



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Heinzen, EL;O'Neill, AC;Zhu, X;Allen, AS;Bahlo, M;Chelly, J;Chen, MH;Dobyns, WB;Freytag, S;Guerrini, R;Leventer, RJ;Poduri, A;Robertson, SP;Walsh, CA;Zhang, M

Title:

De novo and inherited private variants in MAP1B in periventricular nodular heterotopia

Date:

2018-05-01

Citation:

Heinzen, E. L., O'Neill, A. C., Zhu, X., Allen, A. S., Bahlo, M., Chelly, J., Chen, M. H., Dobyns, W. B., Freytag, S., Guerrini, R., Leventer, R. J., Poduri, A., Robertson, S. P., Walsh, C. A. & Zhang, M. (2018). De novo and inherited private variants in MAP1B in periventricular nodular heterotopia. *Plos Genetics*, 14 (5), <https://doi.org/10.1371/journal.pgen.1007281>.

Persistent Link:

<https://hdl.handle.net/11343/250600>

License:

CC BY

RESEARCH ARTICLE

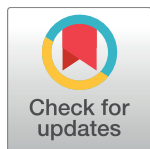
De novo and inherited private variants in *MAP1B* in periventricular nodular heterotopia

Erin L. Heinzen^{1*}, Adam C. O'Neill², Xiaolin Zhu¹, Andrew S. Allen^{3,4}, Melanie Bahlo^{5,6}, Jamel Chelly^{7,8}, Ming Hui Chen⁹, William B. Dobyns^{10,11}, Saskia Freytag¹², Renzo Guerrini¹³, Richard J. Leventer^{14,15,16}, Annapurna Poduri¹⁷, Stephen P. Robertson², Christopher A. Walsh^{18,19,20}, Mengqi Zhang^{4,21}, for the Epi4K Consortium[¶], Epilepsy Phenome/Genome Project[¶]

1 Institute for Genomic Medicine, Columbia University Medical Center, New York, New York, United States of America, **2** Department of Women's and Children's Health, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand, **3** Center for Statistical Genetics and Genomics, Duke University Medical Center, Durham, North Carolina, United States of America, **4** Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, United States of America, **5** Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia, **6** Department of Medical Biology, School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria, Australia, **7** Pôle de Biologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France, **8** IGBMC, INSERM U964, CNRS UMR 7104, Université de Strasbourg, Illkirch, France, **9** Department of Cardiology and Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, United States of America, **10** Departments of Pediatrics and Neurology, University of Washington, Seattle, Washington, United States of America, **11** Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington, United States of America, **12** Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia, **13** Neuroscience Department, Children's Hospital Anna Meyer-University of Florence, Florence, Italy, **14** Department of Neurology Royal Children's Hospital, University of Melbourne, Parkville, Victoria, Australia, **15** Murdoch Children's Research Institute, University of Melbourne, Parkville, Victoria, Australia, **16** Department of Pediatrics, University of Melbourne, Parkville, Victoria, Australia, **17** Department of Neurology, Division of Epilepsy and Clinical Neurophysiology, Boston Children's Hospital, Boston, Massachusetts, United States of America, **18** Division of Genetics and Genomics, Manton Center for Orphan Disease Research and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, Massachusetts, United States of America, **19** Departments of Pediatrics and Neurology, Harvard Medical School, Boston, Massachusetts, United States of America, **20** Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **21** Program in Computational Biology and Bioinformatics, Duke University, Durham, NC, United States of America

[¶] The full list of Epi4K Consortium, Epilepsy Phenome/Genome Project members and contributors are listed in the acknowledgments.

* Corresponding author on behalf of the Epi4K Consortium, epi4k@columbia.edu



 OPEN ACCESS

Citation: Heinzen EL, O'Neill AC, Zhu X, Allen AS, Bahlo M, Chelly J, et al. (2018) *De novo* and inherited private variants in *MAP1B* in periventricular nodular heterotopia. *PLoS Genet* 14(5): e1007281. <https://doi.org/10.1371/journal.pgen.1007281>

Editor: Joachim Hallmayer, Stanford School of Medicine, UNITED STATES

Received: September 25, 2017

Accepted: February 27, 2018

Published: May 8, 2018

Copyright: © 2018 Heinzen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Some sequence data has been restricted from sharing by the IRB boards of The Royal Children's Hospital Melbourne and the Health and Disability Ethics Committee as the consent gained only allows the data generated to be used by the investigators collecting the sample. Requests for data can be sent to the Epi4K Consortium (epi4k@columbia.edu). The rest of the sequence data, 128 out of 202, are available from dbGaP under Study Accession number phs001183.v1.p1.

Abstract

Periventricular nodular heterotopia (PVNH) is a malformation of cortical development commonly associated with epilepsy. We exome sequenced 202 individuals with sporadic PVNH to identify novel genetic risk loci. We first performed a trio-based analysis and identified 219 *de novo* variants. Although no novel genes were implicated in this initial analysis, PVNH cases were found overall to have a significant excess of nonsynonymous *de novo* variants in intolerant genes ($p = 3.27 \times 10^{-7}$), suggesting a role for rare new alleles in genes yet to be associated with the condition. Using a gene-level collapsing analysis comparing cases and controls, we identified a genome-wide significant signal driven by four ultra-rare loss-of-function heterozygous variants in *MAP1B*, including one *de novo* variant. In at least one instance, the *MAP1B* variant was inherited from a parent with previously undiagnosed PVNH. The PVNH was frontally predominant and associated with perisylvian

Funding: This work was supported by grants from the National Institute of Neurological Disorders and Stroke (The Epilepsy Phenome/Genome Project NS053998; Epi4K—Administrative Core NS077274; Epi4K—Sequencing, Biostatistics and Bioinformatics Core NS077303; Epi4K— Project 1 – Epileptic Encephalopathies NS077364, Epi4K—Phenotyping and Clinical Informatics Core NS077276), Curekids NZ and the Health Research Council of NZ 14/200, and the European Union through the Seventh Framework Programme (FP7) under the project DESIRE (N602531). AP was supported by NINDS K23NS069784 and the Boston Children’s Hospital Translational Research Program. R.J.L. is supported by a Melbourne Childrens Clinician Scientist Fellowship. Work by SF and MB was supported by the Victorian Government’s Operational Infrastructure Support Program and Australian Government NHMRC IRIS. MB is funded by NHMRC Senior Research Fellowship 110297 and NHMRC Program Grant 1054618. CAW was supported by the NINDS (RO1 NS 35129). CAW is an Investigator of the Howard Hughes Medical Institute. P.J.L. was supported by an NHMRC Career Development Fellowship (GNT1032364). The collection of control samples and data was funded in part by: Biogen, Inc. (<https://www.biogen.com>); Gilead Sciences, Inc. (<http://www.gilead.com>); UCB (<http://www.ucb-usa.com/Home>); Bryan ADRC NIA P30AG028377 (<https://neurology.duke.edu/research/research-centers/joseph-and-kathleen-bryan-alzheimers-disease-research-center-bryan-adrc-13>); B57 SAIC-Fredrick Inc M11-074; National Institute of Neurological Disorders and Stroke (<https://www.ninds.nih.gov>, RC2NS070344; RC2MH089915; U01NS077303; U01NS053998, U54NS078059, P01HD080642); National Human Genome Research Institute (<http://www.genome.gov>, Yale Mendelian Genomics Center - UM1HG006504, U01HG007672); National Institute of Mental Health (<https://www.nimh.nih.gov>, K01MH098126, R01MH097971, R01MH099216, RC2MH089915); National Institute of Diabetes and Digestive and Kidney Diseases (<https://www.niddk.nih.gov>, R01DK080099); National Institute of Allergy and Infectious Diseases (<https://www.niaid.nih.gov>, Division of Intramural Research, 1R56AI098588-01A1); National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (<https://chavi-id-duke.org>, UM1AI100645, U19AI067854); National Center for Advancing Translational Sciences (<https://ncats.nih.gov>, UL1TR000040); Eunice Kennedy Shriver National Institute of Child Health and Human Development (<https://nichd.nih.gov>, R01HD048805); the Ellison Medical

polymicrogyria. These results implicate *MAP1B* in PVNH. More broadly, our findings suggest that detrimental mutations likely arising in immediately preceding generations with incomplete penetrance may also be responsible for some apparently sporadic diseases.

Author summary

Almost 20 years ago the first gene responsible for periventricular nodular heterotopia (PVNH), a disorder that leads to abnormal migration of neurons during fetal brain development, was discovered. Since that time additional genes have been identified, but collectively they only explain a minority of cases. In this work we sought to further elucidate the genetic basis of this disorder using exome sequencing of 202 individuals with PVNH. We found a clear role for *de novo* mutations in PVNH, although with this analysis alone we were unable to pinpoint which of the *de novo* mutations in novel genes caused the disease. One patient was found to have a *de novo* variant in *MAP1B*, a gene that encodes a protein that plays a role at several key steps of brain development. With further analysis of the exome sequence data we found an additional three cases with a very rare inherited variant in *MAP1B*. This pattern is not expected to occur by chance and therefore indicates that these variants are likely responsible for the PVNH in these patients. Further strengthening the association of *MAP1B* in PVNH, all of the patients with a *MAP1B* variant had a similar brain abnormality, and at least one of the parents who transmitted the variant to their child was also similarly affected. This work adds to a growing list of genes responsible for PVNH, illuminates new genes involved in brain development, and importantly informs us about the types of genetic variants involved in PVNH.

Introduction

Malformations of cortical development are phenotypically heterogeneous and frequently associated with epilepsy, intellectual disability and congenital neurological deficits [1]. Periventricular nodular heterotopia (PVNH) is one such malformation where a population of neurons fails to migrate to the cerebral cortex and instead adopt heterotopic positions along their sites of origin—adjacent to the lateral ventricles [2]. Ten loci [*FLNA*, *ARFGEF2*, *FAT4*, *DCSH1*, *EML1*, *NEDD4L*, *INTS8*, *EML1*, *AKT3*, *MCPH1* and *C6orf70* (also known as *ERMARD*)] are currently implicated in the causation of PVNH [3–11]. Variants in these genes explain approximately 25% of sporadic instances of the brain malformation, with variants in *FLNA* being the most frequently found [3–10]. Despite only a small number of genes identified to date, germline genetic variation is thought to explain a significant fraction of patients, particularly in light of the often bilateral symmetric presentation and the lack of evidence for extrinsic etiologies [12–14].

To further characterize the genetic bases of PVNH, we exome sequenced 202 trios with sporadic PVNH and performed two analyses (Methods). First, we executed a trio-based approach to search for *de novo* risk variants in the patient population. Given the phenotypic heterogeneity of PVNH, ranging from mild, sometimes subclinical, to very severe [15, 16], and the presence of X-linked *FLNA*-positive cases in both sporadic and inherited PVNH, we also sought to evaluate the role of risk alleles agnostic to the mode of inheritance by performing a gene-level case-control collapsing analysis of 196 probands (excluding six individuals sequenced from lymphoblastoid cell line DNA) and controls. In the collapsing analysis, we searched for

Foundation New Scholar award AG-NS-0441-08; the Duke Chancellor's Discovery Program Research Fund 2014; Neil Molberger Brain Research Fund; Endocrine Fellows Foundation Grant (<http://www.endocrinefellows.org>); Bill and Melinda Gates Foundation (<https://www.gatesfoundation.org>); The Murdock Study Community Registry and Biorepository (<https://www.murdock-study.com/ongoing-studies/community-registry-and-biorepository/>); The Stanley Institute for Cognitive Genomics at Cold Spring Harbor Laboratory (<http://www.cshl.edu/Research/Neuroscience/Stanley-Institute-for-Cognitive-Genomics.html>); the Duke Genome Sequencing Clinic; New York-Presbyterian Hospital (<http://www.nyp.org>); Columbia University College Physicians and Surgeons (http://www.cumc.columbia.edu/education/physicians_surgeons); Columbia University Medical Center (<http://www.cumc.columbia.edu>); The J. Willard and Alice S. Marriott Foundation (<https://www.marriottfoundation.org>); The Muscular Dystrophy Association (<https://www.mda.org>); The Nicholas Nunno Foundation (<http://www.nicholasnunnofoundation.org>); The JDM Fund for Mitochondrial Research (<http://www.thejdmfund.org>); The Arturo Estopinan TK2 Research Fund; and The Endocrine Fellows Foundation; Helaine B Allen and Emily Allen Wolff. Data collection and sharing for the WHICAP project (used as controls in this analysis) was supported by the Washington Heights-Inwood Columbia Aging Project (WHICAP, <http://www.cumc.columbia.edu/dept/taub/res-normal.html>, PO1AG07232, R01AG037212, RF1AG054023) funded by the National Institute on Aging (NIA) and by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Number UL1TR001873. This manuscript has been reviewed by WHICAP investigators for scientific content and consistency of data interpretation with previous WHICAP Study publications. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: Stephen Petrou holds equity and is a scientific founder of Praxis Precision Medicines and RogCon. David Goldstein is a founder and holds equity in Pairnomix and Praxis Precision Medicines, and has research funding from Janssen, Gilead, Biogen, AstraZeneca, and UCB. Erin Heinzen has research funding from AstraZeneca.

enrichment of rare, putatively deleterious variants (inherited or *de novo*), within the protein-coding sequence of individual genes [17, 18]. The results of our analyses specifically pinpointed *de novo* and inherited variants in *MAP1B* in PVNH, and more broadly implicate ultra-rare, likely recently acquired variation in the genetic architecture of PVNH. Finally, given the challenges associated with distinguishing disease-relevant variations from background variation in genetically heterogeneous conditions, we utilized human brain-specific transcriptomic data [19, 20] to undertake a systems genetic analysis aimed at further organizing candidates for future investigation.

Results

Given the prominent role for *de novo* variation in severe, sporadic neurodevelopmental disorders [21–26], we first identified *de novo* variants within trios using GATK multi-sample calling as described previously [21, 27]. A total of 219 *de novo* variants were identified in the 202 trios (1.1 per trio, Methods, S1 Table).

Among these *de novo* variants, nine were located in *FLNA*, a previously identified PVNH gene (Table 1)[3]. Consistent with the known role of *FLNA* in PVNH, observing nine *de novo* variants in *FLNA* is extremely unlikely to occur by chance ($p = 3.4 \times 10^{-23}$, FitDNM method [28]). A *de novo* variant was also identified in *NEDD4L*; the genetic findings in this same patient were previously reported in Broix et al [7]. No *de novo* variants were detected in *C6orf70* (also known as *ERMARD*), a previously identified dominant PVNH gene [6]. Three additional genes had multiple *de novo* variants in unrelated individuals (Table 1), including three *de novo* variants in *CHD5* and two each in *UGGT1* and *PLXNC1*. *CHD5* did not have a statistically significant excess of *de novo* variants in a cohort of this size compared to that expected based on the mutability and size of the gene, as assessed using the FitDNM method [28] ($p = 1.6 \times 10^{-5}$). Because we do not have estimates of mutation rates for insertion-deletion variants, and both *UGGT1* and *PLXNC1* harbored one single nucleotide substitution and one insertion-deletion variant, we were not able to formally test for enrichment of *de novo* variants in these genes using the FitDNM method.

We next evaluated if *de novo* variation across the cohort had a distinct profile compared to controls using two orthogonal approaches. First, we performed a hot-zone analysis using previously described methods comparing profiles of *de novo* variation predicted to alter the level or activity of a protein that is encoded by a gene that has less than expected functional variation in the population (intolerant gene) in cases and controls [22] (Methods). We found significant enrichment of hot-zone *de novo* variants in PVNH cases (30.4%) compared to controls (9.6%) ($p = 0.001$). Removing the disease-causing *FLNA* variants from this analysis, a significant enrichment remained in PVNH cases (25%) compared to controls [odds ratio = 3.31 (95% CI: 1.22–9.08), $p = 0.01$]. Second, to further validate the observations from the hot-zone analysis comparing cases to controls, we also developed a likelihood model analysis to evaluate if the distribution of single nucleotide *de novo* variants in affected individuals differs significantly from that expected in the general population using a modified version of previously described methods [22, 27] (S1 Text). The model includes parameters estimating the relative risk associated with types of *de novo* variants and the proportion of the exome that confers PVNH risk. Given the results of the hot-zone analysis we specifically focused on analyses of the distribution of nonsynonymous *de novo* variants in the 4,317 intolerant genes [genes with a Residual Variation Intolerance Score (RIVS score) in the lowest 25th percentile]. We observed a highly significant shift in the distribution from expectation ($p = 3.27 \times 10^{-7}$). Further, point estimates from the *de novo* mutation architecture model suggest that <1% of intolerant genes are involved in PVNH risk and that each individual variant is highly penetrant [γ (relative risk): >4k] (S1

Table 1. Genes with multiple *de novo* variants.

Gene	Proband ID	variant id (chr-position-ref-var, hg19)	ExAC v0.3.1 MAF	RefSeq Transcript ID	Function	CDS position	Protein position	Amino acid substitution	Polyphen2
<i>CHD5</i>	pvhnd29397ly1	1-6185252-G-A	0.00004	NM_015557.2	synonymous	4302	1434	N	-
<i>CHD5</i>	pvhcnw8001bvg1	1-6219435-C-G	0	NM_015557.2	missense	348	116	K/N	benign
<i>CHD5</i>	pvhit130Lbou1	1-6166767-C-A	0	NM_015557.2	missense	5651	1884	R/L	probably_damaging
<i>FLNA</i>	pvhnd39214nu1	X-153592477-G-C	0	NM_001110556.1	stop_gained	2193	731	Y/*	-
<i>FLNA</i>	pvhcnw14103bvj1	X-153590679-G-A	0	NM_001110556.1	stop_gained	2587	863	R/*	-
<i>FLNA</i>	pvhnd32846lz1	X-153590679-G-A	0	NM_001110556.1	stop_gained	2587	863	R/*	-
<i>FLNA</i>	pvhnd39654ajz1	X-153589918-G-A	0	NM_001110556.1	stop_gained	2965	989	Q/*	-
<i>FLNA</i>	pvhnd26332me1	X-153586868-G-A	0	NM_001110556.1	stop_gained	4543	1515	R/*	-
<i>FLNA</i>	pvhnd29397ly1	X-153593004-G-A	0	NM_001110556.1	stop_gained	1912	638	Q/*	-
<i>FLNA</i>	pvhnd25061mw1	X-153588557-C-CCCCG	0	NM_001110556.1	frameshift indel	3605–3606	1202	-	-
<i>FLNA</i>	pvhnd37807nj1	X-153599348-ATCT-A	0	NM_001110556.1	Inframe indel	263–265	88–89	KM/M	-
<i>FLNA</i>	pvhcnw8001bvg1	X-153581372-C-A	0	NM_001456.3	missense	6199	2067	A/S	probably_damaging (0.986)
<i>PLXNC1</i>	pvhnd21601lu1	12-94658948-TAC-T	0	NM_005761.2	frameshift indel	3545–3546	1182	-	-
<i>PLXNC1</i>	pvhit256Jbpb1	12-94648646-C-T	0.00001	NM_005761.2	missense_variant	2965	989	R/W	probably_damaging (0.999)
<i>UGGT1</i>	pvhnd27930mf1	2-128890752-CG-C	0	NM_020120.3	Frameshift indel	1416	472	-	-
<i>UGGT1</i>	pvhnd40500bij1	2-128855098-A-C	0	NM_020120.3	missense	154	52	T/P	benign (0.366)

<https://doi.org/10.1371/journal.pgen.1007281.t001>

Fig). However, we note that the wide confidence intervals on the parameter estimates suggest considerable uncertainty in these estimates and that much larger samples sizes will be needed to refine estimates of genetic architecture parameters in PVNH.

Since PVNH can have highly variable clinical presentations, ranging from subclinical to severe, we hypothesized that in some cases the risk alleles may be inherited from clinically unaffected parents, as is well recognized to occur in *FLNA*-associated PVNH. To address this hypothesis, we performed an association test evaluating for enrichment of rare alleles across individual genes in cases compared to controls using a gene-based collapsing analysis [17]. We started with variant calls from exome sequence data generated from 196 PVNH cases (excluding the six samples where sequencing was from DNA extracted from a lymphoblastoid cell line) and 13,364 controls selected from other studies and non-enriched for neurodevelopmental, neuropsychiatric, or severe pediatric diseases. All ethnicities were included in the analysis and we ensured approximately equal proportions of ethnicities among cases and controls. After a relatedness check and principal component analysis (Methods, S2 Fig), a total of 196 cases and 13,151 controls remained for association analysis.

To identify genes associated with PVNH under the case-control association analysis framework, we performed a genome-wide search for an over- or under-representation of “qualifying variants” in protein-coding genes in cases compared to controls—looking for genes where rare alleles confer risk or protection, respectively (Methods). Under a loss-of-function-only (LoF-only) model, where qualifying variants are required to be LoF variants (Methods), two genes (*FLNA* and *MAP1B*) showed enrichment of qualifying variants in PVNH patients with genome-wide significant p-values (Table 2, Fig 1). As a negative control, we also evaluated synonymous variants and found no enrichment of synonymous variants in cases or controls (S3 Fig). *SON* also had genome-wide significant enrichment of qualifying variants in cases, however this signal was driven by four *de novo* variants that were later found to be sequencing artifacts.

Table 2. Top associations from the gene-level case-control collapsing analyses.

LoF-only						
	Case carrier	%Case carrier	Control carrier	%Control carrier	Fisher's exact test p-value	Rank of p-value among all genes tested
<i>FLNA</i>	7	3.57%	1	0.008%	$1.1 \times 10^{-12*}$	1/18405
<i>MAP1B</i>	4	2.04%	0	0%	$4.5 \times 10^{-8*}$	2/18405
LoF + missense ("probably damaging")						
	Case carrier	%Case carrier	Control carrier	%Control carrier	Fisher's exact test p-value	Rank of p-value among all genes tested
<i>FLNA</i>	12	6.12%	55	0.42%	$2.1 \times 10^{-10*}$	1/18405
<i>MAP1B</i>	4	2.04%	15	0.11%	1.5×10^{-4}	2/18405

*Genome-wide significant (Methods).

<https://doi.org/10.1371/journal.pgen.1007281.t002>

Remarkably, *MAP1B*, a gene not previously known to be associated with PVNH, was the second most significant gene (following *FLNA*) owing to the presence of LoF qualifying variants in 4 of the 196 cases, and the absence of a LoF qualifying variant among 13,151 controls (Table 2). When we expanded the qualifying variants to include "probably damaging" missense variants (Methods), *FLNA* was the only significant association signal due to the additional contribution of missense qualifying variants, which have previously been demonstrated to be pathogenic in PVNH patients (Table 2). However, *MAP1B* became less significant because missense qualifying variants were found in 15 of the 13,151 controls and none of the cases (Table 2, S3 Fig). We further examined each of the four *MAP1B* LoF qualifying variants identified in the four PVNH cases. All four were heterozygous (Table 3), including one *de*

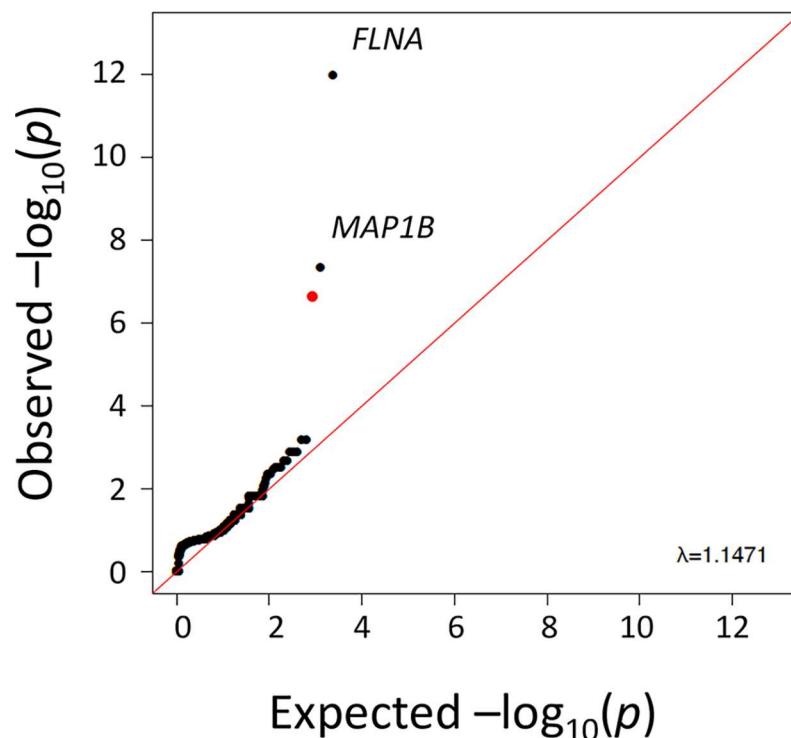


Fig 1. Quantile-quantile plot for gene-level association tests interrogating LoF variants. Black dots represent transformed p values against the expected transformed p values for genes with qualifying LoF variants. The red dot corresponds to the p value associated with *SON* however all four variants driving this signal were found to be false positives with Sanger sequencing. The red line indicates the expectation under the null model of no effect on risk.

<https://doi.org/10.1371/journal.pgen.1007281.g001>

Table 3. MAP1B LoF qualifying variants identified in PVNH patients.

Variant ID (chr-position-ref-var, hg19)	Proband ID	Parental origin	Ethnicity (self-declared, inferred from genetic data)	Variant/reference depth (proband): (transmitting parent)	Annotation*
5-71490089-C-T	pvhnd29281lw1	de novo	American Indian—Alaska Native/ European	63/70: NA	c.907C>T;p.(Arg303*)
5-71490776-C-T	pvhcw12701bvi1	inherited from father	European/European	38/31: 26/29	c.1594C>T;p. (Gln532*)
5-71492498-C-T	pvhnz9000cfc1	inherited from mother	Unknown/European	43/39: 35/26	c.3316C>T;p. (Arg1106*)
5-71489999-GC-G	pvhit1238Pbti1	inherited from father	European/European	44/50: 47/55	c.818delC;p. (Leu274Cysfs*4)

*Annotations are based on the canonical transcript NM_005909.3.

<https://doi.org/10.1371/journal.pgen.1007281.t003>

de novo and three inherited variants. None of the four cases had been resolved by a genetic diagnosis (e.g., *FLNA* or *NEDD4L* variants). All four variants are predicted to cause early premature truncation to the microtubule-associated protein 1B, which is 2,468 amino acids in length (Table 3, Fig 2). *MAP1B* is very intolerant to standing functional variation with an ExAC-based RVIS percentile of 2.27% and only 12 LOF variants (20 alleles) observed in the ExAC and gnomAD databases combined [29]. It is also a LoF-depleted gene achieving a ExAC-based FDR of 1.53×10^{-11} for preferential depletion of LoF variants [30] and a probability of being LoF intolerant (pLI) score of one [29, 31]. One additional de novo LOF variant in *MAP1B* was identified in a patient reported to have a range of phenotypes including an abnormality of the nervous system in the Deciphering Developmental Disorders Study (Fig 2, p. (Glu659Lysfs*22))[32], but MRI data that would allow for a diagnosis of PVNH was unavailable. All *MAP1B* variants were confirmed to be present with Sanger sequencing and the inheritance patterns were correctly inferred from the exome sequence data. *MAP1B* encodes a neuronal microtubule-associated protein that plays a key role in neurogenesis and neuronal migration through its effects on microtubule assembly and axon formation [33–35].

All of our patients with a *MAP1B* variant have anterior PVNH, bilateral and symmetric in three, and two of the four have deep perisylvian / insular polymicrogyria (S2 Table, S4 Fig, Fig 3). The pattern of PVNH is distinctive in that the nodules are frontal-predominant. This compares to the typical *FLNA*-associated PVNH in which the nodules are maximal along the bodies of the lateral ventricles, and the posterior or infrasylvian form of PVNH in which the nodules are maximal along the atria and temporal horns [36]. Seizures, cognitive impairment,

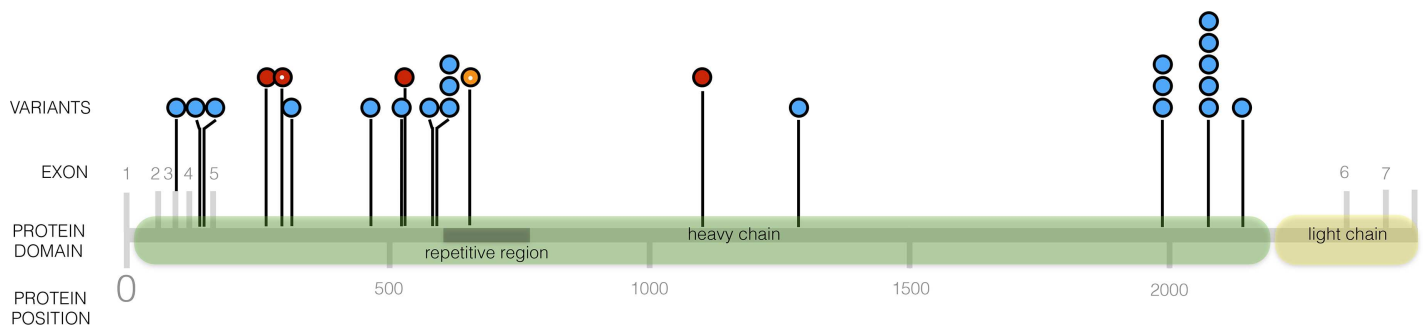


Fig 2. Distribution of *MAP1B* LOF alleles in PVNH cases (red dots), in individuals from ExAC and gnomAD databases (blue dots with number of alleles observed represented by number of dots running vertically at this site), and in the Deciphering Developmental Disorders case (orange dot). A Sanger confirmed de novo variant is indicated with a white dot in the circle.

<https://doi.org/10.1371/journal.pgen.1007281.g002>

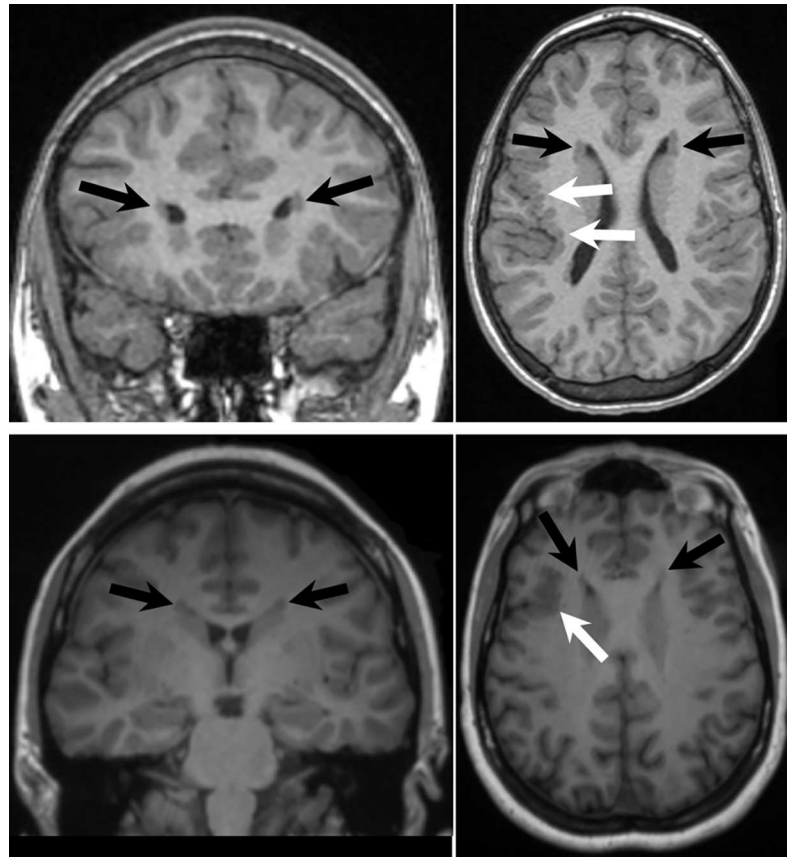


Fig 3. Brain MRI of subjects pvhzn9000cfc1 (top) and mother (bottom). Images are coronal T1-weighted (left column) and axial T1-weighted (right column). The images all show bilateral periventricular nodular grey matter heterotopia maximal in the frontal regions (black arrows). The axial images show over-folded cortex in the deep perisylvian/insular region on the right consistent with polymicrogyria (white arrows).

<https://doi.org/10.1371/journal.pgen.1007281.g003>

and other dysmorphic features were variable across the four patients. Only one of the three transmitting parents (mother of pvhzn9000cfc1) reported having possible neurological symptoms and was available for additional clinical evaluation. Interestingly, both the mother and child in this family had similar neuroimaging findings, consisting of bilateral anterior PVNH and deep perisylvian and insular polymicrogyria (Fig 3), although the extent of the brain malformation was much milder in the mother. This distinctive phenotype in both mother and child with the same *MAP1B* variant, and the rareness of the phenotype in patients with PVNH [37], further implicates *MAP1B* in PVNH. The other two parents from whom probands inherited variants in *MAP1B* did not report neurological symptoms and had not undergone neuroimaging.

We next evaluated if we could detect additional association signals by looking across sets of genes comprising a pathway. To do so, we first removed the two marginally significant genes (*FLNA* and *MAP1B*) as well as the gene generating an artifactual signal (*SON*) from these analyses. We interrogated the 10,705 pathways defined in the GSEA Hallmark and C2 Gene Sets, and Gene Ontology defined gene sets [38–40]. To assess whether there was any evidence of residual non-null signal across genes in each pathway we used a higher criticism approach (Methods) [41] that is especially sensitive to detecting low-level signals across a series of genes. No pathway was significantly associated with PVNH status after removing the gene-level

signals driven by *FLNA* and *MAP1B* variants. These analyses suggest that if additional association signal is present in the dataset, we are either underpowered to detect it or the relevant pathways are not captured amongst the gene sets evaluated.

Two biallelic risk models, one including LoF only and one LoF and “probably damaging” missense variants, were also evaluated considering a combination of both recessive or compound heterozygous qualifying variants using a case-control collapsing approach. No statistically significant association signals were detected and no gene had more than two qualifying genotypes in PVNH cases. No disease-causing biallelic genotypes were detected in the four known recessive PVNH genes *ARFGEF2*, *FAT4*, *DCHS1*, *INTS8*, *MCPH1* or *EML1*. Newly recessive, compound heterozygous, and newly hemizygous genotypes identified in the PVNH trios are provided in S3 and S4 Tables.

Recently, it has been shown that epileptic encephalopathy genes tend to be co-expressed in the brain during development, and that, for this epilepsy phenotype, identifying genes harboring a single *de novo* mutation in a trio-based study that are transcriptionally co-regulated with known disease genes can effectively pinpoint genes that will be associated with the phenotype in larger cohorts [19, 20]. Both the hot-zone and the architecture analyses suggest that there are additional pathogenic *de novo* variants beyond the *de novo* variant in *MAP1B* and *NEDD4L* and those found in *FLNA*. In order to nominate candidate PVNH genes amongst the set with a *de novo* mutation in one of the 202 PVNH cases evaluated in this study, we evaluated human brain development developmental co-expression patterns of *de novo* mutation carrying genes with known PVNH genes. Since the prioritization approach will be most effective in cases where known genes tend to be co-expressed, we first evaluated if a set of human and rodent PVNH genes [(n = 14) Methods], exhibit greater co-expression than random sets of 14 genes during brain development (Methods). We showed that, among 1,000 randomly selected 14-gene sets, PVNH genes, including *MAP1B*, tend to have higher correlation coefficients when evaluating all possible two gene correlations within the 14 PVNH gene set (S5 Fig)[42–44]. We next evaluated the prioritization procedure (Methods) using a leave-one-out approach (S2 Text) where a known PVNH gene is removed from the list and evaluated to see if it would be subsequently prioritized. We found this approach was able to successfully reprioritize more PVNH genes than expected by chance (S5 Table). Based on these analyses, we compiled a list of all genes harboring at least one *de novo* variant predicted to alter the encoded protein and excluded those occurring in any of the 14 known PVNH loci to identify the genes that are co-expressed with known PVNH genes (S2 Text). Using this approach, 14/107 candidate genes exceeded the empirical significance cut-off in both transcriptomic datasets analyzed. These genes included *LRIG3*, *MBNL1*, *ARID4B*, *NLGN1*, *KIFC3*, *SV2A*, *ADAM17*, *KIFAP3*, *FUBP3*, *ARHGAP35*, *PI4KA*, *MCM8*, *EDEM3*, and *DCX* (S6 Table). Co-expression heatmaps showing the modules of co-regulation of known and prioritized PVNH genes are provided in Fig 4. The patterns of co-expression of these candidate PVNH genes with human PVNH genes across development are also provided in S6 Fig. These 14 genes harboring *de novo* variants in our 202 trios that co-express with known PVNH genes should be considered candidate PVNH genes, particularly those with hot-zone variants (*MBNL1*, *ARID4B*, *NLGN1*, *ARHGAP35*, *EDEM3* and *DCX*).

Discussion

The PVNH cohort of 202 cases analyzed in this study was assembled with the goal of identifying novel variants and genes for this disorder. In this cohort we sought to identify disease-causing *de novo* variants considering the standard *de novo* variant paradigm in sporadic disease, and also rare inherited risk alleles since PVNH can exist in patients with subtle clinical or purely radiological presentations.

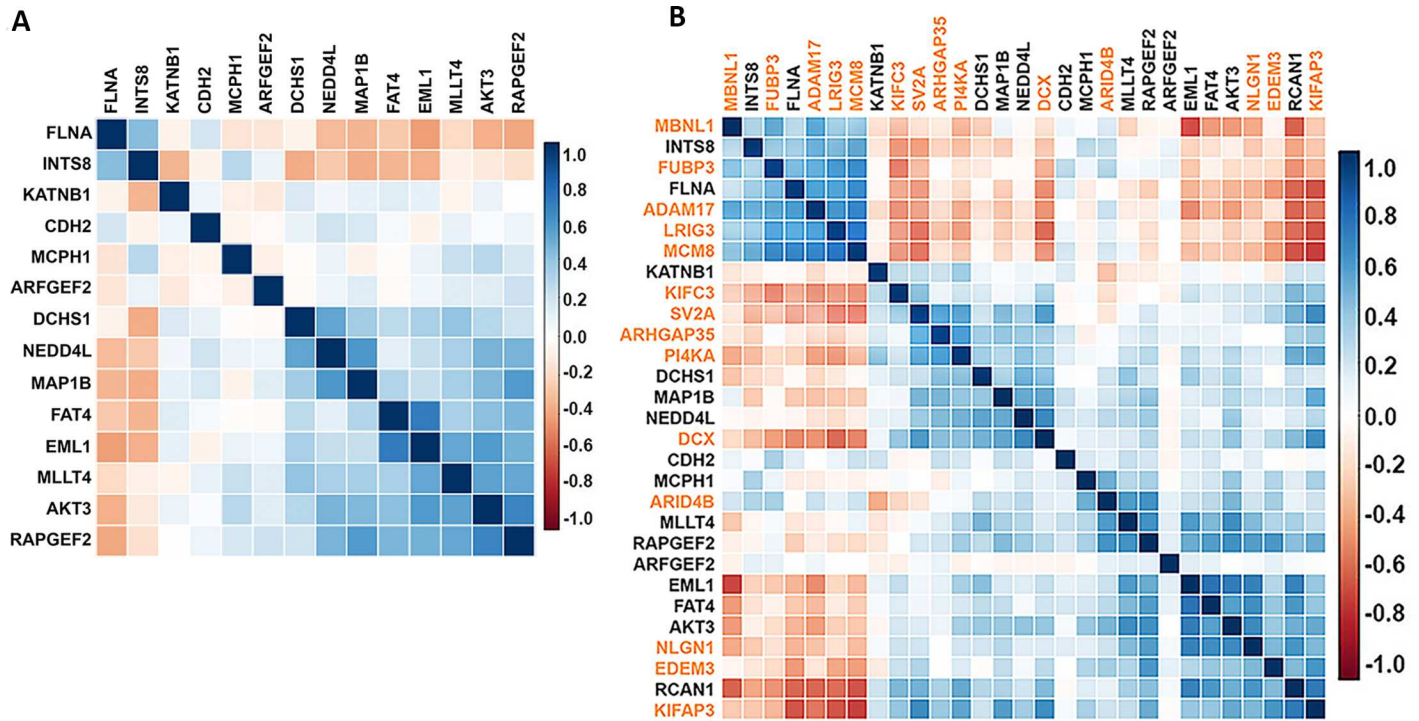


Fig 4. Ordered correlation matrices for the PVNH query and the fourteen loci significantly co-expressing within this node. Pairwise Pearson's correlation represented as a matrix between (a) pairs of the 14 genes within the PVNH gene set (Methods) and (b) the human PVNH query plus the 14 genes whose co-regulatory patterns significantly exceed the eFDR in both the Kang and Miller transcriptomic datasets. Genes are ordered according to hierarchical clustering, with the most positive (+1) and negative (-1) co-regulatory interactions represented as blue and red squares, respectively.

<https://doi.org/10.1371/journal.pgen.1007281.g004>

Using a trio approach, no gene in this study showed a genome-wide significant enrichment of *de novo* variants, other than *FLNA*, an already established PVNH gene. Despite this, results from our hot-zone analyses estimate that approximately 15 patients (7.4% of the cohort) harbor a non-*FLNA* *de novo* pathogenic variant, despite our inability in this small cohort to pinpoint these specific variants. This is further supported by the highly significant enrichment of nonsynonymous *de novo* variants in intolerant genes in the architecture analysis.

While we cannot pinpoint the exact pathogenic *de novo* variants outside of those in known PVNH genes, we suspect that a number of genes harbor pathogenic variants based either on meeting the hot-zone criteria or showing evidence of co-expression in the brain with known PVNH genes during the critical developmental time period. In total we identified 35 variants meeting the hot-zone criteria in 29 genes (S1 Table). Among these hot-zone *de novo* variants was one located in *MAP1B*, a gene implicated in this study through the gene-level collapsing analysis. *CHD5* was also found to harbor one hot-zone *de novo* variant, along with one synonymous *de novo* variant found in 0.004% of controls (S1 Table) and one missense variant predicted to be benign by Polyphen-2. Despite observing three *de novo* variants in *CHD5*, this pattern could occur by chance accounting for the mutability, predicted impact of the variants, and the size of the gene in a cohort of this size. However, this is an interesting candidate gene given what is known about the biological role of this gene. *CHD5* encodes the chromatin-remodeling protein chromodomain helicase DNA binding protein 5, which binds DNA and regulates transcription [45, 46]. *CHD5* expression is restricted to the brain where it activates genes promoting neuron terminal differentiation. Acute knockdown of *CHD5* within the developing mouse cortex, via *in utero* electroporation, impairs radial migration and causes a

failure of cells to reach the cortical plate [47]. Additional studies will be needed to confirm or disprove this candidate association.

Further complementing the hot-zone analyses, we also used brain-specific human transcriptomic resources to nominate candidate genes based on their co-regulatory expression patterns with known PVNH genes. Interestingly, known disease-causing PVNH genes form distinct patterns of co-expression with loci that produce similar phenotypes, suggesting the co-expression networks outlined here are supportive of a common pathway. For example, the expression patterns of *FLNA* and *INTS8* are highly correlated across development (Fig 4). Pathogenic variants in *FLNA* produce a phenotype of symmetrically distributed heterotopia predominantly lining the anterior horns and ventricular bodies of the lateral ventricles. Hypoplasia of the cerebellar vermis and posterior fossa cysts are common accompaniments [12]. A very similar clinical phenotype is produced by variants in *INTS8* [10]. Although the genes significantly co-expressing with the query PVNH set should be viewed only as candidate PVNH loci several are also hot-zone variants, further reinforcing their potential role in PVNH.

Using a gene-level collapsing analysis to assess enrichment for both inherited and *de novo* alleles in PVNH cases, we identified a significant enrichment of loss-of-function variants in *MAP1B* in cases compared to controls, allowing us to clearly implicate this gene in PVNH risk. Interestingly, three of the four *MAP1B* variants driving the association signal were found to be transmitted from a unaffected parent, explaining why it was not identified in the initial trio analysis. None of the three inherited *MAP1B* variants showed evidence of mosaicism based on the number of reads supporting the variant compared to reference (Table 3).

MAP1B, encoding microtubule associated protein 1B, is involved in regulating both microtubule and actin dynamics. Specifically, *MAP1B* is encoded as a single peptide with one cleavage site located near the C-terminus. The subsequent cleavage of *MAP1B* induces the production of a heavy and light chain that can both interact with microtubules [48]. Neurons lacking *MAP1B* have reduced Rac1 and Cdc42 activity, with a concomitant increase in RhoA [49]. Changes in neurite extension and synapse development have also been associated with *MAP1B* modulation [50]. Although *MAP1B* is most commonly associated with roles in postmitotic neurons, a recent study in zebrafish indicates a role for Map1b earlier in neural convergence and neural tube development [51]. This role may also be relevant for the formation of PVNH where early functions in epithelial adherens junction formation have also been implicated [52–54]. *MAP1B* transcripts are predominantly detected in the early stages of cortical development where they are also negatively regulated by the Fragile X mental retardation protein (FMRP), an important cellular process contributing to various neurodevelopmental diseases [55, 56]. Interestingly, PVNH has also been reported in patients with Fragile X syndrome due to marked expansion and instability of the CGG trinucleotide repeat within the *FMR1* gene [57].

In addition to implicating a novel gene in PVNH, one of the most interesting aspects of this work is the idea that sporadic disease may, in some cases, be due to deleterious variants that arise in the germline in earlier antecedents to the proband yet for some reason fail to give rise to a phenotype in these individuals. While non-penetrance is always a consideration in genetic risk, the unique component here is that the *MAP1B* variants identified in this study are very rare (absent from at least 150K samples encompassing all public and internal databases) and loss-of-function variants are virtually absent from the population as well, a bioinformatics signature that is consistent with disease-causing *de novo* variants. This suggests that LOF variants in *MAP1B* would be likely to have occurred in very recent generations. Such a pattern has been documented for rare deleterious copy number variants where high-risk deletions or duplications have been shown to be transmitted from a clinically unaffected first or second degree relative, but this is only very rarely reported for sporadic diseases caused by point or small insertion-deletion variants [58, 59]. In fact this very phenomenon has been described probabilistically and shown not

only to be possible but likely, depending on the disease penetrance and reproductive fitness conferred by the variants in question [60]. Interestingly, Kosmicki et al. recently reported over-transmission of LoF variants in LoF-depleted genes in a large cohort of sporadic autism spectrum disorder, a finding consistent with some transmitted alleles conferring risk [59]. This expansion of the *de novo* paradigm in PVNH may be in part due to the syndrome's potential to result in sub-clinical phenotypes, but it may also represent the tip of the iceberg for a much more widespread effect in sporadic disease risk that has largely not been considered in most trio-based studies performed to date.

Methods

Ethics statement

The study was performed according to the standards of the ethics committees and the institutional review boards at each institute. Columbia University Medical Center's Institutional Review Board centrally reviewed the approvals from each site under protocol number AAAP0052.

Patient ascertainment and phenotyping

PVNH patients were assembled from multiple patient collections sites, including: (1) multiple sites encompassing the Epilepsy Phenome/Genome Project (EPGP, www.epgp.org) Cohort (n = 70), University of Florence's Anna Meyer Children's Hospital (n = 22), Boston Children's Hospital (n = 24), University of Washington (n = 12), University of Otago (n = 65), and the Royal Children's Hospital Melbourne (n = 10). All samples had presumed sporadic disease based on patient and family interview, and all except for a subset in the EPGP cohort were pre-screened either clinically or in the research setting for disease-causing *FLNA* variants. MRIs were reviewed for the EPGP cohort as previously described [14] and for the additional cohorts by the enrolling sites. Patients enrolled into the EPGP cohort were excluded if an *FLNA* variant had been previously identified although not all patients underwent genetic testing for *FLNA* variants. EPGP inclusion criteria included the presence of epilepsy, whereas patients in other cohorts did not necessarily have epilepsy.

For comparison, 13,198 individuals who were sequenced as part of other genetic studies in the IGM were used as controls in this study. Approximately 6900 were neuropsychiatrically normal to our knowledge, and the remaining subjects had conditions where there are no known co-morbidities with epilepsy or brain malformations (S7 Table).

Exome sequencing

Exome sequencing was performed on DNA from 202 probands and their parents at the Institute for Genomic Medicine (IGM, Columbia University), the Dunedin School of Medicine (University of Otago, New Zealand), and the Institute for Applied Genomics (Udine, Italy) (S8 Table). All externally-generated raw data were transferred to the IGM, where a combined analysis was performed using the same alignment and variant calling pipeline. The alignment and variant calling details have been previously reported [21]. Six of the 202 trios studied had one or more samples from the family exome sequenced from DNA extracted from a lymphoblastoid cell line (LCL); all others were sequenced from primary DNA sources (S8 Table).

De novo variant calling

Candidate *de novo* variants were jointly called with the GATK Unified Genotyper for all family members in a trio as described previously [21]. Variants not located in the exonic region or

splice sites (2-basepairs flanking an exon) defined by the Consensus Coding Sequence (CCDS, release 14, GRCh37.p13) were excluded. On average ~20 *de novo* single nucleotide variants were called per individual using this permissive calling approach. To remove the false positives from the dataset, we used Sanger sequencing validation results of a subset of *de novo* single variant calls from a subset of the PVNH trio cohort and from 403 individuals analyzed as part of other trio sequencing studies performed in the IGM to fit a machine-learning model using variant-level, individual-level, and genomic features to predict true positives (S3 Text, S7 Table). Trios sequenced from DNA from LCLs were excluded from this analysis. A fraction of the trios sequenced at another site were also excluded from this model because of insurmountable batch effect issues that confounded the model predictions. The resulting data set was comprised of 401 Sanger validated and 317 Sanger refuted (including inherited variants and variants not confirmed in proband) from 535 (132 PVNH and 403 from other studies) trios. Cross-validation was used to estimate the model's accuracy, which was found to have high sensitivity (98%) and specificity (93%) (S3 Text). Confident in the model's ability to predict true and false *de novo* calls, we then applied the model to the 9,172 *de novo* variant calls where Sanger sequencing was not performed. For each variant evaluated, the model assigned a probability score reflecting how likely the call is a true *de novo* variant. A score approaching one had a high probability of being a true *de novo* variant, and a score approaching zero had a low probability of being a true *de novo* variant (S7 Fig). We then set a threshold probability score for declaring a true *de novo* variant using the expected number of autosomal synonymous *de novo* variants per trio of 0.303 which translated to an expectation of 162 autosomal synonymous *de novo* variants across the cohort of 535 trios. The expected per trio rate of autosomal *de novo* synonymous variants was calculated by taking sum of the estimated trinucleotide mutation rate [27, 61, 62] across all possible substitutions in the autosomal protein-coding sequence that would not result in a change in the protein-coding sequence and multiplying by two to account for the two chromosomes. The threshold probability score for declaring a *de novo* variant true was set to 0.978, which allowed for 162 total *de novo* variants to be accepted as true either via direct Sanger confirmation or by having the highest probability score in the model.

Since *de novo* variants called in LCL trios, indel *de novo* variants, and those from trios beset by confounding batch effects were not analyzed in the model approach, we Sanger sequenced the majority of *de novo* calls that we felt may contribute to PVNH, including those that passed quality control filters and those that were absent in IGM controls and the ExAC and EVS databases. Quality control filters included: single nucleotide variant calls were excluded if with $QD < 2.0$, $MQ < 40.0$, $FS > 60.0$, $HS > 13.0$, $MQRS < -12.5$, or $RPRS < -8.0$; indel variant calls were excluded if with $QD < 2.0$, $RPRS < -20.0$, or $FS > 200$. More than 70% of this subset of calls were confirmed with Sanger sequencing.

A list of all Sanger confirmed and model predicted true *de novo* variants identified in the PVNH cohort ($n = 219$ *de novo* variants) are provided in S1 Table.

Listing of newly recessive, compound heterozygous, and newly hemizygous genotypes in PVNH probands

We first identified all putatively protein-altering (missense, nonsense, or indels) residing in the protein-coding regions (CCDS, release 14, GRCh37.p13) newly recessive, compound heterozygous, and newly hemizygous genotypes in the PVNH probands by assessing the genotypes across the trios. Genotypes were excluded if they had a quality score (QUAL) < 30 and a genotype quality (GQ) score of < 20 in the proband. We also required a minimum coverage of 10-fold at a variant site to call a homozygous reference genotype. Newly recessive, compound

heterozygous, and newly hemizygous genotypes were excluded if any contributing variant had a the minor allele frequency greater than 1% or if a recessive or hemizygous genotype was reported in the internal control cohort or any population in Exome Variant Server (EVS) and Exome Aggregate Consortium (ExAC release 0.3).

Hot-zone analysis

Sanger confirmed and model-predicted true (see [Methods](#)) single nucleotide *de novo* substitutions found in PVNH cases and absent from in-house controls and the EVS and ExAC databases were first scored on their likelihood to alter the encoded protein. Trios with one or more samples sequenced from lymphoblastoid cell lines were excluded from this analysis. Synonymous and loss-of-function (nonsense, and splice acceptor/donor) variants were scored 0 and 1, respectively, and missense variants were scored using their Polyphen-2 score (HumVar). We next scored each *de novo* variant at the gene level using the gene-level residual variation intolerance percentile (RVIS, %RVIS_ExAC_0.05% (all populations), which ranks genes based on their tolerance to polymorphic functional genetic variation [63] on a scale from 0 to 1, with the higher the value the more tolerant the gene is to standing functional variation. For comparison, we also assessed *de novo* variants in previously published healthy control trios (n = 250 [64]) using the same annotations and filtering procedures used in this study. For each case and control sample with more than one *de novo* substitution meeting the aforementioned criteria, only the single most damaging *de novo* variant was used; i.e. the *de novo* variants with the shortest Euclidian distance from the most damaging coordinate [x = 1, y = 0] on a plot of the variant-level vector along the X-axis and the gene-level vector (RVIS percentile score) along the Y-axis. Individuals with no single nucleotide *de novo* substitutions did not contribute to this analysis. A two-tail Fisher's exact was used to test whether the single most damaging *de novo* variants found in PVNH cases preferentially lie in the "hot-zone", defined by a PolyPhen-2 score of ≥ 0.95 and RVIS $\leq 25^{\text{th}}$ percentile [63], compared to control trios.

Gene-level collapsing analyses

Variants for analysis were restricted to the consensus coding sequence public transcripts (CCDS release 14) plus 2 base pair intronic extensions. Variants were further required to have: i) at least 10-fold coverage, ii) quality score (QUAL) of at least 30, iii) genotype quality (GQ) score of at least 20, iv) quality by depth (QD) score of at least 2, v) mapping quality (MQ) score of at least 40, vi) read position rank sum (RPRS) score greater than -3, vii) mapping quality rank sum (MQRS) score greater than -6, viii) indels were required to have a maximum Fisher's strand bias (FS) of 200, ix) variants were screened according to VQSR tranche calculated using the known SNV sites from HapMap v3.3, dbSNP, and the Omni chip array from the 1000 Genomes Project to "PASS" SNVs were required to achieve a tranche of 99.9% for SNVs in genomes and exomes and 99% for indels in genomes, x) for heterozygous genotypes, the alternate allele ratio was required to be $\geq 25\%$. Finally, variants were excluded if they were among a predefined list of known sequencing artifacts or if they were marked by EVS (<http://evs.gs.washington.edu/EVS/>) or ExAC (<http://exac.broadinstitute.org/about>) as being problematic variants. Variants were annotated to Ensembl 73 using SnpEff. All variants meeting these criteria were eligible to be qualifying variants in the gene-based collapsing analyses. Additional filtering based on variant function or per inheritance models being tested were applied depending on the sub-analysis performed. We note that the case and control populations were pre-screened with both KING and PLINK to ensure only unrelated (up to second-degree) samples were used. Any exomes with gender discordance between clinically-reported and X:Y coverage ratios were removed, as were contaminated samples according to VerifyBamID. No PVNH cases were excluded with this filtering.

Before running gene-based collapsing analysis, we implemented both sample- and site-level pruning procedures to minimize the systemic bias in data that might lead to spurious association or reduced power to detect real association. The site-pruning procedure was performed as described previously [18]. Here, we described the sample-level pruning procedure including removing related individuals and population outliers identified in principal component analysis (PCA). To identify related individuals, we generated genotype data in PLINK format and then used KING to calculate pairwise kinship coefficients for all case and control subjects. No individual were found to be related greater than the kinship coefficient 0.1. Next we ran PCA using EIGENSTRAT with a LD-pruned (r^2 threshold 0.2) list of single-nucleotide polymorphisms (SNPs) extracted from exome sequencing data.

Following cleaning of the dataset, we then assessed for enrichment or depletion of “qualifying variants” across cases and controls. A “qualifying variant” was defined by a set of criteria based on allele frequency and functional prediction of that variant, with the criteria designed to capture the characteristics of previously identified pathogenic variants causing PVNH. Specifically, in this study, a variant was determined to be qualifying in the dominant model if it 1) was absent in the Exome Variant Server (EVS) and Exome Aggregate Consortium (ExAC release 0.3), 2) had ≤ 4 copies of variant allele in the 196 cases plus 13,151 controls, and 3) was predicted to be loss-of-function (stop gained, frameshift, splice site acceptor, splice site donor, start lost, or exon deleted) or missense “probably damaging” by PolyPhen-2 (HumDiv).

In the bi-allelic model, that included compound heterozygous and recessive genotypes, a genotype was considered qualifying if 1) the variant site(s) was predicted to be loss-of-function (stop gained, frameshift, splice site acceptor, splice site donor, start lost, or exon deleted) or missense “probably damaging” by PolyPhen-2 (HumDiv), and 2) the variant(s) sites had a minor allele frequency of < 0.001 in the Exome Variant Server (EVS), the Exome Aggregate Consortium (ExAC release 0.3), and across the case-control cohort.

For each gene, an indicator variable (1/0 states) was assigned to each individual based on the presence of at least one qualifying variant (dominant model) or genotype (bi-allelic model) in the gene (state 1) or no qualifying variant/genotype in that gene (state 0). We note that phasing of compound heterozygous variants was not taken into account in the collapsing analyses due to the fact that those data were not available for the control cohort. Two-tailed Fisher’s exact test was used to evaluate statistical significance of genic association. With 18,405 genes tested, we adopted the genome-wide significance level of $p = 6.79 \times 10^{-7}$ using Bonferroni correction correcting for all the genes in the genome and the four different models tested ($0.05/18,405/4$).

Quantile-quantile plots were generated using a permutation-based expectation. To achieve this, for each model (matrix) we randomly permuted the case and control labels of the original configuration: 196 cases and 13,151 controls and then recomputed the Fisher’s Exact test for all genes. This was repeated 1,000 times. For each of the 1,000 permutations we ordered the p-values and then took the mean of each rank-ordered estimate across the 1,000 permutations, i.e., the average 1st order statistic, the average 2nd order statistic, etc. Thus, these represent the empirical estimates of the expected ordered p-values (expected $-\log_{10}(p\text{-values})$). This empirical-based expected p-value distribution no longer depends on an assumption that the p-values are uniformly distributed under the null.

Pathway analyses

Analyses begin with marginal, gene-level p-values from a standard collapsing analysis. However, in order to optimize computational speed we use a standard chi-square test instead of Fisher’s Exact test. Since the ultimate null distribution was computed via permutation, we

expect this change to have minimal impact. Analyses were performed across all 10,705 pathways defined in the GSEA Hallmark and C2 Gene Sets, and Gene Ontology defined gene sets [38–40]. The higher criticism (HC) test is obtained by maximizing a scaled difference between the observed distribution of gene-level p-values across the pathway and the distribution of p-values one would expect if all the genes were null [41]. We use permutation to compute this expectation. The HC test is not only sensitive to extreme p-values but also to more subtle shifts in the p-value distribution. We compute unweighted and weighted versions of the HC test, where the weighted version upweights genes that are especially important (node centrality or low genic intolerance) within the pathway. Since a given gene may be included in multiple pathways, the resulting pathway-level tests will be correlated. Thus, in order to account for the large number of tests conducted while also taking this correlation into account, we use the permutation-based multiplicity adjustment procedure of Ge et al [65].

PVNH gene set

We established a list of genes associated with PVNH and related phenotypes based on evidence from the rodent and human literature for use in the co-expression analyses. To compile this list we specifically included all genes for which mutations reproducibly produce subcortical or periventricular heterotopia in a substantial fraction of individuals, and excluded genes with only single reports of periventricular heterotopia in human. Nine genes have been previously reported to be human PVNH risk loci in more than one individual, including *FLNA*, *FAT4*, *DCHS1*, *ARFGEF2*, *C6orf70*, *AKT3*, *INTS8*, *MCPH1*, and *NEDD4L* [3–11]. Genes for which mutations are reported to cause a partial, diffuse, heterotopic malformation, specifically subcortical band heterotopia were excluded, however we did include genes associated with subcortical heterotopia often presenting with PVNH, including *GPSM2* [66], *EML1* [5] and *KATNB1* [67]. In addition to human PVNH genes, we also included those genes in mice which when conditionally knocked-out induce impaired neuronal migration phenotypes analogous to, or closely resembling, subcortical or PVNH in humans, including *CTNNA1*, *RAPGEF2*, *RCAN1* and *MLLT4* [68–71]. We also include *MAP1B* based on the data presented in this report. *C6orf70*, *RCAN1* and *GPSM2* were not represented in all transcriptomic datasets analysed here and were therefore excluded. The PVNH genelist thus consisted of 14 genes.

Brain transcriptomic datasets

We downloaded three publicly available transcriptomic datasets generated from post-mortem human brain. While all datasets contain only donors deemed to have normal brain development, they are very different with regards to number of sampled regions per donor, number of donors and ages of donors. For the Miller et al [42] dataset, on average 328 regions from 4 fetal brains were assayed. The Colantuoni et al dataset exclusively looked at the prefrontal cortex in 266 brains from fetuses, children and adults [44]. The Kang et al dataset similarly includes donors of all ages but sampled on average 57 regions per brain. Additional details of these datasets are provided in [S9 Table](#). All three transcriptomic datasets contained post-mortem human brain expression data from the disease-relevant time periods (4–38 weeks post conception, [S9 Table](#)), however in some cases the data were limited. Although each dataset is built from various brain structures, for the purposes of these prioritizations this information was not used.

Using methods described previously [72], outlier samples were first removed, followed by normalization using the Removal Of Unwanted Variation (RUV) method (R package *RUVcorr*) that controls for systematic noise using negative control genes [73].

The three datasets used here were subsetted so that only the transcriptomic information within the disease relevant periods (4–38 weeks post conception) were used targeted for use in

the candidate gene prioritization, however only two (Miller and Kang datasets) were found to have sufficient data to be useful in these analyses (S2 Text, S5 Table). Thus, co-expression analyses were limited to just the Miller and Kang transcriptomic datasets.

Prioritization of genes harboring a *de novo* variant based on co-expression with known genes

In this study, we identified 107 genes harboring model predicted true or Sanger confirmed *de novo* variants, excluding *FLNA*, *NEDD4L* and *MAP1B*. To prioritize candidate PVNH gene based on co-expression with the 14 PVNH associated genes (see above), we first estimated the background correlation coefficient for any random 107 gene set whereby 20% of genes in this list would be prioritized. To do this we generated 1,000 sets each containing 107 randomly selected genes. In each random set, the pair-wise absolute weighted correlation between the expression of each of these random genes and all PVNH genes were calculated. The weighted correlation refers to correlations weighted by the inverse of the number of samples contributed by the respective donor. For any single gene in the 107 gene set, only the maximum correlation with a PVNH gene was retained, resulting in 107 correlation coefficients retained for each randomly selected gene set, and 18,200 total correlations across all 1,000 sets of genes. From this distribution of 18,200 maximum absolute correlations, the correlation coefficient threshold was that corresponding to the lowest value for the highest 20th percentile was used as a threshold. We then calculated the correlation coefficient between genes harboring a *de novo* variant and the PVNH genes, and prioritized those with a correlation coefficient with any of the 14 PVNH genes greater than the threshold value.

Supporting information

S1 Text. *De novo* variant architecture analyses.

(PDF)

S2 Text. Evaluation of the co-expression prioritization method.

(PDF)

S3 Text. Predictive model for identifying true *de novo* variants.

(PDF)

S4 Text. Supporting information references.

(PDF)

S5 Text. Control sample acknowledgments.

(PDF)

S6 Text. Specific contributions of authors, members of the Epi4K Consortium and Epilepsy Phenome/Genome Projects, and additional participants involved in this work.

(PDF)

S1 Table. *De novo* variants identified in PVNH cases.

(XLSX)

S2 Table. Phenotypes of patients and transmitting parents where available.

(PDF)

S3 Table. Homozygous and hemizygous variants identified in PVNH cases.

(XLSX)

S4 Table. Compound heterozygous mutations identified in PVNH cases.
(XLSX)

S5 Table. PVNH genes prioritized in the leave one out analysis across the three datasets.
(PDF)

S6 Table. Candidate genes prioritized based on co-regulation with known PVNH genes during the key developmental period (4–38 pcw).
(XLSX)

S7 Table. Control cohort composition.
(PDF)

S8 Table. PVNH cohort summary.
(XLSX)

S9 Table. Overview of transcriptomic datasets used for co-expression analyses.
(PDF)

S10 Table. Features used in the predictive model and their relative influence in the *de novo* variant confirmation model.
(PDF)

S1 Fig. Likelihood surface representing the genetic architecture of periventricular nodular heterotopia.
(PDF)

S2 Fig. Eigenvectors of cases (red dots) and controls (blue dots) across top three principal components from eigenstrat analyses.
(PDF)

S3 Fig. Quantile-quantile plot for gene-level association tests interrogating (A) LoF and “probably damaging” (Polyphen-2) missense variants, and (B) synonymous variants.
(PDF)

S4 Fig. Brain MRI of subjects with LoF *MAP1B* variants. The left image is coronal T1 inversion recovery (pvhit1238Pbti1), the middle image is coronal T2-weighted (pvhnd29281lw1) and the right image is axial T2 weighted (pvhcw12701bvi1). All images show periventricular nodular grey matter heterotopia maximal in the frontal regions (arrows). No polymicrogyria was seen in subjects pvhit1238Pbti1 or pvhnd29281lw1. There was possible right insular polymicrogyria in pvhcw12701bvi1, but the available images were not of sufficient quality to be conclusive. MRI image for fourth patient is provided in Fig 3.
(PDF)

S5 Fig. Cumulative proportion of correlation coefficients for 14 genes implicated in PVNH compared to 1000 randomly selected sets of 14 genes.
(PDF)

S6 Fig. Correlation matrices for prioritized genes compared to each known human PVNH query gene individually and across different periods of brain development. Pairwise Pearson’s correlations between prioritized genes harboring *de novo* mutations based on co-expression profiles and the 8 human query genes through each time period analyzed and presented across each row. Patterns of positive (+1) and reciprocal (-1) co-regulatory interactions are represented as blue and red squares, respectively. Data was derived from the Miller and Kang transcriptomic datasets with the time points in which the data was derived being separated

into specific periods as indicated in by the value below each point; 1, 4-8pcw; 2, 8-10pcw; 3, 10-13pcw; 4, 13-16pcw; 5, 16-19pcw; 6, 19-24pcw; 7, 24-38pcw; 8, birth-6M; 9, 6M-1Y; 10, 1-6Y; 11, 6-12Y; 12, 12-20Y; 13, 20-40Y; 14, 40-60Y; 15, 60Y+. pcw, weeks post conception; M, months after birth; Y, years after birth.

(PDF)

S7 Fig. Histogram of *de novo* variant confirmation probabilities.

(PDF)

Acknowledgments

We are grateful to the patients, their families, clinical research coordinators and referring physicians for participating in the various recruitment sites that provided the phenotype data and DNA samples used in this study. Additional acknowledgments recognizing contributors of the control population used in this work is provided in [S5 Text](#). Specific contributions of authors, members of the Epi4K Consortium and Epilepsy Phenome/Genome Projects, and additional contributors involved in this work are provided in [S6 Text](#).

Epi4K Consortium Members and Additional Contributors

Andrew S. Allen, Melanie Bahlo, Samuel F. Berkovic, Joshua S. Bridgers, Jamel Chelly, Brett Copeland, Patrick Cossette, Francesca Darra, Norman Delanty, Dennis Dlugos, William B. Dobyns, Evan E. Eichler, Michael P. Epstein, Catharine Freyer, Saskia Freytag, Andrew E. Fry, David B. Goldstein, Nicole G. Griffin, Renzo Guerrini, Erin L. Heinzen, Ming Hui Chen, Michael R. Johnson, Sitharthan Kamalakaran, Edwin P. Kirk, Richard J. Leventer, Daniel H. Lowenstein, Ruben Kuzniecky, Paul J. Lockhart, Colin Malone, Anthony G. Marson, George McGillivray, Caroline Mebane, Heather C. Mefford, Davide Mei, Terence J. O'Brien, Adam C. O'Neill, Ruth Ottman, Steven Petrou, Slavé Petrovski, Daniela Pilz, Annapurna Poduri, Stephen P. Robertson, Ingrid E. Scheffer, Elliott H. Sherr, Nicholas Stong, Zhong Ren, Christopher A. Walsh, Mengqi Zhang, Xiaolin Zhu.

Epilepsy Phenome/Genome Project Members

Bassel Abou-Khalil, Dina Amrom, Eva Andermann, Frederick Andermann, Samuel F. Berkovic, Judith Bluvstein, Alexis Boro, Gregory D. Cascino, Damian Consalvo, Pat Crumrine, Orrin Devinsky, Dennis Dlugos, Nathan Fountain, Catharine Freyer, Daniel Friedman, Eric Geller, Tracy Glauser, Simon Glynn, Kevin Haas, Sheryl Haut, Sucheta Joshi, Heidi Kirsch, Robert Knowlton, Eric Kossoff, Ruben Kuzniecky, Daniel H. Lowenstein, Paul V. Motika, Ruth Ottman, Juliann M. Paolicchi, Jack M. Parent, Annapurna Poduri, Ingrid E. Scheffer, Renée A. Shellhaas, Elliott H. Sherr, Jerry J. Shih, Shlomo Shinnar, Rani K Singh, Michael R. Sperling, Michael C. Smith, Joseph Sullivan, Eileen P. G. Vining, Gretchen K. Von Allmen, Peter Widdess-Walsh.

Author Contributions

Conceptualization: Erin L. Heinzen, Adam C. O'Neill, Andrew S. Allen, Jamel Chelly, Ming Hui Chen, William B. Dobyns, Renzo Guerrini, Richard J. Leventer, Annapurna Poduri, Stephen P. Robertson, Christopher A. Walsh, Epi4K Consortium.

Formal analysis: Erin L. Heinzen, Adam C. O'Neill, Xiaolin Zhu, Andrew S. Allen, Saskia Freytag, Richard J. Leventer, Annapurna Poduri, Stephen P. Robertson, Mengqi Zhang.

Supervision: Erin L. Heinzen, Andrew S. Allen, Melanie Bahlo, Jamel Chelly, Ming Hui Chen, William B. Dobyns, Renzo Guerrini, Richard J. Leventer, Annapurna Poduri, Stephen P. Robertson, Christopher A. Walsh, Epi4K Consortium.

Writing – original draft: Erin L. Heinzen, Adam C. O’Neill, Xiaolin Zhu, Andrew S. Allen, Saskia Freytag, Richard J. Leventer, Annapurna Poduri, Stephen P. Robertson.

Writing – review & editing: Erin L. Heinzen, Adam C. O’Neill, Xiaolin Zhu, Andrew S. Allen, Melanie Bahlo, Jamel Chelly, William B. Dobyns, Saskia Freytag, Renzo Guerrini, Richard J. Leventer, Annapurna Poduri, Stephen P. Robertson, Christopher A. Walsh, Mengqi Zhang, Epi4K Consortium, Epilepsy Phenome/Genome Project.

References

1. Chang BS, Katzir T, Liu T, Coriveau K, Barzillai M, Apse KA, et al. A structural basis for reading fluency: white matter defects in a genetic brain malformation. *Neurology*. 2007; 69(23):2146–54. <https://doi.org/10.1212/01.wnl.0000286365.41070.54> PMID: 18056578.
2. Barkovich AJ, Guerrini R, Kuzniecky RI, Jackson GD, Dobyns WB. A developmental and genetic classification for malformations of cortical development: update 2012. *Brain*. 2012; 135(Pt 5):1348–69. <https://doi.org/10.1093/brain/aws019> PMID: 22427329; PubMed Central PMCID: PMC3338922.
3. Fox JW, Lamperti ED, Eksioglu YZ, Hong SE, Feng Y, Graham DA, et al. Mutations in filamin 1 prevent migration of cerebral cortical neurons in human periventricular heterotopia. *Neuron*. 1998; 21(6):1315–25. Epub 1999/01/12. PMID: 9883725.
4. Sheen VL, Ganesh VS, Topcu M, Sebire G, Bodell A, Hill RS, et al. Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex. *Nat Genet*. 2004; 36(1):69–76. <https://doi.org/10.1038/ng1276> PMID: 14647276.
5. Kielar M, Tuy FP, Bizzotto S, Lebrand C, de Juan Romero C, Poirier K, et al. Mutations in *Eml1* lead to ectopic progenitors and neuronal heterotopia in mouse and human. *Nat Neurosci*. 2014; 17(7):923–33. <https://doi.org/10.1038/nn.3729> PMID: 24859200.
6. Conti V, Carabalona A, Pallese-Pocachard E, Parrini E, Leventer RJ, Buhler E, et al. Periventricular heterotopia in 6q terminal deletion syndrome: role of the *C6orf70* gene. *Brain*. 2013; 136(Pt 11):3378–94. <https://doi.org/10.1093/brain/awt249> PMID: 24056535.
7. Broix L, Jagline H, E LI, Schmucker S, Drouot N, Clayton-Smith J, et al. Mutations in the HECT domain of NEDD4L lead to AKT-mTOR pathway deregulation and cause periventricular nodular heterotopia. *Nat Genet*. 2016; 48(11):1349–58. <https://doi.org/10.1038/ng.3676> PMID: 27694961; PubMed Central PMCID: PMC5086093.
8. Cappello S, Bohringer CR, Bergami M, Conzelmann KK, Ghanem A, Tomassy GS, et al. A radial glia-specific role of RhoA in double cortex formation. *Neuron*. 2012; 73(5):911–24. <https://doi.org/10.1016/j.neuron.2011.12.030> PMID: 22405202.
9. Trimborn M, Bell SM, Felix C, Rashid Y, Jafri H, Griffiths PD, et al. Mutations in microcephalin cause aberrant regulation of chromosome condensation. *Am J Hum Genet*. 2004; 75(2):261–6. <https://doi.org/10.1086/422855> PMID: 15199523; PubMed Central PMCID: PMC1216060.
10. Oegema R, Baillat D, Schot R, van Unen LM, Brooks A, Kia SK, et al. Human mutations in integrator complex subunits link transcriptome integrity to brain development. *PLoS Genet*. 2017; 13(5): e1006809. <https://doi.org/10.1371/journal.pgen.1006809> PMID: 28542170; PubMed Central PMCID: PMC5466333.
11. Alcantara D, Timms AE, Gripp K, Baker L, Park K, Collins MS, et al. Mutations of *AKT3* are associated with a wide spectrum of developmental disorders including extreme megalencephaly. *Brain*. 2017; 140(10):2610–2622. <https://doi.org/10.1093/brain/awx203> PMID: 28969385.
12. Parrini E, Ramazzotti A, Dobyns WB, Mei D, Moro F, Veggiotti P, et al. Periventricular heterotopia: phenotypic heterogeneity and correlation with Filamin A mutations. *Brain*. 2006; 129(Pt 7):1892–906. <https://doi.org/10.1093/brain/awl125> PMID: 16684786.
13. Mandelstam SA, Leventer RJ, Sandow A, McGillivray G, van Kogelenberg M, Guerrini R, et al. Bilateral posterior periventricular nodular heterotopia: a recognizable cortical malformation with a spectrum of associated brain abnormalities. *AJNR Am J Neuroradiol*. 2013; 34(2):432–8. <https://doi.org/10.3174/ajnr.A3427> PMID: 23348762.
14. Fallil Z, Pardoe H, Bachman R, Cunningham B, Parulkar I, Shain C, et al. Phenotypic and imaging features of FLNA-negative patients with bilateral periventricular nodular heterotopia and epilepsy. *Epilepsy*

- Behav. 2015; 51:321–7. <https://doi.org/10.1016/j.yebeh.2015.07.041> PMID: 26340046; PubMed Central PMCID: PMC4594191.
15. Dubeau F, Tampieri D, Lee N, Andermann E, Carpenter S, Leblanc R, et al. Periventricular and subcortical nodular heterotopia. A study of 33 patients. *Brain*. 1995; 118 (Pt 5):1273–87. PMID: 7496786.
 16. Barkovich AJ, Kuzniecky RI. Gray matter heterotopia. *Neurology*. 2000; 55(11):1603–8. PMID: 11187088.
 17. Heinzen EL, Neale BM, Traynelis SF, Allen AS, Goldstein DB. The genetics of neuropsychiatric diseases: looking in and beyond the exome. *Annu Rev Neurosci*. 2015; 38:47–68. <https://doi.org/10.1146/annurev-neuro-071714-034136> PMID: 25840007.
 18. Epi4K Consortium, Epilepsy Phenome/Genome Project. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. *Lancet Neurol*. 2017; 16(2):135–43. [https://doi.org/10.1016/S1474-4422\(16\)30359-3](https://doi.org/10.1016/S1474-4422(16)30359-3) PMID: 28102150.
 19. Oliver KL, Lukic V, Freytag S, Scheffer IE, Berkovic SF, Bahlo M. In silico prioritization based on co-expression can aid epileptic encephalopathy gene discovery. *Neurology Genetics*. 2016; 2(1):e51. <https://doi.org/10.1212/NXG.000000000000051> PMID: 27066588; PubMed Central PMCID: PMC4817907.
 20. Oliver KL, Lukic V, Thorne NP, Berkovic SF, Scheffer IE, Bahlo M. Harnessing gene expression networks to prioritize candidate epileptic encephalopathy genes. *PLoS One*. 2014; 9(7):e102079. <https://doi.org/10.1371/journal.pone.0102079> PMID: 25014031; PubMed Central PMCID: PMC4090166.
 21. Epi4K Consortium and Epilepsy Phenome/Genome Project. De novo mutations in epileptic encephalopathies. *Nature*. 2013; 501(7466):217–21. Epub 2013/08/13. <https://doi.org/10.1038/nature12439> PMID: 23934111; PubMed Central PMCID: PMC3773011.
 22. Euro Epinomics- R. E. S. Consortium, Epilepsy Phenome/Genome Project, Epi4k Consortium. De Novo Mutations in Synaptic Transmission Genes Including *DNM1* Cause Epileptic Encephalopathies. *Am J Hum Genet*. 2014; 95(4):360–70. Epub 2014/09/30. <https://doi.org/10.1016/j.ajhg.2014.08.013> PMID: 25262651; PubMed Central PMCID: PMC4185114.
 23. Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*. 2012. [https://doi.org/10.1016/S0140-6736\(12\)61480-9](https://doi.org/10.1016/S0140-6736(12)61480-9) PMID: 23020937.
 24. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012; 74(2):285–99. Epub 2012/05/01. S0896-6273(12)00340-6. <https://doi.org/10.1016/j.neuron.2012.04.009> PMID: 22542183.
 25. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012; 485(7397):237–41. Epub 2012/04/13. <https://doi.org/10.1038/nature10945> PMID: 22495306.
 26. O’Roak BJ, Stessman HA, Boyle EA, Witherspoon KT, Martin B, Lee C, et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nature communications*. 2014; 5:5595. Epub 2014/11/25. <https://doi.org/10.1038/ncomms6595> PMID: 25418537; PubMed Central PMCID: PMC4249945.
 27. The Epi4k Consortium and Epilepsy Phenome/Genome Project. De novo mutations in epileptic encephalopathies. *Nature*. 2013. <https://doi.org/10.1038/nature12439> PMID: 23934111.
 28. Jiang Y, Han Y, Petrovski S, Owzar K, Goldstein DB, Allen AS. Incorporating Functional Information in Tests of Excess De Novo Mutational Load. *Am J Hum Genet*. 2015; 97(2):272–83. <https://doi.org/10.1016/j.ajhg.2015.06.013> PMID: 26235986; PubMed Central PMCID: PMC4573447.
 29. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536(7616):285–91. <https://doi.org/10.1038/nature19057> PMID: 27535533; PubMed Central PMCID: PMC45018207.
 30. Petrovski S, Gussow AB, Wang Q, Halvorsen M, Han Y, Weir WH, et al. The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. *PLoS Genet*. 2015; 11(9):e1005492. <https://doi.org/10.1371/journal.pgen.1005492> PMID: 26332131; PubMed Central PMCID: PMC4557908.
 31. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet*. 2014; 46(9):944–50. Epub 2014/08/05. <https://doi.org/10.1038/ng.3050> PMID: 25086666.
 32. Deciphering Developmental Disorders S. Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. 2017; 542(7642):433–8. <https://doi.org/10.1038/nature21062> PMID: 28135719.
 33. Bouquet C, Soares S, von Boxberg Y, Ravaille-Veron M, Propst F, Nothias F. Microtubule-associated protein 1B controls directionality of growth cone migration and axonal branching in regeneration of adult

- dorsal root ganglia neurons. *J Neurosci*. 2004; 24(32):7204–13. <https://doi.org/10.1523/JNEUROSCI.2254-04.2004> PMID: 15306655.
34. Gonzalez-Billault C, Avila J, Caceres A. Evidence for the role of MAP1B in axon formation. *Mol Biol Cell*. 2001; 12(7):2087–98. PMID: 11452005; PubMed Central PMCID: PMC55658.
 35. Del Rio JA, Gonzalez-Billault C, Urena JM, Jimenez EM, Barallobre MJ, Pascual M, et al. MAP1B is required for Netrin 1 signaling in neuronal migration and axonal guidance. *Curr Biol*. 2004; 14(10):840–50. <https://doi.org/10.1016/j.cub.2004.04.046> PMID: 15186740.
 36. Pisano T, Barkovich AJ, Leventer RJ, Squier W, Scheffer IE, Parrini E, et al. Peritrigonal and temporo-occipital heterotopia with corpus callosum and cerebellar dysgenesis. *Neurology*. 2012; 79(12):1244–51. <https://doi.org/10.1212/WNL.0b013e31826aac88> PMID: 22914838; PubMed Central PMCID: PMC3440449.
 37. Wieck G, Leventer RJ, Squier WM, Jansen A, Andermann E, Dubeau F, et al. Periventricular nodular heterotopia with overlying polymicrogyria. *Brain*. 2005; 128(Pt 12):2811–21. <https://doi.org/10.1093/brain/awh658> PMID: 16311271.
 38. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015; 1(6):417–25. Epub 2016/01/16. <https://doi.org/10.1016/j.cels.2015.12.004> PMID: 26771021; PubMed Central PMCID: PMC4707969.
 39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25(1):25–9. Epub 2000/05/10. <https://doi.org/10.1038/75556> PMID: 10802651; PubMed Central PMCID: PMC3037419.
 40. The Gene Ontology C. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*. 2017; 45(D1):D331–D8. Epub 2016/12/03. <https://doi.org/10.1093/nar/gkw1108> PMID: 27899567; PubMed Central PMCID: PMC5210579.
 41. Donoho D, Jin J. Higher Criticism for Detecting Sparse Heterogeneous Mixtures. *The Annals of Statistics*. 2004; 32(3):962–94.
 42. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014; 508(7495):199–206. <https://doi.org/10.1038/nature13185> PMID: 24695229; PubMed Central PMCID: PMC4105188.
 43. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011; 478(7370):483–9. <https://doi.org/10.1038/nature10523> PMID: 22031440; PubMed Central PMCID: PMC3566780.
 44. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011; 478(7370):519–23. <https://doi.org/10.1038/nature10524> PMID: 22031444; PubMed Central PMCID: PMC3510670.
 45. Quan J, Yusufzai T. The tumor suppressor chromodomain helicase DNA-binding protein 5 (CHD5) remodels nucleosomes by unwrapping. *J Biol Chem*. 2014; 289(30):20717–26. <https://doi.org/10.1074/jbc.M114.568568> PMID: 24923445; PubMed Central PMCID: PMC4110282.
 46. Flaus A, Martin DM, Barton GJ, Owen-Hughes T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res*. 2006; 34(10):2887–905. <https://doi.org/10.1093/nar/gkl295> PMID: 16738128; PubMed Central PMCID: PMC1474054.
 47. Nitarska J, Smith JG, Sherlock WT, Hillege MM, Nott A, Barshop WD, et al. A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell reports*. 2016; 17(6):1683–98. <https://doi.org/10.1016/j.celrep.2016.10.022> PMID: 27806305; PubMed Central PMCID: PMC45149529.
 48. Mei X, Sweatt AJ, Hammarback JA. Regulation of microtubule-associated protein 1B (MAP1B) subunit composition. *J Neurosci Res*. 2000; 62(1):56–64. [https://doi.org/10.1002/1097-4547\(20001001\)62:1<56::AID-JNR6>3.0.CO;2-#](https://doi.org/10.1002/1097-4547(20001001)62:1<56::AID-JNR6>3.0.CO;2-#) PMID: 11002287.
 49. Montenegro-Venegas C, Tortosa E, Rosso S, Peretti D, Bollati F, Bisbal M, et al. MAP1B regulates axonal development by modulating Rho-GTPase Rac1 activity. *Mol Biol Cell*. 2010; 21(20):3518–28. <https://doi.org/10.1091/mbc.E09-08-0709> PMID: 20719958; PubMed Central PMCID: PMC2954117.
 50. Gonzalez-Billault C, Jimenez-Mateos EM, Caceres A, Diaz-Nido J, Wandosell F, Avila J. Microtubule-associated protein 1B function during normal development, regeneration, and pathological conditions in the nervous system. *J Neurobiol*. 2004; 58(1):48–59. <https://doi.org/10.1002/neu.10283> PMID: 14598369.
 51. Jayachandran P, Olmo VN, Sanchez SP, McFarland RJ, Vital E, Werner JM, et al. Microtubule-associated protein 1b is required for shaping the neural tube. *Neural Dev*. 2016; 11:1. <https://doi.org/10.1186/s13064-015-0056-4> PMID: 26782621; PubMed Central PMCID: PMC4717579.

52. Thumkeo D, Shinohara R, Watanabe K, Takebayashi H, Toyoda Y, Tohyama K, et al. Deficiency of mDia, an actin nucleator, disrupts integrity of neuroepithelium and causes periventricular dysplasia. *PLoS One*. 2011; 6(9):e25465. <https://doi.org/10.1371/journal.pone.0025465> PMID: 21980468; PubMed Central PMCID: PMC3182227.
53. Feng Y, Chen MH, Moskowitz IP, Mendonza AM, Vidali L, Nakamura F, et al. Filamin A (FLNA) is required for cell-cell contact in vascular development and cardiac morphogenesis. *Proc Natl Acad Sci U S A*. 2006; 103(52):19836–41. <https://doi.org/10.1073/pnas.0609628104> PMID: 17172441; PubMed Central PMCID: PMC1702530.
54. Lian G, Sheen VL. Cytoskeletal proteins in cortical development and disease: actin associated proteins in periventricular heterotopia. *Front Cell Neurosci*. 2015; 9:99. <https://doi.org/10.3389/fncel.2015.00099> PMID: 25883548; PubMed Central PMCID: PMC34381626.
55. Zalfa F, Giorgi M, Primerano B, Moro A, Di Penta A, Reis S, et al. The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell*. 2003; 112(3):317–27. PMID: 12581522.
56. Lu R, Wang H, Liang Z, Ku L, O'Donnell W T, Li W, et al. The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development. *Proc Natl Acad Sci U S A*. 2004; 101(42):15201–6. <https://doi.org/10.1073/pnas.0404995101> PMID: 15475576; PubMed Central PMCID: PMC524058.
57. Moro F, Pisano T, Bernardina BD, Polli R, Murgia A, Zoccante L, et al. Periventricular heterotopia in fragile X syndrome. *Neurology*. 2006; 67(4):713–5. <https://doi.org/10.1212/01.wnl.0000230223.51595.99> PMID: 16924033.
58. Dimova PS, Kirov A, Todorova A, Todorov T, Mitev V. A novel PCDH19 mutation inherited from an unaffected mother. *Pediatr Neurol*. 2012; 46(6):397–400. <https://doi.org/10.1016/j.pediatrneurol.2012.03.004> PMID: 22633638.
59. Kosmicki JA, Samocha KE, Howrigan DP, Sanders SJ, Slowikowski K, Lek M, et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat Genet*. 2017; 49(4):504–10. <https://doi.org/10.1038/ng.3789> PMID: 28191890.
60. Costain G. Parental expression is overvalued in the interpretation of rare inherited variants. *Eur J Hum Genet*. 2015; 23(1):4–7. <https://doi.org/10.1038/ejhg.2014.64> PMID: 24755951; PubMed Central PMCID: PMC4266746.
61. Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, et al. Genome-wide patterns and properties of de novo mutations in humans. *Nat Genet*. 2015; 47(7):822–6. <https://doi.org/10.1038/ng.3292> PMID: 25985141; PubMed Central PMCID: PMC4485564.
62. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet*. 2007; 80(4):727–39. <https://doi.org/10.1086/513473> PMID: 17357078; PubMed Central PMCID: PMC1852724.
63. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*. 2013; 9(8):e1003709. Epub 2013/08/31. <https://doi.org/10.1371/journal.pgen.1003709> PMID: 23990802; PubMed Central PMCID: PMC3749936.
64. Genome of the Netherlands C. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet*. 2014; 46(8):818–25. <https://doi.org/10.1038/ng.3021> PMID: 24974849.
65. Ge YC, Dudoit S, Speed TP. Resampling-based multiple testing for microarray data analysis. *Test*. 2003; 12(1):1–77. <https://doi.org/10.1007/Bf02595811> PubMed PMID: WOS:000184058300001.
66. Doherty D, Chudley AE, Coghlan G, Ishak GE, Innes AM, Lemire EG, et al. GPSM2 mutations cause the brain malformations and hearing loss in Chudley-McCullough syndrome. *Am J Hum Genet*. 2012; 90(6):1088–93. <https://doi.org/10.1016/j.ajhg.2012.04.008> PMID: 22578326; PubMed Central PMCID: PMC3370271.
67. Mishra-Gorur K, Caglayan AO, Schaffer AE, Chabu C, Henegariu O, Vonhoff F, et al. Mutations in KATNB1 cause complex cerebral malformations by disrupting asymmetrically dividing neural progenitors. *Neuron*. 2014; 84(6):1226–39. <https://doi.org/10.1016/j.neuron.2014.12.014> PMID: 25521378; PubMed Central PMCID: PMC5024344.
68. Gil-Sanz C, Landeira B, Ramos C, Costa MR, Muller U. Proliferative defects and formation of a double cortex in mice lacking *Mlt4* and *Cdh2* in the dorsal telencephalon. *J Neurosci*. 2014; 34(32):10475–87. <https://doi.org/10.1523/JNEUROSCI.1793-14.2014> PMID: 25100583; PubMed Central PMCID: PMC4200106.
69. Schmid MT, Weinandy F, Wilsch-Brauninger M, Huttner WB, Cappello S, Gotz M. The role of alpha-E-catenin in cerebral cortex development: radial glia specific effect on neuronal migration. *Front Cell Neurosci*. 2014; 8:215. <https://doi.org/10.3389/fncel.2014.00215> PMID: 25147501; PubMed Central PMCID: PMC4124588.

70. Li Y, Wang J, Zhou Y, Li D, Xiong ZQ. Rcan1 deficiency impairs neuronal migration and causes periventricular heterotopia. *J Neurosci*. 2015; 35(2):610–20. <https://doi.org/10.1523/JNEUROSCI.1003-14.2015> PMID: 25589755.
71. Maeta K, Edamatsu H, Nishihara K, Ikutomo J, Bilasy SE, Kataoka T. Crucial Role of Rapgef2 and Rapgef6, a Family of Guanine Nucleotide Exchange Factors for Rap1 Small GTPase, in Formation of Apical Surface Adherens Junctions and Neural Progenitor Development in the Mouse Cerebral Cortex. *eNeuro*. 2016; 3(3). <https://doi.org/10.1523/ENEURO.0142-16.2016> PMID: 27390776; PubMed Central PMCID: PMC4917737.
72. Freytag S, Burgess R, Oliver KL, Bahlo M. brain-coX: investigating and visualising gene co-expression in seven human brain transcriptomic datasets. *Genome Med*. 2017; 9(1):55. <https://doi.org/10.1186/s13073-017-0444-y> PMID: 28595657; PubMed Central PMCID: PMC5465565.
73. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostatistics*. 2012; 13(3):539–52. <https://doi.org/10.1093/biostatistics/kxr034> PMID: 22101192; PubMed Central PMCID: PMC3577104.