*, and *DNMT3B* (OMIM#126375, OMIM# 602769, OMIM# 602900). DNMT1 was the first of the DNMT family to be researched and isolated, and

is referred to as a maintenance methyltransferase. This refers to its role transferring methyl groups to hemimethylated DNA, that is, DNA with methylation tags on one strand of the DNA helix, following parental DNA replication. To maintain proper methylation patterns in the daughter cells, *DNMT1* must methylate the opposing strand's CpG [30]. The presence of this enzyme is therefore integral during embryonic development, and genetic disruption results in embryonic lethality in mouse embryo models [30,36]. *DNMT3A* and *DNMT3B* perform methyltransferase activity without the need for the DNA strand to be hemimethylated (although they also methylate hemimethylated DNA with similar efficiency), and as such are termed "de novo" methyltransferases. Capable of modifying methylation patterns throughout the genome, these genes are integrally linked to developmental processes, particularly those related to genomic imprinting. Although they both perform similar activities, *DNMT3A* is expressed ubiquitously, while *DNMT3B* is expressed at its highest levels within the testis, thyroid, bone marrow and thymus [37]. Disruption of these DNMT enzymes is implicated in a number of neurodevelopmental conditions, including adult-onset autosomal dominant cerebellar ataxia, with deafness and narcolepsy (ADCADN) and Tatton-Brown-Rahman syndrome (TBRS). In ADCADN, mutations in *DNMT1* result in a degenerative condition characterized by hearing loss, narcolepsy, cataplexy and cerebellar ataxia, highlighting the myriad of CNS developmental processes affected by the loss of this maintenance methyltransferase [38].

Erasers, on the other hand, reverse this process, removing methyl groups to reduce compaction of chromatin and promote a more transcriptionally active state. One group of erasers, the Ten-Eleven Translocation (TET)   enzymes, can perform this activity, removing methyl groups through their cytosine dioxygenase activity. TET proteins convert 5mC to 5-hydroxymethylcytosine (5hmC), then 5-formylcytosine (5fC), and finally 5-carbocylcytosine (5caC), with each step in the process carried out by TET1, TET2, and TET3 respectively [39]. The methyl groups of 5mC derivatives are more readily removed and repaired using base excision repair (BER) processes, resulting in an unmethylated cytosine residue [40]. This interaction between DNA methylation associated proteins creates a plastic network of epigenetic fine tuning, allowing for modulation of gene regulation throughout an organism's life. Marks can be added and removed in different cells, or at particular developmental periods, fostering a great deal of transcriptional flexibility.

Finally, readers, such as *MECP2*, MBD1, MBD2, MBD3 and MBD4 (OMIM#300005, OMIM#156535, OMIM#603547,OMIM# 603573, OMIM#603574), can recognize the presence of methyl groups with specialized methyl binding domains within their structure [30,35,41], allowing the body to accurately assess and modulate methylation at key developmental stages, resulting in spatial and temporal differentiation through manipulation of transcription [42, 43]. *MECP2* binds to methylated CpG sites, and promotes interaction with several histone modifying enzymes, including histone deacetylase complexes (HDAC)[44, 45] as well as chromatin remodeling complexes, such as the cAMP response element binding protein (CREB)[46]. The effects of these interactions vary depending on the associated complex and genomic context, with transcriptional activation resulting in some instances [44, 46] while others promote transcriptional repression [44, 47]. Epigenetic reader proteins exhibit a wide range of effects on the regulation of epigenetic activity, and the downstream effects of their disruption lead to significant alterations in transcriptional activity. Disruption of *MECP2* results in Rett syndrome (RTT), while duplication of *MECP2* results in MECP2 duplication syndrome, each characterized with significant intellectual disability, autism and developmental regression, highlighting large scale developmental consequences [48].

Interaction with the micro and macroenvironmental exposures further increases the complexity of the epigenetic landscape, and this network of epigenetic machinery allows the body to react to these stimuli. For example, epigenetic mechanisms have been shown to respond to changes in diet, exercise, exposure to chemicals, and a myriad of other environmental stimuli [49]. Epigenetic mechanisms can therefore be described as a conduit between the environment, the genotype, and through their effects on the transcriptional landscape, the ensuing phenotype of the cell, playing a key role in cellular and organismal responses to adaptation to their environment.

My investigation of epigenetic machinery and its effects on the biology of human development focuses on mendelian neurodevelopmental disorders, (NDDs). By assessing the recurrent DNA methylation patterns, I aimed to develop accurate epigenetic biomarkers that can help in resolving ambiguous genetic findings and variable phenotypes, and provide further insights into the molecular pathophysiology of genetic NDD disorders.

1.2 MENDELIAN NEURODEVELOPMENTAL DISORDERS
1.2.1 DEFINING NEURODEVELOPMENTAL DISORDERS

The term neurodevelopmental disorder is a broad one, referring to a large cohort of complex, heterogeneous conditions that involve some form of disruption to brain development, resulting in highly variable difficulties in cognition, learning, speech, motor function and various sensory capabilities. Under the current criteria of Diagnostic and Statistical Manual of Mental Disorders (DSM-5), NDDs are defined as a group of conditions with onset in the development period, inducing deficits that produce impairments of functioning [50]. This necessarily broad definition results in grouping a large group of disorders that includes intellectual disability disorders, communication disorders, autism spectrum disorders and neurodevelopmental motor disorders.

1.2.2 CURRENT DIAGNOSTIC GUIDELINES

NDDs have a global prevalence of approximately 2-3% of the population. Application of whole exome sequencing (WES) and whole genome sequencing (WGS) has greatly improved the diagnostic rate in these conditions [6], however successful diagnosis rates remain relatively low. Current diagnostic guidelines use a combination of genetic information, and assessment by qualified clinicians to aid in the diagnosis of these conditions. Causes of NDDs are highly heterogeneous, including environmental factors, single gene mutations and chromosomal aberrations, with a large number of genes involved in the epigenetic machinery being implicated [51]. Nevertheless, the complexity of the observed phenotype, and a high degree of genetic variability, resulting in low penetrance and expressivity makes diagnosis of these conditions extremely difficult. Overlapping phenotypes, low penetrance, and the inability of current next generation sequencing (NGS) techniques to identify and interpret non-coding regions of the genome all contribute to lack of genetic diagnosis in an estimated two thirds of patients [52-56].

1.2.2.1 AMERICAN COLLEGE OF MEDICAL GENETICS GUIDELINES FOR VARIANT INTERPRETATION

NGS assessment, including targeted gene analysis and WES, provide a conclusive diagnosis in approximately 15-35% of cases in patients with rare neurodevelopmental disorders [52-56]. Although this is an improvement over previous gold standard chromosomal microarray (CMA) techniques, most patients with a suspected diagnosis do not reach a conclusive diagnosis for their condition. Current American College of Medical Genetics (ACMG) standards and guidelines for variant interpretation rely on various sequence interpretation tools and variant classifications, grouping variants detected on NGS platforms based on their perceived pathogenicity. This classification falls into one of five categories for any detected variant, pathogenic, likely pathogenic, uncertain significance, likely benign and benign, depending on the observed effects of the variant in question. These classifications are based on a number of criteria, including mode of inheritance (inherited or de novo), analysis of previously published literature including similar variants, in silico prediction tools, variant type (frameshift, missense, nonsense, splice-site etc.) and functional evidence of variant expression. Genetic testing in NDDs often results in the classification of variants as variants of unknown significance (VUS), with some estimations at upwards of 25% of diagnostic assessments resulting in VUS classification [57,58]. Although the ACMG guidelines can help interpret the effects of a variant following detection in WES, the ability to classify the downstream effects of epigenetic perturbation is low. Interpretations of these variants can be aided with population databases, and in silico prediction tools, but reclassification of genetic VUS in these patients, if possible, is directed by functional studies such as DNA methylation assessments.

1.2.2.2 CLINICAL FEATURES ASSESSMENT

Clinical features are an important source of information for the diagnosis of NDDs. Clinician's role in the process is indispensable, providing an assessment of intellectual disabilities, dysmorphisms, speech pathologies and the various other etiologies of NDDs. Methods of assessment include self-reported surveys from patients or their caretakers, dysmorphic features assessment carried out by clinicians, patient history review and a myriad of tests for various aspects of neurodevelopmental pathology, including fine and gross motor skills,

11

written and oral communication, socio cognitive impairment, eating habits, self-care and daily living proficiencies and many others [59,60]. One such example of gross motor assessment is the Gillette functional activity level, which measures the ability of patients to be mobile in various environments on a ten-point scale, providing an assessment of overall capacity for activity and independence [59,60]. Communication methods and frequency is often assessed, with the Communication Function Classification System (CFCS) providing a scale-based reporting method for patients and their caregivers on a patient's communication abilities [59, 61]. Additional assessments of academic abilities (reading, writing, etc.), and medical history are assessed through checklists in clinics, with some guidelines available through governmental agencies, such as the CDC guidelines on developmental milestones [62] Assessment and identification of the myriad of NDDs currently identified is a difficult and complex process, confounded by many factors, such as sex, age, and ethnicity. Facial gestalts vary greatly between disorders, with specific identifying features for disorders occurring rarely, as many of these NDDs involve highly overlapping phenotypes [35] requiring additional genomic findings to provide a definitive classification or diagnosis.

1.2.3 DISRUPTION OF THE EPIGENETIC MACHINERY THEORY OF NDDS

The term "Epigenetic Machinery" refers to the extensive network of proteins involved in modulating the key epigenetic functions throughout the genome, that in turn have considerable downstream effects on the expressed genome as they modulate expression [35]. As such, disruption of these proteins can result in significant changes to not only the protein in question, but the myriad of genes and protein networks that they interact with. For example, when mutations occur within one of the DNMTs, the effects are widespread, and complex, as it affects the entire cell's ability to effectively translate and transcribe its contents [63]. These epigenetic processes are integral to the normal functioning of many organisms, including humans. Targeted DNMT mutation of embryonic stem cells in mouse models were found to result in embryonic lethality when introduced to the germline [64]. When a mutation occurs in protein machinery related to reading, writing or erasing methylation marks, histone modifications or chromatin remodeling complexes, epigenetic and gene expression profiles are consequently disrupted. The EpiSign project has identified a large number of syndromes associated with the disruption of

12

these epigenetic machineries, highlighting their importance in the establishment of the healthy functioning epigenome. (See Figure 1-2)
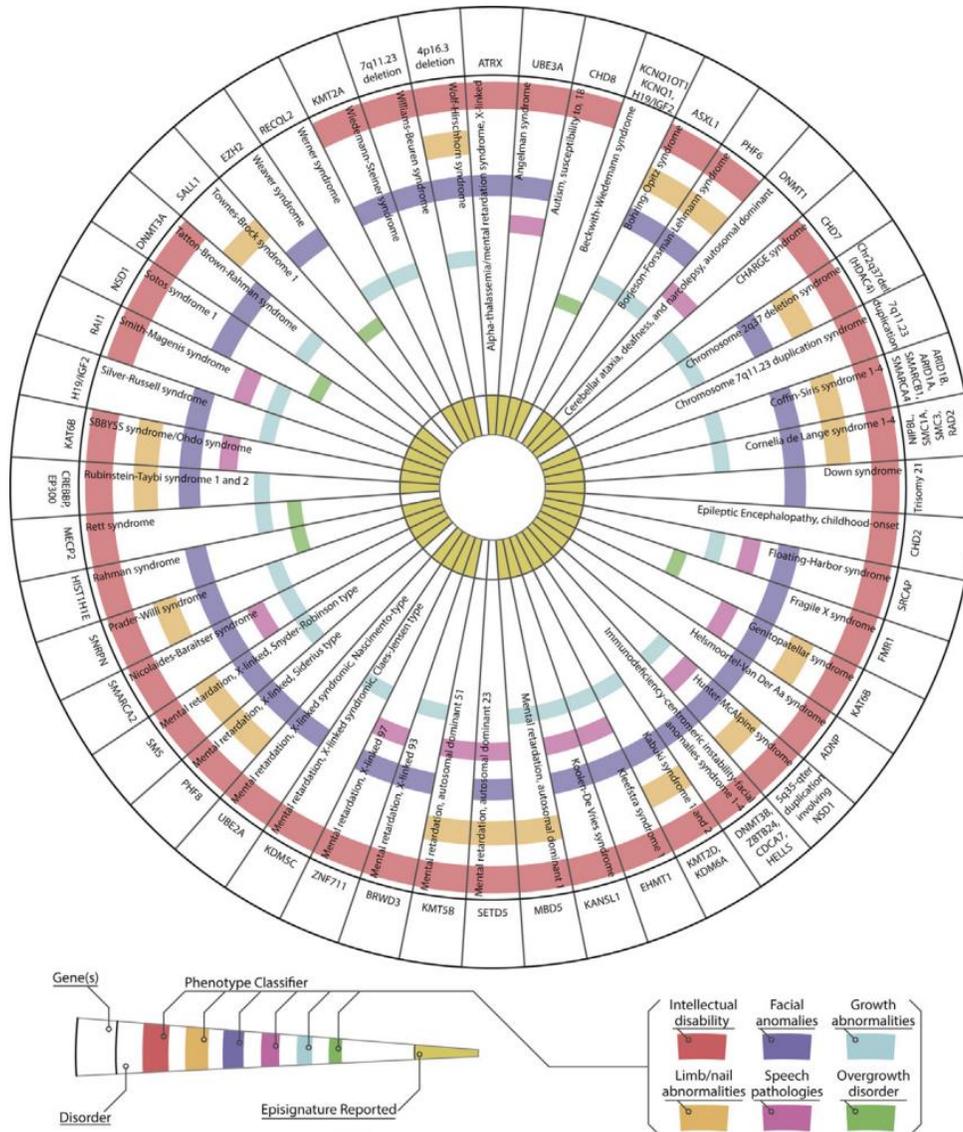
Figure 1-2. A comparison of the phenotypic overlap across conditions associated with disruption of epigenetic machinery. Disorders were categorized into 6 classifiers based on descriptions provided in their MIM profile: intellectual disability (n= 42), facial anomalies (n= 36), growth abnormalities (n=28), limb/nail abnormalities(n= 20), speech pathologies (n= 12), and overgrowth disorders (n= 5). Intellectual disability classifier was applied to disorders whose clinical phenotype contained key words such as intellectual disability, cognitive disability, and

mental retardation. Facial anomalies were assigned to disorders described to have malformations of facial features, including descriptions of abnormal nose, eye, eyelid, and mouth features. Growth abnormalities covered a description of deficiencies in development, including short stature, microcephaly, retardation of somatic development, and poor postnatal growth. Overgrowth disorders were assigned to disorders containing descriptors of acromegaly, macroorchidism, and gigantism. Limb/nail abnormalities category contained disorders associated with malformations of the appendicular skeleton, such as brachydactyly, and absence or hypoplasia of various limb features. The speech pathology category was assigned to disorders that were described with speech delay or absence of speech. Other phenotypic keywords, including epilepsy, cardiac malformations, immune dysfunction, dental malformation, narcolepsy/dementia and blood disorders, were also observed but were excluded from visualization because of low occurrence (n <4). (Figure produced by Gavin Riddolls, Guildenthaw Design.)

1.3 DNA METHYLATION EPISIGNATURES AS DIAGNOSTIC BIOMARKERS

Using genomic DNA methylation analysis in conditions that exhibit unique epigenetic signatures associated with specific genetic defects enables resolution of ambiguous findings and uncertain phenotypes in these patients [65]. Changes in DNA methylation are closely related to the variation in the expression of genes within the human genome and connected to a wide network of genes that directly or indirectly modulate the epigenome through chromatin remodeling, DNA methylation, histone modifications, or more indirect and complex molecular pathways [12, 66, 67, 68]. Variants that occur in early stages of development can therefore have widespread changes in gene expression, and be propagated through cell differentiation and tissue development, which allows for assessment of methylation perturbations in easily accessible tissues such as peripheral blood [12,65]. Peripheral blood samples extracted using standard techniques can be processed using bisulfite conversion techniques and assessed using microarray technology to determine methylation signal intensities at various CpGs throughout the genome. [63,65].

This methylation data is then processed and filtered using established protocols, and a set of age and sex matched controls are identified for comparative analysis. Methylation levels are

then modeled, accounting for blood cell proportions as confounding variables, and significantly differentially methylated probes (>5% methylation signal intensity change and p-value <0.05) are identified in the case cohort and used to create the "episignature". Probe selection parameters included the probe "score", the area under the receiver operating characteristic curve, and a probe-to-probe methylation correlation value. The probe score was derived from the absolute value of the mean methylation difference multiplied by the negative value of the log-transformed adjusted P values for each probe. This score identified the top 1000 probes with distinct methylation changes at statistically significant levels. The 1000 probe threshold is somewhat arbitrary but has proven effective at identifying a large enough set of probes that subsequent filtering steps result in episignatures comprising of ~150-500 probes [12]. This episignature size maintains the ability to provide a flexible biomarker that is informed by biological information from several genomic regions, while remaining computationally efficient. Increasing the number of probes could provide an increase in the number of informative regions identified, however the ensuing increase in computational complexity results in inefficient analysis times. These 1000 probes were then filtered by the AUC derived from ROC analysis of the level of methylation change at the specific probe, identifying probes with a high level of sensitivity and specificity for differentiating cases from controls. Finally, Pearson's correlation coefficients for the remaining probes were conducted, and highly correlated probes were removed, in a pseudo random manner, to avoid over reporting of probes which assay identical regions of the methylome. Additionally, due to the sex related differences in chromosomes, probes on the X and Y chromosomes are filtered from probe selection. Although this reduces the ability of the episignature pipeline to determine sex specific DNA methylation changes, the ensuing methylation patterns can be applied to a larger population, without the need to differentiate samples based on their sex. The resulting episignature is modeled and assessed using supervised machine learning methods to assess the sensitivity and specificity of the probeset to differentiate between case and control samples, as well as the large number of non-case syndrome classified cases, determining the specificity of this episignature for the syndrome of interest (See Figure 1-3).

Modeling methods include multidimensional scaling (MDS), principal component analysis (PCA) and hierarchical clustering heatmaps. Principal component analysis, or PCA, is used primarily in quality control steps of the statistical pipeline, to assess global changes in DNA methylation to identify possible batch effects or improperly labeled samples. This process is a

dimensionality reduction method that reduces the complexity of large datasets to allow for effective visualization of trends in the data, while maintaining as much of the information contained in the larger set as possible. PCA attempts to reduce the complexity of datasets through analysis and detection of orthogonal vectors which represent the variance in data, with the first principal component being one such orthogonal vector which represents the most variance, the second principal component representing the second most variance etc. [118]. Multidimensional scaling or MDS, has a similar goal, reduction of large datasets into more simple and easily visualized mediums, however it differs in how it achieves that goal. MDS uses the pairwise distances in Euclidean space, i.e. the relationships between distinct data points, to create a low dimensional model of the high dimensional dataset, and then uses a stress test to assess whether or not the new model maintains those pairwise distances well [119]. The result is a similar reduction in dimensions, providing a two- or three-dimensional dataset that can be easily visualized and assessed, however the focus of MDS is on maintaining the relationships between data points, while PCA focuses on the preservation of the dimensions through measuring their associated variance.

Where traditional genomic technologies fail to encompass the complexity of phenotypes in relation to the mutation observed in the epigenetic machinery, analysis and comparison of the methylation profile described as an episignature, allow for robust classification of the disorder. Rather than assigning the source of pathogenic condition to a given mutation or set of mutations, the episignature instead provides a genome wide view of the downstream epigenetic effects of a defect in a specific gene involved in the epigenetic machinery.

Figure 1-3. Overview of the Episignature statistical analysis pipeline. Adapted from https://www.diagenode.com/en/categories/bisulfite-conversion, https://en.wikipedia.org/wiki/Support-vector_machine, Haghshenas & Bhai 2020 [15] and Zweig & Campbell 1993 [112]

Investigations of episignatures in the context of inherited disorders has been ongoing for several years, with the earliest reports describing observed patterns of differentially methylated regions (DMRs) in trisomy 21 (Down syndrome) [69]. Expansion of this work has greatly increased in recent years with the current literature reporting 65 episignatures, [12]. These individual episignatures are unique to the disorder in question, which permits the use of a multiclass classification algorithm for concurrent analysis of each methylation pattern. These discoveries, in combination with the previously clinically validated imprinting and trinucleotide repeat disorders has introduced the first clinical genome-wide DNA methylation assay known as EpiSign [63, 70-74].

As previously mentioned, the low diagnostic rate of WES practices in NDDs, and large number of VUS variants indicates the need for improvement in the diagnostic criteria of these complex conditions. DNA methylation assessment can be a powerful functional assay for

resolving these complex cases, benefiting from the ability to identify the unique methylation patterns associated with these conditions from peripheral blood samples.

1.3.1 PERIPHERAL BLOOD AS DIAGNOSTIC TISSUE IN DNA METHYLATION ASSESSMENT

In the context of neurodevelopmental disorders, analysis of brain tissue would provide the clearest identification of biological processes associated with disorder related changes in DNA methylation, however, access to brain tissue is greatly limited, and often derived from postmortem samples. As such, alternate surrogate tissues such as peripheral blood have been explored in research, and eventually implemented for the purposes of our study. It is important to know then, that the comparison of epigenetic markers from peripheral blood is effective in predicting those seen in brain tissues. Several studies have found high concordance between tissues, with Horvath et al, 2013 [113] reporting a correlation of 0.85-0.91 with genome wide methylation patterns from publicly accessed datasets. Additional studies from Davies et al, 2014 [114] and Braun, 2019 [115] found similar results, with high levels of correlation between peripheral blood and brain tissue derived methylation patterns. Due to the emergence of mutations at fertilization, differences in DNA methylation are carried throughout the germline into the various specialized cells that go on to form the brain and other structures, leading to common patterns observed between tissues. Given the ease of access, and cost efficiency of peripheral blood, as well as it's demonstrated value as an effective mirror of the changes in DNA methylation observed at disease associated tissues such as brain tissue, the use of peripheral blood as a surrogate measure provides biologically relevant, and reliable assessment of DNA methylation patterns. Nevertheless, our statistical pipeline utilizes an additional method to account for the variability associated with a heterogenous cell mixture such as peripheral blood. Implementation of the Houseman method involves the estimation of blood cell proportions within a whole blood sample through analysis of cell lineage specific DNA methylation patterns [117]. As each of the various leukocyte lineages have been differentiated into specific roles from their hematopoietic stem cell originators, DNA methylation changes occur at specific loci to modulate gene expression, resulting in a number of differentially methylated regions that can be

used to define a specific cell lineage. These cell proportions can then be considered in the linear modeling of differentially methylated probes as confounding factors to reduce methylation "noise" caused by heterogenous cell mixtures. [117]

1.3.2 EPISIGNATURES IN THE CONTEXT OF INHERITED DISORDERS

Recent studies involving *DNMT1* mutations in a cohort of patients with ADCADN (MIM# 604121), demonstrates the effects of disruption of the epigenetic machinery on the genome [75], Peripheral blood samples from patients with *DNMT1* mutations were shown to have significant differences in genomic methylation when compared to unaffected controls, primarily increased in areas that normally remain unmethylated throughout development [75]. Such hypermethylation of gene promoters can be associated with disruption of gene expression and likely plays a part in the observed pathophysiology of this disorder, namely large-scale multi-organ disruption, particularly in brain tissue. A large and continually growing number of human neurodevelopmental disorders are caused by mutations in the epigenetic machinery genes, each presenting significant difficulties in their diagnosis. Overlapping phenotypes, confounding results from current genetic sequencing techniques, and a limited ability to detect and interpret non-coding genomic regions all contribute to a particularly challenging diagnostic landscape in these conditions.

1.3.3 EPISIGNS AS DIAGNOSTIC MARKERS IN COMPLEX DISORDERS

When variants classified as VUS are matched with an episignature, the variant can be reclassified into likely pathogenic (matching the epigenetic signature). In a recent study, out of 36 patients with VUS variants in a cohort of *KMT2D*, 7 (19%) of the patients were predicted to have a methylation profile matching the episignature for Kabuki syndrome (MIM# 147920), whereas the remaining 29 samples that matched the control cohort methylation profile were predicted to carry likely benign variants [76]. In the same study, 8 out of 16 (50%) of patients with VUS in *NSD1* (OMIM #606681), were predicted as having Sotos syndrome (MIM

#117550) [63]. Assessment of Coffin Siris (CSS; MIM# 135900, 614607) and Nicolaides Baraitser (NCBRS; MIM# 601358) syndromes classified 4 of 18 (22%) of patients with VUS, in genes encoding subunits of the Switch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeling complex, as good clinical matches [77]. Two additional cohorts of patients with VUS in genes associated with the epigenetic machinery were assessed with the Episign multiclass classification algorithm and matched a signature with a predicted pathogenic phenotype in 17/44 (39%) [71] and one out of 9 (11%) patients [70]. As such, episignature assessment is not only enabling reclassification and interpretation of genetic variants, but also expanding the knowledge of the types of genetic variants that can cause genetic disorders. Until recently, variant reports for the *ADNP* gene, associated with Helsmoortel Van der Aa syndrome (HVDAS), (MIM #615873), have been restricted to truncating variants [78,79]. By matching the episignature defined by patients with truncating LOF mutations, 2 unrelated patients with different *ADNP* missense mutations were shown to be affected by HVDAS [80]. Characterizing VUS cases is one way in which episignature analysis is increasing the molecular diagnostic yield. Another approach is to systematically screen patients with NDDs that show negative genomic findings. Genomic analyses, including copy number variation (CNV) arrays and exome sequencing have limitations, including identification of balanced translocations [81,82], allele phasing, mapping problems due to guanine cytosine bias, repetitive elements, and homologous sequences, and are normally restricted to the assessment of coding regions with minimal coverage of intronic or regulatory elements [83]. Methylation profiling can be used as a biomarker to assess patients with negative genetic findings. Methylation profiling can also help to stratify NDDs with phenotypic overlap when diagnosis based on clinical features is difficult. For example, in the case of CHARGE syndrome (MIM#214800) caused by variants in the *CHD7* gene includes Kabuki syndrome (MIM#147920, 300867) as one of the differential diagnoses [84,85]. In this case, matching a given patient's methylation patterns to one of the episignatures associated with these conditions can provide functional evidence supporting diagnosis. In one study, a cohort of 51 patients with phenotypes suggestive of CHARGE syndrome, but lacking a definitive molecular confirmation underwent genome wide DNA methylation analysis. Epigenomic signatures were consistent with CHARGE in 23 patients; 27 patients were ruled out for both signatures, and a final patient had an episignature consistent with Kabuki syndrome [63]. In this case, DNA methylation analysis provided a novel avenue for diagnosis, providing effective classification when traditional

practices of genomic investigation and clinical classification were insufficient. A larger cohort of 965 subjects with a spectrum of neurodevelopmental delays and congenital anomalies, but negative for routine genetic investigations by CMA, and in some cases, targeted gene panels or WES, underwent genome wide methylation profiling. Of the 34 episignatures screened, 16 different conditions were matched across 24 unique samples [63,65], further highlighting the effectiveness and utility of episignature analysis.

1.3.4 EPISIGNATURES AS PHENOTYPIC BIOMARKERS IN NDDS

Disorders involving either direct or indirect perturbation of proteins that regulate epigenetic machinery display significant phenotypic overlap with one another, which may be associated with the downstream effects of altered gene expression that arise when epigenetic patterns are disrupted [86]. The differentially methylated regions in this group of NDDs are genome wide, and can range in numbers from hundreds to thousands with both hypermethylation and hypomethylation changes observed. Changes are interspersed both within and outside gene coding regions, affecting active protein sequences, upstream regulatory elements or deep intronic regions. The extent of change can vary as well, with DMRs that cover areas of single CpGs, to entire CpG islands. The pattern of DMRs within one condition is highly reproducible, which permits mapping of their profile for use as potential diagnostic biomarkers alongside genome wide methylation signatures [63, 70,75, 84-96]. There are shared DMRs across the epigenetic machinery subclass of neurodevelopmental disorders, but typically, less than 10% are linked to more than one condition, and machine learning approaches have been used to accurately classify one condition from another as well as controls using this DMR focused approach. Episignature mapping involves using a training cohort of patients with known pathogenic variants to provide evidence for feature selection and model training, demonstrating accuracies of up to 99.9% [63, 65,70]. Once mapped, an episignature can be validated with a testing cohort of similarly affected individuals. The high accuracy demonstrated by the episignature is further expressed through the high level of sensitivity with these validation cohorts, with over 99% sensitivity in classifying validation cohorts into the correct category, and a high level of specificity, with unaffected control cohorts demonstrating 100% correct classification as control subjects. [63, 65,70]. The importance of accurate diagnosis in these closely related conditions is necessary for creating an appropriate and direct treatment plan for these patients. It has been shown that genetic findings

have changed clinical management in up to 55% of individuals with multiple congenital anomalies, developmental delay, intellectual disability and autism spectrum disorders [97].

1.3.5 INVESTIGATIONS OF PHENOTYPIC AND GENOTYPIC OVERLAP

Episignatures can distinguish conditions from one another, but are also providing insight into the phenotypic variability observed within conditions, or perhaps more specifically, the range of phenotypes observed from variants within the same gene. For example, HVDAS is a common cause of ASD and intellectual disability, with patients with variants in the associated *ADNP* gene for this disorder comprising around 1.7% of ASD cases [78, 79, 98, 99]. Interestingly, a large amount of phenotypic variability is seen within these cohorts however, with a high amount of variable expressivity [78]. Variants in the causative *ADNP* gene have been shown to cluster into two distinct episignatures dependent on the location of the variant, translating to a spectrum of downstream gene expression effects and ultimately offering a possible explanation of the varied phenotypic range observed in HVDAS [80]. In a similar case, variants in *SMARCA2* have historically been associated with NCBRS, but methylation profiling for a subset of *SMARCA2* variants has recently identified an episignature that is predominantly of an opposite pattern to those observed in NCBRS patients, and linked to a divergent phenotype [100].

Although gene level information is often associated with a definitive clinical diagnosis, specific examples of *ADNP* and *SMARCA2* show that nucleotide level variant features, such as location, variant type, or mechanism of action may be required for reaching an accurate diagnosis. Furthermore, in addition to defining a distancing phenotype, variant location can be a factor in milder or less expressive variation of the phenotype associated with these conditions. Revisiting the *SMARCA2* gene as a causative factor in NCBRS, most pathogenic variants map to the ATPase/c-terminal helicase domain of the protein [101]. Patients with variants in *SMARCA2* distal to the helicase domain present with milder neurodevelopmental and atypical features compared with typical NCBRS patients. Downstream methylation effects have been shown to represent an intermediate profile that matches neither NCBRS, nor unaffected controls, but overlaps with a significant number of control DMRs, nearly 50% [87]. This intermediate signature further indicates the possibility of a distinct phenotype, or the possibility of a

"signature scale" corresponding to different levels of methylation perturbation that could explain the variable expressivity observed in some conditions.

Episignatures have begun to guide us towards a more accurate method of determining a patient's diagnosis, while providing insight into the inherent biology that dictates their ensuing phenotype and shedding light on the variable expressivity seen within cohorts that share the same gene of interest. Additionally, DNA methylation provides evidence of overlap between disorders that share molecular interactions, providing common episignatures across multiple genes. BAF complex in CSS and NCBRS syndromes, cohesin complex in Cornelia DeLange Syndrome (CDLS) (MIM# 122470, 300590,610759, 614701,300882) and Cohen Gibson syndrome (MIM #601573), are all examples whereby variants in multiple distinct genes map to common methylation patterns detectable by the episignature process [65].

1.4 MACHINE LEARNING IN EPIGENOMICS

The use of machine learning approaches in bioinformatics has expanded greatly in recent years, making use of "machines", aka algorithms or modeling techniques that employ statistical approaches to find relationships between complex sets of data, and evaluates them, providing interpolated or extrapolated predictions of variables, or classifications [102]. In this way, these machines "learn" continuously improving their ability to provide predictions in each iteration of their use, identifying trends in complex datasets. Machine learning is particularly helpful in the context of complex conditions, such as NDDs, where the classification of a given observation, in this context, the identity of a given NDD, is based on a complex set of characteristics, whose interactions are multifactorial and interconnected, making them difficult to interpret via traditional data analysis [35,86]. In this thesis, I made use of one such machine learning method, known as Support Vector Machines (SVM) based on our previously described methods [63]. SVM can analyze groups of observations, and attempts to find a hyperplane, a boundary within the observed dataset that can separate observations into distinct groups based on their characteristics. The hyperplane is derived through providing observations, in this context

representing methylation profiles form patient samples, to the machine learning algorithm, and mapping them based on similarity of their characteristics, within n-dimensional space, where n is the number of features used to differentiate observations, in this context, DNA methylation probes. Once mapped, the SVM uses support vectors, the most proximal samples within the mapped space to create a hyperplane which splits the sets of observations into the desired categories, i.e., cases vs controls, while maintaining the largest margin between support vectors. This method, also known as a maximal margin classifier, creates a decision boundary, wherein samples on one side of the boundary, represented by a hyperplane spanning the # of desired features in the n dimensional space, are classified as one group, for example controls, while those found on the other side of the boundary are identified as cases. This classifier can then be expanded beyond the dataset used to train it, providing a method for classifying new observations based on their relation to the derived hyperplane with new testing samples [103-107]. This classification method is a binary one, and needs to be modified to provide a continuous scale of the similarity between observations, allowing for more nuanced and detailed results, as such, I then employed Platt's scaling method, which converts the classification into a probability distribution over the classes [108]. Probabilities of being assigned into a given class, either resembling the methylation profile of the disorder training cohort, or not, can provide a more nuanced understanding of different sample types, and allows us to identify samples that are similar to the methylation profile on a continuous scale, rather than a binary one. However, probabilities can differ greatly across different training sets, meaning that without an effective training method, the modeled probabilities are unlikely to reflect empirical probabilities observed in new samples. Platt's scaling method was created specifically for support vector machines to provide a way to create classification probabilities that are trained on a subset of samples to reduce bias and allow for effective classification of new samples. This procedure uses a training set of samples to generate probabilities, which are then calibrated with a logistic regression to the observed scores derived from the SVM classifier. The calibrated probability curve can then be used to evaluate the effectiveness of this calibration, providing a testable model of probabilities that reflect the training group it was based on. For our classifier, we split the training cohort into groups, and provide them as training, calibrating and testing in several iterations, such that each group of samples is used in the training, testing and calibrating categories at least once. Once complete, these scores are then combined into a final MVP score which reflects each iteration,

and provides a highly robust score which reflects each sample's similarity to the methylation profile of the given training group. In this way, we can provide not only a classification into a given disease group following analysis of a selected probeset, but a score for each sample in terms of its similarity to the methylation profile observed in the training group, allowing for substratification of case groups and further avenues of research to identify subgroups with differentiated methylation characteristics within a given disease group.

1.5 CHALLENGES IN DNA METHYLATION ANALYSIS

Episignature detection works on the model that an inherited pathogenic variant in a gene is associated with a unique methylation profile. The assumption with this model is that the inherited changes originated early in development, and an episignature observed in our diagnostic tissue, peripheral blood, is also present in other tissues. The evidence to support this theory is limited, but it has been demonstrated in Sotos syndrome whereby fibroblast samples of affected patients which underwent DNA methylation assessment showed similar patterns to the episignature derived from the peripheral blood tissue of those same patients [92]. Blood is a readily available surrogate tissue that can be used as a biomarker for direct testing when the critical tissues, such as brain tissue in the context of neurodevelopmental disorders , are unavailable. The use of peripheral blood can be limited by tissue mosaicism when the affected tissues for the disorder in question does not include blood, such as in the case of Beckwith Wiedemann syndrome [109], or if the level of mosaicism produces an episignature that is below the threshold of differences of the positive population. This outcome is predicted by the milder profiles observed by carriers of recessive conditions [110] or of carriers with intermediate patterns when compared to affected individuals [87]. In the case of Claes Jensen syndrome (MIM # 300534) caused by variants on the X chromosome gene KDM5C [111], healthy female carriers were shown to exhibit one such intermediate pattern. This intermediate pattern showed some similarities to the affected male patients, as well as with the unaffected control cohorts [111]. The exact limit of detection of mosaicism for episignature analysis will likely be specific to each new signature, and will require systematic investigation to fully understand.

Episignature work has thus far focused primarily on cohorts involving disruption of genes of the epigenetic machinery. The utility of these episignatures is dependent not only on the robustness of methylation changes but also on the diversity of variants and number of positive vases in the cohort. Specific genetic neurodevelopmental disorders are rare, and collection of appropriate numbers of cases or of cases with a diverse variant profile is not always feasible. The type of variant and location of it within genes and their associated domains have been shown to impact the episignature profile [64, 80, 87] and therefore, episignatures based on variants of the same type or within the same protein domain offer challenges in the interpretation of results when alternate variants or regions of the gene are introduced for investigation.

Each of these challenges I have addressed so far are controlled through several quality control steps and statistical processes. These steps are an important part of our statistical pipeline. In the case of batch effect, we assess drift between different batches through the use of a PCA plot which displays the overall methylation profile of all probes assayed upon the EPIC array. This plot displays a low dimension presentation of the trends in DNA methylation across the genome, and allows for identification of particular batches that differ significantly in their global DNA methylation changes. PCA plots which display a dense, homogenous mixture of samples indicate the overall change in methylation between batches is small, while scattered PCA batch plots may flag certain batches for regeneration or further assessment. Overall, the methylation differences between batches are kept low, due to many of our batches being run on site at LHSC, or in collaborating labs which also maintain a high level of adherence to established protocols. Other steps that are key to generating new samples on the classifier include checking the number of failed probes, ensuring that the methylation predicted age (via Horvath clock predictions [113,116]), and predicted sex match the information provided for each patient, and ensuring that the provided sample is from peripheral blood and not some other surrogate tissue, such as fibroblast.

Although DNA methylation signatures or EpiSigns have now evolved beyond scientific concepts to the use in diagnosis of patients with a growing number of neurodevelopmental disorders, much more work remains before this technology can reach its full potential. Collecting cohorts of patients with each of these conditions will take effort and international collaboration for years to come. As technology continues to evolve, it can be expected that targeted approaches

26

such as methylation microarrays may be expanded by the more comprehensive genomic approaches, such as bisulfite genome sequencing. In that context, reference databases will need to evolve to account for the growing data complexity, which may provide an opportunity to reassess conditions with no existing EpiSigns based on methylation microarray analysis. Mapping DNA methylation profiles based on other tissue types, such as buccal swabs, fibroblasts and so forth will provide further understanding of the biology and underpinning mechanisms, as well as providing additional opportunities for clinical utilization of this technology.

1.6 STUDY RATIONALE, HYPOTHESIS AND OBJECTIVES
1.6.1 CONCLUDING REMARKS

The analysis of DNA methylation episignatures provides an important level of epigenetic information that can inform not only more effective diagnosis of complex NDDs, but is inherently implicated in their underlying biology, elucidating mechanisms, pathways and shared characteristics of various conditions. Nevertheless, many questions remain to be explored. Are there readily identifiable episignatures for every genetic disorder? Can the episignature differentiate between syndromes caused by paralogous genes? Can the shared phenotype observed in disorders with distinct genomic origins be correlated to overlapping changes in DNA methylation? Although the identification of effective diagnostic biomarkers in the form of episignatures has been well described, my work expands on the current body of knowledge by assessing genomic and phenotypic correlation of episignatures in Mendelian NDDs.

1.6.2 HYPOTHESIS

I hypothesized that DNA methylation episignatures can be used to provide sensitive and specific classifications in neurodevelopmental disorders, and further stratification of these episignatures can identify key epigenetic patterns that relate to the phenotypic and genotypic variations seen in patients with these disorders.

1.6.3 OBJECTIVES

To achieve my goals of testing this hypothesis, I identified several specific aims that relate to each chapter of my thesis, based on my work over the past several years. 1) I sought to identify syndrome specific episignatures, relating Yin Yang 1 (*YY1,* OMIM#600013) transcription factor variants associated with Gabriele De Vries syndrome (GADEVS, OMIM#617557) , providing a baseline interpretation of episignature development. 2) In an attempt to resolve the effects of gene domain specific variants on the ensuing methylation profile, I investigated variants within the lysine demethylase 2B gene (*KDM2B*, OMIM#609078) , and then stratified the cohort based on specific domains affected by variants within this disease cohort, focusing particularly on those which disrupted the CxxC DNA binding domain within the *KDM2B* sequence. 3) Next, I sought to identify whether or not paralogous genes, Lysine acetyltransferase 6A (*KAT6A*, OMIM#601408) and lysine acetyltransferase 6B (*KAT6B*, OMIM#605880) , with significant shared genetic character and function could be effectively differentiated from one another on the basis of the DNA methylation profiles derived from *KAT6A* patients. 4) Finally, through the analysis of a cohort of two distinct molecular entities, the activity dependent neuroprotector homeobox protein (*ADNP*, OMIM#611386) and the SWI/SNF-related matrix associated actin dependent regulator of chromatin, subfamily A member 2 (*SMARCA2*, OMIM#600014) which both exhibited a particular phenotype.

This work is by no means a comprehensive list of possible avenues of episignature assessment, but provides an extensive expansion of the ways in which researchers can combine genetic, phenotypic and epigenetic evidence that is rapidly rising in prominence and accessibility across the world to inform research. By using DNA methylation assessment, in tandem with traditional methods of genetic and clinical diagnosis, we can further elucidate how changes in the epigenome relate to the specific diagnosis of complex conditions, and the molecular etiologies of disease.

REFERENCES

1. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. *Genes & development.* Apr 1 2009;23(7):781-783.
2. Schenkel LC, Rodenhiser DI, Ainsworth PJ, Pare G, Sadikovic B. DNA methylation analysis in constitutional disorders: Clinical implications of the epigenome. *Critical reviews in clinical laboratory sciences.* 2016;53(3):147-165.
3. Kim JH, Lee JH, Lee IS, Lee SB, Cho KS. Histone Lysine Methylation and Neurodevelopmental Disorders. *International journal of molecular sciences.* Jun 30 2017;18(7).
4. Schenkel LC, Rodenhiser D, Siu V, McCready E, Ainsworth P, Sadikovic B. Constitutional Epi/Genetic Conditions: Genetic, Epigenetic, and Environmental Factors. *Journal of pediatric genetics.* Mar 2017;6(1):30-41.
5. Timpano S, Picketts DJ. Neurodevelopmental Disorders Caused by Defective Chromatin Remodeling: Phenotypic Complexity Is Highlighted by a Review of ATRX Function. *Front Genet*. 2020;11:885. doi:10.3389/fgene.2020.00885
6. Bowman GD, Poirier MG. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chem Rev*. 2015;115(6):2274-2295. doi:10.1021/cr500350x
7. Ho L, Crabtree GR. Chromatin remodelling during development. *Nature*. 2010;463(7280):474-484. doi:10.1038/nature08911
8. Sokpor G, Xie Y, Rosenbusch J, Tuoc T. Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders. *Front Mol Neurosci*. 2017;10:243. doi:10.3389/fnmol.2017.00243
9. Narayanan R, Pirouz M, Kerimoglu C, et al. Loss of BAF (mSWI/SNF) Complexes Causes Global Transcriptional and Chromatin State Changes in Forebrain Development. *Cell Reports*. 2015;13(9):1842-1854. doi:10.1016/j.celrep.2015.10.046
10. Nguyen H, Sokpor G, Pham L, et al. Epigenetic regulation by BAF (mSWI/SNF) chromatin remodeling complexes is indispensable for embryonic development. *Cell Cycle*. 2016;15(10):1317-1324. doi:10.1080/15384101.2016.1160984
11. Alfert A, Moreno N, Kerl K. The BAF complex in development and disease. *Epigenetics Chromatin*. 2019;12(1):19. doi:10.1186/s13072-019-0264-y
12. Levy MA, McConkey H, Kerkhof J, et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. *HGGADVANCE*. 2022;3(1). doi:10.1016/j.xhgg.2021.100075
13. Morrison J, Koeman JM, Johnson BK, et al. Evaluation of whole-genome DNA methylation sequencing library preparation protocols. *Epigenetics Chromatin*. 2021;14(1):28. doi:10.1186/s13072-021-00401-y
14. Li Y, Tollefsbol TO. DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods Mol Biol*. 2011;791:11-21. doi:10.1007/978-1-61779-316-5_2
15. Feng S, Zhong Z, Wang M, Jacobsen SE. Efficient and accurate determination of genome-wide DNA methylation patterns in Arabidopsis thaliana with enzymatic methyl sequencing. *Epigenetics Chromatin*. 2020;13(1):42. doi:10.1186/s13072-020-00361-9
16. Godler DE, Amor DJ. DNA methylation analysis for screening and diagnostic testing in neurodevelopmental disorders. *Essays Biochem*. 2019;63(6):785-795. doi:10.1042/EBC20190056
17. Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462(7271):315-322. doi:10.1038/nature08514

18. Cokus SJ, Feng S, Zhang X, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*. 2008;452(7184):215-219. doi:10.1038/nature06745

19. Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A*. 1992;89(5):1827-1831. doi:10.1073/pnas.89.5.1827

20. Raizis AM, Schmitt F, Jost JP. A bisulfite method of 5-methylcytosine mapping that minimizes template degradation. *Anal Biochem*. 1995;226(1):161-166. doi:10.1006/abio.1995.1204

21. Tanaka K, Okamoto A. Degradation of DNA by bisulfite treatment. *Bioorg Med Chem Lett*. 2007;17(7):1912-1915. doi:10.1016/j.bmcl.2007.01.040

22. Sun, Zhifu et al. "Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis." Epigenomics vol. 7,5 (2015): 813-28. doi:10.2217/epi.15.21

23. Moran, Sebastian et al. "Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences." Epigenomics vol. 8,3 (2016): 389-99. doi:10.2217/epi.15.114

24. Heiss, Jonathan A et al. "Battle of epigenetic proportions: comparing Illumina's EPIC methylation microarrays and TruSeq targeted bisulfite sequencing." Epigenetics vol. 15,1-2 (2020): 174-182. doi:10.1080/15592294.2019.1656159

25. Li, Shizhao, and Trygve O Tollefsbol. "DNA methylation methods: Global DNA methylation and methylomic analyses." Methods (San Diego, Calif.) vol. 187 (2021): 28-43. doi:10.1016/j.ymeth.2020.10.002

26. Vaisvila R, Ponnaluri VKC, Sun Z, et al. EM-seq: Detection of DNA Methylation at Single Base Resolution from Picograms of DNA. Published online May 16, 2020:2019.12.20.884692. doi:10.1101/2019.12.20.884692

27. Wyatt GR. Occurrence of 5-Methyl-Cytosine in Nucleic Acids. *Nature*. 1950;166(4214):237-238. doi:10.1038/166237b0

28. Riggs AD. X chromosome inactivation, differentiation, and DNA methylation revisited, with a tribute to Susumu Ohno. *Cytogenet Genome Res*. 2002;99(1-4):17-24. doi:10.1159/000071569

29. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Science*. 1975;187(4173):226-232.

30. Tost J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Molecular biotechnology*. Jan 2010;44(1):71-81.

31. Hernando-Herraez I, Garcia-Perez R, Sharp AJ, Marques-Bonet T. DNA Methylation: Insights into Human Evolution. *PLoS genetics*. Dec 2015;11(12):e1005661.

32. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013;29(10):569-574. doi:10.1016/j.tig.2013.05.010

33. Jung G, Hernandez-Illan E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nature reviews. Gastroenterology & hepatology*. Feb 2020;17(2):111-130.

34. Linner A, Almgren M. Epigenetic programming-The important first 1000 days. *Acta paediatrica*. Mar 2020;109(3):443-452.

35. Bjornsson HT. The Mendelian disorders of the epigenetic machinery. *Genome research*. Oct 2015;25(10):1473-1481.

36. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*. 1992;69(6):915-926. doi:10.1016/0092-8674(92)90611-f

37.   Yanagisawa Y, Ito E, Yuasa Y, Maruyama K. The human DNA methyltransferases DNMT3A and DNMT3B have two types of promoters with different CpG contents. *Biochim Biophys Acta*. 2002;1577(3):457-465. doi:10.1016/s0167-4781(02)00482-7

38.   Kernohan KD, Cigana Schenkel L, Huang L, et al. Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. Clin Epigenetics. 2016;8:91. doi:10.1186/s13148-016-0254-x

39.   Ito S, Shen L, Dai Q, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 2011;333(6047):1300-1303. doi:10.1126/science.1210597

40.   Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*. 2014;156(1-2):45-68. doi:10.1016/j.cell.2013.12.019

41.   Swanberg SE, Nagarajan RP, Peddada S, Yasui DH, LaSalle JM. Reciprocal co-regulation of EGR2 and MECP2 is disrupted in Rett syndrome and autism. *Hum Mol Genet*. 2009;18(3):525-534. doi:10.1093/hmg/ddn380

42.   Mullegama SV, Elsea SH. Clinical and Molecular Aspects of MBD5-Associated Neurodevelopmental Disorder (MAND). *European journal of human genetics : EJHG.* Aug 2016;24(9):1235-1243.

43.   Liyanage VRB, Rastegar M. Rett Syndrome and MeCP2. Neuromolecular Med. 2014;16(2):231-264. doi:10.1007/s12017-014-8295-9

44.   Robinson HA, Pozzo-Miller L. The role of MeCP2 in learning and memory. Learn Mem. 2019;26(9):343-350. doi:10.1101/lm.048876.118

45.   Maunakea AK, Chepelev I, Cui K, Zhao K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res*. 2013;23(11):1256-1269. doi:10.1038/cr.2013.110

46.   Chahrour M, Jung SY, Shaw C, et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*. 2008;320(5880):1224-1229. doi:10.1126/science.1153252

47.   Long SW, Ooi JYY, Yau PM, Jones PL. A brain-derived MeCP2 complex supports a role for MeCP2 in RNA processing. *Biosci Rep*. 2011;31(5):333-343. doi:10.1042/BSR20100124

48.   Lombardi LM, Baker SA, Zoghbi HY. MECP2 disorders: from the clinic to mice and back. *J Clin Invest*. 2015;125(8):2914-2923. doi:10.1172/JCI78167

49.   Tuscher JJ, Day JJ. Multigenerational epigenetic inheritance: One step forward, two generations back. *Neurobiology of disease.* Dec 2019;132:104591.

50.   Morris-Rosendahl DJ, Crocq MA. Neurodevelopmental disorders-the history and future of a diagnostic concept. *Dialogues Clin Neurosci*. 2020;22(1):65-72. doi:10.31887/DCNS.2020.22.1/macrocq

51.   Iwase S, Bérubé NG, Zhou Z, et al. Epigenetic Etiology of Intellectual Disability. *J Neurosci*. 2017;37(45):10773-10782. doi:10.1523/JNEUROSCI.1840-17.2017

52.   Lindy AS, Stosser MB, Butler E, et al. Diagnostic outcomes for genetic testing of 70 genes in 8565 patients with epilepsy and neurodevelopmental disorders. *Epilepsia.* May 2018;59(5):1062-1071.

53   Grozeva D, Carss K, Spasic-Boskovic O, et al. Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Human mutation.* Dec 2015;36(12):1197-1204.

54.   Reuter MS, Tawamie H, Buchert R, et al. Diagnostic Yield and Novel Candidate Genes by Exome Sequencing in 152 Consanguineous Families With Neurodevelopmental Disorders. *JAMA psychiatry.* Mar 1 2017;74(3):293-299.

55.   Evers C, Staufner C, Granzow M, et al. Impact of clinical exomes in neurodevelopmental and neurometabolic disorders. *Molecular genetics and metabolism.* Aug 2017;121(4):297-307.

56.   Redin C, Gerard B, Lauer J, et al. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of medical genetics.* Nov 2014;51(11):724-736.

57.   Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *Jama.* Nov 12 2014;312(18):1880-1887.

58.   Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *European journal of human genetics : EJHG.* Feb 2017;25(2):176-182.

59.   Van Nuland A, Reddy T, Quassem F, Vassalli JD, Berg AT. PACS1-Neurodevelopmental disorder: clinical features and trial readiness. Orphanet J Rare Dis. 2021;16(1):386. doi:10.1186/s13023-021-02001-1

60.   Ammann-Reiffer C, Bastiaenen CHG, Van Hedel HJA. Measuring change in gait performance of children with motor disorders: assessing the Functional Mobility Scale and the Gillette Functional Assessment Questionnaire walking scale. *Dev Med Child Neurol*. 2019;61(6):717-724. doi:10.1111/dmcn.14071

61.   Paulson A, Vargus-Adams J. Overview of Four Functional Classification Systems Commonly Used in Cerebral Palsy. *Children (Basel)*. 2017;4(4):E30. doi:10.3390/children4040030

62.   National Center on Birth Defects and Developmental Disabilities Division of Human Development and Disability strategic plan FY 2021-2025. Accessed March 13, 2022. https://stacks.cdc.gov/view/cdc/97731

63.   Aref-Eshghi E, Rodenhiser DI, Schenkel LC, et al. Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. *American journal of human genetics.* Jan 4 2018;102(1):156-174.

64.   Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* Jun 12 1992;69(6):915-926.

65.   Aref-Eshghi E, Bend EG, Colaiacovo S, et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. *American journal of human genetics.* Apr 4 2019;104(4):685-700.

66.   Janssen SM, Lorincz MC. Interplay between chromatin marks in development and disease. *Nat Rev Genet*. 2022;23(3):137-153. doi:10.1038/s41576-021-00416-x

67.   Sadikovic B, Aref-Eshghi E, Levy MA, Rodenhiser D. DNA methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. *Epigenomics.* Apr 2019;11(5):563-575.

68.   Haghshenas S, Levy MA, Kerkhof J, et al. Detection of a DNA Methylation Signature for the Intellectual Developmental Disorder, X-Linked, Syndromic, Armfield Type. *Int J Mol Sci*. 2021;22(3):1111. doi:10.3390/ijms22031111

69.   Kerkel K, Schupf N, Hatta K, et al. Altered DNA methylation in leukocytes with trisomy 21. *PLoS genetics.* Nov 18 2010;6(11):e1001212.

70.   Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *American journal of human genetics.* Feb 26 2020.

71.   Schenkel LC, Schwartz C, Skinner C, et al. Clinical Validation of Fragile X Syndrome Screening by DNA Methylation Array. *The Journal of molecular diagnostics : JMD.* Nov 2016;18(6):834-841.

72.   Prickett AR, Ishida M, Bohm S, et al. Genome-wide methylation analysis in Silver-Russell syndrome patients. *Human genetics.* Mar 2015;134(3):317-332.

73.   Wu D, Gong C, Su C. Genome-wide analysis of differential DNA methylation in Silver-Russell syndrome. *Science China. Life sciences.* Jul 2017;60(7):692-699.

74.   Aref-Eshghi E, Schenkel LC, Lin H, et al. Clinical Validation of a Genome-Wide DNA Methylation Assay for Molecular Diagnosis of Imprinting Disorders. *The Journal of molecular diagnostics : JMD.* Nov 2017;19(6):848-856.

75.   Kernohan KD, Cigana Schenkel L, Huang L, et al. Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. *Clinical epigenetics.* 2016;8:91.

76.   Aref-Eshghi E, Schenkel LC, Lin H, et al. The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. *Epigenetics.* 2017;12(11):923-933.

77.   Aref-Eshghi E, Bend EG, Hood RL, et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. *Nature communications.* Nov 20 2018;9(1):4885.

78.   Helsmoortel C, Vulto-van Silfhout AT, Coe BP, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nature genetics.* Apr 2014;46(4):380-384.

79.   Van Dijck A, Vulto-van Silfhout AT, Cappuyns E, et al. Clinical Presentation of a Complex Neurodevelopmental Disorder Caused by Mutations in ADNP. *Biological psychiatry.* Feb 15 2019;85(4):287-297.

80.   Bend EG, Aref-Eshghi E, Everman DB, et al. Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. *Clinical epigenetics.* Apr 27 2019;11(1):64.

81.   Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American journal of human genetics.* May 14 2010;86(5):749-764.

82.   Tammimies K, Marshall CR, Walker S, et al. Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. *Jama.* Sep 1 2015;314(9):895-903.

83.   Suwinski P, Ong C, Ling MHT, Poh YM, Khan AM, Ong HS. Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics. *Frontiers in genetics.* 2019;10:49.

84.   Hsu P, Ma A, Wilson M, et al. CHARGE syndrome: a review. *Journal of paediatrics and child health.* Jul 2014;50(7):504-511.

85.   Butcher DT, Cytrynbaum C, Turinsky AL, et al. CHARGE and Kabuki Syndromes: Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions. *American journal of human genetics.* May 4 2017;100(5):773-788.

86.   Fahrner JA, Bjornsson HT. Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states. *Annual review of genomics and human genetics.* 2014;15:269-293.

87.   Chater-Diehl E, Ejaz R, Cytrynbaum C, et al. New insights into DNA methylation signatures: SMARCA2 variants in Nicolaides-Baraitser syndrome. *BMC medical genomics.* Jul 9 2019;12(1):105.

88.   Krzyzewska IM, Maas SM, Henneman P, et al. A genome-wide DNA methylation signature for SETD1B-related syndrome. *Clinical epigenetics.* Nov 4 2019;11(1):156.

89.   Hood RL, Schenkel LC, Nikkel SM, et al. The defining DNA methylation signature of Floating-Harbor Syndrome. *Scientific reports.* Dec 9 2016;6:38803.

90.   Schenkel LC, Kernohan KD, McBride A, et al. Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. *Epigenetics & chromatin.* 2017;10:10.

91. Ciolfi A, Aref-Eshghi E, Pizzi S, et al. Frameshift mutations at the C-terminus of HIST1H1E result in a specific DNA hypomethylation signature. *Clinical epigenetics.* Jan 7 2020;12(1):7.

92. Choufani S, Cytrynbaum C, Chung BH, et al. NSD1 mutations generate a genome-wide DNA methylation signature. *Nature communications.* Dec 22 2015;6:10207.

93. Siu MT, Butcher DT, Turinsky AL, et al. Functional DNA methylation signatures for autism spectrum disorder genomic risk loci: 16p11.2 deletions and CHD8 variants. *Clinical epigenetics.* Jul 16 2019;11(1):103.

94. Strong E, Butcher DT, Singhania R, et al. Symmetrical Dose-Dependent DNA-Methylation Profiles in Children with Deletion or Duplication of 7q11.23. *American journal of human genetics.* Aug 6 2015;97(2):216-227.

95. Guastafierro T, Bacalini MG, Marcoccia A, et al. Genome-wide DNA methylation analysis in blood cells from patients with Werner syndrome. *Clinical epigenetics.* 2017;9:92.

96. Bacalini MG, Gentilini D, Boattini A, et al. Identification of a DNA methylation signature in blood cells from persons with Down Syndrome. *Aging.* Feb 2015;7(2):82-96.

97. Henderson LB, Applegate CD, Wohler E, Sheridan MB, Hoover-Fong J, Batista DA. The impact of chromosomal microarray on clinical management: a retrospective analysis. *Genetics in medicine : official journal of the American College of Medical Genetics.* Sep 2014;16(9):657-664

98. Deciphering Developmental Disorders S. Prevalence and architecture of de novo mutations in developmental disorders. *Nature.* Feb 23 2017;542(7642):433-438.

99. Arnett AB, Rhoads CL, Hoekzema K, et al. The autism spectrum phenotype in ADNP syndrome. *Autism Res.* 2018;11(9):1300-1310. doi:10.1002/aur.1980

100. Cappuccio G, Sayou C, Tanno PL, et al. De novo SMARCA2 variants clustered outside the helicase domain cause a new recognizable syndrome with intellectual disability and blepharophimosis distinct from Nicolaides-Baraitser syndrome. *Genet Med.* 2020;22(11):1838-1850. doi:10.1038/s41436-020-0898-y

101. Van Houdt JK, Nowakowska BA, Sousa SB, et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nature genetics.* Feb 26 2012;44(4):445-449, S441.

102. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development.* 1959;3(3):210-229. doi:10.1147/rd.33.0210

103. HUANG S, CAI N, PACHECO PP, NARANDES S, WANG Y, XU W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics.* 2017;15(1):41-51. doi:10.21873/cgp.20063

104. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics.* 2018;15(1):41-51. doi:10.21873/cgp.20063

105. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev.* 2012;36(4):1140-1152. doi:10.1016/j.neubiorev.2012.01.004

106. Ozer ME, Sarica PO, Arga KY. New Machine Learning Applications to Accelerate Personalized Medicine in Breast Cancer: Rise of the Support Vector Machines. *OMICS.* 2020;24(5):241-246. doi:10.1089/omi.2020.0001

107. Yan J, Qiu Y, Ribeiro Dos Santos AM, et al. Systematic analysis of binding of transcription factors to noncoding variants. *Nature.* 2021;591(7848):147-151. doi:10.1038/s41586-021-03211-0

108. Platt JC. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In: *Advances in Large Margin Classifiers*. MIT Press; 1999:61-74.

109. Alders M, Maas SM, Kadouch DJ, et al. Methylation analysis in tongue tissue of BWS patients identifies the (EPI)genetic cause in 3 patients with normal methylation levels in blood. *European journal of medical genetics.* May-Jun 2014;57(6):293-297.

110. Schenkel LC, Aref-Eshghi E, Skinner C, et al. Peripheral blood epi-signature of Claes-Jensen syndrome enables sensitive and specific identification of patients and healthy carriers with pathogenic mutations in KDM5C. *Clinical epigenetics.* 2018;10:21.

111. Berletch JB, Yang F, Xu J, Carrel L, Disteche CM. Genes that escape from X inactivation. *Human genetics.* Aug 2011;130(2):237-245.

112. Zweig, M H, and G Campbell. "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine." *Clinical chemistry* vol. 39,4 (1993): 561-77.

113. Braun PR, Han S, Hing B, et al. Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. Transl Psychiatry. 2019;9(1):47. Published 2019 Jan 31. doi:10.1038/s41398-019-0376-y

114. Horvath S, et al. Aging effects on DNA methylation modules in human brain and blood tissue. Genome Biol. 2012;13:R97. doi: 10.1186/gb-2012-13-10-r97.

115. Davies MN, et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol. 2012;13:R43. doi: 10.1186/gb-2012-13-6-r43.

116. Malousi A, Andreou AZ, Georgiou E, Tzimagiorgis G, Kovatsi L, Kouidou S. Age-dependent methylation in epigenetic clock CpGs is associated with G-quadruplex, co-transcriptionally formed RNA structures and tentative splice sites. Epigenetics. 2018;13(8):808-821. doi: 10.1080/15592294.2018.1514232. Epub 2018 Sep 29. PMID: 30270726; PMCID: PMC6224212.

117. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86. doi:10.1186/1471-2105-13-86

118. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol Biol. 2014;1084:193-226. doi: 10.1007/978-1-62703-658-0_11. PMID: 24061923; PMCID: PMC4676806.

119. Hout MC, Papesh MH, Goldinger SD. Multidimensional scaling. Wiley Interdiscip Rev Cogn Sci. 2013;4(1):93-103. doi:10.1002/wcs.1203

Chapter 2: DNA methylation Episignature in Gabriele De Vries syndrome

Florian Cherik, Jack Reilly, Jennifer Kerkhof, Michael Levy, Haley McConkey, Mouna Barat-Houari, Kameryn M. Butler, Christine Coubes, Jennifer A. Lee, Gwenael Le Guyader, Raymond J. Louie, Wesley G. Patterson, Matthew L. Tedder, Mads Bak, Trine Bjørg Hammer, William Craigen, Florence Démurger, Christèle Dubourg, Mélanie Fradin, Rachel Franciskovich, Eirik Frengen, Jennifer Friedman, Nathalie Ruiz Palares, Maria Iascone, Doriana Misceo, Pauline Monin, Sylvie Odent, Christophe Philippe, Flavien Rouxel, Veronica Saletti, Petter Strømme, Perla Cassayre Thulin, Bekim Sadikovic, David Genevieve

Correspondence and requests for materials should be addressed to
Bekim Sadikovic, Department of Pathology and Laboratory Medicine, London Health Sciences Center, 800 Commissioners Rd. East, London N6A5W9, Canada.
 E-mail address: bekim.sadikovic@lhsc.on.ca

David Genevieve, Département de Génétique Médicale, CHRU Arnaud de Villeneuve, 371 avenue du doyen Gaston Giraud, Montpellier 34000, France.
E-mail address: d-genevieve@chu-montpellier.fr

2

PREFACE

In exploring diagnostic avenues available to DNA methylation episignature assessment, one of the most common outcomes observed has been a relationship involving a single gene disruption resulting in a shared episignature for all patients with such a disruption. Although I did identify a single case which indicated the possibility of alternate phenotypes based on variant type and location, due to low sample size, I was unable to fully research the underpinning biology associated with this atypical case in the context of DNA methylation. Nevertheless, it does indicate that future avenues of research, along the lines of the more specialized and stratified episignatures describe in the later chapters of this thesis may indeed be present in Gabriele De Vries (GADEV) patients.

Additionally, with all case samples (with the exception of the atypical case) clustering strongly together in all models, and no evidence of further overlap with other neurodevelopmental disorders, indicted by high scores on the support vector machine-based methylation variant pathogenicity (MVP) scores, this discovered episignature has been shown to be highly sensitive and specific to the disorder in question. This is by no means a new finding, and has been firmly established in a number of our published works over the past several years. Nevertheless, in providing an epigenetic roadmap assessment, it is important to include this baseline outcome to better understand how my work has since expanded on the overall complexity of episignature assessment.

As such, in this chapter, I will be discussing the discovery of a highly sensitive and specific episignature for the neurodevelopmental disorder Gabriele De Vries syndrome. DNA methylation patterns derived from patient samples showed to be effective biomarkers for this disorder, and provided interesting insights into alternate forms of the condition.

ABSTRACT

PURPOSE

Gabriele-de Vries syndrome (GADEVS) is a rare genetic disorder characterized by developmental delay and/or intellectual disability, hypotonia, feeding difficulties, and distinct facial features. To refine the phenotype and to better understand the molecular basis of the syndrome, we analyzed clinical data and performed genome-wide DNA methylation analysis of a series of individuals carrying a YY1 variant.

METHODS

Clinical data were collected for 13 individuals not yet reported through an international call for collaboration. DNA was collected for 11 of these individuals and 2 previously reported individuals in an attempt to delineate a specific DNA methylation signature in GADEVS.

RESULTS

Phenotype in most individuals overlapped with the previously described features. We described 1 individual with atypical phenotype, heterozygous for a missense variant in a domain usually not involved in individuals with YY1 pathogenic missense variations. We also described a specific peripheral blood DNA methylation profile associated with YY1 variants.

CONCLUSIONS

We reported a distinct DNA methylation episignature in GADEVS. We expanded the clinical profile of GADEVS to include thin/sparse hair and cryptorchidism. We also highlighted the utility of DNA methylation episignature analysis for classification of variants of unknown clinical significance.

KEYWORDS

DNA methylation, Epigenetics, Gabriele-de Vries syndrome, Intellectual disability, YY1

INTRODUCTION

Alteration of proteins involved in chromatin regulation is a well-established cause of many neurodevelopmental disorders. Among these conditions, Gabriele-de Vries syndrome (GADEVS, OMIM# 617557) is a rare congenital disorder characterized by variable intellectual disability (ID), various neurological disorders (hypotonia, abnormal movements, behavioral disorders, brain abnormalities), feeding difficulties, ophthalmological abnormalities, significant but not specific facial features, and more rarely cardiac or renal malformations [1-5]. GADEVS is mainly caused by pathogenic missense variants in Ying Yang 1 Transcription Factor gene (*YY1*, OMIM# 600013) and less frequently by truncating variants or whole gene deletions, suggesting haploinsufficiency as the underlying mechanism [1].

YY1 encodes the Ying Yang 1 Transcription factor, which is a ubiquitously expressed transcription factor in mammals. Its name comes from its ability to be both an activator and a repressor of transcription [6]. *YY1* is characterized by four highly conserved C2H2 Zinc fingers located in its C-terminal domain. The N-terminal region corresponds to the transcriptional activation domain. A transcriptional repression domain, including the REPO domain allowing the recruitment of the polycomb complex, is located between the N-terminal region and Zinc fingers domain[7-9].

It has been demonstrated that genetic disorders involving genes related to chromatin regulatory functions exhibit specific DNA methylation signatures, referred to as episignatures [10-12]. DNA methylation episignature analysis has recently been implemented as the diagnostic clinical genomic DNA methylation test EpiSign, in patients with rare disorders, providing strong evidence for clinical utility including the ability to provide conclusive diagnostic findings in the majority of subjects tested[13] In this study, we describe the clinical phenotype of 13 previously unpublished individuals carrying a pathogenic or likely pathogenic variant or a complete deletion of *YY1*, as well as a specific epigenetic signature associated with GADEVS.

MATERIALS AND METHODS

SUBJECTS AND SERIES

We contacted clinicians about 19 individuals carrying a pathogenic or likely pathogenic variant or a deletion of *YY1* through clinical networks (Groupe DI France, AnDDI-RARES http://anddi-rares.org/, ERN ITACHA https://ern-ithaca.eu/) and GeneMatcher (http://www.genematcher.org)[14]. We collected clinical and molecular data, DNA samples, brain MRI and neuropsychological assessment data of individuals from this series, when available. Referring physicians provided the data by filling in a standardized table.

This study was approved by the Institutional Review Board of Montpellier University Hospital (IRBMTP_2020_05_202000459, ClinicalTrial.gov identifier: NCT04381715) and the Western University Research Ethics Board (REB 106302). We obtained informed written consent from all individuals or their legal guardians to participate in the study and to publish their photographs. All samples and records were de-identified. The research was conducted in accordance with the Declaration of Helsinki.

MOLECULAR STUDIES

Diagnostic laboratories performed genetic tests on DNA from blood samples using next-generation sequencing or microarrays. The pathogenicity of point variants was verified according to American College of Medical Genetics and Genomics (ACMG) recommendations [15], using the varsome interface (https://varsome.com/)[16]. The visualization of the variants on the protein sequence was performed with the ProteinPaint tool (https://proteinpaint.stjude.org/), using the canonical isoform NM_003403.

STATISTICAL ANALYSIS

To describe the continuous variables (SD of growth parameters, age of milestones acquisition), we calculated medians, minimums, maximums and interquartile ranges in order to construct corresponding boxplots. We also included data from the literature in these graphs.

METHYLATION ARRAY AND QUALITY CONTROL

DNA methylation analysis and episignature classifier development was performed using a previously established protocol [11,12,17,18]. Genomic DNA was extracted from peripheral blood samples using standard techniques and followed by bisulfite conversion and hybridization to the Illumina Infinium methylation EPIC bead chip arrays, according to manufacturer's protocol. Idat files, containing methylated and unmethylated signal intensity plots (beta values) were produced from these microarrays, and used for analysis in R 4.0.2. Normalization was performed using the Illumina Infinium methylation EPIC array with background correction from the minfi package [19]. Previously defined exclusion criteria [12,17] were used to exclude probes with detection p values >0.01, probes on the X and Y chromosomes, probes known to contain SNPs at the site of CpG interrogation or single nucleotide extension, and probes known to cross react with chromosomal locations other than their target regions. All samples were examined for genome-wide methylation distribution and those deviating from a bimodal distribution were excluded. Factor analysis using a principal component analysis (PCA) was performed to examine batch effects and identify outliers.

DNA METHYLATION PROFILING

Probe methylation levels (beta values), were calculated as the ratio of signal intensity in methylated probes versus total sum of unmethylated and methylated probes, resulting in values ranging from zero to one. To allow for linear regression modeling, beta values were logit transformed using the limma package [20], allowing for the identification of differentially methylated probes. Data were adjusted for the blood cell type composition as per Houseman et al [21]. Estimated blood cell proportion was added to the model matrix of the linear models as a confounding variable [22] Using the eBayes function in the limma package [23], p values were moderated and corrected for multiple testing using the Benjamini Hochberg (BH) method. Probes with the most significant methylation differences were selected using two items from this dataset: the level of methylation difference (relative methylation signal intensity), and the probability that an observed difference is due to random chance (p values). Evaluation of this interaction was carried out by multiplying the absolute methylation difference between affected cases and controls by the negative value of the log transformed p values, and ranking the top 1000 probes with the highest values from this transformation. Next, receiver operating

characteristic analysis (ROC) was performed on each probe, to measure the pairwise correlation coefficient between probes. Probes with low area under curve values from ROC analysis were removed, as well as highly correlated probes, eliminating probes with low sensitivity and specificity, and probes with highly correlated characteristics using Pearson's correlation coefficient. This ensures that the final probeset contains the most differentiating, non-redundant probes that are not influenced by random data structures. Only probes with a methylation difference greater than 5% were included in this analysis. This probe filtering process was designed to avoid reporting of probes with low effect size, and those influenced by technical or random variations as conducted in previous studies [11,12].

SELECTION OF MATCHED CONTROLS FOR METHYLATION PROFILING

For episignature characterization, mapping of probes and feature selection, matched controls were randomly selected from the LHSC EpiSign Knowledge Database (EKD)[12]. All of the GADEVS samples were assayed, therefore all the controls selected for episignature identification were analyzed using the same array type. Samples were matched by age, sex and batch using the MatchIt package. A 4:1 ratio of controls to cases was deemed optimal for this analysis, as previously described [11]. PCA analysis was performed after each attempt at matching to detect outliers and determine data structures for the presence of batch effect. Outlier samples, and those with highly aberrant data structures were removed, and subsequent matching trials were performed until consistent iterations with no outliers in the first two components of the PCA were derived. No such samples were identified for removal in this cohort.

CLUSTERING AND DIMENSION REDUCTION

Hierarchical clustering and multidimensional scaling were used after each iteration of analysis to examine the data structure of the identified episignature. Hierarchical clustering was performed using Ward's method on Euclidean distance by the base stats package in R, and visualized with the ggplot2 package [24,25]. Multi-dimensional scaling provides a visual representation of sample methylation profile similarity based on the scaling of the pairwise Euclidean distances between each sample.

DISCOVERY/TRAINING COHORT SELECTION

Identification of disease-specific episignatures was performed using a randomly selected sub-set of the database, on a 75:25 ratio of discovery: training, using the caTools package in R. Testing samples were used to assess the performance of the classification model developed later in the study. For every disease group in the discovery cohort, a sex and age-matched control group with a sample size at least four times larger was selected from the reference control group using the MatchIT package, and methylation profiles were compared between the two.

CROSS VALIDATION

For each round of validation, one of the 13 selected GADEVS samples was removed from probe selection, alongside matched controls. The remaining GADEVS samples were designated as testing samples, and all three groups were modeled using multidimensional scaling to determine how they cluster/segregate with one another. This process was repeated with different combinations of assigned training and testing samples until all cases had been removed from probe selection and used for testing once (see Figure 2-S2).

CLASSIFICATION MODEL

Specificity of the episignature was assessed using the Methylation Variant Pathogenicity (MVP) score, using all the identified probes. A support vector machine (SVM) used a linear kernel for training on GADEVS cases and controls. Once again, a 4:1 ratio of controls to cases was used to divide both the case and control samples previously matched and used for probe selection into training and testing cohorts for the SVM. Furthermore, the remaining unselected samples from the EKD were also divided similarly (75% training, 25% testing) to allow for comparison and testing of signature robustness against all of the samples in the EKD. Using the e1071 R package, we performed 10-fold cross validation to determine hyperparameters optimal for episignature classification. In this process, the training set was divided into ten folds by random assignment, where the first nine are used for training, and the last used for testing the accuracy of the model. The mean accuracy over all rounds was then calculated, and hyperparameters with the best performance by this metric were selected. The model provides a score ranging from 0-1 for each subject, representing the model's confidence in predicting

43

whether the subject has a DNA methylation profile similar to the GADEVS probe set or not. Conversion of these SVM decision values was done using Platt's scaling method [26], and the class obtaining the greatest score determined the predicted phenotype. A classification as GADEVS was made when a sample received the greatest score for that class (normally greater than 0.5). Finally, the model was applied to both a training set of a large cohort of individuals with clinical and molecular diagnoses of neurodevelopmental disorders, as well as a group of healthy controls to determine its effective specificity.

VALIDATION OF CLASSIFICATION

To ensure the model is not susceptible to the batch structure of the methylation experiment, the classifier was applied to samples assayed on the same batch as the cases used for training. Using methylation data from individuals without a confirmed diagnosis of GADEVS within the EKD and assayed on the same microarray chip as case samples, methylation profiles were modeled to ensure the classifier was not confounded by technical artifacts unique to the given microarray. Specificity was determined by supplying a large number of DNA methylation arrays from unaffected subjects to the model. To further assess the specificity of the GADEVS classifier relative to other neurodevelopmental disorders, we applied it to cases with other patient cohorts exhibiting distinct episignatures within the EKD.

RESULTS

CONSTITUTION OF THE SERIES

We contacted the referring clinicians of 19 individuals carrying a pathogenic or likely pathogenic variant in *YY1*. We excluded three individuals either because they refused to participate in the study or because neither clinical data nor a DNA sample was available. Another individual was excluded because the *YY1* variant was inherited from a healthy parent. We therefore included 15 individuals in this study. For details see Table 2-S1.

Among these 15 individuals, 13 were not previously reported; these 13 individuals were labeled YY1-1 to YY1-13 and constituted the clinical series that allowed us to refine the phenotypic data related to *YY1* variants. The two remaining individuals, "individual 5" and "individual 8," were initially reported by Gabriele et al, 2017 [1]. Regarding the episignature series, we used DNA samples from 11 individuals of the clinical series (samples from

individuals YY1-8 and YY1-9 were not available) along with DNA samples from "individual 5" and "individual 8." The episignature series is detailed in Table 2-S2.

CLINICAL SPECTRUM ASSOCIATED WITH YY1 PATHOGENIC VARIANTS

Clinical data were collected for the 13 individuals (YY1-1 to YY1-13) not previously reported. Detailed clinical data are available in Table 2-S3. Among this series, 12 individuals had a phenotype overlapping with that previously described in the literature. Unfortunately, individual YY1-6 (father of individual YY1-7) died accidentally before being clinically assessed. The only data concerning individual YY1-6 is the presence of ID. Due to the clearly unusual phenotype that was observed in individual YY1-10, we chose to describe him separately. Table 2-1 summarizes the clinical data from this series and the literature. The 12 individuals with phenotype overlapping with the literature presented with variable ID and/or developmental delay. All these individuals presented with craniofacial features among which the most frequent were long face, broad forehead, simple ears, malar hypoplasia and full nasal tip. They also frequently had thin and/or sparse hair. (figure 2-1A).

We also observed various neurological disorders such as hypotonia, behavioral disorders (ASD, low frustration tolerance, anxiety, self-harm, ADHD), and abnormal movement (dystonia). Feeding disorders were present in 10/10 individuals. Frequent additional features include skeletal abnormalities, ophthalmologic abnormalities, and cryptorchidism.

Overall distribution (including data from literature) of ages of growth parameters and milestones achievement is represented in figure 2-1 (respectively B and C). Individual YY1-10 was considered to have an unusual *YY1* phenotype because of overgrowth and obesity (BMI=41kg/m²), slight macrocephaly (HC at 53cm [+2.3SD]) and moderate craniofacial features (See figure 2-1A). Full clinical features of individual YY1-10 are detailed in Table 2-S3.

| | Classical YY1 phenotype | | | Atypical phenotype |
	Present study (n=12)	Literature (n=14)	Total (n=26)	YY1-10
Growth				
IUGR	1/9	4/13	5/22 (23%)	-
Short stature	2/11	2/14	4/25 (16%)	Overgrowth
BMI < -2SD	4/10	3/10	7/20 (35%)	Obesity
Microcephaly	2/10	1/12	3/22 (14%)	Macrocephaly
Development				
Motor delay	8/11	11/14	19/25 (76%)	+
Language delay	10/11	10/12	20/23 (87%)	+
ID	11/12	11/12	22/24 (92%)	+
Neurological features				
Hypotonia	5/11	5/13	10/24 (42%)	+
Behavioral disorders	10/11	7/12	17/23 (74%)	+
Abnormal movement	4/11[a]	7/12	11/23 (48%)	-
Abnormal brain MRI	4/8	8/13	12/21 (57%)	+
Miscellaneous				
Cardiac abnormalities	1/9	4/11	5/20 (25%)	-
Cryptorchidism	3/7	1/5	4/12 (33%)	-
Skeletal abnormalities	9/10[b]	8/13	17/23 (74%)	-
Feeding disorders	10/10	12/13	22/23 (96%)	-
Constipation	4/11	NR	4/11 (36%)	-
Sparse hair	6/10	NR (9/12)*	15/22 (68%)	-
Endocrine abnormalities	2/9	3/14	5/22 (16%)	-
Recurrent infections	2/10	3/14	5/24 (21%)	-
Ophthalmologic abnormalities	9/10[c]	7/13	16/23 (70%)	+
Deafness	1/10	NR	1/10 (10%)	-
Morphological features				
Long face	8/11	NR (7/12)*	15/23 (65%)	-
Facial asymmetry	3/10	9/14	12/24 (50%)	-
Broad forehead	9/11	14/14	23/25 (92%)	-
Ears abnormality	11/11	12/12	23/23 (100%)	-
Up slanting palpebral fissures	4/10	1/11	5/21 (24%)	-
Down slanting palpebral fissures	2/10	6/11	8/21 (38%)	-
Full nasal tip	8/10	11/13	19/23 (83%)	+
Malar hypoplasia	6/10	11/13	17/23 (74%)	-
Smooth philtrum	3/9	NR (2/12)*	5/21 (24%)	Deep
Thin upper lip	5/10	NR (1/12)*	6/22 (27%)	Thick
Thick lower lip	2/10	5/13	7/21 (33%)	+
Pointed chin	3/10	5/12	8/22 (36%)	-
Micrognathia	3/10	NR (3/12)*	6/22 (27%)	-

Table 2-1: Summary of clinical features of individuals carrying a pathogenic variant of *YY1* (this series and the literature). Individuals with an atypical variant are described separately. Frequencies marked by

46

an * are based on our own interpretation of the pictures available in literature. +: feature present; - : feature absent; NR: not reported; NK: not known. a: dystonia, dyskinesia; b: camptodactyly, joint hyperlaxity, scoliosis, plagiocephaly, turricephaly; c: Hyperopia, superficial punctatae keratitis, nystagmus, strabismus

(5/12), astigmatism, myopia, cortical vision abnormalities.
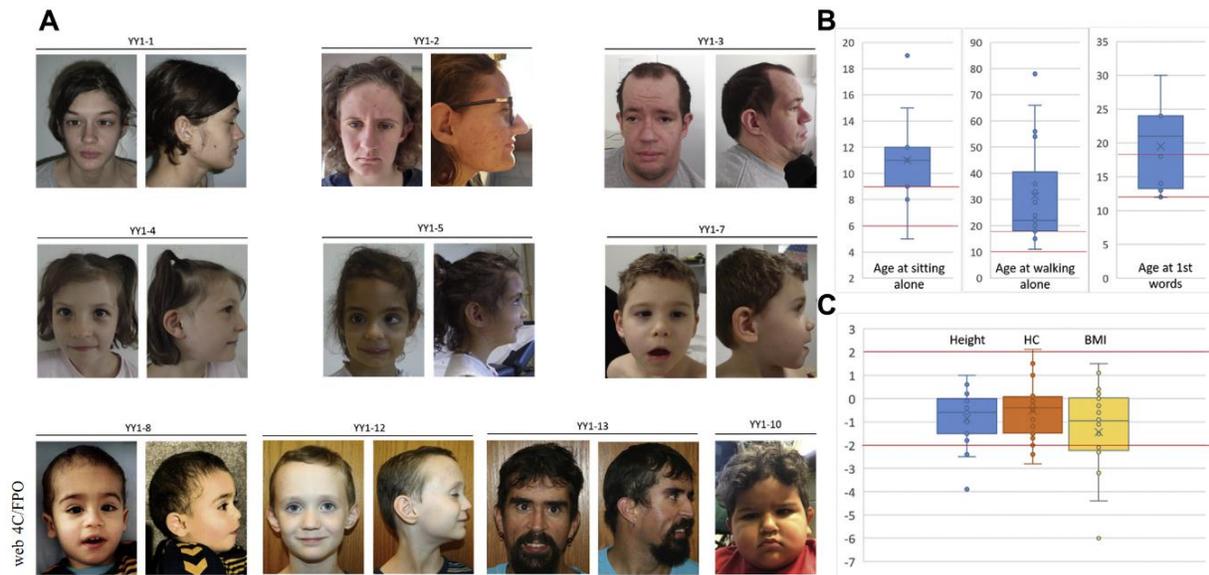
Figure 2-1: Representation of some clinical features related to *YY1*. **A:** Front and lateral view of individuals from this series. Common facial features are long face, broad forehead, simple ears, malar hypoplasia, full nasal tip and sparse hair. **B:** Boxplots showing distribution of ages at sitting alone, walking alone and first words in standard deviation. **C:** Boxplots showing distribution of height, head circumference (HC) and BMI, in standard deviation.

YY1 VARIANTS SPECTRUM

We collected molecular data from 13 unpublished individuals including a father-son pair (individuals YY1-6 and YY1-7). Except for this pair, all variants were *de novo*. The variants of the series and from the literature are represented on the YY1 protein sequence in Figure 2-S1. Among these 13 individuals, 12 carried a pathogenic or likely pathogenic sequencing variant (10 were missense and two were truncating variants). All missense variants were located in zinc finger domain except for individuals YY1-10. The variant p.(Gly176Asp) from individual YY1-10 was located in the transcriptional repression domain. Missense variants located in this domain has never been previously described in the literature to our knowledge. The last individual YY1-3 had a microdeletion encompassing *YY1*, *WARS1* and the 3' end of *EML1*.

DETECTION AND VERIFICATION OF AN EPISIGNATURE FOR YY1/GADEVS

DNA methylation profiles from 13 individuals peripheral blood samples, which all had confirmed molecular variants in the *YY1* gene and clinical presentation of GADEVS, were used to establish a DNA methylation episignature for this disorder. Overall methylation patterns in all 13 patients were assessed for several key features, including sample quality, and similarity of the sample methylation profiles to case samples versus controls. Of these, one sample, YY1-10 segregated consistently with controls, exhibiting methylation patterns more similar to age and sex matched control samples than the rest of the disorder cohort, and was removed from probe selection. Comparisons were carried out, matching GADEVS samples with age, sex and batch-matched controls at a ratio of 4:1 (4 matched controls for each case sample). When compared to controls, significant differences in methylation patterns across 487 probes, which are visualized using a volcano plot (Figure 2-S2) were detected. Selected probes had a minimum methylation difference of 10%, and a multiple testing corrected p value of <0.01 (limma multivariable regression modeling).

VISUALIZATION OF METHYLATION PROFILES INDICATES DISTINCT CLUSTERING PATTERNS OF YY1 CASES

Hierarchical clustering was used to visualize methylation differences based on the selected probes, and was plotted using Ward's method alongside 56 age and sex-matched control samples. This model demonstrated a clear separation of the control and case samples, with the exception of the YY1-10 sample. This sample grouped with control samples in all iterations of the model, indicating that the associated variant in this sample results in a methylation profile more similar to control samples than the other cases with confirmed *YY1* variants. The location and characteristics of the variant are atypical, with a missense mutation within the transcriptional repression domain of YY1, and reported presentation of overgrowth characteristics. Multiple dimensional scaling (MDS) showed similar findings, with cases grouping tightly together away from the control cohorts (see Figure 2-2B). Cross validation using GADEVS samples was performed, showing in the majority of cases that the remaining testing samples clustered with the other GADEVS samples, and segregated from the controls. In three cases, samples YY1-6, YY1-7 and YY1-11, cross validation showed less specific clustering along with lowered MVP scores, suggesting a level of signal heterogeneity and further data structure within the observed common episignature. However, all samples consistently segregated with the case cohort in hierarchical clustering and multidimensional scaling plots, and received high MVP scores when provided to the finalized SVM classifier (see Figure 2-2D and Figure 2-S2).
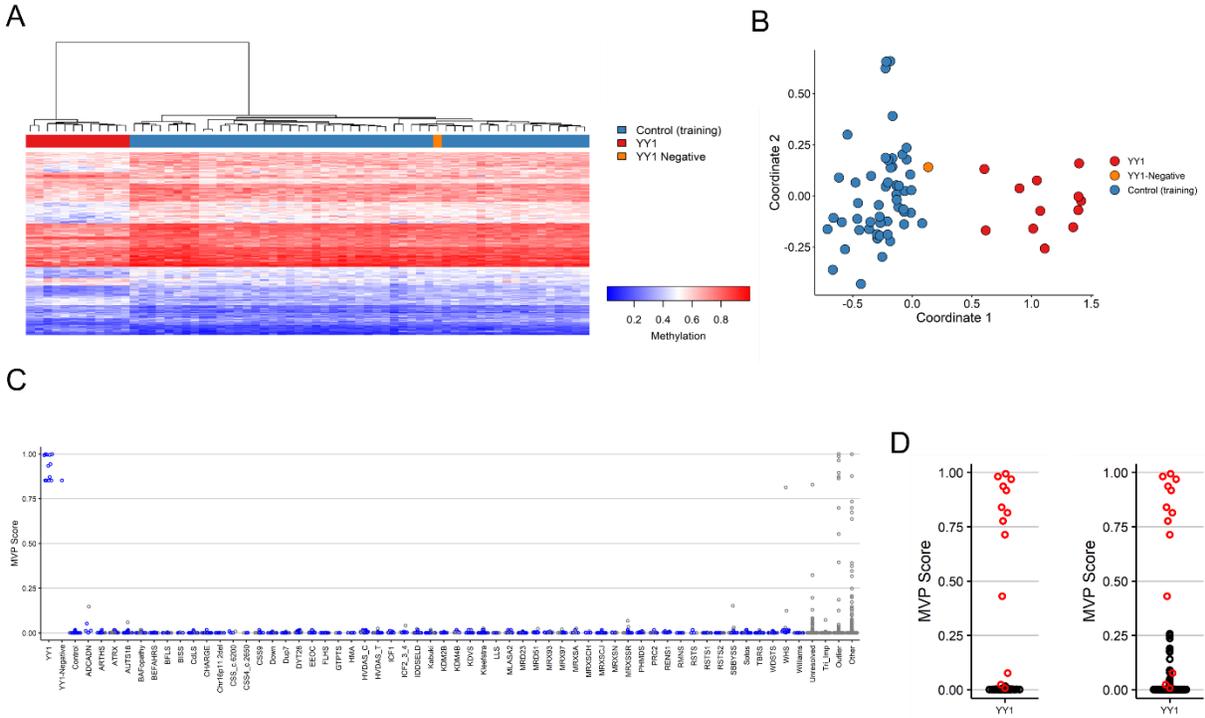
Figure 2-2: **DNA Methylation Profile for GADEVs**

A: DNA methylation signal intensity plot for 13 patients with identified *YY1* mutations sorted by hierarchical clustering. Cases in red represent GADEVS cases, those in blue indicate cases with no phenotypic or genotypic presentation of GADEVS, including samples with confirmed presentation of other syndromes, and the final case in orange, refers to sample YY1-10, which was removed from probe selection following segregation with control samples. **B:** Multidimensional scaling plot representing the dimensions of variation in methylation signal intensity at informative CpG identified for GADEVS. Represents similarity of methylation profiles of GADEVS patients, marked in red. **C:** SVM classifier model for GADEVS. Each sample receives scores for the probability of having a DNA methylation profile similar to cases as compared to samples with a confirmed Episignature in the EKD. Higher value on Y-axis indicates that a sample presents a methylation profile more similar to cases compared to the methylation profiles of patients with other disorders. Thirty-six other syndromes with confirmed Episignatures from the EKD are plotted based on this relative scale of similarity to indicate probeset specificity for the case disorder. **D:** Cross Validation summary representing the MVP scores received for each sample during their respective testing round. Case samples are marked in red, while the remaining samples from the EKD are marked in black. Left side plot contains MVP scores for the EKD samples following training the SVM against controls, while the right contains MCP scores for EKD samples following training the SVM against all samples within the EKD.

50

MVP SCORE DEMONSTRATES SENSITIVITY AND SPECIFICITY OF GADEVS EPISIGNATURE

Samples were provided to a support vector machine binary classifier with a linear kernel to assess the sensitivity and specificity, and the ability of the selected probe set to classify samples. For each sample, the classifier provides a methylation variant pathogenicity (MVP) score between 0 and 1. When plotted against control samples, all GADEVS samples received high scores (>0.8) close to 1, while the control scores remained near 0, indicating the classifier has a high sensitivity for the detection of the GADEVS episignature (see Figure 2-2C). Furthermore, specificity of the classifier was tested by providing it with a large number of subjects with a confirmed diagnosis of a neurodevelopmental disorder of various types with existing episignatures within the EKD. 75% of both case and control samples from other syndromes in the EKD were used for training, with the additional 25% reserved for testing. Case samples scored >0.8, while the remaining non-GADEVS cases scored very low, with no case exceeding a score of 0.5 to be classified as a GADEVS sample, indicating a very high level of specificity for the selected probe set.

DISCUSSION

We describe the phenotype of 12 new individuals carrying a pathogenic or likely pathogenic variant in *YY1*, proposed to lead to *YY1* loss-of-function as reported by Gabriele et al, 2017 [1]. In addition, missense variants in the zinc finger domains and truncating variants both lead to an overall decrease in the occupancy of *YY1* on the genome and a loss of H3K27 acetylation at the active enhancers linked by *YY1*, and consequently to a differential expression of target genes [1]. It was therefore postulated that *YY1* could have an impact on DNA methylation especially since *YY1* has been demonstrated to have the ability to recruit the Polycomb complex [27,28] known to be involved in the control of DNA methylation [29].

We observed a similar phenotype in our series to that described in the literature, such as variable ID and developmental delay, behavioral and abnormal movement disorders, skeletal abnormalities, and ophthalmological abnormalities, associated with craniofacial features and feeding difficulties with a consequent low BMI in individuals with the classical variants.

We also observed some differences including additional clinical features not previously described in the literature, including thin and/or sparse hair (6/10). Looking at pictures from the literature it seems that collectively, 15/22 (68%) of *YY1* individuals had this clinical feature.

In addition, congenital malformations and cardiac malformations seem less frequent in this series. We also observed cryptorchidism in 3/7 males, whereas this feature has been described only once in the literature. However, *YY1*-related disorders are very rare, and it is difficult to make conclusions on such a small number of individuals. Indeed, despite an international call for recruitment we could only identify 13 new individuals with *YY1* pathogenic or likely pathogenic variants according to ACMG classification criteria. This condition is probably still underdiagnosed since the involvement of *YY1* in neurodevelopmental disorders only became recognized in 2017. As the severity of ID seems to be variable in GADEVS, we wanted to study the neurocognitive profile of individuals carrying a *YY1* variant in order to highlight a possible specific pattern. However, data from neuropsychological assessments were largely insufficient, because the data were incomplete or uninterpretable. Additional studies should be performed to this point.

In addition to the clinical features, here we demonstrate the first evidence of a peripheral blood DNA methylation episignature, as a common molecular phenotype in patients presenting with classical features of GADEVS. All samples provided evidence of a common methylation profile for GADEVs, with of limited signal heterogeneity within the cross-validation model for 3 samples (YY1-6, YY1-7, and YY1-11) which received more moderate scores compared to the rest of the cohort. These findings, alongside the atypical sample (YY1-10) indicate the possibility of additional data structure, or sub-signatures, associated with variants in the *YY1* sequence, similar to what is observed in some other genetic conditions [17,18]. Further research with larger sample size will be necessary to study this hypothesis.

Individual YY1-3 carrying a deletion encompassing *YY1* plus two other genes (*WARS1* [MIM 191050] and *EML1* [MIM 602033]) has a similar epigenetic signature to that observed in individuals with pathogenic missense variants, suggesting that his phenotype can be at least partially attributed to *YY1* haploinsufficiency. In addition to dystonia previously described by Gabriele et al, 2017, Carminho-Rodrigues et al, 2020 [3] and Zorzi et al, 2021 [4], individual YY1-3 also has severe spasticity, as well as short stature (-3.9 SD). One of the other two genes, *WARS1*, could explain the additional neurological feature as this gene is associated with a

dominant distal motor neuropathy phenotype, however *WARS1* variants described in this condition are all missense [30-32] so this is unlikely.

Regarding individual with atypical localization of missense variant (individual YY1-10 with p.(Gly176Asp), located in in the transcriptional repression domain), we observed some major differences in phenotype than GADEVS, i.e., overgrowth, obesity and macrocephaly. Moreover, his DNA methylation profile is not specific and does not fit with the GADEVS episignature. The p.(Gly176Asp) variant was initially considered as likely pathogenic according to the ACMG classification (*de novo* variant absent from gnomAD exomes and genomes) but the result of the DNA methylation analysis has allowed us to reclassify this variant to unknown significance related to GADEVS. However, whether this variant is likely benign is not certain, given the possibility of yet to be defined alternate episignatures or lack thereof. The utility of EpiSign analysis in the reclassification of variants of uncertain clinical significance has been recently demonstrated in the clinical setting in a large number of Mendelian disorders with established episignatures [13]. Several studies have been published from our lab thus far involving additional substratification of episignatures [17-18] further highlighting the importance of methylation profiling in elucidating complex presentations of phenotype that remain unexplained by genetic diagnosis alone.

Considering the phenotype of overgrowth in individual YY1-10, the pathophysiological mechanism could be the selective alteration of the transcriptional repression function. However functional analysis or additional individuals with the same p.(Gly176Asp) *YY1* variant should be necessary to definitively rule out or confirm this variant to be responsible for a novel *YY1*-related disorder.

In conclusion, we describe 12 novel individuals with Gabriele-de Vries syndrome. We identified novel features (i.e., thin and/or sparse hair and cryptorchidism in males). We also describe for the first time a highly sensitive and specific DNA methylation episignature for GADEVS and demonstrate the utility of EpiSign in the clinical assessment of variants of uncertain clinical significance. Additional research is necessary to support the expanded clinical spectrum and genotype-phenotype correlations in GADEVS.

ACKNOWLEDGEMENTS

We thank the families for their participation in this study.

AUTHOR INFORMATION

Conceptualization: FC, JR, PM, DG, BS; Data curation: FC, JR, ML, DG; Formal analysis: FC, JR, ML, JK, MB, MBH, KMB, JAL, RJL, MLT, CD, NRP, MI, CP, EW, DJ; Investigation: FC, JR, JK, KB, CC, GLD, WGP, TBH, FD, WC, MF, RF, EF, JF, HG, DM, PM, SO, FR, VS, PS, PT, DG, BS; Methodology: DG, BS; Project administration: DG, BS; Supervision: DG, BS; Validation: FC, JR, DG; Visualization: FC, JR; Writing – original draft: FC, JR, DG, BS; Writing – review & editing: DG, BS

ETHICS DECLARATION

This study was approved by the Institutional Review Board of Montpellier University Hospital (IRBMTP_2020_05_202000459, ClinicalTrial.gov identifier: NCT04381715) and the Western University Research Ethics Board (REB 106302). We obtained informed written consent from all individuals or their legal guardians to participate in the study and to publish their photographs. All samples and records were de-identified. The research was conducted in accordance with the Declaration of Helsinki.

FUNDING

PERMISSIONS

This chapter was adapted for this thesis with permission from Genomics in Medicine.

REFERENCES

1. Gabriele M, Vulto-van Silfhout AT, Germain P-L, et al. YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction. *The American Journal of Human Genetics*. 2017;100(6):907-925. doi:10.1016/j.ajhg.2017.05.006

2. Morales-Rosado JA, Kaiwar C, Smith BE, Klee EW, Dhamija R. A case of *YY1* - associated syndromic learning disability or Gabriele-de Vries syndrome with myasthenia gravis. *Am J Med Genet Part A*. Published online December 14, 2018:ajmg.a.40626. doi:10.1002/ajmg.a.40626

3. Carminho-Rodrigues MT, Steel D, Sousa SB, et al. Complex movement disorder in a patient with heterozygous YY1 mutation (Gabriele-de Vries syndrome). *Am J Med Genet*. 2020;182(9):2129-2132. doi:10.1002/ajmg.a.61731

4. Zorzi G, Keller Sarmiento IJ, Danti FR, et al. *YY1* -Related Dystonia: Clinical Aspects and Long-Term Response to Deep Brain Stimulation. *Mov Disord*. Published online February 27, 2021:mds.28547. doi:10.1002/mds.28547

5. Tan L, Li Y, Liu F, et al. A 9-month-old Chinese patient with Gabriele-de Vries syndrome due to novel germline mutation in the *YY1* gene. *Mol Genet Genomic Med*. Published online December 23, 2020. doi:10.1002/mgg3.1582

6. Shi Y, Seto E, Chang L-S. Transcriptional Repression by YY1, a Human GLI-Kr ippeI-Related Protein, and Relief of Repression by Adenovirus EIA Protein. :12.

7. Sarvagalla S, Kolapalli SP, Vallabhapurapu S. The Two Sides of YY1 in Cancer: A Friend and a Foe. *Front Oncol*. 2019;9:1230. doi:10.3389/fonc.2019.01230

8. He Y, Casaccia-Bonnefil P. The Yin and Yang of YY1 in the nervous system. *Journal of Neurochemistry*. 2008;106(4):1493-1502. doi:10.1111/j.1471-4159.2008.05486.x

9. Atchison ML, Basu A, Zaprazna K, Papasani M. Mechanisms of Yin Yang 1 in Oncogenesis: The Importance of Indirect Effects. *Crit Rev Oncog*. 2011;16(3-4):143-161. doi:10.1615/CritRevOncog.v16.i3-4.20

10. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *The American Journal of Human Genetics*. 2020;106(3):356-370. doi:10.1016/j.ajhg.2020.01.019

11. Aref-Eshghi E, Bend EG, Colaiacovo S, et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. *Am J Hum Genet*. 2019;104(4):685-700. doi:10.1016/j.ajhg.2019.03.008

12. Aref-Eshghi E, Rodenhiser DI, Schenkel LC, et al. Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. *Am J Hum Genet*. 2018;102(1):156-174. doi:10.1016/j.ajhg.2017.12.008

13. Sadikovic B, Levy MA, Kerkhof J, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genet Med*. Published online February 5, 2021. doi:10.1038/s41436-020-01096-4

14. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Human Mutation*. 2015;36(10):928-930. doi:10.1002/humu.22844

15. On behalf of the ACMG Laboratory Quality Assurance Committee, Richards S, Aziz N, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-423. doi:10.1038/gim.2015.30

16. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. Wren J, ed. *Bioinformatics*. 2019;35(11):1978-1980. doi:10.1093/bioinformatics/bty897

17. Bend EG, Aref-Eshghi E, Everman DB, et al. Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. *Clin Epigenetics*. 2019;11(1):64. doi:10.1186/s13148-019-0658-5

18. Aref-Eshghi E, Bend EG, Hood RL, et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin–Siris and Nicolaides–Baraitser syndromes. *Nat Commun*. 2018;9(1):4885. doi:10.1038/s41467-018-07193-y

19. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049

20. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007

21. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86. doi:10.1186/1471-2105-13-86

22. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*. 2012;7(7):e41361. doi:10.1371/journal.pone.0041361

23. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3. doi:10.2202/1544-6115.1027

24. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2009.

25. Joe H. Ward. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*. 1963;58(301):236-244. doi:10.1080/01621459.1963.10500845

26. Smola AJ, Bartlett PJ. *Advances in Large Margin Classifiers*. MIT Press; 2000.

27. Atchison L. Transcription factor YY1 functions as a PcG protein in vivo. *The EMBO Journal*. 2003;22(6):1347-1358. doi:10.1093/emboj/cdg124

28. Wilkinson FH, Park K, Atchison ML. Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proceedings of the National Academy of Sciences*. 2006;103(51):19296-19301. doi:10.1073/pnas.0603564103

29. Viré E, Brenner C, Deplus R, et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*. 2006;439(7078):871-874. doi:10.1038/nature04431

30. Tsai P-C, Soong B-W, Mademan I, et al. A recurrent WARS mutation is a novel cause of autosomal dominant distal hereditary motor neuropathy. *Brain*. 2017;140(5):1252-1266. doi:10.1093/brain/awx058

31. Li J-Q, Dong H-L, Chen C-X, Wu Z-Y. A novel WARS mutation causes distal hereditary motor neuropathy in a Chinese family. *Brain*. 2019;142(9):e49. doi:10.1093/brain/awz218

32. Wang B, Li X, Huang S, et al. A novel WARS mutation (p.Asp314Gly) identified in a Chinese distal hereditary motor neuropathy family. *Clin Genet*. 2019;96(2):176-182. doi:10.1111/cge.13563

SUPPLEMENTARY FIGURES AND TABLES
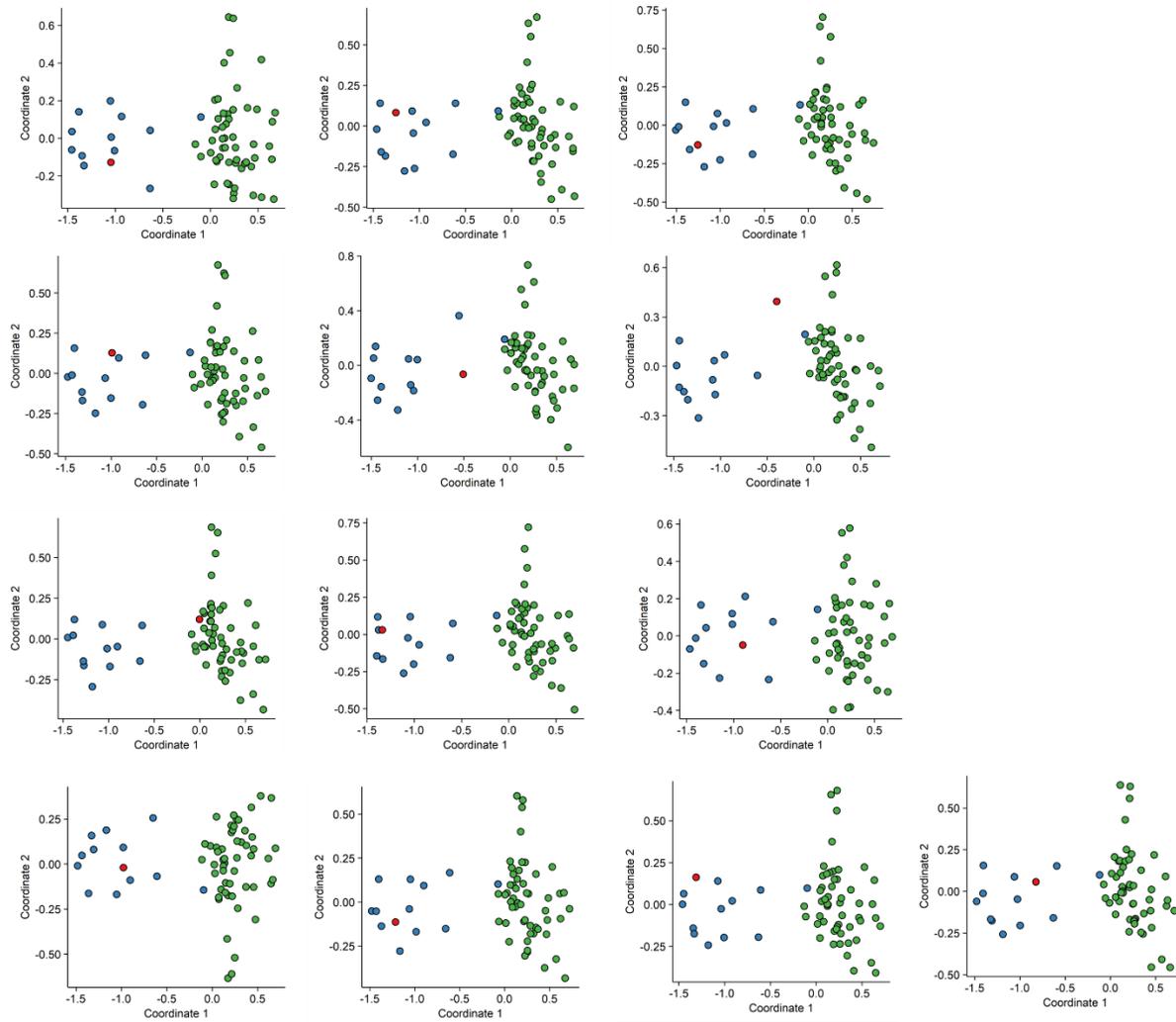
Figure 2-S1: Cross validation. For each round of validation, one of the 13 selected GADEVS samples was removed from probe selection, alongside matched controls. The remaining GADEVS samples were designated as testing samples, and all three groups were modeled using multidimensional scaling to determine how they cluster/segregate with one another. This process was repeated with different combinations of assigned training and testing samples until all cases had been removed from probe selection and used for testing once. Green, control; Blue, probe selection (training); Red, testing/validation. Red samples should cluster with blue samples.
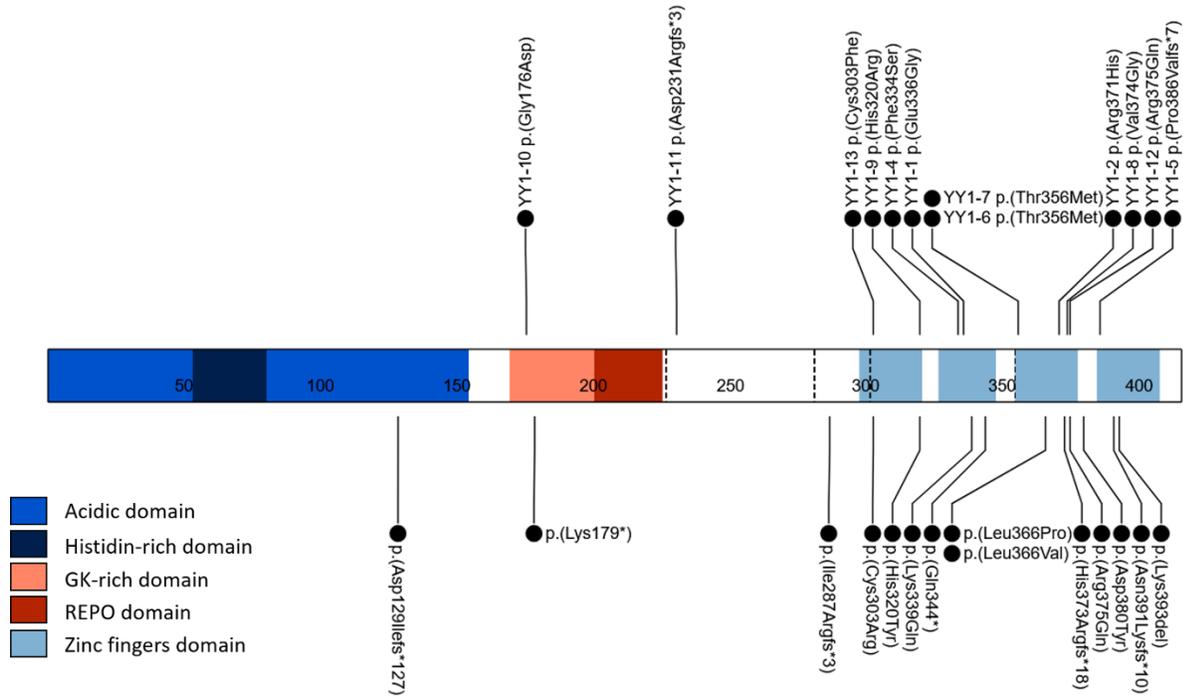
Figure 2-S2: Graphical representation of the YY1 protein and its functional domains. Variants from the series (top) and from the literature[1–5] (bottom) are indicated by tags. The missense variants from this series are mainly located in the Zinc finger domain, with the exception of one variant, which is located in the glycine and lysine rich domain, involved in transcription repression.
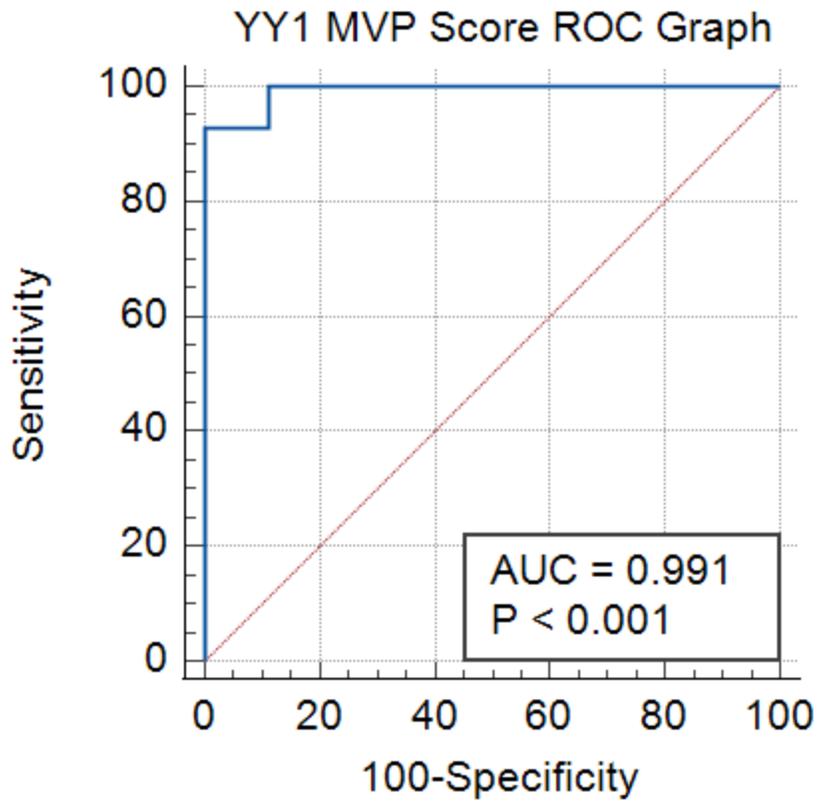
Figure 2-S3: YY1 MVP Score ROC Graph: Receiver operating characteristic curve demonstrating the sensitivity and specificity of the generated MVP scores for the YY1 cohort and the remaining EKD samples used for training.

	Recruited individuals	**Included individuals**
Montpellier University Hospital	3	3
Groupe DI	2	2
Anddi-RARES	2	2
ERN ITHACA	1	1
Clinical trial	1	0
GeneMatcher	8	5
Other	2	2
Total	19	15

Table 2-S1: number of individuals recruited and included according to the network used

Epi-signature ID	Individual ID	Label	Sex	Age at sample collection (years)	Genotype
MS3563	YY1-6	YY1	m	33	YY1(NM_003403.5):c.1067C>T, p.(Thr356Met)
MS3564	YY1-7	YY1	m	2	YY1(NM_003403.5):c.1067C>T, p.(Thr356Met)
MS3565	YY1-3	YY1	m	28	arr[GRCh37]14q32.2(100402364-101351127)x1
MS3567	YY1-2	YY1	f	17	YY1(NM_003403.3):c.1112G>A, p.(Arg371His)
MS3568	YY1-1	YY1	f	5	YY1(NM_003403.4):c.1007A>G, p.(Glu336Gly)
MS3569	Individual 5[+]	YY1	f	17*	YY1:c.535A>T,p.(Lys179*)
MS3570	YY1-5	YY1	f	4	YY1(NM_003403.4):c.1151_1154dup, p.(Pro386Valfs*7)
MS3571	YY1-4	YY1	f	1	YY1(NM_003403.4):c.1001T>C, p.(Phe334Ser)
MS3575	YY1-10	YY1 Negative	m	2.5	YY1(NM_003403.4):c.527G>A, p.(Gly176Asp)
MS4008	YY1-11	YY1	m	7	YY1(NM_003403.5):c.690dup, p.(Asn231Argfs*3)
MS4447	YY1-12	YY1	m	25.6	YY1(NM_003403.5):c.1124G>A, p.(Arg375Gln)
MS4828	YY1-13	YY1	m	19	YY1(NM_003403.4):c.908G>T, p.(Cys303Phe)
MS4881	Individual 8[+]	YY1	f	34	YY1(NM_003403.4):c.385delG, p.(Asp129Ilefs*127)

Table 2-S2: GADEVS samples with accompanying genetic and phenotypic information. DNA was derived from peripheral blood taken from patients with features of GADEVS previously identified through genomic sequencing and clinical assessment. Individuals marked by (a) are individuals already described in Gabriele et al., 2017 [1] Missing age information, marked by (b) was calculated using methylation based molecular estimates based on the Horvath Clock CpGs [6].

	YY1-1	YY1-2	YY1-3	YY1-4	YY1-5	YY1-6	YY1-7	YY1-8	YY1-9	YY1-10	YY1-11	YY1-12	YY1-13
Protein change	p.Glu336Gly	p.Arg371His	NR	p.Phe334Ser	p.Pro386Val fs*7	p.Thr356Met	p.Thr356Met	p.Val374Gly	p.His320Arg	p.Gly176Asp	p.Asp231Arg fs*3	p.Arg375Gln	p.Cys303Phe
Growth													
Genre	F	F	M	F	F	M	M	M	M	M	M	M	M
Age at examination	18y2m	16y7m	31y9m	6y	6y2m	33y	5y9m	4y9m	2y1m	2y10m	8y1m	25y7mo	20y6m
Birth weight in g (p)	1760 (37)	2940 (25)	1560 (0.3)	2880 (10)	1885 (3)	NP	2820 (NP)	3908 (50)	3160 (35)	4167 (94)	2515 (27)	2720 (7)	2920 (NP)
Weight in kg (p)	34.6 (-5 SD)	39.5 (-2,7 SD)	54 (-1,9 SD)	18.6 (-0,6 SD)	15.2 (-2,5 SD)	NP	15 (-2,4 SD)	18 (+0 SD)	9.2 (1y7m) (-2,4 SD)	45.3 (+7,9 SD)	20.5 (-1,7 SD)	44.2 (-3,6 SD)	56.2 (-1,6 SD)
Height in cm (Z-score)	162,5 (0,2 SD)	151,5 (-1,7 SD)	148,5 (-3,9 SD)	115 (+0 SD)	111 (-1 SD)	NP	101.5 (-2,4 SD)	110 (+0,6 SD)	78 (1y4m) (-0,6 SD)	105 (+2,7 SD)	118 (-1,8 SD)	169.4 (-1 SD)	176.5 (-0,1 SD)
BMI in kg/m² (Z-score)	13.2 (-6 SD)	17.2 (-1,5 SD)	24.5 (+0,4 SD)	14.1 (-0,9 SD)	12.3 (-3,2 SD)	NP	14.6 (-0,7 SD)	14.9 (-0,5 SD)	NR	40.8 (+6,8 SD)	14.7 (-0,6 SD)	15.4 (-4,43 SD)	18 (-2,3 SD)
HC in cm (Z-score)	NP	54.5 (0,1 SD)	54.5 (-1,4 SD)	49 (-1,7 SD)	50.5 (-0,5 SD)	NP	48 (-2,5 SD)	48 (1y9m) (-0,3 SD)	48 (-0,8 SD)	53 (+2,4 SD)	50,7 (-1,4 SD)	53 (-2,4 SD)	59 (+2,1 SD)
Psychomotor development													
Age of sitting alone (months)	NP	9	15	11	>9	NP	19	NP	NP	8	NP	9	8
Age of walking alone (months)	29	22	54	18	20	NP	66	54	Absent	18	18	18	15
1st words (months)	NP	NP	NP	24	30	NP	NR	NR	NP	15	12 to 24	18	12
1st sentences	NP	NP	NP	33	36	NP	NR	NR	NR	36	36 to 48	36	NP
Language	Delay	Delay	Simple sentences	Normal	Simple sentences	NP	Absent	Absent	Words	Simple sentences	Normal	Normal	Simple sentences
ID	+	+	+	+	+	+	+	+	+	+	Learning difficulties	+	+
Neurologic features													
Hypotonia	-	-	+	+	+	NP	+	-	+	+	-	-	-
Behavioral disorders	Anxiety, self-harm	Anxiety, self-harm, stereotypies	-	ADHD	Low frustration tolerance	NP	ASD	Autism	ADHD, low frustration tolerance, stereotyped behavior	ASD, aggressive-ness	Trouble focusing	Anxiety	ADHD, ASD, OCD, anxiety, aggressiveness
Abnormal movement	-	-	Severe dystonia, spasticity	Dystonia	-	NP	-	-	-	-	Facial tics, Tourette syndrome	-	Generalized dystonia
Brain MRI	NP	-	-	-	Aspecific FLAIR hyper-intensities	NP	Focal lesions of white matter	benign enlargement of subarachnoid spaces	-	Enlargement of subarachnoid spaces, perimesen-cephalic lipoma	NP	NP	Suspected focal cortical dysplasia, aspecific T2 hyperintensity foci
Other	Dysmetria	Febrile seizures	Oculomotor disorders	-	Amyotrophy	NP	-	--	Congenital torticollis, week tendon reflexes	-	Gait imbalance, exercise-induced	-	-

64

	YY1-1	YY1-2	YY1-3	YY1-4	YY1-5	YY1-6	YY1-7	YY1-8	YY1-9	YY1-10	YY1-11	YY1-12	YY1-13
Miscellaneous													
Cardiac abnormalities	-	-	-	AtSD, VSD	NP	NP	-	-	NP	-	-	-	-
Urologic abnormalities	-	-	Crypt-orchidism	-	NP	NP	Unilateral crypt-orchidism	-	Bilateral crypt-orchidism	-	-	-	-
Skeletal abnormalities	Campto-dactyly, hyperlaxity, scoliosis, long fingers	Hyperlaxity	NP	Hyperlaxity	Finger hyperlaxity, long fingers	NP	Occipital plagiocephaly hyperlaxity	-	Turricephaly	-	Joint hyperlaxity	Scoliosis, joint pain	Scheuerman's kyphosis (spinal fusion)
Gastro-intestinal abnormalities	Feeding disorders, constipation	Severe constipation, Feeding disorders	Feeding disorders in infancy	Feeding disorders in infancy	Feeding disorders	NP	Feeding disorders	Feeding disorders, constipation	NP	Feeding disorders	Feeding disorders	Feeding disorders in infancy Chronic constipation	Feeding disorders, G-tube
Hair abnormalities	-	-	-	Thin hair	Thin hair	NP	Thin hair	Thin hair	Sparse hair	-	Sparse hair	Facial hirsutism	-
Endocrine abnormalities	-	-	-	-	NP	NP	-	-	NP	-	-	Hypo-thyroidism	Thyroid nodule
Immune abnormalities	-	-	-	-	NP	NP	-	-	Recurrent infections	-	-	Recurrent infections	-
Ophthalmologic abnormalities	Hyperopia superficial punctuated keratitis, nystagmus	Hyperopia, astigmatism	-	+ (unspecified)	Convergent strabismus	NP	NP	Convergent strabismus, cortical vision abnormalities	Convergent strabismus	Strabismus, amblyopia, astigmatism	Strabismus	Strabismus Myopia	+ (unspecified)
Deafness	-	Conductive, bilateral 20-35dB	-	-	NP	NP	-	-	-	-	-	-	-

Table 2-S3: Full description of 12 previously unpublished individuals with *YY1* pathogenic variants.: the 13th individual, YY1-6, for whom we have no clinical information, is not reported in this table +: feature present; -: feature absent; ADHD: attention deficit hyperactivity disorder; ASD: autism spectrum disorder; AtSD: atrial septal defect; ID: intellectual disability; NP: not provided; p: percentile; NR: not relevant; SD: standard deviation; VSD: ventricular septal defect

CONFLICT OF INTEREST NOTIFICATION PAGE

The authors declare no conflict of interest

CONCLUSION

This second chapter in my thesis lays the groundwork for the ensuing works described in later chapters. In exploring the identification of a sensitive and specific biomarker for GADEVs, I found a robust episignature capable of classifying a number of patients with YY1 variants on the basis of their distinct methylation profiles. This biomarker can greatly improve the ability to detect this disorder in clinical contexts, adding to the already powerful genetic and phenotypic features that currently help guide diagnosis. Within ClinVar, 78 YY1 variants have been identified, with 20 currently identified as variants of unknown significance (20/78, 26%, See Appendix Table 1). DNA methylation profiling of patients with these variants could provide the functional evidence required to provide an effective diagnosis of GADEVs, thereby increasing the diagnostic yield for this gene sequence.

Furthermore, this episignature's interesting presentation of an atypical sample dissimilar in methylation profile and phenotype provides an interesting insight into the possibility of further subsignatures within the YY1 sequence. This atypical sample was classified as not having a methylation profile similar to other case samples by the hierarchical clustering heatmaps and multidimensional scaling models, as well as receiving low MVP scores within each iteration of our classifier, guiding us to look further into the possible reasons for this atypical presentation. These investigations led to novel insights into the patient in question, revealing phenotypic differences in the presentation of the disorder, characterized with symptoms of overgrowth not seen in other patients. This atypical phenotype was unknown to me in the preliminary stages of the analysis, and was identified through methylation profiling. Further research into the methylation differences observed in this sample, and samples with similar variants may reveal similarly affected pathways that explain the alternate phenotype observed, and should be explored to better understand these patients, as well as those presenting with the more typical phenotype of GADEVs.

Chapter 3: Delineation of a KDM2B-related neurodevelopmental disorder and its associated DNA methylation signature

Richard H. van Jaarsveld[1†], Jack Reilly[2†],, Marie-Claire Cornips[1], Michael A. Hadders[3], Emanuele Agolini[4], Priyanka Ahimaz[5],Kwame Anyane-Yeboa[5], Severine Audebert Bellanger[6], Ellen van Binsbergen1, Marie-Jose van den Boogaard1, Elise Brischoux-Boucher7, Raymond Caylor[8], Andrea Ciolfi[9], Ton AJ van Essen[10¶], Paolo Fontana[11], Saskia Hopman[1], Maria Iascone[12], Margaret M Javier[13], Erik-Jan Kamsteeg[14], Jennifer Kerkhof[15], Jun Kido[16], HyungGoo Kim[17], Tjitske Kleefstra[14], Fortunato Lonardo[11], Abbe Lai[18], Dorit Lev[19], Michael A. Levy[15], Suzanne M.E. Lewis[13], Angie Lichty[8], Naomichi Matsumoto[20], Idit Maya[21,22], Haley McConkey[15], Andre Megarbane[23,24], Vincent Michaud[25], Evalina Miele[26], Marcello Niceta[9], Antonio Novelli[4], Roberta Onesimo[26], Rolph Pfundt[14], Bernt Popp[28], Eloise Prijoles[8], Raissa Relator[15], Sylvia Redon[6], Dmitrijs Rots[14], Karen Rouault[6,29], Ken Saida[20], Jolanda Schieving[30], Marco Tartaglia[9], Romano Tenconi[31], Kevin Uguen[6], Nienke Verbeek[1], Christopher A Walsh[32], Keren Yosovich[33], Christopher J. Yuskaitis[18], Giuseppe Zampino[26,27] Bekim Sadikovic [2,15‡*], Mariëlle Alders[34‡,*], Renske Oegema[1‡*]

1. Department of Genetics, University Medical Centre Utrecht, Utrecht, the Netherlands
2. Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 3K7, Canada
3. Oncode Institute and Center for Molecular Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands
4. Medical Genetics Laboratory, Bambino Gesù Children's Hospital, Rome, Italy
5. Division of Clinical Genetics, Department of Pediatrics, Columbia University, New York, NY, USA
6. Service de Génétique médicale et de biologie de la reproduction, Centre Hospitalier Regional Universitaire Brest, 29200 Brest, France
7. Centre de Génétique Humaine, CHU de Besançon, Université de Franche-Comté,

Besançon, France

8. Greenwood Genetic Center, Greenwood, South Carolina, USA

9. Genetics and Rare Diseases Research Division, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy

10. University Medical Centre Groningen, University of Groningen, Department of Medical Genetics, Groningen, The Netherlands

11. Medical Genetics Unit, A.O.R.N. San Pio, Benevento, Italy

12. Laboratorio di Genetica Medica - ASST Papa Giovanni XXIII, Bergamo, Italy

13. Department of Medical Genetics, BC Children's Hospital Research Institute, The University of British Columbia, Vancouver, BC, Canada

14. Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands

15. Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, ON N6A 5W9, Canada

16. Department of Pediatrics, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan

17. Neurological Disorders Research Center, Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Doha, Qatar

18. Department of Neurology, Boston Children's Hospital, Boston, MA, USA; Division of Epilepsy and Neurophysiology, Boston Children's Hospital, Boston, MA, USA

19. The Rina Mor Institute of Medical Genetics, Wolfson Medical Center, Holon, Israel

20. Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan

21. The Raphael Recanati Genetic Institute, Rabin Medical Center, Beilinson Hospital Petach-Tikva, Israel

22. Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

23. Department of Human Genetics, Gilbert and Rose-Marie Chagoury School of Medicine, Lebanese American University, Lebanon.

24. Institut Jérôme Lejeune, Paris, France.

25. Department of Medical Genetics, CHU Bordeaux, Bordeaux, France

26. Center for Rare Diseases and Congenital Defects, Fondazione Policlinico Universitario

A. Gemelli IRCCS, Rome, Italy

27. Faculty of Medicine and Surgery, Catholic University of Sacred Hearth, Rome, Italy

28. Institute of Human Genetics, Friedrich-Alexander University of Erlangen-Nuremberg, Erlangen, Germany

29. Université de Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France

30. Department of Pediatric Neurology, Radboud University Medical Center, Nijmegen, the Netherlands

31. Department of Pediatrics, Clinical Genetics, Università di Padova, 35122 Padova, Italy

32. Division of Genetics and Genomics and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA

33. Molecular Genetic Laboratory, Edith Wolfson Medical Center, Holon, Israel

34. Amsterdam UMC, Department of Clinical Genetics, Genome Diagnostics laboratory Amsterdam, Reproduction & Development, University of Amsterdam, Amsterdam, The Netherlands

†, ‡ These authors contributed equally

¶ In memoriam.

* Corresponding authors: Renske Oegema: R.Oegema@umcutrecht.nl, Mariëlle Alders: m.alders@amsterdamumc.nl, Bekim Sadikovic: Bekim.Sadikovic@lhsc.on.ca

3

---

3   A version of this chapter has been submitted to the Journal of Clinical Investigation, awaiting review.

PREFACE

In the previous chapter, we discussed the discovery of a novel episignature for Gabriele De Vries syndrome, which encompassed a number of variants within the associated YY1 transcription factor gene sequence. This episignature was common to the entirety of the cohort across the spectrum of genetic variants with the exception of a single atypical case. Due to the small sample size of this atypical signature, I hypothesized the possibility of an additional episignature, associated with an alternate phenotype of overgrowth, requiring further research and similar samples to effectively classify this potential subsignature. Expanding on this potential avenue of multiple episignatures within a cohort of patients involving disruptions of the same gene sequence, I now present to you my findings within a cohort of patients with KDM2B variants, associated with a novel NDD we have coined as KDM2B-related-neurodevelopmental - disorder or K2BNDD. Within this cohort, we identified not only an episignature for variants within the KDM2B gene, similar to the one described in my previous chapter, but a specific DNA methylation pattern associated with disruption of a particular gene domain within the KDM2B sequence.

Disruption of the CxxC DNA binding motif domain within the KDM2B sequence resulted in a distinct DNA methylation subsignature, characterized a much larger magnitude of change in methylation signal intensity with distinct multidimensional scaling models when compared to the matched controls and other KDM2B cases with variants outside this particular domain. Furthermore, phenotypic differences were also observed in these CxxC domain patients, with considerable increases in the incidence of congenital anomalies, not seen in other KDM2B cases. Gathering these findings, I propose the existence of multiple domain specific episignatures that are robust diagnostic biomarkers, unique not only to a particular gene sequence, but even to specific domains within those gene sequences. This chapter demonstrates the assessment of multiple episignatures with multiple specific phenotypes, as a result of domain specific differences, within a cohort derived from a single gene origin.

ABSTRACT

Mutations in genes involved in the epigenetic machinery are an emerging cause for neurodevelopmental disorders (NDDs). Lysine-demethylase 2B (*KDM2B*) encodes an epigenetic regulator but has not been recognized as an NDD gene to date. Here we present a cohort of 21 individuals with heterozygous –likely- pathogenic variants in *KDM2B*. These individuals present with developmental delay and/or intellectual disability, autism, attention deficit /hyperactivity disorder, AD(H)D, congenital organ anomalies and facial dysmorphism. To establish this cohort, we assessed 24 variants in 33 individuals. We applied methylation arrays on blood-derived DNA samples to establish a *KDM2B*-specific epigenetic signature characterized by hypermethylation of CpG-dinucleotides. We identify the CxxC-domain as a mutational hotspot and identify a specific episignature for this subgroup. Importantly, we were able to detect the *KDM2B*-episignature even in the context of a dual diagnosis with the presence of another episignature, demonstrating the robustness of this assay.

INTRODUCTION

Many genes encoding for epigenetic regulators have been implicated as monogenic disease genes in neuro-developmental disorders (NDDs). This group of disorders, collectively referred to as 'Mendelian Disorders of the Epigenetic Machinery' (MDEMs) , is characterized by intellectual disability (ID) and/or growth abnormalities [1]. For an increasing number of MDEMs, distinct genome-wide methylation signatures (or episignatures) have been identified [2]. These signatures are emerging as valuable tools in clinical practice, as they are unique for each disorder and can be detected in peripheral blood samples, providing a robust and easily accessible diagnostic tool [3]. The *KDM2B* gene (lysine-demethylase 2B, a.k.a FBXL10, NDY1, CXXC2 and JHDM1B; OMIM #609078) encodes for a well-studied component of the epigenetic machinery. The canonical, full-length KDM2B protein (KDM2B-Long Form; KDM2B-LF) acts by demethylating lysine residues K4, K36 and K79 of Histone 3 [4,5,6,7,8,9,10,11]. This catalytic activity is provided for by the JmjC-domain , which is conserved from yeast to humans [11,12]. Interestingly, an alternative transcript produces a shorter KDM2B isoform (KDM2B-SF), which lacks the JmjC domain and thus lacks catalytic activity [13]. This short form is highly expressed in mouse embryonic stem cells (mESC) [14] suggesting important functions of *KDM2B* not directly related to lysine demethylation activity.

Apart from the JmjC-domain, both KDM2B isoforms share the same architecture, consisting of four additional domains. The first is the CxxC-domain, a DNA-binding domain that specifically binds unmethylated CpG-dinucleotides and directs KDM2B to promoter regions [15, 16, 17]. The CxxC-domain is stabilized by the adjacent PHD-domain [17], a protein interacting domain that binds to methylated H3K4 and H3K36 residues [8]. The KDM2B proteins are completed by the F-box and LRR domains which are implicated in protein-protein interactions, most notably the CUL1-RING complex and Polycomb Group proteins [18, 19, 20].

KDM2B has been implicated in many biological processes, including cell cycle regulation, metabolic regulation and DNA-damage repair [5, 15, 21, 22]. Moreover, in line with a central role in epigenetic and transcriptional regulation, *KDM2B* is essential for organism development and regulates cellular differentiation [15]. For instance, *KDM2B* can immortalize cells and maintains stemness in mESC [7, 23]. In addition, *KDM2B* is essential for survival of neuronal progenitor cells, and full knock-out causes aberrant neuronal development in mice, and ultimately embryonic lethality [14,15]. Interestingly, re-expression of only KDM2B-SF in a *Kdm2b* knock-out background resets the methylation of CpG-islands to baseline levels and rescues embryonic lethality [15]. Although the molecular mechanisms by which KDM2B operates -and the contributions of each isoform to these functions- remain to be determined, these findings suggest that the functions of *KDM2B* are not limited to lysine demethylation alone.

Despite scarce reports describing individuals carrying *KDM2B* germline variants [24, 25, 26, 27, 28], a *KDM2B*-related human disorder has not been delineated to date and the significance of reported variants remains uncertain. Here, we present a series of 33 individuals with heterozygous *KDM2B* variants, collected through international collaborations and literature review. We establish a *KDM2B*-related episignature and apply this for further characterization of the identified variants. For 21 individuals (representing 16 different variants), variants were classified as (likely) pathogenic. We delineate a novel NDD with or without congenital anomalies, and propose to refer to this novel syndrome as 'KDM2B-related NDD' (K2BNDD).

RESULTS

GENETIC VARIANTS IN KDM2B

The present study was initiated after the identification of a *de novo* c.1912G>A (p.Gly638Ser) variant in *KDM2B* (NM_032590.4; Table 3-1, Extended Data Table 3-1) by diagnostic trio-exome-sequencing (trio-ES)   in the index patient (#1). *KDM2B* presented as an outstanding disease gene candidate as the gene is intolerant for both putative loss-of-function (pLoF; o/e=0.09 [0.05-0.18]) and missense (Z=3.44) variants in the general population [29]. Furthermore, *KDM2B* is a known epigenetic regulator and the patient's phenotype fitted with known MDEMs (Table 2, Extended Data Table 3-2) [1]. In addition, the identified variant was absent from the gnomAD database [29], predicted damaging by multiple algorithms (Table 3-1, Extended data Table 3-1) and affects a well conserved residue (Supplementary Figure 3-S1A) in a known functional domain (i.e. the CxxC-domain; Figure 3-1A & 3-1B). We therefore aimed to collect additional cases carrying *KDM2B* variants and formed the present cohort after online matchmaking using the Genematcher platform [30], literature search, personal communication and in-house database searches.
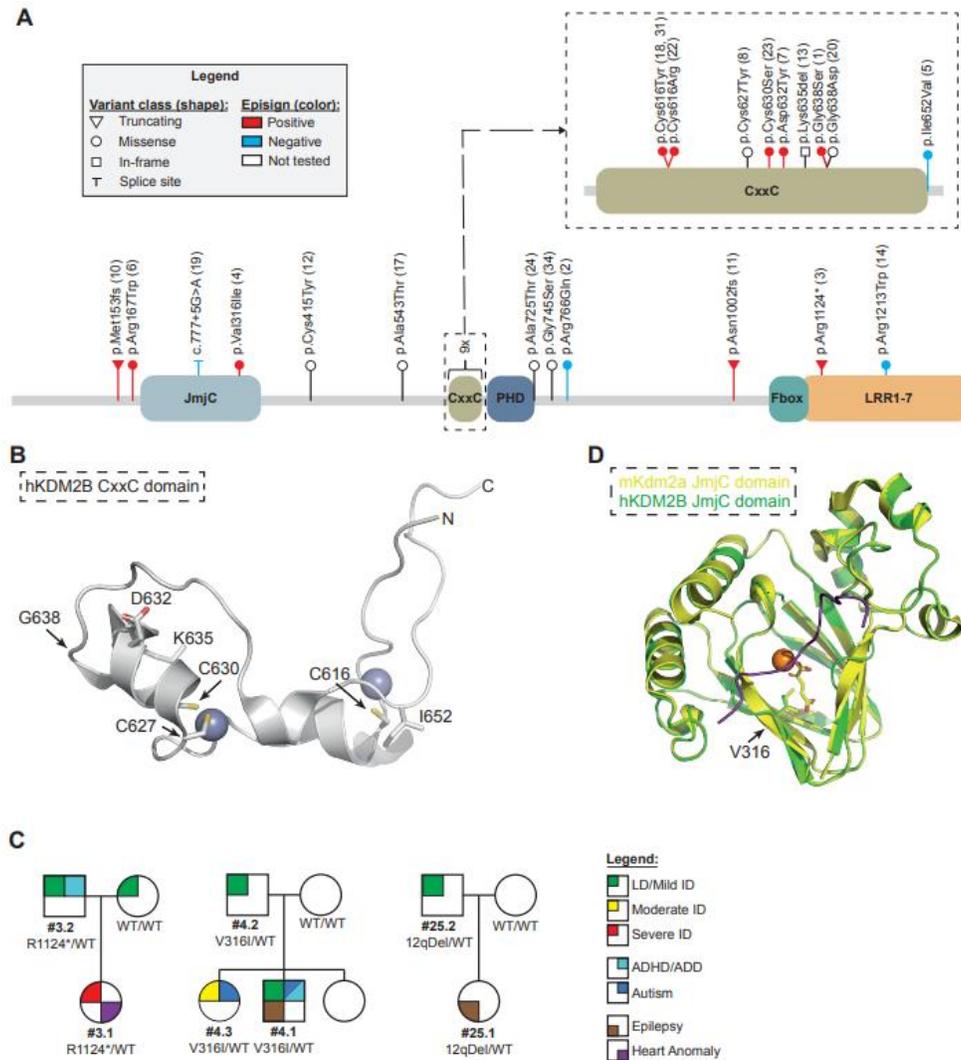
Figure 3-1: A cohort of heterozygous KDM2B variant carriers. A: Schematic representation of the *KDM2B* gene, its known domains and the variants included in this study. Lollipops representing individual variants indicate location, classification of (predicted) impact on the transcript and/or protein (shape) and the classification based on the first analysis of the methylation arrays (color; Supplementary Figure 3-S2). Larger deletions (i.e. cases 25.1, 25.2, 29 and 30) are not shown. **B:** CxxC-domain missense variants on the known crystal structure. Purple spheres represent Zn^{2+} ions. Sidechains of relevant residues are included. **C:** Pedigrees depicting all cases of inherited variants of which the pedigree has not been published before (families 3, 4, 5 and 25). All remaining pedigrees can be found in Supplementary Figure 3-S1. ADHD: Attention Deficit Hyperactivity Disorder; ID: Intellectual Disability; LD: Learning Difficulties; WT: Wild Type . D) Projection of the p.Val316Ile (family 4) variant on the structure of the mouse KDM2A JmjC structure (yellow). The predicted human KDM2B JmjC structure as determined by AlphaFold is shown in green. Orange sphere indicates Fe^{2+} ion and the aKG cofactor is shown as yellow sticks. Purple line indicates target peptide (Histone 3).

We collected data of a total of 33 individuals from 25 families representing 24 different heterozygous variants in *KDM2B* (Figure 3-1, Table 3-1& 3-2, Supplementary Figure 3-S1, Extended data tables 1&2). Our cohort encompasses seven pLoFs (7/24), sixteen missense variants (16/24) and one in-frame deletion (1/24). Eighteen variants were confirmed *de novo* (18/24). Of note, one of these variants was identified twice and thus occurred *de novo* at two independent occasions (c.1847G>A, #18 and #31). Five variants are inherited (5/24; families 3, 4, 5, 24 and 25; Figure 3-1C and Supplementary Figure 3-S1B). Nine individuals have been previously reported in other studies: family 25 [31], family 24 [28], individuals #29 and #30 [32], and #34 [26]. An overview of all variants, including inheritance, presence in gnomAD and summarized results from prediction algorithms, is presented in Table 3-1 (details in Extended Data Table 3-1). Of note, we observed a remarkable clustering of coding-altering variants (8 missense and 1 in-frame deletion) in the CxxC-domain (Figure 3-1A and 3-1B), of which seven missense variants (7/8) were predicted to be damaging by all algorithms. The only exception is p.Ile652Val, which is furthermore the only inherited variant and reported twice in gnomAD.

We performed additional structural modeling of the coding altering variants located in areas where structural data is available. First, we projected the variants located in the CxxC-domain on the known crystal structure (Figure 3-1B) [17]. The missense variants located at positions p.Cys616, p.Cys627 and p.Cys630 are predicted to be damaging as they are likely to influence the interaction with the Zinc ions (Zn^{2+}) located at the respective positions. The p.Ile652Val variant is located near the same Zn^{2+} ion as p.Cys616, however this substitution is less likely to affect the position of the ion as the sidechain is located towards the surface, and the respective loss of a methyl group is not expected to influence the local structure. Functional consequences of this variant are therefore questionable, furthermore strengthened by its presence in gnomAD and inconsistent results from prediction algorithms (Table 3-1). The variants affecting p.Gly638 are expected to impact on the torsion angle at this location, most likely disrupting the local architecture and are therefore expected to be damaging. Finally, the p.Asp632Tyr and p.Lys635del variants are located within an α-helix and are likely to impact on the function of this structure.

Individual	Variant (NM_032590.4)		Inheritance	GnomAD	In silico prediction	EpiSign
Pathogenic variants						
1	c.1912G>A,	p.(Gly638Ser)	de novo	-	4/4	Y
3.1, 3.2	c.3370C>T	p.(Arg1124*)	Paternal	-	LoF	Y; Y
4.1, 4.2, 4.3	c.946G>A	p.(Val316Ile)	Paternal	1x	4/4	Y; U; Y
6	c.499C>T	p.(Arg167Trp)	de novo	-	4/4	Y
10	c.457delA	p.(Met153Cfs*24)	de novo	-	LoF	Y
11	c.3005_3023del19	p.(N1002Sfs*35)	de novo	-	LoF	Y
18, 31	c.1847G>A,	p.(Cys616Tyr)	de novo	-	4/4	Y; NA
20	c.1913G>A	p.(Gly638Asp)	de novo	-	4/4	NA
22	c.1846T>C	p.(Cys616Arg)	de novo	-	4/4	Y
23	c.1889G>C	p.(Cys630Ser)	de novo	-	4/4	Y
25.1, 25.2	12q24.31 deletion		Paternal	NA	LoF	Y
29	12q24.31 deletion		de novo	NA	LoF	Y
30	12q24.31 deletion		de novo	NA	LoF	Y
Likely pathogenic variants						
13	c.1903_1905delAAG	p.(K635del)	de novo	-	NA	NA
Variants of unknown significance (VUS)						
12	c.1244G>A	p.(Cys415Tyr)	unknown	-	1/4	NA
17	c.1627G>A	p.(Ala543Thr)	de novo	2x	4/4	NA
24.1-4	c.2173G>A	p.(Ala725Thr)	inherited	-	4/4	NA
34	c.2233G>A	p.(Gly745Ser)	de novo	-	4/4	NA
2	c.2297G>A	p.(Arg766Gln)	de novo	2x, 7alt	3/4	N
5.1, 5.2	c.1954A>G	p.(Ile652Val)	Maternal	2x	3/4	N; U
14	c.3637C>T	p.(Arg1213Trp)	de novo	5alt	4/4	N
19	c.777+5G>A	Splice site	de novo	-	3/3 reduced	N
Abbreviations: alt=alternative; LoF=Loss of Function; N=No/Negative; NA=Not Applicable/Not Assessed; U=Uncertain; x=times; Y=Yes/Positive; - =absent						

Table 3-1: Overview of KDM2B variants in the cohort. International collaborations resulted in a cohort of 33 individuals representing 24 variants in *KDM2B*. The table indicates genetic details of each variant, appearance of the variant in gnomAD (or alternative variants affecting the same residue, alt), summary of in silico prediction results and inclusion in the KDM2B episignature cohort. alt=alternative; LoF=Loss of Function; N=No/Negative; NA=Not Applicable/Not Assessed; U=Uncertain; x=times; Y=Yes/Positive. Additional and supporting information per variant can be found in Extended Data Table 3-

Ind	Sex, age (year)	Variant (NM_032590.4)	Inheritance	ID/DD	Behavior/ psychiatry	Hypotonia	Microcephaly (OFC < -2 SD)	Cardiac anomalies	Kidney anomalies	other
Pathogenic variants										
1	M, 7	c.1912G>A, p.(Gly638Ser)	De novo	Speech delay, SON-IQ 86	Autism	-	-	VSD, ASD, fetal atrial flutter	-	Familial polydactyly
3.1	F, 7	c.3370C>T, p.(Arg1124*)	Paternal	Severe ID	hyperactivity	+	-	VSD, DORV	NA	Phelan McDermid syndrome: 22q13 deletion
3.2	M, 42	c.3370C>T, p.(Arg1124*)	Unknown	Learning difficulties	ADD	NA	-	-	NA	COPD
4.1	M, 15	c.946G>A, p.(Val316Ile)	Paternal	Mild	Autism, ADHD, tantrums	NA	-	NA	NA	epilepsy
4.2	M, 65	c.946G>A, p.(Val316Ile)	Unknown	Learning difficulties - mild ID	NA	NA	-	NA	NA	decreased renal function, osteoporosis (adult age)
4.3	F, 21	c.946G>A, p.(Val316Ile)	Paternal	Moderate	Autism, tantrums, anxiety	NA	-	NA	NA	
6	M, 9	c.499C>T, p.(Arg167Trp)	De novo	Speech delay, non-verbal IQ 97	ADHD	-	-	-	-	Congenital ptosis, cryptorchidism
7	M, 4	c.1894G>T, p.(Asp632Tyr)	De novo	Learning difficulties	Autism, ADHD, impulsiveness	-	-	PVS, ASD	-	Hypertonia, progressive contractures, inguinal hernia
8	F, 6	c.1880G>A, p.(Cys627Tyr)	De novo	Mild speech delay	-	-	-	ASD, MR, PDA, PVS	Single kidney	Short stature
10	M, 10	c.457del, p.Met153Cysfs*24	De novo	Mild ID, IQ 66	-	-	-	Atrial septal aneurysm, MR	-	SHOC2-related Noonan syndrome
11	F, 5	c.3005_3023del19, p.(Asn1002Sfs35)	De novo	Global DD, moderate ID	Autism, hyperactivity	+	-	-	NA	Epilepsy, MRI abnormalities (MCD)
18	M, 5	c.1847G>A, p.(Cys616Tyr)	De novo	Moderate global DD	-	-	+	-	Single kidney	Coloboma, hypertrichosis, failure to thrive
20	F, 14	c.1913G>A, p.(Gly638Asp)	De novo	Speech delay - learning difficulties	Mild autistic features	NA	-	mild mitral insufficiency	-	Short stature, R oculomotor defect enophtalmus
22	M, 16m	c.1846T>C, p.(Cys616Arg)	De novo	Global DD, speech delay	-	Upper limbs	-	PFO	Single kidney	Brain MRI abnormalities, unilateral anophthalmia, bilateral SNHL, facial asymmetry
23	F, 3	c.1889G>C, p.(Cys630Ser)	De novo	Severe DD, no speech	-	+	NA	ASD	Single kidney, Right VUR	Short stature, poor weight gain, squint,

Ind	Sex, age (year)	Variant (NM_032590.4)	Inheritance	ID/DD	Behavioral difficulties	Hypotonia	Microcephaly (OFC < -2 SD)	Cardiac anomalies	Kidney anomalies	other
										congenital obstructio ductus nasolacrimalis
25.1	F, 12	12q24.31 deletion (including *KDM2B, HNF1A*)	Paternal	Severe, no speech, cannot walk	Not specified	+	+	NA	Normal renal function	Epilepsy, hip dysplasia Published by Chouery et al; Krzyzewska et al.
25.2	M, adult	12q24.31 deletion (including *KDM2B, HNF1A*)	Unknown	Normal	-	-	-	-	-	Insulin-dependent diabetes at 14y. Published by Chouery et al; Krzyzewska et al.
29	F, 12	12q24.31 deletion (including *KDM2B & SETD1B*)	De novo	+	Autism, ADHD	-	-	NA	NA	Preauricular tags, oligodontia, umbilical hernia, published Krzyzewska et al.
30	M	12q24.31 deletion (including *KDM2B & SETD1B*)	De novo	+	Probable autism	+	OFC at 4th percentile	NA	NA	Epilepsy, published Krzyzewska et al., patient 10; Labonne et al.
31	M, 5	c.1847G>A, p.(Cys616Tyr)	De novo	Speech delay, mild- moderate ID	Stereotypies	-	+	ASD	-	cryptorchidism, talus pes, kyphosis. congenital obstruction of ductus nasolacrimalis
Likely pathogenic variant (sample not available for methylation analysis)										
13	F, 4	c.1903_1905delAAG, p.(Lys635del)	De novo	Moderate speech delay, mild ID	-	+	+	ASD (x2), PVS, PDA, PFO	-	Feeding difficulties at birth
Ind	Sex, age (year)	Variant (NM_032590.4)	Inheritance	ID/DD	Behavioral difficulties	Hypotonia	Microcephaly (OFC < -2 SD)	Cardiac anomalies	Kidney anomalies	other
VUS (sample not available for methylation analysis)										
12	M, 0,2	c.1244G>A, p.(Cys415Tyr)	Unknown	UK	UK	+	-	UK	UK	Macrocephaly, polyhydramnios, club foot, contractures, multiple arthrogryposis, undescended testis; published Monies et al: PMID: 31130284
17	F, 7	c.1627G>A, p.(Ala543Thr)	De novo	+	Autism	+	-	PFO	NA	History of failure-to-thrive until age 2, epilepsy, later obesity, MRI abnormalities
24.1	F, 32	c.2173G>A, p.(Ala725Thr)	Maternal	Moderate	SCZ	NA	NA	Incomplete RBBB, normal echocardiogram	NA	IQ 39 after diagnosis of schizophrenia.

										published by Yokotsuka-Ishida et al.
24.2	F, 69	c.2173G>A, p.(Ala725Thr)	Unknown	NA	SCZ	NA	NA	NA	NA	published by Yokotsuka-Ishida et al.
24.3	F, 39	c.2173G>A, p.(Ala725Thr)	Maternal	+	SCZ	NA	NA	NA	NA	Seizures. published by Yokotsuka-Ishida et al.
24.4	M, 34	c.2173G>A, p.(Ala725Thr)	Maternal	Moderate - severe	SCZ	NA	NA	NA	NA	CP, L opaque cornea, L eye blindness. published by Yokotsuka-Ishida et al.
34	NA	p.Gly745Ser	De novo	NA	SCZ	NA	NA	NA	NA	Published Girard et al.
VUS (Variants not showing KDM2B specific episignature)										
2	F, 1,8	c.2297G>A, p.(Arg766Gln)	De novo	-	-	-	NA	-	NA	CL/P, preaxial polydactyly, finger contractures, thumb hypoplasia
5.1	M, 28	c.1954A>G, p.(Ile652Val)	Maternal	Mild- moderate	Autism	NA	-	NA	NA	Scoliosis, hearing loss due to cholesteatoma
5.2	F, 58	c.1954A>G, p.(Ile652Val)	Unknown	-	NA	NA	NA	NA	NA	dyslexia
14	M, 12	c.3637C>T, p.(Arg1213Trp)	De novo	Global DD, limited speech, mild ID (IQ 64)	Hyperactivity, aggressive behavior	+	-	-	-	Macrocephaly, epilepsy, brain MRI abnormalities, hand/finger abnormalities
19	F, 1	c.777+5G>A	De novo	Severe DD, no speech	NA	+	+	ASD		Neonatal seizures, thrombotic angiopathy, SNHL, abnormal renal vasculature

+= presence of feature, = absence of feature, ADHD= attention deficit hyperactivity disorder, ASD= atrial septal defect, CLP= cleft lip/palate, COPD= chronic obstructive pulmonary disease, CP= cerebral paresis, DD= developmental delay, F, female= ID, intellectual disability, DORV= double outlet right ventricle, L= left; M= male; MR= mitral regurgitation; MCD= malformation of cortical development, NA= not assessed, PDA= persistent ductus arteriosus, PFO= persistent foramen ovale, PVS= pulmonary valve stenosis, RBBB= right bundle branch block, SNHL= Sensorineural hearing loss, VSD= ventricular septal defect, VUR= vesicoureteral reflux

Table 3-2: An overview of the phenotypes associated with K2BNDD. This table summarizes the clinical features of individuals with *KDM2B* variants. More extensive data are presented in Extended Data Table 3-2 and the clinical summaries. Abbreviations: +=presence of feature, -=absence of feature, ADHD=attention deficit hyperactivity disorder, ASD=atrial septal defect, CLP=cleft lip/palate, COPD=chronic obstructive pulmonary disease, CP=cerebral paresis, DD=developmental delay, F=female, ID=intellectual disability, DORV=double outlet right ventricle, L=left; M=male; MR=mitral regurgitation; MCD=malformation of cortical development, NA=not assessed, PDA=persistent ductus arteriosus, PFO=persistent foramen ovale, PVS=pulmonary valve stenosis, RBBB=right bundle branch block, SNHL=Sensorineural hearing loss, VSD=ventricular septal defect, VUR=vesicoureteral reflux.

Two other missense variants are located in known domains for which structural data is available. The p.Ala725Thr variant is located at the border of the PHD-domain   and affects a residue located just outside the known structure [17]. This might indicate that this variant resides in an unstructured area, and we are therefore unable to predict functional consequences. The p.Val316Ile variant is located in the JmjC-domain. Since the structure of the *KDM2B* JmjC-domain remains to be resolved, we projected the variant on the known structure of the homologous mouse Kdm2a domain [12] and the human KDM2B AlphaFold model [33]. This residue is located within the active site of the JmjC-domain, near the catalytic metal ion and cofactor alpha-ketoglutarate binding site (Figure 3-1D). The variant is expected to impact on metal ion binding as it results in an increase in size, thus likely interfering with the catalytic activity of the JmjC-domain.

In summary, we collected a total of 24 variants in *KDM2B*. Based on the absence from controls and predicted functional impact on the gene product, the majority was considered promising candidates to explain the patients' phenotypes.

A GENOME WIDE EPISIGNATURE IN KDM2B PATIENTS

We next aimed to determine if the variants had an impact on *KDM2B* function. Due to its role in the epigenetic machinery, we hypothesized that *KDM2B* deficiency leads to genome-wide changes in DNA methylation; an effect which has been observed for >30 other monogenic disease genes involved in chromatin organization [2, 3, 34]. These methylation changes present as disease specific episignatures, which are detectable in peripheral blood. As such, episignatures not only provide fundamental insights into the molecular consequences of genetic variants; they provide easily accessible diagnostic tools to identify syndromes or re-classify variants of unknown significance (VUS) [3].

Under the assumptions that such an episignature also exists for *KDM2B* and the majority of the variants in our cohort disrupt gene function, we set out to determine a *KDM2B*-related episignature. To that end, we generated genome-wide methylation array data for 21 individuals (Table 3-1, Extended Data Table 3-1) according to previously established protocols. We excluded two samples because of technical errors (#2 and #4.2, Supplementary Figure 3-S2D). Another three samples were excluded for the establishment of the episignature, as they failed to

group with case samples after cross validations (#5.1, #14 and #19; Supplementary Figure 3-S2B, D and E). This suggests the respective variants do not impact on *KDM2B* function, or at least not in a similar fashion as the majority of the variants. Of note, all these three variants were marked uncertain based on inheritance and/or presence in gnomAD (Table 3-1, Extended Data Table 3-1). As the #5.1 sample -representing the p.Ile652Val variant- did not pass cross validation testing, we additionally excluded the sample from #5.2 from further analysis.

The remaining 15 samples, representing 13 variants, were used to establish a *KDM2B* episignature (Figure 3-2). To this end, methylation patterns were assessed for sample quality, degree of methylation change and statistical robustness of observed changes at each probe, allowing for effective modeling of the methylation differences observed between case samples and matched controls who do not carry *KDM2B* variants (see Materials and Methods). Comparisons were performed against age and sex matched controls, leading to the identification of 156 statistically differentially methylated probes (Figure 3-2A). Hierarchical clustering based on this probe set showed distinct clustering of case samples away from controls, with all samples presenting a more similar methylation profile to one another as compared matched controls (Figure 3-2B and 3-2C). Cross validation assays, based on the removal of each single sample from the probe selection training process, confirmed the probe set is able to effectively identify *KDM2B* variant carriers, as all case samples remained grouped together on each iteration (Supplementary Figure 3-S3B). In conclusion, we established an episignature able to discriminate *KDM2B* variant carriers from controls. Interestingly, the *KDM2B* associated episignature mainly consists of hypermethylated probes (Figure 3-2A).

A

pvalueAdjustMethod: BH; score: 900; methDiffValue: 0.05; rocValue: 450; corValue: 0.7; corMethod: 1

B

C

D

Figure 3-2: A *KDM2B* specific episignature. After initial analysis (Supplementary Figure 3-S2), fifteen samples, identified as outliers in the initial analysis, were included for the training of a *KDM2B* specific episignature. **A:** Volcano plot indicating selected probes (red) included in the *KDM2B* episignature. **B:** Multidimensional scaling (MDS) plot for selected probes, representing the pairwise distance across samples (red) and controls (blue), based on the top two dimensions. **C:** Heatmap of selected probes and unsupervised hierarchical clustering results indicating the episignature's ability to decipher *KDM2B* variant carriers (red) from controls (blue). **D:** Support Vector Machine (SVM) classifier indicating specificity of the KDM2B episignature. Graph shows summary of 4-fold validation using all fifteen case samples and 75% of unaffected controls and other episignatures for training (blue) and the other 25% for testing (grey). Y-axis: MVP scores as determined by SVM. X-axis: different groups of samples, controls and other known episignatures. Red arrowhead indicates the IDDSELD sample referred to in the text. *Note. Figure differs from previously published version of the paper. Results were regenerated for figure quality improvement, and differ somewhat in quantitative results, however, qualitative results (correct classification of KDM2B samples, outlier IDDSELD sample) remain the same.

We next tested the sensitivity and specificity of the episignature using a support vector machine. For each sample, we determined a methylation variant pathogenicity (MVP) score between zero and one based on matching the *KDM2B* episignature. All *KDM2B* samples included in the training set received scores >0.8 while control samples remained near zero, indicating high sensitivity for the detection of the *KDM2B* episignature (Supplementary Figure 3-S3C). Specificity was tested using a similar classifier that was instead trained against a large number of samples with confirmed diagnoses of a non-*KDM2B* related NDD from our Episign knowledge database. 75% of both case and control samples were used for training the classifier with the remaining 25% reserved for testing (Figure 3-2D). Case samples again scored high (>0.85) while the remainder of samples scored low (<0.5), with few exceptions. The most notable exceptions are cerebellar ataxia, deafness, and narcolepsy (ADCADN; OMIM 604121), a disorder associated with *DNMT1;* Hunter–McAlpine syndrome (HMA; OMIM# 601379), associated with *NSD1* ; and Dystonia 28, childhood-onset (DYT28 ; OMIM# 617284), associated with *KMT2B* (also see Discussion). One other sample amongst the control samples did score remarkably high for the *KDM2B* signature (Figure 3-2D, red arrow head). This sample was previously diagnosed with intellectual developmental disorder with seizures and language delay (IDDSELD , OMIM# 619000), a disorder caused by variants in *SETD1B* . This gene is located close to *KDM2B* at 12q24, and upon closer investigation we identified this sample to originate

from a case reported to have a 12q24.31 deletion which does not include *KDM2B* but might affect regulatory regions [32, 35]. For reference, we included the clinical description of this individual (#33, Extended Data Table 3-2).

Of note, our cohort contains two cases (#29 and #30) that carry larger deletions encompassing both *KDM2B* and *SETD1B* (Table 3-1, Extended Data Table 3-1 and 3-2) who were previously reported based on the *SETD1B* deletion and its associated episignature [32]. In these two samples, we have additionally identified the *KDM2B* signature. Another sample showing co-existing episignatures is #3.1, which was previously diagnosed with Phelan-McDermid syndrome (PHMDS, OMIM 606232) due to a 22q13 deletion, which was also confirmed based on the respective episignature. We here thus identify the *KDM2B* related signature associated with a *KDM2B* nonsense variant as well. These results indicate that multiple episignatures can coexist in a single individual and that the method is able to correctly identify both syndromes independently.

A CXXC-DOMAIN SPECIFIC EPISIGNATURE DISCRIMINATES BETWEEN LOF AND CXXC MISSENSE CARRIERS

Our cohort consists of both coding-altering and LoF variants, and we observed a remarkable clustering of variants in the CxxC domain (Figure 3-1A). One could therefore hypothesize that CxxC missense variants exert different or additional effects as compared to LoF variants. We therefore asked whether CxxC missense variants resulted in a different episignature. To this end, we performed the same analysis as before based on a selection of five samples (Figure 3-3C; Extended Data Table 3-1) carrying missense variants in the CxxC domain for which methylation data was available. The resulting probe set was then used for the hierarchical clustering of these CxxC samples and the LoF samples within the cohort. Interestingly, this probeset correctly differentiates between all *KDM2B* variant carriers (i.e. including the LoF variants) and controls (Figure 3-3B and 3-3C), suggesting that LoF and CxxC variants affect the same genomic regions. However, this probe set also segregates CxxC missense samples from LoF variant samples (Figure 3-3B and 3-3C), indicating that CxxC variants have a distinct impact on DNA methylation as compared to LoF variants. Most notably, 106 hypermethylated probes amongst the 107 significant probes selected for the CxxC-trained episignature present

with an on average increased methylation level even exceeding that of the hypermethylation probes of the pan-KDM2B probeset (mean methylation difference of all hypermethylated probes: 16.56% ±4.21 vs. 10.38% ±3.68, respectively; Figure 3-2A and Figure 3-3A). Similar, yet less pronounced, results were achieved using the LoF variant samples to train for probe selection (Supplementary Figure 3-S4). In conclusion, CxxC missense variants cause a distinct episignature that is associated with increased hypermethylation levels.

A

pvalueAdjustMethod: BH; score: 1000; methDiffValue: 0.05; rocValue: 500; corValue: 0.75; corMethod: 1
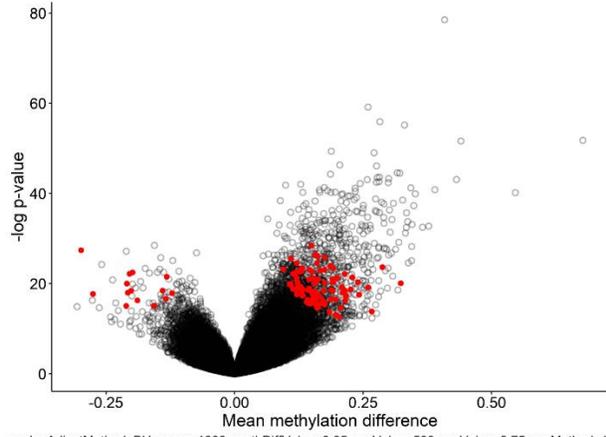
B
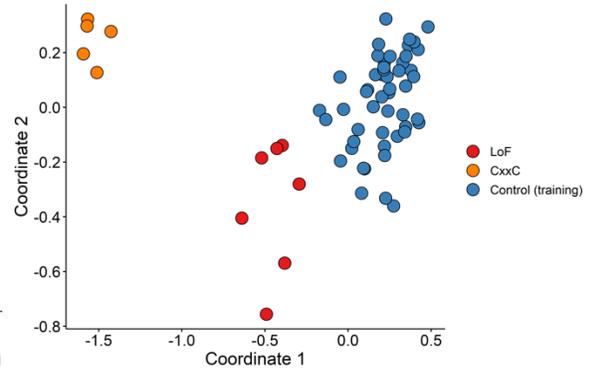
C
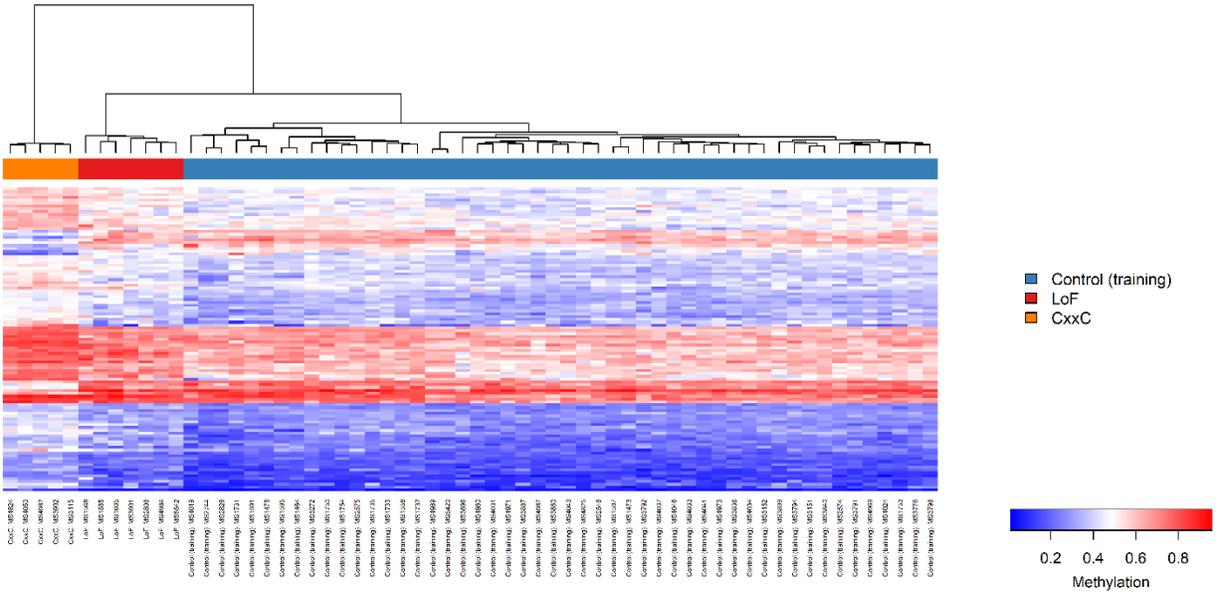
D
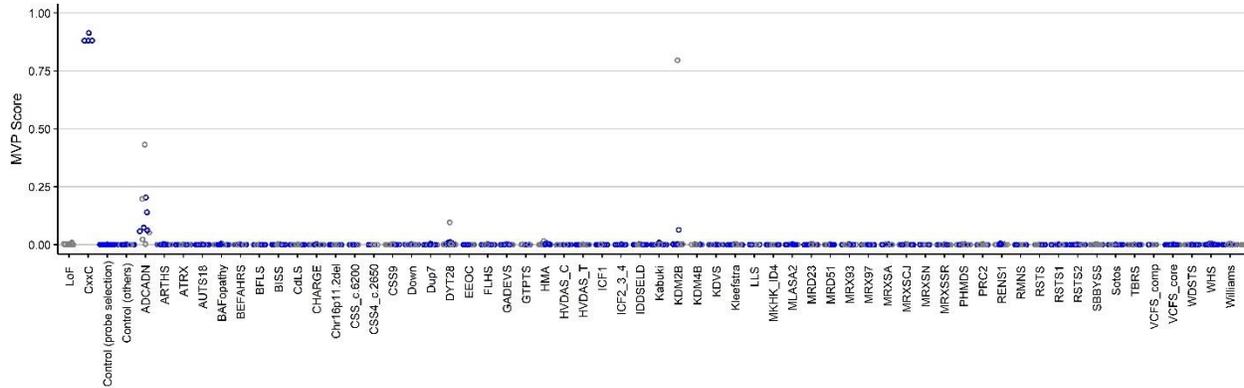
Figure 3-3: A CxxC-variant specific episignature. All samples representing a CxxC-coding variant and included in the *KDM2B* episignature training set, were used to train a CxxC-variant specific episignature. **A:** Volcano plot indicating all selected probes (red) included in the CxxC episignature. **B:** Multidimensional scaling (MDS) plot for selected probes, representing the pairwise distance across CxxC variants (orange), LoF variants (red) and controls (blue), based on the top two dimensions. **C:** Heatmap of selected probes and unsupervised hierarchal clustering results indicating the episignature's ability to decipher *KDM2B* variant carriers (red and orange) from controls (blue), and to decipher CxxC variants (orange) from LoF variants (red). **D:** Support Vector Machine (SVM) classifier indicating specificity of the CxxC episignature. Graph as in Figure 3-2D.

CLASSIFICATION OF KDM2B VARIANTS BASED ON EPISIGNATURE

Based on the pan-*KDM2B* episignature (Figure 3-2), we established a functional effect for 13 variants within our cohort, represented by 15 individuals. Four variants tested negative for the signature. Among the negative samples are two missense variants of which the *a priori* prediction was doubtful (i.e. p.Arg766Gln & p.Ile652Val, Table 3-1 and Extended Data Table 3-1). Although the remaining missense (p.Arg1213Trp) is *de novo* and predicted damaging by all algorithms, two alternative substitutions (i.e. p.Arg1213Gln & p.Arg1213Leu) in five individuals are reported in the gnomAD database. The other negative sample is that of the only splice-site variant in our cohort, possibly indicating the predicted splice effects do not occur, or at least not to a level that interferes with gene functionality. These four variants we consider to remain VUSs, as a negative episignature result does not suffice to infer an absence of functional effects. Based on the ACMG/AMP criteria [36], incorporating the functional evidence provided by the

episignature (criterium PS3) [3], we classified the 13 positive variants as pathogenic (Table 3-1, Extended Data Table 3-1).

As we now consider *KDM2B* as an established disease gene (including arguments from the shared clinical phenotypes described below), we set out to re-classify the variants that were not tested or returned inconclusive results for the episignature, based on the ACMG/AMP guidelines [36] (Extended Data Table 3-1). Importantly, we considered the CxxC-domain as an established hotspot for pathogenic variation in *KDM2B* (criterium PM1), as all tested *de novo* variants within this domain returned positive for the episignature. Amongst the seven variants not tested for the episignature, we re-classified two variants as pathogenic, as they involve *de novo* variants located within the CxxC-domain and are absent from the gnomAD database. In addition, one of these variants (p.Gly638Asp, #20) affects a residue at which a different substitution was confirmed pathogenic based on the episignature (p.Gly638Ser, #1). One variant was re-classified as likely pathogenic (p.K635del, #13); although this variant affects the CxxC-domain as well, it represents the only in-frame deletion in our cohort and no additional functional evidence for pathogenicity can be collected. The four remaining variants we consider VUSs as they are either inherited or of unknown inheritance, are not located in the CxxC-domain and/or are reported in the gnomAD database. In conclusion, based on the ACMG/AMP guidelines and supported by the functional evidence provided by the episignature, we classified 15 variants as pathogenic, one variant as likely pathogenic and eight variants remain VUSs (Table 3-1, Extended Data Table 3-1, Supplementary Figure 3-5).

CLINICAL FEATURES IN (LIKELY) PATHOGENIC KDM2B-VARIANT CARRIERS

We next determined the clinical phenotypes associated with the novel *KDM2B*-related syndrome. Clinical data of all individuals were systematically collected (Table 3-2 and Extended Data Table 3-2) and detailed clinical histories for all individuals are available as supplemental material. To prevent confounding the clinical presentation, we here limit the clinical description to patients that: 1) carry a pathogenic or likely pathogenic variant affecting *KDM2B*; and 2) for whom *KDM2B* represents the only identified genetic disorder. Within this group of 15 individuals, all presented with speech delay, developmental delay (DD), learning difficulties and/or ID. Behavioral concerns such as autism-spectrum-disorder (ASD) and attention-deficit hyperactivity-disorder (ADHD) are common (9/14). Growth parameters were within the normal

range for the majority. We observe several congenital defects, including congenital heart defects (CHD; 7/15), unilateral kidney agenesis (4/15) and ophthalmological anomalies (6/14). Two patients had cryptorchidism, two had epilepsy.

We collected facial photographs of 12 individuals, but no recognizable facial gestalt could be identified by an experienced dysmorphologist (RO) (Figure 3-4). Facial features noted in several individuals with CxxC-domain variants were a broad nasal tip, large ear lobes, and exaggerated Cupid's bow. Interestingly, in the individuals with LoF variants the nose was often more prominent, with a narrow nasal ridge and malar flattening with the exception of #10, who also has a diagnosis of Noonan syndrome.

The *KDM2B* related phenotype thus presents as a NDD of variable expression. Most common features include DD/ID , behavioral abnormalities, congenital defects and facial dysmorphisms. We propose to refer to this novel syndrome as 'KDM2B-related NDD' (K2BNDD).
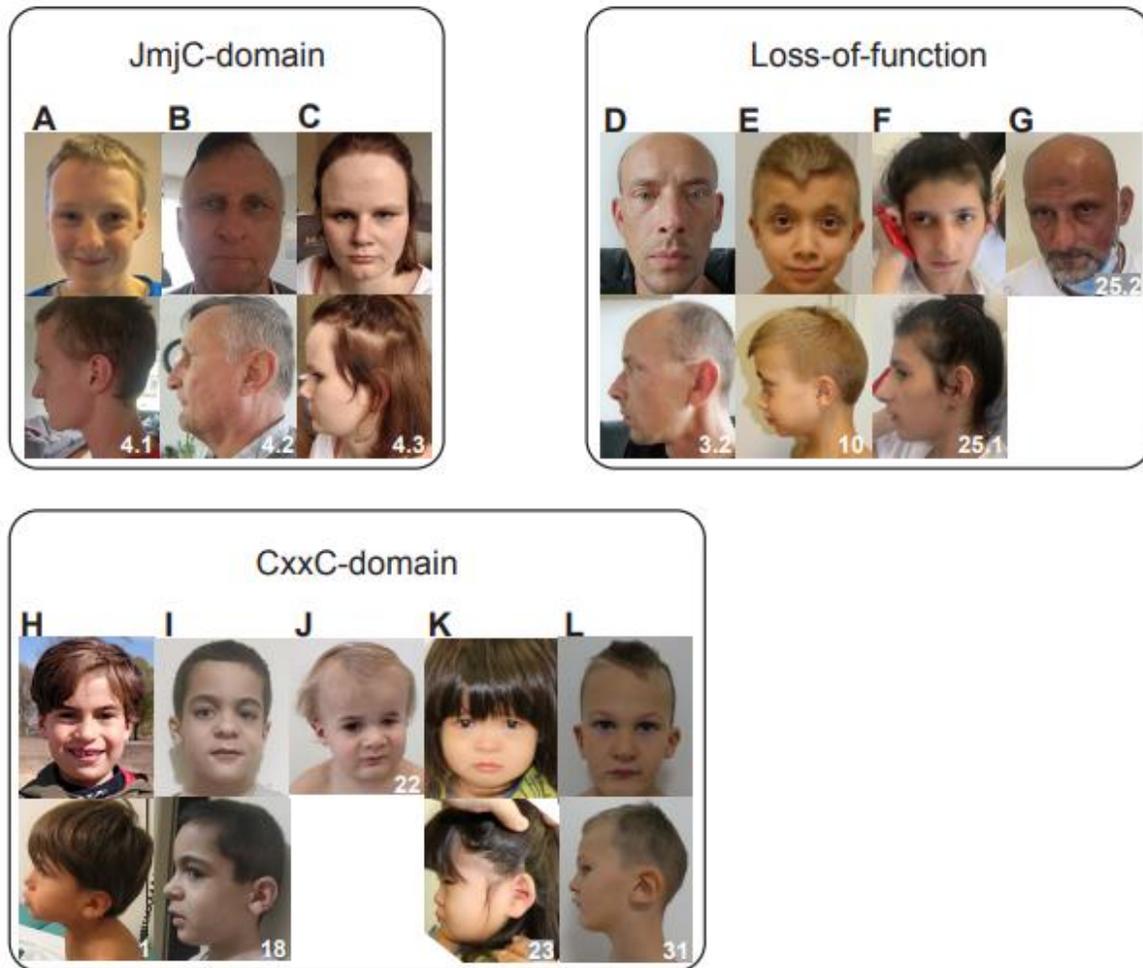
Figure 3-4: Facial photographs of individuals with *KDM2B* pathogenic variants. A-C) Individuals of family 4 with the p.Val316Ile variant, located in the JmjC-domain. D-G) Individuals with loss-of-function variants. Individual 10 (E) is also affected with Noonan syndrome H-L) Individuals with missense variants in the CxxC domain.

POTENTIAL GENOTYPE PHENOTYPE ASSOCIATIONS

As our methylation analysis revealed differences between CxxC and LoF variants, we next performed genotype-based patient stratification in order to support detection of possible genotype-phenotype relationships. Unfortunately, the current number of individuals available for analysis is limited, precluding the establishment of a genotype-phenotype relationship. Importantly, only two LoF variant carriers are not confounded by additional findings (i.e. dual diagnosis or larger deletion affecting multiple genes). We note however that unilateral kidney

agenesis and eye anomaly (coloboma/anophthalmia/congenital obstruction ductus nasolacrimales) was only reported in those individuals carrying a CxxC variant. In addition, congenital heart defects were present in six individuals with a CxxC variant, and only in two with a LoF variant. Of note, both these LoF carriers were diagnosed with another, both of which are associated with congenital heart defects. We thus note that congenital organ anomalies might be overrepresented in CxxC variant carriers, however the currently limited number of available cases precludes drawing any conclusions. Epilepsy did not occur in association with the CxxC domain variants, and did occur in one patient with a JmjC domain variant, one patient with a frameshift variant and two patients with a 12q24.31 deletion.

DISCUSSION

We describe a novel NDD caused by heterozygous pathogenic variants in *KDM2B*, which encodes a well-studied epigenetic regulator with multiple molecular functions, including the demethylation of H3 lysine residues. We collected a cohort of 33 individuals with 24 heterozygous *KDM2B* variants and performed genome-wide methylation profiling. In 15 samples, representing 13 unique *KDM2B* variants, we identified a shared episignature. We utilized this episignature to re-classify the variants in our cohort based on the ACMG/AMP criteria [36] and conclude that 16 variants are (likely) pathogenic. In line with other MDEMs, pathogenic variant carriers present with variable phenotypic expression, including DD/ID, congenital organ anomalies and/or facial dysmorphisms. We refer to this novel syndrome as *KDM2B*-related NDD (K2BNDD).

Given that K2BNDD presents as a heterogeneous disorder with variable severity and phenotypes, larger cohorts are needed to fully encompass the phenotypes associated with the disorder. Furthermore, the limited number of individuals representing potential groups for genotype-phenotype associations precludes drawing conclusions in this regard. We do however observe potential hints towards such associations, especially for CxxC-domain variant carriers and the presence of congenital anomalies. A more severe phenotypic expression of CxxC-

variants would be in line with the enhanced DNA methylation levels observed in this subgroup as compared to LoF-carriers (Figure 3-3).

The episignature identified for K2BNDD adds to the expanding toolbox offered by methylation profiling. Not only can these signatures be used to provide functional evidence to support the evaluation of VUSs; the increasing number of syndromes for which episignatures have been established enables the diagnosis of uncharacterized individuals, as well as the identification of novel pathogenic variants through pinpointing the causal gene [3]. The results presented here furthermore emphasize its utility by demonstrating that multiple episignatures can coexist within a single individual. Moreover, our analysis identified the K2BNDD related signature in a case previously diagnosed with IDDSELD (Figure 3-2D). Upon closer investigation, we noticed this sample carries a larger deletion on the 12q24 region, directly affecting the coding region of *SETD1B*, but not *KDM2B*. We hypothesize therefore that this deletion affects a regulatory region, causing mis regulation of *KDM2B* and thereby the K2BNDD-signature. Alternatively, the presence of two samples with a deletion encompassing both *KDM2B* and *SETD1B* in our episignature training cohort might cause the signature to contain some traces of the IDDSELD signature as well. This sample remains of interest for further investigation.

In line with the overall clinical performance of EpiSign testing [3], the *KDM2B* episignature enables specific and sensitive detection of K2BNDD. For this episignature, we noticed MVP scores over 0.25 for three other disorders (Figure 3-2D). The first is associated with ADCADN, which is caused by mutations in *DNMT1*, a methyltransferase known as the central player in the maintenance of CpG methylation [2, 37]. Interestingly, *DNMT1* has been suggested to regulate H3K4 methylation, providing a direct mechanistic link with *KDM2B* [38]. The second is HMA , a syndrome associated with duplication of 5q35 [39,40]. This region includes *NSD1*, which encodes a lysine methyltransferase known to methylate H3K36 [41], providing a direct functional link between HMA to K2BNDD as well. Finally, DYT28 is associated with *KMT2B* [42], encoding another methyltransferase reported to methylate H3K4 [43], again suggesting a direct link with *KDM2B*. All three disorders might therefore be associated with dysregulation of the same molecular process as for K2BNDD, and as such the same genomic regions might be affected in all three disorders. Of note, the K2BNDD,

93

ADCADN, HMA and DYT28 episignatures are all characterized by hypermethylation [2, 44, 45]. Alternatively, therefore, the elevated MVP scores might reflect a set of loci sensitive to hypermethylation irrespective of the underlying mechanisms. Phenotypically, K2BNDD shares features with HMA, e.g. mild – moderate delay, congenital heart defects and dysmorphism, but present differently from ADCADN and DYT28. Future studies will have to determine if and how K2BNDD, ADCADN, DYT28 and HMA are related, and might provide valuable insights into the etiology of these disorders.

The majority of pathogenic variants are of *de novo* origin, however in three families the variant was inherited from a mildly or unaffected parent (families 3,4 & 25). The more severe presentation in the children could be explained by a second diagnosis in the child, as was identified in family 3. Individual 25.2 is of special interest, he has a 12q24.31 deletion encompassing *KDM2B* and the episignature is present, however does not seem to be clinically affected. Individual #4.2 appears more mildly affected as compared to his affected children as well. These observations might be explained by multiple hypotheses. First, the parents might be mosaic carriers, resulting in a smaller percentage of affected cells and thereby reduced expression of the phenotypes. Alternatively, all inherited pathogenic variants originate from the father, possibly indicating that males are affected less severely. In mice, *Kdm2b* has been shown to be involved in X-chromosome silencing [15], and as such a different clinical expression in males versus females seems plausible. Future studies will have to inform on which hypotheses are true, or whether different explanations underlie these observations.

In line with hypermethylation defining the episignature, elevated DNA methylation levels have also been observed in a mouse *Kdm2b* knock-out model [15]. In addition, our cohort contains several truncating variants and gene deletions, and *KDM2B* constraint metrics indicate the gene is intolerant for LoF variants in the healthy population [29]. Collectively, these observations argue that K2DNDD is likely caused by haplo-insufficiency of *KDM2B*. Coding-altering variants associated with K2DNDD are therefore expected to cause LoF as well. The clustering of coding-altering variants in the CxxC-domain however suggests additional mechanisms by which variants in this domain cause disease. Supporting this notion is the identification of a distinct episignature associated with CxxC-variants and the elevated levels of hypermethylation defining this episignature. These observations are suggestive of a (partial) dominant-negative effect of CxxC-domain associated variants. Interestingly, the CxxC-domain

94

has been specifically implicated in the developmental functions of *KDM2B*, as CxxC mutants fail to rescue cellular differentiation induced by *Kdm2b* depletion in mESC [7] and specific deletion of the CxxC-domain induces developmental defects in the heterozygous state, whereas heterozygous knockouts appear healthy [18, 46]. Future studies will have to determine the molecular mechanisms by which both LoF and CxxC-specific variants lead to hypermethylation and how these mechanisms relate to the associated phenotypes.

In summary, we have delineated a novel syndrome that is caused by heterozygous *KDM2B* variants and characterized clinically by DD/ID, behavioral challenges including autism and ADHD, congenital anomalies mainly of the heart, urogenital system and eyes, and variable facial dysmorphism. KDM2B directly affects gene expression by epigenetic processes, and affected individuals show a distinct episignature. As such, K2BNDD represents a novel addition to the emerging group of MDEMs [1]. The signature can aid in reclassification of VUSs, and the detection of K2BNDD missed during routine diagnostic testing, e.g. due to intronic variants. We observed both *de novo* and dominantly inherited pathogenic variants. As the latter might easily be overseen by standard trio-ES based diagnostic testing, we suggest including the gene in relevant gene panels in order to facilitate the identification of inherited variants.

MATERIALS AND METHODS

ETHICAL CONSIDERATIONS AND APPROVAL, INCLUSION CRITERIA AND DATA COLLECTION

All individuals were included after informed consent forms, stating they agree to participate in research efforts and do agree with publication of their clinical and genetic data, as well as photos for relevant cases, were signed and received by the respective institutions. Patient privacy was respected during the exchange of data amongst researchers and/or clinicians. This study was approved by the medical ethical committee installed by the University Medical Centre Utrecht (TCBIO 20/714, March 18th, 2021).

Individuals were included based on the identification of a heterozygous *KDM2B* suspected to be pathogenic based on *in silico* predictions and/or inheritance. Individuals carrying bi-allelic VUSs were not considered for this study. Individual 1 was the index patient, after a

KDM2B variant was annotated to be of interest after diagnostic trio-ES. Families 2-5 were included after local, in-house database searches. All remaining individuals were included after personal communication, literature search, or resulting from searches using the Genematcher platform [30]. For the published cases, we contacted the original authors for updated clinical information. For all individuals, clinical and genetic data was collected through a standardized spreadsheet which was completed by the respective physicians and/or researchers.

GENETIC VARIANT DETECTION

Variants in individuals/families 24, 25, 29, 30 and 34 were identified as described before [26,28,31,32]. The variant of individual 4.3 was identified by targeted Sanger sequencing. All other variants were detected through clinical and/or research-based exome-sequencing.

ANALYSIS AND CLASSIFICATION OF KDM2B VARIANTS

Structural analysis of variants was performed with Pymol (The Pymol Molecular Graphics System, Version 2.5 Schrödinger, LLC). All figures were generated with Pymol. Four *in silico* prediction algorithms were consulted: SIFT , Metadome, MutationTaster and Polyphen-2 [47, 48, 49, 50]. All variants were manually analyzed using Alamut Visual v2.15 (Sophia Genetics). Variants were classified according to the 2015 ACMG/AMP guidelines [36]. Episign results were used as criterium PS3; PM1 as applied for CxxC-domain variants.

EPISIGN METHODS

METHYLATION ARRAY AND QUALITY CONTROL

DNA methylation analysis and Episignature classifier development was performed using previously established protocol [34, 51, 52, 53]. Stored genomic DNA samples extracted from peripheral blood, previously used for genomic sequencing, were used for bisulfite conversion and hybridization to the Illumina infinium methylation EPIC bead chip arrays, according to manufacturer's protocol. Idat files, containing methylated and unmethylated signal intensity plots (beta values) were produced from these microarrays, and used for analysis in R 4.0.2. Normalization was performed using the Illumina Infinium methylation EPIC array with background correction from the minfi package [54]. Previously defined exclusion criteria [52,53] were used to exclude probes with detection p values >0.01, probes on the x and y chromosomes, probes known to contain SNPs at the site of CpG interrogation or single nucleotide extension, and probes known to cross react with chromosomal locations other than their target regions were removed. All samples were examined for genome wide methylation distribution and those deviating from a bimodal distribution were excluded. Factor analysis using a principal component analysis (PCA) was performed to examine batch effect and identify outliers.

DNA METHYLATION PROFILING

Probe methylation levels (beta values), were calculated as the ratio of signals intensity in methylated probes vs total sum of unmethylated and methylated probes, resulting in values ranging from zero to one. To allow for linear regression modeling, beta values were logit transformed using the limma package [55] , allowing for the identification of differentially methylated probes. Data was adjusted for the blood cell type composition as per Houseman et al [56]. Estimated blood cell proportion was added to the model matrix of the linear models as confounding variables [57]. Using the eBayes function in the limma package [58], p values were moderated and corrected for multiple testing using the Benjamini Hochberg method. Probes with the most significant methylation differences are selected using two facts from this dataset, the level of methylation difference (relative methylation signal intensity), and the probability that an observed difference is due to random chance (p values). Evaluation of this interaction is carried out by multiplying the absolute methylation difference between affected cases and controls by

the negative value of the log transformed p values, and ranking the top 1000 probes with the highest values from this transformation. Next, receiver operating characteristic analysis (ROC) is performed on each probe, to measure the pairwise correlation coefficient between probes. Probes with low area under curve values from ROC analysis are removed, as well as highly correlated probes, eliminating probes with low sensitivity and specificity, and probes with highly correlated characteristics using Pearson's correlation coefficient. This ensures that the final probeset contains the most differentiating, non-redundant probes that are not influenced by random data structures. Only probes with a methylation difference greater than 5% were included in this analysis. This probe filtering process was designed to avoid reporting of probes with low effect size, and those influenced by technical or random variations as conducted in previous studies [52,53].

SELECTION OF MATCHED CONTROLS FOR METHYLATION PROFILING

For episignature characterization, mapping of probes and feature selection, matched controls were randomly selected from the LHSC EpiSign Knowledge Database (EKD) [52]. All of the KDM2B samples were assayed, therefore all the controls selected for episignature identification were analyzed using the same array type. Samples were matched by age, sex and batch using the MatchIt package. A 4:1 ratio of controls to cases was deemed optimal for this analysis, as previously described [34]. Principal Component Analysis (PCA) analysis was performed after each attempt at matching to detect outliers and determine data structures for the presence of batch effect. Outlier samples, and those with highly aberrant data structures were removed, and subsequent matching trials were performed until consistent iterations with no outliers in the first two components of the PCA were derived.

CLUSTERING AND DIMENSION REDUCTION

Hierarchical clustering and multidimensional scaling were used after each iteration of analysis to examine the data structure of the identified episignature. Hierarchical clustering was performed using Ward's method on Euclidean distance by the base stats package in R, and visualized with the ggplot2 package [59, 60]. Multidimensional scaling provides a visual

98

representation of sample methylation profile similarity based on the scaling of the pairwise Euclidean distances between each sample. Observations of study samples' methylation profiles at this stage allowed for further refinement of the cohort used for probe selection training.

DISCOVERY/TRAINING COHORT SELECTION

Identification of disease specific episignatures was performed using a randomly selected sub-setting of the database, on a 75:25 ratio of discovery:training, using the caTools package in R. Testing samples were used to assess the performance of the classification model developed later in the study. For every disease group in the discovery cohort, a sex and age matched control group with a sample size at least 4 times larger was selected from the reference control group using the MatchIT package, and methylation profiles were compared between the two.

CROSS VALIDATION

For each round of validation, one of the selected KDM2B samples was removed from probe selection, alongside matched controls. The remaining KDM2B samples were designated as testing samples, and all three groups were modeled using multidimensional scaling to determine how they cluster/segregate with one another. This process was repeated with different combinations of assigned training and testing samples until all cases had been removed from probe selection and used for testing once.

EPISIGNATURE CLASSIFICATION MODEL

Specificity of the episignature was assessed using the Methylation Variant Pathogenicity (MVP) score, using all the identified probes. A support vector machine (SVM) used a linear kernel for training on KDM2B cases and controls. Once again, a 4:1 ratio of controls to cases was used to divide both the case and control samples previously matched and used for probe selection into training and testing cohorts for the SVM. Furthermore, the remaining unselected samples from the EKD were also divided similarly (75% training, 25% testing) to allow for comparison and testing of signature robustness against all of the samples in the EKD. Using the e1071 R package, we performed 10-fold cross validation to determine hyperparameters optimal for episignature classification. In this process, the training set was divided into ten folds by random assignment, where the first nine are used for training, and the last used for testing the

accuracy of the model. The mean accuracy over all rounds was then calculated, and hyperparameters with the best performance by this metric were selected. The model provides a score ranging from 0-1 for each subject, representing the model's confidence in predicting whether the subject has a DNA methylation profile similar to the KDM2B probe set or not. Conversion of these SVM decision values was done using Platt's scaling method [61], and the class obtaining the greatest score determined the predicted phenotype. A classification as KDM2B was made when a sample received the greatest score for that class (normally greater than 0.5). Finally, the model was applied to both a training set of a large cohort of individuals with clinical and molecular diagnoses of neurodevelopmental disorders, as well as a group of healthy controls to determine its effective specificity.

VALIDATION OF EPISIGN CLASSIFICATION

To ensure the model is not susceptible to the batch structure of the methylation experiment, the classifier was applied to samples assayed on the same batch as the cases used for training. Using methylation data from individuals without a confirmed diagnosis of KDM2B within the EKD assayed on the same microarray chip as case samples, methylation profiles were modeled to ensure the classifier is not confounded by technical artifacts unique to the given microarray. Specificity was determined by supplying a large number of DNA methylation arrays from unaffected subjects to the model. To further assess the specificity of the KDM2B classifier relative to other neurodevelopmental disorders we applied it to cases with other patient cohorts exhibiting distinct episignatures within the EKD.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

RO designed and coordinated the study. RO, RvJ, CC collected and analyzed clinical and genetic data. JR, JK, ML, JH, BS, and MA were involved in creating the episignature. MH depicted the 3D domain structures. RvJ, JR, CC and RO wrote the manuscript. All authors contributed to data generation or analysis and reviewed/edited the manuscript.

REFERENCES

1. Fahrner JA, Bjornsson HT. Mendelian disorders of the epigenetic machinery: postnatal malleability and therapeutic prospects. Human Molecular Genetics. 2019;28(R2):R254-R264. doi:10.1093/hmg/ddz174

2. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. Am J Hum Genet. 2020;106(3):356-370. doi:10.1016/j.ajhg.2020.01.019

3. Sadikovic B, Levy MA, Kerkhof J, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. Genet Med. 2021;23(6):1065-1074. doi:10.1038/s41436-020-01096-4

4. Frescas D, Guardavaccaro D, Bassermann F, Koyama-Nasu R, Pagano M. JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. Nature. 2007;450(7167):309-313. doi:10.1038/nature06255

5. He J, Kallin EM, Tsukada YI, Zhang Y. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). Nat Struct Mol Biol. 2008;15(11):1169-1175. doi:10.1038/nsmb.1499

6. He J, Nguyen AT, Zhang Y. KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. Blood. 2011;117(14):3869-3880. doi:10.1182/blood-2010-10-312736

7. He J, Shen L, Wan M, Taranova O, Wu H, Zhang Y. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat Cell Biol. 2013;15(4):373-384. doi:10.1038/ncb2702

8. Janzer A, Stamm K, Becker A, Zimmer A, Buettner R, Kirfel J. The H3K4me3 histone demethylase Fbxl10 is a regulator of chemokine expression, cellular morphology, and the metabolome of fibroblasts. J Biol Chem. 2012;287(37):30984-30992. doi:10.1074/jbc.M112.341040

9. Kang JY, Kim JY, Kim KB, et al. KDM2B is a histone H3K79 demethylase and induces transcriptional repression via sirtuin-1-mediated chromatin silencing. FASEB J. 2018;32(10):5737-5750. doi:10.1096/fj.201800242R

10. Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet. 2006;7(9):715-727. doi:10.1038/nrg1945

11. Tsukada Y ichi, Fang J, Erdjument-Bromage H, et al. Histone demethylation by a family of JmjC domain-containing proteins. Nature. 2006;439(7078):811-816. doi:10.1038/nature04433

12. Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet. 2006;7(9):715-727. doi:10.1038/nrg1945

13. Vacík T, Laďinović D, Raška I. KDM2A/B lysine demethylases and their alternative isoforms in development and disease. Nucleus. 2018;9(1):431-441. doi:10.1080/19491034.2018.1498707

14. Fukuda T, Tokunaga A, Sakamoto R, Yoshida N. Fbxl10/Kdm2b deficiency accelerates neural progenitor cell death and leads to exencephaly. Mol Cell Neurosci. 2011;46(3):614-624. doi:10.1016/j.mcn.2011.01.001

15. Boulard M, Edwards JR, Bestor TH. Abnormal X chromosome inactivation and sex-specific gene dysregulation after ablation of FBXL10. Epigenetics Chromatin. 2016;9:22. doi:10.1186/s13072-016-0069-1

16. Deiktakis EE, Abrams M, Tsapara A, et al. Identification of Structural Elements of the Lysine Specific Demethylase 2B CxxC Domain Associated with Replicative Senescence Bypass in Primary Mouse Cells. Protein J. 2020;39(3):232-239. doi:10.1007/s10930-020-09895-z

17. Xu C, Liu K, Lei M, et al. DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. Structure. 2018;26(1):85-95.e3. doi:10.1016/j.str.2017.11.022

18. Blackledge NP, Farcas AM, Kondo T, et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell. 2014;157(6):1445-1459. doi:10.1016/j.cell.2014.05.004

19. Han XR, Zha Z, Yuan HX, et al. KDM2B/FBXL10 targets c-Fos for ubiquitylation and degradation in response to mitogenic stimulation. Oncogene. 2016;35(32):4179-4190. doi:10.1038/onc.2015.482

20. Inagaki T, Iwasaki S, Matsumura Y, et al. The FBXL10/KDM2B scaffolding protein associates with novel polycomb repressive complex-1 to regulate adipogenesis. J Biol Chem. 2015;290(7):4163-4177. doi:10.1074/jbc.M114.626929

21. Jiang Y, Qian X, Shen J, et al. Local generation of fumarate promotes DNA repair through inhibition of histone H3 demethylation. Nat Cell Biol. 2015;17(9):1158-1168. doi:10.1038/ncb3209

22. Marcon E, Ni Z, Pu S, et al. Human-Chromatin-Related Protein Interactions Identify a Demethylase Complex Required for Chromosome Segregation. Cell Reports. 2014;8(1):297-310. doi:10.1016/j.celrep.2014.05.050

23. Liang G, He J, Zhang Y. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. Nat Cell Biol. 2012;14(5):457-466. doi:10.1038/ncb2483

24. Charng WL, Karaca E, Coban Akdemir Z, et al. Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. BMC Med Genomics. 2016;9:42. doi:10.1186/s12920-016-0208-3

25. Faundes V, Newman WG, Bernardini L, et al. Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am J Hum Genet. 2018;102(1):175-187. doi:10.1016/j.ajhg.2017.11.013

26. Girard SL, Gauthier J, Noreau A, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nat Genet. 2011;43(9):860-863. doi:10.1038/ng.886

27. Monies D, Abouelhoda M, Assoum M, et al. Lessons Learned from Large-Scale, First-Tier Clinical Exome Sequencing in a Highly Consanguineous Population. Am J Hum Genet. 2019;104(6):1182-1201. doi:10.1016/j.ajhg.2019.04.011

28. Yokotsuka-Ishida S, Nakamura M, Tomiyasu Y, et al. Positional cloning and comprehensive mutation analysis identified a novel KDM2B mutation in a Japanese family with minor malformations, intellectual disability, and schizophrenia. J Hum Genet. 2021;66(6):597-606. doi:10.1038/s10038-020-00889-4

29. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

30. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36(10):928-930. doi:10.1002/humu.22844

31. Chouery E, Choucair N, Abou Ghoch J, El Sabbagh S, Corbani S, Mégarbané A. Report on a patient with a 12q24.31 microdeletion inherited from an insulin-dependent diabetes mellitus father. Mol Syndromol. 2013;4(3):136-142. doi:10.1159/000346473

32. Krzyzewska IM, Maas SM, Henneman P, et al. A genome-wide DNA methylation signature for SETD1B-related syndrome. Clin Epigenetics. 2019;11(1):156. doi:10.1186/s13148-019-0749-3

33. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2

34. Aref-Eshghi E, Bend EG, Colaiacovo S, et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. The American Journal of Human Genetics. 2019;104(4):685-700. doi:10.1016/j.ajhg.2019.03.008

35. Qiao Y, Tyson C, Hrynchak M, et al. Clinical application of 2.7M Cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability. Clin Genet. 2013;83(2):145-154. doi:10.1111/j.1399-0004.2012.01860.x

36. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-424. doi:10.1038/gim.2015.30

37. Winkelmann J, Lin L, Schormair B, et al. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. Hum Mol Genet. 2012;21(10):2205-2210. doi:10.1093/hmg/dds035

38. Sun L, Huang L, Nguyen P, et al. DNA Methyltransferase 1 and 3B Activate BAG-1 Expression via Recruitment of CTCFL/BORIS and Modulation of Promoter Histone Methylation. Cancer Res. 2008;68(8):2726-2735. doi:10.1158/0008-5472.CAN-07-6654

39. Hunter AGW, Dupont B, McLaughlin M, et al. The Hunter-McAlpine syndrome results from duplication 5q35-qter. Clin Genet. 2005;67(1):53-60. doi:10.1111/j.1399-0004.2005.00378.x

40. Hunter AG, McAlpine PJ, Rudd NL, Fraser FC. A "new" syndrome of mental retardation with characteristic facies and brachyphalangy. J Med Genet. 1977;14(6):430-437. doi:10.1136/jmg.14.6.430

41. Rayasam GV, Wendling O, Angrand PO, et al. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. EMBO J. 2003;22(12):3153-3163. doi:10.1093/emboj/cdg288

42. Zech M, Boesch S, Maier EM, et al. Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. Am J Hum Genet. 2016;99(6):1377-1387. doi:10.1016/j.ajhg.2016.10.010

43. Demers C, Chaturvedi CP, Ranish JA, et al. Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. Mol Cell. 2007;27(4):573-584. doi:10.1016/j.molcel.2007.06.022

44. Ciolfi A, Foroutan A, Capuano A, et al. Childhood-onset dystonia-causing KMT2B variants result in a distinctive genomic hypermethylation profile. Clin Epigenetics. 2021;13(1):157. doi:10.1186/s13148-021-01145-y

45. Kernohan KD, Cigana Schenkel L, Huang L, et al. Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. Clin Epigenetics. 2016;8:91. doi:10.1186/s13148-016-0254-x

46. Andricovich J, Kai Y, Peng W, Foudi A, Tzatsos A. Histone demethylase KDM2B regulates lineage commitment in normal and malignant hematopoiesis. The Journal of Clinical Investigation. 2016;126(3):905. doi:10.1172/JCI84014

47. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-249. doi:10.1038/nmeth0410-248

48. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361-362. doi:10.1038/nmeth.2890

49. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016;11(1):1-9. doi:10.1038/nprot.2015.123

50. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Hum Mutat. 2019;40(8):1030-1038. doi:10.1002/humu.23798

51. Aref-Eshghi E, Bend EG, Hood RL, et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. Nat Commun. 2018;9(1):4885. doi:10.1038/s41467-018-07193-y

52. Aref-Eshghi E, Rodenhiser DI, Schenkel LC, et al. Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. Am J Hum Genet. 2018;102(1):156-174. doi:10.1016/j.ajhg.2017.12.008

53. Bend EG, Aref-Eshghi E, Everman DB, et al. Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. Clin Epigenetics. 2019;11(1):64. doi:10.1186/s13148-019-0658-5

54. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049

55. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. doi:10.1093/nar/gkv007

56. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86. doi:10.1186/1471-2105-13-86

57. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. PLOS ONE. 2012;7(7):e41361. doi:10.1371/journal.pone.0041361

58. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3. doi:10.2202/1544-6115.1027

59. Ward JH. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association. 1963;58(301):236-244. doi:10.1080/01621459.1963.10500845

60. Wilkinson L. ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. Biometrics. 2011;67(2):678-679. doi:10.1111/j.1541-0420.2011.01616.x

61. Advances in Large-Margin Classifiers | MIT Press eBooks | IEEE Xplore. Accessed March 13, 2022. https://ieeexplore.ieee.org/book/6267437

SUPPLEMENTAL TABLES AND FIGURES
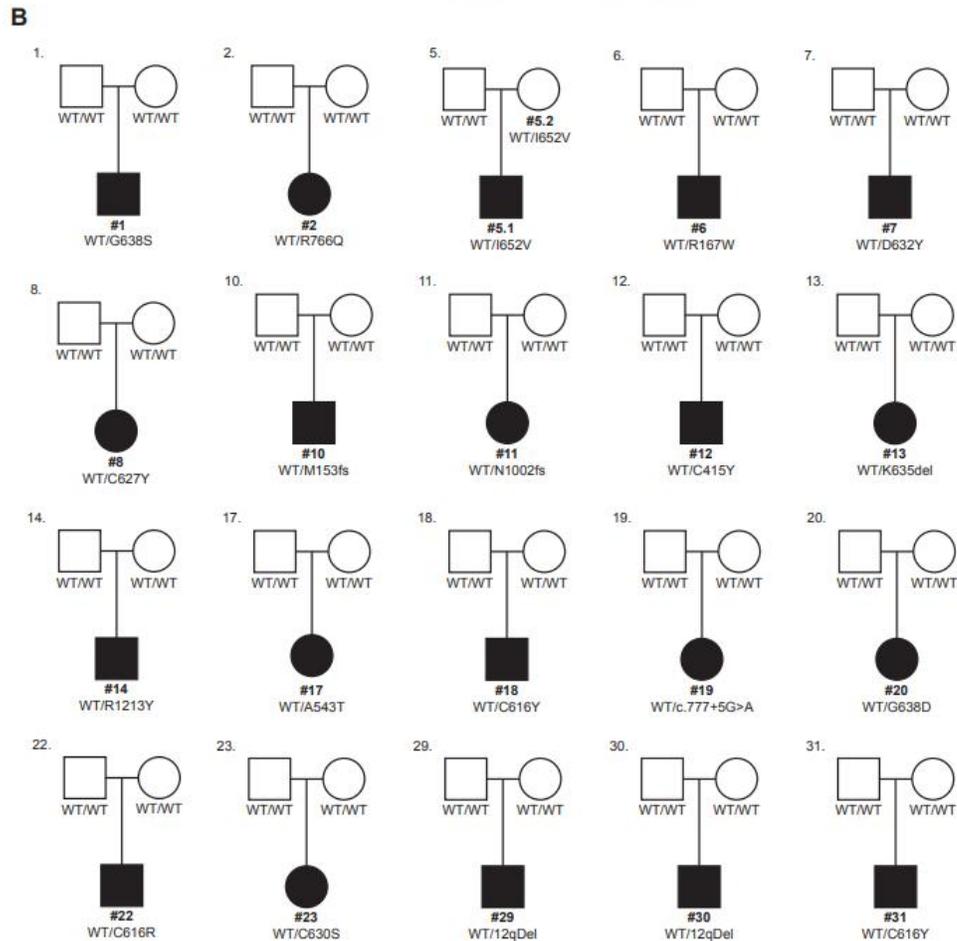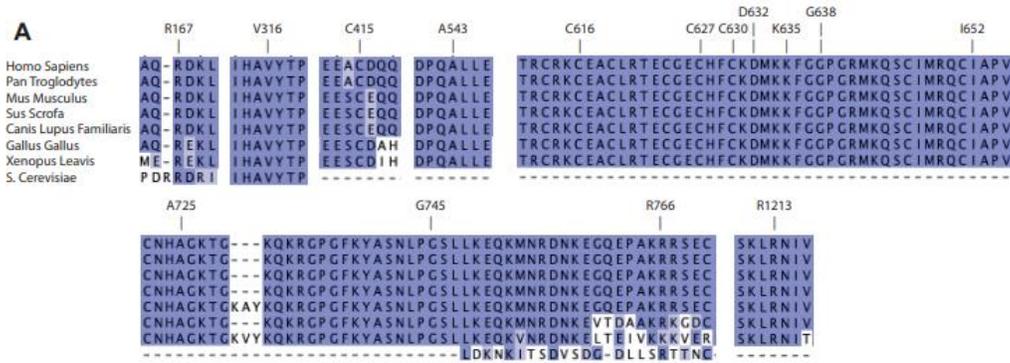
Supplementary Figure 3-1: A cohort of heterozygous KDM2B variant carriers. A) Conservation of *KDM2B* residues affected by coding altering variants in the cohort. Purple color intensity is indicative of conservation across the species included in the figure. B) Pedigrees for the individuals not included in Figure 3-1A and/or not published before. Family 24 has been published elsewhere (Yokotsuka-Ishida et al., 2021).
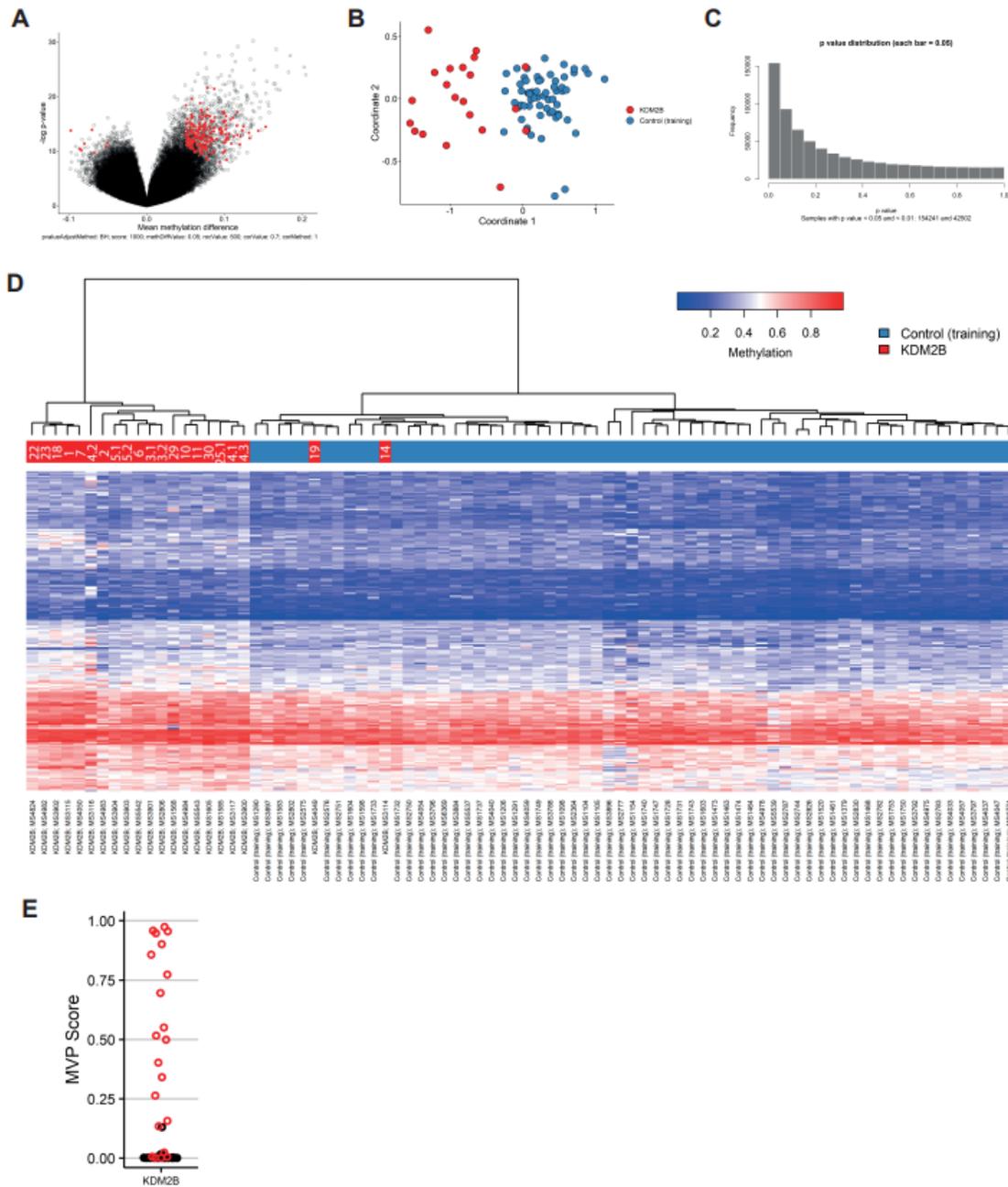
Supplementary Figure 3-2: A shared episignature amongst the majority of *KDM2B* variant carriers. All samples for which methylation array data was available were used to train an episignature. A) Volcano plot indicating all selected probes (red) included in the signature. B) Multidimensional scaling (MDS) plot for selected probes, representing the pairwise distance across samples (red) and controls (blue), based on the top two dimensions. C) Distribution of the (raw) p-values of all probes. D) Heatmap of selected probes and unsupervised hierarchal clustering results indicating the clustering of the majority of variant carriers (red) apart from controls (blue). E) Cross-validation of the episignature. Each test sample (red) was removed from the training cohort and was subsequently tested against the resulting classifier.
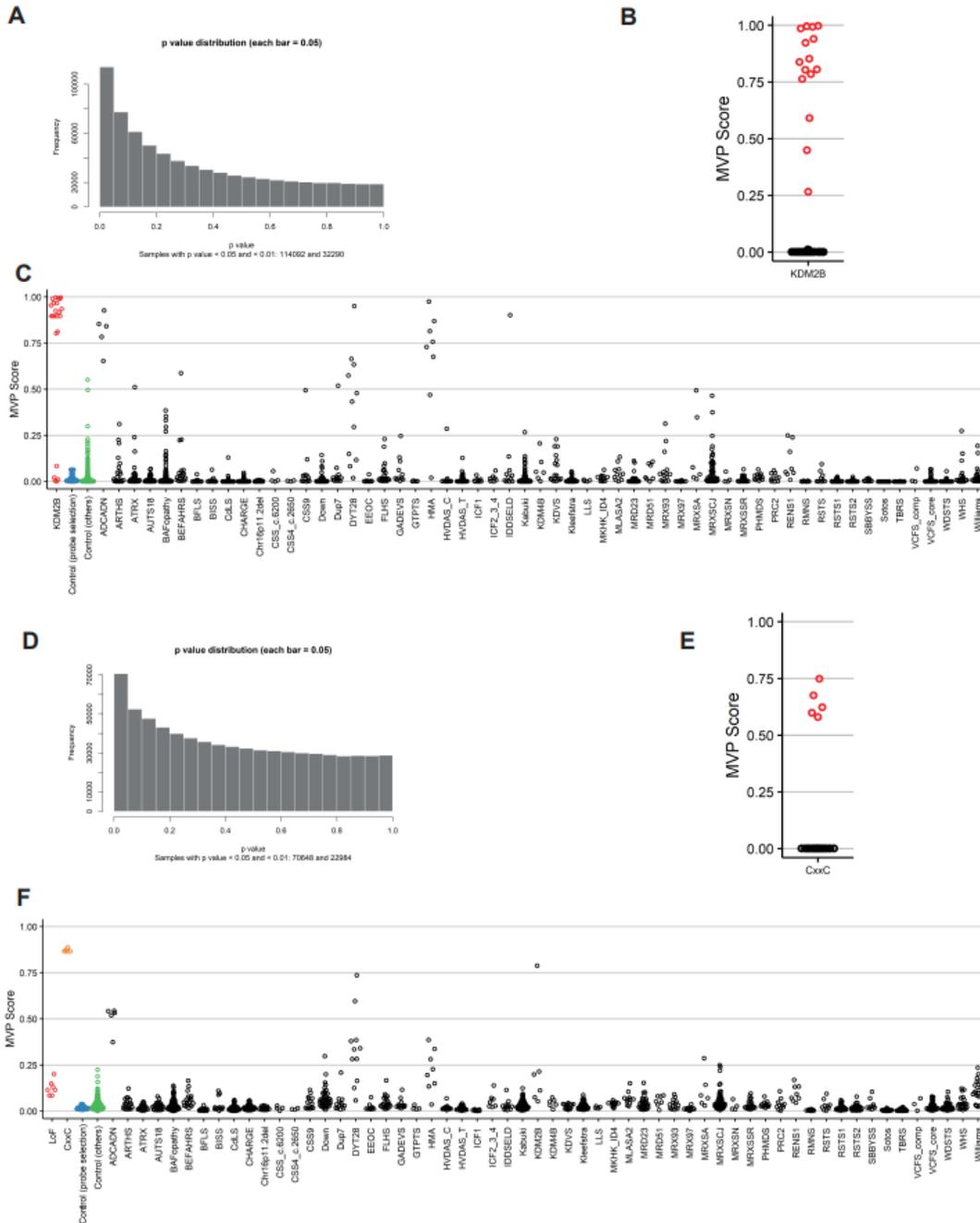
Supplementary Figure 3-3: Supporting data for figures 3-2 and 3-3. A) Distribution of the (raw) p-values of all probes for the samples used to train the *KDM2B* episignature (Figure 3-2). B) Cross-validation of the *KDM2B* episignature. Each test sample (red) was removed from the training cohort and was subsequently tested against the resulting classifier. C) Support Vector Machine (SVM) model trained on the samples included in Figure 2. All case samples and matched controls were used for training. Next, each sample was tested against the provided classifier. D) As (A), E) as (B) and F) as (C) for the CxxC episignature (Figure 3-3).

110
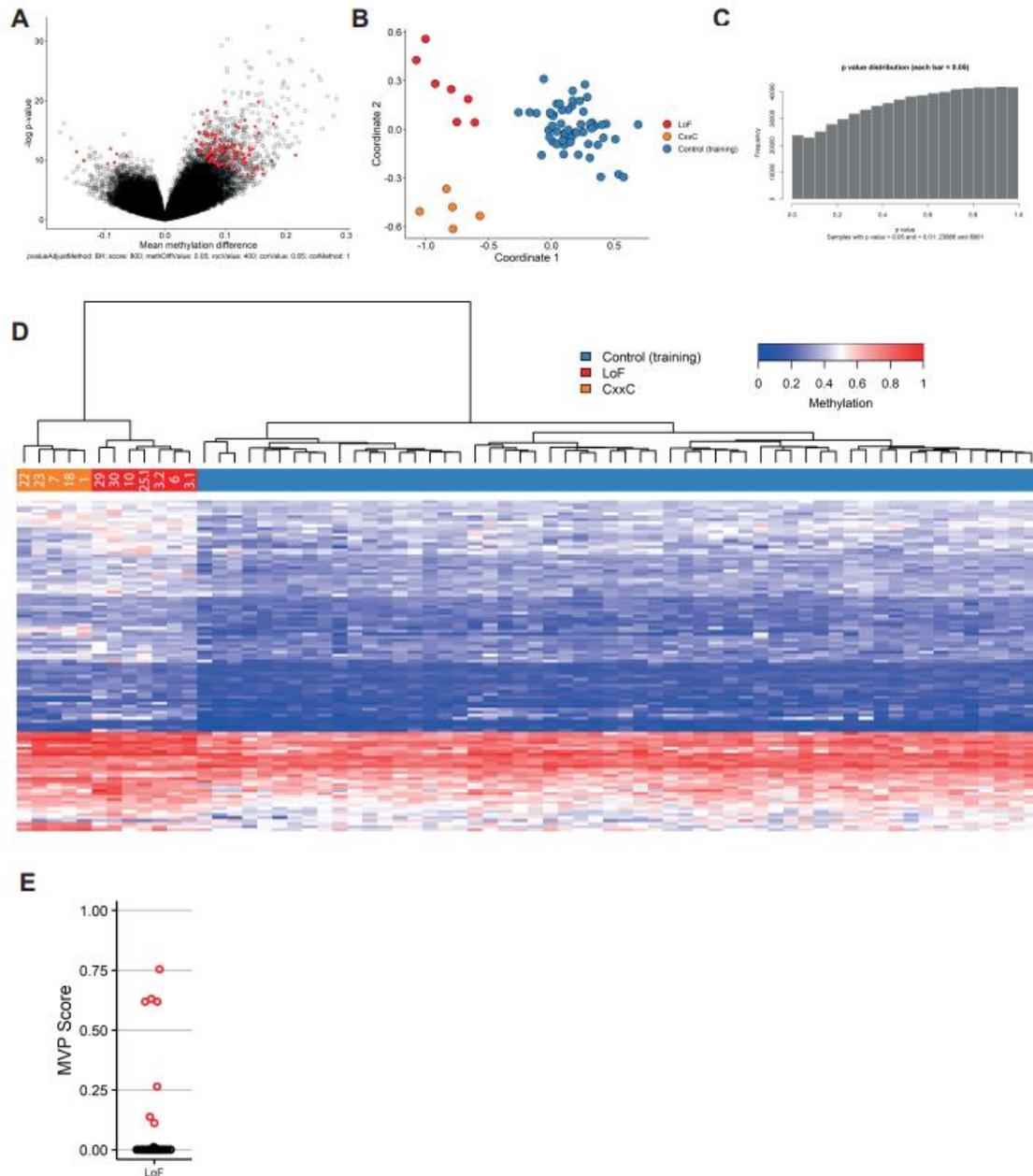
Supplementary Figure 3-4: Identification of a LoF associated episignature. All samples representing a loss-of-function (LoF) variant and included in the *KDM2B* episignature, were used to train a LoF-variant specific episignature. A) Volcano plot indicating all selected probes (red) included in the LoF episignature. B) Multidimensional scaling (MDS) plot for selected probes, representing the pairwise distance across CxxC variants (orange), LoF variants (red) and controls (blue), based on the top two dimensions. C) Distribution of the (raw) p-values of all probes. D) Heatmap of selected probes and unsupervised hierarchal clustering results indicating the episignature's ability to decipher *KDM2B* variant carriers (red and orange) from controls (blue), and to decipher CxxC variants (orange) from LoF variants (red). E) Cross-validation of the LoF episignature. Each test LoF sample (red) was removed from the training cohort and was subsequently tested against the resulting classifier.

111
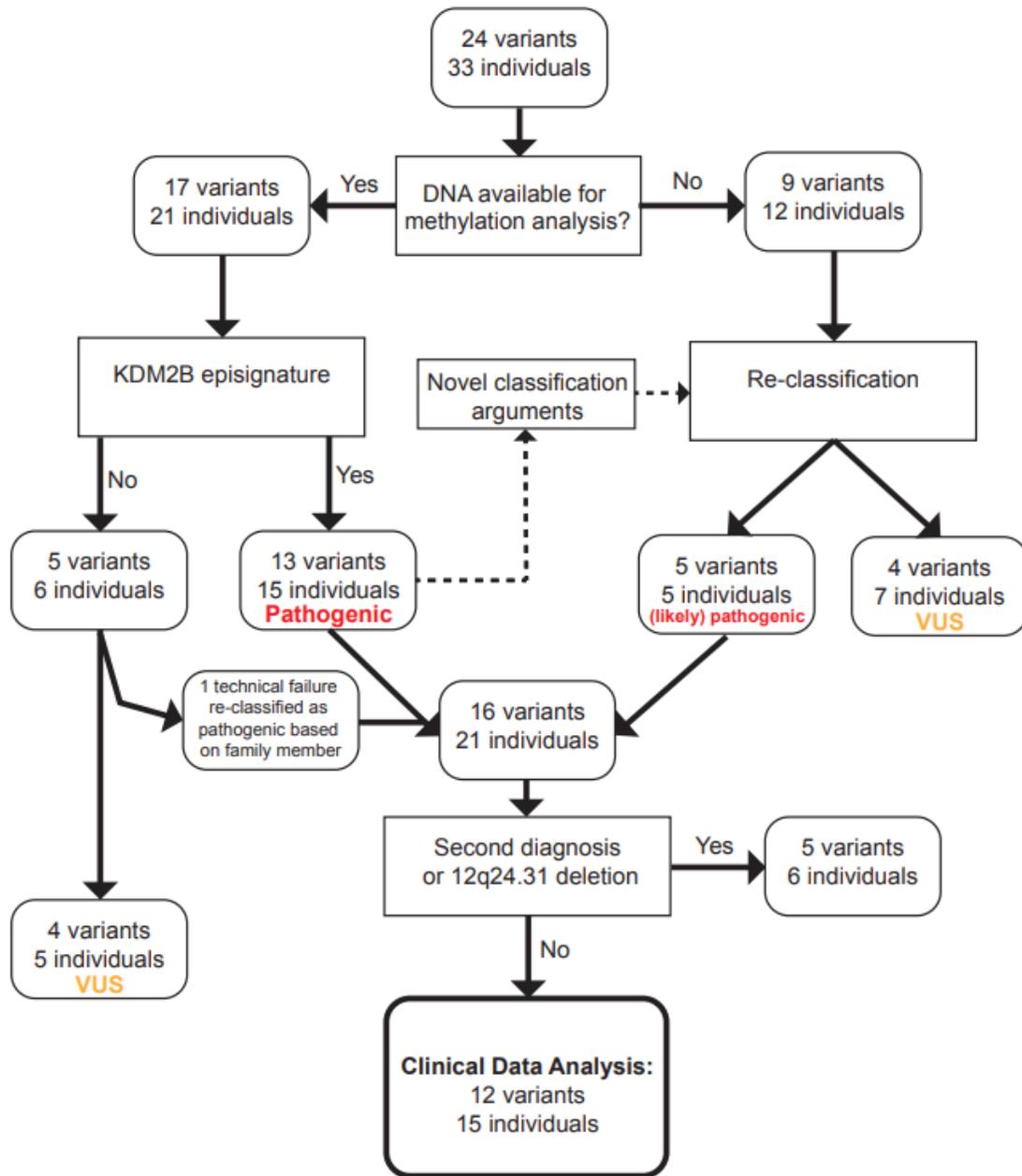
Supplementary Figure 3-5: Patient inclusion flow chart. Flow chart indicating which samples/variants were used in the different sub-cohorts used throughout this study.
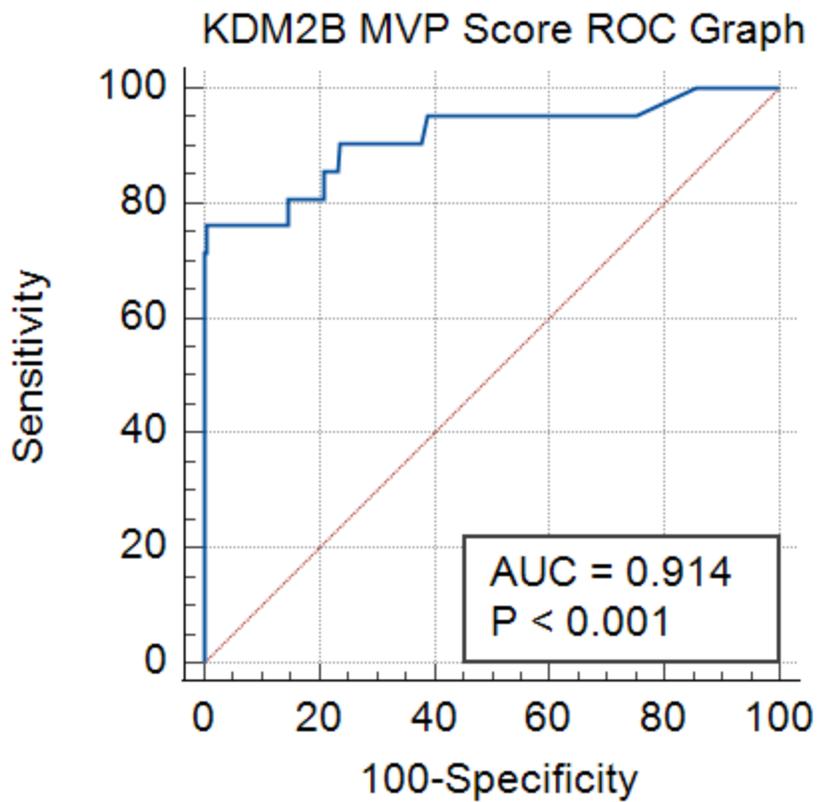
Supplementary Figure 3-6: KDM2B MVP Score ROC Graph. Receiver operating characteristic curve demonstrating the sensitivity and specificity of the generated MVP scores for the KDM2B cohort and the remaining EKD samples used for training.

CONCLUSION

      This chapter provides a definitive example of episignature substratification briefly hypothesized in the first chapter, demonstrating the additional episignatures that can be derived through analysis of specific domains. In the 2019 paper from Bend et al concerning the identification of multiple episignatures within the ADNP gene sequence [53], researchers determined two specific episignatures within the ADNP sequence, one associated with disruptions of the central region of the gene, and another associated with the terminal ends of the sequence. This provided evidence towards the existence of specific DNA methylation patterns that result from disruptions of different loci within a given gene sequence, and helped guide my analysis towards the identification of further substratified episignatures within a cohort of patients with common genetic origins.

      This research proved fruitful upon the analysis of the KDM2B cohort, and showed that domain specific episignatures can exhibit a high level of sensitivity and specificity, and provide interesting insights into the key mechanisms required for gene function. When a specific domain is disrupted, analysis of the epigenome, phenotype and genotype can help compare and contrast the affected individuals carrying mutations in different gene domains. Through this comparison, we can identify how these domains interact with DNA methylation profiles, and the ensuing phenotype through their presentation in not just case samples experiencing disruptions of the domain in question, but cases presenting without domain disruption, and unaffected controls. This approach can prove highly effective in the context of NDDs which are associated with complex variable phenotypes, which may be better explained by the functional evidence that domain specific episignatures provide, tying the variation in presentation to a more focused view of genes and their important domain machineries.

      Most importantly, this work provides evidence of a novel NDD. These findings can also be applied to the reclassification of variants involved in the KDM2B sequence, representing functional information that can identify evidence for pathogenicity in variants of unknown significance. Within ClinVar, a number of such VUS's have been reported amidst other KDM2B variants (9/36, 25%, See Appendix Table 2) which could be reclassified following modeling with the derived KDM2B signature.

Further work is required to understand exactly how the disruption of the CxxC binding domain results in the distinct methylation pattern observed in the KDM2B cohort, but given the functional nature of this domain as a DNA binding motif, we can hypothesize the effect is due to loss of specific DNA binding mediated by the CxxC domain. Functional studies, and an in-depth analysis of how the KDM2B gene and it's CxxC domain interact with DNA to perform demethylase activities will potentially shed light into the exact mechanisms of this interaction, and explain the phenotypic differences observed in the affected patients. Large scale changes in DNA methylation observed in the CxxC patients could result in gene expression changes leading to the increased incidence of congenital anomalies not seen in cases with variants outside this domain. In this way, I describe, using the KDM2B cohort as an example, the analytical process of identifying additional sub stratified episignatures within a single gene sequence, which has resulted in an effective DNA methylation biomarker for a novel NDD, and insights into the functioning and molecular relationships of KDM2B's gene domains.

Chapter 4: Episignature assessment provides high sensitivity and specificity biomarker detection for patients with *KAT6A* and *KAT6B* sequence variants

Niels Vos MD [1], *Jack Reilly [2], M.W. Elting, MD,PhD [1], Philippe M Campeau MD [3], Coman D MD[4,5], Zornitza Stark MD[6,7], Tiong Yang Tan MD[6,7], David J Amor MD[8,9], Benjamin A Kamien MD[10], Chirag Patel MD[11], Matthew L Tedder PhD[12], Giuseppe Merla PhD[13,14] Paolo Prontera MD[15], Marco Castori MD, PhD[16], Kai Muru PhD[17], Felicity Collins MD[18,19], Janine Smith MD [20,21], Bruria Ben Zeev MD [22], Alessandra Murgia MD, PhD[23], Emanuela Leonardi PhD[23], Natacha Esber [24], Antonio Martinez-Monseny MD [25], Matthew Wallis MD [26] **Marielle Alders MD,PhD [1], **Bekim Sadikovic PhD [27]

[1] Department of Human Genetics, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, Netherlands

[2] Department of Pathology and Laboratory Medicine, Western University, London, Canada;

[3] Department of Pediatrics, Sainte-Justine UHC and University of Montreal, Montreal, Canada

[4] Department of Metabolic Medicine, Queensland Children's Hospital, Brisbane

[5] School of Medicine, University of Queensland, Brisbane, Australia

[6] Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Melbourne, Australia

[7] Department of Paediatrics, University of Melbourne, Melbourne, Australia

[8] Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Rd, Parkville, Australia

[9] University of Melbourne Department of Pediatrics, Parkville, Australia

[10] Genetics Services of Western Australia, Perth, Australia

[11] Genetic Health Queensland, Royal Brisbane & Women's Hospital, Brisbane, Australia

[12] Greenwood Genetic Centre, Greenwood, USA

[13] Laboratory of Regulatory and Functional Genomics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Gionvanni Rotonda (Foggia), Italy

[14] Department of Molecular Medicine and Medical Biotechnology, University of Naples Federica II,Naples, Italy

[15] Medical Genetics Unit, University of Perugia Hopsital SM della Misericordia, Perugia, Italy

[16] Division of Medical Genetics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Foggia), Italy

[17] Department of Clinical Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia

[18] Discipline of Child and Adolescent Health and Genomic Medicine, Sydney Medical School, Sydney University, Sydney, NSW, Australia

[19] Department of Clinical Genetics, Western Sydney Genetics Program, Children's Hospital at Westmead, Sydney, NSW, Australia.

[20] Sydney Children's Hospitals Network-Westmead, Sydney, Australia.

[21] University of Sydney, Sydney, Australia.

[22] Sackler School of Medicine Tel Aviv University, Tel Aviv, Israel

[23]. Laboratory of Molecular Genetics of Neurodevelopment, Department of Women's and Children's Health, University of Padua, Via Giustiniani 3, 35128, Padua, Italy

[24] KAT6A Foundation, 3 Louise Dr., West Nyack, NY 10994, USA.

[25] Genetics and Molecular Medicine Department, Rare Disease Pediatric Unit, Hospital Sant Joan de Déu, Barcelona, Spain.

[26] Tasmanian Clinical Genetics Service, Tasmanian Health Service, Royal Hobart Hospital, Hobart, Australia

[27] Molecular Genetics Laboratory, Molecular Diagnostics Division, London Health Sciences Centre, London, ON N6A 5W9, Canada

*Equal contribution with first listed author

**Corresponding and co-senior authors

PREFACE

Throughout this work, I have made reference to the intricate net of biological interactions that tie the genome, epigenome and ensuing phenotype of neurodevelopmental disorders together. In the context of a monogenic disorder, I have demonstrated the existence of multiple methylation profiles that accompany changes in phenotype in line with alternate genetic variants in terms of location, mechanism of effect, and type. It is therefore pertinent to discuss the effects of DNA methylation profiles in patients with multiple genetic origins, and how episignature profiles can guide assessment in the context of paralogous, but distinct molecular entities. As such, I have chosen to describe my research concerning the discovery of an episignature for patients with variants in the lysine acetyltransferase gene *KAT6A*, which shares a paralogous gene, *KAT6B*. These two genes share a common genetic origin, each acting as a subunit in the MOZ/MORF complex, and are associated with similar histone modifying activity however, disruptions of the *KAT6B* sequence are associated with two distinct NDDs, Genitopatellar syndrome (GTPTS) and Say-Beiber-Biesecker-Young-Simpson syndrome (SBBYSS), while *KAT6A* is associated with *KAT6A* Syndrome (AKA Mental Retardation Disorder 32 (MRD32), Arboleda-Tham Syndrome (ARTHS)). For GTPTS and SBBYSS, episignatures were derived in 2018, demonstrating that the disruption of lysine acetyltransferase activity can have distinct effects on the methylation profiles of affected patients [3], however, such an episignature has not been established for *KAT6A* so far. This context provides an interesting challenge to the derivation of an episignature, with shared function and genetic character between these two genes, but association with distinct syndromes that share considerable phenotypic overlap. These confounding factors provide an interesting complexity to our episignature assessment practices, posing several questions. To what degree can the DNA methylation profiles of these different disorders overlap? To what extent can the differences in DNA methylation be correlated to the distinct phenotypes seen in the three disorders? Finally, and most importantly, is there an effective way to fully differentiate the methylation pattern of KAT6A syndrome, even in the context of shared gene function and homology with GTPTS and SBBYSS? Data presented in this chapter will address these questions and elucidate the epigenetic assessment of a disorder while compensating for the confounding effects brought about by paralogous genes. This will provide evidence towards a syndrome specific episignature for KAT6A syndrome, as well as potentially

explaining the overlap in phenotypic characteristics of these lysine acetyltransferase disorders through the lens of epigenetics.

ABSTRACT

Accurate diagnosis for patients living with neurodevelopmental disorders is often met with numerous challenges, related to the ambiguity of findings and lack of specificity in genetic variants leading to pathology. Genome wide DNA methylation analysis has been used to develop highly sensitive and specific "episignatures" as biomarkers capable of differentiating and classifying complex neurodevelopmental disorders. In this manuscript, we describe distinct episignatures for *KAT6A* syndrome, caused by mutations in the lysine acetyltransferase A gene (*KAT6A*), and the lysine acetyl transferase B (*KAT6B*) associated with two other neurodevelopmental disorders. We demonstrate the ability of our models to differentiate between highly overlapping episignatures, increasing the ability to effectively identify and diagnose these conditions.

INTRODUCTION

Hereditary neurodevelopmental disorders (NDD) are relatively common, with a global prevalence of approximately 3-5% [1,2]. A thorough clinical examination and genetic testing can help pinpoint the underlying cause, but as more NDD-related genes are being discovered and examined, more genetic variants of unknown significance are being identified. Segregation analyses and functional assays can sometimes help determine pathogenicity [3,4], but in the majority of cases, clinicians, patients and families are left without definitive diagnoses.

Assessment of DNA methylation profiles provides novel possibilities to confirm diagnoses and explain pathophysiology. Epigenetics describes the heritable changes in gene expression without altering the underlying nucleotide sequence [5]. Epigenetic mechanisms include DNA methylation, histone modifications and the effects of non-coding RNAs (ncRNA) [6,7] Mutations in genes involved in chromatin regulatory processes have been implicated in a growing number of NDDs [3]. Predominantly related to neurobehavioral phenotypes, disorders of the epigenetic machinery have the potential to cause widespread disruption of developmental programs [8]. Chromatin regulatory genes, such as DNA methyltransferases play an important role in modulating cell differentiation during development [8]. Unique DNA methylation profiles have been described in patients with mutations in genes involved in epigenetic and chromatin regulating processes [9]. Highly sensitive and specific algorithms based on disease-associated, differentially methylated CpG dinucleotides, which are effectively detectible using microarray technology, are referred to as "episignatures" or "EpiSigns", and to date, over 50 have been described. These episignatures provide a sensitive and specific molecular technique for diagnosis and variant classification for a growing number of genetic conditions [3,10-22]. In certain conditions, a significant overlap has been described between the methylation profiles of multiple syndromes [20], as well as sub-gene level, or gene domain specific signatures. Recently, DNA methylation testing has been adapted in a clinical setting demonstrating a significant diagnostic yield and utility in diagnosis of Mendelian disorders [20,21].

Histone acetyltransferases (HATs) are enzymes that acetylate the lysine tails of histones. Multiple HAT complexes have been described, including the Gcn5 N-acetyltransferases *(GNAT)* and the MYST (*Moz/Morf, Ybf2/Sas3, Sas2* and *Tip60*) families. MOZ or Lysine (K) Acetyl Transferase 6A (*KAT6A,* OMIM#601408) and MORF or Lysine Acetyltransferase 6B (*KAT6B,* OMIM#605880) form the catalytic subunits of a protein complex (together with *BRPF1/2/3,*

ING5 and *hEAF6*) that specifically acetylates lysine residues on histone H3 tails [23,24,25]. This acetylation alters chromatin structure and generally results in a more active gene expression. A disruption of this *KAT6A/B* complex can result in widespread changes of gene expression. KAT6A syndrome, caused by pathogenic mutations in the *KAT6A* gene, was first described by Arboleda et al and Tham et al., in 2015 [26,27]. A reverse dysmorphology approach (first whole exome sequencing (WES), sanger sequencing and array comparative genomic hybridization (array-CGH) showing *KAT6A* mutations, followed by phenotyping) was used to first describe *KAT6A* syndrome in seven patients. A large follow-up cohort to these studies, from Kennedy et al. [28], identified truncating or nonsense variants, as well as a subset of missense variants located within highly conserved residues in the *KAT6A* sequence. Intellectual disability and speech delay were found in all patients; however the presentation was found to be more severe in those patients with truncating variants in the last two exons of the gene (exons 16 and 17) when compared to early truncating variants (located in exons 1-15) [26, 27,28]. It was found that 95% of late truncating mutations were rated as moderate or severe for intellectual disability, while 60% of the early truncating cases were classified as mild intellectual disability. Interestingly, mutations in the *KAT6B* gene, encoding the other catalytic subunit of the *KAT6A/B* protein complex, can result in multiple distinctive syndromes (Genitopatellar Syndrome (GTPTS, OMIM #606170), Ohdo Syndrome (Say-Beiber-Biesecker-Young-Simpson; SBBYSS variant, OMIM #603736), which share a significant amount of phenotypic overlap with KAT6A syndrome [29,30,31,32,33].

Research of diagnostic rates of genes involved in developmental disorders notes that *KAT6A* syndrome is a common cause of syndromic intellectual disability [1], indicating the importance of providing a reliable molecular biomarker for the identification and classification of *KAT6A* variants. With overlapping clinical features with several disorders as described above, particularly *KAT6B* related disorders, a specific molecular test, such as those we have demonstrated previously with episignatures, could provide a significant improvement to the accurate diagnoses of these conditions and can potentially help to explain the clinical features. In this paper we demonstrate that a unique *KAT6A* syndrome classifier can be derived through the assessment of DNA methylation patterns in patients with identified *KAT6A* variants, providing evidence of this episignatures ability to differentiate between case and control samples. Additionally, because of the partial overlap in clinical features and the similar roles of *KAT6A*

and *KAT6B* as catalytic subunits of their MYST family HAT protein complex, we investigated if clinical similarities and differences between these syndromes can be explained by their methylation profiles.

MATERIALS AND METHODS

SUBJECTS AND CONTROL COHORTS

DNA samples were extracted from peripheral blood of 21 individuals with clinical and molecular features of KAT6A syndrome (See Table 4-3), that are part of the EpiSign Knowledge Databases (EKD) [20] housed at the London Health Sciences Centre (LHSC; Canada), and recruited from Amsterdam University Medical Centers (AUMC, The Netherlands), University of Western Australia, University of Queensland (Australia), University of Sydney (Australia), Victoria Clinical Genetics Services (Australia), Royal Prince Alfred Hospital (Australia), University of Montreal (Canada), Tartu University Hospital (Estonia), Scientific Institute for Research, Hospitalization and Healthcare (IRCCS, Italy), Sheba Medical Centre (Israel) and Sant Joan De Déu Hospital Barcelona (Spain). All samples and records were de-identified. Additionally, 4 samples with confirmed *KAT6B* variants associated with GTPTS from University of Montreal, as well as 10 SBBYSS samples with associated *KAT6B* variants from collaborators at University of Montreal, AUMC and Greenwood Genetics Centre (USA), were included in our models to assess the overlap in methylation profiles between these adjacent conditions.

The research was conducted in accordance with the Declaration of Helsinki. Study protocol has been approved by the Western University Research Ethics Board (REB 106302). Informed consent was obtained by physicians for use of the clinical information of the described patients.

METHYLATION ARRAY AND QUALITY CONTROL

DNA methylation protocol, analysis and episignature construction were performed using a previously established protocol [3,12,21,34]. Genomic DNA was extracted from peripheral blood samples using standard techniques and underwent bisulfite conversion for analysis using the Illumina Infinium methylation EPIC bead chip arrays, according to manufacturer's protocol. Methylated and unmethylated signal intensity plots (beta values) were processed to obtain the idat files for analysis in R 4.0.2. Normalization of the Illumina Infinium methylation EPIC array data was carried out with background correction from the minfi package [35]. Exclusion criteria,

as previously defined [3,21] excluded probes with detection p values >0.01, probes on the x and y chromosomes, and those known to contain SNPs at the site of CpG interrogation or single nucleotide extension, and probes known to cross react with chromosomal locations other than their target regions were removed. All samples were examined for genome wide methylation density and those deviating from a bimodal distribution were excluded. Factor analysis using a principal component analysis (PCA) was performed to examine batch effect and identify outliers.

DNA METHYLATION PROFILING

Methylation levels for probes were measured as beta values, based on the ratio of intensity in methylated signals vs total sum of unmethylated and methylated signals from microarray analysis, represented as a value ranging from zero to one. These values were used for the biological interpretation and visualization of samples. To allow for linear regression modeling, beta values were logit transformed using the limma package [36], allowing for the identification of probes differentially methylated between cases and controls. Blood cell type compositions were used to adjust analysis using an algorithm developed by Houseman et al [37]. Estimated blood cell proportions were added to the model matrix of the linear models as confounding variables [38]. Using the eBayes function in the limma package, p values were moderated and corrected for multiple testing using the Benjamini Hochberg method. The most informative probes were selected using two factors from this dataset, the level of methylation difference (relative methylation signal intensity), and the probability that an observed difference is due to random chance (p values). The most informative probes are selected using two factors from this dataset, the level of methylation difference (relative methylation signal intensity) and the probability that an observed difference is due to random chance (p value distribution). Evaluation of this interaction is carried out by multiplying the absolute methylation difference between affected cases and controls by the negative value of the log transformed p values, and taking a list of 1000 probes with the highest values from this transformation. Next, receiver operating curve characteristic analysis (ROC) is performed on each probe, and measures pairwise correlation coefficient between probes. Probes with low area under the curve values from ROC analysis are removed, as well as highly correlated probes, eliminating probes with low sensitivity and specificity, and probes with highly correlated characteristics using Pearson's correlation coefficient. This ensures that the final probeset contains the most differentiating, non-

redundant probes that are not influenced by random data structures. Furthermore, only probes with a methylation difference greater than 5% were included in the analysis, as investigation of Illumina microarrays has shown that methylation values are prone to technical error when attempts are made to assess methylation differences below 5%. This probe filtering process was designed to avoid reporting of probes with low effect size, and those influenced by technical or random variations as conducted in previous studies [3,12].

SELECTION OF MATCHED CONTROLS FOR METHYLATION PROFILING

For mapping the episignature for probe and feature selection, matched controls were randomly selected from the LHSC EpiSign Knowledge Database (EKD) [3]. All *KAT6A* samples and controls were assayed using the Illumina EPIC array. Matching was done by age, sex and batch using the MatchIt package. A ratio of 4:1 of matched controls to cases was used for each analysis; previous efforts have found that increasing the ratio beyond this compromised the ability of the model to effectively match samples [12]. First assessments found that there was significant overlap between *KAT6A* case samples and the additional 14 *KAT6B* associated samples, (see Figure 4-1B, and 4-1C) and as such, these *KAT6B* samples were added as control samples for training the probe selection models. After each attempt at matching, a rudimentary PCA analysis was performed to detect outliers and determine data structures for possible batch effect and other characteristics. Outlier samples, and those with highly aberrant data structures were removed, and subsequent matching trials were performed until we achieved consistent iterations with no outliers detected in the first two components of the PCA. Overall, 4 samples were removed from analysis due to significant divergence in methylation patterns exhibited, identified in Table 4-1.
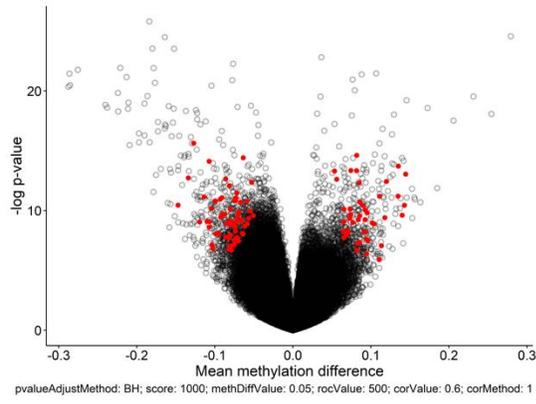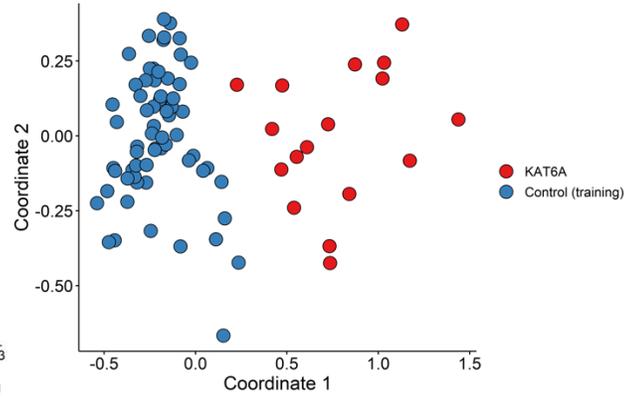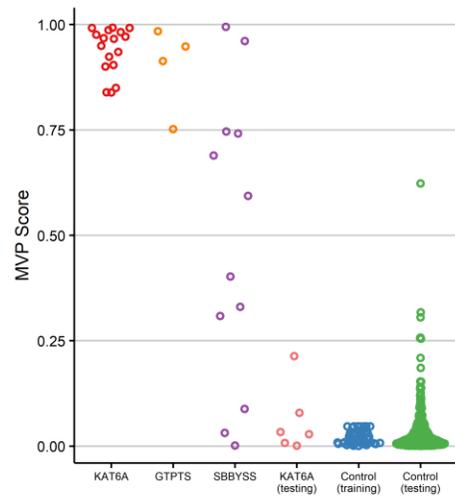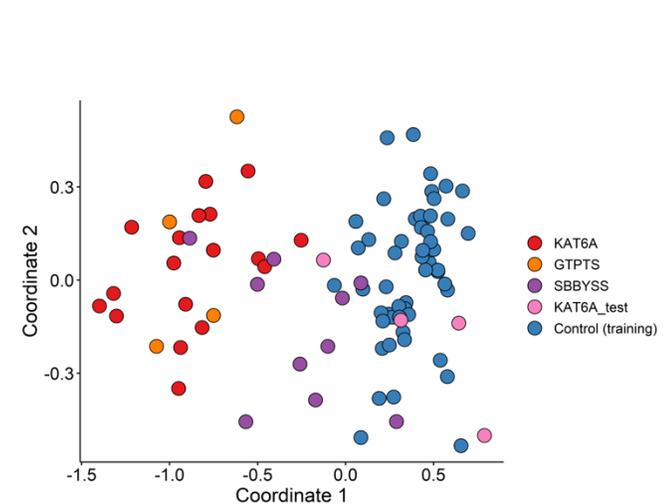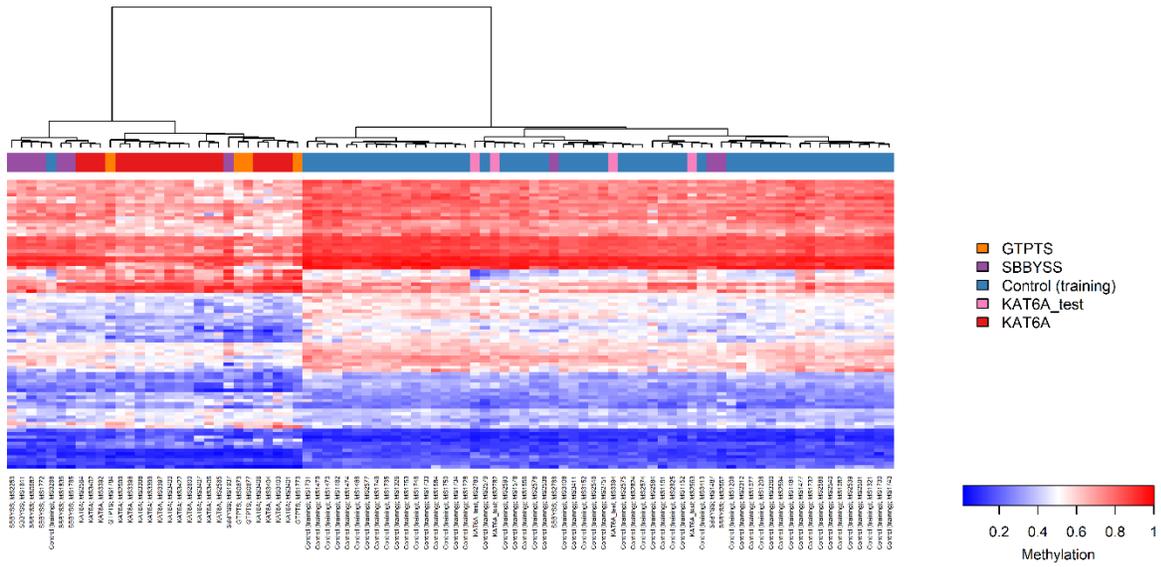
Figure 4-1. Models of Episignature Discovery trained on KAT6A samples and match controls

A: Bimodal distribution plot of mean methylation difference vs -log p-value for each probe, represented as circles on the plot. Probes highlighted in red indicate the probes chosen following preliminary analysis, wherein the most highly differentiated probes with statistically significant p-values are selected for representation

B: Multidimensional scaling plot representing the dimensions of variation in methylation signal intensity at informative CpG identified for *KAT6A*. Represents comparisons of the similarity of methylation profiles of *KAT6A* patients (marked in red) to control samples (marked in blue) which include cases without a confirmed phenotypic and/or genotypic presentation of *KAT6A*, including the added *KAT6B* samples and samples with confirmed presentation of other syndromes.

C: Targeted SVM classifier model for *KAT6A* without inclusion of *KAT6B* in training controls. Each sample receives scores for the probability of having a DNA methylation profile similar to cases as compared to controls. Higher value on Y axis indicates that a sample presents a methylation profile more similar to cases compared to controls. SBBYSS and GTPTS samples from the EpiSign Knowledge Database are plotted based on this relative scale of similarity to indicate probeset overlap between adjacent disorders. Additionally, *KAT6A* samples removed from analysis have been plotted, labelled as *KAT6A* (testing)

D: Multidimensional scaling plot representing the dimensions of variation in methylation signal intensity at informative CpG identified for *KAT6A* without inclusion of KAT6B in training controls. Represents comparisons of *KAT6A* patients with confirmed KAT6 related patients (Genitopatellar Syndrome-GTPTS, Say-Barber-Biesecker-Young-Simpson Syndrome- SBBYSS). Cases marked in red represent *KAT6A* cases, while purple represent SBBYSS cases, orange represent GTPTS cases, pink represents *KAT6A* samples removed from analysis, blue indicate cases with no phenotypic or genotypic presentation of any disorder (Control-Training), including samples with confirmed presentation of other syndromes.

E: DNA methylation signal intensity plot comparing confirmed KAT6A syndrome patients with confirmed *KAT6* related patients without inclusion of *KAT6B* in training controls (Genitopatellar Syndrome-GTPTS, Say-Barber-Biesecker-Young-Simpson Syndrome- SBBYSS), sorted by hierarchical clustering. Cases marked in red atop the figure represent KAT6A syndrome cases, while purple represent GTPTS cases, yellow represent SBBYSS cases, blue indicate cases with no phenotypic or genotypic presentation of *KAT6A*, including samples with confirmed presentation of other syndromes.

CLUSTERING AND DIMENSION REDUCTION

Following each analysis, probes were examined with hierarchical clustering and multiple dimensional scaling (MDS) to examine the structure of the identified episignature. Hierarchical clustering was performed using Ward's method on Euclidean distance by the base stats package in R, and visualized with the ggplot2 package. MDS was performed by scaling of the pairwise Euclidean distances between samples.

DISCOVERY/TRAINING COHORT SELECTION

Identification of disease specific episignatures was performed using a randomly selected subset of the database, on a 75:25 ratio of discovery:training, using the caTools package in R. Testing samples were used to assess the performance of the classification model developed later in the study. For every disease group in the discovery cohort, a sex and age matched control group with a sample size at least 4 times larger was selected from the reference control group using the MatchIT package, and methylation profiles were compared between the two.

CROSS VALIDATION

For each round of validation, one of the 17 selected *KAT6A* samples was removed from probe selection, alongside matched controls and added *KAT6B* samples. The remaining *KAT6A* samples were designated as testing samples, and all three groups were modeled using MDS to determine how they clustered/segregated with one another. This process was repeated 10 times with different combinations of assigned training and testing samples (See Supplemental Figure 4-1).

CLASSIFICATION MODEL

The Methylation Variant Pathogenicity (MVP) score was created to assess the specificity of the identified methylation signature using all of the identified probes. A support vector machine (SVM) classifier used a linear kernel for training on *KAT6A* cases and controls. A 4:1 ratio of controls to cases, with cases and controls that had been used previously for probe selection, paired with 75% of the remaining controls, and 75% of the other syndrome samples from our EKD was chosen for modeling the episignature classification. Once modeled, we tested

the classifier using the remaining 25% of controls and other syndrome samples from the EKD. EKD samples include both 450k and EPIC array data, allowing the classifier to assess both array types, however because the majority of the samples to be tested later were assayed using the EPIC array, we limited the analysis to probes shared by both array types. Training used the e1071 R package, we used 10-fold cross validation to determine hyperparameters optimal to classification. In this process, the training set was divided into ten folds by random assignment, where the first nine are used for training, and the last used for testing the accuracy of the model. The mean accuracy was then calculated, and hyperparameters with the best performance by this metric were selected. The model provides a score ranging from 0-1 for each subject, representing the model's confidence in predicting whether the subject has a DNA methylation profile similar to the *KAT6A* probe set. Conversion of these SVM decision values was done using Platt's scaling method [39], and the class obtaining the greatest score determined the predicted phenotype. Classification as *KAT6A* was made when a sample received the greatest score for that class (normally greater than 0.5). Finally, the model was applied to both a training set of a large cohort of individuals with clinical and molecular diagnoses of neurodevelopmental disorders, as well as a group of healthy controls to determine the model's effective specificity.

VALIDATION OF CLASSIFICATION

To ensure the model is not susceptible to batch structure of the methylation experiment, the classifier was applied to samples assayed on the same batch as the cases used for training. Using downloaded methylation data from isolated cell populations of healthy individuals from the GEO online database to ensure the classifier is not sensitive to the blood cell type compositions, we provided these samples to the classifier and examined the variance of scores across different blood cell types. The model was then applied to the case cohort to evaluate its predictive ability on affected subjects. Specificity was determined by supplying a large number of DNA methylation arrays from healthy subjects to the model. To understand whether or not the model is sensitive to detecting other medical conditions presenting with similar phenotypes of neurodevelopmental disorder and intellectual disability, we tested a large number of subjects with confirmed clinical and molecular diagnoses of similar syndromes with the *KAT6A* classifier model.

DIFFERENTIALLY METHYLATED GENES

   To identify regions of significant methylation changes and the genes associated with them, the DMRcate algorithm was used [40]. P values were calculated for each probe using multivariable limma regression modeling, which were then kernel smoothed to identify regions with a minimum of 3 probes no more than 1 kb apart and an average regional methylation difference of greater than 10%. The Stouffer false-discovery rates (FDR) were used to select regions across the identified differentially methylated regions (DMRs). Analysis was performed on the same set of cases and controls used for methylation profiling and was also adjusted for blood cell type composition.

RESULTS

COMPARISON OF PHENOTYPES

   Common features of *KAT6A* syndrome include developmental (especially speech) delay and intellectual disability, varying in severity, and specific dysmorphic features, like microcephaly, eye abnormalities (ptosis, blepharophimosis and strabismus), low-set ears, a broad nasal tip, a thin upper lip and small peg shaped teeth (see Table 4-1). The majority of patients have a visual defect, heart anomaly and other congenital anomalies (see Table 4-2). Overlapping phenotypic characteristics between *KAT6A* and *KAT6B*-related syndromes include developmental delay and intellectual disability, microcephaly, hypotonia, and feeding problems [31,32,33]. Typical features of GTPTS are, as the name implies, genital and patellar abnormalities, but the vast majority of patients also have corpus callosum abnormalities (agenesis or hypoplasia), heart and renal malformations and limb contractures. Other (congenital) abnormalities have also been described in the literature (see Table 4-3). Facial features partially overlap with *KAT6A* syndrome and include low set and/or posteriorly rotated ears, a flat and/or broad nasal bridge with a broad and bulbous nasal tip and micro- or retrognathia. Mask-like facies and eye abnormalities (ptosis, hypertelorism, strabismus) are described in some patients with GTPTS (see Table 4-1). A characteristic feature of SBBYSS is a long thumb or great toe, but other hand and skeletal abnormalities have also been described. Teeth, thyroid, heart and genital defects are common, as well as hearing loss and lacrimal duct abnormalities. Regarding facial features of patients with SBBYSS, ptosis or blepharophimosis with mask-like facies and a cleft lip and/or palate are frequently seen (see Table 4-1 and 4-2).

Facial dysmorphism:	KAT6A (n=81)	KAT6B – GTPTS (n=26)	KAT6B – SBBYSS (n=73)
Microcephaly	36% (27/75)	65% (17/26)	30% (22/73)
Ptosis / Blepharophimosis	17% (13/78)	8% (2/26)	45% (33/73)
Downslanting palpebral fissures	1% (1/76)	-	8% (6/73)
Hypertelorism	5% (4/76)	12% (3/26)	11% (8/73)
Strabismus	55% (42/76)	8% (2/26)	5% (4/73)
Low-set and/or posteriorly rotated ears	32% (26/81)	19% (5/26)	12% (9/73)
Nasal bridge: flat and/or broad	1% (1/76)	12% (3/26)	16% (12/73)
Nasal bridge: prominent	-	8% (2/26)	3% (2/73)
Nasal tip: Broad, fleshy, bulbous.	85% (61/72)	31% (8/26)	23% (17/73)
Philtrum Short	'Common'	-	1% (1/73)
Long	-	4% (1/26)	10% (7/73)
Lip Thin	62% (42/68)	4% (1/26)	7% (5/73)
Micro/retrognathia	9% (7/81)	23% (6/26)	12% (9/73)
Mask-like facies	-	8% (2/26)	42% (31/73)

Table 1. Overview of the dysmorphic features of patients with *KAT6A* and *KAT6B*-related syndromes as described in literature. GTPTS = genitopatellar syndrome. SBBYSS = Say-Barber-Biesecker-Young-Simpson syndrome. '-' means that the feature has not occurred or has not been described in the specific group.

Anomaly of:	KAT6A (n=81)	KAT6B – GTPTS (n=26)	KAT6B – SBBYSS (n=73)
Corpus callosum	4% (3/76)	96% (25/26)	8% (6/73)
Brain (other)	19% (15/81)	31% (8/26)	18% (13/73)
Visual	65% (42/65)	15% (4/26)	8% (6/73)
Hearing	9% (7/76)	8% (2/26)	18% (13/73)
Lacrimal duct	5% (4/81)	-	16% (12/73)
Cleft lip/palate	3% (2/76)	23% (6/26)	26% (19/73)
Teeth	22%* (17/76)	12% (3/26)	40% (29/73)
Thyroid	3% (2/76)	19% (5/26)	33% (24/73)
Heart	53% (42/79)	65% (17/26)	38% (28/73)
Renal	4% (3/81)	81% (21/26)	4% (3/73)
Genital	10% (8/81)	88% (23/26)	40% (29/73)
Anus	1% (1/76)	27% (7/26)	1% (1/73)
Musculoskeletal			
Craniosynostosis	10% (8/81)	-	-
Thorax	2% (2/81)	19% (5/26)	3% (2/73)
Spine	3% (2/76)	8% (2/26)	1% (1/73)
Respiratory tract	10% (8/81)	31% (8/26)	8% (6/73)
Contractures	3% (2/76)	88% (23/26)	26% (19/73)
Patella	-	69% (18/26)	15% (11/73)
Long dig I (hand)	-	19% (5/26)	62% (45/73)
Hand (other)	10% (8/81)	15% (4/26)	16% (12/73)
Skeletal (other)	-	31% (8/26)	11% (8/73)

Table 2. Overview of the congenital and structural anomalies of patients with *KAT6A* and *KAT6B*-related syndromes as described in literature. GTPTS = genitopatellar syndrome. SBBYSS = Say-Barber-Biesecker-Young-Simpson syndrome. '-' means that the feature has not occurred or has not been described in the specific group.
* mainly small, peg shaped teeth, dental crowding.

Table 4-1,   4-2. Overview o f phenotypes for patients with KAT6A and KAT6B Syndromes

id	target	sex	age (years)	genotype
MS2563	KAT6A	f	6	KAT6A p.Met547Glufs*3
MS2564	KAT6A	m	3	KAT6A p.Gln1873*
MS2565	KAT6A	f	16	KAT6A p.Arg971Profs*5
MS2599	KAT6A	f	4	KAT6A c.3449dup p.Trp1152Metfs*23
MS2600	KAT6A	m	5	KAT6A c.3860_3861del p.Glu1287fs
MS2780	KAT6A	m	2	KAT6A: c.4645G>A p.Gly1549Ser
MS2782	KAT6A	m	1	KAT6A p.Ser1551Arg
MS3394	KAT6A	f	2	KAT6A: c.1961A>G p.Gln654Arg
MS3396	KAT6A	f	*21	KAT6A: c.3385C>T, p.Arg1129*
MS3397	KAT6A	f	*25	KAT6A: c.3820G>T, p.Glu1274*
MS3398	KAT6A	m	*2	KAT6A: c.3399_3400dup; p.Lys1134Argfs*14
MS3399	KAT6A	f	*4	KAT6A: c.3377delC; p.Ser1126Phefs*8
MS3400	KAT6A	m	*19	KAT6A: c.3631_3632delGT; p.Val1211*
MS3401	KAT6A	f	*2	KAT6A: c.4254_4257del; p.Glu1419Trpfs*12
MS3402	KAT6A	m	*13	KAT6A: c.3182T>A; p.Leu1061*
MS3403	KAT6A	m	*5	KAT6A: c.4224dup; p.Leu1409Ilefs*10
MS3404	KAT6A	m	*1	KAT6A: c.4502dup, p.Asn1501Lysfs*6
MS3405	KAT6A	f	*15	KAT6A: c.3070C>T, p.Arg1024*
MS3406	KAT6A	f	*1	KAT6A: c.3434del, p.Pro1145Leufs*2
MS3407	KAT6A	f	*36	KAT6A: c.3034C>T; p.Arg1012*
MS3422	KAT6A	m	9	KAT6A: c.3640A>T,p.Lys1214*
MS0673	GTPTS	f	1 month	KAT6B: c.3578_3585delTCCAGCAT; p.Phe1193Serfs*23
MS0677	GTPTS	m	3	KAT6B: c.3769_3772delTCTA; p.Lys1258Glyfs*13
MS1773	GTPTS	*m	*1	KAT6B: c.3788_3789del; p.Lys1263Argfs*7
MS1784	GTPTS	*f	*5	KAT6B: c.3788_3789del; p.Lys1263Argfs*7
MS0682	SBBYSS	f	6	KAT6B: c.3046del; p.Ser1016Alafs*98
MS1487	SBBYSS	*f	2	KAT6B: c.3147G>A; p.Ala1008Argfs*64
MS1641	SBBYSS	f	26	KAT6B: c.3349_3350del; p.Gln1117Valfs*19
MS1772	SBBYSS	*m	*8	KAT6B: c.3172C>T;p.Arg1058*
MS1785	SBBYSS	*m	*4	KAT6B: c.5492C>G;Ser1831*
MS1835	SBBYSS	*f	*17	KAT6B: c.5502C>G, p.Tyr1834*
MS1937	SBBYSS	f	1 month	KAT6B c.4617_4618del; p.Glu1540Aspfs*30
MS2253	SBBYSS	f	*14	KAT6B: c.4831delG
MS2557	SBBYSS	m	4	KAT6B: c.5795_5798del; p.Leu1932Hisfs*18
MS2788	SBBYSS	f	1	KAT6B: p.Arg9452*

Table 4-3. Overview of the *KAT6A* (n = 21) and *KAT6B* (SBBYSS n = 10, GTPTS n = 4) samples used in our analyses, including information about included patients gender, age and *KAT6A/B* mutation. Samples highlighted in yellow were removed from probe selection following preliminary analysis which indicated their methylation profiles were more similar to matched control samples than other cases. Samples marked with an asterisk (*) did not have age or sex information provided, and thus had their status calculated via methylation profile.

DETECTION AND VERIFICATION OF AN EPISIGNATURE FOR KAT6A

DNA methylation profiles from the peripheral blood of 17 individuals with confirmed clinical and molecular presentations of *KAT6A* syndrome were used to derive the episignature. Sample filtering steps removed 4 of the original 21 samples for several reasons, including sample quality, sample type, and sample clustering patterns (samples clustering with controls based on preliminary assessment are excluded). Samples had fewer than 1000 failed probes and passed quality control requirements.

First analyses where *KAT6A* samples were compared to matched controls indicated the presence of shared methylation patterns between *KAT6A* and highly related syndromic groups based on 97 differentially methylated probes (See Figure 4-1A), namely GTPTS and SBBYSS, neurodevelopmental disorders associated with mutations in the lysine acetyltransferase protein 6B (*KAT6B*) gene. Although the model was capable of differentiating KAT6A samples from controls (See Figure 4-1B), MDS and heatmap models showed significant overlap in clustering patterns between these three syndromes and both the *KAT6B* related disorders scored moderately high scores in the classifier (See Figure 4-1C, 4-1D, 4-1E), representing the significant degree of overlap between these conditions. Full delineation of these highly similar KAT6 family syndromes was achieved when the 14 GTPTS and SBBYSS samples in our database were provided as controls for probe selection, training the selected probeset against the features of this cohort alongside age, sex and array matched controls. This allowed for distinct MDS and heatmap separation of case and control samples, with corresponding high MVP scores for the *KAT6A* cohort and low scores for the remaining syndrome's samples.

When compared to this control set, cases showed significant differences in methylation patterns of 114 probes, which are visualized using a volcano plot (See Figure 4-2A). Probes with a minimum methylation difference of 5% between the two cohorts, and a multiple testing corrected p value of <0.01 (limma multivariable regression modeling) were used for the episignature. P values were adjusted for blood cell type composition to ensure comparability between heterogeneous peripheral blood sample sources. Hierarchical clustering and MDS demonstrate that the selected probeset strongly separates cases and controls (See Figures 4-2B, 4-2D, 4-2E). Cross-validation using *KAT6A* samples was performed to validate the sensitivity of the episignature, showing in each case that the remaining testing samples clustered with the other *KAT6A* samples, and segregated from the controls. Plotting *KAT6B* samples associated with GTPTS and SBBYSS conditions showed distinct clustering of *KAT6A* samples from the *KAT6B* samples, where previous models showed adjacent clustering. Some minor overlap of the SBBYSS cohort in the MDS model was observed, however hierarchical clustering heatmaps showed the capability of the model in differentiating the two syndromes (See Figures 4-2D, 4-2E).

DEVELOPMENT OF AN MVP SCORE

Subjects with *KAT6A* variants, matched controls, and a number of samples from syndrome cohorts previously established in the EKD were used for training a model to test the sensitivity and specificity of the episignature created through previous probe selection steps. The MVP score was set to generate a single score from 0-1 for each sample, with 1 being a methylation pattern highly similar to the case samples, and 0 being a methylation pattern highly similar to matched control samples. The class obtaining the highest score determined the episignature classification. These results were validated by a series of tests to validate their reliability. While all *KAT6A* samples received high scores close to 1, control scores remained near 0, indicating the classifier has a high sensitivity for the detection of the *KAT6A* episignature (See Figure 4-2C, Figure 4-3). Furthermore, specificity of the classifier was tested by providing it with a large number of subjects with confirmed diagnosis of an NDD of various types, including trinucleotide repeat expansion abnormalities, imprinting defect disorders, BAFopathies, Mendelian disorders of the epigenetic machinery, down syndrome as well as subjects with nonsyndromic autism spectrum disorders. The vast majority of EKD samples,

including the *KAT6B* syndromes, GTPTS and SBBYSS, were classified as being highly similar to controls using the *KAT6A* classifier, confirming its efficacy as a sensitive and specific model for the identification of *KAT6A* related episignatures.
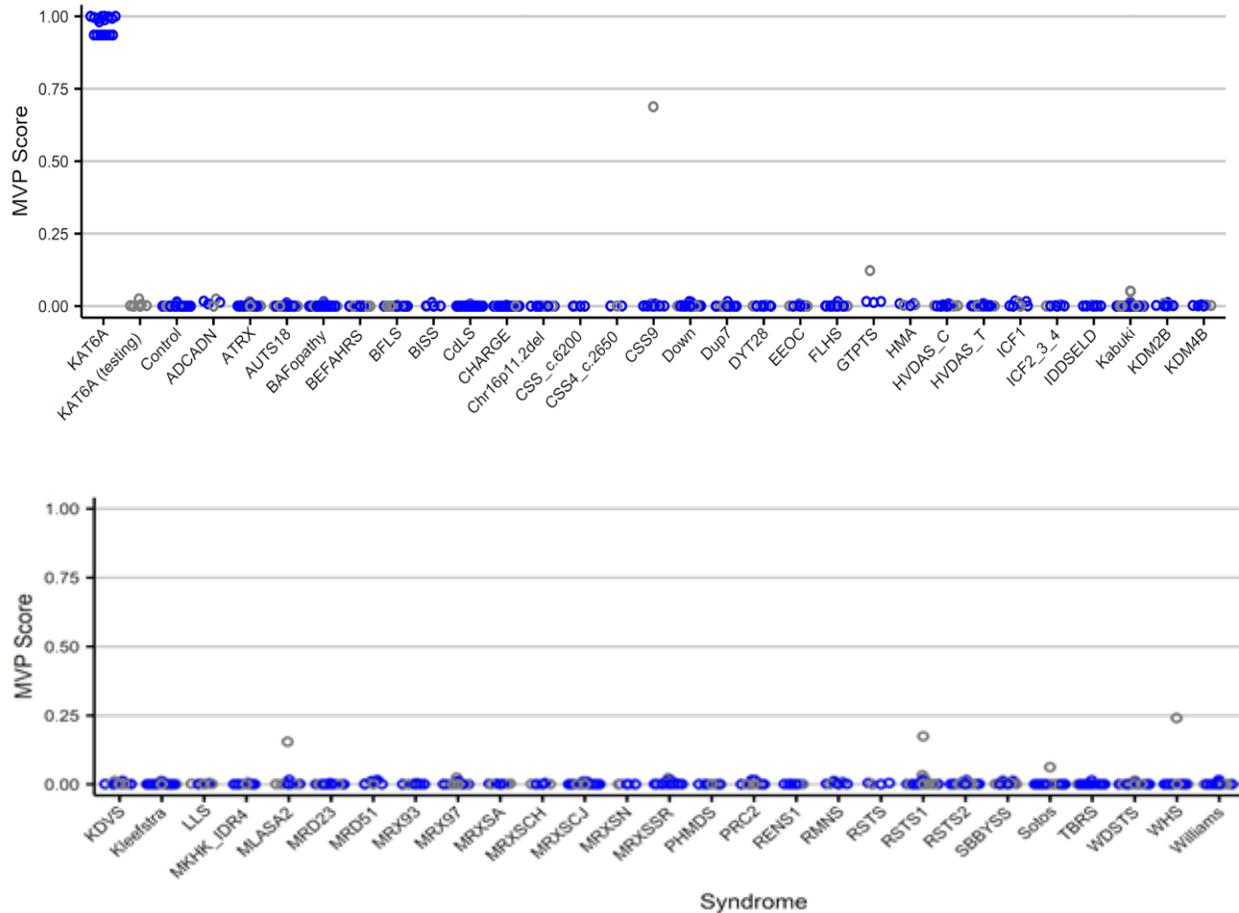
Figure 4-3. SVM classifier model for *KAT6A*. Each sample receives scores for the probability of having a DNA methylation profile similar to cases as compared to every other sample with a confirmed episignature in the EKD. Higher value on Y axis indicates that a sample presents a methylation profile more similar to cases compared to the methylation profiles of patients with other disorders. 53 other syndromes with confirmed episignatures from the EKD are plotted based on this relative scale of similarity to indicate probeset specificity for the case disorder. This classifier also clearly distinguishes KAT6A syndrome from the *KAT6B* syndromes GTPTS and SBBYSS.

DIFFERENTIALLY METHYLATED GENES

In the significantly differentially methylated regions (DMRs), we found a total of 36 differentially methylated genes (DMGs) in the assessed *KAT6A* syndrome, GTPTS and SBBYSS samples, compared to control samples. For *KAT6A* syndrome, this included 15 genes, whilst GTPTS samples had 25 DMGs and SBBYSS 8 DMGs.

The following genes were differentially methylated in *KAT6A* syndrome and GTPTS samples, compared to controls: *BMP4, HEY2, HOTAIRM1, HOXA1, HOXA3, HOXA5, HOXA6, HOXA-AS3, RP11-357H14.17, RP1-170O19.22, RP1-170O19.23*. The *GLI2* gene was only differentially methylated in KAT6A syndrome and SBBYSS samples compared to controls. For a full list of all 36 DMGs in the DMRs of *KAT6A* syndrome, GTPTS and SBBYSS samples (all compared to controls), see supplementary Table 4-1.

DISCUSSION

Recent advances in the assessment of DNA methylation profiles have provided significant improvements to the diagnosis and discovery of rare diseases. To date, 65 neurodevelopmental conditions have been associated with specific DNA methylation profiles, known as episignatures [20,41]. Often, genes with roles involving histone modifications, DNA methylation and chromatin have been implicated, and the resulting downstream changes to the epigenetic profile of these patients can be used to differentiate between conditions exhibited by different patients [20,42,43,44]. These episignatures are complex representations of the corresponding gene expression profiles affected by these various epigenetic modifiers, and in several cases, have provided clarification for variants of unknown significance [43,44], and refinement of the spectrum of variants associated with a given condition [20,41]. This study focuses on a cohort of patients with variants within the *KAT6A* gene sequence, and further expands upon the landscape of epigenetic classifiers, providing evidence of this technology's capability of differentiating syndromes with a high degree of phenotypic and genotypic overlap. These findings provide novel avenues of research and expansion of diagnostic criteria for patients with *KAT6A* variants, as well as it's adjacent sister conditions, GTPTS and SBBYSS.

KAT6A syndrome is a relatively novel condition, first identified in 2015 by Arboleda et al. [26], and shortly after by Tham et al. [27], and presents as an intellectual disability disorder with additional features including facial abnormalities, skeletal abnormalities, speech pathologies

and other (congenital) abnormalities. Based on DNA methylation data derived from the peripheral blood of patients with clinically and molecularly identified *KAT6A* syndrome, we have identified an episignature specific to the syndrome's associated pathology. Our data suggests that not only is there a specific DNA methylation signature for patients with variants in the *KAT6A* sequence, but that significant overlap in methylation differences can occur between similar conditions GTPTS and SBBYSS, potentially elucidating similar mechanisms of action.

DNA methylation signatures vary in genomic locations and intensities across different disorders, with the identified *KAT6A* syndrome signature showing highly robust characteristics, including distinct hierarchical clustering, segregation of case and controls following MDS models, similar to previously described episignatures [3,10,11,12,13, 14, 15,16,17,18,19,20]. Of the 21 samples with (likely pathogenic) variants in *KAT6A*, 17 samples, all with truncating variants, showed a specific signature. Four *KAT6A* syndrome samples grouped predominantly with control samples. Of these four samples, three had missense variants and one had a truncating variant, located more upstream in the *KAT6A* gene than the other samples. This may suggest that, similar to *KAT6B*, multiple episignatures exist for *KAT6A* syndrome, possibly accompanied by a different phenotype, and that the four "negative" samples might have another signature. However, the group number is too small to draw conclusions yet. Similar results exhibiting multiple episignatures from variants within single genes have been described, including *ADNP*, *SMARCA2*, *SRCAP*, and *KMT2D* [21, 45,46, 47]. Locations of variants plays a major role in determining the subsequent methylation patterns seen when patients are assessed, as such it is not unreasonable to suggest a similar phenomenon may be occurring within these more "upstream" *KAT6A* variants. Variants outside previously established "canonical" loci associated with *KMT2D* related Kabuki syndrome, and *SRCAP* related Floating Harbor syndrome were found to exhibit methylation patterns distinct from those found in patients with canonical variants in *SRCAP* exons 33 and 34 for Floating Harbor syndrome, and in *KMT2D* exons outside 38 and 39 for Kabuki syndrome 1. Furthermore, in the assessment of Helsmortel Van der Aa syndrome patient cohorts, patients classified with variants within the central domain of the *ADNP* gene provide a different episignature than those exhibited by patients with variants within the terminal domain [21]. Additionally, location, especially in terms of affected domains associated with the syndrome may play a key role in determining methylation patterns as well, as demonstrated in the assessment of non-Nicolaides Baraitser patients with *SMARCA2* mutations.

Assessment of these novel BIS patients identified large changes in the methylation pattern for patients with variants outside of the helicase domain, that result in a recognizably distinct syndrome from NCBRS, despite being associated with the same *SMARCA2* gene [45]. Further assessment of the variant loci, and the disruption of genetic and epigenetic machinery that comes as a result of these changes in different loci may explain the distinct methylation patterns exhibited by these non-compliant samples.

Evidence of highly sensitive and specific association of the selected probeset with *KAT6A* syndrome was provided following cross validation and SVM classifier training. All case samples received high MVP scores, while controls and individuals from other disorders provided to the model scored very low scores. This indicates that the highly differentiated CpG probes selected through training with the *KAT6A* cohort are capable of reliably classifying this condition in a sensitive and specific manner. The first models were produced without the addition of KAT6B samples to the control training cohort, to test the ability of the *KAT6A* derived episignature to differentiate GTPTS patients using the DMRs derived from patients with *KAT6A* variants. Providing four GTPTS samples to the episignature derived from probe selection using *KAT6A* samples showed GTPTS samples clustered away from the *KAT6A* sample cluster, as well as being visually separable from the remaining control samples. The selected methylated regions seem to contain informative loci for both disorders, providing further evidence of the genetic overlap of these disorders. Additionally, models created using the comparison of *KAT6A* samples and SBBYSS samples, showed similar results, with SBBYSS samples intermixed with the *KAT6A* samples. Interestingly, when GTPTS and SBBYSS were assessed separately in a previous publication [20], it was discovered that despite being caused by mutations within the same gene, the methylation profiles exhibited by these two syndrome groups were quite distinct from one another, and resulted in episignatures for both conditions. When the probeset was trained using *KAT6A* samples only, without the addition of GTPTS and SBBYSS samples to the training control group, intermediate scores were produced in the MVP assessment for both GTPTS and SBBYSS. This indicates a significant overlap in methylation profiles between these similar disorders, a phenomenon which has been observed in assessments of other NDDs, including the combined BAFopathy episignature derived from the assessment of Nicolaides-Baraitser (NCBRS) and Coffin-Siris (CSS) cohorts [34].

Further research is required to determine why these two distinct syndromes are showing methylation profile overlap with *KAT6A* syndrome, a disorder associated with genetic variants in *KAT6A*. As stated before, the two associated genes, *KAT6A* and *KAT6B* are both components of the similar MOZ/MYST complexes, that exert lysine acetyltransferase activity [33]. In an attempt to search for answers we compared the clinical features of these three syndromes (Tables 4-1 and 4-2) and tried to find an explanation for the overlap and differences of features of these syndromes in the methylation profiles. We saw that the *BMP4* gene (OMIM # 112262) is differentially methylated in *KAT6A* syndrome and GTPTS [48]. Mutations in this gene can cause a disorder characterized by cleft lip/palate, brain, eye and skeletal malformations. The *BMP4* gene is also involved in tooth development. All of these features have been described in *KAT6A* syndrome and GTPTS. The *HEY2* gene (OMIM # 604674) is also differentially methylated in both *KAT6A* syndrome and GTPTS samples. This gene is believed to be important for heart development and its differential methylation could perhaps explain the high frequency of heart anomalies in *KAT6A* syndrome (53%) and GTPTS (65%) patients. [49]

Multiple HOX-genes are differentially methylated in KAT6A syndrome and GTPTS samples (*HOTAIRM1, HOXA1, HOXA3, HOXA5, HOXA6* and *HOXA-AS3*) or only in GTPTS samples (*HOXA4, HOXA-AS2, HOXB3, HOXB6, HOXB-AS3* and *HOXB-AS4*). HOX genes are a group of Homeobox genes, that are important for various developmental processes. A regulatory effect of *KAT6A* on HOX gene expression has been described previously [50, 51]. Mutations in some of the HOX genes are associated with specific malformations or syndromic disorders with a wide spectrum of symptoms (hearing loss, developmental delay, cardiovascular and skeletal malformations, dysmorphic features), but for other HOX genes the function and consequences of mutations need to be further elucidated. Samples of *KAT6A* syndrome and SBBYSS patients had a differentially methylated *GLI2* gene (OMIM # 165230) in common. This gene is associated with Holoprosencephaly 9 (OMIM # 610829) and Culler-Jones syndrome (OMIM # 615849) and could explain some of *KAT6A* syndrome and SBBYSS features, like eye/visual abnormalities and microcephaly. It would be interesting to perform deep phenotyping of the patients of which samples are used in our current study and to specifically compare the differentially methylated genes per patient to their clinical features.

An important consideration to make is that the identification of DNA methylation episignatures is achieved using peripheral blood samples from patients, rather than the tissues most relevant to the particular syndrome, in this case neural tissue. Therefore it can be proposed that the DNA methylation patterns seen in the peripheral blood do not necessarily match those within the neural tissues of the same patients. Nonetheless, it has been shown that epigenetic alterations in brain and peripheral tissues are correlated to changes in brain tissue [52,53], which suggests that methylation patterns in peripheral blood appear to reflect methylation patterns in other tissues for at least some genes. Although we have determined that the selected probes used in the *KAT6A* signature are effective at providing evidence towards differentiation of *KAT6A* samples from other cases, not all of them are necessarily related to the phenotype of the condition, but still act as robust biomarkers for the condition. Further work on the overlapping differentially methylated probes for the *KAT6A* and *KAT6B* related episignatures will provide insights into how these differentially methylated probes relate to the pathophysiology of this condition. Additionally, the Illumina EPIC bead chip array used for this assay covers nearly 860,000 human genomic methylation CpG sites, including 99% of published Refseq genes, and 96% of CpG islands [3,12]. Therefore, although the array does not represent the totality of human methylation profiles, it assesses the vast majority of biologically relevant gene sequences and other elements with significant effects on heredity. Further advances in microarray and epigenetic assessment technologies may unveil more active methylation sites for assessment, however with the present capabilities of the platform, reliable and comprehensive assessments of human epigenetic change can be provided.

In conclusion, the discovery of a highly robust *KAT6A* syndrome episignature expands a list of NDDs with DNA methylation episignatures that can be used for screening and diagnosis of patients with rare neurodevelopmental conditions. The highly sensitive and specific DNA methylation profile detected from peripheral blood of patients with *KAT6A* variants enables effective diagnosis, screening and classification of suspected *KAT6A* variants and provides a novel avenue of testing in diagnostic settings. Assessment of the epigenetic overlap between *KAT6A* syndrome, GTPTS and SBBYSS has provided insight into common pathways, that potentially result in the shared characteristics of these conditions. Additional work expanding on

the number of cases with different types of *KAT6A* mutations, across multiple exons, will allow for more extensive refinement of the episignature, and a further assessment of the diagnostic overlap of this model with adjacent syndromes will provide further avenues of research and greater understanding of these complex conditions.

REFERENCES

1. Wakap, S.N.; Lambert, D.M.; Olry, A.; Rodwell, C.; Gueydan, C.; Lanneau, V.; Murphy, D.; Le Cam, Y.;Rath, A. Estimating cumulative point prevalence of rare diseases: Analysis of the Orphanet database.Eur. J.Hum. Genet.2020,28, 165–173.

2. Zablotsky B, Black LI, Maenner MJ, et al. Prevalence and Trends of Developmental Disabilities among Children in the United States: 2009–2017. Pediatrics. 2019;144(4):e20190811

3. Aref-Eshghi E, Rodenhiser DI, Schenkel LC, et al. Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. The American Journal of Human Genetics. 2018;102(1):156-174. doi:10.1016/j.ajhg.2017.12.008

4. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-424. doi:10.1038/gim.2015.30

5. Sadikovic, B.; Aref-Eshghi, E.; Levy, M.A.; Rodenhiser, D. DNA methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. Epigenomics 2019,11,563–575.

6. Skvortsova K, Iovino N, Bogdanović O. Functions and mechanisms of epigenetic inheritance in animals. Nat Rev Mol Cell Biol. 2018 Dec;19(12):774-790. doi: 10.1038/s41580-018-0074-2. PMID: 30425324.

7. Costa FF. Non-coding RNAs, epigenetics and complexity. Gene. 2008 Feb 29;410(1):9-17. doi: 10.1016/j.gene.2007.12.008. Epub 2008 Jan 15. PMID: 18226475.

8. Bjornsson HT. The Mendelian disorders of the epigenetic machinery. Genome Res. 2015;25(10):1473-1481. doi:10.1101/gr.190629.115

9. Barros-Silva, D.; Marques, C.J.; Henrique, R.; Jerónimo, C. Profiling DNA methylation based on next-generation sequencing approaches: New insights and clinical applications. Genes 2018,9, 429.

10. Choufani, S.; Cytrynbaum, C.; Chung, B.H.Y.; Turinsky, A.L.; Grafodatskaya, D.; Chen, Y.A.; Cohen, A.S.A.;Dupuis, L.; Butcher, D.T.; Siu, M.T.; et al. NSD1 mutations generate a genome-wide DNA methylation signature. Nat. Commun.2015,6.

11. Schenkel, L.C.; Kernohan, K.D.; McBride, A.; Reina, D.; Hodge, A.; Ainsworth, P.J.; Rodenhiser, D.I.;Pare, G.; Bérubé, N.G.; Skinner, C.; et al. Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome.Epigenet. Chromatin 2017,10, 1–11.

12. Aref-Eshghi, E.; Bend, E.G.; Colaiacovo, S.; Caudle, M.; Chakrabarti, R.; Napier, M.; Brick, L.; Brady, L.;Carere, D.A.; Levy, M.A.; et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in GeneticallyUnsolved Individuals with Suspected Hereditary Conditions.Am. J. Hum. Genet.2019,104, 685–700.

13. Butcher, D.T.; Cytrynbaum, C.; Turinsky, A.L.; Siu, M.T.; Inbar-Feigenberg, M.; Mendoza-Londono, R.;Chitayat, D.; Walker, S.; Machado, J.; Caluseriu, O.; et al. CHARGE and Kabuki Syndromes: Gene-SpecificDNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically OverlappingConditions.Am. J. Hum. Genet.2017,100, 773–788

14. Hood, R.L.; Schenkel, L.C.; Nikkel, S.M.; Ainsworth, P.J.; Pare, G.; Boycott, K.M.; Bulman, D.E.; Sadikovic, B.The defining DNA methylation signature of Floating-Harbor Syndrome.Sci. Rep.2016,6, 1–9.

15. Schenkel, L.C.; Aref-Eshghi, E.; Skinner, C.; Ainsworth, P.; Lin, H.; Paré, G.; Rodenhiser, D.I.; Schwartz, C.;Sadikovic, B. Peripheral blood epi-signature of Claes-Jensen syndrome enables sensitive and specific identification of patients and healthy carriers with pathogenic mutations in KDM5C.Clin. Epigenet.2018,10,1–11.

16. Schenkel, L.C.; Schwartz, C.; Skinner, C.; Rodenhiser, D.I.; Ainsworth, P.J.; Pare, G.; Sadikovic, B. Clinical Validation of Fragile X Syndrome Screening by DNA Methylation Array.J. Mol. Diagn.2016,18, 834–841.

17. Li, Y.; Chen, J.A.; Sears, R.L.; Gao, F.; Klein, E.D.; Karydas, A.; Geschwind, M.D.; Rosen, H.J.; Boxer, A.L.;Guo, W.; et al. An Epigenetic Signature in Peripheral Blood Associated with the Haplotype on 17q21.31,a Risk Factor for Neurodegenerative Tauopathy.PLoS Genet.2014,10.

18. Aref-Eshghi, E.; Schenkel, L.C.; Lin, H.; Skinner, C.; Ainsworth, P.; Paré, G.; Siu, V.; Rodenhiser, D.;Schwartz, C.; Sadikovic, B. Clinical Validation of a Genome-Wide DNA Methylation Assay for MolecularDiagnosis of Imprinting Disorders.J. Mol. Diagn.2017,19, 848–856.

19. Guastafierro, T.; Bacalini, M.G.; Marcoccia, A.; Gentilini, D.; Pisoni, S.; Di Blasio, A.M.; Corsi, A.; Franceschi, C.;Raimondo, D.; Spanò, A.; et al. Genome-wide DNA methylation analysis in blood cells from patients with Werner syndrome.Clin. Epigenet.2017,9, 1–10.

20. Aref-Eshghi, E.; Kerkhof, J.; Pedro, V.P.; Barat-Houari, M.; Ruiz-Pallares, N.; Andrau, J.C.; Lacombe, D.;Van-Gils, J.; Fergelot, P.; Dubourg, C.; et al. Evaluation of DNA Methylation Episignatures for Diagnosis andPhenotype Correlations in 42 Mendelian Neurodevelopmental Disorders.Am. J. Hum. Genet.2020,106,356–370

21. Bend EG, Aref-Eshghi E, Everman DB, et al. Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. Clin Epigenetics. 2019;11(1):64. Published 2019 Apr 27. doi:10.1186/s13148-019-0658-5

22. Sadikovic, B., Levy, M.A., Kerkhof, J. *et al.* Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genet Med* **23,** 1065–1074 (2021). https://doi.org/10.1038/s41436-020-01096-4

23. Lee KK, Workman JL. Histone acetyltransferase complexes: one size doesn't fit all. Nat Rev Mol Cell Biol. 2007 Apr;8(4):284-95. doi: 10.1038/nrm2145. PMID: 17380162

24. Akatsuki Kimura, Kazuko Matsubara, Masami Horikoshi, A Decade of Histone Acetylation: Marking Eukaryotic Chromosomes with Specific Codes, The Journal of Biochemistry, Volume 138, Issue 6, December 2005, Pages 647–662, https://doi.org/10.1093/jb/mvi184

25. Klein BJ, Lalonde ME, Côté J, Yang XJ, Kutateladze TG. Crosstalk between epigenetic readers regulates the MOZ/MORF HAT complexes. Epigenetics. 2014 Feb;9(2):186-93. doi: 10.4161/epi.26792. Epub 2013 Oct 29. PMID: 24169304; PMCID: PMC3962528.

26. Arboleda VA, Lee H, Dorrani N, et al. De Novo Nonsense Mutations in KAT6A, a Lysine Acetyl-Transferase Gene, Cause a Syndrome Including Microcephaly and Global Developmental Delay. Am J Hum Genet. 2015;96(3):498-506. doi:10.1016/j.ajhg.2015.01.017

27. Tham E, Lindstrand A, Santani A, et al. Dominant Mutations in KAT6A Cause Intellectual Disability with Recognizable Syndromic Features. Am J Hum Genet. 2015;96(3):507-513. doi:10.1016/j.ajhg.2015.01.016

28. Kennedy J, Goudie D, Blair E, et al. KAT6A Syndrome: genotype-phenotype correlation in 76 patients with pathogenic KAT6A variants. Genet Med. 2019;21(4):850-860. doi:10.1038/s41436-018-0259-2

29. Su Y, Liu J, Yu B, Ba R, Zhao C. Brpf1 Haploinsufficiency Impairs Dendritic Arborization and Spine Formation, Leading to Cognitive Deficits. Front Cell Neurosci. 2019;13. doi:10.3389/fncel.2019.00249

30. Yan K, Rousseau J, Littlejohn RO, et al. Mutations in the Chromatin Regulator Gene BRPF1 Cause Syndromic Intellectual Disability and Deficient Histone Acetylation. Am J Hum Genet. 2017;100(1):91-104. doi:10.1016/j.ajhg.2016.11.011

31. Clayton-Smith J, O'Sullivan J, Daly S, et al. Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. Am J Hum Genet. 2011;89(5):675-681. doi:10.1016/j.ajhg.2011.10.008

32. Kraft M, Cirstea IC, Voss AK, et al. Disruption of the histone acetyltransferase MYST4 leads to a Noonan syndrome–like phenotype and hyperactivated MAPK signaling in humans and mice. J Clin Invest. 2011;121(9):3479-3491. doi:10.1172/JCI43428

33. Wiesel-Motiuk N, Assaraf YG. The key roles of the lysine acetyltransferases KAT6A and KAT6B in physiology and pathology. Drug Resist Updates. 2020 Dec;53:100729. doi: 10.1016/j.drup.2020.100729. Epub 2020 Oct 7. PMID: 33130515.

34. Aref-Eshghi, E.; Bend, E.G.; Hood, R.L.; Schenkel, L.C.; Carere, D.A.; Chakrabarti, R.; Nagamani, S.C.S.;Cheung, S.W.; Campeau, P.M.; Prasad, C.; et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin–Siris and Nicolaides–Baraitser syndromes.Nat. Commun.2018,9.

35. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049

36. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. doi:10.1093/nar/gkv007

37. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13(1):86. doi:10.1186/1471-2105-13-86

38. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One. 2012;7(7):e41361. doi:10.1371/journal.pone.0041361

39. Advances in Large-Margin Classifiers. Published online September 29, 2000. doi:10.7551/mitpress/1113.001.0001

40. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin. 2015;8(1):6.

41. Levy, Michael A et al. "Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders." *HGG advances* vol. 3,1 100075. 3 Dec. 2021, doi:10.1016/j.xhgg.2021.100075

42. Haghshenas S, Levy MA, Kerkhof J, Aref-Eshghi E, McConkey H, Balci T, Siu VM, Skinner CD, Stevenson RE, Sadikovic B, Schwartz C. Detection of a DNA Methylation Signature for the Intellectual Developmental Disorder, X-Linked, Syndromic, Armfield Type. Int J Mol Sci. 2021 Jan 23;22(3):1111. doi: 10.3390/ijms22031111. PMID: 33498634; PMCID: PMC7865843.

43. Aref-Eshghi E, Bourque DK, Kerkhof J, Carere DA, Ainsworth P, Sadikovic B, Armour CM, Lin H. Genome-wide DNA methylation and RNA analyses enable reclassification of two variants of uncertain significance in a patient with clinical Kabuki syndrome. Hum Mutat. 2019 Oct;40(10):1684-1689. doi: 10.1002/humu.23833. Epub 2019 Jul 3. PMID: 31268616.

44. Aref-Eshghi E, Schenkel LC, Lin H, Skinner C, Ainsworth P, Paré G, Rodenhiser D, Schwartz C, Sadikovic B. The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. Epigenetics. 2017;12(11):923-933. doi: 10.1080/15592294.2017.1381807. Epub 2017 Nov 7. PMID: 28933623; PMCID: PMC5788422.

45. Cappuccio, Gerarda et al. "De novo SMARCA2 variants clustered outside the helicase domain cause a new recognizable syndrome with intellectual disability and

blepharophimosis distinct from Nicolaides-Baraitser syndrome." *Genetics in medicine : official journal of the American College of Medical Genetics* vol. 22,11 (2020): 1838-1850. doi:10.1038/s41436-020-0898-y

46. Rots, Dmitrijs et al. "Truncating SRCAP variants outside the Floating-Harbor syndrome locus cause a distinct neurodevelopmental disorder with a specific DNA methylation signature." *American journal of human genetics* vol. 108,6 (2021): 1053-1068. doi:10.1016/j.ajhg.2021.04.008

47. Cuvertino, Sara et al. "A restricted spectrum of missense KMT2D variants cause a multiple malformations disorder distinct from Kabuki syndrome." *Genetics in medicine : official journal of the American College of Medical Genetics* vol. 22,5 (2020): 867-877. doi:10.1038/s41436-019-0743-3

48. Blackburn, Patrick R et al. "Variable expressivity of syndromic BMP4-related eye, brain, and digital anomalies: A review of the literature and description of three new cases." European journal of human genetics : EJHG vol. 27,9 (2019): 1379-1388. doi:10.1038/s41431-019-0423-4

49. Reamon-Buettner SM, Borlak J. HEY2 mutations in malformed hearts. Hum Mutat. 2006 Jan;27(1):118. doi: 10.1002/humu.9390. PMID: 16329098.

50. Voss AK, Collin C, Dixon MP, Thomas T. Moz and retinoic acid coordinately regulate H3K9 acetylation, Hox gene expression, and segment identity. Dev Cell. 2009 Nov;17(5):674-86. doi: 10.1016/j.devcel.2009.10.006. PMID: 19922872.

51. Shane C. Quinonez, Jeffrey W. Innis,Human HOX gene disorders,Molecular Genetics and Metabolism, Volume 111, Issue 1,2014, Pages 4-15, ISSN 1096-7192, https://doi.org/10.1016/j.ymgme.2013.10.012.

52. Klengel T, Pape J, Binder EB, Mehta D. The role of DNA methylation in stress-related psychiatric disorders. Neuropharmacology. 2014 May;80:115-32. doi: 10.1016/j.neuropharm.2014.01.013. Epub 2014 Jan 19. PMID: 24452011.

53. Lussier AA, Bodnar TS, Mingay M, Morin AM, Hirst M, Kobor MS, Weinberg J. Prenatal Alcohol Exposure: Profiling Developmental DNA Methylation Patterns in Central and Peripheral Tissues. Front Genet. 2018 Dec 4;9:610. doi: 10.3389/fgene.2018.00610. PMID: 30568673; PMCID: PMC6290329.
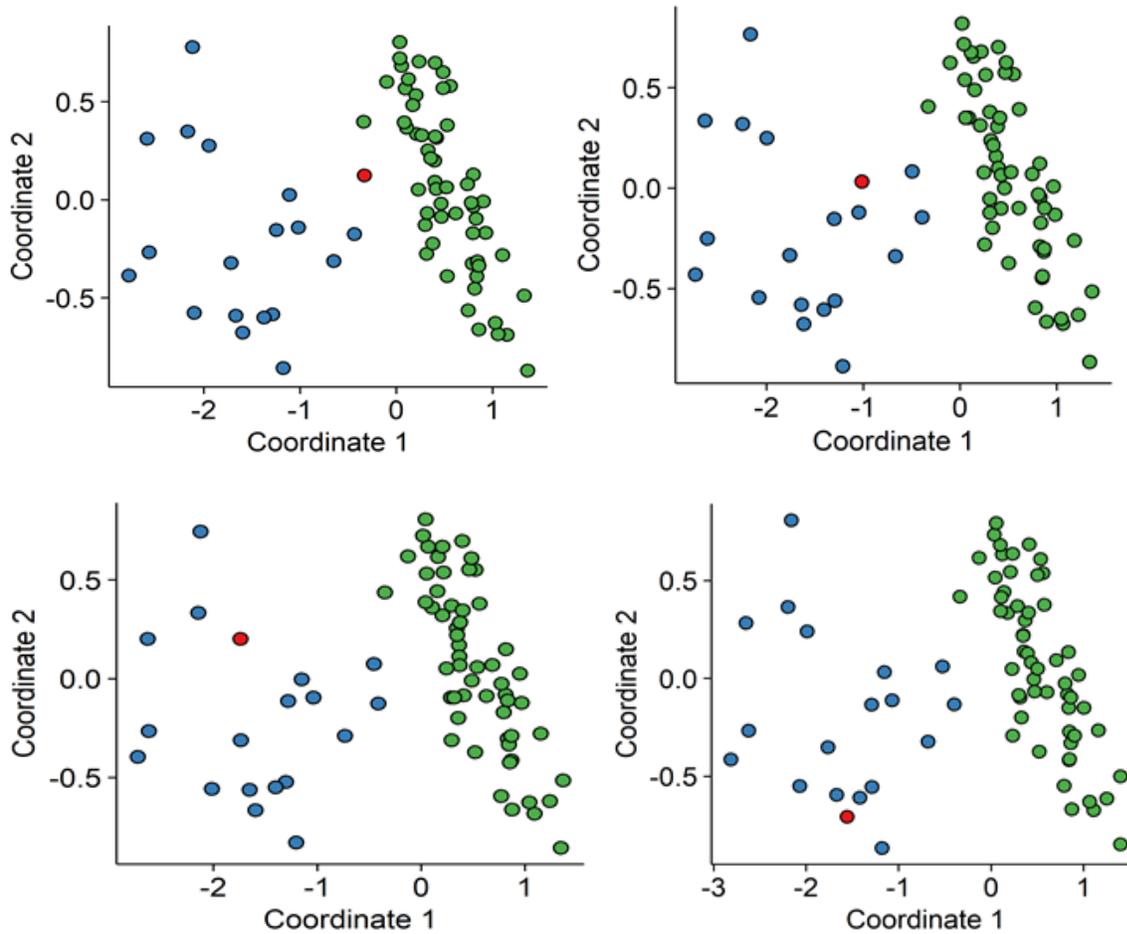
Supplemental Figure 4-1. Multidimensional scaling plot for cross validation, representing the dimensions of variation in methylation signal intensity at informative CpG identified for KAT6A. Represents comparisons of the similarity of methylation profiles of KAT6A patients, with those chosen for probe selection marked in blue, and the cross-validation sample marked in red. Control samples (shown in green) include cases without a confirmed phenotypic and/or genotypic presentation of KAT6A, including samples with confirmed presentation of other syndromes.

KAT6A MVP Score ROC Graph
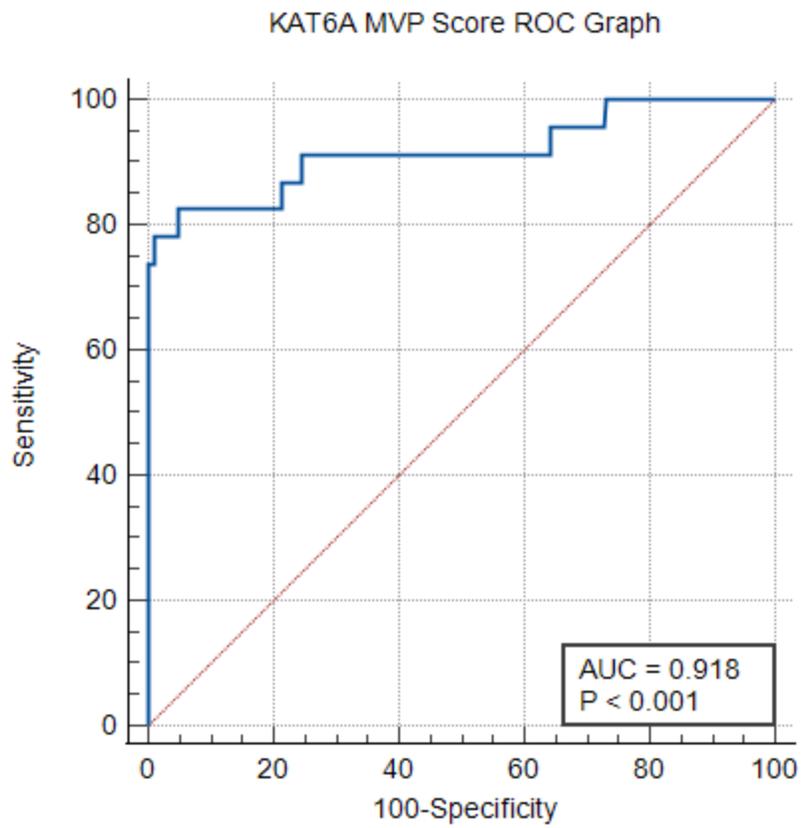
Supplemental Figure 4-2. **KAT6A MVP Score ROC Graph.** Receiver operating characteristic curve demonstrating the sensitivity and specificity of the generated MVP scores for the KAT6A cohort and the remaining EKD samples used for training.

Differentially methylated gene	Syndrome(s)
ANK2	GTPTS
BMP4	KAT6A+GTPTS
COL18A1	GTPTS
CTC-340A15.2	KAT6A
DTNA	SBBYSS
GLI2	KAT6A+SBBYSS
HEY2	KAT6A+GTPTS
HLA-DPA1	SBBYSS
HLA-DPB1	SBBYSS
HOTAIRM1	KAT6A+GTPTS
HOXA1	KAT6A+GTPTS
HOXA3	KAT6A+GTPTS
HOXA4	GTPTS
HOXA5	KAT6A+GTPTS
HOXA6	KAT6A+GTPTS
HOXA-AS2	GTPTS
HOXA-AS3	KAT6A+GTPTS
HOXB3	GTPTS
HOXB6	GTPTS
HOXB-AS3	GTPTS
HOXB-AS4	GTPTS
KIAA1161	SBBYSS
LTBP3	SBBYSS
METTL11B	GTPTS
NKAIN1	SBBYSS
PCCA	GTPTS
PDE4D	KAT6A
PPP2R2C	SBBYSS
RP11-297J22.1	KAT6A
RP11-357H14.17	KAT6A+GTPTS
RP1-155D22.2	GTPTS
RP11-650J17.1	GTPTS
RP1-170O19.22	KAT6A+GTPTS
RP1-170O19.23	KAT6A+GTPTS
TJP1	GTPTS
ZDHHC14	GTPTS

Supplemental Table 4-1. Differentially Methylated Genes for KAT6A Episignature Probeset.

Differentially methylated genes for each syndromic group. Genes were termed as differentially

methylated if a methylation signal intensity at 3 consecutive CpGs within the gene sequence exceeded a 5% change in value compared to controls.

CONCLUSION

This chapter answers important questions often raised in the identification of episignatures for NDDs. How does the presence of a highly similar gene, in this case the paralogous counterpart to *KAT6A, KAT6B*, affect the derivation of a sensitive and specific episignature? Does the similarity methylation of profiles in highly similar genes make identification of specific DNA methylation biomarkers impossible? And finally, does the shared methylation character coincide with shared phenotypic characteristics shared between the NDDs associated with each respective gene? These are important distinctions to make, as it provides evidence towards future assessment of episignatures in the context of homologous genes, but provides a better understanding of DNA methylation profiling in the context of shared etiology. Furthermore, assessment of the regions of overlap in terms of the genome and phenotype provides functional evidence to elucidate the common molecular pathways which result in the commonality seen in these disorders.

In my work, I demonstrated evidence of an episignature derived from the training of classifier models with patients with *KAT6A* variants, and a cohort of age and sex matched controls, however, when compared with data from patients carrying mutations in the paralogous gene *KAT6B* reduced specificity was evident. Understandably, due to their nature as exchangeable subunits within the MOZ/MORF complex, and common function as lysine acetyl transferases, the disruption of both *KAT6A* and *KAT6B* resulted in significant changes in DNA methylation at common loci, perhaps explaining the significant phenotypic overlap observed between these conditions. Identification of a specific episiganture for *KAT6A* was made possible through a modified statistical analysis pipeline by training against *KAT6B* samples along with unaffected controls. Once this was performed, a *KAT6A* episignature with a high degree of sensitivity and specificity was derived. This episignature can be applied to the reclassification of the large number of variants of unknown significance currently reported in ClinVar for *KAT6A* (89/410, 21%, See Appendix Table 4-3). This expansion of the analytical pipeline provides a novel increase in diagnostic power for the derivation of DNA methylation episignatures, providing a concrete example of episignature specificity, even in the context of multiple confounders in terms of genetic and phenotypic overlap.

I believe this process can be further applied to future cohorts, with the understanding that our ability to identify causative genes and explain their effects should be considered in the

context of broader gene function and molecular pathways. *KAT6A* and *KAT6B* have shared molecular function, resulting in a large degree of overlap in their methylation profiles, and clinical phenotypes, which must be accounted for in the search for effective diagnostic biomarkers. The interactions observed between *KAT6A* and *KAT6B* demonstrates a network of molecular pathways that are well described by DNA methylation analysis. Data presented in this chapter connects the genome, epigenome and phenotype in a way that mirrors the interconnected nature of human biology, while providing highly sensitive and specific diagnostic biomarkers. The functional information provided by DNA methylation analysis will be crucial in future work when we begin to combine the layers of diagnostic evidence from genomic and phenotype derived sources. This could result in the elucidation of common pathways that result in shared phenotypes, identifying hitherto unknown molecular interactors, and evaluating the effects of changes in regulatory and non coding sequences of DNA and how they relate to human health.

Chapter 5: Overlapping clinical and episignature phenotypes between BIS and Helsmoortel-Van Der Aa Syndrome

AUTHORS: Jack Reilly1, Gerarda Cappuccio2,3, Bekim Sadikovic1, Jennifer Kherhof1, Nicola Brunetti-Pierri2,3

INSTITUTIONS: 1London Health Sciences Centre and St. Joseph's Health Care London, Scientific and Clinical Director, Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, ON, Canada; 2Department of Translational Medicine, Federico II University of Naples, Naples, Italy; 3Telethon Institute of Genetics and Medicine, Pozzuoli, Italy.

PREFACE

Building on the evidence of epigenetic, genetic and phenotypic interactions discussed in the previous chapter, in this chapter I expand on the connection of shared epigenetic profiles between different genes, but partially overlapping phenotypes. To explore this, the final chapter of this work will discuss DNA methylation patterns in a cohort of patients with variants in two distinct genes that present with a partially overlapped phenotype. Specific variants in the activity dependent neuroprotector protein (*ADNP*), commonly associated with Helsmoortel Van Der AA (HVDAS), and the SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 2 (*SMARCA2*), commonly associated with both Nicolaides Baraitser syndrome (NCBRS), and Blepharophimosis impaired intellectual development syndrome (BIS), were assessed for evidence of a shared episignature in relation to their shared phenotypic features. *ADNP* and *SMARCA2*'s interactions have been investigated, finding that *ADNP* functions as a guide for the BAF complex that *SMARCA2* is a subunit of [4]. Disruption of *ADNP*'s sequence is hypothesized to result in similar disruption of the BAF complexes chromatin remodeling activity, potentially explaining the partially overlapping phenotype seen in this cohort of patients with variants in distinct genes.

In my work, I sought to describe this interaction in the context of the subsequent changes in DNA methylation, attempting to identify a common episignature that would represent the common changes in epigenome profiles. I describe the assessment of an episignature that is common to two distinct molecular entities, united by a common phenotypic feature, providing a novel example of a dysmorphology based approach towards episignature assessment. As I will demonstrate, this episignature was found to be highly sensitive and specific to a subset of *ADNP* variants in the central domain of the protein sequence, and BIS cases with *SMARCA2* variants, with the common blepharophimosis phenotype.

INTRODUCTION

For this work, we describe a novel episignature, involving a novel method of discovery, based on shared phenotype observed in patients with mutations in *SMARCA2*, and *ADNP*. Blepharophimosis-Impaired Intellectual Development syndrome (BIS) is a recently recognized disorder distinct from Nicolaides-Baraister syndrome that presents with distinct facial features of blepharophimosis, and global developmental delay [1,2]. BIS is due to pathogenic variants in *SMARCA2*, that encodes the catalytic subunit of the superfamily II helicase group of the BRG1 and BRM-associated factors (BAF) forming the BAF complex, the mammalian homolog of switch/sucrose nonfermentable (SWI/SNF), a chromatin remodeling complex that regulates expression of several genes involved in chromatin remodeling and gene expression regulation. Individuals bearing variants within the bipartite nuclear localization (BNL) signal domain of *ADNP*, the gene responsible for Helsmoortel-Van Der Aa Syndrome (HVDAS) present with blepharophimosis and epicanthal folds, a striking overlap with the BIS phenotype. Interestingly, *ADNP* was found to interact with several major proteins of the SWI/SNF complex, with protein-protein interactions identified between *ADNP* and several of the BAF complex subunits identified by Mandel and Gozes in 2007 [3,4,5] *ADNP* was found to bind directly to *SMARCA2*, *SMARCA4*, and *SMARCC2* through its C terminal end, providing a molecular pathway through which disturbance of proper *ADNP* function can result in impacts on the BAF complex's ability to provide it's chromatin remodeling capacity.

SMARCA2's role within the cell is to encode a catalytic subunit of the superfamily II helicase group of the BRG1 and BRM-associated factors (BAF), of which there are two [1]. The BAF complex is involved in chromatin remodeling activity, regulating expression of a number of genes, as such, disruption of the *SMARCA2* gene has been seen to be involved in several NDDs, including Nicolaides Baraitser syndrome, and our syndrome of interest, Blepharophimosis-Impaired Intellectual Development syndrome (BIS). BIS is a congenital disorder, with distinct facial features of blepharophimosis, and global developmental delay. Individuals living with BIS exhibit delayed motor skills, difficulties in independent locomotion, and impaired intellectual development with poor or absent speech.

ADNP, or the activity dependent neuroprotector homeobox, is a homeodomain zinc-finger protein with transcription factor activity. This protein, integral to proper brain formation [3] is known to interact with the SWI/SNF complex, and is commonly mutated in autism spectrum disorder (ASD) cases, being detected in 0.17% of individuals with ASD. Individuals with ADNP variants are commonly classified as Helsmoortel-Van Der Aa Syndrome (HVDAS), classically presenting with intellectual disability, developmental delay, motor dysfunction, ASD and facial dysmorphisms [6]. Previous work has shown that depending on the location of variants within the ADNP sequence, different episignatures can be derived, resulting in two episignatures for central domain and terminal domain regions respectively [8].

This manuscript focuses on the emergence of a potential third episignature derived from a cohort of ADNP patients, derived by means of a phenotype first, dysmorphology approach. It has been clinically observed that within a subset of patients with ADNP mutations, with variants with the nuclear bipartite localization domain [7] , symptoms of blepharophimosis were observed, a trait not seen in other ADNP patients, but common in SMARCA2 associated BIS cohorts. This led to the assessment of these molecularly distinct, but phenotypically overlapping cohorts via DNA methylation microarray testing, to determine whether or not the common features seen in these patients could be explained by shared changes in methylation patterning.

METHODS

SUBJECTS AND CONTROLS COHORTS

DNA samples were extracted from peripheral blood of 25 individuals with molecular variants identified in ADNP and SMARCA2. 10 ADNP samples, 8 of which were previously published in Bend et al, 2017 [8], were assessed, with an additional 2 provided by the Telethon Institute which featured characteristics of blepharophimosis. (See Table 5-1). The 15 remaining samples were associated with SMARCA2 variants, and featured clinical features of BIS (See Table 5-2). All samples and records were de-identified. Age and sex matched controls were selected from the Episign Knowledge Database, and represented various other conditions not

associated with ADNP or SMARCA2 related disorders, including NCBRS and other BAFopathy conditions.

The research was conducted in accordance with the Declaration of Helsinki. Study protocol has been approved by the Western University Research Ethics Board (REB 106302). Informed consent was obtained by physicians for use of the clinical information of the described patients.

Case ID	Blepharophimosis	Gender	Age (years)	*ADNP variant*
MS1167	+	Female	12	c.2157C>G (p.Tyr719*)
MS1182	-	Female	3	c.2157C>G (p.Tyr719*)
MS1254	-	Female	5	c.2157C>G (p.Tyr719*)
MS1255	-	Female	4	c.2157C>G (p.Tyr719*)
MS1262	NR	Male	13.4	c.2156dupA (rs1135401808)
MS1265	NR	Male	4.2	c.2188C>T (p.Arg730*)
MS1276	-	Male	5.4	c.2157C>G (p.Tyr719*)
MS1277	-	Female	12	c.2156_2157insA (p.Tyr719*)
MS2687	+	Female	8	c.2157C>G:(p.Tyr719*)
MS4091	+	Female	6	c.2157del, p.(Tyr719*)

Table 5-1. BNL-ADNP Sample Table. +: present; -: absent, NR: Not Reported

Case ID	Gender	Age (year)	SMARCA2 variant
MS1982	male	7	c.1585C>G, p.(Leu529Val)
MS1984	male	12	c.2810G>A;p.Arg937His
MS1988	male	8	c.2810G>A;p.Arg937His
MS1989	male	8	c.2810G>A;p.Arg937His
MS2524	male	0.2	c.2810G>A;p.Arg937His
MS4929	male	0.5	c.2809C>T, p.(Arg937Cys)
MS5212	female	17.4	c.1538G>T, p.(Gly513Val)
MS5213	female	3	c.2809C>T, p.(Arg937Cys)
MS5214	female	3	c.1573C>T, p.(Arg525Cys)
MS5215	female	5.4	c.2566A>G, p.(Met856Val)
MS5216	male	5	c.1573C>T, p.(Arg525Cys)
MS6938	female	13	c.1574G>A, p.(Arg525His)
MS7249	female	11.4	c.1534G>A, p.(Glu512Lys)
MS7416	female	9	c.1585C>G, p.(Leu529Val)
MS7417	female	17	c.6286C>A, p.(Asp510Gly)

Table 5-2. BIS Sample Table. The clinical features of these cases have been previously described in greater detail in Cappuccio et al.[1]

CLINICAL CHARACTERISTICS

Patients were assessed by clinicians at several separate sites, with blepharophimosis being identified in samples MS2687, and MS4091, by Dr Brunetti. An additional 8 samples, previously published in Bend et al, 2017 [8], were assessed via photographs for characteristics of blepharophimosis, however two samples, MS1262, and MS1265 did not have photographs available, and were labeled as "Not Reported". One additional sample was identified from photograph analysis published in the original Bend cohort. This sample, MS1167, was identified as having the blepharophimosis phenotype.

METHYLATION ARRAY AND QUALITY CONTROL

DNA methylation analysis and Episignature classifier development was performed using previously established protocol [9-12]. Genomic DNA was extracted from peripheral blood samples using standard techniques and followed by bisulfite conversion and hybridization to the Illumina infinium methylation EPIC bead chip arrays, according to the manufacturer's protocol. Idat files, containing methylated and unmethylated signal intensity plots (beta values) were produced from these microarrays, and used for analysis in R 4.0.2. Normalization was performed using the Illumina Infinium methylation EPIC array with background correction from the minfi package [12]. Previously defined exclusion criteria [9,10,11] were used to exclude probes with detection p values >0.01, probes on the x and y chromosomes, probes known to contain SNPs at the site of CpG interrogation or single nucleotide extension, and probes known to cross react with chromosomal locations other than their target regions were removed. All samples were examined for genome wide methylation distribution and those deviating from a bimodal distribution were excluded. Factor analysis using a principal component analysis (PCA) was performed to examine batch effect and identify outliers.No case samples were identified for removal.

DNA METHYLATION PROFILING

Probe methylation levels (beta values), were calculated as the ratio of signal intensity in methylated probes vs total sum of unmethylated and methylated probes, resulting in values ranging from zero to one. To allow for linear regression modeling, beta values were logit

transformed using the limma package [13], allowing for the identification of differentially methylated probes. Data was adjusted for the blood cell type composition as per Houseman et al [14]. Estimated blood cell proportion was added to the model matrix of the linear models as confounding variables [15]. Using the eBayes function in the limma package [16], p values were moderated and corrected for multiple testing using the Benjamini Hochberg method. Probes with the most significant methylation differences are selected using two facts from this dataset, the level of methylation difference (relative methylation signal intensity), and the probability that an observed difference is due to random chance (p values). Evaluation of this interaction is carried out by multiplying the absolute methylation difference between affected cases and controls by the negative value of the log transformed p values, and ranking the top 1000 probes with the highest values from this transformation. Next, receiver operating characteristic analysis (ROC) is performed on each probe, to measure the pairwise correlation coefficient between probes. Probes with low area under curve values from ROC analysis are removed, as well as highly correlated probes, eliminating probes with low sensitivity and specificity, and probes with highly correlated characteristics using Pearson's correlation coefficient. This ensures that the final probeset contains the most differentiating, non-redundant probes that are not influenced by random data structures. Only probes with a methylation difference greater than 5% were included in this analysis. This probe filtering process was designed to avoid reporting of probes with low effect size, and those influenced by technical or random variations as conducted in previous studies [9,10,11].

SELECTION OF MATCHED CONTROLS FOR METHYLATION PROFILING

For episignature characterization, mapping of probes and feature selection, matched controls were randomly selected from the LHSC EpiSign Knowledge Database (EKD)[11]. All of the ADNP and BIS samples were assayed, therefore all the controls selected for episignature identification were analyzed using the same array type. Samples were matched by age, sex and batch using the MatchIt package. A 4:1 ratio of controls to cases was deemed optimal for this analysis, as previously described [9,10,11]. PCA analysis was performed after each attempt at matching to detect outliers and determine data structures for the presence of batch effect. Outlier

160

samples, and those with highly aberrant data structures were removed, and subsequent matching trials were performed until consistent iterations with no outliers in the first two components of the PCA were derived.

CLUSTERING AND DIMENSION REDUCTION

Hierarchical clustering and multidimensional scaling were used after each iteration of analysis to examine the data structure of the identified episignature. Hierarchical clustering was performed using Ward's method on Euclidean distance by the base stats package in R, and visualized with the ggplot2 package [17,18]. Multidimensional scaling provides a visual representation of sample methylation profile similarity based on the scaling of the pairwise Euclidean distances between each sample.

DISCOVERY/TRAINING COHORT SELECTION

Identification of disease specific episignatures was performed using a randomly selected sub-setting of the database, on a 75:25 ratio of discovery:training, using the caTools package in R. Testing samples were used to assess the performance of the classification model developed later in the study. For every disease group in the discovery cohort, a sex and age matched control group with a sample size at least 4 times larger was selected from the reference control group using the MatchIT package, and methylation profiles were compared between the two.

CROSS VALIDATION

For each round of validation, one of the 25 selected ADNP-BIS samples was removed from probe selection, alongside matched controls. The remaining ADNP-BIS samples were designated as testing samples, and all three groups were modeled using multidimensional scaling to determine how they cluster/segregate with one another. This process was repeated with different combinations of assigned training and testing samples until all cases had been removed from probe selection and used for testing once. (See Supplemental Figures)

CLASSIFICATION MODEL

Specificity of the episignature was assessed using the Methylation Variant Pathogenicity (MVP) score, using all the identified probes. A support vector machine (SVM) used a linear

kernel for training on ADNP-BIS cases and controls. Once again, a 4:1 ratio of controls to cases was used to divide both the case and control samples previously matched and used for probe selection into training and testing cohorts for the SVM. Furthermore, the remaining unselected samples from the EKD were also divided similarly (75% training, 25% testing) to allow for comparison and testing of signature robustness against all of the samples in the EKD. Using the e1071 R package, we performed 10-fold cross validation to determine hyperparameters optimal for episignature classification. In this process, the training set was divided into ten folds by random assignment, where the first nine are used for training, and the last used for testing the accuracy of the model. The mean accuracy over all rounds was then calculated, and hyperparameters with the best performance by this metric were selected. The model provides a score ranging from 0-1 for each subject, representing the model's confidence in predicting whether the subject has a DNA methylation profile similar to the ADNP-BIS probe set or not. Conversion of these SVM decision values was done using Platt's scaling method [19,20], and the class obtaining the greatest score determined the predicted phenotype. A classification as ADNP-BIS was made when a sample received the greatest score for that class (normally greater than 0.5). Finally, the model was applied to both a training set of a large cohort of individuals with clinical and molecular diagnoses of neurodevelopmental disorders, as well as a group of healthy controls to determine its effective specificity.

VALIDATION OF CLASSIFICATION

To ensure the model is not susceptible to the batch structure of the methylation experiment, the classifier was applied to samples assayed on the same batch as the cases used for training. Using methylation data from individuals without a confirmed diagnosis of ADNP-BIS within the EKD assayed on the same microarray chip as case samples, methylation profiles were modeled to ensure the classifier is not confounded by technical artifacts unique to the given microarray. Specificity was determined by supplying a large number of DNA methylation arrays from unaffected subjects to the model. To further assess the specificity of the ADNP-BIS classifier relative to other neurodevelopmental disorder we applied it to cases with other patient cohorts exhibiting distinct episignatures within the EKD.

RESULTS

DETECTION AND VERIFICATION OF A SHARED EPISIGNATURE FOR ADNP-BIS
PATIENTS

DNA methylation profiles from the peripheral blood of 15 individuals with confirmed
clinical and molecular presentations of BIS syndrome and 10 individuals classified as ADNP
cases with variants within the bipartite nuclear localization domain, were used to derive the
episignature. Sample filtering steps were performed to measure sample quality, sample type, and
sample's clustering patterns (samples clustering with controls based on preliminary assessment
are excluded), however no samples required removal . Samples had fewer than 1000 failed
probes and passed quality control requirements. Comparisons were carried out, matching ADNP-
BIS samples with age, sex and batch-matched controls at a ratio of 4:1 (4 matched controls for
each case sample).

Cases showed significant differences in methylation patterns of 164 probes, which are
visualized using a volcano plot (See Figure 5-1A). Probes with a minimum methylation
difference of 5% between the two cohorts, and a multiple testing corrected p value of <0.01
(limma multivariable regression modeling) were used for the episignature. P values were
adjusted for blood cell type composition to ensure comparability between heterogeneous
peripheral blood sample sources. MDS and heatmap models showed significant overlap in
clustering patterns between the ADNP and BIS cases, with all 25 cases receiving high scores in
the classifier (See Figures 5-1B, 5-1C, and 5-1D), representing the significant degree of overlap
between these conditions. Hierarchical clustering and multiple dimensional scaling (MDS)
demonstrate that the selected probeset strongly separates cases and controls (See Figure 5-1B).
Cross-validation using ADNP-BIS samples was performed to validate the sensitivity of the
episignature, showing in each case that the remaining testing samples clustered with the other
ADNP-BIS samples, and segregated from the controls (See Supplementary Figure 5-1).

COMPARATIVE ANALYSIS OF EPISIGNATURES

Plotting ADNP-BIS samples alongside other ADNP variants in the central domain of the protein, not associated with the ADNP-BNL domain, showed distinct clustering of BIS samples, with some minor overlap with the ADNP-BNL domain variants with other ADNP samples in the MDS model (Figure 5-2A). Conversely, plotting ADNP-BIS samples alongside ADNP variants in the terminal domain of the protein (HVDAS_T), not associated with the ADNP-BNL domain, showed methylation profiles seen in the ADNP-BIS cohort do not match those seen in the HVDAS_T samples (Figure 5-2B). Similarly, plotting ADNP-BIS samples alongside SMARCA2 associated BAFopathy-NCBRS cases, showed methylation profiles seen in the ADNP-BIS cohort do not match those seen in the BAFopathy-NCBRS samples (Figure 5-2C).

A

B

ADNP-BNL
BIS
Control (training)

pvalueAdjustMethod: BH; score: 900; methDiffValue: 0.05; rocValue: 300; corValue: 0.75; corMethod: 1

C

Control (training)
ADNP-BNL
BIS

Methylation

D

Figure 5-1. **ADNP-BIS Episignature Models A**. Bimodal distribution plot of mean methylation difference vs -log p-value for each probe, represented as circles on the plot. Probes highlighted in red indicate the probes chosen following preliminary analysis, wherein the most highly differentiated probes with statistically significant p-values are selected for representation **B**. Multidimensional scaling plot representing the dimensions of variation in methylation signal intensity at informative CpG identified for training using ADNP-NBL and BIS cases. Represents comparisons of ADNP-NBL and BIS patients against the age and sex matched controls from the EKD. Cases marked in red represent ADNP-NBL cases, while orange represent BIS cases, blue indicate Control (training) cases with no phenotypic or genotypic presentation of either case group, including samples with confirmed presentation of other syndromes. Variant labels are noted on each sample. **C**. DNA methylation signal intensity plot comparing confirmed BIS syndrome patients and ADNP-NBL patients against training controls consisting of age and sex matched controls from the episign knowledge database. Samples are sorted by hierarchical clustering using ward's method. Cases marked in red atop the figure represent ADNP-NBL cases, while orange represent BIS cases, blue indicate cases with no phenotypic or genotypic presentation of either case group, including samples with confirmed presentation of other syndromes. **D**. SVM classifier model for ADNP-NBL and BIS cases. Each sample receives scores for the probability of having a DNA methylation profile similar to cases as compared to controls. Higher value on Y axis indicates that a sample presents a methylation profile more similar to cases compared to controls. All case samples used for probe selection (n=24) received high scores (>0.95), one ADNP associated HVDAS C sample received an elevated score (0.43), while the remaining samples within the episign knowledge database received low scores <0.05)
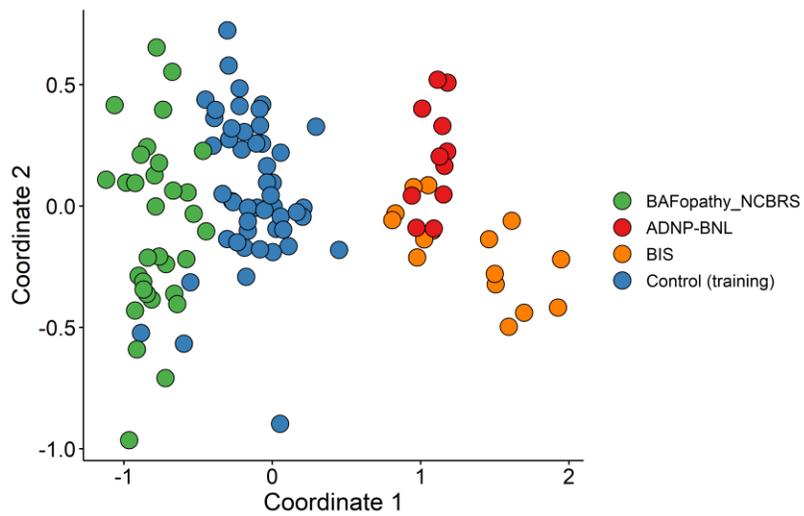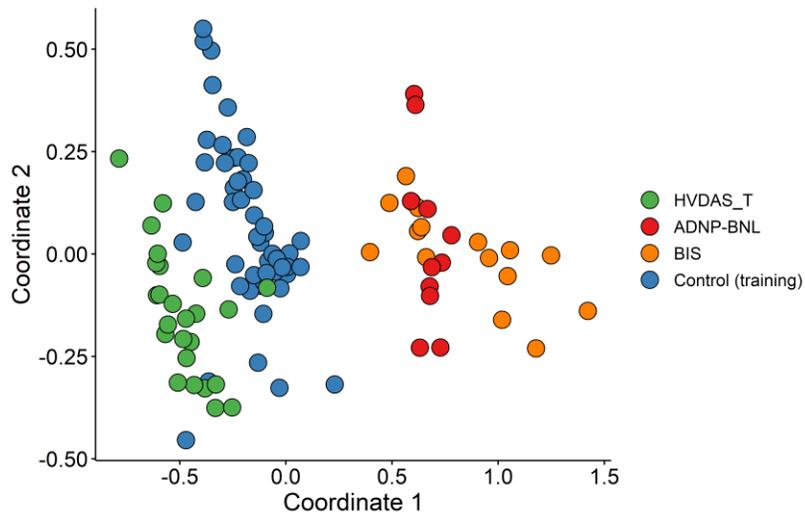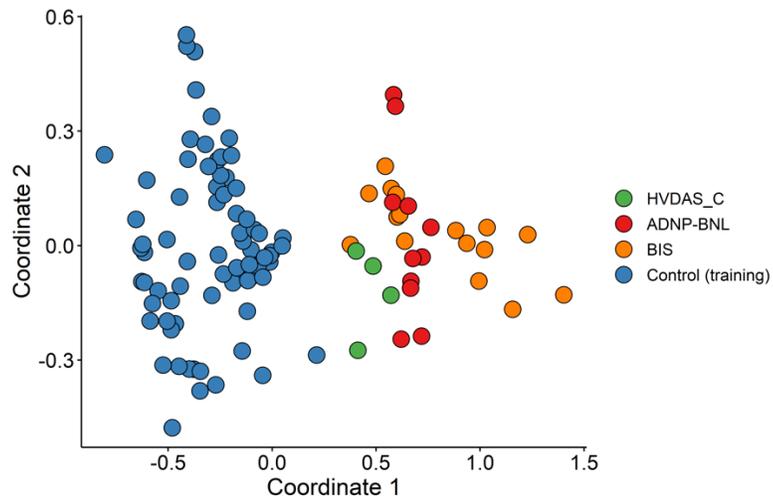
Figure 5-2. Plotting Additional Sample Types onto ADNP-BIS episignature Models

A. Multidimensional scaling plot representing the dimensions of variation in methylation signal intensity at informative CpG identified for training using ADNP-NBL and BIS cases. Represents comparisons of ADNP-NBL and BIS patients against the age and sex matched controls from the EKD as well as central domain ADNP variant associated HVDAS_C cases which were not included in probe selection. Cases marked in red represent ADNP-NBL cases, while orange represent BIS cases, green indicate HVDAS_C cases that involve variants outside of the bipartite nuclear localization domain, and blue indicate Control (training) cases with no phenotypic or genotypic presentation of either case group, including samples with confirmed presentation of other syndromes. **B**. Multidimensional scaling plot representing the dimensions of variation in methylation signal intensity at informative CpG identified for training using ADNP-NBL and BIS cases. Represents comparisons of ADNP-NBL and BIS patients against the age and sex matched controls from the EKD as well as terminal domain associated ADNP variant HVDAS_T cases which were not included in probe selection. Cases marked in red represent ADNP-NBL cases, while orange represent BIS cases, green indicate HVDAS_T cases that involve variants outside of the bipartite nuclear localization domain, and blue indicate Control (training) cases with no phenotypic or genotypic presentation of either case group, including samples with confirmed presentation of other syndromes. **C**. Multidimensional scaling plot representing the dimensions of variation in methylation signal intensity at informative CpG identified for training using ADNP-NBL and BIS cases. Represents comparisons of ADNP-NBL and BIS patients against the age and sex matched controls from the EKD as well as SMARCA2 associated BAFopathy-NCBRS cases which were not included in probe selection. Cases marked in red represent ADNP-NBL cases, while orange represent BIS cases, green indicate BAFopathy-NCBRS cases that involve variants outside of the bipartite nuclear localization domain, and blue indicate Control (training) cases with no phenotypic or genotypic presentation of either case group, including samples with confirmed presentation of other syndromes.

DEVELOPMENT OF AN MVP SCORE

ADNP-BIS subjects, matched controls, and a number of samples from syndrome cohorts previously established in the EKD were used for training a model to test the sensitivity and specificity of the episignature created through previous probe selection steps. The MVP score was set to generate a single score from 0-1 for each sample, with 1 being a methylation pattern highly similar to the case samples, and 0 being a methylation pattern highly similar to matched control samples. The class obtaining the highest score determined the episignature classification. These results were validated by a series of tests to validate their reliability. While all ADNP-BIS samples used for probe selection (not including the 4 BIS samples removed in previous rounds of analysis) received high scores close to 1, control scores remained near 0, indicating the classifier has a high sensitivity for the detection of the shared ADNP-BIS episignature (See Figure 5-6, and Supplemental Figure 5-S2). Furthermore, specificity of the classifier was tested by providing it with a large number of subjects with confirmed diagnosis of an NDD of various types, including trinucleotide repeat expansion abnormalities, imprinting defect disorders, BAFopathies, Mendelian disorders of the epigenetic machinery, down syndrome as well as subjects with nonsyndromic autism spectrum disorders. The vast majority of EKD samples, including the ADNP variant associated episignatures for the central and terminal domain Helsmoortel Van Der AA syndome (HVDAS_C, HVDAS_T), were classified as being highly similar to controls using the ADNP-BIS classifier, confirming its efficacy as a sensitive and specific model for the identification of the shared methylation profiles of BIS and ADNP-BNL cases. One HVDAS C case, involving an ADNP frameshift mutation further downstream from the bipartite nuclear localization domain received an elevated score (0.43), possibly due to the close proximity to the domain of interest and having the same variant type. Investigations of this case did not reveal any evidence of blepharophimosis phenotype.

DISCUSSION

DNA methylation analysis has been used in the analysis of an expanding number of neurodevelopmental disorders (NDD), implicating the changes in methylation distribution to the underlying biology of various complex mendelian conditions. Over sixty NDDs have been identified which exhibit distinct alterations in DNA methylation, referred to as episignatures,

now used for diagnostic clinical testing [21-32]. This process usually begins with genetic variant information, which is then used to inform the selection of samples for episignature discovery at various levels of detail, including gene level, domain specific episignatures and variant type specific signatures, as well as comparisons across cohorts with specific clinical features. Rarely, though growing in incidence as the understanding of phenotypic-epigenetic interactions increases, the process begins with the identification of clinical features that comprise a single disorder, or multiple. In this work, we expand upon the analytical framework developed via the continuing work of the EpiSign project, seeking to describe a novel method of identifying an episiganture for a specific subset of patients with a common phenotype. A common characteristic of blepharophimosis was noted in a subset of ADNP patients with frameshift mutations within the bipartite nuclear localization signal domain of the ADNP gene, required for localization to the nucleus, as well as cellular export and import signals, indicating that this gene has roles throughout the various cellular compartments. [33-36]. These patients matched the blepharophimosis phenotype commonly observed in Blepharophimosis intellectual disability syndrome (BIS), which are associated with variants in SMARCA2. Despite being associated with variants in two distinct molecular entities, these samples were shown to have a common methylation profile when analyzed, providing evidence towards potential shared regulatory pathways which contribute to the shared blepharophimosis. Although the exact mechanism through which ADNP-BNL mutations lead to the observed phenotype is not fully realized at this point, it seems to be specific to the effects of this particular truncation in the ADNP protein [7]. This is further supported by the findings in the 2014 paper from helsmoortel et al [37], that found that despite these frameshift mutations resulting in truncated mRNA transcripts, at least 4 patients were observed to have these transcripts escape nonsense mediated decay. As a result, lacking the ability to enter the nucleus without the BNL domain [38,39], these proteins could exert effects elsewhere within the cell, a capability which has been show in further studies of ADNP's function in cytoplasmic and extracellular compartments [3,39]. Furthermore, given the interactions observed between ADNP and SMARCA2 [4], it is possible that the chromatin remodeling activities of SMARCA2 are at least somewhat reliant on interactions with the ADNP protein, as variants lacking in the BNL domain and unable to transport into the nucleus of the cell for chromatin remodeling activity in tandem with SMARCA2's BAF complex [4,5,11,37], regulating key stages of embryonic development, potentially resulting in similar phenotypes to

those observed in patients with SMARCA2 mutations. ADNP's effect on the embryonic development stage has been shown in mouse knockout models, wherein the importance of ADNP was demonstrated with embryonic lethality in complete deficiency models, as well as significant evidence of chromatin-immunoprecipitation.[4, 33].

Although the blepharophimosis phenotype was not observed in all patients carrying ADNP-BNL domain disrupting variants, the presence of this phenotype, uncharacteristic of the presentation seen in the vast majority of other ADNP related HVDAS patients, provided a novel path towards the derivation of an episignature. Despite a large number of samples recruited for testing after the identification of blepharophimosis in the original cohort not presenting with blepharophimosis, the DNA methylation profiles exhibited by these patients continued to match those seen in SMARCA2 associated BIS patients, providing a novel situation wherein two distinct molecular entities that are not paralogous, or part of a common protein complex showed common DNA methylation changes. [4,11]. Further analysis of the changes in DNA methylation, in tandem with gene expression studies could reveal the source of the variation in phenotype observed despite the seemingly common variant effect observed in the ADNP cohort. Furthermore, studies into the exact interactions of the SMARCA2 and ADNP genes will provide insights into the molecular pathways that dictate their apparent epigenetic and phenotypic overlap.

REFERENCES

1. Cappuccio, Gerarda et al. "De novo SMARCA2 variants clustered outside the helicase domain cause a new recognizable syndrome with intellectual disability and blepharophimosis distinct from Nicolaides-Baraitser syndrome." *Genetics in medicine : official journal of the American College of Medical Genetics* vol. 22,11 (2020): 1838-1850. doi:10.1038/s41436-020-0898-y

2. Zaki, Maha S et al. "Blepharophimosis-ptosis-intellectual disability syndrome: A report of nine Egyptian patients with further expansion of phenotypic and mutational spectrum." *American journal of medical genetics. Part A* vol. 182,12 (2020): 2857-2866. doi:10.1002/ajmg.a.61857

3. Shmuel Mandel, Gideon Rechavi, Illana Gozes,Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis, Developmental Biology, Volume 303, Issue 2, 2007, Pages 814-824, ISSN 0012-1606, https://doi.org/10.1016/j.ydbio.2006.11.039.

4. Vandeweyer, Geert et al. "The transcriptional regulator ADNP links the BAF (SWI/SNF) complexes with autism." *American journal of medical genetics. Part C, Seminars in medical genetics* vol. 166C,3 (2014): 315-26. doi:10.1002/ajmg.c.31413

5. Lessard, Julie et al. "An essential switch in subunit composition of a chromatin remodeling complex during neural development." *Neuron* vol. 55,2 (2007): 201-15. doi:10.1016/j.neuron.2007.06.019

6. Breen, Michael S et al. "Episignatures Stratifying Helsmoortel-Van Der Aa Syndrome Show Modest Correlation with Phenotype." *American journal of human genetics* vol. 107,3 (2020): 555-563. doi:10.1016/j.ajhg.2020.07.003

7. Takenouchi, Toshiki et al. "Further evidence that a blepharophimosis syndrome phenotype is associated with a specific class of mutation in the ADNP gene." *American journal of medical genetics. Part A* vol. 173,6 (2017): 1631-1634. doi:10.1002/ajmg.a.38126

8. Bend, Eric G et al. "Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome." *Clinical epigenetics* vol. 11,1 64. 27 Apr. 2019, doi:10.1186/s13148-019-0658-5

9. Aref-Eshghi E, Rodenhiser DI, Schenkel LC, et al. Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. Am J Hum Genet. 2018;102(1):156-174. doi:10.1016/j.ajhg.2017.12.008

10. Aref-Eshghi E, Bend EG, Colaiacovo S, et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. Am J Hum Genet. 2019;104(4):685-700. doi:10.1016/j.ajhg.2019.03.008

11. Aref-Eshghi E, Bend EG, Hood RL, et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin–Siris and Nicolaides–Baraitser syndromes. Nat Commun. 2018;9(1):4885. doi:10.1038/s41467-018-07193-y

12. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049

13. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. doi:10.1093/nar/gkv007

14. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86. doi:10.1186/1471-2105-13-86

15. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One. 2012;7(7):e41361. doi:10.1371/journal.pone.0041361

16. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3. doi:10.2202/1544-6115.1027

17. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2009.

18. Joe H. Ward. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association. 1963;58(301):236-244. doi:10.1080/01621459.1963.10500845

19. Platt JC. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In: *Advances in Large Margin Classifiers*. MIT Press; 1999:61-74.

20. Smola AJ, Bartlett PJ. Advances in Large Margin Classifiers. MIT Press; 2000.

21. Bjornsson HT. The Mendelian disorders of the epigenetic machinery. Genome Res. 2015;25(10):1473-1481. doi:10.1101/gr.190629.115

22. Choufani, S.; Cytrynbaum, C.; Chung, B.H.Y.; Turinsky, A.L.; Grafodatskaya, D.; Chen, Y.A.; Cohen, A.S.A.;Dupuis, L.; Butcher, D.T.; Siu, M.T.; et al. NSD1 mutations generate a genome-wide DNA methylation signature. Nat. Commun.2015,6.

23. Schenkel, L.C.; Kernohan, K.D.; McBride, A.; Reina, D.; Hodge, A.; Ainsworth, P.J.; Rodenhiser, D.I.;Pare, G.; Bérubé, N.G.; Skinner, C.; et al. Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome.Epigenet. Chromatin 2017,10, 1–11.

24. Aref-Eshghi, E.; Bend, E.G.; Colaiacovo, S.; Caudle, M.; Chakrabarti, R.; Napier, M.; Brick, L.; Brady, L.;Carere, D.A.; Levy, M.A.; et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in GeneticallyUnsolved Individuals with Suspected Hereditary Conditions.Am. J. Hum. Genet.2019,104, 685–700.

25. Butcher, D.T.; Cytrynbaum, C.; Turinsky, A.L.; Siu, M.T.; Inbar-Feigenberg, M.; Mendoza-Londono, R.;Chitayat, D.; Walker, S.; Machado, J.; Caluseriu, O.; et al. CHARGE and Kabuki Syndromes: Gene-SpecificDNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically OverlappingConditions.Am. J. Hum. Genet.2017,100, 773–788

26. Hood, R.L.; Schenkel, L.C.; Nikkel, S.M.; Ainsworth, P.J.; Pare, G.; Boycott, K.M.; Bulman, D.E.; Sadikovic, B.The defining DNA methylation signature of Floating-Harbor Syndrome.Sci. Rep.2016,6, 1–9.

27. Schenkel, L.C.; Aref-Eshghi, E.; Skinner, C.; Ainsworth, P.; Lin, H.; Paré, G.; Rodenhiser, D.I.; Schwartz, C.;Sadikovic, B. Peripheral blood epi-signature of Claes-Jensen syndrome enables sensitive and specific identification of patients and healthy carriers with pathogenic mutations in KDM5C.Clin. Epigenet.2018,10,1–11.

28. Schenkel, L.C.; Schwartz, C.; Skinner, C.; Rodenhiser, D.I.; Ainsworth, P.J.; Pare, G.; Sadikovic, B. Clinical Validation of Fragile X Syndrome Screening by DNA Methylation Array.J. Mol. Diagn.2016,18, 834–841.

29. Li, Y.; Chen, J.A.; Sears, R.L.; Gao, F.; Klein, E.D.; Karydas, A.; Geschwind, M.D.; Rosen, H.J.; Boxer, A.L.;Guo, W.; et al. An Epigenetic Signature in Peripheral Blood Associated with the Haplotype on 17q21.31,a Risk Factor for Neurodegenerative Tauopathy.PLoS Genet.2014,10.

30. Aref-Eshghi, E.; Schenkel, L.C.; Lin, H.; Skinner, C.; Ainsworth, P.; Paré, G.; Siu, V.; Rodenhiser, D.;Schwartz, C.; Sadikovic, B. Clinical Validation of a Genome-Wide DNA Methylation Assay for MolecularDiagnosis of Imprinting Disorders.J. Mol. Diagn.2017,19, 848–856.

31. Guastafierro, T.; Bacalini, M.G.; Marcoccia, A.; Gentilini, D.; Pisoni, S.; Di Blasio, A.M.; Corsi, A.; Franceschi, C.;Raimondo, D.; Spanò, A.; et al. Genome-wide DNA methylation analysis in blood cells from patients with Werner syndrome.Clin. Epigenet.2017,9, 1–10.

32. Aref-Eshghi, E.; Kerkhof, J.; Pedro, V.P.; Barat-Houari, M.; Ruiz-Pallares, N.; Andrau, J.C.; Lacombe, D.;Van-Gils, J.; Fergelot, P.; Dubourg, C.; et al. Evaluation of DNA Methylation Episignatures for Diagnosis andPhenotype Correlations in 42 Mendelian Neurodevelopmental Disorders.Am. J. Hum. Genet.2020,106,356–370

33. Furman, Sharon et al. "Subcellular localization and secretion of activity-dependent neuroprotective protein in astrocytes." *Neuron glia biology* vol. 1,3 (2004): 193-9. doi:10.1017/S1740925X05000013

34. Magico, Adam C, and John B Bell. "Identification of a classical bipartite nuclear localization signal in the Drosophila TEA/ATTS protein scalloped." *PloS one* vol. 6,6 (2011): e21431. doi:10.1371/journal.pone.0021431

35. Canela-Pérez, Israel et al. "Nuclear localization signals in trypanosomal proteins." *Molecular and biochemical parasitology* vol. 229 (2019): 15-23. doi:10.1016/j.molbiopara.2019.02.003

36. Chung, Jeeyun et al. "Nuclear import of hTERT requires a bipartite nuclear localization signal and Akt-mediated phosphorylation." *Journal of cell science* vol. 125,Pt 11 (2012): 2684-97. doi:10.1242/jcs.099267

37. Helsmoortel, Céline et al. "A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP." *Nature genetics* vol. 46,4 (2014): 380-4. doi:10.1038/ng.2899

38. Liang, S H, and M F Clarke. "A bipartite nuclear localization signal is required for p53 nuclear import regulated by a carboxyl-terminal domain." *The Journal of biological chemistry* vol. 274,46 (1999): 32699-703. doi:10.1074/jbc.274.46.32699

39. Furman, Sharon et al. "Subcellular localization and secretion of activity-dependent neuroprotective protein in astrocytes." *Neuron glia biology* vol. 1,3 (2004): 193-9. doi:10.1017/S1740925X05000013
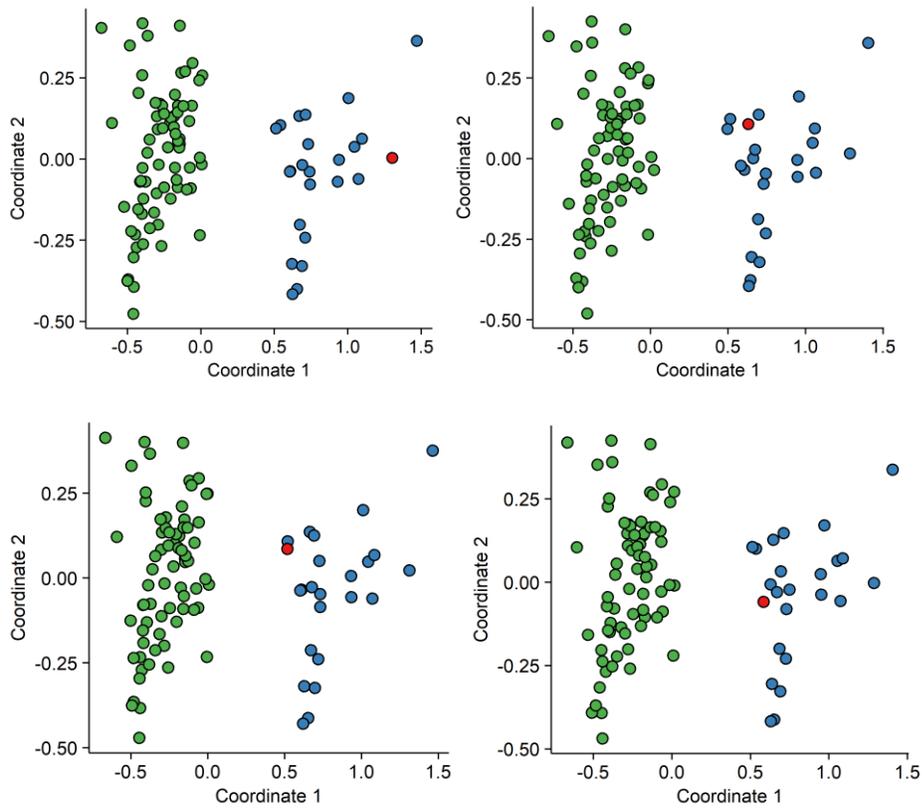
Supplemental Figure 5-1. Multidimensional scaling plot for cross validation, representing the dimensions of variation in methylation signal intensity at informative CpG identified for the shared SMARCA2-ADNP signature. Represents comparisons of the similarity of methylation profiles of SMARCA2-ADNP patients, with those chosen for probe selection marked in blue, and the cross-validation sample marked in red. Control samples (shown in green) include cases without a confirmed phenotypic and/or genotypic presentation of SMARCA2-ADNP, including samples with confirmed presentation of other syndromes.

ADNP BIS MVP Score ROC Graph

AUC = 0.933
P < 0.001
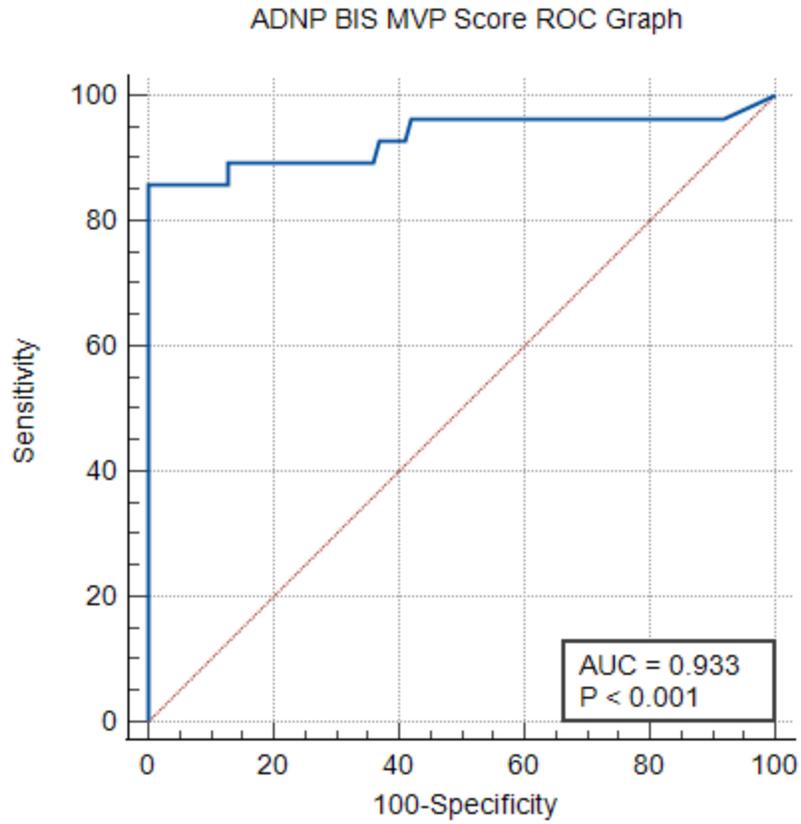
Supplemental Figure 5-2: ADNP BIS MVP Score ROC Graph: Receiver operating characteristic curve demonstrating the sensitivity and specificity of the generated MVP scores for the ADNP-BIS cohort and the remaining EKD samples used for training.

CONCLUSIONS

This chapter, concerning the identification of a common episignature between two distinct molecular entities, each associated with their own respective set of NDDs, on the basis of a shared dysmorphology presents an interesting proposition for the future of episignature discovery. Within the scope of the EpiSign project, a large number of NDDs have received highly sensitive and specific episignatures that are capable of differentiating these conditions on the basis of their DNA methylation profiles. In the vast majority of cases, this process began with the identification of genomic information, with chromosomal abnormalities, gene variants, and variant location being used to sort and provide supervised grouping of samples for the SVM based classifier. This approach led to a large number of gene specific episignatures, as well as interesting cases wherein episignature assessment led to the derivation of additional stratifications within groups of samples that have had their own unique DNA methylation profiles. In the case of HVDAS, associated with variants in the *ADNP* gene, localization of the variants within the central or terminal regions of the gene sequence provided two distinct episignatures [8], and even within this thesis, my third chapter concerning the identification of a domain specific episignature for K2BNDD provided further evidence of the relatively unexplored depths of potential episignatures guided by genomic information. With the success of this approach, as well as the constant emergence of phenotypic trends within various episignatures discussed within this work, it seems pertinent to discuss the possibility of phenotype led discovery of episignatures as well.

Led by the observations of Dr Nicola Brunetti and his lab at the Telethon Institute of Genetics and Medicine, identifying the uncharacteristic appearance of blepharophimosis phenotypes within a subset of *ADNP* patients, I began to analyze these samples in tandem with *SMARCA2* associated *SMARCA2* variants. As a result, I identified a highly sensitive and specific episignature that can distinguish samples from both genetic origins, seemingly tied by the presence of a shared phenotype, opening the door to future work in phenotype led discovery of shared functional characteristics. Within ClinVar, a large repository of *ADNP* variants are classified as variants of unknown significance (56/264, 21%, see Appendix Table 5), which can potentially be resolved through the episignature analysis pipeline into either the ADNP central or terminal region signatures, as well as the newly described *ADNP-SMARCA2* shared episignature. Furthermore, *SMARCA2* variants also have a large number of variants currently classified as

VUSs (138/756, 18%, See Appendix Table 6) which can be similarly reclassified using the BAF complex or shared *ADNP-SMARCA2* episignature. The use of functional information such as DNA methylation is key to the better understanding of these variants which may include effects in non-coding regions, or splice sites. These are difficult to interpret using traditional genome based analysis methods which often depend on disturbance of protein structure or stability to indicate pathogenicity. Through episignature assessment, we can provide a better pathway to diagnosis when we compare the DNA methylation profiles to assess patterns of overlap in gene expression.

Although the phenotype observed in the preliminary cohort of patients was not shared by several of the new subjects used for deriving the episignature, the DNA methylation pattern derived from this cohort was shared between all subjects. The complex variability associated with NDDs, and gene regulatory networks can introduce a large amount of difficulty into the derivation of phenotype based biomarkers, but nonetheless, this study was propagated by phenotypic findings first, setting it apart from the remaining projects described in previous chapters. This type of discovery is rapidly gaining prominence in the field of molecular diagnostics, with various deep phenotyping databases providing significant increase in the ability to analyze and corroborate the functional evidence derived from analysis of various non-coding regions of the DNA, expression profiling, and phenotype ontology. In the future, I believe it will be essential to investigate common phenotypes between disorders, on the basis of functional information such as DNA methylation profiling, which can better account for the complexity observed in these conditions. Further investigations of episignatures through the lens of phenotypic data will assuredly provide novel avenues of understanding into the complex interactions of the genome, epigenome and ensuing phenotype observed in patients. By continuing to provide highly robust biomarkers such as the one described in this work to the expanding landscape of functional annotation databases will allow for disentanglement of the vast web of molecular interactions that dictate human health.

Chapter 6: Conclusions

While there have been major advancements in the understanding of human genetic sequence, , our ability to sequence the entirety of genomic DNA has fallen short of providing the comprehensive blueprint of the molecular etiology of hereditary neurodevelopmental disorders. . Moving beyond the one-gene one-disease paradigm [1], researchers have demonstrated that genetic conditions can be complex, resulting from multifactorial genetic and environmental causes, with confounding characteristics of incomplete penetrance, variable expressivity, and overlapping phenotypes.. Global population prevalence of genetic conditions is 2-3% [2], Clinicians and researchers using a myriad of tools from chromosomal microarrays to whole exome sequencing have attempted to resolve these conditions and provide effective diagnosis to those living with neurodevelopmental disorders.. Presently, conclusive diagnosis through NGS in both WES and targeted gene panels provide conclusive diagnosis in approximately 15-35% of cases [3-8], which despite vast improvements over previous methods, still means that a significant majority of persons living with neurodevelopmental disorders are not receiving a conclusive diagnosis. Often, a so-called diagnostic journey must be undertaken to enable diagnosis, involving extensive molecular testing, clinical phenotypes, and surveys of behavior and medical history, which can frequently result in inconclusive findings, leaving patients and their families seeking answers. Development of new diagnostic tools is necessary to ensure that those living with NDDs receive the accurate diagnosis and level of care they deserve. As such, my research has focused on the assessment of effectiveness of DNA methylation profiling in NDDs and my findings demonstrate it to be a powerful method for increasing the diagnostic yield in this patient population.

Epigenetics and DNA methylation are inherently tied to the function and expression of the genome, uniting the underlying biology dictated by our sequence of nucleotides, with the ensuing phenotypic presentation that we observe as phenotypes[9-12]. Epigenetic studies, such as profiling of patients global DNA methylation patterns, provide a functional link between disruptions of the genetic sequence and their downstream effects on chromatin.. Our lab has developed approaches for DNA methylation profiling to improve the diagnosis and understanding of complex neurodevelopmental disorders. This work has provided effective

biomarkers for over 70 conditions. The so-called episignatures, representing DNA methylation profiles at identified CpG dinucleotides which are unique to the patients with specific genetic disorders [13-23]. Methylation profiling is enabled by comparing data from specific patient cohorts to the EpiSign knowledge database which currently includes >10,000 methylation profiles across a growing number of genetic conditions allowing for the development of both highly sensitive, and specific biomarkers for these disorders. Episignatures can be used for the reclassification of variants of unknown significance, as they represent functional biological evidence of pathogenicity, as DNA methylation profiles which match the derived episignature imply common gene disruption. This enables molecular diagnosis in the absence of traditional genetic sequence findings [21,24,25].

Our database and catalog of episignatures has grown at an increasing pace since its inception, as has our understanding of the complexities of DNA methylation profiles in relation to both phenotype and genotype [13,23, 26, 27]. In my Episignatures are not always a linear one biomarker to one disorder phenomenon, but instead have a branching network of potential uses and outcomes, correlating with the gene level disruptions [14,15], as well as sub-gene level associations such as domain specific episignatures [28]. Furthermore, common episignatures can be associated with multiple genes, belonging to common molecular pathways or gene networks, , as well as shared phenotypes across distinct genetic disorders. My work describes the illustration of an "Episignature Roadmap" detailing the consequences of DNA methylation perturbations in relation to NDDs in a variety of genomic and phenotypic contexts.

In the second chapter of this thesis I described the identification of a gene-specific episignature for Gabriele De Vries syndrome associated with YY1 transcription factor A high level of sensitivity and specificity was found for this episignature, with all samples with the exception of one mapping consistently throughout each model, resulting in high MVP scores (>0.8) in the final classifier model, while the remaining samples within our EKD did not exceed an MVP score of 0.25, simultaneously demonstrating a high level of sensitivity and specificity. This was further enforced through a series of leave one out cross validation assessments, where despite the removal of each sample one by one from case cohort training, samples still received high MVP scores, indicating that the model is not sensitive to changes in the cohort composition, and indeed represents an effective method of identifying GADEVs patients through DNA

methylation. The exceptional case, referred to as YY1-atypical did not fit the episignature derived for this case cohort adding an interesting level of complexity, and indicates avenues of future research. The YY1-atypical case presented a distinctly different methylation profile when compared to other case samples, and upon further assessment of the clinical features, this patient was identified as having a striking atypical presentation of the disorder, with symptoms of overgrowth not seen in the rest of the cohort. Furthermore, the unique variant in the *YY1*-atypical case was not seen in the remainder of the *YY1* cohort, indicating that there is a potential additional signature tied to phenotypic differences, and domain or variant specific DNA methylation changes. As such, this chapter provides an example of an effective gene level episignature, while also exemplifying the necessity of assessing other patients with *YY1* mutations for further substratification according to differences in genetic and phenotypic data.

In the third chapter, I continued to explore the possibilities provided by episignature assessment through the delineation of a *KDM2B*- related episignature for an as of yet unnamed disorder we have termed *KDM2B* related neurodevelopmental disorder, or K2BNDD. Expanding on the presence of additional subsignatures hinted at in the atypical GADEVs case in the previous chapter, I discovered not only a sensitive and specific biomarker for patients with variants the the *KDM2B* sequence, but an additional sub-signature specific to a key domain in the *KDM2B* sequence. Disruption of the CxxC DNA binding motif of the *KDM2B* gene showed large changes in methylation not seen in the remainder of the *KDM2B* cohort, and when assessed, showed distinct clustering of these samples from both the matched controls and other *KDM2B* samples. This result was mirrored in the MVP classifier, with CxxC samples receiving high scores (>0.8) while the remaining samples scored very low, close to 0, indicating a very high level of specificity for the methylation changes observed. Additionally, the subsignature provided evidence of change at the phenotypic level as well, with CxxC patients corresponding to significant increases in the incidence of congenital anomalies and organ malformations, not seen in other cases. This demonstrates the effectiveness of these stratified episignatures, to enable the diagnosis at gene-level, while providing domain specific biomarkers that can explain differences in phenotype at a higher level of resolution.

In the fourth chapter, I shifted my focus from stratifying signatures from a single genetic origin, to assessing the effectiveness of comparing genes of multiple distinct origins with

overlapped episignatures. The first example of this involves the *KAT6A/KAT6B* homologues. I demonstrated the existence of shared methylation profiles between these paralogous genes, which are associated with 3 distinct neurodevelopmental disorders.. These three disorders, *KAT6A* syndrome, GTPTS, and SBBYSS provided interesting insights into the effects of common genetic origin and function in relation to their overlapped DNA methylation profiles The models showed a significant amount of overlap in differentially methylated probes, correlated with the shared phenotype observed across these disorders. Samples in each of the case groups clustered more closely with one another than relative to the age and sex matched controls. Additionally, these disorders shared a large proportion of the differentially methylated genes. Interestingly, *KAT6A* syndrome and GTPTS shared a significantly higher number of differentially methylated genes (n=11) when compared to those shared between *KAT6A* syndrome and SBBYSS (n=1), which could explain some of the trends in shared phenotypes in these disorders. For example, in both GTPTS and *KAT6A* syndrome, *HEY2* , a gene important in heart development was differentially methylated, potentially explaining the increased frequency of heart anomalies in these patients groups (*KAT6A* syndrome; 53%, GTPTS; 65%) as compared to SBBYSS (38%) which did not exhibit differential methylation in this gene. These shared methylation profiles also required a customized strategy in order to derive a specific molecular classifier. Including the *KAT6B* samples to the training control cohort, enabled development of a specific classifier with full differentiation of the *KAT6A* syndrome samples. Hence, even in the presence of shared genomic, epigenomic and phenotypic features this approach allowed me to develop episignatures with a high level of sensitivity and specificity.

The fifth chapter, highlights the shared episignature in patients with *ADNP* and *SMARCA2* mutations. Here, I explored the effectiveness of a dysmorphology-first approach to the discovery of an episignature. Common clinical features of blepharophimosis were observed in a cohort of patients with genetic variants in two genes with unrelated functions, *ADNP* and *SMARCA2*. Although blepharophimosis is commonly associated with patients with *SMARCA2* variants in association with Blepharophimosis and Intellectual Development syndrome (BIS), it is rare to see in patients with *ADNP* variants commonly associated with Helsmoortel Van Der AA syndrome. As a result, by assessing patients with the common clinical phenotypesI aimed to assess evidence of a phenotype-specific DNA methylation episignature that may be common between these two distinct patient groups. I demonstrated a common, highly sensitive and

specific episignature biomarker capable of delineating these patients with a common blepharophimosis phenotype, highlighting the use of such an approach. This provided further evidence that DNA methylation patterns can be linked to both the genomic and phenotypic presentations across different genetic disorders, which may affect common molecular pathways.

Overall, my work demonstrates a strong correlation between genetic, epigenetic and phenotypic patterns observed in patients with genetic neurodevelopmental disorders, along with providing effective molecular diagnostic biomarkers. Changes in DNA methylation reflect the intricate network of the underpinning genetic changes and the consequent phenotypes. Several endeavors are underway already in various laboratories attempting to increase our understanding of the complex networks of interactions between these three levels of diagnostic evidence, providing powerful new databases that integrate variant classification information, transcriptomics, and phenotypic information. The ENCODE project [29] has provided a searchable database of extensive information of various methods of assessing molecular biology, including gene expression analysis through assessment of active chromatin states, DNA methylation levels, and RNA binding, alongside variant information to provide deeper insights into the interaction of these various molecular mechanisms. This work has provided extensive improvements to our understanding of the functional elements of the genome including DNA methylation and histone modifications, and their effects on chromatin accessibility to modulate gene transcription, transcription factor binding networks, and non-coding RNAs [30,31,32]. This database provides a powerful refutation of the claim that the majority of our DNA is "junk" leftover from evolutionary pressures, and instead a vast sea of information for the functional regulation of the transcribed genome. EpiXCan [33,34] provides similar investigative power, assessing the transcriptome in relation to gene-trait associations. This database incorporates epigenomic data alongside variant information, gene expression, and regulatory annotations to provide a better understanding of transcriptomics, facilitating combination of large existing databases of genetic variation in coding regions, alongside less well researched non-coding regions of DNA. Through focusing on trait-associated biological pathways, this project exemplifies the transition towards a more holistic view of molecular interactions, beyond the classical one-gene-one disease paradigm... The Human Phenome Ontology (HPO) takes this work one step further. It includes a standardized vocabulary of phenotypic abnormalities associated with a specific genetic diseases that can be used by researchers to inform

understanding of disparate disorders through a phenotype based view of pathology, by correlating genetic conditions on the basis of their common clinical presentations related to their ontological terms [35]. The HPO database synthesizes model organism data, WGS/WES data, and phenotype ontology terms to network their various interactions, potentially explaining the common etiology of complex disorders such as NDDs. This "deep-phenotyping" approach, is a powerful new tool in the assessment of rare complex disorders, and has been adopted by a number of organizations in their search for pathways that relate to disease etiologies. This work includes advances in phenotype annotations for various conditions and areas of research, such as epilepsy disorders, mitochondrial DNA - phenome correlations, neurology, immunology and even expansion into phenome ontology traits for model organisms such as mouse and zebra fish [36,37,38,39]. Creation of large scale databases of standardized phenotypic characteristics, paired with extensive genome and epigenomic data aids in the molecular diagnosis, but can also be used to further develop ontology terms and their associated disease algorithms, continually self-refining, similar to how our own EKD operates. As such, the creation of analytical frameworks through which we can guide the identification and analysis of these complex disorders and the traits that cause them is paramount in ensuring that we can effectively improve our ability to diagnose persons living with these conditions, as well as better understand the underlying biology which unites them.

Numerous other projects have begun addressing additional concerns associated with transcriptomic reliability and translatability. First and foremost, the effects of different tissue types on the gene expression networks associated with different conditions is well established, with epigenetic patterning such as DNA methylation differing extensively from tissue to tissue within a single organism, in relation to the cell lineage specification of gene expression that results in the various cell types within our bodies. [40,41,42,43]. Within my research I focused on peripheral blood, which has been shown to be an effective substrate for the identification of DNA methylation profiles in relation for disorders where primary clinical phenotypes affect alternate tissues, namely CNV tissues in the context of the NDDs [25, 44,45, 46] . However it can be questioned whether or not gene expression profiles from can de reliant in detecting disorder specific episignatures. Variations in the epigenome arising from germline disruptions in early embryonic stages are propagated through cell lineage specification and represented across

multiple tissues in the developed organism [18, 47], arguing that the epigenetic changes in peripheral blood may represent effective surrogates for the similar disruptions in other tissues. Analysis of tissue specific epigenetic changes has also been carried out by several entities, including the gene ORGANizer project and the Genotype-Tissue Expression (GTEx) database. [48,49]. The Gene ORGANizer project has aided in assessing not only the transcriptomic, genomic and phenotypic expression of traditional model tissues such as peripheral blood, tumor samples, or fibroblasts, but more disease relevant tissue sources, including those which are more difficult to obtain such as brain or CNV tissue, or those which are rarely sampled from (bone, larynx, spinal cord etc) [49]. This is further complicated by the fact that many tissues can be assessed at different developmental stages, taken post mortem, and extracted from particular sections of a given organ.. Gene ORGANizer makes use of several of the databases I have previously mentioned, including use of HPO for gene-phenotype associations, as well as ClinVar, and Uniprot for gene-disease associations. In attempting to improve this relatively unexplored aspect of confounding factors in gene expression analysis, the Gene ORGANizer project seeks to represent phenotype based associations in relation to genomic traits in a wide range of human tissues, connecting the expression profiles for the entirety of the human body.

In summary, my research focused on integration of genetic and phenotypic features to establish the epigenomic correlations and develop diagnostic biomarkers. I hypothesized that DNA methylation episignatures can be used to provide highly robust biomarkers for the classification of neurodevelopmental disorders, and further stratification of these signatures can identify key epigenetic patterns that relate to the phenotypic and genotypic variations seen in patients with these disorders. To demonstrate this, I identified specific aims related to the derivation of DNA methylation episignatures in a variety of molecular contexts, defining syndrome specific, and domain specific episignatures, as well as signatures in the context of homologous genes, and distinct disorders with shared phenotypic characteristics. In chapter 2, I hypothesized the existence of a gene specific episignature for Gabriele De Vries Syndrome. Demonstrated through various models, I found that the methylation patterns displayed by patients with *YY1* disruptions were shared across the cohort, enabling sensitive and specific identification of this disease group. In chapter 3, analysis of K2BNDD was carried out to derive a disorder specific episignature, as well as an investigation into the domain specific trends in DNA methylation caused by disruption of the CxxC DNA binding domain of *KDM2B*. My

results demonstrate a robust signature for the K2BNDD cohort, as well as methylation profiles for patients with disruptions of the CxxC domain, distinct from the remainder of the cohort. Chapter 4 demonstrated the derivation of an episignature in the context of the gene homologues, *KAT6A* and *KAT6B*. This provided a sensitive and specific episignature, even in the presence of highly overlapping molecular and phenotypic characteristics, as well as highlighting potential pathways which explain these shared characteristics. Finally, chapter 5 discussed the assessment of two molecularly distinct disorders with a shared phenotype, identifying common DNA methylation changes between specific ADNP mutations and SMARCA2 associated BIS cases, demonstrating the effectiveness of episignatures in More focused analyses, such as domain specific investigations, or the grouping of syndromes with disparate genetic origins, can be used to demonstrate associations with common phenotypes. My work has provided effective biomarkers that can enable shortening of the diagnostic odyssey in a number of genetic NDD conditions, increasing diagnostic yield in the process. Whether it is beginning with gene-level analysis, or phenotypic observations, distinct episignatures can be derived, revealing a network of intricate molecular connections.

REFERENCES

1. Cerrone, Marina et al. "Beyond the One Gene-One Disease Paradigm: Complex Genetics and Pleiotropy in Inheritable Cardiac Disorders." *Circulation* vol. 140,7 (2019): 595-610. doi:10.1161/CIRCULATIONAHA.118.035954

2. Timpano S, Picketts DJ. Neurodevelopmental Disorders Caused by Defective Chromatin Remodeling: Phenotypic Complexity Is Highlighted by a Review of ATRX Function. *Front Genet*. 2020;11:885. doi:10.3389/fgene.2020.00885

3. Lindy AS, Stosser MB, Butler E, et al. Diagnostic outcomes for genetic testing of 70 genes in 8565 patients with epilepsy and neurodevelopmental disorders. *Epilepsia.* May 2018;59(5):1062-1071.

4. Grozeva D, Carss K, Spasic-Boskovic O, et al. Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Human mutation.* Dec 2015;36(12):1197-1204.

5. Reuter MS, Tawamie H, Buchert R, et al. Diagnostic Yield and Novel Candidate Genes by Exome Sequencing in 152 Consanguineous Families With Neurodevelopmental Disorders. *JAMA psychiatry.* Mar 1 2017;74(3):293-299.

6. Evers C, Staufner C, Granzow M, et al. Impact of clinical exomes in neurodevelopmental and neurometabolic disorders. *Molecular genetics and metabolism.* Aug 2017;121(4):297-307.

7. Redin C, Gerard B, Lauer J, et al. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of medical genetics.* Nov 2014;51(11):724-736.

8. Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American journal of human genetics.* May 14 2010;86(5):749-764.

9. Tost J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Molecular biotechnology.* Jan 2010;44(1):71-81.

10. Fahrner JA, Bjornsson HT. Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states. *Annual review of genomics and human genetics.* 2014;15:269-293.

11. Elhamamsy, Amr Rafat. "Role of DNA methylation in imprinting disorders: an updated review." *Journal of assisted reproduction and genetics* vol. 34,5 (2017): 549-562. doi:10.1007/s10815-017-0895-5

12. Jones, Peter A. "Functions of DNA methylation: islands, start sites, gene bodies and beyond." *Nature reviews. Genetics* vol. 13,7 484-92. 29 May. 2012, doi:10.1038/nrg3230

13. Levy MA, McConkey H, Kerkhof J, et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. *HGGADVANCE*. 2022;3(1). doi:10.1016/j.xhgg.2021.100075

14. Hood RL, Schenkel LC, Nikkel SM, et al. The defining DNA methylation signature of Floating-Harbor Syndrome. *Scientific reports.* Dec 9 2016;6:38803.

15. Haghshenas, Sadegheh et al. "Diagnostic Utility of Genome-Wide DNA Methylation Analysis in Mendelian Neurodevelopmental Disorders." *International journal of molecular sciences* vol. 21,23 9303. 6 Dec. 2020, doi:10.3390/ijms21239303

16. Haghshenas, Sadegheh et al. "Detection of a DNA Methylation Signature for the Intellectual Developmental Disorder, X-Linked, Syndromic, Armfield Type." *International journal of molecular sciences* vol. 22,3 1111. 23 Jan. 2021, doi:10.3390/ijms22031111

17. Ciolfi A, Aref-Eshghi E, Pizzi S, et al. Frameshift mutations at the C-terminus of HIST1H1E result in a specific DNA hypomethylation signature. *Clinical epigenetics.* Jan 7 2020;12(1):7.

18. Schenkel LC, Aref-Eshghi E, Skinner C, et al. Peripheral blood epi-signature of Claes-Jensen syndrome enables sensitive and specific identification of patients and healthy carriers with pathogenic mutations in KDM5C. *Clinical epigenetics.* 2018;10:21.

19. Aref-Eshghi E, Schenkel LC, Lin H, et al. Clinical Validation of a Genome-Wide DNA Methylation Assay for Molecular Diagnosis of Imprinting Disorders. *The Journal of molecular diagnostics : JMD.* Nov 2017;19(6):848-856.

20. Kernohan KD, Cigana Schenkel L, Huang L, et al. Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. *Clinical epigenetics.* 2016;8:91.

21. Aref-Eshghi E, Schenkel LC, Lin H, et al. The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. *Epigenetics.* 2017;12(11):923-933.

22. Aref-Eshghi E, Bend EG, Hood RL, et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. *Nature communications.* Nov 20 2018;9(1):4885.

23. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *American journal of human genetics.* Feb 26 2020.

24. Aref-Eshghi, Erfan et al. "Genome-wide DNA methylation and RNA analyses enable reclassification of two variants of uncertain significance in a patient with clinical Kabuki syndrome." *Human mutation* vol. 40,10 (2019): 1684-1689. doi:10.1002/humu.23833

25. Aref-Eshghi, Erfan et al. "Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions." *American journal of human genetics* vol. 104,4 (2019): 685-700. doi:10.1016/j.ajhg.2019.03.008

26. Kerkhof, Jennifer et al. "DNA methylation episignature testing improves molecular diagnosis of Mendelian chromatinopathies." *Genetics in medicine : official journal of the American College of Medical Genetics* vol. 24,1 (2022): 51-60. doi:10.1016/j.gim.2021.08.007

27. Sadikovic, Bekim et al. "Functional annotation of genomic variation: DNA methylation episignatures in neurodevelopmental Mendelian disorders." *Human molecular genetics* vol. 29,R1 (2020): R27-R32. doi:10.1093/hmg/ddaa144

28. Bend EG, Aref-Eshghi E, Everman DB, et al. Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. *Clinical epigenetics.* Apr 27 2019;11(1):64.

29. ENCODE Project Consortium. "The ENCODE (ENCyclopedia Of DNA Elements) Project." *Science (New York, N.Y.)* vol. 306,5696 (2004): 636-40. doi:10.1126/science.1105136

30. Moraes, Fernanda, and Andréa Góes. "A decade of human genome project conclusion: Scientific diffusion about our genome knowledge." *Biochemistry and molecular biology*

education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology vol. 44,3 (2016): 215-23. doi:10.1002/bmb.20952

31. Tragante, Vinicius et al. "The ENCODE project and perspectives on pathways." *Genetic epidemiology* vol. 38,4 (2014): 275-80. doi:10.1002/gepi.21802

32. Ding, Nan et al. *Yi chuan, " The ENCODE project and functional genomics studies", Hereditas* vol. 36,3 (2014): 237-47.

33. Zhang, W., Voloudakis, G., Rajagopal, V.M. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat Commun* 10, 3834 (2019). https://doi.org/10.1038/s41467-019-11874-7

34. Li, Ling et al. "Transcriptome-wide association study of coronary artery disease identifies novel susceptibility genes." *Basic research in cardiology* vol. 117,1 6. 17 Feb. 2022, doi:10.1007/s00395-022-00917-8

35. Sebastian Köhler, Leigh Carmody, Nicole Vasilevsky, Julius O B Jacobsen, Daniel Danis, Jean-Philippe Gourdine, Michael Gargano, Nomi L Harris, Nicolas Matentzoglu, Julie A McMurry, David Osumi-Sutherland, Valentina Cipriani, James P Balhoff, Tom Conlin, Hannah Blau, Gareth Baynam, Richard Palmer, Dylan Gratian, Hugh Dawkins, Michael Segal, Anna C Jansen, Ahmed Muaz, Willie H Chang, Jenna Bergerson, Stanley J F Laulederkind, Zafer Yüksel, Sergi Beltran, Alexandra F Freeman, Panagiotis I Sergouniotis, Daniel Durkin, Andrea L Storm, Marc Hanauer, Michael Brudno, Susan M Bello, Murat Sincan, Kayli Rageth, Matthew T Wheeler, Renske Oegema, Halima Lourghi, Maria G Della Rocca, Rachel Thompson, Francisco Castellanos, James Priest, Charlotte Cunningham-Rundles, Ayushi Hegde, Ruth C Lovering, Catherine Hajek, Annie Olry, Luigi Notarangelo, Morgan Similuk, Xingmin A Zhang, David Gómez-Andrés, Hanns Lochmüller, Hélène Dollfus, Sergio Rosenzweig, Shruti Marwaha, Ana Rath, Kathleen Sullivan, Cynthia Smith, Joshua D Milner, Dorothée Leroux, Cornelius F Boerkoel, Amy Klion, Melody C Carter, Tudor Groza, Damian Smedley, Melissa A Haendel, Chris Mungall, Peter N Robinson, Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources, Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D1018–D1027, https://doi.org/10.1093/nar/gky1105

36. Köhler, Sebastian et al. "The Human Phenotype Ontology in 2021." *Nucleic acids research* vol. 49,D1 (2021): D1207-D1217. doi:10.1093/nar/gkaa1043

37. Ratnaike, Thiloka E et al. "MitoPhen database: a human phenotype ontology-based approach to identify mitochondrial DNA diseases." *Nucleic acids research* vol. 49,17 (2021): 9686-9695. doi:10.1093/nar/gkab726

38. Haimel, Matthias et al. "Curation and expansion of Human Phenotype Ontology for defined groups of inborn errors of immunity." *The Journal of allergy and clinical immunology* vol. 149,1 (2022): 369-378. doi:10.1016/j.jaci.2021.04.033

39. Köhler, Sebastian et al. "The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data." *Nucleic acids research* vol. 42,Database issue (2014): D966-74. doi:10.1093/nar/gkt1026

40. Ryall, James G et al. "Metabolic Reprogramming of Stem Cell Epigenetics." *Cell stem cell* vol. 17,6 (2015): 651-662. doi:10.1016/j.stem.2015.11.012

41. Ohkura, Naganari et al. "Development and maintenance of regulatory T cells." *Immunity* vol. 38,3 (2013): 414-23. doi:10.1016/j.immuni.2013.03.002

42. Kalish, Jennifer M et al. "Epigenetics and imprinting in human disease." *The International journal of developmental biology* vol. 58,2-4 (2014): 291-8. doi:10.1387/ijdb.140077mb

43. Bishop, Karen S, and Lynnette R Ferguson. "The interaction between epigenetics, nutrition and the development of cancer." *Nutrients* vol. 7,2 922-47. 30 Jan. 2015, doi:10.3390/nu7020922

44. Schenkel, Laila C et al. "Constitutional Epi/Genetic Conditions: Genetic, Epigenetic, and Environmental Factors." *Journal of pediatric genetics* vol. 6,1 (2017): 30-41. doi:10.1055/s-0036-1593849

45. Schenkel, Laila C et al. "DNA methylation analysis in constitutional disorders: Clinical implications of the epigenome." *Critical reviews in clinical laboratory sciences* vol. 53,3 (2016): 147-65. doi:10.3109/10408363.2015.1113496

46. Herceg, Zdenko et al. "Roadmap for investigating epigenome deregulation and environmental origins of cancer." *International journal of cancer* vol. 142,5 (2018): 874-882. doi:10.1002/ijc.31014

47. Barbosa, M., Joshi, R.S., Garg, P. *et al.* Identification of rare de novo epigenetic variations in congenital disorders. *Nat Commun* 9, 2064 (2018). https://doi.org/10.1038/s41467-018-04540-x

48. GTEx Consortium. "The Genotype-Tissue Expression (GTEx) project." *Nature genetics* vol. 45,6 (2013): 580-5. doi:10.1038/ng.2653

49. Gokhman, David et al. "Gene ORGANizer: linking genes to the organs they affect." *Nucleic acids research* vol. 45,W1 (2017): W138-W145. doi:10.1093/nar/gkx302

APPENDICES

Appendix Table 1. Table of reported variants for *YY1,* retrieved from ClinVar databases (https://www.ncbi.nlm.nih.gov/clinvar/?term=YY1%5Bgene%5D&redir=gene)

Name	Clinical significance (Last reviewed)	Accession
GRCh37/hg19 14q32.2(chr14:100704697-100706973)x0	Benign(Last reviewed: Dec 1, 2011)	VCV000614478
GRCh37/hg19 14q32.2(chr14:100691178-100706973)x1	Benign(Last reviewed: Dec 14, 2011)	VCV000614476
NM_003403.5(YY1):c.141G>C (p.Glu47Asp)	Benign(Last reviewed: Dec 31, 2019)	VCV000716403
GRCh37/hg19 14q32.2(chr14:100704886-100706973)x1	Benign(Last reviewed: Feb 16, 2012)	VCV000614479
GRCh37/hg19 14q32.2(chr14:100701417-100717821)x1	Benign(Last reviewed: Oct 20, 2010)	VCV000614477
NM_003403.5(YY1):c.30C>T (p.Ala10=)	Likely benign(Last reviewed: Dec 18, 2018)	VCV000743296
NM_003403.5(YY1):c.306G>A (p.Glu102=)	Likely benign(Last reviewed: Dec 31, 2019)	VCV000795809
NM_003403.5(YY1):c.842+9T>C	Likely benign(Last reviewed: Jul 16, 2018)	VCV000759598
GRCh37/hg19 14q32.2-32.33(chr14:98051841-107285437)x3	Likely pathogenic	VCV000395888
NM_003403.5(YY1):c.527G>A (p.Gly176Asp)	Likely pathogenic(Last reviewed: Dec 11, 2019)	VCV001029703

NM_003403.5(YY1):c.690dup (p.Asp231fs)	Likely pathogenic(Last reviewed: Dec 21, 2020)	VCV001331550
GRCh37/hg19 14q32.2(chr14:100655021-100742092)x1	Likely pathogenic(Last reviewed: Jul 10, 2019)	VCV000980621
NM_003403.5(YY1):c.1118A>G (p.His373Arg)	Likely pathogenic(Last reviewed: Jul 7, 2020)	VCV000976756
NM_003403.5(YY1):c.1220A>G (p.His407Arg)	Likely pathogenic(Last reviewed: Jun 17, 2021)	VCV001184871
NM_003403.5(YY1):c.1015A>C (p.Lys339Gln)	Likely pathogenic(Last reviewed: Jun 9, 2017)	VCV000520982
NM_003403.5(YY1):c.550_551del (p.Ser184fs)	Likely pathogenic(Last reviewed: May 14, 2018)	VCV000545935
GRCh37/hg19 14q32.2-32.31(chr14:99737888-101847855)	Likely pathogenic(Last reviewed: Nov 1, 2018)	VCV000625815
GRCh37/hg19 14q23.2-32.33(chr14:62493932-107285437)x3	Pathogenic	VCV000397361
GRCh37/hg19 14q11.2-32.33(chr14:19794561-107234280)x3	Pathogenic	VCV000395470
GRCh37/hg19 14q32.13-32.31(chr14:95871795-102457523)x1	Pathogenic(Last reviewed: Apr 1, 2021)	VCV001340262
GRCh38/hg38 14q24.3-32.33(chr14:73655772-106879298)x3	Pathogenic(Last reviewed: Apr 30, 2010)	VCV000144518
GRCh38/hg38 14q24.3-32.33(chr14:77222795-106879298)x3	Pathogenic(Last reviewed: Apr 5, 2012)	VCV000149176

GRCh38/hg38 14q32.2-32.33(chr14:97938637-106855263)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057085
GRCh38/hg38 14q32.12-32.33(chr14:91455861-106832642)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000058527
GRCh38/hg38 14q31.3-32.33(chr14:86094030-106832642)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000058526
GRCh38/hg38 14q31.2-32.33(chr14:83912345-106855405)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000058525
NM_003403.5(YY1):c.1096C>G (p.Leu366Val)	Pathogenic(Last reviewed: Aug 17, 2018)	VCV000430619
NM_003403.5(YY1):c.1097T>C (p.Leu366Pro)	Pathogenic(Last reviewed: Aug 17, 2018)	VCV000430618
NM_003403.5(YY1):c.1138G>T (p.Asp380Tyr)	Pathogenic(Last reviewed: Dec 1, 2010)	VCV000430617
GRCh37/hg19 14q32.2-32.33(chr14:96829290-107287663)x1	Pathogenic(Last reviewed: Dec 18, 2019)	VCV000929832
GRCh38/hg38 14q11.2-32.33(chr14:20043514-106877229)x3	Pathogenic(Last reviewed: Dec 2, 2014)	VCV000155306
NM_003403.5(YY1):c.1124G>A (p.Arg375Gln)	Pathogenic(Last reviewed: Dec 21, 2020)	VCV001331552
GRCh38/hg38 14q32.2-32.31(chr14:99930669-101022599)x1	Pathogenic(Last reviewed: Dec 22, 2010)	VCV000154707
GRCh38/hg38 14q32.2-32.33(chr14:99831655-106855263)x1	Pathogenic(Last reviewed: Dec 30, 2009)	VCV000144349
NM_003403.5(YY1):c.385del (p.Asp129fs)	Pathogenic(Last reviewed: Feb 14, 2019)	VCV000817604

GRCh38/hg38 14q32.12-32.33(chr14:92540983-104863658)x3	Pathogenic(Last reviewed: Feb 16, 2011)	VCV000146638
GRCh37/hg19 14q11.1-32.33(chr14:19000422-107289053)x3	Pathogenic(Last reviewed: Jan 1, 2013)	VCV000601776
GRCh38/hg38 14q11.2-32.33(chr14:20151149-106855263)x3	Pathogenic(Last reviewed: Jan 14, 2011)	VCV000146230
NM_003403.5(YY1):c.690del (p.Asp231fs)	Pathogenic(Last reviewed: Jan 14, 2021)	VCV001162319
GRCh37/hg19 14q11.2-32.33(chr14:19280733-107287663)x3	Pathogenic(Last reviewed: Jan 5, 2017)	VCV000601777
GRCh37/hg19 14q11.2-32.33(chr14:20511673-107285437)	Pathogenic(Last reviewed: Jul 14, 2015)	VCV000443977
NM_003403.5(YY1):c.1111C>T (p.Arg371Cys)	Pathogenic(Last reviewed: Jul 19, 2019)	VCV001186721
GRCh38/hg38 14q32.2-32.33(chr14:97638520-106855263)x3	Pathogenic(Last reviewed: Jul 30, 2009)	VCV000146074
NM_003403.5(YY1):c.535A>T (p.Lys179Ter)	Pathogenic(Last reviewed: Jul 7, 2017)	VCV000430621
NM_003403.5(YY1):c.1030C>T (p.Gln344Ter)	Pathogenic(Last reviewed: Jul 7, 2017)	VCV000430620
GRCh37/hg19 14q32.2-32.33(chr14:99794230-107285437)x3	Pathogenic(Last reviewed: Jun 19, 2017)	VCV000688581
GRCh37/hg19 14q24.2-32.33(chr14:73750741-107285437)x3	Pathogenic(Last reviewed: Jun 22, 2015)	VCV000442718
NM_003403.5(YY1):c.860_864del (p.Ile287fs)	Pathogenic(Last reviewed: Jun 28, 2017)	VCV000432981

GRCh37/hg19 14q32.2-32.33(chr14:100661319-107285437)x1	Pathogenic(Last reviewed: May 18, 2015)	VCV000443644
GRCh37/hg19 14q32.2-32.33(chr14:100575917-107281934)	Pathogenic(Last reviewed: Nov 1, 2018)	VCV000625744
NM_003403.5(YY1):c.1040_1041insCGACGGACAACGGCTAGTTTATTTTTACTTGCAGCTTCAAAACCGCCACCTTCCATTGCTTGTCCAGTGATACGGAGACCTTCCTCGGCAGCAAAACGAATCAATTCTGCTGTACG (p.Thr348_Gly349insAspGlyGlnArgLeuValTyrPheTyrLeuGlnLeuGlnAsnArgHisLeuProLeuLeuValGlnTer)	Pathogenic(Last reviewed: Nov 27, 2019)	VCV001323777
GRCh38/hg38 14q32.2-32.31(chr14:99794337-100944567)x1	Pathogenic(Last reviewed: Oct 1, 2010)	VCV000153097
GRCh38/hg38 14q32.2-32.33(chr14:99448012-106850609)x3	Pathogenic(Last reviewed: Oct 20, 2010)	VCV000146615
NM_003403.5(YY1):c.468_483del (p.Gly157fs)	Pathogenic(Last reviewed: Oct 23, 2020)	VCV000985223
GRCh37/hg19 14q32.2(chr14:100317190-101012999)	Pathogenic(Last reviewed: Oct 26, 2018)	VCV000813332
GRCh37/hg19 14q11.2-32.33(chr14:19327823-107287663)x3	Pathogenic(Last reviewed: Oct 31, 2014)	VCV000601778
GRCh37/hg19 14q32.12-32.33(chr14:91969028-107285437)x3	Pathogenic(Last reviewed: Oct 4, 2017)	VCV000687996
GRCh38/hg38 14q32.2-32.31(chr14:100262836-102500697)x1	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000150931

GRCh38/hg38 14q32.13-32.33(chr14:95524407-106879501)x1	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000146793
NM_003403.5(YY1):c.1115C>G (p.Thr372Arg)	Uncertain significance	VCV000091950
NM_003403.5(YY1):c.608A>G (p.Lys203Arg)	Uncertain significance(Last reviewed: Apr 7, 2021)	VCV001341784
GRCh38/hg38 14q32.2(chr14:100031805-100808500)x1	Uncertain significance(Last reviewed: Aug 12, 2011)	VCV000058202
NM_003403.5(YY1):c.514G>T (p.Val172Phe)	Uncertain significance(Last reviewed: Dec 15, 2018)	VCV001333898
GRCh37/hg19 14q32.2(chr14:98924025-101159952)x3	Uncertain significance(Last reviewed: Dec 6, 2017)	VCV000564133
NM_003403.5(YY1):c.562G>A (p.Gly188Ser)	Uncertain significance(Last reviewed: Feb 1, 2019)	VCV001029701
NM_003403.5(YY1):c.1032A>G (p.Gln344=)	Uncertain significance(Last reviewed: Feb 13, 2018)	VCV001032129
NM_003403.5(YY1):c.985G>C (p.Glu329Gln)	Uncertain significance(Last reviewed: Jun 22, 2021)	VCV001329515
NM_003403.5(YY1):c.-5C>T	Uncertain significance(Last reviewed: Mar 5, 2019)	VCV001029702
GRCh37/hg19 14q32.2(chr14:100744400-100910248)x3	Uncertain significance(Last reviewed: May 15, 2018)	VCV000685016
NM_003403.5(YY1):c.1114A>G (p.Thr372Ala)	Uncertain significance(Last reviewed: May 18, 2018)	VCV001053301

NM_003403.5(YY1):c.207CCA[7] (p.His80dup)	Uncertain significance(Last reviewed: May 20, 2019)	VCV000931670
NM_003403.5(YY1):c.956C>T (p.Thr319Ile)	Uncertain significance(Last reviewed: Nov 10, 2017)	VCV000985282
NM_003403.5(YY1):c.1159_1161del (p.Phe387del)	Uncertain significance(Last reviewed: Nov 19, 2018)	VCV000985714
NM_003403.5(YY1):c.742C>T (p.Pro248Ser)	Uncertain significance(Last reviewed: Oct 17, 2019)	VCV001309573
NM_003403.5(YY1):c.193C>T (p.His65Tyr)	Uncertain significance(Last reviewed: Oct 25, 2019)	VCV000954755
NM_003403.5(YY1):c.1106A>G (p.Asn369Ser)	Uncertain significance(Last reviewed: Oct 25, 2021)	VCV001321254
NM_003403.5(YY1):c.202G>A (p.Ala68Thr)	Uncertain significance(Last reviewed: Sep 13, 2019)	VCV000996852
GRCh38/hg38 14q32.2(chr14:100236766-100743192)x3	Uncertain significance(Last reviewed: Sep 27, 2013)	VCV000152326

Appendix Table 2. Table of reported variants for *KDM2B,* retrieved from ClinVar databases (https://www.ncbi.nlm.nih.gov/clinvar/?term=KDM2B%5Bgene%5D&redir=gene)

Name	Clinical significance (Last reviewed)	Accession
NM_032590.5(KDM2B):c.3522G>T (p.Leu1174=)	Benign(Last reviewed: Dec 31, 2019)	VCV000784262
NM_032590.5(KDM2B):c.1287C>T (p.Gly429=)	Benign(Last reviewed: Dec 31, 2019)	VCV000776760
NM_032590.5(KDM2B):c.3174C>T (p.Asp1058=)	Benign(Last reviewed: Dec 31, 2019)	VCV000720488
NM_032590.5(KDM2B):c.397+7C>T	Benign(Last reviewed: Feb 25, 2018)	VCV000715044
NM_032590.5(KDM2B):c.1326C>T (p.Gly442=)	Benign(Last reviewed: Jan 3, 2019)	VCV000773689

NM_025126.4(RNF34):c.530G>A (p.Arg177His)	Benign(Last reviewed: Jun 2, 2018)	VCV000778448
NM_032590.5(KDM2B):c.3050G>A (p.Arg1017His)	Likely benign	VCV000242896
NM_032590.5(KDM2B):c.1605C>T (p.Pro535=)	Likely benign(Last reviewed: Aug 15, 2017)	VCV000714574
NM_032590.5(KDM2B):c.83_84del (p.Thr28fs)	Likely benign(Last reviewed: Dec 29, 2019)	VCV000777690
NM_032590.5(KDM2B):c.56A>G (p.His19Arg)	Likely benign(Last reviewed: Mar 29, 2018)	VCV000770235
NM_032590.5(KDM2B):c.3699C>A (p.Ile1233=)	Likely benign(Last reviewed: Mar 29, 2018)	VCV000738407
GRCh37/hg19 12q24.22-24.33(chr12:117461902-133841395)x3	Likely pathogenic(Last reviewed: Nov 1, 2021)	VCV001330196
GRCh38/hg38 12q24.31(chr12:121471000-122459718)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057611
GRCh38/hg38 12q24.31(chr12:121325874-122505529)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057610
GRCh38/hg38 12q24.31(chr12:120504068-122459718)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057609
GRCh38/hg38 12q24.23-24.33(chr12:118165459-133182322)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057207
GRCh38/hg38 12q24.21-24.33(chr12:115131583-133166920)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059821
GRCh38/hg38 12q24.31-24.32(chr12:120718786-127500215)x1	Pathogenic(Last reviewed: Aug 19, 2010)	VCV000154387

GRCh38/hg38 12q24.31-24.33(chr12:120697672-133202490)x3	Pathogenic(Last reviewed: Aug 2, 2011)	VCV000148578
GRCh37/hg19 12p13.33-q24.33(chr12:173787-133777902)x3	Pathogenic(Last reviewed: Dec 2, 2014)	VCV000441983
GRCh37/hg19 12p13.33-q24.33(chr12:621220-133779118)x3	Pathogenic(Last reviewed: Jan 1, 2013)	VCV000613617
GRCh37/hg19 12p13.33-q24.33(chr12:191619-133777645)x3	Pathogenic(Last reviewed: Jan 5, 2017)	VCV000613610
GRCh37/hg19 12p13.33-q24.33(chr12:173787-133777902)	Pathogenic(Last reviewed: Jul 14, 2015)	VCV000441984
GRCh38/hg38 12q24.21-24.33(chr12:114268403-133201316)x3	Pathogenic(Last reviewed: Jul 18, 2014)	VCV000155589
GRCh37/hg19 12q24.23-24.33(chr12:120367241-133777645)x3	Pathogenic(Last reviewed: Jun 27, 2016)	VCV000601434
GRCh37/hg19 12p13.33-q24.33(chr12:1-133851895)x3	Pathogenic(Last reviewed: Sep 20, 2016)	VCV000268075
GRCh38/hg38 12p13.33-q24.33(chr12:121271-133196807)x3	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000150740
GRCh37/hg19 12q24.31(chr12:121441298-122107345)x3	Uncertain significance	VCV000394621
GRCh37/hg19 12q24.31(chr12:121882818-122666131)x1	Uncertain significance	VCV000395963
GRCh38/hg38 12q24.23-24.31(chr12:119286893-122638552)x3	Uncertain significance(Last reviewed: Aug 12, 2011)	VCV000058232

GRCh37/hg19 12q24.31(chr12:121899406-122234599)x3	Uncertain significance(Last reviewed: Jun 14, 2018)	VCV000815562
NM_032590.5(KDM2B):c.46C>T (p.Arg16Ter)	Uncertain significance(Last reviewed: Mar 29, 2016)	VCV000403007
GRCh37/hg19 12q24.31(chr12:121887337-123386068)x3	Uncertain significance(Last reviewed: Mar 3, 2021)	VCV001328462
GRCh37/hg19 12q24.31(chr12:121970346-122287290)x1	Uncertain significance(Last reviewed: May 1, 2021)	VCV001340538
GRCh37/hg19 12q24.31(chr12:121903358-122234650)x3	Uncertain significance(Last reviewed: May 16, 2018)	VCV000686313
GRCh37/hg19 12q24.31(chr12:121931513-122059588)x1	Uncertain significance(Last reviewed: May 3, 2018)	VCV000563906

Appendix Table 3. Table of reported variants for *ADNP,* retrieved from ClinVar databases (https://www.ncbi.nlm.nih.gov/clinvar/?term=ADNP%5Bgene%5D&redir=gene)

Name	Clinical significance (Last reviewed)	Accession
NM_001282531.3(ADNP):c.3236A>G (p.Asn1079Ser)	Benign(Last reviewed: Apr 1, 2020)	VCV001237918

NM_001282531.3(ADNP):c.2554A>G (p.Lys852Glu)	Benign(Last reviewed: Apr 23, 2018)	VCV000711607
NM_001282531.3(ADNP):c.1122C>T (p.Asn374=)	Benign(Last reviewed: Aug 24, 2020)	VCV001273051
NM_001282531.3(ADNP):c.108+124C>T	Benign(Last reviewed: Aug 31, 2018)	VCV001252442
NM_001282531.3(ADNP):c.1781A>G (p.Gln594Arg)	Benign(Last reviewed: Dec 31, 2019)	VCV000735993
NM_001282531.3(ADNP):c.2943G>T (p.Val981=)	Benign(Last reviewed: Dec 31, 2019)	VCV000712654
NM_001282531.3(ADNP):c.3278_3279dup (p.Gly1094fs)	Benign(Last reviewed: Feb 11, 2016)	VCV000225224
NM_001282531.3(ADNP):c.2666G>C (p.Ser889Thr)	Benign(Last reviewed: Jan 11, 2021)	VCV000588325
NM_001282531.3(ADNP):c.2743G>A (p.Val915Ile)	Benign(Last reviewed: Jan 14, 2021)	VCV000252698
NM_001282531.3(ADNP):c.*50A>C	Benign(Last reviewed: Jul 14, 2021)	VCV001189012
NM_001282531.3(ADNP):c.2067C>T (p.Gly689=)	Benign(Last reviewed: Jul 14, 2021)	VCV000587808
NM_001282531.3(ADNP):c.202-303G>A	Benign(Last reviewed: Jul 27, 2018)	VCV001272506
NM_001282531.3(ADNP):c.108+233G>A	Benign(Last reviewed: Jul 3, 2018)	VCV001260648
NM_001282531.3(ADNP):c.2568C>T (p.Val856=)	Benign(Last reviewed: Jul 3, 2018)	VCV000587839
NM_001282531.3(ADNP):c.909G>A (p.Met303Ile)	Benign(Last reviewed: Jul 30, 2020)	VCV001265979
NM_001282531.3(ADNP):c.2310T>C (p.Phe770=)	Benign(Last reviewed: Jun 16, 2021)	VCV000747523

NM_001282531.3(ADNP):c.3058C>G (p.Gln1020Glu)	Benign(Last reviewed: Mar 29, 2021)	VCV001278536
NM_001282531.3(ADNP):c.535A>G (p.Ile179Val)	Benign(Last reviewed: Mar 3, 2021)	VCV001236127
NM_001282531.3(ADNP):c.801C>G (p.Pro267=)	Benign(Last reviewed: Mar 5, 2020)	VCV001294552
NM_001282531.3(ADNP):c.3095C>G (p.Ser1032Cys)	Benign(Last reviewed: May 5, 2021)	VCV001287084
NM_001282531.3(ADNP):c.833AGA[1] (p.Lys279del)	Benign(Last reviewed: Nov 4, 2020)	VCV000589741
NM_001282531.3(ADNP):c.108+280A>C	Benign(Last reviewed: Oct 5, 2018)	VCV001262451
NM_001282531.3(ADNP):c.2076G>A (p.Lys692=)	Benign(Last reviewed: Sep 14, 2018)	VCV000587838
NM_001282531.3(ADNP):c.1163C>T (p.Ala388Val)	Benign(Last reviewed: Sep 3, 2020)	VCV001245671
NM_001282531.3(ADNP):c.1212G>C (p.Ser404=)	Benign/Likely benign(Last reviewed: Apr 12, 2021)	VCV000588523
NM_001282531.3(ADNP):c.2317A>G (p.Lys773Glu)	Benign/Likely benign(Last reviewed: Apr 13, 2020)	VCV000777582
NM_001282531.3(ADNP):c.2815A>C (p.Ile939Leu)	Benign/Likely benign(Last reviewed: Apr 9, 2019)	VCV000722218
NM_001282531.3(ADNP):c.2782G>C (p.Asp928His)	Benign/Likely benign(Last reviewed: Apr 9, 2021)	VCV000589370
NM_001282531.3(ADNP):c.1075A>G (p.Ile359Val)	Benign/Likely benign(Last reviewed: Dec 14, 2021)	VCV000718326

NM_001282531.3(ADNP):c.3185T>C (p.Ile1062Thr)	Benign/Likely benign(Last reviewed: Dec 15, 2020)	VCV000713739
NM_001282531.3(ADNP):c.2971A>G (p.Met991Val)	Benign/Likely benign(Last reviewed: Dec 29, 2017)	VCV000588466
NM_001282531.3(ADNP):c.2931A>G (p.Gly977=)	Benign/Likely benign(Last reviewed: Dec 31, 2019)	VCV000589201
NM_001282531.3(ADNP):c.422_424dup (p.Ser141dup)	Benign/Likely benign(Last reviewed: Dec 31, 2019)	VCV000434092
NM_001282531.3(ADNP):c.1752A>G (p.Gln584=)	Benign/Likely benign(Last reviewed: Dec 5, 2020)	VCV000588836
NM_001282531.3(ADNP):c.3279C>T (p.Ala1093=)	Benign/Likely benign(Last reviewed: Jan 27, 2020)	VCV000434090
NM_001282531.3(ADNP):c.2475G>T (p.Gly825=)	Benign/Likely benign(Last reviewed: Jul 1, 2021)	VCV000587880
NM_001282531.3(ADNP):c.393G>A (p.Pro131=)	Benign/Likely benign(Last reviewed: Nov 21, 2021)	VCV000588351
NM_001282531.3(ADNP):c.2749C>T (p.Pro917Ser)	Benign/Likely benign(Last reviewed: Sep 23, 2020)	VCV000723547
NM_001282531.3(ADNP):c.2617G>T (p.Asp873Tyr)	Conflicting interpretations of pathogenicity(Last reviewed: Apr 17, 2020)	VCV000434095
NM_001282531.3(ADNP):c.2772G>C (p.Glu924Asp)	Conflicting interpretations of pathogenicity(Last reviewed: Dec 31, 2019)	VCV000434093

NM_001282531.3(ADNP):c.1063G>A (p.Ala355Thr)	Conflicting interpretations of pathogenicity(Last reviewed: Dec 7, 2017)	VCV000499359
NM_001282531.3(ADNP):c.3213_3216del (p.Ser1071fs)	Likely benign	VCV000694537
NM_001282531.3(ADNP):c.2463C>T (p.Gly821=)	Likely benign(Last reviewed: Apr 1, 2021)	VCV000589626
NM_001282531.3(ADNP):c.3092G>A (p.Ser1031Asn)	Likely benign(Last reviewed: Apr 11, 2019)	VCV001336949
NM_001282531.3(ADNP):c.2352A>G (p.Arg784=)	Likely benign(Last reviewed: Apr 17, 2018)	VCV000740826
NM_001282531.3(ADNP):c.2573C>G (p.Ala858Gly)	Likely benign(Last reviewed: Apr 2, 2020)	VCV000931183
NM_001282531.3(ADNP):c.192G>A (p.Thr64=)	Likely benign(Last reviewed: Apr 25, 2018)	VCV000741177
NM_001282531.3(ADNP):c.921C>G (p.Leu307=)	Likely benign(Last reviewed: Aug 1, 2021)	VCV001298870
NM_001282531.3(ADNP):c.2265C>T (p.Asp755=)	Likely benign(Last reviewed: Aug 15, 2018)	VCV000764807
NM_001282531.3(ADNP):c.549C>T (p.His183=)	Likely benign(Last reviewed: Aug 17, 2018)	VCV000764587
NM_001282531.3(ADNP):c.782T>C (p.Val261Ala)	Likely benign(Last reviewed: Aug 18, 2020)	VCV001337659
NM_001282531.3(ADNP):c.609A>G (p.Lys203=)	Likely benign(Last reviewed: Aug 20, 2018)	VCV000762411

NM_001282531.3(ADNP):c.1754A>G (p.Asn585Ser)	Likely benign(Last reviewed: Dec 17, 2018)	VCV000589686
NM_001282531.3(ADNP):c.1234C>G (p.Leu412Val)	Likely benign(Last reviewed: Dec 18, 2020)	VCV001191415
NM_001282531.3(ADNP):c.1932A>G (p.Arg644=)	Likely benign(Last reviewed: Dec 2, 2017)	VCV000587993
NM_001282531.3(ADNP):c.1635T>C (p.Asp545=)	Likely benign(Last reviewed: Dec 31, 2019)	VCV000797708
NM_001282531.3(ADNP):c.723C>T (p.Ile241=)	Likely benign(Last reviewed: Dec 31, 2019)	VCV000743463
NM_001282531.3(ADNP):c.1428A>G (p.Ala476=)	Likely benign(Last reviewed: Dec 31, 2019)	VCV000729217
NM_001282531.3(ADNP):c.2715C>T (p.Asn905=)	Likely benign(Last reviewed: Dec 31, 2019)	VCV000714188
NM_001282531.3(ADNP):c.2725G>A (p.Glu909Lys)	Likely benign(Last reviewed: Dec 4, 2020)	VCV001193365
NM_001282531.3(ADNP):c.1127G>A (p.Arg376Lys)	Likely benign(Last reviewed: Feb 24, 2021)	VCV001193486
NM_001282531.3(ADNP):c.402C>T (p.Ser134=)	Likely benign(Last reviewed: Feb 27, 2018)	VCV000735278
NM_001282531.3(ADNP):c.1747G>T (p.Ala583Ser)	Likely benign(Last reviewed: Feb 3, 2021)	VCV001195748
NM_001282531.3(ADNP):c.1592T>C (p.Met531Thr)	Likely benign(Last reviewed: Jan 1, 2019)	VCV000975320

NM_001282531.3(ADNP):c.1123G>A (p.Gly375Arg)	Likely benign(Last reviewed: Jan 1, 2019)	VCV000975319
NM_001282531.3(ADNP):c.886C>T (p.Arg296Trp)	Likely benign(Last reviewed: Jan 1, 2019)	VCV000975318
NM_001282531.3(ADNP):c.627C>A (p.Val209=)	Likely benign(Last reviewed: Jan 11, 2017)	VCV000589486
NM_001282531.3(ADNP):c.142T>C (p.Leu48=)	Likely benign(Last reviewed: Jan 12, 2018)	VCV000732811
NM_001282531.3(ADNP):c.1847A>G (p.Lys616Arg)	Likely benign(Last reviewed: Jan 12, 2022)	VCV001342109
NM_001282531.3(ADNP):c.2157C>T (p.Tyr719=)	Likely benign(Last reviewed: Jan 2, 2019)	VCV000799506
NM_001282531.3(ADNP):c.855A>G (p.Pro285=)	Likely benign(Last reviewed: Jan 24, 2018)	VCV000722164
NM_001282531.3(ADNP):c.1014C>G (p.Gly338=)	Likely benign(Last reviewed: Jan 27, 2017)	VCV000589532
NM_001282531.3(ADNP):c.-5-232C>T	Likely benign(Last reviewed: Jan 28, 2019)	VCV001205472
NM_001282531.3(ADNP):c.2561C>T (p.Ser854Phe)	Likely benign(Last reviewed: Jan 4, 2021)	VCV001200985
NM_001282531.3(ADNP):c.1062C>T (p.Asn354=)	Likely benign(Last reviewed: Jul 18, 2018)	VCV000741518
NM_001282531.3(ADNP):c.2406C>A (p.Ser802=)	Likely benign(Last reviewed: Jul 2, 2018)	VCV000713746

NM_001282531.3(ADNP):c.1683C>T (p.Asn561=)	Likely benign(Last reviewed: Jul 31, 2018)	VCV000761244
NM_001282531.3(ADNP):c.2535A>C (p.Leu845=)	Likely benign(Last reviewed: Jul 6, 2018)	VCV000756768
NM_001282531.3(ADNP):c.1773A>C (p.Pro591=)	Likely benign(Last reviewed: Jun 1, 2017)	VCV000589109
NM_001282531.3(ADNP):c.1704A>G (p.Thr568=)	Likely benign(Last reviewed: Jun 11, 2021)	VCV001327804
NM_001282531.3(ADNP):c.666C>T (p.His222=)	Likely benign(Last reviewed: Jun 22, 2017)	VCV000434091
NM_001282531.3(ADNP):c.3147T>C (p.Asn1049=)	Likely benign(Last reviewed: Jun 8, 2018)	VCV000749518
NM_001282531.3(ADNP):c.2189G>A (p.Arg730Gln)	Likely benign(Last reviewed: Mar 1, 2021)	VCV001300622
NM_001282531.3(ADNP):c.1893A>G (p.Leu631=)	Likely benign(Last reviewed: Mar 29, 2018)	VCV000737800
NM_001282531.3(ADNP):c.1141G>C (p.Gly381Arg)	Likely benign(Last reviewed: Mar 31, 2017)	VCV000389878
NM_001282531.3(ADNP):c.2856A>C (p.Ala952=)	Likely benign(Last reviewed: May 13, 2018)	VCV000789499
NM_001282531.3(ADNP):c.483G>A (p.Glu161=)	Likely benign(Last reviewed: May 17, 2018)	VCV000589614
NM_001282531.3(ADNP):c.2808C>T (p.Tyr936=)	Likely benign(Last reviewed: May 20, 2018)	VCV000745344

NM_001282531.3(ADNP):c.191C>T (p.Thr64Met)	Likely benign(Last reviewed: May 22, 2021)	VCV001194988
NM_001282531.3(ADNP):c.957C>G (p.Val319=)	Likely benign(Last reviewed: May 4, 2018)	VCV000742535
NM_001282531.3(ADNP):c.1392C>T (p.His464=)	Likely benign(Last reviewed: Nov 13, 2017)	VCV000725817
NM_001282531.3(ADNP):c.2994C>T (p.Asp998=)	Likely benign(Last reviewed: Nov 13, 2020)	VCV001188179
NM_001282531.3(ADNP):c.2574T>C (p.Ala858=)	Likely benign(Last reviewed: Nov 24, 2017)	VCV000728627
NM_001282531.3(ADNP):c.2849A>G (p.His950Arg)	Likely benign(Last reviewed: Nov 29, 2018)	VCV001336819
NM_001282531.3(ADNP):c.356A>G (p.Lys119Arg)	Likely benign(Last reviewed: Oct 20, 2016)	VCV000588925
NM_001282531.3(ADNP):c.285T>C (p.Asn95=)	Likely benign(Last reviewed: Oct 24, 2016)	VCV000588368
NM_001282531.3(ADNP):c.1275T>C (p.Ser425=)	Likely benign(Last reviewed: Oct 9, 2017)	VCV000589712
NM_001282531.3(ADNP):c.2570A>G (p.Asn857Ser)	Likely benign(Last reviewed: Sep 10, 2019)	VCV001210870
NM_001282531.3(ADNP):c.2466G>A (p.Val822=)	Likely benign(Last reviewed: Sep 13, 2017)	VCV001336233
NM_001282531.3(ADNP):c.1584C>T (p.Ala528=)	Likely benign(Last reviewed: Sep 18, 2018)	VCV000751571

NM_001282531.3(ADNP):c.1597A>G (p.Met533Val)	Likely benign(Last reviewed: Sep 27, 2017)	VCV000589403
NM_001282531.3(ADNP):c.1896A>G (p.Lys632=)	Likely benign(Last reviewed: Sep 30, 2018)	VCV000624198
NM_001282531.3(ADNP):c.673C>T (p.Arg225Ter)	Likely pathogenic	VCV000374229
NM_001282531.3(ADNP):c.2250_2274del (p.Val751fs)	Likely pathogenic	VCV000996677
NM_001282531.3(ADNP):c.1717del (p.Asp573fs)	Likely pathogenic(Last reviewed: Apr 22, 2016)	VCV000438275
NM_001282531.3(ADNP):c.2712dup (p.Asn905Ter)	Likely pathogenic(Last reviewed: Apr 26, 2021)	VCV001098391
NM_001282531.3(ADNP):c.1754dup (p.Asn585fs)	Likely pathogenic(Last reviewed: Aug 1, 2016)	VCV000421870
GRCh37/hg19 20q13.13-13.2(chr20:47726521-50427649)x1	Likely pathogenic(Last reviewed: Aug 10, 2015)	VCV000442502
NM_001282531.3(ADNP):c.106dup (p.Glu36fs)	Likely pathogenic(Last reviewed: Aug 17, 2016)	VCV000421945
NM_001282531.3(ADNP):c.709del (p.Val237fs)	Likely pathogenic(Last reviewed: Aug 30, 2017)	VCV000976124
NM_001282531.3(ADNP):c.1046_1047del (p.Leu349fs)	Likely pathogenic(Last reviewed: Dec 23, 2015)	VCV000984840
NM_001282531.3(ADNP):c.1179_1180del (p.Leu394fs)	Likely pathogenic(Last reviewed: Dec 6, 2017)	VCV000503933

NM_001282531.3(ADNP):c.2187dup (p.Arg730fs)	Likely pathogenic(Last reviewed: Feb 2, 2018)	VCV000504292
NM_001282531.3(ADNP):c.1265dup (p.Gln423fs)	Likely pathogenic(Last reviewed: Feb 5, 2020)	VCV000828170
NM_001282531.3(ADNP):c.2938C>T (p.Gln980Ter)	Likely pathogenic(Last reviewed: Jan 1, 2019)	VCV000982697
NM_001282531.3(ADNP):c.321del (p.Asn108fs)	Likely pathogenic(Last reviewed: Jan 3, 2022)	VCV001333276
NM_001282531.3(ADNP):c.3307dup (p.Ter1103LeuextTer?)	Likely pathogenic(Last reviewed: Jan 31, 2017)	VCV000545057
NM_001282531.3(ADNP):c.3069_3072del (p.Arg1023fs)	Likely pathogenic(Last reviewed: Jun 13, 2016)	VCV000988372
NM_001282531.3(ADNP):c.1310dup (p.Gly438fs)	Likely pathogenic(Last reviewed: Jun 29, 2021)	VCV001297054
GRCh38/hg38 20q13.13-13.2(chr20:49989123-51495645)x1	Likely pathogenic(Last reviewed: Mar 21, 2011)	VCV000148193
GRCh37/hg19 20q13.13-13.2(chr20:47682662-49884981)x1	Likely pathogenic(Last reviewed: Mar 25, 2014)	VCV000187826
GRCh37/hg19 20q13.13(chr20:49447090-49510400)x1	Likely pathogenic(Last reviewed: Mar 8, 2018)	VCV000564608
NM_001282531.3(ADNP):c.2946dup (p.Asp983fs)	Likely pathogenic(Last reviewed: Oct 23, 2020)	VCV000987062
NM_001282531.3(ADNP):c.2129dup (p.Ser711fs)	Likely pathogenic(Last reviewed: Oct 24, 2016)	VCV000984843

NM_001282531.3(ADNP):c.201G>C (p.Gln67His)	Likely pathogenic(Last reviewed: Oct 4, 2017)	VCV000521542
NM_001282531.3(ADNP):c.2157del (p.Thr718_Tyr719insTer)	Likely pathogenic(Last reviewed: Sep 26, 2019)	VCV000800950
NM_001282531.3(ADNP):c.1540T>G (p.Cys514Gly)	Likely pathogenic(Last reviewed: Sep 27, 2021)	VCV001319913
NM_001282531.3(ADNP):c.82_85del (p.Leu28fs)	not provided	VCV000973033
NM_001282531.3(ADNP):c.642_651del (p.Asn214fs)	Pathogenic	VCV000981627
NM_001282531.3(ADNP):c.2808del (p.Lys935_Tyr936insTer)	Pathogenic(Last reviewed: Apr 1, 2014)	VCV000139634
NM_001282531.3(ADNP):c.1211C>A (p.Ser404Ter)	Pathogenic(Last reviewed: Apr 1, 2014)	VCV000139633
NM_001282531.3(ADNP):c.733G>T (p.Glu245Ter)	Pathogenic(Last reviewed: Apr 16, 2018)	VCV000619999
NM_001282531.3(ADNP):c.1620_1630dup (p.Thr544fs)	Pathogenic(Last reviewed: Apr 18, 2016)	VCV000986226
NM_001282531.3(ADNP):c.790C>T (p.Arg264Ter)	Pathogenic(Last reviewed: Apr 27, 2016)	VCV000280535
NM_001282531.3(ADNP):c.2498_2499del (p.Lys833fs)	Pathogenic(Last reviewed: Apr 27, 2018)	VCV000985962
NM_001282531.3(ADNP):c.655_656del (p.Glu218_Ser219insTer)	Pathogenic(Last reviewed: Aug 10, 2021)	VCV001299548
GRCh38/hg38 20q13.13-13.2(chr20:49731076-51202566)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000058973
GRCh38/hg38 20q13.13-13.2(chr20:49947237-55875406)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059219

GRCh38/hg38 20q13.12-13.33(chr20:44787704-64277321)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059218
NM_001282531.3(ADNP):c.1287dup (p.Ala430fs)	Pathogenic(Last reviewed: Aug 14, 2017)	VCV000451210
NM_001282531.3(ADNP):c.2239G>T (p.Glu747Ter)	Pathogenic(Last reviewed: Aug 15, 2017)	VCV000489048
NM_001282531.3(ADNP):c.2213C>A (p.Ser738Ter)	Pathogenic(Last reviewed: Aug 19, 2021)	VCV000984838
NM_001282531.3(ADNP):c.2213C>G (p.Ser738Ter)	Pathogenic(Last reviewed: Aug 2, 2017)	VCV000391218
NM_001282531.3(ADNP):c.2287dup (p.Ser763fs)	Pathogenic(Last reviewed: Aug 25, 2017)	VCV000522015
NM_001282531.3(ADNP):c.190dup (p.Thr64fs)	Pathogenic(Last reviewed: Aug 30, 2021)	VCV000280557
NM_001282531.3(ADNP):c.1239_1240del (p.Gln414fs)	Pathogenic(Last reviewed: Dec 1, 2016)	VCV000973124
NM_001282531.3(ADNP):c.1184dup (p.Ser396fs)	Pathogenic(Last reviewed: Dec 1, 2017)	VCV000973125
NM_001282531.3(ADNP):c.67_70del (p.Leu23fs)	Pathogenic(Last reviewed: Dec 1, 2018)	VCV000973123
GRCh37/hg19 20p13-q13.33(chr20:61569-62915555)x3	Pathogenic(Last reviewed: Dec 2, 2014)	VCV000443340
NM_001282531.3(ADNP):c.2318dup (p.Tyr774fs)	Pathogenic(Last reviewed: Dec 8, 2016)	VCV000374212
NM_001282531.3(ADNP):c.712C>T (p.Gln238Ter)	Pathogenic(Last reviewed: Feb 1, 2018)	VCV000872736
NM_001282531.3(ADNP):c.2157C>A (p.Tyr719Ter)	Pathogenic(Last reviewed: Feb 1, 2018)	VCV000280623

NM_001282531.3(ADNP):c.845del (p.Gly282fs)	Pathogenic(Last reviewed: Feb 3, 2017)	VCV000423279
NM_001282531.3(ADNP):c.940_941del (p.Leu314fs)	Pathogenic(Last reviewed: Feb 4, 2019)	VCV000984839
NM_001282531.3(ADNP):c.331del (p.Tyr111fs)	Pathogenic(Last reviewed: Feb 6, 2019)	VCV000985605
NM_001282531.3(ADNP):c.916C>T (p.Arg306Ter)	Pathogenic(Last reviewed: Feb 8, 2019)	VCV000817538
GRCh37/hg19 20p13-q13.33(chr20:63244-62961294)x3	Pathogenic(Last reviewed: Jan 1, 2013)	VCV000604423
NM_001282531.3(ADNP):c.280C>T (p.Arg94Cys)	Pathogenic(Last reviewed: Jan 1, 2019)	VCV000975321
NM_001282531.3(ADNP):c.898dup (p.Ser300fs)	Pathogenic(Last reviewed: Jan 1, 2020)	VCV001174075
NM_001282531.3(ADNP):c.539_542del (p.Val180fs)	Pathogenic(Last reviewed: Jan 1, 2021)	VCV000373314
NM_001282531.3(ADNP):c.56_57del (p.Val19fs)	Pathogenic(Last reviewed: Jan 14, 2022)	VCV000817018
NM_001282531.3(ADNP):c.1033C>T (p.Gln345Ter)	Pathogenic(Last reviewed: Jan 15, 2016)	VCV000984841
NM_001282531.3(ADNP):c.2157C>G (p.Tyr719Ter)	Pathogenic(Last reviewed: Jan 3, 2022)	VCV000139635
NM_001282531.3(ADNP):c.819del (p.Lys274fs)	Pathogenic(Last reviewed: Jan 4, 2017)	VCV000280262
NM_001282531.3(ADNP):c.1235del (p.Leu412fs)	Pathogenic(Last reviewed: Jan 5, 2016)	VCV000280199
GRCh37/hg19 20p13-q13.33(chr20:63244-62948788)x3	Pathogenic(Last reviewed: Jan 5, 2017)	VCV000604422

GRCh37/hg19 20p13-q13.33(chr20:61569-62915555)	Pathogenic(Last reviewed: Jul 14, 2015)	VCV000443339
NM_001282531.3(ADNP):c.2496_2499del (p.Asn832fs)	Pathogenic(Last reviewed: Jul 21, 2021)	VCV000139632
NM_001282531.3(ADNP):c.2287del (p.Ser763fs)	Pathogenic(Last reviewed: Jul 5, 2016)	VCV000521175
NM_001282531.3(ADNP):c.70del (p.Ser24fs)	Pathogenic(Last reviewed: Jun 15, 2015)	VCV000419147
NM_001282531.3(ADNP):c.1807_1808dup (p.Pro604fs)	Pathogenic(Last reviewed: Jun 17, 2021)	VCV001184924
NM_001282531.3(ADNP):c.2175del (p.Leu726fs)	Pathogenic(Last reviewed: Jun 21, 2017)	VCV000432980
NM_001282531.3(ADNP):c.337_340del (p.Thr113fs)	Pathogenic(Last reviewed: Jun 26, 2017)	VCV000450110
NM_001282531.3(ADNP):c.69dup (p.Ser24Ter)	Pathogenic(Last reviewed: Jun 26, 2017)	VCV001335928
NM_001282531.3(ADNP):c.1106_1108delinsCTGT (p.Leu369fs)	Pathogenic(Last reviewed: Jun 6, 2016)	VCV000265590
NM_001282531.3(ADNP):c.1222_1223del (p.Lys408fs)	Pathogenic(Last reviewed: Mar 12, 2015)	VCV000190279
NM_001282531.3(ADNP):c.1936_1937del (p.Arg646fs)	Pathogenic(Last reviewed: Mar 18, 2016)	VCV000619994
GRCh38/hg38 20q13.13-13.2(chr20:50781990-52792847)x1	Pathogenic(Last reviewed: Mar 21, 2011)	VCV000148190
NM_001282531.3(ADNP):c.2865_2868del (p.Ser955fs)	Pathogenic(Last reviewed: Mar 23, 2016)	VCV000984844
NM_001282531.3(ADNP):c.2230G>T (p.Glu744Ter)	Pathogenic(Last reviewed: Mar 27, 2020)	VCV000489245

NM_001282531.3(ADNP):c.517C>T (p.Arg173Ter)	Pathogenic(Last reviewed: Mar 27, 2020)	VCV000431117
NM_001282531.3(ADNP):c.177_178dup (p.Asp60fs)	Pathogenic(Last reviewed: Mar 9, 2017)	VCV000620048
NM_001282531.3(ADNP):c.2195del (p.Lys731_Leu732insTer)	Pathogenic(Last reviewed: May 18, 2017)	VCV000430071
NM_001282531.3(ADNP):c.57dup (p.Lys20fs)	Pathogenic(Last reviewed: May 2, 2017)	VCV000429255
NM_001282531.3(ADNP):c.1337G>A (p.Trp446Ter)	Pathogenic(Last reviewed: May 2, 2017)	VCV000521765
NM_001282531.3(ADNP):c.-5-1G>C	Pathogenic(Last reviewed: May 28, 2019)	VCV000803615
NM_001282531.3(ADNP):c.2157_2159del (p.Tyr719_Glu720delinsTer)	Pathogenic(Last reviewed: May 28, 2019)	VCV000803614
NM_001282531.3(ADNP):c.2499del (p.Val834fs)	Pathogenic(Last reviewed: May 9, 2016)	VCV000520984
NM_001282531.3(ADNP):c.1132dup (p.Tyr378fs)	Pathogenic(Last reviewed: Nov 1, 2019)	VCV000872735
GRCh37/hg19 20q13.13-13.2(chr20:47627844-52045480)x1	Pathogenic(Last reviewed: Nov 12, 2013)	VCV000443773
NM_001282531.3(ADNP):c.1876_1892del (p.Leu626fs)	Pathogenic(Last reviewed: Nov 27, 2017)	VCV000503886
NM_001282531.3(ADNP):c.2454C>G (p.Tyr818Ter)	Pathogenic(Last reviewed: Nov 4, 2021)	VCV001334645
NM_001282531.3(ADNP):c.2491_2494del (p.Leu831fs)	Pathogenic(Last reviewed: Nov 5, 2021)	VCV000139631
GRCh38/hg38 20p13-q13.33(chr20:99557-64277321)x3	Pathogenic(Last reviewed: Oct 19, 2010)	VCV000146596

NM_001282531.3(ADNP):c.1652_1653del (p.Asp551fs)	Pathogenic(Last reviewed: Oct 23, 2020)	VCV000987470
NM_001282531.3(ADNP):c.1666C>T (p.Gln556Ter)	Pathogenic(Last reviewed: Oct 23, 2020)	VCV000987411
NM_001282531.3(ADNP):c.2327dup (p.Asn776fs)	Pathogenic(Last reviewed: Oct 23, 2020)	VCV000987168
NM_001282531.3(ADNP):c.2187del (p.Lys729fs)	Pathogenic(Last reviewed: Oct 23, 2020)	VCV000987030
NM_001282531.3(ADNP):c.2194_2197del (p.Leu732fs)	Pathogenic(Last reviewed: Oct 3, 2017)	VCV000452597
GRCh37/hg19 20p13-q13.33(chr20:63244-62912463)x3	Pathogenic(Last reviewed: Oct 31, 2014)	VCV000604421
NM_001282531.3(ADNP):c.484C>T (p.Gln162Ter)	Pathogenic(Last reviewed: Sep 15, 2016)	VCV000984842
NM_001282531.3(ADNP):c.330dup (p.Tyr111fs)	Pathogenic(Last reviewed: Sep 16, 2016)	VCV000280876
NM_001282531.3(ADNP):c.2387G>A (p.Trp796Ter)	Pathogenic(Last reviewed: Sep 23, 2021)	VCV001338860
NM_001282531.3(ADNP):c.2268dup (p.Lys757fs)	Pathogenic(Last reviewed: Sep 30, 2016)	VCV000280907
NM_001282531.3(ADNP):c.2378T>G (p.Leu793Ter)	Pathogenic(Last reviewed: Sep 30, 2016)	VCV000280874
NM_001282531.3(ADNP):c.64dup (p.Ile22fs)	Pathogenic/Likely pathogenic(Last reviewed: Apr 26, 2021)	VCV000931423
NM_001282531.3(ADNP):c.2188C>T (p.Arg730Ter)	Pathogenic/Likely pathogenic(Last reviewed: Jun 17, 2021)	VCV000279598
NM_001282531.3(ADNP):c.2156dup (p.Tyr719Ter)	Pathogenic/Likely pathogenic(Last reviewed: Oct 23, 2020)	VCV000190278

NM_001282531.3(ADNP):c.583C>T (p.Pro195Ser)	Uncertain significance	VCV000813652
NM_001282531.3(ADNP):c.1931G>A (p.Arg644Gln)	Uncertain significance	VCV000813555
NM_001282531.3(ADNP):c.1721C>T (p.Ala574Val)	Uncertain significance(Last reviewed: Apr 1, 2019)	VCV000809260
NM_001282531.3(ADNP):c.283A>G (p.Asn95Asp)	Uncertain significance(Last reviewed: Apr 1, 2020)	VCV000932382
Single allele	Uncertain significance(Last reviewed: Apr 13, 2018)	VCV000560064
NM_001282531.3(ADNP):c.953G>T (p.Gly318Val)	Uncertain significance(Last reviewed: Apr 16, 2019)	VCV001305248
NM_001282531.3(ADNP):c.197A>G (p.Asn66Ser)	Uncertain significance(Last reviewed: Apr 25, 2018)	VCV001336506
NM_001282531.3(ADNP):c.108+6T >C	Uncertain significance(Last reviewed: Apr 26, 2018)	VCV000816876
NM_001282531.3(ADNP):c.2165T>C (p.Met722Thr)	Uncertain significance(Last reviewed: Apr 5, 2017)	VCV000426751
NM_001282531.3(ADNP):c.298G>T (p.Asp100Tyr)	Uncertain significance(Last reviewed: Apr 7, 2021)	VCV001300678
NM_001282531.3(ADNP):c.2669G>A (p.Gly890Asp)	Uncertain significance(Last reviewed: Aug 1, 2019)	VCV001304829
NM_001282531.3(ADNP):c.2083A>G (p.Asn695Asp)	Uncertain significance(Last reviewed: Aug 14, 2019)	VCV001304872

NM_001282531.3(ADNP):c.638C>T (p.Ser213Leu)	Uncertain significance(Last reviewed: Aug 22, 2019)	VCV001027973
NM_001282531.3(ADNP):c.2474G> A (p.Gly825Glu)	Uncertain significance(Last reviewed: Dec 1, 2016)	VCV000973122
NM_001282531.3(ADNP):c.2632A> G (p.Ser878Gly)	Uncertain significance(Last reviewed: Dec 13, 2016)	VCV000589464
NM_001282531.3(ADNP):c.1345T> C (p.Cys449Arg)	Uncertain significance(Last reviewed: Dec 27, 2019)	VCV001311337
NM_001282531.3(ADNP):c.2012A> G (p.Tyr671Cys)	Uncertain significance(Last reviewed: Dec 6, 2019)	VCV001310944
NM_001282531.3(ADNP):c.3302A> G (p.Gln1101Arg)	Uncertain significance(Last reviewed: Feb 18, 2020)	VCV001312755
NM_001282531.3(ADNP):c.2116C> A (p.Leu706Ile)	Uncertain significance(Last reviewed: Jan 1, 2018)	VCV000546901
NM_001282531.3(ADNP):c.391C>G (p.Pro131Ala)	Uncertain significance(Last reviewed: Jan 1, 2020)	VCV001174100
NM_001282531.3(ADNP):c.2572G> A (p.Ala858Thr)	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000976094
NM_001282531.3(ADNP):c.481G>A (p.Glu161Lys)	Uncertain significance(Last reviewed: Jan 2, 2020)	VCV001027972
NM_001282531.3(ADNP):c.872C>G (p.Ala291Gly)	Uncertain significance(Last reviewed: Jan 25, 2018)	VCV001031904
NM_001282531.3(ADNP):c.2143G> A (p.Val715Met)	Uncertain significance(Last reviewed: Jan 5, 2016)	VCV000434094

NM_001282531.3(ADNP):c.985C>G (p.Gln329Glu)	Uncertain significance(Last reviewed: Jan 6, 2016)	VCV000285662
NM_001282531.3(ADNP):c.839G>A (p.Ser280Asn)	Uncertain significance(Last reviewed: Jul 1, 2018)	VCV000624199
NM_001282531.3(ADNP):c.2600A>G (p.Asn867Ser)	Uncertain significance(Last reviewed: Jul 10, 2019)	VCV001337284
NM_001282531.3(ADNP):c.2716_2718del (p.Asp906del)	Uncertain significance(Last reviewed: Jul 15, 2020)	VCV001306884
NM_001282531.3(ADNP):c.3056T>C (p.Met1019Thr)	Uncertain significance(Last reviewed: Jul 27, 2020)	VCV001027971
NM_001282531.3(ADNP):c.1133A>G (p.Tyr378Cys)	Uncertain significance(Last reviewed: Jul 28, 2016)	VCV000588109
NM_001282531.3(ADNP):c.1102C>T (p.Gln368Ter)	Uncertain significance(Last reviewed: Jun 14, 2016)	VCV000338748
NM_001282531.3(ADNP):c.1677C>A (p.His559Gln)	Uncertain significance(Last reviewed: Jun 20, 2016)	VCV000387173
NM_001282531.3(ADNP):c.2059T>C (p.Cys687Arg)	Uncertain significance(Last reviewed: Jun 26, 2017)	VCV001052737
NM_001282531.3(ADNP):c.3040A>C (p.Lys1014Gln)	Uncertain significance(Last reviewed: Jun 29, 2021)	VCV001334498
NM_001282531.3(ADNP):c.669C>T (p.Cys223=)	Uncertain significance(Last reviewed: Jun 5, 2017)	VCV000432769
NM_001282531.3(ADNP):c.650A>G (p.Glu217Gly)	Uncertain significance(Last reviewed: Mar 24, 2021)	VCV001331626

NM_001282531.3(ADNP):c.736C>T (p.Arg246Cys)	Uncertain significance(Last reviewed: Mar 3, 2020)	VCV001251938
NM_001282531.3(ADNP):c.-5-1G>A	Uncertain significance(Last reviewed: Mar 8, 2017)	VCV000423847
NM_001282531.3(ADNP):c.208C>T (p.Arg70Trp)	Uncertain significance(Last reviewed: May 1, 2021)	VCV001176428
NM_001282531.3(ADNP):c.2749C>G (p.Pro917Ala)	Uncertain significance(Last reviewed: May 14, 2020)	VCV001301725
NM_001282531.3(ADNP):c.1855G>T (p.Val619Phe)	Uncertain significance(Last reviewed: May 16, 2017)	VCV000386594
NM_001282531.3(ADNP):c.2918_2932del (p.Glu973_Ser978delinsAla)	Uncertain significance(Last reviewed: May 17, 2018)	VCV000546231
NM_001282531.3(ADNP):c.2056C>G (p.His686Asp)	Uncertain significance(Last reviewed: May 17, 2021)	VCV001303971
NM_001282531.3(ADNP):c.344A>G (p.Asn115Ser)	Uncertain significance(Last reviewed: May 2, 2016)	VCV000589274
NM_001282531.3(ADNP):c.3248dup (p.Val1084fs)	Uncertain significance(Last reviewed: May 21, 2018)	VCV000546371
NM_001282531.3(ADNP):c.775A>C (p.Asn259His)	Uncertain significance(Last reviewed: May 26, 2017)	VCV000589645
NM_001282531.3(ADNP):c.3097T>C (p.Tyr1033His)	Uncertain significance(Last reviewed: May 29, 2019)	VCV001306158
NM_001282531.3(ADNP):c.724G>A (p.Glu242Lys)	Uncertain significance(Last reviewed: May 5, 2017)	VCV000589090

225

NM_001282531.3(ADNP):c.121T>C (p.Phe41Leu)	Uncertain significance(Last reviewed: Nov 14, 2019)	VCV001310439
NM_001282531.3(ADNP):c.1180C>G (p.Leu394Val)	Uncertain significance(Last reviewed: Nov 20, 2019)	VCV001030496
NM_001282531.3(ADNP):c.2189G>C (p.Arg730Pro)	Uncertain significance(Last reviewed: Nov 5, 2019)	VCV001309778
NM_001282531.3(ADNP):c.2971_2 979del (p.Met991_Pro993del)	Uncertain significance(Last reviewed: Nov 5, 2019)	VCV000966928
NM_001282531.3(ADNP):c.3304G>A (p.Ala1102Thr)	Uncertain significance(Last reviewed: Oct 24, 2019)	VCV000931701
NM_001282531.3(ADNP):c.499T>C (p.Cys167Arg)	Uncertain significance(Last reviewed: Oct 27, 2020)	VCV001304715
NM_001282531.3(ADNP):c.2150G>A (p.Arg717His)	Uncertain significance(Last reviewed: Oct 30, 2019)	VCV000958481
NM_001282531.3(ADNP):c.1142G>T (p.Gly381Val)	Uncertain significance(Last reviewed: Sep 13, 2017)	VCV000588833

Appendix Table 4. Table of reported variants for *SMARCA2,* retrieved from ClinVar databases (https://www.ncbi.nlm.nih.gov/clinvar/?term=SMARCA2%5Bgene%5D&redir=gene)

Name	Clinical significance (Last reviewed)	Accession
NM_003070.5(SMARCA2):c.3293-45T>C	Benign(Last reviewed: Apr 13, 2020)	VCV001249804
NM_003070.5(SMARCA2):c.4199+41G>C	Benign(Last reviewed: Apr 17, 2020)	VCV001224862
NM_003070.5(SMARCA2):c.97C>T (p.Pro33Ser)	Benign(Last reviewed: Apr 18, 2019)	VCV000366193
NM_003070.5(SMARCA2):c.4254-18G>C	Benign(Last reviewed: Apr 20, 2021)	VCV001281533

NM_003070.5(SMARCA2):c.3125+ 47T>A	Benign(Last reviewed: Apr 29, 2020)	VCV001289555
NM_003070.5(SMARCA2):c.1080A >G (p.Glu360=)	Benign(Last reviewed: Apr 3, 2020)	VCV001265175
NM_003070.5(SMARCA2):c.231C> T (p.Ile77=)	Benign(Last reviewed: Apr 3, 2020)	VCV000588510
NM_003070.5(SMARCA2):c.2883+ 37T>C	Benign(Last reviewed: Apr 9, 2020)	VCV001221707
NM_003070.5(SMARCA2):c.3684+ 104A>C	Benign(Last reviewed: Aug 10, 2018)	VCV001287712
NM_003070.5(SMARCA2):c.1047- 217C>T	Benign(Last reviewed: Aug 10, 2018)	VCV001278462
NM_003070.5(SMARCA2):c.2883+ 90A>C	Benign(Last reviewed: Aug 10, 2018)	VCV001270974
NM_003070.5(SMARCA2):c.3762+ 194A>C	Benign(Last reviewed: Aug 10, 2018)	VCV001266909
NM_003070.5(SMARCA2):c.2992- 74G>A	Benign(Last reviewed: Aug 10, 2018)	VCV001229665
NM_003070.5(SMARCA2):c.2770- 109A>C	Benign(Last reviewed: Aug 10, 2018)	VCV001228268
NM_003070.5(SMARCA2):c.4738- 316dup	Benign(Last reviewed: Aug 10, 2019)	VCV001243486
NM_003070.5(SMARCA2):c.4359+ 178C>T	Benign(Last reviewed: Aug 11, 2018)	VCV001263334
NM_003070.5(SMARCA2):c.2348+ 311_2348+312dup	Benign(Last reviewed: Aug 11, 2019)	VCV001269133
NM_003070.5(SMARCA2):c.2883+ 269_2883+281dup	Benign(Last reviewed: Aug 11, 2019)	VCV001181773
NM_003070.5(SMARCA2):c.3510G >T (p.Arg1170=)	Benign(Last reviewed: Aug 11, 2021)	VCV001304563

NM_003070.5(SMARCA2):c.3981+22C>T	Benign(Last reviewed: Aug 20, 2020)	VCV001258850
NM_003070.5(SMARCA2):c.1827A>G (p.Pro609=)	Benign(Last reviewed: Aug 21, 2018)	VCV000126345
NM_003070.5(SMARCA2):c.356-220del	Benign(Last reviewed: Aug 21, 2019)	VCV001282930
NM_003070.5(SMARCA2):c.2526+175TG[22]	Benign(Last reviewed: Aug 21, 2019)	VCV001282148
NM_003070.5(SMARCA2):c.3078+88dup	Benign(Last reviewed: Aug 21, 2019)	VCV001233837
NM_003070.5(SMARCA2):c.226-20A>G	Benign(Last reviewed: Aug 22, 2020)	VCV001249904
NM_003070.5(SMARCA2):c.2526+175TG[17]	Benign(Last reviewed: Aug 23, 2019)	VCV001182066
NM_003070.5(SMARCA2):c.3125+43G>T	Benign(Last reviewed: Aug 24, 2020)	VCV001276859
NM_003070.5(SMARCA2):c.3078+89C>A	Benign(Last reviewed: Aug 26, 2019)	VCV001222148
NM_003070.5(SMARCA2):c.4360-14A>T	Benign(Last reviewed: Aug 27, 2021)	VCV001254119
NM_003070.5(SMARCA2):c.3982-40C>T	Benign(Last reviewed: Aug 28, 2020)	VCV001288554
NM_003070.5(SMARCA2):c.695A>C (p.Gln232Pro)	Benign(Last reviewed: Aug 30, 2019)	VCV000366200
NM_003070.5(SMARCA2):c.4594+102_4594+103insT	Benign(Last reviewed: Aug 31, 2018)	VCV001236852
NM_003070.5(SMARCA2):c.4253+53C>T	Benign(Last reviewed: Aug 31, 2018)	VCV001232653
NM_003070.5(SMARCA2):c.791-133del	Benign(Last reviewed: Aug 6, 2019)	VCV001289984

NM_003070.5(SMARCA2):c.3078+82A>T	Benign(Last reviewed: Aug 6, 2019)	VCV001266423
NM_003070.5(SMARCA2):c.791-133dup	Benign(Last reviewed: Aug 6, 2019)	VCV001266002
NM_003070.5(SMARCA2):c.2348+312GT[21]	Benign(Last reviewed: Aug 7, 2019)	VCV001223916
NM_003070.5(SMARCA2):c.1522-231G>A	Benign(Last reviewed: Aug 8, 2018)	VCV001296691
NM_003070.5(SMARCA2):c.1936-55A>T	Benign(Last reviewed: Aug 8, 2018)	VCV001270525
NM_003070.5(SMARCA2):c.3763-80T>C	Benign(Last reviewed: Aug 8, 2018)	VCV001259756
NM_003070.5(SMARCA2):c.2527-25G>A	Benign(Last reviewed: Aug 8, 2018)	VCV001257198
NM_003070.5(SMARCA2):c.-36-45A>G	Benign(Last reviewed: Aug 8, 2018)	VCV001251909
NM_003070.5(SMARCA2):c.-36-230T>G	Benign(Last reviewed: Aug 8, 2018)	VCV001231248
NM_003070.5(SMARCA2):c.1522-48T>C	Benign(Last reviewed: Aug 8, 2018)	VCV001178694
NM_003070.5(SMARCA2):c.177G>A (p.Thr59=)	Benign(Last reviewed: Aug 8, 2018)	VCV000126344
NM_003070.5(SMARCA2):c.4594+16C>T	Benign(Last reviewed: Dec 1, 2020)	VCV001249925
NM_003070.5(SMARCA2):c.4199+47G>C	Benign(Last reviewed: Dec 1, 2020)	VCV001240429
NM_003070.5(SMARCA2):c.1935+35T>G	Benign(Last reviewed: Dec 13, 2020)	VCV001249356
NM_003070.5(SMARCA2):c.4595-41T>A	Benign(Last reviewed: Dec 15, 2020)	VCV001288135

NM_003070.5(SMARCA2):c.669GCA[14] (p.Gln238dup)	Benign(Last reviewed: Dec 18, 2019)	VCV000436797
NM_003070.5(SMARCA2):c.462G>A (p.Gly154=)	Benign(Last reviewed: Dec 23, 2021)	VCV000126348
NM_003070.5(SMARCA2):c.2770-7C>G	Benign(Last reviewed: Dec 31, 2019)	VCV000366213
NM_003070.5(SMARCA2):c.2931G>A (p.Leu977=)	Benign(Last reviewed: Dec 6, 2021)	VCV001327235
NM_003070.5(SMARCA2):c.2349-15T>A	Benign(Last reviewed: Dec 6, 2021)	VCV001327234
NM_003070.5(SMARCA2):c.3292+24C>T	Benign(Last reviewed: Dec 7, 2020)	VCV001243685
NM_003070.5(SMARCA2):c.1266T>A (p.Thr422=)	Benign(Last reviewed: Dec 7, 2020)	VCV001227377
NM_003070.5(SMARCA2):c.4594+80_4594+98del	Benign(Last reviewed: Dec 9, 2019)	VCV001279806
NM_003070.5(SMARCA2):c.4247G>C (p.Gly1416Ala)	Benign(Last reviewed: Jan 12, 2018)	VCV000914998
NM_003070.5(SMARCA2):c.517C>T (p.Pro173Ser)	Benign(Last reviewed: Jan 12, 2018)	VCV000914908
NM_003070.5(SMARCA2):c.*716A>G	Benign(Last reviewed: Jan 12, 2018)	VCV000366342
NM_003070.5(SMARCA2):c.*698T>C	Benign(Last reviewed: Jan 12, 2018)	VCV000366341
NM_003070.5(SMARCA2):c.*620T>C	Benign(Last reviewed: Jan 12, 2018)	VCV000366339
NM_003070.5(SMARCA2):c.*324G>C	Benign(Last reviewed: Jan 12, 2018)	VCV000366335
NM_003070.5(SMARCA2):c.3939C>T (p.Asp1313=)	Benign(Last reviewed: Jan 12, 2018)	VCV000366222

NM_003070.5(SMARCA2):c.3685-6C>A	Benign(Last reviewed: Jan 12, 2018)	VCV000366221
NM_003070.5(SMARCA2):c.2991+10G>A	Benign(Last reviewed: Jan 12, 2018)	VCV000366216
NM_003070.5(SMARCA2):c.1675A>C (p.Arg559=)	Benign(Last reviewed: Jan 13, 2018)	VCV000914438
NM_003070.5(SMARCA2):c.734A>T (p.Gln245Leu)	Benign(Last reviewed: Jan 13, 2018)	VCV000912952
NM_003070.5(SMARCA2):c.*726T>C	Benign(Last reviewed: Jan 13, 2018)	VCV000366343
NM_003070.5(SMARCA2):c.*431G>A	Benign(Last reviewed: Jan 13, 2018)	VCV000366337
NM_003070.5(SMARCA2):c.*181C>T	Benign(Last reviewed: Jan 13, 2018)	VCV000366333
NM_003070.5(SMARCA2):c.*138C>T	Benign(Last reviewed: Jan 13, 2018)	VCV000366332
NM_003070.5(SMARCA2):c.4679G>A (p.Arg1560Gln)	Benign(Last reviewed: Jan 13, 2018)	VCV000366321
NM_003070.5(SMARCA2):c.4584A>G (p.Ser1528=)	Benign(Last reviewed: Jan 13, 2018)	VCV000366254
NM_003070.5(SMARCA2):c.4516A>T (p.Ile1506Phe)	Benign(Last reviewed: Jan 13, 2018)	VCV000366251
NM_003070.5(SMARCA2):c.2907C>T (p.Asp969=)	Benign(Last reviewed: Jan 13, 2018)	VCV000366214
NM_003070.5(SMARCA2):c.1877+9T>A	Benign(Last reviewed: Jan 13, 2018)	VCV000366209
NM_003070.5(SMARCA2):c.716C>T (p.Pro239Leu)	Benign(Last reviewed: Jan 23, 2021)	VCV001249161
NM_003070.5(SMARCA2):c.4737+18G>A	Benign(Last reviewed: Jan 26, 2021)	VCV001269049

NM_003070.5(SMARCA2):c.4737+13G>A	Benign(Last reviewed: Jan 27, 2021)	VCV000912462
NM_003070.5(SMARCA2):c.708A>G (p.Gln236=)	Benign(Last reviewed: Jan 27, 2021)	VCV000366202
NM_003070.5(SMARCA2):c.483G>A (p.Pro161=)	Benign(Last reviewed: Jan 9, 2021)	VCV001278061
NM_003070.5(SMARCA2):c.4360-79A>G	Benign(Last reviewed: Jul 14, 2021)	VCV001185431
NM_003070.5(SMARCA2):c.1347+16_1347+17insT	Benign(Last reviewed: Jul 14, 2021)	VCV001185430
NM_003070.5(SMARCA2):c.4253+40G>C	Benign(Last reviewed: Jul 14, 2021)	VCV000802456
NM_003070.5(SMARCA2):c.3079-25T>A	Benign(Last reviewed: Jul 15, 2021)	VCV001271266
NM_003070.5(SMARCA2):c.4199+46C>T	Benign(Last reviewed: Jul 15, 2021)	VCV001226808
NM_003070.5(SMARCA2):c.4199+183T>C	Benign(Last reviewed: Jul 26, 2018)	VCV001288648
NM_003070.5(SMARCA2):c.3456+164A>G	Benign(Last reviewed: Jul 26, 2018)	VCV001287511
NM_003070.5(SMARCA2):c.2416-331A>G	Benign(Last reviewed: Jul 26, 2018)	VCV001286901
NM_003070.5(SMARCA2):c.2185-224A>T	Benign(Last reviewed: Jul 26, 2018)	VCV001283518
NM_003070.5(SMARCA2):c.790+73C>T	Benign(Last reviewed: Jul 26, 2018)	VCV001276752
NM_003070.5(SMARCA2):c.1347+191A>C	Benign(Last reviewed: Jul 26, 2018)	VCV001271351
NM_003070.5(SMARCA2):c.3292+256T>G	Benign(Last reviewed: Jul 26, 2018)	VCV001248990

NM_003070.5(SMARCA2):c.3079-136G>A	Benign(Last reviewed: Jul 26, 2018)	VCV001242782
NM_003070.5(SMARCA2):c.1747-50G>C	Benign(Last reviewed: Jul 26, 2018)	VCV001233896
NM_003070.5(SMARCA2):c.3763-196A>G	Benign(Last reviewed: Jul 27, 2018)	VCV001296611
NM_003070.5(SMARCA2):c.2184+112A>G	Benign(Last reviewed: Jul 27, 2018)	VCV001283777
NM_003070.5(SMARCA2):c.3763-188G>A	Benign(Last reviewed: Jul 27, 2018)	VCV001277473
NM_003070.5(SMARCA2):c.1047-188C>T	Benign(Last reviewed: Jul 27, 2018)	VCV001272405
NM_003070.5(SMARCA2):c.3079-248G>A	Benign(Last reviewed: Jul 27, 2018)	VCV001272260
NM_003070.5(SMARCA2):c.3763-284G>A	Benign(Last reviewed: Jul 27, 2018)	VCV001243769
NM_003070.5(SMARCA2):c.1348-158T>C	Benign(Last reviewed: Jul 3, 2018)	VCV001296716
NM_003070.5(SMARCA2):c.1935+216C>T	Benign(Last reviewed: Jul 3, 2018)	VCV001296708
NM_003070.5(SMARCA2):c.1174-109G>A	Benign(Last reviewed: Jul 3, 2018)	VCV001296650
NM_003070.5(SMARCA2):c.3078+114C>T	Benign(Last reviewed: Jul 3, 2018)	VCV001289266
NM_003070.5(SMARCA2):c.3981+203C>T	Benign(Last reviewed: Jul 3, 2018)	VCV001288601
NM_003070.5(SMARCA2):c.2184+99T>C	Benign(Last reviewed: Jul 3, 2018)	VCV001286658
NM_003070.5(SMARCA2):c.3762+73C>G	Benign(Last reviewed: Jul 3, 2018)	VCV001281323

NM_003070.5(SMARCA2):c.3126-225A>C	Benign(Last reviewed: Jul 3, 2018)	VCV001278660
NM_003070.5(SMARCA2):c.4595-20T>C	Benign(Last reviewed: Jul 3, 2018)	VCV001278359
NM_003070.5(SMARCA2):c.790+83A>C	Benign(Last reviewed: Jul 3, 2018)	VCV001270376
NM_003070.5(SMARCA2):c.1935+85G>T	Benign(Last reviewed: Jul 3, 2018)	VCV001258233
NM_003070.5(SMARCA2):c.3126-53T>C	Benign(Last reviewed: Jul 3, 2018)	VCV001257336
NM_003070.5(SMARCA2):c.4737+317C>T	Benign(Last reviewed: Jul 3, 2018)	VCV001249269
NM_003070.5(SMARCA2):c.4461+48G>A	Benign(Last reviewed: Jul 3, 2018)	VCV001248205
NM_003070.5(SMARCA2):c.355+101A>C	Benign(Last reviewed: Jul 3, 2018)	VCV001244769
NM_003070.5(SMARCA2):c.4595-285T>G	Benign(Last reviewed: Jul 3, 2018)	VCV001240618
NM_003070.5(SMARCA2):c.1878-52A>T	Benign(Last reviewed: Jul 3, 2018)	VCV001236919
NM_003070.5(SMARCA2):c.4595-77G>T	Benign(Last reviewed: Jul 3, 2018)	VCV001236353
NM_003070.5(SMARCA2):c.226-107G>A	Benign(Last reviewed: Jul 3, 2018)	VCV001235432
NM_003070.5(SMARCA2):c.225+207C>A	Benign(Last reviewed: Jul 3, 2018)	VCV001232226
NM_003070.5(SMARCA2):c.356-253A>G	Benign(Last reviewed: Jul 3, 2018)	VCV001228940
NM_003070.5(SMARCA2):c.4738-299A>G	Benign(Last reviewed: Jul 3, 2018)	VCV001183914

NM_003070.5(SMARCA2):c.4594+232T>C	Benign(Last reviewed: Jul 3, 2018)	VCV001182132
NM_003070.5(SMARCA2):c.4461+195T>C	Benign(Last reviewed: Jul 3, 2018)	VCV001177711
NM_003070.5(SMARCA2):c.4638C>G (p.Asp1546Glu)	Benign(Last reviewed: Jul 3, 2018)	VCV000126349
NM_003070.5(SMARCA2):c.-5G>A	Benign(Last reviewed: Jul 3, 2018)	VCV000126341
NM_003070.5(SMARCA2):c.2349-10del	Benign(Last reviewed: Jul 30, 2020)	VCV001277235
NM_003070.5(SMARCA2):c.3672G>A (p.Glu1224=)	Benign(Last reviewed: Jul 31, 2018)	VCV000126346
NM_003070.5(SMARCA2):c.683A>C (p.Gln228Pro)	Benign(Last reviewed: Jul 31, 2018)	VCV000126351
NM_003070.5(SMARCA2):c.3762+47C>T	Benign(Last reviewed: Jun 16, 2020)	VCV001250497
NM_003070.5(SMARCA2):c.3292+25G>A	Benign(Last reviewed: Jun 16, 2021)	VCV001247000
GRCh37/hg19 9p24.3(chr9:2143543-2151371)x0	Benign(Last reviewed: Jun 21, 2012)	VCV000611461
NM_003070.5(SMARCA2):c.4218G>A (p.Val1406=)	Benign(Last reviewed: Jun 24, 2019)	VCV001337262
NM_003070.5(SMARCA2):c.225+28C>G	Benign(Last reviewed: Jun 25, 2020)	VCV001273778
NM_003070.5(SMARCA2):c.2185-44C>G	Benign(Last reviewed: Jun 29, 2020)	VCV001236102
NM_003070.5(SMARCA2):c.1806C>T (p.Thr602=)	Benign(Last reviewed: Jun 30, 2021)	VCV000366208
NM_003070.5(SMARCA2):c.689A>C (p.Gln230Pro)	Benign(Last reviewed: Mar 1, 2021)	VCV000912949

NM_003070.5(SMARCA2):c.1422G >A (p.Gln474=)	Benign(Last reviewed: Mar 14, 2019)	VCV000366205
NM_003070.5(SMARCA2):c.4595-33C>T	Benign(Last reviewed: Mar 19, 2020)	VCV001229573
NM_003070.5(SMARCA2):c.4736G >A (p.Arg1579His)	Benign(Last reviewed: Mar 20, 2018)	VCV000912460
NM_003070.5(SMARCA2):c.4595-7G>C	Benign(Last reviewed: Mar 20, 2020)	VCV000366320
NM_003070.5(SMARCA2):c.2349-22G>T	Benign(Last reviewed: Mar 21, 2020)	VCV001243045
NM_003070.5(SMARCA2):c.1510C >A (p.Arg504=)	Benign(Last reviewed: Mar 22, 2021)	VCV001294473
NM_003070.5(SMARCA2):c.597C> T (p.Pro199=)	Benign(Last reviewed: Mar 26, 2018)	VCV000914909
NM_003070.5(SMARCA2):c.3126-17C>T	Benign(Last reviewed: Mar 26, 2020)	VCV001252991
NM_003070.5(SMARCA2):c.4595-22A>G	Benign(Last reviewed: Mar 27, 2020)	VCV001175587
NM_003070.5(SMARCA2):c.1521+ 21G>A	Benign(Last reviewed: Mar 31, 2020)	VCV001256677
NM_003070.5(SMARCA2):c.1962G >A (p.Gln654=)	Benign(Last reviewed: Mar 31, 2020)	VCV000366210
NM_003070.5(SMARCA2):c.750A> G (p.Gln250=)	Benign(Last reviewed: May 13, 2021)	VCV000366203
NM_003070.5(SMARCA2):c.1188G >T (p.Val396=)	Benign(Last reviewed: May 14, 2021)	VCV001276895
NM_003070.5(SMARCA2):c.3763-47C>A	Benign(Last reviewed: May 15, 2021)	VCV001281331
NM_003070.5(SMARCA2):c.669GC A[16] (p.Gln236_Gln238dup)	Benign(Last reviewed: May 17, 2021)	VCV000587963

NM_003070.5(SMARCA2):c.2348+31G>A	Benign(Last reviewed: May 20, 2020)	VCV001230826
NM_003070.5(SMARCA2):c.51G>C (p.Pro17=)	Benign(Last reviewed: May 24, 2019)	VCV001245835
NM_003070.5(SMARCA2):c.1047-35C>A	Benign(Last reviewed: May 29, 2020)	VCV001294504
NM_003070.5(SMARCA2):c.1047-32C>T	Benign(Last reviewed: May 30, 2020)	VCV001278581
NM_003070.5(SMARCA2):c.1747-48G>C	Benign(Last reviewed: May 31, 2020)	VCV001270992
NM_003070.5(SMARCA2):c.2526+13A>G	Benign(Last reviewed: May 31, 2020)	VCV001243855
NM_003070.5(SMARCA2):c.2883+46A>G	Benign(Last reviewed: May 31, 2020)	VCV001174411
NM_003070.5(SMARCA2):c.1046+45C>T	Benign(Last reviewed: May 4, 2020)	VCV001236579
NM_003070.5(SMARCA2):c.1046+275G>C	Benign(Last reviewed: Nov 10, 2018)	VCV001234193
NM_003070.5(SMARCA2):c.3763-48C>A	Benign(Last reviewed: Nov 11, 2020)	VCV001268759
NM_003070.5(SMARCA2):c.1983C>A (p.Leu661=)	Benign(Last reviewed: Nov 23, 2020)	VCV001269159
NM_003070.5(SMARCA2):c.4738-28C>G	Benign(Last reviewed: Nov 25, 2020)	VCV001283447
NM_003070.5(SMARCA2):c.645G>A (p.Leu215=)	Benign(Last reviewed: Nov 26, 2021)	VCV001326519
NM_003070.5(SMARCA2):c.4200-47G>A	Benign(Last reviewed: Nov 4, 2020)	VCV001236464
NM_003070.5(SMARCA2):c.3078+65del	Benign(Last reviewed: Nov 7, 2019)	VCV001235860

NM_003070.5(SMARCA2):c.790+45G>A	Benign(Last reviewed: Nov 7, 2020)	VCV001263993
NM_003070.5(SMARCA2):c.2883+275_2883+281dup	Benign(Last reviewed: Oct 1, 2019)	VCV001241962
NM_003070.5(SMARCA2):c.355+49A>C	Benign(Last reviewed: Oct 13, 2020)	VCV001251903
NM_003070.5(SMARCA2):c.4590C>T (p.Ser1530=)	Benign(Last reviewed: Oct 15, 2021)	VCV000126347
NM_003070.5(SMARCA2):c.3763-58G>A	Benign(Last reviewed: Oct 17, 2018)	VCV001286142
NM_003070.5(SMARCA2):c.2992-8G>A	Benign(Last reviewed: Oct 17, 2018)	VCV000366217
NM_003070.5(SMARCA2):c.1878-30C>T	Benign(Last reviewed: Oct 21, 2018)	VCV001256925
NM_003070.5(SMARCA2):c.567T>C (p.Tyr189=)	Benign(Last reviewed: Oct 27, 2021)	VCV001302893
NM_003070.5(SMARCA2):c.2526+175TG[25]	Benign(Last reviewed: Oct 29, 2019)	VCV001269625
NM_003070.5(SMARCA2):c.4359+98G>C	Benign(Last reviewed: Oct 31, 2018)	VCV001280456
NM_003070.5(SMARCA2):c.4462-122A>C	Benign(Last reviewed: Oct 31, 2018)	VCV001224274
NM_003070.5(SMARCA2):c.1521+44G>A	Benign(Last reviewed: Oct 9, 2018)	VCV001282827
NM_003070.5(SMARCA2):c.2037-185G>T	Benign(Last reviewed: Sep 1, 2018)	VCV001274285
NM_003070.5(SMARCA2):c.3457-79T>C	Benign(Last reviewed: Sep 1, 2018)	VCV001271671
NM_003070.5(SMARCA2):c.3763-34T>G	Benign(Last reviewed: Sep 1, 2018)	VCV001267888

NM_003070.5(SMARCA2):c.1747-86C>G	Benign(Last reviewed: Sep 1, 2018)	VCV001247753
NM_003070.5(SMARCA2):c.1878-84A>G	Benign(Last reviewed: Sep 1, 2018)	VCV001230525
NM_003070.5(SMARCA2):c.3763-35C>T	Benign(Last reviewed: Sep 1, 2018)	VCV001181741
NM_003070.5(SMARCA2):c.1428C>G (p.Leu476=)	Benign(Last reviewed: Sep 14, 2021)	VCV001302814
NM_003070.5(SMARCA2):c.2526+175TG[18]	Benign(Last reviewed: Sep 18, 2020)	VCV001268087
NM_003070.5(SMARCA2):c.2883+270_2883+281dup	Benign(Last reviewed: Sep 2, 2019)	VCV001246625
NM_003070.5(SMARCA2):c.2526+175TG[21]	Benign(Last reviewed: Sep 20, 2019)	VCV001273314
NM_003070.5(SMARCA2):c.3078+88del	Benign(Last reviewed: Sep 23, 2019)	VCV001282504
NM_003070.5(SMARCA2):c.2991+10G>T	Benign(Last reviewed: Sep 23, 2020)	VCV001275606
NM_003070.5(SMARCA2):c.4359+46G>A	Benign(Last reviewed: Sep 23, 2021)	VCV001300281
NM_003070.5(SMARCA2):c.2349-3T>C	Benign(Last reviewed: Sep 29, 2021)	VCV000366211
GRCh37/hg19 9p24.3(chr9:2149063-2151371)x1	Benign(Last reviewed: Sep 30, 2010)	VCV000611462
NM_003070.5(SMARCA2):c.3684+341G>A	Benign(Last reviewed: Sep 4, 2018)	VCV001296684
NM_003070.5(SMARCA2):c.4200-134G>A	Benign(Last reviewed: Sep 4, 2018)	VCV001228734
NM_003070.5(SMARCA2):c.174G>A (p.Pro58=)	Benign(Last reviewed: Sep 4, 2018)	VCV000126343

NM_003070.5(SMARCA2):c.717G>A (p.Pro239=)	Benign(Last reviewed: Sep 5, 2018)	VCV000126354
NM_003070.5(SMARCA2):c.2526+175TG[24]	Benign(Last reviewed: Sep 8, 2019)	VCV001244021
NM_003070.5(SMARCA2):c.2883+274_2883+281dup	Benign(Last reviewed: Sep 9, 2019)	VCV001274602
NM_003070.5(SMARCA2):c.2349-10dup	Benign(Last reviewed: Sep 9, 2019)	VCV001183167
NM_003070.5(SMARCA2):c.669GCA[17] (p.Gln235_Gln238dup)	Benign/Likely benign(Last reviewed: Apr 15, 2021)	VCV000588473
NM_003070.5(SMARCA2):c.677A>C (p.Gln226Pro)	Benign/Likely benign(Last reviewed: Apr 2, 2019)	VCV000366199
NM_003070.5(SMARCA2):c.4257A>C (p.Ser1419=)	Benign/Likely benign(Last reviewed: Apr 25, 2021)	VCV000366236
NM_003070.5(SMARCA2):c.210G>A (p.Met70Ile)	Benign/Likely benign(Last reviewed: Apr 30, 2020)	VCV000588943
NM_003070.5(SMARCA2):c.1737G>A (p.Pro579=)	Benign/Likely benign(Last reviewed: Apr 7, 2021)	VCV000366206
NM_003070.5(SMARCA2):c.669GCA[12] (p.Gln238del)	Benign/Likely benign(Last reviewed: Aug 12, 2019)	VCV000126353
NM_003070.5(SMARCA2):c.791-7C>T	Benign/Likely benign(Last reviewed: Aug 25, 2020)	VCV000366204
NM_003070.5(SMARCA2):c.4029T>G (p.Leu1343=)	Benign/Likely benign(Last reviewed: Dec 1, 2020)	VCV000366231

NM_003070.5(SMARCA2):c.246C>T (p.Asp82=)	Benign/Likely benign(Last reviewed: Dec 12, 2019)	VCV000366195
NM_003070.5(SMARCA2):c.459G>A (p.Pro153=)	Benign/Likely benign(Last reviewed: Dec 2, 2020)	VCV000914397
NM_003070.5(SMARCA2):c.4717G>A (p.Asp1573Asn)	Benign/Likely benign(Last reviewed: Dec 31, 2019)	VCV000366324
NM_003070.5(SMARCA2):c.1122C>G (p.Thr374=)	Benign/Likely benign(Last reviewed: Dec 4, 2020)	VCV000126342
NM_003070.5(SMARCA2):c.669GCA[10] (p.Gln236_Gln238del)	Benign/Likely benign(Last reviewed: Feb 1, 2022)	VCV000212231
NM_003070.5(SMARCA2):c.2733A>G (p.Gln911=)	Benign/Likely benign(Last reviewed: Feb 10, 2022)	VCV000588361
NM_003070.5(SMARCA2):c.4440G>A (p.Thr1480=)	Benign/Likely benign(Last reviewed: Feb 19, 2020)	VCV000366243
NM_003070.5(SMARCA2):c.4761G>A (p.Thr1587=)	Benign/Likely benign(Last reviewed: Feb 4, 2020)	VCV000366327
NM_003070.5(SMARCA2):c.680A>C (p.Gln227Pro)	Benign/Likely benign(Last reviewed: Jan 13, 2018)	VCV000588909
NM_003070.5(SMARCA2):c.3438C>T (p.Ser1146=)	Benign/Likely benign(Last reviewed: Jan 18, 2019)	VCV000366220
NM_003070.5(SMARCA2):c.669G>A (p.Gln223=)	Benign/Likely benign(Last reviewed: Jan 18, 2019)	VCV000366198
NM_003070.5(SMARCA2):c.666A>G (p.Gln222=)	Benign/Likely benign(Last reviewed: Jan 4, 2019)	VCV000366197

NM_003070.5(SMARCA2):c.3282A>G (p.Leu1094=)	Benign/Likely benign(Last reviewed: Jul 8, 2019)	VCV000588468
NM_003070.5(SMARCA2):c.483G>T (p.Pro161=)	Benign/Likely benign(Last reviewed: Jun 1, 2021)	VCV000366196
NM_003070.5(SMARCA2):c.513C>A (p.Pro171=)	Benign/Likely benign(Last reviewed: Jun 10, 2019)	VCV000212229
NM_003070.5(SMARCA2):c.2136G>A (p.Val712=)	Benign/Likely benign(Last reviewed: Jun 26, 2019)	VCV000588159
NM_003070.5(SMARCA2):c.4479C>T (p.Ile1493=)	Benign/Likely benign(Last reviewed: Mar 1, 2021)	VCV000914528
NM_003070.5(SMARCA2):c.4207G>A (p.Val1403Met)	Benign/Likely benign(Last reviewed: Mar 11, 2020)	VCV000366232
NM_003070.5(SMARCA2):c.3843C>T (p.Pro1281=)	Benign/Likely benign(Last reviewed: Mar 25, 2021)	VCV000588558
NM_003070.5(SMARCA2):c.4200-4G>A	Benign/Likely benign(Last reviewed: Mar 5, 2020)	VCV000588141
NM_003070.5(SMARCA2):c.685_686insCGC (p.Gln229_Gln230insPro)	Benign/Likely benign(Last reviewed: May 9, 2020)	VCV000588299
NM_003070.5(SMARCA2):c.4646G>A (p.Arg1549Gln)	Benign/Likely benign(Last reviewed: Nov 2, 2018)	VCV000915171
NM_003070.5(SMARCA2):c.669GCA[15] (p.Gln237_Gln238dup)	Benign/Likely benign(Last reviewed: Nov 20, 2020)	VCV000436800
NM_003070.5(SMARCA2):c.669GCA[8] (p.Gln234_Gln238del)	Benign/Likely benign(Last reviewed: Nov 5, 2018)	VCV000212230

NM_003070.5(SMARCA2):c.4699G >C (p.Val1567Leu)	Benign/Likely benign(Last reviewed: Oct 10, 2019)	VCV000366322
NM_003070.5(SMARCA2):c.4731T >C (p.Asp1577=)	Benign/Likely benign(Last reviewed: Oct 14, 2020)	VCV000588668
NM_003070.5(SMARCA2):c.844G> A (p.Ala282Thr)	Conflicting interpretations of pathogenicity(Last reviewed: Aug 2, 2019)	VCV000588156
NM_003070.5(SMARCA2):c.876C> T (p.Pro292=)	Conflicting interpretations of pathogenicity(Last reviewed: Dec 2, 2020)	VCV000913314
NM_003070.5(SMARCA2):c.2810G >A (p.Arg937His)	Conflicting interpretations of pathogenicity(Last reviewed: Feb 18, 2022)	VCV000827774
NM_003070.5(SMARCA2):c.669GC A[19] (p.Gln233_Gln238dup)	Conflicting interpretations of pathogenicity(Last reviewed: Jan 1, 2019)	VCV000126352
NM_003070.5(SMARCA2):c.175A> T (p.Thr59Ser)	Conflicting interpretations of pathogenicity(Last reviewed: Jan 13, 2018)	VCV000366194
NM_003070.5(SMARCA2):c.3229T >A (p.Ser1077Thr)	Conflicting interpretations of pathogenicity(Last reviewed: Jan 3, 2020)	VCV001311593
NM_003070.5(SMARCA2):c.2361C >A (p.Asn787Lys)	Conflicting interpretations of pathogenicity(Last reviewed: Jun 10, 2021)	VCV000694693

NM_003070.5(SMARCA2):c.2348C>T (p.Ser783Leu)	Conflicting interpretations of pathogenicity(Last reviewed: Jun 10, 2021)	VCV000561113
NM_003070.5(SMARCA2):c.1514G>A (p.Arg505Gln)	Conflicting interpretations of pathogenicity(Last reviewed: Mar 25, 2020)	VCV000829813
NM_003070.5(SMARCA2):c.1854C>T (p.Asp618=)	Conflicting interpretations of pathogenicity(Last reviewed: May 13, 2021)	VCV000212224
NM_003070.5(SMARCA2):c.225+145G>A	Likely benign(Last reviewed: Apr 1, 2019)	VCV001185805
NM_003070.5(SMARCA2):c.3492G>A (p.Gly1164=)	Likely benign(Last reviewed: Apr 15, 2021)	VCV001309084
NM_003070.5(SMARCA2):c.400G>A (p.Val134Ile)	Likely benign(Last reviewed: Apr 21, 2017)	VCV000589374
NM_003070.5(SMARCA2):c.2883+203G>A	Likely benign(Last reviewed: Apr 29, 2019)	VCV001210931
NM_003070.5(SMARCA2):c.669GCA[11] (p.Gln237_Gln238del)	Likely benign(Last reviewed: Apr 4, 2018)	VCV000587879
NM_003070.5(SMARCA2):c.1935+23G>T	Likely benign(Last reviewed: Apr 5, 2021)	VCV001300815
NM_003070.5(SMARCA2):c.4716C>T (p.Ser1572=)	Likely benign(Last reviewed: Apr 7, 2021)	VCV001301136

NM_003070.5(SMARCA2):c.2185-93A>G	Likely benign(Last reviewed: Apr 9, 2019)	VCV001213498
NM_003070.5(SMARCA2):c.1408G>T (p.Ala470Ser)	Likely benign(Last reviewed: Aug 1, 2019)	VCV000810344
NM_003070.5(SMARCA2):c.4200-57G>A	Likely benign(Last reviewed: Aug 10, 2018)	VCV001180191
NM_003070.5(SMARCA2):c.3078+65dup	Likely benign(Last reviewed: Aug 10, 2019)	VCV001215316
NM_003070.5(SMARCA2):c.2348+312GT[23]	Likely benign(Last reviewed: Aug 10, 2019)	VCV001204968
NM_003070.5(SMARCA2):c.2348+312GT[19]	Likely benign(Last reviewed: Aug 11, 2019)	VCV001215668
NM_003070.5(SMARCA2):c.226-234AAAC[7]	Likely benign(Last reviewed: Aug 13, 2019)	VCV001219596
NM_003070.5(SMARCA2):c.4253+135G>A	Likely benign(Last reviewed: Aug 17, 2018)	VCV001219728
NM_003070.5(SMARCA2):c.-36-205T>C	Likely benign(Last reviewed: Aug 17, 2018)	VCV001191112
NM_003070.5(SMARCA2):c.902C>T (p.Ala301Val)	Likely benign(Last reviewed: Aug 25, 2021)	VCV001254900
NM_003070.5(SMARCA2):c.704_705insACAACAGCAGCC (p.236QQQP[3])	Likely benign(Last reviewed: Aug 26, 2021)	VCV001254928
NM_003070.5(SMARCA2):c.4479C>G (p.Ile1493Met)	Likely benign(Last reviewed: Aug 3, 2020)	VCV001337600

NM_003070.5(SMARCA2):c.3882C>G (p.Leu1294=)	Likely benign(Last reviewed: Aug 30, 2016)	VCV000589368
NM_003070.5(SMARCA2):c.3982-24G>A	Likely benign(Last reviewed: Aug 31, 2018)	VCV001219608
NM_003070.5(SMARCA2):c.4461+204G>A	Likely benign(Last reviewed: Aug 31, 2018)	VCV001206728
NM_003070.5(SMARCA2):c.226-123T>C	Likely benign(Last reviewed: Aug 8, 2018)	VCV001209526
NM_003070.5(SMARCA2):c.4359+126G>A	Likely benign(Last reviewed: Aug 8, 2018)	VCV001201328
NM_003070.5(SMARCA2):c.1348-113C>G	Likely benign(Last reviewed: Aug 8, 2018)	VCV001197819
NM_003070.5(SMARCA2):c.2883+107G>C	Likely benign(Last reviewed: Aug 8, 2018)	VCV001194879
NM_003070.5(SMARCA2):c.1522-182G>A	Likely benign(Last reviewed: Dec 1, 2018)	VCV001218462
NM_003070.5(SMARCA2):c.1348-150G>C	Likely benign(Last reviewed: Dec 1, 2018)	VCV001195815
NM_003070.5(SMARCA2):c.1201C>T (p.Arg401Cys)	Likely benign(Last reviewed: Dec 1, 2020)	VCV001197110
NM_003070.5(SMARCA2):c.2929C>T (p.Leu977=)	Likely benign(Last reviewed: Dec 14, 2021)	VCV001327749
NM_003070.5(SMARCA2):c.4360-16C>G	Likely benign(Last reviewed: Dec 17, 2018)	VCV001217641

NM_003070.5(SMARCA2):c.3457-319T>C	Likely benign(Last reviewed: Dec 17, 2018)	VCV001199594
NM_003070.5(SMARCA2):c.1174-122T>C	Likely benign(Last reviewed: Dec 17, 2018)	VCV001197362
NM_003070.5(SMARCA2):c.1173+130G>C	Likely benign(Last reviewed: Dec 17, 2018)	VCV001191289
NM_003070.5(SMARCA2):c.2036+217T>C	Likely benign(Last reviewed: Dec 17, 2018)	VCV001187328
NM_003070.5(SMARCA2):c.666_683del (p.Gln233_Gln238del)	Likely benign(Last reviewed: Dec 17, 2021)	VCV001328645
NM_003070.5(SMARCA2):c.2805C>T (p.Ile935=)	Likely benign(Last reviewed: Dec 23, 2016)	VCV000588587
NM_003070.5(SMARCA2):c.681G>A (p.Gln227=)	Likely benign(Last reviewed: Dec 26, 2019)	VCV001188115
NM_003070.5(SMARCA2):c.3192G>A (p.Ala1064=)	Likely benign(Last reviewed: Dec 30, 2015)	VCV000436799
NM_003070.5(SMARCA2):c.2185-234C>G	Likely benign(Last reviewed: Dec 5, 2018)	VCV001190753
NM_003070.5(SMARCA2):c.2526+52T>G	Likely benign(Last reviewed: Dec 5, 2020)	VCV001197473
NM_003070.5(SMARCA2):c.2154C>G (p.Leu718=)	Likely benign(Last reviewed: Dec 8, 2020)	VCV001320537
NM_003070.5(SMARCA2):c.3267G>A (p.Arg1089=)	Likely benign(Last reviewed: Feb 13, 2020)	VCV000366219

NM_003070.5(SMARCA2):c.3555C>T (p.Leu1185=)	Likely benign(Last reviewed: Feb 23, 2021)	VCV000588117
NM_003070.5(SMARCA2):c.2184+5T>C	Likely benign(Last reviewed: Jan 1, 2017)	VCV000562022
NM_003070.5(SMARCA2):c.3230C>G (p.Ser1077Cys)	Likely benign(Last reviewed: Jan 1, 2019)	VCV000982924
NM_003070.5(SMARCA2):c.4461+3C>G	Likely benign(Last reviewed: Jan 1, 2019)	VCV000982825
NM_003070.5(SMARCA2):c.4080T>G (p.Asp1360Glu)	Likely benign(Last reviewed: Jan 12, 2018)	VCV000914996
NM_003070.5(SMARCA2):c.750A>T (p.Gln250His)	Likely benign(Last reviewed: Jan 12, 2018)	VCV000913313
NM_003070.5(SMARCA2):c.*29C>T	Likely benign(Last reviewed: Jan 12, 2018)	VCV000366331
NM_003070.5(SMARCA2):c.4206C>T (p.Asn1402=)	Likely benign(Last reviewed: Jan 13, 2018)	VCV000914997
NM_003070.5(SMARCA2):c.957C>G (p.Leu319=)	Likely benign(Last reviewed: Jan 13, 2018)	VCV000913316
NM_003070.5(SMARCA2):c.4725G>A (p.Glu1575=)	Likely benign(Last reviewed: Jan 13, 2018)	VCV000366325
NM_003070.5(SMARCA2):c.4029T>A (p.Leu1343=)	Likely benign(Last reviewed: Jan 13, 2018)	VCV000366230
NM_003070.5(SMARCA2):c.1746+10C>T	Likely benign(Last reviewed: Jan 13, 2018)	VCV000366207

NM_003070.5(SMARCA2):c.1747-11C>A	Likely benign(Last reviewed: Jan 25, 2021)	VCV001193234
NM_003070.5(SMARCA2):c.1521+55A>G	Likely benign(Last reviewed: Jan 28, 2019)	VCV001210650
NM_003070.5(SMARCA2):c.2762G>T (p.Gly921Val)	Likely benign(Last reviewed: Jan 8, 2021)	VCV001337248
NM_003070.5(SMARCA2):c.356-55T>C	Likely benign(Last reviewed: Jul 26, 2018)	VCV001211670
NM_003070.5(SMARCA2):c.1878-27C>T	Likely benign(Last reviewed: Jul 27, 2020)	VCV001209310
NM_003070.5(SMARCA2):c.704A>C (p.Gln235Pro)	Likely benign(Last reviewed: Jul 29, 2021)	VCV001219703
NM_003070.5(SMARCA2):c.2883+37dup	Likely benign(Last reviewed: Jul 7, 2020)	VCV001206659
NM_003070.5(SMARCA2):c.*355dup	Likely benign(Last reviewed: Jun 14, 2016)	VCV000366336
NM_003070.5(SMARCA2):c.701A>C (p.Gln234Pro)	Likely benign(Last reviewed: Jun 14, 2016)	VCV000366201
NM_003070.5(SMARCA2):c.4509G>A (p.Arg1503=)	Likely benign(Last reviewed: Jun 16, 2021)	VCV001328668
NM_003070.5(SMARCA2):c.4595-261_4595-254del	Likely benign(Last reviewed: Jun 17, 2020)	VCV001189961
NM_003070.5(SMARCA2):c.667_668insCAG (p.Gln223delinsProGlu)	Likely benign(Last reviewed: Jun 17, 2021)	VCV001328660

NM_003070.5(SMARCA2):c.744G>A (p.Thr248=)	Likely benign(Last reviewed: Jun 18, 2016)	VCV000589294
NM_003070.5(SMARCA2):c.682_683insCGC (p.Gln228_Gln229insPro)	Likely benign(Last reviewed: Jun 22, 2021)	VCV001329651
NM_003070.5(SMARCA2):c.4595-48C>T	Likely benign(Last reviewed: Jun 24, 2021)	VCV001329774
NM_003070.5(SMARCA2):c.4738-281del	Likely benign(Last reviewed: Jun 5, 2019)	VCV001202877
NM_003070.5(SMARCA2):c.4462-312A>G	Likely benign(Last reviewed: Jun 5, 2019)	VCV001187510
NM_003070.5(SMARCA2):c.3729T>A (p.Ile1243=)	Likely benign(Last reviewed: Mar 1, 2018)	VCV000810347
NM_003070.5(SMARCA2):c.3191C>T (p.Ala1064Val)	Likely benign(Last reviewed: Mar 12, 2021)	VCV001254664
NM_003070.5(SMARCA2):c.1676GGA[3] (p.Arg562del)	Likely benign(Last reviewed: Mar 13, 2020)	VCV001218335
NM_003070.5(SMARCA2):c.669GCA[18] (p.Gln234_Gln238dup)	Likely benign(Last reviewed: Mar 17, 2021)	VCV001209185
NM_003070.5(SMARCA2):c.3292+35C>G	Likely benign(Last reviewed: Mar 19, 2019)	VCV001197398
NM_003070.5(SMARCA2):c.*9G>C	Likely benign(Last reviewed: Mar 20, 2020)	VCV000366329
NM_003070.5(SMARCA2):c.791-6C>T	Likely benign(Last reviewed: Mar 21, 2018)	VCV000680332

NM_003070.5(SMARCA2):c.1664A GA[3] (p.Lys558del)	Likely benign(Last reviewed: Mar 25, 2020)	VCV001198505
NM_003070.5(SMARCA2):c.2991+ 40A>G	Likely benign(Last reviewed: Mar 27, 2021)	VCV001300383
NM_003070.5(SMARCA2):c.3963G >A (p.Thr1321=)	Likely benign(Last reviewed: Mar 29, 2017)	VCV000588973
NM_003070.5(SMARCA2):c.4254- 319_4254-317dup	Likely benign(Last reviewed: Mar 29, 2019)	VCV001199480
NM_003070.5(SMARCA2):c.669GC A[9] (p.Gln235_Gln238del)	Likely benign(Last reviewed: Mar 3, 2021)	VCV000588716
NM_003070.5(SMARCA2):c.2883+ 27G>T	Likely benign(Last reviewed: Mar 31, 2020)	VCV001207787
NM_003070.5(SMARCA2):c.695_71 5del (p.Gln232_Gln238del)	Likely benign(Last reviewed: Mar 9, 2020)	VCV001182585
NM_003070.5(SMARCA2):c.4764T >A (p.Asp1588Glu)	Likely benign(Last reviewed: Mar 9, 2021)	VCV001342724
NM_003070.5(SMARCA2):c.2532G >A (p.Arg844=)	Likely benign(Last reviewed: Mar 9, 2021)	VCV001300851
NM_003070.5(SMARCA2):c.2244C >T (p.Ala748=)	Likely benign(Last reviewed: May 18, 2021)	VCV001327195
NM_003070.5(SMARCA2):c.669GC A[3] (p.Gln229_Gln238del)	Likely benign(Last reviewed: May 24, 2021)	VCV001338913
NM_003070.5(SMARCA2):c.4595- 182C>G	Likely benign(Last reviewed: May 26, 2019)	VCV001213233

NM_003070.5(SMARCA2):c.2211C>T (p.Ser737=)	Likely benign(Last reviewed: May 5, 2021)	VCV001327677
NM_003070.5(SMARCA2):c.1347+206dup	Likely benign(Last reviewed: May 8, 2020)	VCV001207419
NM_003070.5(SMARCA2):c.1035G>A (p.Glu345=)	Likely benign(Last reviewed: Nov 1, 2021)	VCV001335724
NM_003070.5(SMARCA2):c.1890C>T (p.Ala630=)	Likely benign(Last reviewed: Nov 11, 2020)	VCV001301177
NM_003070.5(SMARCA2):c.1173+299G>T	Likely benign(Last reviewed: Nov 22, 2018)	VCV001202248
NM_003070.5(SMARCA2):c.3762+250A>G	Likely benign(Last reviewed: Nov 25, 2018)	VCV001200301
NM_003070.5(SMARCA2):c.226-234AAAC[8]	Likely benign(Last reviewed: Nov 3, 2019)	VCV001194512
NM_003070.5(SMARCA2):c.669GCA[6] (p.Gln232_Gln238del)	Likely benign(Last reviewed: Nov 9, 2020)	VCV000587833
NM_003070.5(SMARCA2):c.4594+165C>T	Likely benign(Last reviewed: Oct 14, 2018)	VCV001204791
NM_003070.5(SMARCA2):c.1047-129G>A	Likely benign(Last reviewed: Oct 16, 2018)	VCV001214208
NM_003070.5(SMARCA2):c.2184+45A>G	Likely benign(Last reviewed: Oct 16, 2018)	VCV001190752
NM_003070.5(SMARCA2):c.1522-19G>A	Likely benign(Last reviewed: Oct 16, 2018)	VCV001186042

NM_003070.5(SMARCA2):c.3126-285T>C	Likely benign(Last reviewed: Oct 17, 2018)	VCV001198678
NM_003070.5(SMARCA2):c.3079-139C>T	Likely benign(Last reviewed: Oct 17, 2018)	VCV001179267
NM_003070.5(SMARCA2):c.4200-131_4200-130insCTT	Likely benign(Last reviewed: Oct 20, 2020)	VCV001194536
NM_003070.5(SMARCA2):c.3684+24A>T	Likely benign(Last reviewed: Oct 21, 2018)	VCV001205159
NM_003070.5(SMARCA2):c.2348+312GT[20]	Likely benign(Last reviewed: Oct 28, 2019)	VCV001203724
NM_003070.5(SMARCA2):c.1521+48G>A	Likely benign(Last reviewed: Oct 31, 2018)	VCV001200685
NM_003070.5(SMARCA2):c.3981+235G>A	Likely benign(Last reviewed: Oct 5, 2018)	VCV001208309
NM_003070.5(SMARCA2):c.708_722del (p.Gln238_Gln242del)	Likely benign(Last reviewed: Oct 6, 2020)	VCV001211740
NM_003070.5(SMARCA2):c.3274C>G (p.Leu1092Val)	Likely benign(Last reviewed: Oct 7, 2019)	VCV001202936
NM_003070.5(SMARCA2):c.3292+99T>C	Likely benign(Last reviewed: Oct 9, 2018)	VCV001203402
NM_003070.5(SMARCA2):c.1173+274T>C	Likely benign(Last reviewed: Oct 9, 2018)	VCV001195577
NM_003070.5(SMARCA2):c.4738-271T>G	Likely benign(Last reviewed: Sep 1, 2018)	VCV001180080

NM_003070.5(SMARCA2):c.2416-143C>G	Likely benign(Last reviewed: Sep 2, 2019)	VCV001218779
NM_003070.5(SMARCA2):c.3384T>C (p.Ala1128=)	Likely benign(Last reviewed: Sep 26, 2018)	VCV001336892
NM_003070.5(SMARCA2):c.666_686del (p.Gln232_Gln238del)	Likely benign(Last reviewed: Sep 26, 2020)	VCV001212093
NM_003070.5(SMARCA2):c.356-220dup	Likely benign(Last reviewed: Sep 30, 2019)	VCV001192109
NM_003070.5(SMARCA2):c.3762+182G>A	Likely benign(Last reviewed: Sep 4, 2018)	VCV001213430
NM_003070.5(SMARCA2):c.4594+95C>A	Likely benign(Last reviewed: Sep 4, 2018)	VCV001212574
NM_003070.5(SMARCA2):c.2036+64T>C	Likely benign(Last reviewed: Sep 4, 2018)	VCV001196254
NM_003070.5(SMARCA2):c.4594+152G>T	Likely benign(Last reviewed: Sep 4, 2018)	VCV001193834
NM_003070.5(SMARCA2):c.2185-271C>T	Likely benign(Last reviewed: Sep 4, 2018)	VCV001185727
NM_003070.5(SMARCA2):c.791-134_791-133del	Likely benign(Last reviewed: Sep 5, 2019)	VCV001186369
NM_003070.5(SMARCA2):c.2526+214_2526+215insGGTGTG	Likely benign(Last reviewed: Sep 6, 2019)	VCV001220119
NM_003070.5(SMARCA2):c.2883+271_2883+281dup	Likely benign(Last reviewed: Sep 8, 2019)	VCV001219783

NM_001289396.1:c.[3495G>C(;)3917G>A]	Likely pathogenic	VCV000374221
NM_003070.5(SMARCA2):c.1538G>T (p.Gly513Val)	Likely pathogenic	VCV000829812
NM_003070.5(SMARCA2):c.2809C>T (p.Arg937Cys)	Likely pathogenic	VCV000827771
NM_003070.5(SMARCA2):c.1600G>A (p.Asp534Asn)	Likely pathogenic	VCV000827769
GRCh37/hg19 9p24.3-24.1(chr9:1232387-4611862)x1	Likely pathogenic	VCV000396183
NM_003070.5(SMARCA2):c.787T>A (p.Ser263Thr)	Likely pathogenic(Last reviewed: Apr 1, 2020)	VCV000978583
NM_003070.5(SMARCA2):c.2834T>G (p.Phe945Cys)	Likely pathogenic(Last reviewed: Apr 11, 2017)	VCV000449924
NM_003070.5(SMARCA2):c.3490G>A (p.Gly1164Arg)	Likely pathogenic(Last reviewed: Aug 1, 2021)	VCV001299136
NM_003070.5(SMARCA2):c.3612T>G (p.Phe1204Leu)	Likely pathogenic(Last reviewed: Aug 24, 2016)	VCV000430455
NM_003070.5(SMARCA2):c.1600G>C (p.Asp534His)	Likely pathogenic(Last reviewed: Aug 24, 2016)	VCV000988516
NM_003070.5(SMARCA2):c.3457-2A>T	Likely pathogenic(Last reviewed: Aug 7, 2020)	VCV000981460
GRCh37/hg19 9p24.3-11.2(chr9:204193-44259464)x4	Likely pathogenic(Last reviewed: Dec 19, 2016)	VCV000559575
NM_003070.5(SMARCA2):c.3021C>G (p.Asn1007Lys)	Likely pathogenic(Last reviewed: Dec 29, 2015)	VCV000265528

NM_003070.5(SMARCA2):c.3849G >T (p.Trp1283Cys)	Likely pathogenic(Last reviewed: Dec 8, 2016)	VCV000373701
NM_003070.5(SMARCA2):c.3236T >C (p.Met1079Thr)	Likely pathogenic(Last reviewed: Feb 12, 2016)	VCV000432081
NM_003070.5(SMARCA2):c.1600G >T (p.Asp534Tyr)	Likely pathogenic(Last reviewed: Jul 22, 2014)	VCV000217002
NM_003070.5(SMARCA2):c.2639C >A (p.Thr880Asn)	Likely pathogenic(Last reviewed: Jun 1, 2021)	VCV001176897
NM_003070.5(SMARCA2):c.2254G >A (p.Gly752Arg)	Likely pathogenic(Last reviewed: Jun 10, 2021)	VCV001032869
NM_003070.5(SMARCA2):c.3599A >C (p.Gln1200Pro)	Likely pathogenic(Last reviewed: Jun 24, 2016)	VCV000976221
NM_003070.5(SMARCA2):c.3482A >G (p.His1161Arg)	Likely pathogenic(Last reviewed: Jun 3, 2014)	VCV000217001
NM_003070.5(SMARCA2):c.473del (p.Pro158fs)	Likely pathogenic(Last reviewed: Mar 8, 2019)	VCV000817478
NM_003070.5(SMARCA2):c.3441C >A (p.Asp1147Glu)	Likely pathogenic(Last reviewed: Mar 9, 2020)	VCV000988739
NM_003070.5(SMARCA2):c.3314G >T (p.Arg1105Leu)	Likely pathogenic(Last reviewed: May 1, 2019)	VCV000810345
NM_003070.5(SMARCA2):c.3587A >C (p.Gln1196Pro)	Likely pathogenic(Last reviewed: May 28, 2019)	VCV000802455
NM_003070.5(SMARCA2):c.2737T >C (p.Phe913Leu)	Likely pathogenic(Last reviewed: May 31, 2016)	VCV000521074

NM_003070.5(SMARCA2):c.1553T >C (p.Ile518Thr)	Likely pathogenic(Last reviewed: May 6, 2020)	VCV000992996
NM_003070.5(SMARCA2):c.3456G >C (p.Gln1152His)	Likely pathogenic(Last reviewed: May 8, 2017)	VCV000429419
NM_003070.5(SMARCA2):c.3562G >A (p.Ala1188Thr)	Likely pathogenic(Last reviewed: Nov 10, 2016)	VCV000369656
NM_003070.5(SMARCA2):c.1601A >G (p.Asp534Gly)	Likely pathogenic(Last reviewed: Nov 3, 2021)	VCV001319165
NM_003070.5(SMARCA2):c.1540T >C (p.Tyr514His)	Likely pathogenic(Last reviewed: Oct 12, 2017)	VCV000452666
NM_003070.5(SMARCA2):c.3493C >A (p.Gln1165Lys)	Likely pathogenic(Last reviewed: Oct 24, 2014)	VCV000212227
NM_003070.5(SMARCA2):c.3962C >T (p.Thr1321Met)	Likely pathogenic(Last reviewed: Oct 26, 2020)	VCV001285425
NM_003070.5(SMARCA2):c.2648C >A (p.Pro883Gln)	Likely pathogenic(Last reviewed: Oct 31, 2018)	VCV000436804
NM_003070.5(SMARCA2):c.2326T >C (p.Tyr776His)	Likely pathogenic(Last reviewed: Oct 4, 2016)	VCV000521300
NM_003070.5(SMARCA2):c.2342C >G (p.Pro781Arg)	Likely pathogenic(Last reviewed: Sep 10, 2020)	VCV000981428
NM_003070.5(SMARCA2):c.3623C >G (p.Ser1208Cys)	Likely pathogenic(Last reviewed: Sep 10, 2020)	VCV000981388
NM_003070.5(SMARCA2):c.2348C >G (p.Ser783Trp)	Likely pathogenic(Last reviewed: Sep 10, 2020)	VCV000436803

NM_003070.5(SMARCA2):c.1574G>A (p.Arg525His)	Likely pathogenic(Last reviewed: Sep 10, 2020)	VCV000829814
NM_003070.5(SMARCA2):c.2786A>T (p.Glu929Val)	Likely pathogenic(Last reviewed: Sep 10, 2020)	VCV000827770
NM_003070.5(SMARCA2):c.1538G>A (p.Gly513Asp)	Likely pathogenic(Last reviewed: Sep 16, 2020)	VCV000979174
NM_003070.5(SMARCA2):c.2552A>G (p.Asp851Gly)	Likely pathogenic(Last reviewed: Sep 2, 2016)	VCV000267266
GRCh37/hg19 9p24.3-23(chr9:203861-10700288)x3	Likely pathogenic(Last reviewed: Sep 22, 2014)	VCV000443400
NM_003070.5(SMARCA2):c.3602C>A (p.Ala1201Glu)	Likely pathogenic(Last reviewed: Sep 22, 2019)	VCV001206691
NM_003070.5(SMARCA2):c.1529A>G (p.Asp510Gly)	Likely pathogenic(Last reviewed: Sep 30, 2020)	VCV000982400
NM_003070.5(SMARCA2):c.3495G>C (p.Gln1165His)	no interpretation for the single variant	VCV000374406
NM_003070.5(SMARCA2):c.3917G>A (p.Arg1306Lys)	no interpretation for the single variant	VCV000374405
NM_003070.5(SMARCA2):c.3614A>G (p.Asp1205Gly)	not provided	VCV000068774
NM_003070.5(SMARCA2):c.3562G>C (p.Ala1188Pro)	not provided	VCV000068773
NM_003070.5(SMARCA2):c.3436A>C (p.Ser1146Arg)	not provided	VCV000068772
NM_003070.5(SMARCA2):c.3404T>C (p.Leu1135Pro)	not provided	VCV000068771
NM_003070.5(SMARCA2):c.3314G>C (p.Arg1105Pro)	not provided	VCV000068770

NM_003070.5(SMARCA2):c.2838A>T (p.Leu946Phe)	not provided	VCV000068768
NM_003070.5(SMARCA2):c.2837T>C (p.Leu946Ser)	not provided	VCV000068767
NM_003070.5(SMARCA2):c.2641G>C (p.Gly881Arg)	not provided	VCV000068766
NM_003070.5(SMARCA2):c.2556A>C (p.Glu852Asp)	not provided	VCV000068763
NM_003070.5(SMARCA2):c.2551G>C (p.Asp851His)	not provided	VCV000068761
NM_003070.5(SMARCA2):c.2264A>G (p.Lys755Arg)	not provided	VCV000068759
NM_003070.5(SMARCA2):c.2561A>G (p.His854Arg)	not provided	VCV000068764
NM_003070.5(SMARCA2):c.2563C>G (p.Arg855Gly)	Pathogenic	VCV000068765
GRCh37/hg19 9p24.3-24.2(chr9:213161-3497920)x1	Pathogenic	VCV000395051
GRCh37/hg19 9p24.3-24.1(chr9:203861-5909152)x1	Pathogenic	VCV000396782
GRCh37/hg19 9p24.3-23(chr9:203861-11414732)x1	Pathogenic	VCV000396960
GRCh37/hg19 9p24.3-24.1(chr9:203861-8735462)x1	Pathogenic	VCV000395984
GRCh37/hg19 9p24.3-22.3(chr9:203861-14322268)x1	Pathogenic	VCV000395946
GRCh37/hg19 9p24.3-22.3(chr9:203861-15211277)x1	Pathogenic	VCV000396532
GRCh37/hg19 9p24.3-22.2(chr9:203861-16670878)x1	Pathogenic	VCV000396784

GRCh37/hg19 9p24.3-22.2(chr9:203861-16856907)x1	Pathogenic	VCV000396028
GRCh37/hg19 9p24.3-22.2(chr9:203861-16925108)x1	Pathogenic	VCV000396012
GRCh37/hg19 9p24.3-22.2(chr9:204193-18073357)x1	Pathogenic	VCV000981212
GRCh37/hg19 9p24.3-22.2(chr9:213161-17496750)x1	Pathogenic	VCV000394262
GRCh37/hg19 9p24.3-13.1(chr9:32396-39140211)	Pathogenic	VCV000394346
GRCh37/hg19 9p24.3-q13(chr9:203861-67983174)x4	Pathogenic	VCV000396847
GRCh37/hg19 9p24.3-q13(chr9:203861-68188391)x4	Pathogenic	VCV000396397
GRCh37/hg19 9p24.3-q21.11(chr9:203861-69002883)x3	Pathogenic	VCV000396594
GRCh37/hg19 9p24.3-q21.11(chr9:13997-70919878)x4	Pathogenic	VCV000393860
GRCh37/hg19 9p24.3-q34.3(chr9:203861-141020389)x3	Pathogenic	VCV000397469
GRCh37/hg19 9p24.3-q34.3(chr9:203864-141020389)x3	Pathogenic	VCV000396494
NM_003070.5(SMARCA2):c.3395G>A (p.Gly1132Asp)	Pathogenic(Last reviewed: Apr 1, 2012)	VCV000031687

GRCh38/hg38 9p24.3-24.1(chr9:204104-5695507)x1	Pathogenic(Last reviewed: Apr 14, 2011)	VCV000148301
GRCh38/hg38 9p24.3-23(chr9:204104-11610300)x3	Pathogenic(Last reviewed: Apr 14, 2011)	VCV000148307
GRCh37/hg19 9p24.3-23(chr9:203861-14080419)x1	Pathogenic(Last reviewed: Apr 2, 2019)	VCV000815189
NM_003070.5(SMARCA2):c.3650T>C (p.Leu1217Pro)	Pathogenic(Last reviewed: Apr 20, 2018)	VCV000280814
GRCh37/hg19 9p24.3-q21.33(chr9:203861-88189913)x3	Pathogenic(Last reviewed: Apr 22, 2014)	VCV000442489
GRCh38/hg38 9p24.3-24.1(chr9:204104-8266492)x1	Pathogenic(Last reviewed: Apr 27, 2011)	VCV000148381
GRCh38/hg38 9p24.3-23(chr9:204193-10164955)x1	Pathogenic(Last reviewed: Apr 29, 2013)	VCV000144246
GRCh38/hg38 9p24.3-13.1(chr9:203861-38381642)	Pathogenic(Last reviewed: Apr 29, 2013)	VCV000155344
GRCh38/hg38 9p24.3-24.1(chr9:185579-7635806)x1	Pathogenic(Last reviewed: Apr 30, 2011)	VCV000152906
GRCh38/hg38 9p24.3-24.1(chr9:211086-7444397)x1	Pathogenic(Last reviewed: Apr 30, 2011)	VCV000151916
GRCh38/hg38 9p24.3-23(chr9:211086-11457340)x1	Pathogenic(Last reviewed: Apr 30, 2011)	VCV000152907
GRCh37/hg19 9p24.3-q34.3(chr9:62525-141006407)	Pathogenic(Last reviewed: Apr 30, 2011)	VCV000395707

GRCh38/hg38 9p24.3-24.1(chr9:204104-5426099)x3	Pathogenic(Last reviewed: Apr 8, 2011)	VCV000148264
GRCh38/hg38 9p24.3-24.2(chr9:204193-3468435)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059837
GRCh38/hg38 9p24.3-24.2(chr9:220253-3793376)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057352
GRCh38/hg38 9p24.3-24.2(chr9:280255-3905421)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059066
GRCh38/hg38 9p24.3-24.1(chr9:220253-5140455)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059063
GRCh38/hg38 9p24.3-24.1(chr9:220253-6073001)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059062
GRCh38/hg38 9p24.3-24.1(chr9:211086-6106482)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000060440
GRCh38/hg38 9p24.3-24.1(chr9:220253-6968724)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059065
GRCh38/hg38 9p24.3-23(chr9:1592306-12387899)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057027
GRCh38/hg38 9p24.3-23(chr9:204193-10340779)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000060431
GRCh38/hg38 9p24.3-23(chr9:211086-11867480)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000060441
GRCh38/hg38 9p24.3-23(chr9:195399-11081440)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000060416

GRCh38/hg38 9p24.3-23(chr9:203993-12621562)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000060427
GRCh38/hg38 9p24.3-23(chr9:211087-13754567)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059060
GRCh38/hg38 9p24.3-23(chr9:204193-13974100)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057180
GRCh38/hg38 9p24.3-22.3(chr9:111216-14650762)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000060415
GRCh38/hg38 9p24.3-22.1(chr9:1242978-18957216)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059067
GRCh38/hg38 9p24.3-22.2(chr9:220253-18073359)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059064
GRCh38/hg38 9p24.3-21.3(chr9:204193-22086858)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059836
GRCh38/hg38 9p24.3-13.3(chr9:204193-33284638)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059839
GRCh38/hg38 9p24.3-13.1(chr9:220253-38815419)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059876
GRCh38/hg38 9p24.3-13.1(chr9:204193-38815478)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000160921
GRCh38/hg38 9p24.3-13.1(chr9:203993-38815619)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059835
GRCh38/hg38 9p24.3-13.1(chr9:204193-38741440)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057406

GRCh38/hg38 9p24.3-q34.3(chr9:193412-138114463)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059875
GRCh38/hg38 9p24.3-q34.3(chr9:193412-138124532)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059874
GRCh38/hg38 9p24.3-q34.3(chr9:193412-138179445)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000160862
GRCh37/hg19 9p24.3-21.3(chr9:46587-22012051)x3	Pathogenic(Last reviewed: Aug 12, 2014)	VCV000611426
GRCh38/hg38 9p24.3-24.2(chr9:204193-4210335)x1	Pathogenic(Last reviewed: Aug 18, 2010)	VCV000144708
GRCh38/hg38 9p24.3-23(chr9:204104-14182668)x1	Pathogenic(Last reviewed: Aug 19, 2011)	VCV000148598
GRCh38/hg38 9p24.3-13.3(chr9:204104-34151476)x3	Pathogenic(Last reviewed: Aug 2, 2012)	VCV000150116
GRCh38/hg38 9p24.3-q21.11(chr9:204104-66233120)x4	Pathogenic(Last reviewed: Aug 2, 2012)	VCV000148823
GRCh38/hg38 9p24.3-q21.31(chr9:193412-79877816)x3	Pathogenic(Last reviewed: Aug 20, 2012)	VCV000154945
GRCh37/hg19 9p24.3-23(chr9:46587-13708607)x1	Pathogenic(Last reviewed: Aug 22, 2014)	VCV000611425
GRCh37/hg19 9p24.3-24.1(chr9:203861-7007586)x1	Pathogenic(Last reviewed: Aug 25, 2017)	VCV000563673
GRCh38/hg38 9p24.3-q34.3(chr9:193412-138179445)x3	Pathogenic(Last reviewed: Aug 27, 2010)	VCV000033205

GRCh37/hg19 9p24.3-23(chr9:203861-9306658)x1	Pathogenic(Last reviewed: Aug 28, 2017)	VCV000563675
GRCh37/hg19 9p24.3-q21.11(chr9:203861-70985795)x4	Pathogenic(Last reviewed: Aug 30, 2017)	VCV000563686
GRCh38/hg38 9p24.3-23(chr9:204104-11298187)x1	Pathogenic(Last reviewed: Aug 5, 2011)	VCV000146684
GRCh38/hg38 9p24.3-22.1(chr9:204104-18882281)x1	Pathogenic(Last reviewed: Aug 5, 2011)	VCV000146394
GRCh38/hg38 9p24.3-13.1(chr9:204193-38815478)x3	Pathogenic(Last reviewed: Aug 5, 2011)	VCV000032288
NM_003070.5(SMARCA2):c.2648C>T (p.Pro883Leu)	Pathogenic(Last reviewed: Aug 6, 2021)	VCV000030016
GRCh37/hg19 9p24.3-24.1(chr9:46587-5486856)x1	Pathogenic(Last reviewed: Dec 18, 2014)	VCV000611423
GRCh37/hg19 9p24.3-21.3(chr9:203861-20653468)x3	Pathogenic(Last reviewed: Dec 18, 2017)	VCV000563681
GRCh37/hg19 9p24.3-21.2(chr9:203861-26397133)x3	Pathogenic(Last reviewed: Dec 18, 2017)	VCV000563682
GRCh38/hg38 9p24.3-24.1(chr9:220253-7733826)x1	Pathogenic(Last reviewed: Dec 22, 2010)	VCV000154621
NM_003070.5(SMARCA2):c.3220C>G (p.Gln1074Glu)	Pathogenic(Last reviewed: Dec 23, 2019)	VCV000984920
NM_003070.5(SMARCA2):c.3439G>A (p.Asp1147Asn)	Pathogenic(Last reviewed: Dec 3, 2021)	VCV000279980

GRCh38/hg38 9p24.3-22.2(chr9:204193-18073359)x1	Pathogenic(Last reviewed: Dec 30, 2009)	VCV000144343
NM_003070.5(SMARCA2):c.3313C>T (p.Arg1105Cys)	Pathogenic(Last reviewed: Dec 7, 2020)	VCV000068769
GRCh38/hg38 9p24.3-22.1(chr9:220253-18708805)x1	Pathogenic(Last reviewed: Feb 18, 2011)	VCV000146356
NM_003070.5(SMARCA2):c.2815C>T (p.His939Tyr)	Pathogenic(Last reviewed: Feb 26, 2012)	VCV000030018
NM_003070.5(SMARCA2):c.3476G>T (p.Arg1159Leu)	Pathogenic(Last reviewed: Feb 26, 2012)	VCV000030015
NM_003070.5(SMARCA2):c.2642G>T (p.Gly881Val)	Pathogenic(Last reviewed: Feb 26, 2012)	VCV000030013
NM_003070.5(SMARCA2):c.3473A>T (p.Asp1158Val)	Pathogenic(Last reviewed: Feb 26, 2012)	VCV000030011
NM_003070.5(SMARCA2):c.3604G>T (p.Gly1202Cys)	Pathogenic(Last reviewed: Feb 26, 2012)	VCV000030009
NM_003070.5(SMARCA2):c.3637C>T (p.Arg1213Trp)	Pathogenic(Last reviewed: Feb 26, 2012)	VCV000030008
NM_003070.5(SMARCA2):c.2255G>C (p.Gly752Ala)	Pathogenic(Last reviewed: Feb 26, 2012)	VCV000030019
GRCh38/hg38 9p24.3-q34.3(chr9:204193-138179445)	Pathogenic(Last reviewed: Feb 28, 2010)	VCV000144309
GRCh37/hg19 9p24.3-23(chr9:203861-13486759)x1	Pathogenic(Last reviewed: Feb 29, 2016)	VCV000442671
GRCh37/hg19 9p24.3-q34.3(chr9:10590-141122247)x3	Pathogenic(Last reviewed: Jan 1, 2013)	VCV000611419

GRCh38/hg38 9p24.3-24.1(chr9:204193-6968724)x1	Pathogenic(Last reviewed: Jan 14, 2011)	VCV000146231
NM_003070.5(SMARCA2):c.3446A>G (p.Asn1149Ser)	Pathogenic(Last reviewed: Jan 20, 2016)	VCV000379917
GRCh37/hg19 9p24.3-23(chr9:13997-11376705)x1	Pathogenic(Last reviewed: Jan 20, 2016)	VCV000253580
GRCh37/hg19 9p24.3-22.1(chr9:213161-19450250)x3	Pathogenic(Last reviewed: Jan 20, 2016)	VCV000253670
GRCh37/hg19 9p24.3-13.1(chr9:213161-39092820)x3	Pathogenic(Last reviewed: Jan 20, 2016)	VCV000253667
GRCh37/hg19 9p24.3-13.1(chr9:163131-38763958)x3	Pathogenic(Last reviewed: Jan 20, 2016)	VCV000253633
GRCh37/hg19 9p24.3-11.2(chr9:213161-47212321)x4	Pathogenic(Last reviewed: Jan 20, 2016)	VCV000253592
GRCh37/hg19 9p24.3-q34.3(chr9:163131-141122114)x3	Pathogenic(Last reviewed: Jan 20, 2016)	VCV000253402
GRCh38/hg38 9p24.3-24.1(chr9:204104-5657733)x1	Pathogenic(Last reviewed: Jan 23, 2012)	VCV000149059
GRCh38/hg38 9p24.3-23(chr9:204104-10023901)x1	Pathogenic(Last reviewed: Jan 23, 2012)	VCV000149060
GRCh38/hg38 9p24.3-23(chr9:204193-10852686)x1	Pathogenic(Last reviewed: Jan 24, 2011)	VCV000146254
GRCh38/hg38 9p24.3-24.2(chr9:266045-3346702)x1	Pathogenic(Last reviewed: Jan 27, 2011)	VCV000154563

GRCh37/hg19 9p24.3-23(chr9:203861-14103730)x1	Pathogenic(Last reviewed: Jan 29, 2019)	VCV000815188
GRCh38/hg38 9p24.3-23(chr9:204193-10473327)x1	Pathogenic(Last reviewed: Jan 30, 2010)	VCV000144391
NM_003070.5(SMARCA2):c.3479C>G (p.Ala1160Gly)	Pathogenic(Last reviewed: Jan 30, 2022)	VCV001320261
GRCh37/hg19 9p24.3-q34.3(chr9:46587-141066491)x3	Pathogenic(Last reviewed: Jan 5, 2017)	VCV000611427
NM_003070.5(SMARCA2):c.3298A>C (p.Thr1100Pro)	Pathogenic(Last reviewed: Jan 5, 2022)	VCV001334325
GRCh38/hg38 9p24.3-q34.3(chr9:203861-138125937)x3	Pathogenic(Last reviewed: Jul 1, 2013)	VCV000153518
GRCh37/hg19 9p24.3-22.2(chr9:203861-17789410)x1	Pathogenic(Last reviewed: Jul 11, 2018)	VCV000815186
GRCh37/hg19 9p24.3-q34.3(chr9:203862-141020389)	Pathogenic(Last reviewed: Jul 14, 2015)	VCV000443986
NM_003070.5(SMARCA2):c.3313C>G (p.Arg1105Gly)	Pathogenic(Last reviewed: Jul 14, 2016)	VCV000280726
GRCh38/hg38 9p24.3-q22.1(chr9:203861-88130444)x4	Pathogenic(Last reviewed: Jul 16, 2013)	VCV000153561
GRCh38/hg38 9p24.3-24.1(chr9:203861-5094461)x1	Pathogenic(Last reviewed: Jul 18, 2014)	VCV000155570
NM_003070.5(SMARCA2):c.3386G>A (p.Gly1129Asp)	Pathogenic(Last reviewed: Jul 23, 2019)	VCV000390604

GRCh38/hg38 9p24.3-24.1(chr9:220253-8866675)x1	Pathogenic(Last reviewed: Jul 30, 2009)	VCV000146111
GRCh37/hg19 9p24.3-13.1(chr9:203861-38472979)x3	Pathogenic(Last reviewed: Jul 30, 2018)	VCV000687476
GRCh38/hg38 9p24.3-q21.11(chr9:204104-67549861)x3	Pathogenic(Last reviewed: Jul 5, 2011)	VCV000150341
GRCh38/hg38 9p11.2-q34.3(chr9:193412-138159073)x3	Pathogenic(Last reviewed: Jul 5, 2011)	VCV000150340
GRCh38/hg38 9p24.3-13.1(chr9:204104-38768294)x3	Pathogenic(Last reviewed: Jul 9, 2012)	VCV000150242
GRCh37/hg19 9p24.3-23(chr9:46587-12532584)x1	Pathogenic(Last reviewed: Jul 9, 2016)	VCV000611424
NM_003070.5(SMARCA2):c.2564G>A (p.Arg855Gln)	Pathogenic(Last reviewed: Jun 10, 2021)	VCV001177355
NM_003070.5(SMARCA2):c.2554G>C (p.Glu852Gln)	Pathogenic(Last reviewed: Jun 10, 2021)	VCV001177354
NM_003070.5(SMARCA2):c.3476G>A (p.Arg1159Gln)	Pathogenic(Last reviewed: Jun 10, 2021)	VCV000030010
GRCh37/hg19 9p24.3-q13(chr9:203861-67986965)x3	Pathogenic(Last reviewed: Jun 13, 2017)	VCV000685107
NM_003070.5(SMARCA2):c.3464A>C (p.Gln1155Pro)	Pathogenic(Last reviewed: Jun 18, 2018)	VCV000419659
GRCh38/hg38 9p24.3-q34.3(chr9:203862-138125937)x3	Pathogenic(Last reviewed: Jun 22, 2015)	VCV000155409

GRCh37/hg19 9p24.3-22.2(chr9:203861-17125893)x1	Pathogenic(Last reviewed: Jun 23, 2014)	VCV000441876
GRCh37/hg19 9p24.3-22.2(chr9:203861-17655298)x1	Pathogenic(Last reviewed: Jun 23, 2014)	VCV000442304
GRCh38/hg38 9p24.3-23(chr9:204090-13146846)x1	Pathogenic(Last reviewed: Jun 30, 2010)	VCV000144686
GRCh37/hg19 9p24.3-13.1(chr9:203861-38787480)x3	Pathogenic(Last reviewed: Mar 10, 2016)	VCV000443177
GRCh37/hg19 9p24.3-22.1(chr9:204193-18654812)	Pathogenic(Last reviewed: Mar 15, 2021)	VCV001047891
SMARCA2, 55-KB DEL	Pathogenic(Last reviewed: Mar 18, 2012)	VCV000030020
GRCh38/hg38 9p24.3-24.2(chr9:203861-4585050)x1	Pathogenic(Last reviewed: Mar 18, 2014)	VCV000154196
GRCh38/hg38 9p24.3-24.2(chr9:204104-3367760)x1	Pathogenic(Last reviewed: Mar 21, 2011)	VCV000148184
GRCh37/hg19 9p24.3-23(chr9:203861-9924905)x1	Pathogenic(Last reviewed: Mar 23, 2018)	VCV000563677
GRCh38/hg38 9p24.3-21.1(chr9:203861-31423873)x4	Pathogenic(Last reviewed: Mar 24, 2014)	VCV000153184
GRCh38/hg38 9p24.3-23(chr9:204193-11277770)x1	Pathogenic(Last reviewed: Mar 30, 2010)	VCV000144441
GRCh37/hg19 9p24.3-q13(chr9:203861-67983174)x4	Pathogenic(Last reviewed: Mar 8, 2018)	VCV000563684

GRCh37/hg19 9p24.3-q21.12(chr9:203861-72717793)x3	Pathogenic(Last reviewed: May 15, 2018)	VCV000563687
GRCh37/hg19 9p24.3-24.1(chr9:203861-5081516)x1	Pathogenic(Last reviewed: May 16, 2014)	VCV000442111
GRCh37/hg19 9p24.3-23(chr9:203861-11028975)x1	Pathogenic(Last reviewed: May 19, 2017)	VCV000687390
GRCh37/hg19 9p24.3-q21.11(chr9:203861-70984588)x3	Pathogenic(Last reviewed: May 20, 2019)	VCV000815190
NM_003070.5(SMARCA2):c.2554G>A (p.Glu852Lys)	Pathogenic(Last reviewed: May 24, 2017)	VCV000068762
GRCh37/hg19 9p24.3-24.2(chr9:204090-2430905)x1	Pathogenic(Last reviewed: May 28, 2020)	VCV000983185
GRCh38/hg38 9p24.3-21.1(chr9:220257-29424848)x3	Pathogenic(Last reviewed: May 30, 2010)	VCV000144555
GRCh38/hg38 9p24.3-q34.3(chr9:193412-138159073)x3	Pathogenic(Last reviewed: May 7, 2012)	VCV000150050
GRCh37/hg19 9p24.3-23(chr9:203861-11033228)x1	Pathogenic(Last reviewed: May 9, 2017)	VCV000687364
GRCh37/hg19 9p24.3-q34.3(chr9:10590-141107672)x3	Pathogenic(Last reviewed: Nov 1, 2010)	VCV000611417
GRCh37/hg19 9p24.3-q13(chr9:203861-68262804)x3,4	Pathogenic(Last reviewed: Nov 1, 2017)	VCV000563685
GRCh37/hg19 9p24.3-13.1(chr9:214309-39156958)	Pathogenic(Last reviewed: Nov 1, 2018)	VCV000625636

NM_003070.5(SMARCA2):c.2383T>C (p.Trp795Arg)	Pathogenic(Last reviewed: Nov 10, 2017)	VCV000981747
NM_003070.5(SMARCA2):c.3484C>T (p.Arg1162Cys)	Pathogenic(Last reviewed: Nov 12, 2021)	VCV000373431
GRCh37/hg19 9p24.3-24.2(chr9:203861-2978707)x1	Pathogenic(Last reviewed: Nov 13, 2017)	VCV000689070
GRCh37/hg19 9p24.3-13.1(chr9:203861-38787480)x3	Pathogenic(Last reviewed: Nov 2, 2018)	VCV000563683
GRCh37/hg19 9p24.3-q34.3(chr9:10590-141114095)x2	Pathogenic(Last reviewed: Nov 23, 2011)	VCV000611418
NM_003070.5(SMARCA2):c.2561A>T (p.His854Leu)	Pathogenic(Last reviewed: Nov 23, 2016)	VCV000981748
GRCh38/hg38 9p24.3-22.2(chr9:204193-16897580)x1	Pathogenic(Last reviewed: Nov 30, 2009)	VCV000146719
GRCh38/hg38 9p24.3-q21.13(chr9:193412-74615913)x3	Pathogenic(Last reviewed: Nov 30, 2009)	VCV000146703
GRCh38/hg38 9p24.3-q21.11(chr9:13997-68401065)x3	Pathogenic(Last reviewed: Nov 30, 2010)	VCV000146018
GRCh38/hg38 9p24.3-24.2(chr9:204104-3755031)x1	Pathogenic(Last reviewed: Nov 4, 2011)	VCV000148849
NM_003070.5(SMARCA2):c.3475C>G (p.Arg1159Gly)	Pathogenic(Last reviewed: Nov 4, 2019)	VCV000030012
GRCh37/hg19 9p24.3-23(chr9:203861-10666419)x1	Pathogenic(Last reviewed: Oct 1, 2018)	VCV000815187

GRCh37/hg19 9p24.3-23(chr9:203861-11271239)x1	Pathogenic(Last reviewed: Oct 11, 2017)	VCV000563678
NM_003070.5(SMARCA2):c.3602C>T (p.Ala1201Val)	Pathogenic(Last reviewed: Oct 12, 2021)	VCV000030017
NM_003070.5(SMARCA2):c.2486C>T (p.Thr829Ile)	Pathogenic(Last reviewed: Oct 13, 2017)	VCV000212225
GRCh38/hg38 9p24.3-22.3(chr9:214367-16307944)x1	Pathogenic(Last reviewed: Oct 14, 2010)	VCV000147703
GRCh38/hg38 9p24.3-24.1(chr9:204104-6322471)x1	Pathogenic(Last reviewed: Oct 15, 2012)	VCV000151712
GRCh38/hg38 9p24.3-22.3(chr9:322690-16401656)x1	Pathogenic(Last reviewed: Oct 15, 2012)	VCV000152943
GRCh38/hg38 9p24.3-q21.12(chr9:193412-70630731)x3	Pathogenic(Last reviewed: Oct 20, 2010)	VCV000146931
NM_003070.5(SMARCA2):c.3593T>G (p.Val1198Gly)	Pathogenic(Last reviewed: Oct 21, 2016)	VCV000521357
GRCh37/hg19 9p24.3-22.1(chr9:203861-19448473)x3	Pathogenic(Last reviewed: Sep 10, 2018)	VCV000684956
GRCh38/hg38 9p24.3-24.1(chr9:204104-7133443)x1	Pathogenic(Last reviewed: Sep 12, 2011)	VCV000148677
NM_003070.5(SMARCA2):c.1573C>T (p.Arg525Cys)	Pathogenic(Last reviewed: Sep 15, 2021)	VCV000829811
GRCh38/hg38 9p24.3-23(chr9:204193-11435662)x1	Pathogenic(Last reviewed: Sep 16, 2011)	VCV000147839
NM_003070.5(SMARCA2):c.3314G>A (p.Arg1105His)	Pathogenic(Last reviewed: Sep 2, 2020)	VCV000982859

NM_003070.5(SMARCA2):c.1585C>G (p.Leu529Val)	Pathogenic(Last reviewed: Sep 2, 2020)	VCV000829810
GRCh37/hg19 9p24.3-24.1(chr9:203861-4959039)x1	Pathogenic(Last reviewed: Sep 20, 2017)	VCV000563670
GRCh38/hg38 9p24.3-24.1(chr9:204090-4970154)x3	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000154749
GRCh38/hg38 9p24.3-23(chr9:204090-9282864)x1	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000150751
GRCh38/hg38 9p24.3-22.3(chr9:204090-15260600)x1	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000154897
GRCh38/hg38 9p24.3-21.3(chr9:459131-24207894)x3	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000149484
GRCh38/hg38 9p24.3-q34.3(chr9:193412-138124524)x3	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000149828
GRCh38/hg38 9p24.3-24.1(chr9:203861-8172957)x1	Pathogenic(Last reviewed: Sep 23, 2013)	VCV000153821
GRCh37/hg19 9p24.3-q34.3(chr9:203861-141020388)x3	Pathogenic(Last reviewed: Sep 28, 2018)	VCV000685192
GRCh38/hg38 9p24.3-21.2(chr9:204104-27963369)x3	Pathogenic(Last reviewed: Sep 4, 2012)	VCV000150819
GRCh37/hg19 9p24.3-22.3(chr9:203861-14744606)x1	Pathogenic(Last reviewed: Sep 6, 2017)	VCV000563679
NM_003070.5(SMARCA2):c.2853G>C (p.Lys951Asn)	Pathogenic(Last reviewed: Sep 7, 2016)	VCV000436805

NM_003070.5(SMARCA2):c.2810G>T (p.Arg937Leu)	Pathogenic/Likely pathogenic(Last reviewed: Apr 30, 2021)	VCV000827773
NM_003070.5(SMARCA2):c.3485G>A (p.Arg1162His)	Pathogenic/Likely pathogenic(Last reviewed: Sep 1, 2018)	VCV000030014
NM_003070.5(SMARCA2):c.2156T>C (p.Leu719Pro)	Uncertain significance	VCV000827775
NM_003070.5(SMARCA2):c.2795T>C (p.Ile932Thr)	Uncertain significance	VCV000827828
NM_003070.5(SMARCA2):c.3040A>G (p.Lys1014Glu)	Uncertain significance	VCV000827772
NM_003070.5(SMARCA2):c.4666A>G (p.Lys1556Glu)	Uncertain significance	VCV000813686
NM_003070.5(SMARCA2):c.1450C>A (p.His484Asn)	Uncertain significance	VCV000374228
NM_003070.5(SMARCA2):c.1279C>T (p.Arg427Cys)	Uncertain significance(Last reviewed: Apr 1, 2020)	VCV000916365
NM_003070.5(SMARCA2):c.2425G>C (p.Ala809Pro)	Uncertain significance(Last reviewed: Apr 1, 2021)	VCV001176896
NM_003070.5(SMARCA2):c.182G>C (p.Gly61Ala)	Uncertain significance(Last reviewed: Apr 10, 2019)	VCV000449800
NM_003070.5(SMARCA2):c.4370G>A (p.Arg1457His)	Uncertain significance(Last reviewed: Apr 12, 2019)	VCV001305007
NM_003070.5(SMARCA2):c.3412G>C (p.Ala1138Pro)	Uncertain significance(Last reviewed: Apr 14, 2018)	VCV000985482
NM_003070.5(SMARCA2):c.3519G>T (p.Arg1173Ser)	Uncertain significance(Last reviewed: Apr 16, 2019)	VCV001305100
NM_003070.5(SMARCA2):c.3615C>G (p.Asp1205Glu)	Uncertain significance(Last reviewed: Apr 16, 2019)	VCV001305099

NM_003070.5(SMARCA2):c.3940G >A (p.Val1314Met)	Uncertain significance(Last reviewed: Apr 16, 2021)	VCV001314521
GRCh37/hg19 9p24.3-24.2(chr9:2180509-3128422)x3	Uncertain significance(Last reviewed: Apr 28, 2017)	VCV000686138
NM_003070.5(SMARCA2):c.4262G >A (p.Arg1421Gln)	Uncertain significance(Last reviewed: Apr 30, 2020)	VCV001047931
GRCh38/hg38 9p24.3-24.2(chr9:1845513-3022547)x3	Uncertain significance(Last reviewed: Apr 4, 2013)	VCV000152115
NM_003070.5(SMARCA2):c.344A> T (p.Gln115Leu)	Uncertain significance(Last reviewed: Apr 5, 2019)	VCV001308470
Single allele	Uncertain significance(Last reviewed: Apr 9, 2019)	VCV000635963
NM_003070.5(SMARCA2):c.3379A >G (p.Arg1127Gly)	Uncertain significance(Last reviewed: Aug 1, 2017)	VCV000810346
NM_003070.5(SMARCA2):c.4738-7C>T	Uncertain significance(Last reviewed: Aug 1, 2019)	VCV000872500
NM_003070.5(SMARCA2):c.4103G >A (p.Arg1368Lys)	Uncertain significance(Last reviewed: Aug 1, 2019)	VCV000872499
GRCh38/hg38 9p24.3(chr9:839152-2094920)x3	Uncertain significance(Last reviewed: Aug 12, 2011)	VCV000057235
GRCh38/hg38 9p24.3-24.2(chr9:1998911-2925112)x3	Uncertain significance(Last reviewed: Aug 12, 2011)	VCV000058462
NM_003070.5(SMARCA2):c.3265C >T (p.Arg1089Trp)	Uncertain significance(Last reviewed: Aug 7, 2019)	VCV001307583

NM_003070.5(SMARCA2):c.919G>A (p.Val307Met)	Uncertain significance(Last reviewed: Dec 11, 2019)	VCV000931617
NM_003070.5(SMARCA2):c.1534G>A (p.Glu512Lys)	Uncertain significance(Last reviewed: Dec 11, 2020)	VCV000992210
NM_003070.5(SMARCA2):c.3448C>G (p.Pro1150Ala)	Uncertain significance(Last reviewed: Dec 13, 2019)	VCV001029502
NM_003070.5(SMARCA2):c.31C>A (p.Pro11Thr)	Uncertain significance(Last reviewed: Dec 18, 2017)	VCV000522971
NM_003070.5(SMARCA2):c.497A>G (p.Gln166Arg)	Uncertain significance(Last reviewed: Dec 20, 2019)	VCV001311425
NM_003070.5(SMARCA2):c.3389G>A (p.Gly1130Asp)	Uncertain significance(Last reviewed: Dec 30, 2017)	VCV000623322
NM_003070.5(SMARCA2):c.209T>A (p.Met70Lys)	Uncertain significance(Last reviewed: Dec 31, 2019)	VCV001311613
NM_003070.5(SMARCA2):c.583G>C (p.Gly195Arg)	Uncertain significance(Last reviewed: Dec 6, 2019)	VCV001311059
NM_003070.5(SMARCA2):c.3638G>C (p.Arg1213Pro)	Uncertain significance(Last reviewed: Feb 1, 2018)	VCV000495133
NM_003070.5(SMARCA2):c.482C>T (p.Pro161Leu)	Uncertain significance(Last reviewed: Feb 14, 2014)	VCV000126350
NM_003070.5(SMARCA2):c.669GCA[5] (p.Gln231_Gln238del)	Uncertain significance(Last reviewed: Feb 2, 2017)	VCV000436801
NM_003070.5(SMARCA2):c.4486C>A (p.Gln1496Lys)	Uncertain significance(Last reviewed: Feb 3, 2017)	VCV000521478

NM_003070.5(SMARCA2):c.50C>T (p.Pro17Leu)	Uncertain significance(Last reviewed: Feb 4, 2022)	VCV001186796
NM_003070.5(SMARCA2):c.1211C >T (p.Thr404Met)	Uncertain significance(Last reviewed: Feb 5, 2021)	VCV001314105
NM_003070.5(SMARCA2):c.*587T> C	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000915202
NM_003070.5(SMARCA2):c.2070C >G (p.Ser690Arg)	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000914955
NM_003070.5(SMARCA2):c.2037- 4C>A	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000914954
NM_003070.5(SMARCA2):c.3981+ 11C>G	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000914487
NM_003070.5(SMARCA2):c.*489A> G	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000913964
NM_003070.5(SMARCA2):c.*80T> C	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000913578
NM_003070.5(SMARCA2):c.890C> A (p.Pro297Gln)	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000913315
NM_003070.5(SMARCA2):c.2527- 3T>C	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000913001
NM_003070.5(SMARCA2):c.699G> A (p.Gln233=)	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000912951
NM_003070.5(SMARCA2):c.*876G >C	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000912496

NM_003070.5(SMARCA2):c.*537T>C	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000366338
NM_003070.5(SMARCA2):c.*197A>G	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000366334
NM_003070.5(SMARCA2):c.*5T>C	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000366328
NM_003070.5(SMARCA2):c.2933A>T (p.Tyr978Phe)	Uncertain significance(Last reviewed: Jan 12, 2018)	VCV000366215
NM_003070.5(SMARCA2):c.*670A>G	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000915203
NM_003070.5(SMARCA2):c.4673C>G (p.Pro1558Arg)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000915172
NM_003070.5(SMARCA2):c.1812G>A (p.Lys604=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000914439
NM_003070.5(SMARCA2):c.1599C>T (p.Thr533=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000914437
NM_003070.5(SMARCA2):c.399C>T (p.His133=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000914396
NM_003070.5(SMARCA2):c.324T>A (p.Pro108=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000914395
NM_003070.5(SMARCA2):c.105A>G (p.Pro35=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000914394
NM_003070.5(SMARCA2):c.*230T>C	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913962

NM_003070.5(SMARCA2):c.*204G >A	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913961
NM_003070.5(SMARCA2):c.*192G >C	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913960
NM_003070.5(SMARCA2):c.*101A> C	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913579
NM_003070.5(SMARCA2):c.3768G >A (p.Met1256Ile)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913366
NM_003070.5(SMARCA2):c.3633C >T (p.His1211=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913365
NM_003070.5(SMARCA2):c.3216C >T (p.Phe1072=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913364
NM_003070.5(SMARCA2):c.990C> T (p.Pro330=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913318
NM_003070.5(SMARCA2):c.961C> T (p.Leu321=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000913317
NM_003070.5(SMARCA2):c.693G> A (p.Gln231=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000912950
NM_003070.5(SMARCA2):c.*782G >A	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000912495
NM_003070.5(SMARCA2):c.*765T> G	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000912494
NM_003070.5(SMARCA2):c.4737+ 12C>T	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000912461

NM_003070.5(SMARCA2):c.*694T>G	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000366340
NM_003070.5(SMARCA2):c.*16T>C	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000366330
NM_003070.5(SMARCA2):c.4738-9T>C	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000366326
NM_003070.5(SMARCA2):c.4701G>A (p.Val1567=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000366323
NM_003070.5(SMARCA2):c.4499A>C (p.Lys1500Thr)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000366250
NM_003070.5(SMARCA2):c.3165T>C (p.Leu1055=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000366218
NM_003070.5(SMARCA2):c.2452C>T (p.Leu818=)	Uncertain significance(Last reviewed: Jan 13, 2018)	VCV000366212
NM_003070.5(SMARCA2):c.4765G>C (p.Asp1589His)	Uncertain significance(Last reviewed: Jan 24, 2018)	VCV001032871
NM_003070.5(SMARCA2):c.1931A>C (p.Glu644Ala)	Uncertain significance(Last reviewed: Jan 27, 2020)	VCV001314988
NM_003070.5(SMARCA2):c.1436C>T (p.Ala479Val)	Uncertain significance(Last reviewed: Jan 28, 2020)	VCV001315067
NM_003070.5(SMARCA2):c.4364G>C (p.Arg1455Thr)	Uncertain significance(Last reviewed: Jul 1, 2016)	VCV000374420
NM_003070.5(SMARCA2):c.4595-11T>A	Uncertain significance(Last reviewed: Jul 1, 2019)	VCV001306417

NM_003070.5(SMARCA2):c.4212G >T (p.Glu1404Asp)	Uncertain significance(Last reviewed: Jul 10, 2019)	VCV001304869
NM_003070.5(SMARCA2):c.2932T >C (p.Tyr978His)	Uncertain significance(Last reviewed: Jul 19, 2019)	VCV001304600
NM_003070.5(SMARCA2):c.274G> A (p.Gly92Arg)	Uncertain significance(Last reviewed: Jul 21, 2018)	VCV001032870
NM_003070.5(SMARCA2):c.1625A >G (p.Asn542Ser)	Uncertain significance(Last reviewed: Jul 22, 2019)	VCV001211033
NM_003070.5(SMARCA2):c.2852A >G (p.Lys951Arg)	Uncertain significance(Last reviewed: Jul 27, 2018)	VCV000976071
NM_003070.5(SMARCA2):c.4696G >A (p.Val1566Ile)	Uncertain significance(Last reviewed: Jul 27, 2020)	VCV001029503
NM_003070.5(SMARCA2):c.1333C >T (p.Arg445Cys)	Uncertain significance(Last reviewed: Jul 29, 2021)	VCV001319926
NM_003070.5(SMARCA2):c.4414A >C (p.Met1472Leu)	Uncertain significance(Last reviewed: Jul 7, 2017)	VCV000976210
NM_003070.5(SMARCA2):c.4275A >G (p.Glu1425=)	Uncertain significance(Last reviewed: Jun 1, 2019)	VCV000810348
NM_003070.5(SMARCA2):c.2794A >T (p.Ile932Leu)	Uncertain significance(Last reviewed: Jun 10, 2020)	VCV000430530
NM_003070.5(SMARCA2):c.2267C >T (p.Thr756Ile)	Uncertain significance(Last reviewed: Jun 10, 2021)	VCV000068760
NM_003070.5(SMARCA2):c.2878G >A (p.Glu960Lys)	Uncertain significance(Last reviewed: Jun 11, 2019)	VCV001303289

NM_003070.5(SMARCA2):c.508G>C (p.Gly170Arg)	Uncertain significance(Last reviewed: Jun 11, 2021)	VCV001327907
NM_003070.5(SMARCA2):c.2329C>G (p.Leu777Val)	Uncertain significance(Last reviewed: Jun 17, 2014)	VCV000377393
NM_003070.5(SMARCA2):c.3252T>A (p.Asp1084Glu)	Uncertain significance(Last reviewed: Jun 19, 2019)	VCV000976144
NM_003070.5(SMARCA2):c.4164C>T (p.Asn1388=)	Uncertain significance(Last reviewed: Jun 25, 2015)	VCV000212228
NM_003070.5(SMARCA2):c.1174-10T>C	Uncertain significance(Last reviewed: Jun 25, 2019)	VCV001337253
NM_003070.5(SMARCA2):c.4157_4166del (p.Gln1386fs)	Uncertain significance(Last reviewed: Jun 27, 2019)	VCV001306761
NC_000009.11:g.(?_2029023)_(5300444_?)dup	Uncertain significance(Last reviewed: Jun 30, 2020)	VCV001042801
NM_003070.5(SMARCA2):c.795G>A (p.Pro265=)	Uncertain significance(Last reviewed: Mar 17, 2016)	VCV000436798
NM_003070.5(SMARCA2):c.*307C>T	Uncertain significance(Last reviewed: Mar 2, 2018)	VCV000913963
NM_003070.5(SMARCA2):c.4696G>C (p.Val1566Leu)	Uncertain significance(Last reviewed: Mar 2, 2021)	VCV001342463
NM_003070.5(SMARCA2):c.915C>G (p.Pro305=)	Uncertain significance(Last reviewed: Mar 24, 2016)	VCV000436802
NM_003070.5(SMARCA2):c.2348+8A>C	Uncertain significance(Last reviewed: Mar 26, 2021)	VCV001342334

NM_003070.5(SMARCA2):c.-56G>A	Uncertain significance(Last reviewed: Mar 30, 2018)	VCV000913279
NM_003070.5(SMARCA2):c.4508G>A (p.Arg1503Gln)	Uncertain significance(Last reviewed: May 1, 2018)	VCV000810349
NM_003070.5(SMARCA2):c.880G>A (p.Ala294Thr)	Uncertain significance(Last reviewed: May 12, 2020)	VCV001098643
NM_003070.5(SMARCA2):c.887A>C (p.Gln296Pro)	Uncertain significance(Last reviewed: May 18, 2017)	VCV000588983
NM_003070.5(SMARCA2):c.869C>T (p.Ala290Val)	Uncertain significance(Last reviewed: May 24, 2019)	VCV001302166
NM_003070.5(SMARCA2):c.2082T>G (p.Ser694Arg)	Uncertain significance(Last reviewed: May 26, 2021)	VCV001326395
NM_003070.5(SMARCA2):c.929C>A (p.Pro310Gln)	Uncertain significance(Last reviewed: May 28, 2019)	VCV000802454
NM_003070.5(SMARCA2):c.3796C>G (p.Arg1266Gly)	Uncertain significance(Last reviewed: May 31, 2017)	VCV000589167
NM_003070.5(SMARCA2):c.1232A>G (p.Asn411Ser)	Uncertain significance(Last reviewed: Nov 15, 2019)	VCV001187925
NM_003070.5(SMARCA2):c.2032A>G (p.Ile678Val)	Uncertain significance(Last reviewed: Nov 18, 2020)	VCV001333848
NM_003070.5(SMARCA2):c.1226C>G (p.Ala409Gly)	Uncertain significance(Last reviewed: Nov 21, 2019)	VCV001310623
GRCh37/hg19 9p24.3(chr9:2130392-2185324)x1	Uncertain significance(Last reviewed: Nov 3, 2017)	VCV000563563

NM_003070.5(SMARCA2):c.1325G >A (p.Arg442Lys)	Uncertain significance(Last reviewed: Oct 1, 2018)	VCV000810343
NM_003070.5(SMARCA2):c.1240G >A (p.Ala414Thr)	Uncertain significance(Last reviewed: Oct 10, 2019)	VCV000451242
NM_003070.5(SMARCA2):c.3982- 4A>T	Uncertain significance(Last reviewed: Oct 13, 2020)	VCV001304125
NM_003070.5(SMARCA2):c.1296G >C (p.Leu432=)	Uncertain significance(Last reviewed: Oct 31, 2014)	VCV000212223
NM_003070.5(SMARCA2):c.1064C >G (p.Ala355Gly)	Uncertain significance(Last reviewed: Oct 4, 2019)	VCV001308897
NM_003070.5(SMARCA2):c.3754de l (p.Leu1252fs)	Uncertain significance(Last reviewed: Sep 10, 2020)	VCV000981424
NM_003070.5(SMARCA2):c.1458C >G (p.Asn486Lys)	Uncertain significance(Last reviewed: Sep 10, 2020)	VCV000827829
NM_003070.5(SMARCA2):c.334C> T (p.Pro112Ser)	Uncertain significance(Last reviewed: Sep 12, 2019)	VCV001308195
NM_003070.5(SMARCA2):c.1586T >G (p.Leu529Arg)	Uncertain significance(Last reviewed: Sep 13, 2017)	VCV000522058
NM_003070.5(SMARCA2):c.3385G >C (p.Gly1129Arg)	Uncertain significance(Last reviewed: Sep 23, 2014)	VCV000212226
NM_003070.5(SMARCA2):c.2420C >T (p.Thr807Ile)	Uncertain significance(Last reviewed: Sep 8, 2017)	VCV000451779

Appendix Table 5. Table of reported variants for *KAT6A,* retrieved from ClinVar databases (https://www.ncbi.nlm.nih.gov/clinvar/?term=KDM2B%5Bgene%5D&redir=gene)

Name	Clinical significance (Last reviewed)	Accession
NM_006766.5(KAT6A):c.3392G>A (p.Arg1131His)	Benign(Last reviewed: Apr 12, 2021)	VCV001221162
NM_006766.5(KAT6A):c.1364-9T>A	Benign(Last reviewed: Apr 26, 2021)	VCV001271402
NM_006766.5(KAT6A):c.2533C>T (p.Arg845Cys)	Benign(Last reviewed: Apr 5, 2019)	VCV001263118
NM_006766.5(KAT6A):c.3660C>T (p.Pro1220=)	Benign(Last reviewed: Apr 8, 2021)	VCV000730577
NM_006766.5(KAT6A):c.1996+62G>A	Benign(Last reviewed: Aug 10, 2019)	VCV001274664
NM_006766.5(KAT6A):c.1996+98A>G	Benign(Last reviewed: Aug 10, 2019)	VCV001272501
NM_006766.5(KAT6A):c.1996+90_1996+99del	Benign(Last reviewed: Aug 10, 2019)	VCV001270393
NM_006766.5(KAT6A):c.1598+263dup	Benign(Last reviewed: Aug 10, 2019)	VCV001178308
NM_006766.5(KAT6A):c.1996+100A>G	Benign(Last reviewed: Aug 11, 2019)	VCV001277202
NM_006766.5(KAT6A):c.1996+61_1996+62insTA	Benign(Last reviewed: Aug 13, 2019)	VCV001289176

NM_006766.5(KAT6A):c.1996+88_1996+99del	Benign(Last reviewed: Aug 13, 2019)	VCV001277059
NM_006766.5(KAT6A):c.1740+294ATG[11]	Benign(Last reviewed: Aug 13, 2019)	VCV001267606
NM_006766.5(KAT6A):c.1740+294ATG[10]	Benign(Last reviewed: Aug 20, 2019)	VCV001258284
NM_006766.5(KAT6A):c.1996+94_1996+101del	Benign(Last reviewed: Aug 21, 2019)	VCV001263963
NM_006766.5(KAT6A):c.1599-8G>A	Benign(Last reviewed: Aug 24, 2020)	VCV000786120
NM_006766.5(KAT6A):c.1044-57A>G	Benign(Last reviewed: Aug 3, 2018)	VCV001250240
NM_006766.5(KAT6A):c.1740+294ATG[8]	Benign(Last reviewed: Aug 6, 2019)	VCV001232892
NM_006766.5(KAT6A):c.1740+294ATG[12]	Benign(Last reviewed: Aug 7, 2019)	VCV001291825
NM_006766.5(KAT6A):c.4707C>T (p.Tyr1569=)	Benign(Last reviewed: Dec 15, 2020)	VCV001252501
NM_006766.5(KAT6A):c.4149C>T (p.Ser1383=)	Benign(Last reviewed: Dec 19, 2019)	VCV001272650
NM_006766.5(KAT6A):c.1599-9C>T	Benign(Last reviewed: Dec 31, 2019)	VCV000770571
NM_006766.5(KAT6A):c.3353-31C>A	Benign(Last reviewed: Feb 10, 2021)	VCV001278127
NM_006766.5(KAT6A):c.1364-11T>C	Benign(Last reviewed: Feb 27, 2020)	VCV001281888
NM_006766.5(KAT6A):c.5036C>T (p.Pro1679Leu)	Benign(Last reviewed: Feb 6, 2020)	VCV001248835
NM_006766.5(KAT6A):c.2672	Benign(Last reviewed: Feb 7, 2021)	VCV001285736

C>T (p.Thr891Met)		
NM_006766.5(KAT6A):c.3352 +20C>T	Benign(Last reviewed: Feb 9, 2021)	VCV001223126
NM_006766.5(KAT6A):c.2869 G>T (p.Ala957Ser)	Benign(Last reviewed: Jan 11, 2018)	VCV000731757
NM_006766.5(KAT6A):c.1996 +49T>C	Benign(Last reviewed: Jan 13, 2021)	VCV001244907
NM_006766.5(KAT6A):c.3764 C>G (p.Ala1255Gly)	Benign(Last reviewed: Jan 14, 2020)	VCV001281676
NM_006766.5(KAT6A):c.2437 -318T>C	Benign(Last reviewed: Jul 17, 2018)	VCV001290657
NM_006766.5(KAT6A):c.1902 +96T>G	Benign(Last reviewed: Jul 17, 2018)	VCV001267389
NM_006766.5(KAT6A):c.1043 +233C>T	Benign(Last reviewed: Jul 17, 2018)	VCV001259092
NM_006766.5(KAT6A):c.*185 C>G	Benign(Last reviewed: Jul 17, 2018)	VCV001247574
NM_006766.5(KAT6A):c.-325- 129A>G	Benign(Last reviewed: Jul 17, 2018)	VCV001183258
NM_006766.5(KAT6A):c.2982 G>A (p.Pro994=)	Benign(Last reviewed: Jul 17, 2018)	VCV000587818
NM_006766.5(KAT6A):c.401T >C (p.Leu134Ser)	Benign(Last reviewed: Jul 20, 2018)	VCV000587821
NM_006766.5(KAT6A):c.4455 C>T (p.Ser1485=)	Benign(Last reviewed: Jul 20, 2018)	VCV000587817
NM_006766.5(KAT6A):c.1445 T>A (p.Met482Lys)	Benign(Last reviewed: Jul 26, 2019)	VCV001178084

NM_006766.5(KAT6A):c.3039+158C>A	Benign(Last reviewed: Jul 3, 2018)	VCV001276014
NM_006766.5(KAT6A):c.-199T>A	Benign(Last reviewed: Jul 3, 2018)	VCV001245822
NM_006766.5(KAT6A):c.1902+20G>A	Benign(Last reviewed: Jul 3, 2018)	VCV001242874
NM_006766.5(KAT6A):c.826-58C>T	Benign(Last reviewed: Jul 3, 2018)	VCV001182953
NM_006766.5(KAT6A):c.4914C>T (p.Cys1638=)	Benign(Last reviewed: Jul 3, 2018)	VCV000587956
NM_006766.5(KAT6A):c.4872C>G (p.Val1624=)	Benign(Last reviewed: Jul 3, 2018)	VCV000587916
NM_006766.5(KAT6A):c.3192G>A (p.Thr1064=)	Benign(Last reviewed: Jul 5, 2018)	VCV000587815
NM_006766.5(KAT6A):c.2226C>T (p.Asp742=)	Benign(Last reviewed: Mar 11, 2020)	VCV001258419
NM_006766.5(KAT6A):c.600+5G>T	Benign(Last reviewed: Mar 13, 2020)	VCV000716531
NM_006766.5(KAT6A):c.1996+48C>G	Benign(Last reviewed: Mar 25, 2020)	VCV001177720
NM_006766.5(KAT6A):c.5241A>G (p.Pro1747=)	Benign(Last reviewed: May 29, 2020)	VCV001262345
NM_006766.5(KAT6A):c.5379G>A (p.Gln1793=)	Benign(Last reviewed: Nov 19, 2021)	VCV000588207
NM_006766.5(KAT6A):c.4394A>G (p.Asp1465Gly)	Benign(Last reviewed: Oct 1, 2021)	VCV001300277
NM_006766.5(KAT6A):c.710-14C>G	Benign(Last reviewed: Oct 1, 2021)	VCV001300266

NM_006766.5(KAT6A):c.2694A>G (p.Gln898=)	Benign(Last reviewed: Oct 21, 2018)	VCV000587932
NM_006766.5(KAT6A):c.5597C>G (p.Ser1866Cys)	Benign(Last reviewed: Oct 21, 2019)	VCV001287491
NM_006766.5(KAT6A):c.1482+6G>A	Benign(Last reviewed: Oct 23, 2020)	VCV001222349
NM_006766.5(KAT6A):c.1902+268G>C	Benign(Last reviewed: Sep 22, 2018)	VCV001224567
NM_006766.5(KAT6A):c.3576C>T (p.Ile1192=)	Benign(Last reviewed: Sep 28, 2020)	VCV001273565
NM_006766.5(KAT6A):c.600+283T>A	Benign(Last reviewed: Sep 30, 2019)	VCV001286870
NM_006766.5(KAT6A):c.2627A>C (p.Gln876Pro)	Benign(Last reviewed: Sep 6, 2021)	VCV001269461
NM_006766.5(KAT6A):c.3440A>T (p.Lys1147Ile)	Benign(Last reviewed: Sep 6, 2021)	VCV001049560
NM_006766.5(KAT6A):c.2487A>G (p.Val829=)	Benign(Last reviewed: Sep 8, 2021)	VCV000587953
NM_006766.5(KAT6A):c.2373A>C (p.Glu791Asp)	Benign/Likely benign(Last reviewed: Apr 1, 2019)	VCV000588282
NM_006766.5(KAT6A):c.4779G>A (p.Ser1593=)	Benign/Likely benign(Last reviewed: Apr 1, 2021)	VCV000721663
NM_006766.5(KAT6A):c.3561C>T (p.Cys1187=)	Benign/Likely benign(Last reviewed: Apr 11, 2019)	VCV000588050
NM_006766.5(KAT6A):c.4146G>A (p.Thr1382=)	Benign/Likely benign(Last reviewed: Apr 16, 2020)	VCV000588065

NM_006766.5(KAT6A):c.3242C>G (p.Pro1081Arg)	Benign/Likely benign(Last reviewed: Apr 19, 2021)	VCV000719038
NM_006766.5(KAT6A):c.4940AGC[5] (p.Gln1650dup)	Benign/Likely benign(Last reviewed: Aug 10, 2020)	VCV000588016
NM_006766.5(KAT6A):c.3321AGA[2] (p.Glu1109del)	Benign/Likely benign(Last reviewed: Aug 4, 2020)	VCV000588354
NM_006766.5(KAT6A):c.3577G>A (p.Val1193Ile)	Benign/Likely benign(Last reviewed: Dec 10, 2020)	VCV000587853
NM_006766.5(KAT6A):c.5259A>G (p.Leu1753=)	Benign/Likely benign(Last reviewed: Dec 2, 2020)	VCV000589656
NM_006766.5(KAT6A):c.5526G>A (p.Thr1842=)	Benign/Likely benign(Last reviewed: Dec 23, 2020)	VCV000588011
NM_006766.5(KAT6A):c.5994C>T (p.Asn1998=)	Benign/Likely benign(Last reviewed: Dec 27, 2019)	VCV000588777
NM_006766.5(KAT6A):c.1477C>T (p.Leu493=)	Benign/Likely benign(Last reviewed: Dec 31, 2019)	VCV000588369
NM_006766.5(KAT6A):c.4956A>G (p.Pro1652=)	Benign/Likely benign(Last reviewed: Jul 18, 2019)	VCV000589069
NM_006766.5(KAT6A):c.1134A>G (p.Ser378=)	Benign/Likely benign(Last reviewed: Jul 22, 2021)	VCV000589524
NM_006766.5(KAT6A):c.4953_4976del (p.Pro1652_Pro1659del)	Benign/Likely benign(Last reviewed: Jul 3, 2019)	VCV000588080
NM_006766.5(KAT6A):c.4611C>T (p.Ser1537=)	Benign/Likely benign(Last reviewed: Jul 8, 2020)	VCV000589286

NM_006766.5(KAT6A):c.5913C>T (p.Asn1971=)	Benign/Likely benign(Last reviewed: Jun 10, 2019)	VCV000588835
NM_006766.5(KAT6A):c.648A>G (p.Lys216=)	Benign/Likely benign(Last reviewed: Jun 4, 2021)	VCV000589011
NM_006766.5(KAT6A):c.1157G>A (p.Arg386Gln)	Benign/Likely benign(Last reviewed: May 1, 2021)	VCV000588950
NM_006766.5(KAT6A):c.1029G>T (p.Lys343Asn)	Benign/Likely benign(Last reviewed: May 18, 2020)	VCV000726877
NM_006766.5(KAT6A):c.5040_5051del (p.1677_1680QQPQ[1])	Benign/Likely benign(Last reviewed: May 28, 2019)	VCV000589281
NM_006766.5(KAT6A):c.4952C>T (p.Pro1651Leu)	Benign/Likely benign(Last reviewed: May 28, 2019)	VCV000445624
NM_006766.5(KAT6A):c.1741-10A>G	Benign/Likely benign(Last reviewed: May 9, 2021)	VCV000711395
NM_006766.5(KAT6A):c.3010A>G (p.Ile1004Val)	Benign/Likely benign(Last reviewed: Nov 8, 2020)	VCV000589512
NM_006766.5(KAT6A):c.1185T>G (p.Asp395Glu)	Benign/Likely benign(Last reviewed: Oct 1, 2019)	VCV000588308
NM_006766.5(KAT6A):c.5967T>C (p.Ala1989=)	Benign/Likely benign(Last reviewed: Oct 25, 2020)	VCV000588426
NM_006766.5(KAT6A):c.4324G>A (p.Ala1442Thr)	Benign/Likely benign(Last reviewed: Oct 27, 2020)	VCV000445863
NM_006766.5(KAT6A):c.3505C>T (p.Arg1169Ter)	Conflicting interpretations of pathogenicity(Last reviewed: Apr 20, 2020)	VCV000280873
NM_006766.5(KAT6A):c.5740	Conflicting interpretations of pathogenicity(Last reviewed: Feb 10, 2020)	VCV000445654

293

A>G (p.Met1914Val)		
NM_006766.5(KAT6A):c.3855 G>T (p.Gln1285His)	Conflicting interpretations of pathogenicity(Last reviewed: Feb 7, 2021)	VCV000588899
NM_006766.5(KAT6A):c.893C >T (p.Thr298Ile)	Conflicting interpretations of pathogenicity(Last reviewed: Jan 1, 2019)	VCV000634490
NM_006766.5(KAT6A):c.3859 GAG[2] (p.Glu1289del)	Conflicting interpretations of pathogenicity(Last reviewed: Nov 10, 2020)	VCV000377124
NM_006766.5(KAT6A):c.4968 _4982dup (p.Gln1657_Pro 1661dup)	Conflicting interpretations of pathogenicity(Last reviewed: Oct 1, 2020)	VCV000587892
NM_006766.5(KAT6A):c.1342 A>G (p.Ser448Gly)	Conflicting interpretations of pathogenicity(Last reviewed: Oct 11, 2017)	VCV000497669
NM_006766.5(KAT6A):c.5525 C>G (p.Thr1842Arg)	Likely benign	VCV000265767
NM_006766.5(KAT6A):c.3802 C>T (p.Pro1268Ser)	Likely benign	VCV001206121
NM_006766.5(KAT6A):c.1997 -6C>T	Likely benign(Last reviewed: Apr 1, 2018)	VCV000624322
NM_006766.5(KAT6A):c.709+ 330A>C	Likely benign(Last reviewed: Apr 10, 2019)	VCV001198545
NM_006766.5(KAT6A):c.4982 C>T (p.Pro1661Leu)	Likely benign(Last reviewed: Apr 13, 2021)	VCV001300918
NM_006766.5(KAT6A):c.5601 G>A (p.Ala1867=)	Likely benign(Last reviewed: Apr 2, 2021)	VCV001213003
NM_006766.5(KAT6A):c.1065 A>T (p.Lys355Asn)	Likely benign(Last reviewed: Apr 6, 2021)	VCV001300971

NM_006766.5(KAT6A):c.4722 C>T (p.Gly1574=)	Likely benign(Last reviewed: Apr 8, 2020)	VCV001199128
NM_006766.5(KAT6A):c.3039 +30C>T	Likely benign(Last reviewed: Aug 10, 2018)	VCV001204582
NM_006766.5(KAT6A):c.1996 +61_1996+62in sTGTA	Likely benign(Last reviewed: Aug 10, 2019)	VCV001188463
NM_006766.5(KAT6A):c.1996 +61_1996+62in sTGTGTGTA	Likely benign(Last reviewed: Aug 13, 2019)	VCV001194085
NM_006766.5(KAT6A):c.1996 +50GT[23]	Likely benign(Last reviewed: Aug 15, 2019)	VCV001215549
NM_006766.5(KAT6A):c.1996 +50G>C	Likely benign(Last reviewed: Aug 15, 2020)	VCV001187219
NM_006766.5(KAT6A):c.1996 +94_1996+99d el	Likely benign(Last reviewed: Aug 21, 2019)	VCV001191182
NM_006766.5(KAT6A):c.4962 G>A (p.Pro1654=)	Likely benign(Last reviewed: Aug 22, 2017)	VCV000776322
NM_006766.5(KAT6A):c.1996 +92_1996+101 del	Likely benign(Last reviewed: Aug 23, 2019)	VCV001211380
NM_006766.5(KAT6A):c.4768 G>A (p.Gly1590Ser)	Likely benign(Last reviewed: Aug 25, 2018)	VCV000588852
NM_006766.5(KAT6A):c.3742 G>A (p.Glu1248Lys)	Likely benign(Last reviewed: Aug 3, 2016)	VCV000377054
NM_006766.5(KAT6A):c.1996 +50GT[22]	Likely benign(Last reviewed: Aug 7, 2019)	VCV001204751
NM_006766.5(KAT6A):c.603G >A (p.Pro201=)	Likely benign(Last reviewed: Dec 1, 2020)	VCV000589173
NM_006766.5(KAT6A):c.1741 -182A>C	Likely benign(Last reviewed: Dec 12, 2018)	VCV001188190

NM_006766.5(KAT6A):c.1996+8T>A	Likely benign(Last reviewed: Dec 19, 2017)	VCV000730438
NM_006766.5(KAT6A):c.2517T>C (p.Ala839=)	Likely benign(Last reviewed: Dec 2, 2020)	VCV001201486
NM_006766.5(KAT6A):c.5586G>A (p.Ala1862=)	Likely benign(Last reviewed: Dec 22, 2020)	VCV001207709
NM_006766.5(KAT6A):c.5601G>T (p.Ala1867=)	Likely benign(Last reviewed: Dec 31, 2019)	VCV000719150
NM_006766.5(KAT6A):c.4188C>G (p.His1396Gln)	Likely benign(Last reviewed: Dec 8, 2020)	VCV001186435
NM_006766.5(KAT6A):c.2864C>T (p.Pro955Leu)	Likely benign(Last reviewed: Feb 1, 2021)	VCV000710644
NM_006766.5(KAT6A):c.5730T>C (p.Asn1910=)	Likely benign(Last reviewed: Feb 18, 2021)	VCV000719218
NM_006766.5(KAT6A):c.4977_4985dup (p.Gln1660_Pro1662dup)	Likely benign(Last reviewed: Feb 2, 2018)	VCV000724518
NM_006766.5(KAT6A):c.4182C>T (p.Asp1394=)	Likely benign(Last reviewed: Feb 21, 2020)	VCV001194144
NM_006766.5(KAT6A):c.4767C>T (p.Tyr1589=)	Likely benign(Last reviewed: Feb 22, 2017)	VCV000588810
NM_006766.5(KAT6A):c.4222G>A (p.Glu1408Lys)	Likely benign(Last reviewed: Feb 26, 2021)	VCV001254829
NM_006766.5(KAT6A):c.5027A>C (p.Gln1676Pro)	Likely benign(Last reviewed: Jan 1, 2019)	VCV000975258
NM_006766.5(KAT6A):c.1662	Likely benign(Last reviewed: Jan 1, 2019)	VCV000975257

G>T (p.Gln554His)		
NM_006766.5(KAT6A):c.5572 C>T (p.Arg1858Cys)	Likely benign(Last reviewed: Jan 1, 2019)	VCV000975256
NM_006766.5(KAT6A):c.2710 G>A (p.Glu904Lys)	Likely benign(Last reviewed: Jan 1, 2019)	VCV000975255
NM_006766.5(KAT6A):c.5666 G>A (p.Arg1889His)	Likely benign(Last reviewed: Jan 12, 2018)	VCV000734321
NM_006766.5(KAT6A):c.958C >G (p.Leu320Val)	Likely benign(Last reviewed: Jan 12, 2022)	VCV001254827
NM_006766.5(KAT6A):c.3040 -8T>C	Likely benign(Last reviewed: Jan 13, 2021)	VCV001187477
NM_006766.5(KAT6A):c.1903 -206C>T	Likely benign(Last reviewed: Jan 28, 2019)	VCV001217129
NM_006766.5(KAT6A):c.1043 T>G (p.Val348Gly)	Likely benign(Last reviewed: Jan 31, 2018)	VCV000589247
NM_006766.5(KAT6A):c.1599 -10T>C	Likely benign(Last reviewed: Jan 5, 2021)	VCV001329611
NM_006766.5(KAT6A):c.5185 A>T (p.Ile1729Leu)	Likely benign(Last reviewed: Jul 14, 2020)	VCV001198558
NM_006766.5(KAT6A):c.4503 C>T (p.Asn1501=)	Likely benign(Last reviewed: Jul 26, 2016)	VCV000588093
NM_006766.5(KAT6A):c.4512 C>T (p.Ala1504=)	Likely benign(Last reviewed: Jun 1, 2021)	VCV001176883
NM_006766.5(KAT6A):c.4119 C>T (p.Ser1373=)	Likely benign(Last reviewed: Jun 4, 2021)	VCV001327333
NM_006766.5(KAT6A):c.5494	Likely benign(Last reviewed: Mar 1, 2019)	VCV001209314

A>G (p.Met1832Val)		
NM_006766.5(KAT6A):c.5019 A>G (p.Pro1673=)	Likely benign(Last reviewed: Mar 21, 2017)	VCV000587796
NM_006766.5(KAT6A):c.- 17A>G	Likely benign(Last reviewed: Mar 31, 2019)	VCV001206802
NM_006766.5(KAT6A):c.2983 G>A (p.Glu995Lys)	Likely benign(Last reviewed: Mar 5, 2018)	VCV000548600
NM_006766.5(KAT6A):c.2760 G>C (p.Leu920=)	Likely benign(Last reviewed: Mar 9, 2017)	VCV000589558
NM_006766.5(KAT6A):c.2505 A>G (p.Pro835=)	Likely benign(Last reviewed: May 12, 2021)	VCV001321557
NM_006766.5(KAT6A):c.4988 C>T (p.Pro1663Leu)	Likely benign(Last reviewed: May 14, 2018)	VCV000589048
NM_006766.5(KAT6A):c.4150 G>A (p.Val1384Met)	Likely benign(Last reviewed: May 26, 2021)	VCV001197725
NM_006766.5(KAT6A):c.2633 G>A (p.Arg878His)	Likely benign(Last reviewed: May 27, 2020)	VCV001198352
NM_006766.5(KAT6A):c.5776 C>G (p.Arg1926Gly)	Likely benign(Last reviewed: May 28, 2019)	VCV000802401
NM_006766.5(KAT6A):c.2241 C>A (p.Ile747=)	Likely benign(Last reviewed: May 5, 2021)	VCV001321530
NM_006766.5(KAT6A):c.4560 C>T (p.Ser1520=)	Likely benign(Last reviewed: Nov 1, 2021)	VCV001335709
NM_006766.5(KAT6A):c.5668 G>A (p.Ala1890Thr)	Likely benign(Last reviewed: Nov 15, 2017)	VCV000725886

NM_006766.5(KAT6A):c.3833G>A (p.Arg1278His)	Likely benign(Last reviewed: Nov 24, 2020)	VCV001218506
NM_006766.5(KAT6A):c.1363+65T>C	Likely benign(Last reviewed: Nov 25, 2018)	VCV001199783
NM_006766.5(KAT6A):c.5683C>T (p.Arg1895Cys)	Likely benign(Last reviewed: Nov 25, 2020)	VCV001207257
NM_006766.5(KAT6A):c.5549C>T (p.Pro1850Leu)	Likely benign(Last reviewed: Nov 30, 2017)	VCV000787620
NM_006766.5(KAT6A):c.3040-83A>G	Likely benign(Last reviewed: Nov 5, 2018)	VCV001205810
NM_006766.5(KAT6A):c.3936C>T (p.Asp1312=)	Likely benign(Last reviewed: Nov 6, 2020)	VCV001193761
NM_006766.5(KAT6A):c.3040-9A>G	Likely benign(Last reviewed: Oct 13, 2020)	VCV001211647
NM_006766.5(KAT6A):c.1010G>A (p.Arg337His)	Likely benign(Last reviewed: Oct 15, 2020)	VCV001254687
NM_006766.5(KAT6A):c.1996+278T>G	Likely benign(Last reviewed: Oct 16, 2018)	VCV001219902
NM_006766.5(KAT6A):c.1482+238dup	Likely benign(Last reviewed: Oct 17, 2018)	VCV001208341
NM_006766.5(KAT6A):c.2437-267A>G	Likely benign(Last reviewed: Oct 17, 2018)	VCV001195547
NM_006766.5(KAT6A):c.1044-280C>T	Likely benign(Last reviewed: Oct 5, 2018)	VCV001195795
NM_006766.5(KAT6A):c.3225T>C (p.Asp1075=)	Likely benign(Last reviewed: Oct 5, 2020)	VCV001208736
NM_006766.5(KAT6A):c.3264G>T (p.Leu1088Phe)	Likely benign(Last reviewed: Oct 8, 2020)	VCV001223761

NM_006766.5(KAT6A):c.1996+90_1996+101del	Likely benign(Last reviewed: Oct 9, 2019)	VCV001191665
NM_006766.5(KAT6A):c.2986A>G (p.Ser996Gly)	Likely benign(Last reviewed: Sep 17, 2019)	VCV001189464
NM_006766.5(KAT6A):c.1044-180G>T	Likely benign(Last reviewed: Sep 19, 2019)	VCV001204832
NM_006766.5(KAT6A):c.336G>C (p.Glu112Asp)	Likely benign(Last reviewed: Sep 2, 2019)	VCV001204805
NM_006766.5(KAT6A):c.1996+50GT[25]	Likely benign(Last reviewed: Sep 2, 2019)	VCV001199080
NM_006766.5(KAT6A):c.5101C>T (p.Pro1701Ser)	Likely benign(Last reviewed: Sep 23, 2020)	VCV000984622
NM_006766.5(KAT6A):c.3256C>T (p.Arg1086Cys)	Likely benign(Last reviewed: Sep 24, 2020)	VCV001198511
NM_006766.5(KAT6A):c.825+53G>C	Likely benign(Last reviewed: Sep 26, 2018)	VCV001216193
NM_006766.5(KAT6A):c.826-236_826-232del	Likely benign(Last reviewed: Sep 26, 2018)	VCV001201893
NM_006766.5(KAT6A):c.-192G>C	Likely benign(Last reviewed: Sep 26, 2018)	VCV001195781
NM_006766.5(KAT6A):c.710-215G>A	Likely benign(Last reviewed: Sep 26, 2018)	VCV001193899
NM_006766.5(KAT6A):c.601-183C>T	Likely benign(Last reviewed: Sep 26, 2018)	VCV001191337
NM_006766.5(KAT6A):c.2965GAG[4] (p.Glu993del)	Likely benign(Last reviewed: Sep 29, 2021)	VCV001300574
NM_006766.5(KAT6A):c.5644G>A (p.Val1882Ile)	Likely benign(Last reviewed: Sep 3, 2019)	VCV001186257

NM_006766.5(KAT6A):c.5019ACC[3](p.Pro1675dup)	Likely benign(Last reviewed: Sep 8, 2021)	VCV000589573
NM_006766.5(KAT6A):c.3921_3922del(p.Glu1307fs)	Likely pathogenic	VCV001334428
NM_006766.5(KAT6A):c.1582C>T(p.Pro528Ser)	Likely pathogenic	VCV000996682
NM_006766.5(KAT6A):c.2007del(p.Leu669fs)	Likely pathogenic(Last reviewed: Apr 1, 2018)	VCV000624321
NM_006766.5(KAT6A):c.4688_4689del(p.Asn1562_Tyr1563insTer)	Likely pathogenic(Last reviewed: Apr 7, 2015)	VCV000559923
NM_006766.5(KAT6A):c.296G>A(p.Trp99Ter)	Likely pathogenic(Last reviewed: Aug 1, 2017)	VCV000493478
NM_006766.5(KAT6A):c.4399C>T(p.Gln1467Ter)	Likely pathogenic(Last reviewed: Aug 7, 2019)	VCV000422237
NM_006766.5(KAT6A):c.4861C>T(p.Gln1621Ter)	Likely pathogenic(Last reviewed: Dec 9, 2021)	VCV001334703
NM_006766.5(KAT6A):c.4070del(p.Gln1357fs)	Likely pathogenic(Last reviewed: Jan 1, 2019)	VCV000982773
NM_006766.5(KAT6A):c.4666A>T(p.Ile1556Phe)	Likely pathogenic(Last reviewed: Jan 1, 2019)	VCV000975406
NM_006766.5(KAT6A):c.5123del(p.Asn1708fs)	Likely pathogenic(Last reviewed: Jan 3, 2022)	VCV001333695
NM_006766.5(KAT6A):c.2437-3C>G	Likely pathogenic(Last reviewed: Jan 9, 2017)	VCV000392057
NM_006766.5(KAT6A):c.3596	Likely pathogenic(Last reviewed: Jul 11, 2017)	VCV000559924

del (p.Gly1199fs)		
NM_006766.5(KAT6A):c.4037 del (p.Gly1346fs)	Likely pathogenic(Last reviewed: Jul 18, 2018)	VCV000817161
NM_006766.5(KAT6A):c.3040 -1G>T	Likely pathogenic(Last reviewed: Jun 17, 2016)	VCV000985644
NM_006766.5(KAT6A):c.5701 del (p.Val1901fs)	Likely pathogenic(Last reviewed: Jun 3, 2021)	VCV001321924
NM_006766.5(KAT6A):c.3039 +1G>T	Likely pathogenic(Last reviewed: Mar 1, 2017)	VCV000444748
NM_006766.5(KAT6A):c.5248 _5257del (p.Ala1749_Thr 1750insTer)	Likely pathogenic(Last reviewed: Mar 20, 2018)	VCV000523923
NM_006766.5(KAT6A):c.3184 G>T (p.Glu1062Ter)	Likely pathogenic(Last reviewed: Mar 20, 2018)	VCV000523839
NM_006766.5(KAT6A):c.4085 del (p.Lys1362fs)	Likely pathogenic(Last reviewed: May 1, 2021)	VCV001176884
NM_006766.5(KAT6A):c.4664 G>A (p.Ser1555Asn)	Likely pathogenic(Last reviewed: May 26, 2020)	VCV000976788
NM_006766.5(KAT6A):c.5645 _5646delinsGC TGGCCGTA (p.Val1882fs)	Likely pathogenic(Last reviewed: May 26, 2020)	VCV000817979
NM_006766.5(KAT6A):c.1280 G>T (p.Arg427Leu)	Likely pathogenic(Last reviewed: Nov 19, 2015)	VCV000430275
NM_006766.5(KAT6A):c.3039 +1del	Likely pathogenic(Last reviewed: Oct 19, 2017)	VCV000453041
NM_006766.5(KAT6A):c.2437 -1G>A	Likely pathogenic(Last reviewed: Oct 23, 2020)	VCV000986888

NM_006766.5(KAT6A):c.1907_1908del (p.Lys636fs)	Likely pathogenic(Last reviewed: Oct 5, 2020)	VCV000988723
NM_006766.5(KAT6A):c.3212A>C (p.Glu1071Ala)	Likely pathogenic(Last reviewed: Sep 13, 2015)	VCV000429660
NM_006766.5(KAT6A):c.4819C>T (p.Gln1607Ter)	Likely pathogenic(Last reviewed: Sep 19, 2017)	VCV000451851
NM_006766.5(KAT6A):c.4348_4349del (p.Leu1450fs)	Likely pathogenic(Last reviewed: Sep 25, 2020)	VCV000988752
NM_006766.5(KAT6A):c.5736T>G (p.Asn1912Lys)	not provided	VCV000818162
NM_006766.5(KAT6A):c.1506del (p.Asp503fs)	Pathogenic	VCV001172592
GRCh37/hg19 8p11.21-11.1(chr8:40690198-43388233)x3	Pathogenic	VCV000394062
GRCh37/hg19 8p23.1-11.1(chr8:12580132-43388233)x3	Pathogenic	VCV000394647
GRCh37/hg19 8p23.3-11.1(chr8:158048-43786723)x3	Pathogenic	VCV000397117
GRCh37/hg19 8p23.3-q24.3(chr8:158991-146280828)x3	Pathogenic	VCV000394884
NM_006766.5(KAT6A):c.3039+1G>A	Pathogenic(Last reviewed: Apr 10, 2020)	VCV000984972
NM_006766.5(KAT6A):c.4398_4399del (p.Gln1467fs)	Pathogenic(Last reviewed: Apr 14, 2020)	VCV001323134

NM_006766.5(KAT6A):c.1786_1787del (p.Leu596fs)	Pathogenic(Last reviewed: Apr 20, 2020)	VCV000978867
NM_006766.5(KAT6A):c.4069C>T (p.Gln1357Ter)	Pathogenic(Last reviewed: Apr 20, 2020)	VCV000978866
NM_006766.5(KAT6A):c.3703G>T (p.Glu1235Ter)	Pathogenic(Last reviewed: Apr 20, 2020)	VCV000978865
NM_006766.5(KAT6A):c.1569C>A (p.Tyr523Ter)	Pathogenic(Last reviewed: Apr 20, 2020)	VCV000978864
NM_006766.5(KAT6A):c.3432dup (p.Pro1145fs)	Pathogenic(Last reviewed: Apr 25, 2016)	VCV000280545
NM_006766.5(KAT6A):c.2127del (p.Lys709fs)	Pathogenic(Last reviewed: Apr 25, 2017)	VCV000985477
NM_006766.5(KAT6A):c.4945C>T (p.Gln1649Ter)	Pathogenic(Last reviewed: Apr 3, 2020)	VCV000984649
GRCh38/hg38 8p11.21-q11.21(chr8:41845699-47893948)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059786
GRCh38/hg38 8p11.21-11.1(chr8:39981424-43532444)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059784
GRCh38/hg38 8p11.23-11.21(chr8:37899430-42371734)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059782
GRCh38/hg38 8p11.22-q11.21(chr8:39830633-49209461)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059783

GRCh38/hg38 8p12-11.21(chr8:343 12250-43158901)x1	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057114
GRCh38/hg38 8p12-q11.21(chr8:29 719897-48521849)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057312
GRCh38/hg38 8p21.2-q11.21(chr8:25 832130-48521849)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057237
GRCh38/hg38 8p23.1-11.21(chr8:126 09975-42085703)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000161033
GRCh38/hg38 8p22-q11.21(chr8:14 940110-47929925)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059769
GRCh38/hg38 8p23.1-11.1(chr8:1272 8904-43673207)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000161019
GRCh38/hg38 8p23.1-11.1(chr8:1275 0796-43532444)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059766
GRCh38/hg38 8p23.1-11.1(chr8:1260 9975-43336172)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000059764
GRCh38/hg38 8p23.3-q24.3(chr8:241 530-145049449)x3	Pathogenic(Last reviewed: Aug 12, 2011)	VCV000057496
NM_006766.5(KAT6A):c.442C >T (p.Arg148Ter)	Pathogenic(Last reviewed: Aug 13, 2018)	VCV000589472

GRCh37/hg19 8p23.1-q24.3(chr8:124 90999-146295771)x3	Pathogenic(Last reviewed: Aug 15, 2014)	VCV000443619
GRCh38/hg38 8p23.1-11.1(chr8:1264 6123-43686843)x3	Pathogenic(Last reviewed: Aug 2, 2011)	VCV000151014
GRCh38/hg38 8p23.3-q24.3(chr8:208 048-145070385)x3	Pathogenic(Last reviewed: Aug 2, 2013)	VCV000153413
NM_006766.5(KAT6A):c.1085 _1089dup (p.Arg364fs)	Pathogenic(Last reviewed: Aug 21, 2017)	VCV000451609
NM_006766.5(KAT6A):c.4361 dup (p.Thr1455fs)	Pathogenic(Last reviewed: Aug 22, 2015)	VCV000419571
GRCh38/hg38 8p21.3-q24.3(chr8:212 91522-145070385)x3	Pathogenic(Last reviewed: Aug 26, 2013)	VCV000153737
NM_006766.5(KAT6A):c.2463 del (p.Asn821fs)	Pathogenic(Last reviewed: Aug 29, 2018)	VCV001034308
GRCh38/hg38 8p23.3-q24.3(chr8:226 452-145068712)x3	Pathogenic(Last reviewed: Dec 10, 2012)	VCV000154791
NM_006766.5(KAT6A):c.4228 _4232del (p.Lys1410fs)	Pathogenic(Last reviewed: Dec 13, 2018)	VCV000524135
GRCh38/hg38 8p11.21-11.1(chr8:3996 0531-43673207)x3	Pathogenic(Last reviewed: Dec 22, 2010)	VCV000154596
GRCh38/hg38 8p23.1-11.1(chr8:1272 8904-43673207)x3	Pathogenic(Last reviewed: Dec 22, 2010)	VCV000032941

306

GRCh38/hg38 8p23.1-11.1(chr8:12609975-43673207)x3	Pathogenic(Last reviewed: Dec 22, 2010)	VCV000146507
GRCh38/hg38 8p23.1-11.1(chr8:12182421-43673207)x3	Pathogenic(Last reviewed: Dec 22, 2010)	VCV000148174
NM_006766.5(KAT6A):c.1176del (p.Cys393fs)	Pathogenic(Last reviewed: Feb 27, 2018)	VCV000504378
GRCh37/hg19 8p23.3-q24.3(chr8:10213-146293414)x3	Pathogenic(Last reviewed: Jan 1, 2013)	VCV000610612
NM_006766.5(KAT6A):c.3034C>T (p.Arg1012Ter)	Pathogenic(Last reviewed: Jan 1, 2019)	VCV000975405
NM_006766.5(KAT6A):c.1663del (p.Gln555fs)	Pathogenic(Last reviewed: Jan 1, 2019)	VCV000975259
NM_006766.5(KAT6A):c.3385C>T (p.Arg1129Ter)	Pathogenic(Last reviewed: Jan 1, 2019)	VCV000180229
NM_006766.5(KAT6A):c.4025del (p.Lys1342fs)	Pathogenic(Last reviewed: Jan 10, 2018)	VCV000419562
NM_006766.5(KAT6A):c.3306del (p.Lys1103fs)	Pathogenic(Last reviewed: Jan 16, 2020)	VCV000807616
NM_006766.5(KAT6A):c.805C>T (p.Arg269Ter)	Pathogenic(Last reviewed: Jan 23, 2019)	VCV000620554
GRCh38/hg38 8p23.1-11.1(chr8:12383584-43673207)x3	Pathogenic(Last reviewed: Jan 24, 2011)	VCV000146257
NM_006766.5(KAT6A):c.1473	Pathogenic(Last reviewed: Jan 4, 2019)	VCV000817975

_1476dup (p.Leu493fs)		
GRCh37/hg19 8p23.3-q24.3(chr8:164984-146293414)x3	Pathogenic(Last reviewed: Jan 5, 2017)	VCV000610618
NM_006766.5(KAT6A):c.5566del (p.Ser1856fs)	Pathogenic(Last reviewed: Jan 7, 2019)	VCV000817973
NM_006766.5(KAT6A):c.5546_5555del (p.Met1849fs)	Pathogenic(Last reviewed: Jan 8, 2016)	VCV000280240
NM_006766.5(KAT6A):c.856C>T (p.Arg286Ter)	Pathogenic(Last reviewed: Jul 13, 2018)	VCV000598767
NM_006766.5(KAT6A):c.4381C>T (p.Gln1461Ter)	Pathogenic(Last reviewed: Jul 2, 2020)	VCV000391696
GRCh37/hg19 8p12-q24.3(chr8:31936551-146295771)x3	Pathogenic(Last reviewed: Jul 25, 2018)	VCV000687493
NM_006766.5(KAT6A):c.3318_3319insCT (p.Glu1107fs)	Pathogenic(Last reviewed: Jun 11, 2019)	VCV000807615
NM_006766.5(KAT6A):c.1096C>T (p.Arg366Ter)	Pathogenic(Last reviewed: Jun 18, 2018)	VCV000620229
NM_006766.5(KAT6A):c.3182T>G (p.Leu1061Ter)	Pathogenic(Last reviewed: Jun 21, 2016)	VCV000280659
NM_006766.5(KAT6A):c.3928C>T (p.Gln1310Ter)	Pathogenic(Last reviewed: Jun 28, 2016)	VCV000985176
NM_006766.5(KAT6A):c.5639C>A (p.Ser1880Ter)	Pathogenic(Last reviewed: Jun 29, 2017)	VCV000489019

NM_006766.5(KAT6A):c.3353-3_3353dup	Pathogenic(Last reviewed: Jun 30, 2017)	VCV000450319
NM_006766.5(KAT6A):c.1903-5_1903-2del	Pathogenic(Last reviewed: Mar 1, 2019)	VCV000692079
NM_006766.5(KAT6A):c.3443del (p.Lys1148fs)	Pathogenic(Last reviewed: Mar 22, 2018)	VCV000559638
NM_006766.5(KAT6A):c.3553C>T (p.Gln1185Ter)	Pathogenic(Last reviewed: Mar 29, 2016)	VCV000280462
NM_006766.5(KAT6A):c.1146_1147insG (p.Tyr383fs)	Pathogenic(Last reviewed: Mar 29, 2021)	VCV001048764
NM_006766.5(KAT6A):c.3692del (p.Ala1231fs)	Pathogenic(Last reviewed: Mar 30, 2020)	VCV000985843
NM_006766.5(KAT6A):c.3879dup (p.Glu1294fs)	Pathogenic(Last reviewed: Mar 5, 2015)	VCV000180678
GRCh37/hg19 8p23.3-q24.3(chr8:158049-146295771)x3	Pathogenic(Last reviewed: Mar 5, 2015)	VCV000442201
NM_006766.5:c.3411del	Pathogenic(Last reviewed: May 14, 2021)	VCV001082502
NM_006766.5(KAT6A):c.1928A>G (p.Asn643Ser)	Pathogenic(Last reviewed: May 14, 2021)	VCV000626907
NM_006766.5(KAT6A):c.1900A>T (p.Lys634Ter)	Pathogenic(Last reviewed: May 18, 2017)	VCV000985898
NM_006766.5(KAT6A):c.3780del (p.Pro1261fs)	Pathogenic(Last reviewed: May 2, 2018)	VCV000545910
NM_006766.5(KAT6A):c.3782del (p.Pro1261fs)	Pathogenic(Last reviewed: May 2, 2018)	VCV000524127

NM_006766.5(KAT6A):c.3395_3396del (p.Asp1132fs)	Pathogenic(Last reviewed: May 23, 2016)	VCV000985124
GRCh37/hg19 8p11.22-q12.3(chr8:39555657-64049089)x3	Pathogenic(Last reviewed: May 26, 2017)	VCV000685558
GRCh38/hg38 8p23.1-11.21(chr8:12609975-42085703)x3	Pathogenic(Last reviewed: May 27, 2010)	VCV000034276
NM_006766.5(KAT6A):c.1536dup (p.Glu513Ter)	Pathogenic(Last reviewed: May 27, 2016)	VCV000280641
GRCh37/hg19 8p23.1-11.1(chr8:11935023-43824035)x3	Pathogenic(Last reviewed: May 5, 2014)	VCV000443194
NM_006766.5(KAT6A):c.3116_3117del (p.Ile1038_Ser1039insTer)	Pathogenic(Last reviewed: May 5, 2016)	VCV000162616
NM_006766.5(KAT6A):c.1312C>T (p.Arg438Ter)	Pathogenic(Last reviewed: Nov 21, 2021)	VCV001325410
NM_006766.5(KAT6A):c.4256_4260dup (p.Asp1421fs)	Pathogenic(Last reviewed: Nov 23, 2018)	VCV000817328
NM_006766.5(KAT6A):c.1405C>T (p.Arg469Ter)	Pathogenic(Last reviewed: Nov 23, 2021)	VCV001285569
NM_006766.5(KAT6A):c.4210dup (p.Glu1404fs)	Pathogenic(Last reviewed: Nov 24, 2016)	VCV000373672
NM_006766.5(KAT6A):c.4254_4257del (p.Glu1419fs)	Pathogenic(Last reviewed: Nov 30, 2015)	VCV000369686
GRCh37/hg19 8p11.21(chr8:4	Pathogenic(Last reviewed: Nov 30, 2017)	VCV000687703

1761813-42107108)x1		
NM_006766.5(KAT6A):c.3661G>T (p.Glu1221Ter)	Pathogenic(Last reviewed: Nov 7, 2019)	VCV000280246
GRCh37/hg19 8p23.3-11.1(chr8:176814-43396776)	Pathogenic(Last reviewed: Oct 1, 2020)	VCV000997076
NM_006766.5(KAT6A):c.3055C>T (p.Arg1019Ter)	Pathogenic(Last reviewed: Oct 12, 2019)	VCV000489088
NM_006766.5(KAT6A):c.4372_4373del (p.Ser1458fs)	Pathogenic(Last reviewed: Oct 15, 2019)	VCV000817759
GRCh38/hg38 8p23.3-q24.3(chr8:241530-145054634)x3	Pathogenic(Last reviewed: Oct 19, 2010)	VCV000059738
NM_006766.5(KAT6A):c.3973G>T (p.Glu1325Ter)	Pathogenic(Last reviewed: Oct 23, 2020)	VCV000987501
NM_006766.5(KAT6A):c.1465_1471del (p.Gln489fs)	Pathogenic(Last reviewed: Oct 23, 2020)	VCV000987373
NM_006766.5(KAT6A):c.3286dup (p.Cys1096fs)	Pathogenic(Last reviewed: Oct 9, 2016)	VCV000985670
NM_006766.5(KAT6A):c.3412del (p.Glu1139fs)	Pathogenic(Last reviewed: Oct 9, 2017)	VCV000452579
NM_006766.5(KAT6A):c.907+1del	Pathogenic(Last reviewed: Sep 18, 2017)	VCV000620010
GRCh37/hg19 8p23.3-q24.3(chr8:158048-146295771)x3	Pathogenic(Last reviewed: Sep 18, 2018)	VCV000687787

GRCh38/hg38 8p12-q12.1(chr8:365 80103-59618998)x3	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000150770
GRCh38/hg38 8p23.3-q24.3(chr8:241 605-145054781)x3	Pathogenic(Last reviewed: Sep 21, 2012)	VCV000149660
NM_006766.5(KAT6A):c.658C >T (p.Arg220Ter)	Pathogenic(Last reviewed: Sep 5, 2018)	VCV000620356
NM_006766.5(KAT6A):c.931C >T (p.Arg311Ter)	Pathogenic(Last reviewed: Sep 6, 2019)	VCV000419207
GRCh37/hg19 8p23.3-q24.3(chr8:158 049-146295771)	Pathogenic(Last reviewed: Sep 9, 2015)	VCV000442200
NM_006766.5(KAT6A):c.3070 C>T (p.Arg1024Ter)	Pathogenic(Last reviewed: Sep 9, 2021)	VCV000180230
NM_006766.5(KAT6A):c.3353 -1G>A	Pathogenic/Likely pathogenic(Last reviewed: Apr 26, 2021)	VCV000449582
NM_006766.5(KAT6A):c.2218 del (p.Arg740fs)	Pathogenic/Likely pathogenic(Last reviewed: Feb 5, 2020)	VCV000828153
NM_006766.5(KAT6A):c.4645 G>A (p.Gly1549Ser)	Pathogenic/Likely pathogenic(Last reviewed: Jan 3, 2022)	VCV000419627
NM_006766.5(KAT6A):c.4038 del (p.Val1347fs)	Pathogenic/Likely pathogenic(Last reviewed: Jul 18, 2018)	VCV000617516
NM_006766.5(KAT6A):c.949C >T (p.Arg317Ter)	Pathogenic/Likely pathogenic(Last reviewed: Oct 23, 2020)	VCV000489323
GRCh37/hg19 8p21.2-q12.1(chr8:247 72064-24813176)x3	Uncertain significance	VCV000393993

NM_006766.5(KAT6A):c.4804 A>G (p.Ser1602Gly)	Uncertain significance(Last reviewed: Apr 12, 2021)	VCV001303206
NM_006766.5(KAT6A):c.601- 3C>T	Uncertain significance(Last reviewed: Apr 15, 2020)	VCV001013743
NM_006766.5(KAT6A):c.4776 GTC[1] (p.Ser1597del)	Uncertain significance(Last reviewed: Apr 18, 2018)	VCV000588522
NM_006766.5(KAT6A):c.5974 G>C (p.Val1992Leu)	Uncertain significance(Last reviewed: Apr 20, 2018)	VCV000452008
NM_006766.5(KAT6A):c.5366 C>G (p.Thr1789Arg)	Uncertain significance(Last reviewed: Apr 23, 2019)	VCV001305371
NM_006766.5(KAT6A):c.5560 C>T (p.His1854Tyr)	Uncertain significance(Last reviewed: Apr 27, 2019)	VCV000638390
NM_006766.5(KAT6A):c.4108 G>T (p.Glu1370Ter)	Uncertain significance(Last reviewed: Apr 8, 2013)	VCV000162180
NM_006766.5(KAT6A):c.4594 A>T (p.Met1532Leu)	Uncertain significance(Last reviewed: Aug 16, 2019)	VCV001307891
NM_006766.5(KAT6A):c.2004 G>C (p.Leu668Phe)	Uncertain significance(Last reviewed: Aug 2, 2019)	VCV001307349
NM_006766.5(KAT6A):c.96A> G (p.Ile32Met)	Uncertain significance(Last reviewed: Aug 24, 2019)	VCV000937075
NM_006766.5(KAT6A):c.4210 G>A (p.Glu1404Lys)	Uncertain significance(Last reviewed: Aug 26, 2016)	VCV000589359
NM_006766.5(KAT6A):c.2806 C>A (p.Pro936Thr)	Uncertain significance(Last reviewed: Aug 26, 2021)	VCV001342810
NM_006766.5(KAT6A):c.2135 G>A (p.Ser712Asn)	Uncertain significance(Last reviewed: Aug 27, 2021)	VCV001328184

NM_006766.5(KAT6A):c.4058 A>G (p.Asp1353Gly)	Uncertain significance(Last reviewed: Dec 1, 2018)	VCV000810280
NM_006766.5(KAT6A):c.824C >T (p.Ala275Val)	Uncertain significance(Last reviewed: Dec 12, 2019)	VCV001310917
NM_006766.5(KAT6A):c.5135 C>T (p.Thr1712Ile)	Uncertain significance(Last reviewed: Dec 17, 2019)	VCV001310411
NM_006766.5(KAT6A):c.5809 C>T (p.His1937Tyr)	Uncertain significance(Last reviewed: Dec 26, 2019)	VCV001310678
NM_006766.5(KAT6A):c.4445 C>T (p.Pro1482Leu)	Uncertain significance(Last reviewed: Dec 7, 2019)	VCV001028557
NM_006766.5(KAT6A):c.4520 G>T (p.Ser1507Ile)	Uncertain significance(Last reviewed: Dec 8, 2017)	VCV000589580
NM_006766.5(KAT6A):c.107T >C (p.Val36Ala)	Uncertain significance(Last reviewed: Feb 1, 2019)	VCV000810281
NM_006766.5(KAT6A):c.638T >A (p.Leu213His)	Uncertain significance(Last reviewed: Feb 14, 2020)	VCV001311988
NM_006766.5(KAT6A):c.406G >A (p.Gly136Arg)	Uncertain significance(Last reviewed: Feb 15, 2021)	VCV001223701
NM_006766.5(KAT6A):c.250A >G (p.Lys84Glu)	Uncertain significance(Last reviewed: Feb 17, 2020)	VCV001303125
NM_006766.5(KAT6A):c.5594 C>T (p.Pro1865Leu)	Uncertain significance(Last reviewed: Feb 7, 2020)	VCV001311705
NM_006766.5(KAT6A):c.3778 A>G (p.Ser1260Gly)	Uncertain significance(Last reviewed: Jan 2, 2020)	VCV001028556
NM_006766.5(KAT6A):c.4841	Uncertain significance(Last reviewed: Jan 26, 2021)	VCV001331663

G>A (p.Ser1614Asn)		
NM_006766.5(KAT6A):c.2843 G>A (p.Arg948Gln)	Uncertain significance(Last reviewed: Jan 28, 2020)	VCV000983321
NM_006766.5(KAT6A):c.752G >A (p.Arg251Gln)	Uncertain significance(Last reviewed: Jan 28, 2021)	VCV001285505
NM_006766.5(KAT6A):c.4837 _4854dup (p.Gly1613_Me t1618dup)	Uncertain significance(Last reviewed: Jan 29, 2020)	VCV001206430
NM_006766.5(KAT6A):c.2140 AAG[1] (p.Lys715del)	Uncertain significance(Last reviewed: Jan 31, 2018)	VCV000504267
NM_006766.5(KAT6A):c.571G >T (p.Val191Leu)	Uncertain significance(Last reviewed: Jan 4, 2021)	VCV000389090
NM_006766.5(KAT6A):c.2496 A>G (p.Glu832=)	Uncertain significance(Last reviewed: Jan 6, 2021)	VCV001313851
NM_006766.5(KAT6A):c.2869 G>C (p.Ala957Pro)	Uncertain significance(Last reviewed: Jul 1, 2017)	VCV000493477
NM_006766.5(KAT6A):c.3040 -9_3040-8del	Uncertain significance(Last reviewed: Jul 1, 2017)	VCV000493476
NM_006766.5(KAT6A):c.3485 A>T (p.His1162Leu)	Uncertain significance(Last reviewed: Jul 1, 2020)	VCV001013578
NM_006766.5(KAT6A):c.3187 C>T (p.Pro1063Ser)	Uncertain significance(Last reviewed: Jul 12, 2019)	VCV001307027
NM_006766.5(KAT6A):c.1097 G>A (p.Arg366Gln)	Uncertain significance(Last reviewed: Jul 6, 2016)	VCV000588013
NM_006766.5(KAT6A):c.4697 C>G (p.Pro1566Arg)	Uncertain significance(Last reviewed: Jul 8, 2019)	VCV000985405

NM_006766.5(KAT6A):c.254C>T (p.Pro85Leu)	Uncertain significance(Last reviewed: Jul 9, 2021)	VCV001341746
NM_006766.5(KAT6A):c.4420T>A (p.Cys1474Ser)	Uncertain significance(Last reviewed: Jun 1, 2021)	VCV001013577
NM_006766.5(KAT6A):c.5339A>G (p.Tyr1780Cys)	Uncertain significance(Last reviewed: Jun 11, 2019)	VCV001219823
NM_006766.5(KAT6A):c.4529C>T (p.Thr1510Ile)	Uncertain significance(Last reviewed: Jun 17, 2021)	VCV001328779
NM_006766.5(KAT6A):c.1433C>T (p.Thr478Ile)	Uncertain significance(Last reviewed: Jun 20, 2017)	VCV000432911
NM_006766.5(KAT6A):c.3670_3675del (p.Lys1224_Glu1225del)	Uncertain significance(Last reviewed: Jun 20, 2019)	VCV001302094
NM_006766.5(KAT6A):c.2989C>T (p.Pro997Ser)	Uncertain significance(Last reviewed: Jun 21, 2019)	VCV001028555
NM_006766.5(KAT6A):c.2459A>G (p.Glu820Gly)	Uncertain significance(Last reviewed: Jun 23, 2017)	VCV000445391
NM_006766.5(KAT6A):c.5966C>G (p.Ala1989Gly)	Uncertain significance(Last reviewed: Jun 26, 2019)	VCV001306757
NM_006766.5(KAT6A):c.2911C>T (p.Arg971Cys)	Uncertain significance(Last reviewed: Jun 29, 2020)	VCV001184331
NM_006766.5(KAT6A):c.2008T>C (p.Ser670Pro)	Uncertain significance(Last reviewed: Jun 3, 2021)	VCV001306132
NM_006766.5(KAT6A):c.3989A>G (p.Lys1330Arg)	Uncertain significance(Last reviewed: Jun 5, 2019)	VCV001306017

NM_006766.5(KAT6A):c.1212 G>C (p.Lys404Asn)	Uncertain significance(Last reviewed: Jun 8, 2021)	VCV001327765
NM_006766.5(KAT6A):c.2696 A>G (p.Tyr899Cys)	Uncertain significance(Last reviewed: Mar 1, 2021)	VCV001342656
NM_006766.5(KAT6A):c.3299 C>G (p.Ser1100Cys)	Uncertain significance(Last reviewed: Mar 2, 2017)	VCV000589549
NM_006766.5(KAT6A):c.2791 G>A (p.Gly931Arg)	Uncertain significance(Last reviewed: Mar 2, 2019)	VCV001028553
NM_006766.5(KAT6A):c.2947 G>T (p.Gly983Cys)	Uncertain significance(Last reviewed: Mar 25, 2019)	VCV001028554
NM_006766.5(KAT6A):c.458G >A (p.Arg153His)	Uncertain significance(Last reviewed: Mar 8, 2017)	VCV000424162
NM_006766.5(KAT6A):c.3353 -4A>G	Uncertain significance(Last reviewed: May 17, 2017)	VCV000588978
NM_006766.5(KAT6A):c.523A >G (p.Asn175Asp)	Uncertain significance(Last reviewed: May 17, 2021)	VCV001254350
NM_006766.5(KAT6A):c.2915 G>C (p.Arg972Pro)	Uncertain significance(Last reviewed: May 19, 2016)	VCV000589262
NM_006766.5(KAT6A):c.248C >G (p.Pro83Arg)	Uncertain significance(Last reviewed: May 24, 2021)	VCV001326554
NM_006766.5(KAT6A):c.5921 G>A (p.Gly1974Glu)	Uncertain significance(Last reviewed: May 31, 2017)	VCV000589125
NM_006766.5(KAT6A):c.5299 C>G (p.His1767Asp)	Uncertain significance(Last reviewed: May 4, 2020)	VCV001028560
NM_006766.5(KAT6A):c.3830	Uncertain significance(Last reviewed: May 5, 2014)	VCV000162181

_3831insTT (p.Arg1278fs)		
NM_006766.5(KAT6A):c.5207C>T (p.Pro1736Leu)	Uncertain significance(Last reviewed: May 5, 2020)	VCV001028559
NM_006766.5(KAT6A):c.4235A>G (p.Glu1412Gly)	Uncertain significance(Last reviewed: May 6, 2016)	VCV000587871
NM_006766.5(KAT6A):c.5630G>A (p.Arg1877His)	Uncertain significance(Last reviewed: Nov 1, 2017)	VCV000547034
NM_006766.5(KAT6A):c.26A>G (p.Tyr9Cys)	Uncertain significance(Last reviewed: Nov 1, 2021)	VCV001335710
NM_006766.5(KAT6A):c.247C>T (p.Pro83Ser)	Uncertain significance(Last reviewed: Nov 18, 2016)	VCV000373381
NM_006766.5(KAT6A):c.2366A>G (p.Glu789Gly)	Uncertain significance(Last reviewed: Nov 19, 2019)	VCV001304714
NM_006766.5(KAT6A):c.3032A>C (p.Lys1011Thr)	Uncertain significance(Last reviewed: Nov 26, 2019)	VCV001310343
NM_006766.5(KAT6A):c.3026C>T (p.Thr1009Met)	Uncertain significance(Last reviewed: Nov 29, 2017)	VCV000595202
NM_006766.5(KAT6A):c.766C>T (p.Arg256Trp)	Uncertain significance(Last reviewed: Nov 29, 2019)	VCV000976403
NM_006766.5(KAT6A):c.4949A>C (p.Gln1650Pro)	Uncertain significance(Last reviewed: Nov 3, 2021)	VCV001319833
NM_006766.5(KAT6A):c.2422G>A (p.Glu808Lys)	Uncertain significance(Last reviewed: Nov 4, 2019)	VCV001309699
NM_006766.5(KAT6A):c.4292dup (p.Leu1431fs)	Uncertain significance(Last reviewed: Nov 6, 2013)	VCV000162182

NM_006766.5(KAT6A):c.4598T>C (p.Met1533Thr)	Uncertain significance(Last reviewed: Nov 8, 2018)	VCV001034309
NM_006766.5(KAT6A):c.3412C>T (p.Leu1138Phe)	Uncertain significance(Last reviewed: Oct 1, 2019)	VCV000871561
NM_006766.5(KAT6A):c.4688A>T (p.Tyr1563Phe)	Uncertain significance(Last reviewed: Oct 1, 2019)	VCV000871560
NM_006766.5(KAT6A):c.3982A>G (p.Lys1328Glu)	Uncertain significance(Last reviewed: Oct 12, 2016)	VCV000376989
NM_006766.5(KAT6A):c.4752G>C (p.Gln1584His)	Uncertain significance(Last reviewed: Oct 22, 2020)	VCV001313294
NM_006766.5(KAT6A):c.5878A>C (p.Ser1960Arg)	Uncertain significance(Last reviewed: Oct 23, 2018)	VCV000388829
NM_006766.5(KAT6A):c.68A>G (p.Gln23Arg)	Uncertain significance(Last reviewed: Oct 25, 2019)	VCV001309657
NM_006766.5(KAT6A):c.4960C>A (p.Pro1654Thr)	Uncertain significance(Last reviewed: Oct 3, 2019)	VCV001028558
NC_000008.10:g.(?_41905876)_(42188497_?)dup	Uncertain significance(Last reviewed: Oct 7, 2020)	VCV001061009
NM_006766.5(KAT6A):c.5817C>A (p.Asn1939Lys)	Uncertain significance(Last reviewed: Sep 1, 2019)	VCV001303386
NM_006766.5(KAT6A):c.5270A>G (p.Gln1757Arg)	Uncertain significance(Last reviewed: Sep 1, 2021)	VCV001013576
NM_006766.5(KAT6A):c.205A>C (p.Asn69His)	Uncertain significance(Last reviewed: Sep 16, 2020)	VCV000992311

GRCh37/hg19 8p12-11.1(chr8:36094421-43822214)x3	Uncertain significance(Last reviewed: Sep 17, 2018)	VCV000688131