

# **Molecular pathogenesis underlying syndromic forms of primary microcephaly**

Dissertation  
for the award of the degree  
“Doctor rerum naturalium”  
of the Georg-August-Universität Göttingen

within the doctoral program Molecular Medicine  
of the Georg-August University School of Science (GAUSS)

submitted by  
M.Sc. Nadine Rosin  
  
born in Hamm (Westf.)

Göttingen 2019

## **Thesis Committee:**

**Prof. Dr. Bernd Wollnik**

Institute of Human Genetics, University Medical Center Göttingen

**Prof. Dr. Wolfram-Hubertus Zimmermann**

Institute of Pharmacology and Toxicology, University Medical Center Göttingen

**Prof. Dr. Andre Fischer**

Dep. for Psychiatry and Psychotherapy, German Center for Neurodegenerative Diseases (DZNE) Göttingen

Former member:

**Prof. Dr. Steven Johnson**

Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, USA

## **Members of the Examination Board**

Reviewer

**Prof. Dr. Bernd Wollnik**

Institute of Human Genetics, University Medical Center Göttingen

2<sup>nd</sup> Reviewer

**Prof. Dr. Wolfram-Hubertus Zimmermann**

Institute of Pharmacology and Toxicology, University Medical Center Göttingen

Further members of the Examination Board:

**Prof. Dr. Andre Fischer**

Dep. For Psychiatry and Psychotherapy, German Center for Neurodegenerative Diseases (DZNE) Göttingen

**Prof. Dr. Ralf Dressel**

Institute for Cellular and Molecular Immunology, University Medical Center Göttingen

**Prof. Dr. Michael Sereda**

Molecular and Translational Neurology, Max Planck Institute of Experimental Medicine, Göttingen

**Prof. Dr. Martin Göpfert**

Department of Cellular Neurobiology, University of Göttingen

**Date of Disputation:** 19<sup>th</sup> December 2019

## Table of contents

<b>Abbreviations</b> .....	<b>I</b>
<b>Summary</b> .....	<b>VII</b>
<b>Zusammenfassung</b> .....	<b>IX</b>
<b>1 Introduction</b> .....	<b>1</b>
<b>1.1 The mammalian neocortex</b> .....	<b>1</b>
1.1.1 Developmental stages and characteristics of murine neocortex .....	2
1.1.2 Signaling pathways involved in cortical neurogenesis .....	3
<b>1.2 Isolated and syndromic forms of human microcephaly</b> .....	<b>5</b>
1.2.1 Autosomal recessive and autosomal dominant inheritance .....	5
1.2.2 Model systems for functional analysis of microcephaly.....	6
<b>1.3 Epigenetic modifications in stem cells and during cortigogenesis</b> .....	<b>8</b>
<b>1.4 The KMT2 family – The TrxG histone modifier for H3K4me1/me2/me3</b> .....	<b>9</b>
1.4.1 The KMT2 family – Importance for embryonic development and diseases .....	10
1.4.2 KMT2B - The unique KMT2 family member.....	12
<b>2 Study aims</b> .....	<b>15</b>
<b>3 Material &amp; Methods</b> .....	<b>16</b>
<b>3.1 Material</b> .....	<b>16</b>
3.1.1 Subjects .....	16
3.1.2 Enzymes .....	16
3.1.3 Synthetic oligonucleotides .....	16
3.1.4 Bacteria and plasmids .....	17
3.1.5 Antibodies .....	17
3.1.6 Solutions and buffers .....	18
3.1.7 Kits.....	20
3.1.8 Reference sequences .....	20
3.1.9 Chemicals and expendable items.....	21
3.1.10 Non-expandable equipment .....	21
3.1.11 Online resources and software .....	22
<b>3.2 Methods</b> .....	<b>24</b>
3.2.1 Analysis of nucleic acids .....	24
3.2.1.1 Polymerase Chain Reaction (PCR) .....	24

## Table of contents

3.2.1.2	Quantitative PCR (qPCR) .....	24
3.2.1.3	Reverse Transcription PCR (RT-PCR) .....	25
3.2.1.4	Gel electrophoresis .....	25
3.2.1.5	Purification and Sanger sequencing .....	25
3.2.1.6	DNA extraction .....	25
3.2.1.7	RNA extraction and cDNA synthesis .....	25
3.2.1.8	NGS-based methods.....	26
3.2.1.8.1	Whole exome sequencing (WES).....	26
3.2.1.8.2	Deep amplicon sequencing .....	26
3.2.1.8.3	RNA sequencing (RNAseq) by Poly(A) selection .....	26
3.2.1.8.4	Single cell RNA sequencing (scRNAseq).....	26
3.2.1.9	Cloning of KMT2B .....	27
3.2.1.9.1	Cloning of mutant KMT2B harboring a C-terminal GFP-tag in peGFP-N1 .....	27
3.2.1.9.2	TOPO <sup>®</sup> cloning and pJET1.2 cloning.....	27
3.2.1.9.3	Different KMT2B constructs with N terminal Flag-HA tag in pCS2+ .....	27
3.2.1.9.4	Colony PCR .....	28
3.2.1.9.5	Digestion with restriction enzymes .....	29
3.2.1.9.6	Ligation and Transformation .....	29
3.2.2	Cell culture .....	29
3.2.2.1	Cultivation of fibroblasts from a skin biopsy of index patient K3027 .....	29
3.2.2.2	Reprogramming dermal fibroblasts to human induced pluripotent stem cells (iPSC) 30	
3.2.2.3	Differentiation of iPSC into neural progenitor cells (NPCs) .....	30
3.2.2.4	Passaging cells .....	31
3.2.2.4.1	iPSC .....	31
3.2.2.4.2	Primary fibroblasts, HEK293T and COS-7 cells .....	31
3.2.2.5	Freezing and thawing cells.....	31
3.2.2.5.1	iPSC.....	31
3.2.2.5.2	Primary fibroblasts, HEK293T and COS-7 cells .....	31
3.2.2.6	Cell transfection .....	31
3.2.2.7	Protein extraction .....	31
3.2.2.7.1	Total cell lysate .....	31
3.2.2.7.2	Cell fractionation .....	32
3.2.2.7.3	Measurement of protein concentration .....	32
3.2.2.8	Mass spectrometry.....	32
3.2.2.9	Fluorescence activated cell sorting (FACS).....	32
3.2.3	Immunological approaches .....	32
3.2.3.1	Western Blot.....	32
3.2.3.1.1	Western Blot of the KMT2B protein .....	33
3.2.3.2	Immunocytochemistry (ICC).....	33
3.2.3.3	Immunohistochemistry (IHC).....	34

## Table of contents

3.2.4	The zebrafish <i>Danio rerio</i> – <i>in vivo</i> analysis .....	34
3.2.4.1	Maintenance of zebrafish .....	34
3.2.4.2	Genotyping zebrafish .....	34
3.2.4.3	CRISPR/Cas9-based gene editing in zebrafish .....	35
3.2.4.3.1	Establishment of positive control <i>tyrosinase</i> .....	35
3.2.4.3.2	Injection of CRISPR/Cas9 solutions .....	35
3.2.4.3.3	Generation of stable <i>kmt2ba</i> and <i>kmt2bb</i> knock out lines.....	36
3.2.4.4	Measurement of head area .....	36
3.2.4.5	Hematoxylin and eosin (HE) stain.....	36
3.2.4.6	Whole-mount immunofluorescence (WMI) for activated Caspase 3.....	36
3.2.5	Mus musculus – <i>in vivo</i> analysis .....	37
3.2.5.1	Mouse strains .....	37
3.2.5.2	Genotyping mice .....	37
3.2.5.3	Cryosections .....	37
3.2.6	Statistics .....	37
<b>4</b>	<b>Results .....</b>	<b>38</b>
<b>4.1</b>	<b>Analyses of whole-exome sequencing (WES) data .....</b>	<b>38</b>
4.1.1	<i>DLG3</i> - A novel microcephaly-associated candidate gene.....	38
4.1.1.1	Clinical characterization of index patient K3627 .....	38
4.1.1.2	Somatic mosaicism of <i>DLG3</i> in a family with a syndromic form of microcephaly ...	39
4.1.2	Identification of novel microcephaly causing mutations .....	42
4.1.2.1	<i>DDX11</i> - Novel nonsense mutations underlying syndromic microcephaly .....	42
4.1.2.2	<i>FANCI</i> – Novel mutations affecting splicing in a patient with Fanconi anemia.....	44
<b>4.2</b>	<b>Functional characterization of <i>KMT2B</i>, a novel gene associated with syndromic microcephaly.....</b>	<b>46</b>
4.2.1	Clinical characterization of index patient K3027 .....	46
4.2.2	<i>KMT2B</i> – Identification of a novel candidate gene.....	46
4.2.3	Ectopically expressed <i>KMT2B</i> shows altered cellular localization .....	47
4.2.4	Investigation of patient-derived primary dermal fibroblasts .....	49
4.2.5	Reprogramming primary fibroblasts to iPSC.....	51
4.2.6	Investigation of endogenous <i>KMT2B</i> in patient-derived fibroblasts and iPSC... ..	52
4.2.7	Analyses of RNAseq data in iPSC.....	55
4.2.8	Differentiation of iPSC into neural progenitor cells (NPCs) .....	59
4.2.9	CRISPR/Cas9-mediated Knock Out of <i>kmt2ba</i> and <i>kmt2bb</i> in zebrafish .....	63
4.2.10	Detection of microcephaly in a <i>kmt2bb</i> knock out model .....	64
4.2.11	Investigation of apoptosis in the <i>kmt2bb</i> <sup>-/-</sup> zebrafish model.....	67
4.2.12	Histological investigation of the <i>Emx1-Cre</i> <sup>pos/+</sup> ; <i>Kmt2b</i> <sup>fl/+</sup> mouse model .....	68
4.2.13	scRNAseq analysis in the heterozygous <i>Kmt2b</i> cKO model.....	70

## Table of contents

<b>5</b>	<b>Discussion .....</b>	<b>75</b>
<b>5.1</b>	<b>Novel insights into the pathogenesis of syndromic microcephaly.....</b>	<b>75</b>
5.1.1	The power of NGS .....	75
5.1.1.1	Identification of a low-level somatic mosaicism .....	75
5.1.1.2	Expanding the mutational spectrum of <i>DDX11</i> .....	77
5.1.1.3	Identification of novel <i>FANCI</i> mutations affecting splicing .....	78
<b>5.2</b>	<b>The pathomechanism underlying <i>KMT2B</i>-associated microcephaly .....</b>	<b>81</b>
5.2.1	The methyltransferase KMT2B and its importance for stem cells .....	81
5.2.1.1	Linking WNT-signaling to the function of KMT2B during stem cell differentiation ..	83
5.2.1.2	The role of apoptosis in patient-derived cells.....	84
5.2.2	Haploinsufficiency vs. dominant negative effect .....	86
5.2.2.1	p.Leu1822Thrfs*12 KMT2B is not detectable <i>in vitro</i> .....	86
5.2.2.2	<i>In vivo</i> results favor haploinsufficiency as disease underlying effect.....	87
5.2.3	Early corticogenesis in the heterozygous <i>Kmt2b</i> cKO mouse.....	88
5.2.3.1	The strength of scRNAseq over RNAseq .....	88
5.2.3.2	scRNAseq - Identification of mild differences between proliferating and non-proliferating RGCs .....	88
5.2.3.3	Future perspectives - Detailed characterization of the mild phenotype during corticogenesis .....	90
5.2.3.4	Analyzing heterozygous <i>Kmt2b</i> cKO embryos.....	91
5.2.4	Characterization of KMT2B protein and RNA .....	92
5.2.4.1	Putative SET domain independent functions of the KMT2B protein.....	92
5.2.4.2	The zebrafish – Model to study SET domain dependency for microcephaly development .....	95
5.2.4.3	A lncRNA transcribed from the <i>KMT2B</i> locus .....	96
5.2.5	Future perspectives – Investigating <i>KMT2B</i> -associated phenotypic variations in the zebrafish model .....	97
<b>6</b>	<b>Conclusion.....</b>	<b>101</b>
<b>7</b>	<b>Supplementary.....</b>	<b>102</b>
<b>8</b>	<b>References.....</b>	<b>114</b>
<b>9</b>	<b>List of Figures.....</b>	<b>134</b>
<b>10</b>	<b>List of Tables .....</b>	<b>135</b>
<b>11</b>	<b>Appendix .....</b>	<b>136</b>
	<b>Acknowledgements.....</b>	<b>136</b>
	<b>Affidavit.....</b>	<b>138</b>

## Abbreviations

### Abbreviations

A	Adenine
aa	amino acid
Ala	Alanine
AO	Acridine orange
APC2	Adenomatosis Polyposis Coli 2
Arg	Arginine
ARM	armadillo
Array-CGH	Array-based Comparative Genomic Hybridization
Asn	Asparagine
Asp	Aspartic acid
ASPM	Abnormal Spindle Microtubule Assembly
β	beta
BAF	BRG1-Associated Factor
BCA	Bicinchoninic acid assay
BCL2L1	B-cell lymphoma 2 like 1
BMP	Bone morphogenetic protein
C	Cytosine
Ca <sup>2+</sup>	Calcium
Cas9	CRISPR-associated protein-9 nuclease
CCND1	cyclin D1
CTNNB1	Catenin, beta-1
cDNA	complementary DNA
CCG	Cologne Center for Genomics
CDK2	Cyclin-dependent kinase-2
CIAP	Calf-intestinal alkaline phosphatase
cKO	conditional KO
CLB	Cell lysis buffer
CLSM	confocal laser scanning microscope
CNV	copy-number variation
COL1A1	Collagen, type I, alpha 1
COMPASS	Complex of proteins associated with Set1
COS-7	CV-1 in Origin Simian-7
CP	cortical plate
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDX11	DEAD/H-Box Helicase 11
DDX11L	DEAD/H-Box Helicase 11 like
DE	differentially expressed
DEB	Diepoxybutane
del	Deletion
DEPC	Diethyl pyrocarbonate
DGVDD	Aspartic Acid, Glycine, Valine, Aspartic Acid, Aspartic Acid
dH <sub>2</sub> O	demineralised water
DLG3	Disks Large Homolog 3
Dlx2	Distal-Less Homeobox 2

## Abbreviations

DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxid
DNA	deoxyribonucleic acid
DNE	dominant negative effect
DNMT	DNA methyltransferase
dNTP	deoxynucleotide
dpf	Days post fertilization
dSet1	drosophila Set1
E	Embryonic day
EB	embryoid body
EC	endothelial cells
E.coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EJC	exon junction complex
Emx1	Empty Spiracles Homeobox 1
EN	excitatory neurons
NEU	N-ethyl-N-nitrosourea
ESC	embryonic stem cell
ESE	exonic splice enhancer
EtOH	Ethanol
et al.	et alii (and others)
Exo	Exonuclease
Ezh2	Enhancer of zeste homolog 2
F / fwd	forward
FACS	Fluorescence activated cell sorting
FANCI	Fanconi anemia, complementation group I
FBS	fetal bovine serum
FCS	fetal calf serum
FDR	false discovery rate
FISH	Fluorescence in situ Hybridization
fl	floxed
FOSL1	Fos-related antigen 1
h	hour
HeLa	Henrietta Lacks
hpf	Hours post fertilization
Foxg1	Forkhead Box G1
fs	frameshift
FYRC	Phenylalanine Tyrosine-rich domain C-terminal
FYRN	Phenylalanine Tyrosine -rich domain N-terminal
Fz	Frizzled
G	Guanine
γ	gamma
GFP	green fluorescent protein
Gln	Glutamine
Gly	Glycine
GO	gene ontology
HA	hemagglutinin
HE	Hematoxylin and eosin
HEK293T	human embryonic kidney 293 cells



## Abbreviations

Hes1	Hes Family BHLH Transcription Factor 1
Hes5	Hes Family BHLH Transcription Factor 5
HGMD	Human Gene Mutation Database
HOX	Homeobox
HOXA5	Homeobox A5
HOXB7	Homeobox B7
HRP	horseradish peroxidase
HSC70	Heat shock cognate 71 kDa protein
H3K4me1	monomethylation of lysine 4 on histone 3
H3K4me2	dimethylation of lysine 4 on histone 3
H3K4me3	trimethylation of lysine 4 on histone 3
H3K9me3	trimethylation of lysine 9 on histone 3
H3K27me3	trimethylation of lysine 27 on histone 3
HUWE1	HECT, UBA and WWE Domain containing E3 Ubiquitin protein ligase 1
ICC	Immunocytochemistry
Igfbp2	Insulin-like growth factor-binding protein 2
Igfbp5	Insulin-like growth factor-binding protein 5
IgG	Immunoglobulin G
IHC	Immunohistochemistry
Ile	Isoleucine
I/M	immune cells / macrophages
IN	inhibitory interneurons
iN	induced neuronal cell
ins	Insertion
IPC	intermediate progenitor cell
ips	induced pluripotent stem cell
Isl1	ISL LIM Homeobox 1
IZ	intermediate zone
kb	kilobase
KCl	Potassium chloride
KD	knock down
kDa	kilo Dalton
Kdm6b	Lysine demethylase 6B
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIF2A	Kinesin Family Member 2A
KIF5C	Kinesin Family Member 5C
KMT2	lysine methyltransferase subclass 2
KMT2B	lysine methyltransferase 2B
KO	Knock out
LB	lysogeny broth
Leu	Leucine
LIG4	DNA Ligase 4
lncRNAs	long noncoding RNAs
M	Molarity
Magoh	Mago Homolog
Magohb	Mago Homolog b
Mb	megabase
MCPH	autosomal recessive primary microcephaly
MEF	murine embryonic fibroblast

## Abbreviations

min	Minute
MOCA	mouse organogenesis cell atlas
mRNA	messenger RNA
MZ	marginal zone
N	Normality
n.a.	not applied
NaCl	Sodium chloride
NB	Nuclei buffer
n.d.	not detected
NEC	Neuroepithelial cell
NeuN	neuronal nuclei
NFATC1	Nuclear factor of activated T-cells, cytoplasmic 1
Ngn1	Neurogenin 1
NGS	next-generation sequencing
NLS	nuclear localization signal
NMD	nonsense-mediated mRNA decay
NPC	neural progenitor cell
NPTX1	Neuronal Pentraxin 1
NSB	Nuclei suspension buffer
NSC	Neural stem cell
Oct-3/4	octamer-binding transcription factor 3/4
ORA	over-representation analysis
Otx2	Orthodenticle Homeobox 2
P	Passage
p	Percentile
Pat	Patient
Pax6	Paired Box 6
PBS	phosphate-buffered saline
Pbx3	PBX Homeobox 3
PcG	Polycomb group
PCR	Polymerase Chain Reaction
Pen	Penicillin
PFA	polymeric formaldehyde
PHD2	second PHD finger
PHD3	third PHD finger
PHD4	fourth PHD finger
pH3	phosphorylated H3
PP	preplate
PRM	parallel reaction monitoring technique
PDVF	polyvinylidene difluoride
PI	propidium iodide
pos	positive
qPCR	Quantitative PCR
R / rev	reverse
Rbm8a	RNA Binding Motif Protein 8A
RGC	radial glial cell
Ring1b	Ring Finger Protein 1b
RNA	ribonucleic acid
RNAseq	RNA sequencing

## Abbreviations

ROCK	Rho associated protein kinase
RT	room temperature
scRNAseq	Single cell RNA sequencing
SCU	stem cell unit
SDS	sodium dodecyl sulfate
sgRNA	single-guide RNA
SP	subplate
RT-PCR	Reverse Transcription PCR
Sap	Shrimp Alkaline Phosphatase
SD	Standard deviation
SDS-Page	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
sgRNA	single guide RNA
Shh	Sonic hedgehog
SNP	short neural precursor
SNV	single nucleotide variant
Sox1	SRY-box transcription factor 1
Sox2	SRY-box transcription factor 2
Srg3	Swi3-related gene
Strep	Streptomycin
SVZ	subventricular zone
T	Thymine
TAD	transactivation domain
TAL	transcriptome and genome analysis laboratory
TBE	tris-borate-EDTA
TBS	Tris-buffered saline
TBS-T	TBS-Tween®
TCF/LEF	T-cell factor/lymphoid enhancer-binding factor
Thr	Threonine
TLF	Tübingen long fin
TMEM88	Transmembrane Protein 88
Top2a	DNA Topoisomerase II Alpha
TOUCH58	Touchdown-like 58 PCR
Trp	Tryptophan
TRP53	Transformation-related protein 53
Trr	Trithorax-related
Trx	Thritorax
TrxG	Trithorax group
TSC	Total Spectrum Count
TUBB2B	tubulin beta 2B class IIb
tyr	tyrosinase
UMG	University Medical Center Göttingen
UMI	unique molecular identifiers
VAF	variant allele frequencies
Val	Valine
vs.	versus
VZ	ventricular zone
WB	Western Blot
WDR82	WD Repeat Domain 82

## Abbreviations

WDFY3	WD Repeat And FYVE Domain Containing 3
WES	whole-exome sequencing
WIF1	WNT Inhibitory Factor 1
WGS	whole-genome sequencing
WMI	Whole-mount immunofluorescence
WNT	Wingless-related integration site
WRAD	WDR5, RBBP5, ASH2L and DPY30
wt / WT	wild-type
XRCC2	X-Ray Repair Cross Complementing 2
XRCC4	X-Ray Repair Cross Complementing 4
Zfhx3	Zinc Finger Homeobox 3
+	wild-type allele
*	Stopp

## Summary

Microcephaly describes a reduced brain size in patients that manifests in an evidently reduced head circumference compared to healthy individuals of same age, sex and ethnicity. Even though microcephaly is still considered a rare clinical feature in Europe, hundreds of syndromes manifest with microcephaly indicating that the identification and characterization of microcephaly-associated genes can help to shed light on cellular processes important for brain development, but also beyond.

Since a substantial number of microcephaly patients remain without a genetic diagnosis, I investigated in the first part of my thesis whole-exome sequencing (WES) data of three patients from non-consanguineous families. In the first family of Pakistani descent, different family members were presenting with varying severities of syndromic microcephaly pointing towards an autosomal dominant or X-linked mode of inheritance. Using a combination of WES and deep amplicon sequencing, I identified the c.848delA mutation in the novel and highly convincing microcephaly-associated candidate gene *DLG3* and confirmed a low-level somatic mosaicism in one affected family member. In the second family, I applied WES and identified two novel nonsense mutations (c.442C>T; p.Gln148\* and c.2557C>T; p.Arg879\*) in the *DDX11* gene, associated with Warsaw breakage syndrome. With this, I expanded the spectrum of mutations and phenotypes associated with the *DDX11* gene. In the last family, I identified two novel mutations affecting splicing in the *FANCI* gene (c.2890-2A>G and c.2108A>G; p.Asp703Gly), associated with Fanconi anemia.

In the second part of my thesis, I characterized the novel microcephaly-associated gene *KMT2B* and the previously identified *de novo* c.5462insC (p.Leu1822Thrfs\*12) mutation in more detail. *KMT2B* encodes a H3K4 methyltransferase and exerts essential functions especially during early embryonic developmental stages. Ectopic overexpression of wild-type and p.Leu1822Thrfs\*12 *KMT2B* in COS-7 cells confirmed stability of the mutant protein, albeit the nuclear import was impaired. Further characterization of endogenous RNA and protein levels in patient-derived cells demonstrated a reduction of *KMT2B* RNA and protein. I applied deep amplicon sequencing and verified stability of the c.5462insC transcript. However, I did not detect the p.Leu1822Thrfs\*12 protein, neither by using Western Blot nor by using mass spectrometry, pointing towards protein instability and favoring haploinsufficiency as disease underlying effect. The bulk levels of H3K4me3 were not

## Summary

affected in patient-derived fibroblasts, and cells showed no increased apoptosis activation. Patient-derived fibroblasts were successfully reprogrammed into induced pluripotent stem cells (iPSC), and subsequent RNA sequencing (RNAseq) uncovered differential expression of several WNT-signaling-associated genes. Differentiation of iPSC into early neural progenitor cells (NPC) confirmed a normal course in the expression of pluripotency marker *OCT4* and the NPC markers *SOX1* as well as *PAX6* indicating successful initiation of differentiation. In order to investigate the pathomechanism *in vivo*, I used a CRISPR/Cas9-based strategy and generated a knock out (KO) model of *KMT2B* in the zebrafish *Danio rerio*. Targeting the *KMT2B* homologs *kmt2ba* and *kmt2bb*, I showed that homozygous *kmt2ba* as well as *kmt2bb* KO animals were viable. KO of *kmt2bb*, but not *kmt2ba*, resulted in growth retardation and microcephaly at 14 days post fertilization (dpf), thereby mimicking the phenotype of our patient. Whole-mount immunostainings for activated Caspase 3 illustrated increased activation of apoptosis in affected embryos at 32 hours post fertilization (hpf). In order to characterize early corticogenesis prior to the onset of apoptosis, I investigated a murine conditional heterozygous *Kmt2b* KO model (heterozygous *Kmt2b* cKO) at embryonic day 13.5 (E13.5). Immunohistochemistry did not reveal any differences between control and heterozygous *Kmt2b* cKO embryos. Based on this, only a very mild phenotype was to be expected and the sensitive next-generation sequencing (NGS)-based approach of single-cell RNA sequencing (scRNAseq) was applied for gene expression analysis. Subsequently, analysis of scRNAseq data pointed towards a mild imbalance in the number of proliferating vs. non-proliferating radial glial cells (RGCs) in developing forebrains of heterozygous *Kmt2b* cKO embryos.

In summary, via WES analysis, the data of my PhD thesis end the long odyssey of an unknown diagnosis for three families and provide important information about the underlying genetic cause and open perspectives in the field of therapeutic approaches. The characterization of *KMT2B* and the c.5462insC mutation provides novel insights into the pathomechanism of microcephaly and also into cellular processes beyond.

## Zusammenfassung

Mikrozephalie beschreibt eine verminderte Gehirngröße bei Patienten, welche sich in einem signifikant verringertem Kopfumfang im Vergleich zu gesunden Menschen gleichen Alters, Geschlechts und Ethnizität widerspiegelt. Auch wenn Mikrozephalie in Europa ein seltenes klinisches Merkmal ist, findet man es als Teil des Phänotyps in mehreren Hundert Syndromen. Dies bedeutet, dass die Identifizierung und Charakterisierung Mikrozephalie-assoziiierter Gene neue Einblicke in zelluläre Prozesse der Gehirnentwicklung und weiterer Organe liefern kann.

Da eine genetische Diagnose bei einer beträchtlichen Anzahl an Patienten mit Mikrozephalie immer noch nicht bekannt ist, habe ich im ersten Teil meiner Doktorarbeit *whole-exome* Sequenzierungsdaten (WES) von drei Patienten aus nicht-konsanguinen Familien untersucht. In der ersten Familie pakistanischen Ursprungs zeigten mehrere Familienmitglieder eine unterschiedlich starke Ausprägung an Mikrozephalie, was auf einen autosomal dominanten oder X-Chromosomal gekoppelten Erbgang hindeutete. Mit einer Kombination von WES und *Deep Amplicon Sequencing* konnte ich die c.848delA Mutation in dem neuen Mikrozephalie-Kandidatengen *DLG3* identifizieren und in einem betroffenen Familienmitglied einen somatischen Mosaizismus nachweisen. In der zweiten Familie konnte ich mit WES zwei neue *nonsense* Mutationen (c.442C>T; p.Gln148\* und c.2557C>T; p.Arg879\*) in dem Gen *DDX11* nachweisen, welches bereits mit dem *Warsaw breakage* Syndrom assoziiert ist. In der dritten Familie habe ich schließlich zwei neue Mutationen in dem *FANCI* Gen identifiziert (c.2890-2A>G und c.2108A>G; p.Asp703Gly), die den Prozess des Spleißens beeinflussen.

Im zweiten Teil meiner Arbeit habe ich das neue Mikrozephalie-assoziierte Gen *KMT2B* und die identifizierte *de novo* c.5462insC (p.Leu1822Thrfs\*12) Mutation charakterisiert. *KMT2B* kodiert eine *H3K4 Methyltransferase* und übt essentielle Funktionen vor allem in frühen Phasen der Embryonalentwicklung aus. Eine ektopische Überexpression von Wild-Typ und p.Leu1822Thrfs\*12 *KMT2B* in COS-7 Zellen bestätigte, dass das mutierte Protein zwar stabil ist, seine Fähigkeit in den Zellkern transportiert zu werden allerdings beeinträchtigt ist. Eine weitere Charakterisierung zeigte, dass die Menge an endogener RNA sowie an endogenem Protein in den Patientenzellen reduziert war. Ich konnte mithilfe von *Deep Amplicon Sequencing* aufzeigen, dass das c.5462insC Transkript zwar stabil war, das Protein aber wahrscheinlich nicht, da es weder mit Western Blot noch mit der

## Zusammenfassung

Massenspektrometrie nachgewiesen wurde und somit auf eine zugrundeliegende Haploinsuffizienz hindeutete. Die Gesamtmenge an H3K4me3 war in Patienten-Fibroblasten nicht verändert und diese Zellen zeigten keine erhöhte Apoptose-Aktivierung. Patienten-Fibroblasten wurden im Verlauf der Arbeit erfolgreich in induzierte pluripotente Stammzellen (iPSC) umprogrammiert und eine anschließende RNA Sequenzierung (RNAseq) zeigte, dass einige Gene, die mit dem WNT-Signalweg assoziiert sind, unterschiedlich exprimiert waren. Eine Differenzierung der iPSC in frühe neuronale Vorläuferzellen (NPC) war mit einem normalen Expressionsverlauf des Pluripotenz-Markers *OCT4* sowie der NPC Marker *SOX1* und *PAX6* verbunden. Diese Daten deuteten daher auf eine erfolgreiche Initiation der Differenzierung hin. Um den Pathomechanismus *in vivo* genauer zu untersuchen, habe ich mit der *CRISPR/Cas9*-Methode ein *Knock Out* (KO) Modell für *KMT2B* in dem Zebrafisch *Danio rerio* hergestellt. Die beiden *KMT2B* Homologe *kmt2ba* und *kmt2bb* wurden ähnlich der Position der humanen Mutationsstelle modifiziert. Ich konnte zeigen, dass homozygote *kmt2ba* KO Tiere sowie homozygote *kmt2bb* KO Tiere lebensfähig waren. KO von *kmt2bb*, aber nicht von *kmt2ba*, führte zu Wachstumsverzögerungen und Mikrozephalie. Dies wurde in den Tieren 14 Tage nach der Befruchtung (dpf) sichtbar und ähnelte den Merkmalen unseres Patienten. *Whole-mount* Immunfärbungen für aktiviertes *Caspase 3* zeigte eine erhöhte Apoptose-Aktivierung bei 32 Stunden nach der Befruchtung (hpf). Um die frühe Phase der kortikalen Entwicklung vor dem Auftreten der Apoptose genauer zu charakterisieren, habe ich weitere Analysen in einem Mausmodell durchgeführt. Dafür habe ich ein konditionales heterozygoten *Kmt2b* KO Modell (heterozygoter *Kmt2b* KO) an Embryonalstadium 13.5 (E13.5) untersucht. Immunohistochemische Analysen zeigten keine Unterschiede zwischen Kontrollen und heterozygoten *Kmt2b* KO Embryonen was auf einen schwachen Phänotyp hindeutete. Aus diesem Grund wurde die sensitive *Next-Generation sequencing* (NGS) basierte Methode *single-cell RNA sequencing* (scRNAseq) für die Analyse der Genexpression angewendet. Die Auswertung der scRNAseq Daten wies auf ein leichtes Missverhältnis von teilenden und nicht-teilenden Radial Gliazellen (RGC) in dem sich entwickelnden Vorderhirn der heterozygoten *Kmt2b* KO hin.

Zusammenfassend lässt sich sagen, dass ich mit Hilfe von WES in drei Familien die zugrundeliegende genetische Diagnose der Mikrozephalie identifiziert habe. Meine Ergebnisse liefern den Familien nicht nur die genetische Ursache, sondern eröffnen



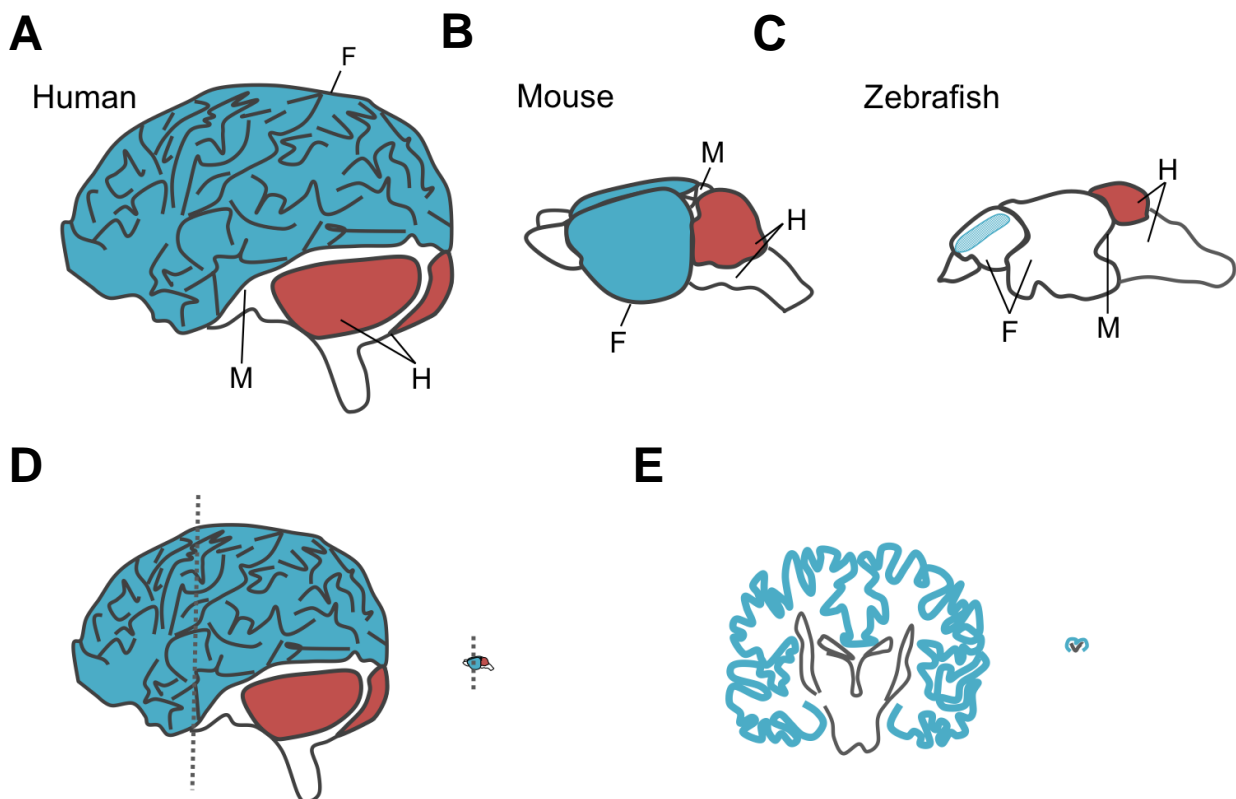
## Zusammenfassung

ihnen auch Perspektiven hinsichtlich Therapiemöglichkeiten. Die Charakterisierung von *KMT2B* und der c.5462insC Mutation ermöglicht neue Einblicke in den Pathomechanismus der Mikrozephalie und auch neue Einblicke in zelluläre Prozesse darüber hinaus.

# 1 Introduction

## 1.1 The mammalian neocortex

The mammalian cerebral cortex consists of the allocortex and the neocortex. The neocortex accounts for the largest part and is responsible for higher cognitive functions in humans (Figure 1A) [1]. Its size expanded tremendously during mammalian evolution, which is reflected mainly via an expansion of the surface area, while cortical thickness did not change significantly (Figure 1D/E) [1]. As a result, the cerebral cortex is arranged in folds (cortical gyration) in some species including humans, while it appears smooth in other species like mice (Figure 1A/B). Even though neocortical tissues of human and mice possess some differences, investigations of physiological and pathological processes have been studied in detail in mice, and have forwarded our knowledge extensively [2,3].



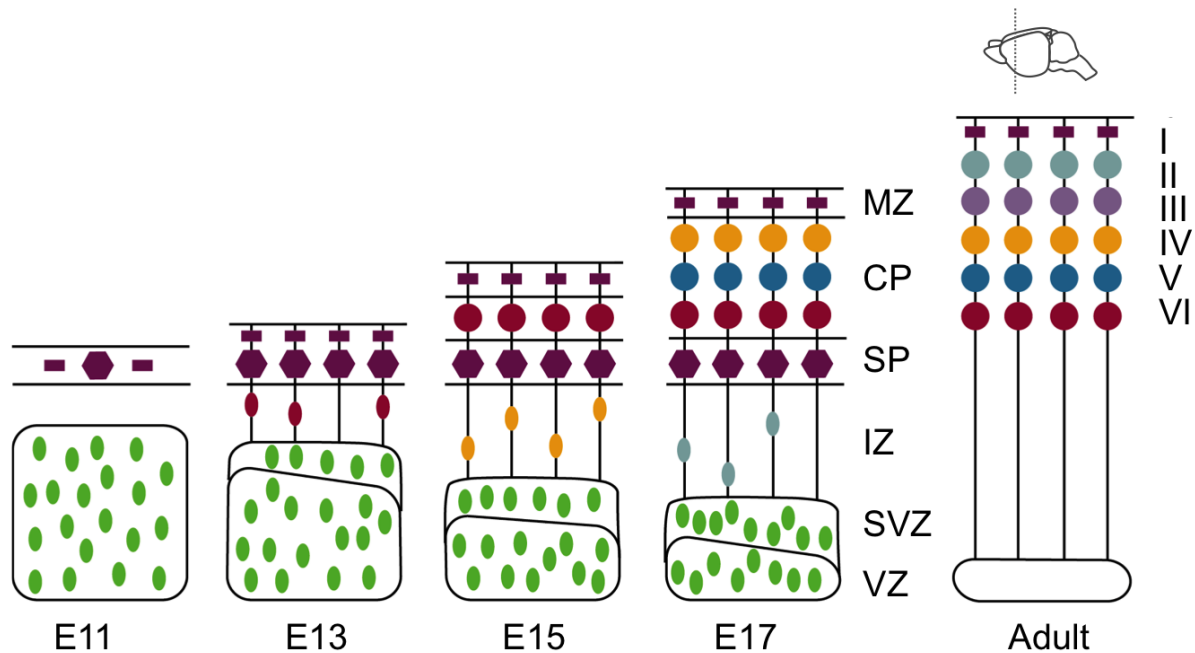
**Figure 1 The cerebral cortex in different species.** The cerebral cortex (blue) in human (A), mouse (B) and zebrafish (C) is part of the forebrain (F). The cerebral cortex faces the outside in human and mouse forebrains, while the dorsal pallium in zebrafish faces the inside (dashed colour). Images do not represent size differences. Midbrain (M), Hindbrain (H). Cerebellum in red. (D) and (E) Size comparison of human and mouse brain. (E) illustrates frontal view of the section plane from (D) (dashed line). Figure modified after Rakic 2009; Miterko et al. 2018; Mueller 2012; Leung et al. 2013 [1,4-6]

### ***1.1.1 Developmental stages and characteristics of murine neocortex***

The neocortex consists mainly of neurons and glial cells, which are generated in a tightly controlled spatial and temporal manner during embryogenesis. Neurons are generated first, followed by the generation of glial cells [7]. The two main classes of neurons in the neocortex are excitatory projection neurons (glutamatergic) and inhibitory neurons (GABAergic), which both can be further subdivided into a wide range of subclasses and subtypes [8-10]. Projection neurons form long axons even to other brain regions, while interneurons build contacts within the neocortex itself [11]. Beside neurons, different oligodendrocytes and astrocytes are components of the neocortex as well.

The source for the majority of cells in the developing neocortex is formed by different types of neural progenitor cells (NPCs), which are already important prior to the onset of neurogenesis that starts at embryonic day 10.5 (E10.5) in the mouse. Neuroepithelial cells (NECs) are the first stem cells in the central nervous system. NECs divide rapidly and perform mainly symmetric proliferative cell divisions thereby generating two daughter cells with identical cellular fate and expanding the pool of progenitors cells in the ventricular zone (VZ) [12]. Between E9 and E10, these NECs transform into radial glial cells (RGCs), which perform mainly asymmetric cell divisions [13,14]. One daughter cell remains as RGC in the VZ (Figure 2). The other daughter cell can either differentiate into an intermediate progenitor cell (IPC) and form the subventricular zone (SVZ), or it can leave the proliferative zones, and finally differentiate into a post-mitotic neuronal cell [15-18]. IPCs in turn, mainly perform symmetric cell divisions yielding either two post-mitotic neuronal daughter cells or two IPCs [16,19,20]. In addition, more progenitors including short neural precursors (SNPs) have been described making the process of cortical neurogenesis even more complex [21].

From E10.5, increasing numbers of post-mitotic neuronal cells leave the VZ and SVZ and migrate along the intermediate zone (IZ) to build up the characteristic layering of the neocortex (Figure 2) [22]. The preplate (PP) can be detected first and it is separated by E13 into the subplate (SP) and marginal zone (MZ). Subsequently, the cortical plate (CP) is formed in an “inside-out” manner indicating that deeper layers are generated prior to upper layers. With progressive corticogenesis, the characteristic six-layered neocortex emerges and remains throughout lifetime [23].



**Figure 2: Steps of murine cortical neurogenesis from E11 to adulthood.** Neural progenitor cells (green) residing in the ventricular zone (VZ) and in the subventricular zone (SVZ). At the beginning of neurogenesis, post-mitotic cells leave these regions and form initially the preplate (PP), and thereafter, the subplate (SP) and marginal zone (MZ). In the course of neurogenesis, migrating neurons form the cortical plate (CP), which can be found in adult mice as six-layered cerebral cortex (layer I-VI). Intermediate Zone (IZ). Figure modified after Gupta et al. 2002; Dehay & Kennedy 2007 [23,24].

### 1.1.2 Signaling pathways involved in cortical neurogenesis

It is still not completely understood how the fate of a differentiating cell is determined during cortical neurogenesis. One challenging question for example is, why progenitors during early embryonic stages produce a broader range of neurons than progenitors during later embryonic stages. Furthermore, it is currently unclear how the cellular switch from neurogenesis to gliogenesis is regulated. However, a wide range of intrinsic and extrinsic factors have been identified that play an important role in these processes. These include different signaling pathways including WNT, Notch, BMP and Hedgehog signaling that have been shown to be involved in the regulation of cortical neurogenesis, and are described in the following.

The Wingless-related integration site (WNT) protein family consists of secreted glycoproteins, which induce an intracellular signaling cascade via binding of Frizzled (Fz) receptor complexes on the outer cellular membrane [25]. This binding stabilizes cytosolic  $\beta$ -Catenin, which translocates to the nucleus and subsequently regulates expression of downstream target genes via interaction with the transcription factor family TCF/LEF (canonical WNT-signaling) [26]. In addition, WNT proteins can also

activate WNT/Ca<sup>2+</sup> and planar cell polarity pathway independent of  $\beta$ -Catenin (non-canonical WNT-signaling) (Figure 12) [25].

Several *Wnt* and *Fz* genes are expressed during different stages of cortical neurogenesis and some genes already as early as E9.5 [27]. Activated WNT-signaling is detectable in the VZ around E14, but is reduced when differentiating post-mitotic neurons leave the VZ and migrate along IZ to form the CP [28]. In addition, canonical WNT-signaling activates expression of *Pax6* and *Ngn1*, two key transcription factors of cortical neurogenesis [29,30]. It seems that WNT-signaling can have different effects on cortical neurogenesis. In RGCs, WNT-signaling promotes symmetric proliferative cell divisions thereby increasing the number of RGCs and inhibiting their differentiation [28,31]. In line with this, overexpression of  $\beta$ -*catenin* results in a delayed generation of IPCs and these mice are characterized by an enlarged cortical surface area with cortical gyrfication similar to human brain morphology [32,33]. In contrast, during later developmental stages, WNT-signaling promotes differentiation of IPCs as well, indicating that effects of WNT-signaling for cortical neurogenesis vary and are dependent on the cellular background [31,34].

Notch-signaling is another pathway and especially important for short range signals since Notch receptors face the extracellular space and can be bound by membrane coupled ligands (Delta-like and Jagged) of neighboring cells [35]. Binding of Notch receptors subsequently results in proteolytic cleavages of the intracellular receptor domain and finally activates expression of downstream target genes including *Hes* genes, a family encoding transcriptional repressors [35]. Notch-signaling has been reported to inhibit cortical neurogenesis via affecting NECs and differentiation of RGCs [14,36]. As a result, an overexpression of activated Notch from E9.5 maintains RGCs and reduces their differentiation into neurons [37]. Similarly, Notch-signaling critically determines the cell fate after asymmetric cell divisions also in zebrafish [38]. BMP (Bone morphogenetic protein) and Hedgehog-signaling are two additional pathways, that have been suggested to take part in diverse functions during cortical neurogenesis. The BMP pathway regulates migration of neurons during corticogenesis [39]. In addition, *in vitro* data point towards an involvement of BMP-signaling in cell fate determination and in the transition from neurogenesis to gliogenesis [40]. The importance of Hedgehog-signaling is reflected by a conditional knock out (cKO) of *Shh*, which alters the cell cycle length of progenitor cells and finally decreases the number of generated neurons [41].

Taken together, these data indicate that the involvement of signaling pathways in cortical neurogenesis is a complex process and far more complicated than illustrated here. One signaling pathway can not only affect different cells and stages of cortical neurogenesis, but also affect the signaling of other pathways thereby forming a complex network required to build up the six-layered neocortex during embryogenesis [42].

### **1.2 Isolated and syndromic forms of human microcephaly**

Microcephaly describes a reduced head circumference in patients of at least 3 SD below the mean compared to healthy individuals of same sex, age and ethnicity [43]. The reduction of the head circumference derives from a reduction of the brain size and mainly from a reduction of the cerebral cortex, which can additionally manifest in simplified cortical gyration and cognitive impairments. The overall structure of the brain is usually not affected [44,45]. Microcephaly is described as isolated when no other malformations are present. In contrast, microcephaly can also develop as clinical feature in a wide range of inherited syndromes with additional severe brain malformations and other affected organs. To date, hundreds of microcephaly-associated syndromes have been described with microcephaly either manifesting during embryogenesis (primary microcephaly) or after birth (secondary microcephaly) ([www.omim.org](http://www.omim.org)). Between 2011 and 2017, microcephaly occurred with a prevalence of 1.89 per 10,000 individuals in the European population and thus, it is still a rare clinical feature in Europe ([www.eurocat-network.eu](http://www.eurocat-network.eu)).

#### **1.2.1 Autosomal recessive and autosomal dominant inheritance**

Cortical neurogenesis involves different cellular processes including cell proliferation, DNA damage response, apoptosis, cell migration, and cell differentiation, which together refine the number of generated cells and cortical layering. Isolated and syndromic forms of primary microcephaly can therefore manifest from mutations in genes, which are associated with any of these cellular processes [46,47]. The microcephaly panel at the Institute of Human Genetics (Göttingen), as an example, consists of 75 microcephaly-associated genes and is used for routine diagnostic testing to date ([www.humangenetik-umg.de](http://www.humangenetik-umg.de)). Most microcephaly-associated genes, including the following examples, show an autosomal recessive mode of inheritance and are thought to cause loss of protein function. *MCPH1* (MIM 251200) was the first identified gene associated with autosomal recessive primary microcephaly (MCPH)

[48]. Depletion of *Mcph1* results in hypersensitivity to  $\gamma$ -irradiation and premature switch from symmetric to asymmetric cell divisions, which is accompanied by a depletion of NPCs [49,50]. Mutations in *ASPM* (MIM 605481) account for most cases of MCPH [44,51]. Similar to *Mcph1*, *Aspm* is important for maintaining symmetric proliferative cell divisions in NPCs [52]. Overexpression of stabilized  $\beta$ -catenin can partially rescue these defects indicating that ASPM is linked to canonical WNT-signaling [53]. Genes associated with DNA damage response like *XRCC2* (MIM 600375), *XRCC4* (MIM 94363) and *LIG4* (MIM 601837) underlie syndromic forms of microcephaly, namely Fanconi anemia, “Seckel-like” and LIG4 syndromes, respectively [54-56]. Depletion of any of these genes results in defective cortical neurogenesis in mice, which is accompanied by excessive apoptosis [57-59].

To date, a few genes underlying autosomal dominant forms of isolated and syndromic microcephaly have been described as well. Missense mutations in the microtubule-associated genes like *KIF5C* (MIM 604593), *KIF2A* (MIM 602591) and *TUBB2B* (MIM 612850) have been identified in patients with complex cortical dysplasia [60,61]. *Tubb2b* was shown to regulate neuronal migration, while a knockdown of *Kif2a* causes impaired proliferation of RGCs followed by premature neuronal differentiation [60,62]. Kadir et al identified a missense mutation in the autophagy related gene *WDFY3* (MIM 617485) in a patient presenting with primary microcephaly [63]. Using different *in vitro* and *in vivo* approaches the authors identified a link between primary microcephaly and canonical WNT-signaling. Their model proposes that WDFY3 attenuates canonical WNT-signaling through degradation of  $\beta$ -Catenin. This reduction is required for the physiological transition of RGCs to IPCs, which is accompanied by a switch from symmetric to asymmetric cell divisions.

### **1.2.2 Model systems for functional analysis of microcephaly**

Several model systems are used to study the pathogenesis of microcephaly and other neurodevelopmental diseases *in vivo*. Zebrafish (*Danio rerio*) and mouse (*Mus musculus*) are among the most frequently used model systems for this achievement. The nervous system of zebrafish is simple, but, similar to mammals, the brain can be subdivided into forebrain, midbrain and hindbrain (Figure 1A-C). The cerebral cortex, as found in mammals, is not detectable in zebrafish. However, the dorsal pallium has been proposed as the corresponding homolog, which faces the inside of the forebrain and not, as in mammals, the surface [64]. The use of zebrafish as a model system

possesses several advantages over other species. The embryos are transparent and enable whole-mount embryo immunofluorescence stainings during the first days of development [65]. The embryogenesis takes in total three days and has been described in detail [66]. After embryogenesis, animals are considered larvae until they reach the juvenile stage and finally become fertile adults [67]. Adult zebrafish can lay batches of 200 – 300 eggs. Fertilization and embryonic development are external, thereby facilitating genetic manipulations. For this, two approaches are mainly used in zebrafish, namely injection of morpholinos in order to knock down (KD) the gene of interest and CRISPR/Cas9 for knock out (KO) studies [68,69]. Due to gene duplications, several human genes have two or even more homologs in zebrafish [70]. These homologs frequently show different expression patterns as well as functions and it is possible that clinical features of the human phenotype are spread over these homologs, when they are depleted separately [68,71-73].

The murine cortical neurogenesis has been extensively studied (see 1.1.1.). It peaks around E13.5 and takes in total eight days (E11 – E19) [24]. Mostly, microcephaly-associated genes are either knocked out or knocked down in order to study their *in vivo* functions, whereas the generation of a knock in is still difficult [59,62]. Since KO of several genes results in early embryonic lethality and hence, prevents proper cortical analysis, the Cre/loxP system is frequently used in order to generate a cKO [74,75]. In these transgenic animals, a KO is only achieved in cerebral cortices when Cre recombinase is expressed under the control of cortex specific promoters like the *Emx1*-promoter [76]. *Emx1* is expressed from E9.5 initially in VZ and with progressing neurogenesis also in IZ, PP, and CP [77]. Cre subsequently excises a targeted exon (flanked by floxP sites) of the gene of interest in developing cortices, while the gene remains intact in other tissues where Cre is not expressed. Depletion of several microcephaly-associated genes recapitulated the severity of the human phenotypes in mice [78-80]. However, for some genes, only a milder or missing phenotype has been reported, which might be explained either by differences between murine and human cortices, or by a higher robustness of mice for genetic changes [24,50,81,82]. Therefore, other *in vivo* models came more and more into the focus of scientists including ferrets, which have gyrified cortices and enable investigations for genes, which do not recapitulate murine phenotypes [83].

To date, *in vitro* cultures also gained more attention based on the introduction of induced pluripotent stem cells (iPSC) and their differentiation into NPCs and cerebral



organoids [84,85]. So far, functional studies using patient-derived material were mainly limited to the use of dermal fibroblasts and even if a molecular phenotype was detectable, the question of the significance for neurogenesis remained unclear [86]. Cerebral organoids might solve this problem. These *in vitro* systems are small three dimensional cultures, which are characterized by several brain-like regions including detectable layers of the cerebral cortex. To date, cerebral organoids have been already used as a model system to study the pathogenesis of microcephaly especially for genes with only mild or missing murine phenotypes [85,87]. In addition, iPSC derived NPCs can be cultured *in vitro* in order to investigate proliferation defects, apoptosis and differentiation ability towards cortical neurons underlying microcephaly [88]. In contrast to the above described *in vivo* animal models, iPSC cultures and differentiated progenies have the strong advantage that the genetic background, including epigenetic effects, of the patient remains unchanged, which might influence the disease severity.

### **1.3 Epigenetic modifications in stem cells and during cortigogenesis**

Epigenetics describes modifications, which are not encoded in the genome itself, but distributed as patterns of DNA methylations and histone modifications all along the genome. Importantly, these modifications influence each other, and thus, they can not be seen as isolated processes [89-91]. Proteins, which are involved in epigenetics can be sub-divided into writers, readers and erasers [92]. Writers modify DNA or histones, readers detect these modifications, and erasers finally remove these modifications.

DNA methylations are accomplished at cytosines by different DNA methyltransferases (DNMTs) in mammals [93]. Depending on the cell type, methylations occur either at CpG dinucleotides or at non-CpG dinucleotides and this modification is mainly associated with gene silencing [94]. Histone modifications, on the other hand, consist of a wide range of modifications like acetylations and methylations that affect the chromatin state and hence, the accessibility for gene expression [95]. Histone modifications are accomplished mainly via two groups of modifiers that act as antagonists in mammals, namely the Polycomb group (PcG) and Trithorax group (TrxG) [96]. The relevance of PcG and TrxG for stem cells and their neuronal differentiation has already been shown [97]. H3K27me3 is considered to mainly repress gene expression and this modification is accomplished by PcG.

TrxG, on the other hand, accomplishes H3K4me3, which is mainly associated with active gene expression [98,99].

The chromatin at developmental genes, including the NPC marker *Pax6*, is bivalently marked in embryonic stem cells (ESCs) via trimethylation of lysine 27 on histone 3 (H3K27me3) and trimethylation of lysine 4 on histone 3 (H3K4me3) [100]. Together, these modifications keep genes in a poised state with no or only very low gene expression (bivalent state) [100-102]. With initiation of differentiation, the promoters of genes important for the neural lineage are de-methylated at H3K27 and these genes are subsequently expressed [103]. Thus, if H3K27me3 is not properly removed in ESCs, these cells have an impaired ability to differentiate towards the neuronal lineage [104]. Pluripotency-associated genes, on the other hand, which are highly expressed in ESCs, lose the activating H3K4me3 modification in the course of neural differentiation and are silenced via H3K9me3 [105].

PcG and TrxG have also been shown to be important during cortical neurogenesis *in vivo*. The PcG members *Ring1b* and *Ezh2*, for instance, are reported to regulate termination of neurogenesis and initiate gliogenesis [106]. Furthermore, Nguyen et al. showed that the TrxG members BAF155 and BAF170 are important for the promotion of neuronal differentiation and repression of cell proliferation during late cortical neurogenesis [74,107]. The authors propose that this regulation is accomplished via interactions of BAF155/BAF170 with demethylases *Kdm1a* and *Kdm6a/b*, which results in the removal of H3K27me3 and H3K4me2, respectively. Finally, genes related to neural differentiation are expressed, while genes related to cell proliferation, including WNT-signaling-associated genes, are repressed.

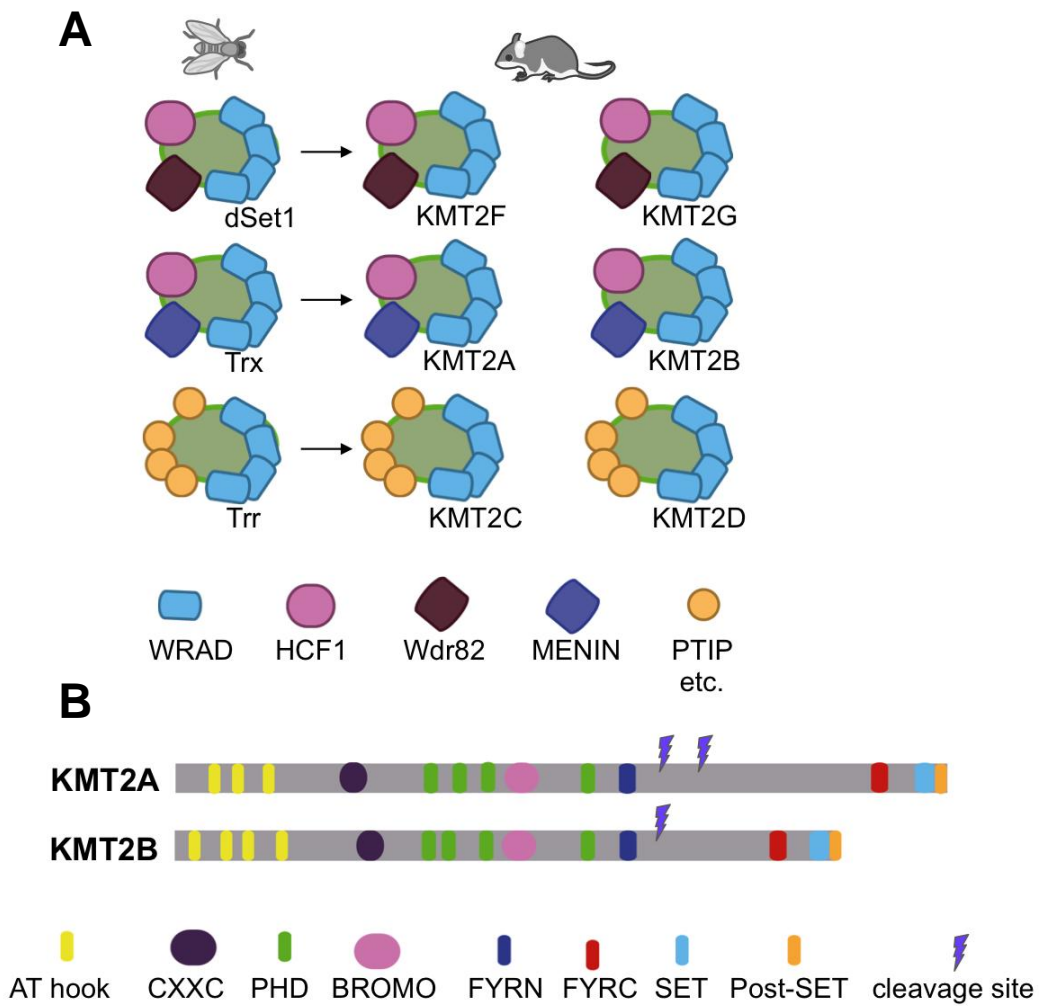
### **1.4 The KMT2 family – The TrxG histone modifier for H3K4me1/me2/me3**

As already described, histone modifications are accomplished by different enzymes including members of the heterogeneous groups PcG and TrxG. TrxG can be further divided into different subclasses including the lysine methyltransferase subclass 2 (KMT2) [108]. Members of KMT2 are related to *Set1*, the only H3K4 methyltransferase found in yeast [109]. The fruit fly *Drosophila melanogaster* contains already three different H3K4 methyltransferases, which evolved after gene duplication events, namely, *drosophila Set1 (dSet1)*, *Thritorax (Trx)* and *Trithorax-related (Trr)* [110]. In mammals, each *Drosophila* gene duplicated once more, giving a total number of six KMT2s (Figure 3A) [111,112]. *KMT2A* (MIM 159555) and *KMT2B* (MIM 606834) evolved from *Trx*, *KMT2C* (MIM 606833) and *KMT2D* (MIM

602113) from *Trr*, and *KMT2F* (MIM 611052) as well as *KMT2G* (MIM 611055) from *dSet1*. All encoded KMT2 proteins are characterized by high molecular weights ranging from 186 kDa (*KMT2F*) up to 593 kDa (*KMT2D*) and by an intrinsic methyltransferase activity in form of the highly conserved SET domain (Figure 3B). *KMT2E*, on the other hand, has no intrinsic methyltransferase activity and shows more homologous features to a different histone modifier family than the KMT2 family [113]. The intrinsic methyltransferase activity of the SET domain, however, is not solely sufficient for proper H3K4me1/me2/me3. Instead, all KMT2 members are found in large multi-protein complexes being called *Complex of proteins associated with Set1* (COMPASS) (with *KMT2F/G* proteins) or COMPASS-like (with *KMT2A/B/C/D* proteins) (Figure 3A) [114-119]. Importantly, COMPASS(-like) complexes share some common cofactors, but they also have additional unique cofactors. The cofactors MENIN, PTIP and WDR82 are specific for *KMT2A/KMT2B*, *KMT2C/KMT2D* and *KMT2F/KMT2G* COMPASS-(like) complexes, respectively [112,114,117]. On the other hand, the proteins WDR5, RBBP5, ASH2L and DPY30 form the WRAD complex, which binds to the SET domain of all COMPASS(-like) complexes (Figure 3A) [120,121]. These cofactors have been shown to be important to increase the enzymatic activity of the SET domain and to guide KMT2s to specific target sites [122-126].

### **1.4.1 The KMT2 family – Importance for embryonic development and diseases**

The fact that mammals need more than one H3K4 methyltransferase immediately raises the question about the reason. A reduction of *Kmt2a*, *Kmt2b* and *Kmt2c* does not affect bulk levels of either H3K4me1, H3K4me2 or H3K4me3 in murine ESCs indicating that these enzymes indeed have some redundant functions [127,128]. KMT2s, however, also acquired unique essential functions, which is highlighted by different propensities for H3K4me modifications and target sites. As an example, the proteins *KMT2A* and *KMT2B*, as well as *KMT2F* and *KMT2G*, show mainly different locations in nuclei, and hence have different target sites [118,129]. Most KMT2s are capable of accomplishing H3K4me1, H3K4me2 and H3K4me3 [130-135]. *KMT2A* and *KMT2B* show high rates of H3K4me1 and H3K4me2, and bind to promoter as well as enhancer regions [133,134,136,137]. Even though their H3K4me3 rates are low, *Kmt2b* is the major modifier for H3K4me3 at promoter regions in murine ESC and importantly, *Kmt2a* seems to be its backup [127,128]. *KMT2C* and *KMT2D*



**Figure 3: The KMT2 family in mammals. (A)** Three pairs of proteins evolved in mammals after gene duplication events. KMT2F and KMT2G evolved from dSet1, KMT2A and KMT2B from Trx and KMT2C and KMT2D from Trr. KMT2s are found in large COMPASS(-like) complexes characterized by unique and overlapping cofactors. **(B)** Comparison of KMT2A and KMT2B protein with similar domain structure and Taspase 1 cleavage sites (purple flash). Figure modified after Hu et al. 2013; Bögershausen et al. 2013; Rao & Dou 2015 [128,143,144].

accomplish H3K4me1 at enhancer regions, whereas KMT2F and KMT2G are linked to H3K4me3 modifications and bind mainly to promoters [117,138-142].

How these complexes are recruited to their target sites is still not entirely understood. However, long non-coding RNAs (lncRNAs), a wide range of cofactors, different patterns of DNA methylations and transcription factors seem to be involved [108].

The different patterns of H3K4me1/me2/me3 regulate essential processes during embryogenesis. KMT2A as well as KMT2B are considered crucial regulators for a different subset of *HOX* genes, which are important to regulate segment identities during embryonic development [145-147]. KO of any *Kmt2* member results in lethality during different stages of embryogenesis and is characterized by different congenital phenotypes. KO of *Kmt2f* results in lethality as early as E7.5, whereas KO of the

other *Kmt2s* ranges from E10.5 (*Kmt2a*) over E11.5 (*Kmt2b*, *Kmt2d* and *Kmt2g*) up to time of birth (*Kmt2c*) [130,141,146,147]. Similarly in humans, no homozygous mutations have been reported in genetic diseases yet, indicating that a complete loss of any *KMT2* member is also in humans not compatible with life. In contrast, heterozygous mutations have been reported in almost all *KMT2* genes, which most probably result in loss of protein function. These genetic developmental diseases include Wiedemann-Steiner syndrome (*KMT2A*), childhood-onset dystonia (*KMT2B*), Kleefstra syndrome 2 (*KMT2C*), Kabuki syndrome (*KMT2D*), and schizophrenia (*KMT2F*) [148-152]. Microcephaly has been reported as a clinical feature in patients with mutations in *KMT2B*, *KMT2C* and *KMT2D*, accounting for up to 58 % of affected individuals [150,153,154].

Finally, somatic mutations have frequently been reported in all *KMT2* genes underlying different types of cancer in tissues like lung, brain and bladder (<https://cancer.sanger.ac.uk/cosmic>). *KMT2C* is considered an important tumor suppressor gene [155]. The proto-oncogene *KMT2A*, on the other hand, is known especially for its involvement in the development of acute leukemia. Loss of *Kmt2a* causes hematopoietic abnormalities, but does not result in the onset of leukemia [147]. Different chromosomal abnormalities, however, cause leukemia via the formation of *KMT2A* fusion proteins with a wide range of fusion partners, that operate as gain-of-function mutants and cause expression of different target genes [156,157]. Even though the pathomechanism is still under investigation, it is already known that the wt *KMT2A* protein plays an essential role in leukemia development, although wt *KMT2A* and *KMT2A* fusion proteins have different target sites [158]. Beside this, *KMT2A* fusion proteins can act as dominant-negative mutants as well. In this scenario, *KMT2A* fusion proteins impair the S phase checkpoint after DNA damage, which prevents stabilization of wt *KMT2A* protein [159].

These data highlight that all *KMT2s* have essential unique functions during embryonic development and diseases. Since *KMT2B* is of great interest for the present PhD thesis, the current knowledge about *KMT2B* is summarized in the following paragraph.

### **1.4.2 *KMT2B* - The unique *KMT2* family member**

*KMT2B* is widely expressed throughout human tissues [160]. The gene consists of roughly 8.2 kilobases (kb), which are packed in 37 coding exons. In addition to one protein coding transcript (ENST00000420124.2), the *KMT2B* gene encodes for an

## Introduction

additional non-coding lncRNA (ENST00000606995.2), which can be found in different cell types, including dermal fibroblasts and iPSC (Figure S1).

Full-length KMT2B has a molecular weight of 298 kDa and possesses several important domains and motifs, however, their functions are not completely characterized (Figure 3B). AT hook and CXXC domains are DNA binding domains at the N-terminus of KMT2B and the latter has been shown to be essential for guidance of Kmt2b to correct target sites [129,137]. KMT2B has four PHD fingers that seem to recognize histone modifications and might be involved in the guidance of KMT2B to correct target sites as well. The second PHD finger (PHD2) has an intrinsic E3 ubiquitin ligase activity and can modify H3 and H4, which raises the possibility that the remaining PHD fingers might also perform other so far unknown tasks [161]. A BROMO domain is located between PHD3 and PHD4, however, its function in KMT2B has not been characterized yet. Subsequently, KMT2B has a FYRN domain, a FYRC domain and finally the SET domain at its C-terminus (Figure 3B).

Full length KMT2B is cleaved by the endopeptidase Taspase 1 [162]. The conserved DGVDD cleavage site is located between the FYRN and FYRC domains of KMT2B and these two domains associate after cleavage [163,164]. Even though the relevance of protein cleavage for protein function is still not completely understood, it might be involved in guidance to target sites [164,165].

Beside its function in murine ESCs, the function of KMT2B has been studied in human ESCs as well. It has been shown that bivalent modifications of developmental genes in human ESCs are not fixed, but are dynamically regulated in the course of the cell cycle via KMT2B [166]. These modifications are accomplished in G1 phase and are regulated via phosphorylation of KMT2B by Cyclin-dependent kinase-2 (CDK2). As a result, developmental genes are shortly transcribed during G1 phase and differentiation is initiated [166,167]. KMT2B has been linked to the cell cycle by different previous studies as well. Takeda et al. reported that the C-terminus of KMT2B can bind to the transcription factors E2Fs *in vitro*, a protein family of key regulators for cell cycle progression [162]. In addition, KMT2B regulates the expression of other cell cycle associated proteins like Cyclin B, Cyclin D, p27, HOXA5 and HOXB7 [168]. Reduced proliferation rates after *KMT2B* depletion, however, were only shown in cancer cell lines like human osteosarcoma U2OS cells, while most other publications reported not a reduced proliferation rate, but an increase in apoptosis as cause for observed proliferation defects [168,169].

*Kmt2b* KO in murine ESCs results in increased apoptosis rates and significantly impaired differentiation *in vitro* mainly towards the ectodermal and mesodermal lineage [170]. This is in line with the phenotype of the *Kmt2b* KO mouse, which is characterized by increased apoptosis rates and embryonic lethality before E11.5 [146]. *Kmt2b* seems to be essential mostly during early embryonic development since Cre/loxP induced KO after E11.5 does not result in embryonic lethality and instead, results in viable mice characterized by sterility [133,171]. The sterility in male mice derives from an impaired differentiation ability of spermatogonial stem cells, which is accompanied again by increased rates of apoptosis [171,172]. Similarly, investigation of oocytes in female mice points towards an increased expression of several apoptosis-associated genes [133].

Cre/loxP induced KO of *Kmt2b* in adult mice results only in mild phenotypic abnormalities [171]. Glaser et al. however failed to induce Cre recombination in brains and as a result, effects of *Kmt2b* KO on whole brains in adults can not be excluded. In line with this, Kerimoglu et al. conditionally knocked out *Kmt2b* only in excitatory forebrain neurons of adult mice and indeed, observed a phenotype [173]. Depletion of *Kmt2b* resulted in reductions of H3K4me2 and H3K4me3 at promoter regions and impaired memory formation [173]. A cKO of *Kmt2a* resulted in impaired memory formation as well, albeit most deregulated target genes differed from *Kmt2b* cKO mice [174].

In contrast, severe phenotypic abnormalities have not been reported in heterozygous *Kmt2b* KO mice. Investigations at molecular level, however, revealed that haploinsufficiency of *Kmt2b* results in the development of diabetes [175]. In humans, heterozygous *KMT2B* mutations cause childhood-onset dystonia. The first patients have been reported in 2016 with suggested haploinsufficiency as the disease underlying effect [148,176]. The mutation spectrum ranges from missense, over nonsense and frameshift to splice site mutations, which are distributed all along the *KMT2B* gene. The phenotypic spectrum varies also extensively. Microcephaly is reported in 21 % of patients and occurs only in association with mutations predicted to have a severe effect on protein function [154]. Recently, Barbagiovanni et al. aimed to investigate the pathomechanism underlying dystonia in more detail [177]. They differentiated *Kmt2b* KO murine embryonic fibroblasts (MEFs) into induced neuronal cells (iNs) and observed that *Kmt2b* is essential for differentiation as well as maturation of iNs and for suppression of myocyte fate genes.

## 2 Study aims

Thanks to next-generation sequencing (NGS)-based approaches like whole-exome sequencing (WES), the number of microcephaly-associated genes has increased drastically during the past years. However, the associated molecular changes causing microcephaly are frequently not known and await investigation. In addition, the underlying genetic cause for a large number of patients has not been identified resulting in patients and families that still have to live without a clear genetic diagnosis. This indicates that even though progress has been made in the identification of microcephaly-associated genes and associated molecular changes during the past years, many scientific questions still await investigation.

Facing this, the aim of my PhD thesis can be divided into two separate parts. In the first part of my PhD thesis, I aimed to identify the genetic causes in three patients with different forms of syndromic microcephaly and yet unknown genetic diagnosis. For this, WES data should be analyzed in order to identify novel microcephaly-associated candidate genes or mutations in already known microcephaly-associated genes.

In the second part of my PhD thesis, I aimed to characterize a recently identified mutation in the novel microcephaly-associated gene *KMT2B* in a patient with syndromic microcephaly and no signs of any movement disorder. The *KMT2B* gene is ubiquitously expressed and encodes for a methyltransferase that is important for the regulation of wide range of target genes, especially during embryonic development. In order to understand the underlying pathomechanism in the affected patient, functional analyses should be performed using different *in vitro* as well as *in vivo* models.

In summary, the data of my PhD thesis should help to expand our current knowledge about genetic causes and associated molecular changes that underlie syndromic forms of microcephaly. In addition, a deeper characterization of the widely expressed *KMT2B* gene should help to better understand physiological as well as pathological epigenetic processes even beyond microcephaly development.



### 3 Material & Methods

#### 3.1 Material

##### 3.1.1 Subjects

The present PhD thesis adheres to the Declaration of Helsinki protocols. Patients and family members gave written approval prior to blood or skin biopsy donation. DNA and RNA were extracted using standard protocols by the laboratory diagnostics of the Institute of Human Genetics (Cologne, Germany) and of the Institute of Human Genetics (Göttingen, Germany). Index patient K3627 was clinically investigated at the Institute of Human Genetics (Düsseldorf, Germany). Index patients K3923, K3929 and K3027 were all clinically investigated at the Marmara University Hospital (Istanbul, Turkey). A skin biopsy of index patient K3027 was collected, which was subsequently used to culture primary dermal fibroblasts. WT fibroblasts were either purchased from PromoCell GmbH (Heidelberg, Germany) or received from a female patient who had a breast reduction at the University Medical Center Göttingen (UMG).

##### 3.1.2 Enzymes

Table 1 lists all enzymes, which were applied in the course of my PhD thesis.

Table 1: Enzymes used during PhD project.

Enzyme	Company
<b>Benzonase<sup>®</sup> Nuclease</b>	Novagen, Merck Chemicals GmbH, Schwalbach, Germany
<b>Calf intestine alkaline phosphatase</b>	Fermentas, Thermo Fisher Scientific, Bonn, Germany
<b>Exonuclease I</b>	New England Biolabs, Frankfurt, Germany
<b>(Fast Digest) restriction enzymes</b>	Thermo Fisher Scientific, Bonn, Germany
<b>Pfu Polymerase</b>	Promega GmbH, Mannheim, Germany
<b>Phosphatase-Inhibitor-Mix II</b>	SERVA Electrophoresis GmbH, Heidelberg, Germany
<b>Halt<sup>™</sup> Protease Inhibitor Cocktail</b>	Thermo Fisher Scientific, Bonn, Germany
<b>Proteinase K</b>	AppliChem, Darmstadt, Germany
<b>Shrimp Alkaline Phosphatase</b>	Promega GmbH, Mannheim, Germany
<b>Taq Polymerases</b>	Thermo Fisher Scientific, Bonn, Germany
<b>T4 DNA Ligase</b>	Thermo Fisher Scientific, Bonn, Germany
<b>Trypsin/EDTA</b>	Gibco <sup>®</sup> , Life Technologies GmbH, Darmstadt, Germany

##### 3.1.3 Synthetic oligonucleotides

All synthetic oligonucleotides were either purchased from Integrated DNA Technologies (Leuven, Belgium) or from Eurofins Genomics (Ebersberg, Germany). List of used synthetic oligonucleotides can be found in the supplementary.

### 3.1.4 Bacteria and plasmids

Tables 2 and 3 summarize bacteria and plasmids, which were applied in the course of my PhD thesis.

Table 2: Bacteria used during PhD project.

Bacteria	Company
<i>E.coli</i> Top10F'	Life Technologies GmbH, Darmstadt, Germany
XL10-Gold® Ultracompetent cells	Agilent Technologies MfG GmbH & Co. KG, Wagenhäusel-Wiesental, Germany

Table 3: Plasmids used during PhD project.

Plasmid	Company
FLAG-HA_KMT2B_pcDNA5/FRT/TO	Kind gift of Prof. Dr. Robert Roeder, Laboratory of Biochemistry and Molecular Biology, Rockefeller University, New York, USA
pCasx hCas9	Kind gift of PD Dr. Roland Dosch, Institute of Human Genetics, Göttingen, Germany
pCR™ Blunt II-TOPO®	Life Technologies GmbH, Darmstadt, Germany
pcS2+	Kind gift of PD Dr. Roland Dosch, Institute of Human Genetics, Göttingen, Germany
peGFP-N1	Clontech, Becton Dickinson, Heidelberg, Germany
pJet1.2blunt	Thermo Fisher Scientific, Bonn, Germany

### 3.1.5 Antibodies

Table 4 lists all antibodies, which were applied in the course of my PhD thesis.

Table 4: Antibodies used during PhD project.

Type	Antibody	Dilution	Company
Primary Antibodies	α-Tubulin	1:10,000 (WB)	unknown
	β-Actin	1:20,000 (WB)	Sigma-Aldrich Chemie GmbH, Schnelldorf, Germany
	Caspase-3 (14220)	1:1,000 (WB)	Cell Signaling Technology Europe, Frankfurt, Germany
	Cleaved Caspase-3 (9664)	1:500 (WMI)	Cell Signaling Technology Europe, Frankfurt, Germany
	Cytochrome C (sc-13156)	1:100 (ICC)	Santa Cruz Biotechnology Inc., Dallas, Texas, USA
	GFP (11814460001)	1:4,000 (WB)	Roche Diagnostics, Indianapolis, USA
	Histone 3 (9715)	1:1,000 (WB)	Cell Signaling Technology Europe, Frankfurt, Germany
	H3K4me3 (ab8580)	1:1,000 (WB)	Abcam, Cambridge, UK
	HSC70 (sc-7298)	1:1,000 (WB)	Santa Cruz Biotechnology Inc., Dallas, Texas, USA
	HuC/HuD (A-21271)	1:20 (IHC)	Thermo Fisher Scientific, Bonn, Germany
	Ki-67 (PA5-19462)	1:100 (ICC)	Thermo Fisher Scientific, Bonn, Germany
	KMT2B/MLL4 (ab56770)	1:500 (WB)	Abcam, Cambridge, UK
	Kmt2b	1:1,000 (WB)	Kind gift of Francis Stewart, BioInnovationsZentrum,

## Material & Methods

			Dresden, Germany
	Lamin A+C (ab108922)	1:1,000 (WB)	Abcam, Cambridge, UK
	LIN-28A (AF3757)	1:1,000 (WB)	R&D Systems, Inc., Minneapolis, USA
	Nestin (4D11)	1:100 (ICC)	Novus Biologicals, Cambridge, UK
	NeuN (MAB377)	1:200 (IHC)	Merck Chemicals GmbH, Darmstadt, Germany
	Oct-3/4 (AF1759)	1:1,000 (WB)	R&D Systems, Inc., Minneapolis, USA
	Pax6 (PRB-278P)	1:100 (IHC)	Developmental Studies Hybridoma Bank, Iowa, USA
	Sox2 (MA1-014)	1:1,000 (WB)	Thermo Fisher Scientific, Carlsbad, USA
	Sox2 (14-9811-82)	1:200 (IHC)	eBioscience, San Diego, USA
	TBR2 / Eomes (14-4875-82)	1:200 (IHC)	Thermo Fisher Scientific, Carlsbad, USA
<b>Secondary Antibodies</b>	Different Alexa conjugated IgG antibodies (488,568,647)	1:400 (IHC) 1:300 (ICC) 1:500 (WMI)	Thermo Fisher Scientific, Carlsbad, USA
	Goat-anti-mouse IgG-HRP	1:10,000 (WB)	Santa Cruz Biotechnology Inc., Dallas, Texas, USA
	Goat-anti-rabbit IgG-HRP	1:10,000 (WB)	Santa Cruz Biotechnology Inc., Dallas, Texas, USA
	Goat-anti-rat IgG-HRP	1:10,000 (WB)	Santa Cruz Biotechnology Inc., Dallas, Texas, USA

### 3.1.6 Solutions and buffers

#### Exo-Sap-Mix (400 µl)

Shrimp Alkaline Phosphatase 10 µl  
Exonuclease I 15 µl  
dH<sub>2</sub>O 375 µl

Mix was stored at -20°C and thawed before usage.

#### Agarose gel (1,0 % gel)

Agarose 1.5 g  
1x TBE-Buffer 150 ml

#### LB medium (0.5 l)

Peptone 5 g  
Yeast extract 2.5 g  
NaCl 2.5 g  
Add dH<sub>2</sub>O to 500 ml

LB medium and agar were autoclaved and stored at 4°C.

#### Agar (0.3 l)

Peptone 3 g  
Yeast extract 1.5 g  
NaCl 1.5 g  
Agar 4.5 g  
Add dH<sub>2</sub>O to 300 ml

#### Antibiotic stocks

Ampicillin (100 mg/ml): Ampicillin 1000 mg  
dH<sub>2</sub>O/EtOH (1:1) 10 ml  
Kanamycin (30 mg/ml): Kanamycin 300 mg  
dH<sub>2</sub>O 10 ml

Stocks were stored at -20°C and diluted 1:1000 with Agar/LB medium for use.

## Material & Methods

### Cell culture medium (Fibroblasts, COS-7, HEK293, HeLa)

DMEM (1x)	500 ml
FCS	50 ml
Pen/Strep (10,000 U Pen/ml; 10,000 µg Strep/ml)	5 ml
Amphotericin B	1.4 ml

### 10x TBS Buffer (1 l)

Tris	24.2 g
NaCl	87.6 g

Add dH<sub>2</sub>O to 1 l

The pH value was adjusted with HCl to 7.4 and stored at room temperature (RT).

### TBST Buffer (1 l)

Tween 20	1 ml
10x TBST	100 ml
dH <sub>2</sub> O	899 ml

### Total cell lysis buffer (10 ml)

TRIS/HCl (1 M, pH 7.5)	0.2 ml
NaCl (1 M)	1.5 ml
EDTA (10 mM)	1 ml
NP 40 (10 %)	1 ml
SDS (10 %)	0.25 ml
Protease inhibitors	0.25 ml (added immediately before use)
dH <sub>2</sub> O	5.8 ml

Total cell lysis buffer was kept on ice.

### Cell fractionation buffer

#### Nuclear fraction buffer (10 ml)

1 M Tris (pH 7.5)	200 µl
0.5 M KCl	200 µl
10 mM EDTA	100 µl
10 % NP40	500 µl
Protease Inhibitor	250 µl

Add dH<sub>2</sub>O to 10 ml

#### Cytosolic fraction buffer (1 ml)

1 M Tris (pH 7.5)	50 µl
1 M NaCl	400 µl
10 mM EDTA	100 µl
10 % NP40	100 µl
10 % SDS	50 µl
Protease Inhibitor	25 µl

Add dH<sub>2</sub>O to 1 ml

### Transfer Buffer KMT2B (10x)

Glycine	144.1 g
Tris	30.3 g

Add dH<sub>2</sub>O to 1 l  
Adjust pH 8.3

### Transfer Buffer KMT2B (1x)

Transfer Buffer KMT2B (10x)	100 ml
100 % Methanol	100 ml
10 % SDS	2.5 ml

Add dH<sub>2</sub>O to 1 l

### Nuclei buffer (NB)

10 mM Tris-HCl (pH 7.4)
10 mM NaCl
3 mM MgCl <sub>2</sub>

stored at 4°C

### Nuclei suspension buffer (NSB) (fresh prepared)

NB
1 % SUPERase in RNase Inhibitor (20 U/µl, Ambion)
2 % BSA (20 mg/ml, New England Biolabs GmbH)

**Cell lysis buffer (CLB)**

NB

0.1 % IGEPAL® CA-630

1 % SUPERase in RNase Inhibitor (20 U/μl, Ambion)

2 % BSA (20 mg/ml, New England Biolabs GmbH)

**Glyoxal solution**

100 % EtOH 1.6 ml

40 % Glyoxal 0.6 ml

Acetic acid 60 μl

dH<sub>2</sub>O 5.74 ml

pH was adjusted with 1 N NaOH to 6.5 and solution was always prepared fresh.

**3.1.7 Kits**

Table 5 summarizes kits, which were applied in the course of my PhD thesis.

Table 5: Kits used during PhD project.

<b>Kits</b>	<b>Company</b>
<b>Big Dye Terminator V1.1 and V3.1 Cycle Sequencing Kit</b>	Life Technologies GmbH, Darmstadt, Germany
<b>CloneJET PCR Cloning Kit</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>DNase I, Amplification Grade</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>DNeasy Blood &amp; Tissue Kit</b>	Qiagen GmbH, Hilden, Germany
<b>First Strand cDNA Synthesis Kit</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>Gibson Assembly Cloning Kit</b>	New England Biolabs GmbH, Frankfurt, Germany
<b>Invisorb® Spin DNA extraction Kit</b>	Stratec Molecular GmbH, Berlin, Germany
<b>MEGAscript® T7 Kit</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>MicroSpin™ G-50 columns</b>	Merck Chemicals GmbH, Darmstadt, Germany
<b>mMESSAGE mMACHINE® kit</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>NucleoBond® Xtra Maxi Kit</b>	Macherey-Nagel GmbH & Co. KG, Düren, Germany
<b>Nucleo Spin® Plasmid</b>	Macherey-Nagel GmbH & Co. KG, Düren, Germany
<b>Pierce® BCA Protein Assay Kit</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>Qiagen Multiplex PCR Kit</b>	Qiagen GmbH, Hilden, Germany
<b>QIAquick Gel Extraction Kit</b>	Qiagen GmbH, Hilden, Germany
<b>QuantiNova SYBR® Green RT-PCR Kit</b>	Qiagen GmbH, Hilden, Germany
<b>QuantiFast SYBR® Green RT-PCR Kit</b>	Qiagen GmbH, Hilden, Germany
<b>RNeasy Mini Kit</b>	Qiagen GmbH, Hilden, Germany
<b>TOPO® TA Cloning® Kit</b>	Life Technologies GmbH, Darmstadt, Germany
<b>Zero Blunt® TOPO® PCR Cloning Kit</b>	Thermo Fisher Scientific, Carlsbad, USA

**3.1.8 Reference sequences**

The following reference sequences were used:

*KMT2B* RefSeq NM\_014727; NP\_055542

*DLG3* RefSeq NM\_021120; NP\_066943

*DDX11* RefSeq NM\_001257144; NP\_152438

*FANCI* RefSeq NM\_001113378; NP\_NP001106849

### 3.1.9 Chemicals and expendable items

#### Chemicals were purchased from the following manufacturers:

Applied Biosystems, Invitrogen, Gibco<sup>®</sup>, Life Technologies GmbH (Darmstadt, Germany), AppliChem GmbH (Darmstadt, Germany), Biochrom AG (Berlin, Germany), Carl Roth GmbH + Co. KG (Karlsruhe, Germany), Fermentas, Thermo Fisher Scientific (Carlsbad, USA), Merck Chemicals GmbH (Schwalbach, Germany), Promega GmbH (Mannheim, Germany), Qiagen GmbH (Hilden, Germany), Roche Diagnostics (Indianapolis, USA), Sigma-Aldrich Chemie GmbH (Schnellendorf, Germany), 4titude Ltd (Wotton, UK).

#### Expendable equipment was purchased from the following manufacturers:

Becton Dickinson (Heidelberg, Germany), Bio-Rad Laboratories Inc. (Hercules, USA), Corning Incorporated (Corning, USA), Eppendorf (Hamburg, Germany), Carl Roth GmbH + Co. KG (Karlsruhe, Germany), GE Healthcare Europe GmbH (Freiburg, Germany), Greiner Bio-One GmbH (Frickenhausen, Germany), Labomedic (Bonn, Germany), Merck Chemicals GmbH (Schwalbach, Germany), Sarstedt AG & Co. KG (Nümbrecht, Germany), Starlab GmbH (Hamburg, Germany), Stem Cell Technologies (Vancouver, Canada), Stratec Molecular GmbH (Berlin, Germany), Thermo Fisher Scientific (Carlsbad, USA), VWR International GmbH (Darmstadt, Germany).

### 3.1.10 Non-expandable equipment

Table 6 lists non-expandable equipment, which was applied in the course of my PhD thesis.

Table 6: Equipment during PhD project.

Equipment		Company
<b>CASY<sup>®</sup> Cell Counter</b>		OLS OMNI Life Science GmbH & Co. KG, Bremen, Germany
<b>Centrifuges</b>	<b>Heraeus<sup>™</sup> Fresco<sup>™</sup> 21</b>	Thermo Fisher Scientific, Carlsbad, USA
	<b>Heraeus<sup>™</sup> Megafuge<sup>™</sup> 16</b>	Thermo Fisher Scientific, Carlsbad, USA
	<b>MiniStar silverline</b>	VWR International GmbH, Langenfeld, Germany
<b>FACSCanto II flow cytometer</b>		Becton Dickinson, Heidelberg, Germany
<b>Gel Electrophoresis Chambers</b>		Thermo Fisher Scientific, Carlsbad, USA
<b>Heracell 240 CO2 Incubator</b>		Heraeus Med GmbH, Meerbusch, Germany
<b>IKA<sup>®</sup> RCT basic IKAMAG<sup>™</sup> Safety Control Magnetic Stirrer</b>		IKA-Werke GmbH & Co. KG, Staufen, Germany
<b>Julabo<sup>™</sup> shaking water bath SW22</b>		JULABO GmbH, Seelbach, Germany
<b>Lab water system arium<sup>®</sup> pro</b>		Sartorius AG, Göttingen, Germany
<b>Leica CM3050 S cryostat</b>		Leica Biosystems Nussloch GmbH, Wetzlar, Germany
<b>Microscopy</b>	<b>Inverted AE2000 Trinocular cell microscope</b>	Motic Europe, Barcelona, Spain
	<b>Olympus FLUOVIEW FV1000 confocal laser scanning microscope</b>	Olympus Europa SE & Co. KG, Hamburg, Germany
	<b>Olympus BX60 Fluorescence microscope</b>	Olympus Europa SE & Co. KG, Hamburg, Germany
	<b>ZEISS Axio Imager 2</b>	Carl Zeiss Microscopy GmbH, Oberkochen, Germany
	<b>Zeiss Axio Imager Z1</b>	Carl Zeiss Microscopy GmbH, Oberkochen, Germany

## Material & Methods

<b>Zeiss SteREO Lumar.V12</b>	Carl Zeiss Microscopy GmbH, Oberkochen, Germany
<b>Microwave NN-E201WM</b>	Panasonic Corporation, Oaza Kadoma, Japan
<b>Mr. Frosty™ Cryo Freezing Container</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>NanoDrop™ One</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>Neubauer improved chamber</b>	Glaswarenfabrik Karl Hecht GmbH & Co. KG, Sondheim/Rhön, Germany
<b>New Brunswick™ Innova® 40</b>	Eppendorf AG, Hamburg, Germany
<b>pH meter Hanna Instruments™ HI2211-02</b>	Hanna Instruments Deutschland GmbH, Vöhringen, Germany
<b>QuantStudio 5 Real-Time PCR System</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>Scale MC-1</b>	Sartorius AG, Göttingen, Germany
<b>S1000 ThermalCycler</b>	Bio-Rad Laboratories, Inc., Hercules, USA
<b>ThermoMixer™ F1.5</b>	Eppendorf AG, Hamburg, Germany
<b>ThermoScientific™ Safe 2020 Class II Biological Safety Cabinet</b>	Thermo Fisher Scientific, Carlsbad, USA
<b>Vortex-Genie2</b>	Scientific Industries, Inc., Bohemia, USA
<b>Western Blot</b>	ChemiDoc™ Touch Imaging System PowerPac™ Basic power supply Mini-PROTEAN Tetra cell Trans-Blot Turbo Transfer System Bio-Rad Laboratories, Inc., Hercules, USA

### 3.1.11 Online resources and software

Tables 7 and 8 list all online resources and software, which were applied in the course of my PhD thesis.

Table 7: Online resources used during PhD project.

Online Resource	Domain
<b>BLAST</b>	<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>
<b>Clustal Omega</b>	<a href="https://www.ebi.ac.uk/Tools/msa/clustalo/">https://www.ebi.ac.uk/Tools/msa/clustalo/</a>
<b>COSMIC</b>	<a href="https://cancer.sanger.ac.uk/cosmic">https://cancer.sanger.ac.uk/cosmic</a>
<b>CRISPRscan</b>	<a href="https://www.crisprscan.org">https://www.crisprscan.org</a>
<b>DoubleDigest Calculator</b>	<a href="https://www.thermofisher.com/de/en/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-scientific/double-digest-calculator-thermo-scientific.html#/legacy=www.thermoscientificbio.com">https://www.thermofisher.com/de/en/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-scientific/double-digest-calculator-thermo-scientific.html#/legacy=www.thermoscientificbio.com</a>
<b>Ensembl</b>	<a href="http://www.ensembl.org/index.html">http://www.ensembl.org/index.html</a>
<b>Eurocat</b>	<a href="https://eu-rd-platform.jrc.ec.europa.eu/eurocat">https://eu-rd-platform.jrc.ec.europa.eu/eurocat</a>
<b>Exac Browser</b>	<a href="http://exac.broadinstitute.org">http://exac.broadinstitute.org</a>
<b>Fugene HD Protocol database</b>	<a href="https://www.promega.com/techserv/tools/FugeneHdTool/">https://www.promega.com/techserv/tools/FugeneHdTool/</a>
<b>GENECODE</b>	<a href="https://www.gencodegenes.org">https://www.gencodegenes.org</a>
<b>Human Gene Mutation Database (HGMD)</b>	<a href="http://www.hgmd.org/">http://www.hgmd.org/</a>
<b>Human splicing finder (HSF)</b>	<a href="http://www.umd.be/HSF/">http://www.umd.be/HSF/</a>
<b>Institute of Human Genetics (Göttingen)</b>	<a href="https://www.humangenetik-umg.de">https://www.humangenetik-umg.de</a>

## Material & Methods

<b>Mouse organogenesis cell atlas</b>	<a href="https://brotmanbaty.org/mouse-organogenesis-cell-atlas-moca/">https://brotmanbaty.org/mouse-organogenesis-cell-atlas-moca/</a>
<b>Mutation Taster</b>	<a href="http://www.mutationtaster.org">http://www.mutationtaster.org</a>
<b>NCBI</b>	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
<b>Oligonucleotide Properties Calculator</b>	<a href="http://biotools.nubic.northwestern.edu/OligoCalc.html">http://biotools.nubic.northwestern.edu/OligoCalc.html</a>
<b>OMIM</b>	<a href="https://www.omim.org">https://www.omim.org</a>
<b>PolyPhen-2</b>	<a href="http://genetics.bwh.harvard.edu/pph2/">http://genetics.bwh.harvard.edu/pph2/</a>
<b>SMART</b>	<a href="http://smart.embl-heidelberg.de">http://smart.embl-heidelberg.de</a>
<b>UCSC Genome Browser</b>	<a href="http://smart.embl-heidelberg.de">http://smart.embl-heidelberg.de</a>
<b>Uniprot</b>	<a href="https://www.uniprot.org">https://www.uniprot.org</a>
<b>Varbank</b>	<a href="https://varbank.ccg.uni-koeln.de">https://varbank.ccg.uni-koeln.de</a>

Table 8: Software used during PhD project.

<b>Software</b>	<b>Application</b>
<b>Adobe® Illustrator® 17.0.0</b>	Image design
<b>Adobe® Photoshop® CS3 and CS6</b>	Data processing and image design
<b>BD FACSDiva 6.1.3</b>	FACS data analysis
<b>DNASTAR® FinchTV 1.5.0</b>	DNA Sequence analysis
<b>GraphPad® Prism® 5.0a</b>	Scientific graphing and statistics
<b>ImageJ version 1.47v</b>	Processing of images
<b>Microsoft® Excel® Mac 2011 14.7.2</b>	Data processing
<b>Microsoft® PowerPoint® Mac 2011 14.7.2</b>	Image design
<b>Microsoft® Word® Mac 2011 14.7.2</b>	Text writing
<b>PyMOL v1.7.0.5</b>	3D visualization of proteins and mutagenesis
<b>R version 3.4.4</b>	Analysis of RNAseq and scRNAseq data
<b>Sequence Pilot</b>	Deep amplicon sequencing analyses
<b>SnapGene Viewer 4.1.6</b>	Visualization of plasmids



## 3.2 Methods

### 3.2.1 Analysis of nucleic acids

#### 3.2.1.1 Polymerase Chain Reaction (PCR)

Target sequences were amplified via Polymerase chain reaction (PCR). Standard PCR conditions were the following:

<b>Standard PCR reaction (25 µl)</b>		<b>Touchdown-like 58 PCR program (TOUCH58)</b>		
10x Taq Buffer	2.5 µl	94°C	3 min	
MgCl <sub>2</sub> (25 mM)	1.5 µl	94°C	30 sec	} 4 cycles
dNTPs (2.5 mM each)	2 µl	62°C	30 sec	
Forward primer (10 µM)	1 µl	72°C	variable	} 4 cycles
Reverse primer (10 µM)	1 µl	94°C	30 sec	
Taq Polymerase (5 U/µl)	0.15 µl	60°C	30 sec	} 4 cycles
DNA (20 ng/µl)	2 µl	72°C	variable	
dH <sub>2</sub> O	14.85 µl	94°C	30 sec	} 31 cycles
		58°C	30 sec	
		72°C	variable	
		72°C	10 min	
		10°C	10 min	

The duration of the extension step varied depending on the length of the amplified sequence. For Taq Polymerase reactions, 1 min was used per 1 kb of fragment size. If target sequences could not be amplified using the standard PCR reaction, the following PCR reaction was applied.

<b>HotStar PCR reaction (20 µl)</b>		<b>Multiplex cyclor program</b>		
Multiplex PCR Master Mix	10 µl	94°C	15 min	
Forward primer (10 µM)	1 µl	94°C	30 sec	} 4 cycles
Reverse primer (10 µM)	1 µl	62°C	30 sec	
DNA (20 ng/µl)/cDNA	2 µl	72°C	variable	} 4 cycles
dH <sub>2</sub> O	6 µl	94°C	30 sec	
		60°C	30 sec	} 4 cycles
		72°C	variable	
		94°C	30 sec	} 36 cycles
		58°C	30 sec	
		72°C	variable	
		72°C	10 min	
		10°C	10 min	

#### 3.2.1.2 Quantitative PCR (qPCR)

cDNA samples were diluted 1:10 with dH<sub>2</sub>O. PCR were performed in 10 µl reaction volumes on 384 well plates using QuantStudio5 (Thermo Fisher Scientific). Either QuantiNova™ SYBR® Green PCR Kit (Qiagen GmbH) or QuantiFast SYBR Green PCR Kit (Qiagen GmbH) were used following manufacturer's instructions. Three technical replicates were always analyzed per run and sample. Data were analyzed using the ΔΔ CT-method and normalised to human or murine *GAPDH*.

### 3.2.1.3 Reverse Transcription PCR (RT-PCR)

After RNA isolation and cDNA synthesis, a standard PCR reaction was applied using 2 µl of synthesized cDNA as template.

### 3.2.1.4 Gel electrophoresis

To examine DNA fragments, 5 µl of each PCR product was mixed with 6x Loading Dye and separated on a 1.0 % agarose gel. 5 µl of a DNA ladder was used as size standard. If fragments had to be extracted, the QIAquick Gel Extraction Kit was used following manufacturer's instructions.

### 3.2.1.5 Purification and Sanger sequencing

In order to prepare PCR products for DNA sequencing, 8 µl of each PCR product was mixed with 2 µl Exo-Sap-Mix and incubated 30 min at 37°C followed by 5 min at 95°C. Samples were diluted with dH<sub>2</sub>O if necessary and used for Sequencing reaction.

<b>Sequencing reaction (10 µl)</b>		<b>Sequencing cyler program</b>	
5x Buffer	2.25 µl	95°C	30sec
BigDye Terminator v1.1 or v3.1	0.25 µl	95°C	10sec
Primer (10 µM)	0.25 µl	55°C	5sec
purified PCR product	0.5 µl	60°C	4min
dH <sub>2</sub> O	6.75 µl	4°C	10min

} 35 cycles

To confirm sequences for cloning approaches, 100-150 ng of purified plasmids were used per sequencing reaction. Sanger sequencing was performed either by the CCG, the laboratory diagnostic of the Institute of Human Genetics (Göttingen) or Microsynth SeqLab (Göttingen) [178]. Sequences were analyzed with DNASTAR<sup>®</sup> FinchTV.

### 3.2.1.6 DNA extraction

In order to perform the Array-CGH analysis in iPSC, DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen GmbH) and following manufacturer's instructions.

### 3.2.1.7 RNA extraction and cDNA synthesis

RNA isolation was either performed using the RNeasy Mini Kit (Qiagen GmbH) and following manufacturer's instructions or after the protocol of the Transcriptome Analysis Labor (TAL) (Göttingen). Briefly, the medium was aspirated from cells and washed once with PBS. Depending on number of cells, cells were lyzed with 0.5 – 1.0 ml TRIzol per dish. Lysates were collected in 1.5 ml reaction tubes on ice and stored at -80 °C until all samples have been collected.

Samples were thawed and incubated 5 min at room temperature (RT) while shaking. 0.2 ml Chloroform was added per 1 ml TRIzol, samples were shaken for 30 seconds, subsequently incubated for 5 min at RT and centrifuged at 12.000 x g for 15 min at 4 °C. In the meantime 1 µl GlycoBlue and 0.5 ml ice-cold isopropyl alcohol per 1 ml TRIzol were added to a new tube on ice. After centrifugation, the upper phase was transferred to the fresh prepared tubes, vortexed and incubated overnight at -20 °C. The following day, samples were centrifuged at 12,000 x g for 30 min. at 4 °C. The supernatant was removed and the pellet was washed once with ice-cold 75 % Ethanol (1 ml per 1 ml TRIzol). Samples were centrifuged again (see above) and washing was repeated once. Finally, the supernatant was removed, pellets were

## Material & Methods

dried at 37 °C, resuspended in 20 µl H<sub>2</sub>O and measured with the NanoDrop (Thermo Fisher Scientific).

1 µg RNA was used for DNase treatment using DNase I, Amplification Grade (Thermo Fisher Scientific) and following manufacturer's instructions. Subsequently, RNA was used to reversely transcribe cDNA with the First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) following manufacturer's instructions. Samples were stored at -80°C and kept on ice during usage.

### **3.2.1.8 NGS-based methods**

#### **3.2.1.8.1 Whole exome sequencing (WES)**

WES was performed as already previously described [179]. Briefly, sequences were enriched using the Agilent sureselect V6 enrichment kit and were run on an Illumina HiSeq2000 or on an Illumina HiSeq4000 sequencer by the Cologne Center for Genomics (CCG). Data were analyzed with the 'Varbank' exome and genome analysis pipeline v.2.1.

#### **3.2.1.8.2 Deep amplicon sequencing**

In order to generate PCR products for deep amplicon sequencing analysis, primers were designed for PCR products of approximately 150 bp and flanking the site of interest. A standard PCR reaction was performed and quality of products was checked on 1 % agarose gels. Deep amplicon sequencing was performed by the CCG. Data were analysed using the Sequence Pilot Software (jsi medical systems GmbH).

#### **3.2.1.8.3 RNA sequencing (RNAseq) by Poly(A) selection**

Extracted RNA from iPSC were used for RNAseq performed by the TAL. Data pre-processing was accomplished by the TAL using  $\log_2\text{foldchange} \geq 1$  or  $\geq -1$  and  $p\text{-value} < 0.05$ . Analyses of candidate gene lists were performed via Gene set enrichment analysis with 'topGO' version 2.30.1 in R version 3.4.4 [180].

#### **3.2.1.8.4 Single cell RNA sequencing (scRNAseq)**

E13.5 embryos were extracted from maternal uteri. Embryos were decapitated, forebrains dissected and snap-frozen in liquid nitrogen. Tissues were stored at -150 °C. Four forebrains of each genotype were pooled for further analysis. Tissues were collected in 2 ml low binding Eppendorf tubes using 1.5 ml Cell lysis buffer (CLB) and pipetted up and down 20 times using a 1000 µl pipette. Suspensions were filtered using a 40 µm cell strainer and thereafter, the cell strainer was washed once with 200 µl CLB. Samples were centrifuged at 4 °C, 500 x g for 10 min and supernatant was discarded. Subsequently, nuclei were re-suspended in 1000 µl Nuclear Suspension Buffer (NSB), centrifuged (see above), pellets were re-suspended in 300 µl NSB and filtered with 40 µm cell strainer. Finally, nuclei were counted using a Neubauer counting chamber and diluted with NSB to 1,000 nuclei per µl. 20 µl were used for further processing. Library preparation was performed by the Max Planck Institute for Molecular Genetics Berlin using the Chromium Single Cell 3' Library & Gel Bead Kit v3 (10x Genomics).

### 3.2.1.9 Cloning of KMT2B

#### 3.2.1.9.1 Cloning of mutant KMT2B harboring a C-terminal GFP-tag in peGFP-N1

Full length KMT2B cDNA with a C-terminal GFP-tag was already generated during my Master's thesis. To mimic the mutation of patient K3027 in KMT2B (c.5462insC; p.L1822Tfs\*12), a fragment of KMT2B was amplified with the Pfu Polymerase Mastermix and TOUCH58 using the primers KMT2Bcl\_F3\_F and KMT2B\_K3027cl\_R and cDNA of patient K3027 as PCR template.

##### Standard Pfu PCR reaction (50 µl)

10x Pfu Buffer with MgSO <sub>4</sub>	5 µl
dNTP (2.5 mM each)	4 µl
Forward Primer (10 µM)	1 µl
Reverse Primer (10 µM)	1 µl
Pfu Polymerase (2.5 U/µl)	1.25 µl
DNA (20 ng/µl)/cDNA	2 µl
dH <sub>2</sub> O	35.75 µl

The amplified fragment was cloned into pCR<sup>®</sup>2.1-TOPO<sup>®</sup> and used for another round of PCR to introduce the restriction site BamHI at the 3 prime region with the reverse primer KMT2B\_K3027\_rev\_BamHI\_GFP\_N1. PCR products were subsequently purified. Full length KMT2B cDNA in peGFP-N1 with GFP-tag were digested with EcoRV and BamHI. Mutant KMT2B fragments were cloned into digested KMT2B peGFP-N1 using the Gibson Assembly<sup>®</sup> Cloning Kit (New England Biolabs) and XL10-Gold<sup>®</sup> Ultracompetent cells following manufacturer's instructions.

#### 3.2.1.9.2 TOPO<sup>®</sup> cloning and pJET1.2 cloning

If fragments had to be cloned into pCR<sup>®</sup>2.1-TOPO<sup>®</sup> as an intermediate step, an Adenine overhang was added to the PCR product, which was not amplified using Taq polymerase.

##### Adenine Overhang (50 µl)

10x Taq Buffer	5 µl
dNTPs (2.5 mM each)	4 µl
MgCl <sub>2</sub> (25 mM)	3 µl
Taq Polymerase (5 U/µl)	1 µl
purified PCR product	37 µl

Samples were incubated for 30 min at 72°C and afterwards purified using the QIAquick Gel Extraction Kit following manufacturer's instructions. Ligation of inserts with pCR<sup>®</sup>2.1-TOPO<sup>®</sup>, pCR<sup>®</sup>4Blunt-TOPO<sup>®</sup> or pJET1.2/blunt were performed according to manufacturer's instructions and subsequently used for transformation of *E. coli* Top10F'.

#### 3.2.1.9.3 Different KMT2B constructs with N terminal Flag-HA tag in pCS2+

Full length KMT2B cDNA with a N-terminal FLAG-HA tag was kindly provided by the lab of Prof. Dr. Robert Roeder from the Rockefeller University, New York, USA. The cDNA was inserted into the pcDNA5/FRT/TO plasmid. Full length KMT2B cDNA was cut off from plasmid by using the restriction enzyme PmeI (Thermo Fisher Scientific) and cloned into Stul (Eco147I) digested pCS2+ plasmid.

## Material & Methods

To mimic the mutation of patient K3027 in KMT2B (c.5462insC; p.L1822Tfs\*12), a fragment of KMT2B was amplified with Pfu Polymerase using the primers KMT2Bcl\_Fr.3\_F and KMT2B\_K3027\_Stopp\_Xbal\_R and mutant KMT2B-peGFP-N1 (c.5462insC; p.L1822Tfs\*12) as PCR template. PCR product was purified, cloned into pJET1.2./blunt plasmid and subsequently digested with EcoRV and XbaI to remove the fragment. Finally, the fragment was ligated into EcoRV/XbaI digested KMT2B-pCS2+.

In order to generate full length KMT2B harbouring a catalytic inactive SET domain (c.7810\_7811delTAinsGC; p.2604Y>A) or mimicking the c.7852G>A mutation (p.Glu2618Lys), the following cloning strategy was applied. A KMT2B fragment was amplified using the primers KMT2B\_7364\_F and KMT2BwtNtagR and full length KMT2B-peGFP-N1 as template. The amplified fragment was purified and cloned into pJET1.2./blunt plasmid. Subsequently, a mutagenesis PCR was performed using the primers KMT2B\_Y-A\_2604\_F and KMT2B\_Y-A\_2604\_R or KMT2B\_2618\_3\_F and KMT2B\_2618\_3\_R. In order to exchange the mutant fragment in wild-type KMT2B, the mutant fragment in pJET1.2./blunt plasmid and full length KMT2B-pCS2+ were digested using AflIII and XbaI (Thermo Fisher Scientific). Since AflIII cuts the full length KMT2B at position c.7705 and additionally at the 5 prime region, a two step ligation procedure was performed. In a first step, the AflIII/XbaI digested fragment was ligated in AflIII/XbaI digested pcs2+. In a second step, the re-ligated plasmid was cut with AflIII and the KMT2B sequence 5 prime of cDNA position 7705 was inserted.

Finally, a KMT2B construct was generated with a premature stop codon 5 prime of the PHD fingers (c.3522insGA; p.F1175\*). A KMT2B fragment was amplified using the primers KMT2B\_Fr.3\_F and KMT2B\_noPHD\_rev and full length KMT2B-peGFP-N1 as template. PCR product was purified, cloned into pJET1.2./blunt plasmid and subsequently digested with EcoRV and XbaI to remove the insert. Finally, the insert was ligated into EcoRV/XbaI digested KMT2B-pCS2+.

For long-term storage of transformed bacteria, 800 µl of individual bacterial colonies were mixed with 200 µl 100 % Glycerol in a cryo vial and stored at -80°C.

### 3.2.1.9.4 Colony PCR

In order to identify colonies carrying the correct sequences, single bacteria colonies were picked and diluted in 20 µl dH<sub>2</sub>O. Primers were chosen that either bind to the appropriate vector or to the inserted sequence. Amplification occurred with the Colony PCR reaction and TOUCH58 program.

#### Colony PCR reaction (25 µl)

10x Taq buffer	2.5 µl
MgCl <sub>2</sub> (25 mM)	1.5 µl
dNTPs (2.5 mM each)	2 µl
Forward Primer (10 µM)	1 µl
Reverse Primer (10 µM)	1 µl
Taq Polymerase (5 U/µl)	0.15 µl
Colony diluted in dH <sub>2</sub> O	5 µl
dH <sub>2</sub> O	11.85 µl

PCR products were separated on 1.0 % agarose gels. Positive colonies were inoculated in 4 ml LB medium containing the appropriate antibiotic and incubated overnight at 37°C while shaking at 225 rpm. Plasmids were isolated with the NucleoSpin® Plasmid Kit, digested with appropriate restriction enzymes and finally prepared for Sanger sequencing.

### 3.2.1.9.5 Digestion with restriction enzymes

For the following sub-cloning of inserts into expression vectors, inserts and target vectors had to be digested with appropriate restriction enzymes. The amount of used enzymes and choice of most suitable buffers based on information from the Thermo Scientific DoubleDigest online tool, and were performed in a 30 µl reaction volume. 500-1500 ng of vectors were used per reaction and incubated for at least 2 h at 37°C. If FastDigest enzymes (Thermo Fisher Scientific) were used, samples were incubated 15-30 min at different temperatures according to manufacturer's instructions. Digested products were separated on 1 % agarose gels, purified with the QIAquick Gel Extraction Kit and stored at -20°C. To prevent re-ligation of single digestion reactions, 1 µl of CIAP was added to the vector-enzyme digestion reaction after 3 h and additionally incubated for 1 h at 37°C.

### 3.2.1.9.6 Ligation and Transformation

The following ligation reaction was prepared and incubated for 2 h at RT.

<b>Ligation reaction (10 µl)</b>	
digested vector	1 µl
10x T4 DNA Ligase buffer	1 µl
T4 DNA Ligase	1 µl
digested insert	7 µl

A negative control (no insert) was always performed to determine the amount of successfully digested vectors. Thereafter, the ligated constructs were used for transformation of *E. coli* bacteria using the heat shock method. For that purpose, 50 µl of competent *E. coli* Top10F' bacteria were thawed, mixed with the ligation reaction and incubated on ice for 30 min. Samples were heat shocked for 20 sec in a 42°C water bath and immediately filled up with 300 µl LB medium. Depending on the positive selection markers Ampicillin and Kanamycin, bacterial cultures were shaken at 225 rpm in the 37°C shaker for 30 min or 60 min, respectively. Finally, cultures were spread on antibiotic containing agar plates and incubated overnight at 37°C.

## 3.2.2 Cell culture

Cells were kept in an incubator in a humidified atmosphere at 37 °C and 5 % CO<sub>2</sub>, unless otherwise stated.

### 3.2.2.1 Cultivation of fibroblasts from a skin biopsy of index patient K3027

A skin biopsy of index patient K3027 was cut using a sterile scalpel, transferred to 10 cm cell culture dish and incubated in cell culture medium which was exchanged if necessary. After a few weeks, fibroblasts became apparent and had to be detached for the first time. For this, the cell culture medium was removed and cells were washed once with 6 ml PBS. 2 ml of 0.02 % EDTA/PBS was subsequently added and dishes were incubated in the incubator for several minutes until cells became bent. The EDTA/PBS solution was discarded and 2 ml Trypsin was added. Upon cell detachment, 8 ml cell culture medium was added per dish. Fibroblasts were transferred to new 10 cm cell culture dishes and designated as passage 1 (P1).

### ***3.2.2.2 Reprogramming dermal fibroblasts to human induced pluripotent stem cells (iPSC)***

Passages 3 of control and patient-derived human primary fibroblasts were used for generation of iPSC following the TeSR™-E7™ episomal protocol of Stem Cell Technologies. Briefly, primary fibroblasts were detached,  $1 \times 10^6$  cells re-suspended in 120  $\mu$ l Buffer R and transferred to a mixture of episomal vectors containing 3  $\mu$ g each (OCT4, SHP53, SOX2, KLF4, LIN28 and L-MYC) and 1  $\mu$ g of GFP encoding control plasmid. Cells were electroporated using the Neon® Transfection System (Invitrogen) and each vial was subsequently distributed onto two matrigel (Corning) pre-coated 10 cm dishes. Cells were cultured in cell culture medium for at least two days and afterwards in TeSR™-E7™ until iPSC colonies emerged. iPSC colonies were passaged manually and each colony was transferred to a separate matrigel pre-coated well of a 6-well plate containing mTeSR™1 medium including 10  $\mu$ M Y-27632. Henceforth, cells were cultured either in mTeSR™1 medium or in StemMACS™ iPS-Brew XF.

### ***3.2.2.3 Differentiation of iPSC into neural progenitor cells (NPCs)***

iPSC were cultured in StemMACS™ iPS-Brew XF medium (Miltenyi Biotec) for several passages prior to the differentiation start at passage 19. Cells were cultured for 4 passages after being thawed before initiation of differentiation. Two days prior to the start of differentiation, iPSC were counted using the CASY® Cell Counter and  $1.4 \times 10^5$ ,  $1.6 \times 10^5$  and  $1.8 \times 10^5$  cells were seeded on matrigel coated 6 well-plates. Differentiation was accomplished following the Monolayer Culture Protocol of STEMdiff™ Neural System (Stem Cell Technologies). Only iPSC wells with confluencies < 85 % were considered for differentiation. After washing cells once with 1 ml of Versene solution (Gibco), 2 ml of Gentle Cell Dissociation Reagent (Stem Cell Technologies) was added per well and plates were incubated for 10 min at 37 °C. Each well was rinsed several times with a 1000  $\mu$ l pipette and cells of corresponding iPSC clones were pooled in one 50 ml Falcon tube. To ensure breakage of cell clusters and generation of single cells, cell suspensions were re-suspended for at least 200 times using the 1000  $\mu$ l pipette. Formation of single cells was always confirmed with a microscope. 10 ml of DMEM/F-12 was added to each tube and cells were centrifuged for 7 min at 300 x g. The supernatant was discarded and cells were re-suspended in 1-4 ml STEMdiff™ Neural Induction Medium (+SMADi and 10  $\mu$ M Y-27632). Subsequently, cells were counted using the CASY® Cell Counter and  $1.5 \times 10^6$  cells were seeded on matrigel coated 6-well plates ( $3 \times 10^5$  for 24-well plates and  $6 \times 10^5$  cells for 12-well plates). A medium change was performed daily (without Y-27632). On day 5, cells were passaged for the first time to obtain P1 of NPCs. Briefly, the medium was removed and 1 ml Accutase was added per 6-well. Cells were incubated for 8 min at 37 °C. Each well was rinsed several times by using a 1000  $\mu$ l pipette and cells of corresponding clones were pooled in one 50 ml Greiner tube. 5 ml DMEM/F-12 (Gibco) was added to each tube and cells were centrifuged for 5 min at 300 x g. The supernatant was discarded and cells were re-suspended in 1-3 ml STEMdiff™ Neural Induction Medium (+SMADi and 10  $\mu$ M Y-27632). Subsequently, cells were counted using the CASY® Cell Counter and  $1.23 \times 10^6$  cells were seeded on matrigel coated 6-well plates ( $2.5 \times 10^5$  for 24-well plates and  $5 \times 10^5$  cells for 12-well plates). A medium change was performed daily (without Y-27632) until cells were either harvested or passaged for the next time.

### **3.2.2.4 *Passaging cells***

#### **3.2.2.4.1 iPSC**

Detachment of iPSC was performed either with ReLeSR™ (Stem Cell Technologies) or Versene (Gibco) following manufacturer's instructions. Size of iPSC colonies was reduced by re-suspending cell suspensions carefully with a 1000 µl pipette. Matrigel coated dishes were prepared at least 3 h prior to usage and stored in the cell incubator, at room temperature or in the fridge up to one week. Cell suspensions were transferred to freshly coated matrigel dishes in different dilutions.

#### **3.2.2.4.2 Primary fibroblasts, HEK293T and COS-7 cells**

Cells were washed once with PBS. After removal of PBS, 2 ml Trypsin was added and dishes were incubated at 37°C until cells detached. 8 ml cell culture medium was added and cell suspensions were distributed to new dishes at desired dilutions.

### **3.2.2.5 *Freezing and thawing cells***

#### **3.2.2.5.1 iPSC**

After detachment of cells (see above), cell suspensions were collected in 15 ml Falcons and mixed with FBS (ES, heat-inactivated) and DMSO. Each cryo vial contained a mixture of 400 µl FBS, 100 µl DMSO and 500 µl cell suspension. Cryo vials were transferred to -80 °C in Mr. Frosty™ Freezing Container. For long-term storage cells were transferred to -150 °C.

For revival, cryo vials were thawed in 37 °C water bath. Cell suspensions were transferred to 15 ml Falcon tubes containing 5 ml iPSC culture medium and centrifuged for 5 min at 300 x g. The supernatant was discarded and the cell pellet carefully re-suspended in iPSC culture medium (+10 µM Y-27632). Finally, cells were transferred to matrigel coated dishes and incubated at 37 °C.

#### **3.2.2.5.2 Primary fibroblasts, HEK293T and COS-7 cells**

Cells were detached and after addition of fresh cell culture medium, the cell suspensions were transferred to 15 ml tubes and centrifuged for 5 min at 1500 rpm. After centrifugation, supernatants were carefully removed and cells were re-suspended in 2 ml freezing medium (90 % FCS and 10 % DMSO) and distributed to cryo vials. Cryo vials were transferred to Mr. Frosty™ Freezing Container and placed into the -80°C freezer. For long-term storage, cells were placed in the liquid nitrogen tank or in the -150 °C freezer.

For revival, cryo vials were immersed in a 37°C water bath until the suspension was almost thawed. The content was pipetted into a 10 cm cell culture dish containing 10 ml cell culture medium. The medium was always exchanged the following day.

### **3.2.2.6 *Cell transfection***

Transfections were performed using FuGENE® HD Transfection reagent (Promega) and following manufacturer's instructions. Cells were analysed 24 - 48 h post transfection.

### **3.2.2.7 *Protein extraction***

#### **3.2.2.7.1 Total cell lysate**

Cells were washed once with PBS while keeping the dishes on ice. 200-1000 µl ice cold total cell lysis buffer was added per dish. Cells were scraped off using a cell scraper and collected



## Material & Methods

in 1.5 ml reaction tubes. If pellets were supposed to be stored, cells were detached, collected in 15 ml Falcon tubes and centrifuged for 5 min at 300 x g. The supernatant was removed and pellets were washed once with PBS. After centrifugation for 5 min at 300 x g, the supernatant was removed and pellets were stored at -80 °C. In order to continue with the protocol, cell pellets were re-suspended in cold lysis buffer. Lysates were incubated on ice for 5 min. 1.0 µl Benzonase was added to each tube and incubated on ice for additional 30 min while vortexing the sample every few minutes. Lysates were stored at -20 °C or -80 °C.

### **3.2.2.7.2 Cell fractionation**

Cells were washed once with PBS while keeping dishes on ice. 200-1000 µl ice cold buffer 1 was added per dish. Cells were scraped off using a cell scraper, collected in 1.5 ml reaction tubes and incubated on ice for 20 min while vortexing every few minutes. Cells were centrifuged for 10 min at 12.000 x g and 4 °C. The supernatant containing the cytoplasmic fraction was frozen at -20 °C or -80 °C. The pellet was re-suspended in 50-100 µl buffer 2. 1 µl Benzonase was added per sample and incubated on ice for 30 min while vortexing every few minutes. Samples were centrifuged for 5 min at 13.000 rpm and 4 °C. The supernatant containing the nuclear fraction was stored at -20 °C or -80 °C.

### **3.2.2.7.3 Measurement of protein concentration**

The total protein concentration of cell lysates was estimated after the Bicinchoninic acid assay (BCA) [181]. The BCA Protein Assay Kit was used following manufacturer's instructions. The light absorbance was measured with a spectrophotometer at 562 nm and concentrations were estimated using a calibration curve.

### **3.2.2.8 Mass spectrometry**

Different cell lysates were separated on 4-20 % Mini-PROTEAN® TGX™ precast protein gels (Bio-Rad Laboratories) following manufacturer's instructions. Subsequently, gels were incubated overnight in Roti®-Blue (Carl Roth GmbH + Co. KG) while shaking at 4°C. Gels were washed several times in dH<sub>2</sub>O and forwarded to the Proteomics Service Facility (Göttingen) who performed targeted mass spectrometry using the parallel reaction monitoring technique (PRM). Data were analyzed using scaffold and skyline software.

### **3.2.2.9 Fluorescence activated cell sorting (FACS)**

Cells were detached, collected in 15 ml Falcons and centrifuged at 300 x g for 5 min. The supernatant was discarded, samples were washed once with 2 ml PBS and centrifuged. The supernatant was discarded again and pellets were dissolved in 300 µl cold PBS. Subsequently, 700 µl ice-cold 100 % Ethanol was added drop-wise while vortexing samples carefully. Samples were stored at -20 °C overnight or up to one week. Subsequently, samples were centrifuged at 300 x g for 5 min, the supernatant discarded and pellets re-suspended in 500 µl PBS. 5 µl PureLink™ RNase A (Thermo Fisher Scientific) and 5 µl propidium iodide (Merck Chemicals GmbH) were added. Samples were incubated for 30 min at 37 °C, protected from light while shaking at 190 rpm and afterwards kept on ice. Data were measured using the FACSCanto II flow cytometer and BD FACSDiva Version 6.1.3.

## **3.2.3 Immunological approaches**

### **3.2.3.1 Western Blot**

To prepare samples for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-Page), 15 - 100 µg of protein lysates were mixed with 4x NuPAGE® LDS Sample Buffer

## Material & Methods

(Thermo Fisher Scientific) and 10x NuPAGE<sup>®</sup> Reducing Agent (Thermo Fisher Scientific), followed by incubation at 95°C for 5 min. Either the Precision Plus Protein All Blue Standards (Bio-Rad Laboratories) or the SeeBlue Plus2 Pre-Stained Standard (Thermo Fisher Scientific) were used as reference. For the Thermo Fisher Scientific system, different gels were used including NuPAGE<sup>®</sup> 4-12 % Bis-Tris Mini Gels with running buffers NuPAGE<sup>®</sup> MOPS SDS and NuPAGE<sup>®</sup> MES SDS or NuPAGE<sup>®</sup> 7% Tris-Acetate Mini Gels with the Tris-Acetate SDS running buffer. The gel electrophoresis was performed with 130-150 V for 1.5 h in the XCell SureLock Electrophoresis Cell (Thermo Fisher Scientific). To transfer proteins to nitrocellulose or polyvinylidene difluoride (PDVF) membranes, a blotting sandwich was prepared in a XCell Blot Module. PVDF membranes were initially activated in 100 % methanol (30 seconds) and subsequently washed in dH<sub>2</sub>O. Several blotting pads, two filter papers and one membrane were completely soaked 1x Transfer Buffer (wet blotting). The chamber was filled up with 1x Transfer Buffer. 25-30 V was applied for 2.5 h to enable blotting.

For the Bio-Rad Laboratories system, 4-20 % Mini-PROTEAN<sup>®</sup> TGX<sup>™</sup> precast protein gels were used with 1x TGS Buffer following manufacture's instructions. Electrophoresis was performed with 80-110 V for 1.5 h in Mini-PROTEAN Tetra cells. Semi dry blotting was performed with the Trans-Blot Turbo Transfer System following manufacturer's instructions.

For blocking, membranes were incubated in 5 % powdered milk/TBST for 1 h at RT. Subsequently, membranes were incubated at 4°C overnight in 2 % powdered milk/TBST including the primary antibody at different dilutions. Membranes were washed three times for 20 min with 2 % powdered milk/TBST followed by incubation in 2 % powdered milk/TBST containing the secondary antibody for 1 h at RT. Thereafter, membranes were washed three times with 2 % powdered milk/TBST for 20 min and three times in PBS for 10 min. Either the SuperSignal West Pico Chemoluminescent Substrate (Thermo Fisher Scientific) or the Clarity Western ECL Substrate (Bio-Rad Laboratories) were used for protein detection. All washing and incubation steps were performed while membranes were gently shaking.

### 3.2.3.1.1 Western Blot of the KMT2B protein

The HiMark Pre-Stained Protein Standard (Thermo Fisher Scientific) was used as reference. For separation, lysates were run with NuPAGE<sup>®</sup> 7% Tris-Acetate Mini Gels and Tris-Acetate SDS running buffer (Thermo Fisher Scientific). For blotting of KMT2B protein, the self-made "1x Transfer buffer KMT2B" was prepared and a wet blot was used with PVDF membranes. The blotting sandwich was prepared in a XCell SureLock Electrophoresis Cell (Thermo Fisher Scientific) and immersed in prepared transfer buffer. A magnetic stirrer was inserted and the blotting chamber was placed in a box surrounded by ice. Blotting was performed at 4°C at 30 V overnight while the magnetic stirrer stirred slowly.

### 3.2.3.2 Immunocytochemistry (ICC)

Cells were seeded in 24-well plates on cover slips at least 24 h prior to harvesting. The medium was removed and cells were washed once with PBS. iPSC were incubated 20 min in 4 % Roti<sup>®</sup> -Histofix (Carl Roth GmbH + Co. KG) and primary fibroblasts 30 min in 4 % polymeric formaldehyde (PFA) at RT. Cells were washed with PBS and stored at 4 °C for up to one week. Subsequently, cells were washed twice with PBS and blocked with 50 mM NH<sub>4</sub>Cl for 10 min at RT. Cells were washed twice with PBS and permeabilized three times for 4 min with 0.2 % Triton in PBS. Primary antibodies were diluted in 0.2 % Triton/PBS (1:40 – 1:300 dilutions) and incubated either 1 h at RT or overnight at 4 °C. Cells were washed three times for 4 min with 0.2 % Triton/PBS and subsequently incubated with the secondary antibody (1:300 – 1:1000 in 0.2 % Triton/PBS) for 1 h at RT (protected from light). Cells were

## Material & Methods

washed three times for 4 min with 0.2 % Triton/PBS, two times for 5 min with PBS and merged in ProLong<sup>®</sup> Diamond Antifade Mountant with DAPI (Thermo Fisher Scientific) on slides. Slides dried overnight at RT (protected from light) and were stored at 4 °C. Pictures were taken either with Zeiss Axio Imager Z1 or Olympus FLUOVIEW FV1000 confocal laser scanning microscope. If Phalloidin stainings were performed, Phalloidin was diluted 1:2000 and incubated for 1 h with the secondary antibody solution. To exclude unspecific binding of secondary antibodies, one sample without primary antibody was used as control.

### 3.2.3.3 Immunohistochemistry (IHC)

Frozen slides were thawed and washed 10 min with PBS in a staining dish while rocking. Sections were surrounded with an ImmEdge Pen (Vector). Thereafter, slides were incubated for 1 h at RT in a humidified box with 50 µl blocking solution (5 % Normal goat serum (Biozol) in PBT Triton 0.1 %). Slides were washed twice with PBS for 5 min while rocking in a staining dish. Primary antibodies were diluted in blocking solution and incubated with samples overnight at 4 °C in a humidified box. Samples were washed twice with PBS for 15 min while rocking in a staining dish. Secondary antibodies were diluted in blocking solution and incubated with samples for 2 h at RT in a humidified box. Samples were washed twice with PBS for 30 min while rocking in a staining dish. For staining of multiple antigens, the previous described steps were repeated beginning with dilution of primary antibodies. Finally, Vectashield with DAPI (Biozol) was added, samples were covered with cover slips and stored at 4 °C. Images were taken using ZEISS Axio Imager 2 and processed with Adobe Photoshop CS3 or CS6.

### 3.2.4 The zebrafish *Danio rerio* – in vivo analysis

#### 3.2.4.1 Maintenance of zebrafish

Maintenance followed the description by Westerfield and was in accordance with regulations of the Georg-August-University Göttingen, Germany [182].

#### 3.2.4.2 Genotyping zebrafish

Fishes were anaesthetized in 200 ml water containing 4 ml 0.4 % Tricaine and 200 µl 1 % Isoflurane. The exterior part of the fin was cut with a scalpel and transferred to a PCR tube. Embryos and tail fins were lysed under the following conditions. After removal of E3 medium, 25 µl 25 mM NaOH, 0.25 mM EDTA was added per sample and incubated for 30 min at 95 °C. Subsequently, 25 µl 40 mM Tris (pH 7.5) was added and samples were stored at -20 °C. 1 µl was used as template for PCR reactions using the primers of Table 9, and genotyping was performed via subsequent Sanger sequencing:

Table 9: Primers used for genotyping of zebrafish.

Gene	Forward primer	Reverse primer
<i>tyr</i>	AGCTCTTCAGCTCGTCTCTC	TGAGTGAGGATACTGCGG
<i>kmt2ba</i>	CACACAATTGTGCACAGTCC	GAGACAGATGAGCACAAAGAC
<i>kmt2bb</i>	GAAGCCCTTGAATTGGGGTT	TGAACAACGTTCAAATCGGTATC

### 3.2.4.3 CRISPR/Cas9-based gene editing in zebrafish

#### 3.2.4.3.1 Establishment of positive control *tyrosinase*

To establish the CRISPR/Cas9-based knock out strategy in zebrafish, a knock out of the *tyrosinase* gene was chosen as positive control following the protocol of Varshney et al. [183]. Briefly, *tyrosinase* sgRNA was amplified in a primer based PCR reaction using the following primers and Mastermix:

Phusion	0.5 µl
Phusion Buffer	10 µl
dNTPs	2 µl
CRISPR_F primer	2.5 µl
tyr_gRNA_4_R primer	2.5 µl
dH <sub>2</sub> O	32.5 µl

PCR product was purified using the Invisorb® Spin DNA extraction Kit (Stratec Molecular GmbH) and quality was evaluated on a 1 % agarose gel. 250 ng of purified PCR product was used as template to reversely transcribe RNA using the Ambion Megascript Kit T7 (Thermo Fisher Scientific) and purified with the Microspin™ G-50 Columns (Merck Chemicals GmbH). Quality of RNA was analyzed with a 1 % agarose gel (appearance of double band) and concentration was measured with NanoDrop. Aliquots were stored at -80 °C until usage.

For preparation of Cas9 enzyme, 5 µg of hCas9 encoding plasmid pCasX was linearized with Asp718I restriction enzyme and checked on a 1 % agarose gel. For purification, digestion solution was filled with RNase free dH<sub>2</sub>O to 100 µl. 100 µl Phenol/Chloroform/Isoamylalcohol (50:49:1) was added and mixed. Solution was centrifuged 5 min at maximal speed. The aqueous phase was transferred to a fresh tube. 10 µl 3 M NaAC (pH 5.2) and 250 µl 100 % EtOH were added and centrifuged 15 min at maximum speed. Pellet was washed with 250 µl 70 % EtOH, dried and re-suspended in 20 µl RNase free dH<sub>2</sub>O. RNA was reversely transcribed using the mMESAGE mMACHINE™ T7 Transcription Kit (Thermo Fisher Scientific) and purified with the Microspin™ G-50 Columns (Merck Chemicals GmbH). Quality of RNA was checked on a 1 % agarose gel and concentration measured with NanoDrop. Aliquots were stored at -80 °C until use.

#### 3.2.4.3.2 Injection of CRISPR/Cas9 solutions

Prior to injections, micro needles were calibrated in order to inject 2 nl per pulse. The following injection solution was prepared on ice and injected into one or two cell stages of wild-type zebrafish embryos (hybrid of ABxTLF) to ensure delivery of 300 pg Cas9 and 50 pg sgRNA:

Cas 9 RNA	x µl (150 ng/µl)
tyr sgRNA	x µl (25 ng/µl)
1 M KCl	1 µl
0.5 % Phenolred	1 µl
<u>RNase free water</u>	<u>x µl</u>
	10 µl

In addition, a solution without sgRNA was always injected to confirm intactness of Cas9 enzyme and un-injected embryos were kept as references. Embryos were incubated in E3 medium for 48 h at 28.5 °C and finally checked for pigmentation alterations with a binocular.

### 3.2.4.3.3 Generation of stable *kmt2ba* and *kmt2bb* knock out lines

Target sites for *kmt2ba* and *kmt2bb* were selected with the designing tool CRISPRSCAN [184]. Exon 28 of each homolog was used as regions of interest to find protospacer adjacent motif (PAM) candidates in order to generate knock outs reflecting the truncating mutation of index patient K3027 (*KMT2B* c.5462insC; p.L1822Tfs\*12) in close proximity. Additionally, exons 28 of *kmt2ba* and *kmt2bb* were sequenced in ABxTLF to exclude impairments at target sites. sgRNAs were amplified as already described above using the reverse primers *kmt2ba\_ex28\_CRSC\_R* and *kmt2bb\_ex28\_CRSC\_R* for *kmt2ba* and *kmt2bb*, respectively. Cas9 synthesis and injections were performed as already described above. Approximately 100 injected embryos per homolog were raised to adulthood, 20 candidates were genotyped via Sanger sequencing to confirm mosaics and positive fish were crossed with ABxTLF. Approximately 100 embryos of the F1 generation were raised to adulthood and genotyping was performed via Sanger sequencing in order to identify candidates for further breeding.

### 3.2.4.4 Measurement of head area

After crossing *kmt2ba*<sup>+/-</sup> x *kmt2ba*<sup>+/-</sup> and *kmt2bb*<sup>+/-</sup> x *kmt2bb*<sup>+/-</sup>, F1 generation was investigated from 1 dpf. Embryos and larvae were fixed in Glyoxal solution. For this, embryos were anaesthetized in 0.4 % Tricaine/1 % Isoflurane, incubated in Glyoxal solution for 30 min on ice and subsequently for 30 min at RT. Animals were either imaged immediately or stored at 4°C for up to 7 days. Subsequently, embryos and larvae were positioned on petri dishes filled with 1 % agarose. Pictures were taken from dorsal view using the Zeiss SteREO Lumar.V12 microscope. Animals were collected in single PCR tubes, labelled according to the order of pictures and genotyped. The head area was determined as illustrated in figure 18B with ImageJ.

### 3.2.4.5 Hematoxylin and eosin (HE) stain

At 12 dpf larvae were anaesthetized, fixed in Glyoxal solution (see 3.2.4.4) and tail biopsies were used for genotyping. Larvae were incubated for 24 h in 4 % PFA at RT, embedded in paraffin and subsequently cut with technical support of Anke Frommhold, Institute of Human Genetics (Göttingen). Slides were incubated for 20 min in Xylol, followed by a declining EtOH series (100 %, 96 % and 70 %) for 1 min each. Slides were washed with dH<sub>2</sub>O and incubated for 15 min in a hematoxylin solution. Thereafter, slides were washed in warm tap water (5 min) and incubated for 20 sec in a 1 % eosin solution. Subsequently, slides were quickly washed in dH<sub>2</sub>O and incubated in ascending EtOH series (70 %, 96 %, 100 %) and in Xylol. Coverslips were attached with one drop of PERTEX<sup>®</sup> (VWR International) and slides were dried for 15 min at 55°C. Pictures were taken using the Olympus BX60 fluorescence microscope.

### 3.2.4.6 Whole-mount immunofluorescence (WMI) for activated Caspase 3

Whole-mount immunofluorescence for activated Caspase 3 was performed at 32 hpf following the protocol of Sorrells et al. [65]. Embryos were embedded in 4% methylcellulose and imaged using the Olympus BX60 fluorescence microscope.

### 3.2.5 *Mus musculus* – in vivo analysis

#### 3.2.5.1 Mouse strains

C57BL/6N mice were provided from the animal facility of the Institute of Human Genetics (Göttingen). *Srg3-Emx1-Cre* and *Mil2CamK-Cre* mice were kindly provided by Dr. Tran Tuoc (Institute of Neuroanatomy, Göttingen, Germany) and Prof. Andre Fischer (German Centre for Neurodegenerative Diseases (DZNE), Göttingen, Germany), respectively. All experiments were performed according to the European and German protection of animal acts. Animals were kept at a 12 h light/dark cycle at 22 °C and 55 ± 5 % relative humidity. *Srg3-Emx1-Cre* were crossed with C57BL/6N mice in order to outcross the floxed *Srg3* allele in animals which were subsequently applied for experiments (*Emx1-Cre*).

#### 3.2.5.2 Genotyping mice

Ear or tail biopsies were incubated overnight at 55 °C in 100 µl DirectPCR-Tail lysis reagent (VWR International) and 5 µl Proteinase K while shaking at 350 rpm. Samples were incubated for 5 min at 95 °C, followed by centrifugation for 5 min at 4 °C and 13.500 rpm. 1 µl was used for PCR reactions using the primers in Table 10.

Table 10: Primers used for genotyping of mice.

Gene	Forward primer	Reverse primer
<b><i>Srg3</i></b>	TGTCATCCATGAGGAGTGGTC	GGTAGCTCACAAATGCCTGT
<b><i>Emx1-Cre</i></b>	ATGCTTCTGTCCGTTTGCCG	CCTGTTTTGCACGTTACCG
<b><i>Kmt2b</i></b>	CGGAGGAAGAGAGCAGTGACG	GGACAGGAGTCACATCTGCTAGG

#### 3.2.5.3 Cryosections

Embryos were extracted from maternal uteri and collected in petri dishes containing cold 1x PBS-DEPC on ice. For E13.5, embryos were decapitated and entire heads were used for further investigation. Tissues were transferred to 12-well plates and washed twice for 15 min in cold 1x PBS-DEPC on ice while rocking. Subsequently tissues were fixed in 4 % PFA. E13.5 heads were incubated 4 h while rocking on ice. Tissues were washed three times for 15 min in cold 1x PBS-DEPC on ice and subsequently incubated in 25 % Sucrose/1xPBS-DEPC at 4 °C overnight while rocking. Tissues were transferred to soft papers to remove residual sucrose, transferred into boxes with Tissue freezing medium (Leica Instruments) and incubated at 4 °C for 1 h. Finally, tissues were positioned in embedding medium, fixed on dry ice and stored at -20 °C. Tissues were sliced using the Leica CM3050 S cryostat in 10 µm thickness. Sections were transferred to slides, dried on a 37 °C heating plate and stored at -20 °C.

#### 3.2.6 Statistics

Data were analyzed using Microsoft Excel and Prism 5 software. The two-tailed student's t-test was applied to evaluate differences between patient and wt samples. Significance is indicated with: \*p<0.05, \*\*p<0.01, \*\*\*p<0.001.

## 4 Results

### 4.1 Analyses of whole-exome sequencing (WES) data

#### 4.1.1 *DLG3 - A novel microcephaly-associated candidate gene*

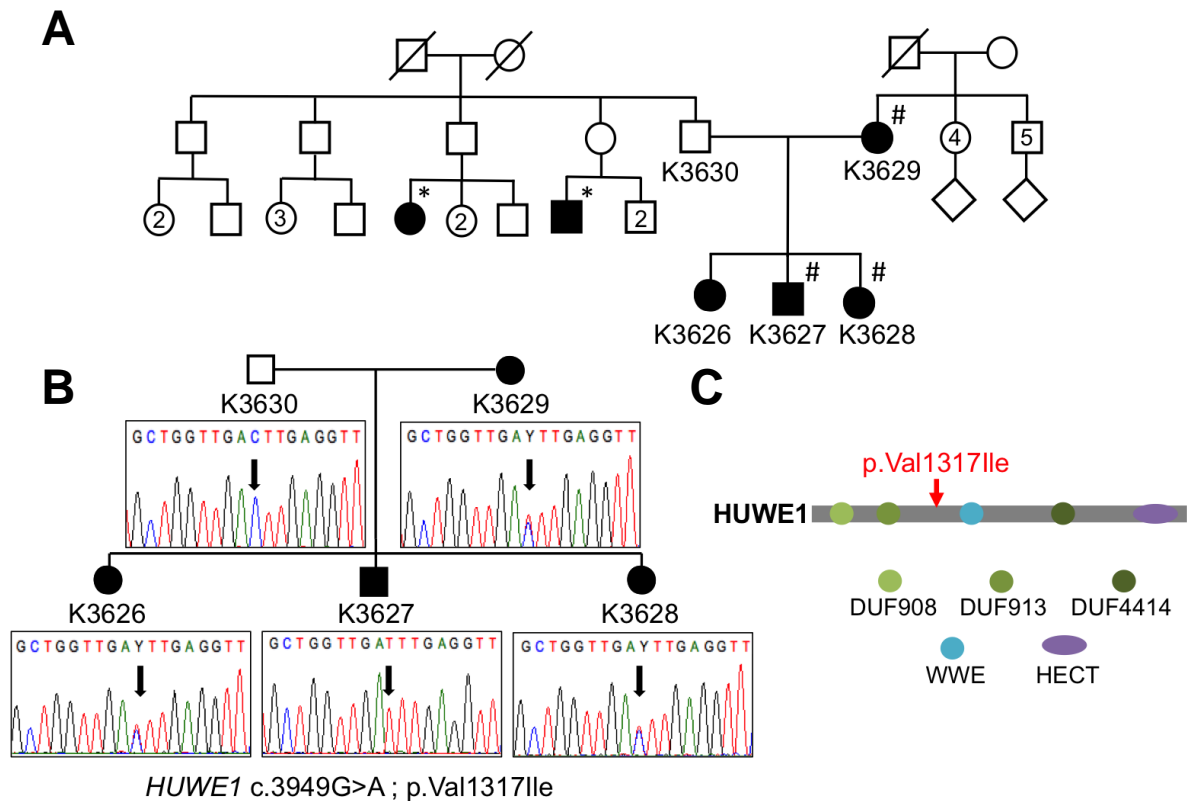
##### 4.1.1.1 *Clinical characterization of index patient K3627*

A non-consanguineous family of Pakistani descent was clinically investigated at the Institute of Human Genetics in Düsseldorf. Both parents and daughters were introduced as healthy, while the male index patient K3627 presented with a syndromic form of microcephaly (Figure 4A). He was the second child and born after an uneventful pregnancy at 37 weeks' of gestation with normal length (48 cm), weight (2.65 kg) and head circumference (34 cm). The motoric development was already impaired in infancy with postural and tonus abnormalities. Physiotherapeutic treatment improved symptoms, so that the patient presented only with a slight form of general muscle hypotonia at the age of 12 7/12 years. His speech development was delayed. Hearing was not impaired. The index patient K3627 had intellectual disability and visited a school for children with learning difficulties. His facial features included longitudinal-oval face, upslanting palpebral fissures, narrow nasal bridge, hyperpigmentation of the oral mucosa, and high-arched palate. At the age of 12 7/12 years his body length (-2 SD), weight (-3 SD) and head circumference (-5 SD) were reduced.

Chromosomal analysis was normal (46, XY) and FISH analysis excluded DiGeorge syndrome. Sanger Sequencing of MCPH-associated genes including *ASPM*, *CASC5*, *CDK5RAP2*, *CENPJ*, *CEP135*, *CEP152*, *MCPH1*, *STIL*, *WDR62*, and *ZNF33* did not identify any disease causing mutations.

A clinical re-evaluation of the family showed that the mother (K3629) and the two sisters (K3626 and K3628) were presenting with microcephaly (K3629 -4.2 SD, K3626 -3.3 SD, K3628 -4.3 SD) as well (Figure 4A). In addition, the older sister (K3626) presented with learning difficulties and growth retardation pointing towards an autosomal dominant or X-linked mode of inheritance.

## Results



**Figure 4: Identification of a variant in *HUWE1*.** (A) Pedigree of a non-consanguineous family of Pakistani descent. The male index patient K3627 was severely affected (filled symbol), while his sisters (K3626 and K3628) and his mother (K3629) were only mildly affected. DNA samples of K3627, K3628 and K3629 were subjected to WES (hash). The father (K3630) was healthy, while some family members of his lineage were presenting with developmental delay (asterisk). (B) Electropherograms depicting the position of the identified c.3949G>A variant in *HUWE1* after Sanger sequencing. (C) Position of the identified p.Val1317Ile variant (red) in *HUWE1* protein characterized by DUF908, DUF913, WWE DUF4414 and HECT domains.

### 4.1.1.2 Somatic mosaicism of *DLG3* in a family with a syndromic form of microcephaly

In order to identify the underlying genetic cause for the syndromic form of microcephaly in the family, DNA from index patient K3627, his younger sister (K3628) and mother (K3629) were subjected to WES, performed by the Cologne Center for Genomics (CCG). Analysis and filtering was conducted with the analysis pipeline 'Varbank' of the CCG. In total, a mean coverage of 74 reads was achieved for index patient K3627, as well as 79 and 78 reads for his sister and mother, respectively. Furthermore, 96.4 % of targets were covered more than 10 x in index patient K3627 and his sister, and 96.3 % in the mother. Based on the family history, I applied filter criteria to search for autosomal dominant and X-linked variants present in all three family members with a severe effect on protein function. However, with this filtering strategy, no convincing candidate gene was identified. Therefore, WES data were re-analyzed with focus on heterozygous missense variants having a milder

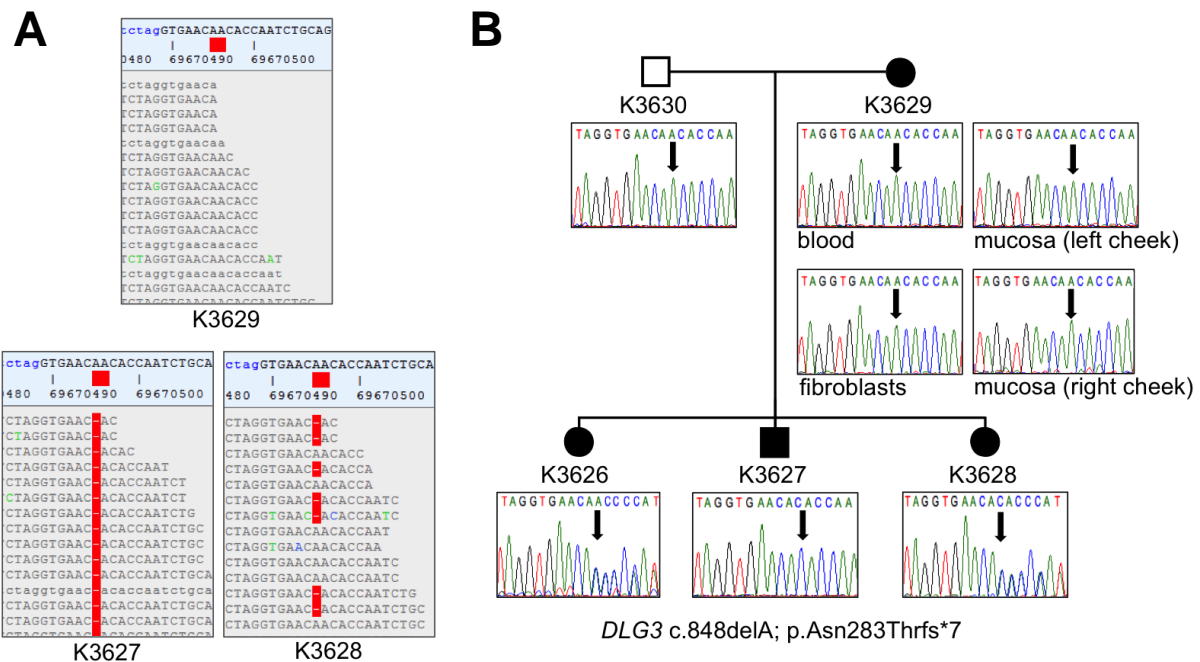


## Results

effect on protein function. This strategy resulted in a candidate list containing 38 heterozygous and hemizygous variants (Table S1). By the use of different analyses tools (see 3.1.11) and published literature, only one promising variant remained, namely the c.3949G>A variant (p.Val1317Ile) in the *HUWE1* gene (MIM 300697). Subsequent Sanger sequencing confirmed that the identified variant was hemizygous in index patient K3627 and heterozygous in his mother and sisters (Figure 4B). The identified missense variant in *HUWE1* was promising because firstly, missense variants in *HUWE1* have already been identified in microcephalic patients with overlapping phenotypic features and affected females have been reported as well [185,186]; secondly, the p.Val1317Ile variant was not listed in the ExAC Browser, which contains 60,706 WES datasets of unrelated individuals (in 2019). Thirdly, *HUWE1* is located on the X Chromosome, explaining a more severe phenotype in index patient K3627 compared to his female family members; and finally, *HUWE1* encodes for an E3 ubiquitin ligase important for the regulation of apoptosis, a common mechanism underlying the pathogenesis of microcephaly [187]. On the other hand, the p.Val1317Ile variant is not located in a known HUWE1 protein domain (Figure 4C), and *in silico* prediction tools differ massively and assign the variant either as benign (PolyPhen-2) or disease causing (MutationTaster).

Therefore, I re-analyzed WES-data to ensure that no other variant was overlooked. This time, filter criteria were applied to search for variants with a severe effect on protein function, detectable in index patient K3627 and his sister, but not in the mother. Using this filter strategy, one severe frameshift mutation was identified in the *DLG3* gene (MIM 300189, RefSeq NM\_021120) predicted to generate a stop codon, namely a deletion of adenine at cDNA position 848 (c.848delA; p.Asn283Thrfs\*7) (Figure 5A). This mutation was highly convincing because firstly, the deletion was not annotated in the ExAC Browser, secondly, the identified mutation introduces a stop codon and hence exerts a severe effect on protein function via truncating the protein. Thirdly, the *DLG3* gene is located on the X Chromosome, explaining a more severe phenotype in index patient K3627 compared to the milder affected sisters. Fourthly, similar mutations in *DLG3* have already been identified in patients with a non-syndromic form of X-linked mental retardation, with overlapping phenotypic features and affected females have been reported as well [188]. Finally, *Dlg3* is expressed during early brain development and involved in signal transduction [189]. A *Dlg3* knock out results in reduced brain weight and reduction of the neocortex area in mice

## Results



**Figure 5: Low-level *DLG3* mosaicism.** (A) Varbank pictures illustrating that the identified c.848delA mutation in *DLG3*, which was identified in index patient K3627 and K3628 but not in the mother K3629. (B) Electropherograms highlighting that the identified c.848delA mutation was not detectable in different tissues of the mother K3629 with Sanger sequencing

[189]. Based on these data, I decided to continue with this mutation and Sanger sequenced all family members. As expected, the hemizygous state of the *DLG3* mutation was verified in index patient K3627, while both milder affected sisters were heterozygous. However, the mutation was not detectable in the mother K3629 (Figure 5B). As a result, a gonosomal mosaicism for the identified *DLG3* mutation was hypothesized in the mother. This type of mosaicism describes mutations that are detectable only in a subset of germ and somatic cells, perfectly matching the segregation of the disease in the present family [190]. DNA derived from blood, from dermal fibroblasts, from mucosa of left cheek and from mucosa of right cheek of the mother were Sanger sequenced; but again, neither tissue was positive for the c.848delA variant (Figure 5B).

The reported variant allele frequencies (VAF) of mosaics ranges enormously in published literature and variants with frequencies of 6 % can not be detected with Sanger sequencing anymore [191,192]. Giving that the c.848delA mutation was not detectable in the mother via Sanger sequencing, the VAF was hypothesized of being less or equal 6 %. Therefore, DNA samples derived from different tissues of the mother and DNA of two control samples (Father K3630 and K3923) were subjected

## Results

to closer examination using NGS-based deep amplicon sequencing, performed by the CCG. For this NGS-based approach, a PCR product surrounding the c.848delA mutation is subjected to parallel sequencing. This strategy increased the coverage at cDNA position 848 from 103 reads (in WES), up to at least 1,805,656 reads (in deep amplicon sequencing) (Table 11). The c.848delA mutation was not detectable in both control samples (0 %), while its frequency ranged from 0.3 % up to 0.9 % in different tissues derived from the mother (Table 11).

**Table 11: Results of deep amplicon sequencing using different DNA samples.** Table specifies total number of reads covering the cDNA position 848 in *DLG3*. Frequency of c.848delA in different DNA samples in percentage. Numbers in brackets indicate amount of c.848delA positive reads.

Sample	total number of reads	c.848delA frequency
Blood K3923	1,805,656	0 % (3)
Blood K3630	3,583,971	0 % (10)
Blood K3629	3,461,401	0.4 % (13651)
Primary Fibroblasts K3629	2,067,722	0.3 % (5175)
Left cheek K3629	2,383,373	0.9 % (21615)
Right cheek K3629	1,843,730	0.3 % (6313)

Taken together, the presented data indicate that *DLG3* is a novel candidate gene underlying microcephaly and the predicted severe effect of the c.848delA variant on protein function makes *DLG3* a more convincing candidate compared to the identified *HUWE1* variant. Further functional analyses will be essential to finally proof, which of the identified variants is the underlying cause for the genetic disease in the presented family.

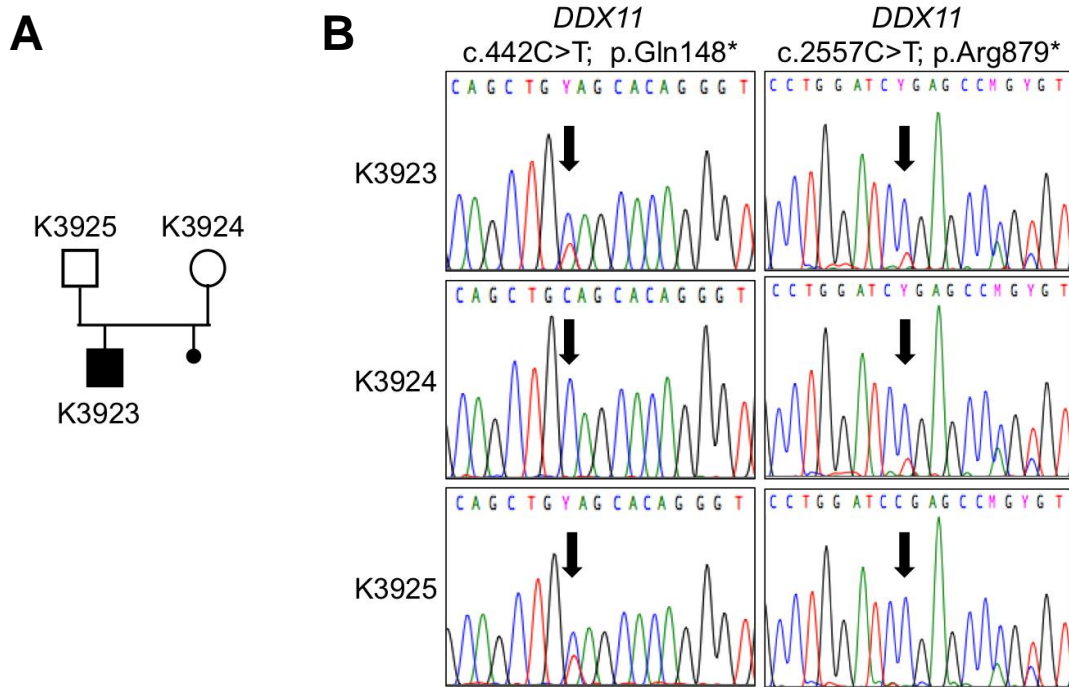
### 4.1.2 Identification of novel microcephaly causing mutations

#### 4.1.2.1 *DDX11* - Novel nonsense mutations underlying syndromic microcephaly

Index patient K3923 was clinically examined in the Department of Pediatrics at the Marmara University Hospital in Istanbul, Turkey. His parents were not related. The mother had a previous miscarriage and index patient K3923 was the only child (Figure 6A). At the age of 18 months he was presenting with deafness, reduced body weight (-2 SD), reduced body length (-2 SD), and reduced head circumference (-6 SD).

In order to identify the underlying genetic cause, DNA of index patient K3923 was subjected to WES, performed by the CCG. WES data were analyzed for variants in

## Results



**Figure 6: Identification of two novel nonsense mutations in *DDX11*.** (A) Pedigree of index patient K3923 illustrating a non-consanguineous family. (B) Electropherograms confirming compound heterozygosity in index patient K3923 for the identified nonsense mutations in exon 4 and 26 of *DDX11*.

known microcephaly-associated genes. Based on the pedigree of the family, filter criteria were applied in order to search for compound heterozygous variants and two variants in the known microcephaly-associated gene *DDX11* (MIM 601150, RefSeq NM\_001257144) were identified, affecting exon 4 and exon 26 of the gene [193]. Both identified mutations are predicted to introduce stop codons in *DDX11* and were not annotated in the ExAC Browser. The c.442C>T mutation in exon 4 is inherited by the father and predicted to terminate translation of *DDX11* protein early at amino acid position 148 (p.Gln148\*) (Figure 6B). The c.2557C>T mutation in exon 26 is inherited by the mother and predicted to introduce a stop codon at amino acid position 879 (p.Arg879\*).

*DDX11* is a DNA helicase important for the proper cohesion of sister chromatids during mitosis [194]. Recessive mutations in *DDX11* underlie Warsaw breakage syndrome and to date only 12 patients have been reported [195]. Patients with Warsaw breakage syndrome present with diverse clinical features, however, almost all reported patients are characterized by severe microcephaly, intrauterine growth retardation, and deafness [195]. The type of *DDX11* mutations reported so far range from missense to nonsense mutations, which are distributed along the *DDX11* gene.

## Results

Based on the severe effect of both identified nonsense mutations and the matching phenotypic overlap with Warsaw breakage syndrome, the presented *DDX11* variants were considered as disease causing.

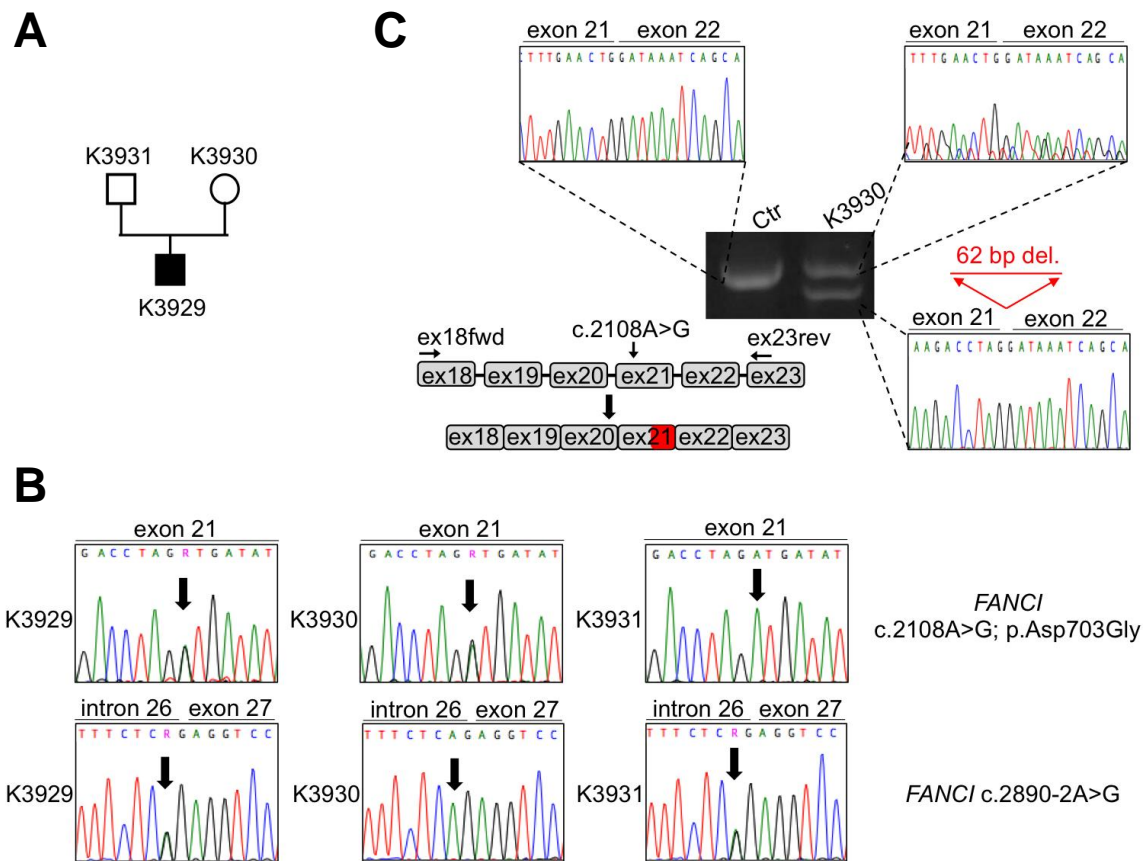
### ***4.1.2.2 FANCI – Novel mutations affecting splicing in a patient with Fanconi anemia***

A 2 year old boy (index patient K3929) was examined in the Department of Pediatrics at the Marmara University Hospital in Istanbul, Turkey (Figure 7A). He was the only child of unrelated parents and presented with a severe form of microcephaly (-6 SD), craniosynostosis, mild trigonocephaly, metopic sutur synostosis, reduced body weight (-2 SD), and reduced body length (-2 SD). He had bilateral clinodactyly and semian creases on both hands. His dysmorphic features included bilateral epicanthus and hyperthelormism. Blood count was normal.

DNA of index patient K3929 was subjected to WES, performed by the CCG. Based on the family pedigree, WES data were analyzed for variants with a severe effect on protein function and either a dominant or compound heterozygous mode of inheritance. Using this strategy, I identified two heterozygous mutations in the known microcephaly-associated gene *FANCI* (MIM 611360, RefSeq NM\_001113378), which were not annotated in the ExAC Browser. One mutation affects the acceptor splice site at the 3' end of intron 26 and changes AG to GG (c.2890-2A>G). The Human Splicing Finder predicts this mutation to most probably affecting the splicing of *FANCI* pre-mRNA. The second mutation is a missense mutation (c.2108A>G; p.Asp703Gly) in exon 21. The prediction tools MutationTaster and Human Splicing Finder propose that the mutation is not only a missense mutation but also a mutation that alters splicing via the activation of an exonic cryptic donor site in exon 21. Sanger sequencing confirmed the compound heterozygous mode of inheritance in the family since the c.2890-2A>G mutation was inherited by the father and the c.2108A>G mutation by the mother (Figure 7B).

Mutations in the *FANCI* gene underlie a phenotypic heterogeneous group of patients with Fanconi anemia and hematological abnormalities as the striking hallmark appearing at different time points during childhood [196]. Since index patient K3929 was only 2 years of age during his clinical examination and no abnormal blood count was detectable, I contacted the responsible clinician and requested a clinical update. Indeed, it was confirmed that index patient K3929 was positively tested for cytopenia

## Results



**Figure 7: Detection of defective splicing of *FANCI* pre-mRNA. (A)** Pedigree of index patient K3929. **(B)** Sanger sequencing confirmed identified mutations in the *FANCI* gene in exon 21 and intron 26. **(C)** Reverse transcribed cDNA derived from the mother K3930 was investigated for a hypothesized splicing defect. Sanger sequencing confirmed formation of an aberrant splice site in exon 21 resulting in the deletion (del.) of 62 bp (red). Cartoon illustrates the position of the c.2108A>G mutation in exon 21, and position of primers used for reverse transcription PCR (RT-PCR).

by the age of 5 years. A Diepoxybutane (DEB) test was positive as well, which further emphasized that *FANCI* was the disease-associated gene.

In order to confirm that the identified c.2108A>G mutation is not only a missense mutation but also a mutation affecting splicing of *FANCI* pre-mRNA, patient-derived RNA was intended to be analyzed in more detail. However, since index patient K3929 received already two bone marrow transplantations, which prevented analysis of blood derived RNA, RNA of his mother (K3930), who carried the mutation, was analyzed instead. In fact, the analysis confirmed that not only the c.2890-2A>G mutation but also the c.2108A>G mutation affects splicing of *FANCI* pre-mRNA. The mutation creates an aberrant exonic splice site in exon 21, which results in the deletion of 62 bp predicted to cause a frameshift and premature deletion of *FANCI* translation (p.Asp703Glyfs\*2) (Figure 7C).

Taken together, compound heterozygous mutations were identified in the *FANCI* gene in index patient K3929 with a predicted severe effect on splicing. The presented

analysis underscores the important teamwork between clinicians and scientists, which is required in order to identify underlying causes of genetic diseases.

### **4.2 Functional characterization of *KMT2B*, a novel gene associated with syndromic microcephaly**

#### ***4.2.1 Clinical characterization of index patient K3027***

Index patient K3027 (hereinafter referred to as “patient”) was clinically examined in the Department of Pediatrics at the Marmara University Hospital in Istanbul, Turkey. The patient is the third daughter of healthy, non-consanguineous Turkish parents (Figure 8A). She was born after an uneventful pregnancy at 39 weeks’ of gestation. Both siblings were healthy. She presented with microcephaly and dysmorphic facial features including a long nose and bilateral epicanthic folds. At the age of 6 years and 6 months, body weight (-5 SD), body length (-3 SD), and head circumference (-9 SD) were reduced. Hematological and biochemical tests were normal. At a follow-up examination at the age of 8 years and 3 months, her body weight was 14.3 kg (-5 SD), body length 108 cm (-3 SD), and head circumference 44.3 cm (-7 SD). Cranial CT was normal. The patient had mild intellectual disability, learning difficulties and was hyperactive. She visited a school for children with learning difficulties. At the age of 10 years, the patient spoke well, was able to run and cycle and hence, she did not present with any movement disorder.

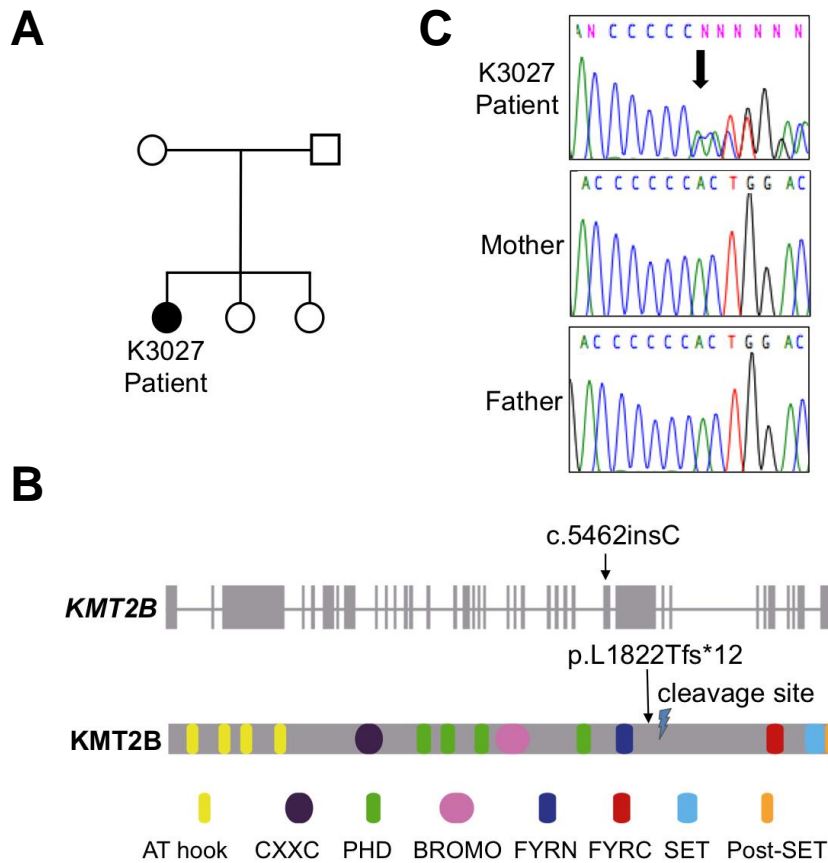
#### ***4.2.2 *KMT2B* – Identification of a novel candidate gene***

Prior to the start of my PhD thesis, DNA of the patient was subjected to WES. After exclusion of variants in known microcephaly-associated genes, a heterozygous mutation in exon 27 of the lysine methyltransferase 2B (*KMT2B*) gene (RefSeq NM\_014727) was identified, namely c.5462insC predicted to result in a frameshift and truncated protein (p.Leu1822Thrfs\*12) (Figure 8B). Parenthood was confirmed via microsatellite marker analysis and subsequent Sanger sequencing confirmed the *de novo* status of the mutation and verified that neither the mother nor the father were carriers (Figure 8C).

At the time when the mutation was identified, *KMT2B* was not associated with any genetic developmental disease, however, the mutation was highly convincing due to the following reasons. Firstly, *de novo* mutations in other members of the *KMT2*



## Results



**Figure 8: The novel microcephaly-associated gene *KMT2B*.** (A) Pedigree of the female index patient K3027. (B) Cartoon illustrating the *KMT2B* gene consisting of 37 coding exons and position of the c.5462insC mutation in exon 27. *KMT2B* protein consists of several domains and motifs. The identified p.Leu1822Thrs\*12 mutation is located between the FYRN domain and the Taspase 1 cleavage site. Cartoon modified after Bögershausen et al. [144] (C) Sanger sequencing illustrating the *de novo* occurrence of the identified c.5462insC mutation.

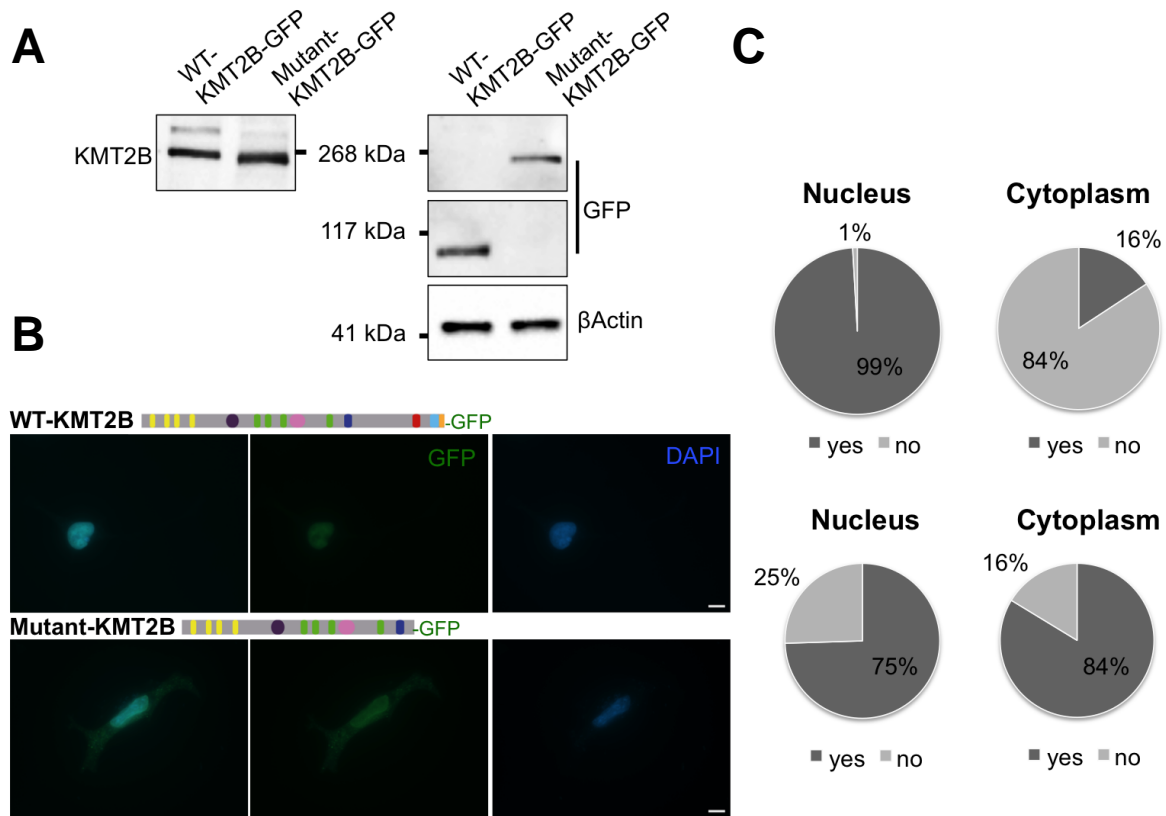
family have already been associated with developmental diseases and microcephaly is a known phenotypic feature among them [149-153]. Secondly, the identified c.5462insC mutation was not annotated in the ExAC Browser and in addition, the ExAC Browser did not contain any nonsense or frameshift mutation in *KMT2B* at all. Finally, the mutation was predicted to introduce a frameshift, which causes loss of the highly conserved SET domain pointing towards loss of protein function (Figure 8B). Based on these data, I decided to functionally investigate *KMT2B* and the identified c.5462insC mutation in the course of my PhD thesis.

### 4.2.3 Ectopically expressed *KMT2B* shows altered cellular localization

The identified c.5462insC mutation (p.Leu1822Thrs\*12) in *KMT2B* is predicted to cause a frameshift and consequently, loss of the highly conserved SET domain. *KMT2B* is initially proteolytically cleaved before N- and C-terminus dimerize via



## Results



**Figure 9: Ectopic overexpression of WT and mutant KMT2B protein.** (A) WB using an N-terminal binding KMT2B antibody and a C-terminal binding GFP antibody in HEK293T cells confirmed stability of mutant protein. WT KMT2B was cleaved, however, mutant protein was not cleaved. (B) Representative immunofluorescence image of WT and mutant KMT2B protein in COS-7 cells. (C) Quantification of the cellular localization of GFP-tagged WT and mutant KMT2B protein in COS-7 cells. At least 60 pictures of each expression construct were analyzed per transfection (n=3).

association of its FYRN and FYRC domains [163,164]. The cleavage site and the FYRC domain are located C-terminally of the identified mutation and p.Leu1822Thrfs\*12 KMT2B is therefore not cleaved (Figure 8B). However, since the FYRN domain is located N-terminally of the identified p.Leu1822Thrfs\*12 mutation and is intact, it is possible that the FYRN domain of p.Leu1822Thrfs\*12 KMT2B associates with the FYRC domain of wild-type (wt) KMT2B. In this scenario, the mutation causes a more severe dominant negative effect and not only loss of protein function (haploinsufficiency). In order to get experimental data for the responsible mutational effect, I initially cloned wt and mutant KMT2B (p.Leu1822Thrfs\*12) harboring a C-terminal GFP-tag into a mammalian expression construct (see 3.2.1.9). Overexpression of constructs in HEK293T cells and subsequent Western Blot (WB) analysis confirmed the integrity of these expression constructs and the stability of the mutant protein (Figure 9A). Use of a N-terminally binding KMT2B antibody and a C-terminally binding GFP antibody detected un-

## Results

cleaved (full length) KMT2B (294 kDa plus GFP-tag) as well as cleaved KMT2B protein (223 kDa N-terminus and 71 kDa C-terminus plus GFP-tag) in wt lysates. In contrast, only one truncated un-cleaved KMT2B protein was detectable after expression of mutant KMT2B protein (199 kDa plus GFP-tag). Next, I overexpressed both KMT2B constructs in COS-7 cells and quantified the cellular localization of the GFP signal. Figure 9B illustrates a representative immunofluorescence image. As expected, wt KMT2B was mostly localized to the nucleus (99 % of cells) and only to a small extend outside the nucleus (16 % of cells) (Figure 9C). Mutant KMT2B showed a reduced nuclear localization (75 % of cells) and an increased localization outside the nucleus (84 % of cells) indicating that the truncation of the KMT2B protein interferes with nuclear translocation.

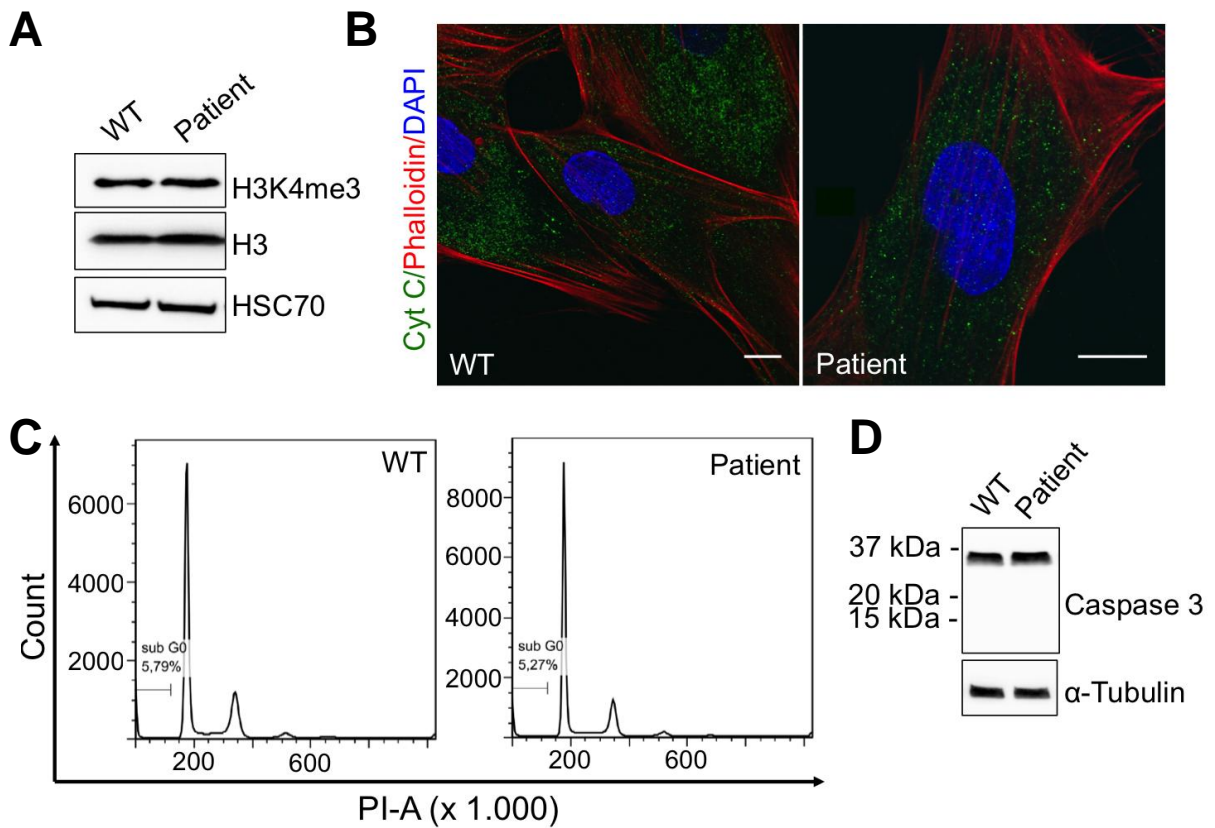
In summary, these data indicate that the ectopically expressed mutant KMT2B protein is stable and characterized by an altered cellular localization thereby leaving the possibility of a dominant negative effect. To further analyze the mutational effect *in vitro*, I aimed to investigate the cellular effects in patient-derived fibroblasts as a next step.

### **4.2.4 Investigation of patient-derived primary dermal fibroblasts**

A skin biopsy of the patient was used to culture primary dermal fibroblasts for subsequent *in vitro* experiments. Since *KMT2B* encodes a methyltransferase, I initially aimed to investigate the methylation level of H3K4 in patient-derived fibroblasts. It has already been shown that a reduction of KMT2 family members does not result in bulk changes of either H3K4me1, H3K4me2 or H3K4me3 [127,128]. In order to confirm this, whole cell lysates of wt and patient-derived primary fibroblasts were analyzed via WB with the focus on H3K4me3. Indeed, in line with the previous publications, differences in bulk levels of H3K4me3 were not detected (Figure 10A).

Since KMT2B is involved in the regulation of gene expression and mutations in *KMT2B* can therefore impact different cellular processes, the published literature was intensively searched for potential pathomechanisms underlying syndromic microcephaly. Interestingly, *in vivo* as well as *in vitro* experiments have already shown that a reduction of *KMT2B*, or its murine homolog, results in increased activation of apoptosis, a known cellular mechanism underlying the pathogenesis of microcephaly [146,170]. Ansari et al., for example, showed in *in vitro* experiments

## Results



**Figure 10: Characterization of patient-derived fibroblasts.** (A) WB for H3K4me3 illustrates that the bulk levels of H3K4me3 were not affected in patient-derived cells. H3 and HSC70 served as loading controls (n=3). (B) Immunofluorescence stainings for Cytochrome C did not detect any increased release to the cytoplasm. Actin filaments were stained with phalloidin (n=3). Scale bars=10  $\mu$ m. (C) Representative image of flow cytometry analysis focusing on the cell cycle distribution of WT and patient-derived fibroblasts with percentage of cells in subG0/1 phase. (n=3) (D) WB using a Caspase 3 antibody that recognizes full length (35 kDa) as well as activated Caspase 3 (17 and 19 kDa). Only full length Caspase 3 was detected in WT as well as patient-derived fibroblasts.  $\alpha$ -Tubulin served as loading control (n=3).

that a knock down of *KMT2B* causes release of the mitochondrial protein Cytochrome C, which is only detectable in the cytoplasm after initiation of apoptosis. Furthermore, they observed an increased population of apoptotic cells (sub-G0/1) with flow cytometry [168].

In order to analyze the effect of the identified mutation in patient-derived fibroblasts, I performed immunostainings for Cytochrome C and imaged cells with a confocal laser scanning microscope (CLSM). However, contrary to Ansari et al., no differences between wt and patient-derived cells were detectable (Figure 10B). To investigate the cell cycle distribution with flow cytometry, cells were fixed and subsequently stained with propidium iodide (PI). Figure 10C depicts one exemplary measurement and illustrates that the number of cells in sub-G0/1 phase was similar. On average 7,2 % of wt and 5,8 % of patient-derived fibroblasts were detectable in the sub-G0/1

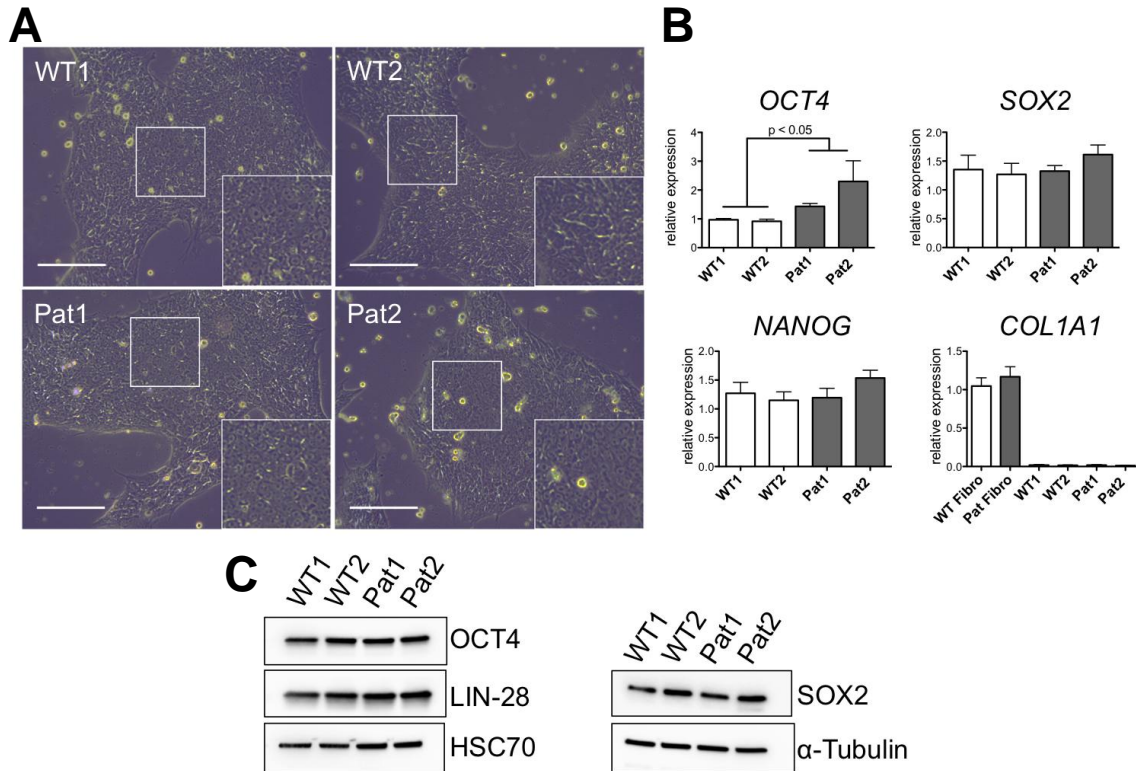
phase ( $p=0.5789$ ). Finally, whole cell lysates of wt and patient-derived fibroblasts were investigated with WB for activated Caspase 3 (17 and 19 kDa), a key player in apoptosis activation. However, similar to the previous results, only the inactive full length form of Caspase 3 (35 kDa) was detectable in patient-derived fibroblasts and not the activated form (Figure 10D).

Taken together, the presented data indicate that no activation of apoptosis was detectable in patient-derived fibroblasts. Since the identified mutation affects mainly cortical development and this at early embryonic stages, I decided to continue with the functional analysis in a different *in vitro* and more similar model and aimed to generate iPSC as a next step.

#### **4.2.5 Reprogramming primary fibroblasts to iPSC**

It has not been investigated whether KMT2B is required for reprogramming of somatic cells to iPSC to date. In collaboration with the lab of Prof. Dr. Jay Gopalakrishnan, I therefore aimed to reprogram wt and patient-derived fibroblasts to iPSC in a next step. Patient-derived cells did not show any obvious differences during the reprogramming process and first colonies, resembling iPSC morphologies, emerged at a similar time as wt colonies (data not shown). Since reprogramming is known to affect the chromosomal integrity of iPSC, an Array-CGH analysis was applied, which was performed by the laboratory diagnostic of the Institute of Human Genetics (Göttingen) [197]. Three wt and two patient-derived iPSC clones were analyzed. The analysis confirmed three chromosomal deletions present in all iPSC clones, which were not detectable in original fibroblasts indicating that reprogramming was the source for these copy-number variations (CNV) and hence, these CNV were iPSC specific (Table S2). Since these deletions did not contain any disease-associated genes or genes related to pluripotency and similar CNVs, which were related to iPSC reprogramming, have already been published, we decided to continue with two wt-derived (WT1 and WT2) and two patient-derived iPSC clones (Pat1 and Pat2) [197]. Figure 11A illustrates the typical iPSC colony morphology on matrigel coated plates. Both wt and patient-derived iPSC clones were characterized by compact cells and high nucleus to cytoplasm ratios. Subsequent quantitative PCRs (qPCRs) revealed that all iPSC clones expressed the pluripotency markers *OCT4*, *SOX2* and *NANOG*, while they did not express the fibroblast marker *COL1A1* anymore (Figure 11B) [198]. In addition, WB experiments confirmed that *OCT4*, *SOX2* and *LIN-28* proteins were detectable to similar extends, which is in line with

## Results



**Figure 11: Characterization of iPSC.** (A) Representative image of iPSC colony morphology after reprogramming wt and patient-derived fibroblasts. Two iPSC clones (WT1,WT2 and Pat1,Pat2) were investigated. scale bars=50  $\mu$ m. (B) qPCR results for expression of pluripotency markers *OCT4*, *SOX2* and *NANOG* (n=4), and fibroblast specific marker *COL1A1* (n=3). Mean  $\pm$  SD. *OCT4* expression: WT1 vs. Pat1 \*\*\*p=0.0002; WT1 vs. Pat2 \*p=0.0102; WT2 vs. Pat1 \*\*\*p=0.0002; WT2 vs. Pat2 \*\*p=0.0087. unpaired t-test. Expression levels were normalized to *GAPDH*. (C) WB analysis for the pluripotency marker *OCT4*, *LIN-28* and *SOX2*. *HSC70* and  $\alpha$ -Tubulin served as loading control (n=5).

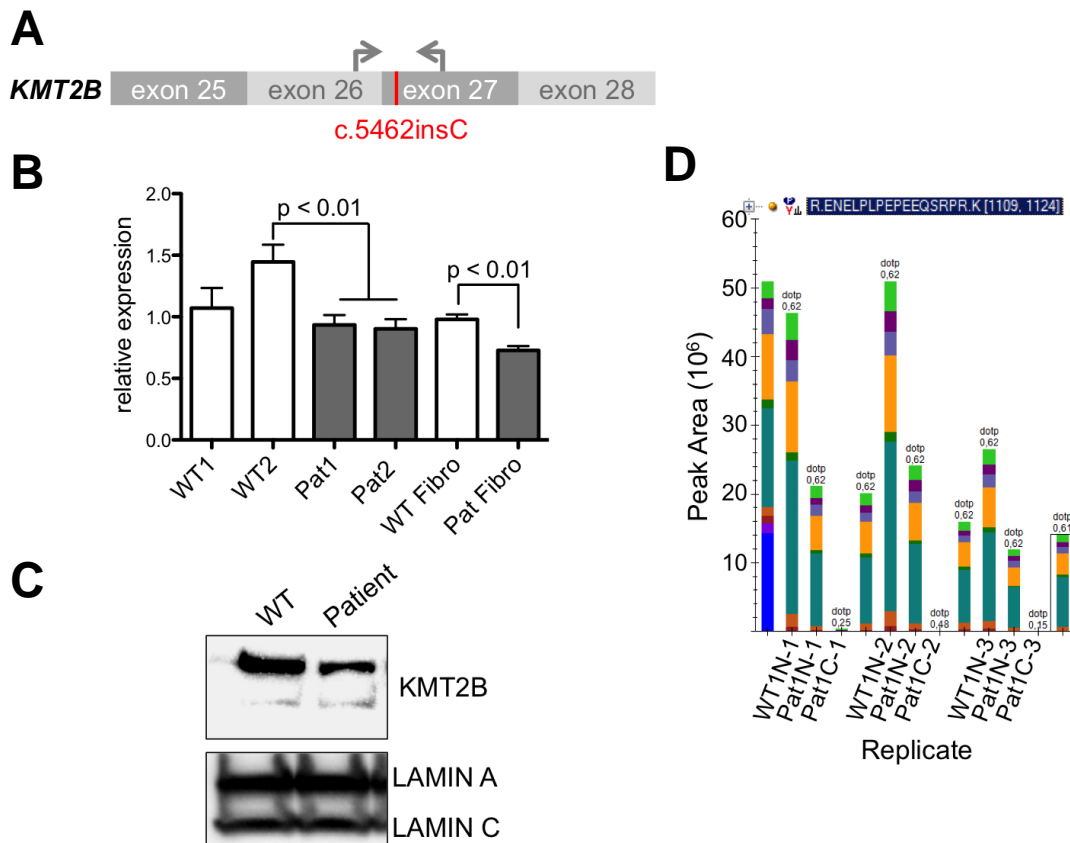
previous publications confirming that expression of pluripotency markers is not down-regulated in *Kmt2b* KO ESCs (Figure 11C) [137,170].

Taken together, these data show that patient-derived fibroblasts were successfully reprogrammed to iPSC indicating that reduced expression of *KMT2B*, due to the identified mutation, did not result in obvious reprogramming impairments. In a next step, patient-derived fibroblasts and patient-derived iPSC were used for a detailed characterization of the identified c.5462insC mutation.

### 4.2.6 Investigation of endogenous *KMT2B* in patient-derived fibroblasts and iPSC

Initially, in order to investigate the expression and transcript stability of endogenous *KMT2B* in patient-derived fibroblasts and iPSC, qPCRs were performed with primers spanning the site of the c.5462insC mutation in exon 27 (Figure 12A). In patient-derived fibroblasts, the expression of *KMT2B* was significantly reduced compared to wt cells by 26 % (\*\*p=0.0012) (Figure 12B). By contrast, a significant reduction of

## Results



**Figure 12: Investigation of endogenous KMT2B.** (A) Position of primers used for qPCR as well as for the PCR. PCR products were subsequently subjected to deep amplicon sequencing. (B) qPCR results of *KMT2B* expression in wt and patient-derived fibroblasts and iPSC (WT1,WT2 and Pat1,Pat2). Mean  $\pm$  SD. WTFibro vs. PatFibro  $**p=0.0012$ ; WT1 vs. Pat1  $p=0.2659$ ; WT1 vs. Pat2  $p=0.1855$ ; WT2 vs. Pat1  $**p=0.0055$ ; WT2 vs. Pat2  $**p=0.0043$  (unpaired t-test). WT1 vs. WT2  $p=0.1600$ ; Pat1 vs. Pat2  $p=0.7414$  (paired t-test) ( $n=3$ ). (C) WB for KMT2B using nuclear lysates of wt and patient-derived fibroblasts. LAMIN A/C served as loading control. ( $n=3$ ) (D) Representative image of one KMT2B specific peptide identified using skyline software in nuclear lysates of WT1 and Pat1, and cytosolic lysates of Pat1. Peak area determines the relative amount of the identified peptide and indicates that KMT2B was reduced in the nuclear fraction of Pat1, whereas KMT2B was not detectable in the cytoplasmic fraction. ( $N=3$ ;  $n=1$ )

*KMT2B* in patient-derived iPSC clones was only detectable in comparison to WT2 (WT2 vs. Pat1  $**p=0.0055$ ; WT2 vs. Pat2  $**p=0.0043$ ). A comparison of both patient-derived iPSC clones with WT1 revealed only slight reductions of 7 % and 10 % (WT1 vs. Pat1  $p=0.2659$ ; WT1 vs. Pat2  $p=0.1855$ ) indicating that the *KMT2B* expression was not consistently reduced in patient-derived iPSC clones and instead, similar to wt cells. Since these qPCR data did not reveal any information about the quantity of wt and c.5462insC transcripts, one PCR product of each iPSC clone was subjected to NGS-based deep amplicon sequencing, performed by the CCG. Using this approach, the total number of reads surrounding the mutation site ranged from  $4.7 \times 10^6$  to  $8.3 \times 10^6$  in WT1 and WT2, respectively, and  $4.0 \times 10^6$  to  $4.1 \times 10^6$  in Pat1 and Pat2, respectively. Both wt samples showed the nucleotide adenine at cDNA position 5462 with a frequency of 99.5 % ( $4.65 \times 10^6$  reads in WT1 and  $8.26 \times 10^6$  reads in WT2), while the missing 0.5 % were distributed in equal amounts to the remaining

## Results

nucleotides indicating artifact background. In addition, the cytosine insertion was not detectable. In contrast, the c.5462insC mutation was detected at a frequency of 26.4 % and 27.5 % in Pat1 and Pat2 ( $1.05 \times 10^6$  reads in Pat1 and  $1.14 \times 10^6$  reads in Pat2), respectively, indicating that the c.5462insC transcript was stable and enabling the possibility for translation of the p.Leu1822Thrfs\*12 KMT2B protein.

Next, I analyzed the stability of wt and c.5462insC KMT2B on protein level. Therefore, I used nuclear protein lysates of wt and patient-derived fibroblasts and performed WB using a N-terminal binding KMT2B antibody. Figure 12C illustrates the reduction of KMT2B in patient-derived fibroblasts compared to wt cells of approximately 50 %. Since the p.Leu1822Thrfs\*12 KMT2B protein was predicted to have roughly the molecular weight of the cleaved N-terminal part of wt KMT2B, it should be detectable in the WB, if present. However, no aberrant protein band in the patient sample was observable indicating that the p.Leu1822Thrfs\*12 protein was not able to translocate to the nucleus, which is in line with the previous results in 4.2.3.

Since overexpression of mutant KMT2B protein showed an increased aberrant localization outside the nucleus (see 4.2.3.), I aimed to investigate the cytoplasmic fraction of protein cell lysates in more detail. Cytoplasmic as well as nuclear fractions of patient-derived iPSC clones were analyzed with mass spectrometry by the proteomics service facility of the University Medical Center Göttingen (UMG). For this, I separated cell lysates under reducing conditions with gel electrophoresis. Gel bands were subsequently excised, purified and spiked with specific amounts of Cytochrome C serving as an internal protein reference. Targeted mass spectrometry using the parallel reaction monitoring technique (PRM) was used in order to detect the specific peptides of the KMT2B protein. Three technical replicates were performed. Amounts of detected peptides were subsequently analyzed with skyline and scaffold software. With this, different KMT2B peptides located N-terminally of the

**Table 12: Mass spectrometry results using scaffold software of three technical replicates (N1-N3).** Relative amounts of all detected KMT2B peptides in Total Spectrum Count [TSC]. WT1 and Pat1 were analyzed first and subsequently WT2 and Pat2 resulting in different relative amounts. Meas.: Measurement

Meas.	Sample	cytoplasmic fraction			nuclear fraction [TSC]		
		N1	N2	N3	N1	N2	N3
1	WT1	n.a.	n.a.	n.a.	163	279	186
	Pat1	n.d.	n.d.	n.d.	78	115	94
2	WT2	n.a.	n.a.	n.a.	6	5	1
	Pat2	n.d.	n.d.	n.d.	4	1	0

n.a.: not applied, n.d.: not detected



## Results

p.Leu1822Thrfs\*12 mutation were identified. Analysis with skyline and scaffold revealed that KMT2B was not detectable in the cytoplasmic fraction of patient samples and only with reduced amounts in the nuclear fraction, namely 45,7 % in the first and 42 % in the second measurement compared to wt samples (Table 12 and Figure 12D).

In summary, the presented data reveal that even though the c.5462insC transcript was detectable, the corresponding p.Leu1822Thrfs\*12 KMT2B protein was not, which indicates that the p.Leu1822Thrfs\*12 protein is likely unstable and hence, favor loss of KMT2B function as disease underlying effect.

### 4.2.7 Analyses of RNAseq data in iPSC

Since KMT2B is a methyltransferase and therefore important for the regulation of gene expression during embryonic developmental stages, gene expression was investigated in patient-derived iPSC in more detail. RNA samples of each wt and patient-derived iPSC clone were extracted at passage 19 and subjected to RNA-sequencing (RNAseq) by Poly(A) selection. RNAseq as well as pre-processing of data ( $\log_2\text{foldchange} \geq 1$  or  $\geq -1$ ;  $p\text{-value} < 0.05$ ) was performed by the transcriptome and genome analysis laboratory (TAL) in Göttingen. Table 13 summarizes the number of differentially expressed (DE) genes and the amount of up- and down-regulated genes after single comparison of each wt-derived iPSC clone with each patient-derived iPSC clone.

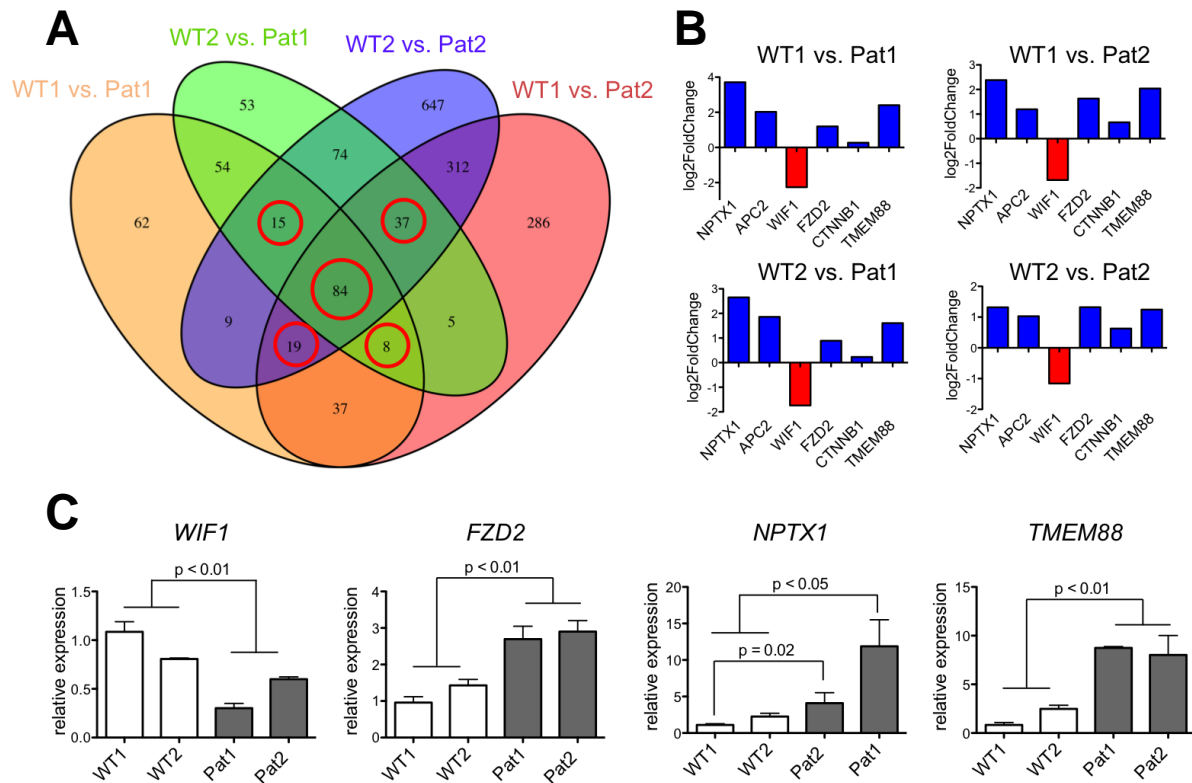
Table 13: Number of DE genes after RNAseq in iPSC and comparison of each wt-derived clone with each patient-derived clone.

Comparison	Number of DE genes	Number of up-regulated genes	Number of down-regulated genes
WT1vs.Pat1	288	189	99
WT1vs.Pat2	788	214	574
WT2vs.Pat1	330	155	175
WT2vs.Pat2	1197	271	926
Shared by all comparisons	84	56	28

In three out of four comparisons, more genes were down- than up-regulated, which is in line with Kmt2b as the main H3K4me3 modifier at promoter regions in stem cells and H3K4me3 being mainly associated with active gene expression [128,199]. The number of DE genes was drastically reduced when candidate lists were screened for



## Results



**Figure 13: RNAseq in iPSC. (A)** Venn diagram depicting the number of differentially expressed genes after single comparison of each wt-derived iPSC clone with each patient-derived iPSC clone. Red circles highlight genes, which were analyzed with gene set enrichment analysis. **(B)** Log2foldChanges of identified differentially expressed genes. **(C)** qPCRs in order to confirm differential expression of identified genes. *WIF1*: WT1 vs. Pat1 \*\*\* $p=0.0003$ ; WT1 vs. Pat2 \*\*\* $p<0.0001$ ; WT2 vs. Pat1 \*\* $p=0.0013$ ; WT2 vs. Pat2 \*\*\* $p=0.0001$ . *FZD2*: WT1 vs. Pat1 \*\* $p=0.0015$ ; WT1 vs. Pat2 \*\*\* $p=0.0006$ ; WT2 vs. Pat1 \*\* $p=0.0048$ ; WT2 vs. Pat2 \*\* $p=0.0017$ . *NPTX1*: WT1 vs. Pat1 \*\* $p=0.0069$ ; WT1 vs. Pat2 \* $p=0.0235$ ; WT2 vs. Pat1 \* $p=0.0104$ ; WT2 vs. Pat2  $p=0.1027$ . *TMEM88*: WT1 vs. Pat1 \*\*\* $p<0.0001$ ; WT1 vs. Pat2 \*\* $p=0.0036$ ; WT2 vs. Pat1 \*\*\* $p<0.0001$ ; WT2 vs. Pat2 \*\* $p=0.0094$ . (n=3; Mean  $\pm$  SD; unpaired t-test)

genes shared by all four comparisons. Then, only 84 genes remained consisting of 56 up- and 28 down-regulated ones (Table 13, Figure 13A). To get a better insight into the function of these shared DE genes, I performed a gene set enrichment analysis using top gene ontology (GO) in R [180]. Different GO pathways were enriched in GO categories “molecular function” and “biological process” (Table S3). Among involved genes were two genes particularly interesting, namely *NPTX1* and *APC2* (Figure 13B). *NPTX1* is an important player for the differentiation of human pluripotent stem cells into the neural lineage [200] and similarly, *APC2* is involved in WNT-signaling, a well known pathway also important for differentiation of stem cells [201-203]. Re-investigation of data showed that beside *APC2* also *CTNNB1* encoding  $\beta$ -Catenin and the WNT-signaling regulator *TMEM88* were increased in each comparison as well (Figure 13B) [204]. An impaired differentiation as potential

## Results

pathomechanism in the present patient is in line with the data of Singh et al., who showed that *KMT2B* is important for the initiation of differentiation in human pluripotent stem cells [166].

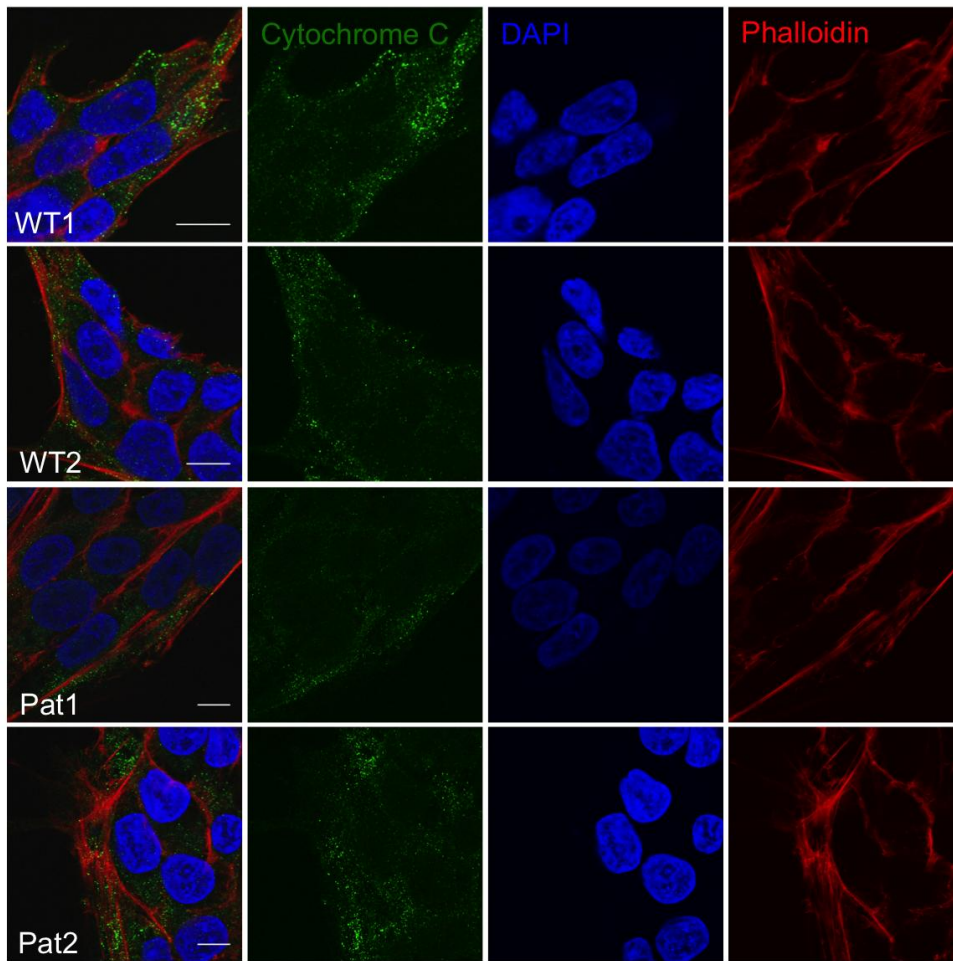
In a next step, I also included DE genes, which were shared by three out of the four comparisons for the gene set enrichment analysis (Figure 13A and Table S4). Using this strategy, more WNT-signaling-associated genes were identified, namely *WIF1* and *FZD2* as well (Figure 13B) [205]. qPCR confirmed that *FZD2* and *TMEM88* were both up- and *WIF1* down-regulated in patient-derived iPSC clones (Figure 13C). Contrary, the significant up-regulation of *NPTX1* in patient samples was not consistent among iPSC clones, even though the trend was apparent (WT1 vs. Pat1 \*\*p=0.0069; WT1 vs. Pat2 \*p=0.0235; WT2 vs. Pat1 \*p=0.0104; WT2 vs. Pat2 p=0.1027). Similarly, an up-regulation of *CTNNB1* and *APC2* was not consistent among iPSC clones (Figure S2).

Since the RNAseq data of patient-derived iPSC clones showed a high variance, which prevented further analyses, I aimed to investigate the RNAseq data of one comparison in more detail (Figure S3). For this, I re-analyzed, with bioinformatician Dr. Alexander Wolff, the data of “WT2 vs. Pat1”. This comparison was chosen because firstly, the biological variance between these samples was low and secondly, RNAseq data of three biological replicates from each iPSC clone enabled a more robust data analysis than analysis of clones with only two biological replicates (Figure S3). An over-representation analysis (ORA) for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was performed with a false discovery rate (FDR) of less than 5 % [206]. This analysis revealed that 1,172 genes were differentially expressed in Pat1 compared to WT2. The top 30 of called KEGG pathways listed the

Table 14: KEGG pathway enrichment of comparison “WT2 vs. Pat1”: DE genes in apoptosis and WNT-signaling pathway. bold letters indicate an up-regulation of genes and thinner letters indicate a down-regulation.

KEGG pathway	DE genes
Apoptosis	<b>TRP53/BAX/CTSF/TNFRSF10B/CASP9/PTPN13/APAF1/CFLAR/MAPK3/CTSC/BBC3/FOS/ACTB/TUBA1A/RIPK1/CTSB/ATF4/BID/KRAS/PIK3CD/FAS/PIDD1/CTSH/PDPK1/CASP6/ITPR3/NFKBIA/TRAF1/AKT3/PMAIP1/PIK3R1/BCL2L1/DFFA/ITPR1/CAPN1/NRAS</b>
WNT-signaling	<b>TRP53/APC2/SFRP1/CCND1/FZD5/FZD7/PRICKLE1/FZD2/FOSL1/FRAT2/DAAM2/GPC4/ROR2/FZD8/PRKCA/WIF1/ROR1/LGR4/SMAD4/SHAH1/RSP04/SFRP5/APC/CTNNB1/NFATC1/NKD2/LRP6/DAAM1/FZD3/TCF7L1/PRKACB/TBL1XR1/SMAD3/PPP3R1/CTBP1/FBXW11/WNT8A/PLCB4/PRICKLE2/SERPINF1</b>

## Results



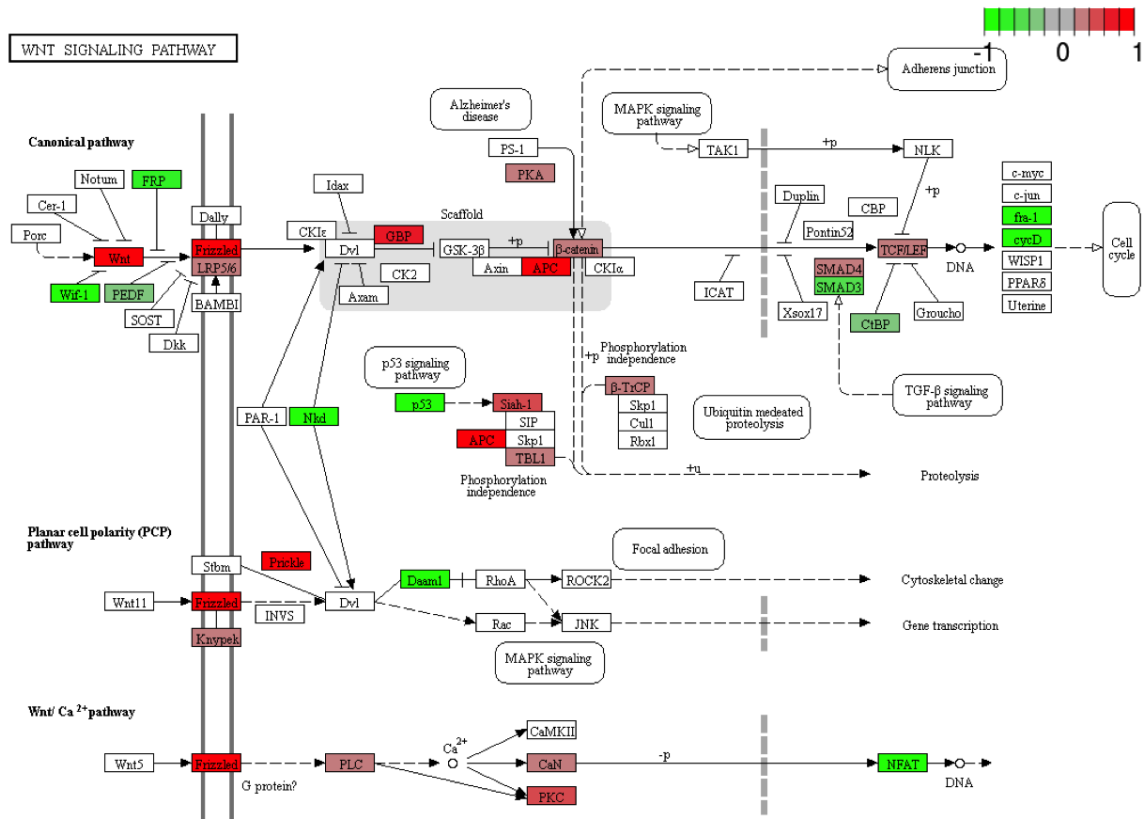
**Figure 14: Immunocytochemistry for Cytochrome C.** No increased release of Cytochrome C to the cytoplasm was detectable in wild-type (WT) and patient-derived iPSC. Actin filaments were stained with phalloidin (n=3). Scale bars=10  $\mu$ m.

apoptosis and WNT-signaling pathway (Table S5). The apoptosis pathway consisted of in total 116 genes. Of these 116 genes were 36 differentially expressed in Pat1 and more precisely, 19 genes were up- and 17 down-regulated (Table 14).

Amongst these DE genes, the pro-apoptic gene *TRP53* was down- and the anti-apoptic gene *BCL2L1* up-regulated [207]. p53 is a master regulator for several cellular processes, including the induction of apoptosis, and has been shown to be involved in the release of Cytochrome C [208]. In line with a down-regulation of *TRP53* in Pat1, immunostainings for Cytochrome C in patient-derived iPSC clones did not show any aberrant apoptosis activation, which is illustrated in Figure 14.

The WNT-signaling pathway in the KEGG analysis consists of 134 genes, which are either associated with canonical or non-canonical WNT-signaling. 40 genes were differentially expressed in Pat1 compared to WT2, namely, 27 genes were up- and

## Results



**Figure 15: KEGG pathway analysis.** Several identified differentially expressed genes in “WT2 vs. Pat1” were associated with WNT-signaling pathway. Green=down-regulation. Red=up-regulation.

13 genes were down-regulated (Figure 15 and Table 14). *WNT8A*, *CTNNB1* as well as different Frizzled class receptor genes (*FZD2*, *FZD3*, *FZD5*, *FZD7* and *FZD8*) were consistently up-regulated, while WNT-signaling inhibitors, like *WIF1* and *NKD2*, were down-regulated. In addition, effectors of WNT-signaling including *CCND1*, encoding Cyclin D1 and the transcription factors *NFATC1* as well as *FOSL1* (*FRA1*) were down-regulated, which further strengthens a putative involvement of WNT-signaling in the underlying pathomechanism of our patient.

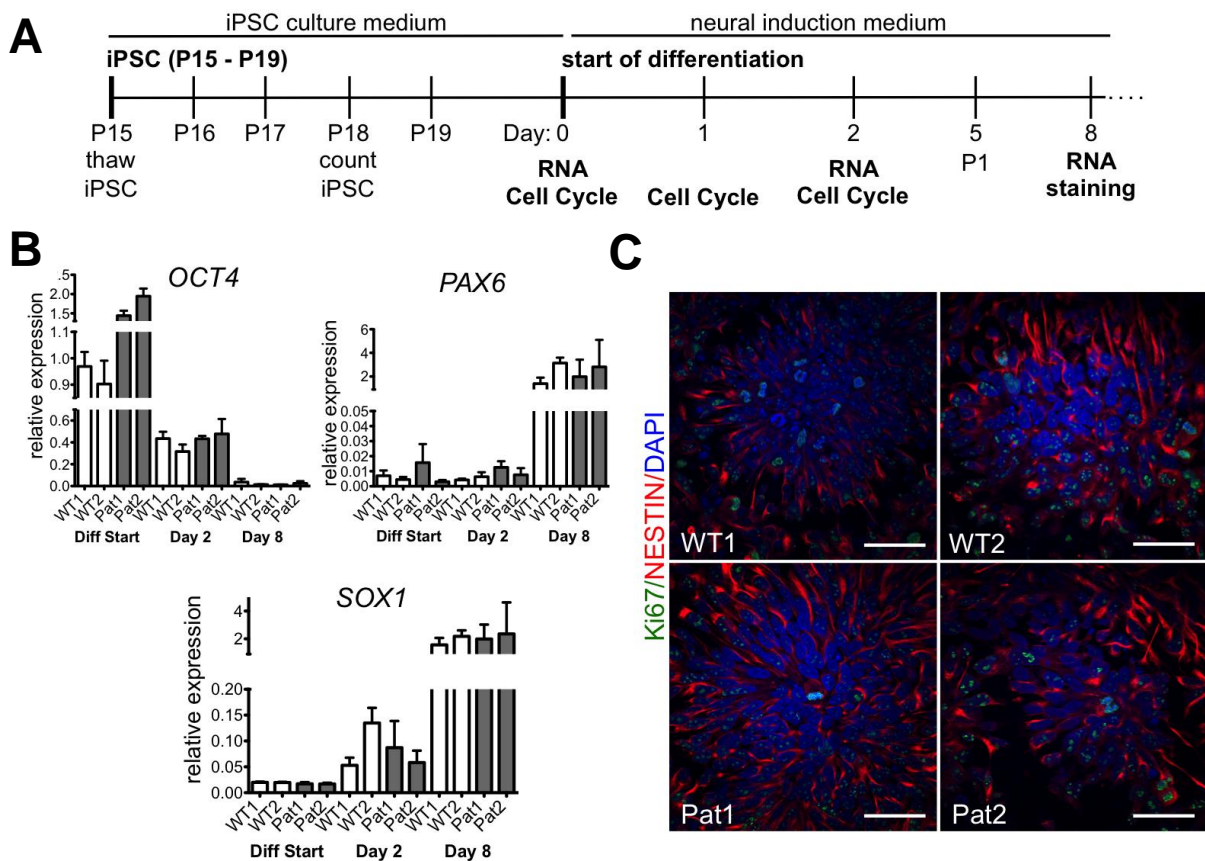
Taken together, the presented RNAseq data give first hints for an impaired differentiation ability of patient-derived iPSC. In order to investigate the differentiation ability of patient-derived iPSC into the neural lineage in more detail, I differentiated iPSC to neural progenitor cells (NPCs) in a next step.

### 4.2.8 Differentiation of iPSC into neural progenitor cells (NPCs)

In order to investigate the differentiation ability of patient-derived iPSC, I differentiated all iPSC clones to early passages of NPCs, in collaboration with the stem cell unit (SCU) at the UMG. Figure 16A summarizes the experimental design

## Results

with different time points of sample collection. Initially, RNA was extracted at three days (day 0, day 2 and day 8) to trace the course of gene expression from initiation of differentiation to passage 1 (day 8) of NPCs. The quantification of the pluripotency marker *OCT4* and NPC markers *PAX6* as well as *SOX1* with qPCR showed that the course of gene expression was similar among all iPSC clones (Figure 16B). While the expression of *OCT4* clearly decreased over time, the expression of *SOX1* and *PAX6* increased to similar extends. In line with this, immunostainings for the NPC marker NESTIN and the proliferation marker Ki67 in passage 1 of NPCs indicated that cells were still proliferating and characterized by expression of NESTIN protein (Figure 16C). In addition, all cell lines showed similar cell morphologies in passage 1 and lost the compact structure of the former iPSC colonies (Figure 17B).



**Figure 16: Differentiation of iPSC into passage 1 of NPCs.** (A) Experimental design for differentiation of iPSC into NPCs. iPSC were thawed at passage 15 (P15). At P18, iPSC were counted prior to passaging. 48 h later, a monolayer protocol was applied and different analyses were performed at specific time points. (B) Time course of *OCT4*, *SOX1* and *PAX6* expression during differentiation. (n=3) (C) Immunofluorescence stainings for NESTIN and proliferation marker Ki67 at day 8 of differentiation into NPCs. Scale bars=50 μm. (n=3).

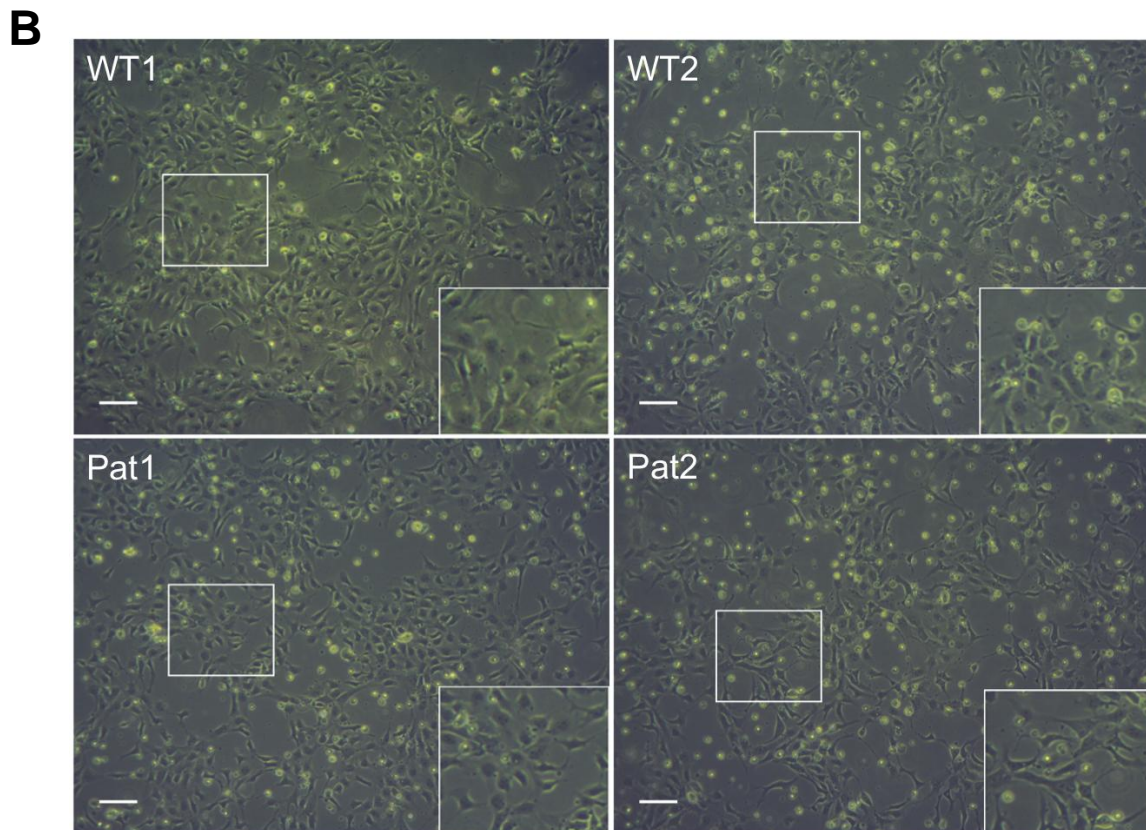
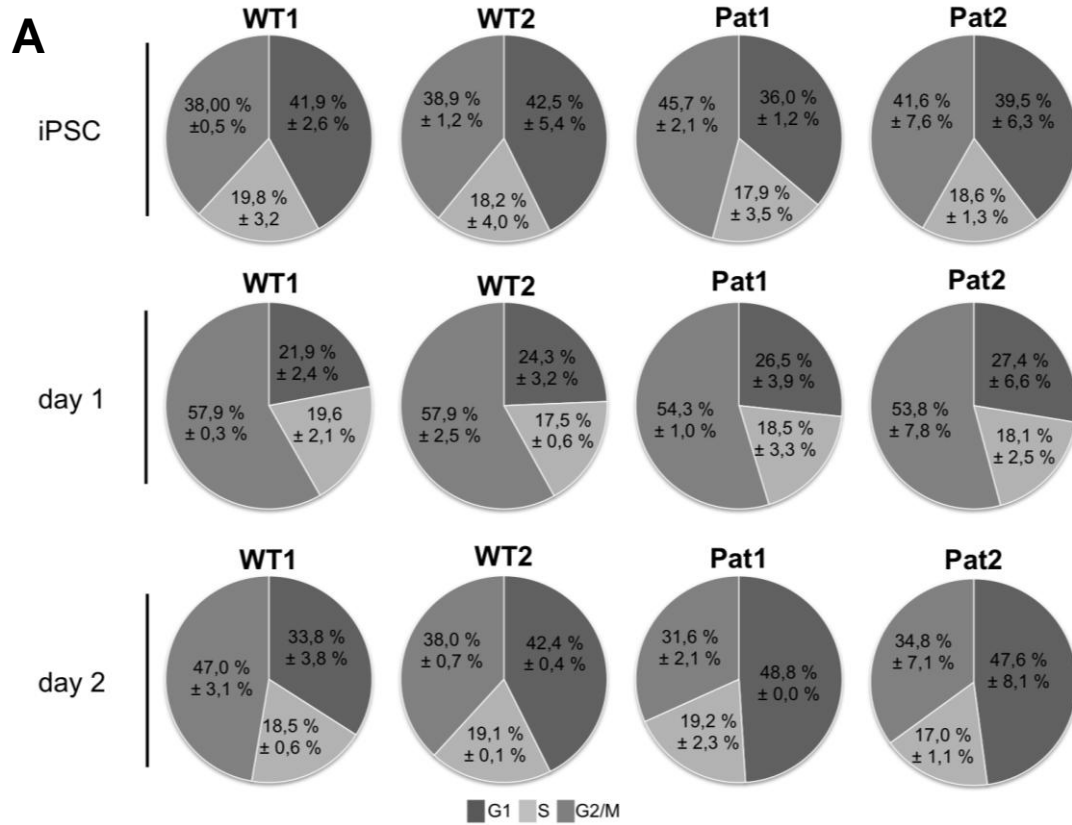
## Results

Since it has already been shown that the differentiation of stem cells, regulated by KMT2B, is linked to the cell cycle, I also investigated the distribution of cells in the cell cycle during first days of differentiation (day 0-2) with flow cytometry [166] (Figure S4). Figure 17A illustrates mean values ( $\pm$ SD) of cells either in G1, S or G2/M phase of the cell cycle from at least two independent experiments. These measurements revealed that the cell cycle distribution in iPSC was similar among all iPSC clones with most cells being either in G1 or G2/M phase. At day 1, the number of cells in G2/M phase increased in iPSC clones to more than 53 % and subsequently, decreased at day 2 in WT2, Pat1 and Pat2 to similar extends. However, only WT1 showed still a higher number of cells in G2/M phase at day 2, which is in line with a deviating behavior of WT1 during differentiation that was observed. More specifically, WT1 had always lowest cell confluencies at day 1, even though same cell numbers were seeded and hence, appeared to proliferate more slowly during the next days (Figure S5, data not shown).

In summary, the presented data show that patient-derived iPSC clones showed normal expression rates of important NPC markers during the first days of differentiation pointing towards a regular ability of these cells for the initial onset of differentiation. Even if, iPSC clone WT1 showed an aberrant cell cycle distribution at day 2, the comparison of iPSC clone WT2 with patient-derived iPSC clones does not reveal any obvious cell cycle differences during the first days of differentiation.



## Results



**Figure 17: Investigating iPSC and differentiating cells. (A)** Pie chart illustrating iPSC and differentiating cells, which were fixed, stained with PI and subsequently analyzed with cell cytometry. G1 phase (dark grey); S phase (middle grey); G2/M phase (light grey). Mean in percentage  $\pm$  SD. iPSC: n=2; day 1: n=3; day 2: n=2. **(B)** Representative image of NPC morphology at P1. Scale bars=50  $\mu$ m.

#### 4.2.9 CRISPR/Cas9-mediated Knock Out of *kmt2ba* and *kmt2bb* in zebrafish

Since patient-derived iPSC did not show any obvious differences during early stages of differentiation into NPCs *in vitro*, I next aimed to investigate the effect of the identified c.5462insC mutation on tissue level in a homozygous *KMT2B* KO. For this, I intended to establish an *in vivo* KO model and aimed to prepare expression constructs, which can be used for subsequent rescue experiments.

The zebrafish *Danio rerio* is a common model system for the investigation of microcephaly [68,69,209]. *KMT2B* has two predicted homologs in zebrafish, namely *kmt2ba* and *kmt2bb*, which have not been extensively investigated yet (Figure S6). Maternal RNA of both homologs is present already at early stages of embryogenesis pointing towards important functions during developmental [72]. In order to knock out both homologs of *KMT2B*, I aimed to use the CRISPR/Cas9 approach and initially established a positive control targeting the *tyrosinase (tyr)* gene in the lab of PD Dr. Roland Dosch [183]. A successful KO of *tyr* causes loss of pigmentation in injected embryos visible with the naked eye (Figure S7). This positive control was therefore a reliable and fast marker to confirm the functional quality of reagents and correct handling during all performed injections.

In order to mimic the human c.5462insC mutation in zebrafish, I aligned the coding sequences of *kmt2ba* and *kmt2bb* with *KMT2B* and designed appropriate single-guide RNA (sgRNA) sequences to target both homologs in close proximity to the identified human c.5462insC mutation. Cas9 RNA and sgRNAs targeting either

Table 15: Position and protein effect of different mutations in human *KMT2B* and zebrafish homologs.

gene	exon	cDNA position	protein effect
<i>KMT2B</i>	27	c.5462insC	p.Leu1822Thrfs*12
<i>kmt2ba</i>	28	c.5192_5199delGGACTCCT	p.Thr1732Valfs*33
<i>kmt2bb</i>	28	c.7020_7024delCGACC	p.Asp2341Trpfs*40

*kmt2ba*, *kmt2bb* or *tyr* were injected into one cell stages of wt hybrids (ABxTLF). Subsequently, injected F0 fish were genotyped and positive mosaic fish were crossed with ABxTLF (Figure S7C). The F1 generation was screened for heterozygous candidate fish with either insertions or deletions predicted to cause a premature truncation of *kmt2ba* or *kmt2bb* protein. Table 15 summarizes the position of mutations in *kmt2ba* and *kmt2bb*, which were finally selected. In order to increase



## Results

the number of heterozygous mutants (*kmt2ba*<sup>+/-</sup> and *kmt2bb*<sup>+/-</sup>), positive fish were crossed with ABxTLF for a second time.

### 4.2.10 Detection of microcephaly in a *kmt2bb* knock out model

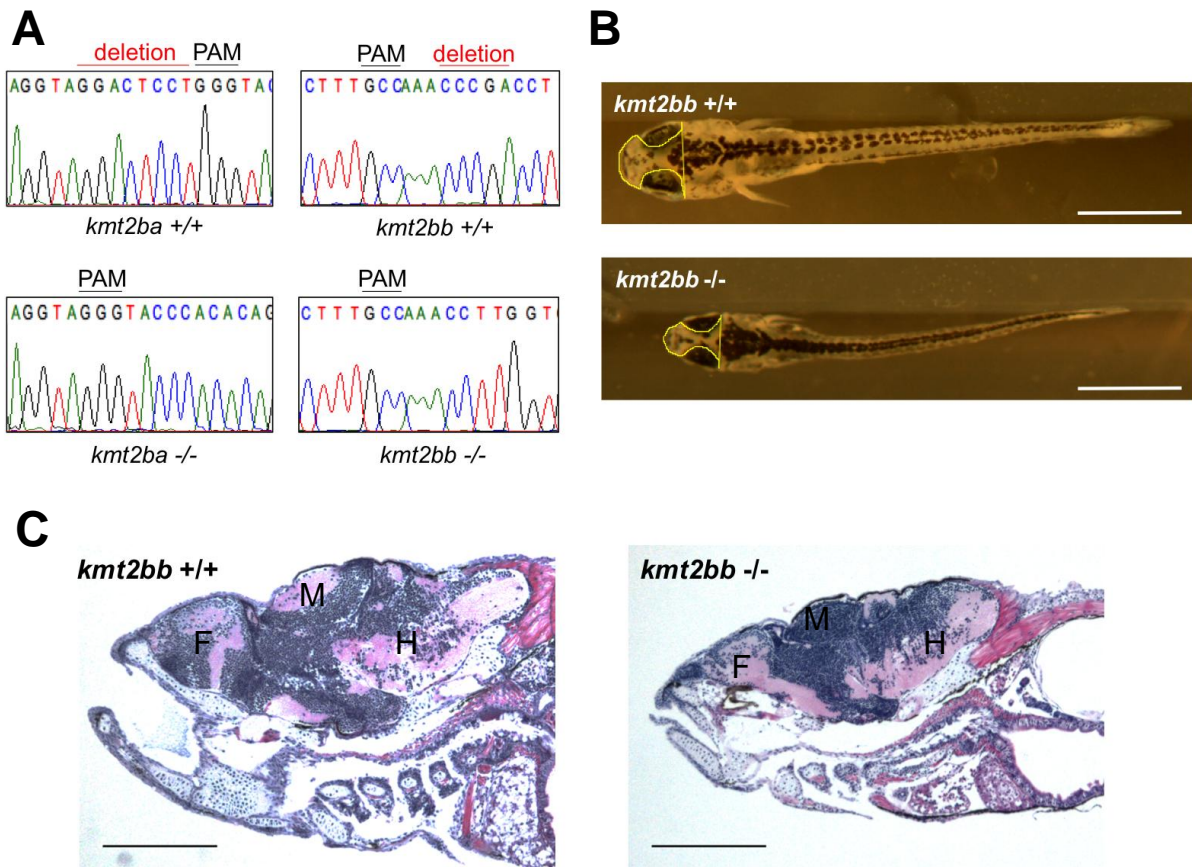
In order to receive homozygous KO fish of each genotype, I crossed adult *kmt2ba*<sup>+/-</sup> fish with *kmt2ba*<sup>+/-</sup> fish and *kmt2bb*<sup>+/-</sup> fish with *kmt2bb*<sup>+/-</sup> fish. Subsequent Sanger sequencing confirmed the expected mendelian pattern of inheritance in the next generation (Figure 18A and Table 16). *kmt2ba* and *kmt2bb* KO embryos were both viable and showed no drastic phenotypic abnormalities within the first five days post fertilization (dpf).

Table 16: Mendelian ratios of wt, heterozygous and homozygous *kmt2ba* as well as *kmt2bb* KO fish after crossing *kmt2ba*<sup>+/-</sup> with *kmt2ba*<sup>+/-</sup> and *kmt2bb*<sup>+/-</sup> with *kmt2bb*<sup>+/-</sup>.

	<i>kmt2ba</i> <sup>+/-</sup> x <i>kmt2ba</i> <sup>+/-</sup>	<i>kmt2bb</i> <sup>+/-</sup> x <i>kmt2bb</i> <sup>+/-</sup>
WT (+/+)	39 (24,8 %)	51 (21,3 %)
heterozygous KO (+/-)	78 (49,7 %)	130 (54,4 %)
homozygous KO (-/-)	40 (25,5 %)	58 (24,3 %)
Total number	157	239

Since the presented patient suffered from a syndrome characterized by features including a reduced head circumference and reduced body length (see 4.2.1), I initially only quantified the head area of fish between 1 dpf and 5 dpf (Figure 18B). No obvious differences were detectable by 5 dpf, neither for *kmt2ba* nor for *kmt2bb* (Figure 19). A closer look, however, revealed that for *kmt2bb* a trend towards a reduced head area was obvious at 5 dpf, even though the reduction was not significant yet (WT vs. KO p=0.06). Therefore, I decided to investigate the head area also at later developmental stages and subjected 14 as well as 21 dpf fish for measurements. Subsequent quantifications revealed that *kmt2ba* KO showed still no significant differences at any investigated time point (Figure 19). In contrast, homozygous *kmt2bb* KO larvae had significant reduced head areas at 14 as well as 21 dpf, and furthermore, the reduction was progressive (mean relative head area 5 dpf=0.90; mean relative head area 14 dpf=0.77; mean relative head area 21 dpf=0.68) (Figure 19). Heterozygous *kmt2bb* KO fish showed only significant differences compared to wt fish at 14 but not at 21 dpf (14 dpf \*\*p=0.0045, 21 dpf p=0.8564). In line with the phenotype of the patient, the total size of *kmt2bb* KO fish

## Results

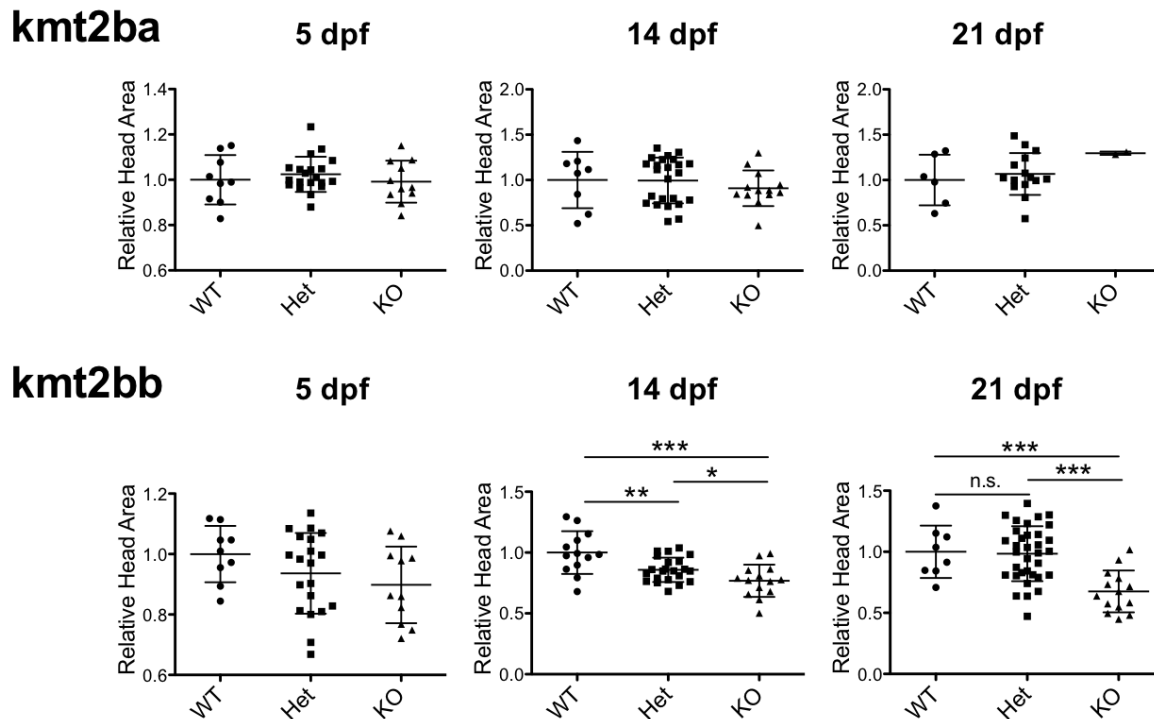


**Figure 18: CRISPR/Cas9-based KO in zebrafish.** (A) Electropherogram depicting the position of PAM sequences and generated deletions in wild-type (+/+) and homozygous *kmt2ba* as well as *kmt2bb* KO (-/-) fish. (B) Complete overview of *kmt2bb* (+/+) and *kmt2bb* (-/-) fish from dorsal perspective at 14 dpf. Yellow line highlights area, which was measured in order to estimate head area of fish. Scale bars=2 mm. (C) Sagittal sections of *kmt2bb* (+/+) and *kmt2bb* (-/-) fish after Hematoxylin and eosin (HE) stain at 12 dpf. Forebrain (F); Midbrain (M); Hindbrain (H). scale bars= 100  $\mu$ m.

as well as the brain size were reduced, while the overall structure of the brain appeared normal (Figures 18B/C).

In order to prove that loss of KMT2B underlies the observed syndromic microcephaly in *kmt2bb* KO fish, I cloned wt KMT2B into the pCS2+ expression vector, which can be used for subsequent rescue experiments. In addition, I prepared four KMT2B expression constructs with varying mutations, in order to characterize the relevance of different protein domains for the rescue of the underlying phenotype in zebrafish (Figures S8 and S9). Two expression constructs have missense mutations in the SET domain of KMT2B (p.Tyr2604Ala and p.Glu2618Lys). The p.Tyr2604Ala mutation abolishes the methyltransferase activity of KMT2B, whereas the p.Glu2618Lys mutation is an unpublished mutation identified in a patient with syndromic microcephaly (data not shown) [137]. Two additional KMT2B constructs express truncated forms of KMT2B including the p.Leu1822Thrfs\*12 mutation, which was identified in the patient presented in this PhD thesis and truncates KMT2B after

## Results



**Figure 19: Microcephaly in homozygous *kmt2bb* KO larvae.** Relative head area was investigated at 5 dpf, 14 dpf and 21 dpf after crossing *kmt2ba*<sup>+/-</sup> x *kmt2ba*<sup>+/-</sup> as well as *kmt2bb*<sup>+/-</sup> x *kmt2bb*<sup>+/-</sup> fish. Genotypes were subsequently determined via Sanger sequencing. *kmt2bb* 5 dpf (total n=40): WT vs. Het p=0.2088; WT vs. KO p=0.0600; Het vs. KO p=0.4417. *kmt2bb* 14 dpf (total n=49): WT vs. Het \*\*p=0.0045; WT vs. KO \*\*\*p=0.0007; Het vs. KO \*p=0.0262. *kmt2bb* 21 dpf (total n=56): WT vs. Het p=0.8564; WT vs. KO \*\*\*p=0.0009; Het vs. KO \*\*\*p<0.0001. (unpaired t-test)

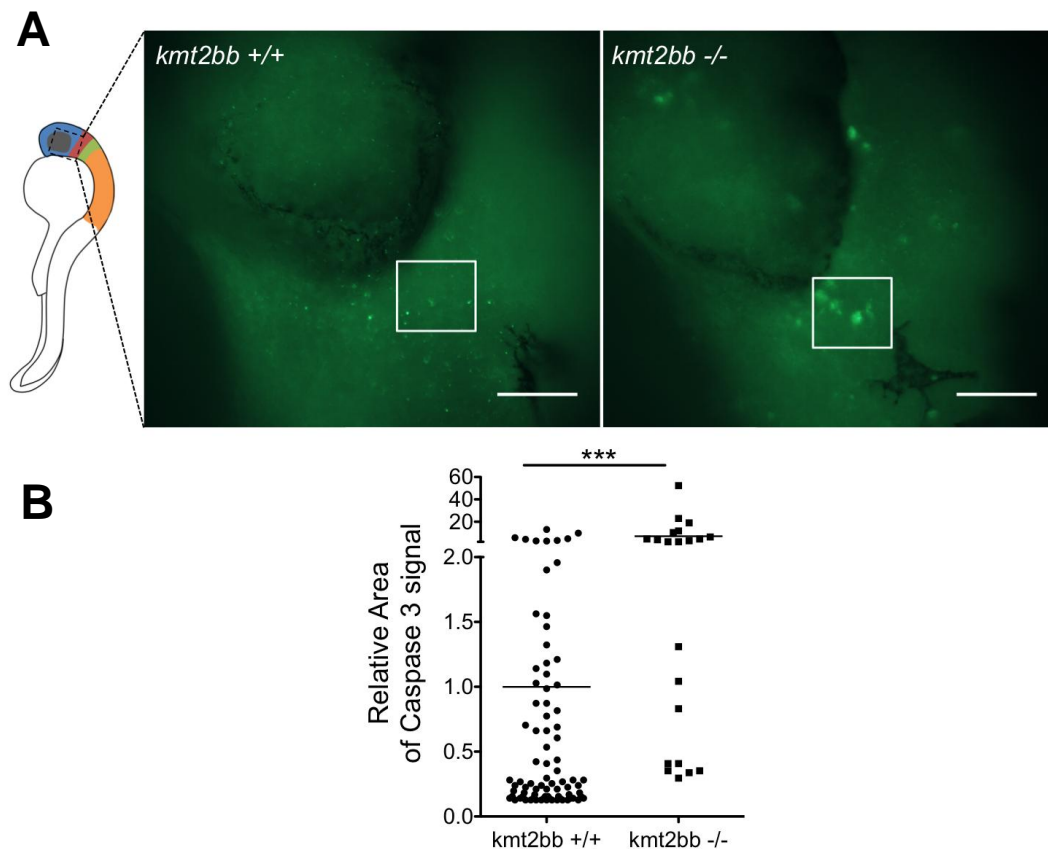
the FYRN domain. The second truncating expression construct carries a nonsense mutation (p.Phe1175\*) leading to an early stop of KMT2B protein translation already after the CXXC motif (Figure S8A). These expression constructs together with the *kmt2bb* KO zebrafish model, which I generated within my PhD thesis, will enable us to firstly, characterize the pathogenic effect of each identified mutation in *KMT2B* and secondly, will enable us to investigate if the microcephaly development of our patient is exclusively dependent on the SET domain of KMT2B.

Taken together, the presented data unravel different important novel findings. The data indicate that both *KMT2B* homologs have different functions in zebrafish embryos. In line with the phenotype of the patient, KO of *kmt2bb* causes microcephaly during early developmental stages. Since homozygous *kmt2bb* KO showed a more severe phenotype than heterozygous *kmt2bb* KO, the presented data indicate that the underlying pathomechanism is rather caused by haploinsufficiency than by a dominant negative effect.

## Results

### 4.2.11 Investigation of apoptosis in the *kmt2bb*<sup>-/-</sup> zebrafish model

In order to investigate whether increased apoptosis underlies the reduced growth rate in *kmt2bb* KO, I investigated apoptosis with immunostainings for activated Caspase 3 in fixed embryos followed by Sanger sequencing. Since the trend for a reduced head area was already detectable in *kmt2bb* KO at 5 dpf (Figure 19), I expected to detect an increased activation of apoptosis, if present, already at earlier time points. Zebrafish embryos are transparent and are thus a nice model system to perform whole-mount immunostainings. However, the skin of older embryos and larvae impedes the antibody to enter the tissue and hence, requires intensive protocol establishment. Based on this, I performed initial whole-mount immunostainings at 32 hours post fertilization (32 hpf) [65]. These immunostainings revealed that *kmt2bb* KO embryos showed large areas of activated Caspase 3 at 32 hpf compared to wt embryos, predominantly detectable at the head, caudal to the



**Figure 20: Investigation of apoptosis in *kmt2bb*<sup>-/-</sup> zebrafish.** (A) Representative immunofluorescence image of *kmt2bb*<sup>+/+</sup> and *kmt2bb*<sup>-/-</sup> embryos at 32 hpf using an antibody recognizing activated Caspase 3. Rectangle highlights area, which was applied for quantification in (B). Scale bars=100  $\mu$ m. Cartoon: lateral perspective of zebrafish embryo at 30 hpf illustrating position of forebrain (blue), midbrain (red), midbrain-hindbrain-boundary (green) and hindbrain (orange). Cartoon was modified after Vaz et al. [210]. (B) Quantification of area from activated Caspase 3 immunofluorescence staining in (A) in *kmt2bb*<sup>+/+</sup> (n=7) and *kmt2bb*<sup>-/-</sup> (n=4) embryos at 32 hpf. \*\*\*p<0.0001 (unpaired t-test).

## Results

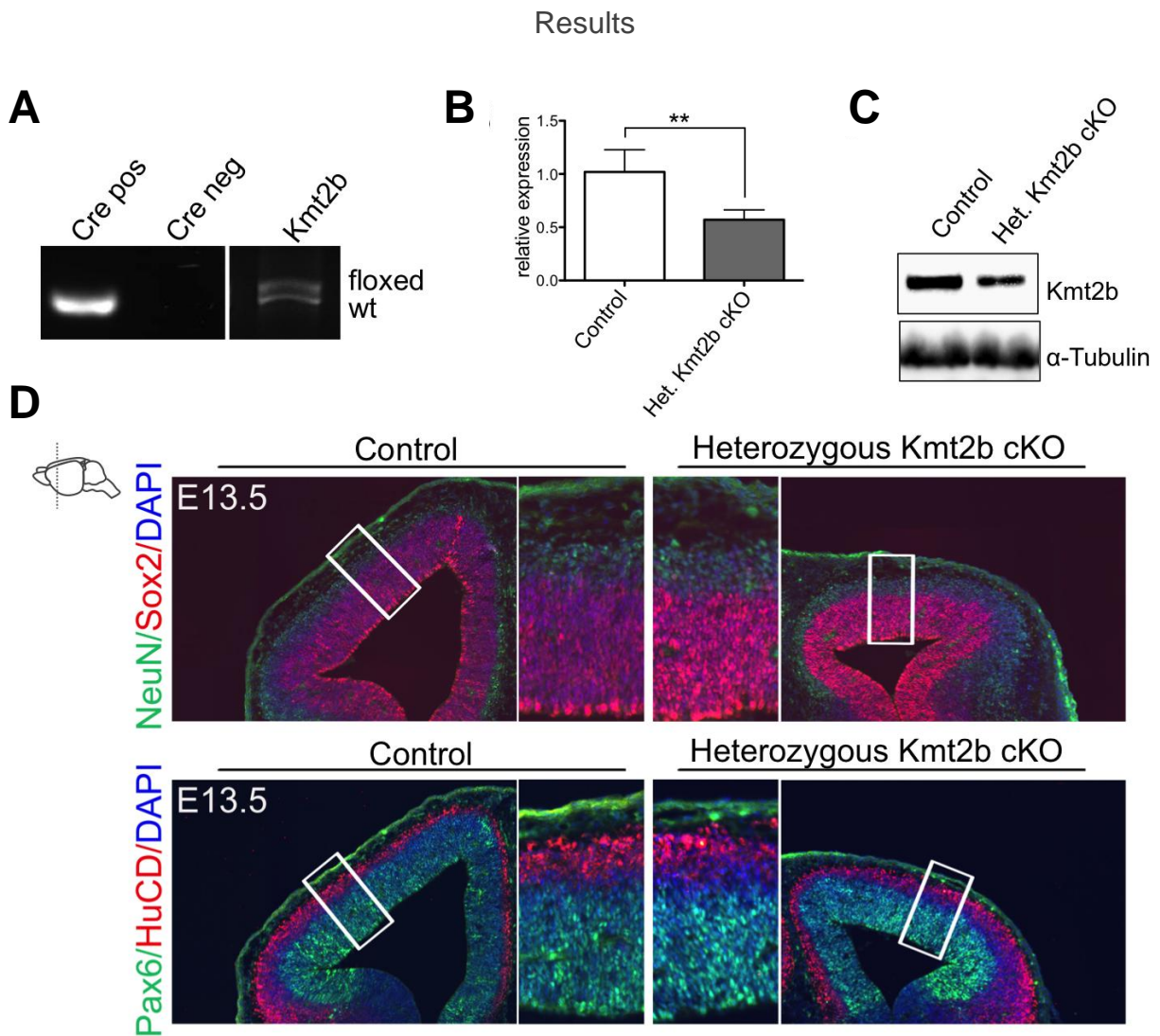
eye. A comparison with published zebrafish data indicated that the described area was likely assigned to the forebrain and/or midbrain of zebrafish embryos (Figure 20A) [210]. Quantification of the signal indicated that the activated Caspase 3 area was indeed significantly increased in *kmt2bb* KO embryos ( $***p < 0.0001$ ) (Figure 20B).

Taken together, the performed immunostainings for activated Caspase 3 point towards an increased activation of apoptosis in *kmt2bb* KO embryos at 32 hpf. Future experiments are necessary to document the course of apoptosis and associated molecular processes in more detail.

### **4.2.12 Histological investigation of the *Emx1-Cre<sup>pos/+</sup>;Kmt2b<sup>fl/+</sup>* mouse model**

The homozygous KO of *kmt2bb* in zebrafish pointed towards an increased activation of apoptosis during embryonic development. Considering the importance of *Kmt2b* during early embryonic stages, I aimed to investigate early corticogenesis and associated progenitor cells in more detail. Since the complexity of the cerebral cortex is unique to mammals, I decided to change the model system and use the mouse model instead. The murine corticogenesis and associated pathological processes are already well characterized [1,211]. Since we aimed to investigate affected cells of the cerebral cortex before they induce apoptosis, I chose the heterozygous *Kmt2b* KO model for initial experiments. According to the mouse organogenesis cell atlas (MOCA), *Kmt2b* is expressed in NPCs and RGCs between E9.5 and E13.5 and hence, we decided to investigate embryos at E13.5 [212].

The previous described zebrafish data pointed towards an underlying haploinsufficiency rather than an underlying dominant negative effect (see 4.1.10.). Based on this, I used the following Cre/loxP system. A floxed *Kmt2b* line (*Kmt2b<sup>fl/fl</sup>*), with exon 2 flanked by loxP sites, was crossed with an *Emx1-Cre* line (*Emx1-Cre<sup>pos/+</sup>*), expressing *Cre* in early cortical progenitor cells from E9.5, thereby causing a conditional heterozygous *Kmt2b* KO in the cerebral cortex [213]. The pregnant female was sacrificed at E13.5 (early corticogenesis) and embryos were subsequently collected and genotyped (Figure 21A). *Emx1-Cre<sup>pos/+</sup>;Kmt2b<sup>fl/+</sup>* embryos (hereinafter referred to as “heterozygous *Kmt2b* cKO”) and *Emx1-Cre<sup>+/+</sup>;Kmt2b<sup>fl/+</sup>* embryos (hereinafter referred to as “control”) were applied for subsequent experiments. qPCR analysis revealed that the expression of *Kmt2b* was significantly reduced in cortices of heterozygous *Kmt2b* cKO embryos by 45 % and this reduction



**Figure 21: Investigating heterozygous *Kmt2b* cKO.** (A) Genotyping of *Kmt2b* embryos after crossing *Emx1-Cre<sup>pos/+</sup>* mice with *Kmt2b<sup>fl/fl</sup>* mice were all heterozygous for the floxed (fl) and wild-type (wt) *Kmt2b* allele and either positive or negative for the *Cre* allele. (B) qPCR confirming reduction of *Kmt2b* expression in cerebral cortices of heterozygous *Kmt2b* cKO embryos. \*\* $p=0.0033$  ( $n=4$ ; Mean  $\pm$  SD; unpaired t-test). (C) WB confirming reduction of *Kmt2b* in cerebral cortices of heterozygous *Kmt2b* cKO embryos.  $\alpha$ -Tubulin served as loading control ( $n=3$ ). (D) Immunohistochemistry for neural stem cell (NSC) markers Sox2 and Pax6 as well as pan-neuronal markers HuCD and NeuN at E13.5. Cartoon (left) illustrating the section plane of isolated brains.

was verified via WB on protein level as well (Figure 21B/C). In collaboration with the lab of Dr. Tran Tuoc at the Institute of Neuroanatomy, Göttingen, I next performed immunohistochemistry (IHC) to investigate progenitor cells and generated neurons at E13.5. IHC for the neural stem cell (NSC) markers Sox2 as well as Pax6 and pan-neuronal markers HuCD and NeuN, however, did not reveal any obvious differences in heterozygous *Kmt2b* cKO embryos (Figure 21D). In order to investigate progenitor cells in more detail, a double staining for Pax6 positive RGCs, of the VZ, and Tbr2 positive IPCs, of the SVZ, was performed. However, also these immunostainings did not reveal any significant differences (Figure S10).



In summary, using the Cre/loxP system, I successfully generated a heterozygous cKO of *Kmt2b* in the murine cerebral cortex. Subsequent IHC at E13.5, however, did not reveal any significant structural differences in analyzed brain areas between control and heterozygous *Kmt2b* cKO embryos indicating that heterozygous *Kmt2b* cKO embryos had either a mild phenotype, which was not detectable with IHC, or that these embryos did not present with a phenotype during early corticogenesis.

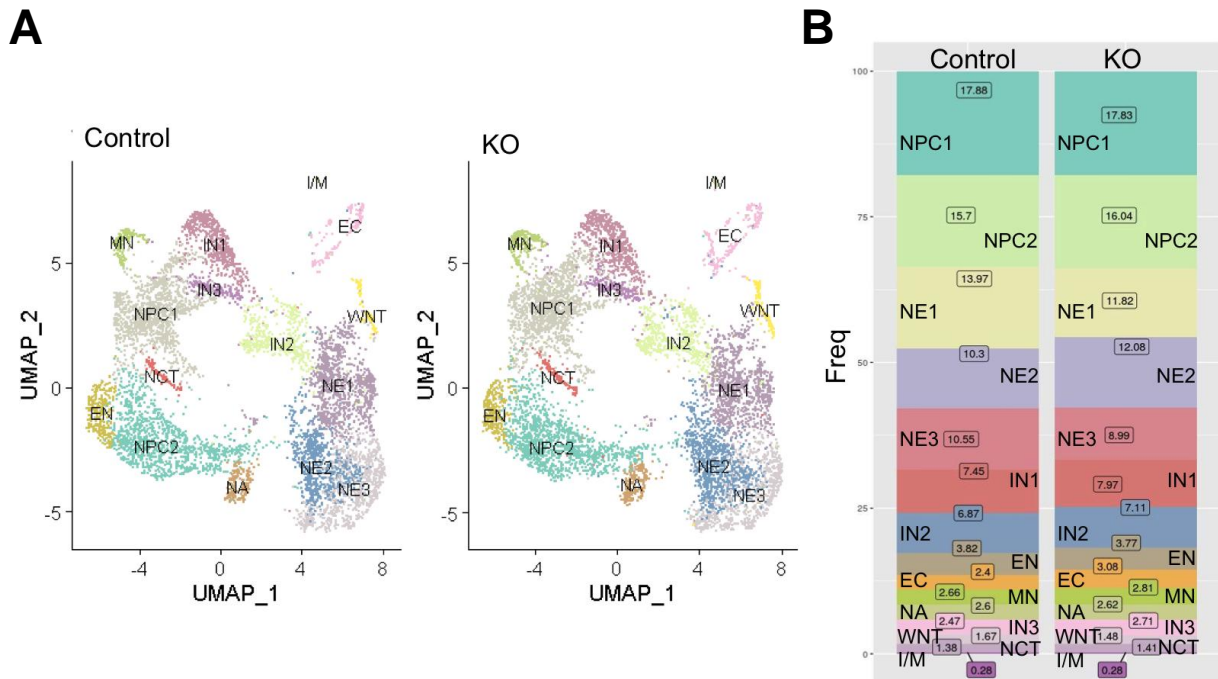
### ***4.2.13 scRNAseq analysis in the heterozygous Kmt2b cKO model***

The presented IHC data did not reveal an obvious phenotype in heterozygous *Kmt2b* cKO embryos. Previous studies, however, already illustrated that the severity of the human phenotype can not always be recapitulated in murine KO models, even if the molecular pathomechanism is detectable [77]. I therefore aimed to investigate heterozygous *Kmt2b* cKO in more detail, with the main focus on gene expression, due to the known function of *Kmt2b* as methyltransferase. Considering the IHC results, I expected only mild differences and decided to investigate gene expression with a more sensitive NGS-based approach than RNAseq and applied scRNAseq.

For this, I extracted forebrains at E13.5, pooled tissues of four embryos after genotyping and subjected samples to scRNAseq, which was performed in collaboration with the lab of PD Dr. Malte Spielmann, at the Max Planck Institute for Molecular Genetics, Berlin. Subsequently, 10,000 nuclei per sample were subjected to library preparation using the Chromium Single Cell 3' Library & Gel Bead Kit v3 (10x Genomics) (see 3.2.1.8.4). Sequencing resulted in approximately 7,700 and 8,600 nuclei in control and heterozygous *Kmt2b* cKO samples, respectively, with an average of 1,300-1,500 unique molecular identifiers (UMI) per nuclei (hereinafter referred to as "cells"). Subsequently, scRNAseq data were analyzed using the R toolkit *seurat* [214,215]. It is important to consider for data analysis that the developing forebrain consists already of different brain regions at E13.5, including cerebral cortex, choroid plexus, and ganglionic eminences [76,216]. This indicates that besides progenitor cells of the cerebral cortex, which produce excitatory glutamatergic projection neurons, also progenitor cells of the ganglionic eminences can be found and these cells produce inhibitory GABAergic neurons [76].

The performed analysis resulted in the identification of 14 cell clusters with most cells at E13.5 being characterized by progenitor-like characteristics, which is in line with E13.5 as an early time point of cortical neurogenesis (Control 68.4 % vs. KO

## Results

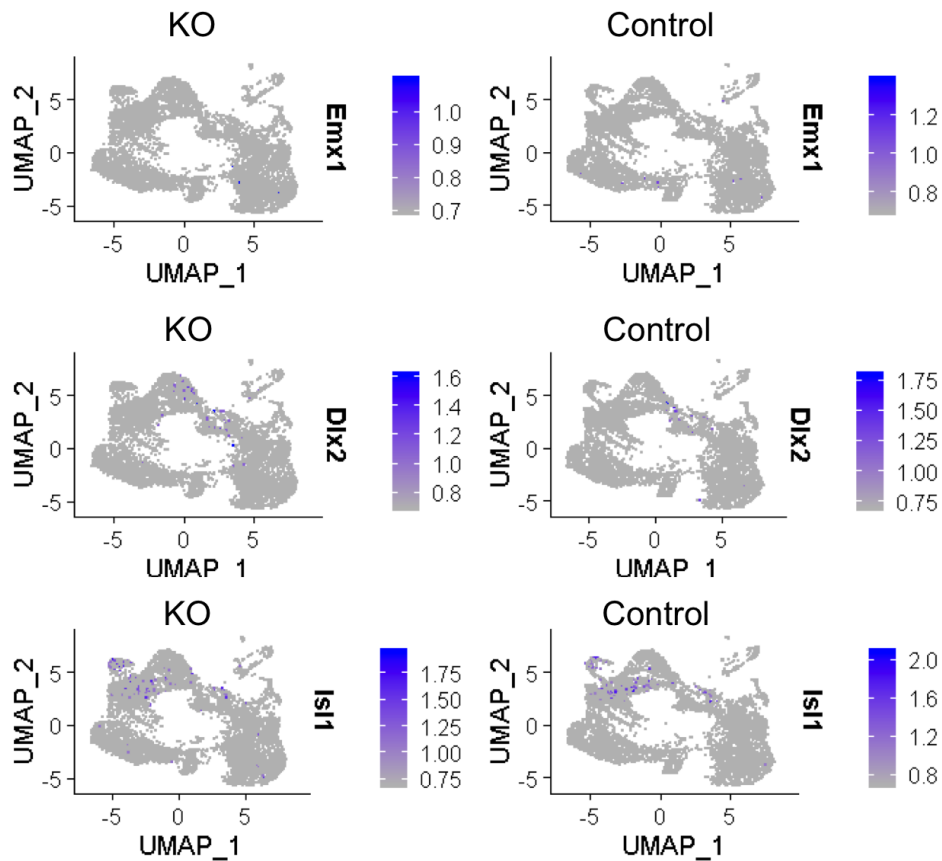


**Figure 22: scRNAseq in heterozygous *Kmt2b* cKO embryos at E13.5. (A)** Uniform manifold approximation and projection (UMAP) plot illustrating clustering of forebrain cells, pooled from four embryos of each genotype. Neural progenitor cell (NPC); neural epithelia cells (NEC); inhibitory neurons (IN); excitatory neurons (EN); endothelial cells (EC); motoneurons (MN); immunocells/macrophages (I/M); WNT-signaling (WNT); indistinct neuronal markers (NCT); not annotated (NA). **(B)** Quantitative comparison of the amount of each identified cell cluster in control and heterozygous *Kmt2b* cKO (KO) embryos. Number in boxes represent percentages of cells.

66.8%) (Figure 22). Among these identified progenitor clusters, cells arranged in either three NEC clusters (Control 34.8 % vs. KO 32.9 %) or in two NPC clusters (Control 33.6 % vs. KO 33.9 %). NEC clusters 1 – 3 expressed RGC markers, which overlap with markers of the VZ [217-219]. NEC1 expressed *Hes5*, *Ednrb* and *Sox2*, while clusters NEC2 and NEC3 expressed *Pax6*. NEC1 and NEC3 expressed proliferation markers *Ki67* and *Top2a*, whereas cells in NEC2 did not express these proliferation markers indicating that clusters NEC1 and NEC3 contained proliferating cells and NEC2 did not. A closer look at the proliferating NEC clusters NEC1 and NEC3 revealed that these clusters were slightly reduced in heterozygous *Kmt2b* cKO embryos compared to control animals (Control 24.5 % vs. KO 20.8 %) (Figure 22B). In contrast, NEC2, which did not express proliferation markers, was slightly increased in heterozygous *Kmt2b* cKO embryos (Control 10.3 % vs. KO 12.1 %).



## Results

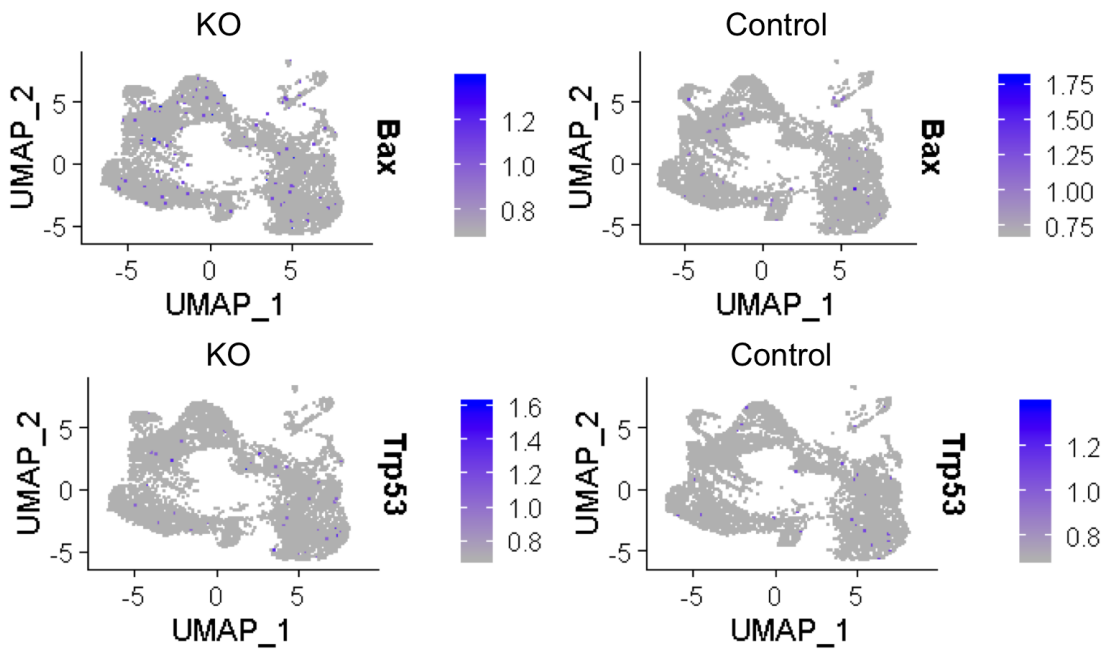


**Figure 23:** Forebrain specific marker genes in heterozygous *Kmt2b* cKO embryos. UMAP plots illustrating cells expressing specific markers of different forebrain regions (*Emx1*, *Dlx2*, *Isl1*) in control and heterozygous *Kmt2b* cKO embryos (KO).

The NPC cluster NPC1 expressed *Pbx3* and *Zfx3* indicating that these cells likely represented precursors of inhibitory neurons, which are generated in ganglionic eminences of forebrains and migrate long distances towards cortical layers (Control 17.9 % vs. KO 17.8 %) [220-222]. Accordingly, NPC1 clustered in close proximity to inhibitory neuronal clusters IN1 and IN3 and in addition, these cells expressed *Isl1*, a gene expressed in cells of ganglionic eminences (Figures 23) [223]. The NPC cluster NPC2 expressed *Eomes (Tbr2)*, a marker of IPCs from the SVZ of the developing neocortex (Control 15.7 % vs. KO 16.0 %) [218,224]. In line with the important role of IPCs for neuron production in the developing neocortex, NPC2 clustered in close proximity to excitatory neurons (EN) (Figure 22A) [225].

Most differentiated neuronal cells at E13.5 were inhibitory neurons (IN), which is in line with previous published forebrain data [226]. Three IN clusters were identified

## Results



**Figure 24: Apoptosis-associated genes in heterozygous *Kmt2b* cKO embryos.** UMAP plots illustrating cells expressing the apoptosis-associated genes *Bax* and *Trp53* in control and heterozygous *Kmt2b* cKO (KO) embryos.

(Control 16.8 % vs. KO 17.8 %) and in agreement these cells expressed the ventral telencephalon specific genes *Isl1* and *Dlx2* (Figure 23) [223]. Moreover, one EN cluster was identified (Control 3.8 % vs. KO 3.8 %). *Emx1* is a marker for cells mainly derived from the cerebral cortex itself and hence, expression of *Emx1* should be detected in RGCs, IPCs and ENs of developing cortices [227]. However, *Emx1* gene expression was barely detectable in scRNAseq data, which can be likely attributed to the high number of dropout events of scRNAseq (Figure 23) [228,229].

In addition, non-neuronal cell clusters were characterized by endothelial cells (EC) (Control 2.4 % vs. KO 3.1 %) and immune cells / macrophages (I/M) (Control 0.3 % vs. KO 0.3 %). Next to the EC cluster, one cluster expressed several WNT-signaling related genes (WNT-cluster) (Control 1.7 % vs. KO 1.5 %). The expression of *Otx2*, *Igfbp2* and *Igfbp5* indicated that cells of the WNT-cluster also had non-neuronal character and were possibly cells of choroid plexus [218].

Since we intended to investigate molecular changes prior to the onset of apoptosis in the heterozygous *Kmt2b* cKO model, apoptosis-associated genes were investigated in more detail. Analyses of identified clusters did not reveal any significant differences in expression of genes associated with apoptosis indicating that apoptosis was not apparently activated at E13.5 (data not shown). Since Andreu-Vieyra et al. confirmed increased expression of *Trp53* and different apoptosis-

## Results

associated genes like *Bax* in conditional *Kmt2b* KO oocytes, expression of these genes was investigated in more detail [133]. The performed analysis revealed that the number of *Bax* expressing cells was slightly increased in heterozygous *Kmt2b* cKO embryos albeit the difference was not obvious for *Trp53* (Figure 24).

In summary, the performed scRNAseq analysis resulted in the identification of 14 cell clusters in developing forebrains at E13.5, which were mainly characterized by progenitor-like characteristics. However, no significant differences between control and heterozygous *Kmt2b* cKO embryos were detectable. A slight shift in the ratio of proliferating vs. non-proliferating RGCs is interesting and points towards initial differences. However, since these experiments were performed at the end of my PhD thesis, future experiments are essential in order to characterize these differences in more detail.

## 5 Discussion

### 5.1 Novel insights into the pathogenesis of syndromic microcephaly

Patients with syndromic microcephaly are characterized by a reduced head circumference and additional varying clinical features. A study in 2014 estimated that beside microcephaly, affected patients frequently show abnormal brain morphologies (76 %), intellectual disability (65 %), epilepsy (43 %), ophtalmological malformations (30 %), facial dysmorphism (19 %), oropharynx (13 %), skeletal (13 %), kidney (13 %), heart (14 %) and gastrointestinal (9 %) malformations [230]. In 29 % of investigated cases, a genetic cause was identified, while the cause remained unknown in 41 % indicating that a substantial number of microcephaly patients remain without a genetic diagnosis.

In the course of my PhD thesis, I aimed to address this issue and firstly, expanded the mutation spectrum in known microcephaly-associated genes in two patients with syndromic microcephaly. Secondly, I identified a novel highly convincing candidate gene for microcephaly and discovered a low-level somatic mosaicism in one affected family member. Finally, I characterized a mutation in a novel microcephaly-associated gene *KMT2B* and provide first experimental data for the underlying pathomechanism, thereby expanding our knowledge about microcephaly and giving novel insights into cellular processes beyond microcephaly.

#### 5.1.1 The power of NGS

My PhD project is a practical example how state-of-the-art technologies can be applied in order to answer different scientific questions. Within the course of my PhD thesis, I applied four different NGS-based approaches, namely WES, deep amplicon sequencing, RNAseq and scRNAseq. Whereas the results of RNAseq and scRNAseq will be in focus later in the discussion, the WES and deep amplicon sequencing results will be discussed in the following.

##### 5.1.1.1 Identification of a low-level somatic mosaicism

In 2010, WES was applied for the first time as gene identification and diagnostic tool in order to identify mutations in patients with Miller syndrome [151]. Since then, its usage has increased enormously and to date, WES is still the method of choice to identify mutations in monogenic diseases. The release of WES data from a large

## Discussion

healthy control group (60,706 individuals) in the ExAC Browser, allows insights into common and rare genetic variants and facilitates the discrimination between mutations and polymorphisms [231]. Even though WES covers not the whole genome, but only the coding exons and surrounding intronic regions, it is assumed that approximately 80 % of mutations are located within these regions [232]. The success rate of WES varies massively in published literature and ranges from 29 % up to 68 % [233-235].

In the course of my PhD, I investigated in total five WES data of patients with syndromic forms of microcephaly and identified causal variants in three analyses.

In the first WES analysis, I investigated a family characterized by varying severities of syndromic microcephaly and identified a frameshift mutation (c.848delA; p.Asn283Thrfs\*7) in the *DLG3* gene in index patient K3627 and his affected sister, but not in the affected mother (Figure 5). These data indicated that the variant did not co-segregate with the phenotype in the family, which usually would exclude *DLG3* as candidate gene. However, since the c.848delA mutation was highly convincing (see 4.1.1.2.), I applied DNA derived from different tissues of the mother to deep amplicon sequencing. The c.848delA mutation was detectable at varying low levels, thereby proving a low-level somatic mosaicism and emphasizing that DNA derived from different tissues should be investigated when mosaics are hypothesized. Even though cDNA position 848 was covered in WES data with 103 reads, the deep amplicon sequencing results illustrated why the c.848delA mutation was not called here. Considering a VAF of 0.4 % in blood derived DNA (Table 11), not one single read is estimated to be positive in 103 reads and instead, for one positive read, a minimum of 250 reads is required. The mean coverage in the WES data of the mother estimated 78 reads and hence, 250 reads exceeds the mean coverage of WES by far, indicating that WES is not a suitable approach to detect low-level mosaicisms.

The presented analysis was the first investigation of low-level mosaicism using deep amplicon sequencing at the Institute of Human Genetics in Göttingen. However, the awareness about the relevance of mosaic mutations for genetic diseases increased enormously during the past years, since firstly, somatic mosaic mutations can manifest phenotypes in affected patients; secondly, mosaic mutations might occur more frequently than expected thereby reducing the number of misinterpreted *de novo* germline mutations; and finally, mosaic mutations might account for incomplete

penetrance and variable expressivity in some pedigrees [191,192,236-238]. Since mosaic mutations escape detection in routine diagnostics, deep amplicon sequencing was subsequently established in the laboratory diagnostic of the Institute of Human Genetics in Göttingen and is offered for genetic testing to date.

### ***5.1.1.2 Expanding the mutational spectrum of DDX11***

In the second WES analyses, I expanded the mutation spectrum of *DDX11*, a known gene associated with Warsaw breakage syndrome and only one dozen reported cases to date (Figure 6B) [193,195]. Warsaw breakage syndrome has an autosomal recessive trait of inheritance and affected patients are clinically very heterogeneous, but they share some common features including microcephaly, intrauterine growth retardation, and deafness [195]. *DDX11* is a known DNA helicase important for the proper cohesion of sister chromatids during mitosis and thus, for maintaining genomic integrity [194]. The protein has an iron-sulfur domain (267 – 350 aa) and six helicase motifs distributed along the *DDX11* protein. The first identified nonsense mutation of index patient K3923 is predicted to truncate *DDX11* already N-terminally of the iron-sulfur domain (p.Gln148\*). The second identified nonsense mutation of index patient K3923 is predicted to introduce a stop codon near the C-terminus of the protein (p.Arg879\*) (*DDX11* total length: 906 aa). Wu et al. investigated a previously reported p.Lys897del mutation in *DDX11* in more detail, which is in close proximity to the identified p.Arg879\* mutation [239]. The authors reported that the p.Lys897del mutation results in an impaired DNA binding and ATPase activity of *DDX11*. Based on these data, it can be hypothesized that the identified nonsense mutations in index patient K3923 result most likely in loss of protein function and genomic instability, which finally underlies the development of microcephaly.

However, the present WES analysis was difficult since the performed investigation was hampered by *DDX11L* sequences. The *DDX11L* gene family evolved as a result of segmental duplication from *DDX11* and 18 copies have been described yet [240]. These copies are distributed in sub-telomeric regions along the genome and show high sequence similarities especially with exons located towards the 3 prime end of the *DDX11* gene (RefSeq NM\_001257144: in total 27 exons). As a result, the c.2557C>T mutation in exon 26 of *DDX11* was not called during the first round of WES analysis, even though the position was covered with more than 100 reads and instead, variant calling required less stringent filter criteria. In worst case, the non-identified *DDX11* mutation could have resulted in complete exclusion of *DDX11* as

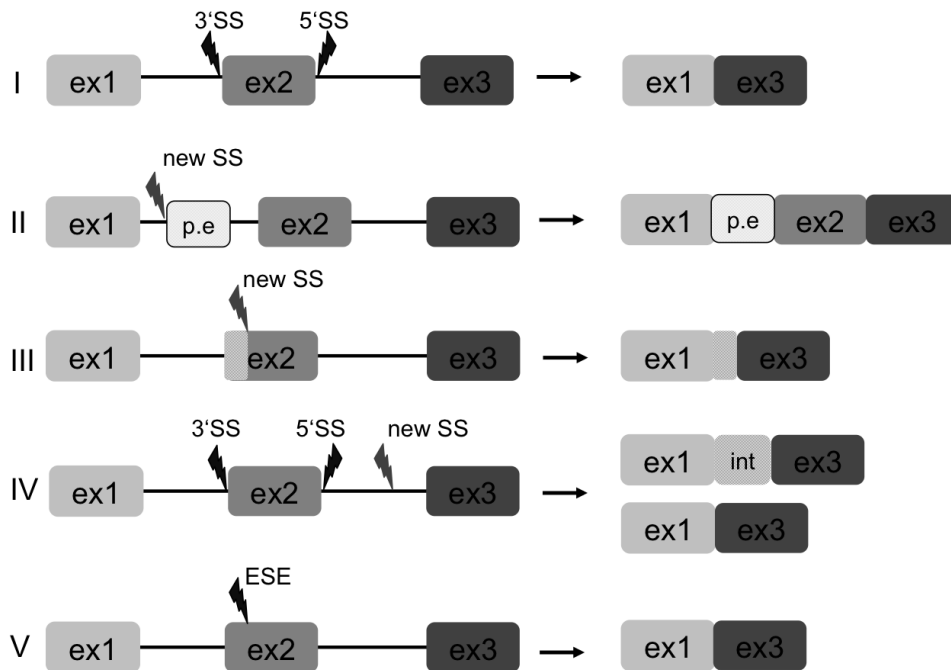
disease-associated gene. Figure 6B depicts the electropherogram after Sanger sequencing of exon 4 and 26 in *DDX11*. In line with the described sequence similarity between *DDX11* and *DDX11L*, peak height, especially of the identified c.2557C>T mutation in exon 26, was not as expected 50 %, but instead, had only a small height indicating that PCR primers amplified not only *DDX11*, but also *DDX11L*. The reduced variant frequency might be misinterpreted as normal sequencing background if data are not treated with sufficient care and it is even worth speculating that *DDX11L* copies negatively influenced the number of reported *DDX11* cases to date.

### **5.1.1.3 Identification of novel *FANCI* mutations affecting splicing**

In the third WES analysis, I identified two novel mutations affecting splicing in the *FANCI* gene, which is associated with Fanconi anemia [196].

The first identified mutation in *FANCI* (c.2890-2A>G) changed the highly conserved acceptor splice site in intron 26 from AG to GG (Figure 7B). It is estimated that 98,7 % of intronic acceptor splice sites show a canonical AG and mutations affecting this position have been frequently reported in genetic diseases to date [241-243]. Based on this, I was convinced about the severe effect of the identified c.2890-2A>G mutation and did not verify its influence on *FANCI* pre-mRNA splicing. Wimmer et al. divided splice site mutations into five categories (Figure 25) [244]: Category (I) describes mutations that affect canonical splice sites in introns resulting in exon skipping. Category (II) describes intronic mutations that generate novel intronic splice sites resulting in insertion of cryptic exons. Category (III) describes exonic mutations that generate cryptic splice sites resulting in partial loss of exonic sequences. Category (IV) is similar to (I) and describes that, after loss of canonical splice sites, a cryptic splice site can be used instead resulting in either partial intronic retention or partial exon skipping. Finally, category (V) describes mutations that destroy exonic splice enhancers (ESE) and cause exon skipping. According to the categories of Wimmer et al., it is likely that the c.2890-2A>G mutation causes more than one aberrant splice product of *FANCI* pre-mRNA since it can be assigned to categories (I) and (IV), both characterized by skipping of exon 27.

## Discussion



**Figure 25: Splicing categories of Wimmer et al. [244].** Five splicing categories (I-V) describe different mutations that result in divergent splicing products. Exon (ex); splice site (SS); pseudo exon (p.e.); intron (int); exonic splice enhancer (ESE). Cartoon modified after Abramowicz & Gos [245].

The second mutation was more difficult to interpret because at first glance it was only a missense mutation (c.2108A>G) in exon 21 (Figure 7B). Investigations of reverse transcribed cDNA, confirmed that the c.2108A>G mutation was the second mutation affecting splicing via generation of a cryptic splice site in exon 21 (Figure 7C). The c.2108A>G mutation can therefore be assigned to category (III) of splicing defects, and resulted in loss of 62 bp from exon 21 (p.Asp703Glyfs\*2) [244].

FANCI is an important protein involved in DNA damage response, a well known mechanism involved in the pathogenesis of microcephaly [54-56]. Smogorzewska et al. reported a putative nuclear localization signal (NLS) located between amino acid 779 and 795 in FANCI [246]. Importantly, the putative NLS is located C-terminally to the predicted p.Asp703Glyfs\*2 mutation and thus is lost in the truncated protein. This indicates that even if the identified p.Asp703Glyfs\*2 protein was stable, it would most likely not be able to translocate into the nucleus in order to exert its physiological function. Moreover, FANCI possesses an armadillo (ARM) repeat close to the C-terminus (985 aa – 1207 aa), which is important for DNA binding [246,247]. This repeat might therefore be impaired or lost in splicing products of the identified c.2890-2A>G mutation and based on these findings, it is likely that both identified



## Discussion

mutations result in loss of protein function and subsequently, in loss of genomic integrity, which underlies the microcephaly development.

The *FANCI* investigation nicely illustrated the essential interplay between clinicians and scientists in order to find the underlying genetic cause. More than 50,000 variants were called in the WES data of patient K3929 prior to the application of filter criteria (data not shown). In addition, even after application of filter criteria, heterozygous variants in several microcephaly-associated genes can possibly remain (Table S1). Considering a monogenic disease, finding the “one” causative mutation can be challenging and requires a combination of appropriately applied filter criteria, variant interpretation and assessment of putative phenotypic outcome. Patient K3929 was only two years of age when he was clinically investigated and at that time, he did not show any abnormal blood count (see 4.2.3). Patients with Fanconi anemia are clinically very heterogeneous, however, hematological abnormalities are key hallmarks that develop during different time points of childhood [248]. A clinical update was therefore a mandatory step in order to decide about future investigations. The c.2108A>G mutation was back then only certainly known as missense mutation (p.Asp703Gly) and a new blood sample was necessary for the extraction of RNA and subsequent splicing analysis. Convincing data were therefore important in order to decide whether the family should be contacted again, since they suffered already enough after patient K3929 received two bone marrow transplantations accompanied by recurrent infections and hospitalization.

In conclusion, using NGS-based WES, I successfully identified four novel mutations in the microcephaly-associated genes *DDX11* and *FANCI*. Using a combination of WES and deep amplicon sequencing, I identified a novel mutation in the *DLG3* gene and proved a low-level somatic mosaicism. Mutations in *DLG3* have only been reported in patients with X-linked mental retardation to date, highlighting that *DLG3* is a novel highly convincing candidate gene underlying microcephaly [188]. Therefore, my results end the long odyssey towards a diagnosis for these families and provide important information about the underlying genetic cause and open perspectives in the field of therapeutic approaches and prenatal diagnostics to plan future pregnancies.

## 5.2 The pathomechanism underlying *KMT2B*-associated microcephaly

The main part of my PhD project focused on the characterization of *KMT2B*, a novel gene associated with syndromic microcephaly. Prior to the start of my thesis, a *de novo* c.5462insC mutation (p.Leu1822Thrfs\*12) was identified in a girl presenting with severe microcephaly, short stature, reduced body weight, and facial dysmorphism (see 4.2.1). During my PhD, I was able to characterize the c.5462insC mutation in more detail and provide first data pointing towards the underlying pathomechanism, which will be discussed in the following.

### 5.2.1 The methyltransferase *KMT2B* and its importance for stem cells

Even before *KMT2B* became associated with a human genetic disease in 2016, its role in human and murine stem cells was already intensively described (see 1.4.2.) [127,128,137,148,166,170,171]. These studies together revealed that *Kmt2b* is important for H3K4me3 at promoter and enhancer regions in stem cells and that loss of *Kmt2b* results in impaired differentiation of cells and increased apoptosis rates. At the beginning of my PhD thesis, I investigated patient-derived fibroblasts, however, I did not observe any obvious phenotype during culturing of cells (see 4.2.4). These data are in line with a depletion of *Wdr5*, an important cofactor of COMPASS(-like) complexes, that also causes defects in stem cells but not in fibroblasts [249]. Based on that, we were convinced that a similar *in vitro* model was crucial for the investigation of the underlying pathomechanism in the present microcephalic patient, with the c.5462insC mutation and decided to generate patient-derived iPSC (see 4.2.5).

Considering the function of *KMT2B* as an important methyltransferase in stem cells, we aimed to investigate gene expression differences in patient-derived iPSC and performed RNAseq. The number of DE genes differed in each single comparison, with the highest numbers of DE genes constantly in comparisons with iPSC clone Pat2 (Table 13). Interestingly, Hu et al. demonstrated that the number of DE genes in *Kmt2b* KO ESCs increased significantly when cells differentiated into primordial germ cells [137]. The number of DE genes was more than doubled in comparison to ESCs, six days after differentiation was initiated and more than tripled, nine days after differentiation was initiated. In line with this, the PCA plot from the present RNAseq experiment illustrated a high biological variance between iPSC clone Pat1 and Pat2 (Figure S3). This variance was even higher than the variance between Pat1 and both wt-derived iPSC clones and prevented a paired data analysis. Considering these

## Discussion

results and the importance of *Kmt2b* for differentiation of stem cells, it can be hypothesized that iPSC clone Pat2 already initially started to differentiate whereas Pat1 did not.

However since only two iPSC clones were investigated, we can not certainly rule out that the high biological variance has other causes and is a result of incomplete quality check after generation of iPSC. Karyotyping is the method of choice to check chromosomal integrity in iPSC, while the resolution ranges from 5 up to 10 megabase (Mb) [250,251]. In contrast, we performed an Array-CGH analysis and received a resolution of 1 Mb in iPSC. Indeed, we identified CNVs in all applied iPSC, which were most likely related to reprogramming (Table S2). The size of these CNVs ranged from 2.7 up to 3.3 Mb indicating that these CNVs would have been most likely missed in studies using karyotyping. We decided to continue with iPSC clones, since no disease or pluripotency-associated genes were affected by these chromosomal aberrations, but of course, we can not exclude any relevance of these CNVs for iPSC. In addition, we can not rule out smaller CNVs and single nucleotide variants (SNVs) in iPSC clones due to the minimal resolution of 1 Mb. More than 1,000 SNVs can be detected in iPSC and from these, a handful can affect coding sequences [252]. SNVs can occur during different stages of iPSC culture. They can either be present already as low-level mosaics in donor cells, occur *de novo* during reprogramming, or occur during culture of iPSC [253,254]. Identifying these SNVs is only possible with NGS-based approaches including whole-genome sequencing (WGS), WES or deep amplicon sequencing, which were not applied here. Finally, we can not exclude that the high biological variance is connected to the reprogramming of fibroblasts to iPSC itself. The epigenome is reconstituted during the process of reprogramming and H3K4me is an important part of it [255,256]. Variations in the epigenome between iPSC lines have been associated with reprogramming, whereas these differences usually decrease with continuous passaging [255,257,258]. We did not investigate DNA methylations or histone modifications during the course of reprogramming or during culturing of iPSC in general. Hence, we can not rule out that the high biological variance between iPSC clone Pat1 and Pat2 is linked to aberrant epigenome reconstitution originated during reprogramming and can be both dependent or independent of *KMT2B*.

### 5.2.1.1 Linking WNT-signaling to the function of KMT2B during stem cell differentiation

The subsequent RNAseq analysis in iPSC identified differential expression of several WNT-signaling-associated genes (see 4.2.7). A connection of WNT-signaling and KMT2B is highly interesting and indeed, different studies pointed already towards a direct regulation. Kerimoglu et al. performed RNAseq and H3K4me3 ChIPseq in hippocampus of adult *Kmt2a* and *Kmt2b* cKO mice [174]. RNAseq did not point towards altered expression of WNT-signaling-associated genes. However, ChIPseq uncovered that promoter regions of several WNT-signaling-associated genes, including *Fzd2* and *Fzd7*, which were de-regulated in Pat1 derived iPSC, had reduced H3K4me3 in *Kmt2b* cKO mice (Table 4). An altered H3K4me3 at these genes was not observed in *Kmt2a* cKO mice, indicating that *Kmt2b* specifically generates H3K4me3 at WNT-signaling-associated genes. In addition, GST-pulldown assays revealed an interaction of  $\beta$ -Catenin and KMT2B pointing towards a putative role of  $\beta$ -Catenin in targeting KMT2B COMPASS-like complexes to WNT-regulated genes [259]. Furthermore, the WNT-signaling pathway has already been linked to the development of primary microcephaly as well. Mutations in the *WDFY3* gene, for instance, have been identified in patients presenting with primary microcephaly and the pathogenesis was linked to a de-regulation of WNT-signaling (see 1.2.1) [63].

Based on these findings, we used the presented RNAseq data of my PhD thesis as an indicator to select the direction of further investigations. A putative de-regulation of WNT-signaling can have different outcomes. Either canonical or non-canonical WNT-signaling pathways are affected (see 1.1.2). Non-canonical WNT-signaling has not been extensively investigated in stem cells yet, whereas canonical WNT-signaling has been widely described. The publications about the roles of canonical WNT-signaling in ESCs are controversial. While some papers reported an importance of activated canonical WNT-signaling for self-renewal [260,261], others reported an importance of activated signaling for differentiation [201,202]. Giving the reported role of *Kmt2b* for differentiation of ESCs and its redundancy for self-renewal, we aimed to investigate differentiation of iPSC in a next step [170].

All iPSC clones were differentiated into P1 of NPCs, accompanied by a normal course of *OCT4*, *SOX1* and *PAX6* gene expression (see 4.2.8). Lubitz et al. reported an impaired differentiation of homozygous *Kmt2b* KO ESCs, which was accompanied by a delayed reduction of *Oct4* during embryoid body (EB) formation [170]. The data

from my PhD thesis did not uncover any delayed reduction of *OCT4* in patient-derived cells, which indicates that the underlying pathomechanism in our patient is either different, or that different approaches account for this discrepancy. Lubitz et al. investigated a homozygous *Kmt2b* KO in murine ESC, while we investigated a heterozygous depletion of *KMT2B* in human iPSC. This emphasizes that different *in vitro* systems characterized by different dosages of *KMT2B* were investigated and of course, these deviant conditions might account for the observed differences. In addition, Lubitz et al. used the EB formation assay and differentiated cells into mature neurons ( $\beta$ -Tubulin III positive). The EB formation assay is commonly used *in vitro* in order to investigate the differentiation ability of stem cells towards all three germ layers, thereby testing their pluripotency [262]. In my PhD thesis, pluripotency of iPSC was not investigated and since we hypothesized an early involvement of *KMT2B* in the microcephaly-associated pathomechanism, iPSC were differentiated into an early stage of NPC (see 4.2.8). It is therefore likely that an impaired differentiation of patient-derived cells becomes obvious during later stages of differentiation *in vitro*, for instance during differentiation into mature neurons, or during differentiation into organoid cultures [85].

In order to investigate an involvement of WNT-signaling in more detail, different small molecules that either activate or inhibit WNT-signaling can be applied, to firstly, perform subsequent rescue experiments, or secondly, to enhance the molecular phenotype of differentiating patient-derived iPSC, also into organoid cultures [263,264].

In summary, the presented data of my PhD thesis point towards a de-regulation of WNT-signaling in patient-derived iPSC, while the significance for the underlying pathomechanism still has to be investigated. Even though patient-derived iPSC showed normal initiation of differentiation into early NPCs, characterization of mature NPCs, their differentiation into neurons and also differentiation of iPSC into the mesodermal and endodermal lineage should be in focus of future investigations, to fully characterize the differentiation potential of patient-derived cells.

### **5.2.1.2 The role of apoptosis in patient-derived cells**

Beside WNT-signaling-associated genes, several apoptosis-associated genes were differentially expressed in Pat1 derived iPSC including down-regulation of the *TRP53* gene. p53 is an important regulator of different cellular processes like apoptosis and is well known as transcription factor to regulate the expression of several

downstream target genes [265]. In addition, p53 is important for the release of Cytochrome C from mitochondria during activation of apoptosis [208]. In line with this, no increased Cytochrome C release was obvious in patient-derived iPSC, pointing towards a physiological or even down-regulated disposition for apoptosis in these cells. However, these results contrast the published data of *Kmt2b* KO ESCs, which indeed stated an increased activation of apoptosis raising the question for the reason [170]. As already mentioned above, Lubitz et al. investigated a homozygous *Kmt2b* KO compared to the heterozygous *KMT2B* mutation of the presented iPSC model from my PhD thesis. The authors did not comment on ESCs with heterozygous *Kmt2b* KO indicating that either no or just marginal apoptotic phenotypes were detectable in these cells.

Since the presented *in vivo* zebrafish data from my PhD thesis point towards an increased apoptosis rate in homozygous *kmt2bb* KO embryos (see 4.2.11), I still hypothesize that increased apoptosis is involved in the pathomechanism underlying the phenotype of our patient. Different scenarios are conceivable that explain why apoptosis was not detectable in the *in vitro* models yet. First of all, it is difficult to compare *in vitro* with *in vivo* data, since environments and conditions differ and can only be partially simulated *in vitro*. Cells have to cope with different cellular stresses like replication stress and DNA damage during cortical development and it has been shown that cells of the VZ and SVZ are very sensitive for the activation of apoptosis [57,266]. Hence, it is possible that stress induction in patient-derived iPSC can evoke obvious differences, which are hidden in stress-free conditions [267,268]. In addition, it has to be considered that only dermal fibroblasts and iPSC were investigated for apoptosis in my PhD thesis *in vitro*. During cortical development, the VZ and SVZ are characterized by a pool of different stem and progenitor cells including NECs, RGCs and IPCs (see 1.1.1). Apoptosis might become therefore obvious in other cell types like mature NPCs or during differentiation of NPCs into neurons *in vitro*. In line with this, Lubitz et al. reported also increased apoptosis rates during differentiation of homozygous *Kmt2b* KO ESCs into neurons ( $\beta$ -Tubulin III positive) [170]. In my PhD thesis, iPSC were differentiated into NPCs. For this protocol, single cell iPSC cultures were required and the Rho-associated protein kinase (ROCK) inhibitor Y-27632 was applied in order to increase cell survival (see 3.2.2.3). Y-27632 is commonly used in iPSC cultures and suppresses mitochondrial dependent apoptosis. However, Y-27632 also affects the metabolism of treated cells [269-271]. Based on that, we can

not exclude that Y-27632 had any impacts on the cellular phenotype during iPSC differentiation.

Taken together, even though increased apoptosis was not detectable in patient-derived iPSC *in vitro*, apoptosis is still favored as involved cellular mechanism in the microcephaly-associated pathomechanism. Characterization of patient-derived mature NPCs and differentiation of these cells into neurons or organoid cultures *in vitro* might be required in order to uncover apoptosis in patient-derived cells.

### **5.2.2 Haploinsufficiency vs. dominant negative effect**

Haploinsufficiency implies that a heterozygous mutation results in loss of protein function, while the amount of the remaining wt protein is not sufficient to fulfill its physiological functions. Dominant negative effect (DNE) has been first defined by Herskowitz in 1987 [272]. In contrast to haploinsufficiency, DNE implies that mutant and wt protein interact with each other, which impairs also the function of the wt protein and finally results in a more severe functional reduction than haploinsufficiency. KMT2B is cleaved by the endopeptidase Taspase 1 and subsequently, the N- and C-termini associate via its FYRN and FYRC domain (see 1.4.2). The cleavage site for Taspase 1 is located C-terminally of the identified p.Leu1822Thrfs\*12 mutation of our patient and hence, the p.Leu1822Thrfs\*12 protein can not be cleaved anymore, while its FYRN domain is intact. As a result, the FYRN domain of mutant KMT2B could possibly interact with the FYRC domain of wt KMT2B, thereby impairing its physiological function and causing a more severe effect (Figure 3). Based on this, it was mandatory during my PhD thesis to investigate whether haploinsufficiency or DNE were the disease underlying effect in our patient.

#### **5.2.2.1 p.Leu1822Thrfs\*12 KMT2B is not detectable *in vitro***

One major part of my PhD thesis focused on the identification of the c.5462insC transcript and p.Leu1822Thrfs\*12 KMT2B protein in patient-derived cells (see 4.2.6). Initially, I applied deep amplicon sequencing to confirm stability of the c.5462insC transcript. Usually, allele specific (q)PCRs are performed for these types of investigations via usage of complementary mismatches in primer/probe sequences, which enable proper allele discrimination [273-275]. The identified c.5462insC mutation, however, affected a cytosine stretch, offering only reduced sequence variation at the mutation site and complicating specificity of primer/probe design.

Deep amplicon sequencing, on the other hand, requires only a regular PCR product with primers flanking the mutation site, which underscores that deep amplicon sequencing was the method of choice for this investigation.

In addition, I aimed to specifically identify the mutant protein with a sensitive approach. For this, I applied mass spectrometry and focused for direct identification of mutant KMT2B protein on the unique eleven amino acid sequence, derived from the frameshift (p.Leu1822Thrfs\*12). Before I applied patient-derived cells for analysis, I aimed to establish the method with overexpressed ectopic mutant KMT2B. During establishment, we nicely identified several peptides of KMT2B, located N-terminally of the mutation site, however, unfortunately, we did not detect the unique frameshift derived peptide (data not shown). Therefore, I could not use mass spectrometry for direct identification of mutant KMT2B protein in patient-derived cells. Instead, I assessed the total amount of KMT2B protein (wt and mutant KMT2B) in nuclear and cytosolic fractions and these results were in line with the reduced amount of KMT2B detected with WB analysis, which together favored an underlying haploinsufficiency as disease underlying effect (see 4.2.3).

### ***5.2.2.2 In vivo results favor haploinsufficiency as disease underlying effect***

In addition, I investigated a KO of the KMT2B homolog *kmt2bb* in zebrafish (see 4.2.10). Homozygous *kmt2bb* larvae showed a more severe reduction of the head area than heterozygous *kmt2bb* KO larvae and hence, favored again an underlying haploinsufficiency as disease underlying effect.

While homozygous *kmt2bb* KO embryos were characterized by a progressive form of microcephaly, heterozygous *kmt2bb* KO larvae showed a significant reduction at 14 dpf, but not at 21 dpf. However, we have to take into consideration that the applied zebrafish lines (hybrid of ABxTLF) had a different genetic background, which might explain this inconsistency (PD Dr. Roland Dosch, Institute of Human Genetics Göttingen, personal communication). In line with this, high variations in head shape were observed during head measurements and these variations were independent of *kmt2ba* and *kmt2bb* genotypes (data not shown).

In order to finally proof haploinsufficiency vs. DNE *in vivo*, I prepared wt and c.5462insC KMT2B expression constructs, which can be subsequently used for transcription of RNA (Figure S8A). Single injection of wt and c.5462insC *KMT2B* RNA and combined injection of wt and c.5462insC *KMT2B* RNA into zebrafish



embryos will finally proof whether the p.Leu1822Thrfs\*12 protein exerts a DNE or not [276]. Since different *Kmt2a* KO studies reported varying lethal phenotypes, which were associated with varying length of Kmt2a protein, I also aimed to investigate a second, earlier truncated KMT2B protein [147,277,278]. The identified p.Leu1822Thrfs\*12 mutation truncates KMT2B after the FYRN domain. For the second truncated KMT2B protein, I therefore introduced a stop codon in exon 10, predicted to result in KMT2B truncation already after the CXXC domain (p.Phe1175\*) (Figure S8A). This mutant protein has its CXXC domain, which is important for recruitment to target sites, however, it lacks all C-terminal located domains including PHD1-4, BROMO and SET domain [137].

In summary, my PhD thesis provides fundamental *in vitro* and *in vivo* data that favor haploinsufficiency over DNE as disease underlying effect. Overexpression of wt and p.Leu1822Thrfs\*12 KMT2B in zebrafish will be essential to finally proof this.

### **5.2.3 Early corticogenesis in the heterozygous *Kmt2b* cKO mouse**

#### **5.2.3.1 The strength of scRNAseq over RNAseq**

The NGS-based approach scRNAseq was first introduced in 2009 [279]. Thereafter, the amount of published data increased vigorously and today, a bunch of scRNAseq data is available including data from specific brain regions and even from whole murine embryos [212,218,279,280]. scRNAseq has the advantage over RNAseq that transcripts from each cell are uniquely barcoded during library preparation. These transcripts can be subsequently assigned to single cells and these can be assigned again to different cell populations characterized by similar expression profiles [228]. In contrast, RNAseq investigates only a pool of cells and hence, data from small cellular populations might remain undetected indicating that scRNAseq is more sensitive in order to identify differences of small cellular populations [281]. Since the IHC results of my PhD thesis did not reveal an obvious phenotype in heterozygous *Kmt2b* cKO embryos, only a mild phenotype was hypothesized and hence, scRNAseq was the method of choice for subsequent gene expression analysis.

#### **5.2.3.2 scRNAseq - Identification of mild differences between proliferating and non-proliferating RGCs**

Roughly 80 % of all neurons in the cerebral cortex are ENs, which are generated in the proliferative zones (VZ and SVZ) of the developing neocortex itself and hence,

## Discussion

abnormalities here are typically investigated in studies focusing on microcephaly development [76,282]. The remaining number of neurons in the cerebral cortex are inhibitory and are not generated in the developing neocortex itself, but instead, mainly in the ganglionic eminences and these cells migrate towards the cortical layers while differentiating [76,283].

We applied forebrains of heterozygous *Kmt2b* cKO embryos to scRNAseq, since haploinsufficiency of *Kmt2b* in the cerebral cortex could possibly impact the surrounding tissue as well. Even though the performed analysis did not uncover any significant differences in heterozygous *Kmt2b* cKO embryos, subsequent in-depth analysis revealed that the ratio of proliferating vs. non-proliferating RGCs was slightly shifted between genotypes (see 4.2.13). Indeed, these differences were only mild and thus, they likely point towards the initial steps of the pathogenesis and are therefore discussed in the following.

RGCs are considered as progenitors of most neurons in the developing neocortex [17,18,284,285]. They mainly undergo asymmetric cell divisions generating daughter cells that either remain as RGC, or differentiate towards IPC, neuron, or later also towards astrocytes [15-18,284,286]. This implies that RGCs play an important role in balancing the number of progenitors and hence, in regulating the number of generated neurons and finally cortical size. In line with this, disturbances in RGCs have been frequently reported underlying the development of microcephaly [287,288]. Premature differentiation of RGCs results in depletion of the progenitor pool and hence, in loss of neurons, which are subsequently formed during neurogenesis. Even though the scRNAseq data did not show any differences in the number of IPCs and ENs at E13.5, it is possible that these differences become apparent during later embryonic stages, which should be investigated in future experiments.

Based on the scRNAseq data, we can not distinguish if the RGCs of clusters NEC1, NEC2 and NEC3 belong exclusively to the cerebral cortex or also to the ventral telencephalon. The identified NEC clusters represent most likely a mixture of RGCs from different telencephalic regions including progenitors from ganglionic eminences. An imbalance of RGCs from the cerebral cortex and ganglionic eminences indicates that the conditional *Kmt2b* KO affected both progenitor pools. Gorski et al. confirmed that the applied *Emx1-Cre* line expresses *Cre* not only in developing cortices, but also in parts of the ganglionic eminences [76]. Therefore, the expression data of

Gorski et al. strengthen the hypothesis that the observed imbalance in RGCs is not a coincidence, but rather specific for the conditional heterozygous *Kmt2b* cKO.

### ***5.2.3.3 Future perspectives - Detailed characterization of the mild phenotype during corticogenesis***

The causes for an imbalance between proliferating vs. non-proliferating RGCs can be diverse. It is conceivable that the regulation of proliferation vs. differentiation of RGCs itself is impaired, or that the cause can be found already earlier in NECs and that the transition of NECs into RGCs is affected [284,289].

Considering the data of my PhD thesis that pointed towards a putative involvement of WNT-signaling, a disturbed balance of proliferation vs. differentiation in RGCs might be interesting, since this process is indeed regulated by WNT-signaling (see 1.1.2) [28,31]. A very mild phenotype in heterozygous *Kmt2b* cKO animals, on the other hand, might point towards a too late onset of *Cre* expression and hence, too late onset of the cKO, which favors earlier developmental stages as important time point. *Emx1* is expressed from E9.5 while NECs start to transform to RGCs between E9.0 and E10.0 indicating that earlier *Cre* expression, for instance under the promoter of *Foxg1* (E8.0), could cause a more severe phenotype in heterozygous *Kmt2b* cKO embryos [13,77,290]. Since *Kmt2b* was already known to be important during early embryonic stages and in order to determine the exact time point of its involvement for cortical neurogenesis, I included exactly these two *Cre* lines in the appropriate animal application [171]. We decided to perform initial experiments with the *Emx1-Cre* line since this mouse line results in a more consistent phenotype compared to the *Foxg1-Cre* line (Dr. Tran Tuoc, Institute of Neuroanatomy Göttingen, personal communication). However, based on the presented scRNAseq data, either an earlier initiated cKO of *Kmt2b*, or even a heterozygous germline KO of *Kmt2b* should be in focus of future experiments.

Even though the scRNAseq data pointed towards an imbalance of proliferating vs. non-proliferating RGCs, no increased expression of apoptosis-associated genes was detected, which at first glance contrasts the increased apoptosis activation observed in homozygous *kmt2bb* KO zebrafish embryos (see 4.2.11). However, apoptosis can still be an important part of the pathomechanism and two scenarios are possible here. Firstly, it is conceivable that *Kmt2b* dosage determines apoptosis activation and therefore apoptosis was not initiated in heterozygous *Kmt2b* cKO embryos. Or secondly, it is conceivable that apoptosis is activated during later developmental time

points, namely in differentiating neurons indicating that regulation of proliferation and apoptosis are both involved in the underlying pathomechanism at different time points [287,291]. KMT2B has been already associated with apoptosis as well as with cell proliferation in previous studies (see 1.4.2). Furthermore, a later onset of apoptosis activation also matches the above discussed apoptosis results from iPSC (see 5.2.3), which together favor a later onset of apoptosis activation during corticogenesis.

### **5.2.3.4 Analyzing heterozygous *Kmt2b* cKO embryos**

We decided to start with initial investigations in heterozygous *Kmt2b* cKO embryos mainly due to two reasons. Firstly, we aimed to investigate affected cells during early corticogenesis before they induce apoptosis and hence, a milder phenotype was intended. Secondly, several published data pointed towards a detectable phenotype already in heterozygous *Kmt2b* cKO animals and gave reason to start with initial investigations here, which will be discussed in the following.

Even though Glaser et al. did not report a phenotype in heterozygous *Kmt2b* KO mice, it was conceivable that a mild form of microcephaly remained undetected in these mice and in line with this, Goldsworthy et al. reported in 2013 that heterozygous germline *Kmt2b* KO animals of Glaser et al. indeed had diabetes, which was not reported yet [146,175]. Additionally, Kerimoglu et al. investigated a conditional *Kmt2b* KO in excitatory neurons of adult mice forebrains (*CamKII-Cre;Kmt2b<sup>fl/fl</sup>*) [173]. Even though the authors intended to generate a homozygous KO, roughly 30 % of *Kmt2b* protein was still detectable and these animals showed impaired memory formation. In addition, a heterozygous KO of the sister gene *Kmt2a* was reported with an obvious phenotype characterized by growth retardation, anemia, hypo-fertility and segment anomalies [147]. However, unlike us, Yu et al. investigated a heterozygous germline KO of *Kmt2a* and not a cKO, which most likely causes a more severe phenotypic manifestation.

Moreover, similar cKO strategies using the same *Emx1-Cre* line of my PhD thesis resulted in heterozygous cKO embryos presenting with microcephaly. Mao et al. investigated a heterozygous cKO of *Rbm8a* (*Emx1-Cre;Rbm8a<sup>fl/+</sup>*) between E11.5 and E14.5 and observed severe cortical abnormalities from E12.5 [287]. In addition, it was previously shown that germline haploinsufficiency of *Magoh* (*Magoh<sup>Mos2/+</sup>*) caused severe cortical abnormalities starting from E12.5 [291]. McMahon et al. established a heterozygous cKO model for the *Magoh* gene (*Emx1-Cre;Magoh<sup>fl/+</sup>*)

and confirmed microcephaly at E16.5 [292]. Since McMahon et al. did not comment on earlier time points, it is not known whether microcephaly was already detectable prior to E16.5 in *Emx1-Cre;Magoh<sup>fl/+</sup>* mice as well. *Rbm8a* and *Magoh* are both components of the exon junction complex (EJC) and hence, are involved in RNA splicing [293,294]. A direct connection of *Kmt2b* with these genes can not be made and in agreement, *Kmt2b* does not regulate expression of *Magoh* in murine ESCs [171]. However, instead, *Kmt2b* regulates expression of the duplicated *Magoh* gene *Magohb* in murine ESCs [171]. MAGOH and MAGOHB are both components of the EJC and involved in nonsense-mediated mRNA decay (NMD) [295]. The proteins show 98 % sequence identity and both genes are ubiquitously expressed in adult mice. Based on their data, Singh et al. hypothesized that these proteins are either involved during different developmental stages or that MAGOHB compensates the function of MAGOH during specific time points [295]. According to MOCA, not only *Magoh* but also *Magohb* is expressed in NPCs and RGCs between E9.5 and E13.5 of murine developing brains [212]. In line with the observed imbalance in RGCs from my PhD thesis, heterozygous cKO of *Rbm8a* and *Magoh* resulted in premature depletion of progenitor cells and additionally, in increased neuronal apoptosis [287,291].

In conclusion, initial murine experiments were performed in heterozygous *Kmt2b* cKO embryos at E13.5, which point towards an imbalance of proliferating vs. non-proliferating RGCs. However, in order to receive a more detailed characterization of the pathomechanism, future experiments should investigate heterozygous as well as homozygous *Kmt2b* (c)KO animals at different time points.

### **5.2.4 Characterization of KMT2B protein and RNA**

#### **5.2.4.1 Putative SET domain independent functions of the KMT2B protein**

Several publications focus on SET domain dependent methyltransferase functions of KMT2B and investigate H3K4 methylation [127,137,173,174]. Of course, these publications consistently confirmed an essential involvement of KMT2B for these epigenetic modifications, however, growing published as well as unpublished data indicate that this is not “the whole story”, and that KMT2B has either SET domain independent functions or non-histone targets as well.

Several KMTs have already been associated with non-histone targets and *Set1*, the yeast homolog of *KMT2B*, targets a non-histone protein as well [296-298]. Apart from

## Discussion

this, the SET domain is not the only enzymatic domain in KMT2B. KMT2B has different domains and motifs that are not completely characterized and leave possibilities for SET domain independent functions (see 1.4.2). In line with this, Wang et al. showed that PHD2 of KMT2B is characterized by an intrinsic E3 ubiquitin ligase activity, which can modify H3 and H4 *in vitro*, while its relevance for non-histone targets has not been investigated yet [161]. Importantly, the PHD2 finger is located N-terminally to the p.Leu1822Thrfs\*12 KMT2B protein from our patient and is therefore intact, which further emphasizes why detection of the mutant protein was an important focus of my PhD thesis (see 4.2.6). However, since the results of my PhD thesis point towards a likely instability of the p.Leu1822Thrfs\*12 protein, also modifications accomplished by PHD2 might be involved in the pathogenesis of the microcephaly development in our patient.

Additional consent for SET independent functions comes from O'Donnell-Luria et al. who just recently discovered heterozygous mutations in *KMT2E*, in patients with neurodevelopmental diseases [299]. The phenotypic spectrum of investigated *KMT2E* patients varied tremendously, while microcephaly was reported in some affected patients. The KMT2E protein has a PHD finger and a SET domain. However, its SET domain has no intrinsic methyltransferase activity *in vitro* indicating that the protein has either different functions or that it requires additional regulations in order to become activated [300].

Furthermore, studies focusing on KMT2A point towards SET independent functions as well and these studies are highly interesting since KMT2A and KMT2B evolved from the same ancestor gene and show several similarities in their protein structure (Figure 3). Ali et al. showed that the transactivation domain (TAD), but not the SET domain of KMT2A, is essential to regulate S phase progression of U2OS cells [169]. The TAD domain is located between amino acid position 2847 and 2855 of KMT2A. A sequence alignment revealed that KMT2B has a highly similar peptide sequence (2111aa - 2119aa), which has not been investigated yet. Importantly, this peptide is located C-terminally of the p.Leu1822Thrfs\*12 KMT2B protein of our patient and therefore, certainly lost, which indicates that the peptide might be involved in the underlying pathomechanism as well.

Finally, also murine *in vivo* data of *Kmt2a* point towards SET domain independent functions. *Kmt2a* has been targeted in different KO models, which were all embryonic lethal, while they showed varying phenotypes. The first study inserted a *lacZ* reporter

## Discussion

into exon 3 of *Kmt2a* [147]. Homozygous *Kmt2a* KO embryos died after E10.5. Heterozygous animals were viable and showed obvious phenotypes including growth retardation, anemia, hypo-fertility, and segment anomalies, which is in line with an important role of *Kmt2a* for *Hox* gene regulation. Deleting *Kmt2a* in exon 5 resulted in earlier embryonic lethality than deletions in exon 3 and these homozygous KOs died already around the two cell stage [277]. Heterozygous animals were alive and showed neurological and skeletal anomalies [277]. The difference between these KO models on *Kmt2a* protein level is that the first study truncated *Kmt2a* after AT hook motifs, whereas the second study truncated *Kmt2a* in the CXXC domain (Figure 3). Replacing exons 12-14 of *Kmt2a* and hence, truncating *Kmt2a* in PHD2, caused lethality between E11.5 and E14.5 in homozygous embryos, while heterozygous animals were alive with mild anemia [301]. In contrast, if only the SET domain was replaced, homozygous KO mice were even viable and showed only skeletal anomalies [278].

In my PhD thesis, as in most *Kmt2b* KO studies, KO of *Kmt2b* was achieved after removal of the second exon causing truncation of the protein early after the first AT hook motif [146]. Goldsworthy et al. introduced a different *Kmt2b* KO mouse, which has the full length *Kmt2b* protein characterized by reduced methyltransferase activity [175]. The authors achieved this with an N-ethyl-N-nitrosourea (ENU) mutagenesis and identified a missense mutation in the SET domain of *Kmt2b* (p.Met2628Lys). Interestingly, they reported a divergent lethal phenotype compared to Glaser et al. [146]. However, since the methyltransferase activity was not completely abolished and other random mutations can not certainly be excluded, the p.Met2628Lys *Kmt2b* KO mouse of Goldsworthy et al. can not be used to study SET domain independent functions of *Kmt2b*.

In conclusion, in contrast to the described *Kmt2a* studies, no other *Kmt2b* KO models, either with varying length of *Kmt2b* proteins, or with a deleted SET domain, have been investigated to date. Since several data point towards SET domain independent functions of KMT2B, it is a mandatory step to investigate if the microcephaly development of our patient is truly SET domain dependent or even SET domain independent.

#### 5.2.4.2 The zebrafish – Model to study SET domain dependency for microcephaly development

Since the above discussed data favor also SET domain independent functions of KMT2B, a deeper characterization of encoded protein domains and target proteins is an essential step. During my PhD thesis, I aimed to address this characterization. For this, I established a *kmt2bb* KO model in zebrafish characterized by significant microcephaly development (see 4.2.11). In order to investigate if the phenotype development is dependent on the SET domain function, I prepared different expression constructs, which can be subsequently used during rescue experiments.

The prepared KMT2B expression constructs harbor different mutations (Figure S8A). Beside wt and p.Leu1822Thrfs\*12 KMT2B, two constructs are characterized by missense mutations. The first missense mutation was reported by Hu et al. in 2017 [137]. The authors introduced this *Kmt2b* mutation (p.Tyr2602Ala, corresponding to p.Tyr2604Ala in human KMT2B) in murine ESCs and confirmed abolishment of the methyltransferase activity, whereas the protein was still recruited to target sites. Therefore, the p.Tyr2604Ala protein has its full protein length and can be used as negative control for the methyltransferase activity. As long as the pathomechanism is dependent on the SET domain, it should not be able to rescue the phenotype. The second missense mutation is an unpublished mutation of our collaborator Dr. Gaetan Lesca from the Hospices Civils de Lyon, Bron, France. To my knowledge, missense mutations in *KMT2B* have only been reported in patients with dystonia, but not in patients with microcephaly [154]. Dr. Lesca identified the c.7852G>A mutation in a patient with syndromic microcephaly (-2.5 SD) and no signs of dystonia by the age of six years, which is similar to the phenotype of our patient (unpublished data). The c.7852G>A mutation (p.Glu2618Lys) is therefore the first *KMT2B* missense mutation causing microcephaly. The p.Glu2618Lys mutation affects the highly conserved SET domain of KMT2B and is in close proximity to the p.Tyr2604Ala mutation of Hu et al. [137]. Both positions are highly conserved among KMT2 proteins and even down to the yeast homolog Set1 (Figure S9). *In silico* mutagenesis using the published SET domain crystal structure of KMT2D uncovered that both missense mutations face the substrate channel of the SET domain indicating that the p.Glu2618Lys mutation abolishes methyltransferase activity in KMT2B as well (Figure S8B) [124].

In summary, during my PhD thesis, I established an *in vivo kmt2bb* KO model and prepared different expression constructs, which will be essential for subsequent



rescue experiments in order to proof whether microcephaly development is dependent or independent of the SET domain in KMT2B.

#### **5.2.4.3 A lncRNA transcribed from the KMT2B locus**

Beside the above discussed SET independent functions of the KMT2B protein, other functions might also be encoded in different *KMT2B* RNA sequences. It has already been reported that RNAs can be bi-functional, indicating that the same RNA transcript can either be translated into protein or perform regulatory functions as RNA [302,303]. Moreover, the awareness for different transcripts after alternative splicing including lncRNAs increased during the past years. These non coding RNAs have typically a length of > 200 bp and are already associated with the development of human diseases [304,305]. lncRNAs of 16,849 genes and 75,141 transcripts have been annotated in GENCODE to date, while roughly 40 % are brain specific [306,307]. lncRNAs have been shown to bind WDR5, an essential component of COMPASS-(like) complexes and are important for the regulation of gene expression including the *HOXA* gene [308,309]. In addition, lncRNAs are required for pluripotency in stem cells, during neuronal differentiation and seem to be involved in cortical development as well [310,311]. However, they have not been identified as cause underlying microcephaly in human patients to date.

In my PhD thesis, I investigated expression of *KMT2B* in patient-derived fibroblasts and iPSC (Figure 12B). The expression of *KMT2B* was significantly reduced in fibroblasts, which is in agreement with the published reduction of *KMT2B* in fibroblasts of dystonia patients [176]. The significant down-regulation, however, was not consistent in iPSC clones. A closer look revealed that expression of *KMT2B* was increased in both patient-derived iPSC clones by approximately 20 % compared to patient-derived fibroblasts, but not consistently in wt-derived iPSC clones (Figure 12B). In order to explain this difference, I looked for additional transcripts of *KMT2B* and indeed found one lncRNA (ENST00000606995.2) annotated in ENSEMBL [312]. This lncRNA is encoded in exon 3 of *KMT2B* and hence, not affected by the identified c.5462insC mutation of our patient (exon 27). Subsequent reverse transcription PCR (RT-PCR) confirmed that the lncRNA as well as the protein coding mRNA were detectable in investigated fibroblasts and iPSC (Figure S1). Even though the lncRNA was not specific for iPSC, its identification might be relevant for disease development allowing different conceivable scenarios.

Firstly, the identified lncRNA might influence pluripotency and differentiation of iPSC. Based on this, expression levels of the identified lncRNA should be quantified in iPSC and during differentiation as a next step. Sheik Mohamed et al. uncovered that expression of the lncRNA AK028326 was linked to mRNA expression of the pluripotency marker *Oct4* and knock down of AK028326 resulted in reduced expression of *Oct4* mRNA [313]. Even though no differences in OCT4 protein level was detectable in my PhD thesis, mRNA expression of *OCT4* was increased in both patient-derived iPSC (Figure 11B/C). Hence, it will be interesting to investigate if a connection between the expression of the *KMT2B* lncRNA and *OCT4* mRNA can be drawn in iPSC and during neuronal differentiation. In order to characterize the identified lncRNA in more detail, its ability for protein binding should be investigated as well. For this, a biotinylated RNA-pulldown experiment can be designed using the biotinylated *KMT2B* lncRNA as bait, which is followed by mass spectrometry in order to identify binding partners [314]. It will be interesting to see whether the *KMT2B* lncRNA binds COMPASS(-like) complexes or COMPASS independent proteins, thereby uncovering novel COMPASS independent functions of *KMT2B*.

Secondly, the identification of the lncRNA enables the hypothesis that other unknown *KMT2B* lncRNAs are present as well. Since 39 % of lncRNAs have a Poly(A) signal, the RNAseq (iPSC) and scRNAseq data (murine forebrains) of my PhD thesis can be re-investigated for *KMT2B* lncRNAs, which are specifically expressed in iPSC and during early cortical development [307].

In conclusion, even though the present thesis does not uncover SET domain independent functions of *KMT2B*, it provides first interesting hints that point towards this direction, which should be investigated in more detail during future experiments.

### ***5.2.5 Future perspectives – Investigating KMT2B-associated phenotypic variations in the zebrafish model***

At the beginning of my PhD thesis, *KMT2B* was not associated with any genetic disease. The patient of the present thesis was characterized by syndromic microcephaly, but showed no characteristics of the neurological movement disorder dystonia. Hence, the focus of the present thesis was not pointed towards dystonia development and instead, exclusively towards microcephaly development. End of 2016, mutations in *KMT2B* were published in patients with dystonia for the first time and only 21 % of affected patients were presenting with microcephaly [148,154,176].

## Discussion

In one reported family, microcephaly showed incomplete penetrance, while the dystonia phenotype was variable expressed [176]. The reported mean age of onset for dystonia estimated 4.1 years and 6.4 years for predicted protein truncation mutations and missense mutations, respectively [176]. Our patient is currently aged 10 years and shows no signs of any movement disorder (see 4.2.1). Still, it can not be excluded that our patient develops dystonia during later developmental stages. However, considering her age, it is unlikely and awaits an explanation.

Zech et al. hypothesized that the dosage of *KMT2B* is crucial for development of dystonia [148]. Based on the data of my PhD thesis, it can be agreed that haploinsufficiency and dosage of *KMT2B* is likely the cause for microcephaly development as well. However, the underlying pathomechanism might be even more complex and the established zebrafish models could help to shed light into this discrepancy.

*KMT2B* has two homologs in zebrafish, namely *kmt2ba* and *kmt2bb* which have been both targeted with CRISPR/Cas9 while only KO of *kmt2bb* resulted in microcephalic larvae (see 4.2.10). Based on published literature, it is conceivable that the human phenotype is spread over these homologs and that dystonia becomes obvious only in *kmt2ba* KO fish [68]. The *kmt2ba* KO larvae showed no obvious phenotype by 21 dpf, also no signs of aberrant muscular contractions. In order to investigate also later developmental stages in more detail, I set up new crosses with heterozygous *kmt2ba* and *kmt2bb* KOs at the end of my PhD project. This progeny can be subsequently used for genotyping and intensive characterization. Adult homozygous *kmt2ba* and *kmt2bb* KO fish should be investigated e.g. for viability, fertility, and swimming behaviors because dystonia characteristics might of course manifest during later developmental stages as well. The zebrafish has already been used to investigate the pathogenesis of dystonia-associated genes. The described zebrafish phenotype varies in published literature, while some publications report swimming and locomotion defects in animals, others do not observe any obvious movement abnormalities [315-318].

In the course of my PhD project, I also knocked down *kmt2ba* and *kmt2bb* with morpholinos in zebrafish larvae. For this, I injected morpholinos into one cell stage embryos that blocked translation of *kmt2ba* and *kmt2bb* mRNA [319]. The results of these injections were not included in the present thesis, since injected control morpholinos also induced a slight phenotype and hence, observations should be

## Discussion

interpreted very carefully (data not shown). During these injections, only injections of morpholinos targeting *kmt2ba* resulted in curved embryos characterized by reduced movements. Thus, these observations might point towards a dystonia related phenotype and await further investigation.

But how is it possible that the dystonia phenotype becomes obvious after morpholino injection, but not in homozygous *kmt2ba* KO embryos of the CRISPR/Cas9 experiments? The difference might be found in a putative maternal effect of *kmt2ba*, which was targeted with morpholinos, but not in the CRISPR/Cas9 experiment. *kmt2ba* RNA is detectable at high levels in zebrafish embryos already very early at 0.75 hpf and decreases around 4 hpf, which is the opposite course of the *kmt2bb* RNA level within the first four hpf [72]. However, this RNA is not expressed by the embryo itself and instead, is inherited from the mother [320]. The embryo starts to express own genes around 3.5 hpf [321]. As a result, maternal inherited RNAs and proteins affect earliest embryonic stages and many are degraded with progressive development when own zygotic genes are expressed [322]. Translation blocking morpholinos target these maternal inherited RNAs [323]. However, in the present thesis, only heterozygous *kmt2ba* KO fish were crossed among each other, underscoring that the heterozygous adult female fish inherited *kmt2ba* RNA to the embryos and hence, maternal effects were not investigated. In order to investigate maternal effects of *kmt2ba* in future experiments, homozygous female *kmt2ba* KO fish have to be used for crossings and could possibly result in detection of dystonia phenotypes in the next generation.

Assuming that *kmt2ba* and *kmt2bb* result in different phenotypes in zebrafish, the next question that immediately arises is the reason for this difference. At the beginning of my PhD thesis, I hypothesized that KO of *kmt2bb* is the preferred candidate for microcephaly development. This hypothesis based on a sequence alignment of human KMT2B with *kmt2ba* and *kmt2bb*. Even though the SET and post SET domains of KMT2B are highly conserved in *kmt2ba* and *kmt2bb*, *kmt2bb* has the conserved Taspase 1 cleavage site while *kmt2ba* has not (Figure S6). KMT2B is known to be cleaved by Taspase 1 while the FYRN and FYRC domain associate afterwards and hence, *kmt2bb* can be cleaved by Taspase 1 but *kmt2ba* remains likely un-cleaved. Even though the importance of protein cleavage is still not known, Hsieh et al. showed that knock down of *Taspase1* results in reduced expression of *HOX* genes, thereby pointing towards a connection of protein cleavage and

## Discussion

regulation of target genes [115,164]. Taspase 1 mediated cleavage is only restricted to KMT2A and KMT2B indicating that cleavage per se is not important for SET domain activity, since KMT2C, KMT2D, KMT2F and KMT2G perform H3K4 methylations without cleavage [163,164]. Kmt2b was shown to be mainly detectable in its cleaved form in murine ESCs [127]. However, the ratio of cleaved vs. un-cleaved protein might be tissue specific, since the un-cleaved form is clearly detectable [146]. In line with this, in my PhD thesis, WB using a N-terminal binding KMT2B antibody detected two protein bands in primary dermal fibroblasts (Figures 12). Furthermore, using mass spectrometry, peptides located C-terminally of the Taspase 1 cleavage site were identified in wt-derived iPSC, indicating that un-cleaved KMT2B is detectable in iPSC as well (data not shown). Considering all these data and a likely dosage effect of *KMT2B* for disease development, it can be hypothesized that the varying phenotypes in KMT2B patients might not simply be dosage dependent, but might be even more complex. The ratio of cleaved vs. un-cleaved KMT2B protein might be important to determine the phenotypic outcome. Cleaved and un-cleaved proteins could possibly have different target sites and hence, explain different phenotypes.

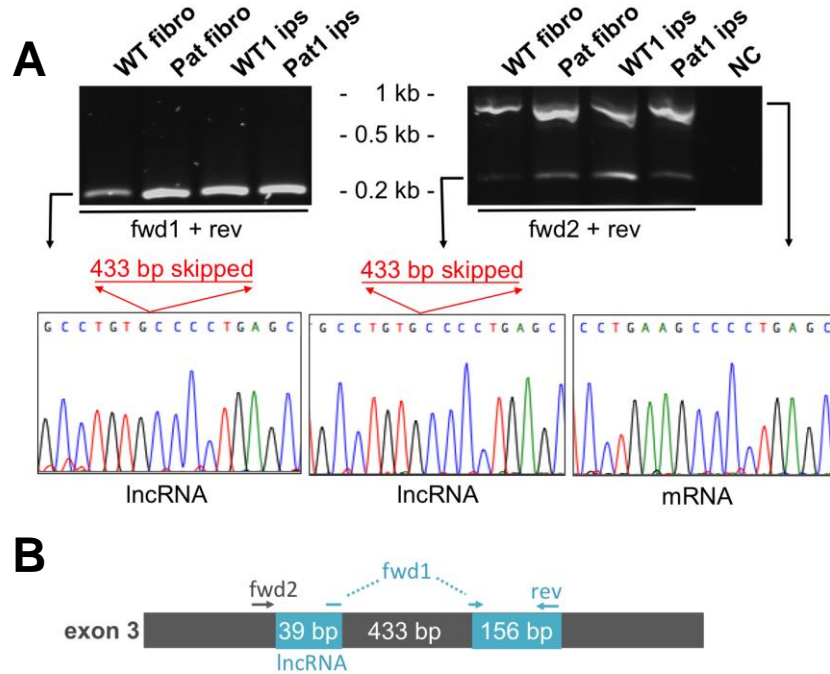
In summary, combining the data and observations of my PhD thesis as well as published literature, it is conceivable that cleavage of KMT2B is involved in phenotypic differences of affected patients. Detailed characterization of target sites and the importance of protein cleavage for target site recruitment should be in focus of future experiments in order to answer this question.

## 6 Conclusion

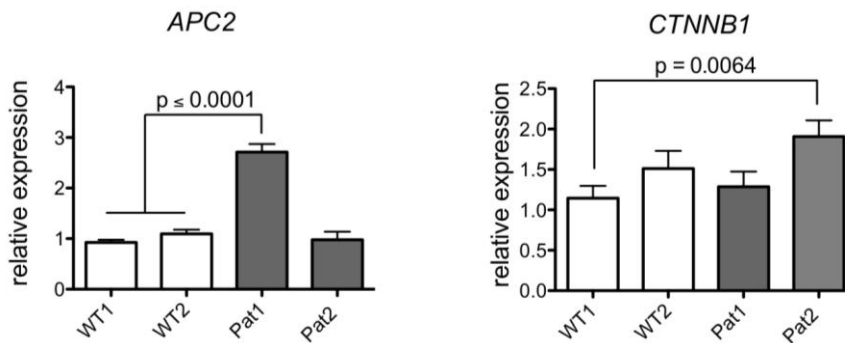
During my PhD thesis, I identified four novel mutations in *DDX11* and *FANCI*, two known microcephaly-associated genes, thereby expanding the mutation spectrum underlying microcephaly development. In addition, I identified a mutation in the novel and highly convincing microcephaly-associated candidate gene *DLG3*, which should be investigated during future analyses thereby promoting our knowledge about the diverse molecular processes underlying microcephaly manifestations. Moreover, the identification of *DLG3* was associated with application of the NGS-based approach deep amplicon sequencing, which I introduced at the Institute of Human Genetics in Göttingen.

Furthermore, I characterized a mutation in the novel microcephaly-associated gene *KMT2B*. My results provide important novel directions for future experiments in order to decipher the KMT2B protein and its functions during brain development in more detail. Since mutations in *KMT2B* are not only associated with microcephaly development, but also with cancer and dystonia, the presented data can be applied to processes and therapeutic opportunities beyond microcephaly as well. A putative link between WNT-signaling and KMT2B offers novel opportunities for drug development in order to fight cancer. In addition, the data of my thesis support that the established zebrafish model is a promising *in vivo* approach to identify small molecules for drug development in order to ameliorate dystonia features in patients with *KMT2B* mutations.

## 7 Supplementary

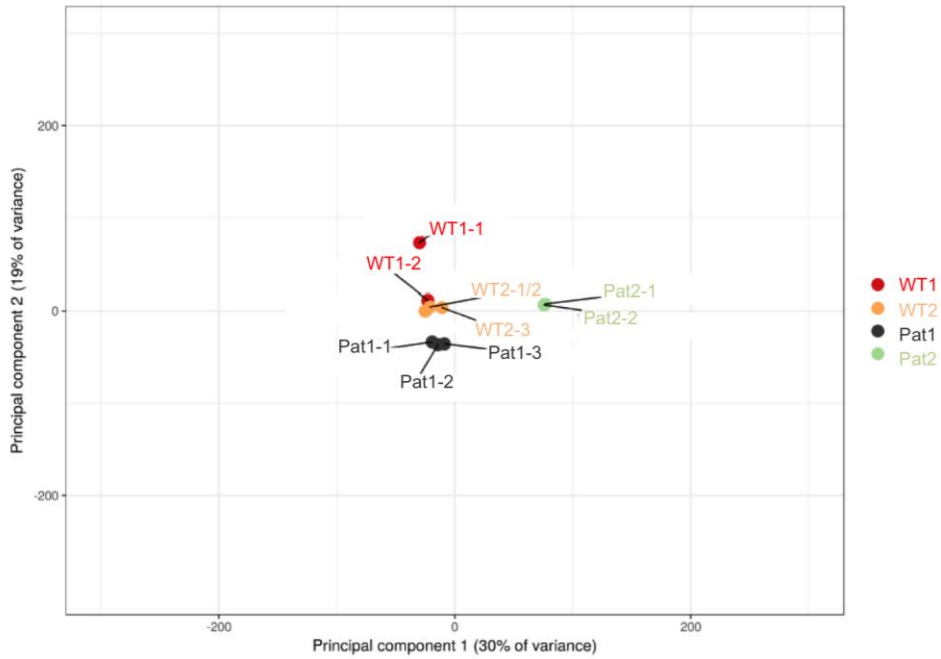


**Figure S1: Verification of *KMT2B* lncRNA.** (A) Two different PCRs were performed using primers either specific for the lncRNA (left) or for the lncRNA as well as the mRNA (right). Subsequent Sanger sequencing confirmed presence of a lncRNA beside the mRNA in exon 3 of *KMT2B* in fibroblasts as well as in iPSC. (B) Cartoon illustrating the binding sites of primers used for PCRs. Primer pair 1 (fwd1 + rev) amplified specifically the lncRNA since the binding site of the fwd1 primer is split in the mRNA sequence due to retention of a 433 bp fragment. In contrast, primer pair 2 (fwd2 + rev) amplified both RNAs since the binding site of fwd2 is located 5 prime of the lncRNA. mRNA sequence (grey and blue); lncRNA sequence (blue).

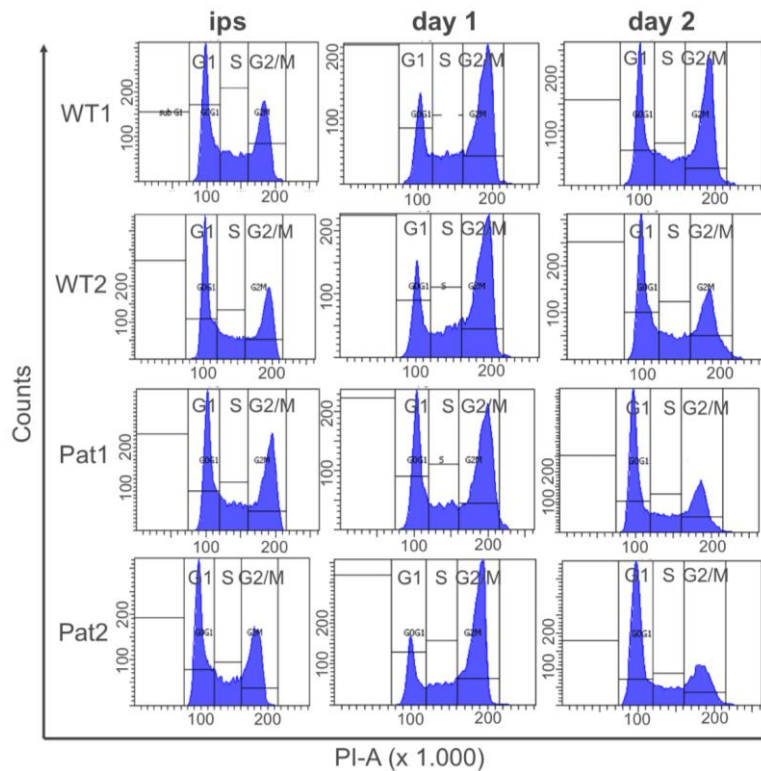


**Figure S2: Expression of *APC2* and *CTNNB1* is not consistently de-regulated in patient-derived iPSC.** *APC2*: WT1 vs. Pat1 \*\*\* $p < 0.0001$ ; WT1 vs. Pat2  $p = 0.6062$ ; WT2 vs. Pat1 \*\*\* $p = 0.0001$ ; WT2 vs. Pat2  $p = 0.3234$ . *CTNNB1*: WT1 vs. Pat1  $p = 0.3656$ ; WT1 vs. Pat2 \*\* $p = 0.0064$ ; WT2 vs. Pat1  $p = 0.2494$ ; WT2 vs. Pat2  $p = 0.0830$  (unpaired t-test; Mean  $\pm$  SD;  $n = 3$ ).

## Supplementary



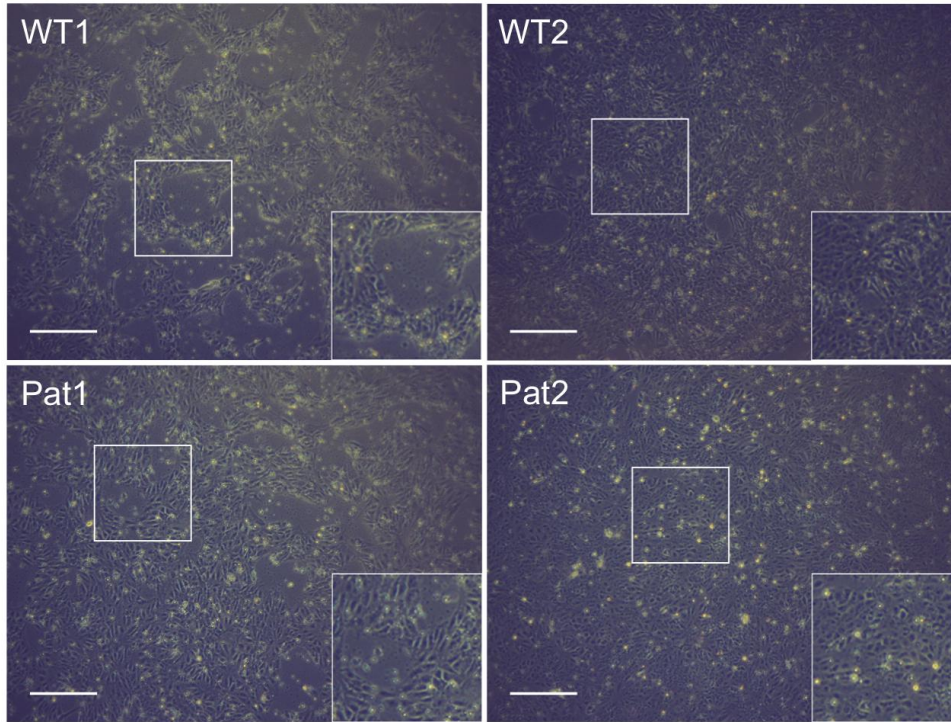
**Figure S3: PCA plot illustrating variance of iPSC samples applied for RNAseq.** Two biological replicates were applied of iPSC clones WT1 and Pat2. Three biological replicates were applied of WT2 and Pat1.



**Figure S4: Cell cycle distribution during iPSC differentiation.** iPSC and differentiating cells (day 1 and day 2) were fixed, stained with PI and investigated with flow cytometry. day 1 (n=3); iPSC and day 2 (n=2).



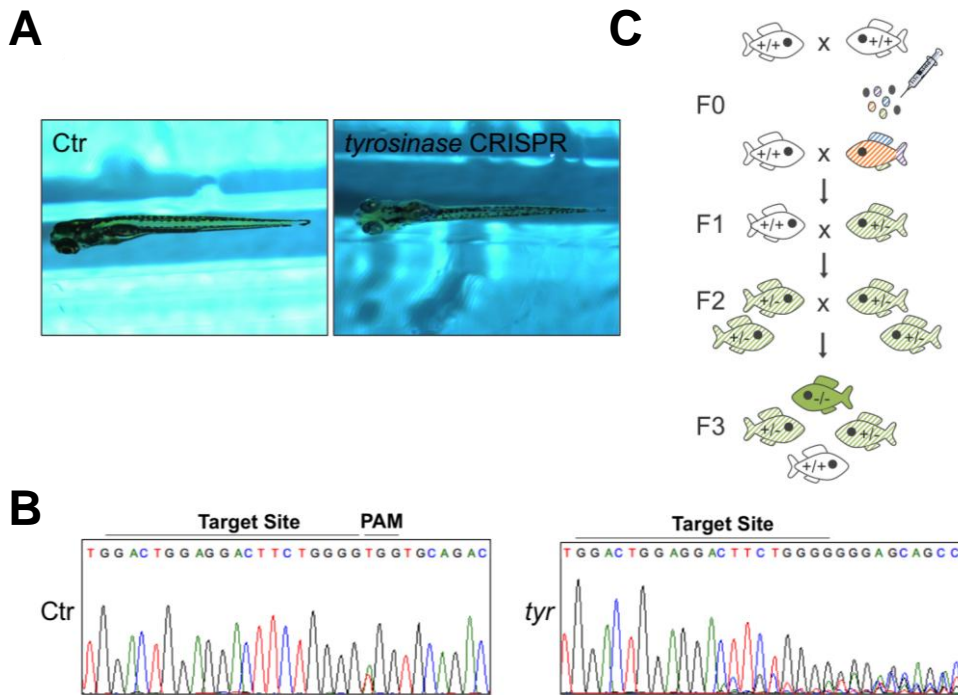
Supplementary



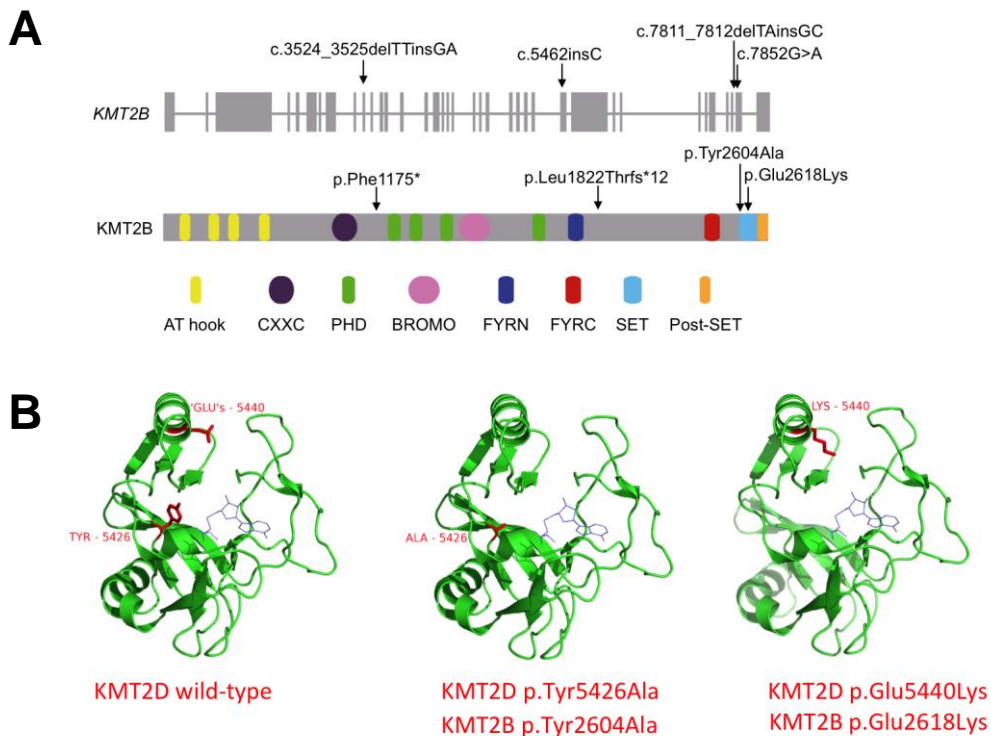
**Figure S5: Representative morphology of differentiating cells at day 1.** At the start of differentiation (day 0), iPSC colonies were broken down to single cells and  $1.5 \times 10^6$  iPSC were transferred to new matrigel coated 6-well plates per iPSC clone. Cells attached at varying degrees. Scale bars=100  $\mu$ m.

KMT2B	KKVFQKHDTLLDGKEIVNPDGFDVLRVRYVDFEGINFKRKFLTGLEPDAINVLIIGSIRI	1737
kmt 2ba	KKVYCYKHRELINSK1-EQNGFVVRVRYVDFDGLISLRKFLTGLEPESVNVMIIGSLHV	1542
kmt 2bb	KKVFCYKHRDLISGKM-ITGQGFVLRVRYVDFEGISLRKFLTGLEPESINLMIGSLLV	2236
	***:* ** :*..* <u>FYRN</u> :***:******:*..:*****:***:*****:..	
KMT2B	DSLGTLSDSLSDCEGRLPFIQYQCSRLYWSTVDARRRCWYRCRILEYRPNWGPREEPAHLEA	1797
kmt 2ba	KNLGVLTELSANSGLYPVGYQCSRWYVSTVDPRRRCRYTKCVSEVRPSTRCRTPSPFMDV	1602
kmt 2bb	EKLGKLSLSACDGKLPVGYECSRWFSTVNLPRRCRYTFRVTEVRPAVQEKQVEESP	2296
	..** *:* ** .:*:*:*:*:* ** :*** * : * * * .	
	:	
	<u>Taspase 1 cleavage site</u>	
KMT2B	SEEESSPSTRYIHFPVTVVSAPGLAP----SATPGAPRIQLDGVDDGDSEAEAVQQP	2078
kmt 2ba	METDDEVADHYLNFSRTVVVCEPAKDQTGLTVLPNSGSIQLDGDNDSESDTGEANAD	1947
kmt 2bb	NSSSDEDMENYFNFTRTVVTCKAQRDSSK--KSPSVRQISQLDGVDDGTESDTSLANS	2692
	. . . . . :* * * * . * . * . * * * . * . * . * . * . * . * . * . * .	
	:	
	<u>FYRC</u>	
KMT2B	SEEEPPSPDDKENQAPKRTGPHLRFEISSDGFVSAESLEGAWRTLIEKVQEARGHARL	2451
kmt 2ba	SSDEESFSPGKQKASLYQPHLIFQITSEDFVSAESMDVAVKAVMDGVQEARIGCRM	2593
kmt 2bb	SDDELSFSDGEEHVPLSRDQPHLRFEIASDDGFSIEADSEVAWSAVLEGVEARAAAGL	3508
	*.:* .:.. . *** *:*:*:*:*:*:*:*:* * : * : * : * : * : * : * : *	
	<u>FYRC</u>	
KMT2B	RHLSFSGMSGARLLGIHDAVIFLAEQLPGAQRQHYKFRYHQQEGEGEPPLNPHGAAR	2511
kmt 2ba	EQLPFNSINGARVGLHDAVLFLEQLQGASLQKHRFRFHQHEIPEEELPENPNGCAR	2653
kmt 2bb	RLLTMCMSGARVMGMLHDAVVYLVEQLQAKRCHRHTFRFHQASQDDDLVNPSCGAR	3568
	. * : .:*****: * : * * * : * * * * . * : * : * : * : * : * * * *	
KMT2B	AEVYLRKCTDFDMFNFLASQHRVLPAGATCDEEDEVQLRSTRRATSLELPMAMRFRHLK	2571
kmt 2ba	AEVYVRKSTDFDMFNFLASQHRQLPESRPCDDDEDDIMLKSSRRATSLELPMAMRFRHLEK	2713
kmt 2bb	SEVYLRKSTDFDMFNFLASQHRQLPDTLDLDEEDEVLLKSTRRATSLELPMAMRFRHLER	3628
	:***:*..***** ** : * : * : * : * : * : * : * : * : * : * : * : * : *	
	<u>SET</u>	
KMT2B	TSKEAVGVYRSAIHGRGLFCKRNI DAGEMVIEYSGIVIRSVLTDKREKPYDGGIGCYMF	2631
kmt 2ba	TSKEAVGVYRSAIHGRGLFCKRNI EAGEMVIEYAGNVIRAVLTDKREKYDSKIGCYMF	2773
kmt 2bb	TSKEAVGVYRSAIHGRGLFCKRNI EAGEMVIEYSGIVIRSVLTDKREKPYDGGIGCYMF	3688
	*****:*..***** ** : * : * : * : * : * : * : * : * : * : * : * : * : *	
	<u>SET</u>	
KMT2B	RMDDFDVVDATMHGNAARF INHSCEPNCFSRVIHVEGQKHIVIFALRRILRGEELTYDYK	2691
kmt 2ba	RIDDFDVVDATMHGNAARF INHSDPCNYSRVINVEGQKHIVIFALRRIYRGEELTYDYK	2833
kmt 2bb	RIDDFDVVDATMHGNAARF INHSCEPNCYSRVINVEGQKHIVIFALRRIYRGEELTYDYK	3748
	*:***** ** : * : * : * : * : * : * : * : * : * : * : * : * : *	
	<u>post SET</u>	
KMT2B	FPIEDASNKPCNGAKRCRRFLN	2715
kmt 2ba	FPIEDADNKLHCNCGARRCRRFLN	2857
kmt 2bb	FPIEDASNKLCNCGAKRCRRFLN	3772
	*****.* ** : * : * : * : * : * : * : * : * : * : * : * : * : *	

**Figure S6: The zebrafish homologs kmt2ba and kmt2bb.** Alignment of KMT2B protein sequence with kmt2ba and kmt2bb.



**Figure S7: CRISPR/Cas9-based strategy in zebrafish.** (A) The established CRISPR/Cas9 positive control targeted the gene *tyrosinase* (*tyr*) and caused loss of pigmentation after successful KO visible with the naked eye. (B) Sanger sequencing confirmed that the *tyr* gene was successfully targeted with injected CRISPR and Cas9 RNAs. (C) Experimental strategy in order to KO *kmt2ba* and *kmt2bb* in zebrafish.



**Figure S8: Expression constructs for subsequent zebrafish experiments.** (A) cDNA and protein position of all prepared *KMT2B* mutations. In-silico mutagenesis of the p.Tyr2604Ala and p.Glu2618Lys mutations in published SET domain of *KMT2D* [124]. Left: wild-type *KMT2D*. Middle: p.Tyr2604Ala in *KMT2B* corresponds to p.Tyr5426Ala in *KMT2D*. Right: p.Glu2618Lys in *KMT2B* corresponds to p.Glu5440Lys in *KMT2D*.





## Supplementary

Table S1: Heterozygous and hemizygous variants

Chr	Gene	Variant pos. cDNA	Variant pos. Protein	Notes
1	<i>DDX59</i>	c.919G>A	p.Val307Ile	
1	<i>TARBP1</i>	c.1691C>T	p.Ser564Phe	
1	<i>CASZ1</i>	c.2464G>A	p.Gly822Ser	
1	<i>DLGAP3</i>	c.37C>T	p.Arg13Cys	
2	<i>LONRF2</i>	c.112G>A	p.Asp38Asn	
2	<i>ZNF638</i>	c.784A>T	p.Met262Leu	
2	<i>ESPNL</i>	c.1687G>A	p.Val563Ile	
5	<i>GIN1</i>	c.200A>G	p.Glu67Gly	
6	<i>BYSL</i>	c.109C>T	p.Arg37Trp	
6	<i>LRP11</i>	c.122C>T	p.Pro41Leu	
8	<i>PPP1R16A</i>	c.494C>T	p.Ala165Val	
9	<i>CD72</i>	c.689-10T>G		
9	<i>FRRS1L</i>	c.151_152delCGinsGC	p.Arg51Ala	
9	<i>FANCC</i>	c.1534-10C>G		1. <i>FANCC</i> genes show recessive trait of inheritance - no second variant identified 2. phenotypic exclusion of gene
10	<i>ANK3</i>	c.2239T>G	p.Tyr747Asp	
10	<i>ARHGAP22</i>	c.1388C>T	p.Pro463Leu	
11	<i>HPS5</i>	c.2414T>C	p.Leu805Ser	
11	<i>INSC</i>	c.1313G>A	p.Arg438Gln	
11	<i>APOA4</i>	c.1147C>A	p.Gln383Lys	
11	<i>ART5</i>	c.800G>C	p.Arg267Pro	
12	<i>ANO6</i>	c.2734C>T	p.Leu912Phe	
12	<i>ABCD2</i>	c.2007A>C	p.Lys669Asn	
12	<i>M6PR</i>	c.321C>G	p.Asn107Lys	
12	<i>ITGA5</i>	c.2528G>A	p.Gly843Asp	
13	<i>ATP11A</i>	c.2236_2238delCTGinsGTC	p.Leu746Val	
14	<i>PTGDR</i>	c.1055C>T	p.Ser352Phe	
15	<i>IL16</i>	c.1375G>A	p.Glu459Lys	
15	<i>IGDCC4</i>	c.1471C>T	p.Arg491Trp	
16	<i>DNASE1L2</i>	c.176A>G	p.Gln59Arg	
16	<i>ITGAX</i>	c.2105G>A	p.Arg702Gln	
16	<i>TIGD7</i>	c.386T>G	p.Phe129Cys	
17	<i>DNAH9</i>	c.5372C>T	p.Ala1791Val	
17	<i>SLC47A2</i>	c.869C>G	p.Pro290Arg	
17	<i>CENPV</i>	c.56_64delGGGCCTCCGins(GGGCCTCCG)2	p.Ser21_Ala22insGlyAlaSer	1. similar variant in ExAC with high allele frequency 2. microtubule-associated genes have recessive trait of inheritance - no second variant identified
19	<i>CPAMD8</i>	c.5288G>A	p.Cys1763Tyr	
20	<i>SNTA1</i>	c.115G>A	p.Val39Met	
20	<i>TMC2</i>	c.933+1G>C		
X	<i>HUWE1</i>	c.3949G>A	p.Val1317Ile	

Supplementary

Table S2: Array CGH-analysis of dermal fibroblasts and iPSC.

	WT Fibro	WT1	WT2	Patient Fibro	Pat1	Pat2	Pat3
Chr.: 12q24.32q24.33 Del: 128,135,196- 130,870,474 bp Size: ca. 2.7 Mb	No	(Yes?)	Yes	No	(Yes?)	Yes	(Yes?)
Chr.: 22q13.31q13.33 Del:47,340,968- 50,131,773 bp Size: ca. 2.8 Mb	No	Yes but smaller size	Yes but smaller size	No	Yes	Yes	Yes
Chr.: 22q13.31q13.33 Del:47,340,968- 50,131,773 bp Size: ca. 2.8 Mb	No	Yes	Yes	No	(Yes)	(Yes)	(Yes)

(Yes) not called by program but likely to be there (optic aspect). (Yes?) not called by program but maybe there (optic aspect)

Table S3: Gene set enrichment analysis with topGO applying DE genes shared by all four comparisons. Table summarizes enriched GO pathways of the GO categories „molecular function“ and „biological process“.

Molecular Function				Biological Process			
GO Pathway	p value	genes/ total	genes	GO Pathway	p value	genes/ total	genes
negative regulation of female gonad development	0.012	1 / 5	<i>AMH</i>	1-phosphatidylinositol 4-kinase activity	0.016	1 / 7	<i>PI4KA P1</i>
axonogenesis involved in innervation	0.012	1 / 5	<i>NPTX1</i>	cysteine-type endopeptidase activity	0.017	2 / 85	<i>USP6, CTSF</i>
activation of GTPase activity	0.002	3 / 105	<i>APC2, USP6, CCDC1 25</i>	carbon-nitrogen ligase activity, with glutamine as amido-N-donor	0.023	1 / 10	<i>CPS1</i>
ribosomal small subunit assembly	0.001	2 / 19	<i>RPS28, PWP2</i>	semaphorin receptor binding	0.023	1 / 10	<i>SEMA 6D</i>
response to oleic acid	0.012	2 / 19	<i>CPS1</i>	cAMP-dependent protein kinase inhibitor activity	0.023	1 / 10	<i>PKIB</i>

Supplementary

Table S4: Gene set enrichment analysis with top GO applying DE genes shared by all four comparisons and by three out of four comparisons. Table summarizes candidate GO pathways of the GO term „Molecular Function“ and „Biological Process“.

Molecular Function				Biological Process			
GO Pathway	p value	genes/ total	genes	GO Pathway	p value	genes/ total	genes
MHC class II receptor activity	0.004	3 / 62	<i>HLA-DQB1</i> , <i>HLA-DRA</i> , <i>HLA-DPA1</i>	T cell costimulation	5.79e-05	6 / 130	<i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-DRB5</i> , <i>HLA-DRA</i> , <i>HLA-DPB1</i> , <i>HLA-DPA1</i>
peptide antigen binding	2.79e-05	6 / 119	<i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-DRB5</i> , <i>HLA-DRA</i> , <i>HLA-DPB1</i> , <i>HLA-DPA1</i>	antigen processing and presentation of exogenous peptide antigen via MHC class II	4.57e-05	7 / 183	<i>CTSF</i> , <i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-DRB5</i> , <i>HLA-DRA</i> , <i>HLA-DPB1</i> , <i>HLA-DPA1</i>
Wnt-protein binding	0.011	2 / 32	<i>WIF1</i> , <i>FZD2</i>	T cell receptor signaling pathway	0.0003	7 / 252	<i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-DRB5</i> , <i>HLA-DRA</i> , <i>HLA-DPB1</i> , <i>HLA-DPA1</i> , <i>PSMB9</i>
translation activator activity	0.001	2 / 10	<i>DAZL</i> , <i>RPS27L</i>	Interferon-gamma-mediated signaling pathway	7.05e-05	7 / 196	<i>NMI</i> , <i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-DRB5</i> , <i>HLA-DRA</i> , <i>HLA-DPB1</i> , <i>HLA-DPA1</i>
metal ion binding	0.006	31 / 4423	<i>CPS1</i> , <i>CA12</i> , <i>LHX5</i> , <i>IRAK3</i> , <i>COMT</i> , <i>PCDHB5</i> , <i>FEZF1</i> , <i>MICAL2</i> , <i>PHC2</i> , <i>ZNF280D</i> , <i>DNAJA4</i> , <i>ZNF300</i> , <i>CHP2</i> , <i>ZNF528</i> , <i>ZNF558</i> , <i>BNC1</i> , <i>NPTX1</i> , <i>ZNF572</i> , <i>RPS27L</i> , <i>ZNF559</i> , <i>PCDH18</i> , <i>ZNF699</i> , <i>ZNF331</i> ,	ribosomal small subunit assembly	0.0001	3 / 19	<i>RPS27L</i> , <i>RPS28</i> , <i>PWP2</i>

Supplementary

ZNF667,  
RNF39,  
MT1H,  
ZNF880,  
ZNF717,  
ZNF736,  
SPON1,  
CBSL

Table S5: Top 30 of KEGG pathway enrichment analysis of comparison "WT2 vs Pat1". GeneRatio: number of DE genes in pathway / total number of DE genes. BgRatio: Total number of genes in pathway / Total number of genes in all pathways

KEGG pathway	GeneRatio	BgRatio	pvalue	FDR
Glycine, serine and threonine metabolism	19/1172	32/5765	1,42608E-06	0,000449215
Hepatocellular carcinoma	52/1172	146/5765	1,0152E-05	0,001344327
Protein digestion and absorption	27/1172	60/5765	1,28031E-05	0,001344327
Phospholipase D signaling pathway	44/1172	122/5765	3,40901E-05	0,002657796
Relaxin signaling pathway	41/1172	112/5765	4,21872E-05	0,002657796
Platinum drug resistance	26/1172	64/5765	0,000153822	0,008075652
Glutathione metabolism	19/1172	42/5765	0,00022633	0,010184868
Chronic myeloid leukemia	28/1172	74/5765	0,000362274	0,014264555
ECM-receptor interaction	26/1172	69/5765	0,000626426	0,019876261
Proteoglycans in cancer	52/1172	168/5765	0,000630992	0,019876261
Cell adhesion molecules (CAMs)	35/1172	103/5765	0,00076822	0,021999017
Arginine and proline metabolism	18/1172	43/5765	0,001030224	0,027043375
Estrogen signaling pathway	33/1172	98/5765	0,001281501	0,031051754
Focal adhesion	52/1172	174/5765	0,001531168	0,034451275
Autoimmune thyroid disease	10/1172	19/5765	0,001767658	0,035847816
Glutamatergic synapse	32/1172	96/5765	0,001820841	0,035847816
Toxoplasmosis	31/1172	93/5765	0,002131331	0,037638696
Asthma	7/1172	11/5765	0,002150783	0,037638696
Viral myocarditis	17/1172	43/5765	0,0029495	0,047569184
Colorectal cancer	28/1172	84/5765	0,003423518	0,047569184
Endometrial cancer	21/1172	58/5765	0,003571521	0,047569184
Leishmaniasis	19/1172	51/5765	0,003794222	0,047569184
Chagas disease (American trypanosomiasis)	26/1172	77/5765	0,003864673	0,047569184
Amyotrophic lateral sclerosis	17/1172	44/5765	0,003914007	0,047569184
Apoptosis	36/1172	116/5765	0,003915571	0,047569184
Gastric cancer	37/1172	120/5765	0,003926345	0,047569184
Graft-versus-host disease	9/1172	18/5765	0,004706704	0,054720355
Long-term depression	19/1172	52/5765	0,004864032	0,054720355
Wnt signaling pathway	40/1172	134/5765	0,005219128	0,056690531
Breast cancer	37/1172	123/5765	0,006158457	0,064663803

Supplementary

Table S6: Applied synthetic oligonucleotides of PhD thesis.

Primer name	Primer sequence (5' – 3')	Application	Source
KMT2BwtNtagRev	TCTAGAGTCGCGGCCGCAAGCTTT CAGTTAAGGAACCGACGGCAG	Sequencing and cloning of different <i>KMT2B</i> expression constructs	
KMT2BmtNtagRev	TCTAGAGTCGCGGCCGCAAGCTTT CAGGAGCTCCAGGGACAAGA		
KMT2BmtNotI_R	GCGGCCGCTCAGGAGCTCCAGGG ACAAGA		
KMT2BmtXhoI_R	CTCGAGTCAGGAGCTCCAGGGACA AGA		
KMT2B_K3027_Stopp_ XbaI_R	CTCTAGATCAGGAGCTCCAGGGAC A		
KMT2B_1951_F	CAGTTTACCCCAAGCGAAG		
KMT2Bcl_F1_rev	CACCATCCGTTCTGTGCC		
KMT2Bcl_F2_fwd	GGCACAGAACGGATGGTG		
KMT2Bcl_F2_rev	GCAACAGCGTCTTCACTATC		
KMT2Bcl_F3_fwd	GATAGTGAAGACGCTGTTGC		
KMT2Bcl_F3_rev	AGAACATCTGTGTCCAGTGG		
KMT2Bcl_F4_fwd	CCA CTG GAC ACA GAT GTT CT		
KMT2Bcl_F4_BamHI_ GFP_N1_rev	GGT GGC GAC CGG TGG ATC CCG GTT AAG GAA CCG ACG GCA G		
KMT2Bcl_Seq_841_F	CGTGGAGGACAGTCAAGC		
KMT2B_Seq_945_fwd	GAT GGG ACA ATT GTC CTT GG		
KMT2B_Seq_1025_rev	CTGTCCAGATCCAACCTCTTC		
KMT2B_Seq_1154_fwd	CTGTAGCTGAGGAGATGATG		
KMT2B_Seq_1987_fwd	GAATCGGTGCTTACTCCTC		
KMT2B_Seq_2056_rev	AGGCTCAGCTGGAGAGTC		
KMT2B_Seq_4044_fwd	CGACTATGAGAGCAAGATGAT		
KMT2B_Seq_4156_rev	TCCGCAGGTGTACAGCAC		
KMT2B_Seq_6522_fwd	GTT AAG CCT TGG CCC TGC		
KMT2B_Seq_6642_rev	CAAAGGTGGCTGCTTCACT		
KMT2B_Seq_7364_fwd	CCT TTA GTG GAA TGA GTG GG		
KMT2B_Seq_7567_rev	GAGGCCAGGAAGTTGAACAT		
KMT2Bcl_Seq_4987_re	CCGGGCACACATGAAGTG		
KMT2Bcl_1987_rev	GAGGAGTAAGCACCCGATTC		
KMT2Bcl_6522_rev	GCAGGGCCAAGGCTTAAC		
KMT2Bcl_fr.2.2-3_fwd	GTGTCAACTGCCTAGACAAG		
KMT2Bcl_fr.2.2-3_rev	GGTCGCTGTTTGACGCAG		
KMT2Bcl_fr.3-4.1_fwd	GAGCACAGTGGATGCTCG		
KMT2Bcl_fr.3-4.1_rev	CGGGATGGCGAGTAGTTG		
KMT2Bcl_F2_rev	GCAACAGCGTCTTCACTATC		
KMT2Bseq1511-F	GAGGCCCTCCGGAAGAC		
KMT2Bseq4261-R	CACCACCTTGGAGCTCAG		
KMT2Bseq7909-F	GAT GTA GTG GAC GCC ACG		
KMT2Bseq4504-F	GAATCTGCGTTCGGCTG		
KMT2Bseq4881-F	GAATGTGCATGCTGCTGTG		
KMT2Bseq336-F	GAG CAG TGA CGG GGA ATC		
KMT2B_K3027_7C_F	GTG AGG ACC CCC CCA CTG		
KMT2B_K3027_7C_R	CAGTGGGGGGGTCCTCAC		
KMT2B_Y-A_2604_F	CAT CGA GGC CTC TGG CAT		
KMT2B_Y-A_2604_R	ATGCCAGAGGCCTCGATG		
KMT2B_2618_3_F	GGA AGA AGT TCT ACG ATG GG		
KMT2B_2618_3_R	CCCATCGTAGAATTCTTCC		
KMT2B_noPHD_rev	CTCTAGATCAATCCACACGGACGC GGT		
Oct4_Fwd	CAGTGCCCGAAACCCACAC	qPCR	[324]
Oct4_Rev	GGAGACCCAGCAGCCTCAAA		
Sox2_Fwd	TACCTCTTCCCTCCACTCCA	qPCR	
Sox2_Rev	GGTAGTGCTGGGACATGTGA		
Nanog_Fwd	CAGAAGGCCTCAGCACCTAC	qPCR	
Nanog_Rev	ATTGTTCCAGGTCTGGTTGC		



Supplementary

COL1A1_Fwd	gggattccctggacctaag	qPCR	[198]
COL1A1_Rev	ggaacacctcgctctcca		
DLG3_Ex6_fwd	TAATTCTTTCTGTGGCCAGG	Verification of	
DLG3_Ex6_rev	ctacagtgccaccactgagc	variant	
PAX6qPCR-F	GTGTCTACCAACCAATTCCACAAC	qPCR	[325]
PAX6qPCR-R	CCCAACATGGAGCCAGATG		
Sox1qPCR-F	GCAAGATGGCCCAGGAGAA	qPCR	
Sox1qPCR-R	CCTCGGACATGACCTTCCA		
GAPDH_qPCR_F	CCATGTTTCGTCATGGGTGTGAAC		
GAPDH_qPCR_R	CGGCCATCACGCCACAGTTT		
kmt2ba-ex28-F	CACACAATTGTGCACAGTCC	Genotyping	
kmt2ba-ex28-R	GAGACAGATGAGCACAAGAC	zebrafish	
DDX11ex26-F	tcgccgtgggaatgtgc	Verification of	
DDX11ex26-R	GAGATGCCACAGTGGATGC	variant	
DDX11ex4-F	CAGCATGGTTCAATGGCTGC	Verification of	
DDX11ex4-R	TCTGGACAGGGGTAGTGGAG	variant	
kmt2bb_ex28_new-F	GAAGCCCTTGAATTGGGGTT	Genotyping	
kmt2bb_ex28_new-R	TGAACAACGTTCAAATCGGTATC	zebrafish	
CRISPR-F	AGC ACC GAC TCG GTG CCA CTT T TT CAA GTTGAT AAC GGA CTA GC C TTA TTT TAA CTT GCT ATTTCT A GC TCT AAA AC	CRISPR/Cas9	
WIF1-qPCR-F	AAGGTTGGCATGGAAGACAC	qPCR	[326]
WIF1-qPCR-R	TTAAGTGAAGGCGTGTGCTG		
APC2-qPCR-F	GAGGAACTGGACCGGGAAC	qPCR	
APC2-qPCR-R	TCGAGAACTGCGTCTCCAC		
FZD2-qPCR-F	GCCCTCATGAACAAGTTCGGT	qPCR	
FZD2-qPCR-R	GAGCTCCGTCCTCGGAGTG		
NPTX1-qPCR-F	GAGACGCTCAGCCAACTCG	qPCR	
NPTX1-qPCR-R	GGAATTGAGGCGGCTGTACT		
CTNNB1-qPCR-F	GGTGGACCCCAAGCTTTAGT	qPCR	
CTNNB1-qPCR-R	GCTTGCATTCCACCAGCTTC		
kmt2ba_ex28_CRSC_R	ctaatacgactcactataGGTCCGGAGGTA GGACTCCTgttttagagctagaaatagcaagtt aaaataa	CRISPR/Cas9	
kmt2bb_ex28_CRSC_R	ctaatacgactcactataGGGGGCACCAAG GTCGGGTTgttttagagctagaaatagcaagtt aaaataa	CRISPR/Cas9	
Tyr_F	AGCTCTTCAGCTCGTCTCTC	CRISPR/Cas9	[183]
Tyr_R	TGAGTGAGGATACTGCGG		
tyr_gRNA_1_R	ctaatacgactcactatagg CCCCAGAAGTCCTCCAGTCC gttttagagctagaaatagcaagttaaaataa		
tyr_gRNA_2_R	ctaatacgactcactatagg GGACTGGAGGACTTCTGGGG gttttagagctagaaatagcaagttaaaataa		
FANCI_ex21-22_F	AATTTGTAACGCCAATGAATC	Verification of	
FANCI_ex21-22_R	GCGCTCTCTAGTGGAAAGAAC	variant	
FANCI_ex27_F	GCTTAGGAAACTACTGCTCAATTC		
FANCI_ex27_R	CACAGGAAAGAATACATCTTATAGG TC		
Srg3_TO666	TGT CAT CCA TGA GGA GTG GTC	Genotyping	Kindly provided
Srg3_TO667	GGT AGC TCA CAA ATG CCT GT	mouse	by Dr. Tran
Emx1_Cre_To89	ATG CTT CTG TCC GTT TGC CG	Genotyping	Tuoc
Emx1_Cre_To90	CCT GTT TTG CAC GTT CAC CG	mouse	
Kmt2b_genotype_F	cggaggaagagagcagtgacg	Genotyping	Kindly provided
Kmt2b_genotype_R	ggacaggagtcacatctgctagg	mouse	by Prof. Andre
qKmt2b_4 for	GGAGGAAGAGAGCAGTGACG	qPCR	Fischer
qKmt2b_4 rev	TCATCTGAATGAAAACCCTGAA		[174]
HUWE1_ex33_F	GCCAATTATTTTGTGCTGG	Verification of	

## Supplementary

HUWE1_ex33_R	AGGAAATCTAAACTGTCTCCTGG	variant	
KMT2B_lncRNA_5P_F	GTGGCCCCAAAAGCACC	Investigation of	
KMT2B_lncRNA_F	ATGCCTGTGCCCTGAGC	lncRNA	
KMT2B_lncRNA_R	TCAGAGCTGGAGCCCCG		

## 8 References

1. Rakic P (2009) Evolution of the neocortex: Perspective from developmental biology. *Nat Rev Neurosci* 10: 724–735.
2. Kornack DR, Rakic P (1998) Changes in cell-cycle kinetics during the development and evolution of primate neocortex. *Proc Natl Acad Sci USA* 95: 1242–1246.
3. Smart IHM, Dehay C, Giroud P, Berland M, Kennedy H (2002) Unique Morphological Features of the Proliferative Zones and Postmitotic Compartments of the Neural Epithelium Giving Rise to Striate and Extrastriate Cortex in the Monkey. *Cerebral Cortex* 12: 37–53.
4. Miterko LN, Lackey EP, Heck DH, Sillitoe RV (2018) Shaping Diversity Into the Brain's Form and Function. *Front Neural Circuits* 12: 83. doi: 10.3389/fncir.2018.00083.
5. Mueller T (2012) What is the thalamus in zebrafish?. *Front Neurosci* 6: 64. doi:10.3389/fnins.2012.00064.
6. Leung LC, Wang GX, Mourrain P (2013) Imaging zebrafish neural circuitry from whole brain to synapse. *Front Neural Circuits* 7:76. doi:10.3389/fncir.2013.00076.
7. Miller FD, Gauthier AS (2007) Timing Is Everything: Making Neurons versus Glia in the Developing Cortex. *Neuron* 54: 357–369.
8. Fame RM, MacDonald JL, Macklis JD (2011) Development, specification, and diversity of callosal projection neurons. *Cell Stem Cell* 34: 41–50.
9. Molyneaux BJ, Arlotta P, Menezes JRL, Macklis JD (2007) Neuronal subtype specification in the cerebral cortex. *Nat Rev Neurosci* 8: 427–437.
10. Miyoshi G, Butt SJB, Takebayashi H, Fishell G (2007) Physiologically Distinct Temporal Cohorts of Cortical Interneurons Arise from Telencephalic Olig2-Expressing Precursors. *J Neurosci* 27: 7786–7798.
11. Franco SJ, Müller U (2013) Shaping Our Minds: Stem and Progenitor Cell Diversity in the Mammalian Neocortex. *Neuron* 77: 19–34.
12. Götz M, Huttner WB (2005) The cell biology of neurogenesis. *Nat Rev Mol Cell Biol* 6: 777–788.
13. Misson J-P, Edwards MA, Yamamoto M, Caviness VS Jr (1988) Identification of radial glial cells within the developing murine central nervous system: studies based upon a new immunohistochemical marker. *Neuron* 44: 95–108.
14. Hatakeyama J, Bessho J, Katoh K, Ookawara S, Fujioka M, Guillemot M, Kageyama R (2004) Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development* 131: 5539–5550.
15. Haubensak W, Attardo A, Denk W, Huttner WB (2004) From The Cover: Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: A major site of neurogenesis. *Proc Natl Acad Sci USA* 101: 3196–3201.
16. Noctor SC, Martínez-Cerdeño V, Ivic L, Kriegstein AR (2004) Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci* 7: 136–144.

## References

17. Noctor SC, Flint AC, Weissman TA, Dammerman RS, Kriegstein AR (2001) Neurons derived from radial glial cells establish radial units in neocortex. *Nature* 409: 714–720.
18. Miyata T, Ayano Kawaguchi, Hideyuki Okano, Ogawa M (2001) Asymmetric Inheritance of Radial Glial Fibers by Cortical Neurons. *Neuron* 13: 727–741.
19. Kowalczyk T, Pontious A, Englund C, Daza RAM, Bedogni F, et al. (2009) Intermediate Neuronal Progenitors (Basal Progenitors) Produce Pyramidal–Projection Neurons for All Layers of Cerebral Cortex. *Cerebral Cortex* 19: 2439–2450.
20. Miyata T, Kawaguchi A, Saito K, Kawano M, Muto T, et al. (2004) Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. *Development* 131: 3133–3145.
21. Gal JS, Morozov YM, Ayoub AE, Chatterjee M, Pasko Rakic, et al. (2006) Molecular and Morphological Heterogeneity of Neural Precursors in the Mouse Neocortical Proliferative Zones. *J Neurosci* 26: 1045–1056.
22. Angevine JB, Sidman RL (1961) Autoradiographic Study of Cell Migration during Histogenesis of Cerebral Cortex in the Mouse. *Nature* 192: 766–768.
23. Gupta A, Tsai L-H, Wynshaw-Boris A (2002) Life is a journey: A genetic look at neocortical development. *Nature Reviews Genetics* 3: 342–355.
24. Dehay C, Kennedy H (2007) Cell-cycle control and cortical development. *Nat Rev Neurosci* 8: 438–450.
25. Komiya Y, Habas R (2008) Wnt signal transduction pathways. *Organogenesis* 4:2. 68–75.
26. Niehrs C (2012) The complex world of WNT receptor signalling. *Nat Rev Mol Cell Biol* 13: 767–779.
27. Harrison-Uy SJ, Pleasure SJ (2012) Wnt Signaling and Forebrain Development. *Cold Spring Harb Perspect Biol* 4:a008094.
28. Woodhead GJ, Mutch CA, Olson EC, Chenn A (2006) Cell-Autonomous beta-Catenin Signaling Regulates Cortical Precursor Proliferation. *J Neurosci* 26: 12620–12630.
29. Gunhaga L, Marklund M, Sjödal M, Hsieh J-C, Jessell TM, et al. (2003) Specification of dorsal telencephalic character by sequential Wnt and FGF signaling. *Nat Neurosci* 6: 701–707.
30. Hirabayashi Y, Itoh Y, Tabata H, Nakajima K, Akiyama T, Masuyama N, Gotoh Y (2004) The Wnt/ -catenin pathway directs neuronal differentiation of cortical neural precursor cells. *Development* 131: 2791–2801.
31. Munji RN, Choe Y, Li G, Siegenthaler JA, Pleasure SJ (2011) Wnt Signaling Regulates Neuronal Differentiation of Cortical Intermediate Progenitors. *J Neurosci* 31: 1676–1687.
32. Chenn A, Walsh CA (2002) Regulation of Cerebral Cortical Size by Control of Cell Cycle Exit in Neural Precursors. *Science* 297: 365–369.
33. Wrobel CN, Mutch CA, Swaminathan S, Taketo MM, Chenn A (2007) Persistent expression of stabilized  $\beta$ -catenin delays maturation of radial glial cells into intermediate progenitors. *Developmental Biology* 309: 285–297.
34. Fang W-Q, Chen W-W, Fu AKY, Ip NY (2013) Axin Directs the Amplification and Differentiation of Intermediate Progenitors in the Developing Cerebral Cortex. *Neuron* 79: 665–679.

## References

35. Lobry C, Oh P, Mansour MR, Look AT, Aifantis I (2014) Notch signaling: switching an oncogene to a tumor suppressor. *Blood* 123: 2451–2459.
36. Ohtsuka T, Sakamoto M, Guillemot F, Kageyama R (2001) Roles of the Basic Helix-Loop-Helix Genes *Hes1* and *Hes5* in Expansion of Neural Stem Cells of the Developing Brain. *Journal of Biological Chemistry* 276: 30467–30474.
37. Gaiano N, Nye JS, Fishell G (2000) Radial Glial Identity Is Promoted by Notch1 Signaling in the Murine Forebrain. *Neuron* 26: 395–404.
38. Dong Z, Yang N, Yeo S-Y, Chitnis A, Guo S (2012) Intralineage Directional Notch Signaling Regulates Self-Renewal and Differentiation of Asymmetrically Dividing Radial Glia. *Neuron* 74: 65–78.
39. Saxena M, Agnihotri N, Sen J (2018) Perturbation of canonical and non-canonical BMP signaling affects migration, polarity and dendritogenesis of mouse cortical neurons. *Development* 145: dev147157. doi: 10.1242/dev.147157.
40. Nakashima K, Takizawa T, Ochiai W, Yanagisawa M, Hisatsune T, et al. (2001) BMP2-mediated alteration in the developmental pathway of fetal mouse brain cells from neurogenesis to astrocytogenesis. *PNAS* 98: 5868–5873.
41. Komada M, Saitsu H, Kinboshi M, Miura T, Shiota K, et al. (2008) Hedgehog signaling is involved in development of the neocortex. *Development* 135: 2717–2727.
42. Dave RK, Ellis T, Toumpas MC, Robson JP, Julian E, et al. (2011) Sonic Hedgehog and Notch Signaling Can Cooperate to Regulate Neurogenic Divisions of Neocortical Progenitors. *PLoS ONE* 6: e14680. doi:10.1371/journal.pone.0014680.
43. Woods CG, Parker A (2013) Investigating microcephaly. *Arch Dis Child* 98: 707–713.
44. Bond J, Roberts E, Mochida GH, Hampshire DJ, Scott S, et al. (2002) ASPM is a major determinant of cerebral cortical size. *Nat Genet* 32: 316–320.
45. Jamieson CR, Fryns J-P, Jacobs J, Matthijs G, Abramowicz MJ (2000) Primary Autosomal Recessive Microcephaly: *MCPH5* Maps to 1q25-q32. *Am J Hum Genet* 67: 1575–1577.
46. Pirozzi F, Nelson B, Mirzaa G (2018) From microcephaly to megalencephaly: determinants of brain size. *Dialogues Clin Neurosci* 20: 267–282.
47. Mahmood S, Ahmad W, Hassan MJ (2011) Autosomal Recessive Primary Microcephaly (MCPH): clinical manifestations, genetic heterogeneity and mutation continuum. *Orphanet J Rare Dis* 6: 39. doi:10.1186/1750-1172-6-39.
48. Jackson AP, Eastwood H, Bell SM, Adu J, Toomes C, et al. (2002) Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am J Hum Genet* 71: 136–142.
49. Liang Y, Gao H, Lin S-Y, Peng G, Huang X, et al. (2010) BRIT1/MCPH1 Is Essential for Mitotic and Meiotic Recombination DNA Repair and Maintaining Genomic Stability in Mice. *PLoS Genet* 6: e1000826. doi:10.1371/journal.pgen.1000826.
50. Gruber R, Zhou Z, Sukchev M, Joeris T, Frappart P-O, et al. (2011) MCPH1 regulates the neuroprogenitor division mode by coupling the centrosomal cycle with mitotic entry through the Chk1–Cdc25 pathway. *Nat Cell Biol* 13: 1325–1334.
51. Nicholas AK, Swanson EA, Cox JJ, Karbani G, Malik S, et al. (2009) The molecular landscape of ASPM mutations in primary microcephaly. *Journal of Medical Genetics* 46: 249–253.

## References

52. Fish JL, Kosodo Y, Enard W, Paabo S, Huttner WB (2006) Aspm specifically maintains symmetric proliferative divisions of neuroepithelial cells. *Proc Natl Acad Sci USA* 103: 10438–10443.
53. Buchman JJ, Durak O, Tsai LH (2011) ASPM regulates Wnt signaling pathway activity in the developing brain. *Genes & Development* 25: 1909–1914.
54. Shamseldin HE, Elfaki M, Alkuraya FS (2012) Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation. *Journal of Medical Genetics* 49: 184–186.
55. Shaheen R, Faqeih E, Ansari S, Abdel-Salam G, Al-Hassnan ZN, et al. (2014) Genomic analysis of primordial dwarfism reveals novel disease genes. *Genome Res* 24: 291–299.
56. O'Driscoll M, Cerosaletti KM, Girard P-M, Dai Y, Stumm M, et al. (2001) DNA Ligase IV Mutations Identified in Patients Exhibiting Developmental Delay and Immunodeficiency. *Mol Cell* 8: 1175–1185.
57. Gatz SA, Ju L, Gruber R, Hoffmann E, Carr AM, et al. (2011) Requirement for DNA Ligase IV during Embryonic Neuronal Development. *J Neurosci* 31: 10088–10100.
58. Gao Y, Sun Y, Frank KM, Dikkes P, Fujiwara Y, et al. (1998) A Critical Role for DNA End-Joining Proteins in Both Lymphogenesis and Neurogenesis. *Cell* 95: 891–902.
59. Deans B, Grifn CS, Maconochie M, Thacker J (2000) *Xrcc2* is required for genetic stability, embryonic neurogenesis and viability in mice. *The EMBO Journal* 19: 6675–6685.
60. Poirier K, Lebrun N, Broix L, Tian G, Saillour Y, et al. (2013) Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly. *Nat Genet* 45: 639–647.
61. Jaglin XH, Poirier K, Saillour Y, Buhler E, Tian G, et al. (2009) Mutations in the  $\beta$ -tubulin gene TUBB2B result in asymmetrical polymicrogyria. *Nat Genet* 41: 746–752.
62. Sun D, Zhou X, Yu H-L, He X-X, Guo W-X, et al. (2017) Regulation of neural stem cell proliferation and differentiation by Kinesin family member 2a. *PLoS ONE* 12: e0179047. doi:10.1371/journal.pone.0179047.
63. Kadir R, Harel T, Markus B, Perez Y, Bakhrat A, et al. (2016) ALFY-Controlled DVL3 Autophagy Regulates Wnt Signaling, Determining Human Brain Size. *PLoS Genet* 12: e1005919. doi:10.1371/journal.pgen.1005919.
64. Mueller T, Donga Z, Berberoglu MA, Guoa S (2011) The Dorsal Pallium in Zebrafish, *Danio rerio* (Cyprinidae, Teleostei). *Brain Research* 1381: 95–105.
65. Sorrells S, Toruno C, Stewart RA, Jette C (2013) Analysis of Apoptosis in Zebrafish Embryos by Whole-mount Immunofluorescence to Detect Activated Caspase 3. *J. Vis. Exp.* 82: e51060. doi:10.3791/51060.
66. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* 203: 253–310.
67. Parichy DM, Elizondo MR, Mills MG, Gordon TN, Engeszer RE (2009) Normal table of postembryonic zebrafish development: Staging by externally visible anatomy of the living fish. *Dev Dyn* 238: 2975–3015.
68. Nakayama T, Al-Maawali A, El-Quessny M, Rajab A, Khalil S, et al. (2015) Mutations in PYCR2, Encoding Pyrroline-5-Carboxylate Reductase 2, Cause Microcephaly and Hypomyelination. *The American Journal of Human Genetics* 96: 709–719.
69. Perez Y, Bar-Yaacov R, Kadir R, Wormser O, Shelef I, et al. (2019) Mutations in the

## References

- microtubule-associated protein MAP11 (C7orf43) cause microcephaly in humans and zebrafish. *Brain* 142: 574–585.
70. Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current Opinion in Cell Biology* 11: 699-704.
  71. Van de Peer Y, Taylor JS, Joseph J, Meyer A (2002) Wanda: a database of duplicated fish genes. *Nucleic Acids Research* 30: 109–112.
  72. Sun X-J, Xu P-F, Zhou T, Hu M, Fu C-T, et al. (2008) Genome-Wide Survey and Developmental Expression Mapping of Zebrafish SET Domain-Containing Genes. *PLoS ONE* 3: e1499. doi:10.1371/journal.pone.0001499.
  73. Laprairie RB, Denovan-Wright EM, Wright JM (2016) Subfunctionalization of peroxisome proliferator response elements accounts for retention of duplicated fabp1 genes in zebrafish. *BMC Evol Biol* 6:147. doi: 10.1186/s12862-016-0717-x.
  74. Nguyen H, Kerimoglu C, Pirouz M, Pham L, Kiszka KA, et al. (2018) Epigenetic Regulation by BAF Complexes Limits Neural Stem Cell Proliferation by Suppressing Wnt Signaling in Late Embryonic Development. *Stem Cell Reports* 10: 1734–1750.
  75. Tuoc TC, Boretius S, Sansom SN, Pitulescu M-E, Frahm J, et al. (2013) Chromatin Regulation by BAF170 Controls Cerebral Cortical Size and Thickness. *Developmental Cell* 25: 256–269.
  76. Gorski JA, Talley T, Qiu M, Puelles L, Rubenstein JLR, et al. (2002) Cortical Excitatory Neurons and Glia, But Not GABAergic Neurons, Are Produced in the Emx1-Expressing Lineage. *J Neurosci* 22: 6309–6314.
  77. Cecchi C, Boncinelli E (2000) Emx homeogenes and mouse brain development. *Cell Stem Cell* 23: 347–352.
  78. Yang YJ, Baltus AE, Mathew RS, Murphy EA, Evrony GD, et al. (2012) Microcephaly Gene Links Trithorax and REST/NRSF to Control Neural Stem Cell Proliferation and Differentiation. *Cell* 151: 1097–1112.
  79. McIntyre RE, Lakshminarasimhan Chavali P, Ismail O, Carragher DM, Sanchez-Andrade G, et al. (2012) Disruption of Mouse Cenpj, a Regulator of Centriole Biogenesis, Phenocopies Seckel Syndrome. *PLoS Genet* 8: e1003022. doi:10.1371/journal.pgen.1003022.
  80. Marjanovic M, Sanchez-Huertas C, Terre B, Gomez R, Scheel JF, et al. (2015) CEP63 deficiency promotes p53-dependent microcephaly and reveals a role for the centrosome in meiotic recombination. *Nature Communications* 6:7676. doi: 10.1038/ncomms8676.
  81. Pulvers JN, Bryk J, Fish JL, Wilsch-Brauninger M, Arai Y, et al. (2010) Mutations in mouse *Aspm* (abnormal spindle-like microcephaly associated) cause not only microcephaly but also major defects in the germline. *Proc Natl Acad Sci USA* 107: 16595–16600.
  82. Barrera JA, Kao L-R, Hammer RE, Seemann J, Fuchs JL, et al. (2010) CDK5RAP2 Regulates Centriole Engagement and Cohesion in Mice. *Developmental Cell* 18: 913–926.
  83. Johnson MB, Sun X, Kodani A, Borges-Monroy R, Girsakis KM, et al. (2018) *Aspm* knockout ferret reveals an evolutionary mechanism governing cerebral cortical size. *Nature* 556: 370–375.
  84. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, et al. (2007) Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* 131: 861–

## References

- 872.
85. Lancaster MA, Renner M, Martin C-A, Wenzel D, Bicknell LS, et al. (2013) Cerebral organoids model human brain development and microcephaly. *Nature* 501: 373–379.
  86. Kalay E, Yigit G, Aslan Y, Brown KE, Pohl E, et al. (2011) CEP152 is a genome maintenance protein disrupted in Seckel syndrome. *Nat Genet* 43: 23–26.
  87. Zhang W, Yang S-L, Yang M, Herrlinger S, Shao Q, et al. (2019) Modeling microcephaly with cerebral organoids reveals a WDR62-CEP170-KIF2A pathway promoting cilium disassembly in neural progenitors. *Nature Communications*: 10: 2612. doi: 10.1038/s41467-019-10497-2
  88. Tang H, Hammack C, Ogden SC, Wen Z, Qian X, et al. (2016) Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. *Stem Cell* 18: 587–590.
  89. Portela A, Esteller M (2010) Epigenetic modifications and human disease. *Nat Biotechnol* 28: 1057–1068.
  90. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, et al. (2006) The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439: 871–874.
  91. Jin B, Li Y, Robertson KD (2011) DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? *Genes & Cancer* 2: 607–617.
  92. Nicholson TB, Veland N, Chen T (2015) Chapter 3. Writers, Readers, and Erasers of Epigenetic Marks. *Epigenetic Cancer Therapy*: 31-66. doi:10.1016/B978-0-12-800206-3.00003-3.
  93. Bestor TH (2000) The DNA methyltransferases of mammals. *Human Molecular Genetics* 9: 2395–2402.
  94. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462: 315–322.
  95. Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21: 381–395.
  96. Schuettengruber B, Bourbon H-M, Di Croce L, Cavalli G (2017) Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* 171: 34–57.
  97. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441: 349–353.
  98. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, et al. (2007) High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell Reports* 129: 823–837.
  99. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, et al. (2005) Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse. *Cell Reports* 120: 169–181.
  100. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, et al. (2006) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125: 315–326.
  101. Ringrose L, Paro R (2004) Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annu Rev Genet* 38: 413–443.
  102. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G (2007) Genome



## References

- Regulation by Polycomb and Trithorax Proteins. *Cell* 128: 735–745.
103. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553–560.
  104. Burgold T, Spreafico F, De Santa F, Totaro MG, Prosperini E, et al. (2008) The Histone H3 Lysine 27-Specific Demethylase Jmjd3 Is Required for Neural Commitment. *PLoS ONE* 3: e3034. doi:10.1371/journal.pone.0003034.
  105. Golebiewska A, Atkinson SP, Lako M, Armstrong L (2009) Epigenetic Landscaping During hESC Differentiation to Neural Cells. *Stem Cells* 27: 1298–1308.
  106. Hirabayashi Y, Suzki N, Tsuboi M, Endo TA, Toyoda T, et al. (2009) Polycomb Limits the Neurogenic Competence of Neural Precursor Cells to Promote Astrogenic Fate Transition. *Neuron* 63: 600–613.
  107. Narayanan R, Pirouz M, Kerimoglu C, Pham L, Wagener RJ, et al. (2015) Loss of BAF (mSWI/SNF) Complexes Causes Global Transcriptional and Chromatin State Changes in Forebrain Development. *Cell Rep* 13: 1842–1854.
  108. Bochyńska A, Lüscher-Firzlaff J, Lüscher B (2018) Modes of Interaction of KMT2 Histone H3 Lysine 4 Methyltransferase/COMPASS Complexes with Chromatin. *Cells* 7:17. doi: 10.3390/cells7030017.
  109. Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, et al. (2001) COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci USA* 98: 12902–12907.
  110. Mohan M, Herz HM, Smith ER, Zhang Y, Jackson J, et al. (2011) The COMPASS Family of H3K4 Methylases in *Drosophila*. *Molecular and Cellular Biology* 31: 4310–4318.
  111. Shilatifard A (2008) Molecular Implementation and Physiological Roles for Histone H3 Lysine 4 (H3K4) Methylation. *Current Opinion in Cell Biology* 20: 341–348.
  112. Hughes CM, Rozenblatt-Rosen O, Milne TA, Copeland TD, Levine SS, et al. (2004) Menin Associates with a Trithorax Family Histone Methyltransferase Complex and with the Hoxc8 Locus. *Mol Cell* 13: 587–597.
  113. Pijnappel WWMP, Schaft D, Roguev A, Shevchenko A, Tekotte H, et al. (2001) The *S. cerevisiae* SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes & Development* 15: 2991–3004.
  114. Cho Y-W, Hong T, Hong S, Guo H, Yu H, et al. (2007) PTIP Associates with MLL3- and MLL4-containing Histone H3 Lysine 4 Methyltransferase Complex. *Journal of Biological Chemistry* 282: 20395–20406.
  115. Yokoyama A, Wang Z, Wysocka J, Sanyal M, Aufiero DJ, et al. (2004) Leukemia Proto-Oncoprotein MLL Forms a SET1-Like Histone Methyltransferase Complex with Menin To Regulate. *Molecular and Cellular Biology* 24: 5639-5649.
  116. Patel SR, Kim D, Levitan I, Dressler GR (2007) The BRCT-Domain Containing Protein PTIP Links PAX2 to a Histone H3, Lysine 4 Methyltransferase Complex. *Developmental Cell* 13: 580–592.
  117. Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, et al. (2008) Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS. *Molecular and Cellular Biology* 28: 7337–7344.
  118. Lee J-H, Tate CM, You J-S, Skalnik DG (2007) Identification and Characterization of the Human Set1BHistone H3-Lys. *J Biol Chem* 282: 13419–13428.

## References

119. Lee J-H, Skalnik DG (2005) CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. *J Biol Chem* 280: 41725–41731.
120. Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, et al. (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* 13: 713–719.
121. Patel A, Dharmarajan V, Vought VE, Cosgrove MS (2009) On the Mechanism of Multiple Lysine Methylation by the Human Mixed Lineage Leukemia Protein-1 (MLL1) Core Complex. *Journal of Biological Chemistry* 284: 24242–24256.
122. Zhang P, Lee H, Brunzelle JS, Couture J-F (2012) The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Research* 40: 4237–4246.
123. Zhang P, Chaturvedi C-P, Tremblay V, Cramet M, Brunzelle JS, et al. (2015) A phosphorylation switch on RbBP5 regulates histone H3 Lys4 methylation. *Genes & Development* 29: 123–128.
124. Zhang Y, Mittal A, Reid J, Reich S, Gamblin SJ, et al. (2015) Evolving Catalytic Properties of the MLL Family SET Domain. *Structure* 23: 1921-1933.
125. Milne TA, Hughes CM, Lloyd R, Yang Z, Rozenblatt-Rosen O, et al. (2005) Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc Natl Acad Sci USA* 102: 749–754.
126. Ullius A, Luscher-Firzlaff J, Costa IG, Walsemann G, Forst AH, et al. (2014) The interaction of MYC with the trithorax protein ASH2L promotes gene transcription by regulating H3K27 modification. *Human Molecular Genetics* 42: 6901–6920.
127. Denissov S, Hofemeister H, Marks H, Kranz A, Ciotta G, et al. (2014) Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* 141: 526–537.
128. Hu D, Garruss AS, Gao X, Morgan MA, Cook M, et al. (2013) The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol* 20: 1093–1097.
129. Bach C, Mueller D, Buhl S, Garcia-Cuellar MP, Slany RK (2008) Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. *Oncogene* 28: 815–823.
130. Bledau AS, Schmidt K, Neumann K, Hill U, Ciotta G, et al. (2014) The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development* 141: 1022–1035.
131. Jang Y, Wang C, Zhuang L, Liu C, Ge K (2017) H3K4 Methyltransferase Activity Is Required for MLL4 Protein Stability. *J Mol Biol* 429: 2046–2054.
132. Placek K, Hu G, Cui K, Zhang D, Ding Y, et al. (2017) MLL4 prepares the enhancer landscape for *Foxp3* induction via chromatin looping. *Nat Immunol* 18: 1035–1045.
133. Andreu-Vieyra CV, Chen R, Agno JE, Glaser S, Anastassiadis K, et al. (2010) MLL2 Is Required in Oocytes for Bulk Histone 3 Lysine 4 Trimethylation and Transcriptional Silencing. *PLoS Biol* 8: e1000453. doi:10.1371/journal.pbio.1000453.
134. Wang P, Lin C, Smith ER, Guo H, Sanderson BW, et al. (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: 6074-6085.

## References

135. Alicea-Velázquez NL, Shinsky SA, Loh DM, Lee J-H, Skalnik DG, et al. (2016) Targeted Disruption of the Interaction between WD-40 Repeat Protein 5 (WDR5) and Mixed Lineage Leukemia (MLL)/SET1 Family Proteins Specifically Inhibits MLL1 and SETd1A Methyltransferase Complexes. *Journal of Biological Chemistry* 291: 22357–22372.
136. Patel A, Vought VE, Dharmarajan V, Cosgrove MS (2008) A Conserved Arginine-containing Motif Crucial for the Assembly and Enzymatic Activity of the Mixed Lineage Leukemia Protein-1 Core Complex. *Journal of Biological Chemistry* 283: 32162–32175.
137. Hu D, Gao X, Cao K, Morgan MA, Mas G, et al. (2017) Not All H3K4 Methylations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification. *Mol Cell* 65: 460–475.
138. Herz H-M, Mohan M, Garruss AS, Liang K, Takahashi Y-H, et al. (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *26*: 2604–2620.
139. Rickels R, Herz H-M, Sze CC, Cao K, Morgan MA, et al. (2017) Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat Genet* 49: 1647-1653.
140. Guo C, Chen LH, Huang Y, Chang C-C, Wang P, et al. (2013) KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. *Oncotarget* 4: 2144–2153.
141. Lee J-E, Wang C, Xu S, Cho Y-W, Wang L, et al. (2013) H3K4 mono- and dimethyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* 2: e01503. doi: 10.7554/eLife.01503.
142. Deng C, Li Y, Liang S, Cui K, Salz T, et al. (2013) USF1 and hSET1A Mediated Epigenetic Modifications Regulate Lineage Differentiation and HoxB4 Transcription. *PLoS Genet* 9: e1003524. doi:10.1371/journal.pgen.1003524.
143. Rao RC, Dou Y (2015) Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer* 15: 334–346.
144. Bögershausen N, Bruford E, Wollnik B (2013) Skirting the pitfalls: a clear-cut nomenclature for H3K4 methyltransferases. *Clin Genet* 83: 212–214.
145. Ernst P, Mabon M, Davidson AJ, Zon LI, Korsmeyer SJ (2004) An Mll-Dependent Hox Program Drives Hematopoietic Progenitor Expansion. *Current Biology* 14: 2063–2069.
146. Glaser S, Schaft J, Lubitz S, Vintersten K, van der Hoeven F, et al. (2006) Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* 133: 1423–1432.
147. Yu BD, Hess JL, Horning SE, Brown GAJ, Korsmeyer SJ (1995) Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* 378: 505–508.
148. Zech M, Boesch S, Maier EM, Borggraefe I, Vill K, et al. (2016) Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. *The American Journal of Human Genetics* 99: 1377–1387.
149. Jones WD, Dafou D, McEntagart M, Woollard WJ, Elmslie FV, et al. (2012) De Novo Mutations in MLL Cause Wiedemann-Steiner Syndrome. *Cell Reports* 91: 358–364.
150. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, et al. (2012) Disruption of an EHMT1-Associated Chromatin-Modification Module Causes Intellectual Disability. *Cell Reports* 91: 73–82.
151. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, et al. (2010) Exome

## References

- sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 42: 790–793.
152. Takata A, Xu B, Ionita-Laza I, Roos JL, Gogos JA, et al. (2014) Loss-of-Function Variants in Schizophrenia Risk and SETD1A as a Candidate Susceptibility Gene. *Neuron* 82: 773–780.
153. Li Y, Bögershausen N, Alanay Y, Simsek Kiper PÖ, Plume N, et al. (2011) A mutation screen in patients with Kabuki syndrome. *Hum Genet* 130: 715–724.
154. Gorman KM, Meyer E, Kurian MA (2017) Review of the phenotype of early-onset generalised progressive dystonia due to mutations in KMT2B. *European Journal of Paediatric Neurology* 22: 245-256.
155. Chen C, Liu Y, Rappaport AR, Kitzing T, Schultz N, et al. (2014) MLL3 Is a Haploinsufficient 7q Tumor Suppressor in Acute Myeloid Leukemia. *Cancer Cell* 25: 652–665.
156. Liu H, Cheng EHY, Hsieh JJD (2009) MLL fusions: Pathways to leukemia. *Cancer Biology & Therapy* 8: 1204–1211.
157. Meyer C, Burmeister T, ger DGO, Tsaur G, Fechina L, et al. (2017) The MLL recombinome of acute leukemias in 2017: *Leukemia* 32: 273-284.
158. Xu J, Li L, Xiong J, denDekker A, Ye A, et al. (2016) MLL1 and MLL1 fusion proteins have distinct functions in regulating leukemic transcription program. *Cell Discov* 2: 16008. doi:10.1038/celldisc.2016.8
159. Liu H, Takeda S, Kumar R, Westergard TD, Brown EJ, et al. (2010) Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature* 467: 343–346.
160. FitzGerald KT, Diaz MO (1999) *MLL2*: A New Mammalian Member of the *trx/MLL* Family of Genes. *Genomics* 59: 187–192.
161. Wang J, Muntean AG, Wu L, Hess JL (2012) A Subset of Mixed Lineage Leukemia Proteins Has PlantHomeodomain (PHD)-mediated E3 Ligase Activity. *J Biol Chem* 287: 43410–43416.
162. Takeda S, Chen DY, Westergard TD, Fisher JK, Rubens JA, et al. (2006) Proteolysis of MLL family proteins is essential for Taspase1-orchestrated cell cycle progression. *Genes & Development* 20: 2397–2409.
163. Hsieh JJD, Ernst P, Erdjument-Bromage H, Tempst P, Korsmeyer SJ (2003) Proteolytic Cleavage of MLL Generates a Complex of N- and C-Terminal Fragments That Confers Protein Stability and Subnuclear Localization. *Molecular and Cellular Biology* 23: 186–194.
164. Hsieh JJD, Cheng EHY, Korsmeyer SJ (2003) Taspase1: A Threonine Aspartase Required for Cleavage of MLL and Proper. *Cell* 115: 293–303.
165. Yokoyama A, Ficara F, Murphy MJ, Meisel C, Hatanaka C, et al. (2013) MLL Becomes Functional through Intra-Molecular Interaction Not by Proteolytic Processing. *PLoS ONE* 8: e73649. doi:10.1371/journal.pone.0073649.
166. Singh AM, Sun Y, Li L, Zhang W, Wu T, et al. (2015) Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. *Stem Cell Reports* 5: 323–336.
167. Singh AM, Chappell J, Trost R, Lin L, Wang T, et al. (2013) Cell-Cycle Control of Developmentally Regulated Transcription Factors Accounts for Heterogeneity in Human

## References

- Pluripotent Cells. *Stem Cell Reports* 1: 532–544.
168. Ansari KI, Kasiri S, Mishra BP, Mandal SS (2012) Mixed lineage leukaemia-4 regulates cell-cycle progression and cell viability and its depletion suppresses growth of xenografted tumour in vivo. *British Journal of Cancer* 107: 315–324.
169. Ali A, Veeranki SN, Tyagi S (2014) A SET-domain-independent role of WRAD complex in cell-cycle regulatory function of mixed lineage leukemia. *Nucleic Acids Research* 42: 7611–7624.
170. Lubitz S, Glaser S, Schaft J, Stewart AF, Anastassiadis K (2007) Increased Apoptosis and Skewed Differentiation in Mouse Embryonic Stem Cells Lacking the Histone Methyltransferase Mll2. *Molecular Biology of the Cell* 18: 2356–2366.
171. Glaser S, Lubitz S, Loveland KL, Ohbo K, Robb L, et al. (2009) The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics Chromatin* 2: 5. doi:10.1186/1756-8935-2-5.
172. Tomizawa S-I, Kobayashi Y, Shirakawa T, Watanabe K, Mizoguchi K, et al. (2018) Kmt2b conveys monovalent and bivalent H3K4me3 in mouse spermatogonial stem cells at germline and embryonic promoters. *Development* 145: dev169102. doi: 10.1242/dev.169102.
173. Kerimoglu C, Agis-Balboa RC, Kranz A, Stilling R, Bahari-Javan S, et al. (2013) Histone-methyltransferase MLL2 (KMT2B) is required for memory formation in mice. *J Neurosci* 33: 3452–3464.
174. Kerimoglu C, Sakib MS, Jain G, Benito E, Burkhardt S, et al. (2017) KMT2A and KMT2B Mediate Memory Function by Affecting Distinct Genomic Regions. *Cell Reports* 20: 538–548.
175. Goldsworthy M, Absalom NL, Schröter D, Matthews HC, Bogani D, et al. (2013) Mutations in Mll2, an H3K4 Methyltransferase, Result in Insulin Resistance and Impaired Glucose Tolerance in Mice. *PLoS ONE* 8: e61870. doi: 10.1371/journal.pone.0061870.
176. Meyer E, Carss KJ, Rankin J, Nichols JME, Grozeva D, et al. (2017) Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. *Nat Genet* 49: 223–237.
177. Barbagiovanni G, Germain P-L, Zech M, Atashpaz S, Pietro Lo Riso, et al. (2018) KMT2B Is Selectively Required for Neuronal Transdifferentiation, and Its Loss Exposes Dystonia Candidate Genes. *Cell Rep* 25: 988–1001.
178. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467.
179. Rosin N, Elcioglu NH, Beleggia F, Isgüven P, Altmüller J, et al. (2015) Mutations in XRCC4 cause primary microcephaly, short stature and increased genomic instability. *Human Molecular Genetics* 24: 3708-3717.
180. Alexa A, Rahnenfuhrer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Human Molecular Genetics* 22: 1600–1607.
181. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, et al. (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* 150: 76–85.
182. Westerfield M (2000) *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. 4th Edition, University of Oregon Press, Eugene.

## References

183. Varshney GV, Carrington B, Pei W, Bishop K, Chen Z, et al. (2016) A High-Throughput functional genomics workflow based on CRISPR/Cas9-mediated targeted mutagenesis in zebrafish. *Nat Protoc* 11: 2357-2375.
184. Moreno-Mateos MA, Vejnár CE, Beaudoin J-D, Fernández JP, Mis EK, et al. (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR/Cas9 targeting in vivo. *Nat Methods* 12: 982–988.
185. Friez MJ, Brooks SS, Stevenson RE, Field M, Basehore MJ, et al. (2016) HUWE1 mutations in Juberg-Marsidi and Brooks syndromes: the results of an X-chromosome exome sequencing study. *BMJ Open* 6: e009537. doi: 10.1136/bmjopen-2015-009537.
186. Froyen G, Corbett M, Vandewalle J, Jarvela I, Lawrence O, et al. (2008) Submicroscopic Duplications of the Hydroxysteroid Dehydrogenase HSD17B10 and the E3 Ubiquitin Ligase HUWE1 Are Associated with Mental Retardation. *The American Journal of Human Genetics* 82: 432–443.
187. Chen D, Kon N, Li M, Zhang W, Qin J, et al. (2005) ARF-BP1/Mule Is a Critical Mediator of the ARF Tumor Suppressor. *Cell* 121: 1071–1083.
188. Tarpey P, Parnau J, Blow M, Woffendin H, Bignell G, et al. (2004) Mutations in the DLG3 Gene Cause Nonsyndromic X-Linked Mental Retardation. *Am J Hum Genet* 75: 318–324.
189. Crocker-Buque A, Currie SP, Luz LL, Grant SG, Duffy KR, et al. (2016) Altered thalamocortical development in the SAP102 knockout model of intellectual disability. *Human Molecular Genetics* 25: 4052–4061.
190. Biesecker LG, Spinner NB (2013) A genomic view of mosaicism and human disease. *Nature Reviews Genetics* 14: 307–320.
191. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, et al. (2015) Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. *The American Journal of Human Genetics* 97: 67–74.
192. Qin L, Wang J, Tian X, Yu H, Truong C, et al. (2016) Detection and Quantification of Mosaic Mutations in Disease Genes by Next-Generation Sequencing. *The Journal of Molecular Diagnostics* 18: 446–453.
193. van der Lelij P, Chrzanowska KH, Godthelp BC, Rooimans MA, Oostra AB, et al. (2010) Warsaw Breakage Syndrome, a Cohesinopathy Associated with Mutations in the XPD Helicase Family Member DDX11/ChIR1. *Cell Reports* 86: 262–266.
194. Parish JL, Rosa J, Wang X, Lahti JM, Doxsey SJ, et al. (2006) The DNA helicase ChIR1 is required for sister chromatid cohesion in mammalian cells. *Journal of Cell Science* 119: 4857–4865.
195. Alkhunaizi E, Shaheen R, Bharti SK, Joseph-George AM, Chong K, et al. (2018) Warsaw breakage syndrome: Further clinical and genetic delineation. *Am J Med Genet* 176: 2404–2418.
196. Dorsman JC, Levitus M, Rockx D, Rooimans MA, Oostra AB, et al. (2007) Identification of the Fanconi Anemia Complementation Group I Gene, FANCI. *Cellular Oncology* 29: 211-218.
197. Lu J, Li H, Hu M, Sasaki T, Baccei A, et al. (2014) The Distribution of Genomic Variations in Human iPSCs Is Related to Replication-Timing Reorganization during Reprogramming. *Cell Reports* 7: 70–78.
198. Brandl C, Zimmermann SJ, Milenkovic VM, Rosendahl SMG, Grassmann F, et al. (2014) In-Depth Characterisation of Retinal Pigment Epithelium (RPE) Cells Derived from

## References

- Human Induced Pluripotent Stem Cells (hiPSC). *Neuromolecular Medicine* 16: 551–564.
199. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, et al. (2005) A high-resolution map of active promoters in the human genome. *Nature* 436: 876–880.
200. Boles NC, Hirsch SE, Le S, Corneo B, Najm F, et al. (2014) NPTX1 Regulates Neural Lineage Specification from Human Pluripotent Stem Cells. *Cell Reports* 6: 724–736.
201. Davidson KC, Adams AM, Goodson JM, McDonald CE, Potter JC, et al. (2012) Wnt/ $\beta$ -catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. *Proc Natl Acad Sci USA* 109: 4485–4490.
202. Abu-Remaileh M, Gerson A, Farago M, Nathan G, Alkalay I, et al. (2010) Oct-3/4 regulates stem cell identity and cell fate decisions by modulating Wnt/ $\beta$ -catenin signalling. *The EMBO Journal* 29: 3236–3248.
203. Schneikert J, Chandra SHV, Ruppert JG, Ray S, Wenzel EM, et al. (2013) Functional Comparison of Human Adenomatous Polyposis Coli (APC) and APC-Like in Targeting Beta-Catenin for Degradation. *PLoS ONE* 8: e68072. doi:10.1371/journal.pone.0068072.
204. Lee H-J, Finkelstein D, Li X, Wu D, Shi D-L, et al. (2010) Identification of Transmembrane Protein 88 (TMEM88) as a Dishevelled-binding Protein. *Journal of Biological Chemistry* 285: 41549–41556.
205. Rubin EM, Guo Y, Tu K, Xie J, Zi X, et al. (2010) Wnt Inhibitory Factor 1 Decreases Tumorigenesis and Metastasis in Osteosarcoma. *Mol Cancer Ther* 9: 731–741.
206. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28: 27–30.
207. Motoyama N, Wang F, Roth K, Sawa H, Nakayama K, et al. (1995) Massive cell death of immature hematopoietic cells and neurons in Bcl-x-deficient mice. *Science* 267: 1506–1510.
208. Schuler M, Bossy-Wetzel E, Goldstein JC, Fitzgerald P, Green DR (2000) p53 Induces Apoptosis by Caspase Activation through Mitochondrial Cytochrome c Release. *J Biol Chem* 275: 7337–7342.
209. Jobst-Schwan T, Schmidt JM, Schneider R, Hoogstraten CA, Ullmann JFP, et al. (2018) Acute multi-sgRNA knockdown of KEOPS complex genes reproduces the microcephaly phenotype of the stable knockout zebrafish model. *PLoS ONE* 13: e0191503. doi:10.1371/journal.pone.0191503.s012.
210. Vaz R, Hofmeister W, Lindstrand AA (2019) Zebrafish Models of Neurodevelopmental Disorders: Limitations and Benefits of Current Tools and Techniques. *Int J Mol Sci* 20: E1296. doi:10.3390/ijms20061296.
211. Martynoga B, Drechsel D, Guillemot F (2012) Molecular Control of Neurogenesis: A View from the Mammalian Cerebral Cortex. *Cold Spring Harbor Perspectives in Biology* 4: a008359. doi:10.1101/cshperspect.a008359.
212. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, et al. (2019) The single cell transcriptional landscape of mammalian organogenesis. *Nature* 566: 496–502.
213. Gulisano M, Broccoli V, Pardini C, Boncinelli E (1996) Emx1 and Emx2 Show Different Patterns of Expression During Proliferation and Differentiation of the Developing Cerebral Cortex in the Mouse. *European Journal of Neuroscience* 8: 1037–1050.
214. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36: 411–420.

## References

215. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, et al. (2019) Comprehensive Integration of Single-Cell Data. *Cell* 177: 1888–1902.e21. doi:10.1016/j.cell.2019.05.031.
216. Chen VS, Morrison JP, Southwell MF, Foley JF, Bolon B, et al. (2017) Histology Atlas of the Developing Prenatal and Postnatal Mouse Central Nervous System, with Emphasis on Prenatal Days E7.5 to E18.5. *Toxicologic Pathology* 45: 705–744.
217. Hutton SR, Pevny LH (2011) SOX2 expression levels distinguish between neural progenitor populations of the developing dorsal telencephalon. *Developmental Biology* 352: 40–47.
218. Loo L, Simon JM, Xing L, McCoy ES, Niehaus JK, et al. (2019) Single-cell transcriptomic analysis of mouse neocortical development. *Nature Communications* 10: 134. doi: 10.1038/s41467-018-08079-9.
219. Götz M, Stoykova A, Gruss P (1998) *Pax6* Controls Radial Glia Differentiation in the Cerebral Cortex. *Neuron* 21: 1031–1044.
220. Wonders CP, Anderson SA (2006) The origin and specification of cortical interneurons. *Nat Rev Neurosci* 7: 687–696. doi:10.1038/nrn892.
221. Chen Y-JJ, Friedman BA, Ha C, Durinck S, Liu J, et al (2017) Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types. *Scientific Reports* 7: 45656. doi: 10.1038/srep45656.
222. Batista-Brito R, Machold R, Klein C, Fishell G (2008) Gene Expression in Cortical Interneuron Precursors is Prescient of their Mature Function. *Cerebral Cortex* 18: 2306–2317.
223. Tole S, Ragsdale CW, Grove EA (2002) Dorsoventral Patterning of the Telencephalon Is Disrupted in the Mouse Mutant *extra-toes* (J). *Developmental Biology* 217: 254–265.
224. Englund C, Fink A, Lau C, Pham D, Daza RAM, et al. (2005) *Pax6*, *Tbr2*, and *Tbr1* Are Expressed Sequentially by Radial Glia, Intermediate Progenitor Cells, and Postmitotic Neurons in Developing Neocortex. *J Neurosci* 25: 247–251.
225. Haubensak W, Attardo A, Denk W, Huttner WB (2004) Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: A major site of neurogenesis. *Proc Natl Acad Sci* 101: 3196–3201.
226. Preissl S, Fang R, Huang H, Zhao Y, Raviram R, et al. (2018) Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat Neurosci* 21: 432–439.
227. Chan CH, Godinho N, Thomaidou D, Tan SS, Gulisano M, Parnavelas JG (2001) *Emx1* is a Marker for Pyramidal Neurons of the Cerebral Cortex. *Human Molecular Genetics* 11: 1191–1198.
228. Haque A, Engel J, Teichmann SA, Lönnberg T (2017) A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 9: 75. doi:10.1186/s13073-017-0467-4.
229. Tracy S, Yuan G-C, Dries R (2019) RESCUE: imputing dropout events in single-cell RNA-sequencing data. *BMC Bioinformatics* 20: 388. doi:10.1186/s12859-019-2977-0.
230. Hagen von der M, Pivarcsi M, Liebe J, Bernuth von H, Didonato N, et al. (2014) Diagnostic approach to microcephaly in childhood: a two-center study and review of the literature. *Dev Med Child Neurol* 56: 732–741.
231. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536: 285–291.



## References

232. Dixon-Salazar TJ, Silhavy JL, Udpa N, Schroth J, Bielas S, et al. (2012) Exome Sequencing Can Improve Diagnosis and Alter Patient Management. *Science Translational Medicine* 4: 138ra78. doi:10.1126/scitranslmed.3003544.
233. Rump P, Jazayeri O, van Dijk-Bos KK, Johansson LF, van Essen AJ, et al. (2016) Whole-exome sequencing is a powerful approach for establishing the etiological diagnosis in patients with intellectual disability and microcephaly. *BMC Medical Genomics* 9:7. doi: 10.1186/s12920-016-0167-8.
234. Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, et al. (2016) Exome Sequencing and the Management of Neurometabolic Disorders. *N Engl J Med* 374: 2246–2255.
235. Mahler EA, Johannsen J, Tsiakas K, Kloth K, Lüttgen S, et al. (2019) Exome sequencing in children. *Dtsch Arztebl Int* 116: 197-204.
236. Kaplan L, Foster R, Shen Y, Parry DM, McMaster ML, et al. (2010) Monozygotic twins discordant for neurofibromatosis 1. *Am J Med Genet* 152A: 601–606.
237. Campbell IM, Yuan B, Robberecht C, Pfundt R, Szafranski P, et al. (2014) Parental Somatic Mosaicism Is Underrecognized and Influences Recurrence Risk of Genomic Disorders. *The American Journal of Human Genetics* 95: 173–182.
238. Lim JS, Kim W-I, Kang H-C, Kim SH, Park AH, et al. (2015) Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. *Nature Medicine* 21: 395-400.
239. Wu Y, Sommers JA, Khan I, de Winter JP, Brosh RM Jr. (2012) Biochemical Characterization of Warsaw Breakage Syndrome Helicase. *Journal of Biological Chemistry* 287: 1007–1021.
240. Costa V, Casamassimi A, Roberto R, Gianfrancesco F, Matarazzo MR, et al. (2009) DDX11L: a novel transcript family emerging from human subtelomeric regions. *BMC Genomics* 10: 250. doi:10.1186/1471-2164-10-250.
241. Vořechovský I (2006) Aberrant 3' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Human Molecular Genetics* 34: 4630–4641.
242. Takahara K, Schwarze U, Imamura Y, Hoffman GG, Toriello H, et al. (2002) Order of Intron Removal Influences Multiple Splice Outcomes, Including a Two-Exon Skip, in a COL5A1 Acceptor-Site Mutation That Results in Abnormal Pro- $\alpha$ 1(V) N-Propeptides and Ehlers-Danlos Syndrome Type I. *Cell Reports* 71: 451–465.
243. Burset M, Seledtsov IA, Solovyev VV (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research* 28: 4364–4375.
244. Wimmer K, Roca X, Beiglböck H, Callens T, Etzler J, et al. (2007) Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5' splice-site disruption. *Hum Mutat* 28: 599–612.
245. Abramowicz A, Gos M (2018) Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics* 59: 253–268.
246. Smogorzewska A, Matsuoka S, Vinciguerra P, McDonald ER III, Hurov KE, et al. (2007) Identification of the FANCI Protein, a Monoubiquitinated FANCD2 Paralog Required for DNA Repair. *Cell Reports* 129: 289–301.
247. Siddiqui MQ, Choudhary RK, Thapa P, Kulkarni N, Rajpurohit YS, et al. (2017) Structural and biophysical properties of h-FANCI ARM repeat protein. *Journal of Biomolecular Structure and Dynamics* 35: 3032-3042.

## References

248. Savage SA, Ballew BJ, Giri N, NCI DCEG Cancer Genomics Research Laboratory, Chandrasekharappa SC, et al. (2015) Novel FANCI mutations in Fanconi anemia with VACTERL association. *Am J Med Genet* 170: 386–391.
249. Ang Y-S, Tsai S-Y, Lee D-F, Monk J, Su J, et al. (2011) Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional Network. *Cell Reports* 145: 183–197.
250. Cui C, Shu W, Li P. (2016) Fluorescence In situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications. *Front Cell Dev Biol* 4:89. doi: 10.3389/fcell.2016.00089.
251. Meisner LF, Johnson JA (2008) Protocols for cytogenetic studies of human embryonic stem cells. *Methods* 45: 133–141.
252. Cheng L, Hansen NF, Zhao L, Du Y, Zou C, et al. (2012) Low Incidence of DNA Sequence Variation in Human Induced Pluripotent Stem Cells Generated by Nonintegrating Plasmid Expression. *Cell Reports* 10: 337–344.
253. Ji J, Ng SH, Sharma V, Neculai D, Hussein S, et al. (2012) Elevated Coding Mutation Rate During the Reprogramming of Human Somatic Cells into Induced Pluripotent Stem Cells. *Stem Cells* 30: 435–440.
254. Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, et al. (2012) Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* 492: 438–442.
255. Mattout A, Biran A, Meshorer E (2011) Global epigenetic changes during somatic cell reprogramming to iPS cells. *Journal of Molecular Cell Biology* 3: 341–350.
256. Koche RP, Smith ZD, Adli M, Gu H, Ku M, et al. (2011) Reprogramming Factor Expression Initiates Widespread Targeted Chromatin Remodeling. *Cell Reports* 8: 96–105.
257. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, et al. (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471: 68–73.
258. Nishino K, Toyoda M, Yamazaki-Inoue M, Fukawatase Y, Chikazawa E, et al. (2011) DNA Methylation Dynamics in Human Induced Pluripotent Stem Cells over Time. *PLoS Genet* 7: e1002085. doi:10.1371/journal.pgen.1002085.
259. Sierra J, Yoshida T, Joazeiro CA, Jones KA (2006) The APC tumor suppressor counteracts  $\beta$ -catenin activation and H3K4 methylation at Wnt target genes. *Genes & Development* 20: 586–600.
260. Berge ten D, Kurek D, Blauwkamp T, Koole W, Maas A, et al. (2011) Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat Cell Biol* 13: 1070–1075.
261. Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH (2003) Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Medicine* 10: 55–63.
262. Sheridan SD, Surampudi V, Rao RR (2012) Analysis of Embryoid Bodies Derived from Human Induced Pluripotent Stem Cells as a Means to Assess Pluripotency. *Stem Cells International* 2012: 738910. doi: 10.1155/2012/738910.
263. Chen B, Dodge ME, Tang W, Lu J, Ma Z, et al. (2009) Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat Chem Biol* 5: 100–107.

## References

264. Rhim JH, Luo X, Xu X, Gao D, Zhou T, et al. (2015) A High-content screen identifies compounds promoting the neuronal differentiation and the midbrain dopamine neuron specification of human neural progenitor cells. *Scientific reports* 5:16237. doi: 10.1038/srep16237.
265. Fischer M (2017) Census and evaluation of p53 target genes. *Oncogene* 36: 3943–3956.
266. Chen H-R, Juan H-C, Wong Y-H, Tsai J-W, Fann M-J (2017) Cdk12 Regulates Neurogenesis and Late-Arising Neuronal Migration in the Developing Cerebral Cortex. *Human Molecular Genetics* 27: 2289–2302.
267. Desmarais JA, Unger C, Damjanov I, Meuth M, Andrews P (2016) Apoptosis and failure of checkpoint kinase 1 activation in human induced pluripotent stem cells under replication stress. *Stem Cell Res Ther* 7:17. doi: 10.1186/s13287-016-0279-2
268. Momcilovic O, Knobloch L, Fornasaglio J, Varum S, Easley C, et al. (2010) DNA Damage Responses in Human Induced Pluripotent Stem Cells and Embryonic Stem Cells. *PLoS ONE* 5: e13410. doi:10.1371/journal.pone.0013410.
269. Ohgushi M, Matsumura M, Eiraku M, Murakami K, Aramaki T, et al. (2010) Molecular Pathway and Cell State Responsible for Dissociation-Induced Apoptosis in Human Pluripotent Stem Cells. *Cell Reports* 7: 225–239.
270. Ishizaki T, Uehata M, Tamechika I, Keel J, Nonomura K, et al. (2000) Pharmacological Properties of Y-27632, a Specific Inhibitor of Rho-Associated Kinases. *Mol Pharmacol* 57: 976–983.
271. Vernardis SI, Terzoudis K, Panoskaltis N, Mantalaris A (2017) Human embryonic and induced pluripotent stem cells maintain phenotype but alter their metabolism after exposure to ROCK inhibitor. *Scientific reports* 7: 42138. doi:10.1038/srep42138.
272. Herskowitz I (1987) Functional inactivation of genes by dominant negative mutations. *Nature* 329: 219–222.
273. Barbaro V, Confalonieri L, Vallini I, Ferrari S, Ponzin D, et al. (2012) Development of an Allele-Specific Real-Time PCR Assay for Discrimination and Quantification of p63 R279H Mutation in EEC Syndrome. *The Journal of Molecular Diagnostics* 14: 38–45.
274. Darawi MN, Ai-Vyrn C, Ramasamy K, Hua PPJ, Pin TM, et al. (2013) Allele-specific polymerase chain reaction for the detection of Alzheimer's disease-related single nucleotide polymorphisms. *BMC Medical Genetics* 14: 27. doi:10.1186/1471-2350-14-27.
275. Wilhelm J, Reuter H, Tews B, Pingoud A, Hahn M (2002) Detection and Quantification of Insertion/Deletion Variations by Allele-Specific Real-Time PCR: Application for Genotyping and Chimerism Analysis. *Biological Chemistry* 383: 1423–1433.
276. Sarparanta J, Jonson PH, Golzio C, Sandell S, Luque H, et al. (2012) Mutations affecting the cytoplasmic functions of the co-chaperone DNAJB6 cause limb-girdle muscular dystrophy. *Nat Genet* 44: 450–455.
277. Ayton P, Sneddon SF, Palmer DB, Rosewell IR, Owen MJ, et al. (2001) Truncation of the MLL gene in exon 5 by gene targeting leads to early preimplantation lethality of homozygous embryos. *Genesis* 30: 201–212.
278. 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. *Proc Natl Acad Sci USA* 103: 6629–6634.
279. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6: 377–382.

## References

280. Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, et al. (2015) Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347: 1138–1142.
281. Miyamoto DT, Zheng Y, Wittner BS, Lee RJ, Zhu H, et al. (2015) RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science* 349: 1351–1356.
282. Sahara S, Yanagawa Y, O'Leary DDM, Stevens CF (2012) The Fraction of Cortical GABAergic Neurons Is Constant from Near the Start of Cortical Neurogenesis to Adulthood. *J Neurosci* 32: 4755–4761.
283. Mayer C, Hafemeister C, Bandler RC, Machold R, Batista-Brito R, et al. (2018) Developmental diversification of cortical inhibitory interneurons. *Nature* 555: 457–462.
284. Noctor SC, Flint AC, Weissman TA, Wong WS, Clinton BK, et al. (2002) Dividing Precursor Cells of the Embryonic Cortical Ventricular Zone Have Morphological and Molecular Characteristics of Radial Glia. *J Neurosci* 22: 3161–3173.
285. Malatesta P, Hartfuss E, Gotz M (2000) Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage. *Development* 127: 5253–5263.
286. Attardo A, Calegari F, Haubensak W, Wilsch-Bräuninger M, Huttner WB (2008) Live Imaging at the Onset of Cortical Neurogenesis Reveals Differential Appearance of the Neuronal Phenotype in Apical versus Basal Progenitor Progeny. *PLoS ONE* 3: e2388. doi:10.1371/journal.pone.0002388.
287. Mao H, Pilaz LJ, McMahon JJ, Golzio C, Wu D, et al. (2015) Rbm8a Haploinsufficiency Disrupts Embryonic Cortical Development Resulting in Microcephaly. *J Neurosci* 35: 7003–7018.
288. Baumgartner M, Olthof AM, Aquino GS, Hyatt KC, Lemoine C, et al. (2018) Minor spliceosome inactivation causes microcephaly, owing to cell cycle defects and death of self-amplifying radial glial cells. *Development* 145: dev166322. doi:10.1242/dev.166322.
289. Hartfuss E, Galli R, Heins N, Götz M (2001) Characterization of CNS Precursor Subtypes and Radial Glia. *Developmental Biology* 229: 15–30.
290. Hebert JM, McConnell SK (2000) Targeting of cre to the Foxg1 (BF-1) Locus Mediates loxP Recombination in the Telencephalon and Other Developing Head Structures. *Developmental Biology* 222: 296–306.
291. Silver DL, Watkins-Chow DE, Schreck KC, Pierfelice TJ, Larson DM, et al. (2010) The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat Neurosci* 13: 551–558.
292. McMahon JJ, Shi L, Silver DL (2014) Generation of a Magoh conditional allele in mice. *Genesis* 52: 752–758.
293. Kataoka N, Diem MD, Kim VN, Yong J, Dreyfuss G (2001) Magoh, a human homolog of *Drosophila mago nashi* protein, is a component of the splicing-dependent exon-exon junction complex. *The EMBO Journal* 20: 6424–6433.
294. Gehring NH, Neu-Yilik G, Schell T, Hentze MW, Kulozik AE (2003) Y14 and hUpf3b Form an NMD-Activating Complex. *Cell Reports* 11: 939–949.
295. Singh KK, Wachsmuth L, Kulozik AE, Gehring NH (2013) Two mammalian MAGOH genes contribute to exon junction complex composition and nonsense-mediated decay. *RNA Biology* 10: 1291–1298.
296. Zhang K, Lin W, Latham JA, Riefler GM, Schumacher JM, et al. (2005) The Set1

## References

- Methyltransferase Opposes Ipl1 Aurora Kinase Functions in Chromosome Segregation. *Cell* 122: 723–734.
297. Latham JA, Chosed RJ, Wang S, Dent SYR (2011) Chromatin Signaling to Kinetochores: Transregulation of Dam1 Methylation by Histone H2B Ubiquitination. *Cell* 146: 709–719.
298. Herz H-M, Garruss A, Shilatifard A (2013) SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends Biochem Sci* 38: 621–639.
299. O'Donnell-Luria AH, Pais LS, Faundes V, Wood JC, Sveden A, et al. (2019) Heterozygous variants in KMT2E cause a spectrum of neurodevelopmental disorders and epilepsy. *Am J Hum Genet* 104: 1210–1222.
300. Mas-y-Mas S, Barbon M, Teyssier C, Déméné H, Carvalho JE, et al. (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: e0165139. doi:10.1371/journal.pone.0165139.
301. Yagi H, Deguchi K, Aono A, Tani Y, Kishimoto T, et al. (1998) Growth Disturbance in Fetal Liver Hematopoiesis of Mll-Mutant Mice. *Blood* 92: 108–117.
302. Candeias MM, Malbert-Colas L, Powell DJ, Daskalogianni C, Maslon MM, et al. (2008) p53 mRNA controls p53 activity by managing Mdm2 functions. *Nature Cell Biology* 10: 1098–1105.
303. Naski N, Gajjar M, Bourougaa K, Malbert-Colas L, Fähræus R, et al. (2014) The p53 mRNA-Mdm2 interaction. *Cell Cycle* 8: 31–34.
304. Tseng Y-Y, Moriarity BS, Gong W, Akiyama R, Tiwari A, et al. (2014) PVT1 dependence in cancer with MYC copy-number increase. *Nature* 512: 82–86.
305. Alvarez-Dominguez JR, Lodish HF (2017) Emerging mechanisms of long noncoding RNA function during normal and malignant hematopoiesis. *Blood* 130: 1965–1975.
306. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, et al. (2012) GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* 22: 1760-1774.
307. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, et al. (2012) The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* 22: 1775-1789.
308. Grote P, Wittler L, Hendrix D, Koch F, Währisch S, et al. (2013) The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental Cell* 24: 206–214.
309. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, et al. (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472: 120–124.
310. Field AR, Jacobs FMJ, Fiddes IT, Phillips APR, Reyes-Ortiz AM, et al. (2019) Structurally Conserved Primate lncRNAs Are Transiently Expressed during Human Cortical Differentiation and Influence Cell-Type-Specific Genes. *Stem Cell Reports* 12: 245–257.
311. Ng S-Y, Johnson R, Stanton LW (2012) Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *The EMBO Journal* 31: 522–533.

## References

312. Hunt SE, McLaren W, Gil L, Thormann A, Schuilenburg H, et al. (2018) Ensembl variation resources. Database (Oxford) 2018. doi: 10.1093/database/bay119
313. Sheik Mohamed J, Gaughwin PM, Lim B, Robson P, Lipovich L (2010) Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA* 16: 324–337.
314. Panda A, Martindale J, Gorospe M (2016) Affinity Pulldown of Biotinylated RNA for Detection of Protein-RNA Complexes. *Bio-protocol* 6: e2062. doi:10.21769/BioProtoc.2062.
315. Friedrich T, Lambert AM, Masino MA, Downes GB (2012) Mutation of zebrafish dihydrolipoamide branched-chain transacylase E2 results in motor dysfunction and models maple syrup urine disease. *Disease Models & Mechanisms* 5: 248–258.
316. Zech M, Lam DD, Francescatto L, Schormair B, Salminen AV, et al. (2015) Recessive Mutations in the  $\alpha 3$  (VI) Collagen Gene COL6A3 Cause Early-Onset Isolated Dystonia. *The American Journal of Human Genetics* 96: 883–893.
317. Sager JJ, Torres GE, Burton EA (2012) The Zebrafish Homologue of the Human DYT1 Dystonia Gene Is Widely Expressed in CNS Neurons but Non-Essential for Early Motor System Development. *PLoS ONE* 7: e45175. doi:10.1371/journal.pone.0045175.
318. Tuschl K, Meyer E, Valdivia LE, Zhao N, Dadswell C, et al. (2016) Mutations in SLC39A14 disrupt manganese homeostasis and cause childhood-onset parkinsonism–dystonia. *Nature Communications* 7: 11601. doi: 10.1038/ncomms11601.
319. Eisen JS, Smith JC (2008) Controlling morpholino experiments: don't stop making antisense. *Development* 135: 1735–1743.
320. Li X (2010) Extending the maternal–zygotic effect with genomic imprinting. *Mol Hum Reprod* 16: 695–703.
321. Harvey SA, Sealy I, Kettleborough R, Fenyes F, White R, et al. (2013) Identification of the zebrafish maternal and paternal transcriptomes. *Development* 140: 2703–2710.
322. Schier AF (2007) The Maternal-Zygotic Transition: Death and Birth of RNAs. *Science* 316: 406–407.
323. Nasevicius A, Ekker SC (2000) Effective targeted gene “knockdown” in zebrafish. *Nat Genet* 26: 216–220.
324. Ebert AD, Yu J, Rose FF, Mattis VB, Lorson CL, et al. (2008) Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* 457: 277–280.
325. Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, et al. (2015) FOXP1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell Reports* 16: 375–390.
326. Lambiv WL, Vassallo I, Delorenzi M, Shay T, Diserens A-C, et al. (2011) The Wnt inhibitory factor 1 (WIF1) is targeted in glioblastoma and has a tumor suppressing function potentially by induction of senescence. *Neuro Oncol* 13: 736–747.

## 9 List of Figures

Figure 1: The cerebral cortex in different species.....	1
Figure 2: Steps of murine cortical neurogenesis from E11 to adulthood .....	3
Figure 3: The KMT2 family in mammals .....	11
Figure 4: Identification of a variant in <i>HUWE1</i> .....	39
Figure 5: Low-level <i>DLG3</i> mosaicism .....	41
Figure 6: Identification of two novel nonsense mutations in <i>DDX11</i> .....	43
Figure 7: Detection of defective splicing of <i>FANCI</i> pre-mRNA.....	45
Figure 8: The novel microcephaly-associated gene <i>KMT2B</i> .....	47
Figure 9: Ectopic overexpression of WT and mutant KMT2B protein. ....	48
Figure 10: Characterization of patient-derived fibroblasts.....	50
Figure 11: Characterization of iPSC. ....	52
Figure 12: Investigation of endogenous KMT2B.....	53
Figure 13: RNAseq in iPSC. ....	56
Figure 14: Immunocytochemistry for Cytochrome C.....	58
Figure 15: KEGG pathway analysis .....	59
Figure 16: Differentiation of iPSC cells into passage 1 of NPCs.....	60
Figure 17: Investigating iPSC and differentiating cells.....	62
Figure 18: CRISPR/Cas9-based KO in zebrafish .....	65
Figure 19: Microcephaly in homozygous <i>kmt2bb</i> KO larvae.....	66
Figure 20: Investigation of apoptosis in <i>kmt2bb</i> <sup>-/-</sup> zebrafish.....	67
Figure 21: Investigating heterozygous <i>Kmt2b</i> cKO.....	69
Figure 22: scRNAseq in heterozygous <i>Kmt2b</i> cKO embryos at E13.5 .....	71
Figure 23: Forebrain specific marker genes in heterozygous <i>Kmt2b</i> cKO embryos. ....	72
Figure 24: Apoptosis-associated genes in heterozygous <i>Kmt2b</i> cKO embryos.....	73
Figure 25: Splicing categories of Wimmer et al. ....	79
Figure S1: Verification of <i>KMT2B</i> lncRNA.....	102
Figure S2: Expression of <i>APC2</i> and <i>CTNNB1</i> is not consistently de-regulated in patient-derived iPSC.....	102
Figure S3: PCA plot illustrating variance of iPSC samples applied for RNAseq.....	103
Figure S4: Cell cycle distribution during iPSC differentiation.....	103
Figure S5: Representative morphology of differentiating cells at day 1.....	104
Figure S6: The zebrafish homologs <i>kmt2ba</i> and <i>kmt2bb</i> .....	104
Figure S7: CRISPR/Cas9-based strategy in zebrafish.....	105
Figure S8: Expression constructs for subsequent zebrafish experiments.....	105
Figure S9: Protein alignment of different KMT2s.....	106
Figure S10: Immunohistochemistry for RGC marker Pax6 as well as IPC marker Tbr2 in control and heterozygous <i>Kmt2b</i> cKO ( <i>Kmt2b</i> fl/+ <sub>Emx1-Cre</sub> ) embryos at E13.5.....	106

## 10 List of Tables

Table 1: Enzymes used during PhD project.....	16
Table 2: Bacteria used during PhD project. ....	17
Table 3: Plasmids used during PhD project.....	17
Table 4: Antibodies used during PhD project.....	17
Table 5: Kits used during PhD project. ....	20
Table 6: Equipment during PhD project.....	21
Table 7: Online resources used during PhD project. ....	22
Table 8: Software used during PhD project. ....	23
Table 9: Primers used for genotyping of zebrafish. ....	34
Table 10: Primers used for genotyping of mice. ....	37
Table 11: Results of deep amplicon sequencing using different DNA samples. ....	42
Table 12: Mass spectrometry results using scaffold software of three technical replicates .....	54
Table 13: Number of DE genes after RNAseq in iPSC and comparison of each wt-derived clone with each patient derived clone. ....	55
Table 14: KEGG pathway enrichment of comparison.....	57
Table 15: Position and protein effect of different mutations in human and zebrafish.63	
Table 16: Mendelian ratios of wt, heterozygous and homozygous <i>kmt2ba</i> as well as <i>kmt2bb</i> KO fish after crossing <i>kmt2ba</i> +/- with <i>kmt2ba</i> +/- and <i>kmt2bb</i> +/- with <i>kmt2bb</i> +/- . ....	64
Table S1: Heterozygous and hemizygous variants.....	107
Table S2: Array CGH-analysis of dermal fibroblasts and iPSC.....	108
Table S3: Gene set enrichment analysis with topGO applying DE genes shared by all four comparisons.....	108
Table S4: Gene set enrichment analysis with top GO applying DE genes shared by all four comparisons and by three out of four comparisons.....	109
Table S5: Top 30 of KEGG pathway enrichment analysis of comparison “WT2 vs Pat1”.....	110
Table S6: Applied synthetic oligonucleotides of PhD thesis.....	111



## 11 Appendix

### Acknowledgements

First, I'd like to thank my supervisor and director of the Institute of Human Genetics (Göttingen) Prof. Dr. Bernd Wollnik for giving me the great opportunity to work on this exciting project. Thanks for always motivating and supervising me in the course of the project. I will never forget your outstanding support that enabled the handling of family and PhD project in parallel, and I am pretty aware that your support can not be taken for granted. I thank the members of my TAC committee, Prof. Dr. Bernd Wollnik, Prof. Dr. Andre Fischer, Prof. Dr. Wolfram Zimmermann and the former member Prof. Dr. Steven Johnson for their guidance during my Phd and all fruitful feedback in TAC meetings. I thank Prof. Dr. Ralf Dressel, Prof. Dr. Martin Göpfert and Prof. Dr. Michael Sereda for their participation in my examination board. I wish to thank my graduate school "Molecular Medicine", GAUSS, GGNB and all involved persons that enabled this PhD project and who were always available for questions.

Since, I met a lot of people, not only in Göttingen but also in Cologne and elsewhere, I'd like to thank the following people in alphabetic order. Special thanks to the Gopalakrishnan lab (Institute of Human Genetics, Düsseldorf, Germany), especially to Dr. Elke Gabriel and Prof. Dr. Jay Gopalakrishnan who provided important knowledge for the reprogramming of iPSC. Elke, even though our work is already long time ago, I gladly think back to it and I am really sorry that it did not last longer. You are really a great scientist and I thank you for all your input. I would like to thank the people from the Institute of Human Genetics (Cologne, Germany), especially I would like to thank the director Prof. Dr. Brunhilde Wirth, RG Beck and the team of the laboratory diagnostics. I really had a great time from the very first day and it was a pity leaving you guys. Next, I thank the people from the Institute of Human Genetics (Göttingen, Germany). Special thanks to my favourite secretary Petra Albers. I always felt welcome at your office and I really appreciate all your support. I will think of you at my new job in Münster, and hopefully I will see you there one day. I wish to thank all members of RG Burfeind/Kaulfuß who offered me working space and adopted me as full member during the first months after moving to Göttingen. Thanks to all technicians of the laboratory diagnosics who always tried to run my sequences as soon as possible. Actually, I don't remember how many sequences I forwarded to you, but it was definitely a lot ☺ I wish to thank RG Dosch, especially PD Dr. Roland Dosch and Gudrun Kracht. I always felt as a full member in your lab and learned a lot about zebrafish and scientific procedures in general. Thanks Roland for all your discussions, ideas and especially your precious time for

## Appendix

single meetings. You're doing a great job as group leader! Thank you Gudrun for all your help, you are really "the mum of RG Dosch" and I don't know how I would have managed all this without you. Thanks to RG Hahn, especially a big thank you to Anke Frommhold for great technical assistance with the zebrafish larvae and to Dr. Julia Bauer and Dominik Botermann. Special thanks to PD Dr. Silke Kaulfuß, Dr. Jessica Nolte, Dr. Inka Praulich and Dr. Roser Ufartes for all advices, discussions and open ears in difficult times. I learned a lot and actually I can not imagine how it would have managed everything without you. I wish to thank Dr. Alexander Wolff for the assistance with RNAseq data analysis. Thanks to Prof. Dr. Robert G. Roeder (Rockefeller University, New York, USA) who kindly shared the wt KMT2B construct with me. Thanks to RG Sachinidis (University of Cologne, Germany), especially to Prof. Dr. Agapios Sachinidis and Dr. Vaiboh Shinde who offered me knowledge and working space for my iPSC. I wish to thank the SCU, especially Dr. Lukas Cyganek and Nadine Gotzmann. I always felt welcome at your lab and enjoyed being there. Thanks for all the knowledge you shared with me about stem cell culture and great technical assistance. I wish to thank RG Spielmann (Max Planck Institute for Molecular Genetics, Berlin, Germany), especially PD Dr. Malte Spielmann and Jana Henck for great assistance with analysis of scRNAseq data. Thanks to Prof. Dr. Francis Stewart (TU Dresden, Dresden, Germany) who kindly shared the self manufactured Kmt2b antibody with me. I wish to thank RG Tuoc, especially to Linh Pham, Dr. Tran Tuoc and Pauline Ulmke for sharing their knowledge about murine *in vivo* experiments, for great technical assistance and for working space for my mouse experiments. I always felt comfortable at your lab. Thanks to the ITAL, especially to Susanne Luthin for great technical assistance with RNA extractions from cells and tissues. I thank all members of RG Wollnik. First, I'd like to thank all former members, especially Esther Milz, Dr. Dr. Shahida Moosa and Dr. Maria Schotik for their friendship and support during bad times. You are really lovely people and I am glad that we met. I missed you a lot in Göttingen and it was never the same again! Thanks to all current members especially to Luise Graichen and Christian Müller for technical assistance. Dr. Lukasz Smorag for all fruitful discussions, help with the animal application and open ears. Thank you Gesa Werner for your support during the second part of my thesis. You're really a great colleague and friend. I wish you all the best! Last but not least, Dr. Gökhan Yigit for all scientific discussions, your support and reading of the thesis. I wish you all the best in Göttingen or wherever you will finally settle down ☺

Finally, I wish to thank my family and friends, especially Andre and Nele for their patience, love and support. Thanks to my lovely parents and parents-in-law who made this journey possible. I am deeply grateful for everything you have done and I know it was not easy for all of us.

**Affidavit**

Here I declare that my doctoral thesis entitled “Molecular pathogenesis underlying syndromic forms of primary microcephaly” has been written independently with no other sources and aids than quoted.

---

Nadine Rosin, M.Sc.

Lippetal, November 2019