

Possible relationship between common genetic variation and white matter development in a pilot study of preterm infants

Michelle L. Krishnan¹, Zi Wang², Matt Silver³, James P. Boardman⁴, Gareth Ball¹, Serena J. Counsell¹, Andrew J. Walley⁵, Giovanni Montana² & Anthony David Edwards¹

¹Centre for the Developing Brain, King's College London, St Thomas' Hospital, London SE1 7EH, UK

²Department of Biomedical Engineering, King's College London, St Thomas' Hospital, London SE1 7EH, UK

³Department of Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

⁴MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh EH16 4TJ, UK

⁵School of Public Health, Faculty of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK

Keywords

Brain development, genomics, lipids, magnetic resonance imaging, metabolic pathways, multivariate analysis, neonatal, single-nucleotide polymorphism

Correspondence

Michelle L. Krishnan, Centre for the Developing Brain, Department of Perinatal Imaging, 1st Floor South Wing, St Thomas' Hospital, London SE1 7EH, UK.
Tel: +44 20 7188 9145; Fax: +44 (0) 20 7188 9154; E-mail: michelle.krishnan@kcl.ac.uk

Funding Information

This study was supported by the NIHR BRCs of Imperial College London and Guy's and St Thomas' NHS Foundation Trust/King's College London, and the MRC through Strategic Grant to ADE and Clinical Training Fellowship to MLK (Grant Ref: MR/L001578/1).

Received: 8 September 2015; Revised: 16 December 2015; Accepted: 19 December 2015

Brain and Behavior, 2016; 0(0), e00434,
doi: 10.1002/brb3.434

Abstract

Background: The consequences of preterm birth are a major public health concern with high rates of ensuing multisystem morbidity, and uncertain biological mechanisms. Common genetic variation may mediate vulnerability to the insult of prematurity and provide opportunities to predict and modify risk. **Objective:** To gain novel biological and therapeutic insights from the integrated analysis of magnetic resonance imaging and genetic data, informed by prior knowledge. **Methods:** We apply our previously validated pathway-based statistical method and a novel network-based method to discover sources of common genetic variation associated with imaging features indicative of structural brain damage. **Results:** Lipid pathways were highly ranked by Pathways Sparse Reduced Rank Regression in a model examining the effect of prematurity, and PPAR (peroxisome proliferator-activated receptor) signaling was the highest ranked pathway once degree of prematurity was accounted for. Within the PPAR pathway, five genes were found by Graph Guided Group Lasso to be highly associated with the phenotype: aquaporin 7 (AQP7), malic enzyme 1, NADP(+)-dependent, cytosolic (ME1), perilipin 1 (PLIN1), solute carrier family 27 (fatty acid transporter), member 1 (SLC27A1), and acetyl-CoA acyltransferase 1 (ACAA1). Expression of four of these (ACAA1, AQP7, ME1, and SLC27A1) is controlled by a common transcription factor, early growth response 4 (EGR-4). **Conclusions:** This suggests an important role for lipid pathways in influencing development of white matter in preterm infants, and in particular a significant role for interindividual genetic variation in PPAR signaling.

Introduction

Preterm birth accounts for 11% of all births (Blencowe et al. 2012), and is the leading global cause of deaths under 5 years of age (March of Dimes, 2012; World Health Organization, 2014). Over 30% of survivors experience cognitive problems (Moore et al. 2012) which last into adulthood, manifesting in a specific manner with

anxiety, inattention, and social and communication problems compared to term-born infants (Hack 2009). This is associated with a higher prevalence of psychopathology with a three to eightfold increased risk of behavioral problems in preterm infants, and a three to eightfold increase in the prevalence of Autism Spectrum Disorders (ASD) compared to the general population (Baird et al. 2006; Williams et al. 2006; Johnson and Wolke 2013), as

well as a risk ratio of 7.4 for bipolar affective disorder among infants born below 32 weeks of gestation (Nosarti *et al.* 2012).

White matter myelination in normal growth is typically restricted to the brain stem, globus pallidus and ventrolateral nucleus of the thalamus up to term (Brody *et al.* 1987; Kinney *et al.* 1988; Tanaka *et al.* 1995), then enters a period of rapid development from 38 weeks' gestation onwards. This continues apace particularly in the first and second years, and extends into adulthood (Yakovlev and LeCours 1967; Bartzokis *et al.* 2010; Groeschel *et al.* 2010; Miller *et al.* 2012). White and gray matter show linked yet characteristic trajectories in normal development (Groeschel *et al.* 2010), with brain volume, cortical thickness, and surface area peaking in growth rate during late childhood/early adolescence (Lenroot *et al.* 2007; Nie *et al.* 2013), and cortical folding seeming to peak earlier in childhood (Zilles *et al.* 1989; Armstrong *et al.* 1995; Nie *et al.* 2013; Li *et al.* 2014). The effect of prematurity on these processes as evaluated with MRI indicates widespread alterations of the white matter that correlate with functional measures (Krishnan *et al.* 2007; Counsell *et al.* 2008; Eikenes *et al.* 2011; van Kooij *et al.* 2012; Groppo *et al.* 2014), paralleled by changes in the overlying gray matter (Ajayi-Obe *et al.* 2000; Ball *et al.* 2013; Vinall *et al.* 2013; Smyser *et al.* 2015).

Preterm brain injury can be considered a broad entity resulting from various factors such as hypoxia, ischemia, infection, and inflammation (Volpe 2009; Jablonska *et al.* 2012) that can have a variable impact on development. The principal neuropathological mechanisms in the preterm infant have been identified as periventricular leukomalacia (PVL) and neuronal/axonal disease, forming the composite of encephalopathy of prematurity. PVL is most commonly of a diffuse form and involves microscopic necroses that later form glial scars which are hard to detect with imaging (Volpe 2009). The more visible aspect of PVL comprises marked astrogliosis and microgliosis, alongside fluctuations in numbers of premyelinating oligodendrocytes and oligodendroglial progenitors (Haynes *et al.* 2003; Back *et al.* 2005; Robinson *et al.* 2006; Billiards *et al.* 2008).

Diffusion MRI (d-MRI) provides measures of white matter structure that are correlated with neurodevelopmental outcome (Counsell *et al.* 2008; van Kooij *et al.* 2012; Ball *et al.* 2015) and highly heritable, such that 60% of the variability in d-MRI measures between individuals in the neonatal period can be attributed to genetic factors and persists into adulthood (Geng *et al.* 2012; Shen *et al.* 2014). The d-MRI measure of fractional anisotropy FA (the fraction of the magnitude of diffusion that can be attributed to directionally dependent diffusion), has been shown to increase during the early

(premyelination) stage of white matter development (Wimberger *et al.* 1995) and subsequent myelination by maturing oligodendrocytes leads to further increases in FA, with early sites including the posterior limb of the internal capsule and the optic radiations (Huppi *et al.* 1998; Neil *et al.* 1998; Drobyshevsky *et al.* 2005). Decreased fractional anisotropy FA in preterm infants is related to cognitive, fine-motor, and gross-motor outcome at 2 years (van Kooij *et al.* 2012), and these alterations of white matter persist into adulthood in very preterm individuals and are associated with cognitive function (Allin *et al.* 2011). Imaging endophenotypes provide a more direct link to genetic underpinnings than the neurodevelopmental or behavioral features of disease, demonstrating higher genetic penetrance and informing on the biological foundation of disease.

Susceptibility to perinatal brain injury is likely to be modulated by the combined effects of multiple genes of individually small effect in response to environmental influences during pregnancy and in the early postnatal period (Dempfle *et al.* 2008; Leviton *et al.* 2015). Common DNA sequence variation is estimated to account for up to 50% of additive genetic variation in complex traits, including neuroanatomical features (Yang *et al.* 2010; Toro *et al.* 2014) as well as neurological disorders including autism (Gaugler *et al.* 2014), epilepsy (Speed *et al.* 2014), and schizophrenia (Arnedo *et al.* 2014). Given that preterm birth poses an extreme challenge to the whole organism, it is plausible that common variation between individuals results in differential vulnerability to adverse stimuli, impacting development.

In this work, we focus on understanding the influence of common genetic variation on white matter development within the preterm population. The stress of premature extrauterine life leads to a broad range of neuroimaging changes and related neurological outcomes, producing contrast between individuals that we are able to use here in a within-group design. The sparse regression approaches used in this study have been developed for the selection of explanatory variables associated with a quantitative trait within a study cohort, and we employ these methods here for that purpose.

Statistical genetics and bioinformatics methods allow the principled joint analysis of the large imaging and genetic datasets involved, and facilitate the biological interpretation of results. In contrast with a hypothesis-based approach that aims to test a specific set of assumptions against a significance threshold, a data-driven technique seeks a principled and biologically informed way to uncover the signal within the data, yielding unbiased and novel insights that can be validated experimentally in a constructive and iterative manner (Robinson *et al.* 2014). A traditional hypothesis-based method is the genome-

wide association study (GWAS) whereby each single-nucleotide polymorphism (SNP) is tested for association with the phenotype, requiring typically hundreds of thousands of hypotheses and stringently adjusted p -values. The regression modeling approach, in comparison, involves fitting a predictive model for the phenotype using all SNPs, while also ranking all SNPs based on their predictive value. Regression modeling is of particular benefit in imaging genomics studies where the space of possible hypotheses is vast, encompassing genetic features (SNPs) multiplied by image features (voxels). This obviates the need for multiple-testing correction and significance thresholds, while producing a meaningful ranking of results (Silver and Montana 2012b).

In these large datasets, the number of subjects n is typically much smaller than the number of features p (e.g., single-nucleotide polymorphisms SNPs), posing a statistical and analytical problem. Two current approaches are to either increase n significantly or find a principled way to reduce p while preserving the underlying signal. We have addressed this problem by developing a pathways-driven sparse regression method (PsRRR) (Silver and Montana 2012b) which we have robustly validated (Silver *et al.* 2013) and extended to multivariate imaging traits (Silver *et al.* 2012a). We have subsequently applied the Graph Guided Group Lasso (GGGL) to improve SNP and gene selection by integrating information from grouping SNPs into genes and organizing genes into a weighted gene network encoding the functional relatedness between all pairs of genes (Wang and Montana 2014). We apply these methods to the preterm population, leveraging prior biological knowledge by using SNPs and genes grouped into biological pathways or networks, which allows the detection of previously unexposed signal (Wang *et al.* 2010) and eases the interpretation of results (Cantor *et al.* 2010). Common genetic variation within biological canonical pathways and functional networks is used to explain interindividual variation in imaging features relevant to neurodevelopmental outcome.

Patients and Methods

Patient characteristics

Participants' characteristics: mean GA 28 + 4 weeks, range 23 + 2 to 32 + 6, mean PMA at scan 40 + 3, range 27 + 4 to 47 + 6 weeks. This cohort has previously been described in detail (Boardman *et al.* 2014). Research was carried out in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), with approval from the NHS Research Ethics Committee and to the standard of the associated granting agencies.

MR Image acquisition and analysis

MR images were acquired for 72 preterm infants (mean gestational age (GA) 28 + 4 weeks, mean postmenstrual age (PMA) at scan 40 + 3 weeks). Imaging was performed on a Philips 3-Tesla system (Philips Medical Systems, Netherlands) using an eight-channel phased array head coil. Single-shot echo-planar diffusion tensor imaging was acquired in the transverse plane in 15 noncollinear directions using the following parameters: repetition time (TR): 8000 msec; echo time (TE): 49 msec; slice thickness: 2 mm; field of view: 224 mm; matrix: 128 × 128 (voxel size: 1.7531 × 1.753 × 2 mm³); b value: 750 sec/mm²; SENSE factor: 2. T1-weighted 3D MPRAGE were acquired with parameters: TR = 17 msec, TE = 4.6 msec, inversion delay = 1500 msec, flip angle = 13°, acquisition plane = sagittal, voxel size = 0.82 × 1.03 × 1.6 mm, FOV = 210 × 167 mm and acquired matrix = 256 × 163. T2-weighted fast spin-echo: TR × 8700 ms, TE = 160 msec, flip angle = 90°, acquisition plane = axial, voxel size = 1.15 × 1.18 × 2 mm, FOV = 220 mm, and acquired matrix = 192 × 186.

Diffusion tensor imaging (DTI) analysis was performed by using FMRIB's Diffusion Toolbox (v2.0; RRID:nif-0000-00305) as implemented in FMRIB's Software Library (FSL v4.1.5; www.fmrib.ox.ac.uk/fsl) (Smith and Nichols 2009). Each infant's diffusion-weighted image (DWI) was registered to their respective nondiffusion-weighted ($b = 0$) image and corrected for differences in spatial distortion owing to eddy currents. Images were brain extracted using Brain Extraction Tool (BET v2.1), diffusion tensors calculated voxelwise, and FA maps generated.

Tract-Based Spatial Statistics (TBSS) (Smith *et al.*, 2006) was performed by using a modified pipeline specifically optimized for neonatal DTI analysis. This included an initial low degrees-of-freedom linear registration step and a second registration to a population-average FA map, which has been shown to reduce global misalignment between neonatal fractional anisotropy (FA) maps (Ball *et al.* 2010). The aligned data were used to create a mean FA map and a mean FA skeleton that represents the center of all white matter tracts common to the group. The FA skeleton was thresholded at FA ≥ 0.2 before each infant's aligned FA data were projected onto it.

Three separate adjustments for clinical variables were made to the TBSS phenotype using the FSL general linear model tool, with permutation-based significance testing using the randomize tool with Threshold-Free Cluster Enhancement. First, imaging data were adjusted for the effect of PMA at scan only; second, the imaging data were adjusted for both PMA and GA, thereby focusing on the effect of genetic variation and environment. Third, there was an adjustment for genetic ancestry. In all cases,

dimension reduction in the phenotype with principal component analysis was carried out for computational efficiency.

Saliva genotyping

The concentration of all the genomic DNA samples was measured using the PicoGreen protocol. 200 ng of genomic DNA was used for each Illumina HumanOmniExpress-12 array according to the manufacturer's instructions. HumanOmniExpress-12 arrays have 730,525 markers with a mean spacing of ~4 kb. 392,197 of those markers are within 10 kb of a known RefSeq gene and there are 15,062 coding SNPs and 7459 MHC markers included in that total. All samples successfully passed quality control.

Genome-wide genotyping

Samples were genotyped on Illumina HumanOmniExpress-12 arrays. The genotype matrix was recoded in terms of minor allele counts, including only SNPs with MAF $\geq 5\%$ and $\geq 99\%$ genotyping rate (Purcell *et al.* 2007). After these filtering steps, 613,186 SNPs remained.

Assessment of population stratification

Whole genome SNP data were used for complete linkage agglomerative clustering, based on pairwise identity-by-state (IBS) distance as implemented in PLINK 1.9 (Chang *et al.* 2015), to assess whether any two individuals belonged to the same population. Dimension reduction in the IBS distance matrix was carried out by principal component analysis, and the first principal component was used as a covariate in TBSS analysis to adjust for population stratification.

Information on self-reported ethnicity (as defined in ISB standard DSCN 11/2008) was collected by asking mothers (and fathers when present) to define themselves according to a list of options. The terms were drawn from Ethnic Category National Codes as in Department of Health Guidance at the time. Parental self-reported ethnicity was summarized into broader categories for the purposes of data visualization by aggregating all "White" subcategories into a single group "White", all "Black" subcategories into "Black", and all "Asian" subcategories into "Asian". In cases where either one parent self-reported as "Mixed" or if there was a discrepancy between maternal and paternal ethnicities, the term "Mixed" was applied. Where parents were both from an ASEAN member state (two cases) the individual was classified by the authors as "SE Asian". These aggregated ethnic categories were used to label the datapoints of the PCA plot of the first two principal components of the

IBS variance-standardized relationship matrix. This illustrates the correspondence between the first two components of genetic ancestry and ethnicity, and provides an overview of the cohort population mixture.

Pathways sparse reduced rank regression

Pathways sparse reduced rank regression estimates the regression coefficients in the linear model with multivariate responses, subjected to constraints. The model accounts for potential biasing factors such as pathway linkage disequilibrium and size by using an adaptive, weight-tuning procedure. Pathway weightings in the regression model are adjusted according to the empirical bias in pathway selection frequencies, obtained by fitting the PsRRR model with a null response. Depending on the degree of penalization, some coefficients are driven to zero, thus performing variable selection so that only SNPs within associated pathways are retained in the model. Pathways are ranked in order of importance using a resampling strategy, with highest ranked pathways having highest selection frequency and highest correlation with the phenotype. SNPs were mapped to genes (NCBI GRCh37) and KEGG pathways (Kyoto Encyclopedia of Genes and Elements), after excluding cancer-related pathways due to high redundancy in gene membership as previously described (Silver *et al.* 2012a), which nonetheless allows the genes to participate in the model as part of other pathways. Parameters used: Model adjusted for PMA; λ 0.99, 100 subsamples, 20 iterations with 2000×10 plus 4000×10 model fits per iteration. Model adjusted for GA and PMA; λ 0.99, 100 subsamples, 20 iterations with 2000×10 plus 4000×10 model fits per iteration.

Graph Guided Group Lasso

Graph Guided Group Lasso incorporates prior information from SNP-gene mapping as well as from the gene functional interaction network to guide variable selection (Wang and Montana 2014). The functional relationships between genes within the top ranked pathway peroxisome proliferator-activated receptor (PPAR) signaling pathway were systematically described by clustering the genes based on their GO BP annotations (Ovaska *et al.* 2008). This resulted in an adjacency matrix based on pairwise semantic similarity of GO terms, which has been shown to correlate with protein sequence similarity (Apweiler *et al.* 2004) and protein family similarity (Couto *et al.* 2006).

The GGGL approach assumes a linear relationship between gene function (described here by GO annotation) and brain imaging features, applying an additional

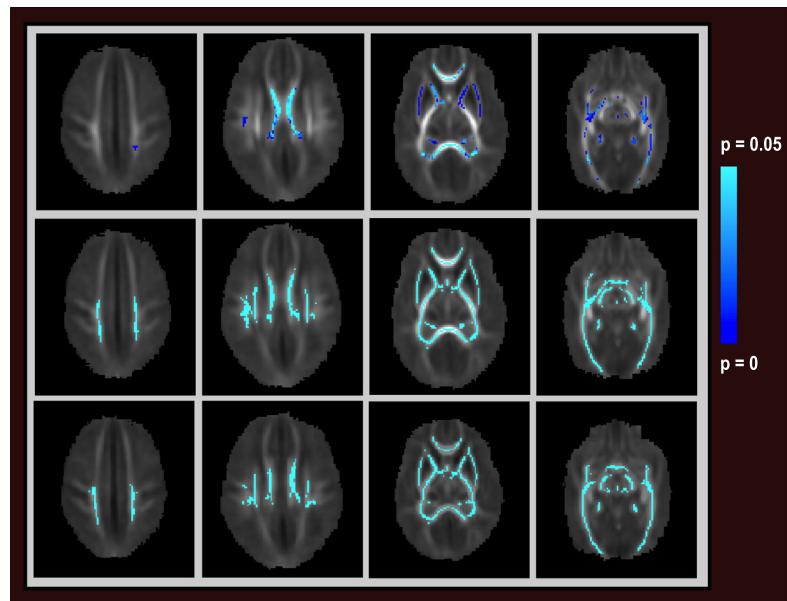


Figure 1. Group white matter DTI skeleton, showing voxels that vary significantly between individuals (corrected $P < 0.05$ for all voxels, darker blue signifies lower P -value). Axial views superior to inferior left to right. Top row: Voxels varying between individuals adjusting for PMA at scan. Middle row: Voxels varying between individuals adjusting for GA at birth and PMA. Bottom row: Voxels varying between individuals adjusting for GA, PMA and genetic ancestry.

penalty function on the regression coefficients that incorporates network structure information. The model selects functionally related genes that are associated with the trait and identifies important SNPs within the selected genes. GGGL-1 further imposes that SNPs in functionally related genes have similar effects on the phenotypic trait.

The regularization parameter μ controls the weight of the prior knowledge added to the squared loss function. When μ is large the model relies more on prior knowledge from the network. Using GGGL-1 with $\mu = 0.1, 1, 10$ resulted in the same set of genes with selection probability > 0.4 (a threshold representing a step change in the probability distribution). Using GGGL-2 with $\mu = 0.1, 1, 10$ also yielded very stable results, with 27 of the top 30 SNPs appearing in all three lists for different values of μ , and the same genes were selected in the three instances. Both versions of the method (GGGL-1 and GGGL-2) produced comparable results, with the same genes and SNPs being ranked highly by both approaches (Tables S8 and S9). Genes with more than one occurrence and a selection probability threshold > 0.4 were retained, at a threshold representing a step change in the probability distribution.

Results

Imaging phenotypes

A group white matter skeleton was constructed and used to extract voxel-wise fractional anisotropy (FA) values for each of the 72 individuals. These values were serially adjusted for important clinical variables: effect of gestational age (GA) at birth, effect of postmenstrual age

(PMA) at scan and genetic ancestry, resulting in three separate phenotypes. In all instances the phenotypes underwent dimension reduction with principal component analysis (PCA). The elbow of the scree plot was selected in each case to determine the number of components used for the subsequent analysis (scree plots in Figure S1).

The statistic images (Fig. 1) indicate that GA at birth had a significant impact on the central white matter in particular, whereas including correction for PMA resulted in a more diffuse effect, and correcting for genetic ancestry in addition to GA and PMA had little effect on the phenotype (distributions of residuals in Figure S2).

Effect of population stratification

We use the term population stratification to refer here to allele frequency differences between subpopulations within a population. Genetic ancestry is a cause of population stratification, and the term here refers to the use of empirical methods to assign an ancestry classification (Ali-Khan *et al.* 2011). Population stratification of the cohort was assessed by calculating pairwise identity by state (IBS) values and using these to perform complete linkage clustering (Methods). This revealed a degree of stratification along the first two components, corresponding with parental self-reported ethnicity (Figure S3).

When the effect of genetic ancestry was included in the imaging phenotype, the PsRRR algorithm did not converge, suggesting a loss of signal in the relationship between genetic predictors and imaging phenotype after adjustment for genetic ancestry. Subsequent analyses therefore refer only to the models adjusting for GA and/or PMA.

Pathway ranking

SNPs were mapped to genes and pathways from the Kyoto Encyclopedia of Genes and Elements (KEGG) as summarized in Table S4.

The PsRRR method was used to identify and rank biological pathways predictive of white matter integrity. Pathways were ranked by stability selection in order of selection frequency, with each adjusted phenotype analyzed in turn (Table 1). These empirical results were compared with null selection frequencies (Fig. 2, Figure S5, Table S6, and Table S7).

Lipid pathways were significantly over-represented in the top ranking pathways adjusted for PMA ($P \leq 0.005$) (Table 1, bold) and although the total number of lipid pathways decreased with adjustment for GA, the empirical selection frequency of the most highly ranked lipid pathway (peroxisome proliferator-activated receptor (PPAR) signaling) increased from 0.09 to 0.2 (Fig. 3).

Gene relationships

The GGGL method was applied to genes in the most highly ranked KEGG pathway (PPAR signaling) (Patients and Methods). This allowed clarification of the association between individual genes and SNPs with the phenotype, which was constrained to a biological pathway of interest. Within the PPAR pathway, the GGGL-1 method selected a subset of genes (5/69) functionally related in terms of Gene Ontology (GO) Biological Process (BP) and linearly correlated with white matter FA. These were aquaporin 7 (AQP7), malic enzyme 1, NADP(+)-dependent, cytosolic (ME1), perilipin 1 (PLIN1), solute carrier family 27 (fatty acid transporter), member 1 (SLC27A1), and acetyl-CoA acyltransferase 1 (ACAA1). Of the top thirty SNPs selected by the GGGL-2 method, all were found within genes with selection probability >0.4 in GGGL-1, indicating strong agreement between the two approaches Table S8, Table S9.

Functional relationships

Analysis of transcriptional regulation using the PASTAA algorithm (Roeder *et al.* 2007) indicated that expression of four of these five genes (ACAA1, AQP7, ME1, and SLC27A1) is controlled by a common transcription factor, early growth response (EGR-4), hypergeometric P -value 7.7×10^{-4} .

The relationships between the five genes highlighted by GGGL were further characterized using the GeneMANIA prediction algorithm (Warde-Farley *et al.* 2010), revealing close coexpression links as well as physical interactions between the seed genes and additional interacting genes

Table 1. Top thirty KEGG pathways ranked by PsRRR. Lipid pathways highlighted in bold. Left column: Results with phenotype adjusted for PMA. Right column: Results adjusted for GA and PMA.

Pathways adjusted for PMA	Pathways adjusted for GA and PMA
Glycine Serine and Threonine Metabolism	PPAR Signaling Pathway
PPAR Signaling Pathway	Dilated Cardiomyopathy
Alpha-Linolenic Acid Metabolism	Glycerolipid Metabolism
Ether Lipid Metabolism	Alpha-Linolenic Acid Metabolism
Glycerophospholipid Metabolism	Pyrimidine Metabolism
SNARE Interactions In Vesicular Transport	Calcium Signaling Pathway
Hypertrophic Cardiomyopathy HCM	Cardiac Muscle Contraction
Glycerolipid Metabolism	Hematopoietic Cell Lineage
Basal Transcription Factors	Complement and Coagulation Cascades
Cardiac Muscle Contraction	Aminoacyl tRNA Biosynthesis
Hematopoietic Cell Lineage	Pancreatic Cancer
Phosphatidylinositol Signaling System	Renin Angiotensin System
Ubiquitin Mediated Proteolysis	Nucleotide Excision Repair
Nucleotide Excision Repair	SNARE Interactions in Vesicular Transport
JAK-STAT Signaling Pathway	Glycosylphosphatidylinositol GPI Anchor Biosynthesis
Adipocytokine Signaling Pathway	Type II Diabetes Mellitus
Glycosylphosphatidylinositol GPI Anchor Biosynthesis	Epithelial Cell Signaling in Helicobacter Pylori Infection
GNRH Signaling Pathway	DNA Replication
Starch and Sucrose Metabolism	Glycine Serine and Threonine Metabolism
Long-Term Depression	Sulfur Metabolism
ABC Transporters	Dorso Ventral Axis Formation
Endocytosis	Peroxisome
Fatty Acid Metabolism	Bladder Cancer
Antigen Processing and Presentation	Primary Immunodeficiency
Ascorbate and Aldarate Metabolism	Ascorbate and Aldarate Metabolism
Lysosome	Lysosome
One Carbon pool by Folate	One Carbon pool by Folate
Fc Epsilon RI Signaling Pathway	Axon Guidance
Viral Myocarditis	Vascular Smooth Muscle Contraction
Complement and Coagulation Cascades	Fc Epsilon RI Signaling Pathway

PMA, postmenstrual age; GA, gestational age; PPAR, peroxisome proliferator-activated receptor; KEGG, Kyoto Encyclopedia of Genes and Elements; PsRRR, pathways-driven sparse regression method

(Fig. 4). This interacting set of 25 genes is significantly enriched for disease associations with fatty liver, hypertriglyceridemia, obesity, insulin resistance, and type 2 diabetes, adjusted P -value $< 5 \times 10^{-6}$ (WEBGestalt) (Wang *et al.* 2013).

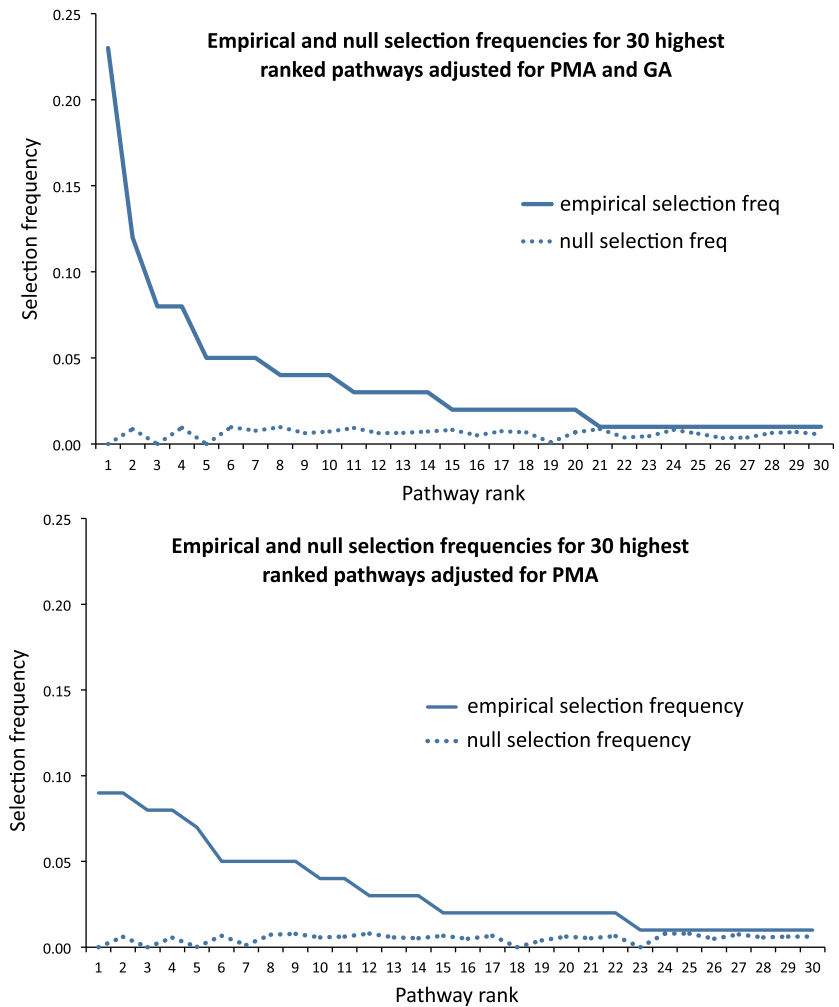


Figure 2. Empirical and null selection frequencies for thirty most predictive pathways, adjusted for PMA (top) and adjusted for GA and PMA (bottom).

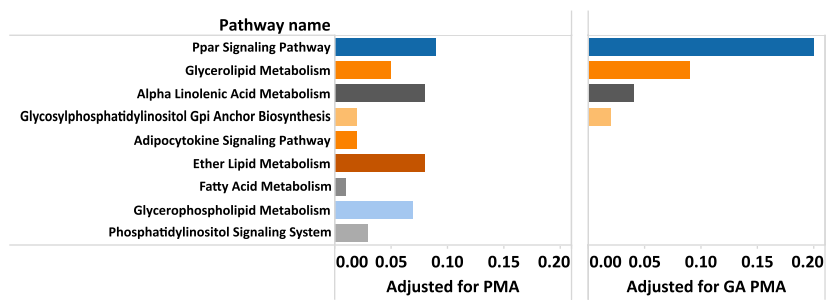


Figure 3. Empirical selection frequencies for lipid pathways among top thirty ranked by PsRRR, sorted by rank postadjustment for GA and PMA. Left: adjusted by postmenstrual age (PMA); Right: Adjusted for gestational age (GA) and postmenstrual age (PMA).

Discussion

This integrated analysis of multivariate imaging and genetic data suggests a relationship between lipid pathways, PPAR signaling particularly, and variability in preterm white matter development. Functional network methods draw a specific focus to a subset of five genes within the PPAR pathway (ACAA1, AQP7, PLIN1, ME1, and SLC27A1),

four of which are under common transcriptional control by EGR-4.

The focus of this study is on white matter development among preterm infants, and in this population TBSS detects variability in white matter features that are related both to adverse perinatal events and later neurodevelopment (Counsell *et al.* 2008; Eikenes *et al.* 2011; van Kooij *et al.* 2012). This provides a basis for within-group com-

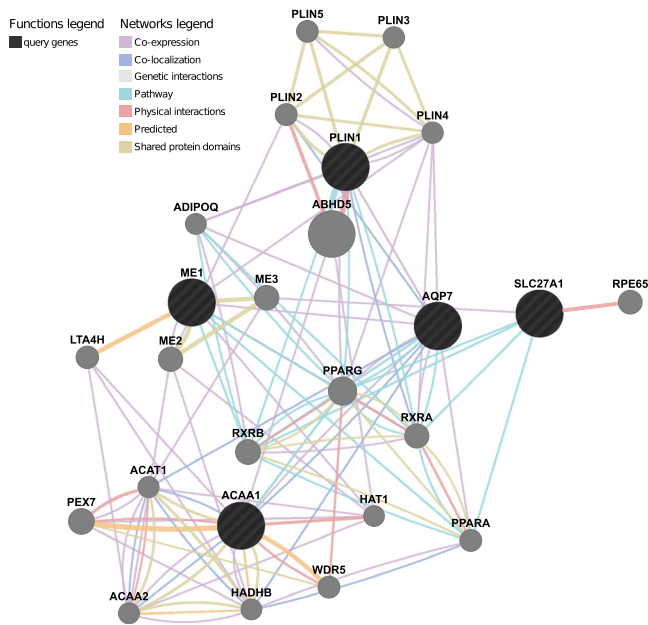


Figure 4. Functional gene relationships based on protein–protein, protein–DNA, and genetic interactions, pathways, reactions, gene and protein expression data, and protein domains. Gene function predictions are based on gene ontology (GO) annotations patterns.

parison of preterm infants; indeed healthy preterm infants are the best-matched controls for preterm infants with less favorable features. However, this design prevents inferences about differences between preterm and term infants. Nevertheless, we have detected a possible relationship in brain development that replicates findings in adults, associating lipid metabolism with variation in white matter FA in healthy adults (Braskie *et al.* 2011; Heise *et al.* 2011).

Anticipated main effects in the model were gestational age (GA) at birth, which can be viewed as degree of prematurity and length of early separation from the placenta, and postmenstrual age at scan (PMA), which captures ongoing development *ex utero*. It is known that FA increases in white matter as part of ongoing development in term and preterm infants (Huppi *et al.* 1998; Neil *et al.* 1998; Drobyshevsky *et al.* 2005; Smyser *et al.* 2015), and therefore PMA was included as a covariate throughout the analysis, in an attempt to focus on variability due to factors other than development. Regarding the implications of these observations for white matter development in later life, alterations of white matter including reduced FA have been shown to persist into adulthood in very preterm individuals and are associated with cognitive function (Allin *et al.* 2011). Given that 60% of the variability in d-MRI measures between individuals in the neonatal period can be attributed to genetic factors (Geng *et al.* 2012), and that this heritability persists into adulthood (Shen *et al.* 2014) we would hypothesize that a predictive relationship identified in the neonatal period would be preserved or even increased in adults (Trzaskowski *et al.* 2014). It has also been found in healthy

adults aged 20–78 years that the genetic effects of the ApoE4 allele on white matter FA were preserved independently of age (Heise *et al.* 2011). To evaluate whether variation in biological pathways is associated with white matter structure as a function of degree of prematurity, we firstly allowed GA to remain in the model while adjusting for PMA. The impact of lipid pathways is substantial in this model, as indicated by the significant over-representation of lipids among the top ranked pathways. When GA is adjusted for alongside PMA, specific lipid pathways such as PPAR signaling and glycerolipid metabolism increase in their degree of importance, although lipid pathways as a group are no longer over-represented.

A possible interpretation of these findings is that interindividual genetic variability in lipid metabolism has an effect on white matter structure (FA) and this is linked to degree of prematurity (GA). On adjustment for GA the main remaining effects are likely to be genetic influence and environmental variables. In this latter scenario, key lipid metabolic pathways remain an important determinant of white matter integrity in the context of a greater variety of biological processes. It is tenable that the genes highlighted here are involved in processes contributing to normal white matter myelination, and that the disruption of these physiological processes in a subset of preterms contributes to the observed variability in FA within the preterm group.

Multiple sources of evidence indicate that a well-balanced and carefully timed fatty acid supply during the neonatal period is a determinant of growth, visual development, and cognitive development (Fleith and Clandinin 2005; Innis 2007), although there is active debate on how

this should be addressed clinically (Adamkin and Radmacher 2014). We have previously highlighted a member of the PPAR pathway (fatty acid desaturase 2, FADS2) in a separate candidate gene analysis of this cohort (Boardman *et al.* 2014). FADS polymorphisms have since been independently associated with behavioral outcomes in children, suggesting a programming effect of PPAR genotype (Jensen *et al.* 2014). Other systemic effects of FADS2 could be mediated via its role in catalyzing the conversion of linoleic acid (LA) into arachidonic acid (AA) and that of alpha-linolenic acid (ALA) into eicosapentaenoic acid (EPA). Polymorphisms in FADS1 and FADS2 have been linked to a proinflammatory phenotype promoting atherosclerosis and coronary artery disease (Martinelli *et al.* 2008; Glaser *et al.* 2011), and there appears to be a genetic regulation of the level of desaturase activity that varies with ethnicity (Merino *et al.* 2011; Sergeant *et al.* 2012; Chilton *et al.* 2014).

The peroxisome proliferator-activated receptors (PPARs) are ligand-activated transcription factors belonging to the superfamily of nuclear hormone receptors, with an important role in nutrient homeostasis. They are involved in cell membrane structure, signaling, inflammation and biotransformation (Rosen and Spiegelman 2001), neuronal and glial differentiation and axon polarity, and neuroprotection (Gray *et al.* 2011; Minghetti *et al.* 2014; Quintanilla *et al.* 2014). Preterm birth is associated with an increased cardiometabolic disease risk in adulthood (Ryckman *et al.* 2013; Bayman *et al.* 2014; Kajantie and Hovi 2014), and PPAR-gamma agonists (thiazolidinediones) are widely used in the treatment of Type 2 Diabetes Mellitus (T2DM) and insulin resistance. T2DM features here among the top 30 pathways adjusted for GA and PMA, and four of the five PPAR pathway genes highlighted by GGGL are linked to T2DM.

ME1 has been associated with sex-specific gene regulation in the offspring as a result of peri-conception maternal obesity (Dahlhoff *et al.* 2014), and identified as a key regulator of a T2DM-specific gene expression network (Zhong *et al.* 2010). PLIN1 regulates droplet formation in lipopolysaccharide-stimulated microglia (Khatchadourian *et al.* 2012) and mutations have been linked to familial lipodystrophy and severe insulin resistance and T2DM (Kozusko *et al.* 2015). SLC27A1 (also known as fatty acid transport molecule 1, FATP1) is localized to mitochondria (Guitart *et al.* 2014), is involved in fatty acid transport across the blood–brain barrier (Mitchell *et al.* 2011) and has been considered as a therapeutic target for insulin resistance (Matsufuji *et al.* 2012). There are indications that AQP7 expression is associated with insulin resistance and obesity (Lebeck 2014), and missense mutations in humans result in a variety of neurological sequelae including severe hypotonia, psychomotor retardation

and/or epilepsy as well as multisystem abnormalities (Goubau *et al.* 2013). ACAA1 is involved in neuronal growth and myelinogenesis (Houdou *et al.* 1993) and may modify immune responses via Toll-like signaling (Sordillo *et al.* 2011).

ACAA1, AQP7, ME1, and SLC27A1 are jointly regulated by the EGR-4 transcription factor, which is important in neuronal maturation (Ludwig *et al.* 2011) and synaptic plasticity (Beckmann and Wilce 1997). EGR4 gene expression is induced by cerebral ischemia and inflammation (Decker *et al.* 2003; Mengozzi *et al.* 2012), key mechanisms in preterm brain injury (Vannucci and Hagberg 2004; Volpe 2012). EGR4 is also upregulated by EGFR signaling (Mayer *et al.* 2009), linked to myelination and remyelination (Aguirre *et al.* 2007). Both fatty acids and EGR signaling have been associated with mental illnesses including schizophrenia (Yamada *et al.* 2007; Matsumata *et al.* 2014) and the interaction of EGR-1 and the PPAR pathway has been described in relation to cardiovascular risk (Fruchart 2009).

Additional considerations

In this analysis we have performed a within-group comparison, using the contrast between preterm infants with different values of white matter FA to highlight differences in genetic profile that could add to the biological understanding of white matter development following premature birth. This is driven in part by the paucity of comparable linked imaging and genetic data from healthy term infants, but as a corollary allows us to specifically ask why preterm infants of a similar gestation and at a similar stage of postnatal development have different white matter features that are expected to be functionally relevant. Adjustment for genetic ancestry appeared to leave no independent associations detected by the PsRRR algorithm, suggesting a role for ethnicity and genetic background (the complete genotype of an organism across all loci) in modifying the phenotypic consequences of an allele and its association with disease risk. The impact of alleles may be modified by the local genetic architecture (Jing *et al.* 2014), and it could be that the effect of ancestry is acting through differences in lipid metabolism. In view of the small sample size involved it would be beyond the scope of this work to make statements about the significance of this observation, but this suggests an exciting avenue for further study.

Conclusion

In this hypothesis generating work, interpretation of our findings suggests that genetic variation in lipid pathways might influence white matter development among pre-

term infants. The relationship between imaging measures of brain development and genetic profile appears to be modulated by degree of prematurity. Given that preterm infants are at increased risk of both mental illness and cardiovascular morbidity in later life, this work suggests a unifying mechanism through which these systemic effects might be mediated.

Acknowledgments

Our thanks to the children and families who participated in the study. This study was supported by the NIHR BRCs of Imperial College London and Guy's and St Thomas' NHS Foundation Trust/King's College London, and the MRC through Strategic Grant to ADE and Clinical Training Fellowship to MLK (Grant Ref: MR/L001578/1).

Conflict of Interest

None declared.

References

- Adamkin, D. H., and P. G. Radmacher. 2014. Current trends and future challenges in neonatal parenteral nutrition. *J. Neonatal Perinatal Med.* 7:157–164.
- Aguirre, A., J. L. Dupree, J. M. Mangin, and V. Gallo. 2007. A functional role for EGFR signaling in myelination and remyelination. *Nat. Neurosci.* 10:990–1002.
- Ajayi-Obe, M., N. Saeed, F. M. Cowan, M. A. Rutherford, and A. D. Edwards. 2000. Reduced development of cerebral cortex in extremely preterm infants. *Lancet* 356:1162–1163.
- Ali-Khan, S. E., T. Krakowski, R. Tahir, and A. S. Daar. 2011. The use of race, ethnicity and ancestry in human genetic research. *HUGO J.* 5:47–63.
- Allin, M. P., D. Kontis, M. Walshe, J. Wyatt, G. J. Barker, R. A. Kanaan, et al. 2011. White matter and cognition in adults who were born preterm. *PLoS One* 6:e24525.
- Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, et al. 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 32:D115–D119.
- Armstrong, E., A. Schleicher, H. Omran, M. Curtis, and K. Zilles. 1995. The ontogeny of human gyrification. *Cereb. Cortex* 5:56–63.
- Arnedo, J., D. M. Svrakic, C. Del Val, R. Romero-Zaliz, H. Hernandez-Cuervo, Molecular Genetics of Schizophrenia C., et al. 2014. Uncovering the hidden risk architecture of the schizophrenias: confirmation in three independent genome-wide association studies. *Am. J. Psychiatry* 172:139–153.
- Back, S. A., N. L. Luo, R. A. Mallinson, J. P. O'Malley, L. D. Wallen, B. Frei, et al. 2005. Selective vulnerability of preterm white matter to oxidative damage defined by F2-isoprostanes. *Ann. Neurol.* 58:108–120.
- Baird, G., E. Simonoff, A. Pickles, S. Chandler, T. Loucas, D. Meldrum, et al. 2006. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet* 368:210–215.
- Ball, G., S. J. Counsell, M. Anjari, N. Merchant, T. Arichi, V. Doria, et al. 2010. An optimised tract-based spatial statistics protocol for neonates: applications to prematurity and chronic lung disease. *NeuroImage* 53:94–102.
- Ball, G., L. Srinivasan, P. Aljabar, S. J. Counsell, G. Durighel, J. V. Hajnal, et al. 2013. Development of cortical microstructure in the preterm human brain. *Proc. Natl Acad. Sci. USA* 110:9541–9546.
- Ball, G., L. Pazderova, A. Chew, N. Tusor, N. Merchant, T. Arichi, et al. 2015. Thalamocortical Connectivity Predicts Cognition in Children Born Preterm. *Cereb. Cortex* 25:4310–4318.
- Bartzokis, G., P. H. Lu, K. Tingus, M. F. Mendez, A. Richard, D. G. Peters, et al. 2010. Lifespan trajectory of myelin integrity and maximum motor speed. *Neurobiol. Aging* 31:1554–1562.
- Bayman, E., A. J. Drake, and C. Piyasena. 2014. Prematurity and programming of cardiovascular disease risk: a future challenge for public health?. *Arch. Dis. Child. Fetal Neonatal Ed.* 99:F510–F514.
- Beckmann, A. M., and P. A. Wilce. 1997. Egr transcription factors in the nervous system. *Neurochem. Int.*, 31:477–510; discussion 517–6.
- Billiards, S. S., R. L. Haynes, R. D. Folkerth, N. S. Borenstein, F. L. Trachtenberg, D. H. Rowitch, et al. 2008. Myelin abnormalities without oligodendrocyte loss in periventricular leukomalacia. *Brain Pathol.* 18:153–163.
- Blencowe, H., S. Cousens, M. Z. Oestergaard, D. Chou, A. B. Moller, R. Narwal, et al. 2012. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet* 379:2162–2172.
- Boardman, J. P., A. Walley, G. Ball, P. Takousis, M. L. Krishnan, L. Hughes-Carre, et al. 2014. Common genetic variants and risk of brain injury after preterm birth. *Pediatrics* 133:e1655–e1663.
- Braskie, M. N., N. Jahanshad, J. L. Stein, M. Barysheva, K. L. McMahon, G. I. de Zubicaray, et al. 2011. Common Alzheimer's disease risk variant within the CLU gene affects white matter microstructure in young adults. *J. Neurosci.* 31:6764–6770.
- Brody, B. A., H. C. Kinney, A. S. Kloman, and F. H. Gilles. 1987. Sequence of central nervous system myelination in human infancy. I. An autopsy study of myelination. *J. Neuropathol. Exp. Neurol.* 46:283–301.
- Cantor, R. M., K. Lange, and J. S. Sinsheimer. 2010. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* 86:6–22.

- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7.
- Chilton, F. H., R. C. Murphy, B. A. Wilson, S. Sergeant, H. Ainsworth, M. C. Seeds, et al. 2014. Diet-gene interactions and PUFA metabolism: a potential contributor to health disparities and human diseases. *Nutrients* 6:1993–2022.
- Counsell, S. J., A. D. Edwards, A. T. Chew, M. Anjari, L. E. Dyet, L. Srinivasan, et al. 2008. Specific relations between neurodevelopmental abilities and white matter microstructure in children born preterm. *Brain*, 131:3201–3208.
- Couto, F. M., M. J. Silva, and P. M. Coutinho. 2006. Measuring semantic similarity between Gene Ontology terms. *Data Knowl. Eng.* 61:137–152.
- Dahlhoff, M., S. Pfister, A. Blutke, J. Rozman, M. Klingenspor, M. J. Deutsch, et al. 2014. Peri-conceptual obesogenic exposure induces sex-specific programming of disease susceptibilities in adult mouse offspring. *Biochim. Biophys. Acta* 1842:304–317.
- Decker, E. L., N. Nehmann, E. Kampen, H. Eibel, P. F. Zipfel, and C. Skerka. 2003. Early growth response proteins (EGR) and nuclear factors of activated T cells (NFAT) form heterodimers and regulate proinflammatory cytokine gene expression. *Nucleic Acids Res.* 31:911–921.
- Dempfle, A., A. Scherag, R. Hein, L. Beckmann, J. Chang-Claude, and H. Schafer. 2008. Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur. J. Hum. Genet.* 16:1164–1172.
- Drobyshevsky, A., S. K. Song, G. Gamkrelidze, A. M. Wyrwicz, M. Derrick, F. Meng, et al. 2005. Developmental changes in diffusion anisotropy coincide with immature oligodendrocyte progression and maturation of compound action potential. *J. Neurosci.* 25:5988–5997.
- Eikenes, L., G. C. Lohaugen, A. M. Brubakk, J. Skranes, and A. K. Haberg. 2011. Young adults born preterm with very low birth weight demonstrate widespread white matter alterations on brain DTI. *NeuroImage* 54:1774–1785.
- Fleith, M., and M. T. Clandinin. 2005. Dietary PUFA for preterm and term infants: review of clinical studies. *Crit. Rev. Food Sci. Nutr.* 45:205–229.
- Fruchart, J. C. 2009. Peroxisome proliferator-activated receptor-alpha (PPARalpha): at the crossroads of obesity, diabetes and cardiovascular disease. *Atherosclerosis* 205: 1–8.
- Gaugler, T., L. Klei, S. J. Sanders, C. A. Bodea, A. P. Goldberg, A. B. Lee, et al. 2014. Most genetic risk for autism resides with common variation. *Nat. Genet.* 46:881–885.
- Geng, X., E. C. Prom-Wormley, J. Perez, T. Kubarych, M. Styner, W. Lin, et al. 2012. White matter heritability using diffusion tensor imaging in neonatal brains. *Twin Res. Hum. Genet.* 15:336–350.
- Glaser, C., E. Lattka, P. Rzehak, C. Steer, and B. Koletzko. 2011. Genetic variation in polyunsaturated fatty acid metabolism and its potential relevance for human development and health. *Matern. Child Nutr.* 7(Suppl 2):27–40.
- Goubau, C., J. Jaeken, E. N. Levchenko, C. Thys, M. Di Michele, G. A. Martens, et al. 2013. Homozygosity for aquaporin 7 G264V in three unrelated children with hyperglyceroluria and a mild platelet secretion defect. *Genet. Med.* 15:55–63.
- Gray, E., M. Ginty, K. Kemp, N. Scolding, and A. Wilkins. 2011. Peroxisome proliferator-activated receptor-alpha agonists protect cortical neurons from inflammatory mediators and improve peroxisomal function. *Eur. J. Neurosci.* 33:1421–1432.
- Groeschel, S., B. Vollmer, M. D. King, and A. Connelly. 2010. Developmental changes in cerebral grey and white matter volume from infancy to adulthood. *Int. J. Dev. Neurosci.* 28:481–489.
- Grosso, M., D. Ricci, L. Bassi, N. Merchant, V. Doria, T. Arichi, et al. 2014. Development of the optic radiations and visual function after premature birth. *Cortex* 56:30–37.
- Guitart, M., O. Osorio-Conles, T. Pentinat, J. Cebria, J. Garcia-Villoria, D. Sala, et al. 2014. Fatty acid transport protein 1 (FATP1) localizes in mitochondria in mouse skeletal muscle and regulates lipid and ketone body disposal. *PLoS One* 9:e98109.
- Hack, M. 2009. Adult outcomes of preterm children. *J. Dev. Behav. Pediatr.* 30:460–470.
- Haynes, R. L., R. D. Folkerth, R. J. Keefe, I. Sung, L. I. Swzeda, P. A. Rosenberg, et al. 2003. Nitrosative and oxidative injury to premyelinating oligodendrocytes in periventricular leukomalacia. *J. Neuropathol. Exp. Neurol.* 62:441–450.
- Heise, V., N. Filippini, K. P. Ebmeier, and C. E. Mackay. 2011. The APOE varepsilon4 allele modulates brain white matter integrity in healthy adults. *Mol. Psychiatry* 16:908–916.
- Houdou, S., S. Takashima, and Y. Suzuki. 1993. Immunohistochemical expression of peroxisomal enzymes in developing human brain. *Mol. Chem. Neuropathol.* 19:235–248.
- Huppi, P. S., S. E. Maier, S. Peled, G. P. Zientara, P. D. Barnes, F. A. Jolesz, et al. 1998. Microstructural development of human newborn cerebral white matter assessed in vivo by diffusion tensor magnetic resonance imaging. *Pediatr. Res.* 44:584–590.
- Innis, S. M. 2007. Fatty acids and early human development. *Early Hum. Dev.* 83:761–766.
- Jablonska, B., J. Scafidi, A. Aguirre, F. Vaccarino, V. Nguyen, E. Borok, et al. 2012. Oligodendrocyte regeneration after neonatal hypoxia requires FoxO1-mediated p27Kip1 expression. *J. Neurosci.* 32:14775–14793.
- Jensen, H. A., L. B. Harslof, M. S. Nielsen, L. B. Christensen, C. Ritz, K. F. Michaelsen, et al. 2014. FADS single-

- nucleotide polymorphisms are associated with behavioral outcomes in children, and the effect varies between sexes and is dependent on PPAR genotype. *Am. J. Clin. Nutr.* 100:826–832.
- Jing, L., L. Su, and B. Z. Ring. 2014. Ethnic background and genetic variation in the evaluation of cancer risk: a systematic review. *PLoS One* 9:e97522.
- Johnson, S., and D. Wolke. 2013. Behavioural outcomes and psychopathology during adolescence. *Early Hum. Dev.* 89:199–207.
- Kajantie, E., and P. Hovi. 2014. Is very preterm birth a risk factor for adult cardiometabolic disease? *Semin. Fetal Neonatal Med.* 19:112–117.
- Khatchadourian, A., S. D. Bourque, V. R. Richard, V. I. Titorenko, and D. Maysinger. 2012. Dynamics and regulation of lipid droplet formation in lipopolysaccharide (LPS)-stimulated microglia. *Biochim. Biophys. Acta* 1821:607–617.
- Kinney, H. C., B. A. Brody, A. S. Kloman, and F. H. Gilles. 1988. Sequence of central nervous system myelination in human infancy. II. Patterns of myelination in autopsied infants. *J. Neuropathol. Exp. Neurol.* 47:217–234.
- van Kooij, B. J., L. S. de Vries, G. Ball, I. C. van Haastert, M. J. Benders, F. Groenendaal, et al. 2012. Neonatal tract-based spatial statistics findings and outcome in preterm infants. *AJNR Am. J. Neuroradiol.* 33:188–194.
- Kozusko, K., V. H. Tsang, W. Bottomley, Y. H. Cho, S. Gandotra, M. Mimmack, et al. 2015. Clinical and molecular characterization of a novel PLIN1 frameshift mutation identified in patients with familial partial lipodystrophy. *Diabetes* 64:299–310.
- Krishnan, M. L., L. E. Dyet, J. P. Boardman, O. Kapellou, J. M. Allsop, F. Cowan, et al. 2007. Relationship between white matter apparent diffusion coefficients in preterm infants at term-equivalent age and developmental outcome at 2 years. *Pediatrics* 120:e604–e609.
- Lebeck, J. 2014. Metabolic impact of the glycerol channels AQP7 and AQP9 in adipose tissue and liver. *J. Mol. Endocrinol.* 52:R165–R178.
- Lenroot, R. K., N. Gogtay, D. K. Greenstein, E. M. Wells, G. L. Wallace, L. S. Clasen, et al. 2007. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *NeuroImage* 36:1065–1073.
- Leviton, A., P. Gressens, O. Wolkenhauer, and O. Dammann. 2015. Systems approach to the study of brain damage in the very preterm newborn. *Front. Syst. Neurosci.* 9:58.
- Li, G., L. Wang, F. Shi, A. E. Lyall, W. Lin, J. H. Gilmore, et al. 2014. Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. *J. Neurosci.* 34:4228–4238.
- Ludwig, A., P. Uvarov, S. Soni, J. Thomas-Crusells, M. S. Airaksinen, and C. Rivera. 2011. Early growth response 4 mediates BDNF induction of potassium chloride cotransporter 2 transcription. *J. Neurosci.* 31:644–649.
- March of Dimes, P., Save the Children, WHO. 2012. Born Too Soon: The Global Action Report on Preterm Birth.
- Martinelli, N., D. Girelli, G. Malerba, P. Guarini, T. Illig, E. Trabetti, et al. 2008. FADS genotypes and desaturase activity estimated by the ratio of arachidonic acid to linoleic acid are associated with inflammation and coronary artery disease. *Am. J. Clin. Nutr.* 88:941–949.
- Matsufuji, T., M. Ikeda, A. Naito, M. Hirouchi, H. Takakusa, S. Kanda, et al. 2012. Discovery and optimization of novel fatty acid transport protein 1 (FATP1) inhibitors. *Bioorg. Med. Chem. Lett.* 22:5067–5070.
- Matsumata, M., H. Inada, and N. Osumi. 2014. Fatty acid binding proteins and the nervous system: their impact on mental conditions. *Neurosci. Res.* doi: 10.1016/j.neures.2014.08.012
- Mayer, S. I., O. G. Rossler, T. Endo, P. Charnay, and G. Thiel. 2009. Epidermal-growth-factor-induced proliferation of astrocytes requires Egr transcription factors. *J. Cell Sci.* 122:3340–3350.
- Mengozi, M., I. Cervellini, P. Villa, Z. Erbayraktar, N. Gokmen, O. Yilmaz, et al. 2012. Erythropoietin-induced changes in brain gene expression reveal induction of synaptic plasticity genes in experimental stroke. *Proc. Natl Acad. Sci. USA* 109:9617–9622.
- Merino, D. M., H. Johnston, S. Clarke, K. Roke, D. Nielsen, A. Badawi, et al. 2011. Polymorphisms in FADS1 and FADS2 alter desaturase activity in young Caucasian and Asian adults. *Mol. Genet. Metab.* 103:171–178.
- Miller, D. J., T. Duka, C. D. Stimpson, S. J. Schapiro, W. B. Baze, M. J. McArthur, et al. 2012. Prolonged myelination in human neocortical evolution. *Proc. Natl Acad. Sci. USA* 109:16480–16485.
- Minghetti, L., R. Salvi, M. Lavinia Salvatori, M. A. Ajmone-Cat, C. De Nuccio, S. Visentin, et al. 2014. Nonenzymatic oxygenated metabolites of alpha-linolenic acid B1- and L1-phytoprostanes protect immature neurons from oxidant injury and promote differentiation of oligodendrocyte progenitors through PPAR-gamma activation. *Free Radic. Biol. Med.* 73:41–50.
- Mitchell, R. W., N. H. On, M. R. Del Bigio, D. W. Miller, and G. M. Hatch. 2011. Fatty acid transport protein expression in human brain and potential role in fatty acid transport across human brain microvessel endothelial cells. *J. Neurochem.* 117:735–746.
- Moore, T., E. M. Hennessy, J. Myles, S. J. Johnson, E. S. Draper, K. L. Costeloe, et al. 2012. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. *BMJ* 345: e7961.
- Neil, J. J., S. I. Shiran, R. C. McKinstry, G. L. Scheff, A. Z. Snyder, C. R. Almli, et al. 1998. Normal brain in human newborns: apparent diffusion coefficient and diffusion anisotropy measured by using diffusion tensor MR imaging. *Radiology* 209:57–66.

- Nie, J., G. Li, and D. Shen. 2013. Development of cortical anatomical properties from early childhood to early adulthood. *NeuroImage* 76:216–224.
- Nosarti, C., A. Reichenberg, R. M. Murray, S. Cnattingius, M. P. Lambe, L. Yin, et al. 2012. Preterm birth and psychiatric disorders in young adult life. *Arch. Gen. Psychiatry* 69:E1–E8.
- Ovaska, K., M. Laakso, and S. Hautaniemi. 2008. Fast gene ontology based clustering for microarray experiments. *BioData Min.* 1:11.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.
- Quintanilla, R. A., E. Utreras, and F. A. Cabezas-Opazo. 2014. Role of PPAR gamma in the differentiation and function of neurons. *PPAR Res.* 2014:768594.
- Robinson, S., Q. Li, A. Dechant, and M. L. Cohen. 2006. Neonatal loss of gamma-aminobutyric acid pathway expression after human perinatal brain injury. *J. Neurosurg.* 104:396–408.
- Robinson, S. W., M. Fernandes, and H. Husi. 2014. Current advances in systems and integrative biology. *Comput. Struct. Biotechnol. J.* 11:35–46.
- Roider, H. G., A. Kanhere, T. Manke, and M. Vingron. 2007. Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics* 23:134–141.
- Rosen, E. D., and B. M. Spiegelman. 2001. PPARgamma: a nuclear regulator of metabolism, differentiation, and cell growth. *J. Biol. Chem.* 276:37731–37734.
- Ryckman, K. K., K. S. Borowski, N. I. Parikh, and A. F. Saftlas. 2013. Pregnancy complications and the risk of metabolic syndrome for the offspring. *Curr. Cardiovasc. Risk Rep.* 7:217–223.
- Sergeant, S., C. E. Hugenschmidt, M. E. Rudock, J. T. Ziegler, P. Ivester, H. C. Ainsworth, et al. 2012. Differences in arachidonic acid levels and fatty acid desaturase (FADS) gene variants in African Americans and European Americans with diabetes or the metabolic syndrome. *Br. J. Nutr.* 107:547–555.
- Shen, K. K., S. Rose, J. Fripp, K. L. McMahon, G. I. de Zubicaray, N. G. Martin, et al. 2014. Investigating brain connectivity heritability in a twin study using diffusion imaging data. *NeuroImage* 100:628–641.
- Silver, M., G. Montana, and Alzheimer's Disease Neuroimaging I. 2012b. Fast identification of biological pathways associated with a quantitative trait using group lasso with overlaps. *Stat. Appl. Genet. Mol. Biol.*, 11:Article 7.
- Silver, M., E. Janousova, X. Hua, P. M. Thompson, G. Montana, and Alzheimer's Disease Neuroimaging I. 2012a. Identification of gene pathways implicated in Alzheimer's disease using longitudinal imaging phenotypes with sparse regression. *NeuroImage*, 63:1681–1694.
- Silver, M., P. Chen, R. Li, C. Y. Cheng, T. Y. Wong, E. S. Tai, et al. 2013. Pathways-driven sparse regression identifies pathways and genes associated with high-density lipoprotein cholesterol in two Asian cohorts. *PLoS Genet.* 9:e1003939.
- Smith, S. M., M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, et al. 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage* 31:1487–1505.
- Smith, S. M., and T. E. Nichols. 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* 44:83–98.
- Smyser, T. A., C. D. Smyser, C. E. Rogers, S. K. Gillespie, T. E. Inder, and J. J. Neil. 2015. Cortical gray and adjacent white matter demonstrate synchronous maturation in very preterm infants. *Cereb. Cortex.* doi: 10.1093/cercor/bhv164
- Sordillo, J. E., S. Sharma, A. Poon, J. Lasky-Su, K. Belanger, D. K. Milton, et al. 2011. Effects of endotoxin exposure on childhood asthma risk are modified by a genetic polymorphism in ACAA1. *BMC Med. Genet.* 12:158.
- Speed, D., T. J. O'Brien, A. Palotie, K. Shkura, A. G. Marson, D. J. Balding, et al. 2014. Describing the genetic architecture of epilepsy through heritability analysis. *Brain*, 137:2680–2689.
- Tanaka, S., T. Mito, and S. Takashima. 1995. Progress of myelination in the human fetal spinal nerve roots, spinal cord and brainstem with myelin basic protein immunohistochemistry. *Early Hum. Dev.* 41:49–59.
- Toro, R., J. B. Poline, G. Huguette, E. Loth, V. Frouin, T. Banaschewski, et al. 2014. Genomic architecture of human neuroanatomical diversity. *Mol. Psychiatry.* doi: 10.1038/mp.2014.99.
- Trzaskowski, M., J. Yang, P. M. Visscher, and R. Plomin. 2014. DNA evidence for strong genetic stability and increasing heritability of intelligence from age 7 to 12. *Mol. Psychiatry* 19:380–384.
- Vannucci, S. J., and H. Hagberg. 2004. Hypoxia-ischemia in the immature brain. *J. Exp. Biol.* 207:3149–3154.
- Vinall, J., R. E. Grunau, R. Brant, V. Chau, K. J. Poskitt, A. R. Synnes, et al. 2013. Slower postnatal growth is associated with delayed cerebral cortical maturation in preterm newborns. *Sci. Transl. Med.*, 5:168ra8.
- Volpe, J. J. 2009. Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. *Lancet Neurol.* 8:110–124.
- Volpe, J. J. 2012. Neonatal encephalopathy: an inadequate term for hypoxic-ischemic encephalopathy. *Ann. Neurol.* 72:156–166.
- Wang, Z., and G. Montana. 2014. The graph-guided group lasso for genome-wide association studies. Pp. 131–153 *in* J. A. K. Suykens, M. Signoretto, and A. Argyrou, eds. *Regularization, optimization, kernels, and support vector machines.* CRC Press, Boca Raton, FL.

- Wang, K., M. Li, and H. Hakonarson. 2010. Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* 11:843–854.
- Wang, J., D. Duncan, Z. Shi, and B. Zhang. 2013. WEB-based gene set analysis toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* 41:W77–W83.
- Warde-Farley, D., S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, et al. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38:W214–W220.
- Williams, J. G., J. P. Higgins, and C. E. Brayne. 2006. Systematic review of prevalence studies of autism spectrum disorders. *Arch. Dis. Child.* 91:8–15.
- Wimberger, D. M., T. P. Roberts, A. J. Barkovich, L. M. Prayer, M. E. Moseley, and J. Kucharczyk. 1995. Identification of “premyelination” by diffusion-weighted MRI. *J. Comput. Assist. Tomogr.* 19:28–33.
- World Health Organization, W.H.O. 2014. Preterm birth. In: Organization, W.H., editor.
- Yakovlev, P. I., and A. R. LeCours. 1967. The myelogenetic cycles of regional maturation of the brain. Pp 3–70 *in* A. Minkowski, ed. *Regional development of the brain in early life*. Blackwell, Oxford.
- Yamada, K., D. J. Gerber, Y. Iwayama, T. Ohnishi, H. Ohba, T. Toyota, et al. 2007. Genetic analysis of the calcineurin pathway identifies members of the EGR gene family, specifically EGR3, as potential susceptibility candidates in schizophrenia. *Proc. Natl Acad. Sci. USA* 104:2815–2820.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–569.
- Zhong, H., J. Beaulaurier, P. Y. Lum, C. Molony, X. Yang, D. J. Macneil, et al. 2010. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet.* 6:e1000932.
- Zilles, K., E. Armstrong, K. H. Moser, A. Schleicher, and H. Stephan. 1989. Gyrification in the cerebral cortex of primates. *Brain Behav. Evol.* 34:143–150.

Supporting Information

Additional supporting information may be found in the online version of this article:

Figure S1. Scree plots of PCA for TBSS phenotype adjusted for PMA (left) and GA plus PMA (right).

Figure S2. Frequency density distributions of residuals in TBSS phenotypes adjusted for PMA, GA and PMA, or GA, PMA, and ethnicity.

Figure S3. First two components from principal component analysis of population stratification based on pairwise identity by state (IBS).

Table S4. Mapping of SNPs to genes and pathways.

Figure S5. Null selection frequencies for all KEGG pathways in the PsRRR model.

Table S6. PsRRR pathway rankings with null and empirical selection frequencies adjusted for PMA.

Table S7. PsRRR pathway rankings with null and empirical selection frequencies adjusted for GA and PMA.

Table S8. GGGL-1 gene selection frequencies >0.4.

Table S9. GGGL-2 top thirty SNPs.