OPEN

Citation: Transl Psychiatry (2017) 7, e1188; doi:10.1038/tp.2017.159

www.nature.com/tp

ORIGINAL ARTICLE

Genome-wide association analysis identifies common variants influencing infant brain volumes

K Xia[1], J Zhang[2], M Ahn[2,9], S Jha[3], JJ Crowley[1,4,5], J Szatkiewicz[1,4], T Li[6], F Zou[2,10], H Zhu[2,6], D Hibar[7,11], P Thompson[7], ENIGMA Consortium, PF Sullivan[1,4,8], M Styner[1], JH Gilmore[1] and RC Knickmeyer[1]

Genome-wide association studies (GWAS) of adolescents and adults are transforming our understanding of how genetic variants impact brain structure and psychiatric risk, but cannot address the reality that psychiatric disorders are unfolding developmental processes with origins in fetal life. To investigate how genetic variation impacts prenatal brain development, we conducted a GWAS of global brain tissue volumes in 561 infants. An intronic single-nucleotide polymorphism (SNP) in *IGFBP7* (rs114518130) achieved genome-wide significance for gray matter volume ($P = 4.15 \times 10^{-10}$). An intronic SNP in *WWOX* (rs10514437) neared genome-wide significance for white matter volume ($P = 1.56 \times 10^{-8}$). Additional loci with small *P*-values included psychiatric GWAS associations and transcription factors expressed in developing brain. Genetic predisposition scores for schizophrenia and ASD, and the number of genes impacted by rare copy number variants (CNV burden) did not predict global brain tissue volumes. Integration of these results with large-scale neuroimaging GWAS in adolescents (PNC) and adults (ENIGMA2) suggests minimal overlap between common variants impacting brain volumes at different ages. Ultimately, by identifying genes contributing to adverse developmental phenotypes, it may be possible to adjust adverse trajectories, preventing or ameliorating psychiatric and developmental disorders.

Translational Psychiatry (2017) **7,** e1188; doi:10.1038/tp.2017.159; published online 1 August 2017

INTRODUCTION

Advances in genomics are providing new insights into the biology of psychiatric disorders.[1] At the same time, advances in magnetic resonance imaging (MRI) are transforming our understanding of the structural and functional organization of the human brain.[2] Classic twin studies, as well as genome-wide association studies (GWAS), show that the structure of the human brain is strongly shaped by genetics.[3,4] Altered brain structure has been reported in individuals with psychiatric disorders[5] and in asymptomatic individuals with a family history,[6] suggesting that genetic influences on disease risk are partly mediated by changes in neuroanatomy. It is thus somewhat surprising that large-scale imaging genetics studies have thus far shown little overlap between genome-wide significant loci associated with brain structure and those associated with psychiatric disorders.[4,7,8] One potential explanation for this discrepancy is that most imaging-genetic studies use data collected in adolescence or adulthood. As such they cannot address the reality that many psychiatric disorders are unfolding processes with roots in prenatal and early postnatal development.[9,10]

The prenatal–early postnatal period represents the foundational phase of human brain development and relies on precise spatiotemporal regulation of gene expression.[11] Many genes associated with mental illness show elevated expression in early life,[10,12] and early aberrations in neurodevelopment relevant to psychiatric disorders can be captured via MRI as evidenced by studies of infants at high familial risk for schizophrenia or autism spectrum disorders (ASD).[13,14] In addition, twin studies indicate infant brain volumes are highly heritable.[15] Consequently, imaging-genetic studies in infants could provide new insights into the origins of psychiatric disorders. Furthermore, the infant brain is highly plastic and a promising target for therapeutic interventions. To investigate how genetic variation impacts prenatal and early postnatal brain development, we conducted a GWAS of global brain tissue volumes in a unique cohort of infants who received high-resolution MRI scans of the brain around 5 weeks of age.

MATERIALS AND METHODS

Subjects

A total of 561 infants (300 male, 261 female) between 0 and 24 weeks of age, encompassing 295 singletons or unpaired twins, 17 sibling pairs and 232 twins (61 same-sex dizygotic (DZ) pairs, 37 same-sex monozygotic (MZ) pairs and 18 opposite-sex DZ pairs). Overall, 63% of subjects are European ancestry; remaining subjects are primarily African ancestry. See Supplementary Methods and Supplementary Table 1 for additional information.

[1]Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA; [2]Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA; [3]Curriculum in Neurobiology, University of North Carolina, Chapel Hill, NC, USA; [4]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; [5]Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden; [6]Department of Biostatistics, MD Andersen Cancer Center, University of Texas, Houston, TX, USA; [7]Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, USC Keck School of Medicine, University of Southern California, Marina del Rey, CA, USA and [8]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden. Correspondence: Dr RC Knickmeyer, Department of Psychiatry, University of North Carolina, 343 Medical Wings C, Campus Box #7160, Chapel Hill, NC 27599-7160, USA.
E-mail: rebecca_knickmeyer@med.unc.edu
[9]Current address: Department of Mathematics and Statistics, University of Nevada, Reno, NV, USA.
[10]Current address: Department of Biostatistics, University of Florida, Gainesville, FL, USA.
[11]Current address: Janssen Pharmaceutical Companies of Johnson & Johnson, San Diego, CA, USA.
Received 23 May 2017; accepted 1 June 2017

Image acquisition and analysis

T1- and T2-weighted scans were acquired with either a Siemens Allegra head-only 3 T scanner ($N = 512$) or a Siemens TIM Trio 3T scanner ($N = 49$) (Siemens Medical Supplies, Erlangen, Germany). See Supplementary Methods for acquisition parameters. ANOVA performed in 561 unrelated subjects indicated platform (Allegra or Trio) had a significant impact on gray matter volume and a marginal effect on intracranial volume (ICV). Acquisition parameters within each platform did not significantly affect brain volumes. Platform was included as a covariate in all analyses. Automatic segmentation of brain tissue into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) was performed using an atlas-based expectation-maximization segmentation algorithm using T1- and T2-weighted images as in ref. 16. ICV represents the sum of GM, WM, and CSF.

Genome-wide genotyping

Genotyping of single-nucleotide polymorphisms (SNPs) was carried out using Affymetrix Axiom Genome-Wide LAT and Exome arrays. Samples were randomized across 96-well plates. Each plate contained a common control sample. Population stratification was assessed using principle component analysis (PCA).[17] See Supplementary Methods for additional details. Imputation was performed with MACH-Admix[18] using the 1000 Genomes Project (1000G) reference panel (phase1_release_v3.20101123).[19,20] To evaluate the quality of imputed SNPs, we computed mean R^2 for varying minor allele frequency (MAF) categories and R^2 cutoffs (Supplementary Table 2).[21] We retained SNPs with average $R^2 > 0.8$ and excluded SNPs with MAF < 0.01.

Statistical methods

We investigated associations between 8 762 422 SNPs and four infant brain volume measures (ICV, GM, WM and CSF). The first three genotypic principal components were included as covariates as was scanner type. ICV was a covariate for GM, WM and CSF. Additional covariates were selected via adaptive lasso from a comprehensive set of demographic and medical history variables (Supplementary Methods). For ICV and WM, we included birth weight, gestational age at birth, sex and age at MRI. For GM, we included birth weight, gestational age at birth and age at MRI. For CSF, we included age at MRI.

To account for subject correlation among MZ and DZ twins, we applied ACE-based linear mixed effects models.[21] The ACE model not only distinguishes correlation among MZ and DZ twins, it also includes common environmental effect as well as random effect in the linear mixed effect model with restricted maximum likelihood (REML) for optimization. Association analysis was performed with two-sided likelihood ratio tests. *P*-values $< 1.25 \times 10^{-8}$ were considered genome-wide significant (Bonferroni correction for four MRI phenotypes, which is from the conventional GWAS threshold of 5×10^{-8}[ref. 22] divided by four). The suggestive significance level for this correction was $P = 1.25 \times 10^{-6}$, which is derived from the conventional suggestive GWAS threshold of 5×10^{-6} [ref. 23] divided by four. The phenotypic variance explained by each SNP was calculated as $h^2 = \beta'^2 \mathrm{var}(X)/(\beta'^2 \mathrm{var}(X) + \sigma_a^2 + \sigma_c^2 + \sigma_e^2)$, where β' is the adjusted effect size of the SNP through an approximate conditional-likelihood approach to correct for the overestimation of effect size caused by winner's curse,[24] var(X) is the variance of SNP genotype, and σ_a^2, σ_c^2 and σ_e^2 are the variance explained by shared genetic effect, shared environment effect and random effect, respectively. We simulated statistical power under a series of major genetic effect β values with varying minor allele frequencies. We set the variance explained by genetic effect to 0.2 and the random error variation σ_e^2 to 1. Sample size is sufficient to discover a genetic effect on the order of ~8% with 80% power (Supplementary Figure 1).

We ran multiple sensitivity analyses to evaluate the robustness of our findings. We examined top SNPs in individuals with European ancestry (based on PCA, $N = 356$), individuals with non-European (primarily African) ancestry (based on PCA, $N = 205$), individuals born after 32 weeks gestation ($N = 530$), a subsample excluding offspring of mothers with a confirmed diagnosis of schizophrenia or bipolar disorder ($N = 526$), a subsample excluding children who spent more than 24 hours in neonatal intensive care ($N = 427$), a subsample consisting only of unrelated individuals ($N = 423$), and a subsample including only the two most common image acquisition sequences ($N = 538$). We also ran a sensitivity analysis using the full sample, but a reduced set of covariates. As some of our phenotypes were moderately correlated (Supplementary Table 3), we also applied a PCA analysis to the brain volumes and used the top three PCs as GWAS phenotypes. Finally, to ensure the validity of our most significant GWAS findings, we performed permutation tests for the two most significant SNPs (Supplementary Methods).

We calculated fixation index (F_{ST}) for all linkage disequilibrium (LD)-independent SNPs with *P*-values $< 1.25 \times 10^{-6}$ to further evaluate potential confounding by population stratification. We checked major and minor allele frequencies within our sample using maternal-reported and genetic ancestry (from PCA), and cross-checked our data with allele frequencies provided through the 1000 Genomes Project (1000G) (phase I, version 3).[25]

After performing LD-pruning (--indep-pairwise 50 5 0.5: 50 bp moving window at step of 5 bp with r^2 threshold of 0.5 in PLINK, see Supplementary Methods), we tested for enrichment of prenatally expressed genes as described in ref. 26 using increasingly liberal *P*-value thresholds. This approach down-weights noisy signal from null SNPs that are invariably incorporated, especially at large *P*-value thresholds, while capturing true, subtle signals that increase power for association.[27,28] Each SNP was assigned to the nearest gene within 2 megabases. Elevated prenatal expression was defined as a fetal/postnatal log2 fold-change of 0.5 or greater. Analyses were performed using both the microarray-based 'BrainCloud' dorsolateral prefrontal cortex transcriptome[29] and the RNA sequencing-based Brainspan transcriptome of 15 brain regions.[30]

Using two-sided exact binomial tests (sign test) after LD-pruning, we compared our data to two large-scale neuroimaging GWAS conducted in older individuals: ENIGMA2,[31] a primarily adult sample, and the Philadelphia Neurodevelopmental Cohort (PNC),[32] a primarily adolescent sample (See Supplementary Methods for details). We used the—score function in PLINK and LD-pruned genotype data from our neonate cohort to generate polygenic scores for ICV, GM and WM at nine different *P*-value thresholds (ambiguous SNPs were not included). These 'neonatal scores' were used to predict the equivalent phenotype in the PNC cohort (covarying for age, sex and first three ancestry PCs). In a similar manner, we generated polygenic scores for ICV, GM and WM using the PNC cohort and used these 'adolescent scores' to predict the equivalent phenotype in neonates. Covariates were the same as in univariate neonate analyses. Finally, we generated polygenic scores for ICV using ENIGMA2 data. Note that data on GM and WM is not available for ENIGMA2. This 'adult ICV score' was used to predict ICV in the neonate cohort and PNC cohorts. Only summary-level GWAS data was available for ENIGMA2; consequently adult ICV never served as a dependent variable. The same procedure was used to generate genetic predisposition scores for schizophrenia[33] and ASD.[34] Resulting risk scores were used as predictors for neonatal brain volumes.

Gene set-enrichment analysis (GSEA) was performed on each brain volume GWAS with MAGMA.[35] We tested candidate pathways implicated in psychiatric disorders including glutamatergic postsynaptic proteins comprising activity-regulated cytoskeleton-associated protein (ARC)[36–38] and *N*-methyl-D-aspartate receptor (NMDAR) complexes,[38,39] genes whose mRNAs bind to fragile X mental retardation protein (FMRP),[36,37,39,40,41] calcium channels[34,40,42] and genes involved in histone methylation.[43] We focused on pathways supported by convergent evidence from genome-wide screens for common and exonic variation and CNVs. Given our univariate results, we also tested genes with *REST* binding sites[44] and Rbfox targets. For the latter analysis, we used genes with RBFOX2 count > 4 or summed RBFOX1 and RBFOX3 > 12.[45]

PennCNV and Partek Genomics Suite 6.6 were used for CNV calling from signal intensity files of the Affymetrix Axiom Genotype array.[46] Included CNVs had SNP coverage > 15, length > 100 kb, and for PennCNV, confidence score $\geqslant 10$. We excluded CNVs with $> 50\%$ overlap with common CNVs (freq $> 1\%$ in EUR population[47] or our population) or regions of known rearrangement (chr14:105573485–107197774 and chr22:22473003–23188329). CNVs occuring in both PennCNV and Partek were considered validated CNVs. See Supplementary Appendix 1 for details. Linear mixed effects models were used to test whether CNV burden (number of CNV-affected genes per subject) predicted ICV, GM, WM and CSF. Covariates were the same as in univariate analyses. Finally, we determined whether any one in the sample carried a rare CNV that overlapped with known neuropsychiatric CNVs (Supplementary Table 4).

Experiments were undertaken with the understanding and written consent of each subject's mother or father, with the approval of the Institutional Review Board of the University of North Carolina (UNC) School of Medicine.

Table 1. LD-independent SNPs with *P*-values smaller than 1.25×10^{-6}

Tissue volume	*Chr*	*Best SNP*	*A1*	*A2*	*MAF*	*Effect size*	*Likelihood ratio*	*P-value*	*Variance explained*	*Diff./allele*	*Closest protein coding gene (±500 kb)*
ICV	15	rs8030297	T	G	0.07	−32 231	28.46	9.56E−08	2.47	−6.73	*KLF13*
	20	rs4809613	A	G	0.17	−19 053	24.38	7.91E−07	0.36	−3.98	*EYA2*
	3	rs9831671	T	G	0.27	20 450	24.03	9.49E−07	0.43	4.27	*LMCD1*
WM	16	rs10514437	G	A	0.03	−6617	31.97	1.56E−08	4.14	−3.76	*WWOX*
	20	rs117068934	G	A	0.03	−5988	26.34	2.86E−07	1.01	−3.41	*SLC13A3*
	6	rs9361254	G	T	0.13	−3106	24.10	9.16E−07	0.31	−1.77	*MEI4*[a], *HTR1B*
	18	rs138549714	G	GT	0.07	−3608	23.96	9.85E−07	0.29	−2.05	*MYOM1*
	12	rs11834561	G	A	0.38	2071	23.92	1.00E−06	0.29	1.18	*USP44*
	21	rs200193477	A	AT	0.26	3577	23.66	1.15E−06	0.27	2.03	—
GM	4	rs114518130[b]	G	A	0.01	14 973	39.04	4.15E−10	6.52	6.13	*IGFBP7*
	7	rs7786147	C	T	0.11	−4135	28.10	1.15E−07	2.27	−1.69	*MPLKIP*[c]
	3	rs200355458	C	CG	0.09	4657	27.74	1.39E−07	1.86	1.91	*OPA1*
	8	rs2565117	C	A	0.01	−10 939	26.33	2.88E−07	1.12	−4.48	*ZMAT4*
	18	rs138087875	A	AC	0.06	4789	25.78	3.83E−07	0.60	1.96	*CNDP1*
	16	rs12919005	C	G	0.03	6973	25.55	4.31E−07	0.51	2.85	*TOX3*
	9	rs77196186	C	T	0.05	−4970	24.77	6.45E−07	0.40	−2.03	*CAAP1*
	16	rs9889138	G	A	0.07	4482	24.39	7.86E−07	0.34	1.83	*C16orf62*
	14	rs1092055	T	C	0.45	2142	24.28	8.32E−07	0.34	0.88	*SLC35F4*
	10	rs11012877	G	A	0.11	3683	24.23	8.54E−07	0.33	1.51	*CACNB2*
	1	rs2992041	C	T	0.43	2516	24.05	9.37E−07	0.32	1.03	*CTH*
	16	rs7206373	G	C	0.46	2084	23.77	1.09E−06	0.29	0.85	*SEC14L5*
CSF	16	rs7195044	C	G	0.09	5619	26.55	2.56E−07	1.33	9.54	*RBFOX1*
	17	rs11867179	C	G	0.24	3042	25.43	4.6E−07	0.53	5.17	*TBX4*
	4	rs113668862	C	T	0.08	−4951	24.95	5.9E−07	0.43	−8.41	*FRG1*
	18	rs11875537	T	G	0.44	2559	24.35	8.05E−07	0.37	4.35	*METTL4*

Abbreviations: CSF, cerebrospinal fluid; GM, gray matter; ICV, intracranial volume; LD, linkage disequalibrium; MAF, minor allele frequency; SNP, single-nucleotide polymorphism; WM, white matter. [a]Provisional or inferred status in RefSeq (next closest protein coding gene within 500 kb is also listed). [b]Genome-wide significant ($P < 1.25 \times 10^{-8}$), Bonferroni correction of 5×10^{-8} for four phenotypes. [c]Also known as c7orf11. Note that all SNPs are imputed with the exception of rs10514437; see Supplementary Figure 2 for a genotyping cluster plot of rs10514437. A1 is the effective allele and A2 is the reference allele. Diff./allele is the ratio of effect size over mean brain volume. We define LD-independent SNPs as those with low LD ($r^2 < 0.1$) to a more significantly associated SNP within a 500 kb window.

Code availability

Computer code used to generate study results can be found at http://www.bios.unc.edu/research/genomic_software/CorrMeta/ace/.

RESULTS

Table 1 displays all LD-independent SNPs with *P*-values $< 1.25 \times 10^{-6}$. Quantile–quantile plots are displayed in Supplementary Figure 2. Locus zoom plots are displayed in Supplementary Figures 3–6. Lambda is 1.01 for ICV and WM, 1.00 for GM and 0.98 for CSF. Lambda-1000 is 1.01 for ICV, WM and GM, and 0.96 for CSF. An intronic SNP in *IGFBP7* (rs114518130) achieved genome-wide significance for GM ($P = 4.15 \times 10^{-10}$, Figure 1 and Supplementary Figure 5). An intronic SNP in *WWOX* (rs10514437) fell just short of genome-wide significance for WM ($P = 1.56 \times 10^{-8}$, Figure 1 and Supplementary Figures 4 and 7). Additional loci nearing genome-wide significance were located near (within 500 kb) transcriptional regulators expressed in developing brain, psychiatric GWAS associations, and genes involved in neurotransmitter signaling and neural development (Supplementary Table 5). LD-independent SNPs with *P*-values $< 1.25 \times 10^{-6}$ were associated with absolute percent differences of 0.85–9.54% in volume per risk allele, explaining 0.27–6.52% of the phenotypic variance (after correction for 'winner's curse'). Effect sizes in primary analyses were highly similar to effect sizes in sensitivity analyses (Supplementary Table 6) with the exception of rs113668862 (association with CSF, $P = 5.9 \times 10^{-7}$), which showed no effect on CSF in non-European populations. Similar genes emerged when using the top three PCs from a PCA analysis of the brain volumes as GWAS phenotypes (Supplementary Table 7).

Permutation testing yielded an empirical *P*-value of 8×10^{-8} for rs10514437's association with WM and an empirical *P*-value of 1×10^{-8} for rs114518130's association with GM. Low empirical *P*-values suggest these findings are reliable and robust.

F_{ST} values > 0.05, indicating potential confounding by ancestry, were observed for rs9831671 (association with ICV, $P = 9.49 \times 10^{-7}$), rs7786147 (association with GM, $P = 1.15 \times 10^{-7}$), rs2565117 (association with GM, $P = 2.88 \times 10^{-7}$) and rs113668862 (association with CSF, $P = 5.9 \times 10^{-7}$). Although F_{ST} values were below 0.05 for our top hits in *IGFBP7* and *WWOX*, we note that the minor (A) allele for the *IGFBP7* SNP is very rare in populations of European ancestry (Supplementary Table 8). We were unable to run sensitivity analysis in individuals of European ancestry as the minor allele did not occur in this group.

Three genes, *TOX3, FRG1* and *c7orf11* (*MPLKIP*) show elevated prenatal expression across multiple brain regions (Supplementary Appendix 2; Supplementary Table 9). *KLF13, WWOX, SLC13A3, HTR1B, MYOM1, ZMAT4, CNDP1, CACNB2* and *SEC14L5* show a strong increase in expression from prenatal to postnatal life. *EYA2, USP44, OPA1, SLC35F4, CTH, RBFOX1* and *METTL4* show elevated prenatal expression in a small subset of brain regions. Examination of exon level expression data from Brainspan suggests *METTL4*, *C16orf62* and *RBFOX1* have specific isoforms with differentially regulated expression across the lifespan (Supplementary Appendix 3). Tests for enrichment of prenatally expressed genes were significant at multiple thresholds across multiple brain regions (Figure 2).

Comparing our neonate cohort to the adolescent cohort, sign tests for WM were significant at thresholds of $P < 0.001$, $P < 0.01$ and $P < 0.1$ (Figure 3). Percent overlap is small with 53.8, 51.2 and 50.6% of SNPs showing the same direction of effect versus 50% expected by chance. Comparing our neonate cohort to the adult cohort, the sign test for ICV is significant at the threshold of $P < 0.1$ with 50.40% overlap. Comparing the adolescent cohort to

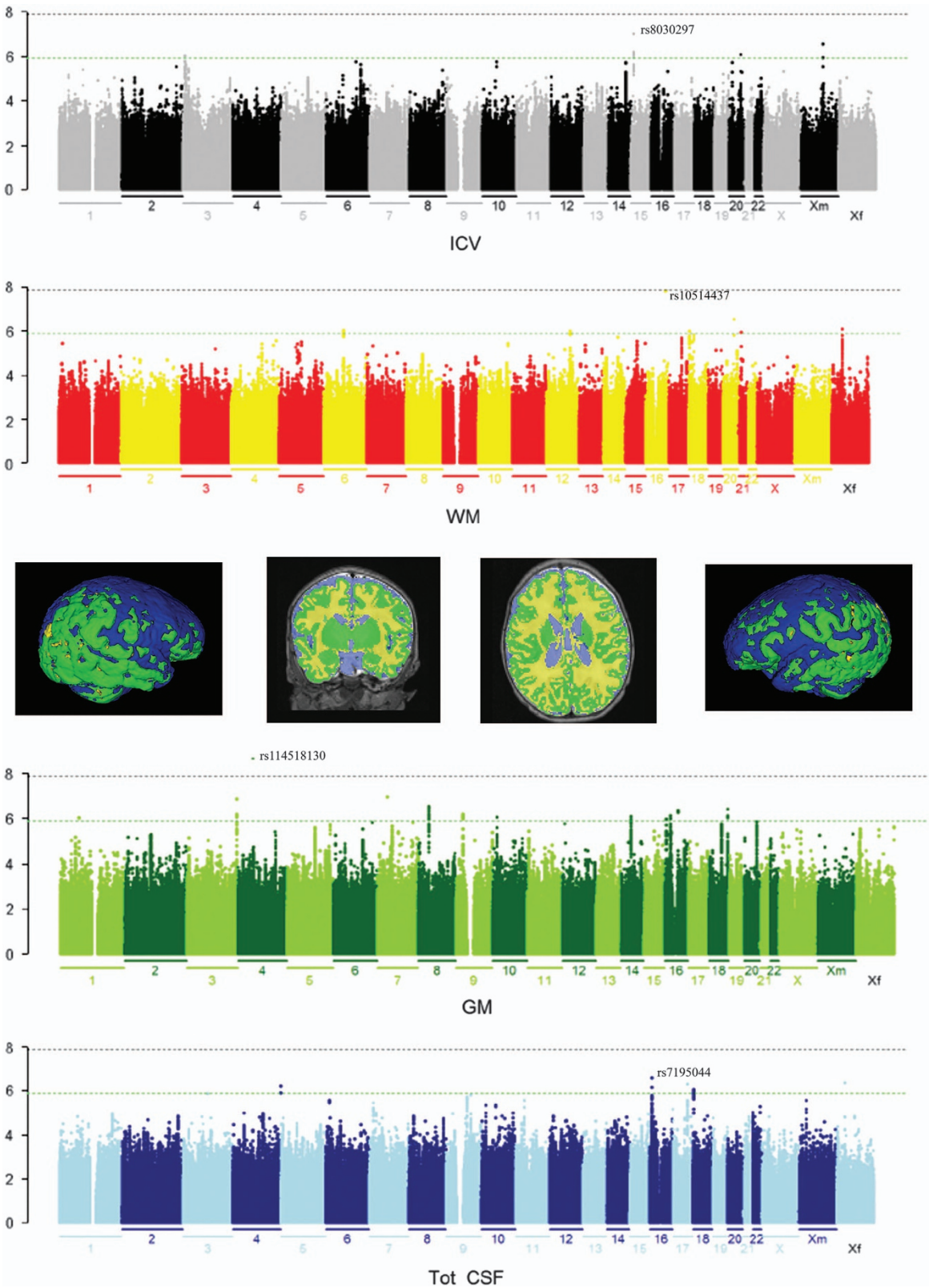

Figure 1. Common genetic variants associated with infant brain volumes. Manhattan plots colored with a scheme that matches the corresponding tissue volume (middle) are shown (yellow is white matter (WM); green is gray matter (GM); blue is cerebrospinal fluid (CSF)). Genome-wide significance is shown for the threshold of $P = 1.25 \times 10^{-8}$ (gray dotted line). The green line indicates the threshold of $P = 1.25 \times 10^{-6}$. The most significant single-nucleotide polymorphism (SNP) for each phenotype is labeled.

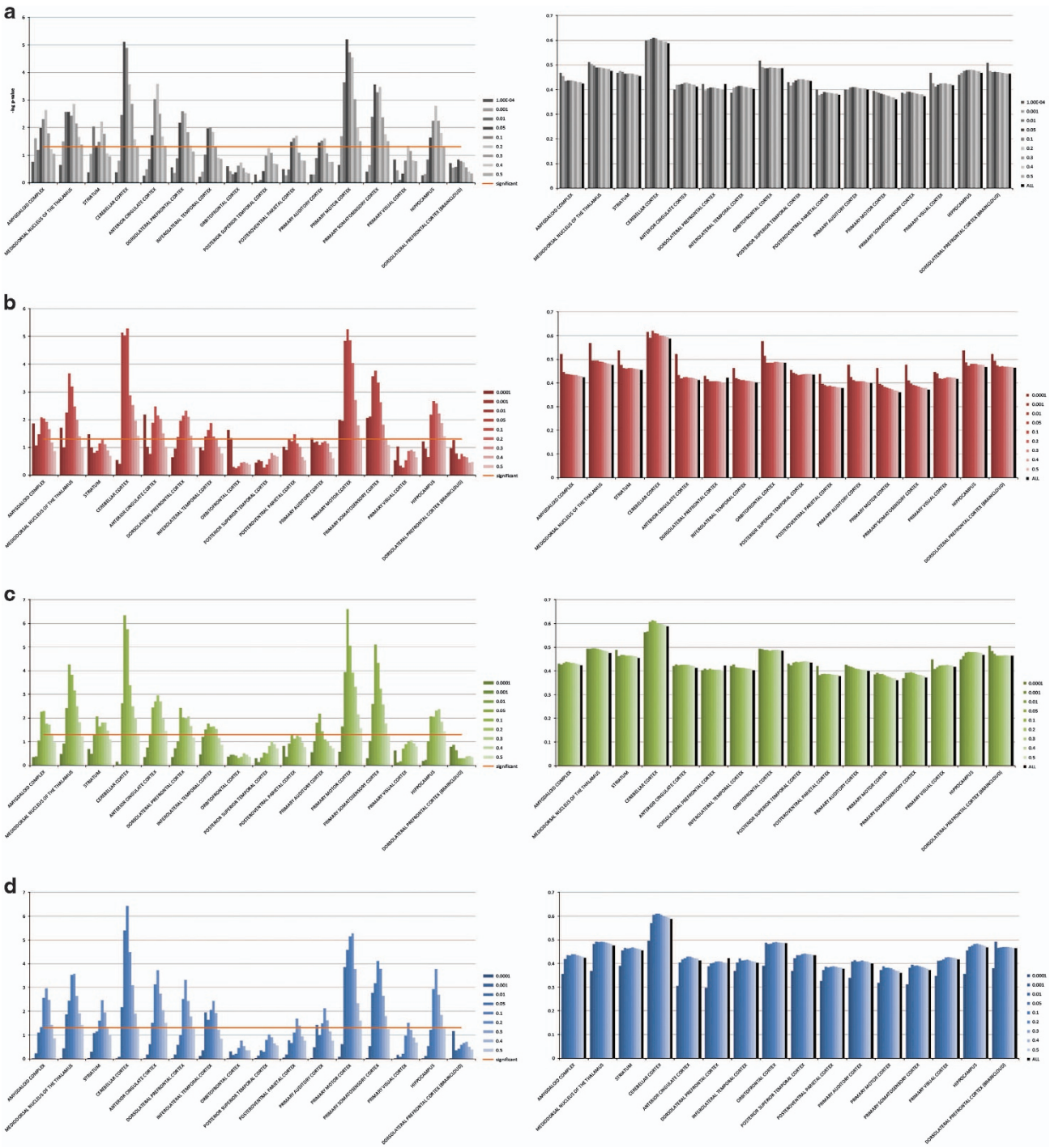

Figure 2. Enrichment of prenatally expressed genes at multiple *P*-value thresholds. After LD-pruning, each single-nucleotide polymorphism (SNP) was assigned to the nearest gene within 2 megabases. Elevated prenatal expression was defined as a fetal/postnatal log2 fold-change threshold of 0.5. Analyses were performed using both the microarray-based 'BrainCloud' dorsolateral prefrontal cortex (DLPFC) transcriptome[31] and the RNA sequencing-based Brainspan transcriptome of 15 different brain regions.[32] Left panels show $-\log_{10}$ *P*-values. Orange line represents the threshold for statistical significance. Right panels show percentage of genes with elevated prenatal expression. (**a**) intracranial volume (ICV). (**b**) white matter (WM). (**c**) gray matter (GM). (**d**) CSF.

the adult cohort, the sign test for ICV is significant at the threshold of $P < 0.1$ with 50.80% overlap.

Neonatal polygenic scores for WM showed positive associations with adolescent WM at thresholds between $P < 0.01$ and $P < 0.5$ (Figure 4). Adolescent WM scores showed positive associations with neonatal WM at 7 of 9 *P*-value thresholds, and adolescent GM scores showed a positive association with neonatal GM at the threshold of $P < 0.1$. Adult polygenic scores for ICV did not predict neonatal ICV, but did predict adolescent ICV at 8 of 9 *P*-value thresholds. When comparing across ages, polygenic scores explained between 0.4 and 1.6% of variance after adjusting for covariates. Scatterplots of significant relationships can be found in

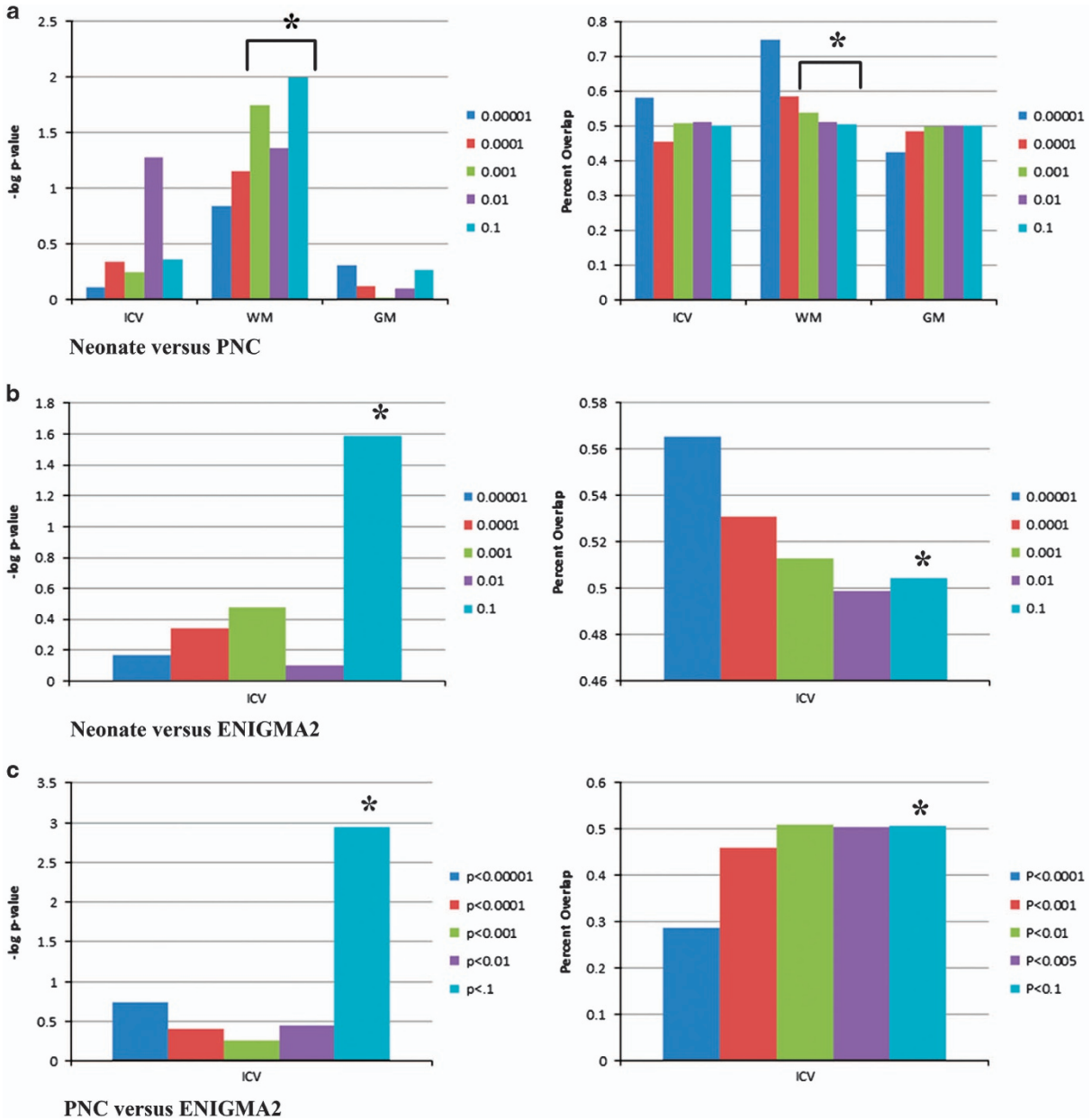

Figure 3. Sign tests suggest minimal overlap between common variants impacting brain volumes in neonates, adolescents and adults. (**a**) Left panel shows $-\log_{10}$ *P*-values for two-sided exact binomial tests comparing single-nucleotide polymorphisms (SNPs) impacting intracranial volume (ICV) and ICV-adjusted white matter (WM) and gray matter (GM) in neonates to the same SNPs in adolescents (PNC). Right panel shows percent overlap. Tests for ICV-adjusted WM are significant ($P = 0.02$, 0.04 and 0.01) at the thresholds of $P < 0.001$, $P < 0.01$, and $P < 0.1$, respectively. (**b**) Left panel shows $-\log_{10}$ *P*-values for two-sided exact binomial tests comparing SNPs impacting ICV in neonates to the same SNPs in adults (ENIGMA2). Right panel shows percent overlap. Test is significant at the threshold of $P < 0.1$ ($P = 0.02$). (**c**) Left panel shows $-\log_{10}$ *P*-values for two-sided exact binomial tests comparing SNPs impacting ICV in adolescents (PNC) to the same SNPs in adults (ENIGMA2). Right panel shows percent overlap. Test is significant at the threshold of $P < 0.1$ ($P = 0.001$). * indicates significance.

Supplementary Figure 8. To put these numbers in context, neonatal polygenic scores explain between 74 and 93% of variance in neonatal brain volumes after adjusting for covariates and adolescent polygenic scores explain between 67 and 93% of variance in adolescent brain volumes after adjusting for covariates.

Supplementary Table 10 lists all LD-independent SNPs with *P*-values $< 1.25 \times 10^{-6}$ in the neonate and compares them to equivalent SNPs and phenotypes in PNC and ENIGMA2. The SNP tagging *EYA2* is nominally significant in ENIGMA2 ($P = 0.048$) with the same direction of effect. No other overlapping SNPs are significant in PNC or ENIGMA2, but some of our top hits are not represented in PNC data as they failed imputation quality control. Previously published genome-wide significant hits for ICV[8,48,49] in adults did not predict ICV in infants (Supplementary Table 11), although we were unable to examine rs9303535 as it was not in the 1000G reference database and hence not imputed. No suitable proxy SNPs were available.

Genetic predisposition scores for schizophrenia and ASD did not predict neonatal brain volumes (Supplementary Figure 9). Regarding GSEA, Rbfox targets showed a nominally significant association with CSF, but no gene sets achieved significance after correction for multiple comparisons (Supplementary Table 12).

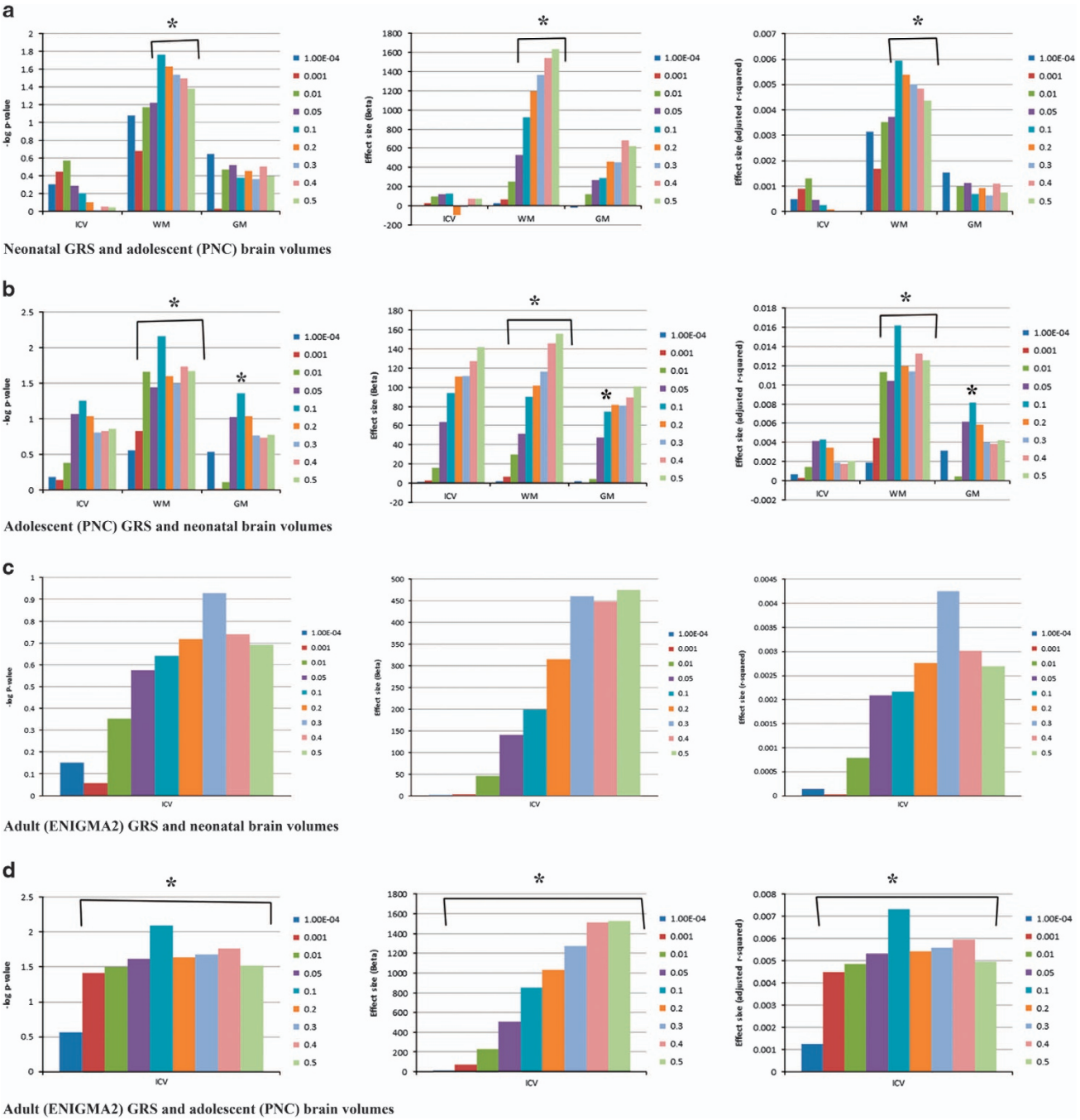

Figure 4. Polygenic scores for neonatal, adolescent and adult brain volumes suggest genetic determinants of global brain volumes are highly distinct at different ages. (**a**) Left panel shows $-\log_{10}$ *P*-values for the association between neonatal polygenic scores for intracranial volume (ICV), ICV-adjusted white matter (WM), and ICV-adjusted gray matter (GM) and the matching phenotype in adolescents (PNC). Middle panel shows the effect size and direction of effect (Beta). Right panel shows the r^2 value (percentage of variance explained after adjusting for covariates). (**b**) Left panel shows $-\log_{10}$ *P*-values for the association between adolescent polygenic scores for ICV, ICV-adjusted WM, and ICV-adjusted GM and the matching phenotype in this neonatal cohort. Middle panel shows the effect size and direction of effect (Beta). Right panel shows the r^2 value (percentage of variance explained after adjusting for covariates). (**c**) Left panel shows $-\log_{10}$ *P*-values for the association between adult (ENIGMA2) polygenic scores for ICV and ICV in this neonatal cohort. Middle panel shows the effect size and direction of effect (Beta). Right panel shows the r^2 value (percentage of variance explained after adjusting for covariates). (**d**) Left panel shows $-\log_{10}$ *P*-values for the association between adult (ENIGMA2) polygenic scores for ICV and ICV in adolescents (PNC). Middle panel shows the effect size and direction of effect (Beta). Right panel shows the r^2 value (percentage of variance explained after adjusting for covariates).* indicates significance.

CNV burden was not significantly associated with any of the neuroimaging traits examined (Supplementary Figure 10). Four individuals carried a rare CNV, which overlapped with a known neuropsychiatric CNV. Percent overlap ranged from 0.27 to 96.29%. All CNVs affected at least one gene also present in the known neuropsychiatric CNV. Two of these children had a history of psychiatric illness in one or both parents (attention deficit disorder or anxiety disorder). Review of medical records indicated

that one child had delayed speech and fine motor skills at 2 years of age. Long-term follow-up data was not available. Examination of individual value plots suggested these individuals had brain volumes in the normal range at this age (Supplementary Figure 11).

DISCUSSION

The prenatal and early postnatal period represents a critical window in brain development with long-term implications for cognition[50,51] and psychiatric risk.[9,10] In this first GWAS of infant neuroimaging phenotypes, an intronic SNP in *IGFBP7* (rs114518130) achieved genome-wide significance for GM. An intronic SNP in *WWOX* (rs10514437) fell just short of genome-wide significance for WM. Although *IGFBP7* has been implicated in learning and memory,[52] it is perhaps even more intriguing that this SNP is within 100 kb of *REST*, which encodes a master negative regulator of neurogenesis.[44] Experimental manipulation of the *REST/NRSF-SWI/SNF* chromatin remodeling complex alters expression of several GWAS-supported schizophrenia genes.[53] However, GSEA analysis of *REST* targets did not reach significance, somewhat diminishing the case that *REST* is the causal element in this association. Mutations in *WWOX* are associated with brain phenotypes including autosomal recessive spinocerebellar ataxia-12 (SCAR12; MIM #614322), epilepsy, mental retardation and microcephaly. MRI studies are limited, but hypomyelination, hypoplasia of the corpus callosum and posterior white matter hyperintensities have been reported. In adult human brains, levels of the protein encoded by *WWOX* (WOX1) are high in neuronal axons and nerve bundles such as corpus callosum and striatal fascicles,[54] consistent with our observed association with WM.

Many top associations tag transcriptional regulators expressed in developing brain (*KLF13, LMCD1, TOX3* and *TBX4*), though none are known eQTLs. Other genes with strong biological plausibility include *CACNB2*, a calcium channel subunit which has emerged in schizophrenia GWAS,[33] *HTR1B*, which encodes the 5-hydroxytryptamine (serotonin) receptor 1B, and *RBFOX1*, an RNA splicing factor, which regulates expression of large genetic networks during early neuronal development.[55] Most of these genes show a strong increase in expression from the prenatal to the postnatal period, whereas *TOX3* shows elevated prenatal expression. Examination of exon level expression data suggests the first several exons of *RBFOX1* show elevated prenatal expression. When examining more permissive *P*-value thresholds, there was evidence of significant prenatal expression enrichment for all phenotypes examined.

We compared our results to two large-scale neuroimaging GWAS, one focused on adolescents (PNC)[32] and one on adults (ENIGMA2).[31] These comparisons do not represent a conventional replication sample. Instead, they represent a test of whether there are shared genetic determinants of global brain volumes at differing ages. Sign tests and polygenic score analyses suggest genetic variants influencing neonatal WM have detectable effects in adolescence. Polygenic score analyses suggest genetic variants influencing ICV in adults have detectable effects in adolescence. However, percent overlap in sign tests was minimal and the percent of variance explained by polygenic scores was small. The overall pattern of results suggests genetic determinants of global brain volumes are highly distinct at different ages. Variants associated with neonatal brain volumes are likely located in genes involved with foundational prenatal processes including neurogenesis, neuronal migration and differentiation, programmed cell death, dendritogenesis and axonogenesis.[56] Variants associated with adolescent brain volumes are likely located in genes involved in postnatal neurodevelopmental processes including synaptogenesis during early childhood and synaptic/dendritic pruning during early adolescence. Variants associated with adult brain volumes likely represent synaptic/dendritic pruning in late adolescence and perhaps neurodegenerative events associated with aging.[56]

Genetic predisposition scores for schizophrenia and ASD did not predict global brain volumes in neonates. Results are similar to a recent study, which found no evidence of genetic overlap between schizophrenia risk and ICV in adults.[7] To our knowledge, this is the first study to apply a similar approach for ASD risk. In addition, although rare genic CNVs are increased in ASD[57] and schizophrenia,[58] CNV burden did not predict global brain volumes in neonates. Null findings should be considered in light of several qualifiers. First, genetic variants associated with disease may influence neurodevelopmental processes occurring at a resolution not captured by MRI. Second, other neuroimaging phenotypes may be more relevant to ASD and schizophrenia (for example, cortical thickness and surface area, ventricular volume, diffusion tensor imaging or functional activity/connectivity).[13,59–63] Third, longitudinal change in neuroimaging phenotypes may be more relevant to ASD and schizophrenia than cross-sectional measures.[64] Finally, although many psychiatric risk genes show elevated expression in early life[10,12] differences in global brain volumes may not arise until later in life. Regarding ASD, 6-month-old infants at high familial risk do not have larger brain volumes than low risk infants, despite reports of generalized enlargement in older children with ASD.[65] Regarding schizophrenia, neonates at high familial risk do have larger GM volume than controls, but this affect is restricted to males.[14] Future studies with larger samples should address how genetic predisposition and CNV burden influence neurodevelopmental trajectories across infancy and childhood, as well as possible interactions with sex. Larger samples would also allow stable estimates of SNP-sense heritability to be calculated.

In conclusion, we have shown that common genetic variation influences infant brain volumes. Several individual loci identified in this study have clear relevance to intellectual disability and mental illness, although genetic predisposition scores for schizophrenia and ASD did not predict variation in our sample. To our knowledge, this is the first study of its kind, consequently independent replication is critical. This is particularly true for the intronic SNP in *IGFBP7,* which has a low minor allele frequency, was imputed, and was not present in the largest racial/ethnic group in the study. Nevertheless, the current results suggest this area of research will be fruitful. Ultimately, by identifying genes contributing to developmental phenotypes, it may be possible to develop treatments that normalize adverse trajectories, preventing the onset of psychiatric disorders or reducing their severity.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (MH064065, MH070890 and HD053000 to JHG, MH083045 to RCK, SES-1357666, DMS-1407655, UL1 RR025747-03 and MH086633-02 to FZ, HD-003110 and EB005149-01 to MS and NS007431 to SJ), the Brain & Behavior Research Foundation, and the John and Polly Sparks Foundation, Bank of America, Trustee (NARSAD Young Investigator Grant to KX). DH and PT were supported in part by an ENIGMA Consortium Grant (U54 EB020403 to PMT) from the NIH Institutes contributing to the Big Data to Knowledge (BD2K) Initiative, including the NIBIB and NCI. Support for the collection of the PNC data sets was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. All PNC subjects were recruited through the Center for Applied Genomics at The Children's Hospital in Philadelphia. The current study also makes use of data generated by the DECIPHER Consortium. A full list of centers who contributed to the generation of the data is available from https://decipher.sanger.ac.uk/ and via email from decipher@sanger.ac.uk. Funding for DECIPHER was provided by the Wellcome Trust. We thank the participating families that made this project possible as well as the staff of the UNC MRI Research Center, the UNC Neuro Image Research and Analysis Laboratories, the UNC Early Brain

Development Program and the UNC Biospecimen Processing Core (BSP), including Joseph Blocher, Dianne Evans and Patricia Basta.

REFERENCES

1 Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* 2010; **68**: 182–186.
2 Paus T. Population neuroscience: why and how. *Hum Brain Mapp* 2010; **31**: 891–903.
3 Blokland GA, de Zubicaray GI, McMahon KL, Wright MJ. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Res Hum Genet* 2012; **15**: 351–371.
4 Strike LT, Couvy-Duchesne B, Hansell NK, Cuellar-Partida G, Medland SE, Wright MJ. Genetics and brain morphology. *Neuropsychol Rev* 2015; **25**: 63–96.
5 van Erp TG, Hibar DP, Rasmussen JM, Glahn DC, Pearlson GD, Andreassen OA *et al.* Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol Psychiatry* 2016; **21**: 547–553.
6 Bois C, Whalley HC, McIntosh AM, Lawrie SM. Structural magnetic resonance imaging markers of susceptibility and transition to schizophrenia: a review of familial and clinical high risk population studies. *J Psychopharmacol* 2015; **29**: 144–154.
7 Franke B, Stein JL, Ripke S, Anttila V, Hibar DP, van Hulzen KJ *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nat Neurosci* 2016; **19**: 420–431.
8 Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM *et al.* Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 2012; **44**: 552–561.
9 Fatemi SH, Folsom TD. The neurodevelopmental hypothesis of schizophrenia, revisited. *Schizophr Bull* 2009; **35**: 528–548.
10 Chen JA, Penagarikano O, Belgard TG, Swarup V, Geschwind DH. The emerging picture of autism spectrum disorder: genetics and pathology. *Annu Rev Pathol* 2015; **10**: 111–144.
11 Silbereis JC, Pochareddy S, Zhu Y, Li M, Sestan N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron* 2016; **89**: 248–268.
12 Birnbaum R, Jaffe AE, Chen Q, Hyde TM, Kleinman JE, Weinberger DR. Investigation of the prenatal expression patterns of 108 schizophrenia-associated genetic loci. *Biol Psychiatry* 2014; **77**: e43–e51.
13 Wolff JJ, Gu H, Gerig G, Elison JT, Styner M, Gouttard S *et al.* Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. *Am J Psychiatry* 2012; **169**: 589–600.
14 Gilmore JH, Kang C, Evans DD, Wolfe HM, Smith JK, Lieberman JA *et al.* Prenatal and neonatal brain structure and white matter maturation in children at high risk for schizophrenia. *Am J Psychiatry* 2010; **167**: 1083–1091.
15 Gilmore JH, Schmitt JE, Knickmeyer RC, Smith JK, Lin W, Styner M *et al.* Genetic and environmental contributions to neonatal brain structure: a twin study. *Hum Brain Mapp* 2010; **31**: 1174–1182.
16 Prastawa M, Gilmore JH, Lin W, Gerig G. Automatic segmentation of MR images of the developing newborn brain. *Med Image Anal* 2005; **9**: 457–466.
17 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
18 Liu EY, Li M, Wang W, Li Y. MaCH-admix: genotype imputation for admixed populations. *Genet Epidemiol* 2013; **37**: 25–37.
19 Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56–65.
20 Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–1073.
21 Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K *et al.* Heritability and genomics of gene expression in peripheral blood. *Nat Genet* 2014; **46**: 430–437.
22 Barsh GS, Copenhaver GP, Gibson G, Williams SM. Guidelines for genome-wide association studies. *PLoS Genet* 2012; **8**: e1002812.
23 Reed E, Nunez S, Kulp D, Qian J, Reilly MP, Foulkes AS. A guide to genome-wide association analysis and post-analytic interrogation. *Stat Med* 2015; **34**: 3769–3792.
24 Ghosh A, Zou F, Wright FA. Estimating odds ratios in genome scans: an approximate conditional likelihood approach. *Am J Hum Genet* 2008; **82**: 1064–1074.
25 Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO *et al.* A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.
26 Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR. Prenatal expression patterns of genes associated with neuropsychiatric disorders. *Am J Psychiatry* 2014; **171**: 758–767.
27 Iyegbe C, Campbell D, Butler A, Ajnakina O, Sham P. The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research. *Soc Psychiatry Psychiatr Epidemiol* 2014; **49**: 169–182.
28 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
29 Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 2011; **478**: 519–523.
30 Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M *et al.* Spatio-temporal transcriptome of the human brain. *Nature* 2011; **478**: 483–489.
31 Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivieres S, Jahanshad N *et al.* Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**: 224–229.
32 Satterthwaite TD, Connolly JJ, Ruparel K, Calkins ME, Jackson C, Elliott MA *et al.* The Philadelphia Neurodevelopmental Cohort: a publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage* 2016; **124**(Pt B): 1115–1119.
33 PGC. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
34 PGC. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**: 1371–1379.
35 de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *Plos Comput Biol* 2015; **11**: e1004219.
36 Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**: 185–190.
37 Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.
38 Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* 2012; **17**: 142–153.
39 Darnell JC, Van Driesche SJ, Zhang C, Hung KY, Mele A, Fraser CE *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 2011; **146**: 247–261.
40 Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM *et al.* Copy number variation in schizophrenia in Sweden. *Mol Psychiatry* 2014; **19**: 762–773.
41 Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**: 285–299.
42 Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
43 Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 2015; **18**: 199–209.
44 Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein–DNA interactions. *Science* 2007; **316**: 1497–1502.
45 Weyn-Vanhentenryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z *et al.* HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep* 2014; **6**: 1139–1152.
46 Szatkiewicz JP, Neale BM, O'Dushlaine C, Fromer M, Goldstein JI, Moran JL *et al.* Detecting large copy number variants using exome genotyping arrays in a large Swedish schizophrenia sample. *Mol Psychiatry* 2013; **18**: 1178–1184.
47 Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet* 2015; **16**: 172–183.
48 Ikram MA, Fornage M, Smith AV, Seshadri S, Schmidt R, Debette S *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat Genet* 2012; **44**: 539–544.
49 Adams HH, Hibar DP, Chouraki V, Stein JL, Nyquist PA, Renteria ME *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci* 2016; **19**: 1569–1582.
50 Short SJ, Elison JT, Goldman BD, Styner M, Gu H, Connelly M *et al.* Associations between white matter microstructure and infants' working memory. *Neuroimage* 2013; **64**: 156–166.
51 Alcauter S, Lin W, Smith JK, Short SJ, Goldman BD, Reznick JS *et al.* Development of thalamocortical connectivity during infancy and its cognitive correlations. *J Neurosci* 2014; **34**: 9067–9075.
52 Agis-Balboa RC, Arcos-Diaz D, Wittnam J, Govindarajan N, Blom K, Burkhardt S *et al.* A hippocampal insulin-growth factor 2 pathway regulates the extinction of fear memories. *EMBO J* 2011; **30**: 4071–4083.
53 Loe-Mie Y, Lepagnol-Bestel AM, Maussion G, Doron-Faigenboim A, Imbeaud S, Delacroix H *et al.* SMARCA2 and other genome-wide supported

schizophrenia-associated genes: regulation by REST/NRSF, network organization and primate-specific evolution. *Hum Mol Genet* 2010; **19**: 2841–2857.
54 Chiang MF, Chen ST, Lo CP, Sze CI, Chang NS, Chen YJ. Expression of WW domain-containing oxidoreductase WOX1 in human nervous system tumors. *Anal Cell Pathol* 2013; **36**: 133–147.
55 Bill BR, Lowe JK, Dybuncio CT, Fogel BL. Orchestration of neurodevelopmental programs by RBFOX1: implications for autism spectrum disorder. *Int Rev Neurobiol* 2013; **113**: 251–267.
56 Stiles J, Jernigan TL. The basics of brain development. *Neuropsychol Rev* 2010; **20**: 327–348.
57 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al.* Functional impact of global rare copy number variation in autism spectrum conditions. *Nature* **466**: 368–372.
58 Stone JL, O'Donovan MC, Gurling H, Kirov GK, Blackwood DHR, Corvin A *et al.* Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**: 237–241.
59 Minzenberg MJ, Laird AR, Thelen S, Carter CS, Glahn DC. Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Arch Gen Psychiatry* 2009; **66**: 811–822.
60 Goghari VM, Truong W, Spilka MJ. A magnetic resonance imaging family study of cortical thickness in schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 2015; **168**: 660–668.
61 Zanetti MV, Schaufelberger MS, Doshi J, Ou Y, Ferreira LK, Menezes PR *et al.* Neuroanatomical pattern classification in a population-based sample of first-episode schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 2013; **43**: 116–125.
62 Karlsgodt KH. Diffusion imaging of white matter in schizophrenia: progress and future directions. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2016; **1**: 209–217.
63 Bernhardt BC, Di Martino A, Valk SL, Wallace GL. Neuroimaging-based phenotyping of the autism spectrum. *Curr Top Behav Neurosci* 2016; **30**: 341–355.
64 Wolff JJ, Piven J. Neurodevelopmental disorders: accelerating progress in autism through developmental research. *Nat Rev Neurol* 2014; **10**: 431–432.
65 Hazlett HC, Gu H, McKinstry RC, Shaw DW, Botteron KN, Dager SR *et al.* Brain volume findings in 6-month-old infants at high familial risk for autism. *Am J Psychiatry* 2012; **169**: 601–608.

Supplementary Information accompanies the paper on the *Translational Psychiatry* website (http://www.nature.com/tp)