



Mechanistic basis and evaluation of neurological malleability of Wiedemann- Steiner syndrome

Tinna Reynisdóttir

Thesis for the degree of Philosophiae Doctor

August 2024

School of Health Sciences

FACULTY OF MEDICINE

UNIVERSITY OF ICELAND

Mechanistic basis and evaluation of neurological malleability of Wiedemann-Steiner syndrome

Tinna Reynisdóttir

Thesis for the degree of Philosophiae Doctor

Supervisor

Professor Hans Tómas Björnsson MD PhD

Doctoral committee

Assistant Professor Ásbjörg Ósk Snorradóttir PhD

Bergþóra Eiríksdóttir DVM

Professor Eiríkur Steingrímsson PhD

Wendy Jones MD PhD

August 2024

School of Health Sciences

FACULTY OF MEDICINE

UNIVERSITY OF ICELAND

Undirliggjandi sjúkdómsmyndun og ákvörðun mótanleika
taugasvipgerðar í Wiedemann-Steiner heilkenni

Tinna Reynisdóttir

Ritgerð til doktorsgráðu

Leiðbeinandi

Prófessor Hans Tómas Björnsson MD PhD

Doktorsnefnd

Lektor Ásbjörg Ósk Snorradóttir PhD

Bergþóra Eiríksdóttir DVM

Prófessor Eiríkur Steingrímsson PhD

Wendy Jones MD PhD

Ágúst 2024

Heilbrigðisvísindasvið

LÆKNADEILD

HÁSKÓLI ÍSLANDS

Thesis for a doctoral degree at the University of Iceland. All right reserved. No part of this publication may be reproduced in any form without the prior permission of the copyright holder.

© Tinna Reynisdóttir 2024

ISBN 978-9935-9781-2-7

ORCID: 0000-0001-5148-4744

Reykjavik, Iceland 2024

Dedication

Þessi ritgerð er tileinkuð ömmu Dýrley

Ágrip

Wiedemann-Steiner heilkenni (WDSTS) er sjaldgæfur erfðasjúkdómur sem veldur greindarskerðingu sem enn reynist ólæknandi og lítið er vitað um sjúkdómsmyndun þess. Orsakir WDSTS eru arfblendnar stökkbreytingar í geni sem táknar fyrir ensíminu lysín-histónumetyltransferasa 2A (KMT2A). KMT2A er hluti af umframerfðakerfinu og sér um að bæta við metýlhóp á lysín amínósýru 4 á histónuhala H3 histónupróteinsins og þannig stýra genatjáningu. Birtingarmynd WDSTS felur í sér margvísleg líkamleg einkenni, t.d. vaxtarskerðingu, vöðvaslappleika, meltingarfæravandamál, afbriðgilega andlitsdrætti og aukinn hárvöxt. Hegðunar- og taugakerfisvandamál eru einnig algeng hjá einstaklingum með WDSTS, líkt og greindarskerðing, þroskaskerðing og aukin árásargirni. Markmið þessarar doktorsritgerðar var að dýpka skilning okkar á sjúkdómsmyndun WDSTS með ýmsum aðferðum og auka þannig þekkingu á þessu sjaldgæfa heilkenni. Auk þess rannsökuðum við mótanleika taugasvipgerðar WDSTS fósturstigi til þess að meta möguleika á þróun meðferða við heilkenninu í framtíðinni.

Í fyrsta lagi notuðum við lífupplýsingafræðilegar aðferðir til þess að skoða áhrif mislestursbreytinga í KMT2A geninu. Við bárum saman dreifingu mislestursbreytinga í KMT2A sem valda WDSTS og þeirra sem finnst í heilbrigðu þýði og valda ekki WDSTS. Við greindum mikla aukningu mislestursbreytinga í CXXC hneppinu í einstaklingum með WDSTS en hneppi þetta sér um DNA bindingu KMT2A. Hinsvegar fundum við enga aukningu á mistlestursbreytingum í SET hneppinu sem sér um ensím virkni próteinsins. Þessar niðurstöður benda til þess að CXXC hneppið sinni mikilvægu hlutverki í sjúkdómsmyndun WDSTS. Næst nýttum við okkur gervigreindartólið AlphaFold2 (AF2) í bland við tilraunaþekkingu okkar til þess að búa til tölvulíkan af ensíminu. Þetta tölvulíkan gerði okkur kleift að spá fyrir um áhrif allra mögulegra mislestursstökkbreytinga á byggingu CXXC hneppisins í KMT2A próteininu, þ.e. hvort hún væri líkleg til þess að valda heilkenninu eða ekki. Þessi aðferð gæti því nýst í framtíðinni í klínískri vinnu við greiningu á því hvort KMT2A stökkbreytingar eru líklegar til að valda WDSTS.

Í öðru lagi þróuðum við nýtt músamódel fyrir WDSTS sem kallast *Kmt2a^{+/LSL}*. Við nýttum okkur margvíslegar aðferðir til að greina áhrif breytingarinnar á svipgerð músanna eins og hegðunarpróf, vefjafræðilegar litanir, kjarnsýrumögnun (PCR), RT-qPCR, tölvusneiðmyndir og heilraðgreiningu erfðamengis. Við sýndum fram á að borið saman við villigerð úr sama goti sýndi *Kmt2a^{+/LSL}* músamódelið lykilsvipgerðir WDSTS heilkennisins og er því gott líkan fyrir WDSTS í mönnum. Þetta gerir músamódelið að mikilvægu tóli fyrir rannsóknir á WDSTS. Auk þess sýna *Kmt2a^{+/LSL}* mýsnaar aukinn hárvöxt og hvítan kviðlægan blett í feldi sem gæti mögulega tengt KMT2A við áður

óþekkt hlutverk í hárvexti og taugakambspróun. Til þess að rannsaka sjúkdómsmyndun WDSTS betur í frumum framkvæmdum við RNA raðgreiningu og CUT&Run greiningu á H3K4me1 umframerfðamerkinu í músa taugaforverafrumum (mNPC). Við greindum vanstjórnun í genatjáningu og aukið aðgengi litnis í $Kmt2a^{+/LSL}$ taugaforverafrumum.

Að lokum var mótanleiki taugasvipgerðar WDSTS músa rannsakað með því að notast við nestin-Cre músamódel. Með þeirri nálgun gátum við endurheimt við fulla tjáningu $Kmt2a$ gensins um miðja meðgöngu í $Nes-Cre^{+/-}Kmt2a^{+/LSL}$ músunum. Með áðurnefndum svipgerðargreiningaraðferðum sýndum við fram á að taugaforverafrumur $Nes-Cre^{+/-}Kmt2a^{+/LSL}$ músanna höfðu leiðrétt afbrigðileikana í genatjáningu og endurheimt jafnvægi í umframerfðakerfinu. Við sýndum einnig fram á að $Nes-Cre^{+/-}Kmt2a^{+/LSL}$ mýsnar leiðrétt þær taugasvipgerðir sem lýst var í $Kmt2a^{+/LSL}$ músunum. Niðurstöður þessar sýna mótanleika taugasvipgerðarinnar í WDSTS músunum á meðgöngu.

Þegar á heildina er litið auka niðurstöður þessarar doktorsritgerðar skilning okkar á WDSTS. Við kynnum nýjar niðurstöður sem sýna fram á hlutverk CXXC hneppisins í sjúkdómsmyndun WDSTS auk þess að varpa ljósi á áður óþekkt hlutverk KMT2A í taugakambspróun. Okkur tókst að búa til nýtt músamódel fyrir WDSTS og lýsa því í smáatriðum ásamt því að sýna fram á mótanleika taugasvipgerðar músamódelisins á fósturstigi. Fyrir vikið ætti að vera fýsilegt að þróa meðferðarúrræði við WDSTS í framtíðinni.

Lykilorð: Umframerfðir, KMT2A, greindarskerðing, músamódel, Wiedemann-Seiner heilkenni

Abstract

Wiedemann-Steiner syndrome (WDSTS) is a rare genetic cause of intellectual disability, with no cure and little research on the disease mechanism to this date. WDSTS is caused by heterozygous variants in a gene encoding the epigenetic regulator lysine-histone methyltransferase 2A (KMT2A). Individuals with WDSTS present with a range of symptoms widely affecting the body such as growth retardation, hypotonia, GI symptoms, distinct craniofacial features and hypertrichosis. Within the nervous system, individuals with WDSTS most commonly present with intellectual disability, developmental delay and aggressive behavior, which can prove difficult to manage and impacts their quality of life. In this PhD thesis, we aimed to gain a deeper understanding of the underlying disease mechanism of Wiedemann-Steiner syndrome (WDSTS) using several different methods and to expand on the current knowledge of this rare syndrome. Furthermore, we investigated the *in utero* malleability of the WDSTS neurological phenotypes in a mouse model, advancing the potential of future therapeutics for the syndrome.

First, using bioinformatic approaches, we wanted to explore the distribution of WDSTS-associated missense variants (MVs) in KMT2A. We identified the DNA binding CXXC domain of KMT2A to be highly enriched for WDSTS causing MVs using a comparative spatial analysis of MVs found in healthy populations and individuals with WDSTS. We found no evidence of MV enrichment in the enzymatic SET domain. These results suggest a novel, non-enzymatic and central role for the CXXC domain in the WDSTS pathogenesis. Utilizing the new AI-based protein prediction software AlphaFold2 (AF2) and prior experimental knowledge, we performed an *in silico* saturation mutagenesis of all potential MVs found within the CXXC domain of KMT2A. This allowed us to classify any future MVs as likely pathogenic or not, providing a new approach to determine the pathogenicity of a variant within the clinic.

Second, we established a novel disease mouse model of WDSTS termed *Kmt2a*^{+/LSL}. We utilized various methods to characterize the mice in detail such as behavior testing, histology staining, PCRs, RT-qPCRs, CT scanning and Genome-wide long-read sequencing. We successfully demonstrated that the *Kmt2a*^{+/LSL} model displays key WDSTS phenotypes compared to their wildtype littermates, and adequately represents the human WDSTS disease, making it a valuable model for any future WDSTS studies. Additionally, we provided evidence of hypertrichosis and a white abdominal spot present in our *Kmt2a*^{+/LSL} model, potentially linking KMT2A to novel roles in hair growth and neural crest development. To further explore the mechanistic basis of WDSTS on a cellular level, we performed RNA sequencing and a CUT&Run of

the H3K4me1 mark of mouse neuronal progenitor cells (mNPCs). We uncovered gene expression dysregulation and increased chromatin accessibility of the *Kmt2a*^{+/*LSL*} mutant mNPCs.

Finally, we tested the malleability of neurological dysfunction using a nestin-Cre model, crossing the mouse strain with our *Kmt2a*^{+/*LSL*} mice after validating its functionality. Using this approach, we restored *Kmt2a* expression around mid-gestation in the Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} mice. With the phenotyping methods described above, we provide robust evidence that the gene expression and epigenetic imbalance is restored in our Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} mNPCs. We also demonstrated that the Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} mice showed phenotypic rescue of the WDSTS-like neurological dysfunction of the *Kmt2a*^{+/*LSL*} mice.

Taken together, the results of this thesis expand on the current knowledge of WDSTS. We provide new evidence for the role of the CXXC domain of KMT2A in WDSTS pathogenesis, as well as shedding light on new roles of KMT2A in biological processes such as neural crest development. We successfully created and characterized a novel mouse model for researching WDSTS as well as providing compelling evidence of *in utero* malleability of the neurological dysfunction, making WDSTS a promising candidate for future therapeutic opportunities.

Keywords: Epigenetics, KMT2A, Intellectual disability, Mouse model, Wiedemann-Steiner syndrome

Acknowledgements

First, I want to thank my supervisor Hans Tómas Björnsson for trusting me with this outstanding project, guidance and mentorship throughout my PhD journey. Your passion and interest in science has inspired me countless times and given me inspiration to continue even when experiments did not go as planned. Under your guidance I have had the freedom to follow my curiosity, plan my experiments, had the opportunity to attend conferences, learned to write papers and so much more, all of which are invaluable to a PhD student. I am deeply grateful for this experience.

Our lab manager, Kimberley Jade Anderson, has played a central role in my PhD journey. Kimberley has taught me the ways around the lab and the mice, and has collaborated hands-on on both my manuscripts, and for that I am forever grateful. More than that, you have supported me in the lows and cheered me on during the highs, it would have been impossible without you. Next, I want to thank Leandros Boukas for providing me with excellent guidance through my first years as a young scientist. You have taught me valuable lessons on critical thinking and pushed me to think outside the box (still logically though). Thank you for taking me on as your student and friend.

Furthermore, I want to express my gratitude to all the members of the Louma G. Foundation Laboratory and our extended lab members in Baltimore, for their input, teamwork and all the fun times in the lab. Katrín Möller and Juan helped me take on the cell culture and guided me every step, your brilliance and knowledge seems endless. Arnhildur, Salvör and Meghna have provided invaluable feedback and assistance. And sometimes just an ear to listen to the difficulties of a PhD. I have been beyond lucky to collaborate and work with such brilliant people who are always willing to lend a helping hand. Thank you.

To Sara, my forever PhD wife. I am not even sure how I can thank you enough. You inspired me to take on this huge task of a PhD and supported me through every step of the way. The ups and the downs, the experiments that worked and the many that did not. You have given me great feedback, helped me brainstorm and troubleshoot for many years now. You are the greatest friend, colleague, partner in crime and lab-rat. I will forever cherish our time spent together inside and outside the lab. The countless mornings sitting with our Breakfast-club member Teitur, discussing science and supporting each other. From the bottom of my heart thank you both.

At ArcticLAS, the lab has been so very fortunate to have the assistance of two outstanding specialists, Bergþóra and Þóra. Working with mice has been very challenging for me and your help has been essential to my PhD progress. All the coffee

break chats, support and willingness to assist in any experiment has been valued. I am very grateful for my enthusiastic doctoral committee: Eiríkur, Ásbjörg Ósk, Wendy Jones and Bergþóra for taking the time out of their busy schedules to attend my committee meetings, providing me with great advice, scientific input and counselling. Your revision of my thesis and comments have been of great help. Also, I want to thank my international collaborators, Aamir R. Zuberi at the Jackson Laboratory and Valerie Burke DeLeon at University of Florida.

To my family and friends. I want to thank you all for cheering me on when I decided to take on such a big, and slightly crazy, journey that a PhD is. I am incredibly fortunate to have such a supportive family. Mom and dad, you have believed in me and supported me through my entire 20 years of school, through every single up and down. You have made this PhD dream of mine possible. To my siblings Gísli Már and Bryndís and my bonus sister Elín, I have looked up to your academic and non-academic achievements since I was young, and your accomplishments have been a great inspiration to me. My best friends: Bergrós, Agnes, Hjördís, Guðrún and Eva, your endless love and constant reassurance throughout these years has been above and beyond. In every single PhD crisis (and there were so many) you picked up the phone and helped me get back on track. You are the definition of true friendship and I hold the memories we have made close to my heart. To Uglá, Alexandra, Ásdís and all my friends that have supported me, thank you for always being there. To Christian, thank you for making my time in Copenhagen unforgettable. You provided me with love and support when I struggled, a hug and a cheer when things worked out. Finally, to everyone that assisted me in translating my abstract to Icelandic, mínar dýpstu þakkir.

This project was carried out at the Louma G. Foundation laboratory at the Faculty of Medicine at the University of Iceland. It was funded and made possible by the WSS Foundation and with a doctoral grant from the Eimskip University Fund, for that I am grateful. Also, I thank Félag Háskólakennara and BHM for funding my international travel allowing me to present my work abroad.

Contents

Dedication.....	i
Ágrip.....	iii
Abstract	v
Acknowledgements.....	vii
Contents	ix
List of Abbreviations	xiii
List of Figures	xix
List of Tables.....	xx
List of Equations	xxi
List of Original Papers.....	xxii
Declaration of Contribution	xxiii
1 Introduction	1
1.1 Neural crest development	1
1.2 Central nervous system development	1
1.2.1 Neuron development	3
1.2.2 Postnatal neurogenesis	4
1.3 Epigenetics and transcription	4
1.3.1 Overview	4
1.3.2 Alternative splicing	5
1.3.3 DNA methylation	6
1.3.4 Histone modification machinery	7
1.3.5 Promoters and enhancers	9
1.3.6 Interaction and compensation through epigenetic machineries	9
1.4 Lysine methyl transferase 2A (KMT2A)	10
1.4.1 KMT2s.....	10
1.4.2 KMT2A structure	11
1.4.3 KMT2A functional role	12
1.5 Epigenetics and neurological disorders.....	13
1.5.1 Mendelian disorders of the epigenetic machinery (MDEMs)	14
1.5.2 Wiedemann-Steiner syndrome.....	16

1.6	Therapeutic development for rare diseases	18
1.6.1	Overview of therapeutic approaches.....	18
1.6.2	Current studies of the MDEMs.....	20
1.6.3	<i>In utero</i> therapeutics	20
1.6.4	Challenges and considerations.....	21
1.7	Cre-loxP system.....	21
2	Aims.....	23
2.1	Aim 1: Exploring the effects of WDSTS missense variants <i>in silico</i>	23
2.2	Aim 2: Creating and characterizing a mouse model of WDSTS (<i>Kmt2a</i> ^{+/LSL}).....	23
2.3	Aim 3: Testing malleability of the neurological dysfunction in WDSTS <i>in vitro</i> and <i>in vivo</i>	24
3	Materials and Methods	25
3.1	Spatial missense variant (MV) analysis and conservation.....	25
3.2	MV classification and saturation mutagenesis.....	25
3.3	Mouse models	25
3.3.1	Mouse husbandry and licenses.....	25
3.3.2	The WDSTS mouse model	26
3.3.3	Nestin-Cre mouse model	26
3.3.4	ROSA-CreER ^{T2} mouse model.....	27
3.3.5	<i>Dct-LacZ</i> mouse model	28
3.4	Mouse experiments	28
3.4.1	Genotyping	28
3.4.2	Inducible Cre-recombination of <i>Kmt2a</i> ^{+/LSL} mice	30
3.4.3	Mouse model phenotyping	33
3.5	Cellular experiments.....	35
3.5.1	mNPCs extraction and culture	35
3.5.2	Incucyte experiments	35
3.6	Sequencing experiments.....	36
3.6.1	RNA sequencing	36
3.6.2	Epigenetic profiling methods	36
3.6.3	Genome-wide sequencing	37
3.7	Statistical testing and figure rendering	37
4	Results	39

4.1	<i>In silico</i> comparative analysis of WDSTS causing MVs (Paper I).....	39
4.1.1	WDSTS- associated MVs preferentially occur in the CXXC domain of KMT2A.....	39
4.1.2	Effects of <i>Kmt2a</i> cKO disruption on histone methylation and gene expression	40
4.1.3	Missense variant classification scheme.....	41
4.2	<i>In utero</i> malleability of neurological dysfunction of WDSTS (Paper II).....	43
4.2.1	Creation and validation of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mouse model.....	43
4.2.2	Phenotyping of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mouse model.....	43
4.2.3	Neurological dysfunction of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mouse model	45
4.2.4	Validation of Nes-Cre ^{+/-} / <i>Kmt2a</i> ^{+/<i>LSL</i>} mouse model	47
4.2.5	Transcriptional disruption resulting from heterozygous <i>Kmt2a</i> loss.....	48
4.2.6	<i>Kmt2a</i> expression in the <i>Kmt2a</i> ^{+/<i>LSL</i>} mNPCs	51
4.2.7	Chromatin abnormalities due to heterozygous <i>Kmt2a</i> loss	52
4.2.8	<i>In utero</i> rescue of the neurological dysfunction of the Nes-Cre ^{+/-} / <i>Kmt2a</i> ^{+/<i>LSL</i>} mice.....	54
4.3	Unpublished data	56
4.3.1	Tamoxifen induced recombination	56
4.3.2	Fur phenotype of <i>Kmt2a</i> ^{+/<i>LSL</i>} mice	60
4.3.3	X-gal stainings	63
4.3.4	Cellular phenotypes of mNPCs.....	64
5	Discussion	67
5.1	Classification of missense variants of the CXXC domain	67
5.2	Technical challenges and considerations of the Cre mouse models.....	68
5.3	Characterization of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mice.....	70
5.3.1	Fur phenotype of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mice	71
5.4	Malleability of the WDSTS neurological phenotype.....	72
5.4.1	Nestin-Cre mouse model	73
5.4.2	<i>In vitro</i> compensation and cellular rescue.....	74
5.5	The CXXC domain vs the SET domain	76
5.6	Implications for future therapeutics for WDSTS	77
5.7	Linking KMT2A to neural crest development	77
5.7.1	White abdominal spot	77

5.7.2	Craniofacial abnormalities and overlap with KS	78
5.7.3	KMT2A gene regulation and the neural crest.....	78
5.7.4	Enteric nervous system defects.....	79
6	Conclusions.....	81
6.1	Future directions.....	81
6.2	Closing remarks.....	82
	References.....	83
	Original Publications	125
	Paper I	127
	Paper II	161

List of Abbreviations

4-OH-tam	4-Hydroxy-tamoxifen
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AAV	Adeno-associated virus
Acetyl-CoA	Acetyl coenzyme A
AD	Alzheimer's disease
AF2	AlphaFold2
AI	Artificial intelligence
ALL	Acute lymphoblastic leukemia
ASD	Autism spectrum disorder
ASD	Autism spectrum disorder
ASH2L	ASH2 like protein
ASO	Antisense oligonucleotides
BBB	Blood-brain-barrier
BLBP	Brain Lipid-Binding Protein
BMP	Bone Morphogenetic Protein
Bp	Base pair
CA	Cornu Ammonis
Cas9	CRISPR Associated Protein 9
CBC	Complete Blood Count
cDNA	Complementary DNA
CF	Cystic Fibrosis
CGI	CpG island
ChIP-Seq	Chromatin Immunoprecipitation followed by sequencing
cKI	Conditional knock-in

cKO	Conditional knock-out
CNS	Central nervous system
CO	Corn oil
CpG	Cytosine—guanine dinucleotide
Cre-ERT2	Cre-recombinase estrogen receptor T2
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSF	Cerebrospinal fluid
CT	Cycle threshold
CUT&Run	Cleavage Under Targets & Release Using Nuclease
Dct	Dopachrome tautomerase
DCX	Doublecortin
DD	Developmental delay
DEG	Differentially expressed gene
DG	Dentate gyrus
DM	Double mutant
DNA	Deoxyribonucleic acid
DNA _m	DNA methylation
DNMT	DNA methyltransferase
DPE	Downstream promoter element
E	Embryonic day
EM gene	Epigenetic machinery gene
EMT	Epithelial-to-mesenchymal transition
ERT	Enzyme replacement therapy
ESC	Embryonic stem cells
EtOH	Ethanol
FDR	False discovery rate
FGF	Fibroblast Growth Factor
FOXD3	Forkhead Box D3
FTT	Failure to thrive
gDNA	Genomic DNA

GFAP	Glial Fibrillary Acid Protein
GI	Gastrointestinal
GLM	Generalized linear models
GTF	General transcription factor
GWS	Genome-wide sequencing
H3K27me	Histone 2 lysine 27 methylation
H3K4me	Histone 3 lysine 4 methylation
H3S10p	Histone 3 serine 10 phosphorylation
HAT	Histone acetyltransferase
HCF1	Host Cell Factor 1
HCF2	Host Cell Factor 2
HD	Huntington's disease
HDAC	Histone deacetylase
HDACi	HDAC inhibitor
HDM	Histone demethylase
HMT	Histone methyltransferase
hnRNP	Heterogeneous nuclear ribonucleoproteins
Hox	Homeobox
HP1	Heterochromatin protein 1
HSC	Hematopoietic stem cell
HSCR	Hirschsprung disease
ID	Intellectual disability
IF	Isoform fraction
IF staining	Immunofluorescence staining
IHC	Immunohistochemistry
Inr	Initiator sequence
IP	Intraperitoneal
IQ	Intelligence quotient
IUGT	In-utero gene therapy
JAX	The Jackson Laboratory

JMJD	Jumonji C domain-containing protein
K	Lysine
Kb	Kilobase
KD	Knock-down
kDa	Kilodalton
KDM	Lysine-specific demethylases
KI	Knock-in
KMT	Lysine methyltransferase
KO	Knock-out
KS	Kabuki syndrome
LFC	Log ₂ fold change
lncRNA	Long non-coding RNA
LoF	Loss-of-function
LoxP	Locus of x-over P1
LSD	Lysine-specific demethylases
LSL	LoxP-stop-loxP
LTP	Long-term potentiation
MAP2	Microtubule-Associated Protein 2
MBD protein	Methyl-CpG-binding domain protein
MDEM	Mendelian disorder of the epigenetic machinery
miRNA	Micro-RNA
MLL	Mixed lineage leukemia
mNPC	Mouse neuronal progenitor cell
mRNA	Messenger ribonucleic acid
MRI	Magnetic resonance imaging
MV	Missense variant
NaOH	Sodium hydroxide
NC	Neural crest
NCC	Neural crest cell
NDD	Neurodevelopmental disorder

NeuN	Neuronal Nuclei antigen
NeuroD1	Neuronal Differentiation 1 protein
NMD	Nonsense-mediated decay
NPC	Neuronal progenitor cells
NSC	Neural stem cell
NT	Neurotransmitter
O/E	Observed-to-expected
OD	Optical density
ONT	Oxford Nanopore Technology
ORA	Over-representation analysis
P	Postnatal day
PAM	Protospacer adjacent motif
PAX3	Paired box 3
PCR	Polymerase chain reaction
PHD	Plant homeodomain
PMT	Post-translational modification
PNS	Peripheral nervous system
Pre-mRNA	Precursor mRNA
PRMT	Protein arginine methyltransferase
R	Arginine
RBBP5	Retinoblastoma Binding Protein 5
RNA pol II	RNA polymerase II
RNA-seq	RNA sequencing
Rpm	Rounds per minute
RT	Room temperature
RT-qPCR	Real-time quantitative PCR
RTS	Rubinstein-Taybi syndrome
SAM	S-adenosyl methionine
SD	Standard deviation
sgRNA	Single guide RNA

SGZ	Subgranular zone
siRNA	Small interfering RNA
SMA	Spinal muscular atrophy
SNAI1	Snail Family Transcriptional Repressors 1
SNAI2	Snail Family Transcriptional Repressors 2
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
SOX10	Sex-determining region Y-Box Transcription Factor 9
SOX9	Sex-determining region Y-Box Transcription Factor 9
SR protein	Serine/arginine rich nuclear phosphoprotein
SV40pA	Polyadenylation signal from the simian virus 40
SVZ	Subventricular zone
TAD	Topologically associated domains
Tam	Tamoxifen
TBP	TATA binding protein
TET	Ten-eleven translocation methylcytosine dioxygenase enzyme
TF	Transcription factor
TSS	Transcription start site
TWIST1	Twist Family BHLH Transcription Factor 1
UHRF protein	Ubiquitin-like containing PHD- and RING-finger domain protein
VZ	Ventricular zone
WDR5	WD repeat domain 5
WDSTS	Wiedemann-Steiner syndrome
Wnt	Wingless and Int-1 ligand
WT	Wild-type
ZEB2	Zinc Finger E-Box Binding Homeobox 2
Δ	Delta

List of Figures

Figure 1. Schematic figure of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mouse model.	26
Figure 2. Nestin-Cre and <i>Kmt2a</i> ^{+/<i>LSL</i>} breeding schematic.	27
Figure 3. Tamoxifen induced Cre-recombination of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mice.	28
Figure 4. The distribution of likely pathogenic Wiedemann-Steiner syndrome missense variants across the different domains of KMT2A.	40
Figure 5. The relationship between disrupted H3K4me1/3, regional observed-to-expected CpG ratio, and gene expression in <i>Kmt2a</i> -deficient mice.	41
Figure 6. An <i>in silico</i> saturation mutagenesis of the CXXC domain of KMT2A.	42
Figure 7. A novel Wiedemann-Steiner syndrome mouse model (<i>Kmt2a</i> ^{+/<i>LSL</i>}) recapitulates core phenotypes seen in individuals with WDSTS.	44
Figure 8. Neurological defects observed in the <i>Kmt2a</i> ^{+/<i>LSL</i>} mouse model.	46
Figure 9. Validation of Nes-Cre ^{+/-} <i>Kmt2a</i> ^{+/<i>LSL</i>} model.	48
Figure 10. Transcriptional consequences of <i>Kmt2a</i> loss and rescue.	50
Figure 11. <i>Kmt2a</i> expression of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mNPCs.	51
Figure 12. Epigenetic imbalance caused by heterozygous <i>Kmt2a</i> loss.	53
Figure 13. <i>In utero</i> rescue of the neurological phenotypes in WDSTS mice.	55
Figure 14. Mouse weight results from optimization 1.	57
Figure 15. Mouse weight results from optimization 2.	58
Figure 16. Relative <i>Kmt2a</i> expression from optimization 2.	59
Figure 17. Mouse weight results from optimization 3.	59
Figure 18. Relative <i>Kmt2a</i> expression from optimization 3.	60
Figure 19. Hair follicle density of <i>Kmt2a</i> ^{+/<i>LSL</i>} mice.	61
Figure 20. Fur coat of the <i>Kmt2a</i> ^{+/<i>LSL</i>} mice.	62
Figure 21. Melanin quantification in the fur and skin of <i>Kmt2a</i> ^{+/<i>LSL</i>} mice.	63
Figure 22. X-gal staining of mouse embryos.	64
Figure 23. Proliferation rate of mNPCs.	65
Figure 24. Scratch wound healing assay using mNPCs.	66

List of Tables

Table 1. Human lysine methyltransferase 2 (KMT2) nomenclature.	10
Table 2. Genotyping reactions for ExtraMix and Q5 reactions.	29
Table 3. Genotyping PCR primers, their sequence and PCR product size.....	30
Table 4. Genotype, treatment and number of mice in each optimization for the inducible Cre-recombination.	32
Table 5. RT-qPCR primer names and sequences.....	33

List of Equations

Equation 1. RT-qPCR relative gene expression calculation.	33
--	----

List of Original Papers

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals:

- I. Reynisdottir, T., Anderson, K. J., Boukas, L., & Bjornsson, H. T. (2022). *Missense variants causing Wiedemann-Steiner syndrome preferentially occur in the KMT2A-CXXC domain and are accurately classified using AlphaFold2*. PLoS Genetics, 18(6), e1010278. <https://doi.org/10.1371/journal.pgen.1010278>
- II. Reynisdottir, T., Anderson, K. J., Brinn, A., Franklin, K., Ouyang, J., Snorraddottir, A. O., Lutz, C. M., Zuberi, A. R., DeLeon, V. B., & Bjornsson, H. T. (2024). *In-utero rescue of neurological dysfunction in a mouse model of Wiedemann-Steiner syndrome*. <https://doi.org/10.1101/2024.07.19.604339>. Submitted for publication.

In addition, some unpublished data are presented.

Papers are reprinted by kind permission of the publishers.

Declaration of Contribution

Paper I: Missense variants causing Wiedemann-Steiner syndrome preferentially occur in the KMT2A-CXXC domain and are accurately classified using AlphaFold2.

In this paper, where I am first author, we explore the mechanistic pathogenesis of WDSTS by utilizing a bioinformatics approaches. I designed the study with my supervisor Hans Tómas Björnsson and post-doc Leandros Boukas. I performed initial analysis for the paper under the guidance of Leandros Boukas, for Figures 1 and 3-4. This includes the comparative analysis of missense variants in WDSTS compared to healthy controls and the development of an AlphaFold2-based variant effect classification scheme for the CXXC domain of KMT2A. I then performed *in silico* saturation mutagenesis of all potential variants within the CXXC domain of KMT2A using the variant classification scheme. I wrote the manuscript and generated figures in collaboration with supervisor and co-authors.

Paper II: *In utero* rescue of neurological dysfunction in a mouse model of Wiedemann-Steiner syndrome.

In this paper we characterize a novel mouse model for Wiedemann-Steiner syndrome (WDSTS) and utilize it to explore the *in utero* malleability of the neurological phenotype of WDSTS. My supervisor Hans Tómas Björnsson designed the novel mouse model that was subsequently created at the Jackson Laboratory. I managed all the mouse colonies, set up breedings, genotyped pups and maintained them. I performed the majority of all mouse experiments, which included mouse behavior testing, regular weight monitoring, neuronal phenotyping, dissections and sample preparation followed by experiments such as RT-qPCRs and PCRs. I performed CT scanning of the mice and we collaborated with Valerie Burke DeLeon for the craniofacial morphology analysis. I cryosectioned perfused brains of the mice under the guidance of Kimberley Jade Anderson, and I stained the brain sections, imaged and analyzed them. Additionally, I performed all statistical analysis of the phenotyping of the mice. I isolated neuronal progenitor cells (NPCs) from newborn pups from the mouse models, cultured and performed all cellular experiments. Kimberly assisted me with the CUT&Run experiment and Juan Ouyang assisted me with the ONT sequencing. I performed nearly all data and statistical analysis of the manuscript, including the major sequencing experiments: CUT&Run, RNA sequencing and genome-wide sequencing, as well as generating most of the figures. Lastly, I wrote manuscript in collaboration with my supervisor and co-authors.

Unpublished data: The unpublished data presented in this thesis are results from experiments using two other mouse models, the ROSA-CreER^{T2} and *Dct-LacZ* models, as well as additional cellular experiments using mNPCs. I performed all the experiments, analyzed the data and generated the figures.

1 Introduction

1.1 Neural crest development

The development of an embryo is defined by a series of critical events. A defining event in vertebrate development is the formation of the neural crest (NC) during neurulation. During this process the neural plate of the ectoderm folds inward and forms the neural tube, separating from the dorsal-most part, which will give rise to the NC (Dickinson et al., 1995). In humans, the NC formation starts around 30 days post fertilization, and in mice this begins at around embryonic day 9-9.5 (E9-9.5) (Betters et al., 2010; Schwarz et al., 2009; Vandamme & Berx, 2019). The neural crest gives rise to multipotent neural crest cells (NCCs) that can be divided by their regional position into cranial, vagal, trunk and sacral NCCs (Betters et al., 2010; Méndez-Maldonado et al., 2020). Based on their position, NCCs migrate to various places of the embryo along specific routes and differentiate into multiple cell types, contributing to the formation of various tissues (Bronner, 2012; Le Douarin & Kalcheim, 1999). This includes the glial and neuronal cells of the peripheral nervous system (PNS, e.g. the enteric nervous system), pigment cells of the epidermis, cells of the cranial cartilage and bone, as well as neuroendocrine cells (Bronner & LeDouarin, 2012). To migrate, a subset of NCCs undergo epithelial-to-mesenchymal transition (EMT) under the regulation of multiple transcription factors (TFs) such as Snail Family Transcriptional Repressors 1 and 2 (SNAI1 and SNAI2), Zinc Finger E-Box Binding Homeobox 2 (ZEB2), and Twist Family BHLH Transcription Factor 1 (TWIST1) during which the cells maintain their stem cell-like properties, lose their adhesion to nearby cells and gain the ability to invade distant regions of the embryo (Bronner & LeDouarin, 2012; Méndez-Maldonado et al., 2020; Thiery et al., 2009). The formation of the NC, and the migration and differentiation of the NCCs into different cell types, is tightly regulated by cell signaling molecules, such as Bone Morphogenetic Proteins (BMPs), Fibroblast Growth Factors (FGFs), Wntless and Int-1 ligands (Wnts), Notch ligands, and retinoic acid. It is also regulated by transcription factors such as Sex-determining region Y-Box Transcription Factors 9 and 10 (SOX9 and SOX10), Forkhead Box D3 (FOXD3), and Paired box 3 (PAX3), though its complex regulation is yet to be fully elucidated (Méndez-Maldonado et al., 2020).

1.2 Central nervous system development

The central nervous system (CNS) originates from the closed neural tube after neurulation of vertebrate embryogenesis. The rostral (front) end of the neural tube gives rise to the brain, and the caudal (tail) end gives rise to the spinal cord. At weeks three

to four of human gestation, the rostral end of the neural tube can be further divided into the three primary brain vesicles: the prosencephalon, the mesencephalon, and the rhombencephalon. By week five of gestation, these vesicles further divide into the five secondary brain vesicles. Specifically, the prosencephalon separates into the telencephalon and the diencephalon, while the rhombencephalon splits into the metencephalon and the myelencephalon. Briefly, the prosencephalon is a precursor to the forebrain, which gives rise to the cerebrum, thalamus, and hypothalamus. The mesencephalon is a precursor to the midbrain, and rhombencephalon to the hindbrain, which will develop into the pons, cerebellum, and medulla oblongata. These brain regions continue to develop throughout embryogenesis and into early childhood forming the complex structure of the brain (Sanes et al., 2019).

Signature characteristics of the human brain are the folds (sulci) and ridges (gyri) of the cerebral cortex, which increase the brain's surface area. Other key characteristics are the gray and white matter. The gray matter is located in the outermost layer of the brain called the neocortex, in deeper brain structures called the basal ganglia, and the innermost layer of the spinal cord. The gray matter is composed of neuronal cell bodies and their dendrites, which appear gray in dissected brains, and serve as relay and processing centers for information from the brain to the body and vice versa. The white matter comprises the rest of the brain (in between and around the basal ganglia structures) as well as the outer layer of the spinal cord and contains the myelinated axons of the neurons. The myelin is a fatty material that gives the white matter its white color in dissected brains. Myelin wraps around the axons to enhance the efficiency of the electrochemical signals between neurons that make up the pathways between different parts of the brain. At the center of the brain, there are intertwined cavities that divide into four main ventricles, termed the ventricular system of the brain. These ventricles are filled with cerebrospinal fluid (CSF), a crucial fluid system within the brain that renews frequently and is responsible for clearing out any cellular debris or waste material and is involved in transportation of hormones. The cellular area that develops nearest to the ventricles is called the ventricular zone (VZ) and it is a place where the majority of neural stem cells (NSCs) reside. In the early weeks of CNS development, the NSCs move from the VZ to the subventricular zone (SVZ) and differentiate into glial and neuronal progenitor cells (NPCs) (Brodal, 2010; Stiles & Jernigan, 2010). The glial progenitor cells give rise to astrocytes, oligodendrocytes, and ependymal cells while the NPCs differentiate into neurons. It was previously believed that the glial cells served solely as an adhesive for the neurons in the brain (hence the name 'glia' which originates from the Greek word for 'glue'), but now we know that the glial cells play very important and distinct roles in the brain (Purves et al., 2001; Stiles & Jernigan, 2010). Briefly, astrocytes are known to maintain the optimal environment for neuronal signaling in many different ways, e.g., through structural support, synapse maintenance, and water homeostasis. Oligodendrocytes are responsible for myelinating the axons of neurons while ependymal cells line the

ventricle walls and contribute to CSF production and waste management (Deng et al., 2023; Jäkel & Dimou, 2017). Lastly, microglia are a subtype of glial cells that originate from the yolk sack of the embryo and function as the brain's immune system. They have phagocytic properties and clear up any debris in their vicinity (Jäkel & Dimou, 2017; Purves et al., 2001).

1.2.1 Neuron development

Neurons are the information-processing cells of the brain. They intertwine and connect to form the intricate neuronal network responsible for our cognition, feelings, thoughts, actions, and physiological functions (Stiles & Jernigan, 2010). The brain is estimated to contain 85-100 billion neurons (Pakkenberg & Gundersen, 1997). Each neuron is able to form connections with more than a thousand other neurons, making the total number of neuron-to-neuron connections, called synapses, in the brain estimated to be around 600 trillion. As NPCs differentiate into neurons, they gain key cellular structures which are essential for their proper function. This includes e.g., their axon, dendrites, synapses and myelin sheath. Neurons communicate via electrochemical signals that are received at the dendrites, travel through the cell as an action potential, and are transmitted to the next neuron through the axon (Stiles & Jernigan, 2010). The signal moves between neurons through small junctions called synapses in the form of neurotransmitters (NTs). NTs are a vast group of molecules that relay the electrochemical signal but moreover, can amplify, diminish or change the signal (Teleanu et al., 2022). As neurons mature and become more specialized in their function within the CNS, they lose their ability to proliferate and enter a post-mitotic state (Stiles & Jernigan, 2010). Neurons within the CNS can broadly be categorized into two groups: projection neurons and interneurons. The projection neurons make long-distance connections to mediate information to, from, and within the CNS, while the interneurons span a shorter range and are involved in more localized processing of information. These groups can then be further classified into subclasses, based on their function and the NTs they release (J. Harris et al., 2015). Through maturation, neuronal cells express certain proteins that help researchers identify at which stage in differentiation they are. Generally, NSCs and NPCs are known to express nestin, Glial Fibrillary Acid Protein (GFAP), SOX2 and Brain Lipid-Binding Protein (BLBP). As cells proliferate, they lose their GFAP, SOX2 and BLBP expression and start expressing later-stage neuronal progenitor markers such as Doublecortin (DCX), Neuronal Differentiation 1 protein (NEUROD1). Immature neurons lose their nestin expression and gain expression of Neuronal Nuclei antigen (NeuN). Finally, mature neurons express markers such as Calbindin, NeuN, and Microtubule-Associated Protein 2 (MAP2) (Boldrini et al., 2018; Boutin et al., 2010; DeGiosio et al., 2022; K.-R. Kim et al., 2021).

1.2.2 Postnatal neurogenesis

The CNS continues to develop throughout mammalian embryogenesis until the early postnatal period and some parts even continue to develop into adulthood. In humans, neurogenesis starts at E42 and by mid-gestation, the majority of the fetus's neurons have been formed (Bayer et al., 1993; Stiles & Jernigan, 2010). Neurogenesis slows down considerably once majority of neurons have moved into a post-mitotic state. Yet, there are NSCs that reside within certain areas of the adult brain, such as the SVZ and the subgranular zone (SGZ) of the dentate gyrus (DG) of the hippocampus. These areas of the brain contribute to the brain's neuroplasticity by producing new neurons after birth, a process termed postnatal neurogenesis (Altman & Das, 1965, 1966; Kempermann & Gage, 1999). This malleability of the brain contrasts with the former belief that all mature neurons were formed *in utero* (Kozareva et al., 2019).

The hippocampus is a structure located in the temporal lobe and is a part of the limbic system. The hippocampal structure can be divided into the DG, Cornu Ammonis (CA), and subiculum. The CA region can be further divided into four subregions (CA1-CA4). These distinct regions interact to serve distinct functions, but the hippocampus as a whole plays a major role in the formation of new memories, memory consolidation, visuospatial memory, and learning (P. E. Gilbert & Brushfield, 2009; Kozareva et al., 2019). Adult hippocampal neurogenesis has been studied in humans, revealing its consistent presence throughout aging and its essential role in maintaining memory consolidation and stress recovery (Sahay et al., 2011; Schloesser et al., 2010).

1.3 Epigenetics and transcription

1.3.1 Overview

In 1958, Francis Crick coined and introduced his 'central dogma' of biology, outlining how the genetic information from deoxyribonucleic acid (DNA) is transcribed into messenger ribonucleic acid (mRNA) and then translated into proteins (Crick, 1958). At the time, transcription was an exciting research topic, with the role of the RNA polymerase II (RNA Pol II) in eukaryotic cells outlined by a few groups in the late 50s and 60s. There was also the discovery of transcription factors (TFs) that guide the RNA pol II to the transcription start sites (TSSs) of the DNA, advancing our knowledge of how gene expression is linked to protein synthesis, a process central to all cellular function (Burgess et al., 1969; Burley & Roeder, 1996; Weiss & Gladstone, 1959). Epigenetics, a term coined in the 1940s, refer to heritable changes in gene expression without changes in the DNA sequence itself (Waddington, 1942). In more recent years, this definition has been acknowledged to include non-heritable changes, since these mechanisms have been shown to play a role in non-dividing cells such as post-mitotic neurons. These epigenetic changes affect chromatin structure and accessibility and are

thought to alter gene expression. These changes can be divided into three major groups of DNA methylation, histone modifications or remodeling and RNA-based mechanisms (Gibney & Nolan, 2010). Transcription and epigenetics are tightly regulated and intertwined, with epigenetic modifications altering gene expression at the transcription level, by making the DNA more or less accessible to the transcription machinery (Alaskhar Alhamwe et al., 2018; Moore et al., 2013).

1.3.2 Alternative splicing

Splicing is a step in RNA processing that describes how the precursor mRNA (pre-mRNA), which includes non-coding introns and coding exons, is transformed into mature mRNA by removing introns at defined splice sites, by the spliceosome (Hoskins & Moore, 2012; Y. Wang et al., 2015). The human spliceosome consists of five small nuclear RNAs (snRNAs) linked to polypeptides to form small nuclear ribonucleoproteins (snRNPs), along with multiple other splicing factors and proteins, creating a highly dynamic complex, responsible for many steps involved in splicing (Hoskins & Moore, 2012; Jurica & Moore, 2003). Alternative splicing was first described in 1978 by Walter Gilbert and describes the process by which a single gene can be transcribed into multiple variations of mRNA and protein, termed isoforms, by splicing different combinations of exons together (W. Gilbert, 1978; Y. Wang et al., 2015). This can fundamentally change the proteins' function by including or excluding certain protein domains or leads to nonsense-mediated decay (NMD) of the mRNA or protein. Through alternative splicing, the protein diversity within a cell is highly enriched without the need to increase the number of genes within its genome (Graveley, 2001; Y. Wang et al., 2015). It has been shown that alternative splicing affects 95% of human genes and expands the ~20.000-25.000 human genes to more than 100.000 proteins estimated to be found in the human body, with the total numbers yet to be determined (Aken et al., 2017; Gaudet et al., 2017; International Human Genome Sequencing Consortium, 2004; Pan et al., 2008). This process is regulated by the interaction of RNA-binding regulatory proteins and regulatory sequences of the DNA itself. These sequences can be divided into splicing enhancers and splicing inhibitors, often located nearby a splice site to either increase or decrease the chance of an exon being spliced by hindering or allowing spliceosome access. Regulatory proteins include activators such as serine/arginine rich nuclear phosphoprotein family (SR protein family) and repressors such as heterogeneous nuclear ribonucleoproteins (hnRNPs) (Gabut et al., 2008; Jiang & Chen, 2021). Additionally, other factors such as micro-RNAs (miRNAs) and long non-coding RNAs (lncRNAs) have been shown to influence alternative splicing regulation (Luco & Misteli, 2011; Y. Wang et al., 2015). Furthermore, alternative splicing can be influenced by environmental factors such as temperature, stress and extracellular signaling molecules e.g. growth factors, hormones and cytokines (Graveley, 2001; Pelisch et al., 2005; Pleiss et al., 2007a, 2007b). Alternative splicing plays a key role in cellular differentiation and development, allowing for rapid changes in protein diversity and regulation of gene expression at the post-transcriptional level (Y.

Wang et al., 2015). Distinct alternative splicing patterns can depend on the cell type, as well as developmental context, both of which are essential for nervous system development and neurogenesis. It has been shown that there is a remarkably high rate of alternative splicing in the mammalian brain, which has been demonstrated to influence neuronal differentiation, morphogenesis, and synapse establishment and plasticity (Norris & Calarco, 2012; Yeo et al., 2004; Zheng & Black, 2013). Abnormalities in alternative splicing in the brain have been associated with multiple neurodevelopmental disorders (NDDs) such as Rett and Fragile X syndromes, as well as autism spectrum disorders (ASD) (R. Li et al., 2016; Shah & Richter, 2021; Stamova et al., 2013; Thacker et al., 2020)

1.3.3 DNA methylation

DNA methylation (DNAm) is an epigenetic modification essential for regulation of gene expression, genomic imprinting and X chromosome inactivation (Holliday & Pugh, 1975; Sharp et al., 2011; Wutz et al., 1997). The modification involves the placement of a methyl group to the 5' cytosine of the DNA, forming 5-methylcytosine (5mC), by DNA methyltransferases (DNMTs) (Christman, 2002; Holliday & Pugh, 1975; Ito et al., 2011). The methylation mark can then be removed passively during cell division, when the new cytosine is not re-methylated during DNA replication, or actively by DNA demethylases also known as ten-eleven translocation methylcytosine dioxygenase enzymes (TETs) (Tahiliani et al., 2009). This process primarily occurs at cytosine–guanine dinucleotides in the DNA, also called CpG sites (Holliday & Pugh, 1975). In the genome, CpG sites are usually methylated, with the exception of CpG clusters, also known as CpG islands (CGIs), which are predominantly located at gene promoter regions (Bird et al., 1985; Saxonov et al., 2006). Furthermore, unmethylated CpG islands have been associated with modified histones that enhance gene expression and are less tightly bound to nucleosomes, allowing TFs to directly access the DNA (Choi, 2010; Tazi & Bird, 1990). Subsequently, the addition of a methyl group at gene promoters or the first exon usually leads to gene silencing by inhibiting the binding of TFs at gene promoters, or by recruiting gene-repressing proteins (Brenet et al., 2011; Mohn et al., 2008; Moore et al., 2013). The placement and removal of DNAm is a very dynamic process that changes throughout cell development and differentiation, leading to distinct DNAm cellular profiles (Meissner et al., 2008; M. Weber et al., 2007). DNAm plays an important role in neuronal development and function, with the prevalence of 5mC being highest in the human brain compared to other tissues yet it makes up only around 1% of the genome (Ehrlich et al., 1982). Studies using knock-out (KO) animal models have demonstrated that the presence of intact DNAm machinery and the maintenance of DNAm at promoters of key neurogenesis and astroglialogenesis regulatory proteins are essential for proper pre- and postnatal development of the CNS (Fan et al., 2001, 2005). Interestingly, the continuous expression of DNMTs in post-mitotic neurons suggest a further role for DNAm in the adult brain (Goto et al., 1994; Inano et al., 2000). Depletion of DNMTs in mouse models has been shown to cause

defects in neuronal synaptic plasticity, memory and learning (Feng et al., 2010; Nguyen et al., 2007).

1.3.3.1 DNAm regulators

DNAm regulators include three groups of enzymes that place, erase, or read the 5mC mark. The family of proteins involved in placing the 5mC mark are known as DNMTs and are four in total; DNMT1, DNMT3A, DNMT3B and DNMT3L (Moore et al., 2013). DNMT3A and DNMT3B can place the methylation mark by transferring a methyl group from a S-adenosyl methionine (SAM) to the fifth carbon of a cytosine residue forming the 5mC (Okano et al., 1999). DNMT3L lacks the catalytic domain shared by the other DNMTs but is needed to stimulate and assist the methyltransferase activity of DNMT3A and DNMT3B (Hata et al., 2002; Suetake et al., 2004). Furthermore, it has been shown to be essential for genomic imprinting and X chromosome inactivation (Kaneda et al., 2004; Zamudio et al., 2011). DNMT1 is responsible for maintaining the 5mC mark during DNA replication during cell division, where the newly added cytosine is momentarily unmethylated (Goto et al., 1994; Pradhan et al., 1999). The TET protein family consisting of TET1, TET2 and TET3 are enzymes responsible for removing the 5mC mark. First, a hydroxyl group is added to the 5mC to form 5-hydroxymethylcytosine (5hmC), which is then oxidized to 5-formylcytosine and finally to 5-carboxycytosine (Ito et al., 2011; Tahiliani et al., 2009). Additionally, there are several families of enzymes that recognize and bind with high affinity to the 5mC mark, affecting gene expression and chromatin structure by impairing binding of TFs to the CpG site. These include the methyl-CpG-binding domain proteins (MBD proteins), ubiquitin-like containing PHD- and RING-finger domain proteins (UHRF proteins) and zinc-finger proteins (Bostick et al., 2007; Filion et al., 2006; Meehan et al., 1989; Moore et al., 2013).

1.3.4 Histone modification machinery

Histones are positively charged proteins that tightly pack the negatively charged DNA strand into chromatin, forming nucleosomes. Around 147 base pairs (bp) of DNA wraps around each nucleosome that consists of two sets of core histone proteins called H2A, H2B, H3, and H4, all of which are highly conserved across species (Olins & Olins, 1974; Wolffe, 1998; Zlatanova et al., 1998). Protruding from the nucleosome are the N-terminal peptide sequences of the histones, called the histone tails, that are subject to chemical post-translational modifications (PTMs) (du Preez & Patterton, 2013). The PTMs of histone tail residues and rearrangements of the chromatin comprise a large group of epigenetic modifications also known as histone modifications. Like DNAm, they play a key role in the regulation of gene expression by influencing DNA accessibility for the transcriptional machinery (Allfrey et al., 1964; Wolffe, 1998). Furthermore, histone PTMs can influence transcription by affecting alternative splicing of the mRNA (Graveley, 2001). The chemical modifications of the histone residues

involve e.g., methylation, acetylation, phosphorylation, sumoylation and ubiquitination, that can either be activating or repressing marks of transcriptional activity, depending on their placement and type (Allfrey et al., 1964; Bannister & Kouzarides, 2011). Furthermore, the vast and diverse group of enzymes that regulate histone modifications can be classified into writers, erasers and readers of histone marks, and remodelers of chromatin. Notably, some of these enzymes have been demonstrated to have overlapping functions, with some writers, erasers, and remodelers additionally having a histone modification 'reading' domain (Boukas et al., 2019).

1.3.4.1 Histone acetylation

Histone acetylation is regulated by two groups of enzymes with contrasting functions, the histone acetyltransferases (HATs) that add acetyl groups to residues of histone tails and histone deacetylases (HDACs) that remove them (Alaskhar Alhamwe et al., 2018; Brownell & Allis, 1995; Kleff et al., 1995; Taunton et al., 1996). HATs catalyze the transfer of an acetyl group, with the assistance of acetyl coenzyme A (acetyl-CoA), to the ϵ -amino group of lysine (K) residues on histone tails (Allfrey et al., 1964, 1968). With the addition of an acetyl group to the positively charged lysine residues, their charge becomes neutral and they partially lose their affinity for the negatively charged DNA. This leads to less tightly bound DNA and more open chromatin, resulting in better access for the transcriptional machinery. Histone acetylation is therefore generally linked to increased gene expression (Brownell et al., 1996; Yang & Seto, 2007).

1.3.4.2 Histone methylation

Another major histone modification is methylation of histone tail residues of lysines (K) and arginines (R), placed by histone methyltransferases (HMTs) (Bannister & Kouzarides, 2011). HMTs can add one, two or three methyl groups to the histone tail residues from cofactor SAM. They can be divided into two main groups of enzymes: the protein arginine methyltransferases (PRMTs), which are responsible for placing a mono-, symmetrical or asymmetrical di-methylation mark on arginine residues, and lysine methyltransferases (KMTs) that place mono-, di- or tri-methylation mark on lysine residues. In contrast to histone acetylation, histone methylation does not affect the charge of the histone tails but works to attract other regulatory proteins to the chromatin and is therefore linked to both repressive and active gene transcription (Alaskhar Alhamwe et al., 2018; Bannister & Kouzarides, 2011). Histone demethylases (HDMs) are a group of enzymes responsible for removing the methylation mark and include lysine-specific demethylases (LSDs or KDMs) and Jumonji C domain-containing proteins (JMJDs) (Alaskhar Alhamwe et al., 2018; Shi et al., 2004; Tsukada et al., 2006).

1.3.5 Promoters and enhancers

Fundamental regulatory elements in transcription are distal- and proximal regulatory elements in the DNA sequence, called enhancers and promoters, respectively. These are sites of transcriptional and epigenetic regulation (Haberle & Stark, 2018). Promoters are regions that recruit and directly bind RNA Pol II and the general transcription factors (GTFs) needed to initiate gene transcription. They are located upstream or at the 5' end of a gene, and usually consist of a core promoter and a proximal promoter element. The core promoter commonly contains a binding site for the TATA binding protein (TBP) of the transcriptional machinery, known as a TATA box, an initiator sequence (Inr) including the TSS and the downstream promoter element (DPE) (Klein & Struhl, 1994; Nikolov & Burley, 1994; Smale & Baltimore, 1989). Proximal promoter elements are located upstream of the promoter and contain DNA sequences that bind to transcriptional activators, promoting gene expression (Hampsey, 1998). Enhancers are distal cis-regulatory elements that can be located far from the promoter region of a gene they influence. They can be upstream or downstream of the target gene, or within its introns. In three-dimensional space they are located in proximity to the target promoter region bind through Mediator/Cohesin complexes and their associated proteins (Dixon et al., 2012; Kagey et al., 2010). Enhancers, as the name suggests, can bind transcriptional activators and regulatory proteins, enhancing the chance of gene transcription starting or increasing the abundance of transcripts (Haberle & Stark, 2018; Karnuta & Scacheri, 2018). Enhancers and promoter regions interact with the histone machinery, for example, active promoters are often marked by histone 3 lysine 4 trimethylation (H3K4me3), poised enhancers by H3K4 monomethylation (H3K4me1) and histone 3 lysine 27 trimethylation (H3K27me3), and active enhancers by H3K4me1 and H3K27 acetylation (H3K27ac) (Heintzman et al., 2007; Zentner et al., 2011). The epigenetic state of enhancers and promoters changes throughout cellular development and differentiation, providing the cell with a precise fine-tuning system of gene expression, further linking epigenetic modifications to transcriptional activity (Haberle & Stark, 2018; Hampsey, 1998).

1.3.6 Interaction and compensation through epigenetic machineries

The epigenetic machinery is a highly complex and intricate system consisting of hundreds of proteins regulating and fine-tuning the transcription of the genome within a cell. There is a strong link between DNAm and PTMs of histones, with DNAm regulators being able to directly bind to histone modifiers and modifications, often to suppress gene expression, or to block their binding (Bannister & Kouzarides, 2011). Histone marks can greatly influence each other by competing to bind the same substrates or by affecting substrates within close proximity, creating a network of regulatory signals. An example of this is the blocking of H3K9me2/3 by heterochromatin protein 1 (HP1) when histone 3 serine 10 (H3S10) is phosphorylated (Fischle et al., 2005). An imbalance in the chromatin state can disrupt the equilibrium

of open and closed chromatin, and recent studies have speculated whether this disruption can cause the cell to compensate by other means (Moore et al., 2013; Y. Wang et al., 2015). To further support the importance of the balance of histone marks, studies have demonstrated that knocking out opposing epigenetic regulators of the same mark can restore the balance (Fahrner & Bjornsson, 2014; Vallianatos et al., 2020). An example of this is a study that showed an impairment of cellular differentiation in KMT2D-deficient embryonic stem cells (ESCs), that was restored by the knock-down (KD) of the opposing enzyme LSD1 (Cao et al., 2018).

1.4 Lysine methyl transferase 2A (KMT2A)

1.4.1 KMT2s

There are eight classes of lysine methyltransferases (KMTs) known in humans, KMT1-8, all of which include an enzymatic SET domain, apart from DOT1L (also known as KMT4). The second class (of eight) KMTs is highly conserved and known as lysine methyltransferases 2 (KMT2s) (Alaskhar Alhamwe et al., 2018; Allis et al., 2007). The KMT2 family contains eight proteins, KMT2A-H, six of which have a *Drosophila melanogaster* homolog (Rao & Dou, 2015). The nomenclature for these enzymes has changed throughout history and they will be referred to in this thesis by their most widely used names, according to Table 1 (Allis et al., 2007).

Table 1. Human lysine methyltransferase 2 (KMT2) nomenclature.

Gene name in thesis	Other gene names in literature
<i>KMT2A</i>	<i>MLL, MLL1, ALL, HRX, TRX1</i>
<i>KMT2B</i>	<i>MLL2, HRX2, TRX2, sometimes MLL4</i>
<i>KMT2C</i>	<i>MLL3</i>
<i>KMT2D</i>	<i>MLL4, ALR, sometimes MLL2</i>
<i>KMT2E</i>	<i>MLL5</i>
<i>SETD1A</i>	<i>KMT2F</i>
<i>SETD1B</i>	<i>KMT2G</i>
<i>ASH1L</i>	<i>KMT2H</i>

The KMT2s all have a slightly different protein structure and functions, yet they share similar qualities. They are often grouped in pairs, based on which *Drosophila* homolog they derived from. This includes KMT2A and KMT2B, KMT2C and KMT2D,

and SETD1A and SETD1B. These six proteins function in COMPASS or COMPASS-like complexes of multiple proteins to enhance their otherwise weak enzymatic functions, all of which share the subunits of ASH2 like protein (ASH2L), Retinoblastoma Binding Protein 5 (RBBP5), WD repeat domain 5 (WDR5) and DPY30 (Rao & Dou, 2015). The KMTs are highly specific enzymes in which lysine residues of the histone tails they methylate. The KMT2s are collectively responsible for methylating histone 3 lysine 4 (H3K4me), an epigenetic mark of increased gene expression, however, they have been shown to be highly specific in which substrate they methylate within the genome (Rao & Dou, 2015; Schuettengruber et al., 2011). This is with the exception of KMT2E, which contains a SET domain but lacks methyltransferase activity (Mas-y-Mas et al., 2016). KMT2C and KMT2D have been shown to be responsible for the majority of H3K4me1 at enhancer regions, while SETD1A and SETD1B are highly enriched at promoters marked by H3K4me2/3. Interestingly, KMT2A and KMT2B have been shown to bind to highly specific promoters as well as enhancer regions (Poreba et al., 2020; P. Wang et al., 2009). To further support the importance of KMT2s in development, studies in animal models have revealed that they all show lethality in homozygosity (Bledau et al., 2014; Glaser et al., 2006; Lee et al., 2013; Y. J. Li et al., 2023; Yu et al., 1995; Zhu et al., 2016).

1.4.2 KMT2A structure

KMT2A was the first KMT2 protein to be discovered in 1992 and found to be a homolog of the *Drosophila melanogaster* trithorax (*trx*) gene and the *Saccharomyces cerevisiae* Set gene (Guenther et al., 2005; Tkachuk et al., 1992). As discussed above, KMT2A and KMT2B are in a COMPASS-like complex, and unlike the other KMT2s, they also bind MENIN and Host Cell Factor 1 or 2 (HCF1 or HCF2) (T. Nakamura et al., 2002; Yokoyama et al., 2004). The gene, *KMT2A*, is positioned on human chromosome 11q23 (mouse chromosome 9) and encodes the large 3969 amino acid and ~500 kilodalton (kDa), multidomain KMT2A protein. These domains include three AT hooks, a CXXC domain, four plant homeodomains (PHD domain), FYR-N and FYR-C domains, a WIN motif, a bromodomain, a TAD domain, the enzymatic SET domain and a post-SET domain (Rao & Dou, 2015). In general, the PHD fingers mediate protein-protein interactions, and the WIN motif binds the WDR5 protein, which is needed for the protein's catalytic activity (Fair et al., 2001; Patel et al., 2008). Additionally, KMT2A is post-translationally regulated by the enzyme Taspase-1 that cleaves the protein into a 320 kDa N-terminal and a 180 kDa C-terminal product at one of two cleavage sites, CS1 or CS2. These protein products subsequently heterodimerize again through their FYR-N and FYR-C domains, to serve the protein's enzymatic function (Hsieh et al., 2003). KMT2A has two different direct DNA-binding domains, AT hooks that bind AT-rich DNA and a CXXC domain that binds unmethylated CpG sites that frequently occur at promoter regions (Allen et al., 2006; Zeleznik-Le et al., 1994). Furthermore, the binding of KMT2A to promoter regions is stabilized by the PHD3 domain and a bromodomain that have a high affinity for H3K4me3 and lysine acetylation, which are

found in abundance around TSS (Z. Wang et al., 2010). KMT2A is able to place mono-, di-, and trimethylation marks on H3K4 (H3K4me1/2/3) through its SET domain and even though it is both found at enhancers and promoter regions, only 5% of global H3K4me3 are fully KMT2A-dependent (Guenther et al., 2005; T. Nakamura et al., 2002; P. Wang et al., 2009).

1.4.3 KMT2A functional role

KMT2A has been involved in many essential biological processes such as transcription, cell differentiation, mitotic retention during cell division, embryonic development, and hematopoiesis (Blobel et al., 2009; Ernst et al., 2004; Yu et al., 1995). In cellular and mouse models, KMT2A is essential to lymphocyte differentiation, more specifically in B-cell differentiation, and is involved in fetal liver hematopoietic stem cell (HSC) generation (Ernst et al., 2004). Somatic mutations in KMT2A are known to play a role in cancer (Kandoth et al., 2013). In fact, KMT2A was originally discovered through cancer research and called ALL (acute lymphoblastic leukemia) or MLL (mixed lineage leukemia) due to its implications in the disease (Ayton & Cleary, 2001; Rao & Dou, 2015; Tkachuk et al., 1992). The most common chromosomal rearrangement type in these cancers is KMT2A translocations, involving the translocation of two chromosomes and the formation of KMT2A-fusion proteins. These leukemias make up around 10% of human leukemias and are associated with poor prognosis in pediatric patients (Mirro et al., 1986; Tkachuk et al., 1992). With over 70 fusion partners known to date, KMT2A-rearranged leukemias remain a significant challenge in cancer research and treatment (Rao & Dou, 2015).

1.4.3.1 KMT2A in development

As previously discussed, KMT2A is important in embryonic development, with its expression starting as early as E7 (Yu et al., 1995). This notion is further reinforced by the fact that homozygous KO of mouse *Kmt2a*^{-/-} leads to embryonic lethality around E10.5-11.5 (Yagi et al., 1998; Yu et al., 1995, 1998). However, mice with targeted homozygous disruptions solely of the SET domain were viable with minor defects, indicating that the KMT2A enzymatic activity by itself is not essential for embryonic development (Yu et al., 1995). Furthermore, prior research has shown KMT2A to play a crucial role in Homeobox (*Hox*) gene regulation, initiation, and maintenance during embryonic development, all of which are essential for embryo pattern formation to establish segment identity in mammals (Terranova et al., 2006; P. Wang et al., 2009; Yu et al., 1995). More specifically, KMT2A has been shown to regulate e.g., *Hoxa7*, *Hoxa9*, *Hoxc8*, and *Hoxc9*, with studies demonstrating DNAm defects at some overlapping loci, highlighting the dynamic epigenetic effects of KMT2A (Milne et al., 2002; Terranova et al., 2006; P. Wang et al., 2009; Yu et al., 1995, 1998). Subsequently, mice with heterozygous KO of *Kmt2a*^{+/-} have demonstrated defects in their axial skeletal structure and sternum (Yu et al., 1995).

1.4.3.2 *KMT2A in the brain*

KMT2A is widely expressed in the brain, where it is needed to maintain the H3K4me1/3 landscape in neurons and plays an essential role in postnatal neurogenesis in the SVZ and hippocampus (Kerimoglu et al., 2017; Lim et al., 2009). Mouse neurons with homozygous *Kmt2a* conditional knock-out (cKO) in the postnatal forebrain showed decreased synaptic plasticity, while pyramidal neurons from heterozygous *Kmt2a* KO mice showed reduced spine density but not dendritic length (Jakovcevski et al., 2015; Vallianatos et al., 2020). *In vivo*, studies have shown that mice with homozygous cKO of *Kmt2a* in adult excitatory forebrain neurons and prefrontal cortex display anxiety-like behavior, defects in associative, spatial, and working memory, as well as heterozygous KO of *Kmt2a* demonstrates defects in contextual fear memory (Gupta et al., 2010; Jakovcevski et al., 2015; Kerimoglu et al., 2017). Collectively, this demonstrates the importance of KMT2A in neuronal structure, memory consolidation and visuospatial memory.

There is some speculation about the redundancy to the vast epigenetic regulatory machinery, with multiple proteins serving similar functions. However, despite many similarities between KMT2A and KMT2B, such as their binding partners and protein structure, they seem to largely cover distinct areas of the brain, as well as regulate mostly non-overlapping genes. This is further supported by different memory defects in mice once behaviorally tested (Kerimoglu et al., 2017). In contrast, KDM5C, a lysine-specific demethylase that is responsible for removing the H3K4me1/2/3 marks widely co-exists with KMT2A in the brain. Interestingly when these opposing enzymes were crossed in mice, creating double mutants with heterozygous KO of *Kmt2a* and *Kdm5c*, the researchers found the mice to have restored defects in neuronal structure and memory function. At the cellular level, they observed restored gene expression in a subset of differentially expressed genes as well as H3K4me3 peaks. This importantly highlights the adaptability of the epigenetic system and opposing enzymatic functions (Vallianatos et al., 2020). Eventhough much is known about KMT2A, this complex protein is speculated to serve many functions that have yet to be discovered.

1.5 Epigenetics and neurological disorders

There is an emerging field within epigenetics termed neuroepigenetics that aims to address the effects of epigenetics on neuronal function, development, cellular morphology, as well as elucidating the role of epigenetics in non-dividing neurons in the adult brain (Day & Sweatt, 2010). Within neuroepigenetics, neurological diseases attributed in part or fully to epigenetic dysfunction are steadily growing, including complex neurodegenerative diseases, psychiatric disorders and congenital cognitive disorders (Grezenko et al., 2023). For example, in the neurodegenerative Alzheimer's disease (AD), research has shown that in a diseased state, DNAm is aberrant along with increased deacetylation of histones, both of which are associated with a more closed

chromatin state and repressed gene expression. The dysfunction in the epigenetic landscape is believed to lead to cognitive impairment and neurodegeneration (Sharma et al., 2020). Similarly, in Huntington's disease (HD), disruption in normal histone acetyltransferase activity has been demonstrated to affect disease progression (Sadri-Vakili & Cha, 2006). Epigenetics have also been found to play a role in psychiatric disorders such as schizophrenia, depression and drug addiction. There are many factors to consider for such complex diseases, e.g., genetic predisposition, genetic defects in epigenetic factors, as well as environmental risk factors that influence the epigenetic landscape such as DNAm (Grezenko et al., 2023; Moore et al., 2013). Lastly, a large number of congenital NDDs have been linked to genetic mutations in epigenetic regulators (discussed further in section 1.5.1) (Fahrner & Bjornsson, 2014). As previously discussed, the epigenetic machinery mediates dynamic and malleable processes in the brain, even in post-mitotic neurons (Inano et al., 2000; Moore et al., 2013). This allows for a promising opportunity for therapeutic intervention, e.g., the use of histone deacetylase inhibitors in the treatment of HD or NDDs in mice (Bjornsson et al., 2014; Mielcarek et al., 2011). Furthermore, epigenetic signatures of various neurological diseases, such as aberrant DNAm profiles, show promising correlation to disease diagnosis, prognosis and clinical outcomes (Younesian et al., 2022).

1.5.1 Mendelian disorders of the epigenetic machinery (MDEMs)

1.5.1.1 Genetic basis

Mendelian disorders of the epigenetic machinery (MDEMs) are a group of disorders caused by mutations in genes encoding epigenetic regulators. These include writers, readers, erasers and remodelers in the DNAm and histone PTM machineries (Fahrner & Bjornsson, 2014). The genes encoding these epigenetic machinery regulators (EM genes) share some common features. It has been shown that a subset of EM genes are highly co-expressed in various tissues (including the brain), in a non-tissue specific manner and thus likely contributing to the epigenetic homeostasis. EM genes have also been demonstrated to be highly intolerant to loss-of-function (LoF) variants, demonstrating haploinsufficiency. With the number of known epigenetic regulators continuously expanding (currently there are around 300 regulators known), the subgroup of those associated with MDEMs also keeps growing (Boukas et al., 2019). Currently, defects in more than 85 regulators have been linked to MDEMs (Fahrner & Bjornsson, 2019; J. R. Harris et al., 2024). MDEMs had been described in case reports and larger cohorts, so in 2014 Fahrner et al. proposed the umbrella term "MDEMs" for these disorders, since they tend to share many genetic and phenotypic characteristics. The majority of MDEMs are linked to autosomal genes and can generally be divided into two groups: MDEMs that are inherited in a dominant manner and affect highly dosage sensitive genes and then MDEMs that are inherited in a recessive manner that affect less dosage sensitive genes (Fahrner & Bjornsson, 2019). Research groups have theorized that the imbalance of the epigenetic state leads to downstream gene

expression dysregulation which plays a central role in the underlying MDEM disease pathogenesis (Fahrner & Bjornsson, 2014). An example of this can be seen in Kabuki syndrome (KS), where a disruption of either the H3K4me writer KMT2D or the H3K4 demethylase KDM6A result in an overlapping phenotype (Miyake et al., 2013; S. B. Ng et al., 2010). Interestingly, in animal models of MDEMs with heterozygous loss of EM genes, the changes in gene expression tend to be subtle, suggesting a potential epigenetic compensation mechanism by the cells and further highlighting the importance of epigenetic balance (Bjornsson et al., 2014; Tudor et al., 2002; Vallianatos et al., 2020).

1.5.1.2 Clinical features

In addition to sharing a similar genetic basis, MDEMs commonly present with shared clinical phenotypes. The most prevalent features of MDEMs are intellectual disability (ID), observed in about 80% of disorders, and growth abnormalities, observed in about 60% of disorders (Fahrner & Bjornsson, 2019). Some more specific features found in many of these disorders include hypotonia, characteristic facial features, obesity, immune dysfunction and cardiac malformations (Fahrner & Bjornsson, 2014, 2019). The overlap of phenotypes leaves researchers questioning whether there might be shared disease mechanisms, such as common downstream targets of gene dysregulation (Fahrner & Bjornsson, 2014). Interestingly, there is a subset of MDEMs that have been linked to defects in neural crest development. These disorders often present with symptoms of dysfunction in neural crest-derived organs, such as cardiac malformations and craniofacial defects (Gabriele et al., 2018). Examples of this are CHARGE syndrome and KS, caused by mutations in the chromatin remodeler CHD7 and histone methyltransferase KMT2D, respectively (Bouazoune & Kingston, 2012). Research has demonstrated the role of these epigenetic regulators in NCC formation and migration (Bajpai et al., 2010; Schwenty-Lara et al., 2020). Furthermore, in mouse models of KS, studies have specifically linked the craniofacial dysmorphism to defects in the neural crest (Shpargel et al., 2017, 2020). Taken together, this might indicate that there are crucial timepoints in embryonic development, where dysfunction in certain embryonic tissues leads to a shared downstream dysregulation and an overlapping phenotype.

In addition to phenotypic description and clinical genetic tests, individuals have also been diagnosed with MDEMs through DNA methylation array technology. Several MDEMs have distinct DNAm signatures, and utilizing this accurate testing method has helped exponentially in the differential diagnosis of syndromes with highly similar phenotypes, e.g. KS and CHARGE syndrome (Butcher et al., 2017). Furthermore, this has also helped to elucidate whether a *de novo* mutation in an EM gene is pathogenic or not, which is currently a major challenge to clinical geneticists (Fahrner & Bjornsson, 2019).

1.5.1.3 Postnatal neurological malleability

Neurological defects in MDEMs are very common, with ID being one of the most common manifestations (Fahrner & Bjornsson, 2014). In stark contrast to the former belief that ID was not a treatable disorder, multiple animal models of MDEMs displaying ID have now demonstrated a reversal of their neurological phenotypes postnatally (Alarcón et al., 2004; Bjornsson et al., 2014; Guy et al., 2007; Korzus et al., 2004). One of the first disorders to demonstrate this malleability was a mouse model of Rett syndrome, a disorder that leads to autism spectrum disorder (ASD), with a conditional knock-in (cKI) of *Mecp2*. The neurological defects of the model, such as decreased long-term potentiation (LTP) were reversed upon postnatal regain of MECP2 expression (Guy et al., 2007). These results fundamentally changed the way researchers approached MDEMs and since then, other MDEMs have also been shown to exhibit postnatal malleability of their neurological dysfunction. So far, these syndromes include for example KS, Rubinstein-Taybi (RST) syndrome and Kleeftstra syndrome (Alarcón et al., 2004; Bjornsson et al., 2014; T. Nakamura et al., 2024). These findings open up a great avenue for research into the therapeutic window of opportunity for rare genetic causes of ID.

1.5.2 Wiedemann-Steiner syndrome

Wiedemann-Steiner syndrome (WDSTS) is a MDEM that was first described in 1989 by a German pediatrician Hans Rudolf Wiedemann. In 2000, it was described again by a Brazilian geneticist Carlos Eduardo Steiner (Steiner & Marques, 2000; Wiedemann et al., 1989). From these initial case reports, the disorder was named Wiedemann-Steiner syndrome. In the years prior to genetic testing, the individuals received a WDSTS diagnosis based on their clinical features. Given the shared overlap between WDSTS and KS, individuals would occasionally get diagnosed with atypical KS (Miyake et al., 2013). In the year 2012, Jones et al. first described the molecular basis of the disease, linking WDSTS to heterozygous *de novo* disease causing mutations in the KMT2A gene (Jones et al., 2012). Currently, the most common way of diagnosis is through exome sequencing and whole genome sequencing. WDSTS causing variants are most commonly frameshift mutations (~40%), followed by nonsense mutations (~30%), missense variants (MVs, ~20%) and splice variants (~10%). WDSTS is inherited in an autosomal dominant manner, with most cases reported being due to *de novo* variants, i.e. not found in the individuals' parents (Sheppard & Quintero-Rivera, 2022).

WDSTS is a congenital disorder that widely affects the body, with common characteristics including ID, distinctive facial features, pre- and postnatal growth retardation, hypotonia and hypertrichosis (Jones et al., 2012; Sheppard et al., 2021). One of the most consistently reported issues in individuals with WDSTS are ID or developmental delay (DD), affecting nearly all individuals diagnosed or approximately 97% of cases (Sheppard & Quintero-Rivera, 2022). Furthermore, there have been

reports of hyperactivity, aggressive behavior and ASD in individuals with WDSTS. A cohort of 52 individuals with WDSTS underwent brain imaging, where ~13.5% of cases demonstrated structural abnormalities such as an altered corpus callosum and abnormal myelination (Sheppard et al., 2021). The characteristic craniofacial features reported in the majority of cases (63-69%) include narrow palpebral fissures, hypertelorism, widened nasal bridge, and a broad tip of the nose. Other common facial features include lateral flare of the eyebrows, downslanted palpebral fissures and drooping eyelid(s). Additionally, dental problems such as advanced dental age were present in a majority of individuals (57.7%). Pre and postnatal growth retardation are common in individuals with WDSTS, with a majority of cases born below the 25th percentile and later on in life present with a short stature. Hypotonia was present in 72.4% of individuals with WDSTS. A majority of cases presented with failure to thrive (FTT) (67.7%), often associated with constipation (63.8%) or feeding difficulties (66.3%), with tube feeding required in about one fourth of cases (25.5%). In addition, growth hormone deficiency has been reported in a few cases, some of which have been linked to pituitary abnormalities detected by magnetic resonance imaging (MRI) (Sheppard et al., 2021). One of WDSTS's distinctive features, separating the disorder from other differential diagnosis, is the excessive growth of non-androgenous related terminal hair known as hypertrichosis terminalis (Jones et al., 2012). This increased hair growth is seen in the majority of individuals affected by WDSTS and is most notable on the face as long eyelashes and thick eyebrows, as well as on the back and of the elbows (hypertrichosis cubiti). Other issues reported in WDSTS include sleep apnea, cardiac abnormalities, seizures, strabismus and musculoskeletal anomalies. However, this is not an exhaustive list of symptoms reported, highlighting the extensive and wide range of phenotypes linked to WDSTS. There has been one genotype-phenotype study that has demonstrated a correlation between LoF variants causing WDSTS to an increased likelihood of hypotonia, while individuals with non-LoF are more likely to have seizures (Sheppard et al., 2021; Sheppard & Quintero-Rivera, 2022).

With a whole spectrum of symptoms, ranging in severity, it is clear that individuals with WDSTS can be heavily affected during their daily lives. Caregivers of these individuals have reported behavioral problems and severe gastrointestinal (GI) symptoms to be the most debilitating factors of the affected individuals' lives. Intelligence quotient (IQ) scores of individuals with WDSTS ranges from 40 to 85 (with a median of 65), with a majority of individuals requiring some form of educational support. With less than 500 cases diagnosed to date which are predominantly children, estimating life expectancy can be a challenge. Furthermore, with the genetic basis of WDSTS only discovered a little more than 10 years ago and limited access to genetic testing in many countries, it is estimated that the syndrome is underdiagnosed in the world's population. Currently, there is no cure or standard treatment regimen for WDSTS. Cases of WDSTS are evaluated on an individual basis and management of the symptoms present is recommended. This may include ID/DD management, feeding

therapy, stool softeners, orthopedic treatment, growth hormone therapy etc, depending on the availability of health care to each individual (Sheppard & Quintero-Rivera, 2022).

As previously mentioned, WDSTS and KS share a considerable phenotypic overlap, such as ID, craniofacial abnormalities, growth retardation, hypotonia and FTT. Although both of these disorders affect epigenetic regulators that target the same lysine residue (H3K4me), leading researchers to hypothesize that there might be common downstream targets and mechanisms, there are some several key differences. On a phenotypic level, hypertrichosis and severe constipation are uncommon symptoms in individuals with KS and there are only a few cases of hearing loss in individuals with WDSTS (Boniel et al., 2021; Sheppard et al., 2021). An interesting difference on a protein level is that although these enzymes (KMT2A and KMT2D) serve a similar function in methylating the H3K4 substrate, mice with a truncated SET domain of the proteins showed that *Kmt2a*^{SET/SET} mice were fertile with minor defects, while *Kmt2d*^{SET/SET} were lethal (Bjornsson et al., 2014; Yu et al., 1995). This highlights some of the unknown differences between these seemingly similar proteins.

1.6 Therapeutic development for rare diseases

1.6.1 Overview of therapeutic approaches

Although each rare disease only makes up a few hundred to a few hundred thousand cases, collectively they are estimated to affect around 400 million people worldwide, while some estimations stating they affect one in every ten people. With rare diseases affecting such a big part of the world's population and only approximately 5% of them having an FDA approved treatment, there is great need for improved research in this field ("RARE Disease Facts," 2024). Different strategies for therapeutic approaches for rare diseases include small molecule drugs, protein-based drugs, gene editing strategies, gene therapy and RNA-based therapies. Small molecules have been well established as drug agents due to their many advantages, such as stability, cost effectiveness, multiple administration routes and easily controlled dosing, leading to FDA approval for diseases such as Cystic Fibrosis (CF) (Keating et al., 2018; Tambuyzer et al., 2020). Protein based therapies aim to restore protein levels in diseases that are caused by a lack of a particular protein or accumulation of a defective protein product. Examples include the supplement of blood clotting factor VIII and factor IX in treatment for hemophilia, as well as administration of exogenous enzymes (enzyme replacement therapy, ERT) in treatment for Gaucher disease. A considerable downside of ERTs is the high costs of enzyme production and purification. RNA based therapies include antisense oligonucleotides (ASOs) and small interfering RNAs (siRNAs), that aim to silence protein regulation at the mRNA level. These therapies have proved effective in a few neurological syndromes such as hereditary amyloidosis and spinal muscular atrophy (SMA), and show much promise for others such as HD

(Tambuyzer et al., 2020). However, ASO's ability to cross the blood brain barrier is restricted and thus requires a more invasive method of induction such as intraventricular or intrathecal administration (Q. Li, 2020).

Gene therapy is a technology that aims to deliver a transgene encoding a protein or RNA into cells, either to be expressed endogenously or to be integrated into the cell's DNA. In the case of diseases caused by a lack of a specific protein, gene therapy can increase the protein production by delivering a transgene under an active promoter. A common way to deliver a transgene is through viral vectors such as adeno-associated virus (AAV) vector, retrovirus vector and lentivirus vector. This can be delivered either *in vivo* to the desired target tissue or *in vitro*. Alternatively, it can be delivered *ex vivo*, where cells are modified outside of the body and then transplanted back, also known as cell therapy. One restriction of gene therapy is the size of which the transgene can be to fit into the vector. AAV vectors can deliver a product up to 5 kilobases (kb) in size, including promoter region and other necessary elements. For larger proteins, this will prove a challenge. Gene therapy is rapidly advancing, with multiple FDA approved treatments for diseases such as SMA, sickle cell disease, hemophilia and beta-thalassemia (Tambuyzer et al., 2020).

Finally, gene editing approaches, such as CRISPR-Cas9 technology (acronym for Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR Associated Protein 9), target the endogenous locus directly for editing. Briefly, the CRISPR-Cas9 system targets the DNA sequence of interest at a protospacer adjacent motifs (PAMs), with a single guide RNA (sgRNA), where the Cas9 cuts the DNA, causing a double-stranded DNA break, allowing for precise gene editing (Redman et al., 2016). This system has revolutionized biomedical science in the field of gene editing, allowing scientists to insert, delete and manipulate genome in a highly accurate manner (Tambuyzer et al., 2020). Delivering the CRISPR/Cas9 system to the affected cells can prove a big challenge. Some additional methods include viral vectors, extracellular vesicle-based systems, physical delivery methods, and chemical delivery methods. Considerable disadvantages to CRISPR-Cas9 gene editing include the potential for off-target effects, which can occur at sequences nearly homologous to the target site (Koo et al., 2015).

Therapeutic developments for rare diseases have advanced in recent years, with better diagnostic tools such as whole genome and exome sequencing as well as medical advancements in understanding disease mechanisms and improved technology. Furthermore, exponential advancements in artificial intelligence (AI) systems are propelling the rate of drug discovery and design, as well as advancing basic molecular knowledge. Pharmaceutical companies, such as Atomwise, Insilico Medicine, and Exscientia increasingly leverage AI for optimizing drug discoveries (Arnold, 2023). Deepmind's AlphaFold2 (AF2) has substantially improved *in silico* predictions of protein-folding and protein interactions with other proteins, DNA, or

other molecules (Jumper et al., 2021).

1.6.2 Current studies of the MDEMs

MDEMs are considered good candidates for therapeutic intervention for several reasons. One key factor is their relative genetic simplicity, meaning that the syndromes are due to disease causing variants in a single known gene, unlike many complex disorders that are caused by interactions of multiple genes. Furthermore, the majority of MDEMs are due to defects in enzymes, which have been shown to be good targets for therapeutic interventions (Fahrner & Bjornsson, 2019). For these diseases to progress to clinical studies, there must be considerable evidence, both *in vitro* and *in vivo*, demonstrating the disease's responsiveness to treatment. For a few MDEMs, there has been considerable research done, with new research constantly emerging that demonstrates postnatal malleability of the brain and improvements upon ID/DD treatment (Alarcón et al., 2004; Bjornsson et al., 2014; Guy et al., 2007). In an effort to restore the epigenetic imbalance of MDEMs *in vivo*, inhibitors of epigenetic regulators have been tested in mouse models. HDAC inhibition has previously been demonstrated to improve memory in AD mouse models (Kilgore et al., 2010). Once applied to MDEMs, they have demonstrated great promise for improving neurological defects, both behavioral and molecular, in mouse models of KS and RST by restoring the chromatin balance (Alarcón et al., 2004; Bjornsson et al., 2014). Histone demethylase (HDM) inhibition has shown advancements in treating neurological phenotypes in *Kmt2c*^{+/-} mouse model of Kleefstra syndrome as well as in *Kmt2d*^{+/-} mouse model of KS (Bjornsson et al., 2014; T. Nakamura et al., 2024). When it comes to chemical inhibition of enzymes, the specificity of the drugs needs to be considered, since many of them interact with multiple HDACs (Bannister & Kouzarides, 2011).

Moving from pre-clinical trials, KS has shown promising progress in drug discoveries. In a study from 2023, 31 small molecule compounds were identified to increase KMT2D levels in cell lines, with thirteen showing great promise. Further validation will be needed in cellular and animal models to estimate their readiness for clinical trials (Ma et al., 2023). Furthermore, there is an ongoing clinical early phase 1 trial for KS, testing the efficacy of a modified Atkins diet in patients. This is based on results from a mouse model of KS showing improved metrics of neurological defects after ketogenic diet treatment, which naturally increases beta-hydroxybutyrate, an endogenous HDAC inhibitor (HDACi) (Benjamin et al., 2017). Although some studies have shown great promise, other therapeutic agents have not met expectations in clinical trials and have therefore not been continued. An example of this is the discontinuation of Fingolimod in treatment for Rett syndrome, where treatment did not have the desired success in mice (Naegelin et al., 2021).

1.6.3 *In utero* therapeutics

In utero therapy options are increasingly more feasible as treatment of genetic disorders

during pregnancy and before birth. Animal studies in mice and primates have shown that there are considerable advantages when it comes to *in utero* gene therapies (IUGTs) (Bose et al., 2021). For one, the small dosage needed due to the fetus's smaller size is highly important for cost efficiency and manufacturing. Secondly, a fetus has a high proportion of progenitor cells which makes it more likely the benefits of the therapy will extend to newly formed cells. Studies have shown that when it comes to neurological diseases, it is highly beneficial that the blood-brain-barrier (BBB) of a fetus is permeable, making treatment more efficient (Massaro et al., 2018; Mattar et al., 2013). Lastly, studies indicate that the fetal immune system tolerates well transgene delivery with gene therapy (Davey et al., 2017). There are also limitations to IUGTs that need further research and assessment. The risks of the procedure must be carefully weighed against the benefits. This makes perinatally and early postnatally fatal or very severe diseases the top candidates for further research. *In utero* enzyme replacement therapies have moved on to phase 1 clinical trial for lysosomal storage diseases, based on promising pre-clinical studies showing great benefits for disease progression and patient survival (Herzeg et al., 2023). However, IUGTs have not yet reached clinical trials, despite the substantial progress that has been made in the past few years (Bose et al., 2021).

1.6.4 Challenges and considerations

Generally, there are a lot of challenges researchers must overcome in developing novel therapeutics. The enormous cost of drug development and the research needed to meet the qualification for a clinical trial often restricts scientists from moving forward with development. The regulatory process of approving new drugs for clinical trials at all phases can be lengthy and complicated with many agencies involved. There are also ethical principles and regulations that must be upheld throughout the process, such as animal welfare ethics and informed consent of patients. Moreover, after a drug has been approved for manufacturing, patents can grant a single company the right to sell the drug at high prices, restricting access to the drug for lower income patients (Yildirim et al., 2016). Finally, rare diseases often take longer than other diseases to move to clinical trials. This might be due to a lack of funding for rare disease research, as *in vitro* and *in vivo* experiments are generally very expensive. This can slow down the generation and validation of pre-clinical results needed to move on to clinical studies. Lastly, recruitment for clinical studies can prove difficult with small patient groups (Coors et al., 2017).

1.7 Cre-loxP system

The Cre-loxP recombination system is a genetic mechanism widely used to induce recombination of the genome. This recombination system was described in 1981 in the bacteriophage P1 (Sternberg & Hamilton, 1981). The 38 kDa Cre recombinase enzyme was discovered to recognize a highly specific 34 bp DNA sequence called a loxP site

(locus of x-over P1), and excise the DNA positioned between them. In mouse studies, the Cre-loxP recombination system has been used to generate various Cre-driver mouse lines, which can either be constitutive or inducible, which are then bred to loxP-containing mice. Constitutive Cre mouse lines express the Cre recombinase under specific promoters or enhancers, and are therefore expressed when those regulatory elements are active. This enables scientists to restrict the Cre recombination to specific tissues or developmental timepoints. Inducible Cre models have been designed to provide scientists with even more control over the timing of Cre recombinase expression. For example, in a tamoxifen-inducible Cre model, Cre recombinase is expressed but restricted to the cytoplasm. Once tamoxifen is administered to the mice, the Cre recombinase is able to enter the nucleus and recombine the DNA at the loxP sites. These systems have advanced genome editing technologies and provided with a highly controlled method of manipulating the genome in a cellular, developmental and tissue specific manner (H. Kim et al., 2018).

2 Aims

The general aim of this PhD study was to investigate the disease mechanism of WDSTS from multiple angles using various approaches, including *in silico*, *in vitro*, and *in vivo* methods. Additionally, we aimed to explore the *in utero* malleability of the neurological dysfunction of WDSTS, with the long-term goal of paving the way for WDSTS to be considered for clinical studies.

2.1 Aim 1: Exploring the effects of WDSTS missense variants *in silico*

In the first aim, we set out to explore whether we could identify specific protein domains of KMT2A that are more likely to contribute to WDSTS pathogenesis using computational approaches. Since disease-causing missense variants tend to cluster in critically important domains, we hypothesized that we would observe an enrichment of WDSTS variants within certain domains of KMT2A. We utilized publicly available data on missense variants found in healthy populations (collected from gnomAD) and in individuals with WDSTS (collected from ClinVar) and performed a comparative analysis. We then aimed to assess where gene expression disruption primarily occurred in the genome due to cKO of *Kmt2a* in a relevant mouse neuronal cell type, using publicly available RNA-seq and ChIP-seq data sets. Focusing on the CXXC domain of KMT2A, we aimed to investigate whether experimental data in combination with newly available protein-folding program AlphaFold2 could accurately predict the effects of missense variants within the domain.

2.2 Aim 2: Creating and characterizing a mouse model of WDSTS (*Kmt2a*^{+/LSL})

The second aim was to create and explore a novel mouse model of WDSTS, *Kmt2a*^{+/LSL}, created in collaboration with the Jackson Laboratory. The mouse model has an insertion of a heterozygous loxP-stop-loxP (LSL) cassette in *Kmt2a* gene, leading to a loss of function KMT2A expression of the affected allele. We wanted to expand the mouse colony and validate the presence of the LSL cassette using PCR, RT-qPCR and genome wide sequencing. To be able to utilize this model as a WDSTS disease model for further experiments, we needed to establish WDSTS-like phenotypes of the mice. We aimed to investigate these phenotypes using a broad range of strategies, such as weight monitoring, histology staining and CT scanning. Furthermore, we proposed to describe the neurological dysfunction of the model, using methods like brain sectioning and behavioral testing.

2.3 Aim 3: Testing malleability of the neurological dysfunction in WDSTS *in vitro* and *in vivo*

The final aim was centered around investigating the malleability of the neurological dysfunction of the mouse model, given that intellectual disability is a prominent characteristic of WDSTS. We tested two strategies to remove the LSL cassette using Cre-recombinase mouse models, an inducible ROSA-CreER^{T2} model and a constitutive nestin-Cre model. To explore the reversal of the WDSTS-like phenotypes *in vitro*, we intended to analyze a relevant cell type, mouse NPCs, cultured from the mouse models. We then utilized methods such as RNA sequencing and CUT&RUN to evaluate the gene expression dysregulation and epigenetic imbalance caused by the heterozygous loss of *Kmt2a*. We then wanted to explore the cellular rescue after the regain of *Kmt2a*. Lastly, we set out to explore the rescue of the neurological phenotypes *in vivo* using the methods listed in aim 2.

3 Materials and Methods

In this chapter, I will outline the methods used in my thesis and briefly discuss the rationale for selecting these specific methods. More detailed description of materials and methods used can be found in each paper (Paper I, Paper II).

3.1 Spatial missense variant (MV) analysis and conservation

A description of how the missense variants were collected and filtered from gnomAD, ClinVar and COSMIC, as well as the variant enrichment and conservation analysis can be found in the Methods in Paper I.

3.2 MV classification and saturation mutagenesis

Detailed description of how the variant classification scheme was developed, validated and deployed for the *in silico* saturation mutagenesis can be found in Methods in Paper I.

3.3 Mouse models

3.3.1 Mouse husbandry and licenses

All mouse lines were housed at ArcticLAS, where all mouse experiments were performed (Reykjavík, Iceland). The mice are kept in individually ventilated cages (Green line Sealsafe cages, Tecniplast), five together at most, on Aspen bedding (Tapvei 4HV, Brogarden, Denmark) and provided with filtered tapwater (Danmil filter, #DA25CSSS02). Furthermore, the mice were kept at 21-23°C with a 12-hour day-night cycle and relative air humidity of 40-50%. The mice were given rodent diet *ad libitum*, either with a breeding or maintenance diet (Altromin NIH#31M or Altromin 1324, respectively, Brogarden, Denmark). All mouse strains were kept on a pure C57BL/6J background. Pups were weaned at 3-4 weeks of age, and marked with ear-clips used for genotyping. A regular FELASA health monitoring check for resident and sentinel animals was performed with samples sent to and analyzed at IDEXX (Germany) during the time of this PhD, with all results negative. The import of mice from The Jackson Laboratory (USA) and all experimental protocols complied with the Icelandic Food and Veterinary Authority, license no. 2208602, and were approved by the National Expert Advisory Board on Animal Welfare.

3.3.2 The WDSTS mouse model

The novel WDSTS mouse model was designed and created in collaboration with The Jackson Laboratory (JAX). For a more detailed technical description of the creation of the strain, see Methods in Paper II. The model has a heterozygous stop cassette inserted in intron 1 of *Kmt2a*, flanked by loxP sites on either side (Figure 1). The loxP-stop-loxP (LSL) cassette gave rise to the model's nomenclature, *Kmt2a*^{+/LSL}. The stop cassette contains a terminating sequence, i.e., a polyadenylation signal from the simian virus 40 (SV40pA). The terminating sequence results in a truncation of the mRNA transcript and subsequent polyadenylation, leading to an obstructed and truncated transcript of the affected allele that then gets degraded (Hans & Alwine, 2000). Therefore, the cassette acts as a LoF variant that results in heterozygous *Kmt2a* expression in the mice, similar to what is found in individuals with WDSTS. Since the cassette is surrounded by loxP sites, we are able to utilize the Cre-lox recombination system to effectively remove the LSL cassette and restore the *Kmt2a* expression. To accomplish this, we need to breed the mice to a Cre-containing mouse model, and in this thesis, we attempted both inducible and constitutive recombination. This model was kept and maintained in heterozygosity. With this mouse model design, we are able to create a functioning WDSTS disease model as well as test malleability of the WDSTS phenotypes.

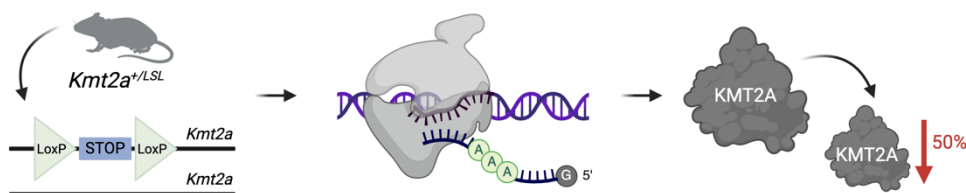


Figure 1. Schematic figure of the *Kmt2a*^{+/LSL} mouse model.

A schematic figure shows the heterozygous LSL cassette insertion into intron 1 of *Kmt2a* (left) in the *Kmt2a*^{+/LSL} mouse model. The LSL cassette leads to a premature polyadenylation and truncation of the mRNA transcript (middle). This results in decreased KMT2A expression in the *Kmt2a*^{+/LSL} mouse model (right).

3.3.3 Nestin-Cre mouse model

The nestin-Cre model was purchased from JAX (B6.Cg-Tg(Nes-cre)1Kln/J, strain #003771). This model has a transgene of the Cre-recombinase expressed under the control of the rat nestin promoter and enhancer, inserted on chromosome 12. This results in a constitutive Cre expression in the CNS and PNS, specifically in the neuronal and glial precursor cells (Tronche et al., 1999). The Cre expression is stated to start as early as E11 by the creators of the model, while other papers have shown that Cre expression does not reach sufficiently high levels to result in effective recombination

until late embryonic development, or even early postnatal days (Liang et al., 2012; Tronche et al., 1999). The nestin-Cre strain was kept in homozygosity ($Nes-Cre^{-/-}$) for maintenance and breeding to $Kmt2a^{+/LSL}$ model but all experiments were conducted on heterozygous mice ($Nes-Cre^{+/-}$) (Figure 2). This model was chosen to restore the $Kmt2a$ expression efficiently *in utero* in the CNS and PNS, allowing us to elucidate which WDSTS phenotypes are neurologically driven as well assess the malleability of the neurological dysfunction of WDSTS.

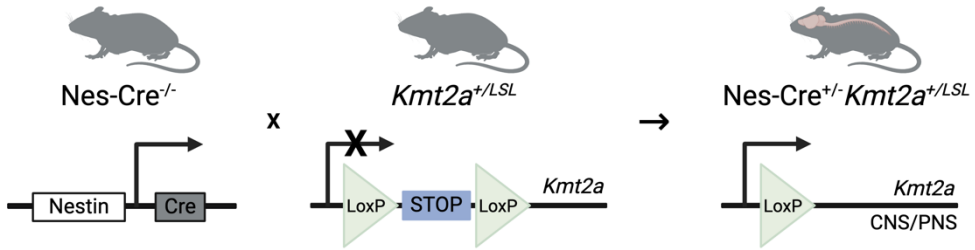


Figure 2. Nestin-Cre and $Kmt2a^{+/LSL}$ breeding schematic.

This figure shows how the breeding cross between the nestin-Cre and $Kmt2a^{+/LSL}$ mouse models yields double-mutant mice $Nes-Cre^{+/-} Kmt2a^{+/LSL}$. The Cre-recombinase is present in nestin-expressing cells in the PNS and CNS, resulting in regained $Kmt2a$ expression of those cells.

3.3.4 ROSA-CreER^{T2} mouse model

The ROSA-CreER^{T2} mouse model was also purchased from JAX (B6.129-Gt(ROSA)26Sor^{tm1(cre/ERT2)Tyj}/J, strain #008463). The model has a Cre-ER^{T2} (cre-recombinase estrogen receptor T2) cassette inserted into intron 1 of the *Gt(ROSA)26Sor* gene, located on chromosome 6, that results in a fusion protein of the Cre-recombinase and modified estrogen receptor (ER^{T2}) under the control of the ubiquitously expressed ROSA26 promoter. The Cre recombinase is restricted to the cytoplasm of the cell by the estrogen receptor, and once exposed to tamoxifen (tam) that gets converted to its active metabolite 4-hydroxy-tamoxifen (4-OH-tam), the Cre-recombinase can move into the nucleus and cause recombination at the loxP sites in the genome (Ventura et al., 2007). The ROSA-CreER^{T2} strain was also kept in homozygosity (ROSA-CreER^{T2}^{-/-}) for maintenance and breeding to $Kmt2a^{+/LSL}$ model with all experiments conducted on heterozygous mice (ROSA-CreER^{T2}^{+/-}) (Figure 3). This model is an inducible Cre mouse line that allows for temporal control of the recombination and was therefore chosen to test the postnatal malleability of WDSTS phenotypes.

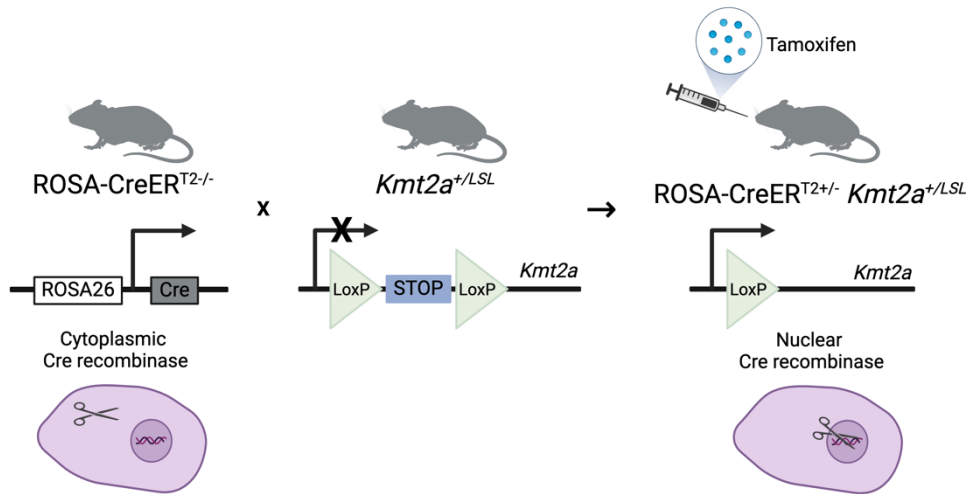


Figure 3. Tamoxifen induced Cre-recombination of the *Kmt2a*^{+/-LSL} mice.

This figure shows the induced recombination of the LSL cassette using the ROSA-CreER^{T2} model. In the ROSA-CreER^{T2}^{-/-} mice the Cre recombinase (depicted as scissors in the figure) is ubiquitously expressed but retained in the cytoplasm of the cells (left). Once the mice are crossed with the *Kmt2a*^{+/-LSL} model (middle) to produce double mutant mice (ROSA-CreER^{T2}^{+/-} *Kmt2a*^{+/-LSL}), administering tamoxifen to the cells will release the Cre-recombinase into the nucleus to cause recombination of the LSL cassette (right).

3.3.5 *Dct-LacZ* mouse model

When characterizing the *Kmt2a*^{+/-LSL} model, we wanted to explore potential neural crest defects by looking into the melanoblast migration using a *Dct-LacZ* mouse model. The model was gifted to us from Dr Eiríkur Steingrímsson, described by Mackenzie et al. (Mackenzie et al., 1997). The model has a bacterial beta-galactosidase gene (*LacZ* gene) under a Dopachrome tautomerase (*Dct*) promoter. This results in beta-galactosidase expression in *Dct*-expressing cells during embryonic development, including melanoblasts, retinal pigment epithelium and neuronal precursor cells of the telencephalon (Jiao et al., 2006), which can be detected and visualized with a beta-galactosidase staining, see section 3.4.3.2. This model was kept in heterozygosity (*Dct-LacZ*^{+/-}) and crossed with the *Kmt2a*^{+/-LSL} model.

3.4 Mouse experiments

3.4.1 Genotyping

All mice were genotyped from DNA samples collected from the ear-clips done at weaning-age. Briefly, samples were collected in an Eppendorf tube and 200 µl of tail lysis buffer (Tris HCl pH 8.5 (100 mM), EDTA (5 mM), SDS (0.2%) and NaCl (200 mM)) was added with 0.5% 20 mg/ml proteinase K (Thermo Scientific, E00491), and

incubated at 55°C for 1+ hours at 1200-1400 rounds per minute (rpm). Samples were centrifuged for 10 minutes (all centrifuge steps were carried out at 13000 rpm) , and supernatant moved to a new Eppendorf tube, along with 200 µl isopropanol (Honeywell, 34863) and 1 µl 20 mg/ml glycogen (Thermo Scientific, D4065) and vortexed thoroughly. DNA was pelleted by centrifugation for 10 minutes and pellet was washed 1-2 times with 1ml 70-75% ethanol (EtOH). Supernatant was removed and DNA was eluted with 150-200 µl TE buffer (provided by deCODE). DNA concentration was measured on NanoDrop spectrophotometers (Thermo Scientific), and samples were stored at -20°C. Genotyping of all mice was performed using polymerase chain reaction (PCR) of the DNA, either with an ExtraMix reaction or a Q5 reaction. Reaction reagents and thermocycler conditions can be found in Table 2. Negative controls were run with every batch, with nuclease free H₂O instead of DNA sample. PCR reactions were run on a MiniAmp Thermal Cycler (Thermo Scientific, A37834).

Table 2. Genotyping reactions for ExtraMix and Q5 reactions.

Genotypes detected	Reaction reagents	Thermocycler conditions
ExtraMix reaction:		
Nestin-Cre <i>Dct-LacZ</i> ROSA-CreER ^{T2}	- 4 µl ExtraMix (provided by deCODE) - 3 µl DNA sample - 2.4 µl nuclease free H ₂ O - 0.3 µl primers (10µM)	- 96°C for 2 minutes 35 cycles of: - 96°C for 30 seconds - Annealing temp* for 30 seconds - 70°C for 1 minute Then: - 72°C for 10 minutes - 4°C hold Annealing temperature for Nestin-Cre reaction was 50°C, for <i>Dct-LacZ</i> reaction was 62°C and for ROSA-CreER ^{T2} was 57°C.
Q5 reaction:		
<i>Kmt2a</i>	- 5 µl NEBNext Ultra II Q5 Master Mix (BioLabs, M0544S) - 3 µl DNA sample - 1 µl nuclease free H ₂ O - 0.5 µl primers (10µM)	- 98°C for 30 seconds 35 cycles of: - 98°C for 10 seconds - 72°C for 2 minutes Then: - 72°C for 2 minutes - 4°C hold

All genotyping primers and the corresponding DNA PCR product sizes can be found in Table 3. All primers were dissolved in nuclease free H₂O to 100 µM stock solutions and stored at -20°C. Working solutions were diluted stock 1:10 in nuclease free H₂O (final concentration of 10 µM). A Gel Loading Dye 6x Purple (BioLabs, B70255) was added to all PCR samples, and they were run on 1.5% agarose gel (1.5% Agarose Basic (AppliChem, A8963), in TAE buffer with 0.005% ethidium bromide (Sigma, E1510)), with a 100 bp or 1kb DNA Ladder (BioLabs, N3231L and N0550S, respectively), at 90V for 20-30 minutes using PowerPac 3000 (BioRad). Gels were imaged using an Universal Hood II (BioRad) and analyzed with the ImageLab software (BioRad).

Table 3. Genotyping PCR primers, their sequence and PCR product size.

Genotype reaction	Primer name	Primer sequence (5' → 3')	PCR product size
Nestin-Cre	nestinCre_wtF	TTGCTAAAGCGCTACATAGGA	Nes-Cre ^{+/+} : 150 bp Nes-Cre ^{+/-} : 150 and 246 bp Nes-Cre ^{-/-} : 246 bp
	nestinCre_R	GCCTTATTGTGGAAGGACTG	
	nestinCre_tgF	CCTTCCTGAAGCAGTAGAGCA	
<i>Dct-LacZ</i>	DctLacZ-F	CAGGACACGGCTTGTCATCATGGTGT	<i>Dct-LacZ</i> ^{+/+} : 300 bp <i>Dct-LacZ</i> ^{+/-} : 300 and 450 bp
	DctLacZ-R	CATTCATCGTCTCTCAGGAATTCA	
	DctWT-F	CAGGACACGGCTTGTCATCATGGTGT	
	DctWT-R	CCACACAGACACCTACCACATGCGT	
ROSA-CreER ^{T2}	Cre_WT_fwd	CTGGCTTCTGAGGACCG	ROSA-CreER ^{T2+/+} : 198 bp ROSA-CreER ^{T2+/-} : 198 and 150 bp ROSA-CreER ^{T2-/-} : 150 bp
	Cre_WT_rev	CGTGATCTGCAACTCCAGTC	
	Cre_MT_fwd	CCGAAAATCTGTGGAAGTC	
	Cre_MT_rev	AGGCAAATTTGGTGTACGG	
<i>Kmt2a</i>	kmt2a_genof1	AACACATGGCTTCCTGGAGG	<i>Kmt2a</i> ^{+/+} : 477 bp <i>Kmt2a</i> ^{+/-LSL} : 1562 bp
	kmt2a_genor1	TGCAGGGTTAACACATGGCT	

3.4.2 Inducible Cre-recombination of *Kmt2a*^{+/-LSL} mice

For inducing the recombination of the LSL cassette in the *Kmt2a*^{+/-LSL} mouse model postnatally, we crossed the *Kmt2a*^{+/-LSL} mouse strain with the ROSA-CreER^{T2} strain. This breeding setup produced offspring of either a wild-type (WT) (ROSA-CreER^{T2+/-}

Kmt2a^{+/+}) or a double mutant (DM) (*ROSA-CreER*^{T2+/-}*Kmt2a*^{+/^{LSL}}) genotype, which were used for all optimizations. Three methods were tested: in Optimization 1 the tamoxifen was given with intraperitoneal injections (IP injections), in Optimization 2 the tamoxifen was given orally to ensure rapid conversion of the drug to the active metabolite 4-OH-tam, and in Optimization 3 the 4-OH-tam was administered directly via IP injections.

3.4.2.1 Optimization 1

For this optimization, two-week-old pups from *ROSA-CreER*^{T2+/-} x *Kmt2a*^{+/^{LSL}} breedings were used. The mice were treated with tamoxifen and controls were treated with corn oil vehicle. The positive controls included non-injected and tamoxifen injected *ROSA-CreER*^{T2+/-}*Kmt2a*^{+/+} mice, while the negative controls were corn oil treated *ROSA-CreER*^{T2+/-}*Kmt2a*^{+/^{LSL}} mice. The tamoxifen powder (Sigma, T5648) was weighed into an Eppendorf tube and 5x 100% EtOH was added for a stock solution with a final concentration of 200 mg/ml. Tubes were placed on a shaker at room temperature (RT) for 15 minutes protected from light. Working solution was made fresh by diluting the stock solution 1:10 in corn oil (Santa-Cruz, sc-214761) for a final concentration of 20 mg/ml and placed on a shaker at 37°C for 15 minutes. The working solution was made fresh every day. The mice were given the tamoxifen or corn oil (CO) with IP injections based on their weight at a 200 mg/kg dose. For the number of mice per group, see Table 4. This administration method was tested, as it is the recommended method by JAX (Madisen et al., 2010).

3.4.2.2 Optimization 2

In this optimization 5 breeding cages were set up with *ROSA-CreER*^{T2+/-} (female) and *Kmt2a*^{+/^{LSL}} (male) mice, resulting in 5 litters of a total of 35 pups. After the females got pregnant, the male was removed. Tamoxifen or the corn oil vehicle was administered with oral gavage to the lactating female mice two times a day for 3 days, for a total of six doses, starting on postnatal day 1 (P1), following guidelines described by Weber and Álvarez-Aznar et al. (Álvarez-Aznar et al., 2020; T. Weber, 2009). Three cages received tamoxifen treatment at a lower concentration than in Optimization 1 (stock solution diluted 1:20 for a final concentration of 10 mg/ml) while two cages received the corn oil vehicles. The *ROSA-CreER*^{T2+/-}*Kmt2a*^{+/+} mice served as positive controls and the corn oil treated *ROSA-CreER*^{T2+/-}*Kmt2a*^{+/^{LSL}} mice served as negative controls. The mice were given the tamoxifen or corn oil with an oral gavage based on their weight, at 100 mg/kg. This administration route was tested to bypass injecting the pups directly which can cause stress and inflict internal wounds if not done properly. For the number of mice per group, see Table 4. Subsequently, the mice were weighed weekly from age 2-5 weeks and then sacrificed and dissected. Cerebellar samples were collected in 700µl TRI-reagent solution (Thermo Scientific, AM9738), see section 3.4.2.4 for sample preparation.

3.4.2.3 Optimization 3

In this optimization 2-3-week-old pups from ROSA-CreER^{T2+/-} x *Kmt2a*^{+/*LSL*} breedings were given the active metabolite 4-OH-tam for two consecutive days. The corn oil treated ROSA-CreER^{T2+/-}/*Kmt2a*^{+/+} mice served as positive controls while the oil treated ROSA-CreER^{T2+/-}/*Kmt2a*^{+/*LSL*} mice served as negative controls. The 25 g of 4-OH-tam (Sigma, H7904) was put in a fresh tube and 1 ml of corn oil was added. The tube was placed on a shaker at 1300 rpm at 55°C for 30 minutes and vortexed regularly, protected from light. The solution was split into two vials with 0.5 ml each and then 0.75 ml of corn oil was added, resulting in a final concentration of 10 mg/ml. The solution was sonicated for 30 seconds on, 30 seconds off, for 10 minutes until the powder was fully dissolved. The mice were given the 4-OH-tam or corn oil via IP injections based on their weight, at 100 mg/kg. For the number of mice per group, see Table 4. Mice were sacrificed and dissected 9 weeks later, with hippocampal samples collected in 300 µl RNA later stabilization solution (Thermo Scientific, AM7020) see section 3.4.2.4 for sample preparation.

Table 4. Genotype, treatment and number of mice in each optimization for the inducible Cre-recombination.

Genotype	Treatment	Optimization 1	Optimization 2	Optimization 3
ROSA-CreER ^{T2+/-} / <i>Kmt2a</i> ^{+/+}	Tamoxifen	10 mice	2 mice	0 mice
	Corn-oil	1 mouse	2 mice	5 mice
	No injection	3 mice	0 mice	0 mice
ROSA-CreER ^{T2+/-} / <i>Kmt2a</i> ^{+/<i>LSL</i>}	Tamoxifen	3 mice	10 mice	4 mice
	Corn-oil	2 mice	3 mice	1 mouse

3.4.2.4 RNA isolation, cDNA synthesis and RT-qPCR

Brain tissue samples from Optimizations 1 and 3 were homogenized in TRI-reagent solution and then the RNA was isolated using the Direct-zol RNA Micro-prep kit (Zymo, R2060) using the manufacturer's protocol and RNA concentration was measured on the NanoDrop. The RNA samples (120-150 ng) were synthesized to complementary cDNA (complementary DNA) using the High-capacity cDNA reverse transcription kit (Thermo Scientific, 4374966) on the MiniAmp Thermal Cycler using the manufacturer's suggested thermal cycler conditions. The cDNA was used as input (5-6.25 ng/reaction) for real-time quantitative PCR (RT-qPCR) for measuring the abundance of a gene's mRNA transcripts. The RT-qPCR was performed using Luna Universal qPCR Master Mix (BioLabs, M3003S), with the reaction and thermal cycler conditions according to the

manufacturer's protocol run on the CFX384 Touch Real-Time PCR System (Bio-Rad). The primers of the target and reference genes used can be found in Table 5.

Table 5. RT-qPCR primer names and sequences.

Gene	Primer name	Primer sequence (5' → 3')
<i>Gapdh</i>	Gapdh_Fwd	TGGCCTTCCGTGTTCTAC
Housekeeping gene	Gapdh_Rev	GAGTTGCTGTTGAAGTCGCA
<i>Kmt2a</i>	Kmt2a_Exon30-32_fwd	GGCTTTCAGATCTGTGCGGAA
Target gene	Kmt2a_Exon30-32_rev	AACCGTTCACACCTGCAAAT

RT-qPCR results were analyzed using CFX Manager (Bio-Rad), where the average cycle threshold (C_T) was calculated for each sample using three technical replicates, excluding technical replicates where the standard deviation (SD) was above 0.1. The delta C_T (ΔC_T) values were calculated for each sample by subtracting the average C_T value from the housekeeping gene and the target gene. The $\Delta\Delta C_T$ values were then calculated by finding the difference between the ΔC_T values of each sample and the average of the C_T values from the target gene of the control genotype (ROSA-CreER^{T2}/*-Kmt2a*^{+/+}). Finally, the relative target gene expression was calculated by using Equation 1. The results were plotted and an unpaired t-test applied to test statistical significance using Prism 9 (GraphPad Software).

$$\text{Relative gene expression (\%)} = 2^{-\Delta\Delta C_T} * 100$$

Equation 1. RT-qPCR relative gene expression calculation.

3.4.3 Mouse model phenotyping

The mouse models were characterized using a variety of phenotyping assays, with an emphasis on comparing the *Kmt2a*^{+/*LSL*} model to their *Kmt2a*^{+/+} littermates. Some of the established phenotypes were then used to assess the phenotypic rescue of ROSA-CreER^{T2} and nestin-Cre models. In short, the mice were weighed regularly to monitor their growth, and their skulls were CT scanned to assess their craniofacial defects. The hypotonia was tested using a hindlimb suspension test and a surface righting test, while the visuospatial memory of adult mice was assessed using a Y-maze. Blood samples from the mice were taken and the Complete Blood Count (CBC) was analyzed. Furthermore, the mice were dissected and brain RNA samples taken to evaluate the relative *Kmt2a* expression (see section 3.4.2.4 above), distal colon samples were taken for calretinin staining and skin samples were taken for H&E staining. A subset of mice was then perfused through their cardiovascular system and the brains dissected,

cryopreserved, and sectioned using a cryostat. The brain samples were then used for immunofluorescence (IF) staining. For a more in-depth description of the mouse model characterization and detailed protocols, see Methods in Paper II.

3.4.3.1 *Fur phenotyping*

In an effort to characterize the fur phenotype of the *Kmt2a^{+/-LSL}* model, we utilized a few different approaches. To evaluate the rate of fur regrowth, we removed hair from the backs of adult mice and imaged the regrowth every few days (see Methods in Paper II). We then wanted to assess the melanin density of the skin on the abdominal area and back of *Kmt2a^{+/-LSL}* pups compared to *Kmt2a^{+/+}* littermates, so the skin sections were immunohistochemistry stained (IHC) for melan-A, a marker for melanocytes (Pitcovski et al., 2017). The staining was performed using the same protocol as the H&E staining described in Methods in Paper II. Briefly, the skin sections were deparaffinized and then rehydrated in xylene and ethanol. Antigen retrieval was performed using Envision Flex Target Retrieval Solution High 9 (Agilent, K800421-2) and then the immunostaining was done using the AutostainerLink 48 (Agilent) system with the EnVision Detection System Peroxidase/DAB, Rabbit/Mouse kit (Agilent, K4065). The primary antibody melan-A (Agilent, M7196) was applied and incubated at room temperature for 30 minutes. Next, the sections were incubated with FLEX Linker for 10 minutes, followed by EnVision FLEX/HRP incubation. Between steps, the sections were washed with Tris-buffered NaCl solution with Tween 20 (pH 7.6) (Agilent, S3306). Then, sections were incubated with 3,3'-diaminobenzidine solution (Agilent, K4065) for 10 minutes and counterstained with hematoxylin for 30 seconds. Finally, sections were dehydrated in 100% ethanol and xylene, and coverslipped using mounting medium (Pertex, Histolab). Images were captured using a NanoZoomer XR scanner (HAMAMATSU). The sections were analyzed with NDP view (HAMAMATSU), where average gray value was calculated from 10 images per biological replicate using Fiji (Schindelin et al., 2012). Results were plotted and analyzed with an unpaired t-test using Prism 9.

To assess the coat color of the adult mice, we measured the fur color using protocol #2 from Berlin et al. (Berlin et al., 2012). The lower back of 7-week-old mice was shaved and their fur collected. In technical triplicates for each mouse, 20 mg of hair was weighed into a fresh 1.5 ml Eppendorf tube and 500 μ l of 1M sodium hydroxide (NaOH) was added. The tubes were placed on a heat block for 2 hours at 83°C and 500 rpm and vortexed every 30 minutes, until fur was lysed. Samples were centrifuged at 14000 rpm for 5 minutes and each technical replicate was split into 3 wells of a 96 well plate and the optical density (OD) absorbance of the lysate was measured at 470 nm with 1M NaOH as blank. The average OD value for each technical replicate was calculated as the mean value from the 3 wells, and then the blank value was subtracted from the average value of each technical replicate. Finally, the three

technical replicates were averaged to get final OD value for each mouse, which were plotted, and an unpaired t-test was applied for statistical testing using Prism 9.

3.4.3.2 X-gal staining

We utilized the X-gal staining to evaluate whether the *Kmt2a*^{+/LSL} mice showed defects in their melanoblast quantity or migration from the neural crest. To do this, we performed a timed breeding with a male *Kmt2a*^{+/LSL} mouse and a female *Dct-LacZ*^{+/-} mouse. The female was sacrificed at E13.5 and the embryos were dissected out, fixed, permeabilized and then X-gal stained according to protocol #13 from Berlin et al. (Berlin et al., 2012). Briefly, the staining of the embryos causes the X-gal chemical to stain blue in the presence of bacterial enzyme beta-galactosidase, which is encoded by the *LacZ* gene, which is only expressed under the *Dct* promoter. We are therefore able to visually detect the quantity and positioning of the *Dct* expressing cells, including melanoblasts of the *Dct-LacZ*^{+/-}*Kmt2a*^{+/+} and *Dct-LacZ*^{+/-}*Kmt2a*^{+/LSL} embryos. The embryos were imaged using Leica EZ4 W Microscope with the 8x and 12.5x magnification lens for visualization. To evaluate the melanoblast migration from the dorsal part of the embryo to the ventral, the torso was divided into top-, middle-, and bottom area between the base of the forelimb and hindlimb. The melanoblasts within each area on both the right and left sides of the embryo were counted 5 times within a fixed size of 280 x 280x pixels using Fiji (Schindelin et al., 2012). The average melanoblast count for each section was calculated from the ten areas and plotted as a percentage of the average total melanoblast count for each embryo. For counting the melanoblasts in each area, we used the Image-based Tool for Counting Nuclei (ITCN) plugin for Fiji. Embryos with torn skin on the sides were excluded from counting. To test significance, an unpaired t-test was applied using Prism 9.

3.5 Cellular experiments

3.5.1 mNPCs extraction and culture

Mouse NPCs (mNPCs) were dissected from the hippocampus of the mouse models, cultured and utilized for *in vitro* experiments. For a detailed description of protocols used, see Methods in Paper II.

3.5.2 Incucyte experiments

3.5.2.1 Proliferation assay

To assess the proliferation rate of the mNPCs from the mouse models, mNPCs were seeded at a density of 50.000 cells/well in 4 technical replicates per biological replicate in a TPP 24-well plate (Sigma, Z707791). The plate was then placed in the Incucyte and 9 images were taken per well every two hours for 96 hours. The average image percentage confluency (%) was calculated for each biological replicate and

plotted for each genotype. To test the statistical significance, a non-linear regression (logistic growth curve) was applied to each genotype and then compared, using Prism 9.

3.5.2.2 Scratch wound healing assay

To evaluate the migratory rate and capabilities of our mNPCs, we used the Incucyte® 96-Well Scratch Wound Cell Migration assay (Sartorius, Germany). We seeded the mNPCs at a 45,000 cells/well density 96-well Incucyte Imagerock plate (Sartorius, BA-04857). We seeded for 10 technical replicates per biological replicate and let the plate reach 90% confluency, which took ~48 hours, before making the wound, using the Incucyte 96-well Woundmaker Tool (Sartorius, 4563). We performed the assay as described in the manufacturer's protocol. After making the wound, we placed the plate in the Incucyte and imaged one image per well every two hours for 36 hours. The analysis was performed using the Incucyte Scratch Wound Analysis Software Module, only using the technical replicates per sample that had a successful wound. The average confluency percentage (%) of the wounds for each genotype was plotted and a non-linear regression (exponential plateau curve) was fitted to each genotype and then compared, using Prism 9.

3.6 Sequencing experiments

3.6.1 RNA sequencing

In Paper I, we utilized a publicly available RNA sequencing (RNA-seq) dataset [GSE99250] from *Kmt2a* cKO cells of mouse hippocampal CA neurons (Kerimoglu et al., 2017). See Methods in Paper I for the description of the RNA-seq analysis. In Paper II we performed our own RNA-seq experiment from mNPCs from the mouse models, see Methods in Paper II for sample preparation, sequencing and downstream analysis.

3.6.2 Epigenetic profiling methods

We used publicly available Chromatin Immunoprecipitation followed by sequencing (ChIP-Seq) data [GSE99250] from mouse hippocampal neurons in *Kmt2a* cKO mice (Kerimoglu et al., 2017). We analyzed both the H3K4me1 and H3K4me3 data, see Methods in Paper I for a more detailed description of the data analysis. We then performed a Cleavage Under Targets & Release Using Nuclease (CUT&Run) of the H3K4me1 mark in mNPCs of our mouse models. For a detailed description of the experimental protocol and the downstream computational analysis, see Methods in Paper II.

3.6.3 Genome-wide sequencing

Genome-Wide Sequencing (GWS) using the Nanopore from Oxford Nanopore Technology (ONT) of *Kmt2a*^{+/LSL} mNPCs genomic DNA (gDNA) was used to sequence and visualize the correct placement and orientation of the LSL cassette, as well as inspect potential CRISPR off target sites. This sequencing method was chosen since the LSL cassette is around 1.5 kb and a long-read sequencing protocol was desired to be able to sequence the entire cassette in one read. This way we could verify its completeness, exact location and orientation. For experimental protocol and the downstream analysis, see Methods in Paper II.

3.7 Statistical testing and figure rendering

Statistical testing, data analysis and plot rendering was performed using Prism 9 (v 9.4.1, GraphPad Software, Boston Massachusetts USA) and R (R Core Team, 2021). In this thesis we use p-value nomenclature as following: ns- not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Some schematic figures were created with bioRender.com.

4 Results

4.1 *In silico* comparative analysis of WDSTS causing MVs (Paper I)

4.1.1 WDSTS- associated MVs preferentially occur in the CXXC domain of KMT2A

Since disease missense variants (MVs) tend to cluster in functionally important protein domains, we wanted to assess the contribution of the multiple domains of KMT2A to WDSTS pathogenesis. We performed a comparative analysis by observing the spatial distribution of MVs within KMT2A domains found in individuals with WDSTS (68 variants) and compared them to MVs found in healthy controls and to somatic cancer variants (1403 and 1788 variants, respectively) (Figure 4A, Supplemental Figure 1 from Paper I). When compared to the healthy population, the results demonstrated a striking 61.9-fold enrichment of WDSTS variants within the CXXC domain (Fisher's exact test, Bonferroni-adjusted $p < 0.0001$), as well as 16-fold enrichment in the PHD4 domain and 6.8-fold enrichment in the PHD1 domain (Fisher's exact test, Bonferroni-adjusted, $p < 0.05$ for PHD1 and PHD3). However, we found no enrichment of WDSTS MVs within the other domains, including the enzymatic SET domain (Figure 4B). Similarly, when compared to the somatic cancer variants, which are mostly non-pathogenic passenger variants, the results remain the same, yet a little attenuated (Supplemental Figure 1 from Paper I). Next, we wanted to explore the CXXC domain further in the context of its role in WDSTS pathogenesis and found that the domain is highly conserved across species, based on its high phyloP score. However, conservation alone does not explain the importance of the CXXC domain since the SET domain displays a similar phyloP conservation score, but no MV enrichment (Supplemental Figure 2 from Paper I). Furthermore, we observed no enrichment of MVs in CXXC domains of other epigenetic regulators; DNMT1, KMT2B and TET3, that have been associated with Mendelian diseases (Figure 4C).

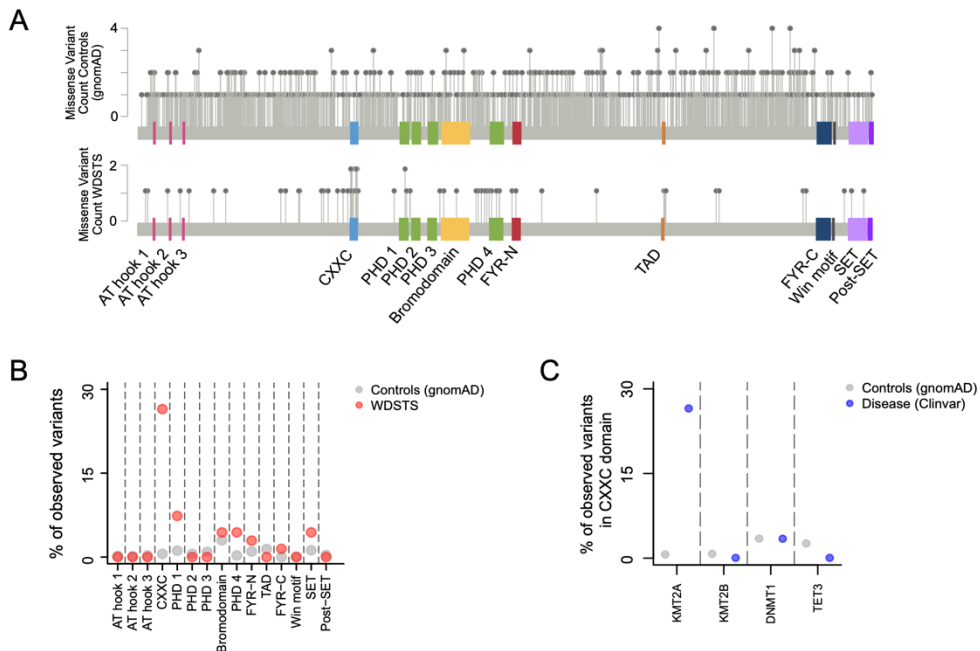


Figure 4. The distribution of likely pathogenic Wiedemann-Steiner syndrome missense variants across the different domains of KMT2A.

(A) KMT2A missense variants in gnomAD (top) and WDSTS (bottom) (B) The percentage of missense variants from gnomAD (gray dots) and WDSTS (red dots) that fall in each of the different domains of KMT2A. (C) The percentage of missense variants in gnomAD (gray dots) and likely pathogenic variants (blue dots) that fall in the CXXC domain of different epigenetic regulators. Figure reprinted and modified from (Reynisdottir et al., 2022) with permission from PLoS Genetics.

4.1.2 Effects of *Kmt2a* cKO disruption on histone methylation and gene expression

Since the CXXC domain is the DNA binding domain of KMT2A (binds to unmethylated CpG sites), we wanted to explore what effect loss of *Kmt2a* has on those regions. We gathered gene expression and histone methylation data from mice with cKO of *Kmt2a* in hippocampal CA neurons and compared them to WT mice. We discovered a strong correlation between disrupted H3K4me1/3 peaks and the peak's observed-to-expected (O/E) CpG ratio, meaning that the top differentially disrupted peaks preferentially occurred at CpG-rich sites (Fisher's exact test, $p < 0.0001$, Figure 5A). We found strong evidence that the two histone marks are mutually present at promoter regions (± 1 kb from TSS), with a decreased intensity in the *Kmt2a* cKO mice (Figure 5B). Even though the peaks with H3K4me1/3 disruptions were enriched at gene promoters (39%, 3.3-fold enrichment and 86%, 1.3-fold enrichment, respectively), the gene promoter disruption had no systematic impact on the corresponding gene expression (Wilcoxon

rank-sum test, H3K4me1: $p=0.91$, H3K3me3: $p=0.13$, Figure 5C-D).

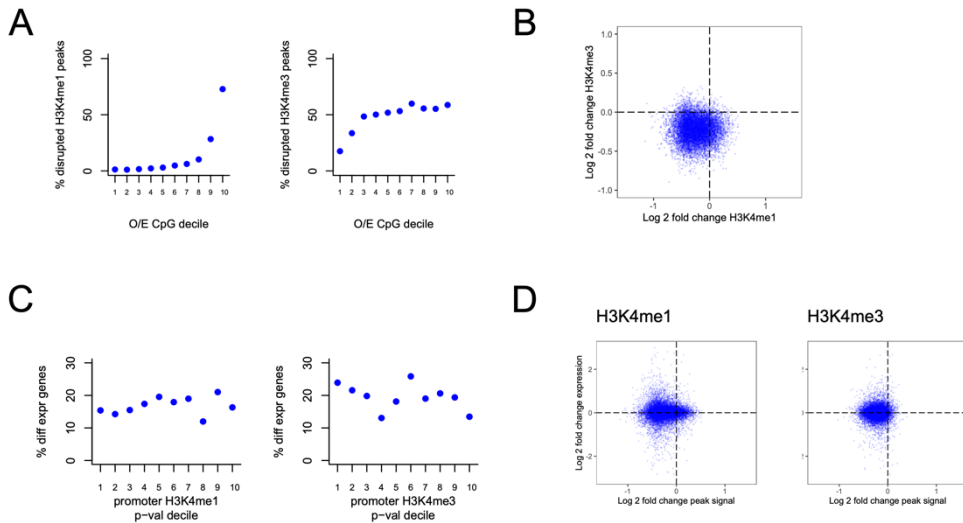


Figure 5. The relationship between disrupted H3K4me1/3, regional observed-to-expected CpG ratio, and gene expression in *Kmt2a*-deficient mice.

(A) The percentage of disrupted H3K4me1 peaks (left) and H3K4me3 peaks (right), stratified based on the observed-to-expected CpG ratio of the underlying peak sequence. (B) Scatterplot of the log₂ fold change of H3K4me1 peaks against the log₂ fold change of H3K4me3 peaks at promoters (+/-1kb from the TSS) that harbor peaks for both marks. (C) The percentage of differentially expressed genes, stratified based on the p-value of associated promoter peaks (+/-1kb from the TSS) from the differential H3K4me1 analysis (left) and H3K4me3 (right) analysis. (D) Scatterplot of the log₂ fold change of H3K4me1 peaks (left) and H3K4me3 (right) against the log₂ fold change of gene expression of the downstream gene. Each point corresponds to a gene-promoter pair. In cases where multiple peaks were present at the same promoter, the average log₂ fold change was computed. Figure reprinted from (Reynisdottir et al., 2022) with permission from PLoS Genetics.

4.1.3 Missense variant classification scheme

Our results have demonstrated a significant role of the CXXC domain in WDSTS pathogenesis, with nearly a third of WDSTS MVs found in this domain. Given the significant role of this relatively small KMT2A protein domain, we combined experimental data with *in silico* approaches (mainly AF2 protein folding predictions, see Methods in Paper I) to create a variant classification scheme that would determine the pathogenic effect of a MV in the CXXC domain. The scheme considered the position of the MV with regard to structurally important residues involved in a salt bridge bond and zinc ion binding, as well as highly conserved and functionally important residues such as the KQ and KFGG sites. The scheme also considered the changes in Coulombic electrostatic potential and the 3D structure results from AF2 (See Figure 3C from Paper I). We validated the scheme using a variant test set, revealing the

scheme to be highly accurate with a positive and negative predictive value higher than 92%. (See Figure 3D from Paper I). Once we had validated the classification scheme, we performed an *in silico* saturation mutagenesis of all potential variants within the CXXC domain that fall, classifying a total of 450 variants (Figure 6A). First, we determined that 92 variants (20.4%) were synonymous and 27 (6%) led to a premature stop codon. We then deployed the scheme to classify the remaining 331 variants, with 51.1% variants predicted to have no effect, 27.2% variants predicted to cause compromised DNA binding, 20.2% were determined to cause protein domain unfolding and for 1.5% of variants we provide no prediction (Figure 6B). To validate these results, we used orthogonal approach, demonstrating that the residues most likely to cause unfolding or disrupted DNA binding are more conserved across species, highlighting their functional relevance (Figure 6C). Finally, we found a significant overrepresentation of WDSTS MVs in the eight cysteine residues that are responsible for zinc ion binding of the domain, essential to uphold the domain's structural integrity (Fisher's exact test, $p < 0.01$).

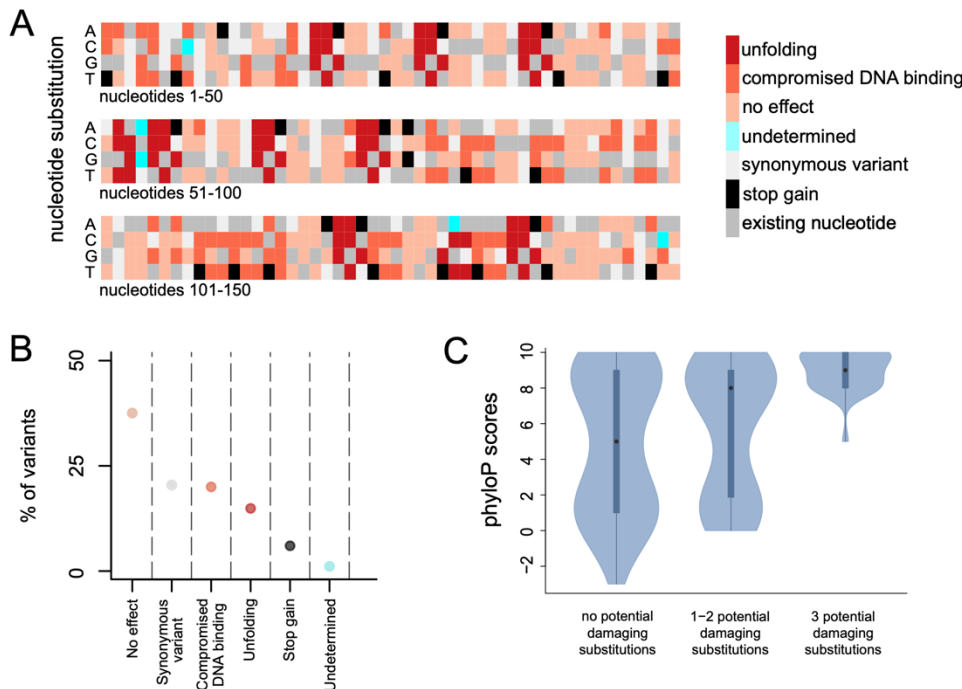


Figure 6. An *in silico* saturation mutagenesis of the CXXC domain of KMT2A.

(A) Heatmap depicting the predicted effect of each nucleotide substitution within the CXXC domain. (B) The percentage of variants for each type of predicted effect. (C) The distribution of the phyloP score of the nucleotides coding for the CXXC domain, stratified according to the number of substitutions predicted to have a damaging effect (unfolding, compromised DNA binding, stop-gain). Figure reprinted from (Reynisdottir et al., 2022) with permission from PLoS Genetics.

4.2 *In utero* malleability of neurological dysfunction of WDSTS (Paper II)

4.2.1 Creation and validation of the *Kmt2a*^{+/LSL} mouse model

In order to study the malleability of WDSTS phenotypes, we created a novel mouse model *Kmt2a*^{+/LSL} in collaboration with the Jackson Laboratory using CRISPR-Cas9 technology. The model has a heterozygous insertion of a cassette containing a terminating sequence flanked with loxP sites, in intron 1 of the *Kmt2a* gene located on mouse chromosome 9. The cassette is a LSL cassette, giving rise to the nomenclature of the mouse model, *Kmt2a*^{+/LSL}. Once we received the mice from the Jackson Laboratory to our mouse facility in Iceland, we performed a long-read ONT sequencing of their gDNA to validate the cassette's correct integration into their genome. The results demonstrated that the LSL cassette was successfully integrated into the mouse *Kmt2a*^{+/LSL} gDNA at the predicted site, with the entire cassette intact and in the correct orientation (Figure 7A). We also explored potential CRISPR off-target sites, revealing no variants nor indels at those sites (Supplemental Figure 1 from Paper II). After validating the LSL cassette we proceeded to phenotyping the model to establish WDSTS-like phenotypes in order to assess their malleability. In breeding the *Kmt2a*^{+/LSL} mice, we observed a significant skewing of the Mendelian ratios (chi-squared test, $p < 0.01$). From the mouse pups being born, there were only 40.9% *Kmt2a*^{+/LSL} pups compared to 59.1% *Kmt2a*^{+/+} pups. The model was lethal in homozygosity, with no *Kmt2a*^{LSL/LSL} pups born. Finally we validated the cassette was causing heterozygous *Kmt2a* expression in the *Kmt2a*^{+/LSL} mice (~50% reduction, $n=3-7$ per genotype, unpaired t-test, $p < 0.05$), using RT-qPCR from nervous tissue (Figure 7B).

4.2.2 Phenotyping of the *Kmt2a*^{+/LSL} mouse model

After validating that the cassette was in the correct location and functioning as expected, we moved on to phenotyping the model. Similar to what has been described in previous mouse models of heterozygous KO of *Kmt2a*, the *Kmt2a*^{+/LSL} mice demonstrated significant growth retardation ($n=14-16$ per genotype, unpaired t-test, $p < 0.001$, Figure 7C) as well as hematopoietic defects such as abnormal lymphocyte and monocyte percentages and platelet distribution width compared to their *Kmt2a*^{+/+} littermates ($n=8$ per genotype, unpaired t-test, $p < 0.05$, Supplemental Figure 3 from Paper II) (Yu et al., 1995). We also discovered an increase in hair follicle density in the abdominal area but not the back of P0 *Kmt2a*^{+/LSL} pups compared to *Kmt2a*^{+/+} littermates (discussed further in section 4.3.2, Figure 7D). Given the prominent craniofacial features of individuals with WDSTS, we CT scanned the skulls of our *Kmt2a*^{+/LSL} model and compared to wild-type littermates. We identified a significant separation between the genotypes on a PCA plot ($n=6$ per genotype, Procrustes ANOVA, $p < 0.001$) characterized by an increased width and height of the neurocranium as well as increased ventral bowing of the cranium in the *Kmt2a*^{+/LSL} mice

(Figure 7E-G). In addition, we discovered the *Kmt2a*^{+/-LSL} mice to have a noticeable gap in their interfrontal suture that is not present in their wild-type littermates (Figure 7H). Finally, we observed that the majority (57.5%) of the *Kmt2a*^{+/-LSL} mice presented with a white abdominal spot, which was not observed in the wild-type littermates, indicative of defects in the migration of melanoblasts from the neural crest, see section 4.3.3 (Figure 7I-J).

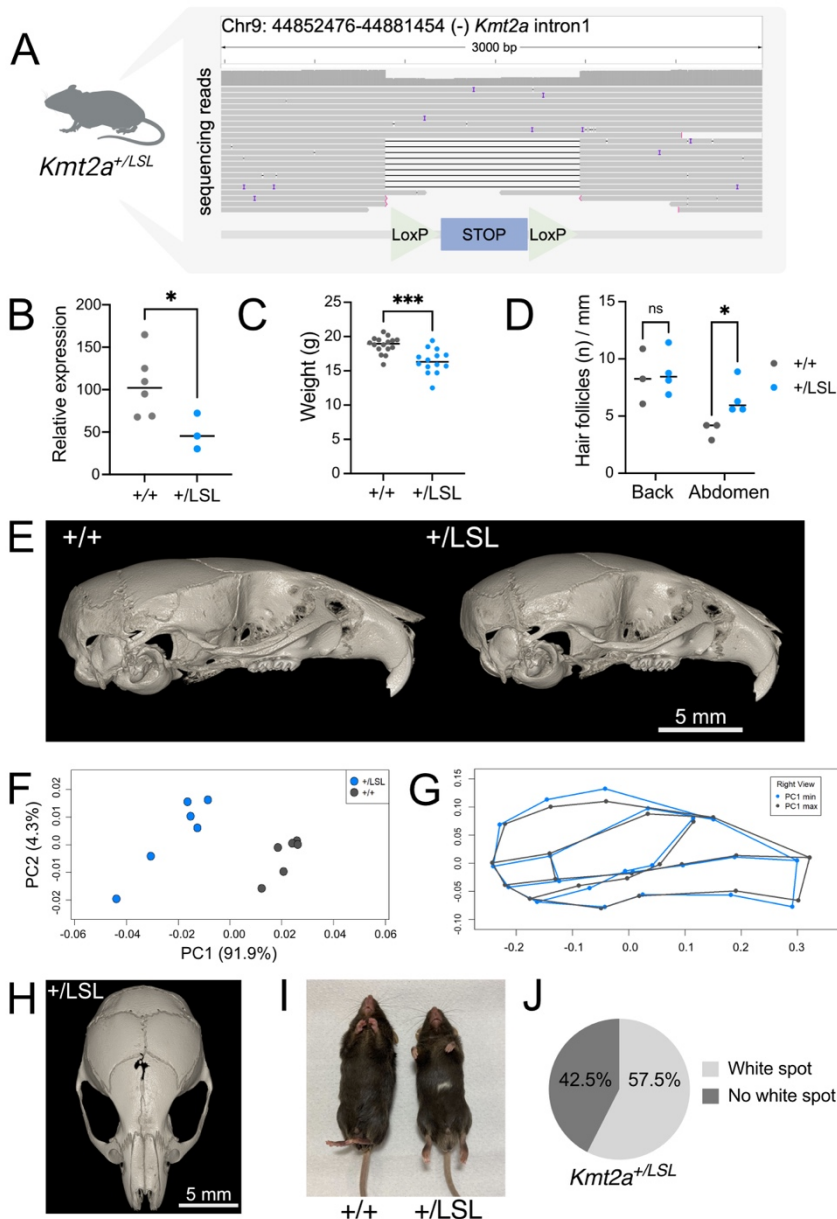


Figure 7. A novel Wiedemann-Steiner syndrome mouse model (*Kmt2a*^{+/-LSL}) recapitulates core phenotypes seen in individuals with WDSTS.

(A) Schematic figure of *Kmt2a*^{+/*LSL*} mouse model with a heterozygous loxP-stop-loxP cassette inserted in intron 1 of *Kmt2a* showing ONT sequencing reads of the cassette. Each line represents an individual long read including several that span the entire cassette as well as boundaries on either side allowing verification of its orientation. (B) Relative total *Kmt2a* mRNA expression is decreased in *Kmt2a*^{+/*LSL*} mice compared to *Kmt2a*^{+/+} littermates (n=7 *Kmt2a*^{+/+} and n=3 *Kmt2a*^{+/*LSL*}). (C) *Kmt2a*^{+/*LSL*} mice weigh significantly less compared to *Kmt2a*^{+/+} littermates (n=14-16 per genotype, 8-week females). (D) Hair follicle count per mm of skin is decreased on the abdominal area but not on the back of newborn *Kmt2a*^{+/*LSL*} pups compared to littermates (n=3-4 per genotype). (E) Representative CT images of *Kmt2a*^{+/+} and *Kmt2a*^{+/*LSL*} mice from a lateral view, depicting the craniofacial phenotype. (F) PCA plot of the craniofacial structure showing a distinct separation of the genotypes along PC1. (G) The *Kmt2a*^{+/*LSL*} mice are positioned at the negative end of PC1 and characterized by increased height and width of the neurocranium, a shortened midface, and ventral bowing of the cranium. (H) Representative figure of the gap of the interfrontal suture present in the *Kmt2a*^{+/*LSL*} mice. (I) Representative figure of the white abdominal spot of a *Kmt2a*^{+/*LSL*} mouse next to a wild-type littermate. (J) Majority of *Kmt2a*^{+/*LSL*} mice (57.5%, n=40) present with a white abdominal spot, a phenotype not present in *Kmt2a*^{+/+} littermates. Figure reprinted from (Reynisdottir et al., 2024).

4.2.3 Neurological dysfunction of the *Kmt2a*^{+/*LSL*} mouse model

Given the consistent reporting of neurological issues such as intellectual disability and developmental delay in individuals with WDSTS, we wanted to focus on establishing the neurological phenotype of our model. To start off with, we performed two hypotonia tests in young pups of both genotypes (P6, n=14-27 per genotype). The first test was a surface righting test, where the *Kmt2a*^{+/*LSL*} mice displayed an increased surface righting time compared to their wild-type littermates, suggesting a defect in their core strength and overall trunk coordination (unpaired t-test, p<0.01, Figure 8A). The second test was a hindlimb suspension test, where compared to littermates, the *Kmt2a*^{+/*LSL*} pups performed significantly worse on the third trial, indicating a lack of strength in their hindlimbs (unpaired t-test, p<0.05, Figure 8B). Next, we tested the memory of adult mice with an open armed Y maze (n=18-24 per genotype, Figure 8C). After letting the mice explore the maze freely for the duration of the test, we did not find a significant difference between the groups when comparing their total arm entries (unpaired t-test, p=0.157, Figure 8D). However, we found that the *Kmt2a*^{+/*LSL*} mice had a significant decrease in the percentage of spontaneous alternations when compared to their wild-type littermates (unpaired t-test, p<0.05, Figure 8E). This behavior test suggests that although the *Kmt2a*^{+/*LSL*} mice don't exhibit a motor phenotype, they exhibit defects in their hippocampal visuospatial and working memory. When looking into the brain structure of the mouse model, more specifically the DG of the hippocampus (Figure 8F), we observed an overall decrease in the granule cell layer area of the DG in *Kmt2a*^{+/*LSL*} mice compared to their *Kmt2a*^{+/+} littermates (n=3-4 per genotype, Figure 8G). To ensure this was not due to the previously described growth retardation of these mice, we corrected the area size to the mouse body weight, revealing a statistically significant difference between the groups (unpaired t-test, p<0.05, Figure 8H).

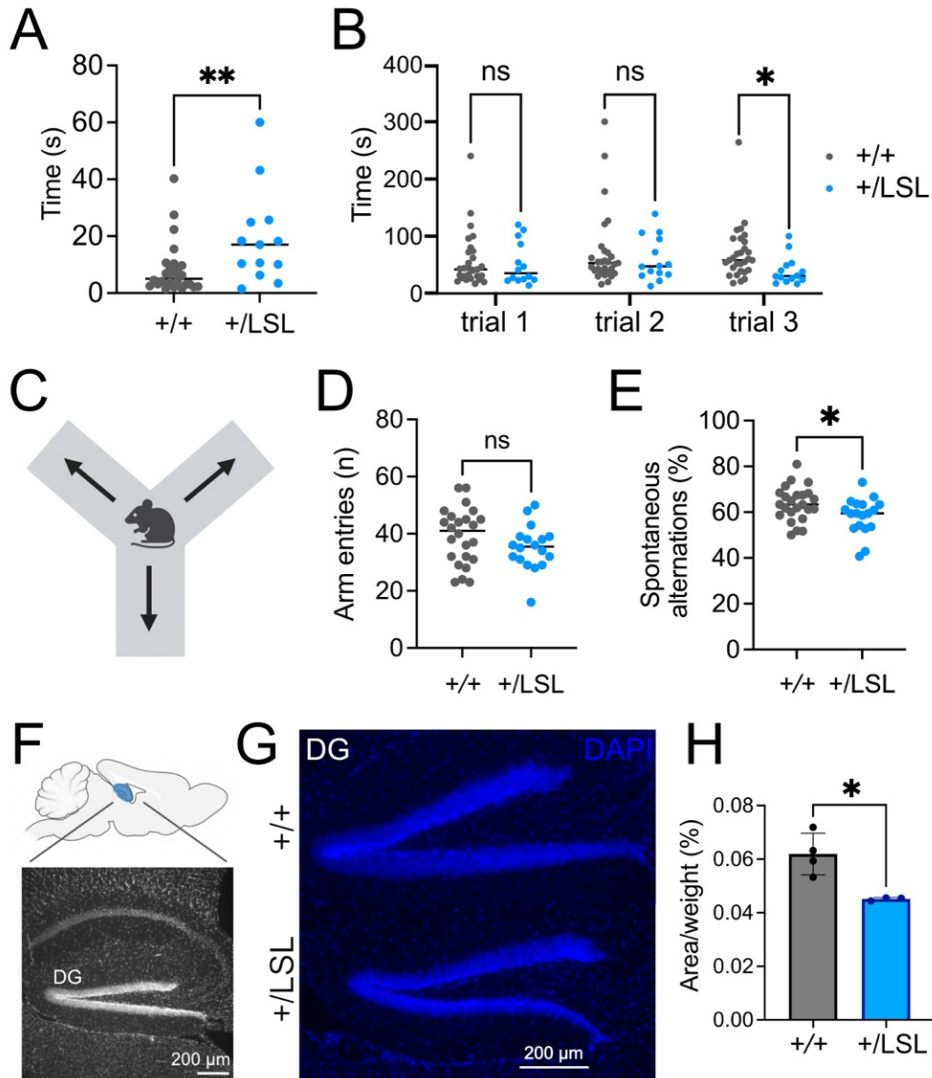


Figure 8. Neurological defects observed in the *Kmt2a*^{+/LSL} mouse model.

(A) Increased surface righting time in *Kmt2a*^{+/LSL} and (B) decreased hindlimb suspension time at the third trial in hypotonia tests (P6, n=27-29, *Kmt2a*^{+/+}, n=14-15, *Kmt2a*^{+/LSL}). (C) Schematic overview of a Y maze. A spontaneous alternation is completed when the mouse has entered each arm of the maze consecutively. (D) Adult *Kmt2a*^{+/LSL} mice do not show a significant decrease in total arm entries in a Y maze (n=24, *Kmt2a*^{+/+}, n=18, *Kmt2a*^{+/LSL}). (E) Adult *Kmt2a*^{+/LSL} mice exhibit a significant decrease in the percentage of spontaneous alternations in a Y maze. (F) Schematic figure of the mouse hippocampus with a representative image showing the DG region of the hippocampus. (G) Representative images of DAPI staining of the DG area of the hippocampus in *Kmt2a*^{+/+} and *Kmt2a*^{+/LSL} mice. (H) Quantification of granule cell layer area of the DG demonstrating a decrease in *Kmt2a*^{+/LSL} mice compared to wild-type littermates when corrected for body weight (n=3-4 per genotype). Figure reprinted from (Reynisdottir et al., 2024).

Lastly, given that severe constipation and other GI symptoms are yet another key phenotype of individuals with WDSTS, we hypothesized that the migration of neural crest derived neurons of the enteric nervous system could potentially be defective. To investigate this notion in our mouse model, we dissected distal colon sections from *Kmt2a*^{+/-LSL} mice and compared them to the *Kmt2a*^{+/+} littermates. We stained the sections for Calretinin, which is commonly used to diagnose Hirschsprung disease (HSCR) and stains for the presence of ganglion cells. HSCR is characterized by the absence of ganglion cells in the colon and rectum and shares some GI-related phenotypes with WDSTS (Zemheri et al., 2021). The results demonstrated a positive Calretinin staining in both genotypes, suggesting that there is normal presence of ganglion cells in the *Kmt2a*^{+/-LSL} model (Supplemental Figure 2D from Paper II).

4.2.4 Validation of Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mouse model

After establishing that the *Kmt2a*^{+/-LSL} model was an adequate disease model for WDSTS, displaying some core WDSTS-like phenotypes we could use these benchmark phenotypes to measure the level of phenotypic rescue in the *Kmt2a*^{+/-LSL} mice after KMT2A restoration. To test the malleability, we chose to go forward with a nestin-Cre mouse model (see section 5.2 for technical considerations), which expresses the Cre-recombinase under the nestin promoter. Once we cross the strains (*Kmt2a*^{+/-LSL} x Nes-Cre^{+/-}), the LSL cassette should be removed from the mouse genome, resulting in a knock-in (KI) of KMT2A at mid gestation in the nervous tissue of the double-mutant Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice (Figure 9A). First, we wanted to validate that the Cre-loxP recombination system of this breeding cross was functioning as expected. Using PCR of the gDNA with primers that target the genome on both sides of the LSL cassette we compared samples from the nervous tissue (hippocampus) and non-nervous tissue (skin sample from ear clips). We could visually confirm that the amount of the LSL cassette was decreased in the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice in the nervous tissue samples, compared to the *Kmt2a*^{+/-LSL} positive control (Figure 9B). Next, we assessed the regain of *Kmt2a* expression in the nervous tissue (whole brain lysates from P0 pups), using RT-qPCR. The results showed that when compared to the Nes-Cre^{+/-}*Kmt2a*^{+/+} littermates, the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} pups showed no significant difference in relative *Kmt2a* expression (n=3 per genotype, unpaired t-test, p=0.903, Figure 9C). Taken together, we concluded that the LSL cassette was being successfully recombined and removed, leading to a regained *Kmt2a* expression in the nervous tissue. These results allowed us to move forward to assess the phenotypic malleability of the *Kmt2a*^{+/-LSL} model using this strategy.

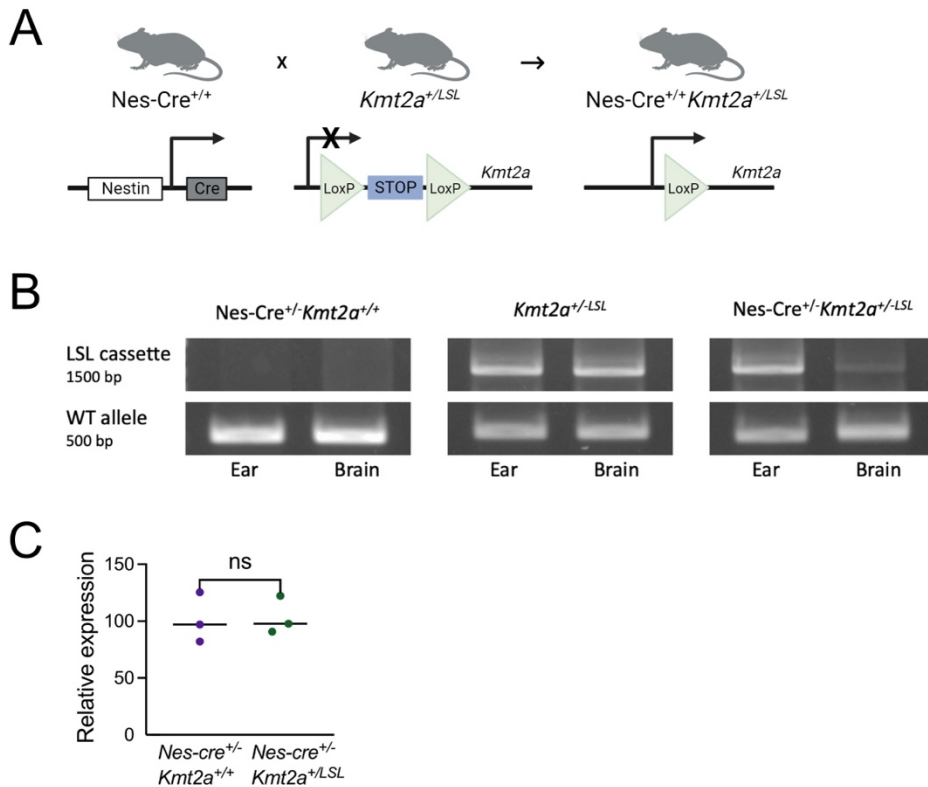


Figure 9. Validation of *Nes-Cre*^{+/+} *Kmt2a*^{+/LSL} model.

(A) A schematic overview of the nestin-Cre strategy. By exposing the *Kmt2a*^{+/LSL} mice to nestin-Cre *Kmt2a* levels are expected to be restored *in utero* in the nervous system of the *Nes-Cre*^{+/+} *Kmt2a*^{+/LSL} mice allowing us to test *in utero* rescue. (B) PCR products from gDNA samples from *Nes-Cre*^{+/+} *Kmt2a*^{+/+}, *Kmt2a*^{+/LSL} and *Nes-Cre*^{+/+} *Kmt2a*^{+/LSL} mice run on an agarose gel, comparing visually the quantity of LSL cassette in the nervous system (hippocampus) and outside the nervous system (ear clips). (C) RT-qPCR demonstrates that relative *Kmt2a* mRNA expression is unchanged in *Nes-Cre*^{+/+} *Kmt2a*^{+/LSL} mice compared to *Nes-Cre*^{+/+} *Kmt2a*^{+/+} littermates (n=3 per genotype) in nervous tissue. Figure reprinted and modified from (Reynisdottir et al., 2024).

4.2.5 Transcriptional disruption resulting from heterozygous *Kmt2a* loss

In order to assess the malleability of the *Nes-Cre*^{+/+} *Kmt2a*^{+/LSL} model, we wanted to have a closer look at what was happening on a cellular level. We dissected and cultured primary hippocampal mNPCs from newborn pups (P0 pups) from our mice. This way we could focus on a homogenous cell population from a relevant cell type to our project. Since KMT2A is a transcriptional regulator, we wanted to explore the transcriptional abnormalities caused by heterozygous loss of *Kmt2a* by performing an RNA-seq on mRNA from the mNPCs (n=3 per genotype). First, we compared the *Kmt2a*^{+/+} and *Kmt2a*^{+/LSL} samples and unveiled 64 significantly differentially expressed genes (DEGs), when applying a stringent cut-off of corrected q value < 0.05 (Figure

10A). Furthermore, we discovered that 74.4% of these genes were downregulated in the *Kmt2a*^{+/-LSL} samples, consistent with the loss of a transcriptional activator. Next, we wanted to see whether the transcriptional dysregulation was restored in our Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} model. To do this, we compared the top-decile DEGs (based on their p value) that overlapped when comparing the *Kmt2a*^{+/+} vs *Kmt2a*^{+/-LSL} and then *Kmt2a*^{+/-LSL} vs Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} samples. Here we applied a more liberal cut-off in order to capture the global transcriptional effect. We found a significant overlap of these two contrasts (Fisher's exact test, $p < 0.0001$) with a total of 438 genes. We then compared the log₂ fold change (LFC) of the genes from each contrast and plotted their correlation (Figure 10B). We found a striking negative correlation, meaning that genes that were downregulated in the *Kmt2a*^{+/-LSL} samples when comparing to the wild-type *Kmt2a*^{+/+} samples, were upregulated and restored in the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} rescue model as well as the other way around ($R^2 = 0.579$). This was a very strong effect, with 96.9% of genes showing rescue in the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mNPCs. We then wanted to have a closer look at which pathways these rescued genes are a part of and performed an over-representation analysis (ORA) revealing significant enrichment of pathways such as neuron development and differentiation (FDR < 0.05, Supplemental Figure 5 from Paper II). When plotting the expression scores (z-scores) of the genes involved in these enriched pathways, we find that the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} and *Kmt2a*^{+/+} samples cluster together, while the *Kmt2a*^{+/-LSL} samples cluster separately (Figure 10C). Moreover, we observe a similar trend as described before, with the majority of the genes (76.9%) being downregulated in the *Kmt2a*^{+/-LSL} samples. Of note, we discovered transcription factors such as *Sox5* and *En2* to be among these downregulated genes. These factors, when downregulated, have been linked to ID and ASD (Gharani et al., 2004; Schanze et al., 2013). More specifically, they have been linked to defects in hippocampal neurogenesis (Durens et al., 2021; L. Li et al., 2022). With this in mind, we asked the question if there were other ASD-risk genes amongst our rescued genes since up to half of individuals with WDSTS have also been diagnosed with ASD (R. Ng et al., 2023), and in fact we observed a significant overlap of 26 ASD-risk genes with the rescued genes (Fisher's exact test, $p < 0.0001$, Figure 10D).

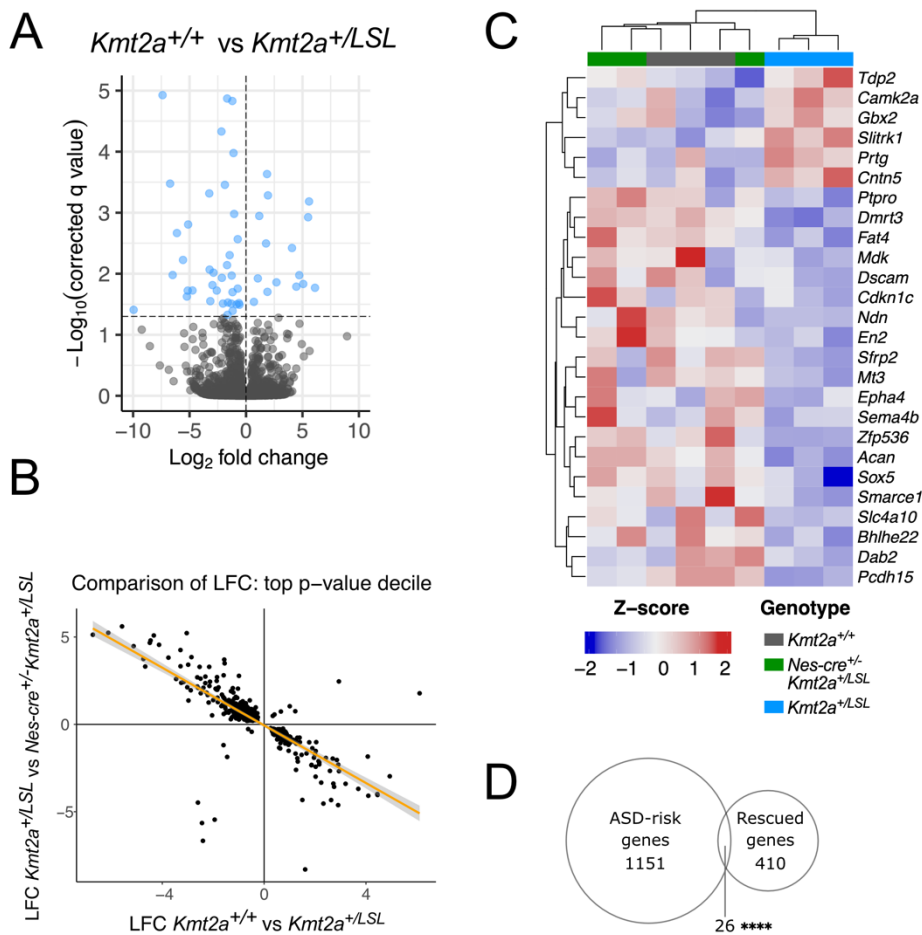


Figure 10. Transcriptional consequences of *Kmt2a* loss and rescue.

(A) Volcano plot showing the log₂ fold changes of differentially expressed genes comparing *Kmt2a*^{+/+} vs *Kmt2a*^{+/LSL} mNPCs with significant genes labeled in blue. (B) Correlation plot showing the negative correlation between log₂ fold changes between *Kmt2a*^{+/+} vs *Kmt2a*^{+/LSL} and then between *Kmt2a*^{+/LSL} vs *Nes-Cre*^{+/LSL}/*Kmt2a*^{+/LSL} mice ($R^2=0.57988$). (C) Heatmap showing clustering of *Kmt2a*^{+/LSL} samples compared to *Kmt2a*^{+/+} and *Nes-Cre*^{+/LSL}/*Kmt2a*^{+/LSL} samples of the 26 rescued genes from significantly overrepresented biological pathways (with FDR<0.05). Expression values were normalized to z-scores and plotted, with increased expression represented in red and decreased expression in blue. (D) Venn diagram of overlapping rescued genes with 1177 known ASD-risk factor genes, with 26 genes overlapping. Figure reprinted and modified from (Reynisdottir et al., 2024).

4.2.6 *Kmt2a* expression in the *Kmt2a*^{+/-LSL} mNPCs

Interestingly, we did not find *Kmt2a* to be one of the significant DEGs from the RNA-seq of our mNPCs (Wald test from DESeq2, corrected p value=0.732 for *Kmt2a*^{+/+} vs *Kmt2a*^{+/-LSL} samples and corrected p value=0.775 for *Kmt2a*^{+/-LSL} vs Nes-Cre^{+/-}/*Kmt2a*^{+/-LSL} samples). Upon further investigation of the RNA sequencing coverage, we were able to visualize an overall decrease in sequencing coverage for the full length *Kmt2a* transcripts, *Kmt2a*-201 and *Kmt2a*-202. However, we discovered a large peak of transcripts in solely the *Kmt2a*^{+/-LSL} cells mapping to the first exon as well as to a second exon, which is not present in the full-length transcripts (Figure 11A). This second exon maps to a truncated protein coding transcript, *Kmt2a*-203, that only has 9/36 exons of the full-length transcript and does not contain the enzymatic C-terminal SET domain (Figure 11B). To quantify this, we performed an isoform switch analysis, that quantifies the ratio of each isoform transcript expressed in the samples, termed isoform fraction (IF). We discovered a significant decrease in the canonical transcript *Kmt2a*-201 in the *Kmt2a*^{+/-LSL} samples (Generalized linear models (GLMs) from DEXSeq, p<0.001 and p<0.01, respectively) and a significant increase of the truncated *Kmt2a*-203 transcript (Generalized linear models (GLMs) from DEXSeq, p<0.0001 for both comparisons) compared to the *Kmt2a*^{+/+} and Nes-Cre^{+/-}/*Kmt2a*^{+/-LSL} samples (Figure 11C). These results indicate that there is a potential compensation mechanism taking place in the *Kmt2a*^{+/-LSL} cells, but more importantly, that this effect is reversed in the Nes-Cre^{+/-}/*Kmt2a*^{+/-LSL} mNPCs.

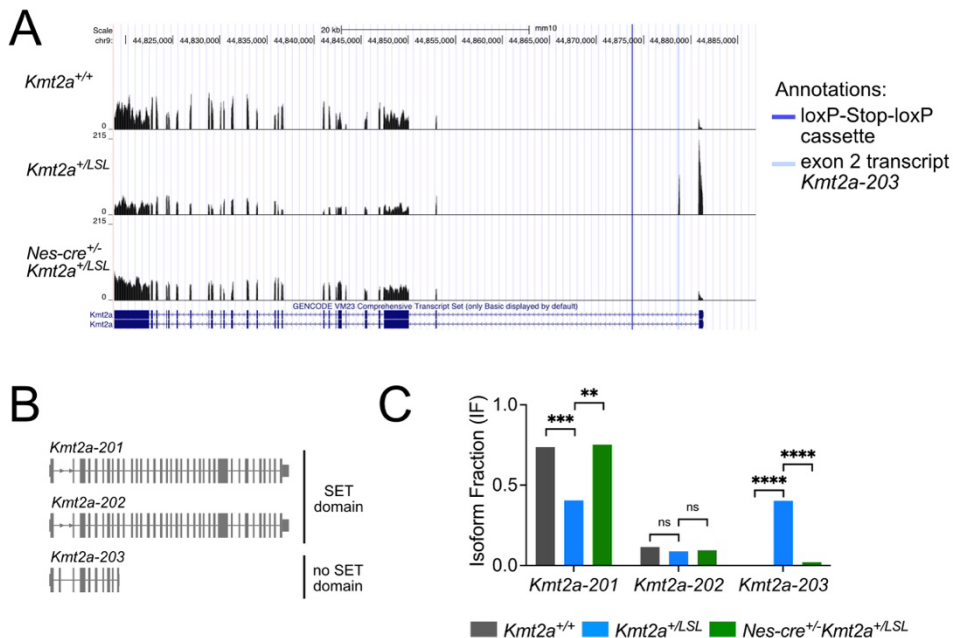


Figure 11. *Kmt2a* expression of the *Kmt2a*^{+/-LSL} mNPCs.

(A) UCSC browser tracks of genome coverage of RNA-seq files showing a decrease of *Kmt2a* coverage in *Kmt2a^{+/-LSL}* cells for full length transcripts. Only in *Kmt2a^{+/-LSL}* cells is there a peak mapped to exon 2 of the short *Kmt2a-203* transcript (B) Representative figure of protein coding *Kmt2a* isoforms and their exons. (C) Isoform switch analysis reveals a shift in isoform fraction (IF) in *Kmt2a^{+/-LSL}* mNPCs from the longer SET domain-containing *Kmt2a* transcript to the short non-SET domain containing *Kmt2a* transcript. Figure reprinted and modified from (Reynisdottir et al., 2024).

4.2.7 Chromatin abnormalities due to heterozygous *Kmt2a* loss

Since KMT2A is a known histone methyltransferase, we wanted to investigate the epigenetic effect of heterozygous *Kmt2a* loss in our *Kmt2a^{+/-LSL}* mNPCs and then assess the rescue in the *Nes-Cre^{+/-}Kmt2a^{+/-LSL}* samples. We performed a CUT&Run of the H3K4me1 mark in our mNPCs, which is known to be placed by KMT2A (n=3-4 per genotype). Interestingly, the majority of differential peaks were gained in the *Kmt2a^{+/-LSL}* mNPCs when comparing the *Kmt2a^{+/+}* vs *Kmt2a^{+/-LSL}* genotypes (70.4%, A). To explore whether the change in chromatin state was the driver for the transcriptional changes outlined in section 4.2.5, we mapped the H3K4me1 peaks to the nearest gene (by distance to TSS, see Methods in Paper II), and found no systematic disruption of expression of the corresponding genes (Error! Reference source not found.B-C). Since H3K4me1 is a known marker of enhancers, we compared the H3K4me1 signal of mouse NPC and neuronal enhancers in the *Kmt2a^{+/+}* and *Kmt2a^{+/-LSL}* samples (Figure 12D), revealing an increased H3K4me1 signal in the *Kmt2a^{+/-LSL}* samples of the neuronal enhancers, but not the NPC enhancers. Once again, this increase in H3K4me1 signal did not correlate with a differential expression change of the corresponding gene (Supplemental Figure 6 from Paper II), implying the gain of epigenetic mark is not driving gene expression changes. After establishing this increase of chromatin accessibility in the *Kmt2a^{+/-LSL}* mNPCs, we wanted to assess whether this effect is rescued in our *Nes-Cre^{+/-}Kmt2a^{+/-LSL}* samples. Strikingly, we find that 98.0% of the differential peaks previously described when comparing *Kmt2a^{+/+}* and *Kmt2a^{+/-LSL}* samples are reversed in the *Nes-Cre^{+/-}Kmt2a^{+/-LSL}* model. We visualized these results on a heatmap, showing the increased signal of *Kmt2a^{+/-LSL}* samples compared to the wild-type and rescue model (Figure 12E, Supplemental Figure 6 from Paper II). To visualize some examples, we plotted the gained H3K4me1 signal of *Kmt2a^{+/-LSL}* samples at neuronal enhancers for the genes *Acsf3* and *Bahcc1* (Figure 12F). To investigate further where the rescued H3K4me1 peaks are within the genome, we overlapped the corresponding genes of the top decile peaks (when comparing the *Kmt2a^{+/+}* to *Kmt2a^{+/-LSL}* samples and then the *Kmt2a^{+/-LSL}* to *Nes-Cre^{+/-}Kmt2a^{+/-LSL}* samples) and performed an ORA. We discovered that the pathways enriched were mainly involved in CNS development and neuron differentiation (Figure 12G), suggesting that the chromatin disruption is still taking place at functionally significant sites. Finally, we did not find a significant overlap of genes that had increased gene expression and

increased H3K4me1 signal in *Kmt2a*^{+/-LSL} samples, nor decreased gene expression and H3K4me1 signal in *Kmt2a*^{+/-LSL} samples (Figure 12H-I).

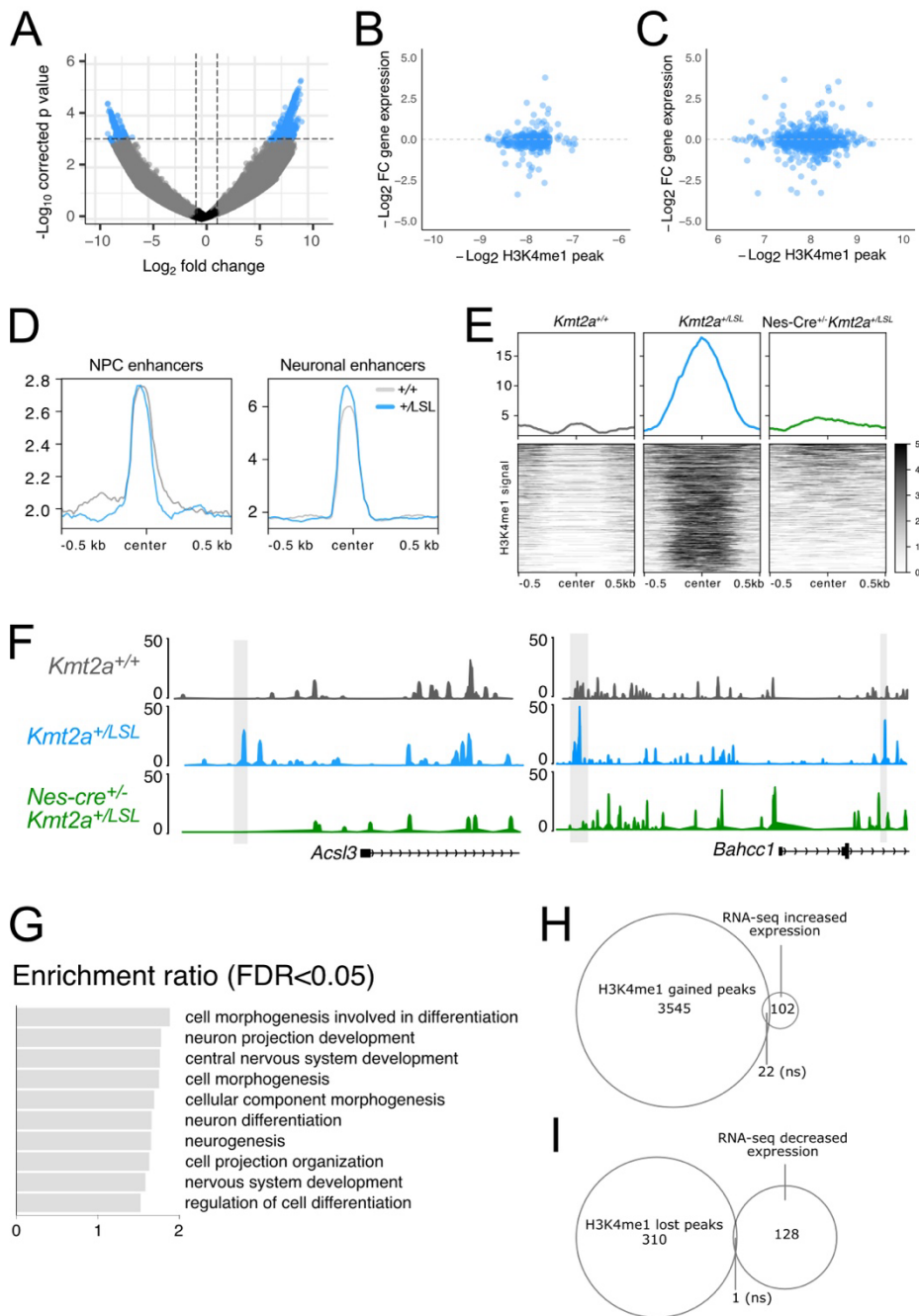


Figure 12. Epigenetic imbalance caused by heterozygous *Kmt2a* loss.

(A) Volcano plot showing differential H3K4me1 peaks comparing *Kmt2a*^{+/+} and *Kmt2a*^{+/^{LSL}} samples. Expression changes for genes with (B) lost H3K4me1 peaks and (C) gained H3K4me1 peaks in *Kmt2a*^{+/^{LSL}} samples. (D) Signal plot of H3K4me1 signal for *Kmt2a*^{+/+} and *Kmt2a*^{+/^{LSL}} samples over NPC and neuronal enhancers. (E) Heatmap of H3K4me1 signal at differential sites when comparing *Kmt2a*^{+/+}, *Kmt2a*^{+/^{LSL}} and *Nes-Cre*^{+/+}*Kmt2a*^{+/^{LSL}} samples. (F) Representative H3K4me1 signal figures, showing gained signal in *Kmt2a*^{+/^{LSL}} samples at neuronal enhancers (gray) at *Acsl3* and *Bahcc1* genes compared to wild-type and rescued samples. (G) Overrepresentation analysis of genes with differential H3K4me1 peak in *Kmt2a*^{+/^{LSL}} samples. The image shows enriched pathways (with FDR<0.05). (H) Overlap of genes with a gained H3K4me1 peak and genes that have an increased expression in *Kmt2a*^{+/^{LSL}} samples. (I) Overlap of genes with a lost H3K4me1 peak and genes that have decreased expression in *Kmt2a*^{+/^{LSL}} samples. Figure reprinted from (Reynisdottir et al., 2024).

4.2.8 *In utero* rescue of the neurological dysfunction of the *Nes-Cre*^{+/+}*Kmt2a*^{+/^{LSL}} mice

Finally, we moved on to validating our findings of the transcriptional and epigenetic rescue *in vivo* by phenotyping the *Nes-Cre*^{+/+}*Kmt2a*^{+/^{LSL}} mice. Unsurprisingly, the *Nes-Cre*^{+/+}*Kmt2a*^{+/^{LSL}} mice presented with non-neurological phenotypes described in the *Kmt2a*^{+/^{LSL}} mice in section 4.2.2, such as significant growth retardation (n=5-9 per genotype, unpaired t-test, p<0.05, Figure 13A) and similar craniofacial abnormalities (n=3-4 per genotype, Procrustes ANOVA, p=0.115) when compared to their wild-type *Nes-Cre*^{+/+}*Kmt2a*^{+/+} controls (Supplemental Figure 7 from Paper II). We discovered that the frequency of the white abdominal spot decreased from 57.5% in the *Kmt2a*^{+/^{LSL}} mice to merely 12.5% in the *Nes-Cre*^{+/+}*Kmt2a*^{+/^{LSL}} mice (Figure 13B), suggesting that there is some malleability of this phenotype in our model. More importantly, when we tested the memory of adult mice utilizing the Y maze (n=14-9, per genotype), we discovered a non-significant difference between the groups for the total arm entries (unpaired t-test, p=0.517), as well as the percentage of spontaneous alternations (unpaired t-test, p=0.961), indicating that the *Nes-Cre*^{+/+}*Kmt2a*^{+/^{LSL}} mice do not demonstrate defects in their hippocampal dependent visuospatial memory, when compared to their wild-type littermates (Figure 13C-D). In concordance, brain sections of the *Nes-Cre*^{+/+}*Kmt2a*^{+/^{LSL}} mice revealed no significant difference in the area size of the granule cell layer of the DG when corrected for body weight, compared to wild-type *Nes-Cre*^{+/+}*Kmt2a*^{+/+} mice (n=3-5 per genotype, unpaired t-test, p=0.748, Figure 13E-F).

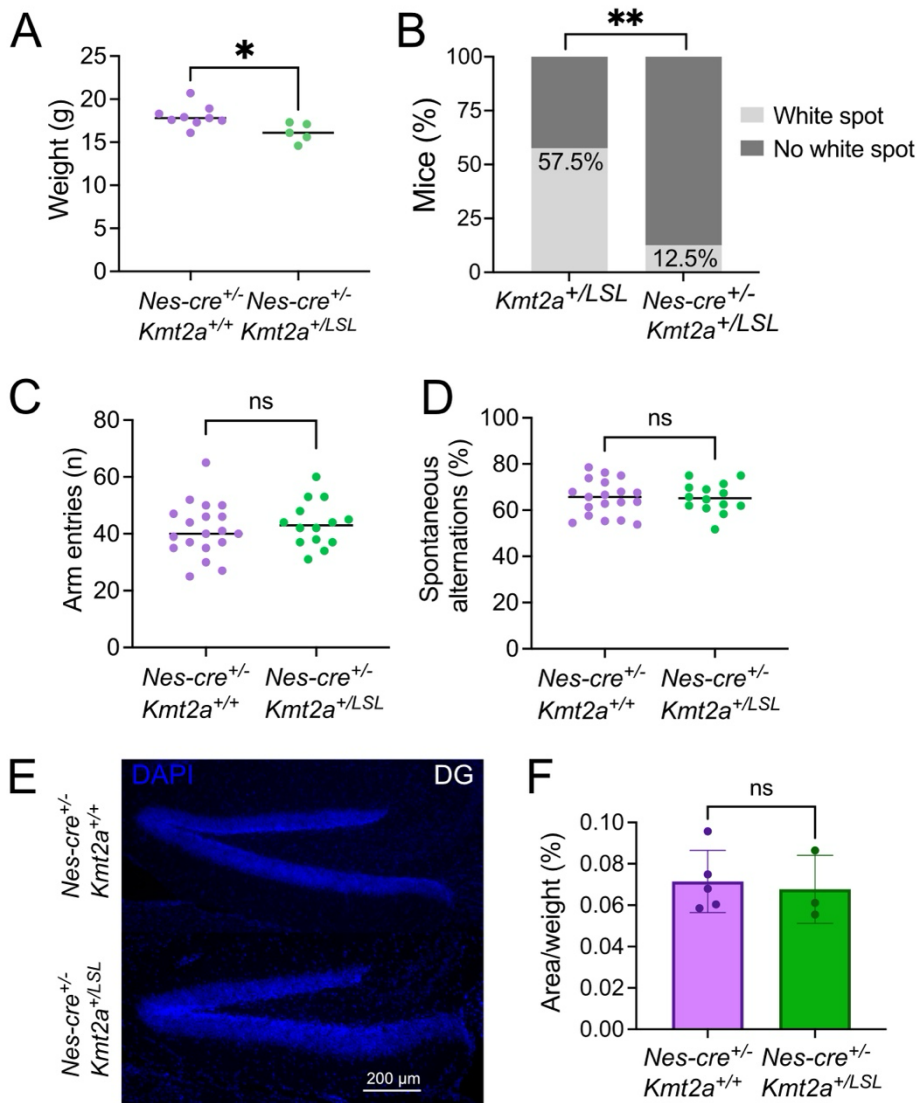


Figure 13. *In utero* rescue of the neurological phenotypes in WDSTS mice.

(A) The $Nes\text{-}Cre^{+/-}Kmt2a^{+/LSL}$ mice demonstrate growth deficiency compared to their $Nes\text{-}Cre^{+/-}Kmt2a^{+/+}$ littermates ($n=5-9$). (B) The presence of the white belly spot is significantly reduced in $Nes\text{-}Cre^{+/-}Kmt2a^{+/LSL}$ mice (12.5%) compared to $Kmt2a^{+/LSL}$ mice (57.5%) (Fisher's exact test). (C) $Nes\text{-}Cre^{+/-}Kmt2a^{+/LSL}$ mice do not show a significant decrease of total arm entries in a Y maze ($n=14-19$ per genotype) (D) $Nes\text{-}Cre^{+/-}Kmt2a^{+/LSL}$ mice do not exhibit differences in spontaneous alternations in the Y-maze, compared to their $Nes\text{-}Cre^{+/-}Kmt2a^{+/+}$ littermates. (E) Representative images of DG of the hippocampus area in $Nes\text{-}Cre^{+/-}Kmt2a^{+/+}$ and $Nes\text{-}Cre^{+/-}Kmt2a^{+/LSL}$ mice. (F) Quantification of granule cell layer area of the DG of the hippocampus demonstrates no difference in size when corrected for body weight ($n=3-5$ per genotype). Figure reprinted from (Reynisdottir et al., 2024).

4.3 Unpublished data

4.3.1 Tamoxifen induced recombination

In an effort to establish the postnatal malleability of the WDSTS phenotypes of the *Kmt2a*^{+/*LSL*} mouse model, we attempted to utilize the inducible Cre-recombinase model ROSA-CreER^{T2}. In this chapter I will discuss the results from the three optimizations I conducted to test the efficacy of this model for postnatal recombination of the LSL cassette. See section 5.2 for the discussion on the technical considerations of this model.

4.3.1.1 Optimization 1

In the first optimization, the mice were administered tamoxifen or the corn oil vehicle for treatment for three consecutive days. Their weight was monitored on the day of injections and then weekly. Weight results demonstrated a significant difference for the males when comparing the tamoxifen treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/*LSL*} mice to the tamoxifen treated ROSA-CreER^{T2+/-} as well as the non-injected ROSA-CreER^{T2+/-} controls (mixed effects model, $p < 0.05$ and $p < 0.001$, respectively, Figure 14A). For the female mice, there was no statistical difference between the tamoxifen and corn oil treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/*LSL*} mice or the tamoxifen treated ROSA-CreER^{T2+/-} mice (mixed effects model, $p = 0.667$ and $p = 0.488$, respectively, Figure 14B). However, these results are calculated with only one tamoxifen treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/*LSL*} mouse. There was not enough power to calculate the statistical difference between the ROSA-CreER^{T2+/-}*Kmt2a*^{+/*LSL*} mice and non-injected controls, since there was only one mouse in each group. Collectively, these results indicate that the tamoxifen treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/*LSL*} mice had not gained weight with treatment, although number of mice in each group are relatively low. Two weeks after the last injection, 7 of the 19 mice from the optimization had been found dead in their cages, along with 3 mice that were in a weak state. There were too few mice of each genotype left to continue the experiment, therefore all mice were sacrificed.

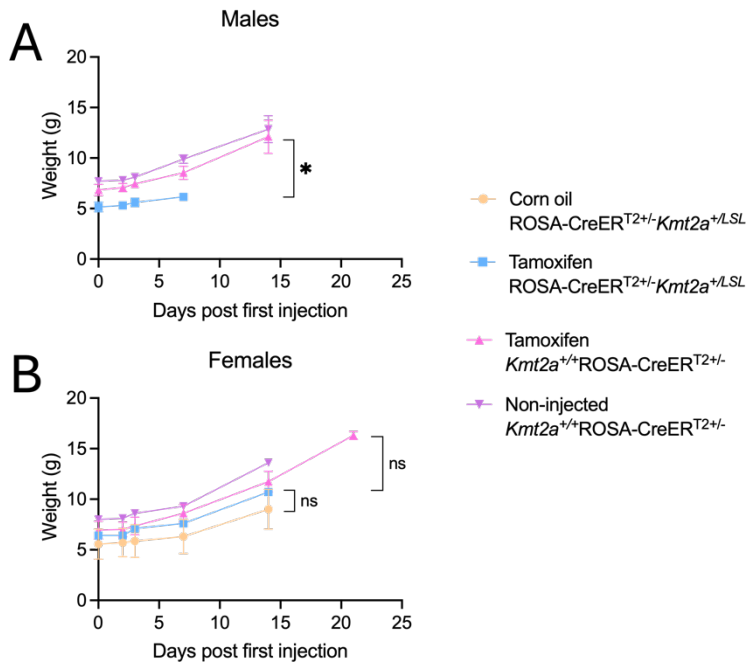


Figure 14. Mouse weight results from optimization 1.

The weight of male (A) and female (B) mice for each genotype and treatment was monitored for three consecutive days and then weekly after the first injection.

4.3.1.2 Optimization 2

In the second optimization, the pups were administered tamoxifen indirectly through the breast milk of their lactating mothers right after birth. Their weight was monitored weekly from 2 weeks of age. Results demonstrated that when comparing the tamoxifen and corn oil treated ROSA-CreERT^{2+/-}Kmt2a^{+/LSL} mice, there was no significant difference (mixed effects model, $p=0.377$ and $p=0.0748$ for females and males, respectively, Figure 15A-B). Furthermore, when comparing the tamoxifen treated ROSA-CreERT^{2+/-}Kmt2a^{+/LSL} mice to the tamoxifen treated ROSA-CreERT^{2+/-}Kmt2a^{+/+} controls, the difference is statistically significant (mixed effects model, $p<0.0001$ and $p<0.001$, for females and males, respectively, Figure 15A-B). Similarly, when comparing the tamoxifen treated ROSA-CreERT^{2+/-}Kmt2a^{+/LSL} mice to the non-injected ROSA-CreERT^{2+/-}Kmt2a^{+/+} controls the difference is statistically significant for the females (mixed effects model, $p<0.0001$). However, the difference is non-significant for the males (mixed effects model, $p=0.102$). This difference is likely due to lack of power in the number of mice. Again, these results demonstrate that the treated ROSA-CreERT^{2+/-}Kmt2a^{+/LSL} mice did not gain weight with tamoxifen treatment.

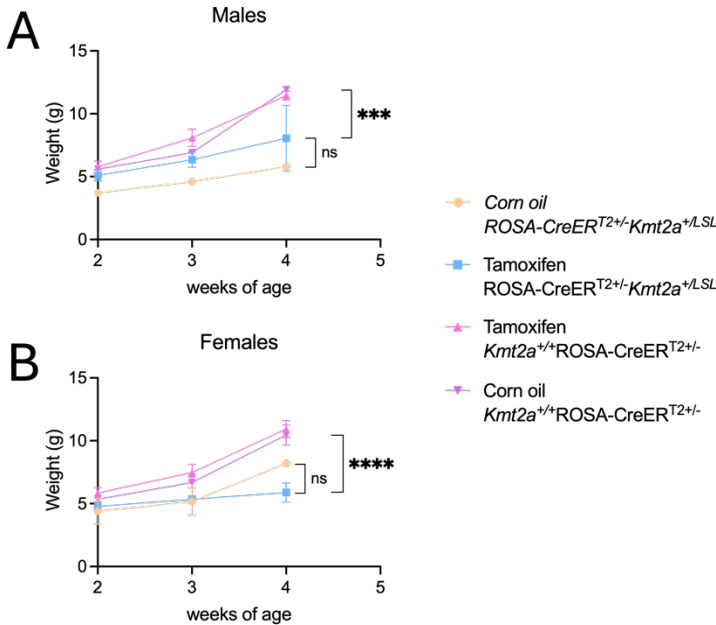


Figure 15. Mouse weight results from optimization 2.

The weight of (A) female and (B) male mice, for each genotype and treatment was monitored weekly from two weeks of age.

When the mice were 6 weeks old, they were sacrificed, dissected and cerebellum samples were collected. To assess the effectiveness of the inducible Cre-recombination in the mice, the relative *Kmt2a* expression was measured using RT-qPCR (Figure 16). The results demonstrated a significant decrease in the relative mRNA expression of tamoxifen treated ROSA-CreERT^{2+/-}Kmt2a^{+/LSL} mice when compared to the tamoxifen treated ROSA-CreERT^{2+/-}Kmt2a^{+/+} mice (unpaired t-test, $p < 0.05$), similar to what was found in the negative controls of corn oil treated ROSA-CreERT^{2+/-}Kmt2a^{+/LSL} mice (unpaired t-test, $p < 0.05$). Furthermore, there was no statistical difference between the relative *Kmt2a* expression between the corn oil and tamoxifen treated ROSA-CreERT^{2+/-}Kmt2a^{+/LSL} mice (unpaired t-test, $p = 0.359$). These results show insufficient LSL cassette recombination and that the *Kmt2a* expression was not restored in tamoxifen treated ROSA-CreERT^{2+/-}Kmt2a^{+/LSL} mice.

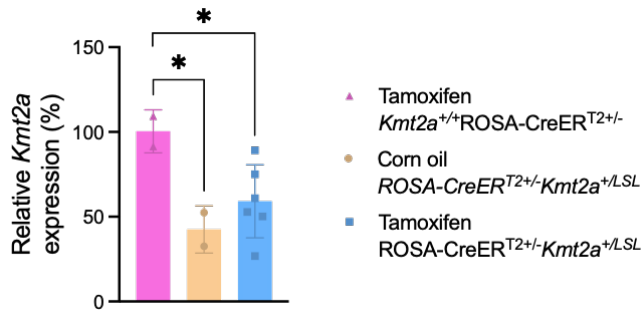


Figure 16. Relative *Kmt2a* expression from optimization 2.

4.3.1.3 Optimization 3

In the final optimization, 4-OH-tam was administered to 2-3-week-old pups for two consecutive days. Their weight was monitored on the day of injection and weekly thereafter until the mice were dissected at 12 weeks of age. For the males, there was no significant difference between the ROSA-CreER^{T2+/-}*Kmt2a*^{+/LSL} and either of the control groups. This is most likely due to a lack of power with such few mice of each genotype and treatment (mixed effects model, $p=0.092$, Figure 17A). This cohort had relatively few mice, but for the females, there was a significant difference between the weight of the tamoxifen treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/LSL} and the corn oil treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/+} mice demonstrating that the mice had not gained weight (mixed effects model, $p<0.05$, Figure 17B).

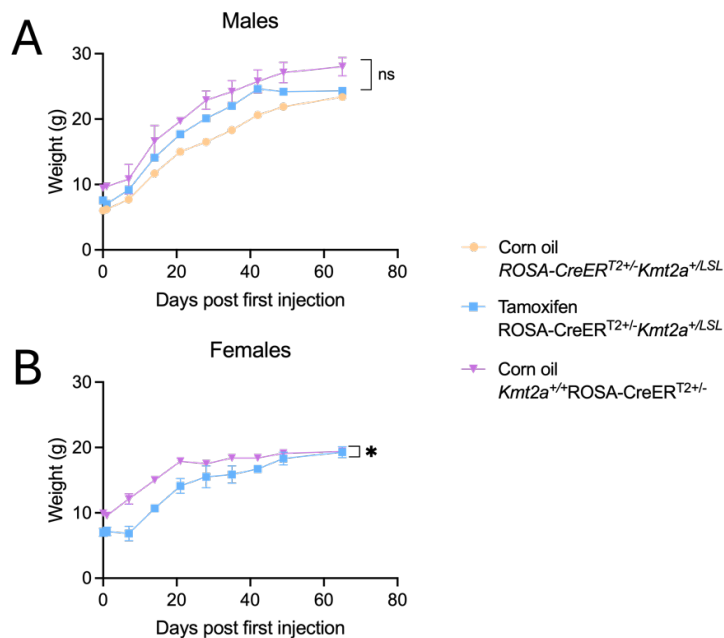


Figure 17. Mouse weight results from optimization 3.

The weight of (A) female and (B) male mice was monitored on the day of the two consecutive injections and then weekly for each genotype and treatment.

Similar to optimization 2, the mice were sacrificed, and hippocampal tissue samples were used to calculate the relative *Kmt2a* expression, to assess the efficacy of the LSL recombination. The RT-qPCR results demonstrate that there was not a significant difference between the corn oil treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/+} mice and the corn oil treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/-LSL} (unpaired t-test, $p=0.137$), however, there was only one sample in the latter group, resulting in low power of the test. Finally, when comparing the corn oil treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/+} mice to the tamoxifen treated ROSA-CreER^{T2+/-}*Kmt2a*^{+/-LSL} there was not a significant difference (unpaired t-test $p=0.062$). However, these results do not convincingly indicate that there was a regain of *Kmt2a* expression of the treated group, with the results still trending towards no rescue of *Kmt2a* expression (Figure 18).

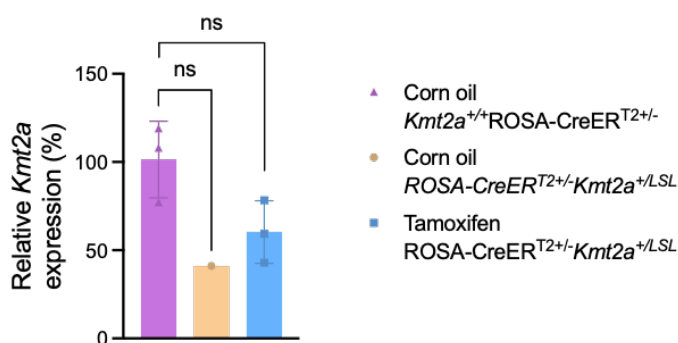


Figure 18. Relative *Kmt2a* expression from optimization 3.

4.3.2 Fur phenotype of *Kmt2a*^{+/-LSL} mice

In Paper II, we explored the hypothesis that the *Kmt2a*^{+/-LSL} mouse model would have hypertrichosis, given the consistent reporting of hypertrichosis in individuals with WDSTS. We uncovered an increased hair follicle density in the abdominal area of the *Kmt2a*^{+/-LSL} P0 pups, compared to their *Kmt2a*^{+/+} littermates suggestive of a potential regional hypertrichosis mechanism in the mice (unpaired t-test, $p<0.05$, Figure 19A-B). To roughly assess whether the *Kmt2a*^{+/-LSL} mice had increased hair growth rate, we removed the hair at the back of adult mice and visually compared the regrowth of hair between genotypes at regular intervals (Figure 19C). We discovered no obvious change in hair regrowth.

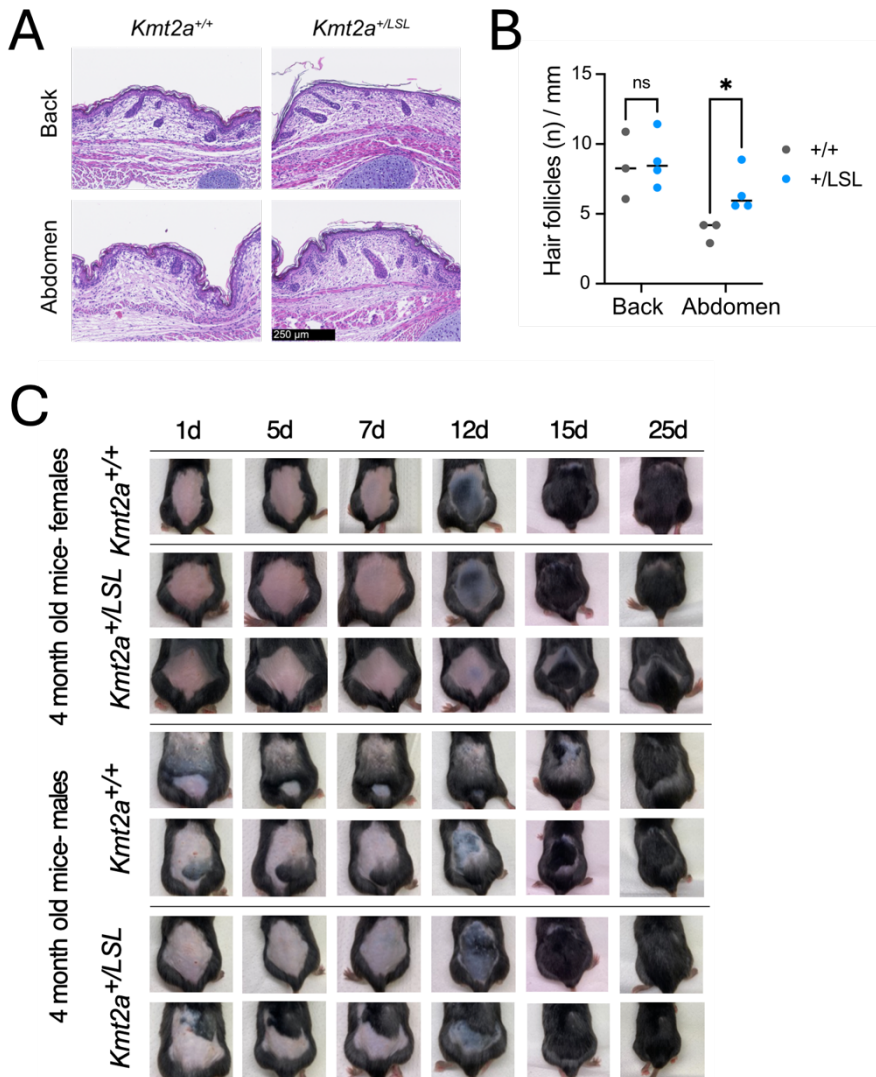


Figure 19. Hair follicle density of *Kmt2a*^{+/-LSL} mice.

(A) H&E staining of the skin of *Kmt2a*^{+/+} and *Kmt2a*^{+/-LSL} P0 pups. Scale bar is 250 μ m. (B) Hair follicle count per mm of skin is decreased in abdominal area but not on the back of newborn *Kmt2a*^{+/-LSL} pups compared to littermates. (C) Hair regrowth after hair removal for 4-month-old mice. Panels B and C are reprinted and modified from (Reynisdottir et al., 2024).

In addition to the phenotype described in Paper II, we also observed that the *Kmt2a*^{+/-LSL} mice exhibited visibly distinguishable differences in their fur compared to their *Kmt2a*^{+/+} littermates. The *Kmt2a*^{+/-LSL} mice presented with a slightly grayer and less shiny coat (Figure 20). The difference was most notable with younger mice, at 2-8 weeks of age. Furthermore, the coat often appeared scruffy and less groomed.

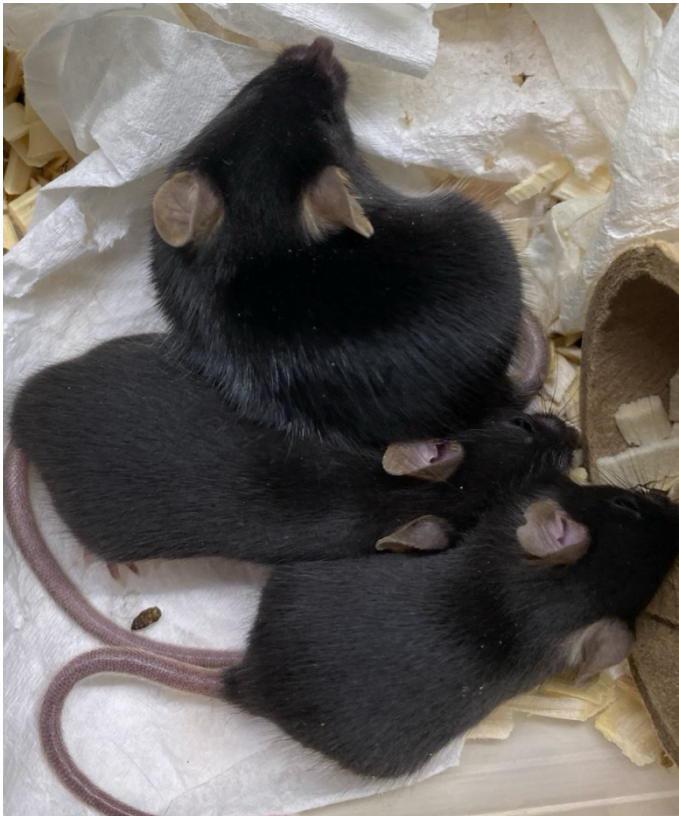


Figure 20. Fur coat of the *Kmt2a*^{+/LSL} mice.

Compared to their wild-type *Kmt2a*^{+/+} littermates (the mouse at the top of the figure), the *Kmt2a*^{+/LSL} mice (the mice in the middle and at the bottom of the figure) present with a grayer and less shiny coat.

In an effort to address whether the seemingly grayer coat is due to decreased melanin in the fur, we measured and compared the optical density (OD) of the fur lysates of young adult mice. There was not a significant difference when comparing the results between the *Kmt2a*^{+/LSL} mice and their *Kmt2a*^{+/+} littermates (unpaired t-test, $p=0.335$, Figure 21A). Next, we wanted to investigate whether we could observe a difference in the melanocytes from skin samples of P0 pups, using an IHC staining of melan-A. The results demonstrate a non-significant difference in the melan-A quantification of the P0 pups when comparing the *Kmt2a*^{+/LSL} pups and their *Kmt2a*^{+/+} littermates (unpaired t-test, $p=0.291$ for their back and $p=0.976$ for their abdominal skin samples, Figure 21B).

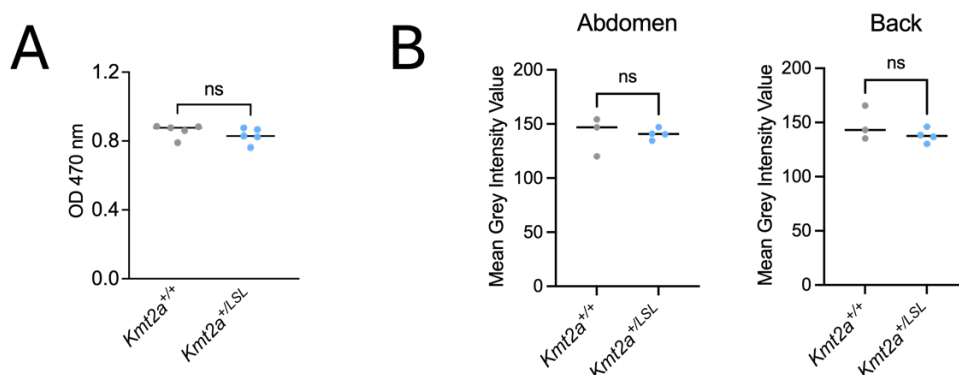


Figure 21. Melanin quantification in the fur and skin of *Kmt2a*^{+/LSL} mice.

(A) Melanin concentration measured by OD from fur lysates in *Kmt2a*^{+/+} and *Kmt2a*^{+/LSL} mice. (B) Melan-A quantification from the abdominal (left) and back (right) skin samples of PO *Kmt2a*^{+/+} and *Kmt2a*^{+/LSL} pups.

4.3.3 X-gal stainings

As discussed in section 4.2.2 above, a majority of our *Kmt2a*^{+/LSL} mice presented with a white abdominal spot, and we hypothesized this could potentially be due to a defect in the melanoblast migration from the neural crest. In an effort to address these potential defects in migration, and also to look further into melanoblast distribution of the embryo skin, we crossed the *Kmt2a*^{+/LSL} model with a *Dct-LacZ* model. We X-gal stained *Dct-LacZ*^{+/+}*Kmt2a*^{+/+} and *Dct-LacZ*^{+/+}*Kmt2a*^{+/LSL} E13.5 embryos and visualized them. We observed no gross abnormalities of the melanoblasts visually. However, we did observe less blue X-gal staining in the telencephalon of 2 out of 3 *Dct-LacZ*^{+/+}*Kmt2a*^{+/LSL} embryos (Figure 22A). When comparing the average number of melanoblasts of the torso of the mice, there was not a significant difference (unpaired t-test, $p=0.836$, Figure 22B). Next, we wanted to assess whether the melanoblasts were migrating at a different pace between genotypes, from the dorsal to the ventral part of the torso. The torso was split into three sections: the top, the middle and the bottom section (Figure 22C). When the melanoblasts within each section were quantified as a percentage of the total number of melanoblasts, there was not a significant difference in any of the sections (unpaired t-test, bottom section $p=0.182$, middle section $p=0.644$, top section $p=0.202$, Figure 22D).

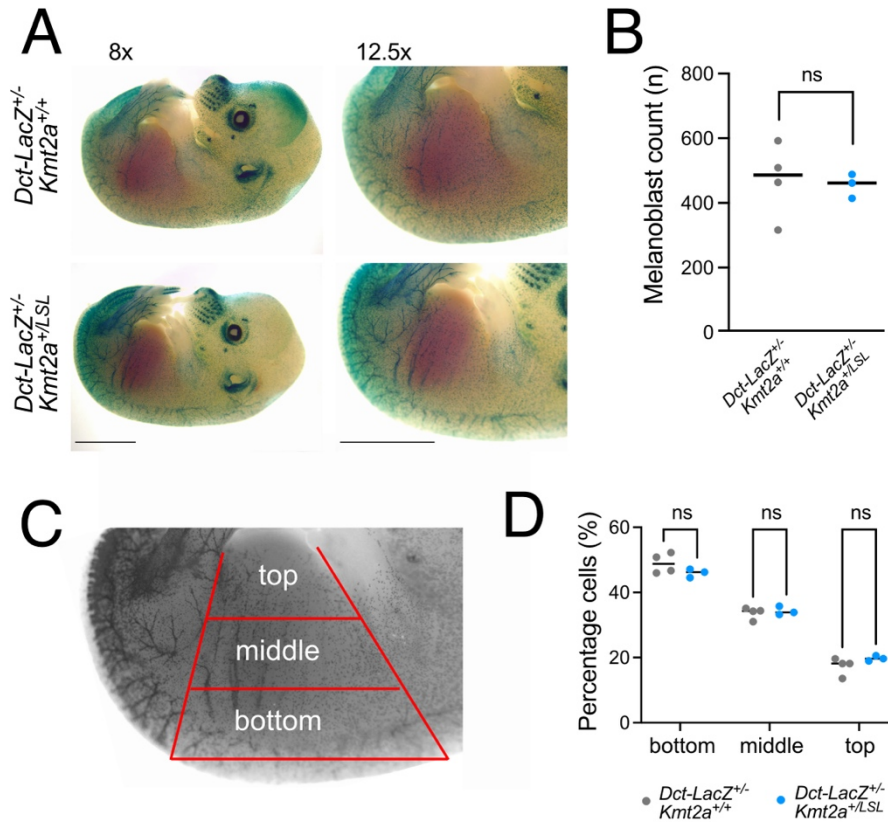


Figure 22. X-gal staining of mouse embryos.

(A) X-gal staining of *Dct-LacZ^{+/-}Kmt2a^{+/+}* (upper panel) and *Dct-LacZ^{+/-}Kmt2a^{+/LSL}* (lower panel) embryos visualized at a 8x and 12.5x magnification. Scale bar is 2.5mm. (B) Average melanoblast count (n) from the torso of *Dct-LacZ^{+/-}Kmt2a^{+/+}* and *Dct-LacZ^{+/-}Kmt2a^{+/LSL}* embryos. (C) A representative image of the three sections used for assessing the melanoblast migration (top, middle and bottom). (D) The percentage of the total melanoblasts within each section in *Dct-LacZ^{+/-}Kmt2a^{+/+}* and *Dct-LacZ^{+/-}Kmt2a^{+/LSL}* embryos.

4.3.4 Cellular phenotypes of mNPCs

As discussed in section 4.2, we harvested and cultured mNPCs from our mouse models (*Kmt2a^{+/+}*, *Kmt2a^{+/LSL}*, *Nes-Cre^{+/-}Kmt2a^{+/+}* and *Nes-Cre^{+/-}Kmt2a^{+/LSL}*). We have previously described extensive phenotyping of the mice as well as the rescue of a subset of phenotypes. We therefore wanted to explore whether the *Kmt2a^{+/LSL}* mNPCs displayed any cellular phenotypes, and whether they were reversed in the *Nes-Cre^{+/-}Kmt2a^{+/LSL}* mNPCs. We performed two phenotyping experiments, however, we only had one biological replicate for the *Kmt2a^{+/LSL}* and *Nes-Cre^{+/-}Kmt2a^{+/LSL}* mutant mNPCs and two biological replicates for the *Kmt2a^{+/+}* and *Nes-Cre^{+/-}Kmt2a^{+/+}* wild-type controls. First, we wanted to assess the proliferation rate of the cells, with a proliferation assay

that measures the growth rate with cellular confluency percentage (%). We found a significant increase in the confluency % of the *Kmt2a*^{+/-LSL} mNPCs compared to the *Kmt2a*^{+/+} controls (non-linear regression, $p < 0.0001$, Figure 23A). When comparing the *Nes-Cre*^{+/-}*Kmt2a*^{+/-LSL} with the *Nes-Cre*^{+/-}*Kmt2a*^{+/+} controls, we find that this effect is attenuated, and while it is still statistically significant, the results are based on few biological replicates (non-linear regression, $p < 0.05$, Figure 23B).

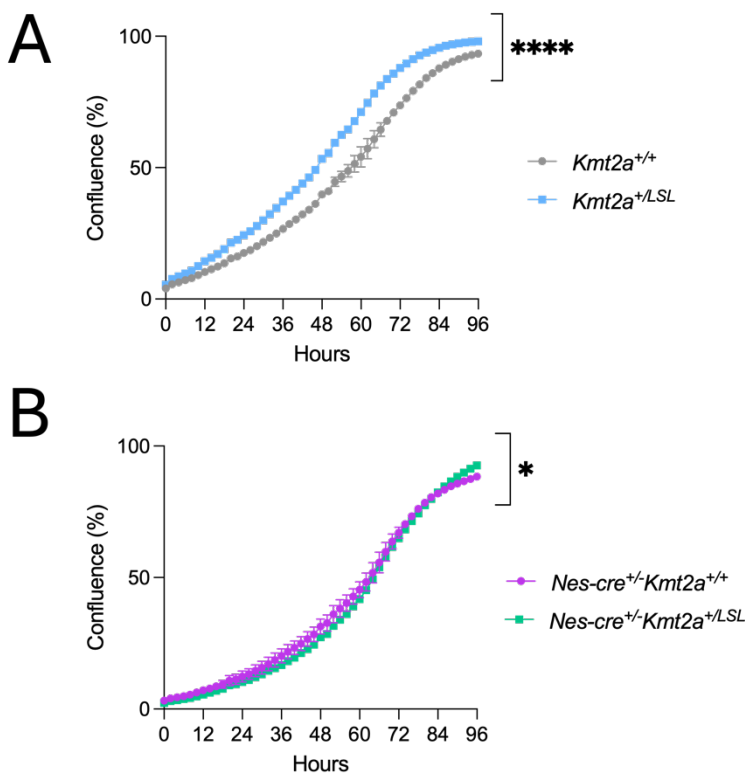


Figure 23. Proliferation rate of mNPCs.

The mNPC proliferation rate as confluency percentage (%) for (A) *Kmt2a*^{+/+} vs *Kmt2a*^{+/-LSL} mNPCs and (B) *Nes-Cre*^{+/-}*Kmt2a*^{+/+} vs *Nes-Cre*^{+/-}*Kmt2a*^{+/-LSL} mNPCs.

Lastly, in light of the frequency of the white abdominal spot of the *Kmt2a*^{+/-LSL} mouse model, as well as the prominent gap of the *Kmt2a*^{+/-LSL} mouse skull at the interfrontal suture, we hypothesized that this might be due to migratory defects of cells harboring the LSL cassette and subsequent heterozygous *Kmt2a* expression. To test this in our mNPCs, we used a scratch wound healing assay, where the cells are grown to 90% confluency and a scratch is made in the wells. The rate of the mNPCs to migrate into the wound and reach full confluency is measured in relative wound density percentage (%) (Figure 24A). We compared the *Kmt2a*^{+/+} and *Kmt2a*^{+/-LSL} mNPCs and found a non-significant difference in their migration rate (non-linear regression,

$p=0.993$, Figure 24B). Similarly, when comparing the $Nes-Cre^{+/-}Kmt2a^{+/-}$ mNPCs to the $Nes-Cre^{+/-}Kmt2a^{+/-LSL}$ mNPCs, we also found no significant difference (non-linear regression, $p=0.993$, Figure 24C). Again, more biological replicates would be needed to fully conclude whether the cells display any migratory defects.

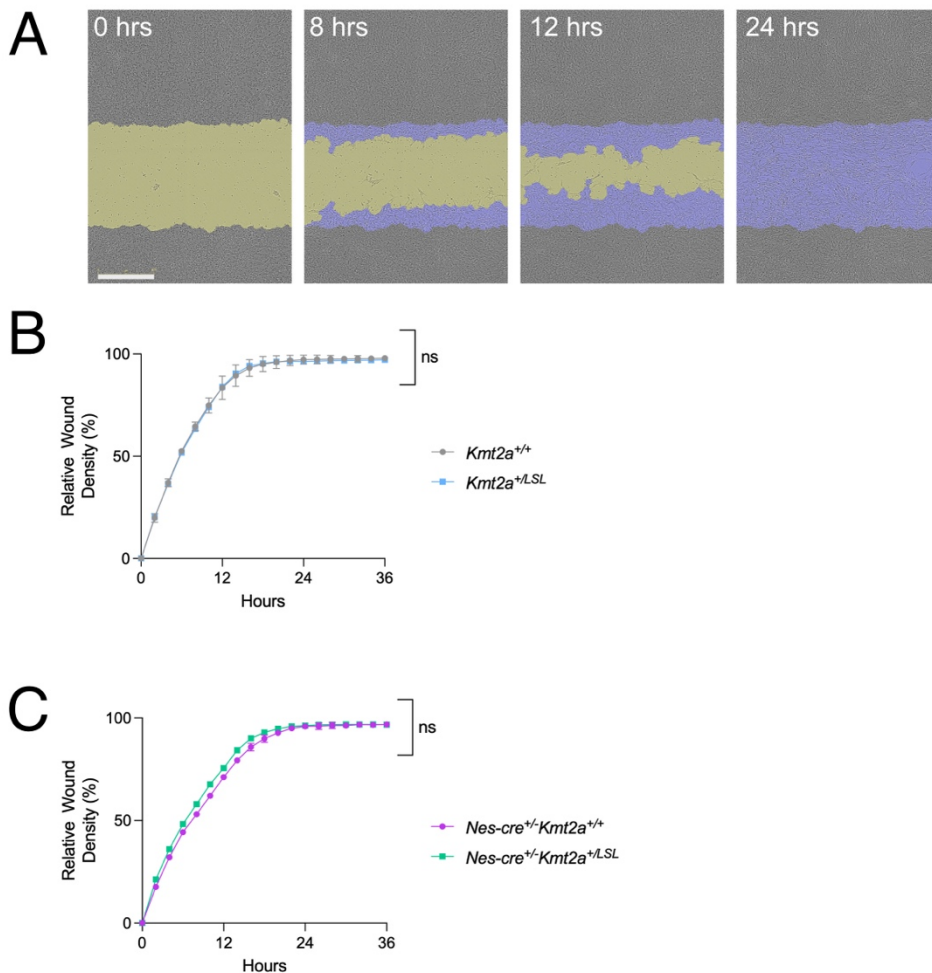


Figure 24. Scratch wound healing assay using mNPCs.

(A) An overview example of the scratch wound healing assay, where the mNPCs migrate into the scratch wound. Images taken at 0, 8, 12 and 24 hour post-scratch. Scale bar is 400 μ m. (B) The relative wound density (%) for $Kmt2a^{+/+}$ and $Kmt2a^{+/LSL}$ mNPCs comparison. (C) The relative wound density (%) for $Nes-Cre^{+/-}Kmt2a^{+/+}$ and $Nes-Cre^{+/-}Kmt2a^{+/LSL}$ mNPCs comparison.

5 Discussion

Collectively, the work described in this PhD thesis presents novel insights regarding the mechanistic basis of WDSTS. We discovered the importance of the CXXC domain in WDSTS pathogenesis and developed a novel tool to characterize potential missense variants within the KMT2A-CXXC domain as benign or disruptive. We established the first disease mouse model for studying WDSTS *in vivo*, *Kmt2a*^{+/LSL}, and characterized it in detail. The results show that the model to be a suitable disease model, displaying core WDSTS phenotypes. Lastly, we established the *in utero* malleability of the neurological dysfunction in *Kmt2a*^{+/LSL} mice. This chapter will discuss the body of work of this PhD as a whole put into context with the literature, in addition to briefly addressing its advantages and limitations.

5.1 Classification of missense variants of the CXXC domain

Classifying *de novo* missense variants as pathogenic or not remains a big challenge for clinical geneticists to this day. In Paper I, we utilized AF2, one of the many new AI tools in the biomedical field, for protein structure prediction and to assist in creating a MV classification scheme. At the time of publication, this was one of the first classifiers that directly utilized the protein prediction from AF2, which does not heavily rely on conservation scores nor population frequency. This enabled our tool to be non-biased towards these known confounding factors. AI tools have revolutionized drug development and pharmaceutical sciences, making significant advancements such as solving the protein prediction problem. They have enabled scientists to prioritize molecules for potential drug development progression and aided in antibody design. However, there are certain limitations to consider. In our variant classification scheme, we also incorporated substantial experimental data available to determine the downstream effect of MVs. We considered chemical bonds that AF2 does not account for, such as zinc ion binding of the cysteine residues as well as the electrostatic potential of residues that directly interact with the DNA. Although AI continues to improve at MV prediction, scientists should be wary of making direct causal links between changes in predicted protein structure and the pathogenicity of a variant. Proteins are known to exist in many conformational states, and therefore one should be hesitant to make that link without further experimental validation. Furthermore, AF2 relies on experimentally validated structures for its neural network that can bias the structure input towards the WT protein structure. Lastly, there are protein size and runtime limitations to AF2, which have now been greatly improved by the newly released AF3, making protein prediction for larger proteins even more accessible to scientists (Abramson et al., 2024). Overall, Paper I addressed Aim 1 of my thesis, with

the classification of MVs found in the CXXC domain of KMT2A comprising the second part of that aim. We successfully developed a novel and highly accurate MV classification scheme that enabled us to perform an *in silico* saturation mutagenesis of any potential MV found in that domain. We believe this will both serve as a relevant strategy for classifying future MVs in other proteins and domains as well as providing supportive evidence of variant pathogenicity to clinical geneticists facing novel MVs in the CXXC domain of KMT2A.

5.2 Technical challenges and considerations of the Cre mouse models

In this thesis, we explored two different approaches for causing recombination of the LSL cassette in the *Kmt2a*^{+/LSL} model to be able to assess the malleability of WDSTS phenotypes in mice. We tested the inducible ROSA-CreER^{T2} mouse model and the constitutive nestin-Cre mouse model, both of which have their advantages and disadvantages. In this chapter, we will compare the two models and discuss some technical considerations.

Our first approach to explore WDSTS disease malleability in the *Kmt2a*^{+/LSL} mice was utilizing the inducible ROSA-CreER^{T2} model. This model gives us the great flexibility to choose a time point to induce the recombination with tamoxifen, opening up an avenue to explore the window of opportunity to treat WDSTS, as well as investigate which WDSTS phenotypes are malleable postnatally. However, it proved challenging to administer the drug to reach sufficient recombination efficiency and restore *Kmt2a* expression. To assess whether the recombination was successful, we monitored the weight of the mice after tamoxifen administration, hypothesizing that rescue of KMT2A might ameliorate this well-established phenotype. However, we were uncertain if this would be the case, and it is possible the growth retardation phenotype may not be postnatally malleable. To ensure that the optimization had resulted in efficient recombination of the cassette, we also assessed the relative *Kmt2a* expression between the treated groups using RT-qPCR. In the first optimization, we administered tamoxifen as suggested by collaborators at JAX. The weight data suggested that there was not a significant increase in the body weight of the treated ROSA-CreER^{T2}^{+/LSL} *Kmt2a*^{+/LSL} in the males. In the females there was not a significant difference, but the number of mice in the cohort was too small to conclude whether the treatment was having an effect. We discovered that the tamoxifen had toxic effects on the mice, with many of them dying and others looking very weak, just weeks after administration. This is likely due, in part, to the challenges of working with tamoxifen, which is very light sensitive and can easily precipitate. It is also worth noting that multiple injections can be stressful for the mice, which can result in weight loss. However, the mice did not demonstrate a significant decrease in their body weight in general, nor did we observe a significant difference of the tamoxifen treated or non-injected ROSA-CreER^{T2}^{+/LSL} *Kmt2a*^{+/+} mice, suggesting that the mice were handling the IP injections well and that

the tamoxifen was not causing weight loss. In the second optimization, we tested a different approach by administering a lower dose of 10 mg/ml tamoxifen directly to the lactating females with an oral gavage, to ensure the conversion of the drug to the active metabolite 4-OH-tam. Here, the pups are not directly handled or injected, reducing the stress on the treated mice. However, we can not be certain how much tamoxifen each pup is receiving. The results demonstrated that the treated ROSA-CreER^{T2}/*Kmt2a*^{+/*LSL*} mice did not show a significant increase in their body weight compared to the corn oil treated negative controls. Furthermore, the RT-qPCR results demonstrated a significant decrease in relative *Kmt2a* expression in the treated group compared to the positive controls. Collectively, these data did not conclusively demonstrate that this approach restored *Kmt2a* expression. In our final optimization, we administered the active metabolite 4-OH-tam via IP injections, to ensure the potency of the drug delivered to the cells was not a confounding factor in the results. The results from the previous optimizations indicated that the effect of tamoxifen on the mice's weight was negligible so due to the high costs of 4-OH-tam, the positive and negative controls were only administered the corn oil vehicle. For the females, the corn oil treated control stopped gaining normal weight three weeks after the first injection, making it appear so that the 4-OH-tam treated mice had caught up in weight. Unfortunately, there were too few mice in this cohort, with only one mouse in three of the five different treatment groups, to confidently conclude the effects of the treatment. Ultimately, the results did not show convincing restoration of the *Kmt2a* expression.

In this thesis, we performed three optimizations to induce Cre-recombination utilizing the ROSA-CreER^{T2} mouse model. We tested tamoxifen and its active metabolite 4-OH-tam, different drug doses and administration routes. However, we did not exhaust other possibilities such as administering tamoxifen in the feed or water, although these methods make it hard to standardize the dosing (Whitfield et al., 2015; Yoshinobu et al., 2021). It is difficult to determine what exactly went wrong with the ROSA-CreER^{T2} model. One possibility is that the drug did not reach high enough potency to efficiently cause recombination of the LSL cassette. This is a very delicate balance since frequent administration of high-volume doses via IP injections or oral gavage can be stressful for laboratory rodents such as rats and mice (Stuart & Robinson, 2015). On the other hand, the drug can be toxic if the corn oil volume is decreased and the tamoxifen concentration is too high, as we saw in optimization 1. Another possibility could be that the ROSA26 promoter is not strong enough and the Cre expression was not high enough within the cells, to efficiently recombine the LSL cassette, once released into the nucleus. The ROSA26 promoter is ubiquitously expressed but is known to be weaker compared to other ubiquitously expressed promoters available i.e., the CAG promoter was found to be 9-10 times stronger and the UbC promoter was found to be 3-5 times stronger in mouse embryonic stem cells (Chen et al., 2011). Originally, our aim with this thesis was to test postnatal malleability of the WDSTS phenotypes in the *Kmt2a*^{+/*LSL*} mice using the ROSA-CreER^{T2} mouse model. Nevertheless, without

convincing results that the LSL cassette was efficiently removed and *Kmt2a* expression was regained with tamoxifen treatment, it was not suitable to move forward in assessing the phenotypic rescue of the mice. We therefore switched Cre-recombination strategies and decided to test a constitutive nestin-Cre mouse model strategy.

The first thing we needed to do with our new model was to confirm its functionality before moving on to other experiments. Since the model expresses Cre recombinase under the nestin-promoter, we confirmed the removal of the LSL cassette by PCR as well as the restoration of *Kmt2a* expression via RT-qPCR specifically in the brain. After we confirmed the model was functioning as expected, we moved on to perform the experiments described in Paper II and successfully demonstrated *in utero* malleability of the neurological dysfunction of the *Kmt2a*^{+/LSL} mice. Furthermore, we validated the function of the LSL cassette and that the Cre-lox system was functioning as expected, leaving us with a novel and valuable reagent to work with in future WDSTS research. The disadvantages to switching over to the nestin-Cre mouse model compared to the ROSA-CreER^{T2} were mainly twofold. First, we could no longer temporally control the knock-in of *Kmt2a*, and needed to change our study aims to an earlier developmental stage, from testing postnatal malleability to testing *in utero* malleability of the WDSTS phenotypes. This changes our trajectory slightly since *in utero* therapies such as IUGT are not as far along in clinical studies compared to postnatal treatments. Additionally, highly severe and early postnatally fatal diseases will be prioritized for this new technology, extending the time to get WDSTS into clinical trials (Bose et al., 2021). Nevertheless, establishing the *in utero* malleability of the WDSTS neurological dysfunction is a pivotal step in advancing this syndrome for future treatment. If the WDSTS phenotypes are not malleable *in utero*, it becomes increasingly less likely that WDSTS can be treated postnatally. Secondly, with the nestin-Cre approach, we were restricted to knocking-in the *Kmt2a* expression in the nervous system and were therefore not able to assess the malleability of WDSTS phenotypes outside the CNS and PNS. For further discussion on the nestin-Cre model, see section 5.4.1.

5.3 Characterization of the *Kmt2a*^{+/LSL} mice

In order to achieve our second and third research aims, we needed to create and characterize a WDSTS mouse model. The model needed to represent the syndrome sufficiently, in order to establish quantifiable metrics that could then be used to assess the rescue using a Cre mouse model cross. We successfully created a novel mouse model (*Kmt2a*^{+/LSL}) in collaboration with JAX through a service contract. They executed our design idea using CRISPR-Cas9 technology, where an loxP-stop-loxP (LSL) cassette was inserted into intron 1 of *Kmt2a*, resulting in a LoF variant in our mice, as commonly seen in individuals with WDSTS. We validated the correct cassette insertion placement and orientation with ONT long-range GWS and observed no variants on predicted CRISPR off-target sites. This was an essential foundation for our project since it enabled

us to confidently credit any of our findings as a result of heterozygous *Kmt2a* expression.

Although we are the first to describe a mouse model designed specifically to study WDSTS *in vivo*, there have been other mouse models published with a heterozygous KO of *Kmt2a*, giving us an opportunity to compare our model to prior studies. As previously described by Yu et al. in 1995, the *Mll^{-/-}* mice show lethality when the allele is bred to homozygosity, as seen in our model. Similar to our *Kmt2a^{+/-LSL}* model, their *Kmt2a* heterozygous mice (*Mll^{+/-}*) demonstrated hematopoietic abnormalities in their B-cell populations and platelets (Yu et al., 1995). They also described defects in their axial skeleton and sternum, however, they did not address whether they had craniofacial abnormalities. Lastly, they demonstrated that their mice had growth retardation, which was also reported by Vallianatos et al in their heterozygous *Kmt2a* cKO mice (Vallianatos et al., 2020; Yu et al., 1995). Furthermore, various neurological defects have been described in mice with heterozygous *Kmt2a* cKO, such as memory defects and altered neuronal structures (Kerimoglu et al., 2017; Vallianatos et al., 2020). Our model was in concordance with these previous findings, where we describe structural differences in the hippocampus along with defects in the hippocampal-dependent visuospatial memory. In our X-gal staining experiments, we observed significantly less blue staining in the forebrains of 2/3 *Dct-LacZ^{+/-}Kmt2a^{+/-LSL}* embryos compared to their wild-type littermates. This was an interesting finding, potentially suggesting that there were fewer neuronal progenitor cells of the telencephalon in *Dct-LacZ^{+/-}Kmt2a^{+/-LSL}* embryos due to heterozygous *Kmt2a* expression. However, this experiment would need to be repeated, and the embryos potentially sectioned, to be able to draw any conclusions from it. It is also a possibility that the mutant embryos randomly had less staining due to the staining process that took place in 15 ml Eppendorf tubes, where the X-gal stain may have not fully surrounded the embryo. Taken together, the similarities between our *Kmt2a^{+/-LSL}* mice to previous mouse models provided us with even more confidence in our model to continue with our in-depth phenotyping.

5.3.1 Fur phenotype of the *Kmt2a^{+/-LSL}* mice

We noticed early on when working with the *Kmt2a^{+/-LSL}* mouse model that its fur demonstrated considerable visual differences compared to their wild-type littermates. Given that hypertrichosis is prevalent in individuals with WDSTS, exploring the fur phenotype of the mice was an intriguing task. By sectioning and staining skin samples of P0 pups, we discovered that the *Kmt2a^{+/-LSL}* mice had increased hair follicle density on the abdominal area but not at the back, compared to wild-type littermates. Increased hair follicle density has been suggested to be one of the reasons for hypertrichosis in humans, along with an abnormal hair growth cycle or increased conversion of the soft and non-pigmented vellus hair to the coarser, pigmented terminal hair (Wendelin et al., 2003). Hypertrichosis in mice has been studied in various mouse models available that

show abnormalities in hair cycling, follicle structure or density (M. Nakamura et al., 2013). Furthermore, studies have also shown increased hair follicle density in mice with the application of external factors such as a 2-deoxy-D-ribose gel or GS-HuMSCs-CM injections (conditioned medium from human umbilical cord mesenchymal stem cells cultivated in gelatin sponge), although the underlying mechanisms are not yet fully understood (Anjum et al., 2024; Ni et al., 2024). In an effort to visualize whether we would find evidence of abnormal hair growth rate in our mice, we removed the hair on their backs using a hair removal cream and imaged the regrowth of hair every few days. We did not observe any obvious differences between genotypes although more thorough tests would be needed to rule out an abnormal hair regrowth in our mice.

The grayish appearance of the *Kmt2a*^{+/LSL} mouse fur led us to start the fur phenotyping by investigating the fur color using a few different approaches. We analyzed the melanin quantity of fur lysates of adult mice, the staining of melanocytes in the skin of newborn pups (melan-A staining), and the melanoblast counts in mouse embryos (X-gal staining). We did not observe a significant difference between genotypes in any of these tests. This could be due to the experiments we performed not being accurate or sensitive enough to detect subtle differences, or that the difference in the fur of the *Kmt2a*^{+/LSL} mice might not be related to the melanin concentration nor the melanocyte or melanoblast quantity. We did not explore other possibilities, like the composition of the different hair types of the mouse fur coat that can briefly be categorized into guard, awl, auchene and zigzag hair (Duverger & Morasso, 2009). Additionally, the scruffy appearance of the fur might be due to a lack of self-grooming in *Kmt2a*^{+/LSL} mice, and not differences in the fur itself. Further behavior testing would be needed to assess this question. Overall, we did not find a potential explanation for the difference in the fur appearance of the *Kmt2a*^{+/LSL} mice. However, this is the first time that loss of *Kmt2a* has been linked to a fur phenotype in mice, that could potentially open up new avenues for specialists in the field to explore the role KMT2A plays in mouse fur phenotypes. Collectively, these results demonstrate that we have successfully completed our second aim with the creation and characterization of the *Kmt2a*^{+/LSL} mouse model.

5.4 Malleability of the WDSTS neurological phenotype

The third and last aim of this study was to assess the malleability of the neurological dysfunction of WDSTS *in vitro* and *in vivo* using our *Kmt2a*^{+/LSL} mouse model. Here we are building upon prior research that has shown other related MDEMs such as Rubinstein-Taybi, Rett, and Kabuki syndromes to be postnatally malleable of their neurological phenotype (Alarcón et al., 2004; Benjamin et al., 2017; Guy et al., 2007). In this thesis we wanted to test whether we could find similar malleability in WDSTS. Vallianatos et al. crossed two mouse strains with heterozygous cKO of the opposing enzymes KMT2A and KDM5C in an effort to restore the enzymatic balance at conception in excitatory neurons. They discovered that by crossing the strains, they

were able to restore some neurological defects such as aggressive behavior and dendritic spine density in the double mutants (Vallianatos et al., 2020). In our approach, we designed our mouse model in a way that the LSL cassette was flanked with loxP sites. By doing this, we were able to temporally knock-in the *Kmt2a* expression in a tissue-specific manner, giving us a unique opportunity to test whether the mice displayed rescue of established phenotypes in a proof-of-concept manner. This way, we can directly attribute improvement in phenotypes in future drug treatment studies (e.g., with an HDACi treatment) to the malleability of the disease, rather than side-effects of the drug administration. Using a nestin-Cre mouse cross, we successfully completed the *in vivo* part of our third aim, showing the *in utero* rescue of neurological defects in our mouse model of WDSTS. We provided robust evidence of the efficient removal of the LSL cassette from the mouse genome, restoration of *Kmt2a* mRNA expression in nervous tissue, and subsequent rescue of structural and behavioral neurological defects in our *Kmt2a*^{+/-LSL} mice.

5.4.1 Nestin-Cre mouse model

In order to explore the malleability of the WDSTS mouse model, we decided to use the nestin-Cre mouse model, as discussed in previous chapters. The nestin-Cre model was successful in our studies, however, there are a few caveats that are important to keep in mind. First off, there is some uncertainty about when exactly the nestin expression starts and Cre recombination takes place during the embryonic development. The creators of the model state that the expression in neuronal and glial precursor cells starts as early as E11 (Tronche et al., 1999). Since then, studies have disagreed with this notion and state that the Cre expression can not cause recombination until late embryonic stages or even early postnatal days (E15.5-17.7 until P0), once Cre recombinase expression reaches high enough levels (Liang et al., 2012). Even though the ambiguity of when exactly the Cre recombination occurs in our model is a disadvantage, our results may suggest that the window of therapeutic opportunity for WDSTS tested spans late gestation and potentially into early postnatal periods. Secondly, the nestin-Cre mice have been shown to have mild hypopituitarism, which is important to keep in mind for the assessment of the rescue of our phenotypes (Declercq et al., 2015; Galichet et al., 2010). The nestin-Cre mice demonstrate decreased body weight and to ensure that the nestin-Cre genotype is not a confounding factor when assessing the rescue of the growth retardation phenotype, we compared the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} model to wild-type Nes-Cre^{+/-}*Kmt2a*^{+/+} littermates. This was also the reason it was important to correct the area size of the granule cell layer of the hippocampus for body weight. Our results still show an effect of the LSL cassette on growth. Similarly, we found that the effect of the LSL cassette on craniofacial structure was present in the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice when compared to Nes-Cre^{+/-}*Kmt2a*^{+/+} littermates. Despite not being statistically significant, the features resembled those of the *Kmt2a*^{+/-LSL} mice, with an increased height of the neurocranium, ventral bowing, and an opening of the interfrontal sutures present in the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice. In the statistical testing of the nestin-Cre

models, there were fewer landmarks collected due to them being missing in the CT scan, which might be the cause of non-significant results. These findings indicate that non-nestin expressing cell types might give rise to these phenotypes e.g., bone and cartilage derived cells, or alternatively, that the growth or craniofacial structure is altered before Cre recombination of the LSL takes place. In either instance, we would not necessarily expect a phenotypic rescue. Finally, the effects of the nestin-Cre genotype have been assessed extensively with behavioral testing by Giusti et al. The results revealed that an impairment in contextual and cued fear conditioning was observed. However, there was otherwise no significant difference in learning, memory, locomotion, exploratory activity nor a change in response to anxiety inducing triggers (Giusti et al., 2014). Reassuringly, we found no effects of the nestin-Cre genotype in our behavioral tests, with no difference between the Nes-Cre^{+/+}*Kmt2a*^{+/+} and the *Kmt2a*^{+/+} mice in the Y-maze. However, to ensure this effect is not skewing our data, we compared the Nes-Cre^{+/+}*Kmt2a*^{+/LSL} mice to Nes-Cre^{+/+}*Kmt2a*^{+/+} littermates. These data emphasizes the importance of having the necessary controls to be able to draw correct conclusions from rescue-type experiments and to be able to evaluate whether there is a confounding genotype effect that needs to be taken into consideration.

5.4.2 *In vitro* compensation and cellular rescue

Scientists should be careful when translating assumptions made from cellular data to animal models or even to human disease. Cellular experiments do not completely replicate *in vivo* environments and often include immortalized cell lines, homogenous cell populations and optimal growth conditions. In this thesis, we cultured primary neuronal progenitor cells directly from our mice and performed all cellular experiments, namely RNA sequencing and CUT&Run of the H3K4me1 mark. The benefit of this approach is that it allows us to replicate the hippocampal mNPCs environment as closely as we could, enabling us to gain a better understanding of potential underlying mechanisms, which can later be validated directly in our mouse model.

We hypothesized that there is some cellular compensation that occurs throughout development to make up for the heterozygous loss of *Kmt2a*, at a transcriptional and epigenetic level. Our RNA sequencing results revealed that in our *Kmt2a*^{+/LSL} mNPCs, where the LSL cassette is inserted into intron 1 of the canonical *Kmt2a* transcript, we observed an increased sequencing coverage of the first exon as well as a shift in isoform usage to a protein-coding but truncated *Kmt2a*-203 transcript. Prior research has shown that the full-length *Kmt2a* transcript needs to be correctly cleaved by Taspase-1 at the designated cleavage sites for subsequent heterodimerization of both parts to function properly (Hsieh et al., 2003). Overexpression of the individual N or C terminal KMT2A products leads to degradation of the products through two different pathways (Yokoyama et al., 2011). We therefore hypothesized that the truncated *Kmt2a*-203 transcript likely goes through a degradation

process like NMD at mRNA or protein level and does not have a functional role. Furthermore, the truncated transcript does not include the cleavage sites and only a few KMT2A domains, not including the enzymatic SET domain. This suggested that the cells were trying to compensate for the KMT2A expression through alternative isoform usage.

Importantly, this effect was reversed in our Nes-Cre^{+/−}*Kmt2a*^{+/LSL} mNPCs once the LSL cassette had been removed. This was a great example demonstrating a limitation to simple RNA-seq analysis using DESeq2, where all transcripts for a gene are collectively quantified together and major shifts in isoform usage within the same gene are not accounted for. It has been shown that switches in isoform usage and alternative splicing of isoforms can have great biological meaning, especially in the brain, and it is therefore important to keep this in mind when analyzing RNA-seq data (Chau et al., 2021). To ensure that when we quantified the relative *Kmt2a* mRNA expression with RT-qPCR of whole brain lysates from our mice, we designed primers that spanned the exon 30-32 junction, capturing only the full-length transcripts predicted to be functional (*Kmt2a*-201 and *Kmt2a*-202). Another finding that suggested a compensatory effect of the gene expression was the absence of a known direct target of *Kmt2a* in the top DEG list. This could mean that some of the genes that KMT2A directly targets and regulates are compensated for through other mechanisms, perhaps resulting in dysregulation of other genes that we then observed in our RNA-seq results. Previously, Vallianatos et al. described restoration of gene expression abnormalities in a mouse model with heterozygous cKO of *Kmt2a* and *Kdm5c*, restoring the enzymatic balance at conception (Vallianatos et al., 2020). Similarly, we discovered a striking reverse correlation of the dysregulated genes in our Nes-Cre^{+/−}*Kmt2a*^{+/LSL} mNPCs, suggesting that the transcriptional disruption due to heterozygous *Kmt2a* loss is restored in our rescue model *in utero*. To evaluate the effects of heterozygous loss of *Kmt2a* on the chromatin level, we performed CUT&Run on the H3K4me1 mark placed by KMT2A. Surprisingly, the directionality of our results show that the majority of differential H3K4me1 peaks when comparing the *Kmt2a*^{+/LSL} to *Kmt2a*^{+/+} samples, shows an increase of the chromatin mark in *Kmt2a*^{+/LSL}, on the contrary to what was expected for an epigenetic writer. However, we found a reversal of this effect with the majority of peaks gained in our *Kmt2a*^{+/LSL} being restored in our Nes-Cre^{+/−}*Kmt2a*^{+/LSL} mNPCs. However, this trend towards increased accessibility of the chromatin has been observed before in models of heterozygous loss of epigenetic regulators. This includes an increase of chromatin accessibility in a mouse model of KS as well as an increase in H3K4me2 in peripheral blood mononuclear cells (PBMCs) in individuals with KS (Jung et al., 2023; Luperchio et al., 2021). Additionally, an increase of the silencing mark H3K9me was observed in mouse models of Kleefstra syndrome (Iacono et al., 2018).

In the past, studies on KMT2A have often been performed in cell lines with homozygous KO of *Kmt2a* and in animal models with homozygous cKO or knock-down (KD) of *Kmt2a* (Huang et al., 2015; Lim et al., 2009). The results from these studies have produced valuable information on the role and function of KMT2A, especially

within the CNS, but should be handled with care when drawing direct conclusions about WDSTS pathogenesis. In reality, complete biallelic loss of KMT2A is lethal in humans and mice (occurring at gestation) and thus a homozygous environment might not truly recapitulate human disease. This might partly explain why we see a difference between the directionality of the chromatin marks in Paper I compared to Paper II. In Paper I, we analyzed a dataset of ChIP-seq data (H3K4me1 and H3K4me3 marks) from homozygous cKO of *Kmt2a* in hippocampal CA neurons and saw a decrease in both marks, whereas we find an increase in the H3K4me1 mark in Paper II (Kerimoglu et al., 2017). Interestingly, both Paper I and Paper II reveal a similar pattern, where the changes in epigenetic marks due to *Kmt2a* loss do not directly affect gene expression, which is globally decreased. Similarly, in a study of KS1, the H3K4me1 mark was found to not correlate with gene expression changes, but instead the changes were reflected in H3K27ac changes (Gabriele et al., 2021). We hypothesized that the increased H3K4me1 mark might be a cellular compensatory effect to try to make up for this global decrease in gene expression rather than a main driver of the gene expression dysregulation. This is in contrast to what has been proposed before, where the changes in the epigenome are attributed to being the drivers of gene expression changes (Luperchio et al., 2021). However, more studies on the interactions of these marks would be needed to fully elucidate this hypothesis as well as to distinguish which epigenetic marks might be causal and which ones might be a secondary effect. Another hypothesis as to why the H3K4me1 mark, a mark of poised enhancers, was increased in our *Kmt2a*^{+/-LSL} samples is that the increase could be due to priming of the mNPCs for precocious differentiation. In concordance with this, we found an increase in the H3K4me1 signal in *Kmt2a*^{+/-LSL} cells at neuronal enhancers but not at NPC enhancers. This could suggest that the neuronal enhancers are being primed towards a premature neuronal state. In support of this, precocious differentiation has been seen in neuronal and chondrocyte cell models of KS and could be a potential mechanism in WDSTS as well (Carosso et al., 2019; Fahrner et al., 2019).

Altogether, our mNPC experiments have yielded meaningful insights into the mechanistic basis of WDSTS on a cellular level, and our results indicate that by restoring full *Kmt2a* mRNA expression *in utero*, one is able to restore the transcriptional dysregulation and epigenetic imbalance, successfully completing the second part of our last aim.

5.5 The CXXC domain vs the SET domain

One of the central discoveries of my PhD was the implication of the CXXC domain in WDSTS pathogenesis. In Paper I, we discovered through the spatial distribution of WDSTS-causing MVs that the CXXC domain is highly enriched for pathological variants. Furthermore, we discovered the first and fourth PHD finger domains of KMT2A were enriched with WDSTS variants. These domains are known to stabilize the heterodimerization of the N- and C-terminal KMT2A protein products, a role not centred

around the enzymatic function (Yokoyama et al., 2011). These results are in contrast to what has been seen for KS, where the SET domain was enriched for MVs (Faundes et al., 2019). This is especially interesting in the light of mouse model data demonstrating that mice with homozygous truncation of the enzymatic SET domain are viable for *Kmt2a*, while being lethal in *Kmt2d* (Bjornsson et al., 2014; Yu et al., 1995). Collectively, these results fulfilled the first part of my first aim by shedding new light on the non-enzymatic role of KMT2A that might influence WDSTS pathogenesis.

5.6 Implications for future therapeutics for WDSTS

With novel diagnostic tools like whole-genome sequencing and exome sequencing, there has been a significant advancement in diagnosing genetic disorders earlier in life (Strande & Berg, 2016). This has led to multiple novel strategies in the treatment of congenital disorders, such as ASO-based therapies, gene therapy, enzyme replacement therapy, and novel CRISPR-Cas9 techniques, with a portion already making their way into clinical phase trials (Herzeg et al., 2023; Morshedzadeh et al., 2023; Scharner & Aznarez, 2021). We believe that our approach utilizing a proof-of-concept method is a suitable strategy to ask whether a disease is malleable and to what extent, as well as testing the therapeutic window of opportunity. Doing this in a systematic manner could help prioritize syndromes that show such malleability for future therapeutic development. This will hopefully provide other rare diseases with guidelines of a useful strategy that can test the malleability of a syndrome in a relatively straightforward manner, since conducting experiments can both be costly and time consuming. Moreover, we hope this research will help advance WDSTS into the next phase of therapeutic research.

5.7 Linking KMT2A to neural crest development

To this date, the role of KMT2A has been studied broadly in the body e.g., in blood and the CNS (in NPCs and neurons), however, it has not been associated with NC development (Ernst et al., 2004; Lim et al., 2009). This project had some interesting findings, potentially suggesting a novel role for KMT2A in NC development.

5.7.1 White abdominal spot

Our first big hint that KMT2A might have a role in NC development, was the prevalence of the white abdominal spot in the majority of our *Kmt2a*^{+/*LSL*} mice. The abdominal spot ranged from a few distinct white hairs to circa 1-2 cm, most commonly found along the midline, or slightly to the left. Various mouse models have been described to have regional white spots in their fur coats e.g. mice with variants in the *Mitf*, *Sox10*, *Kit*, *Kitl* and *Snai2* genes. The reason for the white spotting has been attributed to defects in the melanoblast migration from the neural crest to the rest of the embryo (Vandamme & Berx, 2019). Furthermore, mouse models have been described that specifically present

with a prominent white abdominal spot, the last area that the melanoblasts migrate to, such as mice with variants in *Pax3* and *Myo10* genes (Ohnishi et al., 2017; Ohno et al., 2017; Tokuo et al., 2018). Thus, KMT2A might directly or indirectly affect these genes causing a similar phenotype, or alternatively, be a novel cause of white abdominal spotting in mice. Interestingly, we discovered that the prevalence of the white abdominal spot in our *Kmt2a*^{+/LSL} mice drastically dropped from 57.5% to 12.5% in our *Nes-Cre*^{+/+}*Kmt2a*^{+/LSL} mice but did not fully disappear. These results indicate that there is some malleability of this phenotype and since neural crest stem cells and melanocyte precursors are known to express nestin, this might indicate that our model shows rescue of this potential neural crest phenotype (Amoh et al., 2005; L. Li et al., 2010).

5.7.2 Craniofacial abnormalities and overlap with KS

A prominent overlapping feature of WDSTS and KS are the craniofacial abnormalities. Here, we describe the craniofacial phenotype of WDSTS mouse model for the first time, which in concordance, shares similar characteristics as the abnormalities seen in a mouse model of KS (Fahrner et al., 2019). In contrast, we describe a gap in the interfrontal sutures, which was not seen in KS mice. This suggests that although there is considerable phenotypic overlap, there are some key differences in the role of KMT2A on craniofacial development compared to KMT2D. Research has demonstrated that the cranial cartilage and bone is largely derived from the facial mesenchyme, which is originates from the neural crest (S. F. Gilbert, 2000; Méndez-Maldonado et al., 2020). Moreover, it has been shown that KMT2D is a regulator of facial NCC differentiation and that the facial features of mice with KS type 1 and type 2 have been directly linked to neural crest dysfunction (Shpargel et al., 2017, 2020). This supportive evidence in KS makes it an interesting and feasible hypothesis that KMT2A plays a role in NC development. Taken together, our results could suggest that although there is a difference between KMT2A and KMT2D, there might be some shared direct targets or downstream effects leading to an overlap in phenotype.

5.7.3 KMT2A gene regulation and the neural crest

Although this PhD did not dive deeper into the established *Hox* gene regulation of KMT2A, it is worth mentioning that *Hox* genes are known to play a role in NCC differentiation and NC patterning, especially in the cranial region (Gouti et al., 2011; Gouti & Gavalas, 2008). *Hox* genes are expressed at the same time as NC formation is happening during embryonic development and LoF variants in *Hox* genes have been shown to affect NC derived tissues (Lumsden & Krumlauf, 1996; Odelin et al., 2023). Interestingly, mouse models with heterozygous *Kmt2a* loss show embryonic lethality at approximately E10.5, which is around a similar time point as the neural crest forms (E9-9.5), which potentially suggests that KMT2A is essential to NC formation (Vandamme & Berx, 2019; Yu et al., 1995).

5.7.4 Enteric nervous system defects

The prevalence of severe GI symptoms in individuals with WDSTS is high, affecting approximately 67% of individuals, with constipation and failure to thrive being the most common manifestations (Sheppard & Quintero-Rivera, 2022). Constipation in the context of neurological disorders is not uncommon and is often attributed to malfunction of the enteric nervous system or hypotonia of the smooth muscles in the gut (Camilleri, 2021; Johanson et al., 1992). Even though we did not directly test our mouse model for GI phenotypes, we discovered that the *Kmt2a*^{+/^{LSL}} mice showed indications of hypotonia in tests performed on 6 day old pups, compared to their *Kmt2a*^{+/⁺} littermates. However, after discovering a potential link between WDSTS and defects in the neural crest, we hypothesized that the GI symptoms might be due to defects in the neuronal migration of the enteric nervous system from the neural crest into the gut, as described in Hirschsprung disease (HSCR). The characteristics of HSCR include severe constipation and megacolon, which have been attributed to the absence of ganglion cells in the distal colon and rectum due to defects in neural crest migration (Lotfollahzadeh et al., 2024). To test this, we utilized a diagnostic tool for HSCR and stained the distal colon sections with Calretinin, a marker for neural ganglia, of the *Kmt2a*^{+/^{LSL}} and *Kmt2a*^{+/⁺} mice (Zemheri et al., 2021). Our results demonstrated positive Calretinin staining of both genotypes, suggesting the presence of ganglia in the gut of the *Kmt2a*^{+/^{LSL}} mice. We can therefore draw the tentative conclusion that the reason for the severe constipation in individuals with WDSTS is not due to the absence of ganglia, however, we can not determine yet whether those neurons might be defective. Similarly, we did not find a migratory defect in our mNPCs when tested in a scratch wound healing assay. Although, further experiments of the migratory defects of relevant cell types would be needed to fully exclude a migratory defect.

6 Conclusions

The overarching aim of this PhD thesis was to investigate the underlying mechanistic basis of WDSTS pathogenesis and explore the malleability of the WDSTS neurological phenotypes. To achieve this, we utilized various methods, ranging from bioinformatical approaches using open source data and our own sequencing data, to *in vitro* and *in vivo* experiments on primary neuronal progenitor cells and multiple mouse models. We consider our three aims to be met and research questions answered. However, as often in research, we are now left with multiple new avenues of interesting questions to follow up on.

6.1 Future directions

This work has further emphasized the complexity and intricacy of the epigenetic machinery. There is much research yet to be conducted to elucidate its role further and outline the role of KMT2A function and non-enzymatic function. Diving deeper into the effects of heterozygous *Kmt2a* loss on gene expression regulation and chromatin accessibility would prove very insightful. To assess this, it would be interesting to perform CUT&Run on other epigenetic marks such as H3K4me3 and H3K27ac in our cells and overlap with gene expression data. This could also be useful for investigating cellular compensation of the epigenetic imbalance. Furthermore, following up on the potential precocious differentiation phenotype could be conducted by differentiating the mNPCs further and then performing another RNA sequencing experiment to detect gene expression changes, as well as IF staining the cells with neuronal markers. Another interesting lead to further investigate from this work would be to elucidate further the importance of the CXXC domain in WDSTS pathogenesis. With almost a third of WDSTS-causing MVs within that single small domain, understanding its function better could prove very insightful. A mouse model or a relevant cell line with a MV within the CXXC domain could potentially yield clues to what the basis of the pathogenesis could be and whether that might correlate with a more severe phenotype or not.

The results from the phenotyping of the *Kmt2a*^{+/LSL} mouse model have opened up new and exciting paths to research novel roles of KMT2A, e.g. in NC development, neuronal differentiation and mouse fur phenotype. An interesting start would be to sample nervous tissue from the *Kmt2a*^{+/LSL} model at various time points in early embryogenesis and look into mRNA expression of neural crest and EMT marker genes (such as *Sox9*, *Snai1*, *Twist*, etc). This might yield interesting clues to where KMT2A might play a part in the vast and intricate signaling system of NC development. Also,

exploring how KMT2A regulates *Hox* genes in the context of NC development and disease could contribute to the field of embryology. It might also be interesting to further elucidate the fur phenotype of *Kmt2a*^{+/LSL} mice with more detailed testing of the hair follicles in skin sections and hair types.

Lastly, the major follow-up to this work should be to establish how far the window of opportunity for therapeutic intervention for WDSTS extends. We have now established a well functioning disease model, with multiple phenotypes to use as rescue metrics. This could either be done with another inducible Cre-recombinase mouse model or with drug intervention such as treatment with HDACi or inhibitors of opposing enzymes (e.g. KDM5Ci). These results could move WDSTS into clinical trials even sooner.

6.2 Closing remarks

It has been a privilege to work on this important PhD project. Studying a rare disease like WDSTS has given me the chance to contribute to many fields of science by exploring KMT2A from new angles, developing and testing novel methods and building upon prior research. I hope this PhD will have a positive impact on the lives of people with WDSTS and their families, paving the way for future therapeutic options.

References

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493–500. <https://doi.org/10.1038/s41586-024-07487-w>
- Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsдорff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Juettemann, T., Keenan, S., Laird, M. R., ... Flicek, P. (2017). Ensembl 2017. *Nucleic Acids Research*, 45(D1), D635–D642. <https://doi.org/10.1093/nar/gkw1104>
- Alarcón, J. M., Malleret, G., Touzani, K., Vronskaya, S., Ishii, S., Kandel, E. R., & Barco, A. (2004). Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: A model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration. *Neuron*, 42(6), 947–959. <https://doi.org/10.1016/j.neuron.2004.05.021>
- Alaskhar Alhamwe, B., Khalaila, R., Wolf, J., Von Bülow, V., Harb, H., Alhamdan, F., Hii, C. S., Prescott, S. L., Ferrante, A., Renz, H., Garn, H., & Potaczek, D. P. (2018). Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy, Asthma & Clinical Immunology*, 14(1), 39. <https://doi.org/10.1186/s13223-018-0259-4>
- Allen, M. D., Grummitt, C. G., Hilcenko, C., Min, S. Y., Tonkin, L. M., Johnson, C. M., Freund, S. M., Bycroft, M., & Warren, A. J. (2006). Solution structure of the

nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *The EMBO Journal*, 25(19), 4503–4512.
<https://doi.org/10.1038/sj.emboj.7601340>

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America*, 51(5), 786–794.
<https://doi.org/10.1073/pnas.51.5.786>

Allfrey, V. G., Pogo, B. G., Littau, V. C., Gershey, E. L., & Mirsky, A. E. (1968). Histone acetylation in insect chromosomes. *Science (New York, N.Y.)*, 159(3812), 314–316. <https://doi.org/10.1126/science.159.3812.314>

Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhatar, R., Shilatifard, A., Workman, J., & Zhang, Y. (2007). New Nomenclature for Chromatin-Modifying Enzymes. *Cell*, 131(4), 633–636. <https://doi.org/10.1016/j.cell.2007.10.039>

Altman, J., & Das, G. D. (1965). Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *Journal of Comparative Neurology*, 124(3), 319–335. <https://doi.org/10.1002/cne.901240303>

Altman, J., & Das, G. D. (1966). Autoradiographic and histological studies of postnatal neurogenesis. I. A longitudinal investigation of the kinetics, migration and transformation of cells incorporating tritiated thymidine in neonate rats, with special reference to postnatal neurogenesis in some brain regions. *Journal of Comparative Neurology*, 126(3), 337–389.
<https://doi.org/10.1002/cne.901260302>

Álvarez-Aznar, A., Martínez-Corral, I., Daubel, N., Betsholtz, C., Mäkinen, T., & Gaengel, K. (2020). Tamoxifen-independent recombination of reporter genes

- limits lineage tracing and mosaic analysis using CreERT2 lines. *Transgenic Research*, 29(1), 53–68. <https://doi.org/10.1007/s11248-019-00177-8>
- Amoh, Y., Li, L., Katsuoka, K., Penman, S., & Hoffman, R. M. (2005). Multipotent nestin-positive, keratin-negative hair-follicle bulge stem cells can form neurons. *Proceedings of the National Academy of Sciences*, 102(15), 5530–5534. <https://doi.org/10.1073/pnas.0501263102>
- Anjum, M. A., Zulfiqar, S., Chaudhary, A. A., Rehman, I. U., Bullock, A. J., Yar, M., & MacNeil, S. (2024). Stimulation of hair regrowth in an animal model of androgenic alopecia using 2-deoxy-D-ribose. *Frontiers in Pharmacology*, 15, 1370833. <https://doi.org/10.3389/fphar.2024.1370833>
- Arnold, C. (2023). Inside the nascent industry of AI-designed drugs. *Nature Medicine*, 29(6), 1292–1295. <https://doi.org/10.1038/s41591-023-02361-0>
- Ayton, P. M., & Cleary, M. L. (2001). Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene*, 20(40), 5695–5707. <https://doi.org/10.1038/sj.onc.1204639>
- Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C.-P., Zhao, Y., Swigut, T., & Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, 463(7283), 958–962. <https://doi.org/10.1038/nature08733>
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, 21(3), 381–395. <https://doi.org/10.1038/cr.2011.22>
- Bayer, S. A., Altman, J., Russo, R. J., & Zhang, X. (1993). Timetables of neurogenesis in the human brain based on experimentally determined patterns in the rat. *Neurotoxicology*, 14(1), 83–144.

- Benjamin, J. S., Pilarowski, G. O., Carosso, G. A., Zhang, L., Huso, D. L., Goff, L. A., Vernon, H. J., Hansen, K. D., & Bjornsson, H. T. (2017). A ketogenic diet rescues hippocampal memory defects in a mouse model of Kabuki syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 114(1), 125–130. <https://doi.org/10.1073/pnas.1611431114>
- Berlin, I., Luciani, F., Gallagher, S. J., Rambow, F., Conde-Perez, A., Colombo, S., Champeval, D., Delmas, V., & Larue, L. (2012). General strategy to analyse coat colour phenotypes in mice. *Pigment Cell & Melanoma Research*, 25(1), 117–119. <https://doi.org/10.1111/j.1755-148X.2011.00912.x>
- Betters, E., Liu, Y., Kjaeldgaard, A., Sundström, E., & García-Castro, M. I. (2010). Analysis of early human neural crest development. *Developmental Biology*, 344(2), 578–592. <https://doi.org/10.1016/j.ydbio.2010.05.012>
- Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1), 91–99. [https://doi.org/10.1016/0092-8674\(85\)90312-5](https://doi.org/10.1016/0092-8674(85)90312-5)
- Bjornsson, H. T., Benjamin, J. S., Zhang, L., Weissman, J., Gerber, E. E., Chen, Y.-C., Vaurio, R. G., Potter, M. C., Hansen, K. D., & Dietz, H. C. (2014). Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome. *Science Translational Medicine*, 6(256), 256ra135. <https://doi.org/10.1126/scitranslmed.3009278>
- Bledau, A. S., Schmidt, K., Neumann, K., Hill, U., Ciotta, G., Gupta, A., Torres, D. C., Fu, J., Kranz, A., Stewart, A. F., & Anastassiadis, K. (2014). The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development*, 141(5), 1022–1035. <https://doi.org/10.1242/dev.098152>

- Blobel, G. A., Kadauke, S., Wang, E., Lau, A. W., Zuber, J., Chou, M. M., & Vakoc, C. R. (2009). A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Molecular Cell*, 36(6), 970–983.
<https://doi.org/10.1016/j.molcel.2009.12.001>
- Boldrini, M., Fulmore, C. A., Tartt, A. N., Simeon, L. R., Pavlova, I., Poposka, V., Rosoklija, G. B., Stankov, A., Arango, V., Dwork, A. J., Hen, R., & Mann, J. J. (2018). Human Hippocampal Neurogenesis Persists throughout Aging. *Cell Stem Cell*, 22(4), 589-599.e5. <https://doi.org/10.1016/j.stem.2018.03.015>
- Boniel, S., Szymańska, K., Śmigiel, R., & Szczatuba, K. (2021). Kabuki Syndrome- Clinical Review with Molecular Aspects. *Genes*, 12(4), 468.
<https://doi.org/10.3390/genes12040468>
- Bose, S. K., Menon, P., & Peranteau, W. H. (2021). InUtero Gene Therapy: Progress and Challenges. *Trends in Molecular Medicine*, 27(8), 728–730.
<https://doi.org/10.1016/j.molmed.2021.05.007>
- Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science (New York, N.Y.)*, 317(5845), 1760–1764.
<https://doi.org/10.1126/science.1147939>
- Bouazoune, K., & Kingston, R. E. (2012). Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proceedings of the National Academy of Sciences*, 109(47), 19238–19243.
<https://doi.org/10.1073/pnas.1213825109>
- Boukas, L., Havrilla, J. M., Hickey, P. F., Quinlan, A. R., Bjornsson, H. T., & Hansen, K. D. (2019). Coexpression patterns define epigenetic regulators associated with

- neurological dysfunction. *Genome Research*, 29(4), 532–542.
<https://doi.org/10.1101/gr.239442.118>
- Boutin, C., Hardt, O., De Chevigny, A., Coré, N., Goebbels, S., Seidenfaden, R., Bosio, A., & Cremer, H. (2010). NeuroD1 induces terminal neuronal differentiation in olfactory neurogenesis. *Proceedings of the National Academy of Sciences*, 107(3), 1201–1206. <https://doi.org/10.1073/pnas.0909015107>
- Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A. J., Socci, N. D., & Scandura, J. M. (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. *PloS One*, 6(1), e14524.
<https://doi.org/10.1371/journal.pone.0014524>
- Brodal, P. (2010). *The Central Nervous System*. Oxford University Press.
- Bronner, M. E. (2012). Formation and migration of neural crest cells in the vertebrate embryo. *Histochemistry and Cell Biology*, 138(2), 179–186.
<https://doi.org/10.1007/s00418-012-0999-z>
- Bronner, M. E., & LeDouarin, N. M. (2012). Evolution and Development of the Neural Crest: An Overview. *Developmental Biology*, 366(1), 2–9.
<https://doi.org/10.1016/j.ydbio.2011.12.042>
- Brownell, J. E., & Allis, C. D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *Proceedings of the National Academy of Sciences of the United States of America*, 92(14), 6364–6368. <https://doi.org/10.1073/pnas.92.14.6364>
- Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., & Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, 84(6), 843–851. [https://doi.org/10.1016/s0092-8674\(00\)81063-6](https://doi.org/10.1016/s0092-8674(00)81063-6)

- Burgess, R. R., Travers, A. A., Dunn, J. J., & Bautz, E. K. (1969). Factor stimulating transcription by RNA polymerase. *Nature*, 221(5175), 43–46.
<https://doi.org/10.1038/221043a0>
- Burley, S. K., & Roeder, R. G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *Annual Review of Biochemistry*, 65, 769–799.
<https://doi.org/10.1146/annurev.bi.65.070196.004005>
- 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., Dupuis, L., Grafodatskaya, D., Reardon, W., Gilbert-Dussardier, B., Verloes, A., Bilan, F., Milunsky, J. M., Basran, R., Papsin, B., ... Weksberg, R. (2017). CHARGE and Kabuki Syndromes: Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions. *American Journal of Human Genetics*, 100(5), 773–788.
<https://doi.org/10.1016/j.ajhg.2017.04.004>
- Camilleri, M. (2021). Gastrointestinal motility disorders in neurologic disease. *Journal of Clinical Investigation*, 131(4), e143771. <https://doi.org/10.1172/JCI143771>
- Cao, K., Collings, C. K., Morgan, M. A., Marshall, S. A., Rendleman, E. J., Ozark, P. A., Smith, E. R., & Shilatifard, A. (2018). An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. *Science Advances*, 4(1), eaap8747.
<https://doi.org/10.1126/sciadv.aap8747>
- Carosso, G. A., Boukas, L., Augustin, J. J., Nguyen, H. N., Winer, B. L., Cannon, G. H., Robertson, J. D., Zhang, L., Hansen, K. D., Goff, L. A., & Bjornsson, H. T. (2019). Precocious neuronal differentiation and disrupted oxygen responses in Kabuki syndrome. *JCI Insight*, 4(20), e129375, 129375.
<https://doi.org/10.1172/jci.insight.129375>

- Chau, K. K., Zhang, P., Urresti, J., Amar, M., Pramod, A. B., Chen, J., Thomas, A., Corominas, R., Lin, G. N., & Iakoucheva, L. M. (2021). Full-length isoform transcriptome of the developing human brain provides further insights into autism. *Cell Reports*, 36(9), 109631.
<https://doi.org/10.1016/j.celrep.2021.109631>
- Chen, C., Krohn, J., Bhattacharya, S., & Davies, B. (2011). A comparison of exogenous promoter activity at the ROSA26 locus using a Φ C31 integrase mediated cassette exchange approach in mouse ES cells. *PloS One*, 6(8), e23376.
<https://doi.org/10.1371/journal.pone.0023376>
- Choi, J. K. (2010). Contrasting chromatin organization of CpG islands and exons in the human genome. *Genome Biology*, 11(7), R70. <https://doi.org/10.1186/gb-2010-11-7-r70>
- Christman, J. K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: Mechanistic studies and their implications for cancer therapy. *Oncogene*, 21(35), 5483–5495. <https://doi.org/10.1038/sj.onc.1205699>
- Coors, M., Bauer, L., Edwards, K., Erickson, K., Goldenberg, A., Goodale, J., Goodman, K., Grady, C., Mannino, D., Wanner, A., Wilson, T., Yarborough, M., & Zirkle, M. (2017). Ethical issues related to clinical research and rare diseases: 15th Gordon L. Snider Critical Issues Workshop, April 1, 2016, Bethesda, Maryland. *Translational Science of Rare Diseases*, 2(3–4), 175–194.
<https://doi.org/10.3233/TRD-170013>
- Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, 12, 138–163.
- Davey, M. G., Riley, J. S., Andrews, A., Tyminski, A., Limberis, M., Pogoriler, J. E., Partridge, E., Olive, A., Hedrick, H. L., Flake, A. W., & Peranteau, W. H. (2017). Induction of Immune Tolerance to Foreign Protein via Adeno-

- Associated Viral Vector Gene Transfer in Mid-Gestation Fetal Sheep. *PloS One*, 12(1), e0171132. <https://doi.org/10.1371/journal.pone.0171132>
- Day, J. J., & Sweatt, J. D. (2010). DNA methylation and memory formation. *Nature Neuroscience*, 13(11), 1319–1323. <https://doi.org/10.1038/nn.2666>
- Declercq, J., Brouwers, B., Pruniau, V. P. E. G., Stijnen, P., de Faudeur, G., Tuand, K., Meulemans, S., Serneels, L., Schraenen, A., Schuit, F., & Creemers, J. W. M. (2015). Metabolic and Behavioural Phenotypes in Nestin-Cre Mice Are Caused by Hypothalamic Expression of Human Growth Hormone. *PloS One*, 10(8), e0135502. <https://doi.org/10.1371/journal.pone.0135502>
- DeGiosio, R. A., Grubisha, M. J., MacDonald, M. L., McKinney, B. C., Camacho, C. J., & Sweet, R. A. (2022). More than a marker: Potential pathogenic functions of MAP2. *Frontiers in Molecular Neuroscience*, 15, 974890. <https://doi.org/10.3389/fnmol.2022.974890>
- Deng, S., Gan, L., Liu, C., Xu, T., Zhou, S., Guo, Y., Zhang, Z., Yang, G.-Y., Tian, H., & Tang, Y. (2023). Roles of Ependymal Cells in the Physiology and Pathology of the Central Nervous System. *Aging and Disease*, 14(2), 468–483. <https://doi.org/10.14336/AD.2022.0826-1>
- Dickinson, M. E., Selleck, M. A., McMahon, A. P., & Bronner-Fraser, M. (1995). Dorsalization of the neural tube by the non-neural ectoderm. *Development (Cambridge, England)*, 121(7), 2099–2106. <https://doi.org/10.1242/dev.121.7.2099>
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), 376–380. <https://doi.org/10.1038/nature11082>

- du Preez, L. L., & Patterton, H.-G. (2013). Secondary structures of the core histone N-terminal tails: Their role in regulating chromatin structure. *Sub-Cellular Biochemistry*, 61, 37–55. https://doi.org/10.1007/978-94-007-4525-4_2
- Durens, M., Soliman, M., Millonig, J., & DiCicco-Bloom, E. (2021). Engrailed-2 is a cell autonomous regulator of neurogenesis in cultured hippocampal neural stem cells. *Developmental Neurobiology*, 81(5), 724–735. <https://doi.org/10.1002/dneu.22824>
- Duverger, O., & Morasso, M. I. (2009). Epidermal patterning and induction of different hair types during mouse embryonic development. *Birth Defects Research. Part C, Embryo Today: Reviews*, 87(3), 263–272. <https://doi.org/10.1002/bdrc.20158>
- Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., McCune, R. A., & Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Research*, 10(8), 2709–2721. <https://doi.org/10.1093/nar/10.8.2709>
- Ernst, P., Fisher, J. K., Avery, W., Wade, S., Foy, D., & Korsmeyer, S. J. (2004). Definitive Hematopoiesis Requires the Mixed-Lineage Leukemia Gene. *Developmental Cell*, 6(3), 437–443. [https://doi.org/10.1016/S1534-5807\(04\)00061-9](https://doi.org/10.1016/S1534-5807(04)00061-9)
- Fahrner, J. A., & Bjornsson, H. T. (2014). Mendelian disorders of the epigenetic machinery: Tipping the balance of chromatin states. *Annual Review of Genomics and Human Genetics*, 15, 269–293. <https://doi.org/10.1146/annurev-genom-090613-094245>
- Fahrner, J. A., & Bjornsson, H. T. (2019). Mendelian disorders of the epigenetic machinery: Postnatal malleability and therapeutic prospects. *Human Molecular Genetics*, 28(R2), R254–R264. <https://doi.org/10.1093/hmg/ddz174>

- Fahrner, J. A., Lin, W.-Y., Riddle, R. C., Boukas, L., DeLeon, V. B., Chopra, S., Lad, S. E., Luperchio, T. R., Hansen, K. D., & Bjornsson, H. T. (2019). Precocious chondrocyte differentiation disrupts skeletal growth in Kabuki syndrome mice. *JCI Insight*, 4(20), e129380, 129380.
<https://doi.org/10.1172/jci.insight.129380>
- Fair, K., Anderson, M., Bulanova, E., Mi, H., Tropschug, M., & Diaz, M. O. (2001). Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. *Molecular and Cellular Biology*, 21(10), 3589–3597.
<https://doi.org/10.1128/MCB.21.10.3589-3597.2001>
- Fan, G., Beard, C., Chen, R. Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkiewicz, D., Bates, B., Lee, P. P., Kuhn, R., Trumpp, A., Poon, C., Wilson, C. B., & Jaenisch, R. (2001). DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 21(3), 788–797.
<https://doi.org/10.1523/JNEUROSCI.21-03-00788.2001>
- Fan, G., Martinowich, K., Chin, M. H., He, F., Fouse, S. D., Hutnick, L., Hattori, D., Ge, W., Shen, Y., Wu, H., ten Hoeve, J., Shuai, K., & Sun, Y. E. (2005). DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. *Development (Cambridge, England)*, 132(15), 3345–3356.
<https://doi.org/10.1242/dev.01912>
- Faundes, V., Malone, G., Newman, W. G., & Banka, S. (2019). A comparative analysis of KMT2D missense variants in Kabuki syndrome, cancers and the general population. *Journal of Human Genetics*, 64(2), 161–170.
<https://doi.org/10.1038/s10038-018-0536-6>
- Feng, J., Zhou, Y., Campbell, S. L., Le, T., Li, E., Sweatt, J. D., Silva, A. J., & Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic

- function in adult forebrain neurons. *Nature Neuroscience*, 13(4), 423–430.
<https://doi.org/10.1038/nn.2514>
- Filion, G. J. P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., & Defossez, P.-A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and Cellular Biology*, 26(1), 169–181.
<https://doi.org/10.1128/MCB.26.1.169-181.2006>
- Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H., & Allis, C. D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature*, 438(7071), 1116–1122. <https://doi.org/10.1038/nature04219>
- Gabriele, M., Lopez Tobon, A., D’Agostino, G., & Testa, G. (2018). The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 84, 306–327.
<https://doi.org/10.1016/j.pnpbp.2017.12.013>
- Gabriele, M., Vitriolo, A., Cuvertino, S., Pereira, M. F., Franconi, C., Germain, P.-L., Capocefalo, D., Castaldi, D., Tenderini, E., Bèchet, N. B., Millar, C., Koemans, T., Sabherwal, N., Stumpel, C., Frega, M., Palumbo, O., Carella, M., Malerba, N., Squeo, G. M., ... Testa, G. (2021). *KMT2D haploinsufficiency in Kabuki syndrome disrupts neuronal function through transcriptional and chromatin rewiring independent of H3K4-monomethylation*.
<https://doi.org/10.1101/2021.04.22.440945>
- Gabut, M., Chaudhry, S., & Blencowe, B. J. (2008). SnapShot: The splicing regulatory machinery. *Cell*, 133(1), 192.e1. <https://doi.org/10.1016/j.cell.2008.03.010>
- Galichet, C., Lovell-Badge, R., & Rizzoti, K. (2010). Nestin-Cre mice are affected by hypopituitarism, which is not due to significant activity of the transgene in the

- pituitary gland. *PLoS One*, 5(7), e11443.
<https://doi.org/10.1371/journal.pone.0011443>
- Gaudet, P., Michel, P.-A., Zahn-Zabal, M., Britan, A., Cusin, I., Domagalski, M., Duek, P. D., Gateau, A., Gleizes, A., Hinard, V., Rech de Laval, V., Lin, J., Nikitin, F., Schaeffer, M., Teixeira, D., Lane, L., & Bairoch, A. (2017). The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic Acids Research*, 45(D1), D177–D182. <https://doi.org/10.1093/nar/gkw1062>
- Gharani, N., Benayed, R., Mancuso, V., Brzustowicz, L. M., & Millonig, J. H. (2004). Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. *Molecular Psychiatry*, 9(5), 474–484.
<https://doi.org/10.1038/sj.mp.4001498>
- Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity*, 105(1), 4–13. <https://doi.org/10.1038/hdy.2010.54>
- Gilbert, P. E., & Brushfield, A. M. (2009). The role of the CA3 hippocampal subregion in spatial memory: A process oriented behavioral assessment. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 33(5), 774–781.
<https://doi.org/10.1016/j.pnpbp.2009.03.037>
- Gilbert, S. F. (2000). The Neural Crest. In *Developmental Biology*. 6th edition. Sinauer Associates. <https://www.ncbi.nlm.nih.gov/books/NBK10065/>
- Gilbert, W. (1978). Why genes in pieces? *Nature*, 271(5645), 501.
<https://doi.org/10.1038/271501a0>
- Giusti, S. A., Vercelli, C. A., Vogl, A. M., Kolarz, A. W., Pino, N. S., Deussing, J. M., & Refojo, D. (2014). Behavioral phenotyping of Nestin-Cre mice: Implications for genetic mouse models of psychiatric disorders. *Journal of Psychiatric Research*, 55, 87–95. <https://doi.org/10.1016/j.jpsychires.2014.04.002>

- Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., Van Der Hoeven, F., Tuffeland, K. R., Aasland, R., Anastassiadis, K., Ang, S.-L., & Stewart, A. F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development*, 133(8), 1423–1432.
<https://doi.org/10.1242/dev.02302>
- Goto, K., Numata, M., Komura, J. I., Ono, T., Bestor, T. H., & Kondo, H. (1994). Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. *Differentiation; Research in Biological Diversity*, 56(1–2), 39–44. <https://doi.org/10.1046/j.1432-0436.1994.56120039.x>
- Gouti, M., Briscoe, J., & Gavalas, A. (2011). Anterior Hox genes interact with components of the neural crest specification network to induce neural crest fates. *Stem Cells (Dayton, Ohio)*, 29(5), 858–870.
<https://doi.org/10.1002/stem.630>
- Gouti, M., & Gavalas, A. (2008). Hoxb1 controls cell fate specification and proliferative capacity of neural stem and progenitor cells. *Stem Cells (Dayton, Ohio)*, 26(8), 1985–1997. <https://doi.org/10.1634/stemcells.2008-0182>
- Graveley, B. R. (2001). Alternative splicing: Increasing diversity in the proteomic world. *Trends in Genetics: TIG*, 17(2), 100–107. [https://doi.org/10.1016/s0168-9525\(00\)02176-4](https://doi.org/10.1016/s0168-9525(00)02176-4)
- Grezenko, H., Ekhatior, C., Nwabugwu, N. U., Ganga, H., Affaf, M., Abdelaziz, A. M., Rehman, A., Shehryar, A., Abbasi, F. A., Bellegarde, S. B., & Khaliq, A. S. (2023). Epigenetics in Neurological and Psychiatric Disorders: A Comprehensive Review of Current Understanding and Future Perspectives. *Cureus*, 15(8), e43960. <https://doi.org/10.7759/cureus.43960>

- Guenther, M. G., Jenner, R. G., Chevalier, B., Nakamura, T., Croce, C. M., Canaani, E., & Young, R. A. (2005). Global and Hox-specific roles for the MLL1 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(24), 8603–8608.
<https://doi.org/10.1073/pnas.0503072102>
- Gupta, S., Kim, S. Y., Artis, S., Molfese, D. L., Schumacher, A., Sweatt, J. D., Paylor, R. E., & Lubin, F. D. (2010). Histone methylation regulates memory formation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(10), 3589–3599.
<https://doi.org/10.1523/JNEUROSCI.3732-09.2010>
- Guy, J., Gan, J., Selfridge, J., Cobb, S., & Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science (New York, N.Y.)*, 315(5815), 1143–1147. <https://doi.org/10.1126/science.1138389>
- Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews. Molecular Cell Biology*, 19(10), 621–637. <https://doi.org/10.1038/s41580-018-0028-8>
- Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiology and Molecular Biology Reviews: MMBR*, 62(2), 465–503. <https://doi.org/10.1128/MMBR.62.2.465-503.1998>
- Hans, H., & Alwine, J. C. (2000). Functionally significant secondary structure of the simian virus 40 late polyadenylation signal. *Molecular and Cellular Biology*, 20(8), 2926–2932. <https://doi.org/10.1128/MCB.20.8.2926-2932.2000>
- Harris, J. R., Gao, C. W., Britton, J. F., Applegate, C. D., Bjornsson, H. T., & Fahrner, J. A. (2024). Five years of experience in the Epigenetics and Chromatin Clinic: What have we learned and where do we go from here? *Human Genetics*, 143(4), 607–624. <https://doi.org/10.1007/s00439-023-02537-1>

- Harris, J., Tomassy, G. S., & Arlotta, P. (2015). Building blocks of the cerebral cortex: From development to the dish. *Wiley Interdisciplinary Reviews. Developmental Biology*, 4(5), 529–544. <https://doi.org/10.1002/wdev.192>
- Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development (Cambridge, England)*, 129(8), 1983–1993. <https://doi.org/10.1242/dev.129.8.1983>
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39(3), 311–318. <https://doi.org/10.1038/ng1966>
- Herzeg, A., Borges, B., Lianoglou, B. R., Gonzalez-Velez, J., Canepa, E., Munar, D., Young, S. P., Bali, D., Gelb, M. H., Chakraborty, P., Kishnani, P. S., Harmatz, P., Cohen, J. L., & MacKenzie, T. C. (2023). Intrauterine enzyme replacement therapies for lysosomal storage disorders: Current developments and promising future prospects. *Prenatal Diagnosis*, 43(13), 1638–1649. <https://doi.org/10.1002/pd.6460>
- Holliday, R., & Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science (New York, N.Y.)*, 187(4173), 226–232.
- Hoskins, A. A., & Moore, M. J. (2012). The spliceosome: A flexible, reversible macromolecular machine. *Trends in Biochemical Sciences*, 37(5), 179–188. <https://doi.org/10.1016/j.tibs.2012.02.009>
- Hsieh, J. J.-D., Cheng, E. H.-Y., & Korsmeyer, S. J. (2003). Taspase1: A threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell*, 115(3), 293–303. [https://doi.org/10.1016/s0092-8674\(03\)00816-x](https://doi.org/10.1016/s0092-8674(03)00816-x)

- Huang, Y.-C., Shih, H.-Y., Lin, S.-J., Chiu, C.-C., Ma, T.-L., Yeh, T.-H., & Cheng, Y.-C. (2015). The epigenetic factor Kmt2a/Mll1 regulates neural progenitor proliferation and neuronal and glial differentiation. *Developmental Neurobiology*, 75(5), 452–462. <https://doi.org/10.1002/dneu.22235>
- Iacono, G., Dubos, A., Méziane, H., Benevento, M., Habibi, E., Mandoli, A., Riet, F., Selloum, M., Feil, R., Zhou, H., Kleefstra, T., Kasri, N. N., van Bokhoven, H., Herault, Y., & Stunnenberg, H. G. (2018). Increased H3K9 methylation and impaired expression of Protocadherins are associated with the cognitive dysfunctions of the Kleefstra syndrome. *Nucleic Acids Research*, 46(10), 4950–4965. <https://doi.org/10.1093/nar/gky196>
- Inano, K., Suetake, I., Ueda, T., Miyake, Y., Nakamura, M., Okada, M., & Tajima, S. (2000). Maintenance-type DNA methyltransferase is highly expressed in post-mitotic neurons and localized in the cytoplasmic compartment. *Journal of Biochemistry*, 128(2), 315–321. <https://doi.org/10.1093/oxfordjournals.jbchem.a022755>
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931–945. <https://doi.org/10.1038/nature03001>
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)*, 333(6047), 1300–1303. <https://doi.org/10.1126/science.1210597>
- Jäkel, S., & Dimou, L. (2017). Glial Cells and Their Function in the Adult Brain: A Journey through the History of Their Ablation. *Frontiers in Cellular Neuroscience*, 11. <https://doi.org/10.3389/fncel.2017.00024>

- Jakovcevski, M., Ruan, H., Shen, E. Y., Dincer, A., Javidfar, B., Ma, Q., Peter, C. J., Cheung, I., Mitchell, A. C., Jiang, Y., Lin, C. L., Pothula, V., Stewart, A. F., Ernst, P., Yao, W.-D., & Akbarian, S. (2015). Neuronal Kmt2a/Mll1 histone methyltransferase is essential for prefrontal synaptic plasticity and working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(13), 5097–5108.
<https://doi.org/10.1523/JNEUROSCI.3004-14.2015>
- Jiang, W., & Chen, L. (2021). Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Computational and Structural Biotechnology Journal*, 19, 183–195.
<https://doi.org/10.1016/j.csbj.2020.12.009>
- Jiao, Z., Zhang, Z. G., Hornyak, T. J., Hozeska, A., Zhang, R. L., Wang, Y., Wang, L., Roberts, C., Strickland, F. M., & Chopp, M. (2006). Dopachrome tautomerase (Dct) regulates neural progenitor cell proliferation. *Developmental Biology*, 296(2), 396–408. <https://doi.org/10.1016/j.ydbio.2006.06.006>
- Johanson, J. F., Sonnenberg, A., Koch, T. R., & McCarty, D. J. (1992). Association of constipation with neurologic diseases. *Digestive Diseases and Sciences*, 37(2), 179–186. <https://doi.org/10.1007/BF01308169>
- Jones, W. D., Dafou, D., McEntagart, M., Woollard, W. J., Elmslie, F. V., Holder-Espinasse, M., Irving, M., Sagar, A. K., Smithson, S., Trembath, R. C., Deshpande, C., & Simpson, M. A. (2012). De novo mutations in MLL cause Wiedemann-Steiner syndrome. *American Journal of Human Genetics*, 91(2), 358–364. <https://doi.org/10.1016/j.ajhg.2012.06.008>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein

- structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.
<https://doi.org/10.1038/s41586-021-03819-2>
- Jung, Y. L., Hung, C., Choi, J., Lee, E. A., & Bodamer, O. (2023). Characterizing the molecular impact of KMT2D variants on the epigenetic and transcriptional landscapes in Kabuki syndrome. *Human Molecular Genetics*, 32(13), 2251–2261. <https://doi.org/10.1093/hmg/ddad059>
- Jurica, M. S., & Moore, M. J. (2003). Pre-mRNA splicing: Awash in a sea of proteins. *Molecular Cell*, 12(1), 5–14. [https://doi.org/10.1016/s1097-2765\(03\)00270-3](https://doi.org/10.1016/s1097-2765(03)00270-3)
- Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J., & Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314), 430–435.
<https://doi.org/10.1038/nature09380>
- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J., & Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471), 333–339.
<https://doi.org/10.1038/nature12634>
- Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., & Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, 429(6994), 900–903.
<https://doi.org/10.1038/nature02633>

- Karnuta, J. M., & Scacheri, P. C. (2018). Enhancers: Bridging the gap between gene control and human disease. *Human Molecular Genetics*, 27(R2), R219–R227. <https://doi.org/10.1093/hmg/ddy167>
- Keating, D., Marigowda, G., Burr, L., Daines, C., Mall, M. A., McKone, E. F., Ramsey, B. W., Rowe, S. M., Sass, L. A., Tullis, E., McKee, C. M., Moskowitz, S. M., Robertson, S., Savage, J., Simard, C., Van Goor, F., Waltz, D., Xuan, F., Young, T., ... VX16-445-001 Study Group. (2018). VX-445-Tezacaftor-Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles. *The New England Journal of Medicine*, 379(17), 1612–1620. <https://doi.org/10.1056/NEJMoa1807120>
- Kempermann, G., & Gage, F. H. (1999). New Nerve Cells for the Adult Brain. *Scientific American*, 280(5), 48–53.
- Kerimoglu, C., Sakib, M. S., Jain, G., Benito, E., Burkhardt, S., Capece, V., Kaurani, L., Halder, R., Agís-Balboa, R. C., Stilling, R., Urbanke, H., Kranz, A., Stewart, A. F., & Fischer, A. (2017). KMT2A and KMT2B Mediate Memory Function by Affecting Distinct Genomic Regions. *Cell Reports*, 20(3), 538–548. <https://doi.org/10.1016/j.celrep.2017.06.072>
- Kilgore, M., Miller, C. A., Fass, D. M., Hennig, K. M., Haggarty, S. J., Sweatt, J. D., & Rumbaugh, G. (2010). Inhibitors of class 1 histone deacetylases reverse contextual memory deficits in a mouse model of Alzheimer's disease. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 35(4), 870–880. <https://doi.org/10.1038/npp.2009.197>
- Kim, H., Kim, M., Im, S.-K., & Fang, S. (2018). Mouse Cre-LoxP system: General principles to determine tissue-specific roles of target genes. *Laboratory Animal Research*, 34(4), 147–159. <https://doi.org/10.5625/lar.2018.34.4.147>

- Kim, K.-R., Jeong, H.-J., Kim, Y., Lee, S. Y., Kim, Y., Kim, H.-J., Lee, S.-H., Cho, H., Kang, J.-S., & Ho, W.-K. (2021). Calbindin regulates Kv4.1 trafficking and excitability in dentate granule cells via CaMKII-dependent phosphorylation. *Experimental & Molecular Medicine*, 53(7), 1134–1147.
<https://doi.org/10.1038/s12276-021-00645-4>
- Kleff, S., Andrulis, E. D., Anderson, C. W., & Sternglanz, R. (1995). Identification of a gene encoding a yeast histone H4 acetyltransferase. *The Journal of Biological Chemistry*, 270(42), 24674–24677.
<https://doi.org/10.1074/jbc.270.42.24674>
- Klein, C., & Struhl, K. (1994). Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains in vivo. *Science (New York, N.Y.)*, 266(5183), 280–282. <https://doi.org/10.1126/science.7939664>
- Koo, T., Lee, J., & Kim, J.-S. (2015). Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. *Molecules and Cells*, 38(6), 475–481. <https://doi.org/10.14348/molcells.2015.0103>
- Korzus, E., Rosenfeld, M. G., & Mayford, M. (2004). CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron*, 42(6), 961–972. <https://doi.org/10.1016/j.neuron.2004.06.002>
- Kozareva, D. A., Cryan, J. F., & Nolan, Y. M. (2019). Born this way: Hippocampal neurogenesis across the lifespan. *Aging Cell*, 18(5), e13007.
<https://doi.org/10.1111/accel.13007>
- Le Douarin, N., & Kalcheim, C. (1999). *The Neural Crest* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511897948>
- Lee, J.-E., Wang, C., Xu, S., Cho, Y.-W., Wang, L., Feng, X., Baldrige, A., Sartorelli, V., Zhuang, L., Peng, W., & Ge, K. (2013). H3K4 mono- and di-

- methytransferase MLL4 is required for enhancer activation during cell differentiation. *eLife*, 2, e01503. <https://doi.org/10.7554/eLife.01503>
- Li, L., Fukunaga-Kalabis, M., Yu, H., Xu, X., Kong, J., Lee, J. T., & Herlyn, M. (2010). Human dermal stem cells differentiate into functional epidermal melanocytes. *Journal of Cell Science*, 123(Pt 6), 853–860. <https://doi.org/10.1242/jcs.061598>
- Li, L., Medina-Menéndez, C., García-Corzo, L., Córdoba-Beldad, C. M., Quiroga, A. C., Calleja Barca, E., Zinchuk, V., Muñoz-López, S., Rodríguez-Martín, P., Ciorraga, M., Colmena, I., Fernández, S., Vicario, C., Nicolis, S. K., Lefebvre, V., Mira, H., & Morales, A. V. (2022). SoxD genes are required for adult neural stem cell activation. *Cell Reports*, 38(5), 110313. <https://doi.org/10.1016/j.celrep.2022.110313>
- Li, Q. (2020). Nusinersen as a Therapeutic Agent for Spinal Muscular Atrophy. *Yonsei Medical Journal*, 61(4), 273–283. <https://doi.org/10.3349/ymj.2020.61.4.273>
- Li, R., Dong, Q., Yuan, X., Zeng, X., Gao, Y., Chiao, C., Li, H., Zhao, X., Keles, S., Wang, Z., & Chang, Q. (2016). Misregulation of Alternative Splicing in a Mouse Model of Rett Syndrome. *PLoS Genetics*, 12(6), e1006129. <https://doi.org/10.1371/journal.pgen.1006129>
- Li, Y. J., Li, C.-Y., Li, C.-Y., Hu, D.-X., Xv, Z.-B., Zhang, S.-H., Li, Q., Zhang, P., Tian, B., Lan, X.-L., & Chen, X.-Q. (2023). KMT2E Haploinsufficiency Manifests Autism-Like Behaviors and Amygdala Neuronal Development Dysfunction in Mice. *Molecular Neurobiology*, 60(3), 1609–1625. <https://doi.org/10.1007/s12035-022-03167-w>

- Liang, H., Hippenmeyer, S., & Ghashghaei, H. T. (2012). A Nestin-cre transgenic mouse is insufficient for recombination in early embryonic neural progenitors. *Biology Open*, 1(12), 1200–1203. <https://doi.org/10.1242/bio.20122287>
- Lim, D. A., Huang, Y.-C., Swigut, T., Mirick, A. L., Garcia-Verdugo, J. M., Wysocka, J., Ernst, P., & Alvarez-Buylla, A. (2009). Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. *Nature*, 458(7237), 529–533. <https://doi.org/10.1038/nature07726>
- Lotfollahzadeh, S., Taherian, M., & Anand, S. (2024). Hirschsprung Disease. In *StatPearls*. StatPearls Publishing.
<http://www.ncbi.nlm.nih.gov/books/NBK562142/>
- Luco, R. F., & Misteli, T. (2011). More than a splicing code: Integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. *Current Opinion in Genetics & Development*, 21(4), 366–372.
<https://doi.org/10.1016/j.gde.2011.03.004>
- Lumsden, A., & Krumlauf, R. (1996). Patterning the vertebrate neuraxis. *Science (New York, N.Y.)*, 274(5290), 1109–1115.
<https://doi.org/10.1126/science.274.5290.1109>
- Luperchio, T. R., Boukas, L., Zhang, L., Pilarowski, G., Jiang, J., Kalinousky, A., Hansen, K. D., & Bjornsson, H. T. (2021). Leveraging the Mendelian disorders of the epigenetic machinery to systematically map functional epigenetic variation. *eLife*, 10, e65884. <https://doi.org/10.7554/eLife.65884>
- Ma, X., Hung, C., & Bodamer, O. (2023). P239: Small molecule, high throughput drug discovery for Kabuki syndrome*. *Genetics in Medicine Open*, 1(1), 100267.
<https://doi.org/10.1016/j.gimo.2023.100267>
- Mackenzie, M. A., Jordan, S. A., Budd, P. S., & Jackson, I. J. (1997). Activation of the receptor tyrosine kinase Kit is required for the proliferation of melanoblasts in

the mouse embryo. *Developmental Biology*, 192(1), 99–107.

<https://doi.org/10.1006/dbio.1997.8738>

Madisen, L., Zwingman, T. A., Sunkin, S. M., Oh, S. W., Zariwala, H. A., Gu, H., Ng, L. L., Palmiter, R. D., Hawrylycz, M. J., Jones, A. R., Lein, E. S., & Zeng, H. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neuroscience*, 13(1), 133–140.
<https://doi.org/10.1038/nn.2467>

Massaro, G., Mattar, C. N. Z., Wong, A. M. S., Sirka, E., Buckley, S. M. K., Herbert, B. R., Karlsson, S., Perocheau, D. P., Burke, D., Heales, S., Richard-Londt, A., Brandner, S., Huebecker, M., Priestman, D. A., Platt, F. M., Mills, K., Biswas, A., Cooper, J. D., Chan, J. K. Y., ... Rahim, A. A. (2018). Fetal gene therapy for neurodegenerative disease of infants. *Nature Medicine*, 24(9), 1317–1323.
<https://doi.org/10.1038/s41591-018-0106-7>

Mas-y-Mas, S., Barbon, M., Teyssier, C., Déméné, H., Carvalho, J. E., Bird, L. E., Lebedev, A., Fattori, J., Schubert, M., Dumas, C., Bourguet, W., & Le Maire, A. (2016). The Human Mixed Lineage Leukemia 5 (MLL5), a Sequentially and Structurally Divergent SET Domain-Containing Protein with No Intrinsic Catalytic Activity. *PLOS ONE*, 11(11), e0165139.
<https://doi.org/10.1371/journal.pone.0165139>

Mattar, C. N., Waddington, S. N., Biswas, A., Johana, N., Ng, X. W., Fisk, A. S., Fisk, N. M., Tan, L. G., Rahim, A. A., Buckley, S. M. K., Tan, M. H., Lu, J., Choolani, M., & Chan, J. K. Y. (2013). Systemic delivery of scAAV9 in fetal macaques facilitates neuronal transduction of the central and peripheral nervous systems. *Gene Therapy*, 20(1), 69–83.
<https://doi.org/10.1038/gt.2011.216>

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., & Bird, A. P. (1989).

Identification of a mammalian protein that binds specifically to DNA containing

methylyated CpGs. *Cell*, 58(3), 499–507. [https://doi.org/10.1016/0092-8674\(89\)90430-3](https://doi.org/10.1016/0092-8674(89)90430-3)

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), 766–770.
<https://doi.org/10.1038/nature07107>

Méndez-Maldonado, K., Vega-López, G. A., Aybar, M. J., & Velasco, I. (2020). Neurogenesis From Neural Crest Cells: Molecular Mechanisms in the Formation of Cranial Nerves and Ganglia. *Frontiers in Cell and Developmental Biology*, 8. <https://doi.org/10.3389/fcell.2020.00635>

Mielcarek, M., Benn, C. L., Franklin, S. A., Smith, D. L., Woodman, B., Marks, P. A., & Bates, G. P. (2011). SAHA decreases HDAC 2 and 4 levels in vivo and improves molecular phenotypes in the R6/2 mouse model of Huntington's disease. *PloS One*, 6(11), e27746.
<https://doi.org/10.1371/journal.pone.0027746>

Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D., & Hess, J. L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. *Molecular Cell*, 10(5), 1107–1117.
[https://doi.org/10.1016/s1097-2765\(02\)00741-4](https://doi.org/10.1016/s1097-2765(02)00741-4)

Mirro, J., Kitchingman, G., Williams, D., Lauzon, G. J., Lin, C. C., Callihan, T., & Zipf, T. F. (1986). Clinical and laboratory characteristics of acute leukemia with the 4;11 translocation. *Blood*, 67(3), 689–697.

Miyake, N., Mizuno, S., Okamoto, N., Ohashi, H., Shiina, M., Ogata, K., Tsurusaki, Y., Nakashima, M., Saitsu, H., Niikawa, N., & Matsumoto, N. (2013). KDM6A

- point mutations cause Kabuki syndrome. *Human Mutation*, 34(1), 108–110.
<https://doi.org/10.1002/humu.22229>
- Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M., & Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular Cell*, 30(6), 755–766. <https://doi.org/10.1016/j.molcel.2008.05.007>
- Moore, L. D., Le, T., & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38(1), 23–38.
<https://doi.org/10.1038/npp.2012.112>
- Morshedzadeh, F., Ghanei, M., Lotfi, M., Ghasemi, M., Ahmadi, M., Najari-Hanjani, P., Sharif, S., Mozaffari-Jovin, S., Peymani, M., & Abbaszadegan, M. R. (2023). An Update on the Application of CRISPR Technology in Clinical Practice. *Molecular Biotechnology*, 1–19. <https://doi.org/10.1007/s12033-023-00724-z>
- Naegelin, Y., Kuhle, J., Schädelin, S., Datta, A. N., Magon, S., Amann, M., Barro, C., Ramelli, G. P., Heesom, K., Barde, Y.-A., Weber, P., & Kappos, L. (2021). Fingolimod in children with Rett syndrome: The FINGORETT study. *Orphanet Journal of Rare Diseases*, 16(1), 19. <https://doi.org/10.1186/s13023-020-01655-7>
- Nakamura, M., Schneider, M. R., Schmidt-Ullrich, R., & Paus, R. (2013). Mutant laboratory mice with abnormalities in hair follicle morphogenesis, cycling, and/or structure: An update. *Journal of Dermatological Science*, 69(1), 6–29.
<https://doi.org/10.1016/j.jdermsci.2012.10.001>
- Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., & Canaani, E. (2002). ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in

- transcriptional regulation. *Molecular Cell*, 10(5), 1119–1128.
[https://doi.org/10.1016/s1097-2765\(02\)00740-2](https://doi.org/10.1016/s1097-2765(02)00740-2)
- Nakamura, T., Yoshihara, T., Tanegashima, C., Kadota, M., Kobayashi, Y., Honda, K., Ishiwata, M., Ueda, J., Hara, T., Nakanishi, M., Takumi, T., Itohara, S., Kuraku, S., Asano, M., Kasahara, T., Nakajima, K., Tsuboi, T., Takata, A., & Kato, T. (2024). Transcriptomic dysregulation and autistic-like behaviors in *Kmt2c* haploinsufficient mice rescued by an LSD1 inhibitor. *Molecular Psychiatry*.
<https://doi.org/10.1038/s41380-024-02479-8>
- Ng, R., Kalinousky, A., Fahrner, J. A., Bjornsson, H. T., & Harris, J. (2023). The social phenotype associated with Wiedemann-Steiner-syndrome: Autistic traits juxtaposed with high social drive and prosociality. *American Journal of Medical Genetics Part A*, 191(10), 2591–2601. <https://doi.org/10.1002/ajmg.a.63351>
- Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., Lee, C., Turner, E. H., Smith, J. D., Rieder, M. J., Yoshiura, K.-I., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D. A., ... Shendure, J. (2010). Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nature Genetics*, 42(9), 790–793. <https://doi.org/10.1038/ng.646>
- Nguyen, S., Meletis, K., Fu, D., Jhaveri, S., & Jaenisch, R. (2007). Ablation of de novo DNA methyltransferase *Dnmt3a* in the nervous system leads to neuromuscular defects and shortened lifespan. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 236(6), 1663–1676.
<https://doi.org/10.1002/dvdy.21176>
- Ni, J., Ye, D., Zeng, W., Ma, S., Wang, Z., Kuang, Y., & Yang, L. (2024). Promotion of hair growth by a conditioned medium from human umbilical cord mesenchymal stem cells cultivated in a 3D scaffold of gelatin sponge.

European Journal of Medical Research, 29(1), 270.

<https://doi.org/10.1186/s40001-024-01830-7>

Nikolov, D. B., & Burley, S. K. (1994). 2.1 A resolution refined structure of a TATA box-binding protein (TBP). *Nature Structural Biology*, 1(9), 621–637.

<https://doi.org/10.1038/nsb0994-621>

Norris, A. D., & Calarco, J. A. (2012). Emerging Roles of Alternative Pre-mRNA Splicing Regulation in Neuronal Development and Function. *Frontiers in Neuroscience*, 6, 122. <https://doi.org/10.3389/fnins.2012.00122>

Odelin, G., Faucherre, A., Marchese, D., Pinard, A., Jaouadi, H., Le Scouarnec, S., FranceGenRef Consortium, Deleuze, J.-F., Génin, E., Lindenbaum, P., Redon, R., Schott, J.-J., Chiarelli, R., Achouri, Y., Faure, E., Herbane, M., Théron, A., Avierinos, J.-F., Jopling, C., ... Zaffran, S. (2023). Variations in the poly-histidine repeat motif of HOXA1 contribute to bicuspid aortic valve in mouse and zebrafish. *Nature Communications*, 14(1), 1543.

<https://doi.org/10.1038/s41467-023-37110-x>

Ohnishi, T., Miura, I., Ohba, H., Shimamoto, C., Iwayama, Y., Wakana, S., & Yoshikawa, T. (2017). A spontaneous and novel Pax3 mutant mouse that models Waardenburg syndrome and neural tube defects. *Gene*, 607, 16–22.

<https://doi.org/10.1016/j.gene.2016.12.037>

Ohno, T., Maegawa, T., Katoh, H., Miyasaka, Y., Suzuki, M., Kobayashi, M., & Horio, F. (2017). A new missense mutation in the paired domain of the mouse Pax3 gene. *Experimental Animals*, 66(3), 245–250.

<https://doi.org/10.1538/expanim.17-0013>

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian

- development. *Cell*, 99(3), 247–257. [https://doi.org/10.1016/s0092-8674\(00\)81656-6](https://doi.org/10.1016/s0092-8674(00)81656-6)
- Olins, A. L., & Olins, D. E. (1974). Spheroid chromatin units (v bodies). *Science (New York, N.Y.)*, 183(4122), 330–332.
<https://doi.org/10.1126/science.183.4122.330>
- Pakkenberg, B., & Gundersen, H. J. (1997). Neocortical neuron number in humans: Effect of sex and age. *The Journal of Comparative Neurology*, 384(2), 312–320.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40(12), 1413–1415.
<https://doi.org/10.1038/ng.259>
- Patel, A., Dharmarajan, V., & Cosgrove, M. S. (2008). Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *The Journal of Biological Chemistry*, 283(47), 32158–32161. <https://doi.org/10.1074/jbc.C800164200>
- Pelisch, F., Blaustein, M., Kornblihtt, A. R., & Srebrow, A. (2005). Cross-talk between Signaling Pathways Regulates Alternative Splicing. *Journal of Biological Chemistry*, 280(27), 25461–25469.
<https://doi.org/10.1074/jbc.M412007200>
- Pitcovski, J., Shahar, E., Aizenshtein, E., & Gorodetsky, R. (2017). Melanoma antigens and related immunological markers. *Critical Reviews in Oncology/Hematology*, 115, 36–49. <https://doi.org/10.1016/j.critrevonc.2017.05.001>
- Pleiss, J. A., Whitworth, G. B., Bergkessel, M., & Guthrie, C. (2007a). Rapid, transcript-specific changes in splicing in response to environmental stress. *Molecular Cell*, 27(6), 928–937.
<https://doi.org/10.1016/j.molcel.2007.07.018>

- Pleiss, J. A., Whitworth, G. B., Bergkessel, M., & Guthrie, C. (2007b). Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biology*, 5(4), e90. <https://doi.org/10.1371/journal.pbio.0050090>
- Poreba, E., Lesniewicz, K., & Durzynska, J. (2020). Aberrant Activity of Histone-Lysine N-Methyltransferase 2 (KMT2) Complexes in Oncogenesis. *International Journal of Molecular Sciences*, 21(24), 9340. <https://doi.org/10.3390/ijms21249340>
- Pradhan, S., Bacolla, A., Wells, R. D., & Roberts, R. J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *The Journal of Biological Chemistry*, 274(46), 33002–33010. <https://doi.org/10.1074/jbc.274.46.33002>
- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2001). Neuroglial Cells. In *Neuroscience. 2nd edition*. Sinauer Associates. <https://www.ncbi.nlm.nih.gov/books/NBK10869/>
- R Core Team. (2021). *R: A language and environment for statistical computing*. [Computer software]. <https://www.R-project.org/>
- Rao, R. C., & Dou, Y. (2015). Hijacked in cancer: The KMT2 (MLL) family of methyltransferases. *Nature Reviews. Cancer*, 15(6), 334–346. <https://doi.org/10.1038/nrc3929>
- RARE Disease Facts. (2024). *Global Genes*. <https://globalgenes.org/rare-disease-facts/>
- Redman, M., King, A., Watson, C., & King, D. (2016). What is CRISPR/Cas9? *Archives of Disease in Childhood. Education and Practice Edition*, 101(4), 213–215. <https://doi.org/10.1136/archdischild-2016-310459>

- Reynisdottir, T., Anderson, K. J., Boukas, L., & Bjornsson, H. T. (2022). Missense variants causing Wiedemann-Steiner syndrome preferentially occur in the KMT2A-CXXC domain and are accurately classified using AlphaFold2. *PLOS Genetics*, 18(6), e1010278. <https://doi.org/10.1371/journal.pgen.1010278>
- Reynisdottir, T., Anderson, K. J., Brinn, A., Franklin, K., Ouyang, J., Snorradottir, A. O., Lutz, C. M., Zuberi, A. R., DeLeon, V. B., & Bjornsson, H. T. (2024). *In-utero rescue of neurological dysfunction in a mouse model of Wiedemann-Steiner syndrome*. <https://doi.org/10.1101/2024.07.19.604339>
- Sadri-Vakili, G., & Cha, J.-H. J. (2006). Mechanisms of disease: Histone modifications in Huntington's disease. *Nature Clinical Practice. Neurology*, 2(6), 330–338. <https://doi.org/10.1038/ncpneuro0199>
- Sahay, A., Scobie, K. N., Hill, A. S., O'Carroll, C. M., Kheirbek, M. A., Burghardt, N. S., Fenton, A. A., Dranovsky, A., & Hen, R. (2011). Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature*, 472(7344), 466–470. <https://doi.org/10.1038/nature09817>
- Sanes, D. H., Reh, T. A., Harris, W. A., & Landgraf, M. (2019). *Development of the Nervous System*. Academic Press.
- Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), 1412–1417. <https://doi.org/10.1073/pnas.0510310103>
- Schanze, I., Schanze, D., Bacino, C. A., Douzgou, S., Kerr, B., & Zenker, M. (2013). Haploinsufficiency of SOX5, a member of the SOX (SRY-related HMG-box) family of transcription factors is a cause of intellectual disability. *European*

Journal of Medical Genetics, 56(2), 108–113.

<https://doi.org/10.1016/j.ejmg.2012.11.001>

Scharner, J., & Aznarez, I. (2021). Clinical Applications of Single-Stranded Oligonucleotides: Current Landscape of Approved and In-Development Therapeutics. *Molecular Therapy*, 29(2), 540–554.

<https://doi.org/10.1016/j.ymthe.2020.12.022>

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–682. <https://doi.org/10.1038/nmeth.2019>

Schloesser, R. J., Lehmann, M., Martinowich, K., Manji, H. K., & Herkenham, M. (2010). Environmental enrichment requires adult neurogenesis to facilitate the recovery from psychosocial stress. *Molecular Psychiatry*, 15(12), 1152–1163. <https://doi.org/10.1038/mp.2010.34>

Schuettengruber, B., Martinez, A.-M., Iovino, N., & Cavalli, G. (2011). Trithorax group proteins: Switching genes on and keeping them active. *Nature Reviews. Molecular Cell Biology*, 12(12), 799–814. <https://doi.org/10.1038/nrm3230>

Schwarz, Q., Maden, C. H., Vieira, J. M., & Ruhrberg, C. (2009). Neuropilin 1 signaling guides neural crest cells to coordinate pathway choice with cell specification. *Proceedings of the National Academy of Sciences of the United States of America*, 106(15), 6164–6169. <https://doi.org/10.1073/pnas.0811521106>

Schwenty-Lara, J., Nehl, D., & Borchers, A. (2020). The histone methyltransferase KMT2D, mutated in Kabuki syndrome patients, is required for neural crest cell

- formation and migration. *Human Molecular Genetics*, 29(2), 305–319.
<https://doi.org/10.1093/hmg/ddz284>
- Shah, S., & Richter, J. D. (2021). Do Fragile X Syndrome and Other Intellectual Disorders Converge at Aberrant Pre-mRNA Splicing? *Frontiers in Psychiatry*, 12, 715346. <https://doi.org/10.3389/fpsy.2021.715346>
- Sharma, V. K., Mehta, V., & Singh, T. G. (2020). Alzheimer's Disorder: Epigenetic Connection and Associated Risk Factors. *Current Neuropharmacology*, 18(8), 740–753. <https://doi.org/10.2174/1570159X18666200128125641>
- Sharp, A. J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S. B., Dupre, Y., & Antonarakis, S. E. (2011). DNA methylation profiles of human active and inactive X chromosomes. *Genome Research*, 21(10), 1592–1600.
<https://doi.org/10.1101/gr.112680.110>
- Sheppard, S. E., Campbell, I. M., Harr, M. H., Gold, N., Li, D., Bjornsson, H. T., Cohen, J. S., Fahrner, J. A., Fatemi, A., Harris, J. R., Nowak, C., Stevens, C. A., Grand, K., Au, M., Graham, J. M., Sanchez-Lara, P. A., Campo, M. D., Jones, M. C., Abdul-Rahman, O., ... Quintero-Rivera, F. (2021). Expanding the genotypic and phenotypic spectrum in a diverse cohort of 104 individuals with Wiedemann-Steiner syndrome. *American Journal of Medical Genetics. Part A*, 185(6), 1649–1665. <https://doi.org/10.1002/ajmg.a.62124>
- Sheppard, S. E., & Quintero-Rivera, F. (2022). Wiedemann-Steiner Syndrome. In M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, & A. Amemiya (Eds.), *GeneReviews®*. University of Washington, Seattle. <http://www.ncbi.nlm.nih.gov/books/NBK580718/>
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstone, J. R., Cole, P. A., Casero, R. A., & Shi, Y. (2004). Histone demethylation mediated by the nuclear amine

oxidase homolog LSD1. *Cell*, 119(7), 941–953.

<https://doi.org/10.1016/j.cell.2004.12.012>

Shpargel, K. B., Mangini, C. L., Xie, G., Ge, K., & Magnuson, T. (2020). The KMT2D Kabuki syndrome histone methylase controls neural crest cell differentiation and facial morphology. *Development (Cambridge, England)*, 147(21), dev187997. <https://doi.org/10.1242/dev.187997>

Shpargel, K. B., Starmer, J., Wang, C., Ge, K., & Magnuson, T. (2017). UTX-guided neural crest function underlies craniofacial features of Kabuki syndrome. *Proceedings of the National Academy of Sciences*, 114(43), E9046–E9055. <https://doi.org/10.1073/pnas.1705011114>

Smale, S. T., & Baltimore, D. (1989). The “initiator” as a transcription control element. *Cell*, 57(1), 103–113. [https://doi.org/10.1016/0092-8674\(89\)90176-1](https://doi.org/10.1016/0092-8674(89)90176-1)

Stamova, B. S., Tian, Y., Nordahl, C. W., Shen, M. D., Rogers, S., Amaral, D. G., & Sharp, F. R. (2013). Evidence for differential alternative splicing in blood of young boys with autism spectrum disorders. *Molecular Autism*, 4(1), 30. <https://doi.org/10.1186/2040-2392-4-30>

Steiner, C. E., & Marques, A. P. (2000). Growth deficiency, mental retardation and unusual facies. *Clinical Dysmorphology*, 9(2), 155–156. <https://doi.org/10.1097/00019605-200009020-00021>

Sternberg, N., & Hamilton, D. (1981). Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *Journal of Molecular Biology*, 150(4), 467–486. [https://doi.org/10.1016/0022-2836\(81\)90375-2](https://doi.org/10.1016/0022-2836(81)90375-2)

Stiles, J., & Jernigan, T. L. (2010). The basics of brain development. *Neuropsychology Review*, 20(4), 327–348. <https://doi.org/10.1007/s11065-010-9148-4>

Strande, N. T., & Berg, J. S. (2016). Defining the Clinical Value of a Genomic Diagnosis in the Era of Next-Generation Sequencing. *Annual Review of*

- Genomics and Human Genetics*, 17, 303–332.
<https://doi.org/10.1146/annurev-genom-083115-022348>
- Stuart, S. A., & Robinson, E. S. J. (2015). Reducing the stress of drug administration: Implications for the 3Rs. *Scientific Reports*, 5(1), 14288.
<https://doi.org/10.1038/srep14288>
- Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., & Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *The Journal of Biological Chemistry*, 279(26), 27816–27823.
<https://doi.org/10.1074/jbc.M400181200>
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)*, 324(5929), 930–935.
<https://doi.org/10.1126/science.1170116>
- Tambuyzer, E., Vandendriessche, B., Austin, C. P., Brooks, P. J., Larsson, K., Miller Needleman, K. I., Valentine, J., Davies, K., Groft, S. C., Preti, R., Oprea, T. I., & Prunotto, M. (2020). Therapies for rare diseases: Therapeutic modalities, progress and challenges ahead. *Nature Reviews Drug Discovery*, 19(2), 93–111. <https://doi.org/10.1038/s41573-019-0049-9>
- Taunton, J., Hassig, C. A., & Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science (New York, N.Y.)*, 272(5260), 408–411. <https://doi.org/10.1126/science.272.5260.408>
- Tazi, J., & Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell*, 60(6), 909–920. [https://doi.org/10.1016/0092-8674\(90\)90339-g](https://doi.org/10.1016/0092-8674(90)90339-g)
- Teleanu, R. I., Niculescu, A.-G., Roza, E., Vladăcenco, O., Grumezescu, A. M., & Teleanu, D. M. (2022). Neurotransmitters-Key Factors in Neurological and

- Neurodegenerative Disorders of the Central Nervous System. *International Journal of Molecular Sciences*, 23(11), 5954.
<https://doi.org/10.3390/ijms23115954>
- Terranova, R., Agherbi, H., Boned, A., Meresse, S., & Djabali, M. (2006). Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proceedings of the National Academy of Sciences of the United States of America*, 103(17), 6629–6634.
<https://doi.org/10.1073/pnas.0507425103>
- Thacker, S., Sefyi, M., & Eng, C. (2020). Alternative splicing landscape of the neural transcriptome in a cytoplasmic-predominant Pten expression murine model of autism-like Behavior. *Translational Psychiatry*, 10(1), 380.
<https://doi.org/10.1038/s41398-020-01068-x>
- Thiery, J. P., Acloque, H., Huang, R. Y. J., & Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell*, 139(5), 871–890.
<https://doi.org/10.1016/j.cell.2009.11.007>
- Tkachuk, D. C., Kohler, S., & Cleary, M. L. (1992). Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell*, 71(4), 691–700. [https://doi.org/10.1016/0092-8674\(92\)90602-9](https://doi.org/10.1016/0092-8674(92)90602-9)
- Tokuo, H., Bhawan, J., & Coluccio, L. M. (2018). Myosin X is required for efficient melanoblast migration and melanoma initiation and metastasis. *Scientific Reports*, 8(1), 10449. <https://doi.org/10.1038/s41598-018-28717-y>
- Tronche, F., Kellendonk, C., Kretz, O., Gass, P., Anlag, K., Orban, P. C., Bock, R., Klein, R., & Schütz, G. (1999). Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. *Nature Genetics*, 23(1), 99–103. <https://doi.org/10.1038/12703>

- Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., & Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, 439(7078), 811–816.
<https://doi.org/10.1038/nature04433>
- Tudor, M., Akbarian, S., Chen, R. Z., & Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), 15536–15541. <https://doi.org/10.1073/pnas.242566899>
- Vallianatos, C. N., Raines, B., Porter, R. S., Bonefas, K. M., Wu, M. C., Garay, P. M., Collette, K. M., Seo, Y. A., Dou, Y., Keegan, C. E., Tronson, N. C., & Iwase, S. (2020). Mutually suppressive roles of KMT2A and KDM5C in behaviour, neuronal structure, and histone H3K4 methylation. *Communications Biology*, 3(1), 278. <https://doi.org/10.1038/s42003-020-1001-6>
- Vandamme, N., & Berx, G. (2019). From neural crest cells to melanocytes: Cellular plasticity during development and beyond. *Cellular and Molecular Life Sciences*, 76(10), 1919–1934. <https://doi.org/10.1007/s00018-019-03049-w>
- Ventura, A., Kirsch, D. G., McLaughlin, M. E., Tuveson, D. A., Grimm, J., Lintault, L., Newman, J., Reczek, E. E., Weissleder, R., & Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. *Nature*, 445(7128), 661–665. <https://doi.org/10.1038/nature05541>
- Waddington, C. H. (1942). The epigenotype. *International Journal of Epidemiology*, 41(1), 10–13. <https://doi.org/10.1093/ije/dyr184>
- Wang, P., Lin, C., Smith, E. R., Guo, H., Sanderson, B. W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L. M., Ge, K., Krumlauf, R., & Shilatifard, A. (2009). Global Analysis of H3K4 Methylation Defines MLL Family Member Targets and Points to a Role for MLL1-Mediated H3K4

- Methylation in the Regulation of Transcriptional Initiation by RNA Polymerase II. *Molecular and Cellular Biology*, 29(22), 6074–6085.
<https://doi.org/10.1128/MCB.00924-09>
- Wang, Y., Liu, J., Huang, B. O., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M., & Wang, X.-Z. (2015). Mechanism of alternative splicing and its regulation. *Biomedical Reports*, 3(2), 152–158.
<https://doi.org/10.3892/br.2014.407>
- Wang, Z., Song, J., Milne, T. A., Wang, G. G., Li, H., Allis, C. D., & Patel, D. J. (2010). Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell*, 141(7), 1183–1194.
<https://doi.org/10.1016/j.cell.2010.05.016>
- Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, 39(4), 457–466. <https://doi.org/10.1038/ng1990>
- Weber, T. (2009). Inducible gene manipulations in serotonergic neurons. *Frontiers in Molecular Neuroscience*, 2. <https://doi.org/10.3389/neuro.02.024.2009>
- Weiss, S. B., & Gladstone, L. (1959). A MAMMALIAN SYSTEM FOR THE INCORPORATION OF CYTIDINE TRIPHOSPHATE INTO RIBONUCLEIC ACID¹. *Journal of the American Chemical Society*, 81(15), 4118–4119.
<https://doi.org/10.1021/ja01524a087>
- Wendelin, D. S., Pope, D. N., & Mallory, S. B. (2003). Hypertrichosis. *Journal of the American Academy of Dermatology*, 48(2), 161–179; quiz 180–181.
<https://doi.org/10.1067/mjd.2003.100>

- Whitfield, J., Littlewood, T., & Soucek, L. (2015). Tamoxifen administration to mice. *Cold Spring Harbor Protocols*, 2015(3), 269–271.
<https://doi.org/10.1101/pdb.prot077966>
- Wiedemann, H.-R., Kunze, J., Grosse, F. R., & Dibbern, H. (1989). A syndrome of abnormal facies, short stature, and psychomotor retardation. *Atlas of Clinical Syndromes: A Visual Aid to Diagnosis for Clinicians and Practicing Physicians*, 198–199.
- Wolffe, A. (1998). *Chromatin: Structure and Function*. Academic Press.
- Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., & Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature*, 389(6652), 745–749. <https://doi.org/10.1038/39631>
- Yagi, H., Deguchi, K., Aono, A., Tani, Y., Kishimoto, T., & Komori, T. (1998). Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood*, 92(1), 108–117.
- Yang, X.-J., & Seto, E. (2007). HATs and HDACs: From structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*, 26(37), 5310–5318. <https://doi.org/10.1038/sj.onc.1210599>
- Yeo, G., Holste, D., Kreiman, G., & Burge, C. B. (2004). Variation in alternative splicing across human tissues. *Genome Biology*, 5(10), R74.
<https://doi.org/10.1186/gb-2004-5-10-r74>
- Yildirim, O., Gottwald, M., Schöler, P., & Michel, M. C. (2016). Opportunities and Challenges for Drug Development: Public–Private Partnerships, Adaptive Designs and Big Data. *Frontiers in Pharmacology*, 7.
<https://doi.org/10.3389/fphar.2016.00461>
- Yokoyama, A., Ficara, F., Murphy, M. J., Meisel, C., Naresh, A., Kitabayashi, I., & Cleary, M. L. (2011). Proteolytically cleaved MLL subunits are susceptible to

- distinct degradation pathways. *Journal of Cell Science*, 124(Pt 13), 2208–2219.
<https://doi.org/10.1242/jcs.080523>
- Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., & Cleary, M. L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Molecular and Cellular Biology*, 24(13), 5639–5649.
<https://doi.org/10.1128/MCB.24.13.5639-5649.2004>
- Yoshinobu, K., Araki, M., Morita, A., Araki, M., Kokuba, S., Nakagata, N., & Araki, K. (2021). Tamoxifen feeding method is suitable for efficient conditional knockout. *Experimental Animals*, 70(1), 91–100.
<https://doi.org/10.1538/expanim.19-0138>
- Younesian, S., Yousefi, A.-M., Momeny, M., Ghaffari, S. H., & Bashash, D. (2022). The DNA Methylation in Neurological Diseases. *Cells*, 11(21), 3439.
<https://doi.org/10.3390/cells11213439>
- Yu, B. D., Hanson, R. D., Hess, J. L., Horning, S. E., & Korsmeyer, S. J. (1998). MLL, a mammalian *trithorax*-group gene, functions as a transcriptional maintenance factor in morphogenesis. *Proceedings of the National Academy of Sciences*, 95(18), 10632–10636. <https://doi.org/10.1073/pnas.95.18.10632>
- Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A., & Korsmeyer, S. J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. *Nature*, 378(6556), 505–508. <https://doi.org/10.1038/378505a0>
- Zamudio, N. M., Scott, H. S., Wolski, K., Lo, C.-Y., Law, C., Leong, D., Kinkel, S. A., Chong, S., Jolley, D., Smyth, G. K., de Kretser, D., Whitelaw, E., & O'Bryan, M. K. (2011). DNMT3L is a regulator of X chromosome compaction and post-meiotic gene transcription. *PloS One*, 6(3), e18276.
<https://doi.org/10.1371/journal.pone.0018276>

- Zelevnik-Le, N. J., Harden, A. M., & Rowley, J. D. (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proceedings of the National Academy of Sciences of the United States of America*, 91(22), 10610–10614. <https://doi.org/10.1073/pnas.91.22.10610>
- Zemheri, E., Engin Zerk, P., & Ulukaya Durakbasa, C. (2021). Calretinin immunohistochemical staining in Hirschsprung's disease: An institutional experience. *Northern Clinics of Istanbul*, 8(6), 601–606. <https://doi.org/10.14744/nci.2020.69376>
- Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, 21(8), 1273–1283. <https://doi.org/10.1101/gr.122382.111>
- Zheng, S., & Black, D. L. (2013). Alternative pre-mRNA splicing in neurons: Growing up and extending its reach. *Trends in Genetics: TIG*, 29(8), 442–448. <https://doi.org/10.1016/j.tig.2013.04.003>
- Zhu, T., Liang, C., Li, D., Tian, M., Liu, S., Gao, G., & Guan, J.-S. (2016). Histone methyltransferase Ash1L mediates activity-dependent repression of neurexin-1 α . *Scientific Reports*, 6, 26597. <https://doi.org/10.1038/srep26597>
- Zlatanova, J., Leuba, S. H., & van Holde, K. (1998). Chromatin fiber structure: Morphology, molecular determinants, structural transitions. *Biophysical Journal*, 74(5), 2554–2566. [https://doi.org/10.1016/S0006-3495\(98\)77963-9](https://doi.org/10.1016/S0006-3495(98)77963-9)

Original Publications

Paper I

Paper I

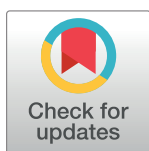
RESEARCH ARTICLE

Missense variants causing Wiedemann-Steiner syndrome preferentially occur in the KMT2A-CXXC domain and are accurately classified using AlphaFold2

Tinna Reynisdottir¹, Kimberley Jade Anderson¹, Leandros Boukas^{2,3*}, Hans Tomas Bjornsson^{1,2,4,5*}

1 Laboratory of Translational Medicine, Faculty of Medicine, University of Iceland, Reykjavik, Iceland, **2** McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **3** Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **4** Department of Pediatrics, Johns Hopkins University, Baltimore, Maryland, United States of America, **5** Department of Genetics and Molecular Medicine, Landspítali University Hospital, Reykjavik, Iceland

* lboukas1@jh.edu (LB); htb@hi.is (HTB)



OPEN ACCESS

Citation: Reynisdottir T, Anderson KJ, Boukas L, Bjornsson HT (2022) Missense variants causing Wiedemann-Steiner syndrome preferentially occur in the KMT2A-CXXC domain and are accurately classified using AlphaFold2. PLoS Genet 18(6): e1010278. <https://doi.org/10.1371/journal.pgen.1010278>

Editor: John M. Greally, Albert Einstein College of Medicine, UNITED STATES

Received: January 16, 2022

Accepted: May 27, 2022

Published: June 21, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1010278>

Copyright: © 2022 Reynisdottir et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Variant data are available in the manuscript and [supplementary](#)

Abstract

Wiedemann-Steiner syndrome (WDSTS) is a neurodevelopmental disorder caused by *de novo* variants in *KMT2A*, which encodes a multi-domain histone methyltransferase. To gain insight into the currently unknown pathogenesis of WDSTS, we examined the spatial distribution of likely WDSTS-causing variants across the 15 different domains of KMT2A. Compared to variants in healthy controls, WDSTS variants exhibit a 61.9-fold overrepresentation within the CXXC domain—which mediates binding to unmethylated CpGs—suggesting a major role for this domain in mediating the phenotype. In contrast, we find no significant overrepresentation within the catalytic SET domain. Corroborating these results, we find that hippocampal neurons from *Kmt2a*-deficient mice demonstrate disrupted histone methylation (H3K4me1 and H3K4me3) preferentially at CpG-rich regions, but this has no systematic impact on gene expression. Motivated by these results, we combine accurate prediction of the CXXC domain structure by AlphaFold2 with prior biological knowledge to develop a classification scheme for missense variants in the CXXC domain. Our classifier achieved 92.6% positive and 92.9% negative predictive value on a hold-out test set. This classification performance enabled us to subsequently perform an *in silico* saturation mutagenesis and classify a total of 445 variants according to their functional effects. Our results yield a novel insight into the mechanistic basis of WDSTS and provide an example of how AlphaFold2 can contribute to the *in silico* characterization of variant effects with very high accuracy, suggesting a paradigm potentially applicable to many other Mendelian disorders.

files. ChIP-seq and RNA-seq are available on GEO GSE99250.

Funding: This work was supported by a grant from the Wiedemann-Steiner Foundation to HTB (salary coverage of TR). HTB is also supported by the Louma G. Foundation, the Icelandic Research Fund (#217988, #195835, #206806) and the Icelandic Technology Development Fund (#2010588). Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health (#R01GM121459 to LB). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: HTB is a consultant for Mahzi therapeutics. No other authors have any potential conflict of interest.

Author summary

Wiedemann-Steiner syndrome (WDSTS) is a neurodevelopmental pediatric disorder caused by the genetic disruption of the histone methyltransferase KMT2A. Since KMT2A has many different domains that perform different functions, we reasoned that by identifying the domains most enriched for WDSTS-causing genetic variants we would gain insights into the incompletely understood molecular pathogenesis of WDSTS. We discovered that the CXXC domain—which binds unmethylated CpGs—shows by far the greatest enrichment, suggesting that loss of the CpG-binding ability of KMT2A plays a central role in WDSTS. Next, to understand specific rules underlying the genetic disruption of the CXXC domain, we combined prior knowledge about the function/structure of the domain with 3D structure prediction by AlphaFold2 to develop an effect classifier for CXXC missense variants. We found that this classifier exhibits accurate performance, and we therefore applied it to provide classifications for any such variant that can possibly arise, in order to aid in the interpretation of such variants in the clinic. Our work provides novel insights into WDSTS and suggests a strategy for missense variant classification that can potentially be applied to many other pediatric genetic disorders.

Introduction

Wiedemann Steiner syndrome (WDSTS, OMIM: 605130) is a Mendelian disorder of the epigenetic machinery. Its phenotypic features include intellectual disability, postnatal growth deficiency, hypertrichosis, and characteristic facial features. WDSTS is typically caused by heterozygous *de novo* variants in the gene encoding the histone-lysine N-methyltransferase 2A (KMT2A) [1]. KMT2A, also known as *MLL/MLL1*, is post-translationally cleaved into N-terminal and a C-terminal fragments, which subsequently heterodimerize. Each of these fragments contains several annotated protein domains. The larger N-terminal fragment contains three AT-hooks, a cysteine-rich CXXC domain, four plant homeodomain (PHD) fingers, a bromodomain and a FYR-N domain. The smaller C-terminal fragment contains a transactivation (TAD) domain, a *Win* motif, FYR-C domain, a SET, and a post-SET domain [2].

In this study, we address two questions. First, how important is the role of the different KMT2A domains in the pathogenesis of WDSTS? The answer would provide important clues into the molecular basis of the disorder, since the different domains mediate different functions. Second, what are the rules that determine how the genetic disruption of the most important domain(s) causes WDSTS? The answer would enable a systematic characterization of WDSTS-causing variants, informing basic biology as well as future clinical decision-making.

To answer the first question, we adopt an unbiased genetic approach. We examine the spatial distribution of likely WDSTS-causing missense variants across the different domains, reasoning that such variants will be enriched in the domains most critical for WDSTS pathogenesis when compared against variants in healthy individuals. To address the second question, we focus on the CXXC domain responsible for binding unmethylated CpGs [3], which we find shows by far the greatest enrichment. By leveraging the recent breakthrough in protein structure prediction from primary amino acid sequence by the deep neural network-based AlphaFold2 [4], we combine accurate 3D structure prediction with existing biological data in order to create an effect classification scheme for missense variants in the CXXC domain. After evaluating our classifier, we deploy it to perform an *in silico* saturation mutagenesis, examining a total of 445 variants.

Results

Preferential occurrence of missense variants likely pathogenic for Wiedemann-Steiner syndrome in the CXXC domain of *KMT2A*

If certain domains of a protein are critical for its function, disease-causing missense variants will tend to preferentially occur in these domains. We thus set out to explore the distribution of *KMT2A* missense variants (MVs) across its 15 different domains. We started by identifying 68 MVs that are likely pathogenic for WDSTS (Methods; Fig 1A). As a control, we used 1403 MVs which are present in gnomAD and are therefore not expected to cause WDSTS (Methods; Fig 1A). We tested each *KMT2A* domain for enrichment of WDSTS MVs relative to gnomAD MVs. We discovered a 61.9-fold enrichment in the CXXC domain (Fig 1B; Fisher's

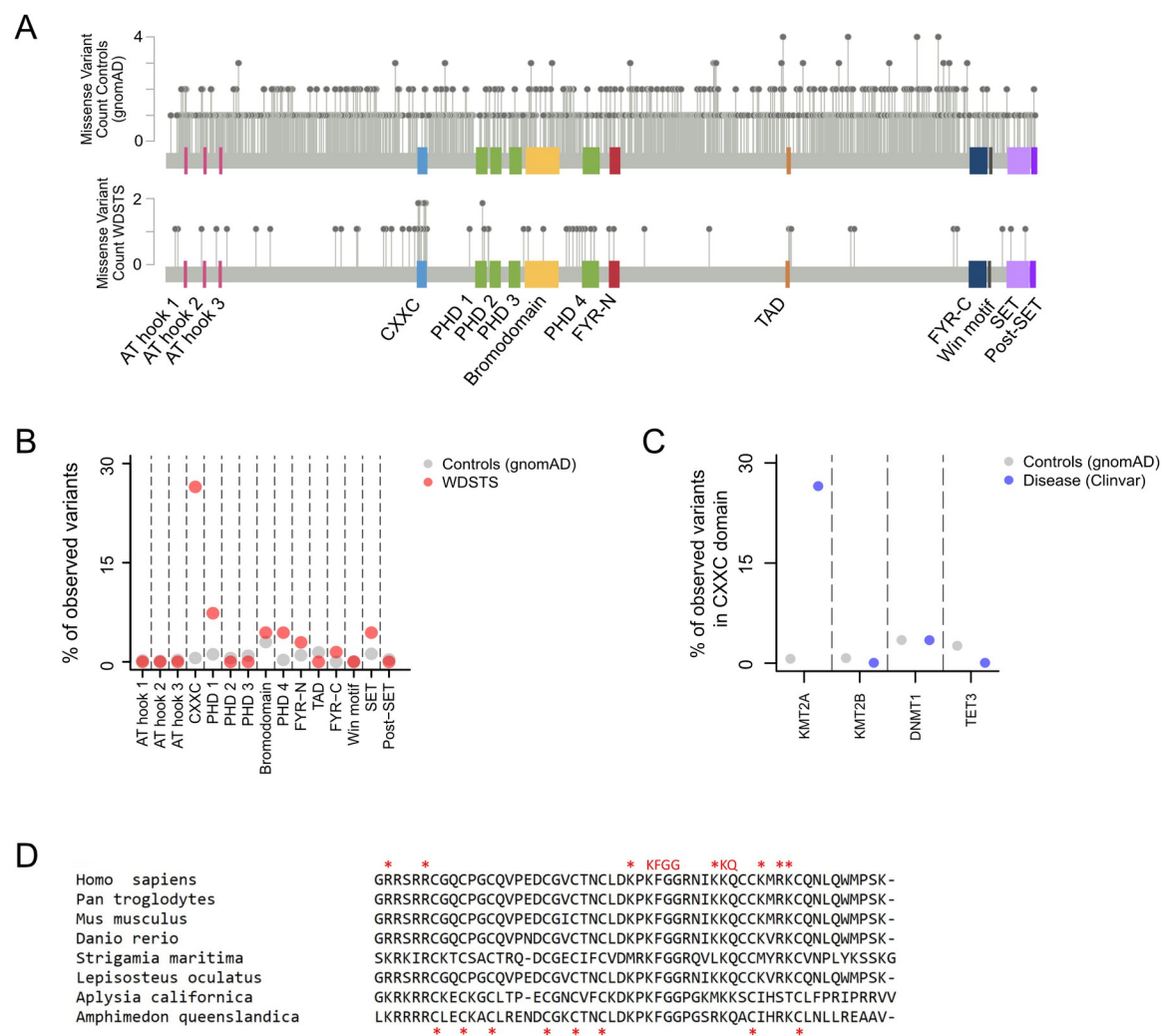


Fig 1. The distribution of likely pathogenic Wiedemann-Steiner syndrome missense variants across the different domains of *KMT2A*. (A) *KMT2A* missense variants in gnomAD (top) and WDSTS (bottom). See Methods for filtering criteria. (B) The percentage of missense variants from gnomAD (grey dots) and WDSTS (red dots) that fall in each of the different domains of *KMT2A*. (C) The percentage of missense variants in gnomAD (grey dots) and likely pathogenic variants (blue dots) that fall in the CXXC domain of different epigenetic regulators. (D) Multiple sequence alignment of the amino-acid sequence of the CXXC domain of *KMT2A* in eight eukaryotic species. Residues known to be important for DNA binding are marked with red asterisks at the top (see Methods for details). The eight zinc ion-binding cysteines are marked with red asterisks at the bottom.

<https://doi.org/10.1371/journal.pgen.1010278.g001>

exact test, Bonferroni-adjusted $p = 1.56 \times 10^{-18}$). This far exceeded the enrichment observed at other domains, with the second most enriched domain being the fourth PHD finger (Fig 1B; 16-fold enrichment; Fisher's exact test, Bonferroni-adjusted $p = 0.043$), and the first PHD finger also showing significant enrichment (Fig 1B; 6.8-fold enrichment; Fisher's exact test, Bonferroni-adjusted $p = 0.031$).

To assess the robustness of our result, we repeated our analysis using a control set of 1788 *KMT2A* somatic variants obtained from sequencing of tumor samples (Methods). While these variants are likely a mix of driver and passenger variants, the latter are expected to comprise the majority, justifying the use of this set of variants as controls alternative to gnomAD. We recapitulated our result, with the rank ordering of the different domains with respect to their enrichment of WDSTS MVs remaining unchanged (S1A–S1C Fig); this is consistent with the pool of somatic variants containing mostly benign passenger variants. However, across all domains, the enrichment estimates are attenuated compared to the gnomAD-vs-WDSTS comparison (S1C Fig). This suggests that the same domains that contribute to WDSTS pathogenesis contribute to the tumorigenic role of *KMT2A* as well. With respect to this tumorigenic role, however, our result should be interpreted with caution; gain-of-function mechanisms that are likely not captured by our analysis are probably also involved, as evidenced from *KMT2A* translocations that act as drivers in certain types of leukemias.

Notably, the catalytic SET domain of *KMT2A* does not show significant enrichment for WDSTS MVs (Fig 1B; Fisher's exact test, $p = 0.899$). In contrast, it shows significant, albeit weak, enrichment, when the somatic cancer variants are compared against the gnomAD controls (odds ratio = 2.87, Fisher's exact test, $p = 8.43 \times 10^{-5}$). We note here that, while the enrichment of WDSTS-causing MVs within the CXXC domain is consistent with the high conservation of its sequence (Figs 1D and S2), conservation alone cannot explain the lack of enrichment in the SET domain, since it is as conserved as the CXXC domain (S2 Fig).

The CXXC domains of other epigenetic regulators do not show enrichment for disease-causing missense variants

We next asked if the preferential occurrence of disease-causing missense variants in the CXXC domain is unique to *KMT2A*, or whether this is a general phenomenon across epigenetic regulators that have this domain. Apart from *KMT2A*, three other CXXC-domain-containing epigenetic regulators have been linked to Mendelian diseases: *KMT2B* (DYS28; OMIM:617284), *DNMT1* (HSN1E; OMIM:614116), and *TET3* (BEFAHRS; OMIM:618798). However, we found that none of these genes exhibits significant enrichment of disease-causing MVs in the CXXC domain (Fig 1C; Methods; Fisher's exact test, *DNMT1* $p = 1$, *KMT2B* $p = 1$, *TET3* $p = 1$), and verified that this is not due to inadequate power (Methods).

The disruption of histone methylation in *Kmt2a*-deficient mice preferentially occurs at CpG-rich regions but has no systematic effect on gene expression

Since the CXXC domain mediates binding to clusters of unmethylated CpG dinucleotides [3], our domain enrichment results imply that *KMT2A* exerts its most important function at CpG-rich regions. Our results also imply that its most important function is not its catalytic activity as a histone methyltransferase, given the lack of MV enrichment in the SET domain. We sought to test these implications, by comparing genome-wide histone methylation (H3K4me1 and H3K4me3) and gene expression patterns in hippocampal CA (Cornu Ammonis) neurons from mice with *Kmt2a* knockout in excitatory neurons (*Kmt2a* cKO) as well as wild-type mice (using ChIP-seq and RNA-seq data; Methods) [5].

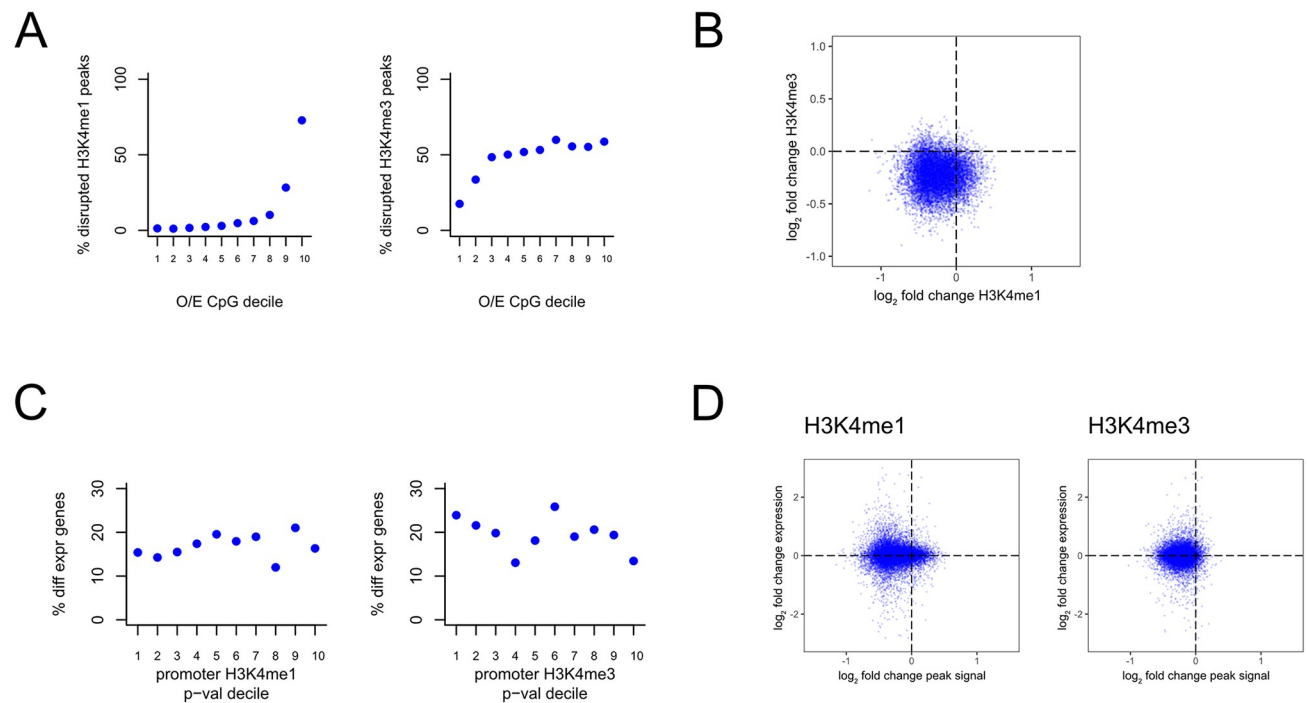


Fig 2. The relationship between disrupted H3K4me1/3, regional observed-to-expected CpG ratio, and gene expression in *Kmt2a*-deficient mice. (A) The percentage of disrupted H3K4me1 peaks (left) and H3K4me3 peaks (right), stratified based on the observed-to-expected CpG ratio of the underlying peak sequence. (B) Scatterplot of the log₂ fold change of H3K4me1 peaks against the log₂ fold change of H3K4me3 peaks at promoters (+/-1kb from the TSS) that harbor peaks for both marks. (C) The percentage of differentially expressed genes, stratified based on the p-value of associated promoter peaks (+/-1kb from the TSS) from the differential H3K4me1 analysis (left) and H3K4me3 (right) analysis. (D) Scatterplot of the log₂ fold change of H3K4me1 peaks (left) and H3K4me3 (right) against the log₂ fold change of gene expression of the downstream gene. Each point corresponds to a gene-promoter pair. In cases where multiple peaks were present at the same promoter, the average log₂ fold change was computed.

<https://doi.org/10.1371/journal.pgen.1010278.g002>

First, we found a strong relationship between the CpG-richness of a region and the probability of H3K4me1/3 disruption. CpG-rich peaks are much more likely to exhibit disruption upon *Kmt2a* cKO compared to CpG-poor peaks (Fig 2A; 5 wild-type vs 3 cKO mice; $p < 2.2 \times 10^{-16}$; Methods). Reassuringly, there is strong concordance between the changes of the two histone marks in the mutant mice; at the vast majority of promoter regions (+/-1kb from TSS) which bear both H3K4me1 and H3K4me3 peaks, both marks show decreased intensity in the mutant mice (Figs 2B and S3A). However, although regions with the strongest evidence for histone methylation disruption (1st p-value decile) most frequently correspond to promoters (39% and 86% for H3K4me1 and H3K4me3, respectively; 3.3-fold and 1.3-fold enrichment compared to regions within the 10th p-value decile, respectively; Fisher's exact test, $p < 2.2 \times 10^{-16}$; S3B Fig), we found no evidence for a systematic impact of promoter H3K4me1 or H3K4me3 disruption on gene expression (5 wild-type vs 6 cKO mice; Methods). Specifically, genes whose promoters have disrupted H3K4me1 or H3K4me3 are not significantly more likely to be differentially expressed compared to genes without histone methylation disruption at their promoter (Fig 2C and 2D; $p = 0.91$ and $p = 0.13$ for H3K4me1 and H3K4me3, respectively, when testing for a shift in p-value distribution between the 1st and 10th deciles with the one tailed Wilcoxon rank-sum test; Methods). Taken together, these results show that, at least in adult excitatory CA hippocampal neurons: a) KMT2A preferentially acts at high-CpG-density regions, and b) its catalytic activity has little effect on gene expression, suggesting that it may also not affect higher-level phenotypes.

An AlphaFold2-based scheme classifies missense variants in the CXXC domain of *KMT2A* with high accuracy

Given our evidence for a central role of the KMT2A-CXXC domain in WDSTS, we sought to develop a variant classification scheme that would: a) label any possible variant in the CXXC domain as pathogenic or not, and b) in the case of pathogenic variants, provide a characterization of their functional effect. We reasoned that such a classifier should take into account the effect of variants on the secondary structures of the domain, the mean Coulombic electrostatic potential of the mutant domain structure, and the ability of the mutated domain to form hydrogen bonds with the DNA backbone.

To assess the feasibility of our approach, we first examined if AlphaFold2 (AF2)—which has recently enabled the determination of 3D-protein structures with experimental-level accuracy—accurately predicts the structure of the CXXC domain. We observed a highly confident prediction (Fig 3A; pLDDT > 70 for 96.5% of residues and pLDDT > 90 for 82.5% of residues;

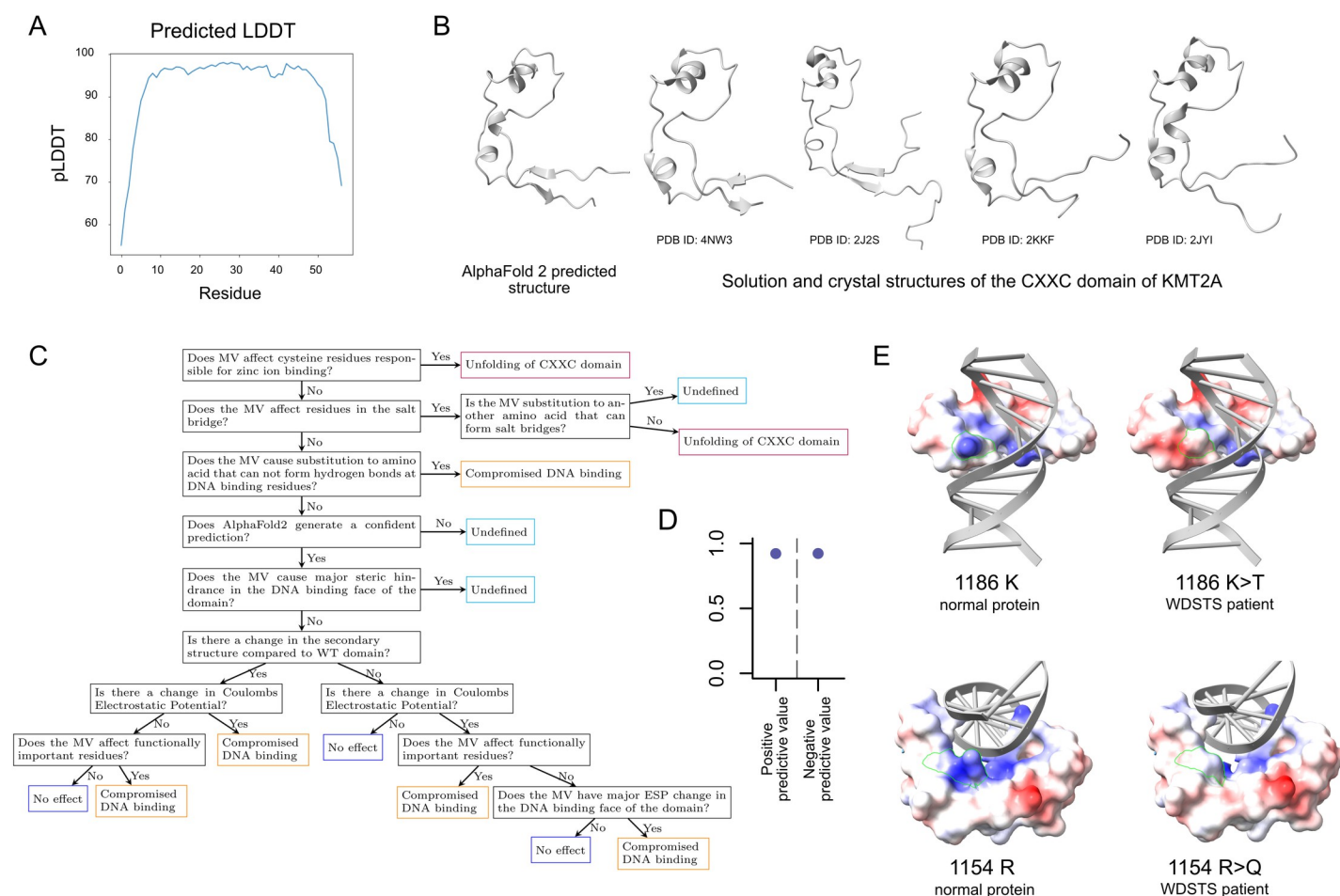


Fig 3. An AlphaFold2-based variant effect classification scheme for the CXXC domain of *KMT2A*. (A) Predicted LDDT values for the CXXC domain of *KMT2A*. (B) The AlphaFold2-predicted and experimentally determined structures of the CXXC domain of *KMT2A*. (C) The variant effect classification scheme. See [Methods](#) for details on the derivation of the scheme. (D) The positive and negative predictive value that the classifier shown in (C) attained on a hold-out test set consisting of 41 missense variants (see main text and [Methods](#) for details). (E) The position of two different missense variants in the 3D structure of the CXXC domain, in conjunction with a 3D representation of the engaged DNA backbone. For comparison, the same representation is shown for the normal protein as well (PDB ID: 4NW3). The surface of the domain is color-coded based on the electrostatic potential, with red indicating a negative charge and blue a positive charge. Both variants lead to a decreased electrostatic potential in a residue important for DNA binding.

<https://doi.org/10.1371/journal.pgen.1010278.g003>

Methods). In addition, all features previously identified in solution and crystal structures of the domain (ID: 2J2S, 2JYI, 4NW3, 2KKF) are present in the predicted structure: a crescent overall shape, two antiparallel beta sheets at the N and C terminals, and four alpha helices (Fig 3B). While no single experimentally derived structure contains all these features, this can be explained by the domain existing in multiple conformational states. We then proceeded to derive a classification scheme, using a training set of 14 MVs with experimentally determined effects and prior biological knowledge about the domain (Figs 3C and S6A, S3 Table; **Methods**) [6]. The first decision rules of the scheme pertain to residues whose substitution effects are not captured by AF2 (**Methods**): the eight cysteines responsible for zinc ion binding, the residues that form direct hydrogen bonds with the DNA, and the residues forming the salt bridge. The rest of the scheme is then divided into two cascades, based on the AF2-predicted secondary structure of the mutant domain and its mean Coulombic electrostatic potential (**Methods**). Variants receive different classifications depending on their effect on secondary structure/electrostatic potential, as well as their position.

To evaluate the performance of our classifier, we used a hold-out test set consisting of 41 MVs. Of these, 18 are MVs with strong evidence for pathogenicity for WDSTS (**Methods**), 13 are MVs with experimentally determined effects [6,7], and 10 are MVs seen in gnomAD/TOPMed, and are thus expected to be benign (S6B Fig, S3 Table). On this test set, our classifier attained a 92.6% positive and a 92.9% negative predictive value (Figs 3D, S6B and S6C). As an example, in Fig 3E we depict two WDSTS pathogenic variants which our classifier correctly labels as compromising DNA binding because they decrease the electrostatic potential at residues important for the formation of contacts with the DNA backbone. Notably, out of the 18 WDSTS MVs present in the CXXC domain of KMT2A, nine are positioned at cysteine residues responsible for zinc ion binding compared to none of the control variants (Fisher's exact test, $p = 0.009816$).

***In silico* saturation mutagenesis classifies 445 variants in the CXXC domain of KMT2A**

Clinical geneticists seeking to establish a diagnosis for patients are often confronted with missense variants that have not been seen before; this can make it hard to assess if they are pathogenic or not. The accuracy with which our classifier performs motivated us to perform an *in silico* saturation mutagenesis for the CXXC domain of KMT2A. Our goal was to create a resource that will enable rapid classification of any newly encountered missense variant in this domain, so that these classifications can be used as supporting evidence in the clinical setting. We focused on the 50/57 residues that have pLDDT > 70 and high MSA coverage. In total, we assessed 450 variants (Fig 4A).

Out of these 450 variants, 92 (20.4%) are synonymous, and 27 variants (6%) lead to premature stop codons. The remaining 331 variants (331/450, 73.6%) were classified based on our variant classification scheme. 169 variants (169/331, 51.1%) were predicted to have no effect, 90 (90/331, 27.2%) variants were classified as causing compromised DNA binding, and 67 variants (67/331, 20.2%) were classified as causing unfolding of the domain. For 5 variants (1.5%), we provide no prediction (Fig 4B; S4 Table). To obtain orthogonal validation for our classifications, we examined the conservation of individual nucleotides coding for the CXXC domain stratified according to the number of substitutions predicted to be damaging, and found strong concordance; sites with a greater number of predicted damaging substitutions are more conserved (Fig 4C).

Discussion

Our contribution in this work is twofold. First, we provide strong genetic evidence that the domains most important for mediating the causal role of KMT2A in Wiedemann-Steiner

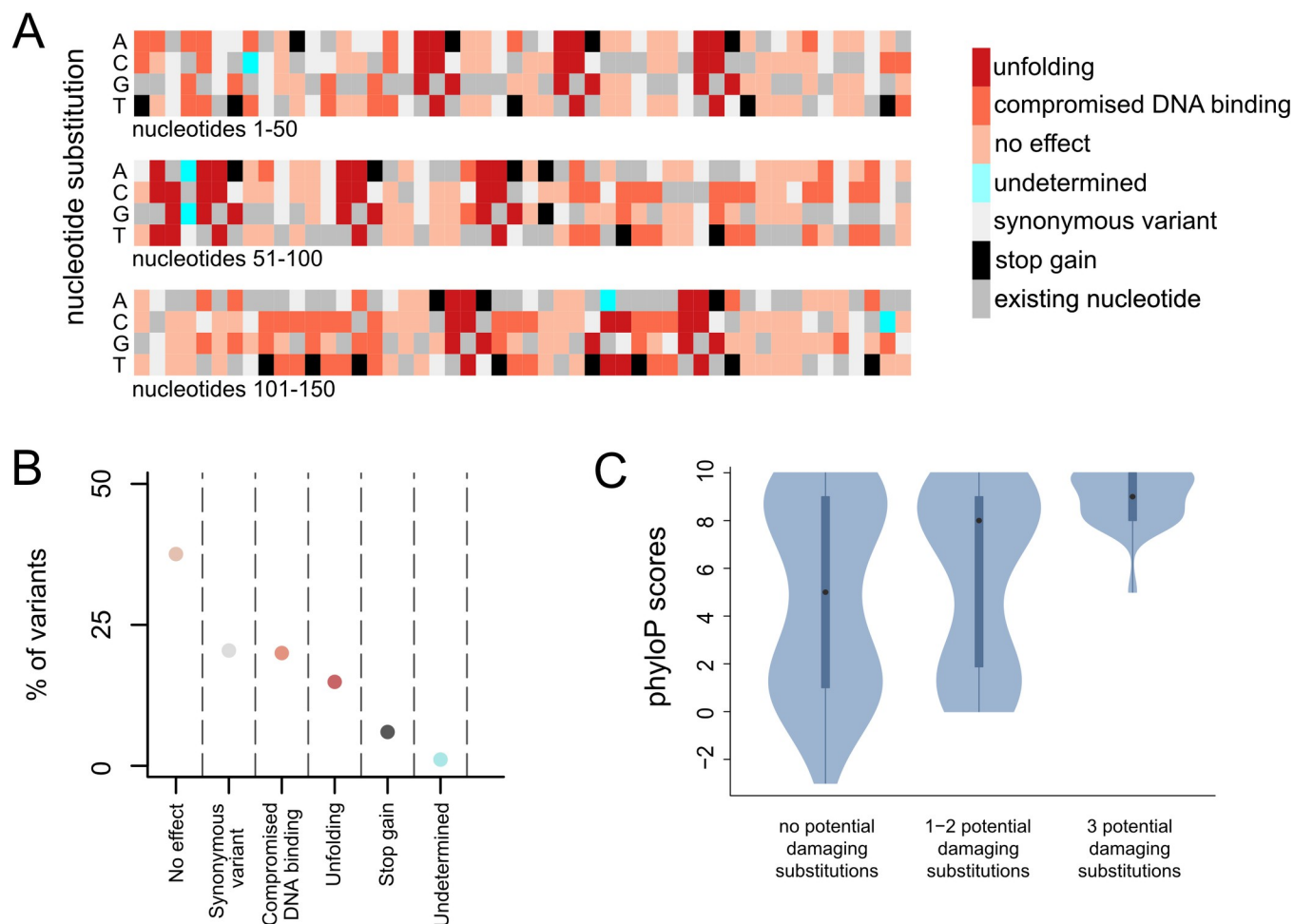


Fig 4. An *in silico* saturation mutagenesis of the CXXC domain of KMT2A. (A) Heatmap depicting the predicted effect of each nucleotide substitution within the CXXC domain. (B) The percentage of variants for each type of predicted effect. (C) The distribution of the phyloP score of the nucleotides coding for the CXXC domain, stratified according to the number of substitutions predicted to have a damaging effect (unfolding, compromised DNA binding, stop-gain).

<https://doi.org/10.1371/journal.pgen.1010278.g004>

syndrome are the CXXC domain and—to a lesser extent—the fourth and first PHD fingers. It is noteworthy that these PHD fingers have been shown to be important for stabilizing the interaction between the N- and the C-terminal KMT2A fragments [8]; their enrichment for WDSTS-causing missense variants may thus not be attributable to their histone-binding function. We emphasize here that our domain fold-enrichment estimates are based on a relatively low number of WDSTS variants and may change once more variants are reported. However, we anticipate the rank-ordering of the different domains, and the finding that the CXXC is by far the most enriched domain, to remain unchanged.

Our ChIP- and RNA-seq results suggest that lack of KMT2A recruitment to CpG-rich locations—either by missense variants in the CXXC domain or by loss-of-function variants—is central to WDSTS pathogenesis. However, they also suggest that the phenotype is not mediated via ensuing systematic defects in histone methylation-dependent gene expression, but rather that alternative mechanisms might be at play. Such mechanisms may involve defects in polymerase loading, which has previously been shown to depend on the presence, but not the catalytic activity, of some epigenetic regulators [9]. Alternatively, KMT2A may serve as a recruiter of other regulatory factors. Our results emphasize that more research aimed at elucidating

such alternative pathways is warranted, and has the potential to contribute to our understanding of WDSTS. This is also consistent with the lack of overrepresentation of WDSTS missense variants within the enzymatic SET domain of KMT2A, and stands in contrast to what was previously observed in Kabuki syndrome, a Mendelian disorder of the epigenetic machinery with considerable phenotypic overlap with WDSTS caused by variants in KMT2D, which is structurally similar to KMT2A [10]. We emphasize here, however, that our findings are based only on data from adult excitatory neurons from the CA region of the hippocampus, whereas gene expression and histone marks are known to be dynamic during development and vary between different cell types. Consequently, future work is needed to assess whether our conclusions generalize to other cell types (neuronal or not) involved in WDSTS pathogenesis, and to earlier developmental stages, when the disease process most likely initiates.

Our second contribution is the demonstration that the recent breakthrough in protein structure prediction by AlphaFold2 can be leveraged in order to classify the effects of missense variants with high accuracy. We highlight that here we use AF2 to directly predict the structure of the mutant proteins; based on these mutant structures, we then assess the effect of variants on secondary structural features and electrostatic potential. This is in contrast to recent work that only uses the wild-type structures as input to algorithms that predict the effect of variants on biophysical attributes like $\Delta\Delta G$ [11]. Prior to our study, it has been unclear if AlphaFold2 is capable of predicting mutant structures accurately, partly because of the use of the multiple sequence alignment. While we do not provide direct evidence, our results indirectly suggest that variant effects on local structural features can be reliably predicted. This conclusion is supported by the high concordance between our predictions and experimentally validated effects, as well as by the severe drop in performance when we use our classifier with inaccurate structure predictions as input (31% negative predictive value; S7 Fig) [12]. We do not however, expect AlphaFold2-based classifications to be able to capture global effects of variants, such as destabilization of the entire domain.

There were three variants in our hold-out test set that were classified erroneously. One of them (from the WDSTS cohort) is in fact annotated as a variant of uncertain significance, suggesting that our classification is not necessarily inaccurate in that case, but rather the “true” label may be incorrect. The other two erroneously classified variants may yield insights into possible limitations of our method. The TOPMed variant misclassified as resulting in compromised DNA binding may reflect the inability of our classifier to place effects along a gradient of severity; in other words, this variant may indeed affect DNA binding as we predict, but only to a mild extent not enough to cause severe disease, and is thus still present among healthy individuals. On the other hand, the pathogenic variant misclassified as benign is located in the structurally important KFGG site in the distal loop of the domain. This raises the possibility that our classifier could be improved by including an earlier decision rule assessing whether a variant affects the KFGG site, as we currently only utilize this information in later steps.

Within the context of the ACMG variant interpretation guidelines, our classifier can be used as an *in silico* tool that provides supporting evidence of benign impact (BP4 evidence class) or of pathogenicity (PP3 evidence class). We highlight that it is the first such tool that incorporates accurate 3D protein structure prediction by AlphaFold2, and does not directly use inputs such as conservation and population frequency, which are used by most other tools. Thus, we believe the use of our classifier in combination with other tools may prove particularly powerful for predicting the effect of variants in the CXXC domain of KMT2A.

In summary, our work yields insights into the pathogenesis of Wiedemann-Steiner syndrome and presents a strategy for characterizing variant effects using AlphaFold2 that we anticipate will be broadly applicable to many other disease-relevant proteins.

Materials and methods

Missense variants

Missense variants (MVs) present in the general population were obtained from the Genome Aggregation Database (gnomAD; version 2.1.1 and 3.1.1), which does not include individuals with severe pediatric disorders like WDSTS [13]. Somatic MVs were obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC; version 94) [14]. MVs present in individuals with disease phenotypes (Wiedemann-Steiner syndrome [KMT2A], Hereditary sensory neuropathy type 1E [DNMT1], Childhood-onset dystonia [KMT2B], and Beck-Fahrner syndrome [TET3]) were obtained from ClinVar [15]. To increase our power, we chose to include ClinVar variants labeled as Pathogenic, Likely Pathogenic, or Variants of Uncertain Significance (VUS), and filtered them using the phred-like CADD scores, acquired from Ensembl Variant Effect Predictor (VEP) [16]. Specifically, we only retained MVs with a phred-like CADD score above 20 [17]. In the case of WDSTS, we obtained 11 additional variants from the Human Gene Mutation Database (HGMD) with CADD score above 20, as well as 16 MVs from Lebrun et al, Baer et al, Miyake et al, WD Jones (all studies of individuals with a WDSTS clinical phenotype; [S1 Table](#)) [18–22]. Disease MVs that are also present in gnomAD were excluded from subsequent analyses.

To ensure that our domain enrichment estimates are not artifacts driven by the inclusion of VUS's, we also performed the domain enrichment analysis after excluding VUS's and obtained very similar results ([S1D Fig](#)). Without VUS's, the enrichment in the CXXC domain is even greater, but the confidence interval around the point estimate is wider ([S1D Fig](#); confidence interval = 49.2–442.8 without VUS's vs 24.3–173.4 when including VUS's). The same is true for the 1st and 4th PHD finger, which again show significant enrichment as well ([S1D Fig](#); For PHD finger 4; confidence interval = 1.9–163.0 without VUS's vs 2.3–97.0 when including VUS's and for PHD finger 1; confidence interval = 1.6–35.0 without VUS's vs 1.9–20.4 when including VUS's). Together, these results are consistent with the notion that, by excluding VUS's, we are increasing our signal-to-noise ratio, since we are not including any variants with uncertain pathogenicity. However, this comes at the expense of less power (reflected in the greater uncertainty around the enrichment point estimates), since this analysis is inevitably conducted using fewer variants.

Finally, since it is possible that very rare MVs in gnomAD may in fact be pathogenic, we assessed whether our result still holds when excluding low-frequency variants, and found this to be true. Using only variants with MAF greater than $10e-5$ (452 variants), we found that the CXXC domain still shows the greatest enrichment (odds ratio = 159.4, Fisher's exact test, $p = 3.66e-15$), followed by PHD fingers 4 and 1 (for PHD finger 4; odds ratio = inf, Fisher's exact test, $p = 0.0322$ and for PHD finger 1; odds ratio = 5.8, Fisher's exact test, $p = 0.122$), while the SET domain shows no significant enrichment (Fisher's exact test, $p = 0.762$), [S1E Fig](#).

For the enrichment analysis in the CXXC domain of KMT2B and TET3 ([Fig 1C](#)), we tested if the observed lack of enrichment can be attributed to the low number of total counts in these genes (22 and 16, respectively). In both cases, we observe 0 MVs in the CXXC domain. In contrast, out of the MVs in gnomAD, 0.7% and 2.5% fall within the CXXC domain of KMT2B and TET3, respectively. Using these gnomAD percentages and the formula for the probability mass function of the binomial distribution, we calculated that, even if the true ratio of the percentage of disease variants falling in the CXXC domain to the corresponding percentage of gnomAD variants is 3 times less compared to the ratio for KMT2A, the probability of observing at least one MV in the CXXC domain is 84.5% for KMT2B and 99.9% for TET3. These estimates make inadequate power an unlikely explanation for the lack of MV enrichment in the

CXXC domain of KMT2B and TET3. The coordinates of protein domains were attained using InterPro [23] (S5 Table).

For Figs 1A and S1A, variants in *KMT2A* were plotted using the Mutation Mapper tool from cBioPortal [24].

Evolutionary conservation

The amino acid sequence of the CXXC domain of KMT2A orthologs in the eukaryotic species shown in Fig 1D were obtained using the STRING Database [25]. The alignment of the KMT2A orthologous proteins was performed using the Clustal Omega multiple sequence alignment tool from EMBL-EBI [26], with default parameters. The phyloP conservation scores of the SET and CXXC domain nucleotides across 100 vertebrates were obtained using the GenomicScores R package [27].

ChIP-seq analysis

Raw ChIP-seq sequencing data (fastq files containing unpaired reads) were downloaded from GSE99250 [5]. The reads were aligned to the mouse mm10 (GRCm38) reference genome with Bowtie2 using the default settings with the -U option for aligning unpaired reads, generating a sam file output for each fastq file [28]. The sam files were converted to bam files using Samtools view with the -b option, then sorted using Samtools sort with the -O BAM option for outputting bam files [29]. Peaks were called using MACS2, using a threshold of q-value (-q option) 0.1 for significant peaks, and options “—nomodel—extsize 200” [30]. The sorted bam files and the lists of significant peaks (xls files from MACS2) were used as input to the DiffBind package in R [31]. DiffBind was then run with default settings, except for the options “fragmentSize = 0, RemoveDuplicates = TRUE, filterFun = mean, score = DBA_SCORE_READS” in the dba.count command. Differential peaks obtained in DiffBind were given genomic annotations and converted to GRanges objects using the ChIPseeker package in R [32]. Sequences of peaks were obtained using the getSeq() function from the BSgenome.Mmusculus.UCSC.mm10 package in R [33]. The observed-to-expected CpG ratio was calculated using the following formula [34]:

$$o/e \text{ CpG ratio} = \frac{p(\text{CpG})}{p(C)p(G)}$$

, where $p(\text{CpG})$ represents the proportion of CpGs in a given region (similarly for $p(C)$ and $p(G)$).

RNA-seq analysis

Raw RNA-seq data (fastq files) were downloaded from GSE99250 [5]. Reads were pseudo-aligned to the mouse mm10 (GRCm38) reference transcriptome using Kallisto [35] (kallisto quant command) with the options—single for single-end reads, -l 150 as an approximation of fragment length, -s 20 as an approximation of fragment length standard deviation, and -b 100 for running 100 bootstraps. The Kallisto output was imported into R using the tximport package [36], with transcripts mapped to genes using the BiomaRt package in R [37]. The tximport software was run with options, type = “Kallisto” and ignoreTxVersion = TRUE. The data was filtered to exclude genes with less than 10 counts. Differential expression analysis was performed using DESeq2 with default settings [38].

To evaluate the robustness and reliability of our differential expression analysis, we performed the following quality control checks. First, we examined the distribution of the

resulting p-values (S4A Fig), which is consistent with a two-component mixture; one component corresponding to non-differential genes (p-values distributed uniformly between 0 and 1), and one component corresponding to differential genes (p-values concentrated close to 0). Such a distribution indicates a well-calibrated differential expression test, and the existence of true differential hits. Second, we performed a principal component analysis using the expression matrix after applying a regularized log transformation, as implemented in the `rlog()` function in `DeSeq2` with the setting “blind = TRUE”; this revealed that the mutants are separate from the wild-type on PCA space (S4B Fig). Finally, an MA plot (S4C Fig) indicates no obvious systematic biases in the expression data.

For Fig 2A and 2C, the percentage of differentially marked H3K4me1/3 peaks and differentially expressed genes were estimated from the corresponding p-value distributions using Storey's method [39,40], as implemented in the `qvalue` R package. Specifically, we used the `pi0est()` function, with the “`pi0.method`” parameter set to “bootstrap”.

Derivation of the variant effect classifier

AF2 structure predictions. Structure predictions from AF2 were generated using AF2 Colaboratory (v2.0), which does not use templates (which would be expected to bias the mutant structure towards the structure of the wild-type domain [4]). We do not provide a prediction for variants which result in a structure where: a) the pLDDT value of the beta sheets, positioned at the distal ends, drops below 70, or b) the pLDDT value drops to 85 or lower for more than two residues that are not positioned at the distal ends of the domain, or at residues that fall within a secondary structure that is predicted to be absent.

Electrostatic potential calculations. Secondary structure visualization and computation of mean Coulombic values were performed using UCSF ChimeraX (version 1.2.5) [41], with the AF2 predicted structure as input. MVs were classified as causing a change in electrostatic potential (ESP) if the mean electrostatic potential value of the mutant structure deviated by more than 0.2 (in either direction) from the wild-type domain, whose mean value is equal to 6.28.

Derivation of the decision rules defining the variant effect classification scheme. We obtained the experimentally determined effects of MVs in the CXXC domain of KMT2A from Allen et al and Cierpicki et al [6,7]. We then used 14 of these variants to form our training set. These variants were chosen to ensure a representation of different types of disruption (unfolding of the domain/defective DNA binding) within the training set. Within a group of variants causing the same type of disruption, a random subset was selected for inclusion in the training set (4 out of 8 variants causing unfolding, 6 out of 11 variants compromising DNA binding), whereas the rest were included in the test set used for evaluating the performance of our classifier (see section below). Based on the training set, as well as existing biological knowledge about the structure and function of the domain [6,7], we derived the decision rules defining our classifier as follows.

Our initial decision rule classifies variants which lead to substitutions of the zinc-ion binding cysteine residues as causing unfolding of the domain, since this was the case for all such variants in the training set (C1161A, C1173A, C1189A, D1166A, R1192A). Our next decision rule then classifies variants which lead to substitutions of amino acids responsible for forming direct hydrogen bonds with the DNA (S2 Table; Cierpicki et al [7]) to amino acids that are incapable of forming hydrogen bonds (Ala, Cys, Gly, Ile, Leu, Met, Phe, Pro, Val) as resulting in compromised DNA binding [42]; this was the case with two such variants in the training set (K1186A and Q1187A) and is consistent with the known chemistry underlying protein-DNA contacts.

We subsequently divide the classification scheme into two cascades, based on whether the variant to be classified affects the secondary structure of the domain. For the first cascade, pertaining to variants that do not cause a change in CXXC structure (as determined by AlphaFold2; see previous section), we use the change in ESP (computed using ChimeraX; see previous section) and the position of the variant to determine its effect. There were two variants (N1172A, C1188A) in our training set that were experimentally shown to have no effect on the folding of the domain or on DNA binding. We found that these variants caused neither a change in secondary structure, nor a change in ESP. Therefore, we included a decision rule classifying variants that affect neither the secondary structure nor the ESP as benign. There was also one variant (R1153A) that was experimentally shown to have no effect on folding or DNA binding, which we found caused an ESP change without a concomitant secondary structure change, but the ESP change was at a functionally non-important residue (facing away from the DNA backbone and thus unlikely to impact DNA binding). We thus decided to label variants causing no change in secondary structure and changes in ESP at functionally non-important residues as benign. By contrast, there were two variants (K1176A and R1151A) experimentally shown to compromise DNA binding, which we found caused no change in secondary structure, but caused an ESP change at a functionally important residue (that is, a residue that has been experimentally implicated in electrostatic interaction with the DNA backbone, or a residue forming hydrogen bonds with the DNA, or a residue at the structurally important KFGG site (S2 Table) [7,43,44]). Therefore, we included a decision rule classifying variants which cause no change in secondary structure but cause ESP changes at these sites as resulting in compromised DNA binding.

Finally, for the second cascade, variants that demonstrate a change in secondary structure as well as a change in ESP are classified as compromising DNA binding based on the experimentally determined effect of variants R1154A (loss of beta sheets) and D1175A (loss of an alpha helix). Additionally, similar to the prior cascade, we included a decision rule labeling variants that cause a secondary structure change, do not affect ESP, but affect functionally important residues, as resulting in compromised DNA binding, based on prior knowledge about the function of the domain. Finally, variants that cause secondary structure changes but have no impact on ESP and are not positioned at functionally important residues are classified as having no effect.

Test set. As described above, we included a random subset of variants from each potential type of disruption (unfolding/compromised DNA binding). In addition, we included 18 variants seen in patients with phenotypic features of WDSTS [15,18,19,21,22], and 7 and 3 variants from gnomAD and TOPMed [13,45], respectively, which are expected to be benign. With regards to the WDSTS variants, we know that they are pathogenic, but we do not know their precise damaging effect. Therefore, for these variants we are not able to assess whether the precise label that our classifier assigns (unfolding/compromised DNA binding) is true. However, we can still assess the positive predictive value and true positive rate by asking if known damaging variants (pathogenic WDSTS variants or variants with an experimentally determined detrimental effect) are correctly classified as causing some type of damage (either unfolding or disruption of DNA binding).

Saturation mutagenesis. In our saturation mutagenesis, we chose to not provide a prediction for the single variant (R1151P) which causes major steric hindrance in the DNA binding face of the domain, since we do not have data to determine the potential impact of this variant. With respect to the MVs that were classified as disrupting the electrostatic potential, the majority caused changes approximately equal to 0.5, with the minimum being 0.31.

Supporting information

S1 Fig. (A) *KMT2A* missense variants in COSMIC (top) and ClinVar (bottom). (B) The percentage of missense variants in COSMIC (grey dots) and WDSTS patients (red dots) that fall in each of the different domains of *KMT2A*. (C) Correlation between the fold-enrichment (odds ratio) of WDSTS MVs compared to COSMIC MVs and gnomAD MVs. (D) The percentage of missense variants in gnomAD (grey dots) and WDSTS patients (red dots) that fall in each of the different domains of *KMT2A*, after excluding variants of uncertain significance. (E) The percentage of missense variants in gnomAD (grey dots) and WDSTS patients (red dots) that fall in each of the different domains of *KMT2A*, after excluding gnomAD variants with $MAF < 10e-5$.

(PDF)

S2 Fig. PhyloP score distribution for the SET and CXXC domain nucleotides of *KMT2A*.

(PDF)

S3 Fig. (A) Venn diagram depicting the overlap between promoters (± 1 kb from TSS) harboring H3K4me1 peaks and those harboring H3K4me3 peaks. (B) Genomic annotation of peaks within the 1st (yellow dots) and 10th p-value decile (purple dots) for H3K4me1 peaks (left) and H3K4me3 peaks (right).

(PDF)

S4 Fig. (A) The histogram of p-values from the differential expression RNA-seq analysis. (B) PCA plot based on the expression matrix, after a variance stabilizing transformation (see [Methods](#)). (C) MA plot of the $\log_2$ fold-change against the mean of normalized counts from the differential expression analysis. Differentially expressed genes are colored in blue.

(PDF)

S5 Fig. (A) Multiple sequence alignment depth plot and (B) predicted alignment error of *KMT2A* CXXC wild-type domain prediction from AlphaFold2.

(PDF)

S6 Fig. Variant classification scheme from the (A) training set and (B) test set. Variants that do not fall under the correct classification according to the scheme are underlined. (C) Confusion matrix from test set results.

(PNG)

S7 Fig. Predicted structure of the CXXC domain of *KMT2A* using ColabFold without using the multiple sequence alignment or templates as input.

(PDF)

S1 Table. WDSTS variant information.

(CSV)

S2 Table. Functionally important residues of the CXXC domain of *KMT2A*.

(PDF)

S3 Table. Variant information from training and test set.

(TXT)

S4 Table. Saturation mutagenesis results.

(CSV)

S5 Table. Amino acid co-ordinates of the proteins domains of *KMT2A*.

(PDF)

Author Contributions

Conceptualization: Tinna Reynisdottir, Hans Tomas Bjornsson.

Data curation: Tinna Reynisdottir, Kimberley Jade Anderson.

Formal analysis: Leandros Boukas.

Funding acquisition: Hans Tomas Bjornsson.

Supervision: Leandros Boukas, Hans Tomas Bjornsson.

Visualization: Tinna Reynisdottir, Kimberley Jade Anderson.

Writing – original draft: Tinna Reynisdottir, Leandros Boukas.

Writing – review & editing: Leandros Boukas, Hans Tomas Bjornsson.

References

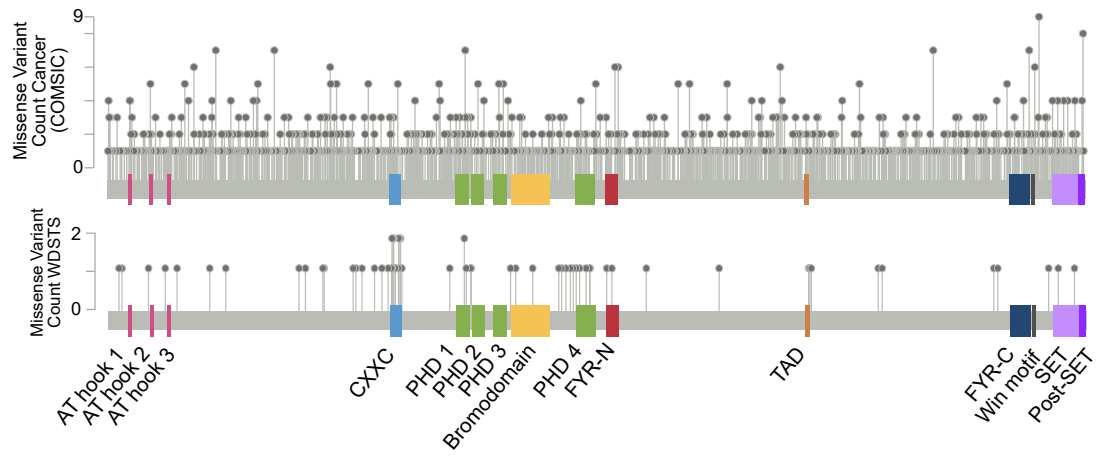
1. Jones Wendy D., et al., De Novo Mutations in MLL Cause Wiedemann-Steiner Syndrome. *The American Journal of Human Genetics*, 2012. 91(2): p. 358–364. <https://doi.org/10.1016/j.ajhg.2012.06.008> PMID: 22795537
2. Slany R.K., The molecular biology of mixed lineage leukemia. *Haematologica*, 2009. 94(7): p. 984–993. <https://doi.org/10.3324/haematol.2008.002436> PMID: 19535349
3. Voo K.S., et al., Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol*, 2000. 20(6): p. 2108–21. <https://doi.org/10.1128/MCB.20.6.2108-2121.2000> PMID: 10688657
4. Jumper J., et al., Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021. 596(7873): p. 583–589. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
5. Kerimoglu C., et al., KMT2A and KMT2B Mediate Memory Function by Affecting Distinct Genomic Regions. *Cell Reports*, 2017. 20(3): p. 538–548. <https://doi.org/10.1016/j.celrep.2017.06.072> PMID: 28723559
6. Allen M.D., et al., Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *The EMBO Journal*, 2006. 25(19): p. 4503–4512. <https://doi.org/10.1038/sj.emboj.7601340> PMID: 16990798
7. Cierpicki T., et al., Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. *Nature structural & molecular biology*, 2010. 17(1): p. 62–68. <https://doi.org/10.1038/nsmb.1714> PMID: 20010842
8. Yokoyama A., et al., Proteolytically cleaved MLL subunits are susceptible to distinct degradation pathways. *Journal of Cell Science*, 2011. 124(13): p. 2208–2219. <https://doi.org/10.1242/jcs.080523> PMID: 21670200
9. Dorigi K.M., et al., Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell*, 2017. 66(4): p. 568–576.e4. <https://doi.org/10.1016/j.molcel.2017.04.018> PMID: 28483418
10. Faundes V., et al., A comparative analysis of KMT2D missense variants in Kabuki syndrome, cancers and the general population. *Journal of Human Genetics*, 2019. 64(2): p. 161–170. <https://doi.org/10.1038/s10038-018-0536-6> PMID: 30459467
11. Akdel M., et al., A structural biology community assessment of AlphaFold 2 applications. *bioRxiv*, 2021: p. 2021.09.26.461876.
12. Mirdita M., et al., ColabFold—Making protein folding accessible to all. *bioRxiv*, 2022: p. 2021.08.15.456425.
13. Karczewski K.J., et al., The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 2020. 581(7809): p. 434–443. <https://doi.org/10.1038/s41586-020-2308-7> PMID: 32461654
14. Tate J.G., et al., COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, 2019. 47(D1): p. D941–D947. <https://doi.org/10.1093/nar/gky1015> PMID: 30371878
15. Landrum M.J., et al., ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic acids research*, 2018. 46(D1): p. D1062–D1067. <https://doi.org/10.1093/nar/gkx1153> PMID: 29165669

16. McLaren W., et al., Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* (Oxford, England), 2010. 26(16): p. 2069–2070. <https://doi.org/10.1093/bioinformatics/btq330> PMID: 20562413
17. Kircher M., et al., A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 2014. 46(3): p. 310–315. <https://doi.org/10.1038/ng.2892> PMID: 24487276
18. Lebrun N., et al., Molecular and cellular issues of KMT2A variants involved in Wiedemann-Steiner syndrome. *Eur J Hum Genet*, 2018. 26(1): p. 107–116. <https://doi.org/10.1038/s41431-017-0033-y> PMID: 29203834
19. Baer S., et al., Wiedemann-Steiner syndrome as a major cause of syndromic intellectual disability: A study of 33 French cases. *Clin Genet*, 2018. 94(1): p. 141–152. <https://doi.org/10.1111/cge.13254> PMID: 29574747
20. Miyake N., et al., Delineation of clinical features in Wiedemann-Steiner syndrome caused by KMT2A mutations. *Clinical Genetics*, 2016. 89(1): p. 115–119. <https://doi.org/10.1111/cge.12586> PMID: 25810209
21. Jones W.D., Genetic and phenotypic investigations into developmental disorders, in Wellcome Trust Sanger Institute. 2017, University of Cambridge: Newnham College.
22. Stenson P.D., et al., Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*, 2003. 21(6): p. 577–81. <https://doi.org/10.1002/humu.10212> PMID: 12754702
23. Blum M., et al., The InterPro protein families and domains database: 20 years on. *Nucleic acids research*, 2021. 49(D1): p. D344–D354. <https://doi.org/10.1093/nar/gkaa977> PMID: 33156333
24. Cerami E., et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*, 2012. 2(5): p. 401–404. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID: 22588877
25. Jensen L.J., et al., STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, 2009. 37(Database issue): p. D412–6. <https://doi.org/10.1093/nar/gkn760> PMID: 18940858
26. Madeira F., et al., The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, 2019. 47(W1): p. W636–W641. <https://doi.org/10.1093/nar/gkz268> PMID: 30976793
27. Puigdevall P., Castelo R., GenomicScores: seamless access to genomewide position-specific scores from R and Bioconductor. *Bioinformatics*, 2018. 34(18): p. 3208–3210. <https://doi.org/10.1093/bioinformatics/bty311> PMID: 29718111
28. Langmead B., Salzberg S.L., Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 2012. 9(4): p. 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
29. Li H., et al., The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 2009. 25(16): p. 2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
30. Zhang Y., et al., Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 2008. 9(9): p. R137. <https://doi.org/10.1186/gb-2008-9-9-r137> PMID: 18798982
31. Stark R., Brown G., DiffBind: differential binding analysis of ChIP-Seq peak data. 2011.
32. Yu G., Wang L.-G., He Q.-Y., ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 2015. 31(14): p. 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145> PMID: 25765347
33. TD T., BSgenome.Mmusculus.UCSC.mm10: Full genome sequences for *Mus musculus* (UCSC version mm10, based on GRCm38.p6), in R package version 1.4.3. 2021.
34. Gardiner-Garden M., Frommer M., CpG islands in vertebrate genomes. *J Mol Biol*, 1987. 196(2): p. 261–82. [https://doi.org/10.1016/0022-2836\(87\)90689-9](https://doi.org/10.1016/0022-2836(87)90689-9) PMID: 3656447
35. Bray N.L., et al., Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 2016. 34(5): p. 525–527. <https://doi.org/10.1038/nbt.3519> PMID: 27043002
36. Sonesson C., Love M.I., Robinson M.D., Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*, 2015. 4: p. 1521. <https://doi.org/10.12688/f1000research.7563.2> PMID: 26925227
37. Durinck S., et al., Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 2009. 4(8): p. 1184–1191. <https://doi.org/10.1038/nprot.2009.97> PMID: 19617889
38. Love M.I., Huber W., Anders S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 2014. 15(12): p. 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
39. Storey J.D., A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2002. 64(3): p. 479–498.

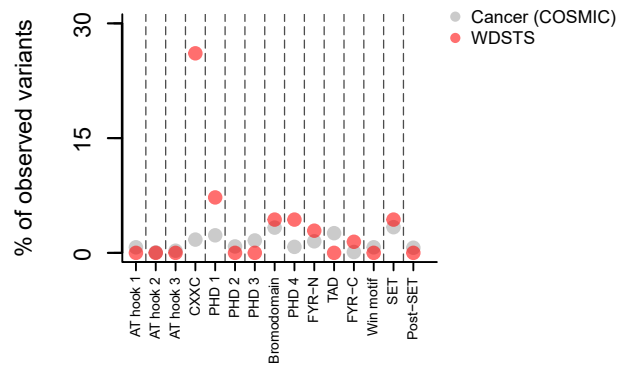
40. Storey J.D., Tibshirani R., Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 2003. 100(16): p. 9440–9445. <https://doi.org/10.1073/pnas.1530509100> PMID: [12883005](https://pubmed.ncbi.nlm.nih.gov/12883005/)
41. Pettersen E.F., et al., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science: a publication of the Protein Society*, 2021. 30(1): p. 70–82.
42. Hubbard R.E., Kamran Haider M., *Hydrogen Bonds in Proteins: Role and Strength*. 2010.
43. Frauer C., et al., Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1. *PLOS ONE*, 2011. 6(2): p. e16627. <https://doi.org/10.1371/journal.pone.0016627> PMID: [21311766](https://pubmed.ncbi.nlm.nih.gov/21311766/)
44. Ayton P.M., Chen E.H., Cleary M.L., Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Molecular and cellular biology*, 2004. 24(23): p. 10470–10478. <https://doi.org/10.1128/MCB.24.23.10470-10478.2004> PMID: [15542854](https://pubmed.ncbi.nlm.nih.gov/15542854/)
45. Taliun D., et al., Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, 2021. 590(7845): p. 290–299. <https://doi.org/10.1038/s41586-021-03205-y> PMID: [33568819](https://pubmed.ncbi.nlm.nih.gov/33568819/)

S1 Fig

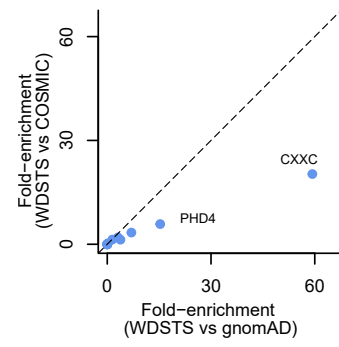
A



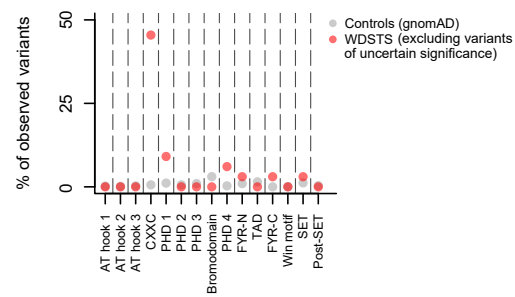
B



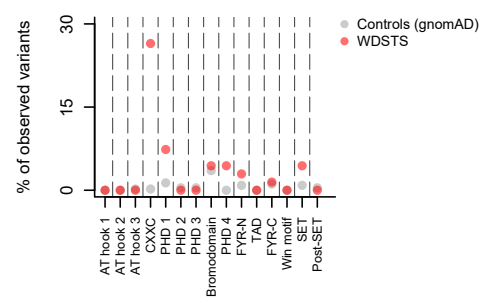
C



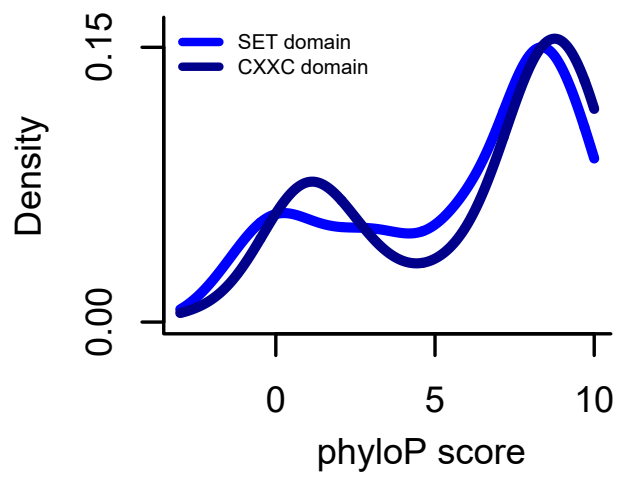
D



E

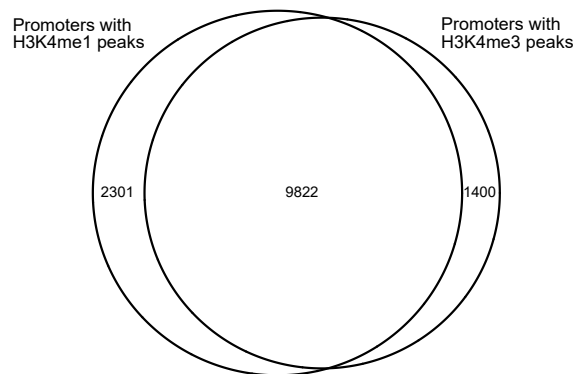


S2 Fig

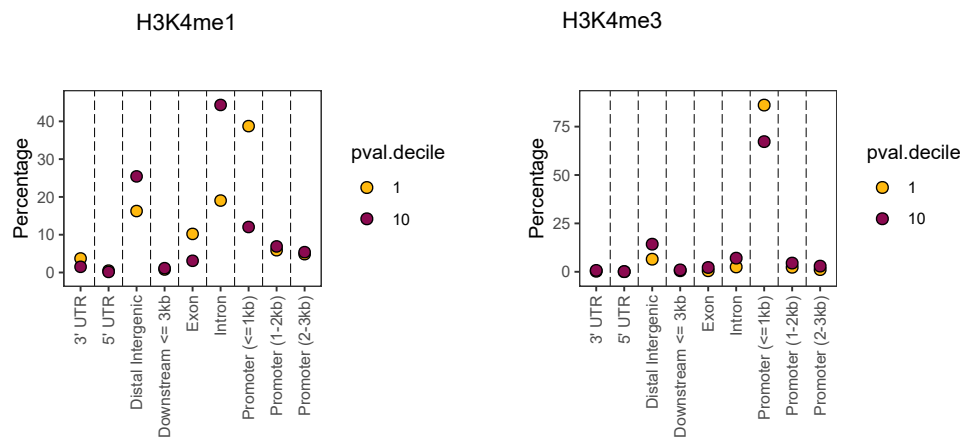


S3 Fig

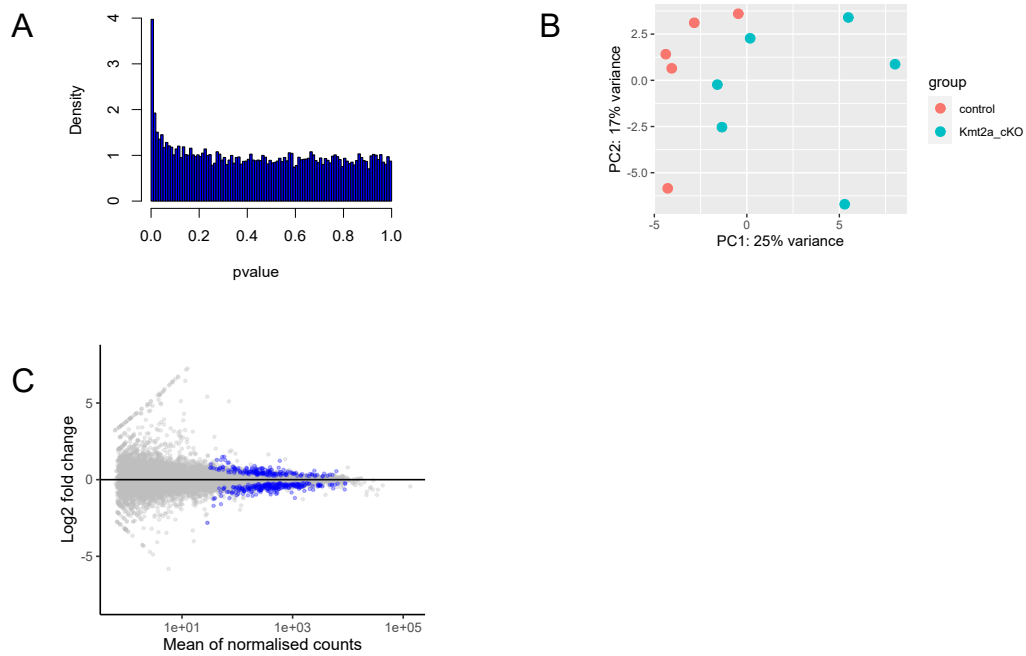
A



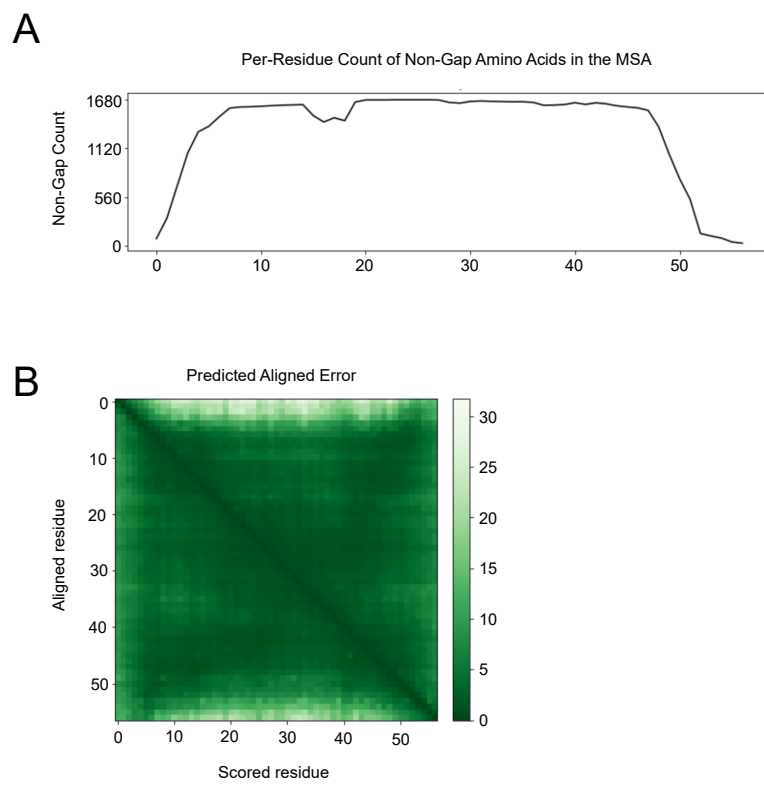
B



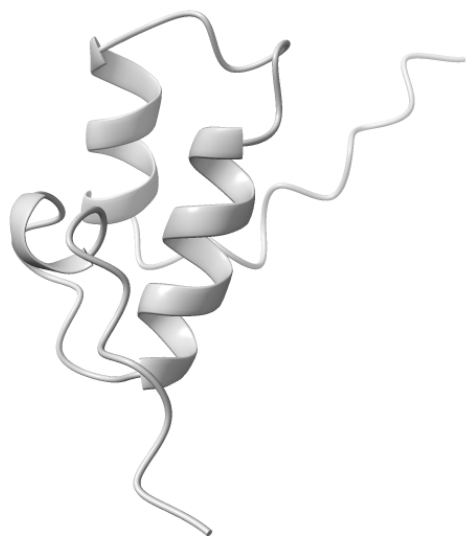
S4 Fig



S5 Fig

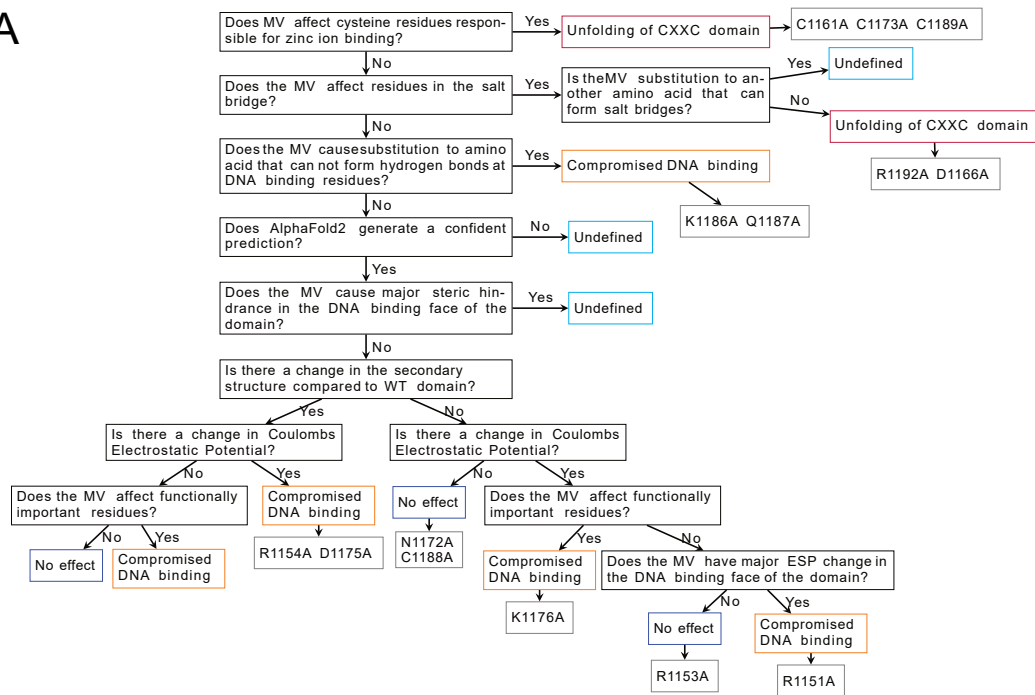


S6 Fig

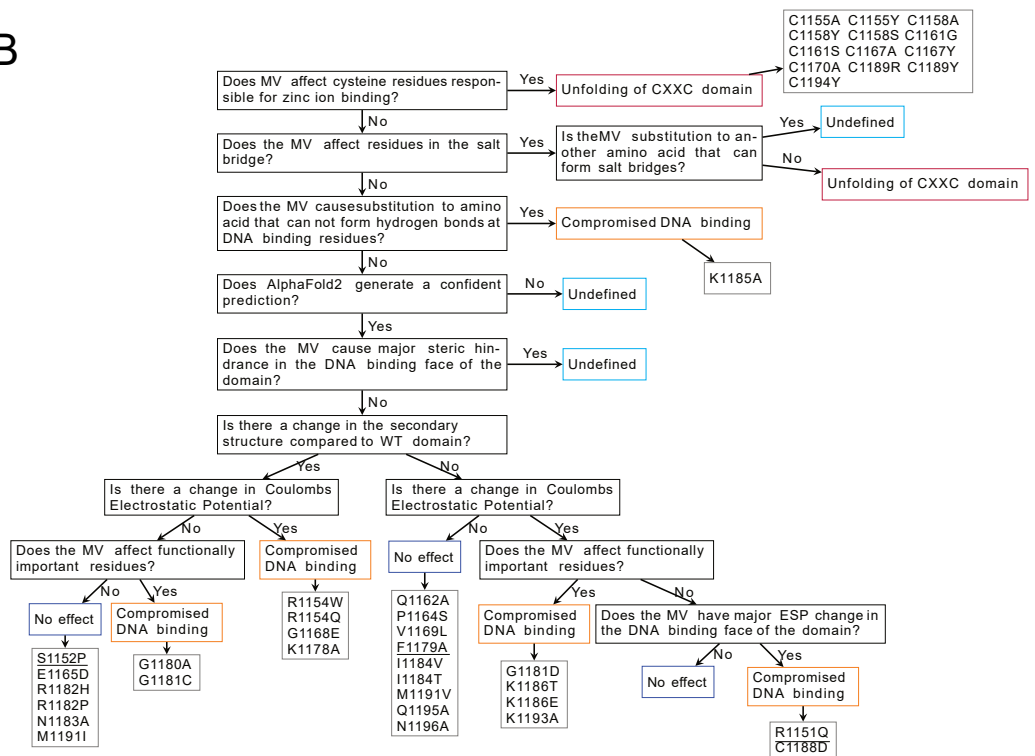


S7 Fig

A



B



C

		Predicted		
		benign	damaging	total
True	benign	25	1	26
	damaging	2	13	15
	total	27	14	

S1 Table

Variant	Nucleotide change (NM_001197104.2)	Domain	Source	Annotation (Clinvar/HGMD)	Phred-like CADD score
P44T	c.133C>T	No domain	Clinvar	Uncertain Significance	21.4
S56Y	c.167C>A	No domain	Clinvar	Uncertain Significance	22
R164K	c.491G>A	No domain	Clinvar	Uncertain Significance	22.1
K232E	c.694A>G	No domain	Clinvar	Uncertain Significance	23.6
P280T	c.838C>A	No domain	Miyake et al 2016		24.4
V413I	c.1237G>A	No domain	HGMD(Faundes et al 2017)	Uncertain Significance	25.4
R478Q	c.1433G>A	No domain	Clinvar	Uncertain Significance	27.9
V776A	c.2327T>C	No domain	Clinvar	Uncertain Significance	21.1
S802C	c.2405C>G	No domain	Clinvar	Uncertain Significance	25
S873N	c.2618G>A	No domain	Clinvar	Uncertain Significance	21.1
R878Q	c.2633G>A	No domain	HGMD(Fokstuen et al 2016)	likely benign/Uncertain significance	29
S996P	c.2986T>C	No domain	Clinvar	Uncertain Significance	20.2
G1007C	c.3019G>T	No domain	Clinvar	Conflicting interpretations of pathogenicity	31
A1032S	c.3094G>T	No domain	Clinvar	Uncertain Significance	31
R1083Q	c.3248G>A	No domain	HGMD(Monies et al 2017)		31
L1113R	c.3338A>G	No domain	HGMD(Faundes et al 2017)	likely pathogenic	31
N1138S	c.3413A>G	No domain	Clinvar	Uncertain Significance	25.8
S1152P	c.3454T>C	CXXC	HGMD(Faundes et al 2017)	Uncertain Significance	26.9
R1154W	c.3460C>T	CXXC	Clinvar	Pathogenic/likely pathogenic	23.8
R1154Q	c.3461G>A	CXXC	Clinvar	Pathogenic/likely pathogenic	29.8
C1155Y	c.3464G>A	CXXC	Clinvar	Pathogenic/likely pathogenic	31
C1158Y	c.3473G>A	CXXC	Clinvar	Likely pathogenic	29.5
C1158S	c.3473G>C	CXXC	Clinvar	Pathogenic	28.8
C1161G	c.3481T>G	CXXC	Lebrun N et al 2018		29.0
C1161S	c.3482G>C	CXXC	Jones WD 2017		28.6
C1167Y	c.3500G>A	CXXC	Jones WD 2017		29.4
G1168D	c.3504G>A	CXXC	Lebrun N et al 2018		28.9
G1180A	c.3539G>C	CXXC	HGMD(Dillon et al 2018)		27.5
G1181D	c.3542G>A	CXXC	Baer S et al 2018		28.6
G1181C	c.3541G>T	CXXC	Clinvar	Likely pathogenic	31
K1186E	c.3556A>G	CXXC	Jones WD 2017		29.7
K1186T	c.3557A>C	CXXC	Clinvar	Pathogenic	29.2
C1189R	c.3565T>C	CXXC	Clinvar	Pathogenic	31
C1189Y	c.3566G>A	CXXC	Clinvar	Pathogenic	29.3
C1194Y	c.3581G>A	CXXC	Clinvar	Uncertain Significance	30

S1389C	c.4166C>G	No domain	HGMD(Liu et al 2017)		25.7
C1448R	c.4342T>C	PHD finger 1	Clinvar	Likely pathogenic	29.7
C1448Y	c.4343G>A	PHD finger 1	Clinvar	Pathogenic	29.9
H1456R	c.4367A>G	PHD finger 1	Clinvar	Likely pathogenic	26
W1474R	c.4420T>A	PHD finger 1	Clinvar	Uncertain Significance	31
C1476S	c.4426T>A	PHD finger 1	Clinvar	Uncertain Significance	27.9
R1636Q	c.4907G>A	Bromodomain	HGMD(Faundes et al 2017)	Uncertain Significance	29.5
S1657A	c.4969T>G	Bromodomain	Clinvar	Uncertain Significance	26.1
V1726M	c.5176G>A	Bromodomain	Clinvar	Uncertain Significance	32
P1832T	c.5494C>A	No domain	Clinvar	Likely pathogenic	22.6
S1861I	c.5582G>T	No domain	Clinvar	Uncertain Significance	26.6
C1879R	c.5635T>C	No domain	Clinvar	Uncertain Significance	26.8
G1891V	c.5672G>T	No domain	Jones WD 2017	likely pathogenic	28.2
H1903R	c.5672G>T	No domain	Jones WD 2017		25.8
D1917Y	c.5749G>T	No domain	Jones WD 2017		29.3
G1943E	c.5828G>A	PHD finger 4	Clinvar	Likely pathogenic	28.3
H1958R	c.5873A>G	PHD finger 4	Baer S et al 2018		26.7
F1959C	c.5876T>G	PHD finger 4	Clinvar	Uncertain Significance	31
G2027E	c.6080G>A	FYR-N	Clinvar	Likely pathogenic	33
F2048V	c.6142T>G	FYR-N	Clinvar	Uncertain Significance	25.9
S2188C	c.6563C>G	No Domain	Clinvar	Uncertain Significance	26
E2419K	c.7255G>A	No domain	Lebrun N et al 2018		25.4
C2844F	c.8531G>T	No domain	HGMD(Lelieveld et al 2016)		27.7
L2848P	c.8543T>C	No domain	Clinvar	Likely pathogenic	28.2
M2853R	c.8558T>G	TAD	Lebrun N et al 2018		26.8
N2859K	c.8577T>A	No domain	Jones WD 2017		22.5
G3131S	c.9382G>A	No domain	HGMD(Liu et al 2017)	benign/likely benign	24.1
S3147F	c.9440C>T	No domain	Baer S et al 2018		23.8
Q3601R	c.10802A>G	No domain	Clinvar	Uncertain Significance	20.6
L3617P	c.10850T>C	No domain	Baer S et al 2018		23.5
R3824W	c.9382G>A	SET domain	HGMD(Faundes et al 2017)	Uncertain significance	32
G3863S	c.11587G>A	SET domain	Clinvar	Uncertain Significance	29.8
I3929L	c.11785A>C	SET domain	Clinvar	Likely pathogenic	27.6

S2 Table

Functionally important residues in the CXXC domain of KMT2A

Residues implicated in electrostatic interaction with the DNA	Arg1150, Arg1154, Lys1176, Lys1178, , Lys1190, Arg1192, Lys1193
Residues responsible for hydrogen bond formation with the DNA	Lys1185, Lys1186, Gln1187
Residues at the KFGG site	Lys1178, Phe1179, Gly1180, Gly1181

S3 Table

References

1. Allen, M.D., et al., Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. The EMBO Journal, 2006. 25(19): p. 4503-4512.
2. Cierpicki, T., et al., Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nature structural & molecular biology, 2010. 17(1): p. 62-68.
3. Lebrun, N., et al., Molecular and cellular issues of KMT2A variants involved in Wiedemann-Steiner syndrome. Eur J Hum Genet, 2018. 26(1): p. 107-116.
4. Jones, W.D., Genetic and phenotypic investigations into developmental disorders, in Wellcome Trust Sanger Institute. 2017, University of Cambridge: Newnham College.
5. Baer, S., et al., Wiedemann-Steiner syndrome as a major cause of syndromic intellectual disability: A study of 33 French cases. Clin Genet, 2018. 94(1): p. 141-152.

Set	Variant (ENSP00000436786.1)	Source	Determined effect	Variant classification result	Positioned at a functionally important residue	Change in secondary structure	Mean electrostatic potential
Trainig set	p.R1151A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	no	no	5.77
Trainig set	p.R1154A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	yes	yes (loss of beta sheets)	5.66
Trainig set	p.C1161A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		
Trainig set	p.D1166A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		
Trainig set	p.C1173A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		
Trainig set	p.D1175A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	no	yes (loss of alpha helix)	6.84
Trainig set	p.K1176A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	yes	no	5.69
Trainig set	p.K1186A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	yes	yes (loss of alpha helix)	5.88
Trainig set	p.K1187A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	yes	yes (loss of alpha helix)	6.33
Trainig set	p.C1189A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		

Trainig set	p.R1192A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		
Trainig set	p.N1172A	Allen et al., 2006 [1]	No effect	No effect	no	no	6.31
Trainig set	p.C1188A	Allen et al., 2006 [1]	No effect	No effect	no	no	6.33
Trainig set	p.R1153A	Allen et al., 2006 [1]	No effect	No effect	no	no	5.73
Test set	p.C1155A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		
Test set	p.C1158A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		
Test set	p.C1167A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		
Test set	p.C1170A	Allen et al., 2006 [1]	Unfolding of CXXC domain	Unfolding of CXXC domain	no		
Test set	p.K1178A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	yes	yes (loss of alpha helix)	5.74
Test set	p.F1179A	Allen et al., 2006 [1]	Experimentally determined	No effect	yes	no	6.38
Test set	p.K1185A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	yes	no	5.78
Test set	p.C1188D	Cierpicki T et al., 2010 [2]	Experimentally determined	Compromised DNA binding	no	no	5.60
Test set	p.K1193A	Allen et al., 2006 [1]	Experimentally determined	Compromised DNA binding	yes	no	5.77
Test set	p.S1152P	HGMD	Not precisely determined: WDSTS patient	Compromised DNA binding	no	yes (extended beta sheets)	6.35
Test set	p.R1154W	Clinvar	Not precisely determined: WDSTS patient	Compromised DNA binding	yes	yes (loss of beta sheets)	5.70
Test set	p.R1154Q	Clinvar	Not precisely determined: WDSTS patient	Compromised DNA binding	yes	yes (loss of beta sheets)	5.70
Test set	p.C1155Y	Clinvar	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.C1158Y	Clinvar	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.C1158S	Clinvar	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.C1161G	Lebrun N et al., 2018 [3]	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.C1161S	Jones WD, 2017 [4]	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.C1167Y	Jones WD, 2017 [4]	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.G1168D	Lebrun N et al., 2018 [3]	Not precisely determined: WDSTS patient	Compromised DNA binding	no	yes (loss of alpha helix)	5.79
Test set	p.G1180A	HGMD	Not precisely determined: WDSTS patient	Compromised DNA binding	yes	yes (loss of alpha helix)	6.28
Test set	p.G1181D	Baer et al., 2018 [5]	Not precisely determined: WDSTS patient	Compromised DNA binding	yes	no	5.66

Test set	p.G1181C	Clinvar	Not precisely determined: WDSTS patient	Compromised DNA binding	yes	yes (loss of beta sheets)	6.30
Test set	p.K1186E	Jones WD, 2017 [4]	Not precisely determined: WDSTS patient	Compromised DNA binding	yes	no	5.37
Test set	p.K1186T	Clinvar	Not precisely determined: WDSTS patient	Compromised DNA binding	yes	no	5.86
Test set	p.C1189R	Clinvar	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.C1189Y	Clinvar	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.C1194Y	Clinvar	Not precisely determined: WDSTS patient	Unfolding of CXXC domain	no		
Test set	p.N1196A	Allen et al., 2006 [1]	No effect	No effect	no	no	6.30
Test set	p.Q1162A	Allen et al., 2006 [1]	No effect	No effect	no	no	6.33
Test set	p.N1183A	Allen et al., 2006 [1]	No effect	No effect	no	yes (loss of alpha helix)	6.24
Test set	p.Q1195A	Allen et al., 2006 [1]	No effect	No effect	no	no	6.35
Test set	p.P1164S	gnomAD	No effect (Healthy population)	No effect	no	no	6.26
Test set	p.V1169L	gnomAD	No effect (Healthy population)	No effect	no	no	6.27
Test set	p.R1182H	gnomAD	No effect (Healthy population)	No effect	no	no	5.95
Test set	p.R1182P	gnomAD	No effect (Healthy population)	No effect	no	no	5.94
Test set	p.I1184V	gnomAD	No effect (Healthy population)	No effect	no	no	6.32
Test set	p.I1184T	gnomAD	No effect (Healthy population)	No effect	no	no	6.31
Test set	p.M1191V	gnomAD	No effect (Healthy population)	No effect	no	no	6.35
Test set	p.R1151Q	TOPmed	No effect (Healthy population)	Compromised DNA binding	no	no	5.8
Test set	p.E1165D	TOPmed	No effect (Healthy population)	No effect	no	yes (loss of alpha helix)	6.34
Test set	p.M1191I	TOPmed	No effect (Healthy population)	No effect	no	yes (loss of alpha helix)	6.32

S4 Table

hg38 coordinates	Nucleotide (ENST00000 534358.8)	amino acid position (ENSP00000 436786.1)	A	C	G	T
118478077	c.3445G	p.G1149	Compromised DNA binding	Compromised DNA binding	Excisting nucleotide	Stop gain
118478078	c.3446G	p.G1149	Compromised DNA binding	No effect	Excisting nucleotide	No effect
118478079	c.3447A	p.G1149	Excisting nucleotide	Synonymous mutation	Synonymous mutation	Synonymous mutation
118478080	c.3448C	p.R1150	Compromised DNA binding	Excisting nucleotide	Compromised DNA binding	Compromised DNA binding
118478081	c.3449G	p.R1150	Compromised DNA binding	Compromised DNA binding	Excisting nucleotide	Compromised DNA binding
118478082	c.3450T	p.R1150	Synonymous mutation	Synonymous mutation	Synonymous mutation	Excisting nucleotide
118478083	c.3451C	p.R1151	Synonymous mutation	Excisting nucleotide	Compromised DNA binding	Stop gain
118478084	c.3452G	p.R1151	Compromised DNA binding	Undefined	Excisting nucleotide	Compromised DNA binding
118478085	c.3453A	p.R1151	Excisting nucleotide	Synonymous mutation	Synonymous mutation	Synonymous mutation
118478086	c.3454T	p.S1152	No effect	No effect	No effect	Excisting nucleotide
118478087	c.3455C	p.S1152	Stop gain	Excisting nucleotide	No effect	No effect
118478088	c.3456G	p.S1152	Synonymous mutation	Synonymous mutation	Excisting nucleotide	Synonymous mutation
118478089	c.3457A	p.R1153	Excisting nucleotide	Synonymous mutation	Compromised DNA binding	Compromised DNA binding
118478090	c.3458G	p.R1153	No effect	No effect	Excisting nucleotide	No effect
118478091	c.3459G	p.R1153	Synonymous mutation	No effect	Excisting nucleotide	No effect
118478092	c.3460C	p.R1154	Synonymous mutation	Excisting nucleotide	Compromised DNA binding	Compromised DNA binding
118478093	c.3461G	p.R1154	Compromised DNA binding	Compromised DNA binding	Excisting nucleotide	Compromised DNA binding
118478094	c.3462G	p.R1154	Synonymous mutation	Synonymous mutation	Excisting nucleotide	Synonymous mutation
118478095	c.3463T	p.C1155	Unfolding	Unfolding	Unfolding	Excisting nucleotide
118478096	c.3464G	p.C1155	Unfolding	Unfolding	Excisting nucleotide	Unfolding
118478097	c.3465T	p.C1155	Stop gain	Synonymous mutation	Unfolding	Excisting nucleotide
118478098	c.3466G	p.G1156	No effect	No effect	Excisting nucleotide	No effect
118478099	c.3467G	p.G1156	No effect	No effect	Excisting nucleotide	No effect
118478100	c.3468G	p.G1156	Synonymous mutation	Synonymous mutation	Excisting nucleotide	Synonymous mutation
118478101	c.3469C	p.Q1157	Compromised DNA binding	Excisting nucleotide	No effect	Stop gain
118478102	c.3470A	p.Q1157	Excisting nucleotide	No effect	No effect	No effect
118478103	c.3471G	p.Q1157	Synonymous mutation	No effect	Excisting nucleotide	No effect
118478104	c.3472T	p.C1158	Unfolding	Unfolding	Unfolding	Excisting nucleotide

118478105	c.3473G	p.C1158	Unfolding	Unfolding	Existing nucleotide	Unfolding
118478106	c.3474T	p.C1158	Stop gain	Synonymous mutation	Unfolding	Existing nucleotide
118478107	c.3475C	p.P1159	No effect	Existing nucleotide	No effect	No effect
118478108	c.3476C	p.P1159	No effect	Existing nucleotide	No effect	No effect
118478109	c.3477C	p.P1159	Synonymous mutation	Existing nucleotide	Synonymous mutation	Synonymous mutation
118478110	c.3478G	p.G1160	No effect	No effect	Existing nucleotide	No effect
118478111	c.3479G	p.G1160	No effect	No effect	Existing nucleotide	No effect
118478112	c.3480C	p.G1160	Synonymous mutation	Existing nucleotide	Synonymous mutation	Synonymous mutation
118478113	c.3481T	p.C1161	Unfolding	Unfolding	Unfolding	Existing nucleotide
118478114	c.3482G	p.C1161	Unfolding	Unfolding	Existing nucleotide	Unfolding
118478115	c.3483C	p.C1161	Stop gain	Existing nucleotide	Unfolding	Synonymous mutation
118478116	c.3484C	p.Q1162	No effect	Existing nucleotide	No effect	Stop gain
118478117	c.3485A	p.Q1162	Existing nucleotide	No effect	No effect	No effect
118478118	c.3486G	p.Q1162	Synonymous mutation	No effect	Existing nucleotide	No effect
118478119	c.3487G	p.V1163	No effect	No effect	Existing nucleotide	No effect
118478120	c.3488T	p.V1163	Compromised DNA binding	No effect	No effect	Existing nucleotide
118478121	c.3489G	p.V1163	Synonymous mutation	Synonymous mutation	Existing nucleotide	Synonymous mutation
118478122	c.3490C	p.P1164	No effect	Existing nucleotide	No effect	No effect
118478123	c.3491C	p.P1164	No effect	Existing nucleotide	No effect	No effect
118478124	c.3492T	p.P1164	Synonymous mutation	Synonymous mutation	Synonymous mutation	Existing nucleotide
118478125	c.3493G	p.E1165	No effect	Compromised DNA binding	Existing nucleotide	Stop gain
118478126	c.3494A	p.E1165	Existing nucleotide	Compromised DNA binding	No effect	Compromised DNA binding
118478127	c.3495G	p.E1165	Synonymous mutation	No effect	Existing nucleotide	No effect
118478128	c.3496G	p.D1166	Unfolding	Unfolding	Existing nucleotide	Unfolding
118478129	c.3497A	p.D1166	Existing nucleotide	Unfolding	Unfolding	Unfolding
118478130	c.3498C	p.D1166	Undefined	Existing nucleotide	Undefined	Synonymous mutation
118478131	c.3499T	p.C1167	Unfolding	Unfolding	Unfolding	Existing nucleotide
118478132	c.3500G	p.C1167	Unfolding	Unfolding	Existing nucleotide	Unfolding
118478133	c.3501T	p.C1167	Stop gain	Synonymous mutation	Unfolding	Existing nucleotide
118478134	c.3502G	p.G1168	No effect	No effect	Existing nucleotide	No effect
118478135	c.3503G	p.G1168	Compromised DNA binding	No effect	Existing nucleotide	No effect

118478136	c.3504T	p.G1168	Synonymous mutation	Synonymous mutation	Synonymous mutation	Existing nucleotide
118478137	c.3505G	p.V1169	No effect	No effect	Existing nucleotide	No effect
118478138	c.3506T	p.V1169	No effect	No effect	No effect	Existing nucleotide
118478139	c.3507T	p.V1169	Synonymous mutation	Synonymous mutation	Synonymous mutation	Existing nucleotide
118478140	c.3508T	p.C1170	Unfolding	Unfolding	Unfolding	Existing nucleotide
118478141	c.3509G	p.C1170	Unfolding	Unfolding	Existing nucleotide	Unfolding
118478142	c.3510T	p.C1170	Stop gain	Synonymous mutation	Unfolding	Existing nucleotide
118478143	c.3511A	p.T1171	Existing nucleotide	No effect	No effect	No effect
118478144	c.3512C	p.T1171	No effect	Existing nucleotide	No effect	No effect
118478145	c.3513T	p.T1171	Synonymous mutation	Synonymous mutation	Synonymous mutation	Existing nucleotide
118478146	c.3514A	p.N1172	Existing nucleotide	No effect	No effect	No effect
118478147	c.3515A	p.N1172	Existing nucleotide	No effect	No effect	No effect
118478148	c.3516T	p.N1172	Compromised DNA binding	Synonymous mutation	Compromised DNA binding	Existing nucleotide
118478149	c.3517T	p.C1173	Unfolding	Unfolding	Unfolding	Existing nucleotide
118478150	c.3518G	p.C1173	Unfolding	Unfolding	Existing nucleotide	Unfolding
118478151	c.3519C	p.C1173	Stop gain	Existing nucleotide	Unfolding	Synonymous mutation
118478152	c.3520T	p.L1174	No effect	Synonymous mutation	No effect	Existing nucleotide
118478153	c.3521T	p.L1174	Stop gain	No effect	Stop gain	Existing nucleotide
118478154	c.3522A	p.L1174	Existing nucleotide	No effect	Synonymous mutation	No effect
118478155	c.3523G	p.D1175	Compromised DNA binding	Compromised DNA binding	Existing nucleotide	Compromised DNA binding
118478156	c.3524A	p.D1175	Existing nucleotide	Compromised DNA binding	No effect	Compromised DNA binding
118478157	c.3525T	p.D1175	No effect	Synonymous mutation	No effect	Existing nucleotide
118478158	c.3526A	p.K1176	Existing nucleotide	Compromised DNA binding	Compromised DNA binding	Stop gain
118478159	c.3527A	p.K1176	Existing nucleotide	Compromised DNA binding	No effect	Compromised DNA binding
118478160	c.3528G	p.K1176	Synonymous mutation	Compromised DNA binding	Existing nucleotide	Compromised DNA binding
118478161	c.3529C	p.P1177	No effect	Existing nucleotide	No effect	No effect
118478162	c.3530C	p.P1177	No effect	Existing nucleotide	No effect	No effect
118478163	c.3531C	p.P1177	Synonymous mutation	Existing nucleotide	Synonymous mutation	Synonymous mutation
118478164	c.3532A	p.K1178	Existing nucleotide	Compromised DNA binding	Compromised DNA binding	Stop gain
118478165	c.3533A	p.K1178	Existing nucleotide	Compromised DNA binding	No effect	Compromised DNA binding
118478166	c.3534G	p.K1178	Synonymous mutation	Compromised DNA binding	Existing nucleotide	Compromised DNA binding

118478167	c.3535T	p.F1179	No effect	No effect	No effect	Existing nucleotide
118478168	c.3536T	p.F1179	No effect	No effect	No effect	Existing nucleotide
118478169	c.3537T	p.F1179	No effect	Synonymous mutation	No effect	Existing nucleotide
118478170	c.3538G	p.G1180	No effect	Compromised DNA binding	Existing nucleotide	Compromised DNA binding
118478171	c.3539G	p.G1180	Compromised DNA binding	Compromised DNA binding	Existing nucleotide	No effect
118478172	c.3540T	p.G1180	Synonymous mutation	Synonymous mutation	Synonymous mutation	Existing nucleotide
118478173	c.3541G	p.G1181	No effect	Compromised DNA binding	Existing nucleotide	Compromised DNA binding
118478174	c.3542G	p.G1181	Compromised DNA binding	Compromised DNA binding	Existing nucleotide	Compromised DNA binding
118478175	c.3543T	p.G1181	Synonymous mutation	Synonymous mutation	Synonymous mutation	Existing nucleotide
118478176	c.3544C	p.R1182	No effect	Existing nucleotide	No effect	No effect
118478177	c.3545G	p.R1182	No effect	No effect	Existing nucleotide	No effect
118478178	c.3546C	p.R1182	Synonymous mutation	Existing nucleotide	Synonymous mutation	Synonymous mutation
118478179	c.3547A	p.N1183	Existing nucleotide	No effect	No effect	No effect
118478180	c.3548A	p.N1183	Existing nucleotide	No effect	No effect	No effect
118478181	c.3549T	p.N1183	Compromised DNA binding	Synonymous mutation	Compromised DNA binding	Existing nucleotide
118478182	c.3550A	p.I1184	Existing nucleotide	No effect	No effect	No effect
118478183	c.3551T	p.I1184	Compromised DNA binding	No effect	Compromised DNA binding	Existing nucleotide
118478184	c.3552A	p.I1184	Existing nucleotide	Synonymous mutation	No effect	Synonymous mutation
118478185	c.3553A	p.K1185	Existing nucleotide	Compromised DNA binding	Compromised DNA binding	Stop gain
118478186	c.3554A	p.K1185	Existing nucleotide	Compromised DNA binding	No effect	Compromised DNA binding
118478187	c.3555G	p.K1185	Synonymous mutation	Compromised DNA binding	Existing nucleotide	Compromised DNA binding
118478188	c.3556A	p.K1186	Existing nucleotide	Compromised DNA binding	Compromised DNA binding	Stop gain
118478189	c.3557A	p.K1186	Existing nucleotide	Compromised DNA binding	No effect	Compromised DNA binding
118478190	c.3558G	p.K1186	Synonymous mutation	Compromised DNA binding	Existing nucleotide	Compromised DNA binding
118478191	c.3559C	p.Q1187	Compromised DNA binding	Existing nucleotide	Compromised DNA binding	Stop gain
118478192	c.3560A	p.Q1187	Existing nucleotide	Compromised DNA binding	Compromised DNA binding	Compromised DNA binding
118478193	c.3561G	p.Q1187	Synonymous mutation	No effect	Existing nucleotide	No effect
118478194	c.3562T	p.C1188	No effect	No effect	No effect	Existing nucleotide
118478195	c.3563G	p.C1188	No effect	No effect	Existing nucleotide	No effect
118478196	c.3564C	p.C1188	Stop gain	Existing nucleotide	No effect	Synonymous mutation
118478197	c.3565T	p.C1189	Unfolding	Unfolding	Unfolding	Existing nucleotide

118478198	c.3566G	p.C1189	Unfolding	Unfolding	Existing nucleotide	Unfolding
118478199	c.3567C	p.C1189	Stop gain	Existing nucleotide	Unfolding	Synonymous mutation
118478200	c.3568A	p.K1190	Existing nucleotide	Compromised DNA binding	Compromised DNA binding	Stop gain
118478201	c.3569A	p.K1190	Existing nucleotide	Compromised DNA binding	No effect	Compromised DNA binding
118478202	c.3570G	p.K1190	Synonymous mutation	Compromised DNA binding	Existing nucleotide	Compromised DNA binding
118478203	c.3571A	p.M1191	Existing nucleotide	No effect	No effect	No effect
118478204	c.3572T	p.M1191	No effect	No effect	No effect	Existing nucleotide
118478205	c.3573G	p.M1191	No effect	No effect	Existing nucleotide	No effect
118478206	c.3574A	p.R1192	Existing nucleotide	Synonymous mutation	Unfolding	Stop gain
118478207	c.3575G	p.R1192	Undefined	Unfolding	Existing nucleotide	Unfolding
118478208	c.3576A	p.R1192	Existing nucleotide	Unfolding	Synonymous mutation	Unfolding
118478209	c.3577A	p.K1193	Existing nucleotide	Compromised DNA binding	Compromised DNA binding	Stop gain
118478210	c.3578A	p.K1193	Existing nucleotide	Compromised DNA binding	No effect	Compromised DNA binding
118478211	c.3579A	p.K1193	Existing nucleotide	Compromised DNA binding	Synonymous mutation	Compromised DNA binding
118478212	c.3580T	p.C1194	Unfolding	Unfolding	Unfolding	Existing nucleotide
118478213	c.3581G	p.C1194	Unfolding	Unfolding	Existing nucleotide	Unfolding
118478214	c.3582T	p.C1194	Stop gain	Synonymous mutation	Unfolding	Existing nucleotide
118478215	c.3583C	p.Q1195	Compromised DNA binding	Existing nucleotide	No effect	Stop gain
118478216	c.3584A	p.Q1195	Existing nucleotide	No effect	No effect	No effect
118478217	c.3585G	p.Q1195	Synonymous mutation	No effect	Existing nucleotide	No effect
118478218	c.3586A	p.N1196	Existing nucleotide	No effect	No effect	No effect
118478219	c.3587A	p.N1196	Existing nucleotide	No effect	No effect	No effect
118478220	c.3588T	p.N1196	No effect	Synonymous mutation	No effect	Existing nucleotide
118478221	c.3589C	p.L1197	No effect	Existing nucleotide	No effect	Synonymous mutation
118478222	c.3590T	p.L1197	No effect	No effect	Compromised DNA binding	Existing nucleotide
118478223	c.3591A	p.L1197	Existing nucleotide	Synonymous mutation	Synonymous mutation	Synonymous mutation
118478224	c.3592C	p.Q1198	Compromised DNA binding	Existing nucleotide	No effect	Stop gain
118478225	c.3593A	p.Q1198	Existing nucleotide	Undefined	Compromised DNA binding	No effect
118478226	c.3594A	p.Q1198	Existing nucleotide	No effect	Synonymous mutation	No effect

S5 Table

Domain	Amino acid co-ordinates ENSP00000436786.1
AT hook 1	169-180
AT hook 2	217-227
AT hook 3	301-309
CXXC Domain	1147-1203
PHD finger 1	1433-1480
PHD finger 2	1481-1531
PHD finger 3	1568-1625
Bromodomain	1636-1770
PHD finger 4	1935-1981
FYR-N Domain	2021-2077
TAD Domain	2850-2858
FYR-C Domain	3669-3756
Win Motif	3765-3773
SET Domain	3832-3954
Post-SET Domain	3956-3972

Paper II

Paper II

In-utero rescue of neurological dysfunction in a mouse model of Wiedemann-Steiner syndrome

Authors

Tinna Reynisdottir¹, Kimberley Jade Anderson^{1,2}, Andrew Brinn³, Katie Franklin³, Juan Ouyang¹, Asbjorg Osk Snorraddottir^{4,5}, Cathleen M. Lutz⁶, Aamir R. Zuberi⁶, Valerie Burke DeLeon³, Hans Tomas Bjornsson^{1,2,7*}

Institutes

¹Louma G. Laboratory of Epigenetic Research, Faculty of Medicine, University of Iceland;

²Department of Genetics and Molecular Medicine, Landspítali University Hospital, Reykjavik, Iceland;

³Department of Anthropology, University of Florida;

⁴Faculty of Medicine, University of Iceland, Reykjavik, Iceland;

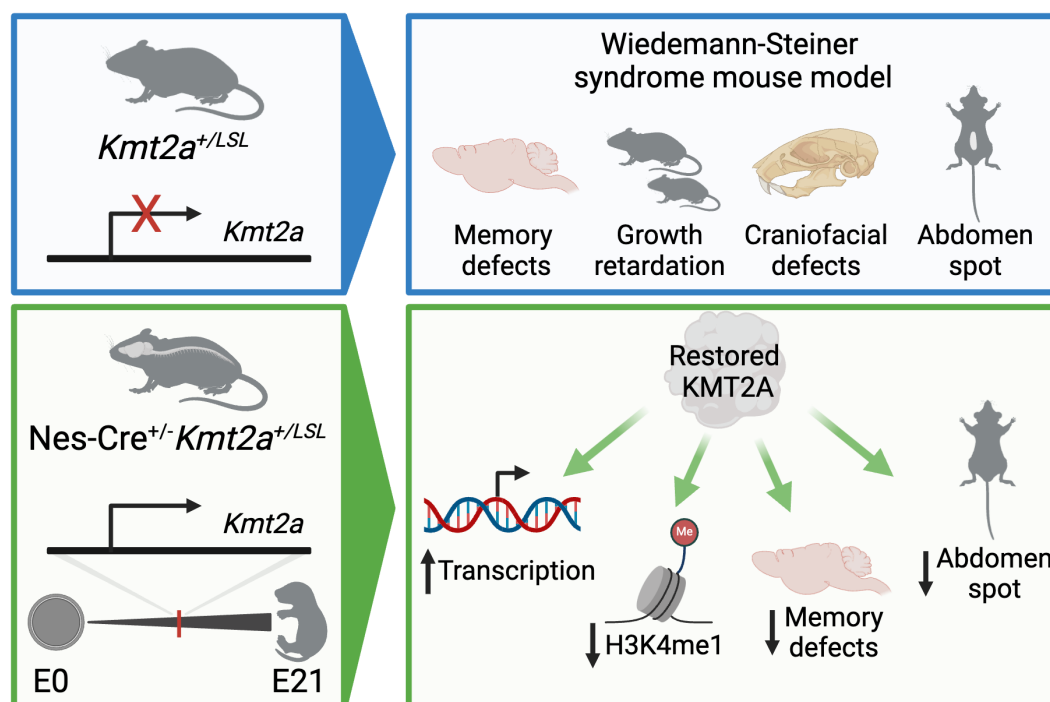
⁵Department of Pathology, Landspítali University Hospital, Reykjavik, Iceland;

⁶The Jackson Laboratory, Bar Harbor;

⁷McKusick-Nathans Departments of Genetic Medicine and Pediatrics, Johns Hopkins University.

*Corresponding author: Hans Tomas Bjornsson (htb@hi.is)

Graphical abstract



Abstract

Wiedemann-Steiner syndrome (WDSTS) is a rare genetic cause of intellectual disability primarily caused by heterozygous loss of function variants in the gene encoding the histone methyltransferase KMT2A. Prior studies have shown successful postnatal amelioration of disease phenotypes for Rett, Rubinstein-Taybi and Kabuki syndromes, related Mendelian disorders of the epigenetic machinery. To explore whether the neurological phenotype in WDSTS is treatable *in-utero*, we created a novel mouse model carrying a loss of function variant in between two loxP sites. *Kmt2a*^{+/*LSL*} mice demonstrate core features of WDSTS including growth retardation, craniofacial abnormalities, and hypertrichosis as well as hippocampal memory defects. The neurological phenotypes show rescue upon restoration of *KMT2A in-utero* following breeding to a nestin-Cre. Together, our data provide a novel mouse model to explore the therapeutic window in WDSTS. Our work suggests that WDSTS has a window of opportunity extending at least until the mid-point of *in-utero* development, making WDSTS an ideal candidate for future therapeutic strategies.

Keywords: *KMT2A, genetic rescue, Nestin-Cre, WDSTS, Intellectual disability*

Introduction

Wiedemann Steiner syndrome (WDSTS, OMIM: #605130) is a Mendelian disorder of the epigenetic machinery (MDEM), inherited in an autosomal dominant manner. WDSTS is primarily caused by *de novo* heterozygous variants in the gene encoding the histone-lysine methyltransferase 2A (KMT2A, previously MLL, MLL-1 or ALL in literature) [1]. KMT2A plays a crucial role in the mono-, di- and trimethylation of lysine 4 of histone 3 (H3K4me1/2/3) [2], modifications generally associated with active gene expression [3], [4], occurring at enhancers and promoters. KMT2A has been shown to be essential for postnatal neurogenesis in mice, and as a regulator of neural progenitor proliferation and differentiation [5], [6]. Key features of WDSTS include intellectual disability, growth retardation, hypotonia, craniofacial defects, gastrointestinal problems, and hypertrichosis. Previous mouse studies have shown postnatal rescue of neurological defects in several other MDEMs, including Rubinstein-Taybi [7], Rett [8] and Kabuki syndromes [9], [10]. To extend these findings to WDSTS, we have created a

novel mouse model for WDSTS ($Kmt2a^{+/LSL}$). Earlier mouse models have demonstrated homozygous knock-out (KO) of $Kmt2a$ ($Kmt2a^{-/-}$ mice) is lethal, while heterozygous mice ($Kmt2a^{+/-}$) exhibited growth retardation, skeletal defects, hemopoietic abnormalities [11] as well as neurological defects in visuospatial- and associative memory [12] and increased aggressive behavior [13]. Notably, it has been shown that joint constitutive KO of the writer-eraser duo KMT2A and KDM5C rescues disease phenotypes seen in either disease model, including rescue of transcriptome abnormalities, neuronal structure and aggressive behavior. This finding suggests that the enzymatic function of KMT2A plays a role in disease and that a specific balance is required among these histone machinery components [13], [14]. In contrast, homozygous mice with a mutation specifically altering the C-terminal enzymatic SET domain remain viable and fertile [15]. In concordance with the latter, our previous work has shown that missense variants causing WDSTS are mainly clustered outside of the enzymatic SET domain, and within the DNA binding CXXC domain [16], suggesting that non-enzymatic roles of KMT2A may be key in WDSTS pathogenesis.

To explore whether WDSTS is treatable *in-utero*, we created and characterized a novel mouse model ($Kmt2a^{+/LSL}$) which carries a heterozygous insertion of a loxP-stop-loxP (LSL) cassette in intron 1 of $Kmt2a$ resulting in heterozygous KO of the affected allele. The model ($Kmt2a^{+/LSL}$) serves as a novel disease model for WDSTS as well as allowing temporal control of $Kmt2a$. By crossing the model to a nestin-Cre model, we demonstrated that *in-utero* rescue of KMT2A levels could rescue WDSTS neurological phenotypes in primary neuronal progenitor cells (NPCs) from these mice and validated this finding by *in vivo* studies.

Results

Targeted knock-out of *Kmt2a* recapitulates key features of WDSTS

WDSTS broadly affects patients with common clinical features within and outside the nervous system. To get an unbiased view of WDSTS phenotypes we have summarized available data from an ongoing RARE-X WDSTS collection. The majority of the patients exhibit neurological issues (93.8%), most commonly intellectual disability and/or developmental delay, with a large part exhibiting behavioral and/or psychiatric problems (75.8%). Other common features include growth retardation (95.4%), dental and/or oral problems (79.4%) and hypotonia (60.0%). Distinctive features of WDSTS also include digestive system issues (76.9, most commonly constipation) and hypertrichosis cubiti (57%) (Table 1).

Clinical feature	WDSTS patients % (n)
Growth retardation	95.4% (62/65)
Brain and/or neurological issues	93.8% (60/64)
Eye and/or vision issues	84.7% (50/59)
Sleep issues	80.3% (53/66)
Dental and/or oral issues	79.4% (50/63)
Digestive system issues	76.9% (50/65)
Behavior and/or psychiatric issues	75.8% (47/62)
Hypotonia	60.0% (33/55)
Hypertrichosis cubiti	57.0% (57/100)
Head, face and/or neck issues	56.3% (36/62)

Table 1: Defining clinical features of WDSTS patients including percentage and count of patients demonstrating the observed phenotype. Data gathered from RARE-X patient database and literature [1] [17].

To elucidate the *in-utero* neurological malleability of WDSTS, we created a novel mouse model (*Kmt2a*^{+/LSL}) with a heterozygous loxP-stop-loxP (LSL) cassette insertion in intron 1 of *Kmt2a* (Figure 1A). The LSL cassette induces an early poly-A signal leading to immediate truncation of the mRNA transcript in intron 1 of the affected allele, allowing rescue of expression with exposure to Cre-recombinase. The mouse model was created via a service contract with Jackson Laboratory who tested functionality of the cassette mice by breeding *Kmt2a*^{+/LSL} mice to a Sox2-Cre containing mouse model which demonstrated that loxP sites were functional (see Methods). To further verify the orientation and completeness of the

cassette, we performed Oxford nanopore technology (ONT) long-read sequencing on genomic DNA (gDNA) from *Kmt2a*^{+/*LSL*} mice, from which single reads revealed the entirety of the cassette integrated into intron 1 of *Kmt2a* in the correct orientation (Figure 1A). Since this experiment provides whole-genome coverage we also interrogated two predicted potential off-target sites from cassette integration and observe no obvious variants or indels at these sites (Supplemental Figure 1A-D). The offspring of *Kmt2a*^{+/*LSL*} mice show skewing of the expected Mendelian ratio with a significant bias against the inheritance of the mutant allele ($p < 0.01$, chi-squared test), with 40.9% (90/220) *Kmt2a*^{+/*LSL*} mice born and 59.1% (130/220) *Kmt2a*^{+/*+*} wild-type littermates. Similar to other *Kmt2a*-deficient models [11], *Kmt2a*^{+/*LSL*} mice demonstrate lethality in homozygosity, with no *Kmt2a*^{*LSL*/*LSL*} mice born (Supplemental Figure 1E-F). As predicted, the *Kmt2a*^{+/*LSL*} mice demonstrate ~50% reduction of *Kmt2a* mRNA transcripts as quantified by real-time quantitative polymerase chain reaction (RT-qPCR) ($n=3-7$, Figure 1B). In addition to observing appropriate integration of the LSL cassette, we observed characteristic WDSTS phenotypes of the *Kmt2a*^{+/*LSL*} mice, supporting the notion that these mice adequately recapitulate the human disease. As previously reported in *Kmt2a*^{+/*-*} mice [11], *Kmt2a*^{+/*LSL*} mice exhibit growth deficiency with reduced weight compared to *Kmt2a*^{+/*+*} littermates (Figure 1C, Supplemental Figure 2A-B), alongside hemopoietic abnormalities similar to what has been previously described [11], including a decreased lymphocyte percentage, increase in monocyte percentages and an increase in platelet distribution width (Supplemental Figure 3). Given the consistent description of hypertrichosis in individuals with WDSTS [17], we decided to explore the skin histology of newborn P1 (postnatal day 1) pups. *Kmt2a*^{+/*LSL*} mice show increased hair follicle count on the abdomen compared to their *Kmt2a*^{+/*+*} littermates, while there was no significant increase in hair follicle count on their back (Figure 1D). Furthermore, we did not observe any obvious differences in hair regrowth upon removal of hair from the backs of the *Kmt2a*^{+/*LSL*} mice compared to their wild-type littermates (Supplemental Figure 2C). This may indicate that the basis of hypertrichosis in WDSTS may involve increased number of hair follicle in contrast to increased rate of hair regrowth. Since severe constipation and other GI symptoms are a common phenotype in WDSTS patients, we stained the distal colon sections of *Kmt2a*^{+/*+*} and *Kmt2a*^{+/*LSL*} mice for Calretinin, a common diagnostic marker for Hirschsprung's disease (HD) that marks ganglion cells [18]. Our results showed positive staining in both genotypes (Supplemental Figure 2D), indicative of normal presence of ganglions and nerve fibers in our WDSTS model.

However, this does not rule out more subtle abnormalities of the enteric nervous system or problems relating to smooth muscle of the intestines.

To evaluate whether craniofacial abnormalities are present, we performed CT scans on the *Kmt2a*^{+/*LSL*} mice (n=6) mice and their *Kmt2a*^{+/*+*} littermates (n=6). A Procrustes ANOVA yielded a significant effect of genotype on shape (Figure 1E, $p < 0.001$). Principal Components Analysis (PCA) showed a clear separation between the groups along PC1, which describes 91.9% of the symmetric shape variance (Figure 1F). The *Kmt2a*^{+/*LSL*} mice are characterized by having increased height and width of the neurocranium and ventral bowing of the cranium compared to littermates (Figure 1G). Additionally, we observed a prominent gap at the interfrontal suture in the *Kmt2a*^{+/*LSL*} mice, which was not observed in any of the *Kmt2a*^{+/*+*} littermates (Figure 1H). Similar craniofacial defects observed in Kabuki syndrome have been linked to neural crest defects [19], [20], [21]. Given the shared characteristics of these syndromes, it raises the question whether neural crest defects could also contribute to abnormalities found in WDSTS. Interestingly, in our model, a majority (57.5%) of the *Kmt2a*^{+/*LSL*} mice present with a white abdominal spot, ranging from few distinct hairs to spots ~1-2 cm in length, most common along the midline (Figure 1I-J, Supplemental Figure 2E), suggestive of defects in the melanoblast migration from the neural crest [22]. In contrast, we have never observed such a white spot in wild-type littermates.

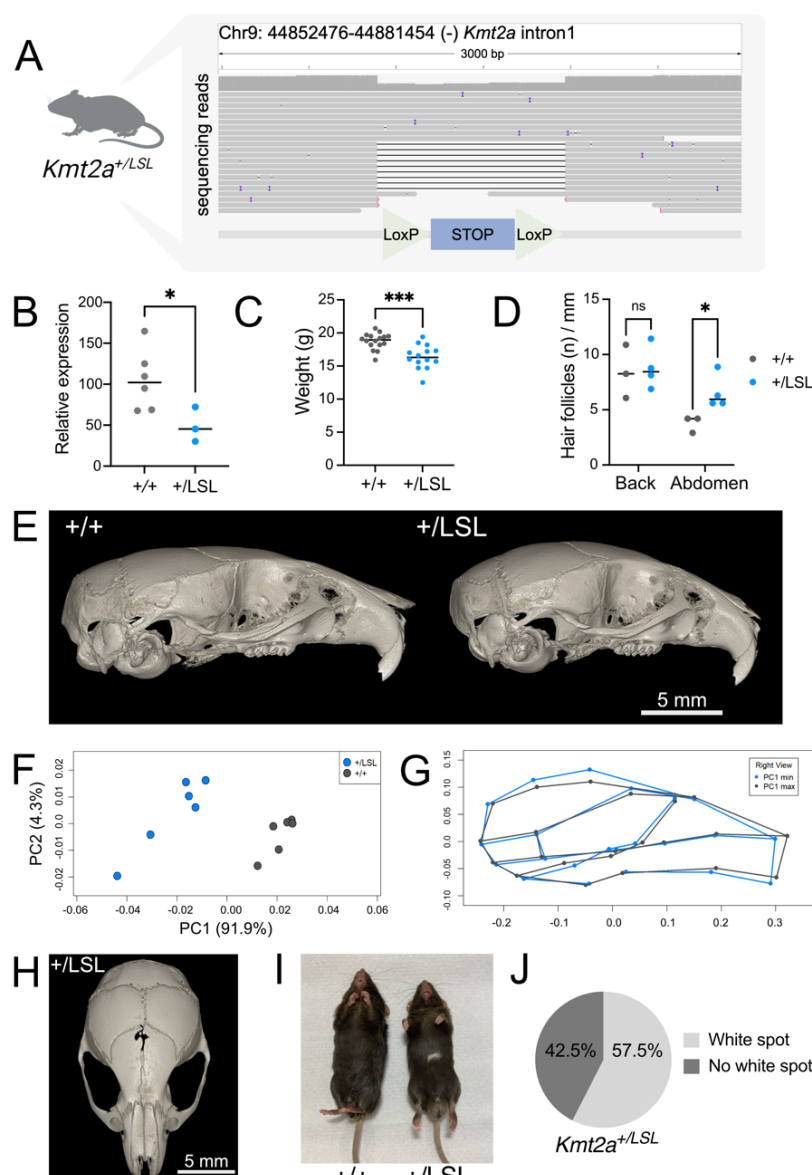


Figure 1: A novel Wiedemann-Steiner syndrome mouse model (*Kmt2a*^{+/LSL}) recapitulates core phenotypes seen in individuals with WDSTS.

(A) Schematic figure of *Kmt2a*^{+/LSL} mouse model with a heterozygous loxP-stop-loxP cassette inserted in intron 1 of *Kmt2a* showing ONT sequencing reads of the cassette. Each line represents an individual long read including several that span the entire cassette as well as boundaries on either side allowing verification of its orientation. (B) Relative total *Kmt2a* mRNA expression is decreased in *Kmt2a*^{+/LSL} mice compared to *Kmt2a*^{+/+} littermates (n=7 *Kmt2a*^{+/+} and n= 3 *Kmt2a*^{+/LSL}). (C) *Kmt2a*^{+/LSL} mice weigh significantly less compared to *Kmt2a*^{+/+} littermates (n=14-16 per genotype, 8-week females). (D) Hair follicle count per mm of skin is decreased on abdomen but not on the back of newborn *Kmt2a*^{+/LSL} pups compared to littermates (n=3-4 per genotype). (E) Representative CT images of *Kmt2a*^{+/+} and *Kmt2a*^{+/LSL} mice from lateral view, the depicting the craniofacial phenotype. (F) PCA plot of the craniofacial structure showing a distinct separation of the genotypes along PC1. (G) The *Kmt2a*^{+/LSL} mice are positioned at the negative end of PC1 and characterized by increased height and width of the neurocranium, a shortened midface, and ventral bowing of the cranium. (H) Representative figure of the gap of the interfrontal suture present in the *Kmt2a*^{+/LSL} mice. (I) Representative figure of the white abdominal spot of a *Kmt2a*^{+/LSL} mouse next to a wild-type littermate. (J) Majority of *Kmt2a*^{+/LSL} mice (57.5%, n=40) present with a white abdominal spot, a phenotype not present in *Kmt2a*^{+/+} littermates. ns: not significant, *p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001.

Observed neurological defects in the *Kmt2a*^{+/*LSL*} mouse model

Consistent with the hypotonia observed in some WDSTS patients, *Kmt2a*^{+/*LSL*} mice exhibit significantly longer (p-value < 0.01, unpaired t-test) surface righting time compared to *Kmt2a*^{+/*+*} littermates, indicating of lack of core strength and coordination (Figure 2A) and a significant decrease (p-value < 0.01, unpaired t-test) in hindlimb suspension time compared to *Kmt2a*^{+/*+*}, indicative of increased fatigue in hindlimbs (Figure 2B), tested on postnatal day 6 (PD6, n=14-27). The visuospatial memory function in adult *Kmt2a*^{+/*LSL*} mice was tested in an open Y maze (Figure 2C). The *Kmt2a*^{+/*LSL*} mice did not exhibit a significant decrease of arms entered during the test time (p-value = 0.1576, unpaired t-test), suggesting the adult mice do not exhibit a gross motor phenotype (Figure 2D). However, the *Kmt2a*^{+/*LSL*} mice demonstrate significant deficits (p-value < 0.01, unpaired t-test) in spontaneous alternations compared to *Kmt2a*^{+/*+*} littermates, indicative of defective working memory reliant on the hippocampus and prefrontal cortex (n=18-24 per genotype, Figure 2E) [23], [24]. These findings are in concordance of previous studies describing the importance of *Kmt2a* for memory consolidation [12]. Histologically, when we measured the size of the granule cell layer of the dentate gyrus (DG) of the hippocampus with DAPI staining, the total area was significantly smaller (p-value < 0.05, unpaired t-test) in the *Kmt2a*^{+/*LSL*} mice, compared to *Kmt2a*^{+/*+*} littermates, when corrected for body weight (n=3-4 per genotype, Figure 2F-H). A decreased size of the dentate gyrus has been previously described in a mouse model of Kabuki syndrome [9] and later validated in individuals with Kabuki syndrome [25]. Taken together, our findings support that the *Kmt2a*^{+/*LSL*} mouse model recapitulates core human WDSTS disease phenotypes and allows for quantitative measurement of a number of metrics related to the neurological phenotype that can be used to investigate the *in-utero* malleability.

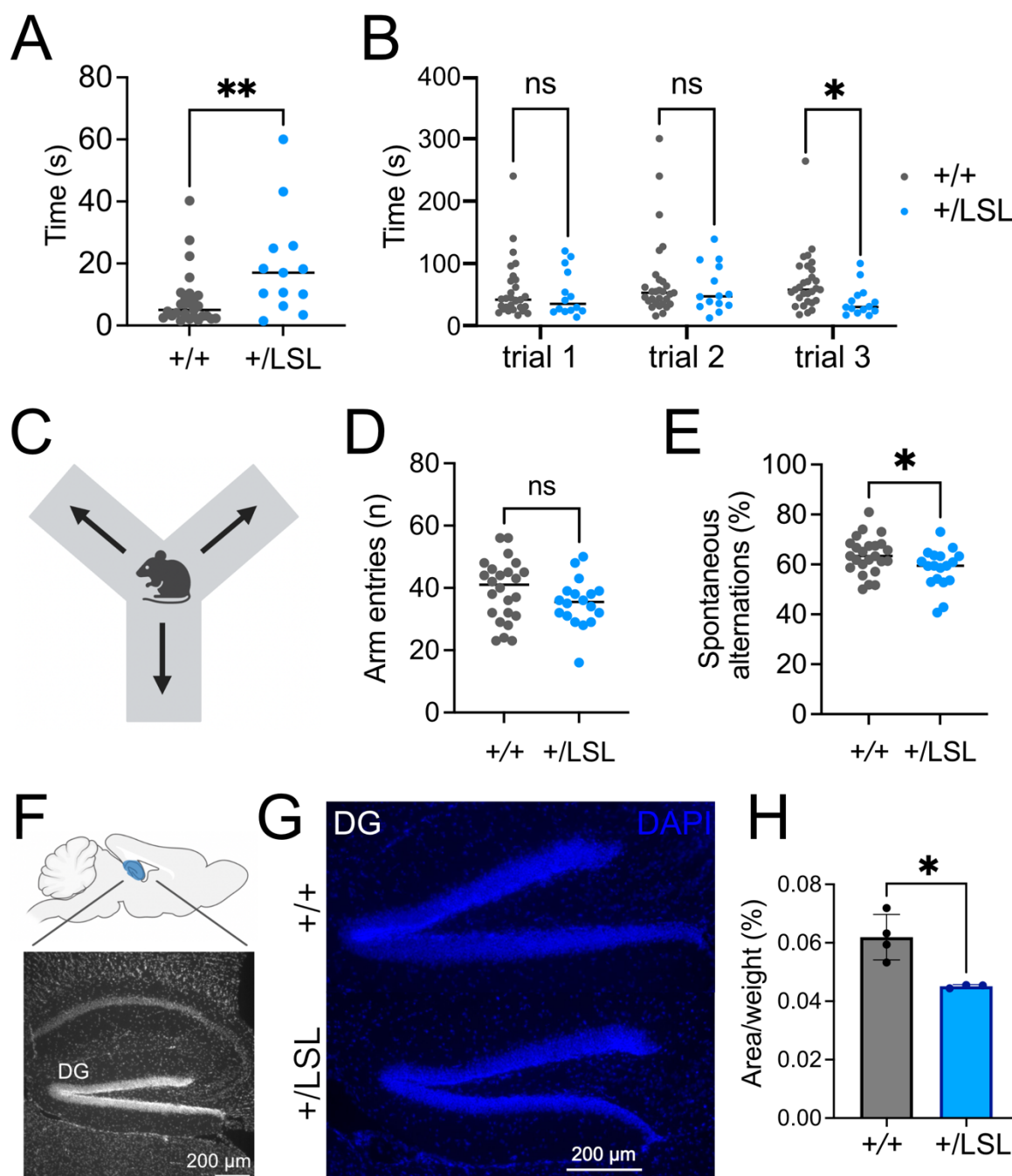


Figure 2: Neurological defects observed in the *Kmt2a*^{+/LSL} mouse model.

(A) Increased surface righting time in *Kmt2a*^{+/LSL} and (B) decreased hindlimb suspension time at the third trial in hypotonia tests (PD6, n= 27-29, *Kmt2a*^{+/+}, n=14-15, *Kmt2a*^{+/LSL}). (C) Schematic overview of a Y maze. A spontaneous alternation is completed when mouse has entered each arm of the maze consecutively. (D) Adult *Kmt2a*^{+/LSL} mice do not show a significant decrease of total arm entries in a Y maze (n= 24, *Kmt2a*^{+/+}, n= 18, *Kmt2a*^{+/LSL}). (E) Adult *Kmt2a*^{+/LSL} mice exhibit a significant decrease in percentage of spontaneous alternations in a Y maze. (F) Schematic figure of the mouse hippocampus with a representative image showing the DG region of the hippocampus. (G) Representative images of a DAPI staining of the DG area of hippocampus in *Kmt2a*^{+/+} and *Kmt2a*^{+/LSL} mice. (H) Quantification of granule cell layer area of the DG demonstrating a decrease in *Kmt2a*^{+/LSL} mice compared to wild-type littermates when corrected for body weight (n=3-4 per genotype). ns: not significant, *p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001.

Nes-Cre^{+/-}Kmt2a^{+/-LSL} mice exhibit successful cassette removal

To test *in-utero* rescue of the neurological phenotypes, we crossed our Kmt2a^{+/-LSL} model with a nestin-Cre model (Nes-Cre^{+/+}) which should allow *in-utero* rescue of Kmt2a expression in a cell-type specific manner in the nervous system at mid to late gestation (Figure 3A). We hypothesize that with a knock in of Kmt2a expression in nervous tissue, the markers of neurological dysfunction will show rescue in Nes-Cre^{+/-}Kmt2a^{+/-LSL} mice, with a prior study [13] demonstrating rescue at the time of conception. To validate the efficacy of the LSL cassette recombination in Nes-Cre^{+/-}Kmt2a^{+/-LSL} mice in the nervous system we performed PCR with primers targeting the LSL cassette in the hippocampus and skin tissue from the ear using gDNA and compared to Nes-Cre^{+/-}Kmt2a^{+/+} controls. We observe an obvious decrease of the amount of LSL cassette in the nervous tissue but not in the skin tissue, of Nes-Cre^{+/-}Kmt2a^{+/-LSL} mice compared to controls (Supplemental Figure 4). To further validate the removal of the LSL cassette and the ability to regain normal Kmt2a expression, we quantified the relative mRNA expression in whole brain lysates of P1 pups and observed no significant difference between Nes-Cre^{+/-}Kmt2a^{+/+} and Nes-Cre^{+/-}Kmt2a^{+/-LSL} mice (Figure 3B, n=3 per genotype) in contrast to a significant difference between the genotypes without the Cre (Figure 1B).

Gene expression abnormalities are rescued in Nes-Cre^{+/-}Kmt2a^{+/-LSL} mNPCs

To evaluate rescue of transcriptional abnormalities caused by heterozygous Kmt2a-deficiency in a homogenous cell population, we dissected the hippocampus from newborn pups (P0) and cultured mouse neuronal progenitor cells (mNPCs). We performed RNA-sequencing using mRNA from mNPCs from both genotypes to identify gene expression changes between Kmt2a^{+/+} and Kmt2a^{+/-LSL} NPCs (n=3 per genotype). We observed 64 significant differentially expressed genes (DEGs) when comparing Kmt2a^{+/+} and Kmt2a^{+/-LSL} samples with a majority of genes downregulated (47/64 genes), in concordance with the loss of a transcriptional activator (Figure 3C, Supplemental table 1).

Next, we wanted to assess whether there was an *in-utero* reversal of the abnormalities we uncovered in the RNA-seq dataset. First, we observe a significant overlap of 438 genes (p-value < 0.0001, Fisher's exact test) within the top-decile of p-values for comparisons between both Kmt2a^{+/+} vs Kmt2a^{+/-LSL} and Kmt2a^{+/-LSL} vs Nes-Cre^{+/-}Kmt2a^{+/-LSL}. Secondly, there is a striking

negative correlation of the log₂-fold change (LFC) of these genes between the comparisons, where 96.9% of genes (423/438 genes) demonstrated such an effect ($R^2=0.57988$, Figure 3D, Supplemental table 2). In essence, these results indicate that downregulated genes from *Kmt2a*^{+/-LSL} vs *Kmt2a*^{+/-} show upregulation in *Kmt2a*^{+/-LSL} mice in presence of the nestin-Cre and vice versa. We performed an over-representation analysis (ORA) of the genes showing such rescue and observed significant overrepresentation of pathways implicated in neuron development and neuron differentiation (Supplemental Figure 5A). We pooled the genes from the overrepresented pathways together and can observe a similar trend, with 20/26 genes downregulated in *Kmt2a*^{+/-LSL} samples (Figure 3E, see filtering criteria in Methods).

Notably, we observed down-regulation of a transcription factors, such as *En2* and *Sox5*, in *Kmt2a*^{+/-LSL} mNPCs compared to the *Kmt2a*^{+/-} mNPCs. Disruption of these transcription factors has previously been linked to defects in hippocampal neurogenesis [26], [27] as well as being associated with intellectual disability and autism-spectrum-disorder (ASD) [28][29]. Since up to 50% of WDSTS patients have ASD [30], we compared the rescued genes to known ASD-risk genes [31] and discovered a significant overlap of 26 genes (p-value < 0.0001, Fisher's exact test, Figure 3F). Interestingly, *Kmt2a* was not among the significant DEGs in this dataset. However, upon further examination of individual exons in *Kmt2a*^{+/-LSL} cells, the majority of reads were limited to the first exon of *Kmt2a*, with an overall decrease in coverage for the full length *Kmt2a* transcript (Supplementary figure 5B-C). We then preformed an isoform-specific analysis that revealed a significant shift of *Kmt2a* isoforms in the *Kmt2a*^{+/-LSL} samples, from the full length canonical *Kmt2a*-201 isoform to the truncated *Kmt2a*-203 isoform, that does not contain the enzymatic SET domain (Figure 5G-H). More notably, upon the LSL cassette removal in the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} cells, these effects are reversed, confirming functionality of the cassette. Taken together, our findings support the notion that a heterozygous loss of *Kmt2a* leads to aberrant mRNA expression in mNPCs. These changes at the molecular level are then reversed *in-utero* in the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} model.

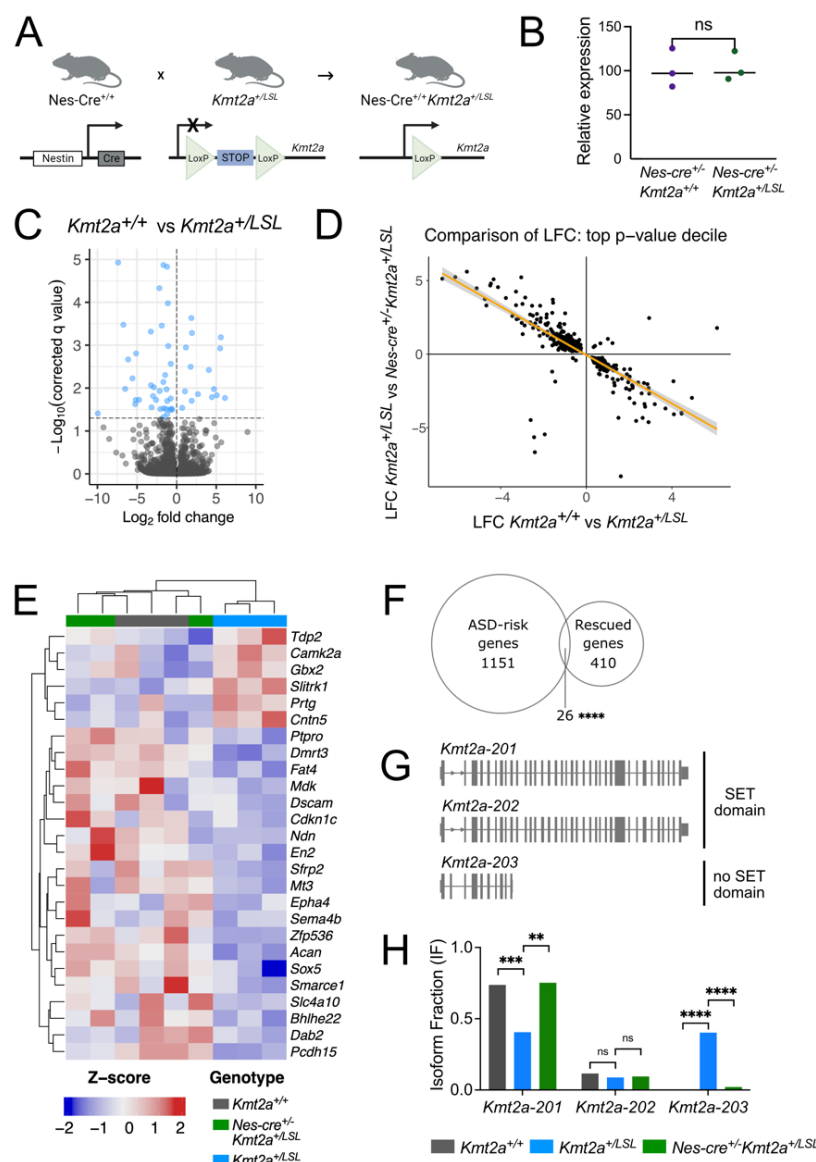


Figure 3: Transcriptional consequences of *Kmt2a* loss and rescue.

(A) A schematic overview of the nestin-Cre strategy. By exposing the *Kmt2a^{+/-LSL}* mice to nestin-Cre *Kmt2a* levels are expected to be restored *in-utero* in the nervous system of the *Nes-Cre^{+/+} Kmt2a^{+/-LSL}* mice *in-utero* allowing us to test *in-utero* rescue. (B) RT-qPCR demonstrates that relative *Kmt2a* mRNA expression is unchanged in *Nes-Cre^{+/+} Kmt2a^{+/-LSL}* mice compared to *Nes-Cre^{+/+} Kmt2a^{+/+}* littermates (n=3 per genotype) in nervous tissue. (C) Volcano plot showing the log₂ fold changes of differentially expressed genes comparing *Kmt2a^{+/+}* vs *Kmt2a^{+/-LSL}* mNPCs with significant genes labelled in blue. (D) Correlation plot showing the negative correlation between log₂ fold changes between *Kmt2a^{+/+}* vs *Kmt2a^{+/-LSL}* and then between *Kmt2a^{+/-LSL}* vs *Nes-Cre^{+/+} Kmt2a^{+/-LSL}* mice ($R^2=0.57988$). (E) Heatmap showing clustering of *Kmt2a^{+/-LSL}* samples compared to *Kmt2a^{+/+}* and *Nes-Cre^{+/+} Kmt2a^{+/-LSL}* samples of the 26 rescued genes from significantly overrepresented biological pathways (with FDR<0.05). Expression values were normalized to z-scores and plotted, with increased expression represented in red and decreased expression in blue. (F) Venn diagram of overlapping rescued genes with 1177 known ASD-risk factor genes, with 26 genes overlapping. (G) Representative figure of *Kmt2a* isoforms and their exons. (H) Isoform switch analysis reveals a shift in isoform fraction (IF) in *Kmt2a^{+/-LSL}* mNPCs from the longer SET domain-containing *Kmt2a* transcript to the short non-SET domain containing *Kmt2a* transcript. ns: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

H3K4me1 signal is rescued but does not drive gene expression changes in Nes-Cre^{+/-}

Kmt2a^{+/*LSL*} mNPCs

To look further into the effects of heterozygous *Kmt2a* loss at the chromatin level, we performed a quantification of the H3K4me1 mark using CUT&RUN in mNPCs (n=3 *Kmt2a*^{+/+}, n=4 *Kmt2a*^{+/*LSL*} and Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*}). When comparing *Kmt2a*^{+/+} and *Kmt2a*^{+/*LSL*} samples, the majority of differential peaks (70.4%, 1380/1962 peaks) show a gain of H3K4me1 in *Kmt2a*^{+/*LSL*} samples (Figure 4A, Supplemental Table 3). While mapping enhancer regions to genes is not trivial, when we linked the disrupted peaks to the nearest gene by distance, we found no correlation between the gene expression changes and the H3K4me1 epigenetic mark (Figure 4B-C, See Methods). Investigating the increased H3K4me1 signal further, we compared the signal over mouse NPC enhancer and neural enhancer regions from EnhancerAtlas [32]. This revealed an increased signal at the neuronal enhancers but not at the NPC enhancers for *Kmt2a*^{+/*LSL*} samples (Figure 4D). Again, the increased signal does not correlate with the nearest gene's expression changes (Supplemental Figure 6A). To assess the rescue of this epigenetic mark in Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} samples, we observe a reversal of 98.0% of peaks (Supplemental Figure 6B) and when visualizing the overall H3K4me1 signal at the differential peaks on a heatmap, we observe an increased signal in *Kmt2a*^{+/*LSL*} samples compared to *Kmt2a*^{+/+} wild-types, which is then reduced in Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} samples (Figure 4E). This effect can be more specifically visualized on tracks showing gained H3K4me1 signal at neuronal enhancers for genes *Acsf3* and *Bahcc1* for *Kmt2a*^{+/*LSL*} samples (Figure 4F). When applying a more relaxed cut-off, the 4963 peaks that fell into the top decile of p-values when overlapping the results from *Kmt2a*^{+/*LSL*} vs *Kmt2a*^{+/+} samples and the *Kmt2a*^{+/*LSL*} and Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} samples, map to 3646 individual genes. When performing an over-representation analysis (ORA) of these genes, we observe pathways involved in CNS differentiation, development, and neurogenesis (Figure 4G), highlighting that the disruption is occurring at functionally relevant sites. When examining the genes with restored H3K4me1 signal and gene expression, we observe a non-significant overlap of genes with a gained H3K4me1 peak and increased expression in *Kmt2a*^{+/*LSL*} samples (Figure 4H), as well as a non-significant overlap of genes with a lost H3K4me1 peak and decreased expression in *Kmt2a*^{+/*LSL*} samples (Figure 4I). Collectively, we observe an increased H3K4me1 signal in *Kmt2a*^{+/*LSL*} samples, albeit with no global effects on gene expression.

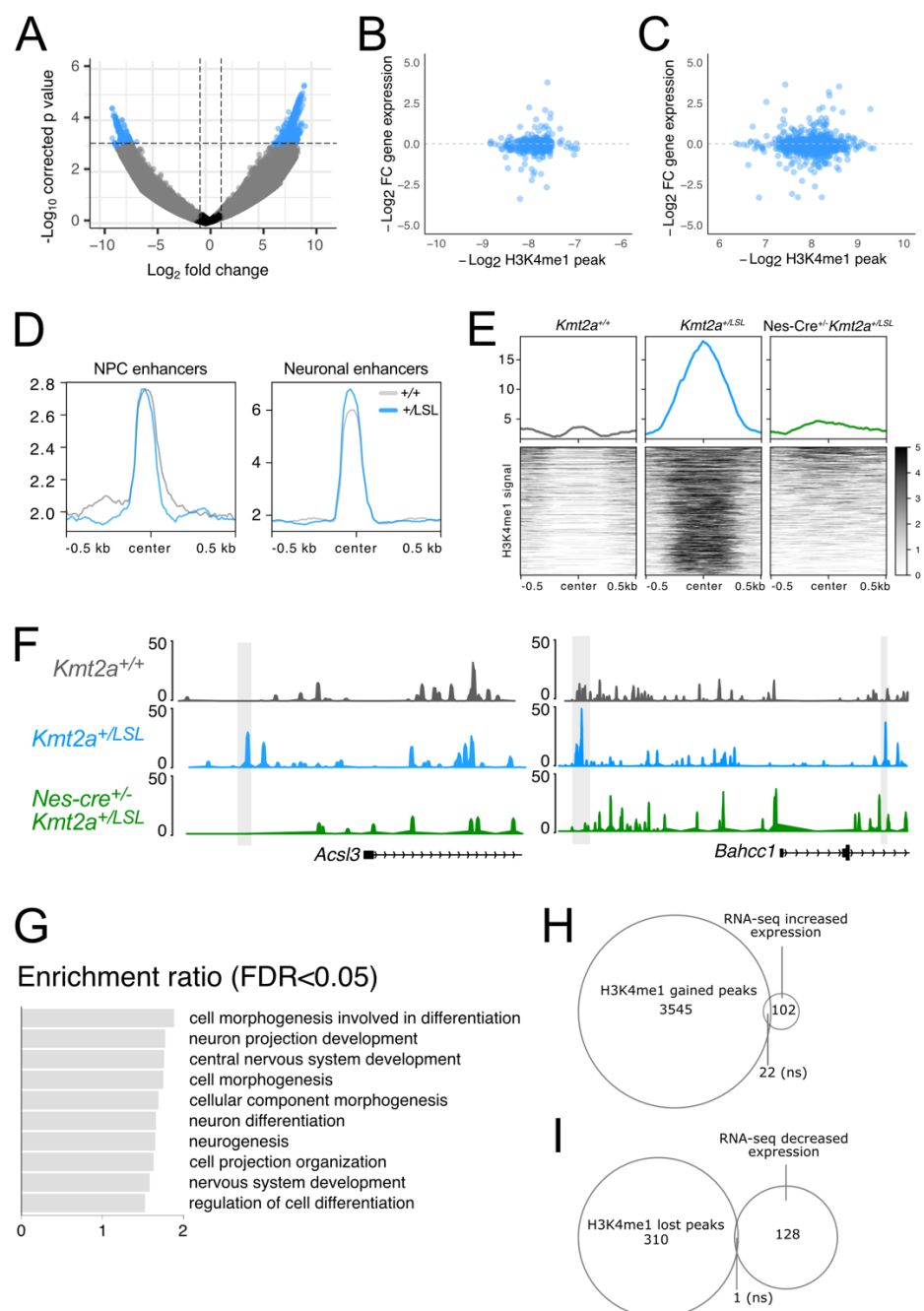


Figure 4: Epigenetic imbalance caused by heterozygous *Kmt2a* loss.

(A) Volcano plot showing differential H3K4me1 peaks comparing *Kmt2a*^{+/+} and *Kmt2a*^{+/-LSL} samples. Expression changes for genes with (B) lost H3K4me1 peaks and (C) gained H3K4me1 peaks in *Kmt2a*^{+/-LSL} samples. (D) Signal plot of H3K4me1 signal for *Kmt2a*^{+/+} and *Kmt2a*^{+/-LSL} samples over NPC and neuronal enhancers. (E) Heatmap of H3K4me1 signal at differential sites when comparing *Kmt2a*^{+/+}, *Kmt2a*^{+/-LSL} and *Nes-Cre*^{+/+}*Kmt2a*^{+/-LSL} samples. (F) Representative H3K4me1 signal figures, showing gained signal in *Kmt2a*^{+/-LSL} samples at neuronal enhancers (gray) at *Acs13* and *Bahcc1* genes compared to wild-type and rescued samples. (G) Over-representation analysis of genes with differential H3K4me1 peak in *Kmt2a*^{+/-LSL} samples. Image shows enriched pathways (with FDR<0.05). (H) Overlap of genes with a gained H3K4me1 peak and genes that have an increased expression in *Kmt2a*^{+/-LSL} samples. (I) Overlap of genes with a lost H3K4me1 peak and genes that have decreased expression in *Kmt2a*^{+/-LSL} samples. ns: not significant, *p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001.

WDSTS phenotypes are rescued *in-utero*

After demonstrating that nestin-Cre leads to sufficient recombination of the LSL cassette and rescued *Kmt2a* expression (see Figure 3, Supplemental Figure 4) we characterized the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice. The Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice demonstrate growth deficiency compared to their Nes-Cre^{+/-}*Kmt2a*^{+/+} littermates (n=5-9, Figure 5A, Supplemental figure 7A-B). Furthermore, we found similar craniofacial defects of the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} as previously described in *Kmt2a*^{+/-LSL} mice (Figure 1E-H) indicating that there is no rescue of this phenotype (Supplemental figure 7C-D). This would suggest that either the growth defects of the *Kmt2a*^{+/-LSL} mice are not solely driven by nestin-expressing cell types, or that the cause of the growth phenotypes occurs prior to when the rescue of *Kmt2a* takes place *in-utero*. In contrast, the prevalence of the white abdominal spot of the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice drastically decreased from 57.5% to 12.5% (Figure 5B) compared to the 57.5% of the *Kmt2a*^{+/-LSL} mice, indicating malleability of this phenotype. Behavioral testing of adult mice demonstrates no difference of arm entries between Nes-Cre^{+/-}*Kmt2a*^{+/+} and Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} (n=14-19, Figure 5C), nor of the percentage of spontaneous alterations in the Y-maze (Figure 5D), indicative of non-impaired working memory of Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice. Similarly, the brain histology of the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice demonstrate normal hippocampal structure, with a non-significant difference in the size of the area of the granule layer of the DG, when compared to Nes-Cre^{+/-}*Kmt2a*^{+/+} littermates and corrected for body weight (Figure 5E-F). Taken together, when bred to a nestin-Cre mouse model, we observe histological and behavioral rescue of the neurological defects of the Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} mice.

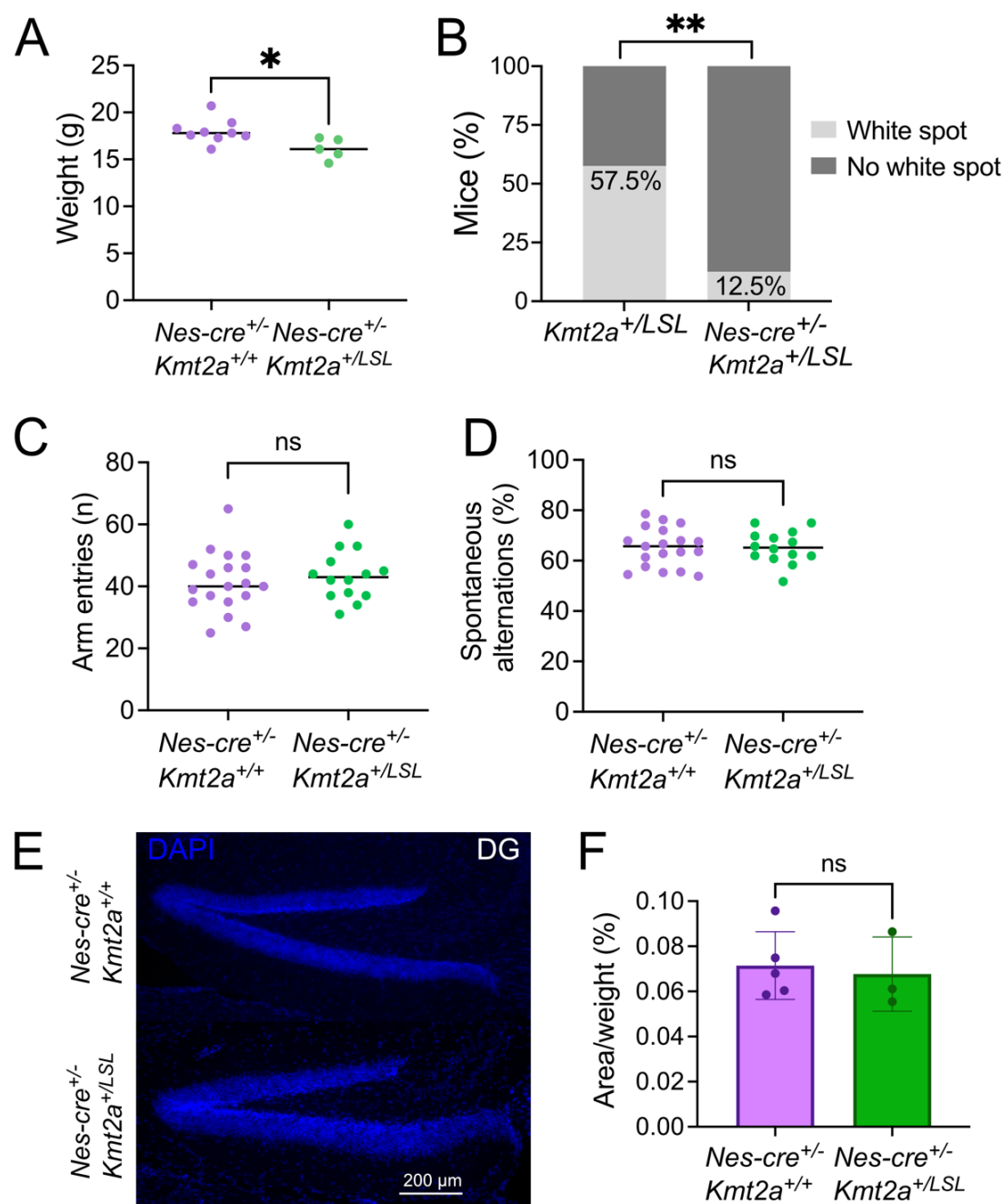


Figure 5: In-utero rescue of the neurological phenotypes in WDSTS mice

(A) The $Nes-cre^{+/-} Kmt2a^{+/LSL}$ mice demonstrate growth deficiency compared to their $Nes-cre^{+/-} Kmt2a^{+/+}$ littermates (n=5-9). (B) The presence of the white belly spot is significantly reduced in $Nes-cre^{+/-} Kmt2a^{+/LSL}$ mice (12.5%) compared to $Kmt2a^{+/LSL}$ mice (57.5%) (Fisher's exact test). (C) $Nes-cre^{+/-} Kmt2a^{+/LSL}$ mice do not show a significant decrease of total arm entries in a Y maze (n=14-19 per genotype) (D) $Nes-cre^{+/-} Kmt2a^{+/LSL}$ mice do not exhibit differences in spontaneous alternations in the Y-maze, compared to their $Nes-cre^{+/-} Kmt2a^{+/+}$ littermates. (E) Representative images of DG of the hippocampus area in $Nes-cre^{+/-} Kmt2a^{+/+}$ and $Nes-cre^{+/-} Kmt2a^{+/LSL}$ mice. (F) Quantification of granule cell layer area of the DG of the hippocampus demonstrating no difference in size when corrected for body weight (n=3-5 per genotype). ns: not significant, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

Discussion

In this paper we have created and characterized a novel WDSTS mouse model (*Kmt2a*^{+/*LSL*}). This model depicts many of the human WDSTS phenotypes, validating it as a disease model for WDSTS. We observed rescue of the neurological phenotype of the mouse model after crossing the mice to a nestin-Cre model, demonstrating *in-utero* malleability of WDSTS. Genetic disorders are being diagnosed earlier than before with the onset of whole genome sequencing and exome sequencing [33] and there is an explosion of strategies that have successfully made it to the clinic including small molecules, ASO-based strategies and CRISPR-Cas9 based strategies [34], [35]. Using a genetically based strategy to examine the likelihood, the window of opportunity, and which aspects of the phenotype are malleable could be of great value for prioritizing any Mendelian disorder for therapeutic development as well as offering an excellent genetic control for such studies. Our work uses one such strategy which could be used for other congenital disorders to test the timing and effectiveness of early manipulation. We demonstrate rescue of the altered *Kmt2a* allele and its consequences on multiple levels: genetic, molecular, and phenotypic. Our work also provides temporal insights into the timing of disease onset in the CNS and neural crest. There is some debate in the literature surrounding the exact timing of action in the nestin-Cre mouse model, with the original paper suggesting recombination occurring as early as E11 in the nervous system [36], [37]. However, more recent studies have shown that Cre expression only reaches sufficiently high levels in NSCs and NPCs to cause recombination during late embryonic (E15.5-17.5) and early postnatal periods [36].

In vitro, we find robust evidence that the transcriptional changes caused by the loss of *Kmt2a* are reversed, with the majority of genes downregulated upon its loss and then restored. Surprisingly, the directionality of the chromatin-rescue occurs in the opposite direction from what would be expected for an H3K4 methyltransferase, i.e. the sites that show the biggest effect have increased levels of the H3K4me1 mark for *Kmt2a*^{+/*LSL*} samples, that then normalizes upon genetic rescue. A similar chromatin effect, with a trend towards increased chromatin accessibility, has previously been observed in studies of patients and animal models of Kabuki syndromes [38], [39] and in cells from animal models of Rubinstein-Taybi and Kleeftstra syndromes [7], [40]. Moreover, the increased H3K4me1 signal at neuronal enhancers could suggest the cells are priming these enhancers towards a premature neuronal

phenotype, similar to what has been described in Kabuki syndrome [41], [42]. Another potential reason for this observation could be that the increased H3K4 methylation involves a compensatory response. For instance, KMT2A might act as a direct or indirect negative regulator of another epigenetic regulators that promote open chromatin. We also observe signs of increased transcriptional initiation of *Kmt2a* and isoform switching in *Kmt2a*^{+/*LSL*} samples, showing attempted compensation is taking place at least at the locus itself. Alternatively, we hypothesize KMT2A might be occupying binding sites, effectively blocking other epigenetic- or transcriptional factors from binding at those loci. We also find no correlation of gene expression linked to the changes in H3K4me1 signal, suggesting that the lack of KMT2A's enzymatic function is not the sole driver of WDSTS pathogenesis. This aligns with our prior finding that WDSTS missense variants do not show enrichment within the enzymatic SET domain of KMT2A in contrast to e.g. Kabuki syndrome causing missense variants that cluster in the KMT2D-SET domain [16][43]. This is emphasized even further with the fact that enzyme-specific homozygous KO of *Kmt2a* yields viable and fertile mice while such KOs lead to lethality in other *Kmt2* mouse models e.g. *Kmt2d* [11][44]. Further studies will be needed to understand the underlying mechanism of this effect in more detail. Our findings suggest that disruption of gene expression and a global imbalance in the epigenetic machinery are not solely caused by the loss of the enzymatic activity of KMT2A and may result in cellular compensation, leading to downstream secondary effects, which could play a major role in MDEM disease mechanisms.

In addition to previously described features, the *Kmt2a*^{+/*LSL*} model demonstrates multiple additional common WDSTS-like phenotypes such as craniofacial defects, hypotonia and hypertrichosis, one of the defining features of WDSTS (Supplemental Table 4). Interestingly, our mice appear to show increased hair follicle density in some regions, without an obvious rate of regrowth of the hair, and could therefore open a novel avenue of research into the mechanistic basis of hypertrichosis in humans. We observe no rescue of the growth retardation of Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} mice. However, it is important to note that the nestin-Cre mice have previously been shown to demonstrate mild hypopituitarism, accompanied by decreased body weight, and to ensure that these results could not be due to the nestin-Cre genotype, we compared their weight to the Nes-Cre^{+/-}*Kmt2a*^{+/+} littermates [45]. Similarly, we observed no rescue of the craniofacial structure abnormalities in the Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} mice. This could be explained in two ways. Either the growth retardation might be related to

dysfunction in cell populations that do not express nestin, or secondly, the growth retardation might occur early enough in development to not be rescued by the nestin-Cre expression. In both scenarios we would not expect a rescue of these phenotypes.

Previous studies have linked abnormal craniofacial features to neural crest dysfunction in KS in mice [19], [20], [21], a prominent phenotype shared between KS and WDSTS. In concordance, we observe a prominent white abdominal spot in the majority of *Kmt2a*^{+/*LSL*} mice, with a striking decrease in the Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} mice, without fully eliminating it. This could suggest the timing of rescue of the neural crest defects is around the same time as nestin expression starts, showing partial rescue in our model. Ultimately, our findings suggest a role for *Kmt2a* in neural crest development with more research needed to fully elucidate its role. Lastly, we observe rescue of the neurological dysfunction in the Nes-Cre^{+/-}*Kmt2a*^{+/*LSL*} mice both histologically and in behavior. We conclude the neurological phenotype of WDSTS is malleable *in-utero* and that WDSTS is yet another treatable cause of ID from the group of Mendelian disorders of the epigenetic machinery.

Acknowledgements

We thank the Wiedemann-Steiner foundation for supporting this work. We thank Leandros Boukas, Katrín Möller and Hilmar Örn Gunnlaugsson for valuable input. Some figures were created using BioRender.com. Data utilized in this publication was provided by RARE-X, a research program of Global Genes. RARE-X provides a collaborative platform for global data sharing and analysis to accelerate research for rare diseases. Global Genes is a 501(c)(3) non-profit organization dedicated to eliminating the burdens and challenges of rare diseases for patients and families globally.

Funding

This work was specifically funded by a grant to HTB from the Wiedemann-Steiner Foundation. HTB and JO are also funded by a grant from the Icelandic research fund (217988-051). TR and JO are funded by individual grants from the University of Iceland doctoral fund grants. Mouse model strain development, strain curation and cryopreservation of the mouse strains was supported by the Jackson Laboratory Precision Genetics Grant from the National Institutes of Health, U540D030157 to C.M.L.

Author contributions

Conceptualization: HTB; Methodology: HTB, TR, KJA, JO, AOS, VBD, ARZ; Validation: TR, KJA, ARZ; Formal analysis: TR, KJA, VBD, AB, KPF; Investigation: TR, KJA, AOS, JO, ARZ; Resources: HTB; Data curation: TR, KJA, LB, VBD, AB, KPF; Writing – original draft: TR; Writing – review & editing: HTB, TR, KJA; Visualization: TR, KJA, VBD; Supervision: HTB; Project administration: TR, HTB, KJA; Funding acquisition: HTB, CML.

Competing interests

HTB is a consultant for Mahzi Therapeutics and founder of Kaldur Therapeutics. AOS is an employee of Arctic Therapeutics. Other authors declare that they have no competing interests.

Methods

Animals

The mouse model was created by the Jackson Laboratory, by microinjection of C57BL/6J single cell zygotes with CRISPR/Cas9 using guides 5'-CTGCAGCGAGAGACTGTATG-3' and 5'-GAGACTGTATGAGGTATCAG-3' with a donor plasmid containing a lox-STOP-lox element flanked by *Kmt2a* intron 1 genomic DNA sequences of 1.4-kb and 1.8-kb, respectively. Of 38 mice recovered, six founders were identified with the desired genome edited allele at the *Kmt2a* locus, as determined by genomic long-range PCR. Two lines were established after mating of the founders to C57BL/6J mice to confirm germ line transmission of the edited lox-STOP-lox knock-in (KI) allele in *Kmt2a* and subsequently expanded by an additional backcross to C57BL/6J mice (JAX Stock #664) prior to intercrossing and strain validation [46]. These two conditional KI (cKI) lines, designated as stock numbers #35724 C57BL/6J-*Kmt2a*^{em7Lutzy}/J and #37525 C57BL/6J-*Kmt2a*^{em8Lutzy}/J each contain the same allele but were derived from different founders. The functionality of the loxP elements within each KI allele was confirmed by test mating of heterozygous #35724 and #35725 male mice with hemizygous B6.Cg-*Edil3*^{Tg(Sox2-cre)1Amc}/J female Cre-expressing mice (JAX Stock #8454) [47]. Whereas both strains were mated with Sox2-Cre to confirm loxP functions, postCre Het x Het matings were only performed using the em7.1 allele (JAX Stock #35727, now extinct). Sox2-Cre Tg did not segregate in these mice (Supplemental Figure 1G). This validation revealed both stocks #35724 and #37525 behaved the same way in subsequent matings and therefore only Stock #35724, named *Kmt2a*^{+/^{LSL}}, was characterized further with respect to expression profiling in the pre-and post Cre mice. Jackson laboratory Stock #35724 can be purchased from the Jackson Laboratory mouse mutant repository (<https://mice.jax.org>). Genotyping was performed from DNA isolated from ear clips using primers *Kmt2a*-genoF1-(5'-ACA CAT GGC TTC CTG GAG G-3') and *Kmt2a*-genoR1-(5'-CTG TTT GCA GTC GGA AAG CC-3'). Nestin-Cre^{+/⁺} mice (JAX stock #3771) were acquired from the Jackson Laboratory [48], [49]. Nestin-Cre model was genotyped using primers NesCre-WT-F-(5'-TTG CTA AAG CGC TAC ATA GGA -3'), NesCre-Tg-F-(5'-CCT TCC TGA AGC AGT AGA GCA-3'), NesCre-R-(5'-GCC TTA TTG TGG AAG GAC TG -3'). All mouse strains were maintained on pure C57BL/6J background and provided food (Altromin NIH#31 M (breeding) or Altromin 1324 (maintenance), Brogaarden) and filtered water *ad libitum*. The animals were maintained under a 12-hour light/dark cycle at room temperature (21-23°C) with relative humidity around 40%. Regular FELASA health monitoring was conducted with samples taken from sentinel and resident animals and analyzed by IDEXX Germany, yielding negative results during experimental period. Import of mice and experiment protocols were in accordance with Icelandic Food and Veterinary Authority (license no. 2208602) and accepted by National Expert Advisory Board on Animal Welfare.

Behavioral testing

The hypotonia tests (surface righting and hindlimb suspension tests) were performed on P6 pups as described by Feather-Schussler et al [50]. For the surface righting test, the pups were placed on their backs and held in place for 5 seconds, recording the time for the pups to return to a prone position. The test was conducted in 3 trials, giving the pup a 30 second rest in between each trial and a time-out at 60 seconds, with mean scores over all trials plotted. For the hindlimb suspension test, the pups were placed on their hindlimbs on the rim of a 50 ml conical tube, with the latency to fall recorded. The test was conducted in 3 trials, giving the pup a 60 second rest in between each trial and a time-out at 5 minutes. All trials were plotted,

with each trial tiring the mice more. The spontaneous alternation Y-maze was performed on 8–10-week-old mice according to protocol described by Kraeuter et al [51]. Spontaneous alternation percentage is calculated with the following formula:

$$\% \text{ spontaneous alternation} = \frac{\text{Number of Alternations}}{(\text{Total number of arm entries} - 2)} \times 100$$

CT scan

CT scans were performed on 2-3.5-month-old mice (3-6 mice per genotype) using 50 kV tube voltage, 0.21 mA tube current and 1712 mGy dose estimate using the pre-clinical optical imaging and CT system from MILabs (Houten, Netherlands). Virtual reconstructions of the skull were produced for each specimen by importing the CT data to 3dSlicer at half resolution (0.030 x 0.030 x 0.030 mm³ cubic voxels), and generating a model of the skull surface based on thresholding. A geometric morphometric approach allowed us to estimate cranial *shape* based on 3D coordinate data for a set of 31 biologically homologous landmarks (Supplemental Table 5) evenly distributed across the cranium. Landmark coordinate data were exported from 3dSlicer [52] and we used the R *geomorph* package (v.4.0.7) for all morphometric analyses [53]. We assumed object symmetry for each specimen using the *bilat.symmetry()* function, which removes the effect of asymmetric variation [54]. For a subset of our data, rostral and caudal ends of the crania were inadvertently truncated in the original CT scanning episode. This resulted in missing data for nasale (nal) in six specimens, and for opisthion (opi), basion (bas), and occipital condyles (rfmc and lfmc) in 11 specimens. Missing data were estimated using the *estimate.missing()* function, which predicts the location of those landmarks based on the average of other individuals in the sample [55]. Both *Kmt2a*^{+/*LSL*} and wild-type genotypes were equally represented in this average sample. For each analysis, we conducted a new Procrustes superimposition and tested the effect of genotype using a Procrustes ANOVA with the *procD.lm()* function. We visualized the results using Principal Components Analysis (PCA).

Histology

Samples were taken from the belly and backs of newborn pups, fixed in 4% paraformaldehyde, and then processed at the Department of Pathology, Landspítali National University Hospital, Reykjavik, Iceland. The tissue samples were paraffin-embedded and cut into 1.5µm sections for hematoxylin and eosin (H&E) staining and 3µm sections for immunohistochemistry (IHC). H&E staining was done using standard methods. Sections for immunohistochemistry were de-paraffinized and rehydrated in xylene and ethanol. Antigen retrieval was done with Envision Flex Target Retrieval Solution, high 9 (Agilent, K800421-2). Sections were immunostained in AutostainerLink 48 from Agilent with EnVision Detection System Peroxidase/DAB, Rabbit/Mouse kit (Agilent, K4065). Incubation with primary antibody Calretinin (Agilent, M7245) was performed at room temperature for 30 min. After incubation with a primary antibody, sections were incubated with EnVision FLEX/HRP. Sections were washed between steps with Tris-buffered NaCl solution with Polysorbate 20 pH 7.6 (Agilent, S3306). All sections were incubated with 3,3'-diaminobenzidine solution (Agilent, K4065) for 10 min. Sections were counterstained with hematoxylin for 30 seconds. Finally, sections were dehydrated with 100% ethanol and xylene followed by cover slipping with mounting medium (Pertex, Histolab). Images were acquired with NanoZoomer XR (HAMAMATSU) scanner. Sections were analyzed using NDP view (version 2.7.25). Hair follicles were counted on 8 mm

of skin, each section was counted 4 times, and the average number was plotted per mm of skin.

CBC measurement and hair removal

Blood samples for Complete Blood Count (CBC) were collected in EDTA blood tubes from 3-5 m/o mice (n=8 per genotype) and measured on VetScan HM5 v2.1 (Abaxis Inc, USA). Hair was removed from the backs of adult mice using Veet Hair Removal Crème (Reckitt Benckiser).

Perfusion, cryosectioning, IF staining

Mice were euthanized 9 weeks old with Xylazine (4 mg/ml final concentration) Ketamine (16 mg/ml concentration) combination in 0.9% NaCl and perfused intravascularly with 4% PFA after first flushing blood system with 1x PBS. Brains were dissected out and cryopreserved in a two-step gradient of 30% and 15% sucrose solution (sucrose in 0.1M phosphate buffer pH=7.2). Brains were cut sagittally and frozen in Tissue-Tek OCT compound. Brains were sectioned in 20µm serial sections on Leica CM1850 cryostat (Leica Biosystems). Sections were stored at -80°C until stained. Every 2nd, 6th, 10th, and 15th brain section were used for DAPI staining. Samples were washed 2x10 minutes with TBS-T (1x TBS (Santa Cruz, sc-24951) with 0.05% Triton X-100 (Sigma-Aldrich, T8787-100ML)) and then treated in DAKO (sodium citrate buffer pH=6) at 95°C for 20 minutes for antigen retrieval, washed 3x10 minutes in TBS-T and then blocked for an hour in TBS-T+ (TBS-T with 3% normal goat serum (Abcam, ab7481)) at room temperature. Samples were incubated in primary antibody solution (antibody in TBS-T+) overnight at 4°C. Slides were washed 3x 10 minutes in TBST-T and mounted with Fluoromount-G with DAPI (Thermo Fisher Scientific, 00-4959-52). Primary antibodies used Alexa Fluor® 647 Anti-NeuN antibody (Abcam, ab190565 1:500 dilution). Brain samples were imaged using CrestOptics CICERO spinning disk with a 10x and 40x objective. Images were processed using Fiji [56]. DAPI staining were used for area quantification, representative NeuN stainings in Supplemental Figure 2F.

RNA isolation from brain and Real Time Quantitative Polymerase Chain Reaction (RT-qPCR)

Total RNA was isolated from whole brain tissue of P1 pups using TRIzol™ Reagent (Thermo Fisher Scientific, #15596026) according to the manufacturer's instructions. RNA concentration was measured on NanoDrop (Thermo Fisher Scientific), and 0.5 ng of RNA was used for input for cDNA synthesis using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, #4368814), run on the MiniAmp Thermal Cycler (Applied Biosystems). Luna Universal qPCR Master Mix (NEB, M3003) was used for RT-qPCR reactions and run on CFX384 Real-Time PCR Detection System (Bio-Rad). Each biological replicate of the RT-qPCR assay in this study was carried out in technical triplicates.

Mouse NPC extraction and culture

Mouse NPCs were dissected from hippocampus of P0 pups and cultured in an adapted protocol from Bernas et al [57]. Dissected hippocampi were dissociated in 1X TrypLE™ (Thermo Fisher Scientific, A1217701), washed in Neurobasal growth medium: Neurobasal medium (Thermo Fisher Scientific, 21103049) with 1X B27 supplement (Thermo Fisher Scientific, 17504044) and filtered through a 70 µm cell strainer (Miltenyi Biotech, 130-110-916). The cell suspension was centrifuged at 350g for 5 minutes, resuspended in 500µl Neurobasal growth medium containing 1X Penicillin/Streptomycin (Sigma, P4333-100ML), 1X Glutamax (Thermo Fisher Scientific, 35050038), 20ng/mL FGF-2 (Peprotech, 100-18B), 20

ng/mL EFG (Peprotech, AF-100-15), and 2 µg/mL heparin (MP Biomedicals, 210193125) and seeded onto Matrigel (Corning, 354234) coated plates that were previously incubated at 37°C for 2 hours (Matrigel diluted 1:100 in DMEM/F-12 with GlutaMAX™ Supplement (Thermo Fisher Scientific, 31331028)). After first splitting of cells, mNPCs were cultured in media conditions listed above on plates coated firstly with poly-D-Lysine (PDL) hydrobromide (Sigma, P7280) overnight followed by laminin (Roche, 11243217001) coating for 2 hours at 37°C.

RNA-seq

mNPCs were harvested in TRIzol™ Reagent (Thermo Fisher Scientific, #15596026) and RNA isolated with Direct-zol™ RNA Microprep kit (Zymo, #R2060). RNA concentration and purity was measured on a Nanodrop and with Agilent RNA 6000 Nano Kit (Agilent Technologies, 5067-1511) on Agilent 2100 Bioanalyzer. All samples had RIN>9. Samples were sequenced by sequencing company Novogene (Cambridge, UK). Raw RNA-seq data (fastq files) were downloaded from Novogene and raw reads were pseudoaligned to the mouse mm10 reference transcriptome (GRCm38) using Kallisto (version 0.48.0) [58], running the kallisto quant command with -b 100 for 100 bootstraps. For viewing sequencing coverage, fastq files were aligned to mm10 reference transcriptome using kallisto quant using -genomebam, creating bam files that were converted to bigWig files using bamCoverage with default settings from deepTools2 (version 3.5.1) [59]. Genome coverage bigwig files were viewed in the UCSC browser [60]. Aligned data was imported to R [61] using tximport (version 1.30.0) [62] and mapped to mm10 genes with the BiomaRt package [63]. Differential expression analysis was performed on samples containing >10 counts and genes with only one sample containing counts were excluded. We used DESeq2 (version 1.42.0) using default settings [64]. To correct for false discovery rates, the z-scores returned from DESeq2 were used as an input for fdrtool with statistic="normal" and statistical cut of was set at corrected q-value < 0.05 [65]. For further analysis we used three genotypes: *Kmt2a*^{+/+} for wildtype samples, *Kmt2a*^{+/-LSL} for WDSTS samples and Nes-Cre^{+/-}*Kmt2a*^{+/-LSL} for assessing the *in-utero* rescue, excluding the Nes-Cre^{+/-}*Kmt2a*^{+/+} cells as they served only as another wildtype control. To assess the quality of the differential analysis we performed three quality control checks. First, we inspected the histogram of corrected p-values (Supplemental Figure 5D), revealing an even distribution of values between 0 and 1, with a clustering of values close to 0. This suggests a calibrated differential analysis. Secondly, we observed the genotype differences on a principal component analysis (PCA) plot (Supplemental Figure 5E), after performing regularized log transformation with rlog() function on expression matrix. PCA plot reveals slight variance between biological replicates within genotypes. Lastly, an MA plot demonstrates no obvious systematic issues of the data (Supplementary Figure 5F). WebGestalt was used for over-representation analysis (ORA) [66]. Autism-spectrum-disease (ASD) risk genes were acquired from SFARI Gene database (<https://gene.sfari.org/>, version Q1 2024) [31]. The Complex Heatmaps package was used for generating heatmaps for gene expression visualization (version 2.18.0) [67]. Expression values were normalized to z-scores and outliers with z score>2 were filtered out, with the rest plotted. IsoformSwitchAnalyzeR was used for identification of isoform switches (version 2.2.0) [68]. Venn diagrams were created with DeepVenn [69] and Fisher's test was applied for statistical analysis using GeneOverlap [70].

CUT&RUN

CUT&RUN was performed according to CUTANA™ CUT&RUN Protocol (v.1.7 EpiCypher) from 160.000 cells per sample, with n=3 *Kmt2a*^{+/+}, n=4 *Kmt2a*^{+/*LSL*} and n=4 Nes-Cre^{+/+}/*Kmt2a*^{+/*LSL*}. Cells were permeabilized with 0.01% digitonin buffer (Sigma, D141) and E. Coli DNA (EpiCypher, 23618-1401) was used as spike-in for normalization (0.2 ng final concentration per sample). The antibodies used were Rabbit IgG Antibody (EpiCypher, 23613-0042, 0.5μl per sample) and H3K4me1 (Abcam, ab8895, 0.25μl per sample). DNA was isolated using Genomic DNA Clean & Concentrator (Zymo, D4065). Library preparation was done using TruSeq RNA CD Index plate (Illumina, 20019792). Quality of samples were assessed using Agilent Bioanalyzer DNA high sensitivity kit (Agilent Technologies, 5067-4626). DNA samples were pooled to a 1.8pM concentration and sequenced on a NextSeq 550Dx instrument (Illumina) using NextSeq 500/550 High output kit v2.5 (Illumina, 20024907). Demultiplexing was performed using Illumina bcl2fastq command with the options --ignore-missing-control - --ignore-missing-positions --ignore-missing-filter. Fastq sample files were aligned to the mouse mm10 genome using Bowtie2 (version 2.4.4) [71] for paired end reading. Aligned sam files were sorted, indexed and converted to bam files using SAMtools (version 1.17) [72]. For visualizing H3K4me1 peaks in UCSC, bam files for all samples for each genotype were merged with SAMtools. They were then converted to bigwig files with deepTools (version 3.5.1) and normalized to 1x mouse genome with --normalizeUsing RPGC [59]. Peaks were called for merged genotypes against the merged IgG samples for each genotype using MACS3 (version 3.0.0) with settings -q 0.1, -maxgap-300, -keep-dup all [73]. First, the raw reads for each sample were counted using dba.count() function from Diffbind (version 3.12.0) with „score= DBA_SCORE_READS“, followed up with retrieving the matrix with dba.peakset() function, with settings „bRetrieve=TRUE“ [74]. Raw count matrix was filtered to only include peaks where the was greater than 10 for at least one genotype. Count matrix was used as input for DESeq2 where differential analysis was performed. FDRtool with statistic=“normal” was applied to correct for false discovery rates, with a cut off at corrected p-value <0.001. H3K4me1 peaks were annotated to nearest gene using annotatePeak() function from ChIPseeker package (version 1.5.1) [75]. Mouse NPC and Neuronal enhancer sites were retrieved from Enhancer Atlas (version 2.0) [32]. Heatmaps and plotprofiles were plotted using deepTools2 (version 3.5.1) [59].

GWS

DNA from mNPCs in culture were harvested from *Kmt2a*^{+/*LSL*} cell line cell line (1 million cells) using Monarch® HMW DNA Extraction Kit (NEB, T3050). We sheared the high molecular weight (HMW) DNA to achieve consistent and narrow size ranges to 20kb using the Megaruptor 3 Shearing kit (E07010003, Diagenode). Megaruptor 3 was performed using customized settings, including a concentration of 50ng/μL, a volume of 70μL, and a speed of 30. Subsequently, we measured the sheared DNA size using the HS Large Fragment 50kb kit (DNF-464-0500, Agilent) on a Fragment Analyzer and analyzed by the High Sensitivity Large Fragment 50 kb Analysis software. We performed library preparation with a Ligation Sequencing kit V14 (SQK-LSK114, Oxford Nanopore Technologies, ONT), and sequenced by loading the sample into an R10.4.1 flow cell (FLO-PRO114M, ONT) loaded on a PromethION 24 sequencer based on the manufacture's guidelines (ONT). Nanopore sequencing was conducted using customized run settings, which included a Run Limit of 96 hours, a Minimum Read Length of 200 bp, and Super-accurate Base calling. To optimize output, we washed the

flow cell once after 22 hours of sequencing using the Flow Cell Wash kit XL (EXP-WAS004-XL, ONT), followed by a 1.5-hour wait period and flushing of the flow cell with Flow Cell Priming Mix twice by using Sequencing Auxiliary Vials V14 (EXP-AUX003, ONT). We utilized the remaining DNA library from the previous preparation for loading. Data acquisition and base calling in real-time using PromethION device and MinKNOW software developed by ONT. Nanopore sequencing yielded approximately 86.43 Gb of estimated bases with an average read length of 23.5 kb. All fastq files were merged into one fastq file using cat command in Linux. A sam file was created by aligning the fastq file to the mouse genome using minimap2 with –map-ont function for Oxford Nanopore Technologies data [76] and reference genome mm10. Sam file was converted to a bam file, sorted and indexed using samtools [72]. Sequencing results were viewed in IGV [77].

Statistical testing

Data was analyzed using Prism 9 (v 9.4.1, GraphPad Software, Boston Massachusetts USA) and R [61]. Unless otherwise stated, significance between two groups was calculated with an unpaired t-test using a P-value nomenclature; *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

References

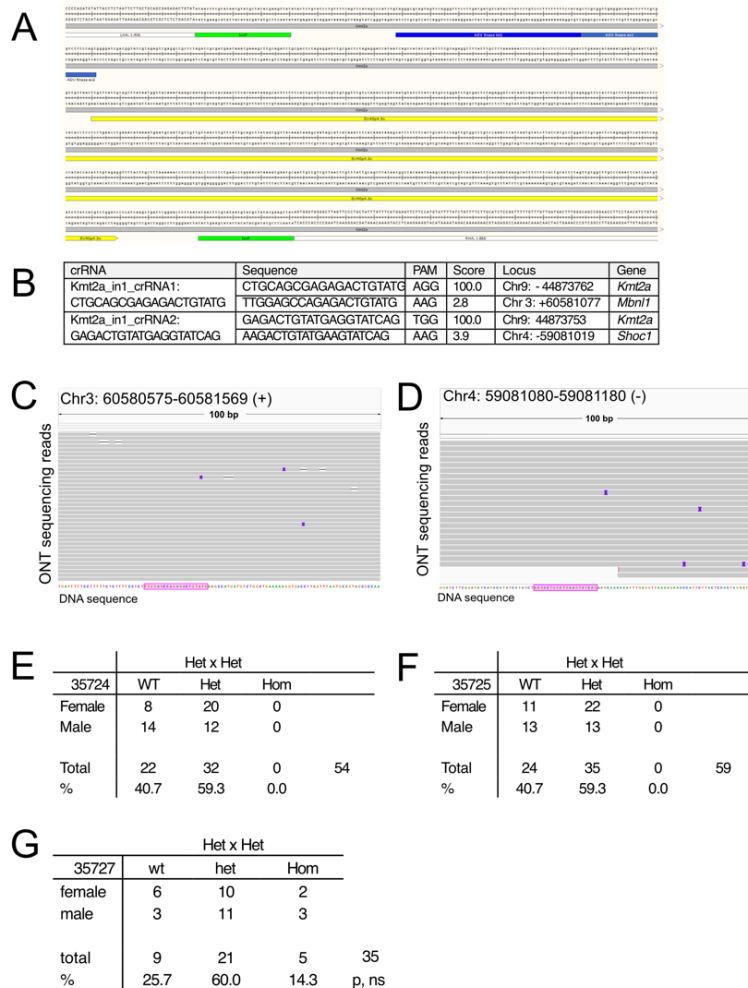
- [1] W. D. Jones *et al.*, “De novo mutations in MLL cause Wiedemann-Steiner syndrome,” *Am. J. Hum. Genet.*, vol. 91, no. 2, pp. 358–364, Aug. 2012, doi: 10.1016/j.ajhg.2012.06.008.
- [2] T. A. Milne *et al.*, “MLL targets SET domain methyltransferase activity to Hox gene promoters,” *Mol. Cell*, vol. 10, no. 5, pp. 1107–1117, Nov. 2002, doi: 10.1016/s1097-2765(02)00741-4.
- [3] T. Kusch, “Histone H3 lysine 4 methylation revisited,” *Transcription*, vol. 3, no. 6, pp. 310–314, 2012, doi: 10.4161/trns.21911.
- [4] H. Santos-Rosa *et al.*, “Active genes are tri-methylated at K4 of histone H3,” *Nature*, vol. 419, no. 6905, pp. 407–411, Sep. 2002, doi: 10.1038/nature01080.
- [5] Y.-C. Huang *et al.*, “The epigenetic factor Kmt2a/Mll1 regulates neural progenitor proliferation and neuronal and glial differentiation,” *Dev. Neurobiol.*, vol. 75, no. 5, pp. 452–462, May 2015, doi: 10.1002/dneu.22235.
- [6] D. A. Lim *et al.*, “Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells,” *Nature*, vol. 458, no. 7237, pp. 529–533, Mar. 2009, doi: 10.1038/nature07726.
- [7] J. M. Alarcón *et al.*, “Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration,” *Neuron*, vol. 42, no. 6, pp. 947–959, Jun. 2004, doi: 10.1016/j.neuron.2004.05.021.
- [8] J. Guy, J. Gan, J. Selfridge, S. Cobb, and A. Bird, “Reversal of neurological defects in a mouse model of Rett syndrome,” *Science*, vol. 315, no. 5815, pp. 1143–1147, Feb. 2007, doi: 10.1126/science.1138389.
- [9] H. T. Bjornsson *et al.*, “Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome,” *Sci. Transl. Med.*, vol. 6, no. 256, p. 256ra135, Oct. 2014, doi: 10.1126/scitranslmed.3009278.
- [10] L. Zhang *et al.*, “Inhibition of KDM1A activity restores adult neurogenesis and improves hippocampal memory in a mouse model of Kabuki syndrome,” *Mol. Ther. Methods Clin. Dev.*, vol. 20, pp. 779–791, Mar. 2021, doi: 10.1016/j.omtm.2021.02.011.
- [11] B. D. Yu, J. L. Hess, S. E. Horning, G. A. Brown, and S. J. Korsmeyer, “Altered Hox expression and segmental identity in Mll-mutant mice,” *Nature*, vol. 378, no. 6556, pp. 505–508, Nov. 1995, doi: 10.1038/378505a0.
- [12] C. Kerimoglu *et al.*, “KMT2A and KMT2B Mediate Memory Function by Affecting Distinct Genomic Regions,” *Cell Rep.*, vol. 20, no. 3, pp. 538–548, Jul. 2017, doi: 10.1016/j.celrep.2017.06.072.
- [13] C. N. Vallianatos *et al.*, “Mutually suppressive roles of KMT2A and KDM5C in behaviour, neuronal structure, and histone H3K4 methylation,” *Commun. Biol.*, vol. 3, no. 1, p. 278, Jun. 2020, doi: 10.1038/s42003-020-1001-6.
- [14] J. A. Fahrner and H. T. Bjornsson, “Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states,” *Annu. Rev. Genomics Hum. Genet.*, vol. 15, pp. 269–293, 2014, doi: 10.1146/annurev-genom-090613-094245.
- [15] R. Terranova, H. Agherbi, A. Boned, S. Meresse, and M. Djabali, “Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 17, pp. 6629–6634, Apr. 2006, doi: 10.1073/pnas.0507425103.

- [16] T. Reynisdottir, K. J. Anderson, L. Boukas, and H. T. Bjornsson, "Missense variants causing Wiedemann-Steiner syndrome preferentially occur in the KMT2A-CXXC domain and are accurately classified using AlphaFold2," *PLOS Genet.*, vol. 18, no. 6, p. e1010278, Jun. 2022, doi: 10.1371/journal.pgen.1010278.
- [17] S. E. Sheppard *et al.*, "Expanding the genotypic and phenotypic spectrum in a diverse cohort of 104 individuals with Wiedemann-Steiner syndrome," *Am. J. Med. Genet. A.*, vol. 185, no. 6, pp. 1649–1665, Jun. 2021, doi: 10.1002/ajmg.a.62124.
- [18] E. Zemheri, P. Engin Zerk, and C. Ulukaya Durakbasa, "Calretinin immunohistochemical staining in Hirschsprung's disease: An institutional experience," *North. Clin. Istanb.*, vol. 8, no. 6, pp. 601–606, 2021, doi: 10.14744/nci.2020.69376.
- [19] K. B. Shpargel, J. Starmer, C. Wang, K. Ge, and T. Magnuson, "UTX-guided neural crest function underlies craniofacial features of Kabuki syndrome," *Proc. Natl. Acad. Sci.*, vol. 114, no. 43, pp. E9046–E9055, Oct. 2017, doi: 10.1073/pnas.1705011114.
- [20] J. Schwenty-Lara, D. Nehl, and A. Borchers, "The histone methyltransferase KMT2D, mutated in Kabuki syndrome patients, is required for neural crest cell formation and migration," *Hum. Mol. Genet.*, vol. 29, no. 2, pp. 305–319, Jan. 2020, doi: 10.1093/hmg/ddz284.
- [21] K. B. Shpargel, C. L. Mangini, G. Xie, K. Ge, and T. Magnuson, "The KMT2D Kabuki syndrome histone methylase controls neural crest cell differentiation and facial morphology," *Dev. Camb. Engl.*, vol. 147, no. 21, p. dev187997, Jul. 2020, doi: 10.1242/dev.187997.
- [22] L. L. Baxter, L. Hou, S. K. Loftus, and W. J. Pavan, "Spotlight on spotted mice: a review of white spotting mouse mutants and associated human pigmentation disorders," *Pigment Cell Res.*, vol. 17, no. 3, pp. 215–224, Jun. 2004, doi: 10.1111/j.1600-0749.2004.00147.x.
- [23] Z. Sarnyai, E. L. Sibille, C. Pavlides, R. J. Fenster, B. S. McEwen, and M. Toth, "Impaired hippocampal-dependent learning and functional abnormalities in the hippocampus in mice lacking serotonin(1A) receptors," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 26, pp. 14731–14736, Dec. 2000, doi: 10.1073/pnas.97.26.14731.
- [24] A. K. Swonger and R. H. Rech, "Serotonergic and cholinergic involvement in habituation of activity and spontaneous alternation of rats in a Y maze," *J. Comp. Physiol. Psychol.*, vol. 81, no. 3, pp. 509–522, Dec. 1972, doi: 10.1037/h0033690.
- [25] J. Boisgontier *et al.*, "Anatomical and functional abnormalities on MRI in kabuki syndrome," *NeuroImage Clin.*, vol. 21, p. 101610, 2019, doi: 10.1016/j.nicl.2018.11.020.
- [26] L. Li *et al.*, "SoxD genes are required for adult neural stem cell activation," *Cell Rep.*, vol. 38, no. 5, p. 110313, Feb. 2022, doi: 10.1016/j.celrep.2022.110313.
- [27] M. Durens, M. Soliman, J. Millonig, and E. DiCicco-Bloom, "Engrailed-2 is a cell autonomous regulator of neurogenesis in cultured hippocampal neural stem cells," *Dev. Neurobiol.*, vol. 81, no. 5, pp. 724–735, Jul. 2021, doi: 10.1002/dneu.22824.
- [28] I. Schanze, D. Schanze, C. A. Bacino, S. Douzgou, B. Kerr, and M. Zenker, "Haploinsufficiency of SOX5, a member of the SOX (SRY-related HMG-box) family of transcription factors is a cause of intellectual disability," *Eur. J. Med. Genet.*, vol. 56, no. 2, pp. 108–113, Feb. 2013, doi: 10.1016/j.ejmg.2012.11.001.
- [29] N. Gharani, R. Benayed, V. Mancuso, L. M. Brzustowicz, and J. H. Millonig, "Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder," *Mol. Psychiatry*, vol. 9, no. 5, pp. 474–484, May 2004, doi: 10.1038/sj.mp.4001498.

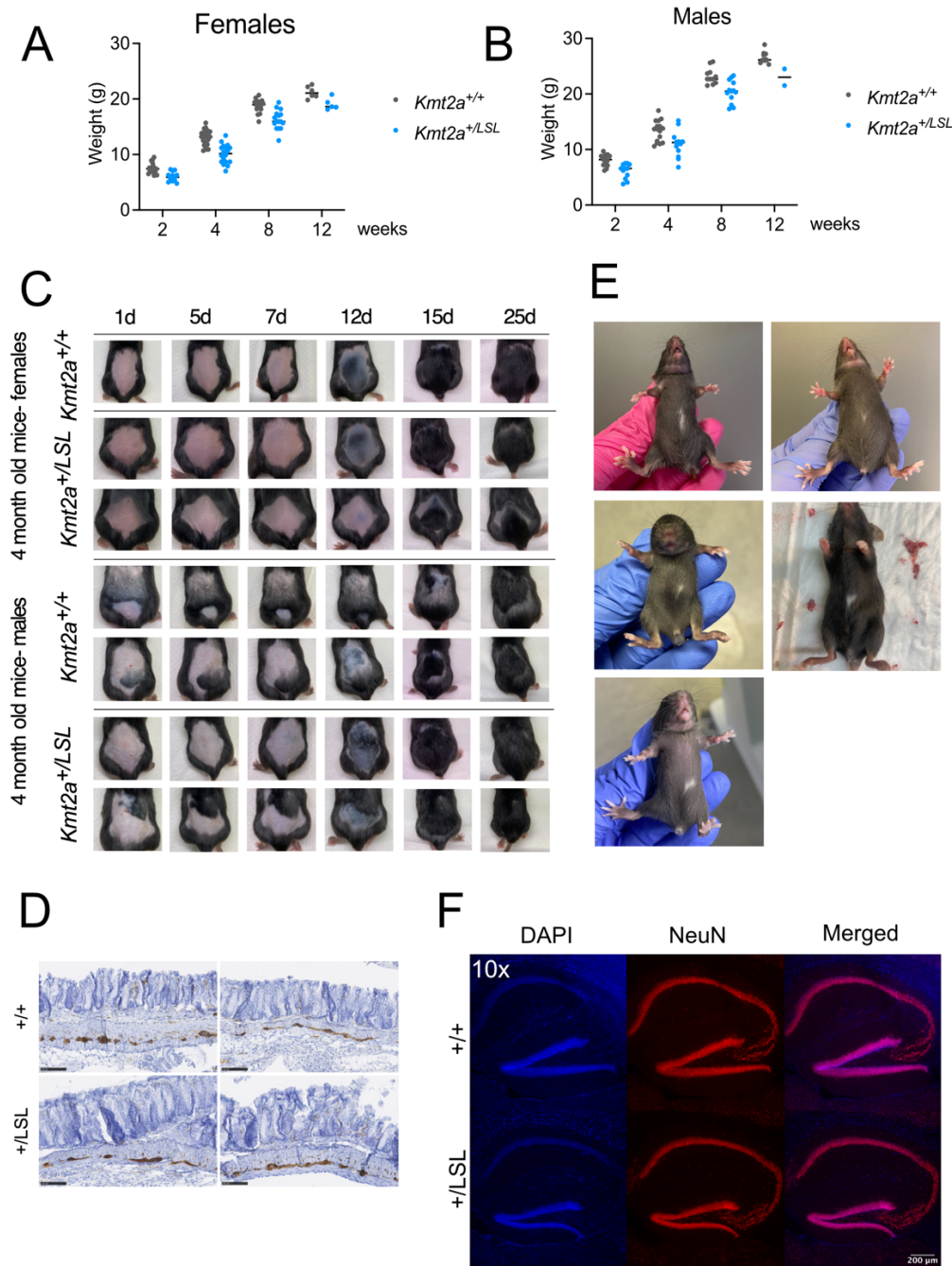
- [30] R. Ng, A. Kalinousky, J. A. Fahrner, H. T. Bjornsson, and J. Harris, "The social phenotype associated with Wiedemann-Steiner-syndrome: Autistic traits juxtaposed with high social drive and prosociality," *Am. J. Med. Genet. A.*, vol. 191, no. 10, pp. 2591–2601, Oct. 2023, doi: 10.1002/ajmg.a.63351.
- [31] S. Banerjee-Basu and A. Packer, "SFARI Gene: an evolving database for the autism research community," *Dis. Model. Mech.*, vol. 3, no. 3–4, pp. 133–135, Mar. 2010, doi: 10.1242/dmm.005439.
- [32] T. Gao, B. He, S. Liu, H. Zhu, K. Tan, and J. Qian, "EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types," *Bioinforma. Oxf. Engl.*, vol. 32, no. 23, pp. 3543–3551, Dec. 2016, doi: 10.1093/bioinformatics/btw495.
- [33] N. T. Strande and J. S. Berg, "Defining the Clinical Value of a Genomic Diagnosis in the Era of Next-Generation Sequencing," *Annu. Rev. Genomics Hum. Genet.*, vol. 17, pp. 303–332, Aug. 2016, doi: 10.1146/annurev-genom-083115-022348.
- [34] F. Morshedzadeh *et al.*, "An Update on the Application of CRISPR Technology in Clinical Practice," *Mol. Biotechnol.*, pp. 1–19, Jun. 2023, doi: 10.1007/s12033-023-00724-z.
- [35] J. Scharner and I. Aznarez, "Clinical Applications of Single-Stranded Oligonucleotides: Current Landscape of Approved and In-Development Therapeutics," *Mol. Ther.*, vol. 29, no. 2, pp. 540–554, Feb. 2021, doi: 10.1016/j.ymthe.2020.12.022.
- [36] H. Liang, S. Hippenmeyer, and H. T. Ghashghaei, "A Nestin-cre transgenic mouse is insufficient for recombination in early embryonic neural progenitors," *Biol. Open*, vol. 1, no. 12, pp. 1200–1203, Sep. 2012, doi: 10.1242/bio.20122287.
- [37] S. A. Giusti *et al.*, "Behavioral phenotyping of Nestin-Cre mice: implications for genetic mouse models of psychiatric disorders," *J. Psychiatr. Res.*, vol. 55, pp. 87–95, Aug. 2014, doi: 10.1016/j.jpsychires.2014.04.002.
- [38] Y. L. Jung, C. Hung, J. Choi, E. A. Lee, and O. Bodamer, "Characterizing the molecular impact of KMT2D variants on the epigenetic and transcriptional landscapes in Kabuki syndrome," *Hum. Mol. Genet.*, vol. 32, no. 13, pp. 2251–2261, Jun. 2023, doi: 10.1093/hmg/ddad059.
- [39] T. R. Luperchio *et al.*, "Leveraging the Mendelian disorders of the epigenetic machinery to systematically map functional epigenetic variation," *eLife*, vol. 10, p. e65884, Aug. 2021, doi: 10.7554/eLife.65884.
- [40] G. Iacono *et al.*, "Increased H3K9 methylation and impaired expression of Protocadherins are associated with the cognitive dysfunctions of the Kleefstra syndrome," *Nucleic Acids Res.*, vol. 46, no. 10, pp. 4950–4965, Jun. 2018, doi: 10.1093/nar/gky196.
- [41] G. A. Carosso *et al.*, "Precocious neuronal differentiation and disrupted oxygen responses in Kabuki syndrome," *JCI Insight*, vol. 4, no. 20, pp. e129375, 129375, Oct. 2019, doi: 10.1172/jci.insight.129375.
- [42] J. A. Fahrner *et al.*, "Precocious chondrocyte differentiation disrupts skeletal growth in Kabuki syndrome mice," *JCI Insight*, vol. 4, no. 20, pp. e129380, 129380, Oct. 2019, doi: 10.1172/jci.insight.129380.
- [43] V. Faundes, G. Malone, W. G. Newman, and S. Banka, "A comparative analysis of KMT2D missense variants in Kabuki syndrome, cancers and the general population," *J. Hum. Genet.*, vol. 64, no. 2, pp. 161–170, Feb. 2019, doi: 10.1038/s10038-018-0536-6.

- [44] H. T. Bjornsson *et al.*, “Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome,” *Sci. Transl. Med.*, vol. 6, no. 256, p. 256ra135, Oct. 2014, doi: 10.1126/scitranslmed.3009278.
- [45] C. Galichet, R. Lovell-Badge, and K. Rizzoti, “Nestin-Cre mice are affected by hypopituitarism, which is not due to significant activity of the transgene in the pituitary gland,” *PloS One*, vol. 5, no. 7, p. e11443, Jul. 2010, doi: 10.1371/journal.pone.0011443.
- [46] B. Paigen, A. Morrow, C. Brandon, D. Mitchell, and P. Holmes, “Variation in susceptibility to atherosclerosis among inbred strains of mice,” *Atherosclerosis*, vol. 57, no. 1, pp. 65–73, Oct. 1985, doi: 10.1016/0021-9150(85)90138-8.
- [47] S. Hayashi, P. Lewis, L. Pevny, and A. P. McMahon, “Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain,” *Mech. Dev.*, vol. 119 Suppl 1, pp. S97–S101, Dec. 2002, doi: 10.1016/s0925-4773(03)00099-6.
- [48] F. Tronche *et al.*, “Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety,” *Nat. Genet.*, vol. 23, no. 1, pp. 99–103, Sep. 1999, doi: 10.1038/12703.
- [49] A. Ventura *et al.*, “Restoration of p53 function leads to tumour regression in vivo,” *Nature*, vol. 445, no. 7128, pp. 661–665, Feb. 2007, doi: 10.1038/nature05541.
- [50] D. N. Feather-Schussler and T. S. Ferguson, “A Battery of Motor Tests in a Neonatal Mouse Model of Cerebral Palsy,” *J. Vis. Exp. JoVE*, no. 117, p. 53569, Nov. 2016, doi: 10.3791/53569.
- [51] A.-K. Kraeuter, P. C. Guest, and Z. Sarnyai, “The Y-Maze for Assessment of Spatial Working and Reference Memory in Mice,” *Methods Mol. Biol. Clifton NJ*, vol. 1916, pp. 105–111, 2019, doi: 10.1007/978-1-4939-8994-2_10.
- [52] A. Fedorov *et al.*, “3D Slicer as an image computing platform for the Quantitative Imaging Network,” *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012, doi: 10.1016/j.mri.2012.05.001.
- [53] Adams D, Collyer M, Kaliontzopoulou A, Baken E, “Geomorph: Software for geometric morphometric analyses. R package version 4.0.7.” 2024. [Online]. Available: <https://cran.r-project.org/package=geomorph>.
- [54] C. P. Klingenberg, M. Barluenga, and A. Meyer, “Shape Analysis of Symmetric Structures: Quantifying Variation Among Individuals and Asymmetry,” *Evolution*, vol. 56, no. 10, pp. 1909–1920, 2002, doi: 10.1111/j.0014-3820.2002.tb00117.x.
- [55] P. Gunz, P. Mitteroecker, S. Neubauer, G. W. Weber, and F. L. Bookstein, “Principles for the virtual reconstruction of hominin crania,” *J. Hum. Evol.*, vol. 57, no. 1, pp. 48–62, Jul. 2009, doi: 10.1016/j.jhevol.2009.04.004.
- [56] J. Schindelin *et al.*, “Fiji: an open-source platform for biological-image analysis,” *Nat. Methods*, vol. 9, no. 7, pp. 676–682, Jul. 2012, doi: 10.1038/nmeth.2019.
- [57] S. N. Bernas, O. Leiter, T. L. Walker, and G. Kempermann, “Isolation, Culture and Differentiation of Adult Hippocampal Precursor Cells,” *Bio-Protoc.*, vol. 7, no. 21, p. e2603, Nov. 2017, doi: 10.21769/BioProtoc.2603.
- [58] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, “Near-optimal probabilistic RNA-seq quantification,” *Nat. Biotechnol.*, vol. 34, no. 5, pp. 525–527, May 2016, doi: 10.1038/nbt.3519.
- [59] F. Ramírez *et al.*, “deepTools2: a next generation web server for deep-sequencing data analysis,” *Nucleic Acids Res.*, vol. 44, no. W1, pp. W160-165, Jul. 2016, doi: 10.1093/nar/gkw257.

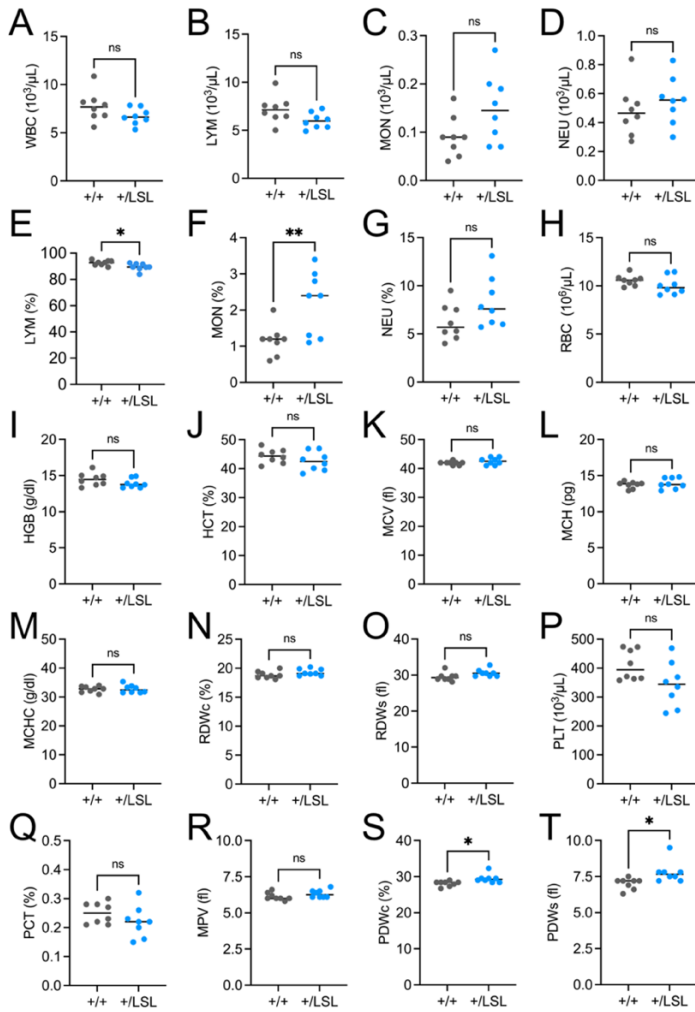
- [60] W. J. Kent *et al.*, “The human genome browser at UCSC,” *Genome Res.*, vol. 12, no. 6, pp. 996–1006, Jun. 2002, doi: 10.1101/gr.229102.
- [61] R Core Team, “R: A language and environment for statistical computing.” R Foundation for Statistical Computing, Vienna, Austria., 2021. [Online]. Available: <https://www.R-project.org/>
- [62] C. Soneson, M. I. Love, and M. D. Robinson, “Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences,” *F1000Research*, vol. 4, p. 1521, 2015, doi: 10.12688/f1000research.7563.2.
- [63] S. Durinck, P. T. Spellman, E. Birney, and W. Huber, “Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt,” *Nat. Protoc.*, vol. 4, no. 8, pp. 1184–1191, 2009, doi: 10.1038/nprot.2009.97.
- [64] M. I. Love, W. Huber, and S. Anders, “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2,” *Genome Biol.*, vol. 15, no. 12, p. 550, 2014, doi: 10.1186/s13059-014-0550-8.
- [65] K. Strimmer, “fdrtool: a versatile R package for estimating local and tail area-based false discovery rates,” *Bioinforma. Oxf. Engl.*, vol. 24, no. 12, pp. 1461–1462, Jun. 2008, doi: 10.1093/bioinformatics/btn209.
- [66] Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, and B. Zhang, “WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs,” *Nucleic Acids Res.*, vol. 47, no. W1, pp. W199–W205, Jul. 2019, doi: 10.1093/nar/gkz401.
- [67] Z. Gu, R. Eils, and M. Schlesner, “Complex heatmaps reveal patterns and correlations in multidimensional genomic data,” *Bioinforma. Oxf. Engl.*, vol. 32, no. 18, pp. 2847–2849, Sep. 2016, doi: 10.1093/bioinformatics/btw313.
- [68] “The Landscape of Isoform Switches in Human Cancers - PubMed.” Accessed: May 13, 2024. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28584021/>
- [69] T. Hulsen, “DeepVenn – a web application for the creation of area-proportional Venn diagrams using the deep learning framework Tensorflow.js.” Department of Hospital Services & Informatics, Philips Research, Eindhoven, the Netherlands, 2022.
- [70] Shen L, Sinai I SoMaM, “GeneOverlap: Test and visualize gene overlaps.” 2024.
- [71] B. Langmead and S. L. Salzberg, “Fast gapped-read alignment with Bowtie 2,” *Nat. Methods*, vol. 9, no. 4, pp. 357–359, Mar. 2012, doi: 10.1038/nmeth.1923.
- [72] H. Li *et al.*, “The Sequence Alignment/Map format and SAMtools,” *Bioinforma. Oxf. Engl.*, vol. 25, no. 16, pp. 2078–2079, Aug. 2009, doi: 10.1093/bioinformatics/btp352.
- [73] Y. Zhang *et al.*, “Model-based analysis of ChIP-Seq (MACS),” *Genome Biol.*, vol. 9, no. 9, p. R137, 2008, doi: 10.1186/gb-2008-9-9-r137.
- [74] R. Stark and G. Brown, “DiffBind: Differential binding analysis of ChIP-Seq peak data”.
- [75] “Exploring Epigenomic Datasets by ChIPseeker - PubMed.” Accessed: May 13, 2024. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/36286622/>
- [76] H. Li, “Minimap2: pairwise alignment for nucleotide sequences,” *Bioinforma. Oxf. Engl.*, vol. 34, no. 18, pp. 3094–3100, Sep. 2018, doi: 10.1093/bioinformatics/bty191.
- [77] J. T. Robinson *et al.*, “Integrative genomics viewer,” *Nat. Biotechnol.*, vol. 29, no. 1, pp. 24–26, Jan. 2011, doi: 10.1038/nbt.1754.



Supplemental Figure 1: Mouse DNA sequencing data (A) DNA sequence of loxP-stop-loxP cassette insertion in intron 1 of *Kmt2a* out listed by Jackson Laboratory. (B) Table lists DNA sequence of CRISPR RNA (crRNA) for loxP-stop-loxP cassette insertion using CRISPR/Cas9, the DNA sequence targeted by the crRNA, their PAM site, off targeting score and location. Off-target scoring was performed by Jackson Laboratory, with higher off-target editing risk set at score >20.0 for unlinked non-canonical PAM (NAG) in coding region. Oxford Nanopore Technologies (ONT) sequencing results of potential CRISPR off-target sites viewed in Integrative Genomics Viewer (IGV) showing no off-target effects for (C) chromosome 3 and (D) chromosome 4. Het-het breeding ratio of *Kmt2a*^{+/*LSL*} mice, (E) stock #35724 C57BL/6J-*Kmt2a*^{em7*Lutzy*/J} and (F) stock #37525 C57BL/6J-*Kmt2a*^{em8*Lutzy*/J}, demonstrating no homozygous pups born, confirming the model's lethality in homozygosity. (G) Breeding ratio of post-Cre het-het mice (Stock #35727 C57BL/6J-*Kmt2a*^{em7.1*Lutzy*/J}), showing Sox2-Cre Tg is not segregating in these mice.



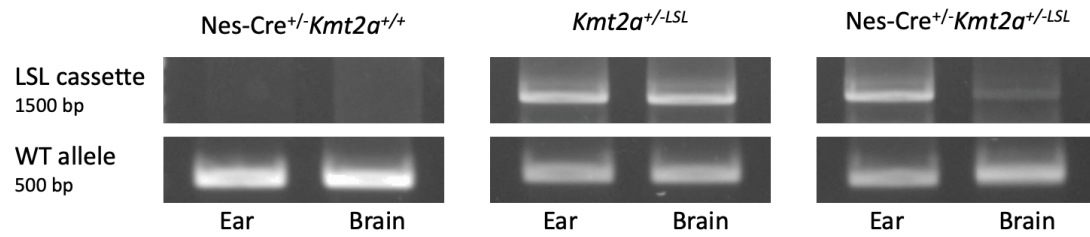
Supplemental Figure 2: Additional phenotyping data of $Kmt2a^{+/LSL}$ mice. Weight data from $Kmt2a^{+/+}$ and $Kmt2a^{+/LSL}$ mice at 2, 4, 8 and 12 weeks for (A) females and (B) males. (C) Hair regrowth after hair removal for 4-month-old mice (n=2 for each genotype per sex). (D) Calretinin staining of the distal colon of $Kmt2a^{+/+}$ and $Kmt2a^{+/LSL}$ mice. Both genotypes show positive calretinin staining (brown), indicative of the presence ganglion cells. Scale bar is 100 μ m. (E) Examples of white belly spots present in $Kmt2a^{+/LSL}$ mice. (F) Representative figures from NeuN staining, scale bar is 200 μ m.



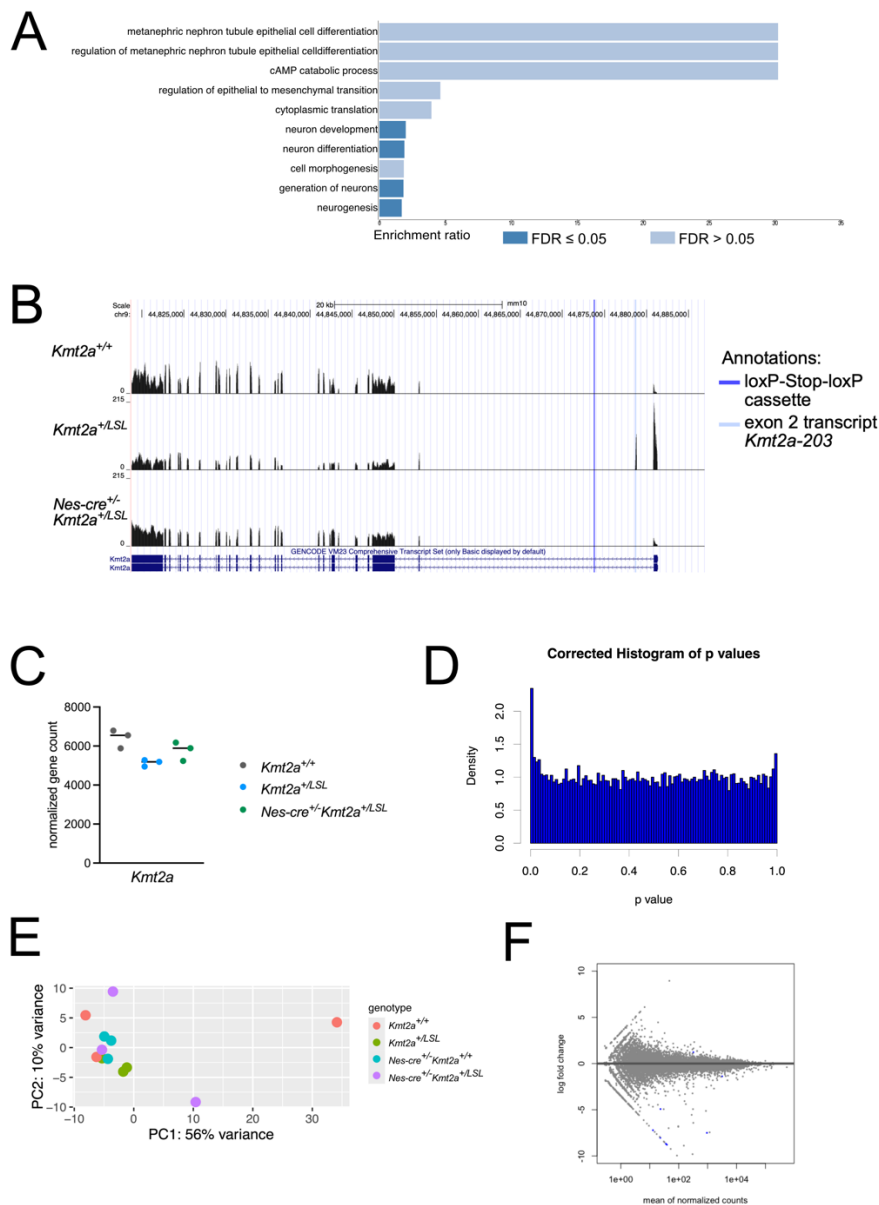
23

24 **Supplemental Figure 3: Results of Complete Blood Count (CBC) of $Kmt2a^{+/LSL}$ mice compared**
 25 **to $Kmt2a^{+/+}$ littermates (n=8 per genotype) (A) WBC - Total white blood cell count ($10^3/\mu L$) (B)**
 26 **LYM - lymphocyte count ($10^3/\mu L$) (C) MON - monocyte count ($10^3/\mu L$) (D) NEU - neutrophil**
 27 **count ($10^3/\mu L$) (E) LYM% — lymphocyte percentage (%) (F) MON% — monocyte percentage**
 28 **(%) (G) NEU (%) — neutrophil percentage (%) (H) RBC — red blood cell count ($10^6/\mu L$) (I) HGB**
 29 **— hemoglobin (g/dl) (J) HCT — hematocrit (%) (K) MCV — mean corpuscular volume (fl) (L)**
 30 **MCH — mean corpuscular hemoglobin (pg) (M) MCHC — mean corpuscular hemoglobin**
 31 **concentration (g/dl) (N) RDWc, red cell distribution width, coefficient of variation (%) (O)**
 32 **RDWs red cell distribution width, std dev (fl) (P) PLT — platelet count ($10^3/\mu L$) (Q) PCT —**
 33 **platelet crit (%) (R) MPV — mean platelet volume (fl) (S) PDWc, — platelet distribution width,**
 34 **coefficient of variation (%) (T) PDWs, platelet distribution width, std dev (fl). ns: not**
 35 **significant, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.**

36

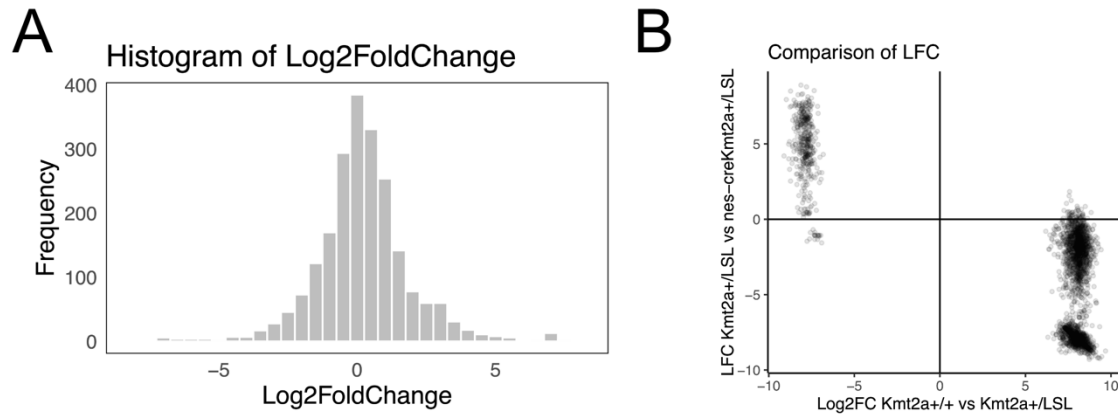


Supplemental Figure 4: PCR products from gDNA samples from Nes-Cre^{+/-}Kmt2a^{+/+}, Kmt2a^{+/-}LSL and Nes-Cre^{+/-}Kmt2a^{+/-}LSL mice run on an agarose gel, comparing visually the presence of LSL cassette in the nervous system (hippocampus) and outside the nervous system (ear clips). There is removal of the LSL cassette in the Nes-Cre^{+/-}Kmt2a^{+/-}LSL mice.



44

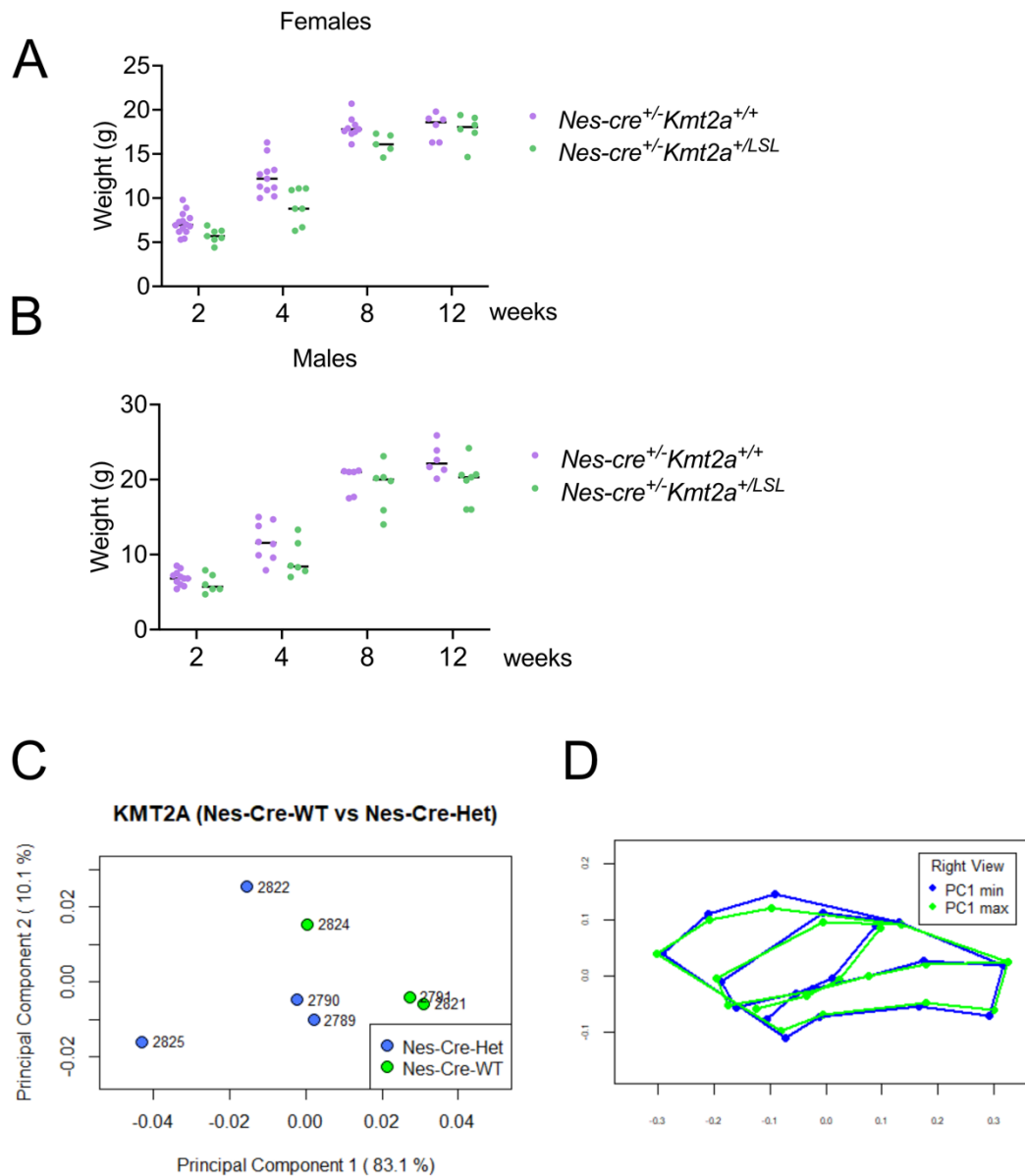
45 **Supplemental Figure 5: RNA sequencing results.** (A) Over representation analysis of the top
 46 decile rescued genes. Figure shows pathways enriched in dark blue (with FDR<0.05) (B) UCSC
 47 browser tracks of genome coverage of RNA seq files showing a decrease of *Kmt2a* coverage
 48 in *Kmt2a*^{+/-LSL} cells for full length transcripts. Only in *Kmt2a*^{+/-LSL} cells is there a presence of a
 49 peak mapped to exon 2 of the short *Kmt2a-203* transcript. (C) Normalized gene count values
 50 for *Kmt2a* from RNA seq data. (D) Histogram of corrected p-values from differential
 51 expression analysis of RNA seq data. (E) PCA plot of the expression matrix of all genotypes (F)
 52 MA plot of log2 fold change values plotted against mean of normalized counts.



53

54 **Supplemental Figure 6:** CUT&RUN results. (A) Histogram of log₂ fold changes for gene
 55 expression of genes linked to neuronal enhancers. (B) Correlation plot showing the
 56 correlation between H3K4me1 changes between *Kmt2a*^{+/+} vs *Kmt2a*^{+/LSL} and then between
 57 *Kmt2a*^{+/LSL} vs Nes-Cre^{+/+} *Kmt2a*^{+/LSL} contrasts.

58



Supplemental Figure 7: Additional phenotyping data from the $Nes-Cre^{+/-}Kmt2a^{+/LSL}$ model compared to littermates. Weight data from $Nes-Cre^{+/-}Kmt2a^{+/LSL}$ and $Nes-Cre^{+/-}Kmt2a^{+/+}$ (A) female and (B) male mice. (C) PCA of craniofacial abnormalities of $Nes-Cre^{+/-}Kmt2a^{+/LSL}$ vs. $Nes-Cre^{+/-}Kmt2a^{+/+}$ shows separation of genotypes along PC1. (D) Similar to the previous analysis (Figure 1E-H), the $Nes-Cre^{+/-}Kmt2a^{+/LSL}$ mice are positioned at the negative end of PC1 and characterized by increased height and width of the neurocranium, a shortened midface, and ventral bowing of the cranium.

Supplemental Table 1: Significant DEGs between *Kmt2a*^{+/+} and *Kmt2a*^{+/*LSL*} samples with their corrected p value and corrected q value.

gene	cor.pval	cor.qval
Grb10	2,22E-02	3,84E+02
Uck2	3,77E-01	3,27E+03
Hspa2	4,95E-01	2,34E+04
Gm10705	5,86E+00	2,54E+04
Nek7	2,49E+00	8,61E+04
Npcd	4,11E+04	1,17E+07
Gm10108	7,50E+04	1,85E+08
Gm14494	1,76E+05	3,55E+08
Ppp1r3b	2,03E+02	3,90E+08
Psd4	4,35E+05	7,30E+08
Fat1	5,56E+05	8,74E+08
Nav1	8,48E+05	1,19E+09
Nav1	8,48E+05	1,19E+09
Ppp1r12a	1,03E+06	1,35E+09
Dmrt3	1,20E+06	1,48E+09
Mtus2	4,05E+06	4,67E+09
Ulk4	9,75E+06	0.000105415484501709
Rpgrip1	2,34E+07	0.000232711304448866
Gm11585	3,62E+07	0.000333856437024141
Gm10146	3,84E+07	0.000349626501343152
Xlr3a	5,73E+07	0.000483111694768594
Ankef1	6,31E+07	0.000520298986354538
Nrg1	8,32E+07	0.000654097876379912
Nacad	1,46E+08	0.00104756306019184
Slc37a4	1,61E+08	0.00113207581224977
Slc26a8	1,72E+08	0.00118784184076258
Gm10052	2,34E+07	0.00155937239945776
AC132444.6	3,37E+08	0.00215623552823898
Mt3	4,40E+08	0.00271434552172039
Car10	5,35E+08	0.00319099208705152
Aim2	6,57E+08	0.00378842894922863
Zfp536	8,92E+08	0.0049609157295213
Retn	1,10E+09	0.00593975471499562
Gm6206	1,40E+09	0.00722455224923782
Gm6984	1,74E+09	0.00856300820852473
Clhc1	2,03E+09	0.00961037169338642
Zfp964	2,29E+09	0.0104944557404484
Gm14824	2,31E+09	0.0105628304819891
Pcdh15	2,36E+09	0.0107111928879198
Hpgds	2,67E+09	0.0116734310069822

Pmaip1	2,76E+09	0.0119263786899909
Prom2	3,37E+09	0.0138355071486223
Gm12439	3,67E+08	0.0147033301790239
Mcmdc2	3,90E+09	0.0153361224771457
Tmem151a	4,26E+07	0.0162845100784377
Rmdn2	4,54E+09	0.0170013954908102
Cachd1	4,75E+09	0.0175118860285515
Sftpb	5,32E+09	0.0188024030130518
Thbs4	5,34E+09	0.0188485859279124
Npsr1	5,36E+09	0.0188890460364386
4930503B20Rik	5,79E+09	0.0197961633065591
Ccdc177	5,91E+09	0.0200394127163123
Leap2	7,29E+09	0.023634832188955
Gm10012	9,21E+09	0.0281250984410246
Slitrk1	9,56E+09	0.0288959677479833
mt-Atp8	9,89E+09	0.0296071011870187
MIlt10	0.000101994143252737	0.0302507417582987
Wdr7	0.000104657435637368	0.0308003041259233
Map3k5	0.000105129884838684	0.0308969050438718
Sfrp2	0.000109330134651087	0.0317442124725863
Dph7	0.000111719994918102	0.0322172627971495
Klf12	0.000141310553636309	0.0390139085289729
Kif13b	0.000148996684610214	0.0406855980319235
Nlrc5	0.000180071696960882	0.0470869103804284

Supplemental Table 2: Gene list of top decile genes based on p-value between *Kmt2a*^{+/+} vs *Kmt2a*^{+/LSL} and then between *Kmt2a*^{+/LSL} vs Nes-Cre^{+/+}-*Kmt2a*^{+/LSL}.

1110004E09Rik	Col18a1	Hspb8	Prdx2	Tsnaxip1
1500009L16Rik	Col28a1	Hspe1	Prelp	Ttc25
1700016D06Rik	Col6a4	Htra1	Prkca	Ttc9
1700021F05Rik	Col7a1	Ifrd1	Prom2	Txnl4a
1700024J04Rik	Cox16	Ift81	Prtg	Tyw3
1700037C18Rik	Cox6c	Igsf9b	Psma2	Uap1l1
1700109H08Rik	Cox7b	Il2ra	Psmb9	Uba52
2010107E04Rik	Cox7c	Ino80d	Psme1	Ubl5
2900026A02Rik	Cpxm2	Irf7	Ptpro	Ulk4
3110052M02Rik	Crlf1	Islr	Pus10	Unc5d
4930453N24Rik	Csrp2	Itgb3bp	Pxdc1	Uqcrb
4930503B20Rik	Ctnnd2	Itpr2	Rab25	Uqcrh
5330417C22Rik	Cutal	Junb	Rab40b	Usp3
A830010M20Rik	Cyb5r2	Kcnma1	Ranbp9	Usp35
A830039N20Rik	Cyba	Kirrel2	Rasgef1c	Uvrag

Aard	D130040H23Rik	Krcc1	Rdh19	Uvssa
Abrac1	D430041D05Rik	Lama2	Reck	Vit
AC132444.6	Dab2	Lctl	Recql5	Vmn1r227
Acan	Dcps	Lect1	Retn	Vmn1r84
Acy3	Dennd3	Lif	Rftn2	Vmn2r102
Adamts9	Denr	Lpin1	Rmdn2	Vps45
Adcy3	Dhrs9	Lpl	Rpl18a	Vwf
Adnp	Dmrt3	Lrp1b	Rpl30	Wdr7
Ago3	Dock9	Lrrc24	Rpl37	Wisp1
Agtpbp1	Dph3	Lrrc7	Rpl37a	Xlr3a
Aifm2	Dscam	Lrrk1	Rpl37rt	Zan
Aim2	Eif1ax	Magi2	Rpl38	Zbed5
Akr1b7	En1	Magix	Rpl39	Zbtb42
Akr1b8	En2	Map3k6	Rps12	Zc3hav1
Akr1d1	Epha4	Marc2	Rps20	Zcchc2
Aldh1a7	Erich2	Masp2	Rps27	Zeb1
Amph	Fam49a	Mbd1	Rps28	Zfp109
Ang	Fanca	Mcmdc2	Rps29	Zfp354c
Angptl1	Fap	Mdk	Rrp1b	Zfp433
Ank3	Fat4	Mdn1	Sap30bp	Zfp536
Arhgap15	Fbxl17	Mfap4	Schip1	Zfp777
Arhgap39	Fbxo17	Mfsd2a	Sec22c	Zic3
Arl6ip6	Fech	Mfsd9	Sec24d	Zpr1
Arrdc3	Fgf14	Mier1	Sema4b	
Ascl1	Fkbp11	Miip	Sfrp1	
Atp2a3	Flt1	MIph	Sfrp2	
Atp5b	Fndc7	Morc4	Sftpb	
Atp5h	Frmpd3	Mpc2	She	
Atp5j	Frs3	Mpv17l	Shfm1	
AW551984	Fryl	Mrpl47	Sipa1l2	
B230118H07Rik	Fyb	mt-Atp8	Slc22a17	
B2m	Gbx2	Mt3	Slc35f1	
Bai1	Gcnt2	Mtcl1	Slc4a10	
BC024139	Gm10012	Mtus2	Slc4a1ap	
Bcas2	Gm10052	Myh13	Slc9b1	
Bcat1	Gm10146	Myl12b	Slco1a5	
Bckdhb	Gm10481	Naaa	Slirp	
Bcl2l11	Gm1070	Nap1l1	Slitrk1	
Bcl2l15	Gm10985	Nckap5l	Smarce1	
Bcor	Gm11244	Ndn	Snrpg	
Bhlhe22	Gm11459	Ndufa1	Sntg1	
Bnpl	Gm11585	Ndufa6	Sostdc1	
Bola2	Gm11960	Ndufab1	Sox5	

Bola3	Gm12481	Ndufc1	Sp9
Brcc3	Gm12734	Nek7	Spaca6
Brpf3	Gm13136	Neto1	Spata5
Btd	Gm13378	Nmd3	Sqrdl
C2cd5	Gm13597	Nmi	Ssu72
C77370	Gm13722	Nr3c2	St3gal1
Cadps2	Gm15425	Nrbp2	St6gal2
Camk2a	Gm16487	Nsf	Stab2
Camk4	Gm17131	Ntn4	Stambpl1
Capn11	Gm17132	Nxpe2	Stk32b
Car10	Gm20517	Oas3	Syt12
Car6	Gm21119	Olfr292	Tbx15
Carns1	Gm21897	Pacrg	Tcf15
Cask	Gm27029	Palb2	Tcf15
Casp8	Gm3604	Palm3	Tdp2
Casp8ap2	Gm5403	Parg	Tef
Ccdc101	Gm5884	Pax8	Tgfb1i1
Ccdc114	Gm6863	Pcbp3	Thbs4
Ccdc177	Gm7061	Pcdh15	Thg1l
Ccdc67	Gm7120	Pcdhb19	Timm8b
Ccdc73	Gm7332	Pcdhb3	Tm7sf2
Ccdc8	Gm7964	Pdcd5	Tmem132e
Ccdc80	Gm9864	Pde7a	Tmem140
Ccdc85a	Gng5	Pde7b	Tmem150b
Cd164l2	Gpr26	Pde8a	Tmem151a
Cdipt	Gpx7	Pdpn	Tmem176a
Cdk5rap1	Grb10	Pdzd3	Tmem33
Cdkl4	Gsn	Phxr2	Tmem35
Cdkn1c	Gstt1	Pkd2l1	Tmprss5
Ceacam18	Gstt3	Pkhd1	Tnfrsf22
Cenpu	H2-T10	Pkp4	Tnfsf12Tnfsf13
Cenpv	Hecw2	Pla2g16	Toporsos
Cept1	Heyl	Plagl1	Tpi1
Cfap44	Hipk1	Pld5	Tpm4
Cfap46	Hist1h3e	Plekhg1	Tpte
Cfap46	Hlx	Plp1	Traf3ip2
Cfap46	Hmga1	Pnrc1	Trim25
Cfap46	Hnrnpa3	Pole4	Trim6
Cherp	Hnrnph3	Ppfibp1	Trim65
Chn1	Hpn	Ppm1h	Trp53i11
Clhc1	Hsbp1	Ppp1r3b	Trpc3
Cntn5	Hsbp1l1	Prdm4	Tsku

Supplemental Table 3: Genomic location of significant differential H3K4me1 peaks between *Kmt2a*^{+/+} and *Kmt2a*^{+/-LSL} samples, corrected p-value and the genes they map to (.csv).

Chr	Start	Stop	Corrected_pvalue	Gene_id	Chr	Start	Stop	Corrected_pvalue	Gene_id
chr1	3120835	3121235	0.00097196515011766	Xkr4	chr1	28428839	28429239	0.000832917078844142	Glis3
chr1	3300768	3301168	0.000962118234820952	Xkr4	chr1	28583490	28583890	0.000903023385756319	Glis3
chr1	4616454	4616854	0.000445006575764406	Sox17	chr1	28895761	28896161	0.000924659668215089	Slc1a1
chr1	4993765	4994165	0.000344452915516102	Rgs20	chr1	29648976	29649376	0.000237135078951844	Ermp1
chr1	9738727	9739127	0.00090908843052051	1700034P13Rik	chr1	31085837	31086237	0.000551414128252992	Cstf2t
chr1	12887293	12887693	0.000911480149082378	Slco5a1	chr1	31315451	31315851	0.000400568151442959	Cstf2t
chr1	14202732	14203132	6,33E+08	Eya1	chr1	32774454	32774854	0.000652095609755543	Pten
chr1	16304141	16304541	0.000273871575472429	Stau2	chr1	34550232	34550632	0.000286759333585129	Ifit2
chr1	21896868	21897268	0.000552840189420145	Kcnq5	chr1	34877279	34877679	0.000788123667968321	Pank1
chr1	31014365	31014765	0.000814492241065334	Ptp4a1	chr1	36895708	36896108	0.000861078337799359	Tnks2
chr1	34312260	34312660	0.000396658940466033	Dst	chr1	36903679	36904079	0.00049894430589803	Fgfbp3
chr1	34726424	34726824	0.000565504231857084	Arhgef4	chr1	37033897	37034297	0.000870888448555984	Btaf1
chr1	37432960	37433360	0.000987675656124232	Unc50	chr1	37088705	37089105	0.000922705836872595	A330032B11Rik
chr1	37891115	37891515	0.000327645588079184	Mrpl30	chr1	46128296	46128696	0.000584803267431822	Elovl3
chr1	37892179	37892579	0.000844755842559275	Mrpl30	chr1	46168557	46168957	1,89E+09	Gbf1
chr1	38110426	38110826	0.000302795726847771	Rev1	chr1	46477091	46477491	0.000126497520506419	Sufu
chr1	39023203	39023603	0.000643534545389901	Pdcl3	chr1	46600242	46600642	0.000958252773394674	Wbp1l
chr1	39550222	39550622	0.000969716481320937	Snord89	chr1	46690370	46690770	0.00021129551131005	Borcs7
chr1	41156437	41156837	0.000741072456208913	4930448I06Rik	chr1	46789264	46789664	0.000822877206354988	Cnnm2
chr1	41284239	41284639	0.000549679449168394	4930448I06Rik	chr1	47179903	47180303	0.000691819408377148	Neur1a
chr1	41413112	41413512	0.000969576439182607	4930448I06Rik	chr1	47200916	47201316	0.000642601849980551	Neur1a
chr1	41907785	41908185	0.000236333542880995	4930448I06Rik	chr1	49267558	49267958	0.000599524149148989	Sorcs3
chr1	42873481	42873881	0.000652880679296475	Mrps9	chr1	50233551	50233951	0.000764738977032886	Sorcs1
chr1	43266454	43266854	0.000763498697585518	Fhl2	chr1	50668381	50668781	0.000422468640350226	Sorcs1
chr1	43316922	43317322	0.000864184678810442	Fhl2	chr1	51868485	51868885	0.000728438460790182	Ins1
chr1	45797797	45798197	0.000869731360305437	Wdr75	chr1	53443784	53444184	0.000927717418317986	Mirt1
chr1	50865373	50865773	0.000294133035817934	Tmeff2	chr1	53452286	53452686	0.000749082962298697	Mirt1
chr1	51272668	51273068	0.000680806476867746	Cavin2	chr1	53455498	53455898	0.000434365454708763	Mirt1
chr1	52292559	52292959	0.000513651595826836	Gls	chr1	54076880	54077280	0.000754128121104536	Adra2a
chr1	53426106	53426506	0.000603423674402803	Asnsd1	chr1	55555741	55556141	7,78E+09	Vti1a
chr1	57267284	57267684	9,42E+09	BC055402	chr1	56874628	56875028	0.000179628757132511	Vwa2
chr1	58961340	58961740	0.000436679561633291	Trak2	chr1	57015659	57016059	0.000362542546954892	Afap1l2
chr1	59229613	59230013	0.000914268613557034	Als2	chr1	58217533	58217933	8,87E+08	Mir5623
chr1	61473114	61473514	0.000646197484807765	9530026F06Rik	chr2	4381485	4381885	0.000955767842558686	Frmd4a
chr1	61648721	61649121	0.000598250472406914	Pard3b	chr2	6574929	6575329	0.000102198933154218	Celf2
chr1	62701101	62701501	0.000877279831634992	Nrp2	chr2	6681227	6681627	0.000609838575436727	Celf2
chr1	62708170	62708570	0.000406529023994961	Nrp2	chr2	6794077	6794477	0.000792545634463293	Celf2
chr1	62783144	62783544	0.00041049081238187	Nrp2	chr2	8673183	8673583	0.000638423347628203	1700061F12Rik

chr1	6402281 4	64023214	0.000632700804064 523	Mir6899	chr2	11150885	11151285	0.000940044096860 104	Prkcq
chr1	6829567 2	68296072	0.000593093856809 723	ErbB4	chr2	11426268	11426668	4,52E+09	8030442B05R ik
chr1	6959424 5	69594645	0.000902875124741 52	Ikzf2	chr2	14615404	14615804	0.000221106868748 322	Cacnb2
chr1	7147734 4	71477744	0.000191669299651 975	Abca12	chr2	17644541	17644941	0.000263025587709 187	Mir6419
chr1	7157146 8	71571868	0.000184300560047 479	Atic	chr2	17900886	17901286	0.000878355130553 787	H2al2a
chr1	7180082 5	71801225	0.000734317476540 18	Gm8883	chr2	18374304	18374704	0.000361397016590 326	Dnajc1
chr1	7235714 0	72357540	0.000395151221511 68	Xrcc5	chr2	19661494	19661894	0.000251274583129 346	Otud1
chr1	7274066 0	72741060	0.000251400358015 63	Rpl37a	chr2	21206923	21207323	0.000172582107906 383	Thnsl1
chr1	7308541 8	73085818	0.000867523769723 988	Teshl	chr2	21475082	21475482	0.000736838889378 921	Gpr158
chr1	7324414 5	73244545	0.000415857973516 509	D530049I02R ik	chr2	24468910	24469310	0.000412198517995 233	Pax8
chr1	7374290 3	73743303	0.000177911130344 599	6030407O03 Rik	chr2	24534696	24535096	0.000994981607256 307	Pax8
chr1	7396180 7	73962207	0.000825721936286 605	Tns1	chr2	26912474	26912874	0.000254754136361 113	Rpl7a
chr1	7434009 5	74340495	0.000162227086277 555	Pnkd	chr2	27867128	27867528	0.000900468832111 923	Col5a1
chr1	7477852 7	74778927	0.000832386971610 921	Wnt6	chr2	30649525	30649925	0.000178169449540 411	Gm14486
chr1	7608587 2	76086272	0.000634475518314 748	Gm17751	chr2	34469249	34469649	0.000495924523176 15	Mapkap1
chr1	7716477 3	77165173	0.000780156921094 344	Epha4	chr2	35490596	35490996	0.000321860798818 019	Ggta1
chr1	7737717 6	77377576	8,38E+09	Epha4	chr2	37708499	37708899	0.000293205039758 826	Strbp
chr1	7758513 3	77585533	0.000852743182571 558	Epha4	chr2	38018710	38019110	0.000651942107927 184	Dennd1a
chr1	7942526 4	79425664	0.000628478188976 933	Scg2	chr2	38581893	38582293	0.000578117671921 508	Psmb7
chr1	8047616 5	80476565	0.000549872636422 366	1700016L21R ik	chr2	38669031	38669431	0.000452979322787 872	Psmb7
chr1	8062993 3	80630333	0.000494170395465 51	Dock10	chr2	40820784	40821184	0.000906039333149 922	Lrp1b
chr1	8066991 8	80670318	5,81E+09	Dock10	chr2	41229835	41230235	0.000921675623755 78	Lrp1b
chr1	8070315 4	80703554	0.000654641862887 217	Dock10	chr2	42059602	42060002	0.000825290613583 718	Lrp1b
chr1	8086079 7	80861197	1,66E+09	Dock10	chr2	44419817	44420217	0.000699362596696 673	Gtdc1
chr1	8191297 3	81913373	0.000492505302398 083	Mir6344	chr2	45007197	45007597	0.000275748517941 876	Zeb2
chr1	8268770 1	82688101	0.000967101651141 444	Col4a3	chr2	52593781	52594181	0.000870795857216 278	Cacnb4
chr1	8441927 3	84419673	0.000494915698690 424	Pid1	chr2	52908859	52909259	0.000171142121736 745	Fmnl2
chr1	8674440 3	86744803	0.000748058944253 671	Dis3l2	chr2	53037911	53038311	0.000394777879235 031	Prpf40a
chr1	8826954 2	88269942	0.000311738443334 431	Hjulp	chr2	56114264	56114664	0.000294478257267 583	Mir195b
chr1	8863021 1	88630611	6,19E+09	Arl4c	chr2	59315320	59315720	0.000532808251911 598	Pkp4
chr1	8981915 9	89819559	0.000843418101226 678	Agap1	chr2	59617765	59618165	0.000240464958640 496	Tanc1
chr1	9148324 9	91483649	0.000465946459712 541	Traf3ip1	chr2	59705428	59705828	0.000806007165708 111	Tanc1
chr1	9217608 8	92176488	0.000335783328883 51	Hdac4	chr2	60284474	60284874	0.000646275359347 914	Cd302
chr1	9267693 4	92677334	0.000747055532809 737	Otos	chr2	60613353	60613753	0.000585954637504 438	Itgb6
chr1	9334675 7	93347157	0.000511574508726 076	Mir6902	chr2	62437175	62437575	0.000288456659965 242	Dpp4
chr1	9792370 9	97924109	0.000496633257211 831	Pam	chr2	62643984	62644384	0.000542432611265 298	Ifih1
chr1	1017936 47	10179404	0.000728699570823 11	Gm20268	chr2	67874953	67875353	0.000475582188526 502	Mir6335
chr1	1046137 50	10461415 0	7,52E+09	Cdh20	chr2	70208308	70208708	0.000780250162688 123	Myo3b
chr1	1066899 45	10669034 5	0.000920375759720 926	Bcl2	chr2	71756121	71756521	2,22E+09	Itga6
chr1	1256884 23	12568882 3	9,55E+09	Gpr39	chr2	72028598	72028998	0.000630386851336 873	Rapgef4

chr1	1257618 65	12576226 5	0.000354862832943 947	Gpr39	chr2	76206517	76206917	0.000347030633071 51	Pde11a
chr1	1264403 59	12644075 9	0.000343228588384 87	Nckap5	chr2	76313277	76313677	0.000593925649950 977	Pde11a
chr1	1283323 79	12833277 9	0.000652107734142 371	Lct	chr2	76639117	76639517	0.000950290385992 814	Prkra
chr1	1284441 33	12844453 3	0.000512853457764 972	Dars1	chr2	77169244	77169644	0.000177323531512 119	Ccdc141
chr1	1329106 63	13291106 3	0.000532161935764 774	Lrn2	chr2	77519782	77520182	0.000816545882768 249	Zfp385b
chr1	1351475 52	13514795 2	0.000411400986646 715	Arl8a	chr2	77774357	77774757	0.000252814861148 92	Zfp385b
chr1	1354131 60	13541356 0	0.000572069084880 944	Ipo9	chr2	79049995	79050395	0.000911283465845 347	Ube2e3
chr1	1356911 00	13569150 0	0.000614943032240 189	Nav1	chr2	79806814	79807214	0.000516758787742 377	Ppp1r1c
chr1	1385032 42	13850364 2	0.000104176625913 555	Nek7	chr2	79855305	79855705	0.000652026331281 697	Pde1a
chr1	1389869 67	13898736 7	0.000900712164940 698	4930596I21Ri k	chr2	80268020	80268420	7,50E+09	Prdx6b
chr1	1390513 12	13905171 2	0.000706440896343 974	Dennd1b	chr2	84383569	84383969	0.000452753022238 461	Calcr1
chr1	1391079 83	13910838 3	0.000521884018421 126	Dennd1b	chr2	84850648	84851048	0.000792551172759 337	Slc43a1
chr1	1391354 57	13913585 7	9,96E+09	Dennd1b	chr2	90667344	90667744	0.000624859268862 91	Nup160
chr1	1428949 64	14289536 4	0.000785713031118 407	Cdc73	chr2	91027076	91027476	0.000516758858847 943	Rapsn
chr1	1435405 29	14354092 9	8,32E+09	Cdc73	chr2	91247316	91247716	0.000337851351509 322	Pacsin3
chr1	1437067 94	14370719 4	0.000998069050238 692	Cdc73	chr2	91389675	91390075	0.000845010450756 689	Cstpp1
chr1	1443387 27	14433912 7	0.000903100704021 886	Rgs1	chr2	91851564	91851964	3,03E+09	Ambra1
chr1	1444883 55	14448875 5	4,82E+08	Rgs21	chr2	92023010	92023410	4,96E+09	Creb3l1
chr1	1516295 36	15162993 6	0.000311286554882 662	Niban1	chr2	92200220	92200620	0.000932195074281 772	Mir1955
chr1	1519003 62	15190076 2	0.000861401520470 65	1700025G04 Rik	chr2	92299601	92300001	0.000566642327976 208	Phf21a
chr1	1529379 66	15293836 6	0.000485356628367 795	Nmnat2	chr2	92555833	92556233	0.000996966098006 48	Mir7221
chr1	1529562 90	15295669 0	8,46E+09	Nmnat2	chr2	93150289	93150689	0.000286571224190 846	Trp53i11
chr1	1552765 72	15527697 2	0.000654052800983 518	BC034090	chr2	93165692	93166092	0.000986036010993 452	Trp53i11
chr1	1553986 68	15539906 8	0.000939032551261 265	Xpr1	chr2	96666634	96667034	0.000841576893883 866	Lrrc4c
chr1	1560286 20	15602902 0	0.000714090702535 586	Tor1aip1	chr2	97335924	97336324	0.000426737340513 794	Lrrc4c
chr1	1562033 70	15620377 0	0.000234999976383 099	Fam163a	chr2	10245009	10245049	0.000993200245334 114	Fjx1
chr1	1562050 89	15620548 9	0.000859827572976 757	Fam163a	chr2	10288098	10288138	6,22E+08	Cd44
chr1	1572090 09	15720940 9	0.000692540698945 754	Rasal2	chr2	10371927	10371967	0.000854922247559 342	Abtb2
chr1	1577564 62	15775686 2	0.000676748969253 405	Sec16b	chr2	10575373	10575413	0.000520006779453 1	Elp4
chr1	1587319 18	15873231 8	1,31E+08	Astn1	chr2	10598646	10598686	0.000117703613602 638	Dnajc24
chr1	1594036 96	15940409 6	0.000116776910620 731	Cop1	chr2	10689288	10689328	0.000417796696782 435	Arl14ep
chr1	1612533 51	16125375 1	9,10E+09	Prdx6	chr2	10973464	10973504	0.000849990563735 137	Bdnf
chr1	1623621 61	16236256 1	0.000797674860335 551	Dnm3	chr2	10977144	10977184	0.000410857222718 83	Bdnf
chr1	1625680 68	16256846 8	0.000704658275565 873	Vamp4	chr2	11065665	11065705	0.000496483827290 95	Slc5a12
chr1	1628831 86	16288358 6	0.000841557477930 799	Fmo2	chr2	11302204	11302244	0.000224005166624 242	Ryr3
chr1	1646205 78	16462097 8	0.000493131499119 004	Gm26685	chr2	11353575	11353615	0.000334115722850 248	Fmn1
chr1	1647181 45	16471854 5	0.000215403006991 41	Dpt	chr2	11560597	11560637	0.000552454155956 283	Cdin1
chr1	1682373 90	16823779 0	0.000524395868973 349	Pbx1	chr2	11715748	11715788	5,66E+09	Mir674
chr1	1682795 26	16827992 6	0.000528834133008 527	Pbx1	chr2	12265700	12265740	3,26E+09	Mir147
chr1	1719823 96	17198279 6	0.000475315289854 006	Vangl2	chr2	12276555	12276595	0.000574376235552 565	Sqor

chr1	1730625 58	17306295 8	0.000325505826668 682	Or10j27	chr2	12318027 6	12318067 6	0.000499724497898 368	Sqor
chr1	1744851 92	17448559 2	0.000777622528708 166	Fmn2	chr2	12414885 1	12414925 1	0.000930780682950 605	Sema6d
chr1	1757029 51	17570335 1	0.000808712938932 254	Wdr64	chr2	12438001 3	12438041 3	7,65E+09	4930583P06R ik
chr1	1762757 10	17627611 0	0.000753960150015 054	Pld5	chr2	12438177 1	12438217 1	0.000601062520200 424	4930583P06R ik
chr1	1764135 23	17641392 3	0.000231801400755 627	Gm36536	chr2	12494793 8	12494833 8	0.000733037949509 185	Slc24a5
chr1	1799646 50	17996505 0	0.000569429386393 372	Cdc42bpa	chr2	12543244 5	12543284 5	0.000503614896521 842	Fbn1
chr1	1809459 68	18094636 8	0.000166558675046 78	Tmem63a	chr2	12601565 7	12601605 7	0.000519900972672 627	Fgf7
chr1	1861855 70	18618597 0	0.000959911553895 854	Lyplal1	chr2	12827650 5	12827690 5	0.000175208186707 243	Bcl2l11
chr1	1862253 97	18622579 7	0.000160684228363 017	Lyplal1	chr2	12868551 1	12868591 1	0.000278467122954 762	Anapc1
chr1	1862614 14	18626181 4	0.000268151467808 009	Lyplal1	chr2	12874679 4	12874719 4	0.000727168676058 598	Mertk
chr1	1871855 49	18718594 9	0.000795601197621 787	Spata17	chr2	12956133 1	12956173 1	0.000145589812054 37	F830045P16R ik
chr1	1876724 15	18767281 5	0.000588598354041 237	Esrrg	chr2	12960411 9	12960451 9	0.000630950277338 727	Sirpa
chr1	1877576 99	18775809 9	0.000643055977042 284	Esrrg	chr2	13104479 7	13104519 7	0.000785120351400 659	Gfra4
chr1	1878998 02	18790020 2	0.000690972760425 845	Esrrg	chr2	13118976 3	13119016 3	0.000685746304403 212	Cdc25b
chr1	1881362 25	18813662 5	0.000588725795687 806	Ush2a	chr2	13126469 4	13126509 4	0.000285544568276 341	Pank2
chr1	1911049 24	19110532 4	0.000444885213842 339	Batf3	chr2	13189600 7	13189640 7	0.000756318391789 801	Prnp
chr1	1943214 18	19432181 8	0.000913613537857 35	4930503O07 Rik	chr2	13240987 4	13241027 4	0.000241973985573 907	4921508D12 Rik
chr1	1943999 45	19440034 5	0.000769358536527 154	4930503O07 Rik	chr2	13268821 5	13268861 5	0.000421222159708 634	Shld1
chr1	1946187 89	19461918 9	0.000664781592368 824	Plxna2	chr2	13348401 3	13348441 3	0.000205507052037 701	Bmp2
chr1	1946663 42	19466674 2	0.000870886749344 546	Plxna2	chr2	13511468 4	13511508 4	0.000763109800599 349	4930545L23R ik
chr1	1947339 26	19473432 6	2,18E+09	Plxna2	chr2	13556473 9	13556513 9	5,82E+09	9630028H03 Rik
chr1	1949779 03	19497830 3	0.000339423702432 562	Gm16897	chr2	13563440 8	13563480 8	0.000344808108505 346	Plcb4
chr1	8307308 0	8307708 0	0.000503467218993 636	Ust	chr2	13658889 1	13658929 1	3,12E+09	Ankef1
chr1	8516766 0	8517166 0	0.000765946759849 534	Ust	chr2	13716476 0	13716516 0	0.000361218163385 635	Jag1
chr1	9235757 0	9236157 0	0.000839764764815 509	Sash1	chr2	13800763 6	13800803 6	0.000497019690168 354	Btbd3
chr1	1355334 8	13553748 0	0.000433538580865 367	Pex3	chr2	14074033 2	14074073 2	0.000923447886668 649	Flrt3
chr1	1652406 6	16524466 0	2,53E+09	Gm20125	chr2	14385326 7	14385366 7	0.000919459513679 888	Bfsp1
chr1	1904373 4	19044134 0	0.000598016133076 307	Tnfaip3	chr2	14408597 6	14408637 6	0.000993258339010 65	Banf2
chr1	2067689 2	20677292 0	0.000971811419428 814	Pde7b	chr2	14441041 1	14441081 1	0.000404680773641 086	Kat14
chr1	2360315 4	23603554 0	5,80E+09	Snora33	chr2	14667794 0	14667834 0	0.000125172789403 605	4933406D12 Rik
chr1	2414892 1	24149321 0	0.000249253415781 414	Stx7	chr2	15076644 9	15076684 9	0.000128761716596 371	Entpd6
chr1	2495914 3	24959543 0	0.000126350573866 452	Arg1	chr2	15211509 8	15211549 8	0.000320718278132 848	Srxn1
chr1	2546129 0	25461690 0	0.000264102958102 974	Epb41l2	chr2	15255432 3	15255472 3	0.000681902258167	Defb29
chr1	2573497 9	25735379 0	0.000529510518362 697	4930401C15 Rik	chr2	15340715 9	15340755 9	0.000494516267071 31	Asxl1
chr1	2609933 0	26099730 0	0.000918419098558 276	Tmem200a	chr2	15341460 8	15341500 8	9,60E+09	Asxl1
chr1	2690226 8	26902668 0	0.000933336972130 05	Arhgap18	chr2	15422884 9	15422924 9	0.000921322480039 51	Bpifb5
chr1	2692011 3	26920513 0	0.000510632028364 1	Arhgap18	chr2	15567502 6	15567542 6	0.000425460103772 313	Trpc4ap
chr1	2700176 7	27002167 0	0.000739704808481 401	Lama2	chr2	15651447 6	15651487 6	0.000673794790195 803	Epb41l1
chr1	2726691 9	27267319 0	6,73E+09	Lama2	chr2	15702744 5	15702784 5	0.000212684926065 076	Soga1
chr1	2823309 5	28233495 0	0.000416550355756 051	Ptpkr	chr2	15718705 2	15718745 2	0.000468351477906 026	Rbl1

chr1	2859304	28593440	0.000546376908764	Ptpk	chr2	15718798	15718838	0.000959150608489	Rbl1
0	0		135		6	6	834		
chr1	3018343	30183837	0.000714171983636	Cenpw	chr2	15739921	15739961	0.000134877768714	Src
0	7		325		8	8	015		
chr1	3083933	30839730	0.000840045120896	Hey2	chr2	15741725	15741765	0.000633827269535	Src
0	0		749		1	1	869		
chr1	3319386	33194268	0.000978524048655	Trdn	chr2	15755672	15755712	0.000399997228626	Blcap
0	8		061		6	6	381		
chr1	3861553	38615934	0.000696259472399	Rfpl4b	chr2	15837970	15838010	0.000432731624604	Snhg11
0	4		596		6	6	443		
chr1	3953834	39538742	0.000637000286923	Fyn	chr2	15873860	15873900	0.000324162053191	Fam83d
0	2		595		2	2	301		
chr1	3973629	39736697	8,10E+09	Rev3l	chr2	15925064	15925104	0.000483739634767	Dhx35
0	7				8	8	666		
chr1	4094381	40944212	0.000956149874089	Wasf1	chr2	16105759	16105799	9,00E+08	Chd6
0	2		54		9	9			
chr1	4224576	42246169	0.000372906194065	Foxo3	chr2	16161936	16161976	4,03E+08	Ptprt
0	9		159		8	8			
chr1	4368946	43689861	0.000176379680734	1700027J07R	chr2	16176412	16176452	0.000932763504147	Ptprt
0	1		806	ik	chr2	16176412	16176452	0.000932763504147	Ptprt
chr1	4481102	44811422	0.000763895557164	Prep	chr2	16586361	16586401	0.000171277989918	Zmynd8
0	2		274		8	8	117		
chr1	4969438	49694788	0.000979929458170	Grik2	chr2	16784784	16784824	0.000552803570557	Ptpn1
0	8		714		6	6	98		
chr1	5282044	52820843	0.000517264209119	Slc35f1	chr2	16785116	16785156	0.000787389965926	Ptpn1
0	3		35		5	5	533		
chr1	5285540	52855808	7,32E+09	Slc35f1	chr2	16798475	16798515	0.000948720992000	Ripor3
0	8				0	0	274		
chr1	5675200	56752404	0.000407780453099	Gm36595	chr2	16801053	16801093	0.000808330858437	Ripor3
0	4		971		0	0	678		
chr1	5861552	58615923	0.000563279409260	Edar	chr2	16860385	16860425	0.000788729351583	Nfatc2
0	3		264		1	1	806		
chr1	5879812	58798529	0.000298331119755	Sh3rf3	chr2	16895379	16895419	0.000179138252509	Zfp64
0	9		968		9	9	148		
chr1	5964149	59641890	0.000677422698237	Mcu	chr2	17040702	17040742	0.000129532946312	Bcas1
0	0		784		5	5	017		
chr1	5989836	59898765	0.000704551147082	Dnajb12	chr2	17200399	17200439	0.000285144084313	Cbln4
0	5		899		6	6	682		
chr1	6302674	63027143	0.000660331690107	Pbld2	chr2	17854573	17854613	0.000893962057740	Cdh26
0	3		485		0	0	838		
chr1	6317411	63174513	0.000499796526729	Mypn	chr2	17972979	17973019	0.000453228674000	Cdh4
0	3		712		5	5	217		
chr1	6324806	63248462	0.000535890393690	Herc4	chr2	18042911	18042951	0.000716939278279	Slco4a1
0	2		73		8	8	094		
chr1	7535726	75357660	0.000127632095256	Gm5779	chr2	18171359	18171399	0.000723893861902	Opr1
0	0		569		2	2	081		
chr1	7537371	75374116	0.000895622019952	Gm5779	chr2	18183909	18183949	0.000849093957576	Pcmt2
0	6		302		1	1	486		
chr1	7551943	75519830	0.000820108666889	Snrpd3	chr3	8162594	8162994	0.000378207388332	1700010I02Ri
0	0		752				497	k	
chr1	7841327	78413671	0.000915023383607	Rrp1	chr3	8944815	8945215	0.000987700683646	Tpd52
0	1		272				286		
chr1	7990895	79909355	0.000505597747667	Med16	chr3	9550481	9550881	0.000529282966015	Zfp704
0	5		341				51		
chr1	8159801	81598411	3,73E+09	Tle6	chr3	14609410	14609810	0.000773508891513	Rbis
0	1						064		
chr1	8291966	82920068	0.000761287689480	Txnrd1	chr3	15193530	15193930	0.000968635315675	Sirpd
0	8		25				703		
chr1	8292339	82923795	0.000186743678673	Txnrd1	chr3	18052524	18052924	0.000576976306972	Bhlhe22
0	5		329				51		
chr1	8418739	84187797	3,73E+09	Gm34184	chr3	18999326	18999726	0.000515261414288	Armc1
0	7						926		
chr1	8613800	86138407	0.000794127832420	4930486F22R	chr3	19300586	19300986	0.000628583950025	Pde7a
0	7		388	ik			297		
chr1	8670664	86707048	4,24E+09	Ttc41	chr3	20880908	20881308	0.000115458101694	Agtr1b
0	8						221		
chr1	8685209	86852497	0.000309596754519	Stab2	chr3	21553766	21554166	0.000841413197855	Tbl1xr1
0	7		026				134		
chr1	8697321	86973615	0.000587746785654	Stab2	chr3	22157386	22157786	0.000693892197781	Tbl1xr1
0	5		764				709		
chr1	8762946	87629862	0.000928455160694	Pah	chr3	27743445	27743845	0.000127332310567	Fndc3b
0	2		508				052		
chr1	8767865	87679054	0.000579633356063	Pah	chr3	28038194	28038594	0.000458924128478	Pld1
0	4		258				477		
chr1	9050657	90506975	0.000913119829552	Anks1b	chr3	28365223	28365623	0.000859298426863	Tnik
0	5		667				164		
chr1	9118315	91183553	0.000270519416759	Tmpo	chr3	28372864	28373264	0.000102422998155	Tnik
0	3		374				66		
chr1	9307087	93071270	0.000506244238386	Cfap54	chr3	30619855	30620255	0.000527720100552	Mynn
0	0		3				02		

chr1 0	9394748 4	93947884	0.000613623795501 894	Mir331	chr3 34103598	34103998	0.000946531403412 72	Sox2ot
chr1 0	9439816 8	94398568	0.000866458662590 386	Tmcc3	chr3 35066637	35067037	0.000846768900651 94	Mir6378
chr1 0	9532133 1	95321731	0.000395128030790 337	Cradd	chr3 35243085	35243485	0.000837147974727 115	Mir6378
chr1 0	9655572 5	96556125	0.000857684365535 683	Btg1	chr3 36862134	36862534	0.000300163029460 121	Bltp1
chr1 0	9688357 3	96883973	0.000718247619103 884	4930556N09 Rik	chr3 37636194	37636594	0.000830222348109 366	Spry1
chr1 0	9943999 9	99440399	0.000403973545814 562	Gm20110	chr3 37650682	37651082	0.000525715878607 613	Spry1
chr1 0	9945013 7	99450537	0.000454500139916 325	Gm20110	chr3 39062142	39062542	0.000105205263263 075	Fat4
chr1 0	9962756 6	99627966	0.000920311908768 623	Gm18409	chr3 39081644	39082044	0.000279692187979 519	Fat4
chr1 0	1001065 37	10010693	8,68E+09	Kitl	chr3 44601245	44601645	0.000584901856856 224	2610316D01 Rik
chr1 0	1003003 76	10030077	0.000203421138163 407	Gm4301	chr3 46399614	46400014	0.000445317474666 806	Pabpc4l
chr1 0	1024355 85	10243598	0.000629319225649 239	Nts	chr3 50433291	50433691	0.000649642999999 367	Slc7a11
chr1 0	1025542 91	10255469	0.000968493675854 72	Rassf9	chr3 52140123	52140523	5,75E+09	Maml3
chr1 0	1033573 74	10335777	0.000510844705974 955	Slc6a15	chr3 52955661	52956061	0.000549457519340 768	Gm20750
chr1 0	1034481 07	10344850	0.000711389409862 262	Slc6a15	chr3 53145571	53145971	0.000470734903986 969	Lhfp16
chr1 0	1056849 68	10568536	0.000928468680579 497	Gm15663	chr3 55337228	55337628	0.000872946271196 273	Dcl1
chr1 0	1104564 32	11045683	0.000567543629882 206	Nav3	chr3 55445155	55445555	0.000342217988210 88	Dcl1
chr1 0	1106581 84	11065858	0.000278746364730 686	9230102K24 Rik	chr3 57949967	57950367	0.000640969082803 977	Pfn2
chr1 0	1110726 00	11107300	0.000185134149659 705	Zdhc17	chr3 57980679	57981079	0.000324601525279 578	Pfn2
chr1 0	1115275 71	11152797	0.000343124503599 634	Phlda1	chr3 60893180	60893580	0.000691972173292 31	P2ry1
chr1 0	1119755 80	11197598	0.000575300655726 085	Krr1	chr3 64964506	64964906	0.000696043456549 278	Kcnab1
chr1 0	1153600 39	11536043	0.000859992333778 425	Tmem19	chr3 65574155	65574555	0.00089883658768 002	Tiparp
chr1 0	1160965 76	11609697	0.000357777252792 513	Taf7l2	chr3 65730151	65730551	0.000986023755857 213	Lekr1
chr1 0	1198376 31	11983803	0.000790343570373 553	Grip1	chr3 69827005	69827405	2,92E+09	Sptssb
chr1 0	1202022 59	12020265	0.000466722067238 523	Irak3	chr3 70102558	70102958	0.000914041089770 112	Otol1
chr1 0	1204070 03	12040740	0.000636638486906 76	9230105E05R ik	chr3 70192015	70192415	0.000646776540282 89	Otol1
chr1 0	1205852 74	12058567	0.000775055686271 12	Hmga2	chr3 70458668	70459068	0.000531019655932 496	Otol1
chr1 0	1208779 05	12087830	0.000581431035460 822	Msrb3	chr3 70649712	70650112	0.000963526373110 479	Otol1
chr1 0	1214121 87	12141258	0.000253881581974 369	Gns	chr3 73706516	73706916	0.000743467331899 783	Bche
chr1 0	1217901 82	12179058	0.000805743933884 218	Srgap1	chr3 75647665	75648065	5,68E+09	Platr10
chr1 0	1227157 62	12271616	0.000909837869326 635	Mir8104	chr3 79222952	79223352	0.000486800731178 239	4921511C10 Rik
chr1 0	1276762 40	12767664	1,51E+09	Nemp1	chr3 79231338	79231738	0.000230396129213 828	4921511C10 Rik
chr1 0	1277487 65	12774916	0.000696199233697 126	Gpr182	chr3 79810145	79810545	0.000259388698349 294	Tmem144
chr1 0	1278286 44	12782904	0.000957812746869 902	Rdh18-ps	chr3 80695843	80696243	0.000248841105003 672	Gria2
chr1 0	1286087 05	12860910	0.000776882906026 843	Rps26	chr3 81049242	81049642	0.000874976707169 628	Pdgfc
chr1 1	3418850	3419250	0.000288933728248 342	Limk2	chr3 82131627	82132027	0.000993798222231 002	Gucy1a1
chr1 1	4629834	4630234	0.000525368372418 678	Gm11961	chr3 83145504	83145904	0.000686572630364 424	Gm10710
chr1 1	5000894	5001294	0.000820513542042 489	Ap1b1	chr3 83586633	83587033	0.000237674994825 854	1700028M03 Rik
chr1 1	5848928	5849328	0.000144108329160 897	Polm	chr3 83765391	83765791	0.000181927998898 601	Sfrp2
chr1 1	6094486	6094886	0.000548081581672 744	Nudcd3	chr3 83895152	83895552	0.000682629848549 832	Tmem131l
chr1 1	6560283	6560683	0.000909986847525 524	Ccm2	chr3 84209353	84209753	0.000624252079866 272	Trim2

chr1 1	6639031	6639431	0.000529929189464 706	Wap	chr3 84845900	84846300	0.000117436446491 537	Fbxw7
chr1 1	7046278	7046678	0.000101512527500 569	Adcy1	chr3 85413840	85414240	0.000153698374812 938	Mir466k
chr1 1	8464046	8464446	0.000233044771466 018	Tns3	chr3 88713823	88714223	0.000688016980485 484	Rit1
chr1 1	1657419	16574593	0.000632219242849 841	Sec61g	chr3 89100163	89100563	0.000572908664735 028	Rusc1
chr1 1	1671101	16711414	0.000741741765603 265	Egfr	chr3 89308796	89309196	0.000609360372644 696	Efna3
chr1 1	1683614	16836549	0.000270056817347 486	Egfr	chr3 90629689	90630089	0.000128584545985 3	S100a6
chr1 1	1898378	18984187	0.000255508201899 346	Meis1	chr3 96008906	96009306	0.000622290952406 068	Plekho1
chr1 1	2001158	20011982	0.000271091559901 571	Spred2	chr3 96057420	96057820	0.000643701078050 452	Vps45
chr1 1	2025514	20255548	0.000774356291739 942	Cep68	chr3 96425494	96425894	0.000656901295018 386	Terc
chr1 1	2059416	20594561	0.000750984580136 56	Sertad2	chr3 97254552	97254952	0.000689720395373 739	Bcl9
chr1 1	2128605	21286452	0.000229063779103 589	Vps54	chr3 97291741	97292141	0.000920103889457 069	Bcl9
chr1 1	2143167	21432075	0.000482477628789 724	Ugp2	chr3 97729932	97730332	0.000327072013905 028	Pde4dip
chr1 1	2231169	22312092	0.000720017211299 773	Ehbp1	chr3 98993444	98993844	0.000702304792442 376	Gm12440
chr1 1	2389206	23892460	0.000404625120985 713	Papolg	chr3 10110388	10110428	0.000841012566516 808	Ptgrn
chr1 1	2401245	24012852	0.000143671824606 617	Bcl11a	chr3 10217672	10217712	0.000872225103745 3	Vangl1
chr1 1	2482933	24829739	0.000665117581355 901	Gm10466	chr3 10241306	10241346	0.000189155433218 957	Ngf
chr1 1	2586234	25862747	0.000797609764699 558	5730522E02R ik	chr3 10373576	10373616	0.000163913263692 139	Olfm13
chr1 1	2936337	29363774	0.000156598750227 444	Ccdc88a	chr3 10452298	10452338	5,62E+09	Lrig2
chr1 1	2972513	29725535	0.000462017361368 305	Rtn4	chr3 10725808	10725848	0.000518468182599 152	Prok1
chr1 1	3001982	30020224	0.000230724657102 721	Eml6	chr3 10746081	10746121	0.000507051318799 467	Kcnc4
chr1 1	3021124	30211641	0.000267300355214 495	Sptbn1	chr3 10770457	10770497	0.000640734271498 022	Ahcyl1
chr1 1	3037193	30372331	0.000480538223730 509	4930505A04 Rik	chr3 10906584	10906624	0.000183547613028 301	Stc25a54
chr1 1	3098490	30985308	0.000531082989325 915	Chac2	chr3 10962256	10962296	7,76E+09	Vav3
chr1 1	3132764	31328043	0.000852946842416 991	Stc2	chr3 10999780	10999820	0.000857759807151 526	Ntng1
chr1 1	3172347	31723871	0.000883353795135 156	Bod1	chr3 11597004	11597044	0.000814208328947 075	Mir669n
chr1 1	3271208	32712480	0.000497692756926 46	Fbxw11	chr3 11705353	11705393	0.000470881767507 425	1700061117Ri k
chr1 1	3306199	33062397	0.000444034899842 682	Fgf18	chr3 11783273	11783313	0.000927482378947 353	Snx7
chr1 1	3663532	36635728	1,95E+09	Tenm2	chr3 11917365	11917405	0.000107007317235 341	Dpyd
chr1 1	3889206	38892468	0.000294412705308 389	Gm12130	chr3 11968488	11968528	0.000475657207597 902	Ptbp2
chr1 1	4378281	43783213	0.000734467387167 426	Ttc1	chr3 12048795	12048835	0.000187931629753 058	6530403H02 Rik
chr1 1	4405757	44057975	0.000453043935171 982	Adra1b	chr3 12141766	12141806	0.000624532260795 84	Cnn3
chr1 1	4608940	46089801	0.000571519138351 295	Mir8100	chr3 12160562	12160602	0.000404237612629 421	A730020M07 Rik
chr1 1	4862225	48622658	0.000812642012114 839	Rack1	chr3 12216927	12216967	0.000203052018721 284	Abca4
chr1 1	5153532	51535726	8,09E+09	Col23a1	chr3 12243320	12243360	0.000906906550709 685	Bcar3
chr1 1	5158367	51584074	0.000229367202516 428	Phykpl	chr3 12677602	12677642	0.000803769434228 219	Camk2d
chr1 1	5196456	51964968	0.000662452244748 923	Cdkn2aipnl	chr3 12700139	12700179	0.000432317178308 494	Ank2
chr1 1	5331184	53312242	0.000523231871437 879	Hspa4	chr3 12929704	12929744	0.000548569476467 975	Enpep
chr1 1	5350521	53505610	0.000708593104101 096	Septin8	chr3 13200643	13200683	0.000241419113755 859	Dkk2
chr1 1	5474889	54749291	0.000293208098042 452	Cdc42se2	chr3 13374279	13374319	0.000158830307489 088	Tet2
chr1 1	5494049	54940899	0.000146853723504 536	Tnlp1	chr3 13566192	13566232	0.000830165942102 612	Nfkb1

chr1	5494449	54944896	0.000468575391277	Tnip1	chr3	13591834	13591874	0.000598894604568	Slc39a8
chr1	6		87		chr3	8	8	967	
chr1	5496292	54963322	0.000636274535756	Tnip1	chr3	13658400	13658440	0.000820313387791	Ppp3ca
chr1	2		431		chr3	5	5	316	
chr1	5637563	56376031	0.000692853878553	Nmur2	chr3	13689241	13689281	0.000390300200290	Ppp3ca
chr1	1		779		chr3	6	6	739	
chr1	5682731	56827713	0.000521955149754	Gria1	chr3	13780771	13780811	0.000227091817628	Mir6380
chr1	3		172		chr3	0	0	811	
chr1	5700493	57005332	0.000534676916608	Gria1	chr3	13878784	13878824	0.000896664645477	Tspan5
chr1	2		356		chr3	8	8	552	
chr1	5833051	58330913	0.000191559281020	Sh3bp5l	chr3	14337853	14337893	0.000105647605086	A830019L24R
chr1	3		215		chr3	9	9	912	ik
chr1	6151644	61516847	7,28E+09	B9d1	chr3	14631523	14631563	0.000440303298516	Gm10636
chr1	7				chr3	9	9	742	
chr1	6159050	61590905	4,73E+08	Epn2	chr3	14680429	14680469	0.000160166255286	Prkacb
chr1	5				chr3	8	8	809	
chr1	6287595	62876352	0.000208774083114	Fbxw10	chr3	14708098	14708138	0.000600018052876	Ttll7
chr1	2		482		chr3	8	8	905	
chr1	6291077	62911171	5,04E+09	Tvp23b	chr3	14783462	14783502	0.000983621579223	Ttll7
chr1	1				chr3	5	5	826	
chr1	6472266	64723068	0.000607938452029	Elac2	chr3	14892664	14892704	0.000669053166166	Adgrl2
chr1	8		666		chr3	2	2	833	
chr1	6593752	65937920	0.000602684159625	Dnah9	chr3	14900391	14900431	0.000124941496346	Adgrl2
chr1	0		388		chr3	2	2	898	
chr1	6617768	66178081	0.000994068238054	Dnah9	chr3	14935368	14935408	0.000158825765720	Adgrl2
chr1	1		036		chr3	8	8	794	
chr1	6715211	67152514	0.000811835763325	Myh2	chr3	15171757	15171797	0.000507557171655	Ifi44
chr1	4		996		chr3	8	8	071	
chr1	6951658	69516989	0.000692731654127	Dnah2	chr3	15267710	15267750	0.000628071732626	Ak5
chr1	9		643		chr3	5	5	578	
chr1	7242616	72426564	0.000766702251493	Ggt6	chr3	15569138	15569178	7,69E+09	9330178D15
chr1	4		667		chr3	0	0		Rik
chr1	7272262	72723025	0.000184769816583	Ankfy1	chr4	4775116	4775516	0.000430735297296	Bpnt2
chr1	5		33		chr4			98	
chr1	7317540	73175809	0.000698035967874	Emc6	chr4	5646853	5647253	0.000424737537378	Fam110b
chr1	9		988		chr4			705	
chr1	7533495	75335358	0.000448424263540	Rpa1	chr4	8809764	8810164	0.000465070755811	Chd7
chr1	8		97		chr4			148	
chr1	7562773	75628137	0.000800653767184	Inpp5k	chr4	9337231	9337631	0.000411155986884	Clvs1
chr1	7		301		chr4			459	
chr1	7617682	76177220	0.000610526099630	Vps53	chr4	9666707	9667107	0.000266770154944	Asph
chr1	0		748		chr4			362	
chr1	7628486	76285261	0.000644555619529	Mrm3	chr4	14041196	14041596	0.000142151431808	Runx1t1
chr1	1		585		chr4			735	
chr1	7640277	76403176	0.000614141932087	Nxn	chr4	14967683	14968083	0.000860205403126	Gm11837
chr1	6		753		chr4			368	
chr1	7721496	77215361	0.000738849049289	Ssh2	chr4	16038254	16038654	0.000176772428297	Osgin2
chr1	1		225		chr4			278	
chr1	7868962	78690029	0.000864346794733	Nlk	chr4	18220804	18221204	4,31E+08	Mmp16
chr1	9		795		chr4				
chr1	7903638	79036785	0.000410292559359	Gm11201	chr4	20629053	20629453	0.000809220591995	Nkain3
chr1	5		204		chr4			397	
chr1	7922515	79225554	0.000343284835246	Wsb1	chr4	22881580	22881980	0.000667620724783	Gm30731
chr1	4		26		chr4			324	
chr1	8189759	81897993	0.000275425398140	5530401A14	chr4	31402335	31402735	0.000612768081380	Map3k7
chr1	3		072	Rik	chr4			555	
chr1	8269250	82692903	0.000517288945809	Gm11426	chr4	32307147	32307547	0.000829023939371	Bach2
chr1	3		158		chr4			343	
chr1	8695863	86959039	0.000407623167580	Ypel2	chr4	35226212	35226612	0.000901877486038	C9orf72
chr1	9		029		chr4			316	
chr1	8775964	87760041	0.000217735883247	Tspoap1	chr4	35372368	35372768	0.000360277035582	C9orf72
chr1	1		361		chr4			39	
chr1	8841519	88415596	0.000583483001364	Mir378b	chr4	41570969	41571369	0.000866922502845	Dnai1
chr1	6		238		chr4			702	
chr1	9126270	91263105	0.000965933313236	Kif2b	chr4	43309445	43309845	0.000603489878215	Atp8b5
chr1	5		578		chr4			457	
chr1	9418950	94189904	0.000959031439503	Tob1	chr4	44260501	44260901	0.000814469174352	Rnf38
chr1	4		377		chr4			35	
chr1	9457060	94571006	0.000484111503371	Acsf2	chr4	44504289	44504689	0.000636190855260	Mir5120
chr1	6		354		chr4			343	
chr1	9488169	94882092	0.000118723787342	Col1a1	chr4	47609046	47609446	0.000610123723076	Sec61b
chr1	2		567		chr4			894	
chr1	9547600	95476409	0.000741713781929	Spop	chr4	49347479	49347879	0.000342386675640	Acnat2
chr1	9		665		chr4			238	
chr1	9702957	97029970	0.000103550251070	Scrn2	chr4	49575429	49575829	0.000508293353780	Pgap4
chr1	0		672		chr4			997	
chr1	1008973	10089774	0.000695529335819	Stat3	chr4	53462069	53462469	0.000658234743810	Slc44a1
chr1	42		955		chr4			437	
chr1	1013391	10133959	0.000325850152458	Aoc3	chr4	57190134	57190534	0.000527828985558	Gm12530
chr1	97		34		chr4			18	

chr1	1020698	10207021	0.000943744616712	Mpp2	chr4	58668421	58668821	0.000911915590108	Lpar1
1	18	8	205					947	
chr1	1020723	10207279	0.000560937757750	Mpp2	chr4	59393542	59393942	0.000150565629771	SUSD1
1	90	0	679					338	
chr1	1028790	10287943	0.000425674989002	Eftud2	chr4	62584915	62585315	9.70E+09	Rgs3
1	39	9	811						
chr1	1029711	10297156	0.000176825547945	2410004I01Ri	chr4	64227798	64228198	4.95E+09	Gm11217
1	65	5	736						
chr1	1036488	10364928	0.000739436422438	Rprml	chr4	64305544	64305944	0.000269523858831	Gm11217
1	82	2	011					278	
chr1	1056285	10562898	0.000121654574519	Tanc2	chr4	64635635	64636035	0.000396083625805	Gm11217
1	86	6	07					455	
chr1	1057744	10577487	0.000431219427421	Tanc2	chr4	65125444	65125844	0.000332268887323	Pappa
1	76	6	858					073	
chr1	1059449	10594535	0.000301315878328	Cyb561	chr4	65681376	65681776	0.000682840242449	Trim32
1	54	4	268					911	
chr1	1087350	10873541	0.000478874250370	Cep112	chr4	66658715	66659115	0.000778756416702	Tlr4
1	14	4	964					109	
chr1	1088458	10884626	0.000140671783546	Axi2	chr4	66917455	66917855	0.000165783279524	Tlr4
1	62	2	154					989	
chr1	1091161	10911656	0.000245328705649	E030025P04	chr4	72029560	72029960	0.000524587517366	Tle1
1	63	3	195					06	
chr1	1098688	10986925	0.000414109182544	1700023C21	chr4	76153734	76154134	0.000838276623671	Ptprd
1	50	0	253					419	
chr1	1140164	11401684	0.000573258711898	Sdk2	chr4	81003319	81003719	0.000572389909729	Lurap1l
1	48	8	283					409	
chr1	1140552	11405567	0.000287049714751	Sdk2	chr4	81498500	81498900	2.41E+09	Mpdz
1	76	6	575						
chr1	1148215	11482192	0.000686232709640	Gpr142	chr4	83124944	83125344	0.000802378807303	Frem1
1	24	4	189					716	
chr1	1151605	11516093	0.000121282068971	Nherf1	chr4	83906297	83906697	0.000979716113866	6030471H07
1	39	9	995					491	
chr1	1157380	11573846	0.000236929745051	Tmem94	chr4	84407566	84407966	0.000265762044116	Bnc2
1	65	5	295					435	
chr1	1176110	11761142	0.000946233998835	2900041M22	chr4	84637714	84638114	0.000560333023614	Bnc2
1	24	4	311					357	
chr1	1190477	11904819	0.000426854663543	Cbx8	chr4	85222329	85222729	0.000716926817984	Sh3gl2
1	95	5	53					964	
chr1	1196050	11960547	0.000370488004237	Rptor	chr4	85941802	85942202	0.000631570677979	Adamts1
1	70	0	89					809	
chr1	1197106	11971104	0.000661796466988	Rptor	chr4	88691500	88691900	0.000821417303444	Ifnab
1	40	0	163					427	
chr1	1202677	12026815	0.000344742006076	Bahcc1	chr4	89305618	89306018	0.000128935432368	Cdkn2b
1	58	8	26					804	
chr1	1203508	12035122	0.000208950883349	0610009L18R	chr4	94593848	94594248	0.000608787765783	Plaa
1	20	0	601					436	
chr1	1206393	12063972	6.77E+09	Pycr1	chr4	95082437	95082837	0.000806035660598	Jun
1	21	1						337	
chr1	3931006	3931406	2.10E+09	Pomc	chr4	95144429	95144829	0.000323318884087	Jun
2								209	
chr1	8775589	8775989	0.000357889359262	Sdc1	chr4	97693800	97694200	0.000952733235702	E130114P18R
2			842					663	
chr1	9966058	9966458	0.000255113887756	Osr1	chr4	97791640	97792040	6.54E+09	E130114P18R
2			897						
chr1	1039210	10392506	0.000826695558202	Rdh14	chr4	98886612	98887012	0.000834638468847	Usp1
2	6		584					126	
chr1	1410374	14104147	0.000316127281104	Lratd1	chr4	99134710	99135110	2.29E+09	Dock7
2	7		528						
chr1	1583478	15835180	0.000241826940958	Trib2	chr4	99185392	99185792	4.08E+09	Atg4c
2	0		312						
chr1	1589360	15894008	0.000832730613671	Trib2	chr4	99511144	99511544	0.000734624411062	Foxd3
2	8		062					423	
chr1	1676433	16764734	0.000603657528121	Greb1	chr4	99904933	99905333	0.000180114563688	Pgm1
2	4		637					827	
chr1	1701633	17016735	0.000329627159519	2410004P03	chr4	10010639	10010679	7.76E+09	Ror1
2	5		452			9	9		
chr1	1705356	17053969	0.000566160488295	2410004P03	chr4	10016129	10016169	0.000979680717054	Ror1
2	9		164			5	5	565	
chr1	2869765	28698051	0.000427827052575	Trappc12	chr4	10225656	10225696	0.000223257786146	Pde4b
2	1		491			4	4	513	
chr1	3124365	31244050	5.25E+09	Lamb1	chr4	10253007	10253047	0.000483225927309	Pde4b
2	0					4	4	983	
chr1	3840671	38407110	0.000648147929546	4930428E07R	chr4	10255971	10256011	0.000993921352474	Pde4b
2	0		816			2	2	78	
chr1	3846984	38470243	0.000785053509090	4930428E07R	chr4	10256128	10256168	0.000268922599474	Pde4b
2	3		083			8	8	086	
chr1	3859704	38597442	0.000198858378198	4930428E07R	chr4	10262501	10262541	0.000780127252500	Pde4b
2	2		885			2	2	54	
chr1	3885316	38853560	0.000401480545131	Etv1	chr4	10332375	10332415	0.000412765855186	Oma1
2	0		561			4	4	015	

chr1	3931003	39310434	0.000188929816681	Etv1	chr4	10382240	10382280	0.000886605912199	Dab1
chr1	2	4	505		chr4	1	1	245	
chr1	3992040	39920804	0.000171059340157	Arl4a	chr4	10548338	10548378	0.000293190039754	Plpp3
chr1	2	4	376		chr4	5	5	92	
chr1	4076972	40770127	0.000249239778502	Dock4	chr4	10656030	10656070	0.000188675655042	Dhcr24
chr1	2	7	33		chr4	0	0	214	
chr1	4131247	41312876	0.000249973557236	Lrrn3	chr4	10699272	10699312	3,36E+07	Ssbp3
chr1	2	6	771		chr4	7	7		
chr1	4441025	44410650	0.000321103605236	Nrcam	chr4	10712801	10712841	0.000460966784094	Cdcp2
chr1	2	0	184		chr4	7	7	685	
chr1	4597007	45970478	0.000687564765172	Nova1	chr4	10754538	10754578	0.000929663940381	Glis1
chr1	2	8	954		chr4	2	2	964	
chr1	5182617	51826574	0.000265731603363	Hectd1	chr4	10775712	10775752	0.000657888487374	Lrp8
chr1	2	4	628		chr4	7	7	692	
chr1	5293675	52937151	0.000224541975680	Akap6	chr4	10914477	10914517	0.000757193856121	Osbpl9
chr1	2	1	115		chr4	6	6	94	
chr1	5357287	53573273	0.000400713394009	Npas3	chr4	11034008	11034048	0.000742878406313	Elavl4
chr1	2	3	6		chr4	7	7	041	
chr1	5363083	53631230	5,48E+09	Npas3	chr4	11160595	11160635	0.000109729874240	Agbl4
chr1	2	0			chr4	0	0	96	
chr1	5381095	53811359	6,36E+08	Egln3	chr4	11166743	11166783	0.000813752397510	Spata6
chr1	2	9			chr4	4	4	248	
chr1	5417057	54170970	0.000737384043251	Egln3	chr4	11812377	11812417	0.000824823419880	St3gal3
chr1	2	0	188		chr4	1	1	882	
chr1	5435015	54350550	0.000554471008442	Gm7550	chr4	11901942	11901982	0.000643371209193	Gm12866
chr1	2	0	192		chr4	6	6	999	
chr1	5483976	54840161	0.000880629178776	Cfl2	chr4	12485166	12485206	0.000874475166727	Yrdc
chr1	2	1	404		chr4	4	4	34	
chr1	5548083	55481232	0.000443184977504	Nfkbia	chr4	12531907	12531947	0.000173595125040	Grik3
chr1	2	2	174		chr4	3	3	338	
chr1	5629901	56299417	0.000762637375387	Gm19990	chr4	12718831	12718871	0.000911947048409	Dlgap3
chr1	2	7	287		chr4	0	0	134	
chr1	5853016	58530567	0.000154377283928	Clec14a	chr4	12724953	12724993	0.000479963902306	Smim12
chr1	2	7	309		chr4	8	8	535	
chr1	6700044	67000845	0.000212130751939	Mdga2	chr4	12972492	12972532	0.000127742910285	Khdrbs1
chr1	2	5	155		chr4	9	9	5	
chr1	6711852	67118922	0.000326733078388	Mdga2	chr4	13003651	13003691	5,81E+09	Col16a1
chr1	2	2	486		chr4	3	3		
chr1	7046401	70464417	0.000137304330229	Tmx1	chr4	13199756	13199796	0.000493851279244	Epb41
chr1	2	7	959		chr4	9	9	506	
chr1	7046564	70466040	0.000719068570018	Tmx1	chr4	13429183	13429223	0.000342499854323	Pdik1l
chr1	2	0	022		chr4	4	4	108	
chr1	7086728	70867687	0.000345659834070	Frmd6	chr4	13485455	13485495	8,73E+09	Maco1
chr1	2	7	977		chr4	1	1		
chr1	7124886	71249261	0.000580261617078	Dact1	chr4	13712087	13712127	0.000440977994574	Gm13003
chr1	2	1	914		chr4	5	5	654	
chr1	7139963	71400037	0.000953991009694	1700083H02	chr4	13900466	13900506	0.000444497731369	Tmco4
chr1	2	7	288	Rik	chr4	3	3	564	
chr1	7202069	72021096	0.000586786263226	Gpr135	chr4	13904916	13904956	0.000540613257538	Htr6
chr1	2	6	502		chr4	2	2	514	
chr1	7268773	72688130	0.000883332018251	Dhrs7	chr4	13923614	13923654	0.000280829216362	Capzb
chr1	2	0	304		chr4	6	6	077	
chr1	7325437	73254775	0.000549198691618	Trmt5	chr4	14107592	14107632	0.000174580883792	Necap2
chr1	2	5	624		chr4	9	9	075	
chr1	7340886	73409266	0.000255515833030	D830013O20	chr4	14136622	14136662	0.000171613973969	Fam131c
chr1	2	6	628	Rik	chr4	6	6	142	
chr1	7390957	73909975	0.000204100005593	Hif1a	chr4	14862384	14862424	0.000103174481695	Tardbp
chr1	2	5	455		chr4	8	8	5	
chr1	7555599	75556393	0.000687464053300	Ppp2r5e	chr4	14928482	14928522	0.000396753612589	Kif1b
chr1	2	3	557		chr4	7	7	729	
chr1	7599507	75995472	0.000432111370460	Syne2	chr4	15008768	15008808	0.000591602879541	Gpr157
chr1	2	2	175		chr4	3	3	631	
chr1	7649361	76494014	3,87E+09	Plekkg3	chr4	15020685	15020725	7,47E+09	Car6
chr1	2	4			chr4	5	5		
chr1	7807570	78076100	0.000338143271219	Gphn	chr4	15037989	15038029	0.000699108478245	Rere
chr1	2	0	948		chr4	3	3	725	
chr1	7811906	78119467	0.000698186926971	Gphn	chr4	15208790	15208830	0.000642231705248	Plekkg5
chr1	2	7	767		chr4	3	3	131	
chr1	7990763	79908038	0.000618371932899	9430078K24	chr4	15220463	15220503	0.000827603114920	Acot7
chr1	2	8	375	Rik	chr4	9	9	139	
chr1	8570051	85700918	0.000786158139498	Batf	chr4	15228454	15228494	0.000839868495512	Gpr153
chr1	2	8	783		chr4	7	7	43	
chr1	8577793	85778339	0.000505220536107	Flvcr2	chr4	15347854	15347894	0.000483836045078	Ajap1
chr1	2	9	576		chr4	1	1	068	
chr1	8626284	86263244	0.000599659550405	Gpatch2l	chr4	15436594	15436634	0.000513205729057	Prdm16
chr1	2	4	418		chr4	2	2	997	
chr1	8714207	87142474	0.000859578626237	Gstz1	chr4	15583036	15583076	0.000591618546665	Aurkaip1
chr1	2	4	756		chr4	9	9	007	
chr1	8743542	87435820	0.000964027868976	Mir3068	chr5	3340363	3340763	0.000337263202946	Cdk6
chr1	2	0	425					358	

chr1	9087365	90874056	0.000622835980009	Dio2	chr5	3889722	3890122	0.000635715878190	Mterf1a
2	6		562					002	
chr1	9191553	91915938	0.000556855407190	Sel1l	chr5	9149711	9150111	0.000889897718207	Dmtf1
2	8		637					022	
chr1	9447104	94471449	0.000432126268726	Mir8099-1	chr5	9177734	9178134	0.000938332415210	Dmtf1
2	9		13					796	
chr1	9454627	94546671	0.000535907265075	Mir8099-1	chr5	12938337	12938737	0.000375673901833	Sema3a
2	1		464					018	
chr1	9747422	97474626	0.000110836115017	Galc	chr5	13510019	13510419	0.000395498492491	Sema3a
2	6		943					209	
chr1	9831901	98319410	0.000425126866925	Gpr65	chr5	16874252	16874652	0.000184502022280	Hgf
2	0		529					375	
chr1	9955835	99558750	1,12E+09	Foxn3	chr5	17883413	17883813	0.000771199161442	Cd36
2	0							748	
chr1	9971168	99712086	0.000639337061963	1700064M15	chr5	20814543	20814943	0.000729112991734	Phtf2
2	6		507	Rik				171	
chr1	1004438	10044429	0.000109864356137	Ttc7b	chr5	23445009	23445409	0.000552040567124	Kmt2e
2	92	2	912					212	
chr1	1021406	10214105	0.000661369940315	Slc24a4	chr5	24103313	24103713	0.000767400562601	Klhl7
2	54	4	737					939	
chr1	1023055	10230596	2,13E+09	Rin3	chr5	24473566	24473966	0.000987436847977	Agap3
2	67	7						641	
chr1	1046136	10461408	0.000707118263401	Dicer1	chr5	25051939	25052339	0.000436924284235	Prkag2
2	81	1	307					252	
chr1	1049107	10491112	0.000959880977094	Clmn	chr5	25413500	25413900	0.000412904052283	Kmt2c
2	21	1	54					336	
chr1	1057861	10578658	0.000738184382650	Papola	chr5	25496507	25496907	0.000587123105640	Kmt2c
2	88	8	83					242	
chr1	1081449	10814536	0.000363391151493	Setd3	chr5	25612534	25612934	0.000167131467427	Xrcc2
2	67	7	309					417	
chr1	1115920	11159240	0.000479926941187	Mark3	chr5	25665797	25666197	0.000290324345366	Xrcc2
2	08	8	259					91	
chr1	1118641	11186454	0.000720773092473	Ppp1r13b	chr5	27312489	27312889	0.000825851767567	Dpp6
2	49	9	648					221	
chr1	1172420	11724241	8,32E+09	Mir153	chr5	30237534	30237934	0.000469442828817	Selenoi
2	12	2						89	
chr1	1174808	11748125	0.000194678558829	Rapgef5	chr5	31250657	31251057	0.000838740584071	Nrbp1
2	50	0	597					962	
chr1	1188597	11886019	0.000991646501856	Sp8	chr5	31804973	31805373	0.000266071219723	Babam2
2	99	9	902					774	
chr1	1191603	11916078	0.000845181959248	Itgb8	chr5	32390272	32390672	0.000837618669548	Plb1
2	85	5	32					013	
chr1	5433675	5434075	0.000654080423419	1700016G22	chr5	33625176	33625576	9,65E+09	Fam53a
3			856	Rik					
chr1	5546006	5546406	0.000523328718962	1700016G22	chr5	35019890	35020290	0.000565429636027	Rgs12
3			768	Rik				126	
chr1	5859597	5859997	0.000822784321714	Klf6	chr5	35059923	35060323	0.000241916092019	Dok7
3			121					956	
chr1	9064543	9064943	0.000597597200105	Larp4b	chr5	35927707	35928107	1,31E+09	Afap1
3			104						
chr1	9382816	9383216	0.000764130262363	Dip2c	chr5	36176512	36176912	0.000732809318869	Psapl1
3			771					654	
chr1	9423317	9423717	0.000626776429637	Dip2c	chr5	36472089	36472489	0.000971694649890	Grpel1
3			71					775	
chr1	9742288	9742688	0.000124567562949	Zmynd11	chr5	36986090	36986490	0.000825823344017	Wfs1
3			096					929	
chr1	9897583	9897983	0.000152836451587	Zmynd11	chr5	42777295	42777695	0.000286309594775	Cpeb2
3			923					602	
chr1	1036117	10361575	0.000544939670095	Chrm3	chr5	43156636	43157036	4,76E+09	Cpeb2
3	5		39						
chr1	1193358	11933986	0.000415908166580	Ryr2	chr5	43281682	43282082	0.000841790120097	Cpeb2
3	6		214					441	
chr1	1354919	13549599	0.000315370002609	Nid1	chr5	43481911	43482311	0.000546727767212	C1qtnf7
3	9		505					873	
chr1	1379289	13793292	0.000493752387390	Gng4	chr5	44662136	44662536	0.000505545470607	Ldb2
3	2		689					771	
chr1	1551297	15513379	0.000279362787819	Gli3	chr5	48219965	48220365	5,50E+09	Slit2
3	9		704						
chr1	1566431	15664712	0.000340011179008	4933412O06	chr5	50180811	50181211	0.000307725995319	4930448I18Ri
3	2		171	Rik				03	k
chr1	1675350	16753900	0.000293577511004	B230303A05	chr5	50196984	50197384	0.000598981744533	4930448I18Ri
3	0		672	Rik				322	k
chr1	1813027	18130670	0.000907819869665	4930448F12R	chr5	50309746	50310146	0.000430335681935	Mir8117
3	0		705	ik				512	
chr1	1901273	19013134	0.000728937928438	Amph	chr5	50762228	50762628	0.000466033084798	Mir8117
3	4		311					978	
chr1	1939668	19397080	0.000207906690130	Stard3nl	chr5	52217953	52218353	0.000840753747018	C10083M11
3	0		155					974	Rik
chr1	1961440	19614807	0.000415327053284	Epdr1	chr5	53406607	53407007	0.000573833764848	Rbpj
3	7		731					43	

chr1	1978092	19781320	0.000241948752677	Gpr141	chr5	57767502	57767902	0.000664354275301	Pcdh7
chr1	3	0	912		chr5			315	
chr1	2004189	20042298	0.000970031371970	Elmo1	chr5	58177923	58178323	0.000282631660085	4930459L07R
chr1	3	8	714					686	ik
chr1	2116151	21161915	0.000613328265397	Trim27	chr5	62745538	62745938	9,95E+08	Arap2
chr1	3	5	95						
chr1	2118208	21182488	0.000329872298959	Trim27	chr5	62757038	62757438	0.000973114587374	Arap2
chr1	3	8	755					59	
chr1	2122573	21226131	0.000385906451445	Trim27	chr5	64181445	64181845	0.000906469133345	Tbc1d1
chr1	3	1	539					8	
chr1	2126838	21268786	0.000313945307051	Gpx5	chr5	64253664	64254064	0.000578643364230	Tbc1d1
chr1	3	6	089					969	
chr1	2132703	21327437	0.000433410774580	Gpx6	chr5	64508580	64508980	0.000879366934492	C030018K13
chr1	3	7	478					198	Rik
chr1	2192427	21924679	0.000444902510467	Zfp184	chr5	66728829	66729229	0.000503063488493	Limch1
chr1	3	9	582					938	
chr1	2194186	21942269	0.000374646085326	Zfp184	chr5	66769365	66769765	0.000138903233106	Limch1
chr1	3	9	225					058	
chr1	2336499	23365392	0.000675646951255	Zfp322a	chr5	68112685	68113085	0.000656507218639	Grxcr1
chr1	3	2	215					302	
chr1	2352355	23523953	6,95E+09	H4c8	chr5	71950053	71950453	0.000122937378354	Commmd8
chr1	3	3						004	
chr1	2357303	23573438	0.000461366038529	H2bc7	chr5	73221792	73222192	0.000807062587658	Fryl
chr1	3	8	309					836	
chr1	2421016	24210566	0.000202606988189	Carmil1	chr5	74995135	74995535	0.000539528502840	Gm6116
chr1	3	6	302					669	
chr1	2492919	24929597	0.000720017211299	9330162012R	chr5	75046256	75046656	0.000532505343795	Chic2
chr1	3	7	773	ik				897	
chr1	2922059	29220998	0.000420644978106	A330102110Ri	chr5	81661484	81661884	0.000122922955891	Adgrl3
chr1	3	8	122	k				003	
chr1	2935850	29358901	0.000180387737131	Cdkal1	chr5	81909861	81910261	0.000615455075259	Adgrl3
chr1	3	1	138					508	
chr1	2974694	29747345	0.000409503663225	Cdkal1	chr5	83866305	83866705	0.000787239171532	Tecrl
chr1	3	5	008					894	
chr1	3052492	30525321	0.000633095238628	Uqcrfs1	chr5	84999130	84999530	0.000677767595647	Epha5
chr1	3	1	201					083	
chr1	3419029	34190691	0.000424479358851	Slc22a23	chr5	85535429	85535829	0.000418666251752	Cenpc1
chr1	3	1	704					02	
chr1	3440291	34403314	0.000323373038543	Slc22a23	chr5	89754952	89755352	0.000767620516646	Adamts3
chr1	3	4	24					472	
chr1	3480578	34806182	0.000769522551179	Fam50b	chr5	91444704	91445104	0.000975069868207	Btc
chr1	3	2	68					39	
chr1	3526001	35260411	0.000850787373043	1700019C18	chr5	96234903	96235303	0.000970630844591	Mrpl1
chr1	3	1	671	Rik				813	
chr1	3653815	36538551	0.000203935230780	Fars2	chr5	96345451	96345851	0.000798720969067	Fras1
chr1	3	1	994					412	
chr1	3710045	37100858	0.000509921170888	F13a1	chr5	96912630	96913030	0.000375136981566	Gm8013
chr1	3	8	106					548	
chr1	3780225	37802650	0.000336064764342	Rreb1	chr5	97831468	97831868	0.000114066752915	Antrx2
chr1	3	0	447					365	
chr1	3784223	37842633	0.000967346506856	Rreb1	chr5	99734157	99734557	0.000561761327270	A930011G23
chr1	3	3	259					816	Rik
chr1	3963447	39634875	0.000334921902030	Ofcc1	chr5	10196975	10197015	0.000705961589569	Wdfy3
chr1	3	5	993					037	
chr1	3982219	39822593	0.000989390986573	Ofcc1	chr5	10287501	10287541	0.000674431697068	Arhgap24
chr1	3	3	913					822	
chr1	4350337	43503777	0.000142088839421	Ranbp9	chr5	10319451	10319491	9,28E+09	Mapk10
chr1	3	7	683					6	
chr1	4551524	45515643	0.000945340170838	Gmpr	chr5	10688801	10688841	0.000227085801001	Hfm1
chr1	3	3	36					836	
chr1	4646700	46467406	0.000718314796999	Cap2	chr5	11002378	11002418	7,31E+09	Zfp932
chr1	3	6	131					0	
chr1	4669610	46696502	0.000311233159357	Nup153	chr5	11119582	11119655	0.000940960185671	Ttc28
chr1	3	2	993					9	
chr1	4781863	47819035	9,25E+09	G630093K05	chr5	11232084	11232124	0.000930826978581	Ttip11
chr1	3	5		Rik				4	
chr1	4865587	48656278	0.000409071210808	Barx1	chr5	11266151	11266191	0.000586630433790	1700034G24
chr1	3	8	687					0	Rik
chr1	4949078	49491183	0.000988623964153	Ecm2	chr5	11276122	11276162	0.000360861932023	Myo18b
chr1	3	3	224					893	
chr1	5161738	51617780	0.000905569873676	Cks2	chr5	11291676	11291716	0.000182331114557	Grk3
chr1	3	0	626					0	
chr1	5390982	53910229	9,77E+09	E130119H09	chr5	11515959	11515999	0.000794138277940	Mlec
chr1	3	9		Rik				8	
chr1	5459076	54591168	0.000963543255791	Higd2a	chr5	11533942	11533982	2,39E+09	Gatc
chr1	3	8	066					8	
chr1	5469233	54692737	0.000300853343339	Rnf44	chr5	11703164	11703204	0.000506845385628	Suds3
chr1	3	7	513					6	
chr1	5652266	56523069	0.000643872268669	Fbxl21	chr5	12145139	12145179	0.000481486031737	Tmem116
chr1	3	9	909					5	

chr1	5676261	56763011	0.000662969565783	Smad5
3	1		156	
chr1	5702541	57025818	0.000656792472995	Trpc7
3	8		882	
chr1	5846407	58464470	0.000694681778972	Rmi1
3	0		583	
chr1	5849160	58492000	0.000673057730116	Slc28a3
3	0		611	
chr1	6307223	63072911	0.000413389228628	Aopep
3	5		214	
chr1	6340329	63403696	0.000975729289520	Fancc
3	6		944	
chr1	6341898	63419388	0.000490395052441	Fancc
3	8		306	
chr1	6352635	63526751	0.000723623456656	Ptch1
3	1		997	
chr1	6360392	63604323	0.000993802020579	Ptch1
3	3		748	
chr1	6368978	63690187	0.000373372174163	1700024I08Ri
3	7		444	k
chr1	6974004	69740447	0.000796303659428	Ube2ql1
3	7		288	
chr1	6995892	69959322	0.000291970460585	4933416O17
3	2		933	Rik
chr1	7137030	71370702	0.000872289711118	Mir466f-4
3	2		635	
chr1	7700810	77008509	0.000641338254792	Slf1
3	9		29	
chr1	7713601	77136413	0.000295809898086	2210408I21Ri
3	3		308	k
chr1	7739790	77398307	0.000234262936386	2210408I21Ri
3	7		287	k
chr1	7844927	78449671	0.000885366359913	A830082K12
3	1		986	Rik
chr1	7868584	78686244	0.000640272157279	A830082K12
3	4		798	Rik
chr1	7986736	79867769	0.000645045465505	Arrdc3
3	9		145	
chr1	8094023	80940631	0.000207647070181	5430425K12
3	1		949	Rik
chr1	8321857	83218970	0.000105271826714	Mef2c
3	0		937	
chr1	8352632	83526723	7,76E+09	Mef2c
3	3			
chr1	8424813	84248532	0.000802846123038	Tmem161b
3	2		27	
chr1	8600405	86004452	0.000611615498143	Cox7c
3	2		481	
chr1	8988100	89881402	1,16E+09	Vcan
3	2			
chr1	9030470	90305106	0.000758521155569	Tmem167
3	6		625	
chr1	9061367	90614076	0.000560021344564	Atp6ap1l
3	6		454	
chr1	9099283	90993237	0.000188071611115	Rps23
3	7		115	
chr1	9290290	92903309	0.000542726135984	Mtx3
3	9		051	
chr1	9299304	92993446	0.000644675055631	Gm4814
3	6		927	
chr1	9360862	93609022	0.000799790311982	Bhmt
3	2		518	
chr1	9377241	93772812	0.000434894354091	Arsb
3	2		053	
chr1	9394113	93941536	0.000669037500408	Lhfp12
3	6		809	
chr1	9479128	94791685	0.000277946771193	Tbca
3	5		749	
chr1	9588038	95880782	0.000801329155796	lqgap2
3	2		004	
chr1	9682374	96824144	0.000889978960129	Ankrd31
3	4		101	
chr1	9705086	97051261	0.000250971224314	Fam169a
3	1		833	
chr1	9720450	97204902	0.000401574335965	Hexb
3	2		549	
chr1	9765590	97656309	2,80E+09	5330416C01
3	9			Rik
chr1	9945973	99460135	0.000914074733794	Map1b
3	5		568	

chr5	12156410	12156450	9,37E+09	Aldh2
8	8			
chr5	12237982	12238022	0.000404368296345	Gpn3
9	9		27	
chr5	12290975	12291015	0.000395779535418	Kdm2b
9	9		805	
chr5	12445231	12445271	0.000666740917585	Kmt5a
2	2		784	
chr5	13207598	13207638	0.000973421229364	Auts2
4	4		169	
chr5	13617060	13617100	0.000217604959177	Orai2
7	7		331	
chr5	13691880	13691920	0.000234464067521	4933404O12
6	6		145	Rik
chr5	13742576	13742616	0.000615261047070	Zan
0	0		659	
chr5	13833881	13833921	0.000113205875734	Gm10874
7	7		712	
chr5	13939905	13939945	0.000987519692891	C130050O18
7	7		903	Rik
chr5	14037405	14037445	0.000658505009693	Snx8
3	3		444	
chr5	14063618	14063658	0.000222899040429	Ttyh3
8	8		159	
chr5	14082075	14082115	0.000153803507332	Gna12
5	5		961	
chr5	14227507	14227547	0.000622584840171	Foxk1
8	8		303	
chr5	14254489	14254529	0.000578208586693	Radil
6	6		885	
chr5	14269576	14269616	0.000790397162879	Slc29a4
1	1		541	
chr5	14689134	14689174	0.000724668301873	Rasl11a
6	6		788	
chr5	14916109	14916149	0.000524620922483	Gm8615
0	0		843	
chr5	14918762	14918802	0.000827101157804	Uspl1
7	7		044	
chr5	15054723	15054763	0.000107287158926	Brca2
0	0		717	
chr5	15091976	15092016	0.000969163028966	Kl
8	8		769	
chr6	3539523	3539923	0.000532302953033	Vps50
			303	
chr6	3860428	3860828	0.000907385561937	Calcr
			701	
chr6	6456157	6456557	7,15E+09	Sem1
chr6	6870920	6871320	0.000920569669121	Dlx6os1
			241	
chr6	8974813	8975213	0.000916341940665	Nxph1
			255	
chr6	9563754	9564154	0.000600686647524	Nxph1
			551	
chr6	11961055	11961455	0.000945259608674	Phf14
			975	
chr6	12037975	12038375	9,33E+09	Phf14
chr6	17721991	17722391	0.000790703851508	St7
			701	
chr6	22000786	22001186	0.000895806730038	Cped1
			418	
chr6	22574705	22575105	0.000590948701420	Fam3c
			402	
chr6	22590503	22590903	0.000646238894106	Fam3c
			421	
chr6	25260090	25260490	0.000302108272006	Tmem229a
			718	
chr6	25466762	25467162	4,68E+09	Gpr37
chr6	28656904	28657304	0.000840303898813	Snd1
			399	
chr6	28685194	28685594	0.000465078812643	Snd1
			016	
chr6	29467396	29467796	0.000326313428196	Atp6v1f
			556	
chr6	32439491	32439891	0.000225352885542	Plxna4os1
			485	
chr6	33399738	33400138	0.000375654532426	Exoc4
			273	

chr1	1014266	10142703	0.000356965729924	4933416M06	chr6	33739752	33740152	0.000769987789031	Exoc4
chr1	31	1	719	Rik				035	
chr1	1014818	10148223	0.000271704287776	Pik3r1	chr6	33808869	33809269	0.000302930281453	Exoc4
chr1	33	3	231					007	
chr1	1017520	10175241	0.000286637827750	Pik3r1	chr6	33824794	33825194	0.000536881033254	Exoc4
chr1	11	1	402					222	
chr1	1022906	10229107	0.000997928437916	Cd180	chr6	33826409	33826809	0.000763108737123	Exoc4
chr1	72	2	462					597	
chr1	1023229	10232330	0.000691864175092	Cd180	chr6	33882798	33883198	0.000709477351373	Exoc4
chr1	9	9	15					611	
chr1	1025446	10254505	0.000472445016439	Cd180	chr6	33924910	33925310	0.000147241175175	Exoc4
chr1	56	6	282					949	
chr1	1028648	10286521	0.000279968804379	Mast4	chr6	34836064	34836464	8,36E+09	Tmem140
chr1	3	4	749						
chr1	1076353	10763572	0.000749777930908	1700006H21	chr6	38807092	38807492	0.000931588680332	Hipk2
chr1	3	2	468	Rik				94	
chr1	1092150	10921544	0.000641581783069	Pde4d	chr6	38832652	38833052	0.000335280903456	Hipk2
chr1	48	8	167					026	
chr1	1092385	10923896	0.000982284884853	Pde4d	chr6	47174687	47175087	0.000909166709186	Cntnap2
chr1	3	1	08					06	
chr1	1102360	11023642	0.000806637475917	Rab3c	chr6	47203804	47204204	0.000844525083611	Cntnap2
chr1	23	3	249					991	
chr1	1121867	11218715	0.000797378370914	Ankrd55	chr6	47564934	47565334	0.000256069945586	Ezh2
chr1	3	3	981					912	
chr1	1132586	11325903	0.000882667704230	Gzmk	chr6	51108716	51109116	0.000596842393436	Mir148a
chr1	3	7	4					603	
chr1	1135901	11359054	5,94E+09	Snx18	chr6	52443124	52443524	0.000632470532788	1700094M24
chr1	3	8						254	Rik
chr1	1145238	11452428	0.000245709324587	Fst	chr6	53302474	53302874	0.000297317183183	Creb5
chr1	3	9	784					488	
chr1	1145710	11457144	0.000237795328211	4930467J12R	chr6	53413971	53414371	0.000698475161937	Creb5
chr1	3	9	998	ik				218	
chr1	1150855	11508592	0.000884402644636	Pelo	chr6	53602605	53603005	0.000905706254093	Creb5
chr1	3	9	184					808	
chr1	1045269	10453098	0.000668297112671	Fhit	chr6	54760801	54761201	3,63E+09	Znrf2
chr1	4	8	246						
chr1	1188579	11886194	0.000347667946951	Ptprg	chr6	55277431	55277831	0.000471683634794	Mindy4
chr1	4	4	18					218	
chr1	1199749	11997891	0.000706938286071	Ptprg	chr6	56222621	56223021	0.000735519126667	Pde1c
chr1	4	1	601					361	
chr1	1224024	12240640	0.000804099440437	Ptprg	chr6	57581906	57582306	0.000232476039456	Herc6
chr1	4	0	689					975	
chr1	1251740	12517802	0.000126483337341	Cadps	chr6	57582556	57582956	0.000923184507323	Herc6
chr1	4	2	332					427	
chr1	1278115	12781556	4,86E+08	Cadps	chr6	60790014	60790414	0.000710782588811	Snca
chr1	4	6						742	
chr1	1280223	12802636	0.000442343844047	Cadps	chr6	67115575	67115975	0.000674909198326	E230016M11
chr1	4	6	732					912	Rik
chr1	1324294	13243343	0.000232180766434	Gm5087	chr6	71938716	71939116	0.000243954377168	Polr1a
chr1	4	3	276					62	
chr1	1348848	13488885	0.000850756840341	Synpr	chr6	73306754	73307154	0.000985216228749	Suclg1
chr1	4	5	083					808	
chr1	1372049	13720899	0.000679760548317	Sntn	chr6	78516255	78516655	0.000448204974988	Reg3g
chr1	4	9	976					542	
chr1	1430919	14309597	0.000922008533651	Or2t1	chr6	79400740	79401140	6,52E+09	Gm20362
chr1	4	7	653						
chr1	1483770	14838104	0.000379284579142	Nek10	chr6	81756613	81757013	0.000449536186291	Mir7232
chr1	4	4	602					51	
chr1	1506313	15063539	0.000747852387931	Nek10	chr6	83267170	83267570	8,51E+08	Slc4a5
chr1	4	9	847						
chr1	1646782	16468229	0.000283917032938	Top2b	chr6	84098412	84098812	0.000825424392306	Dysf
chr1	4	9	952					706	
chr1	2096007	20960477	0.000878072164248	Vcl	chr6	85434105	85434505	0.000743780316492	Smyd5
chr1	4	7	733					693	
chr1	2115296	21153362	0.000223688860002	Adk	chr6	85971924	85972324	0.000461568430926	Dusp11
chr1	4	2	236					595	
chr1	2167329	21673698	0.000315802073160	Dusp29	chr6	86044930	86045330	0.000824496646953	Add2
chr1	4	8	176					499	
chr1	2236826	22368662	0.000433499937411	Lrmda	chr6	86504380	86504780	0.000586711879095	2310040G24
chr1	4	2	375					869	Rik
chr1	2245172	22452128	0.000608802985393	4930405A10	chr6	87216507	87216907	0.000639669880573	Antrx1
chr1	4	8	453	Rik				734	
chr1	2311461	23115011	0.000429732662806	Gm10248	chr6	87578293	87578693	0.000102827022575	Prokr1
chr1	4	1	215					58	
chr1	2381541	23815811	0.000333256612627	Kcnma1	chr6	87625364	87625764	0.000451923078585	Prokr1
chr1	4	1	73					326	
chr1	2465545	24655858	6,72E+09	Rps24	chr6	87916509	87916909	0.000270132128582	Hmces
chr1	4	8						734	
chr1	2491585	24916259	0.000197535199511	Gm32123	chr6	88384124	88384524	0.000460564157163	Eefsec
chr1	4	9	822					557	

chr1	2498262	24983028	0.000400079022062	Gm32123	chr6	88764573	88764973	0.000182117852765	Mgll
chr1	4	8	183		chr6	89659563	89659963	0.000121834043796	Txnrd3
chr1	4	2935788	0.000882296393724	Cacna2d3	chr6	90907904	90908304	0.000595680788028	lqsec1
chr1	4	5	724		chr6	93723696	93724096	0.000152804875967	Magi1
chr1	4	2940034	8,75E+09	Cacna2d3	chr6	94257076	94257476	0.000511000793451	4930511A08R
chr1	4	1			chr6	94508741	94509141	0.000787697845275	ik
chr1	4	3007542	0.000120745177187	Cacna1d	chr6	99267456	99267856	0.000203842203441	Slc25a26
chr1	4	3	409		chr6	99809143	99809543	0.000625667080474	Foxp1
chr1	4	3190235	0.000297720588178	Ankrd28	chr6	99923152	99923552	0.000930305568085	Prok2
chr1	4	3	008		chr6	10056802	10056842	0.000103631887294	Prok2
chr1	4	3190631	0.000566692472175	Ankrd28	chr6	10066855	10066895	0.000497740978944	1700049E22R
chr1	4	7	667		chr6	10255027	10255067	0.000571661802734	ik
chr1	4	3439334	0.000928167464153	Mmrn2	chr6	10447600	10447640	0.000974209472450	Shq1
chr1	4	2	029		chr6	10851164	10851204	0.000704086431320	Cntn3
chr1	4	3483385	0.000911092744360	Grid1	chr6	11256683	11256723	0.000704086431320	Cntn6
chr1	4	2	742		chr6	11279065	11279105	0.000694159187511	ltpr1
chr1	4	3522698	0.000617761383432	Grid1	chr6	11348451	11348491	0.000825114042207	Oxtr
chr1	4	7	613		chr6	11369062	11369102	0.000825114042207	Srgap3
chr1	4	3540722	0.000793426298416	Grid1	chr6	11423861	11423901	0.000253411466297	Crelld1
chr1	4	6	833		chr6	11441336	11441376	0.000253411466297	Irak2
chr1	4	3698780	0.000406817691650	Ccser2	chr6	11458465	11458505	0.000938023284526	Slc6a1
chr1	4	4	296		chr6	11469383	11469423	0.000302357022228	Hrh1
chr1	4	4590546	0.000295702844788	Ddhd1	chr6	11484606	11484646	0.000654629852761	Atg7
chr1	4	3	748		chr6	11547159	11547199	0.000756331402154	Atg7
chr1	4	4692773	0.000978690003130	Mir378c	chr6	11690557	11690597	0.000938181881005	Atg7
chr1	4	2	334		chr6	11727983	11728023	0.000756331402154	Pparg
chr1	4	4713724	0.000620308039752	Gm10371	chr6	11806931	11806971	0.000938181881005	Tmem72
chr1	4	1	215		chr6	11902267	11902307	0.000617244742556	Cxcl12
chr1	4	4790128	0.000300988801763	4930447J18R	chr6	11974479	11974519	0.000617244742556	Rasgef1a
chr1	4	5	769	ik	chr6	11987126	11987166	0.000435392853180	Cacna1c
chr1	4	4852281	0.000935862248426	Tmem260	chr6	11992376	11992416	0.000106061732447	Erc1
chr1	4	1	567		chr6	12000667	12000707	0.000914119503210	Erc1
chr1	4	4890367	4,82E+09	4933429O19	chr6	12114559	12114599	0.000424015929753	Wnk1
chr1	4	9		Rik	chr6	12128957	12128997	0.000424015929753	Wnk1
chr1	4	5187091	0.000414521668730	Mettl17	chr6	12542500	12542540	0.000879659316094	Mical3
chr1	4	9	794		chr6	12558097	12558137	0.000879659316094	Slc6a13
chr1	4	5230751	0.000644394639609	Mettl3	chr6	12698541	12698581	0.000931715393753	Cd9
chr1	4	8	08		chr6	12713894	12713934	0.000164284528561	Vwf
chr1	4	5505176	0.000750805431350	Zfhx2os	chr6	12805648	12805688	0.000164284528561	D6Wsu163e
chr1	4	3	976		chr6	12936579	12936619	0.000456904452949	Ccnd2
chr1	4	5538224	0.000218035126074	Dhrs4	chr6	0	0	0.000456904452949	Tspan9
chr1	4	4	989		chr6	0	0	0.000456904452949	Clec12a
chr1	4	5912183	0.000462863325857	1700129C05					
chr1	4	2	654	Rik					
chr1	4	5954455	0.000744887563719	Cdadcl					
chr1	4	5	931						
chr1	4	5980715	7,90E+09	Shisa2					
chr1	4	2							
chr1	4	5983684	0.000679968323101	Shisa2					
chr1	4	8	132						
chr1	4	6026604	0.000922025315856	Mtmer6					
chr1	4	5	612						
chr1	4	6107465	0.000485315372658	Tnfrsf19					
chr1	4	1	057						
chr1	4	6202344	0.000933793815862	Rnaseh2b					
chr1	4	3	671						
chr1	4	6246158	0.000998513678036	Gucy1b2					
chr1	4	3	606						
chr1	4	6353498	0.000718878660789	Mtmer9					
chr1	4	6	679						
chr1	4	6396597	0.000125571631574	4930578I06Ri					
chr1	4	9	006	k					
chr1	4	6499422	3,30E+09	Ints9					
chr1	4	8							
chr1	4	6517151	0.000833474386541	Extl3					
chr1	4	5	244						
chr1	4	6541674	0.000381595881531	Pnoc					
chr1	4	7	416						
chr1	4	6693371	0.000281152762930	Pnma2					
chr1	4	8	503						
chr1	4	6709535	0.000309500825953	Ppp2r2a					
chr1	4	8	49						
chr1	4	6767045	0.000881111906091	Cdca2					
chr1	4	4	636						
chr1	4	6905768	0.000439422753377	Stc1					
chr1	4	4	05						
chr1	4	7343041	0.000628796446065	Itm2b					
chr1	4	4	394						
chr1	4	7395961	0.000273375644856	Sucla2					
chr1	4	2	522						

chr1	7400379	74004191	0.000133609703600	4933402/15R	chr6	12939834	12939874	0.000362619301283	Clec1b
chr1	1		879	ik	chr6	7	7	34	
chr1	7507833	75078735	0.000403914941399	Rubcnl	chr6	13129218	13129258	0.000333386086001	Magohb
chr1	5		464		chr6	2	2	086	
chr1	7527808	75278481	0.000324615555342	Cpb2	chr6	13702484	13702524	0.000966297826360	Pde6h
chr1	1		023		chr6	5	5	263	
chr1	7661676	76617162	0.000835178587194	Serp2	chr6	13727195	13727235	0.000709032304950	Ptpro
chr1	2		951		chr6	0	0	563	
chr1	7762529	77625694	0.000210068038122	Enox1	chr6	13770654	13770694	0.000520314674940	Strap
chr1	4		291		chr6	6	6	625	
chr1	7888547	78885879	0.000174434119336	Vwa8	chr6	14060691	14060731	0.000915604343535	Aebp2
chr1	9		375		chr6	1	1	881	
chr1	7900716	79007566	0.000359807199599	Vwa8	chr6	14125144	14125184	0.000193327482803	Pde3a
chr1	6		16		chr6	1	1	207	
chr1	8432107	84321472	0.000724470409120	Pcdh17	chr6	14304796	14304836	0.000885880038467	C2cd5
chr1	2		936		chr6	2	2	235	
chr1	8615046	86150868	0.000920051202513	4930529K09	chr6	14413029	14413069	0.000710462003496	Sox5
chr1	8		15	Rik	chr6	4	4	115	
chr1	8695102	86951429	6.15E+08	Diaph3	chr6	14620270	14620310	0.000374886262459	Itpr2
chr1	9				chr6	5	5	295	
chr1	8703895	87039350	0.000741815761278	Diaph3	chr6	14674064	14674104	0.000304114673495	Stk38l
chr1	0		852		chr6	9	9	448	
chr1	8972383	89724239	0.000792088456714	4930474H20	chr6	14739046	14739086	0.000902693586805	Ccdc91
chr1	9		08	Rik	chr6	9	9	059	
chr1	9107042	91070828	0.000975095202301	4930474H20	chr6	14859855	14859895	0.000157442411207	Gm6313
chr1	8		324	Rik	chr6	5	5	853	
chr1	9369339	93693794	0.000674612311524	Pcdh9	chr7	12950016	12950416	0.000766495840196	Zfp324
chr1	4		88		chr7			724	
chr1	1019836	10198405	0.000981198382124	Lmo7	chr7	16874552	16874952	0.000204923460497	9330104G04
chr1	52		626		chr7			053	Rik
chr1	1035565	10355699	0.000235622697684	Scel	chr7	19307098	19307498	0.000808214431665	Fosb
chr1	99		734		chr7			411	
chr1	1048697	10487019	0.000766665301609	D130009I18R	chr7	28487678	28488078	0.000692751077312	1700028B04R
chr1	4		182	ik	chr7			925	ik
chr1	1053078	10530826	0.000267537266275	Ndfip2	chr7	28629824	28630224	0.000698928733708	Acp7
chr1	61		708		chr7			692	
chr1	1057988	10579929	0.000351416821191	Spry2	chr7	28772113	28772513	0.000608413875339	Sirt2
chr1	4		108		chr7			942	
chr1	1061783	10617875	0.000647551016696	Trim52	chr7	29516313	29516713	0.000280588222297	Sipa1l3
chr1	4		113		chr7			595	
chr1	1104690	11046943	0.000615875529404	Slitrk6	chr7	30007627	30008027	0.000551296210518	Zfp568
chr1	35		12		chr7			526	
chr1	1154508	11545122	0.000338066954319	4930505G20	chr7	30168179	30168579	0.000989716749539	Zfp146
chr1	26		816	Rik	chr7			626	
chr1	1176575	11765796	0.000710083899838	Mir6239	chr7	35487734	35488134	0.000839296639276	Tdrd12
chr1	61		62		chr7			288	
chr1	1187116	11871204	0.000131171020493	Abcc4	chr7	35565443	35565843	0.000828989695129	Nudt19
chr1	43		648		chr7			012	
chr1	1193014	11930188	0.000384623728800	Hs6st3	chr7	36913076	36913476	0.000314437574801	Tshz3
chr1	83		548		chr7			651	
chr1	1196656	11966600	0.000581098290173	1700006F04R	chr7	37241348	37241748	0.000359652345405	Zfp536
chr1	4		943	ik	chr7			248	
chr1	1196691	11966953	0.000900212640187	1700006F04R	chr7	37374318	37374718	0.000656918363367	Zfp536
chr1	4		581	ik	chr7			12	
chr1	1199895	11998995	0.000858331483153	Oxgr1	chr7	43766815	43767215	0.000323348438929	Klk12
chr1	52		174		chr7			116	
chr1	1203057	12030612	0.000720150231846	Mbnl2	chr7	44495003	44495403	0.000808229413545	Emc10
chr1	26		128		chr7			192	
chr1	1204768	12047725	0.000365827574948	Rap2a	chr7	48807875	48808275	3.67E+09	Zdhhc13
chr1	4		296		chr7				
chr1	1208057	12080614	0.000828103297466	lpo5	chr7	49305387	49305787	0.000467105049857	Nav2
chr1	4		853		chr7			919	
chr1	1212512	12125163	7.97E+09	Farp1	chr7	49389950	49390350	0.000217189694911	Nav2
chr1	4		1		chr7			69	
chr1	1213743	12137470	0.000771709091062	Stk24	chr7	54954148	54954548	0.000751436070797	Luzp2
chr1	4		611		chr7			862	
chr1	1236267	12362719	1.55E+09	Nalcn	chr7	55189751	55190151	0.000513551448997	Luzp2
chr1	4		0		chr7			784	
chr1	1238431	12384353	0.000521838609862	1700024B18	chr7	63563882	63564282	0.000536080674098	Otud7a
chr1	4		657	Rik	chr7			827	
chr1	6575015	6575415	0.000978223254336	Fyb	chr7	64156572	64156972	0.000336908420076	Trpm1
chr1	5		941		chr7			831	
chr1	6716684	6717084	0.000102468371531	Rictor	chr7	64456041	64456441	0.000363951278877	Apba2
chr1	5		605		chr7			117	
chr1	7131074	7131474	0.000591756977623	Lifr	chr7	65449598	65449998	0.000789307709345	Tjp1
chr1	5		925		chr7			266	
chr1	8863669	8864069	0.000202991119336	Ranbp3l	chr7	65919900	65920300	0.000248204902224	Pcsk6
chr1	5		637		chr7			991	
chr1	1247400	12474407	0.000609271084040	Pdzd2	chr7	72837914	72838314	0.000510312367868	Gm6567
chr1	5		186		chr7			693	

chr1	1321050	13210901	0.000155241762094	Cdh6	chr7	76942686	76943086	0.000105773278116	Agbl1
5	1		693					228	
chr1	1378296	13783368	0.000704320060863	Cdh6	chr7	77980825	77981225	0.000928408465108	Ntrk3
5	8		85					665	
chr1	2528342	25283828	0.000392421724251	Basp1	chr7	78063901	78064301	0.000835849427535	Ntrk3
5	8		912					562	
chr1	2748169	27482090	0.000816531019777	Ank	chr7	78338458	78338858	0.000849250675782	Ntrk3
5	0		633					232	
chr1	2810595	28106353	0.000457404038836	Trio	chr7	78451518	78451918	0.000667908327746	Ntrk3
5	3		762					547	
chr1	2993200	29932401	9,96E+09	Mir3964	chr7	78632497	78632897	0.000876383322198	E430016F16R
5	1							082	ik
chr1	3005101	30051414	0.000646438523690	Ctnnd2	chr7	80374323	80374723	0.000283778342097	Man2a2
5	4		618					03	
chr1	3240198	32402384	0.000998453673406	Sema5a	chr7	80637809	80638209	0.000516915823274	Gm15880
5	4		008					289	
chr1	3243314	32433548	0.000480870082557	Sema5a	chr7	82424413	82424813	0.000667050309150	Adamts13
5	8		15					152	
chr1	3248791	32488310	0.000711857394175	Sema5a	chr7	82485004	82485404	0.000183087383548	Adamts13
5	0		786					441	
chr1	3665784	36658243	0.000940508231588	Pabpc1	chr7	83220484	83220884	0.000718421726467	A530021J07Ri
5	3		533					522	k
chr1	3721546	37215860	0.000472898634249	Grhl2	chr7	83238410	83238810	0.000625892645223	A530021J07Ri
5	0		708					17	k
chr1	3743153	37431937	0.000902662737094	Gm15941	chr7	83258683	83259083	0.000122934060970	A530021J07Ri
5	7		495					303	k
chr1	3751552	37515926	0.000396980614312	Ncald	chr7	83881539	83881939	0.000634537726889	Tlnrd1
5	6		481					262	
chr1	3833781	38338219	0.000452155240693	Klf10	chr7	87770802	87771202	0.000687854819863	Grm5
5	9		086					665	
chr1	4003339	40033798	0.000793154752050	Gm33251	chr7	88048098	88048498	7,13E+09	Ctsc
5	8		773						
chr1	4082303	40823434	0.000681867136552	Zfpm2	chr7	88720188	88720588	0.000921290536536	Rab38
5	4		716					556	
chr1	4103380	41034200	8,00E+09	4930555K19	chr7	89116084	89116484	0.000154276567090	Tmem135
5	0			Rik				106	
chr1	4171892	41719322	0.000811778453523	Oxr1	chr7	91567115	91567515	0.000600709689514	Dlg2
5	2		537					867	
chr1	4287186	42872269	0.000849850737774	Angpt1	chr7	92323044	92323444	0.000920918869637	Dlg2
5	9		771					966	
chr1	4319600	43196408	0.000923688112796	Rsp02	chr7	93012323	93012723	5,67E+08	Prcp
5	8		75						
chr1	4935506	49355465	0.000865542331301	Csmd3	chr7	95228836	95229236	0.000129092832548	Tenm4
5	5		095					139	
chr1	5023706	50237464	0.000305730083993	Trps1	chr7	95477720	95478120	0.000922704204626	Tenm4
5	4		261					035	
chr1	5097081	50971216	0.000352682990634	Trps1	chr7	96678041	96678441	0.000167856898427	Tenm4
5	6		845					354	
chr1	5346029	53460694	7,83E+09	Ext1	chr7	97257375	97257775	0.000746165675540	Usp35
5	4							089	
chr1	5493454	54934946	0.000591188744055	Enpp2	chr7	97413653	97414053	0.000860850930597	Thrsp
5	6		06					179	
chr1	5555821	55558615	0.000646083949492	Mrpl13	chr7	97697922	97698322	4,80E+09	Rsf1os2
5	5		615						
chr1	5584352	55843925	0.000238135418927	Sntb1	chr7	98195157	98195557	0.000191829394416	B3gnt6
5	5		81					532	
chr1	5600416	56004560	0.000875221996130	Sntb1	chr7	99217001	99217401	0.000520083135381	Mogat2
5	0		326					544	
chr1	5611902	56119421	0.000770198591925	Sntb1	chr7	10076919	10076959	0.000558530023464	Fam168a
5	1		952			9	9	51	
chr1	5841473	58415139	0.000751653347735	Fam91a1	chr7	10110742	10110782	0.000968361035923	Fchsd2
5	9		637			7	7	238	
chr1	5945364	59454048	0.000244830690925	Nsmce2	chr7	10112850	10112890	0.000113148569202	Fchsd2
5	8		291			4	4	526	
chr1	5945485	59455258	0.000899588292217	Nsmce2	chr7	10201001	10201041	0.000497356017467	Il18bp
5	8		812			7	7	92	
chr1	5947217	59472571	0.000996403267777	Nsmce2	chr7	10579617	10579657	0.000756981475515	Dchs1
5	1		835			6	6	506	
chr1	5995447	59954873	0.000476175497383	4933412E24R	chr7	10768018	10768058	0.000263389369832	Ppfibp2
5	3		879	ik		1	1	945	
chr1	6050931	60509711	0.000590757252581	9930014A18	chr7	11115603	11115643	0.000886691536578	Zbed5-ps
5	1		398	Rik		4	4	388	
chr1	6195919	61959593	0.000162391920608	Myc	chr7	11129330	11129370	0.000869614295924	Zbed5-ps
5	3		407			3	3	137	
chr1	6219896	62199364	0.000721086119941	Pvt1	chr7	11151356	11151396	0.000779051943929	Galnt18
5	4		372			3	3	243	
chr1	6362901	63629415	4,38E+09	Gm20740	chr7	11268869	11268909	1,72E+09	Tead1
5	5					3	3		
chr1	6408466	64085068	8,81E+09	Asap1	chr7	11286546	11286586	0.000238991614995	Rassf10
5	8					8	8	41	

chr1	6591042	65910823	0.000956415659481	Oc90
5	3		236	
chr1	6682332	66823728	0.000407339477724	Sla
5	8		2	
chr1	6835188	68352281	0.000965791435136	Mir30d
5	1		643	
chr1	6924797	69248377	0.000908742237985	4930504C09
5	7		649	Rik
chr1	7060080	70601200	0.000327235398430	Gm19782
5	0		936	
chr1	7205898	72059389	0.000960006836495	Col22a1
5	9		664	
chr1	7266759	72667993	0.000444251151874	Kcnk9
5	3		964	
chr1	7289727	72897677	0.000924318560032	Peg13
5	7		944	
chr1	7373947	73739874	0.000874998215027	Ptp4a3
5	4		345	
chr1	7453034	74530745	0.000824629392521	Adgrb1
5	5		148	
chr1	7469629	74696695	0.000659409209859	4933427E11R
5	5		074	ik
chr1	7632347	76323875	0.000382498126057	Exosc4
5	5		84	
chr1	7716617	77166576	7.52E+09	Rbfox2
5	6			
chr1	7920113	79201530	0.000809952129322	Gm10863
5	0		555	
chr1	7930785	79308256	0.000885083350171	Pla2g6
5	6		678	
chr1	7988120	79881603	0.000291402503527	Apobec3
5	3		882	
chr1	8058646	80586863	0.000305413084555	Grap2
5	3		45	
chr1	8180301	81803416	0.000889035086881	Tef
5	6		984	
chr1	8290532	82905721	0.000141385970837	Tcf20
5	1		238	
chr1	8336532	83365720	0.000247592852755	1700001L05R
5	0		663	ik
chr1	8453179	84532198	0.000931447110573	Rtl6
5	8		691	
chr1	8496390	84964308	0.000943260175778	5031439G07
5	8		171	Rik
chr1	8501747	85017872	0.000339096912740	Upk3a
5	2		811	
chr1	8629031	86290711	0.000828906457302	Tbc1d22a
5	1		159	
chr1	8633578	86336184	0.000221009950592	Tbc1d22a
5	4		377	
chr1	9125283	91253235	0.000313201713123	Abcd2
5	5		545	
chr1	9179668	91797088	0.000512697608694	Lrrk2
5	8		879	
chr1	9251509	92515499	6.90E+09	Pdznr4
5	9			
chr1	9373279	93733196	0.000911335611229	Prickle1
5	6		58	
chr1	9481486	94815268	0.000967386712210	Tmem117
5	8		638	
chr1	9591495	95915352	0.000668149650703	Ano6
5	2		72	
chr1	9593070	95931105	0.000633360019330	Ano6
5	5		009	
chr1	9606000	96060404	0.000636725085815	D030018L15
5	4		026	Rik
chr1	9776282	97763220	0.000620973471664	Rapgef3
5	0		676	
chr1	9782071	97821113	0.000934685181812	Hdac7
5	3		611	
chr1	9940095	99401354	0.000776636709922	Tmbim6
5	4		895	
chr1	1001261	10012659	0.000859482394529	Dip2b
5	96	6	82	
chr1	1013683	10136873	0.000648646867761	Krt80
5	30	0	116	
chr1	1014386	10143901	0.000958203115960	Krt87
5	18	8	732	
chr1	1022017	10220211	0.000743787552784	Zfp740
5	12	2	259	

chr7	11325722	11325762	0.000633160874659	Bmal1
5	5	7		
chr7	11334730	11334770	0.000945502029040	Btbd10
1	1	757		
chr7	11373485	11373525	0.000899901192449	Spon1
3	3	759		
chr7	11406618	11406658	0.000969725972651	Rras2
2	2	769		
chr7	11418100	11418140	0.000114650296878	Copb1
9	9	738		
chr7	11500335	11500375	0.000635172313007	A730082K24R
9	9	137	ik	
chr7	11592046	11592086	0.000258189896050	Sox6
1	1	531		
chr7	11644133	11644173	0.000319273348184	Pik3c2a
5	5	712		
chr7	11769791	11769831	0.000646715248403	Xylt1
9	9	762		
chr7	11773806	11773846	0.000804253888715	Xylt1
0	0	634		
chr7	11962993	11963033	0.000252835290934	Acsn1
9	9	783		
chr7	12148526	12148566	0.000330870911203	Hs3st2
5	5	762		
chr7	12206375	12206415	0.000610316945103	Ears2
6	6	939		
chr7	12467052	12467092	0.000984832528572	3100003L05R
2	2	621	ik	
chr7	12751408	12751448	0.000703244364340	Srcap
8	8	216		
chr7	13000177	13000217	0.000540651989402	Gm4265
4	4	452		
chr7	13011406	13011446	0.000367081911510	Fgfr2
5	5	914		
chr7	13015297	13015337	0.000300418181552	Fgfr2
4	4	908		
chr7	13255506	13255546	0.000744845917047	Oat
7	7	82		
chr7	13347941	13347981	0.000470527470026	Tex36
8	8	294		
chr7	13411442	13411482	0.000293168787015	Adam12
0	0	13		
chr7	13414955	13414995	0.000742717983190	Adam12
1	1	033		
chr7	13430914	13430954	0.000606921623204	D7Ertdd443e
2	2	082		
chr7	13700211	13700251	0.000259448280358	Mgmt
9	9	177		
chr7	13714729	13714769	0.000669904806364	Gm10578
8	8	53		
chr7	13734620	13734660	0.000423658677620	Ebf3
0	0	287		
chr7	13886350	13886390	0.000534829950065	Ppp2r2d
8	8	152		
chr7	14141482	14141522	1.93E+08	Gatd1
8	8			
chr7	14195080	14195120	9.67E+09	Brsk2
1	1			
chr7	14342834	14342874	0.000223564289929	Cdkn1c
5	5	312		
chr7	14454406	14454446	0.000259859598917	Ppfia1
3	3	167		
chr7	14515933	14515973	0.000188483953548	Smim38
2	2	868		
chr8	7484771	7485171	0.000732477700327	Efnb2
		916		
chr8	8131713	8132113	0.000672213473807	Efnb2
		748		
chr8	8476649	8477049	0.000935862901030	Efnb2
		76		
chr8	8883176	8883576	0.000440158409161	Arglu1
		784		
chr8	10259934	10260334	0.000840394091482	Myo16
		954		
chr8	11324390	11324790	0.000455436332192	Col4a1
		916		
chr8	11780215	11780615	0.000920097162185	Arhgef7
		44		
chr8	12233930	12234330	0.000352312512087	Sox1ot
		005		

chr1 5	1032366 79	10323707 9	0.000615876426368 844	Cbx5	chr8 13461409	13461809	0.000729749900123 089	Tmem255b
chr1 6	5049366	5049766	0.000832649360095 417	Glyr1	chr8 13915404	13915804	0.000254732351173 237	Fbxo25
chr1 6	5930315	5930715	0.000539189681462 16	Rbfox1	chr8 14072060	14072460	0.000821105857270 465	Erich1
chr1 6	8708938	8709338	0.000785457185212 39	Usp7	chr8 15994226	15994626	0.000799755327008 933	Csmd1
chr1 6	8926383	8926783	0.000536876422682 742	Hapstr1	chr8 16249090	16249490	0.000782892082427 988	Mir3106
chr1 6	9132818	9133218	0.000569920464596 363	Hapstr1	chr8 22660646	22661046	0.000892961868007 358	Ikbkb
chr1 6	1029586 4	10296264	0.000634243736980 178	Emp2	chr8 23477694	23478094	0.000246040691175 953	Sfrp1
chr1 6	1034742 4	10347824	0.000831599673146 277	Gm1600	chr8 24418156	24418556	0.000482332973537 059	Tcim
chr1 6	1094555 5	10945955	0.000372408128166 191	Gm21859	chr8 26880149	26880549	0.000799102117463 59	2310008N11 Rik
chr1 6	1117267 5	11173075	0.000112694685972 858	Zc3h7a	chr8 27912250	27912650	0.000702871094511 615	Gm8096
chr1 6	1153098 5	11531385	0.000391158573703 31	Snx29	chr8 28095441	28095841	0.000404571646970 942	Gm8096
chr1 6	1367294 4	13673344	0.000709124680565 72	Bfar	chr8 28333169	28333569	0.000574495063652 591	Gm8096
chr1 6	1402919 9	14029599	0.000696564255604 049	Ifitm7	chr8 30362564	30362964	0.000595174719905 556	4933433F19R ik
chr1 6	1519573 2	15196132	0.000735740060163 081	Clxn	chr8 32290543	32290943	0.000486418315400 305	Nrg1
chr1 6	1621183 1	16212231	0.000247970416351 384	Pkp2	chr8 32317247	32317647	0.000428895009872 31	Nrg1
chr1 6	1692868 6	16929086	0.000693104769261 765	Ppm1f	chr8 33873828	33874228	0.000918366988550 412	Rbpms
chr1 6	1711002 1	17110421	0.000113655477455 01	Ppil2	chr8 33898887	33899287	0.000906514188766 083	Rbpms
chr1 6	1765655 4	17656954	8,56E+08	Med15	chr8 34158587	34158987	0.000980074186042 179	Saraf
chr1 6	1818216 0	18182560	0.000450760878012 879	Mir6366	chr8 34695692	34696092	0.000151255935739 503	B930018H19 Rik
chr1 6	1835160 8	18352008	0.000525079406031 637	Arvcf	chr8 35697927	35698327	0.000259542277185 387	Gm19410
chr1 6	2006887 7	20069277	0.000155749451997 478	Klhl24	chr8 35886585	35886985	0.000748766342729 69	5430403N17 Rik
chr1 6	2011553 8	20115938	0.000744360843759 173	Klhl24	chr8 35907863	35908263	0.000430239861828 33	5430403N17 Rik
chr1 6	2047688 0	20477280	0.000953066711574 913	Eif2b5	chr8 35920145	35920545	0.000180347885237 797	5430403N17 Rik
chr1 6	2120193 0	21202330	0.000601135338572 645	Ephb3	chr8 36911321	36911721	0.000537914095987 135	Dlc1
chr1 6	2212273 6	22123136	0.000408853360857 497	Igf2bp2	chr8 47266902	47267302	0.000335364522121 573	Stox2
chr1 6	2248854 1	22488941	0.000295737267843 155	9230117E06R ik	chr8 49288909	49289309	0.000305138251972 981	Tenm3
chr1 6	2323108 1	23231481	0.000264223821825 116	St6gal1	chr8 50926366	50926766	0.000763834645505 446	Gm2516
chr1 6	2413243 9	24132839	0.000551156521238 383	Gm46545	chr8 56478098	56478498	0.000319703940665 05	4930518J21Ri k
chr1 6	2428487 9	24285279	0.000867262123502 854	Lpp	chr8 56784453	56784853	0.000828492595853 492	Fbxo8
chr1 6	2436603 6	24366436	0.000909356224804 37	Lpp	chr8 57363022	57363422	7,28E+09	Hdnr
chr1 6	2439958 3	24399983	0.000730956107682 701	Lpp	chr8 58032166	58032566	0.000130007557851 641	Galntl6
chr1 6	2443319 2	24433592	0.000701083354688 77	Lpp	chr8 59439307	59439707	0.000911497046155 23	4930470O06 Rik
chr1 6	2462466 0	24625060	0.000429088531299 859	Morf4l1-ps1	chr8 60995276	60995676	0.000963475911240 685	Nek1
chr1 6	2607010 6	26070506	0.000118276326635 858	P3h2	chr8 61431260	61431660	0.000745831481458 57	Cbr4
chr1 6	2610615 8	26106558	0.000146943626273 099	P3h2	chr8 66409059	66409459	0.000517136782668 404	Marchf1
chr1 6	2667163 3	26672033	0.000425222374556 444	Il1rap	chr8 66722241	66722641	0.000704503021828 229	Npy1r
chr1 6	2819017 0	28190570	0.000972813356825 197	Fgf12	chr8 67085693	67086093	0.000290569528073 936	Naf1
chr1 6	2972846 0	29728860	0.000687546261280 536	Opa1	chr8 68519836	68520236	0.000633084095152 503	Csgalnact1
chr1 6	3016070 7	30161107	0.000822798763211 23	Hes1	chr8 70045101	70045501	0.000927978706058 541	Sugp1
chr1 6	3018490 5	30185305	0.000135148529390 339	Cpn2	chr8 70198189	70198589	0.000314670617222 124	Slc25a42

chr1	3081827	30818671	0.000782158435771	Fam43a	chr8	70720657	70721057	0.000436478158611	Gm3336
chr1	1		008		chr8			253	
chr1	3293167	32932076	0.000262998380750	Lrch3	chr8	71378977	71379377	0.000402709661130	Nr2f6
chr1	6		945		chr8			096	
chr1	3419557	34195978	0.000840762371725	Kalrn	chr8	72543665	72544065	0.000762495582607	Med26
chr1	8		701		chr8			547	
chr1	3441726	34417666	0.000567627909305	Kalrn	chr8	77621385	77621785	0.000438838550874	1700092C02
chr1	6		424		chr8			898	Rik
chr1	3778016	37780561	0.000120717922220	Fstl1	chr8	78026633	78027033	0.000363352034312	Ttc29
chr1	1		237		chr8			037	
chr1	3895928	38959680	0.000506136681081	Igsf11	chr8	78518566	78518966	0.000788585010151	Slc10a7
chr1	0		903		chr8				
chr1	3926732	39267726	0.000292077769844	Gm36742	chr8	80257197	80257597	0.000713842661693	Hhip
chr1	6		079		chr8			054	
chr1	3942510	39425507	0.000825274127151	Gm36742	chr8	81354756	81355156	4,72E+08	Inpp4b
chr1	7		276		chr8				
chr1	4005306	40053467	0.000824963334382	Lsamp	chr8	81826951	81827351	0.000830071521999	Inpp4b
chr1	6		423		chr8			098	
chr1	4215704	42157440	0.000589289672767	Lsamp	chr8	82346747	82347147	0.000838023337922	Il15
chr1	0		279		chr8			156	
chr1	4267730	42677708	0.000430462130425	4932412D23	chr8	82402061	82402461	0.000988321639225	Il15
chr1	8		857	Rik	chr8			465	
chr1	4361525	43615651	0.000417960204249	Mir568	chr8	82695877	82696277	0.000706580499499	Zfp330
chr1	1		779		chr8			454	
chr1	4579152	45791926	0.000107504719673	Phldb2	chr8	82700467	82700867	0.000560460661422	Zfp330
chr1	6		246		chr8			013	
chr1	4642809	46428494	0.000927328850786	Nectin3	chr8	83274656	83275056	0.000692576236748	Ucp1
chr1	6		946		chr8			01	
chr1	4687695	46877351	0.000325693914125	Nectin3	chr8	85677852	85678252	0.000427449026589	Neto2
chr1	1		624		chr8			2	
chr1	4927486	49275265	0.000348100877190	Myh15	chr8	86864583	86864983	0.000609463672103	N4bp1
chr1	6		172		chr8			461	
chr1	5061558	50615985	0.000269269974683	Ccdc54	chr8	87182848	87183248	7,23E+09	Gm10637
chr1	5		795		chr8				
chr1	5325106	53251461	0.000783736737232	4930404A05	chr8	87874653	87875053	0.000528082708792	Zfp423
chr1	1		088	Rik	chr8			121	
chr1	5731729	57317697	0.000291947553208	Tmem30c	chr8	88922612	88923012	0.000542296092283	Mir8110
chr1	6		161		chr8			401	
chr1	6242761	62428010	0.000428858891457	Nsun3	chr8	89023510	89023910	0.000679257138523	Mir8110
chr1	0		606		chr8			459	
chr1	6784422	67844624	0.000473910688613	Cadm2	chr8	89027000	89027400	0.000254784687412	Mir8110
chr1	6		799		chr8			357	
chr1	6984766	69848064	0.000805242741836	Speer2	chr8	91455736	91456136	0.000913347165975	Fto
chr1	6		798		chr8			908	
chr1	6998045	69980858	0.000846499603179	Speer2	chr8	94624931	94625331	0.000462329683242	Rspry1
chr1	8		174		chr8			224	
chr1	7279549	72795890	0.000307115525379	Robo1	chr8	94704785	94705185	0.000571689485411	Plip
chr1	6		986		chr8			087	
chr1	7385362	73854024	0.000131651267626	Robo2	chr8	95661500	95661900	0.000307998777694	Ndrgr4
chr1	4		376		chr8			607	
chr1	7437784	74378243	0.000828123795931	Robo2	chr8	96297537	96297937	3,54E+09	Got2
chr1	6		113		chr8				
chr1	7485707	74857477	0.000334300809009	Robo2	chr8	10380812	10380852	0.000780561649908	4933400L20R
chr1	6		092		chr8	9	9	584	ik
chr1	7764385	77644257	0.000482331880490	Mir125b-2	chr8	10594209	10594249	0.000527960723808	Lcat
chr1	7		738		chr8	2	2	896	
chr1	7766437	77664778	0.000973508397566	Mir125b-2	chr8	10633850	10633890	0.000759912405148	Smpd3
chr1	6		38		chr8	2	2	055	
chr1	7766491	77665319	0.000531042728792	Mir125b-2	chr8	10675328	10675368	0.000446014472251	Tango6
chr1	9		563		chr8	8	8	344	
chr1	8705056	87050965	0.000443536958271	Gm32865	chr8	10686151	10686191	0.000621244224295	Has3
chr1	5		595		chr8	3	3	88	
chr1	8899369	88994091	0.000732602921014	Krtap22-2	chr8	10694720	10694760	0.000503028889269	Sntb2
chr1	6		847		chr8	9	9	119	
chr1	8949362	89494022	0.000251632220825	Krtap8-1	chr8	10731524	10731564	0.000138656080758	Nfat5
chr1	2		249		chr8	0	0	137	
chr1	8967658	89676984	0.000305490410316	Krtap11-1	chr8	10757715	10757755	0.000403262643756	Psmad7
chr1	6		356		chr8	5	5	674	
chr1	8988983	89890235	0.000108987536255	Tiam1	chr8	10820137	10820177	0.000545086619381	Zfhx3
chr1	5		212		chr8	3	3	685	
chr1	9048357	90483973	0.000803266715152	Hunk	chr8	10950389	10950429	0.000292224251195	Pmfbp1
chr1	3		962		chr8	0	0	794	
chr1	9084452	90844922	0.000819234487041	Eva1c	chr8	10982512	10982552	6,65E+09	Snord71
chr1	2		687		chr8	3	3		
chr1	9085783	90858237	0.000918166102240	Eva1c	chr8	10995384	10995424	0.000218433007712	Marveld3
chr1	6		825		chr8	5	5	182	
chr1	9107257	91072975	0.000939211386850	4932438H23	chr8	11066727	11066767	0.000247260240739	Vac14
chr1	5		047	Rik	chr8	8	8	067	
chr1	9172671	91727114	0.000795843750805	Cryz11	chr8	11073276	11073316	0.000694365688691	Mtss2
chr1	6		325		chr8	9	9	523	

chr1	9230010	92300502	0.000634073421698	Smim11	chr8	11156159	11156199	0.000210236808926	Znrf1
chr1	6	2	966		chr8	7	7	29	
chr1	9372180	93722205	0.000110404452111	Dop1b	chr8	11169297	11169337	0.000841356842402	Ctrb1
chr1	6	5	19		chr8	0	0	385	
chr1	9373247	93732870	0.000784922932586	Dop1b	chr8	11192327	11192367	0.000534082402247	Tmem231
chr1	6	0	918		chr8	3	3	06	
chr1	9435993	94360337	0.000976248822146	Pigp	chr8	11430653	11430693	0.000668433701366	Vat1l
chr1	6	7	975		chr8	2	2	045	
chr1	9447448	94474887	0.000970557312723	Ttc3	chr8	11450696	11450736	0.000439510149652	Wwox
chr1	6	7	219		chr8	5	5	6	
chr1	9452539	94525794	0.000881114781729	Vps26c	chr8	11500729	11500769	0.000380680905468	Wwox
chr1	6	4	586		chr8	0	0	5	
chr1	9459152	94591928	0.000386771385394	Dyrk1a	chr8	11557113	11557153	0.000943183154396	Maf
chr1	6	8	669		chr8	0	0	898	
chr1	9461823	94618635	0.000227236428499	Dyrk1a	chr8	11582669	11582709	7,75E+07	Maf
chr1	6	5	866		chr8	3	3		
chr1	9618985	96190250	0.000742817492423	Lca5l	chr8	11604296	11604336	0.000698815191383	Maf
chr1	6	0	775		chr8	0	0	373	
chr1	9713669	97137092	0.000215435144772	Dscam	chr8	11736310	11736350	0.000270514162050	Cmip
chr1	6	2	735		chr8	1	1	262	
chr1	5191502	5191902	0.000974119693878	Arid1b	chr8	11882976	11883016	0.000227115757018	Cdh13
chr1	7		617		chr8	3	3	392	
chr1	5942267	5942667	1,50E+09	Synj2	chr8	11943224	11943264	0.000191722754896	Osgin1
chr1	7				chr8	1	1	206	
chr1	7578993	7579393	0.000256851502981	Fndc1	chr8	11966105	11966145	0.000326543475415	Wfdc1
chr1	7		27		chr8	8	8	164	
chr1	8240237	8240637	0.000786468345104	Ccr6	chr8	11991633	11991673	0.000305224663671	Usp10
chr1	7		652		chr8	9	9		
chr1	8448341	8448741	0.000742521907797	T	chr8	12034254	12034294	0.000472050900634	Gse1
chr1	7		222		chr8	5	5	846	
chr1	8604961	8605361	0.000757787415609	Pde10a	chr8	12180856	12180896	0.000830037694611	Klhdc4
chr1	7		279		chr8	5	5	319	
chr1	9679492	9679892	0.000941630804613	Pabpc6	chr8	12218180	12218220	0.000751759433662	Zfp469
chr1	7		736		chr8	7	7	125	
chr1	9926510	9926910	3,29E+09	4930452A19	chr8	12222621	12222661	0.000942242572621	Zfp469
chr1	7			Rik	chr8	6	6	321	
chr1	9946603	9947003	0.000953470155009	4930452A19	chr8	12233292	12233332	0.000442827162030	Zfpm1
chr1	7		795	Rik	chr8	8	8	923	
chr1	1016243	10162834	0.000272888494464	Qki	chr8	12295010	12295050	0.000444178701606	Ankrd11
chr1	7		052		chr8	7	7	512	
chr1	1040217	10402579	0.000902969468968	A230009B12	chr8	12343591	12343631	0.000578234086739	Def8
chr1	7		84	Rik	chr8	4	4	071	
chr1	1229067	12291073	0.000205412218009	Map3k4	chr8	12592035	12592075	0.000906166793010	Map3k21
chr1	7		071		chr8	4	4	588	
chr1	1236413	12364532	0.000343086622904	Plg	chr8	12668675	12668715	0.000703292081512	Irf2bp2
chr1	7		609		chr8	2	2	43	
chr1	1355011	13550513	8,36E+09	2700054A10	chr8	12691290	12691330	0.000746188235335	Tomm20
chr1	7			Rik	chr8	3	3	676	
chr1	1417340	14173804	0.000328156330188	Dact2	chr8	12725831	12725871	0.000998023202442	Pard3
chr1	7		012		chr8	6	6	91	
chr1	1439298	14393382	0.000275003904264	4930474M22	chr9	3623056	3623456	0.000197170535934	Gucy1a2
chr1	7		493	Rik	chr9			71	
chr1	2196361	21964017	0.000560977102431	Zfp943	chr9	4457478	4457878	0.000567863589016	Msantd4
chr1	7		526		chr9			948	
chr1	2222786	22228262	0.000254774866467	Zfp994	chr9	4537072	4537472	0.000315997575707	Msantd4
chr1	7		411		chr9			216	
chr1	2387453	23874933	0.000629706913677	Prss21	chr9	6084135	6084535	0.000808962865452	Pdgfd
chr1	7		586		chr9			179	
chr1	2458346	24583864	0.000157236732428	Pkd1	chr9	6959006	6959406	0.000494643634328	Dync2h1
chr1	7		512		chr9			717	
chr1	2480473	24805135	0.000521639785429	Meiob	chr9	8634013	8634413	0.000826302537156	Trpc6
chr1	7		739		chr9			209	
chr1	2509939	25099794	0.000559871988199	Telo2	chr9	9180901	9181301	0.000877627177765	Gm16833
chr1	7		401		chr9			072	
chr1	2521161	25212013	0.000459676760566	Unkl	chr9	13475654	13476054	0.000326244563870	Phxr4
chr1	7		907		chr9			537	
chr1	2767386	27674260	0.000915697786240	Pacsin1	chr9	13569433	13569833	0.000626326125339	Maml2
chr1	7		566		chr9			638	
chr1	2823443	28234839	2,87E+09	Ppard	chr9	13676331	13676731	0.000905584220229	Maml2
chr1	7				chr9			461	
chr1	3083498	30835387	0.000151364075592	Dnah8	chr9	14279526	14279926	0.000208570283851	Sesn3
chr1	7		575		chr9			461	
chr1	3107113	31071531	0.000150820344232	Abcg1	chr9	14398787	14399187	0.000828302264028	Endod1
chr1	7		416		chr9			746	
chr1	3135323	31353638	0.000945591156954	Slc37a1	chr9	14571634	14572034	0.000336100633614	Amotl1
chr1	7		695		chr9			445	
chr1	3206282	32063227	0.000792761747103	Rrp1b	chr9	14713096	14713496	0.000935004544123	Piwi4
chr1	7		743		chr9			208	
chr1	3217759	32177990	0.000354166382891	Notch3	chr9	14907484	14907884	0.000795325803668	Izumo1r
chr1	7		43					367	

chr1	3375872	33759124	0.000871349693992	Rab11b	chr9	16383553	16383953	0.000879803882368	Fat3
chr1	4		94					08	
chr1	3518220	35182603	0.000285544543861	Lst1	chr9	22047863	22048263	2,06E+09	Elavl3
chr1	3		205						
chr1	4084712	40847528	0.000894415428768	9130008F23R	chr9	22327340	22327740	1,93E+09	Gm6607
chr1	8		225	ik					
chr1	4258060	42581002	0.000887833277420	Opn5	chr9	25479968	25480368	0.000120647986803	Eepd1
chr1	2		924					89	
chr1	4283978	42840185	0.000178757374705	Cd2ap	chr9	28442922	28443322	0.000297918097525	Opcml
chr1	5		629					773	
chr1	4602791	46028310	0.000789476347926	Vegfa	chr9	29259927	29260327	0.000608932129286	Ntm
chr1	0		588					81	
chr1	4840771	48408110	0.000446125521153	Nfya	chr9	29668050	29668450	0.000447717355569	Ntm
chr1	0		64					921	
chr1	5114341	51143811	0.000566365297317	Tbc1d5	chr9	30632955	30633355	0.000276900583778	Snx19
chr1	1		35					161	
chr1	5126321	51263612	7,94E+09	Tbc1d5	chr9	31174313	31174713	0.000540036677644	Aplp2
chr1	2							24	
chr1	5127623	51276631	0.000155064794671	Tbc1d5	chr9	34496191	34496591	0.000664001746857	Kirrel3
chr1	1		533					99	
chr1	5545806	55458465	0.000235660513757	St6gal2	chr9	36935294	36935694	0.000560812023290	Pknox2
chr1	5		23					28	
chr1	5666764	56668045	0.000101073397142	Ranbp3	chr9	36969596	36969996	0.000756047813573	Pknox2
chr1	5		049					968	
chr1	6272840	62728808	0.000658306505832	Efn5	chr9	37400174	37400574	0.000338839476403	Robo4
chr1	8		984					23	
chr1	6290580	62906206	0.000918437301586	Efn5	chr9	37568061	37568461	8,33E+09	Siae
chr1	6		028						
chr1	6333394	63334347	0.000812462513645	4930405O22	chr9	38328873	38329273	0.000431153436731	Or8c9
chr1	7		862	Rik				835	
chr1	6408338	64083784	0.000174602537456	Fer	chr9	40196163	40196563	0.000378325539047	Zfp202
chr1	4		27					486	
chr1	6489736	64897765	0.000650566300977	4930583I09Ri	chr9	40451082	40451482	0.000687979812863	Gramd1b
chr1	5		706	k				393	
chr1	6570382	65704228	0.000404424828010	Txndc2	chr9	41331028	41331428	0.000155913743195	Mir100hg
chr1	8		175					235	
chr1	6734730	67347704	0.000759432052690	Ptprm	chr9	41755500	41755900	0.000592772891134	Mir100hg
chr1	4		69					13	
chr1	6738434	67384744	0.000944529881388	Ptprm	chr9	42262185	42262585	0.000525435345682	Sc5d
chr1	4		6					723	
chr1	7052259	70522992	0.000911758469241	Dlgap1	chr9	44375531	44375931	0.000500048939326	Hyou1
chr1	2		963					314	
chr1	7147989	71480293	0.000671240639357	Smchd1	chr9	44491346	44491746	0.000517723641490	Bcl9l
chr1	3		196					164	
chr1	7247782	72478228	0.000566601475583	Alk	chr9	44765482	44765882	0.000871255140135	Arcn1
chr1	8		184					085	
chr1	7380532	73805723	0.000716097751694	Ehd3	chr9	44916775	44917175	0.000162237109392	Atp5mg
chr1	3		456					338	
chr1	7381849	73818894	0.000569853235392	Ehd3	chr9	44966243	44966643	0.000309958175545	Ube4a
chr1	4		603					91	
chr1	7426254	74262947	0.000283576019339	Memo1	chr9	47856801	47857201	0.000607108545091	Nxpe4
chr1	7		815					606	
chr1	7507411	75074511	0.000343457335530	Ltbp1	chr9	48989094	48989494	5,50E+09	Usp28
chr1	1		095						
chr1	7971412	79714523	0.000117277371054	Cyp1b1	chr9	49829391	49829791	0.000552076442487	Ncam1
chr1	3		936					115	
chr1	8096475	80965155	0.000848822148575	Tmem178	chr9	49979137	49979537	0.000167228138456	Ncam1
chr1	5		179					821	
chr1	8394281	83943218	0.000141271337562	8430430B14	chr9	51030058	51030458	0.000485877154325	Sik2
chr1	8		143	Rik				243	
chr1	8449331	84493713	0.000342574575753	Plekhh2	chr9	55699775	55700175	0.000558074711249	Scaper
chr1	3		707					024	
chr1	8536425	85364651	0.000451938252574	Six3	chr9	55845572	55845972	0.000990189924630	Scaper
chr1	1		946					958	
chr1	8870167	88702078	0.000869563492477	Gtf2a1l	chr9	56730079	56730479	0.000488390479925	Lingo1
chr1	8		132					771	
chr1	8876233	88762737	0.000591933394285	Lhcgr	chr9	57759649	57760049	0.000839359459367	Clk3
chr1	7		871					195	
chr1	9092680	90927206	0.000683762746130	4930480K15	chr9	57996639	57997039	0.000200324503239	Cyp11a1
chr1	6		467	Rik				813	
chr1	3275506	3275906	0.000717933780722	Crem	chr9	58063948	58064348	0.000673871906415	Stra6
chr1	8		602					213	
chr1	3383956	3384356	0.000225677631254	Cul2	chr9	58169136	58169536	0.000900507232435	Islr
chr1	8		362					51	
chr1	4668960	4669360	0.000525683831420	Jcad	chr9	58432682	58433082	0.000309740164289	4930461G14
chr1	8		265					751	Rik
chr1	5136739	5137139	0.000595198736994	Svil	chr9	60793503	60793903	0.000205767491560	Uaca
chr1	8		984					982	
chr1	5586614	5587014	0.000138518418057	Zeb1	chr9	62503151	62503551	0.000342204762582	Coro2b
chr1	8		565					51	

chr1 8	7626785	7627185	0.000998800434359 737	Mpp7	chr9	63862941	63863341	0.000980224284038 211	Smad3
chr1 8	8645974	8646374	0.000433107952099 387	Fzd8	chr9	64456511	64456911	0.000868005478237 288	Megf11
chr1 8	1196974	11970141	0.000767221950743 968	Tmem241	chr9	64570818	64571218	0.000277561301198 492	Megf11
chr1 8	1229398	12294382	0.000251581736414 375	Ankrd29	chr9	65143306	65143706	0.000417967201606 473	Igdcc3
chr1 8	1382362	13824029	0.000526272377121 773	Zfp521	chr9	66355669	66356069	0.000369973388296 03	Herc1
chr1 8	1382643	13826839	0.000308206363520 691	Zfp521	chr9	66733514	66733914	0.000882405712715 917	Car12
chr1 8	1395319	13953594	0.000970351718937 001	Zfp521	chr9	67869676	67870076	0.000781463516989 378	Vps13c
chr1 8	1396592	13966322	0.000103475951783 505	Zfp521	chr9	68132934	68133334	0.000239076678747 008	Gm20745
chr1 8	1540581	15406215	0.000775044108350 276	Aqp4	chr9	68768127	68768527	0.000797881815158 83	Rora
chr1 8	1576022	15760625	0.000773994858189 964	Chst9	chr9	69224195	69224595	0.000313578236568 19	Rora
chr1 8	2077773	20778139	0.000385982549305 108	B4galt6	chr9	69323990	69324390	0.000286820405294 597	Rora
chr1 8	2383663	23837031	0.000271725889384 955	Mapre2	chr9	69987637	69988037	0.000552409968236 134	Bnip2
chr1 8	2385883	23859237	0.000615087051967	Mapre2	chr9	71054570	71054970	0.000574849052618 065	Lipc
chr1 8	2402579	24026192	0.000586197888196 915	Zfp24	chr9	71957291	71957691	0.000704750381911 978	Tcf12
chr1 8	2407906	24079465	0.000607662951493 504	Ino80c	chr9	73967148	73967548	8,63E+09	Unc13c
chr1 8	2503188	25032289	0.000411213016702 883	Fhod3	chr9	74197912	74198312	0.000553571907505 379	Wdr72
chr1 8	3006541	30065817	0.000738282309385 33	Pik3c3	chr9	74326213	74326613	0.000696298504256 01	Wdr72
chr1 8	3077689	30777294	0.000755166195211 21	Rit2	chr9	74698996	74699396	0.000207209544033 082	Gm16551
chr1 8	3175796	31758368	0.000639747344772 79	Ammecr1l	chr9	75979592	75979992	0.000366362771671 014	Hmgcll1
chr1 8	3215282	32153221	0.000297219475660 304	Map3k2	chr9	76993810	76994210	0.000384794930431 42	Tinag
chr1 8	3315134	33151746	0.000553369042080 032	Stard4	chr9	77055966	77056366	0.000601369125894 369	Tinag
chr1 8	3334952	33349922	0.000142462330159 132	Nrep	chr9	80083318	80083718	0.000400826478345 495	Senp6
chr1 8	3492205	34922453	0.000714040406391 891	Etf1	chr9	83540743	83541143	0.000475770741563 242	Sh3bgrl2
chr1 8	3554451	35544915	0.000740555610839 3	Snhg4	chr9	83608149	83608549	0.000429557176110 151	Sh3bgrl2
chr1 8	3562044	35620841	0.000568722398069 71	Slc23a1	chr9	84627055	84627455	0.000549524687483 149	4930554C24 Rik
chr1 8	3564035	35640759	0.000645920865140 592	Mzb1	chr9	87558997	87559397	0.000274067181348 592	Tbx18
chr1 8	3588720	35887600	0.000793171391985 092	Cxxc5	chr9	88124136	88124536	6,28E+09	Nt5e
chr1 8	3600432	36004727	0.000562686675410 484	Psd2	chr9	88376929	88377329	0.000645960222665 831	Nt5e
chr1 8	3608902	36089425	0.000108334341465 399	4930471G03 Rik	chr9	91254312	91254712	0.000650845807735 001	Zic4
chr1 8	3609434	36094742	0.000117965432776 312	4930471G03 Rik	chr9	91382536	91382936	0.000842978636100 611	Zic4
chr1 8	3619979	36200190	0.000601186025049 216	Nrg2	chr9	91412353	91412753	0.000651666384870 575	Zic4
chr1 8	3628365	36284058	0.000700691254524 655	Pura	chr9	92177871	92178271	5,94E+09	Plscr5
chr1 8	3713148	37131888	0.000344081350071 246	Pcdhac2	chr9	93545675	93546075	0.000849804839923 252	1700034K08R ik
chr1 8	3793381	37934215	0.000167513346317 705	Diaph1	chr9	94740316	94740716	0.000391983174742 76	Mir7656
chr1 8	3836706	38367460	0.000996306058125 507	Gnpda1	chr9	95159476	95159876	0.000778397311068 879	Slc9a9
chr1 8	3837200	38372406	0.000975293556566 337	Gnpda1	chr9	99535273	99535673	0.000180517994950 113	Armc8
chr1 8	3850796	38508362	0.000676808221636 849	Ndfip1	chr9	99783173	99783573	0.000744035693434 508	Cldn18
chr1 8	3853439	38534790	0.000789899389158 855	Spry4	chr9	10128723	10128763	0.000162083433456 583	Ppp2r3a
chr1 8	3891100	38911400	0.000916240911984 056	Fgf1	chr9	10366131	10366171	0.00098000910625 334	Tmem108
chr1 8	3905793	39058333	0.000285773791352 639	Arhgap26	chr9	10423563	10423603	0.000840682525150 127	Dnajc13

chr1	4268564	42686044	0.000550378808850	Gpr151	chr9	10677950	10677990	0.000582334843106	Rad54l2
chr1	4		338		chr9	8	8	658	
chr1	4337937	43379778	0.000858715410681	Dpysl3	chr9	10811121	10811161	0.000976421099984	Bsn
chr1	8		61		chr9	7	7	526	
chr1	4346300	43463404	0.000905840287230	Dpysl3	chr9	11135213	11135253	0.000388705401442	Trank1
chr1	4		841		chr9	5	5	646	
chr1	4431097	44311371	0.000470051761076	Myot	chr9	11317163	11317203	0.000263601508425	AU023762
chr1	8		062		chr9	0	0	859	
chr1	4771770	47718103	0.000509122997559	Gm4146	chr9	11373764	11373804	0.000323972228776	Clasp2
chr1	8		119		chr9	9	9	004	
chr1	5005557	50055974	0.000177889464815	C030005K06	chr9	11439602	11439642	0.000245004013275	Glb1
chr1	8		228	Rik	chr9	1	1	54	
chr1	5156965	51570057	0.000438902660553	Prr16	chr9	11660970	11661010	0.000422735852277	Rbms3
chr1	8		131		chr9	1	1	478	
chr1	5283946	52839863	0.000576830918049	Sncaip	chr9	11666499	11666539	0.000694806965828	Rbms3
chr1	8		165		chr9	1	1	132	
chr1	5293202	52932422	0.000464572599771	Sncaip	chr9	11736206	11736246	0.000361708590118	Rbms3
chr1	8		84		chr9	1	1	415	
chr1	5412350	54123908	0.000959194071263	Csnk1g3	chr9	12049629	12049669	0.000361357861145	Eif1b
chr1	8		658		chr9	5	5	661	
chr1	5459875	54599157	0.000746457694920	9330117O12	chr9	12075176	12075216	8,55E+09	1700020M21
chr1	8		988	Rik	chr9	0	0		Rik
chr1	5584479	55845199	0.000763841358785	Gramd2b	chr9	12077625	12077665	0.000103403179888	1700020M21
chr1	8		378		chr9	9	9	328	Rik
chr1	5683938	56839783	0.000627692622487	Marchf3	chr9	12222937	12222977	0.000207187467015	Ano10
chr1	8		075		chr9	4	4	979	
chr1	5888986	58890262	0.000304449912602	Adamts19	chr9	12325913	12325953	0.000214902463012	Tmem158
chr1	8		367		chr9	7	7	923	
chr1	6088559	60885993	0.000110243793722	Arsi	chr9	12348089	12348129	0.000518219356564	Limd1
chr1	8		375		chr9	5	5	709	
chr1	6468568	64686088	0.000877828505833	Atp8b1	chr9	12348186	12348226	0.000118342981266	Limd1
chr1	8		067		chr9	4	4	784	
chr1	6593645	65936852	0.000286446289840	Rax	chrX	9695736	9696136	0.000790055648514	Dynlt3
chr1	8		647		chrX	31			
chr1	6601413	66014532	0.000601340020894	Lman1	chrX	10063800	10064200	0.000175658899591	Srpx
chr1	8		847		chrX			813	
chr1	6623242	66232825	0.000816039532584	Mir694	chrX	11577211	11577611	6,42E+09	H2a1j
chr1	8		911		chrX				
chr1	6852653	68526939	9,93E+09	Mc2r	chrX	11578289	11578689	0.000689509424629	H2a1j
chr1	8				chrX			975	
chr1	6882270	68823108	6,69E+09	Mir6356	chrX	12838878	12839278	0.000209349583811	Gm14634
chr1	8				chrX			802	
chr1	6920015	69200552	0.000590005263572	Tcf4	chrX	14671634	14672034	0.000695298301228	Ppp1r2-ps9
chr1	8		779		chrX			942	
chr1	7183346	71833867	0.000784158671472	Dcc	chrX	20161326	20161726	0.000426283430410	Slc9a7
chr1	8		746		chrX			956	
chr1	7359608	73596484	0.000496681964965	Mex3c	chrX	48512361	48512761	0.000736254541039	Aifm1
chr1	8		662		chrX			649	
chr1	7406660	74067009	1,50E+09	Mapk4	chrX	51360919	51361319	0.000906881300724	Mbnl3
chr1	8				chrX			915	
chr1	7413024	74130640	0.000326588174788	Mapk4	chrX	52460827	52461227	0.000968886848987	Mir717
chr1	8		256		chrX			216	
chr1	7546307	75463473	1,98E+09	Gm10532	chrX	52819836	52820236	0.000262592313387	Ccdc160
chr1	8				chrX			211	
chr1	7603988	76040280	0.000111467786199	Zbtb7c	chrX	53210265	53210665	0.000818261345833	Plac1
chr1	8		526		chrX			626	
chr1	7801502	78015421	0.000430705362288	Epg5	chrX	56478199	56478599	0.000218543103553	Ints6l
chr1	8		73		chrX			293	
chr1	7848156	78481961	6,69E+09	Slc14a2	chrX	56639935	56640335	0.000799242376888	Slc9a6
chr1	8				chrX			332	
chr1	8059926	80599667	0.000158391917714	Nfatc1	chrX	56707769	56708169	0.000794797284273	Fhl1
chr1	8		162		chrX			008	
chr1	8065393	80654337	0.000249513880412	Nfatc1	chrX	69413303	69413703	6,09E+09	Aff2
chr1	8		611		chrX				
chr1	8128348	81283887	0.000976100621193	Sall3	chrX	76010700	76011100	0.000682044354160	Pls3
chr1	8		243		chrX			638	
chr1	8183939	81839792	0.000592526667909	Mir5127	chrX	77049660	77050060	0.000766890471361	Cldn34d
chr1	8		21		chrX			137	
chr1	8252777	82528178	0.000719128755595	Mbp	chrX	84937392	84937792	0.000656863677533	Dmd
chr1	8		433		chrX			925	
chr1	8256418	82564584	0.000936849303547	Mbp	chrX	92223496	92223896	0.000371231870455	Mageb1
chr1	8		094		chrX			18	
chr1	8334061	83341016	0.000535596574377	4921531P14	chrX	93049574	93049974	0.000720803543437	BC061195
chr1	8		184	Rik	chrX			62	
chr1	8512554	85125943	0.000866744660976	Fbxo15	chrX	93891670	93892070	0.000894819942734	Pdk3
chr1	8		915		chrX			104	
chr1	8515463	85155030	0.000947631888445	Fbxo15	chrX	94670662	94671062	0.000870117751436	Gspt2
chr1	8		599		chrX			306	
chr1	8635349	86353891	0.000781205966332	Neto1	chrX	95195010	95195410	0.000187737986889	Arhgef9
chr1	8		132		chrX			536	

chr1 8	8892267 8	88923078	0.000385224190200 706	Socs6	chrX 95457246	95457646	0.000124270844720 22	Amer1
chr1 9	3544848	3545248	0.000103906487328 498	Ppp6r3	chrX 10081486	10081526	0.000282503659590 816	Dlg3
chr1 9	4963080	4963480	0.000895555073831 211	Mrpl11	chrX 10691802	10691842	0.000257663600804 703	Lpar4
chr1 9	4991594	4991994	0.000541796506502 656	Npas4	chrX 13157194	13157234	0.000717108742870 121	Diaph2
chr1 9	5705836	5706236	0.000825906256335 829	Kcnk7	chrX 13352010	13352050	0.000759707155773 67	Pcdh19
chr1 9	1071455 4	10714954	0.000671358446059 145	Cd5	chrX 13369275	13369315	0.000476944112219 213	Pcdh19
chr1 9	1152745 6	11527856	0.000751766589780 489	Ms4a6b	chrX 13847910	13847950	0.000562165440163 298	Tex13a
chr1 9	1247814 9	12478549	0.000924442007870 896	Mpeg1	chrX 13886687	13886727	0.000288989242527 071	Nrk
chr1 9	1403363 7	14034037	0.000472614429989 582	Or9i1b	chrX 14098284	14098324	0.000414852819802 913	Atg4a
chr1 9	1603413 2	16034532	0.000242870947724 683	C130060C02 Rik	chrX 14185419	14185459	0.000683620984144 095	Irs4
chr1 9	1727009 5	17270495	0.000698331353767 845	Gcnt1	chrX 14382251	14382291	0.000802693700712 398	Capn6
chr1 9	1757788 5	17578285	0.000578107815088 602	Rfk	chrX 14386625	14386665	0.000963612635638 489	Capn6
chr1 9	1928860 2	19289002	0.000427385998356 122	Rorb	chrX 14393331	14393371	0.000319148987072 415	Dcx
chr1 9	2081378 3	20814183	0.000899726416817 792	Tmc1	chrX 15043895	15043935	0.000686995864299 655	Tmem29
chr1 9	2209654 8	22096948	0.000228291938128 899	Trpm3	chrX 15232762	15232802	4,52E+09	Kantr
chr1 9	2336072 0	23361120	0.000835678464941 703	Smc5	chrX 15264396	15264436	0.000341358101021 205	Shroom2
chr1 9	2415980 8	24160208	2,33E+09	Tjp2	chrX 15300998	15301038	0.000384029282377 574	Usp51
chr1 9	2417333 0	24173730	0.000621714063843 593	Tjp2	chrX 15727863	15727903	6,65E+09	Phex
chr1 9	2469859 1	24698991	0.000815598915017 59	Tmem252	chrX 16190947	16190987	0.000174436948952 472	Rai2
chr1 9	2633079 5	26331195	0.000104986002360 086	Smarca2	chrX 16288495	16288535	0.000537415279720 177	Syp1
chr1 9	2697752 0	26977920	0.000988770463262 512	Gm815	chrX 16704990	16705030	0.000127002570597 945	Gm8817
chr1 9	2737274 9	27373149	0.000348707104924 406	Pum3	chrX 16990579	16990619	0.000570748032183 488	Mid1
chr1 9	2821433 1	28214731	0.000372669388067 681	Glis3	chrX 16992071	16992111	0.000496821343931 808	Mid1

Supplemental Table 4: Table lists clinical features of WDSTS and prevalence in RARE-X data base and literature [1], [17]. Last column represents whether features are present in *Kmt2a*^{+/*LSL*} mouse model.

Clinical feature	WDSTS patients % (n)	Presence in mouse model
Growth retardation	95.4% (62/65)	present
Brain and/or neurological issues	93.8% (60/64)	present
Eye and/or vision issues	84.7% (50/59)	unknown
Sleep issues	80.3% (53/66)	unknown
Dental and/or oral issues	79.4% (50/63)	unknown
Digestive system issues	76.9% (50/65)	unknown
Behavior and/or psychiatric issues	75.8% (47/62)	present
Hypotonia	60.0% (33/55)	present
Hypertrichosis cubiti	57.0% (57/100)	present
Head, face and/or neck issues	56.3% (36/62)	present
Ear and/or hearing issues	44.3% (27/61)	unknown
Lungs and/or breathing issues	40.0% (26/65)	unknown
Issues with bone, cartilage and/or connective tissue	39.0% (23/59)	present

Hormones and/or hormone-producing glands issues	33.3% (19/57)	unknown
Kidney, bladder and/or genital issues	31.1% (19/61)	unknown
Immune system issues	26.3% (15/57)	unknown
Skin issues	24.6% (15/61)	unknown
Hear and/or blood vessel system issues	23.8% (15/63)	unknown
Blood or bleeding issues	1.6% (1/63)	present

Supplemental table 5: Three-dimensional coordinates for k=31 biologically homologous landmarks that were used to estimate shape of the cranium. (.xlsx)

Midline	
nal	nasale
ids	interdentale superior
pns	posterior nasal spine
nas	nasion
brg	bregma
lam	lambda
int	interparietal
opi	opisthion
pcg	posterior crista galli
sem	sphenoethmoid synchondrosis
pbm	intrasphenoid synchondrosis
bom	sphenooccipital synchondrosis
bas	basion
Bilateral	
rnot,lnot	infraorbital notch
rmjs,lmjs	maxillojugal superior
rjss,ljss	jugosquamosal superior
rfsc,lfsc	frontosphenoid crest
ream,leam	external auditory meatus
rast,last	asterion
rpmm,lpmm	premaxillo-maxillary suture
rptg,lptg	pterygoid hamulus
rfmc,lfmc	condyle at foramen magnum

