Systems-level elucidation of the pathogenesis of cerebral arachnoid cysts

Adam J. Kundishora[1†], Garrett Allington[2,3†], Stephen McGee[4†], Kedous Y. Mekbib[1,3†], Vladimir Gainullin[4], Andrew T. Timberlake[1], Carol Nelson-Williams[5], Andres Moreno-De Luca[6], Emre Kiziltug[1], Hannah Smith[1,3], Jack Ocken[1], John Shohfi[1], August Allocco[1], Phan Q. Duy[1], Aladine Elsamadicy[1], Weilai Dong[7], Shujuan Zhao[8], Yung-Chun Wang[8], Hanya Qureshi[1], Michael L. Diluna[1], Shrikant Mane[8,9], Irina R. Tikhonova[10], Po-Ying Fu[8], Christopher Castaldi[9], Francesc López-Giráldez[9], James R. Knight[9], Charuta G. Furey[1], Bob S. Carter[3], Shozeb Haider[12], Seth L. Alper[13], Murat Gunel[1], Francisca Millan[4], Richard P. Lifton[7], Rebecca I. Torene[4†], Sheng Chih Jin[8,14†], and Kristopher T. Kahle[1,3,15,16#†]

1. Department of Neurosurgery, Yale University School of Medicine, New Haven, CT, USA
2. Department of Pathology, Yale University School of Medicine, New Haven, CT, USA
3. Department of Neurosurgery, Massachusetts General Hospital, Boston, MA, USA
4. GeneDx, Gaithersburg, MD, USA
5. Department of Genetics, Yale University School of Medicine, New Haven, CT, USA
6. Department of Radiology, Autism & Developmental Medicine Institute, Genomic Medicine Institute, Geisinger, Danville, PA, USA
7. Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, NY, USA
8. Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA
9. Yale Center for Genomic Analysis, Yale University, West Haven, CT, USA
10. School of Pharmacy, Queen's University Belfast, Belfast, UK
11. Department of Neuroscience and Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, CT, USA
12. School of Pharmacy, University College London, London, UK
13. Division of Nephrology and Vascular Biology Research Center, Beth Israel Deaconess Medical Center; and Department of Medicine, Harvard Medical School, Boston, MA USA
14. Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA
15. Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA
16. Broad Institute of MIT and Harvard, Cambridge, MA, USA

[†] These authors contributed equally

[#] Correspondence: kahle.kristopher@mgh.harvard.edu

ABSTRACT

Cerebral arachnoid cysts (ACs) are one of the most common and poorly understood developmental brain lesions in humans. To begin to elucidate AC pathogenesis, we performed an integrated analysis of 617 proband-parent exomes, 152,898 brain and meningeal single-cell RNAseq transcriptomes, and natural language processing data of proband medical records. We found damaging *de novo* variants (DNVs) were highly enriched in AC probands versus controls ($P=1.57\times10^{-33}$). Seven genes harbored an exome-wide significant DNV burden. AC-associated genes are enriched for chromatin modifiers and converged in midgestational transcription networks essential for neural and meningeal development. Unsupervised clustering of patient phenotypes identified four AC subtypes and clinical severity correlated with the presence of a damaging DNV. These results shed new insight into the coordinated regulation of brain and meningeal development and implicate epigenomic dysregulation due to DNVs in AC pathogenesis. In the appropriate clinical context, ACs may be considered radiographic harbingers of neurodevelopmental pathology warranting genetic testing and neurobehavioral follow-up. These data highlight the utility of a systems-level, multi-omics approach to elucidate sporadic structural brain disease.

INTRODUCTION

The convoluted folding patterns of the primate brain cortex, arising during the last two-thirds of gestation, serve to increase neural processing power by augmenting cortical surface area allocated to neurons [1]. Brain folding is tightly coordinated with leptomeningeal membrane development, including the epithelial-like arachnoid barrier cells whose underlying trabeculated subarachnoid space interdigitates in the sulcal troughs between cortical gyri and the deeper cerebral fissures. This complex structure is especially evident at the most distinctive surface landmark of the human brain, the Sylvian fissure, which separates the frontal and parietal lobes from the temporal lobe. While mechanisms of human gyrification and sulcation have been illuminated by studies of patients with pachygyria and polymicrogyria (genetic disorders of impaired neuronal migration[2]), the genetic regulation of leptomeningeal development has remained elusive despite its important role in bi-directional signaling between underlying cortex and overlying calvarium [3-5].

Arachnoid cysts (ACs) are leptomeningeal-lined, cerebrospinal fluid (CSF)-filled sacs located within the major sulcal folds of the developing brain. ACs occur in larger mammals with a frequency proportional to the degree of brain gyrification and sulcation; they are the most common space-occupying intracranial lesion in humans (~0.6% of individuals) [6] but are absent in smooth-brained rodents. Signs and symptoms of AC mass effect depend on cyst location and size and may include headache, nausea and vomiting, and visual disturbances, among others [7]. ACs most frequently localize to the middle cranial fossa where they invaginate into and widen the Sylvian fissure (~30-40% of cases) [6]. Many ACs are discovered during prenatal development or in early childhood [8]. Although affected individuals may manifest symptoms in the first year, many never develop symptoms that require neurosurgical treatment [9]. Questions of whether and how to treat ACs remain controversial [10,11]. Neurosurgical AC fenestration is reserved most often for cases associated with hydrocephalus (accumulation of excessive CSF in the brain) that result from obstruction of intraventricular CSF flow [11,12]. In the absence of hydrocephalus, ACs have often been considered radiographic “incidentalomas” not warranting further evaluation or follow-up [12].

The pathogenesis of most ACs is unknown. While some cases have been attributed to secondary factors such as brain hemorrhage, trauma, or infection [12-14], most cases are now considered primary, congenital, or idiopathic [14]. A classic mechanism of AC pathogenesis posits a “developmental splitting” of the bilayered arachnoid membrane layer with subsequent expansion of the intra-arachnoid space due to CSF accumulation from ball-valve-like siphoning of CSF [12,15]. Interestingly, ACs have been frequently described in the setting of developmental delay, epilepsy, and autism, particularly in children, which can complicate neurosurgical decision-making in the absence of obstructive hydrocephalus [16-20]. Indeed, the most common indication for brain MRI in 309 pediatric patients found to have an AC was a concern for seizure or cognitive dysfunction or developmental delay [20] Interestingly, ACs are often accompanied by underdevelopment (hypoplasia) of the adjacent cortex, particularly in middle fossa ACs which are situated adjacent to critical memory and language areas in the temporal lobe [21]. These observations suggest ACs may represent anatomic correlates of a more fundamental developmental brain defect. In this context, associated neurodevelopmental phenotypes could result not from AC mass effect *per se*, but rather from a related etiologic factor. Consistent with this idea, AC fenestration for non-hydrocephalic phenotypes such as epilepsy has shown, at best, inconsistent clinical benefit [22].

Gaps in our understanding of AC pathogenesis impede the development of improved diagnostic, prognostic, and therapeutic measures for patients [23]. ACs are encountered more frequently in the setting of Mendelian conditions such as Acrocallosal, Chudley-McCullough, Aicardi, Pallister-Hall syndromes, and the mucopolysaccharidoses [24-26]. Moreover, >38 families have been described in which multiple family members have ACs; many of these are radiographically similar, and sometimes bilateral [27,28]. These data implicate genetic contribution to AC pathogenesis. Nonetheless, no large, systematic human genomic-phenomic study of cerebral ACs has been performed to date. The discovery of rare, large-effect gene variants associated with ACs could help elucidate the genetic regulation of human leptomeningeal-cortical development, provide insight

into disease pathogenesis, explain the heterogeneity of AC phenotypes and therapeutic responses, and aid physicians in clinical decision-making. Nevertheless, the sporadic nature of most AC cases limits the utility of traditional genetic approaches. This has motivated whole-exome sequencing of large patient cohorts, searching for genes with rare, damaging variants in probands more often than expected by chance. This approach has aided the study of other brain malformations [29-32], congenital heart diseases [33], congenital hydrocephalus [34,35], and other genetically heterogeneous neurodevelopmental disorders, including autism and epilepsy [36-41].

Here, we aimed to integrate parent-proband (trio)-based exome sequencing data of AC families, single-cell (sc)-RNAseq data of the developing brain and meninges, and text-mined data from patient medical records in a systems-level approach to elucidate AC pathogenesis. ACs are the most common and perhaps one of the most distinctive structural brain defects in humans that we suspect might serve as a window into the genetic coordination of brain and leptomeningeal development. We hypothesized that: (i) multiple novel AC genes harboring *de novo* variants (DNVs) will be discovered using trio-based exome sequencing; (ii) AC genes will spatiotemporally converge in co-expression modules, cell types, and biological pathways pertinent to the regulation of fetal brain and meningeal development; and (iii) the systematic comparison of phenotypic data from individual AC cases will assist gene discovery by clustering cases with similar endophenotypes, thereby defining clinically-relevant disease subclasses.

RESULTS

ACs are associated with DNVs in genes that are intolerant to mutation and highly expressed in the fetal brain

To systematically identify novel genetic causes of AC, we ascertained 617 proband-parent trios with radiographically defined cerebral ACs (**Supplementary Figure 1**) for Institutional Review Board (IRB)-approved study (see **Methods**). Trios comprised predominantly pediatric patients from a clinical laboratory referral center and an academic healthcare-based center. The clinical laboratory referral cohort included 1,599 samples (560 parent-offspring trios) with a median proband age of 7 years (interquartile range, 3-12 years; range, 0-74 years), The academic healthcare-based cohort included 171 samples (57 parent-offspring trios) with a median proband age of 4.5 years (interquartile range, 1-12 years; range, 0-45 years). Developmental delay, autism, and seizures were more highly represented in the clinical laboratory referral cohort, consistent with their reason for laboratory referral, whereas hydrocephalus and Chiari I malformation were more highly represented in the neurosurgery-based academic healthcare-based cohort (**Supplementary Table 1**). Consistent with the literature [6], ACs most frequently localized to the temporal lobe (middle cranial fossa and Sylvian fissure) (**Figure 1A)**.

Exome sequencing was performed, and variants were called and analyzed as described previously (see **Methods** for sequence alignment, variant calling, calibration, annotation, and validation information) [35,42]. 1,798 control trios (comprising unaffected siblings and parents of patients with autism spectrum disorder [ASD] from the Simons Simplex Collection [SSC] cohort) were analyzed in parallel [43]. Stringent quality control was performed on variants and samples (see **Methods**) to obtain 789 *de novo* variants (DNVs) in 617 individuals, representing a total DNV rate of 1.28 per subject (**Supplementary Figure 2**). The number of observed DNVs in cases and controls resembled previous results with similar sequencing platforms [35,42] and approximated the Poisson distribution (**Supplementary Figure 2**).

Analysis showed protein-damaging (LoF + deleterious [D]-mis)) DNVs (see **Methods**), but not synonymous or missense DNVs inferred to be tolerated (T-mis), were highly enriched in AC cases (1.96-fold enrichment, Poisson $P = 1.34 \times 10^{-25}$; **Table 1**). Greater enrichment was seen in genes intolerant of LoF mutations (pLI ≥ 0.9 in ExAC) (3.31-fold enrichment, Poisson $P = 1.57 \times 10^{-33}$) and high brain expression genes (HBE; those in the top quartile of mouse brain bulk RNA-seq expression at embryonic day 9.5; 2.83-fold enrichment, Poisson

$P = 6.75 \times 10^{-23}$) **Table 1, Figure 1B, C**; see **Methods**). Enrichment was greatest in genes meeting both criteria (3.82-fold enrichment, Poisson $P = 2.13 \times 10^{-23}$; **Table 1**). Controls showed no enrichment of DNVs in any evaluated gene class (**Table 1**).

To detect gene-specific enrichments of DNVs, we applied DeNovoWEST (*De Novo* Weighted Enrichment Simulation Test) (see **Methods**) to all individuals in our cohort and identified seven genes with a pLI > 0.9 that surpassed exome-wide significance thresholds (p-value threshold of 8.61×10^{-7} after correction for triplicate testing of 19,347 reference sequence (RefSeq) genes by one-tailed Poisson test, see **Methods**) (**Extended Data Table 1**). Exome-wide significant genes included *ADNP*, *ARID1B*, *KDM5C*, *PURA*, *FOXP1*, *MAP2K1*, and *SCN2A*. The application of DeNovoWEST to synonymous DNVs identified no significantly enriched genes (**Supplementary Figure 3A**). Orthogonal analyses using DenovolyzeR (see **Methods**) showed *FOXP1*, *ADNP*, *PURA*, *MAP2K1*, and *DDX3X* each had significant enrichment of protein-damaging DNVs in cases and surpassed the multiple-testing correction threshold for exome-wide significance (p-value threshold of 8.61×10^{-7}; **Figure 1, Table 2, Extended Data Table 2**). *KDM5C* harbored an exome-wide significant enrichment of protein-altering (LoF + all missense) DNVs (**Extended Data Table 2**, **Table 2**). Quantile-quantile plots for all variant classes are shown in **Supplementary Figure 3**. Greater enrichment was observed in LoF-intolerant genes with multiple DNVs (12.6-fold enrichment; $P = 3.3 \times 10^{-17}$ by 1M permutations; **Table 2**), supporting these as causal AC disease genes. Two genes with pLI > 0.9 (*ADNP* and *MAP2K1*) had recurrent damaging DNVs (**Table 2**).

Each high-pLI AC gene surpassing exome-wide significance thresholds in both DeNovoWEST and DenovolyzeR analyses also surpassed the evidence threshold set by the Online Mendelian Inheritance in Man (OMIM) database to be considered *a bona fide* causal gene of at least one OMIM disease (**Supplementary Table 2**). Indeed, damaging DNVs in high-pLI genes linked to diseases in OMIM were markedly enriched in AC cases but not in controls (6.95-fold enrichment; $P = 3.68 \times 10^{-56}$ by 1M permutations; **Table 3**). From the observed excess of damaging DNVs in intolerant genes, we estimate that these genes account for ~16.0% of cases studied (see **Methods**). Taken together, these results indicate that ACs are highly associated with DNVs in high-pLI OMIM genes with elevated fetal brain expression.

AC-associated genes encode transcriptional regulators with critical roles in brain development

Each exome-wide significant AC gene is a transcriptional regulator that plays a critical role in brain development (**Supplementary Table 2**). For example, chromatin modifiers *ARID1B*, *ADNP*, and *KDM5C* contained 15 total protein-damaging or -altering DNVs (**Figure 1D** and **Table 2**). *ARID1B* and *ADNP* encode interacting subunits of the neural-specific ATP-dependent SWI/SNF (BAF) chromatin remodeling complex [44]. *ARID1B*, mutated in AD Coffin-Siris syndrome type 1 (OMIM # 135900), contained six LoF DNVs (three stop-gain, three frameshifts) and one D-mis DNV (p.E2082K). *ADNP*, a homeodomain-containing zinc finger transcription factor implicated in AD Helsmoortel-van der Aa type of intellectual disability (OMIM # 615873), contained four LoF DNVs (three stop-gain, one frameshift), including the recurrent p.Y719X mutation in two unrelated AC probands. *KDM5C*, encoding an H3K4me3 and H3K4me2 demethylase mutated in X-linked Claes-Jensen-type intellectual disability (OMIM# 300534), contained three D-mis DNVs (p.E463K, p.R638H, and p.T646M) that all impact conserved residues of the protein's JmjC histone demethylase domain (**Figure 1D**) that are predicted to result in significant ionic repulsion or steric clashes that inhibit demethylase activity (**Supplementary Figure 4**).

Transcriptional regulators *FOXP1*, *DDX3X*, and *PURA* contained ten total protein-damaging DNVs (**Figure 1D and Table 2**). *FOXP1*, encoding a forkhead box (FOX) transcription factor associated with AD mental retardation with language impairment, with or without autistic features (OMIM # 613670) (**Supplementary Table 2**), harbored two LoF (p.S448X and p.Q211X) and two D-Mis DNVs (p.R513H and p.Y469C). The latter impacts conserved residues of the main forkhead domain of the protein and are predicted to disrupt DNA binding (**Figure 1D**, **Table 2, Supplementary Figure 4**) [45]. *DDX3X*, a DEAD-box ATP-dependent RNA helicase associated with X-linked Snijders Blok-type intellectual disability (OMIM# 300958) (**Supplementary**

Table 2), contained one LoF (p.Q27fs) and two D-Mis DNVs (p.R310L and p.T482I) that both impact conserved residues in the helicase domain and predicted to disrupt separation of double-stranded DNA strands (**Figure 1D**, **Table 2, Supplementary Figure 2**). *PURA*, encoding the Transcriptional Activator Protein Pur-Alpha mutated in AD neurodevelopmental disorder with neonatal respiratory insufficiency, hypotonia, and feeding difficulties (OMIM # 616158), contained one LoF (p.A13fs) and two D-mis DNVs (p.F73S and p.I188T) (**Figure 1D**, **Table 2**). p.F73S and p.I188T impact highly conserved residues in the DNA- and RNA-binding PUR II domain of the protein critical for transcriptional regulation (**Figure 1D**).

MAP2K1, encoding mitogen-activated protein kinase (MEK1), contained three D-mis DNVs, including p.T55P and the recurrent p.Y130C mutation in two unrelated AC probands. p.T55P and p.Y130C are predicted to destabilize the ATP binding site of the MEK1 kinase domain (**Figure 1D** and **Table 2**, **Supplementary Figure 4**). MEK1 mediates FGF2-induced chromatin remodeling by regulating H3 methylation [46] and is required for gliogenesis by regulating the expression of transcription factors Etv5/Erm and CoupTF-II [47]. Somatic *MAP2K1* variants have been identified in several cancers [48] and germline variants in *MAP2K1*, including p.Y130C, have been described in AD Cardiofaciocutaneous syndrome type 3 (OMIM # 615279) (**Supplementary Table 2**) [49]. The orthologous *Mek1*^{Y130C} mutation in mice leads to MAPK hyperactivation sensitive to MEK and RAF inhibition [49] and recapitulates specific cardiofaciocutaneous phenotypes [50].

Notably, while many AC probands harboring variants in these exome-wide significant genes exhibited developmental delay, and some had other findings suggestive of a syndrome, few probands showed the constellation of phenotypes that would be typical of a single specific syndrome (**Supplementary Figure 5**). For example, the unrelated AC probands harboring the identical *MAP2K1 de novo* p.Y130C variant, while both having cerebral palsy and seizures, did not have the other classic features characteristic of cardiofaciocutaneous syndrome, which may, in part, explain why the physician referred these patients for exome genetic testing rather than more targeted gene panel testing. These findings underscore the phenotypic heterogeneity among patients with mutations in these genes and show how molecular diagnosis can assist when the clinical diagnosis is complicated by an atypical presentation.

DNVs impact genes in pathways regulating the epigenome

To identify other potential AC genes and gain insight into the molecular pathways impacted by their DNVs, we performed GOrilla pathway analysis (see **Methods**) on genes with a pLI > 0.9 that harbored ≥2 protein-altering DNVs with at least one protein-damaging DNV ("high-confidence (hc)-AC genes" [28 total genes]) and ≥1 damaging DNVs ("possible (p)-AC genes" [124 total genes]) (**Supplementary Table 3**). Among top-scoring gene ontology (GO) terms, we identified the most significant enrichment in biological processes related to nervous system development (GO:0007399; 6.48-fold enrichment; modified Fisher's exact test $P = 3.03 \times 10^{-5}$ after Bonferroni correction), DNA-templated transcription regulation (GO:0045893; 5.21-fold enrichment; $P = 6.04 \times 10^{-4}$), chromatin remodeling (GO:0006338; 3.24-fold enrichment; P = 0.003), and histone-lysine N-methyltransferase activity (GO:0018024; 3.07-fold enrichment; P = 0.004). (**Supplementary Figure 6**) Restricting pathway analysis to hc-AC genes revealed the top-scoring overall enriched biological process to be chromatin remodeling (GO:0006338; 4.12-fold enrichment; P = 0.006) (**Figure 2A**). Accordingly, damaging DNVs in genes of GO:0045893 (3.22-fold enrichment; $P = 2.62 \times 10^{-13}$), GO: 0006338 (9.27-fold enrichment; $P = 1.89 \times 10^{-17}$), and GO:0071565 (30.7-fold enrichment; $P = 3.35 \times 10^{-11}$) were significantly enriched in AC cases but not in controls (**Figure 2A**; **Supplementary Figure 6**).

20 out of 296 damaging DNVs were identified in genes included in GO:0045893 (positive regulation of DNA-templated transcription), including nine LoF and 11 D-mis variants in a total of seven genes (**Supplementary Table 4**, **Supplementary Figure 6**). Multiple GO:0045893 genes were also represented in the enriched histone-lysine N-methyltransferase activity GO:0018024 gene set and the chromatin remodeling GO:0006338 gene set. For example, two damaging DNVs (one nonsense, one splice-site) were identified in *ARID1A*, a component of the same neural-specific nBAF ATP-dependent chromatin remodeling complex that includes two

other exome-wide significant AC genes, *ARID1B* and *ADNP* (**Figure 2B**) [51]. Additionally, damaging DNVs were identified in *KMT2D* (1 frameshift) *and KMT2A* (1 splice-site). Like exome-wide significant AC gene *KDM5C* (**Figure 2B**), *KMT2D* and *KMT2A* modify chromatin via H3K4 methylation [52]. These results implicate epigenomic dysregulation due to germline DNV in AC pathogenesis (**Figure 2B**).

AC genes converge in midgestational neural precursors and arachnoid cells

To gain insight into the developmental time points, cell types, and molecular pathways involved in AC biology, we performed spatiotemporal consensus Weighted Gene Co-expression Network Analysis (WGCNA) [53], leveraging a large bulk RNAseq data set encompassing multiple human brain regions across development and into early childhood [54]. We constructed 88 modules characterized by genes that share highly similar expression patterns during brain development across different cortical regions and therefore likely to be involved in similar functions [55]. Each module was assessed for relative enrichment of genes for AC, autism, epilepsy, congenital heart disease, and human height (as a negative control) using logistic regression (please see **Methods** for details regarding gene lists).

Both hc-AC and p-AC genes converged in several modules in the midgestational human cortex [54]. The most significant of these was the post-conceptional week (PCW) 12-17 “royal-blue” module (adjusted $P = 1.15 \times 10^{-12}$; 6.99-fold enrichment; **Figure 2C, D**). Notably, this module also exhibited the highest enrichment for autism genes but was not enriched for epilepsy genes (**Figure 2C**). The “brown” module was also enriched with AC genes (**Figure 2C**). Both the royal blue and brown modules exhibited peak gene expression early in development at PCW 12-17 (**Figure 2D**). In contrast, the only module to show enrichment for both AC genes and epilepsy genes, “lightcyan”, is expressed much later in neurodevelopment at 10-12 months (**Figure 2D**). Human height genes were not enriched in any of the defined modules. Notably, the royal blue module was the only enriched module when the analysis was restricted to hcAC genes alone (6.89-fold enrichment; adjusted $P = 1.86 \times 10^{-04}$; **Figure 2C**). Genes expressed in the royal blue module are enriched for those involved in histone regulation (WP:2369; adjusted $P = 1.21 \times 10^{-04}$), including H3 acetylation (GO:0043966; enrichment 9.72-fold; adjusted $P = 1.73 \times 10^{-07}$) and methyltransferase activity (GO:0018024; 11.2-fold enrichment; adjusted $P = 8.91 \times 10^{-07}$), including H3-trimethyl-K4-specific modifications (GO:0042800; 8.14-fold enrichment; adjusted $P = 1.19 \times 10^{-05}$) (**Figure 2**).

We next studied AC gene expression using two different sc-RNAseq transcriptomic atlases. One comprised 144,047 total cells from 20 brain regions and 11 developmental periods [56]; the second included 8,851 total cells from all three layers of the embryonic forebrain meninges [57]. We again constructed modules comprised of co-expressed genes which we then mapped to specific cell type clusters. These were then assessed for enrichment of AC and other disease genes (see **Methods**). Of the 42 cell subtypes of brain parenchymal cells [56] (**Figure 2E, F**), hcAC genes were most significantly enriched in two subtypes: “ExN9” cells (adjusted $P = 1.99 \times 10^{-04}$), an excitatory embryonic neuron sub-cluster, and “InN4b” cells (adjusted $P = 7.94 \times 10^{-04}$), middle layer interneuron clusters originating from the medial ganglionic eminence [56] (**Figure 2**). Corroborating the overlap between hcAC genes and autism genes noted in our bulk RNAseq WGCNA analysis (see **Figure 2C**), both ExN9 and InN4b cell types are among the most highly enriched cell types for autism genes (**Figure 2H**). hcAC genes were also enriched in the “OPC3” oligodendrocyte progenitor cells (OPC), a subtype of glial cells responsible for myelin regeneration (adjusted $P = 2.00 \times 10^{-03}$) (**Figure 2H**). Human height genes were not enriched in any of the 42 cell subtypes.

In a parallel analysis performed on the meningeal dataset [57] (**Figure 2F**), p-AC genes were enriched in a cluster of “meninges 3 (arachnoid)” cells ($P = 2.31 \times 10^{-03}$) as well as “meninges 4 (dura)” cells ($P = 4.62 \times 10^{-03}$) (**Figure 2I**). Module 3 was the only module both enriched in hcAC genes and significantly enriched in specific meningeal cell types (meninges 3, $P = 3.65 \times 10^{-3}$, and meninges 4, $P = 1.21 \times 10^{-2}$ **Supplementary Figure 7**). Genes in this module were enriched for pathways related to the structural integrity of the arachnoid cell barrier [58,59], including the regulation of stress fiber assembly (GO:0051492; 6.5-fold enrichment; Adjusted P =

8.15 x 10^{-5}), focal adhesion (GO:0005925; 3.73-fold enrichment; Adjusted P = 2.33 x 10^{-05}), and cell-cell adhesion (GO:0098742; 5.0-fold enrichment; adjusted P = 3.81 x 10^{-7}) (**Figure 2J. Supplementary Figure 7**).

A classification of ACs derived from unsupervised clustering of phenome data

Since the majority of our AC patients came from a pediatric exome referral population, we examined the AC genes for whether they are associated with additional phenotypes and known or suspected Mendelian diseases that could explain the patients' complex presentations. The overlap of exome-wide significant AC genes with multiple OMIM disease genes (**Supplementary Table 2**) prompted our examination of AC genes with other diseases using DisGeNET [60] (**Methods**). The results demonstrate a highly significant overlap between AC genes and genes implicated in muscle hypotonia (P = 4.08 x 10^{-57}: one-tailed Fisher's exact test), dysmorphic features (P = 9.71 x 10^{-57}), intellectual disability (1.51 x 10^{-55}), delayed speech and language development (1.73 x 10^{-52}), and others (**Supplementary Table 5** and **Figure 3A**). We therefore systematically characterized the associated phenotypes of our AC probands by examining their health histories using a natural language processing algorithm (txt2hpo: https://github.com/GeneDx/txt2hpo) (**Methods**). Several HPO terms were frequently present in our AC probands, including "developmental delay" (71.7%, HP:0012758; ranging from mild forms such as delayed social development [HP:0012434], mild global developmental delay [HP0011342], and delayed ability to stand [HP:0025335], to more severe forms such as severe receptive/expressive language delay [HP:0011352/HP:0006863] and profound global developmental delay [HP:0012736]), "craniofacial anomalies" (45.7%, HP0000234; encompassing but not limited to abnormal skull morphology [HP:0000929]), abnormality of the fontanelles of cranial sutures [HP0000235], and decreased calvarial ossification [HP:0005474], often seen overlying ACs); "seizures" (34.8%, HP:0001250); "macrocephaly" (18.1%, HP:0000256); and "autism" (17.8%, HP:0000729) (**Supplementary Table 1**).

After running UMAP (Uniform Manifold Approximation and Projection) on patient phenotypes (**Methods**), AC probands were clustered using HDBSCAN into four clinical clusters (**Figure 3B**) [61]. Clusters were characterized using two different statistical metrics that highlight the commonality of certain terms (phenotypic traits) within a cluster: Term frequency (Tf) and Term frequency – Inverse Document Frequency (Tf-IDF) (**Methods**). As expected, all clusters had a Tf of 1.0 for the term "arachnoid cyst" (HP:0100702) (**Supplementary Table 6**). Cluster 1 was not distinguished by any outlier Tf-IDF rates (as defined by a Tf-IDF >1.5 times the IQR above the third quartile) and thus had no distinctive non-AC phenotypic marker (**Figure 3, Supplementary Table 6**). In contrast, the Tf-IDF statistic identified "seizures" (Tf=1.0; HP:0001250) as an outlying phenotype for cluster 2, "hypotonia" (Tf=0.99, HP:0001290) for cluster 3, and "hypotonia" and "seizures" for cluster 4 (**Figure 3**). A complete list of Tfs per cluster can be found in **Extended Data Table 3.** Based on these data, we classified cluster 1 ACs as low phenotypic complexity ("simple"), clusters 2 and 3 as medium phenotypic complexity ("mixed"), and cluster 4 as high phenotypic complexity ("complex"). Consistent with severe neurodevelopmental phenotypes being easier to diagnose at earlier ages, the average age of probands with complex ACs was younger (cluster 4; 8.1 $\pm$ 0.05 yrs) than those with simple ACs (cluster 1; 11.5 $\pm$ 0.1 yrs; $p<0.0001$ by two-tailed t-test).

Genes with synonymous DNVs were not enriched in any specific clinical AC cluster (**Extended Data Table 4**). Compared to simple (cluster 1) and mixed (cluster 2 and 3) ACs, complex (cluster 4) ACs were more highly associated with genes harboring damaging DNVs, including hcAC and exome-wide significant genes (**Figure 3, Extended Data Table 4**). *ARID1B, DDX3X, KDM5C,* and *PURA* contributed disproportionately to this signal, with >50% of DNVs in these genes localizing to the complex AC cluster (**Figure 3)**. Higher-complexity AC clusters were also enriched for DNVs in genes associated with 9 of our top 10 DisGeNET phenotypes (**Figure 3, Extended Data Table 4**). DNVs in genes associated with specific DisGeNET phenotypes were enriched in the corresponding clusters defined by the matching outlier Tf-IDF phenotype terms. For example, Clusters 3 and 4, uniquely identified by Tf-IDF as hypotonia clusters, also showed significant enrichment for DNVs in known DisGeNET hypotonia genes (**Figure 3, Extended Data Table 4**). These results suggest unsupervised clustering can segregate AC probands into meaningful phenomic-genomic subtypes.

DISCUSSION

To date, damaging genomic variants have not been considered major contributors to ACs, but our findings challenge this dogma. We provide cohort-level statistical evidence that rare, damaging DNVs are associated with AC. The cohort-wide enrichment of DNVs is consistent with the observation that most AC cases are sporadic [62]. The observed fraction of patients with damaging variants allows the inference that ~20% of studied cases studied are attributable to damaging DNVs (see **Methods**). These data indicate that genomic variants represent an important, independent, and historically overlooked contributor to the etiology of AC.

We found evidence in our cohort for both known disease genes and genes previously unassociated with human phenotypes. Identification of independently arising, identical DNVs in *ADNP* and *MAP2K1* indicates that monogenic contributions to AC exist but could be under-recognized. Notably, all seven of the exome-wide significant AC genes are highly expressed in the fetal brain and all are associated with an OMIM-catalogued Mendelian disease that features prominent neurodevelopmental phenotypes as part of the syndrome (see **Supplementary Table 2**). ACs may therefore represent anatomic correlates or radiographic biomarkers of more fundamental defects in brain development, and the associated neurodevelopmental phenotypes in AC patients may arise not from cyst mass effect *per se* but rather from impairment of neuronal connectivity secondary to germline DNV. The high association of AC genes with other DisGeNET phenotypes [60], the relative lack of efficacy of AC fenestration for non-hydrocephalic phenotypes such as epilepsy [63], and the increased rates of ACs in Mendelian diseases such as Acrocallosal, Chudley-McCullough, Aicardi, and Pallister-Hall syndromes, among others [24-26] all support this hypothesis.

Chromatin is dynamically modulated by proteins that epigenetically modify DNA and histones, as well as by larger protein complexes that physically alter genome accessibility and topology. Spatiotemporal coordination of chromatin activity through its regulation of gene transcription [64] is essential for the fundamental cellular processes that underlie brain morphogenesis and neuronal connectivity [65]. Our gene- and cohort-wide network biology analyses support a major role for chromatin modifiers and other transcriptional regulators in AC pathobiology (**Figure 2B**). These include interacting components of the neural-specific ATP-dependent BAF (SWI/SNF) chromatin remodeling complex [66,67], and multiple enzymes involved in H3K4 methylation editing [68]. These genes overlap with those containing the greatest number of sequence variants in DECIPHER, a clinical and genetic database of individuals with global developmental delay, intellectual disability, microcephaly, and other developmental conditions [69]. The above findings also imply dosage sensitivity for these chromatin marks in AC, recalling the haploinsufficiency for chromatin-modifying and -remodeling genes previously implicated in autism [70], congenital heart disease [71], and diverse cancers [72]. Investigation of the consequences of these mutations on specific enhancers and promoters and the genes they regulate should provide further insight into AC pathogenesis.

Integrated exome and transcriptome analysis (including scRNAseq data from >152,898 cells of the developing human brain and meninges) shed insight into the temporal dynamics and spatial patterns of AC gene expression. AC genes converge in neural precursor cells and developmental epochs that overlap with autism genes, and to a lesser extent, epilepsy genes, and are related to the function of developing excitatory and inhibitory circuits. This may shed light on the association of ACs with, and lack of surgical efficacy for, these and other co-morbid neurodevelopmental phenotypes in AC patients [73 74,75]. AC gene expression in arachnoid barrier cells, and their potential regulation of pathways in these cells that are pertinent to the mechanotransduction apparatus that links arachnoid cells to one another and the extracellular matrix [76] is consistent with the six-decade-old developmental “splitting” hypothesis of the bilayered arachnoid epithelium in AC pathogenesis [77]. These hypotheses will be difficult to test in mice and other smooth-brained model systems, in which ACs are absent. Nonetheless, the next experimental steps could include scRNA-seq interrogation of control and patient-derived tissue to define impaired transcriptional networks; ChIP-seq to characterize chromatin modifications and binding proteins; and deep exome sequencing to detect somatic

variants that may explain the unilaterality and phenotypic heterogeneity of cysts and associated phenotypes [78].

Our agnostic, comprehensive analysis of physician-reported phenotypes revealed high rates of neurodevelopmental phenotypes such as developmental delay, epilepsy, and autism in our AC probands (**Supplementary Table 1, Extended Data Table 3**). This is consistent with earlier studies reporting the association of neurodevelopmental phenotypes in ACs [20], albeit with a higher frequency in our study. The difference likely reflects our predominantly pediatric AC cohort, the inclusion of a large number of cases from a clinical referral laboratory, and our systematic phenotype assessment of probands using natural language processing. As most ACs lacking obstructive hydrocephalus are not usually referred by neurosurgeons for formal neurodevelopmental assessment to uncover mild or more subtle cognitive and motor delays, we hypothesize that prior rates of incidence of these phenotypes are underestimates. Continued implementation of natural language processing to assist in deep phenotyping of large cohorts may reveal similar trends in other neurosurgical diseases.

A limitation of our study is the failure to consider other types of genetic variants and non-genetic inheritance (i.e., epigenetics). Rare, *de novo* copy number variants may account for another small fraction; rare or common transmitted variants may account for many more. Non-coding mutations cannot be dismissed. Evidence of dosage sensitivity of many chromatin-modifying genes raises the possibility that environmental perturbations of these pathways in critical developmental windows might phenocopy the effects of these mutations. Still another limitation is our focus on DNVs without evaluating epistatic effects and other types of complex genetic variations. Future cohorts combining whole-genome sequencing, transcriptomics, and epigenomics will help elucidate these questions.

Our findings have clinical implications worthy of validation in other, larger cohorts. In cases where ACs cause obstructive hydrocephalus from mass effect, the decision for neurosurgical decompression or shunting is straightforward and can be lifesaving. However, the decision to operate on ACs in the context of seizures, language or motor delay, or neurobehavioral and psychiatric symptoms can be challenging [63,79-81]. Nonetheless, a survey of neurosurgeons indicated >30% would recommend surgery for a Sylvian fissure AC if the patient presented with psychomotor retardation [82]. Surgical outcomes in such cases are inconsistent [83]. Exome sequencing in these scenarios may be a useful adjunct to help guide treatment decisions and prognostication. For example, the discovery of a damaging DNV in a high-pLI OMIM gene might increase the threshold for a neurosurgeon to offer surgery, even in the presence of very impressive MRI findings (e.g., see **Figure 1A**), because signs and symptoms in these cases may more directly reflect an inborn genetic insult than pathological sequelae of AC-dependent mass effect. The patient and family may be better served by referral for genetic follow-up and counseling and early interventions for speech, neurobehavioral, and physical therapies.

Neurosurgery is predicated on operative precision to achieve its therapeutic goals while limiting morbidity and mortality. The history of neurosurgical progress has been dependent on technological breakthroughs allowing ever-increasing operative precision. Examples include the introduction of microsurgical techniques enabled by advances in optics, and intra-operative cranial navigation driven by innovations in neuroimaging. Recent advances in DNA sequencing, single-cell transcriptomics, and natural language processing, along with the now almost standard use of the electronic medical record, have coalesced to make the present an opportune moment to increase knowledge of ACs and other historically understudied sporadic structural brain diseases , including hydrocephalus [35,84], Chiari malformations [85], neurovascular malformations [32,86,87], and craniosynostosis [88,89]. These notoriously heterogeneous disorders have often been treated with one-size-fits-all operative approaches, leading to varied and unpredictable treatment responses, sometimes with devastating consequences. Continued study of these disorders with systems-level, integrative multi-omics approaches may replace or supplement current antiquated, anatomically-based disease classification systems

with a molecular nomenclature that could increase precision for counseling, prognostication, and surgical treatment stratification – including when not to operate (*primum non nocere*).

METHODS

Patients and data harmonization

Trios comprised patients from an academic healthcare-based center (Yale-CMG) and a clinical laboratory referral center (GeneDx). Regarding the former, all study procedures and protocols comply with Yale University's Human Investigation Committee and Human Research Protection Program. Written informed consent for genetic studies was obtained from all participants. Patients and participating family members provided buccal swab samples (Isohelix SK-2S DNA buccal swab kits), medical records, neuroimaging studies, operative reports, and phenotypic data. For GeneDx, all study procedures and protocols comply with guidelines set forth by the Western Institutional Review Board, Puyallup, WA (WIRB 20162523). Informed consent was obtained from all individuals undergoing genetic testing and/or medical record review, and WIRB waived authorization to use de-identified aggregate data. Individuals or institutions who opted out of this type of data use were excluded. Samples and patient data were provided to GeneDx as described [90]. Criteria for inclusion of GeneDx and Yale-CMG probands in this study was the presence of radiographically defined primary (i.e., idiopathic) cerebral ACs, as read by a neuroradiologist and confirmed by a neurosurgeon when needed. Participants were excluded if their medical records indicated the AC was of a secondary origin and therefore likely to be a brain cyst of a different nature (e.g., a porencephalic cyst secondary to brain hemorrhage, or a tumor-related cyst).

Medical record data from patients were abstracted into Human Phenotype Ontology (HPO), which include positive and negative confirmation across over 14,000 phenotypes in the HPO library (August 2021 release date; https://hpo.jax.org/app/). Extraction was performed using *txt2hpo* (https://github.com/GeneDx/txt2hpo), a python library for extracting HPO-phenotypes from text, with natural language capabilities including inflection (i.e., 'hypotonia' vs. 'hypotonic') and negation (i.e., 'hydrocephalus' vs. 'no hydrocephalus'). To avoid sampling bias, phenotypic data underwent a process of harmonization. Because phenotyping of the GeneDx cohort was performed with *txt2hpo* on complete medical records, it cataloged the absence and presence of phenotypic traits across over 14,000 variables. The Yale-CMG cohort was phenotyped across 860 variables due to the survey-based nature of patient data collection. We thereby restricted the GeneDx phenotyping to the identical variable set queried in the Yale-CMG cohort (**Extended Data Table 5**). This avoided variables unique to the GeneDx cohort becoming drivers of the UMAP/clustering process and obviated any cohort effect. Yale-CMG variables were then translated into HPO language (output of *txt2hpo*) after concatenating both datasets. Phenotypic clustering using the UMAP/HBDSCAN method was performed (See below).

Controls consisted of 1,798 unaffected siblings of people with ASD and unaffected parents from SSC [43]. Permission to access the genomic data in the SSC on the National Institute of Mental Health Data Repository was obtained. Written informed consent for all participants was provided by the Simons Foundation Autism Research Initiative.

Exome sequencing and variant calling

For Yale-CMG cohort, exome capture was performed on genomic DNA samples derived from saliva or blood using the NimbleGen SeqCap EZ MedExome Target Enrichment kit (Roche) or the xGen target capture kit (IDT), followed by 101 or 148 base paired-end sequencing on Illumina platforms as described [35]. Sequence reads were aligned to human reference genome GRCh37/hg19 using BWA-MEM. Single nucleotide variants and small indels were called using GATK HaplotypeCaller [91,92] and FreeBayes [93] and annotated with ANNOVAR [94]. Minor allele frequencies were annotated using the gnomAD and Bravo databases [95-97]. For the GeneDx cohort, samples were sequenced and aligned with either SureSelect Human All Exon v4 (Agilent),

Clinical Research Exome (Agilent), or xGen Exome Research Panel v1.0 (IDT), and sequenced with either 2x100 or 2x150bp reads on Illumina HiSeq 2000, 2500, 4000, or NovaSeq 6000 as previously described [90]. Aligned BAM files (GRCh37/hg19) were converted to CRAM format with Samtools version 1.3.1 and indexed. Individual gVCF files were called with GATK v3.7-0 HaplotypeCaller [91,98] in gVCF mode by restricting output regions to plus/minus 50bp of the RefGene primary coding regions. Single-sample gVCF files were then combined into multisample gVCF files, with each combined file containing 200 samples. These multi-sample GVCF files were then jointly genotyped using GATK GenotypeGVCFs [91]. GATK VariantRecalibrator (VQSR) was applied for both SNPs and small INDELs, with known SNPs from 1000 Genomes phase 1 high confidence set and "gold standard" INDELs from Mills et al [99]. Variants in VQSR VCF files were then annotated with ANNOVAR as previously described.

De novo calling and filtering

DNVs were called as previously described [42]. DNV calls were filtered to create a high-quality set using the parameters below:

- Read depth (DP) >10 in the proband and both parents
- Variant allele frequency (VAF) > 0.15 in the proband for SNVs and VAF > 0.25 for indels
- > 3 reads supporting the alternative allele
- Genotype Quality (GQ) score > 40
- Log odds of being a true variant versus being false from VQSR > -10 outputted from GATK
- Any variant with a general population frequency above 0.01 was also excluded based on 1000 Genomes and ExAC variant population frequency data.
- Filter out DNVs called > 4 times in the parental samples in the cohort
- Filter out DNVs with VAF < 0.3 and VQSLOD < 7
- Filter out *de novo* indels > 100bp
- Filter out *de novo* variants not on chromosome X, with a VAF of 1

The impact of non-synonymous variants on protein function was inferred using DeNovoWEST [42] (https://github.com/queenjobo/DeNovoWEST), which scores all classes of sequence variants on a unified severity scale based on empirically estimated positive predictive pathogenicity values. This software was utilized to assess gene-wise DNV enrichment. Orthogonal analysis with DenovolyzeR classified missense variants as "deleterious" (referred to as D-mis) when predicted as deleterious by MetaSVM or with MPC score ≥ 2 [100-102]. Inferred loss-of-function (LoF) variants include stop-gains, stop-losses, frameshift-insertions and -deletions, and canonical splice site variants. Protein-damaging variants included LoF and D-mis variants and protein-altering variants comprised of LoF and all missense variants.

Kinship analysis

The relationship between proband and parents was estimated with pairwise identity-by-descent (IBD) calculation in PLINK [103]. The IBD sharing between the proband and parents in all trios is between 45% and 55%.

De novo enrichment analysis

The burden of DNVs in AC cases and in unaffected ASD controls was respectively determined using the DeNovoWest [42] and the denovolyzeR tools [102]. DeNovoWEST is the testing framework we used to assess gene-wise DNV enrichment (https://github.com/queenjobo/DeNovoWEST). For DeNovoWest, each observed DNV in our dataset was assigned a mutation severity score, a proxy for expected deleteriousness of the mutation[42]. For each gene, we then calculated a gene severity score, the sum of all mutation severity scores for that gene. There are two components to DeNovoWEST: the overall enrichment test, which includes all

variant consequences, and the gain-of-function specific test, which assesses enrichment and clustering of missense variants only. Each gene was subjected to an enrichment test using all non-synonymous DNVs, followed by a test designed to detect genes probably acting by an altered function mechanism (combining of missense enrichment test with a missense clustering test). We then applied a Bonferroni multiple-testing correction accounting for the number of genes (n = 19,347) and two tests per gene. The value of correction for DeNovoWest is $\alpha = 1.29 \times 10^{-6}$ = (0.05/[2 tests x 19,347 genes])

For DenovolyzeR, the expected number of DNVs in the case and control cohorts across each functional class was calculated by taking the sum of each functional class-specific probability multiplied by the number of probands in the study x 2 (diploid genomes). Then, the expected number of DNVs across functional classes was compared to the observed number in each study using a Poisson test [104]. Gene-set enrichment analyses considered only mutations observed or expected in genes within the specified gene set (i.e., high brain-expressed, LoF-intolerant).

To examine whether any individual gene contains more protein-altering DNVs than expected, the expected number of protein-altering DNVs was calculated from the corresponding probability, adjusting for cohort size. The Poisson test was then used to compare the observed DNMs for each gene versus expected. As separate tests were performed for protein-altering, protein-damaging and LoF DNVs, the relevant Bonferroni multiple-testing threshold is, therefore, equal to $\alpha = 8.6 \times 10^{-7}$ (= 0.05/[3 tests x19,347 genes]).

Gene lists for specific diseases

Genes were classified as high-confidence AC genes if they reached multiple testing-corrected exome-wide significance in our analysis, or if they had two or more protein-altering DNVs with at least one damaging variant in predictive loss-of-function-intolerant (pLI) genes. Genes were classified as "possible" if they harbored ≥ 1 damaging DNV in a gene with pLI score greater than 0.9 in ExAC. As a result, we identified 28 high-confidence AC genes and 124 possible AC genes (see **Supplementary Table 3**). Gene lists not related to AC were adapted from publications referenced [33,105,106].

Weighted Gene Co-expression Network Analysis (WGCNA)

A processed bulk-RNAseq expression data set encompassing sixteen human brain regions across the human development [107] was used for robust consensus WGCNA. Analysis was limited to timepoints between gestational week 9 and postnatal year 3. We removed samples > 3 standard deviations above the mean sample network connectivity. Network analysis was performed with WGCNA [108], assigning genes to specific modules based on bi-weight mid-correlations among genes. Soft threshold power of 10 was chosen to achieve scale-free topology (r2 > 0.9). Then a signed co-expression network was generated. The topological overlap matrix was clustered hierarchically using average linkage hierarchical clustering (using `1 – TOM` as a dis-similarity measure). The topological overlap dendrogram was used to define modules using a minimum module size of 40, a deep split of 4, merge threshold of 0.1.

Module enrichment analysis

Module gene lists were obtained via WGCNA as described above. In a background set of all genes categorized in co-expression modules, logistic regression was used for an indicator-based enrichment: is.disease ~ is.module + gene covariates (GC content, gene length, and mean expression in bulk RNA-seq atlas), as described previously (Walker, Ramaswami et al. 2019). Of the 88 WGCNA modules, the gray module, by WGCNA convention (Li, Santpere et al. 2018), contains all genes that do not co-express and are consequently unassigned to a co-expression network. Thus, the gray module was excluded from enrichment testing, and enrichment significance was defined at the Bonferroni multiple-testing cutoff ($\alpha = 5.68 \times 10^{-04}$).

Cell-type enrichment analysis

Cell-type-enriched genes (cell type markers) were obtained from a scRNA-seq atlas of the human brain spanning from early fetal development into adulthood and from a scRNA-seq atlas of the mouse meninges. In a background set of all genes expressed in ≥3 cells of the scRNA-seq atlas, logistic regression was applied for indicator-based enrichment analysis: is.cell.type ~ is.disease + gene covariates (GC content, gene length). All p-values were adjusted with Bonferroni correction ($\alpha = 1.19 \times 10^{-03}$ for the brain parenchyma and $\alpha = 8.33 \times 10^{-03}$ for the meninges)

Gene Ontology enrichment analysis

Using the identified AC gene lists and gene modules significantly enriched with the AC gene list, enrichment analysis was performed for gene ontologies (biological processes, cellular components, and molecular functions), biological pathways (Wiki pathways Human and Mouse), and upstream transcription factors (TRANSFAC and JASPAR) using the EnrichR R package Version 3.0 [109]. The top 10 terms with the lowest adjusted p-values were reported in the order from highest to lowest combined Z-scores. Similar analyses were also conducted with these lists using GOrilla [110], and QIAGEN Ingenuity Pathway Analysis [111]. Adjusted p-values < 0.05 were considered significant.

Meninges single-cell transcriptomic analysis

Single-cell gene expression in 8,581 cells was recorded from all layers of the mouse meninges. Seurat version 4.0.3 was used to cluster cells into tissue layers characterized by specific meningeal layer gene markers as well as sub-clusters delineating meningeal layer cell subtypes within each cluster, as described previously [57]. The resulting Seurat object was then imported into Monocle3 version 0.2.3 [112] where 24 gene modules were created based upon relative co-expression, excluding genes found not to have any significant differential or co-regulatory expression. Bonferroni-corrected anatomical enrichment analysis was performed for each of the 24 modules to identify specific meningeal tissue layers and meningeal cell subtypes enriched for each module. Further, Bonferroni-corrected hypergeometric enrichment analysis was performed for the pAC, hcAC, and exome-wide significant gene lists in each of the 24 gene modules to identify AC gene list-enriched modules. AC-enriched gene modules were then imported for gene ontology analysis in EnrichR version 3.0 as previously described [109].

Human Phenotype Ontology (HPO) Uniform Manifold Approximation and Projection (UMAP) analysis

Given the comprehensive nature of the HPO vocabulary, and therefore, the relative paucity of some phenotypes within our cohort (~99% of all possible known phenotypes with no representation across our study population), the entire HPO vocabulary was collapsed into 1,000 clinically relevant parent phenotype groups in a process we termed pre-UMAP dimensionality reduction. Groups were based primarily on organ system (e.g., “inflammatory/rheumatic conditions”). Groups with any positive representation amongst our cohort are in **Extended Data Table 5**). Phenotypes relating to the neuroaxis, and therefore highly relevant to AC pathology, were excluded from dimensionality reduction to maintain a high degree of resolution on subtle differences in patients’ clinical neurological presentations (**Extended Data Table 5**). Participants' HPO term/group matrices were subjected to uniform manifold approximation and projection (UMAP) [111] and displayed against two arbitrary X (UMAP1) and Y (UMAP2) axes. Clusters were assigned using HDBSCAN, a high-density based clustering algorithm [113].

Cluster differentiation was defined in two ways:
1) Term frequency (Tf) for each HPO-phenotype was calculated follows:

$$Tf(t,c) = \frac{f_{t,c}}{\sum_{t' \in c} f_{t'c}}$$

where $f_{(t,c)}$ represents the raw count of a study participant phenotype matrices containing term '*t*' within a cluster '*c*' group, and $\sum_{t' \in c} f_{t'c}$ represents the raw count of all study participant matrices in the cluster containing any term. Tf of all terms were then ranked in order with top Tfs displayed in Extended Data Table 5.

2) Term Frequency-Inverse Document Frequency (Tf-IDF), which accounts for relative over-representation of terms that are common across clusters (i.e., all clusters have 1.0 Tf for arachnoid cyst phenotype (HP:0100702), was also calculated in the following way:

$$Tf - IDF = Tf(t,c) \times idf(t,C)$$

where

$$IDF(t,C) = log\ \frac{N}{|(c \in C : t \in c)|}$$

and where *N* is the total number of study participant matrices in the cohort (*C*) and $|(c \in C : t \in c)|$ = the number of study participant matrices in which the term *'t'* appears. Thus if term *'t'* is present in all participant matrices in the cohort, *IDF(t,C)* = 0 and Tf-IDF = 0. Statistical outliers were defined as values falling above an additional 1.5 times the interquartile range of the Tf-IDF value distribution.

Estimating the number of AC genes

A maximum likelihood approach was used to estimate the number of genes contributing to AC via *de novo* events as described previously [35]. We defined K to be the number of observed protein-damaging DNVs in LoF-intolerant genes among cases. R1 indicates the number of LoF-intolerant mutated exactly twice in cases and R2 indicates the number of LoF-intolerant genes mutated three times or more. We set the proportion (E) of protein-damaging mutations in genes based on the point estimate of enrichment in cases compared to expectation (E = (M1–M2)/M1, where M1 and M2 are the observed and expected count of protein-damaging variants per trio, respectively). We then simulated the likelihood function as follows: First, we randomly selected G genes from the of LoF-intolerant gene set. Next, we simulated the number of contributing protein-damaging variants in genes, i.e., C, by sampling once from Binomial (K,E) distribution. Then, we simulated C contributing protein-damaging variants in G genes and K-C non-contributing protein-damaging variants in the complete LoF-intolerant gene set, using each gene's protein-damaging mutability score as probability weights. We performed 20,000 simulations for G from 3 to 300 and calculated the likelihood function L(G) as the proportion of simulations in which the number of genes with two protein-damaging DNV equals to R1 and the number of genes with three or more protein-damaging DNVs equals to R2. We then estimated the number of genes using the maximum likelihood estimate (MLE). Based on the likelihood function, we calculated the Fisher information and constructed the confidence interval based on the MLE and estimated Fisher information using the following equation.

$$MLE \pm 1.96 \times \left(\frac{1}{\sqrt{Fisher\ Information}} \right)$$

ACKNOWLEDGMENTS

We are grateful to the patients and families who participated in this research for their invaluable role in this study.

Funding: This work is supported by the Yale-National Institutes of Health (NIH) Center for Mendelian Genomics (5U54HG006504); R01 NS111029-01A1, R01 NS109358, K12 228168, and the Rudi Schulte Research Institute (K.K.); NIH Medical Scientist Training Program (NIH/National Institute of General Medical Sciences Grant T32GM007205); NIH Clinical and Translational Science Award from the National Center for Advancing Translational Science (TL1 TR001864); the K99/R00 Pathway to Independence Award (K99HL143036 and R00HL143036-02), the Vernon W. Lippard Research Fellowship, the Hydrocephalus Foundation, the March of Dimes, the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

Study design and conceptualization: A.J.K. and K.T.K.
Cohort ascertainment, recruitment, and phenotypic characterization: A.J.K., G.A., S.M., K.M., V.G., E.K., P.Q.D., H.S., J.O., J.S., A.A., M.L.D., C.G.F., A.T.T., H.Q., A.E., P.R., B.S.C., S.L.P., C.A.W., M.G., R.P.L., F.M.Z., R.I.T., S.C.J., K.T.K.
Exome sequencing production and validation: I.R.T, C.C., F.L.-G., S.M.
Exome sequencing analysis: G.A., S.M., V.G., A.J.K., S.Z., Y.-C.W., A.T.T., J.K., P.-Y.F., W.D., F.M.Z., R.I.T., S.C.J., K.T.K.
Integrative genomics analysis: G.A., E.K., K.T.K.
Phenomics analysis: A.J.K., S.M., K.Y.M., V.G., A.M.-D.L., K.T.K.
Statistical analysis: G.A., A.J.K., S.C.J., W.D., S.P., B.L., Q.L.
Sanger sequencing validation: C.N.W.
Neuroimaging characterization: A.J.K., A.M.D., K.T.K.
Biophysical simulation: S.H.
Resources: C.N.W., S.M., C.A.W., M.G., R.P.L, R.I.T., S.C.J., K.T.K.
Writing and review of manuscript: A.J.K., G.A., S.M., K.M., E.K., S.L.A., C.A.W., M.G., R.P.L., F.M.Z., R.I.T., S.C.J., K.T.K.
Project administration: A.J.K., G.A., S.M., K.M., R.I.T., S.C.J., K.T.K.
Funding acquisition and supervision: R.P.L., S.C.J., K.T.K.

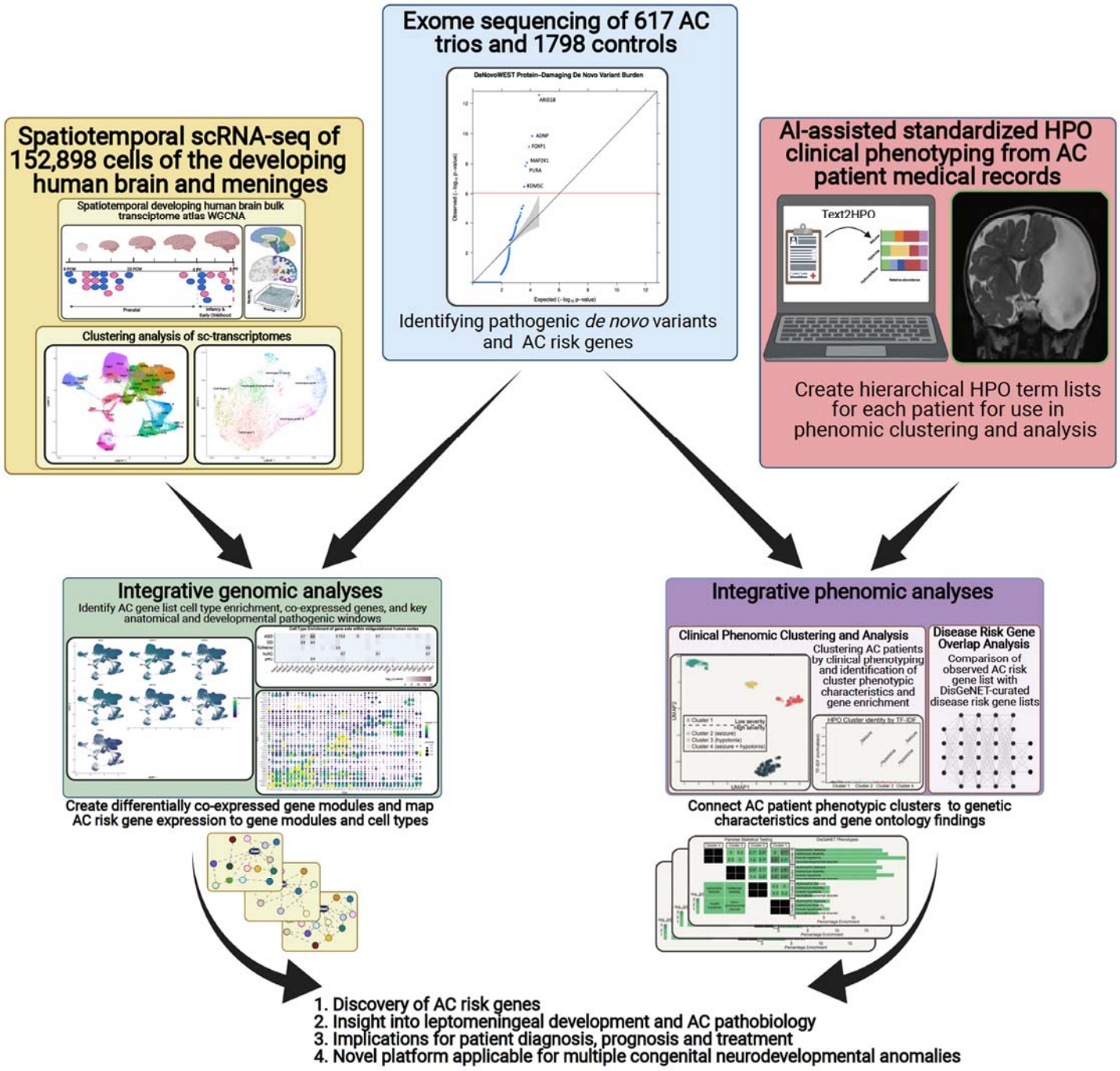
COMPETING INTERESTS

The authors declare no competing interests.

DATA AND SOFTWARE AVAILABILITY

The sequencing data for all congenital hydrocephalus parent-offspring trios and singletons from the healthcare-acquired cohort have been deposited in the NCBI database of Genotypes and Phenotypes and AnVIL (https://anvilproject.org/data?query=consortium%3DCMG) under the accession number phs000744.v4.p2. Sequencing data from GeneDx cannot be made available due to the nature of consent for clinical genetic testing. GeneDx has provided de-identified patient data to this study to improve clinical interpretation of genomic data, in accordance with the patient consent and the ACMG position statement on genomic data sharing. Our in-house pipelines and codes will be available upon request.

Supplementary Figure



758 **Figure 1.**

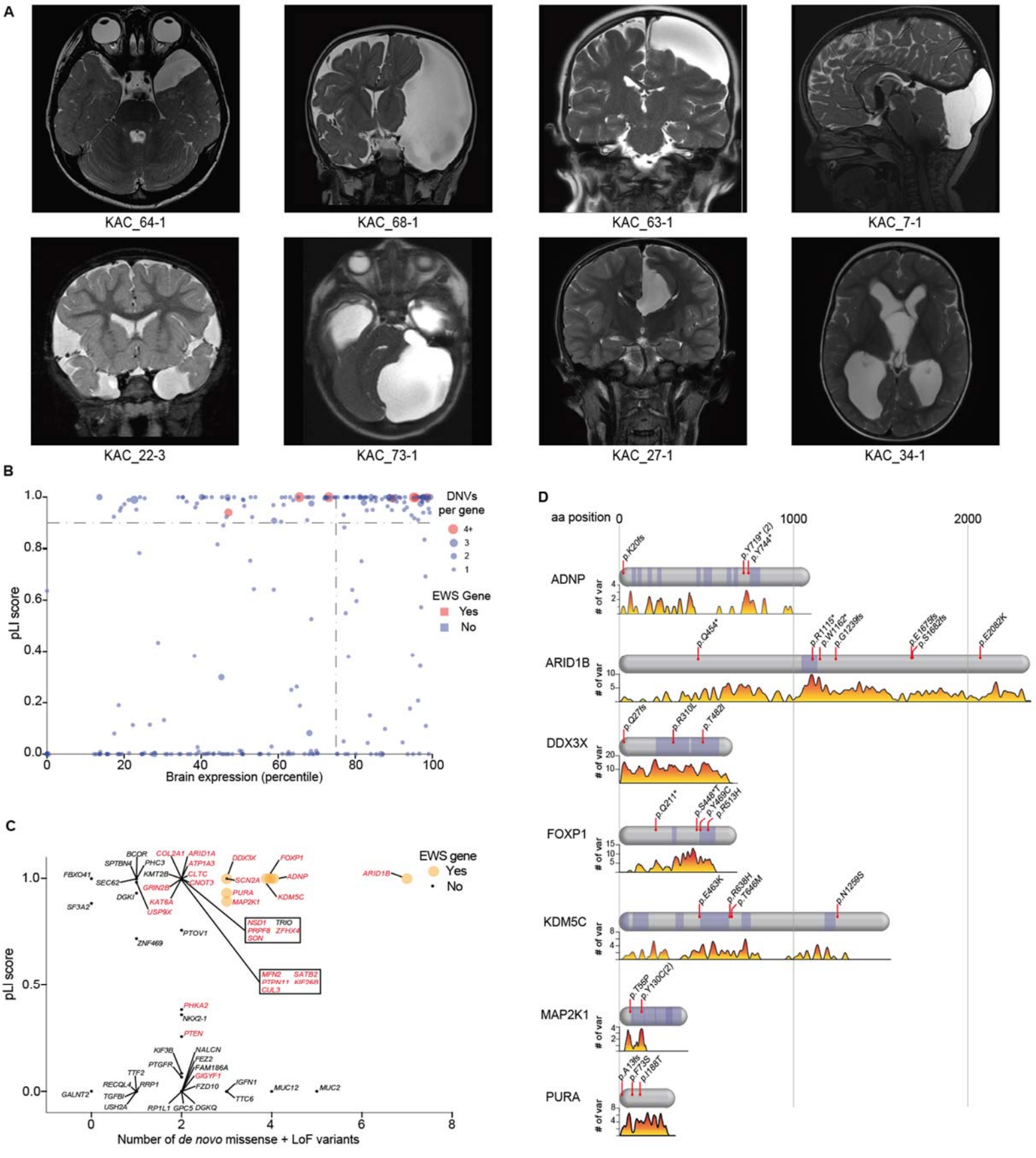
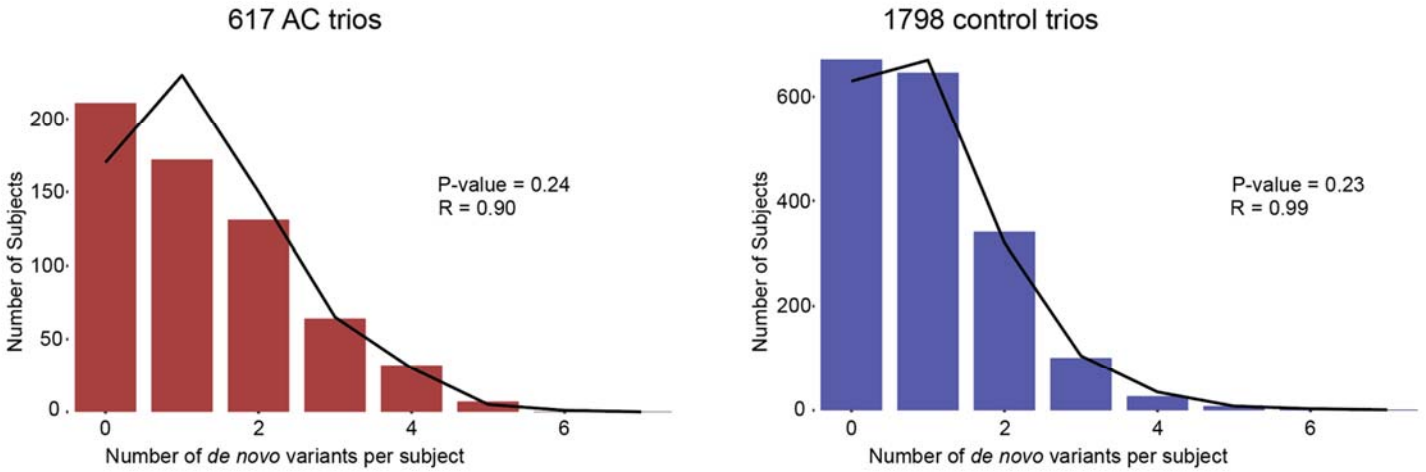

A
KAC_64-1
KAC_68-1
KAC_63-1
KAC_7-1
KAC_22-3
KAC_73-1
KAC_27-1
KAC_34-1
B
pLI score
Brain expression (percentile)
DNVs per gene
4+
3
2
1
EWS Gene
Yes
No
C
pLI score
Number of de novo missense + LoF variants
EWS gene
Yes
No
D
aa position
ADNP
ARID1B
DDX3X
FOXP1
KDM5C
MAP2K1
PURA
# of var

Figure 1. Arachnoid cysts (ACs) are associated with *de novo* variants (DNVs) in high brain-expressed (HBE) genes highly intolerant to loss-of-function (LoF) variants. (A) Representative imaging of AC probands. Axial, coronal, and sagittal T2-weighted brain magnetic resonance images (MRIs) demonstrate the locations and morphologies of representative ACs within our cohort. An anterior temporal AC in proband KAC_64-1. A holohemispheric AC eliciting a significant mass effect and adjacent cortical hypoplasia in KAC_68-1. A frontal convexity AC in KAC_63-1. A midline posterior fossa AC eliciting calvarial remodeling and significant mass effect on the cerebellum in KAC_7-1. A lateral posterior fossa AC causing effacement of the fourth ventricle and associated obstructive hydrocephalus in KAC_73-1. An interhemispheric AC with corpus callosum hypoplasia in KAC 27-1. Bilateral Sylvian fissure and anterior temporal ACs in KAC_22-3. A third ventricular AC causing obstructive hydrocephalus in KAC_34-1. **(B) AC risk genes with HBE highly intolerant to LoF.** LoF variants comprise premature termination, frameshift or splice site variants. pLI is a gene-wide constraint metric that estimates the probability of being intolerant to LoF variation. Mouse brain expression is determined by bulk RNA-seq expression at embryonic day (E) 9.5. The top quartile of brain expression is represented by the vertical dashed line and top 10% of LoF intolerance is represented by the horizontal dashed line. Individual dots represent individual genes in the AC cohort harboring at least one DNV. The size of the dot correlates with the number of DNVs identified. The red color represents genes reaching exome-wide significance (EWS). **(C) AC risk genes with higher numbers of DNVs are predominantly those with high pLI values.** Genes visualized are those with ≥2 DNVs. Large, orange dots represent those that surpass EWS thresholds. Genes in red font indicate those in the high-confidence AC gene set. **(D) DNVs in EWS AC genes cluster to mutational hotspots.** Locations of identified protein-altering and protein-damaging DNVs in EWS AC genes are represented as red dots (locations reported above) in relation to critical functional domains of the gene as reported in the UniProt database (https://www.uniprot.org/: January 2022). "# of var" represents a running average of all reported variants at each position of the gene, including all pathogenic and likely-pathogenetic variants reported in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar: January 2022). Calculation of the running average included reported variants at 10 amino acid residues preceding and following each position.

Supplementary Table 1. Neuroaxial phenotypes in AC probands		
Phenotype	**Clinical laboratory cohort (%)**	**Academic institution cohort (%)**
Developmental delay	76.8	48.8
Autism*	25.7	10.7
Macrocephaly	20.8	26.2
Hydrocephalus	7.4	26.2
Seizures	40.8	19.1
Deafness	15.3	9.5
Ophthalmological abnormalities	11.0	2.4
Chiari malformation	2.0	6.0
Cerebral palsy	7.8	2.4

Supplementary Figure 2.



Supplementary Figure 2. *De novo* variation (DNV) rate closely approximated Poisson distribution in AC cases and controls. The observed number of DNVs per subject (bars) compared to the numbers expected (lines) from the Poisson distribution in the case (red) and control (blue) cohorts. 'p' denotes chi-squared p-value.

Table 1. *De novo* variant enrichment of variant classes and relevent gene groupings

	De novo enrichment for 617 AC cases						*De novo* enrichment for 1798 control cases					
	Observed		Expected				Observed		Expected			
	n	Rate	n	Rate	Enrichment	P	n	Rate	n	Rate	Enrichment	P
All genes							**All genes**					
Total	789	1.28	683.6	1.11	1.15	**4.41E-05**	Total	1839	1.02	1977.1	1.11	0.93	9.99E-01
Synonymous	159	0.26	193.5	0.31	0.822	9.95E-01	Synonymous	492	0.27	559.8	0.31	0.879	9.98E-01
T-mis	334	0.54	339.1	0.55	0.98	6.16E-01	T-mis	949	0.53	979.3	0.55	0.97	8.37E-01
D-mis	175	0.28	91.6	0.15	1.91	**6.74E-15**	D-mis	248	0.14	266.7	0.15	0.93	8.81E-01
LoF	121	0.2	59.4	0.1	2.04	**1.60E-12**	LoF	150	0.08	171.3	0.1	0.876	9.54E-01
Protein-altering	630	1.02	490.1	0.79	1.29	**7.87E-10**	Protein-altering	1347	0.75	1417.3	0.79	0.95	9.71E-01
Damaging	296	0.48	151	0.24	1.96	**1.34E-25**	Damaging	398	0.22	438	0.24	0.909	9.75E-01
Genes with pLI > 0.9 (n = 3049)							**Genes with pLI > 0.9 (n = 3049)**					
Total	266	0.43	164	0.27	1.62	**1.72E-13**	Total	456	0.25	476.2	0.27	0.958	8.29E-01
Synonymous	35	0.06	46.2	0.07	0.758	9.62E-01	Synonymous	115	0.06	134.1	0.07	0.858	9.57E-01
T-mis	86	0.14	74.1	0.12	1.16	9.50E-02	T-mis	233	0.13	214.9	0.12	1.08	1.16E-01
D-mis	83	0.13	29.1	0.05	2.85	**2.87E-16**	D-mis	75	0.04	84.8	0.05	0.884	8.69E-01
LoF	62	0.1	14.6	0.02	4.23	**3.42E-20**	LoF	33	0.02	42.4	0.02	0.778	9.41E-01
Protein-altering	231	0.37	117.9	0.19	1.96	**2.24E-20**	Protein-altering	341	0.19	342.1	0.19	0.997	5.31E-01
Damaging	145	0.24	43.8	0.07	3.31	**1.57E-33**	Damaging	108	0.06	127.2	0.07	0.85	9.63E-01
HBE genes (n = 4491)							**HBE genes (n = 4491)**					
Total	258	0.42	176.3	0.29	1.46	**4.97E-09**	Total	515	0.29	511.6	0.29	1.01	4.47E-01
Synonymous	38	0.06	49.3	0.08	0.772	9.58E-01	Synonymous	128	0.07	143	0.08	0.895	9.04E-01
T-mis	98	0.16	84	0.14	1.17	7.30E-02	T-mis	274	0.15	243.6	0.14	1.12	2.90E-02
D-mis	73	0.12	27.1	0.04	2.69	**2.37E-13**	D-mis	74	0.04	78.9	0.04	0.938	7.23E-01
LoF	49	0.08	15.9	0.03	3.07	**2.45E-11**	LoF	39	0.02	46.2	0.03	0.845	8.72E-01
Protein-altering	220	0.36	127	0.21	1.73	**4.98E-14**	Protein-altering	387	0.22	368.6	0.21	1.05	1.76E-01
Damaging	122	0.2	43	0.07	2.83	**6.75E-23**	Damaging	113	0.06	125	0.07	0.904	8.69E-01
HBE genes with pLI > 0.9 (n = 1506)							**HBE genes with pLI > 0.9 (n = 1506)**					
Total	141	0.23	78.4	0.13	1.8	**1.33E-10**	Total	228	0.13	227.9	0.13	1	5.06E-01
Synonymous	15	0.02	21.8	0.04	0.689	9.47E-01	Synonymous	58	0.03	63.3	0.04	0.917	7.62E-01
T-mis	44	0.07	35.2	0.06	1.25	8.40E-02	T-mis	112	0.06	102.2	0.06	1.1	1.78E-01
D-mis	47	0.08	14.2	0.02	3.3	**5.84E-12**	D-mis	41	0.02	41.5	0.02	0.989	5.50E-01
LoF	35	0.06	7.2	0.01	4.84	**1.01E-13**	LoF	17	0.01	20.9	0.01	0.812	8.34E-01
Protein-altering	126	0.2	56.7	0.09	2.22	**1.53E-15**	Protein-altering	170	0.09	164.6	0.09	1.03	3.48E-01
Damaging	82	0.13	21.5	0.03	3.82	**2.13E-23**	Damaging	58	0.03	62.4	0.03	0.929	7.29E-01

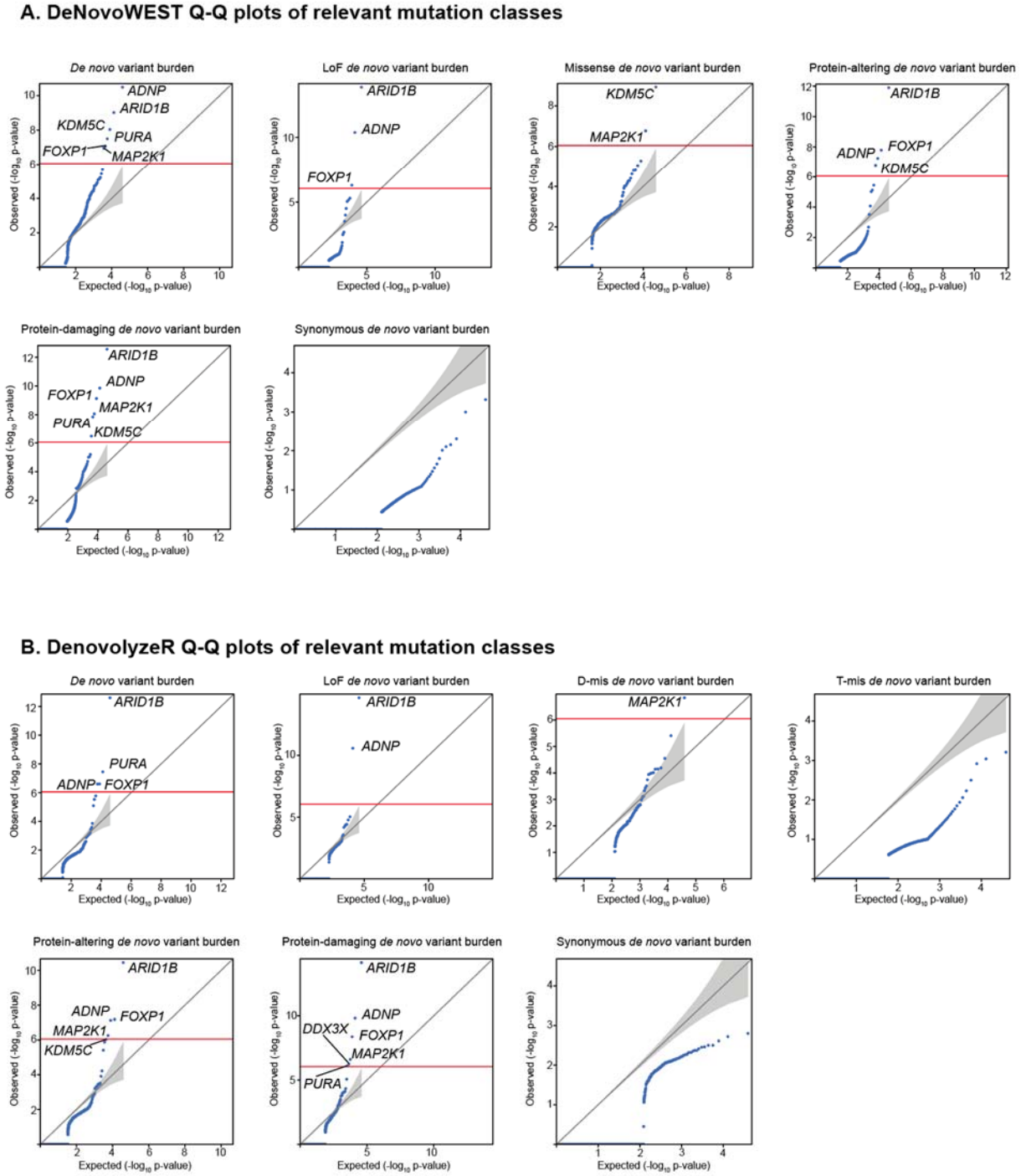
n: number of *de novo* variants (DNVs); Rate: number of DNVs per subject; Enrichment: ratio of observed to expected numbers of DNVs; D-mis: damaging missense variants as predicted by MetaSVM or MPC >2; T-mis: tolerated missense variants as predicted by MetaSVM or MPC <2; LoF: loss-of-function variants comprised of premature termination, frameshift, or splice-site variants; HBE genes: high brain-expressed genes denote genes in the top quartile of expression in the developing mouse brain.

Extended Data Table 1; DenovoWEST output of top 50 significant genes

symbol	hgnc_id	expected	observed	p-value	LoF (n)	Mis (n)	syn + inframe (n)	total DNVs (n)	mis_z	lof_z	pLI
ADNP	15766	0.0073129	2.1628284	3.23E-11	4	0	2	4	2.0722	5.6124	1
ARID1B	18040	0.0349378	3.3958834	9.64E-10	6	1	0	9	2.5854	8.4054	1
KDM5C	11114	0.0130372	2.1857556	9.22E-09	0	4	1	4	5.1452	6.3789	1
PURA	9701	0.0138569	2.5462248	3.17E-08	1	2	0	4	3.3704	2.7693	0.94029
FOXP1	3823	0.0088499	1.6325108	8.80E-08	2	2	0	4	2.2759	5.7478	0.99998
MAP2K1	6840	0.0100613	2.4154865	1.11E-07	0	3	0	3	3.1085	3.5883	0.89669
SCN2A	10588	0.0317986	2.1824072	2.10E-06	0	3	0	3	6.4584	7.9155	1
HPCA	5144	0.0026358	0.7614782	3.61E-06	0	1	0	1	2.2027	2.2914	0.6783
SATB2	21637	0.014011	1.6570378	6.03E-06	0	2	0	2	4.0511	5.3152	0.99999
FGF12	3668	0.0028641	0.7614782	7.27E-06	0	1	0	1	1.523	2.6313	0.62294
MFN2	16877	0.0136769	1.5698062	7.80E-06	0	2	0	2	1.6575	4.9033	0.99415
TRNAU1AP	30813	0.0028285	0.7355346	8.51E-06	0	1	0	1	1.6728	2.1317	0.000244
GPC4	4452	0.0029856	0.6998851	1.20E-05	0	1	0	1	1.6555	3.3877	0.92613
NKX2-1	11825	0.005563	0.8438614	1.36E-05	1	1	0	2	1.8053	2.6071	0.35874
TTPAL	16114	0.0033927	0.7047438	1.59E-05	0	1	0	1	1.4242	2.4404	0.116
GABRA5	4079	0.004959	0.7614782	1.71E-05	0	1	0	1	3.2308	3.6186	0.90535
PHKA2	8926	0.0069098	0.926398	1.84E-05	1	1	0	2	1.8249	5.0133	0.3839
DDX3X	2745	0.012459	1.6490744	1.89E-05	1	2	0	3	4.3295	4.6537	0.99986
TPM4	12013	0.0043043	0.7121939	1.94E-05	0	1	0	1	0.95062	3.0548	0.38852
GRAP2	4563	0.003876	0.6710668	2.76E-05	0	1	0	1	1.1236	2.3944	0.0001526
ATG4B	20790	0.0051782	0.6968524	3.62E-05	0	1	0	1	1.4141	3.8689	0.95648
IL4I1	19094	0.0062581	0.6869851	3.75E-05	0	1	0	1	1.7783	1.8584	0.0039601
NFKBIL1	7800	0.0048896	0.6978748	3.84E-05	0	1	0	1	2.2891	3.0507	0.025515
GIGYF1	9126	0.0118635	0.9529625	3.98E-05	1	1	0	2	-0.1491	4.9349	3.42E-06
ATP6V1A	851	0.0062301	0.6947691	4.24E-05	0	1	0	1	3.3731	4.3595	0.86801
DES	2770	0.0071113	0.7213408	4.53E-05	0	1	0	1	1.7729	3.0521	0.0088282
PTPN11	9644	0.0112827	1.2607814	4.83E-05	0	2	0	2	3.1293	5.3387	0.99998
KIF3B	6320	0.0069448	0.7594118	5.30E-05	0	2	0	2	2.9986	4.0581	0.084213
ARID1A	11110	0.0341655	1.6314278	5.64E-05	2	0	0	2	3.6582	8.5475	1
LMBRD2	25287	0.006544	0.7204218	6.21E-05	0	1	0	1	2.2723	3.8074	0.000164
PACS2	23794	0.0095624	0.7511225	6.46E-05	0	1	0	1	2.2358	5.4113	0.99583
SYT7	11514	0.0084953	0.6988855	6.89E-05	0	1	0	1	2.2108	3.6276	0.96342
CNOT3	7879	0.0199699	1.6212085	7.34E-05	0	2	0	2	3.7642	5.8127	1
BEND3	23040	0.0076358	0.6915357	7.53E-05	0	1	0	1	2.3139	4.0338	0.93555
KAT6A	13013	0.0158791	0.908303	0.0001028	0	2	0	2	2.0718	8.6737	1
TRPC7	20754	0.0097331	0.6647425	0.0001114	0	1	0	1	2.3053	3.8526	0.0001948
AP1G1	555	0.0092372	0.7614782	0.0001152	0	1	0	1	2.9839	6.0506	0.99998
EPB41L3	3380	0.0113523	0.7565325	0.0001183	0	1	0	1	1.4795	4.7822	4.32E-05
HIST1H2BD	4747	0.0014073	0.5544088	0.0001222	0	1	0	1	0.013773	-2.4292	4.21E-06
HIST1H1E	4718	0.0017599	0.6470143	0.0001307	1	0	0	1	-6.7151	0.91262	0.18608
NEDD4L	7728	0.0110577	0.7614782	0.0001385	0	1	0	1	3.7252	6.7479	1
EMC1	28957	0.0103168	0.6754329	0.0001485	0	1	1	1	1.3519	1.4928	6.73E-26
DGKI	2855	0.0112976	0.7047883	0.0001634	0	1	0	2	3.0849	6.0347	0.93134
GRIN2B	4586	0.0344769	1.6350267	0.0001654	0	2	0	2	5.4168	6.483	1
OPA1	8140	0.0108856	0.683359	0.0001911	0	1	0	1	1.9741	6.334	0.98636
BMP1	1067	0.0143094	0.7292523	0.0002184	0	1	0	1	1.9858	4.735	0.0008856
COL2A1	2200	0.0159324	0.7450983	0.0002603	1	1	0	2	3.2926	8.1212	1
TRIO	12303	0.0587955	1.4697011	0.0003051	0	2	1	2	5.3161	10.607	1
USP9X	12632	0.0245208	1.2633573	0.0003146	0	1	0	2	6.4105	8.8633	1
ATP1A3	801	0.0333469	1.5666886	0.0003171	0	2	0	2	6.3327	6.3973	1

Rows above the double line indicate those genes that surpass the exome-wide significance threshold. hgnc_id: HUGO Gene Nomenclature Committee identification number; n: number of *de novo* variants (DNVs); LoF: loss-of-function variants comprised of premature termination, frameshift, or splice-site variants. Mis: missense variants; syn: synonymous variants; inframe: in-frame variants; mis_z is a gene-wide constraint metric that estimates the probability of being intolerant to missense variants. lof_z is a transcript-wide constraint metric that estimates the probability of being intolerant to LoF variants; pLI is a gene-wide constraint metric that estimates the probability of being intolerant to LoF variants. pLI and mis-z values are based on gnomAD v2.1.1.

Supplementary Figure 3.



Supplementary Figure 3. (A) Quantile-quantile (Q-Q) plot comparing observed versus expected p-values (DeNovoWEST derived) for *de novo* variants (DNVs) in each gene in 617 AC cases. *ADNP*, *ARIDB1*, *KDM5C*, *PURA*, *FOXP1*, and *MAP2K1* exhibit exome-wide significant enrichment for all DNVs in AC cases. *ARID1B*, *ADNP*, and *FOXP1* exhibit significant enrichment of loss-of-function (LoF) DNVs comprising premature termination, frameshift, or splice-site variants. *KDM5C* and *MAP2K1* exhibit significant enrichment of missense variants. *ARID1B*, *FOXP1*, *ADNP*, and *KDM5C* exhibit significant enrichment of protein-altering variants, including missense and predictive LoF DNVs. *ARID1B*, *ADNP*, *FOXP1*, *MAP2K1*, *PURA*, and *KDM5C* exhibit significant enrichment of protein-damaging variants, including D-mis and LoF DNVs. There is no significant enrichment of synonymous DNVs among the 617 cases. Grey areas within graphs represents 95% confidence interval for expected values. (**B) Q-Q plot comparing observed versus expected p-values (DenovolyzeR derived) for DNVs in each gene in 617 AC cases**. *ARID1B*, *PURA*, *ADNP*, and *FOXP1* exhibit exome-wide significant enrichment for all DNVs in AC cases. *ARID1B* and *ADNP* exhibit significant enrichment of LoF DNVs. *MAP2K1* exhibits significant enrichment of damaging-missense (D-mis) variants (MetaSVM= 'D' or MPC>2 damaging missense). *ARID1B*, *ADNP*, *FOXP1*, *MAP2K1*, and *KDM5C* exhibit significant enrichment of protein-altering variants. *ARID1B*, *ADNP*, *FOXP1*, *MAP2K1*, and *DDX3X* exhibit significant enrichment of protein-damaging variants. There is no significant enrichment of tolerated-missense (T-mis) DNVs or synonymous DNVs among the 617 cases. The grey areas within graphs represents 95% confidence intervals for expected values.

Extended Data Table 2. DenovolyzeR output of top 50 significant genes

symbol	LoF observed	LoF expected	LoF p-value	protein altering observed	protein altering expected	protein altering p-value	protein damaging observed	protein damaging expected	protein damaging p-value
ARID1B	6	0	2.14E-15	7	0.1	3.51E-11	6	0	7.12E-15
FOXP1	2	0	1.70E-05	4	0	6.68E-08	4	0	4.29E-09
ADNP	4	0	2.61E-11	4	0	7.66E-08	4	0	1.51E-10
KDM5C	0	0	1	4	0.1	8.27E-07	3	0	8.58E-06
PURA	1	0	0.00148	3	0	1.34E-06	3	0	7.78E-07
MAP2K1	0	0	1	3	0	5.61E-07	3	0	2.46E-07
DDX3X	1	0	0.00357	3	0	3.88E-06	3	0	5.62E-07
SCN2A	0	0	1	3	0.1	6.02E-05	3	0.1	4.60E-05
PTPN11	0	0	1	2	0	0.000303	2	0	0.000163
PTEN	1	0	0.0032	2	0	0.000355	2	0	8.90E-05
NKX2-1	1	0	0.00216	2	0	0.000357	2	0	9.85E-05
CUL3	2	0	9.30E-06	2	0	0.000448	2	0	0.000178
SATB2	0	0	1	2	0	0.000497	2	0	0.000154
FZD10	0	0	1	2	0	0.000585	2	0	9.56E-05
CNOT3	0	0	1	2	0	0.00059	2	0	0.000171
MFN2	0	0	1	2	0	0.000611	2	0	0.000425
ATP1A3	0	0	1	2	0	0.0012	2	0	0.000919
GIGYF1	1	0	0.0077	2	0.1	0.00139	2	0	0.000316
PHKA2	1	0	0.00685	2	0.1	0.00139	2	0	0.000825
CLTC	2	0	3.76E-05	2	0.1	0.00178	2	0	0.000834
GRIN2B	0	0	1	2	0.1	0.00192	2	0	0.00054
NALCN	0	0	1	2	0.1	0.00259	2	0.1	0.0024
KAT6A	0	0	1	2	0.1	0.00304	2	0	0.00047
SON	2	0	5.59E-05	2	0.1	0.00426	2	0	8.39E-05
RP1L1	2	0	4.38E-05	2	0.1	0.00431	2	0	0.000108
NSD1	0	0	1	2	0.1	0.00437	2	0.1	0.00173
PRPF8	0	0	1	2	0.1	0.00504	2	0.1	0.00315
ARID1A	2	0	6.92E-05	2	0.1	0.00571	2	0	7.63E-05
KIF26B	0	0	1	2	0.1	0.0063	2	0.1	0.00127
FEZ2	1	0	0.00179	2	0	0.000124	1	0	0.00185
KIF3B	0	0	1	2	0	0.000461	1	0	0.016
COL2A1	1	0	0.00898	2	0.1	0.00212	1	0.1	0.0538
USP9X	0	0	1	2	0.1	0.004	1	0.1	0.0507
ZFHX4	1	0	0.0137	2	0.1	0.00782	1	0	0.0259
FAM92A	1	0	0.000352	1	0	0.000352	1	0	0.000352
CBX5	0	0	1	1	0	0.00614	1	0	0.0042
HIST1H2BD	0	0	1	1	0	0.00671	1	0	0.000744
RALA	0	0	1	1	0	0.00738	1	0	0.00545
ATP6V1E1	0	0	1	1	0	0.00818	1	0	0.00288
FTL	1	0	0.000576	1	0	0.0083	1	0	0.00175
H3F3A	0	0	1	1	0	0.00862	1	0	0.00729
RAC1	0	0	1	1	0	0.00914	1	0	0.00662
TLDC2	1	0	0.00121	1	0	0.0094	1	0	0.00176
OR4L1	1	0	0.000978	1	0	0.00978	1	0	0.00119
GJB2	0	0	1	1	0	0.01	1	0	0.00897
NAA10	0	0	1	1	0	0.0101	1	0	0.00607
ARL4D	1	0	0.00115	1	0	0.0102	1	0	0.00598
HIST1H1E	1	0	0.000275	1	0	0.0106	1	0	0.000681
OR10K1	1	0	0.000899	1	0	0.0107	1	0	0.00121
TMBIM4	1	0	0.00138	1	0	0.0108	1	0	0.0021

Rows above the double line indicate those genes that surpass exome-wide significance threshold. Protein altering *de novo* variants (DNVs) include missense and predictive loss-of-function (LoF). Protein damaging DNVs include MetaSVM= ‘D’ or MPC>2 damaging missense and predictive LoF.

Table 2. Protein-damaging and protein-altering variant information for exome-wide significant genes

gene	chrom	position	reference nucleotide	alternate nucleotide	cDNA change	amino acid change	MetaSVM consequence	MPC SCORE	Phenotypes associated with literature reported variants
ADNP	20	49509021	C	A	c.2230G>T	p.E744*	nonsense	45	.
		49509094	G	T	c.2157C>A	p.Y719*	nonsense	38	Intellectual disability, Helsmoortel - van der aa syndrome (OMIM # 615873)
		49509094	G	C	c.2157C>G	p.Y719*	nonsense	38	Intellectual disability, Helsmoortel - van der aa syndrome (OMIM # 615873)
		49520476	TCA	T	c.56_57del	p.K20fs	frameshift	.	.
ARID1B	6	157505384	G	A	c.3485G>A	p.W1162*	nonsense	39	.
		157502190	C	T	c.3343C>T	p.R1115*	nonsense	36	Coffin-Siris syndrome 1 (OMIM # 135900)
		157528399	G	A	c.6244G>A	p.E2082K	T-mis	22.1	.
		157100423	C	T	c.1360C>T	p.Q454*	nonsense	19.1	.
		157510820	GGC	G	c.3716_3717del	p.G1239fs	frameshift	.	.
		157525007	TG	T	c.5023delG	p.E1675fs	frameshift	.	.
		157525028	GTCCCTTA	G	c.5044_5050del	p.S1682fs	frameshift	.	.
DDX3X	X	41196694	.	A	c.80dupA	p.Q27fs	frameshift	.	.
		41203604	G	T	c.929G>T	p.R310L	D-mis	35	Mental retardation, X-linked 102 (OMIM # 300958)
		41205659	C	T	c.1445C>T	p.T482I	D-mis	26.1	Mental retardation, X-linked 102 (OMIM # 300958)
FOXP1	3	71096126	G	A	c.C631T	p.Q211*	nonsense	37	.
		71021817	C	T	c.1238G>A	p.R513H	D-mis	23.4	Mental retardation with language impairment w/wo autistic features (OMIM # 613670)
		71026813	T	C	c.1406A>G	p.Y469C	D-mis	17.5	.
		71026984	G	GTTT	c.1342_1343insAAA	p.S448*T	nonsense	.	.
KDM5C	X	53228288	C	T	c.1913G>A	p.R638H	D-mis	26.8	Mental retardation, syndromic, Claes Jensen type, X-linked (OMIM # 300534)
		53239754	C	T	c.1387G>A	p.E463K	D-mis	23.7	.
		53228264	G	A	c.1937C>T	p.T646M	D-mis	22.6	X-linked intellectual disability
		53223382	T	C	c.3776A>G	p.N1259S	T-mis	0	.
MAP2K1	15	66729181	A	G	c.389A>G	p.Y130C	D-mis	24.9	Cardio-facio-cutaneous syndrome 3 (OMIM # 615279)
		66729181	A	G	c.389A>G	p.Y130C	D-mis	24.9	Cardio-facio-cutaneous syndrome 3 (OMIM # 615279)
		66727447	A	C	c.163A>C	p.T55P	D-mis	21.6	.
PURA	5	139494329	T	C	c.563T>C	p.I188T	D-mis	29.1	Mental retardation, Autosomal Dominant 31 (OMIM # 616158)
		139493984	T	C	c.218T>C	p.F73S	D-mis	23.8	.
		139493802	TGCGGCGCT GGGTTCGG	T	c.37_52del	p.A13fs	frameshift	.	.

Chrom: chromosome. Phenotypes associated with literature-reported variants in genes include those with high-quality evidence in OMIM (Online Mendelian Inheritance in Man).

Supplementary Table 2. Roles and phenotype of GWS genes

Gene	Role in regulation of gene expression	General role	Associated phenotype	Phenotypes associated with gene variant in cohort	Citations
ADNP	Transcription factor - Forms inaccessible chromatin around its DNA binding site	Neurogenesis/organogenesis & repression of endoderm genes during specific stage in embryologic development. (PMID: 17222401)	Helsmoortel-van der AA syndrome (OMIM # 615873)	Global Developmental Delay Nystagmus Generalized Hypotonia Macrocepahly	PMID: 17363064 PMID: 17878164 PMID: 24531329
ARID1B	Chromatin modifier - Direct binding and nucleosome disruption	Widely expressed chromatin modifier and important contributor to neurodevelopment. (PMID: 22405089)	Coffin-Siris syndrome (OMIM # 135900)	Global Developmental Delay Motor Delay Delayed Speech and Language Development Generalized Hypotonia Cryptorchidism Seizure	PMID: 11988099 PMID: 12665591 PMID: 23906836
DDX3X	Transcriptional activator - Bind to and upregulate promoter activity	Multifunctional ATP-dependent RNA helicase essential in the RNAi pathway and key regulator of the WNT pathway. (PMID: 26235985)	Intellectual developmental disorder, X-linked, syndrome, Snijders Blok type (OMIM # 300958)	Global Developmental Delay Generalized Hypotonia	PMID: 26235985 PMID: 23413191
FOXP1	Transcription repressor - Binding to consensus site in promoter region	Critical role in monocyte differentiation and macrophage function. (PMID: 18799727)	Mental retardation with language impairment with or without autistic features (OMIM # 613670)	Delayed Speech and Language Development Global Developmental Delay Cryptorchidism Macrocephaly Generalized Hypotonia	PMID: 18799727 PMID: 12692134 PMID: 20950788
KDM5C	Chromatin modifier - Histone demethylation	Important role in normal brain function and neuronal gene regulation. (PMID: 15586325)	Mental retardation, X-linked, syndromic, Claes-Jensen type (OMIM # 300534)	Global Developmental Delay Generalized Hypotonia Macrocephaly Autistic Behavior Delayed Speech and Language Development Frontal Bossing	PMID: 17468742 PMID: 10982473
MAP2K1	Transcription repressor - Binding to and inhibiting gene transcriptional complex	Critical activator of MAP kinase involved in various biochemical signaling pathways for growth and DNA synthesis. (PMID: 1411546)	Cardiofaciocutaneous syndrome 3 (OMIM # 615279)	Delayed Speech and Language Development Global Developmental Delay Generalized Hypotonia Failure to Thrive Growth Delay Cerebral Palsy Seizure	PMID: 11545732 PMID: 18042262 PMID: 15489854
PURA	Transcription activator - Direct binding of initiation site	Essential for normal brain development, synapse formation and proliferation of neurons, oligodendrocytes, and astrocytes in the CNS. (PMID: 27148565)	Mental retardation, autosomal dominant 31 (OMIM # 616158)	Generalized Hypotonia Global Developmental Delay Macrocephaly Constipation	PMID: 25342064 PMID: 7862639 PMID: 27148565 PMID: 25342064

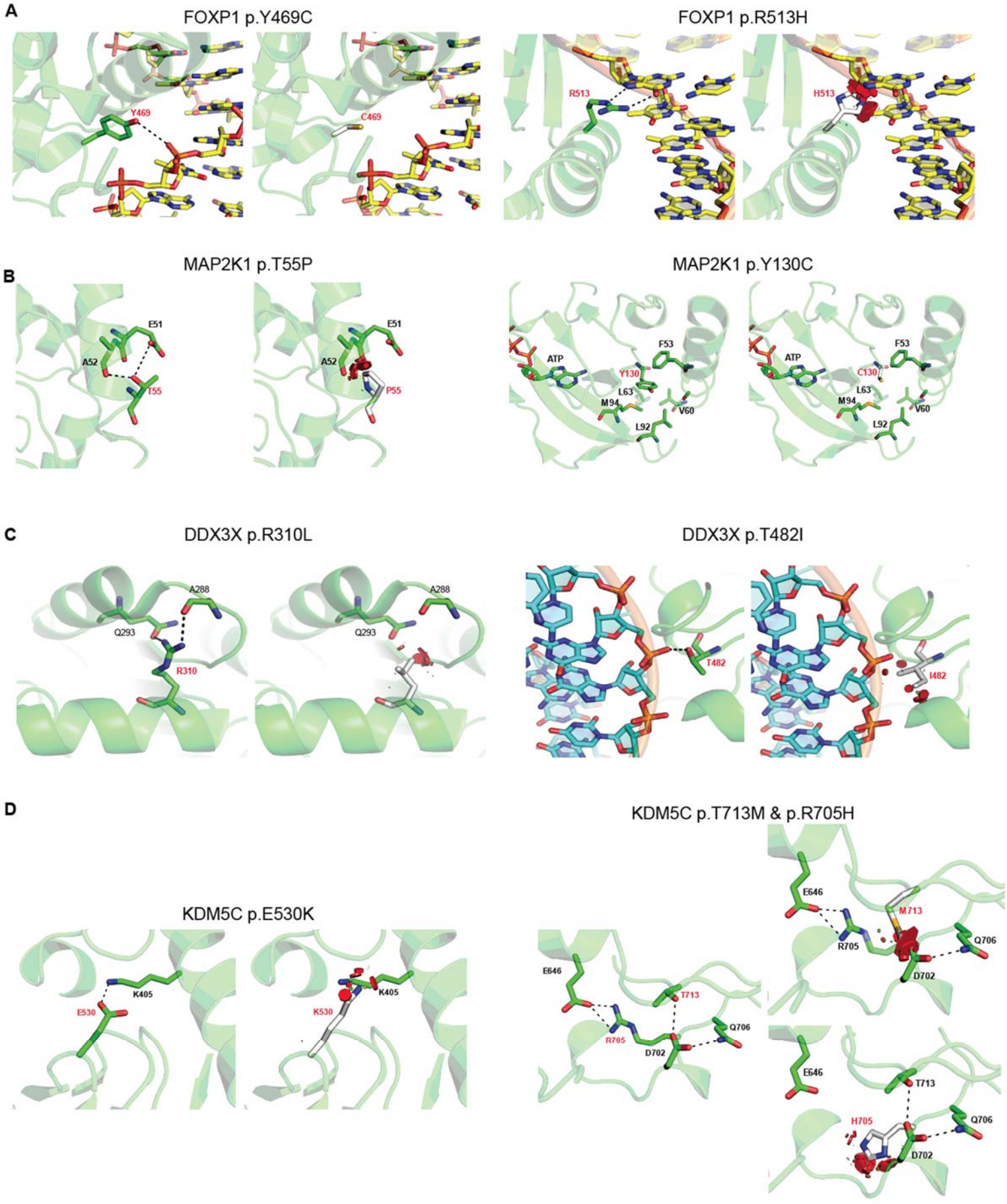
Associated phenotypes represent those that are linked with autosomal dominant or X-linked inheritance of gene variants within the Online Mendelian Inheritance in Man (OMIM) database (https://www.omim.org/). Phenotypes associated with gene variants in the cohort are those phenotypes identified with text2hpo (https://github.com/GeneDx/txt2hpo) that are present in AC probands with specified gene variants.

Table 3. *De novo* variant enrichment of OMIM disease genes

	De novo enrichment for 617 AC cases						*De novo* enrichment for 1798 control cases					
	Observed		Expected				Observed		Expected			
	n	Rate	n	Rate	Enrichment	P	n	Rate	n	Rate	Enrichment	P
Genes with associated dominant OMIM phenotypes (n = 3351)							**Genes with associated dominant OMIM phenotypes (n = 3351)**					
Total	320	0.52	157	0.25	2.04	**2.98E-30**	Total	457	0.25	456.2	0.26	1	4.91E-01
Synonymous	55	0.09	44.3	0.07	1.24	6.65E-02	Synonymous	128	0.07	128.8	0.07	0.994	5.38E-01
T-mis	90	0.15	67	0.11	1.34	4.00E-03	T-mis	200	0.11	194.5	0.11	1.03	3.56E-01
D-mis	105	0.17	31.9	0.05	3.29	**1.50E-24**	D-mis	90	0.05	92.9	0.05	0.968	6.33E-01
LoF	70	0.11	13.8	0.02	5.06	**7.27E-27**	LoF	39	0.02	40.1	0.02	0.973	5.90E-01
Protein-altering	265	0.43	112.7	0.18	2.35	**2.38E-34**	Protein-altering	329	0.18	327.5	0.18	1	4.73E-01
Damaging	175	0.28	45.7	0.07	3.83	**5.81E-48**	Damaging	129	0.07	133	0.07	0.97	6.48E-01
OMIM genes with pLI > 0.9 (n = 792)							**OMIM genes with pLI > 0.9 (n = 792)**					
Total	165	0.27	49.4	0.08	3.34	**2.81E-38**	Total	151	0.08	143.6	0.08	1.05	0.281
Synonymous	16	0.03	14	0.02	1.15	3.27E-01	Synonymous	39	0.02	40.6	0.02	0.961	6.20E-01
T-mis	34	0.06	18.9	0.03	1.8	1.00E-03	T-mis	70	0.04	54.9	0.03	1.28	2.80E-02
D-mis	65	0.11	12.2	0.02	5.32	**3.26E-26**	D-mis	32	0.02	35.6	0.02	0.899	7.48E-01
LoF	50	0.08	4.3	0.01	11.5	**3.25E-35**	LoF	10	0.01	12.6	0.01	0.796	8.04E-01
Protein-altering	149	0.24	35.5	0.06	4.2	**1.11E-45**	Protein-altering	112	0.06	103.1	0.06	1.09	2.01E-01
Damaging	115	0.19	16.5	0.03	6.95	**3.68E-56**	Damaging	42	0.02	48.2	0.03	0.872	8.31E-01

n: number of *de novo* variants (DNVs); Rate: number of DNVs per subject; Enrichment: ratio of observed to expected numbers of DNVs; D-mis: damaging missense DNVs as predicted by MetaSVM or MPC >2; T-mis: tolerated missense DNVs as predicted by MetaSVM or MPC >2; LoF: Loss-of-function DNVs comprised of premature termination, frameshift or splice site variants; Genes with associated dominant OMIM phenotype (OMIM genes) denotes genes report to be associated with an OMIM disorder with autosomal dominant or X-linked inheritance pattern.

Supplementary Figure 4.



Supplementary Figure 4. Structural alterations of *de novo* damaging missense (D-mis) variants in exome-wide significant AC genes. (A) FOXP1 (Forkhead box protein 1) is a transcriptional repressor directly interacting with DNA. The available NMR structure of FOXP1 (PDB id 2KIU) does not include bound DNA. Therefore, to study FOXP1 interactions with DNA, a model of the complex (DNA + FOXP1) was generated based on FOXP2 (PDB id 2AS5; 87% sequence identity). Residues Y469 and R513 are highly conserved in FOXP1, FOXP2, and FOXP3. The hydroxyl group in the Y469 side chain makes a hydrogen bond with the sugar-phosphate backbone of DNA (yellow). Substitution of Cys abolishes this bond. The $\Delta\Delta G_{stability}$ is 3.22. The side chain of R513 makes hydrogen bonds with the guanine base. The shorter side chain of H513 misaligns to make steric clashes with the DNA backbone. The red discs indicate significant pairwise overlap of van der Waals radii and highlight clashes between atoms. The $\Delta\Delta G_{stability}$ is 0.95. **(B)** MAP2K1 (mitogen-activated protein kinase 1) is a dual-specificity protein kinase and plays a central role in the MAP kinase signal transduction pathway (PDB id 6U2G). T55 is positioned on helix 1. The hydroxyl group in the side chain makes hydrogen bonds with the side chain of Glu51 and the backbone oxygen of A52. These hydrogen bonds help to further stabilize the helix in addition to the intra-helical bonds. Proline residues are considered helix-breakers due to their inability to contribute to intra-helical interactions. The cyclic side chain of proline also makes steric clashes with the helical structure. Thus, a T55P variant is energetically unfavorable at this position ($\Delta\Delta G_{stability}$ = 0.75). Residue Y130 is sandwiched between helix 1 and ®-strand 5. The Y130 side chain is involved in packing interactions and surrounded by hydrophobic residues F53, V60, L63, L92, and M94. These residues also lie adjacent to the ATP binding site. A Cys substitution results in loss of the packing interactions, destabilizing the ATP binding site and surrounding region. The $\Delta\Delta G_{stability}$ is 4.58. **(C)** DDX3X (dead box; PDB id 6O5F) is a multifunctional ATP-dependent RNA helicase that regulates RNA biogenesis by unwinding short RNA duplexes in three stages. In the first step, the DDX binds to short double-stranded RNA to form a complex. In the second step, ATP binding to the DDX: dsRNA complex results in dsRNA unwinding. In the final step, ATP hydrolysis results in the release of the unwound ssRNA. The guanidinium side chain of R310 makes hydrogen bonds with the backbone carbonyl oxygen of A288 and with the side chain of Q293 (α-helix 8). A Leu substitution results in the loss of the two hydrogen bonds. The aliphatic side chain of Leu also makes steric clashes with the adjacent loop, which would result in the destabilizing secondary structural elements. The $\Delta\Delta G_{stability}$ of R310L is 1.51. T482 forms a part of the Helicase C-terminal domain. The hydroxyl side chain of T482 forms a direct hydrogen bond with the phosphate-oxygen in the phosphodiester backbone of the dsRNA. Ile substitution at this position results in loss of this hydrogen bond. Additionally, the hydrophobic branched aliphatic side chain in Ile makes steric clashes with the RNA backbone and with residues in the adjacent α-helix 15, resulting in regional destabilization. The $\Delta\Delta G_{stability}$ of T482I is 1.06. **(D)** KDM5C is a zinc finger lysine-specific demethylase 5C that specifically demethylates 'Lys-4' of histone H3. It participates in transcriptional repression of neuronal genes by recruiting histone deacetylases and REST at neuron-restrictive silencer elements. E530 forms an ion pair interaction with K405. The variant E530K would result in electrostatic repulsion between two positively charged lysyl side chains, abolishing the attractive ion pair interaction. The $\Delta\Delta G_{stability}$ is 0.58. The side chains of R705 and T713 are close to one another. R705 makes ion-pair interactions with E664, while the T side-chain hydroxyl hydrogen bonds with D702 and Q706. The $\Delta\Delta G_{stability}$ is 0.58. A variant to H705 would result in steric clashes with the backbone. For the M713 variant, the longer side chain of T713 cannot fit into the loop structure and makes steric clashes with the side chain of D702. The $\Delta\Delta G_{stability}$ is 0.26.

Supplementary Figure 5.

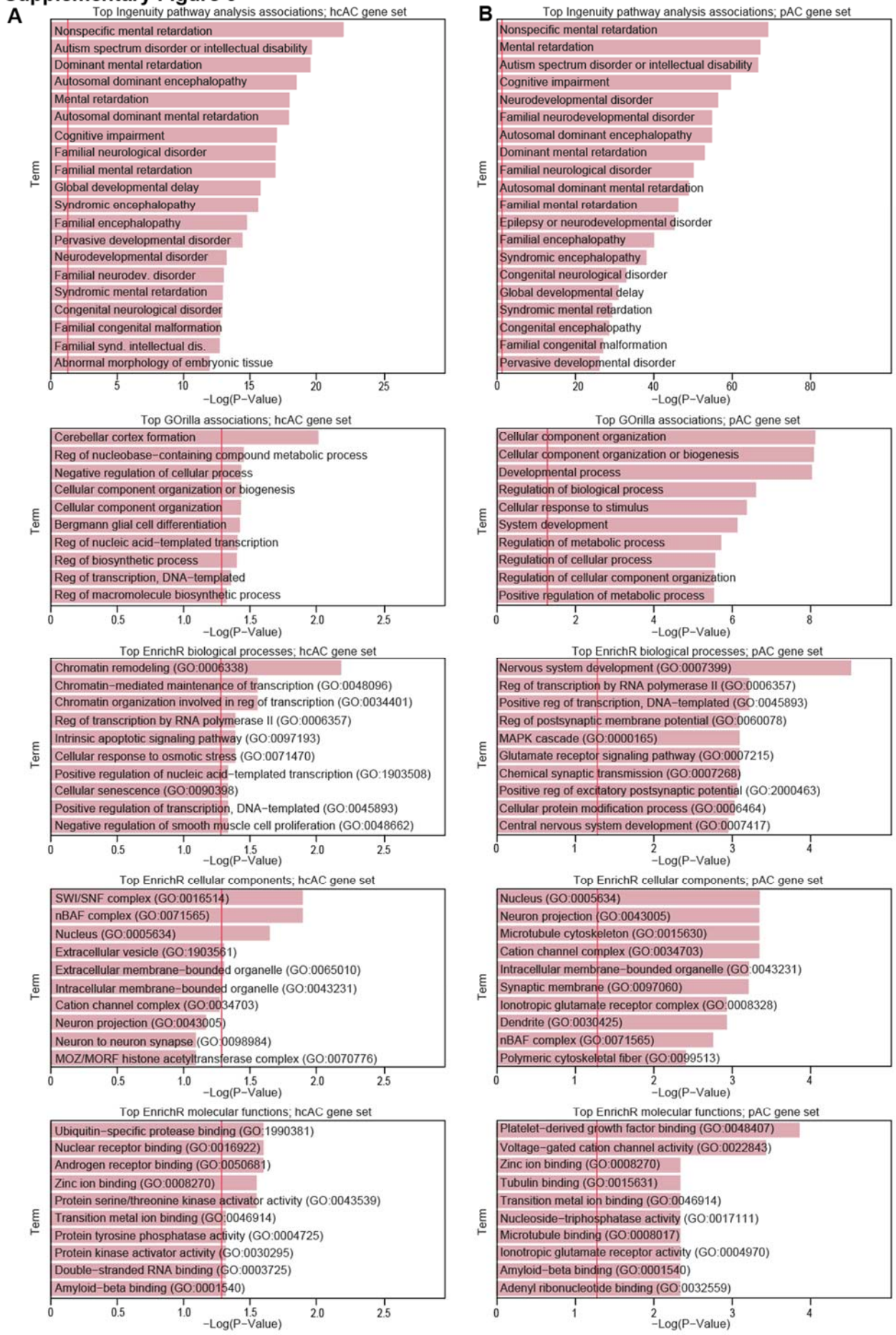
Supplementary Figure 5. Phenomic heat map of traits identified in AC patients harboring *de novo* variants (DNVs) in exome-wide significant AC risk genes. Subject phenotypes were determined by text2HPO natural language processing of medical record data (https://github.com/GeneDx/txt2hpo).

Supplementary Table 3; Candidate AC risk genes

Exome-wide significant (EWS) AC genes			
ANDP	*DDX3X*	*KDM5C*	*PURA*
ARID1B	*FOXP1*	*MAP2K1*	
High-confidence AC (hcAC) genes			
ADNP	*CUL3*	*KIF26B*	*PTPN11*
ARID1A	*DDX3X*	*MAP2K1*	*PURA*
ARID1B	*FOXP1*	*MFN2*	*SATB2*
ATP1A3	*GIGYF1*	*NSD1*	*SCN2A*
CLTC	*GRIN2B*	*PHKA2*	*SON*
CNOT3	*KAT6A*	*PRPF8*	*USP9X*
COL2A1	*KDM5C*	*PTEN*	*ZFHX4*
Possible AC (pAC) genes			
ACLY	*DPF2*	*MAGEL2*	*PURA*
ADNP	*DYNC1H1*	*MAP2K1*	*RALA*
AKT1	*EHBP1*	*MAP2K4*	*RASA1*
ANKRD11	*EZR*	*MED13L*	*REV3L*
ANKRD12	*FBXL19*	*MFN2*	*RIMS2*
AP1G1	*FBXO11*	*MTF1*	*SATB2*
APC	*FOXP1*	*MTOR*	*SCN2A*
APC2	*FURIN*	*MTPAP*	*SCN8A*
ARID1A	*GABRA5*	*MYH14*	*SETBP1*
ARID1B	*GABRG2*	*MYO18A*	*SETD5*
ARID2	*GATAD2B*	*MYT1L*	*SLC2A1*
ATG4B	*GIGYF1*	*NAV1*	*SON*
ATP1A3	*GNAQ*	*NF1*	*SOS1*
ATP6V1A	*GNB1*	*NFIX*	*SOX5*
BAZ1B	*GPC4*	*NR5A1*	*STXBP1*
BMPR1B	*GRIA2*	*NSD1*	*SYT7*
CACNA1A	*GRIN1*	*NUP155*	*TAF1*
CACNA1C	*GRIN2B*	*OPA1*	*TAOK1*
CAMSAP2	*HDAC8*	*PDE2A*	*TCF4*
CASK	*IQSEC2*	*PDGFRB*	*TLK2*
CFLAR	*IRF2BPL*	*PHKA2*	*TRABD*
CLTC	*KAT6A*	*PKD2*	*TRIM8*
CNOT3	*KCNB1*	*POGZ*	*TRIP12*
COL12A1	*KCNQ2*	*POLR2A*	*UBE3A*
COL2A1	*KDM5C*	*POLR2B*	*ULK1*
COL3A1	*KIF1A*	*PPP1CB*	*USP9X*
COL4A1	*KIF26B*	*PPP2CA*	*WBP11*
CTCF	*KIF5A*	*PPP2R5D*	*WDR26*
CUL3	*KMT2A*	*PRPF8*	*WDR45*
DDX3X	*KMT2D*	*PTEN*	*XPO7*
DLG4	*LRRTM2*	*PTPN11*	*ZFHX4*

Exome-wide significant (EWS) genes include those genes surpassing multiple testing corrected significance thresholds for enrichment in protein-altering or protein-damaging DNVs. High-confidence AC (hcAC) genes include EWS genes plus those harboring two or more protein-altering DNVs with at least one damaging variant in predictive loss-of-function intolerant (pLI) genes (pLI score ≥ 0.9 in ExAC). Possible AC genes included EWS genes, hcAC genes, and those genes that harbored ≥ one damaging DNV with a pLI ≥ 0.9.

Supplementary Figure 6



Supplementary Figure 6. Ingenuity pathway, EnrichR, and GOrilla enrichment analysis for high-confidence (hcAC) and probable (pAC) AC gene sets. -Log_{10}(p-values) for hcAC (left column) and pAC (right column) gene sets are shown. Significance thresholds, adjusted for multiple comparisons, are denoted by a vertical red line. Top enriched terms are visualized for readability. Reg: regulation

Figure 2.

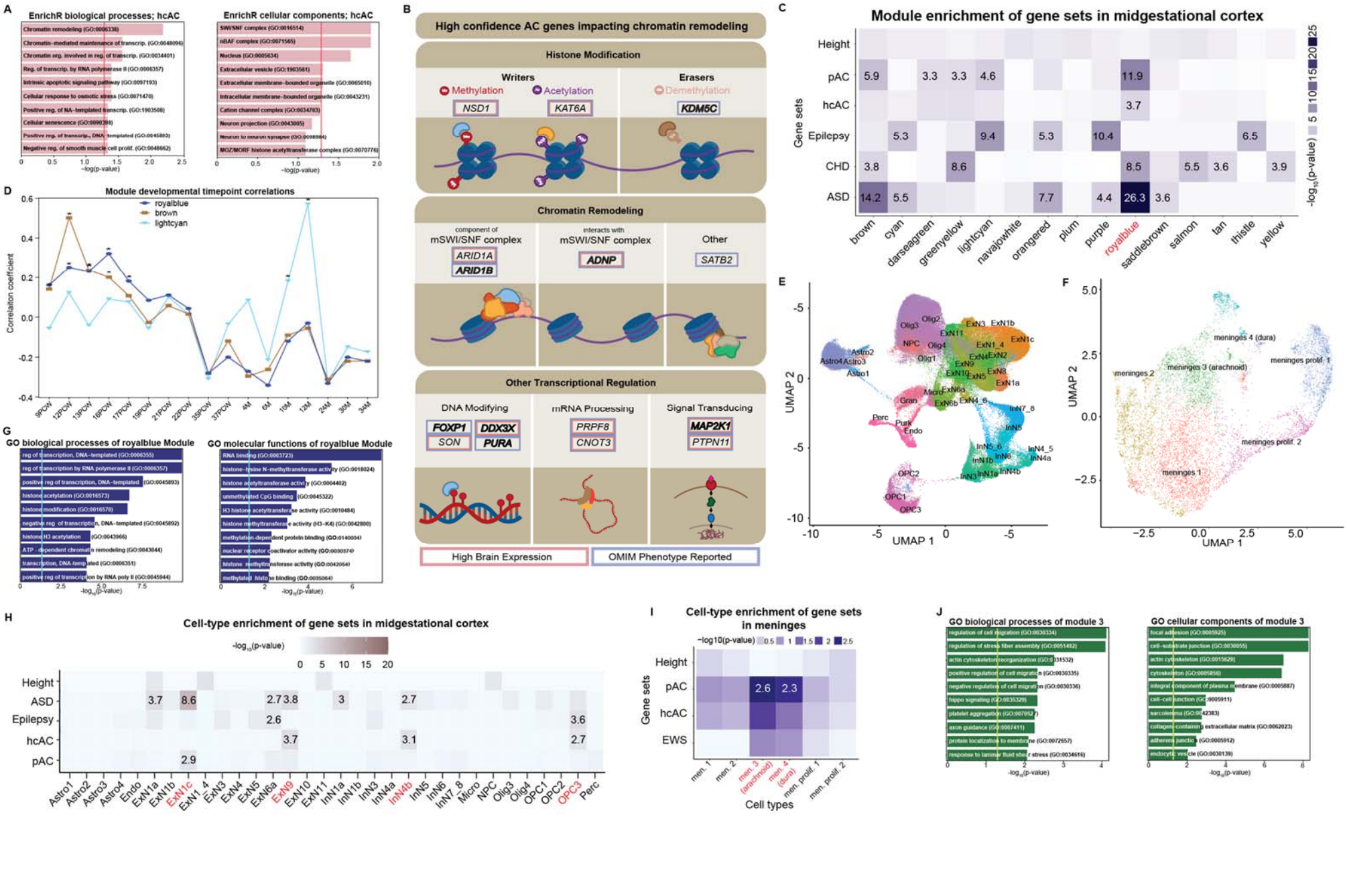

A
EnrichR biological processes; hcAC
Chromatin remodeling (GO:0006338)
Chromatin−mediated maintenance of transcrip. (GO:0048096)
Chromatin org. involved in reg. of transcrip. (GO:0034401)
Reg. of transcrip. by RNA polymerase II (GO:0006357)
Intrinsic apoptotic signaling pathway (GO:0097193)
Cellular response to osmotic stress (GO:0071470)
Positive reg. of NA−templated transcrip. (GO:1903508)
Cellular senescence (GO:0090398)
Positive reg. of transcrip., DNA−templated (GO:0045893)
Negative reg. of smooth muscle cell prolif. (GO:0048662)
−log(p-value)
EnrichR cellular components; hcAC
SWI/SNF complex (GO:0016514)
nBAF complex (GO:0071565)
Nucleus (GO:0005634)
Extracellular vesicle (GO:1903561)
Extracellular membrane−bounded organelle (GO:0065010)
Intracellular membrane−bounded organelle (GO:0043231)
Cation channel complex (GO:0034703)
Neuron projection (GO:0043005)
Neuron to neuron synapse (GO:0098984)
MOZ/MORF histone acetyltransferase complex (GO:0070776)
−log(p-value)
B
High confidence AC genes impacting chromatin remodeling
Histone Modification
Writers
Methylation
Acetylation
NSD1
KAT6A
Erasers
Demethylation
KDM5C
Chromatin Remodeling
component of
mSWI/SNF complex
ARID1A
ARID1B
interacts with
mSWI/SNF complex
ADNP
Other
SATB2
Other Transcriptional Regulation
DNA Modifying
FOXP1
DDX3X
SON
PURA
mRNA Processing
PRPF8
CNOT3
Signal Transducing
MAP2K1
PTPN11
High Brain Expression
OMIM Phenotype Reported
C
Module enrichment of gene sets in midgestational cortex
Gene sets
Height
pAC
hcAC
Epilepsy
CHD
ASD
5.9 3.3 3.3 4.6 11.9
3.7
5.3 9.4 5.3 10.4 6.5
3.8 8.6 8.5 5.5 3.6 3.9
14.2 5.5 7.7 4.4 26.3 3.6
brown cyan darseagreen greenyellow lightcyan navajowhite orangered plum purple royalblue saddlebrown salmon tan thistle yellow
−log10(p-value) 5 10 15 20 25
D
Module developmental timepoint correlations
royalblue
brown
lightcyan
Correlaiton coefficient
0.6 0.4 0.2 0.0 −0.2 −0.4
8PCW 12PCW 13PCW 16PCW 17PCW 19PCW 21PCW 22PCW 35PCW 37PCW 4M 6M 10M 12M 24M 30M 34M
E
UMAP 1
UMAP 2
Astro1 Astro2 Astro3 Astro4 Endo NPC Olig1 Olig2 Olig3 Olig4 Micro Gran Perc Purk Endo OPC1 OPC2 OPC3 ExN1_4 ExN1a ExN1b ExN1c ExN2 ExN3 ExN4 ExN5 ExN6a ExN6b ExN8 ExN9 ExN10 ExN11 InN1a InN1b InN3 InN4a InN4b InN5 InN5_6 InN6 InN7_8
F
UMAP 1
UMAP 2
meninges 1
meninges 2
meninges 3 (arachnoid)
meninges 4 (dura)
meninges prolif. 1
meninges prolif. 2
G
GO biological processes of royalblue Module
reg of transcription, DNA−templated (GO:0006355)
reg of transcription by RNA polymerase II (GO:0006357)
positive reg of transcription, DNA−templated (GO:0045893)
histone acetylation (GO:0016573)
histone modification (GO:0016570)
negative reg. of transcription, DNA−templated (GO:0045892)
histone H3 acetylation (GO:0043966)
ATP−dependent chromatin remodeling (GO:0043044)
transcription, DNA−templated (GO:0006351)
positive reg of transcription by RNA poly II (GO:0045944)
−log10(p-value)
GO molecular functions of royalblue Module
RNA binding (GO:0003723)
histone−lysine N−methyltransferase activity (GO:0018024)
histone acetyltransferase activity (GO:0004402)
unmethylated CpG binding (GO:0045322)
H3 histone acetyltransferase activity (GO:0010484)
histone methyltransferase activity (H3−K4) (GO:0042800)
methylation−dependent protein binding (GO:0140034)
nuclear receptor coactivator activity (GO:0030374)
histone−methyltransferase activity (GO:0042054)
methylated histone binding (GO:0035064)
−log10(p-value)
H
Cell-type enrichment of gene sets in midgestational cortex
−log10(p-value)
0 5 10 15 20
Height
ASD
Epilepsy
hcAC
pAC
3.7 8.6 2.7 3.8 3 2.7
2.6 3.6
3.7 3.1 2.7
2.9
Astro1 Astro2 Astro3 Astro4 Endo ExN1a ExN1b ExN1c ExN1_4 ExN3 ExN4 ExN5 ExN6a ExN8 ExN10 ExN11 InN1a InN1b InN3 InN4a InN4b InN5 InN6 InN7_8 Micro NPC Olig3 Olig4 OPC1 OPC2 OPC3 Perc
I
Cell-type enrichment of gene sets in meninges
−log10(p-value) 0.5 1 1.5 2 2.5
Gene sets
Height
pAC
hcAC
EWS
2.6 2.3
men. 1 men. 2 men. 3 (arachnoid) men. 4 (dura) men. prolif. 1 men. prolif. 2
Cell types
J
GO biological processes of module 3
regulation of cell migration (GO:0030334)
regulation of stress fiber assembly (GO:0051492)
actin cytoskeleton reorganization (GO:0031532)
positive regulation of cell migration (GO:0030335)
negative regulation of cell migration (GO:0030336)
hippo signaling (GO:0035329)
platelet aggregation (GO:0070527)
axon guidance (GO:0007411)
protein localization to membrane (GO:0072657)
response to laminar fluid shear stress (GO:0034616)
−log10(p-value)
GO cellular components of module 3
focal adhesion (GO:0005925)
cell−substrate junction (GO:0030055)
actin cytoskeleton (GO:0015629)
cytoskeleton (GO:0005856)
integral component of plasma membrane (GO:0005887)
cell−cell junction (GO:0005911)
sarcolemma (GO:0042383)
collagen−containing extracellular matrix (GO:0062023)
adherens junction (GO:0005912)
endocytic vesicle (GO:0030139)
−log10(p-value)

Figure 2. *De novo* variants in AC genes disrupt epigenomic regulation and impact midgestational neural precursors and arachnoid cells. (A) EnrichR biological and cellular processes of high-confidence (hcAC) genes converge on pathways regulating chromatin modification. -Log_{10}(p-values) for hcAC genes are shown for EnrichR analysis. Significance thresholds, adjusted for multiple comparisons, are denoted by a vertical red line. Top enriched terms are visualized for readability. **(B) hcAC genes with roles in chromatin modification and transcriptional regulation.** Individual histone proteins are represented as blue circles. DNA is represented as a purple strand. Genes with high brain expression (HBE) in developing mouse brain (embryonic day 9) are boxed in red. Genes associated with autosomal dominant or X-linked OMIM disorders are boxed in blue. Boldface genes surpassed exome-wide significance (ESW) thresholds in DenovolyzeR analysis. Me: methyl group; Acyl: acetyl group. **(C) Enrichment of AC genes in gene modules of the midgestational human cortex compared to other disease genes.** Numbers displayed exceed the Bonferroni-corrected statistical significance threshold and are -log_{10}(p-value). Gene modules in red signify significant enrichment of hcAC genes. Height: human height gene set; pAC: probable AC gene set; hcAC: high-confidence AC gene set; Epilepsy: epilepsy gene set; CHD: congenital heart disease gene set; ASD: autism spectrum disorder gene set (see **Methods** for gene set determination details). **(D) Temporal dynamics of modules enriched with AC genes.** Peak expression of brown and royal blue modules is at post-conception week (PCW) 12-16. Peak expression of the light cyan module is at ~10-12 months (M). Asterisks represent significant differences in expression across modules at specific time points. **(E) Clusters of cell types in the midgestational human cortex defined by scRNAseq. (F) Clusters of cell types in the developing meninges defined by scRNAseq.** Different cell types are noted, see text for details. **(G) GOrilla biological and molecular functions of royal blue module converge on histone regulation.** The significance threshold, adjusted for multiple comparisons, is denoted by the vertical blue line. Top enriched terms are visualized for readability. **(H) Cell-type enrichment of AC and other disease genes in the midgestational human brain.** Numbers displayed exceed the Bonferroni-corrected statistical significance threshold and are -log_{10}(p-value). Different cell types are noted on the x-axis, see text for details. Cell types in red represent significant enrichment with hcAC or pAC gene sets. **(I) Cell-type enrichment of AC genes in the developing meninges. (J) GOrilla pathway analyses of enriched meningeal module 3 (arachnoid cells).** Significance threshold, adjusted for multiple comparisons, denoted by the vertical yellow line. Top enriched terms are visualized for readability.

Supplementary Table 4. *De novo* variant enrichment of GOrilla-term gene sets

	De novo enrichment for 617 arachnoid cyst patients						*De novo* enrichment for 1798 unaffected siblings of autism specturm disorder patients					
	Observed		Expected				Observed		Expected			
	n	Rate	n	Rate	Enrichment	P	n	Rate	n	Rate	Enrichment	P
GO:0034647 histone demethylase activity (H3-trimethyl-K4 specific) (n = 5)							**GO:0034647 histone demethylase activity (H3-trimethyl-K4 specific) (n = 5)**					
Total	4	0.01	0.4	0	11.3	**4.90E-04**	Total	5	0	1	0	4.86	4.00E-03
Synonymous	0	0	0.7	0	0	1.00E+00	Synonymous	0	0	0.7	0	0	1.00E+00
T-mis	1	0	0.1	0	10	9.50E-02	T-mis	2	0	0.5	0	4	9.00E-02
D-mis	3	0	0.1	0	41.6	**5.93E-05**	D-mis	1	0	0.2	0	4.77	1.89E-01
LoF	0	0	0	0	0	1.00E+00	LoF	2	0	0.1	0	16.4	7.00E-03
Protein-altering	4	0.01	0.3	0	15	**1.72E-04**	Protein-altering	5	0	0.8	0	6.42	1.00E-03
Damaging	3	0	0.1	0	26.3	**2.26E-04**	Damaging	3	0	0.3	0	9.05	5.00E-03
GO:0018024 histone-lysine N-methyltransferase activity (n = 42)							**GO:0018024 histone-lysine N-methyltransferase activity (n = 42)**					
Total	9	0.01	3.1	0	2.89	**4.78E-03**	Total	10	0.01	9.1	0.01	1.1	4.20E-01
Synonymous	0	0	1.2	0	0	1.00E+00	Synonymous	4	0	2.5	0	1.63	2.34E-01
T-mis	4	0.01	1.2	0	3.33	3.40E-02	T-mis	3	0	3.3	0	0.91	6.41E-01
D-mis	2	0	0.8	0	2.38	2.06E-01	D-mis	3	0	2.5	0	1.22	4.44E-01
LoF	3	0	0.3	0	10.4	**3.22E-03**	LoF	0	0	0.8	0	0	1.00E+00
Protein-altering	9	0.01	2.3	0	3.97	**5.76E-04**	Protein-altering	6	0	6.6	0	0.91	6.44E-01
Damaging	5	0.01	1.1	0	4.42	**6.07E-03**	Damaging	3	0	3.3	0	0.91	6.39E-01
GO:0071565 nBAF complex (n = 15)							**GO:0071565 nBAF complex (n = 15)**					
Total	13	0.02	1	0	13.5	**4.05E-11**	Total	1	0	2.8	0	0.36	9.39E-01
Synonymous	2	0	0.3	0	6.96	3.42E-02	Synonymous	0	0	1.3	0	0	1.00E+00
T-mis	2	0	0.4	0	5	6.20E-02	T-mis	0	0	1.1	0	0	1.00E+00
D-mis	1	0	0.2	0	4.83	**1.87E-01**	D-mis	1	0	0.6	0	1.66	4.52E-01
LoF	8	0.01	0.1	0	93.3	**6.70E-14**	LoF	0	0	0.3	0	0	1.00E+00
Protein-altering	11	0.02	0.7	0	16.3	**1.81E-10**	Protein-altering	1	0	2	0	0.51	8.59E-01
Damaging	9	0.01	0.3	0	30.7	**3.35E-11**	Damaging	1	0	0.9	0	1.18	5.73E-01
GO:0045893 positive regulation of transcription, DNA-templated (n = 1595)							**GO:0045893 positive regulation of transcription, DNA-templated (n = 1595)**					
Total	105	0.17	64.7	0.04	1.62	**2.60E-06**	Total	148	0.08	187.8	0.1	0.79	9.99E-01
Synonymous	14	0.02	18.8	0.01	0.74	8.95E-01	Synonymous	43	0.02	54.7	0.03	0.79	9.55E-01
T-mis	36	0.06	28.8	0.02	1.25	1.09E-01	T-mis	66	0.04	83.5	0.05	0.79	9.79E-01
D-mis	28	0.05	11.7	0.01	2.38	**3.87E-05**	D-mis	24	0.01	34.2	0.02	0.7	9.72E-01
LoF	27	0.04	5.3	0	5.06	**2.36E-11**	LoF	15	0.01	15.4	0.01	0.97	5.79E-01
Protein-altering	91	0.15	45.9	0.03	1.98	**2.80E-09**	Protein-altering	105	0.06	133.2	0.07	0.79	9.95E-01
Damaging	55	0.09	17.1	0.01	3.22	**2.62E-13**	Damaging	39	0.02	49.6	0.03	0.79	9.48E-01
GO:0006338 Chromatin remodeling (n = 264)							**GO:0006338 Chromatin remodeling (n = 264)**					
Total	42	0.07	9.8	0.02	4.3	**1.97E-14**	Total	32	0.02	28.3	0.02	1.13	2.69E-01
Synonymous	9	0.01	2.7	0	3.38	1.74E-03	Synonymous	7	0	7.7	0	0.91	6.50E-01
T-mis	6	0.01	4.2	0.01	1.43	2.15E-01	T-mis	14	0.01	12.2	0.01	1.15	3.40E-01
D-mis	12	0.02	2	0	6.1	**1.14E-06**	D-mis	7	0	5.7	0	1.22	3.50E-01
LoF	15	0.02	0.9	0	15.9	**1.36E-13**	LoF	4	0	2.7	0	1.46	2.94E-01
Protein-altering	33	0.05	7.1	0.01	4.64	**1.52E-12**	Protein-altering	25	0.01	20.6	0.01	1.21	1.93E-01
Damaging	27	0.04	2.9	0	9.27	**1.89E-17**	Damaging	11	0.01	8.5	0	1.3	2.33E-01
GO:0006338 Chromatin remodeling with pLI > 0.9 (n = 105)							**GO:0006338 Chromatin remodeling with pLI > 0.9 (n = 105)**					
Total	39	0.06	6.7	0.01	5.84	**1.09E-17**	Total	21	0.01	19.4	0.01	1.08	3.89E-01
Synonymous	6	0.01	1.8	0	3.29	1.10E-02	Synonymous	4	0	5.3	0	0.76	7.74E-01
T-mis	6	0.01	2.7	0	2.22	5.67E-02	T-mis	11	0.01	7.7	0	1.43	1.55E-01
D-mis	12	0.02	1.5	0	7.75	**9.59E-08**	D-mis	4	0	4.5	0	0.89	6.60E-01
LoF	15	0.02	0.7	0	22.7	**8.24E-16**	LoF	2	0	1.9	0	1.04	5.71E-01
Protein-altering	33	0.05	4.9	0.01	6.79	**4.69E-17**	Protein-altering	17	0.01	14.1	0.01	1.2	2.54E-01
Damaging	27	0.04	2.2	0	12.2	**2.17E-20**	Damaging	6	0	6.4	0	0.93	6.21E-01
GO:0006338 Chromatin remodeling with pLI > 0.9 and OMIM phenotype (n = 40)							**GO:0006338 Chromatin remodeling with pLI > 0.9 and OMIM phenotype (n = 40)**					
Total	37	0.06	3.1	0.01	11.9	**6.42E-27**	Total	9	0.01	9.1	0.01	0.99	5.52E-01
Synonymous	6	0.01	0.9	0	7.04	2.60E-04	Synonymous	1	0	2.5	0	0.4	9.16E-01
T-mis	5	0.01	1.2	0	4.17	7.75E-03	T-mis	6	0	3.3	0	1.82	1.17E-01
D-mis	12	0.02	0.8	0	14.6	**8.97E-11**	D-mis	1	0	2.4	0	0.42	9.08E-01
LoF	14	0.02	0.3	0	45.9	**5.18E-19**	LoF	1	0	0.9	0	1.13	5.88E-01
Protein-altering	31	0.05	2.3	0	13.7	**1.37E-24**	Protein-altering	8	0	6.6	0	1.22	3.39E-01
Damaging	26	0.04	1.1	0	23.1	**1.76E-26**	Damaging	2	0	3.3	0	0.61	8.38E-01

n: number of *de novo* variants (DNVs); Rate: number of DNVs per subject; Enrichment: ratio of observed to expected numbers of DNVs; D-mis: damaging missense variants as predicted by MetaSVM or MPC >2; T-mis: tolerated missense variants as predicted by MetaSVM or MPC >2; LoF: loss-of-function variants comprised of premature termination, frameshift or splice site variants; histone demethylase activity (H3-trimethyl-K4 specific) denotes genes associated with GOrilla term GO:0034647; histone-lysine N-methyltransferase activity denotes genes associated with GOrilla term GO:0018024); nBAF complex denotes genes associated with the neuronal BAF (barrier-to-autointegration) complex Gorilla term GO:0071565; positive regulation of transcription, DNA-templated represents genes associated with GOrilla term GO:0045893; chromatin remodeling represents genes associated with GOrilla term GO:0006338 (Eden et al. 2009).

Supplemental Figure 7

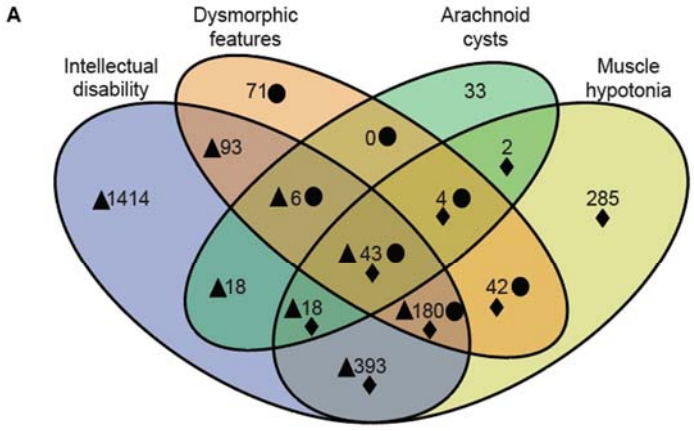
Supplementary Figure 7. (A) Enrichment of AC genes in meningeal gene modules. Numbers displayed exceed the Bonferroni-corrected statistical significance threshold and are -$\log_{10}$(p-value). pAC: possible AC gene set; hcAC; high-confidence AC gene set; EWS exome-wide significant; Mod: module**. (B) GOrilla and WikiPathways analyses of enriched arachnoid cell module 3.** The significance threshold, adjusted for multiple comparisons, is denoted by the vertical yellow line. The top terms are visualized for readability. **(C) Enrichment of gene modules in specific meningeal cell types.** Modules in red have similar meningeal cell-type enrichment compared to AC risk gene meningeal cell-type enrichment. The red asterisk highlights significant enrichment for cell types in the pAC gene set.

Supplementary Table 5. DisGeNET phenotypic overlap with hcAC and pAC gene list

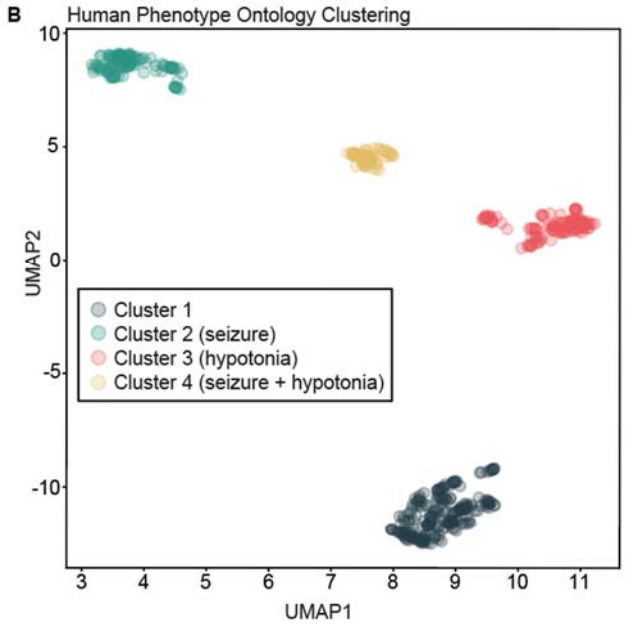
DisGeNET Phenotype	Genes in DisGeNET	DisGeNET Genes Overlapping with 124 hcAC & pAC Genes	Adj. P-Value
Muscle hypotonia	967	67/124	4.08 x 10-57
Dysmorphic features	439	53/124	9.71 x 10-57
Intellectual Disability	2165	85/124	1.51 x 10-55
Delayed speech and language development	560	54/124	1.73 x 10-52
Global developmental delay	1825	78/124	2.28 x 10-52
Neurodevelopmental disorders	535	51/124	4.57 x 10-49
Strabismus	716	54/124	1.08 x 10-46
Generalized hypotonia	955	56/124	1.41 x 10-42
Seizures	2152	72/124	4.39 x 10-40
Poor school performance	211	34/124	4.42 x 10-40
Downward slant of palpebral fissure	391	40/124	3.70 x 10-39
Developmental delay (disorder)	584	45/124	9.59 x 10-39
Multiple congenital anomalies	251	32/124	2.79 x 10-34
Feeding difficulties	473	38/124	3.97 x 10-33
Short stature	1127	51/124	7.03 x 10-33
Depressed nasal bridge	426	36/124	4.06 x 10-32
Movement Disorders	362	34/124	7.47 x 10-32
Autistic Disorder	1112	49/124	6.26 x 10-31
Absent speech	232	29/124	8.58 x 10-31
Microcephaly	1064	48/124	1.06 x 10-30
Byzanthine arch palate	497	36/124	9.80 x 10-30
Autistic behavior	261	29/124	2.85 x 10-29
Short Stature, CTCAE	1010	45/124	2.11 x 10-28
Severe intellectual disability	429	33/124	4.70 x 10-28
Blepharoptosis	595	36/124	5.33 x 10-27

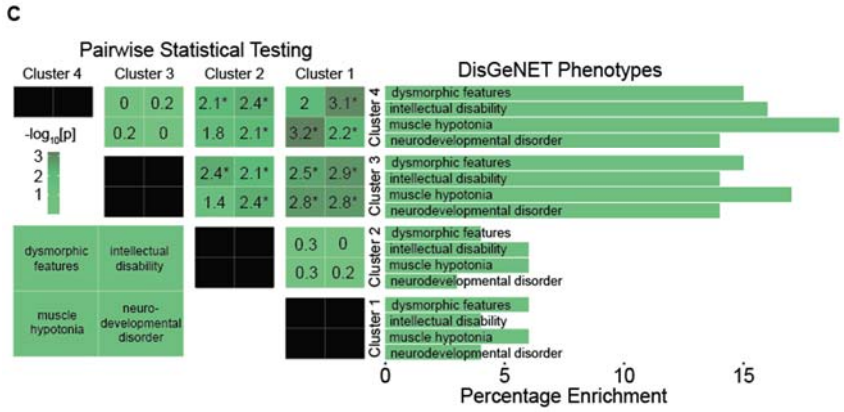
"Genes in DisGeNET" represent all known genes that associate with the listed DisGeNET phenotypes. hcAC: high-confidence AC gene set; pAC: probable AC gene set.

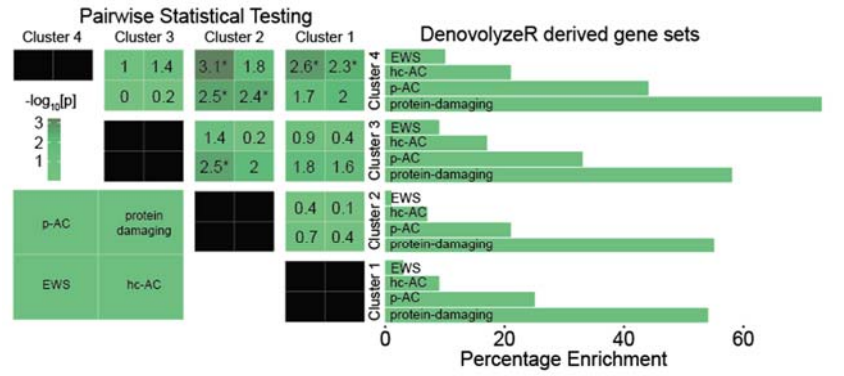
Figure 3.

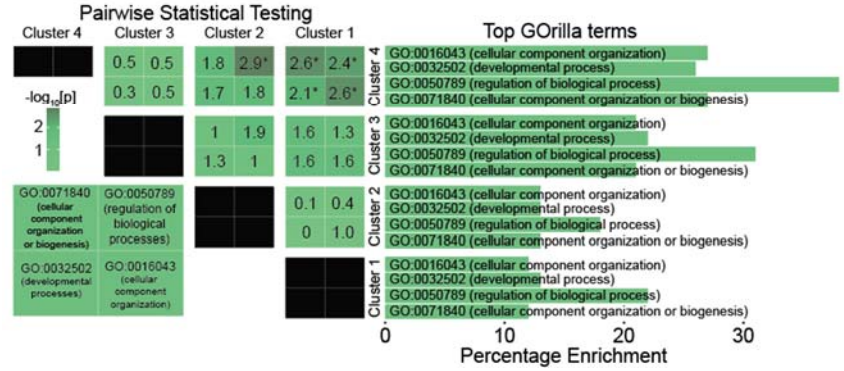


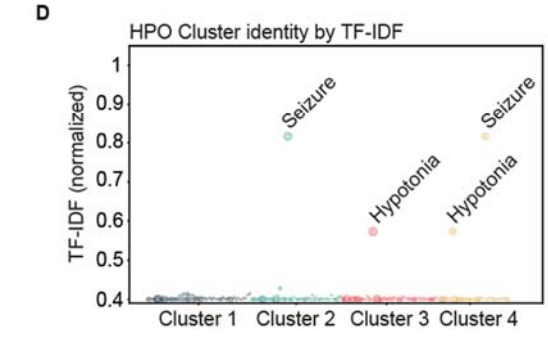
Disease	Genes in DisGeNET	DisGeNET genes overlapping p-AC genes	P-value
muscle hypotonia	967	67/124	♦ 4.08×10^{-57}
dysmorphic features	439	53/124	● 9.71×10^{-57}
intellectual disability	2165	85/124	▲ 1.51×10^{-55}











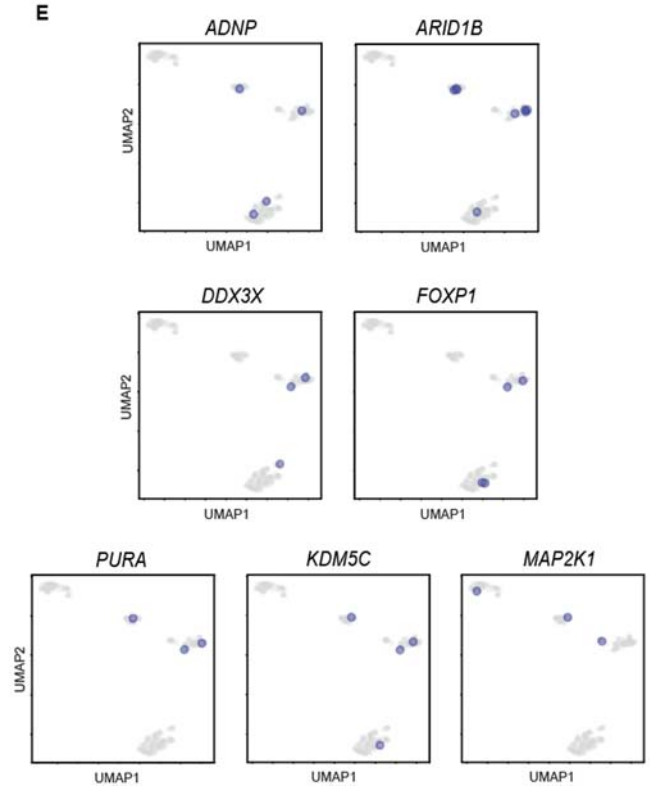


Figure 3. Unsupervised clustering of phenotype data identified clinical AC subtypes that correlate with genomic results. (A) DisGeNET phenotypic overlap of AC probands. Visualization of overlapping individual genes in the probable AC gene set (pAC) and the top-three DisGeNET-curated term gene lists enriched for pAC genes. The table below quantifies overlap and enrichment (one-sided Fisher's exact test). **(B) Human Phenotype Ontology (HPO) clustering of AC cohort.** Clustering of all patients with available phenotypic information reveals four major phenotypic clusters of AC probands. Phenotypic traits of probands were determined via text2hpo assessment of proband medical records. **(C) HPO cluster enrichment.** DisGeNET phenotype enrichment by cluster (top graph), relevant gene-set enrichment (middle graph) and GOrilla term enrichment (bottom graph) are represented. Pairwise statistical testing (student's T-test) is represented to the left of the bar graphs. Significant differences are denoted with an asterisk. Values shown are $-\log_{10}$(p-value) and significant is Bonferroni corrected. hcAC: high-confidence AC gene set; EWS: exome-wide significant **(D) HPO cluster identity by Term Frequency – Inverse Document Frequency (TF-IDF).** Labeled phenotypes are those that qualify as statistical outliers (See Methods). **(E) EWS variant distribution in HPO clusters.**

Supplementary Table 6. Top 20 Term Frequecies (TFs) per cluster

Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Term	Frequency	Term	Frequency	Term	Frequency	Term	Frequency
Arachnoid cyst	1.00	Arachnoid cyst	1.00	Arachnoid cyst	1.00	Arachnoid cyst	1.00
Developmental delay	0.56	Seizures	1.00	Generalized hypotonia	0.99	Generalized hypotonia	1.00
Craniofacial abnormalities	0.44	Developmental delay	0.69	Developmental delay	0.90	Seizures	1.00
Global developmental delay	0.43	Global developmental delay	0.55	Global developmental delay	0.80	Developmental delay	0.92
Delayed speech and language development	0.28	Craniofacial abnormalities	0.38	Craniofacial abnormalities	0.52	Global developmental delay	0.85
Macrocephaly	0.17	Delayed speech and language development	0.31	Delayed speech and language development	0.49	Craniofacial abnormalities	0.51
Gastrointestinal abnormalities	0.17	Autism	0.24	Motor delay	0.31	Delayed speech and language development	0.42
Autism	0.16	Autistic behavior	0.24	FTT	0.24	Gastrointestinal abnormalities	0.30
Autistic behavior	0.16	Macrocephaly	0.18	Constipation	0.24	Constipation	0.29
Respiratory abnormalities	0.15	Respiratory abnormalities	0.16	Failure to thrive	0.24	Autism	0.26
Constipation	0.14	Gastrointestinal abnormalities	0.16	Macrocephaly	0.21	Autistic behavior	0.26
Motor delay	0.14	Headache	0.15	Gastrointestinal abnormalities	0.20	FTT	0.22
Attention deficit hyperactivity disorder	0.13	Motor delay	0.14	Respiratory abnormalities	0.17	Failure to thrive	0.22
Microcephaly	0.13	Constipation	0.13	Urogenitial abnormalities	0.17	Microcephaly	0.22
Abnormality of calvarial morphology	0.12	Hydrocephalus	0.13	Cardiac abnormalities	0.14	Motor delay	0.21
Anxiety	0.12	Attention deficit hyperactivity disorder	0.13	Frontal bossing	0.14	Ophthalmological abnormalities	0.19
Gastroesophageal reflux	0.12	Agenesis of corpus callosum	0.11	Growth delay	0.14	Gastroesophageal reflux	0.19
Urogenitial abnormalities	0.12	Asthma	0.11	Microcephaly	0.14	Macrocephaly	0.16
Hydrocephalus	0.11	Microcephaly	0.11	Gastroesophageal reflux	0.14	Dysphagia	0.16

Term frequencies determined by text2hpo review of AC cohort patient medical records and medical surveys.

Extended Data Table 3. Term Frequecies (TFs) per cluster

Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Term	Frequency	Term	Frequency	Term	Frequency	Term	Frequency
Arachnoid cyst	1.00	Arachnoid cyst	1.00	Arachnoid cyst	1.00	Arachnoid cyst	1.00
Developmental delay	0.56	Seizures	1.00	Generalized hypotonia	0.99	Generalized hypotonia	1.00
Craniofacial abnormalities	0.44	Developmental delay	0.69	Developmental delay	0.90	Seizures	1.00
Global developmental delay	0.43	Global developmental delay	0.55	Global developmental delay	0.80	Developmental delay	0.92
Delayed speech and language development	0.28	Craniofacial abnormalities	0.38	Craniofacial abnormalities	0.52	Global developmental delay	0.85
Macrocephaly	0.17	Delayed speech and language development	0.31	Delayed speech and language development	0.49	Craniofacial abnormalities	0.51
Gastrointestinal abnormalities	0.17	Autism	0.24	Motor delay	0.31	Delayed speech and language development	0.42
Autism	0.16	Autistic behavior	0.24	FTT	0.24	Gastrointestinal abnormalities	0.30
Autistic behavior	0.16	Macrocephaly	0.18	Constipation	0.24	Constipation	0.29
Respiratory abnormalities	0.15	Respiratory abnormalities	0.16	Failure to thrive	0.24	Autism	0.26
Constipation	0.14	Gastrointestinal abnormalities	0.16	Macrocephaly	0.21	Autistic behavior	0.26
Motor delay	0.14	Headache	0.15	Gastrointestinal abnormalities	0.20	FTT	0.22
Attention deficit hyperactivity disorder	0.13	Motor delay	0.14	Respiratory abnormalities	0.17	Failure to thrive	0.22
Microcephaly	0.13	Constipation	0.13	Urogenitial abnormalities	0.17	Microcephaly	0.22
Abnormality of calvarial morphology	0.12	Hydrocephalus	0.13	Cardiac abnormalities	0.14	Motor delay	0.21
Anxiety	0.12	Attention deficit hyperactivity disorder	0.13	Frontal bossing	0.14	Ophthalmological abnormalities	0.19
Gastroesophageal reflux	0.12	Agenesis of corpus callosum	0.11	Growth delay	0.14	Gastroesophageal reflux	0.19
Urogenitial abnormalities	0.12	Asthma	0.11	Microcephaly	0.14	Macrocephaly	0.16
Hydrocephalus	0.11	Microcephaly	0.11	Gastroesophageal reflux	0.14	Dysphagia	0.16
Cardiac abnormalities	0.11	Ophthalmological abnormalities	0.10	Joint hypermobility	0.14	Hydrocephalus	0.15
Asthma	0.11	CP	0.10	Premature birth	0.14	Respiratory abnormalities	0.15
Premature birth	0.11	Cerebral palsy	0.10	Strabismus	0.13	Urogenitial abnormalities	0.15
FTT	0.10	Deafness	0.09	Deafness	0.12	Strabismus	0.15
Failure to thrive	0.10	Anxiety	0.09	Scoliosis	0.12	CP	0.14
Deafness	0.09	Hearing impairment	0.09	Hearing impairment	0.12	Cardiac abnormalities	0.14
Hearing impairment	0.09	FTT	0.08	Muscle weakness	0.12	Cerebral palsy	0.14
Headache	0.08	Scoliosis	0.08	Autism	0.11	Anxiety	0.12
Ophthalmological abnormalities	0.07	Failure to thrive	0.08	Autistic behavior	0.11	Nystagmus	0.12
Scoliosis	0.07	Strabismus	0.08	Ophthalmological abnormalities	0.11	Attention deficit hyperactivity disorder	0.11
Ataxia	0.07	Gastroesophageal reflux	0.08	Asthma	0.11	Asthma	0.10
Growth delay	0.07	Limb abnormalities	0.07	Dysphagia	0.11	Ataxia	0.10
Muscle weakness	0.06	Growth delay	0.07	Plagiocephaly	0.11	Cryptorchidism	0.10
Chiari	0.06	Cardiac abnormalities	0.06	Ataxia	0.09	Joint hypermobility	0.10
Spontaneous abortion	0.06	Abnormality of calvarial morphology	0.06	Ventriculomegaly	0.09	Torticollis	0.10
Hypertension	0.05	Dysphagia	0.06	Attention deficit hyperactivity disorder	0.08	Deafness	0.08
Jaundice	0.05	Polymicrogyria	0.06	Cryptorchidism	0.08	Scoliosis	0.08
Nystagmus	0.05	Ataxia	0.05	Torticollis	0.08	Hearing impairment	0.08
Strabismus	0.05	Joint hypermobility	0.05	CP	0.07	Micrognathia	0.08
Ventriculomegaly	0.05	Muscle weakness	0.05	Atrial septal defect	0.07	Sacral dimple	0.08
Limb abnormalities	0.05	Facial asymmetry	0.04	Cerebral palsy	0.07	Ventriculomegaly	0.08
Arnold-Chiari malformation	0.05	Heart murmur	0.04	Jaundice	0.07	Cerebral visual impairment	0.07
Atrial septal defect	0.05	Micrognathia	0.04	Heart murmur	0.06	Facial asymmetry	0.07
Joint hypermobility	0.05	Nystagmus	0.04	Hypospadias	0.06	Frontal bossing	0.07
Ventricular septal defect	0.05	Obsessive-compulsive behavior	0.04	Nystagmus	0.06	Growth delay	0.07
CP	0.04	Premature birth	0.04	Ventricular septal defect	0.06	Heart murmur	0.07
Aplasia/Hypoplasia of the cerebral white matt	0.04	Respiratory distress	0.04	Micrognathia	0.05	Muscle weakness	0.07
Cerebral palsy	0.04	Ventriculomegaly	0.04	Patent ductus arteriosus	0.05	Polyhydramnios	0.07
Depressivity	0.04	Urogenitial	0.03	Sacral dimple	0.05	Premature birth	0.07
Dysphagia	0.04	Depressivity	0.03	Hydrocephalus	0.05	Agenesis of corpus callosum	0.05
Heart murmur	0.04	Frontal bossing	0.03	Ankyloglossia	0.05	Atrial septal defect	0.05
Maternal diabetes	0.04	Hypothyroidism	0.03	Headache	0.05	Irritability	0.05
Patent ductus arteriosus	0.04	Jaundice	0.03	Obsessive-compulsive behavior	0.05	Ventricular septal defect	0.05
Plagiocephaly	0.04	Nuchal cord	0.03	Anxiety	0.04	Renal	0.04
Respiratory distress	0.04	Schizencephaly	0.03	Dysarthria	0.04	Blindness	0.04
Renal abnormalities	0.04	Renal	0.03	Facial asymmetry	0.04	Chronic lung disease	0.04
Hypoplasia of the corpus callosum	0.04	Abnormal heart morphology	0.03	Limb	0.03	Depressivity	0.04
Agenesis of corpus callosum	0.03	Abnormality of brain morphology	0.03	Renal	0.03	Hypoplasia of the corpus callosum	0.04
Cryptorchidism	0.03	Abnormality of limbs	0.03	Agenesis of corpus callosum	0.03	Hypothyroidism	0.04
Hypothyroidism	0.03	Blindness	0.03	Bradycardia	0.03	Jaundice	0.04

Term frequencies determined by text2hpo review of AC cohort patient medical records and medical surveys.

Extended Data Table 4; Cluster enrichment analysis of all trios with phenomic data, N=551

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		*p*- value of pairwise testing						
	N	Rate	N	Rate	N	Rate	N	Rate	1v2v3v4	1v2	1v3	1v4	2v3	2v4	3v4
Genes of interest															
EWS	7	3%	1	1%	12	9%	7	10%	**$2.6x10^{-3}$**	0.19	$1.4x10^{-2}$	$2.2x10^{-2}$	**$3.2x10^{-3}$**	**$3.2x10^{-3}$**	0.91
hcAC	21	9%	8	7%	23	17%	15	21%	**$4.0x10^{-3}$**	0.42	$2.3x10^{-2}$	$9.8x10^{-3}$	$1.0x10^{-2}$	**$4.2x10^{-3}$**	0.58
pAC	57	25%	25	21%	43	33%	32	44%	**$2.8x10^{-3}$**	0.39	0.13	**$2.3x10^{-3}$**	$3.9x10^{-2}$	**$7.8x10^{-4}$**	0.11
Variant class															
Damaging (Mis+LoF)	123	54%	66	55%	77	58%	53	73%	**$4.5x10^{-2}$**	0.82	0.45	**$5.4x10^{-3}$**	0.65	$1.8x10^{-2}$	$4.2x10^{-2}$
Synonymous	18	8%	12	10%	10	8%	2	3%	0.32	0.50	0.90	0.12	0.48	$5.7x10^{-2}$	0.16
GOrilla go-terms															
GO:0032502 (developmental process)	29	13%	15	13%	29	22%	19	26%	**$1.1x10^{-2}$**	0.96	$2.2x10^{-2}$	**$7.2x10^{-3}$**	$5.1x10^{-2}$	$1.8x10^{-2}$	0.51
GO:0016043 (cellular component organization)	28	12%	16	13%	28	21%	20	27%	**$7.5x10^{-3}$**	0.77	$2.5x10^{-2}$	**$2.3x10^{-3}$**	0.11	$1.6x10^{-2}$	0.32
GO:0071840 (cellular component organization or biogenesis)	28	12%	16	13%	28	21%	20	27%	**$7.5x10^{-3}$**	0.77	$2.5x10^{-2}$	**$2.3x10^{-3}$**	0.11	$1.6x10^{-2}$	0.32
GO:0050789 (regulation of biological process)	49	22%	21	18%	41	31%	28	38%	**$2.6x10^{-3}$**	0.39	$4.6x10^{-2}$	**$4.3x10^{-3}$**	$1.4x10^{-2}$	**$1.4x10^{-3}$**	0.29
GO:0051716 (cellular response to stimulus)	24	11%	10	8%	21	16%	13	18%	0.11	0.52	0.14	0.10	$7.1x10^{-2}$	$5.1x10^{-2}$	0.73
EnrichR go-terms															
GO:0006338 (chromatin remodeling)	1	0%	0	0%	3	2%	4	5%	**$6.5x10^{-3}$**	0.47	0.11	**$3.4x10^{-3}$**	0.10	$9.9x10^{-3}$	0.23
GO:0048096 (chromatin-mediated maintenance of transcription)	1	0%	0	0%	3	2%	4	5%	**$6.5x10^{-3}$**	0.47	0.11	**$3.4x10^{-3}$**	0.10	$9.9x10^{-3}$	0.23
GO:0034401 (chromatin organization involved in regulation of transcription)	0	0%	1	1%	1	1%	1	1%	0.48	0.17	0.19	$7.7x10^{-2}$	0.94	0.73	0.67
GO:0097193 (intrinsic apoptotic signaling pathway)	0	0%	0	0%	2	2%	0	0%	$9.5x10^{-2}$	NA	$6.3x10^{-2}$	NA	0.18	NA	0.29
GO:0006357 (regulation of transcription by RNA polymerase II)	2	1%	0	0%	2	2%	0	0%	0.45	0.30	0.58	0.42	0.18	NA	0.29
Disgenet															
Dysmorphic features	13	6%	5	4%	20	15%	11	15%	**$1.2x10^{-3}$**	0.54	**$2.9x10^{-3}$**	$1.0x10^{-2}$	**$3.8x10^{-3}$**	**$8.2x10^{-3}$**	0.99
Delayed speech and language development	10	4%	4	3%	19	14%	13	18%	**$3.5x10^{-5}$**	0.64	**$8.1x10^{-4}$**	**$1.8x10^{-4}$**	**$2.5x10^{-3}$**	**$6.2x10^{-4}$**	0.52
Muscle hypotonia	10	4%	7	6%	18	14%	12	16%	**$9.9x10^{-4}$**	0.55	**$1.7x10^{-3}$**	**$6.0x10^{-4}$**	$4.1x10^{-2}$	$1.7x10^{-2}$	0.59
Multiple congenital anomalies	10	4%	3	3%	14	11%	8	11%	**$1.2x10^{-2}$**	0.38	$2.3x10^{-2}$	$4.0x10^{-2}$	$1.1x10^{-2}$	$1.5x10^{-2}$	0.94
Intellectual disability	14	6%	7	6%	22	17%	14	19%	**$3.2x10^{-4}$**	0.92	**$1.4x10^{-3}$**	**$8.9x10^{-4}$**	**$7.6x10^{-3}$**	**$4.2x10^{-3}$**	0.65
Neurodevelopmental disorders	10	4%	4	3%	18	14%	10	14%	**$7.0x10^{-4}$**	0.64	**$1.7x10^{-3}$**	**$5.6x10^{-3}$**	**$4.0x10^{-3}$**	**$7.5x10^{-3}$**	0.99
Abnormality of the dentition	5	2%	2	2%	10	8%	9	12%	**$5.7x10^{-4}$**	0.74	$1.4x10^{-2}$	**$3.6x10^{-4}$**	$2.9x10^{-2}$	**$2.1x10^{-3}$**	0.26
Short stature	9	4%	4	3%	19	14%	12	16%	**$4.2x10^{-5}$**	0.78	**$3.8x10^{-4}$**	**$2.8x10^{-4}$**	**$2.5x10^{-3}$**	**$1.5x10^{-3}$**	0.70
Severe intellectual disability	8	4%	5	4%	17	13%	9	12%	**$1.3x10^{-3}$**	0.75	**$7.9x10^{-4}$**	**$4.7x10^{-3}$**	$1.5x10^{-2}$	$3.6x10^{-2}$	0.91
Global developmental delay	10	4%	6	5%	23	17%	12	16%	**$2.7x10^{-5}$**	0.79	**$3.8x10^{-5}$**	**$6.0x10^{-4}$**	**$2.2x10^{-3}$**	$8.5x10^{-3}$	0.86

N: number of probands within a cluster harboring the variants of interest. Rate: N/total number of probands within a cluster; EWS: exome-wide significant; hcAC: high-confidence AC gene set (see **Methods**); pAC: probable AC gene set (see **Methods**); Mis: missense variants; LoF: loss-of-function variants comprising premature termination, frameshift, or splice site variants. Pairwise testing was performed with chi-squared testing between groups. p-values reported are uncorrected and emboldened p-values surpass the Bonferroni correction threshold for multiple comparisons.

REFERENCES

White, T., Su, S., Schmidt, M., Kao, C.Y. & Sapiro, G. The development of gyrification in childhood and adolescence. *Brain Cogn* **72**, 36-45 (2010).

Juric-Sekhar, G. & Hevner, R.F. Malformations of Cerebral Cortex Development: Molecules and Mechanisms. *Annu Rev Pathol* **14**, 293-318 (2019).

Boucherie, C., *et al.* Neural progenitor fate decision defects, cortical hypoplasia and behavioral impairment in Celsr1-deficient mice. *Molecular psychiatry* **23**, 723-734 (2018).

Siegenthaler, J.A., *et al.* Retinoic acid from the meninges regulates cortical neuron generation. *Cell* **139**, 597-609 (2009).

Borrell, V. & Marin, O. Meninges control tangential migration of hem-derived Cajal-Retzius cells via CXCL12/CXCR4 signaling. *Nature neuroscience* **9**, 1284-1293 (2006).

Al-Holou, W.N., *et al.* Prevalence and natural history of arachnoid cysts in adults. *J Neurosurg* **118**, 222-231 (2013).

Mustansir, F., Bashir, S. & Darbar, A. Management of Arachnoid Cysts: A Comprehensive Review. *Cureus* **10**, e2458 (2018).

De Keersmaecker, B., *et al.* Outcome of 12 antenatally diagnosed fetal arachnoid cysts: case series and review of the literature. *European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society* **19**, 114-121 (2015).

Katzman, G.L., Dagher, A.P. & Patronas, N.J. Incidental findings on brain magnetic resonance imaging from 1000 asymptomatic volunteers. *JAMA : the journal of the American Medical Association* **282**, 36-39 (1999).

Koch, C.A., Moore, J.L. & Voth, D. Arachnoid cysts: how do postsurgical cyst size and seizure outcome correlate? *Neurosurg Rev* **21**, 14-22 (1998).

Hayes, M.J., TerMaath, S.C., Crook, T.R. & Killeffer, J.A. A Review on the Effectiveness of Surgical Intervention for Symptomatic Intracranial Arachnoid Cysts in Adults. *World Neurosurg* **123**, e259-e272 (2019).

Jafrani, R., Raskin, J.S., Kaufman, A. & Lam, S. Intracranial arachnoid cysts: Pediatric neurosurgery update. *Surg Neurol Int* **10**, 15 (2019).

Little, J.R., Gomez, M.R. & MacCarty, C.S. Infratentorial arachnoid cysts. *J Neurosurg* **39**, 380-386 (1973).

Choi, J.U. & Kim, D.S. Pathogenesis of arachnoid cyst: congenital or traumatic? *Pediatr Neurosurg* **29**, 260-266 (1998).

Starkman, S.P., Brown, T.C. & Linell, E.A. Cerebral arachnoid cysts. *J Neuropathol Exp Neurol* **17**, 484-500 (1958).

Zeegers, M., *et al.* Radiological findings in autistic and developmentally delayed children. *Brain Dev* **28**, 495-499 (2006).

Martinez-Lage, J.F., Perez-Espejo, M.A., Almagro, M.J. & Lopez-Guerrero, A.L. Hydrocephalus and arachnoid cysts. *Childs Nerv Syst* **27**, 1643-1652 (2011).

Huang, J.H., Mei, W.Z., Chen, Y., Chen, J.W. & Lin, Z.X. Analysis on clinical characteristics of intracranial Arachnoid Cysts in 488 pediatric cases. *Int J Clin Exp Med* **8**, 18343-18350 (2015).

Nikolic, I., *et al.* The association of arachnoid cysts and focal epilepsy: Hospital based case control study. *Clin Neurol Neurosurg* **159**, 39-41 (2017).

Al-Holou, W.N., *et al.* Prevalence and natural history of arachnoid cysts in children. *J Neurosurg Pediatr* **5**, 578-585 (2010).

Wiener, S.N., Pearlstein, A.E. & Eiber, A. MR imaging of intracranial arachnoid cysts. *J Comput Assist Tomogr* **11**, 236-241 (1987).

Gosalakkal, J.A. Intracranial arachnoid cysts in children: a review of pathogenesis, clinical features, and management. *Pediatr Neurol* **26**, 93-98 (2002).

Hall, S., *et al.* Clinical and radiological outcomes following surgical treatment for intra-cranial arachnoid cysts. *Clin Neurol Neurosurg* **177**, 42-46 (2019).

1648 Cilluffo, J.M., Gomez, M.R., Reese, D.F., Onofrio, B.M. & Miller, R.H. Idiopathic ("congenital") spinal
1649 arachnoid diverticula. Clinical diagnosis and surgical results. *Mayo Clin Proc* **56**, 93-101 (1981).
1650 Krauer, F., *et al.* Growth of arachnoid cysts in patients with autosomal dominant polycystic kidney
1651 disease: serial imaging and clinical relevance. *Clin Kidney J* **5**, 405-411 (2012).
1652 Zafeiriou, D.I. & Batzios, S.P. Brain and spinal MR imaging findings in mucopolysaccharidoses: a
1653 review. *AJNR Am J Neuroradiol* **34**, 5-13 (2013).
1654 Plaisier, E., *et al.* Novel COL4A1 mutations associated with HANAC syndrome: a role for the triple
1655 helical CB3[IV] domain. *American journal of medical genetics. Part A* **152A**, 2550-2555 (2010).
1656 Ogura, Y., *et al.* FOXC2 mutations in familial and sporadic spinal extradural arachnoid cyst. *PloS one* **8**,
1657 e80548 (2013).
1658 Bilguvar, K., *et al.* Whole-exome sequencing identifies recessive WDR62 mutations in severe brain
1659 malformations. *Nature* **467**, 207-210 (2010).
1660 Barak, T., *et al.* Recessive LAMC3 mutations cause malformations of occipital cortical development.
1661 *Nature genetics* **43**, 590-594 (2011).
1662 Mishra-Gorur, K., *et al.* Mutations in KATNB1 cause complex cerebral malformations by disrupting
1663 asymmetrically dividing neural progenitors. *Neuron* **84**, 1226-1239 (2014).
1664 Kundishora, A.J., *et al.* DIAPH1 Variants in Non-East Asian Patients With Sporadic Moyamoya
1665 Disease. *JAMA neurology* **78**, 993-1003 (2021).
1666 Jin, S.C., *et al.* Contribution of rare inherited and de novo variants in 2,871 congenital heart disease
1667 probands. *Nature genetics* **49**, 1593-1601 (2017).
1668 Furey, C.G., *et al.* De Novo Mutation in Genes Regulating Neural Stem Cell Fate in Human Congenital
1669 Hydrocephalus. *Neuron* **99**, 302-314 e304 (2018).
1670 Jin, S.C., *et al.* Exome sequencing implicates genetic disruption of prenatal neuro-gliogenesis in
1671 sporadic congenital hydrocephalus. *Nat Med* **26**, 1754-1765 (2020).
1672 de Ligt, J., *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *The New*
1673 *England journal of medicine* **367**, 1921-1929 (2012).
1674 Neale, B.M., *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*
1675 **485**, 242-245 (2012).
1676 Rauch, A., *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual
1677 disability: an exome sequencing study. *Lancet* **380**, 1674-1682 (2012).
1678 Timberlake, A.T., *et al.* Two locus inheritance of non-syndromic midline craniosynostosis via rare
1679 SMAD6 and common BMP2 alleles. *Elife* **5**(2016).
1680 Willsey, A.J., *et al.* De Novo Coding Variants Are Strongly Associated with Tourette Disorder. *Neuron*
1681 **94**, 486-499 e489 (2017).
1682 Jin, S.C., *et al.* Mutations disrupting neuritogenesis genes confer risk for cerebral palsy. *Nature genetics*
1683 **52**, 1046-1056 (2020).
1684 Kaplanis, J., *et al.* Evidence for 28 genetic disorders discovered by combining healthcare and research
1685 data. *Nature* **586**, 757-762 (2020).
1686 Krumm, N., *et al.* Excess of rare, inherited truncating mutations in autism. *Nature genetics* **47**, 582-588
1687 (2015).
1688 Lopez, A.J. & Wood, M.A. Role of nucleosome remodeling in neurodevelopmental and intellectual
1689 disability disorders. *Front Behav Neurosci* **9**, 100 (2015).
1690 Shi, C., *et al.* Down-regulation of the forkhead transcription factor Foxp1 is required for monocyte
1691 differentiation and macrophage function. *Blood* **112**, 4699-4711 (2008).
1692 Song, M.R. & Ghosh, A. FGF2-induced chromatin remodeling regulates CNTF-mediated gene
1693 expression and astrocyte differentiation. *Nature neuroscience* **7**, 229-235 (2004).
1694 Li, X., *et al.* MEK Is a Key Regulator of Gliogenesis in the Developing Brain. *Neuron* **75**, 1035-1050
1695 (2012).
1696 Chakraborty, R., *et al.* Mutually exclusive recurrent somatic mutations in MAP2K1 and BRAF support a
1697 central role for ERK activation in LCH pathogenesis. *Blood* **124**, 3007-3015 (2014).

Senawong, T., *et al.* Germline mutations of MEK in cardio-facio-cutaneous syndrome are sensitive to MEK and RAF inhibition: implications for therapeutic options. *Human molecular genetics* **17**, 419-430 (2008).
Aoidi, R., *et al.* Mek1(Y130C) mice recapitulate aspects of human cardio-facio-cutaneous syndrome. *Dis Model Mech* **11**(2018).
Nie, Z., *et al.* A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. *Mol Cell Biol* **20**, 8879-8888 (2000).
Tumber, A., *et al.* Potent and Selective KDM5 Inhibitor Stops Cellular Demethylation of H3K4me3 at Transcription Start Sites and Proliferation of MM1S Myeloma Cells. *Cell Chem Biol* **24**, 371-380 (2017).
Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
Walker, R.L., *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771 e722 (2019).
Stuart, J.M., Segal, E., Koller, D. & Kim, S.K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249-255 (2003).
Song, L., *et al.* STAB: a spatio-temporal cell atlas of the human brain. *Nucleic Acids Res* **49**, D1029-D1037 (2021).
DeSisto, J., *et al.* Single-Cell Transcriptomic Analyses of the Developing Meninges Reveal Meningeal Fibroblast Diversity and Function. *Dev Cell* **54**, 43-59 e44 (2020).
Castro Dias, M., Mapunda, J.A., Vladymyrov, M. & Engelhardt, B. Structure and Junctional Complexes of Endothelial, Epithelial and Glial Brain Barriers. *Int J Mol Sci* **20**(2019).
Kishi, Y. & Gotoh, Y. Regulation of Chromatin Structure During Neural Development. *Front Neurosci* **12**, 874 (2018).
Pinero, J., *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 (2017).
Sanchez-Rico, M. & Alvarado, J.M. A Machine Learning Approach for Studying the Comorbidities of Complex Diagnoses. *Behav Sci (Basel)* **9**(2019).
Arriola, G., de Castro, P. & Verdu, A. Familial arachnoid cysts. *Pediatr Neurol* **33**, 146-148 (2005).
Martinez, J.O., *et al.* Intracranial arachnoid cysts and epilepsy in children: Should this be treated surgically? Our 29-year experience and review of the literature. *Neurocirugia (Astur : Engl Ed)* (2021).
Shaban, H.A. & Seeber, A. Monitoring the spatio-temporal organization and dynamics of the genome. *Nucleic Acids Res* **48**, 3423-3434 (2020).
Valencia, A.M. & Pasca, S.P. Chromatin dynamics in human brain development and disease. *Trends Cell Biol* **32**, 98-101 (2022).
Sokpor, G., Xie, Y., Rosenbusch, J. & Tuoc, T. Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders. *Front Mol Neurosci* **10**, 243 (2017).
Alfert, A., Moreno, N. & Kerl, K. The BAF complex in development and disease. *Epigenetics Chromatin* **12**, 19 (2019).
Eissenberg, J.C. & Shilatifard, A. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental biology* **339**, 240-249 (2010).
Bragin, E., *et al.* DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res* **42**, D993-D1000 (2014).
De Rubeis, S., *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-215 (2014).
Zaidi, S., *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220-223 (2013).
Kadoch, C. & Crabtree, G.R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv* **1**, e1500447 (2015).
Rylaarsdam, L. & Guemez-Gamboa, A. Genetic Causes and Modifiers of Autism Spectrum Disorder. *Front Cell Neurosci* **13**, 385 (2019).

1749 Chen, T., Giri, M., Xia, Z., Subedi, Y.N. & Li, Y. Genetic and epigenetic mechanisms of epilepsy: a
1750 review. *Neuropsychiatr Dis Treat* **13**, 1841-1859 (2017).
1751 Lisi, E.C. & Cohn, R.D. Genetic evaluation of the pediatric patient with hypotonia: perspective from a
1752 hypotonia specialty clinic and review of the literature. *Dev Med Child Neurol* **53**, 586-599 (2011).
1753 Jahed, Z., Shams, H., Mehrbod, M. & Mofrad, M.R. Mechanotransduction pathways linking the
1754 extracellular matrix to the nucleus. *Int Rev Cell Mol Biol* **310**, 171-220 (2014).
1755 Rengachary, S.S. & Watanabe, I. Ultrastructure and pathogenesis of intracranial arachnoid cysts. *J*
1756 *Neuropathol Exp Neurol* **40**, 61-83 (1981).
1757 Kanton, S., *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain
1758 development. *Nature* **574**, 418-422 (2019).
1759 Rabiei, K., Hogfeldt, M.J., Doria-Medina, R. & Tisell, M. Surgery for intracranial arachnoid cysts in
1760 children-a prospective long-term study. *Childs Nerv Syst* **32**, 1257-1263 (2016).
1761 Gjerde, P.B., *et al.* Anxiety and Depression in Patients with Intracranial Arachnoid Cysts-A Prospective
1762 Study. *World Neurosurg* **132**, e645-e653 (2019).
1763 Kwiatkowska, K., *et al.* Are arachnoid cysts actually clinically mute in relation to neuropsychological
1764 symptoms? Cognitive functioning in children with AC of middle and cranial fossa. *Clin Neurol*
1765 *Neurosurg* **208**, 106825 (2021).
1766 Tamburrini, G., Dal Fabbro, M. & Di Rocco, C. Sylvian fissure arachnoid cysts: a survey on their
1767 diagnostic workout and practical management. *Childs Nerv Syst* **24**, 593-604 (2008).
1768 Schulz, M., *et al.* Surgical management of intracranial arachnoid cysts in pediatric patients: radiological
1769 and clinical outcome. *J Neurosurg Pediatr*, 1-11 (2021).
1770 Kousi, M. & Katsanis, N. The Genetic Basis of Hydrocephalus. *Annu Rev Neurosci* **39**, 409-435 (2016).
1771 Sadler, B., *et al.* Rare and de novo coding variants in chromodomain genes in Chiari I malformation.
1772 *American journal of human genetics* **108**, 100-114 (2021).
1773 Duran, D., *et al.* Mutations in Chromatin Modifier and Ephrin Signaling Genes in Vein of Galen
1774 Malformation. *Neuron* **101**, 429-443 e424 (2019).
1775 Nikolaev, S.I., *et al.* Somatic Activating KRAS Mutations in Arteriovenous Malformations of the Brain.
1776 *The New England journal of medicine* **378**, 250-261 (2018).
1777 Twigg, S.R. & Wilkie, A.O. A Genetic-Pathophysiological Framework for Craniosynostosis. *American*
1778 *journal of human genetics* **97**, 359-377 (2015).
1779 Timberlake, A.T., *et al.* Genetic Influence on Neurodevelopment in Nonsyndromic Craniosynostosis.
1780 *Plast Reconstr Surg* (2022).
1781 Retterer, K., *et al.* Clinical application of whole-exome sequencing across clinical indications. *Genet*
1782 *Med* **18**, 696-704 (2016).
1783 McKenna, A., *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
1784 generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).
1785 Van der Auwera, G.A., *et al.* From FastQ data to high confidence variant calls: the Genome Analysis
1786 Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33 (2013).
1787 Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint*
1788 *arXiv:1207.3907* (2012).
1789 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-
1790 throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
1791 Lek, M., *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291
1792 (2016).
1793 Karczewski, K.J., *et al.* The mutational constraint spectrum quantified from variation in 141,456
1794 humans. *Nature* **581**, 434-443 (2020).
1795 Taliun, D., *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*
1796 **590**, 290-299 (2021).
1797 DePristo, M.A., *et al.* A framework for variation discovery and genotyping using next-generation DNA
1798 sequencing data. *Nature genetics* **43**, 491-498 (2011).
1799 Mills, R.E., *et al.* Natural genetic variation caused by small insertions and deletions in the human
1800 genome. *Genome research* **21**, 830-839 (2011).

1800. Samocha, K.E., *et al.* Regional missense constraint improves variant deleteriousness prediction. *bioRxiv*, 148353 (2017).
1803. Dong, C., *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human molecular genetics* **24**, 2125-2137 (2015).
1805. Ware, J.S., Samocha, K.E., Homsy, J. & Daly, M.J. Interpreting de novo Variation in Human Disease Using denovolyzeR. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* **87**, 7 25 21-15 (2015).
1808. Purcell, S., *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575 (2007).
1810. Samocha, K.E., *et al.* A framework for the interpretation of de novo mutation in human disease. *Nature genetics* **46**, 944-950 (2014).
1813. Lango Allen, H., *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838 (2010).
1816. Grove, J., *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431-444 (2019).
1817. Li, M., *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**(2018).
1818. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4**, Article17 (2005).
1820. Kuleshov, M.V., *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016).
1822. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
1824. Kramer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530 (2014).
1826. Cao, J., *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
1828. Campello, R.J.G.B., Moulavi, D. & Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. in *Advances in Knowledge Discovery and Data Mining* (eds. Pei, J., Tseng, V.S., Cao, L., Motoda, H. & Xu, G.) 160-172 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).