An epigenomic and omics approach to neurodevelopmental disorders

Namitha Mohandas

ORCID: 0000-0002-8743-8384

Doctor of Philosophy

July 2020

Faculty of Medicine, Dentistry and Health Sciences

School of Paediatrics

Submitted in total fulfilment for the degree of Doctor of Philosophy

This page has been left blank intentionally

Abstract

Neurodevelopmental disorders such as cerebral palsy (CP) and epilepsy are some of the most prevalent childhood neurological disorders caused by damage to the growth and development of the brain. Early life environments predispose children to later health outcomes evidenced by the developmental origins of health and disease (DOHaD) phenomenon. Epigenetics, which refers to modifications of DNA without change in DNA sequence, is one way by which environmental exposures may contribute to development of disease. DNA methylation, arguably the most highly studied epigenetic mark, has been correlated with early life environmental exposures and have implications in both disease mechanisms as well as clinical biomarkers of neurodevelopmental diseases. These modifications most likely originate in utero, in line with the DOHaD hypothesis. The study of monozygotic (MZ) twins, in which genetics, age, sex, parental factors and shared environment are controlled for, helps in distinguishing the extent of effect of genetics and environment. Discordance for neurodevelopmental disorders has been recorded in MZ twins indicating a potential role of non-shared factors in disease risk. The aim of this PhD was to utilise the discordant MZ twin model to understand epigenetic changes associated with neurodevelopmental disorders.

Genome-wide DNA methylation was measured within MZ twin cohorts discordant for CP or epilepsy using Illumina's Infinium HumanMethylation450 and EPIC arrays. Statistical and bioinformatics pipelines were applied to evaluate the association of DNA methylation data to disease phenotypes.

As detailed in Chapter Three of this thesis, DNA methylation analysis of CP-discordant twin pairs provides the first evidence that environmentally mediated differential methylation in genes involved in known processes such as hypoxia and inflammation, and processes such as cell adhesion, may contribute to the development of CP.

As detailed in Chapter Four, an epigenome-wide analysis of epilepsy discordant MZ twin pairs revealed distinct patterns of DNA methylation within subtypes of epilepsies of unknown cause. Differentially methylated genes within epilepsy subtypes included those with a role in metabolic pathways, voltage-gated channel signalling and neurotransmitter processes.

This research paves the way for future larger studies, as understanding DNA methylation profiles associated with neurodevelopmental disorders, may facilitate biomarkers for earlier diagnosis. This could lead to possible intervention strategies for patients suffering from a broad spectrum of disorders. Analysing epigenetic data from disease discordant twins provides an elegant study design and has the power to explore non-shared environmental factors that further refine models of disease mechanisms and biomarkers.

The findings of this thesis suggest that epigenetic factors may play a role regulating biological pathways that underlie neurodevelopmental disorders, some of which arise as early as the prenatal period. Replication in other larger and similar cohorts of discordant twin pairs may provide novel targets for biomarker development, thereby allowing for early interventions and helping the health of children.

Declaration

This is to certify that:

(i) This thesis comprises only my original work towards the PhD except where indicated in the preface.

(ii) Due acknowledgement has been made in the text to all materials used.

(iii) The thesis is fewer than the maximum word limit (100,000 words) in length, exclusive of tables, bibliographies, and appendices

Date: 15 July 2020

Signed: Namitha Mohandas

Namitha Mohandas

Department of Paediatrics

Faculty of Medicine, Dentistry and Health Sciences

Preface

During the last few years it has become apparent that most chronic health conditions, from heart disease to psychiatric disorders, originate early in life. The basis of this doctoral thesis is UNIQUE cohorts of monozygotic (identical) twin pairs, in which one of the twins is affected by cerebral palsy or epilepsy. These cohorts have been established by A/Prof. Jeff Craig, Dr. Kylie Crompton and A/Prof. Lata Vadlamudi and have been recruited by the supervisor team to investigate specific epigenetic markers and characterise the variation recorded within each twin pair. The study involved bioinformatics and bio statistical analyses for DNA methylation data with a minor component of lab work, but with the majority of lab work performed by Dr. Yuk J. Loke and past Honours students. My involvement with the project commenced in December 2015, during the initial stages of DNA methylation data processing of UNIQUE CP twin cohort.

The following people contributed towards this work and I would like to acknowledge their contributions:

Associate Professor Jeff Craig, Dr. Kylie Crompton, Professor Dinah Reddihough for conceiving and designing the UNIQUE CP study, which include designing the study, setting up the protocols, organising collaborations, and contributing to data interpretation. Dr. Kylie Crompton, Dr. Sue M. Reid and Dr. David J Amor were involved with ethics, patient recruitment and sample collections. Dr. Yuk J. Loke and Sebastian Bass-Stringer performed the lab work required for this study.

Associate Professor Lata Vadlamudi for designing and establishing the epilepsy study, including ethics, patient recruitment and sample collections. Dr. Yuk J. Loke and Stephanie Hopkins performed the lab work required for this study. Professor Sam Berkovic for assisting with twin recruitment (Epilepsy Research Centre).

Associate Professor Alicia Oshlack and Dr. Jovana Maksimovic from the Bioinformatics group, Murdoch Children's Research Institute (MCRI), for the advice given for statistical analysis in Chapters 3, 4 and 5.

Sebastian Bass-Stringer and Stephanie Hopkins for the validation of significant differentially methylated regions on an independent methylation platform (Chapter 3 and Chapter 4).

Research Assistants Lisa Mackenzie and Carmen Bennett for helping with recruitment and collection of study samples (Chapter 4 and Chapter 5).

The Victorian Cerebral Palsy Register staff for their help with consents and recruitment of participants included in this study. Prof John Hopper (Twins Research Australia); Prof Nick Martin (Queensland Institute of Medical Research Berghofer (QIMR-B)) for their assistance with twin recruitment for the epilepsy study.

Ethics clearance

Cerebral palsy study: Approval was obtained by The Royal Children's Hospital Human Research Ethics Committee (project ID: 33050), and involved written informed consent from each participating family was collected.

Epilepsy study: Informed consent was obtained from each twin individual. A multi-centre ethics approval was obtained with the lead site being Mater Health Services (ethics approval number HREC/13/MHS/114).

Funding and Grants

The UNIQUE CP cohort study was supported by grant from the Financial Markets Foundation for Children (grant number 2013-207). The epilepsy cohort study was

supported by grant from the Financial Markets Foundation for Children (grant number 2013-207). The University of Queensland Mayne Bequest Funds, Royal Brisbane and Women's Foundation Grant, Viertal-Charitable Foundation, Ramaciotti Establishment Grant and The Servier Barry Young Fellowship in Neurology also supported this work.

My PhD candidature was supported by the National Health and Medicine Research Council (NHMRC) Dora Lush Postgraduate Research Scholarship.

Acknowledgements

I would like to express my sincere gratitude to my principal supervisor, Associate Professor Jeffrey Craig, for his continuous support and immense knowledge throughout my PhD. His guidance along with his motivation and enthusiasm for scientific research has inspired me on many occasions. I would also like to thank him for his patience throughout the process of writing this thesis. Without your understanding, this PhD would not have been possible. Thank you for also supporting me financially during the last six months of my PhD.

Besides my supervisor, I would like to thank my co-supervisors Dr. Kylie Crompton, Associate Professor Lata Vadlamudi and Associate Professor Alicia Oshlack for their expert advice – I have learnt so much from our discussions. Thank you for giving me the opportunity to work with you and your team, and for the patience you have shown during some difficult times in my PhD.

My sincere thanks to my advisory committee chair, Professor Richard Saffery for your valuable suggestions and for being supportive throughout my PhD journey. I would also like to thank my advisory committee members, both past and present, Dr. David Martino, Associate Professor Justine Ellis, Dr. Boris Novakovic and Dr. Jane Loke for giving me thoughtful feedbacks during my progress review meetings.

I would like to thank the twin participants and their families, without whom this research would not have been possible and the recruitment teams for reaching out to them.

I can't thank the MCRI Early Life Epigenetics team enough (Dr, Pamela Leong, Samantha Dawson, Mihiri Silva, Lada Staskova, Jennifer Snowball) for being the nicest, kindest, most inspiring people to work with. Special thanks to Dr. Jane Loke for being my mentor and for doing all the lab work that was involved in this PhD project. I

am deeply grateful for our friendship and for all the advice you have given me over the years, both on an academic and a personal level. I also want to thank my advisors, peers and ex-colleagues whose stimulating discussions, life-lessons and friendship has made this a very memorable journey.

To my close friends, thank you for being loving, caring friends and my biggest support system. Thanks for cheering me on when life got tough and for never missing a chance to celebrate my little joys of success.

My heartfelt thanks to my mom, dad, brother and sister-in-law for their constant love, support and encouragement. Thank you for always being there for me and for believing in me.

Thank you to my husband, Shyam, for all the sacrifices you have made so that I could achieve my dream.

Publications arising from this thesis

1. **Mohandas N**, Loke YJ, Mackenzie L, Bennett C, Berkovic SF, Craig JM, Vadlamudi L. Deciphering the role of epigenetics in self-limited epilepsy with centrotemporal spikes. Epilepsy Research 2019. 156:106163.

2. **Mohandas N**, Loke YJ, Hopkins S, Mackenzie L, Bennett C, Berkovic SF, Vadlamudi L, Craig JM. Evidence for type-specific DNA methylation patterns in epilepsy: a discordant monozygotic twin approach. Epigenomics 2019. 11:951-968.

3. **Mohandas N**, Bass-Stringer S, Maksimovic J, Crompton K, Loke JY, Walstab J, Reid SM, Amor DJ, Reddihough D, Craig JM. Epigenome-wide analysis in newborn blood spots from monozygotic twin pairs discordant for cerebral palsy reveals consistent regional differences in DNA methylation. Clinical Epigenetics. 2018 10, 25.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Neurodevelopmental disorders (NDD) describe a wide range of disabilities of motor function, cognition, behaviour or communication impairments or psychiatric disorders resulting from dysfunction in the growth or development of the brain. Common NDDs include communication disorders such as autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), intellectual disability; motor disorders such as developmental coordination disorder (DCD) and cerebral palsy (CP); and other genetic disorders such as Rett syndrome or epilepsy. These conditions are likely to have an early onset in early childhood and can range from developmental deficits affecting specific learning and control of executive functions to global impairments of social skills or intelligence. Global prevalence of communication disorders typically ranges from 0.5 to 2% of the population across countries. Early diagnosis is often difficult given the symptoms and behaviours of neurodevelopment often evolve, as the child grows older.

Epilepsy is a neurological condition associated with increased morbidity and mortality. The social and psychological repercussions can be stigmatizing for people affected by this condition. Furthermore, epilepsy carries a substantial economic burden for health systems, individuals and their families.

Genetics has always played an important role in neuroscience research and massive progress has been made to understand the inheritance patterns of various neurological diseases. A salient feature in many studies is that the genetic aetiology of such diseases is immensely complex (Petrovski & Kwan, 2013). However, there is increasing understanding that most human diseases result from gene-environmental interactions, the latter being most potent in early life. Distinguishing the extent of effect of nature (genetics) and nurture (environment) is confounded by a large number of variables. The most elegant way of approaching this difficulty has been by studying twins. Such studies open up the possibility of singling out environmental and developmental effects that may contribute to disease aetiology and facilitate in biomarker development. Biomarkers can be of two types, *diagnostic biomarkers* that provide discrete and objective indication of diagnostic status or *screening biomarkers* that would allow determination of risk status of a condition. Such biomarkers can be of significant

translational value as they allow the determination of diagnostic risk prior to the appearance of behaviour symptoms, thereby making early detection and intervention possible in the case of cerebral palsy. Minimising the functional and social impacts of neurodevelopmental disorders, such as epilepsy, would represent a significant advancement for large numbers of children with these life-long conditions which not only impact on the individual child but on their families too.

This introduction will provide a broad overview of the epidemiology and impact of **neurodevelopmental disorders with a focus on cerebral palsy and epilepsy**. It will provide a comprehensive review of the current evidence regarding genetic and environmental risk and aetiological factors with particular focus on the unique potential of twin studies in such complex disorders.

1.1 Neurodevelopmental disorders

1.1.1 Types, Symptoms, Prevalence

Neurodevelopmental disorders are quite complex and in the past decade they have undergone substantial progress in diagnosis and classification. In 2014, the new Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (American Psychiatric Association, 2013) classified neurodevelopmental disorders based on adaptive functioning instead of IQ scores (JC, 2013; Salvador-Carulla L, 2011). Communication disorders were classified as a separate category and motor disorders were re-organised. The classification system also reflects a developmental approach to the classification of each of the disorders. The global burden of disease study in 2016 reported that although the burden of mortality among children younger than 5 years decreased by half between 1990 and 2016, there was no corresponding improvement in non-fatal health outcomes among children with developmental disabilities globally (GBD 2016).

Cerebral palsy and epilepsy are two of the key prevalent disorders of neurodevelopment in children. Both these disorders have been described to coexist, with studies reporting epilepsy diagnosis in up to 50% of children with cerebral palsy (Hadjipanayis, Hadjichristodoulou, & Youroukos, 1997; Singhi, Jagirdar, Khandelwal, & Malhi, 2003). Most neurodevelopmental disorders are known to co-occur and rarely present itself in isolation. However, the symptoms and behaviours associated with a particular neurodevelopmental disorder can change as children grow and mature. As a result, diagnosis of these disorders can be challenging. The prevalence rates of cerebral palsy was reported to be 1.5 to more than 4 per 1,000 live births (Moshe Stavsky, 2017) and 1 in 26 will develop epilepsy during their lifetime (DC Hesdorffer, 2011). In Australia, rates are estimated as 1 in 700 for CP (https://www.cerebralpalsy.org.au/sstposts/StoryId1543990187907) and around 250,000 people or 1% of the population are currently diagnosed with epilepsy (https://www.epilepsy.org.au/about-epilepsy/).

1.1.1.1 Cerebral palsy

Cerebral palsy (CP) is a clinical description of a group of heterogeneous motor (movement) impairment syndromes resulting from lesions or anomalies of the brain. CP is usually non-progressive although changes in the severity of the disorder can occur. The condition generally arises in the early stages of development and is therefore considered to most likely have causes occurring *in utero* (P. Rosenbaum, 2007). CP is the most common physical disabilities occurring in childhood (Novak, Hines, Goldsmith, & Barclay, 2012) and preterm infants are at a higher risk of the disease (Elovitz, Mrinalini, & Sammel, 2006). CP risk is usually diagnosed around 2-3 years of age, as the symptoms of this disorder are heterogeneous in nature thereby making it hard to define the disease. Primarily, CP affects movements and posture limiting physical capabilities. It can often also be accompanied by difficulties in sensation, perception, cognition, communication and behaviour as well as secondary musculoskeletal problems (P. Rosenbaum, 2007).

Cerebral palsy definition and diagnosis

The diagnosis of CP is primarily dependent on clinical assessment of the individual and is typically based on parental reports of the child's developing motor skills. The definition of cerebral palsy therefore encompasses the limitations of development of movement and posture that are attributed to the disturbances occurring in the infant brain. The recent international working group of cerebral palsy defined it is as follows: "Cerebral palsy describes a group of permanent disorders of the development of movement and posture, causing activity limitation, that are attributed to non-progressive disturbances that occurred in the developing fetal or infant brain".

Clinical diagnosis usually involves both neurological as well as motor skills assessment and repeated examination of children as some neurological abnormalities observed in the early months of preterm born children may not be associated with motor impairment and may resolve during the early years of life. Categorisation of motor skills is based on a widely used standardised measure of motor function, the Gross Motor Function Classification System (GMFCS) (K. Bjornson, 1998; Russell DJ, 1989), which categorises CP individuals into groups based on severity (**Figure 1.1** and **Figure 1.2**) even though each case may have unique characteristics. The grouping was traditionally based on the pattern of distribution of the affected limbs and then further grouped by the type of movement abnormality.

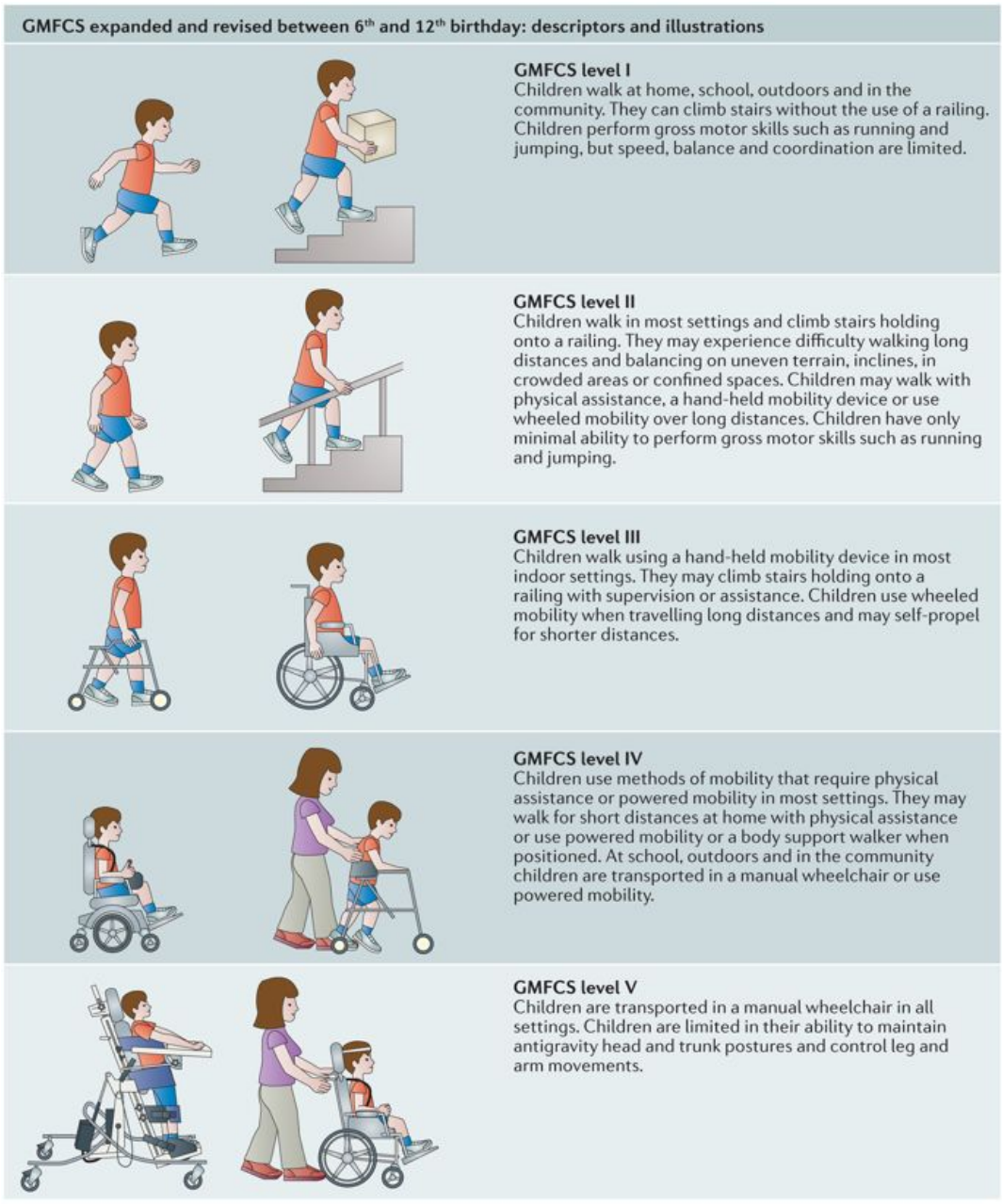


Figure 1.1: Gross Motor Function Classification System (GMFCS) for children with cerebral palsy

(Taken from Graham et al., 2016: The Gross Motor Function Classification System (GMFCS) has become the gold standard to classify motor function in children with cerebral palsy. The GMFCS is an ordinal classification in which different descriptors are used according to the age of the child. The descriptors for children 6–12 years of age are shown. GMFCS has been shown to be valid, reliable, stable and predictive of long-term gross motor function. Images are courtesy of B. Reid, A. Harvey and H.K.G., The Royal Children's Hospital, Melbourne, Victoria, Australia. (Graham et al., 2016)).

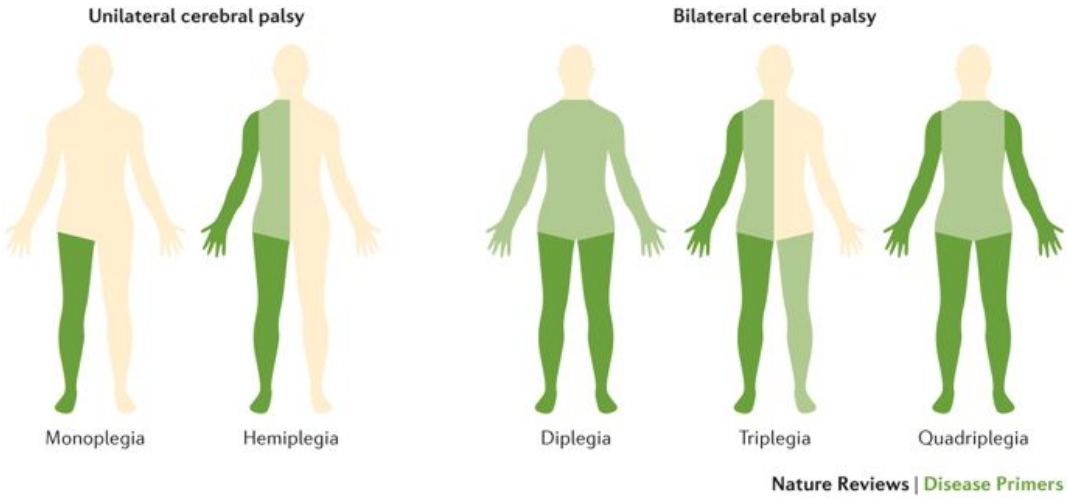


Figure 1.2: Topographical description in cerebral palsy

(Taken from Graham et al., 2016: In monoplegia, one limb is affected and it is more often the lower limb. In hemiplegia, one side of the body is affected and the upper limb is usually more affected than the lower limb. In diplegia, all limbs are affected, but the lower limbs are much more affected than the upper limbs, which frequently only show fine motor impairment. In triplegia, the usual pattern is unilateral upper limb involvement and bilateral (asymmetrical) lower limb involvement. The lower limb is invariably more affected on the same side as the upper limb involvement. In quadriplegia, all four limbs and the trunk are involved. Diplegia, triplegia and quadriplegia are covered by the term bilateral cerebral palsy (Graham et al., 2016))

Cerebral palsy prevalence and aetiology

A meta-analysis recently reported the worldwide prevalence rates of CP as 2.11 per 1000 live births (Oskoui, Coutinho, Dykeman, Jette, & Pringsheim, 2013). This makes cerebral palsy the most common motor disability in childhood. It was once considered that all children with cerebral palsy were either born premature or had a difficult labour associated with neonatal asphyxia. These were considered to be directly linked to the cause of CP however has now been considered to be risk factors in the development of CP (T. R. Han, Bang, Lim, Yoon, & Kim, 2002; Murphy DJ, 1997; Sukhov, Wu, Xing, Smith, & Gilbert, 2012). Various other risk factors include intrauterine infection,

changes in blood pressure, and conditions associated with increased clotting (O'Callaghan et al., 2011). It is known that gestational age is an important risk factor for CP with the risk steadily increasing with declining gestational age. Along with an early gestational age at birth, CP is 50% more likely to develop if associated with white matter injury or other brain injuries (Pinto-Martin JA, 1995). Perinatal factors such as chorioamnionitis (intra-amniotic infection) or other evidence of perinatal inflammation; transient hypothyroxinaemia (low maternal thyroid hormone levels) have been associated with the development of CP in premature infants. However whether these factors act via brain damage or if a direct link exists, is unclear. In full term infants that develop CP, factors such as low Apgar score have also been shown to associate with increased risk of developing CP (Nelson KB, 1984). It is also unclear if the same factors have an equal risk in different subtypes of CP, given CP is such a heterogeneous condition.

Similar to other neurodevelopmental disorders, CP is considered to have both a genetic and environmental origin although the extent of the influence of each is not clearly understood. Twins in general have a shorter mean gestational age and lower birth weight than singletons and have higher chances of preterm delivery. There have been a number of cases where the incidence of CP has a 5.1 fold increase in twins compared to that of singletons (Tollanes, Wilcox, Lie, & Moster, 2014), indicating that being a twin itself can increase the risk of CP (Bonellie, Currie, & Chalmers, 2005). Twin specific risk factors may include unequal sharing of blood in monochorionic twins leading to twin-to-twin transfusion syndrome (McNamara, Kane, Craig, Short, & Umstad, 2016), or the unequal sharing of placenta (cord insertion, placental vasculature, etc) (Lewi L, 2010). Population based studies that investigate the patterns of heritability of cerebral palsy have shown that families with a history of CP have a higher risk of the disorder in future generations depending on the degree of relatedness (Tollanes et al., 2014). This shows that cerebral palsy has a genetic component.

Various studies have indicated that single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) in some genes to be linked to cerebral palsy (K. B. Nelson et al., 2005; Oskoui et al., 2015). Until recently however, there have been no genetic or epigenetic study demonstrating the use of twin models to study the underlying biology processes that may be involved in cerebral palsy pathophysiology.

1.1.1.2 Epilepsy

Epilepsy is a complex brain disorder of heterogeneous aetiology, presentation and tractability. It is characterised by the occurrence of seizures. Around 30% of cases experience symptoms that remain hard to control with anti-epileptic drugs. The burden of epilepsy is associated with increasing morbidity and mortality. Daily living conditions such as ability to drive, going to work and work-related tasks need to be adjusted significantly in order to ensure safety of patients suffering from epilepsy. Apart from clinical manifestations of the disease and social repercussions, there is a significant psychological impact on the people suffering with epilepsy and their families.

Epilepsy definition and diagnosis

Epilepsy are a group of brain disorders marked by sudden, recurrent episodes of motor or non-motor disturbance, loss of consciousness or convulsions, associated with excessive electrical activity in the brain. Around 35% of epilepsies have a clear acquired cause such as head injury or stroke and the remaining 65% have unknown causes which were historically termed “idiopathic” because of presumed genetic or unknown origin. Idiopathic epilepsy is defined the absence of an underlying structural brain abnormality and can be classified into generalised and focal epilepsies. Focal seizures originate from one region of the brain, and generalised seizures originate bilaterally, and are a result of abnormal activity involving most of the brain.

The clinical classification of idiopathic epilepsies into focal and generalised is based on clinical assessment with electroencephalogram (EEG) and neuroimaging techniques such as MRI, CT or PET scans of the brain. **Figure 1.3** shows the EEGs from generalised epilepsy (sharp waves seen over the entire brain) and **Figure 1.4** shows EEGs from focal epilepsy (sharp waves seen over one region of the brain). Understanding if an epilepsy is focal or generalised will assist in tailoring treatment choices, as some medications can aggravate certain epilepsy types. Anti-epileptic medications can treat some of these cases; however, depending on the type of epilepsy, the choice of anti-epileptic medication can aggravate certain epilepsy types.

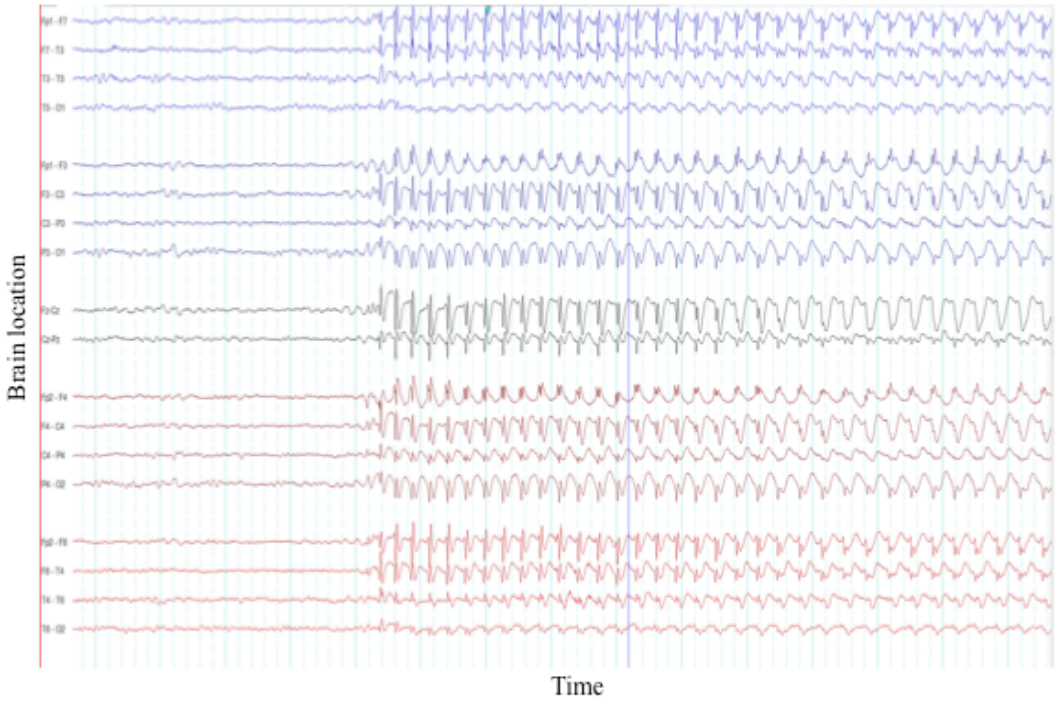


Figure 1.3: EEG showing generalised seizure

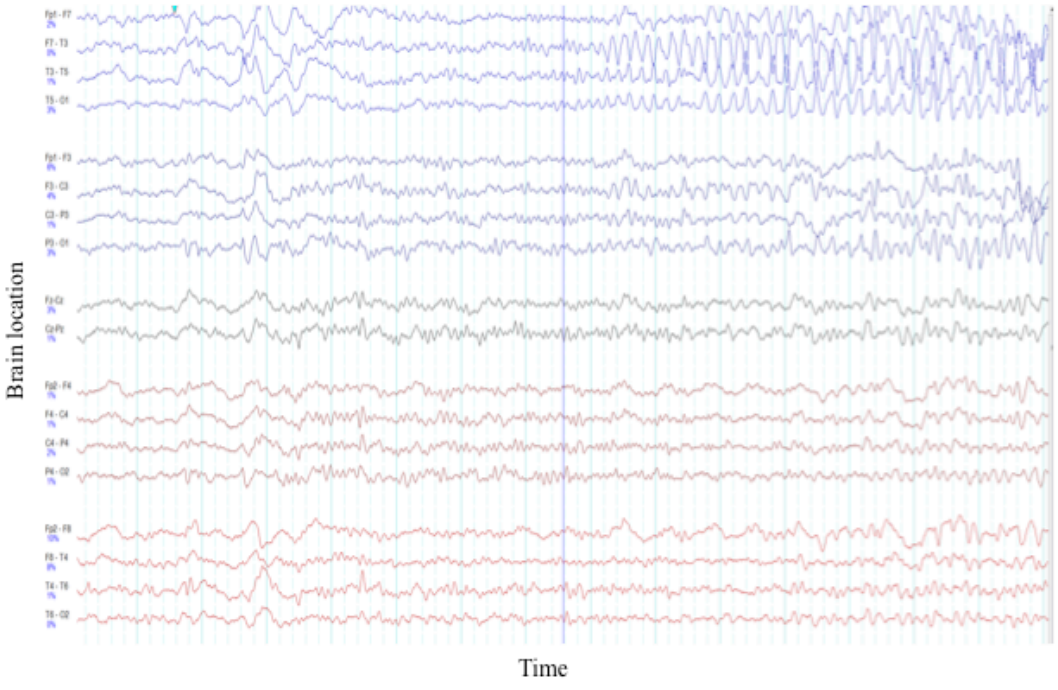


Figure 1.4: EEG showing focal seizure

Epilepsy prevalence and aetiology

Globally, over 50 million people are affected by epilepsy. The World Health Organisation's 2010 Global Burden of Disease study ranks epilepsy as the second most onerous neurologic disorder worldwide in terms of disability-adjusted life years (Murray et al., 2012). The prevalence rates for epilepsy in Australia were estimated to be 1 in 26 (https://www.epilepsy.org.au/about-epilepsy/facts-and-statistics/). The aetiology of epilepsy includes a wide variety of phenomena, including traumatic brain or head injury, developmental brain malformations, and genetic causes. The role of genetics in epilepsy has long been described with various studies characterising the patterns of inheritance of epilepsy, and the molecular explanations for these patterns (Berkovic, Howell, Hay, & Hopper, 1998; MI, 2010; Speed D, 2014; Vadlamudi et al., 2014).

Molecular investigations have uncovered many genes underlying different rare monogenic types of epilepsy, with the first epilepsy gene, cholinergic receptor nicotinic alpha 4 subunit (*CHRNA4*), being identified in 1995, as a cause of autosomal dominant frontal lobe epilepsy (Steinlein OK, 1995). After 2008, high throughput sequencing and the application of next generation sequencing revealed an overabundance of genes associated with epilepsy, clearly defining a role of genetics in the disorder (Helbig, Scheffer, Mulley, & Berkovic, 2008; Vadlamudi et al., 2014). This is particularly true of the relatively rare group of epileptic encephalopathies, which are now emerging as a genetically heterogeneous group of largely de novo dominant disorders. In contrast, monogenic variants account for 1% of common forms of epilepsy. 65% of cases of epilepsy remain clinically idiopathic, leaving a great number of patients without an explanation for their condition (Hauser, Annegers, & Kurland, 1993). These common forms are likely multifactorial, with a significant and complex genetic architecture. The marked phenotypic and genotypic heterogeneity in the idiopathic epilepsies imply other causative mechanisms must be involved.

1.1.2 Genetics of neurodevelopmental disorders

Studies of neurodevelopmental disorders have advanced considerably from detecting chromosomal abnormalities such as aneuploidy and microdeletions to single gene defects and those with complex aetiology (Shen J, 2014). The advances in genetic technologies over the years demonstrate not just the range of genetic abnormalities associating to the phenotype but also the complexity and variability of neurodevelopmental disorders. Teasing apart the complexity of genetic patterns of inheritance of neurodevelopmental disorders largely involves twin and family studies that established whether the disorder is of genetic origin. Such studies provided considerable success in evaluating risk factors at disease specific loci and genes, though they did not explain the phenotypic variation associated with the disorder.

1.1.2.1 Genetic background of cerebral palsy

With the advent of next generation sequencing technologies, a large number of causative genetic variations have been identified in individuals with neurodevelopmental disorders. However, it is increasingly being understood that these neurodevelopmental disorders likely result due to complex genetic and genetic-environmental interaction. Population-based studies that investigate the patterns of genetic aetiology of cerebral palsy have shown that families with a history of CP have a higher risk of the disorder in future generations depending on the degree of relatedness (Tollanes et al., 2014). Various studies have indicated that SNPs and CNVs are associated to cerebral palsy. For example, SNPs associated with thrombophilic genes, apolipoprotein E (APOE) gene, cytokines, etc. have been indicated to play a role in cerebral palsy (O'Callaghan, MacLennan, Haan, Dekker, & South Australian Cerebral Palsy Research, 2009). This large systematic review was performed on 22 papers that examined SNP association with CP outcome. Association of CP to thromobophilic SNPs were not validated among studies due to limited size or heterogeneity in study design such as multiple ethnicities, gestational age or infection. Three studies that investigated APOE SNPs and CP association found no relationship and warranted further clarification. SNPs in cytokines such as tumor necrosis factor (TNFA) and lymphotoxin alpha (LTA) gene were significantly associated to CP outcomes in five studies. However, these studies need replication in larger datasets and further analysis

corresponding to CP subtypes. Several studies report that CNVs may explain CP in about 10% to 20% of cases (McMichael et al., 2015; Oskoui et al., 2015; Segel et al., 2015). McMichael et al. identified putatively deleterious CNVs affecting 20% of a CP cohort. Candidate genes implicated included genes involved in microcephaly such as microcephalin MCPH1 and some implicated in spastic paraplegia like spastic paraplegia-6 gene, SPG6/NIPA1. Segel et al. identified putatively pathogenic CNVs in 19% of their cohort, and identified CNVs that they considered likely to be pathogenic in another 12%. De novo pathogenic CNVs were also found in some patients. The investigators implicated gene alterations in KANK1 (also reported by (Lerer I, 2005)). Oskoui et al. found de novo deleterious CNVs in 7% of their patients and, in total, about 10% carried CNVs thought to account for their symptoms. They also identified copy number abnormalities affecting KANK1. Recent genetic studies of CP cases using exome sequencing show that 14% of cases have likely causative single-gene mutations (McMichael et al., 2015). Exome sequencing from a large cohort of ataxic CP cases had reported de novo point mutations in the genes potassium voltage-gated channel subfamily C member 3 (KCNC3), inositol 1,4,5-trisphosphate receptor type 1 (ITPR1), and spectrin beta non-erythrocytic 2 (SPTBN2) (Badawi & Keogh, 2013). This work also suggested that de novo mutations may significantly contribute to CP pathogenesis, as has been shown for related neurodevelopmental disorders, including intellectual disability, epilepsy, and autism. However, many of these studies have not had follow-up functional studies that are essential to confirm CP pathogenicity. Another complexity is the fact that CP is highly heterogeneous. Therefore, it is likely that CP cases are explained by more complex genetics and environmental factors and not just a single major gene effect (O'Callaghan et al., 2009).

1.1.2.2 Genetic background of epilepsy

The genetic component of epilepsies has been defined by familial and twin studies prior to the genomic era, and further strengthened by the discovery of genes mutated in epilepsy. Using twin studies, the components of variance in epileptic phenotypes, like all other human disorders, can be broken down into additive genetics (similar to narrow sense heritability), common environment and non-shared environment (individual environment plus developmental factors). One of the first twin studies in epilepsy was

by Sillanpää in 1998 (Sillanpaa, Jalava, Kaleva, & Shinnar, 1998), with a cohort consisting of over 27,000 twin pairs in total from the Finnish Twin Cohort Study and including 316 cases of epileptic seizures occurring in 310 twin pairs, both MZ and DZ. The results suggested that there was a higher concordance rate in (observed to expected ratio = 5.48) in MZ twins as opposed to DZ pairs (observed to expected ratio = 2.12). This suggested that most of the epileptic seizures were associated with genetic variability. Ganor et al., 2005 (Ganor, Freilinger, Dulac, & Levite, 2005) studied the role of cytokines and autoimmunity in epilepsy. Based on previous studies, they specifically looked at antibodies against glutamate receptor, and four key cytokines namely IFN-gamma, TNF-alpha, IL-4 and IL-10. They compared the serum levels of the antibodies against these molecules in 4 discordant MZ twins and 49 non-epileptic controls. They found significantly higher levels of specific antibodies in the serum of affected twins compared to that of the healthy controls. Interestingly, within twin pairs they found high levels of antibodies to the receptor but low levels of cytokines compared to the unaffected twin. Genome-wide expression study was also investigated in five discordant and four concordant MZ twins with idiopathic absence epilepsy and healthy controls (Helbig et al., 2008). Using microarrays, they identified genes that were differentially expressed and 16 of those genes were validated using qRT-PCR. Nine of these genes were also replicated in an independent cohort. Two genes early growth response protein 1 (EGR1) and reticulocalbin 2 (RCN2) were significantly up regulated and have been implicated in seizures.

In the last few decades, advances have been made in identifying genetic variants that associate with epilepsy (Helbig et al., 2008). Variants in SCN1A and SCN1B, genes encoding sodium channels, have been shown to be causative of generalised epilepsy with febrile seizures plus (GEFS+), a familial syndrome of autosomal dominant inheritance that presents as variable epilepsy phenotypes across different individuals of the same family (Wallace et al., 2001). These discoveries are of great importance to neurology, reflected by the current-day practice of screening eligible families for SCN1A mutations (Hirose et al., 2013). Mutations of other genes such as protocadherin 19 (*PCDH19*), glutamate receptor subunit epsilon-1 (*GRIN2A*), DEP domain-containing 5 (*DEPDC5*) and potassium channel subfamily T, member 1 (*KCNT1*) are also such examples. Non-coding intronic region variations and copy number variations (CNV)

have also evolved as important genetic mechanisms in epilepsy (Leu, Coppola, & Sisodiya, 2016; Liu et al., 2016). Genome-wide association studies associate multiple common variants with epilepsy phenotype. A recent study (ILAE, 2018) reported multiple genome-wide loci as biologically plausible candidates in genetic generalised epilepsy and as targets of anti-epileptic medication. However, genetic variants associated with epilepsy that have been discovered so far account for only a minor portion of the aetiology all epilepsies. The complicated genotype-phenotype correlations in epilepsy indicate that other modifying factors are responsible for determining the specific subtype. Moreover, twin studies have shown differing case wise concordance estimates (proportion of epilepsy-concordant pairs as a proportion of concordant and discordant pairs combined) for idiopathic generalised epilepsies (MZ= 0.77; DZ= 0.35) and focal epilepsies (MZ= 0.40; DZ= 0.03) (Vadlamudi et al., 2014). The lack of complete case wise concordance estimates in MZ twin pairs who are genetically identical indicates that presence of other factors such as the non-shared environment that may contribute to discordance (Kjeldsen, Kyvik, Christensen, & Friis, 2001).

1.1.3 Epigenetics

In 1942, Conrad Waddington coined the term ‘epigenetics’ which literally means the layer (of regulation) above genes (Waddington, 1956). Waddington also hypothesised that the epigenetics responds to the environment. Initially, epigenetics was defined as the study of ‘changes in organisms caused by the modification of gene expression rather than an alteration of the genetic code itself’ (Holliday & Pugh, 1975; Riggs, 1975). Studies attempting to understand the various mechanisms involving the effect of environment in the development of an organism and its interactions among genes was becoming increasingly popular in the later years. Epigenetics has since changed its definition multiple times from ‘the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail change in DNA sequence’ (C. Wu & Morris, 2001) to the most recent being ‘the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states’ (Bird, 2007).

One of the best examples of the role of epigenetics in regulating gene expression was demonstrated by the experiments of Michael Meaney and Moshe Szyf showing that maternal behaviour of rats to their new born pups causes epigenetic alterations in the pup's brain, leading to phenotypic differences in their adulthood (Meaney MJ, 2005). They revealed that maternal behaviour could produce epigenetic changes in the brain that lead to the increased production of glucocorticoid receptor, NR3C1, in the pups brain, thereby leading to enhanced stress resistance as they develop into adults (Weaver et al., 2004). Other studies also demonstrated that a similar effect was seen in monkeys and humans (Champagne & Curley, 2009; Fairbanks, 1989).

1.1.3.1 Developmental origins of health and disease (DOHaD)

Abnormalities that originate from environment in early life relate to maternal factors, nutrition, among others and some of these aspects may affect embryonic development. Early statistical analysis carried out by David Barker in the 1980s laid the foundation of this theory (Barker et al., 1993; Barker & Osmond, 1986; Barker, Winter, Osmond, Margetts, & Simmonds, 1989). He compared the low-birth weight babies with their later health outcomes such as cardiovascular diseases and hypothesised that the low birth weight was due to its adaptive nature to its environment, which may have the low levels of fetal nutrients. He hypothesised that the adaptive response in the fetus would channel essential nutrients to the development of the key organs such as the brain, and neglect the development of other organs causing in utero growth restrictions and leading to higher chances of chronic illnesses like diabetes in the future (Barker et al., 1993; Burdge GC, 2010); (Hales & Barker, 2013). Neel, called the 'thrifty genotype' hypothesis, also records this adaptation in 1962. It explains that in a food scarce environment, the fetus would adapt to the low food environment by selecting 'thrifty genes', so that the child can survive in a food scarce situation in future. However, food may be in abundance at the later life stages, and the child's body may not be able to adapt to food abundance environment, thus leading to diabetes and related health complications (Neel, 1962). This adaptive mechanism was termed the "Barker Hypothesis" was later named "Fetal Origins Hypothesis" and is now known as the "Developmental origins of health and disease (DOHaD)" hypothesis. Research is now being focussed on the mechanisms during gestation and the role that it plays in

development. Epigenetics is considered to be one of the main factors that may explain the DOHaD concept.

1.1.3.2 DNA methylation, histone modification and non-coding RNA

Understanding epigenetics, the "switch" that turns genes on and off, will help to identify how the environment brings changes to the genes that expose itself to disease risk. There are many epigenetic modifications described in the literature, the best understood being addition of the methyl molecule (CH_3) to the cytosine-guanine (CpG) dinucleotide of DNA (Gibney ER, 2010; Vaissière T, 2008) by DNA methyltransferase (DNMT) enzymes. The methylated cytosine residues lie immediately adjacent to a guanine nucleotide linked by a phosphodiester bond denoted by a "p", there by placing two methylated cytosine molecules diagonal to each other on opposing DNA strands. There are different types of methyltransferase enzymes depending on their functions. *De novo* DNMTs (DNMT3A and DNMT3B) are responsible to initiate the pattern of methylation on a DNA sequence. These enzymes add new methyl groups to the cytosine residues. *Maintenance* DNMT (DNMT1), on the other hand, adds methylation to the DNA sequence where one strand is already methylated. These enzymes maintain the modification on the DNA sequence established by *de novo* methyltransferases.

DNA methylation has different roles and targets in different organisms. For example, in plants more than 50% of the cytosine residues are methylated whereas in fungi only the repetitive DNA sequences are methylated (Phillips, 2008). In some species, DNA methylation is completely absent or present only in the transposons in the genome (Phillips, 2008). DNA methylation has been observed to have global distribution patterns as well as 'mosaic' pattern of methylation, where regions of heavily methylated DNA are interspersed with non-methylated regions (Simmen et al., 1999; Weber & Schubeler, 2007). It is estimated that 70%-80% of CpG dinucleotides in the human genome are methylated, with most of the unmethylated CpG dinucleotides located in CpG "islands" (Davuluri, Grosse, & Zhang, 2001; Ohlsson & Kanduri, 2002; Simmons, 2011). CpG islands are variously defined as >200bp stretches of DNA with significantly higher concentration of CpG dinucleotides, than the genome average (Gardiner-Garden & Frommer, 1987) and tend to be in the promoter region of a gene indicating that they play a regulatory role in gene expression. At their promoter, genes are silenced by

methylation via blocking of transcription factors (Inamdar, Ehrlich, & Ehrlich, 1991; Xinsheng Nan, 1997) and also recruiting factors that encourage a local inactive chromatin configuration (S. Eden, Hashimshony, Keshet, Cedar, & Thorne, 1998; Mehler, 2008). Methylation states that differ between cell types, tissue types and individuals can account for variations in gene expression that may play a role in the expression of certain phenotypes.

DNA methylation has been shown to be essential for mammalian embryonic development as shown in an experiment in mice, where lack of DNMT enzymes led to decreased DNA methylation and ultimately death in mice (Li, Bestor, & Jaenisch, 1992; Lyko, 2018). The importance of DNA methylation was tested in a study that looked at the regulation of the *agouti* gene in mice (Dolinoy, 2008). They noticed that the set of mice that were fed a diet rich in vitamin B and folic acid (acting as methyl donors), resulted in *agouti* mothers to produce healthy, normal weight pups, not prone to diabetes, where as its counterpart that were not fed a specific diet, resulted in *agouti*, diabetic prone offspring (Morgan, Sutherland, Martin, & Whitelaw, 1999). This study demonstrated how the environment can have an affect on gene expression by altering DNA methylation. Several studies in model organisms have since recorded similar effects of the environment on the expression of genes and thereby on phenotype (Giacomo Cavalli, 1998; Menon & Meller, 2014; Seth et al., 2013). DNA methylation also plays an important role in a number of cellular processes such as genomic imprinting, X-chromosome inactivation and chromosome stability (A. Eden, Gaudet, Waghmare, & Jaenisch, 2003; Li, Beard, & Jaenisch, 1993; T. Mohandas, Sparkes, & Shapiro, 1981).

Histone modification is another epigenetic mechanism that affects gene expression by modifying the chromatin structure and function. Histones are proteins found in the eukaryotic cell that help in packaging and ordering DNA into structural units called nucleosomes. They undergo post-translational modifications such as acetylation, methylation, phosphorylation, ubiquitylation and sumoylation. Histone acetylation is a well-studied process and is involved in the ‘opening’ of the chromatin structure thereby helping the transcriptional machinery to access the DNA (Bannister & Kouzarides, 2011). Numerous studies have established a link between DNA methylation and histone

acetylation (Bartke et al., 2010; Tamaru H, 2003). For example, the hypermethylation of CpG islands in gene promoters triggers deacetylation of histones (Javaid N, 2017). It has also been observed that low levels of histone acetylation may induce DNA methylation thereby silencing gene expression and offering many routes to alter gene expression. The histone modification machinery are regulated by three major classes of enzymes - histone acetyltransferases (HATs) and deacetylases (HDACs), which introduce and remove acetyl groups, respectively; histone methyltransferases (HMTs) introduce methyl groups and histone demethylases (HDMs), on the other hand, remove methyl groups. The disruption of the activity of these enzymes has been reported to hamper recruitment of transcription factors, transcriptional machinery, and therefore gene expression (Gallinari, Di Marco, Jones, Pallaoro, & Steinkuhler, 2007; Wada et al., 2009).

Non-coding RNA (ncRNA) molecules such as miRNA, siRNA, piRNA and lncRNA are also known to demonstrate epigenetic regulation. MicroRNAs (miRNA) are generally seen to bind to messenger RNA targets to induce cleavage, degradation or block translation (Chuang & Jones, 2007; Yao, Chen, & Zhou, 2019). Short interfering RNA (siRNA) also has similar mechanism of action where targeted mRNA is degraded due to gene silencing (Verdel, Vavasseur, Le Gorrec, & Touat-Todeschini, 2009). They also have been shown to induce chromatin condensation and histone methylation (Joh, Palmieri, Hill, & Motamedi, 2014). Piwi-interacting RNAs (piRNA) interacts with the piwi family of proteins specifically, and suppresses transposon regulation in germline and somatic cells (Niemitz, 2012; Yamanaka et al., 2014). Long non-coding RNAs (lncRNA), which form the majority of ncRNAs, can form complexes with chromatin-modifying proteins and affect their catalytic activity, therefore influencing chromatin states and affecting gene expression (P. Han & Chang, 2015; Zhang, Vielle, Espinosa, & Zhao, 2019).

Epigenetic modifications in disease phenotypes can be due to genetic, developmental or environmental factors but the exact mechanism is unknown in many cases. Importantly, many studies have shown that prenatal environment can influence epigenetic state, which is in turn related to risk for chronic disease (Gluckman, Hanson, & Low, 2011;

Gluckman, Hanson, & Mitchell, 2010; Joubert et al., 2016). Furthermore, such epigenetic states have been shown to be maintained for many years after an initial environmental event, examples of which are famine, maternal nutrient intake and their links with obesity and heart disease (Godfrey, Costello, & Lillycrop, 2015; Heijmans, Tobi, Lumey, & Slagboom, 2014). Studies in animals have shown that disease-associated epigenetic state can be reversible after birth (Weaver et al., 2005). This information has huge implications for addressing human disease, namely that epigenetic assays could be used to supplement other biomarkers to assist with (i) diagnosis of chronic diseases; (ii) assess risk in very early childhood before onset of overt symptoms and (iii) designing therapeutic interventions.

1.1.3.3 Epigenetics and neurological diseases

Epigenetic modification has been implicated as both a mediator and potential biomarker for neurodevelopmental diseases (Petronis, 2010; Qureshi & Mehler, 2010; Urdinguio, Sanchez-Mut, & Esteller, 2009), though these have been less well studied than other chronic diseases (Loke, Hannan, & Craig, 2015; Mikeska & Craig, 2014). DNA methylation has been shown to be important for all aspects of brain development, homeostasis, plasticity, and response to injury (Mehler, 2008). Prenatal damage to the growth and development of the brain can result in serious neurological disorders, as discussed in the previous *section 1.1.3.1*. Understanding the underlying disease mechanisms and the development of potential biomarkers are key to better prediction and early diagnosis of such chronic conditions. Genetics play an important role in understanding the inheritance patterns of such disorders (discussed above in *section 1.1.2*); however, there is increasing understanding that these disorders, like other chronic conditions, result from a combination of genes, environment and developmental variation, the latter two being most prominent in early life.

Twins play an important role in differentiating phenotypic variance into genetic, shared and non-shared (or developmental) environments. Although shared factors are an important contributor to chronic disease risk, non-shared factors are of a greater magnitude and are likely to be more influential for health (Boomsma, Busjahn, & Peltonen, 2002; Petronis, 2010).

1.1.4 Twins

The study of twins offers a new opportunity to study epigenetics since the genetic variability between identical or monozygotic twins is known to be very low or even absent (van Dongen, Slagboom, Draisma, Martin, & Boomsma, 2012; A. H. Wong, Gottesman, & Petronis, 2005). Monozygotic (MZ) twins develop from one single zygote that splits into two during the first few days of life. Dizygotic (DZ) twins develop from two different eggs fertilized by two separate sperm cells. The mechanism of MZ twining are not fully understood; however, it has been hypothesised that the time at which MZ embryos separate, determines the development of each embryo and affects growth and development (McNamara et al., 2016). According to this hypothesis, embryos that separate in the first four days of pregnancy develop with separate amniotic sacs and chorionic casings called dichorionic twins. They are like singletons in the womb as they have separate placentas and umbilical cords and have the highest survival rate among twins. They account for 25% of MZ twins. On the other hand, eggs separating between 5-8 days share a single placenta and chorion but have two amniotic sacs and umbilical cords called monochorionic twins, which make up 75% of MZ twins. In this case, one twin can have an advantage over the other. The umbilical cord may be centrally located on one sac and marginally on the other, thereby limiting the nutrients for one twin compared to the other. Although this happens in-utero, the lighter twin can catch up in early life with a nurturing environment. Sharing of the placenta can also cause vascular anastomoses, in which shared blood supply can result in twin-twin transfusion syndrome leading to growth risks and even mortality. According to the above hypothesis, twins may also result from a split between 8-12 days of conception and share an amniotic sac, therefore called monoamniotic twins. The chance of survival of both twins in this instance is low as it is very likely that the cords can get entangled (Dickinson, 2005) leading to restricted blood flow and complications in both twins. Conjoint twins are formed as a result of the egg splitting after 12 days.

Twin pregnancies are often accompanied with complications to the baby as well as the mother including pre-term delivery, restricted fetal growth, increased nausea, gestational hypertension and pre-eclampsia. Placental sharing can be one of the factors that determine the risk in twin pregnancies (B, 2018; Hubinont et al., 2015; Lewi L,

2010). Compared to singletons, twin pregnancies are also at a higher risk to fetal anomalies and chromosomal abnormalities (Hubinont et al., 2015; Khong JJ, 2006; Lopriore E, 2006).

Although twin pregnancies are associated with higher risk of health complications in general, they can be valuable in understanding in importance of genetic and environmental influences on complex trait variation. The classical twin model was a popular study design, where phenotypic differences between MZ and DZ twins were compared to distinguish the proportion of variance in a disease due to genetic variation within a population (Boomsma et al., 2002). Such studies had important implications for the medical community as neurological disorders such as epilepsy and autism were found to have a heritable component, overturning previous assumptions that it originated primarily in perinatal factors (Bailey A, 1995; Samuel F Berkovic, 1993). More sophisticated twin models like the ACE model involved estimating proportion of variance due to additive genetic effects (A), common or shared factors (C) and non-shared or unique factors (E) for a given phenotype. In chronic diseases, the largest variance is expected from the non-shared factors or twin-specific environmental factors because the A (genetics) stays equal in MZ twin pairs and the C (shared factors) is relatively small in difference within twin pairs. This non-shared environment can also include the normal developmental noise that would be different between twins. Non-shared factors refer largely to the fetoplacental unit where there may be differences in placental size, morphology, location of cord insertion, blood flow, infection etc. between the twin pairs.

The comparison of discordant monozygotic twins, called the co-twin control model, is a powerful tool that can aid in the detection of biomarkers for various disorders, as it controls for shared genetic and shared environmental factors, enabling focus on non-shared factors as mechanisms to explain the disease discordance. Twin registries have helped greatly in these studies by establishing collections of longitudinal and cross sectional data in twins across age categories from birth (Hur & Craig, 2013). Twin studies also continue to recruit cohorts specific to a disease phenotype, enabling very large twin studies to be carried out worldwide (Sahu & Prasuna, 2016). The advantage of using monozygotic twins is that they are perfectly matched for age, genetic background and sex. The differences between them can be analysed for association to unique biological pathways that may be responsible for the discordance. It also allows

for the exploration of environmental influences on twins during early life. Diverse epigenetic profiles and gene expression has been reported in previous studies indicating the role of stochastic and environmental factors in utero and in early life (van Dongen et al., 2012).

Monozygotic twins originate from one zygote, but some studies have shown that their somatic cells are not always identical at the sequence level. Mutations may occur at various stages of development (Vadlamudi L, 2010; Vadlamudi et al., 2014). Structural variations such as copy number variations have been seen to vary with age and also with cell type. Studies comparing structural components in buccal and blood samples of twins have observed different results, sometimes a pre-twinning duplication in a twin pair or a post-twinning deletion in concordant twins or no difference at all depending on the disease phenotype (Ehli et al., 2012; Forsberg et al., 2012; Veenma et al., 2012). The technique used to measure these differences is also important. Most studies involve microarrays with the limitation of not being able to achieve total coverage across the whole genome. The application of whole genome sequencing techniques may give rise to more accurate and extensive sequence differences with each twin pair (Baranzini et al., 2010). Understanding the time of occurrence of de novo mutations is very important as it informs us of the risk of recurrence in multiple births. It can also be crucial to understand the biological mechanisms underlying the disease due to the mutagenesis. Studying twins can be a powerful tool that enables us to identify DNA sequence variants, epigenetic discordance for disease as well as gene expression and other aspects. The value of twin studies can be a constant platform of debate as they are naturally vulnerable to risk, yet they provide power by uniform genetics thereby helping us look at additional factors and better pinpoint the disease mechanisms and pathways. Twin models have been especially used to study major neurodevelopmental and neurocognitive conditions to understand the epigenetic landscape of disease discordance. This is a valuable model, especially in paediatric neurodevelopmental disorders, such as cerebral palsy and epilepsy, where the disorder is heterogeneous and disease mechanisms involved are complex.

1.1.5 Epigenetic studies in cerebral palsy

There is growing evidence that risk for CP is mediated by epigenetic mechanisms (Fleiss & Gressens, 2012). Dysregulation of methylation capacity and folate single-carbon metabolism has been reported in children affected with severe CP (Schoendorfer et al., 2012). Folate or single carbon metabolism provides the carbon substrate (methyl group) required for DNA methylation. Moreover, alterations in DNA methylation patterns are indicated in bacterial infection, inflammation, growth restriction and early life stress (Blaze & Roth, 2015; Houtepen et al., 2016). These are also considered risk factors to cerebral palsy. Several studies have looked at DNA methylation patterns in CP affected and unaffected individuals to potentially help create a diagnostic biomarker that allow for early interventions. One such study was (Crowgey, Marsh, Robinson, Yeager, & Akins, 2018) where they evaluated DNA methylation profiles using machine learning based classification model from 22 spastic CP affected individuals against 21 controls, and found significantly differentially methylated CpG sites across the groups. They were also able to validate their findings in a similar cohort with 73% accuracy. Another recent study with a similar approach of using artificial intelligence, identified 230 significantly differentially methylated CpG loci in 258 genes had a 95% sensitivity and 94.4% specificity for newborn prediction of CP (Bahado-Singh et al., 2019). Although both studies found significant association of differentially methylated CpGs to CP relevant biological processes, the limitations of both studies were its limited sample size. Studies of phenotypically discordant monozygotic twins have allowed for inter-twin comparisons allowing the genetic and environmental components of complex human disorders to be better identified.

1.1.6 Epigenetic studies in epilepsy

The role of genetics in epilepsy has long been described, and has been applied to the heritable component of epilepsy. Various studies have confirmed and characterised patterns of inheritance of epilepsy, and the molecular explanations for these patterns (Berkovic et al., 1998; MI, 2010; Speed D, 2014; Vadlamudi et al., 2014).

There is growing evidence that epigenetics plays a major role in brain maturation and brain function as well as epileptogenesis (Kobow & Blumcke, 2014). For example, mutation in the methyl-CpG-binding protein 2 (*MECP2*) gene located on the Xq28 chromosome was found to regulate neuronal function in Rett syndrome. This early discovery suggested that epigenetics may have a broader role in neurodevelopmental disorders. The first evidence for a role for DNA methylation in the pathogenesis in seizures was produced by in vitro studies, which demonstrated a decrease in spontaneous excitatory neurotransmission following inhibition of DNMT enzymes in hippocampal slices (Levenson et al., 2006), and then in hippocampal primary neurons (E. D. Nelson, Kavalali, & Monteggia, 2008). The study by Levenson et al. (Levenson et al., 2006) provided the first evidence that DNMTs play a role in modulating adult central nervous system function. Firstly, rapid changes in the DNA methylation were observed in two genes implicated in synaptic plasticity, BDNF and Reelin. Moreover, DNMT inhibition blocked induction of long-term potentiation, suggesting that DNMT activity is required for synaptic plasticity. These findings were validated by a similar study conducted by Nelson et al. (E. D. Nelson et al., 2008). They demonstrated that DNA methylation changes in BDNF, as well as a reduction in the frequency of excitatory postsynaptic currents and synaptic vesicle fusion. Together, these findings provided preliminary evidence that spontaneous synaptic transmission may be regulated by DNA methylation. Altered DNA methylation has been detected in a number of genes previously associated with neuronal hyperactivity and seizures. These include *GRIA2* (Machnes et al., 2013), *GRIN2b* (Ryley Parrish et al., 2013), and *BDNF* (Martinowich et al., 2003). Martinowich et al. were the first to describe dynamic methylation changes in the BDNF promoter, suggesting that methylation-mediated remodelling may play a critical role in long-lasting neuronal adaptive responses (Martinowich et al., 2003). Candidate gene studies have been used to investigate diverse epigenetic alterations in epilepsy, several contributing significant research to the field. A key example is a 2009 study by Kobow et al. (Kobow et al., 2009), which investigated the role of Reelin promoter methylation in association with granule cell dispersion in temporal lobe epilepsy. Reelin is an extracellular matrix protein that plays a key role in cell position and neuronal migration during brain development, and synaptic function and plasticity in adult life (Pulido Fontes, Quesada Jimenez, & Mendioroz Iriarte, 2015; Qureshi & Mehler, 2010). Reelin promoter methylation was analysed from three hippocampal subregions from eight temporal lobe epilepsy and three matched controls, and was

found to be greater in all patients with temporal lobe epilepsy compared with controls. These data suggest that increased promoter methylation of the Reelin gene may contribute to lower Reelin expression in temporal lobe epilepsy with granule cell dispersion, and that epigenetic regulation is likely to play a role in this seizure disorder. These findings contribute to the growing body of evidence to support a role for DNA methylation in human epilepsies. More recently, in temporal lobe epilepsy, a candidate gene approach was used to investigate CPA6 promoter methylation (Belhedi et al., 2014). Zhu et al. also investigated temporal lobe epilepsy, by examining DNMT1 and DNMT3A expression in 25 patients with temporal lobe epilepsy compared to ten healthy controls (Zhu et al., 2012). DNMT1 and DNMT3a expression was measured, and was significantly increased in patients with temporal lobe epilepsy. It is known that DNMT can regulate synaptic plasticity in the hippocampus and that DNMT inhibition causes deficits in excitatory synaptic transmission (Levenson et al., 2006; E. D. Nelson et al., 2008). Therefore, it was suggested that the increase in DNMT expression in this study may be related to epilepsy via an alteration of synaptic plasticity and neuronal network excitability (Levenson et al., 2006; E. D. Nelson et al., 2008; Zhu et al., 2012). This was the first study to report the change of DNA methyltransferase expression in the epileptic brain, and supports the principle that there is a association between chronic temporal lobe epilepsy and DNA methylation.

Although there have been numerous studies focusing on epigenetics in epilepsy, there is a lack of studies examining epigenome wide associations in this disorder. The only epigenome wide association study in human epilepsy to date was published recently by Long et al 2017. This group compared whole blood genomic DNA methylation patterns in mesial temporal lobe epilepsy patients, relative to healthy controls. The cases and control groups were found to have significantly different DNA methylation profiles at 216 sites. A subset of sites was cross validated, and all sites were compared to known functional genes. Whilst this recent study fills an important gap in the body of research in this field, it has numerous limitations, particularly in the analytical approach. It was evident from twin studies that the high level of discordance of epilepsy in MZ twins suggested the role of non-genetic factors in the aetiology of the disease. However there have been no epigenetic studies undertaken using the twin model thus far. This raises important questions on integrated genetic and epigenetic analyses using the power of twins in understanding the potential mechanism of epilepsy.

1.2 Study Aims

1.2.1 Research Questions

1. Are there epigenetic variations in neurological diseases such as cerebral palsy and epilepsy that may shed light on the biological pathways underlying the diseases?
2. Does the combined knowledge of epigenetic markers and gene expression help in the prediction of these neurological diseases at birth?

1.2.2 Research aims

1. To identify within-pair differences in DNA methylation within MZ twin pairs discordant for cerebral palsy
2. To explore gene expression profiles from the same CP discordant MZ twin pairs from the cerebral palsy cohort. Further, to compare the results of gene expression against CP-specific DNA methylation biomarkers generated from preliminary epigenome wide association analysis of MZ twins discordant for CP.
3. To identify significant within-pair differences in DNA methylation within MZ twin pairs discordant for idiopathic epilepsy
4. To identify epigenetic signatures that could explain the role of epigenetics in self-limited epilepsy with centrotemporal spikes (SECTS)

Chapter 2 Methods

This chapter details the methods used in this thesis. As a PhD by publication, relevant sections from this chapter have been used in publications and are therefore also included as part of chapters 3, 4 and 5.

Section 2.1 discusses the study cohorts used in the study, the process by which the cohorts were recruited, and the bio samples collected; section 2.2 details the genomic DNA extraction method from multiple tissues; section 2.3 describes the DNA methylation array platforms used in this study; section 2.4 outlines the steps involved in epigenome-wide association study (EWAS) detailing the statistical DNA methylation analysis workflow; finally, section 2.5 discusses how the data was validated on an independent array platform.

2.1 The UNIQUE twins cohort

The UNIQUE twin cohorts are made up of clinical cohorts of monozygotic twins discordant for cerebral palsy and epilepsy.

Table 2.1 Characteristics of UNIQUE cohorts used in this study

Study cohort characteristics	Cerebral palsy	Epilepsy
No. of twin pairs	15	15
Tissue types		
Buccal cells		✔
Blood sample		✔
Guthrie spots	✔	✔
Subtypes of disorder	5	2
Age group	8-20	14-67
Sex	Both	Both
Geographical Location	Victoria	Australia

2.1.1 Recruitment

2.1.1.1 Cerebral palsy

Participants were sought from the Victorian Cerebral Palsy Register (VCPR) at RCH (RCH; HREC 33050), which in June 2016 included a total of 91 children with CP known to be of twin birth. Potential participants were enrolled in two stages
(i) Initial contact from the VCPR sending out study cover letters inviting the families to contact the research team if interested. The register only stores information on the child with CP, therefore the status of the other child of each twin pair was unknown. Inclusion Criteria: Twin birth, both children alive, one child only of the pair diagnosed with cerebral palsy, aged between 3-17 years and with available newborn screening ("Guthrie") card.
(ii) Screening through telephone interview. If the twins were known to be dizygotic (via zygosity test or differing sex) or were probably dizygotic (different hair and eye colour) they were excluded from the project. If the twins were known to be monozygotic (via zygosity test) or are probably monozygotic (same sex, same hair and eye colour), they continued through recruitment. Medical conditions of both twins were noted, and any child that did not have CP but had another neurological condition was excluded.
Clinical characteristics of the twin pairs were extracted from the newborn screening database and from VPCR. Further details of the cohort are explained in Chapter 3.

2.1.1.2 Epilepsy

Ethics approval for this research was obtained from Mater Health Services (ethics approval number HREC/13/MHS/114). Potentially eligible participants were obtained from the Epilepsy Research Centre (ERC) database, referral from Queensland Neurologists, the Queensland Institute of Medical Research Berghofer (QIMR-B), Epilepsy Queensland and the Australian Twin Registry (HREC /13/MHS/114 AM09). The Epilepsy Research Centre based at the Melbourne Brain Institute, and associated with Austin Health, Victoria maintains a database of epilepsy research participants. Discordant monozygotic twin pairs found through this database were contacted and asked if they would like to participate. Once participants expressed willingness to be

contacted by the research group, participant information sheets (PIS) and participant consent forms (PCF) were distributed to obtain permission to proceed with the study. Monozygotic twin pairs, aged between 14-67 years, where the affected twin has epilepsy from a non-acquired cause and included the following diagnoses were recruited for this study:

- Genetic Generalised Epilepsy (GGE)
- Self-limited Epilepsy with Centro-Temporal Spikes (SECTS)
- Temporal Lobe Epilepsy with Hippocampal Sclerosis (TLE/HS)
- Non-lesional Temporal Lobe Epilepsy (NLTLE)
- Self-limited occipital epilepsy

Each consenting participant underwent a systematic review of medical history, including seizure history and characterisation, review of relevant investigations such as EEG and MRI, and medication history. A research assistant engaged in this study filled questionnaires for each patient through a telephonic or direct interview process. The affected questionnaire involved an extensive history of the seizure types and utilized a validated questionnaire (Reutens, Howell, Gebert, & Berkovic, 1992). The unaffected questionnaire was aimed at excluding unrecognised seizures in the unaffected co-twin.

Other points of focus for the questionnaires included birth, development and medications history (detailed in Chapter 4).

2.1.2 Bio samples

Guthrie cards or newborn blood spots: For the CP cohort, the Victorian Clinical Genetic Service (VCGS) was contacted for access to neonatal newborn screening card or guthrie card, collected 2 to 4 days after birth. For the epilepsy cohort, registries across all Australian states depending on where the patient was born were contacted. Where the guthrie cards have not yet been destroyed, punch-outs from these cards were requested. Whole 10mm diameter blood spots were obtained, from which six to ten 3mm samples were punched out for DNA extraction.

Buccal cells: Epithelial cells from the buccal surface of the cheek used for epigenetic analysis were collected during the research visits or through swab kits mailed to the families with written instructions. Samples collected were frozen until batch processing for DNA extraction using standard techniques and storage at -80°C.

Whole Blood: Venous blood was collected from twins by trained phlebotomists using venipuncture or finger pricks at the time of presentation for clinical assessments. For blood samples collected, buffy coat was frozen viably in liquid nitrogen and for future studies, plasma was stored at -80°C.

2.2 Genomic DNA extraction

2.2.1 Genomic DNA extraction from blood

Genomic DNA (gDNA) was extracted from the dried blood spot samples using the ZR DNA Card Extraction Kit (Zymo Research, Irvine, CA, USA) with some modifications to the manufacturer's protocol. Briefly, eight 3-mm punches were from each 1-cm-diameter blood spot and were transferred to a 2-mL Eppendorf Safe-Lock microcentrifuge tube (Merck, Darmstadt, Germany) containing ZR BashingBeads. Four hundred microliters of PBS containing 40 µL of 20 mg/mL proteinase K (Sigma-Aldrich, St. Louis, Missouri, USA) was added, and samples were vortexed and centrifuged briefly, followed by incubation overnight at 37 °C. Following incubation, 400 µL of ZR lysis solution was added to each tube. Punches were homogenised for 30 s at 4 m/s2 using Thermo Savant FastPrep 120 Cell Disrupter System (Global Medical Instrumentation (GMI) Incorporation, Minnesota, USA). Tubes were centrifuged for 1 min at 10,000 rpm, and 390 µL of 2×digestion buffer and 10 µL of 20 mg/mL proteinase K were added. Tubes were mixed by inversion and incubated for 30 min at 55 °C, then left to cool at ambient temperature for 3–4 min before centrifuging for 1 min at 8000 rpm. Six hundred fifty microliters of supernatant was added to 1.3 mL of DNA isolation buffer contained in a 5-mL Falcon tube (Thermo Fisher Scientific, MA, USA). This mixture was passed through the Zymo-Spin IC column by centrifuging for 1 min at 14,000 rpm, followed by the discard of flow-through liquid. The spin column was then washed twice by adding 200 µL of DNA

wash buffer and centrifuged for 1 min at 14,000 rpm. Finally, 20 μL of DNA elution buffer (pre-warmed at 55 °C) was added to the column and incubated at ambient temperature for 15 min before final centrifugation for 2 min at 14,000 rpm. This was repeated, resulting in a final elution volume of 40 μL containing genomic DNA. DNA concentration was measured by spectrophotometry (Nanodrop, Wilmington, DE, USA) to allow calculation of the required volume of each sample for array analysis. The quality of the extracted gDNA samples was visualised using agarose gel electrophoresis.

2.2.2 Genomic DNA extraction from buccal

Genomic DNA extraction from buccal sample was performed with the NucleoBond CB20 DNA Extraction Kit (Macherey-Nagel; Du¨ren, Germany). For blood samples, genomic DNA was extracted using the Flexigene kit (Qiagen; Hilden, Germany) following manufacturer's protocol. Following genomic DNA extraction from each of the samples, the concentration of extracted DNA was measured by spectrophotometry, and the quality of the extracted samples was determined by gel electrophoresis.

2.3 DNA methylation arrays

Chemical modifications to the DNA sequence and histones can be complex to comprehend and decipher. The growth of high throughput sequencing technologies has enabled the assessment of methylation marks across the whole genome in mammalian organisms. Sequence based assays developed to detect methylation patterns can be broadly divided based on affinity enrichment, chemical conversion or enzymatic restriction. In this study, the Illumina Infinium platforms were used which are described below:

2.3.1 Illumina Infinium Human Methylation450 BeadChip Array (HM450)

Illumina Infinium HumanMethylation450 BeadChip (HM450) arrays are a cost-effective alternative to whole genome bisulfite sequencing, and as such have been widely used to profile DNA methylation, particularly for studies with large numbers of samples. Illumina first initiated measurement of DNA methylation by Infinium technology on the HumanMethylation27 array (L. J. Bibikova M, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL., 2009), which measured methylation at approximately 27,000 CpGs, primarily in promoter regions. Like bisulfite sequencing, the Infinium assay detects methylation status at single base resolution. However, because of its limited coverage the array was not truly considered to represent the "true picture of the genome" until the HM450 array was established (Illumina, 2015). The HM450 array increased the genomic coverage of the platform to 485,000 sites representing genes by combining the original Infinium I assay with the novel Infinium II probes. The array covers 99% of RefSeq genes, with an average of 17 CpG sites per gene region distributed across the promoter, 5'UTR, first exon, gene body, and 3'UTR. It covers 96% of CpG islands, with additional coverage in island shores and the regions flanking them. The assay interrogates these chemically differentiated loci using two site-specific probes, one designed for the methylated locus (M bead type) and another for the unmethylated locus (U bead type) (**Figure 2.1**). Single-base extension of the probes incorporates a labelled ddNTP (dideoxynucleotides triphosphates), which is subsequently stained with a fluorescence reagent. The level of methylation for the interrogated locus can be determined by calculating the ratio of the fluorescent signals from the methylated versus unmethylated sites.

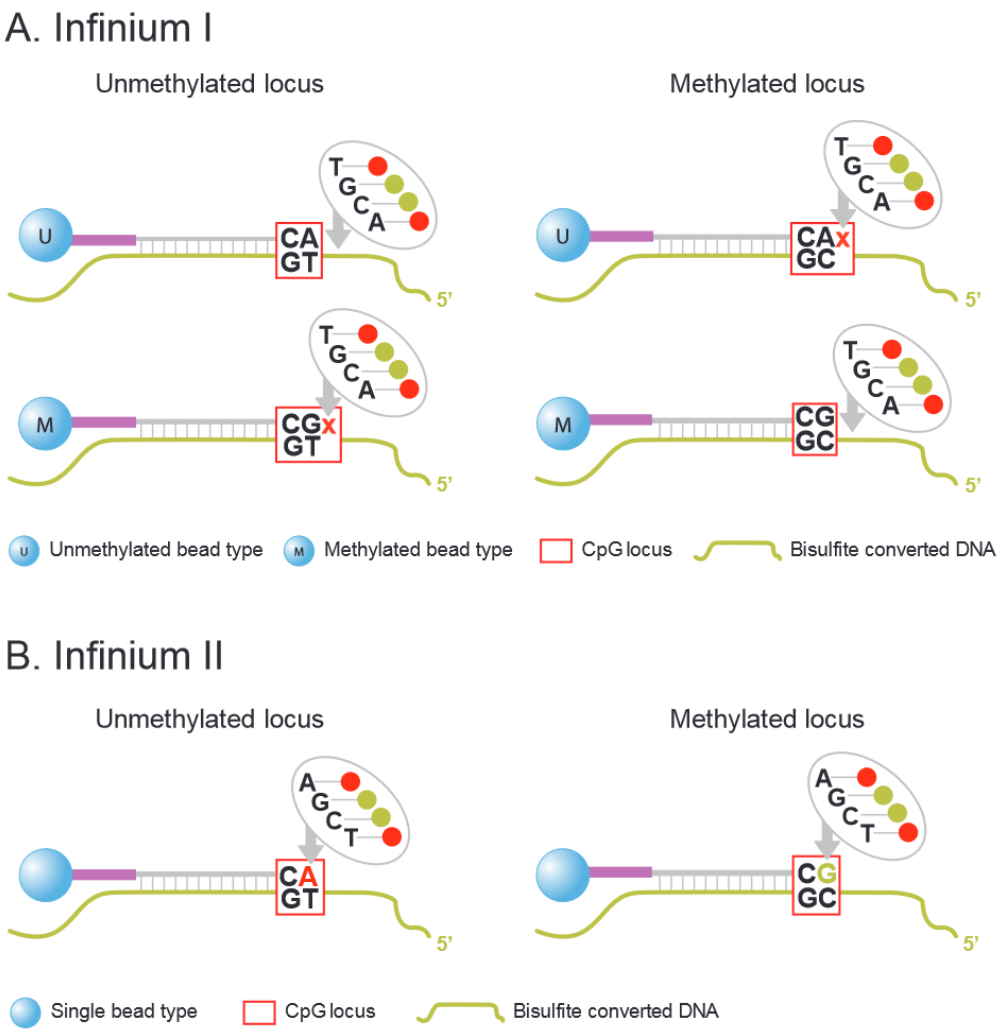


Figure 2.1 Illumina Infinium HumanMethylation BeadChip array chemistry (Figure from Illumina datasheet, 2015): The HumanMethylation450 BeadChip employs both Infinium I and Infinium II assays, enhancing its breadth of coverage. Infinium I assay design employs two bead types per CpG locus, one each for the methylated and unmethylated states. The Infinium II design uses one bead type, with the methylated state determined at the single base extension step after hybridization.

2.3.2 Illumina Infinium MethylationEPIC BeadChip array (EPIC)

Illumina have recently release a new platform, the Illumina Infinium MethylationEPIC BeadChip array (EPIC), which increases the number of CpG probes from ~450,000 to ~850,000. The EPIC array also uses a combination of the Infinium I

and II assays but achieves additional coverage by increasing the size of each array. It builds on the Infinium HM450 arrays with > 90% of the original CpGs plus an additional 350,000 CpGs in enhancer regions. Although the number of CpGs covered is increased on the new array, the same probe chemistry is being used and thus the principles for analysing the new EPIC array is similar, to a large extent, to the HM450 array.

2.4 Statistical analysis workflow

Statistical analysis of DNA methylation data was performed using the R programming language (http://www.R-project.org). Specifically, R Bioconductor packages developed for analysing DNA methylation array data such as *minfi*, *missMethyl, limma* and *methylumi* were used (Aryee et al., 2014; Phipson, Maksimovic, & Oshlack, 2016; Ritchie et al., 2015). Raw DNA methylation data generated from the array platform is stored as IDAT files. These files along with a metadata comma separated file describing each sample, is loaded into the R environment for analysis. Along with these files, specific R package libraries required for the analysis and reference annotation information from Illumina are loaded as well.

2.4.1 Quality Control

Once the data is loaded into the R environment, the data quality is evaluated. Based on the sample material, in some cases, DNA quality, quantity and sample homogeneity may vary between individual samples. Similar to quality control standards used in genome-wide association studies, epigenome-wide analyses also have standardised quality control measures. The detection p-value reported by the Illumina software (Bead Studio) is used as a quality control measure of probe performance. This value is defined as 1 - p-value computed from the background model characterising the chance that the signal was distinguishable from negative control probes. According to Illumina standards, the probes that have a detection p-value of greater than 0.05 are usually recommended to be excluded (Bibikova et al., 2011). Depending on the data however, this value may vary. Detection p-values are

generated for every CpG in every sample, indicative of the quality of methylation signal. Very small p-values are indicative of a reliable signal whilst large p-values generally indicate a poor quality signal.

2.4.2 Normalization

Normalisation addresses the removal of sources of experimental artefacts, random noise and technical variation caused by the array technology. If the dataset is not normalised, it has the potential to overshadow the true biological differences (S. Sun, Huang, Yan, Huang, & Lin, 2011). There are various types of normalisation techniques developed for methylation arrays however in this study, two different types of normalisation are considered: (1) between-array normalisation, removing technical variations between samples on different arrays, and (2) within-array normalisation, correcting for array features such as intensity-related dye biases or probe types (Siegmund, 2011). Pre-processing can be different depending on the different R packages used and can be used to smooth out datasets with global methylation differences or different tissue types, or to normalise data in a single tissue type, where there is no global variability. The packages currently used for normalization are described below:

2.4.2.1 SWAN

Subset-quantile within-array normalization (SWAN) defines an average quantile distribution using probes that are biologically similar based on their total CpG content and allows both the Infinium I and II probes to be normalised together (Maksimovic, Gordon, & Oshlack, 2012).

2.4.2.2 SQN

Subset quantile normalisation (SQN) (Touleimat N, 2012), calculates subset of probes using the genomic locations of CpGs and applies quantile normalisation. The reference quantiles used in this approach are based only on type I Infinium probes with significant detection p-values.

2.4.2.3 BMIQ

The β-mixture quantile dilation normalisation (BMIQ) method (Ruth Pidsley, 2013), is within array normalisation method, using quantiles to normalise the type II infinium probe values into a distribution comparable to the type I infinium probes using a β-mixture model that fits type I and type II probes separately and then quantile normalises the distributions of type II probes corresponding to those of type I probe (Teschendorff et al., 2013). This method uses a three-methylation-state β-mixture model but does not use fit to the middle 'hemimethylated' component in the normalisation; therefore, it requires a dual distribution (Teschendorff et al., 2013). With this method, probes that have matching biological characteristics are not selected and improve normalisation of the data.

2.4.3 Data exploration

2.4.3.1 Principal Component Analysis

The use of principal component analysis in methylation analysis was initially described in the genome-wide HumanMethylation27k methylation analysis (Teschendorff et al., 2013). Principal component analysis is used to develop a small number of artificial variables, called principal components, which account for most of the variance in the observed variables of a sample (Jolliffe & Cadima, 2016). The first few components are kept as significant predictors for statistical analyses; however, additional principal components may be of biological significance due to strong confounding variables in the dataset. The main disadvantage with PCA lies in the poor interpretability of the resulting principal components and the requirement of a large sample size in order to obtain reliable results.

2.4.3.2 Multi-dimensional Scaling

Multi-dimensional scaling (MDS) plots are widely used and highly popular for visualising data. The foundation of MDS plots come from principal components analysis, which is an unsupervised method for looking at the similarities and differences between the various samples. Generally, samples that are more similar to each other cluster together, and samples that are very different are distant to each other on the plot. There are various dimensions that help visualise the degree of

variability within the samples. The first dimension (or principal component one) takes into account the largest source of variation in the data, second dimension takes the second largest orthogonal source of variation in the data and so on. By highlighting the data with known factors of interest, it is easy to identify the where the variation comes from in the data. One of the greatest uses of MDS plots is that it can help identify samples mix-ups right at the beginning of the analysis. By understanding the factors causing variability in the sample, the analysis can be driven to take into account these factors and adjust for unwanted bias.

2.4.3.3 Filtering

Poor probes

Probes that have failed in one or more samples are removed for better accuracy of statistical testing. Based on detection p-values, the low performing probes are filtered out from the analysis.

Cross-reactive probes

Probes that are cross-reactive across the genome are usually removed from all methylation analysis. This is due to the cross-hybridisation of infinium probes on the 450K array (Chen et al., 2013) or the EPIC array (Pidsley et al., 2016) that may account for the large spurious effects during analysis.

Single nucleotide polymorphism (SNP)

Another issue during quality control at the probe level arises from some probes within the CpG loci, which include single-nucleotide polymorphisms (SNPs) near or within the probe sequence or even in the target CpG dinucleotide. There may be up to 25% probes on the 450K array that are affected by a SNP. It is important to remove the SNP-associated loci from the data as this may have implications in downstream analysis.

Sex-specific probes

Probes present on the X and Y chromosomes are removed as part of data filtering when samples include both males females. This is done to remove unwanted bias in data analysis as males and females have different methylation of the X chromosome

which is usually much larger than the effect of interest. Filtering out these chromosomes enables comparison of both males and females. This step depends on the samples and the questions being asked during analysis.

Cell type heterogeneity

Since DNA methylation varies largely between all types, it is important to check the consistency of cell type proportions in samples and correct for heterogeneity if necessary. Cell-type percentages may be derived directly through blood counts as part of clinical diagnosis and this can be used as covariates in the regression model. If this information is absent, cell-type components are predicted using the Houseman method (Houseman et al., 2012). This method identifies proportions of Tcell, Bcell, granulocytes, natural-killer cells, etc and is common practice in methylation analysis.

2.4.4 Regression analysis (differentially methylated probes (DMP))

For each CpG, there are two measurements: a methylated intensity (denoted by M) and an unmethylated intensity (denoted by U). These intensity values can be used to determine the proportion of methylation at each CpG locus. Methylation levels are commonly reported as either beta values ($\beta = M/(M + U + \alpha)$) where α is a constant of absolute value 100, to regularise beta when both M and U values are small (L. J. Bibikova M, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL., 2009; L. Z. Bibikova M, Zhou L, Chudin E, 2006) or M-values ($Mvalue = \log2(M/U)$). Beta values and M-values are related through a logit transformation. Beta values are generally preferable for describing the level of methylation at a locus or for graphical presentation because percentage methylation is easily interpretable. However, due to their distributional properties, M-values are more appropriate for statistical testing (Du et al., 2010).

Differential methylation analysis allows identifying probe-wise methylation differences within a specific group (within monozygotic twin pairs) or between specific groups (such as cases and controls). Specifically, the *limma* model uses an empirical Bayes moderated *t*-test, computed for each probe, which is similar to a paired *t*-test, except that the residual standard deviation is pooled across all samples (Ritchie et al., 2015). This "pooling of information" from many probes makes it powerful when there are only few replicates.

In this study, differential methylation analysis was performed using the Remove unwanted variation (RUVm) method (Maksimovic, Gagnon-Bartsch, Speed, & Oshlack, 2015), implemented using the *RUVadjust* function in the *missMethyl* R package (Phipson et al., 2016). RUV is a data-driven method of controlling for unwanted technical and biological variation in regression analyses using an empirically determined set of control probes. This method was first developed for gene expression analysis (Gagnon-Bartsch JA, 2012) and was later adapted for methylation analysis (Maksimovic et al., 2015). RUVm works on the concept of estimating the unwanted variation using negative control features that should not be associated with the factor of interest but are affected by the unwanted variation instead. RUVm uses factor analysis of the negative control features to estimate the components of unwanted variation and are then included in a linear model to perform the adjustment. Differentially methylated probes that had a false discovery rate (FDR) of < 0.05 after correcting for multiple testing using the Benjamini-Hochberg method (Hochberg, 1995) was considered statistically significant.

2.4.5 Statistical testing (differentially methylated regions (DMR))

Often a *probe-wise* analysis is performed, however it is also important to know whether multiple CpGs, in close proximity to each other, are concordantly differentially methylated. The identification of differentially methylated *regions* can be done using several bioconductor packages. Some of the most popular are the *bumphunter* function in *minfi* package (Aryee et al., 2014; Jaffe et al., 2012), and, the recently published *dmrcate* in the *DMRcate* package (Peters et al., 2015). They are each based on different statistical methods. *DMRcate* is based on the *limma* model that takes into account the moderated *t-test* scores and the confounding variables used in the regression analysis. DMRcate applies a kernel smoothing process and Stouffer's method to compute p-values that are then corrected by Benjamini-Hochberg method (Hochberg, 1995) and considered significant CpGs if within a certain threshold. Regions for DMRs are identified by collapsing contiguous significant CpGs that are within a defined number of nucleotides from each other.

The *bumphunter* function calculates 'bumps' in the data by running user-defined number of permutations and calculating significant p-value. In comparison to *DMRcate*, the process can be slow to run or the run needs to be parallelised. The

parameters used in the *bumphunter* function for each study are detailed in the respective chapters.

2.4.6 Gene Ontology testing

Gene Ontology (GO) provides a detailed representation of functional relationships between biological processes, molecular function and cellular components (Ashburner et al., 2000). To test for enrichment of gene ontology terms, all differentially methylated CpGs were ranked by p-value and the resulting ranked gene list was supplied to the R package, *gometh* from the *missMethyl* package. *GOmeth* performs GO enrichment analyses taking into account the varying numbers of CpGs associated with genes. For example, in the HM450 array, the numbers of CpGs mapping to genes can vary from as few as 1 to as many as 1200. The genes that have more CpGs associated with them will have a higher probability of being identified as differentially methylated compared to genes with fewer CpGs. However, *gometh* takes care of this bias in the data by adjusting for the number of CpGs associated with each gene. It considers the significant CpGs and calculates the probability of a gene being selected given the number of associated CpGs. A Wallenius' noncentral hypergeometric distribution test is then performed for each gene ontology (GO) category. Significant GO terms were selected among the category of biological process based on p-value and FDR threshold of 0.05.

2.4.7 Biological pathway analysis

Pathway analysis was performed using the *gometh* function from the *missMethyl* package, which takes into account the variable number of probes associated with each gene. The KEGG option of the *gometh* function in *missMethyl* was used to provide further insights into relevant pathways associated with the differentially methylated CpGs.

2.5 Validation on an independent platform

Site-specific validation was attempted using the Sequenom MassArray EpiTYPER (Agena Biosciences). The Sequenom EpiTYPER is a DNA methylation analysis technology that allows looking through hundreds on CpGs and amplifying sequences up to 600 base pairs to detect around 5% methylation differences. The method employs base specific cleavage and matrix assisted desorption/ionization mass spectrometry (MALDI-TOF MS). DNA is bisulphite-converted followed by PCR amplification of an area of interest (Suchiman et al., 2015). RNA transcription is performed on the reverse stand followed by base specific cleavage. Analysis of the weight of the cleavage products results in a signal pattern distinct to both methylated and non-methylated DNA fragment.

In-silico assay prediction was performed using the BiocLite MassArray package. The source and batch of DNA used for validation was the same as that used for the Infinium array analysis. Bisulphite treatment of genomic DNA was performed using the EZ-96 DNA Methylation-Lightning MagPrep kit (Zymo Research; Irvine, USA) according to the manufacturer's instructions. T7 tagged primers were designed for gene regions of interest using the Sequenom EpiDesigner package (Suchiman et al., 2015). PCR amplification was performed on the bisulphite converted DNA samples. Each DNA sample was amplified in triplicate, to ensure that the results were reliable and reproducible. The amplification reaction consisted of 5.3 μL of nuclease free water, 7.5 μL of FastStart Master Mix (Roche, Mannhein, Germany), 0.6 μL of 10 μM forward primer, and 0.6 μL of 10 μM reverse primer with thermal cycling conditions as follows: 95 °C for 10 minutes; 5 cycles of 95 °C for 10 seconds, annealing temperature determined by gradient optimisation for 30 seconds and 72 °C for 2 minutes; 35 cycles of 95 °C for 10 seconds, annealing temperature determined by gradient optimisation for 30 seconds and 72 °C for 90 seconds and final extension at 72 °C for 7 minutes.. Each 96 well PCR plate also included three negative controls, where nuclease free water replaced DNA. To reduce the impact of batch effect, all samples for each DMR were amplified in a single PCR plate.

Raw data generated from the MassArray EpiTYPER were cleaned using a Microsoft Excel macro developed in-house (Cruickshank et al., 2013; Ollikainen et al., 2010). The median value of triplicates was determined and any replicates >10 % from the median were removed as previously described (Cruickshank et al., 2013; Ollikainen et al., 2010). Pearson's correlation coefficients and p-values were calculated to assess

the cross-platform correlation. A correlation coefficient >0.50 was used as the cut-off for a strong correlation, and a p-value ≤ 0.05 indicated a significant correlation.

Chapter 3 Epigenome-wide analysis in newborn blood spots from monozygotic twins discordant for cerebral palsy reveals consistent regional differences in DNA methylation

This chapter comprises entirely of a publication:

Mohandas N, Bass-Stringer S, Maksimovic J, Crompton K, Loke JY, Walstab J, Reid SM, Amor DJ, Reddihough D, Craig JM. Epigenome-wide analysis in newborn blood spots from monozygotic twin pairs discordant for cerebral palsy reveals consistent regional differences in DNA methylation. Clinical Epigenetics. 2018 10, 25.

Supplementary material is listed in Appendices 1 - 11 (Page xiii - xi)

Mohandas *et al. Clinical Epigenetics* (2018) 10:25
https://doi.org/10.1186/s13148-018-0457-4

Clinical Epigenetics

RESEARCH Open Access

Epigenome-wide analysis in newborn blood spots from monozygotic twins discordant for cerebral palsy reveals consistent regional differences in DNA methylation

Namitha Mohandas[1,2†], Sebastian Bass-Stringer[1†], Jovana Maksimovic[2,3], Kylie Crompton[2,4,5], Yuk J. Loke[1], Janet Walstab[2,4], Susan M. Reid[2,4,5], David J. Amor[2,4,5], Dinah Reddihough[2,4,5] and Jeffrey M. Craig[1,2,6*]

Abstract

Background: Cerebral palsy (CP) is a clinical description for a group of motor disorders that are heterogeneous with respect to causes, symptoms and severity. A diagnosis of CP cannot usually be made at birth and in some cases may be delayed until 2–3 years of age. This limits opportunities for early intervention that could otherwise improve long-term outcomes. CP has been recorded in monozygotic twins discordant for the disorder, indicating a potential role of non-genetic factors such as intrauterine infection, hypoxia-ischaemia, haemorrhage and thrombosis. The aim of this exploratory study was to utilise the discordant monozygotic twin model to understand and measure epigenetic changes associated with the development of CP.

Methods: We performed a genome-wide analysis of DNA methylation using the Illumina Infinium Human Methylation 450 BeadChip array with DNA from newborn blood spots of 15 monozygotic twin pairs who later became discordant for CP. Quality control and data preprocessing were undertaken using the *minfi* R package. Differential methylation analysis was performed using the remove unwanted variation (RUVm) method, taking twin pairing into account in order to identify CP-specific differentially methylated probes (DMPs), and *bumphunter* was performed to identify differentially methylated regions (DMRs).

Results: We identified 33 top-ranked DMPs based on a nominal p value cut-off of $p < 1 \times 10^{-4}$ and two DMRs ($p < 1 \times 10^{-3}$) associated with CP. The top-ranked probes related to 25 genes including *HNRNPL*, *RASSF5*, *CD3D* and *KALRN* involved in immune signalling pathways, in addition to *TBC1D24*, *FBXO9* and *VIPR2* previously linked to epileptic encephalopathy. Gene ontology and pathway analysis of top-ranked DMP-associated genes revealed enrichment of inflammatory signalling pathways, regulation of cytokine secretion and regulation of leukocyte-mediated immunity. We also identified two top-ranked DMRs including one on chromosome 6 within the promoter region of *LTA* gene encoding tumour necrosis factor-beta (TNF-β), an important regulator of inflammation and brain development. The second was within the transcription start site of the *LIME1* gene, which plays a key role in inflammatory pathways such as MAPK signalling. CP-specific differential DNA methylation within one of our two top DMRs was validated using an independent platform, MassArray EpiTyper.

(Continued on next page)

* Correspondence: jeff.craig@mcri.edu.au
†Equal contributors
[1]Environmental and Genetic Epidemiology Research, Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia
[2]Department of Paediatrics, The University of Melbourne, Flemington Road, Parkville, Victoria 3052, Australia
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: Ours is the first epigenome-wide association study of CP in disease-discordant monozygotic twin pairs and suggests a potential role for immune dysfunction in this condition.

Keywords: Cerebral palsy, DNA methylation, Inflammation, Epigenetics, Discordant twins

Background

Cerebral palsy (CP) describes a group of motor impairment syndromes caused by lesions or anomalies of the developing brain [1]. It is non-progressive, but the severity of symptoms may change over time [2]. CP is the most common childhood physical disability [3] with a worldwide prevalence of 2.11 per 1000 live births [4]. In preterm infants (< 37 weeks gestation), the prevalence is higher, ranging from 5 to 92 per 1000 depending on gestational age [5]. The prevalence of CP in multiple births is almost four times that of singletons [6]. There are many factors that may be responsible for this increased risk in multiple birth pregnancies, with the most likely being low birth weight and preterm birth, both known risk factors for CP.

The brain insult or anomaly resulting in CP may occur during the prenatal, perinatal or early postnatal period [1, 7], and in many cases, the timing is unknown. Although newborns may be recognised as being at risk of CP, less than half of all children who are ultimately diagnosed are identified before 1 year of age, and only three quarters are identified before age 2 [8].

CP has a multifactorial pathogenesis and risk factors including intrauterine growth restriction or infection, placental abnormalities, inflammation, signs of fetal distress and genetic variation [9, 10]. Although the latter explains a proportion of CP cases, particularly cerebral maldevelopments [9, 11, 12], many non-genetic factors likely play a role [7] though their respective contributions have not been comprehensively addressed [13–15].

Genetically identical monozygotic (MZ) twin pairs discordant for CP highlight the role of non-shared factors in the pathogenesis of CP [16]. Non-shared factors are often described as the differences in the intrauterine environment that influences the development of individual members of a twin pair [17, 18]. Such within-pair variation can arise from differences in the length and morphology of the umbilical cord or placenta. This can affect growth rate and development of the individual twins leading to discordance in infection or inflammation, leading to disease discordance [13, 19, 20]. More generally, the study of phenotypically discordant MZ twins, matched for genetic variation, sex, gestational age and maternal factors, provides a great opportunity to examine the role of epigenetics in disease aetiology by allowing us to isolate the effect of such non-shared environmental factors [21, 22].

Epigenetics refers to a range of modifications and processes that regulate the activity of DNA, including gene expression. Epigenetic variation has emerged as a candidate mediator of a range of health outcomes beginning in early life as part of the 'Developmental Origins of Health and Disease' (DOHaD) phenomenon. [23–25]. In DOHaD, environmental conditions in utero and during infancy alter the developmental trajectory of an individual, which manifests as specific chronic health phenotypes later in life.

DNA methylation is the most widely studied epigenetic process and is one of the mechanisms that are involved in tissue differentiation during early development. Despite this, previous studies have investigated the concordance in DNA methylation state between the brain and peripheral tissues, revealing many similarities [26–28]. Several epigenome-wide association studies (EWAS) have identified DNA methylation variation in the cord blood in association with later neurocognitive function and behaviour [29–31]. Similarly, the whole blood has been used to detect differential methylation patterns between affected and unaffected individuals in brain disorders such as schizophrenia [27], bipolar disorder [32, 33] and Alzheimer's disease [34]. Animal studies have also reported that environmental factors affecting brain processes leave biomarker signatures in the blood with consistent methylation status across the brain and peripheral tissues [35]. We have previously shown that DNA methylation varies within pairs of MZ twins from birth [36, 37]. In this study, we hypothesised that early life non-shared factors that play a role in the aetiology of CP in discordant MZ twins may be reflected in differences in DNA methylation across tissues including neonatal blood. Furthermore, we hypothesised that a subset of differential methylation will be located in genes previously implicated in CP aetiology [38–42], in particular, pathways involved in inflammation, hypoxia-ischaemia and thrombosis.

Methods

Samples and DNA extraction

Participant CP-discordant twin pairs, who were suspected to be MZ on the basis of same sex and questionnaire data that measured concordance for hair and eye colour, were recruited through the Victorian Cerebral Palsy Register, a

population-based registry of individuals born or receiving medical services in the Australian state of Victoria. Participants were excluded if either twin presented with another known neurological disorder. Written informed consent from the families was obtained. A single 1-cm-diameter dried blood spot was acquired from all participants from a neonatal newborn screening card collected 2 to 4 days after birth and then stored within the Victorian Clinical Genetics Service.

Genomic DNA (gDNA) was extracted from the dried blood spot samples using the ZR DNA Card Extraction Kit (Zymo Research, Irvine, CA, USA) with some modifications to the manufacturer's protocol. Briefly, eight 3-mm punches were from each 1-cm-diameter blood spot and were transferred to a 2-mL Eppendorf Safe-Lock microcentrifuge tube (Merck, Darmstadt, Germany) containing ZR BashingBeads. Four hundred microliters of PBS containing 40 μL of 20 mg/mL proteinase K (Sigma-Aldrich, St. Louis, Missouri, USA) was added, and samples were vortexed and centrifuged briefly, followed by incubation overnight at 37 °C. Following incubation, 400 μL of ZR lysis solution was added to each tube. Punches were homogenised for 30 s at 4 m/s^2 using Thermo Savant FastPrep 120 Cell Disrupter System (Global Medical Instrumentation (GMI) Incorporation, Minnesota, USA). Tubes were centrifuged for 1 min at 10,000 rpm, and 390 μL of 2× digestion buffer and 10 μL of 20 mg/mL proteinase K were added. Tubes were mixed by inversion and incubated for 30 min at 55 °C, then left to cool at ambient temperature for 3–4 min before centrifuging for 1 min at 8000 rpm. Six hundred fifty microliters of supernatant was added to 1.3 mL of DNA isolation buffer contained in a 5-mL Falcon tube (Thermo Fisher Scientific, MA, USA). This mixture was passed through the Zymo-Spin IC column by centrifuging for 1 min at 14,000 rpm, followed by the discard of flow-through liquid. The spin column was then washed twice by adding 200 μL of DNA wash buffer and centrifuged for 1 min at 14,000 rpm. Finally, 20 μL of DNA elution buffer (prewarmed at 55 °C) was added to the column and incubated at ambient temperature for 15 min before final centrifugation for 2 min at 14,000 rpm. This was repeated, resulting in a final elution volume of 40 μL containing genomic DNA. DNA concentration was measured by spectrophotometry (Nanodrop, Wilmington, DE, USA) to allow calculation of the required volume of each sample for array analysis. The quality of the extracted gDNA samples was visualised using agarose gel electrophoresis.

Illumina Infinium HumanMethylation450 arrays

Following bisulphite conversion of genomic DNA, genome-wide analysis of DNA methylation was assessed using HM450 (Illumina, San Diego, CA, USA), at the Department of Pathology, University of Melbourne. Hybridisation and scanning were performed following the manufacturer's instructions. Statistical analysis was performed using the R statistical programming language (http://www.R-project.org) in conjunction with Bioconductor packages developed for the analysis of methylation arrays.

Preprocessing of Illumina Infinium 450K array data

The raw intensity data (IDAT files) were imported into R (3.3.1; http://cran.r-project.org/). Data quality was assessed using the *minfi* (v1.20.2) Bioconductor package [43]. From 485,512 HM450 probes, 67,120 were removed based on either (1) poor performance (mean detection p value of >0.01, $n = 14{,}056$); (2) probes containing either a single nucleotide polymorphism (SNP) at the target CpG site or at the single nucleotide extension site ($n = 16{,}307$); (3) probes that map to multiple locations in the genome ($n = 27{,}120$), [44]; and (4) or to sex chromosomes ($n = 9637$). Samples were also evaluated using a modified version of the Houseman method [45, 46] implemented in *minfi*, to estimate the cell type composition. The Wilcoxon signed-rank statistical test was used to compare the difference in cell type proportion between CP cases and controls. The analysis was completed before a cord blood reference panel was widely available, so cohorts used an adult whole blood reference [43] to estimate the proportion of B cells, CD8+ T-cells, CD4+ T-cells, granulocytes, NK cells and monocytes in each sample. The data was normalised using subset-quantile within array normalisation (SWAN) [47]. Covariates such as birth weight, birth order and postnatal age (in days) at which newborn screening cards were created were assessed as potential confounders.

Differential methylation analysis

Beta (β) values (proportion of the methylated signal over the total signal) were converted to M-values, the $\log_2$ ratio of the intensities of the methylated signal versus the unmethylated signal. Differential methylation analysis was performed using remove unwanted variation (RUVm) [48], implemented in the *missMethyl* R package [49] taking into account the twin relationships. RUVm is a data-driven method of controlling for unwanted technical and biological variation in regression analyses using an empirically determined set of negative control probes assumed not to be associated with the biological factor of interest. p values were adjusted to control for the false discovery rate (FDR) using the Benjamini-Hochberg method [50]. Differentially methylated probes (DMPs) were considered significant if they fell within the FDR threshold of 0.1. We also investigated the top-ranked DMPs with an unadjusted p value less than 1×10^{-4} [51].

Identification of differentially methylated regions

Differentially methylated regions (DMRs) were identified using the *bumphunter* package [43, 52]. The cut-off value, which is a user-defined numeric value that determines the upper and lower bounds of the genomic profiles that will be used as candidate regions, was set to 0.02 and the number of permutations set to 1000.

Functional annotation and pathway analysis

Gene ontology and pathway analysis were performed using the *gometh* function from the *missMethyl* package [49], which appropriately takes into account the variable number of HM450 probes associated with each gene. Gene ontology enrichment was performed for the 1000 top-ranked DMP-associated genes. The KEGG option of the *gometh* function in *missMethyl* was used to provide further insights into relevant biological processes associated with the top-ranked DMPs.

Validation of differentially methylated regions

Site-specific validation was performed using the Sequenom MassArray EpiTYPER (Agena Biosciences). T7-tagged primers were designed for two regions (Additional file 1) using the Sequenom EpiDesigner package [53]. Forward primer sequences contained a 10 base 5′ tag (AGGAAGAGAG) and reverse primers a 31 base 5′ tag (CAGTAATACGACTCACTATAGGGAGAAGGCT). In silico, assay prediction was performed using the BiocLite *MassArray* package. DNA used for validation was the same as that used for the HM450 analysis. Bisulphite treatment of genomic DNA was accomplished using the MethylEasy *Xceed* Kit (Human Genetic Signatures, North Ryde, Australia). One microliter of bisulphite-converted product was amplified in triplicate for each sample using the FastStart kit (Roche, Mannheim, Germany) in 15 μL of reagents with thermal cycling conditions as follows: 95 °C for 10 min; 5 cycles of 95 °C for 10 s, 60 for 30 s and 72 °C for 2 min; 40 cycles of 95 °C for 10 s, 62 for 30 s and 72 °C for 90 s; and final extension at 72 °C for 7 min. Raw data generated from the MassArray EpiTYPER was cleaned using a Microsoft Excel macro developed in-house [54, 55]. The median value of triplicates was determined, and any replicates > 10% from the median were removed as previously described [54, 55].

Within-twin pair analysis

To explore the top-ranked CpGs within each twin pair and compare them across twin pair groups, probes were ranked according to delta beta ($\Delta\beta$, the difference in DNA methylation of the CP minus non-CP twin) within pairs, and the top-ranked 100 probes were compared across all 15 twin pairs. The genes corresponding to all probes with a $\Delta\beta$ value > 0.5 were then compared across the 15 twin pairs. Gene ontology analysis was performed on the top 1000 probes from each twin pair, and common ontologies between twin pairs were identified.

Results

Subject characteristics

The study cohort consisted of 16 CP-discordant twin pairs (ten male and six female) for which pre-screening suggested a high probability of monozygosity (Table 1). All were tested for genetic zygosity using data from 65 SNPs from the Infinium arrays. The variability of SNPs for one twin pair (pair no. 9003) was substantially larger than the remaining samples and was therefore assigned as dizygotic (DZ). This pair was excluded from further analysis. Five subtypes of CP were reported: spastic diplegia (6), spastic quadriplegia (3), spastic hemiplegia (3), dyskinesia (2) and ataxia (1). The severity of CP ranged from mild (independently ambulant) to severe (wheelchair dependent), and the underlying neuropathology included white matter (11), grey matter brain injury (2) and both white and grey matter mixed injury (2). Three twin pairs were born at term (37–41 weeks), while all other twin pairs were born preterm (< 37 weeks).

Global DNA methylation profiles in CP-discordant monozygotic twins

Global DNA methylation (average β value across all probes) within twin pairs was compared using a pairwise Pearson correlation for the 418,392 probes remaining after filtering and quality control (see the 'Methods' section) for all 15 twin pairs. Within-pair methylation

Table 1 Descriptive characteristics of the study cohort

Twin ID	Type of CP	Gestational age (weeks)	Brain Injury
9001	Spastic diplegia	27 (Preterm)	WMI
9002	Spastic diplegia	28 (Preterm)	WMI
9003*	Spastic diplegia	29 (Preterm)	WMI
9004	Spastic diplegia	29 (Preterm)	WMI
9005	Spastic quadriplegia	31 (Preterm)	WMI
9006	Spastic diplegia	32 (Preterm)	WMI
9007	Spastic hemiplegia	32 (Preterm)	WMI
9008	Spastic quadriplegia	33 (Preterm)	WMI
9009	Spastic diplegia	33 (Preterm)	WMI
9010	Spastic diplegia	36 (Preterm)	M
9011	Dyskinetic: dystonic	38 (Term)	GMI
9012	Dyskinetic: hypotonia	32 (Preterm)	WMI
9013	Spastic quadriplegia	38 (Term)	M
9014	Ataxia	34 (Preterm)	GMI
9015	Spastic hemiplegia	32 (Preterm)	WMI
9016	Spastic hemiplegia	38 (Term)	WMI

WMI white matter injury, *M* miscellaneous, *GMI* grey matter injury
*Twin pair 9003 was later confirmed not to be MZ and removed from the analysis

correlation coefficients ranged from 0.980 to 0.996 (Additional file 2) compared to 0.976 to 0.995 between unrelated unaffected individuals.

Top-ranked CP-associated DMPs

Cleaned data was explored by principal component (PC) analysis which revealed few (6/54) significant correlations ($p < 0.05$, $r < 0.6$ shaded in Additional file 3) between the top six principal components of DNA methylation and nine technical (e.g. age at which Guthrie card was created) and biological (e.g. sex, subtype of CP) covariates. This suggested that none of the covariates tested was consistently associated with DNA methylation. In addition, none were associated with the largest principal component of variation within the dataset, PC1. Multi dimensional scaling (MDS) plots of the first three dimensions of the processed methylation data also showed that chip location and position on the 450K array were not found to affect methylation data (Additional file 4).

Apart from the above covariates, it is known that cell-type heterogeneity within the whole blood can confound epigenome-wide analyses. Therefore cell-type composition within CP-discordant pairs was evaluated. The levels of $CD8^+$ T cells (CD8T) and $CD4^+$ T cells (CD4T), B cells, natural-killer (NK) cells, monocytes and granulocytes were compared between the two groups (CP cases and normal co-twins). There was no statistically significant difference in the estimated cell-type proportions of $CD8^+$ T cells, NK cells, B cells and monocytes ($p > 0.05$). However, the proportion of $CD4^+$ T cells was lower ($p = 0.002$), and the proportion of granulocytes was found higher ($p = 0.021$) in CP cases relative to normal co-twins (Additional file 5).

To take into account potential sources of unwanted variation (such as cell-type composition), the genome-wide analysis was performed using RUVm, which adjusts for biological and technical variation using a set of data-driven negative control probes [46, 48, 56, 57]. This analysis did not identify any significant CP-associated DMPs after adjusting for multiple testing. Nevertheless, as this is an exploratory epigenome-wide study of CP, we focused on the characteristics of the top-ranked DMPs based on a nominal p-value cut-off of $p < 1 \times 10^{-4}$ as used by others [51]. This resulted in a list of 33 top-ranked DMPs, corresponding to 25 genes (Table 2), most of which showed a consistent direction in the majority of twin pairs (> 12/15). The average difference in methylation ($\Delta\beta$ = CP twin minus unaffected twin) ranged from + 0.6 to + 11.9% and from − 2.5 to − 12.4% (Table 2). Figure 1 shows the within-pair differences in methylation for the top ten DMPs.

The top-ranked probe *cg00376816* (average $\Delta\beta = 11.6\%$, $p = 4.57 \times 10^{-6}$) was located on chromosome 19, within the gene body of the *HNRNPL* gene encoding the heterogeneous nuclear ribonucleoprotein L. Others included *cg04242728* (in the 5′ end of the *TBC1D24* gene, ranked 3) and *cg19607845* (in the gene body of *FBXO9* gene, ranked 13). Probes located near the gene body of immune and inflammatory genes, such as Ras association domain family member 5 (*RASSF5*), major histocompatibility complex DM alpha-chain (*HLA-DMA*), *CD3D* and kalirin (*KALRN*) genes, were also among the top-ranked 33 DMPs.

To identify enriched biological processes or molecular functions, we performed gene ontology analysis on genes associated with the top-ranked 1000 DMPs. The top 20 gene ontologies ranked by nominal p-value were 'regulation of immune response', 'lymphocyte activation', 'differentiation and aggregation and T cell activation' with the top two ontologies related to cell-cell adhesion processes (Table 3; Additional file 6). Enriched disease pathways, as reported by KEGG analysis, included MAPK signalling (19 associated genes from 245 genes in the KEGG pathway list; p value 3.6×10^{-10}), cytokine-cytokine receptor interaction (13 associated genes from 240 genes; p value 1.3×10^{-08}) and Ras signalling (15 associated genes from 218 genes; p value: 9.8×10^{-08}).

We also identified DMRs associated with CP [52] (Table 4). The top-ranked DMR, spanning 434 bp and with a p value of 5.6×10^{-4}, was located on chromosome 6 (Fig. 2). This DMR spans 12 probes (average $\Delta\beta = 3.7\%$) within the coding region of the *LTA* gene, approximately ~ 800 bp downstream of the transcription start site (TSS). *LTA* codes for the lymphotoxin-alpha protein otherwise known as tumour necrosis factor beta (TNF-β). Other top DMRs include those within *LTBP1*, *CD300*, *CHST11* and *LIME1*. Gene ontology analysis of the top-ranked DMR-associated genes revealed similar findings to top-ranked DMPs (Table 5; Additional file 7). We found an over-representation of inflammatory signalling pathways, also similar to that of the top-ranked DMPs. The top pathways included TNF and TGF-beta signalling and cytokine-cytokine receptor interaction. The nuclear factor kappa-light-chain-enhancer of activated B cell (NF-κB) signalling pathway was also enriched.

Validation of DMRs

LIME1 and *LTA* DMRs were selected for validation as they were highly ranked, had large, consistent effect sizes across all pairs and were biologically relevant to CP. Three CpGs from the HM450 platform contained within three CpG units on the MassArray Epityper (consisting of seven CpG sites in total) were tested for the *LIME1* DMR, and four CpGs contained within three units on the MassArray Epityper (four CpG sites in total) were tested for the *LTA* DMR, both in regions being approximately 200 base pairs upstream of the transcriptional start site (TSS200) and likely to be in gene promoters. Scatter plots were generated to assess the validity of the

Table 2 The top-ranked DMPs (ranked on unadjusted *p* value) in cerebral palsy discordant twins

Probe ID	Rank	Chromosome position	Gene	Location	*P* value	Methylation difference (%)
cg00376816	1	chr19: 39332571	*HNRNPL*		4.57E-06	11.64
cg25011252	2	chr8: 61777859	*CHD7*		6.33E-06	7.64
cg04242728	3	chr16: 2536153	*TBC1D24*		1.46E-05	− 8.20
cg03907855	4	chr8: 1106363	–	Island	1.52E-05	− 3.20
cg05707458	5	chr1: 205424829	–	Island	1.93E-05	2.39
cg18369327	6	chr5: 109220986	*LOC100289673*		2.07E-05	− 2.54
cg18768238	7	chr10: 49892829	*WDFY4*		2.52E-05	− 8.58
cg14163311	8	chr1: 206730397	*RASSF5*		3.21E-05	8.20
cg16602500	9	chr1: 52889575	*ZCCHC11*		3.64E-05	6.25
cg06419846	10	chr11: 66083572	*CD248*		3.65E-05	5.95
cg03929569	11	chr13: 30688864	–	Island	4.35E-05	0.60
cg14414943	12	chr1: 111770718	*CHI3L2*		4.65E-05	6.21
cg19607845	13	chr6: 52930050	*FBXO9*		4.79E-05	− 9.61
cg19942731	14	chr22: 45609421	*C22orf9*		5.11E-05	− 10.38
cg14073571	15	chr7: 158823178	*VIPR2*		5.14E-05	− 7.59
cg03106245	16	chr11: 47399788	*SPI1*		5.33E-05	− 11.45
cg08936645	17	chr4: 37910273	*TBC1D1*		5.36E-05	− 7.84
cg02806715	18	chr6: 32920567	*HLA-DMA*		5.69E-05	− 6.58
cg17512380	19	chr10: 18971598	–	OpenSea	5.81E-05	5.67
ch.10.89216809R	20	chr10: 89226829	–	OpenSea	6.52E-05	− 3.85
cg15613292	21	chr2: 232478359	–	Shore	6.89E-05	2.62
cg00263248	22	chr14: 98444151	*C14orf64*		6.94E-05	6.40
cg19540797	23	chr2: 3605489	*RNASEH1*		7.30E-05	2.27
cg08360638	24	chr13: 20781097	–	OpenSea	7.52E-05	7.09
cg07728874	25	chr11: 118213272	*CD3D*		7.93E-05	9.14
cg11348257	26	chr1: 76556226	*ST6GALNAC3*		8.03E-05	− 12.35
cg07011093	27	chr3: 123987726	*KALRN*		8.38E-05	− 5.03
cg09335613	28	chr6: 71998106	*OGFRL1*		8.85E-05	− 6.68
cg22230912	29	chr3: 16331335	*OXNAD1*		8.90E-05	11.86
cg13505608	30	chr9: 140128562	*SLC34A3*		9.22E-05	2.53
cg04975778	31	chr2: 62732758	*TMEM17*		9.26E-05	2.31
cg12306086	32	chr4: 106117747	*TET2*		9.52E-05	− 10.22

Methylation difference (%) was calculated as the mean of the DNA methylation levels of the CP twin minus the unaffected twin ($\Delta\beta$)

data for the three *LIME1* and *LTA* probes within a 250-base-pair region across both the HM450 and the EpiTYPER platforms (Additional file 8). Pearson's correlation coefficients were determined, and the significance of the correlation was assessed for each probe-CpG unit comparison. Five out of six probes (*cg21201401*, *cg06653796*, *cg14597739*, *cg11586857*, *cg21999229*) had a positive correlation between the two platforms. Among these, one probe out of three for *LIME1* and two of three for *LTA* had moderate correlations ($r > 0.5$). All moderate correlations were also significant with $p < 0.05$ ($r = 0.88$, $p = 1.6 \times 10^{-4}$; $r = 0.40$, $p = 0.197$; $r = 0.12$, $p = 0.65$ for *LIME1* and $r = 0.69$, $p = 1.4 \times 10^{-4}$; $r = 0.59$, $p = 0.043$; $r = 0.21$, $p = 0.34$ for *LTA*.). The $\Delta\beta$ values were calculated for two of the three probes within the *LTA* gene region, with correlation coefficient values of $r = 0.69$ for *cg14597739* and $r = 0.21$ for *cg21999229*. The $\Delta\beta$ for probes within the *LIME1* gene region were not calculated due to insufficient data points, resulting from limited material remaining, for a valid within-pair analysis.

Differential methylation analysis within individual twin pairs

Since CP is a highly heterogeneous disorder [58], it is possible that a subset of disease-associated DNA methylation

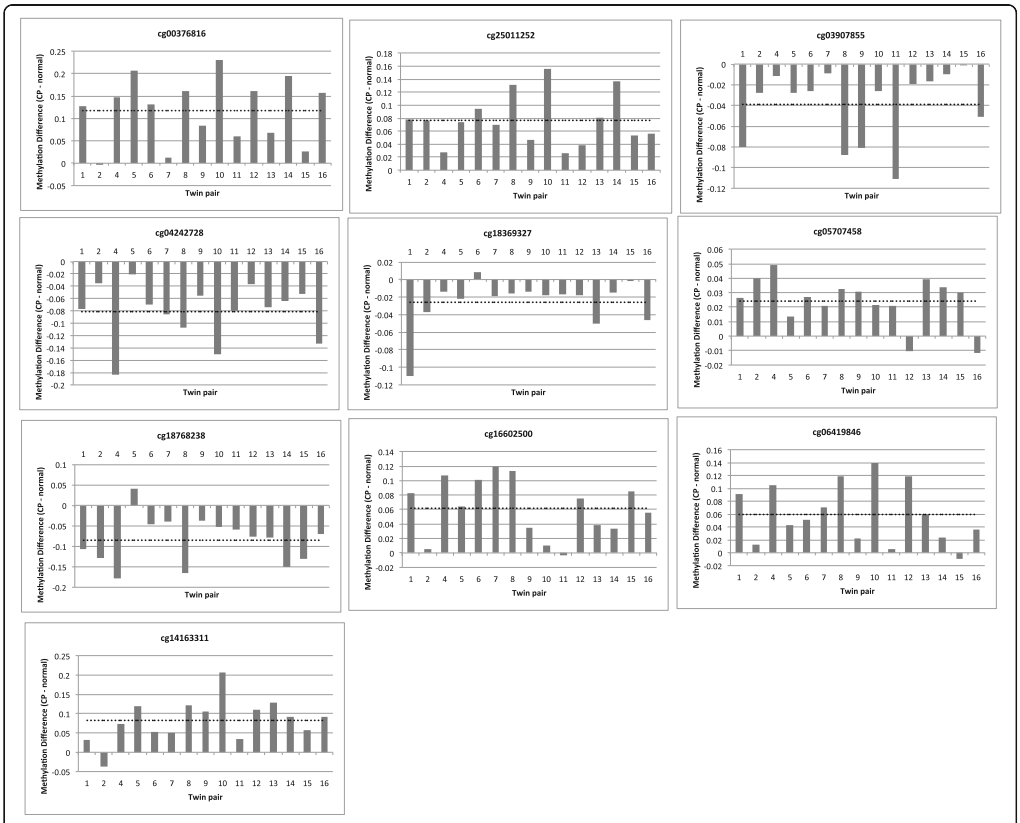


Fig. 1 DNA methylation differences (CP minus normal twin) for the top ten CP-specific DMPs. The direction of methylation difference for each probe was consistently in the same direction (hyper- or hypomethylated) across > 12/15 twin pairs. The dotted line represents the average methylation difference, and methylation difference is represented in absolute value (effect size of 0.1 is 10% methylation difference)

patterns may be specific to each proband. To test this [59], we determined the top CP-associated CpGs for each twin pair ranked by absolute differences in DNA methylation ($\Delta\beta > 0.5$) and looked at significant gene ontologies common to multiple pairs (Additional file 9). Gene ontologies corresponding to cell adhesion were found in 5/15 twin pairs (Additional file 10). Similarly, common CpGs with a within-pair methylation difference > 50% were found in multiple twin pairs (Table 6) in genes such as *BICD2*, *HLA-DPB2*, *RPTOR* and *PIK3CG* (Additional file 11), involved in neuronal cell migration, muscular atrophy or muscle contraction pathways and immune response and inflammatory pathways, respectively [60–63]. Notably, two different CpG sites within twin pairs 4 and 8 corresponding to the *WWTR1* gene had an absolute methylation difference of greater than 50% with the same direction of effect. Affected CpG sites were located near the 5′ end of the gene in both pairs (Table 6).

Discussion

This exploratory study represents an initial step towards investigating potential CP-associated epigenetic differences, with the longer-term aim of identifying predictive biomarkers with clinical utility. We identified DNA methylation differences in dried blood spots from 15 CP-discordant MZ twin pairs and found differential methylation at several gene loci associated with hypoxia signalling, inflammation and cell adhesion. These pathways had been previously linked to CP, consistent with part of our hypothesis.

Pairwise global DNA methylation difference between CP and non-CP members of each pair measured for

Table 3 Top 20 gene ontology (GO) terms (BP = biological process) analysed for the 1000 top-ranked CP-associated DMPs

GO ID	GO term	Ontology	No. of genes	DM genes	Unadjusted *p* value
GO:0098609	Cell-cell adhesion	BP	1091	90	8.27E-06
GO:0007156	Homophilic cell adhesion via plasma membrane adhesion molecules	BP	149	25	1.70E-05
GO:0042098	T cell proliferation	BP	157	18	2.91E-05
GO:0042129	Regulation of T cell proliferation	BP	133	16	3.28E-05
GO:0046649	Lymphocyte activation	BP	557	46	6.47E-05
GO:0032729	Positive regulation of interferon-gamma production	BP	58	10	7.55E-05
GO:0070661	Leukocyte proliferation	BP	241	23	9.24E-05
GO:0001775	Cell activation	BP	811	61	9.70E-05
GO:0050863	Regulation of T cell activation	BP	270	26	1.27E-04
GO:0022610	Biological adhesion	BP	1611	117	1.28E-04
GO:1903037	Regulation of leukocyte cell-cell adhesion	BP	284	27	1.30E-04
GO:0070663	Regulation of leukocyte proliferation	BP	187	19	1.66E-04
GO:0045321	Leukocyte activation	BP	656	50	1.93E-04
GO:0007159	Leukocyte cell-cell adhesion	BP	442	37	2.27E-04
GO:0098742	Cell-cell adhesion via plasma-membrane adhesion molecules	BP	214	27	2.32E-04
GO:0050865	Regulation of cell activation	BP	437	37	2.37E-04
GO:0016337	Single organismal cell-cell adhesion	BP	669	54	2.49E-04
GO:0032649	Regulation of interferon-gamma production	BP	85	11	2.51E-04
GO:0007155	Cell adhesion	BP	1606	115	2.56E-04
GO:0045601	Regulation of endothelial cell differentiation	BP	28	7	2.75E-04

DM differentially methylated, *FDR* false discovery rate

each twin group and comparison of the size of DNA methylation difference made between groups allowed for an assessment of how variable the differences in methylation may be for different cases of CP. Our results are consistent with previous studies (e.g. [59]) that have indicated that neurodevelopmental disorders such as autism spectrum disorder are not associated with systemic within-twin pair differences in global DNA methylation.

We tested site-specific DNA methylation patterns across the genome for their association with CP, taking CP status within discordant MZ pairs into account. Although no probe reached an adjusted statistical significance at FDR < 0.1, the top-ranked DMPs, at a nominal cut-off of $p < 1 \times 10^{-4}$, were enriched for the cellular processes of inflammation, cell adhesion and immune response. These showed the direction of effect across most or all discordant twin pairs. This was in accordance with EWAS of other neurodevelopmental disorders including twins discordant for autism spectrum disorders [59], asthma [51], depression [64] and aggressive behaviour [65].

Table 4 Differentially methylated regions (DMRs) from Bumphunter analysis

Chromosome position	Gene	Gene description	*P* value	Fwer
chr6: 31539735-31540169	*LTA*	Lymphotoxin-alpha	0.00056	0.988
chr12: 31799116-31799118	–	–	0.00125	0.992
chr11: 47399813-47400146	*SPI1*	Spi-1 Proto-Oncogene	0.00129	0.996
chr2: 33359198-33359688	*LTBP1*	Latent transforming growth factor beta binding protein 1	0.00166	0.996
chr17: 72527607-72527724	*CD300LB*	CD300 molecule like family member b	0.00244	0.996
chr12: 105071483-105071483	*CHST11*	Carbohydrate chondroitin 4 sulfotransferase 11	0.00259	0.996
chr9: 130868874-130868874	*SLC25A25*	Solute carrier family 25 member 25	0.00269	0.996
chr11: 2722062-2722195	*KCNQ1*	Potassium voltage-gated channel subfamily Q member 1	0.00164	1
chr1: 27961563-27962037	*FGR*	FGR proto-oncogene, Src family tyrosine kinase	0.00223	1
chr20: 62367805-62367893	*LIME1*	Lck interacting transmembrane adaptor 1	0.00167	1

fwer family-wise error rate

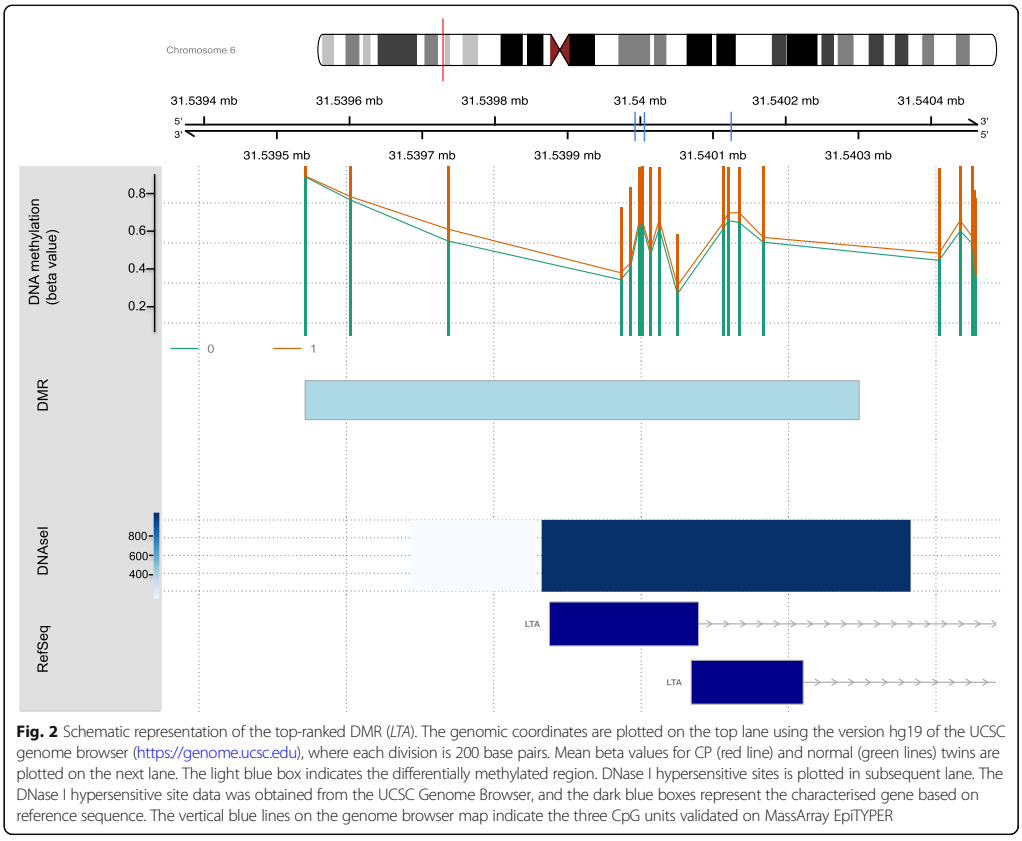


Fig. 2 Schematic representation of the top-ranked DMR (*LTA*). The genomic coordinates are plotted on the top lane using the version hg19 of the UCSC genome browser (https://genome.ucsc.edu), where each division is 200 base pairs. Mean beta values for CP (red line) and normal (green lines) twins are plotted on the next lane. The light blue box indicates the differentially methylated region. DNase I hypersensitive sites is plotted in subsequent lane. The DNase I hypersensitive site data was obtained from the UCSC Genome Browser, and the dark blue boxes represent the characterised gene based on reference sequence. The vertical blue lines on the genome browser map indicate the three CpG units validated on MassArray EpiTYPER

Average within-pair DNA methylation differences of up to 12.5% were observed, comparable to previous findings in neurodevelopmental disorders, ranging from 1.5 to 12%. Furthermore, we validated five of the six CpG probes, with positive correlations across platforms. Three of these five probes (one from *LIME1* and two from *LTA*) showed a strong and significant cross-platform correlation, indicating the validity of methylation values between platforms.

Genes associated with top-ranked DMPs and DMRs were enriched for similar ontologies and pathways, namely immune response, lymphocyte-mediated immunity, interferon-gamma production and regulation of immune response.

Our study revealed top-ranked DMRs associated with genes that play a role in inflammation, such as *LTA/TNF*β and *LIME1*, supporting part of our hypothesis that inflammation plays a key role in CP aetiology. Genetic variants of *LTA* have been implicated in multiple studies as being associated with risk for CP [66–68]. *LTA* plays an important role in inflammation and brain development, mediating preterm birth and white matter brain injury [69]. It has been implicated that inflammation [70] and increased levels of its isoform TNF-α were found in children with CP compared to healthy controls [71]. *LIME1* gene links T and B cell signalling to the activation of tyrosine and MAP kinases [72]. Other top DMR-associated genes such as *LTBP1* and *CD300* are also known mediators of inflammatory pathways such as ERK signalling pathway, interleukin-3,-5 pathway, B cell receptor signalling pathway and other chemokine signalling [72–74]. Genes associated with top-ranked DMPs also showed involvement in other key inflammatory pathways such as Ras signalling (*WDFY4*), MAP kinase signalling (*CD3D*) and interleukin-3,-5 signalling (*KALRN*). Analysis within individual twin pairs also revealed associations to the *WWTR1* gene, which is involved in the activation of TGF-β signalling pathway, an inflammatory pathway that regulates neural survival and death [75, 76]. Gene ontology analysis of top-ranked DMPs showed enrichment of genes involved in regulation of immune response pathways such as those involved in signalling or T-cell activation. It is noteworthy that

Table 5 Gene ontology (GO) analysis for cerebral palsy-associated DMRs. (DM: differentially methylated; FDR: false discovery rate)

GO ID	GO term	Ontology	No. of genes	DM genes	Unadjusted *p* value	FDR
GO:0002705	Positive regulation of leukocyte mediated immunity	BP	80	5	4.57E-10	9.60E-06
GO:0002703	Regulation of leukocyte mediated immunity	BP	139	5	7.48E-09	5.26E-05
GO:0002699	Positive regulation of immune effector process	BP	142	5	7.52E-09	5.26E-05
GO:0050715	Positive regulation of cytokine secretion	BP	91	4	8.69E-08	4.56E-04
GO:0002443	Leukocyte mediated immunity	BP	258	5	1.42E-07	5.95E-04
GO:0002697	Regulation of immune effector process	BP	291	5	4.04E-07	1.09E-03
GO:0042742	Defence response to bacterium	BP	204	4	4.13E-07	1.09E-03
GO:0050778	Positive regulation of immune response	BP	571	6	4.17E-07	1.09E-03
GO:0050707	Regulation of cytokine secretion	BP	135	4	5.96E-07	1.39E-03
GO:0050663	Cytokine secretion	BP	154	4	9.84E-07	2.06E-03
GO:0001819	Positive regulation of cytokine production	BP	360	5	1.32E-06	2.52E-03
GO:0002876	Positive regulation of chronic inflammatory Response to antigenic stimulus	BP	2	2	1.61E-06	2.81E-03
GO:0050776	Regulation of immune response	BP	769	6	1.82E-06	2.82E-03
GO:0002682	Regulation of immune system process	BP	1191	7	1.88E-06	2.82E-03
GO:0002874	Regulation of chronic inflammatory response to antigenic stimulus	BP	3	2	2.15E-06	3.00E-03
GO:0050830	Defence response to Gram-positive bacterium	BP	68	3	2.45E-06	3.22E-03
GO:0002925	Positive regulation of humoral immune response mediated by circulating immunoglobulin	BP	4	2	2.82E-06	3.48E-03
GO:0002678	Positive regulation of chronic inflammatory response	BP	4	2	3.23E-06	3.76E-03
GO:0002718	Regulation of cytokine production involved in immune response	BP	52	3	3.84E-06	4.24E-03
GO:0002684	Positive regulation of immune system process	BP	820	6	4.24E-06	4.44E-03

intrauterine infection is a known risk factor for CP and that many inflammatory cytokines have been shown to be critical to the risks associated with CP [38, 41, 60].

Perinatal brain injury can be induced by a range of insults such as hypoxic-ischaemic injury or infection [77]. An in utero infection such as chorioamnionitis may trigger an innate immune response, resulting in elevated cytokine levels. Cytokines in the fetal blood may contribute to a systemic fetal inflammatory response with eventual penetration across the blood-brain barrier resulting in an inflammatory cascade in the fetal brain [78]. Brain injury induced by neonatal hypoxia-ischaemia also involves key components of inflammation such as immune cells, chemokines, cytokines and cell adhesion molecules [79]. Therefore, we suggest that inflammation may play a role in perinatal brain damage associated with CP.

We also observed an enrichment of CP-associated DMPs and DMRs in gene ontologies associated with cell adhesion, and this was also observed in individual twin pairs. Aberrant expression of cell adhesion molecules has been reported in muscle biopsies of both children and adults with CP [39]. Previous whole exome and whole genome sequencing studies have also illustrated the potential role of cell adhesion in CP by identifying genetic variants in novel candidate genes which function as neural adhesion molecules essential for neurite outgrowth and axon guidance [9, 12]. The NF-κB transcription factor signalling pathway was common in both DMPs and DMRs.

Our results are consistent with a previous gene expression study in newborn blood spot samples from children with CP [80], which identified up-regulation of inflammatory pathways in preterm children who later developed the disorder. Other similarities between the two studies include variation in genes involved in T-cell and B-cell receptor signalling pathways and cytokine-cytokine receptor interaction, all of which were shown to have a dysregulation in CP cases [80]. However, we found no evidence for an association with increased thyroid function in preterm-born CP cases as hypothesised and reported previously [80].

Another top-ranked DMP was *HNRNPL*, which likely plays a role in response to hypoxia via regulation of the vascular endothelial growth factor (*VEGF*) gene [81]. Hypoxia is known to hinder normal development and maturation of the brain and can cause white matter injury in preterm born infants [82] resulting in CP [83]. This finding supports our hypothesis that epigenetic alterations in genes involved in hypoxic pathways play a role in the aetiology of CP.

Table 6 Genes/probes with an absolute methylation difference of > 50% common to multiple twin pairs

Twin pair group	Methylation difference	Common genes
TW7	0.784	*MED27* (*cg26228569*)
TW15	− 0.749	
TW5	0.824	*BICD2* (*cg14341177*)
TW6	− 0.796	
TW7	0.705	
TW1	0.534	*HLA-DPB2* (*cg01309395*)
TW2	− 0.627	
TW11	0.693	
TW10	0.585	
TW2	− 0.768	*CNOT6L* (*cg11671265*)
TW4	0.745	
TW2	0.634	*SIM1* (*cg00736459*)
TW4	0.556	
TW2	− 0.752	*NR2C2* (*ch.3.343413R*)
TW14	0.584	
TW2	0.753	*LRP11* (*cg24761195*)
TW16	0.811	
TW4	− 0.554	*WWTR1* (*cg02134705*)
TW8	− 0.773	*WWTR1* (*cg19547293*)
TW8	− 0.746	*HYAL3;NAT6* (*cg13682223*)
TW14	0.872	
TW5	− 0.707	
TW9	0.797	*RPTOR* (*cg08905415*)
TW14	− 0.748	
TW4	0.615	*PIK3CG* (*cg08779777*)
TW16	− 0.505	
TW10	− 0.678	

Two high-ranking DMPs lie within the *TBC1D24* and *FBOX9* genes respectively, and both have previously been associated with epilepsy [84, 85]. In 29% of CP cases in Victoria, Australia, epilepsy is comorbid with CP [5]. These results may suggest a potentially shared aetiology between epilepsy and CP [86, 87].

Our findings agree with those of others showing a link between early life DNA methylation state and neurodevelopmental and cognitive outcomes [29, 88], which would allow for early diagnosis and facilitate timely intervention. Currently, MRI scans, assessment tests such as the General Movements Assessment and interventions such as environmental enrichment, early developmental, early motor and physiotherapy interventions are used to inform strategies for early intervention in high-risk groups, such as preterm born children, [89–93].

Given that CP is a highly heterogeneous condition, this study highlights the importance of using epigenetic biomarkers to distinguish and detect underlying pathways across the disorder. For individuals, such an epigenetic state at birth could be used to estimate risk for subsequent development of overt CP.

The strength of using MZ twins is that they are matched for parental age, age, sex, season of birth and genetic factors. Although twins have a higher risk of CP than singletons, the causative mechanisms, such as thrombosis, and infection in the mother, the placenta or the umbilical cords are likely to be similar [7]. Studying twins discordant for CP allows genetic and environmental components to be partitioned from each other and provides a unique opportunity to evaluate the importance of non-shared environmental factors such as umbilical cord or placental function during early development in isolation. It is possible that only twin of a pair may develop an infection or inflammation of the umbilical cord or placenta ([13, 19, 20, 94]). Such non-shared environmental factors are known to influence the development of individual members of a twin pair. This pilot study also highlights the importance of analysing DNA methylation in dried blood spots, which are collected at birth and stored by many countries and the potential for developing future predictive diagnostic tests [55, 95, 96].

Although each tissue has a subset of CpGs whose DNA methylation patterns are tissue-specific, DNA methylation changes that are concordant between the blood and brain have been detected in previous studies [32, 97, 98]. One example is where they identified parallel changes in DNA methylation between the brain and blood in 30% of the genes implicated in Parkinson's disease [98]. It was also shown that a DNA methylation module exists in key ageing-related regulatory genes both in the brain and blood [99]. In addition, animal studies have reported that an early environment resulting in a brain disorder can alter DNA methylation in the same gene across the brain and peripheral tissues [35].

There are some limitations to this study. While similar in sample size to many comparable twin studies of brain-related disorders [58, 59, 64, 100], we acknowledge that larger sample sizes of 25 twin pairs or more are preferable to detect a mean effect size of at least 8% methylation (FDR = 0.05) [101]. As CP is a heterogeneous condition, the small sample size of our cohort, with a lack of CP concordant and healthy twin pairs for comparison, also limits our capability to understand the biological mechanisms of brain injury that may be specifically associated with CP subtype. The use of peripheral tissue and blood also limits our capability to pinpoint the mechanism of CP. Despite the fact that the brain and blood arise from separate cell lineages, and are thought to be epigenetically distinct, many epigenetic studies are often conducted in the blood due to ease of availability [102]. Previous investigations of methylomic variation across the blood and brain tissue from different

regions of the brain have found distinct differences in gene expression and DNA methylation patterns [26, 103–105]. Studies have also shown the inconsistencies in DNA methylation markers from the blood in predicting brain DNA methylation status [27, 106]. However, evidence from animal studies have also shown that blood DNA methylation patterns may in fact reflect patterns in the brain in a subset of genes [35, 107], suggesting that peripheral epigenetic marks may reflect disease mechanisms in some cases. Examples where methylation levels correlate between blood and brain have been reported in Parkinson's disease, depression, schizophrenia, bipolar disorder and autism [107, 108]. The blood is also particularly useful in investigating disease biomarkers [98] and is an important peripheral tissue to consider for neurological disorders, as it is easily accessible to assist in diagnosis. Another limitation is that our data apply to twins only, and we cannot yet generalise our findings more broadly, as there is evidence that risk factors and associated mechanisms leading to CP may be different in twins compared to singletons [109]. To overcome this, our analysis will be repeated in further sets of twins and singletons. Only with this information can we then start to put together risk models for predicting CP at the time of birth. This approach will provide a unique opportunity to identify a biomarker to predict neurodevelopmental outcomes such as CP.

Conclusion

This study provides the first evidence that environment-mediated differential methylation in genes involved in known processes such as hypoxia and inflammation, and perhaps processes such as cell adhesion, may contribute to the development of CP. Our data also pave the way for larger studies to use DNA methylation data in risk models to help predict CP before the onset of overt symptoms and therefore provide a chance for timely ameliorative interventions.

Additional file

Additional file 1: Primer sequences used in site-specific validation using MassArray EpiTYPER. (XLSX 8 kb)

Additional file 2: Scatter plots of genome-wide DNA methylation discordance within twin groups. (PPTX 331 kb)

Additional file 3: Heat map of the associations between the six largest principal components and specified covariates. The heat map provides a score of the strength of the association between DNA methylation (using *M* values) and each covariate, with positive and negative correlations ranging according to the magnitude (red positive, blue negative). The values in brackets for each association represent the *p*-value of the correlation. Of the six significant ($p < 0.05$) associations, all are weak (correlation < 0.6). Abbreviations: CP, cerebral palsy; PC, principal component; PIC, person in charge of performing DNA extraction; GA, gestational age; GMFCS, gross motor function classification system; Guthrie age, age in postnatal days when Guthrie card was made. (PDF 53 kb)

Additional file 4: MDS plots for preprocessed data. Samples are coloured based on chip location ranging from 1 to 3. The figure represents similarities between samples' 1000 most variable probes based on Euclidean distance (sum of squared differences). Dimension 1 represents the largest variation in the dataset, and 2 and 3 are the second and third largest, respectively. (PDF 42 kb)

Additional file 5: Comparison of cell type composition of cerebral palsy cases versus normal individuals. CD8T and CD4T: cytotoxic T cells; NK: natural-killer cells; B cell: B cell or B lymphocytes; Mono: monocytes; Gran: granulocytes. (PDF 87 kb)

Additional file 6: Gene ontology (GO) analysis for top-ranked 1000 DMPs ranked by *p* value. (XLSX 20 kb)

Additional file 7: Gene ontology (GO) analysis for top DMRs ranked by p-value (XLSX 38 kb)

Additional file 8: Cross-platform validation of the two top DMRs, *LTA* and *LIME1*, between HM450 and EpiTYPER platforms. Pearson's correlation coefficients for each probe are shown. The scale of both axes reflects a methylation value between 0 and 1 (β). The regression lines are shown in black. Based on the *r* value (correlation coefficient), correlations across both platforms are shown. The *p*-value indicates the significance of the correlation. (ZIP 126 kb)

Additional file 9: DNA methylation differences within each discordant CP twin pair, identifying numerous loci showing large DNA methylation differences within each discordant twin pair. (PPTX 3104 kb)

Additional file 10: Gene ontology (biological process) common to multiple twin pairs. (XLSX 21 kb)

Additional file 11: CpG sites (probes) within each twin pair group with an absolute methylation difference > 0.5 and their corresponding genes. Genes are colour coded to highlight overlaps between twin pair groups. (XLSX 30 kb)

Abbreviations

CP: Cerebral palsy; DMP: Differentially methylated probe; DMR: Differentially methylated region; EWAS: Epigenome-wide association study; HM450: Human Methylation 450 BeadChip array; MZ: Monozygotic; RUV: Remove unwanted variation; SNP: Single nucleotide polymorphism; SWAN: Subset-quantile within array normalisation

Acknowledgements

We thank the Victorian Cerebral Palsy Register staff for their help with the consent and recruitment of participants included in this study. The Victorian Cerebral Palsy Register receives funding from the Victorian Department of Health and Human Services. We thank Braydon Meyer, MCRI, for developing a Microsoft Excel macro to clean MassArray EpiTYPER data and Eric Joo from the Department of Pathology, University of Melbourne for the bisulphite conversion of genomic DNA and generation of methylation profiles using the HM450 array. We thank Alicia Oshlack from the Murdoch Children's Research Institute for her support and feedback with the data analysis and interpretation of the data. We thank Richard Saffery from the Murdoch Children's Research Institute for his feedback and comments on the manuscript. SMR received support through an Australian National Health and Medical Research Foundation Early Career Fellowship. The authors are also grateful for the support from the Murdoch Children's Research Institute (MCRI), which is supported in part by the Victorian Government's Operational Infrastructure Support Program.

Funding

This work was supported by a grant from the Financial Markets Foundation for Children (grant number 2013-207).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on request.

Authors' contributions

JMC, KC, DR, JW and JM conceived and designed the study. KC, SMR and DJA were involved with the ethics, patient recruitment and sample collections. YJL and SBS performed the lab work required for this study. SBS, NM, JM and DJA implemented the analysis and interpretation. SBS, NM and

JM wrote the code and performed the data analysis. NM, SBS and JMC wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
This study was approved by The Royal Children's Hospital Human Research Ethics Committee (project ID: 33050) and involved written informed consent from each participating family.

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Environmental and Genetic Epidemiology Research, Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia. [2]Department of Paediatrics, The University of Melbourne, Flemington Road, Parkville, Victoria 3052, Australia. [3]Bioinformatics Group, Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia. [4]Developmental Disability and Rehabilitation Research, Murdoch Children's Research Institute, Flemington Road, Parkville, Victoria 3052, Australia. [5]Neurodevelopment and Disability, The Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia. [6]Centre for Molecular and Medical Research, School of Medicine, Deakin University, Geelong, Victoria 3220, Australia.

Received: 10 October 2017 Accepted: 12 February 2018
Published online: 23 February 2018

References

1. Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M. A report: the definition and classification of cerebral palsy April 2006. Dev Med Child Neurol Suppl. 2007;109:8–14.
2. Morris C. Definition and classification of cerebral palsy: a historical perspective. Dev Med Child Neurol Suppl. 2007;109:3–7.
3. Novak I, Hines M, Goldsmith S, Barclay R. Clinical prognostic messages from a systematic review on cerebral palsy. Pediatrics. 2012;130:e1285–312.
4. Oskoui M, Coutinho F, Dykeman J, Jette N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. Dev Med Child Neurol. 2013;55:509–19.
5. Reid SM, Meehan E, McIntyre S, Goldsmith S, Badawi N, Reddihough DS, Australian Cerebral Palsy Register G. Temporal trends in cerebral palsy by impairment severity and birth gestation. Dev Med Child Neurol. 2016; 58(Suppl 2):25–35.
6. Smithers-Sheedy H, McIntyre S, Gibson C, Meehan E, Scott H, Goldsmith S, Watson L, Badawi N, Walker K, Novak I, et al. A special supplement: findings from the Australian Cerebral Palsy Register, birth years 1993 to 2006. Dev Med Child Neurol. 2016;58(Suppl 2):5–10.
7. Pharoah PO, Dundar Y. Monozygotic twinning, cerebral palsy and congenital anomalies. Hum Reprod Update. 2009;15:639–48.
8. ACPR group. Report of the Australian Cerebral Palsy Register, birth years 1993-2009. Sydney: Cerebral Palsy Alliance Research Institute; 2016.
9. McMichael G, Bainbridge MN, Haan E, Corbett M, Gardner A, Thompson S, van Bon BW, van Eyk CL, Broadbent J, Reynolds C, et al. Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. Mol Psychiatry. 2015;20:176–82.
10. Keogh JM, Badawi N. The origins of cerebral palsy. Curr Opin Neurol. 2006; 19:129–34.
11. O'Callaghan ME, MacLennan AH, Haan EA, Dekker G, South Australian Cerebral Palsy Research G. The genomic basis of cerebral palsy: a HuGE systematic literature review. Hum Genet. 2009;126:149–72.
12. Oskoui M, Gazzellone MJ, Thiruvahindrapuram B, Zarrei M, Andersen J, Wei J, Wang Z, Wintle RF, Marshall CR, Cohn RD, et al. Clinically relevant copy number variations detected in cerebral palsy. Nat Commun. 2015;6:7949.
13. Phung DT, Blickstein I, Goldman RD, Machin GA, LoSasso RD, Keith LG. The northwestern twin Chorionicity study I. Discordant inflammatory findings that are related to chorionicity in presenting versus nonpresenting twins. Am J Obstet Gynecol. 2002;186:1041–5.
14. Livinec F, Ancel PY, Marret S, Arnaud C, Fresson J, Pierrat V, Roze JC, Escande B, Thiriez G, Larroque B, et al. Prenatal risk factors for cerebral palsy in very preterm singletons and twins. Obstet Gynecol. 2005;105:1341–7.
15. Benirschke K. The biology of the twinning process: how placentation influences outcome. Semin Perinatol. 1995;19:342–50.
16. Vinnars MT, Vollmer B, Nasiell J, Papadogiannakis N, Westgren M. Association between cerebral palsy and microscopically verified placental infarction in extremely preterm infants. Acta Obstet Gynecol Scand. 2015;94:976–82.
17. Stromswold K. Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. Cognition. 2006;101:333–84.
18. Plomin R: Commentary: why are children in the same family so different? Non-shared environment three decades later. 2011.
19. Dickinson JE, Keil AD, Charles AK. Discordant fetal infection for parvovirus B19 in a dichorionic twin pregnancy. Twin Res Hum Genet. 2006;9:456–9.
20. Kamran Yusuf HJK: The fetus, not the mother, elicits maternal immunologic rejection: lessons from discordant dizygotic twin placentas. 2008.
21. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet. 2002;3:872–82.
22. van Dongen J, Slagboom PE, Draisma HH, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. Nat Rev Genet. 2012;13:640–53.
23. Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. Lancet. 1989;2:577–80.
24. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. Lancet. 1986;1:1077–81.
25. Barker DJ, Gluckman PD, Godfrey KM, Harding JE, Owens JA, Robinson JS. Fetal nutrition and cardiovascular disease in adult life. Lancet. 1993;341:938–41.
26. Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, Coarfa C, Harris RA, Milosavljevic A, Troakes C, et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol. 2012;13:R43.
27. Walton E, Hass J, Liu J, Roffman JL, Bernardoni F, Roessner V, Kirsch M, Schackert G, Calhoun V, Ehrlich S. Correspondence of DNA methylation between blood and brain tissue and its application to schizophrenia research. Schizophr Bull. 2016;42:406–14.
28. Tylee DS, Kawaguchi DM, Glatt SJ. On the outside, looking in: a review and evaluation of the comparability of blood and brain "-omes". Am J Med Genet B Neuropsychiatr Genet. 2013;162B:595–603.
29. Lillycrop KA, Costello PM, Teh AL, Murray RJ, Clarke-Harris R, Barton SJ, Garratt ES, Ngo S, Sheppard AM, Wong J, et al. Association between perinatal methylation of the neuronal differentiation regulator HES1 and later childhood neurocognitive function and behaviour. Int J Epidemiol. 2015;44:1263–76.
30. Hodyl NA, Roberts CT, Bianco-Miotto T. Cord blood DNA methylation biomarkers for predicting neurodevelopmental outcomes. Genes (Basel). 2016;7(12):117.
31. Nemoda Z, Massart R, Suderman M, Hallett M, Li T, Coote M, Cody N, Sun ZS, Soares CN, Turecki G, et al. Maternal depression is associated with DNA methylation changes in cord blood T lymphocytes and adult hippocampi. Transl Psychiatry. 2015;5(4):e545.
32. Teroganova N, Girshkin L, Suter CM, Green MJ. DNA methylation in peripheral tissue of schizophrenia and bipolar disorder: a systematic review. BMC Genet. 2016;17:27.
33. Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, Kalidindi S, Picchioni M, Kravariti E, Toulopoulou T, et al. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum Mol Genet. 2011;20:4786–96.
34. Di Francesco A, Arosio B, Falconi A, Micioni Di Bonaventura MV, Karimi M, Mari D, Casati M, Maccarrone M, D'Addario C. Global changes in DNA methylation in Alzheimer's disease peripheral blood mononuclear cells. Brain Behav Immun. 2015;45:139–44.
35. Aberg KA, Xie LY, McClay JL, Nerella S, Vunck S, Snider S, Beardsley PM, van den Oord EJCG. Testing two models describing how methylome-wide studies in blood are informative for psychiatric conditions. Epigenomics. 2013;5:367–77.
36. Gordon L, Joo JHE, Andronikos R, Ollikainen M, Wallace EM, Umstad MP, Permezel M, Oshlack A, Morley R, Carlin JB, et al. Expression discordance of

monozygotic twins at birth effect of intrauterine environment and a possible mechanism for fetal programming. Epigenetics. 2011;6:579–92.
37. Gordon L, Joo JE, Powel JE, Ollikainen M, Novakovic B, Li X, Andronikos R, Cruickshank MN, Conneely KN, Smith AK, et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. Genome Res. 2012;22:1395–406.
38. Kadhim H, SÉBire G. Immune mechanisms in the pathogenesis of cerebral palsy: implication of proinflammatory cytokines and T lymphocytes. Eur J Paediatr Neurol. 2002;6:139–42.
39. Marbini A, Ferrari A, Cioni G, Bellanova MF, Fusco C, Gemignani F. Immunohistochemical study of muscle biopsy in children with cerebral palsy. Brain Dev. 2002;24:63–6.
40. Girard S, Kadhim H, Roy M, Lavoie K, Brochu ME, Larouche A, Sebire G. Role of perinatal inflammation in cerebral palsy. Pediatr Neurol. 2009;40:168–74.
41. Fleiss B, Gressens P. Tertiary mechanisms of brain damage: a new hope for treatment of cerebral palsy? Lancet Neurol. 2012;11:556–66.
42. Kang M, Min K, Jang J, Kim SC, Kang MS, Jang SJ, Lee JY, Kim SH, Kim MK, An SA, Kim M. Involvement of immune responses in the efficacy of cord blood cell therapy for cerebral palsy. Stem Cells Dev. 2015;24:2259–68.
43. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.
44. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203–9.
45. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.
46. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15(2): R31.
47. Maksimovic J, Gordon L, Oshlack A. SWAN: subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol. 2012;13:R44.
48. Maksimovic J, Gagnon-Bartsch JA, Speed TP, Oshlack A. Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. Nucleic Acids Res. 2015;43:e106.
49. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32:286–8.
50. Hochberg YBaY: Controlling the false discovery rate: a practical and powerful approach to multiple testing. 1995.
51. Murphy TM, Wong CC, Arseneault L, Burrage J, Macdonald R, Hannon E, Fisher HL, Ambler A, Moffitt TE, Caspi A, Mill J. Methylomic markers of persistent childhood asthma: a longitudinal study of asthma-discordant monozygotic twins. Clin Epigenetics. 2015;7:130.
52. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol. 2012;41:200–9.
53. Suchiman HE, Slieker RC, Kremer D, Slagboom PE, Heijmans BT, Tobi EW. Design, measurement and processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. Front Genet. 2015;6:287.
54. Ollikainen M, Smith KR, Joo EJ, Ng HK, Andronikos R, Novakovic B, Abdul Aziz NK, Carlin JB, Morley R, Saffery R, Craig JM. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum Mol Genet. 2010;19:4176–88.
55. Cruickshank MN, Oshlack A, Theda C, Davis PG, Martino D, Sheehan P, Dai Y, Saffery R, Doyle LW, Craig JM. Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. Genome Med. 2013;5(10):96.
56. Lemire M, Zaidi SH, Zanke BW, Gallinger S, Hudson TJ, Cleary SP. The effect of 5-fluorouracil/leucovorin chemotherapy on CpG methylation, or the confounding role of leukocyte heterogeneity: an illustration. Genomics. 2015;106:340–7.
57. Li Yim AYF, Duijvis NW, Zhao J, J de Jonge W, D'Haens G, Mannens M, Mul A, A te Velde A, Henneman P. Peripheral blood methylation profiling of female Crohn's disease patients. Clin Epigenetics. 2016;8:65.
58. Kaut O, Schmitt I, Tost J, Busato F, Liu Y, Hofmann P, Witt SH, Rietschel M, Frohlich H, Wullner U. Epigenome-wide DNA methylation analysis in siblings and monozygotic twins discordant for sporadic Parkinson's disease revealed different epigenetic patterns in peripheral blood mononuclear cells. Neurogenetics. 2016;18:7–22.
59. Wong CC, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, Plomin R, Mill J. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. Mol Psychiatry. 2014;19:495–503.
60. Di H, He Q. The role of inflammatory cytokines in the pathogenesis of cerebral palsy. Gynecol Obstet. 2016;6:360.
61. Huang L, Sherchan P, Wang Y, Reis C, Applegate RL 2nd, Tang J, Zhang JH. Phosphoinositide 3-kinase gamma contributes to neuroinflammation in a rat model of surgical brain injury. J Neurosci. 2015;35:10390–401.
62. Smith LR, Meyer G, Lieber RL. Systems analysis of biological networks in skeletal muscle function. Wiley Interdiscip Rev Syst Biol Med. 2013;5:55–71.
63. Jaarsma D, van den Berg R, Wulf PS, van Erp S, Keijzer N, Schlager MA, de Graaff E, De Zeeuw CI, Pasterkamp RJ, Akhmanova A, Hoogenraad CC. A role for Bicaudal-D2 in radial cerebellar granule cell migration. Nat Commun. 2014;5:3411.
64. Fisher HL, Murphy TM, Arseneault L, Caspi A, Moffitt TE, Viana J, Hannon E, Pidsley R, Burrage J, Dempster EL, et al. Methylomic analysis of monozygotic twins discordant for childhood psychotic symptoms. Epigenetics. 2015;10:1014–23.
65. van Dongen J, Nivard MG, Baselmans BM, Zilhao NR, Ligthart L, Consortium B, Heijmans BT, Bartels M, Boomsma DI. Epigenome-wide association study of aggressive behavior. Twin Res Hum Genet. 2015;18:686–98.
66. Nelson KB, Dambrosia JM, Iovannisci DM, Cheng S, Grether JK, Lammer E. Genetic polymorphisms and cerebral palsy in very preterm infants. Pediatr Res. 2005;57:494–9.
67. Gibson CS, MacLennan AH, Dekker GA, Goldwater PN, Sullivan TR, Munroe DJ, Tsang S, Stewart C, Nelson KB. Candidate genes and cerebral palsy: a population-based study. Pediatrics. 2008;122:1079–85.
68. Yoon BH, Park CW, Chaiworapongsa T. Intrauterine infection and the development of cerebral palsy. BJOG. 2003;110:124–7.
69. Leviton A. Preterm birth and cerebral palsy: is tumor necrosis factor the missing link? Dev Med Child Neurol. 1993;35:553–558.
70. Asselbergs FW, Pai JK, Rexrode KM, Hunter DJ, Rimm EB. Effects of lymphotoxin-alpha gene and galectin-2 gene polymorphisms on inflammatory biomarkers, cellular adhesion molecules and risk of coronary heart disease. Clin Sci (Lond). 2007;112:291–8.
71. Wu J, Li X. Plasma tumor necrosis factor-alpha (TNF-α) levels correlate with disease severity in spastic diplegia, triplegia, and quadriplegia in children with cerebral palsy. Med Sci Monit. 2015;21:3868–74.
72. Hur EM, Son M, Lee OH, Choi YB, Park C, Lee H, Yun Y. LIME, a novel transmembrane adaptor protein, associates with p56lck and mediates T cell activation. J Exp Med. 2003;198:1463–73.
73. Borrego F. The CD300 molecules: an emerging family of regulators of the immune system. Blood. 2013;121:1951–60.
74. Kantola AK, Keski-Oja J, Koli K. Induction of human LTBP-3 promoter activity by TGF-beta1 is mediated by Smad3/4 and AP-1 binding elements. Gene. 2005;363:142–50.
75. Kerstin Krieglstein JS, Schober A, Sullivan A, Unsicker K. TGF-Beta and the regulation of neuron survival and death. J Physiol Paris. 2002;96:25–30.
76. Dobolyi A, Vincze C, Pal G, Lovas G. The neuroprotective functions of transforming growth factor beta proteins. Int J Mol Sci. 2012;13:8219–58.
77. Hagberg H, Edwards AD, Groenendaal F. Perinatal brain damage: the term infant. Neurobiol Dis. 2016;92:102–12.
78. Malaeb S, Dammann O: Fetal inflammatory response and brain injury in the preterm newborn. 2009.
79. Liu F, Mccullough LD: Inflammatory responses in hypoxic ischemic encephalopathy. 2013.
80. Ho NT, Furge K, Fu W, Busik J, Khoo SK, Lu Q, Lenski M, Wirth J, Hurvitz E, Dodge N, et al. Gene expression in archived newborn blood spots distinguishes infants who will later develop cerebral palsy from matched controls. Pediatr Res. 2013;73:450–6.
81. Shih SC, Claffey KP. Regulation of human vascular endothelial growth factor mRNA stability in hypoxia by heterogeneous nuclear ribonucleoprotein L. J Biol Chem. 1999;274:1359–65.
82. Sivakumar V, Ling EA, Lu J, Kaur C. Role of glutamate and its receptors and insulin-like growth factors in hypoxia induced periventricular white matter injury. Glia. 2010;58:507–23.
83. McIntyre S: How low can we go? Recognizing infants at high risk of cerebral palsy earlier. 2015.

84. Suzuki T, Delgado-Escueta AV, Alonso ME, Morita R, Okamura N, Sugimoto Y, Bai D, Medina MT, Bailey JN, Rasmussen A, et al. Mutation analyses of genes on 6p12-p11 in patients with juvenile myoclonic epilepsy. Neurosci Lett. 2006;405:126–31.
85. Afawi Z, Mandelstam S, Korczyn AD, Kivity S, Walid S, Shalata A, Oliver KL, Corbett M, Gecz J, Berkovic SF, Jackson GD. TBC1D24 mutation associated with focal epilepsy, cognitive impairment and a distinctive cerebro-cerebellar malformation. Epilepsy Res. 2013;105:240–4.
86. Ottman R, Annegers JF, Risch N, Hauser WA, Susser M. Relations of genetic and environmental factors in the etiology of epilepsy. Ann Neurol. 1996;39:442–9.
87. Bruck I, Antoniuk SA, Spessatto A, de Bem RS, Hausberger R, Pacheco CG. Epilepsy in children with cerebral palsy. Arq Neuropsiquiatr. 2001;59:35–9.
88. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12:529–41.
89. Baque E, Sakzewski L, Barber L, Boyd RN. Systematic review of physiotherapy interventions to improve gross motor capacity and performance in children and adolescents with an acquired brain injury. Brain Inj. 2016;30:948–59.
90. Degerstedt F, Wiklund M, Enberg B: Physiotherapeutic interventions and physical activity for children in Northern Sweden with cerebral palsy: a register study from equity and gender perspectives. 2017.
91. Morgan C, Novak I, Badawi N. Enriched environments and motor outcomes in cerebral palsy: systematic review and meta-analysis. Pediatrics. 2013;132:e735–46.
92. Spittle A, Orton J, Anderson PJ, Boyd R, Doyle LW: Early developmental intervention programmes provided posthospital discharge to prevent motor and cognitive impairmentin preterm infants. 2015.
93. Morgan C DJ, Gordon AM, Harbourne R, Spittle A, Johnson R, Fetters L: Effectiveness of motor interventions in infants with cerebralpalsy: a systematic review. 2016.
94. Jacques SM, Qureshi F. Chronic villitis of unknown etiology in twin gestations. Pediatr Pathol. 1994;14:575–84.
95. Wong N, Morley R, Saffery R, Craig J. Archived Guthrie blood spots as a novel source for quantitative DNA methylation analysis. BioTechniques. 2008;45:423–4. 6, 8 passim
96. Cruickshank MN, Pitt J, Craig JM. Going back to the future with Guthrie-powered epigenome-wide association studies. Genome Med. 2012;4:83.
97. Wockner LF, Noble EP, Lawford BR, Young RM, Morris CP, Whitehall VL, Voisey J. Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. Transl Psychiatry. 2014;4:e339.
98. Masliah E, Dumaop W, Galasko D, Desplats P. Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes. Epigenetics. 2013;8:1030–8.
99. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. Genome Biol. 2012;13:R97.
100. Dempster EL, Wong CC, Lester KJ, Burrage J, Gregory AM, Mill J, Eley TC. Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. Biol Psychiatry. 2014;76:977–83.
101. Tsai PC, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int J Epidemiol. 2015;44:1429–1441.
102. Smith AK, Kilaru V, Klengel T, Mercer KB, Bradley B, Conneely KN, Ressler KJ, Binder EB. DNA extracted from saliva for methylation studies of psychiatric traits: evidence for tissue specificity and relatedness to brain. Am J Med Genet B Neuropsychiatr Genet. 2015;168:36–44.
103. Rollins B, Martin MV, Morgan L, Vawter MP. Analysis of whole genome biomarker expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet. 2010;153b:919–36.
104. Xin YR, Chanrion B, Liu MM, Galfalvy H, Costa R, Ilievski B, Rosoklija G, Arango V, Dwork AJ, Mann JJ, et al. Genome-wide divergence of DNA methylation marks in cerebral and cerebellar cortices. PLoS One. 2010;5:e11357.
105. Ladd-Acosta C, Pevsner J, Sabunciyan S, Yolken RH, Webster MJ, Dinkins T, Callinan PA, Fan JB, Potash JB, Feinberg AP. DNA methylation signatures within the human brain. Am J Hum Genet. 2007;81:1304–15.
106. Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. Epigenetics. 2015;10:1024–32.
107. Kundakovic M, Gudsnuk K, Herbstman JB, Tang D, Perera FP, Champagne FA. DNA methylation of BDNF as a biomarker of early-life adversity. Proc Natl Acad Sci U S A. 2015;112:6807–13.
108. Pihlstrom L, Berge V, Rengmark A, Toft M. Parkinson's disease correlates with promoter methylation in the alpha-synuclein gene. Mov Disord. 2015;30:577–80.
109. Bonellie SR, Currie D, Chalmers J: Comparison of risk factors for cerebral palsy in twins and singletons. 2005.

Chapter 4 Evidence for type-specific DNA methylation patterns in epilepsy: a discordant monozygotic twin approach.

This chapter comprises entirely of a publication:

Mohandas N, Loke YJ, Hopkins S, Mackenzie L, Bennett C, Berkovic SF, Vadlamudi L, Craig JM. Evidence for type-specific DNA methylation patterns in epilepsy: a discordant monozygotic twin approach. Epigenomics 2019. 11:951-968.

Supplementary material is listed in Appendices 12 - 20 (Page xiii - xxv)

Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Epigenomics

Evidence for type-specific DNA methylation patterns in epilepsy: a discordant monozygotic twin approach

Namitha Mohandas[1,2], Yuk Jing Loke[1], Stephanie Hopkins[1,7], Lisa Mackenzie[3], Carmen Bennett[3], Samuel F Berkovic[5], Lata Vadlamudi[3,4] & Jeffrey M Craig*[,1,2,6]

[1]Environmental & Genetic Epidemiology Research, Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria, Australia

[2]Department of Paediatrics, University of Melbourne, Flemington Road, Parkville, Victoria, Australia

[3]Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Queensland, Australia

[4]Royal Brisbane & Women's Hospital, Queensland, Australia

[5]Epilepsy Research Centre, University of Melbourne, Austin Health, Victoria, Australia

[6]Centre for Molecular & Medical Research, School of Medicine, Deakin University, Geelong, Victoria 3220, Australia

[7]School of Medicine & Public Health, University of Newcastle, Newcastle, New South Wales, Australia

*Author for correspondence: jeffrey.craig@deakin.edu.au

Aim: Epilepsy is a common neurological disorder characterized by recurrent seizures. We performed epigenetic analyses between and within 15 monozygotic (MZ) twin pairs discordant for focal or generalized epilepsy. **Methods:** DNA methylation analysis was performed using Illumina Infinium MethylationEPIC arrays, in blood and buccal samples. **Results:** Differentially methylated regions between epilepsy types associated with *PM20D1* and *GFPT2* genes in both tissues. Within MZ discordant twin pairs, differentially methylated regions associated with *OTX1* and *ARID5B* genes for generalized epilepsy and *TTC39C* and *DLX5* genes for focal epilepsy. **Conclusion:** This is the first epigenome-wide association study, utilizing the discordant MZ co-twin model, to deepen our understanding of the neurobiology of epilepsy.

First draft submitted: 23 August 2018; Accepted for publication: 18 February 2019; Published online: 5 June 2019

Keywords: discordant monozygotic twins • DNA methylation • epigenetics • epilepsy

Epilepsy is a group of common and severe neurological disorders associated with recurrent and unprovoked seizures. It affects over 50 million people worldwide. One in 26 people will develop epilepsy during their lifetime [1], which can have substantial psychological and socioeconomic impacts, impairing quality of life. Only 35% of epilepsy cases have a clear extraneously-acquired cause such as head injury or stroke [2]. The remaining 65% are of unknown cause and were historically termed 'idiopathic', defined as 'no known or suspected etiology other than possible hereditary predisposition' [3]. Epilepsies can be classified into focal, defined as 'originating within networks limited to one cerebral hemisphere'; and generalized as 'originating at some point within and rapidly engaging, bilaterally distributed networks' [4].

The genetic component of epilepsies was clearly recognized in familial and twin studies prior to the genomic era [5,6]. In 1995, the first epilepsy gene was discovered, *CHRNA4* [7], launching the era of discovery of monogenic familial epilepsy syndromes. More recently large-scale molecular genetic studies have led to the identification of a plethora of epilepsy-associated genes with large effect sizes, usually described using the odds ratio, which corresponds roughly to the increase in disease risk. Many genes have exhibited epilepsy type specificity, for example, pathogenic variants in genes such as *SCN1A*, and *GABRG2* are enriched in generalized epilepsy and *DEPDC5, LGI* and *GRIN2A* variants are more prevalent in focal epilepsy [8]. Collectively, these genetic studies suggest that the epilepsy types may have distinct underlying genetic etiology, at least for epilepsies displaying mendelian inheritance.

The genotype-phenotype correlations in epilepsies are complicated, even where major genes are known, and are likely to be more so in the commoner epilepsies where complex inheritance is present [9,10]. Complex inheritance

Future Medicine

implies an interaction of many genetic and environmental factors, and the latter are likely mediated by epigenetic mechanisms [11].

There is growing evidence for the role of epigenetic change (changes in gene expression without DNA sequence change that are heritable through mitosis) in normal and aberrant neurodevelopment, including epileptogenesis and its progression. Research has focused on investigating the role of DNA methylation in temporal lobe epilepsy [12–14], histone modification changes during epileptic seizures [15] and miRNAs in epileptogenic processes [16,17].

Twin studies are ideal for the study of genetic and environmental causes of disease. In particular, the disease discordant monozygotic (MZ) co-twin model controls for shared environment and shared genetics. This model allows us to isolate non-shared environmental factors, which represent the highest component of variance of neurodevelopmental disorders [18,19].

In this exploratory study, we use DNA from two tissues from disease-discordant MZ twins to evaluate DNA methylation profiles in epilepsy. We investigate the differences in DNA methylation patterns between the two common epilepsy types and we further perform two separate epigenetic analyses within MZ twin pairs discordant for focal or generalized epilepsy. To the best of our knowledge, there have been no studies of the epigenetic profile of generalized and focal epilepsies. Understanding these DNA methylation profiles is imperative to understanding their neurobiology as well as for new treatment approaches.

Materials & methods

Samples & DNA extraction

Participants within epilepsy-discordant MZ twin pairs were recruited through Twins Research Australia (https://www.twins.org.au), the Epilepsy Research Centre Database, Melbourne (http://www.epilepsyresearch.org.au), referral from Queensland Neurologists, the Twin Database at the Queensland Institute of Medical Research Berghofer (https://www.qimrberghofer.edu.au/qtwin/) and Epilepsy Queensland (https://www.epilepsyqueensland.com.au). In the cohort, analyses of red cell antigens and human leukocyte antigen markers were performed to confirm zygosity of the older twins pairs, whereas in the younger twin pairs at least 9 different polymorphic DNA marker assay and a 12-marker microsatellite test was performed. If all markers were concordant, the twin pairs were designated MZ. After providing consent, each participant went through a systematic review of medical history, including seizure history and characterization, review of relevant investigations such as electroencephalography and magnetic resonance imaging, and medication history. Both the affected and unaffected twins were interviewed with a validated questionnaire [20]. Blood and buccal biosamples were taken from all participants: epithelial cells from the buccal surface of the cheek were obtained through OraCollect OCR-100 swabs (DNA Genotek, Canada). These kits were mailed to participants with written instructions. Such swabs typically contain 90% buccal epithelial cells [21]. Buccal samples were frozen at -20°C until extraction. Blood samples were obtained at a collection centre local to each individual and stored in 4 ml EDTA tubes. Samples were processed within 24 h to extract the buffy coat layer from blood samples. These samples were then stored in liquid nitrogen until extraction.

Genomic DNA extraction from buccal sample was performed with the NucleoBond CB20 DNA Extraction Kit (Macherey-Nagel; Düren, Germany). For blood samples, genomic DNA was extracted using the Flexigene kit (Qiagen; Hilden, Germany) following manufacturer's protocol. Following genomic DNA extraction from each of the samples, the concentration of extracted DNA was measured by spectrophotometry, and the quality of the extracted samples was determined by gel electrophoresis.

Illumina Infinium EPIC arrays

Following bisulphite conversion of genomic DNA, genome-wide analysis of DNA methylation was assessed using Illumina Infinium EPIC arrays (Illumina, CA, USA) with probes for over 850,000 methylation sites, at the Australian Genome Research Facility (AGRF). Hybridization and scanning was performed following the manufacturer's instructions. Statistical analysis was performed using the R statistical programming language (http://www.R-project.org) in conjunction with Bioconductor packages developed for the analysis of methylation arrays.

Preprocessing of Illumina Infinium EPIC array data

The raw intensity data (IDAT files) were imported into R (3.3.1; http://cran.r-project.org/). Data quality was assessed using the *minfi* (v1.20.2) Bioconductor package [22]. Data from EPIC probes were filtered by removing those based on: poor performance (mean detection p-value of >0.01); probes containing either a single nucleotide polymorphism at the target CpG site or at the single nucleotide extension site; probes that map to multiple locations

in the genome [23]; and probes that map to sex chromosomes (for accurate analysis across all pairs, including males and females). The data were normalized using subset-quantile within array normalization [24]. Covariates such as use of illicit drugs or alcohol, smoking status, presence of febrile seizures, type of anti-epileptic medication, gestational age and birth order of the twins, were assessed as potential confounders.

Differential methylation analysis

Beta (β) values (proportion of the methylated signal over the total signal) were converted to M-values, the $\log_2$ ratio of the intensities of the methylated signal versus the unmethylated signal. Differential methylation analysis was performed using a data-driven method of controlling for unwanted technical and biological variation called Remove Unwanted Variation for differential methylation (RUVm) [25], implemented in the *missMethyl* R package [26] taking twin relationships into account. P-values were adjusted for multiple correction using the Benjamini–Hochberg method [27]. Differentially methylated probes (DMPs) were considered significant if they were lower than the false discovery rate (FDR) threshold of 0.1. We investigated the top-ranked CpGs with an unadjusted p-value less than 1×10^{-4} [28].

Identification of differentially methylated regions

Differentially methylated regions (DMRs) were identified using the *bumphunter* package [22,29]. The cut-off value, which is a user-defined numeric value that determines the upper and lower bounds of the genomic profiles that will be used as candidate regions, was set to 0.05 and the number of permutations set to 250.

Functional annotation & pathway analysis

Gene ontology and pathway analysis were performed using the *gometh* function from the *missMethyl* package [26], which takes into account the variable number of EPIC probes associated with each gene. Gene ontology enrichment was performed for the 1000 top-ranked DMP-associated genes. The KEGG option of the *gometh* function in *missMethyl* was used to provide further insights into relevant biological processes associated with the top-ranked DMPs.

Validation of DMRs

Site-specific validation was attempted using the Sequenom MassArray EpiTYPER (Agena Biosciences). *In silico* assay prediction was performed using the BiocLite *MassArray* package. The source and batch of DNA used for validation was the same as that used for the EPIC analysis. Bisulphite treatment of genomic DNA was performed using the EZ-96 DNA Methylation-Lightning MagPrep kit (Zymo Research; Irvine, USA) according to the manufacturer's instructions. T7 tagged primers were designed for regions (Supplementary Table 1) using the Sequenom EpiDesigner package [30]. PCR amplification was performed on the bisulphite converted DNA samples. Each DNA sample was amplified in triplicate, to ensure that the results were reliable and reproducible. The amplification reaction consisted of 5.3 µl of nuclease free water, 7.5 µl of FastStart Master Mix (Roche, Mannhein, Germany), 0.6 µl of 10 µM forward primer and 0.6 µl of 10 µM reverse primer with thermal cycling conditions as follows: 95°C for 10 min; five cycles of 95°C for 10 s, annealing temperature determined by gradient optimization for 30 s and 72°C for 2 min; 35 cycles of 95°C for 10 s, annealing temperature determined by gradient optimization for 30 s and 72°C for 90 s and final extension at 72°C for 7 min. Each 96 well PCR plate also included three negative controls, where nuclease free water replaced DNA. To reduce the impact of batch effect, all samples for each DMR were amplified in a single PCR plate.

Raw data generated from the MassArray EpiTYPER were cleaned using a Microsoft Excel macro developed in-house [31,32]. The median value of triplicates was determined and any replicates >10% from the median were removed as previously described [31,32]. Pearson's correlation coefficients and p-values were calculated to assess the cross-platform correlation. A correlation coefficient >0.50 was used as the cut-off for a strong correlation, and a p-value ≤0.05 indicated a significant correlation.

Results

Cohort characteristics

Our cohort consisted of 15 MZ twin pairs (ten female and five male) aged between 14 and 67 who were discordant for epilepsy without a known acquired cause (Table 1). The cases were divided into generalized (n = 6) or focal (n = 9) types, based on clinical history and investigations. The focal epilepsies included four with self-limited

Table 1. Descriptive characteristics of the study cohort.

Case	Sex	Current age	Gestational age	Seizure onset (years)	Febrile seizures	Medication history	Family history†	Epilepsy type
1	Female	23	Preterm	8	No	Sodium valproate	Yes	Focal‡
2	Female	49	Preterm	11	No	Carbamazepine	No	Focal§
3	Female	45	Term	9	No	Carbamazepine	No	Focal¶
4	Female	53	Preterm	27	No	Sodium valproate	No	Focal¶
5	Male	67	Term	7	Yes	Dilantin, phenobarbital	Yes	Focal‡
6	Male	14	Term	10	No	Sodium valproate	No	Focal‡
7	Female	29	Term	6	No	Carbamazepine	No	Focal‡
8	Female	65	Term	30	No	Dilantin	Yes	Focal¶
9	Female	67	Preterm	16	Yes	Dilantin	Yes	Focal§
10	Female	64	Preterm	14	No	Sodium valproate	No	Generalized
11	Male	51	Preterm	28	No	Sodium valproate	No	Generalized
12	Male	41	Term	19	No	Sodium valproate	No	Generalized
13	Male	38	Term	7	No	Ethosuximide	No	Generalized
14	Female	48	Preterm	23	No	Sodium valproate	Yes	Generalized
15	Female	55	Preterm	3	Yes	Dilantin	Yes	Generalized

†Family history is defined as at least one relative (up to third degree) previously diagnosed with epilepsy.
Focal epilepsy:
‡Self-limited epilepsy with centrotemporal spikes.
§Self-limited occipital epilepsy.
¶Temporal lobe epilepsy without a structural etiology.

epilepsy with centrotemporal spikes, three with temporal lobe epilepsy without a structural etiology, and two with self-limited occipital epilepsy.

Three cases were reported to have had febrile seizures. Seven affected twins were on sodium valproate (n = 4 generalized group and n = 3 focal group), while all other affected twins were on other anticonvulsant medications such as carbamazepine, phenytoin, ethosuximide and phenobarbitone (Table 1). Five twin pairs reported at least one relative (up to third degree) previously diagnosed with epilepsy.

Within-pair analysis of blood & buccal tissue in all twins

Data exploration using principal component analysis (PCA) and multi-dimensional scaling (MDS) was performed to estimate the variation in the data. Known covariates including age, sex, anticonvulsant medication, alcohol, drugs, birth weight, birth order, gestational age, socioeconomic status (SES) and depression and anxiety stress scale scoring (DASS) were tested to capture similarities and variations between samples of the data and none were shown to be consistently associated with DNA methylation (Figure 1). The cell-type proportions for whole blood samples were calculated using a modified version of the Houseman method [33,34] and no statistically significant differences were observed between the levels of $CD8^+$ T cells (CD8T) and $CD4^+$ T-cells (CD4T), B cells, natural-killer (NK) cells, monocytes and granulocytes between affected and unaffected co-twins (data not shown).

Genome-wide differential methylation analysis of all 15 twin pairs found no statistically significant epilepsy-associated probes with an adjusted FDR value of <0.1 in either tissue. Since this was an exploratory EWAS, we then lowered the threshold of significance to a nominal p-value cut-off of $p < 1 \times 10^{-4}$ as used in previous studies [28,35]. This resulted in a list of 34 top-ranked CpGs in buccal tissue and 21 in whole blood.

To minimise type 1 errors, we focussed on identifying DMRs each identified using multiple proximal probes (more than three probes within a distance of <1 kb) [29] and effect sizes of greater than 5%. This analysis identified six DMRs in buccal samples with multiple contiguous probes with an average DNA methylation difference ranging from 5 to 10%. Analysis of whole blood samples resulted in no statistically significant DMRs, indicating minimal evidence of differential methylation analysis within twin pairs discordant for all epilepsies.

Analysis between generalized & focal epilepsy cases in buccal & blood samples

Next we compared DNA methylation patterns between the two major epilepsy types in buccal and blood samples. CpGs that were differentially methylated between cases of generalized and focal epilepsy in both tissues using RUVm were identified and explored by cluster analysis. There were 62 top-ranked CpGs in buccal and 145 top-

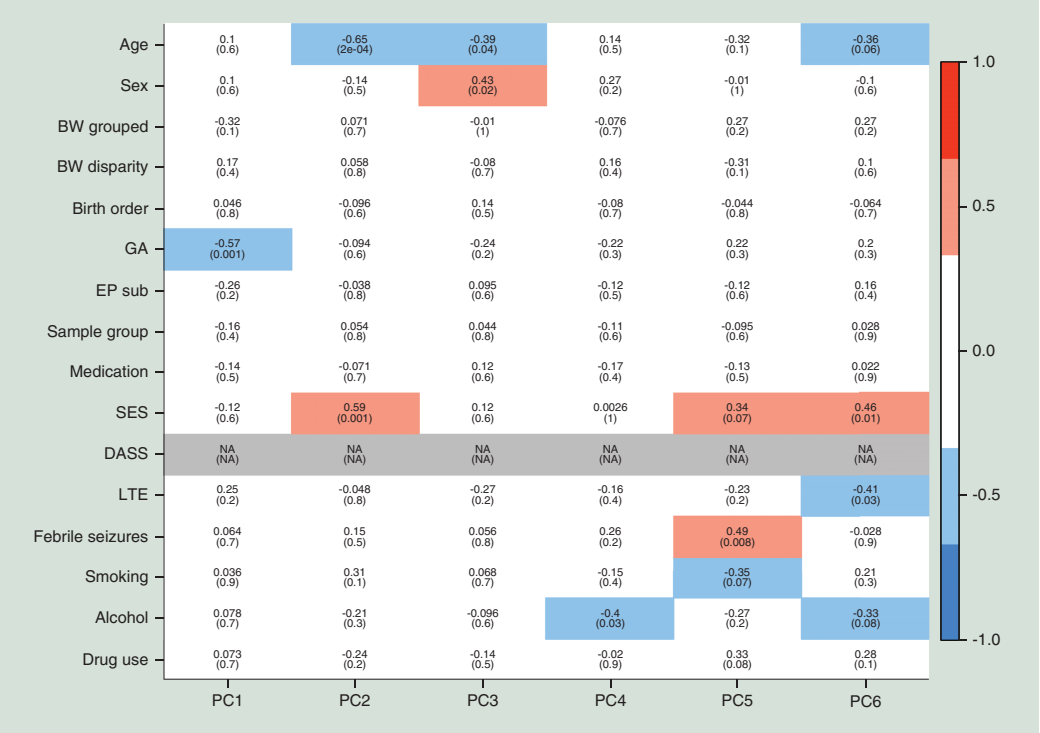


Figure 1. Heat map of the associations between the six largest principal components and specified covariates. The heat map provides a score of the strength of the association between DNA methylation (using M-values) and each covariate, with positive and negative correlations ranging according to magnitude (red positive, blue negative). The values in brackets for each association represent the p-value of the correlation. Of the ten significant ($p < 0.05$) associations, all are weak (correlation <0.65).
BW: Birth weight; DASS: Depression anxiety stress scales and LTE, stress over the past year; EP sub: Epilepsy type; GA: Gestational age; PC: Principal component; SES: Socioeconomic status.

ranked CpGs in blood (nominal p-value $<1 \times 10^{-4}$) (Supplementary Tables 2A & B). A heatmap was generated based on the top 1000 CpG sites ranked by unadjusted p-value after linear regression of epilepsy types within affected individuals (Figure 2). This revealed a distinct differential DNA methylation pattern involving contrasting hypo- and hypermethylation between cases of focal and generalized epilepsy supporting our hypothesis that DNA methylation state in epilepsy types potentially demonstrate specific patterns associated with clinical phenotypes.

We also identified epilepsy type-specific DMRs that ranked high in both sample types (Table 2). The two top-ranked DMRs in both tissues were associated with two genes, *PM20D1* with p-values of 6.8×10^{-5} (buccal) and 1.0×10^{-5} (blood) and *GFPT2* with p-values of 1.35×10^{-5} (buccal) and 1.4×10^{-5} (blood). The average difference in DNA methylation level of *PM20D1* and *GFPT2* was higher in focal epilepsy compared with generalized epilepsy (buccal = 13.9–20%; blood = 18–23%) in both buccal tissue as well as blood. The *PM20D1* DMR located on chromosome 1 spans 653bp with 12 probes likely within the gene promoter region, approximately ~200bp downstream of the transcription start site. The *GFPT2* DMR located on chromosome 5 spans 377bp with four probes within the coding region of the gene.

Given the distinct DNA methylation differences observed between generalized and focal epilepsy, we next searched for epilepsy-specific DNA methylation states within disease-discordant MZ twin pairs, but this time, separately for each epilepsy type.

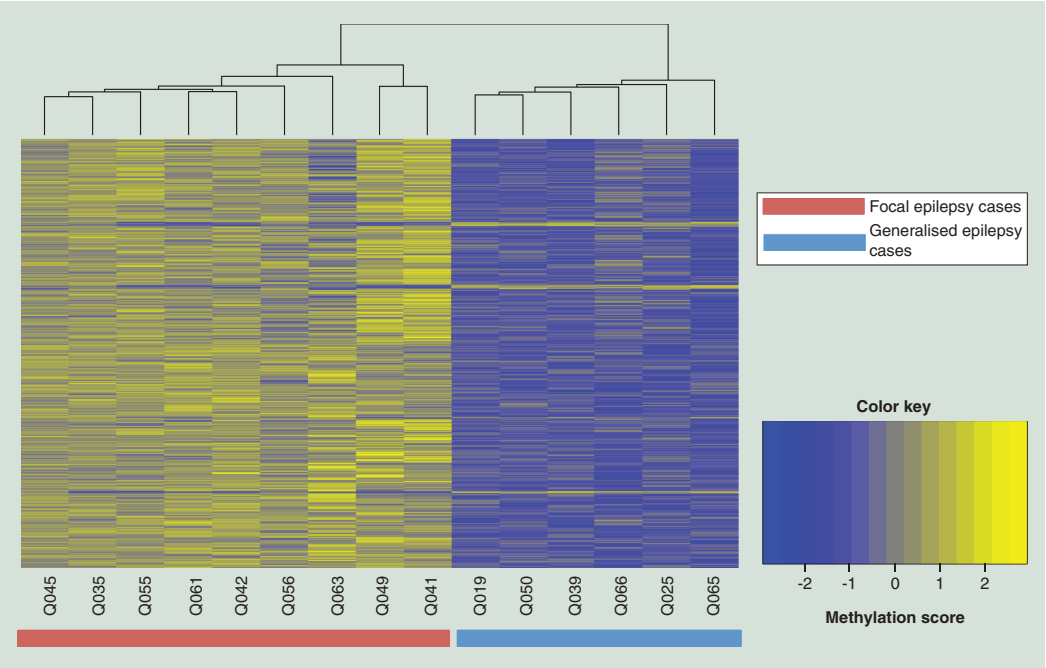


Figure 2. Dendogram highlighting epilepsy type clustering within the top 1000 CpG sites showing contrasting methylation patterns across focal and generalized epilepsy. The top 1000 CpG sites ranked by p-value are represented. Blue to yellow indicates hypomethylation through hypermethylation. The color coded bar indicates the epilepsy types of the affected twin, red: focal epilepsy and blue: generalized epilepsy; numerals on the X-axis indicates affected twins from each twin pair.

Generalized epilepsy: within-twin pair analysis of discordant pairs

Within-pair differences in DNA methylation levels in buccal and blood samples were analysed at individual CpG sites across the genome within six MZ twin pairs discordant for generalized epilepsy. Data were explored for technical and biological variation in the dataset using PCA and MDS plots (see the 'Methods' section), which revealed no consistent association between the covariates tested and DNA methylation (data not shown). EWAS, performed using RUVm, identified no statistically significant DMPs (FDR <0.1). However, 35 DMPs in buccal and seven DMPs in blood were identified at the nominal cut-off of $p < 1.0 \times 10^{-4}$) (Supplementary Tables 2A & B). The average within-pair difference in DNA methylation (case minus control) ranged from +1.2 % to +10.8 % and from -0.5 % to -12.5 %.

We next investigated DMRs using *bumphunter*, and identified regions associated with *ARID5B*, *TAP1* and *TRIM39* genes in blood and within *LY6D*, *OTX1* and *GDNF* genes in buccal tissue. The top-ranked DMR in blood, *ARID5B* (family-wise error rate; fwer = 0.15; p-value = 3.2×10^{-6}), on chromosome 10, consists of 11 CpGs within a distance of 1153bp. Another DMR, ranked second, identified in buccal tissue, *OTX1* (fwer = 0.38; p-value = 1.37×10^{-5}), on chromosome 2, consists of 20 CpG sites within the region (Figure 3). The other top-ranked DMRs are shown in Table 3.

To identify enriched biological processes, we performed gene ontology analysis on genes associated with the top-ranked 1000 DMPs. Gene ontology terms representing biological processes corresponding to 'regulation of cellular metabolic process', 'regulation of tyrosine phosphorylation of Stat5 protein' and 'regulation of spontaneous neurotransmitter secretion' were found (Supplementary Tables 3A & B). The top pathways, as reported by KEGG analysis, included metabolic pathways, were 'TNF signaling', 'dopaminergic synapse', 'neuroactive ligand-receptor interaction' and 'mTOR signaling' (Supplementary Tables 4A & B).

Table 2. The top-ranked differentially methylated regions (ranked on unadjusted p-value) in generalized versus focal cases.

Gene	p-value	fwer	Probes within region (chromosome: location)	Distance between probes (bp)	No of probes in region
Focal vs generalized epilepsy (buccal)					
SLC17A3	3.27E-06	0.072	cg15691649 chr6:25882328		5
			cg03264133 chr6:25882463	135	
			cg01615050 chr6:25882559	96	
			cg03517284 chr6:25882590	31	
			cg24065597 chr6:25882633	43	
PM20D1	6.87E-05	0.368	cg17178900 chr1:205818956		12
			cg26354017 chr1:205819088	132	
			cg14159672 chr1:205819179	91	
			cg14893161 chr1:205819251	72	
			cg07533224 chr1:205819345	94	
			cg12898220 chr1:205819356	11	
			cg05841700 chr1:205819383	27	
			cg11965913 chr1:205819406	23	
			cg07167872 chr1:205819463	57	
			cg24503407 chr1:205819492	29	
			cg16334093 chr1:205819600	108	
			cg07157834 chr1:205819609	9	
GFPT2	1.35E-04	0.484	cg23221052 chr5:179740743		4
			cg13944838 chr5:179740914	171	
			cg23248424 chr5:179741104	190	
			cg02891314 chr5:179741120	16	
Focal vs generalized epilepsy (blood)					
GPR88	8.70E-06	0.22	cg09408571 chr1:101003634		3
			cg06223162 chr1:101003688	54	
			cg07412545 chr1:101003924	236	
PM20D1	1.01E-05	0.232	cg17178900 chr1:205818956		12
			cg26354017 chr1:205819088	132	
			cg14159672 chr1:205819179	91	
			cg14893161 chr1:205819251	72	
			cg07533224 chr1:205819345	94	
			cg12898220 chr1:205819356	11	
			cg05841700 chr1:205819383	27	
			cg11965913 chr1:205819406	23	
			cg07167872 chr1:205819463	57	
			cg24503407 chr1:205819492	29	
			cg16334093 chr1:205819600	108	
			cg07157834 chr1:205819609	9	
GFPT2	1.40E-05	0.32	cg23221052 chr5:179740743		4
			cg13944838 chr5:179740914	171	
			cg23248424 chr5:179741104	190	
			cg02891314 chr5:179741120	16	

Focal epilepsy: within-twin pair analysis of discordant pairs

Exploratory data analysis including PCA and MDS plots similar to our previous analysis were applied, with none of the covariates associated with DNA methylation (data not shown), and regression analysis using RUVm within twin pairs was used to determine focal epilepsy-associated DMPs. We identified 37 DMPs within the nominal p-value threshold of $p < 1 \times 10^{-4}$ in buccal tissue and 22 in blood (Supplementary Tables 2A & B), corresponding

Table 3. The top-ranked differentially methylated regions (ranked on unadjusted p-value) in generalized epilepsy discordant twin pairs.

Gene	p-value	fwer	Probes within region (chromosome: location)	Distance between probes (bp)	No of probes in region
Within twin pair analysis of generalized epilepsy discordant twin pairs (buccal)					
LY6D	6.67E-06	0.288	cg21307747 chr8:143866716		21
			cg25451586 chr8:143866729	13	
			cg05660914 chr8:143866786	57	
			cg17252645 chr8:143867129	343	
			cg20032794 chr8:143867383	254	
			cg05800321 chr8:143867799	416	
			cg24495007 chr8:143867989	190	
			cg07572435 chr8:143868013	24	
			cg08209691 chr8:143868049	36	
			cg04811942 chr8:143868062	13	
			cg09964873 chr8:143868136	74	
			cg14585892 chr8:143868158	22	
			cg20100987 chr8:143868262	104	
			cg26466375 chr8:143868530	268	
			cg02918648 chr8:143868844	314	
			cg17080283 chr8:143868856	12	
			cg02679028 chr8:143868917	61	
			cg23436058 chr8:143869231	314	
			cg26042490 chr8:143869328	97	
			cg01054454 chr8:143869334	6	
			cg14031465 chr8:143869473	139	
			cg05585897 chr8:143869842	369	
OTX1	1.3692E-05	0.384	cg07938743 chr2:63283939		20
			cg21472506 chr2:63283967	28	
			cg23229261 chr2:63284066	99	
			cg10122865 chr2:63284132	66	
			cg25371919 chr2:63284481	349	
			cg19763461 chr2:63284557	76	
			cg22249789 chr2:63284574	17	
			cg22666373 chr2:63284768	194	
			cg22303418 chr2:63284784	16	
			cg17145370 chr2:63285097	313	
			cg05092308 chr2:63285365	268	
			cg19351026 chr2:63285425	60	
			cg13768290 chr2:63285491	66	
			cg04371726 chr2:63285645	154	
			cg18947951 chr2:63285739	94	
			cg07617338 chr2:63285846	107	
			cg27315333 chr2:63285950	104	
			cg11536474 chr2:63286049	99	
			cg25168494 chr2:63286355	306	
			cg10927449 chr2:63286621	266	
GDNF	1.42E-05	0.4	cg07423205 chr5:37834672		20
			cg11396695 chr5:37834848	176	
			cg21590264 chr5:37834850	2	
			cg26473844 chr5:37834909	59	
			cg18725867 chr5:37834958	49	
			cg03266646 chr5:37834999	41	

Table 3. The top-ranked differentially methylated regions (ranked on unadjusted p-value) in generalized epilepsy discordant twin pairs (cont.).

Gene	p-value	fwer	Probes within region (chromosome: location)	Distance between probes (bp)	No of probes in region
			cg24398268 chr5:37835021	22	
			cg08204023 chr5:37835109	88	
			cg18182111 chr5:37835111	2	
			cg20683765 chr5:37835168	57	
			cg26559974 chr5:37835506	338	
			cg00712841 chr5:37835548	42	
			cg19717018 chr5:37835636	88	
			cg23400942 chr5:37835684	48	
			cg17383727 chr5:37835701	17	
			cg19622474 chr5:37835784	83	
			cg15368455 chr5:37835958	174	
			cg02331025 chr5:37835965	7	
			cg08746486 chr5:37835968	3	
			cg00761985 chr5:37835999	31	
Within twin pair analysis of generalized epilepsy discordant twin pairs (blood)					
ARID5B	3.20E-06	0.152	cg02479789 chr10:63807920		11
			cg17633222 chr10:63808249	329	
			cg20728881 chr10:63808314	65	
			cg23468816 chr10:63808748	434	
			cg08783988 chr10:63808750	2	
			cg17973778 chr10:63808852	102	
			cg22846816 chr10:63808857	5	
			cg08241406 chr10:63808875	18	
			cg06456847 chr10:63808883	8	
			cg01969278 chr10:63808948	65	
			cg14789659 chr10:63809073	125	
TAP1	1.34E-05	0.268	chr6:cg26033526		9
			chr6:cg01673307	53	
			chr6:cg24111025	10	
			chr6:cg25042789	43	
			chr6:cg02181920	65	
			chr6:cg06473288	73	
			chr6:cg26234900	112	
			chr6:cg10666909	35	
			chr6:cg08818207	106	
TRIM39	1.38E-05	0.28	cg03094134		9
			cg13079571	83	
			cg09020199	63	
			cg04425551	18	
			cg17080697	44	
			cg07905808	7	
			cg14066298	176	
			cg06230847	29	
			cg01383911	33	

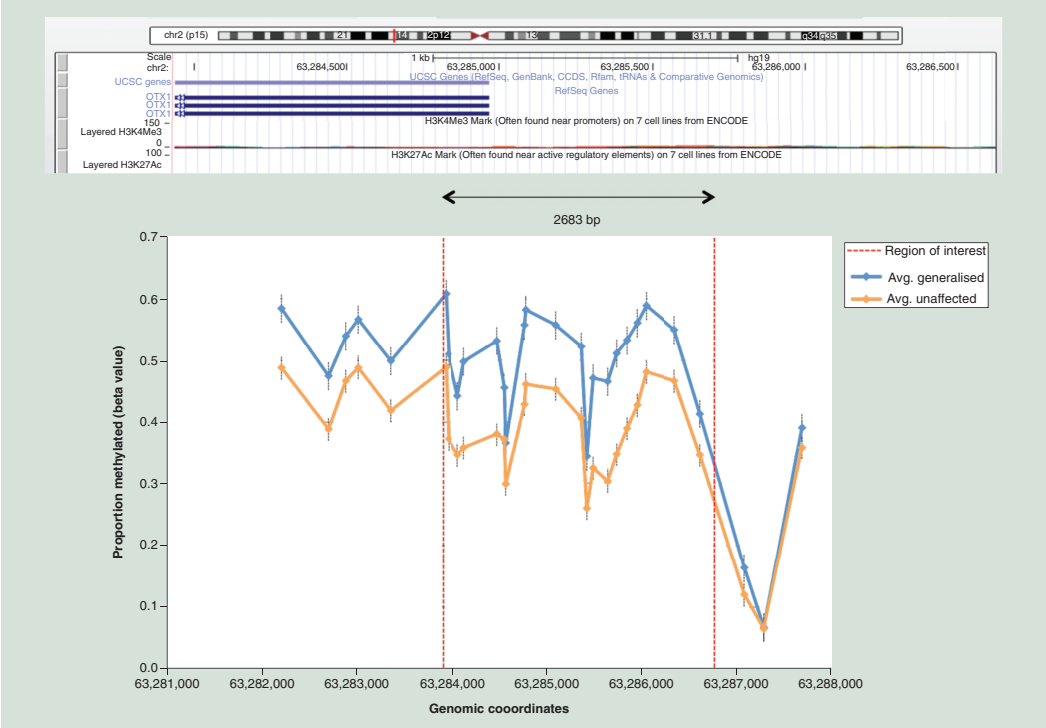


Figure 3. Mean DNA methylation values from buccal samples of discordant monozygotic twin pairs for *OTX1* differentially methylated region. Dotted lines represent error bars with standard errors.

to 31 genes in buccal and 16 in blood. The average within-pair difference in DNA methylation (case minus control) ranged from +0.7 to +6.2% and from -1.1 to -12.8 %.

We next set out to identify DMRs between focal epilepsy affected and unaffected twin pairs. The top three DMRs identified in buccal tissue were all within the *DLX5* gene on chromosome 7 and the DMR, ranked fourth, was within *DLX5* gene on chromosome 14, all spanning from five to seven probes and within 659–1290 base pairs. The average DNA methylation difference ranged from 5.9 to 7.4% (Figure 4). The top-ranked DMRs in whole blood were within genes *ZC3H3*, *TTC39C* and *FGD4*. The *TTC39C* DMR, on chromosome 18, consists of five probes and is located 200bp downstream of the transcription start site and within the promoter of this gene. The average Δβ ranged from 5.9 to 7.8%. The top-ranked DMRs are shown in Table 4.

The top 20 gene ontologies ranked by nominal p-value included biological processes such as 'cerebral cortex tangential migration', 'positive regulation of axon extension involved in axon guidance', 'negative regulation of glial cell apoptotic process' with the top two ontologies related to 'homophilic-cell adhesion via plasma membrane adhesion molecule' and 'cell–cell adhesion via plasma membrane adhesion molecule' in both buccal as well as whole blood tissue analysis (Supplementary Tables 3A & B). Enriched disease pathways included MAPK signaling, PI3K-Akt signaling, axon guidance, TNF signaling pathway and metabolic pathways (Supplementary Tables 4A & B).

Validation of DMRs

Six DMRs were chosen for technical validation with an independent platform, Sequenom MassArray EpiTyper (Supplementary Table 5). *GFPT2* and *PM20D1* were selected from the analysis of generalized and focal epilepsy

Figure 4. Mean DNA methylation values from buccal samples of discordant monozygotic twin pairs for *DLX5* differentially methylated region. Dotted lines represent error bars with standard errors.

cases. One probe-specific CpG was chosen for the *GFPT2* DMR, and three were chosen for *PM20D1* for both buccal and blood samples. All probe-associated CpGs were found to be strongly and significantly ($r = 0.78$ to 0.97; $p < 2.7 \times 10^{-03}$) correlated between platforms (Supplementary Figure 1A & Supplementary Table 5). For the within-twin pair analysis for generalized epilepsy, *ARID5B* was chosen from blood, and *OTX1* from buccal analyses. Five probe-associated CpGs were analysed for *ARID5B*, and two for *OTX1*. As this was a within-twin pair analysis, both beta values and within-twin pair DNA methylation differences were compared. For *ARID5B* and *OTX1*, the beta values were strongly and significantly correlated between platforms ($r = 0.62$ to 0.85, $p = 2.94 \times 10^{-02}$ to 3.81×10^{-04}). However, the within-twin pair differences were found to be strongly but not significantly correlated for most CpGs ($r = 0.53$ to 0.91, $p = 0.05$ to 0.27) (Supplementary Figure 1B & Supplementary Table 5). For the within-twin pair analysis for focal epilepsy, *TTC39C* was selected from blood, and *DLX5* from buccal samples. As in the previous analysis, both absolute and between-group differences in DNA methylation were investigated. *TTC39C* did not meet the criteria for validation, with weak, nonsignificant correlation ($r = 0.1$ to 0.5, $p = 0.04$ to 0.6) between platforms (Supplementary Figure 1). However, results for *DLX5* were strongly, significantly correlated ($r = 0.96$, $p = 1 \times 10^{-05}$) (Supplementary Figure 1).

Discussion

We identified DNA methylation signatures that differentiated focal and generalized epilepsy types and differentiated each epilepsy type compared with their nonepileptic co-twin.

Analysis between generalized & focal epilepsy cases in buccal & blood tissue

The top-ranked DMPs revealed separate clustering of DNA methylation patterns across focal and generalized epilepsy cases. The identification of the *PM20D1* and *GFPT2* DMRs in focal and generalized epilepsy respectively, in both whole blood and buccal tissues, strengthens this observation.

Table 4. The top-ranked differentially methylated regions (ranked on unadjusted p-value) in focal epilepsy discordant twin pairs.

Gene	p-value	fwer	Probes within region (chromosome: location)	Distance between probes (bp)	No of probes in region
Within twin pair analysis of focal epilepsy discordant twin pairs (buccal)					
DLX5	6.11E-04	0.844	cg08835113 chr7:96650192		7
			cg00400832 chr7:96650323	131	
			cg15339231 chr7:96650407	84	
			cg00503840 chr7:96650509	102	
			cg20377305 chr7:96650668	159	
			cg20080624 chr7:96651111	443	
			cg13462129 chr7:96651281	170	
DLX5	6.12E-04	0.844	cg25172279 chr7:96646598		7
			cg25928986 chr7:96647079	481	
			cg01680010 chr7:96647117	38	
			cg00388195 chr7:96647323	206	
			cg04694035 chr7:96647440	117	
			cg26718878 chr7:96647882	442	
			cg08482436 chr7:96647888	6	
DLX5	6.17E-04	0.848	cg27016494 chr7:96651586		7
			cg18873386 chr7:96651915	329	
			cg27032146 chr7:96652115	200	
			cg11500797 chr7:96652123	8	
			cg11891395 chr7:96652153	30	
			cg08101303 chr7:96652222	69	
			cg02101486 chr7:96652245	23	
DLX5	2.21E-03	0.924	cg12049236 chr14:38057893		5
			cg27143688 chr14:38058243	350	
			cg07465387 chr14:38058263	20	
			cg19645221 chr14:38058284	21	
			cg12212453 chr14:38058639	355	
Within twin pair analysis of focal epilepsy discordant twin pairs (blood)					
ZC3H3	2.26E-04	0.612	cg11848483 chr8:144543486		5
			cg04180114 chr8:144543568	82	
			cg20156237 chr8:144543750	182	
			cg11210357 chr8:144543951	201	
			cg24010658 chr8:144544041	90	
TTC39C	3.02E-04	0.66	cg27399348 chr18:21572366		5
			cg12639429 chr18:21572622	256	
			cg05401069 chr18:21572634	12	
			cg18719665 chr18:21572656	22	
			cg17205313 chr18:21572748	92	
FGD4	9.75E-04	0.876	cg07120158 chr12:32638669		4
			cg03418231 chr12:32638773	104	
			cg20470583 chr12:32638830	57	
			cg02251680 chr12:32638839	9	

It is noteworthy that *PM20D1* is one of five genes located within the *PARK16* locus, which is involved in neuronal growth and development. A significant decrease in the expression of *PM20D1*, associated with hypermethylation, has recently been implicated in Alzheimers disease [36], whereas the overexpression of *PM20D1* reduces cell death, decrease β amyloid (Aβ) levels and improves cognitive performance. A subset of four probes within the *PM20D1*

DMR are located within enhancer elements and promoter regions as documented in the Infinium MethylationEPIC annotation data, which suggests a role in gene regulation.

The *PM20D1* DMR has been classified as a metastable epiallele [37], in other words, one which is variably methylated within MZ twin pairs but with consistent DNA methylation levels in more than one tissue from each individual. Typically, epigenetic modifications at metastable epialleles occur stochastically during early embryonic development, affecting all tissues of the body and prominent interindividual variation. Metastable epialleles can indicate the potential importance of environmentally-driven epigenetic changes in the etiology of the diseases such as epilepsy.

The second-ranked DMR-associated gene, *GFPT2*, is involved in neurite outgrowth, early neuronal cell development and neuropeptide signaling/synthesis [38]. This gene was reported to have a splice site mutation detected in a whole exome sequencing study of Rett syndrome, a disorder also associated with seizures [39].

Generalized epilepsy: within-twin pair analysis

Our analysis of MZ twin pairs discordant for generalized epilepsy revealed top-ranked DMRs in genes such as *LY6D*, *OTX1* and *GDNF* in buccal tissue and *ARID5B*, *TAP1* and *TRIM39* in blood. *OTX1* gene is a transcription factor and is fundamental to brain and sensory organ development [40]. A mouse knockdown of *Otx1* demonstrated spontaneous epileptic behaviour, indicating that the encoded protein is required for proper neurological development [41,42]. The third-ranked DMR for generalized epilepsy-discordant twin pairs in buccal tissue, *GDNF*, is another interesting gene that is reported to have seizure-suppressant action in animal models of epilepsy [43].

In blood, the top-ranked DMR, ARID5B, a DNA binding protein, modulates histone methylation and gene transcription by forming a complex with protein kinase A-dependent histone lysine demethylase [44,45]. H3K9 methylation imbalance has been linked to cognitive impairment and to mental retardation, epileptic seizures and autism [46]. The second-ranked DMR-associated gene, *TAP1*, encodes a transporter associated with antigen processing. Polymorphisms of *TAP1* have been found in juvenile myoclonic epilepsy [47]. *TAP1* is also located in close genomic proximity to the bromodomain-containing gene *BRD2*. Multiple linkage and population association studies have connected *BRD2* to forms of genetic generalized epilepsy [47,48].

In comparisons within generalized epilepsy-discordant twin pairs, we observed enrichment in biological processes such as those involved in metabolic and neurotransmitter biosynthetic processes. Previous studies have marked the critical role of metabolic pathways in maintaining neuronal excitability and the impairment of such regulation leading to seizure activity associated with epileptic conditions [49,50].

Focal epilepsy: within-twin pair analysis of discordant pairs

Differential DNA methylation analysis of focal epilepsy-discordant twin pairs revealed top-ranked DMRs associated with genes involving neuronal signaling. Four top-ranking DMRs in buccal tissue were associated to the *DLX5* gene, a transcription factor. *DLX5* is associated with mechanistic target of rapamycin serine/threonine kinase (mTOR) pathway, which regulates the generation of cortical neurons during brain development [51]. The expression of other members of this transcription factor family, in particular, *Dlx1* and *Dlx2* is critical for production of the GABAergic cells in mice [52,53]. Disruptions in neural development can impair the functionality of these regions, leading to the instability of neural networks and ultimately, seizures.

Gene ontology terms associated with focal epilepsy were more likely to play a role in cell-adhesion in both buccal and blood. Recent studies have shown that cell adhesion receptors such as integrins are required for the regulation of GABAergic transmission and can mediate seizure-induced cell death [54,55];). Cell adhesion proteins also play a role in guidance and motor mechanisms of axon growth and synaptic stability, required for neuroplasticity [56,57].

Comparisons to previous DNA methylation studies in epilepsy

We compared our DMRs to previous DNA methylation studies of epilepsy [12–14,58–61] and found no overlap of differentially expressed genes (data not shown). This may be explained by different study designs whereby previous studies examined a specific type of epilepsy, most often temporal lobe epilepsy [12,14,61]; were either performed in animal models [58,59] or focused on epilepsy candidate genes [13,14,60]. However, our pathway analysis within twin pairs discordant for generalized epilepsy revealed similarities such as calcium signaling and neuroactive ligand–receptor interaction that were consistent with those found in DNA methylation analysis of refractory epilepsy [61]. The top-ranking pathways within focal epilepsy discordant twin pairs were enriched in calcium signaling, MAPK

signaling, axon guidance, antigen processing and presentation and endocytosis. These were consistent with pathways enriched in downregulated genes in chronic temporal lobe epileptic hippocampi in rats [58].

Strengths of our study

This is the first study to explore DNA methylation profiles in epilepsies and presents support that generalized and focal epilepsy types differ in their DNA methylation patterns. One of the strengths of this study is the use of multiple easily accessible peripheral tissues. Peripheral tissues such as buccal and blood can optimise the discovery of clinical biomarkers [62,63]. Epigenetic state of accessible tissue is already being used as a clinical biomarker in some disorders [63,64], but not yet in epilepsy. By definition, a biomarker must be from an easily-accessible tissue and need not necessarily resemble the gene expression patterns in the brain. Therefore, using both blood and buccal samples will maximize the potential of finding novel biomarkers [65]. In addition, studying MZ twins discordant for epilepsy highlights the interplay between genetics, shared and non-shared environment leading to phenotypic differences between twins.

Limitations of our study

While the co-twin model is an advantage of this study, it also comes with limitations. MZ co-twins that are carefully assessed for discordance of epilepsy phenotype can be difficult to recruit and screen, which leads to a limitation in the sample size of the cohort. Although similar to sample sizes in many comparable twin studies of brain-related disorders [66–69], we acknowledge that larger sample sizes of 25 twin pairs or more are preferable to detect a mean effect size of at least 8% methylation (FDR = 0.05) [70]. MZ co-twins also are likely to share genetic liability hence, despite their differences in the clinical phenotype, discordant MZ twins can be quite similar in terms of underlying disease susceptibility.

Epigenetic state is known to have a tissue-specific component; however, many recent studies have used accessible peripheral tissues in developing clinical biomarkers for chronic disorders [71–76]. While brain samples remain the gold standard tissue of choice for epigenetic analyses of disease mechanisms in epilepsy, the inaccessibility of brain precludes its feasibility and utility in a clinical biomarker setting. And although there is evidence that peripheral biomarker-associated genes have functions in the brain [62,77], such a relationship is not necessary for a good clinical biomarker.

Whether an epigenetic biomarker represents cause versus effect of a disease remains a challenging subject and longitudinal studies would ideally allow delineation of biomarkers as potential causes of disease (present before disease onset) versus effects of disease (present after disease onset).

Conclusion

To our knowledge this is the first twin epigenome-wide association study to investigate the role of DNA methylation profiles in epilepsy, identifying distinctive methylation clusters that differ between generalized and focal epilepsies. Our data leads the way for larger studies as understanding DNA methylation profiles associated with epilepsy types is central to understanding their neurobiology in the search for new treatment approaches.

Future perspective

This study of MZ twins discordant for epilepsy highlights the interplay between genetic and epigenetic factors. Replication in further twin and singleton cohorts and is a necessary step toward developing peripheral epigenetic biomarkers for epilepsy. However, the specificity of future predictive biomarkers would be enhanced by the analysis of parallel multi-omic data. Such an integrated approach will help identify the underlying biological pathways affected by aberrant DNA methylation in epilepsy. In addition, a deeper understanding of the neurobiology of epilepsy, including epigenetic state, will enhance targeted treatment approaches for this debilitating disorder.

Authors' contributions

L Vadlamudi, SF Berkovic and JM Craig conceived and designed the study. SF Berkovic assisted with twin recruitment (Epilepsy Research Centre). C Bennett, L Vadlamudi, L Mackenzie were involved with ethics, patient recruitment and sample collections. YJ Loke and S Hopkins performed the lab work required for this study. N Mohandas and S Hopkins implemented the analysis and interpretation. N Mohandas wrote the code and performed data analysis. N Mohandas, L Vadlamudi, S Hopkins and JM Craig wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the participants and their families. The authors also wish to thank J Hopper (Twins Research Australia); N Martin (Queensland Institute of Medical Research Berghofer) for their assistance with twin recruitment. The authors thank B Meyer, MCRI, for developing a Microsoft Excel macro to clean MassArray EpiTYPER data. The authors are also grateful for the support from the Murdoch Children's Research Institute (MCRI), which is supported in part by the Victorian Government's Operational Infrastructure Support Program.

Financial & competing interests disclosure

This work was supported by grant from the Financial Markets Foundation for Children (grant number 2013–207). L Vadlamudi was supported by a Queensland Health Research Fellowship. This work was also supported by The University of Queensland Mayne Bequest Funds, Royal Brisbane and Women's Foundation Grant, Viertal-Charitable Foundation, Ramaciotti Establishment Grant and The Servier Barry Young Fellowship in Neurology. The authors have no other relevant affiliations or financialinvolvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical consent disclosure

Informed consent was obtained from each twin individual. A multicentre ethics approval was obtained with the lead site being Mater Health Services (ethics approval number HREC/13/MHS/114). Consent has been obtained from all participants that their de-identified results may be published.

Availability of data & material

The datasets used and/or analysed during the current study are available from the corresponding author on request.

Summary points

- Epilepsy is a common neurological disorder characterized by recurrent seizures and classified into focal and generalized epilepsy types.
- DNA methylation has been implicated as a mediator of early life effects in offspring.
- We measured DNA methylation in two peripheral tissues in 15 monozygotic (MZ) twin pairs discordant for epilepsy using Illumina Infinium EPIC arrays.
- Analysis of all affected epilepsy cases versus unaffected twins found no differentially methylated regions (DMRs).
- Analysis between focal and generalized epilepsy cases showed that each clustered separately using DNA methylation status.
- Analyses within MZ twin pairs discordant for focal epilepsy identified DMRs in *DLX5* (buccal) and *ARID5B* (blood).
- Analyses within MZ twin pairs discordant for generalized epilepsy identified DMRs in *OTX1* (buccal) and *TT39C* (blood).

References

Papers of special note have been highlighted as: • of interest

1. Hesdorffer DC, Logroscino G, Benn EK, Katri N, Cascino G, Hauser WA. Estimating risk for developing epilepsy: a population-based study in Rochester, Minnesota. *Neurology* 76(1), 23–27 (2011).
2. Hauser WA, Annegers JF, Kurland LT. Incidence of epilepsy and unprovoked seizures in rochester, Minnesota – 1935–1984. *Epilepsia* 34(3), 453–468 (1993).
3. Epilepsy COCaTOTILA. Proposal for classification of epilepsies and epileptic syndromes. *Epilepsia* 6(3), 268–278 (1985).
4. Fisher RS, Acevedo C, Arzimanoglou A *et al.* ILAE official report: a practical clinical definition of epilepsy. *Epilepsia* 55(4), 475–482 (2014).
5. Steinlein OK. Genetics and epilepsy. *Dialog. Clin, Neurosci.* 10(1), 29–38 (2008).
6. Vadlamudi L, Milne RL, Lawrence K *et al.* Genetics of epilepsy: the testimony of twins in the molecular era. *Neurology* 83(12), 1042–1048 (2014).
7. Steinlein Ok MJ, Propping P, Wallace RH, Phillips HA, Sutherland GR, Scheffer IE, Berkovic SF. A missense mutation in the neuronal nicotinic acetylcholine receptor alpha 4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy. (1995).
8. Helbig I, Heinzen EL, Mefford HC, Commission IG. Primer Part 1-the building blocks of epilepsy genetics. *Epilepsia* 57(6), 861–868 (2016).

9. Koeleman BPC. Genetics of common forms of epilepsy. *Lancet Neurol.* 16(2), 101–102 (2017).
10. Helbig I, Lowenstein DH. Genetics of the epilepsies: where are we and where are we going? *Curr. Opin. Neurol.* 26(2), 179–185 (2013).

- **Describes the rationale for studying factors other than genetics in epilepsy.**

11. Rutten BP, Mill J. Epigenetic mediation of environmental influences in major psychotic disorders. *Schizophr. Bull.* 35(6), 1045–1056 (2009).
12. Long HY, Feng L, Kang J *et al.* Blood DNA methylation pattern is altered in mesial temporal lobe epilepsy. *Sci. Rep.* 7, 43810 (2017).
13. Miller-Delaney SF, Bryan K, Das S *et al.* Differential DNA methylation profiles of coding and non-coding genes define hippocampal sclerosis in human temporal lobe epilepsy. *Brain* 138(Pt 3), 616–631 (2015).
14. Kobow K JI, Hildebrandt M, Hauke J *et al.* Increased reelin promoter methylation is associated with granule cell dispersion in human temporal lobe epilepsy. *J. Neuropathol. Exp. Neurol.* 68(4), 356–364 (2009).

- **First study to show a potential role of DNA methylation in epilepsy.**

15. Pulido Fontes L, Quesada Jimenez P, Mendioroz Iriarte M. Epigenetics and epilepsy. *Neurología (Engl. Ed.)* 30(2), 111–118 (2015).
16. Liu DZ, Tian Y, Ander BP *et al.* Brain and blood microRNA expression profiling of ischemic stroke, intracerebral hemorrhage, and kainate seizures. *J. Cereb. Blood Flow Metab.* 30(1), 92–101 (2010).
17. Song YJ, Tian XB, Zhang S *et al.* Temporal lobe epilepsy induces differential expression of hippocampal miRNAs including let-7e and miR-23a/b. *Brain Res.* 1387, 134–140 (2011).
18. Rappaport SM. Genetic factors are not the major causes of chronic diseases. *PLoS ONE* 11(4), e0154387 (2016).
19. Van Dongen J, Slagboom PE, Draisma HH, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. *Nat. Rev. Genet.* 13(9), 640–653 (2012).

- **Describes the importance of using unique co-twin model in epigenetic studies.**

20. Reutens DC, Howell RA, Gebert KE, Berkovic SF. Validation of a questionnaire for clinical seizure diagnosis. *Epilepsia* 33(6), 1065–1071 (1992).
21. Theda C, Hwang SH, Czajko A, Loke YJ, Leong P, Craig JM. Quantitation of the cellular content of saliva and buccal swab samples. *Sci. Rep.* 8(1), 6944 (2018).
22. Aryee MJ, Jaffe AE, Corrada-Bravo H *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30(10), 1363–1369 (2014).
23. Chen YA, Lemire M, Choufani S *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 8(2), 203–209 (2013).
24. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* 13(6), R44 (2012).
25. Maksimovic J, Gagnon-Bartsch JA, Speed TP, Oshlack A. Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Res.* 43(16), e106 (2015).
26. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* 32(2), 286–288 (2016).

- **Description of the method used in analyzing DNA methylation.**

27. Hochberg YBaY. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57(1), 289- 300 (1995).
28. Murphy TM, Wong CC, Arseneault L *et al.* Methylomic markers of persistent childhood asthma: a longitudinal study of asthma-discordant monozygotic twins. *Clin. Epigenetics* 7, 130 (2015).
29. Jaffe AE, Murakami P, Lee H *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int. J. Epidemiol.* 41(1), 200–209 (2012).
30. Suchiman HE, Slieker RC, Kremer D, Slagboom PE, Heijmans BT, Tobi EW. Design, measurement and processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. *Front Genet.* 6, 287 (2015).
31. Ollikainen M, Smith KR, Joo EJ *et al.* DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum. Mol. Genet.* 19(21), 4176–4188 (2010).
32. Cruickshank MN, Oshlack A, Theda C *et al.* Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. *Genome Med.* 5(10), 96 (2013).
33. Houseman EA, Accomando WP, Koestler DC *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformat.* 13, 86 (2012).
34. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 15(2), (2014).
35. Van Dongen J, Nivard MG, Baselmans BM *et al.* Epigenome-wide association study of aggressive behavior. *Twin Res. Hum. Genet.* 18(6), 686–698 (2015).

36. Sanchez-Mut JV, Heyn H, Silva BA *et al.* PM20D1 is a quantitative trait locus associated with Alzheimer's disease. *Nat. Med.* 24(5), 598–603 (2018).
37. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. *Trends Genet.* 18(7), 348–351 (2002).
38. Tondreau T, Dejeneffe M, Meuleman N *et al.* Gene expression pattern of functional neuronal cells derived from human bone marrow mesenchymal stromal cells. *BMC Genomics* 9, 166 (2008).
39. Grillo E, Lo Rizzo C, Bianciardi L *et al.* Revealing the complexity of a monogenic disease: rett syndrome exome sequencing. *PLoS One* 8(2), e56599 (2013).
40. Acampora D, Postiglione MP, Avantaggiato V, Di Bonito M, Simeone A. The role of Otx and Otp genes in brain development. *Int. J. Dev. Biol.* 44(6), 669–677 (2000).
41. Acampora D, Mazan S, Avantaggiato V *et al.* Epilepsy and brain abnormalities in mice lacking the Otx1 gene. *Nat. Genet.* 14(2), 218–222 (1996).
42. Avanzini G, Spreafico R, Cipelletti B Acampora D, Simeone A. *et al.* Synaptic properties of neocortical neurons in epileptic mice lacking the Otxl gene. *Epilepsia* Suppl 6, S200- S2005 (2000).
43. Kanter-Schlifke I, Georgievska B, Kirik D, Kokaia M. Seizure suppression by *GDNF* gene therapy in animal models of epilepsy. *Mol. Ther.* 15(6), 1106–1113 (2007).
44. Baba A, Ohtake F, Okuno Y *et al.* PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B. *Nat. Cell Biol.* 13(6), 668–675 (2011).
45. Di Stefano L, Dyson NJ. The emerging roles for histone demethylases in the modulation of signaling pathways. *Biomol. Concepts* 4(1), 13–27 (2013).
46. Black JC, Van Rechem C, Whetstine JR. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell* 48(4), 491–507 (2012).
47. Layouni S, Buresi C, Thomas P, Malafosse A, Dogui M. *BRD2* and *TAP-1* genes and juvenile myoclonic epilepsy. *Neurol. Sci.* 31(1), 53–56 (2010).
48. Velisek L, Shang E, Veliskova J *et al.* GABAergic neuron deficit as an idiopathic generalized epilepsy mechanism: the role of BRD2 haploinsufficiency in juvenile myoclonic epilepsy. *PLoS ONE* 6(8), e23656 (2011).
49. Falkowska A, Gutowska I, Goschorska M, Nowacki P, Chlubek D, Baranowska-Bosiacka I. Energy metabolism of the brain, including the cooperation between astrocytes and neurons, especially in the context of glycogen metabolism. *Int. J. Mol. Sci.* 16(11), 25959–25981 (2015).
50. Vielhaber S, Niessen HG, Debska-Vielhaber G *et al.* Subfield-specific loss of hippocampal N-acetyl aspartate in temporal lobe epilepsy. *Epilepsia* 49(1), 40–50 (2008).
51. Ka M, Smith AL, Kim WY. MTOR controls genesis and autophagy of GABAergic interneurons during brain development. *Autophagy* 13(8), 1348–1363 (2017).
52. Cobos I, Calcagnotto ME, Vilaythong AJ *et al.* Mice lacking Dlx1 show subtype-specific loss of interneurons, reduced inhibition and epilepsy. *Nat. Neurosci.* 8(8), 1059–1068 (2005).
53. Le TN, Zhou QP, Cobos I *et al.* GABAergic interneuron differentiation in the basal forebrain is mediated through direct regulation of glutamic acid decarboxylase isoforms by Dlx homeobox transcription factors. *J. Neurosci.* 37(36), 8816–8829 (2017).
54. Wu X, Reddy DS. Integrins as receptor targets for neurological disorders. *Pharmacol. Ther.* 134(1), 68–81 (2012).
55. Gall CM, Lynch G. Integrins, synaptic plasticity and epileptogenesis. *Adv. Exp. Med. Biol.* 548, 12–33 (2004).
56. Giagtzoglou N, Ly CV, Bellen HJ. Cell adhesion, the backbone of the synapse: "vertebrate" and "invertebrate" perspectives. *Cold Spring Harb. Perspect. Biol.* 1(4), a003079 (2009).
57. Shen K, Cowan CW. Guidance molecules in synapse formation and plasticity. *Cold Spring Harb. Perspect. Biol.* 2(4), a001842 (2010).
58. Kobow K, Kaspi A, Harikrishnan KN *et al.* Deep sequencing reveals increased DNA methylation in chronic rat epilepsy. *Acta Neuropathol.* 126(5), 741–756 (2013).
59. Miller-Delaney SF, Das S, Sano T *et al.* Differential DNA methylation patterns define status epilepticus and epileptic tolerance. *J. Neurosci.* 32(5), 1577–1588 (2012).
60. Belhedi N, Perroud N, Karege F, Vessaz M, Malafosse A, Salzmann A. Increased CPA6 promoter methylation in focal epilepsy and in febrile seizures. *Epilepsy Res.* 108(1), 144–148 (2014).
61. Liu X, Ou S, Xu T *et al.* New differentially expressed genes and differential DNA methylation underlying refractory epilepsy. *Oncotarget* 7(52), 87402–87416 (2016).
62. Masliah E, Dumaop W, Galasko D, Desplats P. Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes. *Epigenetics* 8(10), 1030–1038 (2013).
63. Mikeska T, Craig JM. DNA methylation biomarkers: cancer and beyond. *Genes (Basel)* 5(3), 821–864 (2014).

64. Sabine as Langie MM, Ken Declerck, Gudrun Koppen *et al.* Salivary DNA methylation profiling: aspects to consider for biomarker identification. *Basic Clin. Pharmacol. Toxicol.* Suppl 3, 93–101 (2017).
65. Aberg KA, Xie LY, Mcclay JL *et al.* Testing two models describing how methylome-wide studies in blood are informative for psychiatric conditions. *Epigenomics* 5(4), 367–377 (2013).
66. Wong CC, Meaburn EL, Ronald A *et al.* Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry* 19(4), 495–503 (2014).
67. Dempster EL, Wong CC, Lester KJ *et al.* Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. *Biol Psychiatry* 76(12), 977–983 (2014).
68. Fisher HL, Murphy TM, Arseneault L *et al.* Methylomic analysis of monozygotic twins discordant for childhood psychotic symptoms. *Epigenetics* 10(11), 1014–1023 (2015).
69. Kaut O, Schmitt I, Tost J *et al.* Epigenome-wide DNA methylation analysis in siblings and monozygotic twins discordant for sporadic Parkinson's disease revealed different epigenetic patterns in peripheral blood mononuclear cells. *Neurogenetics* 18(1), 7–22 (2016).
70. Pei-Chien Tsai JTB. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int.J. Epidemiol.* 441(4), 1429– 1441 2015).
71. Yuen AWC, Keezer MR, Sander JW. Epilepsy is a neurological and a systemic disorder. *Epilepsy Behav.* 78, 57–61 (2018).
72. Pinto-Medel MJ, Oliver-Martos B, Urbaneja-Romero P *et al.* Global methylation correlates with clinical status in multiple sclerosis patients in the first year of IFNbeta treatment. *Sci. Rep.* 7(1), 8727 (2017).
73. Fiori LM, Turecki G. Investigating epigenetic consequences of early-life adversity: some methodological considerations. *Eur. J. Psychotraumatol.* 7, 31593 (2016).
74. Karsten SL, Kudo LC, Bragin AJ. Use of peripheral blood transcriptome biomarkers for epilepsy prediction. *Neurosci. Lett.* 497(3), 213–217 (2011).
75. Breen MS, Uhlmann A, Nday CM *et al.* Candidate gene networks and blood biomarkers of methamphetamine-associated psychosis: an integrative RNA-sequencing report. *Transl. Psychiatry* 6, e802 (2016).
76. Bakulski KM, Halladay A, Hu VW, Mill J, Fallin MD. Epigenetic research in neuropsychiatric disorders: the "tissue issue". *Curr. Behav. Neurosci. Rep.* 3(3), 264–274 (2016).
77. Wockner LF, Noble EP, Lawford BR *et al.* Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Transl. Psychiatry* 4, e339 (2014).

Chapter 5 Deciphering the role of epigenetics in self-limited epilepsy with centrotemporal spikes.

This chapter comprises entirely of a publication:

Mohandas N, Loke YJ, Mackenzie L, Bennett C, Berkovic SF, Craig JM, Vadlamudi L. Deciphering the role of epigenetics in self-limited epilepsy with centrotemporal spikes. Epilepsy Research 2019. 156:106163.

Epilepsy Research 156 (2019) 106163

Contents lists available at ScienceDirect

Epilepsy Research

journal homepage: www.elsevier.com/locate/epilepsyres

Deciphering the role of epigenetics in self-limited epilepsy with centrotemporal spikes

Namitha Mohandas[a,b], Yuk Jing Loke[a], Lisa Mackenzie[c], Carmen Bennett[c], Samuel F. Berkovic[e], Jeffrey M. Craig[a,b,f], Lata Vadlamudi[c,d,*]

[a] Environmental & Genetic Epidemiology Research, Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria, Australia
[b] Department of Paediatrics, University of Melbourne, Flemington Road, Parkville, Victoria, Australia
[c] Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Queensland, Australia
[d] Royal Brisbane and Women's Hospital, Queensland, Australia
[e] Epilepsy Research Centre, University of Melbourne, Austin Health, Victoria, Australia
[f] Centre for Molecular and Medical Research, School of Medicine, Deakin University, Geelong, Victoria 3220, Australia

ARTICLE INFO

Keywords:
Self-limited epilepsy with centrotemporal spikes
Epigenetics
Monozygotic twins

ABSTRACT

Objective: The aetiology of self-limited epilepsy with centro-temporal spikes (SECTS) remains controversial and a strong genetic basis has long been presumed. The discordant monozygotic twin (MZ) model controls for shared genetic and environmental factors, enabling focus on the potential role of the non-shared environment.
Methods: DNA methylation data was acquired from DNA extracted from three discordant MZ twin pairs, from both new born blood spots before epilepsy onset, and blood samples taken after epilepsy onset. An epigenome-wide analysis was performed, using the Illumina Infinium EPIC array. Differentially methylated regions (DMR) were identified using the *bumphunter* package in R. Comparative analyses were undertaken at the two different time points as well as a combined analysis independent of time.
Results: Many of the top DMR-associated genes have previously been described in neurodevelopmental disorders. The *LYPD8* gene was associated with a top-ranked DMR both at birth and across the two time points.
Conclusion: We have demonstrated the novel utility of the longitudinal, discordant MZ twin model, to facilitate a deeper appreciation of the complex neurobiology of SECTS. The genetic architecture of SECTS is complex and is likely to involve an interplay between genes and environment, in part mediated by epigenetics.

1. Introduction

Self-limited epilepsy with centrotemporal spikes (SECTS), previously termed benign rolandic epilepsy, is an age-dependent focal epilepsy of childhood. Onset is usually between 5 and 10 years of age and characterized by unilateral sensorimotor seizures, normal neurological development and the electroencephalogram (EEG) trait of centrotemporal spikes.

Historically, SECTS was thought to be caused by genetic factors (Bray and Wiser, 1965; Heijbel et al., 1975). A purely genetic basis for SECTS has been challenged over the years by the paucity of affected relatives and lack of concordance in monozygotic (MZ) twins (Vadlamudi et al., 2004, 2006).

To date, genes such as *GRIN2A*, *DEPDC5*, *ELP4*, *BDNF* and *KCNQ2* have been implicated in SECTS. For a number of these genes the data is not strong or the observations are in cases on the severe end of the phenotypic spectrum of focal epilepsies, often with associated neurological impairment (Xiong and Zhou, 2017).

The role of environmental factors needs consideration in SECTS, potentially mediated by epigenetic mechanisms such as DNA methylation. Recently, epigenetic analyses of MZ twins discordant for neurodevelopmental disorders such as cerebral palsy (Mohandas et al., 2018) have identified differences in DNA methylation in relevant genes. Whilst epigenetic states have a significant tissue-specific component, recent evidence has shown that for some brain disorders including epilepsies, epigenetic analyses of peripheral tissues are of value

Abbreviations: MZ, Monozygotic; DMP, Differentially methylated probe; DMR, Differentially methylated region; EWAS, Epigenome-wide association study

* Corresponding author at: University of Queensland Centre for Clinical Research (UQCCR), Royal Brisbane and Women's Hospital, Butterfield Street, Herston, Qld 4029, Australia
E-mail addresses: namitha.mohandas@mcri.edu.au (N. Mohandas), jane.loke@mcri.edu.au (Y.J. Loke), lisa.mackenzie@uq.edu.au (L. Mackenzie), c.bennett@uq.edu.au (C. Bennett), s.berkovic@unimelb.edu.au (S.F. Berkovic), jeffrey.craig@deakin.edu.au (J.M. Craig), l.vadlamudi@uq.edu.au (L. Vadlamudi).
URL: http://carmenbennett8@gmail.comc.bennett@uq.edu.au (C. Bennett).

https://doi.org/10.1016/j.eplepsyres.2019.106163
Received 5 April 2019; Received in revised form 12 June 2019; Accepted 3 July 2019
Available online 04 July 2019

Table 1
Characteristics of the affected twins with SECTS.

Twin individual	1*	2	3*
Sex	Female	Male	Female
Age of seizure onset (years)	8	10	7
Age at last seizure (years)	11	14	10
Age when second samples taken (years)	23	14	29
Seizure semiology	Gagging noises, drooling, focal seizures	drooling, focal seziures	focal seizures
EEG	Left centro-temporal spikes	Right centro-temporal spikes	Right centro-temporal spikes
Brain MRI & Neurological examination	Normal	Normal	Normal
Family history	Maternal aunt seizures of uncertain nature, mother nocturnal events	No	No

*Vadlamudi et al. 2004.

(Karsten et al., 2011; Yuen et al., 2018).

DNA methylation analysis of Guthrie cards (heel-prick test cards routinely taken shortly after birth) enables identification of epigenetic marks that occur prior to the disease onset supporting the mark being a potential cause rather than effect of the disease, an inference further supported if the mark persists after disease onset.

In this pilot study, we analysed genome-wide DNA methylation in three pairs of MZ twins discordant for SECTS. Our primary aim was to identify gene regions that show SECTS-specific differences in DNA methylation across all pairs at birth and further observe if any of these or other differences was present post diagnosis.

2. Material and methods

2.1. Study participants

Ethics approval was obtained from the Mater Health Services (ethics approval number HREC/13/MHS/114). SECTS-discordant MZ twin pairs were recruited through Twins Research Australia (https://www.twins.org.au) and the Epilepsy Research Centre Database (http://www.epilepsyresearch.org.au). Zygosity testing confirmed monozygosity. Informed consent was obtained from each twin individual. Diagnoses were confirmed with clinical and EEG findings. Blood spots from stored neonatal Guthrie cards were sourced and blood samples post diagnosis (aged 14–29 years) were collected (Table 1).

2.2. DNA methylation analysis

Following genomic DNA extraction from all of the samples, an epigenome-wide analysis was performed, using the Illumina Infinium EPIC array. Data was analysed through a Bioconductor workflow for analyzing DNA methylation array data (Mohandas et al., 2018). Differentially methylated probes (DMPs) were considered significant if the false discovery rate (FDR) threshold was < 0.1 and if the unadjusted p-value $< 1 \times 10^{-4}$. Differentially methylated regions (DMRs), which are multiple proximal CpGs that are concordantly differentially methylated, were identified using the *bumphunter* package in R and ranked by bootstrap-based family-wide error rates (fwer), which is the proportion of permutations that have at least one region as extreme as the observed region.

3. Results

3.1. DNA methylation differences at birth

Data exploration using principal component analysis and multi-dimensional scaling was performed to estimate the sources of variation in the data. Known covariates including age, sex, birth weight, birth order, medication and gestational age were tested to capture similarities and variations between data samples. DMPs were identified but none were considered statistically significant, as they were not within the false discovery rate threshold of 0.1.

To minimise type 1 errors, we focussed on identifying DMRs, each identified as more than 3 probes within a distance of < 1 kb and effect sizes (difference in DNA methylation) of greater than 5%. With our limited sample size, we found no significant DMRs (family-wise error rate; fwer < 0.1) at birth.

Within the top ten DMR-associated genes ranked by fwer, those previously associated with neurodevelopmental disorders, included *TRIM39*, *RNF144A* and *S100A8*. The other top-ranked DMRs are shown in Table 2. The top-ranked DMR, *LYPD8* (fwer = 0.78; p-value = 4.63×10^{-05}) is located on chromosome 1, and included seven probes with average within-pair DNA methylation differences between 8.2% and 13.6%.

3.2. Post-diagnostic DNA methylation analysis

Within the top ten DMR-associated genes ranked by fwer, those previously associated with neurodevelopmental disorders included *OR8U8* and *ITGBL1*, *TBCD* and *ABCC5*. The other top-ranked DMRs are shown in Table 2.

3.3. DNA methylation analysis independent of time

To identify epigenetic differences that differ between SECTS-affected twins and their co-twins from birth through to diagnosis, we compared all SECTS-discordant twin pairs, independent of sample collection time. This enabled greater power for analysis through a technical replication of samples in each group (samples from the same twin taken at the different time points).

Within the top ten DMR-associated genes ranked by fwer, those previously associated with neurodevelopmental disorders included, *BAIAP2, RXRA, SP110* and *RNH1*. The top-ranked DMR, *LYPD8* (family-wise error rate; fwer = 0.83; p-value = 5.22×10^{-04}), is located on chromosome 1 and included six probes with average DNA methylation differences between 5.5% and 8.8%. This gene was also identified as the top DMR in analysis at birth but was not identified as a top ranked DMR post diagnosis. For this DMR, there were greater within-pair DNA methylation differences seen at birth compared with post-diagnosis (average DNA methylation differences of 11% and 1.8% respectively) and across both time points. Comparison of the average differences in DNA methylation is shown in Fig. 1. Other top-ranked DMRs are shown in Table 2.

Table 2
Top-ranking DMRs from each analysis with average DNA methylation difference, p-value, associated gene and number of probes within differentially methylated gene region. "+" means affected hypermethylated compared with the unaffected and "-" means unaffected hypermethylated compared with the affected.

	At birth				Post-diagnosis				Combined			
	Gene (chromosome: location)	p.value	fwer	Average DNA methylation difference (%)	Gene (chromosome: location)	p.value	fwer	Average DNA methylation difference (%)	Gene	p.value	fwer	Average DNA methylation difference (%)
DMRs	LYPD8 (7 probes)	4.63E-05	0.784	10.98	KRTAP6-2 (4 probes)	6.83E-04	0.98	− 11.97	LYPD8 (6 probes)	5.22E-04	0.825	6.73
	TRIM39 (7 probes)	4.97E-05	0.784	10.13	AF127936.7 (4 probes)	7.64E-04	0.996	− 10.44	MARCH8 (4 probes)	2.05E-03	0.94	9.07
	RNF144A (7 probes)	5.11E-05	0.788	9.71	OR8U8 (3 probes)	3.12E-03	1	− 10.79	DYRK4 (3 probes)	3.32E-03	0.96	11.45
	S100A8 (6 probes)	1.03E-04	0.812	12.94	ITGBL1 (3 probes)	3.44E-03	1	− 10.44	SP110 (3 probes)	3.47E-03	0.96	11.37
	TREM1 (7 probes)	5.34E-05	0.824	7.68	RP11-359E19.1 (3 probes)	3.70E-03	1	− 11.90	RXRA (4 probes)	2.39E-03	0.955	8.36
	ZIM2/PEG3 (7 probes)	5.35E-05	0.824	− 7.49	TBCD (3 probes)	3.74E-03	1	− 12.48	BAIAP2 (4 probes)	2.39E-03	0.955	8.29
					CCDC172 (3 probes)	4.15E-03	1	− 10.36	RNH1 (5 probes)	1.17E-03	0.945	6.43

4. Discussion

The genetic architecture of SECTS is complex and is likely to involve an interplay between genes and environment, in part mediated by epigenetics. Whilst no gene regions reached our stringent threshold for significance in our limited discordant twin sample, the potential role of epigenetics is supported by the observation that many of the identified genes have been previously reported in epilepsies and also have large effect sizes.

The *LYPD8* gene was associated with a top-ranked DMR at birth and also across both time points. LYPD8 has a wide range of cellular and immune functions and is expressed in the colon and prefrontal cortex (Loughner et al., 2016). To date, there is no clinical/phenotypic data for gene variants for this gene. The overlap of DMR findings associated with *LYPD8* gene at different time points, with a reduction in this difference with time, suggests this change may have been present from early life and could be a transient effect, which is interesting in light of the fact that SECTS is an age-dependant disorder.

The general lack of overlap of DMRs at birth and post diagnosis may indicate a short-lived effect of SECTS-specific epigenetic programming at birth or the alteration of epigenetic marks by other environmental influences such as age or medications after birth.

We acknowledge that the sample size is small and limits understanding of epigenetic mechanisms associated with SECTS. Larger sample sizes of 25 twin pairs or more are preferable to detect a mean DNA methylation effect size of at least 8% methylation (FDR = 0.05) (Pei-Chien Tsai, 2015). The use of blood rather than brain tissue, can limit understanding of biological mechanisms underlying SECTS, however, epigenetic patterns in blood have been shown to be of value (Karsten et al., 2011; Yuen et al., 2018) and blood is readily accessible.

5. Conclusions

Although limited by small numbers, we have demonstrated the novel utility of the longitudinal, discordant MZ twin model, to facilitate a deeper appreciation of the complex neurobiology of SECTS.

Ethics approval and consent to participate

Informed consent was obtained from each twin individual. A multicentre ethics approval was obtained with the lead site being Mater Health Services (ethics approval number HREC/13/MHS/114).

Availability of data

The datasets used and/or analysed during the current study are available from the corresponding author on request.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Funding

This work was supported by grant from the Financial Markets Foundation for Children (grant number 2013-207). LV was supported by a Queensland Health Research Fellowship. This work was also supported by The University of Queensland Mayne Bequest Funds, Royal Brisbane and Women's Foundation Grant, Viertal-Charitable Foundation, Ramaciotti Establishment Grant and The Servier Barry Young Fellowship in Neurology.

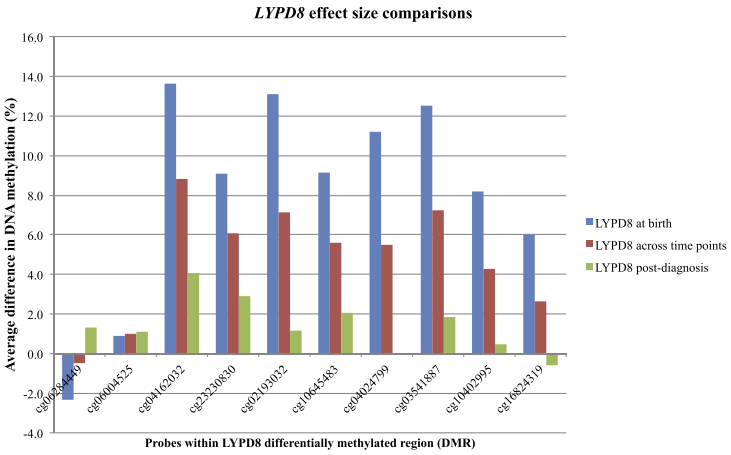


Fig. 1. Comparison of average DNA methylation differences (SECTS affected minus unaffected) for *LYPD8* DMR at birth, post-diagnosis and across both time points. X-axis represents the probes within the differentially methylated region of 556 base pairs (cg04162032 to cg03541887) and the neighbouring probes surrounding the DMR and Y-axis represents the average difference in DNA methylation.

Acknowledgements

The authors thank the participants and their families. The authors are also grateful for the support from the Murdoch Children's Research Institute (MCRI), which is supported in part by the Victorian Government's Operational Infrastructure Support Program.

References

Bray, P.F., Wiser, W.C., 1965. Hereditary characteristics of familial temporal-central focal epilepsy. Pediatrics 36, 207–211.
Heijbel, J., Blom, S., Rasmuson, M., 1975. Benign epilepsy of childhood with centrotemporal EEG foci: a genetic study. Epilepsia 16, 285–293.
Karsten, S.L., Kudo, L.C., Bragin, A.J., 2011. Use of peripheral blood transcriptome biomarkers for epilepsy prediction. Neurosci. Lett. 497, 213–217.
Loughner, C.L., Bruford, E.A., McAndrews, M.S., Delp, E.E., Swamynathan, S., Swamynathan, S.K., 2016. Organization, evolution and functions of the human and mouse Ly6/uPAR family genes. Hum. Genomics 10, 10.
Mohandas, N., Bass-Stringer, S., Maksimovic, J., Crompton, K., Loke, Y.J., Walstab, J., Reid, S.M., Amor, D.J., Reddihough, D., Craig, J.M., 2018. Epigenome-wide analysis in newborn blood spots from monozygotic twins discordant for cerebral palsy reveals consistent regional differences in DNA methylation. Clin. Epigenetics 10, 25.
Pei-Chien Tsai, J.T.B., 2015. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int. J. Epidemiol.
Vadlamudi, L., Harvey, A.S., Connellan, M.M., Milne, R.L., Hopper, J.L., Scheffer, I.E., Berkovic, S.F., 2004. Is benign rolandic epilepsy genetically determined? Ann. Neurol. 56, 129–132.
Vadlamudi, L., Kjeldsen, M.J., Corey, L.A., Solaas, M.H., Friis, M.L., Pellock, J.M., Nakken, K.O., Milne, R.L., Scheffer, I.E., Harvey, A.S., Hopper, J.L., Berkovic, S.F., 2006. Analyzing the etiology of benign rolandic epilepsy: a multicenter twin collaboration. Epilepsia 47, 550–555.
Xiong, W., Zhou, D., 2017. Progress in unraveling the genetic etiology of rolandic epilepsy. Seizure 47, 99–104.
Yuen, A.W.C., Keezer, M.R., Sander, J.W., 2018. Epilepsy is a neurological and a systemic disorder. Epilepsy Behav. 78, 57–61.

Chapter 6 Comparative analysis of DNA methylation and gene expression of cerebral palsy discordant monozygotic twin pairs

6.1 Introduction

Cerebral palsy (CP) is an umbrella term for a group of heterogeneous, non-progressive motor impairment conditions with the onset of symptoms occurring either in utero or during the early stages of life. The symptoms of this disorder are heterogeneous thereby making it hard to define the disease. Primarily, CP affects movements and posture limiting physical capabilities. It can often also be accompanied by difficulties in sensation, perception, cognition, communication and behaviour as well as secondary musculoskeletal problems (Rosenbaum et al., 2007). These impairments may worsen over lifespan in some cases. CP is often not diagnosed until 2-3 years of age, which limits the period of early intervention (Smithers-Sheedy et al., 2016). It is one of the leading cause of physical disabilities occurring in childhood, with its prevalence worldwide being 2–3 per 1000 live births (Clover A, 2014). Nearly half of the affected children are born preterm and a higher incidence of the disorder is reported in twins (Pharoah & Cooke, 1996).

Around 30% of CP cases are known to have underlying genetic mutations associated with the disorder (Fahey, Maclennan, Kretzschmar, Gecz, & Kruer, 2017). Single gene variants and copy number variations (CNVs) identified by genome-wide association studies are associated with CP (MacLennan, Thompson, & Gecz, 2015), (Oskoui et al., 2015). However, the combined effect of these genetic variation accounts for only a proportion of CP cases, and indicates the involvement of additional genetic or epigenetic factors. Evidence for the role of epigenetics in CP disease risk comes from the fact that identical monozygotic (MZ) twin pairs can be discordant for cerebral palsy (Pharoah & Dundar, 2009). MZ twins are an attractive model to explore the epigenetic landscape in disease. Epigenetic marks, such as DNA methylation, provides an insight into the molecular regulatory mechanisms through which the non-shared factors may modulate CP risk in discordant MZ twin pairs. In

Chapter 3, DNA methylation differences around genes specific to immune function, cell adhesion and inflammatory signalling was identified (N. Mohandas et al., 2018). In the study using DNA extracted from newborn blood spots (Guthrie cards), 33 genomic sites were differentially methylated within 15 CP discordant twin pairs and these were associated with 25 genes, enriched in immune response signalling. A differentially methylated region (DMR) within the *LTA* gene, which has been reported to play a role in inflammation and brain development, mediating preterm birth and white mater brain injury, was also found to be consistently differentially methylated in CP cases across all twin pairs. Likewise, a 10kb gene region within the *LIME1* gene was differentially methylated in CP-affected twins across all pairs. Subsequent EWAS studies (Crowgey et al., 2018); (Bahado-Singh et al., 2019) have also implicated genes involved in immune signalling to be associated to the pathophysiology of CP. While these previous studies have successfully captured the DNA methylation patterns associated with CP, they have been limited in their ability to connect the DNA methylation differences to corresponding gene expression changes.

In this chapter, the integration of genome-wide analysis of DNA methylation profiles and gene expression profiles in CP-discordant twin pairs was performed to investigate the relationship of the methylation and expression profile in blood taken at birth. The aim of the study was to explore gene expression profiles from 14 discordant MZ twin pairs from the Victorian UNIQUE cerebral palsy cohort. Further, I aimed to compare the results of gene expression against CP-specific DNA methylation biomarkers generated from preliminary epigenome wide association analysis of MZ twins discordant for CP (N. Mohandas et al., 2018).

6.2 Materials and Methods

6.2.1 Study subjects

CP-discordant twin pairs were recruited from the Victorian Cerebral Palsy Registry based on a written questionnaire and screening assessment (as described in Chapter

3). Twin pairs where one of the twin was reported to have a neurological disorder other than CP, was excluded from the study. Twin ID 9016 was also excluded from this study, as appropriate consent was not received. The zygosity of the twin pairs was also confirmed to be monozygotic. A single 1-cm-diameter dried blood spot was acquired from all participants from a neonatal newborn screening card, stored within the Victorian Clinical Genetics Service that is collected 2 to 4 days after birth.

6.2.2 RNA microarray

RNA isolation, cDNA synthesis, and microarray on blood spot samples was performed at the Laboratory of Microarray Technology at VARI as per published protocol (Ho et al., 2013), (Khoo et al., 2011). Briefly, purified and labelled cDNA was applied onto an 8×60K whole human genome gene expression microarray (Agilent). Each array contains 60,000 oligonucleotide probes (60bp probe) covering 27,958 *entrez* gene RNAs and 7,419 long intergenic non-coding RNAs. The arrays were hybridized for 17 hours at 65 °C and 10 rpm rotation speed, then washed for 2 min each with wash buffer 1 and 2 and scanned with an Agilent G3 high-resolution scanner. Probe features were extracted from the microarray scan data using Feature Extraction software v.10.7.3.1 (Agilent).

6.2.3 Gene expression data analysis

Gene expression was performed on single channel Agilent (G4851B) human whole genome GE8x60K arrays. Data processing and analysis were performed using statistical software R (version 3.5). The raw microarray intensity data was imported into R. The background intensities were corrected followed by between-array quantile normalisation, which is recommended for background correction of non-specific signals. Gene annotation was assigned to gene symbols and *entrez* gene ids from the probe ids. Genes were filtered by removing control probes as indicated by the ControlType column. Probes with no *entrez* gene id or gene symbol were also filtered. All probes that didn't appear to be expressed were excluded and only those above background intensity (>0) on at least 4 arrays were included. Multidimensional scaling plots were used to explore the variation present in the data. Following batch effect correction using surrogate variable analysis (SVA), differential expression

analysis of individual genes was performed with the moderated paired t-test of linear model and an empirical Bayes method implemented in R package *limma* (Ritchie et al., 2015). The significance of gene expression was corrected for multiple testing, setting the false discovery rate (FDR) at 0.05. Significant differentially expressed genes (DEGs) were selected with criteria of $P<0.05$ and log2 fold change $(FC)| \geq 1.1$.

Gene ontology (GO) and pathway analysis of the top-ranking 1000 differentially expressed genes were conducted using the *goana* and *kegga* function of the *limma* package (Ritchie et al., 2015). Gene-set enrichment analysis (GSEA) of differentially expressed genes were also analysed using the *fgsea* package in R (Sergushichev, 2016). GSEA was performed to identify pathways that contain many co-regulated top differentially expressed genes but with small individual effects. It differs from gene ontology analysis in that it considers all genes in contrast to taking only significantly differentially expressed genes. Firstly all pathways from the Reactome pathway database was downloaded (https://reactome.org/). Differentially expressed genes and the corresponding t-statistic was extracted from the analysis. Duplicates of genes was removed and the maximum value of t-statistic from multiple probes on same gene was used.

6.2.4 Protein-Protein Interaction (PPI) and function analysis of the methylation-related DEGs

Although gene ontology analysis of differentially expressed genes provides an understanding of enriched biological processes, I performed protein-protein interaction (PPI) network analysis using a combination of databases: the MINT database (Chatr-aryamontri et al., 2007), from the human protein reference database (HPRD) (Peri S, 2004) and the BioGrid database (Stark et al., 2006), which includes protein-protein interactions, genetic interactions, chemical interactions, and post-translational modifications. Together, these databases have a compiled set of human interactome, which allows the mapping of differentially expressed genes onto the interactome to determine the scores of interactions and identify key hubs of functional significance. The top-ranking DEGs were uploaded to the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database 10.5 (https://string-

db.org/) (Szklarczyk et al., 2015), to construct a PPI network. This database calculates the direct (physical) and indirect (functional) associations of proteins in a given organism, providing critical assessment and integration of PPIs. The protein pairs with a combined confidence score > 0.4 were collected for PPI network construction using Cytoscape software 3.4.0 (http://cytoscape.org/) (Shannon et al., 2003). As nodes with a high degree of connectivity contribute more to the stability of the network, the connectivity degree of each protein node in the PPI network was calculated and the top five were identified as the hub nodes using the Cytoscape plugin NetworkAnalyzer (release 2.7; med.bioinf.mpi-inf.mpg.de/netanalyzer/index.php).

6.3 Results

6.3.1 Clinical characteristics of participants

From a total of 15 MZ twin pairs discordant for CP from the Victorian CP registry, 14 pairs were consented for gene expression study (twin ID 9016 was excluded for the reason provided above (section 6.2.1). Within the cohort, 12 twin pairs were born preterm (less than 38 weeks gestation) and CP diagnosis was heterogeneous including spastic diplegia, spastic quadriplegia, spastic hemiplegia, dyskinetic dystonic, dyskinetic hypotonia and ataxia types. Further details of the clinical characteristics of the cohort are described elsewhere (N. Mohandas et al., 2018).

6.3.2 Differential gene expression analysis

Data filtering and quality control steps outlined in Methods (section 2.4.3), differential expression analysis identified no statistically significant differentially expressed genes across CP cases and unaffected twins with an adjusted p-value of < 0.05. However, among a total of 144 top-ranking differentially expressed genes based on a log fold change (logFC) cut-off of >1.1, I identified 84 genes that were upregulated and 60 genes that were downregulated (**Table 6.1**, **Figure 6.1**). The relative differences in expression were modest, with most genes being differing less than 2-fold (**Table 6.1**). The most downregulated gene was *CBFA2T2* (CBFA2/RUNX1 partner transcriptional co-repressor 2) with a logFC of -1.76. It was

previously reported in a case study of intellectual disability and motor developmental delay as a result of a de novo microdeletion (Hanafusa et al., 2017). The *TASP1* (taspase 1) gene with a logFC of -1.78, has been reported to have copy number variants indicated in attention deficit hyperactivity disorder (ADHD) (Martin et al., 2014). Another top-ranking differentially expressed gene, ranking fifth and being the most upregulated, *TTC1* (tetratricopeptide repeat domain 1), with a logFC of 1.96, was identified as a candidate gene with a single nucleotide polymorphism in brain malformation and functionally affected patients with neurological conditions (Karaca E, 2015).

Table 6.1 Genes differentially expressed (fold change ≥1.1, ≤-1.1) between cerebral palsy cases and unaffected co-twins.

Upregulated genes				
Gene symbol	**logFC**	**AveExpr**	**P.Value**	**adj.P.Val**
TTC1	1.959	6.581	1.58E-04	4.91E-01
RHOD	1.880	4.238	3.91E-02	6.01E-01
DDX28	1.837	4.191	3.85E-02	6.00E-01
CNOT11	1.833	5.269	6.61E-03	5.18E-01
CNOT11	1.754	5.311	1.56E-03	4.91E-01
LPCAT3	1.605	6.046	2.58E-04	4.91E-01
CNOT11	1.584	5.308	4.39E-03	4.95E-01
RMDN1	1.543	5.139	5.55E-03	5.05E-01
CNOT11	1.460	5.358	2.24E-03	4.91E-01
AHCYL1	1.453	3.876	1.38E-03	4.91E-01
ARHGAP25	1.447	6.091	2.24E-03	4.91E-01
CNOT11	1.443	5.208	2.47E-02	5.83E-01
RETN	1.430	7.397	8.07E-05	4.91E-01
MBD2	1.409	3.878	1.44E-03	4.91E-01
SERHL2	1.396	4.094	3.84E-02	6.00E-01
MRPL34	1.391	4.487	2.51E-04	4.91E-01
MYH9	1.391	6.838	7.16E-03	5.26E-01
LHFPL2	1.390	4.100	2.65E-04	4.91E-01
STEAP4	1.375	6.811	1.99E-04	4.91E-01
LAIR1	1.367	5.315	2.32E-04	4.91E-01
DNAJB2	1.365	4.889	2.04E-01	7.49E-01
CNOT11	1.345	5.270	7.51E-03	5.34E-01
HIST1H2BM	1.341	5.700	5.72E-05	4.91E-01
PXYLP1	1.335	4.438	7.19E-03	5.26E-01
SLC22A16	1.335	4.033	2.01E-01	7.47E-01
TRPM3	1.327	4.410	1.34E-01	7.01E-01
KCNK7	1.326	3.657	1.62E-03	4.91E-01
UBALD1	1.323	5.842	5.23E-04	4.91E-01
MAP3K20	1.301	4.539	2.52E-03	4.91E-01
GFOD1	1.301	4.023	5.59E-03	5.05E-01
PLSCR4	1.288	3.588	2.65E-03	4.91E-01
HP	1.288	6.604	1.68E-03	4.91E-01
PRPF40A	1.287	6.095	7.90E-03	5.34E-01
GFOD1	1.285	4.119	1.30E-02	5.60E-01
GPATCH2	1.285	4.350	2.32E-03	4.91E-01
RNF149	1.282	5.916	2.66E-03	4.91E-01
LSP1	1.281	7.116	1.38E-02	5.63E-01
INO80D	1.281	4.999	2.42E-02	5.83E-01
SLC10A3	1.272	4.026	2.87E-03	4.91E-01
CRACR2A	1.249	4.604	8.73E-03	5.37E-01
RNF149	1.243	5.917	7.32E-04	4.91E-01

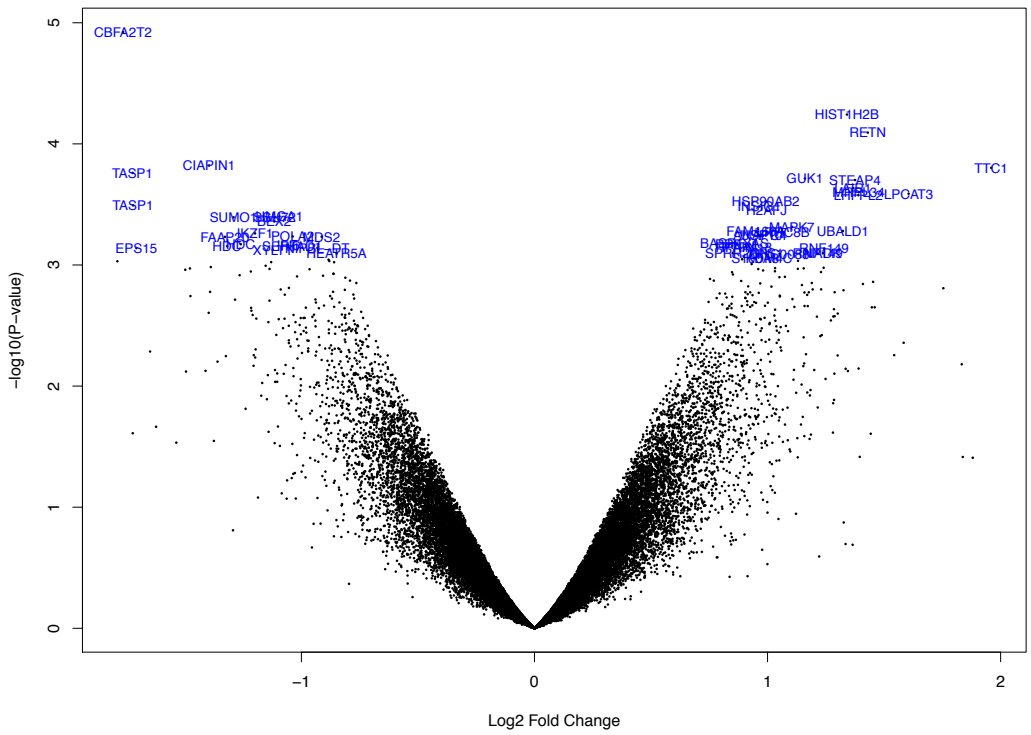
Gene symbol	logFC	AveExpr	P.Value	adj.P.Val
RTN4	1.241	4.671	1.05E-03	4.91E-01
NADK	1.237	5.251	8.00E-04	4.91E-01
CYSTM1	1.233	4.576	3.79E-03	4.91E-01
DSC2	1.230	3.896	1.59E-03	4.91E-01
CNBP	1.229	5.828	1.14E-03	4.91E-01
DHTKD1	1.226	3.620	3.87E-03	4.91E-01
UNC93B1	1.222	4.986	2.55E-01	7.76E-01
RNF149	1.216	5.952	8.08E-04	4.91E-01
TRNT1	1.215	5.422	1.50E-02	5.67E-01
YIPF3	1.212	4.725	2.67E-03	4.91E-01
RNF149	1.200	5.698	7.37E-03	5.33E-01
NRBF2	1.194	5.431	1.59E-03	4.91E-01
PSMD7	1.190	4.469	1.41E-03	4.91E-01
CNOT11	1.186	5.302	2.51E-02	5.83E-01
CNOT11	1.184	5.406	1.08E-02	5.46E-01
MBTPS2	1.182	3.840	1.20E-02	5.51E-01
CNOT11	1.180	5.420	2.73E-02	5.85E-01
CAPN10	1.177	5.306	1.68E-03	4.91E-01
NT5C2	1.173	5.162	2.19E-03	4.91E-01
CNOT11	1.171	5.283	1.13E-02	5.51E-01
AKAP8L	1.170	4.714	1.77E-03	4.91E-01
SSH2	1.165	3.929	5.46E-03	5.05E-01
GUK1	1.160	4.529	1.93E-04	4.91E-01
RNF149	1.159	5.978	2.18E-03	4.91E-01
HIST1H3F	1.158	4.078	1.69E-03	4.91E-01
BICD2	1.158	3.872	2.59E-02	5.84E-01
TPM3	1.157	4.441	6.50E-03	5.18E-01
CSF2RB	1.155	5.690	1.07E-03	4.91E-01
EHBP1L1	1.155	5.488	3.96E-03	4.91E-01
CRTC1	1.153	3.800	6.84E-03	5.19E-01
CUL5	1.151	4.025	1.80E-02	5.67E-01
TRIM4	1.147	3.588	1.46E-03	4.91E-01
DEFA3	1.146	7.846	9.71E-03	5.40E-01
BBX	1.143	3.592	2.08E-02	5.73E-01
PLXND1	1.131	5.024	9.24E-04	4.91E-01
OS9	1.130	6.492	1.08E-03	4.91E-01
MCOLN1	1.130	5.123	3.88E-02	6.00E-01
IL1B	1.128	4.375	2.69E-02	5.85E-01
SHISA5	1.123	8.169	1.14E-02	5.51E-01
PRDX1	1.122	5.364	1.13E-01	6.88E-01
PSMF1	1.115	4.891	2.22E-03	4.91E-01
PBRM1	1.113	5.262	7.35E-03	5.33E-01
BOLA2B	1.111	5.530	8.86E-03	5.37E-01

Downregulated genes

Gene symbol	logFC	AveExpr	P.Value	adj.P.Val
TASP1	-1.789	4.434	9.32E-04	4.91E-01
CBFA2T2	-1.765	4.471	1.19E-05	4.30E-01
TASP1	-1.724	4.530	1.74E-04	4.91E-01
BCOR	-1.724	4.415	2.45E-02	5.83E-01
TASP1	-1.722	4.485	3.22E-04	4.91E-01
EPS15	-1.708	5.442	7.25E-04	4.91E-01
GRK5	-1.649	4.721	5.18E-03	5.05E-01
SPARCL1	-1.623	3.456	2.17E-02	5.75E-01
ARSA	-1.537	7.885	2.93E-02	5.85E-01
SLC25A19	-1.498	4.554	1.09E-03	4.91E-01
TASP1	-1.495	4.299	7.59E-03	5.34E-01
HDC	-1.478	5.758	1.07E-03	4.91E-01
HDC	-1.476	5.725	1.81E-03	4.91E-01
GRAPL	-1.411	5.261	7.48E-03	5.34E-01
U2AF2	-1.398	5.606	2.48E-03	4.91E-01
CIAPIN1	-1.396	2.932	1.51E-04	4.91E-01
SNHG9	-1.392	6.378	1.67E-03	4.91E-01
HDC	-1.389	5.817	1.04E-03	4.91E-01
SERPINB6	-1.375	5.042	2.84E-02	5.85E-01
TASP1	-1.360	4.582	6.28E-03	5.17E-01
FAAP20	-1.327	4.912	5.86E-04	4.91E-01
CHKB	-1.324	6.587	5.66E-03	5.05E-01
HDC	-1.319	5.663	7.00E-04	4.91E-01
SCMH1	-1.293	4.195	1.55E-01	7.19E-01
SUMO1	-1.291	5.693	4.04E-04	4.91E-01
RARB	-1.285	4.009	1.93E-03	4.91E-01
HDC	-1.265	5.706	1.21E-03	4.91E-01
HDC	-1.260	5.739	6.74E-04	4.91E-01
VSTM1	-1.239	6.697	1.54E-02	5.67E-01
AMY1C	-1.217	4.358	1.13E-03	4.91E-01
SPTAN1	-1.216	6.002	2.26E-03	4.91E-01
HDC	-1.214	5.779	2.38E-03	4.91E-01
GNL1	-1.209	3.378	2.58E-03	4.91E-01
TASP1	-1.205	4.622	6.80E-03	5.19E-01
ADPRM	-1.204	3.605	5.55E-03	5.05E-01
IKZF1	-1.199	6.630	5.49E-04	4.91E-01
IGF2	-1.198	4.339	5.81E-03	5.09E-01
WDR53	-1.196	4.445	4.98E-03	5.05E-01
CYFIP2	-1.193	3.836	2.77E-03	4.91E-01
CBR1	-1.186	3.837	8.32E-02	6.61E-01
PVRIG	-1.172	6.238	1.20E-02	5.51E-01

Gene symbol	logFC	AveExpr	P.Value	adj.P.Val
DCAF8	-1.168	4.330	9.42E-03	5.37E-01
PDCD4	-1.153	4.775	1.02E-03	4.91E-01
EIF4A2	-1.148	6.381	1.28E-02	5.58E-01
PUM1	-1.144	3.293	8.12E-03	5.34E-01
LINC01242	-1.142	3.611	2.97E-03	4.91E-01
TASP1	-1.141	4.218	2.38E-02	5.83E-01
DDI2	-1.137	3.699	1.97E-03	4.91E-01
NMI	-1.137	4.041	1.08E-03	4.91E-01
POR	-1.130	3.558	9.47E-04	4.91E-01
XYLT1	-1.123	3.626	7.51E-04	4.91E-01
BEX2	-1.116	4.229	4.37E-04	4.91E-01
TUBB8	-1.115	3.852	2.96E-02	5.85E-01
SSU72	-1.114	4.247	4.03E-04	4.91E-01
SDCBPP2	-1.113	6.422	7.37E-03	5.33E-01
TYW1	-1.113	4.150	9.23E-03	5.37E-01
PRKXP1	-1.108	4.824	3.78E-03	4.91E-01
OTUD6B	-1.104	4.120	9.21E-03	5.37E-01
P2RX4	-1.104	6.431	7.86E-03	5.34E-01
USP15	-1.100	7.486	2.16E-02	5.75E-01

Figure 6.1 Volcano plot visualisation of differential gene expression.



The x-axis is the log of the fold change and the y-axis is the log of the p-value. Each dot corresponds to a specific gene – gene names that are blue are nominally significant (p-value < 0.05 and logFC ≥1.1 or ≤-1.1) top-ranking differentially expressed genes.

6.3.3 Gene ontology and KEGG pathway analysis

The top-ranking differentially expressed genes were enriched in gene ontology annotations related to immune functions, with the top four significant biological process terms associating to orbitofrontal cortex development, immune-response-activating signal transduction, immune response-regulating signalling pathway and activation of immune response (**Table 6.2**). KEGG pathway analysis revealed upregulated genes involved in pathways related to infection with the top five significant KEGG pathways enriched in pathogenic *E coli* infection, shigellosis, bacterial invasion of epithelial cells, salmonella infection and Fc gamma R-mediated phagocytosis (**Table 6.3**).

Gene set enrichment analysis across the Reactome pathway revealed significantly perturbed curated gene sets (**Table 6.4; Figure 6.2**). The most prominently altered

were those related to immune system, interleukin signalling and signalling by Rho GTPases.

Table 6.2 Gene ontology biological process enrichment analyses of differentially expressed genes for CP discordant monozygotic twin pairs.

GO ID	Biological process term (BP)	N	Up	P.Up
GO:0021769	orbitofrontal cortex development	4	1	3.83E-04
GO:0002757	immune response-activating signal transduction	500	2	5.73E-04
GO:0002764	immune response-regulating signaling pathway	531	2	6.47E-04
GO:0002253	activation of immune response	572	2	7.50E-04
GO:0070255	regulation of mucus secretion	9	1	8.63E-04
GO:0051639	actin filament network formation	10	1	9.58E-04
GO:0070254	mucus secretion	11	1	1.05E-03
GO:0002862	negative regulation of inflammatory response to antigenic stimulus	11	1	1.05E-03
GO:0050778	positive regulation of immune response	720	2	1.19E-03
GO:0030011	maintenance of cell polarity	17	1	1.63E-03
GO:0002861	regulation of inflammatory response to antigenic stimulus	21	1	2.01E-03
GO:0050776	regulation of immune response	949	2	2.07E-03
GO:0002684	positive regulation of immune system process	995	2	2.27E-03
GO:0061842	microtubule organizing center localization	26	1	2.49E-03
GO:0034122	negative regulation of toll-like receptor signaling pathway	33	1	3.16E-03

'N' represents the number of genes in the GO term; 'Up' represents the number of up-regulated differentially expressed genes; 'P.Up' represents p-value for over-representation of GO term in up-regulated genes before adjusting for multiple testing.

Table 6.3 KEGG enrichment analyses of differentially expressed genes for CP discordant monozygotic twin pairs.

KEGG pathway	N	Up	P.Up
Pathogenic Escherichia coli infection	54	1	5.17E-03
Shigellosis	64	1	6.13E-03
Bacterial invasion of epithelial cells	74	1	7.08E-03
Salmonella infection	84	1	8.04E-03
Fc gamma R-mediated phagocytosis	88	1	8.42E-03
NOD-like receptor signaling pathway	162	1	1.55E-02
Regulation of actin cytoskeleton	207	1	1.97E-02
Endocytosis	242	1	2.31E-02

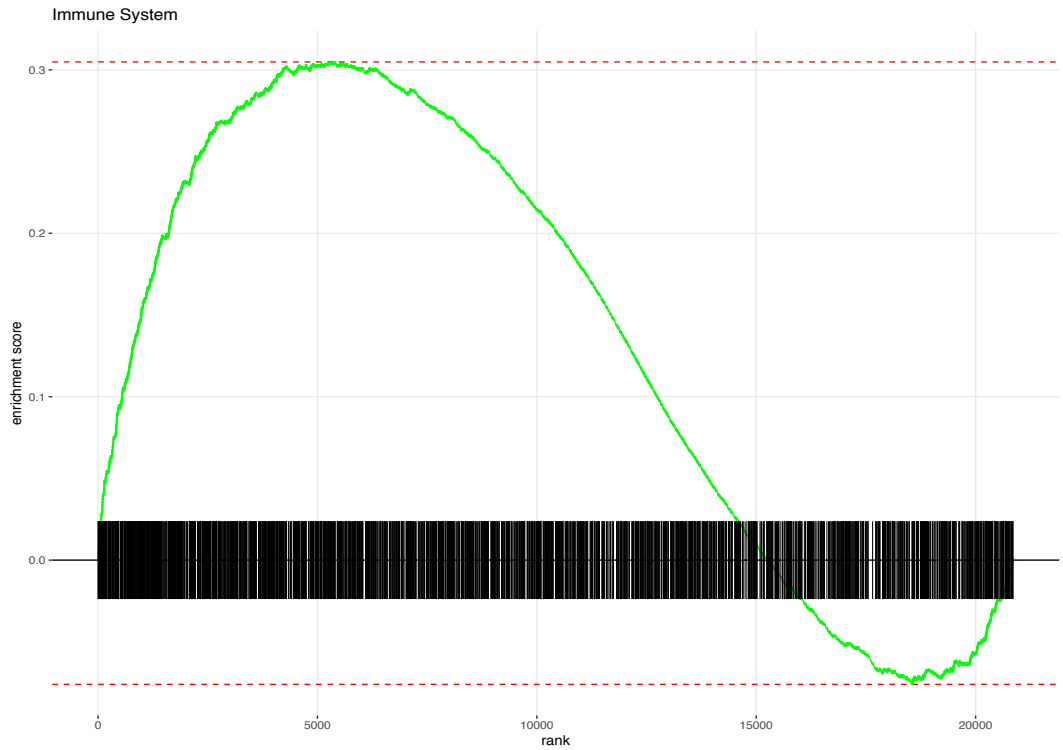
'N' represents the number of genes in the KEGG pathway; 'Up' represents the number of up-regulated differentially expressed genes; 'P.Up' represents p-value for over-representation of GO term in up-regulated genes before adjusting for multiple testing.

Table 6.4 Top ten enriched pathways after gene-set enrichment testing with top-ranking differentially expressed genes.

Pathway	pval	padj	ES	leadingEdge	Core genes
Immune System	1.00E-04	1.05E-02	0.304844	1925	RETN,LAIR1,MAPK7,ANAPC4,PTPN18,S100A8,...
Innate Immune System	1.00E-04	1.05E-02	0.3443865	1020	RETN,LAIR1,MAPK7,S100A8,AHCYL1,PSMD7,...
Transport of small molecules	1.00E-04	1.05E-02	0.3142116	716	LRRC8B,OS9,ATP9B,PSMD7,PSMF1,TRPM6,...
Neutrophil degranulation	1.02E-04	1.05E-02	0.4161991	468	RETN,LAIR1,S100A8,PSMD7,CEACAM3,HP,...
Signaling by Interleukins	1.02E-04	1.05E-02	0.3457324	446	MAPK7,PTPN18,CSF2RB,PSMD7,HSP90AA1,PSMF1,...
Signaling by Receptor Tyrosine Kinases	1.02E-04	1.05E-02	0.3218355	441	MAPK7,PTPN18,AHCYL1,HSP90AA1,NCOR1,ITGB1,...
Signaling by Rho GTPases	1.03E-04	1.05E-02	0.3762591	397	HIST1H2BM,H2AFJ,PREX1,S100A8,KDM4C,NCOA2,...
RHO GTPase Effectors	1.06E-04	1.05E-02	0.382221	269	HIST1H2BM,H2AFJ,S100A8,KDM4C,NCOA2,S100A9,...
Cell Cycle Checkpoints	1.06E-04	1.05E-02	0.3830295	268	HIST1H2BM,ANAPC4,PSMD7,PSMF1,DYNC1LI1,MDM4,...
Chromatin modifying enzymes	1.07E-04	1.05E-02	0.3776612	234	HIST1H2BM,H2AFJ,KDM4C,NCOA2,NCOR1,HIST1H2BD,...

'ES' represents an enrichment score. Leading edges correspond to genes that drive the enrichment as given by *fgsea*.

Figure 6.2 Gene-set enrichment plot for top-ranking differentially expressed genes.

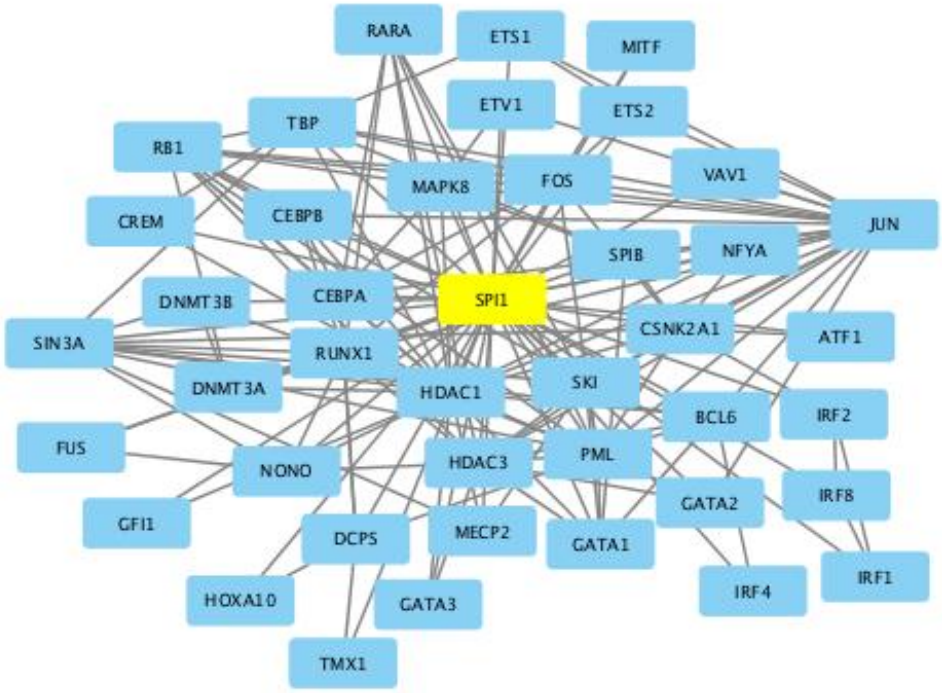


The x-axis shows the ranked list of genes, and the vertical bars on the x-axis show the genes that belong to gene set, which in this case is the "Immune system" set and the y-axis shows the enrichment score. The green line represents the genes enriched in immune system regulation.

6.3.4 Protein-protein interaction network analysis

The protein-protein interaction network of differentially expressed genes included 42 nodes, which represent proteins and 156 edges (**Figure 6.3**), which represent protein-protein associations, with a PPI enrichment p-value of 0.664 based on the STRING database. A module consisting of MAPKs (mitogen activated protein kinase), HDACs (histone deacetylases), DNMT3A and 3B (DNA methyltransferase), *CEBP8* (CCAAT/enhancer-binding protein beta) and IRFs (interferon regulatory factor) was recognized as a significant cluster associating to *SPI1* in the PPI network.

Figure 6.3 Protein-protein interaction network of differentially expressed genes.



Network includes 156 edges (interaction in grey lines) between 42 nodes (blue nodes) respectively. The node with yellow colour represents the key gene in the network module.

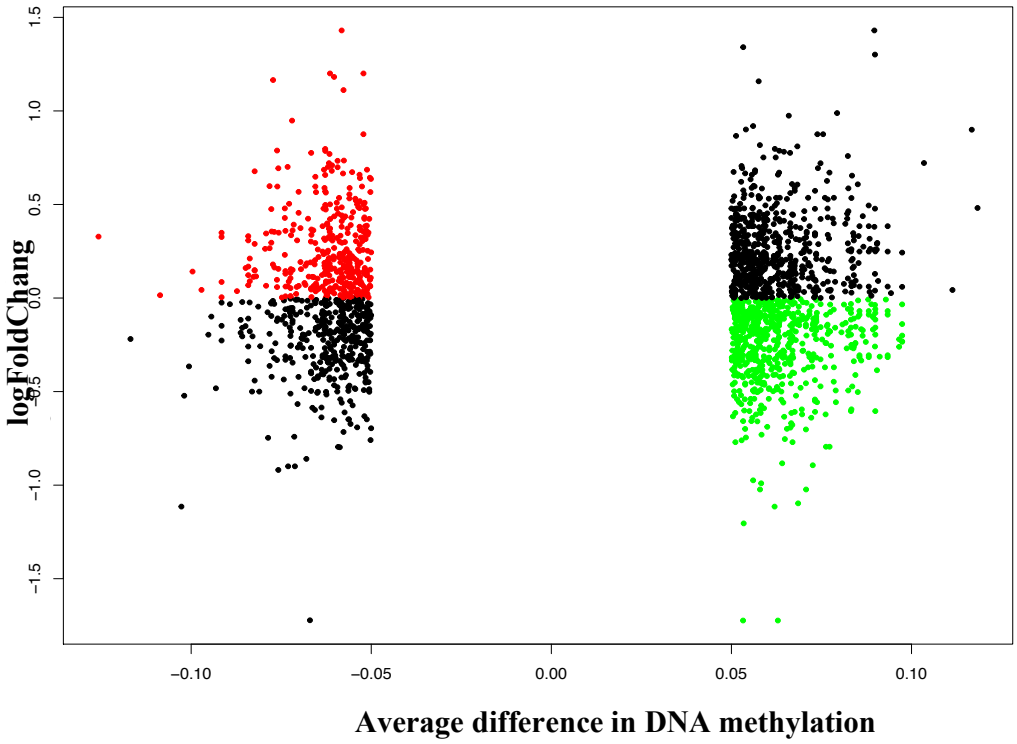
6.3.5 Comparison with differential DNA methylation analysis

To explore the relationship between DNA methylation and gene expression, I compared the top-ranking 100 differentially expressed genes (ranked by unadjusted p value) and the top-ranking 100 differentially methylated probe-associated genes (ranked by unadjusted p value) and found no overlap between the two datasets.

To further investigate the relationship of DNA methylation variation with gene expression patterns, the gene expression profiles of genes at the site of differentially methylation were analysed in more detail. As DNA methylation is known to inhibit gene expression, genes that showed a negative correlation between DNA methylation and gene expression were focused on. I found 1012 genes that displayed an inverse relation, among which, 383 genes were found to be hypomethylated and upregulated and 629 genes were hypermethylated and downregulated (**Figure 6.4**). Out of the 33

top-ranking DMPs identified in the CP EWAS study, 10 genes were found to have an inverse correlation with gene expression. Among the enriched gene ontology terms in both gene expression and methylation analysis, an overlap of 10 GO terms, all relating to immune response/signalling pathways was found (**Table 6.5**).

Figure 6.4 Starburst plot integrating alterations in DNA methylation and gene expression.



The x-axis is the difference in DNA methylation levels (CP affected twin minus unaffected) (ΔM); the y-axis is the difference in gene expression (log 2 FC); green dots represent the hypomethylated/up-regulated genes (383 gnes); red dots represent the hypermethylated/down-regulated genes (629 genes).

Table 6.5 Overlapping enriched gene ontology terms between gene expression and DNA methylation analysis

GO ID	Biological process term
GO:0002764	immune response-regulating signaling pathway
GO:0050776	regulation of immune response
GO:0002768	immune response-regulating cell surface receptor signaling pathway
GO:0006955	immune response
GO:0048584	positive regulation of response to stimulus
GO:0002429	immune response-activating cell surface receptor signaling pathway
GO:0002376	immune system process
GO:0045321	leukocyte activation
GO:0001775	cell activation
GO:0002443	leukocyte mediated immunity

6.4 Discussion

The present chapter identified 144 DEGs as being differentially expressed within CP discordant twin pairs. Both differential expression and DNA methylation analysis of the same cohort of CP affected individuals, supports the dysregulation of immune signalling pathway as a common mechanism. Although the causes of CP may vary, there is enough evidence to indicate the role of prematurity and hypoxic brain damage in this condition (Leviton, 1993),(Hagberg, Edwards, & Groenendaal, 2016). Ischaemic insults can cause recurring neuroinflammatory responses leading to neonatal brain injury, which are often thought to be the mechanisms underlying CP (Lin et al., 2010),(Fleiss & Gressens, 2012). Proinflammatory cytokines are thought to be major mediators in brain injury in neonates with perinatal asphyxia, bacterial infection, or both (Girard et al., 2009). Gene ontology analysis also identified enrichment in orbitofrontal cortex development, which is striking as findings from functional neuroimaging studies have shown that executive functions is primarily mediated by frontal lobes, particularly the prefrontal cortex. Such executive functions deficits have been reported in many CP cases (Bodimeade, Whittingham, Lloyd, & Boyd, 2013), (Bottcher, Flachs, & Uldall, 2010), (Jenks, de Moor, & van Lieshout, 2009).

Interestingly, KEGG analysis revealed enrichment in pathways related to infection. Inflammatory cytokines released during the course of intrauterine infection play a central role in the genesis of preterm parturition, fetal white matter brain injury, and cerebral palsy (Yoon BH, 1997). Increased levels of *IL6* and *TNFA* in umbilical cord blood has been reported as predictor for white matter lesions (Szpecht, Wiak, Braszak, Szymankiewicz, & Gadzinowski, 2016). Animal models of *E coli* infection have been also used to examine causal relationships between intrauterine infection and white matter lesions (Debillon T, 2000). The study showed that all cases with brain white matter lesions had evidence of intrauterine inflammation. In another study for *E coli* infection, the infection found in placenta of cases was significantly higher than in controls (Chmielarczyk A, 2014). Another recent study (Bear & Wu, 2016) showed that a diagnosis of infection was more common in mothers of more than 8000 infants with cerebral palsy than in mothers of unaffected children (13.7% vs. 5.5%, $P<0.001$). All three types of maternal infections studied (chorioamnionitis, other genitourinary infection and respiratory infection) were associated with cerebral palsy in multivariable analyses. Maternal extra-amniotic infections, whether diagnosed during prenatal or birth hospitalizations, conferred an increased risk of cerebral palsy (Yoon BH, 1997).

In the PPI network, an important network module consisting of SPI1 was identified. *SPI1* was identified as a top-ranking differentially methylated region, ranking third in the EWAS study. *SPI1* activates gene expression during myeloid development in microglia and macrophages (Huang KL, 2017). A reduction in the expression of *SPI1* has been found decrease the viability of microglia (Huang KL, 2017). Discordance of DNA methylation for *SPI1* may suggest a difference in the expression of genes associated with myeloid development and in turn a change in the overall viability of microglia. Changes in microglial activity have been shown to lead to an inflammatory cytokine response, which in turn can lead to neuronal injury in the brain (Nardone et al., 2014).

Our results are consistent with a previous gene expression study of CP cases versus controls (van Eyk CL, 2018), in which authors reported that 124 upregulated genes

were enriched for GO categories relating to immune response. Overlap of specific gene ontology term GO:0006955 was also observed with our study. Another study, (Ho et al., 2013), reported *IL1B* to be upregulated in CP cases and this was found to be upregulated in our analysis too (logFC = 1.13; p-value = 0.002).

This study is the first to study both gene expression and DNA methylation data of CP discordant monozygotic twins and demonstrates the dysregulation of common molecular pathways involved in clinically diverse CP, across two omic platforms. Although the study is limited by its sample size and bio sample used, it provides a basis for further investigation of specific genes and signalling pathways in future genomic and functional studies. Transcriptomic studies in neurodevelopmental disorders such as autism spectrum disorder and cerebral palsy have identified gene expression patterns in different tissues such as lymphoblastoid cell lines, peripheral blood, pluripotent stem-cell dervied neurons (Kong et al., 2013), (Ansel, Rosenzweig, Zisman, Melamed, & Gesundheit, 2016), (van Eyk CL, 2018), (Liu et al., 2017). While these tissues does not necessarily reflect the gene expression signatures in the brain, they provide a valuable source for biomarker analysis, with many studies showing the relevance to neuronal function (Nishimura et al., 2007), (Bittel, Kibiryeva, & Butler, 2007), (Bittel DC, 2007). The observed enrichment in the immune response pathways in CP cases may indicate increased susceptibility to neuronal damage resulting from environmental insults in utero or in early post-natal life. It may also indicate a shared systemic immune response due to infection or ischemia in a different tissue. Although our study revealed correlations between DNA methylation and gene expression, future studies are necessary to establish the effect of DNA methylation and gene expression on the phenotype of cerebral palsy. Integrating other omic platforms such as proteomic and metabolomic data is necessary to pinpoint disease mechanisms and disease risk (Higdon et al., 2015), (Y. V. Sun & Hu, 2016), (Ghaemi MS, 2018).

Animal models have also been used to investigate biological pathways during development and to better understand the effects of prenatal infections on neurodevelopment (Meyer, Feldon, & Fatemi, 2009), (Khandaker, Zimbron, Lewis, & Jones, 2013), (Boksa, 2010). Elevated levels of pro inflammatory cytokines during

pregnancy have been reported to be associated with reduced neuronal survival and decreased grey- and white-matter volume in animal studies (Meyer et al., 2009), (Short et al., 2010). This association suggests that perhaps there is a causal relationship between prenatal infection and neural system disorders in humans. Further validation and in depth characterisation of neural changes due to prenatal immune activation in animal models is needed. Extending studies to species other than rat and mouse, such as guinea pig, which has a more mature central nervous system during gestation, may increase relevance of findings to humans. Since neurodevelopmental disorders like cerebral palsy and autism, are of multifactorial aetiology with a strong genetic component, it is vital to integrate whole genome studies with other omic studies to comprehend the interaction of prenatal infection with genetic background. The discordant twin design also implies that these findings are independent of genetics, which usually confounds studies of singletons. Identifying the environmental and genetic interactions of this molecular aberration is an important step in understanding the aetiology of CP and predicting those at greater risk.

Acknowledgements

I thank the Victorian Cerebral Palsy Register staff for their help with the consent and recruitment of participants included in this study. The Victorian Cerebral Palsy Register receives funding from the Victorian Department of Health and Human Services. The authors are also grateful for the support from the Murdoch Children's Research Institute (MCRI), which is supported in part by the Victorian Government's Operational Infrastructure Support Program.

References

Ansel, A., Rosenzweig, J. P., Zisman, P. D., Melamed, M., & Gesundheit, B. (2016). Variation in Gene Expression in Autism Spectrum Disorders: An Extensive Review of Transcriptomic Studies. *Front Neurosci, 10*, 601. doi:10.3389/fnins.2016.00601

Bahado-Singh, R. O., Vishweswaraiah, S., Aydas, B., Mishra, N. K., Guda, C., & Radhakrishna, U. (2019). Deep Learning/Artificial Intelligence and Blood-Based DNA Epigenomic Prediction of Cerebral Palsy. *Int J Mol Sci, 20*(9). doi:10.3390/ijms20092075

Bear, J. J., & Wu, Y. W. (2016). Maternal Infections During Pregnancy and Cerebral Palsy in the Child. *Pediatr Neurol, 57*, 74-79. doi:10.1016/j.pediatrneurol.2015.12.018

Bittel, D. C., Kibiryeva, N., & Butler, M. G. (2007). Whole genome microarray analysis of gene expression in subjects with fragile X syndrome. *Genet Med, 9*(7), 464-472. doi:10.1097/gim.0b013e3180ca9a9a

Bittel DC, K. N., Sell SM, Strong TV, Butler MG. (2007). Whole genome microarray analysis of gene expression in Prader-Willi syndrome.

Bodimeade, H. L., Whittingham, K., Lloyd, O., & Boyd, R. N. (2013). Executive functioning in children with unilateral cerebral palsy: protocol for a cross-sectional study. *BMJ Open, 3*(4). doi:10.1136/bmjopen-2012-002500

Boksa, P. (2010). Effects of prenatal infection on brain development and behavior: A review of findings from animal models.

Bottcher, L., Flachs, E. M., & Uldall, P. (2010). Attentional and executive impairments in children with spastic cerebral palsy. *Dev Med Child Neurol, 52*(2), e42-47. doi:10.1111/j.1469-8749.2009.03533.x

Chatr-aryamontri, A., Ceol, A., Palazzi, L. M., Nardelli, G., Schneider, M. V., Castagnoli, L., & Cesareni, G. (2007). MINT: the Molecular INTeraction database. *Nucleic Acids Res, 35*(Database issue), D572-574. doi:10.1093/nar/gkl950

Chmielarczyk A, W.-M. J., Romaniszyn D, Adamski P, Helwich E, Lauterbach R, Pobiega M, Borszewska-Kornacka M, Gulczyńska E, Kordek A, Heczko PB. (2014). Mode of delivery and other risk factors for Escherichia coli infections in very low birth weight infants.

Clover A, F. C., Pharoah P. (2014). Cerebral Palsy. *The Lancet, 383*(9924), 1240-1249.

Crowgey, E. L., Marsh, A. G., Robinson, K. G., Yeager, S. K., & Akins, R. E. (2018). Epigenetic machine learning: utilizing DNA methylation patterns to predict spastic cerebral palsy. *BMC Bioinformatics, 19*(1), 225. doi:10.1186/s12859-018-2224-0

Debillon T, G.-L. C., Vérielle V, Winer N, Caillon J, Rozé JC, Gressens P. (2000). Intrauterine infection induces programmed cell death in rabbit periventricular white matter.

Fahey, M. C., Maclennan, A. H., Kretzschmar, D., Gecz, J., & Kruer, M. C. (2017). The genetic basis of cerebral palsy. *Dev Med Child Neurol, 59*(5), 462-469. doi:10.1111/dmcn.13363

Fleiss, B., & Gressens, P. (2012). Tertiary mechanisms of brain damage: a new hope for treatment of cerebral palsy? *The Lancet Neurology, 11*(6), 556-566. doi:10.1016/s1474-4422(12)70058-3

Ghaemi MS, D. D., Contrepois K, Callahan B, Ngo TTM, Lee-McMullen B, Lehallier B, Robaczewska A, Mcilwain D, Rosenberg-Hasson Y, Wong RJ, Quaintance C, Culos A, Stanley N, Tanada A, Tsai A, Gaudilliere D, et al. (2018). Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy.

Girard, S., Kadhim, H., Roy, M., Lavoie, K., Brochu, M. E., Larouche, A., & Sebire, G. (2009). Role of Perinatal Inflammation in Cerebral Palsy. *Pediatric Neurology, 40*(3), 168-174. doi:10.1016/j.pediatrneurol.2008.09.016

Hagberg, H., Edwards, A. D., & Groenendaal, F. (2016). Perinatal brain damage: The term infant. *Neurobiology of Disease, 92*, 102-112. doi:10.1016/j.nbd.2015.09.011

Hanafusa, H., Morisada, N., Ishida, Y., Sakata, R., Morita, K., Miura, S., . . . Iijima, K. (2017). The smallest de novo 20q11.2 microdeletion causing intellectual disability and dysmorphic features. *Hum Genome Var, 4*, 17050. doi:10.1038/hgv.2017.50

Higdon, R., Earl, R. K., Stanberry, L., Hudac, C. M., Montague, E., Stewart, E., . . . Kolker, E. (2015). The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders. *OMICS, 19*(4), 197-208. doi:10.1089/omi.2015.0020

Ho, N. T., Furge, K., Fu, W., Busik, J., Khoo, S. K., Lu, Q., . . . Paneth, N. (2013). Gene expression in archived newborn blood spots distinguishes infants who will later develop cerebral palsy from matched controls. *Pediatr Res, 73*(4 Pt 1), 450-456. doi:10.1038/pr.2012.200

Huang KL, M. E., Pimenova AA, Di Narzo AF, Kapoor M, Jin SC, Harari O, Bertelsen S, Fairfax BP, Czajkowski J. (2017). A common haplotype lowers SPI1 (PU.1) expression in myeloid cells and delays age at
onset for Alzheimer's disease

Jenks, K. M., de Moor, J., & van Lieshout, E. C. (2009). Arithmetic difficulties in children with cerebral palsy are related to executive function and working memory. *J Child Psychol Psychiatry, 50*(7), 824-833. doi:10.1111/j.1469-7610.2008.02031.x

Karaca E, H. T., Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, Gonzaga-Jauregui C, Erdin S, Bayram Y, Campbell IM, Hunter JV, Atik MM, Van Esch H, Yuan B, Wiszniewski W, Isikay S, Yesil G, Yuregir OO, Tug Bozdogan S, Aslan H, Aydin H, et al. (2015). Genes that affect brain structure and function identified by rare variant analyses of Mendelian neurologic disease.

Khandaker, G. M., Zimbron, J., Lewis, G., & Jones, P. B. (2013). Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychol Med, 43*(2), 239-257. doi:10.1017/S0033291712000736

Khoo, S. K., Dykema, K., Vadlapatla, N. M., LaHaie, D., Valle, S., Satterthwaite, D., . . . Resau, J. H. (2011). Acquiring genome-wide gene expression profiles in Guthrie card blood spots using microarrays. *Pathol Int, 61*(1), 1-6. doi:10.1111/j.1440-1827.2010.02611.x

Kong, S. W., Shimizu-Motohashi, Y., Campbell, M. G., Lee, I. H., Collins, C. D., Brewster, S. J., . . . Kunkel, L. M. (2013). Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. *Neurogenetics, 14*(2), 143-152. doi:10.1007/s10048-013-0363-z

Leviton, A. (1993). Preterm birth and cerebral palsy: is tumor necrosis factor the missing link? *Developmental Medicine & Child Neurology*.

Lin, C. Y., Chang, Y. C., Wang, S. T., Lee, T. Y., Lin, C. F., & Huang, C. C. (2010). Altered inflammatory responses in preterm children with cerebral palsy. *Ann Neurol, 68*(2), 204-212. doi:10.1002/ana.22049

Liu, X., Campanac, E., Cheung, H. H., Ziats, M. N., Canterel-Thouennon, L., Raygada, M., . . . Rennert, O. M. (2017). Idiopathic Autism: Cellular and Molecular Phenotypes in Pluripotent Stem Cell-Derived Neurons. *Mol Neurobiol, 54*(6), 4507-4523. doi:10.1007/s12035-016-9961-8

MacLennan, A. H., Thompson, S. C., & Gecz, J. (2015). Cerebral palsy: causes, pathways, and the role of genetic variants. *Am J Obstet Gynecol, 213*(6), 779-788. doi:10.1016/j.ajog.2015.05.034

Martin, J., Cooper, M., Hamshere, M. L., Pocklington, A., Scherer, S. W., Kent, L., . . . Holmans, P. (2014). Biological overlap of attention-deficit/hyperactivity disorder and autism spectrum disorder: evidence from copy number variants. *J Am Acad Child Adolesc Psychiatry, 53*(7), 761-770 e726. doi:10.1016/j.jaac.2014.03.004

Meyer, U., Feldon, J., & Fatemi, S. H. (2009). In-vivo rodent models for the experimental investigation of prenatal immune activation effects in neurodevelopmental brain disorders. *Neurosci Biobehav Rev, 33*(7), 1061-1079. doi:10.1016/j.neubiorev.2009.05.001

Mohandas, N., Bass-Stringer, S., Maksimovic, J., Crompton, K., Loke, Y. J., Walstab, J., . . . Craig, J. M. (2018). Epigenome-wide analysis in newborn blood spots from monozygotic twins discordant for cerebral palsy reveals consistent regional differences in DNA methylation. *Clin Epigenetics, 10*, 25. doi:10.1186/s13148-018-0457-4

Nardone, S., Sams, D. S., Reuveni, E., Getselter, D., Oron, O., Karpuj, M., & Elliott, E. (2014). DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Transl Psychiatry, 4*, e433. doi:10.1038/tp.2014.70

Nishimura, Y., Martin, C. L., Vazquez-Lopez, A., Spence, S. J., Alvarez-Retuerto, A. I., Sigman, M., . . . Geschwind, D. H. (2007). Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum Mol Genet, 16*(14), 1682-1698. doi:10.1093/hmg/ddm116

Oskoui, M., Gazzellone, M. J., Thiruvahindrapuram, B., Zarrei, M., Andersen, J., Wei, J., . . . Scherer, S. W. (2015). Clinically relevant copy number variations detected in cerebral palsy. *Nat Commun, 6*, 7949. doi:10.1038/ncomms8949

Peri S, N. J., Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N, Suresh S, Rashmi BP, Shanker K, Padma N, Niranjan V, Harsha HC, Talreja N, Vrushabendra BM, Ramya MA, Yatish AJ, Joy M, Shivashankar HN. (2004). Human protein reference database as a discovery resource for proteomics.

Pharoah, P. O., & Cooke, T. (1996). Cerebral palsy and multiple births. *Arch Dis Child Fetal Neonatal Ed, 75*(3), F174-177.

Pharoah, P. O., & Dundar, Y. (2009). Monozygotic twinning, cerebral palsy and congenital anomalies. *Hum Reprod Update, 15*(6), 639-648. doi:10.1093/humupd/dmp019

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res, 43*(7), e47. doi:10.1093/nar/gkv007

Rosenbaum, P., Paneth, N., Leviton, A., Goldstein, M., Bax, M., Damiano, D., . . . Jacobsson, B. (2007). A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl, 109*, 8-14.

Sergushichev, A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res, 13*(11), 2498-2504. doi:10.1101/gr.1239303

Short, S. J., Lubach, G. R., Karasin, A. I., Olsen, C. W., Styner, M., Knickmeyer, R. C., . . . Coe, C. L. (2010). Maternal influenza infection during pregnancy impacts postnatal brain development in the rhesus monkey. *Biol Psychiatry, 67*(10), 965-973. doi:10.1016/j.biopsych.2009.11.026

Smithers-Sheedy, H., McIntyre, S., Gibson, C., Meehan, E., Scott, H., Goldsmith, S., . . . Australian Cerebral Palsy Register, G. (2016). A special supplement: findings from the Australian Cerebral Palsy Register, birth years 1993 to 2006. *Dev Med Child Neurol, 58 Suppl 2*, 5-10. doi:10.1111/dmcn.13026

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res, 34*(Database issue), D535-539. doi:10.1093/nar/gkj109

Sun, Y. V., & Hu, Y. J. (2016). Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv Genet, 93*, 147-190. doi:10.1016/bs.adgen.2015.11.004

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res, 43*(Database issue), D447-452. doi:10.1093/nar/gku1003

Szpecht, D., Wiak, K., Braszak, A., Szymankiewicz, M., & Gadzinowski, J. (2016). Role of selected cytokines in the etiopathogenesis of intraventricular hemorrhage in preterm newborns. *Childs Nerv Syst, 32*(11), 2097-2103. doi:10.1007/s00381-016-3217-9

van Eyk CL, C. M., Gardner A, van Bon BW, Broadbent JL, Harper K, MacLennan AH, Gecz J. (2018). Analysis of 182 cerebral palsy transcriptomes points to dysregulation of trophic signalling pathways and overlap with autism.

Yoon BH, J. J., Romero R, Park KH, Gomez R, Choi JH, Kim IO. (1997). Amniotic fluid inflammatory cytokines (interleukin-6, interleukin-1beta, and tumor necrosis factor-alpha), neonatal brain white matter lesions, and cerebral palsy.

Chapter 7 Discussion and Implications

7.1 Overview

This thesis encompasses three major themes: twins, neurodevelopmental disorders and epigenetics. Together, these point to the importance of understanding the effect of genetic and non-genetic factors in neurodevelopmental disorders and also the importance of intrauterine environment for the health of the offspring. Previous studies have reported associations of DNA methylation at birth with neurodevelopmental disorders (Corley et al., 2019; Lewis & Kroll, 2018; Loke et al., 2015; Szyf, 2011). However, most of these studies only looked at affected singletons in a case-control study model or in one or two cell types. Chapters 3-5 of this thesis focus on this field of research under the major aim of investigating for associations of DNA methylation with neurodevelopmental disorders using monozygotic twins. Our studies used a statistical approach to identify evidence of associations between DNA methylation of genes and neurodevelopmental phenotypes. Our studies used linear and logistic regression analysis to analyse gene-specific methylation status of functionally important regions across the genome. The advantages and limitations of each of the studies have been discussed within the respective chapters.

7.2 Role of epigenetics in neurodevelopmental disorders

Chapter 3 involved identifying differentially methylated genes associated with cerebral palsy. We identified two gene-associated differentially methylated regions (DMRs) in *LTA* and *LIME1*, which are associated with inflammation, in a peripheral tissue, blood, at birth. Both DMRs, as well as the genes associated with top-ranking differentially methylated probes (DMPs) were associated to immune signaling and were found to be consistently differentially methylated in CP-affected individuals across all twin pairs. The most unexpected finding was identification of the differential methylation in the CP affected-twin in twelve probes within the promoter region of the *LTA* gene, encoding tumor necrosis factor-beta (TNFB), which is an important regulator of inflammation and brain development. This gene has been reported to mediate preterm birth and brain

injury (Leviton, 1993), and also elevated levels of its isoform have been indicated in cerebral palsy (J. Wu & Li, 2015). These findings provide an important insight that epigenetic marks such as DNA methylation, could be influenced by the early non-shared environment in the womb. Furthermore, my preliminary findings indicate that differences in DNA methylation are associated with CP in monozygotic twins and these changes appear to predominately associate with processes involved in immune response. Future studies will be needed to attempt to replicate these findings, both in larger cohorts of similar CP-discordant monozygotic twins, as well as in non-twin cohorts.

Chapter 4 follows a similar kind of analysis using discordant monozygotic twins, in this case, pairs in which one twin was diagnosed with idiopathic epilepsy. This study used two peripheral tissues, buccal epithelium and blood, and the cohort consisted of two types of epilepsy: generalised and focal. With a clear distinction between epilepsy phenotypes within the cohort, differences in DNA methylation patterns corresponding to each of the epilepsy types were identified. Due to this finding, DNA methylation was measured within twin pairs discordant for each epilepsy type separately. Of the top DMPs and DMRs found associating with generalised epilepsy, many genes have previously been associated with epilepsy and in neurotransmitter signaling. Similarly, top DMPs and DMRs identified within focal epilepsy discordant twin pairs associated with genes involved in cell adhesion and neuronal signaling and development. Data generated in this study revealed distinct DNA methylation differences between the two types of epilepsy. The results were also validated in another independent platform, Sequenom Epityper, with high degree of correlations between the two.

Chapter 5 follows on from Chapter 4 to further investigate epigenetics in self-limited epilepsy with centrotemporal spikes (SECTS), a subtype of focal epilepsy. The novelty of this study is the use of a longitudinal discordant monozygotic twin model, which allows to deepen our understanding of the neurobiology of this type of epilepsy. Although the sample numbers in this preliminary study is limited, the analysis revealed top DMR genes previously implicated in neurodevelopmental disorders. The *LYPD8* gene was shown to be associated both at birth and at a later time point post diagnosis of epilepsy. This suggests that perhaps there is a potential role of epigenetic mechanism in

SECTS. Interestingly, DNA methylation differences of the *LYPD8* gene within twin pairs at birth was higher than post diagnosis of epilepsy. Given SECTS is an age-dependent disorder; these results indicate a transient effect of DNA methylation, with a reduction in DNA methylation difference with time.

The final chapter of my PhD aimed to look at gene expression differences in cerebral palsy discordant twins and comparison to DNA methylation signatures within the same cohort. Interestingly, an overlap of gene ontologies enriched in immune responses was observed between top-ranking differentially expressed and methylated genes. The implication of these findings is that the dysregulation of immune signaling pathways act as a common target in cerebral palsy, and this may be an important risk factor in early postnatal life. Integrated analysis using multiple omics platforms such as methylomic and transcriptional analysis demonstrates the presence of convergent molecular signaling pathways in clinically heterogeneous disorders such as CP. This work provides a basis from which to conduct future multi-omic studies, prioritising genes for further functional and genomic studies.

7.3 The systems biology approach

'Systems biology' refers to the comprehensive assessment of biological molecules and includes genomics, transcriptomics, methylomics, proteomics and metabolomics. In the past decade, studies utilising omic platforms, mainly genomics, have enabled identification of thousands of genetic variants contributing to disease. However, with the large value of such studies published, it is increasingly being understood that most diseases results not just from genetic variations but from changes to gene regulation influenced by the environment and stochastic processes during development. Combining data from multi-omics studies has the potential to give insight into the biological pathways underlying the disease as well as provide disease markers to facilitate early diagnosis. Complex disorders such as neurodevelopmental disorders can benefit from such studies by assisting with a clear diagnosis, prognosis and progression and guiding personalised treatments.

A challenging aspect of multi-omic studies, however, is the reproducibility of results as most neurodevelopmental conditions are compounded by data heterogeneity and a lack of standard clinical assessments and clinical data linked to molecular data. With recent technological advances, omic datasets are typically large and complex. Analysis of such omic datasets requires tailored statistical approaches to answers specific biological questions. Data collection and design of the study are also important steps in integrated multi-omic studies to ensure true biological signals are detected. An initial study power calculation is necessary to ensure sufficient sample size in large-scale omic studies.

7.4 Study sample size and power of analytical analysis

This thesis employed two cohorts of twin pairs discordant for neurodevelopmental disorders to explore the role of DNA methylation in these conditions. Both twin cohorts represent small sample size of 15 discordant twin pairs. Although the power of epigenetic studies depends on large sample sizes as with other omics studies, it is known that the proportion of variance explained by single epigenetic variants is often larger than with genetic variants (Godfrey, Inskip, & Hanson, 2011; Relton et al., 2012). Typical published sample sizes have been growing from two to four digit numbers over the past few years and necessary sample size is dependent on effect size (Rakyan, Down, Balding, & Beck, 2011).

Recent evidence has shown that sample size of 500 cases and 500 controls can detect an effect of 2% with greater than 80% power in 81% of sites (Mansell et al., 2019). Epigenome-wide association studies (EWAS) involving MZ twins offer greater power than studies of singletons, especially discordant MZ co-twin studies. The comparison of discordant MZ twins offers an alternative to the traditional case–control study.

Tsai and Bell (Pei-Chien Tsai, 2015) have shown that sample sizes of 25 twin pairs or more are preferable to detect a mean effect size of 8% methylation with statistical significance of 0.05 after adjusting for multiple testing. Although this thesis includes cohorts with samples sizes of 15 twin pairs each, similar sample sizes have been used in many comparable twin studies of brain-related disorders (Dempster et al., 2014; Fisher et al., 2015; Kaut et al., 2016; C. C. Wong et al., 2014). Discordant MZ co-twin studies allows for smaller sample sizes because within-pair analysis controls for sex, age, parents, family environment, and genetics.

7.5 The discordant co-twin approach

All the studies in this thesis highlight the use of discordant monozygotic twins in understanding the epigenetics of neurodevelopmental disorders. The use of information from discordant twins was first used in a study to understand the association between smoking and lung cancer (Friberg, Cederlof, Lundman, & Olsson, 1970). The main advantage of studying monozygotic twin pairs discordant for neurodevelopmental disorders is that we can potentially tease out the causal mechanism of the disorder or casual pathways associated with the disorder. It can also help in identifying epigenetic factors that differ between cases and controls, when they are perfectly matched for age, sex, genetics and parental factors. Twins can also be used in longitudinal studies where methylation patterns at various time points can be measured, which may enable us to understand the onset of disease if not already present at birth and go beyond association studies to shed light on causation. Epigenetic data produced by such studies could contribute towards the development of predictive, diagnostic and prognostic biomarkers for complex neurodevelopmental disorders. However, some aspects to consider when studying twins is that twins in general have a shorter mean gestational age than singletons and have higher chances of preterm delivery. This becomes an issue when factors associated with gestational age at birth influence health outcomes of interest, as is the case for cerebral palsy (Arpino et al., 2005; Elovitz et al., 2006). Lower birth weight is linked to preterm birth as well, which might indicate that this factor also likely increases the risk of developing a neurodevelopmental disorder (Leviton, 1993). Some other factors, such as the process of twinning itself, are sometimes considered as risk factors (K Williams, 1996). In epilepsy, it has been shown that the process of twinning is not related to the disorder (Samuel F Berkovic, 1993). Utilising twins as a model to study neurodevelopmental disorders has many advantages but care needs to be taken when extrapolating findings in twins to singletons.

7.6 Peripheral tissues and neurodevelopmental disorders

This PhD also utilises peripheral tissues – blood and buccal epithelium, to study neurodevelopmental disorders. Although brain tissue is the biologically relevant tissue type in epigenetic analysis involving neurodevelopmental disorders, obtaining high quality brain tissue samples is often not feasible and includes its own challenges. Most often, studies that include brain tissues are obtained from post mortem samples. This makes it harder to get this tissue when studies involve living individuals. A common alternative is to use peripheral tissues such as blood, buccal or saliva, which are generally considered to be strong indicators of biological mechanisms in the brain (Masliah, Dumaop, Galasko, & Desplats, 2013; Wockner et al., 2014). However, where possible, the use of peripheral tissues must be validated by studies using post-mortem brain tissues or animal models to fully understand the effect of confounding factors such as cell composition and diverse biological pathways.

Studies have shown that peripheral tissues can be used effectively to identify biomarkers of neurodevelopmental disorders that may or may not mirror mechanisms in the brain. In many cases, blood has been used to detect differentially methylation patterns between affected and unaffected individuals in disorders such as schizophrenia (Wockner et al., 2014), bipolar disorder (Teroganova, Girshkin, Suter, & Green, 2016) and Parkinsons disease (Masliah et al., 2013). The epigenetic profiles from blood tissue have been shown to have a large overlap with the profiles detected from brain samples (Masliah et al., 2013; Wockner et al., 2014). Moreover, tissues used for predicting biomarkers do not necessarily have to be from the brain. The utility of a biomarker is the indication of a dysfunctional biological process that can be measured easily and non-invasively. The predicted biomarker could mirror a specific pathway in the brain or be a cumulative effect of a number of biological processes being affected in a disease. Although epigenetic changes are known to be tissue-specific, the response to early life exposures can be system wide such as immune cells found in circulating blood, where multiple tissues respond to early life adversity.

Recently the use of buccal tissues as an alternative to blood has been reported to be effective in EWAS. A study (Lowe R, 2013) compared the methylome of buccal versus blood and found a higher association of DNA methylation to disease phenotype in buccal cells compared to that of blood. Buccal tissues are also reported to be a better proxy to study brain related disorders as they exhibit closer similarities to brain DNA methylation patterns than blood (Lowe R, 2013; Smith et al., 2015). Many studies also identified strong correlations between the differentially methylated CpG sites identified in brain tissue compared with buccal tissues in both diseased population as well as healthy cohorts looking at childhood stress and adverse effects (Essex et al., 2013; Hagerty, Bidwell, Harlaar, & Hutchison, 2016). Buccal cells are also considered a better tissue to use to study neurodevelopmental disorders as they originate from ectodermal cell lineage that is the same as the brain, and have been used to identify potential epigenetic biomarkers for neurological outcomes (Francois, Leifert, Martins, Thomas, & Fenech, 2014; Sabine AS Langie, 2017). This is especially relevant for studies looking at early life exposures that take place before buccal and brain cells differentiate from a common germinal epithelium.

In summary, given the difficulty to sample brain tissues for studies using living individuals, the most practical solution for epigenetic studies is to make use of the appropriate peripheral tissues such as blood or buccal epithelium, or both.

7.7 Future implications for individuals and populations

With the analyses presented in this thesis, my hope is that these findings shed some light on the importance of epigenetic regulation such as DNA methylation in neurodevelopmental disorders. More importantly, DNA methylation state at birth may be able to predict risk factors or disease prior to its onset leading to critical early diagnosis and interventions. During the last few years it has become apparent that most chronic health conditions, from heart disease to psychiatric disorders, originate early in life. The exploratory studies included in this thesis serves as a platform to further carry out epigenome studies of larger cohorts and definitely warrants replication of results in non-twin population. An essential follow up to these epigenome-wide association studies to enhance the specificity and interpretability of such studies is to integrate

parallel multi-omic data within the same cohorts (Higdon et al., 2015; Wang, Tan, Tan, & Yu, 2019; Zhao et al., 2019).

Minimising the functional and social impacts of neurodevelopmental diseases would represent a significant advance for large numbers of children with these life-long conditions which not only impact on the individual child but on their families too.

Chapter 8 References

Ansel, A., Rosenzweig, J. P., Zisman, P. D., Melamed, M., & Gesundheit, B. (2016). Variation in Gene Expression in Autism Spectrum Disorders: An Extensive Review of Transcriptomic Studies. *Front Neurosci, 10*, 601. doi:10.3389/fnins.2016.00601

Arpino, C., D'Argenzio, L., Ticconi, C., Di Paolo, A., Stellin, V., Lopez, L., & Curatolo, P. (2005). Brain damage in preterm infants: etiological pathways. *Ann Ist Super Sanita, 41*(2), 229-237.

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics, 30*(10), 1363-1369. doi:10.1093/bioinformatics/btu049

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet, 25*(1), 25-29. doi:10.1038/75556

B, F. (2018). Histopathological examination of the placenta in twin pregnancies.

Badawi, N., & Keogh, J. M. (2013). Causal pathways in cerebral palsy. *J Paediatr Child Health, 49*(1), 5-8. doi:10.1111/jpc.12068

Bahado-Singh, R. O., Vishweswaraiah, S., Aydas, B., Mishra, N. K., Guda, C., & Radhakrishna, U. (2019). Deep Learning/Artificial Intelligence and Blood-Based DNA Epigenomic Prediction of Cerebral Palsy. *Int J Mol Sci, 20*(9). doi:10.3390/ijms20092075

Bailey A, L. C. A., Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study.

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res, 21*(3), 381-395. doi:10.1038/cr.2011.22

Baranzini, S. E., Mudge, J., van Velkinburgh, J. C., Khankhanian, P., Khrebtukova, I., Miller, N. A., . . . Kingsmore, S. F. (2010). Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature, 464*(7293), 1351-1356. doi:10.1038/nature08990

Barker, D. J., Gluckman, P. D., Godfrey, K. M., Harding, J. E., Owens, J. A., & Robinson, J. S. (1993). Fetal nutrition and cardiovascular disease in adult life. *Lancet, 341*(8850), 938-941.

Barker, D. J., & Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet, 1*(8489), 1077-1081.

Barker, D. J., Winter, P. D., Osmond, C., Margetts, B., & Simmonds, S. J. (1989). Weight in infancy and death from ischaemic heart disease. *Lancet, 2*(8663), 577-580.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M., & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell, 143*(3), 470-484. doi:10.1016/j.cell.2010.10.012

Bear, J. J., & Wu, Y. W. (2016). Maternal Infections During Pregnancy and Cerebral Palsy in the Child. *Pediatr Neurol, 57*, 74-79. doi:10.1016/j.pediatrneurol.2015.12.018

Belhedi, N., Perroud, N., Karege, F., Vessaz, M., Malafosse, A., & Salzmann, A. (2014). Increased CPA6 promoter methylation in focal epilepsy and in febrile seizures. *Epilepsy Res, 108*(1), 144-148. doi:10.1016/j.eplepsyres.2013.10.007

Berkovic, S. F., Howell, R. A., Hay, D. A., & Hopper, J. L. (1998). Epilepsies in twins: genetics of the major epilepsy syndromes. *Ann Neurol, 43*(4), 435-445. doi:10.1002/ana.410430405

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., . . . Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics, 98*(4), 288-295. doi:10.1016/j.ygeno.2011.07.007

Bibikova M, L. J., Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL. (2009). genome-wide DNA methylation profiling using Infinium assay.

Bibikova M, L. Z., Zhou L, Chudin E. (2006). High-throughput DNA methylation profiling using universal bead arrays.

Bird, A. (2007). Perceptions of epigenetics. *Nature, 447*(7143), 396-398. doi:10.1038/nature05913

Bittel, D. C., Kibiryeva, N., & Butler, M. G. (2007). Whole genome microarray analysis of gene expression in subjects with fragile X syndrome. *Genet Med, 9*(7), 464-472. doi:10.1097/gim.0b013e3180ca9a9a

Bittel DC, K. N., Sell SM, Strong TV, Butler MG. (2007). Whole genome microarray analysis of gene expression in Prader-Willi syndrome.

Blaze, J., & Roth, T. L. (2015). Evidence from clinical and animal model studies of the long-term and transgenerational impact of stress on DNA methylation. *Semin Cell Dev Biol, 43*, 76-84. doi:10.1016/j.semcdb.2015.04.004

Bodimeade, H. L., Whittingham, K., Lloyd, O., & Boyd, R. N. (2013). Executive functioning in children with unilateral cerebral palsy: protocol for a cross-sectional study. *BMJ Open, 3*(4). doi:10.1136/bmjopen-2012-002500

Boksa, P. (2010). Effects of prenatal infection on brain development and behavior: A review of findings from animal models.

Bonellie, S. R., Currie, D., & Chalmers, J. (2005). Comparison of risk factors for cerebral palsy in twins and singletons. *Dev Med Child Neurol, 47*(9), 587-591.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nat Rev Genet, 3*(11), 872-882. doi:10.1038/nrg932

Bottcher, L., Flachs, E. M., & Uldall, P. (2010). Attentional and executive impairments in children with spastic cerebral palsy. *Dev Med Child Neurol, 52*(2), e42-47. doi:10.1111/j.1469-8749.2009.03533.x

Burdge GC, L. K. (2010). Bridging the gap between epigenetics research and nutritional public health interventions.

Champagne, F. A., & Curley, J. P. (2009). Epigenetic mechanisms mediating the long-term effects of maternal care on development. *Neurosci Biobehav Rev, 33*(4), 593-600. doi:10.1016/j.neubiorev.2007.10.009

Chatr-aryamontri, A., Ceol, A., Palazzi, L. M., Nardelli, G., Schneider, M. V., Castagnoli, L., & Cesareni, G. (2007). MINT: the Molecular INTeraction database. *Nucleic Acids Res, 35*(Database issue), D572-574. doi:10.1093/nar/gkl950

Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., . . . Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic

CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics, 8*(2), 203-209. doi:10.4161/epi.23470
Chmielarczyk A, W.-M. J., Romaniszyn D, Adamski P, Helwich E, Lauterbach R, Pobiega M, Borszewska-Kornacka M, Gulczyńska E, Kordek A, Heczko PB. (2014). Mode of delivery and other risk factors for Escherichia coli infections in very low birth weight infants.
Chuang, J. C., & Jones, P. A. (2007). Epigenetics and microRNAs. *Pediatr Res, 61*(5 Pt 2), 24R-29R. doi:10.1203/pdr.0b013e3180457684
Clover A, F. C., Pharoah P. (2014). Cerebral Palsy. *The Lancet, 383*(9924), 1240-1249.
Corley, M. J., Vargas-Maya, N., Pang, A. P. S., Lum-Jones, A., Li, D., Khadka, V., . . . Maunakea, A. K. (2019). Epigenetic Delay in the Neurodevelopmental Trajectory of DNA Methylation States in Autism Spectrum Disorders. *Front Genet, 10*, 907. doi:10.3389/fgene.2019.00907
Crowgey, E. L., Marsh, A. G., Robinson, K. G., Yeager, S. K., & Akins, R. E. (2018). Epigenetic machine learning: utilizing DNA methylation patterns to predict spastic cerebral palsy. *BMC Bioinformatics, 19*(1), 225. doi:10.1186/s12859-018-2224-0
Cruickshank, M. N., Oshlack, A., Theda, C., Davis, P. G., Martino, D., Sheehan, P., . . . Craig, J. M. (2013). Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. *Genome Medicine, 5*. doi:ARTN 96
10.1186/gm500
Davuluri, R. V., Grosse, I., & Zhang, M. Q. (2001). Computational identification of promoters and first exons in the human genome. *Nat Genet, 29*(4), 412-417. doi:10.1038/ng780
DC Hesdorffer, G. L., EKT Benn, N Katri, G Cascino, WA Hauser. (2011). Estimating risk for developing epilepsy.
Debillon T, G.-L. C., Vérielle V, Winer N, Caillon J, Rozé JC, Gressens P. (2000). Intrauterine infection induces programmed cell death in rabbit periventricular white matter.
Dempster, E. L., Wong, C. C., Lester, K. J., Burrage, J., Gregory, A. M., Mill, J., & Eley, T. C. (2014). Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. *Biol Psychiatry, 76*(12), 977-983. doi:10.1016/j.biopsych.2014.04.013
Dickinson, J. (2005). Monoamniotic twin pregnancy: a review of contemporary practice.
Dolinoy, D. C. (2008). The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutr Rev, 66 Suppl 1*, S7-11. doi:10.1111/j.1753-4887.2008.00056.x
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. doi:10.1186/1471-2105-11-587
Eden, A., Gaudet, F., Waghmare, A., & Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. *Science, 300*(5618), 455. doi:10.1126/science.1083557

Eden, S., Hashimshony, T., Keshet, I., Cedar, H., & Thorne, A. W. (1998). DNA methylation models histone acetylation. *Nature, 394*(6696), 842. doi:10.1038/29680

Ehli, E. A., Abdellaoui, A., Hu, Y., Hottenga, J. J., Kattenberg, M., van Beijsterveldt, T., . . . Davies, G. E. (2012). De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. *Eur J Hum Genet, 20*(10), 1037-1043. doi:10.1038/ejhg.2012.49

Elovitz, M. A., Mrinalini, C., & Sammel, M. D. (2006). Elucidating the early signal transduction pathways leading to fetal brain injury in preterm birth. *Pediatr Res, 59*(1), 50-55. doi:10.1203/01.pdr.0000191141.21932.b6

Essex, M. J., Boyce, W. T., Hertzman, C., Lam, L. L., Armstrong, J. M., Neumann, S. M., & Kobor, M. S. (2013). Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence. *Child Dev, 84*(1), 58-75. doi:10.1111/j.1467-8624.2011.01641.x

Fahey, M. C., Maclennan, A. H., Kretzschmar, D., Gecz, J., & Kruer, M. C. (2017). The genetic basis of cerebral palsy. *Dev Med Child Neurol, 59*(5), 462-469. doi:10.1111/dmcn.13363

Fairbanks, L. A. (1989). Early experience and cross-generational continuity of mother-infant contact in vervet monkeys. *Dev Psychobiol, 22*(7), 669-681. doi:10.1002/dev.420220703

Fisher, H. L., Murphy, T. M., Arseneault, L., Caspi, A., Moffitt, T. E., Viana, J., . . . Mill, J. (2015). Methylomic analysis of monozygotic twins discordant for childhood psychotic symptoms. *Epigenetics, 10*(11), 1014-1023. doi:10.1080/15592294.2015.1099797

Fleiss, B., & Gressens, P. (2012). Tertiary mechanisms of brain damage: a new hope for treatment of cerebral palsy? *The Lancet Neurology, 11*(6), 556-566. doi:10.1016/s1474-4422(12)70058-3

Forsberg, L. A., Rasi, C., Razzaghian, H. R., Pakalapati, G., Waite, L., Thilbeault, K. S., . . . Dumanski, J. P. (2012). Age-related somatic structural changes in the nuclear genome of human blood cells. *Am J Hum Genet, 90*(2), 217-228. doi:10.1016/j.ajhg.2011.12.009

Francois, M., Leifert, W., Martins, R., Thomas, P., & Fenech, M. (2014). Biomarkers of Alzheimer's disease risk in peripheral tissues; focus on buccal cells. *Curr Alzheimer Res, 11*(6), 519-531. doi:10.2174/1567205011666140618103827

Friberg, L., Cederlof, R., Lundman, T., & Olsson, H. (1970). Mortality in smoking discordant monozygotic and dizygotic twins. A study on the Swedish Twin Registry. *Arch Environ Health, 21*(4), 508-513. doi:10.1080/00039896.1970.10667280

Gagnon-Bartsch JA, S. T. (2012). Using control genes to correct for unwanted variation in microarray data.

Gallinari, P., Di Marco, S., Jones, P., Pallaoro, M., & Steinkuhler, C. (2007). HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics. *Cell Res, 17*(3), 195-211. doi:10.1038/sj.cr.7310149

Ganor, Y., Freilinger, M., Dulac, O., & Levite, M. (2005). Monozygotic twins discordant for epilepsy differ in the levels of potentially pathogenic autoantibodies and cytokines. *Autoimmunity, 38*(2), 139-150. doi:10.1080/08916930500100825

Gardiner-Garden, M., & Frommer, M. (1987). CpG islands in vertebrate genomes. *J Mol Biol, 196*(2), 261-282. doi:10.1016/0022-2836(87)90689-9
Ghaemi MS, D. D., Contrepois K, Callahan B, Ngo TTM, Lee-McMullen B, Lehallier B, Robaczewska A, Mcilwain D, Rosenberg-Hasson Y, Wong RJ, Quaintance C, Culos A, Stanley N, Tanada A, Tsai A, Gaudilliere D, et al. (2018). Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy.
Giacomo Cavalli, R. P. (1998). The Drosophila Fab-7 Chromosomal Element Conveys Epigenetic Inheritance during Mitosis and Meiosis.
Gibney ER, N. C. (2010). Epigenetics and gene expression.
Girard, S., Kadhim, H., Roy, M., Lavoie, K., Brochu, M. E., Larouche, A., & Sebire, G. (2009). Role of Perinatal Inflammation in Cerebral Palsy. *Pediatric Neurology, 40*(3), 168-174. doi:10.1016/j.pediatrneurol.2008.09.016
Gluckman, P. D., Hanson, M. A., & Low, F. M. (2011). The role of developmental plasticity and epigenetics in human health. *Birth Defects Res C Embryo Today, 93*(1), 12-18. doi:10.1002/bdrc.20198
Gluckman, P. D., Hanson, M. A., & Mitchell, M. D. (2010). Developmental origins of health and disease: reducing the burden of chronic disease in the next generation. *Genome Medicine, 2*. doi:ARTN 14
10.1186/gm135
Godfrey, K. M., Costello, P. M., & Lillycrop, K. A. (2015). The developmental environment, epigenetic biomarkers and long-term health. *J Dev Orig Health Dis, 6*(5), 399-406. doi:10.1017/S204017441500121X
Godfrey, K. M., Inskip, H. M., & Hanson, M. A. (2011). The long-term effects of prenatal development on growth and metabolism. *Semin Reprod Med, 29*(3), 257-265. doi:10.1055/s-0031-1275518
Graham, H. K., Rosenbaum, P., Paneth, N., Dan, B., Lin, J. P., Damiano, D. L., . . . Lieber, R. L. (2016). Cerebral palsy. *Nat Rev Dis Primers, 2*, 15082. doi:10.1038/nrdp.2015.82
Hadjipanayis, A., Hadjichristodoulou, C., & Youroukos, S. (1997). Epilepsy in patients with cerebral palsy. *Dev Med Child Neurol, 39*(10), 659-663. doi:10.1111/j.1469-8749.1997.tb07359.x
Hagberg, H., Edwards, A. D., & Groenendaal, F. (2016). Perinatal brain damage: The term infant. *Neurobiology of Disease, 92*, 102-112. doi:10.1016/j.nbd.2015.09.011
Hagerty, S. L., Bidwell, L. C., Harlaar, N., & Hutchison, K. E. (2016). An Exploratory Association Study of Alcohol Use Disorder and DNA Methylation. *Alcohol Clin Exp Res, 40*(8), 1633-1640. doi:10.1111/acer.13138
Hales, C. N., & Barker, D. J. (2013). Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. 1992. *Int J Epidemiol, 42*(5), 1215-1222. doi:10.1093/ije/dyt133
Han, P., & Chang, C. P. (2015). Long non-coding RNA and chromatin remodeling. *RNA Biol, 12*(10), 1094-1098. doi:10.1080/15476286.2015.1063770
Han, T. R., Bang, M. S., Lim, J. Y., Yoon, B. H., & Kim, I. W. (2002). Risk factors of cerebral palsy in preterm infants. *Am J Phys Med Rehabil, 81*(4), 297-303. doi:10.1097/00002060-200204000-00011
Hanafusa, H., Morisada, N., Ishida, Y., Sakata, R., Morita, K., Miura, S., . . . Iijima, K. (2017). The smallest de novo 20q11.2 microdeletion causing intellectual

disability and dysmorphic features. *Hum Genome Var, 4*, 17050. doi:10.1038/hgv.2017.50
Hauser, W. A., Annegers, J. F., & Kurland, L. T. (1993). Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935-1984. *Epilepsia, 34*(3), 453-468.
Heijmans, B. T., Tobi, E. W., Lumey, L. H., & Slagboom, P. E. (2014). The epigenome: Archive of the prenatal environment. *Epigenetics, 4*(8), 526-531. doi:10.4161/epi.4.8.10265
Helbig, I., Scheffer, I. E., Mulley, J. C., & Berkovic, S. F. (2008). Navigating the channels and beyond: unravelling the genetics of the epilepsies. *Lancet Neurol, 7*(3), 231-245. doi:10.1016/s1474-4422(08)70039-5
Higdon, R., Earl, R. K., Stanberry, L., Hudac, C. M., Montague, E., Stewart, E., . . . Kolker, E. (2015). The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders. *OMICS, 19*(4), 197-208. doi:10.1089/omi.2015.0020
Hirose, S., Scheffer, I. E., Marini, C., De Jonghe, P., Andermann, E., Goldman, A. M., . . . Genetics Commission of the International League Against, E. (2013). SCN1A testing for epilepsy: application in clinical practice. *Epilepsia, 54*(5), 946-952. doi:10.1111/epi.12168
Ho, N. T., Furge, K., Fu, W., Busik, J., Khoo, S. K., Lu, Q., . . . Paneth, N. (2013). Gene expression in archived newborn blood spots distinguishes infants who will later develop cerebral palsy from matched controls. *Pediatr Res, 73*(4 Pt 1), 450-456. doi:10.1038/pr.2012.200
Hochberg, Y. B. a. Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.
Holliday, R., & Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science, 187*(4173), 226-232.
Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., . . . Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics, 13*, 86. doi:10.1186/1471-2105-13-86
Houtepen, L. C., Vinkers, C. H., Carrillo-Roa, T., Hiemstra, M., van Lier, P. A., Meeus, W., . . . Boks, M. P. (2016). Genome-wide DNA methylation levels and altered cortisol stress reactivity following childhood trauma in humans. *Nat Commun, 7*, 10967. doi:10.1038/ncomms10967
Huang KL, M. E., Pimenova AA, Di Narzo AF, Kapoor M, Jin SC, Harari O, Bertelsen S, Fairfax BP, Czajkowski J. (2017). A common haplotype lowers SPI1 (PU.1) expression in myeloid cells and delays age at
onset for Alzheimer's disease
Hubinont, C., Lewi, L., Bernard, P., Marbaix, E., Debieve, F., & Jauniaux, E. (2015). Anomalies of the placenta and umbilical cord in twin gestations. *Am J Obstet Gynecol, 213*(4 Suppl), S91-S102. doi:10.1016/j.ajog.2015.06.054
Hur, Y. M., & Craig, J. M. (2013). Twin registries worldwide: an important resource for scientific research. *Twin Res Hum Genet, 16*(1), 1-12. doi:10.1017/thg.2012.147
Inamdar, N. M., Ehrlich, K. C., & Ehrlich, M. (1991). Cpg Methylation Inhibits Binding of Several Sequence-Specific DNA-Binding Proteins from Pea, Wheat, Soybean and Cauliflower. *Plant Molecular Biology, 17*(1), 111-123. doi:Doi 10.1007/Bf00036811

Jaffe, A. E., Murakami, P., Lee, H., Leek, J. T., Fallin, M. D., Feinberg, A. P., & Irizarry, R. A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol, 41*(1), 200-209. doi:10.1093/ije/dyr238

Javaid N, C. S. (2017). Acetylation- and Methylation-Related Epigenetic Proteins in the Context of Their Targets.

JC, H. (2013). New terminology for mental retardation in DSM-5 and ICD-11.

Jenks, K. M., de Moor, J., & van Lieshout, E. C. (2009). Arithmetic difficulties in children with cerebral palsy are related to executive function and working memory. *J Child Psychol Psychiatry, 50*(7), 824-833. doi:10.1111/j.1469-7610.2008.02031.x

Joh, R. I., Palmieri, C. M., Hill, I. T., & Motamedi, M. (2014). Regulation of histone methylation by noncoding RNAs. *Biochim Biophys Acta, 1839*(12), 1385-1394. doi:10.1016/j.bbagrm.2014.06.006

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci, 374*(2065), 20150202. doi:10.1098/rsta.2015.0202

Joubert, B. R., den Dekker, H. T., Felix, J. F., Bohlin, J., Ligthart, S., Beckett, E., . . . London, S. J. (2016). Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. *Nat Commun, 7*, 10577. doi:10.1038/ncomms10577

K Williams, E. H., E Alberman. (1996). Cerebral palsy: effects of twinning, birthweight, and gestational age.

K. Bjornson, C. G., V. Buford, J. McLaughlin. (1998). Validity of the gross motor function measure.

Karaca E, H. T., Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, Gonzaga-Jauregui C, Erdin S, Bayram Y, Campbell IM, Hunter JV, Atik MM, Van Esch H, Yuan B, Wiszniewski W, Isikay S, Yesil G, Yuregir OO, Tug Bozdogan S, Aslan H, Aydin H, et al. (2015). Genes that affect brain structure and function identified by rare variant analyses of Mendelian neurologic disease.

Kaut, O., Schmitt, I., Tost, J., Busato, F., Liu, Y., Hofmann, P., . . . Wullner, U. (2016). Epigenome-wide DNA methylation analysis in siblings and monozygotic twins discordant for sporadic Parkinson's disease revealed different epigenetic patterns in peripheral blood mononuclear cells. *Neurogenetics*. doi:10.1007/s10048-016-0497-x

Khandaker, G. M., Zimbron, J., Lewis, G., & Jones, P. B. (2013). Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychol Med, 43*(2), 239-257. doi:10.1017/S0033291712000736

Khong JJ, B. K., Lu Y, Laurie K, Leonardos L, Baird PN, Sahebjada S, Walsh JP, Gajdatsy A, Ebeling PR, Hamblin PS, Wong R, Forehan SP, Fourlanos S, Roberts AP, Doogue M, Selva D, Montgomery GW, Macgregor S, Craig JE. (2006). Pooled Genome Wide Association Detects Association Upstream of FCRL3 With Graves' Disease.

Khoo, S. K., Dykema, K., Vadlapatla, N. M., LaHaie, D., Valle, S., Satterthwaite, D., . . . Resau, J. H. (2011). Acquiring genome-wide gene expression profiles in Guthrie card blood spots using microarrays. *Pathol Int, 61*(1), 1-6. doi:10.1111/j.1440-1827.2010.02611.x

Kjeldsen, M. J., Kyvik, K. O., Christensen, K., & Friis, M. L. (2001). Genetic and environmental factors in epilepsy: a population-based study of 11900 Danish twin pairs. *Epilepsy Res, 44*(2-3), 167-178. doi:10.1016/s0920-1211(01)00196-6

Kobow, K., & Blumcke, I. (2014). Epigenetic mechanisms in epilepsy. *Prog Brain Res, 213*, 279-316. doi:10.1016/B978-0-444-63326-2.00014-4

Kobow, K., Jeske, I., Hildebrandt, M., J., H., Hahnen, E., Buslei, R., . . . Blumcke, I. (2009). Increased reelin promoter methylation is associated with granule cell dispersion in human temporal lobe epilepsy. *Journal of Neuropathology & Experimental Neurology, 68*(4), 356-364.

Kong, S. W., Shimizu-Motohashi, Y., Campbell, M. G., Lee, I. H., Collins, C. D., Brewster, S. J., . . . Kunkel, L. M. (2013). Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. *Neurogenetics, 14*(2), 143-152. doi:10.1007/s10048-013-0363-z

Lerer I, S. M., Meiner V, Cohen T, Zlotogora J, Abeliovich D. (2005). Deletion of the ANKRD15 gene at 9p24.3 causes parent-of-origin-dependent inheritance of familial cerebral palsy.

Leu, C., Coppola, A., & Sisodiya, S. M. (2016). Progress from genome-wide association studies and copy number variant studies in epilepsy. *Curr Opin Neurol, 29*(2), 158-167. doi:10.1097/WCO.0000000000000296

Levenson, J. M., Roth, T. L., Lubin, F. D., Miller, C. A., Huang, I. C., Desai, P., . . . Sweatt, J. D. (2006). Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus. *J Biol Chem, 281*(23), 15763-15773. doi:10.1074/jbc.M511767200

Leviton, A. (1993). Preterm birth and cerebral palsy: is tumor necrosis factor the missing link? *Developmental Medicine & Child Neurology*.

Lewi L, G. L., Van Mieghem T, de Koninck P, Beck V, Medek H, Van Schoubroeck D, Devlieger R, De Catte L, Deprest J. (2010). Monochorionic Diamniotic Twin Pregnancies: Natural History and Risk Stratification.

Lewis, E. M., & Kroll, K. L. (2018). Development and disease in a dish: the epigenetics of neurodevelopmental disorders. *Epigenomics, 10*(2), 219-231. doi:10.2217/epi-2017-0113

Li, E., Beard, C., & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature, 366*(6453), 362-365. doi:10.1038/366362a0

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

Lin, C. Y., Chang, Y. C., Wang, S. T., Lee, T. Y., Lin, C. F., & Huang, C. C. (2010). Altered inflammatory responses in preterm children with cerebral palsy. *Ann Neurol, 68*(2), 204-212. doi:10.1002/ana.22049

Liu, X., Campanac, E., Cheung, H. H., Ziats, M. N., Canterel-Thouennon, L., Raygada, M., . . . Rennert, O. M. (2017). Idiopathic Autism: Cellular and Molecular Phenotypes in Pluripotent Stem Cell-Derived Neurons. *Mol Neurobiol, 54*(6), 4507-4523. doi:10.1007/s12035-016-9961-8

Liu, X., Ou, S., Xu, T., Liu, S. Y., Yuan, J. X., Huang, H., . . . Chen, Y. M. (2016). New differentially expressed genes and differential DNA methylation underlying refractory epilepsy. *Oncotarget, 7*(52), 87402-87416. doi:10.18632/oncotarget.13642

Loke, Y. J., Hannan, A. J., & Craig, J. M. (2015). The Role of Epigenetic Change in Autism Spectrum Disorders. *Front Neurol, 6*, 107. doi:10.3389/fneur.2015.00107

Lopriore E, S. M., Middeldorp JM, Klumper F, Oepkes D, Vandenbussche FP. (2006). Twin pregnancies with two separate placental masses can still be monochorionic and have vascular anastomoses.

Lowe R, G. C., Beyan H, Hawa MI, Bazeos A, Leslie RD, Montpetit A, Rakyan VK, Ramagopalan SV. (2013). Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies.

Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation.

Machnes, Z. M., Huang, T. C., Chang, P. K., Gill, R., Reist, N., Dezsi, G., . . . Szyf, M. (2013). DNA methylation mediates persistent epileptiform activity in vitro and in vivo. *PLoS One, 8*(10), e76299. doi:10.1371/journal.pone.0076299

MacLennan, A. H., Thompson, S. C., & Gecz, J. (2015). Cerebral palsy: causes, pathways, and the role of genetic variants. *Am J Obstet Gynecol, 213*(6), 779-788. doi:10.1016/j.ajog.2015.05.034

Maksimovic, J., Gagnon-Bartsch, J. A., Speed, T. P., & Oshlack, A. (2015). Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Res, 43*(16), e106. doi:10.1093/nar/gkv526

Maksimovic, J., Gordon, L., & Oshlack, A. (2012). SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol, 13*(6), R44. doi:10.1186/gb-2012-13-6-r44

Mansell, G., Gorrie-Stone, T. J., Bao, Y., Kumari, M., Schalkwyk, L. S., Mill, J., & Hannon, E. (2019). Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics, 20*(1), 366. doi:10.1186/s12864-019-5761-7

Martin, J., Cooper, M., Hamshere, M. L., Pocklington, A., Scherer, S. W., Kent, L., . . . Holmans, P. (2014). Biological overlap of attention-deficit/hyperactivity disorder and autism spectrum disorder: evidence from copy number variants. *J Am Acad Child Adolesc Psychiatry, 53*(7), 761-770 e726. doi:10.1016/j.jaac.2014.03.004

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., . . . Sun, Y. E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science, 302*(5646), 890-893. doi:10.1126/science.1090842

Masliah, E., Dumaop, W., Galasko, D., & Desplats, P. (2013). Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes. *Epigenetics, 8*(10), 1030-1038. doi:10.4161/epi.25865

McMichael, G., Bainbridge, M. N., Haan, E., Corbett, M., Gardner, A., Thompson, S., . . . MacLennan, A. H. (2015). Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. *Mol Psychiatry, 20*(2), 176-182. doi:10.1038/mp.2014.189

McNamara, H. C., Kane, S. C., Craig, J. M., Short, R. V., & Umstad, M. P. (2016). A review of the mechanisms and evidence for typical and atypical twinning. *Am J Obstet Gynecol, 214*(2), 172-191. doi:10.1016/j.ajog.2015.10.930

Meaney MJ, S. M. (2005). Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome.

Mehler, M. F. (2008). Epigenetic principles and mechanisms underlying nervous system functions in health and disease. *Prog Neurobiol, 86*(4), 305-341. doi:10.1016/j.pneurobio.2008.10.001

Menon, D. U., & Meller, V. (2014). Germ line imprinting in Drosophila: Epigenetics in search of function. *Fly, 4*(1), 48-52. doi:10.4161/fly.4.1.10881

Meyer, U., Feldon, J., & Fatemi, S. H. (2009). In-vivo rodent models for the experimental investigation of prenatal immune activation effects in neurodevelopmental brain disorders. *Neurosci Biobehav Rev, 33*(7), 1061-1079. doi:10.1016/j.neubiorev.2009.05.001

MI, R. (2010). The genetics of epilepsy--the past, the present and future.

Mikeska, T., & Craig, J. M. (2014). DNA methylation biomarkers: cancer and beyond. *Genes (Basel), 5*(3), 821-864. doi:10.3390/genes5030821

Mohandas, N., Bass-Stringer, S., Maksimovic, J., Crompton, K., Loke, Y. J., Walstab, J., . . . Craig, J. M. (2018). Epigenome-wide analysis in newborn blood spots from monozygotic twins discordant for cerebral palsy reveals consistent regional differences in DNA methylation. *Clin Epigenetics, 10*, 25. doi:10.1186/s13148-018-0457-4

Mohandas, T., Sparkes, R. S., & Shapiro, L. J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science, 211*(4480), 393-396.

Morgan, H. D., Sutherland, H. G., Martin, D. I., & Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet, 23*(3), 314-318. doi:10.1038/15490

Moshe Stavsky, O. M., Salvatore Andrea Mastrolia, Shirley Greenbaum, Nandor Gabor Than, Offer Erez. (2017). Cerebral Palsy—Trends in Epidemiology and Recent Development in Prenatal Mechanisms of Disease, Treatment, and Prevention.

Murphy DJ, H. P., Johnson A. (1997). Neonatal risk factors for cerebral palsy in very preterm babies: case-control study.

Murray, C. J., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., . . . Memish, Z. A. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet, 380*(9859), 2197-2223. doi:10.1016/S0140-6736(12)61689-4

Nardone, S., Sams, D. S., Reuveni, E., Getselter, D., Oron, O., Karpuj, M., & Elliott, E. (2014). DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Transl Psychiatry, 4*, e433. doi:10.1038/tp.2014.70

Neel, J. (1962). Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"?

Nelson, E. D., Kavalali, E. T., & Monteggia, L. M. (2008). Activity-dependent suppression of miniature neurotransmission through the regulation of DNA methylation. *J Neurosci, 28*(2), 395-406. doi:10.1523/jneurosci.3796-07.2008

Nelson, K. B., Dambrosia, J. M., Iovannisci, D. M., Cheng, S., Grether, J. K., & Lammer, E. (2005). Genetic polymorphisms and cerebral palsy in very preterm

infants. *Pediatric Research, 57*(4), 494-499. doi:10.1203/01.Pdr.0000156477.00386.E7
Nelson KB, E. J. (1984). Obstetric Complications as Risk Factors for Cerebral Palsy or Seizure Disorders.
Nishimura, Y., Martin, C. L., Vazquez-Lopez, A., Spence, S. J., Alvarez-Retuerto, A. I., Sigman, M., . . . Geschwind, D. H. (2007). Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum Mol Genet, 16*(14), 1682-1698. doi:10.1093/hmg/ddm116
Novak, I., Hines, M., Goldsmith, S., & Barclay, R. (2012). Clinical prognostic messages from a systematic review on cerebral palsy. *Pediatrics, 130*(5), e1285-1312. doi:10.1542/peds.2012-0924
O'Callaghan, M. E., MacLennan, A. H., Gibson, C. S., McMichael, G. L., Haan, E. A., Broadbent, J. L., . . . Australian Collaborative Cerebral Palsy Research, G. (2011). Epidemiologic associations with cerebral palsy. *Obstet Gynecol, 118*(3), 576-582. doi:10.1097/AOG.0b013e31822ad2dc
O'Callaghan, M. E., MacLennan, A. H., Haan, E. A., Dekker, G., & South Australian Cerebral Palsy Research, G. (2009). The genomic basis of cerebral palsy: a HuGE systematic literature review. *Hum Genet, 126*(1), 149-172. doi:10.1007/s00439-009-0638-5
Ohlsson, R., & Kanduri, C. (2002). New twists on the epigenetics of CpG islands. *Genome Res, 12*(4), 525-526. doi:10.1101/gr.18002
Ollikainen, M., Smith, K. R., Joo, E. J., Ng, H. K., Andronikos, R., Novakovic, B., . . . Craig, J. M. (2010). DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum Mol Genet, 19*(21), 4176-4188. doi:10.1093/hmg/ddq336
Oskoui, M., Coutinho, F., Dykeman, J., Jette, N., & Pringsheim, T. (2013). An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. *Dev Med Child Neurol, 55*(6), 509-519. doi:10.1111/dmcn.12080
Oskoui, M., Gazzellone, M. J., Thiruvahindrapuram, B., Zarrei, M., Andersen, J., Wei, J., . . . Scherer, S. W. (2015). Clinically relevant copy number variations detected in cerebral palsy. *Nat Commun, 6*, 7949. doi:10.1038/ncomms8949
P. Rosenbaum, N. P., A. Leviton, M. Goldstein, B. Martin. (2007). A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl., 109*, 8-14.
Pei-Chien Tsai, J. T. B. (2015). Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *International Journal of Epidemiology*.
Peri S, N. J., Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N, Suresh S, Rashmi BP, Shanker K, Padma N, Niranjan V, Harsha HC, Talreja N, Vrushabendra BM, Ramya MA, Yatish AJ, Joy M, Shivashankar HN. (2004). Human protein reference database as a discovery resource for proteomics.
Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., R, V. L., . . . Molloy, P. L. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin, 8*, 6. doi:10.1186/1756-8935-8-6

Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature, 465*(7299), 721-727. doi:10.1038/nature09230
Petrovski, S., & Kwan, P. (2013). Unraveling the genetics of common epilepsies: approaches, platforms, and caveats. *Epilepsy Behav, 26*(3), 229-233. doi:10.1016/j.yebeh.2012.09.007
Pharoah, P. O., & Cooke, T. (1996). Cerebral palsy and multiple births. *Arch Dis Child Fetal Neonatal Ed, 75*(3), F174-177.
Pharoah, P. O., & Dundar, Y. (2009). Monozygotic twinning, cerebral palsy and congenital anomalies. *Hum Reprod Update, 15*(6), 639-648. doi:10.1093/humupd/dmp019
Phipson, B., Maksimovic, J., & Oshlack, A. (2016). missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics, 32*(2), 286-288. doi:10.1093/bioinformatics/btv560
Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., . . . Clark, S. J. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol, 17*(1), 208. doi:10.1186/s13059-016-1066-1
Pinto-Martin JA, R. S., Cnaan A, Holzman C, Susser MW, Paneth N. (1995). Cranial ultrasound prediction of disabling and nondisabling cerebral palsy at age two in a low birth weight population.
Pulido Fontes, L., Quesada Jimenez, P., & Mendioroz Iriarte, M. (2015). Epigenetics and epilepsy. *Neurologia, 30*(2), 111-118. doi:10.1016/j.nrl.2014.03.012
Qureshi, I. A., & Mehler, M. F. (2010). Epigenetic mechanisms underlying human epileptic disorders and the process of epileptogenesis. *Neurobiol Dis, 39*(1), 53-60. doi:10.1016/j.nbd.2010.02.005
Rakyan, V. K., Down, T. A., Balding, D. J., & Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nat Rev Genet, 12*(8), 529-541. doi:10.1038/nrg3000
Relton, C. L., Groom, A., St Pourcain, B., Sayers, A. E., Swan, D. C., Embleton, N. D., . . . Davey Smith, G. (2012). DNA methylation patterns in cord blood DNA and body size in childhood. *PLoS One, 7*(3), e31821. doi:10.1371/journal.pone.0031821
Reutens, D. C., Howell, R. A., Gebert, K. E., & Berkovic, S. F. (1992). Validation of a questionnaire for clinical seizure diagnosis. *Epilepsia, 33*(6), 1065-1071.
Riggs, A. (1975). X inactivation, differentiation and DNA methylation.
Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res, 43*(7), e47. doi:10.1093/nar/gkv007
Rosenbaum, P., Paneth, N., Leviton, A., Goldstein, M., Bax, M., Damiano, D., . . . Jacobsson, B. (2007). A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl, 109*, 8-14.
Russell DJ, R. P., Cadman DT, Gowland C, Hardy S, Jarvis S. (1989). The gross motor function measure a means to evaluate the effects of physical therapy.
Ruth Pidsley, C. C. Y. W., Manuela Volta, Katie Lunnon, Jonathan Mill, Leonard C Schalkwyk. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data.
Ryley Parrish, R., Albertson, A. J., Buckingham, S. C., Hablitz, J. J., Mascia, K. L., Davis Haselden, W., & Lubin, F. D. (2013). Status epilepticus triggers early and late alterations in brain-derived neurotrophic factor and NMDA glutamate

receptor Grin2b DNA methylation levels in the hippocampus. *Neuroscience, 248*, 602-619. doi:10.1016/j.neuroscience.2013.06.029
Sabine AS Langie, M. M., Ken Declerck, Gudrun Koppen, Lode Godderis, Wim Vanden Berghe, Stacy Drury, Patrick De Boever. (2017). Salivary DNA Methylation Profiling: Aspects to Consider for Biomarker Identification.
Sahu, M., & Prasuna, J. G. (2016). Twin Studies: A Unique Epidemiological Tool. *Indian J Community Med, 41*(3), 177-182. doi:10.4103/0970-0218.183593
Salvador-Carulla L, R. G., Vaez-Azizi LM, Cooper SA, Martinez-Leal R, Bertelli M, Adnams C, Cooray S, Deb S, Akoury-Dirani L, Girimaji SC, Katz G, Kwok H, Luckasson R, Simeonsson R, Walsh C, Munir K, Saxena S. (2011). Intellectual developmental disorders: towards a new name, definition and framework for "mental retardation/intellectual disability" in ICD-11.
Samuel F Berkovic, A. H., David A Hay, John L Hopper. (1993). Twin birth is not a risk factor for seizures.
Schoendorfer, N. C., Obeid, R., Moxon-Lester, L., Sharp, N., Vitetta, L., Boyd, R. N., & Davies, P. S. (2012). Methylation capacity in children with severe cerebral palsy. *Eur J Clin Invest, 42*(7), 768-776. doi:10.1111/j.1365-2362.2011.02644.x
Segel, R., Ben-Pazi, H., Zeligson, S., Fatal-Valevski, A., Aran, A., Gross-Tsur, V., . . . Levy-Lahad, E. (2015). Copy number variations in cryptogenic cerebral palsy. *Neurology, 84*(16), 1660-1668. doi:10.1212/WNL.0000000000001494
Sergushichev, A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation.
Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte, D., Jr., & Mello, C. C. (2013). The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Dev Cell, 27*(6), 656-663. doi:10.1016/j.devcel.2013.11.014
Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res, 13*(11), 2498-2504. doi:10.1101/gr.1239303
Shen J, M. D. (2014). Advances in Genetic Diagnosis of Autism Spectrum Disorders.
Short, S. J., Lubach, G. R., Karasin, A. I., Olsen, C. W., Styner, M., Knickmeyer, R. C., . . . Coe, C. L. (2010). Maternal influenza infection during pregnancy impacts postnatal brain development in the rhesus monkey. *Biol Psychiatry, 67*(10), 965-973. doi:10.1016/j.biopsych.2009.11.026
Siegmund, K. D. (2011). Statistical approaches for the analysis of DNA methylation microarray data. *Hum Genet, 129*(6), 585-595. doi:10.1007/s00439-011-0993-x
Sillanpaa, M., Jalava, M., Kaleva, O., & Shinnar, S. (1998). Long-term prognosis of seizures with onset in childhood. *N Engl J Med, 338*(24), 1715-1722. doi:10.1056/NEJM199806113382402
Simmen, M. W., Leitgeb, S., Charlton, J., Jones, S. J., Harris, B. R., Clark, V. H., & Bird, A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. *Science, 283*(5405), 1164-1167. doi:10.1126/science.283.5405.1164
Simmons, R. (2011). Epigenetics and maternal nutrition: nature v. nurture.

Singhi, P., Jagirdar, S., Khandelwal, N., & Malhi, P. (2003). Epilepsy in children with cerebral palsy. *J Child Neurol, 18*(3), 174-179. doi:10.1177/08830738030180030601

Smith, A. K., Kilaru, V., Klengel, T., Mercer, K. B., Bradley, B., Conneely, K. N., . . . Binder, E. B. (2015). DNA Extracted From Saliva for Methylation Studies of Psychiatric Traits: Evidence for Tissue Specificity and Relatedness to Brain. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 168*(1), 36-44. doi:10.1002/ajmg.b.32278

Smithers-Sheedy, H., McIntyre, S., Gibson, C., Meehan, E., Scott, H., Goldsmith, S., . . . Australian Cerebral Palsy Register, G. (2016). A special supplement: findings from the Australian Cerebral Palsy Register, birth years 1993 to 2006. *Dev Med Child Neurol, 58 Suppl 2*, 5-10. doi:10.1111/dmcn.13026

Speed D, O. B. T., Palotie A, Shkura K, Marson AG, Balding DJ, Johnson MR. (2014). Describing the genetic architecture of epilepsy through heritability analysis.

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res, 34*(Database issue), D535-539. doi:10.1093/nar/gkj109

Steinlein OK, M. J., Propping P, Wallace RH, Phillips HA, Sutherland GR, Scheffer IE, Berkovic SF. (1995). A missense mutation in the neuronal nicotinic acetylcholine receptor alpha 4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy.

Suchiman, H. E., Slieker, R. C., Kremer, D., Slagboom, P. E., Heijmans, B. T., & Tobi, E. W. (2015). Design, measurement and processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. *Front Genet, 6*, 287. doi:10.3389/fgene.2015.00287

Sukhov, A., Wu, Y., Xing, G., Smith, L. H., & Gilbert, W. M. (2012). Risk factors associated with cerebral palsy in preterm infants. *J Matern Fetal Neonatal Med, 25*(1), 53-57. doi:10.3109/14767058.2011.564689

Sun, S., Huang, Y. W., Yan, P. S., Huang, T. H., & Lin, S. (2011). Preprocessing differential methylation hybridization microarray data. *BioData Min, 4*, 13. doi:10.1186/1756-0381-4-13

Sun, Y. V., & Hu, Y. J. (2016). Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv Genet, 93*, 147-190. doi:10.1016/bs.adgen.2015.11.004

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res, 43*(Database issue), D447-452. doi:10.1093/nar/gku1003

Szpecht, D., Wiak, K., Braszak, A., Szymankiewicz, M., & Gadzinowski, J. (2016). Role of selected cytokines in the etiopathogenesis of intraventricular hemorrhage in preterm newborns. *Childs Nerv Syst, 32*(11), 2097-2103. doi:10.1007/s00381-016-3217-9

Szyf, M. (2011). DNA methylation, the early-life social environment and behavioral disorders.

Tamaru H, Z. X., McMillen D, Singh PB, Nakayama J, Grewal SI, Allis CD, Cheng X, Selker EU. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa.

Teroganova, N., Girshkin, L., Suter, C. M., & Green, M. J. (2016). DNA methylation in peripheral tissue of schizophrenia and bipolar disorder: a systematic review. *Bmc Genetics, 17*. doi:ARTN 27
10.1186/s12863-016-0332-2
Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., & Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics, 29*(2), 189-196. doi:10.1093/bioinformatics/bts680
Tollanes, M. C., Wilcox, A. J., Lie, R. T., & Moster, D. (2014). Familial risk of cerebral palsy: population based cohort study. *BMJ, 349*, g4294. doi:10.1136/bmj.g4294
Touleimat N, T. J. (2012). Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation.
Urdinguio, R. G., Sanchez-Mut, J. V., & Esteller, M. (2009). Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol, 8*(11), 1056-1072. doi:10.1016/S1474-4422(09)70262-5
Vadlamudi L, D. L., Lawrence KM, Iona X, McMahon JM, Murrell W, Mackay-Sim A, Scheffer IE, Berkovic SF. (2010). Timing of de novo mutagenesis--a twin study of sodium-channel mutations.
Vadlamudi, L., Milne, R. L., Lawrence, K., Heron, S. E., Eckhaus, J., Keay, D., . . . Berkovic, S. F. (2014). Genetics of epilepsy: The testimony of twins in the molecular era. *Neurology, 83*(12), 1042-1048. doi:10.1212/WNL.0000000000000790
Vaissière T, S. C., Herceg Z. (2008). Epigenetic interplay between histone modifications and DNA methylation in gene silencing.
van Dongen, J., Slagboom, P. E., Draisma, H. H., Martin, N. G., & Boomsma, D. I. (2012). The continuing value of twin studies in the omics era. *Nat Rev Genet, 13*(9), 640-653. doi:10.1038/nrg3243
van Eyk CL, C. M., Gardner A, van Bon BW, Broadbent JL, Harper K, MacLennan AH, Gecz J. (2018). Analysis of 182 cerebral palsy transcriptomes points to dysregulation of trophic signalling pathways and overlap with autism.
Veenma, D., Brosens, E., de Jong, E., van de Ven, C., Meeussen, C., Cohen-Overbeek, T., . . . de Klein, A. (2012). Copy number detection in discordant monozygotic twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) cohorts. *Eur J Hum Genet, 20*(3), 298-304. doi:10.1038/ejhg.2011.194
Verdel, A., Vavasseur, A., Le Gorrec, M., & Touat-Todeschini, L. (2009). Common themes in siRNA-mediated epigenetic silencing pathways. *Int J Dev Biol, 53*(2-3), 245-257. doi:10.1387/ijdb.082691av
Wada, T., Kikuchi, J., Nishimura, N., Shimizu, R., Kitamura, T., & Furukawa, Y. (2009). Expression levels of histone deacetylases determine the cell fate of hematopoietic progenitors. *J Biol Chem, 284*(44), 30673-30683. doi:10.1074/jbc.M109.042242
Waddington, C. H. (1956). Genetic Assimilation of the Bithorax Phenotype. *Evolution, 10*(1), 1-13. doi:Doi 10.2307/2406091
Wallace, R. H., Scheffer, I. E., Barnett, S., Richards, M., Dibbens, L., Desai, R. R., . . . Berkovic, S. F. (2001). Neuronal sodium-channel alpha1-subunit mutations

in generalized epilepsy with febrile seizures plus. *Am J Hum Genet, 68*(4), 859-865. doi:10.1086/319516
Wang, Z. T., Tan, C. C., Tan, L., & Yu, J. T. (2019). Systems biology and gene networks in Alzheimer's disease. *Neurosci Biobehav Rev, 96*, 31-44. doi:10.1016/j.neubiorev.2018.11.007
Weaver, I. C., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., . . . Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nat Neurosci, 7*(8), 847-854. doi:10.1038/nn1276
Weaver, I. C., Champagne, F. A., Brown, S. E., Dymov, S., Sharma, S., Meaney, M. J., & Szyf, M. (2005). Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: altering epigenetic marking later in life. *J Neurosci, 25*(47), 11045-11054. doi:10.1523/JNEUROSCI.3652-05.2005
Weber, M., & Schubeler, D. (2007). Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol, 19*(3), 273-280. doi:10.1016/j.ceb.2007.04.011
Wockner, L. F., Noble, E. P., Lawford, B. R., Young, R. M., Morris, C. P., Whitehall, V. L., & Voisey, J. (2014). Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Transl Psychiatry, 4*, e339. doi:10.1038/tp.2013.111
Wong, A. H., Gottesman, II, & Petronis, A. (2005). Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum Mol Genet, 14 Spec No 1*, R11-18. doi:10.1093/hmg/ddi116
Wong, C. C., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., . . . Mill, J. (2014). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry, 19*(4), 495-503. doi:10.1038/mp.2013.41
Wu, C., & Morris, J. R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science, 293*(5532), 1103-1105. doi:10.1126/science.293.5532.1103
Wu, J., & Li, X. (2015). Plasma Tumor Necrosis Factor-alpha (TNF-α) Levels Correlate with Disease Severity in Spastic Diplegia, Triplegia, and Quadriplegia in Children with Cerebral Palsy. *Medical Science Monitor, 21*, 3868-3874. doi:10.12659/msm.895400
Xinsheng Nan, F. J. C., Adrian Bird. (1997). MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin.
Yao, Q., Chen, Y., & Zhou, X. (2019). The roles of microRNAs in epigenetic regulation. *Curr Opin Chem Biol, 51*, 11-17. doi:10.1016/j.cbpa.2019.01.024
Yoon BH, J. J., Romero R, Park KH, Gomez R, Choi JH, Kim IO. (1997). Amniotic fluid inflammatory cytokines (interleukin-6, interleukin-1beta, and tumor necrosis factor-alpha), neonatal brain white matter lesions, and cerebral palsy.
Zhang, L., Vielle, A., Espinosa, S., & Zhao, R. (2019). RNAs in the spliceosome: Insight from cryoEM structures. *Wiley Interdiscip Rev RNA, 10*(3), e1523. doi:10.1002/wrna.1523
Zhao, Y., Jhamb, D., Shu, L., Arneson, D., Rajpal, D. K., & Yang, X. (2019). Multi-omics integration reveals molecular networks and regulators of psoriasis. *BMC Syst Biol, 13*(1), 8. doi:10.1186/s12918-018-0671-x
Zhu, Q., Wang, L., Zhang, Y., Zhao, F. H., Luo, J., Xiao, Z., . . . Wang, X. F. (2012). Increased expression of DNA methyltransferase 1 and 3a in human

temporal lobe epilepsy. *J Mol Neurosci, 46*(2), 420-426. doi:10.1007/s12031-011-9602-7

Chapter 9 Appendices

Supplementary Material for Chapter 3

Supplementary material for:

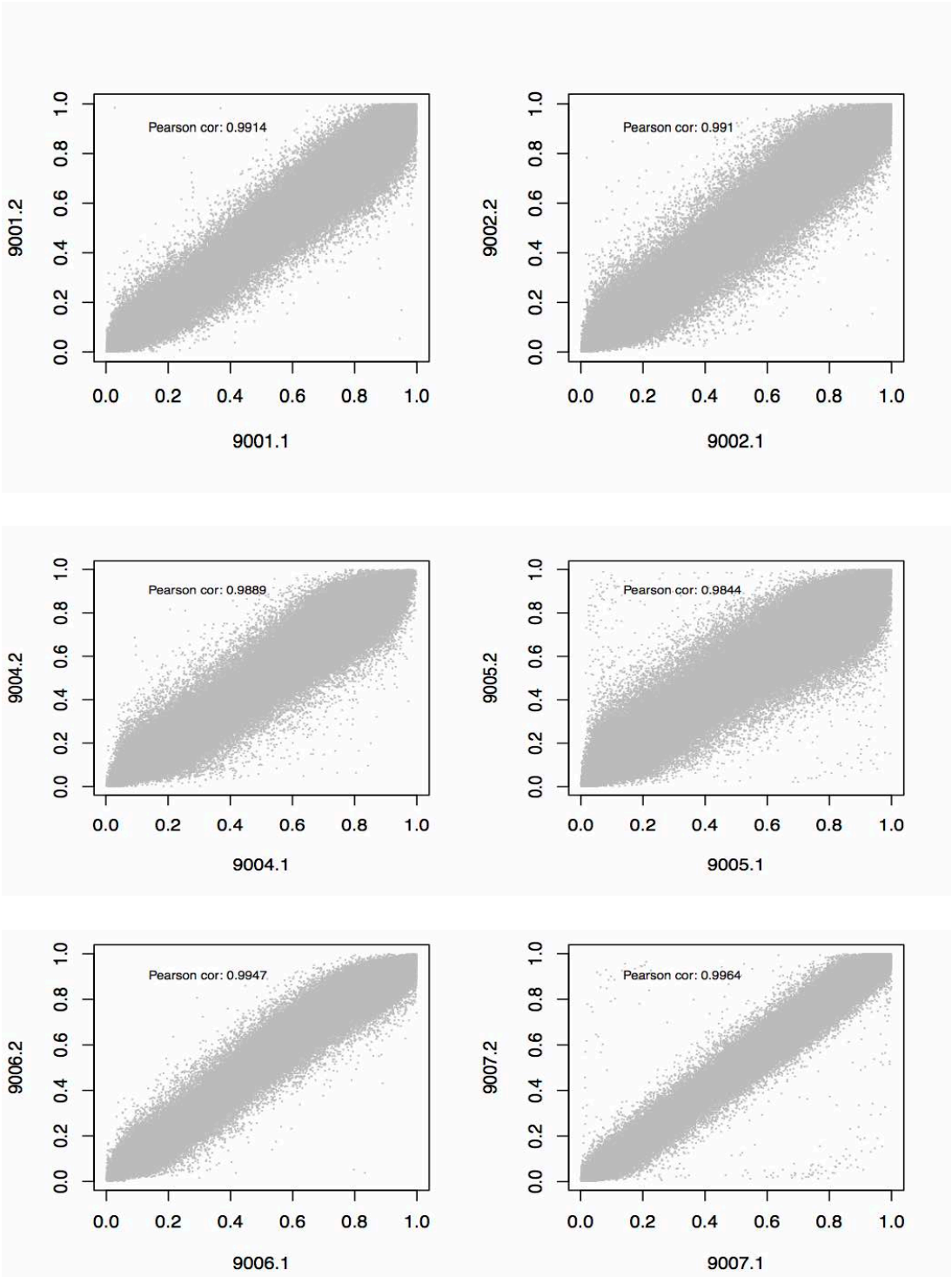
Epigenome-wide analysis in newborn blood spots from monozygotic twins discordant for cerebral palsy reveals consistent regional differences in DNA methylation

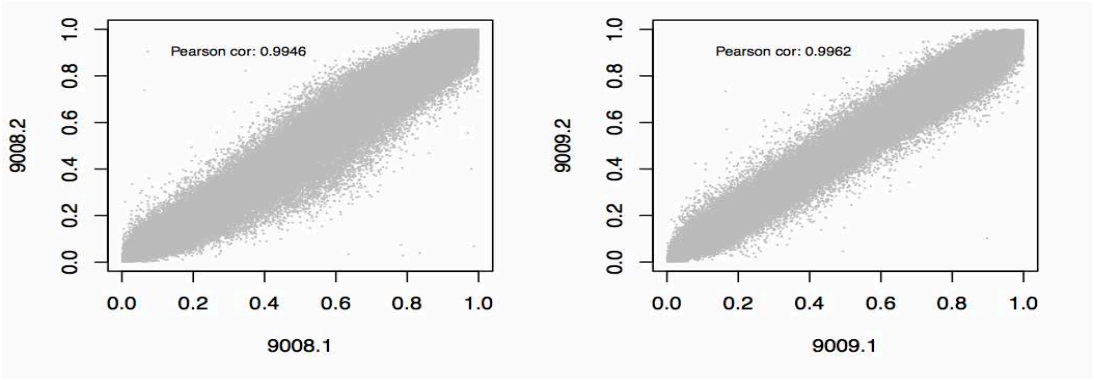
Namitha Mohandas, Sebastian Bass-Stringer, Jovana Maksimovic, Kylie Crompton, Yuk J. Loke, Janet Walstab, Susan M. Reid, David J. Amor, Dinah Reddihough, Jeffrey M Craig

Appendix 1: Primer sequences used in site-specific validation using MassArray EpiTYPER

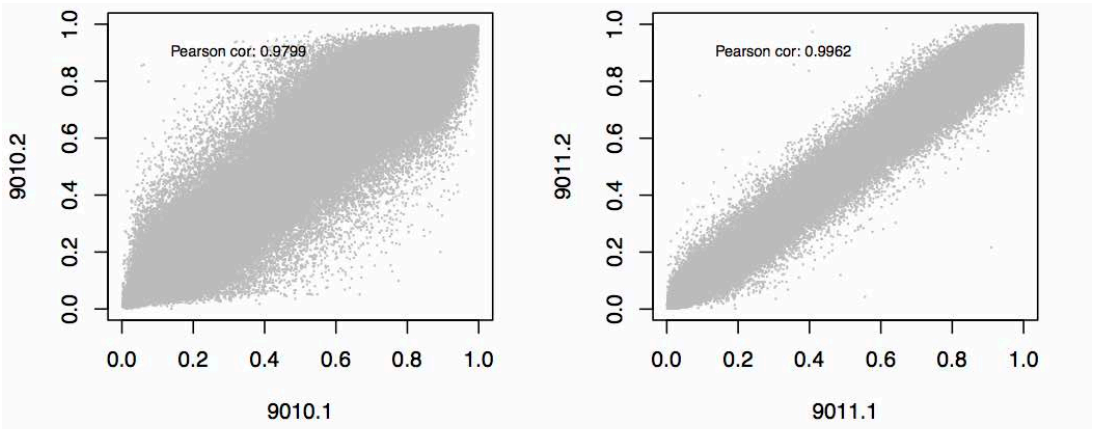
Amplicon	Left Primer	Right Primer
LIME1	GATTTGGTTTTGTGGGTTTGG	CCAAAAAAATCTTTATACAACCCTC
LTA	GGGGATTTAGGTAGTAGGTGTAGGA	CTTCTAAACCCTAAAAACTTCCCC

Appendix 2: Scatter plots of genome-wide DNA methylation discordance within twin groups.

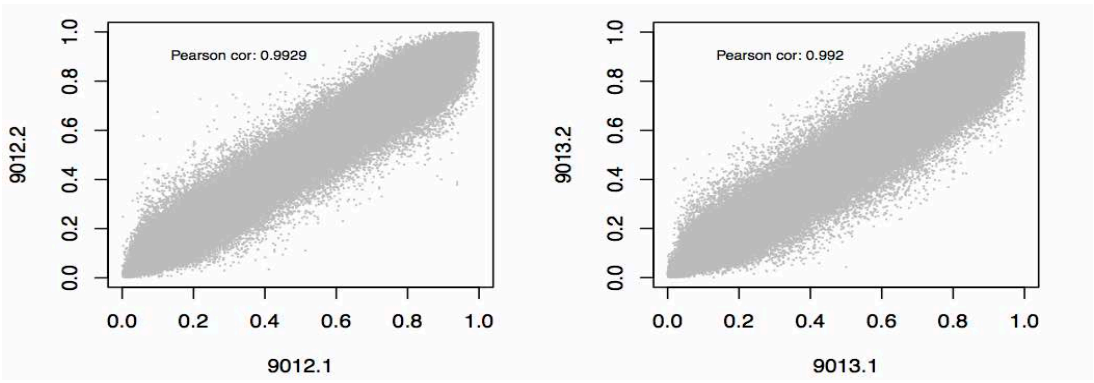



Pearson cor: 0.9946
9008.2
9008.1
0.0
0.2
0.4
0.6
0.8
1.0

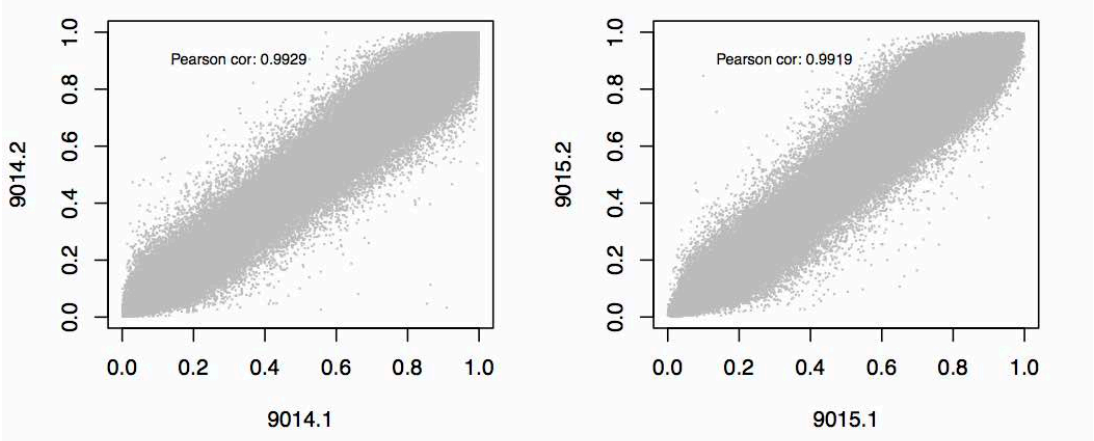
Pearson cor: 0.9962
9009.2
9009.1
0.0
0.2
0.4
0.6
0.8
1.0


Pearson cor: 0.9799
9010.2
9010.1
0.0
0.2
0.4
0.6
0.8
1.0

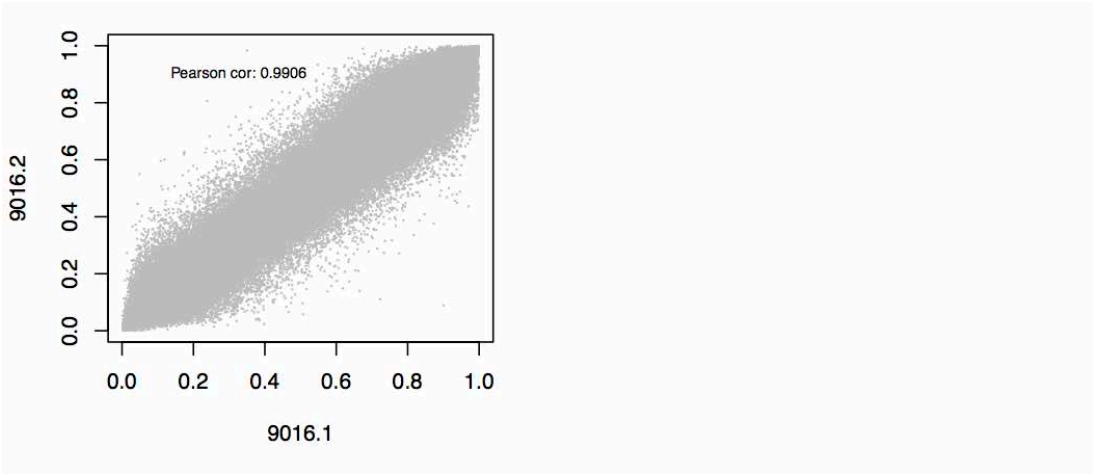
Pearson cor: 0.9962
9011.2
9011.1
0.0
0.2
0.4
0.6
0.8
1.0


Pearson cor: 0.9929
9012.2
9012.1
0.0
0.2
0.4
0.6
0.8
1.0

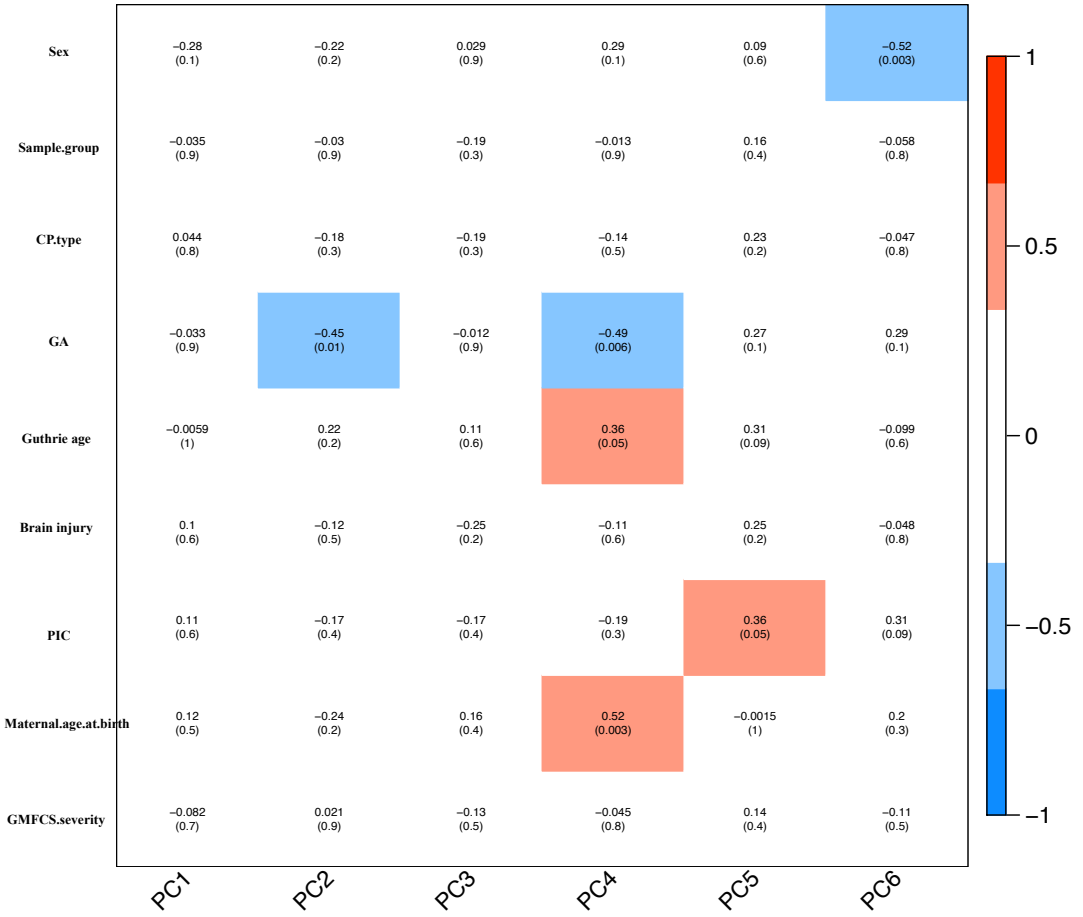
Pearson cor: 0.992
9013.2
9013.1
0.0
0.2
0.4
0.6
0.8
1.0


Pearson cor: 0.9929
9014.2
9014.1
0.0
0.2
0.4
0.6
0.8
1.0

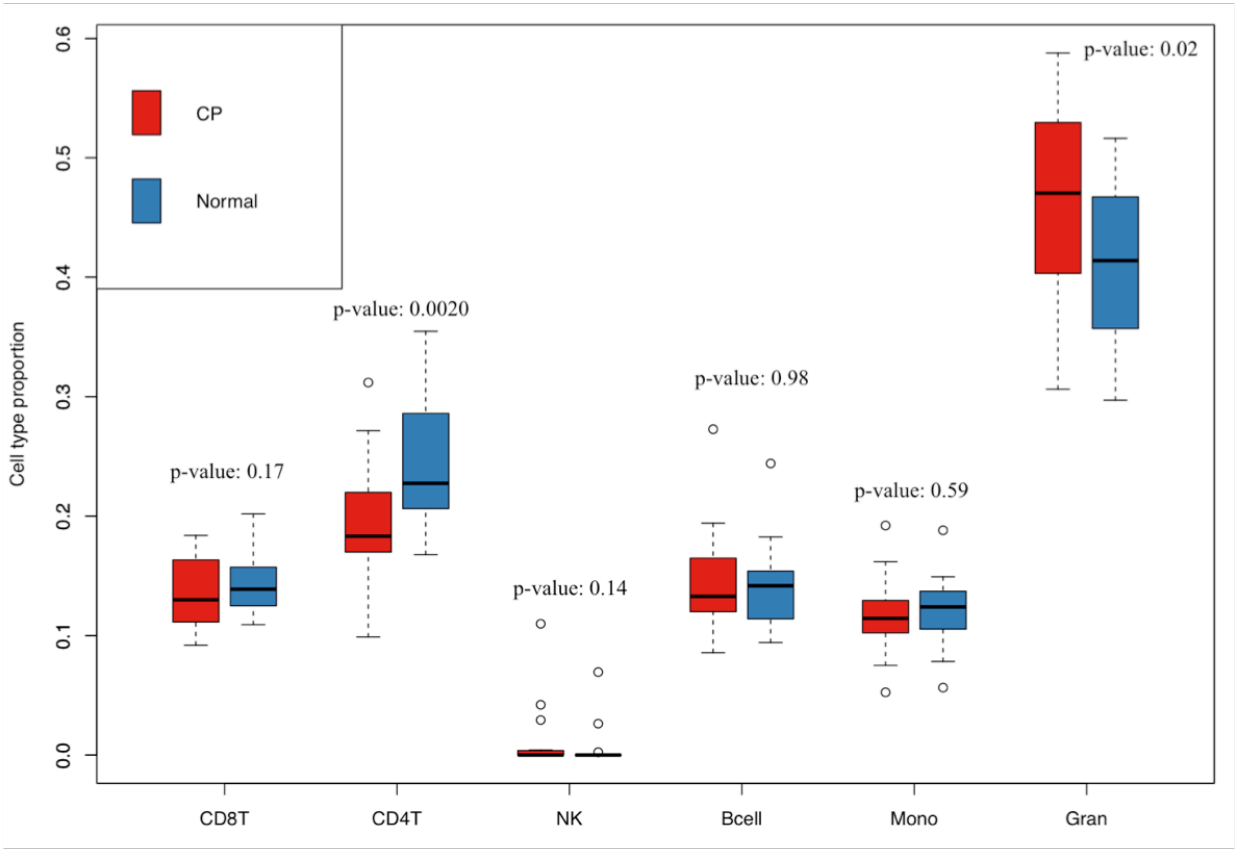
Pearson cor: 0.9919
9015.2
9015.1
0.0
0.2
0.4
0.6
0.8
1.0


Pearson cor: 0.9906
9016.2
9016.1
0.0
0.2
0.4
0.6
0.8
1.0

Appendix 3: Heat map of the associations between the six largest principal components and specified covariates. The heat map provides a score of the strength of the association between DNA methylation (using M-values) and each covariate, with positive and negative correlations ranging according to magnitude (red positive, blue negative). The values in brackets for each association represent the p-value of the correlation. Of the six significant ($p<0.05$) associations, all are weak (correlation < 0.6). Abbreviations: CP, cerebral palsy; PC, principal component; PIC, person in charge of performing DNA extraction; GA, gestational age; GMFCS, gross motor function classification system; Guthrie age, age in postnatal days when Guthrie card was made.



Appendix 4: Comparison of cell type composition of cerebral palsy cases versus normal individuals. CD8T and CD4T: Cytotoxic T cells; NK: natural-killer cells; B-cell: B cell or B lymphocytes; Mono: monocytes; Gran: granulocytes.



Appendix 5: Gene Ontology (GO) analysis for top-ranked 1000 DMPs ranked by P-value.

GO ID	GO Function	Ontology	P. Value
GO:0098609	cell-cell adhesion	BP	5.60E-06
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	BP	1.75E-05
GO:0042129	regulation of T cell proliferation	BP	2.30E-05
GO:0042098	T cell proliferation	BP	3.12E-05
GO:0050776	regulation of immune response	BP	3.78E-05
GO:0046649	lymphocyte activation	BP	4.16E-05
GO:0050863	regulation of T cell activation	BP	8.47E-05
GO:0070661	leukocyte proliferation	BP	1.13E-04
GO:0007159	leukocyte cell-cell adhesion	BP	1.20E-04
GO:0002768	immune response-regulating cell surface receptor signaling pathway	BP	1.33E-04
GO:0030098	lymphocyte differentiation	BP	1.35E-04
GO:1903037	regulation of leukocyte cell-cell adhesion	BP	1.53E-04
GO:0045321	leukocyte activation	BP	1.56E-04
GO:0001775	cell activation	BP	1.58E-04
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	BP	1.70E-04
GO:0070663	regulation of leukocyte proliferation	BP	1.71E-04
GO:0042110	T cell activation	BP	2.05E-04
GO:0070489	T cell aggregation	BP	2.05E-04
GO:0071593	lymphocyte aggregation	BP	2.28E-04
GO:0022610	biological adhesion	BP	2.49E-04
GO:0050670	regulation of lymphocyte proliferation	BP	2.73E-04
GO:0045601	regulation of endothelial cell differentiation	BP	3.11E-04
GO:0032944	regulation of mononuclear cell proliferation	BP	3.27E-04
GO:0070486	leukocyte aggregation	BP	3.31E-04
GO:0007155	cell adhesion	BP	3.43E-04
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	BP	3.44E-04
GO:0044459	plasma membrane part	CC	3.65E-04
GO:0048856	anatomical structure development	BP	4.15E-04
GO:0022407	regulation of cell-cell adhesion	BP	4.42E-04
GO:0098552	side of membrane	CC	4.44E-04
GO:0044767	single-organism developmental process	BP	4.48E-04
GO:0034110	regulation of homotypic cell-cell adhesion	BP	4.54E-04
GO:0050865	regulation of cell activation	BP	4.55E-04
GO:0002703	regulation of leukocyte mediated immunity	BP	5.07E-04
GO:0015732	prostaglandin transport	BP	5.10E-04
GO:0016337	single organismal cell-cell adhesion	BP	5.13E-04
GO:0031329	regulation of cellular catabolic process	BP	5.15E-04
GO:0032502	developmental process	BP	5.20E-04
GO:0046640	regulation of alpha-beta T cell proliferation	BP	5.35E-04
GO:0032729	positive regulation of interferon-gamma production	BP	5.36E-04
GO:0098562	cytoplasmic side of membrane	CC	5.49E-04
GO:0071363	cellular response to growth factor stimulus	BP	6.80E-04
GO:0046651	lymphocyte proliferation	BP	6.81E-04
GO:1901550	regulation of endothelial cell development	BP	6.97E-04
GO:1903140	regulation of establishment of endothelial barrier	BP	6.97E-04
GO:0002764	immune response-regulating signaling pathway	BP	6.98E-04
GO:0009894	regulation of catabolic process	BP	7.52E-04
GO:0002449	lymphocyte mediated immunity	BP	8.29E-04
GO:0032943	mononuclear cell proliferation	BP	8.37E-04
GO:0034109	homotypic cell-cell adhesion	BP	8.38E-04
GO:0061484	hematopoietic stem cell homeostasis	BP	8.49E-04
GO:0046634	regulation of alpha-beta T cell activation	BP	8.63E-04
GO:0051249	regulation of lymphocyte activation	BP	8.64E-04
GO:0070848	response to growth factor	BP	8.69E-04
GO:0002706	regulation of lymphocyte mediated immunity	BP	8.88E-04
GO:0042346	positive regulation of NF-kappaB import into nucleus	BP	9.00E-04
GO:0048731	system development	BP	9.36E-04
GO:0032501	multicellular organismal process	BP	9.83E-04
GO:0002250	adaptive immune response	BP	9.96E-04
GO:0006955	immune response	BP	1.00E-03
GO:0038083	peptidyl-tyrosine autophosphorylation	BP	1.00E-03
GO:0002521	leukocyte differentiation	BP	1.06E-03
GO:0045619	regulation of lymphocyte differentiation	BP	1.06E-03
GO:0022409	positive regulation of cell-cell adhesion	BP	1.10E-03
GO:0002694	regulation of leukocyte activation	BP	1.12E-03
GO:0030217	T cell differentiation	BP	1.29E-03
GO:0031624	ubiquitin conjugating enzyme binding	MF	1.30E-03
GO:0001205	transcriptional activator activity, RNA polymerase II distal enhancer sequence-specific binding	MF	1.31E-03
GO:0031235	intrinsic component of the cytoplasmic side of the plasma membrane	CC	1.32E-03
GO:0032609	interferon-gamma production	BP	1.35E-03
GO:0016404	15-hydroxyprostaglandin dehydrogenase (NAD+) activity	MF	1.42E-03
GO:1900750	oligopeptide binding	MF	1.43E-03
GO:0045580	regulation of T cell differentiation	BP	1.45E-03
GO:0032310	prostaglandin secretion	BP	1.46E-03
GO:0005509	calcium ion binding	MF	1.46E-03
GO:0046633	alpha-beta T cell proliferation	BP	1.48E-03
GO:0002709	regulation of T cell mediated immunity	BP	1.53E-03
GO:0048584	positive regulation of response to stimulus	BP	1.55E-03
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	BP	1.61E-03

GO ID	GO Function	Ontology	P. Value
GO:0032649	regulation of interferon-gamma production	BP	1.63E-03
GO:0050870	positive regulation of T cell activation	BP	1.64E-03
GO:0009966	regulation of signal transduction	BP	1.64E-03
GO:0009898	cytoplasmic side of plasma membrane	CC	1.65E-03
GO:0002429	immune response-activating cell surface receptor signaling pathway	BP	1.74E-03
GO:0098602	single organism cell adhesion	BP	1.76E-03
GO:0002819	regulation of adaptive immune response	BP	1.79E-03
GO:2000377	regulation of reactive oxygen species metabolic process	BP	1.80E-03
GO:0032436	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	BP	1.88E-03
GO:0045785	positive regulation of cell adhesion	BP	1.90E-03
GO:1903036	positive regulation of response to wounding	BP	1.93E-03
GO:2000330	positive regulation of T-helper 17 cell lineage commitment	BP	1.97E-03
GO:0044390	ubiquitin-like protein conjugating enzyme binding	MF	2.08E-03
GO:0030155	regulation of cell adhesion	BP	2.11E-03
GO:0007275	multicellular organism development	BP	2.16E-03
GO:0002133	polycystin complex	CC	2.19E-03
GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	BP	2.27E-03
GO:0002376	immune system process	BP	2.33E-03
GO:0002443	leukocyte mediated immunity	BP	2.33E-03
GO:0034112	positive regulation of homotypic cell-cell adhesion	BP	2.36E-03
GO:0042165	neurotransmitter binding	MF	2.37E-03
GO:0004715	non-membrane spanning protein tyrosine kinase activity	MF	2.37E-03
GO:1903039	positive regulation of leukocyte cell-cell adhesion	BP	2.42E-03
GO:0042130	negative regulation of T cell proliferation	BP	2.44E-03
GO:0048518	positive regulation of biological process	BP	2.52E-03
GO:0043368	positive T cell selection	BP	2.58E-03
GO:0044448	cell cortex part	CC	2.70E-03
GO:0072311	glomerular epithelial cell differentiation	BP	2.80E-03
GO:2000328	regulation of T-helper 17 cell lineage commitment	BP	2.93E-03
GO:0003158	endothelium development	BP	2.93E-03
GO:0051140	regulation of NK T cell proliferation	BP	3.02E-03
GO:0051142	positive regulation of NK T cell proliferation	BP	3.02E-03
GO:0002834	regulation of response to tumor cell	BP	3.03E-03
GO:0002837	regulation of immune response to tumor cell	BP	3.03E-03
GO:0042127	regulation of cell proliferation	BP	3.05E-03
GO:0000145	exocyst	CC	3.08E-03
GO:0050851	antigen receptor-mediated signaling pathway	BP	3.14E-03
GO:0038127	ERBB signaling pathway	BP	3.21E-03
GO:0051251	positive regulation of lymphocyte activation	BP	3.23E-03
GO:0007399	nervous system development	BP	3.23E-03
GO:0007167	enzyme linked receptor protein signaling pathway	BP	3.38E-03
GO:0048534	hematopoietic or lymphoid organ development	BP	3.52E-03
GO:0043584	nose development	BP	3.62E-03
GO:0072010	glomerular epithelium development	BP	3.70E-03
GO:0006952	defense response	BP	3.72E-03
GO:0002696	positive regulation of leukocyte activation	BP	3.76E-03
GO:0090241	negative regulation of histone H4 acetylation	BP	3.86E-03
GO:0010647	positive regulation of cell communication	BP	3.87E-03
GO:0008283	cell proliferation	BP	3.90E-03
GO:0007166	cell surface receptor signaling pathway	BP	4.05E-03
GO:0002711	positive regulation of T cell mediated immunity	BP	4.10E-03
GO:0007165	signal transduction	BP	4.13E-03
GO:0031226	intrinsic component of plasma membrane	CC	4.36E-03
GO:2000321	positive regulation of T-helper 17 cell differentiation	BP	4.39E-03
GO:0046631	alpha-beta T cell activation	BP	4.44E-03
GO:0050868	negative regulation of T cell activation	BP	4.47E-03
GO:0015629	actin cytoskeleton	CC	4.53E-03
GO:0035502	metanephric part of ureteric bud development	BP	4.60E-03
GO:0009893	positive regulation of metabolic process	BP	4.64E-03
GO:2000344	positive regulation of acrosome reaction	BP	4.65E-03
GO:1901800	positive regulation of proteasomal protein catabolic process	BP	4.69E-03
GO:0046136	positive regulation of vitamin metabolic process	BP	4.69E-03
GO:0060557	positive regulation of vitamin D biosynthetic process	BP	4.69E-03
GO:0060559	positive regulation of calcidiol 1-monooxygenase activity	BP	4.69E-03
GO:0070664	negative regulation of leukocyte proliferation	BP	4.99E-03
GO:0043547	positive regulation of GTPase activity	BP	5.22E-03
GO:0042345	regulation of NF-kappaB import into nucleus	BP	5.23E-03
GO:0042348	NF-kappaB import into nucleus	BP	5.23E-03
GO:0050867	positive regulation of cell activation	BP	5.29E-03
GO:0043085	positive regulation of catalytic activity	BP	5.36E-03
GO:0035556	intracellular signal transduction	BP	5.38E-03
GO:0006468	protein phosphorylation	BP	5.40E-03
GO:0009967	positive regulation of signal transduction	BP	5.44E-03
GO:0002682	regulation of immune system process	BP	5.56E-03
GO:0048844	artery morphogenesis	BP	5.62E-03
GO:0042102	positive regulation of T cell proliferation	BP	5.72E-03
GO:0071495	cellular response to endogenous stimulus	BP	5.83E-03
GO:0072074	kidney mesenchyme development	BP	6.19E-03
GO:1903038	negative regulation of leukocyte cell-cell adhesion	BP	6.44E-03
GO:0035601	protein deacylation	BP	6.44E-03
GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	BP	6.61E-03
GO:0006464	cellular protein modification process	BP	6.64E-03

GO ID	GO Function	Ontology	P. Value
GO:0036211	protein modification process	BP	6.64E-03
GO:0022412	cellular process involved in reproduction in multicellular organism	BP	6.69E-03
GO:0001866	NK T cell proliferation	BP	6.70E-03
GO:0031622	positive regulation of fever generation	BP	6.78E-03
GO:0010949	negative regulation of intestinal phytosterol absorption	BP	6.82E-03
GO:0045796	negative regulation of intestinal cholesterol absorption	BP	6.82E-03
GO:0060752	intestinal phytosterol absorption	BP	6.82E-03
GO:1904730	negative regulation of intestinal lipid absorption	BP	6.82E-03
GO:0050853	B cell receptor signaling pathway	BP	7.04E-03
GO:0046632	alpha-beta T cell differentiation	BP	7.07E-03
GO:0032306	regulation of prostaglandin secretion	BP	7.15E-03
GO:0032308	positive regulation of prostaglandin secretion	BP	7.15E-03
GO:0023056	positive regulation of signaling	BP	7.16E-03
GO:0019953	sexual reproduction	BP	7.22E-03
GO:0071310	cellular response to organic substance	BP	7.34E-03
GO:2000319	regulation of T-helper 17 cell differentiation	BP	7.39E-03
GO:0031933	telomeric heterochromatin	CC	7.50E-03
GO:0031234	extrinsic component of cytoplasmic side of plasma membrane	CC	7.51E-03
GO:0046637	regulation of alpha-beta T cell differentiation	BP	7.58E-03
GO:0043382	positive regulation of memory T cell differentiation	BP	7.71E-03
GO:0042474	middle ear morphogenesis	BP	7.81E-03
GO:0098732	macromolecule deacylation	BP	7.83E-03
GO:0010893	positive regulation of steroid biosynthetic process	BP	7.88E-03
GO:0007173	epidermal growth factor receptor signaling pathway	BP	7.98E-03
GO:0072021	ascending thin limb development	BP	8.18E-03
GO:0072218	metanephric ascending thin limb development	BP	8.18E-03
GO:0043370	regulation of CD4-positive, alpha-beta T cell differentiation	BP	8.32E-03
GO:0045732	positive regulation of protein catabolic process	BP	8.32E-03
GO:0042993	positive regulation of transcription factor import into nucleus	BP	8.52E-03
GO:0033077	T cell differentiation in thymus	BP	8.61E-03
GO:0071594	thymocyte aggregation	BP	8.61E-03
GO:0002520	immune system development	BP	8.64E-03
GO:0071944	cell periphery	CC	8.70E-03
GO:0048522	positive regulation of cellular process	BP	8.80E-03
GO:1902932	positive regulation of alcohol biosynthetic process	BP	8.83E-03
GO:0043065	positive regulation of apoptotic process	BP	8.92E-03
GO:1902107	positive regulation of leukocyte differentiation	BP	8.94E-03
GO:0010033	response to organic substance	BP	9.12E-03
GO:0002418	immune response to tumor cell	BP	9.43E-03
GO:0034045	pre-autophagosomal structure membrane	CC	9.48E-03
GO:0015671	oxygen transport	BP	9.63E-03
GO:0030010	establishment of cell polarity	BP	9.67E-03
GO:0032872	regulation of stress-activated MAPK cascade	BP	9.75E-03
GO:0007264	small GTPase mediated signal transduction	BP	9.81E-03
GO:0003779	actin binding	MF	9.93E-03

Appendix 6: Gene Ontology (GO) analysis for top DMRs ranked by P-value.

GO ID	GO Function	Ontology	P. Value	FDR
GO:0002705	positive regulation of leukocyte mediated immunity	BP	3.31E-10	6.84E-06
GO:0002925	positive regulation of humoral immune response mediated by circulating immunoglobulin	BP	3.38E-09	3.50E-05
GO:0002699	positive regulation of immune effector process	BP	5.99E-09	4.13E-05
GO:0002703	regulation of leukocyte mediated immunity	BP	8.49E-09	4.39E-05
GO:0002824	positive regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	BP	2.50E-08	7.36E-05
GO:0002708	positive regulation of lymphocyte mediated immunity	BP	2.82E-08	7.36E-05
GO:0050830	defense response to Gram-positive bacterium	BP	3.12E-08	7.36E-05
GO:0002821	positive regulation of adaptive immune response	BP	3.33E-08	7.36E-05
GO:0002923	regulation of humoral immune response mediated by circulating immunoglobulin	BP	3.46E-08	7.36E-05
GO:0002922	positive regulation of humoral immune response	BP	3.68E-08	7.36E-05
GO:0002697	regulation of immune effector process	BP	3.91E-08	7.36E-05
GO:0044130	negative regulation of growth of symbiont in host	BP	7.69E-08	1.24E-04
GO:0044146	negative regulation of growth of symbiont involved in interaction with host	BP	7.82E-08	1.24E-04
GO:0050715	positive regulation of cytokine secretion	BP	8.46E-08	1.24E-04
GO:0044126	regulation of growth of symbiont in host	BP	9.46E-08	1.24E-04
GO:0044144	modulation of growth of symbiont involved in interaction with host	BP	9.59E-08	1.24E-04
GO:0044117	growth of symbiont in host	BP	1.39E-07	1.53E-04
GO:0044110	growth involved in symbiotic interaction	BP	1.40E-07	1.53E-04
GO:0044116	growth of symbiont involved in interaction with host	BP	1.40E-07	1.53E-04
GO:0002861	regulation of inflammatory response to antigenic stimulus	BP	1.52E-07	1.57E-04
GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	BP	1.94E-07	1.83E-04
GO:0002443	leukocyte mediated immunity	BP	1.95E-07	1.83E-04
GO:0002714	positive regulation of B cell mediated immunity	BP	2.30E-07	1.98E-04
GO:0002891	positive regulation of immunoglobulin mediated immune response	BP	2.30E-07	1.98E-04
GO:0002706	regulation of lymphocyte mediated immunity	BP	2.48E-07	2.05E-04
GO:0002250	adaptive immune response	BP	3.08E-07	2.45E-04
GO:0002819	regulation of adaptive immune response	BP	3.22E-07	2.47E-04
GO:0002682	regulation of immune system process	BP	4.92E-07	3.63E-04
GO:0042742	defense response to bacterium	BP	5.38E-07	3.84E-04
GO:0050707	regulation of cytokine secretion	BP	5.77E-07	3.98E-04
GO:0002920	regulation of humoral immune response	BP	6.62E-07	4.42E-04
GO:0002455	humoral immune response mediated by circulating immunoglobulin	BP	7.87E-07	5.09E-04
GO:0050778	positive regulation of immune response	BP	8.41E-07	5.27E-04
GO:0002712	regulation of B cell mediated immunity	BP	9.84E-07	5.68E-04
GO:0002889	regulation of immunoglobulin mediated immune response	BP	9.84E-07	5.68E-04
GO:0050663	cytokine secretion	BP	9.88E-07	5.68E-04
GO:0002252	immune effector process	BP	1.42E-06	7.93E-04
GO:0001819	positive regulation of cytokine production	BP	1.51E-06	8.20E-04
GO:0002876	positive regulation of chronic inflammatory response to antigenic stimulus	BP	1.59E-06	8.43E-04
GO:0002874	regulation of chronic inflammatory response to antigenic stimulus	BP	2.13E-06	1.10E-03
GO:0098542	defense response to other organism	BP	2.32E-06	1.17E-03
GO:0002449	lymphocyte mediated immunity	BP	2.97E-06	1.47E-03
GO:0002437	inflammatory response to antigenic stimulus	BP	3.08E-06	1.48E-03
GO:0002678	positive regulation of chronic inflammatory response	BP	3.20E-06	1.50E-03
GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	BP	3.28E-06	1.51E-03
GO:0002439	chronic inflammatory response to antigenic stimulus	BP	5.69E-06	2.52E-03
GO:0002367	cytokine production involved in immune response	BP	5.72E-06	2.52E-03
GO:0050714	positive regulation of protein secretion	BP	6.25E-06	2.69E-03
GO:0002684	positive regulation of immune system process	BP	7.24E-06	3.06E-03
GO:0032649	regulation of interferon-gamma production	BP	9.66E-06	4.00E-03
GO:0001817	regulation of cytokine production	BP	1.12E-05	4.46E-03
GO:0002676	regulation of chronic inflammatory response	BP	1.12E-05	4.46E-03
GO:0050729	positive regulation of inflammatory response	BP	1.15E-05	4.49E-03
GO:0050776	regulation of immune response	BP	1.20E-05	4.59E-03
GO:0016064	immunoglobulin mediated immune response	BP	1.22E-05	4.59E-03
GO:0019724	B cell mediated immunity	BP	1.34E-05	4.91E-03
GO:0002700	regulation of production of molecular mediator of immune response	BP	1.35E-05	4.91E-03
GO:0032609	interferon-gamma production	BP	1.56E-05	5.55E-03
GO:0002830	positive regulation of type 2 immune response	BP	1.80E-05	6.24E-03
GO:0050764	regulation of phagocytosis	BP	1.83E-05	6.24E-03
GO:0001816	cytokine production	BP	1.84E-05	6.24E-03
GO:0002863	positive regulation of inflammatory response to antigenic stimulus	BP	2.01E-05	6.70E-03
GO:0002376	immune system process	BP	2.50E-05	8.20E-03
GO:0002374	cytokine secretion involved in immune response	BP	2.63E-05	8.51E-03
GO:0006959	humoral immune response	BP	4.03E-05	1.28E-02
GO:0002544	chronic inflammatory response	BP	4.12E-05	1.29E-02
GO:0002440	production of molecular mediator of immune response	BP	4.63E-05	1.43E-02
GO:0031347	regulation of defense response	BP	4.82E-05	1.47E-02
GO:1903036	positive regulation of response to wounding	BP	5.40E-05	1.62E-02
GO:1903532	positive regulation of secretion by cell	BP	5.53E-05	1.64E-02
GO:0002828	regulation of type 2 immune response	BP	6.06E-05	1.77E-02
GO:0032733	positive regulation of interleukin-10 production	BP	6.54E-05	1.88E-02
GO:0043207	response to external biotic stimulus	BP	6.88E-05	1.92E-02
GO:0051707	response to other organism	BP	6.88E-05	1.92E-02
GO:0051046	regulation of secretion	BP	7.22E-05	1.99E-02
GO:0051047	positive regulation of secretion	BP	8.30E-05	2.23E-02
GO:0051092	positive regulation of NF-kappaB transcription factor activity	BP	8.33E-05	2.23E-02
GO:0009607	response to biotic stimulus	BP	8.42E-05	2.23E-02
GO:0043901	negative regulation of multi-organism process	BP	8.77E-05	2.30E-02
GO:0009617	response to bacterium	BP	1.12E-04	2.90E-02
GO:0042092	type 2 immune response	BP	1.23E-04	3.14E-02
GO:0045834	positive regulation of lipid metabolic process	BP	1.33E-04	3.36E-02
GO:0006955	immune response	BP	1.37E-04	3.41E-02
GO:1901224	positive regulation of NIK/NF-kappaB signaling	BP	1.46E-04	3.60E-02
GO:0050708	regulation of protein secretion	BP	1.48E-04	3.60E-02
GO:0051222	positive regulation of protein transport	BP	1.52E-04	3.65E-02
GO:0005164	tumor necrosis factor receptor binding	MF	1.55E-04	3.69E-02
GO:0070374	positive regulation of ERK1 and ERK2 cascade	BP	1.64E-04	3.86E-02
GO:0043900	regulation of multi-organism process	BP	1.74E-04	4.04E-02
GO:0032653	regulation of interleukin-10 production	BP	1.96E-04	4.50E-02
GO:0006954	inflammatory response	BP	2.06E-04	4.69E-02
GO:0006952	defense response	BP	2.23E-04	5.01E-02
GO:0032613	interleukin-10 production	BP	2.35E-04	5.23E-02
GO:0002709	regulation of T cell mediated immunity	BP	2.56E-04	5.64E-02
GO:1904951	positive regulation of establishment of protein localization	BP	2.68E-04	5.83E-02
GO:1901222	regulation of NIK/NF-kappaB signaling	BP	2.85E-04	6.06E-02

GO ID	GO Function	Ontology	P. Value	FDR
GO:0051050	positive regulation of transport	BP	2.86E-04	6.06E-02
GO:0009306	protein secretion	BP	2.87E-04	6.06E-02
GO:0032757	positive regulation of interleukin-8 production	BP	3.16E-04	6.60E-02
GO:0002702	positive regulation of production of molecular mediator of immune response	BP	3.52E-04	7.28E-02
GO:0050727	regulation of inflammatory response	BP	3.61E-04	7.39E-02
GO:0051240	positive regulation of multicellular organismal process	BP	3.68E-04	7.47E-02
GO:0043903	regulation of symbiosis, encompassing mutualism through parasitism	BP	3.76E-04	7.56E-02
GO:0032813	tumor necrosis factor receptor superfamily binding	MF	3.87E-04	7.70E-02
GO:0002718	regulation of cytokine production involved in immune response	BP	4.12E-04	8.12E-02
GO:0043552	positive regulation of phosphatidylinositol 3-kinase activity	BP	4.50E-04	8.76E-02
GO:0032602	chemokine production	BP	4.53E-04	8.76E-02
GO:0030100	regulation of endocytosis	BP	4.59E-04	8.77E-02
GO:1901700	response to oxygen-containing compound	BP	4.62E-04	8.77E-02
GO:0032755	positive regulation of interleukin-6 production	BP	4.74E-04	8.92E-02
GO:0032103	positive regulation of response to external stimulus	BP	4.79E-04	8.92E-02
GO:0071310	cellular response to organic substance	BP	4.85E-04	8.96E-02
GO:0032729	positive regulation of interferon-gamma production	BP	4.92E-04	9.00E-02
GO:0070372	regulation of ERK1 and ERK2 cascade	BP	5.04E-04	9.15E-02
GO:0009611	response to wounding	BP	5.35E-04	9.63E-02
GO:0090218	positive regulation of lipid kinase activity	BP	5.44E-04	9.71E-02
GO:0006909	phagocytosis	BP	5.57E-04	9.84E-02
GO:0070371	ERK1 and ERK2 cascade	BP	5.70E-04	9.97E-02
GO:0051353	positive regulation of oxidoreductase activity	BP	5.78E-04	9.97E-02
GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity	BP	5.78E-04	9.97E-02
GO:0032677	regulation of interleukin-8 production	BP	6.07E-04	1.04E-01
GO:0046903	secretion	BP	6.15E-04	1.04E-01
GO:0022409	positive regulation of cell-cell adhesion	BP	6.30E-04	1.06E-01
GO:0050766	positive regulation of phagocytosis	BP	6.43E-04	1.07E-01
GO:0032637	interleukin-8 production	BP	6.95E-04	1.15E-01
GO:1903727	positive regulation of phospholipid metabolic process	BP	7.02E-04	1.15E-01
GO:0040007	growth	BP	7.63E-04	1.24E-01
GO:0002696	positive regulation of leukocyte activation	BP	7.64E-04	1.24E-01
GO:1901701	cellular response to oxygen-containing compound	BP	7.85E-04	1.25E-01
GO:1903530	regulation of secretion by cell	BP	7.87E-04	1.25E-01
GO:0050867	positive regulation of cell activation	BP	8.31E-04	1.31E-01
GO:0045321	leukocyte activation	BP	8.42E-04	1.32E-01
GO:0043551	regulation of phosphatidylinositol 3-kinase activity	BP	8.73E-04	1.35E-01
GO:0045926	negative regulation of growth	BP	8.82E-04	1.35E-01
GO:0002377	immunoglobulin production	BP	8.87E-04	1.35E-01
GO:0019901	protein kinase binding	MF	8.92E-04	1.35E-01
GO:0002456	T cell mediated immunity	BP	9.21E-04	1.35E-01
GO:0048583	regulation of response to stimulus	BP	9.73E-04	1.35E-01
GO:0002805	regulation of antimicrobial peptide biosynthetic process	BP	9.85E-04	1.35E-01
GO:0002807	positive regulation of antimicrobial peptide biosynthetic process	BP	9.85E-04	1.35E-01
GO:0002808	regulation of antibacterial peptide biosynthetic process	BP	9.85E-04	1.35E-01
GO:0002815	biosynthetic process of antibacterial peptides active against Gram-positive bacteria	BP	9.85E-04	1.35E-01
GO:0002816	regulation of biosynthetic process of antibacterial peptides active against Gram-positive bacteria	BP	9.85E-04	1.35E-01
GO:0006963	positive regulation of antibacterial peptide biosynthetic process	BP	9.85E-04	1.35E-01
GO:0006965	positive regulation of biosynthetic process of antibacterial peptides active against Gram-positive bacteria	BP	9.85E-04	1.35E-01
GO:0032498	detection of muramyl dipeptide	BP	9.85E-04	1.35E-01
GO:2000361	regulation of prostaglandin-E synthase activity	BP	9.85E-04	1.35E-01
GO:2000363	positive regulation of prostaglandin-E synthase activity	BP	9.85E-04	1.35E-01
GO:0032500	muramyl dipeptide binding	MF	9.85E-04	1.35E-01
GO:0002777	antimicrobial peptide biosynthetic process	BP	9.91E-04	1.35E-01
GO:0002780	antibacterial peptide biosynthetic process	BP	9.91E-04	1.35E-01
GO:0032496	response to lipopolysaccharide	BP	9.95E-04	1.35E-01
GO:0032499	detection of peptidoglycan	BP	1.02E-03	1.38E-01
GO:0032675	regulation of interleukin-6 production	BP	1.06E-03	1.42E-01
GO:0002237	response to molecule of bacterial origin	BP	1.07E-03	1.43E-01
GO:0032701	negative regulation of interleukin-18 production	BP	1.08E-03	1.43E-01
GO:0032635	interleukin-6 production	BP	1.14E-03	1.51E-01
GO:0071608	macrophage inflammatory protein-1 alpha production	BP	1.17E-03	1.53E-01
GO:0046324	regulation of glucose import	BP	1.19E-03	1.54E-01
GO:0070887	cellular response to chemical stimulus	BP	1.21E-03	1.57E-01
GO:0098589	membrane region	CC	1.25E-03	1.60E-01
GO:0033674	positive regulation of kinase activity	BP	1.25E-03	1.60E-01
GO:0043550	regulation of lipid kinase activity	BP	1.27E-03	1.60E-01
GO:0060584	regulation of prostaglandin-endoperoxide synthase activity	BP	1.28E-03	1.60E-01
GO:0060585	positive regulation of prostaglandin-endoperoxide synthase activity	BP	1.28E-03	1.60E-01
GO:1903706	regulation of hemopoiesis	BP	1.32E-03	1.64E-01
GO:0034988	Fc-gamma receptor I complex binding	MF	1.32E-03	1.64E-01
GO:0051239	regulation of multicellular organismal process	BP	1.34E-03	1.64E-01
GO:0019900	kinase binding	MF	1.34E-03	1.64E-01
GO:0071224	cellular response to peptidoglycan	BP	1.37E-03	1.64E-01
GO:0002225	positive regulation of antimicrobial peptide production	BP	1.40E-03	1.64E-01
GO:0002760	positive regulation of antimicrobial humoral response	BP	1.40E-03	1.64E-01
GO:0002803	positive regulation of antibacterial peptide production	BP	1.40E-03	1.64E-01
GO:0002784	regulation of antimicrobial peptide production	BP	1.40E-03	1.64E-01
GO:0002786	regulation of antibacterial peptide production	BP	1.40E-03	1.64E-01
GO:0002775	antimicrobial peptide production	BP	1.40E-03	1.64E-01
GO:0002778	antibacterial peptide production	BP	1.40E-03	1.64E-01
GO:0048584	positive regulation of response to stimulus	BP	1.41E-03	1.64E-01
GO:0009615	response to virus	BP	1.42E-03	1.64E-01
GO:0071417	cellular response to organonitrogen compound	BP	1.44E-03	1.65E-01
GO:0019216	regulation of lipid metabolic process	BP	1.44E-03	1.65E-01
GO:0051223	regulation of protein transport	BP	1.48E-03	1.68E-01
GO:0007166	cell surface receptor signaling pathway	BP	1.52E-03	1.71E-01
GO:1903034	regulation of response to wounding	BP	1.52E-03	1.71E-01
GO:0061048	negative regulation of branching involved in lung morphogenesis	BP	1.55E-03	1.73E-01
GO:0046323	glucose import	BP	1.55E-03	1.73E-01
GO:0080134	regulation of response to stress	BP	1.58E-03	1.74E-01
GO:0002759	regulation of antimicrobial humoral response	BP	1.59E-03	1.74E-01
GO:0002698	negative regulation of immune effector process	BP	1.61E-03	1.74E-01
GO:0046136	positive regulation of vitamin metabolic process	BP	1.62E-03	1.74E-01
GO:0060557	positive regulation of vitamin D biosynthetic process	BP	1.62E-03	1.74E-01
GO:0060559	positive regulation of calcidiol 1-monooxygenase activity	BP	1.62E-03	1.74E-01
GO:0007267	cell-cell signaling	BP	1.68E-03	1.80E-01
GO:0031349	positive regulation of defense response	BP	1.70E-03	1.80E-01
GO:0032661	regulation of interleukin-18 production	BP	1.71E-03	1.80E-01
GO:0038061	NIK/NF-kappaB signaling	BP	1.72E-03	1.80E-01
GO:0018193	peptidyl-amino acid modification	BP	1.72E-03	1.80E-01
GO:0045121	membrane raft	CC	1.73E-03	1.80E-01
GO:0098857	membrane microdomain	CC	1.73E-03	1.80E-01

GO ID	GO Function	Ontology	P. Value	FDR
GO:0010033	response to organic substance	BP	1.77E-03	1.83E-01
GO:0005125	cytokine activity	MF	1.78E-03	1.83E-01
GO:1903725	regulation of phospholipid metabolic process	BP	1.80E-03	1.83E-01
GO:0071677	positive regulation of mononuclear cell migration	BP	1.80E-03	1.83E-01
GO:0022600	digestive system process	BP	1.80E-03	1.83E-01
GO:0051347	positive regulation of transferase activity	BP	1.82E-03	1.84E-01
GO:0002685	regulation of leukocyte migration	BP	1.90E-03	1.91E-01
GO:1901699	cellular response to nitrogen compound	BP	1.91E-03	1.91E-01
GO:0048534	hematopoietic or lymphoid organ development	BP	1.94E-03	1.93E-01
GO:0050436	microfibril binding	MF	1.96E-03	1.94E-01
GO:0032101	regulation of response to external stimulus	BP	2.00E-03	1.97E-01
GO:0050777	negative regulation of immune response	BP	2.01E-03	1.97E-01
GO:0051341	regulation of oxidoreductase activity	BP	2.04E-03	1.99E-01
GO:0032621	interleukin-18 production	BP	2.13E-03	2.06E-01
GO:0070201	regulation of establishment of protein localization	BP	2.15E-03	2.08E-01
GO:0022407	regulation of cell-cell adhesion	BP	2.19E-03	2.10E-01
GO:0034987	immunoglobulin receptor binding	MF	2.19E-03	2.10E-01
GO:0002520	immune system development	BP	2.25E-03	2.15E-01
GO:0002694	regulation of leukocyte activation	BP	2.31E-03	2.20E-01
GO:0045416	positive regulation of interleukin-8 biosynthetic process	BP	2.33E-03	2.20E-01
GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	BP	2.34E-03	2.20E-01
GO:0018108	peptidyl-tyrosine phosphorylation	BP	2.36E-03	2.21E-01
GO:1901671	positive regulation of superoxide dismutase activity	BP	2.40E-03	2.23E-01
GO:1904833	positive regulation of removal of superoxide radicals	BP	2.40E-03	2.23E-01
GO:0086089	voltage-gated potassium channel activity involved in atrial cardiac muscle cell action potential repolarization	MF	2.42E-03	2.23E-01
GO:0018212	peptidyl-tyrosine modification	BP	2.43E-03	2.24E-01
GO:0010827	regulation of glucose transport	BP	2.45E-03	2.24E-01
GO:0008284	positive regulation of cell proliferation	BP	2.46E-03	2.24E-01
GO:0002740	negative regulation of cytokine secretion involved in immune response	BP	2.61E-03	2.37E-01
GO:0035419	activation of MAPK activity involved in innate immune response	BP	2.63E-03	2.38E-01
GO:1903347	negative regulation of bicellular tight junction assembly	BP	2.66E-03	2.39E-01
GO:0035509	negative regulation of myosin-light-chain-phosphatase activity	BP	2.67E-03	2.39E-01
GO:2000551	regulation of T-helper 2 cell cytokine production	BP	2.69E-03	2.39E-01
GO:2000553	positive regulation of T-helper 2 cell cytokine production	BP	2.69E-03	2.39E-01
GO:0001775	cell activation	BP	2.71E-03	2.40E-01
GO:0045785	positive regulation of cell adhesion	BP	2.73E-03	2.41E-01
GO:2000343	positive regulation of chemokine (C-X-C motif) ligand 2 production	BP	2.75E-03	2.41E-01
GO:0034250	positive regulation of cellular amide metabolic process	BP	2.78E-03	2.43E-01
GO:0060627	regulation of vesicle-mediated transport	BP	2.88E-03	2.50E-01
GO:0007586	digestion	BP	2.91E-03	2.52E-01
GO:0071495	cellular response to endogenous stimulus	BP	2.98E-03	2.57E-01
GO:0045347	negative regulation of MHC class II biosynthetic process	BP	2.99E-03	2.57E-01
GO:0032494	response to peptidoglycan	BP	3.03E-03	2.59E-01
GO:0050865	regulation of cell activation	BP	3.20E-03	2.73E-01
GO:0007159	leukocyte cell-cell adhesion	BP	3.22E-03	2.73E-01
GO:0002757	immune response-activating signal transduction	BP	3.26E-03	2.75E-01
GO:1900426	positive regulation of defense response to bacterium	BP	3.28E-03	2.75E-01
GO:0008104	protein localization	BP	3.33E-03	2.79E-01
GO:1900017	positive regulation of cytokine production involved in inflammatory response	BP	3.42E-03	2.83E-01
GO:0005886	plasma membrane	CC	3.43E-03	2.83E-01
GO:0002521	leukocyte differentiation	BP	3.43E-03	2.83E-01
GO:1901033	positive regulation of response to reactive oxygen species	BP	3.45E-03	2.83E-01
GO:0042127	regulation of cell proliferation	BP	3.46E-03	2.83E-01
GO:0045994	positive regulation of translational initiation by iron	BP	3.46E-03	2.83E-01
GO:0071225	cellular response to muramyl dipeptide	BP	3.54E-03	2.88E-01
GO:0097527	necroptotic signaling pathway	BP	3.63E-03	2.94E-01
GO:0035745	T-helper 2 cell cytokine production	BP	3.66E-03	2.94E-01
GO:0046330	positive regulation of JNK cascade	BP	3.66E-03	2.94E-01
GO:0010693	negative regulation of alkaline phosphatase activity	BP	3.67E-03	2.94E-01
GO:0045859	regulation of protein kinase activity	BP	3.69E-03	2.95E-01
GO:0002606	positive regulation of dendritic cell antigen processing and presentation	BP	3.72E-03	2.95E-01
GO:0002253	activation of immune response	BP	3.73E-03	2.95E-01
GO:0045179	apical cortex	CC	3.78E-03	2.98E-01
GO:0045807	positive regulation of endocytosis	BP	3.82E-03	3.00E-01
GO:0032940	secretion by cell	BP	3.82E-03	3.00E-01
GO:0034136	negative regulation of toll-like receptor 2 signaling pathway	BP	3.85E-03	3.01E-01
GO:2000110	negative regulation of macrophage apoptotic process	BP	3.87E-03	3.01E-01
GO:0044027	hypermethylation of CpG island	BP	3.88E-03	3.01E-01
GO:1902107	positive regulation of leukocyte differentiation	BP	3.92E-03	3.03E-01
GO:0035743	CD4-positive, alpha-beta T cell cytokine production	BP	3.98E-03	3.06E-01
GO:0044700	single organism signaling	BP	4.05E-03	3.06E-01
GO:1901668	regulation of superoxide dismutase activity	BP	4.05E-03	3.06E-01
GO:0002274	myeloid leukocyte activation	BP	4.06E-03	3.06E-01
GO:0045080	positive regulation of chemokine biosynthetic process	BP	4.08E-03	3.06E-01
GO:0032880	regulation of protein localization	BP	4.09E-03	3.06E-01
GO:0042834	peptidoglycan binding	MF	4.09E-03	3.06E-01
GO:0071944	cell periphery	CC	4.09E-03	3.06E-01
GO:0023052	signaling	BP	4.10E-03	3.06E-01
GO:0051525	NFAT protein binding	MF	4.14E-03	3.08E-01
GO:0046645	positive regulation of gamma-delta T cell activation	BP	4.19E-03	3.10E-01
GO:2001181	positive regulation of interleukin-10 secretion	BP	4.25E-03	3.13E-01
GO:0002862	negative regulation of inflammatory response to antigenic stimulus	BP	4.26E-03	3.13E-01
GO:0015758	glucose transport	BP	4.27E-03	3.13E-01
GO:0006447	regulation of translational initiation by iron	BP	4.34E-03	3.15E-01
GO:1902533	positive regulation of intracellular signal transduction	BP	4.34E-03	3.15E-01
GO:2000341	regulation of chemokine (C-X-C motif) ligand 2 production	BP	4.34E-03	3.15E-01
GO:0097623	potassium ion export across plasma membrane	BP	4.36E-03	3.15E-01
GO:0008645	hexose transport	BP	4.37E-03	3.15E-01
GO:0015749	monosaccharide transport	BP	4.44E-03	3.19E-01
GO:0007154	cell communication	BP	4.50E-03	3.22E-01
GO:0010243	response to organonitrogen compound	BP	4.52E-03	3.22E-01
GO:0043549	regulation of kinase activity	BP	4.57E-03	3.25E-01
GO:0033209	tumor necrosis factor-mediated signaling pathway	BP	4.61E-03	3.27E-01
GO:0031584	activation of phospholipase D activity	BP	4.64E-03	3.27E-01
GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB signaling	BP	4.70E-03	3.30E-01
GO:0015031	protein transport	BP	4.73E-03	3.30E-01
GO:0071219	cellular response to molecule of bacterial origin	BP	4.73E-03	3.30E-01
GO:0044026	DNA hypermethylation	BP	4.75E-03	3.30E-01
GO:0045346	regulation of MHC class II biosynthetic process	BP	4.76E-03	3.30E-01
GO:0035507	regulation of myosin-light-chain-phosphatase activity	BP	4.78E-03	3.30E-01
GO:0032490	detection of molecule of bacterial origin	BP	4.78E-03	3.30E-01
GO:0051770	positive regulation of nitric-oxide synthase biosynthetic process	BP	4.94E-03	3.40E-01

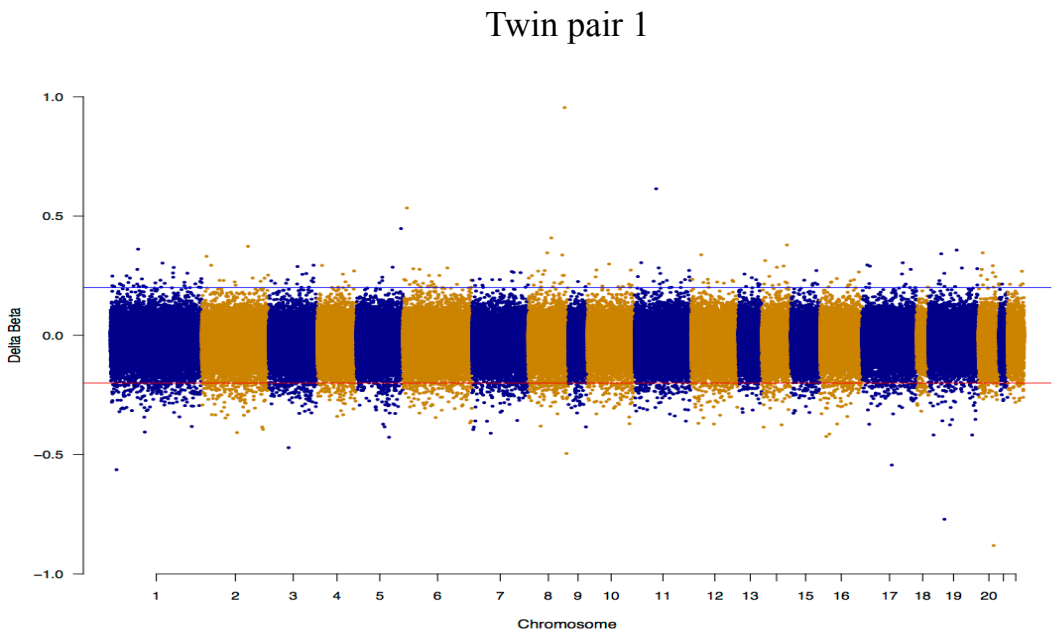
GO ID	GO Function	Ontology	P. Value	FDR
GO:0032874	positive regulation of stress-activated MAPK cascade	BP	4.98E-03	3.41E-01
GO:0002227	innate immune response in mucosa	BP	5.00E-03	3.41E-01
GO:0000187	activation of MAPK activity	BP	5.05E-03	3.43E-01
GO:0035583	sequestering of TGFbeta in extracellular matrix	BP	5.06E-03	3.43E-01
GO:0098590	plasma membrane region	CC	5.08E-03	3.43E-01
GO:0044092	negative regulation of molecular function	BP	5.10E-03	3.43E-01
GO:0070304	positive regulation of stress-activated protein kinase signaling cascade	BP	5.12E-03	3.44E-01
GO:0010911	regulation of isomerase activity	BP	5.16E-03	3.44E-01
GO:0010912	positive regulation of isomerase activity	BP	5.16E-03	3.44E-01
GO:0002710	negative regulation of T cell mediated immunity	BP	5.20E-03	3.46E-01
GO:0045630	positive regulation of T-helper 2 cell differentiation	BP	5.22E-03	3.46E-01
GO:2000664	positive regulation of interleukin-5 secretion	BP	5.32E-03	3.51E-01
GO:2000667	positive regulation of interleukin-13 secretion	BP	5.32E-03	3.51E-01
GO:0045414	regulation of interleukin-8 biosynthetic process	BP	5.37E-03	3.52E-01
GO:0032147	activation of protein kinase activity	BP	5.41E-03	3.53E-01
GO:0002468	dendritic cell antigen processing and presentation	BP	5.43E-03	3.53E-01
GO:0002604	regulation of dendritic cell antigen processing and presentation	BP	5.43E-03	3.53E-01
GO:0071803	positive regulation of podosome assembly	BP	5.49E-03	3.56E-01
GO:0060558	regulation of calcidiol 1-monooxygenase activity	BP	5.59E-03	3.58E-01
GO:0060372	regulation of atrial cardiac muscle cell membrane repolarization	BP	5.61E-03	3.58E-01
GO:0099624	atrial cardiac muscle cell membrane repolarization	BP	5.61E-03	3.58E-01
GO:0042327	positive regulation of phosphorylation	BP	5.62E-03	3.58E-01
GO:1902282	voltage-gated potassium channel activity involved in ventricular cardiac muscle cell action potential repolarization	MF	5.62E-03	3.58E-01
GO:0065008	regulation of biological quality	BP	5.62E-03	3.58E-01
GO:0006927	transformed cell apoptotic process	BP	5.70E-03	3.62E-01
GO:0008643	carbohydrate transport	BP	5.74E-03	3.62E-01
GO:1901698	response to nitrogen compound	BP	5.75E-03	3.62E-01
GO:0002366	leukocyte activation involved in immune response	BP	5.81E-03	3.64E-01
GO:0071692	protein localization to extracellular region	BP	5.82E-03	3.64E-01
GO:0071694	maintenance of protein location in extracellular region	BP	5.82E-03	3.64E-01
GO:0032740	positive regulation of interleukin-17 production	BP	5.86E-03	3.64E-01
GO:0045342	MHC class II biosynthetic process	BP	5.88E-03	3.64E-01
GO:1903708	positive regulation of hemopoiesis	BP	5.90E-03	3.64E-01
GO:2001179	regulation of interleukin-10 secretion	BP	5.90E-03	3.64E-01
GO:0002263	cell activation involved in immune response	BP	5.91E-03	3.64E-01
GO:0071216	cellular response to biotic stimulus	BP	5.97E-03	3.67E-01
GO:0060556	regulation of vitamin D biosynthetic process	BP	6.03E-03	3.68E-01
GO:0002726	positive regulation of T cell cytokine production	BP	6.06E-03	3.68E-01
GO:0060664	epithelial cell proliferation involved in salivary gland morphogenesis	BP	6.07E-03	3.68E-01
GO:0051049	regulation of transport	BP	6.09E-03	3.68E-01
GO:0031622	positive regulation of fever generation	BP	6.09E-03	3.68E-01
GO:0033008	positive regulation of mast cell activation involved in immune response	BP	6.20E-03	3.73E-01
GO:0043306	positive regulation of mast cell degranulation	BP	6.20E-03	3.73E-01
GO:1902262	apoptotic process involved in patterning of blood vessels	BP	6.25E-03	3.74E-01
GO:0086008	voltage-gated potassium channel activity involved in cardiac muscle cell action potential repolarization	MF	6.26E-03	3.74E-01
GO:0030656	regulation of vitamin metabolic process	BP	6.29E-03	3.74E-01
GO:0002385	mucosal immune response	BP	6.30E-03	3.74E-01
GO:0051798	positive regulation of hair follicle development	BP	6.30E-03	3.74E-01
GO:0090241	negative regulation of histone H4 acetylation	BP	6.33E-03	3.74E-01
GO:0042368	vitamin D biosynthetic process	BP	6.34E-03	3.74E-01
GO:0046649	lymphocyte activation	BP	6.38E-03	3.74E-01
GO:0033036	macromolecule localization	BP	6.40E-03	3.74E-01
GO:0045073	regulation of chemokine biosynthetic process	BP	6.41E-03	3.74E-01
GO:0071675	regulation of mononuclear cell migration	BP	6.44E-03	3.74E-01
GO:0042033	chemokine biosynthetic process	BP	6.44E-03	3.74E-01
GO:1900424	regulation of defense response to bacterium	BP	6.47E-03	3.74E-01
GO:0005126	cytokine receptor binding	MF	6.50E-03	3.74E-01
GO:0042228	interleukin-8 biosynthetic process	BP	6.51E-03	3.74E-01
GO:0050755	chemokine metabolic process	BP	6.51E-03	3.74E-01
GO:0031399	regulation of protein modification process	BP	6.55E-03	3.75E-01
GO:0071674	mononuclear cell migration	BP	6.59E-03	3.77E-01
GO:0050789	regulation of biological process	BP	6.63E-03	3.77E-01
GO:0002739	regulation of cytokine secretion involved in immune response	BP	6.65E-03	3.77E-01
GO:0034135	regulation of toll-like receptor 2 signaling pathway	BP	6.69E-03	3.77E-01
GO:0046643	regulation of gamma-delta T cell activation	BP	6.75E-03	3.77E-01
GO:0072608	interleukin-10 secretion	BP	6.77E-03	3.77E-01
GO:0032660	regulation of interleukin-17 production	BP	6.78E-03	3.77E-01
GO:0072603	interleukin-5 secretion	BP	6.78E-03	3.77E-01
GO:2000662	regulation of interleukin-5 secretion	BP	6.78E-03	3.77E-01
GO:0072611	interleukin-13 secretion	BP	6.78E-03	3.77E-01
GO:2000665	regulation of interleukin-13 secretion	BP	6.78E-03	3.77E-01
GO:0016045	detection of bacterium	BP	6.82E-03	3.77E-01
GO:0098543	detection of other organism	BP	6.82E-03	3.77E-01
GO:0061046	regulation of branching involved in lung morphogenesis	BP	6.94E-03	3.82E-01
GO:1901550	regulation of endothelial cell development	BP	6.96E-03	3.82E-01
GO:1903140	regulation of establishment of endothelial barrier	BP	6.96E-03	3.82E-01
GO:0019899	enzyme binding	MF	6.98E-03	3.82E-01
GO:2000010	positive regulation of protein localization to cell surface	BP	7.02E-03	3.82E-01
GO:0050700	CARD domain binding	MF	7.04E-03	3.82E-01
GO:0001818	negative regulation of cytokine production	BP	7.04E-03	3.82E-01
GO:0051338	regulation of transferase activity	BP	7.06E-03	3.82E-01
GO:2000109	regulation of macrophage apoptotic process	BP	7.14E-03	3.83E-01
GO:0045184	establishment of protein localization	BP	7.15E-03	3.83E-01
GO:0051767	nitric-oxide synthase biosynthetic process	BP	7.18E-03	3.83E-01
GO:0051769	regulation of nitric-oxide synthase biosynthetic process	BP	7.18E-03	3.83E-01
GO:0070431	nucleotide-binding oligomerization domain containing 2 signaling pathway	BP	7.19E-03	3.83E-01
GO:0060452	positive regulation of cardiac muscle contraction	BP	7.19E-03	3.83E-01
GO:0072567	chemokine (C-X-C motif) ligand 2 production	BP	7.39E-03	3.93E-01
GO:0035581	sequestering of extracellular ligand from receptor	BP	7.42E-03	3.94E-01
GO:0006915	apoptotic process	BP	7.45E-03	3.94E-01
GO:0033005	positive regulation of mast cell activation	BP	7.48E-03	3.95E-01
GO:0042362	fat-soluble vitamin biosynthetic process	BP	7.53E-03	3.97E-01
GO:0044707	single-multicellular organism process	BP	7.64E-03	4.01E-01
GO:0046328	regulation of JNK cascade	BP	7.71E-03	4.02E-01
GO:0050870	positive regulation of T cell activation	BP	7.74E-03	4.02E-01
GO:0051023	regulation of immunoglobulin secretion	BP	7.75E-03	4.02E-01
GO:0010562	positive regulation of phosphorus metabolic process	BP	7.76E-03	4.02E-01
GO:0045937	positive regulation of phosphate metabolic process	BP	7.76E-03	4.02E-01
GO:0012501	programmed cell death	BP	7.81E-03	4.02E-01
GO:0086013	membrane repolarization during cardiac muscle cell action potential	BP	7.81E-03	4.02E-01
GO:0009719	response to endogenous stimulus	BP	7.86E-03	4.02E-01
GO:1902260	negative regulation of delayed rectifier potassium channel activity	BP	7.87E-03	4.02E-01

GO ID	GO Function	Ontology	P. Value	FDR
GO:0043203	axon hillock	CC	7.87E-03	4.02E-01
GO:0002251	organ or tissue specific immune response	BP	7.88E-03	4.02E-01
GO:0032515	negative regulation of phosphoprotein phosphatase activity	BP	7.91E-03	4.03E-01
GO:0031620	regulation of fever generation	BP	8.11E-03	4.12E-01
GO:0071888	macrophage apoptotic process	BP	8.13E-03	4.12E-01
GO:0045637	regulation of myeloid cell differentiation	BP	8.20E-03	4.15E-01
GO:2000121	regulation of removal of superoxide radicals	BP	8.25E-03	4.16E-01
GO:0034112	positive regulation of homotypic cell-cell adhesion	BP	8.35E-03	4.19E-01
GO:1903039	positive regulation of leukocyte cell-cell adhesion	BP	8.36E-03	4.19E-01
GO:0002573	myeloid leukocyte differentiation	BP	8.36E-03	4.19E-01
GO:0009968	negative regulation of signal transduction	BP	8.38E-03	4.19E-01
GO:0000185	activation of MAPKKK activity	BP	8.41E-03	4.19E-01
GO:0042635	positive regulation of hair cycle	BP	8.45E-03	4.20E-01
GO:0071316	cellular response to nicotine	BP	8.47E-03	4.20E-01
GO:0050901	leukocyte tethering or rolling	BP	8.52E-03	4.20E-01
GO:0061756	leukocyte adhesion to vascular endothelial cell	BP	8.52E-03	4.20E-01
GO:0051533	positive regulation of NFAT protein import into nucleus	BP	8.52E-03	4.20E-01
GO:0043122	regulation of I-kappaB kinase/NF-kappaB signaling	BP	8.58E-03	4.21E-01
GO:0071550	death-inducing signaling complex assembly	BP	8.59E-03	4.21E-01
GO:0002724	regulation of T cell cytokine production	BP	8.63E-03	4.21E-01
GO:0034116	positive regulation of heterotypic cell-cell adhesion	BP	8.65E-03	4.21E-01
GO:0032695	negative regulation of interleukin-12 production	BP	8.66E-03	4.21E-01
GO:0032620	interleukin-17 production	BP	8.67E-03	4.21E-01
GO:0010692	regulation of alkaline phosphatase activity	BP	8.82E-03	4.26E-01
GO:0051704	multi-organism process	BP	8.82E-03	4.26E-01
GO:0098581	detection of external biotic stimulus	BP	8.86E-03	4.26E-01
GO:0050718	positive regulation of interleukin-1 beta secretion	BP	8.87E-03	4.26E-01
GO:0050716	positive regulation of interleukin-1 secretion	BP	8.87E-03	4.26E-01
GO:0002831	regulation of response to biotic stimulus	BP	8.90E-03	4.26E-01
GO:0051716	cellular response to stimulus	BP	8.94E-03	4.27E-01
GO:0060081	membrane hyperpolarization	BP	8.95E-03	4.27E-01
GO:0002888	positive regulation of myeloid leukocyte mediated immunity	BP	9.04E-03	4.29E-01
GO:0071801	regulation of podosome assembly	BP	9.06E-03	4.29E-01
GO:0045624	positive regulation of T-helper cell differentiation	BP	9.06E-03	4.29E-01
GO:2000810	regulation of bicellular tight junction assembly	BP	9.11E-03	4.30E-01
GO:0006897	endocytosis	BP	9.13E-03	4.30E-01
GO:1903817	negative regulation of voltage-gated potassium channel activity	BP	9.20E-03	4.32E-01
GO:0043372	positive regulation of CD4-positive, alpha-beta T cell differentiation	BP	9.21E-03	4.32E-01
GO:0043410	positive regulation of MAPK cascade	BP	9.28E-03	4.34E-01
GO:0042221	response to chemical	BP	9.30E-03	4.35E-01
GO:0042346	positive regulation of NF-kappaB import into nucleus	BP	9.39E-03	4.37E-01
GO:0015271	outward rectifier potassium channel activity	MF	9.53E-03	4.40E-01
GO:0008219	cell death	BP	9.58E-03	4.40E-01
GO:0016265	death	BP	9.58E-03	4.40E-01
GO:1900015	regulation of cytokine production involved in inflammatory response	BP	9.61E-03	4.40E-01
GO:0031652	positive regulation of heat generation	BP	9.63E-03	4.40E-01
GO:0008283	cell proliferation	BP	9.64E-03	4.40E-01
GO:0086014	atrial cardiac muscle cell action potential	BP	9.65E-03	4.40E-01
GO:0086026	atrial cardiac muscle cell to AV node cell signaling	BP	9.65E-03	4.40E-01
GO:0086066	atrial cardiac muscle cell to AV node cell communication	BP	9.65E-03	4.40E-01
GO:1902895	positive regulation of pri-miRNA transcription from RNA polymerase II promoter	BP	9.65E-03	4.40E-01
GO:0045628	regulation of T-helper 2 cell differentiation	BP	9.77E-03	4.44E-01
GO:0071356	cellular response to tumor necrosis factor	BP	9.81E-03	4.44E-01
GO:0002764	immune response-regulating signaling pathway	BP	9.82E-03	4.44E-01
GO:0060693	regulation of branching involved in salivary gland morphogenesis	BP	9.90E-03	4.47E-01
GO:0032754	positive regulation of interleukin-5 production	BP	9.91E-03	4.47E-01
GO:0043302	positive regulation of leukocyte degranulation	BP	9.95E-03	4.47E-01
GO:0032689	negative regulation of interferon-gamma production	BP	9.97E-03	4.47E-01
GO:0071435	potassium ion export	BP	9.99E-03	4.47E-01

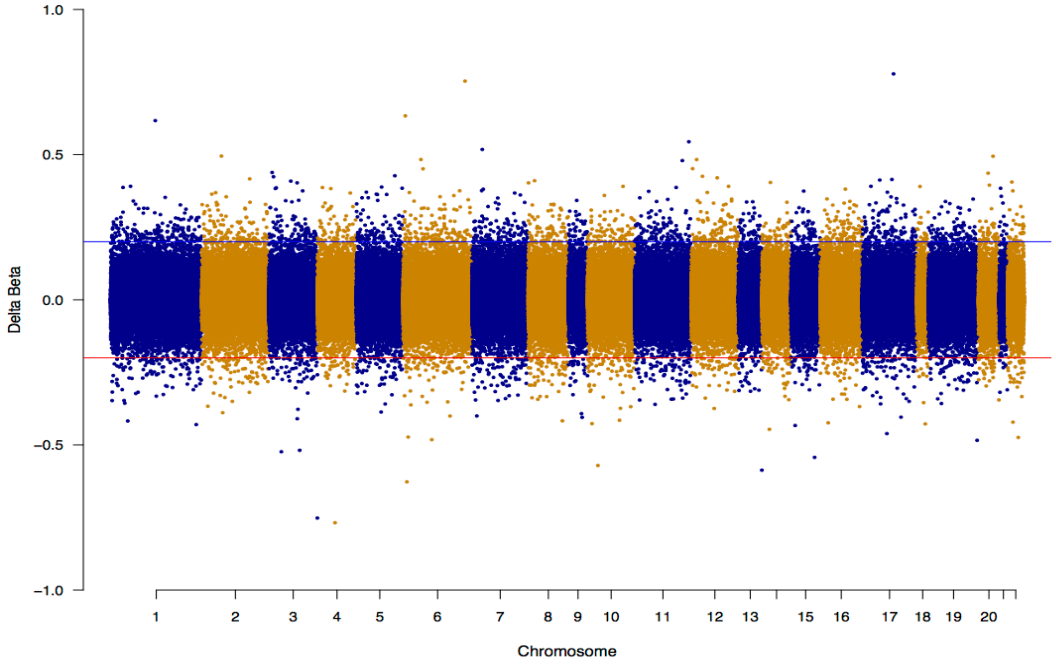
Appendix 7: Cross-platform validation of the two top DMRs, *LTA* and *LIME1* between HM450 and EpiTYPER platforms. Pearson's correlation coefficients for each probe are shown. The scale of both axes reflects a methylation value between 0 and 1 (β). The regression lines are shown in black. Based on the r-value (correlation coefficient), correlations across both platforms are shown. The p-value indicates the significance of the correlation.

Gene	HM450 probe	Sequenom CpG unit	Genomic coordinates	Correlation coefficient (r)	P-value of correlation coefficient (r)
LIME1	cg24631526	CpG_10.11	chr20:62367961	0.11	0.65
LIME1	cg06653796	CpG_1.2	chr20:62367805	0.4	0.2
LIME1	cg21201401	CpG_6.7.8	chr20:62367884	0.88	0.00016
LTA	cg14597739 / cg16219283	CG10.11	chr6: 31539998	0.69	0.00014
LTA	cg21999229	CPG9	chr6: 31540014	0.21	0.34
LTA	cg11586857	CPG2	chr6: 31540136	0.59	0.043

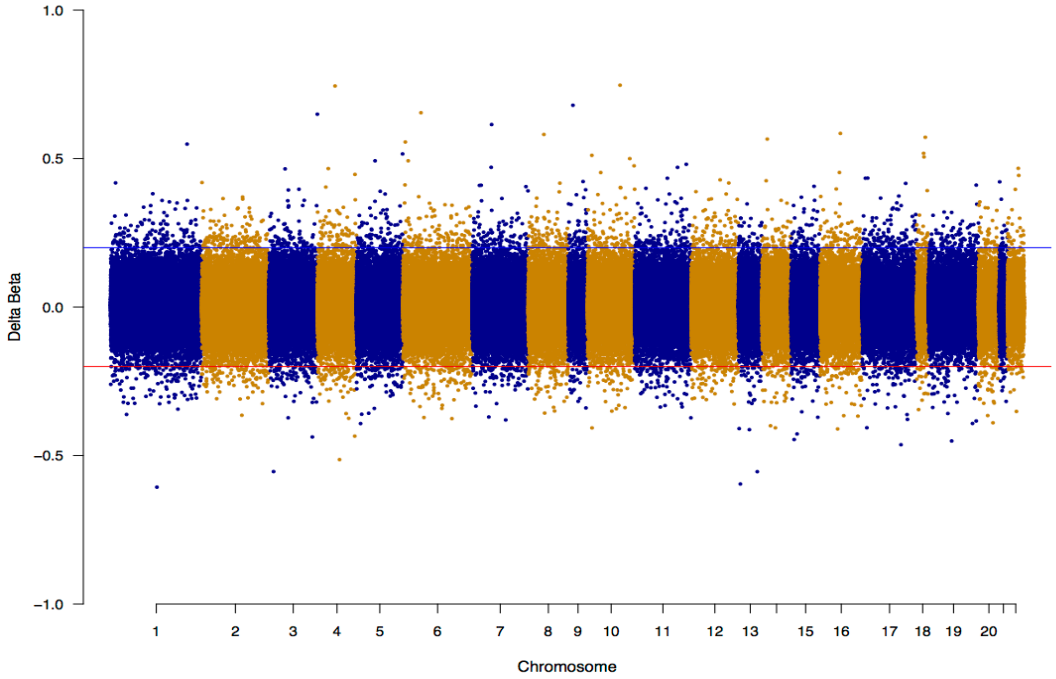
Appendix 8. DNA methylation differences within each discordant CP twin pair, identifying numerous loci showing large DNA methylation differences within each discordant twin pair.



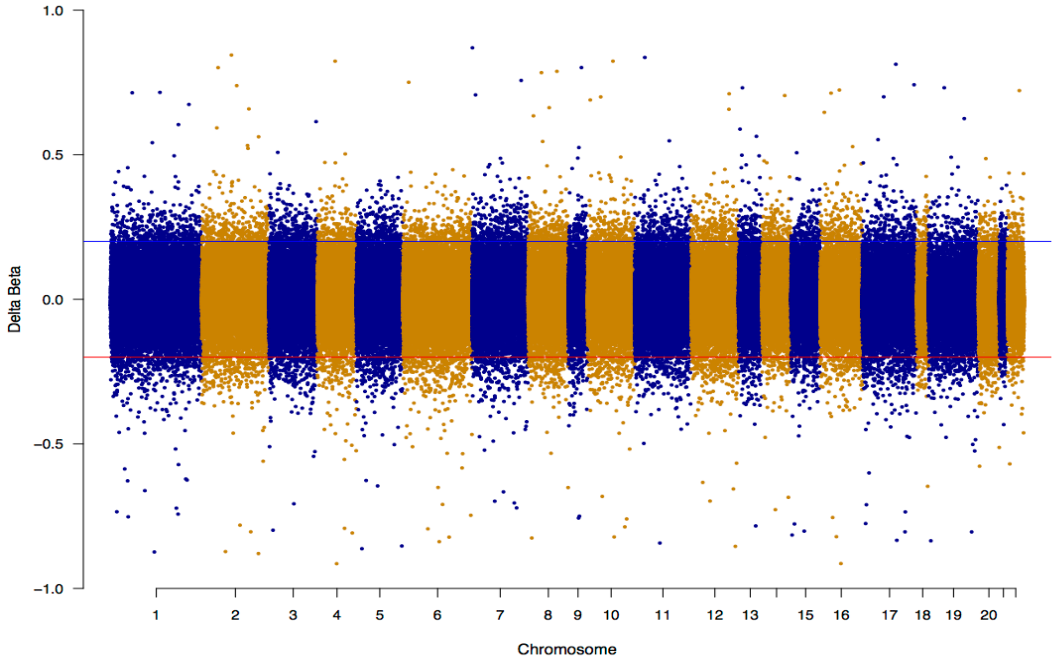
Twin pair 2



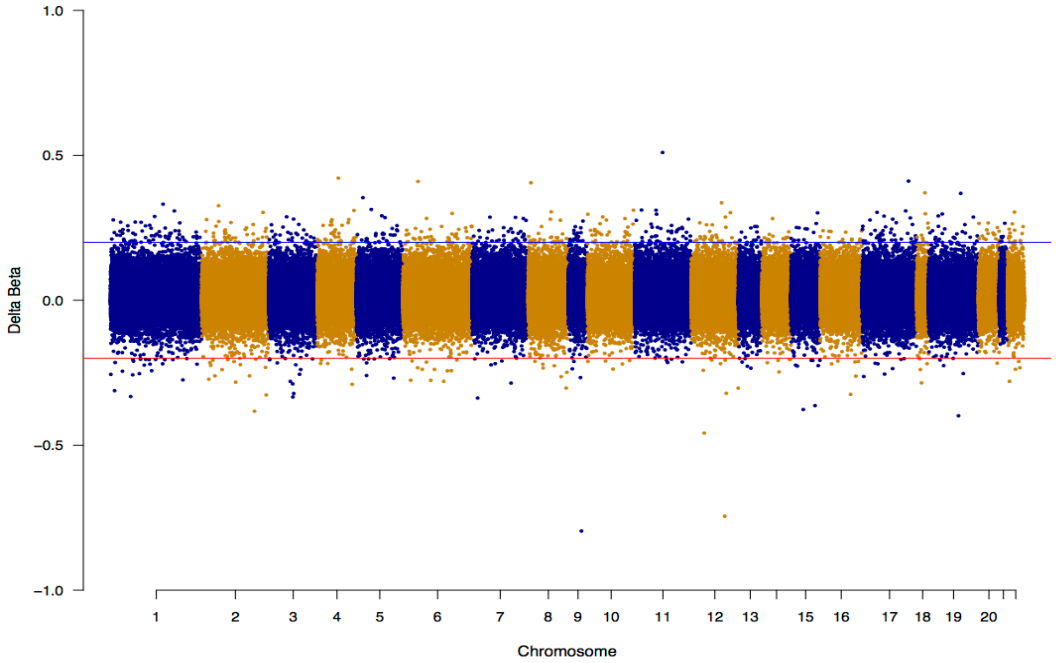
Twin pair 4



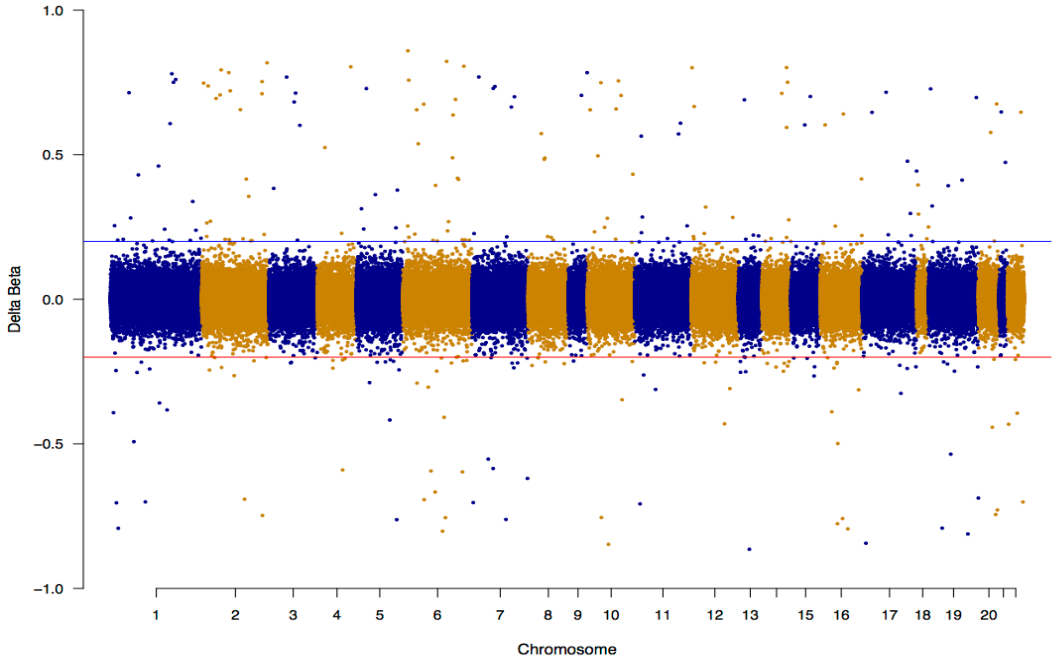
Twin pair 5



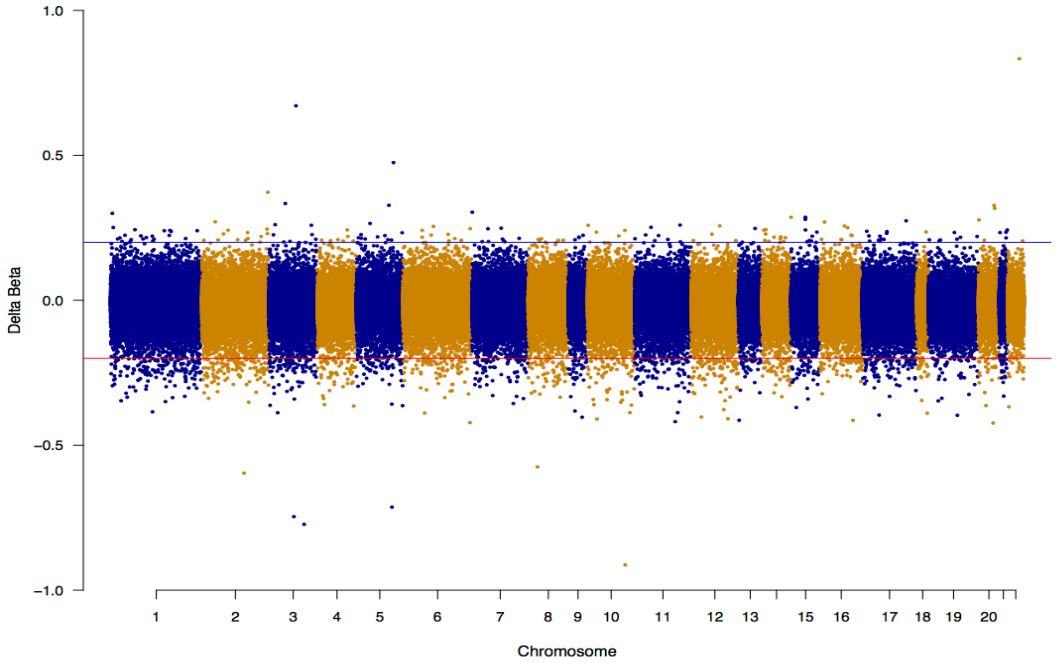
Twin pair 6



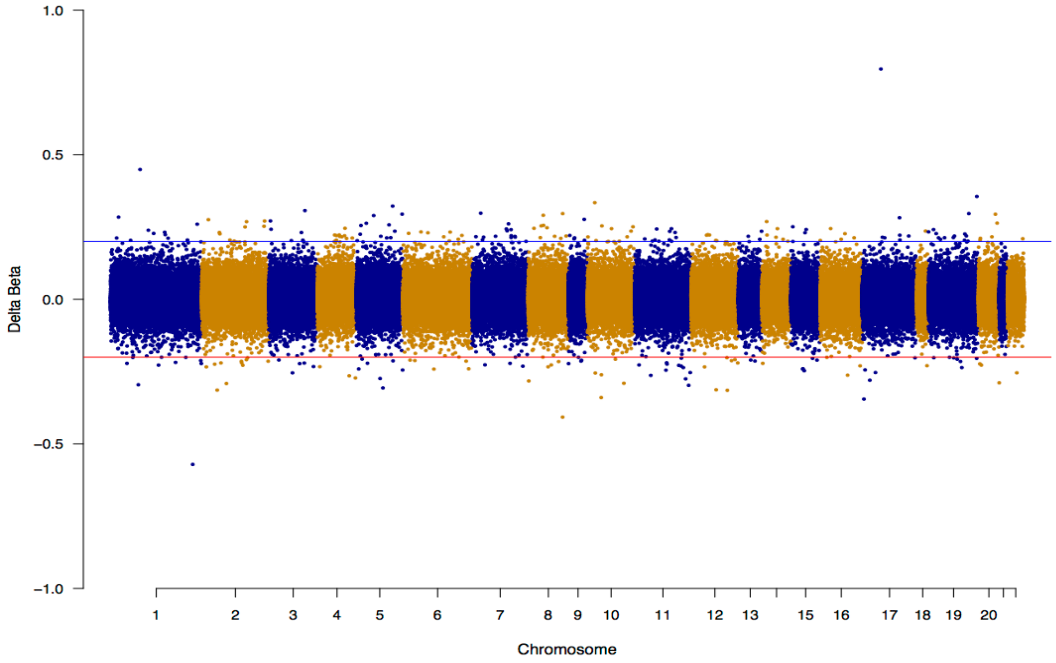
Twin pair 7



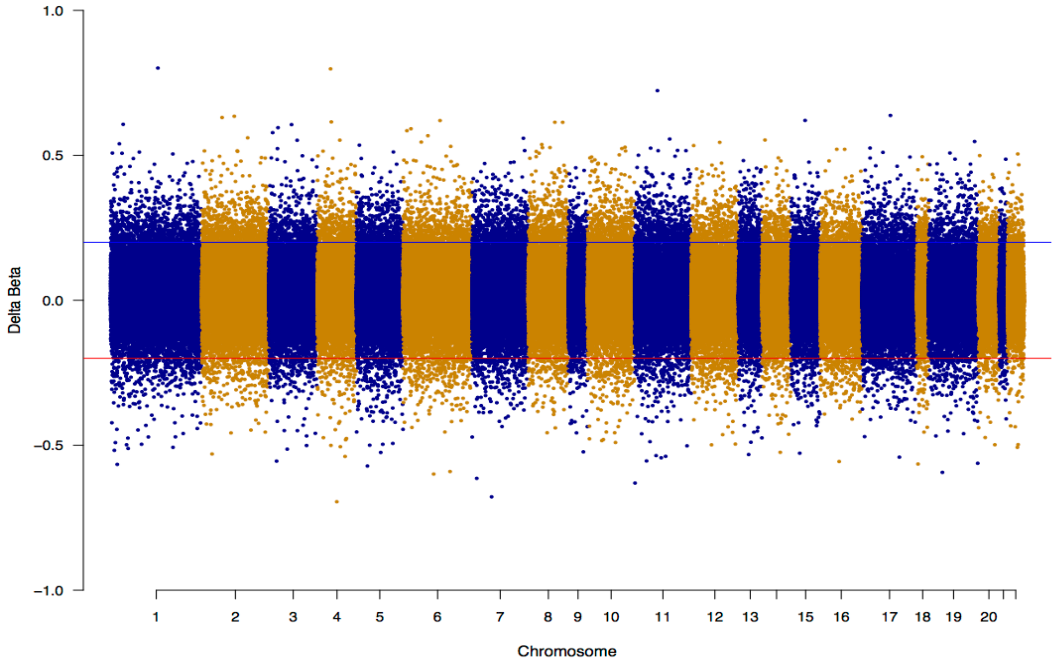
Twin pair 8



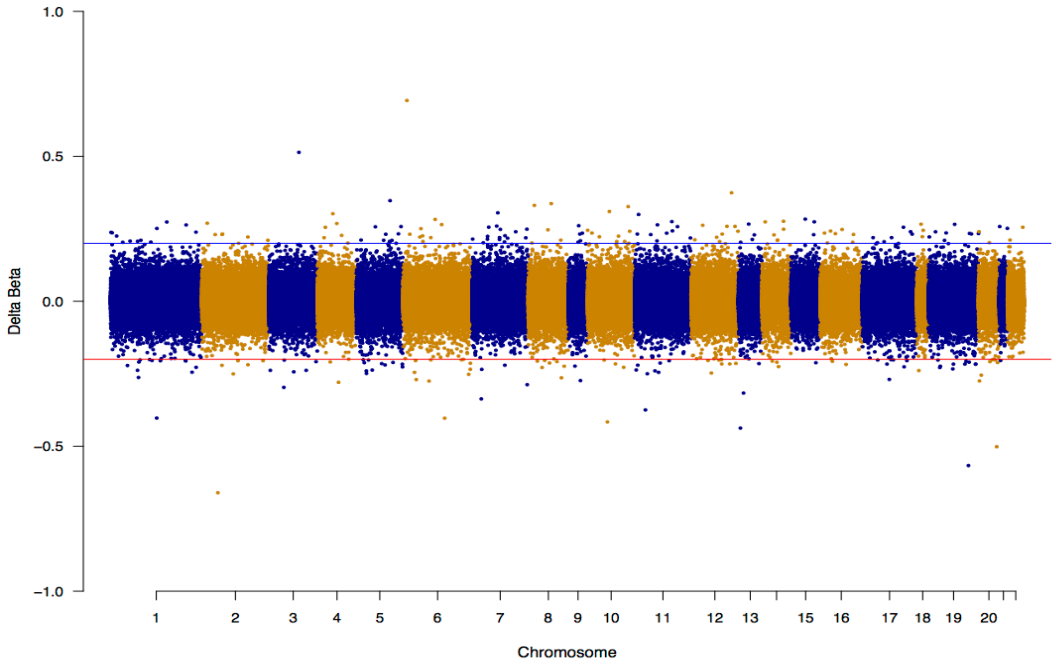
Twin pair 9



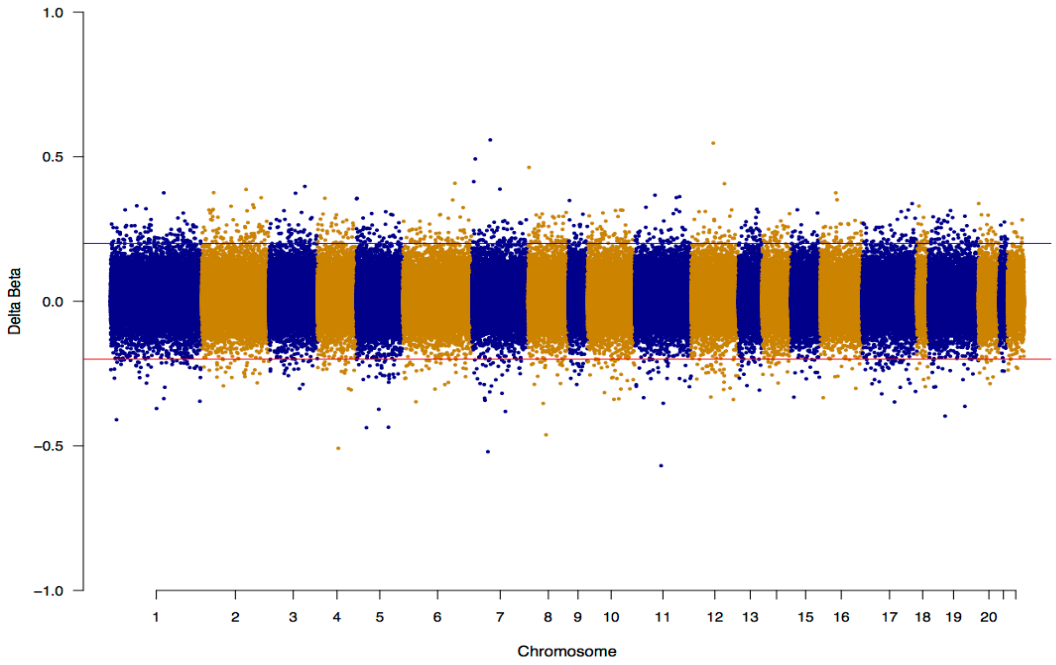
Twin pair 10



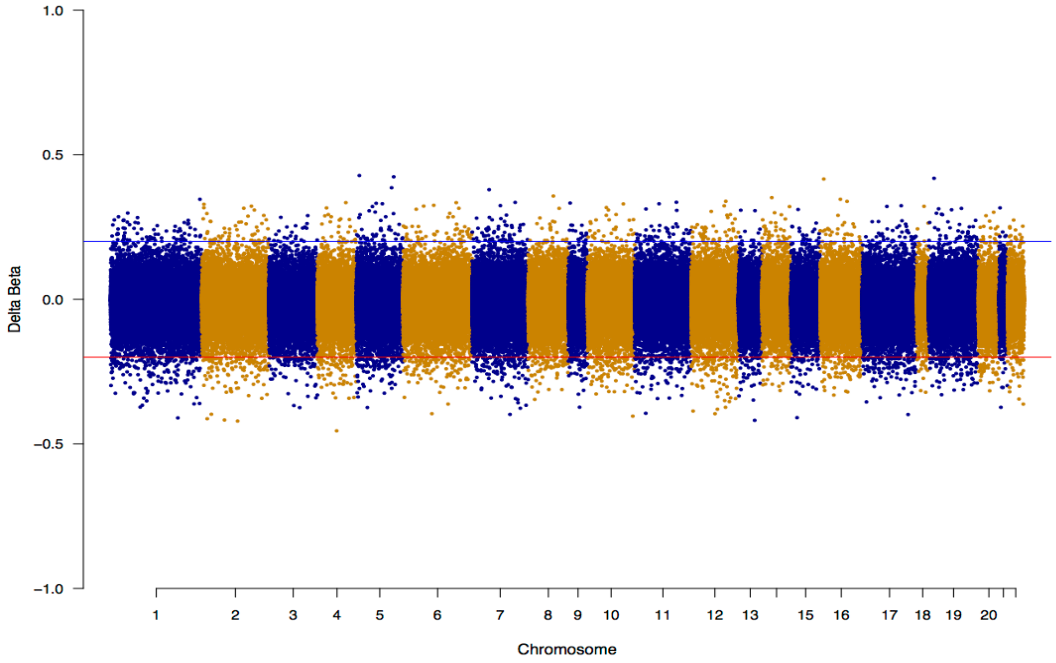
Twin pair 11



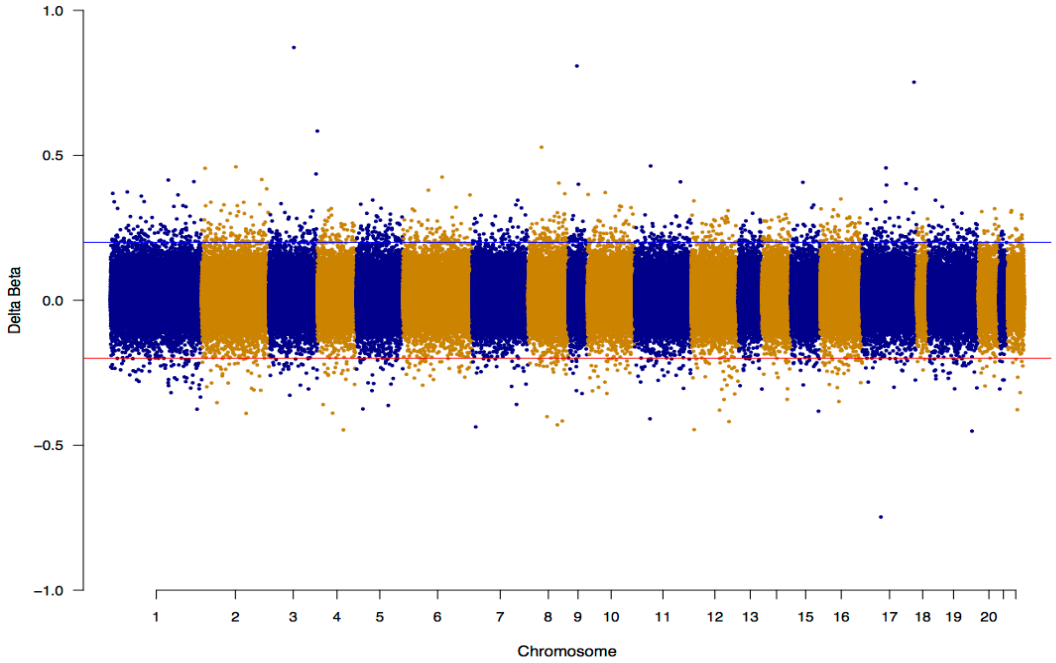
Twin pair 12



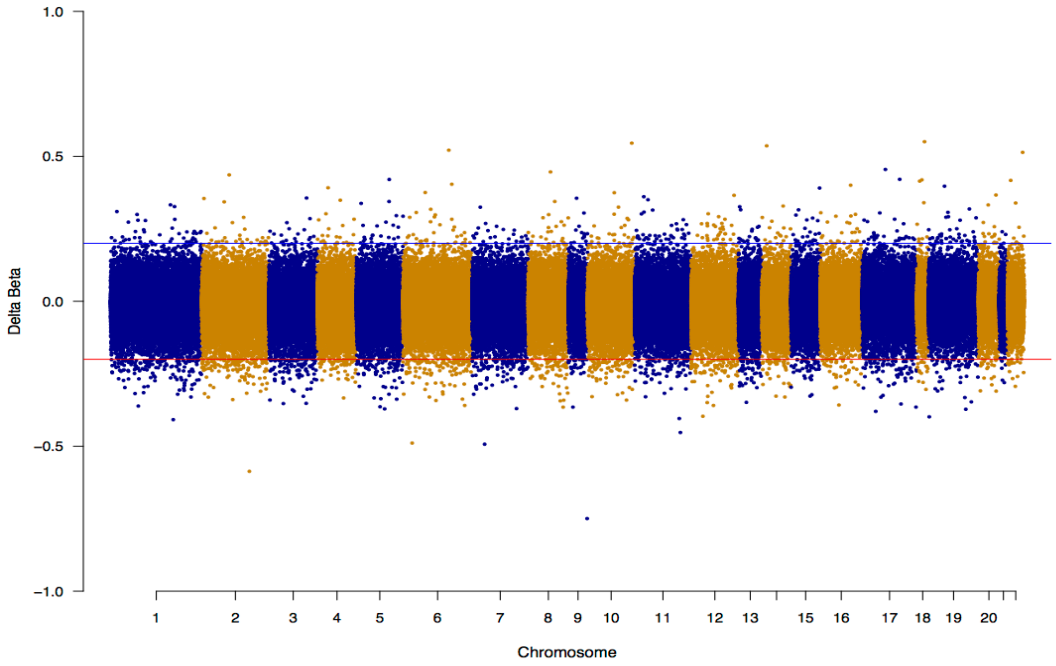
Twin pair 13



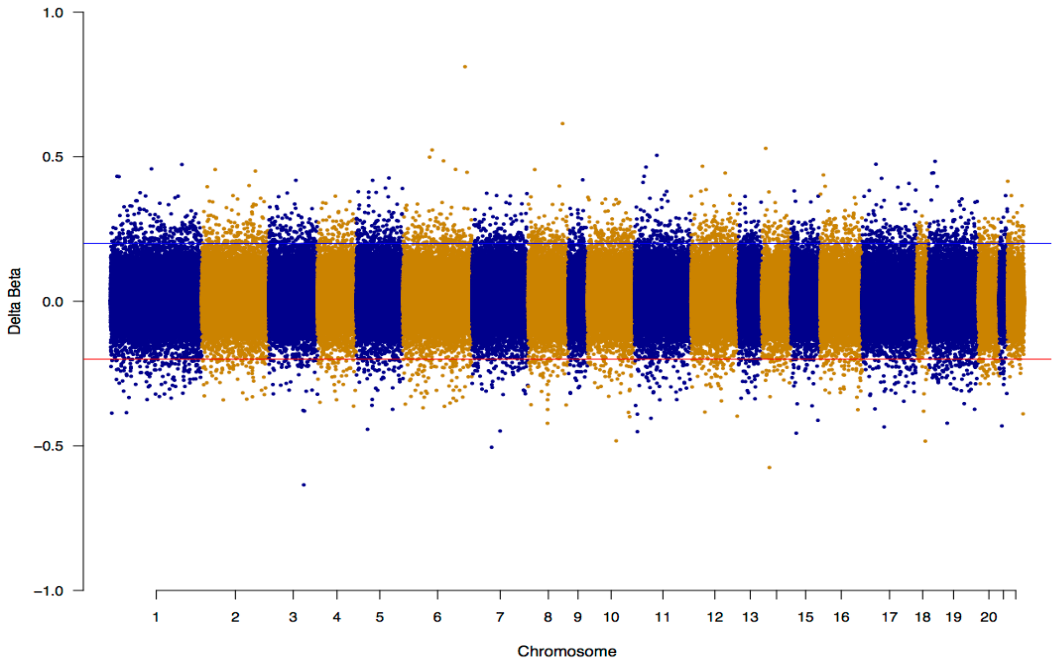
Twin pair 14



Twin pair 15



Twin pair 16



Appendix 9. Gene ontology (biological process) common to multiple twin pairs.

GO ID	Gene ontology terms (biological process)	Twin pairs				
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	TW4	TW10	TW6	TW14	TW15
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	TW6	TW10	TW15		
GO:0098609	cell-cell adhesion	TW6	TW10	TW15		
GO:0007155	cell adhesion	TW6	TW10	TW15		
GO:0022610	biological adhesion	TW10	TW15			
GO:0015821	methionine transport	TW4	TW5			
GO:0070165	positive regulation of adiponectin secretion	TW5	TW16			
GO:0010224	response to UV-B	TW5	TW14			
GO:0031443	fast-twitch skeletal muscle fiber contraction	TW6	TW12			
GO:0031448	positive regulation of fast-twitch skeletal muscle fiber contraction	TW6	TW12			
GO:0031446	regulation of fast-twitch skeletal muscle fiber contraction	TW6	TW12			
GO:2001235	positive regulation of apoptotic signaling pathway	TW9	TW13			
GO:0030683	evasion or tolerance by virus of host immune response	TW10	TW14			
GO:0051807	evasion or tolerance of defense response of other organism involved in symbiotic interaction	TW10	TW14			
GO:0030682	evasion or tolerance of host defense response	TW10	TW14			

Twin pair 1	GO terms (biological process)
GO:0022027	interkinetic nuclear migration
GO:0010818	T cell chemotaxis
GO:1900152	negative regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay
GO:0060212	negative regulation of nuclear-transcribed mRNA poly(A) tail shortening
GO:0009435	NAD biosynthetic process
GO:0042396	phosphagen biosynthetic process
GO:0006599	phosphagen metabolic process
GO:0046314	phosphocreatine biosynthetic process
GO:0006603	phosphocreatine metabolic process
GO:0019359	nicotinamide nucleotide biosynthetic process
GO:0019363	pyridine nucleotide biosynthetic process
GO:0019357	nicotinate nucleotide biosynthetic process
GO:0046497	nicotinate nucleotide metabolic process
GO:0019358	nicotinate nucleotide salvage
GO:0019365	pyridine nucleotide salvage
GO:0034453	microtubule anchoring
GO:0072525	pyridine-containing compound biosynthetic process
GO:0045200	establishment of neuroblast polarity
GO:0045196	establishment or maintenance of neuroblast polarity
GO:1903640	negative regulation of gastrin-induced gastric acid secretion

Twin pair 2	GO terms (biological process)
GO:0061218	negative regulation of mesonephros development
GO:0072179	nephric duct formation
GO:0072172	mesonephric tubule formation
GO:0080058	protein deglutathionylation
GO:0048172	regulation of short-term neuronal synaptic plasticity
GO:0072178	nephric duct morphogenesis
GO:0048168	regulation of neuronal synaptic plasticity
GO:0006641	triglyceride metabolic process
GO:0006639	acylglycerol metabolic process
GO:0072176	nephric duct development
GO:0006638	neutral lipid metabolic process
GO:0009415	response to water
GO:0050435	beta-amyloid metabolic process
GO:0072079	nephron tubule formation
GO:0001676	long-chain fatty acid metabolic process
GO:0097285	cell-type specific apoptotic process
GO:0090185	negative regulation of kidney development
GO:0033559	unsaturated fatty acid metabolic process
GO:0035695	mitophagy by induced vacuole formation
GO:0007174	epidermal growth factor catabolic process

Twin pair 4	GO terms (biological process)
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules
GO:0098609	cell-cell adhesion
GO:0007155	cell adhesion
GO:0022610	biological adhesion
GO:0035747	natural killer cell chemotaxis
GO:0002679	respiratory burst involved in defense response
GO:0072672	neutrophil extravasation
GO:0010818	T cell chemotaxis
GO:0071435	potassium ion export
GO:0097284	hepatocyte apoptotic process
GO:0015821	methionine transport
GO:0036372	opsin transport
GO:0014065	phosphatidylinositol 3-kinase signaling
GO:0036092	phosphatidylinositol-3-phosphate biosynthetic process
GO:0001779	natural killer cell differentiation
GO:0045730	respiratory burst
GO:1900087	positive regulation of G1/S transition of mitotic cell cycle
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway
GO:0072678	T cell migration

Twin pair 5	GO terms (biological process)
GO:0042756	drinking behavior
GO:0045084	positive regulation of interleukin-12 biosynthetic process
GO:0042090	interleukin-12 biosynthetic process
GO:0045075	regulation of interleukin-12 biosynthetic process
GO:0033031	positive regulation of neutrophil apoptotic process
GO:0060083	smooth muscle contraction involved in micturition
GO:0032228	regulation of synaptic transmission, GABAergic
GO:0046541	saliva secretion
GO:0051967	negative regulation of synaptic transmission, glutamatergic
GO:0051932	synaptic transmission, GABAergic
GO:0014832	urinary bladder smooth muscle contraction
GO:0014848	urinary tract smooth muscle contraction
GO:0009451	RNA modification
GO:0060073	micturition
GO:0070165	positive regulation of adiponectin secretion
GO:0006400	tRNA modification
GO:0015821	methionine transport
GO:0090335	regulation of brown fat cell differentiation
GO:0010224	response to UV-B
GO:0007585	respiratory gaseous exchange

Twin pair 6	GO terms (biological process)
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules
GO:2000049	positive regulation of cell-cell adhesion mediated by cadherin
GO:0045989	positive regulation of striated muscle contraction
GO:2000047	regulation of cell-cell adhesion mediated by cadherin
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules
GO:0044331	cell-cell adhesion mediated by cadherin
GO:0042346	positive regulation of NF-kappaB import into nucleus
GO:0032469	endoplasmic reticulum calcium ion homeostasis
GO:0031443	fast-twitch skeletal muscle fiber contraction
GO:0031448	positive regulation of fast-twitch skeletal muscle fiber contraction
GO:0031446	regulation of fast-twitch skeletal muscle fiber contraction
GO:0098609	cell-cell adhesion
GO:0035425	autocrine signaling
GO:2000330	positive regulation of T-helper 17 cell lineage commitment
GO:0003342	proepicardium development
GO:0003343	septum transversum development
GO:0006942	regulation of striated muscle contraction
GO:1900039	positive regulation of cellular response to hypoxia
GO:0042520	positive regulation of tyrosine phosphorylation of Stat4 protein
GO:0007155	cell adhesion

Twin pair 7	GO terms (biological process)
GO:0003183	mitral valve morphogenesis
GO:0003174	mitral valve development
GO:0001701	in utero embryonic development
GO:0003181	atrioventricular valve morphogenesis
GO:0043009	chordate embryonic development
GO:0009792	embryo development ending in birth or egg hatching
GO:0003171	atrioventricular valve development
GO:0042110	T cell activation
GO:0070489	T cell aggregation
GO:0071593	lymphocyte aggregation
GO:0070486	leukocyte aggregation
GO:0003179	heart valve morphogenesis
GO:0070231	T cell apoptotic process
GO:0003197	endocardial cushion development
GO:0007159	leukocyte cell-cell adhesion
GO:0002366	leukocyte activation involved in immune response
GO:0046649	lymphocyte activation
GO:0002263	cell activation involved in immune response
GO:0050863	regulation of T cell activation
GO:0003170	heart valve development

Twin pair 8	GO terms (biological process)
GO:0072386	plus-end-directed organelle transport along microtubule
GO:0072383	plus-end-directed vesicle transport along microtubule
GO:0010628	positive regulation of gene expression
GO:1902680	positive regulation of RNA biosynthetic process
GO:0046640	regulation of alpha-beta T cell proliferation
GO:1990765	colon smooth muscle contraction
GO:1990768	gastric mucosal blood circulation
GO:1990764	myofibroblast contraction
GO:1904326	negative regulation of circadian sleep/wake cycle, wakefulness
GO:1904343	positive regulation of colon smooth muscle contraction
GO:1904346	positive regulation of gastric mucosal blood circulation
GO:1904325	positive regulation of inhibitory G-protein coupled receptor phosphorylation
GO:1904330	positive regulation of myofibroblast contraction
GO:1904320	positive regulation of smooth muscle contraction involved in micturition
GO:1990767	prostaglandin receptor internalization
GO:1904341	regulation of colon smooth muscle contraction
GO:1904344	regulation of gastric mucosal blood circulation
GO:1904323	regulation of inhibitory G-protein coupled receptor phosphorylation
GO:1904328	regulation of myofibroblast contraction
GO:1904318	regulation of smooth muscle contraction involved in micturition

Twin pair 9	GO terms (biological process)
GO:0019233	sensory perception of pain
GO:0019725	cellular homeostasis
GO:0055082	cellular chemical homeostasis
GO:1901216	positive regulation of neuron death
GO:0071287	cellular response to manganese ion
GO:0050432	catecholamine secretion
GO:0048878	chemical homeostasis
GO:0050877	neurological system process
GO:0007600	sensory perception
GO:0046296	glycolate catabolic process
GO:0042415	norepinephrine metabolic process
GO:2001235	positive regulation of apoptotic signaling pathway
GO:0072507	divalent inorganic cation homeostasis
GO:0051937	catecholamine transport
GO:0055065	metal ion homeostasis
GO:0010942	positive regulation of cell death
GO:0015844	monoamine transport
GO:2000016	negative regulation of determination of dorsal identity
GO:0043525	positive regulation of neuron apoptotic process
GO:0003008	system process

Twin pair 10	GO terms (biological process)
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules
GO:0098609	cell-cell adhesion
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules
GO:0007155	cell adhesion
GO:0022610	biological adhesion
GO:0043113	receptor clustering
GO:0097120	receptor localization to synapse
GO:0030683	evasion or tolerance by virus of host immune response
GO:0051807	evasion or tolerance of defense response of other organism involved in symbiotic interaction
GO:0030682	evasion or tolerance of host defense response
GO:0020012	evasion or tolerance of host immune response
GO:0051805	evasion or tolerance of immune response of other organism involved in symbiotic interaction
GO:0070292	N-acylphosphatidylethanolamine metabolic process
GO:0051249	regulation of lymphocyte activation
GO:0007127	meiosis I
GO:0071425	hematopoietic stem cell proliferation
GO:0031627	telomeric loop formation
GO:0002694	regulation of leukocyte activation
GO:0051701	interaction with host
GO:0007129	synapsis

Twin pair 11	GO terms (biological process)
GO:0097118	neuroligin clustering involved in postsynaptic membrane assembly
GO:0046085	adenosine metabolic process
GO:0051898	negative regulation of protein kinase B signaling
GO:2000649	regulation of sodium ion transmembrane transporter activity
GO:0007045	cell-substrate adherens junction assembly
GO:0048041	focal adhesion assembly
GO:0031589	cell-substrate adhesion
GO:0097104	postsynaptic membrane assembly
GO:0034333	adherens junction assembly
GO:0032233	positive regulation of actin filament bundle assembly
GO:1902305	regulation of sodium ion transmembrane transport
GO:0051495	positive regulation of cytoskeleton organization
GO:0007044	cell-substrate junction assembly
GO:0016311	dephosphorylation
GO:0022898	regulation of transmembrane transporter activity
GO:0071602	phytosphingosine biosynthetic process
GO:0006671	phytosphingosine metabolic process
GO:0032409	regulation of transporter activity
GO:0046086	adenosine biosynthetic process
GO:0072009	nephron epithelium development

Twin pair 12	GO terms (biological process)
GO:0014855	striated muscle cell proliferation
GO:0097264	self proteolysis
GO:0002879	positive regulation of acute inflammatory response to non-antigenic stimulus
GO:0060038	cardiac muscle cell proliferation
GO:0010738	regulation of protein kinase A signaling
GO:0002232	leukocyte chemotaxis involved in inflammatory response
GO:0018872	arsonoacetate metabolic process
GO:0000915	actomyosin contractile ring assembly
GO:0044837	actomyosin contractile ring organization
GO:0000912	assembly of actomyosin apparatus involved in cytokinesis
GO:0010737	protein kinase A signaling
GO:0002752	cell surface pattern recognition receptor signaling pathway
GO:0055017	cardiac muscle tissue growth
GO:0002357	defense response to tumor cell
GO:0031443	fast-twitch skeletal muscle fiber contraction
GO:0031448	positive regulation of fast-twitch skeletal muscle fiber contraction
GO:0031446	regulation of fast-twitch skeletal muscle fiber contraction
GO:1902728	positive regulation of growth factor dependent skeletal muscle satellite cell proliferation
GO:1902724	positive regulation of skeletal muscle satellite cell proliferation
GO:0097089	methyl-branched fatty acid metabolic process

Twin pair 13	GO terms (biological process)
GO:2001235	positive regulation of apoptotic signaling pathway
GO:0045471	response to ethanol
GO:0023056	positive regulation of signaling
GO:0010647	positive regulation of cell communication
GO:0051128	regulation of cellular component organization
GO:0007179	transforming growth factor beta receptor signaling pathway
GO:0002523	leukocyte migration involved in inflammatory response
GO:0034121	regulation of toll-like receptor signaling pathway
GO:0009967	positive regulation of signal transduction
GO:0007210	serotonin receptor signaling pathway
GO:0048584	positive regulation of response to stimulus
GO:0031532	actin cytoskeleton reorganization
GO:0097305	response to alcohol
GO:0010941	regulation of cell death
GO:2001233	regulation of apoptotic signaling pathway
GO:0048284	organelle fusion
GO:0071560	cellular response to transforming growth factor beta stimulus
GO:0046628	positive regulation of insulin receptor signaling pathway
GO:0071559	response to transforming growth factor beta
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway

Twin pair 14	GO terms (biological process)
GO:0045600	positive regulation of fat cell differentiation
GO:0034629	cellular protein complex localization
GO:0044332	Wnt signaling pathway involved in dorsal/ventral axis specification
GO:0048242	epinephrine secretion
GO:0061448	connective tissue development
GO:0048241	epinephrine transport
GO:0045444	fat cell differentiation
GO:0071493	cellular response to UV-B
GO:0030853	negative regulation of granulocyte differentiation
GO:0030214	hyaluronan catabolic process
GO:0010224	response to UV-B
GO:0045598	regulation of fat cell differentiation
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules
GO:0042744	hydrogen peroxide catabolic process
GO:1902466	positive regulation of histone H3-K27 trimethylation
GO:1902464	regulation of histone H3-K27 trimethylation
GO:0031295	T cell costimulation
GO:0030683	evasion or tolerance by virus of host immune response
GO:0051807	evasion or tolerance of defense response of other organism involved in symbiotic interaction
GO:0030682	evasion or tolerance of host defense response

Twin pair 15	GO terms (biological process)
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules
GO:0007155	cell adhesion
GO:0022610	biological adhesion
GO:0098609	cell-cell adhesion
GO:0032929	negative regulation of superoxide anion generation
GO:0014029	neural crest formation
GO:0061418	regulation of transcription from RNA polymerase II promoter in response to hypoxia
GO:0071456	cellular response to hypoxia
GO:0032928	regulation of superoxide anion generation
GO:0036294	cellular response to decreased oxygen levels
GO:0035455	response to interferon-alpha
GO:0071453	cellular response to oxygen levels
GO:0032695	negative regulation of interleukin-12 production
GO:2000314	negative regulation of fibroblast growth factor receptor signaling pathway involved in neural plate anterior/posterior pattern formation
GO:0090301	negative regulation of neural crest formation
GO:0090299	regulation of neural crest formation
GO:0042554	superoxide anion generation
GO:0061744	motor behavior
GO:0036343	psychomotor behavior

Twin pair 16	GO terms (biological process)
GO:0003360	brainstem development
GO:0071156	regulation of cell cycle arrest
GO:0090305	nucleic acid phosphodiester bond hydrolysis
GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent
GO:0071257	cellular response to electrical stimulus
GO:0006886	intracellular protein transport
GO:1990167	protein K27-linked deubiquitination
GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I
GO:0021535	cell migration in hindbrain
GO:0070165	positive regulation of adiponectin secretion
GO:0000075	cell cycle checkpoint
GO:0048894	efferent axon development in a lateral line nerve
GO:0048892	lateral line nerve development
GO:0048925	lateral line system development
GO:0021723	medullary reticular formation development
GO:0061451	retrotrapezoid nucleus development
GO:0061452	retrotrapezoid nucleus neuron differentiation
GO:0034613	cellular protein localization
GO:0035523	protein K29-linked deubiquitination
GO:1990168	protein K33-linked deubiquitination

Appendix 10: CpG sites (probes) within each twin pair group with an absolute methylation difference >0.5 and their corresponding genes. Genes are colour coded to highlight overlaps between twin pair groups

Probes	TW1	Methylation difference
cg19727767	*ISM1*	-0.881
cg08587775		-0.771
cg01419577		-0.563
cg14494812	*TOB1*	-0.544
cg01309395	*HLA-DPB2*	0.534
cg10333023	*KCNJ5;C11orf45*	0.614
cg25115537	*ZHX2*	0.955

Probes	TW2	Methylation difference
cg11671265	*CNOT6L*	-0.768
ch.3.343413R	*NR2C2*	-0.752
cg01309395	*HLA-DPB2*	-0.627
ch.13.1320446R		-0.587
cg05626013	*CTNNA3;LRRTM3*	-0.571
cg22724998		-0.543
cg06180356	*PLOD2*	-0.523
cg16958524	*KTELC1*	-0.519
cg04369350		0.518
cg26619099		0.544
cg13066703	*TRAF5*	0.617
cg00736459	*SIM1*	0.634
cg24761195	*LRP11*	0.753
cg15407262	*SLC2A9*	0.778

Probes	TW4	Methylation difference
cg13457410	*PIK3CD*	-0.606
cg01929377	*FAM48A*	-0.596
cg22088594	*FAM48A*	-0.554
cg02134705	*WWTR1*	-0.554
cg15102770	*ADAMTS3*	-0.514
cg18086635	*FAM59A*	0.506
cg02535060	*LRRC27*	0.511
cg27582059	*PCDHB16;PCDHB8*	0.516
cg16942681	*RIOK3*	0.518
cg23068989	*UBE2U*	0.549
cg00736459	*SIM1*	0.556
cg04584103	*ASB2*	0.566
cg20796556	*RIOK3*	0.572
cg10061374		0.581
cg06361531	*DUS2L;DDX28*	0.585
cg08779777	*PIK3CG*	0.615
ch.3.343413R	*NR2C2*	0.650
cg06512249	*GABBR1*	0.654
cg10861135		0.680
cg11671265	*CNOT6L*	0.745
cg18193259	*SLIT1*	0.747

Probes	TW5	Methylation difference
cg12950911		-0.914
cg06525293	*RHBDL1*	-0.914
cg23210365	*TACR1*	-0.879
cg12692722	*TCEA3*	-0.874
cg08901662		-0.873
cg02657836		-0.863
cg25931671	*METTL1;FAM119B*	-0.854
cg26976187	*DDX41*	-0.853
cg12114392	*KCTD21;USP35*	-0.843
cg13973820	*HMGCLL1*	-0.838
cg01287132		-0.835
cg17255878	*PSMB6*	-0.834
cg01919374		-0.826
cg18019017	*HTR1B*	-0.823
cg14846060	*PAOX*	-0.822
cg05131766	*PHKB;ITFG1*	-0.821
cg00786747		-0.815
cg24722073		-0.808
cg24154786	*MYPOP*	-0.805
cg21583412	*NUFIP2*	-0.804
cg19906093	*C2orf88*	-0.804
cg12379383		-0.802
cg01910272	*TNNC1;NISCH*	-0.799
cg09323975		-0.794
cg18948380	*LARP7*	-0.792
cg21096915	*LOC84856*	-0.787
cg20237707		-0.784
cg15048802	*C2orf60;C2orf47*	-0.781
cg02888867	*TRPM7*	-0.777
cg01684881	*FZD2*	-0.776
cg23344688	*MLLT10*	-0.760
cg14000104	*PTCH1*	-0.756
cg03965340	*TCEB2*	-0.755
cg04879451	*C1orf122;YRDC*	-0.752
cg14095101	*TRPM6*	-0.750
cg27245606	*LTB*	-0.747
cg20168230	*GRIK3*	-0.743
cg21735374	*SC65;FKBP10*	-0.735
cg01513967	*RERE*	-0.735
cg12790592	*FOXN3*	-0.728
cg19637330		-0.722
cg21718653	*HBP1*	-0.721
cg02054559	*DUSP3;C17orf105*	-0.710
cg15268184	*PDCD2*	-0.710
cg13682223	*HYAL3;NAT6*	-0.707
cg20495206	*RHEB*	-0.704
cg10348193	*LRRC61;ACTR3C*	-0.698
cg24151229	*GOLGA5*	
cg07956775	*CALHM2*	

Probes	TW5	Methylation difference
cg14933494		-0.666
cg09887955	*PTPN7*	-0.662
cg24686009	*RAP1B*	-0.656
cg27648020	*ATAD2*	-0.651
cg13479358		-0.651
cg25964954	*AFG3L2*	-0.647
cg11901248		-0.645
cg06561366	*PAN2*	-0.633
cg04715245	*IL6R*	-0.628
cg05527185	*THBS4*	-0.627
cg23057040	*RCOR3*	-0.625
cg22698629	*VWA5B1*	-0.621
cg03218192	*AP2B1*	-0.600
cg03940848		-0.587
cg23645885		-0.583
cg01017090	*GNAS*	-0.577
cg20300343		-0.572
cg02801583	*RIBC2;SMC1B*	-0.569
cg26646290	*CCDC53*	-0.567
cg25310310		-0.560
cg18772071	*NAP1L5;HERC3*	-0.554
cg25441771	*MCCC1*	-0.543
cg23861803	*NEU1*	-0.534
cg17245576	*SERINC1;PKIB*	-0.533
cg15248835	*LOC157627;MIR124-1*	-0.532
cg26106948		-0.527
cg25795456	*PRPF31;TFPT*	-0.525
ch.4.122145192R		-0.524
cg05369791	*MESTIT1;MEST*	-0.522
cg25023095	*RRP12*	-0.518
cg19314945	*HPCAL4*	-0.518
ch.20.378824F	*C20orf72*	-0.512
cg03306368	*FBXL4*	-0.510
cg00259834	*RFTN1*	-0.510
cg24371114	*CCDC149*	-0.505
cg22637837	*PCDHGB1;PCDHGA2;PCDHGA1;PCDHGA3*	-0.503
cg24797471	*IER2*	-0.502
cg19455965		0.503
cg05129489	*ADAMTS17*	0.507
cg04331813		0.508
cg18794882	*IAH1;CPSF3*	0.522
cg14032070		0.525
cg10227678	*FTO*	0.528
cg18542495	*THADA*	0.532
cg12073436		0.542
cg09073443		0.546

Probes	TW5	Methylation difference
cg16517593	*MUC6*	0.562
cg07434008	*RPTOR*	0.564
cg23257697	*FAM176A*	0.589
cg21056036	*TUBA3C*	0.593
cg01564165		0.604
cg05317396	*RAMP1*	0.615
cg20288836		0.625
cg27103603	*MOBP*	0.635
cg19920161	*ZNF880*	0.647
cg02979703		0.657
cg01295034	*N4BP1*	0.659
cg21919903	*GPR109A*	0.663
cg19141132	*ARPC2*	0.674
cg13879655		0.690
cg23731742	*NOC2L;SAMD11*	0.700
cg01335658		0.700
cg07121021	*ANKRD30A*	0.705
cg10431950	*GALK1*	0.707
cg20897046	*WDR20*	0.711
cg01410316	*ADAP1*	0.713
cg22147968	*CUX2*	0.714
cg03465061	*BAIAP3*	0.716
cg06187080	*SIPA1L2*	0.722
cg14380230	*PRDM16*	0.724
cg19265040	*MED15*	0.732
cg06073139	*SLC38A7*	0.732
cg04213384	*FARP1*	0.739
cg08516217		0.742
cg13797673	*PASK*	0.751
cg26300322	*CYB561*	0.757
cg01949684	*AGPAT1*	0.784
cg24084481	*SND1;LRRC4*	0.788
cg08254315		0.801
cg19727381		0.802
cg05880944	*DTYMK*	0.813
cg14341177	*BICD2*	0.824
cg16740067	*UBTF*	0.824
cg11777917	*FGFR3*	0.836
cg14194367	*KCNMA1*	0.845
cg04962756	*RELA*	0.870
cg11444668	*NRBP1*	-0.698
cg00108617		-0.685
cg10800464		-0.682

Probes	TW6	Methylation difference
cg14341177	*BICD2*	-0.796
cg19611616	*STK38L*	-0.745
cg13390800	*DCHS1*	0.510

Probes	TW7	Methylation difference
cg12394800	*ATP11A*	-0.865
cg11534938	*TUBGCP2*	-0.848
cg01792081	*NLE1*	-0.844
cg22367886	*TRMT1*	-0.812
cg15279616	*FRMD1*	-0.802
cg08694923		-0.794
cg01928145	*SMPDL3B*	-0.792
cg07791834	*CCDC114*	-0.791
cg05446615	*ZFPM1*	-0.776
cg24118773	*ERGIC1*	-0.762
cg16261619	*ZPBP*	-0.761
cg07053964	*PIGQ*	-0.758
cg16487292	*BTNL2*	-0.755
cg07491809		-0.755
cg25000021		-0.748
cg22749859		-0.744
cg25006249	*C20orf96*	-0.728
cg02878486	*SART1*	-0.707
cg01413566	*C1orf210*	-0.704
cg00435063	*PTPRN2*	-0.703
cg25934700	*SELO*	-0.701
cg10057528	*C1orf86;PRKCZ*	-0.701
cg07880384	*C2;CFB*	-0.693
cg17157205		-0.691
cg27657532		-0.687
cg12291586		-0.667
cg27450153	*GLI3*	-0.620
cg23794200		-0.597
cg10620911	*AGPAT1*	-0.594
cg17445155	*PPEF2*	-0.590
cg09549073	*HOXA5*	-0.585
cg07095530	*ZNHIT1;PLOD3*	-0.553
cg12277416		-0.536
cg04638468	*TMEM155;LOC100192379*	0.525
cg05455825	*DOM3Z;STK19*	0.538
cg03374729		0.564
cg21209219		0.572
cg08194989	*ANK1*	0.573
cg15525341	*DUSP15*	0.577
cg23106158	*RPL36AL;MGAT2*	0.594
cg17054708	*FBLN2*	0.602
cg12822975	*NUSAP1;OIP5*	0.603
cg01558909	*HBM*	0.603
cg17470061	*FNDC5*	0.608
cg22389375	*KCNQ1DN*	0.609
cg19708803		0.637

Probes	TW7	Methylation difference
cg07304536		0.641
cg04594242		0.647
cg21844316		0.647
cg05441651	*ABCG1*	0.648
cg01303723	*NKX6-2*	0.655
cg04861640	*ZNF187*	0.656
cg15165122	*ANKRD53*	0.656
cg15970600	*KIAA0913*	0.658
cg18961681		0.665
cg01623627	*ERBB3*	0.667
cg07710266		0.675
cg23936900	*TOP1*	0.676
cg13872831	*RASSF1*	0.682
cg06537257	*KBTBD7*	0.690
cg20822858	*SGK1*	0.691
cg05044134	*VWA3B*	0.695
cg26747301	*C19orf63;FAM71E1*	0.698
cg20569593		0.701
cg18724430		0.701
cg18722557	*NSMCE4A*	0.705
cg14341177	*BICD2*	0.705
cg06716807	*PPM1G*	0.707
cg24766821		0.711
cg18426060	*AKT1*	0.712
cg14787946	*CEP70*	0.713
cg05053440		0.715
cg11642829	*PELP1*	0.716
cg10884341	*HSPE1;HSPD1*	0.721
cg01077846	*GDF15*	0.728
cg05625559		0.729
cg09596116	*MIR148A*	0.730
cg10420952	*TWIST1*	0.736
cg02089380	*ACVR1*	0.738
cg00547758	*MSH2*	0.748
cg07204937	*GOLGA7B*	0.749
cg18344652	*CNN3*	0.750
cg23627301		0.751
cg24824366		0.753
cg17288102	*XPNPEP1*	0.756
cg01953134	*LYRM2*	0.758
cg19359627	*NEXN*	0.760
cg09342567	*KCNMB3*	0.769
cg02843503	*C7orf28B*	0.769
cg17940673	*RTCD1*	0.780
cg10239074	*BCL2L11*	0.784
cg26228569	*MED27*	0.784
cg07083806	*MYCN;MYCNOS*	0.794
cg00470766	*PRICKLE1*	0.801
cg23082454	*TMEM179*	0.802
cg23675401		0.804
cg24322333	*AKAP12*	0.806
cg27088072		0.818
cg17013193	*BRD2*	0.823
cg01654446	*TAPBP*	0.860

Probes	TW8	Methylation difference
cg22554796	*KIF5B*	-0.913
cg19547293	*WWTR1*	-0.773
cg13682223	*HYAL3;NAT6*	-0.746
cg21036194	*SNCAIP*	-0.713
cg16944964		-0.596
cg05614657	*ZMAT4*	-0.575
cg14984943	*CACNA2D3*	0.671
cg19257200	*SOX10*	0.833

Probes	TW9	Methylation difference
cg24702040	*KIAA1804*	-0.571
cg08905415	*RPTOR*	0.797

Probes	TW10	Methylation difference
cg13054321	*JAKMIP1*	-0.695
cg08779777	*PIK3CG*	-0.678
cg00287711	*TMEM126B;DLG2*	-0.630
cg01947415	*UNC84A*	-0.614
cg11682724		-0.599
cg07860043	*LMTK3*	-0.594
cg18419271		-0.591
cg06421774		-0.572
cg01620569	*AK2*	-0.566
cg03026306	*CABYR*	-0.565
cg27344791	*ZNF433*	-0.562
cg05988645	*C16orf68*	-0.556
cg03623835	*CLDN1*	-0.555
cg05705813	*ME3*	-0.554
cg12755421	*PRKCDBP*	-0.543
cg18596621	*PITPNC1*	-0.541
cg19350469	*BST1*	-0.539
cg14838248		-0.538
cg10287786	*DSCAML1*	-0.536
cg11513637	*CCNA1*	-0.532
cg03485672		-0.530
cg08069370	*SNX1*	-0.527
cg14167415		-0.525
cg16528585	*FNTB*	-0.524
cg14561904	*PTGR1*	-0.523
cg00809863	*KIAA1026*	-0.518
cg09626521		-0.513
cg04872610	*AP4B1;DCLRE1B*	-0.511
cg18222544		-0.507
cg15504662	*LIMK2*	-0.507
cg15786280	*CD38*	-0.505
cg25740457	*CCKBR*	-0.502

Probes	TW10	Methylation difference
cg21120436	*NLGN1*	-0.501
cg08721802		-0.501
cg02157083	*APOA5*	0.500
cg17782954	*ECHDC3*	0.502
cg18187675	*NFIA*	0.505
cg16377162	*EMID1*	0.505
cg03098721	*TTLL7*	0.507
cg06777815		0.507
cg00341060		0.508
cg11964338	*ZNF610*	0.508
cg10008571	*CANT1*	0.510
cg08134856	*STMN1*	0.511
cg22011370	*TSSK1B;MCC*	0.512
cg04744713	*FLJ43860*	0.513
cg14126601	*EIF2AK2*	0.514
cg03444122	*TRIML1*	0.514
cg22753482	*JAKMIP3*	0.515
cg00792015	*DIRC3*	0.515
cg14035807	*IQCA1*	0.516
cg11252310	*KIRREL3*	0.516
cg25160769	*HRASLS5*	0.516
cg25975369	*GRB10*	0.516
cg20737388	*DNAJB13*	0.517
cg07719096		0.521
cg05253165	*NECAB2*	0.521
cg04114269	*C3orf26;FILIP1L;MIR548G*	0.523
cg19179099		0.524
cg07545925	*CD3G*	0.525
cg03728799		0.525
cg08923114		0.527
cg14487271	*WWP1*	0.527

cg21781988		0.528
cg18702715		0.531
cg01313387	*DTX1*	0.534
cg01561869	*PCDHGA2;PCDHGA3;PCDHGA1*	0.535
cg08506869	*KIAA0146*	0.537
cg02288964	*AGRN*	0.540
cg16397563		0.545
cg06638966	*COL19A1*	0.546
cg25757140		0.548
cg15579883		0.552
cg15345393		0.552
cg02160611		0.553
cg16706324	*TRIM78P;TRIM5*	0.556
cg25323444	*MAD1L1*	0.559
cg18639238		0.560
cg09334776		0.568
cg01721313		0.578
cg01309395	*HLA-DPB2*	0.585
cg02774282	*CCDC162*	0.592
cg04414509	*MECOM*	0.596
cg12375380	*EFHB*	0.606
cg03441844		0.607
cg23673360	*XKR6*	0.614
cg17919004		0.614
cg08985817		0.615
cg14263924	*RXRB*	0.620
cg13159361		0.620
cg07493834		0.630
cg12731010	*TNS1*	0.635
cg13800349	*B3GNTL1*	0.638
cg10909185	*KLHL35*	0.723
cg08466517	*TNIP3*	0.798
cg13801402	*BCL2L15*	0.801

Probes	**TW11**	**Methylation difference**
cg05766802		-0.660
cg22620746	*ZNF257*	-0.567
cg24076502		-0.502
cg16595484	*HSPBAP1*	0.514
cg01309395	*HLA-DPB2*	0.693

Probes	**TW12**	**Methylation difference**
cg12728517	*ZNF215*	-0.569
cg06948947	*TBL2*	-0.520
cg14189571	*ZFP42*	-0.508
cg12594237	*CCND2*	0.547
cg08116711	*CUX1*	0.558

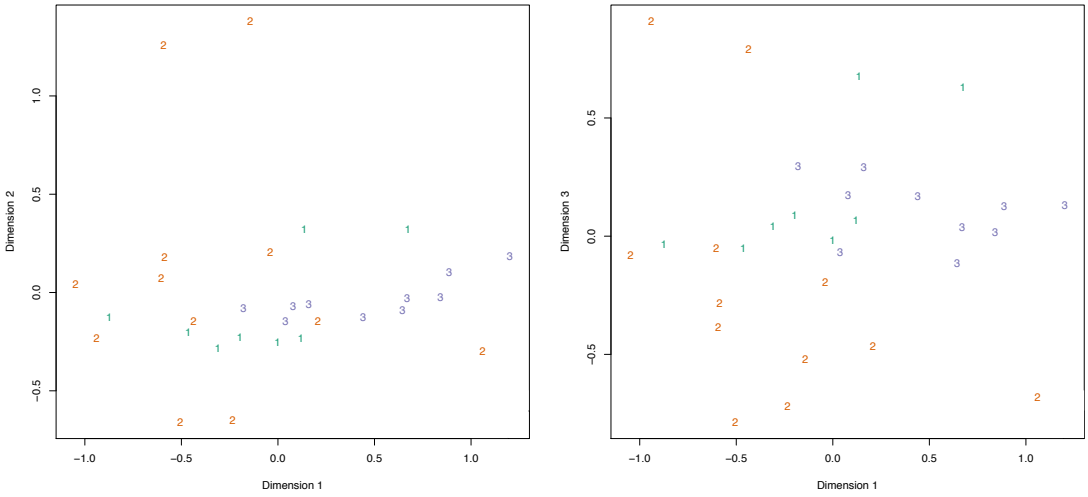
Probes	TW13	Methylation difference
cg12940965		-0.455
cg14131038	*NCK2*	-0.421
cg19023589	*TMCO3*	-0.419
cg08446121	*ANKRD20B*	-0.417
cg01771737	*HDAC4*	-0.413
cg20070090	*S100A8*	-0.410
cg05390077		-0.409
cg26767986	*FAM13C*	-0.404
cg01115243	*VPS4A*	0.416
cg02889672	*SSC5D*	0.418
cg22173752	*TMED7-TICAM2;TICAM2*	0.424
cg01554060	*TMED7-TICAM2;TICAM2*	0.428

Probes	TW14	Methylation difference
cg08905415	*RPTOR*	-0.748
cg08283882	*EBF2*	0.528
ch.3.343413R	*NR2C2*	0.584
cg26287080	*EXOC7*	0.752
cg13803559	*COL5A1*	0.808
cg13682223	*HYAL3;NAT6*	0.872

Probes	TW15	Methylation difference
cg26228569	*MED27*	-0.749
cg19350270	*DPP4*	-0.587
cg25406755	*TFIP11*	0.514
cg17879189	*PRDM13*	0.521
cg04215800		0.536
cg26212180	*VAX1*	0.546
cg19174643		0.551

Probes	TW16	Methylation difference
cg19401033		-0.635
cg06779009		-0.575
cg08779777	*PIK3CG*	-0.505
cg10605064		0.505
cg11065271	*SYTL3*	0.524
cg02943420	*ZFYVE21*	0.529
cg23599843	*JPH1*	0.615
cg24761195	*LRP11*	0.811

Appendix 11: MDS plots for pre-processed data. Samples are colored based on chip location ranging from 1 to 3. The figure represents similarities between samples' 1000 most variable probes based on euclidean distance (sum of squared differences). Dimension 1 represents the largest variation in the dataset and 2 and 3 are the second and third largest respectively



Supplementary Material for Chapter 4

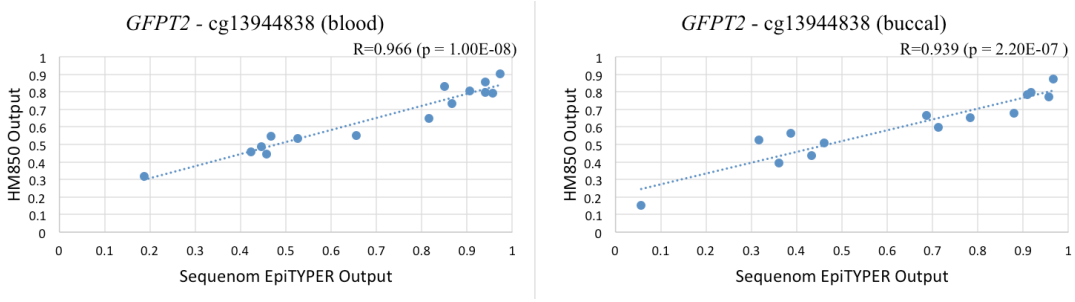
Supplementary material for:

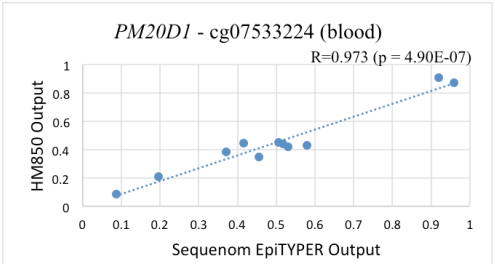
Evidence for type-specific DNA methylation patterns in epilepsy: a discordant monozygotic twin approach

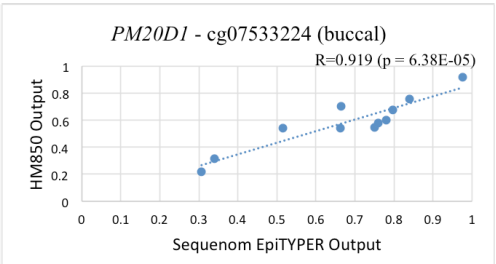
Namitha Mohandas, Yuk Jing Loke, Stephanie Hopkins, Lisa Mackenzie, Carmen Bennett, Samuel F Berkovic, Lata Vadlamudi, Jeffrey M Craig

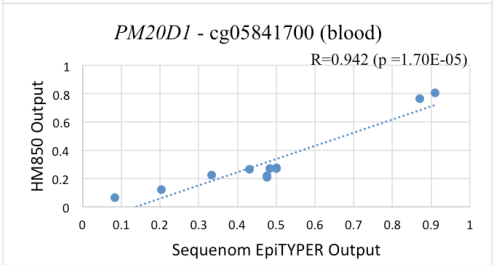
Appendix 12: Cross-platform validation of the top DMRs, GFPT2, PM20D1, ARID5B, OTX1, TTC39C and DLX5 between HM450 and EpiTYPER platforms. Pearson's correlation coefficients for each probe are shown. The scale of both axes reflects a methylation value between 0 and 1 (β) in Figure 12a and difference in methylation values (affected twin minus the unaffected twin) in Figure 12b. Based on the r-value (correlation coefficient), correlations across both platforms are shown. The p-value indicates the significance of the correlation.

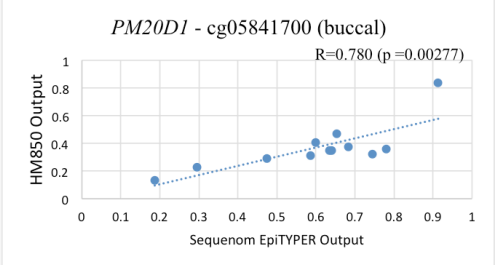
Figure 12a

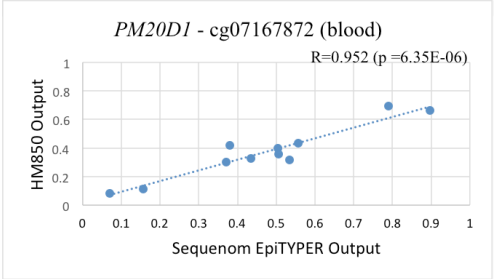


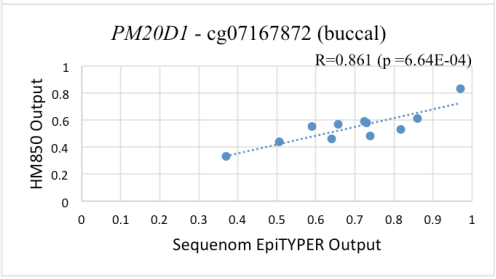


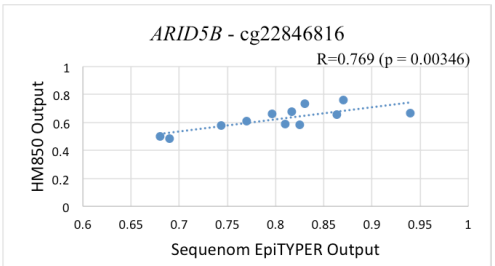


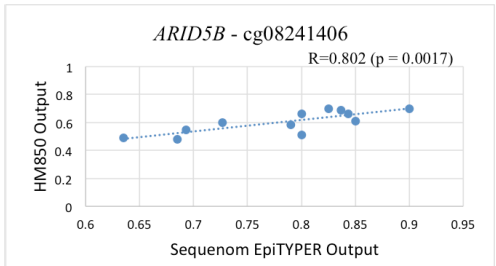


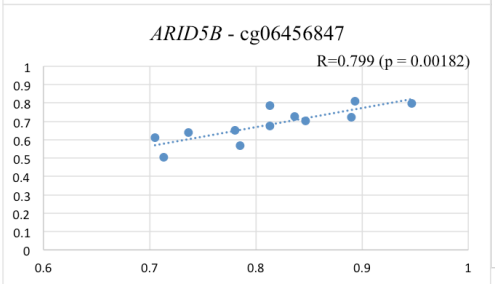


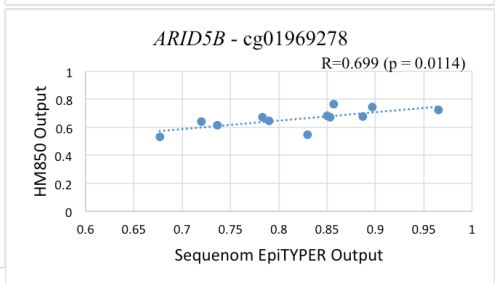




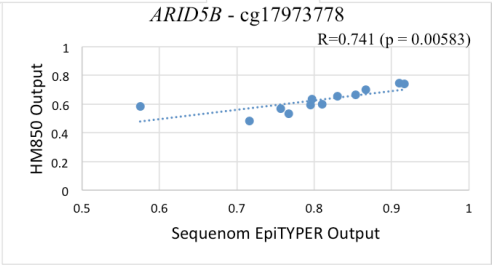


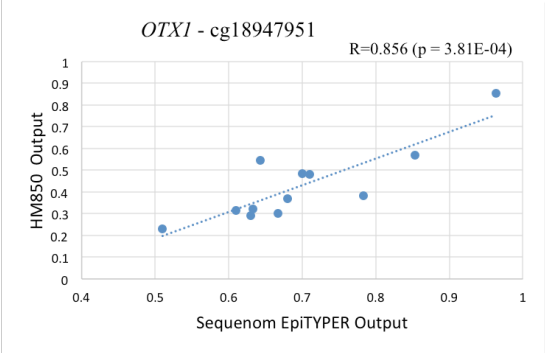


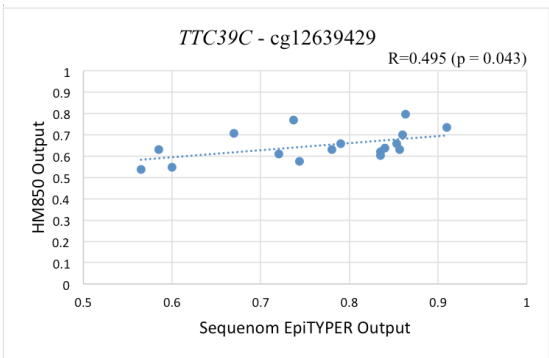


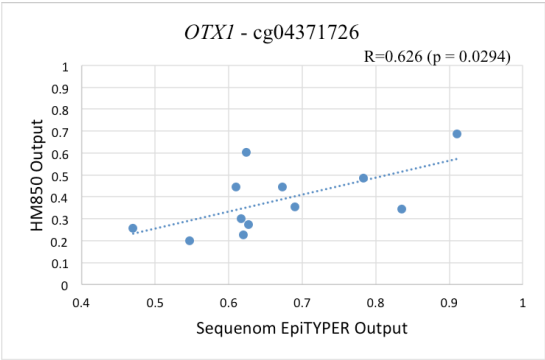


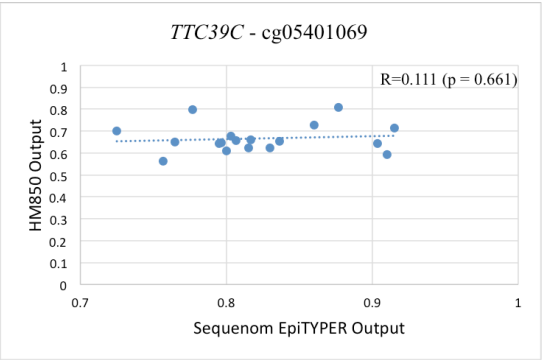
ARID5B beta values





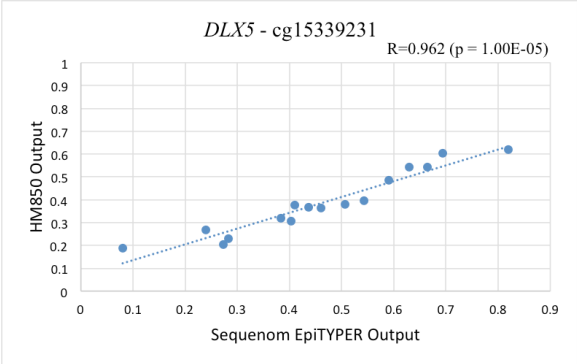






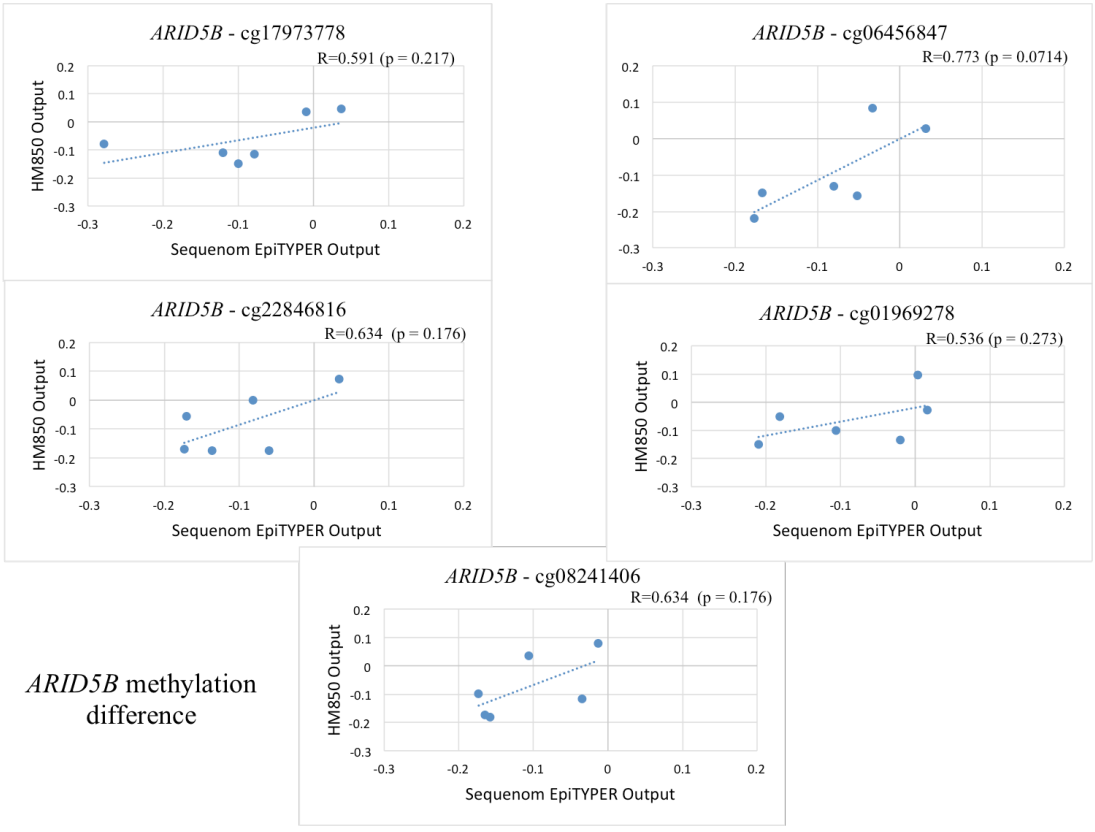
OTX1 beta values

TTC39C beta values

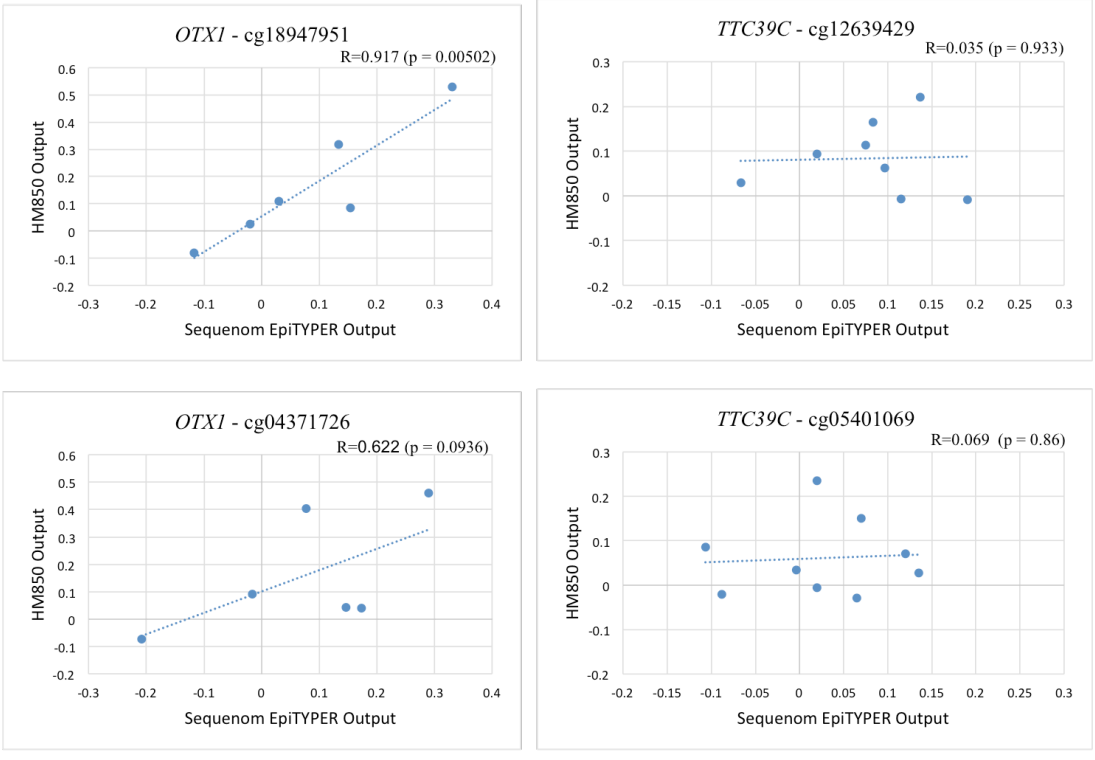


DLX5 beta values

Figure 12b



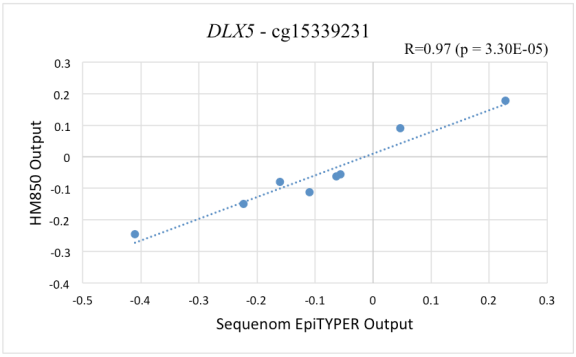
ARID5B methylation difference



OTX1 methylation difference

TTC39C methylation difference

DLX5 methylation difference



Appendix 13. Primer sequences used in site-specific validation using Sequenom EpiTYPER.

Amplicon	Primer Name	Primer Sequence	Amplicon size (base pairs)
GFPT2	Forward	GGTATTGAGAGGGTTTTTGGG	197
	Reverse	AACTCCTCCTACCTTAATACTAACCA	
PM20D1	Forward	GGTAGGGAAAATTAGGAGTAGTATAGTTATTAG	354
	Reverse	AACCAAACATAACTCTAACCTAAAACCC	
ARID5B	Forward	GGGTTAGTTGTGGAAATATTTTTGT	207
	Reverse	AAAAAACTCTACTCTCTCCACCTCC	
OTX1	Forward	GTTTTAGGTTTTGTATTTTTTGGGG	243
	Reverse	AAATATCCCTCTAAACCCTCAACTC	
TTC39C	Forward	GGAGTAATTATGTGTTTTTGATATTTTTG	177
	Reverse	TGTTTTAGGTTTAAGGGTTGTGTTT	
DLX5	Forward	TTAGATTTTTATTTGAGTTGGGGAA	201
	Reverse	CCCTCTAAAACCAAAACTCCATAAT	

Appendix 14. The top ranked DMPs (ranked on unadjusted p-value) in epilepsy affected cases and within discordant twin pairs in buccal tissue.

	Probe ID	Rank	Chromosome position	Gene	Location	P-value	Methylation difference (absolute β value)	
							β GE twin	β FE twin
	cg03810616	1	chr9	SARDH	OpenSea	1.31E-06	0.337	0.620
	cg13113552	2	chr14	EGLN3	S_Shore	1.38E-06	0.035	0.070
	cg05635388	3	chr4	EXOSC9	N_Shore	2.38E-06	0.614	0.806
	cg16345366	4	chr10		OpenSea	2.69E-06	0.353	0.542
	cg25884399	5	chr9	GPSM1	Island	3.04E-06	0.028	0.060
	cg07154356	6	chr19		S_Shore	3.45E-06	0.602	0.836
	cg03004286	7	chr7		OpenSea	7.38E-06	0.754	0.440
	cg21792156	8	chr6	HIST1H2AI	Island	8.97E-06	0.018	0.032
	cg02780120	9	chr2	SF3B1	S_Shore	1.09E-05	0.090	0.179
	cg27285720	10	chr1	GBP4	OpenSea	1.10E-05	0.065	0.139
	cg08316699	11	chr6		N_Shore	1.11E-05		
	cg20946741	12	chr6		OpenSea	1.12E-05		
	cg14765933	13	chr5	SLC6A3	N_Shore	1.12E-05		
	cg06238316	14	chr19	ZNF714	Island	1.21E-05		
	cg14085189	15	chr16	LOC81691	OpenSea	1.52E-05		
	cg22544971	16	chr5		OpenSea	1.56E-05		
	cg06713675	17	chr4	EXOSC9	N_Shore	1.65E-05		
	cg24376776	18	chr10		Island	1.68E-05		
	cg18383160	19	chr14	FAM177A1	Island	1.82E-05		
	cg23997132	20	chr10		OpenSea	1.84E-05		
	cg12672189	21	chr6		OpenSea	1.98E-05		
	cg06202585	22	chr2		OpenSea	1.99E-05		
	cg25488308	23	chr2		OpenSea	2.06E-05		
	cg17436861	24	chr5	DBN1	OpenSea	2.06E-05		
	cg06299328	25	chr8	ESCO2	S_Shore	2.12E-05		
	cg14345882	26	chr6	BTN3A2	OpenSea	2.18E-05		
	cg15362140	27	chr4		S_Shore	2.21E-05		
	cg16361168	28	chr14	THTPA	N_Shore	2.28E-05		
Analysis of focal and generalised epilepsy cases (buccal)	cg05276469	29	chr17		Island	2.32E-05		
	cg10325678	30	chr6	DAAM2	S_Shore	2.53E-05		
	cg09352518	31	chr19	ZNF714	Island	2.68E-05		
	cg09381737	32	chr9	C9orf167	Island	2.74E-05		
	cg21809081	33	chr2	ATL2	N_Shore	2.79E-05		
	cg27218565	34	chr9	PIP5KL1	Island	3.15E-05		
	cg08884395	35	chr6	ESR1	N_Shore	3.32E-05		
	cg09675542	36	chr7		OpenSea	3.38E-05		
	cg20213869	37	chr7	SCIN	OpenSea	3.54E-05		
	cg01483656	38	chr19	ZNF714	N_Shore	3.73E-05		
	cg23230486	39	chr1	GPR3	N_Shore	3.82E-05		
	ch.1.1079286F	40	chr1	S100PBP	OpenSea	3.92E-05		
	cg01046070	41	chr16	MLKL	N_Shelf	3.93E-05		
	cg08135379	42	chr12	AMIGO2	S_Shore	4.76E-05		
	cg22659953	43	chr5	SLC6A3	N_Shelf	4.77E-05		
	cg18624290	44	chr11		OpenSea	4.77E-05		
	cg25176297	45	chr2	HEATR5B	OpenSea	4.92E-05		
	cg19693031	46	chr1	TXNIP	OpenSea	5.11E-05		
	cg10556520	47	chr6		OpenSea	5.40E-05		
	cg14228831	48	chr4	HADH	N_Shore	5.41E-05		
	cg06595320	49	chr3		N_Shore	6.52E-05		
	cg06788267	50	chr1		OpenSea	6.85E-05		
	cg24919342	51	chr3	ABTB1	OpenSea	7.08E-05		
	cg09295738	52	chr11		N_Shore	7.16E-05		
	cg04893899	53	chr21	PTTG1IP	Island	7.58E-05		
	cg12443119	54	chr4		OpenSea	7.58E-05		
	cg26826402	55	chr13	STARD13	OpenSea	7.74E-05		
	cg05991743	56	chr2	CCDC85A	Island	7.92E-05		
	cg06680201	57	chr11	C11orf51	S_Shore	8.73E-05		
	cg04206484	58	chr7	C7orf50	S_Shore	8.96E-05		
	cg01060016	59	chr13		S_Shore	9.11E-05		
	cg17736936	60	chr14		OpenSea	9.28E-05		
	cg16621790	61	chr3	ZMYND10	S_Shore	9.31E-05		
	cg09722212	62	chr2	HEATR5B	OpenSea	9.32E-05		

	Probe ID	Rank	Chromosome position	Gene	Location	P-value	Methylation difference
Within twin pair analysis of generalised epilepsy discordant twin pairs (buccal)	cg02009256	1	chr12	GLS2	Island	3.92E-06	-0.009
	cg17277890	2	chr11	TRIM22	OpenSea	5.22E-06	0.012
	cg20901245	3	chr2	CEBPZOS	Island	1.24E-05	0.063
	ch.5.1443044F	4	chr5	AGGF1	OpenSea	1.29E-05	-0.006
	cg01526911	5	chr5	ZSWIM6	OpenSea	1.43E-05	0.079
	cg23940234	6	chr18	LINC00669	OpenSea	1.58E-05	0.081
	cg01874640	7	chr3	HGD;HGD	OpenSea	1.58E-05	0.070
	cg21931717	8	chr5	SDHAP3	Island	2.00E-05	-0.043
	cg25445017	9	chr15	CTXN2	OpenSea	2.13E-05	-0.074
	cg06901091	10	chr3		OpenSea	3.66E-05	0.096
	cg13677162	11	chr6	EMIS;RP11-73(	OpenSea	3.68E-05	
	cg13818127	12	chr7	SDK1	Island	3.80E-05	
	cg15224432	13	chr19	IGFL3	OpenSea	4.10E-05	
	cg22544971	14	chr5		OpenSea	4.32E-05	
	cg21173044	15	chr13	SACS	Island	4.36E-05	
	cg10553028	16	chr3	TM8;RP11-384I	Island	4.39E-05	
	cg06102690	17	chr4	CCDC149	OpenSea	4.80E-05	
	ch.2.12139298-	18	chr2	GLI2	OpenSea	4.91E-05	
	cg02079693	19	chr6		OpenSea	5.08E-05	
	cg14828574	20	chr10		OpenSea	5.34E-05	
	cg02541746	21	chr5	GHR	OpenSea	5.36E-05	
	cg03647607	22	chr9	IL33	OpenSea	5.39E-05	
	cg00483994	23	chr2	SLC16A14	N_Shelf	5.72E-05	
	cg23280614	24	chr4	ZCCHC4	OpenSea	5.77E-05	
	cg11350048	25	chr2	AC096559.1	OpenSea	6.38E-05	
	cg20696051	26	chr7	WDR60	Island	6.39E-05	
	cg13772158	27	chr2		OpenSea	6.80E-05	
	cg05855741	28	chr9	ACTL7A	OpenSea	6.98E-05	
	cg07993132	29	chr2	ZC3H15	S_Shore	7.02E-05	
	cg17327394	30	chr14	EXOC5	OpenSea	7.73E-05	
	cg15017411	31	chr7	CCDC71L	S_Shore	7.90E-05	
	cg20995564	32	chr2		OpenSea	7.91E-05	
	cg19558654	33	chr22	C22orf45;UPB1	OpenSea	8.25E-05	
	cg18085807	34	chr6	RPP21;TRIM39	Island	8.45E-05	
	cg09666237	35	chr17	LLGL1	Island	8.95E-05	

	Probe ID	Rank	Chromosome position	Gene	Location	P-value	Methylation difference
Within twin pair analysis of focal epilepsy discordant twin pairs (buccal)	cg01665058	1	chr16	PAPD5	OpenSea	4.22E-06	-0.082
	cg24822446	2	chr6	ZFAND3	Island	1.01E-05	0.030
	cg21114148	3	chr12	CEP290	OpenSea	1.02E-05	-0.059
	cg08597698	4	chr4	CBR4	Island	1.08E-05	0.011
	cg23205521	5	chr1	LRP8	Island	1.22E-05	-0.104
	cg16588007	6	chr3	PTPLB	N_Shelf	1.40E-05	-0.090
	cg12510708	7	chr7	NFE2L3	S_Shore	2.49E-05	0.022
	cg21122684	8	chr6	PDSS2	Island	2.55E-05	0.014
	cg08277382	9	chr16	TMCO7	N_Shore	2.59E-05	0.021
	cg13927251	10	chr5	FLT4	S_Shore	3.08E-05	0.022
	cg23455897	11	chr20	KIF16B	Island	3.94E-05	
	cg06861370	12	chr1	EPB41	OpenSea	4.00E-05	
	cg00950418	13	chr7	SRPK2	Island	4.13E-05	
	cg22834211	14	chr9	PTPRD-AS1	S_Shelf	4.42E-05	
	cg26421448	15	chr1	WDR26	N_Shore	4.45E-05	
	cg02348711	16	chr17	ETV4	Island	4.46E-05	
	cg08985812	17	chr1	SLC45A3	Island	4.54E-05	
	cg16704958	18	chr21	DSCR3	S_Shore	4.64E-05	
	cg12266024	19	chr7	FAM126A	Island	4.72E-05	
	cg05487589	20	chr8	EBF2	N_Shore	5.15E-05	
	cg00364778	21	chr11	TRIM68	Island	5.16E-05	
	cg08135379	22	chr12	AMIGO2	S_Shore	5.19E-05	
	cg15865175	23	chr13	BRCA2	S_Shore	5.38E-05	
	cg25244476	24	chr11	ETS1	N_Shore	5.63E-05	
	cg10013356	25	chr6	RAET1G	S_Shore	6.34E-05	
	cg26884557	26	chr11	C105376671;LC	N_Shore	6.48E-05	
	cg06249417	27	chr11	PAFAH1B2	OpenSea	6.49E-05	
	cg25882830	28	chr4	STOX2	Island	6.86E-05	
	cg04821708	29	chr16	SDR42E1	Island	7.02E-05	
	cg00786406	30	chr1	HIPK1	S_Shore	7.23E-05	
	cg03915740	31	chr8	PVT1	OpenSea	7.32E-05	
	cg06218165	32	chr3		OpenSea	7.56E-05	
	cg00107819	33	chr6	FOXP4-AS1	Island	7.64E-05	
	cg08178940	34	chr6	TFAP2B	N_Shelf	7.89E-05	
	cg13371826	35	chr15	SEMA6D	OpenSea	8.67E-05	
	cg21649442	36	chr16	SDR42E1	Island	8.82E-05	
	cg01345395	37	chr6		OpenSea	9.73E-05	

Appendix 15. The top ranked DMPs (ranked on unadjusted p-value) in epilepsy affected cases and within discordant twin pairs in blood tissue.

	Probe ID	Rank	Chromosome position	Gene	Location	P-value	Methylation difference (%)	
							β GE twin	β FE twin
	cg20983004	1	chr20		OpenSea	1.12E-06	0.036	0.108
	cg09747182	2	chr16	XYLT1	OpenSea	3.46E-06	0.020	0.034
	cg17408849	3	chr14	KCNH5	OpenSea	4.06E-06	0.492	0.639
	cg25777896	4	chr13	NC01232;TM9S	N_Shore	5.07E-06	0.209	0.317
	cg15720089	5	chr10	CACNB2	OpenSea	6.22E-06	0.715	0.862
	cg18978275	6	chr1	PLD5	OpenSea	6.38E-06	0.507	0.676
	cg00834796	7	chr5	JAKMIP2	OpenSea	6.48E-06	0.022	0.041
	cg27128734	8	chr19	KLK4	Island	6.67E-06	0.615	0.766
	cg04367427	9	chr2		N_Shore	6.98E-06	0.592	0.712
	cg25569396	10	chr6	NEDD9	OpenSea	7.58E-06	0.023	0.043
	cg19494249	11	chr10	SLC39A12	OpenSea	8.34E-06		
	cg19066391	12	chr3	CCDC71	S_Shore	9.05E-06		
	cg03267342	13	chr9	KLF4	Island	9.08E-06		
	cg18009000	14	chr10	H2AFY2	N_Shore	9.19E-06		
	cg14507193	15	chr5	PLK2	Island	9.22E-06		
	cg16614020	16	chr5	SLC6A3	N_Shore	9.63E-06		
	cg16725535	17	chr7	SDK1	OpenSea	9.67E-06		
	cg04880105	18	chr4		OpenSea	1.03E-05		
	cg17033723	19	chr5		OpenSea	1.24E-05		
	cg00435063	20	chr7	PTPRN2	Island	1.25E-05		
	cg08421207	21	chr7		OpenSea	1.39E-05		
	cg09921810	22	chr16	TRIM72;PYDC1	Island	1.68E-05		
	cg16705929	23	chr2	LRRFIP1	S_Shelf	1.84E-05		
	cg16466397	24	chr19		OpenSea	1.87E-05		
	cg19774368	25	chr2	C1QL2	S_Shore	1.94E-05		
	ch.12.78192401	26	chr12		OpenSea	1.97E-05		
	cg20212912	27	chr4		N_Shore	2.00E-05		
	cg03103158	28	chr8	NTS9;HMBOX	Island	2.07E-05		
	cg27113326	29	chr7	PTPRN2	Island	2.17E-05		
	ch.5.92314382I	30	chr5		OpenSea	2.36E-05		
	cg07343187	31	chr15	ACAN	OpenSea	2.49E-05		
Analysis of focal and generalised epilepsy cases (blood)	cg00880741	32	chr2		N_Shore	2.50E-05		
	cg05499111	33	chr15		OpenSea	2.83E-05		
	cg05411428	34	chr2	AFTPH	N_Shore	2.87E-05		
	cg00304388	35	chr1	CACNA1E	N_Shore	3.10E-05		
	cg06959340	36	chr14	JUB	Island	3.22E-05		
	cg13981838	37	chr15	MFAP1	N_Shore	3.25E-05		
	cg13665102	38	chr11		OpenSea	3.34E-05		
	cg02512920	39	chr10		N_Shore	3.37E-05		
	cg10577630	40	chr12		OpenSea	3.41E-05		
	cg26191704	41	chr5	RARS	N_Shore	3.44E-05		
	cg02423659	42	chr4	GAK	Island	3.51E-05		
	cg13897914	43	chr15	NDUFAF1	S_Shore	3.75E-05		
	cg20395040	44	chr13		OpenSea	3.75E-05		
	cg01573067	45	chr6		S_Shelf	3.75E-05		
	cg00329014	46	chr3	CDC51;CCDC7	N_Shore	3.80E-05		
	cg08716018	47	chr17	NAJC7;NKIRA1	S_Shore	3.87E-05		
	cg21479898	48	chr20		OpenSea	3.94E-05		
	cg24989739	49	chr6	POPDC3	S_Shore	4.02E-05		
	cg24031967	50	chr4	TLL1	N_Shore	4.08E-05		
	cg11203771	51	chr9	HSDL2	N_Shore	4.13E-05		
	cg13106758	52	chr15		OpenSea	4.34E-05		
	cg11829872	53	chr14	STRN3;AP4S1	N_Shore	4.36E-05		
	cg01011764	54	chr22	MAPK11	Island	4.54E-05		
	cg01323954	55	chr15	FAM103A1	N_Shore	4.59E-05		
	cg01949798	56	chr8	NIPAL2	S_Shore	4.65E-05		
	cg23695504	57	chr1	C1orf229	Island	4.81E-05		
	cg10758037	58	chr6	EHMT2;C2	S_Shore	4.88E-05		
	cg26040903	59	chr10	HSPA12A	S_Shore	4.88E-05		
	cg17994536	60	chr19	RDH13	S_Shore	4.90E-05		
	cg23706422	61	chr10	PAPSS2	Island	4.93E-05		
	cg05967710	62	chr12	ANKS1B	N_Shelf	5.11E-05		
	cg08381997	63	chr6	6orf120;WDR2	N_Shore	5.15E-05		
	cg02187874	64	chr7		OpenSea	5.28E-05		
	cg20529969	65	chr8	GLI4	Island	5.34E-05		
	cg17446142	66	chr5	GDF9;UQCRQ	N_Shore	5.35E-05		
	cg24313364	67	chr2	METTL21A	Island	5.39E-05		
	cg07325168	68	chr7	VPS41	OpenSea	5.39E-05		

	Probe ID	Rank	Chromosome position	Gene	Location	P-value
	cg01784406	69	chr5	DDX46	OpenSea	5.49E-05
	cg10576150	70	chr18		OpenSea	5.58E-05
	cg07061913	71	chr10	DNAJC9	N_Shore	5.65E-05
	cg13291208	72	chr8	SULF1	OpenSea	5.70E-05
	cg21051232	73	chr18	NETO1	N_Shore	5.72E-05
	cg17220002	74	chr7	LUC7L2	Island	5.92E-05
	cg02617165	75	chr8		OpenSea	5.93E-05
	cg07408552	76	chr17	ATP6V0A1	S_Shore	5.94E-05
	cg06249486	77	chr16	BBS2	N_Shore	6.02E-05
	cg04516975	78	chr7	FAM133B	Island	6.03E-05
	cg16497018	79	chr10	C10orf75	Island	6.47E-05
	cg07792427	80	chr1	CDC14A	OpenSea	6.57E-05
	cg08905567	81	chr14		OpenSea	6.69E-05
	cg14118175	82	chr13	GPC6	OpenSea	6.73E-05
	cg03683256	83	chr5	ARAP3	OpenSea	6.80E-05
	cg00362285	84	chr1	ERO1LB	N_Shore	6.89E-05
	cg08440934	85	chr5	UBLCP1	N_Shore	6.93E-05
	cg15004182	86	chr20	DEFB116	OpenSea	6.97E-05
	cg14683738	87	chr19	ZNF585B	OpenSea	7.04E-05
	cg11904686	88	chr3	ITIH1	OpenSea	7.06E-05
	cg13450576	89	chr11	GAB2	OpenSea	7.20E-05
	cg25215292	90	chr5	FBXL21	N_Shore	7.22E-05
	cg10052687	91	chr7	CLIP2	N_Shelf	7.26E-05
	cg11976052	92	chr19	KLK4	Island	7.28E-05
Analysis of focal and generalised epilepsy cases (blood)	cg05766510	93	chr7	PTPRN2	Island	7.30E-05
	cg17396989	94	chr4	WHSC2	Island	7.31E-05
	cg00046156	95	chr5		N_Shelf	7.32E-05
	cg10548981	96	chr21		OpenSea	7.34E-05
	cg21826699	97	chr12		OpenSea	7.35E-05
	cg07892785	98	chr3	MYRIP	Island	7.35E-05
	cg09409311	99	chr19	CLPTM1	N_Shore	7.43E-05
	cg16539111	100	chr4	CPE	OpenSea	7.47E-05
	cg02631286	101	chr1	SPOCD1	OpenSea	7.57E-05
	cg15667697	102	chr21	LOC101928233	N_Shore	7.58E-05
	cg16171534	103	chr3	CCDC71	S_Shore	7.58E-05
	cg17310600	104	chr4	36951	S_Shore	7.62E-05
	cg08717880	105	chr2	IOXD4;MIR10I	N_Shore	7.64E-05
	cg20098074	106	chr3	TPRA1	Island	7.67E-05
	cg18785816	107	chr6		OpenSea	7.74E-05
	cg26765567	108	chr11		N_Shore	7.76E-05
	cg09246754	109	chr10	P4HA1	N_Shore	7.82E-05
	cg26351104	110	chr1	TTC39A	Island	7.86E-05
	cg01899029	111	chr15	AGBL1	OpenSea	7.96E-05
	cg10043073	112	chr2	101927907;LRR	OpenSea	8.10E-05
	cg13475583	113	chr1		OpenSea	8.19E-05
	cg14476013	114	chr12	NDUFA9;AKAP	N_Shore	8.23E-05
	cg22326754	115	chr16	ATF7IP2	N_Shore	8.28E-05
	cg03188118	116	chr14	RGS6	S_Shore	8.29E-05
	cg25810036	117	chr17	ITGB3	OpenSea	8.31E-05
	cg06587257	118	chr12	ACCN2	S_Shore	8.37E-05
	cg09138368	119	chr14	ACOT4	S_Shore	8.37E-05
	cg06741896	120	chr11	CCND1	N_Shore	8.38E-05

	Probe ID	Rank	Chromosome position	Gene	Location	P-value
Analysis of focal and generalised epilepsy cases (blood)	cg06916446	121	chr5		N_Shore	8.39E-05
	cg06461373	122	chr11		OpenSea	8.39E-05
	cg07739803	123	chr1	LC30A7;EXTL	N_Shore	8.42E-05
	cg01070987	124	chr3	PFN2	N_Shore	8.43E-05
	cg24065982	125	chr14		OpenSea	8.48E-05
	cg02366953	126	chr16		OpenSea	8.53E-05
	cg04434244	127	chr6		OpenSea	8.61E-05
	cg15726514	128	chr2	C2orf43	S_Shore	8.75E-05
	cg15628956	129	chr12	TENC1	Island	8.78E-05
	cg13726459	130	chr12	C4;HOXC6;HC	S_Shore	8.85E-05
	cg11089595	131	chr8	RALYL	N_Shelf	8.98E-05
	cg06312469	132	chr10	CNNM1	N_Shore	8.98E-05
	cg25290227	133	chr10	GPAM	S_Shore	8.99E-05
	cg08940669	134	chr16	SPNS1	Island	9.04E-05
	cg15139153	135	chr6	RNGTT	OpenSea	9.04E-05
	cg14563196	136	chr1	GBP4	OpenSea	9.15E-05
	cg06189459	137	chr11	OR10AG1	OpenSea	9.42E-05
	cg19683913	138	chr2		OpenSea	9.46E-05
	cg03004286	139	chr7		OpenSea	9.50E-05
	cg15375469	140	chr7	CPVL	OpenSea	9.57E-05
	cg23367857	141	chr7	CALD1	OpenSea	9.63E-05
	cg11865119	142	chr7	MEST	N_Shore	9.72E-05
	cg13388466	143	chr17		Island	9.72E-05
	cg17121140	144	chr13		N_Shore	9.85E-05
	cg13577188	145	chr1	MED18	S_Shore	9.98E-05

	Probe ID	Rank	Chromosome position	Gene	Location	P-value	Methylation difference (%)
Within twin pair analysis of generalised epilepsy discordant twin pairs (blood)	cg06488988	1	chr11	JRKL;CCDC82	Island	2.85E-05	-0.014
	cg12732284	2	chr3	PLD1	OpenSea	3.00E-05	0.076
	cg24008825	3	chr15	ADAM10	OpenSea	3.67E-05	-0.125
	cg23555866	4	chr2	ARHGAP15	OpenSea	5.50E-05	-0.086
	cg12429188	5	chr15	BBS4	OpenSea	7.66E-05	0.108
	cg13595528	6	chr12	R3HDM2	OpenSea	7.81E-05	0.095
	cg02832414	7	chr17	CALCOCO2	OpenSea	9.49E-05	-0.063

	Probe ID	Rank	Chromosome position	Gene	Location	P-value	Methylation difference (%)	
Within twin pair analysis of focal epilepsy discordant twin pairs (blood)	cg17977435	1	chr6		OpenSea	3.56E-06	-0.044	-0.044
	cg02800252	2	chr3	SLMAP	OpenSea	8.2982E-06	-0.128	
	ch.10.2414746R	3	chr10	TCF7L2	OpenSea	9.2092E-06	0.018	
	cg22148904	4	chr1	PER3	OpenSea	2.04E-05	-0.101	-0.101
	cg23664553	5	chr2	ARHGEF4	OpenSea	2.71E-05	0.055	0.055
	cg20363989	6	chr15	fIR548H4;GLC	Island	2.96E-05	0.012	0.012
	cg00167883	7	chr14		OpenSea	3.68E-05	-0.060	-0.060
	cg10161352	8	chr5	ATP6AP1L	OpenSea	3.83E-05	-0.058	-0.058
	cg13285686	9	chr14	FUT8	OpenSea	4.20E-05	-0.091	-0.091
	cg23623214	10	chr18		S_Shelf	4.50E-05	-0.082	-0.082
	cg01773061	11	chr1	LSM10	N_Shore	5.18E-05		
	cg15616358	12	chr11	CPT1A	N_Shore	5.68E-05		
	cg02379352	13	chr2	AFF3	OpenSea	6.19E-05		
	ch.13.10810901	14	chr13	KLF12	OpenSea	6.36E-05		
	cg00719211	15	chr6		Island	6.51E-05		
	cg18135379	16	chr1	OLFML2B	OpenSea	8.00E-05		
	cg20854374	17	chr11		OpenSea	8.23E-05		
	cg04108061	18	chr10		OpenSea	8.65E-05		
	cg20569302	19	chr9	FBXW2	OpenSea	8.92E-05		
	cg07352345	20	chr5	ARL10	Island	9.42E-05		
	cg22826226	21	chr12	SLC11A2	S_Shore	9.64E-05		
	cg10634964	22	chr2	ıNKZF1;ATG9A	Island	9.84E-05		

Appendix 16. Top 20 Gene ontology (GO) terms (BP = biological process) analysed for the 1000 top ranked CP-associated DMPs in buccal.

	GO ID	Term	Ont	N	DE	P.DE	FDR
Focal vs generalised	GO:0006805	xenobiotic metabolic process	BP	91	11	3.01E-05	0.244
	GO:0052697	xenobiotic glucuronidation	BP	9	5	3.20E-05	0.244
	GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	BP	214	22	3.49E-05	0.244
	GO:0071466	cellular response to xenobiotic stimulus	BP	95	11	5.51E-05	0.254
	GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	BP	149	18	6.28E-05	0.254
	GO:0051552	flavone metabolic process	BP	6	4	7.26E-05	0.254
	GO:0009410	response to xenobiotic stimulus	BP	103	11	1.36E-04	0.406
	GO:0052696	flavonoid glucuronidation	BP	19	5	1.82E-04	0.478
	GO:0052695	cellular glucuronidation	BP	21	5	3.78E-04	0.881
	GO:0009812	flavonoid metabolic process	BP	24	5	6.36E-04	1
	GO:0006720	isoprenoid metabolic process	BP	120	11	1.04E-03	1
	GO:0019585	glucuronate metabolic process	BP	26	5	1.05E-03	1
	GO:0006063	uronic acid metabolic process	BP	26	5	1.05E-03	1
	GO:0006575	cellular modified amino acid metabolic process	BP	143	11	1.98E-03	1
	GO:0030178	negative regulation of Wnt signaling pathway	BP	191	16	2.27E-03	1
	GO:0042440	pigment metabolic process	BP	61	7	2.49E-03	1
	GO:0098609	cell-cell adhesion	BP	1091	55	3.22E-03	1
	GO:1904646	cellular response to beta-amyloid	BP	3	2	3.46E-03	1
	GO:1904645	response to beta-amyloid	BP	3	2	3.46E-03	1
	GO:0051436	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	BP	70	7	3.55E-03	1
Within twin pair analysis of generalised epilepsy discordant twin pairs (buccal)	GO:0051171	regulation of nitrogen compound metabolic process	BP	4012	147	1.13E-04	0.755
	GO:0031323	regulation of cellular metabolic process	BP	5331	183	1.54E-04	0.755
	GO:0031326	regulation of cellular biosynthetic process	BP	3897	142	1.62E-04	0.755
	GO:0019222	regulation of metabolic process	BP	5630	191	1.82E-04	0.755
	GO:0009889	regulation of biosynthetic process	BP	3946	143	1.96E-04	0.755
	GO:0010556	regulation of macromolecule biosynthetic process	BP	3726	136	2.51E-04	0.755
	GO:0080090	regulation of primary metabolic process	BP	5290	180	2.84E-04	0.755
	GO:2000112	regulation of cellular macromolecule biosynthetic process	BP	3624	133	2.88E-04	0.755
	GO:0010608	posttranscriptional regulation of gene expression	BP	445	25	8.55E-04	1
	GO:0033030	negative regulation of neutrophil apoptotic process	BP	2	2	9.14E-04	1
	GO:0060255	regulation of macromolecule metabolic process	BP	5309	177	1.11E-03	1
	GO:0046825	regulation of protein export from nucleus	BP	31	5	1.18E-03	1
	GO:0034645	cellular macromolecule biosynthetic process	BP	4535	154	1.50E-03	1
	GO:0051348	negative regulation of transferase activity	BP	331	20	1.63E-03	1
	GO:0042136	neurotransmitter biosynthetic process	BP	15	4	1.78E-03	1
	GO:0044260	cellular macromolecule metabolic process	BP	7619	237	1.78E-03	1
	GO:0033029	regulation of neutrophil apoptotic process	BP	4	2	1.82E-03	1
	GO:0016578	histone deubiquitination	BP	19	4	1.83E-03	1
	GO:0044267	cellular protein metabolic process	BP	4459	149	1.94E-03	1
	GO:0009059	macromolecule biosynthetic process	BP	4685	157	2.06E-03	1
Within twin pair analysis of focal epilepsy discordant twin pairs (buccal)	GO:0007156	homophilic-cell adhesion via plasma membrane adhesion molecule	BP	149	22	1.87E-07	3.92E-03
	GO:0098742	cell-cell adhesion via plasma membrane adhesion molecule	BP	214	24	2.25E-06	2.36E-02
	GO:0006543	glutamine catabolic process	BP	3	2	1.60E-03	1
	GO:0051704	multi-organism process	BP	2074	83	1.80E-03	1
	GO:0034351	negative regulation of glial cell apoptotic process	BP	7	3	2.12E-03	1
	GO:0034350	regulation of glial cell apoptotic process	BP	7	3	2.12E-03	1
	GO:0045184	establishment of protein localization	BP	1871	82	2.14E-03	1
	GO:0031666	positive regulation of lipopolysaccharide-mediated signaling pathway	BP	8	3	2.18E-03	1
	GO:0048643	positive regulation of skeletal muscle tissue development	BP	23	5	2.32E-03	1
	GO:0097178	ruffle assembly	BP	28	5	3.20E-03	1
	GO:0016202	regulation of striated muscle tissue development	BP	108	11	3.53E-03	1
	GO:0007519	skeletal muscle tissue development	BP	144	13	3.55E-03	1
	GO:0048742	regulation of skeletal muscle fiber development	BP	9	3	3.60E-03	1
	GO:0031664	regulation of lipopolysaccharide-mediated signaling pathway	BP	18	4	3.61E-03	1
	GO:0043542	endothelial cell migration	BP	146	13	3.66E-03	1
	GO:0048636	positive regulation of muscle organ development	BP	60	8	3.75E-03	1
	GO:0045844	positive regulation of striated muscle tissue development	BP	60	8	3.75E-03	1
	GO:0006743	ubiquinone metabolic process	BP	13	3	3.87E-03	1
	GO:0048634	regulation of muscle organ development	BP	110	11	3.91E-03	1
	GO:1901863	positive regulation of muscle tissue development	BP	61	8	4.01E-03	1

Appendix 17. Top 20 Gene ontology (GO) terms (BP = biological process) analysed for the 1000 top ranked CP-associated DMPs in blood.

	GO ID	Term	Ont	N	DE	P.DE	FDR
	GO:0007626	locomotory behavior	BP	181	16	5.33E-04	1
	GO:0021554	optic nerve development	BP	12	4	7.15E-04	1
	GO:0051001	negative regulation of nitric-oxide synthase activity	BP	6	3	7.59E-04	1
	GO:1902525	regulation of protein monoubiquitination	BP	3	2	1.01E-03	1
	GO:1903543	positive regulation of exosomal secretion	BP	15	4	1.11E-03	1
	GO:1903541	regulation of exosomal secretion	BP	15	4	1.11E-03	1
	GO:0009186	deoxyribonucleoside diphosphate metabolic process	BP	8	3	1.67E-03	1
	GO:0021631	optic nerve morphogenesis	BP	6	3	1.67E-03	1
Focal vs	GO:1990182	exosomal secretion	BP	18	4	2.19E-03	1
generalised	GO:1901205	negative regulation of adrenergic receptor signaling pathway involved in heart process	BP	2	2	2.23E-03	1
epilepsy	GO:1901204	regulation of adrenergic receptor signaling pathway involved in heart process	BP	2	2	2.23E-03	1
	GO:0018198	peptidyl-cysteine modification	BP	24	4	3.58E-03	1
	GO:0051046	regulation of secretion	BP	624	31	3.89E-03	1
	GO:0015949	nucleobase-containing small molecule interconversion	BP	24	4	4.22E-03	1
	GO:0086103	G-protein coupled receptor signaling pathway involved in heart process	BP	9	3	4.27E-03	1
	GO:0071878	negative regulation of adrenergic receptor signaling pathway	BP	3	2	4.36E-03	1
	GO:0032769	negative regulation of monooxygenase activity	BP	10	3	4.48E-03	1
	GO:0006527	arginine catabolic process	BP	10	3	4.69E-03	1
	GO:0036093	germ cell proliferation	BP	6	2	4.92E-03	1
	GO:0002176	male germ cell proliferation	BP	6	2	4.92E-03	1
	GO:0042523	positive regulation of tyrosine phosphorylation of Stat5 protein	BP	16	4	1.39E-04	1
	GO:0042522	regulation of tyrosine phosphorylation of Stat5 protein	BP	19	4	4.10E-04	1
	GO:0031591	wybutosine biosynthetic process	BP	2	2	6.93E-04	1
	GO:0031590	wybutosine metabolic process	BP	2	2	6.93E-04	1
	GO:0042506	tyrosine phosphorylation of Stat5 protein	BP	22	4	8.15E-04	1
	GO:0072539	T-helper 17 cell differentiation	BP	14	3	1.78E-03	1
Within	GO:0046415	urate metabolic process	BP	12	3	1.85E-03	1
twin pair	GO:1904049	negative regulation of spontaneous neurotransmitter secretion	BP	2	2	1.89E-03	1
analysis of	GO:1904048	regulation of spontaneous neurotransmitter secretion	BP	2	2	1.89E-03	1
generalised	GO:2000319	regulation of T-helper 17 cell differentiation	BP	7	2	2.68E-03	1
epilepsy	GO:0043116	negative regulation of vascular permeability	BP	10	3	2.70E-03	1
discordant	GO:2000346	negative regulation of hepatocyte proliferation	BP	4	2	2.82E-03	1
twin pairs	GO:0072538	T-helper 17 type immune response	BP	16	3	2.89E-03	1
(blood)	GO:0006065	UDP-glucuronate biosynthetic process	BP	3	2	2.94E-03	1
	GO:0090331	negative regulation of platelet aggregation	BP	9	3	3.12E-03	1
	GO:0061737	leukotriene signaling pathway	BP	5	2	3.41E-03	1
	GO:0009240	isopentenyl diphosphate biosynthetic process	BP	3	2	3.62E-03	1
	GO:0019287	isopentenyl diphosphate biosynthetic process, mevalonate pathway	BP	3	2	3.62E-03	1
	GO:0008299	isoprenoid biosynthetic process	BP	27	4	3.97E-03	1
	GO:0007599	hemostasis	BP	331	17	4.14E-03	1
	GO:0007156	homophilic-cell adhesion via plasma membrane adhesion molecule	BP	149	29	5.54E-13	1.16E-08
	GO:0098742	cell-cell adhesion via plasma membrane adhesion molecule	BP	214	29	6.63E-10	6.95E-06
	GO:0021800	cerebral cortex tangential migration	BP	7	4	1.17E-04	4.91E-01
	GO:0010039	response to iron ion	BP	28	6	2.52E-04	7.55E-01
	GO:0061726	mitcochondrion disassembly	BP	79	9	9.66E-04	1
	GO:0000422	mitophagy	BP	79	9	9.66E-04	1
Within	GO:0032655	regulation of interleukin-12 production	BP	48	6	1.10E-03	1
twin pair	GO:0032615	interleukin-12 production	BP	50	6	1.42E-03	1
analysis of	GO:2000310	regulation of N-methyl-D-aspartate selective glutamate receptor activity	BP	13	4	1.49E-03	1
focal	GO:0006369	termination of RNA polymerase II transcription	BP	59	7	1.64E-03	1
epilepsy	GO:0007155	cell adhesion	BP	1605	72	2.00E-03	1
discordant	GO:0006397	mRNA processing	BP	434	24	2.01E-03	1
twin pairs	GO:0080163	regulation of protein serine/threonine phosphatase activity	BP	2	2	2.07E-03	1
(blood)	GO:0022610	biological adhesion	BP	1610	72	2.19E-03	1
	GO:0048842	positive regulation of axon extension involved in axon guidance	BP	7	3	2.66E-03	1
	GO:1902669	positive regulation of axon guidance	BP	7	3	2.66E-03	1
	GO:0050921	positive regulation of chemotaxis	BP	112	9	2.90E-03	1
	GO:0021795	cerebral cortex cell migration	BP	37	6	2.96E-03	1
	GO:0016071	mRNA metabolic process	BP	612	30	3.05E-03	1
	GO:0008334	histone mRNA metabolic process	BP	24	4	3.30E-03	1

Appendix 18. Top 20 hits from pathway analysis (KEGG) search analysed for the 1000 top ranked CP-associated DMPs in buccal.

	KEGG ID	Pathway	N	DE	P.DE	FDR
	path:hsa01100	Metabolic pathways	1192	34	1.08E-13	3.42E-11
	path:hsa00980	Metabolism of xenobiotics by cytochrome P450	69	8	1.25E-09	1.98E-07
	path:hsa00860	Porphyrin and chlorophyll metabolism	36	7	2.61E-09	2.75E-07
	path:hsa00140	Steroid hormone biosynthesis	53	7	1.17E-08	9.30E-07
	path:hsa00040	Pentose and glucuronate interconversions	30	6	4.38E-08	2.78E-06
	path:hsa05204	Chemical carcinogenesis	75	7	5.46E-08	2.89E-06
	path:hsa00983	Drug metabolism - other enzymes	41	6	1.65E-07	7.48E-06
	path:hsa05206	MicroRNAs in cancer	279	12	2.02E-07	7.99E-06
Focal vs	path:hsa00053	Ascorbate and aldarate metabolism	23	5	2.75E-07	9.67E-06
generalised	path:hsa00982	Drug metabolism - cytochrome P450	63	6	3.27E-07	1.04E-05
epilepsy	path:hsa05205	Proteoglycans in cancer	198	11	1.08E-06	3.11E-05
	path:hsa00830	Retinol metabolism	57	5	9.87E-06	2.61E-04
	path:hsa04550	Signaling pathways regulating pluripotency of stem cells	135	8	2.01E-05	4.90E-04
	path:hsa04310	Wnt signaling pathway	138	8	2.65E-05	6.01E-04
	path:hsa04145	Phagosome	141	7	3.95E-05	8.35E-04
	path:hsa03018	RNA degradation	73	5	1.13E-04	2.24E-03
	path:hsa04142	Lysosome	118	6	1.86E-04	3.46E-03
	path:hsa04940	Type I diabetes mellitus	41	4	2.38E-04	4.19E-03
	path:hsa05418	Fluid shear stress and atherosclerosis	136	6	3.00E-04	5.01E-03
	path:hsa05200	Pathways in cancer	384	11	4.04E-04	6.01E-03
	path:hsa01100	Metabolic pathways	1192	26	3.70E-09	1.17E-06
	path:hsa05203	Viral carcinogenesis	191	10	5.66E-07	8.98E-05
	path:hsa04211	Longevity regulating pathway	88	7	4.46E-06	2.96E-04
	path:hsa04668	TNF signaling pathway	106	7	4.79E-06	2.96E-04
	path:hsa05217	Basal cell carcinoma	55	6	5.52E-06	2.96E-04
	path:hsa05200	Pathways in cancer	384	13	5.61E-06	2.96E-04
Within	path:hsa04728	Dopaminergic synapse	126	7	3.69E-05	1.67E-03
twin pair	path:hsa05169	Epstein-Barr virus infection	194	8	4.85E-05	1.92E-03
analysis of	path:hsa04144	Endocytosis	252	9	7.14E-05	2.40E-03
generalised	path:hsa03018	RNA degradation	73	5	7.58E-05	2.40E-03
epilepsy	path:hsa04150	mTOR signaling pathway	147	7	1.02E-04	2.95E-03
discordant	path:hsa05205	Proteoglycans in cancer	198	8	1.22E-04	3.22E-03
twin pairs	path:hsa04110	cell cycle	121	6	1.46E-04	3.56E-03
(buccal)	path:hsa03013	RNA transport	149	6	1.82E-04	3.93E-03
	path:hsa04390	Hippo signaling pathway	151	7	1.86E-04	3.93E-03
	path:hsa05202	ranscriptional misregulation in cancer	168	7	2.15E-04	4.25E-03
	path:hsa04020	Calcium signaling pathway	172	7	2.61E-04	4.86E-03
	path:hsa04922	Glucagon signaling pathway	96	5	3.47E-04	5.52E-03
	path:hsa05166	HTLV-I infection	248	8	3.54E-04	5.52E-03
	path:hsa04925	Aldosterone synthesis and secretion	81	5	3.55E-04	5.52E-03
	path:hsa05206	MicroRNAs in cancer	278	13	1.40E-08	2.97E-06
	path:hsa01100	Metabolic pathways	1192	26	1.87E-08	2.97E-06
	path:hsa04210	Apoptosis	134	10	2.88E-08	3.04E-06
	path:hsa04144	Endocytosis	252	13	6.92E-08	5.49E-06
	path:hsa05200	Pathways in cancer	384	16	8.83E-08	5.60E-06
	path:hsa00220	Arginine biosynthesis	18	4	5.30E-06	2.80E-04
Within	path:hsa05016	Huntington's disease	180	8	2.38E-05	9.89E-04
twin pair	path:hsa04971	Gastric acid secretion	74	6	2.49E-05	9.89E-04
analysis of	path:hsa04151	PI3K-Akt signaling pathway	317	11	2.94E-05	1.04E-03
focal	path:hsa05014	Amyotrophic lateral sclerosis (ALS)	51	5	4.11E-05	1.10E-03
epilepsy	path:hsa04724	Glutamatergic synapse	113	7	4.34E-05	1.10E-03
discordant	path:hsa04015	Rap1 signaling pathway	206	9	4.49E-05	1.10E-03
twin pairs	path:hsa04215	Apoptosis - multiple species	31	4	4.86E-05	1.10E-03
(buccal)	path:hsa00471	D-Glutamine and D-glutamate metabolism	3	2	4.98E-05	1.10E-03
	path:hsa04152	AMPK signaling pathway	117	7	5.20E-05	1.10E-03
	path:hsa05164	Influenza A	152	7	5.70E-05	1.13E-03
	path:hsa05215	Prostate cancer	84	6	7.80E-05	1.45E-03
	path:hsa05010	Alzheimer's disease	159	7	9.07E-05	1.60E-03
	path:hsa04750	Inflammatory mediator regulation of TRP channels	96	6	1.02E-04	1.70E-03
	path:hsa05168	Herpes simplex infection	164	7	1.14E-04	1.81E-03

Appendix 19. Top 20 hits from pathway analysis (KEGG) search analysed for the 1000 top ranked CP-associated DMPs in blood.

	KEGG ID	Pathway	N	DE	P.DE	FDR
Focal vs generalised epilepsy	path:hsa01100	Metabolic pathways	1192	25	8.80E-08	2.79E-05
	path:hsa04024	cAMP signaling pathway	192	10	2.17E-06	3.44E-04
	path:hsa05206	MicroRNAs in cancer	279	10	7.81E-06	8.25E-04
	path:hsa03013	RNA transport	149	7	3.08E-05	2.44E-03
	path:hsa05034	Alcoholism	165	7	9.53E-05	5.96E-03
	path:hsa04144	Endocytosis	252	9	1.42E-04	5.96E-03
	path:hsa04115	p53 signaling pathway	68	5	1.49E-04	5.96E-03
	path:hsa04151	PI3K-Akt signaling pathway	317	10	1.57E-04	5.96E-03
	path:hsa04010	MAPK signaling pathway	245	9	1.69E-04	5.96E-03
	path:hsa04510	Focal adhesion	191	8	2.02E-04	6.39E-03
	path:hsa04110	Cell cycle	121	6	2.24E-04	6.45E-03
	path:hsa04970	Salivary secretion	83	5	2.55E-04	6.72E-03
	path:hsa05014	Amyotrophic lateral sclerosis (ALS)	51	4	6.21E-04	1.51E-02
	path:hsa05416	Viral myocarditis	54	4	7.64E-04	1.73E-02
	path:hsa04360	Axon guidance	168	7	9.04E-04	1.91E-02
	path:hsa05205	Proteoglycans in cancer	198	7	1.23E-03	2.32E-02
	path:hsa04146	Peroxisome	81	4	1.24E-03	2.32E-02
	path:hsa04621	NOD-like receptor signaling pathway	152	5	1.57E-03	2.77E-02
	path:hsa04919	Thyroid hormone signaling pathway	112	5	2.16E-03	3.60E-02
	path:hsa04020	Calcium signaling pathway	172	6	2.62E-03	3.79E-02
Within twin pair analysis of generalised epilepsy discordant twin pairs (blood)	path:hsa01100	Metabolic pathways	1193	29	2.55E-12	8.07E-10
	path:hsa04144	Endocytosis	252	13	8.93E-09	1.42E-06
	path:hsa05322	Systemic lupus erythematosus	119	8	2.27E-08	2.40E-06
	path:hsa04072	Phospholipase D signaling pathway	142	9	6.97E-07	5.52E-05
	path:hsa04080	Neuroactive ligand-receptor interaction	256	10	1.06E-06	6.70E-05
	path:hsa04014	Ras signaling pathway	217	10	1.44E-06	7.61E-05
	path:hsa05200	Pathways in cancer	384	13	2.14E-06	9.67E-05
	path:hsa04015	Rap1 signaling pathway	206	9	1.20E-05	4.77E-04
	path:hsa05321	Inflammatory bowel disease (IBD)	61	5	1.50E-05	5.28E-04
	path:hsa05203	Viral carcinogenesis	191	8	2.15E-05	6.82E-04
	path:hsa04659	Th17 cell differentiation	102	6	2.65E-05	7.53E-04
	path:hsa05034	Alcoholism	165	7	2.85E-05	7.53E-04
	path:hsa04151	PI3K-Akt signaling pathway	317	10	3.55E-05	8.66E-04
	path:hsa03030	DNA replication	34	4	4.19E-05	9.48E-04
	path:hsa04150	mTOR signaling pathway	147	7	6.23E-05	1.32E-03
	path:hsa04022	cGMP-PKG signaling pathway	157	7	7.39E-05	1.46E-03
	path:hsa04071	Sphingolipid signaling pathway	117	6	1.07E-04	2.00E-03
	path:hsa05169	Epstein-Barr virus infection	194	7	1.94E-04	3.42E-03
	path:hsa00900	Terpenoid backbone biosynthesis	22	3	2.24E-04	3.73E-03
	path:hsa04660	T cell receptor signaling pathway	99	5	3.12E-04	4.95E-03

	path:hsa03040	Spliceosome	126	11	1.08E-09	3.42E-07
	path:hsa04151	PI3K-Akt signaling pathway	317	13	1.47E-06	1.89E-04
	path:hsa05203	Viral carcinogenesis	191	10	1.79E-06	1.89E-04
	path:hsa04137	Mitophagy - animal	61	6	6.95E-06	4.68E-04
	path:hsa04010	MAPK signaling pathway	245	11	7.39E-06	4.68E-04
	path:hsa04668	TNF signaling pathway	106	7	9.38E-06	4.96E-04
Within	path:hsa05169	Epstein-Barr virus infection	194	9	1.57E-05	7.10E-04
twin pair	path:hsa05200	Pathways in cancer	384	13	2.39E-05	9.48E-04
analysis of	path:hsa04350	TGF-beta signaling pathway	83	6	3.51E-05	1.12E-03
focal	path:hsa04360	Axon guidance	168	9	3.53E-05	1.12E-03
epilepsy	path:hsa04390	Hippo signaling pathway	151	8	6.87E-05	1.54E-03
discordant	path:hsa05416	Viral myocarditis	54	5	6.88E-05	1.54E-03
twin pairs	path:hsa04136	Autophagy - other	29	4	7.06E-05	1.54E-03
(blood)	path:hsa01100	Metabolic pathways	1192	20	7.16E-05	1.54E-03
	path:hsa04152	AMPK signaling pathway	117	7	7.30E-05	1.54E-03
	path:hsa04612	Antigen processing and presentation	67	5	1.03E-04	2.04E-03
	path:hsa04211	Longevity regulating pathway	88	6	1.16E-04	2.16E-03
	path:hsa05166	HTLV-I infection	248	9	1.65E-04	2.90E-03
	path:hsa04144	Endocytosis	252	9	1.96E-04	3.26E-03
	path:hsa04530	Tight junction	163	7	2.10E-04	3.32E-03

Appendix 20. Cross-platform validation of CpG sites of top-ranking DMRs.

Gene	Tissue	HM850 probe	Sequenom CpG unit	Correlation coefficient (r)	P-value of correlation coefficient (r)
GFPT2	Blood	cg13944838	CpG 7.8.9	0.966	1.00E-08
GFPT2	Buccal	cg13944838	CpG 7.8.9	0.939	2.20E-07
PM20D1	Blood	cg07533224	CpG9.10	0.973	4.90E-07
PM20D1	Blood	cg05841700	CpG12	0.942	1.70E-05
PM20D1	Blood	cg07167872	CpG15	0.952	6.35E-06
PM20D1	Buccal	cg07533224	CpG9.10	0.919	6.38E-05
PM20D1	Buccal	cg05841700	CpG12	0.78	2.77E-03
PM20D1	Buccal	cg07167872	CpG15	0.861	6.64E-04
ARID5B	Blood	cg17973778	CpG1	0.741	5.83E-03
ARID5B	Blood	cg22846816	CpG2	0.769	3.46E-03
ARID5B	Blood	cg08241406	CpG 3	0.802	1.70E-03
ARID5B	Blood	cg08241406	CpG 4	0.799	1.82E-03
ARID5B	Blood	cg01969278	CpG 6.7	0.699	1.14E-02
OTX1	Buccal	cg18947951	CpG 1	0.856	3.81E-04
OTX1	Buccal	cg04371726	CpG 3.4	0.626	2.94E-02
TTC39C	Blood	cg12639429	CpG1	0.495	4.30E-02
TTC39C	Blood	cg05401069	CpG2	0.111	6.61E-01
DLX5	Buccal	cg15339231	CpG1	0.962	1.00E-05