5-4-2021

Multi-omic analysis elucidates the genetic basis of hydrocephalus

Andrew T Hale

Lisa Bastarache

Diego M Morales

John C Wellons III

David D Limbrick Jr.

See next page for additional authors

Authors

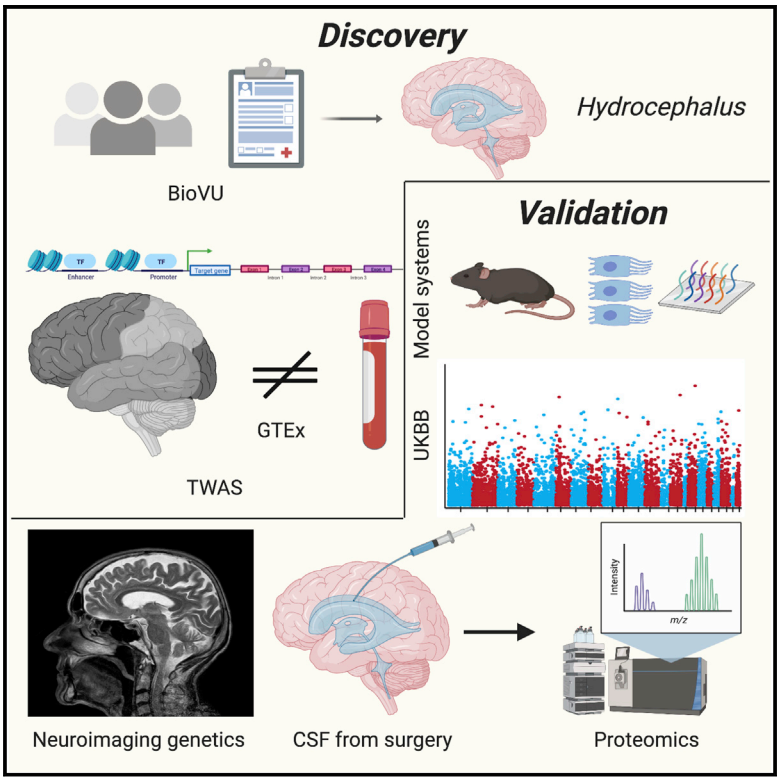
Andrew T Hale, Lisa Bastarache, Diego M Morales, John C Wellons III, David D Limbrick Jr., and Eric R Gamazon

Cell Reports

Article

Multi-omic analysis elucidates the genetic basis of hydrocephalus

Graphical abstract



Highlights

- We conducted TWASs of hydrocephalus in 10 brain regions and whole blood
- The top genes were enriched for TWAS associations with brain structure and integrity
- Validation in murine model and human proteomic data suggested polygenic architecture
- *MAEL* attained experiment-wide significance and independent replication

Authors

Andrew T. Hale, Lisa Bastarache, Diego M. Morales, John C. Wellons III, David D. Limbrick, Jr., Eric R. Gamazon

Correspondence

andrew.hale@vanderbilt.edu (A.T.H.), eric.gamazon@vumc.org (E.R.G.)

In brief

Hale et al. present an integrated omics approach to characterize the genetic basis of hydrocephalus. They reveal tissue-specific genetic associations and enrichment of genes associated with human brain structure phenotypes. Validation of hydrocephalus-associated genes in mouse choroid plexus and human cerebrospinal fluid supports polygenic contributions to hydrocephalus risk.

Hale et al., 2021, Cell Reports *35*, 109085
May 4, 2021
https://doi.org/10.1016/j.celrep.2021.109085

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Multi-omic analysis elucidates the genetic basis of hydrocephalus

Andrew T. Hale,[1,2,3,*] Lisa Bastarache,[4] Diego M. Morales,[5] John C. Wellons III,[6] David D. Limbrick, Jr.,[5] and Eric R. Gamazon[2,3,7,8,9,10,*]

[1]Vanderbilt University School of Medicine, Medical Scientist Training Program, Nashville, TN 37232, USA
[2]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[3]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[4]Department of Bioinformatics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA
[5]Division of Pediatric Neurosurgery, St. Louis Children's Hospital, St. Louis, MO 63110, USA
[6]Division of Pediatric Neurosurgery, Monroe Carell Jr. Children's Hospital of Vanderbilt University, Nashville, TN 37232, USA
[7]Data Science Institute, Vanderbilt University, Nashville, TN 37232, USA
[8]Clare Hall, University of Cambridge, Cambridge CB3 9AL, UK
[9]MRC Epidemiology Unit, University of Cambridge, Cambridge CB3 9AL, UK
[10]Lead contact
*Correspondence: andrew.hale@vanderbilt.edu (A.T.H.), eric.gamazon@vumc.org (E.R.G.)
https://doi.org/10.1016/j.celrep.2021.109085

SUMMARY

We conducted PrediXcan analysis of hydrocephalus risk in ten neurological tissues and whole blood. Decreased expression of *MAEL* in the brain was significantly associated (Bonferroni-adjusted $p < 0.05$) with hydrocephalus. PrediXcan analysis of brain imaging and genomics data in the independent UK Biobank ($N = 8,428$) revealed that *MAEL* expression in the frontal cortex is associated with white matter and total brain volumes. Among the top differentially expressed genes in brain, we observed a significant enrichment for gene-level associations with these structural phenotypes, suggesting an effect on disease risk through regulation of brain structure and integrity. We found additional support for these genes through analysis of the choroid plexus transcriptome of a murine model of hydrocephalus. Finally, differential protein expression analysis in patient cerebrospinal fluid recapitulated disease-associated expression changes in neurological tissues, but not in whole blood. Our findings provide convergent evidence highlighting the importance of tissue-specific pathways and mechanisms in the pathophysiology of hydrocephalus.

INTRODUCTION

Hydrocephalus is a heterogeneous disease resulting from abnormal accumulation of cerebrospinal fluid (CSF) and subsequent elevations in intracranial pressure resulting in impaired neurodevelopment and morbidity (Kahle et al., 2016; Tomycz et al., 2017). Hydrocephalus affects nearly 1 in 1,000 babies born in the United States (Simon et al., 2008), yet the genetic basis of the disease is largely unknown (Kousi and Katsanis, 2016). While clinical trials have attempted pharmacological strategies to treat hydrocephalus (Whitelaw et al., 2001), no pharmacological approaches have been successful. The current treatments for hydrocephalus are surgical interventions such as insertion of a ventriculoperitoneal (VP) shunt as well as endoscopic third ventriculostomy (ETV) with or without choroid plexus cauterization (CPC) (Kahle et al., 2016; Kulkarni et al., 2017). While many studies have evaluated the efficacy and cost of these procedures (Lim et al., 2018), long-term morbidity remains high.

Hydrocephalus can be a secondary consequence of intraventricular hemorrhage (IVH), spina bifida, infection, brain tumor, or congenital form. In addition, anatomic obstruction (i.e., aqueductal stenosis) impairing the flow of CSF can be caused by a rare X-linked mutation in L1 cell adhesion molecule (L1CAM) (Rosenthal et al., 1992). Proposed pathophysiological mechanisms of hydrocephalus include impaired development of the neural stem cell niche (Furey et al., 2018; Lehtinen et al., 2011, 2013; Lehtinen and Walsh, 2011; Carter et al., 2012), abnormal ciliated ependymal cells (Takagishi et al., 2017; Wilson et al., 2010; Wodarczyk et al., 2009), disruption of the ventricular zone (Castaneyra-Ruiz et al., 2018; McAllister et al., 2017), and dysfunction of CSF absorption and/or secretion (Karimy et al., 2017; Lun et al., 2015b). However, very little is known about the underlying germline genetic contributions to hydrocephalus (Kousi and Katsanis, 2016; Zhang et al., 2006).

While numerous studies have sought to identify causative genetic mechanisms leading to hydrocephalus, largely based on isolated human case studies and murine models (Kousi and Katsanis, 2016), critical limitations include cost, patient/family recruitment, number of patients (small by standards of population genetics), individual variant validation (typically *de novo* mutations), and very important species differences between model organisms and human disease. As hydrocephalus is a

component of a wide array of genetic syndromes and Mendelian disorders, as well as a secondary consequence of many pathologies, we hypothesized that hydrocephalus is a polygenic and complex disease, but may converge on a limited number of pathways. Thus, elucidation of the genetic basis of hydrocephalus may lead to new insights into pathophysiological mechanisms and identification of targets for pharmacological intervention.

Genome-wide association studies (GWASs) have become ubiquitous as a tool for identifying genetic predispositions to complex traits, but these analyses require very large sample sizes. Importantly, the underlying mechanisms for the identified loci are, for the most part, unclear. Elucidating the mechanistic basis of a complex, polygenic disorder (Manolio et al., 2009) requires understanding the molecular events that give rise to the disease process. We hypothesized that the summative risk for hydrocephalus results from small variation in the expression of many genes, leading to alterations of a limited number of pathways or biological processes that ultimately predispose individuals to disease. Because of the unavoidably smaller sample size for hydrocephalus (relative to other complex disorders), we aimed to test genes whose expression can be reliably determined using genetic variation (rather than millions of genetic variants, as in conventional GWASs, with unknown gene targets) in order to substantially improve statistical power.

To this end, we used whole-genome genetic data (from blood collected for routine clinical care) linked to the deidentified electronic medical record (BioVU; Roden et al., 2008) to perform the largest genetic analysis of hydrocephalus to date. We hypothesized that genetically determined gene expression contributes to the development of hydrocephalus. Thus, we applied PrediXcan (Gamazon et al., 2015), a gene-based method that utilizes the genetic component of gene expression for disease gene identification, leveraging imputation models derived from a reference human transcriptome panel of neurological tissues (10 brain regions from 889 individuals) and whole blood (from 338 patients) (Battle et al., 2017). Our approach identified genetically determined gene expression traits and pathways associated with hydrocephalus. Although correlation does not imply causation, a disease association with genetically determined expression substantially improves on a SNP-based association from conventional GWASs (in which the relevant gene is generally unknown; Nicolae et al., 2010). We considered whether the gene-level associations were driven by linkage disequilibrium (LD) contamination. We observed a notable degree of tissue specificity in the association of genetically determined gene expression with hydrocephalus risk. We replicated our experiment-wide significant finding in an independent GWAS dataset (UK Biobank) (Bycroft et al., 2018). Furthermore, we sought functional support for our findings using PrediXcan analysis of imaging-based phenotypes in the UK Biobank (Elliott et al., 2018) and transcriptome analysis of a murine model of hydrocephalus. Finally, we compared protein expression analysis from CSF isolated from infants with hydrocephalus to gene expression changes identified by PrediXcan, mirroring our findings in neurological tissues, but not in whole blood. In sum, we illuminate the complex genetic architecture of hydrocephalus and offer crucial insights into the pathophysiological basis of the disease.

RESULTS

A diagram illustrating our study design and approach can be found in Figure 1. In this study, we performed a systematic genetic study of hydrocephalus using PrediXcan (Gamazon et al., 2015, 2018), a methodology that estimates the genetic component of gene expression, to identify gene-level associations with disease. We performed systematic validation of our genetic findings using independent replication in the UK Biobank, analysis of brain structural imaging phenotypes linked to genetic information in the UK Biobank, transcriptome analysis of choroid plexus isolated from a murine model of hydrocephalus, and comparison of genetically determined expression changes to directly measured differential proteomic expression analysis of CSF isolated from infants with hydrocephalus.

Generation of imputed transcriptome (PrediXcan) to identify hydrocephalus-associated genes

Estimation of the genetically determined transcriptome in 10 neurological tissues and whole blood was performed using PrediXcan (Figure 1; see STAR Methods). Differential expression analysis in the frontal cortex highlighted maelstrom spermatogenic transposon silencer (*MAEL*) and lysine demethylase 1A (*KDM1A*) as outliers (Figure 2A). After multiple testing correction (Benjamini-Hochberg adjusted $p < 0.05$), *MAEL*, but not *KDM1A*, in the frontal cortex was significantly associated with hydrocephalus (Figure 2B). Notably, *MAEL* was experiment-wide significant, satisfying a highly stringent Bonferroni threshold for statistical significance ($p_{Bonferroni} < 0.05$) based on the total number of gene-tissue pairs tested (see STAR Methods). We note that the p value threshold for significance for PrediXcan is not the same as that for traditional GWAS, with PrediXcan enjoying a substantially reduced multiple testing burden (Gamazon et al., 2015; see STAR Methods). Genetically determined *MAEL* expression did not differ between varying etiologies of hydrocephalus (i.e., idiopathic congenital, neural tube defects, and others), and the *MAEL* association was not specific to any one specific etiology. Finally, as a Manhattan plot illustrates, *MAEL* was uniquely associated with hydrocephalus within the locus and across all genes tested in the frontal cortex (Figure 2C), suggesting that the *MAEL* association was not merely the result of LD with a different causal gene within the locus. The *MAEL* association was also observed in the hypothalamus (Figures 2D–2F), based on the total number of genes tested in this tissue. In both frontal cortex and hypothalamus, decreased expression of *MAEL* conferred increased predisposition to hydrocephalus. A complete list of all SNPs (and their relative weights) in the *MAEL cis* region (i.e., within 1 Mb) included in the PrediXcan models for frontal cortex and hypothalamus are included in Table S1.

Interestingly, gene expression differences between cases and controls in whole blood, the tissue most often acquired for human genetics studies due to easy accessibility and low cost, showed no significant departure from null expectation (Figures 2G–2I) despite its much larger sample size and, hence, statistical power, demonstrating the importance of considering tissue-specific expression differences in relevant tissues when performing genetic screens of hydrocephalus. However, whole-blood analysis revealed nominally significant associations with previously reported candidate

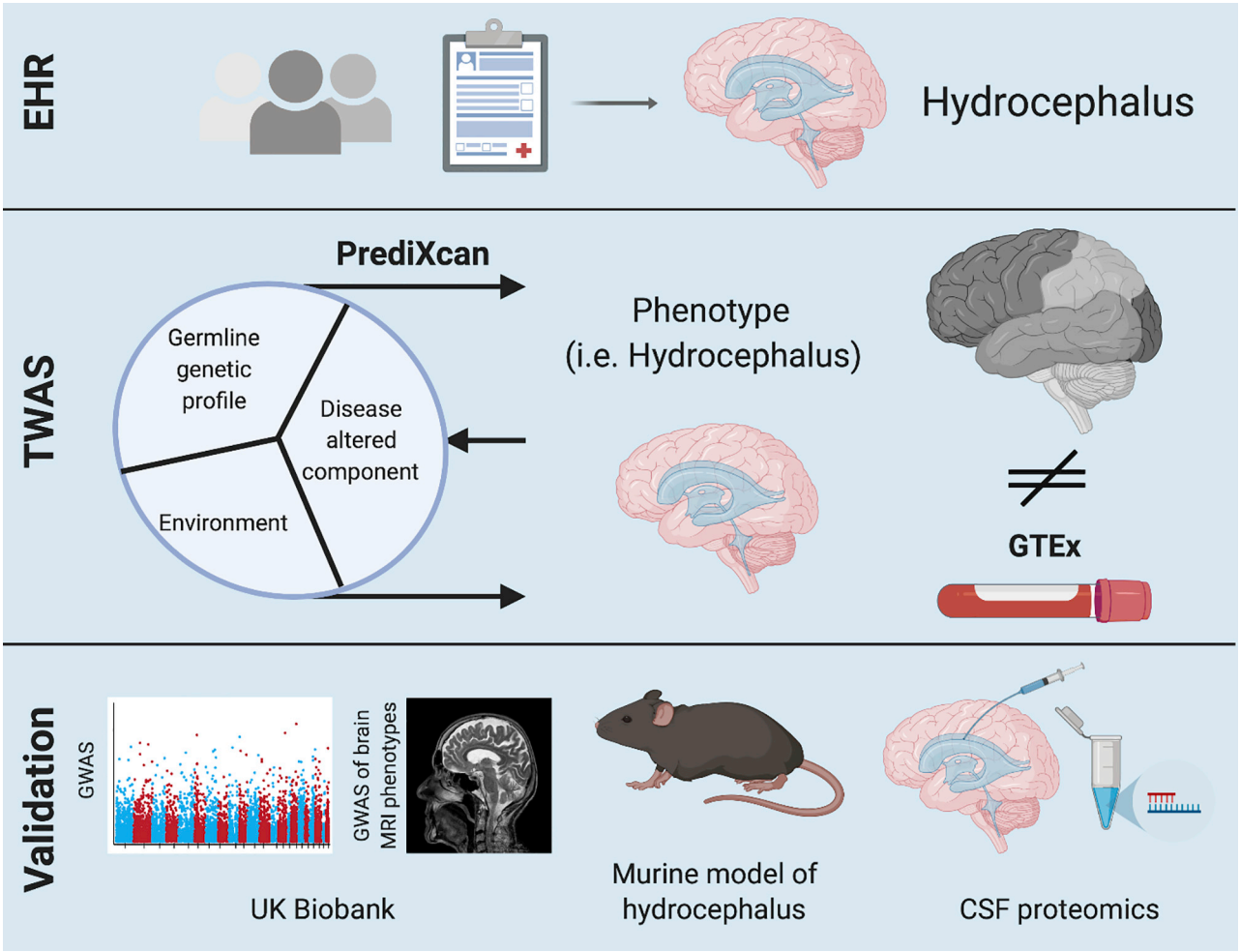


Figure 1. Overview of our approach

We leverage a large DNA biobank, BioVU (Roden et al., 2008), linked to deidentified electronic health record (EHR) data. We applied PrediXcan (Gamazon et al., 2015, 2018), which estimates the tissue-specific, genetically determined component of gene expression (i.e., the "germline genetic profile" of the gene expression trait) based on common variants (minor allele frequency >1%) and imputation against a reference transcriptome panel. For this study, we used GTEx transcriptome data in 10 neurological tissues and whole blood (Battle et al., 2017). The genetic component of expression was then tested for association with phenotype to identify gene-level associations. We performed systematic validation using independent replication of genetic results in the UK Biobank (Bycroft et al., 2018), analysis of structural brain magnetic resonance imaging (MRI) phenotypes in the UK Biobank (Elliott et al., 2018), and analysis of choroid plexus isolated from a mouse model of hydrocephalus (Robledo et al., 2008) and compared genetically determined gene expression changes to proteomic analysis of cerebrospinal fluid (CSF) isolated from infants with hydrocephalus compared to non-affected controls. A summary of each phenotype and the corresponding sample size can be found in Table S4.

genes—*CCDC28A* (Cardenas-Rodriguez et al., 2013), *MARCKS* (Chen et al., 1996; Lang et al., 2006), and *TNFRSF10D* (Habiyaremye et al., 2017; Jiménez et al., 2014)—that have been independently implicated in human and mouse studies of hydrocephalus.

The list of genes nominally associated ($p < 0.05$) with hydrocephalus disease status and the distribution of association p values varied across neurological tissues (Figure 3A; Data S1). Gene expression differences between cases and controls in the frontal cortex were the most significant (Figure 3A). The distribution of gene effect sizes on hydrocephalus differed across neurological tissues (Figure 3B), highlighting the importance of examining tissue-specific effects on disease risk in neurological tissues. We performed hierarchical clustering of nominally associated genes across neurological tissues (Figure 3C). Using multiscale bootstrap resampling (with "average" as the agglomerative method, "correlation" as distance method, and 1,000 bootstrap replicates; Shimodaira, 2004), we found that, in contrast to the genes nominally associated with hydrocephalus ($p < 0.05$), randomly sampled genes did not lead to stable clusters, i.e., the hypothesis "the cluster does not exist" cannot be rejected at the significance level of 0.05 (with an "approximately unbiased" value of zero), and do not recapitulate known relationships between Genotype-Tissue Expression (GTEx) tissues (between the cerebellar hemisphere and cerebellum and between the frontal cortex and cortex) (Battle et al., 2017). The gene-level associations (nominal $p < 0.05$) are highly tissue-specific, with the majority detected in one tissue

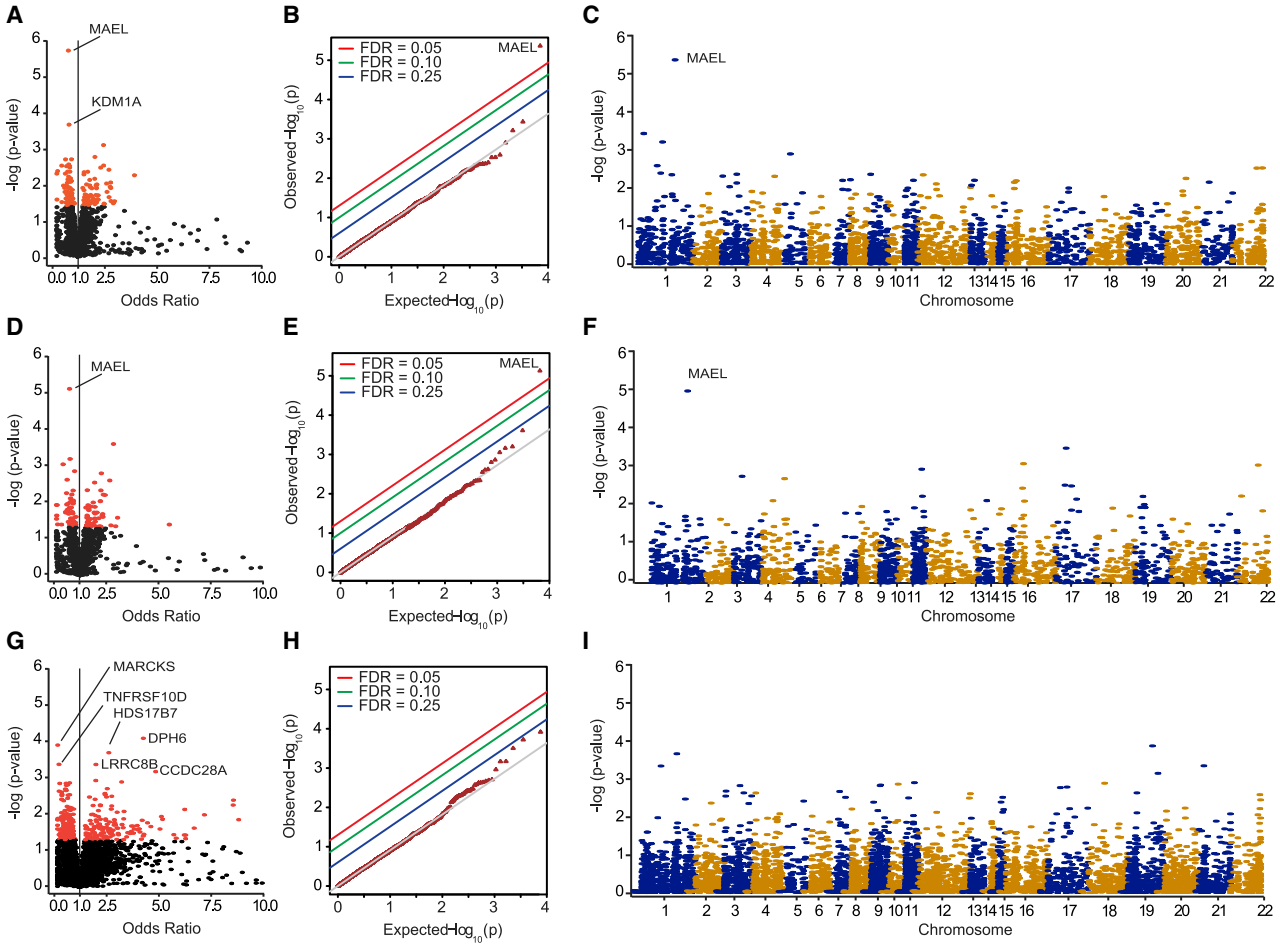


Figure 2. Genome-wide scan identifies tissue-specific gene-level associations with hydrocephalus

(A) Volcano plot showing odds ratio (OR, x axis) versus −log (p value, y axis) for gene expression differences between cases and controls in the frontal cortex.

(B) Q-Q plot demonstrating a significant association between *MAEL* and hydrocephalus after correction using Benjamini-Hochberg FDR in the frontal cortex. *MAEL* is study-wide significant (adjusted $p < 0.05$) after Bonferroni adjustment for the number of gene-tissue pairs tested in the study.

(C) Manhattan plot showing the gene-level association p values and chromosomal location of the signal from *MAEL* in the frontal cortex. The experiment-wide significant gene *MAEL* is also the unique gene in the *cis* region with a nominal gene-level association ($p < 0.05$) with hydrocephalus.

(D–I) Analogous analyses were performed in hypothalamus tissue (D–F) and whole blood (G–I).

See also Data S1.

(Figure 3D). We provide a list of all nominally significant ($p < 0.05$) genes in neurological tissues for reference (Data S1).

Contrary to MAEL's most commonly studied role in spermatogenesis, *MAEL* is actually expressed in both males and females, albeit absent in female sex organs, and is highest in the frontal cortex (the neurological tissue deviating the most from null expectation, Figure 1), outside of the testis (Figure 4A). Interestingly, MAEL-deficient mice have been generated, but they do not develop hydrocephalus (Soper et al., 2008). However, since *MAEL* expression is one of the most highly tissue-specific genes (in expression profile) in the genome (Figure 4B; see STAR Methods) and because MAEL function differs substantially across species (Chen et al., 2015), further mechanistic validation of MAEL in hydrocephalus needs to be performed specifically on human neurological tissue samples, building on our results from human GWASs.

Since the experiment-wide significant genetic signal to emerge from our analyses was decreased expression of *MAEL* in the frontal cortex, we present the canonical model for *MAEL* function (Figure 4C). MAEL is a critical regulator of PIWI-interacting RNA (piRNA)-mediated repression of transposable elements and methylation (Soper et al., 2008). Thus, decreased MAEL leads to increased transposon mobilization and differential methylation patterns, leading to broad changes in gene expression. However, the specific genes in the brain targeted by MAEL-mediated transposon movement and histone modification H3K9me3, which plays a role in targeting DNA methylation (Lehnertz et al., 2003), are unknown. Thus, our study implicates transposon and histone modification (trimethylation) in the pathophysiology of hydrocephalus; however, additional molecular studies are needed to confirm this finding.

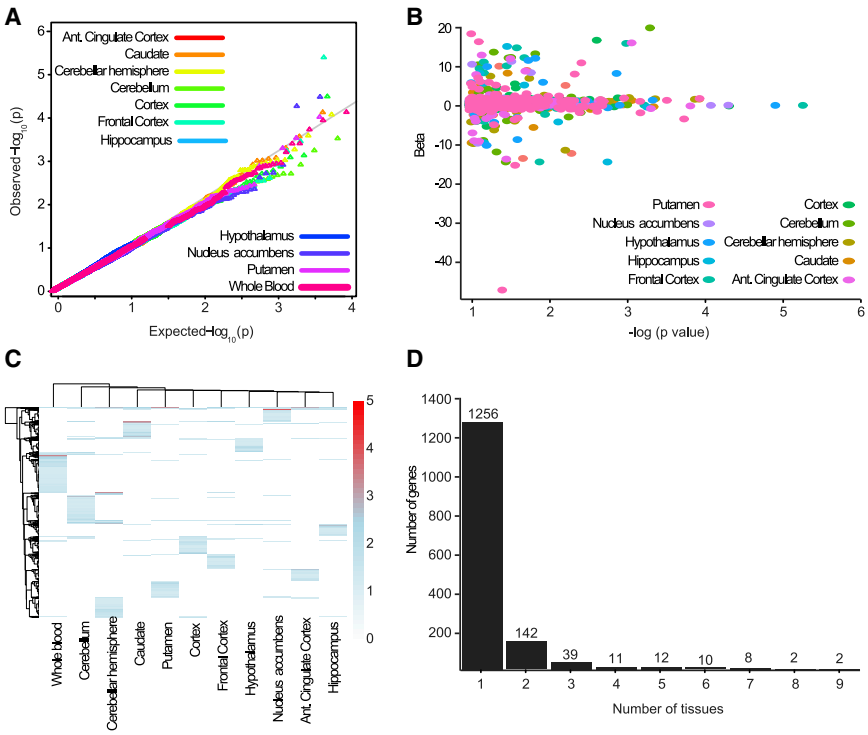


Figure 3. Differentially expressed genes in neurological tissues and whole blood

(A) Gene-level associations (PrediXcan) with hydrocephalus status, determined by logistic regression (with sex and age and the genotype-based principal components as covariates), in each neurological tissue and whole blood, including genes that depart from null expectation. *MAEL* expression in frontal cortex was experiment-wide significant (Bonferroni-adjusted $p < 0.05$) across all tissue-gene pairs tested.

(B) Significance and effect size of gene-level associations in neurological tissues identifying outliers. Gene associations within each neurological tissue are color coded.

(C) Hierarchical gene clustering of the nominally significant gene-level associations ($p < 0.05$), with whole blood an outlier relative to the neurological tissues.

(D) Number of tissues in which gene-level nominal associations ($p < 0.05$) with hydrocephalus are detected.

See also Data S1.

Replication in the UK Biobank

Applying PrediXcan to brain magnetic resonance imaging (MRI) white matter and total brain volume data in the UK Biobank (Elliott et al., 2018; Miller et al., 2016) (N = 8,428; see STAR Methods), we found that *MAEL* expression in the frontal cortex was significantly associated with white matter volume ($p = 0.011$) and total brain volume ($p = 0.015$) at Bonferroni-adjusted $p < 0.05$. Various structural brain alterations have been implicated in the pathophysiology of hydrocephalus (Del Bigio, 2010). Remarkably, we found a significant enrichment for gene-level associations with white matter (Figure 5A) and total brain volume (Figure 5B) among the top differentially expressed genes ($p < 0.05$) in frontal cortex.

For additional support, we analyzed SNPs within the *MAEL cis* region (i.e., the *MAEL* locus that is used in PrediXcan analysis) and their association with hydrocephalus in the UK Biobank. We conducted extensive quality control (QC) on the SNPs within the locus and excluded all low-confidence variants (see STAR Methods). We identified nine variants, in LD ($r^2 > 0.70$), nominally associated with hydrocephalus ($p < 0.05$, Table S2). The most significant SNP (rs75008967) overlaps an enhancer element in fetal brain (male and female) and in several brain regions (Figure S1, H3K27ac) (Creyghton et al., 2010), including hippocampus, substantia nigra, anterior caudate, cingulate gyrus, and dorsolateral prefrontal cortex, based on Roadmap chromatin immunoprecipitation sequencing (ChIP-seq) data (Ward and Kellis, 2012). In addition, a variant in the *cis* region of *MAEL* (rs72687818), independent of the lead SNP above ($r^2 < 0.20$), was the most significant association with a common cause of hydrocephalus, namely "subarachnoid hemorrhage from intracranial artery" (Germanwala et al., 2010; Graff-Radford et al., 1989) in the UK Biobank ($p = 1.1 \times 10^{-20}$). We evaluated the gene-level significance from the SNP-level association results in the *cis* region with hydrocephalus ($p = 0.03$) and with subarachnoid hemorrhage ($p = 4.77 \times 10^{-12}$), strongly confirming the discovery signal. Collectively, these results provide robust support to the importance of the *MAEL* locus for hydrocephalus predisposition.

MAEL-mediated associations with hydrocephalus-related neurological traits

To explore how genetically determined expression of *MAEL* may exert its phenotypic effect on hydrocephalus predisposition in these specific neurological tissues, we identified a cohort of patients in BioVU with related neurological traits that may facilitate further insights into hydrocephalus pathophysiology (see STAR Methods). Consistent with the hydrocephalus associations, decreased MAEL expression in frontal cortex and hypothalamus was found to be associated with "cerebral edema and compression of brain" ($p = 0.0021$ in frontal cortex and $p = 0.0018$ in hypothalamus), "other cerebral degenerations" ($p = 0.0009$ in frontal cortex and $p = 0.001$ in hypothalamus), and intracerebral hemorrhage ($p = 0.016$ in frontal cortex and $p = 0.021$ in hypothalamus). These data are consistent with the association of *MAEL* expression with regulation of white matter and total brain volumes obtained through analysis of brain MRI phenotypes linked to genetic information (Figure 5) in the independent UK Biobank.

In addition, STRING analysis (version 10.5; Szklarczyk et al., 2017) revealed a number of direct interacting partners that may play a role in *MAEL*-mediated hydrocephalus pathophysiology, offering insights into potential targets for intervention (Figure S2A). Notably, the other genes in the proposed biological network from the STRING analysis showed nominally

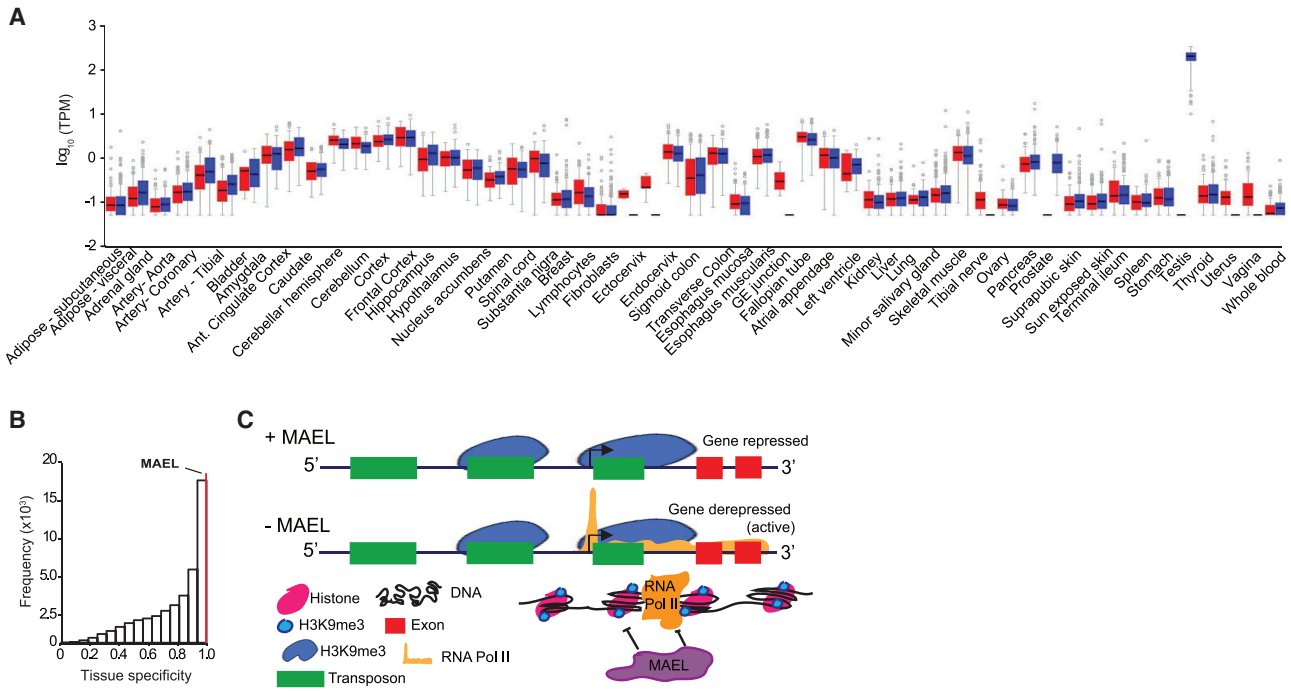


Figure 4. *MAEL* expression profile and model of *MAEL*-mediated trait effect

(A) Population-based expression of *MAEL* in males (blue) and females (red) across 44 tissues included in GTEx. *MAEL* is most highly expressed in frontal cortex among brain regions and displays low expression in whole blood. Box edges show interquartile range, whiskers 1.5 × the interquartile range, and center lines the median.

(B) Tissue specificity (x axis) as quantified by τ (see STAR methods) versus frequency of genes (y axis) across the genome. *MAEL* is one of the most tissue-specific genes in the genome (τ = 0.997).

(C) Canonical model of *MAEL*-mediated alteration of transposon movement and depression of H3K9me3, leading to recruitment of RNA polymerase II (RNA Pol II) and transcription of previously repressed genes.

significant associations with neurological phenotypes (Figure S2B) in BioVU (see STAR Methods), including "type 2 diabetes with neurological manifestations" and "polyneuropathy in diabetes for decreased TDRKH expression" ($p = 7.9 \times 10^{-4}$ and $p = 2.2 \times 10^{-3}$, respectively) in whole blood and multiple sclerosis for increased *DDX4* expression ($p = 0.017$) in frontal cortex.

To assess the significance of the observed *MAEL*-mediated neurological trait associations, we performed permutation analysis (n = 1,000) that preserves the pairwise (gene-gene) correlation (within each tissue) of the estimated genetic component of expression (as well as the gene count) among the tested genes. We observed a significant enrichment (empirical $p < 0.001$) for associations with neurological phenotypes among *MAEL*'s direct interacting partners. Thus, there may be some shared underlying risk for other neurological disorders in patients with alteration of *MAEL*-dependent biological networks, which lends additional support for (as noted above) further validation studies specifically in humans to elucidate *MAEL*'s role in hydrocephalus risk.

Transcriptome analysis and validation in a murine model of hydrocephalus

Many laboratories focus on the use of model organisms in studying hydrocephalus for pre-clinical translational studies. Thus, we conducted differential expression analysis (see STAR Methods) of ChIP-based gene expression data from the choroid plexus of α-adducin knockout mice, which develop hydrocephalus secondary to IVH (Robledo et al., 2008), one of the most common causes of hydrocephalus. Interestingly, there was a marked degree of concordance in both the genes involved and the direction of effect between the transcriptome of α-adducin knockout mouse choroid plexus and tissue-specific gene expression differences observed in our human cohort (Table S2). Although *MAEL* was validated in human GWASs but not in this murine model, this is consistent with the substantial evidence describing the species and tissue-specificity of *MAEL* gene expression and regulation (Chen et al., 2015; Genzor and Bortvin, 2015). Finally, nominally significant associations for *TMEM50B* and *KCTD21* (in frontal cortex and hypothalamus) were replicated in mice with a concordant direction of effect (Table S3).

Genetically determined pathway associations with hydrocephalus

To probe relevant biological processes and molecular functions, we utilized the full set of nominally significant associations ($p < 0.05$) of genetically determined expression with hydrocephalus in each neurological tissue, across all tissues, and in whole blood (Data S1). We hypothesized that pathway analysis of nominally significant genetically determined expression changes (see

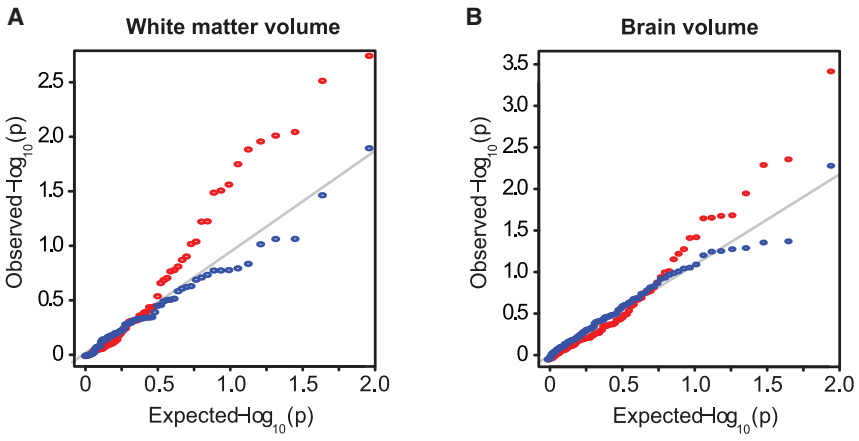


Figure 5. Genomic analysis of brain MRI data in the independent UK Biobank validates gene-level associations with hydrocephalus in the same (significant) discovery tissue

Using PrediXcan analysis of brain imaging and genomic data, we validated the study-wide significant association of *MAEL* in frontal cortex (Bonferroni-adjusted replication $p < 0.05$). We then considered the associations with the imaging-based phenotypes of the top differentially expressed genes in frontal cortex.

(A) For the hydrocephalus-associated genes ($p < 0.05$; in red) in the frontal cortex from the BioVU analysis, the Q-Q plots show the PrediXcan p values for their association with white matter volume (A) and total brain volume (B) in the UK Biobank. The departure from the diagonal line indicates enrichment for gene-level associations with the imaging-based phenotypes among the hydrocephalus-associated genes. For comparison, a Q-Q plot for a random set of genes (of equal count; in blue) is included.

STAR methods) would confirm a number of pathways and biological processes previously associated with hydrocephalus, and identify additional pathways associated with the disease. This approach is in contrast to conventional approaches for differential expression analysis (which utilize the total, directly measured gene expression), enabling us to identify "genetically determined" and potentially "causative," rather than consequential disease-relevant networks (see Figure 1).

We performed gene set enrichment analysis (GSEA) (Subramanian et al., 2005) using the Molecular Signatures Database (MSigDB) on nominally significant genetically determined gene expression changes in the frontal cortex, hypothalamus, whole blood (Data S2 and S3), and all other neurological tissues individually as well as cross-tissue analysis (Data S2 and S3, which present significantly enriched gene sets at a false discovery rate [FDR] of <0.05 in each tissue), as has been previously applied to GWAS data (Wang et al., 2007, 2010). Experiment-wide significant gene sets were identified using Bonferroni correction ($p < 2.9 \times 10^{-7}$) (see STAR Methods). The level of significance for a gene set is not correlated with the number of genes in the set (e.g., Spearman correlation p = 0.68 in frontal cortex). Many of our top gene-level associations were identified in two or more neurological tissues (Figure S3; Figure 2D; Data S2 and S3). GSEA/MSigDB (with default options) was performed on the 225 genes that were nominally significant in two or more tissues as the input set (Data S2 and S3), recapitulating the involvement of many pathways identified by single-tissue analysis.

Exome scan and functional genomics suggest potentially pathogenic variants associated with hydrocephalus

We analyzed exome data for additional support for the role of the top differentially expressed genes in hydrocephalus risk. To this end, we tested the nominally significant hydrocephalus-associated genes detected in at least five tissue types (Figure S3; see STAR Methods) for coding variation effects on disease susceptibility, using a cohort of 29,713 patients (Figure S4A; Data S4). *TMEM50B*, which was also supported in our transcriptome analysis of a murine model of hydrocephalus (Table S3), contains a rare missense variant (rs34327244 or A139T, minor allele frequency [MAF] = 0.4% in Europeans [non-Finnish]) with nominally significant association with hydrocephalus ($p = 7.4 \times 10^{-3}$) (Figure S4B). After Benjamini-Hochberg correction, this association corresponded to a FDR of 11%, warranting additional functional studies given the convergent evidence for the gene. Interestingly, we found that rs34327244 was associated with CSF volume (normalized for overall head size) in the independent UK Biobank (p = 0.016).

In addition, using ChIP-seq data derived from human embryonic stem cells (HUES64) as part of Roadmap Epigenomics (Leung et al., 2015), we found that A139T disrupts an active or primed enhancer element marked by monomethylation of histone H3 at lysine 4 (H3K4me1), suggesting a role for the variant in regulation of transcription. Furthermore, *TMEM50B* is co-expressed with aquaporin 1 (*AQP1*) in the frontal cortex (Figure S4C), and A139T alters a regulatory motif that leads to differential allelic affinity of the NKx-2 homeodomain containing transcription factor thyroid transcription factor 1 (TTF1) (Ward and Kellis, 2012). Since decreased expression of TTF1 leads to decreased expression of AQP1 in the apical membrane of the choroid plexus and alteration in CSF formation (Kim et al., 2007), it is possible that A139T leads to decreased availability of TTF1 to promote *AQP1* expression (Figure S4D), although additional experimental validation is needed. These studies provide preliminary evidence for aquaporin dysregulation in genetically determined risk for hydrocephalus.

CSF proteomic data from patients with hydrocephalus mirror PrediXcan results

Next, we analyzed proteomic data from CSF isolated from patients with hydrocephalus secondary to IVH, from a previously published study (Morales et al., 2012) for mass spectroscopic analysis (see STAR Methods), and compared the differentially expressed proteins to the differentially expressed genes identified by PrediXcan. This analysis revealed a significantly greater overlap between differentially expressed genes in the frontal

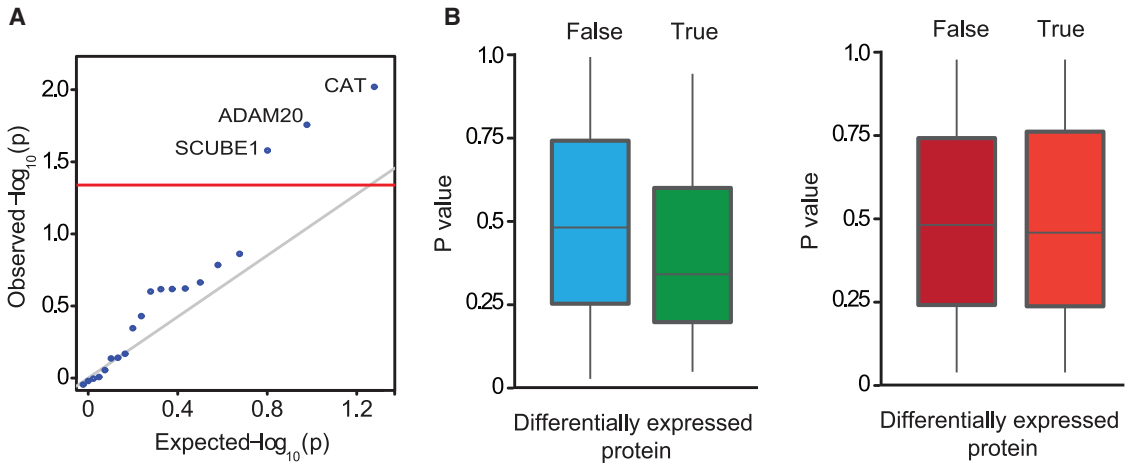


Figure 6. CSF proteomic signature of patients with hydrocephalus recapitulates gene expression associations identified by PrediXcan

(A) Q-Q plot showing significance for three proteins implicated by LC-MS analysis of CSF isolated from patients with hydrocephalus in the association of their genetically determined expression in the frontal cortex ($-\log_{10}$ p value from the PrediXcan analysis shown on the y axis).

(B) PrediXcan p values in frontal cortex of the proteomic signature from CSF (true) versus the remaining proteins (false), demonstrating greater statistical significance (i.e., lower p value from PrediXcan) for the proteins in the CSF signature in frontal cortex (left, p = 0.04) but not in whole blood (right, p = 0.83). Significance was assessed using a Mann-Whitney U test. Box edges show interquartile range, whiskers 1.5 × the interquartile range, and center lines the median.

cortex and the proteomic signature identified by liquid chromatography-mass spectrometry (LC-MS) than expected by chance (enrichment p = 0.016), identifying three proteins that deviate from null expectation for differential expression in frontal cortex between cases and controls (Figure 6A). The most significant of the three proteins, catalase (CAT), showed a 500-fold decreased expression ($p = 1.56 \times 10^{-48}$) between patients with hydrocephalus versus controls. CSF protein levels of ADAM20 and SCUBE1 also showed nominal association with hydrocephalus (PrediXcan $p < 0.05$).

The proteomic signature identified in the CSF tended to have significantly lower (i.e., more statistically significant) p values from the PrediXcan analysis of frontal cortex (Mann-Whitney U test, p = 0.04, Figure 6B, left) than the remaining proteins. In contrast, we observed no significant expression difference between cases and controls in whole blood (Mann-Whitney U test, p = 0.83, Figure 6B, right) for the proteins identified by LC-MS, further showing the limitation of whole blood as a substrate for hydrocephalus genetic analyses. Overall, these results demonstrate that while CSF proteomic analyses may be useful in identifying biomarkers for prognostic purposes, its use, without integration of genetic data, as a discovery platform of genetically determined pathogenic mechanisms is limited.

DISCUSSION

We describe the largest genetic study of hydrocephalus and identify trait-associated genes and signaling pathways, laying the groundwork for molecular studies of hydrocephalus. We report genetically determined contributions to hydrocephalus, the functional consequences of identified genes on brain structure, the transcriptomic signature of differentially expressed genes in choroid plexus of a murine model of hydrocephalus, and hydrocephalus-associated genetically determined expression changes for proteins as measured by an unbiased proteomic screen of CSF isolated from patients with hydrocephalus. The use of genetic information can be used to disentangle changes in expression that influence disease versus secondary expression changes as a result of hydrocephalus, a question that has remained elusive for more than 30 years (Del Bigio, 1989).

PrediXcan imputes the genetic component of gene expression in tissues from which it is nearly impossible to obtain clinical samples, such as neurological tissues. We find a remarkable degree of tissue specificity in gene regulation in hydrocephalus. However, the degree of tissue, cell-type, and single-cell variation in hydrocephalus has yet to be fully appreciated. While common variants and *de novo* mutations have been shown to play a role in a range of neurodevelopmental disorders, including hydrocephalus (Niemi et al., 2018; Short et al., 2018; Furey et al., 2018), this study implicates common variant-mediated regulation of tissue-specific gene expression as a potential driver of hydrocephalus.

Our analysis revealed a significant association between differential expression of *MAEL* and hydrocephalus status. MAEL contains two domains, that is, (1) a high mobility group (HMG)-box domain and (2) an RNase H-fold domain that lacks catalytic residues conserved in RNA-H nucleases, but it displays single-stranded RNA (ssRNA)-specific endonuclease activity (Genzor and Bortvin, 2015; Matsumoto et al., 2015). MAEL is essential for piRNA-mediated transcriptional transposon silencing. Interestingly, transposons are DNA sequences capable of changing their position in the genome, with the potential to induce mutations, as well as change a cell's identity and genomic size. Loss of maelstrom, the MAEL homolog in *Drosophila melanogaster*, has been shown to perturb RNA polymerase II recruitment, nascent RNA output, and steady-state RNA levels of transposons, leading to increased heterochromatin spreading despite modest changes in H3K9me3 patterns (Sienski et al., 2012),

which suggests that MAEL may act independently or downstream of H3K9me3. Consistent with this function, we identified H3K9me3-dependent changes as one of the most significantly enriched curated gene sets for differentially expressed genes (between hydrocephalus cases and controls) across neurological tissues (Data S2 and S3). However, the expression of *MAEL* is one of the most tissue-specific in the genome (Figure 4B), and a human-specific role for *MAEL* in neurological disorders has not previously been described. Thus, additional molecular studies of *MAEL* in hydrocephalus risk and pathogenesis should be performed in human neurological tissue.

Epigenetic modification (H3K9me3) underlying the *MAEL* association with hydrocephalus is supported by mechanistic data on the role of piRNA biogenesis factor Mili (a target of MAEL) in mice. Mili-deficient mice demonstrate broad changes in CpG hypomethylation across the genome (Nandi et al., 2016). Extending this observation, one of the significantly enriched gene sets (FDR < 0.05) across neurological tissues with hydrocephalus was alteration in CpG methylation (Data S2 and S3), providing convergent functional evidence for the potential role of these pathways in human hydrocephalus. Furthermore, deletion of PIWI (another piRNA biogenesis factor) inhibits axon regeneration that is dependent on the slicer domain of PIWI, indicating that post-transcriptional gene silencing may be involved (Kim et al., 2018). Since loss of MAEL in mice results in selective transposon insertions (Aravin et al., 2008), this is certainly a plausible mechanism. However, the complete repertoire of cell type-specific transposon insertion sites in humans remains largely unknown (Elbarbary et al., 2016). Furthermore, there is evidence that transposon elements are enriched in neural stem cells (Upton et al., 2015), a cell type widely hypothesized to be dysfunctional in hydrocephalus.

Collectively, PrediXcan analysis, rare-variant exome scan, a murine model of hydrocephalus, and genomic analysis of imaging-based brain structural phenotypes justify additional functional follow-up studies on the role of *TMEM50B* in hydrocephalus. Interestingly, the chromatin state annotation suggests that A136T (rs34327244) may play a role in transcriptional regulation. Indeed, ChIP-seq analysis shows that rs34327244 overlaps an enhancer region that is marked by monomethylation of H3K4me1. Analysis of transcription factor binding profiles revealed that rs34327244 alters a regulatory motif, resulting in differential allelic affinity for NKx-2 homeodomain containing transcription factor TTF1 (Ward and Kellis, 2012). Intriguingly, AQP1, a critical regulator of CSF formation and intracranial water movement (Iliff et al., 2012), is a direct transcriptional target of TTF1, and AQP1 transcript levels are directly correlated with TTF1 expression (Kim et al., 2007). Remarkably, we discovered that rs34327244 was associated with normalized CSF volume in the UK Biobank. These data suggest that in conferring hydrocephalus risk, *AQP1* dysregulation may be causative; however, additional detailed molecular studies are required to definitively clarify the role of *TMEM50B* in tissue-specific AQP1 regulation and hydrocephalus pathophysiology. Nonetheless, our findings suggest that integrated analysis of brain imaging and functional genomics (human and model organism) data can be used to identify a previously inaccessible molecular mechanism relevant to hydrocephalus pathophysiology.

Several functional and structural features on brain MRI have been associated with hydrocephalus (Del Bigio, 2010); however, the genetic basis for these observations is not known. A repository of brain imaging phenotypes from 8,428 individuals tied to GWAS data in the UK Biobank provides a unique resource to probe these questions (Elliott et al., 2018). Since CSF, white matter, and total brain volumes have been independently associated with hydrocephalus (Mandell et al., 2015), we considered the differentially expressed genes in the frontal cortex identified by PrediXcan (Figures 2A–2C; Data S1) and their PrediXcan associations with these imaging-derived traits. We observed a significant enrichment for top associations with these brain structural phenotypes among the hydrocephalus-associated genes. Testing the effect of hydrocephalus-associated genes on brain-imaging traits may be relevant to understanding both normative brain development and neurocognitive outcomes in hydrocephalus (Mandell et al., 2010, 2015; Peterson et al., 2018). Specifically for the aims of our study, understanding the role of these genes in conferring a risk for increased (e.g., macrocephaly) and decreased (e.g., cerebral degeneration) brain volume enhances our mechanistic understanding of hydrocephalus risk, as brain volume influences CSF circulation (Brinker et al., 2014). Alterations in white matter volume have also been observed in congenital hydrocephalus (Lockwood Estrin et al., 2016), suggesting a shared genetic component. Collectively, our data suggest that hydrocephalus-associated genes may exert their effect on disease risk through their role in regulating brain structure and integrity.

Human and murine studies have demonstrated the importance of the CSF proteome in neurodevelopmental disorders as well as in hydrocephalus. We determined that decreased catalase (*CAT*) gene expression in the frontal cortex had a consistent effect on hydrocephalus as decreased CAT protein in the CSF (Figure 6A). The canonical role of CAT is to cleave hydrogen peroxide into water and oxygen, an important biological process mediating production of reactive oxygen species. CAT has been shown to regulate DNA damage, which causes chromosomal aberrations at regions of the genome centered around transposons (Argueso et al., 2008). These data provide potential links to chromosomal modifications in hydrocephalus pathophysiology.

CSF-derived signals have been shown to play a major role in mediating neural tube closure by signaling to the neural stem cell niche (Chau et al., 2015). In addition, the CSF proteome mediates the localization of Igf1R to the apical membrane of the choroid plexus, in part through regulation by PTEN (Lehtinen et al., 2011). Interestingly, CSF from patients with glioblastoma multiforme (GBM) selectively alters this response (Lehtinen et al., 2011), suggesting that underlying genetic risk may lead to selective development of hydrocephalus. Notably, PTEN signaling is recapitulated here by pathway analysis of genetically determined expression changes, consistent with previous reports on the role of PTEN/phosphatidylinositol 3-kinase (PI3K) signaling in hydrocephalus (Kousi and Katsanis, 2016; Yung et al., 2011; Zheng et al., 2018). Age-dependent alterations in the CSF proteome can also influence adult neural stem cells (Silva-Vargas et al., 2016), which could potentially underlie development of normal pressure hydrocephalus (NPH) in adults. Interestingly, there is a correlation with neurodegeneration-associated proteins in both CSF and cortical biopsies with NPH,

suggesting some overlapping pathophysiology (Jeppsson et al., 2016; Leinonen et al., 2012).

It has already been demonstrated that alterations in choroid plexus gene expression (Lun et al., 2015a) and the neural stem cell niche (Carter et al., 2012; Furey et al., 2018) play significant roles in hydrocephalus. The relationship between neuronal and choroid plexus development, CSF dynamics, and genetic regulation underpin a highly complex physiological system with multiple modes of regulation (Lun et al., 2015b). Thus, while we were not able to directly analyze epithelial cells of the choroid plexus and cells from the neural stem cell niche in the ventricular/subventricular zone from humans, we identified a number of pathways in those compartments that had previously been associated with hydrocephalus through targeted molecular and genetic studies. For instance, analysis of genetically determined expression changes in frontal cortex revealed protein biogenesis as being significantly associated with hydrocephalus. Notably, regulation of the genes encoding the protein biosynthetic machine has been shown to be essential in forebrain development and development of macrocephaly (Chau et al., 2018). These authors showed mechanistic evidence detailing mTOR and MYC signaling as critical regulators of forebrain development, recapitulated here through pathway analysis of genetically determined expression changes (Data S2 and S3). Furthermore, mTOR signaling-mediated regulation of cilia has been shown to be required for ventricular morphogenesis (Foerster et al., 2017), further highlighting the potential importance of mTOR in hydrocephalus.

Our PrediXcan analysis is currently limited to common variant-mediated gene expression. However, as our understanding of the contribution of rare variants to human genetic regulation increases (with the requisite increase in study sample sizes and investment in functional genomic studies in diverse ancestries; Zhong et al., 2019), so too will our comprehension of genetically determined expression across the full range of allele frequency spectrum. Despite this limitation, our analysis of rare variants in the most differentially expressed genes in several neurological tissues provides another layer of functional evidence for the contribution of these genes to disease susceptibility.

Electronic health records linked to DNA biobanks, as vast repositories of disease and medication data, will enable rapid discovery and replication of genetic associations, as UK Biobank replication studies in this study have provided support for the role of *MAEL* in conferring hydrocephalus risk. Finally, validation studies of the implicated genetic components of gene expression in related phenotypes—such as the association between decreased *MAEL* expression and "cerebral edema and compression of brain" and "other cerebral degenerations" as well as the significant enrichment for neurological trait associations among *MAEL*'s direct interacting partners—may provide additional mechanistic insights into hydrocephalus pathophysiology.

We present the largest genetic analysis of hydrocephalus and extensive genomic analyses to identify hydrocephalus-associated genes and pathways. Our study highlights the complexity of hydrocephalus pathophysiology and the polygenicity of its genetic architecture. We propose transposon-mediated genetic regulation through *MAEL* for future mechanistic validation in hydrocephalus. In addition, we identify a potential molecular basis for the role of aquaporin in hydrocephalus risk, a mechanism long hypothesized to be involved in human hydrocephalus. We integrate PrediXcan analysis of imaging-based structural brain phenotypes related to hydrocephalus and demonstrate an enrichment of genes associated with alterations in brain structure and integrity among risk genes. Finally, we observed enrichment for germline-genetic expression changes in neurological tissues among the differentially expressed proteins in CSF between hydrocephalus patients and controls, demonstrating the value of our methodology for the discovery and characterization of the genetic determinants of hydrocephalus.

Limitations of study

We conducted a transcriptome-wide association study (TWAS) of hydrocephalus in 10 brain regions and whole blood. Among the top differentially expressed genes in the brain, we observed an enrichment for gene-level associations with neuroimaging phenotypes, indicating an effect on disease risk through regulation of brain structure and integrity. The molecular mediators that underlie the functional and structural brain variability contributing to disease risk require detailed follow-up studies. Choroid plexus transcriptome analysis of a murine model of hydrocephalus and proteomic analysis of CSF isolated from patients provide additional evidence for the top differentially expressed genes that reach only nominal significance. A larger sample size will improve the resolution to detect disease-associated genes. In the present study, the gene *MAEL* in frontal cortex attained experiment-wide statistical significance in BioVU with additional support in the UK Biobank. Future studies should be aimed at elucidating the underlying molecular mechanism in human tissue samples, given the high degree of evolutionary divergence for the gene in the tissues (i.e., neurological) of interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Estimating the genetically determined expression
  - Genetically determined pathways and networks
  - BioVU analysis of hydrocephalus-related neurological traits
  - UK Biobank replication analysis
  - Differential expression in mice
  - Rare variant exome scan
  - CSF proteomic data from patients with hydrocephalus

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109085.

ACKNOWLEDGMENTS

A.T.H. is supported by the Vanderbilt University Medical Scientist Training Program (T32GM007347) and the National Institutes of Health (F30HL143826 and R35HG010718). E.R.G. is supported by the National Institutes of Health under award numbers R35HG010718, R01HG011138, R01GM140287, and R01HL133559. Funding sources for BioVU are listed at https://victr.vanderbilt.edu/pub/biovu/. E.R.G. has also significantly benefitted from a Fellowship at Clare Hall, University of Cambridge (UK) and is grateful to the President and Fellows of the college for a stimulating intellectual home. Cerebrospinal fluid proteomic analysis was supported by the Washington University Institute of Clinical and Translational Sciences grant UL1TR002345 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). D.D.L. is also supported by a career development award (NIH/NINDS K23NS075151). A.T.H. and E.R.G. would also like to thank the Vanderbilt Institute for Clinical and Translational Research (VICTR) for the micro-grant (VR52639) supporting processing of imputation data. A.T.H. would like to thank the Surgical Outcomes Center for Kids (SOCKs) at Monroe Carell Jr. Children's Hospital for administrative support.

AUTHOR CONTRIBUTIONS

Conceptualization, A.T.H. and E.R.G.; methodology, A.T.H., L.B., D.D.L., and E.R.G.; investigation, A.T.H., L.B., D.M.M., D.D.L., and E.R.G.; writing – original draft, A.T.H. and E.R.G.; writing – review and editing, A.T.H., J.C.W., D.D.L., and E.R.G.; funding acquisition, A.T.H. and E.R.G.; supervision, E.R.G.

DECLARATION OF INTERESTS

E.R.G. receives an honorarium from the journal *Circulation Research* of the American Heart Association as a member of the Editorial Board.

Received: February 28, 2019
Revised: July 1, 2019
Accepted: April 14, 2021
Published: May 4, 2021

REFERENCES

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol. Cell *31*, 785–799.

Argueso, J.L., Westmoreland, J., Mieczkowski, P.A., Gawel, M., Petes, T.D., and Resnick, M.A. (2008). Double-strand breaks associated with repetitive DNA can reshape the genome. Proc. Natl. Acad. Sci. USA *105*, 11845–11850.

Battle, A., Brown, C.D., Engelhardt, B.E., and Montgomery, S.B.GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

Brinker, T., Stopa, E., Morrison, J., and Klinge, P. (2014). A new look at cerebrospinal fluid circulation. Fluids Barriers CNS *11*, 10.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.

Cardenas-Rodriguez, M., Osborn, D.P., Irigoín, F., Graña, M., Romero, H., Beales, P.L., and Badano, J.L. (2013). Characterization of CCDC28B reveals its role in ciliogenesis and provides insight to understand its modifier effect on Bardet-Biedl syndrome. Hum. Genet. *132*, 91–105.

Carter, C.S., Vogel, T.W., Zhang, Q., Seo, S., Swiderski, R.E., Moninger, T.O., Cassell, M.D., Thedens, D.R., Keppler-Noreuil, K.M., Nopoulos, P., et al. (2012). Abnormal development of NG2^{+}PDGFR-α^{+} neural progenitor cells leads to neonatal hydrocephalus in a ciliopathy mouse model. Nat. Med. *18*, 1797–1804.

Castaneyra-Ruiz, L., Morales, D.M., McAllister, J.P., Brody, S.L., Isaacs, A.M., Strahle, J.M., Dahiya, S.M., and Limbrick, D.D. (2018). Blood exposure causes ventricular zone disruption and glial activation in vitro. J. Neuropathol. Exp. Neurol. *77*, 803–813.

Chau, K.F., Springel, M.W., Broadbelt, K.G., Park, H.Y., Topal, S., Lun, M.P., Mullan, H., Maynard, T., Steen, H., LaMantia, A.S., and Lehtinen, M.K. (2015). Progressive differentiation and instructive capacities of amniotic fluid and cerebrospinal fluid proteomes following neural tube closure. Dev. Cell *35*, 789–802.

Chau, K.F., Shannon, M.L., Fame, R.M., Fonseca, E., Mullan, H., Johnson, M.B., Sendamarai, A.K., Springel, M.W., Laurent, B., and Lehtinen, M.K. (2018). Downregulation of ribosome biogenesis during early forebrain development. eLife *7*, e36998.

Chen, J., Chang, S., Duncan, S.A., Okano, H.J., Fishell, G., and Aderem, A. (1996). Disruption of the MacMARCKS gene prevents cranial neural tube closure and results in anencephaly. Proc. Natl. Acad. Sci. USA *93*, 6275–6279.

Chen, K.M., Campbell, E., Pandey, R.R., Yang, Z., McCarthy, A.A., and Pillai, R.S. (2015). Metazoan Maelstrom is an RNA-binding protein that has evolved from an ancient nuclease active in protists. RNA *21*, 833–839.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Del Bigio, M.R. (1989). Hydrocephalus-induced changes in the composition of cerebrospinal fluid. Neurosurgery *25*, 416–423.

Del Bigio, M.R. (2010). Neuropathology and structural changes in hydrocephalus. Dev. Disabil. Res. Rev. *16*, 16–22.

Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M., and Crawford, D.C. (2010). PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics *26*, 1205–1210.

Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat. Biotechnol. *31*, 1102–1110.

Derks, E.M., Zwinderman, A.H., and Gamazon, E.R. (2017). The relation between inflation in type-I and type-II error rate and population divergence in genome-wide association analysis of multi-ethnic populations. Behav. Genet. *47*, 360–368.

Elbarbary, R.A., Lucas, B.A., and Maquat, L.E. (2016). Retrotransposons as regulators of gene expression. Science *351*, aac7247.

Elliott, L.T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K.L., Douaud, G., Marchini, J., and Smith, S.M. (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature *562*, 210–216.

Foerster, P., Daclin, M., Asm, S., Faucourt, M., Boletta, A., Genovesio, A., and Spassky, N. (2017). mTORC1 signaling and primary cilia are required for brain ventricle morphogenesis. Development *144*, 201–210.

Furey, C.G., Choi, J., Jin, S.C., Zeng, X., Timberlake, A.T., Nelson-Williams, C., Mansuri, M.S., Lu, Q., Duran, D., Panchagnula, S., et al. (2018). De novo mutation in genes regulating neural stem cell fate in human congenital hydrocephalus. Neuron *99*, 302–314.e4.

Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., and Im,

H.K.; GTEx Consortium (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.

Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., et al.; GTEx Consortium (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat. Genet. *50*, 956–967.

Ge, T., Chen, C.Y., Neale, B.M., Sabuncu, M.R., and Smoller, J.W. (2017). Phenome-wide heritability analysis of the UK Biobank. PLoS Genet. *13*, e1006711.

Genzor, P., and Bortvin, A. (2015). A unique HMG-box domain of mouse Maelstrom binds structured RNA but not double stranded DNA. PLoS ONE *10*, e0120268.

Germanwala, A.V., Huang, J., and Tamargo, R.J. (2010). Hydrocephalus after aneurysmal subarachnoid hemorrhage. Neurosurg. Clin. N. Am. *21*, 263–270.

Graff-Radford, N.R., Torner, J., Adams, H.P., Jr., and Kassell, N.F. (1989). Factors associated with hydrocephalus after subarachnoid hemorrhage. A report of the Cooperative Aneurysm Study. Arch. Neurol. *46*, 744–752.

GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science *348*, 648–660.

GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

Habiyaremye, G., Morales, D.M., Morgan, C.D., McAllister, J.P., CreveCoeur, T.S., Han, R.H., Gabir, M., Baksh, B., Mercer, D., and Limbrick, D.D., Jr. (2017). Chemokine and cytokine levels in the lumbar cerebrospinal fluid of preterm infants with post-hemorrhagic hydrocephalus. Fluids Barriers CNS *14*, 35.

Iliff, J.J., Wang, M., Liao, Y., Plogg, B.A., Peng, W., Gundersen, G.A., Benveniste, H., Vates, G.E., Deane, R., Goldman, S.A., et al. (2012). A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. Sci. Transl. Med. *4*, 147ra111.

Jeppsson, A., Höltta, M., Zetterberg, H., Blennow, K., Wikkelsø, C., and Tullberg, M. (2016). Amyloid mis-metabolism in idiopathic normal pressure hydrocephalus. Fluids Barriers CNS *13*, 13.

Jiménez, A.J., Rodríguez-Pérez, L.M., Domínguez-Pinos, M.D., Gómez-Roldán, M.C., García-Bonilla, M., Ho-Plagaro, A., Roales-Buján, R., Jiménez, S., Roquero-Mañueco, M.C., Martínez-León, M.I., et al. (2014). Increased levels of tumour necrosis factor alpha (TNFα) but not transforming growth factor-beta 1 (TGFβ1) are associated with the severity of congenital hydrocephalus in the hyh mouse. Neuropathol. Appl. Neurobiol. *40*, 911–932.

Kahle, K.T., Kulkarni, A.V., Limbrick, D.D., Jr., and Warf, B.C. (2016). Hydrocephalus in children. Lancet *387*, 788–799.

Karimy, J.K., Zhang, J., Kurland, D.B., Theriault, B.C., Duran, D., Stokum, J.A., Furey, C.G., Zhou, X., Mansuri, M.S., Montejo, J., et al. (2017). Inflammation-dependent cerebrospinal fluid hypersecretion by the choroid plexus epithelium in posthemorrhagic hydrocephalus. Nat. Med. *23*, 997–1003.

Karnes, J.H., Bastarache, L., Shaffer, C.M., Gaudieri, S., Xu, Y., Glazer, A.M., Mosley, J.D., Zhao, S., Raychaudhuri, S., Mallal, S., et al. (2017). Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. Sci. Transl. Med. *9*, eaai8708.

Kim, J.G., Son, Y.J., Yun, C.H., Kim, Y.I., Nam-Goong, I.S., Park, J.H., Park, S.K., Ojeda, S.R., D'Elia, A.V., Damante, G., and Lee, B.J. (2007). Thyroid transcription factor-1 facilitates cerebrospinal fluid formation by regulating aquaporin-1 synthesis in the brain. J. Biol. Chem. *282*, 14923–14931.

Kim, K.W., Tang, N.H., Andrusiak, M.G., Wu, Z., Chisholm, A.D., and Jin, Y. (2018). A neuronal piRNA pathway inhibits axon regeneration in *C. elegans*. Neuron *97*, 511–519.e6.

Kousi, M., and Katsanis, N. (2016). The genetic basis of hydrocephalus. Annu. Rev. Neurosci. *39*, 409–435.

Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. Brief. Bioinform. *18*, 205–214.

Kulkarni, A.V., Schiff, S.J., Mbabazi-Kabachelor, E., Mugamba, J., Ssenyonga, P., Donnelly, R., Levenbach, J., Monga, V., Peterson, M., MacDonald, M., et al. (2017). Endoscopic treatment versus shunting for infant hydrocephalus in Uganda. N. Engl. J. Med. *377*, 2456–2464.

Lang, B., Song, B., Davidson, W., MacKenzie, A., Smith, N., McCaig, C.D., Harmar, A.J., and Shen, S. (2006). Expression of the human PAC1 receptor leads to dose-dependent hydrocephalus-related abnormalities in mice. J. Clin. Invest. *116*, 1924–1934.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol. *13*, 1192–1200.

Lehtinen, M.K., and Walsh, C.A. (2011). Neurogenesis at the brain-cerebrospinal fluid interface. Annu. Rev. Cell Dev. Biol. *27*, 653–679.

Lehtinen, M.K., Zappaterra, M.W., Chen, X., Yang, Y.J., Hill, A.D., Lun, M., Maynard, T., Gonzalez, D., Kim, S., Ye, P., et al. (2011). The cerebrospinal fluid provides a proliferative niche for neural progenitor cells. Neuron *69*, 893–905.

Lehtinen, M.K., Bjornsson, C.S., Dymecki, S.M., Gilbertson, R.J., Holtzman, D.M., and Monuki, E.S. (2013). The choroid plexus and cerebrospinal fluid: Emerging roles in development, disease, and therapy. J. Neurosci. *33*, 17553–17559.

Leinonen, V., Koivisto, A.M., Alafuzoff, I., Pyykkö, O.T., Rummukainen, J., von Und Zu Fraunberg, M., Jääskeläinen, J.E., Soininen, H., Rinne, J., and Savolainen, S. (2012). Cortical brain biopsy in long-term prognostication of 468 patients with possible normal pressure hydrocephalus. Neurodegener. Dis. *10*, 166–169.

Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A.Y., Yen, C.-A., Lin, S., Lin, Y., Qiu, Y., et al. (2015). Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature *518*, 350–354.

Lim, J., Tang, A.R., Liles, C., Hysong, A.A., Hale, A.T., Bonfield, C.M., Naftel, R.P., Wellons, J.C., and Shannon, C.N. (2018). The cost of hydrocephalus: A cost-effectiveness model for evaluating surgical techniques. J. Neurosurg. Pediatr. *23*, 109–118.

Lockwood Estrin, G., Kyriakopoulou, V., Makropoulos, A., Ball, G., Kuhendran, L., Chew, A., Hagberg, B., Martinez-Biarge, M., Allsop, J., Fox, M., et al. (2016). Altered white matter and cortical structure in neonates with antenatally diagnosed isolated ventriculomegaly. Neuroimage Clin. *11*, 139–148.

Lun, M.P., Johnson, M.B., Broadbelt, K.G., Watanabe, M., Kang, Y.J., Chau, K.F., Springel, M.W., Malesz, A., Sousa, A.M., Pletikos, M., et al. (2015a). Spatially heterogeneous choroid plexus transcriptomes encode positional identity and contribute to regional CSF production. J. Neurosci. *35*, 4903–4916.

Lun, M.P., Monuki, E.S., and Lehtinen, M.K. (2015b). Development and functions of the choroid plexus-cerebrospinal fluid system. Nat. Rev. Neurosci. *16*, 445–457.

Mandell, J.G., Neuberger, T., Drapaca, C.S., Webb, A.G., and Schiff, S.J. (2010). The dynamics of brain and cerebrospinal fluid growth in normal versus hydrocephalic mice. J. Neurosurg. Pediatr. *6*, 1–10.

Mandell, J.G., Kulkarni, A.V., Warf, B.C., and Schiff, S.J. (2015). Volumetric brain analysis in neurosurgery: Part 2. Brain and CSF volumes discriminate neurocognitive outcomes in hydrocephalus. J. Neurosurg. Pediatr. *15*, 125–132.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.

Matsumoto, N., Sato, K., Nishimasu, H., Namba, Y., Miyakubi, K., Dohmae, N., Ishitani, R., Siomi, H., Siomi, M.C., and Nureki, O. (2015). Crystal structure and activity of the endoribonuclease domain of the piRNA pathway factor Maelstrom. Cell Rep. *11*, 366–375.

McAllister, J.P., Guerra, M.M., Ruiz, L.C., Jimenez, A.J., Dominguez-Pinos, D., Sival, D., den Dunnen, W., Morales, D.M., Schmidt, R.E., Rodriguez, E.M., and

Limbrick, D.D. (2017). Ventricular zone disruption in human neonates with intraventricular hemorrhage. J. Neuropathol. Exp. Neurol. *76*, 358–375.

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci. *19*, 1523–1536.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.

Morales, D.M., Townsend, R.R., Malone, J.P., Ewersmann, C.A., Macy, E.M., Inder, T.E., and Limbrick, D.D., Jr. (2012). Alterations in protein regulators of neurodevelopment in the cerebrospinal fluid of infants with posthemorrhagic hydrocephalus of prematurity. Mol. Cell. Proteomics *11*, M111.011973.

Nandi, S., Chandramohan, D., Fioriti, L., Melnick, A.M., Hébert, J.M., Mason, C.E., Rajasethupathy, P., and Kandel, E.R. (2016). Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. Proc. Natl. Acad. Sci. USA *113*, 12697–12702.

Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. PLoS Genet. *6*, e1000888.

Niemi, M.E.K., Martin, H.C., Rice, D.L., Gallone, G., Gordon, S., Kelemen, M., McAloney, K., McRae, J., Radford, E.J., Yu, S., et al. (2018). Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. Nature *562*, 268–271.

Peterson, M., Warf, B.C., and Schiff, S.J. (2018). Normative human brain volume growth. J. Neurosurg. Pediatr. *21*, 478–485.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.

Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. *11*, 459–463.

Robledo, R.F., Ciciotte, S.L., Gwynn, B., Sahr, K.E., Gilligan, D.M., Mohandas, N., and Peters, L.L. (2008). Targeted deletion of α-adducin results in absent β- and γ-adducin, compensated hemolytic anemia, and lethal hydrocephalus in mice. Blood *112*, 4298–4307.

Roden, D.M., Pulley, J.M., Basford, M.A., Bernard, G.R., Clayton, E.W., Balser, J.R., and Masys, D.R. (2008). Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin. Pharmacol. Ther. *84*, 362–369.

Rosenthal, A., Jouet, M., and Kenwrick, S. (1992). Aberrant splicing of neural cell adhesion molecule L1 mRNA in a family with X-linked hydrocephalus. Nat. Genet. *2*, 107–112.

Shimodaira, H. (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. Ann. Stat. *32*, 2616–2641.

Short, P.J., McRae, J.F., Gallone, G., Sifrim, A., Won, H., Geschwind, D.H., Wright, C.F., Firth, H.V., FitzPatrick, D.R., Barrett, J.C., and Hurles, M.E. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. Nature *555*, 611–616.

Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell *151*, 964–980.

Silva-Vargas, V., Maldonado-Soto, A.R., Mizrak, D., Codega, P., and Doetsch, F. (2016). Age-Dependent niche signals from the choroid plexus regulate adult neural stem cells. Cell Stem Cell *19*, 643–652.

Simon, T.D., Riva-Cambrin, J., Srivastava, R., Bratton, S.L., Dean, J.M., and Kestle, J.R.; Hydrocephalus Clinical Research Network (2008). Hospital care for children with hydrocephalus in the United States: Utilization, charges, comorbidities, and deaths. J. Neurosurg. Pediatr. *1*, 131–137.

Simonti, C.N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D.S., Chisholm, R.L., Crosslin, D.R., Hebbring, S.J., Jarvik, G.P., Kullo, I.J., et al. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. Science *351*, 737–741.

Soper, S.F., van der Heijden, G.W., Hardiman, T.C., Goodheart, M., Martin, S.L., de Boer, P., and Bortvin, A. (2008). Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. Dev. Cell *15*, 285–297.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45* (*D1*), D362–D368.

Takagishi, M., Sawada, M., Ohata, S., Asai, N., Enomoto, A., Takahashi, K., Weng, L., Ushida, K., Ara, H., Matsui, S., et al. (2017). Daple coordinates planar polarized microtubule dynamics in ependymal cells and contributes to hydrocephalus. Cell Rep. *20*, 960–972.

Tomycz, L.D., Hale, A.T., and George, T.M. (2017). Emerging insights and new perspectives on the nature of hydrocephalus. Pediatr. Neurosurg. *52*, 361–368.

Upton, K.R., Gerhardt, D.J., Jesuadian, J.S., Richardson, S.R., Sánchez-Luque, F.J., Bodea, G.O., Ewing, A.D., Salvador-Palomeque, C., van der Knaap, M.S., Brennan, P.M., et al. (2015). Ubiquitous L1 mosaicism in hippocampal neurons. Cell *161*, 228–239.

Wang, K., Li, M., and Bucan, M. (2007). Pathway-based approaches for analysis of genomewide association studies. Am. J. Hum. Genet. *81*, 1278–1283.

Wang, K., Li, M., and Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. Nat. Rev. Genet. *11*, 843–854.

Ward, L.D., and Kellis, M. (2012). HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40*, D930–D934.

Wei, W.Q., Bastarache, L.A., Carroll, R.J., Marlo, J.E., Osterman, T.J., Gamazon, E.R., Cox, N.J., Roden, D.M., and Denny, J.C. (2017). Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. PLoS ONE *12*, e0175508.

Whitelaw, A., Kennedy, C.R., and Brion, L.P. (2001). Diuretic therapy for newborn infants with posthemorrhagic ventricular dilatation. Cochrane Database Syst. Rev. (*2*), CD002270.

Wilson, G.R., Wang, H.X., Egan, G.F., Robinson, P.J., Delatycki, M.B., O'Bryan, M.K., and Lockhart, P.J. (2010). Deletion of the Parkin co-regulated gene causes defects in ependymal ciliary motility and hydrocephalus in the quakingviable mutant mouse. Hum. Mol. Genet. *19*, 1593–1602.

Wodarczyk, C., Rowe, I., Chiaravalli, M., Pema, M., Qian, F., and Boletta, A. (2009). A novel mouse model reveals that polycystin-1 deficiency in ependyma and choroid plexus results in dysfunctional cilia and hydrocephalus. PLoS ONE *4*, e7137.

Yung, Y.C., Mutoh, T., Lin, M.E., Noguchi, K., Rivera, R.R., Choi, J.W., Kingsbury, M.A., and Chun, J. (2011). Lysophosphatidic acid signaling may initiate fetal hydrocephalus. Sci. Transl. Med. *3*, 99ra87.

Zhang, J., Williams, M.A., and Rigamonti, D. (2006). Genetics of human hydrocephalus. J. Neurol. *253*, 1255–1266.

Zheng, H., Yu, W.M., Waclaw, R.R., Kontaridis, M.I., Neel, B.G., and Qu, C.K. (2018). Gain-of-function mutations in the gene encoding the tyrosine phosphatase SHP2 induce hydrocephalus in a catalytically dependent manner. Sci. Signal. *11*, eaao1591.

Zhong, Y., Perera, M.A., and Gamazon, E.R. (2019). On using local ancestry to characterize the genetic architecture of human traits: Genetic regulation of gene expression in multiethnic or admixed populations. Am. J. Hum. Genet. *104*, 1097–1115.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
BioVU	Denny et al., 2010; Roden et al., 2008	https://victr.vanderbilt.edu/pub/biovu/
PrediXcan	Gamazon et al., 2015, 2018	https://github.com/hakyimlab/PrediXcan
GTEx	GTEx Consortium, 2013, 2015, 2020	https://gtexportal.org/home/
Gene Set Enrichment Analysis	Subramanian et al., 2005	http://www.gsea-msigdb.org/gsea/index.jsp https://www.gsea-msigdb.org/gsea/index.jsp
Biological samples and deposited data		
PrediXcan genetic associations for communicating hydrocephalus	This paper	https://victr.vanderbilt.edu/pub/biovu/
ChIP-seq data from human embryonic stem cells	Leung et al., 2015	Roadmap Epigenomics
Cerebrospinal fluid LC-MS dataset	This paper and Morales et al., 2012	https://pubmed.ncbi.nlm.nih.gov/22186713/
Differential gene expression in mouse choroid plexus	Robledo et al., 2008	Affymetrix Mouse Gene 1.0 ST Array (GEO: GSE37098)

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Eric R. Gamazon (eric.gamazon@vumc.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All results are available in Data S1, S2,S3, and S4.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

BioVU, one of the largest DNA biobanks tied to an electronic health records database containing 2.6 million unique patient records, is a genomics resource at Vanderbilt University Medical Center (Roden et al., 2008). Detailed information on the construction, utilization, ethics and policies of the BioVU resource is described elsewhere (Roden et al., 2008). Per the policies of BioVU, use of these data fall under non-human subject determination and are approved by the Vanderbilt University IRB (#170502). We leveraged this resource to identify patients with a diagnosis of communicating hydrocephalus (Phecode: 331.1, https://phewascatalog.org; Denny et al., 2013; Wei et al., 2017) who have undergone permanent CSF diversion (VP shunt, ETV, or ETV/CPC). In BioVU, we identified patients of European ancestry (287 patients with hydrocephalus and 18,740 controls).

The median age at first CSF diversion operation (VP shunt, ETV/CPC, or ETV) was 4.1 years [0.46-11.13, interquartile range] and 58% of patients were male. Seventy-four patients (26%) had post-hemorrhagic etiologies of hydrocephalus, whereas 95 (33%) of patients were diagnosed with idiopathic (congenital) hydrocephalus. Eighty-five patients (29%) of patients were diagnosed with neural tube defects. The remaining 33 patients (12%) had other etiologies of hydrocephalus (post-infectious, brain tumor, Chiari Malformation Type I and Dandy-Walker syndrome). Control patients had no diagnosis of hydrocephalus or any other neurological or developmental disorder. Genomic ancestry was quantified using principal components analysis (Price et al., 2010). To avoid potential confounding due to population stratification (Derks et al., 2017), we performed our genetic analyses only on patients of European ancestry. We included 3 genotype-based principal components within the European-ancestry dataset as covariates in downstream analyses (Price et al., 2006).

METHOD DETAILS

Estimating the genetically determined expression

We implemented PrediXcan (Gamazon et al., 2015), a gene-based method that estimates the genetic component of expression using imputation models derived from the GTEx reference transcriptome panel in 10 neurological tissues (for a total of 889 brain samples), including frontal cortex and hypothalamus, and whole blood (from 338 individuals, Data S1) (Battle et al., 2017; Gamazon et al., 2018; Battle et al., 2017). PrediXcan utilizes a patient's germline genetic profile to estimate the genetic component of gene expression in target tissues of interest. The weight (beta) from the imputation model and the number of effect alleles X_{ij} at the variant j for individual i are used to infer the genetic component of gene expression for the ith patient:

$$\widehat{G_i} = \sum_j X_{ij}\widehat{\beta}_j$$

To identify genes associated with communicating hydrocephalus, we performed logistic regression with the genetically-determined expression as independent variable and disease status, with sex and age as covariates. Although the reference transcriptome panel (GTEx) consists of healthy controls, the genetically-determined expression was tested for association with disease status in the GWAS (BioVU) samples. PrediXcan seeks to test the effect on disease risk of the genetic component of gene expression (which is to be distinguished from the disease-altered component). One advantage of the gene-based test is statistical, i.e., the reduced multiple testing burden compared to conventional (SNP-based) GWAS (~5-10M statistical tests) (Gamazon et al., 2015). Another advantage is the ability to explicitly test a biologically-meaningful mechanism (gene expression regulation), that is known to be contributory to complex diseases (Gamazon et al., 2018; GTEx Consortium, 2015) versus genetic variants with mostly unknown gene targets. We emphasize that even if measured gene expression (RNA-seq) is available in the GWAS samples for differential expression analysis, this in no way negates the importance of estimating (and then testing for association with disease risk) just the *genetic component* of gene expression, as PrediXcan aims to do.

Odds ratios were calculated, as in case-control studies, and genes with nominally significant $p < 0.05$ are reported. Experiment-wide significance for a gene association was evaluated using Bonferroni correction for the total number of gene-tissue pairs tested (n = 9,868) for which expression imputation quality in a tissue (assessed in the reference GTEx panel) satisfied $r^2 > 0.01$. False discovery rate (Benjamini-Hochberg) was set at 0.05 for each tissue.

Tissue specificity of gene expression was quantified using the tau statistic (Kryuchkova-Mostacci and Robinson-Rechavi, 2017) applied to the GTEx tissues:

$$\tau = \sum_{i=1}^{n} \frac{1 - \widehat{x_i}}{n - 1}$$

where

$$\widehat{x_i} = x_i \Big/ \max_{1 \le j \le n} x_j$$

Here, x_i provides the expression values for the gene ***i***, and ***n*** is the number of tissues.

Genetically determined pathways and networks

Since the heritability of traits that have been interrogated through GWAS has been shown to be enriched for regulatory variation (Gamazon et al., 2018), we hypothesized that using the genetically determined component of gene expression, as quantified through PrediXcan, rather than the total expression (with potential environmental and technical confounding components) would enhance statistical power to identify pathophysiologically relevant pathways and networks. For each tissue type, we used as input the genes whose genetic component showed nominal association ($p < 0.05$) with hydrocephalus. Gene Set Enrichment Analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) was performed by determining enrichment of input genes in the curated gene sets in the Molecular Signatures Database (MSigDB; http://www.gsea-msigdb.org/gsea/msigdb/index.jsp). This list of input genes was compared against the MSigDB gene sets (C2: Curated Gene Sets, C5: Gene Ontology, C6: Oncogenic Signature, and C1: Chromosomal location). We used the default options in GSEA/MSigDB (which internally uses the hypergeometric distribution) to perform the enrichment analysis. Significance of enrichment was assessed using Benjamini-Hochberg adjusted $p < 0.05$ in each tissue. Experiment-wide significant gene sets were identified using Bonferroni adjustment (adjusted $p < 0.05$, which corresponded to $p < 4.1 \times 10^{-7}$) based on the total number of gene sets (4,762 curated sets; 5,917 GO sets; 326 chromosomal locations; and 189 oncogenic signatures; total count = 11,194) and the total number of tissues (n = 11).

BioVU analysis of hydrocephalus-related neurological traits

For our experiment-wide significant finding, we considered the PrediXcan associations in the exact discovery tissue (frontal cortex) with "Other cerebral degenerations" (Phecode: 331; 417 cases and 17,257 controls) and "Cerebral edema and compression of

brain'' (Phecode: 348.2; 635 cases and 17,257 controls) within BioVU to gain further mechanistic insights into the pathogenesis of hydrocephalus.

UK Biobank replication analysis

Integrated analysis of brain imaging and genomics data can facilitate validation and additional insights into the functional consequences of identified disease-associated genes. We performed PrediXcan analysis in the exact discovery tissue (frontal cortex) on CSF, white matter, and total brain volumes in the UK Biobank (N = 8,428) to validate our experiment-wide significant finding. From the T1 MRI structural image and normalized for overall head size, these brain imaging traits had been quantified (Elliott et al., 2018); we used these phenotypes in the PrediXcan validation analysis. The significance of replication was assessed using Bonferroni-adjusted $p < 0.05$ based on the total number of gene-tissue-phenotype tuples (1*1*3 = 3) tested. We also tested for enrichment of gene-level associations with these imaging phenotypes among the top differentially expressed genes identified in the frontal cortex ($p < 0.05$). In addition, because PrediXcan, the rare-variant exome scan, and a murine model of hydrocephalus appear to converge on *TMEM50B* (albeit only nominally), we interrogated the UK Biobank data for the association of the rare variant rs34327244 (A139T, MAF = 0.4%) within *TMEM50B*, with Bonferroni-adjusted $p < 0.05$ (adjusting for the 3 imaging phenotypes) as the replication significance criterion.

For additional support, we tested the SNPs within the *MAEL cis* region for association with hydrocephalus (Phenotype code = G6_HYDROCEPH) in a much larger UK Biobank dataset (N = 361,194, of which 133 are cases). We utilized public association results from the Benjamin Neale/HAIL team (Bycroft et al., 2018; Ge et al., 2017). Because of the much smaller number of cases than number of controls in this dataset, we followed a widely used variant quality control (QC) set of recommendations. Out of 10,629 SNPs in the region, we excluded all low-confidence variants, defined as follows: (a) MAF < 0.10%; (b) 2*(MAF)*(number of cases) < 25. We used Benjamini-Hochberg adjusted $p < 0.05$ as the cutoff for significance. We also tested the SNPs within the *MAEL cis* region (excluding low-confidence variants) for association with subarachnoid hemorrhage (N = 361,070, of which 124 are cases), a common cause of hydrocephalus (Germanwala et al., 2010; Graff-Radford et al., 1989).

Differential expression in mice

We leveraged publicly available choroid plexus expression data in mice (GEO: GSE37098), with and without hydrocephalus, on the C57BL/6J strain to replicate some of our top findings in the two tissues, i.e., frontal cortex and hypothalamus, with significant signals (Benjamini-Hochberg adjusted $p < 0.05$ within a tissue) in human patients. Only frontal cortex had an experiment-wide significant signal, Bonferroni-adjusted $p < 0.05$ across tested gene-tissue pairs. Gene expression had been previously quantified using Affymetrix Mouse Gene 1.0 ST Array (Robledo et al., 2008). We conducted differential expression analysis using an empirical Bayes method, as implemented in Linear Models for Microarray Data (*limma*), to identify gene expression changes associated with hydrocephalus in these mice. We report the genes that were differentially expressed ($p < 0.05$) in mice among our top genes ($p < 0.05$) in human frontal cortex and hypothalamus (evaluated separately).

Rare variant exome scan

The goal of this analysis was to identify associations between rare exonic variants in genes detected in at least 5 tissues by PrediXcan and hydrocephalus risk. We developed a cohort of 3,890 children and 25,823 adult patients of European ancestry who had previously undergone genotyping using Illumina Infinium Human Exome Bead Chip platforms. ICD9 codes for hospital billing were used to algorithmically define cases as well as controls, as previously described (Denny et al., 2013, 2010, Karnes et al., 2017, Wei et al., 2017). Fisher's Exact test and Bonferroni correction were used to detect rare variant associations with hydrocephalus (Denny et al., 2013, 2010, Karnes et al., 2017, Simonti et al., 2016). We tested a total of 15 rare variants (i.e., MAF < 1%, including 1 in *TMEM50B*).

CSF proteomic data from patients with hydrocephalus

We analyzed a proteomic signature consisting of proteins that were found in a mass spectroscopic analysis of CSF (Morales et al., 2012). We evaluated the genetically determined expression of these genes in frontal cortex for their (PrediXcan) association with hydrocephalus to determine whether the proteins showed significant departure from the null. We then tested whether the signature had greater statistical significance (i.e., lower p value) than the remaining genes (using Mann-Whitney U test). For the latter, we performed the comparisons in frontal cortex (the significant discovery tissue) and whole blood (an easily accessible tissue) to further explore the tissue specificity of these results.

Cell Reports, Volume *35*

Supplemental information

Multi-omic analysis elucidates the genetic basis of hydrocephalus

Andrew T. Hale, Lisa Bastarache, Diego M. Morales, John C. Wellons III, David D. Limbrick Jr., and Eric R. Gamazon

SUPPLEMENTARY INFORMATION

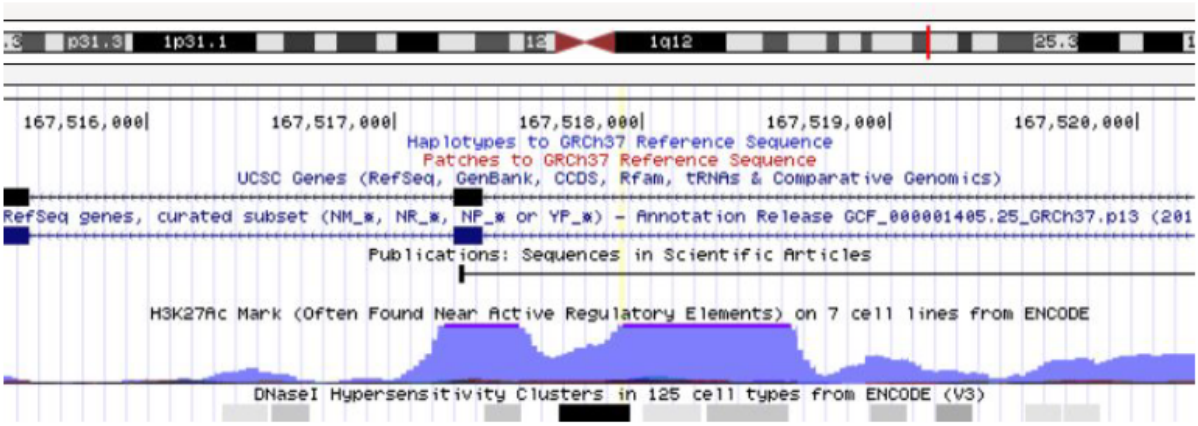
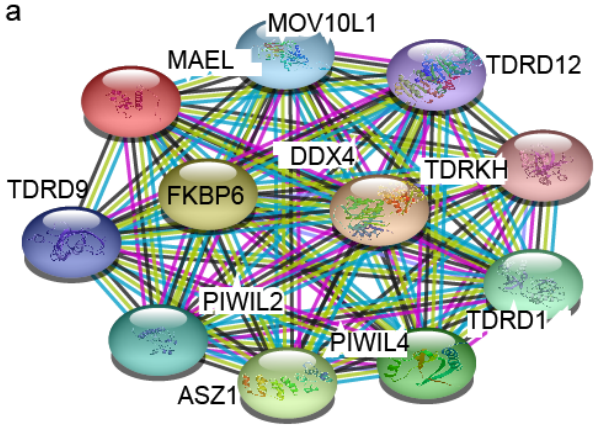


Figure S1. The most significant SNP (rs75008967, yellow line) within the *MAEL* cis region overlaps an enhancer element in fetal brain (male and female) and in several brain regions (H3K27ac), including prefrontal cortex, based on Roadmap Chip-seq data (http://www.roadmapepigenomics.org), Related to Figure 4.

a



b

Gene	Phecode	P value
TDRKH	Type II diabetes with neurological complications and polyneuropathy	7.9×10^{-4}
DDX4	Multiple sclerosis	2.2×10^{-3}

Figure S2. Genetically determined associations with neurological disorders for MAEL interacting partners, Related to Figure 4. **(A)** STRING analysis revealing MAEL interaction network derived from various lines of evidence: curated databases (aqua), experimental determined (pink), gene neighborhood (dark green), gene fusions (red), gene co-occurrence (dark blue), text mining (light green), co-expression (black), and protein homology (light blue). The individual proteins shown in the diagram are predicted functional partners. Interestingly, all proteins shown highlighted here have some degree of protein structural data. **(B)** Genetically determined expression associated with other neurological disorders for proteins in the MAEL interaction network depicted in **(A)**.

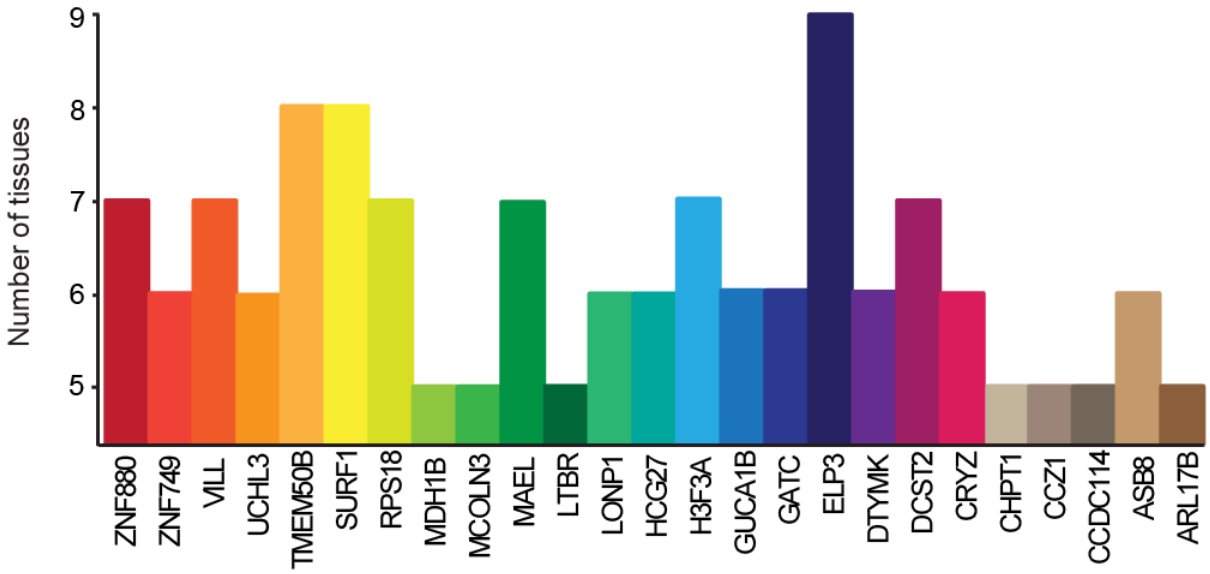


Figure S3. Top 25 genes across multiple neurological tissues, Related to Figures 3-4. Gene with nominal association ($p<0.05$) with hydrocephalus (x-axis) vs. number of neurological tissues in GTEx (y-axis), demonstrating a small subset of tissue-shared hydrocephalus-associated genes.

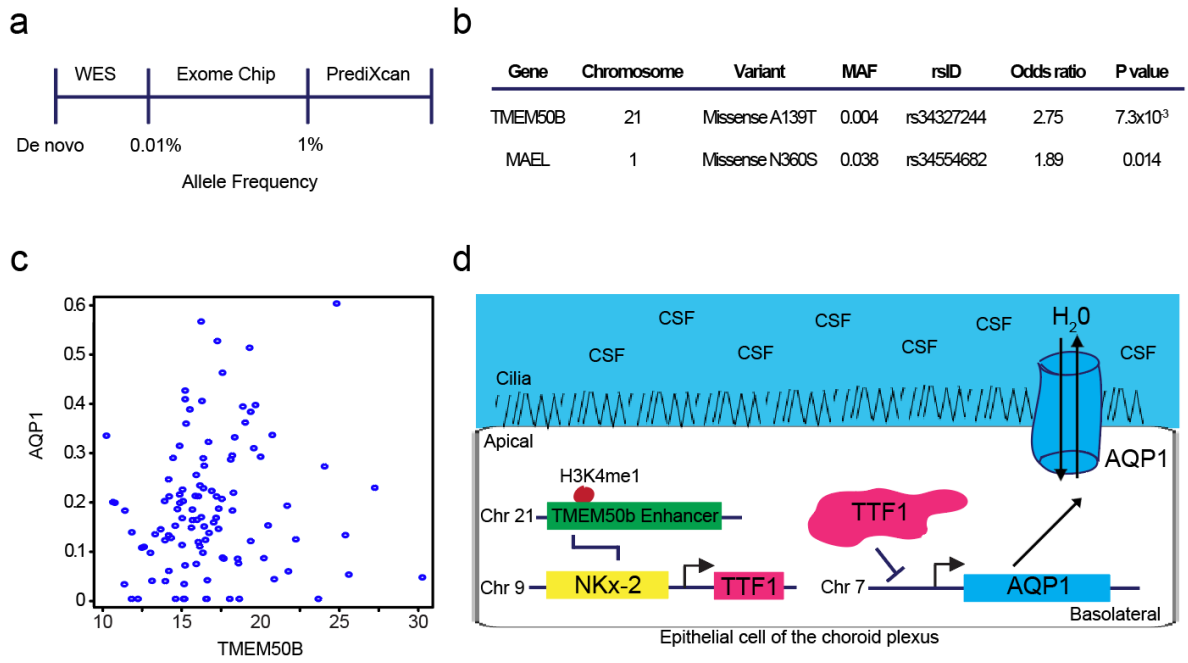


Figure S4: Exome scan reveals nominally significant rare variant associations in the most differentially-expressed genes identified by PrediXcan, Related to Figure 4. (**A**) Illustration of genetic strategies for elucidating the genetic architecture of disease. Whole exome sequencing (WES), Exome-Chip based approaches (Rare variant test), and common-variant associations based on differential gene regulation (PrediXcan) are used to probe the entire allele frequency spectrum. (**B**) Exome scan (in a cohort of 29,713 patients of European ancestry) identifies rare variant associations with hydrocephalus in the most differentially expressed genes across 5 or more neurological tissues. *TMEMB50B* is also supported by differential expression (choroid plexus transcriptome) analysis we performed in a murine model of hydrocephalus. (**C**) Correlation between *TMEM50B* and *AQP1* expression in the frontal cortex (Spearman's correlation= 0.21, p-value= 0.029), consistent with their effect on hydrocephalus. (**D**) Potential mechanism by which *TMEM50B* may contribute to hydrocephalus predisposition. TTF1 is a direct transcriptional regulator of the aquaporin 1 (AQP1) in epithelial cells of the choroid plexus (Kim et al., 2007). rs34327244, identified by the exome scan, disrupts an enhancer element (marked by H3K4me1) in *TMEM50B* and a regulatory motif (Nkx2), leading to increased allelic affinity of TTF1, a direct transcriptional regulator of *AQP1*, and thus decreased availability of TTF1 to promote *AQP1* expression.

Table S1: MAEL SNP predictors of hydrocephalus in the frontal cortex and hypothalamus, Related to Figure 4.

Frontal Cortex				***Hypothalamus***			
SNP	**Weight**	**Reference allele**	**Effect Allele**	**SNP**	**Weight**	**Reference allele**	**Effect Allele**
rs12081670	-0.163	T	C	rs12742385	-0.178	A	G
rs12742385	-0.163	A	G	rs12757615	-0.156	G	A
rs202253	-0.101	T	G	rs17430677	-0.132	T	C
rs437846	-0.094	G	A	rs6670426	-0.121	T	C
rs12757615	-0.070	G	A	rs4656560	-0.075	T	C
rs1890129	-0.052	C	T	rs2269673	-0.075	G	T
rs2269678	-0.049	T	C	rs12741275	-0.075	T	C
rs4657619	-0.048	A	G	rs742469	-0.052	G	A
rs561187	-0.033	A	G	rs34925982	-0.050	A	G
rs472293	-0.029	T	C	rs3738236	-0.048	G	A
rs520560	-0.029	A	G	rs520560	-0.029	A	G
rs517419	-0.028	G	A	rs517419	-0.028	G	A
rs16858233	-0.028	T	C	rs16858233	-0.028	T	C
rs1890130	-0.028	A	G	rs1890130	-0.028	A	G
rs12124524	-0.027	G	A	rs12124524	-0.027	G	A
rs12731305	-0.027	G	A	rs12731305	-0.027	G	A
rs1845459	-0.026	A	G	rs1845459	-0.026	A	G
rs4656550	-0.015	C	T	rs4656550	-0.015	C	T
rs1120205	-0.013	T	C	rs1120205	-0.013	T	C
rs1120604	-0.011	T	C	rs1120604	-0.011	T	C
rs858557	-0.009	C	T	rs858557	-0.009	C	T
rs4657563	-0.008	G	A	rs4657563	-0.008	G	A
rs6680946	-0.008	G	T	rs6680946	-0.008	G	T
rs6427065	-0.003	T	G	rs6427065	-0.003	T	G
rs203846	-0.002	T	C	rs203846	-0.002	T	C
rs500242	-0.001	A	G	rs500242	-0.001	A	G
rs17344586	-0.001	G	A	rs17344586	-0.001	G	A
rs566841	-0.001	T	C	rs566841	-0.001	T	C
rs7531189	0.009	T	C	rs7531189	0.009	T	C
rs12139875	0.012	C	T	rs12139875	0.012	C	T
rs1002160	0.013	C	T	rs1002160	0.013	C	T
rs12123839	0.014	C	T	rs12123839	0.014	C	T
rs6694465	0.014	T	G	rs6694465	0.014	T	G
rs6661553	0.014	A	G	rs6661553	0.014	A	G
rs12067635	0.014	G	A	rs12067635	0.014	G	A
rs4657621	0.015	A	G	rs4657621	0.015	A	G
rs10918372	0.015	A	C	rs10918372	0.015	A	C
rs10489192	0.015	T	C	rs10489192	0.015	T	C
rs6676134	0.022	C	T	rs6676134	0.022	C	T
rs6691431	0.035	T	C	rs6691431	0.035	T	C
rs7545806	0.036	T	C	rs7545806	0.036	T	C
rs10918603	0.038	T	C	rs10918603	0.038	T	C
rs742048	0.038	G	T	rs742048	0.038	G	T
rs6673622	0.046	A	G	rs6673622	0.046	A	G
rs11803833	0.048	A	G	rs11803833	0.048	A	G
rs1214606	0.055	G	T	rs1214606	0.055	G	T
rs2480678	0.067	T	G	rs2480678	0.067	T	G
rs4657543	0.068	T	C	rs4657543	0.068	T	C
rs17331024	0.073	A	G	rs17331024	0.073	A	G
rs10918329	0.085	G	A	rs10918329	0.085	G	A
rs402031	0.094	A	G	rs402031	0.094	A	G
rs7538944	0.103	C	T	rs7538944	0.103	C	T
rs4657608	0.154	C	T	rs4657608	0.154	C	T
rs10489200	0.163	C	T	rs10489200	0.163	C	T
rs12141731	0.169	C	T	rs12141731	0.169	C	T
rs393025	0.182	C	A	rs393025	0.182	C	A
rs269686	0.242	G	A	rs269686	0.242	G	A
rs269684	0.284	C	T	rs269684	0.284	C	T

Table S2. Replication analysis in the independent UK Biobank, Related to Figure 4-5. SNPs within the *MAEL* cis region (within 1 Mb), excluding low-confidence variants (see Methods), were tested for association with hydrocephalus. We identified 9 variants with Benjamini-Hochberg FDR<0.05. All 9 variants are common (MAF > 10%) and in linkage disequilibrium (r^2>0.70). The most significant SNP (rs75008967) overlaps an enhancer element in several brain regions (Figure S1), suggesting regulatory effect.

SNP	Allele 1	Allele 2	P value
rs75008967	A	C	0.000352
rs72695306	T	C	0.000569
rs11810854	A	G	0.000802
rs74796115	G	A	0.000827
rs4330902	T	G	0.000831
rs74119851	G	A	0.000891
rs75065369	C	T	0.000924
rs112620640	A	G	0.000956
rs12410167	G	A	0.000972

Table S3: Mouse choroid plexus gene expression changes compared to human imputed gene expression by PrediXcan, Related to Figure 3.

Gene	P value (human)	Beta (human)	Tissue	P value (mice)	-log fold change (mice)	t (mice)
TMEM50b	0.016/0.015	-0.728/-0.584	Frontal Cortex & Hypothalamus	$1.14x10^{-4}$	$3.36x10^{-1}$	-4.97
Kctd21	0.040/0.037	-0.365/-0.683	Frontal Cortex & Hypothalamus	$3.02x10^{-2}$	$1.68x10^{-1}$	-2.38
Cry1	0.005	0.754	Frontal Cortex	$6.94x10^{-4}$	$-4.35x10^{-1}$	4.21
Vill	0.010	0.467	Frontal Cortex	$4.82x10^{-2}$	$-1.46x10^{-1}$	2.14
Pkd2	0.013	0.880	Frontal Cortex	$2.47x10^{-2}$	$-1.42x10^{-1}$	2.48
Alg8	0.016	-0.752	Frontal Cortex	$9.56x10^{-4}$	$3.01x10^{-1}$	-4.05
Calm2	0.018	10.68	Frontal Cortex	$7.07x10^{-3}$	$-1.47x10^{-1}$	3.10
Sqle	0.023	-0.307	Frontal Cortex	$5.37x10^{-3}$	$3.37x10^{-1}$	-3.23
Usp14	0.023	-0.734	Frontal Cortex	$1.00x10^{-2}$	$1.72x10^{-1}$	-2.93
Eif2d	0.038	-0.612	Frontal Cortex	$4.61x10^{-2}$	$-1.03x10^{-1}$	2.17
Mdh1b	0.041	-0.390	Frontal Cortex	$2.04x10^{-4}$	$5.31x10^{-1}$	-4.18
Chchd1	0.042	-0.155	Frontal Cortex	$2.34x10^{-2}$	$1.31x10^{-1}$	-2.51
Gas6	0.030	-0.490	Hypothalamus	$2.40x10^{-3}$	$2.18x10^{-1}$	-3.61
Ano10	0.003	0.491	Hypothalamus	$2.88x10^{-3}$	$-1.99x10^{-1}$	3.53
Fam171b	0.044	-1.546	Hypothalamus	$5.40x10^{-3}$	$7.44x10^{-3}$	-3.07
Usp42	0.021	0.551	Hypothalamus	$1.95x10^{-2}$	$-1.74x10^{-1}$	2.60
Sfpq	0.010	12.24	Hypothalamus	$1.99x10^{-2}$	$-1.69x10^{-1}$	2.59
Lrrfip2	0.046	0.848	Hypothalamus	$4.88x10^{-2}$	$-1.38x10^{-1}$	2.14

Supplemental Table S4. Number of subjects and phenotypes included in this study, Related to Figure 1.

Data source and phenotype	Number of subjects	Discovery or validation?
BioVU (PheCode: 331.1, Hydrocephalus)	N= 19,027, of which 287 are cases	Discovery
UK Biobank Hydrocephalus GWAS (G6_HYDROCEPH)	N= 361,194, of which 133 are cases	Validation
UK Biobank Brain Imaging GWAS	N= 8,428	Validation
Human cerebrospinal fluid (CSF) proteomic analysis	N= 12 cases	Validation
Mouse choroid plexus microarray information	N= 12, of which 6 developed hydrocephalus	Validation