

Genome-wide association study of 23,500 individuals identifies 7 loci associated with brain ventricular volume

Dina Vojinovic, Hieab H. Adams, Xueqiu Jian, Qiong Yang, Albert Vernon Smith, Joshua C. Bis, Alexander Teumer, Markus Scholz, Nicola J. Armstrong, Edith Hofer, Yasaman Saba, Michelle Luciano, Manon Bernard, Stella Trompet, Jingyun Yang, Nathan A. Gillespie, Sven J. van der Lee, Alexander Neumann, Shahzad Ahmad, Ole A. Andreassen, David Ames, Najaf Amin, Konstantinos Arfanakis, Mark E. Bastin, Diane M. Becker, Alexa S. Beiser, Frauke Beyer, Henry Brodaty, R. Nick Bryan, Robin Bülow, Anders M. Dale, Philip L. De Jager, Ian J. Deary, Charles DeCarli, Debra A. Fleischman, Rebecca F. Gottesman, Jeroen van der Grond, Vilmundur Gudnason, Tamara B. Harris, Georg Homuth, David S. Knopman, John B. Kwok, Cora E. Lewis, Shuo Li, Markus Loeffler, Oscar L. Lopez, Pauline Maillard, Hanan El Marroun, Karen A. Mather, Thomas H. Mosley, Ryan Muetzel, Matthias Nauck, Paul A. Nyquist, Matthew S. Panizzon, Zdenka Pausova, Bruce M. Psaty, Ken Rice, Jerome I. Rotter, Natalie Royle, Claudia L. Satizabal, Reinhold Schmidt, Peter R. Schofield, Pamela J. Schreiner, Stephen Sidney, David J. Stott, Anbupalam Thalamuthu, Andre G. Uitterlinden, Maria C. Valdés Hernández, Meike W. Vernooij, Wei Wen, Tonya White, A. Veronica Witte, Katharina Wittfeld, Margaret J. Wright, Lisa R. Yanek, Henning Tiemeier, William S. Kremen, David A. Bennett, J. Wouter Jukema, Tomas Paus, Joanna M. Wardlaw, Helena Schmidt, Perminder S. Sachdev, Arno Villringer, Hans Jörgen Grabe, WT Longstreth, Cornelia M. van Duijn, Lenore J. Launer, Sudha Seshadri, M. Arfan Ikram, Myriam Fornage

This chapter is accepted for publication in Nature Communications.

The supplemental information for this paper is available at https://drive.google.com/drive/folders/1-2X-Nx3tdaeX4EO_bAWANKo7X_umEx2X?usp=sharing

ABSTRACT

The volume of the lateral ventricles (LV) increases with age and their abnormal enlargement is a key feature of several neurological and psychiatric diseases. Although lateral ventricular volume is heritable, a comprehensive investigation of its genetic determinants is lacking. In this meta-analysis of genome-wide association studies of 23,533 healthy middle-aged to elderly individuals from 26 population-based cohorts, we identify 7 genetic loci associated with LV volume. These loci map to chromosomes 3q28, 7p22.3, 10p12.31, 11q23.1, 12q23.3, 16q24.2, and 22q13.1 and implicate pathways related to tau pathology, S1P signaling, and cytoskeleton organization. We also report a significant genetic overlap between the thalamus and LV volumes ($\rho_{\text{genetic}} = -0.59$, $p\text{-value} = 3.14 \times 10^{-6}$), suggesting that these brain structures may share a common biology. These genetic associations of LV volume provide insights into brain morphology.

INTRODUCTION

The volume of lateral ventricles increases in normal aging.¹⁻⁴ The enlargement of lateral ventricles has also been suggested in various complex neurological disorders such as Alzheimer's disease, vascular dementia and Parkinson's disease⁵⁻⁸ as well as psychiatric disorders such as schizophrenia and bipolar disorder.⁹⁻¹¹ Furthermore, ventricular enlargement has been associated with poor cognitive functioning and cerebral small vessel disease pathology.¹²⁻¹⁴ Even though it might be intuitive to interpret ventricular expansion primarily as an indicator of brain shrinkage after the onset of the disorder, recent studies have provided evidence against this notion.^{15,16} The size of lateral ventricles' is influenced by genetic factors with heritability estimated to be 54%, on average,¹⁶ but changing with age, from 32-35% in childhood to about 75% in late middle and older age.¹⁶ Even though the size of surrounding gray matter structures is also heritable,¹⁷⁻¹⁹ ventricular volume is reported to be genetically independent of other brain regions surrounding the ventricles.²⁰ Similarly, ventricular enlargement in schizophrenia does not appear to be linked to volume reduction in the surrounding structures.¹⁵

Elucidating the genetic contribution to inter-individual variation in lateral ventricular volume can thus provide important insights and better understanding of the complex genetic architecture of brain structures and related neurological and psychiatric disorders. Candidate gene studies have identified single nucleotide polymorphisms (SNPs) mapping to Catechol-O-Methyltransferase (*COMT*) and Neuregulin 1 (*NRG1*) genes as associated with larger lateral ventricular volume in patients with the first episode of non-affective psychosis.^{21,22} However, a comprehensive investigation of the genetic determinants of lateral ventricular volume is lacking.

Here, we perform a genome-wide association (GWA) meta-analysis of 23,533 middle-aged to elderly individuals from population-based cohorts participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium in order to identify common genetic variants that influence lateral ventricle volume. We apply a commonly used two-stage GWA design followed by a joint analysis approach that combines information across the stages and provides greater power.²³ We identify 7 genetic loci associated with lateral ventricular volume and report genome-wide overlap with thalamus volume.

RESULTS

Genome-wide association results

The overview of study design is illustrated in **Supplementary Fig. 1**. The GWA results from 12 studies were combined in stage 1 and subsequently evaluated in an independent sample from 14 studies in stage 2. Finally, the results of stage 1 and stage 2 analyses were combined in stage 3. Detailed information on study participants, image acquisition and genotyping is provided in **Supplementary Note 1** and **Supplementary Data 1-3**.

The results of the stage 1 meta-analysis ($N = 11,396$) are illustrated in **Supplementary Fig. 2**. The quantile-quantile plot suggests that potential population stratification and/or cryptic relatedness are well controlled after genomic correction ($\lambda = 1.04$) (**Supplementary Fig. 2, Supplementary Table 1**). The stage 1 meta-analysis identified 146 significant variant associations, mapping to three chromosomal regions at 3q28, 7p22.3, and 16q24.2 (**Table 1**). All 146 stage 1 significant associations replicated in the stage 2 meta-analysis ($N = 12,137$) with the same direction of effect at Bonferroni adjusted significance ($p\text{-value} = 5 \times 10^{-3}$, **Supplementary Data 4**), except one SNP ($p\text{-value} = 7.6 \times 10^{-3}$). Subsequently, the results from all individual studies were combined in the stage 3 GWA meta-analysis ($N = 23,533$). The quantile-quantile plot showed again adequate control of population stratification or relatedness (**Supplementary Fig. 3**). The combined stage 3 GWA meta-analysis identified 314 additional significant associations mapping to four additional chromosomal regions at 10p12.31, 11q23.1, 12q23.3, and 22q13.1 (**Fig. 1-2, Table 1**). The effect size for the lead variant mapped to 10p12.31 locus was correlated with mean age of the cohort ($r = 0.50$, $p\text{-value} = 0.03$) (**Supplementary Fig. 4**). No correlation was found for the other lead variants (**Supplementary Fig. 5-10**).

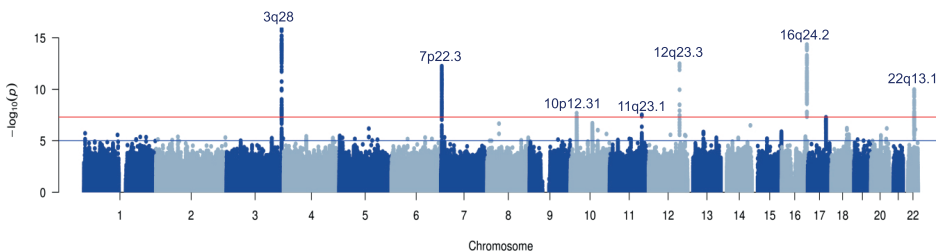


Figure 1. Manhattan plot for stage 3 genome-wide association meta-analysis. Each dot represents a variant. The plot shows $-\log_{10} p$ -values for all variants. Red line represents the genome-wide significance threshold ($p\text{-value} < 5 \times 10^{-8}$), whereas blue line denotes suggestive threshold ($p\text{-value} < 1 \times 10^{-5}$).

Table 1. Genome-wide significant results from the meta-analyses of lateral ventricular volume. Variant that showed the lowest *p*-value in the fixed effect sample-size weighted Z-score meta-analysis for each locus is shown.

SNP	Chr	Annotation	Gene(s)	A1/A2	Stage 1		Stage 2		Stage 3 combined	
					Zscore	P	Zscore	P	Zscore	P
rs34113929*	3q28	intergenic	<i>SNAR-I,OSTN</i>	A/G	-6.84	7.70E-12	-5.05	4.44E-07	-8.27	1.37E-16
rs9937293*	16q24.2	intergenic	<i>FOXL1,C16orf95</i>	A/G	5.65	1.63E-08	5.61	2.03E-08	7.84	4.45E-15
7:2760334-C-CT*	7p22.3	intergenic	<i>AMZ1,GNA12</i>	D/I	-5.88	4.21E-09	-4.48	7.34E-06	-7.21	5.61E-13
rs12146713	12q23.3	intronic	<i>NUAK1</i>	T/C	-5.01	5.57E-07	-5.44	5.32E-08	-7.28	3.25E-13
rs4820299	22q13.1	intronic	<i>TRIOBP</i>	T/C	-4.79	1.71E-06	-4.49	7.04E-06	-6.46	1.05E-10
rs35587371	10p12.31	intronic	<i>MLLT10</i>	A/T	-4.89	1.03E-06	-3.32	9.12E-04	-5.61	2.07E-08
rs7936534	11q23.1	intergenic	<i>ARHGAP20,C11orf53</i>	A/G	4.25	2.12E-05	3.71	2.04E-04	5.54	2.96E-08

Abbreviations: SNP - single nucleotide polymorphism; Chr - chromosome; A1/A2 - effect allele/other allele; Freq - Frequency of effect allele; Zscore - Z score from METAL; P - *p*-value;

*Variants that surpassed genome-wide significance threshold in stage 1 meta-analysis; remaining SNPs listed in the table reached genome-wide significance threshold in combined, stage 3, meta-analysis;



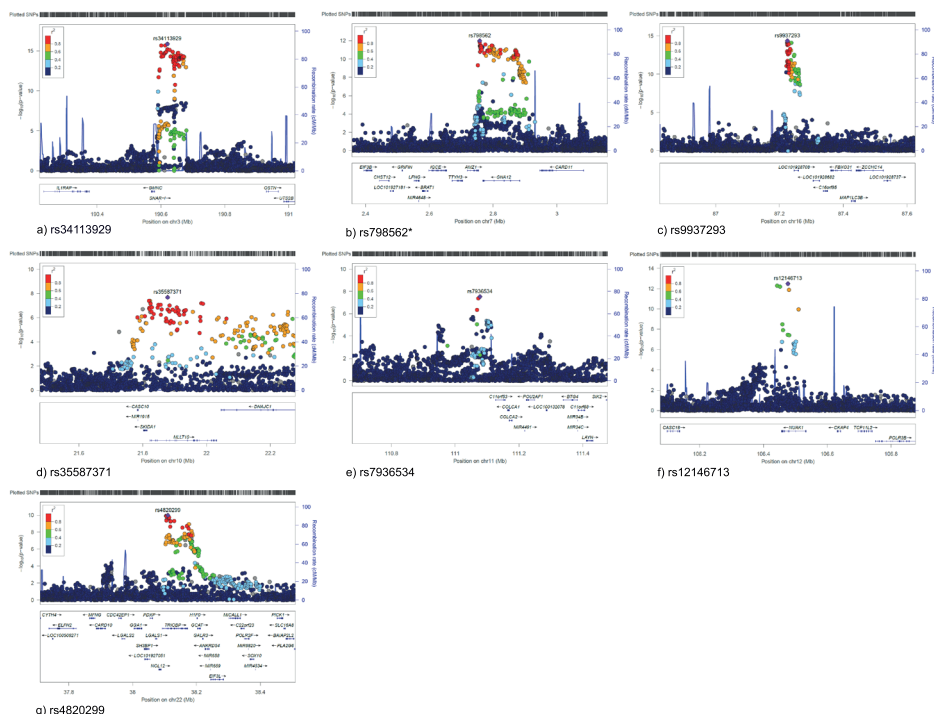


Figure 2. Regional association and recombination plots in combined stage 3 GWA meta-analysis. The left axis represents $-\log_{10} p$ -values for association with total later ventricular volume. The right axis represents the recombination rate, and the x-axis represents chromosomal position (hg19 genomic position). The most significant SNPs of the regions are denoted with a purple diamond. Surrounding SNPs are colored according to their pairwise correlation (r^2) with the top-associated SNP of the region. The gene annotations are below the figure.

Even though cohorts of European (EA) and African-American (AA) ancestry were included, all significant associations were mainly driven by EA samples (**Supplementary Fig. 11-12**). The direction of effect size across the EA cohorts for the 7 lead variants was generally concordant and showed no evidence of any single cohort driving the associations (**Supplementary Fig. 11**). Despite the different methods of phenotyping across the cohorts, the cohorts with different phenotyping methods showed evidence of effect suggesting that there is limited heterogeneity in effects (**Supplementary Fig. 12**).

To investigate whether 7 lead variants have an effect in early life, childhood, the analyses were carried out in a children's cohort of 1,141 participants from Generation R study. The percentage of lead variants showing consistent direction of effect with stage 3 was 85.7% (6 out of 7, binomial p -value = 0.05) (**Supplementary Data 4**), and a variant mapped to the 12q23.3 region showed nominal association with lateral ventricular volume in the children's cohort (effect = -0.15, p -value = 0.01). Additionally, three out of seven lead

variants (or their proxies; $r^2 > 0.7$) showed pleiotropic association (p -value $< 5 \times 10^{-8}$) with other traits according to the PhenoScanner database (**Supplementary Data 5**).²⁴

To capture gender-based differences, sex-stratified GWA analysis was performed ($N_{\text{men}} = 10,358$; $N_{\text{women}} = 12,872$). None of the 15,660,719 variants that were tested for heterogeneity between men and women reached genome-wide significance threshold (**Supplementary Fig. 13**). However, an indel located at 4q35.2 showed suggestive evidence of association in men (4:187559262:C_CAA, p -value $= 5.43 \times 10^{-8}$) but not in women (p -value $= 0.88$).

Independent signals within loci

The conditional and joint (COJO) analysis using the Genome-wide Complex Trait Analysis (GCTA) identified no other additional variants, after conditioning on the lead variant at the locus 3q28, 7p22.3, 10p12.31, 11q23.1, 12q23.3, 16q24.2, or 22q13.1.

Functional annotation

A large proportion of genome-wide significant variants were intergenic (335/460) (**Supplementary Fig. 14**). Variants with the highest probability of having a regulatory function based on RegulomeDB score (Category 1 RegulomeDB score) were located at 7p22.3 and at 22q13.1 (**Supplementary Data 6**). Of 7 lead variants, 4 were intergenic, 4 were in an active chromatin state and 3 showed expression quantitative trait (eQTL) effects (**Supplementary Data 6**). The lead SNP at 22q13.1 (rs4820299) was associated with differential expression of the largest number of genes ($n = 6$). In brain tissue, the alternate allele of this SNP was associated with higher expression of *TRIOBP* suggesting that higher expression was associated with smaller lateral ventricles (**Supplementary Fig. 15**).

Partitioned heritability

SNP-based heritability in the sample of European ancestry participants was estimated at 0.20 ($SE = 0.02$) using LD score regression, and this was higher in women (0.19 ($SE = 0.04$)) than in men (0.15 ($SE = 0.05$)). The 7 lead variants explained 1.5% of total variance in lateral ventricular volume. Partitioning of heritability based on functional annotation using LD score regression, revealed significant enrichment of SNPs within 500 bp of highly active enhancers, where 17% of SNPs accounted for 54% of the heritability (p -value $= 7.9 \times 10^{-6}$, **Supplementary Table 2**). Significant enrichment was also found for histone marks including H3K27ac (which indicates enhancer and promoter regions), H3K9ac (which highlights promoters), H3K4me3 (which indicates promoters/transcription starts), and H3K4me1 (which highlights enhancers) (**Supplementary Table 2**).^{25,26}

Functional enrichment analysis

Functional enrichment analysis using regulatory regions from the ENCODE and Roadmap projects using the GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction (GARFIELD) method revealed that SNPs associated with lateral ventricular volume at p -value threshold $< 10^{-5}$ were more often located in genomic regions harboring histone marks (H3K9ac (associated with promoters) and H3K36me3 (associated with transcribed regions))²⁵ and DNaseI hypersensitivity sites (DHS) than a permuted background (**Figure 3, Supplementary Data 7**).

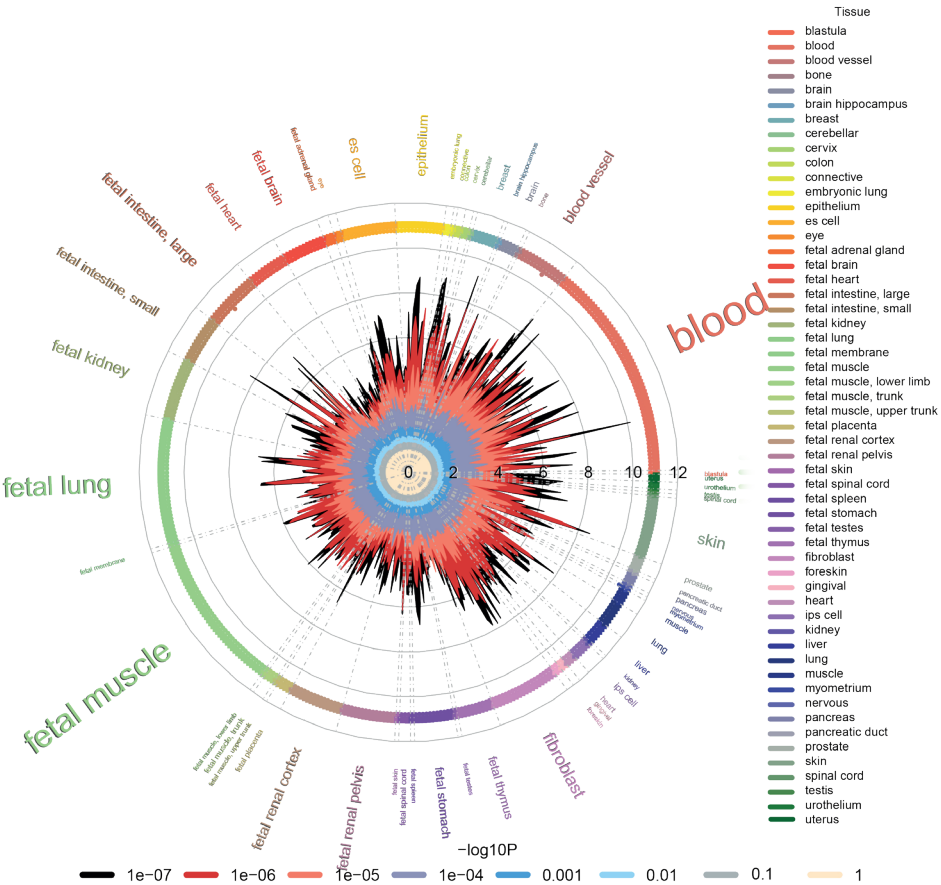


Figure 3. Functional enrichment analysis of lateral ventricular volume loci within DNaseI hypersensitivity spots. The radial lines show fold enrichment (FE) at eight GWA p -value thresholds. The results are shown for each of 424 cell types which are sorted by tissue, represented along the outer circle of the plot. The font size is proportional to the number of cell types from the tissue. FE values are plotted with different colors with respect to different GWA thresholds. Significant enrichment for a given cell type is denoted along the outer circle of the plot from a GWA p -value threshold $< 10^{-5}$ (outermost) to GWA p -value threshold $< 10^{-8}$ (innermost). The results show ubiquitous enrichment.

Integration of gene expression data

Integration of functional data from the Genotype-Tissues Expression (GTEx) project using the MetaXcan method revealed two significant associations between genetically predicted expression in brain tissue and lateral ventricular volume (**Supplementary Fig. 16**). Expression levels of *TRIOBP* at the locus 22q13.1 (p -value = 3.2×10^{-6}) and *MRPS16* at the locus 10q22.2 (p -value = 1.8×10^{-6}) were associated with lateral ventricular volume.

Gene annotation and pathway analysis

The results of gene-based and pathway analyses are illustrated in **Supplementary Table 3** and **Supplementary Data 8**. The pathway analysis identified “regulation of cytoskeleton organization” (GO:0051493) gene-set to be significantly enriched (p -value = 6×10^{-6}). Genes of the “regulation of cytoskeleton organization” pathway have previously been implicated in various neurological or cardiovascular diseases (**Supplementary Data 9**). Furthermore, pathways that pointed towards sphingosine 1 phosphate (S1P) signaling showed suggestive enrichment (**Supplementary Data 8**).

Genetic correlation

Additionally, we examined the genetic overlap between lateral ventricular volume and other traits (**Table 2**). We found that genetically-determined components of thalamus and lateral ventricular volumes appear to be negatively correlated ($\rho_{\text{genetic}} = -0.59$, p -value = 3.14×10^{-6}). This finding was also confirmed at the phenotype level (**Supplementary Table 4**). Weaker genetic overlap was observed with infant head circumference ($\rho_{\text{genetic}} = 0.28$, p -value = 8.7×10^{-3}), intracranial volume ($\rho_{\text{genetic}} = 0.35$, p -value = 9×10^{-3}), height ($\rho_{\text{genetic}} = -0.14$, p -value = 5.7×10^{-3}), and mean pallidum ($\rho_{\text{genetic}} = -0.29$, p -value = 2.5×10^{-2}), whereas no significant genetic overlap was found with neurological diseases, psychiatric diseases, or personality traits.

Genetic risk score

We next examined the association of genetic risk scores (GRS) for Alzheimer’s disease, Parkinson’s disease, schizophrenia, bipolar disorder, cerebral small vessel disease, and tau-related pathology, including tau and phosphorylated tau levels in cerebrospinal fluid, amyotrophic lateral sclerosis (ALS) and progressive supranuclear palsy (PSP), using the lead SNPs from the largest published GWA study and lateral ventricular volume (**Supplementary Data 10**). We found a suggestive association of GRS for tau levels in cerebrospinal fluid (p -value = 9.59×10^{-3}) and lateral ventricular volume (**Supplementary Table 5**). The association was driven by one SNP (**Supplementary Table 6**). No association was observed with other examined phenotypes (**Supplementary Table 5**).

Table 2. The results of genetic correlation between the lateral ventricular volume and anthropometric traits, brain volumes, neurological and psychiatric diseases and personality traits.

Category	Phenotype	PMID	N	rg	SE	P
Anthropometric						
	Height	20881960	133859	-0.135	0.049	5.70E-03
	Infant head circumference	22504419	10768	0.284	0.108	8.70E-03
	Child birth length	25281659	28459	-0.133	0.089	1.34E-01
	Child birth weight	23202124	26836	-0.118	0.102	2.47E-01
Brain volume						
	Mean Thalamus	25607358	13193	-0.591	0.127	3.14E-06
	Mean Pallidum	25607358	13142	-0.29	0.129	2.47E-02
	ICV	25607358	11373	0.347	0.133	9.00E-03
	Mean Accumbens	25607358	13112	-0.29	0.158	6.64E-02
	Mean Putamen	25607358	13145	-0.15	0.089	9.13E-02
	Mean Hippocampus	25607358	13163	-0.204	0.132	1.20E-01
	Mean Caudate	25607358	13171	0.012	0.105	9.06E-01
Neurological diseases						
	Alzheimer's disease	24162737	54162	0.181	0.11	9.87E-02
	Parkinson's disease	19915575	5691	-0.096	0.084	2.55E-01
	Amyotrophic lateral sclerosis	27455348	36052	-0.032	0.128	8.04E-01
	White matter hyperintensities	25663218	17940	0.100	0.094	2.87E-01
Personality traits						
	Neo-conscientiousness	21173776	17375	-0.359	0.158	2.27E-02
	Neo-openness to experience	21173776	17375	0.088	0.118	4.56E-01
	Neuroticism	27089181	170911	-0.03	0.065	6.45E-01
Psychiatric traits						
	ADHD	20732625	5422	-0.276	0.152	6.90E-02
	PGC cross-disorder analysis	23453885	61220	-0.121	0.071	8.65E-02
	Major depressive disorder	22472876	18759	-0.165	0.102	1.05E-01
	Schizophrenia	25056061	77096	-0.067	0.044	1.30E-01
	Subjective well being	27089181	298420	0.087	0.075	2.50E-01
	ADHD (No GC)	27663945	17666	-0.151	0.149	3.11E-01
	Depressive symptoms	27089181	161460	-0.038	0.071	5.93E-01
	Autism spectrum disorder	0	10263	0.041	0.092	6.53E-01
	Anorexia Nervosa	24514567	17767	0.011	0.056	8.43E-01
	Bipolar disorder	21926972	16731	0.009	0.078	9.12E-01

Abbreviations: rg – genetic correlation; SE - standard error; P - *p*-value;

DISCUSSION

We have performed the first genome-wide association study of lateral ventricular volume including up to 23,533 individuals. We identified statistically significant association between lateral ventricular volume and variants at 7 loci. Additionally, we found that genetically-determined components of thalamus and lateral ventricular volume are correlated.

The strongest association was observed at the intergenic 3q28 locus between non-coding RNA *SNAR-1* and *OSTN*. This region has previously been associated with cerebrospinal fluid tau/ptau levels and Alzheimer's disease risk, tangle pathology and cognitive decline.²⁷ Similarly, the genome-wide significant locus at 12q23.3 encompasses *NUAK1*, which has also been associated with tau pathology. Nuak1 modulates tau levels in human cells and animal models and associates with tau accumulation in different tauopathies.²⁸ *NUAK1* is most prominently expressed in the brain where it has a role in mediating axon growth and branching in cortical neurons.²⁹ The lead SNP of the 12q23.3 locus mapped to an intron of *NUAK1*. This SNP is among the top 1% of most deleterious variants in the human genome based on its Combined Annotation Dependent Depletion (CADD) score of 21.5 and is located in an enhancer region (**Supplementary Data 6**). Interestingly, this variant also showed an effect in early life.

In our data, the significant variants of 7p22.3 region had the highest probability of being regulatory based on the RegulomeDB score (1b). The lead variant at 7p22.3 was in an active chromatin state and was associated with differential expression of *GNA12* (**Supplementary Data 6**). The *GNA12* gene is involved in various transmembrane signaling systems.³⁰⁻³³ Interestingly, this gene was part of S1P signaling pathways identified to be enriched among genes associated with lateral ventricular volume. S1P, a bioactive sphingolipid metabolite, regulates nervous system development³⁴ such as neuronal survival, neurite outgrowth, and axon guidance,^{35,36} and plays a role in neurotransmitter release.³⁷ It also plays a role in regulating the development of germinal matrix (GM) vasculature.³⁸ Disruption of S1P regulation results in defective angiogenesis in GM, hemorrhage, and enlarged ventricles.³⁸

The other identified locus, 16q24.2, has previously been connected with small vessel disease and white-matter lesions formation.³⁹ Further, the alternate allele of the lead SNP at 22q13.1 in *TRIOBP* is associated with higher expression of the same gene in basal ganglia and brain cortex, and the same allele is associated with smaller lateral ventricular volume. Interestingly, predicted expression of this gene in cerebral cortex was significantly associated with lateral ventricular volume, suggesting a causal functional

role of the gene. The same analysis revealed significant association of the expression of *MRPS16* in frontal cortex with lateral ventricular volume. This gene was previously related to agenesis/hypoplasia of corpus callosum and enlarged ventricles.⁴⁰

Finally, the lead intergenic SNP at 11q23.1 maps between *C11orf53* and *ARHGAP20*, whereas the 10p12.31 region encompasses *MLLT10* which has been linked to various leukemias, ovarian cancer, and meningioma.^{41,42} The effect size of this variant on lateral ventricular volume was correlated with mean cohort age, with the effect being near zero at younger age and larger at older ages.

The gene-enrichment analysis highlighted “regulation of cytoskeleton organization” (GO:0051493) pathway. Genes that are part of this pathway have previously been implicated in various neurological diseases such as Parkinson’s disease (*PARK2*), frontotemporal dementia (*MAPT*), neurofibromatosis 2 (*NF2*), tuberous sclerosis (*TSC1*) (**Supplementary Data 9**). The cytoskeleton is essentially involved in all cellular processes, and therefore crucial for processes in the brain such as cell proliferation, differentiation, migration, and signaling. Dysfunction of cytoskeleton has been associated with neurodevelopmental, psychiatric and neurodegenerative diseases.⁴³⁻⁴⁵

Previous studies showed significant sex-specific differences in lateral ventricular volume.^{46,47} In our study we did not observe sex-specific differences; as for the lead 7 variants, both males and females were contributing to the association signal. However, we observed only one suggestive association at 4q35.2 that showed association in men only. The lead variant (indel) is mapped to *FAT1* which encodes atypical cadherins. Mutation in this gene causes a defect in cranial neural tube closure in a mouse model and an increase in radial precursor proliferation in the cortex.⁴⁸ However, the SNP-based heritability estimates were slightly higher in females. This may be explained by the differences in sample size in male and female-specific analyses implying that there is lower precision.

We estimated that 20% of genetic variance in lateral ventricular volume could be explained by common genetic variants, suggesting that common variants represent a substantial fraction of overall genetic component of variance. Moreover, the most statistically significant effect occurred in the regions of highly active enhancers and histone marks, suggesting their involvement in gene expression. Using the LD score regression method, we found a significant negative genetic correlation between lateral ventricular volume and thalamus volume. However, these may not be independent events, but inverse reflections of the same biology. Even though not strictly significant, we also observed trends for genetic correlations with other brain volumetric measures. Fur-

thermore, no genome-wide overlap was found between lateral ventricular volume and various neurological or psychiatric diseases. Given that enlargement of lateral ventricles has been suggested in Alzheimer's disease, we examined the association of *APOE* alleles and found no association between the *APOE* $\epsilon 4$ (p -value = 0.86) or *APOE* $\epsilon 2$ (p -value = 0.81) and lateral ventricular volume in our study population.

As we identified loci underlying lateral ventricular volume at the genome-wide level, but also genes and common pathways, our results provide various insights into the genetic contribution to lateral ventricular volume variability and a better understanding of the complex genetic architecture of brain structures. The genes with variants that we found to be associated with lateral ventricular volume are relevant to neurological aging given the characteristics of the study population which is relatively free from the disease as participants with stroke, traumatic brain injury and dementia at the time of magnetic resonance imaging (MRI) were excluded. This is in line with the previously published work of Pfefferbaum *et al.* who showed that the stability of lateral ventricles is genetically determined, whereas other factors such as normal aging or trauma and disease play a role in its change.^{1,16}

However, while studying genetic overlap of lateral ventricular volume and various neurological or psychiatric disorders at multiple levels (LD score regression/polygenic, GRS/oligogenic, GWA hits/monogenic), we found evidence that some single genetic variants have pleiotropic effect on lateral ventricular volume and biochemical markers for a neurological disease (AD) or meningioma (**Supplementary Data 5**), while no evidence was found for genetic overlap with other neurological or psychiatric disorders (**Table 2**, **Supplementary Table 5**). The pattern of association between lateral ventricular volume and psychiatric disorder i.e. schizophrenia on multiple scales is similar to the findings of Franke *et al.* who evaluated association of various subcortical brain volumes and schizophrenia and reported no evidence of genetic overlap.⁴⁹ Even though our study does not provide a definite statement regarding the relationship between lateral ventricular volume and neurological or psychiatric disorders, it lays the foundation for future studies which should disentangle whether lateral ventricular volume is genetically related or unrelated to various neurological and psychiatric disorders (e.g. result from reverse causation). Novel insights may be revealed by improving the power of the studies, studying homogeneous samples with harmonized phenotype assessment methods along with evaluation of common and rare variants.

The strengths of our study are the large sample, population-based design and the use of quantitative MRI. Our study also has several limitations. Despite the effort to harmonize phenotype assessment, the methods used to quantify lateral ventricular volume

differ across cohorts. Because of this phenotypic heterogeneity, association results of participating cohorts were combined using a sample-size weighted meta-analysis, thus limiting discussion on effect sizes. Secondly, phenotypic heterogeneity may have caused the loss of statistical power. However, despite heterogeneity in the phenotype assessment, the association signals were coming from several studies irrespective of the method of phenotype assessment, which suggests robustness of our findings. Furthermore, although we made an effort to include cohorts of EA and AA ancestry, the study comprised predominately of individuals of European origin (22,045 individuals of EA and 1,488 of AA ancestry). Given the disparity in sample size, it is difficult to distinguish whether any inconsistency in results between the 2 groups stems from true genetic differences or from differential power to detect genetic effects. Indeed, this is also exemplified by the plots of the Z-scores (**Supplementary Fig. 11**) showing that direction of effect size in AA cohorts is often inconsistent with the direction of effect size in EA cohorts. However, the same inconsistency can be observed with European cohorts of equally small sample size. This inconsistency may be due to small sample size rather than ethnic background but we cannot rule out that racial-ethnic specific effects may exist. This limitation underscores the need for expanding research studies in non-European populations. Finally, as some loci only reached the genome-wide significance in the combined meta-analysis, they should be considered as highly probable findings and would still require independent replication.

To conclude, we identified genetic associations of lateral ventricular volume with variants mapping to 7 loci and implicating several pathways, including pathway related to tau pathology, cytoskeleton organization, and S1P signaling. These data provide new insights into understanding brain morphology.

METHODS

Study design

The overview of study design is illustrated in **Supplementary Fig. 1**. We performed a GWA meta-analysis of 11,396 participants of mainly European ancestry from 12 studies (stage 1) that contributed summary statistic data before a certain deadline. The deadline was set prior to data inspection and was not influenced by the results of the GWA meta-analysis. Variants that surpassed the genome-wide significance threshold ($p\text{-value} < 5 \times 10^{-8}$) were subsequently evaluated in an independent sample of 12,137 participants of mainly European ancestry from 14 studies (stage 2). Finally, we performed a meta-analysis of all stage 1 and stage 2 studies (stage 3).

Study population

All participating studies are part of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium.⁵⁰ A detailed description of participating studies can be found in **Supplementary Note 1**. General characteristics of study participants are provided in **Supplementary Data 1**. Written informed consent was obtained from all participants. Each study was approved by local ethical committees or the institutional review boards (see **Supplementary Note 1** for details).

Imaging

Each study performed magnetic resonance imaging (MRI) and estimated the volume of the lateral ventricles and intracranial volume (ICV). The field strength of scanners ranged from 0.35 to 3 T. Information on scanner manufacturers and measurement methods is provided in **Supplementary Data 2**. While most of the studies quantified lateral ventricular volume using validated automated segmentation methods, some studies used validated visual grading scales. The visual and volumetric scales were compared previously and showed high agreement for lateral ventricular volume.² The assessment of consistency of lateral ventricular volume on volumetric scale across time and different versions of software (freesurfer v4.5, v5.1, and v6.0), revealed high intraclass correlation ($ICC > 0.98$) in a subset of participants from the Rotterdam Study. Participants with dementia at the time of MRI, traumatic brain injury, prior or current stroke or intracranial tumors were excluded.

Genotyping and imputation

Information on genotyping platforms, quality control procedures and imputations methods for each participating study are provided in **Supplementary Data 3**. All studies used commercially available genotyping arrays, including Illumina or Affymetrix arrays. Similar quality control procedures were applied for each study (**Supplementary Data 3**). Using the validated software (Minimac,⁵¹ IMPUTE,⁵² BEAGLE⁵³), each study performed genotype imputations using mostly the 1000 Genome phase 1 v3 reference panel.

Genome-wide association (GWA) analysis

Each participating study performed the GWA analysis of total lateral ventricular volume under an additive model using variant allele dosage as predictors and natural logarithm of the total lateral ventricular volume as the dependent variable. Transformation of the lateral ventricular volume was applied to obtain approximately normal distribution (**Supplementary Fig. 17**). The association analyses were adjusted for age, sex, total intracranial volume, age² if significant, population stratification, familial relationship (family-based studies) or study site (multi-site studies). Population stratification was controlled for by including principal components derived from genome-wide genotype

data. Study-specific details on covariates and software used are provided in **Supplementary Data 3**. Quality control (QC) was conducted for all participating studies using a standardized protocol provided by Winkler *et al.*⁵⁴ Variants with low imputation quality $r^2 < 0.3$ or minor allele count (MAC) ≤ 6 were filtered out. The association results of participating studies were combined using a fixed-effect sample-size weighted Z-score meta-analysis in METAL because of the difference in measurement methods of lateral ventricular volume.⁵⁵ Genomic control was applied to account for small amounts of population stratification or unaccounted relatedness. After the meta-analysis, variants with information in less than half the total sample size were excluded. Meta-analyses were performed separately for each of the stages. In the stage 1 meta-analysis, a p -value $< 5 \times 10^{-8}$ was considered significant. Variants that surpassed the threshold were evaluated in the stage 2 meta-analysis. In order to model linkage disequilibrium (LD) between those variants, we first calculated the number of independent tests using the eigenvalues of a correlation matrix using the Matrix Spectral Decomposition (matSpDlite) software.⁵⁶ Subsequently, a Bonferroni correction was applied for the effective number of independent tests ($0.05/10$ independent SNPs = 5×10^{-3}). Additionally, all analyses were stratified by sex. Following the same QC steps as for overall analyses, the sex-stratified association results of participating studies were combined using a fixed-effect sample-size weighted Z-score meta-analysis in METAL while applying genomic control.⁵⁵ The variants were assessed only if test statistics (Z-score) were heterogeneous between males and females (p -value < 0.1) and if the association in a sex-combined analysis did not reach genome-wide significance threshold.⁵⁷

Conditional analysis

In order to identify variants that were independently associated with lateral ventricular volume, we performed conditional and joint (COJO) GWA analysis using Genome-wide Complex Trait Analysis (GCTA), version 1.26.0.⁵⁸ LD pattern was calculated based on 1000 Genome phase 1v3 imputed data of 6,291 individuals from the Rotterdam Study I.

Functional annotation

To annotate genome-wide significant variants with regulatory information, we used HaploReg v4.1⁵⁹, RegulomeDB v1.1⁶⁰, and Combined Annotation Dependent Depletion (CADD) tools.⁶¹ To determine whether they have an effect on gene expression, we used GTEx data.⁶² For the lead variants, we explored 5 chromatin marks assayed in 127 epigenomes (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3) of RoadMap data.⁶³ To search for pleiotropic associations between our lead variants and their proxies ($r^2 > 0.7$) with other traits, we used the PhenoScanner database designed to facilitate the cross-referencing of genetic variants with many phenotypes.⁹ The association results with genome-wide significance at 5×10^{-8} were extracted.

Variance explained

The proportion of variance in lateral ventricular volume explained by each lead variant was calculated using Pearson's phi coefficient squared as explained in Draisma *et al.*⁶⁴ The total proportion of variance in lateral ventricular volume was calculated by adding up the proportions of variance in lateral ventricular volume explained by each lead association signal.

Partitioned heritability

SNP-based heritability and partitioned heritability analyses were performed using LD score regression following the previously described method.⁶⁵ Partitioned heritability analysis determines enrichment of heritability in SNPs partitioned into 24 functional classes as reported in Finucane *et al.*⁶⁵ To avoid bias, an additional 500 bp window was included around the variants included in the functional classes. Only the HapMap3 variants were included as these seem to be well-imputed across cohorts.

Functional enrichment analysis

We performed functional enrichment analysis using regulatory regions from the ENCODE and Roadmap projects using GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction (GARFIELD) method.⁶⁶ The method provides fold enrichment (FE) statistics at various GWA *p*-value thresholds after taking into account LD, minor allele frequency, and local gene density.⁶⁶ The FE statistics were calculated at eight GWA *p*-value thresholds (0.1 to 1×10^{-8}). The associations were tested for various regulatory elements including DNase-I hypersensitivity sites, histone modifications, chromatin states and transcription factor binding sites in over 1000 cell and tissue-specific annotations.⁶⁶ The significance threshold calculated based on the number of annotations used was set at 4.97×10^{-5} .

Integration of gene expression

To integrate functional data in the context of our meta-analysis results, we used the MetaXcan method, which evaluated the association between lateral ventricular volume and brain-specific gene-expression levels predicted by genetic variants using the data from GTEx project.^{62,67} This method is an extension of PrediXcan method modified to use summary statistic data from meta-analysis.⁶⁷ Based on a total number of genes tested, the Bonferroni corrected significance threshold was set to $0.05/12,379 = 4 \times 10^{-6}$.

Gene annotation and pathway-based analysis

The gene-based test statistics were computed using VEGAS2 software which tests for enrichment of multiple single variants within the genes while accounting for LD structure.⁶⁸ LD structure was computed based on the 1000 Genomes phase 3 population.

Variants within 10 kb of the 5' and 3' untranslated regions were included in this analysis in order to maintain regulatory variants.⁶⁸ Subsequently, the gene-based scores were used to perform gene-set enrichment analysis using VEGAS2pathway.⁶⁹ VEGAS2Pathway approach accounts for LD between variants within a gene, and between neighboring genes, gene size, and pathway size.⁶⁹ It uses computationally predicted Gene Ontology pathways and curated gene-sets from the MSigDB, PANTHER, and pathway commons databases.⁶⁹ The pathway-based significance threshold was set to the p -value = 1×10^{-5} while taking into account the multiple testing of correlated pathways (0.05/5,000 independent tests).⁶⁹

Genetic correlation

We used the LD score regression method to estimate genetic correlations between lateral ventricular volume and various traits including anthropometric traits, brain volumes, neurological and psychiatric diseases and personality traits. The analyses were performed using a centralized database of summary-level GWA study results and a web interface for LD score regression, the LD-hub.⁷⁰ Summary-level GWA study results for white matter hyperintensities were obtained from the CHARGE consortium⁷¹ and the analyses were performed using the ldsc tool (<https://github.com/bulik/ldsc>).

Genetic risk scores

We generated genetic risk scores (GRS) for Alzheimer's disease, amyotrophic lateral sclerosis (ALS), Parkinson's disease, bipolar disorder, schizophrenia, white matter lesions and tau-related phenotypes. The tau-related phenotypes, including tau and phosphorylated tau levels in cerebrospinal fluid, and progressive supranuclear palsy (PSP), were studied in relatively small sample and are therefore not appropriate for LD score regression. We extracted the lead genome-wide significantly associated SNPs and their effect estimates from the largest published GWA studies (**Supplementary Data 10**). For white matter lesions burden, effect estimate and standard errors were estimated from Z-statistics using the previously published formula.⁷² The allele associated with an increased risk in corresponding traits was considered to be the effect allele. The weighted GRS was constructed as the sum of products of effect sizes as weights and respective allele dosages from 1000 Genome imputed data of Rotterdam Study using R software version 3.2.5 (<https://www.R-project.org>). Variants with low imputation quality ($r^2 < 0.3$) were excluded. Subsequently, the GRS was tested for association with lateral ventricular volume in three cohorts of Rotterdam Study while adjusting for age, sex, total intracranial volume, age² and population stratification. The significance threshold for genetic risk score association was set to p -value = 5×10^{-3} (0.05/10) based on the number of genetic risk scores tested.

Data availability

The summary statistics will be made available upon the publication on the CHARGE dbGaP site under the accession number phs000930.v7.p1.

REFERENCES

1. Pfefferbaum, A., Sullivan, E.V. & Carmelli, D. Morphological changes in aging brain structures are differentially affected by time-linked environmental influences despite strong genetic stability. *Neurobiol Aging* **25**, 175-83 (2004).
2. Carmichael, O.T. *et al.* Ventricular volume and dementia progression in the Cardiovascular Health Study. *Neurobiol Aging* **28**, 389-97 (2007).
3. Apostolova, L.G. *et al.* Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (MCI), and Alzheimer Disease. *Alzheimer Dis Assoc Disord* **26**, 17-27 (2012).
4. Long, X. *et al.* Healthy aging: an automatic analysis of global and regional morphological alterations of human brain. *Acad Radiol* **19**, 785-93 (2012).
5. Nestor, S.M. *et al.* Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. *Brain: A Journal of Neurology* **131**, pp (2008).
6. Kuller, L.H. *et al.* Determinants of vascular dementia in the Cardiovascular Health Cognition Study. *Neurology* **64**, 1548-52 (2005).
7. Mak, E. *et al.* Longitudinal whole-brain atrophy and ventricular enlargement in nondemented Parkinson's disease. *Neurobiol Aging* **55**, 78-90 (2017).
8. Kuller, L.H., Lopez, O.L., Becker, J.T., Chang, Y. & Newman, A.B. Risk of dementia and death in the long-term follow-up of the Pittsburgh Cardiovascular Health Study-Cognition Study. *Alzheimers Dement* **12**, 170-183 (2016).
9. Vita, A., De Peri, L., Silenzi, C. & Dieci, M. Brain morphology in first-episode schizophrenia: a meta-analysis of quantitative magnetic resonance imaging studies. *Schizophr Res* **82**, 75-88 (2006).
10. Kempton, M.J., Geddes, J.R., Ettinger, U., Williams, S.C. & Grasby, P.M. Meta-analysis, database, and meta-regression of 98 structural imaging studies in bipolar disorder. *Arch Gen Psychiatry* **65**, 1017-32 (2008).
11. Olabi, B. *et al.* Are there progressive brain changes in schizophrenia? A meta-analysis of structural magnetic resonance imaging studies. *Biological Psychiatry* **70.1**, 88-96 (2011).
12. Mosley, T.H., Jr. *et al.* Cerebral MRI findings and cognitive functioning: the Atherosclerosis Risk in Communities study. *Neurology* **64**, 2056-62 (2005).
13. Appelman, A.P. *et al.* White matter lesions and lacunar infarcts are independently and differently associated with brain atrophy: the SMART-MR study. *Cerebrovasc Dis* **29**, 28-35 (2010).
14. Geerlings, M.I. *et al.* Brain volumes and cerebrovascular lesions on MRI in patients with atherosclerotic disease. The SMART-MR study. *Atherosclerosis* **210**, 130-6 (2010).
15. Horga, G. *et al.* Correlations between ventricular enlargement and gray and white matter volumes of cortex, thalamus, striatum, and internal capsule in schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience* **261.7**, 467-476 (2011).
16. Kremen, W.S. *et al.* Heritability of brain ventricle volume: converging evidence from inconsistent results. *Neurobiol Aging* **33**, 1-8 (2012).
17. Peper, J.S., Brouwer, R.M., Boomsma, D.I., Kahn, R.S. & Hulshoff Pol, H.E. Genetic influences on human brain structure: a review of brain imaging studies in twins. *Hum Brain Mapp* **28**, 464-73 (2007).
18. Schmitt, J.E. *et al.* Review of twin and family studies on neuroanatomic phenotypes and typical neurodevelopment. *Twin Res Hum Genet* **10**, 683-94 (2007).
19. Kremen, W.S. *et al.* Genetic and environmental influences on the size of specific brain regions in midlife: the VETSA MRI study. *Neuroimage* **49**, 1213-23 (2010).

20. Eyler, L.T. *et al.* Genetic patterns of correlation among subcortical volumes in humans: Results from a magnetic resonance imaging twin study. *Human Brain Mapping* **32.4**, 641-653 (2011).
21. Mata, I. *et al.* A neuregulin 1 variant is associated with increased lateral ventricle volume in patients with first-episode schizophrenia. *Biological Psychiatry* **65.6**, 535-540 (2009).
22. Crespo-Facorro, B. *et al.* Low-activity allele of Catechol-O-Methyltransferase (COMT) is associated with increased lateral ventricles in patients with first episode non-affective psychosis. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* **31**, 1514-1518 (2007).
23. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics* **38**, 209-213 (2006).
24. Staley, J.R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207-3209 (2016).
25. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
26. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
27. Cruchaga, C. *et al.* GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* **78**, 256-68 (2013).
28. Lasagna-Reeves, C.A. *et al.* Reduction of Nuak1 Decreases Tau and Reverses Phenotypes in a Tauopathy Mouse Model. *Neuron* **92**, 407-418 (2016).
29. Courchet, J. *et al.* Terminal Axon Branching Is Regulated by the LKB1-NUAK1 Kinase Pathway via Presynaptic Mitochondrial Capture. *Cell* **153**, 1510-1525 (2013).
30. Yanamadala, V., Negoro, H., Gunaratnam, L., Kong, T. & Denker, B.M. Galpha12 stimulates apoptosis in epithelial cells through JNK1-mediated Bcl-2 degradation and up-regulation of IkappaBalpha. *J Biol Chem* **282**, 24352-63 (2007).
31. Kelly, P. *et al.* The G12 family of heterotrimeric G proteins promotes breast cancer invasion and metastasis. *Proc Natl Acad Sci U S A* **103**, 8173-8 (2006).
32. Krakstad, B.F., Ardawati, V.V. & Aragay, A.M. A role for Galpha12/Galpha13 in p120ctn regulation. *Proc Natl Acad Sci U S A* **101**, 10314-9 (2004).
33. Zhu, D., Kosik, K.S., Meigs, T.E., Yanamadala, V. & Denker, B.M. Galpha12 directly interacts with PP2A: evidence FOR Galpha12-stimulated PP2A phosphatase activity and dephosphorylation of microtubule-associated protein, tau. *J Biol Chem* **279**, 54983-6 (2004).
34. Blaho, V.A. & Hla, T. An update on the biology of sphingosine 1-phosphate receptors. *J Lipid Res* **55**, 1596-608 (2014).
35. Strohlic, L., Dwivedy, A., van Horck, F.P., Falk, J. & Holt, C.E. A role for S1P signalling in axon guidance in the Xenopus visual system. *Development* **135**, 333-42 (2008).
36. Herr, D.R. *et al.* Sphingosine 1-phosphate (S1P) signaling is required for maintenance of hair cells mainly via activation of S1P2. *J Neurosci* **27**, 1474-8 (2007).
37. Shen, H. *et al.* Coupling between endocytosis and sphingosine kinase 1 recruitment. *Nat Cell Biol* **16**, 652-62 (2014).
38. Ma, S., Santhosh, D., Kumar, T.P. & Huang, Z. A Brain-Region-Specific Neural Pathway Regulating Germinal Matrix Angiogenesis. *Dev Cell* **41**, 366-381 e4 (2017).
39. Traylor, M. *et al.* Genetic variation at 16q24.2 is associated with small vessel stroke. *Annals of Neurology* **81.3**, 383-394 (2017).
40. Miller, C. *et al.* Defective mitochondrial translation caused by a ribosomal protein (MRPS16) mutation. *Ann Neurol* **56**, 734-8 (2004).

41. Pharoah, P.D.P. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nature Genetics* **45**, 362-370 (2013).
42. Egan, K.M. *et al.* Brain tumor risk according to germ-line variation in the MLLT10 locus. *European Journal of Human Genetics* **23**, 132-134 (2015).
43. Paus, T., Pesaresi, M. & French, L. White matter as a transport system. *Neuroscience* **276**, 117-25 (2014).
44. McMurray, C.T. Neurodegeneration: diseases of the cytoskeleton? *Cell Death Differ* **7**, 861-5 (2000).
45. Cairns, N.J., Lee, V.M. & Trojanowski, J.Q. The cytoskeleton in neurodegenerative diseases. *J Pathol* **204**, 438-49 (2004).
46. Hasan, K.M., Moeller, F.G. & Narayana, P.A. DTI-based segmentation and quantification of human brain lateral ventricular CSF volumetry and mean diffusivity: validation, age, gender effects and biophysical implications. *Magn Reson Imaging* **32**, 405-12 (2014).
47. Pfefferbaum, A. *et al.* Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85 years) measured with atlas-based parcellation of MRI. *Neuroimage* **65**, 176-93 (2013).
48. Badouel, C. *et al.* Fat1 interacts with Fat4 to regulate neural tube closure, neural progenitor proliferation and apical constriction during mouse brain development. *Development* **142**, 2781-91 (2015).
49. Franke, B. *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nat Neurosci* **19**, 420-431 (2016).
50. Psaty, B.M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* **2**, 73-80 (2009).
51. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
52. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
53. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-97 (2007).
54. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).
55. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
56. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221-227 (2005).
57. Zeggini, E. & Ioannidis, J.P. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10**, 191-201 (2009).
58. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
59. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-4 (2012).
60. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).

61. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
62. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
63. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).
64. Draisma, H.H.M. *et al.* Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat Commun* **6**, 7208 (2015).
65. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* **47**, 1228 (2015).
66. Iotchkova, V. *et al.* GARFIELD - GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction. *bioRxiv* (2016).
67. Barbeira, A. *et al.* MetaXcan: Summary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results. *bioRxiv* (2016).
68. Mishra, A. & Macgregor, S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet* **18**, 86-91 (2015).
69. Mishra, A. & MacGregor, S. A Novel Approach for Pathway Analysis of GWAS Data Highlights Role of BMP Signaling and Muscle Cell Differentiation in Colorectal Cancer Susceptibility. *Twin Res Hum Genet* **20**, 1-9 (2017).
70. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
71. Verhaaren, B.F. *et al.* Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ Cardiovasc Genet* **8**, 398-409 (2015).
72. Chauhan, G. *et al.* Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. *Neurobiol Aging* **36**, 1765 e7-1765 e16 (2015).