

**Integrative functional genomic search for  
regulatory DNA sequence polymorphisms  
influencing DNA methylation and mRNA  
expression in hippocampal brain tissue**

**I n a u g u r a l - D i s s e r t a t i o n**

zur

Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität zu Köln

vorgelegt von

**Ann-Kathrin Ruppert**

aus Köln

**2015**

Berichtersteller:  
Prof. Dr. Peter Nürnberg  
Prof. Dr. Peter M. Schneider

Vorsitz der Prüfung:  
Prof. Dr. Peter Kloppenburg

Beisitzerin:  
Dr. Birgit Budde

Tag der mündlichen Prüfung: 15.06.2015

## ZUSAMMENFASSUNG

Obleich bekannt ist, dass die Ausprägung zahlreicher neuropsychiatrischer Erkrankungen genetisch bedingt ist, sind die grundlegenden Mechanismen dieses Zusammenhangs noch weitestgehend unbekannt. Eine Methode, um Einblicke in die Genetik neuropsychiatrischer Erkrankungen zu erhalten, sind genomweite Assoziationsstudien (GWASs). Mit Hilfe dieser konnten bisher über 2.000 Loci für genetische Risikofaktoren von Hirnerkrankungen identifiziert werden. Die Mehrheit dieser Loci befindet sich in nicht-codierenden DNA-Bereichen, was ihre funktionelle Erforschung erschwert. Die vorliegende Arbeit geht der Fragestellung nach, inwieweit regulatorische Sequenzvarianten, welche DNA-Methylierung und Genexpression beeinflussen, zur genetischen Disposition von neuropsychiatrischen Erkrankungen beitragen.

Meine Studie nutzt einen integrativen Ansatz der funktionellen Genomik, um epigenetische Regulationen im hippocampalen Hirngewebe bei Patienten mit pharmakoresistenter mesialer Temporallappenepilepsie zu untersuchen. Hierzu wurden SNP-Genotypen mit genomweiter CpG-Methylierung und mRNA Genexpression korreliert. Die daraus resultierenden »genomweiten Landkarten« von quantitativen Methylierungs Trait Loci (meQTLs) und quantitativen Expressions Trait Loci (eQTLs) wurden zur Lokalisation von regulatorischen SNPs (rSNPs) verwendet die in Zusammenhang mit einigen Hirnerkrankungen stehen (488 GWAS Katalog Einträge,  $P < 5,0 \times 10^{-8}$ ).

Die vorliegende Arbeit stellt die erste meQTL Studie dar, welche den leistungsfähigen *Human Methylation450 array* auf Basis von frisch-gefrorenem menschlichem Hirngewebe verwendet. Mit Hilfe einer linearen Regressionsanalyse und unter Berücksichtigung einer Korrektur für die Gewebeheterogenität, wurden insgesamt 19.954 (8,5% der 362.000 CpGs) cis-regulierte meQTLs identifiziert. Dies entspricht einer Versechsfachung der bisher bekannten meQTLs aus postmortalem Hirngewebe. Eine signifikante Anreicherung der meQTLs in der 5'-regulatorischen Region vor den Genpromotoren (TSS201-1500;  $P = 7,7 \times 10^{-61}$ ) spiegelt den funktionellen Einfluss dieser Region wider, welche Enhancer als auch Insulatoren beherbergt. Es hat sich gezeigt, dass einige der hoch signifikanten cis-meQTLs bekannte Kandidaten Gene für neurologische Entwicklungsstörungen beeinflussen (*ADARB2*, *HDAC4*, *NAPRT1*, *MAD1L1*, *PTPRN2* und *RIMBP2*). Die Gewebespezifität wurde anhand einer weiteren meQTL-Analyse, unter Beibehaltung gleicher experimenteller Bedingungen, in Blutzellen von 496 deutsch stämmigen Kontrollproben ohne neuropsychiatrische Erkrankungen untersucht. 65% der im Hirngewebe identifizierten meQTLs konnten auch in Blutzellen wiedergefunden werden (Spearman Rank Koeffizient = 0,42). Diese nennenswerte Übereinstimmung eröffnet die Möglichkeit, epigenetische Biomarker für komplexe Hirnerkrankungen in einfach zugänglichem Gewebe auszuwählen. Die zusätzlich zur meQTL Analyse durchgeführte eQTL Analyse konnte unter den 31.000 mRNA-Sonden insgesamt 734 signifikant cis-wirkende eQTLs identifizieren.

In einer weiteren Analyse wurden CpG-Methylierung und Genexpression korreliert – diese stellt die erste systematische Untersuchung dieser Form in frisch-gefrorenem Hirngewebe dar. Hierbei wurden sowohl negative (73%) als auch positive (27%) Korrelationen beobachtet. Die stärksten negativen Korrelationen wurden bei dem Gen *NAPRT1*, welches die Nicotinat Phosphoribosyltransferase kodiert, beobachtet. Des Weiteren konnte bei den mit *NAPRT1* assoziierten meQTLs und eQTLs eine genetische Beeinflussung durch ein und denselben SNP rs9657360 festgestellt werden. Die Korrelations-Ergebnisse kombiniert mit der genetischen Beeinflussung des SNPs zeigten auf, dass das minor C Allel eben dieses SNPs mit einer erhöhten Methylierung in der *NAPRT1*-Promoterregion und einer verminderten Genexpression assoziiert ist. Die Kombination aus einer tumorspezifischen Hypermethylierung einer in der Promoterregion gelegenen CpG island mit gleichzeitiger Verminderung der *NAPRT1* Expression wurde ebenfalls in der Krebsforschung erkannt: *NAPRT1* kann als prädiktiver Biomarker zur Therapie von Karzinomen mit *NAMPT* Inhibitoren eingesetzt werden. Durch den innovativen Ansatz, translationale Auswirkungen der epigenetischen Regulation der Genexpression in Kombination mit meQTLs und eQTLs zu testen, wurde zusätzliche eine genetische Determination erkannt. Diese ist von großer klinischer Bedeutung, da sie einen Ansatz zur Erfassung von Patienten erlaubt, die von der Gabe von *NAMPT* Inhibitoren profitieren können.

Zusätzlich wurde eine Imprinting meQTL (imeQTL) Analyse durchgeführt, um das Potential der Kombination aus Imprinting und Methylierung zu untersuchen. Zur Erfassung der imeQTLs wurde der Methylierungsstatus von 269 Individuen (auf Basis von Blutzellen), stratifiziert nach den elterlich inversen heterozygoten Genotypen, verglichen. Insgesamt konnten 177 CpGs an 31 genomischen Loci identifiziert werden, von denen 22 bisher unbekannte Imprinting Regionen darstellen. Die stärkste Auswirkung von Parent-of-Origin-Effekte auf Methylierung wurde in Regionen beobachtet, die Gene für neurologische Entwicklungsstörungen beherbergen, sowie im chromosomalen Segment 3p21.1, welches eine GWAS Kandidaten Region für affektive Störungen ist. Positionelle Gene in Imprinting Regionen sind aussichtsreiche Kandidaten Gene aufgrund ihrer potentiell monoallelischen Genexpression. Hierdurch wird es möglich, potentiell rezessive Erkrankungsmutationen zu ermitteln.

Die Enrichment Analysen von cis-meQTL assoziierten Genen ergab eine Überrepräsentation von Genen, die positionell im Bereich von GWAS Loci liegen ( $P = 5,8 \times 10^{-4}$ ). Potentielle rSNPs wurden in der GWAS Kandidaten Region von 1q31.2 (*RGS1*) und von 3p21.1 (*PRBM1*) lokalisiert. Die allelische Veränderung von Transkriptionsfaktor-Bindungsstellen durch potentielle rSNPs führt zu quantitativen Änderungen der Gentranskription oder Spleißprozessen, welche wiederum zu pathogenen Verläufen von neuropsychiatrischen Erkrankungen beitragen. Die aus diesen Studien hervorgehende Datenbank von autosomalen meQTLs, imeQTLs und eQTLs in Hirngewebe stellt eine wertvolle Quelle dar, um rSNPs zu identifizieren und deren Beteiligung an Erkrankungsprozessen aufzuklären.

## ABSTRACT

Neuropsychiatric disorders have a strong genetic predisposition, but their genetic basis remains elusive. Genome-wide association studies (GWASs) have mapped more than 2,000 susceptibility loci that were shown to increase the risk of common brain disorders. However, the majority of these susceptibility loci reside in non-coding regions and their functional consequences are unknown. The present study addresses the question whether regulatory sequence variants, affecting DNA methylation and gene expression, may be causal susceptibility alleles.

I used an integrative functional genomics approach to investigate epigenetic regulation phenomena in human hippocampal brain of 115 European patients with pharmacoresistant mesial temporal lobe epilepsy. High-density SNP genotypes were correlated with genome-wide quantitative CpG methylation and mRNA expression levels using the *Human Methylation450 array* (HM450) and the *Human HT-12 v3 array*. Subsequently, a genome-wide map of methylation quantitative trait loci (meQTLs) and expression quantitative trait loci (eQTLs) was used to dissect regulatory SNPs (rSNPs) that confer susceptibility to common brain disorders at 488 known GWAS hits ( $P < 5.0 \times 10^{-8}$ ).

This is the first meQTL study of brain tissue applying the high-density HM450 array in specimens of fresh frozen human brain tissue obtained by epilepsy surgery at large scale. Linear regression analysis of this study implementing a correction for cell-type heterogeneity, identified 19,954 (8.5% of 362k CpGs) cis-acting meQTLs at a false-discovery rate (FDR) of 5%, which is a six-fold increase compared to previous meQTL studies that all investigated postmortem brain tissue. Specifically, cis-meQTLs were strongly enriched upstream of the gene promoter region (TSS201-1500;  $P = 7.7 \times 10^{-61}$ ), highlighting the functional impact of this 5'-regulatory region that harbors binding sites of enhancers and insulators. Some of the most significant cis-meQTLs affected high-ranking candidate genes (*ADARB2*, *HDAC4*, *NAPRT1*, *MAD1L1*, *PTPRN2* and *RIMBP2*) for neurodevelopmental disorders. To explore tissue specificity, the same approach was repeated in an additional meQTL analysis of whole blood cells originating from 496 German population controls without neuropsychiatric disorders. Results show that 65% of the meQTLs in brain tissues were also present in whole blood cells (Spearman's Rank coefficient = 0.42). The present database of cis-meQTLs in brain and blood cells provides a key to select accessible epigenetic biomarkers for brain disorders in whole blood cells. The performed eQTL study identified 734 out of 31k mRNA probes at which expression levels were significantly influenced by cis-acting SNPs (FDR < 5%).

Apart from meQTL and eQTL analyses, additionally a CpG methylation to gene expression correlation analysis was performed. This represents the first systematic delineation of methylation-driven genes in fresh frozen brain tissue. Both inverse correlations (73%) and positive correlations (27%) were

observed, whereby the strongest inverse correlations were detected at *NAPRT1*, the gene encoding Nicotinate Phosphoribosyltransferase. Furthermore, the *NAPRT1*-associated meQTLs and eQTL were both genetically regulated by SNP rs9657360. The minor C allele of that very SNP was significantly associated with high methylation levels in the *NAPRT1* promoter region and simultaneously associated with low gene expression of *NAPRT1*. Both, the tumor-specific hypermethylation of a promoter CpG island as well as loss of *NAPRT1* expression have been previously proposed as predictive biomarkers for the therapy of carcinomas using *NAMPT* inhibitors. The additionally genetic risk constellation which has been identified by my approach – combining meQTLs and eQTLs to unravel the translational impact of epigenetic regulation of gene expression – is of high clinical relevance. It enables a diagnostically driven clinical strategy in tumorigenesis including the selection of patients which likely benefit from the administration of *NAMPT* inhibitors.

To dissect imprinted meQTLs (imeQTLs) exhibiting differential methylation in a Parent-of-Origin (PofO) dependent manner, the CpG methylation states of blood cells in groups of 269 individuals stratified by parentally inverse heterozygous genotypes of nearby SNPs were compared. The imeQTL analysis revealed 177 CpGs at 31 genomic loci of which 22 were previously unknown. The strongest PofO effects were observed at loci harboring neurodevelopmental genes and on chromosome 3p21.1, which is a GWAS candidate region for mood disorders. Genes at genomic loci that show imprinting effects are promising candidate genes because of their potentially monoallelic gene expression which may unmask recessive susceptibility alleles.

Enrichment analyses of genes associated with cis-meQTLs revealed an overrepresentation of genes implicated in GWAS hits of brain disorders ( $P = 5.8 \times 10^{-4}$ ). Potential rSNPs at the GWAS candidate loci 1q31.2 (*RGS1* gene locus) and 3p21.1 (*PRBM1* gene locus) were identified. The allelic alteration of transcription factor binding sites by potential rSNPs is likely to result in changes of gene transcription or splicing processes which could contribute to pathogenic pathways underlying neuropsychiatric disorders. As exemplified in this thesis, the created database of autosomal meQTLs, imeQTLs and eQTLs in brain tissue provides a valuable resource to dissect rSNPs at GWAS hits and to decipher their functional effects.

# TABLE OF CONTENTS

<b>1. Introduction</b>	<b>1</b>
<b>1.1 Neuropsychiatric disorders</b>	<b>1</b>
<b>1.2 Genetic strategies for the dissection of the genetic architecture of neuropsychiatric disorders</b>	<b>2</b>
<b>1.3 Genome- and epigenome-wide projects</b>	<b>3</b>
<b>1.4 DNA methylation</b>	<b>4</b>
<b>1.5 Quantitative trait loci (QTLs)</b>	<b>5</b>
1.5.1 Methylation quantitative trait loci (meQTLs)	6
1.5.2 Expression quantitative trait loci (eQTLs)	7
1.5.3 Imprinted methylation quantitative trait loci (imeQTLs)	7
<b>1.6 Perspectives</b>	<b>8</b>
<b>1.7 Objectives</b>	<b>9</b>
<b>2. Materials and methods</b>	<b>10</b>
<b>2.1. Study participants and surgical specimens</b>	<b>10</b>
2.1.1 Fresh frozen hippocampal brain tissue	10
2.1.2 Whole blood cells from German population controls	11
2.1.3 Whole blood cells from parent-offspring trios with GGE	11
<b>2.2 Genome-wide high-density SNP genotyping</b>	<b>12</b>
2.2.1 SNP imputing	12
<b>2.3 DNA methylation analysis</b>	<b>13</b>
2.3.1 Bisulfite treatment of genomic DNA	13
2.3.2 Array-based genome-wide assessment of CpG methylation	13
2.3.3 Assessment of signal intensities	14
2.3.4 Normalization of signal intensities	14
2.3.5 Quality control filters of CpG methylation profiles	16
<b>2.4 mRNA expression analysis</b>	<b>16</b>
<b>2.5 Statistical quantitative trait loci (QTL) analyses</b>	<b>17</b>
2.5.1 meQTL and eQTL analyses in hippocampal brain tissue	18
2.5.2 Methylation-driven gene expression in hippocampal brain tissue	20
2.5.3 meQTL analysis in blood cells	20
2.5.4 Imprinting methylation QTL (imeQTL) analysis in blood cells	21

<b>2.6 Exploration of the genomic features of QTL analyses.....</b>	<b>24</b>
2.6.1 Genomic distribution of cis-meQTLs .....	24
2.6.2 Enrichment analysis of cis-meQTLs and cis-eQTLs .....	24
<b>2.7 Exploration of the role of QTL-associated rSNPs in common neuropsychiatric disorders.....</b>	<b>24</b>
<b>3. Results .....</b>	<b>26</b>
<b>3.1 Methylation QTL analyses .....</b>	<b>26</b>
3.1.1 Cis-meQTL analysis in hippocampal brain tissue .....	26
3.1.2 Comparative cis-meQTL analysis of brain cells and whole blood cells.....	32
3.1.3 Trans-meQTL analysis in hippocampal brain tissue.....	37
3.1.4 Imprinting methylation QTL analysis in blood .....	38
<b>3.2 Expression QTL analyses .....</b>	<b>45</b>
3.2.1 Cis-eQTL analysis in hippocampal brain tissue .....	45
3.2.2 Trans-eQTL analysis in hippocampal brain tissue.....	48
<b>3.3 Delineation of methylation-driven gene expression in hippocampal brain tissue .....</b>	<b>49</b>
3.3.1 Correlation analysis of CpG methylation with gene expression.....	49
<b>3.4 Exploration of the role of QTL-associated rSNPs in common neuropsychiatric disorders.....</b>	<b>54</b>
3.4.1 Co-occurrence of cis-acting meQTLs and GWAS hits .....	54
3.4.2 Co-occurrence of cis-acting eQTLs and GWAS hits.....	56
<b>4. Discussion .....</b>	<b>57</b>
<b>4.1 Autosomal map of cis-meQTLs in human hippocampal brain tissue .....</b>	<b>58</b>
4.1.1 Gene-centric distribution of cis-meQTLs.....	59
4.1.2 Enrichment analysis of cis-meQTLs .....	60
4.1.3 Tissue-specificity of cis-meQTLs in hippocampal brain cells.....	60
<b>4.2 Autosomal map of trans-meQTLs in human hippocampal brain tissue .....</b>	<b>61</b>
<b>4.3 Genome-wide assessment of imprinting meQTLs in GGE parent-offspring trios.....</b>	<b>61</b>
<b>4.4 Autosomal map of cis-eQTLs in human hippocampal brain tissue .....</b>	<b>63</b>
<b>4.5 Autosomal map of trans-eQTLs in human hippocampal brain tissue .....</b>	<b>64</b>
<b>4.6 Delineation of methylation-driven gene expression in hippocampal brain tissue .....</b>	<b>64</b>
4.6.1 Functional implications of meQTL and eQTL effects on <i>NAPRT1</i> expression.....	65
<b>4.7 Dissection of rSNPs involved in common neuropsychiatric disorders .....</b>	<b>66</b>
<b>4.8 Outlook .....</b>	<b>69</b>



<b>5. References .....</b>	<b>72</b>
<b>6. Appendix .....</b>	<b>87</b>
6.1 Clinical parameters of 115 mTLE patients .....	87
6.2 Confirmation analyses of cis-meQTLs and cis-eQTLs in hippocampal brain tissue and whole blood cells .....	87
6.3 Master regulatory loci of trans-meQTL SNPs .....	91
6.4 QQ-plots of QTL analyses .....	93
6.4.1 QQ-plots of meQTL analysis .....	93
6.4.2 QQ-plots of imeQTL analysis .....	93
6.4.3 QQ-plot of eQTL analysis .....	94
<b>7. Danksagung.....</b>	<b>95</b>
<b>8. Erklärung.....</b>	<b>97</b>
<b>9. Curriculum vitae.....</b>	<b>98</b>

## LIST OF FIGURES

<i>Figure 2-1: Normalization and transformation effects.</i> .....	15
<i>Figure 2-2: Scree plots.</i> .....	19
<i>Figure 2-3: PC variance for four levels of month of methylotyping.</i> .....	21
<i>Figure 2-4: Determination of parental origin and PofO specific association testing for a hypothetical SNP.</i> .....	22
<i>Figure 3-1: Manhattan plot for cis-meQTLs identified in hippocampal brain tissue samples.</i> .....	27
<i>Figure 3-2: Functional localization of genomic regions of meQTLs in hippocampal brain tissue.</i> .....	28
<i>Figure 3-3: P-value scatter plot of hippocampal brain tissue and blood cells.</i> .....	33
<i>Figure 3-4: Functional localization of CpGs in GM12878.</i> .....	36
<i>Figure 3-5: Circos plot of 115 hippocampal brain tissue for trans-meQTLs.</i> .....	37
<i>Figure 3-6: Manhattan plot for cis-imeQTLs of 269 trio offspring.</i> .....	39
<i>Figure 3-7: Quantification of the allele-specific imeQTL effect.</i> .....	42
<i>Figure 3-8: Manhattan plot for cis-eQTLs identified in 115 hippocampal brain tissue samples.</i> .....	45
<i>Figure 3-9: DNA methylation and gene expression of NAPRT1 plotted by genotypes of rs9657360.</i> .....	53
<i>Figure 3-10: NAPRT1 region.</i> .....	54
<i>Figure 4-1: Chromosome 1q31.2 region.</i> .....	67
<i>Figure 6-1: QQ-plot of CpG methylation in 115 hippocampal brain tissue samples and 496 blood cell samples.</i>	93
<i>Figure 6-2: QQ-plot of CpG methylation in 269 parent-offspring trios (heterozygote test).</i> .....	93
<i>Figure 6-3: QQ-plot of CpG methylation in 269 parent-offspring trios (maternal and paternal imeQTL analysis).</i> .....	94
<i>Figure 6-4: QQ-plot of mRNA expression in 115 hippocampal brain tissue samples.</i> .....	94

## LIST OF TABLES

<i>Table 2-1: Survey of the data sets integrated for QTL analyses.</i> .....	17
<i>Table 2-2: Parameters and counts of clinical parameters for hippocampal brain tissue.</i> .....	18
<i>Table 2-3: Spearman rank correlation rhos and nominal P-values of pairs of parameters.</i> .....	19
<i>Table 2-4: Known imprinted clusters and associated allelically methylated regions.</i> .....	23
<i>Table 3-1: Distribution of meQTLs within different gene-centric regions in hippocampal brain tissue.</i> .....	28
<i>Table 3-2: Gene-set enrichment analyses of cis-meQTLs in hippocampal brain tissue.</i> .....	29
<i>Table 3-3: Overlaps with comparable human meQTL studies.</i> .....	30
<i>Table 3-4: Top hits of meQTLs in hippocampal brain tissue.</i> .....	31
<i>Table 3-5: Significant cis-meQTLs of high-ranking candidate genes in hippocampal brain tissue.</i> .....	32
<i>Table 3-6: Frequency of cis-meQTLs for different significance thresholds for hippocampal brain tissue and blood cells.</i> .....	33
<i>Table 3-7: Brain tissue-specific cis-meQTLs not found in blood cells.</i> .....	34
<i>Table 3-8: Genomic distribution of cis-meQTL associated CpGs in hippocampal brain tissue and blood cells.</i> ....	35
<i>Table 3-9: Top 10 imeQTLs and specification of their preferential parental expression.</i> .....	40
<i>Table 3-10: Eight imeQTLs with specification according to their parental expression loci consistent with Crowley et al. (2015).</i> .....	44
<i>Table 3-11: Cis-eQTL overlaps with comparable human eQTL studies.</i> .....	46
<i>Table 3-12: Top hits of cis-eQTLs in hippocampal brain tissue.</i> .....	47
<i>Table 3-13: Master regulatory sites – trans-eQTL SNPs with simultaneous impact on the expression of two genes.</i> .....	48
<i>Table 3-14: Genomic distribution of CpGs influencing gene expression in gene features in hippocampal brain tissue.</i> .....	50
<i>Table 3-15: Top hits of CpG methylation-driven gene expression in hippocampal brain tissue.</i> .....	52
<i>Table 3-16: GWAS lead SNPs in strong LD with cis-acting meQTLs in hippocampal brain tissue.</i> .....	55
<i>Table 6-1: Distribution of clinical parameters within sample group.</i> .....	87
<i>Table 6-2: Reference meQTL and eQTL studies exploited for confirmation analyses.</i> .....	89
<i>Table 6-3: Master regulatory sites – trans-meQTL SNPs with simultaneous impact on the methylation of at least four genes.</i> .....	91

# 1. INTRODUCTION

## 1.1 Neuropsychiatric disorders

Neuropsychiatric disorders such as epilepsy, schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder, attention deficit-hyperactivity disorder and substance use disorders represent 13% of the global burden of diseases, surpassing cardiovascular disease and cancer (World Health Organization 2008; Collins et al. 2011). They cause enormous personal and social burdens (Collins et al. 2011; Labrie et al. 2012) and have a lifetime prevalence that ranges from 0.1% for autism spectrum disorder to approximately 1% for schizophrenia and up to 24% for nicotine dependence (Sullivan et al. 2012). Family studies, including twin and adoption studies provide consistent evidence that genetic factors contribute to the risk of neuropsychiatric disorders (Kendler et al. 2005). These family studies assessed heritability estimates ranging from 37% for major depressive disorder to more than 80% for schizophrenia and bipolar disorder (Cardno et al. 1999; Sullivan et al. 2012) and show empirical evidence of a shared genetic etiology across neuropsychiatric disorders (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013). Overall, almost all neuropsychiatric traits display a complex genetic predisposition (Labrie et al. 2012; Gelernter 2015) and only a small fraction follows mendelian inheritance patterns (Lander & Schork 1994). There is no common genetic architecture for the set of complex psychiatric traits. Generally, the genetic architecture of the vast majority of neuropsychiatric disorders is composed by highly polygenic and heterogeneous factors including multiple risk alleles, epistatic and epigenetic effects. Risk alleles can individually be common or rare, and can include, for example, single nucleotide polymorphisms (SNPs) and copy number variants (CNVs, Gelernter 2015).

Common neuropsychiatric disorders are among the most complex and poorly understood conditions affecting the human body. In recent years there have been major research efforts to improve our understanding of their complex genetic predisposition. Initial studies included genetic linkage studies, candidate gene association studies and targeted sequencing studies, which follow a hypothesis-driven approach. Although these “traditional” approaches have identified a few susceptibility genes, these studies did not succeed to identify common major susceptibility loci. The scenario started to change with the advent of genome-wide association studies (GWASs) which allowed a more systematic, hypothesis-free exploration of the genetic basis of neuropsychiatric disorders. The hypothesis-free approach of a GWAS offered the opportunity to overcome difficulties and obstacles enforced upon the incomplete understanding of the pathophysiology of the disease (Kitsios & Zintzaras 2009).

## 1.2 Genetic strategies for the dissection of the genetic architecture of neuropsychiatric disorders

A GWAS is usually designed as a case-control association study, in which allelic variation of SNPs is compared between individuals with a particular disease and unaffected individuals. The strategy of the GWAS approach is mainly directed on the discovery of common variants conferring low/moderate risks following the “common disease/common variant” hypothesis (Reich & Lander 2001). This hypothesis predicts that the genetic risk for common diseases will often be due to disease-predisposing alleles with relatively high frequencies. In the last few years, a huge number of GWASs have been performed to dissect the genetic basis of many different complex diseases and traits. GWASs investigating the genetic architecture of neuropsychiatric disorders have identified many susceptibility variants (Psychiatric GWAS Consortium Bipolar Disorder Working Group 2011; Sullivan et al. 2012). The National Human Genome Research Institute (NHGRI) catalog of published GWAS hits harbors more than 15,000 SNPs associated with human diseases, of which more than 2,000 SNPs are associated with neuropsychiatric disorders (Welter et al. 2014). However, the identified genetic variants characteristically explain only a modest proportion of the total heritability of these traits. This has led to the common question, how the “missing heritability” of complex diseases can be explained (Eichler et al. 2010). One plausible explanation is that most of GWASs have SNPs with minor allele frequencies of more than 5% implying that many rare variants has not been ascertained by current GWASs. According to the alternative “common disease/rare variant” hypothesis, complex traits are caused collectively by multiple rare DNA sequence variants, each with moderate to high penetrance (Marian 2012). To test this hypothesis, gene sequencing studies have been carried out by next-generation sequencing (NGS) methods: either as target candidate gene studies and whole exome or genome studies (Bamshad et al. 2011). Although NGS has identified many deleterious gene mutations, the current findings do not close the gap of the missing heritability (Petronis 2010; Liu & Leal 2012). One problem arises by the difficulty to interpret the biological effect of variants identified by GWASs (Westra & Franke 2014). Typically, a candidate region identified by a GWAS contains more than one gene and multiple sequence variants form a linkage disequilibrium (LD) block (Albert & Kruglyak 2015). Although variants that alter coding sequences are obvious candidates, the majority of loci identified in GWASs is found in non-coding regions and probably affects regulatory elements (Maurano et al. 2012). It is difficult to unequivocally identify the causal variant for each locus by using the traditional fine-mapping methods. Several lines of evidence suggest that many sequence variants in non-coding regions influence regulatory processes controlling mRNA transcription.

### 1.3 Genome- and epigenome-wide projects

I presume that the understanding of the molecular mechanisms underlying common diseases will be improved by an integrated functional genomics strategy. By 2004, large-scale genome projects already indicate that genome sequences alone cannot explain the whole diversity of life, because they are very similar within and across species (cf. Ptashne et al. 2010). Instead, epigenetics may explain how these similar genetic codes are differentially expressed in different cell-types within different environmental conditions and at different times (Ptashne et al. 2010). Epigenetics refers to heritable changes in gene expression caused by alterations in DNA methylation and chromatin structure (Henikoff & Matzke 1997). Epigenetic factors have been linked to developmental processes and play a critical role in normal cellular differentiation during embryogenesis (Li 2002). They have been implicated to play an important role in several human diseases, including cancer and neuropsychiatric disorders (Dalton et al. 2014). Epigenomics is the science of functional elements regulating gene expression in cells. The epigenome consists of the complete collection of epigenetic marks, such as DNA methylation, histone modifications and non-coding RNAs that exist in a cell at any given point (Romanoski et al. 2015).

Ongoing projects such as ENCODE (Encyclopedia of DNA Elements), the International Human Epigenome Consortium (Bae 2013) and the US National Institutes of Health Epigenomics Roadmap are generating cell-specific reference data sets that provide a basis for delineating the complex interplay between epigenomic processes and the transcriptome. The ENCODE project aimed to catalog the regulatory elements in human cells and to study the epigenomic signatures of cells which are grown in culture (ENCODE Project Consortium 2004, 2007, 2012). Additional approaches and projects further benefit from the ENCODE project (Civelek & Lusk 2014): Systematic maps of transcription factor binding sites and chromatin modifications have been generated as have databases and web-tools such as GWAS3D (Li et al. 2013), which help to automate some of the processes involved. In addition, the Roadmap Epigenomics Project extends the ENCODE project and aims to elucidate how epigenetic processes contribute to human biology and disease (Kundaje et al. 2015). The researchers have linked epigenomic signatures to the corresponding genetic information, producing reference epigenomes for several human tissues and cell-types. The result is a comprehensive landscape of epigenomic elements regulating gene expression in the human body (Romanoski et al. 2015). Kundaje and co-workers (2015) enable insights into the epigenomic landscape, its dynamics across cell-types or tissues and development. Their epigenomic data sets, regulatory annotation and integrative analyses have resulted in the most comprehensive map of the human epigenomic landscapes so far and cover the largest collection of primary cells and tissues (Kundaje et al. 2015). Some of the most widely studied mechanisms of epigenetic regulation include

DNA methylation, histone modifications as well as non-coding RNAs. DNA methylation is the most stable of all epigenetic modifications and the most studied epigenetic modification to date (Beck & Rakyant 2008; Dalton et al. 2014).

## 1.4 DNA methylation

DNA methylation is the only epigenetic mark for which a detailed mechanism of mitotic inheritance has been described (Bird 2002). The most common form of DNA methylation in vertebrates is 5-methylcytosine (5mC), which arises by the addition of a methyl group to cytosine nucleotides (C) and affects 70 to 80% of CpGs in the human genome (Ehrlich et al. 1982). High levels of 5mC in promoter regions that are CpG-rich are strongly associated with transcriptional repression, whereas genomic regions that are CpG-poor exhibit a more complex relationship between DNA methylation and transcriptional activity (Jones 2012). DNA methylation has become an important tool in the emerging systems approach to explore and better understand genome function in health and disease. DNA methylation has been extensively studied for its role in several biological processes such as for example genomic imprinting (Barlow 2011) and characteristic changes in DNA methylation have been reported for cancer (Baylin & Jones 2011; Shames et al. 2013) and several other diseases (Feinberg 2007).

Over the past decade, numerous approaches have been evolved for methylation analysis. Recent advances in NGS and microarray technology allow mapping of DNA methylation at a high genome-wide resolution and in a large number of samples (Laird 2010). These new methods create enormous opportunities for research of the epigenome (Bock 2012). Key advantages of the NGS technology are its comprehensive genomic coverage, high quantitative accuracy and excellent reproducibility (Bock 2012). But the most widely-used approach for epigenome-wide DNA methylation analysis is the *Illumina Infinium Human Methylation450 BeadChip microarray* (HM450). The HM450 microarray offers a powerful tool to assess DNA methylation across the genome. The genomic coverage of the Infinium assay is more limited than that of most bisulfite-sequencing based methods, but the compatibility with existing genotyping pipelines, the lower per-sample cost compared with whole-genome bisulfite sequencing and the simpler analysis and interpretation of methylation data makes it an attractive approach for large-scale sample collections (Bock 2012; Morris & Beck 2015). Verification and validation are usually done using locus-specific DNA methylation assays at a small number of CpGs in many samples to reduce the cost of studying large validation cohorts. For this purpose the most popular method is pyrosequencing (Tost & Gut 2007; Potapova et al. 2011).

## 1.5 Quantitative trait loci (QTLs)

In 2001, Jansen and Nap introduced the concept for a strategy, coined 'genetical genomics' to identify which genes are regulated by genetic variation (Jansen & Nap 2001). By correlating genetic variants with intermediate molecular quantitative traits, such as methylation levels or gene expression levels, it is possible to identify quantitative trait loci (QTLs). To identify the variants that influence DNA methylation or gene expression, two types of data must be collected from each individual. First, the genotype data of each individual are required. Second, in each individual the DNA methylation is measured using an array platform considering genome-wide patterns or the expression of each gene in the genome is measured using either expression microarrays or RNA sequencing. The QTLs are then identified by comparing the genotypes with the methylation or expression levels using a statistical association test. During that course, individuals are grouped according to the allele they carry. A significantly higher methylation or expression level for a gene in one group than in the other group suggests that the variant (or another variant in LD) influences the methylation or expression of this gene. The test is repeated at every DNA variant in the genome which results in a genome scan for methylation quantitative trait loci (meQTLs) or expression quantitative trait loci (eQTLs) for this gene (Albert & Kruglyak 2015).

QTLs can be divided into those that have local effects (cis-QTLs), meaning that the genetic variant is located near the genomic probe and those with distant effects (trans-QTLs), meaning that the genetic variant is located further away from the genomic probe (e.g. >10 Mbps apart or on a different chromosome; Januar et al. 2015). Recent studies have examined the association between genetic variants and the quantitative traits in both cis and trans, of which cis-acting QTLs predominate (Gibbs et al. 2010; Zhang et al. 2010; Bell et al. 2011; Numata et al. 2012; Westra et al. 2013; Ramasamy et al. 2014; Schramm et al. 2014). Distant QTLs have smaller effects sizes and seem also to be more tissue-specific than local QTLs which are often conserved among various tissues (Petretto et al. 2006; van Nas et al. 2010; Fairfax et al. 2012), which further complicates the detection of those trans-acting QTLs. Overall, previous reported studies show that the genetic regulation of gene expression is complex and differs widely across cell-types and tissues, especially for genetic variants that are disease-associated.

The choice of tissue type is a major challenge that distinguishes genetic and epigenetic studies. The tissue type is largely irrelevant for genetic studies of germline genetic variation in contrast to epigenetic studies where often the primary disease- or exposure-relevant target tissue (or cell-type) is available on a limited scale, e.g. brain tissue (Mill & Heijmans 2013). So far, several QTL studies have investigated human brain tissue in neuropsychiatric disorders (Gibbs et al. 2010; Zhang et al. 2010; Numata et al. 2012; Gamazon et al. 2013; Kim et al. 2014; Numata et al. 2014; Ramasamy et al.



2014; Smith et al. 2014), but all of them performed QTL analyses in postmortem brain tissue. Postmortem studies of brain tissue have several limitations: postmortem brain tissue samples have restrictions with respect to mRNA conservation (Bray et al. 2003; Webster et al. 2009) and methods of tissue preservation alter the quality of biomolecules obtained (Januar et al. 2015). A delay of the autopsy of postmortem brain tissue alters DNA methylation profiles from baseline (Miller-Delaney et al. 2015). Additionally, postmortem studies of brain tissue are critical for understanding the disease aetiology (Januar et al. 2015). On the one hand these aspects indicate the need for fresh frozen human brain tissue obtained from living patients. On the other hand it is necessary to explore the extent to which easily accessible cells obtained from tissues such as whole blood can be used to address questions about epigenomic variation in inaccessible tissues such as the brain.

Known and unknown confounders can contribute significantly to the dataset variance in quantitative epigenomic high throughput analyses, so the integration of major confounding factors are required to perform a successful analysis of QTL data (Januar et al. 2015). Potential confounders include age, gender and ethnic diversity. Further, medical histories of participants, such as antiepileptic treatments, antipsychotic or antidepressant medications are useful and necessary. Studies that use postmortem brain tissue also have to consider confounders such as antemortem history, medication use and cause of death or postmortem delay. Another important issue when performing QTL studies is cellular heterogeneity. To overcome this issue, methods of adjusting for DNA methylation variability associated with cell composition differences have been developed for brain as well as for blood cells and can be incorporated in statistical models for adjustment (Houseman et al. 2012; Guintivano et al. 2013; Jaffe & Irizarry 2014). Cellular heterogeneity provides an important issue but only a few current studies start to consider this confounding factor.

#### 1.5.1 Methylation quantitative trait loci (meQTLs)

The association between genetic variants and the disorder could be mediated via its ability to influence DNA methylation (meQTLs). Because sequence variation can directly influence DNA methylation in cis (Schalkwyk et al. 2010) and evidence already provides an enrichment of meQTLs at loci for several disorders identified by GWAS (Numata et al. 2012; Gamazon et al. 2013; Smith et al. 2014), it is likely that interpretation of genetic data can be largely improved by integrating allele-specific epigenetic information into the analyses (Meaburn et al. 2010).

Early meQTL studies focused on methylation data from relatively few CpGs showing a strong bias towards promoter regions. The comprehensive array platform considering genome-wide patterns enable recent studies to cover much more meQTLs. The present study uses the popular platform of the HM450 array that is the best known platform capable of high-throughput work. However,

technical artifacts need to be taken into account, especially where Infinium probes overlap with positions of known DNA variants (Barrow & Byun 2014). When using microarray platforms and studying meQTLs regarding SNPs whose genotype correlates with DNA methylation, potential SNP artifacts represent a substantial challenge (Barrow & Byun 2014). Removing of all SNP-associated probes from the analysis process would not be appropriate as 56% of the probes on the Infinium array contain SNPs. Definite factors such as the distance of an SNP within the probe and the minor allele frequency within the ethnicity of the study population should be considered. It is important to give careful consideration to what parameters should be set and potential confounding factors should be subsequently excluded without minimizing the HM450 probe set excessively. Up to date only few meQTLs have been reported to change gene expression (Gibbs et al. 2010; Gamazon et al. 2013; Gutierrez-Arcelus et al. 2013). The present study addresses this promising issue.

#### 1.5.2 Expression quantitative trait loci (eQTLs)

Expression quantitative trait loci (eQTLs) are regions of the genome which contain DNA sequence variants that influence the expression level of one or more genes (Albert & Kruglyak 2015). The genetics of expression variation of single genes has been studied for a long time, at least since 1962 (Schwartz 1962). Maps of eQTLs are being built in large-scale studies in humans for different cell-types or tissues, such as blood cells and brain tissue (Gibbs et al. 2010; Westra et al. 2013; Kim et al. 2014; Ramasamy et al. 2014; Schramm et al. 2014).

Beyond the hitherto description of large eQTL catalogs, the understanding of the role of regulator variation is currently being expanded in two directions. Typically, eQTLs were identified as 'loci', statistical associations between regions of the genome and the expression of genes. Recently, eQTLs are being used to identify the causal variants and their molecular mechanism of action. One immediate application of eQTLs lies in the interpretation of GWAS risk loci. Large eQTL studies can help to prioritize potential causal variants among multiple polymorphisms within the GWAS candidate regions (Albert & Kruglyak 2015).

#### 1.5.3 Imprinted methylation quantitative trait loci (imeQTLs)

Imprinting meQTL (imeQTL) analyses imply the association of SNP genotypes of defined parental origin with methylation levels. Genes that show a parental bias in methylation will undergo differential regulatory effects from the paternal and maternal alleles. As standard association studies treat both alleles equally, they are unable to detect effects such as imprinting in which the two alleles are differentially regulated.

Genomic imprinting is a form of epigenetic variation whereby Parent-of-Origin (PofO) specific epigenetic modifications are inherited by offspring, resulting in mono-allelic gene expression.

Aberrations in normal imprinting patterns have been linked to congenital disorders such as Prader-Willi / Angelman syndromes (Nicholls et al. 1989; Clayton-Smith 1993), and Beckwith-Wiedemann syndrome (Reik et al. 1995). This finding is consistent with the fact that many imprinted genes have prominent roles in growth and development (Georgiades et al. 2001; Lambertini et al. 2012). Recent GWASs have taken into account potential PofO effects and have uncovered important contributions of imprinting to common complex diseases (Kong et al. 2009). In addition to genetic studies, epidemiological data in many common diseases, including multiple sclerosis (Chao et al. 2010), asthma (Carroll et al. 2005), and bipolar disorder (McMahon et al. 1995; Kornberg et al. 2000) further imply that PofO and imprinting-mediated effects on disease may be pervasive. But despite evidence for the significant impact of imprinting in genome function and disease, catalogs of imprinted genes in the human genome are almost certainly incomplete. At present, there are >120 known imprinted genes in mice; yet in humans, only approximately 85 have been confirmed ([www.geneimprint.com](http://www.geneimprint.com); Wei et al. 2014), 10% of which have been identified since 2012 (Barbaux et al. 2012). The imprinting meQTLs of the present study may help to identify additional imprinted genes which display both allele specific CpG methylation and genomic imprinting. The imeQTL study of this work will help to identify PofO dependent susceptibility effects that may improve the analytical power of GWASs to dissect the complex genetic architecture of common brain disorders and may explain a substantial fraction of the missing genetic heritability.

## 1.6 Perspectives

The field of epigenetics is expanding at an exponential rate and projects like the Roadmap Epigenomics Consortium (Romanoski et al. 2015) makes huge efforts to fill the gap of epigenomic studies and its association with human disease. Systems genetics studies are aiming to identify the role of regulatory variation in complex traits by integrating intermediate phenotypes, such as CpG methylation, transcript, protein or metabolite levels (Civelek & Lusis 2014; Albert & Kruglyak 2015). Recent reviews present a survey about the current knowledge of the molecular architecture of complex traits and are useful for the identification of genes and pathways that underlie common human diseases. However, our understanding of the exact mechanisms by which epigenetic changes modify a phenotype, particularly in association with common neuropsychiatric disorders in humans, is still limited. Current research in this area is limited by a number of factors including difficulties in establishing functional causality, tissue heterogeneity and other confounding factors such as spatial and temporal effects (Januar et al. 2015). The understanding of the role of the methylome in regulating human health and disease is yet at the beginning. This is particularly true in the field of

neuropsychiatric disorders for which access to pathogenic brain tissue is usually not available in living patients.

## 1.7 Objectives

The present study aims to get a deeper insight into the role of regulatory variation in complex traits, especially in neuropsychiatric disorders, with a particular interest in mapping the effects of common genetic variants on gene expression and DNA methylation. An integrative functional model of genomics and epigenomics will be optimal for understanding the etiological pathways to common disorders with complex genetic predisposition, such as neuropsychiatric disorders. An integrative functional approach will further help to trace molecular changes through layers of biological information to the disease outcome.

Functional genomics studies of common brain disorders are difficult because human brain tissue is usually not available. Temporal lobe epilepsy (TLE) offers a unique opportunity to obtain bioptic brain specimens when epilepsy surgery is necessary to control pharmaco-resistant seizures (Grote et al. 2015). I used this intriguing resource to perform the first meQTL and eQTL study of bioptic human brain tissues of pharmaco-resistant mesial TLE (mTLE) patients. Additionally, the approach of the present study allows a systematic examination of the influence of cis-meQTLs on gene expression to delineate methylation-driven genes in fresh frozen brain tissue.

Currently, there are no QTL studies, which consider cellular heterogeneity among different brain regions and focus on neuropsychiatric disorders. Thus, the present study considers two major aspects of epigenetic studies including tissue relevance, as I can use tissue in which the given disease emerges, and cellular heterogeneity.

An intra-individual cross-tissue study concluded that between-tissue variation in DNA methylation greatly exceeds inter-individual differences for any one tissue but suggested that some inter-individual variation in DNA methylation may be correlated between brain regions and blood (Davies et al. 2012). To explore tissue-specificity of the cis-meQTLs detected in hippocampal brain tissue, meQTL analyses in whole blood cells of 496 German population controls without neuropsychiatric disorders were performed, using the same array platform and analytical procedures. This may offer the opportunity to dissect accessible biomarkers in blood DNA sources for common brain disorders.

One major achievement of this study will be a public release (publication in preparation) of a database reporting the complete meQTL and eQTL findings of this thesis. These genome-wide maps of meQTLs and eQTLs will improve the prospects to elucidate the genetic mechanisms determining DNA methylation and gene expression in common brain disorders.

## 2. MATERIALS AND METHODS

Genome-wide methylation quantitative trait loci (meQTL) and expression quantitative trait loci (eQTL) analyses were performed in fresh frozen hippocampal brain tissue of European patients with pharmaco-resistant mesial temporal lobe epilepsy (mTLE). Individual high-density SNP genotypes were correlated with individual quantitative methylation states and individual expression levels. To delineate accessible epigenetic biomarkers, genome-wide maps of meQTLs were also generated in whole blood cells from German population controls using the same methylation microarray. In addition, blood cell DNA samples of parent-offspring trios with Genetic Generalized Epilepsy (GGE) were investigated to screen for differentially methylated parental genomic regions (imprinting meQTLs, imeQTLs).

### 2.1. Study participants and surgical specimens

#### 2.1.1 Fresh frozen hippocampal brain tissue

The present study included 117 pharmaco-resistant mTLE patients of European descent who underwent surgical treatment in the Epilepsy Surgery Program at the University of Bonn Medical Center (Wiebe et al. 2001). In all patients, presurgical evaluation using a combination of noninvasive and invasive procedures revealed that seizures originated in the mesial temporal lobe (Kral et al. 2002). Surgical resection of the hippocampus was clinically indicated in every case (Pernhorst et al. 2011). Informed written approval was obtained from all patients and procedures were in accordance with the Declaration of Helsinki and approved by the local ethics committee.

All fresh frozen hippocampus tissue samples were from identical regions of the hippocampus. Fresh frozen sections were analyzed according to international standards and the diagnostic classification was established by an experienced neuropathologist according to international criteria (Becker et al. 2003; Blumcke et al. 2007). The majority (> 60%) of the hippocampi specimens displayed Ammon's horn sclerosis (AHS; segmental neuronal cell loss and concomitant astrogliosis and microglia activation). A smaller proportion of the specimens showed predominantly lesional alterations such as cortical dysplasia or tumors. Up to five 20 µm thick tissue sections were used for the preparation of genomic DNA and mRNA. DNA was isolated from tissue specimens using the *DNeasy Blood and Tissue Kit* (Qiagen, Hilden, Germany), according to the manufacturer's protocol (Schonberger et al. 2009). mRNA was isolated from tissue specimens using the *Dynabeads mRNA Direct Micro Kit* (Dyna, Oslo, Norway) following the manufacturer's protocol (Fassunke et al. 2008). Complementary DNA

(cDNA) was synthesized by reverse transcription of total mRNA using the *RevertAid First-Strand cDNA Synthase Kit* (Fermentas, St. Leon-Rot, Germany) according to the manufacturer's protocol.

All 117 mTLE patients fulfilled the inclusion criteria by having mRNA-expression data, DNA methylation data and genotypes in sufficient quality. Two patients were excluded because of their outstanding disease type (Rasmussen's encephalitis). Conclusively, analyses have been performed on 115 hippocampal brain tissue samples (60 males, 55 females; range of age of seizure onset: 1-47 years, average age: 11.1). The mTLE patient group was clinically characterized with respect to seizure manifestation. The clinical parameters of the mTLE patients are summarized in Appendix, Table 6-1.

#### 2.1.2 Whole blood cells from German population controls

A population-based cohort of 498 unrelated German population controls without neuropsychiatric disorder (273 males, 225 females; age range: 54 – 74 years, average age: 57.4) was collected from the Western regions of Germany (Ruhr area) within the framework of the cardiovascular longitudinal Heinz Nixdorf RECALL study (HNR; Schmermund et al. 2002). Of the 498 population controls two individuals did not fulfill the inclusion criteria having DNA methylation data and genotypes in sufficient quality. Hence, analyses have been performed on 496 population controls.

#### 2.1.3 Whole blood cells from parent-offspring trios with GGE

Epilepsy patients of European ancestry with common GGE syndromes, including genetic absence epilepsies (GAE), juvenile myoclonic epilepsy (JME) and epilepsies with generalized tonic clonic seizures alone (EGTCS) exhibiting generalized spike and wave discharges in their resting electroencephalogram (gsw-EEG), were recruited in a multi-center effort from the European EPICURE Project (<http://www.epicureproject.eu>). The diagnostic classification of GGE syndromes were prepared according to EPICURE guidelines and standardized phenotyping protocols (<http://portal.ccg.uni-koeln.de/ccg/research/epilepsy-genetics/sampling-procedure/>; ILAE 1989; Nordli 2005; Berg et al. 2010). Individuals with a history of major psychiatric disorders (autism spectrum disorder, schizophrenia or affective disorder) or severe intellectual disabilities were excluded (Trucks 2013). In total, 269 parent-offspring trios of European origin with offspring affected by GGE were available for imeQTL analysis (103 male trio children, 166 females; range of age of seizure onset: 2 – 21 years, average age: 9.8). Trios were recruited from different European countries or countries with European ancestry, including Australia (N = 69), Bulgaria (N=2), Denmark (N = 20), Germany (N = 14), Italy (N = 138), and Turkey (N = 26). The offspring trios were affected by the following GGE syndromes: 175 GAE, 92 JME, and 2 EGTCS alone. All study participants gave informed consent according to the regulations at their local institutional review boards.

## 2.2 Genome-wide high-density SNP genotyping

Genome-wide high-density SNP genotyping was carried out by using different SNP genotyping arrays for the different study cohorts. For the mTLE patients SNP genotyping was performed using the *Illumina HumanHap550 SNP array* (550k SNPs; Illumina, San Diego, CA, USA). For the HNR controls two genotyping arrays, *Illumina HumanOmniExpress 12 v1.1* and *Illumina HumanOmni1 Quad v1.0* (overlap of 539k SNPs; Illumina, San Diego, CA, USA) were combined. For the GGE offspring trios the *Affymetrix Axiom Genome Wide Human genotyping array* (567k SNPs; Affymetrix, Santa Clara, CA, USA) was used. To ensure a high accuracy of the SNP genotype calls, several quality filters were applied for the individual array and the single SNP. Exclusion criteria for SNP arrays were: call rate per array < 97%, and autosomal heterozygosity rate > 29%. Exclusion criteria for SNPs were: i) non-autosomal position, ii) missing hg19 annotation, iii) call rate per SNP < 97%, and iv) European minor allele frequency (MAF) < 5% from the 1000 Genomes Project. In addition, the trio pedigree structure and the relationship of all trio members was checked with the *PedigreeExplorer* (<http://pedigreeexplorer.meb.uni-bonn.de/>) and by IBD-estimation and was further visualized with GRR (graphical representation of relationship errors; Abecasis et al. 2001). All quality control (QC) procedures were carried out using Plink, version 1.9 (<http://pngu.mgh.harvard.edu/purcell/plink/>, Purcell et al. 2007).

### 2.2.1 SNP imputing

SNP imputing is a useful method that can detect causal variants that use the linkage disequilibrium (LD) structure in a genomic segment to infer the alleles of SNPs which are not directly genotyped (Marchini et al. 2007). The pre-phasing based imputation was carried out using a combination of the programs SHAPEIT2 and IMPUTE2 (Howie et al. 2012). Imputation with IMPUTE2 was based on the reference panel: 1000 Genomes Phase I release of NCBI build 37 (hg19). The imputed SNP genotypes were additionally quality filtered using SNPTTESTv2 (Marchini et al. 2007). Imputed SNPs were excluded from further analyses according the following criteria: i) SNPTtest info quality value < 0.9, ii) MAF < 5%, iii) missing data proportion > 3% using Plink 1.9. Furthermore, a LD-based SNP pruning was performed, considering a window of 50 SNPs, followed by a LD calculation between each pair of SNPs in the window and removal of one SNP of SNP pairs with a LD  $r^2 > 0.8$ .

## 2.3 DNA methylation analysis

### 2.3.1 Bisulfite treatment of genomic DNA

All methylation profiling technologies of Illumina are based on genotyping bisulfite-converted DNA. The *EZ DNA Methylation Kit* from Zymo Research (Zymo Research, Irvine, CA, USA) was used for bisulfite treatment of genomic DNA samples. The kit is based on the divergent reaction of unmethylated vs. methylated cytosine and sodium bisulfite: unmethylated cytosine is converted into uracil while methylated cytosine is protected and remains cytosine (Zilberman & Henikoff 2007). Following PCR, the converted uracil nucleotides will be detected as thymine.

A standardized bisulfite conversion protocol was used. Genomic DNA was applied according to the manufacturer's protocol of the *Zymo EZ DNA Methylation kit* (#D5001). Alternative incubation conditions are recommended, which differ from the normal manufacturer's protocol, when the Illumina *Infinium methylation assay* is used. This step based upon Illumina's feedback to Zymo Research that bisulfite conversion efficiency can be improved by incorporating a cyclic denaturation protocol during the process of conversion.

### 2.3.2 Array-based genome-wide assessment of CpG methylation

For mapping CpG methylation level of genomic bisulfite-converted DNA the *Infinium Human Methylation450k BeadChip array* (HM450; Illumina, San Diego, CA, USA) was used. The HM450 array assesses the methylation levels of 485,577 CpG sites (482,421 CpG sites, 3,091 non-CpG sites and 65 random SNPs) and uses the *Infinium methylation assay*.

The CpG probes of the HM450 BeadChip are located in 21,231 RefSeq genes and 26,658 UCSC annotated CpG islands (CGI), as well as in genomic regions such as 5' and 3' UTRs, gene body and promoter. The 5'-regulatory gene region was divided into two blocks of 200 bps (TSS200) and 1,500 bps (TSS201-TSS1500) upstream of the transcription start site (TSS). The CGI region was further extended by including the 2 kb regions flanking CpG island shores (N = 26,249) as well as the CpG island shelves (2 kb regions upstream and downstream of the CpG island shores; N = 24,018; Bibikova et al. 2011).

The *Human Methylation450 BeadChip* applies both Infinium I and II assay chemistry technologies to quantitatively assess the methylation state of bisulfite-treated CpG sites. Both Infinium probes are 50 bases long, but detection of the methylation levels occur by different mechanisms. One bead type of the Infinium I assay corresponds to methylated (C), another bead type to unmethylated (T) state of the CpG site. Both bead types for the same CpG locus incorporate the same type of labeled nucleotide, determined by the base preceding the interrogated "C" in the CpG locus, and will be



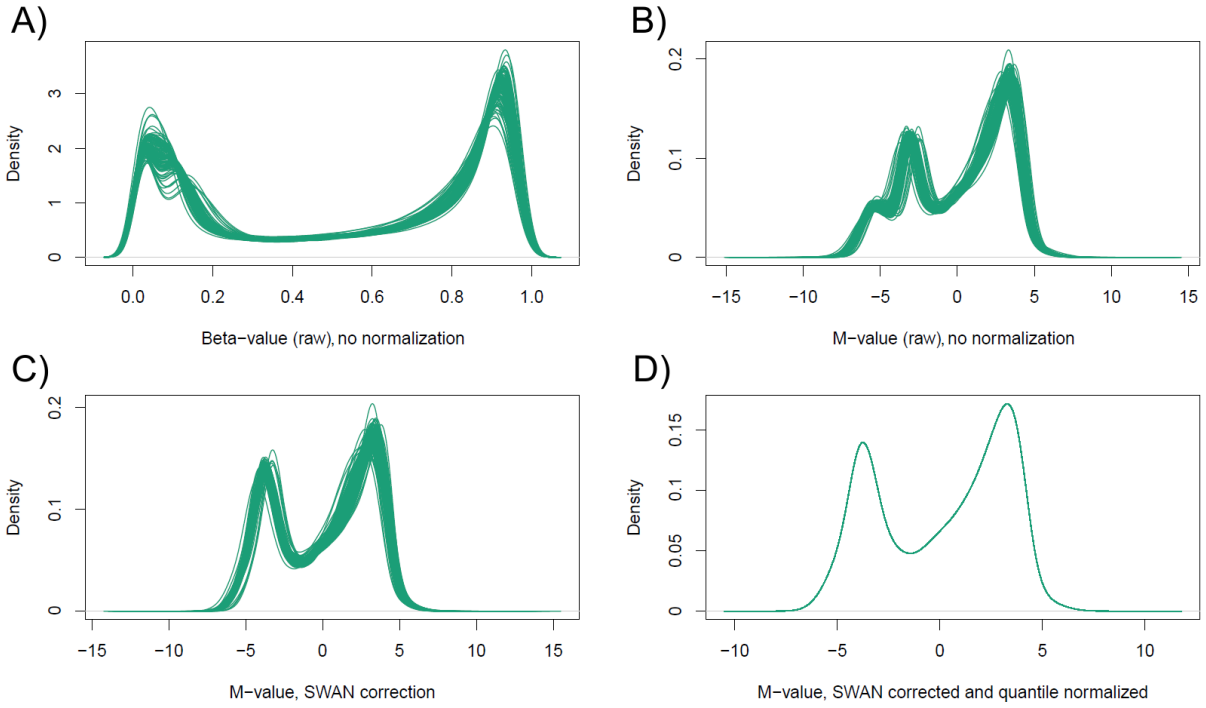
detected in the same color channel. For the Infinium II assay only one bead type corresponds to each CpG locus. Each locus will be detected in two colors (red and green fluorescence signals) by single-base extension (SBE) which reflects the methylation state (Bibikova et al. 2011). The Infinium methylation assay was performed following the standard Infinium protocol.

### 2.3.3 Assessment of signal intensities

Infinium methylation data was processed with the Methylation Module of the *GenomeStudio* software (v2011.1) using *HumanMethylation450 manifest v1.1*. The *GenomeStudio Methylation Module* calculates methylation level of each CpG locus as methylation beta-value ( $\beta$ -value, see below) using the ratio of intensities between methylated and unmethylated alleles (Bibikova et al. 2011). These  $\beta$ -values were exported and used for further analyses. In addition, the *GenomeStudio Methylation Module* has an Infinium Methylation Controls Dashboard which provides a couple of quality parameter like staining, hybridization, extension, bisulfite conversion and specificity. This Control Dashboard gives an overview of the technical quality of the array and the array run.

### 2.3.4 Normalization of signal intensities

The brain and blood methylation arrays were SWAN (subset-quantile within array normalization; Maksimovic et al. 2012) corrected and quantile normalized using the *Bioconductor* R-packages *preprocessCore* and *minfi*. The SWAN performs an independent normalization of six probe categories. The categories are defined by the differentiation of type I and II Infinium probes and in combination with the CpG number in the probe-body (one to three). Using the SWAN method technical variability within and between arrays could be reduced. To increase the performance in terms of detection and true positive rate of highly methylated and unmethylated CpG sites (Du et al. 2010)  $\beta$ -values were transformed to M-values (see below) and subsequently quantile normalized between the arrays (Figure 2-1). The resulting M-values were further checked for general signal deviances in a principle component analysis (PCA). PCA was performed using a correlation dispersion matrix and normalized Eigenvector scaling (*prcomb*, *stats* R-package).



**Figure 2-1: Normalization and transformation effects.**

The distribution of signal values without normalization (A, B), using within array SWAN normalization (C), using within array SWAN normalization and between arrays quantile normalization (D) and using M-value transformation (B - D) is plotted.

The  $\beta$ -value is defined as the ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities), ranging from 0 to 1. For an  $i^{\text{th}}$  interrogated CpG site  $\beta$ -value is defined as:

$$\beta_i = \frac{\max(y_{i,\text{methy}}, 0)}{\max(y_{i,\text{unmethy}}, 0) + \max(y_{i,\text{methy}}, 0) + \alpha}$$

$y_{i,\text{methy}}$  and  $y_{i,\text{unmethy}}$  are the intensities measured by the  $i^{\text{th}}$  methylated and unmethylated probes, respectively. Ideally, a value of zero indicates that all copies of a CpG site are completely unmethylated (no methylated molecules are measured) and a value of one indicates that every copy of the site is methylated. In contrast, the M-value is calculated as the  $\log_2$  ratio of the intensities of methylated probe versus unmethylated probe as shown in the following equation:

$$M_i = \log_2 \left( \frac{\max(y_{i,\text{methy}}, 0) + \alpha}{\max(y_{i,\text{unmethy}}, 0) + \alpha} \right)$$

It has a range from positive to negative values. An M-value close to 0 indicates a similar intensity between the methylated and unmethylated probes, which means that the CpG site is about half-methylated. The meaning of positive M-values is that more molecules are methylated than

unmethylated, while negative M-values mean the opposite. M-values have been widely used in expression microarray analysis, especially two-color microarray analysis and are more statistically valid for differential methylation analysis of methylation levels of CpG loci. In the present work M-values are particularly used for demonstrating methylation levels.

### 2.3.5 Quality control filters of CpG methylation profiles

To ensure a high accuracy of the CpG methylation profiles several quality control filters were applied – mainly based on the previous work of Chen and colleagues (2013): i) CpG probe sets were filtered according to hg19 autosomal representation, ii) CpG detection signal had to be sufficient ( $P$ -value  $> 0.01$  in  $> 5\%$  of the samples), and iii) SNPs in the cg 50mer probe with MAF  $> 1\%$  using the 1000 Genomes Project (release 20110521) were excluded. Additionally, CpGs that overlap known SNPs, so called polymorphic CpGs, were excluded. A CpG site was defined to be polymorphic if a SNP resided at the position of the cytosine or guanine on either strand, and for Infinium I assays, if a SNP resided at the position where SBE occurs (base before C). Furthermore, nonspecific probes (aka as cross-reactive probes, Chen et al. 2013) on the Illumina 450K array, i.e. probes which co-hybridize to alternate sequences which are highly homologous to the intended targets (ca. 6%) were excluded. In total, a set of QC-filtered 362,722 CpG probes remained of the original 485,577 CpG probe set for the regression statistics (see below 2.5).

## 2.4 mRNA expression analysis

For mapping mRNA expression of the 115 mTLE patients the *Illumina HumanHT-12 v3 Expression BeadChip* (Illumina, San Diego, CA, USA) was used. Each array on the *HumanHT-12 v3 Expression BeadChip* targets more than 48,000 probes which were derived from the National Center for Biotechnology Information Reference Sequence (NCBI, RefSeq) (Build 36.2, Release 22) and the UniGene (Build 199).

Raw intensity values for each mRNA expression probe were generated in the Department of Genomics, Research Center Life & Brain at Bonn University, Germany, using the *HumanHT-12 v3 BeadChips*. Subsequently, mRNA expression data were quantile normalized on probe level and without background correction using Illumina *GenomeStudio*. The resulting signals were  $\log_2$  transformations after offset addition (+16). Exclusion criteria for Illumina *HT12* probes included: i) probes which were only marginal or not expressed (minimum Illumina detection  $P$ -value  $> 0.05$ ), and ii) probes without autosomal (hg19) positioning. Additionally, ambiguous and SNP-containing probes

(MAF > 1% in dbSNP 137, 1000 Genomes release 20110521) were excluded for further investigations. A set of QC-filtered 31,146 mRNA expression probes was used for regression statistics.

## 2.5 Statistical quantitative trait loci (QTL) analyses

QTL analyses investigate the influence of genetic variation (SNPs) and a quantitative trait (CpG methylation or mRNA expression). A survey of the integrated data sets for QTL analyses are listed in Table 2-1.

The meQTL and eQTL analyses were performed using the linear regression model of the R-package *MatrixEQTL* (Shabalín 2012). *MatrixEQTL* is designed for fast and memory efficient QTL analyses on large data sets and provides the opportunity of covariate integration without loss of speed. Multiple testing corrections were performed in cis and in trans using the false discovery rate (FDR) step up procedure of *MatrixEQTL* (Benjamini & Hochberg 1995). A  $\pm 1$  Mbps cis window and a standard 5% FDR threshold were used throughout this study. A cis-meQTL or cis-eQTL refer to the most associated SNP/CpG or SNP/mRNA pair within the 1 Mbps windows flanking the target CpG or target mRNA probe, respectively. Trans-meQTLs or trans-eQTLs will refer to all SNP/CpG or SNP/mRNA pairs outside the 1 Mbps windows flanking the target CpG or target mRNA probe respectively. Nevertheless, FDR-correction was done over the complete cis or trans data matrix respectively.

**Table 2-1: Survey of the data sets integrated for QTL analyses.**

The available sample cohorts with the number of filtered individuals, CpG sites, mRNA probes and SNPs are listed.

Condition	Count
<b>Fresh frozen hippocampal brain tissue of mTLE patients:</b>	<b>115</b>
<b>meQTL and eQTL analyses with samples having</b>	
- genotype data: HumanHap550, autosomal QC-filtered SNPs (imputed & LD-pruned)	<b>643,195</b>
- methylation data: HM450, autosomal QC-filtered CpG sites	<b>362,722</b>
- expression data: HumanHT12, autosomal QC-filtered mRNA probes	<b>31,146</b>
<b>Blood cells of HNR controls:</b>	<b>496</b>
<b>meQTL analysis with samples having</b>	
- genotype data: OmniExpress-12 & Omni1-Quad, autosomal QC-filtered SNPs	<b>539,936</b>
- methylation data: HM450, autosomal QC-filtered CpG sites	<b>362,722</b>
<b>Blood cells of GGE parent-offspring-trios:</b>	<b>269</b>
<b>imeQTL analysis with samples having</b>	
- genotype data: Axiom, autosomal QC-filtered SNPs (imputed & LD-pruned)	<b>886,110</b>
- methylation data: HM450, autosomal QC-filtered CpG sites	<b>362,722</b>

### 2.5.1 meQTL and eQTL analyses in hippocampal brain tissue

meQTL and eQTL analyses in 115 fresh frozen hippocampal brain tissue were performed using gender, age at epilepsy surgery and neuronal proportion as cofactors. Gender (Zhang et al. 2010), age at epilepsy surgery (Martino et al. 2013) and neuronal proportion (Guintivano et al. 2013) are known confounders with impact on the general CpG methylome. Existence of additional unknown confounders with general impact on DNA methylation or RNA expression was investigated by PCA on linear regression gender/age/neuronal-proportion residuals. Component loading of the first or first and second principal component (PC) were used as additional cofactors in meQTL and eQTL analyses, respectively.

#### 2.5.1.1 Quantification of cell-type heterogeneity

To quantify neuronal proportions and to generate in silico neuronal profiles following removal of cell-type heterogeneity bias from DNA methylation profiles (Guintivano et al. 2013) the *CETS* package in R was used.

The resulting measures were matched with the pathological diagnosis of the proportion of neuronal cells in the hippocampal brain tissue. A Spearman rank correlation rho of the clinical parameters "nerve cell loss" and "pathology" and the CETS prediction of neuronal cells was calculated. The nerve cell loss parameters are positive correlated to each other. The status of lesional mTLE was found to be positively correlated to the neuronal cell proportion calculated by the R-package *CETS*. From a clinical point of view, the observed decrease of the neuronal proportion in AHS is plausible and in concordance to the pathophysiological observed nerve cell loss in AHS. The CETS proportion of neurons is negatively correlated to the nerve cell loss parameters but only significant in CA1 and CA3 (Table 2-2 & 2-3).

**Table 2-2: Parameters and counts of clinical parameters for hippocampal brain tissue.**

Parameter	Total count	Feature count
Pathology	N = 115	AHS: N = 79; lesional mTLE: N = 36
Nerve cell loss CA1	N = 73	AHS: N = 65; lesional mTLE: N = 8
Nerve cell loss CA2	N = 68	AHS: N = 59; lesional mTLE: N = 9
Nerve cell loss CA3	N = 67	AHS: N = 58; lesional mTLE: N = 7
Nerve cell loss CA4	N = 74	AHS: N = 65; lesional mTLE: N = 9

*Abbreviations: AHS, Ammon's Horn sclerosis; mTLE, mesial temporal lobe epilepsy; CA1 – CA4, hippocampal regions of the cornu Ammonis (Ammon's horn; Blumcke et al. 2007).*

Table 2-3: Spearman rank correlation rhos and nominal P-values of pairs of parameters.

	Lesional mTLE	CETS neurons (proportion of neurons)	Nerve cell loss CA1	Nerve cell loss CA2	Nerve cell loss CA3	Nerve cell loss CA4
Lesional mTLE		0.53	-0.76	-0.40	-0.42	-0.37
CETS neurons	1.44E-09**		-0.28	-0.02	-0.23	-0.18
Nerve cell loss CA1	3.12E-23**	2.44E-03*		0.35	0.31	0.23
Nerve cell loss CA2	1.13E-05**	8.28E-01	1.48E-04**		0.37	0.36
Nerve cell loss CA3	2.85E-06**	1.53E-02*	8.16E-04**	3.65E-05**		0.85
Nerve cell loss CA4	3.93E-05**	5.64E-02	1.32E-02*	8.99E-05**	2.01E-33**	

\*\*  $P < 0.001$ ; \*  $P < 0.05$ ; blue, Spearman rank correlations and red, nominal P-values of pairs of parameters.

### 2.5.1.2 Principal component analysis (PCA)

PCAs were calculated using *prcomb* (*stats* R-package) on residuals of known confounders (gender, age and neuronal proportion) generated by *lm* (*stats* R-package). Selection of relevant PC was done by scree plots. The scree plots show the proportion of variance accounted for each individual PC. Based on the scree plots PC1 was chosen as additional cofactor in meQTL analysis as the amount of explained variation dropped after the first component and PC2 as additional cofactor in eQTL analysis as the amount of explained variation dropped after the second component (Figure 2-2). These PCs have general impact on DNA methylation and mRNA expression, respectively; model for *meQTL analysis signal*  $\sim$  gender + age at epilepsy surgery + neuronal proportion + PC1, model for *eQTL analysis signal*  $\sim$  gender + age at epilepsy surgery + neuronal proportion + PC1 + PC2.

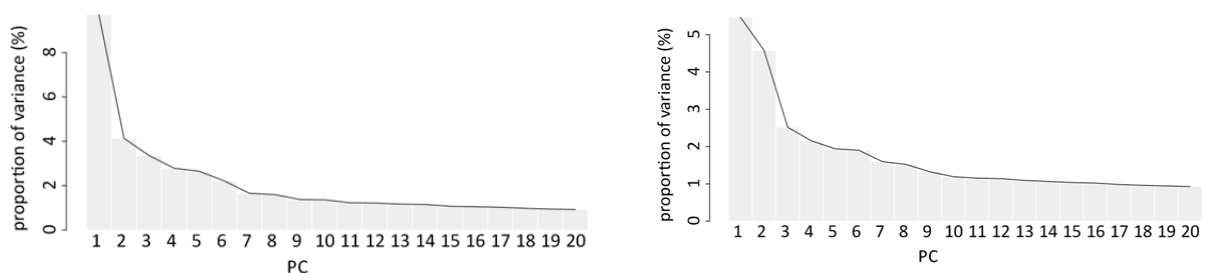


Figure 2-2: Scree plots.

On the left side: PCA variance proportion of CpG methylation; on the right side: PCA variance proportion of mRNA expression.

A successive addition of the PCs one to 10 to the model followed by additional cis-meQTL and cis-eQTL calculations respectively was used to estimate the influence of the first ten PCs to  $P$ -value inflation. The successive addition of PCs led to a very moderate increase of  $\lambda$  only. The lambda median for meQTL analysis (CpG) varied from 1.031 to 1.035 and the lambda median for eQTL analysis (mRNA) varied from 1.016 to 1.020. In conclusion, I can state that the first ten PCs do not

include potential hidden factors which led to a substantial decrease of inflation in cis and can be considered robust.

### 2.5.2 Methylation-driven gene expression in hippocampal brain tissue

DNA methylation and gene expression patterns of 115 fresh frozen hippocampal brain tissue samples were combined by performing CpG methylation to mRNA gene expression correlation analyses. For correlation analyses of meQTL and eQTL of hippocampal brain tissue, the same models as before were used (model for *CpG methylation to gene expression signal*  $\sim$  *gender + age at epilepsy surgery + neuronal proportion + PC1*; model for *gene expression to CpG methylation signal*  $\sim$  *gender + age at epilepsy surgery + neuronal proportion + PC 1 + PC2*). Pearson correlation analyses were calculated on residuals of the confounders generated by the *MatrixEQTL* R-package. Multiple testing corrections were performed in cis ( $\pm$  1Mbps) using FDR step up procedure of *MatrixEQTL* (Benjamini & Hochberg 1995). For the most associated CpG/mRNA pairs within the 1 Mbps windows flanking the target CpG or target mRNA probe, respectively, Pearson correlation coefficients (r-values) were calculated using *cor* (*stats* R-package).

Significant CpG/mRNA probe pairs were selected and these pairs were expanded into triplets of SNP and CpG/mRNA pairs. The SNPs were significantly correlated in cis with the CpG methylation site (cis-meQTL finding) and the mRNA transcript probe (cis-eQTL finding) of the respective CpG/mRNA pair. The meQTL-SNPs and eQTL-SNPs had to be in LD ( $r^2 \geq 0.2$ ).

### 2.5.3 meQTL analysis in blood cells

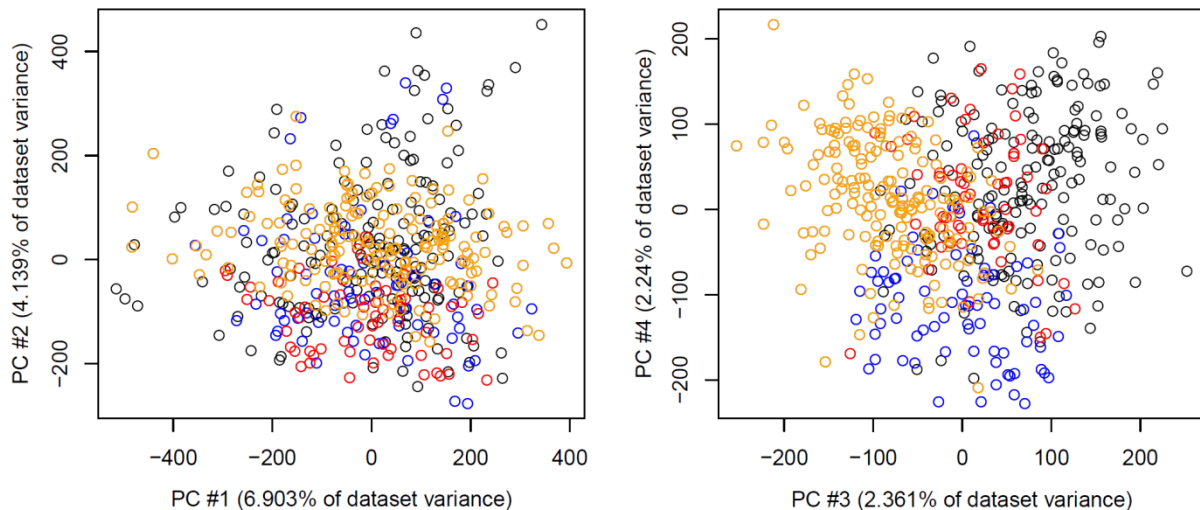
meQTL analysis in 496 HNR blood cells was performed using gender, age of blood withdrawal, cell-type composition (6 levels), and month of methylotyping (3 of 4 levels) as cofactors. Gender and age of blood withdrawal, as well as cell-type composition (Houseman et al. 2012; Jaffe & Irizarry 2014) and month of methylotyping (as batch effect; Harper et al. 2013) are known confounders with impact on the general CpG methylome. Existence of additional unknown confounders with general impact on DNA methylation was investigated by PCA on linear regression gender/age/ cell-type composition/ month of methylotyping residuals.

#### 2.5.3.1 Determination of cell-type composition

Cell-type composition was performed according to Jaffe and Irizarry (2014) and Houseman et al. (2012) which estimate relative proportions of cell-type components in whole blood. Six cell-types including CD4+ effector T-cells (CD4T), B-cells (Bcell), Cytotoxic T-cell (CD8T), Natural killer cells (NK), Mononuclear white blood cells (Mono) and Granulocytes (Gran) were used. The statistical method by Houseman et al. (2012) is virtually the same statistical approach described in Guintivano et al. (2013) for estimating neuron and non-neuron components in brain samples.

### 2.5.3.2 Determination of month of methylotyping

PCA considering four levels of month of methylotyping revealed a strong agglomeration of the factor, mainly in the third and fourth PC. The third and fourth PC reflect 2.36% and 2.24% of data set variance, respectively (Figure 2-3). Therefore, month of methylotyping was added as cofactor.



**Figure 2-3: PC variance for four levels of month of methylotyping.**

On the left side: variance of PC 1 & of PC 2; on the right side: variance of PC 3 & of PC 4. The four colors indicate the four different months of methylotyping.

### 2.5.3.3 Principal component analysis (PCA)

PCAs were calculated using *prcomb* (*stats* R-package) on residuals of known confounders (gender, age of blood withdrawal, cell-type composition, and month of methylotyping) generated by *lm* (*stats* R-package). No PC was added as additional cofactor in meQTL analysis of the blood cells as the proportion of variance remained steady for all PCs. The PCs have no general impact on DNA methylation; model for *meQTL analysis signal*  $\sim$  *gender + age of blood withdrawal + cell-type composition + month of methylotyping*. A successive addition of the PCs one to 10 to the model followed by additional cis-meQTL calculations was used to estimate the influence of the first ten PCs to *P*-value inflation. The successive addition of PCs led to a very moderate increase of  $\lambda$  only. The lambda median for meQTL analysis (CpG) varied from 1.111 to 1.118. In conclusion, I can state that the first ten PCs do not include potential hidden factors which led to a substantial decrease of inflation in cis and that models can be considered robust.

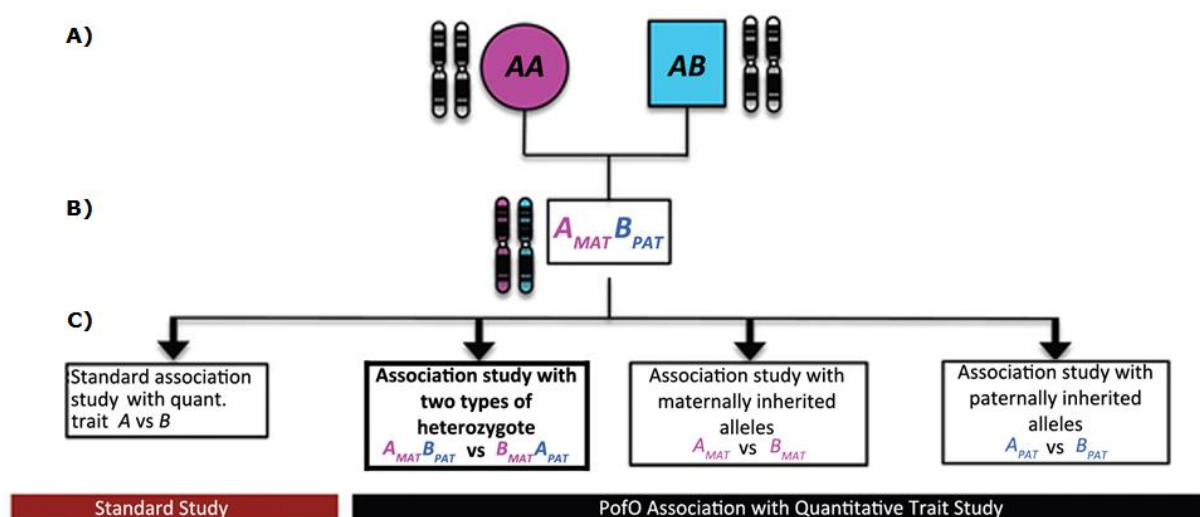
### 2.5.4 Imprinting methylation QTL (imeQTL) analysis in blood cells

I used a novel approach to identify meQTLs with a Parent-of-Origin (PoFO) bias. Parent-offspring trios with GGE were used for the dissection of differentially methylated parental genomic regions. imeQTL analysis in 269 parent-offspring trios was performed using gender, age of blood withdrawal, cell-type



composition (6 levels), month of methylotyping (4 of 5 levels) and origin as cofactors. Gender, age of blood withdrawal, cell-type composition, month of methylotyping and origin of sample are known confounders with impact on the general CpG methylome (Zhang et al. 2010). Cofactor origin included the different European countries (or countries with European ancestry) from where parent-offspring trios were recruited. The cofactors were AUS (Australia) + central Europe (Denmark & Germany) + IT (Italy) + TR (Turkey). Existence of additional unknown confounders with general impact on DNA methylation was investigated by PCA on linear regression gender/age/cell-type composition/month of methylotyping/origin residuals. PCAs were calculated using *prcomb* (*stats* R-package) on residuals of known confounders generated by *lm* (*stats* R-package). No PC was added as additional cofactor in imeQTL analysis as the proportion of variance remained steady across PCs and hence it can be assumed that the PCs have no general impact on DNA methylation; model for *imeQTL analysis signal*  $\sim$  gender + age of blood withdrawal + cell-type composition + month of methylotyping + origin.

With trios, by analyzing parent and child genotypes, rules of Mendelian inheritance allow parental origin of SNP alleles in the child to be defined, as long as at least one member of the trio is homozygous. Three cis-association tests were performed to determine the parental origin and to test PofO specific association for a SNP: i) an association study using only maternally inherited SNPs, ii) an association study using only paternally inherited SNPs, and iii) a comparison of the frequencies of the two classes of reciprocal heterozygote in which parental origin is reversed ( $A_{MAT}/B_{PAT}$  vs.  $A_{PAT}/B_{MAT}$ ; Garg et al. 2012). The method is summarized in Figure 2-4.



**Figure 2-4: Determination of parental origin and PofO specific association testing for a hypothetical SNP.**

A) & B) Assignment of allele parental origin at a given SNP (alleles A and B): the homozygous mother can only contribute allele A to her offspring (MAT, pink), and thus child's B allele must be paternally inherited (PAT, blue). C) Standard association study between SNP genotypes and quantitative trait compared to the proposed scheme, which considers the parental origin of the inherited SNPs present in the child (Figure adapted from Garg et al. 2012).

To detect the differential paternal or maternal associations the following thresholds were defined: 1) one of the parental FDR corrected *P*-values has to be significant (FDR < 0.05) whereas the other, native *P*-value has to be non-significant (native *P*-value > 0.05). The difference between the maternal and paternal *P*-values had to be greater than 10<sup>-6</sup>.

#### 2.5.4.1 Known imprinted genes

The definition of known imprinted genes was adopted from Fang et al. (2012). Fang and colleagues analyzed the methylome for a diverse set of human cell-types, including cultured and uncultured differentiated cells, embryonic stem cells and induced pluripotent stem cells to generate a genomic landscape of human allele-specific DNA methylation (Table 2-4). Validated imprinting regions were double-checked and ensured by using the information repository of mammalian imprinted genes, *MetalImprint* (Wei et al. 2014).

**Table 2-4: Known imprinted clusters and associated allelically methylated regions.**

*Clusters are named according to a representative gene; approximately 65 validated human imprinting genes reside in 32 imprinted clusters were identified. In total, 21 of the clusters contained validated allelically methylated region (AMRs) in four of five uncultured cells. Columns also indicate whether the selected representative AMR was previously identified, and whether it is a known imprinted control region (ICR).*

Imprinting Cluster	hg19 location	Size (kb)	Uncultured	Previously identified	Known ICR	Total human
<i>DIRAS3</i>	chr1:68515537-68517691	2154	5	Yes	Yes	19
<i>NAP1L5</i>	chr4:89617864-89619549	1685	5	Yes	Yes	21
<i>FAM50B</i>	chr6:3848745-3850911	2166	4	Yes	Yes	12
<i>PLAGL1/HYMAI</i>	chr6:144327845-144330191	2346	4	Yes	Yes	18
<i>DDC/GRB10</i>	chr7:50849470-50851331	1861	4	Yes	Yes	19
<i>SGCE/PEG10</i>	chr7:94284018-94289851	5833	5	Yes	Yes	22
<i>MESTIT1/MEST</i>	chr7:130129559-130133682	4123	5	Yes	Yes	22
<i>KCNK9</i>	chr8:141108018-141111481	3463	5	Yes	Yes	20
<i>FANK1</i>	chr10:127584249-127588031	3782	5	Yes	Yes	18
<i>INS-IGF2-H19</i>	chr11:2016476-2024739	8263	5	Yes	Yes	20
<i>KCNQ1OT1</i>	chr11:2719386-2722440	3054	5	Yes	Yes	21
<i>RB1</i>	chr13:48890275-48895948	5673	5	Yes	Yes	7
<i>DLK1/MEG3</i>	chr14:101290239-101295152	4913	3	Yes	Yes	5
<i>SNRPN/SNURF</i>	chr15:25199298-25202152	2854	5	Yes	Yes	22
<i>ZNF597,NAA60</i>	chr16:3492777-3494769	1992	5	Yes	Yes	22
<i>ZIM2/PEG3</i>	chr19:57348719-57353128	4409	5	Yes	Yes	12
<i>PSIMCT-1/HM13</i>	chr20:30134590-30135902	1312	5	Yes	Yes	20
<i>BLCAP/NNAT</i>	chr20:36148274-36151269	2995	5	Yes	Yes	14
<i>L3MBTL</i>	chr20:42142223-42144439	2216	4	Yes	Yes	19
<i>GNAS</i>	chr20:57414794-57486250	71456	5	Yes	Yes	22
<i>TCEB3C</i>	chr18:44548657-44550534	1877	4	Yes	Yes	11

## 2.6 Exploration of the genomic features of QTL analyses

### 2.6.1 Genomic distribution of cis-meQTLs

To explore the distribution of methylated sites associated with cis-meQTLs 12 chromatin state annotations in a B-lymphoblastoid cell line from the blood of a Caucasian female donor (GM12878; Ernst & Kellis 2010) were used. Ernst et al. (2011) distinguished six broad classes of chromatin states, which are referred to as promoter, enhancer, and insulator, as well as transcribed, repressed and inactive states. Within these, active, weak and poised promoters differ in expression level, strong and weak candidate enhancers differ in expression of proximal genes, and strongly and weakly transcribed regions also differ in their positional enrichments along transcripts. Polycomb-repressed regions differ from heterochromatin and repetitive states (Ernst et al. 2011).

### 2.6.2 Enrichment analysis of cis-meQTLs and cis-eQTLs

The enrichment analyses were performed using the R-package *goseq*, which allows selection-unbiased testing for category enrichment amongst differently expressed or methylated genes. Any category may be tested using *goseq*. The following gene-sets, compiled by their expression in brain and their involvement in neuropsychiatric disorders, were used: i) brain expressed genes (N = 8,852; Pinto et al. 2014), ii) genes involved in neurodevelopmental disorders (N = 1,689; Krumm et al. 2013), iii) genes involved in autism spectrum disorder (N = 1,559; Uddin et al. 2014), and iv) susceptibility genes (N = 1,200) at GWAS loci published in the NHGRI GWAS catalog ( $P < 1 \times 10^{-6}$ ; Welter et al. 2014). A correction for the numbers of CpG sites located in different classes of genes according to Gleeleher et al. (2013) was performed. *goseq* used CpG counts as bias data. For calculating the *P*-values the Wallenius approximation was used. The name of the category and the *P*-value for the associated category being over represented amongst differentially methylated genes was documented.

## 2.7 Exploration of the role of QTL-associated rSNPs in common neuropsychiatric disorders

Five filtering steps were carried out to explore whether regulatory SNPs (rSNPs) affecting transcription factor binding sites (TFBS) underlie both the cis-acting QTLs (meQTLs and eQTLs) and the regional GWAS hits. The step-wise filtering procedures included: 1) positional screening of cis-QTLs close by (1 Mbps window) the lead SNPs of GWAS hits ( $P < 5 \times 10^{-8}$ ); 2) dissection of those SNP-pairs (cis-QTL-SNP / GWAS-SNP) showing a LD-relationship with  $r^2 > 0.5$ ; 3) extension of the candidate rSNPs by capturing all SNPs in LD  $r^2 > 0.8$  with the GWAS lead SNP; 4) prioritization of

potential rSNPs affecting TFBS by assessing the allelic alteration of their binding affinity using the GWAS3D software tool (<http://jjwanglab.org/gwas3d>, Li et al. 2013); and 5) quantitative estimates of pathogenic effects across a wide range of functional categories based on i) the combined annotation dependent depletion score (CADD score, Kircher et al. 2014), and ii) the GERP++ score assessing evolutionary sequence conservation (Davydov et al. 2010) implemented in the SNP annotation tool SNIIPA (<http://snipa.helmholtz-muenchen.de/snipa/>, Arnold et al. 2014).

The detected transcription factors (TFs) all belong to TRANSFAC (TRANScription FACtor database) which is a database on eukaryotic transcriptional regulation comprising data on TFs, their target genes and regulatory binding sites (Matys 2003).

### 3. RESULTS

My analyses addressed the following objectives:

- 1) Identification of rSNPs quantitatively influencing CpG methylation and mRNA expression in fresh frozen hippocampal brain tissue.
- 2) Delineation of a genome-wide meQTL map in fresh frozen hippocampal brain tissue and blood cells, offering the dissection of accessible biomarkers.
- 3) Genome-wide assessment of imeQTLs to identify loci that are prone for susceptibility effects by their Parent-of-Origin PofO dependent monoallelic expression.
- 4) Generation of a genome-wide eQTL map in fresh frozen hippocampal brain tissue and delineation of methylation-driven gene expression.
- 5) Identification of rSNPs involved in common brain disorders.

Genome-wide methylation and expression analyses were performed in fresh frozen hippocampal brain tissue of 115 European (descent) patients with resistant mesial temporal lobe epilepsy (mTLE). These analyses were meant to gain fundamental insights into epigenetic regulation in human hippocampal brain tissue and to dissect those epigenetic alterations involved in common neuropsychiatric disorders. For this goal, individual high-density SNP genotypes were correlated with both (i) the individual quantitative methylation states of 362k CpG sites assessed by the *Human Methylation450 array* (HM450) and (ii) the individual expression level of 31k gene transcripts using the *Human HT-12 v3 array*. To delineate accessible epigenetic biomarkers, genome-wide maps of meQTLs were also generated in whole blood cells from 496 German population controls using the same 362k HM450 CpG sites.

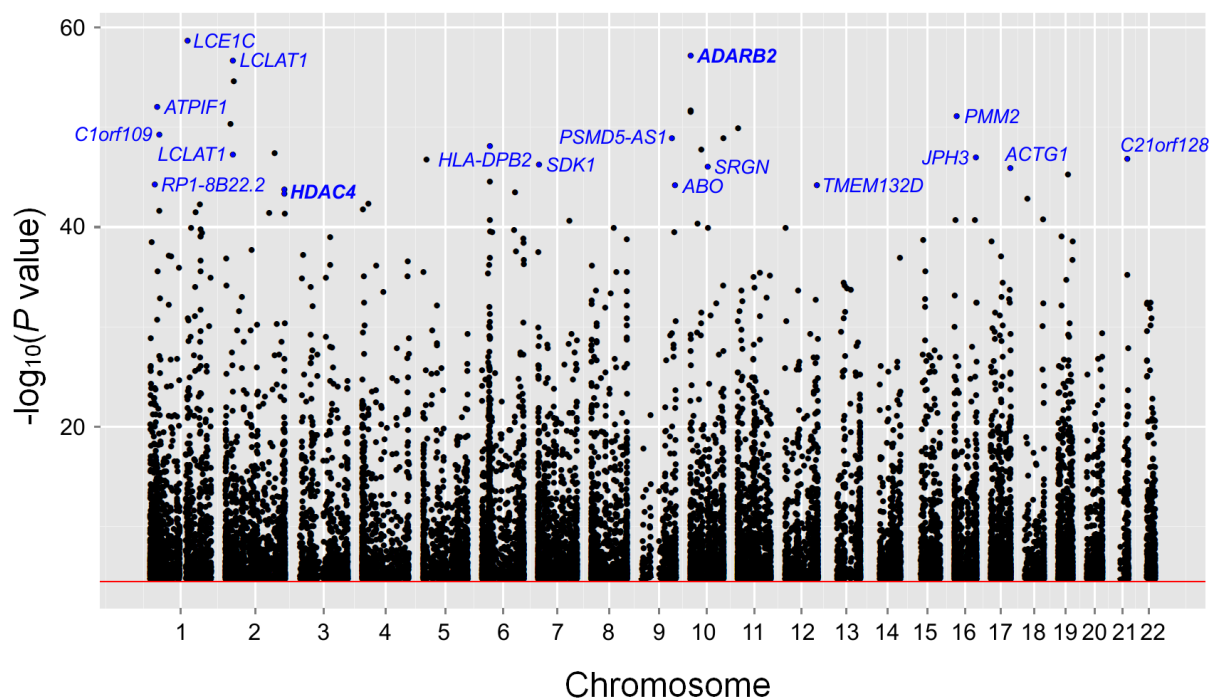
The significance threshold applied throughout this study refers to a standard FDR threshold  $< 0.05$  based on all examined SNP/CpG or SNP/mRNA pairs. In the following, a cis-meQTL or cis-eQTL will refer to the most associated SNP/CpG or SNP/mRNA pair within the 1 Mbps windows flanking the target CpG or target mRNA probe. Trans-meQTLs or trans-eQTLs will refer to all SNP/CpG or SNP/mRNA pairs outside the 1 Mbps windows flanking the target CpG or target mRNA probe.

#### 3.1 Methylation QTL analyses

##### 3.1.1 Cis-meQTL analysis in hippocampal brain tissue

The meQTL analysis examined the influence of cis-acting genetic variations on the quantitative state of CpG methylation. After stringent quality control, 362,722 CpGs and 643,195 LD-pruned SNPs were

correlated for cis-meQTL analysis, resulting in  $1.75 \times 10^8$  tests corresponding to a nominal significance threshold of  $P = 2.95 \times 10^{-5}$  at a FDR of 5%. At the FDR threshold of 5%, 19,954 (5.5%) out of 362,722 CpG sites account for significant cis-acting meQTLs ( $P$ -value range =  $1.99 \times 10^{-59}$  to  $2.95 \times 10^{-5}$ , Figure 3-1). A QQ-plot of the CpG methylation in 115 hippocampal brain tissue samples is shown in Appendix, Figure 6-1.



**Figure 3-1: Manhattan plot for cis-meQTLs identified in hippocampal brain tissue samples.**

The x-axis shows the genomic position of CpGs on chromosomes 1 to 22 (NCBI built 37.3, hg19); the y-axis shows the negative log<sub>10</sub> of the  $P$ -value per SNP/CpG pair. The red line represents the threshold for significant cis-meQTLs with FDR < 0.05. The genes of the 20 most significant cis-meQTLs (shown in Table 3-4) are highlighted.

In total, 13,355 out of 19,954 cis-meQTLs were located intragenic of 7,827 Ensembl genes (hg19). For 1,934 cis-meQTL CpGs a signal with two or more different gene transcripts was observed, while 6,599 cis-meQTLs could not be localized to an Ensembl gene annotation. 11,253 SNPs showed cis-association with two or more CpG sites. The positions of SNPs with significant cis-meQTLs were preferentially located near the CpG sites, 76% within a range of < 150 kb.

### 3.1.1.1 Genomic features associated with cis-meQTLs

Previous studies have shown that SNPs are weakly correlated with the methylation status of CpG islands (Bock et al. 2007; Gibbs et al. 2010). Consistently, cis-meQTLs in hippocampal brain tissue were found to be more likely outside of CpG islands than within (as defined in Gardiner-Garden & Frommer 1987). While 55% of 362k CpG sites are within a CpG island, only 23% of the CpG sites of a significant meQTL for mTLE patients were in islands ( $P < 2.2 \times 10^{-16}$ ). These results focused on CpG

islands themselves without considering CpG island shores (0-2 kb from CpG island) and CpG island shelves (2-4 kb from CpG island).

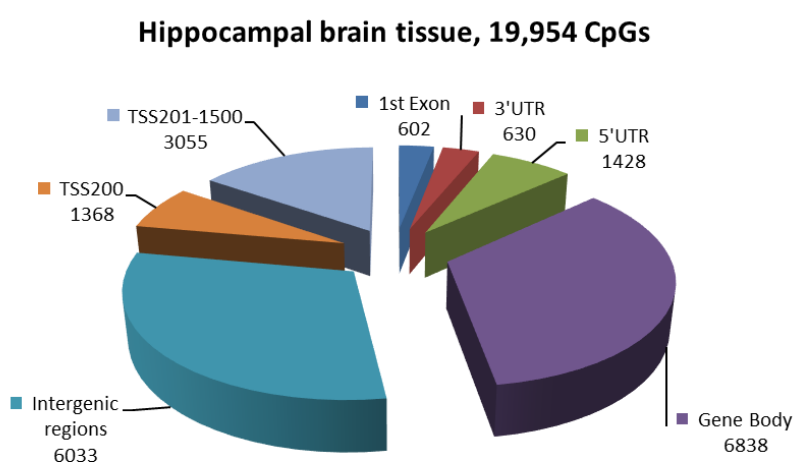
Exploration of the cis-meQTLs in gene-centric regions revealed a moderate enrichment in the TSS1500 region encompassing the regulatory 5'-region 1500 bps upstream from the transcription start site (TSS), compared to the distribution of all quality-filtered 362k CpGs ( $P = 4.41 \times 10^{-3}$ ; Table 3-1 & Figure 3-2). For the other genomic regions a depletion of sites associated with meQTLs were found. The findings revealed a particular accumulation of cis-meQTLs in the gene-centric region TSS201-TSS1500. In total, 3,055 (8.45%) cis-meQTLs were found in the TSS201-TSS1500 region compared to 3.6% of all 362k CpG sites ( $P = 7.72 \times 10^{-61}$ ).

**Table 3-1: Distribution of meQTLs within different gene-centric regions in hippocampal brain tissue.**

The proportion of all quality-filtered 362,722 CpGs in gene-centric regions, as specified by Illumina is shown. Pearson's Chi-squared test with Yates' continuity correction was used to calculate the distribution of 19,954 cis-meQTLs (FDR < 0.05) in different gene-centric regions in hippocampal brain tissue compared to all 362k CpG sites. Red colored background indicates depletion, green colored background an enrichment of meQTLs compared to all 362k CpGs. TSS200 & TS1500, promoter regions were divided into two bins of 200 bps and 1,500 bps blocks upstream of the TSS.

Illumina Group	All, 362,722 CpGs	cis-meQTL, brain	P-values, cis-meQTL, brain
1st Exon	4.88%	3.02% ***	4.70E-36
3'UTR	3.72%	3.16% **	1.60E-05
5'UTR	9.15%	7.16% ***	9.45E-24
Gene Body	34.06%	34.27%	5.40E-01
TSS200	10.98%	6.86% ***	1.07E-81
TSS201-1500	14.62%	15.31% *	4.41E-03
Intergenic regions	22.59%	30.22%***	2.50E-155

Chi-square test: \*\*\* $P < 10^{-50}$ ; \*\* $p < 10^{-5}$ ; \* $p < 0.05$



**Figure 3-2: Functional localization of genomic regions of meQTLs in hippocampal brain tissue.**

Distribution of different genomic regions for the significant 19,954 meQTLs obtained from hippocampal brain tissue.

### 3.1.1.2 Enrichment analysis

To investigate the biological function of the obtained cis-meQTL signals I looked for an enrichment of gene-sets compiled by their expression in brain and their involvement in neuropsychiatric disorders. The gene-sets included brain expressed genes (N = 8,852; Pinto et al. 2014), genes involved in neurodevelopmental disorders (N = 1,689; Krumm et al. 2013) and autism spectrum disorder (N = 1,559; Uddin et al. 2014). In addition, susceptibility genes (N = 1,200) at GWAS loci published in the NHGRI GWAS catalog ( $P < 1 \times 10^{-6}$ , Welter et al. 2014) were examined for the following common neuropsychiatric disorders: Alzheimer's disease, attention-deficit hyperactivity disorder, autism, bipolar disorder, epilepsy, major depressive disorder, migraine, nicotine dependence, Parkinson's disease and schizophrenia. This gene-set will further be referred as Brain-Disorder-GWAS gene-set. Because functional categories are biased by differing CpG contents of the gene-sets, a CpG bias correction was applied. Therefore, the number of CpG sites located in different classes of genes was adjusted according to the procedure recommended by Geeleher et al. (2013). The meQTLs in hippocampal brain tissue were significantly enriched in the Brain-Disorder-GWAS gene-set ( $P = 5.79 \times 10^{-4}$ ) and in the gene-set previously implicated in neurodevelopment disorders ( $P = 0.04$ ). No significant enrichment was observed for brain-expressed and for autism spectrum disorder genes after correcting of the CpG bias (Table 3-2). The significance of these results was determined empirically using sample randomization of 1,000 randomly distributed genes.

**Table 3-2: Gene-set enrichment analyses of cis-meQTLs in hippocampal brain tissue.**

*Enrichment analyses performed in hippocampal brain tissue samples. The analyses applied a correction of the number of CpG sites per gene, respectively. For each category a P-value is given for an enrichment of the compiled gene-set category.*

<b>Category</b>	<b>Overrepresented P-value</b>
Brain-Disorder-GWAS Genes (N = 1,200)	<b>5.79E-04</b>
Neurodevelopment Genes (N = 1,689)	<b>3.90E-02</b>
Autism Spectrum Disorder Genes (N = 1,559)	6.12E-01
Brain Expressed Genes (N = 8,852)	1.00E+00

### 3.1.1.3 Comparison of cis-meQTLs in hippocampal brain tissue and top cis-meQTLs

I compared cis-meQTL findings obtained from recent studies analyzing cis-meQTLs with quality-filtered 362k CpGs (Gibbs et al. 2010; Zhang et al. 2010; Numata et al. 2014; Smith et al. 2014; for an overview on studies that are frequently mentioned for comparison reasons, please refer to Table 6-2 in Appendix). Notable, all comparable meQTL studies assessed cis-methylation levels of postmortem brain tissue and peripheral blood cells with the Infinium HumanMethylation27 BeadChip (HM27). In total, 69% of the cis-meQTLs generated by Gibbs et al. (2010), 71% of the cis-meQTLs generated by Zhang et al. (2010), 67% of the cis-meQTLs generated by Numata et al. (2014), and 77% of the cis-meQTLs generated by Smith et al. (2014) could be identified in the 362k CpGs (Table 3-3). Overall,



66% of the cis-meQTLs generated by the four previous studies were also identified as significant cis-meQTLs in the present study.

*Table 3-3: Overlaps with comparable human meQTL studies.*

<b>Epigenetic study</b>	<b>Cell-type or tissue studied</b>	<b>Overlap with 362k CpGs</b>
Gibbs et al. (2010)	Postmortem brain tissue	69% (615 out of 887)
Zhang et al. (2010)	Postmortem brain tissue	71% (1,446 out of 2,046)
Numata et al. (2014)	Postmortem brain tissue	67% (2,427 out of 3,612)
Smith et al. (2014)	Whole Blood & postmortem brain tissue	77% (710 out of 926)

A major aim of the present meQTL study was a comprehensive map of meQTL findings of all quality-checked 362k CpGs for 115 hippocampal brain tissue samples and 496 blood cell samples in a publically accessible database. These databases will be reported in context with the publication of this study (manuscript in preparation).

The 20 most significant cis-meQTLs for hippocampal brain tissue are shown in Table 3-4. Additionally, FDR values of the cis-meQTLs for the respective CpG site in blood cells are listed.

**Table 3-4: Top hits of meQTLs in hippocampal brain tissue.**

CpG	Chr	CpG Pos	Gene Symbol	meQTL SNP	SNP Pos	FDR, brain	FDR, blood
cg07796016	1	152779584	<i>LCE1C(0.5k)</i>	rs7524281	152783255	3.48E-51	2.29E-202
cg09316607	10	1511024	<i>ADARB2</i>	rs1533486	1511786	5.52E-50	> 5.00E-02
cg15652532	2	30669759	<i>LCLAT1(0.3k)</i>	rs4952148	30777908	1.12E-49	3.02E-240
cg04993605	1	28573052	<i>ATPIF1</i>	rs2938867	28515292	1.92E-45	> 5.00E-02
cg08308162	16	8889244	<i>PMM2</i>	rs45515994	8876845	1.28E-44	5.29E-21
cg24657347	2	20261756	<i>AC098828.2(1.9k)</i>	rs7570953	20255015	6.25E-44	> 5.00E-02
cg06917450	1	38156652	<i>C1orf109</i>	rs768659	38162707	6.51E-43	1.67E-189
cg21717724	9	123604514	<i>PSMD5-AS1</i>	rs3793638	123582697	1.15E-42	2.82E-159
cg02375585	6	33091111	<i>HLA-DPB2</i>	rs114476417	33094663	6.13E-42	1.22E-39
cg23649088	2	200775458	<i>AC073043.1</i>	rs10180850	200747690	2.94E-41	3.29E-202
cg12454169	2	30669597	<i>LCLAT1(0.5k)</i>	rs4952148	30777908	3.91E-41	1.45E-196
cg09496634	16	87712353	<i>JPH3</i>	rs888652	87712366	7.60E-41	2.41E-119
cg19766460	21	43528205	<i>C21orf128</i>	rs2051401	43530856	9.01E-41	8.15E-238
cg24441899	7	4244372	<i>SDK1</i>	rs57656159	4244666	3.41E-40	9.79E-103
cg14418922	10	70824840	<i>SRGN(23.0k)</i>	rs17558246	70851873	4.79E-40	1.65E-217
cg19825371	17	79454562	<i>ACTG1(22.4k)</i>	rs62075991	79459113	7.16E-40	1.98E-126
cg18795169	1	18902165	<i>RP1-8B22.2(19.7k)</i>	rs2745311	18904033	2.85E-38	8.01E-114
cg14440550	9	136131118	<i>ABO</i>	rs8176751	136131022	3.07E-38	2.16E-139
cg01290755	12	129554587	<i>TMEM132D(1.7k)</i>	rs11060115	129554644	3.07E-38	5.92E-153
cg07673080	2	240241154	<i>HDAC4</i>	rs12476996	240257894	7.23E-38	2.62E-133

Abbreviations: CpG, Illumina annotation of the CpG dinucleotide; Chr, chromosome; CpG Pos, cytosine base-pair position of CpG site; Gene Symbol, gene which contains the CpG site; meQTL SNP brain, SNP of the most significant meQTL in hippocampal brain tissue; SNP Pos brain, base-pair position of SNP in hippocampal brain tissue; FDR brain & blood, most significant FDR for a SNP-CpG pair in hippocampal brain tissue samples and in blood cell samples. All annotations refer to genome build hg19.

## 3.1.1.4 Neurodevelopmental candidate genes of cis-meQTLs

Some of the most significant cis-meQTLs were identified in genomic regions of several high-ranking candidate genes for neurodevelopmental disorders: *ADARB2* ( $P_{min} = 6.33 \times 10^{-58}$ ), *HDAC4* ( $P_{min} = 1.62 \times 10^{-44}$ ), *NAPRT1* ( $P_{min} = 1.61 \times 10^{-39}$ ), *MAD1L1* ( $P_{min} = 8.35 \times 10^{-29}$ ), *PTPRN2* ( $P_{min} = 1.67 \times 10^{-26}$ ), and *RIMBP2* ( $P_{min} = 2.32 \times 10^{-24}$ ; Table 3-5).

**Table 3-5: Significant cis-meQTLs of high-ranking candidate genes in hippocampal brain tissue.**

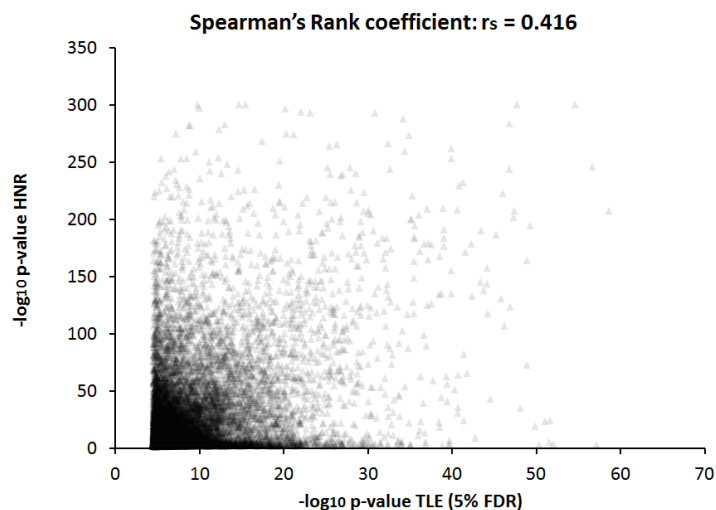
CpG	Chr	CpG Pos	Gene Symbol	meQTL SNP	SNP Pos	FDR	Rank cis-meQTL
cg09316607	10	1511024	<i>ADARB2</i>	rs1533486	1511786	5.52E-50	2
cg07673080	2	240257894	<i>HDAC4</i>	rs12476996	240257894	7.23E-38	29
cg17524265	8	144659883	<i>NAPRT1</i>	rs9657360	144676862	3.20E-33	61
cg13165778	7	1952518	<i>MAD1L1</i>	rs6950151	1953521	3.52E-23	247
cg24764310	7	157512201	<i>PTPRN2</i>	rs11762822	157507684	4.70E-21	351
cg20111978	12	130959558	<i>RIMBP2</i>	rs6486542	130952209	4.54E-19	481

Abbreviations: Rank cis-meQTL, position of the significant cis-meQTL compared to all detected cis-meQTLs; other abbreviations see Table 3-4.

## 3.1.2 Comparative cis-meQTL analysis of brain cells and whole blood cells

To discover accessible epigenetic biomarkers and to test for tissue specificity in meQTL analyses, a genome-wide map of cis-meQTLs were also generated in whole blood cells from 496 German population controls using the same 362k HM450 CpG sites. After stringent quality control, 362,722 CpGs and 539,936 LD-pruned SNPs were correlated for cis-meQTL analysis in blood cell samples, resulting in  $1.60 \times 10^8$  tests corresponding to a nominal significance threshold of  $P = 4.40 \times 10^{-4}$  at a FDR of 5%. At the FDR threshold of 5%, 108,628 (30%) out of 362,722 CpG sites were identified as cis-meQTLs ( $P$ -value range =  $2.23 \times 10^{-308}$  to  $4.40 \times 10^{-4}$ ). A QQ-plot of the meQTL analysis in blood cell samples is shown in Appendix, Figure 6-1.

In total, more than 65% of the significant cis-meQTLs in hippocampal brain tissue were also found in blood cells. Overall, 12,974 out of 19,954 CpGs with significant meQTLs in hippocampal brain tissue represent also meQTLs in blood cells. Accordingly, a moderate correlation of the significant cis-meQTL results in the hippocampal brain tissue and the cis-meQTL results in blood cells (Spearman's Rank coefficient = 0.416; Figure 3-3) was observed.



**Figure 3-3: P-value scatter plot of hippocampal brain tissue and blood cells.**

The linear correlation between 115 hippocampal brain tissue samples and 496 blood cell samples is shown. The negative  $\log_{10}$  of the P-value per SNP/CpG pair is plotted.

Cis-acting meQTLs with the best value per CpG at different significance thresholds for hippocampal brain tissue as well as for blood cells are given in Table 3-6.

**Table 3-6: Frequency of cis-meQTLs for different significance thresholds for hippocampal brain tissue and blood cells.**

Threshold	Brain cis FDR	Blood cis FDR
Value < 5.00E-02	19,954	108,628
Value < 1.00E-10	1,899	24,346
Value < 1.00E-20	378	13,908
Value < 1.00E-30	84	9,177
Value < 1.00E-40	19	6,548
Value < 1.00E-50	1	4,891

Whereas the majority of cis-meQTLs seems to be consistent across tissues – 65% of the significant meQTLs in hippocampal brain tissue represent meQTLs in blood cells –also hippocampal brain tissue-specific meQTLs were observed. The 10 most significant tissue-specific cis-meQTLs for hippocampal brain tissue are shown in Table 3-7. An extra column also indicates the non-significant P-values of the respective CpG sites of the cis-meQTLs in blood cells.

**Table 3-7: Brain tissue-specific cis-meQTLs not found in blood cells.**

<b>CpG</b>	<b>Chr</b>	<b>CpG Pos</b>	<b>Gene Symbol</b>	<b>meQTL SNP</b>	<b>SNP Pos</b>	<b>FDR, brain</b>	<b>Rank cis-meQTL</b>	<b>P-value, blood</b>
cg09316607	10	1511024	<i>ADARB2</i>	rs1533486	1511786	5.52E-50	2	1.27E-03
cg04993605	1	28573052	<i>ATPIF1</i>	rs2938867	28515292	1.92E-45	5	4.46E-04
cg24657347	2	20261756	<i>AC098828.2(1.9k)</i>	rs7570953	20255015	6.25E-44	9	1.54E-03
cg13560919	6	33536144	<i>BAK1(4.2k)</i>	rs17627049	33537802	6.06E-34	52	2.14E-03
cg05315633	6	168628719	<i>RP11-503C24.2(1.5k)</i>	rs73028658	168623545	2.80E-33	60	7.16E-03
cg23020514	14	103360112	<i>TRAF3</i>	rs7147331	103380208	1.97E-31	75	1.13E-03
cg05431684	11	130781542	<i>SNX19</i>	rs73028875	130780595	9.17E-30	96	7.88E-04
cg16307866	4	7129517	<i>RN7SKP36(15.4k)</i>	rs4689604	7129556	1.03E-29	98	3.11E-03
cg07744270	11	64186719	<i>AP003774.5(1.1k)</i>	rs10897503	64186979	1.11E-29	99	9.58E-04
cg13144783	3	46249795	<i>CCR3</i>	rs2201150	46258902	1.06E-28	110	1.14E-03

Abbreviations: *P-value, blood*, refers to the most significant *P-value* for a SNP-CpG pair in blood cell samples ( $FDR < 0.05$ ); other abbreviations see Table 3-4. All annotations refer to genome build hg19.

## 3.1.2.1 Genomic distribution of cis-meQTLs

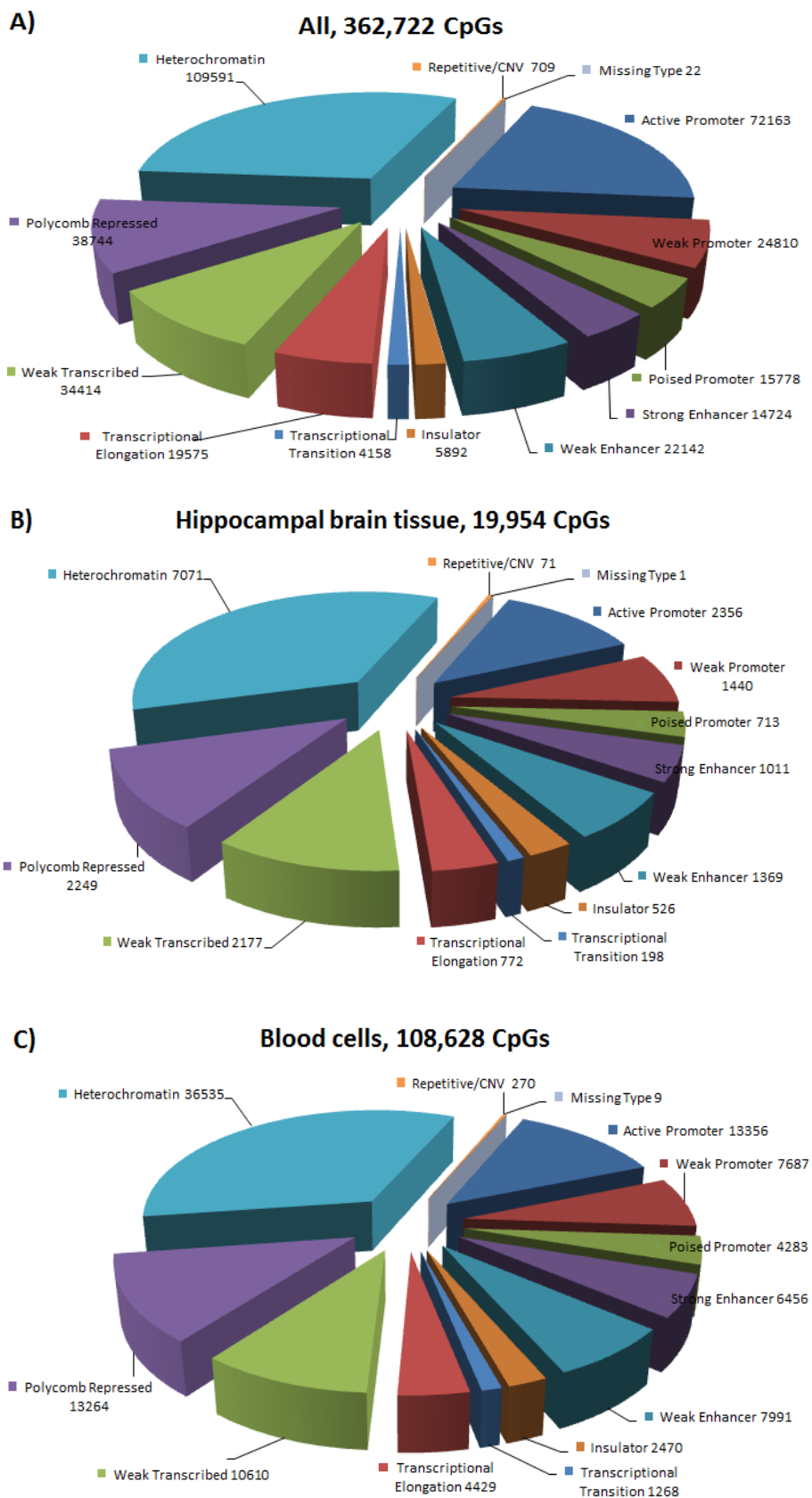
I compared the distribution of methylated sites associated with cis-meQTLs of all assayed 362k CpG methylation sites in B-lymphoblastoid cells (GM12878) and further compared the distribution of significant cis-meQTL associated CpGs between hippocampal brain tissue and blood cells. A strong significant depletion of sites associated with meQTLs at active promoters ( $P = 7.17 \times 10^{-101}$  for brain meQTLs and  $P < 10^{-300}$  for blood meQTLs), as well as a moderate enrichment at weak promoters ( $P = 7.16 \times 10^{-3}$  for brain meQTLs and  $P = 2.3 \times 10^{-4}$  for blood meQTLs) was found. In addition, a relative enrichment of meQTLs at insulators ( $P = 1.72 \times 10^{-26}$  for brain meQTLs and  $P = 7.07 \times 10^{-91}$  for blood meQTLs) and at enhancers ( $P = 4.63 \times 10^{-12}$  for brain meQTLs and  $P < 10^{-250}$  for blood meQTLs) were found. In particular, these results show that the distribution of the meQTLs obtained from blood cells display a similar distribution as those found in hippocampal brain tissue. An overlap of 90% of meQTL associated CpGs between hippocampal brain tissue and blood cells (Table 3-8 & Figure 3-4) was observed.

**Table 3-8: Genomic distribution of cis-meQTL associated CpGs in hippocampal brain tissue and blood cells.**

Specified genomic features from cell line GM12878 were counted for all quality-filtered 362,722 CpGs. Pearson's Chi-squared test with Yates' continuity correction was used to calculate the distribution of the genomic features of CpGs of significant cis-meQTLs ( $FDR < 0.05$ ) in hippocampal brain tissue and blood cell versus all 362k CpG sites. The last column represents the concordance of meQTL associated CpGs within the genomic feature between hippocampal brain tissue and blood cells. Red colored background indicates depletion, green colored background an enrichment of meQTL associated CpGs versus all CpGs.

Cell-type (GM12878)	All, 362k CpGs	cis-meQTL, brain	P-values, cis-meQTL, brain	cis-meQTL, blood	Brain vs. blood
Active Promoter	19.89%	11.81% ****	7.17E-101	12.30% ****	96%
Weak Promoter	6.84%	7.22% *	7.16E-03	7.08% *	98%
Poised Promoter	4.35%	3.57% **	5.05E-09	3.94% **	91%
Strong Enhancer	4.06%	5.07% **	4.63E-12	5.94% ****	85%
Weak Enhancer	6.10%	6.86% **	1.51E-06	7.36% ***	93%
Insulator	1.62%	2.64% **	1.72E-26	2.27% ***	86%
Transcriptional Transition	1.15%	0.99% *	6.45E-03	1.17%	85%
Transcriptional Elongation	5.40%	3.87% **	7.49E-23	4.08% ***	95%
Weak Transcribed	9.49%	10.91% **	2.45E-19	9.77% *	90%
Polycomb Repressed	10.68%	11.27% **	8.95E-08	12.21% ***	92%
Heterochromatin	30.21%	35.44% ***	6.40E-55	33.63% ***	95%
Repetitive/CNV	0.20%	0.36% **	2.47E-06	0.25% **	69%
Missing Type	0.01%	0.01%	3.70E-01	0.01%	100%

Chi-square test: \*\*\*\* $P < 10^{-100}$ ; \*\*\* $P < 10^{-50}$ ; \*\* $P < 10^{-5}$ ; \*  $P < 0.05$



**Figure 3-4: Functional localization of CpGs in GM12878.**

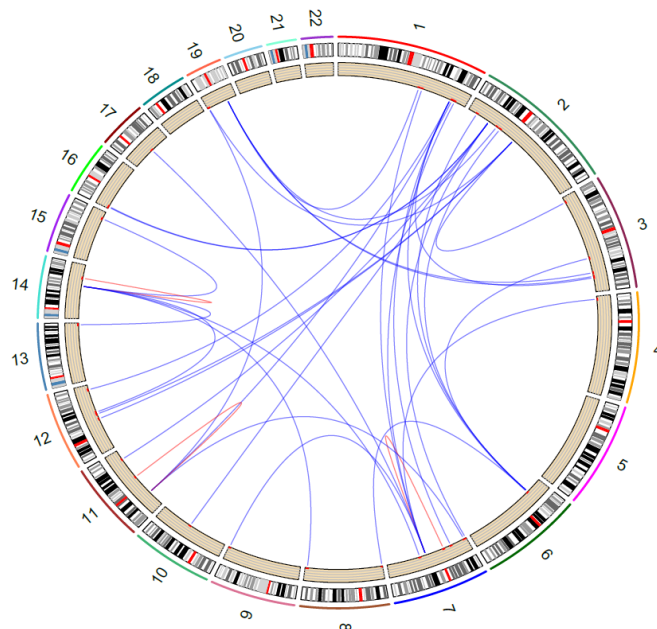
Distribution of different regulatory annotations in GM12878 for A) all 362,722 CpGs, B) the significant 19,954 meQTLs obtained from hippocampal brain tissue, and C) the significant 108,628 meQTLs obtained from blood cells.

### 3.1.3 Trans-meQTL analysis in hippocampal brain tissue

After stringent quality control, 362,722 CpGs and 643,195 LD-pruned SNPs were correlated for trans-meQTL analysis, resulting in  $2.33 \times 10^{11}$  tests corresponding to a nominal significance threshold of  $P = 1.93 \times 10^{-9}$  at a FDR of 5%. In addition to the CpG quality filters applied for meQTL analysis, a more stringent filter was used to sort out CpG probes that show cross-hybridization at other genomic locations (Zhang et al. 2014) and may lead to ambiguous CpG methylation states. Accordingly, trans-meQTLs analysis was carried out with 268,397 CpGs. At the FDR threshold of 5%, 3,927 out of 268,397 CpG sites account for significant trans-acting meQTLs ( $P$ -value range =  $2.93 \times 10^{-38}$  to  $1.93 \times 10^{-9}$ ).

As in previous studies comparing QTL  $P$ -value distributions in cis and trans, more strong associations in cis were found (Gibbs et al. 2010; Gamazon et al. 2013; Ramasamy et al. 2014). In total, 348 trans-meQTL SNPs were significantly associated with methylation levels of two or more genes. Among those, eight “master regulatory loci” with significant simultaneous impact on methylation level of four or more genes in trans were identified (Figure3- 5).

To eliminate long-range ( $>\pm 1$  Mbps) cis-acting meQTLs, it was checked whether trans-CpGs and markers were located on the same chromosome. This pattern was observed for 189 trans-meQTLs, but the trans-markers were more than 10 Mbps away from the CpG sites. Thus, for these SNPs, the effect on the regulation of genes in trans are unlikely to result from long-range cis-relationships.



**Figure 3-5: Circos plot of 115 hippocampal brain tissue for trans-meQTLs.**

Criteria for plotting trans-meQTLs: CpG to SNP distance  $> 10$  Mbps, stack size  $\geq 4$  CpGs per SNP and FDR corrected  $P$ -value ( $P < 0.05$ ), 1 Mbps frame in 100k sliding window. The  $-\log_{10}$   $P$ -values are plotted at CpG position.



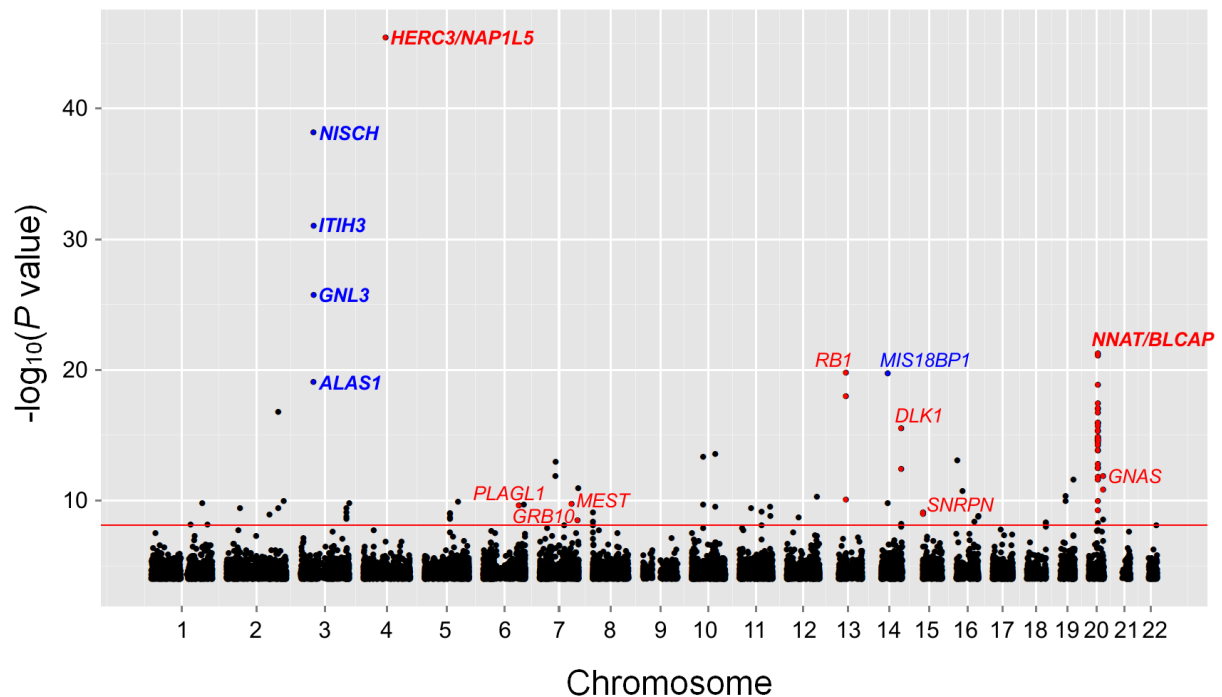
Except for two SNPs (rs114267096 and rs1273196), all trans-acting meQTL SNPs with simultaneous impact on the regulation of four or more genes were not significantly associated with the methylation of a gene located in cis in the underlying data. However, the effects in trans were barely significant; it is difficult to cover robust trans-meQTL findings.

Comparison of trans-meQTL results with other studies testing trans-associations was sparsely useful. Comparable studies assessed methylation levels of postmortem brain tissue with the HM27 BeadChips and utilized different CpG quality filters (Gibbs et al. 2010; Zhang et al. 2010; Numata et al. 2012). A table with the eight master regulatory loci with significant simultaneous impact on methylation level of four or more genes in trans can be found in Appendix, Table 6-3.

#### 3.1.4 Imprinting methylation QTL analysis in blood

In total, blood cell DNA samples of 269 parent-offspring trios with GGE were investigated to screen for differentially methylated parental genomic regions. Screening for imeQTLs was performed by comparing CpG methylation states of GGE trio offspring with nearby reciprocal heterozygote SNP genotypes of inverse parental origin ( $A_{MAT}/B_{PAT}$  vs.  $A_{PAT}/B_{MAT}$ ; Garg et al. 2012). Thus, any difference in the CpG methylation state detected between these two groups strongly suggests an underlying PofO effect. To specify the parental methylation status of an imeQTL, association analyses using only i) maternally inherited and ii) paternally inherited SNP alleles were carried out. The predominant parental CpG methylation state is indicated by the segregation of the alleles.

After stringent quality control, 362,722 CpGs and 886,110 LD-pruned SNPs were correlated for the imeQTL analysis, resulting in  $1.92 \times 10^8$  tests corresponding to a nominal significance threshold of  $P = 1.68 \times 10^{-7}$  at a FDR of 5%. At the FDR threshold of 5%, 177 out of 362,722 CpG sites were identified as autosomal imeQTLs ( $P$ -value range =  $3.47 \times 10^{-46}$  to  $1.68 \times 10^{-7}$ , Figure 3-6). QQ-plots of the CpG methylation in parent-offspring trios are shown in Appendix, Figure 6-2 and Figure 6-3.



**Figure 3-6: Manhattan plot for cis-imeQTLs of 269 trio offspring.**

The x-axis shows the genomic position of CpGs on chromosomes 1 to 22 (NCBI built 37.3, hg19); the y-axis shows the negative log<sub>10</sub> of the P-value per SNP/CpG pair. The red line represents the significance threshold of  $P < 1.68 \times 10^{-7}$  corresponding to a FDR of 5%. Red dots and in red highlighted gene names indicate CpGs in validated imprinted regions ( $P < 1.68 \times 10^{-7}$ ). In blue highlighted gene names indicate CpGs in novel imprinted regions which are shown in Table 3-9.

Genome-wide screening of imeQTLs revealed significant imeQTLs for 177 CpGs at 31 genomic loci ( $P < 1.68 \times 10^{-7}$ ). Altogether, nine known and 22 potentially novel imprinted regions were identified and subsequently specified for their predominant parental methylation state. In total, 29% of the imeQTLs ( $N = 52$ ) reside in nine known imprinting regions (Fang et al. 2012; Wei et al. 2014). The top-ranked imprinted loci were observed in the region 4q22.1 at the known imprinted *HERC3/NAP1L5* locus ( $P_{min} = 3.47 \times 10^{-46}$ ), at chromosome 20q11.23 at the known imprinted *NNAT/BLCAP* locus ( $P_{min} = 5.41 \times 10^{-22}$ ), and at a novel locus on chromosome 3p21.1. The 3p21.1 locus harbors GWAS candidate regions for among others bipolar disorder and major depressive disorder including the genes *NISCH/STAB1* ( $P_{min} = 6.56 \times 10^{-39}$ ), *ITIH3* ( $P_{min} = 9.24 \times 10^{-32}$ ) and *GNL3* ( $P_{min} = 1.71 \times 10^{-26}$ ).

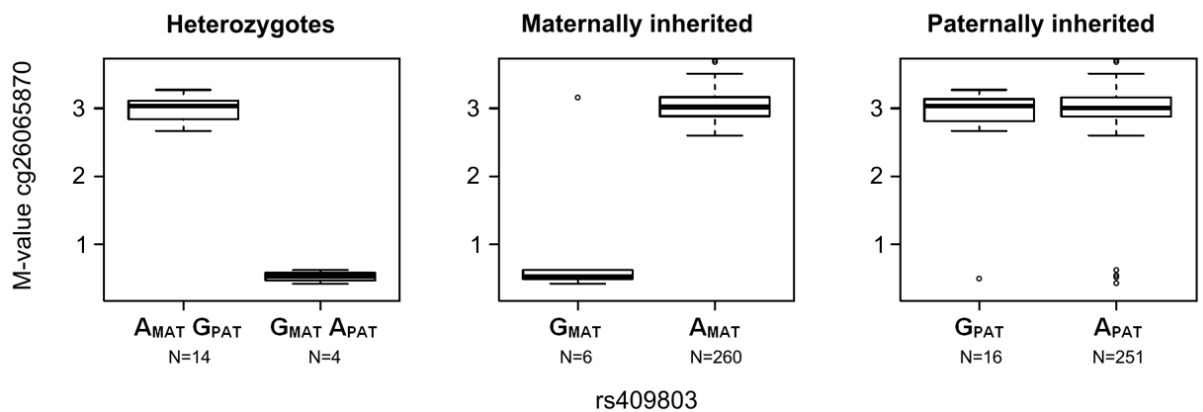
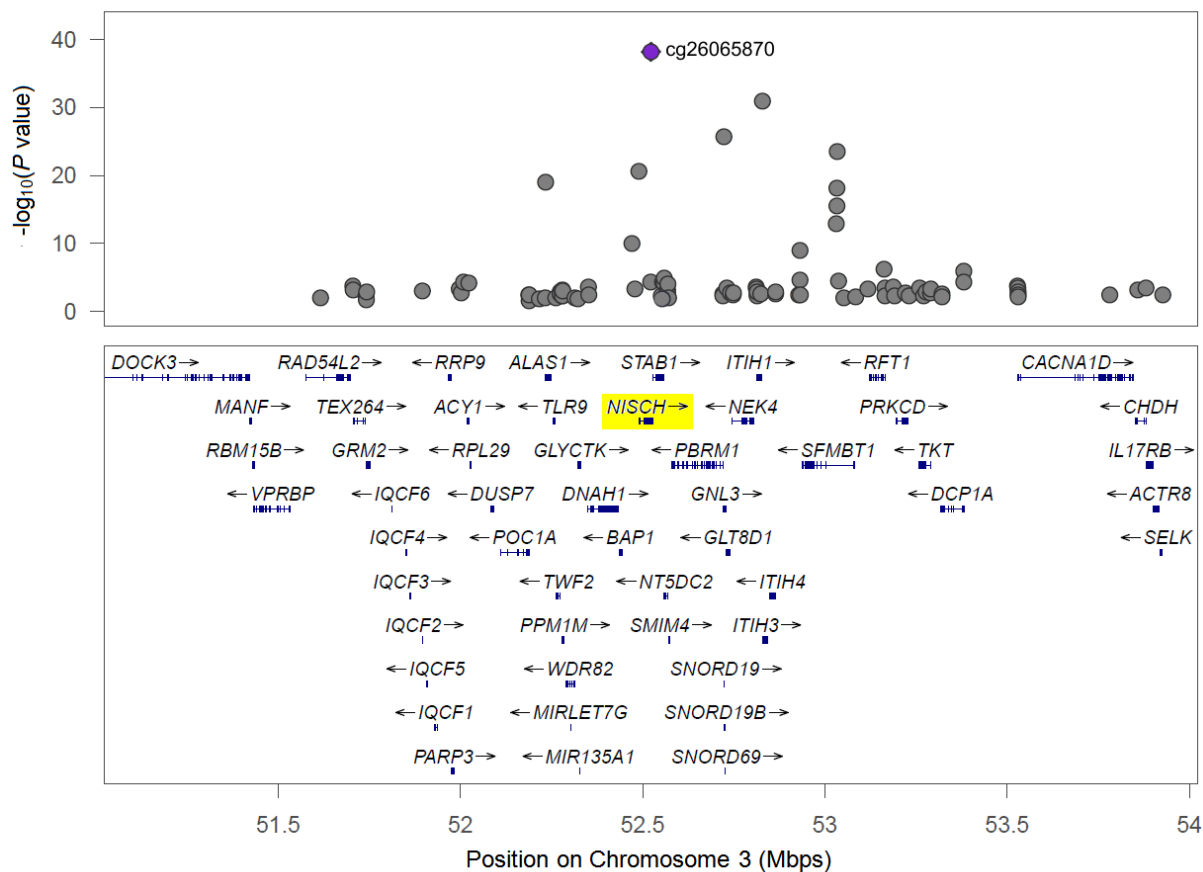
The 10 most significant imeQTLs including several high-ranking candidate genes emphasized for neurodevelopmental disorders are shown and listed in Table 3-9. The most significantly imeQTLs exhibited a combination of both; PofO and allele-specific effects on CpG methylation. The delineation of the parental methylation status of an imprinted CpG and the quantification of the allele-specific imeQTL effect is shown exemplary in Figure 3-7 for two top imeQTL signals; at chromosome 3p21.1 (*NISCH*) and at chromosome 20q11.23 (*NNAT/BLCAP*).

**Table 3-9: Top 10 imeQTLs and specification of their preferential parental expression.**

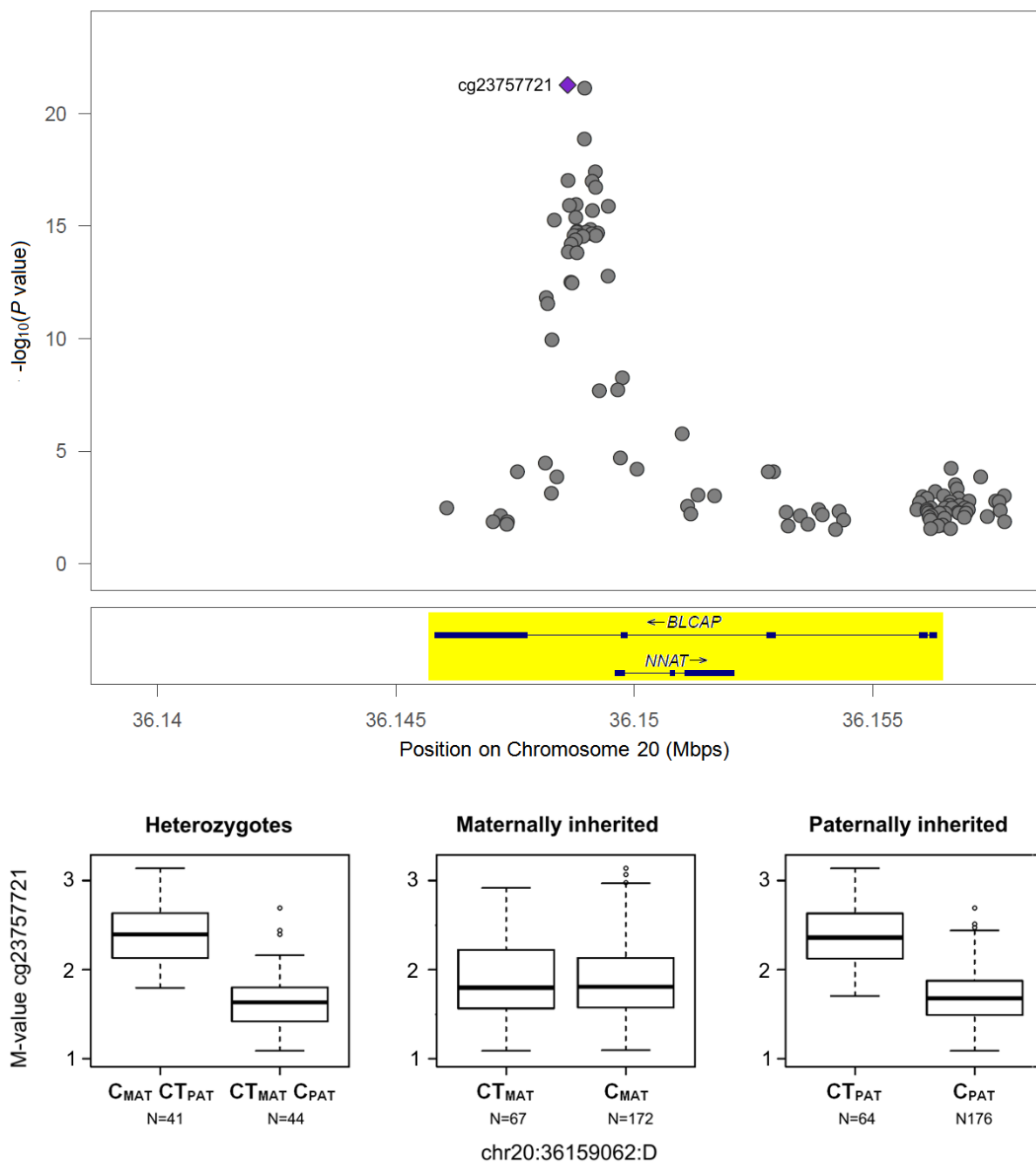
CpG	Chr	CpG Pos	Gene Symbol	imeQTL SNP	SNP Pos	FDR (het)	Impr locus	Meth allele
cg26065870	3	52522591	<i>NISCH</i>	rs409803	52443280	4.21E-31	Novel	Mat
cg18182844	3	52828292	<i>ITIH3(0.5k)</i>	rs409803	52443280	4.45E-24	Novel	Mat
cg00447581	3	52724578	<i>GNL3</i>	rs409803	52443280	6.58E-19	Novel	Mat
cg07598930	3	52233294	<i>ALAS1</i>	rs409803	52443280	1.34E-12	Novel	Mat
cg01026744	4	89619053	<i>HERC3/NAP1L5</i>	rs61737091	89671721	6.68E-38	Known	Pat
cg11408952	13	48892244	<i>RB1</i>	rs3825417	48891836	3.84E-13	Known	Pat
cg02273647	14	45722745	<i>MIS18BP1</i>	rs114370731	45386388	4.12E-13	Novel	Pat
cg23757721	20	36148604	<i>NNAT/BLCAP</i>	chr20:36159062:D	36159062	1.74E-14	Known	Pat
cg08402058	20	36148961	<i>NNAT/BLCAP</i>	chr20:36159062:D	36159062	2.06E-14	Known	Pat
cg14765818	20	36148954	<i>NNAT/BLCAP</i>	chr20:36159062:D	36159062	1.98E-12	Known	Pat

Abbreviations: *FDR (het)*, FDR-value of the heterozygote test; *Impr locus*, indicating if the imeQTL is at a known or novel imprinting locus. The predominant parental CpG methylation state is indicated by the parental methylated allele identified by the transmission analysis ('Meth allele'), whereas 'Pat' indicates the paternally-derived allele and 'Mat' indicates the maternally-derived allele. All annotations refer to the genome build hg19.

A)



B)



**Figure 3-7: Quantification of the allele-specific imeQTL effect.**

The upper plots present regional detail plots of the negative  $\log_{10}$  of P-values of the imeQTL analysis; the x-axes show the genomic position of CpGs at a specific locus. Genes close to the detected association are presented as well. The lower box-plots show the distribution of the methylation states for a specific CpG. In A) a regional detail plot of the negative  $\log_{10}$  P-value for CpG cg26065870 on chromosome 3p21.1 and a box-plot of the paternally imprinted CpG cg26065870 at the NISCH locus is shown; in B) a regional detail plot of the negative  $\log_{10}$  P-value for CpG cg23757721 on chromosome 20q11.23 and a box-plot of the maternally imprinted CpG cg23757721 at the NNAT/BLCAP locus.

The boxplots show the distribution of the methylation states for the CpGs, respectively with respect to maternally and paternally inherited SNPs. The x-axis shows the genotypes of the parental origin of the heterozygous SNP alleles, respectively: for example in A)  $G_{mat}$  and  $A_{mat}$  for maternally inherited SNPs and  $G_{pat}$  and  $A_{pat}$  for paternally inherited SNPs. At the y-axis the methylation level, expressed as M-value of the CpGs is shown. The M-value ranges among +3 and -3 whereas a positive M-value indicates that the allele is highly methylated and therefore inactivated. An allele-specific effect was observed. Furthermore the observed difference between heterozygous individuals of reciprocal parental origin indicated a PofO effect at these loci, as these individuals showed a difference in methylation despite having the same genotype.

The region of chromosome 3p21.1 is of special interest as several studies listed in the NHGRI GWAS catalog found genome-wide significant evidence that SNPs in this region were associated with various neuropsychiatric diseases including major mood disorder (composed of bipolar disorder and major depressive disorder), autism spectrum disorder, attention-deficit hyperactivity disorder and schizophrenia. Additionally, Witt and colleagues (2014) identified *STAB1* as the potentially causative gene for bipolar disorder. This gene is located in close proximity to known mood disorder genes of the 3p21.1 region including *PRBM1* (McMahon et al. 2010) and the *NEK4-ITIH1-ITIH3-ITIH4* multigene region (Psychiatric GWAS Consortium Bipolar Disorder Working Group 2012).

Further, the imeQTL results in 269 blood cell trios were compared with a recently published eQTL study in highly divergent mouse crosses (Crowley et al. 2015). Crowley and co-workers (2015) observed that classical imprinting is under genetic control and found a global allelic imbalance in expression favoring the paternal allele; imprinted genes were 1.5 times more likely to be expressed from the paternal than the maternal allele. My imeQTL results were consistent with these observations. A 1.8 times increased likelihood of imprinted genes to be expressed from the paternal than the maternal allele was observed (79 among 123 divergent parental imeQTLs were expressed from the paternal allele, 64%).

In total, 138 among 177 imeQTL CpGs were located intragenic of 78 Ensembl genes (hg19). Among these 78 genes, eight (10%) were consistent with the imprinted genes in mouse brain reported by Crowley et al. (2015). Yet, half ( $N = 4$ ) of the overlapping genes were in the top 55 of the heterozygous imprinting analysis. All overlapping genes were detected by a threshold of  $P < 1.68 \times 10^{-7}$  (Table 3-10). Additionally, by using the paternal-specific association analysis the *GABRB3* locus which is known to be involved in epileptogenesis (Urak et al. 2006; Pernhorst et al. 2011) was detected. This locus showed evidence of association with paternally inherited alleles but no association with maternally inherited alleles. *GABRB3* was also included in the imprinting gene list of Crowley et al. (2015) as a paternally expressed allele.

**Table 3-10: Eight imeQTLs with specification according to their parental expression loci consistent with Crowley et al. (2015).**

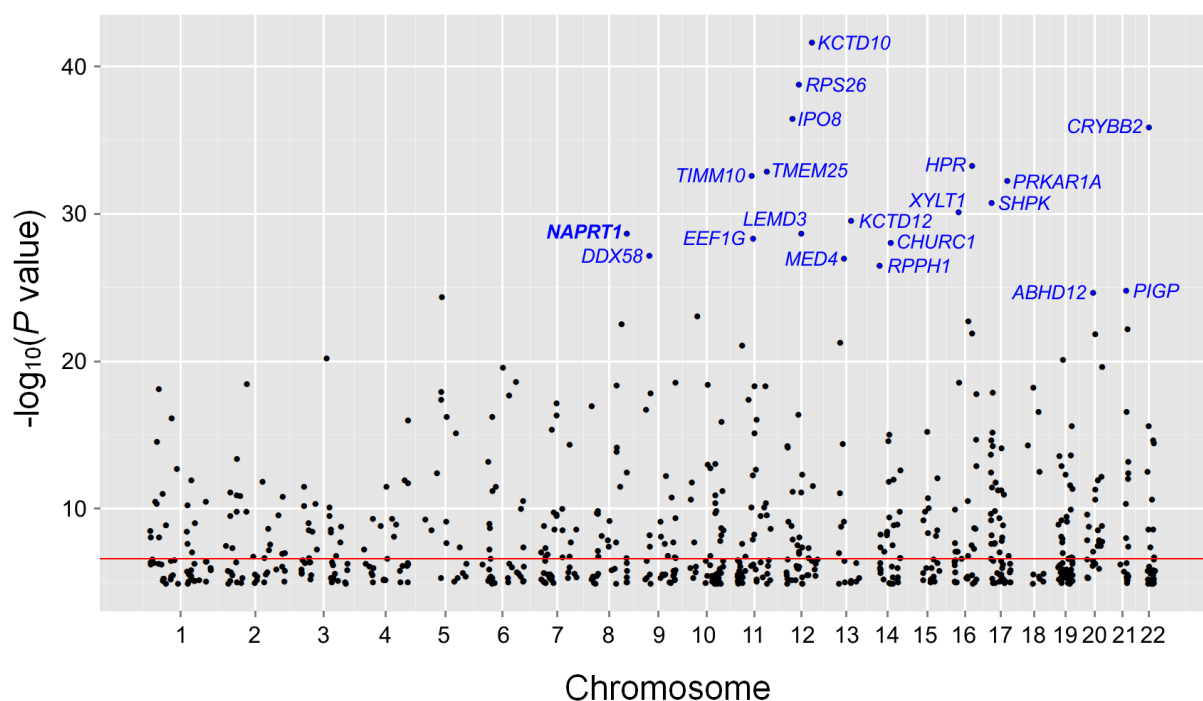
CpG	Chr	CpG Pos	Gene Symbol	imeQTL SNP	SNP Pos	FDR (het)	Rank imeQTL	Impr locus	Meth allele	P-value, Mat	P-value, Pat
cg01026744	4	89619053	<i>HERC3/NAP1L5</i>	rs61737091	89671721	6.68E-38	1	Known	Pat	5.88E-04	3.55E-164
cg23757721	20	36148604	<i>NNAT/BLCAP</i>	chr20:36159062:D	36159062	1.74E-14	5	Known	Pat	8.78E-03	4.58E-39
cg18279536	14	101194748	<i>DLK1</i>	rs10144381	101173297	3.09E-16	21	Known	Pat	9.29E-05	2.59E-14
cg17696847	20	57414217	<i>GNAS</i>	rs6070619	57383157	1.27E-05	52	Known	Mat	3.24E-19	2.42E-07
cg05260959	7	130133110	<i>MEST</i>	rs992859	130120122	1.29E-04	64	Known	Pat	1.06E-03	3.68E-17
cg23460430	6	144329887	<i>PLAGL1</i>	rs55769736	144333557	1.53E-04	67	Known	Pat	2.50E-03	1.56E-18
cg27644292	15	25123287	<i>SNRPN</i>	rs73371266	25128475	6.04E-04	80	Known	Pat	5.05E-04	4.94E-22
cg16349612	7	50849723	<i>GRB10</i>	rs1019002	50810582	1.85E-02	141	Known	Pat	3.28E-04	3.94E-13

Abbreviations: FDR (het), FDR-value of the heterozygote test; Impr locus, indicating if the imeQTL is at a known or novel imprinting locus. The predominant parental CpG methylation state is indicated by the parental methylated allele identified by the transmission analysis ('Meth allele'), whereas 'Pat' indicates the paternally-derived allele and 'Mat' indicates the maternally-derived allele. All annotations refer to genome build hg19.

## 3.2 Expression QTL analyses

### 3.2.1 Cis-eQTL analysis in hippocampal brain tissue

The eQTL analysis examined the influence of cis-acting genetic variations on the quantitative state of mRNA gene expression using the *Illumina HumanHT-12 v3* (HT12v3) array. After stringent quality control, 31,146 mRNA probes and 643,195 LD-pruned SNPs were correlated for cis-eQTL analysis, resulting in  $8.46 \times 10^6$  tests corresponding to a nominal significance threshold of  $P = 2.10 \times 10^{-5}$  at a FDR of 5%. At the FDR threshold of 5%, 734 (2.4%) out of 31,146 mRNA probes account for significant cis-acting eQTLs ( $P$ -value range =  $2.34 \times 10^{-42}$  to  $2.10 \times 10^{-5}$ , Figure 3-8). A QQ-plot of the mRNA expression in hippocampal brain tissue samples is shown in Appendix, Figure 6-4.



**Figure 3-8: Manhattan plot for cis-eQTLs identified in 115 hippocampal brain tissue samples.**

The x-axis shows the genomic position of gene transcript start sites on chromosomes 1 to 22 (NCBI built 37.3, hg19); the y-axis shows the negative log<sub>10</sub> of the P-value per SNP/mRNA pair. The red line represents the threshold for significant cis-eQTLs with FDR < 0.05. The genes of the 20 most significant cis-eQTLs (shown in Table 3-12) are highlighted.

cis-eQTL confirmation analyses were performed by comparing my cis-eQTL analysis with three recent cis-eQTL studies investigating diverse postmortem brain tissues (Gibbs et al. 2010; Kim et al. 2014; Ramasamy et al. 2014). All studies used different assays to profile mRNA transcript in comparison to my analysis (for an overview on studies that are frequently mentioned for comparison reasons, please refer to Table 6-2 in Appendix). Therefore, the significant cis-eQTL SNPs from all three studies had to be within 1 Mbps of my mRNA transcription site and further their SNPs had to be in LD of  $r^2 > 0.2$  with my cis-eQTL SNPs. In total, an overlap of 14% was observed in the significant cis-eQTLs



obtained from fresh frozen hippocampal brain tissue with those generated by Gibbs et al. (2010). Comparing the results of the present study to those found in the eQTL meta-analysis in brain cells (Kim et al. 2014) an overlap of 65% (474 out of 734) cis-eQTLs was observed. But the best overlap was detected by comparing results to those found in the eQTL analysis in ten regions of the human brain (Ramasamy et al. 2014): an overlap of 71% (522 out of 734) cis-eQTLs was observable, considering that a cis-eQTL signal was identified in more than one brain region. Considering only the hippocampal brain region, an overlap of 27% (197 out of 734) cis-eQTLs was identified (Table 3-11).

In addition, Schramm and co-workers (2014) performed a whole-genome eQTL analysis in blood cell samples using the HT12v3 expression array, and the meta-eQTL analysis in blood cells carried out by Westra and colleagues (2013) was restricted to probes present on the HT12v3 array. I found that 50% of cis-eQTLs in hippocampal brain tissue matched to the cis-eQTLs in blood cells detected by Schramm et al. (2014), and 58% matched with those reported by Westra et al. (2013).

**Table 3-11: Cis-eQTL overlaps with comparable human eQTL studies.**

<b>Epigenetic study</b>	<b>Cell-type or tissue studied</b>	<b>Overlap with cis-eQTL (FDR &lt; 0.05)</b>
Gibbs et al. (2010)	Postmortem brain tissue	14% (105 out of 734)
Kim et al. (2014)	Postmortem brain tissue	65% (474 out of 734)
Ramasamy et al. (2014)	Postmortem brain tissue	71% (522 out of 734)
Schramm et al. (2014)	Whole Blood	50% (367 out of 734)
Westra et al. (2013)	Whole Blood	58% (423 out of 734)

The 20 most significant cis-eQTLs for 115 hippocampal brain tissue are shown in Table 3-12, including columns confirming the cis-eQTL results of the present study in diverse postmortem brain tissues (Gibbs et al. 2010; Kim et al. 2014; Ramasamy et al. 2014) as well as in blood cells (Westra et al. 2013; Schramm et al. 2014).

**Table 3-12: Top hits of cis-eQTLs in hippocampal brain tissue.**

mRNA probe	mRNA Location	Gene Symbol	eQTL SNP	SNP Pos	FDR	Detec P-value > 0.05	P-value (Brain <sup>1</sup> )	P-value (Brain <sup>2</sup> )	P-value (Brain <sup>3</sup> )	P-value (Blood <sup>4</sup> )	P-value (Blood <sup>5</sup> )
ILMN_1719064	12:109886649	KCTD10	rs4766601	109890080	3.32E-35	115	5.02E-47	7.61E-09	1.25E-23	1.91E-230	9.81E-198
ILMN_2209027	12:56436251	RPS26	rs10876864	56401085	1.17E-32	115	NA	6.99E-06	1.43E-15	NA	NA
ILMN_1753164	12:30782283	IPO8	rs10771758	30856139	1.66E-30	111	NA	6.53E-08	9.14E-19	< 1E-270	9.81E-198
ILMN_1760708	22:25625496	CRYBB2	rs113130491	25861797	4.83E-30	110	NA	NA	NA	NA	NA
ILMN_1809212	16:72110608	HPR	rs2000999	72108093	1.68E-27	112	3.64E-10	1.06E-26	NA	NA	NA
ILMN_1651745	11:118406322	TMEM25	rs56403966	118425051	3.37E-27	81	NA	5.42E-08	NA	NA	1.12E-04
ILMN_1765332	11:57296041	TIMM10	rs2649652	57324428	5.52E-27	115	2.89E-22	1.58E-14	8.78E-32	2.25E-198	9.81E-198
ILMN_2389590	17:66528779	PRKAR1A	rs2952272	66524565	1.05E-26	109	NA	1.68E-07	NA	1.42E-190	9.81E-198
ILMN_2216918	17:3511949	SHPK	rs62069929	3515275	2.78E-25	115	NA	1.11E-14	4.81E-14	1.66E-36	NA
ILMN_1830462	16:17196448	XYLT1	rs9934313	17198524	1.14E-24	84	NA	NA	NA	1.98E-174	NA
ILMN_2229649	13:77454497	KCTD12	rs73544577	77434063	4.02E-24	20	NA	2.87E-08	NA	NA	3.76E-78
ILMN_1710752	8:144656991	NAPRT1	rs9657360	144676862	2.20E-23	114	3.27E-18	2.18E-30	6.23E-32	3.24E-163	9.81E-198
ILMN_2183938	12:65641975	LEMD3	rs193244911	65622186	2.20E-23	115	NA	NA	NA	6.88E-239	9.81E-198
ILMN_2262288	11:62334905	EEF1G	chr11:62369880:l	62369880	4.25E-23	115	NA	2.30E-23	NA	1.91E-254	NA
ILMN_1798177	14:65401520	CHURC1	rs4902345	65403904	8.00E-23	115	5.77E-41	1.92E-29	3.79E-37	< 1E-270	9.81E-198
ILMN_1851020	9:32454228	DDX58(1.0k)	rs1360171	32454348	5.43E-22	44	NA	NA	NA	NA	NA
ILMN_1664641	13:48650456	MED4	rs4942723	48525457	7.31E-22	115	NA	3.69E-07	4.49E-09	1.73E-239	9.81E-198
ILMN_1704056	14:20811256	RPPH1	rs79030532	20814865	1.99E-21	83	NA	7.02E-08	NA	NA	NA
ILMN_1774949	21:38445331	PIGP	rs1793870	38481436	6.26E-20	104	NA	3.78E-12	1.75E-24	3.42E-75	NA
ILMN_1745116	20:25275591	ABHD12	rs6107052	25530838	8.38E-20	115	NA	7.52E-39	7.11E-10	6.61E-46	9.81E-198

Abbreviations: mRNA probe, Illumina annotation of the mRNA probe; mRNA Location, chromosome of the mRNA probe and the average of the start and end position of the mRNA probe; Gene Symbol, gene which contains the mRNA probe; eQTL SNP, SNP of most significant eQTL of a specific mRNA probe; SNP Pos, base-pair position of SNP; FDR, most significant FDR for a SNP/mRNA pair; Detec P-value > 0.05, number of samples with detection P-value > 0.05; P-value Brain1, cis-eQTL P-value according to Gibbs et al. (2010); P-value Brain2, P-value according to Ramasamy et al. (2014); P-value Brain3, cis-eQTL P-value according to Kim et al. (2014), meta-analysis; P-value Blood4, cis-eQTL P-value according to Schramm et al. (2014); P-value Blood5, cis-eQTL P-value according to Westra et al. (2013), meta-analysis. All annotations refer to genome build hg19.

### 3.2.1.1 Enrichment analysis of cis-eQTLs

Enrichment analyses for the cis-eQTLs were performed using the same gene-sets as for the cis-meQTL enrichment analysis; brain expressed genes (N = 8,852; Pinto et al. 2014), genes involved in neurodevelopmental disorders (N = 1,689) and autism spectrum disorder (N = 1,559), as well as the previously described Brain-Disorder-GWAS gene-set (N = 1,200). The analysis revealed no significant enrichment in the gene-sets. Here, the significant enrichment in brain expressed genes ( $P = 1.35 \times 10^{-12}$ ) could be seen as a kind of validation for performing the enrichment analysis in these cis-eQTLs.

### 3.2.2 Trans-eQTL analysis in hippocampal brain tissue

After stringent quality control, 31,146 mRNA probes and 643,195 LD-pruned SNPs were correlated for trans-eQTL analysis, resulting in  $1.68 \times 10^{10}$  tests corresponding to a nominal significance threshold of  $3.79 \times 10^{-11}$  at a FDR of 5%. At the FDR threshold of 5%, 16 significant trans-acting eQTLs were observed ( $P$ -value range =  $3.06 \times 10^{-35}$  to  $3.79 \times 10^{-11}$ ). Overall, 15 of the trans-eQTLs had SNPs from different chromosomes than the respective mRNA probe while one was on the same chromosome but more than 70 Mbps away from the target mRNA probe. Thus, there was no evidence for long-range cis-eQTLs.

With regard to the rare trans-eQTL findings, only four trans-eQTL SNPs were significantly associated with expression levels of two genes (Table 3-13). None of the obtained trans-eQTL genes overlap with those reported in previously published eQTL studies (Gibbs et al. 2010; Westra et al. 2013; Ramasamy et al. 2014; Schramm et al. 2014).

**Table 3-13: Master regulatory sites – trans-eQTL SNPs with simultaneous impact on the expression of two genes.**

SNP	Chr SNP	Gene of SNP	SNP Pos	mRNA	Gene of mRNA	mRNA Location	FDR
rs10876864	12	<i>IKZF4(0.4k)</i>	56401085	ILMN_1678522	<i>LOC644934</i>	15:64885219	6.12E-25
	12	<i>IKZF4(0.4k)</i>	56401085	ILMN_2310703	<i>RPS26L</i>	13:101192220	3.30E-24
rs17718169	5	<i>ATP6VOE1(4.9k)</i>	172405897	ILMN_1666845	<i>KRT17</i>	17:39775733	3.89E-06
	5	<i>ATP6VOE1(4.9k)</i>	172405897	ILMN_1689515	<i>CPLX3</i>	15:75123999	5.41E-03
rs2069408	12	<i>DGKA(16.5k)</i>	56364321	ILMN_2310703	<i>RPS26L</i>	13:101192220	8.81E-05
	12	<i>DGKA(16.5k)</i>	56364321	ILMN_1678522	<i>LOC644934</i>	15:64885219	1.92E-02
rs12340642	9	NA	79893032	ILMN_1713668	<i>TSNAX</i>	1:231700328	1.09E-02
	9	NA	79893032	ILMN_1718712	<i>C20orf177</i>	20:58522158	4.13E-02

Abbreviations see Table3-12.

### 3.3 Delineation of methylation-driven gene expression in hippocampal brain tissue

#### 3.3.1 Correlation analysis of CpG methylation with gene expression

By correlating local CpG methylation with mRNA gene expression, 565 CpG-expression probe pairs (comprising 565 CpG sites and 292 expression probes of 378 genes) displayed a significant correlation ( $P$ -value range =  $7.58 \times 10^{-30}$  to  $2.38 \times 10^{-6}$ ), of which 152 were positively correlated (Pearson correlation coefficient  $r = +0.42$  to  $+0.80$ ) and 413 showed an inverse correlation ( $r$  ranged from  $-0.83$  to  $-0.42$ ).

Overall, 27% (152 of 565) of the significant CpG-mRNA pairs were correlated in the same direction (positive correlation). Namely, a pattern whereby high methylation levels were associated with high expression levels was observed. This pattern was observed both for methylation sites located within and outside gene bodies. Yet, CpG sites whose methylation levels were positively correlated with expression levels of nearby genes were more distant from the gene's TSS than CpG sites whose methylation levels were inversely correlated with expression levels of nearby genes (median distance of 8,954 bps and 5,766 bps for positive correlation and inverse correlation, respectively;  $P = 0.038$ ).

In total, 73% (413 of 565 sites) of the significant cis CpG-expression probe pairs were observed with an inverse correlation, demonstrating that high methylation states result in low mRNA expression levels. When considering only those CpG-mRNA probe pairs where the CpG site was within a CpG island (as defined in Gardiner-Garden & Frommer 1987) 96 CpG sites were identified. Of those 96 CpG sites, 71% (68 out of 96) of the CpG-mRNA probe pairs were inversely correlated.

The correlation analysis of CpG methylation to gene expression demonstrated that the CpG sites affecting mRNA expression were enriched in gene bodies ( $P = 1.85 \times 10^{-3}$ ) and at strong enhancers ( $P = 3.41 \times 10^{-38}$ ). A depletion of CpG-mRNA probe pairs compared to all 362k CpG sites was observed in the TSS200 region ( $P = 0.04$ ). Interestingly, further differences between positive and inverse correlation of the CpG-expression probe pairs were observed. As mentioned, I was able to identify enrichment in gene bodies but did not observe a difference in the direction of correlations between CpGs within or outside gene bodies. When comparing all 362k CpGs sites with the distribution of 565 correlated CpG-mRNA probe pairs at active promoters, no significant difference was observed. But when comparing positive with inverse correlation at active promoters a significant difference was observable ( $P = 9.43 \times 10^{-3}$ ): CpG-expression probe pairs showing inverse correlation displayed enrichment at active promoters, CpG-expression probe pairs showing positive correlation displayed depletion at active promoters. In addition, the same distribution was observed for the TSS200 region where a moderate difference between the correlation types was detected ( $P = 0.01$ ) and at strong

enhancers which revealed the strongest difference between the correlation types ( $P = 1.58 \times 10^{-4}$ ; Table 3-14).

**Table 3-14: Genomic distribution of CpGs influencing gene expression in gene features in hippocampal brain tissue.**

First table: proportion of all quality-filtered 362,722 CpGs in gene features. Pearson's Chi-squared test with Yates' continuity correction was used to calculate the distribution of 565 CpG-mRNA probe pairs ( $FDR < 0.05$ ) in different gene features in hippocampal brain tissue compared to all 362k CpG sites. Red colored background indicates depletion, green colored background an enrichment of CpG-mRNA probe pairs compared to all 362k CpGs. The second table shows the genomic distribution of positive or inverse correlated CpG-mRNA probe pairs. Differences of the correlation types are highlighted.

Illumina Group	All, 362,722 CpGs	Correlation CpG-mRNA	P-values, correlation
3'UTR	3.72%	3.19%	5.74E-01
5'UTR	9.15%	11.33%	8.50E-02
Gene Body	34.06%	40.35%*	1.85E-03
TSS200	10.98%	8.14%*	3.70E-02
TSS201-1500	14.62%	17.35%	7.50E-02
Active Promoter <sup>1</sup>	19.89%	20.35%	8.25E-01
Strong Enhancer <sup>1</sup>	4.06%	14.87%***	3.41E-38
Insulator <sup>1</sup>	1.62%	1.24%	5.76E-01

<sup>1</sup>These genomic features are from cell line GM12878.

Chi-square test: \*\*\* $P < 10^{-30}$ ; \*\* $P < 10^{-5}$ ; \*  $P < 0.05$

Illumina Group	Correlation CpG-mRNA	Inverse correlation	Positive correlation	P-values, correlation
3'UTR	3.19%	2.66%	4.61%	2.80E-01
5'UTR	11.33%	11.14%	11.84%	8.81E-01
Gene Body	40.35%	39.47%	42.76%	4.99E-01
TSS200	8.14%	9.20%*	5.26%*	1.00E-02
TSS201-1500	17.35%	16.46%	19.74%	3.81E-01
Active Promoter <sup>1</sup>	20.35%	23.00%*	13.16%*	9.43E-03
Strong Enhancer <sup>1</sup>	14.87%	18.16%*	5.92%*	1.58E-04
Insulator <sup>1</sup>	1.24%	0.73%	2.63%	8.80E-02

<sup>1</sup>These genomic features are from cell line GM12878.

Chi-square test: \*\*\* $P < 10^{-30}$ ; \*\* $P < 10^{-5}$ ; \*  $P < 0.05$

To further explore the correlation of CpG methylation with mRNA gene expression the allelic status of these was tested. The significant CpG-mRNA probe pairs were selected where both the CpG methylation site and the mRNA transcript probe had a positional overlapping cis-meQTL or cis-eQTL finding. Overall, 102 of 565 CpG-expression probe pairs (comprising 102 CpG sites and 45 expression probes) had a positional overlapping cis-meQTL and cis-eQTL finding. For only one of the 102 correlations of CpG methylation with mRNA gene expression, neither the same meQTL-eQTL SNP nor a meQTL SNP in LD ( $r^2 \geq 0.2$ ) with an eQTL SNP was found. In total, for 18% of the CpG methylation-expression correlation the CpG methylation and – presumably indirect – the RNA expression was significant genetically determined.

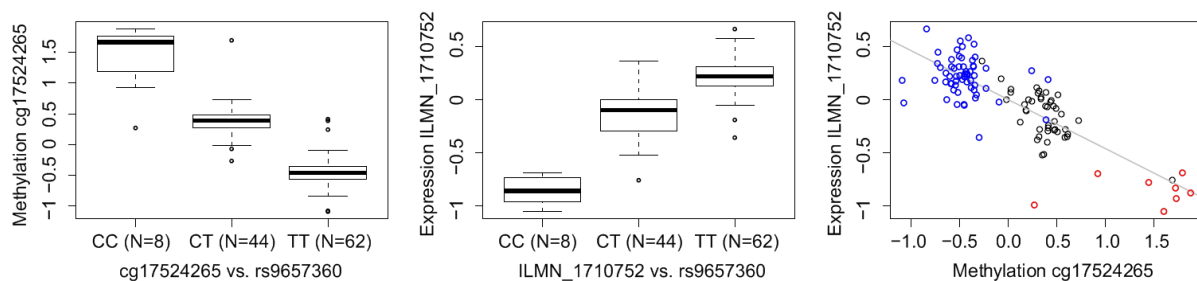
The 10 most significant correlations of CpG methylation with mRNA gene expression and their associated significant cis-meQTLs and cis-eQTLs are shown in Table 3-15. In total, 44 genes showed SNP-methylation-expression three-way associations; the same SNP (or two times an LD SNP) simultaneously showed significant association with both DNA methylation and gene expression, while DNA methylation was significantly correlated with gene expression. The 44 genes involved 101 CpG sites, 44 expression probes, and 340 SNPs.

**Table 3-15: Top hits of CpG methylation-driven gene expression in hippocampal brain tissue.**

Probe	CpG	Chr	Gene Symbol	R	FDR	CpG Pos	mRNA Location	Correlated SNP	SNP Pos	FDR, cis-meQTL	FDR, cis-eQTL
ILMN_1710752	cg17524265	8	<i>NAPRT1</i>	-0.83	1.06E-22	144659883	8:144656991	rs9657360	144676862	3.20E-33	2.20E-23
ILMN_1730054	cg111141652	22	<i>GSTT1</i>	0.80	4.44E-20	24348549	22:24376401	rs738809	24405492	1.42E-06	1.88E-11
ILMN_2326324	cg07415359	11	<i>LDHC</i>	-0.80	8.07E-20	18434354	11:18472557	rs4757651	18417583	2.20E-27	1.55E-16
ILMN_1736184	cg10807101	1	<i>GSTM5</i>	-0.69	1.76E-11	110282274	1:110276633	rs1887547	110295772	2.91E-15	8.75E-09
ILMN_2173294	cg04051697	2	<i>THNSL2</i>	-0.66	8.56E-10	88470813	2:88486079	rs7582011	88452740	1.67E-06	4.64E-14
ILMN_1746646	cg24786847	15	<i>CHRM5</i>	-0.64	6.87E-09	34261172	15:34357171	rs618868	34293378	1.35E-06	1.05E-05
ILMN_1899338	cg19520046	6	<i>KIF25</i>	-0.63	2.26E-08	168397360	6:168376777	rs11752225	168409629	1.53E-04	7.32E-07
ILMN_1732089	cg25755428	19	<i>MRI1(0.1k)</i>	-0.62	5.50E-08	13875111	19:13884988	rs371671	13870153	5.32E-27	5.53E-09
ILMN_1774949	cg21832243	21	<i>PIGP</i>	0.61	3.21E-07	38444223	21:38445331	rs1793870	38481436	3.61E-14	6.26E-20
ILMN_1809212	cg23815491	16	<i>HPR</i>	0.59	1.14E-06	72088622	16:72110608	rs2000999	72108093	1.10E-10	1.68E-27

Abbreviations: R, Pearson correlation coefficient; Genes in the 'Gene Symbol' column indicate the gene of the mRNA probe. All annotations refer to genome build hg19.

One of these genes encodes the Nicotinate Phosphoribosyltransferase (*NAPRT1*). A cis-eQTL regulating *NAPRT1* expression has been reported in several meta-eQTL analyses in brain cells (Gibbs et al. 2010; Kim et al. 2014), in whole blood cells (Westra et al. 2013; Battle et al. 2014) as well as in CD14<sup>+</sup> monocytes (Raj et al. 2014; Gjoneska et al. 2015). The inverse correlation ( $r$ -value range = -0.83 to -0.72) of seven successive CpG probes in the promoter region of *NAPRT1* with the *NAPRT1* 3' expression probe ILMN\_1710752 was the strongest methylation-driven expression in the present study ( $P$ -value range =  $7.58 \times 10^{-30}$  to  $7.0 \times 10^{-20}$ ). The seven CpG probes (cg17524265, cg08017634, cg16316162, cg13282195, cg19357499, cg01494348, cg21187068; chr8:144,659,627-144,660,772) flank the TSS of *NAPRT1*. Five of these CpGs reside in a CpG island (chr8:144,659,745-144,660,635; assigned as CpG: 94). Two additional CpGs are located in the 2 kb region flanking CpG island shores. Moreover, the *NAPRT1*-associated meQTLs and eQTL were genetically regulated by the same SNP rs9657360 ( $P_{min} = 1.61 \times 10^{-39}$  and  $P = 2.15 \times 10^{-29}$ , respectively). The minor C allele of SNP rs9657360 was significantly associated with high methylation levels in the *NAPRT1* promoter region and simultaneously associated with low gene expression of *NAPRT1*. The box plot of methylation (cg17524265) and expression (ILMN\_1710752) in the *NAPRT1* gene with SNP rs9657360 is shown in Figure 3-9 as well as the significant inverse correlation between methylation and expression. The methylation and expression levels are represented by their residuals.



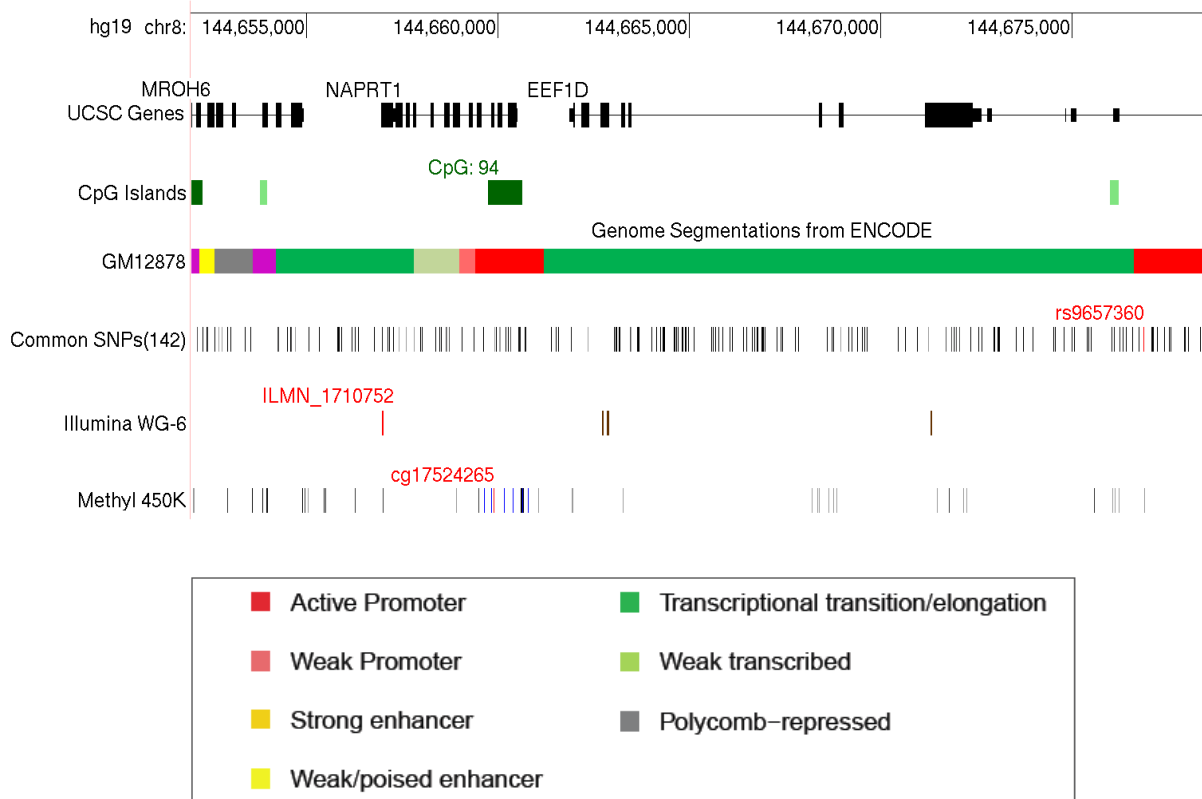
**Figure 3-9: DNA methylation and gene expression of *NAPRT1* plotted by genotypes of rs9657360.**

*DNA methylation (left) and gene expression (middle) of *NAPRT1* is correlated with genotypes of SNP rs9657360. The x axis harbors three genotypes of rs9657360, and the y axis displays residuals representing DNA methylation and gene expression levels. Minor allele C is significantly associated with high methylation level of *NAPRT1* in a cis-manner and simultaneously associated with low gene expression of *NAPRT1*. The scatter plot (right) indicates the linear inverse correlation between methylation and expression and differentiates the three genotype clusters of the CpG methylation and mRNA expression profiles. Blue dots, genotype CC; black dots, genotype CT, and red dots, genotype TT.*

The SWAN corrected and quantile normalized  $\beta$ -values representing methylation levels of cg17524265 show the following values for the genotypes of SNP rs9657360: CC,  $\beta$ -value of 0.71; CT,  $\beta$ -value of 0.54 and TT,  $\beta$ -value of 0.40. Equally, the quantile normalized log<sub>2</sub>-values representing expression levels of ILMN\_1710752 show the following values for the genotypes of SNP rs9657360: CC, log<sub>2</sub>-ratio of 6.95; CT, log<sub>2</sub>-ratio of 7.64 and TT, log<sub>2</sub>-ratio of 8.03.



The *NAPRT1* region including CpG cg17524265 (plus additional six CpG probes), expression probe ILMN\_1710752 and SNP rs9657360 is illustrated in Figure 3-10.



**Figure 3-10: *NAPRT1* region.**

The SNP-methylation-expression three-way association of the *NAPRT1* region is illustrated. CpG cg17524265, mRNA probe ILMN\_1710752 and SNP rs9657360 are highlighted in red. The additional six CpGs which are associated with the mRNA probe are marked in blue (blue lines). It is shown that the CpG probes reside in a CpG island (CpG: 94) encompassing the TSS and the promoter region of *NAPRT1*.

### 3.4 Exploration of the role of QTL-associated rSNPs in common neuropsychiatric disorders

The results of the enrichment analyses demonstrated that meQTLs in hippocampal brain tissue were positional overrepresented in the Brain-Disorder-GWAS gene-set ( $P = 5.79 \times 10^{-4}$ ). Therefore, it was examined, whether rSNPs affecting transcription factor binding sites underlie both the cis-acting QTLs (meQTLs and eQTLs) and the regional GWAS hits, to pinpoint causative rSNPs.

#### 3.4.1 Co-occurrence of cis-acting meQTLs and GWAS hits

The analysis was restricted to GWAS signals with  $P < 5 \times 10^{-8}$ , meQTLs with FDR  $< 0.05$  and LD of the GWAS lead SNP and meQTL SNP with  $r^2 \geq 0.5$ . Overall, 52 GWAS lead SNPs ( $P < 5 \times 10^{-8}$ ) were in strong LD ( $r^2 \geq 0.5$ ) with one of the identified cis-meQTL SNPs. Strong LD of the GWAS lead SNPs and cis-meQTL SNPs is shown for six cis-meQTLs in hippocampal brain tissue in Table 3-16.

**Table 3-16: GWAS lead SNPs in strong LD with cis-acting meQTLs in hippocampal brain tissue.**

CpG	Chr	CpG Pos	meQTL SNP	SNP Pos	FDR	LD SNP GWAS-meQTL	GWAS SNP	GWAS P-value	Disease	Gene Symbol	PMID
cg02586212	1	192544902	rs1359062	192541472	2.21E-17	0.99	rs1323292	2.0E-08	MS	<i>RGS1</i>	21833088
cg22674798	1	3096360	rs207200	3073555	6.17E-07	0.80	rs2651899	4.0E-14	Migraine	<i>PRDM16</i>	23793025
cg18099408	3	52552593	rs6445529	52662722	1.38E-14	0.96	rs2251219	2.0E-09	BD,MDD	<i>PBRM1</i>	20081856
cg18404041	3	52824283	rs2710331	52837855	3.25E-14	0.88	rs2535629	3.0E-12	AUT,ADHD,BD,MDD,SCZ	<i>ITIH3</i>	23453885
cg22298860	6	29690822	rs1362125	29691090	3.26E-21	0.75	rs2523393	1.0E-17	MS	<i>HLA-F</i>	19525953
cg13165778	7	1952518	rs6950151	1953521	3.52E-23	0.74	rs6461049	6.0E-13	SCZ	<i>MAD1L1</i>	23974872

Abbreviations: ADHD, attention-deficit hyperactivity disorder; AUT, autism; BD, bipolar disorder; MDD, major depressive disorder; MS, multiple sclerosis; SCZ, schizophrenia. GWAS associations exceeding  $P$ -values  $< 5 \times 10^{-8}$  from the NHGRI GWAS catalog, LD of GWAS SNPs and meQTL SNPs with  $r^2 \geq 0.5$  and an FDR of 5% for meQTLs.

### 3.4.2 Co-occurrence of cis-acting eQTLs and GWAS hits

My exploration was focused on GWAS hits with  $P < 5 \times 10^{-8}$ , eQTLs with  $FDR < 0.05$  and LD of GWAS SNP and eQTL SNP pairs with  $r^2 \geq 0.5$ . Two SNPs of eQTLs were in strong LD with two GWAS lead SNPs, respectively. These findings included the following traits: schizophrenia (*AS3MT*; GWAS *P-value* =  $4 \times 10^{-14}$ ; Ripke et al. 2013) and multiple sclerosis (*CLECL1*; GWAS *P-value* =  $3 \times 10^{-10}$ ; International Multiple Sclerosis Genetics Consortium 2010).

## 4. DISCUSSION

One of the main goals of human genetics is to elucidate molecular mechanisms underlying disease susceptibility. Up until today, neuropsychiatric disorders are among the most complex and poorly understood conditions. Genome-wide association studies (GWASs) have captured more than two thousand susceptibility loci increasing risk of brain disorders (Kim et al. 2014). However, the majority of these susceptibility loci reside in non-coding regions and their functional consequences remain elusive. Therefore, the question arises whether regulatory sequence variants affecting DNA methylation and gene expression may be the causal susceptibility alleles (Maurano et al. 2012).

The present study aimed to identify regulatory SNPs (rSNPs) that quantitatively influence CpG methylation (meQTL) and mRNA expression (eQTL) and aimed to dissect those rSNPs that confer susceptibility to common brain disorders. To achieve this aim, large-scale data sets of SNPs, CpG methylation states and gene expression profiles of 115 individual samples of fresh frozen hippocampal brain tissue obtained by epilepsy surgery have been integrated using a novel and promising approach. A particular strength of the present study arises from the application of the *Human Methylation450k* (HM450) array and the superior quality of fresh frozen brain tissue from a uniform region of the hippocampus (Grote et al. 2015). These advantages resulted in a six-fold increase of identified meQTLs compared to previous studies with postmortem brain tissue using the *Human Methylation27k* (HM27) array (Gibbs et al. 2010; Zhang et al. 2010; Numata et al. 2012; Gamazon et al. 2013; Numata et al. 2014; Smith et al. 2014). The present study is the first that systematically examined the influence of cis-meQTLs on gene expression in fresh frozen brain tissue to delineate methylation-driven genes. Based on the advances of the 1000 Genomes Project (1000 Genomes Project Consortium 2010), I was also able to improve the quality check (QC) filtering procedures for the CpG methylation (Barrow and Byun 2014) and mRNA expression probes (Westra et al. 2013). In detail, I was successful in removing artificial findings more accurately and refined the quantitative assessment of methylation and expression profiles by implementing a correction for cell-type heterogeneity in the QTL regression analyses (Guintivano et al. 2013). Although recent studies start to consider cellular heterogeneity among different brain regions (De Jager et al. 2014; Pidsley et al. 2014; Xiao et al. 2014), there are yet no comparable studies which focus on neuropsychiatric disorders and perform simultaneously large QTL analyses.

Another major achievement of this study will be the public release of a database reporting the complete meQTL and eQTL findings in context of the publication of this study. These genome-wide maps of meQTLs and eQTLs will improve the prospects to elucidate the genetic mechanisms determining DNA methylation and gene expression, to explore the interplay of DNA methylation and

gene expression, and to extend the allelic spectrum of disease genes of common disorders with complex genetic predisposition. Using the same array platform and analytical procedures, a genome-wide map of cis- and trans-meQTLs in whole blood cells of 496 German population controls without neuropsychiatric disorders was generated as well. With regard to the strong overlap of meQTLs in brain and blood cells, this thesis provides the prerequisites to establish easily accessible epigenetic biomarkers in blood cells for disease-associated cis-meQTLs in brain.

#### 4.1 Autosomal map of cis-meQTLs in human hippocampal brain tissue

This study represents the largest meQTL study in human brain tissue reported so far. To the best of my knowledge, this study is the first study which uses the HM450 array on specimens of fresh frozen human brain tissue. After careful and highly conservative quality filtering procedures, 362,722 autosomal CpGs were included in the meQTL analysis. Previous meQTL studies have been used the smaller HM27 array (27,578 CpGs) and investigations were based on postmortem brain tissue (Gibbs et al. 2010; Zhang et al. 2010; Numata et al. 2014; Smith et al. 2014; for an overview on studies that are frequently mentioned for comparison reasons, please refer to Table 6-2 in Appendix), what might entail unwanted side effects (e.g. alterations in DNA methylation profiles from baseline; cf. 1.5 Quantitative trait loci). Four previous publications of meQTL studies using the HM27 array in brain tissue revealed at the maximum about three thousand (range per study: 887 - 3,612) cis-meQTLs per study (Gibbs et al. 2010; Zhang et al. 2010; Numata et al. 2014; Smith et al. 2014). However, with the present layout I was able to identify 19,954 (out of 362k CpG sites, see above) at which methylation levels are significantly influenced by nearby genetic variations, demonstrating that cis-acting meQTLs are abundant in the autosomal genome. The vast majority (76%) of cis-acting meQTL SNPs displaying the strongest impact on CpG methylation were located within a genomic region of 150 kb around the CpG site. In total, 67% of the identified cis-meQTLs are located intragenic or nearby 7,827 Ensembl genes ( $\pm$  50 kb). Notably, a strong enrichment of cis-meQTLs upstream (TSS201-TSS1500) of the gene promoter region (Table 3-1) was found, highlighting the functional impact of this 5'-regulatory region in the epigenetic regulation of gene expression.

The before mentioned four previous studies have reported 5,061 cis-meQTLs in postmortem brain tissue altogether, of which 3,566 quality-filtered CpG sites were also assessed in my study and 2,353 cis-meQTLs were also identified as significant cis-meQTL. This strong overlap of the cis-meQTLs identified in previous studies of brain tissue support the validity of the overlapping cis-meQTLs observed in the present study.

The enrichment analyses of genes associated with cis-meQTLs indicated an overrepresentation of genes implicated in neurodevelopmental disorders (Krumm et al. 2013) and GWAS hits of brain disorders (Welter et al. 2014). Some of the most significant cis-meQTLs identified within my work are located in genomic regions of several high-ranking candidate genes for neurodevelopmental disorders, such as *ADARB2*, *HDAC4*, *NAPRT1*, *MAD1L1*, *PTPRN2*, and *RIMBP2* (cf. Table 3-5).

The present cis-meQTL and eQTL findings strongly suggest that sequence variants affecting regulatory elements of gene expression or mRNA splicing contribute to the allelic spectrum of disease mutations besides the traditional deleterious exonic and splice-site variants detected by next-generation sequencing (NGS) based whole exome sequencing approaches. At the gene level, these cis-meQTL and -eQTL findings will guide the identification of the causative regulatory sequence variants.

Notably, the SNP showing the strongest correlation with the individual CpG methylation states is not necessarily the regulatory SNP causing the modulating effect. The identified meQTL SNP could be in strong linkage disequilibrium (LD) with the 'real' rSNP which may not be included in the SNP screening set. Consequently, all known SNPs in strong LD with the identified meQTL SNP would have to be examined for their functional alteration of regulatory elements, such as binding sites for transcription factors (TFs), microRNAs, DNaseI hypersensitive sites and regions of histone modifications (Li et al. 2013). Advances of the systematic exploration of regulatory elements of the human genome (Kundaje et al. 2015) have accelerated progress in the dissection of rSNPs. However, deeper insights into the regulatory mechanisms are still required to differentiate the role of potential rSNPs in context of spatial and temporal aspects of gene expression (Januar et al. 2015).

#### 4.1.1 Gene-centric distribution of cis-meQTLs

To explore the gene-centric distribution of the 19k cis-meQTLs identified in hippocampal brain tissue, the localization of the CpGs of the identified cis-meQTLs with the distribution of all QC-filtered 362k CpG sites was compared. A relative positional depletion of cis-meQTLs in the promoter region of genes (TSS200 region) was observed, whereas a relative enrichment was found in intergenic regions and in the 5'-regulatory region of genes upstream of the promoter region (TSS201-TSS1500). The TSS201-1500 region includes well known regulatory elements, such as the insulator CTCF, which is known to exhibit a variable binding affinity of CTCF due to differential CpG methylation at its binding site (Wendt et al. 2008). Specifically, 8.45% (N = 3,055) cis-meQTLs were detected in the TSS201-TSS1500 region compared to 3.6% of all 362k CpG sites ( $P = 7.7 \times 10^{-61}$ ). These findings indicate that methylation-driven gene regulation occurs more likely in regions upstream of the gene promoter, a region which harbors important regulatory elements, such as enhancers and insulators (Jones 2012).

The observed depletion of meQTLs at active promoters is in line with findings of previous studies (Gibbs et al. 2010; Gutierrez-Arcelus et al. 2013; Banovich et al. 2014; De Jager et al. 2014; Kundaje et al. 2015), demonstrating that DNA methylation levels at promoters seem to be less variable and have a lower average methylation level compared with other genomic regions, especially enhancers.

It should also be mentioned that a similar gene-centric distribution of cis-meQTLs in hippocampal brain tissue and blood cells was observed, suggesting that a large proportion of meQTLs share similar methylation control mechanisms in both tissues.

#### 4.1.2 Enrichment analysis of cis-meQTLs

It was evaluated whether the found significant cis-meQTLs were enriched in four gene-sets comprising genes implicated in neurodevelopmental disorders and common traits of the central nervous system. Enrichment analysis showed that cis-meQTLs of the hippocampal brain tissue were significantly enriched in susceptibility genes of GWAS hits of brain disorders ( $P = 5.8 \times 10^{-4}$ ) published in the NHGRI GWAS catalog (Welter et al. 2014) and in genes previously implicated in neurodevelopment disorders ( $P = 0.04$ , Krumm et al. 2013). As diverse functional gene-set categories, e.g. GO categories, are biased according to gene size and the variable coverage of meQTLs (Geeleher et al. 2013; Lockett et al. 2013; Adams et al. 2014; Berko et al. 2014; Delahaye et al. 2014), enrichment results were corrected for the numbers of CpG sites located in different classes of genes.

The enrichment of susceptibility genes implicated by GWAS hits of brain disorders and neurodevelopmental disorders emphasizes, that regulatory SNPs and their effects on CpG methylation and gene expression may confer a substantial contribution to the genetic basis of common neuropsychiatric disorders.

#### 4.1.3 Tissue-specificity of cis-meQTLs in hippocampal brain cells

To explore tissue-specificity of the cis-meQTLs detected in hippocampal brain tissue, the same array platform and analytical procedures were also applied in an additional meQTL analysis of whole blood cells originated from 496 German population controls without neuropsychiatric disorders. The main reasoning behind this analysis is the mere instance that whole blood cells are much more easily accessible as compared to fresh frozen or postmortem brain tissue. In fact, whole blood cells are the most convenient source of tissue for the establishment of epigenetic biomarkers.

Within the analysis more than 108k cis-meQTLs were identified in whole blood cells at a FDR < 0.05. Comparison of cis-meQTLs detected in hippocampal brain tissue with those found in blood cells showed an 65% overlap with a substantial correlation (Spearman's Rank coefficient = 0.42). However,

65% overlap entail that 35% of the CpG sites of cis-meQTLs in brain tissue had no cis-meQTL counterpart in the blood cells. Furthermore, these discordant brain-specific cis-meQTLs includes genes involved in epileptogenesis, such as *GABRB3* (Epi4K Consortium & Epilepsy Phenome/Genome Project 2013), *GRIN2A* (Lemke et al. 2013), *CHRNA2* (Steinlein 2000) and *RBFOX1* (Lal et al. 2013). Hence, this finding demonstrates that brain-specific meQTL analyses are inevitable to identify all rSNPs contributing to common brain disorder. However, results suggest that meQTLs found in whole blood cells are frequently representative for meQTLs in brain tissue and may be suitable as epigenetic biomarkers for brain disorders. The present database of cis-meQTLs in brain and blood cells allows differentiating those cis-meQTLs that are present in both tissues or those that only occur in brain-tissue but not in whole blood cells. Epigenetic biomarkers could lead to improvements in diagnosis and options for early intervention therapies (Januar et al. 2015).

#### 4.2 Autosomal map of trans-meQTLs in human hippocampal brain tissue

Altogether, 2,436 CpG sites at which methylation levels in hippocampal brain tissue are significantly associated with genetic variations in trans (3,927 trans-meQTLs) were identified. For trans-meQTL analysis it was tried to minimize ambiguous trans-signals due to cross-hybridization of CpG probes at additional genomic localizations. Therefore, the HM450 set of CpGs was reduced to 268,397 CpGs, according to more robust and conservative quality-filters applied by Zhang and colleagues (2014) which included a careful genome-wide screen for cross-hybridization of the HM450 CpG probes. Furthermore, the potential effect that SNPs on the regulation of genes in trans resulted from long-range cis-relationships was checked and excluded. These additional filters reduced the number of potentially trans-regulated CpG sites from 3,579 to 2,436 CpGs (and 8,989 to 3,927 for trans-meQTLs). This small sample size used for generating trans-meQTLs largely limited the power to detect trans signals that had previously been implicated in methylation, rendering these analyses even more difficult to interpret. These results demonstrate the complexity of handling trans-acting QTL results and the possibility of a large number of potentially spurious trans-meQTL findings.

#### 4.3 Genome-wide assessment of imprinting meQTLs in GGE parent-offspring trios

The objective of the imprinting meQTL (imeQTL) analysis was the dissection of CpG sites which are differentially methylated depending on their parental origin. Imprinted genes are of particular



interest because they are prone for susceptibility effects due to their Parent-of-Origin (PofO) dependent monoallelic gene expression.

My imeQTL analysis in 269 parent-offspring trios revealed 177 CpGs at 31 genomic loci exhibiting differential methylation in a parent-of-origin dependent manner, of which nine loci were already known (Fang et al. 2012; Wei et al. 2014) and 22 were novel imprinted loci. The detection of nine known imprinting loci can be considered as proof-of-principle of the applied screening approach which compared the CpG methylation states of individuals with parentally inverse heterozygous genotypes of nearby SNPs (Garg et al. 2012). A 1.8-fold increase of paternally imprinted genes compared to maternally imprinted genes was observed. Of the nine known imprinting regions eight imprinted genes are expressed by the paternal allele, and only one is expressed from the maternal allele.

PofO effects are present when the phenotypic effect of an allele depended upon its paternal origin, hence, whether it was inherited from the mother or the father. Strong PofO methylation effects at loci encompassing neurodevelopmental genes, such as the Nucleosome Assembly Protein 1-Like 5 retrogene (*NAP1L5*), Neuronatin (*NNAT*) and Nischarin (*NISCH*) were observed. *NAP1L5* is a chromatin-modifying protein critical for neuronal development, which is most abundantly expressed in neural tissues (Smith et al. 2003). *NNAT* is a dendritically expressed  $Ca^{2+}$  regulator and may play a crucial role in synaptic and possibly cognitive function in the hippocampus (Oyang et al. 2011). *NISCH* is differentially expressed in rat brain, regulates neuronal migration (Ding et al. 2013) and has been implicated to play a role in brain development, dementia, brain cancers and neurodegenerative disorders (Ding et al. 2013; Ostrow et al. 2013).

Those imeQTLs detected on chromosome 3p21.1 are of special interest, considering that GWAS findings of several neuropsychiatric diseases, such as major mood disorder (McMahon et al. 2010), schizophrenia and bipolar disorder (Cross-Disorder Group of the Psychiatric Genomics Consortium et al. 2013; Sleiman et al. 2013), as well as autism spectrum disorder and attention deficit-hyperactivity disorder (Cross-Disorder Group of the Psychiatric Genomics Consortium et al. 2013) have been mapped to this candidate region.

A recently published study investigates the effects of genetic variation and parental origin on gene expression in mice (Crowley et al. 2015). Crowley and co-workers (2015) showed that cis-regulated genes in human blood often have a counterpart in the mouse orthologue. Comparison of my imeQTLs in blood cells with those eQTLs displaying imprinting effects in mice revealed a substantial overlap: 138 out of 177 imeQTL CpGs were located intragenic of 78 Ensembl genes, of which eight (10%) overlapped with imprinted genes detected in mouse brain by Crowley and co-workers (2015).

Additionally, Crowley and co-workers (2015) observed a global allelic imbalance in expression favoring the paternal allele comparable to the allelic imbalance of CpG methylation detected in my imeQTL analysis.

Overall, only few PofO specific allelic associations at imprinted loci have been described for complex traits so far (Lawson et al. 2013). Genes at genomic loci showing imprinting effects should be prioritized as promising candidate genes because of their potentially monoallelic gene expression which may unmask recessive susceptibility alleles.

#### 4.4 Autosomal map of cis-eQTLs in human hippocampal brain tissue

To examine the influence of cis- and trans-acting genetic variations on mRNA expression in human brain tissue, the *Human HT-12 v3* (HT12v3) array was used to measure the expression profiles of 31,146 QC-filtered mRNA target sequences in 115 samples of fresh frozen hippocampal brain tissue. Altogether 734 mRNA probes were identified for which expression levels were significantly influenced by cis-acting SNPs.

Up until now, only few eQTL studies of human brain tissue focus on common brain disorders while employing none candidate gene approaches. The few published studies using brain tissue suffer from small sample sizes and they do not replicate well (Kim et al. 2014). However, notable overlap was found comparing the 734 cis-eQTL results with those reported in previous eQTL studies (Gibbs et al. 2010; Kim et al. 2014; Ramasamy et al. 2014). With respect to an eQTL meta-analysis in postmortem brain cells (Kim et al. 2014), a 65% overlap was found (474 out of 734). An even higher overlap of 71% was found with an eQTL study based on human brain tissue from ten different regions using the *Affymetrix Human Exon 1.0 ST array* (Ramasamy et al. 2014), considering that a cis-eQTL signal was identified in more than one brain region. Altogether, these overlapping eQTL findings underline the validity of the cis-eQTLs identified in this study.

Also, comparisons with previous eQTL meta-analysis in blood cells render interesting and notable overlap (Westra et al. 2013; Schramm et al. 2014). Schramm and co-workers (2014) performed a whole-genome eQTL analysis in 890 blood cell samples using the HT12v3 expression array, and the meta-eQTL analysis in 5,311 blood cells carried out by Westra and colleagues (2013) was restricted to probes present on the HT12v3 array. I found that 50% (367 among 734) of cis-eQTLs in hippocampal brain tissue matched to the cis-eQTLs in blood cells detected by Schramm et al. (2014), and 58% (423 among 734) matched with those reported by Westra et al. (2013). Altogether, some cross-tissue similarities within cis-eQTLs from hippocampal brain tissue and whole blood cells were observed.

These offer the option to delineate accessible transcriptomic biomarkers for at least some eQTLs observed in brain.

#### 4.5 Autosomal map of trans-eQTLs in human hippocampal brain tissue

Altogether, 16 trans-eQTLs with a low significance level were observed of which none has been previously reported (cf. Gibbs et al. 2010; Westra et al. 2013; Ramasamy et al. 2014; Schramm et al. 2014). The available sample size seems to be too small to detect larger effect sizes of trans-eQTLs – this is consistent with general trends in the literature. Westra and Franke (2014) reported that effect sizes of trans-eQTLs are generally small in contrast to cis-eQTL effects.

#### 4.6 Delineation of methylation-driven gene expression in hippocampal brain tissue

Correlation analysis of CpG methylation with mRNA gene expression revealed both inverse and positive correlations. Overall, 73% of CpG-mRNA pairs were inversely correlated and showed an enrichment of methylated CpG sites at active promoters in combination with a reduced gene expression. This is in congruence with previous studies, which demonstrated that DNA methylation at promoter regions reduces gene expression levels (Jones 2012; Gutierrez-Arcelus et al. 2013). Furthermore, results also show that CpG-mRNA pairs with inverse correlation were frequently associated with methylated CpG sites at strong enhancers. Additionally, the observed results suggest that CpG sites affecting gene expression are more likely enriched in non-CpG island (CGI) regions than within CGIs ( $P < 2.2 \times 10^{-16}$ ). CpG sites within CGIs were predominately unmethylated and do not substantially contribute to the variance of gene expression. When analyzing CpG-mRNA pairs for which the target methylation site resided within a CGI, the classical inverse relationship of increased CpG methylation associated with decreased expression levels was observed (in 71% of these cases).

There are only few studies reporting that DNA methylation was positively correlated with gene expression levels (Gibbs et al. 2010; Bell et al. 2011; Gutierrez-Arcelus et al. 2013; Banovich et al. 2014). Previous studies demonstrated that DNA methylation in gene bodies is often associated with activating histone mark and increased expression levels (Hahn et al. 2011; Jones 2012). In contrast, I did not observe a difference in the direction of correlations between CpG sites within or outside gene bodies. Instead, it was found that the CpG and transcription start site (TSS) tend to be more distant from each other in positive correlated CpG-mRNA pairs. This finding was evident with the observed depletion at active promoters and in the TSS200 region for the positive correlated CpG-mRNA pairs.

Exploring the functional effects of SNPs on both DNA methylation and gene expression, only 44 genes were identified fitting this constellation. There may be several explanations for this, including that i) the coverage of the HT12v3 expression array is limited, ii) gene expression levels are controlled by many other factors such as TF binding, chromatin state including histone marks and nucleosome positioning, and by regulation of small RNAs, and iii) the sample size of brain tissue of 115 mTLE patients only provided modest power for eQTL and meQTL mapping. However, compared to the four existing studies which correlated CpG methylation to mRNA gene expression in postmortem brain tissue (Gibbs et al. 2010; Zhang et al. 2010; Gamazon et al. 2013; Numata et al. 2014), the present study reveals more CpG methylation and gene expression correlations whereby the CpG methylation and the mRNA expression was significant genetically determined. This is due to the superior genomic coverage of CpG sites investigated by the HM450 array compared to the HM27 array employed by the reference studies.

#### 4.6.1 Functional implications of meQTL and eQTL effects on *NAPRT1* expression

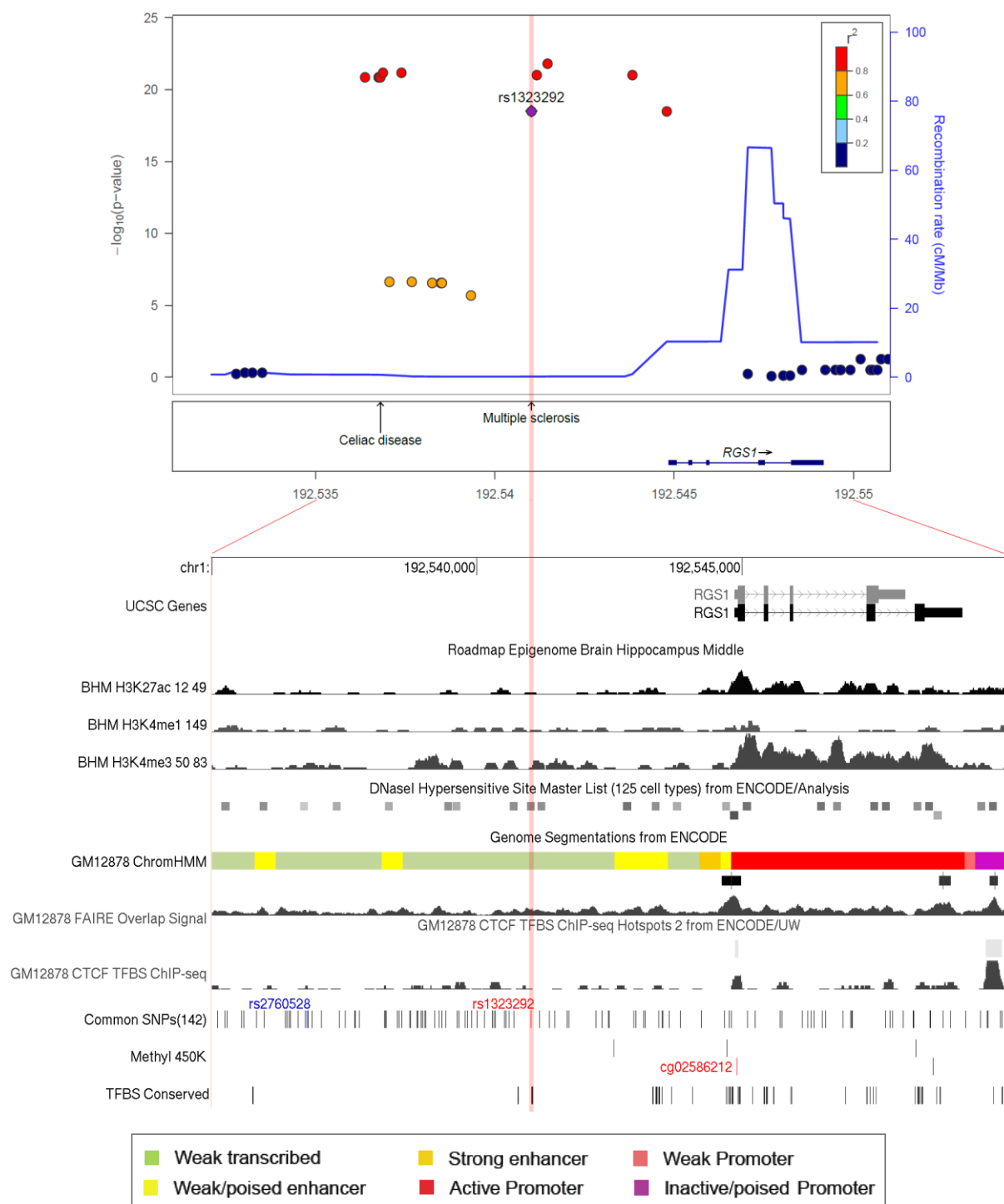
An example of the translational impact of the epigenetic regulation of gene expression is illustrated by using meQTL and eQTL findings on the *NAPRT1* gene, which encodes Nicotinate Phosphoribosyltransferase. I was able to demonstrate that the strong inverse correlations of seven CpGs in the promoter region of *NAPRT1* (correlation coefficient range = -0.83 to -0.72) were genetically determined and controlled via a common rSNP.

In detail, the correlation analysis revealed that five of the seven CpGs reside in a CpG island (chr8:144,659,745-144,660,635; assigned as CpG: 94) and overlap with the TSS of *NAPRT1*. Furthermore, the present meQTL and eQTL results suggest that methylation-driven gene expression of *NAPRT1* is modulated by the genotype of SNP rs9657360 or a nearby rSNP in strong LD with the screening SNP. The minor C-allele (MAF(C)-EUR: 0.20) of the common meQTL/eQTL lead SNP rs9657360 is significantly associated with high methylation state level of the CpG: 94 island in the *NAPRT1* promoter region and low gene expression of *NAPRT1*. Both, the tumor-specific hypermethylation of the promoter CpG: 94 island as well as loss of *NAPRT1* expression has been previously proposed as predictive biomarkers for the therapy of carcinomas using *NAMPT* inhibitors (Shames et al. 2013; Sampath et al. 2015). Considering that a low expression level of *NAPRT1* increases the therapeutic index of *NAMPT* inhibitors in cancer, approximately 4% of the general population carrying a homozygous C/C genotype of SNP rs9657360 should be responsive for *NAMPT* inhibitors. Thus, this genetic risk constellation is of clinical relevance as it enables a diagnostically driven clinical strategy in tumorigenesis including the selection of patients which likely benefit from the administration of *NAMPT* inhibitors (Shames et al. 2013; Sampath et al. 2015).

#### 4.7 Dissection of rSNPs involved in common neuropsychiatric disorders

It was aimed to identify causative rSNPs, which increase the risk of common neuropsychiatric disorders. Previous studies have demonstrated that transcription factor binding sites (TFBSs) influence DNA methylation and gene expression (Stadler et al. 2011; Ziller et al. 2013; Tsankov et al. 2015). To decipher the role of potential rSNPs in neuropsychiatric disorders, it was examined whether SNPs in strong LD ( $r^2 > 0.8$ ) with the GWAS lead SNPs and nearby cis-QTLs cause allelic alterations of TFBSs and change the binding affinity of TFs. Two examples are highlighted below.

The first example involves a GWAS susceptibility locus on chromosome 1q31.2 (*RGS1* gene locus) that has been associated with the risk of multiple sclerosis (MS; Sawcer with the International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2 2011). MS is a chronic inflammatory demyelinating disease of the central nervous system (CNS). All cis-meQTL SNPs in high LD ( $r^2 > 0.8$ ) with the GWAS lead SNP rs1323292 were also significantly associated with meQTL-CpG cg02586212 ( $N = 17$ ,  $P < 2.4 \times 10^{-14}$ ). The most promising LD SNP rs2760528 (CADD 6.8) resides in an evolutionarily conserved sequence (GERP++ score: 3.33) and causes an alteration of the binding affinity of five closely related TRANSFAC binding motifs of the TF STAT (Signal Transducer and Activator of Transcription, Figure 4-1). The strongest allelic alteration of the STAT binding affinities was found for the STAT3 ( $P$ -value =  $4.93 \times 10^{-5}$ ) and the STAT5A ( $P$ -value =  $5.74 \times 10^{-5}$ ) TFBS. STAT family proteins are major regulators of general immune response mediated by the interplay of STAT family proteins with diverse interleukines (IL) and other cytokines regulating T-cell differentiation (Adamson et al. 2009). An abnormal humoral immune response has been described in MS patients (Hafler et al. 2005). The strong LD of SNP rs2760528 with the MS GWAS lead SNP rs1323292 (distance = 4.2 kb; LD = 0.99) implicates a potential functional role of the gene *RGS1*, encoding the regulator of G-protein signaling 1. Previous studies provided evidence for an involvement of *RGS1* in the pathogenesis of MS (International Multiple Sclerosis Genetics Consortium 2010; Tran et al. 2010) and Sawcer and co-workers (2011) identified for the first time *RGS1* as candidate gene for MS via a GWAS. It has been shown that *RGS1* is expressed in immune cells and that overexpression in B-cells attenuates chemokine signaling which plays a major role in attracting immune cells to the site of inflammation (Moratz et al. 2004). An upregulation of *RGS1* measured by gene expression profiling in MS patients highlights the possible importance of *RGS1* for the regulation of chemokine activity in the treatment of MS (Tran et al. 2010). Alterations of *RGS1* expression, potentially through an allelic alteration of the binding affinity of the TFs STAT3/STAT5, could impair the migratory capability of B-cells and possibly alter their migration into the CNS (International Multiple Sclerosis Genetics Consortium 2010). Well fitting, several reports provide evidence for a role of B-cells in the immunopathology of MS (Lucchinetti et al. 2000; Cross et al. 2001; Hafler et al. 2005).



**Figure 4-1: Chromosome 1q31.2 region.**

The GWAS lead SNP *rs1323292* associated with the risk of MS is illustrated and highlighted in red, as well as the cis-meQTL CpG *cg02586212*. The LD SNP *rs2760528* is highlighted in blue. Further, regulatory elements such as DNaseI Hypersensitive sites, genome segmentations and conserved TFBS for the chromosome 1q31.2 region are shown.

The second example involves a GWAS locus on chromosome 3p21.1 (gene complex *NEK4/PRBM1/ITIH3/GNL3*) that is associated with the risk of several neuropsychiatric disorders including bipolar disorder and major depressive disorder (McMahon et al. 2010), autism spectrum disorder and schizophrenia (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013). All cis-meQTL SNPs in high LD ( $r^2 > 0.8$ ) with the GWAS lead SNP rs2251219 (*PRBM1*) for major mood disorder on chromosome 3 were also significantly associated with the meQTL CpG cg18099408 ( $N = 216$ ,  $P < 6.6 \times 10^{-11}$ ). The cis-meQTL SNP rs2251219 was prioritized as top-ranking variant because the synonymous SNP rs2251219 in the *PRBM1* gene affects many regulatory elements including two active enhancer marks (H3K27ac and H3K4me1), resides in an evolutionarily conserved sequence (GERP++ score: 3.54), and encompasses binding sites for four TFs (TRANSFAC binding motifs). The strongest TFBS affinity was obtained for the Tal1 motif ( $P$ -value =  $4.28 \times 10^{-3}$ ) which is involved in the differentiating of midbrain and hindbrain GABAergic neurons (Achim et al. 2012; Achim et al. 2013). The GWAS lead SNP rs2251219 displays a CADD score of 9 indicating that this potential rSNP is ranked among the 10% most deleterious substitutions in the human genome. The GWAS lead SNP rs2251219 is located in *PRBM1*, which encodes polybromo-1 and is important in chromatin remodeling (Thompson 2009). McMahon and co-workers (2010) demonstrated that *PRBM1* is overexpressed in the dorsolateral prefrontal cortex of patients with bipolar disorder. The cis-regulation of gene *PRBM1* could take place potentially through an allelic alteration of the binding affinity of the Tal1 TFBS. Additionally, cis-eQTLs of SNP rs2251219 have been reported with mRNA expression of *ITIH4*, *GLT8D1*, *NT5DC2*, *PRBM1* and *STAB1* (Xia et al. 2012; Witt et al. 2014). Witt and colleagues (2014) emphasized *STAB1* as a new and highly promising candidate gene for bipolar disorder in the chromosome 3p21.1 region based on gene-expression studies in postmortem tissue of the dorsolateral prefrontal cortex which revealed differential expression of *STAB1* in bipolar disorder and schizophrenia patients compared with controls. *STAB1* encodes a multifunctional scavenger receptor, and its expression is induced during chronic inflammation and tumor progression (Kzhyshkowska 2010). *STAB1* is located in close proximity to gene *PRBM1* (distance = 21 kb). The strong LD across the chromosomal *ITIH4/GLT8D1/NT5DC2/PRBM1/STAB1* segment ( $r^2 > 0.9$ ) hinders the efforts to dissect rSNPs of single target genes for neuropsychiatric disorders (McMahon et al. 2010). Complementarily, deleterious non-synonymous SNPs affecting the protein structure facilitate the identification of underlying disease genes. The non-synonymous SNP variant rs4434138 (p.I2282V), located in the *STAB1* gene, is in high LD ( $r^2 > 0.78$ ) with the GWAS lead SNP rs2251219 and display a high CADD score of 11.3, supporting *STAB1* as promising target gene for neuropsychiatric disorders, especially bipolar disorder. Noteworthy, additional non-synonymous LD SNPs associated with meQTLs in the 3p21.1 GWAS candidate region, display high CADD and GERP++

scores and could be considered as potential susceptibility variants (e.g. SNP rs1029871, *NEK4* p.P225A, CADD 29.2, GERP++ 5.9; SNP rs678, *ITIH1* p.E585V, CADD 17.2, GERP++ 5.3).

My explorations highlight several potential rSNPs within the GWAS candidate loci on chromosome 1q31.2 (*RGS1* gene locus) and 3p21.1 (*PRBM1* gene locus) based on their allelic alteration of TFBSs, strong enhancer marks, DNaseI-hypersensitive sites and gene expression. The allelic alteration of TFBSs by potential rSNPs is likely to result in changes of the binding affinities of TFs and thereby gene transcription or splicing processes. Notably, rSNPs extend the allelic spectrum of disease genes and may explain a substantial fraction of the missing heritability.

## 4.8 Outlook

The rapidly growing number of catalogs of QTLs provides deeper insights into the functional impact of regulatory variation and epigenetic processes in neuropsychiatric disorders (Westra & Franke 2014; Albert & Kruglyak 2015; Januar et al. 2015; Kundaje et al. 2015). The Roadmap Epigenetics Project of the US National Institute of Health has generated the most comprehensive human epigenomic landscape across the largest collection of primary cells and tissues (Kundaje et al. 2015), which provides the key to integrate genomic variability with regulatory processes in a spatial and temporal context. However, our understanding of the exact mechanisms by which epigenetic changes modify phenotype is still very limited and further experimental work is necessary.

Large sample sizes and tissue collections will be crucial to ensure that no relevant QTLs will be missed, especially the trans-acting variants, whose effects are likely to be small. To find more trans-QTLs, meta-analyses are required to scale up QTL studies. These meta-analyses will further provide a higher-resolution overview of the downstream effects of common and rare SNPs and will permit causal inference (Westra & Franke 2014). Larger tissue and cell-type specific data sets will specify the cell-type specific QTL effects. To address this need, the Roadmap Epigenomics Consortium has linked epigenomic data to genetic information and produced reference epigenomes for 127 tissue and cell-types (Kundaje et al. 2015; Romanoski et al. 2015). It is important to generate QTL maps under a variety of physiologically important conditions to ensure that the relevant biology is captured (Albert & Kruglyak 2015). The examination of changes in the disease over time and longitudinal epigenetic data at various points in time is crucial. The optimum would involve a longitudinal study of disease-associated tissue of patients collected prior to disease onset and followed up at multiple times throughout the course of the disease. A number of such studies are currently in progress (Pembrey et al. 2014).



The connections along the causal chain from DNA variant to altered expression and trait variation is complex and the identification of large catalogs of causal variants that underlie meQTLs or eQTLs still remains a major challenge (Albert & Kruglyak 2015). To maximize the chance to link GWAS hits with intermediate QTLs, it is necessary to map QTLs in disease-associated tissues (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014; Torres et al. 2014). Currently, all QTL studies focusing on common brain disorder used postmortem brain tissue (cf. Table 6-2) to measure methylation and expression levels what entails several limitations. Fresh frozen hippocampal brain tissue would be the first choice to perform QTL analyses to study neuropsychiatric disorders. Alternatively, if tissues are difficult to obtain or if they represent complex mixtures of cell-types, such as brain tissues, differentiated induced pluripotent stem (iPSC) cells provide a promising option (Sterneckert et al. 2014). In addition, to prioritize putative causal links between DNA variation, expression and phenotypes further development of advanced data integration will be required.

Another important factor arises from the recent and rather rapid advances in technology, which ultimately enabled the acquisition of huge epigenomic data. NGS and microarray technology now offer an option for automated high-throughput data generation and allow mapping of DNA methylation at a high genome-wide resolution and in a large number of samples (Laird 2010). The HM450 array is the most popular platform to profile DNA methylation. However, the platform does not differentiate between methylation and 5-Hydroxymethylcytosine (5hmC). This seems to be a limitation as 5hmC appears to be abundant in brain tissue (Szulwach et al. 2011; Lister et al. 2013) and has recently been discovered to play a critical role in dynamic regulation of genes silenced by methylation (Bhutani et al. 2011; Numata et al. 2012). Furthermore, 5hmC has been implicated in neurodevelopment and could be particularly important in neuropsychiatric disorders (Tseng et al. 2014). Recent studies hypothesize that positive correlation between CpG methylation and gene expression levels is due to 5hmC as it has been shown that 5hmC has activating effects on transcription (Yu et al. 2012; Banovich et al. 2014). Histone acetylation and chromatin marks are other epigenetic modifications, which seem to play an important role in the regulation of gene expression. Histone mark combinations show distinct levels of DNA methylation and predict differences in RNA expression levels that are not reflected in DNA methylation (Kundaje et al. 2015). ChIP-Seq studies permit the targeting of pertinent transcription factors that regulate gene expression within susceptibility loci (Bennett et al. 2015). All these results will inform the identification and refinement of molecular networks that can lead to neuropsychiatric disorders and are influenced by their pathology.

Most published eQTL mapping studies have measured gene expression levels using microarray technology (Westra & Franke 2014) and there are also several eQTL studies available which used the

HT12v3 array (Schurmann et al. 2012; Westra et al. 2013; Schramm et al. 2014). However, the HT13v3 array has been designed to measure whole gene expression and has limitations to differentiate gene isoforms (Pickrell et al. 2010). Recent advances in RNA sequencing (RNAseq) enable the relatively unbiased measurements of expression levels across the entire length of a transcript (Wang et al. 2009). RNAseq comprehensively covers of the entire transcriptome including transcript isoforms, splicing aberrations and RNA-editing modifications (Liang et al. 2013). In addition, RNAseq offers a higher resolution of gene expression quantification than microarrays (Westra & Franke 2014) and allows better mapping of cis-eQTLs within exons (Montgomery et al. 2010; Westra & Franke 2014).

Considering the high complexity of the brain and neuropsychiatric disorders including epilepsy, autism spectrum disorder, bipolar disorder and schizophrenia, integrative functional genomic studies improve the prospects to dissect the complex gene regulation processes in human brain disorders. My study has integrated high-density data sets of SNPs, CpG methylation and gene expression on a genome wide level from the unique resource of fresh frozen human epileptic brain tissue. It was demonstrated that methylation as well as expression levels are affected by genetic variation at a large number of loci across the genome. The presented results provide an initial basis for understanding how genetic variance in humans influences epigenetic marks and expression and forms the basis for further comprehensive analyses.

## 5. REFERENCES

- Abecasis, G.R., Cherny, S.S., Cookson, W. & Cardon, L.R. 2001. GRR: graphical representation of relationship errors. *Bioinformatics* 17: 742-743.
- Achim, K., Peltopuro, P., Lahti, L., Li, J., Salminen, M. & Partanen, J. 2012. Distinct developmental origins and regulatory mechanisms for GABAergic neurons associated with dopaminergic nuclei in the ventral mesodiencephalic region. *Development* 139: 2360-2370.
- Achim, K., Peltopuro, P., Lahti, L., Tsai, H.H., Zachariah, A., Astrand, M., Salminen, M., Rowitch, D. & Partanen, J. 2013. The role of Tal2 and Tal1 in the differentiation of midbrain GABAergic neuron precursors. *Biology Open* 2: 990-997.
- Adams, A.T., Kennedy, N.A., Hansen, R., Ventham, N.T., O'Leary, K.R., Drummond, H.E., Noble, C.L., El-Omar, E., Russell, R.K., Wilson, D.C., Nimmo, E.R., Hold, G.L. & Satsangi, J. 2014. Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory Bowel Diseases* 20: 1784-1793.
- Adamson, A.S., Collins, K., Laurence, A. & O'Shea, J.J. 2009. The Current STATUS of lymphocyte signaling: new roles for old players. *Current Opinion in Immunology* 21: 161-166.
- Albert, F.W. & Kruglyak, L. 2015. The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics* 16: 197-212.
- Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmüller, G. 2014. SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics*: btu779.
- Bae, J.B. 2013. Perspectives of international human epigenome consortium. *Genomics & Informatics* 11: 7-14.
- Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A. & Shendure, J. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* 12: 745-755.
- Banovich, N.E., Lan, X., McVicker, G., van de Geijn, B., Degner, J.F., Blischak, J.D., Roux, J., Pritchard, J.K. & Gilad, Y. 2014. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS Genetics* 10: e1004663.
- Barboux, S., Gascoin-Lachambre, G., Buffat, C., Monnier, P., Mondon, F., Tonanny, M.-B., Pinard, A., Auer, J., Bessières, B., Barlier, A., Jacques, S., Simeoni, U., Dandolo, L., Letourneur, F., Jammes, H. & Vaiman, D. 2012. A genome-wide approach reveals novel imprinted genes expressed in the human placenta. *Epigenetics* 7: 1079-1090.
- Barlow, D.P. 2011. Genomic imprinting: a mammalian epigenetic discovery model. *Annual Reviews of Genetics* 45: 379-403.
- Barrow, T.M. & Byun, H.-M. 2014. Single nucleotide polymorphisms on DNA methylation microarrays: precautions against confounding. *Epigenomics* 6: 577-579.
- Battle, A., Mostafavi, S., Zhu, X., Potash, J.B., Weissman, M.M., McCormick, C., Haudenschild, C.D., Beckman, K.B., Shi, J., Mei, R., Urban, A.E., Montgomery, S.B., Levinson, D.F. & Koller, D. 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Research* 24: 14-24.
- Baylin, S.B. & Jones, P.A. 2011. A decade of exploring the cancer epigenome - biological and translational implications. *Nature Reviews Cancer* 11: 726-734.
- Beck, S. & Rakyán, V.K. 2008. The methylome: approaches for global DNA methylation profiling. *TRENDS in Genetics* 24: 231-237.
- Becker, A.J., Chen, J., Zien, A., Sochivko, D., Normann, S., Schramm, J., Elger, C.E., Wiestler, O.D. & Bluëmcke, I. 2003. Correlated stage-and subfield-associated hippocampal gene expression patterns in experimental and human temporal lobe epilepsy. *European Journal of Neuroscience* 18: 2792-2802.

- Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y. & Pritchard, J.K. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology* 12: R10.
- Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57: 289-300.
- Bennett, D.A., Yu, L., Yang, J., Srivastava, G.P., Aubin, C. & De Jager, P.L. 2015. Epigenomics of Alzheimer's disease. *Translational Research* 165: 200-220.
- Berg, A.T., Berkovic, S.F., Brodie, M.J., Buchhalter, J., Cross, J.H., van Emde Boas, W., Engel, J., French, J., Glauser, T.A., Mathern, G.W., Moshe, S.L., Nordli, D., Plouin, P. & Scheffer, I.E. 2010. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia* 51: 676-685.
- Berko, E.R., Suzuki, M., Beren, F., Lemetre, C., Alaimo, C.M., Calder, R.B., Ballaban-Gil, K., Gounder, B., Kampf, K., Kirschen, J., Maqbool, S.B., Momin, Z., Reynolds, D.M., Russo, N., Shulman, L., Stasiak, E., Tozour, J., Valicenti-McDermott, M., Wang, S., Abrahams, B.S., Hargitai, J., Inbar, D., Zhang, Z., Buxbaum, J.D., Molholm, S., Foxe, J.J., Marion, R.W., Auton, A. & Grealley, J.M. 2014. Mosaic Epigenetic Dysregulation of Ectodermal Cells in Autism Spectrum Disorder. *PLoS Genetics* 10: e1004402.
- Bhutani, N., Burns, D.M. & Blau, H.M. 2011. DNA demethylation dynamics. *Cell* 146: 866-872.
- Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., Fan, J.B. & Shen, R. 2011. High density DNA methylation array with single CpG site resolution. *Genomics* 98: 288-295.
- Bird, A. 2002. DNA methylation patterns and epigenetic memory. *Genes & development* 16: 6-21.
- Blumcke, I., Pauli, E., Clusmann, H., Schramm, J., Becker, A., Elger, C., Merschhemke, M., Meencke, H.J., Lehmann, T., von Deimling, A., Scheiwe, C., Zentner, J., Volk, B., Romstock, J., Stefan, H. & Hildebrandt, M. 2007. A new clinico-pathological classification system for mesial temporal sclerosis. *Acta Neuropathologica* 113: 235-244.
- Bock, C. 2012. Analysing and interpreting DNA methylation data. *Nature Reviews Genetics* 13: 705-719.
- Bock, C., Walter, J., Paulsen, M. & Lengauer, T. 2007. CpG island mapping by epigenome prediction. *PLoS Computational Biology* 3: e110.
- Bray, N.J., Buckland, P.R., Williams, N.M., Williams, H.J., Norton, N., Owen, M.J. & O'Donovan, M.C. 2003. A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. *The American Journal of Human Genetics* 73: 152-161.
- Cardno, A.G., Marshall, E.J., Coid, B., Macdonald, A.M., Ribchester, T.R., Davies, N.J., Venturi, P., Jones, L.A., Lewis, S.W. & Sham, P.C. 1999. Heritability estimates for psychotic disorders: The maudslay twin psychosis series. *Archives of General Psychiatry* 56: 162-168.
- Carroll, W., Lenney, W., Child, F., Strange, R., Jones, P. & Fryer, A. 2005. Maternal glutathione S-transferase GSTP1 genotype is a specific predictor of phenotype in children with asthma. *Pediatric Allergy and Immunology* 16: 32-39.
- Chao, M.J., Herrera, B.M., Ramagopalan, S.V., Deluca, G., Handunnetthi, L., Orton, S.M., Lincoln, M.R., Sadovnick, A.D. & Ebers, G.C. 2010. Parent-of-origin effects at the major histocompatibility complex in multiple sclerosis. *Human Molecular Genetics* 19: 3679-3689.
- Chen, Y.A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J. & Weksberg, R. 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 8: 203-209.
- Civelek, M. & Lusis, A.J. 2014. Systems genetics approaches to understand complex traits. *Nature Reviews Genetics* 15: 34-48.
- Clayton-Smith, J. 1993. Clinical research on Angelman syndrome in the United Kingdom: observations on 82 affected individuals. *American Journal of Medical Genetics* 46: 12-15.
- Collins, P.Y., Patel, V., Joestl, S.S., March, D., Insel, T.R., Daar, A.S., Scientific Advisory, B., the Executive Committee of the Grand Challenges on Global Mental, H., Anderson, W., Dhansay,

- M.A., Phillips, A., Shurin, S., Walport, M., Ewart, W., Savill, S.J., Bordin, I.A., Costello, E.J., Durkin, M., Fairburn, C., Glass, R.I., Hall, W., Huang, Y., Hyman, S.E., Jamison, K., Kaaya, S., Kapur, S., Kleinman, A., Ogunniyi, A., Otero-Ojeda, A., Poo, M.M., Ravindranath, V., Sahakian, B.J., Saxena, S., Singer, P.A. & Stein, D.J. 2011. Grand challenges in global mental health. *Nature* 475: 27-30.
- Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O.A., Anjorin, A., Anney, R., Anttila, V., Arking, D.E., Asherson, P., Azevedo, M.H., Backlund, L., Badner, J.A., Bailey, A.J., Banaschewski, T., Barchas, J.D., Barnes, M.R., Barrett, T.B., Bass, N., Battaglia, A., Bauer, M., Bayes, M., Bellivier, F., Bergen, S.E., Berrettini, W., Betancur, C., Bettecken, T., Biederman, J., Binder, E.B., Black, D.W., Blackwood, D.H., Bloss, C.S., Boehnke, M., Boomsma, D.I., Breen, G., Breuer, R., Bruggeman, R., Cormican, P., Buccola, N.G., Buitelaar, J.K., Bunney, W.E., Buxbaum, J.D., Byerley, W.F., Byrne, E.M., Caesar, S., Cahn, W., Cantor, R.M., Casas, M., Chakravarti, A., Chambert, K., Choudhury, K., Cichon, S., Cloninger, C.R., Collier, D.A., Cook, E.H., Coon, H., Cormand, B., Corvin, A., Coryell, W.H., Craig, D.W., Craig, I.W., Crosbie, J., Cuccaro, M.L., Curtis, D., Czamara, D., Datta, S., Dawson, G., Day, R., De Geus, E.J., Degenhardt, F., Djurovic, S., Donohoe, G.J., Doyle, A.E., Duan, J., Dudbridge, F., Duketis, E., Ebstein, R.P., Edenberg, H.J., Elia, J., Ennis, S., Etain, B., Fanous, A., Farmer, A.E., Ferrier, I.N., Flickinger, M., Fombonne, E., Foroud, T., Frank, J., Franke, B., Fraser, C., Freedman, R., Freimer, N.B., Freitag, C.M., Friedl, M., Frisen, L., Gallagher, L., Gejman, P.V., Georgieva, L., Gershon, E.S., Geschwind, D.H., Giegling, I., Gill, M., Gordon, S.D., Gordon-Smith, K., Green, E.K., Greenwood, T.A., Grice, D.E., Gross, M., Grozeva, D., Guan, W., Gurling, H., De Haan, L., Haines, J.L., Hakonarson, H., Hallmayer, J., Hamilton, S.P., Hamshere, M.L., Hansen, T.F., Hartmann, A.M., Hautzinger, M., Heath, A.C., Henders, A.K., Herms, S., Hickie, I.B., Hipolito, M., Hoefels, S., Holmans, P.A., Holsboer, F., Hoogendijk, W.J., Hottenga, J.J., Hultman, C.M., Hus, V., Ingason, A., Ising, M., Jamain, S., Jones, E.G., Jones, I., Jones, L., Tzeng, J.Y., Kahler, A.K., Kahn, R.S., Kandaswamy, R., Keller, M.C., Kennedy, J.L., Kenny, E., Kent, L., Kim, Y., Kirov, G.K., Klauk, S.M., Klei, L., Knowles, J.A., Kohli, M.A., Koller, D.L., Konte, B., Korszun, A., Krabbendam, L., Krasucki, R., Kuntsi, J., Kwan, P., Landen, M., Langstrom, N., Lathrop, M., Lawrence, J., Lawson, W.B., Leboyer, M., Ledbetter, D.H., Lee, P.H., Lencz, T., Lesch, K.P., Levinson, D.F., Lewis, C.M., Li, J., Lichtenstein, P., Lieberman, J.A., Lin, D.Y., Linszen, D.H., Liu, C., Lohoff, F.W., Loo, S.K., Lord, C., Lowe, J.K., Lucae, S., MacIntyre, D.J., Madden, P.A., Maestrini, E., Magnusson, P.K., Mahon, P.B., Maier, W., Malhotra, A.K., Mane, S.M., Martin, C.L., Martin, N.G., Mattheisen, M., Matthews, K., Mattingsdal, M., McCarroll, S.A., McGhee, K.A., McGough, J.J., McGrath, P.J., McGuffin, P., McInnis, M.G., McIntosh, A., McKinney, R., McLean, A.W., McMahon, F.J., McMahon, W.M., McQuillin, A., Medeiros, H., Medland, S.E., Meier, S., Melle, I., Meng, F., Meyer, J., Middeldorp, C.M., Middleton, L., Milanova, V., Miranda, A., Monaco, A.P., Montgomery, G.W., Moran, J.L., Moreno-De-Luca, D., Morken, G., Morris, D.W., Morrow, E.M., Moskvina, V., Muglia, P., Muhleisen, T.W., Muir, W.J., Muller-Myhsok, B., Murtha, M., Myers, R.M., Myin-Germeys, I., Neale, M.C., Nelson, S.F., Nievergelt, C.M., Nikolov, I., Nimgaonkar, V., Nolen, W.A., Nothen, M.M., Nurnberger, J.I., Nwulia, E.A., Nyholt, D.R., O'Dushlaine, C., Oades, R.D., Olincy, A., Oliveira, G., Olsen, L., Ophoff, R.A., Osby, U., Owen, M.J., Palotie, A., Parr, J.R., Paterson, A.D., Pato, C.N., Pato, M.T., Penninx, B.W., Pergadia, M.L., Pericak-Vance, M.A., Pickard, B.S., Pimm, J., Piven, J., Posthuma, D., Potash, J.B., Poustka, F., Propping, P., Puri, V., Quedsted, D.J., Quinn, E.M., Ramos-Quiroga, J.A., Rasmussen, H.B., Raychaudhuri, S., Rehnstrom, K., Reif, A., Ribases, M., Rice, J.P., Rietschel, M., Roeder, K., Roeyers, H., Rossin, L., Rothenberger, A., Rouleau, G., Ruderfer, D., Rujescu, D., Sanders, A.R., Sanders, S.J., Santangelo, S.L., Sergeant, J.A., Schachar, R., Schalling, M., Schatzberg, A.F., Scheftner, W.A., Schellenberg, G.D., Scherer, S.W., Schork, N.J., Schulze, T.G., Schumacher, J., Schwarz, M., Scolnick, E., Scott, L.J., Shi, J., Shilling, P.D., Shyn, S.I., Silverman, J.M., Slager, S.L., Smalley, S.L., Smit, J.H., Smith, E.N., Sonuga-Barke, E.J.,

- St Clair, D., State, M., Steffens, M., Steinhausen, H.C., Strauss, J.S., Strohmaier, J., Stroup, T.S., Sutcliffe, J.S., Szatmari, P., Szelinger, S., Thirumalai, S., Thompson, R.C., Todorov, A.A., Tozzi, F., Treutlein, J., Uhr, M., van den Oord, E.J., Van Grootheest, G., Van Os, J., Vicente, A.M., Vieland, V.J., Vincent, J.B., Visscher, P.M., Walsh, C.A., Wassink, T.H., Watson, S.J., Weissman, M.M., Werge, T., Wienker, T.F., Wijsman, E.M., Willemsen, G., Williams, N., Willsey, A.J., Witt, S.H., Xu, W., Young, A.H., Yu, T.W., Zammit, S., Zandi, P.P., Zhang, P., Zitman, F.G., Zollner, S., International Inflammatory Bowel Disease Genetics, C., Devlin, B., Kelsoe, J.R., Sklar, P., Daly, M.J., O'Donovan, M.C., Craddock, N., Sullivan, P.F., Smoller, J.W., Kendler, K.S. & Wray, N.R. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* 45: 984-994.
- Cross, A.H., Trotter, J.L. & Lyons, J.-A. 2001. B cells and antibodies in CNS demyelinating disease. *Journal of Neuroimmunology* 112: 1-14.
- Crowley, J.J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I.K., Kim, Y., Wang, J.R., Morgan, A.P., Calaway, J.D., Aylor, D.L., Yun, Z., Bell, T.A., Buus, R.J., Calaway, M.E., Didion, J.P., Gooch, T.J., Hansen, S.D., Robinson, N.N., Shaw, G.D., Spence, J.S., Quackenbush, C.R., Barrick, C.J., Nonneman, R.J., Kim, K., Xenakis, J., Xie, Y., Valdar, W., Lenarcic, A.B., Wang, W., Welsh, C.E., Fu, C.P., Zhang, Z., Holt, J., Guo, Z., Threadgill, D.W., Tarantino, L.M., Miller, D.R., Zou, F., McMillan, L., Sullivan, P.F. & Pardo-Manuel de Villena, F. 2015. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nature Genetics* 47: 353-360.
- Dalton, V.S., Kolshus, E. & McLoughlin, D.M. 2014. Epigenetics and depression: return of the repressed. *Journal of Affective Disorders* 155: 1-12.
- Davies, M.N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., Harris, R.A., Milosavljevic, A., Troakes, C., Al-Sarraj, S., Dobson, R., Schalkwyk, L.C. & Mill, J. 2012. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biology* 13: R43.
- Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A. & Batzoglou, S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology* 6: e1001025.
- De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C., Tang, A., Raj, T., Replogle, J., Brodeur, W., Gabriel, S., Chai, H.S., Younkin, C., Younkin, S.G., Zou, F., Szyf, M., Epstein, C.B., Schneider, J.A., Bernstein, B.E., Meissner, A., Ertekin-Taner, N., Chibnik, L.B., Kellis, M., Mill, J. & Bennett, D.A. 2014. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nature Neuroscience* 17: 1156-1163.
- Delahaye, F., Wijetunga, N.A., Heo, H.J., Tozour, J.N., Zhao, Y.M., Grealley, J.M. & Einstein, F.H. 2014. Sexual dimorphism in epigenomic responses of stem cells to extreme fetal growth. *Nature Communications* 5: 5187.
- Ding, Y., Zhang, R., Zhang, K., Lv, X., Chen, Y., Li, A., Wang, L., Zhang, X. & Xia, Q. 2013. Nischarin is differentially expressed in rat brain and regulates neuronal migration. *PLoS One* 8: e54563.
- Du, P., Zhang, X., Huang, C.C., Jafari, N., Kibbe, W.A., Hou, L. & Lin, S.M. 2010. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11: 587.
- Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., McCune, R.A. & Gehrke, C. 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research* 10: 2709-2721.
- Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H. & Nadeau, J.H. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics* 11: 446-450.
- ENCODE Project Consortium 2004. The ENCODE (ENCyclopedia of DNA elements) project. *Science* 306: 636-640.

- ENCODE Project Consortium 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799-816.
- ENCODE Project Consortium 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.
- Epi4K Consortium & Epilepsy Phenome/Genome Project, Epilepsy Phenome/Genome Project, Allen, A.S., Berkovic, S.F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E.E., Epstein, M.P., Glauser, T., Goldstein, D.B., Han, Y., Heinzen, E.L., Hitomi, Y., Howell, K.B., Johnson, M.R., Kuzniecky, R., Lowenstein, D.H., Lu, Y.F., Madou, M.R., Marson, A.G., Mefford, H.C., Esmaeli Nieh, S., O'Brien, T.J., Ottman, R., Petrovski, S., Poduri, A., Ruzzo, E.K., Scheffer, I.E., Sherr, E.H., Yuskaitis, C.J., Abou-Khalil, B., Alldredge, B.K., Bautista, J.F., Berkovic, S.F., Boro, A., Cascino, G.D., Consalvo, D., Crumrine, P., Devinsky, O., Dlugos, D., Epstein, M.P., Fiol, M., Fountain, N.B., French, J., Friedman, D., Geller, E.B., Glauser, T., Glynn, S., Haut, S.R., Hayward, J., Helmers, S.L., Joshi, S., Kanner, A., Kirsch, H.E., Knowlton, R.C., Kossoff, E.H., Kuperman, R., Kuzniecky, R., Lowenstein, D.H., McGuire, S.M., Motika, P.V., Novotny, E.J., Ottman, R., Paolicchi, J.M., Parent, J.M., Park, K., Poduri, A., Scheffer, I.E., Shellhaas, R.A., Sherr, E.H., Shih, J.J., Singh, R., Sirven, J., Smith, M.C., Sullivan, J., Lin Thio, L., Venkat, A., Vining, E.P., Von Allmen, G.K., Weisenberg, J.L., Widdess-Walsh, P. & Winawer, M.R. 2013. De novo mutations in epileptic encephalopathies. *Nature* 501: 217-221.
- Ernst, J. & Kellis, M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnology* 28: 817-825.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M. & Bernstein, B.E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473: 43-49.
- Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O. & Knight, J.C. 2012. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genetics* 44: 502-510.
- Fang, F., Hodges, E., Molaro, A., Dean, M., Hannon, G.J. & Smith, A.D. 2012. Genomic landscape of human allele-specific DNA methylation. *Proceedings of the National Academy of Sciences* 109: 7332-7337.
- Fassunke, J., Majores, M., Tresch, A., Niehusmann, P., Grote, A., Schoch, S. & Becker, A.J. 2008. Array analysis of epilepsy-associated gangliogliomas reveals expression patterns related to aberrant development of neuronal precursors. *Brain* 131: 3034-3050.
- Feinberg, A.P. 2007. Phenotypic plasticity and the epigenetics of human disease. *Nature* 447: 433-440.
- Gamazon, E.R., Badner, J.A., Cheng, L., Zhang, C., Zhang, D., Cox, N.J., Gershon, E.S., Kelsoe, J.R., Greenwood, T.A., Nievergelt, C.M., Chen, C., McKinney, R., Shilling, P.D., Schork, N.J., Smith, E.N., Bloss, C.S., Nurnberger, J.I., Edenberg, H.J., Foroud, T., Koller, D.L., Scheftner, W.A., Coryell, W., Rice, J., Lawson, W.B., Nwulia, E.A., Hipolito, M., Byerley, W., McMahon, F.J., Schulze, T.G., Berrettini, W.H., Potash, J.B., Zandi, P.P., Mahon, P.B., McInnis, M.G., Zollner, S., Zhang, P., Craig, D.W., Szelinger, S., Barrett, T.B. & Liu, C. 2013. Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Molecular Psychiatry* 18: 340-346.
- Gardiner-Garden, M. & Frommer, M. 1987. CpG islands in vertebrate genomes. *Journal of Molecular Biology* 196: 261-282.
- Garg, P., Borel, C. & Sharp, A.J. 2012. Detection of parent-of-origin specific expression quantitative trait loci by cis-association analysis of gene expression in trios. *PLoS One* 7: e41695.
- Geeleher, P., Hartnett, L., Egan, L.J., Golden, A., Ali, R.A.R. & Seoighe, C. 2013. Gene-set analysis is severely biased when applied to genome-wide methylation data. *Bioinformatics* 29: 1851-1857.
- Gelernter, J. 2015. Genetics of complex traits in psychiatry. *Biology Psychiatry* 77: 36-42.

- Genomes Project Consortium 2010. A map of human genome variation from population-scale sequencing. *Nature* 467: 1061-1073.
- Georgiades, P., Watkins, M., Burton, G.J. & Ferguson-Smith, A.C. 2001. Roles for genomic imprinting and the zygotic genome in placental development. *Proceedings of the National Academy of Sciences* 98: 4522-4527.
- Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J., Johnson, R., Zielke, H.R., Ferrucci, L., Longo, D.L., Cookson, M.R. & Singleton, A.B. 2010. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genetics* 6: e1000952.
- Gjoneska, E., Pfenning, A.R., Mathys, H., Quon, G., Kundaje, A., Tsai, L.H. & Kellis, M. 2015. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* 518: 365-369.
- Grote, A., Schoch, S. & Becker, A.J. 2015. Temporal lobe epilepsy: a unique window into living human brain epigenetic gene regulation. *Brain* 138: 509-511.
- Guintivano, J., Aryee, M.J. & Kaminsky, Z.A. 2013. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* 8: 290-302.
- Gutierrez-Arcelus, M., Lappalainen, T., Montgomery, S.B., Buil, A., Ongen, H., Yurovsky, A., Bryois, J., Giger, T., Romano, L., Planchon, A., Falconnet, E., Bielser, D., Gagnebin, M., Padioleau, I., Borel, C., Letourneau, A., Makrythanasis, P., Guipponi, M., Gehrig, C., Antonarakis, S.E. & Dermitzakis, E.T. 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* 2: e00523.
- Hafler, D.A., Slavik, J.M., Anderson, D.E., O'Connor, K.C., De Jager, P. & Baecher-Allan, C. 2005. Multiple sclerosis. *Immunological Reviews* 204: 208-231.
- Hahn, M.A., Wu, X., Li, A.X., Hahn, T. & Pfeifer, G.P. 2011. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLoS One* 6: e18844.
- Harper, K.N., Peters, B.A. & Gamble, M.V. 2013. Batch effects and pathway analysis: two potential perils in cancer studies involving DNA methylation array analysis. *Cancer Epidemiology Biomarkers & Prevention* 22: 1052-1060.
- Henikoff, S. & Matzke, M.A. 1997. Exploring and explaining epigenetic effects. *TRENDS in Genetics* 13.
- Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K. & Kelsey, K.T. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13: 86.
- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics* 44: 955-959.
- ILAE 1989. Proposal for Revised Classification of Epilepsies and Epileptic Syndromes. *Epilepsia* 30: 389-399.
- International Multiple Sclerosis Genetics Consortium 2010. IL12A, MPHOSPH9/CDK2AP1 and RGS1 are novel multiple sclerosis susceptibility loci. *Genes & Immunity* 11: 397-405.
- International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium, Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E., Edkins, S., Gray, E., Booth, D.R., Potter, S.C., Goris, A., Band, G., Oturai, A.B., Strange, A., Saarela, J., Bellenguez, C., Fontaine, B., Gillman, M., Hemmer, B., Gwilliam, R., Zipp, F., Jayakumar, A., Martin, R., Leslie, S., Hawkins, S., Giannoulatou, E., D'Alfonso, S., Blackburn, H., Martinelli Boneschi, F., Liddle, J., Harbo, H.F., Perez, M.L., Spurkland, A., Waller, M.J., Mycko, M.P., Ricketts, M., Comabella, M., Hammond, N., Kockum, I., McCann, O.T., Ban, M., Whittaker, P., Kempainen, A., Weston, P., Hawkins, C., Widaa, S., Zajicek, J., Dronov, S., Robertson, N., Bumpstead, S.J., Barcellos, L.F., Ravindrarajah, R., Abraham, R., Alfredsson, L., Ardlie, K., Aubin, C., Baker, A., Baker, K., Baranzini, S.E., Bergamaschi, L., Bergamaschi, R., Bernstein, A., Berthele, A., Boggild, M., Bradfield, J.P., Brassat, D., Broadley, S.A., Buck, D., Butzkueven, H., Capra, R., Carroll, W.M.,



- Cavalla, P., Celius, E.G., Cepok, S., Chiavacci, R., Clerget-Darpoux, F., Clysters, K., Comi, G., Cossburn, M., Cournu-Rebeix, I., Cox, M.B., Cozen, W., Cree, B.A., Cross, A.H., Cusi, D., Daly, M.J., Davis, E., de Bakker, P.I., Debouverie, M., D'Hooghe M, B., Dixon, K., Dobosi, R., Dubois, B., Ellinghaus, D., Elovaara, I., Esposito, F., Fontenille, C., Foote, S., Franke, A., Galimberti, D., Ghezzi, A., Glessner, J., Gomez, R., Gout, O., Graham, C., Grant, S.F., Guerini, F.R., Hakonarson, H., Hall, P., Hamsten, A., Hartung, H.P., Heard, R.N., Heath, S., Hobart, J., Hoshi, M., Infante-Duarte, C., Ingram, G., Ingram, W., Islam, T., Jagodic, M., Kabesch, M., Kermodé, A.G., Kilpatrick, T.J., Kim, C., Klopp, N., Koivisto, K., Larsson, M., Lathrop, M., Lechner-Scott, J.S., Leone, M.A., Leppä, V., Liljedahl, U., Bomfim, I.L., Lincoln, R.R., Link, J., Liu, J., Lorentzen, A.R., Lupoli, S., Macciardi, F., Mack, T., Marriotti, M., Martinelli, V., Mason, D., McCauley, J.L., Mentch, F., Mero, I.L., Mihalova, T., Montalban, X., Mottershead, J., Myhr, K.M., Naldi, P., Ollier, W., Page, A., Palotie, A., Pelletier, J., Piccio, L., Pickersgill, T., Piehl, F., Pobywajlo, S., Quach, H.L., Ramsay, P.P., Reunanen, M., Reynolds, R., Rioux, J.D., Rodegher, M., Roesner, S., Rubio, J.P., Ruckert, I.M., Salvetti, M., Salvi, E., Santaniello, A., Schaefer, C.A., Schreiber, S., Schulze, C., Scott, R.J., Sellebjerg, F., Selmaj, K.W., Sexton, D., Shen, L., Simms-Acuna, B., Skidmore, S., Sleiman, P.M., Smestad, C., Sorensen, P.S., Sondergaard, H.B., Stankovich, J., Strange, R.C., Sulonen, A.M., Sundqvist, E., Syvanen, A.C., Taddeo, F., Taylor, B., Blackwell, J.M., Tienari, P., Bramon, E., Tourbah, A., Brown, M.A., Tronczynska, E., Casas, J.P., Tubridy, N., Corvin, A., Vickery, J., Jankowski, J., Villoslada, P., Markus, H.S., Wang, K., Mathew, C.G., Wason, J., Palmer, C.N., Wichmann, H.E., Plomin, R., Willoughby, E., Rautanen, A., Winkelmann, J., Wittig, M., Trembath, R.C., Yaouanq, J., Viswanathan, A.C., Zhang, H., Wood, N.W., Zuvich, R., Deloukas, P., Langford, C., Duncanson, A., Oksenberg, J.R., Pericak-Vance, M.A., Haines, J.L., Olsson, T., Hillert, J., Ivinson, A.J., De Jager, P.L., Peltonen, L., Stewart, G.J., Hafler, D.A., Hauser, S.L., McVean, G., Donnelly, P. & Compston, A. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476: 214-219.
- Jaffe, A.E. & Irizarry, R.A. 2014. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biology* 15: R31.
- Jansen, R.C. & Nap, J.-P. 2001. Genetical genomics: the added value from segregation. *TRENDS in Genetics* 17: 388-391.
- Januar, V., Saffery, R. & Ryan, J. 2015. Epigenetics and depressive disorders: a review of current progress and future directions. *International Journal of Epidemiology*.
- Jones, P.A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* 13: 484-492.
- Kendler, K.S., Kuhn, J.W., Vittum, J., Prescott, C.A. & Riley, B. 2005. The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: a replication. *Archives of General Psychiatry* 62: 529-535.
- Kim, Y., Xia, K., Tao, R., Giusti-Rodriguez, P., Vladimirov, V., van den Oord, E. & Sullivan, P.F. 2014. A meta-analysis of gene expression quantitative trait loci in brain. *Translational Psychiatry* 4: e459.
- Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. & Shendure, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* 46: 310-315.
- Kitsios, G.D. & Zintzaras, E. 2009. Genome-wide association studies: hypothesis-"free" or "engaged"? *Translational Research* 154: 161-164.
- Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Jonasdottir, A., Frigge, M.L., Gylfason, A., Olason, P.I., Gudjonsson, S.A., Sverrisson, S., Stacey, S.N., Sigurgeirsson, B., Benediktsson, K.R., Sigurdsson, H., Jonsson, T., Benediktsson, R., Olafsson, J.H., Johannsson, O.T., Hreidarsson, A.B., Sigurdsson, G., Consortium, D., Ferguson-Smith, A.C., Gudbjartsson, D.F., Thorsteinsdottir, U. & Stefansson, K. 2009. Parental origin of sequence variants associated with complex diseases. *Nature* 462: 868-874.

- Kornberg, J.R., Brown, J.L., Sadovnick, A.D., Remick, R.A., Keck Jr, P.E., McElroy, S.L., Rapaport, M.H., Thompson, P.M., Kaul, J.B., Vrabel, C.M., Schommer, S.C., Wilson, T., Pizzuco, D., Jameson, S., Schibuk, L. & Kelsoe, J.R. 2000. Evaluating the parent-of-origin effect in bipolar affective disorder: Is a more penetrant subtype transmitted paternally? *Journal of Affective Disorders* 59: 183-192.
- Kral, T., Clusmann, H., Urbach, J., Schramm, J., Elger, C., Kurthen, M. & Grunwald, T. 2002. Preoperative evaluation for epilepsy surgery (Bonn Algorithm). *Zentralblatt für Neurochirurgie* 63: 106-110.
- Krumm, N., O'Roak, B.J., Karakoc, E., Mohajeri, K., Nelson, B., Vives, L., Jacquemont, S., Munson, J., Bernier, R. & Eichler, E.E. 2013. Transmission disequilibrium of small CNVs in simplex autism. *American Journal of Human Genetics* 93: 595-606.
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfening, A.R., Wang, X., Clausnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shores, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthal, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M. & Roadmap Epigenomics, C. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317-330.
- Kzhyshkowska, J. 2010. Multifunctional Receptor Stabilin-1 in Homeostasis and Disease. *TheScientificWorldJOURNAL* 10: 2039-2053.
- Labrie, V., Pai, S. & Petronis, A. 2012. Epigenetics of major psychosis: progress, problems and perspectives. *TRENDS in Genetics* 28: 427-435.
- Laird, P.W. 2010. Principles and challenges of genomewide DNA methylation analysis. *Nature Reviews Genetics* 11: 191-203.
- Lal, D., Trucks, H., Moller, R.S., Hjalgrim, H., Koeleman, B.P., de Kovel, C.G., Visscher, F., Weber, Y.G., Lerche, H., Becker, F., Schankin, C.J., Neubauer, B.A., Surges, R., Kunz, W.S., Zimprich, F., Franke, A., Illig, T., Ried, J.S., Leu, C., Nurnberg, P., Sander, T., Consortium, E.M. & Consortium, E. 2013. Rare exonic deletions of the RBF1 gene increase risk of idiopathic generalized epilepsy. *Epilepsia* 54: 265-271.
- Lambertini, L., Marsit, C.J., Sharma, P., Maccani, M., Ma, Y., Hu, J. & Chen, J. 2012. Imprinted gene expression in fetal growth and development. *Placenta* 33: 480-486.
- Lander, E. & Schork, N. 1994. Genetic dissection of complex traits. *Science* 265: 2037-2048.
- Lawson, H.A., Cheverud, J.M. & Wolf, J.B. 2013. Genomic imprinting and parent-of-origin effects on complex traits. *Nature Reviews Genetics* 14: 609-617.
- Lemke, J.R., Lal, D., Reinthaler, E.M., Steiner, I., Nothnagel, M., Alber, M., Geider, K., Laube, B., Schwake, M., Finsterwalder, K., Franke, A., Schilhabel, M., Jahn, J.A., Muhle, H., Boor, R., Van Paesschen, W., Caraballo, R., Fejerman, N., Weckhuysen, S., De Jonghe, P., Larsen, J., Moller, R.S., Hjalgrim, H., Addis, L., Tang, S., Hughes, E., Pal, D.K., Veri, K., Vaher, U., Talvik, T., Dimova, P., Guerrero Lopez, R., Serratosa, J.M., Linnankivi, T., Lehesjoki, A.E., Ruf, S., Wolff, M., Buerki, S., Wohlrab, G., Kroell, J., Datta, A.N., Fiedler, B., Kurlemann, G., Kluger, G., Hahn, A., Haberlandt, D.E., Kutzer, C., Sperner, J., Becker, F., Weber, Y.G., Feucht, M., Steinbock, H., Neophythou, B., Ronen, G.M., Gruber-Sedlmayr, U., Geldner, J., Harvey, R.J., Hoffmann, P., Herms, S., Altmuller, J., Toliat, M.R., Thiele, H., Nurnberg, P., Wilhelm, C., Stephani, U., Helbig, I., Lerche, H., Zimprich, F., Neubauer, B.A., Biskup, S. & von Spiczak, S. 2013.

- Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. *Nature Genetics* 45: 1067-1072.
- Li, E. 2002. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics* 3: 662-673.
- Li, M.J., Wang, L.Y., Xia, Z., Sham, P.C. & Wang, J. 2013. GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic Acids Research* 41: W150-158.
- Liang, L., Morar, N., Dixon, A.L., Lathrop, G.M., Abecasis, G.R., Moffatt, M.F. & Cookson, W.O. 2013. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Research* 23: 716-726.
- Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J.C., Rao, A., Esteller, M., He, C., Haghghi, F.G., Sejnowski, T.J., Behrens, M.M. & Ecker, J.R. 2013. Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* 341.
- Liu, D.J. & Leal, S.M. 2012. Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. *American Journal of Human Genetics* 91: 585-596.
- Lockett, G.A., Patil, V.K., Soto-Ramírez, N., Ziyab, A.H., Holloway, J.W. & Karmaus, W. 2013. Epigenomics and allergic disease. *Epigenomics* 5: 685-699.
- Lucchinetti, C., Bruck, W., Parisi, J., Scheithauer, B., Rodriguez, M. & Lassman, H. 2000. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Annals of Neurology* 47: 707-717.
- Maksimovic, J., Gordon, L. & Oshlack, A. 2012. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biology* 13: R44.
- Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics* 39: 906-913.
- Marian, A.J. 2012. Molecular genetic studies of complex phenotypes. *Translational Research* 159: 64-79.
- Martino, J., Hamer, P.C.D.W., Berger, M.S., Lawton, M.T., Arnold, C.M., de Lucas, E.M. & Duffau, H. 2013. Analysis of the subcomponents and cortical terminations of the perisylvian superior longitudinal fasciculus: a fiber dissection and DTI tractography study. *Brain Structure and Function* 218: 105-121.
- Matys, V. 2003. TRANSFAC(R): transcriptional regulation, from patterns to profiles. *Nucleic Acids Research* 31: 374-378.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutayavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R. & Stamatoyannopoulos, J.A. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337: 1190-1195.
- McMahon, F.J., Akula, N., Schulze, T.G., Muglia, P., Tozzi, F., Detera-Wadleigh, S.D., Steele, C.J., Breuer, R., Strohmaier, J., Wendland, J.R., Mattheisen, M., Muhleisen, T.W., Maier, W., Nothen, M.M., Cichon, S., Farmer, A., Vincent, J.B., Holsboer, F., Preisig, M., Rietschel, M. & Bipolar Disorder Genome Study, C. 2010. Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nature Genetics* 42: 128-131.
- McMahon, F.J., Stine, O.C., Meyers, D.A., Simpson, S.G. & DePaulo, J.R. 1995. Patterns of maternal transmission in bipolar affective disorder. *American Journal of Human Genetics* 56: 1277.
- Meaburn, E.L., Schalkwyk, L.C. & Mill, J. 2010. Allele-specific methylation in the human genome: implications for genetic studies of complex disease. *Epigenetics* 5: 578-582.
- Mill, J. & Heijmans, B.T. 2013. From promises to practical strategies in epigenetic epidemiology. *Nature Reviews Genetics* 14: 585-594.

- Miller-Delaney, S.F., Bryan, K., Das, S., McKiernan, R.C., Bray, I.M., Reynolds, J.P., Gwinn, R., Stallings, R.L. & Henshall, D.C. 2015. Differential DNA methylation profiles of coding and non-coding genes define hippocampal sclerosis in human temporal lobe epilepsy. *Brain* 138: 616-631.
- Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R. & Dermitzakis, E.T. 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464: 773-777.
- Moratz, C., Hayman, J.R., Gu, H. & Kehrl, J.H. 2004. Abnormal B-cell responses to chemokines, disturbed plasma cell localization, and distorted immune tissue architecture in Rgs1<sup>-/-</sup> mice. *Molecular Cell Biology* 24: 5767-5775.
- Morris, T.J. & Beck, S. 2015. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods* 72: 3-8.
- Nicholls, R.D., Knoll, J.H., Butler, M.G., Karam, S. & Lalande, M. 1989. Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. *Nature* 342: 281-285.
- Nordli, D.R. 2005. Idiopathic generalized epilepsies recognized by the International League Against Epilepsy. *Epilepsia* 46: 48-56.
- Numata, S., Ye, T., Herman, M. & Lipska, B.K. 2014. DNA methylation changes in the postmortem dorsolateral prefrontal cortex of patients with schizophrenia. *Frontiers in Genetics* 5: 280.
- Numata, S., Ye, T., Hyde, T.M., Guitart-Navarro, X., Tao, R., Wininger, M., Colantuoni, C., Weinberger, D.R., Kleinman, J.E. & Lipska, B.K. 2012. DNA methylation signatures in development and aging of the human prefrontal cortex. *The American Journal of Human Genetics* 90: 260-272.
- Ostrow, K.L., Michailidi, C., Guerrero-Preston, R., Hoque, M.O., Greenberg, A., Rom, W. & Sidransky, D. 2013. Cigarette smoke induces methylation of the tumor suppressor gene NISCH. *Epigenetics* 8: 383-388.
- Oyang, E.L., Davidson, B.C., Lee, W. & Poon, M.M. 2011. Functional characterization of the dendritically localized mRNA neuronatin in hippocampal neurons. *PLoS One* 6: e24879.
- Pembrey, M., Saffery, R., Bygren, L.O., Network in Epigenetic, E. & Network in Epigenetic, E. 2014. Human transgenerational responses to early-life experience: potential impact on development, health and biomedical research. *Journal of Medical Genetics* 51: 563-572.
- Pernhorst, K., Raabe, A., Niehusmann, P., Van Loo, K.M., Grote, A., Hoffmann, P., Cichon, S., Sander, T., Schoch, S. & Becker, A.J. 2011. Promoter variants determine  $\gamma$ -aminobutyric acid homeostasis-related gene transcription in human epileptic hippocampi. *Journal of Neuropathology & Experimental Neurology* 70: 1080-1088.
- Petretto, E., Mangion, J., Dickens, N.J., Cook, S.A., Kumaran, M.K., Lu, H., Fischer, J., Maatz, H., Kren, V., Pravenec, M., Hubner, N. & Aitman, T.J. 2006. Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genetics* 2: e172.
- Petronis, A. 2010. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 465: 721-727.
- Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y. & Pritchard, J.K. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464: 768-772.
- Pidsley, R., Viana, J., Hannon, E., Spiers, H., Troakes, C., Al-Saraj, S., Mechawar, N., Turecki, G., Schalkwyk, L., Bray, N. & Mill, J. 2014. Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. *Genome Biology* 15: 483.
- Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., Vorstman, J.A., Thompson, A., Regan, R., Pilorge, M., Pellecchia, G., Pagnamenta, A.T., Oliveira, B., Marshall, C.R., Magalhaes, T.R., Lowe, J.K., Howe, J.L., Griswold, A.J., Gilbert, J., Duketis, E., Dombroski, B.A., De Jonge, M.V., Cuccaro, M., Crawford, E.L., Correia, C.T., Conroy, J., Conceicao, I.C., Chiocchetti, A.G., Casey, J.P., Cai, G., Cabrol, C., Bolshakova, N., Bacchelli, E., Anney, R., Gallinger, S., Cotterchio, M., Casey, G., Zwaigenbaum, L., Wittmeyer, K., Wing, K., Wallace, S., van Engeland, H., Tryfon, A., Thomson, S., Soorya, L., Roge, B., Roberts, W., Poustka, F., Moug, S., Minshew, N., McInnes, L.A., McGrew, S.G., Lord, C., Leboyer, M., Le Couteur, A.S., Kolevzon, A., Jimenez Gonzalez, P., Jacob, S., Holt, R.,

- Guter, S., Green, J., Green, A., Gillberg, C., Fernandez, B.A., Duque, F., Delorme, R., Dawson, G., Chaste, P., Cafe, C., Brennan, S., Bourgeron, T., Bolton, P.F., Bolte, S., Bernier, R., Baird, G., Bailey, A.J., Anagnostou, E., Almeida, J., Wijsman, E.M., Vieland, V.J., Vicente, A.M., Schellenberg, G.D., Pericak-Vance, M., Paterson, A.D., Parr, J.R., Oliveira, G., Nurnberger, J.I., Monaco, A.P., Maestrini, E., Klauck, S.M., Hakonarson, H., Haines, J.L., Geschwind, D.H., Freitag, C.M., Folstein, S.E., Ennis, S., Coon, H., Battaglia, A., Szatmari, P., Sutcliffe, J.S., Hallmayer, J., Gill, M., Cook, E.H., Buxbaum, J.D., Devlin, B., Gallagher, L., Betancur, C. & Scherer, S.W. 2014. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics* 94: 677-694.
- Potapova, A., Albat, C., Hasemeier, B., Haeussler, K., Lamprecht, S., Suerbaum, S., Kreipe, H. & Lehmann, U. 2011. Systematic cross-validation of 454 sequencing and pyrosequencing for the exact quantification of DNA methylation patterns with single CpG resolution. *BMC Biotechnology* 11: 6.
- Psychiatric GWAS Consortium Bipolar Disorder Working Group 2012. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics* 44: 1072-1072.
- Psychiatric GWAS Consortium Bipolar Disorder Working Group 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics* 43: 977-983.
- Ptashne, M., Hobert, O. & Davidson, E. 2010. Questions over the scientific basis of epigenome project. *Nature* 464: 487-487.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I. & Daly, M.J. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81: 559-575.
- Raj, T., Ryan, K.J., Replogle, J.M., Chibnik, L.B., Rosenkrantz, L., Tang, A., Rothamel, K., Stranger, B.E., Bennett, D.A., Evans, D.A., De Jager, P.L. & Bradshaw, E.M. 2014. CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. *Human Molecular Genetics* 23: 2729-2736.
- Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., U. K. Brain Expression Consortium, North American Brain Expression Consortium, Coin, L., de Silva, R., Cookson, M.R., Singleton, A.B., Hardy, J., Ryten, M. & Weale, M.E. 2014. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature Neuroscience* 17: 1418-1428.
- Reich, D.E. & Lander, E.S. 2001. On the allelic spectrum of human disease. *TRENDS in Genetics* 17: 502-510.
- Reik, W., Brown, K.W., Schneid, H., Le Bouc, Y., Bickmore, W. & Maher, E.R. 1995. Imprinting mutations in the Beckwith—Wiedemann syndrome suggested by an altered imprinting pattern in the IGF2—H19 domain. *Human Molecular Genetics* 4: 2379-2385.
- Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., Kim, Y., Lee, S.H., Magnusson, P.K., Sanchez, N., Stahl, E.A., Williams, S., Wray, N.R., Xia, K., Bettella, F., Borglum, A.D., Bulik-Sullivan, B.K., Cormican, P., Craddock, N., de Leeuw, C., Durmishi, N., Gill, M., Golimbet, V., Hamshere, M.L., Holmans, P., Hougaard, D.M., Kendler, K.S., Lin, K., Morris, D.W., Mors, O., Mortensen, P.B., Neale, B.M., O'Neill, F.A., Owen, M.J., Milovancevic, M.P., Posthuma, D., Powell, J., Richards, A.L., Riley, B.P., Ruderfer, D., Rujescu, D., Sigurdsson, E., Silagadze, T., Smit, A.B., Stefansson, H., Steinberg, S., Suvisaari, J., Tosato, S., Verhage, M., Walters, J.T., Multicenter Genetic Studies of Schizophrenia Consortium, Levinson, D.F., Gejman, P.V., Kendler, K.S., Laurent, C., Mowry, B.J., O'Donovan, M.C., Owen, M.J., Pulver, A.E., Riley, B.P., Schwab, S.G., Wildenauer, D.B., Dudbridge, F., Holmans, P., Shi, J., Albus, M., Alexander, M., Campion, D., Cohen, D., Dikeos, D., Duan, J., Eichhammer, P., Godard, S., Hansen, M., Lerer, F.B., Liang, K.Y., Maier, W., Mallet, J., Nertney, D.A., Nestadt, G., Norton, N., O'Neill, F.A., Papadimitriou, G.N., Ribble, R., Sanders, A.R., Silverman, J.M., Walsh, D., Williams, N.M., Wormley, B., Psychosis

- Endophenotypes International Consortium, Arranz, M.J., Bakker, S., Bender, S., Bramon, E., Collier, D., Crespo-Facorro, B., Hall, J., Iyegbe, C., Jablensky, A., Kahn, R.S., Kalaydjieva, L., Lawrie, S., Lewis, C.M., Lin, K., Linszen, D.H., Mata, I., McIntosh, A., Murray, R.M., Ophoff, R.A., Powell, J., Rujescu, D., Van Os, J., Walshe, M., Weisbrod, M., Wiersma, D., Wellcome Trust Case Control Consortium, Donnelly, P., Barroso, I., Blackwell, J.M., Bramon, E., Brown, M.A., Casas, J.P., Corvin, A.P., Deloukas, P., Duncanson, A., Jankowski, J., Markus, H.S., Mathew, C.G., Palmer, C.N., Plomin, R., Rautanen, A., Sawcer, S.J., Trembath, R.C., Viswanathan, A.C., Wood, N.W., Spencer, C.C., Band, G., Bellenguez, C., Freeman, C., Hellenthal, G., Giannoulatou, E., Pirinen, M., Pearson, R.D., Strange, A., Su, Z., Vukcevic, D., Donnelly, P., Langford, C., Hunt, S.E., Edkins, S., Gwilliam, R., Blackburn, H., Bumpstead, S.J., Dronov, S., Gillman, M., Gray, E., Hammond, N., Jayakumar, A., McCann, O.T., Liddle, J., Potter, S.C., Ravindrarajah, R., Ricketts, M., Tashakkori-Ghanbaria, A., Waller, M.J., Weston, P., Widaa, S., Whittaker, P., Barroso, I., Deloukas, P., Mathew, C.G., Blackwell, J.M., Brown, M.A., Corvin, A.P., McCarthy, M.I., Spencer, C.C., Bramon, E., Corvin, A.P., O'Donovan, M.C., Stefansson, K., Scolnick, E., Purcell, S., McCarroll, S.A., Sklar, P., Hultman, C.M. & Sullivan, P.F. 2013. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics* 45: 1150-1159.
- Romanoski, C.E., Glass, C.K., Stunnenberg, H.G., Wilson, L. & Almouzni, G. 2015. Epigenomics: Roadmap for regulation. *Nature* 518: 314-316.
- Sampath, D., Zabka, T.S., Misner, D.L., O'Brien, T. & Dragovich, P.S. 2015. Inhibition of nicotinamide phosphoribosyltransferase (NAMPT) as a therapeutic strategy in cancer. *Pharmacology & Therapeutics*.
- Schalkwyk, L.C., Meaburn, E.L., Smith, R., Dempster, E.L., Jeffries, A.R., Davies, M.N., Plomin, R. & Mill, J. 2010. Allelic skewing of DNA methylation is widespread across the genome. *American Journal of Human Genetics* 86: 196-212.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511: 421-427.
- Schmermund, A., Möhlenkamp, S., Stang, A., Grönemeyer, D., Seibel, R., Hirche, H., Mann, K., Siffert, W., Lauterbach, K., Siegrist, J., Jöckel, K.-H. & Erbel, R. 2002. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: Rationale and design of the Heinz Nixdorf RECALL Study. *American Heart Journal* 144: 212-218.
- Schonberger, A., Niehusmann, P., Urbach, H., Majores, M., Grote, A., Holthausen, H., Blumcke, I., Deckert, M. & Becker, A.J. 2009. Increased frequency of distinct TSC2 allelic variants in focal cortical dysplasias with balloon cells and mineralization. *Neuropathology* 29: 559-565.
- Schramm, K., Marzi, C., Schurmann, C., Carstensen, M., Reinmaa, E., Biffar, R., Eckstein, G., Gieger, C., Grabe, H.J., Homuth, G., Kastenmuller, G., Magi, R., Metspalu, A., Mihailov, E., Peters, A., Petersmann, A., Roden, M., Strauch, K., Suhre, K., Teumer, A., Volker, U., Volzke, H., Wang-Sattler, R., Waldenberger, M., Meitinger, T., Illig, T., Herder, C., Grallert, H. & Prokisch, H. 2014. Mapping the genetic architecture of gene regulation in whole blood. *PLoS One* 9: e93844.
- Schurmann, C., Heim, K., Schillert, A., Blankenberg, S., Carstensen, M., Dorr, M., Endlich, K., Felix, S.B., Gieger, C., Grallert, H., Herder, C., Hoffmann, W., Homuth, G., Illig, T., Kruppa, J., Meitinger, T., Muller, C., Nauck, M., Peters, A., Rettig, R., Roden, M., Strauch, K., Volker, U., Volzke, H., Wahl, S., Wallaschofski, H., Wild, P.S., Zeller, T., Teumer, A., Prokisch, H. & Ziegler, A. 2012. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. *PLoS One* 7: e50938.
- Schwartz, D. 1962. Genetic studies on mutant enzymes in maize. III. Control of gene action in the synthesis of pH 7.5 esterase. *Genetics* 47: 1609.
- Shabalin, A.A. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28: 1353-1358.

- Shames, D.S., Elkins, K., Walter, K., Holcomb, T., Du, P., Mohl, D., Xiao, Y., Pham, T., Haverty, P.M., Liederer, B., Liang, X., Yauch, R.L., O'Brien, T., Bourgon, R., Koeppen, H. & Belmont, L.D. 2013. Loss of NAPRT1 expression by tumor-specific promoter methylation provides a novel predictive biomarker for NAMPT inhibitors. *Clinical Cancer Research* 19: 6912-6923.
- Sleiman, P., Wang, D., Glessner, J., Hadley, D., Gur, R.E., Cohen, N., Li, Q., Hakonarson, H. & Janssen Chop Neuropsychiatric Genomics Working Group 2013. GWAS meta analysis identifies TSNARE1 as a novel Schizophrenia / Bipolar susceptibility locus. *Scientific Reports* 3: 3075.
- Smith, A.K., Kilaru, V., Kocak, M., Almli, L.M., Mercer, K.B., Ressler, K.J., Tylavsky, F.A. & Conneely, K.N. 2014. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics* 15: 145.
- Smith, R.J., Dean, W., Konfortova, G. & Kelsey, G. 2003. Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Research* 13: 558-569.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K. & Schubeler, D. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480: 490-495.
- Steinlein, O.K. 2000. Neuronal nicotinic receptors in human epilepsy. *European Journal of Pharmacology* 393: 243-247.
- Sterneckert, J.L., Reinhardt, P. & Scholer, H.R. 2014. Investigating human disease using stem cell models. *Nature Reviews Genetics* 15: 625-639.
- Sullivan, P.F., Daly, M.J. & O'Donovan, M. 2012. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Reviews Genetics* 13: 537-551.
- Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., Vasanthakumar, A., Godley, L.A., Chang, Q., Cheng, X., He, C. & Jin, P. 2011. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature Neuroscience* 14: 1607-1616.
- Thompson, M. 2009. Polybromo-1: the chromatin targeting subunit of the PBAF complex. *Biochimie* 91: 309-319.
- Torres, J.M., Gamazon, E.R., Parra, E.J., Below, J.E., Valladares-Salgado, A., Wachter, N., Cruz, M., Hanis, C.L. & Cox, N.J. 2014. Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. *American Journal of Human Genetics* 95: 521-534.
- Tost, J. & Gut, I.G. 2007. DNA methylation analysis by pyrosequencing. *Nature Protocols* 2: 2265-2275.
- Tran, T., Paz, P., Velichko, S., Cifrese, J., Belur, P., Yamaguchi, K.D., Ku, K., Mirshahpanah, P., Reder, A.T. & Croze, E. 2010. Interferonbeta-1b Induces the Expression of RGS1 a Negative Regulator of G-Protein Signaling. *International Journal of Cell Biology* 2010: 529376.
- Trucks, H.S. 2013. *Association mapping of genomic microdeletions and common susceptibility variants predisposing to genetic generalized epilepsies*. PhD Doctoral Thesis, Universität zu Köln.
- Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A. & Meissner, A. 2015. Transcription factor binding dynamics during human ES cell differentiation. *Nature* 518: 344-349.
- Tseng, P.T., Lin, P.Y., Lee, Y., Hung, C.F., Lung, F.W., Chen, C.S. & Chong, M.Y. 2014. Age-associated decrease in global DNA methylation in patients with major depression. *Neuropsychiatric Disease and Treatment* 10: 2105-2114.
- Uddin, M., Tammimies, K., Pellecchia, G., Alipanahi, B., Hu, P., Wang, Z., Pinto, D., Lau, L., Nalpathamkalam, T., Marshall, C.R., Blencowe, B.J., Frey, B.J., Merico, D., Yuen, R.K. & Scherer, S.W. 2014. Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. *Nature Genetics* 46: 742-747.
- Urak, L., Feucht, M., Fathi, N., Hornik, K. & Fuchs, K. 2006. A GABRB3 promoter haplotype associated with childhood absence epilepsy impairs transcriptional activity. *Human Molecular Genetics* 15: 2533-2541.

- van Nas, A., Ingram-Drake, L., Sinsheimer, J.S., Wang, S.S., Schadt, E.E., Drake, T. & Lusis, A.J. 2010. Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. *Genetics* 185: 1059-1068.
- Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57-63.
- Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., McCorquodale, D.S., 3rd, Cuello, C., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Joshipura, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J.V., N. ACC-Neuropathology Group, Heward, C.B., Reiman, E.M., Stephan, D., Hardy, J. & Myers, A.J. 2009. Genetic control of human brain transcript expression in Alzheimer disease. *American Journal of Human Genetics* 84: 445-458.
- Wei, Y., Su, J., Liu, H., Lv, J., Wang, F., Yan, H., Wen, Y., Liu, H., Wu, Q. & Zhang, Y. 2014. MetalImprint: an information repository of mammalian imprinted genes. *Development* 141: 2516-2523.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L. & Parkinson, H. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* 42: D1001-1006.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K. & Peters, J.M. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451: 796-801.
- Westra, H.J. & Franke, L. 2014. From genome to function by studying eQTLs. *Biochimica et Biophysica Acta* 1842: 1896-1902.
- Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., Zernakova, A., Zernakova, D.V., Veldink, J.H., Van den Berg, L.H., Karjalainen, J., Withoff, S., Uitterlinden, A.G., Hofman, A., Rivadeneira, F., t Hoen, P.A., Reinmaa, E., Fischer, K., Nelis, M., Milani, L., Melzer, D., Ferrucci, L., Singleton, A.B., Hernandez, D.G., Nalls, M.A., Homuth, G., Nauck, M., Radke, D., Volker, U., Perola, M., Salomaa, V., Brody, J., Suchy-Dicey, A., Gharib, S.A., Enquobahrie, D.A., Lumley, T., Montgomery, G.W., Makino, S., Prokisch, H., Herder, C., Roden, M., Grallert, H., Meitinger, T., Strauch, K., Li, Y., Jansen, R.C., Visscher, P.M., Knight, J.C., Psaty, B.M., Ripatti, S., Teumer, A., Frayling, T.M., Metspalu, A., van Meurs, J.B. & Franke, L. 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics* 45: 1238-1243.
- Wiebe, S., Blume, W.T., Girvin, J.P. & Eliasziw, M. 2001. A randomized, controlled trial of surgery for temporal-lobe epilepsy. *New England Journal of Medicine* 345: 311-318.
- Witt, S.H., Juraeva, D., Sticht, C., Strohmaier, J., Meier, S., Treutlein, J., Dukal, H., Frank, J., Lang, M., Deuschle, M., Schulze, T.G., Degenhardt, F., Mattheisen, M., Brors, B., Cichon, S., Nothen, M.M., Witt, C.C. & Rietschel, M. 2014. Investigation of manic and euthymic episodes identifies state- and trait-specific gene expression and STAB1 as a new candidate gene for bipolar disorder. *Translational Psychiatry* 4: e426.
- World Health Organization 2008. *The global burden of disease: 2004 update*. World Health Organization, Geneva, Switzerland.
- Xia, K., Shabalin, A.A., Huang, S., Madar, V., Zhou, Y.H., Wang, W., Zou, F., Sun, W., Sullivan, P.F. & Wright, F.A. 2012. seeQTL: a searchable database for human eQTLs. *Bioinformatics* 28: 451-452.
- Xiao, Y., Camarillo, C., Ping, Y., Arana, T.B., Zhao, H., Thompson, P.M., Xu, C., Su, B.B., Fan, H., Ordonez, J., Wang, L., Mao, C., Zhang, Y., Cruz, D., Escamilla, M.A., Li, X. & Xu, C. 2014. The DNA Methylome and Transcriptome of Different Brain Regions in Schizophrenia and Bipolar Disorder. *PLoS One* 9: e95875.
- Ye, T., Lipska, B.K., Tao, R., Hyde, T.M., Wang, L., Li, C., Choi, K.H., Straub, R.E., Kleinman, J.E. & Weinberger, D.R. 2012. Analysis of copy number variations in brain DNA from patients with schizophrenia and other psychiatric disorders. *Biological Psychiatry* 72: 651-654.



- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., Min, J.H., Jin, P., Ren, B. & He, C. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149: 1368-1380.
- Zhang, D., Cheng, L., Badner, J.A., Chen, C., Chen, Q., Luo, W., Craig, D.W., Redman, M., Gershon, E.S. & Liu, C. 2010. Genetic control of individual differences in gene-specific methylation in human brain. *American Journal of Human Genetics* 86: 411-419.
- Zhang, X., Moen, E.L., Liu, C., Mu, W., Gamazon, E.R., Delaney, S.M., Wing, C., Godley, L.A., Dolan, M.E. & Zhang, W. 2014. Linking the genetic architecture of cytosine modifications with human complex traits. *Human Molecular Genetics* 23: 5893-5905.
- Zilberman, D. & Henikoff, S. 2007. Genome-wide analysis of DNA methylation patterns. *Development* 134: 3959-3965.
- Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., Gnirke, A. & Meissner, A. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500: 477-481.

## 6. APPENDIX

### 6.1 Clinical parameters of 115 mTLE patients

Parameters including gender, pathology and drug therapy are presented in percentage values. Age at epilepsy surgery in years and age at seizure onset in years are presented in mean  $\pm$  SEM values. Biopsy specimens were analyzed according to international standards and the diagnostic classification was established by an experienced neuropathologist according to international criteria (Becker et al. 2003; Blumcke et al. 2007). The hippocampi were stratified according to the pathological pattern of the patient into two groups: Ammon's horn sclerosis (segmental neuronal cell loss and concomitant astrogliosis and microglia activation) and lesion associated (cortical dysplasia or tumors).

**Table 6-1: Distribution of clinical parameters within sample group.**

<b>Number of mTLE patients</b>	115
<b>Gender (male vs. female)</b>	52.2% vs. 47.8%
<b>Age at seizure onset in years</b>	11.1 $\pm$ 8.7
<b>Drug therapy (Sodium-channel blockers monotherapy vs. levetiracetam combinations vs. non-levetiracetam combinations)</b>	19.1% vs. 36.6% vs. 44.3%
<b>Age at epilepsy surgery in years</b>	31.5 $\pm$ 14.2
<b>Pathology (Ammon's horn sclerosis vs. lesion associated)</b>	68.7% vs. 31.3%

### 6.2 Confirmation analyses of cis-meQTLs and cis-eQTLs in hippocampal brain tissue and whole blood cells

In recent years, most meQTL studies which assess methylation levels of postmortem brain tissue and whole peripheral blood cells used the *Infinium Human Methylation27 BeadChip* (HM27; Gibbs et al. 2010; Zhang et al. 2010; Numata et al. 2012; Gamazon et al. 2013; Smith et al. 2014; Numata et al. 2014), while only few used the HM450 array in different blood cell lines (Gutierrez-Arcelus et al. 2013; Banovich et al. 2014).

eQTL studies which assess expression levels of postmortem brain tissue use different arrays to profile mRNA transcript in comparison with my analysis (Gibbs et al. 2010; Gamazon et al. 2013; Ramasamy et al. 2014; Kim et al. 2014). As confirmation criteria the significant cis-eQTL SNPs from the four studies had to be within 1 Mbps of the mRNA transcription site. Additionally, the SNPs had to be in LD of  $r^2 > 0.2$  with my cis-eQTL SNPs. An eQTL analysis in whole blood cell samples was performed using also the *Illumina HumanHT-12 v3* (HT12v3) *expression BeadChip* (Schramm et al. 2014) and a meta-eQTL analysis in whole blood cells was performed by restricting their analysis to sequences

present on the HT12v3 platform (Westra et al. 2013). The previous meQTL and eQTL studies are described in Table 6-2.

*Table 6-2: Reference meQTL and eQTL studies exploited for confirmation analyses.*

Epigenetic study	Reference	Statistics	Array type	Tissue-/Cell-type	Number of individuals in QTL analyses	Phenotypic trait*
Methylation	Banovich et al. 2014	cis-meQTL	HM450	<b>Blood</b> , LCLs	64	none specific phenotypic trait
Methylation	Smith et al. 2014	cis-meQTL	HM27	<b>Whole blood &amp; 4 brain regions</b> , postmortem (CRBLM, FCTX, PONS, TCTX)	90 (blood), 105 (CRBLM), 111 (FCTX), 106 (PONS), 125 (TCTX)	none specific phenotypic trait
Methylation	Numata et al. 2012	cis- & trans-meQTL	HM27	<b>Brain</b> , postmortem (PFC)	108	none specific phenotypic trait
Methylation & Expression	Numata et al. 2014	cis-meQTL, correlation methylation/expression	HM27 & IL Human HT-12 v3 <sup>a</sup>	<b>Brain</b> , postmortem (DLPFC)	216 (106 with schizophrenia, 110 non-psychiatric controls)	Schizophrenia
Methylation & Expression	Gamazon et al. 2013	cis-meQTL & cis-eQTL, correlation methylation/expression	HM27 & Affy Human Gene 1.0 ST	<b>Brain</b> , postmortem (CRBLM)	153 (132 for correlation analysis)	Bipolar disorder
Methylation & Expression	Gutierrez-Arcelus et al. 2013	cis- meQTL & cis-eQTL, correlation methylation/expression	HM450 & RNAseq	<b>Blood</b> , fibroblasts, LCLs, T-cells	107-183 (fibroblasts), 11-185 (LCLs), 66-186 (T-cells)	none specific phenotypic trait
Methylation & Expression	Bell et al. 2011	cis- & trans meQTL, correlation methylation/expression	HM27 & RNAseq <sup>b</sup>	<b>Blood</b> , LCLs	77 (69 for correlation analysis)	none specific phenotypic trait
Methylation & Expression	Zhang et al. 2010	cis- & trans meQTL, correlation methylation/expression	HM27 & Affy HGU95Av2 <sup>c</sup>	<b>Brain</b> , postmortem (CRBLM)	153 (45 for correlation analysis)	none specific phenotypic trait
Methylation & Expression	Gibbs et al. 2010	cis- & trans-meQTL, cis- & trans-eQTL, correlation methylation/expression	HM27 & IL HumanRef-8	<b>Brain</b> , 4 regions, postmortem (CRBLM, FCTX, PONS, TCTX)	150 (from each human brain region)	none specific phenotypic trait
Expression	Ramasamy et al. 2014	cis- & trans-eQTL	Human exon 1.0 ST	<b>Brain</b> , 10 regions, postmortem <sup>d</sup>	134	Parkinson's disease and other brain disorders
Expression	Kim et al. 2014	Meta-analysis (5 studies), cis-eQTL	RNAseq, IL Human 49k oligo, IL HumanRef-8 v2, Affy HG-U133a	<b>Brain</b> , postmortem (FCTX, TCTX)	424 (235 in FCTX, 189 in FCTX 6 TCTX)	Psychiatric disorders

**Table 6-2 (continued): Reference meQTL and eQTL studies exploited for confirmation analyses.**

<b>Epigenetic study</b>	<b>Reference</b>	<b>Statistics</b>	<b>Array type</b>	<b>Tissue-/Cell-type</b>	<b>Number of individuals</b>	<b>Phenotypic trait*</b>
Expression	Schramm et al. 2014	cis- & trans eQTL	IL Human HT-12 v3	<b>Whole Blood</b>	890	Immune response & metabolic traits
Expression	Westra et al. 2013 <sup>e</sup>	Meta-analysis (7 studies), cis- & trans-eQTL	IL Human HT-12 v3, IL Human HT-12 v4, IL HumanRef-8 v2	<b>Whole Blood</b>	5,311	Type 1 diabetes & cholesterol metabolism

Abbreviations: eQTL, expression quantitative trait loci; meQTL, methylation quantitative trait loci; HM27, Illumina Human Methylation27k BeadChip; HM450, Illumina Human Methylation450k BeadChip; IL, Illumina; Affy, Affymetrix; RNA seq, RNA sequencing; CRBLM, cerebellum; DLPFC, dorsolateral prefrontal cortex; FCTX, frontal cortex; PFC, prefrontal cortex; PONS, caudal pons; TCTX, temporal cortex; LCL, lymphoblastoid cell line. \*Selections of traits from those highlighted in the given paper are shown; meQTLs and eQTLs are often compared to many more traits. <sup>a</sup> Numata et al. (2014) utilized previously published expression data from the DLPFC obtained using HumanHT-12\_V3 Illumina BeadArrays as described in detail in Ye et al. (2012). <sup>b</sup> Bell et al. (2011) used the RNA-sequencing data which were obtained for LCLs from 69 individuals in the study from Pickrell et al. (2010). <sup>c</sup> Zhang et al. (2010) used expression data from cerebellum for 45 of the same individuals, available from the SMRI Online Genomics Database. <sup>d</sup> The ten human brain regions from Ramasamy et al. (2014) included cerebellar cortex (CRBL), frontal cortex (FCTX), hippocampus (HIPPP), inferior olivary nucleus (sub-dissected from the medulla, MEDU), occipital cortex (OCTX), putamen (PUTM), substantia nigra (SNIG), temporal cortex (TCTX), thalamus (THAL), intralobular white matter (WHMT). <sup>e</sup> Westra et al. (2013) represents the largest eQTL meta-analysis so far.

### 6.3 Master regulatory loci of trans-meQTL SNPs

**Table 6-3: Master regulatory sites – trans-meQTL SNPs with simultaneous impact on the methylation of at least four genes.**

SNP	Chr SNP	SNP Pos	Gene of SNP	CpG	Chr CpG	CpG Pos	Gene of CpG	FDR
rs12137847	1	218227046	<i>RP11-152L7.1(20.0k)</i>	cg27508046	12	124249428	<i>DNAH10</i>	1.78E-03
	1	218227046	<i>RP11-152L7.1(20.0k)</i>	cg17892401	7	32529540	<i>LSM5</i>	1.16E-02
	1	218227046	<i>RP11-152L7.1(20.0k)</i>	cg03026929	6	32806028	<i>TAP2</i>	3.48E-02
	1	218227046	<i>RP11-152L7.1(20.0k)</i>	cg22684443	3	169756200	<i>GPR160</i>	4.90E-02
rs12464646	2	45339491	NA	cg09603795	10	75256221	<i>RP11-137L10.6</i>	2.87E-02
	2	45339491	NA	cg07805911	7	818287	<i>HEATR2</i>	3.74E-02
	2	45339491	NA	cg11637718	16	4029254	<i>ADCY9</i>	3.98E-02
	2	45339491	NA	cg00299396	11	115483607	<i>AP000997.1(15.2k)</i>	4.43E-02
	2	45339491	NA	cg05262711	16	2807476	<i>SRRM2</i>	4.92E-02
rs73944570	2	104844126	NA	cg14061772	12	73266969	<i>AC131213.1(82.5k)</i>	3.01E-03
	2	104844126	NA	cg20701799	3	10549576	<i>ATP2B2</i>	1.23E-02
	2	104844126	NA	cg22538778	12	63294404	<i>PPM1H</i>	4.22E-02
	2	104844126	NA	cg14047244	19	2457194	<i>GADD45B(18.9k)</i>	4.40E-02
rs114267096	6	31572801	<i>AIF1(10.2k)</i>	cg26221494	8	575388	<i>ERICH1</i>	5.10E-03
	6	31572801	<i>AIF1(10.2k)</i>	cg20814026	2	71205520	<i>AC007040.11</i>	1.32E-02
	6	31572801	<i>AIF1(10.2k)</i>	cg21685048	2	17771856	<i>VSNL1</i>	2.93E-02
	6	31572801	<i>AIF1(10.2k)</i>	cg14720996	3	130278864	<i>COL6A6(0.3k)</i>	4.78E-02
rs59758536	7	83391843	NA	cg12615916	1	227505854	<i>CDC42BPA</i>	5.03E-04
	7	83391843	NA	cg10519543	7	47694732	<i>C7orf65(0.1k)</i>	1.83E-03
	7	83391843	NA	cg13642142	9	139572213	<i>AGPAT2</i>	4.59E-03
	7	83391843	NA	cg04512931	17	43341971	<i>MAP3K14</i>	2.49E-02
	7	83391843	NA	cg19600667	4	6030497	<i>JAKMIP1</i>	4.80E-02
	7	83391843	NA	cg17013990	1	161091682	<i>DEDD</i>	4.94E-02
rs3920533	11	32297640	NA	cg24870476	1	221910211	<i>DUSP10</i>	3.14E-04
	11	32297640	NA	cg18496271	19	1388932	<i>NDUFS7</i>	2.29E-02

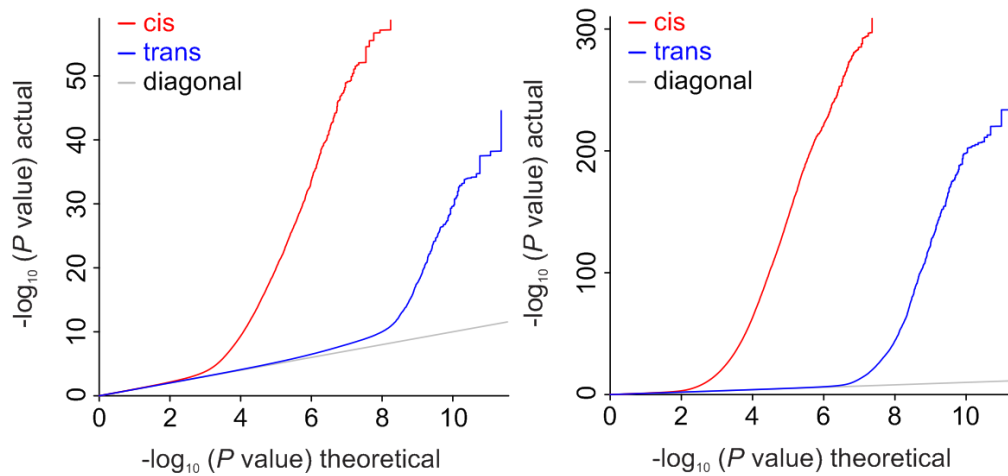
**Table 6-3 (continued): Master regulatory sites – trans-meQTL SNPs with simultaneous impact on the methylation of at least four genes.**

SNP	Chr SNP	SNP Pos	Gene of SNP	CpG	Chr CpG	CpG Pos	Gene of CpG	FDR
rs1273196	11	32297640	NA	cg11931953	7	4781767	FO XK1	4.44E-02
	11	32297640	NA	cg03906031	11	72489173	ARAP1	4.52E-02
	14	64739505	SYNE2(46.3k)	cg08822494	15	88279757	RP11-648K4.2(32.7k)	9.61E-04
	14	64739505	SYNE2(46.3k)	cg01845013	14	80697777	DIO2-AS1	1.18E-02
	14	64739505	SYNE2(46.3k)	cg18450582	7	95546539	DYNC111	1.71E-02
	14	64739505	SYNE2(46.3k)	cg26714129	8	142310710	SLC45A4	2.27E-02
	14	64739505	SYNE2(46.3k)	cg11377054	13	112189885	RP11-65D24.2(50.7k)	2.30E-02
chr19:40667344:D	14	64739505	SYNE2(46.3k)	cg21484956	12	77273469	RP11-461F16.3	2.46E-02
	19	40667344	MAP3K10(30.3k)	cg08008562	1	154530682	UBE2Q1	3.02E-02
	19	40667344	MAP3K10(30.3k)	cg04322363	3	159757071	CTD-2049J23.2	4.39E-02
	19	40667344	MAP3K10(30.3k)	cg10487970	2	63277030	OTX1(0.2k)	4.46E-02
	19	40667344	MAP3K10(30.3k)	cg22539294	3	167033938	ZBBX	4.57E-02

Abbreviations: SNP, SNP of the significant trans-meQTLs in hippocampal brain tissue; Chr SNP, chromosome of the respective SNP; SNP Pos, base-pair position of SNP in hippocampal brain tissue; Gene of SNP, gene which contains the SNP site; CpG, Illumina annotation of the CpG dinucleotide; Chr CpG, chromosome of the respective CpG; CpG Pos, cytosine base-pair position of CpG site; Gene of CpG, gene which contains the CpG site; FDR, significant FDR for a SNP-CpG pair in hippocampal brain tissue samples. All annotations refer to genome build hg19.

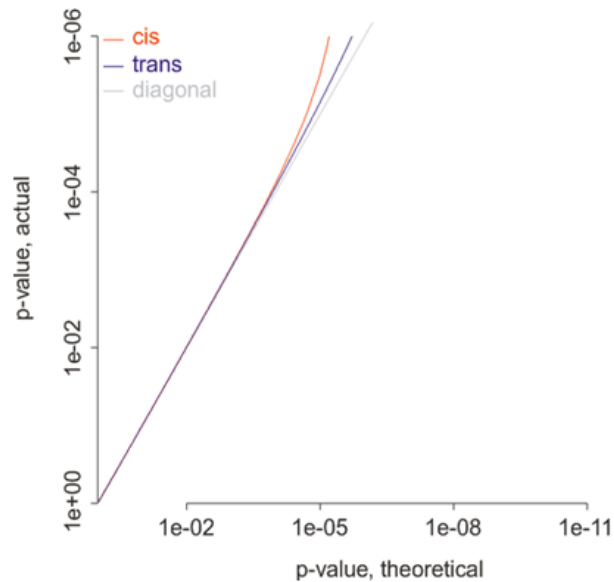
## 6.4 QQ-plots of QTL analyses

### 6.4.1 QQ-plots of meQTL analysis



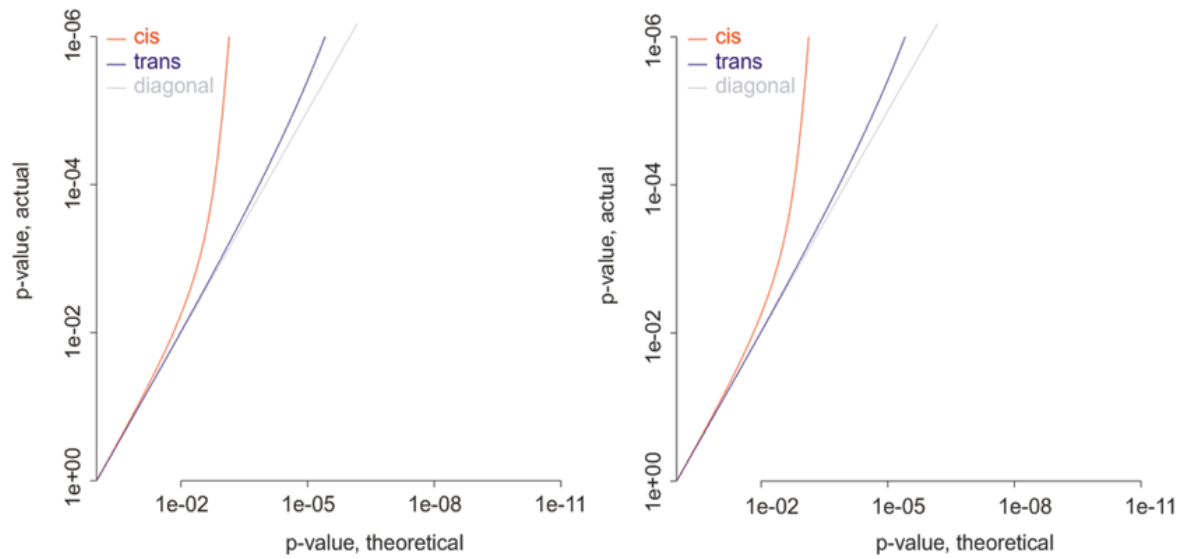
**Figure 6-1: QQ-plot of CpG methylation in 115 hippocampal brain tissue samples and 496 blood cell samples.** QQ-plots in 115 hippocampal brain tissue samples (left side) and 496 blood cell samples (right side). Distribution of cis-meQTLs ( $\pm 1$  Mbps) is in red; distribution of trans-meQTLs is in blue.

### 6.4.2 QQ-plots of imeQTL analysis



**Figure 6-2: QQ-plot of CpG methylation in 269 parent-offspring trios (heterozygote test).** QQ-plots in 269 parent-offspring trios of the imeQTL heterozygote test. Distribution of cis-imeQTLs ( $\pm 1$  Mbps) is in red; distribution of trans-imeQTLs is in blue.

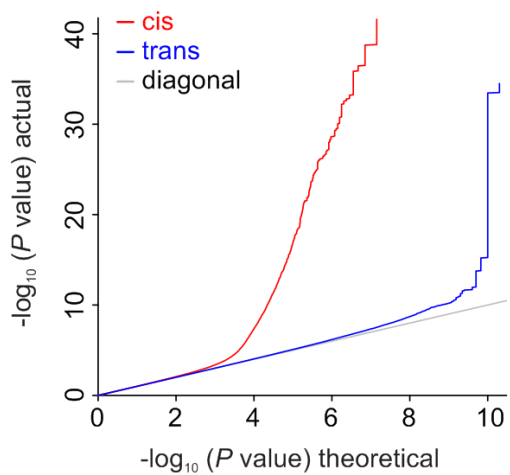




**Figure 6-3: QQ-plot of CpG methylation in 269 parent-offspring trios (maternal and paternal imeQTL analysis).**

QQ-plots in 269 parent-offspring trios of the maternal imeQTL analysis (left side) and the paternal imeQTL analysis (right side). Distribution of cis-imeQTLs ( $\pm 1$  Mb) is in red; distribution of trans-imeQTLs is in blue.

#### 6.4.3 QQ-plot of eQTL analysis



**Figure 6-4: QQ-plot of mRNA expression in 115 hippocampal brain tissue samples.**

Distribution of cis-eQTLs ( $\pm 1$  Mb) is in red; distribution of trans-eQTLs is in blue.

## 7. DANKSAGUNG

Zuallererst möchte ich mich ganz herzlich bei meinem Doktorvater Prof. Dr. Peter Nürnberg für die Übernahme der Erstkorrektur und die Begutachtung bedanken. Des Weiteren bedanke ich mich für seine Unterstützung bei meiner Dissertation und die Bereitstellung eines hochmodernen Labors mit seinem kompetenten Team. Vielen Dank für die Ermöglichung der vorliegenden Arbeit.

Prof. Dr. Peter Schneider danke ich sehr für die Übernahme der Zweitkorrektur und Begutachtung dieser Arbeit. Bei Prof. Dr. Peter Kloppenburg bedanke ich mich für die Übernahme des Vorsitzes in meiner mündlichen Prüfung.

Vor allem danke ich auch Dr. habil. Thomas Sander, der mir unter intensiver und exzellenter Betreuung und Anleitung die Möglichkeit gegeben hat, in seiner Arbeitsgruppe meine Doktorarbeit anzufertigen. Darüber hinaus vielen Dank für die zahlreichen Anregungen und interessanten Diskussionen rund um das Thema Wissenschaft und Forschung und letztendlich für das Korrekturlesen dieser Arbeit. Seine Kooperationsfähigkeit mit vielen Epilepsiezentren hat mir die Arbeit an diesem einmaligen Patientenkollektiv ermöglicht.

Ein ganz besonderer Dank geht auch an Dr. Herbert Schulz, bei dem ich mich sehr für die großartige Unterstützung, die zahlreichen Anregungen und Hilfestellungen bedanken möchte. Erst von Berlin aus und dann vor Ort in Köln. Vielen Dank für die Einführung in R und die zahlreichen Skripte. Ferner gilt mein Dank der ganzen AG Sander, mittlerweile alles Ehemalige: danke vor allem an Dr. Dennis Lal und Dr. Holger Trucks für die Einführung in die Epilepsie Genetik, die gute Zusammenarbeit und auch nette gemeinsame Unternehmungen. Danke auch an Dr. Costin Leu, Kerstin Wodecki, Anna Schlizio und Sebastian Fey.

In besonderer Weise möchte ich auch Prof. Dr. Sven Cichon, Dr. Per Hoffmann und Dr. Andrea Hofmann aus dem Life & Brain sowie Prof. Dr. Albert Becker und Dr. Katharina Pernhorst aus der Epileptologie in Bonn danken – die Zusammenarbeit und der Austausch mit euch sowie eure Hilfestellungen haben einen großen Anteil zu meiner endgültigen Arbeit beigetragen. In schöner Erinnerung sind mir die vielen gemeinsamen Meetings. Ich freue mich auf die weitere Zusammenarbeit mit euch.

Außerdem danke ich dem gesamten CCG für die freundliche und warme Arbeitsatmosphäre. Vor allem Mohammad Toliat und Nina Dalibor danke ich für die stets freundliche und erfahrene Unterstützung im Labor und Wilfried Gunia für die technische Unterstützung.

Nicht zuletzt möchte ich mich sehr bei meiner Mutter bedanken die mich immer vorbehaltlos unterstützt hat und immer für mich da ist. Abschließend möchte ich mich in ganz besonderer Weise bei meinem Mann Jan bedanken – danke für deine Liebe, deine Geduld, deinen Zuspruch und deine großartige Unterstützung. Danke, dass du bei mir bist. Mein Dank gilt weiterhin meinen Freunden, die mir auch in stressigen Zeiten immer wieder eine willkommene Ablenkung geboten haben.

## 8. ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Peter Nürnberg betreut worden.

Köln, den 11.05.2015

---

(Ann-Kathrin Ruppert)

## 9. CURICULUM VITAE

# Ann-Kathrin Ruppert

## (geb. Kriegel)

\* 11. Dezember 1984  
 in Köln, NRW  
 Mayener Str. 8  
 50935 Köln - Sülz  
 ☎ 0178 – 1304002  
 ✉ a.ruppert@uni-koeln.de

### Familienstand

verheiratet, keine Kinder

### Sprachkenntnisse

Englisch (fließend in Wort und Schrift)  
 Spanisch (sehr gute Kenntnisse in Wort und Schrift)  
 Französisch (Grundkenntnisse)

### Promotionsstudium

*Seit 2011*

**Promotionsstudentin der Biologie** an der Mathematisch-Naturwissenschaftlichen-Fakultät, Universität zu Köln  
 Angestrebter Titel: Dr. rer. nat.

Titel der Dissertation:

**"Integrative functional genomic search for regulatory DNA sequence polymorphisms influencing DNA methylation and mRNA expression in hippocampal brain tissue"**, Doktorvater: Prof. Dr. Peter Nürnberg, Epilepsie Genetik, Cologne Center for Genomics

### Universitäres Studium

*2004 – 2010*

**Diplomstudiengang Biologie**, Universität zu Köln  
 Abschlussnote: sehr gut

Diplomarbeit:

**„Entwicklung eines autosomalen SNP-Multiplex zur Differenzierung West-Eurasischer Populationen für forensisch-genetische Anwendungen“**, Betreuer: Prof. Dr. Peter Schneider, Fachbereich Molekulargenetik, Institut für Rechtsmedizin, Universitätsklinikum Köln  
 Note der Diplomarbeit: sehr gut (1,3)

*2006 – 2010*

Stipendiatin der **Studienstiftung des Deutschen Volkes**

### Praktika und

### Auslandsaufenthalte

*05/2009 – 06/2009*

- 5-wöchiges Fachpraktikum am **Institut für Rechtsmedizin, Santiago de Compostela, Spanien**. im Rahmen einer Forschungs Kooperation mit dem Institut für Rechtsmedizin des Universitätsklinikums Köln  
 Schwerpunkt: Autosomale SNP-Multiplex-Analysen

*07/2008 – 08/2008*

- 5-wöchiges Fachpraktikum im **Institut für Rechtsmedizin des Universitätsklinikums Köln**  
 Schwerpunkt: Forensische Molekularbiologie

<p>02/2008 – 04/2008</p> <p>09/2007</p> <p><b>Gymnasialzeit</b> 1995 – 2004</p>	<ul style="list-style-type: none"> <li>– Zweieinhalb monatiger Aufenthalt in Mexiko, inklusive Sprachkurs und 6-wöchigem Fachpraktikum im <b>Centro de Investigación de Biomédica de Occidente (CIBO)</b>, Abteilung Neurologie, <b>Guadalajara, Mexiko</b> Schwerpunkt: Untersuchungen zu Alzheimer und anderen neurologisch bedingten Krankheiten</li> <li>– 3-wöchiges Fachpraktikum im <b>Institut für Neurophysiologie, Medizinische Fakultät des Universitätsklinikums Köln</b> Schwerpunkt: Ionenkanal-Krankheiten</li> </ul> <p>Gymnasium Köln Nippes, Abschlussnote: gut (2,0) Leistungskurse: Spanisch und Chemie</p>
<p><b>Mitgliedschaften in wissenschaftlichen Konsortien</b></p>	<p>ESF EUROCORES <b>EuroEPINOMICS (EpiGENet)</b> <b>ILAE</b> (International League Against Epilepsy) <b>Consortium on Complex Epilepsies</b></p>
<p><b>Publikationen</b></p>	<p>Lal D*, <b>Ruppert AK*</b>, Trucks H, Schulz H, de Kovel CG, Kasteleijn-Nolst Trenité D et al. 2015. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalized epilepsies. <i>PLoS Genet</i> 11(5): e1005226.</p> <p>Galizia EC*, Myers CT*, Leu C*, de Kovel CG, Afrikanova T, ..., <b>Ruppert AK</b>, ... et al. 2015. CHD2 variants are a risk factor for photosensitivity in epilepsy. <i>Brain</i>: awv052.</p> <p>Hallmann K, Zsurka G, Moskau-Hartmann S, Kirschner J, Korinthenberg R, <b>Ruppert AK</b> et al. 2014. A homozygous splice-site mutation in CARS2 is associated with progressive myoclonic epilepsy. <i>Neurology</i> 83(23): 2183-7.</p> <p>International League Against Epilepsy Consortium on Complex Epilepsies. 2014. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. <i>Lancet Neurol</i> 13(9): 893-903.</p> <p>EPICURE Consortium, EMINet Consortium, Steffens M, Leu C, <b>Ruppert AK</b>, Zara F et al. 2012. Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. <i>Hum Mol Genet</i> 21(24): 5359-72.</p> <p>Phillips C, Freire Aradas A, <b>Kriegel AK</b>, Fondevila M, Bulbul O, Santos C et al. 2013. Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries. <i>Forensic Sci Int Genet</i> 7(3): 359-66.</p> <p>Bulbul O, Filoglu G, Altuncul H, Freire Aradas A, Ruiz Y, ..., <b>Kriegel AK</b>, ... et al. 2011. A SNP multiplex for the simultaneous prediction of biogeographic ancestry and pigmentation type. <i>Forensic Sci Int Gene Suppl</i> 3(1): e500–e501.</p> <p>Freire-Aradas A, Fondevila M, <b>Kriegel AK</b>, Phillips C, Gill P, Prieto L et al. 2011. A new SNP assay for identification of highly degraded human DNA. <i>Forensic Sci Int Genet</i> 6(3): 341-9.</p> <p>Tomas C, Axler-DiPerte G, Budimlija ZM, Børsting C, Coble MD, ..., <b>Kriegel AK</b>, ... et al. 2011. Autosomal SNP typing of forensic samples with the GenPlex™ HID System: results of a collaborative study. <i>Forensic Sci Int Genet</i> 5(5): 369-75.</p>