Aus dem Department für Diagnostische Labormedizin der
Universität Tübingen

Institut für Medizinische Genetik und angewandte Genomik

Haploinsufficiency of KMT2E Results in Transcriptional Changes and Leads to Non-syndromic Intellectual Disability

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Medizin

der Medizinischen Fakultät
der Eberhard Karls Universität
zu Tübingen

vorgelegt von

Melchinger, Esther Ursula

2022

Dekan: Professor Dr. B. Pichler

1. Berichterstatter: Professor Dr. O. Rieß
2. Berichterstatter: Professor Dr. H. Lerche

Tag der Disputation: 30.04.2021

Meinen Eltern.

CONTENTS

LIST OF TABLES IV

LIST OF FIGURES V

LIST OF ABBREVIATIONS VII

1. INTRODUCTION AND OBJECTIVE 1

1.1 Introduction 1
1.1.1 Intellectual disability 1
1.1.2 Epigenetics and neuronal processes 2
1.1.3 Histone H3 lysine 4 methylation in neurodevelopment 3
1.1.4 Dynamics of H3K4 methylation: writers, erasers and readers 4
1.1.5 Role of H3K4 methyltransferases in neurological diseases 5
1.1.6 Molecular basis of *KMT2E* 6
1.1.7 Biological function of KMT2E 10
1.1.8 Role of KMT2E in neurodevelopmental disorders and other diseases 11

1.2 Objectives 15

2. MATERIALS AND METHODS 16

2.1 Patient sample 16

2.2 Materials 16
2.2.1 Consumables 16
2.2.2 Equipment 18
2.2.3 Kits 20
2.2.4 Chemicals and substances 21
2.2.5 Buffers and solutions 24
2.2.6 Enzymes 28
2.2.7 Enzyme buffers 29
2.2.8 Bacteria 30
2.2.9 Cell line 30
2.2.10 Plasmids 30
2.2.11 sgRNA pairs for CRISPR/Cas9 system 31
2.2.12 ssODNs for HDR 31
2.2.13 Antibodies 32
2.2.14 Software 33

2.3 Methods 35
2.3.1 Cell-biological methods 35

2.3.1.1 HEK 293 cells 35
2.3.1.2 Bacteria 37
2.3.2 Molecular biological methods 37
2.3.2.1 CRISPR/Cas9 method 37
2.3.2.2 Modulation of KMT2E variants in HEK293 cells 40
2.3.2.3 Design of sgRNA 41
2.3.2.4 Design of repair template: ssODN 41
2.3.2.5 CRISPR/Cas9 construct delivery 41
2.3.2.6 Transformation 44
2.3.2.7 Sanger sequencing: sequence validation of cloned sSpCas9(BB)-2A-Puro 44
2.3.2.8 Glycerine culture 46
2.3.2.9 Transfection: Genome editing of HEK293 cells with CRISPR/Cas9 46
2.3.3 Transcriptome analysis 53
2.3.3.1 RNA-Seq method 54
2.3.3.2 RNA isolation 55
2.3.3.3 Implementation of RNA-Seq 55
2.3.3.4 Implementation of NGS data analysis 56
2.3.4 Validation of prioritized candidate genes by qRT-PC 56
2.3.5 Biochemical methods 62
2.2.5.1. Western Blotting 62
2.3.6 Biometric analysis 67
2.3.6.1 qRT-PCR 67
2.3.6.2 Western blotting 69

3. RESULTS 70

3.1 Genetic and clinical data of our patient 70

3.2 Successful application of CRISPR/Cas9 in HEK293 cell lines 70
3.2.1 Design of sgRNA in close proximity to the variants of three patients 70
3.2.2 Validation of sgRNA in CRISPR/Cas9 expressing plasmid PX459 by Sanger sequencing 73
3.2.3 Validation of Xfect as a suitable transfection agent 75
3.2.4 Validation of Puromycin as a selection marker 76
3.2.5 Identification of successfully genetic modified HEK293 cell clones via Sanger sequencing 77

3.3 Transcriptome analysis 78
3.3.1 RNA-Seq processing and alignment 78
3.3.2 Detection of 30 genes with differential gene expression analysis in *KMT2E*-KO und -WT cell lines 78
3.3.3 Clustering of gene expression profiles and pathway analysis 82

3.4 qRT-PCR confirmed higher gene expression of *PDE4A* in *KMT2E*-KO cell lines 85

3.5 Western blotting indicated an increased histone methylation level of H3K4me3 in *KMT2E*-KO cell lines ... 89

4. DISCUSSION ... 90

4.1 Experimental design: methodological aspects and questions ... 90
4.1.1 Choice of patient variants ... 90
4.1.2 Cell module: HEK293 cells ... 90
4.1.3 CRISPR/Cas9 ... 92
4.1.4 Quality control of RNA-Seq, qRT-PCR and Western blotting ... 95

4.2 Comparison of results with literature on *KMT2E* ... 96
4.2.1 RNA-Seq ... 96
4.2.2 Western blotting ... 99

4.3 Do mutations in *KMT2E* cause ID? ... 101
4.3.1 Results from our experiments ... 101
4.3.2 Affirmation from a contributed paper ... 101

4.4 Open questions and possible experiments to answer them ... 103

5. OUTLOOK ... 104

6. SUMMARY ... 107

7. DEUTSCHE ZUSAMMENFASSUNG ... 108

8. APPENDIX ... 110

8.1 Primer ... 110

9. BIBLIOGRAPHY ... 115

9.1 Papers and books ... 115

9.2 Websites ... 129

10. ERKLÄRUNG ZUM EIGENANTEIL DER DISSERTATIONSSCHRIFT ... 130

11. PUBLICATIONS ... 131

12. DANKSAGUNG ... 132

List of tables

Table 1 Overview of H3K4 methyltransferases and their association with ID syndromes and other neurodevelopmental disorders 7

Table 2 Summary of the two reported patients with *de novo KMT2E* variants and the newly found variant in the Tübingen patient 14

Table 3 Experimental design 62

Table 4 Components of discontinuous SDS-polyacrylamide gel 65

Table 5 Efficiency of CRISPR/Cas9 experiment 78

Table 6 Genes with genome-wide significantly different expression level 78

Table 7 Genome-wide differently expressed genes 81

Table 8 Top ten differentially expressed genes and their correlations with disease complexes 83

List of figures

Figure 1 Composition of chromatin 2

Figure 2 Histone methylation residues in histone H3 tails 4

Figure 3 Regulation of H3 lysine methylation by different lysine methyl-transferases and demethylases 5

Figure 4 Molecular biological structure and protein form of KMT2E 9

Figure 5 Expression level of KMT2E in different tissues (UCSC genome browser) 9

Figure 6 The CRISPR/Cas9 adaptive immune system in bacteria and archaea 38

Figure 7 Human codon-optimized CRISPR/Cas9 39

Figure 8 Genome modifications through NHEJ or HDR 39

Figure 9 Experimental outline for generating KMT2E knockout cell lines 40

Figure 10 Model of CRISPR/Cas9 expression plasmid ppSPCas9(BB)-2A-Puro (PX45)) 42

Figure 11 Schematic overview of PCR 50

Figure 12 Principle of RNA-Seq 54

Figure 13 RNA-Seq workflow 55

Figure 14 qRT-PCR workflow 56

Figure 15 Example of a standard curve for testing the efficiency of a primer pair 59

Figure 16 Example of a melting curve 60

Figure 17 Workflow of classic western blotting 63

Figure 18 sgRNA designs 71

Figure 19 Manually designed ssODN for HDR 72

Figure 20 Demonstration of sequencing PX459 73

Figure 21 Sanger sequencing results of ligated insert sgRNA into CRISPR/Cas9 plasmid PX459 74

Figure 22 peGFPN2 transfected HEK293 cells, stained with DAPI..... 75

Figure 23 Puromycin assay 76

Figure 24 Sanger sequencing results of single-cell clones 77

Figure 25 Ingenuity pathway analysis of KMT2E-KO affected pathways in HEK293cells..... 82

Figure 26 Association of PDE4A with different autism candidate genes..... 84

Figure 27 Differences in relative gene expression for single genes measured by qRT-PCR..... 86

Figure 28 Estimates of the variance (± standard error) of technical replicates (set = 1.0) relative to the variance of biological replicates and standard error of each estimated variance for the 11 target genes 87

Figure 29 Relative gene expression between KMT2E-WT (set=1.0) and KMT2E-KO cell lines together with the corresponding standard errors 88

Figure 30 Example of a western blotting for all three conditions..... 89

Figure 31 Comparison of histone methylation investigation in western blotting between KMT2E-WT and KMT2E-KO HEK293 cells for H3K4me1, H3K4me2, H3K4me3 89

List of abbreviations

AML	Acute myeloid leukemia
ANOVA	Analysis of variance
ASD	Autism spectrum disorder
bp	Base pair
BSA	Bovine serum albumin
cDNA	Complementary deoxynucleic acid
CNV	Copy number variants
CP	Crossing point
CPC	Chromosomal passenger complex
CpG	Cytosine-phosphatidyl-Guanin
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DAC	Decitabine
DAPI	4´,6´-diamidino-2-phenylindole
ddH_20	Double-distilled water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	Deoxynucleic acid
DTT	Dithiothreitol
ddNTP	Dedesoxynucleotide
dNTP	Desoxynucleotide
E	Efficiency

EDTA	Ethylendiaminetetraacetic acid
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FMRP	Fragile X protein
GFP	Green fluorescent protein
H	Histone
H3K4	Histone 3 lysine 4
H3K4me1	Mono-methylated histone 3 lysine 4
H3K4me2	Di-methylated histone 3 lysine 4
H3K4me3	Tri-methylated histone 3 lysine 4
HMT	Histone methyltransferase
HOX	Homeotic gene
HCl	Hydrochloric acid
HDR	Homology directed repair
ID	Intellectual disability
IPS	Induced pluripotent stem cells
IQ	Intelligence quotient
KMD	Histone demethylase
KMT	Histone methyltransferase
KO	Knockout genotype
LB	Luria Bertani
LGD	Likely gene disruptive
MDS	Myeloid dysplastic syndrome
MIM	Mendelian Inheritance in Man
MLL	Mixed-lineage leukemia

MLPA	Multiplex ligation-dependent probe amplification
mRNA	Messenger ribonulcleic acid
MTOC	Microtubule organization center
NAC	N-acetylcysteine
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NMD	Non-sense mediated decay
OMIM	Online medelian inheritance in men
PAM	Protospacer adjacent motive
PBS	Dulbecco's Phosphate-Buffered Saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
pH	pH value
PHD	Plant homeodomain zinc finger
qRT-PCR	Quantitative real-time polymerase chain reaction
r	Pearson correlation coefficient
RNA-Seq	RNA sequencing or Whole transcriptome shotgun sequencing
ROS	Reactive oxygen species
SAM	Methyl-donor S-adenosylmethionin
SDS-Page	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SET	Su(var) 3-9, enhancer of zeste, trithorax (SET) domain
SNP	Single nucleotide polymorphism

sgRNA	Single guide RNA
TALEN	Transcriptor activator-like effector nuclease
TBE	Tris borat ethylendiaminacetid acid
TBS	Tris-buffered saline
TBST	Tris-buffered saline, 0.1% Tween 20
TEMED	Tetramethylendiamin
tracrRNA	Trans-activating crRNA
Thr	Threonine
WES	Whole exome sequencing
WHO	World Health Organization
WT	Wildtype genotype
ZNF	Zinc finger nucleases

1. Introduction and objective

1.1 Introduction

1.1.1 Intellectual disability

The World Health Organization (WHO) defines intellectual disability (ID) as a condition of incomplete brain development, leading to a lack of cognitive, social, language or motor abilities (World Health Organization. 2004). Mentally disabled individuals score an intelligence quotient (IQ) under 70 and suffer from two deficits of abilities in daily life (van Bokhoven 2011; McKenzie et al. 2016). ID can further be classified on the level of either mental impairment or the phenotype (van Bokhoven 2011). Patients can clinically present a syndromic form, where mental impairment is accompanied by additional congenital malformations (*e.g. Down Syndrome)* or neurological features (*e.g. epilepsy, autism spectrum disorder*) (Vissers, Gilissen, and Veltman 2016).

ID is the most common developmental disorder in humans (Srivastava and Schwartz 2014). It represents a global disease with an average prevalence of 2-3% and can impose a heavy burden on families and society (Kramer and van Bokhoven 2009). It gives rise to the highest costs in healthcare (Prince et al. 2007).

However, little is known about the possible causes of ID. Apart from environmental influences and injuries, genetic modifications seem to play an important role (Patel et al. 2010). Genetic modifications can be divided into chromosomal aneuploidies, chromosomal structural modifications and monogenetic variations. Recent next-generation sequencing (NGS) technologies can link 50-60% of moderate-to-severe ID cases to particular genes (Rauch et al. 2006). More than 700 genes have been identified which can lead to ID when mutated. The total number of genes contributing to ID is estimated to range between 1,500-22,000 genes and make ID an important research area because genetic diagnostics can provide information about prognosis, complications, treatment and inheritance (Vissers, Gilissen, and Veltman 2016).

Pavlowsky and colleagues showed functional connections between the products of individual ID genes on the one hand and common molecular pathways on the other (Pavlowsky, Chelly, and Billuart 2012). Neurogenesis, neuronal migration, synaptic function, cytoskeleton dynamics or cellular signaling are examples of these shared pathways. The functional relationships between ID gene products and common molecular pathways allow neuronal cells to form complex networks that can be modified in response to learning and experience, enabling cognitive function (Crews 2008). Interestingly, many ID genes are involved in epigenetic mechanisms as regulators of chromatin structure or chromatin-mediated transcriptional regulation (Kramer and van Bokhoven 2009).

1.1.2 Epigenetics and neuronal processes

The DNA (deoxyribonucleic acid) encodes the genetic information of all living organisms, based on four organic bases (adenine, cytosine, guanosine, thymine) (Watson and Crick 1953). Nucleic acid consists of two chains, which form a double helix, and is organized as chromatin in the nucleus **(Figure 1)** (Kouzarides 2007). It is coiled around nucleosomes like a chain around beats. Each nucleosome consists of a histone (H) octamer, composed of two histones H2a, H2b, H3 and H4. Histone H1, located in-between these components, is linked to the DNA. Chromatin can be remodeled in a dynamic process through four different methyl-cytosines and over 100 different posttranslational modifications of histones. They are highly site-specific and include *inter alia* methylation, acetylation, phosphorylation or hydroxylation.

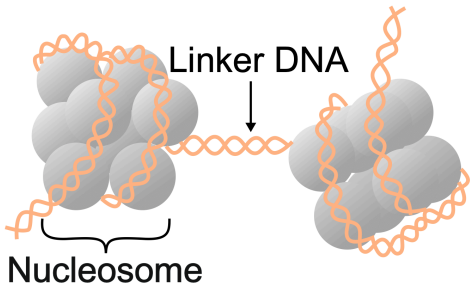


Figure 1 Composition of chromatin
DNA is coiled around nucleosomes which consist of eight histones (two from each H2a, H2b, H3, H4), and linked to histone H1 in-between. These units together form chromatin (modified in line with (Kouzarides 2007)).

These modifications determine the three-dimensional structure of DNA in the nucleus and can regulate gene expression via changes in DNA accessibility. They are summarized as the “epigenome” and remain dynamic throughout their

lifetime, even in neurons and post-mitotic cells of the brain (Cheung and Lau 2005). The posttranscriptional modifications can be altered by various enzymes and exert their influence through different “readers” (Taverna et al. 2007). In addition, these modifications, which alter gene expression without changing the DNA sequence, can even be inherited, as described by the term “epigenetics” (Chen et al. 2014). The structure and regulation of chromatin have a crucial role in neurogenesis. They can affect the induction of synaptic plasticity and memory formation in the hippocampus and prefrontal cortex (Ronan, Wu, and Crabtree 2013). Therefore, it is not surprising that over 50 genes affecting the epigenome have already been associated with neurodevelopmental disorders, autism, schizophrenia and neurodegenerative diseases.

1.1.3 Histone H3 lysine 4 methylation in neurodevelopment

One form of posttranslational modification stands out in the neuronal context (Parkel, Lopez-Atalaya, and Barco 2013): histone H3 lysine 4 (H3K4) methylation seems to be especially important in brain development and is associated with cognitive abilities. In general, methylation can occur on lysine or arginine residues due to their ammonium groups **(Figure 2)** (Kouzarides 2007). The ammonium group of lysine can carry between one and three methyl groups, encoding different information. Lysine methylation can occur on different histone residues, including H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20 (Zhang, Wen, and Shi 2012). Depending on the location, number of methyl groups, histone residues and combination with other posttranslational modifications, histone methylation can signal transcriptional initiation and elongation as well as heterochromatic silencing (Zhou et al. 2011). Overall, H3K4 methylation marks promotors and enhancers and signals gene activation (Barski et al. 2007).

Figure 2 Histone methylation residues in histone H3 tails
The beginning of the amino acid sequence of histone 3 is shown as an example. The amino acids lysine (K) and arginine (R) are marked red and can carry up to three methyl groups which are illustrated by different numbers of blue ME circles. Together with other residues, they form the epigenome and play an important role in neurodevelopment (modified in line with (Kim et al. 2017)).

H3K4 methylation corresponds with defined transcriptional states of a gene, which suits this posttranscriptional modification, especially as a basis for memory storage (Parkel, Lopez-Atalaya, and Barco 2013). It is regulated dynamically during prenatal and early childhood, especially in the prefrontal cortex (Vallianatos and Iwase 2015).

H3K4 methylation also plays an important role in the pathology of neurodevelopmental and neurodegenerative diseases. Genome-wide changes in H3K4 methylation have been observed in patients with autism, schizophrenia and depression (Ronan, Wu, and Crabtree 2013; Ernst, Chen, and Turecki 2009; Cruceanu et al. 2013).

In addition, different H3K4 methylation levels were detected in different tissues with especially high H3K4me1 levels in the brain of healthy individuals, where it seems to be of utmost importance to maintain open chromatin states (Shen et al. 2014).

1.1.4 Dynamics of H3K4 methylation: writers, erasers and readers

The complex balance of H3K4 methylation is achieved by two enzyme families: lysine methyltransferases (KMT) and lysine demethylases (KMD) (Parkel, Lopez-Atalaya, and Barco 2013). So far, thirteen H3K4 specific methyltransferases (“writers”) and eight demethylases (“erasers”) as well as various H3K4 binding partners (“readers”) have been identified (**Figure 3)**.

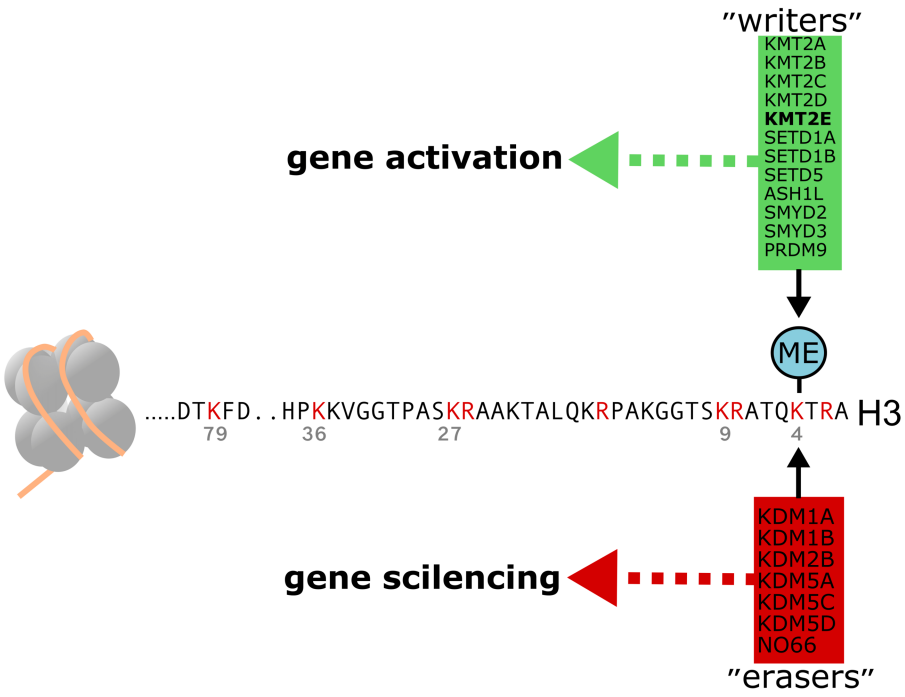


Figure 3 Regulation of H3 lysine methylation by different lysine methyl-transferases and demethylases
Different H3K4 specific methyltransferases and demethylases regulate H3K4 methylation states. H3K4 methylation is usually associated with gene activation, whereas H3K4 demethylation results in gene silencing (modified in line with (Parkel, Lopez-Atalaya et al. 2013)).

The antagonistic enzyme families have probably evolved to shape the delicate epigenetic landscape which is needed for developmental and cellular processes in complex organs such as the brain (Vallianatos and Iwase 2015). They possess a set of functional domains and binding partners: e.g. KMT2A-mediated H3K4 methylation is bound to the WRAD complex, consisting of WDR5, RbBP5, ASH2L, DPY30 (Del Rizzo and Trievel 2011). Each protein functions differently and interacts with other cofactors and chromatin-associated proteins (Wu et al. 2008). Some enzymes target only a subset of genomic loci (e.g. KMT2A-KMT2D), whereas other enzymes function as general regulators of H3K4 methylation (e.g. SET1A/B) (Black, Van Rechem, and Whetstine 2012).

1.1.5 Role of H3K4 methyltransferases in neurological diseases

Research about the role of KTMs and KDMs in intellectual development and cognition began only recently , while their importance for cancer development has been under investigation for much longer time (Parkel, Lopez-Atalaya, and

Barco 2013). A large number of modifications in H3K4 methylation regulators was deciphered in ID-affected individuals in recent years (Kim et al. 2017; Faundes et al. 2018). They complement the class of ID genes involved in the regulation of chromatin structure and chromatin-mediated transcription (Shen et al. 2014). Focusing on H3K4 methylation sites, various H3K4 methyltransferases have been associated with several neurodevelopmental and neurodegenerative disorders. **Table 1** gives an overview of H3K4 methyltransferase which have been identified in ID patients **(Table 1)**.

Genetic studies on H3K4 methyltransferases revealed two surprising findings: First, no accumulation of mutations was observed around specific locations within genes (e.g. functional domains). Second, most pathogenic variants in ID cases occur *de novo* despite a significant number of familial cases (Jones et al. 2012). The versatility of variants in histone methylation enzymes in ID patients supports the hypothesis that H3K4 methylation is fundamental for a healthy brain development. The gain and loss of H3K4 methylation result in similar phenotypes such as ID, ASD and seizures (Shen et al. 2014).

1.1.6 Molecular basis of *KMT2E*

Table 1 illustrates all H3K4 methyltransferases of the KMT2 family that are known to be associated with neurological diseases, but the knowledge level differs regarding the importance of each enzyme in brain development. A rather unknown member is histone lysine methyltransferase 2 E (KMT2E), also known as mixed-lineage leukemia 5 (MLL5).

KMT2E is located at position 104,940,948-105,115,018 on the positive strand of chromosome 7 according to human chromosome annotation 19 **(hg19, Figure 4)**. *KMT2E* covers 174,070 bp, including 27 exons (UCSC Genome browser (2013), UCSC, accessed 14 February 2020, < https://genome.ucsc.edu/index.html>). It is translated into a ~ 200 kDa protein of 1858 amino acid residues (M. Ali, 2013) and, in accordance with other KMT2 family members, it contains two functional domains (Emerling et al. 2002): a N-terminal plant homeodomain (PHD) zinc finger and a central Su(var) 3-9, enhancer of zeste, trithorax (SET) domain.

Table 1 Overview of H3K4 methyltransferases and their association with ID syndromes and other neurodevelopmental disorders

Gene	Syndrome or associated disease	OMIM	Phenotype	Annotated variants and inheritance
KMT2A	Wiedemann-Steiner syndrome	#605139/ +159555	Mild to moderate ID, behavioral difficulties such as autism-like phenotypes and aggression, short stature, distinct facial features (*e.g.,* long lashes on thick arched eyebrow, wide nasal bridge, narrow palpebral fissures), hypertrichosis on elbow and back (Miyake et al. 2016)	Autosomal dominant. Different heterozygous *de novo* truncating mutations (nonsense, frameshift, missense) (Jones et al. 2012)
KMT2B	Childhood-onset dystonia 28	#617284	Developmental delay, dystonia (starting in the lower extremities and beginning in the first decade of life), short stature (Zech et al. 2016; Faundes et al. 2018)	Autosomal dominant with incomplete penetrance. Truncating mutations (nonsense, frameshift, missense, splice site mutations) (Meyer et al. 2017)
KMT2C	Kleefstra-syndrome 2/ ASD	#617768/ *606833	Variable ID, autism, epilepsy, delayed psycho-motor development, mild dysmorphic features, (Koemans et al. 2017; Faundes et al. 2018)	Autosomal dominant with heterogenic truncating mutations in four functionally related genes (KMT2C, MBD5, SMARCB1, NR1I3) (Kleefstra et al. 2012)

Table 1 ***Continued***

Gene	Syndrome or associated disease	OMIM	Phenotype	Annotated variants
KMT2D	Kabuki-syndrome 1	#147920	ID, behavioral abnormalities (ASD, seizures, microcephaly), postnatal dwarfism, musculo-skeletal abnormalities (radiographic abnormalities of vertebrae, hands and hip joints, short fifth finger, persistence of finger pads), distinct facial features (broad nose, prominent earlobes, cleft palate long, palpebral fissures) (Niikawa et al. 1981)	Microdeletion (8p23.1-p22), interstitial duplications (breakpoints 1p13.1 and 1p22.1), inversions (8p23.1-p22), truncating mutations in KMT2D (nonsense, frameshift) (Ng et al. 2010)
KMT2E	ASD	#618512/ *608444	No ID patients, described as ASD risk factor (Iossifov et al. 2012; Dong et al. 2014)	Truncating mutations
KMT2F	Schizophrenia	*611052	Schizophrenia	*De novo* heterozygous loss of function (Takata et al. 2016; Takata et al. 2014; Singh et al. 2016)

* Gene/Locus MIM number; # Phenotype MIM (Mendelian Inheritance in Man) number; + gene and phenotype combined MIM number

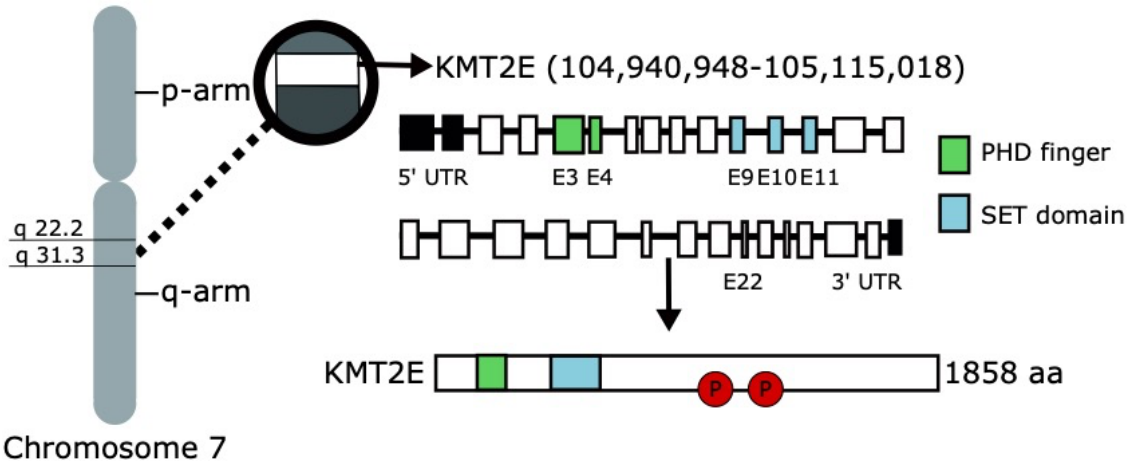


Figure 4 Molecular biological structure and protein form of *KMT2E*
Emerling and colleagues identified *KMT2E* (former MLL5) in 2002. It is located on the q-arm of Chromosome 7 and consists of 27 exons, with a PHD finger in Exon 3 and 4 and a SET domain in Exon 9-11. *KMT2E* translates into an 1858 amino acid long protein, the main isoform. Its two possible phosphorylation sites at AA 861 and AA912 are indicated by an orange circle and serve as binding partners for other proteins (modified in line with (Emerling, Bonifas et al. 2002) (Zhang et al. 2017)).

KMT2E is localized in the nucleus and ubiquitously expressed in embryonic and adult tissues (**Figure 5**) (Deng, Chiu, and Strominger 2004) (Consortium 2013).

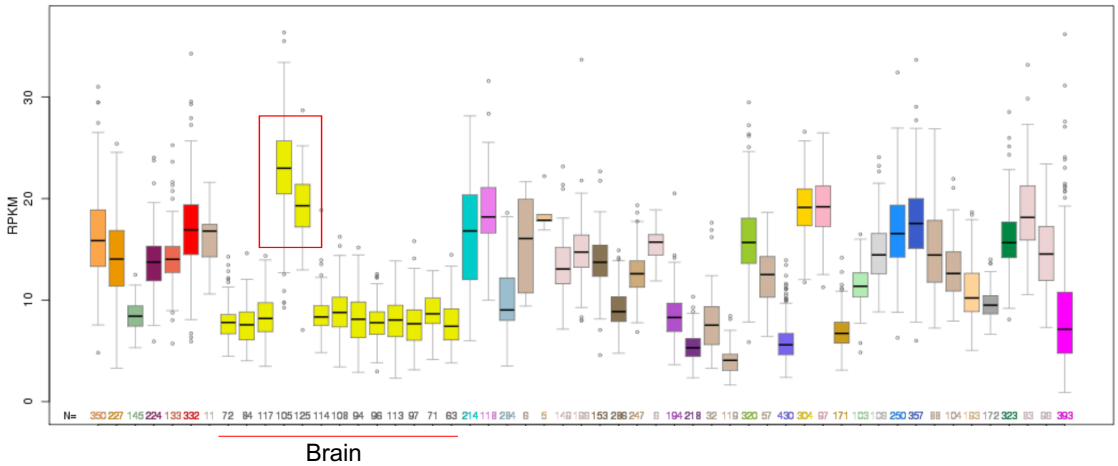


Figure 5 Expression level of *KMT2E* in different tissues (UCSC genome browser)
The bar chart shows the *KMT2E* expression levels in different tissues where the height of a bar indicates the median expression across all samples for a specific tissue. The expression level of *KMT2E* is especially high in the cerebellum and the cerebellar hemisphere of the adult brain (from UCSC Genome browser (2013), UCSC, accessed 14 February 2020, < https://genome.ucsc.edu/index.html >).

KMT2E is evolutionary distant to the other members of the KMT2 enzyme family, making it an interesting study subject. Its SET domain lacks intrinsic histone methyltransferase (HMT) activity towards histone substrates (Sebastian et al. 2009). Two subsequent amino acids (Asn-408 and His-409) are modified to an arginine in the most conserved sequence motif of the SET domain (Mas et al. 2016). Accordingly, the isolated SET domain of KMT2E does not interact with the methyl-donor SAM or any H3K4 histone peptide (Mas et al. 2016).

Furthermore, it was suggested that the KMT2 family requires the trithorax scaffold complex WDR5-RBBP5-ASH2L to exert its methyltransferase activity (Li et al. 2016). However, KMT2E lacks the WDR5 interaction (WIN) motif which enables this interaction (Wu et al. 2008). This raises the question of how KMT2E exerts its HMT activity or whether it has a totally different biological function.

1.1.7 Biological function of KMT2E

The biological functions of KMT2E on gene expression and histone methylation is still disputed (Shen et al. 2014). Different research projects had found diverse effects of KMT2E but a coherent explanation behind KMT2E´s molecular mechanisms and resulting functions did not exist. Previously, KMT2E was identified to play a critical role in various biological processes such as genomic expression, genomic stability, cell cycle progression and hematopoiesis in adults (Zhang et al. 2017).

Early functional investigations found that KMT2E is expressed throughout the cell cycle. In human cell lines, KMT2E overexpression caused G1 arrest (Deng, Chiu, and Strominger 2004), whereas KMT2E knockdown resulted in G1 and G2/M-arrest (Cheng et al. 2008). This implies that an exact balance of KMT2E quantity is of utmost importance for cell cycle progression. Another necessity for the beginning of the G2/M phase was the phosphorylation of Thr-912 in the central domain of KMT2E through the protein Cdc2, which caused KMT2E to dissociate from chromatin (Liu et al. 2010).

KMT2E was shown to form a repressor complex of cytokinesis with TBL1X and NCOR2 (Kittler et al. 2007). The homologous proteins of KMT2E in *S. cereveciae*

also builds a histone deacetylase-recruiting complex which repressed the expression of genes, necessary for cytokinesis, and maintained transcriptional activation states (Pijnappel et al. 2001). Possibly, KMT2E could have the same function in humans.

KMT2E was also found to maintain the determination in quiescent cells (Sebastian et al. 2009). Only when chromosomes were hyper-phosphorylated at position H3T3ph and H3T6ph, KMT2E dissociated from H3K4me3 during mitosis and accumulated at centromeres (Ali et al. 2013). Hence, KMT2E was suggested to exercise the function of a developmental regulator and suppressor for targeting H3K4me3, like its ortholog UpSET in drosophila (Emerling et al. 2002; Rincon-Arano et al. 2012).

Further studies showed that KMT2E maintains genomic stability. KMT2E regulated the stability of the chromosomal passenger complex (CPC), controlling chromosome alignment and segregation to prevent aneuploidy (Liu, Cheng, and Deng 2012). The interaction of KMT2E with PLK1 prevented the appearance of additional microtubule organization centers (MTOCs) and controlled correct spindle assembly (Zhao et al. 2016).

KMT2E knockout mice suffered from increased postnatal lethality, growth retardation, male sterility, small thymus, spleen and lymph nodes (Madan et al. 2009). This demonstrated that KMT2E plays an important role in mammals and the complete depletion of KMT2E has serious health consequences.

All these results of previous studies were collated and defined the functional studies of this thesis to unravel KMT2E´s role in neurodevelopment.

1.1.8 Role of KMT2E in neurodevelopmental disorders and other diseases

KMT2E has already been associated with illnesses besides neurodevelopmental disorders, especially hematologic diseases and various tumors. When KMT2E was first described, the protein was suspected to be a tumor suppressor (Emerling et al. 2002). The coding region of KMT2E on chromosome 7 was often deleted in patients with hematological diseases like myeloid dysplastic syndrome (MDS), acute myeloid leukemia (AML) or therapy-induced leukemia (Zhang et al.

2009). These authors reported that affected patients had a 30% decline in hematopoietic stem cell numbers, defective HSC self-renewal mechanisms and myeloid differentiation. However, subsequent studies failed to support this theory. Hence, the role of KMT2E in leukemia remains debatable. However, KMT2E is still a prognostic marker in leukemia as low *KMT2E* transcription levels are associated with a worse outcome in terms of relapse-free time and overall survival in cytogenetically normal acute myeloid leukemia and core-binding factor leukemia (Damm et al. 2011). In addition, low *KMT2E* transcription levels correspond to lower sensibility to decitabine (DAC), a DNA hypomethylating cytostatic (Yun et al. 2014).

In addition, KMT2E was found to be a key regulator in blood formation. It is involved in the terminal myeloid differentiation and self-renewal of hematopoietic stem cell through its role in DNA methylation (Madan et al. 2009).

Furthermore, a significant number of KMT2E missense mutations and significant expression changes of KMT2E were found in a large range of different solid tumors (Forbes et al. 2015; Rabello Ddo et al. 2013). KMT2E levels were upregulated in large intestine, ovary, stomach or central nervous system tumors but downregulated in pancreas, thyroid or breast cancer. Hitherto, the molecular mechanism behind these observations is still elusive but it is suspected that changes in KMT2E expression lead to modified cell proliferation and genome instability, favoring carcinogenesis.

In addition, dysregulated KMT2E has been associated with atherosclerosis and coronary artery disease. If plasma cholesterol was reduced, KMT2E mediated a decrease in atherosclerotic plaques (Bjorkegren et al. 2014).

The mentioned illnesses show that KMT2E is of great relevance for the human body and changes in KMT2E probably result in serious health problems although until recently, there was no connection of KMT2E with neurodevelopmental disorders.

KMT2E was first suspected to be connected to neurodevelopmental disorders, when a large exome sequencing study found a variant in *KMT2E* to be a risk

factor for autism spectrum disorders (ASD) (Iossifov et al. 2012), which is a highly heritable neurodevelopmental disease of unknown etiology (O'Roak et al. 2012; Willsey et al. 2013). Iossifov and colleagues described a new nonsense variant in *KMT2E* (NM_182931.2: c.3198delC, p.Trp1067Glyfs*2) in their *de novo* gene disruption study of children with ASD **(Table 2)** (Iossifov et al. 2012). KMT2E was reported to be a target gene of the fragile X protein (FMRP), emphasizing connections between ASD and synaptic plasticity. As FMRP target genes are under stronger selection pressure due to their single-allele accessible version, they are extraordinarily dosage-sensitive targets of cognitive disorders. In fact, *KMT2E* belongs to a group of genes in which *de novo* likely gene-disruptive (LGD) variants have a high vulnerability to result in ASD (Iossifov et al. 2015).

Another whole exome sequencing (WES) study found an additional *de novo* variant in *KMT2E* (NM_182931.2: c.167delA p.Tyr56Serfs*34, **Table 2**) (Dong et al. 2014). They showed that *de novo* loss of function variants contributed to ASD. These variants were more common among female ASD patients and enriched among FMRP targets that originated from the paternal chromosome. The underlying mechanism is suspected to lie in the function of KMT2E as a chromatin regulator. Defective chromatin regulation was shown to be a key risk factor for ASD and *KMT2E* is indeed highly expressed throughout the brain, especially during fetal development (Kang et al. 2011).

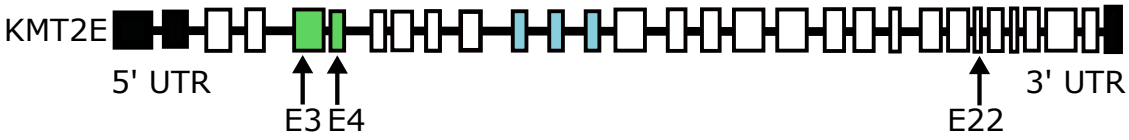
Furthermore, a large deletion including *KMT2E* was found in a patient with intellectual disability, epilepsy, macrosomia, peculiar facial features, corpus callosum hypoplasia, enlarged cisterna magna and right cerebellar hypoplasia (Uliana et al. 2010). The patient suffered from a 3.2-Mb deletion of 7q22.2-7q22.3, resulting in the rearrangement of 15 genes. Taking into account KMT2E's role as a cell cycle controller, it was hypothesized that the patient´s overgrowth symptoms were due to the haploinsufficiency of KMT2E.

We conducted exome sequencing for a cognitively disabled patient, who was presented in Tübingen´s human genetic outpatient clinic. Karyotyping including subtelomere screening via multiplex ligation-dependent probe amplification (MLPA) as well as array analysis (Human Mapping 6.0, Affymetrix) was

unremarkable. A panel analysis of genes, associated with intellectual disability, revealed a *de novo* variant in *KMT2E* (NM_182931.2: c.280delA, p.Thr94Leufs*25). Deletion of the purine adenine at position 280 of the cDNA causes a frameshift, leading to a premature stop coding 25 amino acids after. This mostly caused a premature stop codon, most likely triggering NMD (nonsense mediated decay). NMD is a mechanism which causes degeneration of premature stop codon containing mRNA and results in a complete lack of the protein product. This was taken as an impetus to investigate the role of KMT2E in neurodevelopment through functional studies.

Table 2 Summary of the two reported patients with *de novo KMT2E* variants and the newly found variant in the Tübingen patient

The affected exons found in the patients are marked in the illustration. The table below gives a short overview of each variant, including cDNA position, affected exon, type of variant, effect on translation, effect on amino acid and source (modified in line with (Dong, Walker et al. 2014)).



Gene	cDNA position	Exon	Variant	Effect	Effect on amino-acid sequence	Source
KMT2E	c.167	3	-A	frameshift	p.Tyr56Ser*34	(Dong et al. 2014)
KMT2E	c.280	4	-A	frameshift	p.Thr94Leufs*25	Tübingen patient
KMT2E	c.3198	22	-C	frameshift	p.Trp1067Glyf*2	(Iossifov et al. 2012)

1.2 Objectives

KMT2E represents an interesting candidate gene for ID as several mutations in PHD and SET domain-containing proteins and other members of the histone-lysine methyltransferase protein family have been described to play a causative role in the pathogenesis of several neurodevelopmental disorders, including ID. This thesis aims to establish *KMT2E* as an ID gene through different functional studies.

Hence, the functional effects of possible pathogenic *KMT2E* variants on histone methylation and expression levels were investigated. For this reason, both *KMT2E* variants, described in literature, and the *KMT2E* variant of the Tübingen patient were planned to be recreated in HEK cell lines via CRISPR/Cas9. Afterwards, the difference in H3K4 methylation between the knockout cell lines and wildtype cell lines were to be compared by immunoblotting. The differences in expression levels were to be detected through RNA sequencing.

Our main objectives were to identify the target genes of *KMT2E* and changes in metabolic and signaling pathways by comparing our acquired data with data from other genes of the intellectual disability spectrum. Specifically, we wanted to identify new candidate genes in neurodevelopmental disorders with our results and help to uncover their pathogenesis. Our idea was that our findings could open up new avenues for identifying possible therapeutic approaches.

2. Materials and methods

2.1 Patient sample

The patient´s legal guardians agreed to make the results of the genetic investigation available to research since the variant found in KMT2E was of unclear significance. The University´s ethic counsel had a positive vote prior to the start of the investigations (Project number 592/2017B02).

In a physical examination, the patient showed the following symptoms: the female patient, born in 2005, suffered from intellectual disability with cognitive-verbal developmental disorder, macrocephaly, hyperopia, gross motor coordination disorder, muscular dystonia, knick flatfoot in addition to behavioral disorder with lack of distance, restlessness and distractibility.

2.2 Materials

2.2.1 Consumables

Item	City	Country	Producer
Cell culture flasks 25 cm^2, 75 cm^2	Kremsmünster	Austria	Greiner Bio-One GmbH
Cover film	Dreieich	Germany	Ratiolab GmbH
Cover slips	St. Louis	USA	Sigma Aldrich
Cuvette UVette ®	Hamburg	Germany	Eppendorf AG
Cryo freezing tubes	Kremsmünster	Austria	Greiner Bio-One GmbH
Falcon tubes, polypropylene 15 ml, 50 ml	Kremsmünster	Austria	Greiner Bio-One GmbH
Gloves	Melsungen	Germany	B. Braun

Magnetic beats	Waddinxveen	Netherlands	Clean NA ®
Microscope slides	St. Louis	USA	Sigma Aldrich
Needle	Melsungen	Germany	Sterican ®
Parafilm	Franklin Lakes	USA	Becton Dickinson Labware
PCR plates, 96 wells	Radnor	USA	VWR International
qPCR plates, 384 wells	Wotton	UK	4titude ®
8-strip PCR tubes	Waltham	USA	Fischer Scientific
Petri dish 35×10 mm	Kremsmünster	Austria	Greiner Bio- One GmbH
Petri dish 60× 15 mm	Kremsmünster	Austria	Greiner Bio-One GmbH
Pipette tips 10 µl, 100 µl, 1000 µl	Hamburg	Germany	Eppendorf AG
Pipette tips with filter 10 µl,100 µl, 1000 µl	Hamburg	Germany	StarLab GmBH
Reaction vessel 1.5 ml, 2 ml	Kremsmünster	Austria	Greiner Bio-One GmbH
Safe-lock tubes, 1.5ml	Hamburg	Germany	Eppendorf AG

Sealer	Wotton	UK	4titude®
Sequencing plates	Corning	USA	Corning
Serological pipettes 5 ml, 10 ml, 25 ml	Corning	USA	Corning
Syringe, 20 ml	Meslungen	Germany	B. Braun
Tissue culture plate, 6 wells, 24 wells, 96 wells	Franklin Lakes	USA	BD Falcon

2.2.2 Equipment

Device	City	Country	Producer
µCuvette 1.0	Hamburg	Germany	Eppendorf AG
3730xl DNA analyzer	Foster City	USA	Applied Biosystems
Agarose combs	Erlangen	Germany	PeqLab
Agarose chambers	Erlangen	Germany	PeqLab
Autoclave	Linden	Germany	Systec GmbH
BioPhotometer ® plus	Hamburg	Germany	Eppendorf AG
Centrifuge 5810R, 5423R	Hamburg	Germany	Eppendorf AG

CO_2 incubator CB210	Tuttlingen	Germany	BINDER GmbH
CoolCell ® cell freezing containers	San Rafael	USA	BioCision
Digital gel imaging systems	Cambridge	UK	Biocompare
DNA analyzer ABI3730	Waltham	USA	Thermo Scientific
Fluorescence microscope	Oberkochen	Germany	Zeiss
Heat sealer	Hamburg	Germany	Eppendorf AG
Hirschmann ® pipetus ®	Eberstadt	Germany	Hirschmann Laborgeräte GmbH & Co.KG
LI-COR	Bad Homburg	Germany	L-COR Biosciences
Light cycler ® 480 instrument II	Penzberg	Germany	Roche Life Science
Light microscope Nikon Eclipse TS100	Tokyo	Japan	Nikon Instruments Europe BV
Microscope Axioplan 2	Oberkochen	Germany	Zeiss

Microwave	Güterloh	Germany	Sewerin Elektrogeräte GmbH
Multipette ® M4	Hamburg	Germany	Eppendorf AG
Neubauer counting chamber	Lauda-Königshofen	Germany	Paul Marienfeld GmbH & Co.KG
pH meter	Columbus	USA	Mettler Toledo SevenEasy S20 pH Meter
Precision dare	Balingen	Germany	Kern & Sohn GmbH
Safety workbench	Hanau	Germany	Heraeus
Thermocycler PCR system 2700	Foster City	Germany	Applied Biosystem
Thermomixer compact	Hamburg	Germany	Eppendorf AG
Thermostatplus	Hamburg	Germany	Eppendorf AG
Transilluminator Bio View UV light	Burkhardtsdorf	Germany	Biostep ® GmbH

2.2.3 Kits

Kit	City	Country	Producer
BigDye™ terminator v3.1	Waltham	USA	Thermo Fisher Scientific

cycle sequencing kit			
Clean PCR kit (50 ml)	Waddinxveen	Netherlands	Clean NA
QIAprep spin miniprep kit (250)	Venlo	Netherlands	Qiagen
QIAquick gel extraction kit (250)	Venlo	Netherlands	Qiagen
QuantiTect reverse transcription kit (200)	Venlo	Netherlands	Qiagen
QuantiTect SYBR Green PCR kit (250)	Venlo	Netherlands	Qiagen
RNase-Free DNase set (50)	Venlo	Netherlands	Qiagen
RNeasy mini kit (50)	Venlo	Netherlands	Qiagen

2.2.4 Chemicals and substances

Chemical	City	Country	Producer
Agar-agar	Karlsruhe	Germany	Carl Roth GmbH & Co.KG
Agarose	Basel	Switzerland	Lonza Group

Ampicillin sodium salt (371.39 g/mol)	Karlsruhe	Germany	Carl Roth GmbH & Co.KG
Antibiotic-antimycotic (100x)	Waltham	USA	Thermo Fisher Scientific
Betaine solution (5 M)	St. Louis	USA	Sigma Aldrich
Boric acid	Darmstadt	Germany	Merck KG
Bromophenol	St. Louis	USA	Sigma Aldrich
DAPI	St. Louis	USA	Sigma Aldrich
Distilled water	Darmstadt	Germany	Fresenius Kabi
DMEM (Dulbecco´s Modified Medium)	St. Louis	USA	Sigma-Aldrich
DMSO (dimethylsulfoxid)	St. Louis	USA	Sigma-Aldrich
DNA ladder mix pegGOLD, 100-1000bp	Radnor	USA	VWR Chemicals Prolab ®
DNA gel loading dye (6x)	Waltham	USA	Thermo Fisher Scientific
Dulbecco´s PBS (phosphate-buffered saline)	Waltham	USA	Thermo Fisher Scientific
DTT (dithiothreitol)	Waltham	USA	Thermo Fisher Scientific

EDTA (ethylendiamintetraacetic acid)	Karlsruhe	Germany	Carl Roth GmbH & Co.KG
Ethanol absolute 100%	Radnor	USA	VWR Chemicals Prolab ®
Ethidium bromide dye	Waltham	USA	Thermo Fisher Scientific
FBS (fetal bovine serum)	Waltham	USA	Thermo Fisher Scientific
Kanamycinsulfat (582.58 g/mol)	Karlsruhe	Germany	Carl Roth GmbH & Co.KG
Luria-Bertani (LB) medium	Franklin Lakes	USA	Becton, Dickinson and Company
PCR grade nucleotide mix	Basel	Switzerland	F. Hoffmann-La Roche
Penicilline-streptomycin antibiotics	Waltham	USA	Thermo Fisher Scientific
PFA	St. Louis	USA	Sigma-Aldrich
Poly-lysine	St. Louis	USA	Sigma-Aldrich
Protease-inhibitor	Basel	Switzerland	Roche

Puromycin 50 mg, solid	Dallas	USA	Santa Cruz Biotechnologies
Q-solution	Venlo	Netherlands	Qiagen
Rizma ® base	St. Louis	USA	Sigma-Aldrich
Trypsin/EDTA 0.25%	Waltham	USA	Thermo Fisher Scientific
Western blot marker	Waltham	USA	Page Ruler™ Prestained Protein Ladder, Thermo Scientific™
Xfect transfection reagent	Mpitain View	USA	Clontech

2.2.5 Buffers and solutions

Solution	Ingredients
Agar plates	500 ml of LB-Medium 7.5 g of Agar-Agar autoclave, cool-down to 60°C addition of 500 µl of Ampicillin or Kanamycin, production of plates
Agarose gel (1%)	0.6 g of agarose 1xTBE up to 40 ml, 1 µl Ethidiumbromid

	production of gel
Ampicillin (100 mg/ml)	1 g of Ampicillin 10 ml of ddH_2O sterile filtering, storage at -20°C
2M $CaCl_2$	2 M $CaCl_2$ ddH_20 up to 15 ml
Homogenization buffer	
50 mM Tris-HCl @pH 7.5	500 µl of 1 M Tris-HCl
25 mM KCl	250 µl of 1 M KCl
250 mM Sucrose	855.8 mg of Sucrose
0.5 mM PMSF	25 µl of 0.2 M PMSF
Protease Inhibitor	100 µl of Protease inhibitor
ddH_20	9.13 ml
2× HPS @pH 7.05	0.28 M HCl 0.05 M HEPES (4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid) 0.012 M Dextrose 0.010 M KCl 0.0015 M Na_2HPO_4 ddH_20 up to 50 ml
Kanamycin (50 mg/ml)	0.5 g of Kanamycin 10 ml of ddH_2O sterile filtering, storage at -20°C

1M KCl	3.73 g of KCl Add 50 ml of ddH_2O
LB medium	25 g of LB-powder Add 1 l of ddH_2O autoclave
0.2 M Phenylmethylsulphonylfluoride	34.8 g of PMSF 1 ml of EtOH
0.1% Ponceau (w/v) in 1% (v/v) acetic acid	10 ml of ddH_20 0.3 ml of glacial acetic acid 0.033 g of Ponceau S add 30 ml of ddH_20
0.4 N H_2SO_4	111.12 µl of 18 M H_2SO_4 9.88 ml of ddH_2O
2× SDS sample buffer	65.8 mM of Tris-HCl at pH 6.8 2.1% SDS 0.01% of Bromphenol Blue 26.3% (w/v) glycerol
Standard medium	DMEM 100X antibiotic-antimycotic 100X amino acid 10% FBS
10x TBE (Tris-Borat EDTA)	890 mM of Tris

	20 mM of EDTA 890 mM of boric acid
10x TBS (Tris-buffered saline)	24 g of Tris base 88 g of NaCl Dissolve in 900 ml of ddH_2O pH to 7.6 with 12 M of HCl Add 1 l ddH_2O
1x TBST (Tris-buffered saline, 0.1% Tween 20)	For 1 l: 100 ml 10x TBS 900 ml ddH_2O 1 ml of Tween 20
0.1×TE @pH 7.05	0.001 M of Tris-HCl ($(HOCH_2)_3CNH_2$-HCl) 0.0001 M of EDTA (Ethylenediaminetetra- acetic acid) ddH_20 up to 50 ml
1 M Tris-HCl @pH 7.4	Dissolve 121.1 g of Tris base in 800 ml of DEPC-H_20 Adjust with concentrated HCl (approx. 70 ml of HCl) ddH_20 up to 1000 ml

2.2.6 Enzymes

Enzyme name	Enzymatic activity	Buffer	Producer	Cat. no.
Antarctic phosphatase	5 000 units/ml	Antarctic phosphatase	NEB	M0289S
BbsI	10 000 units/ml	Cutsmart	NEB	R0539S
Exonuclease T7	10 000 units/ml	NEB buffer™ 4	NEB	M0263S
Cloned Pfu DNA polymerase	2 500 units/ml	10x cloned Pfu buffer	Agilent	600154
Taq DNA polymerase	5 000 units/ml	Qiagen PCR buffer	Qiagen	139299897
T4 DNA ligase	400 000 units/ml	T4 DNA ligase reaction buffer	NEB	M0202S
T4 polynucleotide kinase (PNK)	10 000 units/ml	T4 DNA ligase reaction buffer	NEB	M0201S
Trypsin (0.05%)	-	Solution in HBSS	Neolab	1444ML100

All enzymes stored at -20°C.

2.2.7 Enzyme buffers

Buffer name	Component (1*)	Manufacturer	Cat. no
Antarctic phosphatase reaction buffer (10*)	50 mM of Bis-Tris-propane-HCl 1 mM of $MgCl_2$ 0.1 mM of $ZnCl_2$ pH 6@25 °C	NEB	B0289S
CutSmart® buffer (10*)	50 mM of potassium acetate 20 mM of Tris-acetate 10 mM of magnesium Acetate 100 μg/ml BSA pH 7.9 @25°C	NEB	B7204S
1x NEBuffer™ 4 (10*)	50 mM of potassium-acetate 20 mM of Tris-acetate 10 mM of magnesium-acetate 1 mM of DTT µ pH 7.9 @25°C	NEB	B7004S
10x Cloned Pfu DNA polymerase reaction buffer	200 mM of Tris-HCl 100 mM of KCl 100 mM of $(NH_4)_2SO_4$	Agilent	600154

	20 mM of $MgSO_4$ 1.0% Triton ® X-100 1 mg/ml of nuclease-free BSA		
Qiagen PCR buffer	KCl $(NH_4)_2SO_4$	Qiagen	139299897
T4 DNA ligase reaction buffer (10)	10 mM of $MgCl_2$ 50 mM of Tris-HCl 1 mM ATP 10 mM of DTT pH 7.5 @25°C	NEB	B0202S

All enzyme buffers stored at -20°C.

2.2.8 Bacteria

DH5α E. coli acquired from NEB® (Catalog #C2987I) and bottled into aliquots.

2.2.9 Cell line

HEK293 cells acquired from ATCC (ATCC® CRL-1573™) and routinely tested for mycoplasma contamination.

2.2.10 Plasmids

Plasmid	Antibiotics resistance	Com-mercial insert	Source	Self - ligated insert	Restriction-enzyme
pSpCas9(BB)-2A-Puro	Amp^R, $Puro^R$	None	Addgene (ID: 48139)	KMT2E x3 oligo pair	BbsI

(pX459) V 2.0			Annotated GenBank files on http://www.genome-engineering.org/	KMT2E x4 oligo pair	BbsI
				KMT2E x22 oligo pair	BbsI
pEGFP-N1	Kana^{R}	GFP	Addgene (ID: 39299)	None	-

2.2.11 sgRNA pairs for CRISPR/Cas9 system

All oligonucleotides were obtained from Metabion.

Name	Sequence (5´to 3´)
KMT2E_X3_S1	CACCGCATCATTCACACAGTTACAT
KMT2E_X3_AS1	AAACATGTAACTGTGTGAATGATG
KMT2E_X4_S1	CACCGAGTTTCATCAAAATTAGGAG
KMT2E_X4_AS1	AAACCTCCTAATTTTGATGAAACTC
KMT2E_X22_S1	CACCGGGGCACAATGTATTCTTCC
KMT2E_X22_AS1	AAACGGAAGAATACATTGTGCCCC

2.2.12 ssODNs for HDR

All ssODNs were obtained from Metabion.

Name	Sequence (5´to 3´)
CRISPR-Donor c.167delC,	tatacctcgagctcacatcattcacacagttaattcctttgccctatgcggtaagtgttaaacattctttg

p.Tyr56Serfs*34 (KMT2E_Dx3_S1)	
CRISPR-Donor c.167delC, p.Tyr56Serfs*34 (KMT2E_Dx3_AS1)	caaagaagtgtttaacacttaccgcatagggcaaagcaaagcaattaa ctgtgtgaatgatgtgagctgaggtata
CRISPR-Donor c.280delA, p.Thr94Leufs*25 (KMT2E_D_x4_S1)	cctcccccgagagtccttattagcaaaaatgaagtagccatatttaccctc ctaattttgatgaaacttcaa
CRISPR-Donor c.280delA, p.Thr94Leufs*25 (KMT2E_D_x4_AS1)	tggaaggtttcatcaaaattaggaggagggtaaataaatatggctacttcattt ttgctaataaggactctcgggggagg
CRISPR-Donor c.3198delC, p.Trp1067Glyfs*2 (KMT2E_Dx22_S1)	cgactcactctggacggggcacaatgtattcttatgcgtaaagagccat ggcagaacaggagttaacttctc
CRISPR-Donor c.3198delC, p.Trp1067Glyfs*2 (KMT2E_Dx22_AS1)	gagaagttaactcctgttctgccatggctctttacgcagaagaatacattgtg ccccgtccagagtgagtcg

2.2.13 Antibodies

Name	Company	Catalogue no.	Host
Anti-histone H3 antibody (nuclear loading control and ChIP grade)	Abcam	ab24834	Mouse

Anti-histone H3 (mono methyl K4) antibody (ChIP grade)	Abcam	ab8895	Rabbit
Anti-histone H3 (di methyl K4) antibody (ChIP grade)	Abcam	ab32356	Rabbit
Anti-histone H3 (tri methyl K4) antibody (ChIP grade)	Abcam	ab8580	Rabbit

2.2.14 Software

Software	access	use
Cctop (Stemmer et al. 2015; Labuhn et al. 2017)	http://tools-genome-engineering.org	CRISPR design tool
Ingenuity Pathway Analysis	http://www.qiagenbioinformatics.lcom/products/indigenuity-pathway-analysis/	Evaluation RNAseq data
NetPrimer	http://www.premierbiosoft.com/netprimer/netprlaunch/Help/xnetprlkaunch.html	Primer design
Plabstat	https://plant-breeding.uni-hohenheim.de/en/83113	Statistical analysis
R	https://www.r-project.org	Statistical analysis

Reverse complement	https://www.bioinformatics.org/sms/rev_comp.html	Conversion of DNA sequence into its reverse sequence
Standard Nucleotide BLAST	Https://blast.ncbi.nlm.nih.gov/Blast.cgi	Identification of similarity between nucleotide sequences
Standard protein BLAST	Https://blast.ncbi.nlm.nih.gov/Blast.cgi	Identification of similarity between protein sequences
UCSC genome browser	https://genome.ucsc.edu	Combined data from the Genome Browser database

2.3 Methods

2.3.1 Cell-biological methods

2.3.1.1 HEK 293 cells

The adherent cell line HEK293 comprises human embryonic kidney cells which have been transformed permanently by sheared Ad type 5 DNA (Graham et al. 1977). It is suspected that they represent a neuronal lineage of an original kidney culture due to similarities with immature neurons (Shaw et al. 2002). HEK293 cells are hypo-triploid with 64 model chromosomes (Lin, Boone, et al. 2014). They were used for all downstream experiments because they grow reliably and can be transfected easily.

Maintenance and splitting of HEK 293 cells

HEK293 cells were cultured in D10 medium, consisting of DMEM, 10% heat-inactivated fetal bovine serum, penicillin g (100 units/ml) and streptomycin (100 µg/ml), on 10 cm Petri dishes. The cells were incubated at 37°C and 5% CO_2. For cell transfer and splitting, D10 was removed and the cells were rinsed carefully with PBS. 2 ml of Trypsin was added to loosen the adherent cells. The reaction was stopped by adding D10. The dissociated cells could now be reseeded in new culture containers. The cells were split at 70% confluency.

Thawing of HEK 293 cells

After taking HEK293 cells out of the nitrogen tank, the HEK 293 cells were transported on dry ice to a 37°C water bath to thaw. Afterwards, the cells were centrifuged in a 20 ml Falcon with 10 ml of preheated DMEM with 10% of FBS and Pen-Strep at 300 g for 5 min. The supernatant was discarded, and the cells were re-suspended in 10 ml of preheated DMEM with 10% FBS and penicillin-Streptomycin on a 10 cm Petri dish. The cells were incubated at 37°C and in 5% CO_2.

Cryopreservation of HEK 293 cells

Cryopreservation enables the long-term preservation of cells by cooling them down to –80°C (Pegg 2007). This stops any possible harmful enzymatic or

chemical reaction. However, ice formation, dehydration and an increase in the soluble particles can risk cell integrity.

100%-confluent HEK293 cells were loosened from a 10 cm Petri dish by adding 2 ml of Trypsin and incubating them at 37°C and in 5% CO_2 for 5 min. The reaction was stopped by adding 10 ml of D10 medium. The cell suspension was transferred to a 20 ml Falcon and centrifuged at 300 g for 5 min. The supernatant was discarded, the HEK293 cells were re-suspended with 10 ml of PBS and centrifuged again at 300 g for 5 min. This step was repeated again. The cells were finally re-suspended in 900 µl of D10 medium and stored at -80°C.

Pellet production of HEK 293 cells

After confluent HEK293 cells had been rinsed with 10 ml of PBS, the adherent cells were loosened from their 10 cm Petri dish by adding 2 ml of Trypsin and incubating them at 37°C and in 5% CO_2 for 5 min. The reaction was stopped by adding 10 ml of DMEM. The cells were transferred to a 15 ml Falcon and centrifuged at 300 g for 5 min. The supernatant was carefully discarded. The remaining cells were re-suspended in 1 ml of PBS, transferred into a 1.5 ml Eppendorf cup and centrifuged again at 300 g for 5 min. The supernatant was aspired, and the pellets were stored at -80°C.

Production of HEK293-conditioned D10 medium

Conditioned D10 medium was produced to enable the cultivation of single cells. The cell secretome (conditioned medium) contains signal peptides which are proteins shed from the cell surface and released intracellular proteins (Dowling and Clynes 2011). It consists of enzymes, growth factors, cytokines and hormones and influences cell growth fundamentally by regulating cell-to-cell and cell-to-extracellular matrix interactions. It can improve single-cell recovery by providing vital molecules.

The consumed D10 medium of a 70%-confluent HEK293 cell plate was removed, sterile filtered, diluted 1:1 with fresh D10 medium and stored at -20°C until using it for cultivating HEK293 single clones.

Determination of cell numbers

The number of cells was determined using a Neubauer counting chamber. The chamber was filled with the cell suspension of unknown concentration. The cells in all four quadrants of the chamber were counted and the mean cell count was calculated. The mean cell count was multiplied by the dilution factor, ventricular factor and total volume of the cell suspension to obtain the total cell count:

$$Total\ cell\ count =$$

$$mean\ cell\ count\ \times\ dilution\ factor\ \times\ ventricular\ factor\ (10^4) \times total\ volume$$

2.3.1.2 Bacteria

Production of Ampicillin-LB-agar plates

25 g/l of LB powder, 15 g/l of agar and deionized water – filled up to 1 l – were mixed and autoclaved. After cooling down to 60°C, the required antibiotic was added (Ampicillin 1:1000, stock solution 100 mg/ml; Kanamycin 1:1000, stock solution 50 mg/ml). The fluid medium was poured into 100 mm Petri dishes, which were stored at 4°C.

2.3.2 Molecular biological methods

2.3.2.1 CRISPR/Cas9 method

CRISPR/Cas9 is a recently invented genome editing tool (Sander and Joung 2014). Before adapting the CRISPR/Cas9 method as a molecular biological method, genome modifications could only be implemented at precise locations by using engineered nucleases such as zinc finger nucleases (ZFN) or transcription activator-like effectors (TALEN). However, these methods required time-consuming designs of proteins, specifically targeting a certain genomic location. These limitations were overcome with the CRISPR/Cas9 system and due to its relevance, it was named Science´s “Breakthrough of the Year 2015” (Miller et al. 2007; Clark and Pazdernik 2016; Wood et al. 2011).

CRISPR stands for clustered regularly interspaced palindromic repeats. CRISPR/Cas9 is derived from an adaptive immune system of bacteria and archaea against viral and phagal infections **(Figure 6)** (Mojica et al. 2000). It

consists of CRISPR-associated genes (cas), auxiliary trans-activating crRNA (tracrRNA) and CRISPR RNA (crRNA) (Deveau, Garneau, and Moineau 2010). crRNA comprises repetitive elements, interspaced by short variable sequences derived from exogenous DNA targets – so-called protospacers. Each protospacer is immediately followed by a specific protospacer-adjacent motif (PAM) in the target sequence, depending on the CRISPR system's bacterial origin.

In our experiment, the *Streptococcus pyogenes* type II CRISPR system was applied, associated with a 5´-NGG PAM sequence. After an RNAse has processed the crRNA array into a sequence, consisting of the 20-nucleotide target sequence and a fractional sequence of the repetitive element, the crRNA with the tracrRNA directs the nuclease (i.e. cas9) to its viral DNA target.

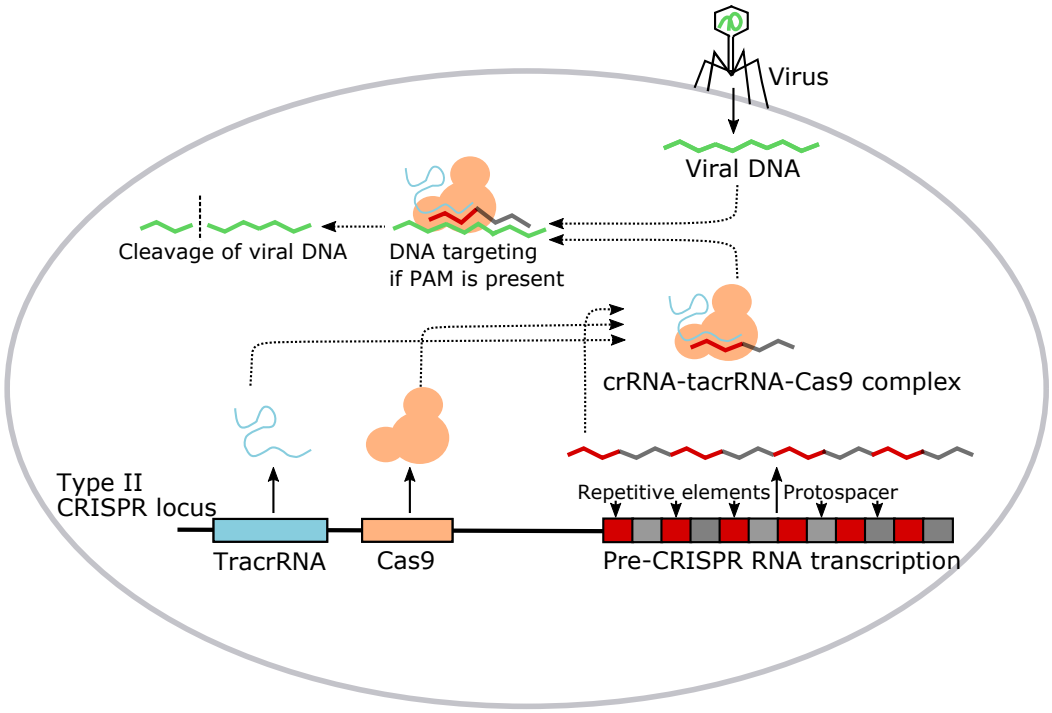


Figure 6 The CRISPR/Cas9 adaptive immune system in bacteria and archaea
Processed crRNA, tracrRNA and cas9 form the complex CRISPR/Cas9. crRNA consists of repetitive elements, which are complementary to parts of the tracrRNA, and protospacers, which are derived from foreign DNA (e.g. virus, phage) and mediate immunity. In case of a viral infection, a 20 bp sequence of the protospacer binds to the viral DNA ("target") by Watson-Crick base pairing if both sequences are complementary. If the target sequence is followed by the specific PAM sequence of the protospacer on the viral DNA, a double-strand break is induced and the infection is made harmless (modified in line with (Mali, Esvelt, and Church 2013)).

The RNA-guided nuclease function of the CRISPR system was transferred to mammalian cells (Jinek et al. 2012): the nuclear localization sequence-marked cas9 was human codon-optimized and the single guide RNA (sgRNA) was fused from crRNA and tracRNA to enable rapid modifications of the target sequence **(Figure 7).** Thus, Cas9 can target any DNA sequence that goes along with its specific PAM sequence (Cong et al. 2013). The 20-nucleotide targeting sequence can be specified within the sgRNA.

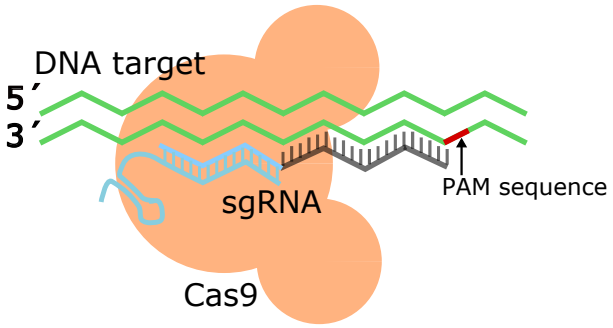


Figure 7 Human codon-optimized CRISPR/Cas9
(modified in line with (Ran et al. 2013)).

After the nuclease Cas9 has cleaved the targeted genomic locus between the 17th and 18th base of the target sequence, both double-strand ends are joined by one of two different DNA damage repair pathways **(Figure 8)** (Ran et al. 2013):

Non-homologous end joining (NHEJ): NHEJ enables an error-prone ligation of both strands, leading to insertions, deletions or switch of nucleotides. These mutations can lead through frameshift mutations and premature stop codons to gene knockouts. Premature stop codons trigger non-sense-mediated decay (NMD). This conserved mechanism leads to the degradation of mRNAs. However, NMD does not occur

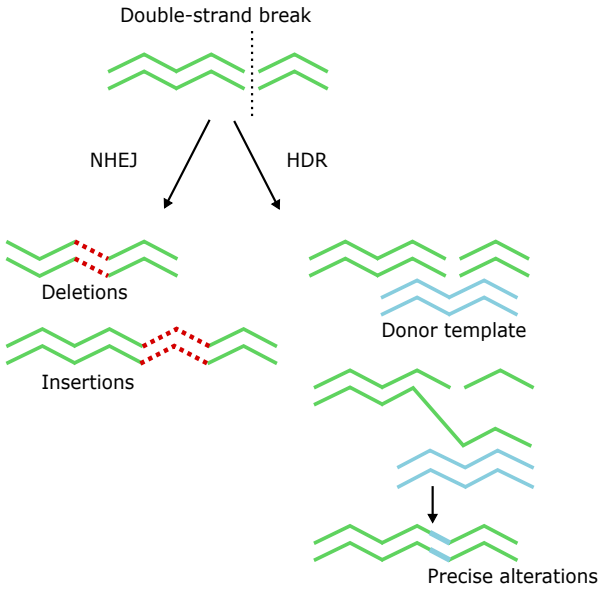


Figure 8 Genome modifications through NHEJ or HDR
After precise double-strand breaks through CRISPR/Cas9, the induction of a cell´s natural DNA damage repair pathways induces genomic modifications (modified in line with (Sander and Joung 2014)).

if the premature stop codon is located within the last exon. (Lindeboom, Supek, and Lehner 2016).

Homology-directed repair (HDR): In the presence of a repair template in the form of a double-stranded DNA with homology arms around the insertion or single-stranded DNA oligonucleotide (ssODN), HDR enables the introduction of precise modifications at the target locus. The repair template can either be the second allele or an exogenously imported synthetic construct. HDR occurs variably, depending on the cell type, cell cycle, target sequence and repair template (Rothkamm et al. 2003; Ran et al. 2013).

2.3.2.2 Modulation of KMT2E variants in HEK293 cells

In order to investigate the effect of the three described patient variants on different cell functions such as histone methylation and transcription, CRISPR/Cas9 was used to generate three separate KMT2E knock out HEK293 cell lines – each modulating a patient´s variant **(Figure 9).**

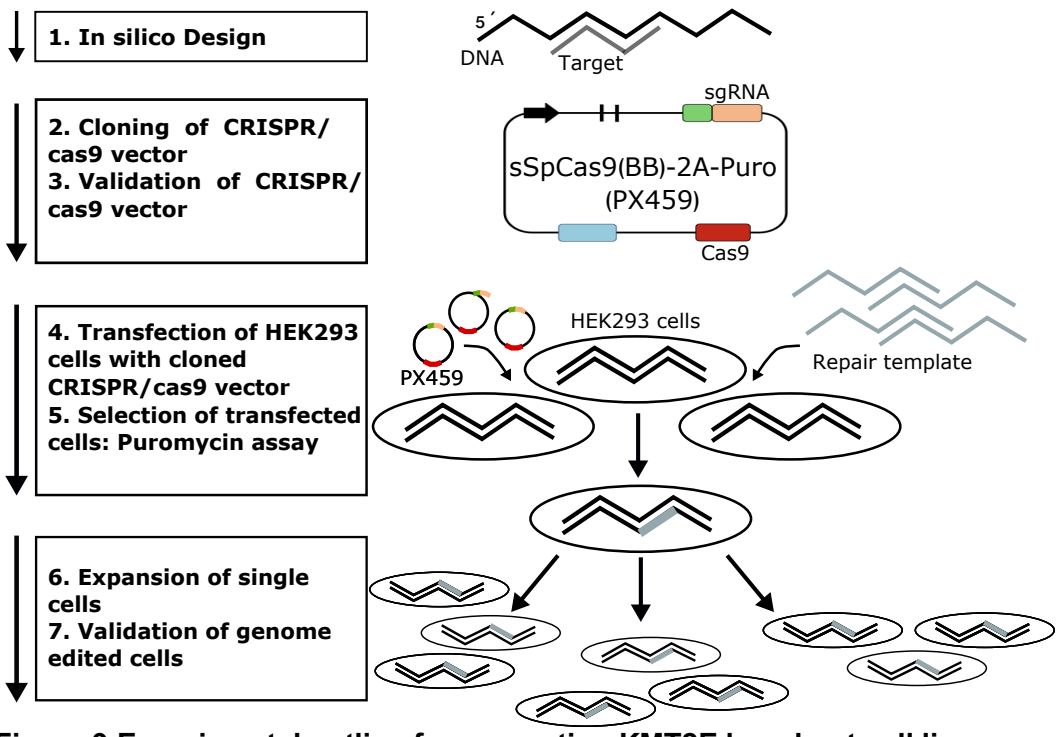


Figure 9 Experimental outline for generating KMT2E knockout cell lines
First, appropriate targets close to the desired variant and ssODN template are designed *in silico*. Second, the CRISPR/Cas9 vectors are cloned and transfected on HEK293 cells. Successfully transfected cells are selected via Puromycin resistance and single cells expanded. Third, Sanger sequencing is used to validate genetic modifications and exclude off-targets (modified in line with (Ran et al. 2013; Stemmer et al. 2015)).

2.3.2.3 Design of sgRNA

The following two aspects need to be considered when designing the 20-nt guide sequence (Ran et al. 2013):

(i) The PAM sequence of the used CRISPR/Cas9 system must immediately follow the 20-nt target sequence.

(ii) The Cas9 nuclease should only cleave at its target sequence; off-targets should be excluded. Base mismatches are unequally tolerable to the CRISPR/Cas9 system, depending on their number and location (Cong et al. 2013). In particular, nucleotides 8-14 before the PAM sequence are less tolerable to mismatches.

The online CRISPR design tool (cctop) was used to identify a suitable target sequence close to the described variants and predict possible off-targets (Stemmer et al. 2015; Labuhn et al. 2017). The following parameters were chosen: (i) the CRISPR/Cas9 system of *Streptococcus pyogenes* with the PAM sequence "NGG-" and (ii) U6 for *in vitro* transcription.

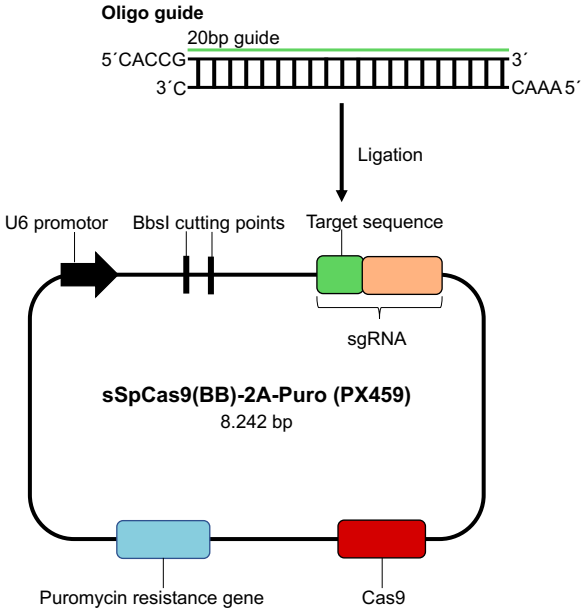
2.3.2.4 Design of repair template: ssODN

ssODN were used to insert the described variants without any need for further cloning (Ran et al. 2013). To increase high HDR efficiency, the single-base change should be within 10 bp (base-pair) of the target sequence of the CRISPR-Cas9 construct. Otherwise, the efficiency of HDR drops enormously. The ssODNs contain at least 40 homologous bp on either side of the target locus and can be orientated into either sense or anti-sense direction. ssODNs are manually designed by adding around 40 bp of the original sequence on either side of the double-strand breaks and only differ in the single-base exchange of the described variant.

2.3.2.5 CRISPR/Cas9 construct delivery

The sgRNAs were delivered using the sgRNA-expression plasmid (Ran et al. 2013): The oligo-pair, encoding the 20-nucleotide target sequence, was annealed and ligated into the restricted plasmid PX459 **(Figure 10)**. Apart from the sgRNA expression cassette, PX459 encodes a human U6 promotor and the expression cassette of the human-optimized nuclease Cas9 hSpCas9, derived from

Streptococcus pyogenes. The plasmid also includes a selection marker: either an antibiotic resistance or fluorescence gene. Here, puromycin resistance was used to select the positively transfected cells.



a. sgRNA for KMT2E Exon 3 (c.167delC, p.Tyr56Serfs*34)

b. sgRNA for KMT2E Exon 4 (c.280delA, p.Thr94Leufs*25)

c. sgRNA for KMT2E Exon 22 (c.3198delC, p.Trp1067Glyfs*2)

Figure 10 Model of CRISPR/Cas9 expression plasmid ppSPCas9(BB)-2A-Puro (PX45))
For each KMT2E variant, a seperate sgRNA and ssODN was designed and ordered (Figure modified in line with (Ran et al. 2013)).

i. Restriction of sSpCas9(BB)-2A-Puro (PX459)

Component	Amount (μl)
2 μg Plasmid (μg/ml)	4
CutSmart ® buffer	5
20 U BbsI	2
H_2O	Ad 50 µl

Each reaction was incubated at 37°C in a heat block for 90 min, followed by heat inactivation at 65°C for 20 min.

ii. Purification of restricted sSpCas9(BB)-2A-Puro with QIAquick Gel extraction kit

First, 300 μl of QG and 100 μl of isopropanol were added to the restricted plasmid PX459. The mixture was transferred to a column and centrifuged at 16,100 g for 1 min. Second, 500 μl of QG Buffer were added and centrifuged at 16,100 g for 1 min. Third, 750 μl of PE buffer were added and centrifuged at 16,100 g for 1 min. Fourth, the flow-through was discarded, the column was dried in air and the filter tube was placed on an Eppendorf cup. Finally, 20 μl of ddH_20 were used to dissolve the purified product and centrifuged at 16,100 g for 1 min. The DNA concentration was determined photometrically by measuring the optical density at 260 nm according to the Lambert Beer law. The quality can be monitored by forming the ratio of 260/280 (pure DNA 1.8), indicating protein contamination, and the ratio of 260/230 (pure DNA 2.3), indicating carbohydrate contamination.

iii. Oligo annealing

For each sgRNA pair, 5 μl of 100 µM sense oligonucleotides and 100 µM antisense oligonucleotides were mixed and incubated at 95°C for 5 min and cooled down to 4°C at -0.1°C/s.

iv. Ligation of sgRNA pairs into purified PX459

Each sgRNA oligo pair was diluted 1:200. The following ligation reaction was set up for each sgRNA oligo pair and one negative control (without sgRNA oligo pair) in an Eppendorf tube:

Component	**Amount (μl)**
Restricted, purified PX459 (25ng/µl)	2
Each sgRNA oligo pair (X3, X4, X22)	1
ddH_20	5
Total	8

After the reagents had been heated to 55°C for 1 min, the mixture was cooled down. Then, 1 µl of T4 DNA ligase reaction buffer and 1 µl of T4 ligase were added. The mixture was incubated at room temperature for 30 min. Afterwards,

5 µl of each reaction were directly transformed into *E.coli DH5α* cells while the rest was incubated at 4°C over night as a backup.

2.3.2.6 Transformation

Chemically competent *E. coli* cells can be obtained in the presence of calcium chloride which enhances cell wall permeability and enables DNA intake (Chan et al. 2013). For each plasmid, 50 μl of competent DH5∝ E. coli were thawed on ice and 5 μl of the ligation product were added and incubated on ice for 10 min. The following 45 s heat shock further increased cell wall permeability further and improved transfection efficiency. The reagent was returned to ice for 2 min. 100 μl of 37°C warm LB medium were added and the bacteria incubated in the shaker at 37°C for 30 min. Then the bacteria were plated onto LB-agar plates containing 100 μg/ml of ampicillin. The plates were incubated overnight at 37°C.

2.3.2.7 Sanger sequencing: sequence validation of cloned sSpCas9(BB)-2A-Puro

i. Overnight culture of single clones

Two colonies were picked from each plate: a sterile pipette tip was used to inoculate a single colony into a 7.5 ml culture of LB medium supplemented with 100 μg/ml of ampicillin. The culture was incubated overnight in a shaker at 230 rpm at 37°C.

ii. Plasmid preparation (Mini QIAprep Spin Miniprep Kit (250) ®)

Plasmid preparation is a method to isolate plasmid DNA from a bacterial culture. Here, the QIAprep spin miniprep kit was used according to the manufacturer´s protocol which enables the extraction of up to 20 µg of plasmid DNA.

iii. Sanger sequencing of plasmid

Sanger sequencing is a sequencing method where a DNA polymerase incorporates desoxynucleotides (dNTP) and dedesoxynucleotides (ddNTP) coincidentally during replication (Sanger, Nicklen, and Coulson 1977). The ddNTPs cause a strand break and are marked by four different fluorescence dyes to differentiate the four different nucleotides (adenine, cytosine, guanosine,

thymine). The resulting DNA products are heat-denaturated, separated by chromatography and visualized by laser.

Here, Sanger sequencing was used to validate the successfully ligated sgRNAs: (a) Exon 3 (X3), (b) Exon 4 (X4) and (c) Exon 22 (X22) within PX459.

The following reagents were prepared in a 96-well plate:

Component	Amount (μl)
Forward or reverse primer (10 pmol)	0,5
Plasmid DNA (100 ng/μl)	3,0
Big dye (dNTPs)	1,0
Big dye Buffer	1,5
H_2O	4,0
Total	10

The 96 well plate was put into a thermocycler with the following temperature protocol:

Temperature	Time	Number of cycles
96°C	2 min	× 1
96°C	10 s	
58°C	5 s	× 30
60°C	3 min	
94°C	1 min	
59°C	1 min	× 10
72°C	1 min	
60°C	1 min	× 1
10°C	Hold	

Afterwards, paramagnetic beads were used for the purification of sequencing PCR products. The DNA binds selectively to the surface of the paramagnetic beads by using ethanol and is collected by applying a magnetic plate. Other particles (e.g. primers, taq DNA polymerase or dNTPs) stay in solution. After the particles have been removed by repeated washing steps, the plasmid DNA was re-suspended by adding water.

First, 10 μl of bead solution were added to the 96-well plate. Second, 42 μl of 85% EtOH were added, inverted several times and placed on the magnetic plate, incubated for 3 to 5 min and discarded again. 100 μl of EtOH were added again and discarded after 30 sec. This step was repeated twice. Finally, 40 μl of H_20 were added to resuspend the DNA. The 96-well plate was placed on the magnetic plate again to bind the paramagnetic beads. 25 μl were transferred to the sequencing plate, which was placed in the Sanger sequencing machine (ABI3730).

After the reaction had taken place in the Sanger sequencing machine, the obtained sequences were examined for agreement with the sequence of the respective sgRNA. In case of a positive result, the corresponding clone was used for the CRISPR/Cas9 transfection.

2.3.2.8 Glycerine culture

After successful ligation of the different sgRNA into PX459 had been validated, the sequenced plasmid was transformed into DH5∝ *E. coli* (according to Section 2.2.2.6). A glycerine culture was made from each ligation product for long-term storage of each bacterial clone.

For this purpose, 850 µl of the transformed DH5∝ *E. coli* suspension were added to 150 µl of glycerine in a cryo-tube on ice and immediately stored at -80°C.

2.3.2.9 Transfection: Genome editing of HEK293 cells with CRISPR/Cas9

Transfection is the transfer of foreign DNA into eukaryotic cells. The different available methods can be divided into viral and non-viral categories. Non-viral approaches include physical methods such as electroporation or microinjection

and chemical methods like calcium phosphate, lipofection or cationic polymers (Kim and Eberwine 2010).

Here Xfect, a biodegradable transfection polymer, was used. Negatively charged DNA forms a complex with the polycation and is absorbed by clathrin-mediated endocytosis/pinocytosis.

i. HEK293 transfection efficiency: GFP transfection

HEK293 cells were transfected with the plasmid pEGFPN1 to measure their transfection efficiency. pEGFPN1 encodes the green fluorescent protein (GFP) which emits green light with an emission peak at 509 nm that can be detected by a fluorescence microscope. This is due to its ability to form internal chromophores when excited by blue or UV light (excitation peaks at 395 nm and 475 nm) (Remington 2011).

The prepared pEGFPN1-plasmid was transfected with Xfect according to the manufacture´s protocol on 60%-confluent HEK293 cells which had been sowed at a concentration of 0.0715 x 10^6/well on Poly-L-lysine coated cover slips in 6-well plates the previous day.

After incubating for 24 hours, the transfected cells were washed with 1 ml of PBS and fixed with 200 μl of 4% paraformaldehyde (PFA) at room temperature for 10 min. After PFA had been discarded, the fixed cells were washed twice with 1 ml of PBS. The fixed cells were counterstained with DAPI, a fluorescent stain that binds to AT-rich regions of double-stranded DNA of alive and fixed cells.

The fixed cells were placed under the fluorescence microscope. In a fluorescence microscope, the light from a light source is passed through an excitation filter (Sanderson et al. 2014). It allows the wavelength to pass through, which is necessary for excitation of the fluorescent dye from an object of interest (excitation peaks: DAPI at 358nm; GFP at 395nm and 475nm) but does not pass the wavelength spectrum in which the fluorescence dye emits light (emission peaks: DAPI at 461nm; GFP at 509nm) (Kapuscinski 1995). The fluorescence generated by the object of interest is measured with a detector.

Then the green fluorescent cells and the blue fluorescent cells were counted in an image. The green fluorescent cells were the successfully transfected cells, while the blue fluorescent cells indicated all living cells. Transfection efficiency was calculated by measuring the proportion of green fluorescent cells on all blue fluorescent cells.

ii. Intrinsic HEK293 resistance: puromycin assay

The antibiotic puromycin inhibits protein synthesis of prokaryotic and eukaryotic cells by causing premature chain termination during translation (Kandala et al. 2018). A puromycin assay was conducted to examine the natural resistance of HEK293 cells against puromycin. The determined minimum deadly concentration was later used to select successfully transfected HEK 293 cells from non-transfected HEK 293 cells.

For this purpose, HEK293 cells were first dissociated from the 10 cm Petri dish and the cell concentration was ascertained by a Neubauer counting chamber. Second, the cell solution was diluted with D10 to a final concentration of 0.0530 $\times 10^6$ cells/ml and 0.5 ml of the ultimate cell solution was pipetted to each well of a 24-well plate. Three wells were transfected with PX459 using Xfect. After incubating for 24 h, the HEK293 reached 60% confluency and the different puromycin amounts (μg/ml) were added to the 24-well plate (0,25 µg, 0,5 µg, 1 µg, 2 µg, 3 µg, 4 µg, 5 µg).

The three wells, containing PX459 transfected HEK293 cells, served as a positive check. The plasmid included the PAC gene which supplied the cells with a puromycin N-aceythyltransferase. Hence, these cells should have survived even a high puromycin concentration.

After incubating for 24 h, the medium was removed and the cells were rinsed with PBS and dissociated with 100 μl of Trypsin/EDTA. 400 μl of D10 stopped the reaction and the cell concentration was determined with a Neubauer counting chamber.

iii. Transfection of 293 HEK cells with cloned PX459

24 h prior to transfection, 500 μl of 1.3 $\times 10^6$/ml well-dissociated HEK293 cells were placed into each plate of six 24-well plates in D10 medium. For each plasmid (sSpCas9(BB)-2A-Puro ligated with sgRNA targeting Exon 3; sSpCas9(BB)-2A-Puro ligated with sgRNA targeting Exon 4; sSpCas9(BB)-2A-Puro ligated with sgRNA targeting Exon 22), one 24-well plate was transfected with Xfect: 1 μg of plasmid DNA was mixed with Xfect reaction buffer to a final volume of 25 μl and 0.3 μl of Xfect polymer was added. Next, the mixture was vortexed and incubated for 10 min. Finally, the solution was added dropwise to the 60%-confluent HEK293 cells of the 24 well plate and incubated for 24 h. In addition, 1 μl of 10μM ssODN template was added to each well to enable HDR, for each sgRNA. After 24 h of incubation, 3 μg of puromycin were added to each well to select successfully transfected HEK293 cells.

iv. Single clone extraction by dilution

Single cells were generated by dilution to receive genetically homogenous cell lines which can be tested for successful genetic modifications later on. 24h after puromycin selection, D10 medium was removed from the transfected cells of the 24 well plate. The cells were rinsed with PBS and dissociated with 100 μl of Trypsin/EDTA. The reaction was stopped by adding 900 µl of D10 and the cell concentration was determined with a Neubauer counting chamber. The cell suspension was diluted to a final concentration of 1 cell/10 μl. 10 μl of each condition were added to the matching 96-well plate, which had already been prepared with 100 µl of 1:1 fresh and conditioned D10 medium. For each condition, eight 96-well plates were prepared. The 96-well plates were incubated at 37°C and in 5% CO_2 for three weeks.

v. Culturing of successfully transfected HEK293 for back-up

When reaching 90-100% confluency, each cell clone of a 96-well plate was expanded. After rinsing the cells with 1 ml of PBS, the cells were dissociated with 100 µl of Trypsin and re-suspended in 900 µl of D10 medium. 200 µl of the cell suspension were transferred to a 6-well backup plate, filled with 800 µl of preheated D 10 medium. The remaining 800 µl of the cell suspension were

transferred to another 6-well plate and filled with 200 µl of preheated D10 medium for subsequent sequencing. The HEK293 cells from the 6-well plate were expanded later on to ten 10 cm Petri dishes after several splitting steps.

vi. Preparation of genomic DNA (QIAamp® DNA extraction kit)

About 90 to 100%-confluent HEK293 cells were harvested from the 6-well plate for sequencing by dissociating them with 200 µl of Trypsin and re-suspending them in preheated D10 medium. Next, the cell suspension was spun down at 200 g for 5 min. Then the remaining medium was discarded. The DNA was extracted from the pellet, using the QIAamp ® DNA extraction kit according to the manufacturer´s protocol.

vii. Polymerase chain reaction (PCR)

The DNA obtained from each cell clone was amplified in a PCR to generate more template DNA for subsequent sequencing. PCR is a molecular biological method, where a DNA template can be amplified *in vitro* (Saiki et al. 1988): In repeated cycles, a thermostable DNA polymerase, derived from *Thermus aquaticus,* synthesizes the complementary DNA strand of a template using dNTPs during elongation after denaturation of the template strand and primer annealing **(Figure 11).**

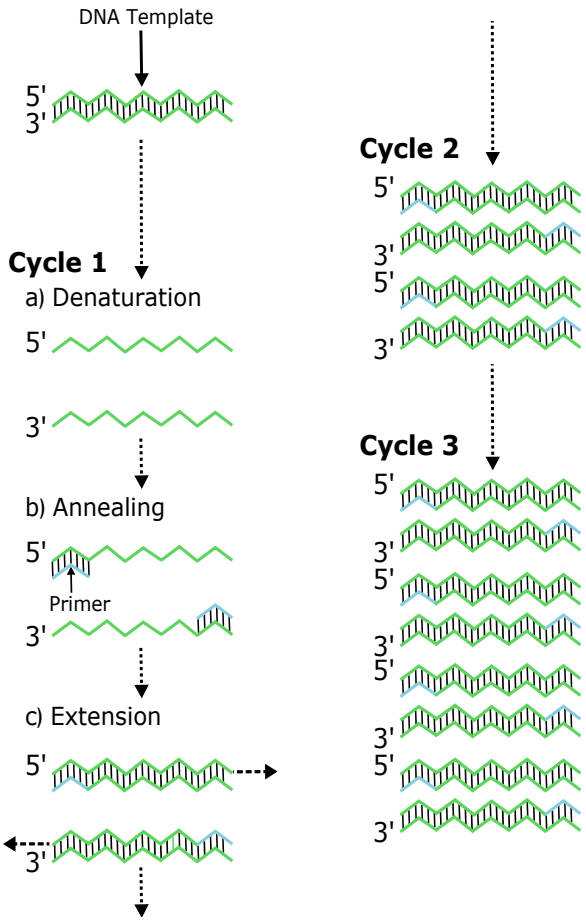


Figure 11 Schematic overview of PCR
(modified in line with (Garibyan and Avashia 2013)).

For each extracted DNA sample, the following reagents were prepared in a 96-well plate:

Component	Amount (µl)
dNTPs	0.4
Taq polymerase buffer	2.0
Taq polymerase	0.2
DNA	1.0
Primer forward (10pM)	0.8
Primer reverse (10 pM)	0.8
H_20	14.8
Total	20.0

After the sample plate had been sealed, the following temperature protocol was applied in a thermocycler:

Temperature	Time	Number of cycles
94 °C	3 min	× 1
94 °C	1 min	
68-59 °C	1 min	× 30
72 °C	1 min	
94 °C	1 min	
59 °C	1 min	× 10
72 °C	1 min	
72 °C	10 min	× 1
15 °C	Hold	

viii. Primer design

The corresponding primers had to be designed for the PCT. This was done in the following steps. In general, PCR primers should consist of 15-25 nucleotides (primer for cDNA 18-20 bp). PCR primers should have a guanin-cytosine content of 50-60%, a melting temperature of between 55°C to 65°C and the difference in melting temperature between forward and reverse primer should not exceed 5°C. A primer should end with a guanine and cytosine nucleotide (cg-clamp) to increase primer annealing efficiency. Further, a primer should not bind to intramolecular secondary structures or dimers. The selected primers were tested for these parameters, using the NetPrimer online software (http://www.premierbiosoft.com/NetPrimer/AnalyzePrimer.jsp). The specific binding of each primer pair was ensured, using UCSC *in silico* PCR (https://genome.ucsc.edu/cgi-bin/hgPcr).

ix. Agarose gel-electrophoresis

The PCR products were applied to a 1.5% agarose gel for amplification control. In an agarose gel-electrophoresis, nucleotide acids can be separated by mass and size due to different running speeds through the agarose pores in an electric field (Lee et al. 2012). Moreover, they can be visualized by the DNA-binding agent ethidium bromide. The size of an unknown DNA can be estimated by a DNA ladder which includes fragments of a known size that result in a specific ribbon pattern.

For the 1.5% agarose gel, 1.3 g of agarose were dissolved in 80 ml of TBE by being heated in a microwave until no gel-remains were visible. Subsequently, 2 µl of 1% ethidium bromide solution were added. The solution was put into a gel-electrophoresis chamber, gel combs attached and the gel cooled down for 20 min.

Next, 5 µl of PCR product and 1 µl of loading buffer were mixed and applied to each gel pocket. 120V was applied to the gel chamber for 45 min.

x. Exonuclease digestion

An exonuclease digestion of the PCR products was run to remove primers and nucleotides (Boissinot et al. 2007). The used exonuclease removes nucleotides in the 3´ to 5´ direction but leaves 5´-terminal dinucleotides, which prevents a breakdown of the entire PCR product. This step is important because the PCR primers could otherwise interfere with the sequencing primers in the subsequent sequencing reaction. In addition, the PCR nucleotides could falsify nucleotide concentrations, resulting in false sequencing results.

The following reagents were prepared:

Component	Amount (µl)
H_20	1.0
Antarctic phosphatase buffer	1.25
Antarctic phosphatase	1.0
Exonuclease buffer	1.25
Exonuclease I	0.5
Total	5.0

The reaction was incubated at 37°C for 30 min, inactivated at 80°C for 15 min and cooled down at 10°C for 15 min.

Sanger sequencing of the respective CRISPR/Cas9 target region was conducted to examine whether Cas9 had successfully cleaved the target locus and had led to the desired genomic alteration. Sequencing was conducted as already described in Section 2.2.2.7.

2.3.3 Transcriptome analysis

A RNA-Seq of three wildtype KMT2E cell lines versus three CRISPR KMT2E knockout cell lines was implemented to investigate whether KMT2E affects gene expression or splicing through alterations of chromatin structure.

2.3.3.1 RNA-Seq method

RNA-Seq uses next-generation sequencing technologies for transcriptome profiling. It also provides information about the quantity of different mRNA transcripts, splicing variants, post-transcriptional modifications, non-coding RNAs or small RNAs at a precise moment (Wang, Gerstein, and Snyder 2009). After RNA isolation, RNA is first converted into complementary DNA (cDNA) by a retroviral reverse transcriptase, as RNA sequencing is technically challenging. Next, the cDNA is fragmented and sequencing adaptors are ligated to the one or both ends. After a possible amplification, depending on the used sequencing platform, each DNA fragment is sequenced in a high-throughput form in order to gain short sequences from one (single-end sequencing) or both ends (pair-end

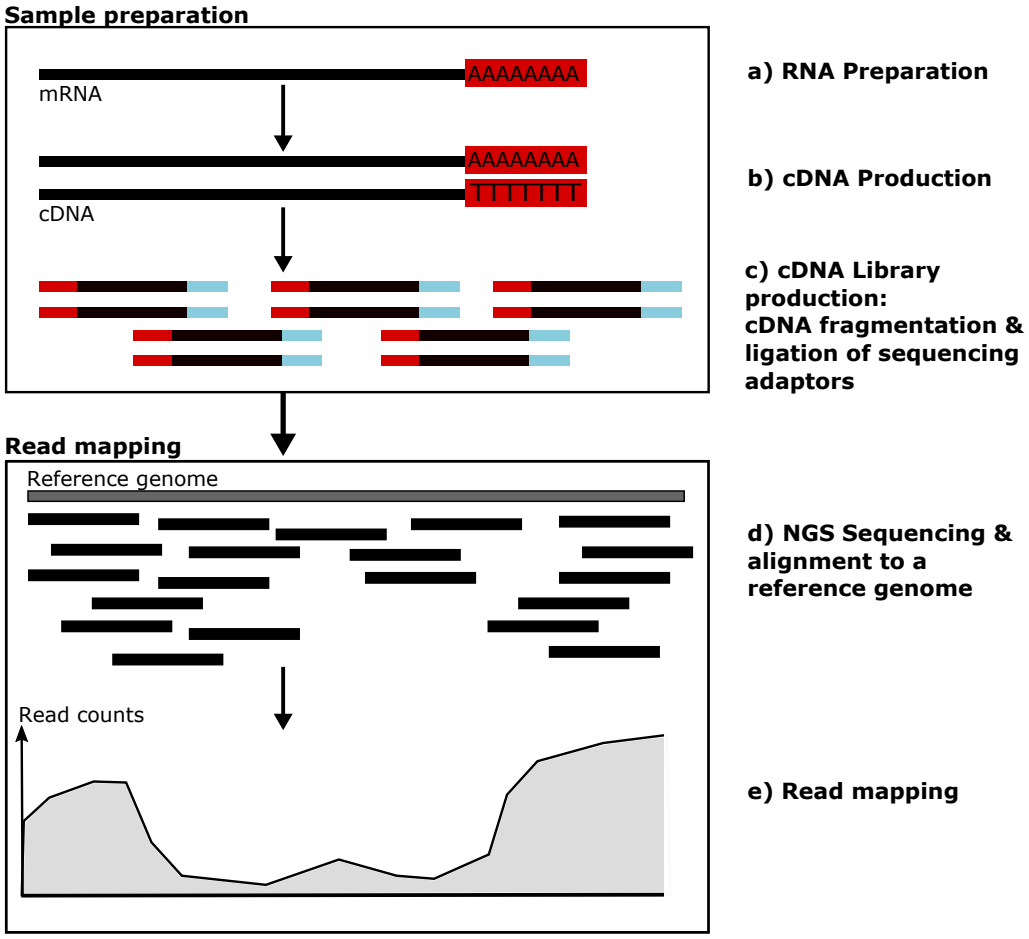


Figure 12 Principle of RNA-Seq
After RNA has been transcribed into cDNA, it is fragmented and sequencing adaptors are ligated to each fragment which can be sequenced by NGS. The sequence reads are aligned to a reference genome and classified as exonic, junction or poly(A) end-reads which create a base-resolution expression profile for each gene (figure in line with (Wang, Gerstein, and Snyder 2009)).

sequencing). The resulting 30-400 bp long reads are then aligned to a reference genome and categorized as exonic reads, junction reads or poly-A-end reads. These reads are then used to generate a base-resolution expression profile for each gene **(Figure 12).**

The RNA-Seq workflow was modified and applied as demonstrated in (**Figure 13)**.

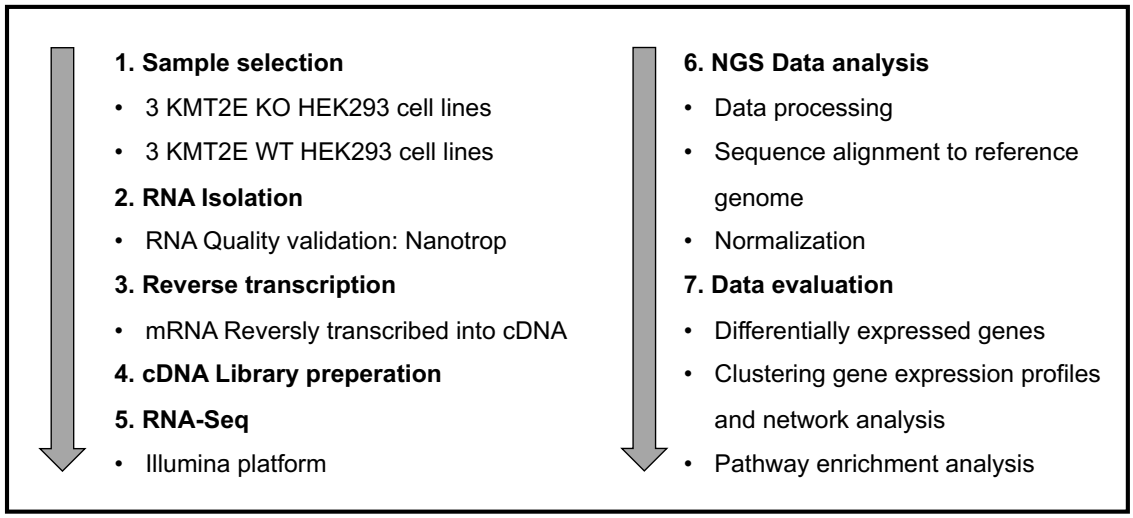


Figure 13 RNA-Seq workflow

2.3.3.2 RNA isolation

Before sequencing RNA, it has to be isolated from cells. For each cell line, a 100%-confluent 10 cm Petri dish was pelleted according to the previously described protocol in order to obtain 1 µg of RNA. A RNeasy mini-kit from Qiagen was used according to the manufacturer´s protocol for RNA isolation. Subsequently, a RNase-Free DNase set was used to remove any traces of DNA. The RNA was kept on ice at all times to prevent degradation by RNAse due to its short biological half-lifetime.

2.3.3.3 Implementation of RNA-Seq

The RNA-Seq was carried out by the c.ATG Core Facility for NGS and Microarrays Tübingen using an Illumina HiSeq 2500 sequencer in paired-end sequencing mode (2x86 bp).

2.3.3.4 Implementation of NGS data analysis

The aligned reads were re-processed by AG Epigenetik to identify genome-wide significantly differentially expressed genes. They used the following bioinformatic tools for their analysis: *STAR* v2.5.4b was used for read mapping against the homo sapiens genome from *Ensembl* v91. *MultiQC* and *FastQC* were employed for quality control of the raw data and mapping data. *GenomicAlignments* package and s*ummarizeOverlaps* were applied for read counting of the aligned data. Differential expression was calculated with DESeq2.

2.3.4 Validation of prioritized candidate genes by qRT-PC

qRT-PCR was implemented as a gold standard to validate the RNA-Seq data and the analyses of the expression profiles, the gold standard for validation (Bustin et al. 2009). qRT-PCR is a laboratory method which measures the concentration of a PCR template during amplification in real time. The two common methods detect the PCR template either (i) by the intercalation of a non-specific fluorescent dye in any double-stranded DNA or (ii) by DNA target-specific fluorescent reporters which emit light only after hybridization. The emitted fluorescence rises in accordance with an increase in double-stranded DNA. Here, the non-specific fluorescent Cyanine dye SYBR Green I was used, which forms a strong fluorescent complex (absorption maximum 497 nm; emission maximum 520 nm) with double-stranded DNA and RNA. For subsequent mathematical calculation of the expression ratio, the crossing point (Cp) is determined for each sample. Cp is defined as the number of amplification cycles, at which the fluorescence rises noticeably above the background

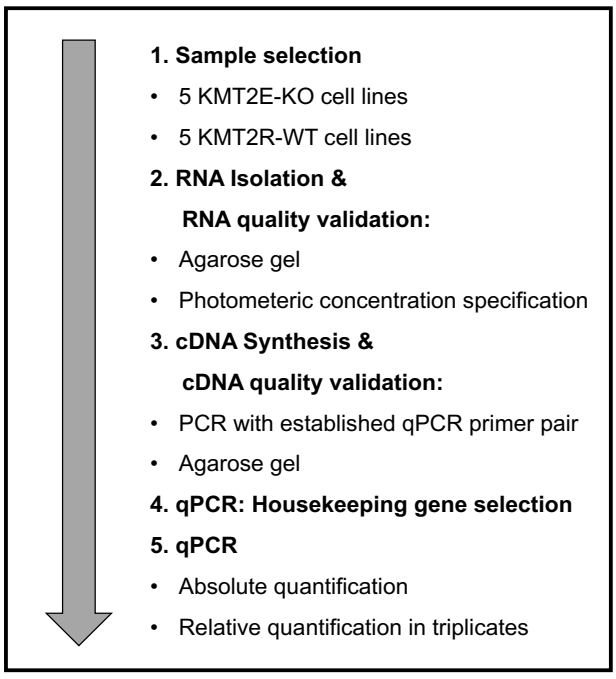


Figure 14 qRT-PCR workflow

fluorescence. Afterwards, the DNA can be quantified in relative terms by normalizing the measured values against an endogenous check.

The gene expression level of 11 selected candidate genes was determined in five different knockout and wildtype cell lines, including the cell lines used for RNA-Seq. All PCR primers (Section 8.1) were designed as described in Section 2.2.5. The qRT-PCR experiment workflow is shown in (**Figure 14**).

i. RNA isolation

RNA isolation according to Section 2.2.3.2.

ii. RNA quality control

The quality of the extracted RNA was checked by determining the ratio of 260/280nm for protein contamination (~2.0 for RNA) and the ratio of 260/230 nm (~2.0 for RNA) for carbohydrate contamination in a spectrophotometric analysis, using the photometer.

iii. Complementary DNA (cDNA) synthesis: reverse transcription

Since a classical PCR can only be conducted with DNA, RNA is transcribed into complementary DNA (cDNA) by the viral-derived enzyme reverse transcriptase. A subsequent DNA wipeout by a DNAse prevents any genomic DNA contamination of the qRT-PCR.

The QuantiTect reverse transcription kit from Qiagen was used according to the manufacturer´s protocol for all 30 samples. For each reaction, 2 µl of gDNA wipeout buffer were mixed with 1 µl of template RNA and 11 µl of RNase-free water in an Eppendorf cup on ice. The reaction was incubated at 42°C for 2 min and immediately placed on ice. 1 µl of reverse-transcription master mix, 4 µl of Quantiscript RT buffer and 1 µl of RT primer mix were added and mixed on ice. First, the reaction was incubated at 42°C for 15 min and then, the reverse transcriptase was inactivated at 95°C for 3 min. The samples were stored at -20°C.

iv. cDNA quality validation: PCR and agarose gel

A PCR was conducted with 2 µl of each obtained cDNA sample according to the protocol described in Section 2.2.9 (vii) with an established cDNA primer. Each

PCR product was applied to 1.5% agarose gel for amplification control according to the protocol from Section 2.2.9 (viii).

v. qRT-PCR

The SYBR Green ® kit from Qiagen was prepared in the following composition for each reaction in a 384-well plate:

Component	Amount (in µl)
SYBR Green Master Mix	5
Forward primer	0.5
Reverse primer	0.5
H_20	2
Total	8

1 µl of the matching cDNA samples was added separately to the 384-well plate.

Each qRT-PCR was run in a 384 Light Cycler using the following temperature program:

Temperature	Time	Number of cycles
95°C	2 min	1
95°C	5 s	40
60°C	10 s	40

Intra-assay precision, accuracy and reproducibility were determined by triplicates of each condition within a cycle. Inter-assay variation was determined by three different, independent experimental runs, using biological replicates and different premix samples of the Light Cycler and SYBR Green kit.

vi. Determination of primer efficiency by calculating amplification efficiency

The efficiency of each primer pair was determined because the subsequent calculations of the expression levels of a gene of interest relies on the assumption that each template is replicated in each PCR cycle with a primer efficiency of

100% (Svec et al. 2015). If the primer efficiency actually differed from that efficiency, the subsequent calculations would have been incorrect, resulting in false expression level calculations.

Here, the efficiency (*E*) of a PCR was referred to the proportion of templates that were replicated in each PCR cycle. It should have exceeded 90% (Lalam 2006). The PCR efficiency was estimated by calibration with a relatively controlled quantity of cDNA. The standard curve was generated by serial dilution of the cDNA using 10-fold dilution steps. The Cp values of these prepared standard samples were measured in a qRT-PCR according to the protocol described in Section 2.2.11 (v). In a diagram, the Cp values were plotted against $\log_{10}$ of the diluted sample concentration **(Figure 15)**. Next, the correlation r of the data points against the straight line was determined to monitor the reliability of the experiment (e.g. exact dilution). The Pearson correlation coefficient r should have exceeded 0.955. Each standard curve was measured in technical triplicates. The efficiency *(E)* of a PCR primer was then easily calculated, inserting the slope of the obtained straight line in the following formula:

$$E = 10^{-\left(\frac{1}{slope}\right)} - 1 \qquad (1)$$

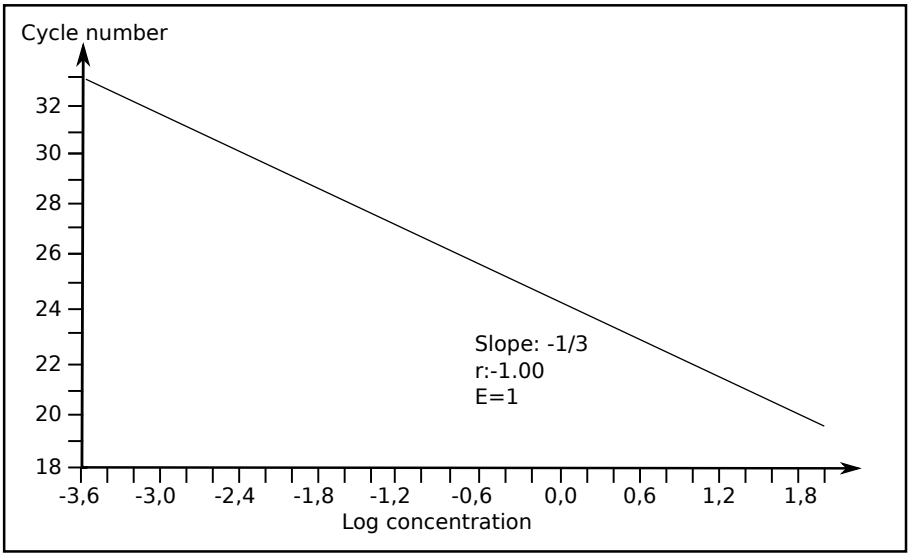


Figure 15 Example of a standard curve for testing the efficiency of a primer pair
The x-axis shows the log concentration of the cDNA; the y-axis shows the measured Cp values.

vii. Melting curve analysis

After each qRT-PCR, a melting curve analysis was performed to verify the specificity of the PCR product because SYBR Green I binds non-specifically to both short and long chain molecules in the minor groove (Pfaffl 2001). For this purpose, the fluorescence was detected during a temperature decrease from 90°C to 50°C after each amplification. When reaching the specific melting temperature, determined by molecule length and nucleotide composition, the fluorescent dye is released suddenly and almost completely because the double-stranded DNA denatures. The relationship between changes in fluorescence against temperature can be illustrated in a diagram, called the "melting curve" **(Figure 16)**. In general, the number of peaks corresponds to the number of existing amplicons. An exception are products which possess a GC-rich sequence, since these amplicons are more stable and do not melt immediately but maintain their double-strand configuration up to a higher temperature, resulting in more than one melting phase.

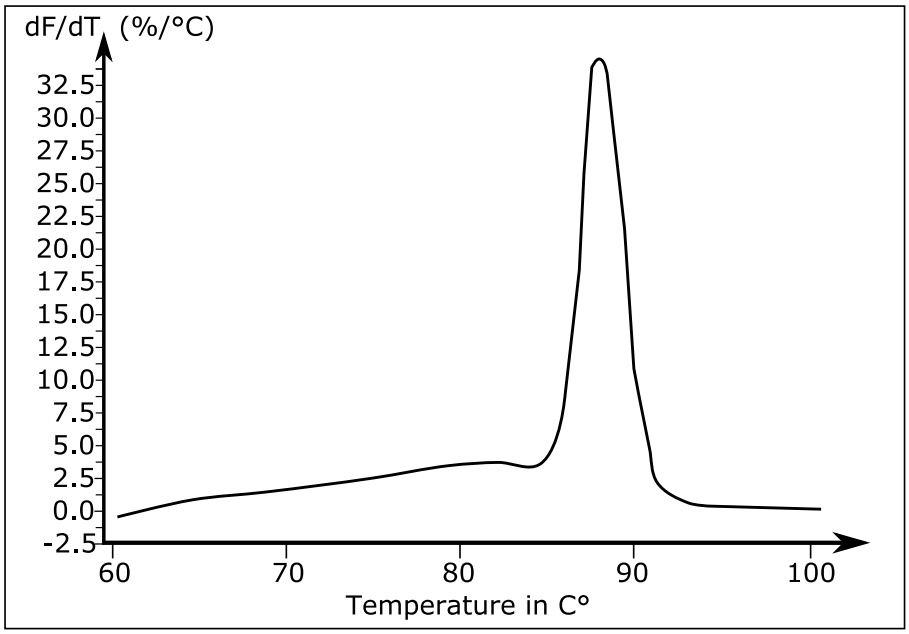


Figure 16 Example of a melting curve
The specificity of a PCR product was checked with a melting curve after each qRT-PCR (Pfaffl 2001). For this purpose, the change in fluorescence (y-axis) was detected during a temperature decrease (x-axis). Each double-stranded product denatures at a specific temperature and releases all its bound fluorescence at once, resulting in a fluorescence peak. However, if non-specific products are produced, there would be different peaks, depending on the number of unspecific PCR products.

viii. Calculation of expression levels of the target genes

Two different methods can be used to determine the expression level of a target gene:

(i) Absolute quantification

In absolute quantification, the expression level of a target gene is measured relative to a set of standards which are used to construct a standard curve (Bustin 2000).

(ii) Relative quantification: $2^{-\Delta\Delta Ct}$ method

In relative quantification, the expression level of a target gene is quantified relative to an endogenous control gene. This method was applied here. Before the expression ratio of a target gene was calculated, the variation of the Cp values was investigated by an analysis of variance (ANOVA) to check for outliers, according to Section 2.2.7.1. The adjusted Cp values were then used for the downstream calculations.

For this calculation, only the PCR efficiencies and the Cp differences between the target gene and the endogenous control are needed (Pfaffl 2001; Rao et al. 2013).

Using the notation system from **Table 3**, ΔCp for the knockout sample is CT_D- CT_B and ΔCp for the wildtype sample is CT_C- CT_A:

ΔCp = Cp (target gene)- Cp (endogenous control)

ΔΔCp is the difference in ΔCp between the knockout and wildtype sample:

ΔΔCp= ΔCp (knockout sample)- ΔCp (wildtype sample) = (CT_D- CT_B)-(CT_C- CT_A)

Finally, the result is presented as the fold change of the target gene expression in a knockout sample relative to a wildtype sample, normalized to an endogenous control. The relative gene expression is set to 1 for the first wildtype sample because here ΔΔCp is equal to 0 and hence 2^0 is equal to 1. The method assumes a PCR efficiency of 100% across all samples, which has been previously tested in Section 2.2.11 (vi). The value 2 is q plus the PCR efficiency of 1.

Table 3 Experimental design

	Wildtype sample	Knockout sample
Reference gene (Endogenous control)	A	B
Target gene	C	D

ix. Selection of appropriate endogenous controls

For the purpose of obtaining reliable estimates of the gene expression levels, so-called “reference genes” are included as endogenous controls, because their evenly expression throughout different cell lines qualifies them as a suitable normalization system (Eisenberg and Levanon 2013; Chapman and Waldenstrom 2015). Irrespective of the initial template concentrations, expression changes for a target gene can be calculated when target genes and endogenous controls are measured simultaneously. Hence, reference genes should be independent of experimental treatments and unbiased by the expression of the target genes.

For that purpose, prior to measuring the expression of the identified target genes from the RNA-Seq, several genes were tested for their suitability as reference genes. Widespread-used reference genes and evenly expressed genes in knockout and wild-type samples from the RNA-Seq were tested. From the tested genes, three genes (*GAPDH*, *ACTB1*, *C1orf43*) were chosen as reference genes and added to each qRT-PCR plate.

2.3.5 Biochemical methods

2.2.5.1. Western Blotting

Western blotting is a method to identify and quantify proteins through specific antibodies **(Figure 17)** (Clark and Pazdernik 2016). This biochemical method was applied to compare the global methylation levels regarding H3K4me1, H3K4me2, H3K4me3 between 3 KMT2E-WT and 3 KMT2E-KO HEK293 cell lines. Prior to

western blotting, each examined HEK293 was harvested and pelleted at 100% confluency from 10 cm Petri dishes.

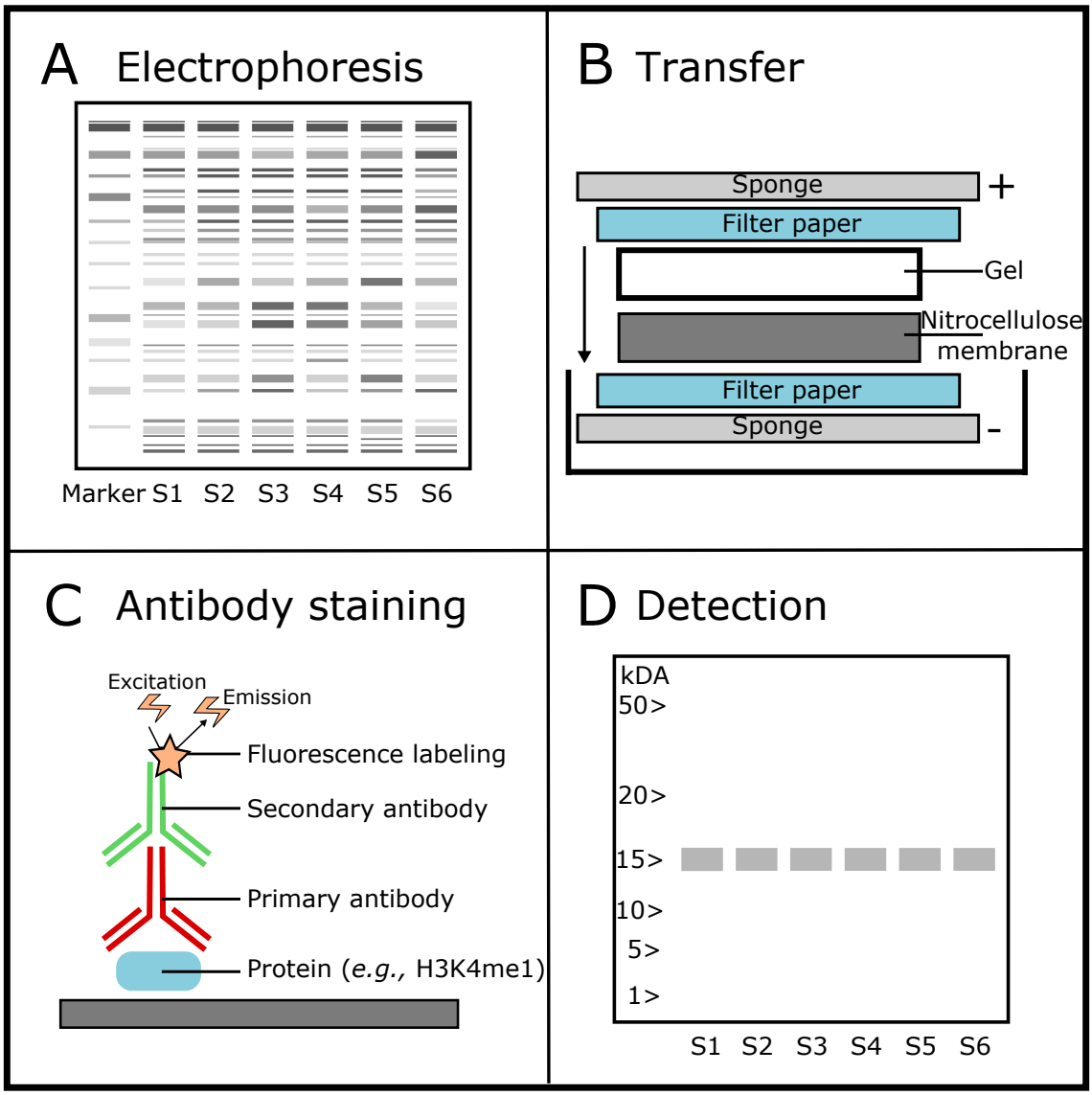


Figure 17 Workflow of classic western blotting
After proteins have been separated in a gel electrophoresis by their molecular mass (A), they are transferred to a membrane (B). A protein of interest is detected by a primary antibody which is identified by a fluorescent secondary antibody (C). The fluorescence can be detected by a luminometer (D) (modified in line with (Ghosh, Gilda, and Gomes 2014)).

Histone extraction

Before the global histone methylation levels could be examined by means of western blotting, the histones had to be extracted according to an established

protocol (Rumbaugh and Miller 2011). One tablet of protease inhibitor was added to the homogenization buffer before use. All steps were carried out on ice.

500 µl of homogenization buffer was added to each pellet which was thawed on ice. The samples were homogenized by pulling a syringe with a 0.5 mm needle up and down five times. Then, the homogenate was centrifuged at 7,700 g at 4°C for 1 min to pellet the nuclei, the supernatant was removed and 500 µl of 0.4 NH_2SO_4 were added. Next, the samples were incubated on ice for 30 min, interrupted by short vortexing every 10 min. Subsequently, all samples were centrifuged at 16,800 g at 4°C for 10 min. The supernatant was transferred to a 1.5 ml Eppendorf cup and the pellet discarded. 250 µl of trichloracetic acid and sodium deoxycholoate were added to the supernatant to precipitate the histones. The samples were inverted slowly by hand until the suspension turned cloudy. Next, each sample was incubated on ice for 30 min. Afterwards, the suspension was centrifuged at 16,000 g at 4°C for 30 min and the supernatant discarded. The remaining pellet was rinsed with 1 ml of cold (4°C) 100% acetone to remove acid from the solution without damaging the histones. After carefully inverting the samples and incubating 5 min on ice, the samples were centrifuged at 16,000 g at 4°C for 5 min. The supernatant was discarded and the pellet dried for 5 min. Then, the histones were re-suspended in 7 µl of DTT, 15 µl of 1x SDS-sample buffer and 50 µl of 10 mM Tris-HCl (pH 8.0). Finally, the samples were stored at -80°C.

SDS-polyacrylamide gel electrophoresis (Laemmli 1970)

Prior to the first step of the western blotting, a discontinuous SDS-polyacrylamide gel – consisting of a separating and stacking gel – was prepared in a provided holder according to **Table 4**. The stacking gel is less concentrated and has a lower pH value. It concentrates all proteins so that all proteins start running from the same point.

Table 4 Components of discontinuous SDS-polyacrylamide gel

Component	15% separating gel (16 ml)	4% stacking gel (10 ml)
ddH_20	3.7 ml	6 ml
30% acryl/bis-acrylamide solution	8 ml	1.33 ml
1.5 M Tris pH 8.8	4 ml	
0.5 M Tris pH 6,8		2.5 ml
10 % SDS	160 µl	100 µl
When ready to use		
TEMED	16 µl	10 µl
10% APS	160 µl	100 µl

The gel was transferred to an SDS-page electrophoresis chamber which was filled up with transfer buffer and 0.1% SDS. The transfer buffer consisted of 3.03 g/l of Tris and 14.41 g/l of glycine. SDS binds to the hydrophobic regions of the proteins, causing a negative loading of all proteins. Hence, the proteins migrate to the anode at a different speed, which depends only on their size but not on their charge. Moreover, proteins with the same migration behavior should appear as a single band on the same level. Each chamber was filled with 10 µl of a sample from the previous histone extraction and Bromophenol-blue. Bromophenol-blue migrates before the protein front and indicates the progress of electrophoresis. The last chamber was filled with a molecular marker (Page Ruler™ Prestained Protein Ladder, Thermo Scientific™) in order to be able to identify the molecular weight. Gel electrophoresis was run at 150 V for 30-45 min until the Bromophenol color front had vanished.

Western blotting (Towbin, Staehelin, and Gordon 1979)

After SDS-PAGE gel electrophoresis, the proteins were separated according to size into single protein bands. The negatively charged protein bands were transferred from the gel to a positively charged nitrocellulose membrane by

applying an electrical field vertical to the SDS-PAGE gel. The gel and nitrocellulose membrane were placed between filter paper in a blotting cassette. The proteins bind on the nitrocellulose membrane in the same pattern as on the SDS-PAGE gel due to hydrophobic interactions. However, the proteins bind more strongly to the nitrocellulose membrane than the gel, making it more suitable for further methods.

The blotting buffer consisted of 20% of methanol, 3.03g/l of Tris and 14.41 g/l of glycine. The blot was run at 110 V for 60 min.

Ponceau staining (Bannur et al. 1999)

A Ponceau stain was used to check whether the protein transfer was successful. It stains proteins as red bands against a clear background. The staining is reversible and allows further immunological detection methods. The nitrocellulose membrane was covered in the Ponceau stain for 5 min. Afterwards, the Ponceau stain was removed, the red protein bands evaluated and washed off with TBS repeatedly until the color fainted.

Immune coloring

Afterwards, the nitrocellulose membrane was incubated with 5% BSA to saturate surplus binding positions and prevent unspecific antibody binding. For this purpose, BSA was added to 1x TBS. Next, the primary antibody Anti-Histone 3 and either Anti-Histone3-monomethylationK4 (H3K4me1), Anti-Histone3-dimethylationK4 (H3K4me2) or Anti-Histone3-trimethylationK4 (H3K4me3) was added 1: 1 000 to the 5% BSA solution and incubated for at 4°C for 45 min. The antibody against H3 serves as a loading control because it exists evenly in all samples whereas the antibodies H3K4me1, H3K4me2 or H3K4me3 are directed against the protein of interest. Afterwards, the 5% BSA solution containing both the antibody against the protein of interest as well as the antibody against the uniform existing protein was removed and the nitrocellulose membrane was washed with TBST three times for 5 min. After 5% BSA had been solved in TBST, the secondary fluorescent marked antibody was added 1: 10 000 to the solution. It is conjugated with fluorophore and binds to the heavy chain of the primary

antibodies, visualizing the target protein and loading control. The secondary antibody was incubated with the nitrocellulose membrane for 45 min. Then the secondary antibody solution was removed, and the nitrocellulose membrane washed with TBST three times for 5 min. Finally, the nitrocellulose membrane was washed with TBS briefly. The protein of interest, e.g. monomethylated lysine 4 of histone 3, was detected by our LI-COR.

2.3.6 Biometric analysis

2.3.6.1 qRT-PCR

In the qRT-PCR, the data obtained for the endogenous control genes (*ACTB1*, *C1orf43*, *GAPDH*) was first subjected to an analysis of variance (ANOVA) using the following model (Snedecor 1980):

$$y_{ijkl} = \mu + r_i + g_j + e_k + rg_{ij} + ge_{jk} + \varepsilon_{ijkl} \qquad (1)$$

where

y_{ijkl} is the C_P value observed on replicate i for gene j in entry k in replication l

μ the general mean

r_i the effect of replicate i, $i = 1,2,3$

g_j the effect of gene j, $j = 1,2,3$

e_k … the effect of entry k, where $k = 1,...,5$ refers to wildtype (WT) 1 to WT5 and $k = 6,...,10$ to knockout (KO)1 to KO5

rg_{ij} the interaction effect of gene j with replicate i

re_{ik} the interaction effect of entry k with replicate i

ε_{ijkl} the experimental error of the l^{th} measurement of gene j in entry k in replicate i

All terms in this model were assumed to be fixed effects except the error term ε_{ijkl} which was assumed to be independently distributed (iid) with ε_{ijkl} : $N\left(0, \sigma_\varepsilon^2\right)$.

By using the test of Anscombe and Tukey (Tukey 1963), we detected one outlier which was treated as a missing value for calculating least squares estimates of $\hat{y}_{i.k.} = \hat{\mu} + \hat{r}_i + \hat{e}_k + \hat{e}_{ik}$.

For each target gene (*EFNB3*, *HSPG2*, *Notch3*, *NRTN*, *PDEA4*, *SCN4B*), the adjusted C_P values y_{ikl} (obtained by subtracting the means $\hat{y}_{i.k.}$ of the endogenous controls from the original C_P values) were subject to a similar ANOVA as for Experiment 1 using the following model:

$$y_{ikl} = \mu + r_i + e_k + \varepsilon_{ikl} \ , \qquad (2)$$

where μ and r_i are defined as in Eq. (1) and ε_{ikl} is iid with ε_{ikl} : $N\left(0, \sigma_\varepsilon^2\right)$. Following Experiment 1, we assumed e_k : $N\left(\mu_W, \sigma_{e_W}^2\right)$ for k = 1, …5 (WT genotypes) and e_k : $N\left(\mu_K, \sigma_{e_K}^2\right)$ for k = 6, …10 (KO genotypes). Checking for outliers and homogeneity of the variances $\sigma_{e_W}^2$ and $\sigma_{e_K}^2$ and calculating estimates of μ_W and μ_K with corresponding standard errors was performed as described above. Finally, the null hypothesis $\mu_W = \mu_K$ was tested with a two-sided t-test with 8 degrees of freedom (Snedecor 1980).

From the means $y_{ik.}$, we calculated the relative quantification (RQ) of gene expression $z_{ik.}$ according to (Yuan et al. 2006). These values were subject to an ANOVA using the same model as in Eq. (1) and applying exactly the same tests as described above. For graphical representation, the RQ values were re-scaled by the mean value of the WT genotypes.

2.3.6.2 Western blotting

In Experiment 1, each wildtype (WT) and knockout (KO) genotype was analyzed for the expression level of (i) a control protein (*H3 loading*) and (ii) the three proteins of interest (H3K4me1, H3K4me2, H3K4me3).

For each protein $H3K4meI$ ($I = 1,2,3$), an ANOVA was conducted with normalized values of the ratio $\frac{H3K4meI}{H3\ loading}$ using the following statistical model:

$$y_{ik} = \mu + r_i + e_k + \varepsilon_{ik} \quad , \qquad (1)$$

where

y_{ik} is the data point for technical replicate i of entry k

μ the general mean

r_i the effect of technical replicate i, $i = 1,2,3$

e_k … the effect of entry k, where $k = 1,2,3$ refers to WT1 to WT3 and $k = 4,5,6$ to KO1 to KO3

ε_{ik} the experimental error of technical replicate i of entry k

While the effect r_i was treated as fixed, all other effects were considered as random with e_k being iid with e_k : $N\left(\mu_W, \sigma^2_{e_W}\right)$ for $k = 1,2,3$ and e_k : $N\left(\mu_K, \sigma^2_{e_K}\right)$ for $k = 4,5,6$, and ε_{ik} iid with ε_{ik} : $N\left(0, \sigma^2_{\varepsilon}\right)$. First, we performed an outlier test according to Anscombe and Tukey (1963), but no outlier was detected. Second, after checking for homogeneity of the variances $\sigma^2_{e_W}$ and $\sigma^2_{e_K}$ with an F-test, we obtained least squares (LS) estimates of μ_W and μ_K with corresponding standard errors. Finally, the null hypothesis $\mu_W = \mu_K$ was tested using a two-sided t-test with 4 degrees of freedom. For graphical representation, all results were re-scaled by the mean value of the WT genotypes.

3. Results

3.1 Genetic and clinical data of our patient

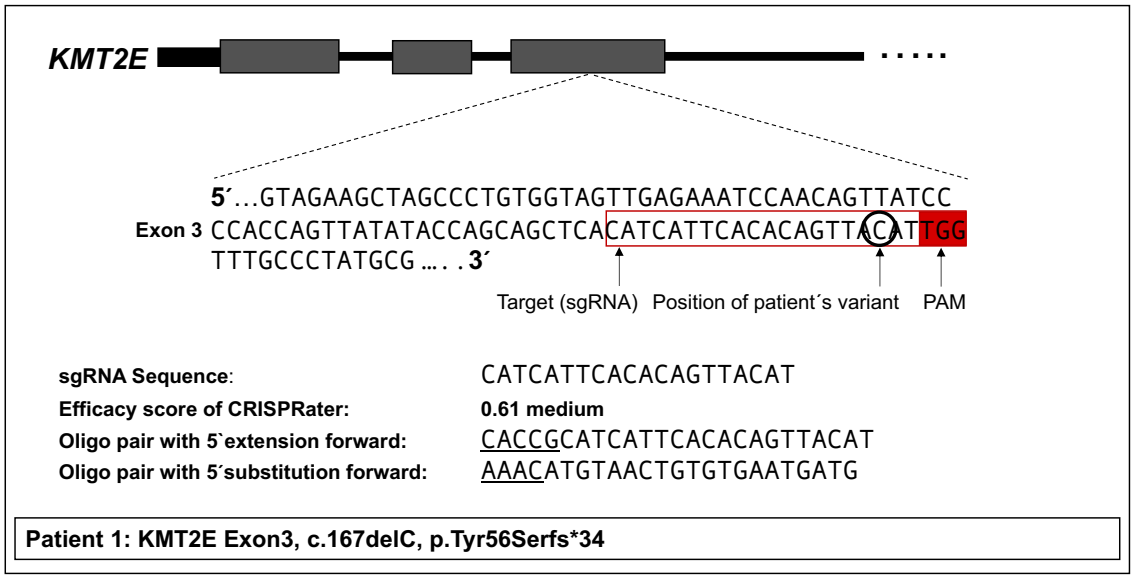
The patient studied here (listed as patient 2 in the following) was included in a genetic study about heterozygous *KMT2E* variants with 37 other patients (O'Donnell-Luria et al. 2019). The study was conducted by an international consortium to which our working group contributed. The variants were assembled through the GeneMatcher networks the Matchmaker Exchange Network and MyGene 2 from September 2016 to August 2018. The genetic and clinical data of our patient can be found in Figure1, Table1 and Table S2 of the paper.

3.2 Successful application of CRISPR/Cas9 in HEK293 cell lines

The *KMT2E* variants of known patients were modelled successfully in HEK293 cells using the CRISPR/Cas9 method.

3.2.1 Design of sgRNA in close proximity to the variants of three patients

A sgRNA, targeting the region of the variant, was first successfully designed with cctop (http://tools-genome-engineering.org). This tool suggested the following sgRNAs in close agreement within the patients' variants **(Figure 18)**.



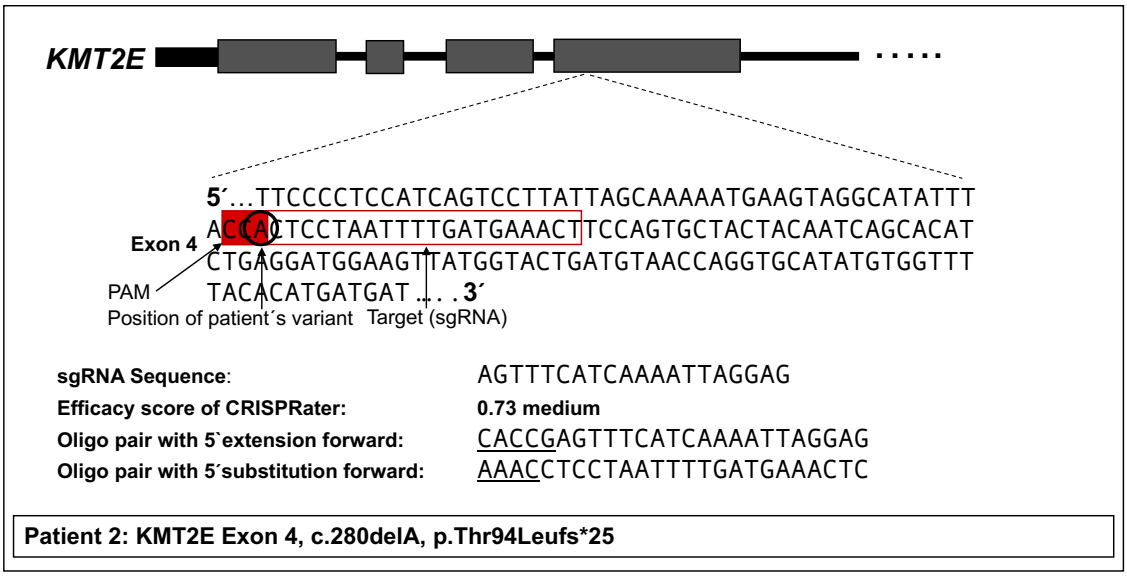


Figure 18 sgRNA designs
A section of the exon sequence is shown, where the respective patient has a variant, marked by a circle. The sgRNAs were designed to be as close to the variant as possible in order to increase the rate of successful HDR with the designed ssODNs which contain the patient's variant. The target sequence is circled red together with the PAM sequence. The calculated efficiencies are given beneath, together with the used oligo pairs.

Design of repair template: ssODN

The manually designed ssODN which served as a repair template for HDR after CRISPR/Cas9 had introduced a double-strand break as close as possible to the position of the patients' variant were successfully constructed **(Figure 19).**

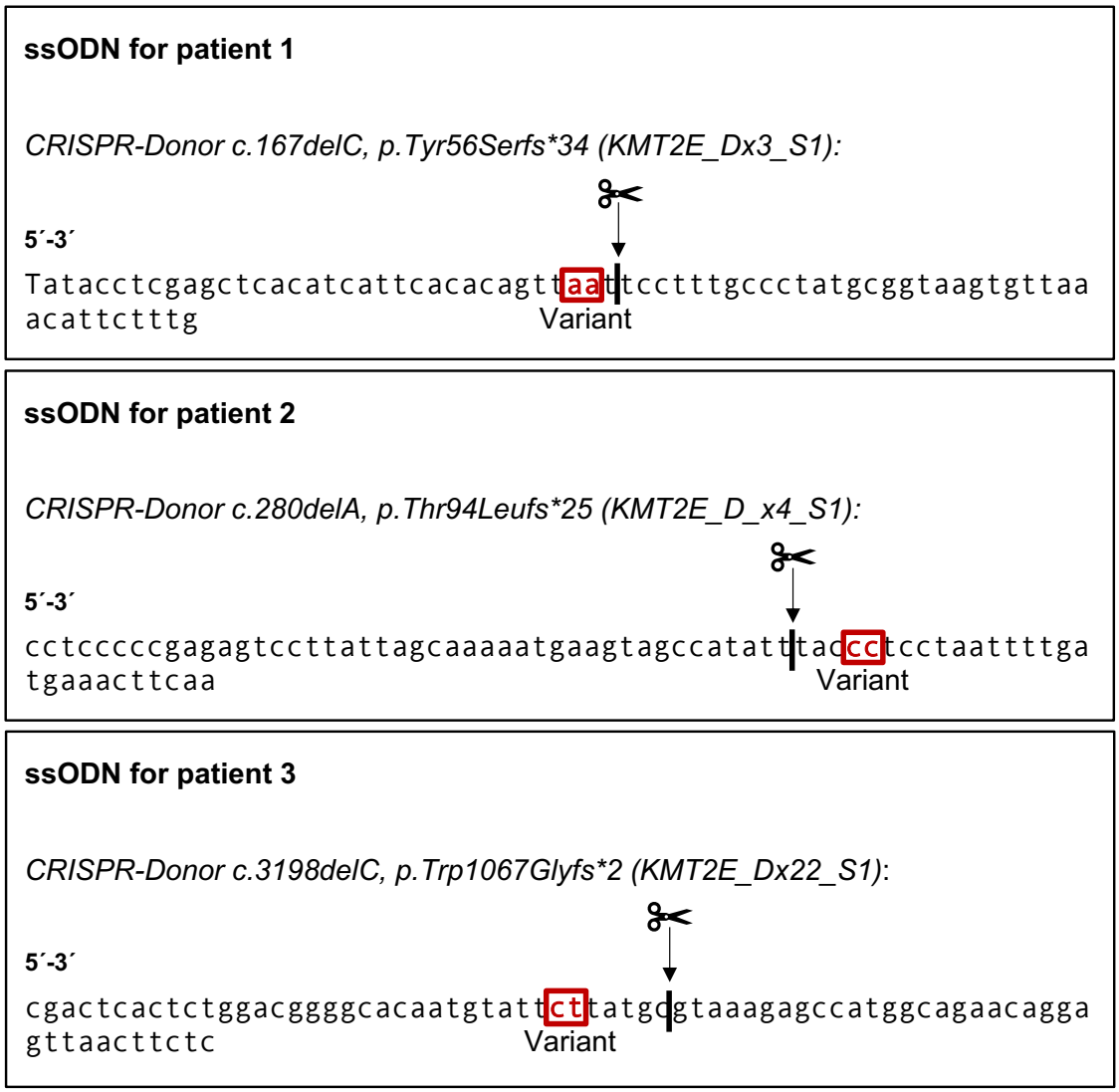


Figure 19 Manually designed ssODN for HDR
The variant of the respective patient is marked in each ssODN. In addition, the double-strand breaks are labeled, where the CRISPR/Cas9 system introduces the double-strand on the DNA.

3.2.2 Validation of sgRNA in CRISPR/Cas9 expressing plasmid PX459 by Sanger sequencing

First, one of the three specific sgRNAs was correctly cloned into the CRISPR/Cas9 expression vector PX459. After transformation, the plasmids were successfully purified from the bacteria culture. For each patient, several successfully transformed plasmids with the respective sgRNA were detected. The correct insert and orientation were verified by Sanger sequencing **(Figure 20, Figure 21)**.

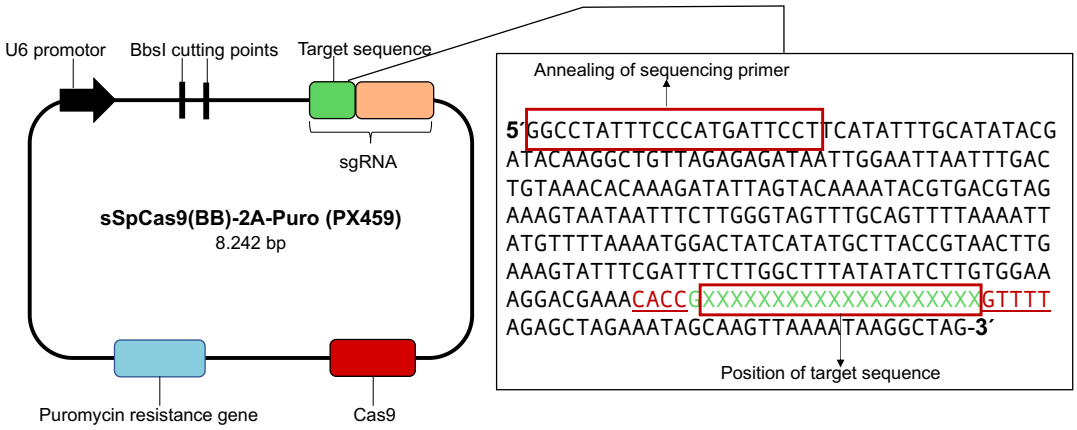


Figure 20 Demonstration of sequencing PX459
After extraction of the plasmid from each bacterial clone, it was amplified with a sequencing primer in a standard PCR.

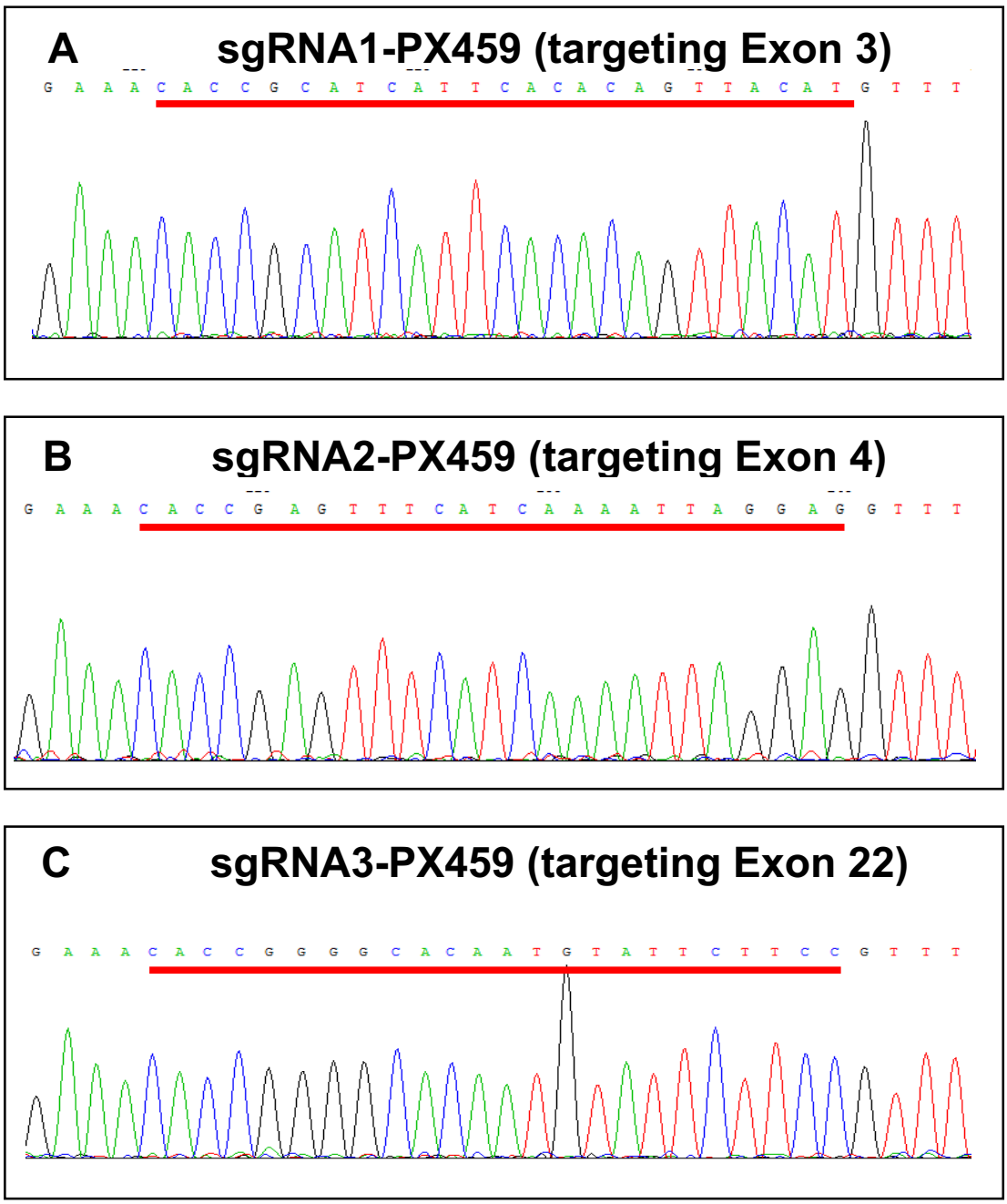


Figure 21 Sanger sequencing results of ligated insert sgRNA into CRISPR/Cas9 plasmid PX459
For each variant, positive clones with the correct inserted sgRNA sequence and orientation were identified, which are marked with a red line.

3.2.3 Validation of Xfect as a suitable transfection agent

HEK293 cells were successfully transfected with the GFP-expressing plasmid pEGFPN2 using Xfect. The transfection efficiency was between 50% and 90% in different experimental approaches, proving Xfect as an appropriate transfection agent for HEK293 cells **(Figure 22)**.

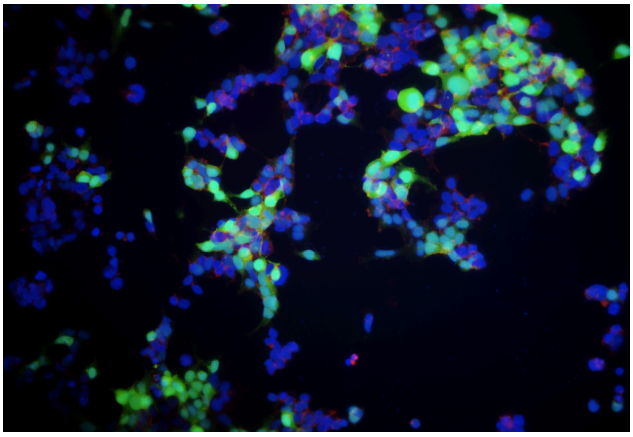


Figure 22 peGFPN2 transfected HEK293 cells, stained with DAPI
The green fluorescent HEK293 cells were an indicator that they had been successfully transfected with pEGFPN2 through Xfect. DAPI stained the cell nucleus. The proportion of green fluorescent cells among all blue fluorescent cells displays the transfection efficiency.

3.2.4 Validation of Puromycin as a selection marker

A puromycin assay successfully tested the intrinsic resistance of the HEK293 cells against the antibiotic and established the puromycin concentration for the CRISPR/Cas9 experiment, where puromycin is used as a selection marker for successfully transfected cells. The experiment showed that HEK293 cells exhibit no internal puromycin resistance with a concentration around 3 µg/ml **(Figure 23).**

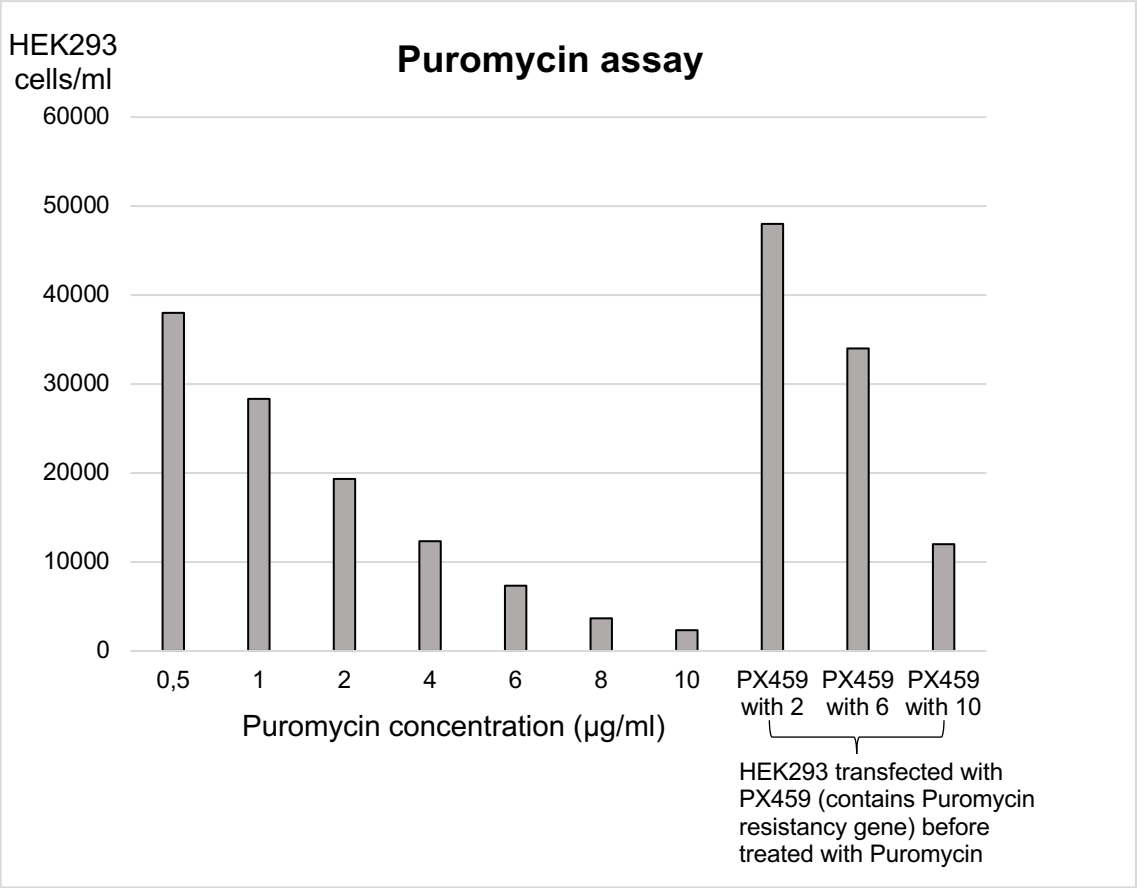


Figure 23 Puromycin assay
HEK293 cells were transfected with puromycin to test them for internal resistance. 3µg/ml was chosen as the minimal inhibitory concentration.

3.2.5 Identification of successfully genetic modified HEK293 cell clones via Sanger sequencing

After transfection of the CRISPR/Cas9 vector in HEK293 cells, the single cells grew to clones and the DNA from over 70 single-cell clones was successfully extracted. After the extracted DNA was amplified in a standard PCR, the amplification was checked with an agarose gel. Then, Sanger sequencing of the extracted DNA revealed seven cell clones with genetic modifications **(Figure 24)**. This translated into a CRISPR efficiency of less than 10% **(Table 5)**. All positive clones showed small deletions or insertions on all three *KMT2E* alleles, resulting in a complete knock out of *KMT2E***.** The desired variants of the patients were not found – HDR via the ssODN had not been successful. Instead, NHEJ had repaired the double-strand breaks which had been introduced by Cas9.

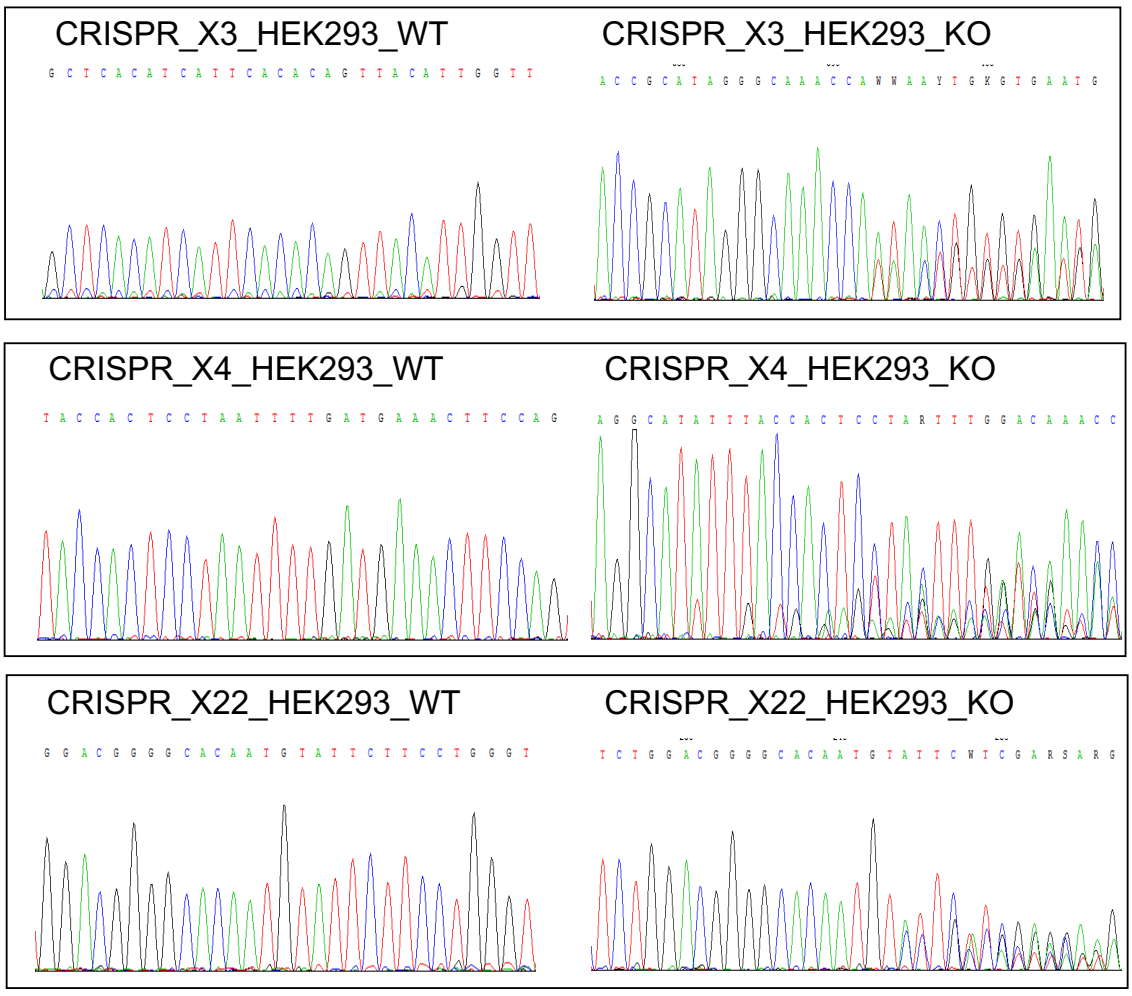


Figure 24 Sanger sequencing results of single-cell clones
Sanger sequencing results from the HEK293 cell clones which were later used for RNA-Seq. All knockout cell lines show different small insertions or deletions on each allele

the different traces of color show. At least 2 of the knockout cell lines have 3 *KMT2E* alleles as shown by the triple peaks. The patient variants were not detected in any clone.

Table 5 Efficiency of CRISPR/Cas9 experiment

Number of wildtype and knockout cell lines which were produced for each variant in the CRISPR/Cas9 experiment. The efficiency for each exon was on average 8.9%.

Target	WT 293 cells	KO 293 cells	CRISPR efficiency
sgRNA1 (exon 3)	21	2	9.5 %
sgRNA2 (exon 4)	42	4	9.8 %
sgRNA3 (exon 22)	16	1	6.3%

3.3 Transcriptome analysis

3.3.1 RNA-Seq processing and alignment

The RNA-Seq analysis of the 3 WT *KMT2E* vs. 3 KO *KMT2E* cell lines resulted in a total number of 24.6 to 29.3 million reads for each library.

3.3.2 Detection of 30 genes with differential gene expression analysis in *KMT2E*-KO und -WT cell lines

The RNA-Seq data between the 3 WT *KMT2E* vs. 3 KO *KMT2E* cell lines was compared. A differential gene expression analysis revealed 30 genes which were differentially expressed by a fold change of over 1.3 or less than 0.7 at the $P \leq 0.05$ significance level **(Table 6)**. Afterwards, the normalized counts in each sample were compared to test whether wildtype and knockout samples cluster within a gene.

Table 6 Genes with genome-wide significantly different expression level

The table shows the 30 genes which were differently expressed by at least 30% between the three *KMT2E* wildtype and knockout HEK293 cell lines. In addition, the function of each gene and the exact log fold change (log FC) are given. Brain expression during development and adulthood was also compared, using information about expression from the UCSC genome browser. It was suggested that the pathology of the patients was caused by shared pathways of KMT2E with other highly brain-expressed genes, the expression of which was influenced by the *KMT2E* variants. Further, the raw RNA reads

data was visually screened to check if the expression clustered strongly between the two groups (*KMT2E*-WT vs. KO). This is a strong indicator of an actual difference in gene expression and a strong hint that these genes can be validated in the following qRT-PCR. The genes shown in bold face were later used for qRT-PCR validation.

Gene	Function	Brain expression	logFC	Cluster
CHST	Carbohydrate sulfotransferase	High	1.58	Moderate
CPNE7	Calcium-dependent phospholipid-binding protein	High	1.02	Moderate
CSMD3	Regulation of dendrite development	High	0.87	Low
FOS	Transcription factor	Low	-1.72	Moderate
FOSB	Transcription factor	None	-2.22	Moderate
GABBR1	GABA-receptor	High	-0.68	Moderate
IGFN1		High	-2.07	Moderate
KCNJ4	**Potassium channel**	**High**	**1.25**	**High**
LTBP4	Latent transforming growth factor beta binding protein	Low	1.11	Moderate
PABPC1L	Polyadenylate-binding protein 1-like	Low	-0.85	Moderate
PDE4A	**Phosphodiesterase 4a**	**High**	**1.18**	**High**
PLPPR3	**Lipid phosphate phosphatase- related protein**	**High**	**1.46**	**High**
RANBP3L	Nuclear export factor	High	1.94	High
RASGRP2	Ras guanyl releasing protein 2	High	1.07	Moderate

S1PR5	**Lysosphingolipid sphingosine 1-phosphate**	**High**	**1.30**	**High**
SCN4B	**Sodium channel, voltage-gated, type IV**	**High**	**1.07**	**High**
SLC1A3	**Glutamate transporter family**	**High**	**0.97**	**High**
SLC25A43	Mitochondrial transporter	Low	1.13	Low
SORCS1	**Neurexin**	**High**	**1.10**	**High**
SULT1C2	Sulfotransferase	Low	-1.65	High
TCF24	Transcription factor	Low	1.17	High
TDRP	Testis development related protein	High	0.84	Low
TMEM38A	Transmembrane protein 38A	High	0.94	High
TSPO	Tryptophan-rich sensory protein	None	1.12	Moderate
ZNF253	Zinc finger protein 253	High	1.46	Moderate
ZNF347	Zinc finger protein 347	High	1.93	Low
ZNF471	Zinc finger protein 471	High	2.43	Low
ZNF667-AS1	ZNF667 Antisense RNA 1	High	3.50	Low
ZNF883	Zinc finger 883	high	1.02	Low

Taking into account fold change, brain expression and clustering between both groups, the following genes were chosen as strong candidates for qRT-PCR validation: *KCNJ4*, *PDE4A*, *PLPP3*, *SCN4B*, *SLC1A3*, *SORCS1*, *S1PR5*.

Four additional genes (*EFNB3, HSPG2, Notch3, NRTN*) were also tested in the qRT-PCR, which did not meet the criterion of at least 30%-fold change because they strongly clustered between both groups or were interesting candidates for interacting with KMT2E due to their high expression in the brain **(Table 7).**

Table 7 Genome-wide differently expressed genes

Four additional genes which were tested for validation in the qRT-PCR did not meet the criteria of at least 30%-fold change in the expression level between *KMT2E*-WT and -KO HEK293 cell lines. Nonetheless, they were interesting because they either clustered strongly between both groups or were highly expressed in the brain during development or in adulthood. Consequently, they were interesting candidates as they could potentially interact with KMT2E and provide an explanation for the phenotype ID/ASD of the patients.

Gene	Function	Brain expression	Cluster
EFNB3	Receptor- protein tyrosine kinase	High	High
HSPG2	Heperan sulfate proteoglycane 2	Low	High
Notch3	Neurogenic locus notch homolog protein 3	Medium	High
NRTN	Neurotrophic factor	Medium	High

3.3.3 Clustering of gene expression profiles and pathway analysis

We used ingenuity pathway analysis assigning the differentially expressed genes from RNAseq to metabolic and signaling pathways as well as associated diseases. This helped to characterize the transcriptional network of *KMT2E* in general and especially in neurodevelopment. Transcriptome profiling revealed that *KMT2E* knockout modified the expression of genes that were particularly enriched in dolichol and dolichyl phosphate biosynthesis, expoxysqualene biosynthesis, sperimine and spermidine biosynthesis **(Figure 25)**.

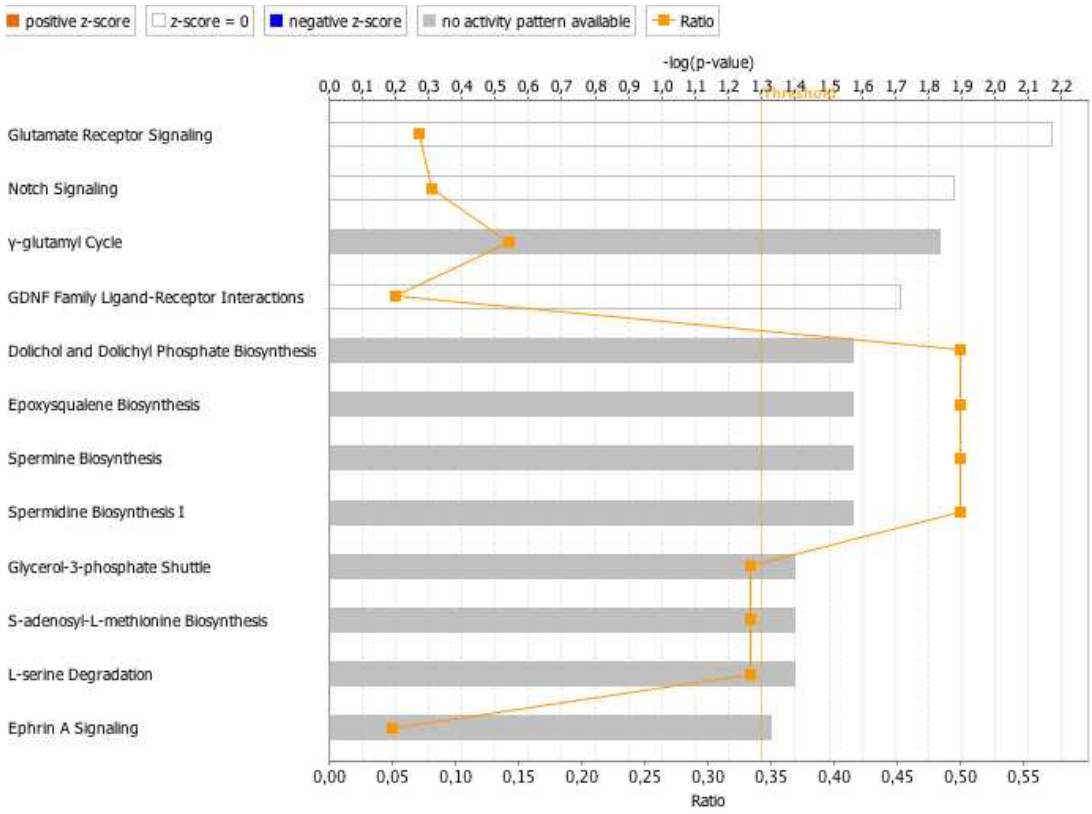


Figure 25 Ingenuity pathway analysis of KMT2E-KO affected pathways in HEK293cells

The difference in expression between KMT2E KO- and WT HEK293 cells is shown. Especially the dolichol and dolichyl phosphate biosynthesis, spermine and sperimidine I biosynthesis pathway are affected.

KMT2E KO seems to especially alter many genes associated to a certain disease complex like organismal movement or neurological diseases as shown in a synopsis **(Table 8)**.

Table 8 Top ten differentially expressed genes and their correlations with disease complexes

The table gives an overview of the most affected disease complexes which include the genes that were differently expressed in the RNA-Seq. The genes of the disease complexes, which were expressed differently in the RNA-Seq, are shown in boldface. The analysis was performed by the IPA software.

AGAP2, ***BACE1***, ***BSG***, *Cg, Creb,* ***CTSB***, ***DDIT3***, *E2f,* ***EPHA4***, *ERK,* ***ESM1***, ***FAM120A***, ***FOS***, ***FOSB***, ***FSCN1***, *Hdac,* ***HEY1***, ***IL1RAP***, ***KCNJ4***, ***LRP2***, ***MYCN***, *NFκB, Notch,* ***NOTCH3***, *PI3K,* ***PIM3***, *Pkc(s),* ***PRP11***, ***RARG***, ***ROR2***, ***RRM2***, ***ST3GAL1***, ***STC1***, ***TNFRSF11A***, *Vegf*	39	25	Organismal development, cellular movement, cancer
Akt, ***ASA H1****, BCR,* ***CSF2R8***, ***CYR61***, ***EFNA3****, ERK1/2, FSH,* ***GADD45A***, ***HELZ2***, ***HSPG2***, ***HYOU1***, ***IFITM1***, ***IGFBPS****, IgG,* ***INHBA****, Interferon alpha,* ***ITPR1****, Jnk, LDL, Lh, Mek,* ***NEU1****, P38 MAPK, PDGF 88,* ***PMAIP1***, ***RAP1GAP****, RNA polymerase II,* ***SEMA3C***, ***SMAD1****, TCF, TCR,* ***TOB2***, ***TRIB1***, ***TSPO***	29	20	Neurological disease, organismal injury and abnormalities, cellular movement
CAPN1, ***CHAC1****, DHCR7, DUSP4, EGR3,* ***EME1****,* ***EPM2AIP1****, FBP1,* ***FOS***, ***GPRIN1****, HHRNPC, IGFBP7, KCNJ2,KRT15, mir-145, MITF,* ***MLH1***, ***NFIC***, ***OSBPL7****,* ***OXCT1****, PHLDA2, S100A2, S8DS,* ***SEMA3C****, SERPINB2, SLC35C2, SMARCA4, SPP1,* ***TEM154***, ***TNFRSF9****, TOP1, TP53, TREM1,* ***ZBED6CL***, ***ZNF503***	19	15	Cellular assembly and organization, cellular development, cellular growth and proliferation
ARRDC3*, CCN1, CDC42EP3, CDCA7L,* ***CDH10***, ***EFNA4****,* ***EFNB3***, ***ENOSF1****, ERBB2, EREG, estrogen receptor, FGFR4,* ***FOS***, ***FSTL3***, ***HISTT1H4H****, ITGAS, ITGB4, JUND, KRT7, LAMC2,* ***MFAP2****, MMP14,* ***NRBP2****, NTSE, POLI, POLR3A,* ***POLR3C****, POLR3F,* ***POLR3H****, PPP1R1B, RAC1,* ***RASGRP2****, SATB1*	17	14	Cellular development, cellular growth and proliferarion
ACTN1, ***APBB3****, APOE, APP,* ***ATF5****,* ***AZI2****, CD40LG, CHRNA7,* ***DEAF1***, ***GABBR1****, GLUL,* ***GPCPD1****, HFE, Hsp90, IFNB1IL12A,* ***ITM2C****, LMNA,* ***MAT1A****,* ***MEGF9****, NGFR, NR3C1, NTRK1, NUPR1,* ***PMPCA****, PRMT5,* ***PTMS***, ***RIC3****, RTN4, SAFB, SNCA, SREBF1, TNFSF9,* ***TUBA8***, ***ZSWIM6***	17	14	Psychological Disorders, nervous system development and function, neurological disease

An IPA analysis of PDE4A, which could be validated as differently expressed between KMT2E-KO and WT HEK 293 cells both in the RNA-Seq and qPCR, showed different strong correlations with other ASD causing genes **(Figure 26)**.

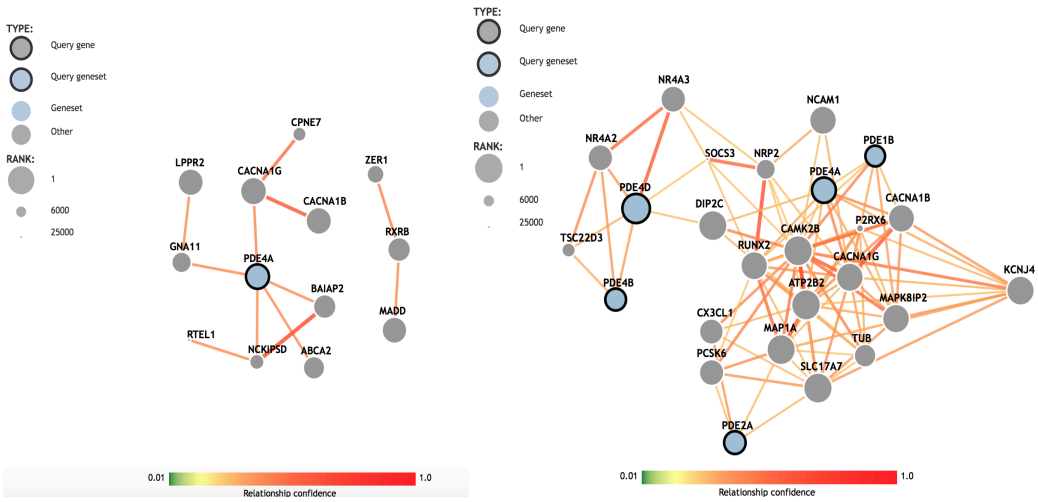
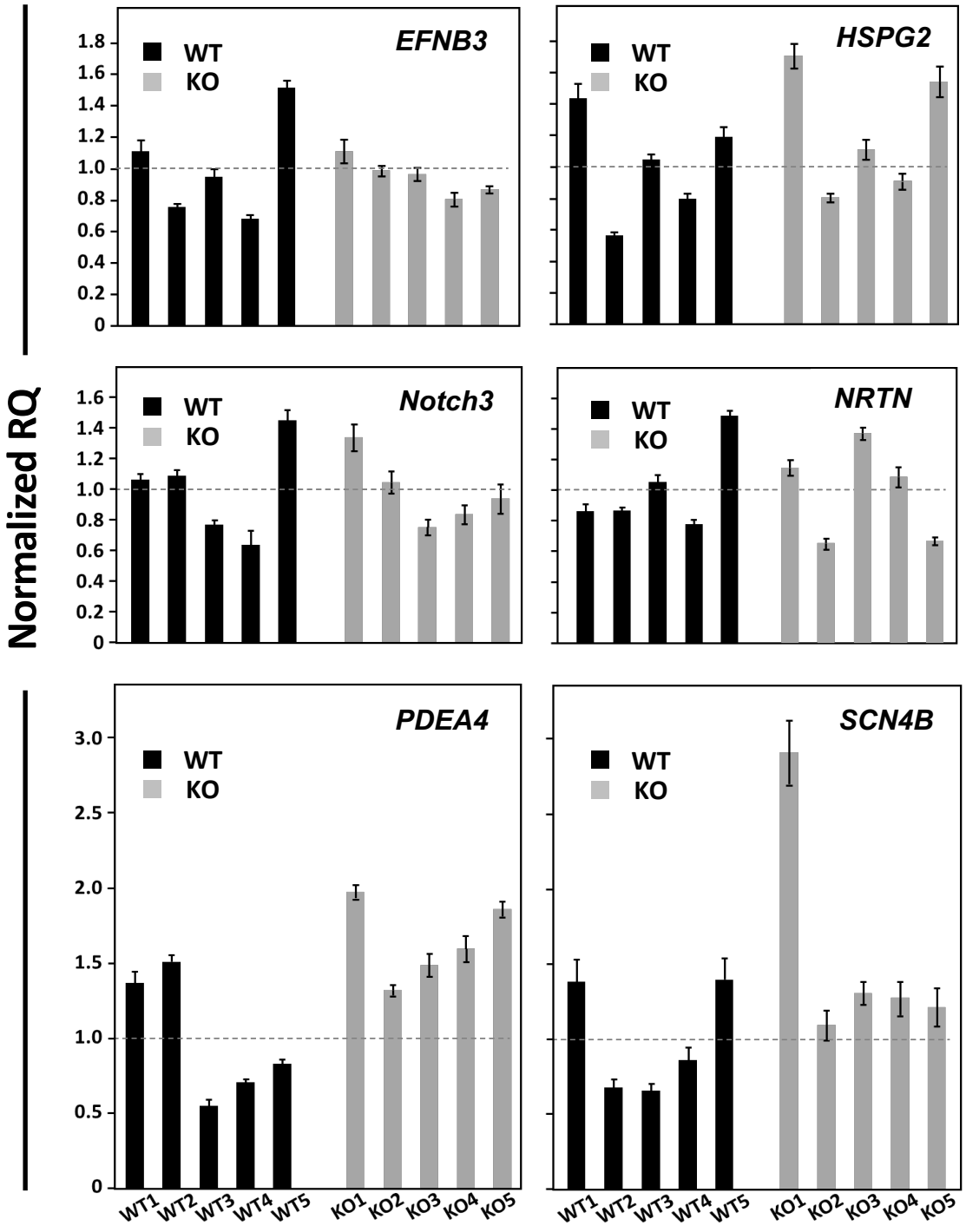


Figure 26 Association of PDE4A with different autism candidate genes
IPA associated PDE4A with different autism candidate genes and determined their relationship confidence.

3.4 qRT-PCR confirmed higher gene expression of *PDE4A* in *KMT2E*-KO cell lines

For validation of the RNA-Seq data, qRT-PCR was conducted with eleven genes chosen from the RNA-Seq according to the criteria described in Section 3.2. Significant expression differences between the *KMT2E*-WT and *KMT2E*-KO HEK293 cell lines were tested in qRT-PCR **(Figure 27)**.



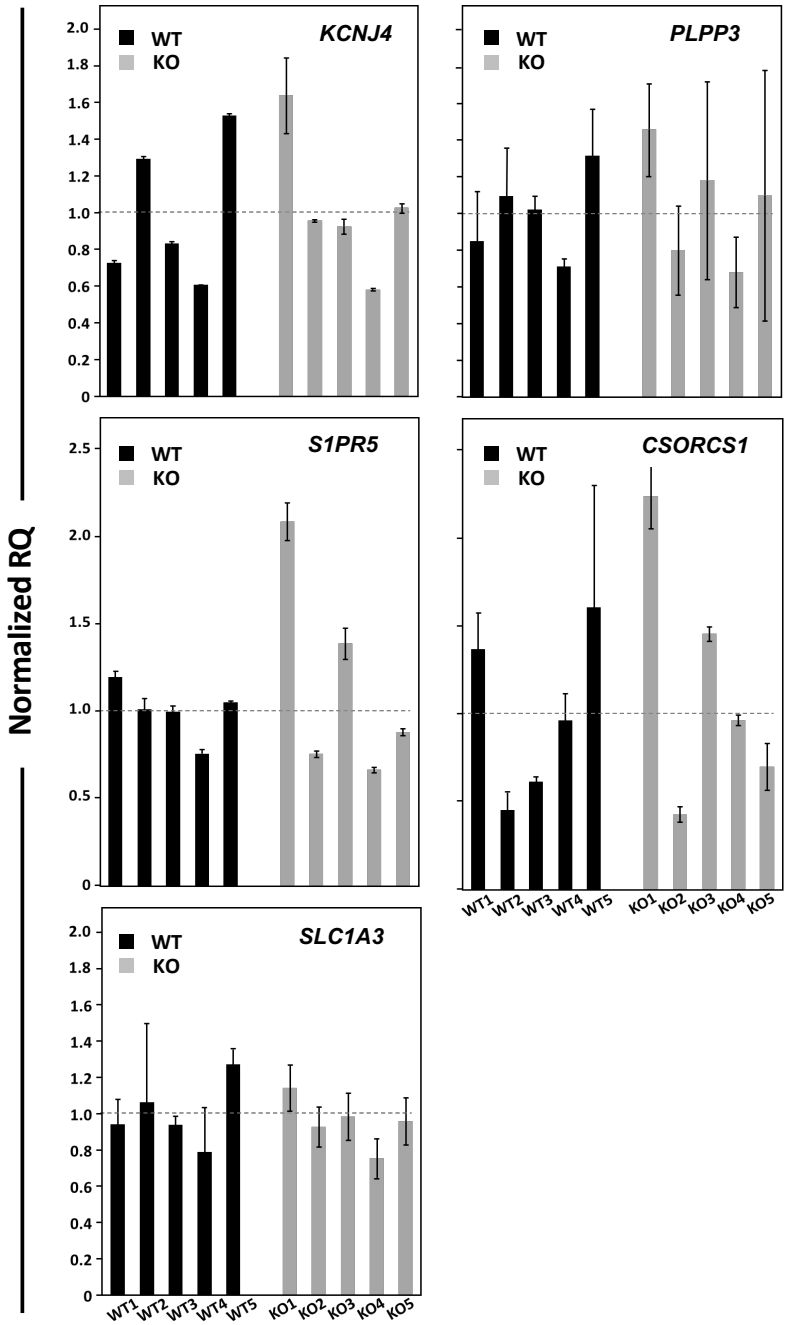


Figure 27 Differences in relative gene expression for single genes measured by qRT-PCR

The difference in relative gene expression between five KMT2E-WT and -KO HEK293 cell lines is shown for eleven different genes. After the primer efficiency and specificity of the qRT-PCR had been successfully tested, the number or cycles needed to overcome background fluorescence was measured. The result was compared to selected endogenous controls which were evenly expressed in all cell lines, including WT and KO. This method was repeated three times for each cell line. The average relative gene expression across all ten cell lines was calculated and the normalized RQ set to 1. The mean expression and corresponding standard deviation of each cell line are represented relative to this value.

Before the difference in relative gene expression between the *KMT2E*-WT and -KO HEK293 cells were tested with a two-sided t-test, the measured values from the endogenous controls and target genes were checked for outliers and homogeneity of variances **(Figure 28)**. One outlier was detected among the endogenous controls and treated as a missing value.

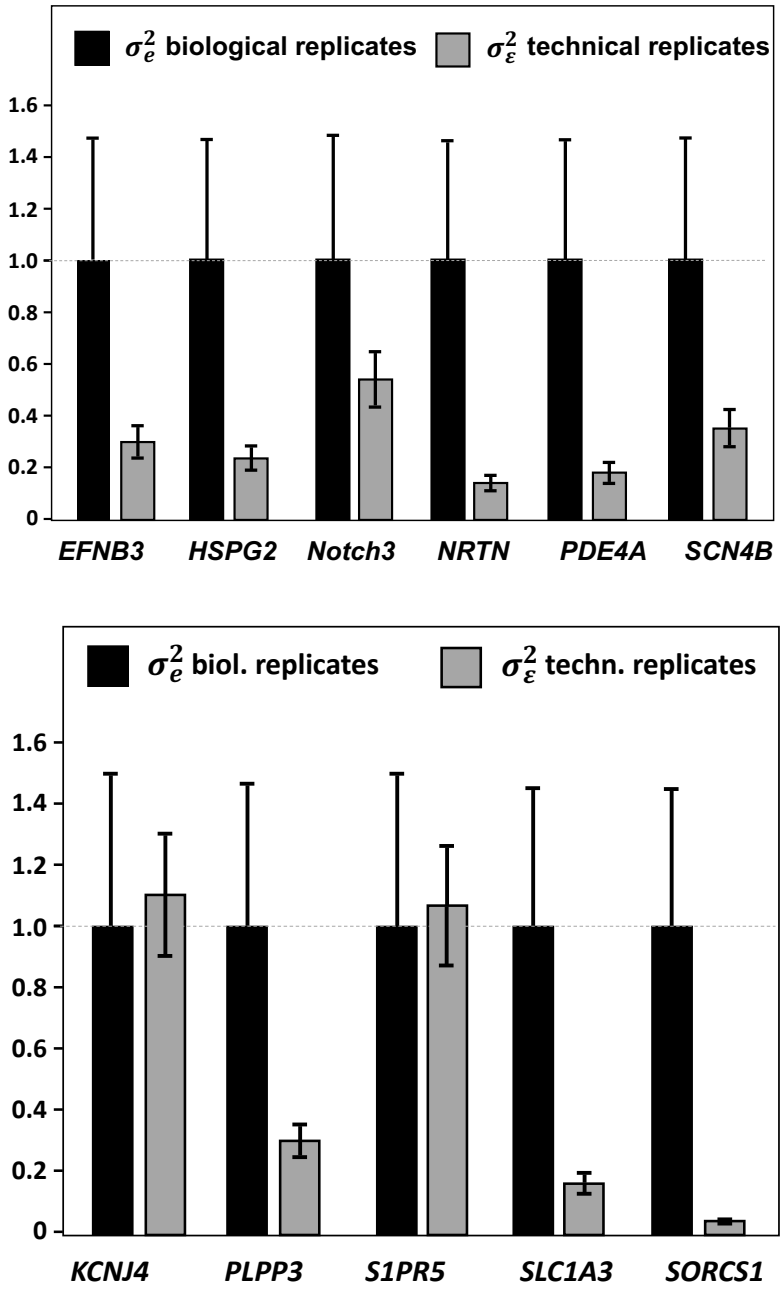


Figure 28 Estimates of the variance (± standard error) of technical replicates (set = 1.0) relative to the variance of biological replicates and standard error of each estimated variance for the 11 target genes

The two-sided t-test for testing the difference in relative gene expression between the *KMT2E*-WT and *KMT2E*-KO HEK293 cell lines was significant ($P < 0.05$) for one out of the 11 genes compared, in which case the KO-cell lines showed an increased expression for the target gene *PDE4A* ($P = 0.043$). The other ten differently expressed target genes from the RNA-Seq could not be verified by qRT-PCR **(Figure 29)**.

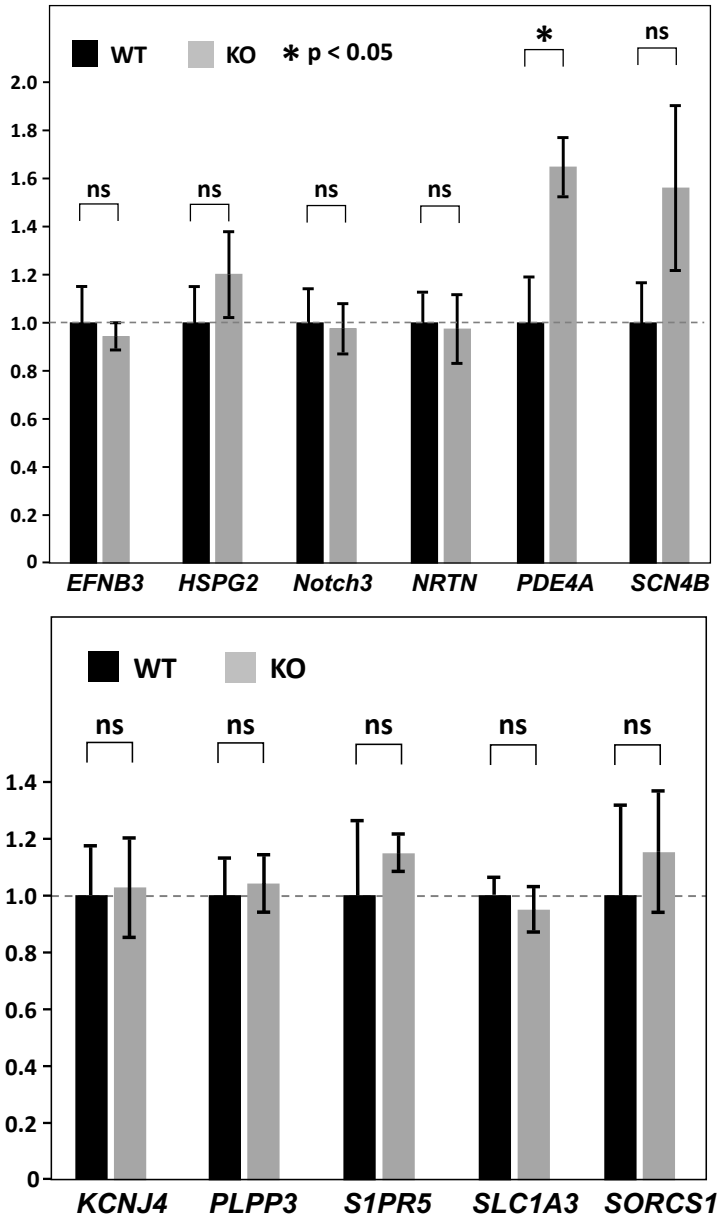


Figure 29 Relative gene expression between *KMT2E*-WT (set=1.0) and *KMT2E*-KO cell lines together with the corresponding standard errors
The result of the two-sided t-test with eight degrees of freedom is shown above the two columns. PDEA4 was significantly ($P < 0.05$) higher expressed in the KO cell lines. For all other target genes, the difference in gene expression found by RNA-Seq could not be validated by qRT-PCR.

3.5 Western blotting indicated an increased histone methylation level of H3K4me3 in *KMT2E*-KO cell lines

The global methylation levels of H3K4me1, H3K4me2 and H3K4me3 were measured via western blotting in the three different KMT2E-WT and KMT2E-KO HEK29 cells lines which were also used for RNA-Seq **(Figure 30)**. The objective of this experiment was to investigate the effect of KMT2E on histone methylation and to elucidate the underlying molecular mechanisms and the functional relevance of KMT2E. A modified histone methylation would affect chromosome organization and most likely gene expression.

The ANOVA revealed no outliers for all three western blots. No significant difference in histone 3 mono- and dimethylation of Lysine 4 was detected by western blotting **(Figure 31)**. Histone 3 trimethylation showed a statistic trend (P = 0.091), where histone 3 trimethylation was increased in KMT2E-KO HEK293 cells in comparison to KMT2E-WT HEK293 cells.

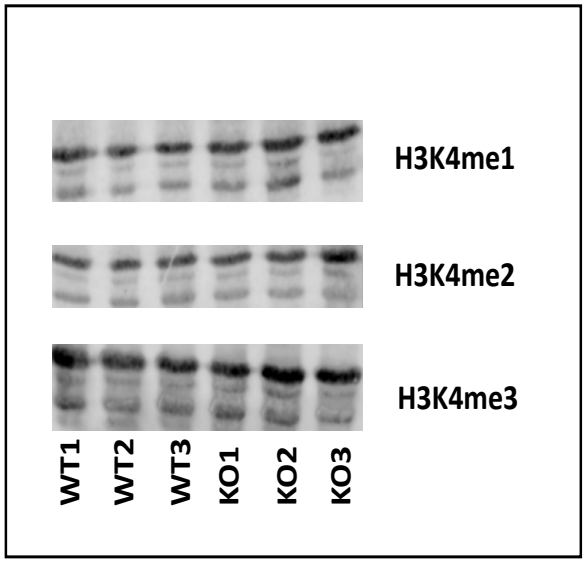


Figure 30 Example of a western blotting for all three conditions
The figure shows a western blotting for all three conditions (H3K4me1, H3K4me2, H3K4me3). H3 served as a loading control for all measurements.

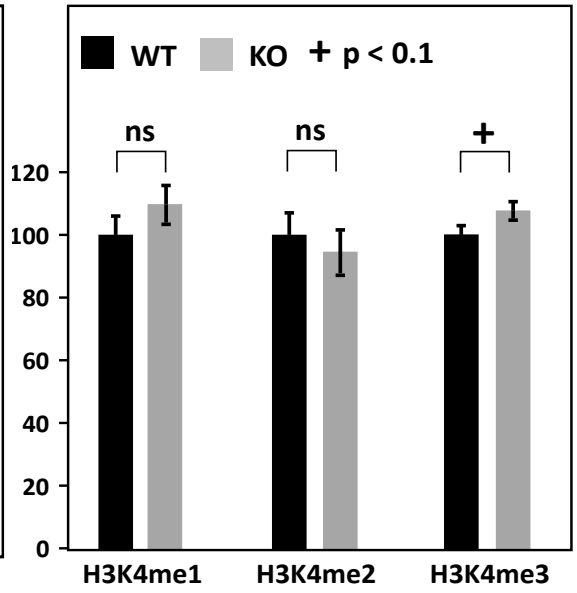


Figure 31 Comparison of histone methylation investigation in western blotting between KMT2E-WT and KMT2E-KO HEK293 cells for H3K4me1, H3K4me2, H3K4me3
Histone methylation of KMT2E-WT and KMT2E-KO HEK293 showed a statistic trend (P < 0.10) only for H3K4me3, but not for H3Kme1 or H3K4me2.

4. Discussion

The present research was part of a larger pilot study conducted in Dr. rer. nat. Rebecca Buchert's mental retardation working group at the University of Tübingen's Institute of Medical Genetics and Applied Genomics. The scheduled project investigated pathogenic variants in *KMT2E* and their effects on histone methylation and gene expression. With this objective in mind, three different *KMT2E* variants were introduced into a cell line and their effects on histone methylation and gene expression were analyzed. In the following, the design of the experiments is reviewed with possible changes for improvement. Afterwards, the results are discussed in light of other in literature reported studies. In particular, our results about *KMT2E* are compared with related proteins. Finally, suggestions for further investigations and experiments are made.

4.1 Experimental design: methodological aspects and questions

4.1.1 Choice of patient variants

In addition to the variant of our Tübingen ID patient, two other KMT2E variants of ASD patients from published studies were used (Dong et al. 2014; Iossifov et al. 2012). This allowed to broaden the inference basis beyond our patient's variant and to arrive at more general conclusions about truncating mutations in KMT2E and the relevance of the protein. Although ASD and ID are heterogeneous diseases, both neurodevelopmental disorders have overlapping phenotypes and often similar pathogenic biochemical and molecular mechanisms (Srivastava and Schwartz 2014). They often coexist in the same patient and the same mutations can cause either cause ID or ASD, although the reason for this observation is not yet fully understood (Laumonnier et al. 2004). This allowed the variants of the ASD patients to be included into the functional studies, although our primary subject of investigation was ID.

4.1.2 Cell module: HEK293 cells

Sequence modifications were introduced into the same exons of HEK293 cells, a secondary cultured human cell line, in which the three patients with KMT2E variants had modifications. Cultured human cell lines themselves are suitable for functional studies on molecular and biochemical issues such as pathway analysis

or histone methylation (Cooper 2000). They are easier, faster and cheaper to breed and maintain than complex model organisms such as drosophila, zebra fish or mice.

For the purpose of answering specific questions in cell biology, highly differentiated cell types with special characteristics are most suitable (e.g. nerve cells for questions on ion transporters or cytoskeleton transportation). The HEK293 cells used in our study are kidney epithelial cells and resemble fibroblasts but have no tissue-specific gene expression (Stepanenko and Dmitrenko 2015). They have similar gene expression profiles such as renal progenitor, adrenal glandular or neuronal cells, which complicates the assignment of HEK293 cells to a specific tissue.

A more suitable model organism for neuronal development could have been Lund human mesencephalic (LUHMES) cells (Edwards and Bloom 2019). This is an embryonic neuronal precursor cell line that can be maintained as proliferating cells or as postmitotic neurons after a treatment with tetracycline, dibutryl cAMP and glial cell-derived neutrophic factor.

Induced pluripotent stem (iPS) cells could have been another alternative to investigate KMT2E variants on a cellular level (Bayart and Cohen-Haguenauer 2013). This would have avoided working with tumor-derived cell lines and thus resemble the characteristics of neuronal cells *in vivo*. In this method, pluripotent stem cells are produced from human somatic cells (e.g. fibroblasts) which later on can be differentiated into other cell types such as neurons. In addition, iPS offers the opportunity to use fibroblasts from an affected patient and thus to generate models with exactly the same genetic basis as the patient's (van Diepen et al. 2018). However, IPS is a challenging laboratory method which requires ample experience with this method in order to apply it successfully. In addition, we had no consent from our patient´s parents to perform a biopsy to acquire her fibroblasts.

A further challenge was the unstable, polyploid karyotype of HEK293 cells (Stepanenko and Dmitrenko 2015). The average chromosome number and aberrations vary among different HEK293 cells from different cell banks.

However, since the aim was to mimic the heterogenous variants of the three-diploid patients, this was not possible in this particular setting. Instead, a diploid cell line such as RPE1 would have been an alternative (Kuznetsova, Kurinov, and Aleksandrova 2014).

4.1.3 CRISPR/Cas9

CRISPR/Cas9 is a fairly new method (Deveau, Garneau, and Moineau 2010) which has been further optimized in recent years for various applications. Sequence modifications were introduced into the same exons of HEK293 cells, in which the three patients with KMT2E variants had modifications, using CRISPR/Cas9. In addition, attempts were made to recreate the exact variants with ssODNs which contained the patient variants and could have served as a template for HDR. However, this objective was not met. In the following, some suggestions are listed to further improve its application in future experiments.

The first issue was the low efficiency with which CRISPR/Cas9 introduced genetic modifications inside all three targets of KMT2E **(Table 5)**. This required millions of HEK293 cells to be treated with the CRISPR/Cas9 constructs. Consequently, additional time, material and costs had to be invested in expanding the single cells to single cell clones until they could be sequenced. Subsequent troubleshooting can increase the CRISPR/Cas9 efficiency and thereby, save time and money.

A strategy to increase the percentage of HEK293cells with genetic modifications is not to optimize the efficiency of the CRISPR/Cas9 system itself, but successful transfection. The latter depends first of all on the choice of the cell line. HEK293 cells are an immortalized cell line and are easy to transfect and divide frequently (Thomas and Smart 2005). This gives CRISPR/Cas9 multiple options to cleave during mitosis, making HEK293 cells an optimal choice.

Another strategy to improve CRISPR/Cas9 efficiency is the correct selection of successfully transfected cells, which was performed with puromycin. First of all, the used cell line must be sensitive to the chosen selection marker (i.e. puromycin). It has been reported in other studies that a small number of HEK293 cells were resistant to even higher concentrations of puromycin (Iwaki and

Umemura 2011). Hence, it is prudent to use further selection markers within the CRISPR/Cas9 vector. In a selection assay, zeocyin and hygromycin g proved to be better selection markers for HEK293 cells than puromycin (Lanza, Kim, and Alper 2013). Another possibility to increase the selection of successfully selected cells would be to separate the cells using fluorescence-activated cell sorting (FACS) instead of a serial solution. In addition, one could use a GFP-bearing CRISPR/Cas9 vector (e.g. PX458) and select successfully transfected cells using FACS (Ran et al. 2013).

One other important reason for low CRISPR/Cas9 efficiency could have been the low rate of inserted double-strand breaks by the CRISPR/Cas9 system, although the HEK293 cells had been successfully transfected. Other working groups experienced the same problem even more seriously in transgenic animals (Jang et al. 2018). A strategy to improve the efficiency (up to 38%) of the CRISPR/Cas9 system itself has been reported by using overlapping sgRNAs (Jang et al. 2018). The best outcome was achieved by taking multiple sgRNA (at least two to three) for the same target region. If the sgRNA has at least 5 common bp in the target sequence, this improves the rate of HDR without influencing the rate of double-strand breaks. This could have been a possible solution to improve the CRISPR/Cas9 efficiency of KMT2E.

The second issue was the failure to introduce the exact patient variants into KMT2E. This most likely did not influence downstream experiments because both, the desired variant and the change in sequence through NHEJ resulted in NMD (Kervestin and Jacobson 2012). NMD is a control mechanism detecting unwanted premature stop codons and preventing the expression of undesired proteins is prevented. If the exact patient variants were desired to be incorporated into the HEK293 cells, it would have been necessary to increase the rate of successful HDR (Liu et al. 2019). There are several possible strategies for achieving this aim. For example, HDR is preferred to NHEJ by inhibiting key proteins (e.g. DNA ligase IV) with chemicals such as SCR7 or enhancing HDR with chemicals such as RS-1 (Maruyama et al. 2015; Chu et al. 2015). RS-1 stabilizes the association of DNA and Rad51, a main actor of homologous

recombination in HDR. In addition, timing of the CRISPR/Cas9 system delivery is crucial because HDR is restricted to the S and G2 phases, which requires synchronization of the cell cycle (Lin, Staahl, et al. 2014).

Besides increasing the rate of HDR itself, the most important step for HDR is optimizing the ssODNs as HDR templates (Richardson et al. 2016). It was demonstrated that the nuclease Cas9 binds to the non-target strand, whereas the complexed sgRNA binds to the target strand (Richardson et al. 2016). This is relevant because Cas9 dissociates asymmetrically from the double-stranded DNA. At the beginning, the nuclease sets the 3´-strand free, which is farthest from the PAM sequence and non-complementary to the sgRNA. If the ssODN are now complementary to the first released 3´-strand (i.e. corresponding to the target sequence), this increases the HDR success rate. It was also demonstrated that such a designed ssODN does not compete for Cas9 binding with the complementary strand. Successful ssODN binding to DNA depends on the ratio of ssODN relative to the CRISPR/Cas9 and sgRNA system, which could be tested for each KMT2E sgRNA construct with a concentration curve (Liu et al. 2019).

Another strategy for successful HDR is optimization of the ssODN design (Richardson et al. 2016). These authors found ssODNs with 36 bp overlapping on the distal side of the PAM sequence and 91 bp overlapping on the proximal-side of the PAM sequence, counted from the site of the double strand break, to reach the highest HDR frequency of around 57% (Richardson et al. 2016). For future experiments, we recommend this strategy for constructing the ssODNs with the patient variants.

An alternative strategy is the use of different Cas9 nucleases. While catalytic inactive nuclease does not improve the rate of HDR, it makes CRISPR/Cas9 more reliable, because it does not edit the target sequence but displaces DNA and enables ssODN to anneal to its complementary sequence, also resulting in successful HDR. Catalytic inactive Cas9 has mutations in the HNH and RuvC domain. In return, Cas9 nickase has only one mutation in one of both domains and cleaves one strand of the target sequence, depending on which domain is

mutated (Richardson et al. (2016). Using normal or catalytic inactive Cas9 nuclease with the N-terminal fused part of CtIP could have been a further option. The fused CtIP were directed to the site of the double-strand break and induced HDR, because it is a key protein for this mechanism (Charpentier et al. 2018).

The third issue was the unsuccessful heterozygous incorporation of the KMT2E patient variants, a common problem encountered in modeling heterozygous human disorders (EauClaire and Webb 2019). This issue has been investigated by several researchers who found various strategies to overcome this challenge. One strategy for inserting heterozygous mutations with CRISPR/Cas9 is using mixed ssODNs of equimolar concentration, where half of the templates have the desired mutation and half have the WT sequence.

These three strategies could be investigated for the purpose of generating heterozygous KMT2E variants in future experiments.

4.1.4 Quality control of RNA-Seq, qRT-PCR and Western blotting

In order to optimize quality control in the RNA-Seq, the technical variation was minimized by preparing all samples simultaneously and diluting them to the same concentration. Both experimental groups were later run on the same sequencing lane. Separately prepared biological replicates were chosen for each experimental group instead of pooled biological replicates because the biological variance was expected to be low in this set-up (vs. post-mortem or clinical samples). We chose two-paired end sequencing because this yields a greater depth and includes the length of the product in the alignment, which increases coverage and improves quality. Thus, it provides more information about transcript variants than the single strand sequencing mode.

For the qRT-PCR, many intervening steps were followed according to the international MIQE guidelines to exclude mistakes and to check for interim results to ensure the quality of the received data (Bustin et al. 2009) because qRT-PCR is a very sensitive method but prone to random and systematic error (Taylor et al. 2019).

The five main categories of errors in Western blotting were checked in troubleshooting to ensure highest quality of the obtained data. Common sources of errors include unusual bands due to protease degradation, high voltage or air bubbles present during transfer to the gel or an excessively low resistance of the gel (Mahmood and Yang 2012)

4.2 Comparison of results with literature on *KMT2E*

4.2.1 RNA-Seq

A differential gene expression analysis of the RNA-Seq data between the *KMT2E* WT and KO cells revealed 30 genes which were differentially expressed by a 0.3-fold change at the $P \leq 0.05$ significance level. Previous analyses of genomic-wide KMT2E occupancy and KMT2E knockdown with shRNA in myoblast cells showed that KMT2E plays a role in multiple cell processes such as cell cycle, differentiation and gene expression (Sebastian et al. 2009; Ali et al. 2013). Therefore, it was not surprising that the *KMT2E* knockout resulted in a modified transcriptome.

However, the mechanisms behind this observation are still elusive but previous studies reported the following observations. A DamID study showed that KMT2E binds to 15,000 genomic regions and 13,000 genes in myoblasts, especially to transcriptional starting sites and CpG-islands which, for the most part, match with promotors (Ali et al. 2013). The study also compared the KMT2E-binding sites with RNA-polymerase-bound promotor regions and found a high consistency of 66% between RNA-polymerase-bound and KMT2E-bound locations. RNA-polymerase-bound promotors are a sign of actively transcribed genes. Another indicator of KMT2E´s role in gene expression is the fact that KMT2E binds to promotors which contain marks of actively transcribed regions, including H3K4me3, H3K4me2 and H3K9ac but not to promotors which contain marks of actively repressed regions such as H3K9me3. From these studies, it can be concluded that KMT2E only binds to the promotors and transcription starting sites of actively transcribed genes and probably plays an important role in gene expression. Hence, these studies are in accordance with our finding of different gene expression between *KMT2E* KO and WT HEK293 cells.

The mechanism how KMT2E binds to histones in actively transcribed regions is already understood (Ali et al. 2013). KMT2E is recruited to euchromatic regions downstream of transcriptional starting sites of active genes via a rare binding mechanism of Tryptophan W141 and Aspartate D128 of the PHD finger to H3K4me3 despite a lack of catalytic activity. In general, PHD fingers act as an effector in chromatin-remodeling proteins by recruiting transcription factors and nucleosome-associated complexes to chromatin through binding histone peptides (Musselman and Kutateladze 2009). KMT2E is released again through the phosphorylation of H3T3 and H3T6, a mechanism observed in mitotic cells which indicate that the release of KMT2E from active genes plays an important role throughout mitosis and development. The importance of this mechanism is supported by the fact that this mechanism is conserved in the *KMT2E* drosophila ortholog *UpSET* (Rincon-Arano et al. 2012).

The conservation of this mechanism could suggest that KMT2E plays a similar transcriptional role as its orthologue UpSET. Homologous genes have been identified in transcription-associated deacetylase Rdp/Sin3repressor complexes (Rincon-Arano et al. 2012; Pijnappel et al. 2001). It has been demonstrated that they keep a tight chromatin configuration around actively transcribed genes by decreasing H3K9 and H4K16 acetylation as well as H3K4me1 and H3K4me2 methylation. Thus, they suppress the gene activation of silent neighboring genes around their target genes. For example, UpSET interacts with histone deacetylases to regulate open chromatin and adjust gene expression (Pijnappel et al. 2001). The reduction or loss of UpSET alters the chromatin architecture of promotors, including the broadening of H3K4me3, and leads to an upregulation of off-target genes (Shen et al. 2014; Rincon-Arano et al. 2012). KMT2E was already shown to be part of the histone deacetylase complex X-CoR which binds to active genes and ensures proper gene activation (Kittler et al. 2007). Hence, the differences in gene expression between the KMT2E KO and WT HEK293 cell lines could be explained by KMT2E´s role in the histone deacetylase complex and represent such off-target genes.

KMT2E was also found to inhibit the demethylase LSD1 and monomethyltransferase SET7/9 (Sebastian et al. 2009). By inhibiting LSD1, the amount of H3K4me2 increases, which indicates active transcription sites. By inhibiting SET7/9, the amount of H3K4me1 decreases, which indicates repressed transcription sites. Through the control of these histone modifying genes, KMT2E could affect H3K4 methylation indirectly (Sebastian et al. 2009). This observation could suggest that a KMT2E knockout lowers gene expression in general due to the lack of active histone markers and abundance of inactive histone markers. This trend cannot be recognized from our RNA-Seq data because the KMT2E knockout is likely to influence other regulators of gene expression.

***PDE4A*: an ASD candidate gene**

Phosphodiesterase 4A (*PDE4A*) was found to be significantly higher expressed in KO cells than in WT cells of KMT2E HEK293 and was validated in the qRT-PCR (P= 0.05) **(Table 6, Figure 29)**. This gene is particularly interesting: *PDE4A* is a phosphodiesterase and belongs to the PDE family. It controls the levels of cAMP by hydrolytic degradation, a key cellular second messenger which controls a variety of cell functions such as phagocytosis (Serezani et al. 2008). *PDE4A* has already been described in synaptic plasticity, memory and cognition processes (Ye, Jackson, and O'Donnell 2000). According to the Human Protein Atlas, PDE4A is most highly expressed in the brain and second highly expressed in the testis.

PDE4A has already been connected to neurological diseases. In an exome-sequencing study, one patient had a *de novo PDE4A* variant, which was associated with ASD according to Autism KB, but no other possible disease-causing mutations (Griesi-Oliveira et al. 2015). Another study measured the levels of *PDE4A* in three brain regions of ASD patients against a healthy control group (Braun et al. 2007). They found different expression patterns between both groups with an increased expression in Brodmann's area BA9 (superior frontal cortex), an unchanged expression pattern in Brodmann´s area B40 (parietal cortex), and a reduced expression pattern in the cerebellum of ASD patients against a heathy control group. The difference in expression could be due to

different functions in different regions. The different expression patterns are an interesting observation because *PDE4A* inhibitors such as rolipram, sertraline or fluoxetine already exist and could offer a therapeutic option for ASD which is an untreatable disorder so far. The protein family PDE4 has already been named as a therapeutically target in different neurodevelopmental diseases such as Alzheimer's or Parkinson's and psychiatric diseases such as schizophrenia and depression (Millar et al. 2005; Houslay, Schafer, and Zhang 2005).

The validated different expression of *PDE4A* supports the hypothesis that KMT2E variants disturb the transcriptional networks of neuronal development. The results from clustering of gene expression profiles and pathway analysis with IPA are consistent with previous studies which described the role of KMT2E in spermatogenesis (Yap et al. 2011).

4.2.2 Western blotting

The study of histone methylation levels in western blot showed the tendency of an increase in H3K4me3 methylation levels in the KMT2E-KO HEK293 cells ($P \leq 0.10$ significance level). This result is in contradiction to a KMT2E knockdown experiment in quiescent myoblasts from another research group (Sebastian et al. 2009). They found that H3K4me3 and H3K4me2 levels were reduced, but H3K4me1 levels were increased and H3K9 levels remained unchanged in a KMT2E knockdown with sgRNA in myoblasts. From their results, they concluded that KMT2E is specific to H3K4.

Considering these opposing results, it has to be kept in mind that functional studies *in vivo* are challenging because the dysregulation of one type of post-translational histone modification causes secondary changes in other types of post-translational histone modification (Shen et al. 2014). Normally, SET domain containing enzymes can methylate histones and the SET domain is a regular element in chromatin regulators (Del Rizzo and Trievel 2011). Likewise, the SET domain of KMT2A-D and KMT2F-G catalyzes the methylation of histone H3 on lysine 4 (H3K4). However, the SET domain of KMT2E lacks methyltransferase (HMT) activity (Madan et al. 2009). The KMT2E homologous genes murine Mll5, set3 of yeast and UpSET of drosophila, with which KMT2E shares long

homologous sequences, also lack HMT activity (Glaser et al. 2006; Eissenberg and Shilatifard 2010; Madan et al. 2009). Instead of the KMT2 protein family with methyltransferase activity, it was suggested that KMT2E forms a protein family together with the mammalian protein SET containing-domain 5 (SETD5) (Mas et al. 2016). SETD5 interacts with the co-transcriptional complex PAF1 and is part of the histone deacetylase complex NCoR, suggesting that SETD5 is essential for the regulation of histone acetylation during gene transcription (Osipovich et al. 2016). Likewise, it was speculated that KMT2E also has also a regulatory function, because it was detected in the Rpd/Sin3 repressor complex which regulates promotor-associated histone acetylation and H3K4 di- and trimethylation (Shen et al. 2014).

This indicates that KMT2E does not have a direct HMT activity but a regulative function and thus, depending on the state of the cell, controls an up- or downregulation of H3K4 methylation. Consequently, the different histone methylation level results between both experiments do not necessarily contradict each other but may show different regulatory states of KMT2E in different tissues.

Another interesting observation was made in ASD patients in comparison to a healthy control. ASD patients had a spread of histone marks away from TSS into gene bodies, and upstream sequences were observed in prefrontal neurons (Shulha et al. 2012). The same spread of histone marks away from TSS was observed in lost or decreased UpSET drosophila cells. Hence, this comparison indicates that KMT2E could influence H3K4 methylation broadening in ASD patients with *KMT2E* variants.

4.3 Do mutations in *KMT2E* cause ID?

4.3.1 Results from our experiments

Mutations in genes, especially those involved in epigenetic regulation and chromatin remodeling (e.g. *CHD8, MECP2, AUTS2, TRAPPC9*), have been identified in a wide variety of neurodevelopmental disorders (Shen et al. 2014). This study investigated the role of *KMT2E* in neuronal development and ID. *KMT2E* was previously a relatively poorly studied gene of the KMT2 protein family, members of which have been described in various ID syndromes (Zhang et al. 2017; Kim et al. 2017). We were successful in establishing *KMT2E* KO cell lines that could serve as a valuable resource for further investigations on the molecular and cellular function of KMT2E. Transcriptome profiling showed that *KMT2E*-KO changes over 30 genes, which were especially enriched in several neurodevelopmental disease complexes. *KMT2E*-KO significantly increased the expression of the autism candidate gene *PDE4A* which had already been associated with changed expression patterns in ASD and a *de novo* variant was described in an ASD patient. This supports the hypothesis that KMT2E plays a significant role in transcriptional networks of special importance for neuronal development. The finding of PDE4A could be of great interest for *KMT2E* ID patients because PDE4A inhibitors already exist, which are used in the treatment of psychiatric disorders, and could offer a therapeutic option for ID patients. However, our RNA-Seq data needs to be validated in further experiments before clinical tests are designed.

4.3.2 Affirmation from a contributed paper

Our functional results were affirmed by a genetic study to which we contributed with genetic and clinical data from our patient (patient 2) and that was published during the development of this thesis (O'Donnell-Luria et al. 2019). The study established former candidate gene *KMT2E* as an ID gene and found heterozygous variants in *KMT2E* as a cause of a spectrum of neurodevelopmental disorders and epilepsy. We reported on 38 patients in 36 families, including three patients already mentioned in literature (Dong et al. 2014; Iossifov et al. 2012; Uliana et al. 2010). 4 patients with microdeletions enclosing *KMT2E* and 34 patients with 31 different heterozygous variants in

KMT2E which occurred *de novo* in the majority of cases were identified. The detected *KMT2E* variants encompassed different protein-truncating variants, including indels, nonsense, essential splice site and protein extension mutations. Almost all variants were predicted to result in NMD. In addition, missense mutations were identified in 4 patients, 3 of which were found in highly conserved positions of the gene. Structure prediction tools anticipated altered binding capacities for missense mutations through changing rotamers, electrical charge and hydrophobicity which would explain the ID phenotype of the patients although *KMT2E* missense mutations also occur in the general population.

The extent of the presenting disorder seemed to depend on the underlying genetic modification. Developmental delay in patients with microdeletions was greater than in patients with protein-truncating mutation but patients with missense mutations showed the most severe cases of developmental delay. This could indicate a gain-of-function or dominant-negative effect of missense mutations whereas protein-truncating mutations and microdeletions could result in haploinsuffiency. Further, 70% of identified patients were male and expressivity depended on sex. Clinically, most patients displayed mild ID (mean IQ 74 for tested patients), one quarter ASD and one fifth epilepsy. Other common characteristics included early developmental delay (delayed independent walking and acquired first word), behavioral problems (e.g. attention-deficit/hyperactivity disorder, stereotypies, skin-picking behavior, self-injurious behavior, aggressivity, anxiety), hypotonia, functional gastrointestinal abnormalities (e.g. reflux, vomiting, bowel motility issues), macrocephaly and unique facial features (e.g. dolichocephaly, large forehead, deep-set eyes, periorbital fullness, prominent cheeks, prominent nasolabial folds).

Moreover, a large database with over 140,000 individuals for *KMT2E* variants was examined to estimate the frequency of *KMT2E* variants in the general population. *KMT2E* turned out to be a candidate haploinsufficiency gene with a probability of loss-of-function intolerance score of 1.0, rarely showing any loss-of-function *KMT2E* variants in the cohort. Most *KMT2E* variants, found in the dataset, were allocated to annotation artifacts, sequence errors, somatic

mosaicism and splice-site rescue, not expected to result in protein-truncating variants like the majority of *KMT2E* variants in the patient cohort.

We concluded that *KMT2E* is an underrated cause of neurodevelopmental disorders with haploinsufficiency as the disease mechanism. Hence, the cooperative effort to establish KMT2E as ID-gene supports the approach of this study. . By establishing *KMT2E* as an ID gene, we showed *KMT2E´*s relevance for the human body through the number of patients with corresponding genotype and phenotype and the depletion of *KMT2E* protein-truncating variants in the general population. These findings supported the relevance of our results from functional studies in *KMT2E* KO HEK293 cell lines and was upheld by a mouse model of *KMT2E*-depleted mice which showed an increased postnatal lethality and surviving animals showed retarded growth (Zhang et al. 2009). Further clinical and functional studies are necessary to fully understand the pathogenesis of *KMT2E* variants in affected patients and to offer therapeutic approaches.

4.4 Open questions and possible experiments to answer them

Building on previous studies, we were able to extend our knowledge about the function of *KMT2E* by our investigations. To further characterize *KMT2E* and validate our RNA-Seq results, a promising avenue would be to convert a *KMT2E* ID patient's fibroblast into neurons via IPC and perform a subsequent RNA-Seq.

In addition, the recruitment of KMT2E to promotors (e.g. *PDE4A*) and its effect on gene expression through selective binding of H3K4me3 could be further investigated with a luciferase assay (Ali et al. 2013) which enables gene expression studies at the transcriptional level. The promotor of a gene of interest (e.g*. PDE4A*) could be cloned into a plasmid, upstream of the firefly luciferase gene. A second ranilla plasmid would be co-transfected, which can be used for normalization because it is permanently expressed. At the same time, KMT2E could be incorporated into a pEGFPN2-containing plasmid for KMT2E-WT-pEGFPN2 and an additionally mutated KMT2E-pEGFPN2 gene generated via site-directed mutagenesis. If KMT2E affects gene expression by binding to the promotor, luciferase activity should be high in the WT KMT2E-pEGFPN2 sample

and no luciferase activity should be observed in the mutated KMT2E-pEGFPN2 sample.

An alternative experiment could be the construction of N-terminal GFP containing KMT2E because there is no KMT2E binding antibody available. The pEGFPN2 KMT2E containing plasmid could be used as a template for HDR when introducing DSB via CRISPR/Cas9 at the N-terminal end of KMT2E in HEK293 cells. A KMT2E-GFP Chip with an GFP-antibody could be carried out within HEK293cells to make KMT2E DNA occupancy visible. The obtained results could then be compared with the RNA-Seq data, with particular attention to the promotors of the expression-modified genes with modified expression.

5. Outlook

The primary aim of this project, to study KMT2E as a histone methyltransferase and epigenetic regulator in the framework of ID, will be pursued in further projects. Dr. rer. nat. Rebecca Buchert and her working group have started to investigate pathogenic variants in histone methylation factors such as JARID2 and HMG20A. This will broaden our understanding of H3K4 changes in neurodevelopmental disorders and help to create common maps of methylation changes in these disease complexes.

Basic research on ID genes themselves as well as on commonly shared pathways and networks among these genes might offer promising therapeutic approaches (Berdasco and Esteller 2013). These therapeutic treatments try to enhance intelligence levels by improving cognitive and behavioral skills. This is a long-awaited message for ID patients, their family members and health care providers because neurodevelopmental disorders such as ID or ASD have long been considered irreversible. In particular, the identification of commonly shared pathways and networks is a key for treatment because ID is too heterogeneous for targeting individual genes (van Bokhoven 2011).

One of these common targets of different ID genes is H3K4 methylation which undergoes significant remodeling during childhood. Interestingly, different studies

showed that H3K4 methylation can also be influenced by environmental influences such as maternal immune activation, parental care or monoaminergic drugs (Shen et al. 2014). This indicates that H3K4 methylation can be modified by internal and external factors. If disturbed H3K4 methylation is present because of defective genes, this could be corrected by external treatment. Drugs affecting histone methylation are already known such as the psychiatric inhibitory drugs clozapine, valproate or phenazine (Peter and Akbarian 2011). They could be used to regain the fragile balance, which is normally maintained by opposing enzyme families, but commonly affected in ID patients (Peter and Akbarian 2011).

N-acetylcysteine (NAC) could represent a *KMT2E*-specific therapy (O'Donnell-Luria et al. 2019). *KMT2E*-depleted mice showed increased levels of reactive oxygen species (ROS) which could be reversed by NAC (Tasdogan et al. 2016). NAC is already therapeutically used for other indications and could therefore be applied in clinical studies. In order to test this possibility, ROS levels in KMT2E haploinsufficient patients would have to be measured and the therapeutic use of NAC in clinical trials evaluated.

Another important tool to test possible therapeutics are animal models. A successful therapeutic intervention could already be carried out in mouse models: the histone deacetylase inhibitor AR-42 was able to rescue structural and functional brain deficits in heterozygous KMT2D mice, a model for Kabuki syndrome (Bjornsson et al. 2014). The authors suspected that other genetic diseases, influencing histone modifications, might also be therapeutically treatable with histone deacetalyse (HDAC) inhibitors. However, the precise role of H3K4 methylation and its modifying enzymes is still too elusive for targeted treatment in humans and needs further investigation (Parkel, Lopez-Atalaya, and Barco 2013).

Alternatively, defective cells could be replaced by stem cells (Derecki et al. 2012). This newly evolving technique has already proven successful in Mecp2 knockout mice, a model organism for the Rett syndrome.

Another option could be not to target core outcomes such as cognition but to achieve behavioral improvements by reducing irritability (Verpelli et al. 2014). For example, a double-blind randomized trial found that the atypical neuroleptic risperidone improved behavioral problems more than placebo did.

These therapeutic approaches are promising for the large number of affected patients and their relatives, for whom only symptomatic treatments are available in most cases. They could improve their quality of life including reducing symptoms such as intellectual impairment (van Bokhoven 2011).

6. Summary

This thesis investigated *KMT2E* as a candidate gene for intellectual disability (ID) with different functional studies. The starting point for this project was a female ID patient with a *de novo* variant in *KMT2E* (NM_182931.2: c.280delA, p.Thr94Leufs*25). A literature review provided no hints for the connection of *KMT2E* with ID but reported *KMT2E* variants in two ASD patients.

KMT2E, a PHD and SET domain containing protein, belongs to the KMT2 family of histone methyltransferases and has been linked to hematologic cancer and coronary artery disease. However, *KMT2E* had not been associated with neuronal development or neuronal diseases although most KMT2 family members had been identified in ID.

In order to unravel the role of KMT2E in ID, we inserted double-strand breaks with CRISPR/Cas9 into HEK293 cells at the positions of the *KMT2E* variants from the ID and two ASD patients. Afterwards, the role of *KMT2E* variants in gene expression profiles was studied with RNA-Seq. A comparison between three *KMT2E*-WT and *KMT2E*-KO cell lines revealed 30 differently expressed genes. For validation of the RNA-Seq results, eleven of these genes were tested by qRT-PCR. The gene *PDEA4* showed a significantly ($P < 0.05$) higher expression level in the *KMT2E*-KO-cell lines than in the WT-cell lines. *PDEA4*, a member of the cyclic nucleotide phosphodiesterase family, acts as a cAMP regulator in many basic cell pathways and modified expression has already been found in ASD patients. Furthermore, the impact of *KMT2E* variants on histone methylation was examined with Western Blots for H3K4me1, H3K4me2 and H3K4me3. In the course of our experiments, a publication was prepared with 38 probands, including our patient, who had ID or ASD and genetic modifications in *KMT2E*.

Our study identified *KMT2E* as a possible cause of ID but further research is warranted to fully understand the function of this protein. These findings promise new therapeutic approaches for affected patients and broaden our knowledge on ID.

7. Deutsche Zusammenfassung

Diese Arbeit untersuchte *KMT2E* als Kandidatengen für geistige Behinderung mit Hilfe verschiedener funktioneller Untersuchungen. Ausgangspunkt war eine geistig behinderte Patientin mit einer *de novo* Variante in *KMT2E* (NM_182931.2: c.280delA, p.Thr94Leufs*25). Eine Literaturrecherche ergab keinerlei Hinweise auf eine Verbindung zwischen *KMT2E* und geistiger Behinderung, allerdings wurden *KMT2E* Varianten bereits in zwei Patienten mit Autismus-Spektrum-Störungen beschrieben.

KMT2E, welches eine PHD und SET Domäne enthält, gehört zur KMT2 Familie der Histonmethyltransferasen und wurde mit Erkrankungen wie koronarer Herzkrankheit und hämatologischen Krebsarten assoziiert. Allerdings wurde *KMT2E* noch nicht mit neuronaler Entwicklung in Verbindung gebracht, obwohl Varianten der meisten KMT2-Proteinfamilienmitglieder in Patienten mit geistiger Behinderung identifiziert wurden.

Um die Rolle von *KMT2E* bei geistiger Behinderung zu erforschen, wurden Doppelstrangbrüche mit CRISPR/Cas9 in HEK293 Zellen jeweils an den Positionen der *KMT2E* Varianten der geistig behinderten Patientin und zweier Autismus-Spektrum Patienten inseriert. Die Rolle von *KMT2E* auf Genexpressionsprofile wurde mit Hilfe von RNA-Seq untersucht. In einem Vergleich von drei *KMT2E*-Wildtyp und *KMT2E*-Knockout HEK293 Zelllinien wurden 30 Gene mit veränderter Expression gefunden. Um die Ergebnisse der RNA-Sequenzierung zu validieren, wurden elf dieser Gene in einer qRT-PCR getestet. Das Gen *PDEA4* zeigte eine signifikant ($P < 0,05$) höhere Expression in den *KMT2E*-Knockout HEK293-Zelllinien als in den *KMT2E*-Wildtyp HEK293-Zelllinien. *PDEA4*, ein Mitglied der zyklischen Nukleotid-Phosphodiesterase-Familie, wirkt als cAMP Regulator in vielen verschiedenen grundlegenden Zellsignalwegen und wurde in der Literatur mit Autismus-Spektrum-Störung in Verbindung gebracht. Mittels Western Blot wurde der Einfluss von *KMT2E* Varianten auf die Histonmethylierung für H3K4me1, H3K4me2, H3K4me3 untersucht. Unsere Forschungsergebnisse fanden Eingang in eine Publikation

mit 38 Probanden inklusive unserer Patientin, welche eine geistige Behinderung oder eine Autismus-Spektrum-Störung und eine *KMT2E* Variante zeigten.

Die Ergebnisse unserer Studie identifizierten *KMT2E* als mögliche Ursache für geistige Behinderung. Weiterführende Forschungsarbeiten sind jedoch angezeigt, um die Funktion dieses Proteins vollständig aufzuklären. Diese Befunde versprechen neue therapeutische Ansätze für betroffene Patienten und erweitern unser Verständnis für geistige Behinderung.

8. Appendix

8.1 Primer

All primers were obtained from Metabion.

Sequencing primers

Product	Sequence (5´to 3´)
PX459-forward primer	gagggcctatttcccatgattcc
PX459-reverse primer	ccggtgtttcgtcctttccacaag
*KMT2E*_X3_F	ctagttttctcagtgccgtaggg
*KMT2E*_X3_R	tatgtctgacacataacaagggctc
*KMT2E*_X4_F	taatggtattaaggaaggaatgagc
*KMT2E*_X4_R	accaaaggttatcagcattcaag
*KMT2E*_X22_F	atcttcaaccatgttaacattggg
*KMT2E*_X22_R	gatacaagcccatcttcggaag

qRT-PCR primers

Product	Sequence (5´to 3´)
Endogenous controls	
*ACTB*_qPCR_f1	AAGAGCTACGAGCTGCCTGAC
*ACTB*_qPCR_r1	AAGGTAGTTTCGTGGATGCCAC
*ACTB*_qPCR_f2	GATCACTGCCCTGGCACC
*ACTB*_qPCR_r2	GCCGATCCACACGGAGTAC

*C1ORF43*_qPCR_f1	CAAGTATGAGCCCCAGCTCC
*C1ORF43*_qPCR_r1	CTCAGAGGTACGAATGGCATC
*C1ORF43*_qPCR_f2	CTATCCGGGCTTTGACTCC
*C1ORF43*_qPCR_r2	ATTTCATTGCAAAGCGCATG
*GAPDH*_qPCR_f1	GGAGCGAGGCTAGCTGGC
*GAPDH*_qPCR_r1	GGCAACAATATCCACTTTACCAG
*GAPDH*_qPCR_f2	GATGGGTGGAGTCGCGTG
*GAPDH*_qPCR_r2	AAGCAGCCCTGGTGACCAG
*GUSB*_qPCR_f1	GACCTGCGCACAAGAGTGG
*GUSB*_qPCR_r1	CTAGCGTGTCGACCCCATTC
*GUSB*_qPCR_f2	TGGTTGGAGAGCTCATTTGG
*GUSB*_qPCR_r2	TTGGTTGTCTCTGCCGAGTG
*PDHB*_qPCR_f1	AGCCATTCGGGATAACAATCC
*PDHB*_qPCR_r1	TTGACTGAGCTTCCGGAGG
*PDHB*_qPCR_f2	TGAAATCTGTGCCAGGATCATG
*PDHB*_qPCR_r2	TTGCATAAGGCATAGGGACATC
*PSMB4*_qPCR_f1	CGTCTGCACTTTACAGAGGTCC
*PSMB4*_qPCR_r1	CTCGAACTTAACGCCGAGG
*PSMB4*_qPCR_f2	AGTACCATGCTGGGTGCCTC
*PSMB4*_qPCR_r2	CTGTGTCCATCTCCCAGAAGC
*SDHA*_qPCR_f1	CAGGTCACACACTGTTGCAGC

*SDHA*_qPCR_r1	TGCTCCGTCATGTAGTGGATG
*SDHA*_qPCR_f2	GTCATCGCACTGTGCATAGAGG
*SDHA*_qPCR_r2	GCAGACGTGCAGCTGAAGTAG
*VPS29*_qPCR_f1	TGGGGCATATAATGCCTTGG
*VPS29*_qPCR_r1	AGCTGATACACATAGGTGACCACTG
*VPS29*_qPCR_f2	TCCAGGTTCTGCCACTGGG
*VPS29*_qPCR_r2	TACTTTCACATCATCTCCAATTAGC
*ywhaz*_qPCR_f1	TACCGTTACTTGGCTGAGGTTG
*ywhaz*_qPCR_r2	GGTATGCTTGTTGTGACTGATCG
*ywhaz*_qPCR_f2	CAGCACGCTAATAATGCAATTAC
*ywhaz*_qPCR_r2	CCTCCTTCTCCTGCTTCAGC
*PSMB2*_qPCR_f1	TGGCCGCCAGCAATATTG
*PSMB2*_qPCR_r1	CAGTGTCTCCAGCCTCTCCAAC
*PSMB2*_qPCR_f2	CACGGCTATGGTGCCTTCC
*PSMB2*_qPCR_r2	CATTTCCTAAGGAGTTCCACTGC
Target genes	
*NOTCH3*_qPCR_f1	CTGCTGCTGCTGCTAGCG
*NOTCH3*_qPCR_r1	GCAGCTGGGTGCAACGAC
*NOTCH3*_qPCR_f2	GTCTGCCAGAGTTCAGTGGTG
*NOTCH3*_qPCR_r2	CAGGGATCTGGCAGGGAG
*EFNB3*_qPCR_f1	GCTCAGCCTGGAGCCTGTC

*EFNB3*_qPCR_r1	CCCGATCTGAGGGTACAGCAC
*EFNB3*_qPCR_f2	CATGAAGGTGCTTCTCCGAG
*EFNB3*_qPCR_r2	AGGCTCCAGGCTGTGGG
*HSPG2*_qPCR_f1	GCTGAGGGCATACGATGGC
*HSPG2*_qPCR_r1	GGTCCCCACTGCCCAGG
*HSPG2*_qPCR_f2	CCGACTGCAGGGACATGTC
*HSPG2*_qPCR_r2	CCGGGGCGGTAAAGATG
*NRTN*_qPCR_f1	CACCGCCTGCCTCGAAC
*NRTN*_qPCR_r1	AGCTCGCGCAGCTCCATC
*NRTN*_qPCR_f2	TCGGACCTGCGCTGGTC
*NRTN*_qPCR_r2	GCCCAGGGCGTCAGCTC
*KCNJ4*_qPCR_f1	ACGTGCGTGCCCAGCTG
*KCNJ4*_qPCR_r1	AAGCCCACATCCACATCCTG
*KCNJ4*_qPCR_f2	TTCAGCGAGAACGCCGTC
*KCNJ4*_qPCR_r2	GTACTCACCCTCTGGGGTCAC
*PDE4A*_qPCR_f1	CCTGTACCGCTCAGACAGCG
*PDE4A*_qPCR_r1	TTACGATGAGGTCTTCAGCGTG
*PDE4A*_qPCR_f2	GAGCGTGGCATGGAAATCA
*PDE4A*_qPCR_r2	GCATCTGGGTGGACAAGGTC
*PLPPR3*_qPCR_f1	ACGCAGTACCGCAGCCAC
*PLPPR3*_qPCR_r1	AAGTTGCCCACCGCGTG

*PLPPR3*_qPCR_f2	CCTACATCACGCAGGACATCTG
*PLPPR3*_qPCR_r2	GCGAAGGCTGACAGCGTG
*SCN4B*_qPCR_f1	AGGACCTCCACTTCCGGTG
*SCN4B*_qPCR_r1	GTCACCTTGGGGTCAGACTTC
*SCN4B*_qPCR_f2	CCACCATCTTCCTCCAAGTCG
*SCN4B*_qPCR_r2	CAGCCAGGATGATGAGTGTCAC
*SLC1A3*_qPCR_f1	CCTGGGGAACTTCTGATGAGG
*SLC1A3*_qPCR_r1	CACCACAGCAATGATGGTGG
*SLC1A3*_qPCR_f2	ACATCTGTCGGCCTGCCC
*SLC1A3*_qPCR_r2	CTCCCAGGGAGTCTCCCAGTAC
*SORCS1*_qPCR_f1	CTATAACCTGGGGAGCATCACAG
*SORCS1*_qPCR_r1	CTCATAGGTTGTTCCATAATCGG
*SORCS1*_qPCR_f2	CACCCCAAACAAGAAGACTGG
*SORCS1*_qPCR_r2	GATAAGCTGCCATCTTCTCCC

9. Bibliography

9.1 Papers and books

Ali, M., H. Rincon-Arano, W. Zhao, S. B. Rothbart, Q. Tong, S. M. Parkhurst, B. D. Strahl, L. W. Deng, M. Groudine, and T. G. Kutateladze. 2013. 'Molecular basis for chromatin binding and regulation of MLL5', *Proceedings of the National Academy of Sciences of the United States of America*, 110: 11296-301.

Bannur, S. V., S. V. Kulgod, S. S. Metkar, S. K. Mahajan, and J. K. Sainis. 1999. 'Protein determination by ponceau S using digital color image analysis of protein spots on nitrocellulose membranes', *Anal Biochem*, 267: 382-9.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. 'High-resolution profiling of histone methylations in the human genome', *Cell*, 129: 823-37.

Bayart, Emilie, and Odile Cohen-Haguenauer. 2013. 'Technological overview of iPS induction from human adult somatic cells', *Current gene therapy*, 13: 73-92.

Berdasco, M., and M. Esteller. 2013. 'Genetic syndromes caused by mutations in epigenetic genes', *Hum Genet*, 132: 359-83.

Bjorkegren, J. L., S. Hagg, H. A. Talukdar, H. Foroughi Asl, R. K. Jain, C. Cedergren, M. M. Shang, A. Rossignoli, R. Takolander, O. Melander, A. Hamsten, T. Michoel, and J. Skogsberg. 2014. 'Plasma cholesterol-induced lesion networks activated before regression of early, mature, and advanced atherosclerosis', *PLoS Genet*, 10: e1004201.

Bjornsson, H. T., J. S. Benjamin, L. Zhang, J. Weissman, E. E. Gerber, Y. C. Chen, R. G. Vaurio, M. C. Potter, K. D. Hansen, and H. C. Dietz. 2014. 'Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome', *Sci Transl Med*, 6: 256ra135.

Black, J. C., C. Van Rechem, and J. R. Whetstine. 2012. 'Histone lysine methylation dynamics: establishment, regulation, and biological impact', *Mol Cell*, 48: 491-507.

Boissinot, K., A. Huletsky, R. Peytavi, S. Turcotte, V. Veillette, M. Boissinot, F. J. Picard, Artel Ea m, and M. G. Bergeron. 2007. 'Rapid exonuclease digestion of PCR-amplified targets for improved microarray hybridization', *Clin Chem*, 53: 2020-3.

Braun, N. N., T. J. Reutiman, S. Lee, T. D. Folsom, and S. H. Fatemi. 2007. 'Expression of phosphodiesterase 4 is altered in the brains of subjects with autism', *Neuroreport*, 18: 1841-4.

Bustin, S. A. 2000. 'Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays', *J Mol Endocrinol*, 25: 169-93.

Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, and C. T. Wittwer. 2009. 'The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments', *Clin Chem*, 55: 611-22.

Chan, W. T., C. S. Verma, D. P. Lane, and S. K. Gan. 2013. 'A comparison and optimization of methods and factors affecting the transformation of Escherichia coli', *Biosci Rep*, 33.

Chapman, J. R., and J. Waldenstrom. 2015. 'With Reference to Reference Genes: A Systematic Review of Endogenous Controls in Gene Expression Studies', *PLoS ONE*, 10: e0141853.

Charpentier, M., A. H. Y. Khedher, S. Menoret, A. Brion, K. Lamribet, E. Dardillac, C. Boix, L. Perrouault, L. Tesson, S. Geny, A. De Cian, J. M. Itier, I. Anegon, B. Lopez, C. Giovannangeli, and J. P. Concordet. 2018. 'CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair', *Nat Commun*, 9: 1133.

Chen, Q. W., X. Y. Zhu, Y. Y. Li, and Z. Q. Meng. 2014. 'Epigenetic regulation and cancer (review)', *Oncol Rep*, 31: 523-32.

Cheng, F., J. Liu, S. H. Zhou, X. N. Wang, J. F. Chew, and L. W. Deng. 2008. 'RNA interference against mixed lineage leukemia 5 resulted in cell cycle arrest', *Int J Biochem Cell Biol*, 40: 2472-81.

Cheung, P., and P. Lau. 2005. 'Epigenetic regulation by histone methylation and histone variants', *Mol Endocrinol*, 19: 563-73.

Chu, V. T., T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky, and R. Kuhn. 2015. 'Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells', *Nat Biotechnol*, 33: 543-8.

Clark, David P., and Nanette Jean Pazdernik. 2016. *Biotechnology* (Elsevier/AP Cell, Academic Cell is an imprint of Elsevier: Amsterdam ; Boston).

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang. 2013. 'Multiplex genome engineering using CRISPR/Cas systems', *Science*, 339: 819-23.

Consortium, G. TEx. 2013. 'The Genotype-Tissue Expression (GTEx) project', *Nat Genet*, 45: 580-5.

Cooper, G. M. 2000. 'The Cell: A Molecular Approach, 2nd edition'.

Crews, D. 2008. 'Epigenetics and its implications for behavioral neuroendocrinology', *Front Neuroendocrinol*, 29: 344-57.

Cruceanu, C., M. Alda, C. Nagy, E. Freemantle, G. A. Rouleau, and G. Turecki. 2013. 'H3K4 tri-methylation in synapsin genes leads to different expression patterns in bipolar disorder and major depression', *Int J Neuropsychopharmacol*, 16: 289-99.

Damm, F., T. Oberacker, F. Thol, E. Surdziel, K. Wagner, A. Chaturvedi, M. Morgan, K. Bomm, G. Gohring, M. Lubbert, L. Kanz, W. Fiedler, B. Schlegelberger, G. Heil, R. F. Schlenk, K. Dohner, H. Dohner, J. Krauter, A. Ganser, and M. Heuser. 2011. 'Prognostic importance of histone methyltransferase MLL5 expression in acute myeloid leukemia', *J Clin Oncol*, 29: 682-9.

Del Rizzo, P. A., and R. C. Trievel. 2011. 'Substrate and product specificities of SET domain methyltransferases', *Epigenetics*, 6: 1059-67.

Deng, L. W., I. Chiu, and J. L. Strominger. 2004. 'MLL 5 protein forms intranuclear foci, and overexpression inhibits cell cycle progression', *Proceedings of the National Academy of Sciences of the United States of America*, 101: 757-62.

Derecki, N. C., J. C. Cronk, Z. Lu, E. Xu, S. B. Abbott, P. G. Guyenet, and J. Kipnis. 2012. 'Wild-type microglia arrest pathology in a mouse model of Rett syndrome', *Nature*, 484: 105-9.

Deveau, H., J. E. Garneau, and S. Moineau. 2010. 'CRISPR/Cas system and its role in phage-bacteria interactions', *Annu Rev Microbiol*, 64: 475-93.

Dong, S., M. F. Walker, N. J. Carriero, M. DiCola, A. J. Willsey, A. Y. Ye, Z. Waqar, L. E. Gonzalez, J. D. Overton, S. Frahm, J. F. Keaney, 3rd, N. A. Teran, J. Dea, J. D. Mandell, V. Hus Bal, C. A. Sullivan, N. M. DiLullo, R. O. Khalil, J. Gockley, Z. Yuksel, S. M. Sertel, A. G. Ercan-Sencicek, A. R. Gupta, S. M. Mane, M. Sheldon, A. I. Brooks, K. Roeder, B. Devlin, M. W. State, L. Wei, and S. J. Sanders. 2014. 'De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder', *Cell Rep*, 9: 16-23.

Dowling, P., and M. Clynes. 2011. 'Conditioned media from cell lines: a complementary model to clinical specimens for the discovery of disease-specific biomarkers', *Proteomics*, 11: 794-804.

EauClaire, S. F., and C. J. Webb. 2019. 'A CRISPR/Cas9 method to generate heterozygous alleles in Saccharomyces cerevisiae', *Yeast*.

Edwards, T. G., and D. C. Bloom. 2019. 'Lund Human Mesencephalic (LUHMES) Neuronal Cell Line Supports Herpes Simplex Virus 1 Latency In Vitro', *J Virol*, 93.

Eisenberg, E., and E. Y. Levanon. 2013. 'Human housekeeping genes, revisited', *Trends Genet*, 29: 569-74.

Eissenberg, J. C., and A. Shilatifard. 2010. 'Histone H3 lysine 4 (H3K4) methylation in development and differentiation', *Dev Biol*, 339: 240-9.

Emerling, B. M., J. Bonifas, C. P. Kratz, S. Donovan, B. R. Taylor, E. D. Green, M. M. Le Beau, and K. M. Shannon. 2002. 'MLL5, a homolog of Drosophila trithorax located within a segment of chromosome band 7q22 implicated in myeloid leukemia', *Oncogene*, 21: 4849-54.

Ernst, C., E. S. Chen, and G. Turecki. 2009. 'Histone methylation and decreased expression of TrkB.T1 in orbital frontal cortex of suicide completers', *Mol Psychiatry*, 14: 830-2.

Faundes, V., W. G. Newman, L. Bernardini, N. Canham, J. Clayton-Smith, B. Dallapiccola, S. J. Davies, M. K. Demos, A. Goldman, H. Gill, R. Horton, B. Kerr, D. Kumar, A. Lehman, S. McKee, J. Morton, M. J. Parker, J. Rankin, L. Robertson, I. K. Temple, Sequencing Clinical Assessment of the Utility of, Study Evaluation as a Service, Study Deciphering Developmental Disorders, and S. Banka. 2018. 'Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders', *Am J Hum Genet*, 102: 175-87.

Forbes, S. A., D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, C. Y. Kok, M. Jia, T. De, J. W. Teague, M. R. Stratton, U. McDermott, and P. J. Campbell. 2015. 'COSMIC: exploring the world's knowledge of somatic mutations in human cancer', *Nucleic Acids Res*, 43: D805-11.

Garibyan, L., and N. Avashia. 2013. 'Polymerase chain reaction', *J Invest Dermatol*, 133: 1-4.

Ghosh, R., J. E. Gilda, and A. V. Gomes. 2014. 'The necessity of and strategies for improving confidence in the accuracy of western blots', *Expert Rev Proteomics*, 11: 549-60.

Glaser, S., J. Schaft, S. Lubitz, K. Vintersten, F. van der Hoeven, K. R. Tufteland, R. Aasland, K. Anastassiadis, S. L. Ang, and A. F. Stewart. 2006. 'Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development', *Development*, 133: 1423-32.

Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn. 1977. 'Characteristics of a human cell line transformed by DNA from human adenovirus type 5', *J Gen Virol*, 36: 59-74.

Griesi-Oliveira, K., A. Acab, A. R. Gupta, D. Y. Sunaga, T. Chailangkarn, X. Nicol, Y. Nunez, M. F. Walker, J. D. Murdoch, S. J. Sanders, T. V. Fernandez, W. Ji, R. P. Lifton, E. Vadasz, A. Dietrich, D. Pradhan, H. Song, G. L. Ming, X. Gu, G. Haddad, M. C. N. Marchetto, N. Spitzer, M. R. Passos-Bueno, M. W. State, and A. R. Muotri. 2015. 'Modeling non-syndromic autism and the impact of TRPC6 disruption in human neurons', *Molecular Psychiatry*, 20: 1350-65.

Houslay, M. D., P. Schafer, and K. Y. Zhang. 2005. 'Keynote review: phosphodiesterase-4 as a therapeutic target', *Drug Discov Today*, 10: 1503-19.

Iossifov, I., D. Levy, J. Allen, K. Ye, M. Ronemus, Y. H. Lee, B. Yamrom, and M. Wigler. 2015. 'Low load for disruptive mutations in autism genes and their biased transmission', *Proceedings of the National Academy of Sciences of the United States of America*, 112: E5600-7.

Iossifov, I., M. Ronemus, D. Levy, Z. Wang, I. Hakker, J. Rosenbaum, B. Yamrom, Y. H. Lee, G. Narzisi, A. Leotta, J. Kendall, E. Grabowska, B. Ma, S. Marks, L. Rodgers, A. Stepansky, J. Troge, P. Andrews, M. Bekritsky, K. Pradhan, E. Ghiban, M. Kramer, J. Parla, R. Demeter, L. L. Fulton, R. S. Fulton, V. J. Magrini, K. Ye, J. C. Darnell, R. B. Darnell, E. R. Mardis, R. K. Wilson, M. C. Schatz, W. R. McCombie, and M. Wigler. 2012. 'De novo gene disruptions in children on the autistic spectrum', *Neuron*, 74: 285-99.

Iwaki, Takayuki, and Kazuo Umemura. 2011. 'A single plasmid transfection that offers a significant advantage associated with puromycin selection, fluorescence-assisted cell sorting, and doxycycline-inducible protein expression in mammalian cells', *Cytotechnology*, 63: 337-43.

Jang, Da Eun, Jae Young Lee, Jae Hoon Lee, Ok Jae Koo, Hee Sook Bae, Min Hee Jung, Ji Hyun Bae, Woo Sung Hwang, Yoo Jin Chang, Yoon Hoo Lee, Han Woong Lee, and Su Cheong Yeom. 2018. 'Multiple sgRNAs with overlapping sequences enhance CRISPR/Cas9-mediated knock-in efficiency', *Experimental & Molecular Medicine*, 50: 16.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. 2012. 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*, 337: 816-21.

Jones, W. D., D. Dafou, M. McEntagart, W. J. Woollard, F. V. Elmslie, M. Holder-Espinasse, M. Irving, A. K. Saggar, S. Smithson, R. C. Trembath, C. Deshpande, and M. A. Simpson. 2012. 'De novo mutations in MLL cause Wiedemann-Steiner syndrome', *Am J Hum Genet*, 91: 358-64.

Kandala, Divya T., Alessia Del Piano, Luca Minati, and Massimiliano Clamer. 2018. 'Targeting translation activity at the ribosome interface with UV-active small molecules', *bioRxiv*: 436311.

Kang, H. J., Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak, T. Guennel, Y. Shin, M. B. Johnson, Z. Krsnik, S. Mayer, S. Fertuzinhos, S. Umlauf, S. N. Lisgo, A. Vortmeyer, D. R. Weinberger, S. Mane, T. M. Hyde, A. Huttner, M. Reimers, J. E. Kleinman, and N. Sestan. 2011. 'Spatio-temporal transcriptome of the human brain', *Nature*, 478: 483-9.

Kapuscinski, J. 1995. 'DAPI: a DNA-specific fluorescent probe', *Biotech Histochem*, 70: 220-33.

Kervestin, S., and A. Jacobson. 2012. 'NMD: a multifaceted response to premature translational termination', *Nat Rev Mol Cell Biol*, 13: 700-12.

Kim, J. H., J. H. Lee, I. S. Lee, S. B. Lee, and K. S. Cho. 2017. 'Histone Lysine Methylation and Neurodevelopmental Disorders', *Int J Mol Sci*, 18.

Kim, Tae Kyung, and James H. Eberwine. 2010. 'Mammalian cell transfection: the present and the future', *Analytical and bioanalytical chemistry*, 397: 3173-78.

Kittler, R., L. Pelletier, A. K. Heninger, M. Slabicki, M. Theis, L. Miroslaw, I. Poser, S. Lawo, H. Grabner, K. Kozak, J. Wagner, V. Surendranath, C. Richter, W. Bowen, A. L. Jackson, B. Habermann, A. A. Hyman, and F. Buchholz. 2007. 'Genome-scale RNAi profiling of cell division in human tissue culture cells', *Nat Cell Biol*, 9: 1401-12.

Kleefstra, T., J. M. Kramer, K. Neveling, M. H. Willemsen, T. S. Koemans, L. E. Vissers, W. Wissink-Lindhout, M. Fenckova, W. M. van den Akker, N. N. Kasri, W. M. Nillesen, T. Prescott, R. D. Clark, K. Devriendt, J. van Reeuwijk, A. P. de Brouwer, C. Gilissen, H. Zhou, H. G. Brunner, J. A. Veltman, A. Schenck, and H. van Bokhoven. 2012. 'Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability', *Am J Hum Genet*, 91: 73-82.

Koemans, T. S., T. Kleefstra, M. C. Chubak, M. H. Stone, M. R. F. Reijnders, S. de Munnik, M. H. Willemsen, M. Fenckova, Ctrm Stumpel, L. A. Bok, M. Sifuentes Saenz, K. A. Byerly, L. B. Baughn, A. P. A. Stegmann, R. Pfundt, H. Zhou, H. van Bokhoven, A. Schenck, and J. M. Kramer. 2017. 'Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder', *PLoS Genet*, 13: e1006864.

Kouzarides, T. 2007. 'Chromatin modifications and their function', *Cell*, 128: 693-705.

Kramer, J. M., and H. van Bokhoven. 2009. 'Genetic and epigenetic defects in mental retardation', *Int J Biochem Cell Biol*, 41: 96-107.

Kuznetsova, Alla V., Alexander M. Kurinov, and Maria A. Aleksandrova. 2014. 'Cell models to study regulation of cell transformation in pathologies of retinal pigment epithelium', *Journal of ophthalmology*, 2014: 801787-87.

Labuhn, M., F. F. Adams, M. Ng, S. Knoess, A. Schambach, E. M. Charpentier, A. Schwarzer, J. L. Mateo, J. H. Klusmann, and D. Heckl. 2017. 'Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications', *Nucleic Acids Res*.

Laemmli, U. K. 1970. 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, 227: 680-5.

Lalam, N. 2006. 'Estimation of the reaction efficiency in polymerase chain reaction', *J Theor Biol*, 242: 947-53.

Lanza, A. M., D. S. Kim, and H. S. Alper. 2013. 'Evaluating the influence of selection markers on obtaining selected pools and stable cell lines in human cells', *Biotechnol J*, 8: 811-21.

Laumonnier, F., F. Bonnet-Brilhault, M. Gomot, R. Blanc, A. David, M. P. Moizard, M. Raynaud, N. Ronce, E. Lemonnier, P. Calvas, B. Laudier, J. Chelly, J. P. Fryns, H. H. Ropers, B. C. Hamel, C. Andres, C. Barthelemy, C. Moraine, and S. Briault. 2004. 'X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family', *Am J Hum Genet*, 74: 552-7.

Lee, P. Y., J. Costumbrado, C. Y. Hsu, and Y. H. Kim. 2012. 'Agarose gel electrophoresis for the separation of DNA fragments', *J Vis Exp*.

Li, Y., J. Han, Y. Zhang, F. Cao, Z. Liu, S. Li, J. Wu, C. Hu, Y. Wang, J. Shuai, J. Chen, L. Cao, D. Li, P. Shi, C. Tian, J. Zhang, Y. Dou, G. Li, Y. Chen, and M. Lei. 2016. 'Structural basis for activity regulation of MLL family methyltransferases', *Nature*, 530: 447-52.

Lin, S., B. T. Staahl, R. K. Alla, and J. A. Doudna. 2014. 'Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery', *Elife*, 3: e04766.

Lin, Y. C., M. Boone, L. Meuris, I. Lemmens, N. Van Roy, A. Soete, J. Reumers, M. Moisse, S. Plaisance, R. Drmanac, J. Chen, F. Speleman, D. Lambrechts, Y. Van de Peer, J. Tavernier, and N. Callewaert. 2014. 'Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations', *Nat Commun*, 5: 4767.

Lindeboom, R. G., F. Supek, and B. Lehner. 2016. 'The rules and impact of nonsense-mediated mRNA decay in human cancers', *Nat Genet*, 48: 1112-8.

Liu, J., F. Cheng, and L. W. Deng. 2012. 'MLL5 maintains genomic integrity by regulating the stability of the chromosomal passenger complex through a functional interaction with Borealin', *J Cell Sci*, 125: 4676-85.

Liu, J., X. N. Wang, F. Cheng, Y. C. Liou, and L. W. Deng. 2010. 'Phosphorylation of mixed lineage leukemia 5 by CDC2 affects its cellular distribution and is required for mitotic entry', *J Biol Chem*, 285: 20904-14.

Liu, Mingjie, Saad Rehman, Xidian Tang, Kui Gu, Qinlei Fan, Dekun Chen, and Wentao Ma. 2019. 'Methodologies for Improving HDR Efficiency', *Frontiers in genetics*, 9: 691-91.

Madan, V., B. Madan, U. Brykczynska, F. Zilbermann, K. Hogeveen, K. Dohner, H. Dohner, O. Weber, C. Blum, H. R. Rodewald, P. Sassone-Corsi, A. H. Peters, and H. J. Fehling. 2009. 'Impaired function of primitive hematopoietic cells in mice lacking the Mixed-Lineage-Leukemia homolog MLL5', *Blood*, 113: 1444-54.

Mahmood, Tahrin, and Ping-Chang Yang. 2012. 'Western blot: technique, theory, and trouble shooting', *North American journal of medical sciences*, 4: 429-34.

Mali, P., K. M. Esvelt, and G. M. Church. 2013. 'Cas9 as a versatile tool for engineering biology', *Nat Methods*, 10: 957-63.

Maruyama, T., S. K. Dougan, M. C. Truttmann, A. M. Bilate, J. R. Ingram, and H. L. Ploegh. 2015. 'Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining', *Nat Biotechnol*, 33: 538-42.

Mas, Y. Mas S., M. Barbon, C. Teyssier, H. Demene, J. E. Carvalho, L. E. Bird, A. Lebedev, J. Fattori, M. Schubert, C. Dumas, W. Bourguet, and A. le

Maire. 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.

McKenzie, Katherine, Meagan Milton, Glenys Smith, and Hélène Ouellette-Kuntz. 2016. 'Systematic Review of the Prevalence and Incidence of Intellectual Disabilities: Current Trends and Issues', *Current Developmental Disorders Reports*, 3: 104-15.

Meyer, E., K. J. Carss, J. Rankin, J. M. Nichols, D. Grozeva, A. P. Joseph, N. E. Mencacci, A. Papandreou, J. Ng, S. Barral, A. Ngoh, H. Ben-Pazi, M. A. Willemsen, D. Arkadir, A. Barnicoat, H. Bergman, S. Bhate, A. Boys, N. Darin, N. Foulds, N. Gutowski, A. Hills, H. Houlden, J. A. Hurst, Z. Israel, M. Kaminska, P. Limousin, D. Lumsden, S. McKee, S. Misra, S. S. Mohammed, V. Nakou, J. Nicolai, M. Nilsson, H. Pall, K. J. Peall, G. B. Peters, P. Prabhakar, M. S. Reuter, P. Rump, R. Segel, M. Sinnema, M. Smith, P. Turnpenny, S. M. White, D. Wieczorek, S. Wiethoff, B. T. Wilson, G. Winter, C. Wragg, S. Pope, S. J. Heales, D. Morrogh, Uk K. Consortium, Study Deciphering Developmental Disorders, Nihr BioResource Rare Diseases Consortium, A. Pittman, L. J. Carr, B. Perez-Duenas, J. P. Lin, A. Reis, W. A. Gahl, C. Toro, K. P. Bhatia, N. W. Wood, E. J. Kamsteeg, W. K. Chong, P. Gissen, M. Topf, R. C. Dale, J. R. Chubb, F. L. Raymond, and M. A. Kurian. 2017. 'Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia', *Nat Genet*, 49: 223-37.

Millar, J. K., B. S. Pickard, S. Mackie, R. James, S. Christie, S. R. Buchanan, M. P. Malloy, J. E. Chubb, E. Huston, G. S. Baillie, P. A. Thomson, E. V. Hill, N. J. Brandon, J. C. Rain, L. M. Camargo, P. J. Whiting, M. D. Houslay, D. H. Blackwood, W. J. Muir, and D. J. Porteous. 2005. 'DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling', *Science*, 310: 1187-91.

Miller, J. C., M. C. Holmes, J. Wang, D. Y. Guschin, Y. L. Lee, I. Rupniewski, C. M. Beausejour, A. J. Waite, N. S. Wang, K. A. Kim, P. D. Gregory, C. O. Pabo, and E. J. Rebar. 2007. 'An improved zinc-finger nuclease architecture for highly specific genome editing', *Nat Biotechnol*, 25: 778-85.

Miyake, N., Y. Tsurusaki, E. Koshimizu, N. Okamoto, T. Kosho, N. J. Brown, T. Y. Tan, P. J. Yap, H. Suzumura, T. Tanaka, T. Nagai, M. Nakashima, H. Saitsu, N. Niikawa, and N. Matsumoto. 2016. 'Delineation of clinical features in Wiedemann-Steiner syndrome caused by KMT2A mutations', *Clin Genet*, 89: 115-9.

Mojica, F. J., C. Diez-Villasenor, E. Soria, and G. Juez. 2000. 'Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria', *Mol Microbiol*, 36: 244-6.

Musselman, C. A., and T. G. Kutateladze. 2009. 'PHD fingers: epigenetic effectors and potential drug targets', *Mol Interv*, 9: 314-23.

Ng, S. B., A. W. Bigham, K. J. Buckingham, M. C. Hannibal, M. J. McMillin, H. I. Gildersleeve, A. E. Beck, H. K. Tabor, G. M. Cooper, H. C. Mefford, C. Lee, E. H. Turner, J. D. Smith, M. J. Rieder, K. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D. A. Nickerson, M. J. Bamshad, and J. Shendure. 2010. 'Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome', *Nat Genet*, 42: 790-3.

Niikawa, N., N. Matsuura, Y. Fukushima, T. Ohsawa, and T. Kajii. 1981. 'Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency', *J Pediatr*, 99: 565-9.

O'Donnell-Luria, A. H., L. S. Pais, V. Faundes, J. C. Wood, A. Sveden, V. Luria, R. Abou Jamra, A. Accogli, K. Amburgey, B. M. Anderlid, S. Azzarello-Burri, A. A. Basinger, C. Bianchini, L. M. Bird, R. Buchert, W. Carre, S. Ceulemans, P. Charles, H. Cox, L. Culliton, A. Curro, Study Deciphering Developmental Disorders, F. Demurger, J. J. Dowling, B. Duban-Bedu, C. Dubourg, S. E. Eiset, L. F. Escobar, A. Ferrarini, T. B. Haack, M. Hashim, S. Heide, K. L. Helbig, I. Helbig, R. Heredia, D. Heron, B. Isidor, A. R. Jonasson, P. Joset, B. Keren, F. Kok, H. Y. Kroes, A. Lavillaureix, X. Lu, S. M. Maas, G. H. B. Maegawa, C. L. M. Marcelis, P. R. Mark, M. R. Masruha, H. M. McLaughlin, K. McWalter, E. U. Melchinger, S. Mercimek-Andrews, C. Nava, M. Pendziwiat, R. Person, G. P. Ramelli, L. L. P. Ramos, A. Rauch, C. Reavey, A. Renieri, A. Riess, A. Sanchez-Valle, S. Sattar, C. Saunders, N. Schwarz, T. Smol, M. Srour, K. Steindl, S. Syrbe, J. C. Taylor, A. Telegrafi, I. Thiffault, D. A. Trauner, H. van der Linden, Jr., S. van Koningsbruggen, L. Villard, I. Vogel, J. Vogt, Y. G. Weber, I. M. Wentzensen, E. Widjaja, J. Zak, S. Baxter, S. Banka, and L. H. Rodan. 2019. 'Heterozygous Variants in KMT2E Cause a Spectrum of Neurodevelopmental Disorders and Epilepsy', *Am J Hum Genet*, 104: 1210-22.

O'Roak, B. J., L. Vives, W. Fu, J. D. Egertson, I. B. Stanaway, I. G. Phelps, G. Carvill, A. Kumar, C. Lee, K. Ankenman, J. Munson, J. B. Hiatt, E. H. Turner, R. Levy, D. R. O'Day, N. Krumm, B. P. Coe, B. K. Martin, E. Borenstein, D. A. Nickerson, H. C. Mefford, D. Doherty, J. M. Akey, R. Bernier, E. E. Eichler, and J. Shendure. 2012. 'Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders', *Science*, 338: 1619-22.

Osipovich, A. B., R. Gangula, P. G. Vianna, and M. A. Magnuson. 2016. 'Setd5 is essential for mammalian development and the co-transcriptional regulation of histone acetylation', *Development*, 143: 4595-607.

Parkel, S., J. P. Lopez-Atalaya, and A. Barco. 2013. 'Histone H3 lysine methylation in cognition and intellectual disability disorders', *Learn Mem*, 20: 570-9.

Patel, D. R., D. E. Greydanus, J. L. Calles, Jr., and H. D. Pratt. 2010. 'Developmental disabilities across the lifespan', *Dis Mon*, 56: 304-97.

Pavlowsky, A., J. Chelly, and P. Billuart. 2012. 'Major synaptic signaling pathways involved in intellectual disability', *Molecular Psychiatry*, 17: 663-63.

Pegg, D. E. 2007. 'Principles of cryopreservation', *Methods Mol Biol*, 368: 39-57.

Peter, C. J., and S. Akbarian. 2011. 'Balancing histone methylation activities in psychiatric disorders', *Trends Mol Med*, 17: 372-9.

Pfaffl, M. W. 2001. 'A new mathematical model for relative quantification in real-time RT-PCR', *Nucleic Acids Res*, 29: e45.

Pijnappel, W. W., D. Schaft, A. Roguev, A. Shevchenko, H. Tekotte, M. Wilm, G. Rigaut, B. Seraphin, R. Aasland, and A. F. Stewart. 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 Dev*, 15: 2991-3004.

Prince, M., V. Patel, S. Saxena, M. Maj, J. Maselko, M. R. Phillips, and A. Rahman. 2007. 'No health without mental health', *Lancet*, 370: 859-77.

Rabello Ddo, A., C. A. de Moura, R. V. de Andrade, A. B. Motoyama, and F. P. Silva. 2013. 'Altered expression of MLL methyltransferase family genes in breast cancer', *Int J Oncol*, 43: 653-60.

Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang. 2013. 'Genome engineering using the CRISPR-Cas9 system', *Nat Protoc*, 8: 2281-308.

Rao, X., X. Huang, Z. Zhou, and X. Lin. 2013. 'An improvement of the 2^(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis', *Biostat Bioinforma Biomath*, 3: 71-85.

Rauch, A., J. Hoyer, S. Guth, C. Zweier, C. Kraus, C. Becker, M. Zenker, U. Huffmeier, C. Thiel, F. Ruschendorf, P. Nurnberg, A. Reis, and U. Trautmann. 2006. 'Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation', *Am J Med Genet A*, 140: 2063-74.

Remington, S. J. 2011. 'Green fluorescent protein: a perspective', *Protein Sci*, 20: 1509-19.

Richardson, C. D., G. J. Ray, M. A. DeWitt, G. L. Curie, and J. E. Corn. 2016. 'Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA', *Nat Biotechnol*, 34: 339-44.

Rincon-Arano, H., J. Halow, J. J. Delrow, S. M. Parkhurst, and M. Groudine. 2012. 'UpSET recruits HDAC complexes and restricts chromatin accessibility and acetylation at promoter regions', *Cell*, 151: 1214-28.

Ronan, J. L., W. Wu, and G. R. Crabtree. 2013. 'From neural development to cognition: unexpected roles for chromatin', *Nat Rev Genet*, 14: 347-59.

Rothkamm, K., I. Kruger, L. H. Thompson, and M. Lobrich. 2003. 'Pathways of DNA double-strand break repair during the mammalian cell cycle', *Mol Cell Biol*, 23: 5706-15.

Rumbaugh, G., and C. A. Miller. 2011. 'Epigenetic changes in the brain: measuring global histone modifications', *Methods Mol Biol*, 670: 263-74.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. 'Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase', *Science*, 239: 487-91.

Sander, J. D., and J. K. Joung. 2014. 'CRISPR-Cas systems for editing, regulating and targeting genomes', *Nat Biotechnol*, 32: 347-55.

Sanderson, M. J., I. Smith, I. Parker, and M. D. Bootman. 2014. 'Fluorescence microscopy', *Cold Spring Harb Protoc*, 2014: pdb top071795.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. 'DNA sequencing with chain-terminating inhibitors', *Proceedings of the National Academy of Sciences of the United States of America*, 74: 5463-7.

Sebastian, S., P. Sreenivas, R. Sambasivan, S. Cheedipudi, P. Kandalla, G. K. Pavlath, and J. Dhawan. 2009. 'MLL5, a trithorax homolog, indirectly regulates H3K4 methylation, represses cyclin A2 expression, and promotes myogenic differentiation', *Proceedings of the National Academy of Sciences of the United States of America*, 106: 4719-24.

Serezani, Carlos H., Megan N. Ballinger, David M. Aronoff, and Marc Peters-Golden. 2008. 'Cyclic AMP: master regulator of innate immune cell function', *American journal of respiratory cell and molecular biology*, 39: 127-32.

Shaw, G., S. Morse, M. Ararat, and F. L. Graham. 2002. 'Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells', *FASEB J*, 16: 869-71.

Shen, E., H. Shulha, Z. Weng, and S. Akbarian. 2014. 'Regulation of histone H3K4 methylation in brain development and disease', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 369.

Shulha, H. P., I. Cheung, C. Whittle, J. Wang, D. Virgil, C. L. Lin, Y. Guo, A. Lessard, S. Akbarian, and Z. Weng. 2012. 'Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons', *Arch Gen Psychiatry*, 69: 314-24.

Singh, T., M. I. Kurki, D. Curtis, S. M. Purcell, L. Crooks, J. McRae, J. Suvisaari, H. Chheda, D. Blackwood, G. Breen, O. Pietilainen, S. S. Gerety, M. Ayub, M. Blyth, T. Cole, D. Collier, E. L. Coomber, N. Craddock, M. J. Daly, J. Danesh, M. DiForti, A. Foster, N. B. Freimer, D. Geschwind, M. Johnstone, S. Joss, G. Kirov, J. Korkko, O. Kuismin, P. Holmans, C. M. Hultman, C. Iyegbe, J. Lonnqvist, M. Mannikko, S. A. McCarroll, P. McGuffin, A. M. McIntosh, A. McQuillin, J. S. Moilanen, C. Moore, R. M. Murray, R. Newbury-Ecob, W. Ouwehand, T. Paunio, E. Prigmore, E. Rees, D.

Roberts, J. Sambrook, P. Sklar, D. St Clair, J. Veijola, J. T. Walters, H. Williams, Study Swedish Schizophrenia, Interval Study, D. D. D. Study, Uk K. Consortium, P. F. Sullivan, M. E. Hurles, M. C. O'Donovan, A. Palotie, M. J. Owen, and J. C. Barrett. 2016. 'Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders', *Nat Neurosci*, 19: 571-7.

Snedecor, G.W. and Cochran, W.G. . 1980. 'Statistical Methods', 7th Edition, Iowa State University Press, Ames.

Srivastava, A. K., and C. E. Schwartz. 2014. 'Intellectual disability and autism spectrum disorders: causal genes and molecular mechanisms', *Neurosci Biobehav Rev*, 46 Pt 2: 161-74.

Stemmer, M., T. Thumberger, M. Del Sol Keyer, J. Wittbrodt, and J. L. Mateo. 2015. 'CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool', *PLoS ONE*, 10: e0124633.

Stepanenko, A. A., and V. V. Dmitrenko. 2015. 'HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution', *Gene*, 569: 182-90.

Svec, D., A. Tichopad, V. Novosadova, M. W. Pfaffl, and M. Kubista. 2015. 'How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments', *Biomol Detect Quantif*, 3: 9-16.

Takata, A., I. Ionita-Laza, J. A. Gogos, B. Xu, and M. Karayiorgou. 2016. 'De Novo Synonymous Mutations in Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia', *Neuron*, 89: 940-7.

Takata, A., B. Xu, I. Ionita-Laza, J. L. Roos, J. A. Gogos, and M. Karayiorgou. 2014. 'Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene', *Neuron*, 82: 773-80.

Tasdogan, A., S. Kumar, G. Allies, J. Bausinger, F. Beckel, H. Hofemeister, M. Mulaw, V. Madan, K. Scharffetter-Kochanek, M. Feuring-Buske, K. Doehner, G. Speit, A. F. Stewart, and H. J. Fehling. 2016. 'DNA Damage-Induced HSPC Malfunction Depends on ROS Accumulation Downstream of IFN-1 Signaling and Bid Mobilization', *Cell Stem Cell*, 19: 752-67.

Taverna, S. D., H. Li, A. J. Ruthenburg, C. D. Allis, and D. J. Patel. 2007. 'How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers', *Nat Struct Mol Biol*, 14: 1025-40.

Taylor, S. C., K. Nadeau, M. Abbasi, C. Lachance, M. Nguyen, and J. Fenrich. 2019. 'The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time', *Trends Biotechnol*, 37: 761-74.

Thomas, P., and T. G. Smart. 2005. 'HEK293 cell line: a vehicle for the expression of recombinant proteins', *J Pharmacol Toxicol Methods*, 51: 187-200.

Towbin, H., T. Staehelin, and J. Gordon. 1979. 'Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and

some applications', *Proceedings of the National Academy of Sciences of the United States of America*, 76: 4350-4.

Tukey, F. J. Anscombe and John W. 1963. 'The Examination and Analysis of Residuals', *Technometrics*, 5: pp. 141-60.

Uliana, V., S. Grosso, M. Cioni, F. Ariani, F. T. Papa, S. Tamburello, E. Rossi, E. Katzaki, M. Mucciolo, A. Marozza, M. Pollazzon, M. A. Mencarelli, F. Mari, P. Balestri, and A. Renieri. 2010. '3.2 Mb microdeletion in chromosome 7 bands q22.2-q22.3 associated with overgrowth and delayed bone age', *Eur J Med Genet*, 53: 168-70.

Vallianatos, C. N., and S. Iwase. 2015. 'Disrupted intricacy of histone H3K4 methylation in neurodevelopmental disorders', *Epigenomics*, 7: 503-19.

van Bokhoven, H. 2011. 'Genetic and epigenetic networks in intellectual disabilities', *Annu Rev Genet*, 45: 81-104.

van Diepen, Laura, Falk F. R. Buettner, Dirk Hoffmann, Christina T. Thiesler, Oliver von Bohlen und Halbach, Viola von Bohlen und Halbach, Lars R. Jensen, Doris Steinemann, Simon Edvardson, Orly Elpeleg, Axel Schambach, Rita Gerardy-Schahn, and Andreas W. Kuss. 2018. 'A patient-specific induced pluripotent stem cell model for West syndrome caused by ST3GAL3 deficiency', *European Journal of Human Genetics*, 26: 1773-83.

Verpelli, C., I. Galimberti, B. Gomez-Mancilla, and C. Sala. 2014. 'Molecular basis for prospective pharmacological treatment strategies in intellectual disability syndromes', *Dev Neurobiol*, 74: 197-206.

Vissers, L. E., C. Gilissen, and J. A. Veltman. 2016. 'Genetic studies in intellectual disability and related disorders', *Nat Rev Genet*, 17: 9-18.

Wang, Z., M. Gerstein, and M. Snyder. 2009. 'RNA-Seq: a revolutionary tool for transcriptomics', *Nat Rev Genet*, 10: 57-63.

Watson, J. D., and F. H. Crick. 1953. 'Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid', *Nature*, 171: 737-8.

Willsey, A. J., S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan, and M. W. State. 2013. 'Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism', *Cell*, 155: 997-1007.

Wood, A. J., T. W. Lo, B. Zeitler, C. S. Pickle, E. J. Ralston, A. H. Lee, R. Amora, J. C. Miller, E. Leung, X. Meng, L. Zhang, E. J. Rebar, P. D. Gregory, F. D. Urnov, and B. J. Meyer. 2011. 'Targeted genome editing across species using ZFNs and TALENs', *Science*, 333: 307.

World Health Organization. 2004. *International statistical classification of diseases and related health problems* (World Health Organization: Geneva).

Wu, M., P. F. Wang, J. S. Lee, S. Martin-Brown, L. Florens, M. Washburn, and A. Shilatifard. 2008. 'Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS', *Mol Cell Biol*, 28: 7337-44.

Yap, Damian B., David C. Walker, Leah M. Prentice, Steven McKinney, Gulisa Turashvili, Katrin Mooslehner-Allen, Teresa Ruiz de Algara, John Fee, Xavier d'Anglemont de Tassigny, William H. Colledge, and Samuel Aparicio. 2011. 'Mll5 is required for normal spermatogenesis', *PLoS ONE*, 6: e27127-e27.

Ye, Y., K. Jackson, and J. M. O'Donnell. 2000. 'Effects of repeated antidepressant treatment of type 4A phosphodiesterase (PDE4A) in rat brain', *J Neurochem*, 74: 1257-62.

Yun, H., F. Damm, D. Yap, A. Schwarzer, A. Chaturvedi, N. Jyotsana, M. Lubbert, L. Bullinger, K. Dohner, R. Geffers, S. Aparicio, R. K. Humphries, A. Ganser, and M. Heuser. 2014. 'Impact of MLL5 expression on decitabine efficacy and DNA methylation in acute myeloid leukemia', *Haematologica*, 99: 1456-64.

Zech, M., S. Boesch, E. M. Maier, I. Borggraefe, K. Vill, F. Laccone, V. Pilshofer, A. Ceballos-Baumann, B. Alhaddad, R. Berutti, W. Poewe, T. B. Haack, B. Haslinger, T. M. Strom, and J. Winkelmann. 2016. 'Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia', *Am J Hum Genet*, 99: 1377-87.

Zhang, X., W. Novera, Y. Zhang, and L. W. Deng. 2017. 'MLL5 (KMT2E): structure, function, and clinical relevance', *Cell Mol Life Sci*, 74: 2333-44.

Zhang, X., H. Wen, and X. Shi. 2012. 'Lysine methylation: beyond histones', *Acta Biochim Biophys Sin (Shanghai)*, 44: 14-27.

Zhang, Y., J. Wong, M. Klinger, M. T. Tran, K. M. Shannon, and N. Killeen. 2009. 'MLL5 contributes to hematopoietic stem cell fitness and homeostasis', *Blood*, 113: 1455-63.

Zhao, W., J. Liu, X. Zhang, and L. W. Deng. 2016. 'MLL5 maintains spindle bipolarity by preventing aberrant cytosolic aggregation of PLK1', *J Cell Biol*, 212: 829-43.

Zhou, Y., J. Kim, X. Yuan, and T. Braun. 2011. 'Epigenetic modifications of stem cells: a paradigm for the control of cardiac progenitor cells', *Circ Res*, 109: 1067-81.

9.2 Websites

https://blast.ncbi.nlm.nih.gov/Blast.cgi [accesss 04.05.2020]

https://blast.ncbi.nlm.nih.gov/Blast.cgi [accesss 04.05.2020]

https://genome.ucsc.edu [accesss 04.05.2020]

http://tools-genome-engineering.org [accesss 04.05.2020]

https://www.bioinformatics.org/sms/rev_comp.html [accesss 04.05.2020]

https://www.bio-rad.com/webroot/web/pdf/lsr/literature/4006028.pdf
[accesss 05.05.2020]
http://www.premierbiosoft.com/netprimer/netprlaunch/Help/xnetprlkaunch.html
[accesss 04.05.2020]

http://www.qiagenbioinformatics.lcom/products/indigenuity-pathway-analysis/
[accesss 04.05.2020]

https://www.r-project.org [accesss 04.05.2020]

10. Erklärung zum Eigenanteil der Dissertationsschrift

Die Arbeit wurde am Institut für Medizinische Genetik und angewandte Genomik der Eberhardt Karls Universität Tübingen unter Betreuung von Prof. Dr. Olaf Rieß und Dr. rer. nat. Rebecca Buchert durchgeführt.

Die Konzeption der Versuche erfolgte durch Dr. rer. nat. Rebecca Buchert als Betreuerin und mich, cand. med. Esther Ursula Melchinger.

Sämtliche Versuche wurden nach Einarbeitung durch Dr. rer. nat Rebecca Buchert und Labormitglied Philipp Koch von mir eigenständig durchgeführt. Die RNA-Seq wurde von der c.ATG Core Facility for NGS and Microarrays Tübingen unseres Instituts durchgeführt und die RNA-Seq Daten bioinformatisch durch die AG Epigenetik analysiert.

Die statistische Auswertung der Daten erfolgte nach Anleitung durch Prof. Dr. Dr. h.c. Albrecht Melchinger durch mich.

Ausgangspunkt dieses Promotionsprojekts waren die Exomsequenzierungs-daten einer Tübinger Patientin, welche Eingang fanden in eine Publikation, an der ich als Mitautorin beteiligt wurde. Die Ergebnisse dieser Promotionsarbeit flossen jedoch nicht unmittelbar in die Publikation ein. Insofern in der Promotionsarbeit auf diese Publikation Bezug genommen wird, sind diese Stellen als Zitate kenntlich gemacht.

Ich versichere, das Manuskript selbständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen verwendet zu haben.

11. Publications

Parts of this dissertation project have been incorporated in the following publication:

O'Donnell-Luria, A. H., L. S. Pais, V. Faundes, J. C. Wood, A. Sveden, V. Luria, R. Abou Jamra, A. Accogli, K. Amburgey, B. M. Anderlid, S. Azzarello-Burri, A. A. Basinger, C. Bianchini, L. M. Bird, R. Buchert, W. Carre, S. Ceulemans, P. Charles, H. Cox, L. Culliton, A. Curro, Study Deciphering Developmental Disorders, F. Demurger, J. J. Dowling, B. Duban-Bedu, C. Dubourg, S. E. Eiset, L. F. Escobar, A. Ferrarini, T. B. Haack, M. Hashim, S. Heide, K. L. Helbig, I. Helbig, R. Heredia, D. Heron, B. Isidor, A. R. Jonasson, P. Joset, B. Keren, F. Kok, H. Y. Kroes, A. Lavillaureix, X. Lu, S. M. Maas, G. H. B. Maegawa, C. L. M. Marcelis, P. R. Mark, M. R. Masruha, H. M. McLaughlin, K. McWalter, **E. U. Melchinger**, S. Mercimek-Andrews, C. Nava, M. Pendziwiat, R. Person, G. P. Ramelli, L. L. P. Ramos, A. Rauch, C. Reavey, A. Renieri, A. Riess, A. Sanchez-Valle, S. Sattar, C. Saunders, N. Schwarz, T. Smol, M. Srour, K. Steindl, S. Syrbe, J. C. Taylor, A. Telegrafi, I. Thiffault, D. A. Trauner, H. van der Linden, Jr., S. van Koningsbruggen, L. Villard, I. Vogel, J. Vogt, Y. G. Weber, I. M. Wentzensen, E. Widjaja, J. Zak, S. Baxter, S. Banka, and L. H. Rodan. 2019. 'Heterozygous Variants in KMT2E Cause a Spectrum of Neurodevelopmental Disorders and Epilepsy', *Am J Hum Genet*, 104: 1210-22.

12. Danksagung

Ich möchte mich herzlich bei meinem Doktorvater Prof Dr. Olaf Rieß für die Überlassung des Themas und die Begutachtung bedanken.

Des weiteren bin ich meiner Betreuerin Dr. rer. nat. Rebecca Buchert zu sehr großem Dank verpflichtet für die freundliche Betreuung und tatkräftige Unterstützung in allen Belangen der Arbeit.

Außerdem bedanke möchte ich mich bei Labormitglied Philipp Koch für seine hilfreiche Unterstützung und Geduld während der Einarbeitungsphase.

Ich bedanke mich außerdem bei Dr. Julia Schulze-Hentrich und ihrer Arbeitsgruppe für die Unterstützung bei der Auswertung der RNA-Seq Daten.