Integration of genetic, transcriptomic, and clinical data provides insight into 16p11.2 and 22q11.2 CNV genes

Mikhail Vysotskiy[1,2,3,4], Xue Zhong[5,6], Tyne W. Miller-Fleming[5,6], Dan Zhou[5,6], Autism Working Group of the Psychiatric Genomics Consortium^, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium^, Schizophrenia Working Group of the Psychiatric Genomics Consortium^, Nancy J. Cox[5,6], Lauren A Weiss[1,2,3*]

[1]Department of Psychiatry, University of California San Francisco, San Francisco, CA, 94143 USA

[2]Institute for Human Genetics, University of California San Francisco, San Francisco, CA, 94143 USA

[3]Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, 94143, USA

[4]Pharmaceutical Sciences and Pharmacogenomics Graduate Program, University of California San Francisco, San Francisco, CA, 94143 USA

[5]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, 37232, USA.

[6]Vanderbilt Genetics Institute, Nashville, TN, 37232, USA

* Corresponding author, lauren.weiss@ucsf.edu

^ A list of consortia members is provided in the supplement

ABSTRACT

Deletions and duplications of the 16p11.2 and 22q11.2 copy number variant (CNV) regions are associated with brain-related disorders including autism spectrum disorder, schizophrenia, bipolar disorder, obesity, and intellectual disability. The contribution of individual genes in each CNV to each phenotype is unknown. We propose a novel *in silico* approach to systematically interrogate each gene's effect on brain-related disorders. As CNVs affect RNA copy number, we attempted to fine-map each CNV region by asking whether natural gene expression variation in any individual gene would increase risk for the same disorder(s). We first used large genotyped cohorts for five CNV-associated traits: autism spectrum disorder, schizophrenia, bipolar disorder, BMI, and IQ. Imputing the transcriptome from the genome, we performed association testing for CNV genes with the genotype datasets. We found individual genes within 16p11.2 associated with schizophrenia (*TMEM219, INO80E, YPEL3*), BMI (*TMEM219, SPN, TAOK2, INO80E*), and IQ (*SPN*), with independent replication. Using a conditional analysis, we showed that *INO80E* explains the schizophrenia association. Second, we used a biobank containing electronic health data and genotypes for 23,000 individuals to perform phenome-wide association studies with predicted expressions of 16p11.2 and 22q11.2 genes and over 1,500 health traits, finding eight significant gene-trait pairs. Third, we compared the medical phenome of CNV carriers to controls within an independent group of 3 million individuals. We identified CNV traits may have indicated genetic testing, other traits previously observed in CNV carriers, and novel traits. Our results provide new insight into the contribution of individual CNV genes to CNV-associated traits.

INTRODUCTION

Multi-gene copy number variants (CNVs), including a 600kb region at 16p11.2 and a 3Mb region at 22q11.2, are known causes of brain-related disorders. The 16p11.2 CNV, originally identified as a risk factor for autism spectrum disorder (ASD), has also been associated with other brain-related disorders, such as schizophrenia, bipolar disorder, intellectual disability, and obesity[1–5]. The 22q11.2 CNV, identified as the cause of DiGeorge (velocardiofacial) syndrome, is associated with ASD, schizophrenia, bipolar disorder, intellectual disability, and obesity, as well [6–11]. The effects of these two CNVs on brain-related disorders can be further subdivided into the effects of deletions vs. duplications. Some brain-related disorders are shared among carriers of deletions and duplications of the same region and others show opposite associations. For instance, ASD and intellectual disability are observed in both deletion and duplication carriers in both 16p11.2 and 22q11.2 [3–8,12–14]. Other traits are specific to one direction of the copy number change: schizophrenia and bipolar disorder are observed in 16p11.2 duplication carriers, but not deletion carriers [15]. A third category of 16p11.2 and 22q11.2-associated traits are "mirrored". 16p11.2 deletion carriers show increased rates of obesity, while duplication carriers tend to be underweight. 22q11.2 duplication carriers show reduced rates of

schizophrenia, as opposed to increased rates in deletion carriers [1,16,17]. The question of which specific genes drive brain-related traits associated with 16p11.2 or 22q11.2 CNVs remains unanswered. Identifying the gene(s) responsible for each trait will clarify the biological processes that go awry as a result of these mutations. Knowledge of the genes and processes involved would, in turn, provide opportunities to develop targeted treatments.

Three of the traditional ways to map CNV genes to disorders are identifying loss-of-function mutations in these genes, analyzing smaller subsets of the entire region, and finding mutations in animal models that are sufficient to recapitulate the phenotype. The first method was used to fine-map the 17p11.2 CNV, another CNV associated with intellectual disability, as well as other behavioral and non-behavioral traits [18,19]. By screening non-CNV-carriers who presented with a phenotype similar to deletion carriers for mutations in individual genes, *RAI1* was identified as a likely driver gene. Most of the features of the deletion syndrome, including intellectual disability, are represented in individuals who carry a defective copy of the *RAI1* gene, but do not carry the entire CNV [20]. Duplications of *Rai1* appear to explain body weight and behavior abnormalities in mouse models of 17p11.2 duplications [21]. Another example is the Williams syndrome CNV at 7q11.23, for which deletion carriers present with cardiovascular defects and atypical behavioral features, as well as intellectual disability [22,23]. The cardiac traits associated with this syndrome are present in individuals with only one functional copy of the *ELN* gene [24,25]. The second method, of finding a smaller “critical region” was used to fine-map the 17q21.31 CNV, associated with developmental delay, friendliness, and facial dysmorphisms [26,27]. By comparing patients who had similar symptoms with overlapping cytogenetic profiles, the common breakpoints of the CNV region were refined to a region containing only six genes [27]. Later, Koolen *et al* identified patients showing intellectual disability and facial dysmorphisms characteristic of this CNV, with disruptive mutations in one of the six genes, *KANSL1,* successfully fine-mapping the region from six genes to one[28]. The third method of recapitulating similar phenotypes in animal models was successful in identifying *TBX1* as a gene important for some of the physical traits involved with 22q11.2 deletions. Mice with heterozygous mutations in the *TBX1* gene show cardiac outflow tract anomalies, similar to human 22q11.2 deletion carriers [29–31]. However, it is unclear that *TBX1* is sufficient to explain brain-related disorders in 22q11.2 carriers [32,33].

The 16p11.2 and 22q11.2 CNVs have been resistant to these traditional approaches for fine-mapping of brain-related traits. To date, no highly penetrant point mutations in 16p11.2 or 22q11.2 genes have been shown to be sufficient for a brain-related disorder. No small subsets of 16p11.2 or 22q11.2 genes have been proven necessary and sufficient to cause a brain-related disorder. A subregion of 22q11.2 has been proposed to explain ASD associated with deletions [34]. As this subset of 22q11.2 contains approximately 20 genes, it is likely that further fine-mapping within this subset is possible. At 16p11.2, a subset of five deleted genes was isolated in a family with a history of ASD [35]. However, this mutation neither caused ASD in all deletion carriers, nor was responsible for ASD in some non-carrier family members. Non-human models for the 16p11.2 and 22q11.2 CNVs, as well as knockouts for individual genes are available in mouse, zebrafish, and drosophila [36–41]. Although mouse models successfully tied *TBX1* mutations at 22q11.2 to cardiac defects, model organism studies have not been

successful in consistently mapping individual genes in these CNVs to brain-related traits [29–31]. Different zebrafish studies of 16p11.2 homologs have implicated different genes as phenotype drivers, as well as showed that most were involved in nervous system development [37,38,42]. The brain-related traits associated with these CNVs are unlikely to be fully captured in model organisms. Hallucinations, a common symptom of schizophrenia, can only be identified in humans. There may be other aspects of 16p11.2 and 22q11.2 CNV biology that are human-specific. For example, mice carrying 16p11.2 duplications are obese, while obesity is associated with deletions in humans [43]. Given the insufficiency of previous approaches, new approaches for fine-mapping genes in these regions to brain-related traits are necessary.

The motivation behind our approach is that in 16p11.2 and 22q11.2 CNV carriers, variation in gene copy number is expected to lead to variation in RNA copy number (with possible downstream effects on protein product). Expression measurements in mouse or human cell lines carrying 16p11.2 and 22q11.2 deletions and duplications confirm that for nearly all genes, duplication carriers have increased expression of individual CNV genes compared to controls and deletion carriers have reduced expression compared to controls [44–49]. As the breakpoints of these CNVs do not disrupt the coding regions of individual genes, we believe that the variation in expression of one or more of the genes is the cause of pathogenicity. While these CNVs significantly disrupt gene expression levels, most genes' expression levels vary among the general population, sometimes by a factor of two or more, as studies such as the Genotype-Tissue Expression Consortium (GTEx) have shown [50–53]. This variation can be, in part, attributed to common genetic polymorphisms (expression quantitative trait loci, eQTLs). If large expression deviation in duplication and deletion carriers is a risk factor for a disorder, we hypothesize that more modest expression variation in the same genes among non-carriers will be a modest risk factor for the same disorder. This idea is analogous to the observation that common polymorphisms of small effect that are associated with a common trait also overlap with Mendelian genes for a similar trait [54–56].

Here, we perform three *in silico* studies of the effect of expression variation of each 16p11.2 and 22q11.2 gene on human traits in non-carriers (Fig 1). First, we identify genes that drive CNV-associated brain-related disorders through expression variation. We focus on five brain-related traits that are common to the 16p11.2 and 22q11.2 CNVs for which large amounts of genetic data have been amassed: ASD, schizophrenia, bipolar disorder, BMI, and IQ [57–61]. The availability of large genetic cohorts is vital for our study, as these large datasets allow us to detect modest effect sizes. While these data are missing expression information, recent tools have leveraged the heritability of gene expression allowing us to "impute" gene expression for genotyped individuals using eQTLs [62,63]. We perform association testing between imputed expression levels of genes in the two CNVs and five brain-related traits to identify gene(s) driving each trait. We found that within large GWAS cohorts, genetically-driven expression differences of at least one 16p11.2 gene were associated with schizophrenia case-control status, IQ, and BMI. Second, we leverage a biobank containing electronic health records for over 3 million individuals, of whom over 23,000 are genotyped [64]. Using a phenome-wide association study across 1,800 traits, we determine the most common medical traits occurring in deletion and duplication carriers, finding both previously known and unexpected associations.

We thirdly impute gene expression for the genotyped individuals in this health system and perform a phenome-wide association study to explore the phenotypic consequences of expression variation of individual 16p11.2 and 22q11.2 genes on the medical phenome, focusing on traits overlapping with those associated with carrier status.

RESULTS

We studied the impact of 16p11.2 and 22q11.2 genes on brain-related traits in three ways (Fig 1). First, we used large GWAS cohorts to identify genes associated with schizophrenia, bipolar disorder, autism spectrum disorder, IQ, and BMI (S1 Table) [57–61]. We performed additional fine-mapping and replication studies where possible. Then, we used electronic health records to determine which traits, both brain-related and not, are over-represented in carriers of 16p11.2 and 22q11.2 duplications and deletions [65]. Finally, we performed a phenome-wide association study between each 16p11.2 and 22q11.2 gene and the clinical traits available in the biobank.

Individual genes at 16p11.2 are associated with schizophrenia, IQ, and BMI

In order to find genes at copy number variant loci driving brain-related disorders, we performed an association analysis between imputed gene expression levels and five traits: ASD, bipolar disorder, schizophrenia, body mass index, and IQ. It has been observed that the 16p11.2 and 22q11.2 copy number variants affect expression of nearby genes, so we included genes flanking the CNV locus by 200kb in both directions [44,45,66]. Overall, we tested 83 coding and noncoding genes at or near 22q11.2 and 46 genes at or near 16p11.2 for which a predictive model was available (S2 Table, S1 Fig). As cis-eQTL are often shared among tissues, we pooled together information from all tissues in GTEx to boost our power to detect brain-related traits [63].

Two genes at 16p11.2 had predicted expression positively associated with schizophrenia (Fig 2, S3 Table): *TMEM219* ($P = 1.5\text{x}10^{-5}$) and *INO80E* ($P = 5.3\text{x}10^{-10}$). These genes had a positive direction of effect, consistent with the association between 16p11.2 duplications and schizophrenia [2]. An additional gene, *YPEL3*, was significantly associated with schizophrenia in the negative direction ($P = 4.9\text{x}10^{-6}$). For IQ, there was one strong positive association at the 16p11.2 locus (Fig 2, S3 Table): *SPN* ($P = 2.9\text{x}10^{-22}$). Intellectual disability is observed in both deletions and duplications of 16p11.2, as such there was no expected direction of effect [3,14]. Four genes showed negative association with BMI (Fig 2, S3 Table): *SPN* ($P = 6.2\text{x}10^{-18}$), *TMEM219* ($P = 2.2\text{x}10^{-5}$), *TAOK2* ($P = 8.5\text{x}10^{-11}$), and *INO80E* ($P = 1.0\text{x}10^{-7}$). We focused on genes with negative associations with BMI because, in humans, obesity is associated with deletions at 16p11.2 [1,17]. Two additional genes, *KCTD13* ($P = 9.5\text{x}10^{-6}$) and *MVP* ($P = 2.1\text{x}10^{-5}$), were significantly associated with BMI in the positive direction. No gene at 16p11.2 was significantly associated with ASD or bipolar disorder (S3 Table, S2 Fig). No individual genes at or near 22q11.2 had predicted expression significantly associated with any of the five traits (S3 Table, S3 Fig).

Follow-up analyses for association results identify INO80E as a driver of schizophrenia

We performed an additional fine-mapping study on the three genes associated with schizophrenia. Linkage disequilibrium between the eQTL SNPs in the predictive models may lead to correlation among predicted expressions for nearby genes, so it is possible not all three detected association signals are independent. The predicted expressions of *INO80E, YPEL3*, and *TMEM219* were moderately correlated (the correlation of *INO80E* to the other genes is in the range of -0.4 to 0.37 across GTEx tissues, for example), consistent with the relationships between the observed expressions of these genes (measured expression of *INO80E* is correlated to measured expression of the other genes in the range -0.36 to 0.31). In order to pick out the gene(s) driving the association signal, we used a conditional analysis approach (S4 Table). We observed that after adjusting the predicted expression of the other CNV genes for the predicted expression of *INO80E*, no gene was significantly associated with schizophrenia. However, when we adjusted the predicted expression of *INO80E* by the predicted expressions of the other two highly associated genes, *INO80E* remained significantly associated with schizophrenia ($P = 2.3\text{x}10^{-6}$). The same pattern was not observed for *TMEM219* or *YPEL3*, suggesting *INO80E* explains the entire 16p11.2 signal for schizophrenia. BMI was also a trait where multiple signals at 16p11.2 were observed, however, as we lacked a well-powered individual-level BMI dataset, we could not perform the equivalent conditional analysis.

Replication of schizophrenia, BMI, and IQ results

For replication, we used a large cohort for which GWAS summary statistics were available for multiple brain-related traits (S1 Table) [67]. The predicted expression of *INO80E* was positively correlated with having an ICD10 diagnosis of schizophrenia (ICD10: F20, 198 cases, $P = 0.04$, S5 Table). The UK Biobank GWAS of BMI is highly inflated in the 16p11.2 region, and nearly every 16p11.2 gene shows significant association at the previously used threshold ($P < 7.6\text{x}10^{-5}$). Using a permutation-based approach, we adjusted the significance threshold to $5.7\text{x}10^{-24}$. All four genes from the discovery analysis replicated (S5 Table): *SPN* ($P = 1.3\text{x}10^{-24}$), *TMEM219* ($P = 1.6\text{x}10^{-29}$), *TAOK2* ($P = 1.2\text{x}10^{-26}$), and *INO80E* ($P = 7.8\text{x}10^{-25}$). We were not able to replicate the IQ result in the UK Biobank, because the UK Biobank sample overlapped with our discovery GWAS. Instead, we found that *SPN* replicated in the BioVU biobank phenotype *mild cognitive impairment* ($P = 0.048$, S5 Table).

Phenome-wide association studies identify previously known and novel traits associated with 16p11.2 and 22q11.2 carrier status

Electronic health records provide a way to interrogate phenotypic variation in large and diverse populations. We queried the de-identified electronic health records for 3.1 million patients at VUMC [65]. Manual chart review was performed to identify potential carriers of 16p11.2 deletions, 16p11.2 duplications, 22q11.2 deletions, and 22q11.2 duplications. We performed a phenome-wide association study (PheWAS) between 16p11.2 and 22q11.2 deletion/duplication carriers and controls against 1,796 medical phenotype codes (Figs 3 and 4) [68,69]. Significant PheWAS associations fell into three major categories: (1) known CNV clinical features, including

possible reasons for the referral of the patient for genetic testing (i.e. neurodevelopmental concerns, epilepsy, congenital heart defects), (2) secondary CNV features known to be present in carriers but unlikely to be a primary reason for referral for genetic testing, (3) novel diagnoses not previously reported (S6 Table).

16p11.2 deletion carrier status was associated with developmental diagnoses (Fig 3): *lack of normal physiological development* (P= 4.6x10^{-18}), *developmental delays and disorders* (P= 6.9x10^{-10}), *delayed milestones* (P= 1.9x10^{-11}) [3]. In addition, 16p11.2 deletion carrier status was associated with *autism* (P= 1.7x10^{-10}) and *mental retardation* (P= 1.6x10^{-12}) [5]. Epilepsy has been observed in 16p11.2 duplication and deletion carriers, and 16p11.2 deletions were associated with *transient alterations of awareness* (P= 1.9x10^{-08}) [3,70]. The association between 16p11.2 deletions and *anterior horn cell disease* (P= 4.6x10^{-6}), a diagnosis encompassing several neuromuscular disorders of the spinal cord, was novel.

16p11.2 duplication carrier status was similarly associated with developmental diagnoses (Fig 3): *lack of normal physiological development* (P= 8.9x10^{-15}), *developmental delays and disorders* (P= 2.9x10^{-13}), *delayed milestones* (P= 1.3x10^{-12}), *autism* (P= 1.8x10^{-12}), and *mental retardation* (P= 2.7x10^{-7}) [3,5]. Like deletions, 16p11.2 duplications were associated with an epilepsy-related diagnosis of *convulsions* (P= 4.3x10^{-8}) [3,70]. 16p11.2 duplication status was also significantly associated with novel dermatologic (*vascular disorders of the skin* (P= 6.0x10^{-9}) and *telogen effluvium* (P= 2.9x10^{-5})) and genitourinary (*retention of urine* (P= 3.2x10^{-4})) diagnoses.

For 22q11.2 deletion carriers, the canonical associated features were cardiac defects such as *cardiomegaly* (P = 1.3x10^{-256}) and *cardiac shunt/heart septal defects* (P = 8.6x10^{-285}) (Fig 4) [6,7]. Other highly associated diagnoses were developmental: *lack of normal physiological development* (P = 1.8x10^{-46}), *developmental delays and disorders* (P = 1.0x10^{-28}), *delayed milestones* (P = 1.1x10^{-10}) [6,7]. The secondary known traits for 22q11.2 deletion carriers included *immunity deficiency* (P < 10^{-284}), and *disorders involving the immune mechanism* (P < 10^{-284}). Previously, it has been reported that 50% of 22q11.2 deletion carriers have T-cell dysfunction and 17% have humoral dysfunction [7]. Very few traits over-represented in 22q11.2 deletion carriers were novel; these included *hyperpotassemia* (P = 3.5x10^{-10}) and *alkalosis* (P = 3.4x10^{-7}).

22q11.2 duplication carrier status was also associated with developmental diagnoses (Fig 4): *delayed milestones* (P = 1.6x10^{-13}), l*ack of normal physiological development* (P = 1.4x10^{-12}), *pervasive developmental disorders* (P = 1.5x10^{-6}) [8]. 22q11.2 duplication status was associated with cardiac phenotypes such as *rheumatic disease of the heart valves* (P = 3.2x10^{-7}) and *cardiomegaly* (P = 4.8x10^{-5}). Cardiac features have not as often been reported in 22q11.2 duplication carriers compared to 22q11.2 deletion carriers [8]. The diagnosis of *chronic inflammatory pelvic disease* (P = 7.4x10^{-5}) appears to be novel.

Phenome-wide association studies identify medical consequences of expression variation in 16p11.2 and 22q11.2 genes

In order to determine the effect of 16p11.2 and 22q11.2 gene expression on medical traits in our biobank, we generated predicted expression for CNV genes, as in the initial analysis, for the 23,000 individuals genotyped in BioVU. We tested the predicted expression of each gene within or flanking the CNVs for association with 1,534 medical phenotypic codes. There were five phenome-wide significant ($P < 3.3\text{x}10^{-5}$) gene-trait associations at 16p11.2: *SULT1A4* with *valvular heart disease/ heart chambers* ($P = 6.5\text{x}10^{-6}$), *TBC1D10B* with *varicose veins of lower extremity, symptomatic* ($P = 3.2\text{x}10^{-5}$), *YPEL3* with *benign neoplasm of breast* ($P = 2.3\text{x}10^{-7}$), *disorders of urea cycle metabolism* ($P = 3.5\text{x}10^{-7}$), and *other specified benign mammary dysplasias* ($P = 1.2\text{x}10^{-5}$). There were three phenome-wide significant gene-trait associations at 22q11.2: *SLC25A1* with *disorders of bilirubin excretion* ($P = 3.8\ \text{x}10^{-6}$), *SNAP29* with *angiodysplasia of intestine (without mention of hemorrhage)* ($P = 7.5\ \text{x}\ 10^{-6}$), *TXNRD2* with *upper gastrointestinal congenital anomalies* ($P = 2.1\text{x}10^{-5}$). None of these traits were associated with CNV carrier status. As few gene-trait pairs reached phenome-wide significance, we considered traits that did not meet the significance threshold in our analysis, but were in the top 1% of phenotypic associations for a given gene. Tables 1 and 2 contain selected gene-trait pairs from this analysis, and S7 Table contains the top 1% of phenotypic associations for each gene studied. Gene-trait pairs that were selected for inclusion in Tables 1 and 2 met at least one of the following three criteria: (a) genes associated with traits which were also associated with case-control status for the respective CNV, (b) genes associated with brain-related traits, (c) gene-trait pairs reaching phenome-wide significance.

Table 1. Selected 16p11.2 gene associations with GWAS and PheWAS traits

Gene	*GWAS Trait*	*PheWAS trait*	*P-value*[A]	*Reason for inclusion*[B]
NPIPB11		Chromosomal anomalies	0.0033	del/dup
NPIPB11		Chromosomal anomalies and genetic disorders	0.0062	del/dup
SULT1A4		Valvular heart disease/ heart chambers	$6.5\text{x}10^{-6}$*	dup, PheWS
SLC7A5P1		Other and unspecified disorders of the nervous system	0.0014	del, brain-related
SLC7A5P1		Other persistent mental disorders due to conditions classified elsewhere	0.0033	brain-related
SPN	BMI[r]		$6.2\text{x}10^{-18}$	brain-related
SPN	IQ[r]		$2.9\text{x}10^{-22}$	brain-related
SPN		Other conditions of brain	0.0059	del
SPN		Generalized anxiety disorder	0.00018	brain-related
SPN		Swelling, mass, or lump in head and neck [Space-	0.0034	brain-related

		occupying lesion, intracranial NOS]		
QPRT		Primary/intrinsic cardiomyopathies	0.002	dup
PRRT2		Dementia with cerebral degenerations	0.00093	brain-related
MVP		Valvular heart disease/ heart chambers	0.00023	dup
MVP		Cardiac pacemaker/device in situ	0.0032	dup
MVP		Heart failure with reduced EF [Systolic or combined heart failure]	0.0018	dup
MVP		Substance addiction and disorders	0.0053	brain-related
CDIPT		Symptoms involving head and neck	0.0025	brain-related
CDIPT		Hearing loss	0.0035	del/dup
CDIPT		Other degenerative diseases of the basal ganglia	0.0039	dup, brain-related
SEZ6L2		Chromosomal anomalies	0.01	del/dup
SEZ6L2		Cardiac arrest and ventricular fibrillation	0.0059	del
TMEM219	Schizophrenia[ns]		$1.5x10^{-5}$	brain-related
TMEM219	BMI[r]		$2.2x10^{-5}$	brain-related
ASPHD1		Symptoms involving head and neck	0.0023	brain-related
ASPHD1		Other degenerative diseases of the basal ganglia	0.0058	dup, brain-related
TAOK2	BMI[r]		$8.5x10^{-11}$	
TAOK2		Valvular heart disease/ heart chambers	0.00084	dup
TAOK2		Other and unspecified congenital anomalies	0.0025	dup
DOC2A		Extrapyramidal disease and abnormal movement disorders	0.0045	del, brain-related
DOC2A		Osteoporosis NOS	0.0051	del
INO80E	Schizophrenia[r]		$5.3x10^{-10}$	brain-related
INO80E	BMI[r]		$4.2x10^{-6}$	brain-related
FAM57B		Valvular heart disease/ heart chambers	0.004	dup
PPP4C		Alcoholic liver damage	0.0013	brain-related

PPP4C		Alcohol-related disorders	0.0023	brain-related
PPP4C		Other specified peripheral vascular diseases	0.0057	dup
YPEL3	Schizophrenia[ns]		4.9×10^{-6}	brain-related
YPEL3		Valvular heart disease/ heart chambers	0.00048	dup
YPEL3		Cardiac pacemaker/device in situ	0.00095	dup
YPEL3		Cardiac arrest	0.0016	del
YPEL3		Benign neoplasm of breast	2.3×10^{-7}*	PheWS
YPEL3		Disorders of urea cycle metabolism	3.5×10^{-7}*	PheWS
YPEL3		Other specified benign mammary dysplasias	1.1×10^{-5}*	PheWS
GDPD3		Valvular heart disease/ heart chambers	0.00059	dup
GDPD3		Somatoform disorder	0.0061	brain-related
MAPK3		Chromosomal anomalies	0.01	del/dup
MAPK3		Other and unspecified disorders of the nervous system	0.00089	del, brain-related
SLX1A		Arteritis NOS	0.0042	dup
SLX1A		Disorders of sacrum	0.009	del
SULT1A3		Mental retardation[r]	0.0019	del/dup, brain-related
SULT1A3		Other and unspecified disorders of the nervous system	0.0058	del, brain-related
SULT1A3		Other specified peripheral vascular diseases	0.007	dup
CD2BP2		Other and unspecified disorders of the nervous system	0.0034	del, brain-related
TBC1D10B		Varicose veins of lower extremity, symptomatic	3.2×10^{-5}*	PheWS
ZNF48		Unspecified diffuse connective tissue disease	0.0088	dup
ZNF48		Sleep apnea	0.0012	del
SEPT1		Respiratory abnormalities	0.0017	dup

[A]Traits in the top 1% of PheWAS results for each gene are included here, regardless of phenome-wide significance

[B]Possible reasons for inclusion are (1) del, dup, or del/dup: trait is over-represented in 16p11.2 deletion carriers, duplication carriers, or both (at a 5% false discovery rate); (2) brain-related trait; (3) PheWS, phenome-wide significant

* Phenome-wide significant ($P < 3.3x10^{-5}$)
[r] This association replicated ($P < 0.1$) in an independent cohort
[ns]not significant after conditional analysis

Table 2. Selected 22q11.2 gene associations with PheWAS traits

Gene	*PheWAS trait*	*P-value*[A]	*Reason for inclusion*[B]
TUBA8	Pericarditis	0.0015	del
TUBA8	Delirium dementia and amnestic and other cognitive disorders	0.0022	del, brain-related
TUBA8	Aortic aneurysm	0.0029	del
TUBA8	Other conditions of brain, NOS	0.004	del, brain-related
TUBA8	Dyspepsia and other specified disorders of function of stomach	0.0052	del
TUBA8	Aphasia	0.0044	brain-related
USP18	Other and unspecified congenital anomalies	0.0068	del/dup
USP18	Chromosomal anomalies	0.007	del/dup
USP18	Lack of normal physiological development, unspecified	0.0078	del/dup
USP18	Developmental delays and disorders	0.0094	del/dup, brain-related
USP18	Dysphagia	0.0037	del
USP18	Delay in sexual development and puberty NEC	0.0062	del
DGCR6	Other conditions of brain	0.0019	del, brain-related
PRODH	Peritoneal adhesions (postoperative) (postinfection)	0.0066	del/dup
PRODH	Delayed milestones	0.0076	del/dup, brain-related
PRODH	Hemorrhage NOS	0.00055	del
PRODH	Respiratory complications	0.0012	del
PRODH	Nausea and vomiting	0.002	del
PRODH	Autoimmune hemolytic anemias	0.0063	del
DGCR9	Ascites (non malignant)	$8.0x10^{-4}$	del
DGCR9	Other disorders of stomach and duodenum	0.0013	del
CA15P1	Hemorrhage or hematoma complicating a procedure	0.0017	del
CA15P1	Complication due to other implant and internal device	0.002	del
CA15P1	Other congenital musculoskeletal anomalies	0.0022	del
CA15P1	Nausea and vomiting	0.0092	del

CA15P1	Obsessive-compulsive disorders	0.0024	brain-related
DGCR2	Nausea and vomiting	0.0016	del
DGCR2	Otorrhea	0.0064	del
DGCR2	Deep vein thrombosis [DVT]	0.0078	del
DGCR2	Other disorders of stomach and duodenum	0.012	del
TSSK1A	Other congenital musculoskeletal anomalies	0.00075	del
TSSK1A	Congestive heart failure; nonhypertensive	0.0097	del
SLC25A1	Dyspepsia and other specified disorders of function of stomach	0.0027	del
SLC25A1	Disorders of bilirubin excretion	3.8×10^{-6}*	PheWS
SLC25A1	Hammer toe (acquired)	0.0051	del
C22orf39	Congenital anomalies of peripheral vascular system	0.00089	del
C22orf39	Impacted cerumen	0.0013	del
C22orf39	Iatrogenic hypotension	0.0055	del
C22orf39	Symptoms involving cardiovascular system	0.0072	del
C22orf39	Schizophrenia	0.0094	del, brain-related
MRPL40	Diseases of the oral soft tissues, excluding lesions specific for gingiva and tongue	8.7×10^{-5}	del
MRPL40	Impacted cerumen	0.00037	del
MRPL40	Other congenital musculoskeletal anomalies	0.0019	del
CDC45	Otorrhea	3.3×10^{-5}	del
CDC45	Congenital anomalies of peripheral vascular system	0.0018	del
CDC45	Tobacco use disorder	0.0074	brain-related
CLDN5	Poisoning/allergy of sulfonamides	9.0×10^{-4}	del
SEPT5	Endocarditis	0.0056	del
TBX1	Aortic ectasia	0.0031	del
TBX1	Renal failure	0.004	del
GNB1L	Symptoms involving cardiovascular system	0.008	del
TXNRD2	Upper gastrointestinal congenital anomalies	2.1×10^{-5}*	del; PheWS
TXNRD2	Digestive congenital anomalies	7.0×10^{-4}	del
TXNRD2	Curvature of spine	0.0027	del
TXNRD2	Tension headache	3.3×10^{-5}	brain-related
TXNRD2	Other mental disorder	0.0048	brain-related

TXNRD2	Conduct disorders	0.0052	brain-related
COMT	Pericarditis	0.0025	del
COMT	Stricture of artery	0.0033	del
COMT	Complication due to other implant and internal device	0.0085	del
COMT	Iron deficiency anemia secondary to blood loss (chronic)	0.0081	del
COMT	Other persistent mental disorders due to conditions classified elsewhere	0.0091	brain-related
ARVCF	Speech and language disorder[r]	0.0011	del, brain-related
ARVCF	Diseases of the oral soft tissues, excluding lesions specific for gingiva and tongue	0.0023	del
ARVCF	Curvature of spine	0.0061	del
ARVCF	Pericarditis	0.0073	del
TANGO2	Supraventricular premature beats	0.00034	del
TANGO2	Other pulmonary inflamation or edema	0.00069	del
TANGO2	Psychogenic disorder	0.0083	brain-related
DGCR8	Abdominal pain	0.0013	del
DGCR8	Ventricular fibrillation and flutter	0.012	del
DGCR8	Psychogenic disorder	0.0052	brain-related
DGCR8	Suicidal ideation[r]	0.012	brain-related
TRMT2A	Acute vascular insufficiency of intestine	0.0091	del
TRMT2A	Psychogenic disorder	0.007	brain-related
RANBP1	Respiratory arrest	0.00084	del
RANBP1	Acute vascular insufficiency of intestine	0.0012	del
RANBP1	Hypotension	0.0014	del
RANBP1	Other infectious and parasitic diseases	0.0057	del
RANBP1	Kyphoscoliosis and scoliosis	0.0078	del
ZDHHC8	Congenital anomalies of face and neck	0.0059	del
LINC00896	Cardiac complications, not elsewhere classified	0.0014	del
LINC00896	Speech and language disorder[r]	0.011	del, brain-related
LINC00896	Aneurysm and dissection of heart	0.013	del

RTN4R	Allergy/adverse effect of penicillin	0.0072	del/dup
RTN4R	Hemorrhage or hematoma complicating a procedure	0.0094	del
DGCR6L	Bacteremia	0.0092	del
USP41	Cardiac congenital anomalies	0.0073	del/dup
USP41	Cardiac shunt/ heart septal defect	0.0085	del/dup
USP41	Hypercalcemia	0.0023	del
USP41	Heart valve replaced	0.0026	del
USP41	Cardiac complications, not elsewhere classified	0.0035	del
USP41	Other CNS infection and poliomyelitis	0.0092	del
USP41	Eustachian tube disorders	0.011	del
ZNF74	Chromosomal anomalies and genetic disorders	0.013	del/dup
ZNF74	Diverticulum of esophagus, acquired	0.013	del
SCARF2	Nystagmus and other irregular eye movements	0.0072	dup
SCARF2	Congenital anomalies of genital organs	8.1×10^{-5}	del
SCARF2	Congenital anomalies of male genital organs	9.3×10^{-5}	del
SCARF2	Sprains and strains	0.0066	del
SCARF2	Obsessive-compulsive disorders	0.0025	brain-related
KLHL22	Rheumatic disease of the heart valves	0.0061	del/dup
KLHL22	Ill-defined descriptions and complications of heart disease	0.0012	del
KLHL22	Symptoms involving head and neck	0.011	brain-related
MED15	Epilepsy, recurrent seizures, convulsions	0.0052	del/dup, brain-related
MED15	Other disorders of circulatory system	0.0026	del
MED15	Convulsions	0.0036	del, brain-related
MED15	Hemorrhage NOS	0.0038	del
MED15	Substance addiction and disorders	0.0034	brain-related
MED15	Bariatric surgery[ns]	0.0012	brain-related
SMPD4P1	Hearing loss	0.0013	del/dup
SMPD4P1	Other and unspecified congenital anomalies	0.0049	del/dup
SMPD4P1	Chronic inflammatory pelvic disease	0.0058	dup

SMPD4P1	Peripheral vascular disease	0.0026	del
SMPD4P1	Circulatory disease NEC	0.0035	del
SMPD4P1	Substance addiction and disorders	0.006	brain-related
ABHD17AP4	Defibrination syndrome	0.002	del
ABHD17AP4	Encounter for long-term (current) use of anticoagulants	0.0022	del
ABHD17AP4	Abnormal involuntary movements	0.0094	del, brain-related
ABHD17AP4	Osteoarthrosis involving more than one site, but not specified as generalized	0.0029	del
POM121L4P	Psychophysical visual disturbances	0.00096	del, brain-related
POM121L4P	Immunity deficiency	0.0097	del
PI4KA	Other specified congenital anomalies of kidney	0.0033	del/dup
PI4KA	Other disorders of intestine	0.0046	del
PI4KA	Acute bronchitis and bronchiolitis	0.011	del
SERPIND1	Primary pulmonary hypertension	0.0092	del/dup
SNAP29	Paralysis/spasm of vocal cords or larynx	0.00075	del
SNAP29	Mobitz II AV block	0.0029	del
SNAP29	Angiodysplasia of intestine (without mention of hemorrhage)	7.5×10^{-6}*	PheWS
CRKL	Paralysis/spasm of vocal cords or larynx	0.0073	del
CRKL	Posttraumatic stress disorder	0.0067	brain-related
AIFM3	Nausea and vomiting	0.00064	del
AIFM3	Hypovolemia	0.0018	del
AIFM3	Hypotension NOS	0.002	del
AIFM3	Congenital anomalies of genital organs	0.0018	del
AIFM3	Abdominal pain	0.0026	del
LZTR1	Disease of tricuspid valve	0.002	del/dup
LZTR1	Cardiomegaly	0.0033	del/dup
LZTR1	Septicemia	0.00098	del
LZTR1	Bacterial infection NOS	0.0028	del
LZTR1	Congenital anomalies of limbs	0.0036	del
LZTR1	Acute vascular insufficiency of intestine	0.0031	del
LZTR1	Hypotension NOS	0.0074	del
LZTR1	Conduct disorders[ns,r]	0.00075	brain-related
THAP7	Heartburn	0.0067	dup

THAP7	Abdominal pain	0.0026	del
THAP7	Nausea and vomiting	0.0045	del
THAP7	Vascular dementia	0.0041	brain-related
TUBA3FP	Aortic aneurysm	0.011	del
TUBA3FP	Substance addiction and disorders	0.0017	brain-related
P2RX6	Nervous system congenital anomalies	0.0062	del/dup, brain-related
P2RX6	Adjustment reaction	0.011	brain-related
P2RX6	Congenital anomalies of limbs	0.0028	del
P2RX6	Congenital anomalies of genital organs	0.007	del
SLC7A4	Dementias	0.00014	del, brain-related
SLC7A4	Disorders of parathyroid gland	0.0014	del
SLC7A4	Delirium dementia and amnestic and other cognitive disorders	0.0024	del, brain-related
SLC7A4	Hyperparathyroidism	0.0045	del
SLC7A4	Chronic vascular insufficiency of intestine	0.0046	del
SLC7A4	Apnea	0.0055	del
P2RX6P	Abnormal movement	0.0078	del/dup, brain-related
P2RX6P	Conduct disorders[ns]	0.0014	brain-related
P2RX6P	Somatoform disorder	0.0021	brain-related
P2RX6P	Congenital anomalies of genital organs	0.0017	del
P2RX6P	Mood disorders[r]	0.003	brain-related
P2RX6P	Depression	0.0059	brain-related
BCRP2	Cardiac arrest and ventricular fibrillation	0.00018	del
BCRP2	Speech and language disorder[r]	0.00042	del, brain-related
BCRP2	Lack of coordination	0.0028	del, brain-related
GGT2	Hypovolemia	0.00082	del
HIC2	Other disorders of stomach and duodenum	0.00089	del
HIC2	Diseases of hard tissues of teeth	0.0016	del
HIC2	Other diseases of the teeth and supporting structures	0.0015	del
HIC2	Dental caries	0.0017	del
HIC2	Fracture of upper limb	0.0029	del
HIC2	Bariatric surgery[r]	0.0034	brain-related
TMEM191C	Other congenital musculoskeletal anomalies	0.0011	del
TMEM191C	Neural tube defects	0.0087	dup, brain-related

TMEM191C	Second degree AV block	0.0036	del
TMEM191C	Complication due to other implant and internal device	0.0053	del
TMEM191C	First degree AV block	0.011	del
PI4KAP2	Other disorders of intestine	0.0078	del
PI4KAP2	Nystagmus and other irregular eye movements	0.0033	dup
UBE2L3	Cardiac congenital anomalies	0.014	del/dup
UBE2L3	Premature beats	0.0025	del
UBE2L3	Coagulation defects	0.0082	del
UBE2L3	Pneumococcal pneumonia	0.013	del
YDJC	Hypotension NOS	0.01	del
YDJC	Otitis media and Eustachian tube disorders	0.0056	del
YDJC	Substance addiction and disorders	0.012	brain-related
YDJC	Suicidal ideation	0.013	brain-related
CCDC116	Disorders of mineral metabolism	8.7×10^{-5}	del
CCDC116	Disorders of calcium/phosphorus metabolism	0.00018	del
CCDC116	Hypotension NOS	0.0015	del
CCDC116	Other congenital musculoskeletal anomalies	0.0018	del
CCDC116	Posttraumatic stress disorder	0.0032	brain-related
SDF2L1	Chronic inflammatory pelvic disease	0.0034	dup
SDF2L1	Other disorders of peritoneum	0.0069	del
PPIL2	Nausea and vomiting	0.00024	del
PPIL2	Hypovolemia	0.0026	del
PPIL2	Other disorders of stomach and duodenum	0.0041	del
PPIL2	Pain in joint	0.0084	del
PPIL2	Conduct disorders[ns,r]	0.0025	brain-related

[A]Traits in the top 1% of PheWAS results for each gene are included here, regardless of phenome-wide significance

[B]Possible reasons for inclusion are (1) del, dup, or del/dup: trait is over-represented in 22q11.2 deletion carriers, duplication carriers, or both (at a 5% false discovery rate); (2) brain-related trait; (3) PheWS: phenome-wide significant

* Phenome-wide significant ($P < 3.3 \times 10^{-5}$)

[r] This association replicated ($P < 0.1$) in an independent cohort

[ns]not significant after conditional analysis

dup trait: trait over-represented in 22q11.2 duplication carriers (5% false discovery rate); del trait: trait over-represented in 22q11.2 deletion carriers; del/dup trait: trait over-represented in both 22q11.2 deletion and duplication carriers; PheWS: phenome-wide significant

Although no associations with brain-related traits reached phenome-wide significance, there were several genes with such traits in their top 1% of association results. These included traits overlapping with the five we had originally chosen. At 16p11.2 there was one such association (Table 1): *SULT1A3* with *mental retardation* (P = 0.0019) . Although this did not overlap with our results in the IQ analysis, *SULT1A3* had a p-value of 0.064 in the UK Biobank IQ replication set. At 22q11.2, there were three associations with canonical 22q11.2-associated brain-related traits (Table 2): *C22orf39* with *schizophrenia* (P = 0.0094), *P2RX6P* with *mood disorders* (P = 0.003), *USP18* with *developmental delays and disorders* (P = 0.0094). We did not identify any of these associations in the ascertained datasets. Other genes at 22q11.2 were associated with diagnoses that are related to or are symptoms of the five canonical brain-related traits we originally tested (ASD, schizophrenia, bipolar disorder, obesity, and intellectual disability) (Table 2): *speech and language disorder* (*ARVCF* [P = 0.0011], *BCRP2* [P = 0.00042], *LINC00896* [P = 0.011]), *conduct disorder* (*LZTR1* [P = 0.00075] , *P2RX6P* [P = 0.0014], *TXNRD2* [P = 0.0052], *PPIL2* [P = 0.0025]), *suicidal ideation* (*DGCR8* [P = 0.012], *YDJC* [P = 0.013]), *bariatric surgery* (*HIC2* [P = 0.0034], *MED15* [P = 0.0012]), *depression* (*P2RX6P, P* = 0.0059). Nine of these 14 per-gene PheWAS associations replicated in our original discovery or replication datasets at a $P < 0.1$ level (S5 Table).

As predicted expressions may be correlated between nearby genes, multiple genes can share a PheWAS trait association due to correlation alone. We selected several notable traits that appeared in multiple genes to test for independence, in the same way as we had tested for independence between three schizophrenia-associated genes in our GWAS cohort analysis (S5 Table). *Bariatric surgery* was a top association for two genes *MED15* (P = 0.0012) and *HIC2* (P = 0.0034) (Table 2). After conditioning each gene on the other, the association between *MED15* and *bariatric surgery* weakened (P = 0.0051), while the association between *HIC2* and *bariatric surgery* remained nearly identical (P = 0.0033), suggesting that the association between *MED15* and this trait was driven by correlation. Six genes at 16p11.2 were associated with *valvular heart disease/ heart chambers*: *FAM57B* (P = 0.004), *GDPD3* (P = 0.00058), *MVP* (P = 0.00023), *SULT1A4* ($P = 6.5\times10^{-6}$), *YPEL3* (P = 0.00048), *TAOK2* (P = 0.00084) (Table 1). *TAOK2* did not have predictive models in any tissues overlapping tissues with predictive models for the other five genes, so it was excluded from conditional analysis. *SULT1A4* and *FAM57B* had correlated predicted expression in brain-hippocampus, while *YPEL3* and *MVP* had correlated predicted expression in thyroid (the correlation between *GDPD3* and either of these genes in thyroid was much lower than the correlation between *YPEL3* and *MVP*). The association between *SULT1A4 and valvular heart disease/ heart chambers* remained nearly the same ($P = 6.4\times10^{-6}$), as did *FAM57B* (P = 0.004), as did *MVP* (P = 0.00018). Thus, we are unable to narrow down the list of potential 16p11.2 drivers of *valvular heart disease/ heart chambers*. *Speech and language disorder* is a top PheWAS trait for *ARVCF* (P = 0.0011), *BCRP2* (P = 0.00042), and *LINC00896* (P = 0.011) (Table 2). *LINC00896* did not have

predictive models in any tissues overlapping tissues with predictive models for the other five genes, so it was excluded from conditional analysis. After conditioning *ARVCF* and *BCRP2* on each other, both the association between *ARVCF* and *speech and language disorder* (*P* = 0.0011), and the association between *BCRP2* and *speech and language disorder* (*P* = 0.00068) remained consistent. Four genes at 16p11.2 had *other and unspecified disorders of the nervous system* as top PheWAS associations: *CD2BP2* (*P* = 0.0034), *MAPK3* (*P* = 0.00089), *SLC7A5P1* (*P* = 0.0014), *SULT1A3* (*P* = 0.0058) (Table 1). *MAPK3* did not have predictive models in any tissues overlapping tissues with predictive models for the other five genes, so was excluded from conditional analysis. After conditional analysis, none of the other three genes, *C2DBP2* (*P* = 0.0042), *SLC7A5P1* (*P* = 0.0013), *SULT1A3* (*P* = 0.0056) showed a noticeable reduction in association with *other and unspecified disorders of the nervous system*. The final frequently-occurring PheWAS trait we tested was *conduct disorders*, a top trait for *LZTR1* (*P* = 0.00075), *P2RX6P* (*P* = 0.0014), *PPIL2* (*P* = 0.0025), and *TXNRD2* (*P* = 0.0052) (Table 2). After conditioning each gene on the other three, *TXNRD2* remained equally associated with *conduct disorders*, while the other three genes, *LZTR1* (*P* = 0.55), *P2RX6P* (*P* = 0.0043), and *PPIL2* (*P* = 0.0071) showed reduced signal, providing evidence that *TXNRD2* may be the independent driver of this trait.

Genes in 16p11.2 and 22q11.2 are associated with traits that are also over-represented in carriers

We also note the traits identified in the PheWAS of 16p11.2 or 22q11.2 CNV carrier status which overlapped with a top PheWAS result of an individual gene in the respective region, and whether the individual gene has been implicated in driving similar traits in the literature. First, we list the 16p11.2 deletion-associated traits overlapping with at least one 16p11.2 single-gene association result (Table 1). *SLX1A* (*P* = 0.009) was associated with *disorders of sacrum*. *CD2BP2* (*P* = 0.0034)*, MAPK3* (*P* = 0.00089)*, SLC7A5P1* (*P* = 0.0014)*, SULT1A3* (*P* = 0.0054) were associated with *other and unspecified disorders of the nervous system* (which includes ICD9 diagnoses such as *secondary parkinsonism*, *dural tear*, *cerebrospinal fluid rhinorrhea*). *SPN* (*P* = 0.0059) was associated with the 16p11.2 deletion-associated diagnosis of *other conditions of brain (*which includes ICD9 diagnoses such as *anoxic brain damage, temporal sclerosis, brain death)*. *SPN* was associated with BMI and IQ in our initial study. None of these gene-trait associations have been previously reported.

Next, we found 16p11.2 genes overlapping with at least one 16p11.2 duplication-associated trait (Table 1). *SLX1A* was associated with the secondary cardiac trait of *arteritis NOS* (*P* = 0.0042) [71]*. QPRT* was associated with *primary/intrinsic cardiomyopathies* (*P* = 0.002)*.* Six genes, *FAM57B* (*P* = 0.004)*, GDPD3* (*P* = 0.00059)*, MVP* (*P* = 0.00023)*, SULT1A4* (*P* = 6.5×10^{-6})*, TAOK2* (*P* = 0.00084)*, YPEL3* (*P* = 0.00048) were associated with *valvular heart disease/ heart chambers*, and the association with *SULT1A4* reached phenome-wide significance. *TAOK2* was associated with *other and unspecified congenital anomalies* (*P* = 0.0025), the flanking gene *ZNF48* was associated with *unspecified diffuse connective tissue disease* (*P* = 0.0088) and *SLX1A* was associated with the *unspecified erythematous condition* (*P* = 0.0024). None of these gene-trait associations have been previously reported.

Additional genes were associated with traits common to both 16p11.2 duplications and deletions (Table 1). *SULT1A3* was associated with *mental retardation* (*P* = 0.0019). *ZNF48*, in the flanking region, was associated with *sleep apnea* (*P* = 0.0012). Neither of these gene-trait associations have been reported previously.

More than half of the 22q11.2 genes we tested (49/85) had a trait in their top 1% PheWAS results overlapping with a trait over-represented in 22q11.2 deletion carriers, and the full list of the gene-trait pairs are in Table 2. *LZTR1* was associated with cardiac and circulatory system traits (*disease of tricuspid valve* [*P* = 0.002]*, cardiomegaly* [*P* = 0.0033]*, hypotension NOS* [*P* = 0.0074]*)*, digestive traits (*acute vascular insufficiency of intestine* [*P* = 0.0031]), and infectious disease traits (*septicemia* [*P* = 0.00098], *bacterial infection NOS* [*P* = 0.0028]). *LZTR1* is a known driver of cardiac defects; digestive and immune defects due to this gene have not been described [72]. *TBX1*, considered to be a driver of cardiac defects, overlaps with the 22q11.2 deletion over-represented trait of *aortic ectasia* (*P* = 0.0031) [29–31]. *USP41* was associated with cardiac traits (*cardiac shunt/ heart septal defect* [*P* = 0.0085], *cardiac congenital anomalies* [*P* = 0.0073], *heart valve replaced* [*P* = 0.0026], *cardiac complications, not elsewhere classified*) [*P* = 0.0035], *hypercalcemia* [*P* = 0.0023], and *eustachian tube disorders* [*P* = 0.0026], which were all over-represented in 22q11.2 deletion carriers. *USP41* has not been previously tied to any of these disorders. The association between *TXNRD2* and a 22q11.2-deletion associated trait *upper gastrointestinal congenital anomalies* (*P* = 2.1×10^{-5}) reached phenome-wide significance; this association has not been previously reported.

Next, we found genes associated with 22q11.2 duplication-associated traits (Table 2). *LZTR1* was associated with the cardiac traits *disease of tricuspid valve* (*P* = 0.002) and *cardiomegaly (P* = 0.0033). *LZTR1* mutations have previously been associated with heart defects [72]. *PRODH* was associated with the trait of *delayed milestones (P* = 0.0076). *PRODH* has been a candidate gene for intelligence phenotypes associated with 22q11.2, but this association has been disputed [73]. *USP18* was associated with the primary known trait of *lack of normal physiological development, unspecified* (*P* = 0.0078) and *other and unspecified congenital anomalies* (*P* = 0.0068). These gene-trait associations were novel. *SMPD4P1* was also associated with *other and unspecified congenital anomalies (P* = 0.0049). This gene is a pseudogene of *SMPD4*, which is associated with structural brain anomalies [74]. *RTN4R* and *GGT2* were associated with the novel traits *allergy/adverse effect of penicillin* (*P* = 0.072) and *personal history of allergy to medicinal agents (P* = 0.0046), respectively. These genes have not been previously associated with any similar traits.

Finally, we tested whether the top individual gene PheWAS results were enriched within the top carrier status PheWAS results. We did this by analyzing where the top PheWAS traits associated with CNV genes were ranked within the top carrier status PheWAS results. We found no evidence for enrichment in any of 16p11.2 duplications, 16p11.2 deletions, 22q11.2 duplications, 22q11.2 deletions (S4 Fig).

The 16p11.2 gene-trait PheWAS pairs that meet the three selection criteria (brain-related traits, traits that were over-represented in 16p11.2 duplication or deletion carriers, or that reach phenome-wide significance) are depicted in Fig 5 (left), weighted by the strength of the association. About half of the 16p11.2 genes we tested (24/46) contribute to 36 traits meeting the selection criteria. The vast majority of these traits (28/36) had a single gene associated with them. At 16p11.2, *YPEL3,* and *MVP* are the largest contributors, by association strength, to traits that meet the criteria. Other than the association between *YPEL3* and *valvular heart disease/heart chambers*, these genes are the sole 16p11.2 genes associated with these traits. The traits that are associated with the most 16p11.2 genes are *valvular heart disease/heart chambers* (a 16p11.2 duplication-associated trait) and *other and unspecified disorders of the nervous system* (a 16p11.2 deletion-associated trait). As previously mentioned, we were unable to determine whether any these multi-gene associations are due to correlation alone, but we do notice that the genes associated with both of these traits are spread out along the CNV region.

The 22q11.2 gene-trait PheWAS pairs that meet the three selection criteria (brain-related traits, traits that were over-represented in 22q11.2 duplication or deletion carriers, or that reach phenome-wide significance) are depicted in Fig 5 (right), weighted by the strength of the association. About two-thirds of the 22q11.2 genes we tested (56/83) contribute to 140 traits meeting the selection criteria. Most traits (103/140) were associated with one single 22q11.2 gene. At 22q11.2, *TXNRD2* and *LZTR1* are the largest contributors, by association strength, to traits that meet the criteria. The traits that are associated with the greatest number of 22q11.2 genes are *other congenital musculoskeletal anomalies* (a 22q11.2 deletion-associated trait), *nausea and vomiting* (a 22q11.2 deletion-associated trait), and *congenital anomalies of genital organs* (a 22q11.2 deletion-associated trait). Genes associated with these traits are not all adjacent, and spread out along the CNV region.

DISCUSSION

We sought to identify individual genes in the 16p11.2 and 22q11.2 regions driving brain-related disorders, focusing on ASD, schizophrenia, bipolar disorder, obesity, and intellectual disability. In this novel *in-silico* approach, we tested whether genetically-driven predicted expression variation of the individual genes in each CNV was associated with brain-related disorders in the general population. We identified and replicated individual genes at 16p11.2 whose expression was associated with schizophrenia (*INO80E*), IQ (*SPN*), and BMI (*SPN, INO80E, TMEM219, TAOK2*) in the expected direction based on known 16p11.2 biology. Third, we used a biobank containing over 1,500 medical traits to explore the consequences of expression variation of 16p11.2 and 22q11.2 CNV genes across the medical phenome. We used electronic health record data to find traits overrepresented in 16p11.2 and 22q11.2 carriers, confirming previously known traits and finding novel traits.

INO80E, the gene we identified as a driver of schizophrenia and BMI, is a chromatin remodeling gene and has not been previously identified as a candidate gene for brain-related disorders. Mice heterozygous for this gene have shown abnormal locomotor activation [75]. Locomotor activity in mice is a frequently measured phenotypic test for brain-related disorders

including schizophrenia [76]. Our results are consistent with a previous observation that eQTLs from dorsolateral prefrontal cortex for *INO80E* co-localize with schizophrenia GWAS SNPs [77]. In addition, an analogous imputed expression based transcriptome-wide association study observed association between *INO80E* and schizophrenia using summary statistics [78]. By focusing on a specific schizophrenia-associated region, using individual level data, and performing a conditional analysis, we have obtained additional precision, and were able to fine-map the signal at 16p11.2 down to a single gene. Our study differs from Gusev *et al* in the expression prediction models used: we used 48 tissue models from the Genotype-Tissue Expression consortium, while Gusev *et al* used brain tissues from the CommonMind Consortium, blood tissues from Young Finns Study and Netherlands Twin Registry, and adipose tissues from the Metabolic Syndrome in Men study. The overlap in association results shows that our approach is robust to variation in predictive models. Due to lack of appropriate data, we were unable to determine whether, as with schizophrenia, the association between *INO80E* and BMI was independent of the other three genes associated with BMI. Mice with a heterozygous mutation in *Ino80e* showed increased body weight, consistent with our association result [75].

SPN, a gene highly associated with both IQ and BMI, is active in immune cells and is not known to play a role in brain-related disorders [79,80]. In the PheWAS study, however, we find *other conditions of brain* and *generalized anxiety disorder* among the top phenotypic associations of *SPN*, suggesting the role of this gene in brain-related disorders is worth further study. We note that the association p-values for *SPN* are much lower than for any other genes showing association signal. This may be because our approach detected relatively few eQTLs for *SPN* (12 SNPs in two tissues), many of which overlapped with highly associated GWAS SNPs for both IQ and BMI, rather than contributing to noise.

Our results give evidence that pleiotropy is involved in the pathogenicity of 16p11.2, as opposed to a strictly “one gene, one trait” model. Specifically, *INO80E* was associated with both schizophrenia and BMI, and *SPN* was associated with both BMI and IQ. Genetic correlations of at least -0.05 and as much as -0.5 have been estimated for the BMI/IQ and SCZ/BMI pairs, suggesting that pleiotropy may play a general role in these disorders [81–84]. Consistent with the genetic correlations, most (8/12) eQTL SNPs in our prediction models for *SPN* drove the associations with both IQ and BMI.

While most associations we detected were in the expected direction given previous knowledge, *MVP* and *KCTD13* were associated with BMI in the opposite (positive) direction, and *YPEL3* with schizophrenia in the negative direction. We resolved the unexpected association between *YPEL3* and schizophrenia by doing a conditional analysis, where we found that *YPEL3* was associated with schizophrenia simply due to correlation with *INO80E*. For *MVP* and *KCTD13* where we could not perform a conditional analysis, we offer three explanations: these results may be false-positives due to correlation-based “hitchhiking”, they may demonstrate a limitation of our approach, or they may have a true BMI-increasing effect. First, it is possible that these two genes are not independent associations with BMI, as was the case with *YPEL3* with schizophrenia. We noticed that 40% of the eQTLs of *KCTD13* used for predicting expression have a different direction of effect on *KCTD13* vs. their other target genes. Similarly, we note

that in five of the eight tissues where a predictive model for *KCTD13* exists, the true expression levels of *KCTD13* have a significant negative correlation with at least one of the genes that are negatively associated with BMI. These observations are consistent with a case where *KCTD13* "hitchhikes" to statistical significance with other negatively-associated genes due to correlation, but does not contribute to BMI itself. While we could not perform a conditional analysis on the BMI data in our study, testing for independence of association can provide evidence for or against this interpretation. Second, this result might represent a limitation of our eQTL-based method. *KCTD13* is a highly brain-expressed gene, but had no high-quality brain prediction models [50]. The direction of the eQTLs regulating *KCTD13* expression in the brain may be brain-specific, and brain may be the only relevant tissue for the effect of *KCTD13* on BMI. That is, *KCTD13* may actually have a strong negative correlation with BMI, but falsely appears positive due to the specific eQTLs used for expression prediction. Such tissue-specific eQTL directions of effect have been observed for at least 2,000 genes [85]. Improved brain-specific prediction models will resolve this limitation. Third, these genes could have a true BMI-increasing effect. Consistent with our *in silico* finding, *MVP* is upregulated in adipocytes and cells co-cultured with adipocytes from obese individuals [86,87]. If *KCTD13* and *MVP* have a true BMI-increasing effect, the 16p11.2 region contains both BMI-increasing and BMI-decreasing genes, and the effect of the BMI-decreasing genes is stronger. Such a model is a potential explanation for the observation that duplications at 16p11.2 in mice, unlike humans, are associated with obesity [43]. One set of genes may be the more influential determinant of the obesity trait in each organism.

Our PheWAS of traits overrepresented in 16p11.2 and 22q11.2 carriers served as a validation of our biobank EHR approach via replication of previously-identified CNV-associated traits. Brain-related traits, such as *delayed milestones, mental retardation* and *pervasive developmental disorders*, were among the top over-represented traits in both 16p11.2 and 22q11.2 CNV carriers. 22q11.2 deletion carriers were strongly associated with *cardiac congenital anomalies* and *cleft palate*, two of the hallmark features of the CNV. Even though the total number of CNV carriers within the biobank was relatively small, the traits listed in their medical records were consistent enough to let us pick up these differences. At the same time, we identified novel traits that may be confirmed in larger samples of CNV carriers. Our 16p11.2 duplication carriers had two significant dermatological trait associations: *vascular disorders of skin* and *telogen effluvium*. To the best of our knowledge, dermatological traits have not been previously recognized as associated with 16p11.2 duplications or deletions. *Telogen effluvium* (stress-induced hair loss) can be potentially explained as a symptom of brain-related disorders, or a side effect of drugs used to treat bipolar disorder or schizophrenia [88,89]. Similarly, the novel diagnosis of *retention of urine* may refer to a side effect of drugs used to treat epilepsy, schizophrenia, and bipolar disorder [90]. Although novel, the association between 16p11.2 deletions and *anterior horn cell disease* could demonstrate another impact of 16p11.2 deletions on the functions of the nervous system. The gene *FUS* at 16p11.2, about 1MB away from the CNV region, is a known cause of anterior horn cell disease [91]. If deletion of the 16p11.2 CNV disrupts an enhancer to *FUS*, we could observe over-representation of *anterior horn cell disease* in the PheWAS.

Our PheWAS between the predicted expressions of 16p11.2 and 22q11.2 genes and 1,500 medical phenotypic codes resulted in eight phenome-wide significant gene-trait pairs. Nearly all of these associations are novel. There is previous evidence that *YPEL3* expression is downregulated in estrogen receptor positive breast cancer cells, and here, this gene has a significant PheWAS association with *other specified benign mammary dysplasias* [92]. However, there has not been any previous evidence that 16p11.2 CNV carriers have increased rates of breast cancer. The sole phenome-wide significant association that overlaps with traits in the carrier vs. non-carrier PheWAS is between *SULT1A4* and *valvular heart disease/heart chambers*. Atrial septal defect, a congenital defect of the separation of heart chambers, has been observed in approximately 10% of 16p11.2 deletion carriers [93,94]. In addition, there has been one case report of a 16p11.2 duplication carrier with atrial septal defect [71]. Because few genes had any associations which were phenome-wide significant, we elected to analyze the top 1% of associations of each gene. We noticed that our gene-by-gene PheWAS recapitulated the effect of two 22q11.2 genes known to lead to disorders of the circulatory system: *TBX1* and *LZTR1* [29–31,95].

We were not able to assess direction of effect in the gene-by-gene PheWAS analysis. The problem of assigning a direction to a cross-tissue association analysis is a difficult one, as direction of effect is often inconsistent across tissues. Some of the potential factors underlying this inconsistency are different directions of eQTLs across tissues and variation among prediction model qualities [85]. By estimating cross-tissue sign as an average across single-tissue signs, we did observe that our significant associations with schizophrenia and BMI were mostly in the expected direction based on prior knowledge. However, there were genes such as significantly associated with both phenotypes that did not follow the expected direction; some potential reasons for this are described earlier. The inconsistency in direction was even more common for genes with suggestive p-values in the 0.0001-0.05 range, the same range as the majority of the PheWAS trait associations.

While many of our gene-specific PheWAS results are not traits previously associated with the 16p11.2 or 22q11.2 CNVs, it does not mean that these traits are irrelevant to CNV biology. Identified 16p11.2 carriers are typically young (average age 13 in our biobank), and are not likely to have developed disorders with older onset age. For example, 16p11.2 carriers are unlikely to have developed breast cancer before adulthood, even if they are at relatively high risk. Because 16p11.2 carriers carry a mutation in a gene associated with breast tumors, physicians may recommend to start mammographic screening early, similar to other high-risk populations. Second, there are genes in both 16p11.2 and 22q11.2 regions that are associated with dementia. While dementia is not a classic feature associated with either CNV, most carriers of these CNVs are still below the age of developing such disorders. Overall, these observations show a need for longitudinal studies of carrier cohorts and studies of carriers in older age. Such additional data may point to additional common features of 16p11.2 and 22q11.2 carriers.

Limitations

The 16p11.2 and 22q11.2 CNVs are significant risk factors for ASD and schizophrenia, respectively, and yet no individual genes in either CNV were associated with case-control status for the associated trait in the best-powered datasets available to us. Assuming the true causal gene(s) for these disorders do exist within the CNV, limitations in our approach may preclude us from discovering them. Predictive models for gene expression are imperfect; they capture a small subset of the variation in expression of these genes (the largest single-tissue prediction R^2 for our genes is 0.45, and the average R^2 is 0.07). Moreover, there are genes in both regions for which no high-quality models exist. If the causal gene is among the genes that cannot be well predicted, we cannot detect this gene by our approach. One category of genes that are not represented in our study are microRNAs. 22q11.2 carriers have a unique microRNA signature, and the contribution of microRNA to 22q11.2-CNV associated schizophrenia has been previously hypothesized [96,97]. If the microRNAs are important regulatory elements for 22q11.2-associated traits, our approach is insufficient to detect them.

A further limitation is that eQTLs generally have small effect sizes. As a result, the variation in expression that can be modeled using eQTLs may be considerably smaller for some genes than the effect of deletions and duplications. For example, there may be a gene at 22q11.2 for which decreases in expression contribute to schizophrenia, but only when expression levels reach a threshold of near 50% of the expression levels of non-carriers. Furthermore, the expression predictions of these genes are calculated solely using *cis*-eQTLs within 1MB of the gene [62]. It may be necessary to consider the effect of *trans*-eQTLs to explore the effect of expression variation accurately.

A third limitation of this approach is that we have so far considered the effect of each CNV gene independently, when, in reality, the genes may not be acting independently. A Drosophila model for 16p11.2 genes has shown evidence of epistasis between genes within a CNV as a modifier of phenotype [39]. If there are 16p11.2 traits in humans also driven by epistasis, our single-gene screen would not have detected the appropriate genes for those traits.

Conclusions

In developing our approach, we hypothesized that naturally occurring variation in gene expression of CNV genes in non-carriers would convey risk for traits seen in CNV carriers. We found that this was true for at least three 16p11.2 associated traits: BMI, schizophrenia, and IQ. Promisingly, the direction of association was generally consistent with whether the trait was found in duplication or deletion carriers. Our approach is computationally efficient, extendable to other CNV-trait pairs, and overcomes one limitation of animal models by testing the effect of CNV genes specifically in humans.

Rather than focusing on any specific tissue(s), we chose to perform a cross-tissue analysis, an approach that improves power to detect gene-trait associations and detected 16p11.2 genes associated with schizophrenia, IQ, and BMI [63]. An underlying assumption behind the use of all tissues (rather than just brain) for these mental disorders is that eQTLs for our genes of interest are shared across tissues, and that the same eQTLs affect the expression of a gene in the brain

as in other tissues. This assumption often, but not always, holds true: in GTEx correlation of eQTL effect sizes between brain and non-brain tissues was 0.499 (Spearman) [51]. We believe our use of the multi-tissue approach is justified as it allows us to probe the effects of genes with no brain expression prediction models.

In this study, we synthesized information from both large GWAS studies and biobanks, benefiting from the strengths of both approaches. Psychiatric brain-related disorders such as autism, schizophrenia, and bipolar disorder have a population frequency below 5%, so large datasets specifically ascertained for brain-related disorders are better at providing sufficient statistical power for association analysis, especially when the effect of each gene is small. On the other hand, the presence of many diagnostic codes in a biobank help identify brain-related traits that may be relevant to CNVs but not the primary reported symptoms, such as speech and language disorder and dementia. An advantage of our biobank approach was our ability to carry out two studies using different aspects of the same dataset. The presence of CNV carrier status in the biobank allowed us to probe the phenotypic consequences of the entire deletion or duplication. Then, we were able to test each CNV gene for association with the same diagnostic descriptions.

Our novel approach provided new insights into how individual genes in the 16p11.2 and 22q11.2 CNVs may drive health and behavior in a human population. Expression imputation methods allowed us to study the predicted effects of individual CNV genes in large human populations. The incorporation of medical records into biobanks provided a way to determine clinical symptoms and diagnoses to which expression differences in the genes may contribute. We expect our ability to detect genes with this type of approach to increase in the coming years, as more individuals in biobanks are genotyped, the number of individuals contributing to large cohorts grow, and the methods to more finely and accurately predict gene expression improve. Additional experiments on our newly prioritized genes are necessary to determine their specific functional impact on brain-related disorders and to evaluate their value as putative therapeutic targets.

METHODS

Case-control and quantitative genotypic data

We obtained the imputed individual-level genotypes for ASD, bipolar disorder, and schizophrenia from the Psychiatric Genomics Consortium in PLINK format (S1 Table). These datasets include mainly European populations and are comprised of several independent cohorts: 30 in bipolar disorder (N = 19,202 cases 30,472 controls, downloaded July 2019), 46 in schizophrenia (N = 31,507 cases 40,230 controls, downloaded July 2018), 14 in ASD (N = 7,386 cases, 8,566 controls, downloaded May 2019) [58,59,98]. For two additional traits, we used publicly available summary statistics: BMI from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (2015, both sexes, n=339,224, downloaded June 2019) and IQ from Savage *et al* 2018 (n=269,867, downloaded May 2019) [60,61].

For replication studies, we used the publicly available GWAS summary statistics for schizophrenia and BMI from the UK Biobank (http://www.nealelab.is/uk-biobank). We could not use the UK Biobank IQ data for replication, as it overlapped with our discovery data. The list of UK Biobank phenotypes used is in S1 Table.

Expression prediction models

In order to impute gene expression, we obtained PrediXcan models for 48 tissues based on GTEx v7 Europeans from predictdb.org [62,63,99]. These models were generated by training an elastic net model that finds the best set of cis-SNP predictors for the expression of a gene in a tissue in the GTEx cohorts [62]. Only models with predicted-observed correlation $R^2 > 0.01$ and $P < 0.05$ are kept.

Genes studied

We studied all coding and noncoding genes at the 16p11.2 and 22q11.2 copy number variant loci for which expression prediction models were available. We included flanking genes in a 200kb window upstream and downstream of the CNV breakpoints. Not every gene in the CNV region was available to be analyzed through our methods. 31 genes at or near 16p11.2 lacked high-quality prediction models in every tissue. Overall, 46 coding and noncoding genes at or near 16p11.2 were tested. 99 genes at or near 22q11.2 lacked high-quality prediction models in every tissue. Overall, 83 coding and noncoding genes at or near 22q11.2 were tested (S2 Table, S1 Fig).

Comparison of observed expression correlations with predicted expression correlations

Observed expression correlations were calculated at a tissue-specific level on data from GTEx v7 [100]. Tissue-specific predicted expression was calculated by applying the appropriate GTEx predictive model on the GTEx v6p genotypes (dbgap id: phs000424.v6.p1) for 450 individuals. To minimize spurious correlations, the predicted expression levels were rigorously filtered and normalized. Specifically, the expression levels were filtered for outliers (values above 1.5*interquartile range, in either direction), adjusted for the principal components of both the predicted expression levels and the first 20 PCs of the GTEx genotypes, inverse-quantile normalized, re-adjusted for principal components, and re-filtered for outliers. We observed that normalization of the predicted expression reintroduced correlation between expression and the genotypic PCs, leading us to perform the correction twice.

Association analysis in the PGC ASD, schizophrenia, and bipolar disorder individual-level data

All individual-level data analysis was performed on the LISA cluster, part of SURFSara. Each of the three PGC collections went through quality control, filtering, and PCA calculation, as described previously [57–59]. For each cohort, the convert_plink_to_dosage.py script in PrediXcan was used to convert chromosome 16 and 22 genotypes from PLINK to dosage

format, keeping only the SNPs used in at least one predictive model. Using these dosages, the --predict function in PrediXcan was used to generate predicted expressions of CNV genes for each individual. Genes with predicted expression of 0 for all individuals in a single tissue were filtered out. The average number of genes filtered out across tissues and cohorts was 0.89; the maximum was 11 in Thyroid in the Japanese cohort. Cross-tissue association studies between predicted expression and case-control status were performed using MultiXcan. In brief, MultiXcan takes the matrix of predicted expressions across tissues for a single gene, calculates the principal components of this matrix to adjust for collinearity, fits a model between these PCs and case-control status, and reports the results of the overall fit [63].

Meta-analysis with METAL on the p-values from MultiXcan, weighted by the sample size of each cohort, was used to calculate a cross-cohort association statistic [101]. The joint fit in MultiXcan generates an F-statistic that is always greater than zero, while some of our traits of interest have a specific expected direction (only in deletion or only a duplication). Thus, a direction was assigned to each MultiXcan result. This was done by running a tissue-specific PrediXcan association analysis between predicted expressions and case-control status (using --logistic), which calculates a signed association Z-score for every gene. The sign of the mean Z-score for that gene across all tissues was the direction of association used for meta-analysis.

Association analysis in summary-level data

Analyses involving summary-level data were performed on the Wynton cluster at UCSF. Both the single-tissue PrediXcan and the multi-tissue MultiXcan methods have been extended to estimate the association results between genetically regulated expression and a trait if only the summary statistics for the trait are available. For each trait's summary statistics, the summary version of PrediXcan (S-PrediXcan) and the associated MetaMany.py script was used to calculate the per-tissue association results for each gene in 48 GTEx tissues. Association results were aggregated across tissues using the summary version of MultiXcan (S-MultiXcan). The mean single-tissue Z-score (as reported in the zmean column in the S-MultiXcan output) was used as the direction of association. The UK Biobank replication studies were performed in the same way.

Conditional analysis to fine-map associations

In order to adapt the multi-tissue association analysis to perform conditional testing, "conditioned predicted expressions" were generated for a set of genes associated with the same trait. As an example, take the set of three genes [*INO80E, YPEL3, TMEM219*] associated with schizophrenia. In order to condition on *INO80E*, for example, the predicted expression of *INO80E* was regressed out of the predicted expressions of *YPEL3* and *TMEM219*. Conditioning was only done in tissues where the predicted expressions of the genes were correlated (Spearman correlation $P < 0.05$). Another set of conditioned predicted expressions was generated by adjusting the predicted expression of *INO80E* by the predicted expressions of [*TMEM219, YPEL3*]. Separately, these per-tissue conditioned predicted expressions were used as inputs for a MultiXcan analysis and METAL meta-analysis on schizophrenia as described

earlier. All three individually associated genes were tested in this manner. The same analysis was later used to test for independence of association between traits in the PheWAS.

Phenome-wide association studies

Vanderbilt University Medical Center (VUMC) houses de-identified phenotypic data in the form of the electronic health records (EHR) within the synthetic derivative (SD) system [65]. The SD contains EHR data including ICD9/10 billing codes, physician notes, lab results, and similar documentation for 3.1 million individuals. BioVU is a biobank at VUMC that is composed of a subset of individuals from the SD that have de-identified DNA samples linked to their EHR phenotype information. The clinical information is updated every 1-3 months for the de-identified EHRs. Detailed description of program operations, ethical considerations, and continuing oversight and patient engagement have been published [65]. At time of analysis, the biobank contained approximately 23,000 individuals who have been genotyped [102]. DNA samples were genotyped with genome-wide arrays including the Multi-Ethnic Global (MEGA) array, and the genotype data were imputed into the HRC reference panel [103] using the Michigan imputation server [104]. Imputed data and the 1000 Genome Project data were combined to carry out principal component analysis (PCA) and European ancestry samples were extracted for analysis based on the PCA plot. GTEx v7 models from PredictDB were applied to the samples to calculate genetically regulated expression (GReX).

Phenome-wide association studies (PheWAS) were carried out for each tissue-specific predicted expression using ‘phecodes’, phenotypes derived from the International Code for Diseases version 9 (ICD-9) billing codes of EHRs. The PheWAS package for R, version 0.11.2-3 (2017) was used to define case, control and exclusion criteria [69,105]. We required two codes on different visit days to define a case for all conditions. Logistic regression was performed for each phecode with at least 20 cases, while adjusting for covariates of age, gender, genotyping array type/batch and three principal components of ancestry. The single-tissue PheWAS results were combined across tissues using MultiXcan, as was done to analyze individual-level GWAS data from the Psychiatric Genomics Consortium [63].

Determining traits over-represented in carriers

3.1 million electronic medical records from the SD at VUMC were queried for keywords corresponding to copy number variations at 16p11.2 and 22q11.2 (S8 Table). Individual charts identified as containing the keywords were manually reviewed and patients were labeled as cases if their medical records provided evidence of CNV carrier status. Patients identified in the queries with insufficient evidence of CNV carrier status were excluded from the analysis. Cases with positive 16p11.2 and 22q11.2 CNV carrier status were identified as: “16p11.2 duplication” (n=48), “16p11.2 deletion” (n=48), ”22q11.2 duplication” (n=43). Additional individuals in the 22q11.2 deletion case group were identified by querying the medical records for alternate terms including: “velocardiofacial”, “DiGeorge”, “conotruncal anomaly face”, “Cayler”, “Opitz G/BBB”, “Shprintzen”, and “CATCH22” (n=388) . Individuals were excluded from case groups if their

records included a mention of additional CNVs. Individuals within the 16p11.2 case groups were also excluded if the size of the reported CNV was 200-250 kb. Individuals within the 22q11.2 case group were excluded if the size of the CNV was smaller than 500 kb or if there was a mention of “distal” when referring to the deletion or duplication. PheWAS was carried out, with each of the four carrier categories as cases and 700,000 medical home individuals as controls. The medical home individuals are patients seen at a Vanderbilt affiliated clinic on five different occasions over the course of three years.

Test for enrichment of gene-specific PheWAS for traits over-represented in carriers

For each of 16p11.2 duplications, 16p11.2 deletions, 22q11.2 duplications, 22q11.2 deletions, the entire carrier vs. non-carrier PheWAS results were ranked. All of the traits in the top 1% of per-gene 16p11.2 and 22q11.2 PheWAS results were converted to a value corresponding to the rank of the trait in the carrier vs. non-carrier PheWAS. To determine whether the CNV genes’ PheWAS top traits were distributed nonrandomly within the ranks of all traits, the distribution of the ranks of the each CNV’s per-gene PheWAS top traits was compared to the ranks of all carrier vs. noncarrier PheWAS traits for the same CNV (a uniform distribution) using a Kolmogorov-Smirnov test.

Significance threshold for association studies

The significance threshold used for each discovery MultiXcan or S-MultiXcan association study and conditional analysis was 0.05/(number of traits*number of CNV genes tested). In practice, this usually meant 5 traits and 132 CNV genes, for a threshold of $P < 7.6\text{x}10^{-5}$. For replication studies, the significance threshold was set at 0.05 in order to test a single gene. The exception was in the BMI UK Biobank dataset. For BMI, 100 random subsets of adjacent genes of approximately the same length and gene count as the CNV were tested. The 99th percentile of the S-MultiXcan p-values for these genes was used as a permutation-based significance threshold.

In the gene-based PheWAS study, there were 1,534 phecodes (each with at least 20 cases) tested overall, corresponding to a Bonferroni-corrected phenome-wide significance threshold of $3.3\text{x}10^{-5}$. For genes having no phenome-wide significant results, their top 15 associations, corresponding to the top 1% of the 1,534 phecodes, were used. In the carrier vs. non-carrier PheWAS, there were 1,796 phecodes tested overall, corresponding to a Bonferroni-corrected phenome-wide significance threshold of $2.78\text{x}10^{-5}$. Additional traits meeting a false discovery rate threshold of 0.05 were considered in identifying traits both over-represented in carriers and represented in individual gene PheWAS.

Graphical summary of selected PheWAS results

The *chordDiagram* method in the *circlize* package was used to generate the circle summary plots [106]. The gene-trait pairs we selected for Tables 1 and 2 were used as inputs, with the -log10 p-value of association used as the weighting to determine the edge width. For the

22q11.2 circle plot, only associations with $P < 5x10^{-3}$ were used in order to create a legible plot. Descriptions were cut off at 55 characters; to read the entire descriptions see Tables 1 and 2.

FUNDING

This work was sponsored by the National Institute of Mental Health [R01 MH107467 to LAW, R01 MH113362 to NJC]; as well as the National Human Genome Research Institute [U01 HG009086 to NJC]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGEMENTS

We would like to acknowledge the consortia and individuals that provided the individual data we used in this study: BioVU at Vanderbilt University, the Genotype-Tissue Expression Consortium, and the Autism, Schizophrenia, and Bipolar working groups of the Psychiatric Genomics Consortium. PGC working group authors are listed in S1 Note. We would also like to thank Megan Sheuy for providing cleaned BMI data for BioVU participants.

Fig 1: An overview of the three components of this study.
We probed the effects of individual genes in the 16p11.2 and 22q11.2 CNVs on phenotype in two ways. First (bottom left), we used large GWAS datasets for brain-related traits associated with both CNVs to determine whether variation in predicted expression in any of the individual genes in each CNV was associated with case-control status for each trait. Second (bottom right), we used a biobank containing clinical and genotypic data, and performed a phenome-wide association study to determine clinical traits that are driven by the predicted expression of individual CNV genes. In the third component of this study (top), we identified the individuals with 16p11.2 and 22q11.2 duplications or deletions present in the biobank and determined the clinical traits that were over-represented in CNV-carriers, as well as whether these traits overlapped with the PheWAS traits driven by the individual genes.

Fig 2. Association between 16p11.2 genes and three brain-related traits.
Association between predicted expression of 16p11.2 genes and schizophrenia (left), BMI (middle), IQ (right), using MultiXcan (schizophrenia) and S-MultiXcan (BMI, IQ). Genes are listed on the horizontal access in order of chromosomal position. The -log10 p-values on the vertical axis are given a positive or negative direction based on the average direction of the single-tissue results. The significance threshold is a Bonferroni correction on the total number of 16p11.2 and 22q11.2 genes tested across 5 traits. Genes exceeding the significance threshold in the expected direction (positive for schizophrenia, negative for BMI, either for IQ) are denoted as x's.

Fig 3. Clinical traits over-represented in 16p11.2 deletion and duplication carriers.
CNV carriers were identified in the EHR by keyword search and chart review (left, 16p11.2 deletions [n=48], right, 16p11.2 duplications [n=48], see Methods). Controls included all individuals without the CNV within the medical home population at Vanderbilt (n~734,000). The

x-axis represents the PheWAS codes that are mapped from ICD-9/ICD-10 codes, grouped and color-coded by organ system. The y-axis represents the level of significance ($-\log_{10}p$). The horizontal red line indicates a Bonferroni correction for the number of phenotypes tested in this PheWAS (p = 0.05/1,796 = 2.784E-5); the horizontal blue line indicates p = 0.05. Each triangle represents a phenotype. Triangles represent direction of effect; upward pointing arrows indicate phenotypes more common in cases. Covariates included age, gender, and self-reported race extracted from the EHR. Phenotypes reaching Bonferroni-corrected significance level are labeled in plot.

Fig 4. Clinical traits over-represented in 22q11.2 deletion and duplication carriers.
CNV carriers were identified in the EHR by keyword search and chart review (left, 22q11.2 deletions [n=388], right, 22q11.2 duplications [n=43], see Methods). Controls included all individuals without the CNV within the medical home population at Vanderbilt (n~734,000). The x-axis represents the PheWAS codes that are mapped from ICD-9/ICD-10 codes, grouped and color-coded by organ system. The y-axis represents the level of significance ($-\log_{10}p$). The horizontal red line indicates a Bonferroni correction for the number of phenotypes tested in this PheWAS (p = 0.05/1,796 = 2.784E-5); the horizontal blue line indicates p = 0.05. Each triangle represents a phenotype. Triangles represent direction of effect; upward pointing arrows indicate phenotypes more common in cases. Covariates included age, gender, and self-reported race extracted from the EHR. Phenotypes reaching Bonferroni-corrected significance level are labeled in plot.

Fig 5. Graphical summary of selected PheWAS results by gene.
Each circle contains the CNV genes, in chromosomal order, on the bottom, and their associated PheWAS traits at the top. Genes are connected to their PheWAS-associated traits, with the width of the line proportional to the -log10 p-value of the association. If a trait is also over-represented in duplication and/or deletion carriers it is marked with a + (duplications), - (deletions), or +/- (both). For 22q11.2 genes, only associations with $P < 5\text{x}10^{-3}$ are shown for clarity. The complete list of gene-trait pairs can be found in Tables 1 and 2, and Supplementary Table S7.

S1 Fig: Position of the 16p11.2 and 22q11.2 CNVs on chromosomes 16 and 22 and list of genes in these regions. Bolded genes had predictive models available and were considered in our study. 22q11.2 genes are staggered on either side of the chromosome for visual clarity.

S2 Fig: Association between 16p11.2 genes and the remaining two brain-related traits.
Association between predicted expression of 16p11.2 genes and autism (left), bipolar disorder (right). Genes are listed on the horizontal axis in order of chromosomal position. The -log10 p-values on the vertical axis are given a positive or negative direction based on the average direction of the single-tissue results.

S3 Fig: Association between 22q11.2 genes and five brain-related traits.
Association between predicted expression of 22q11.2 genes and, from bottom to top, autism, bipolar disorder, schizophrenia, BMI, and IQ, using MultiXcan (autism, bipolar disorder,

schizophrenia) and S-MultiXcan (BMI, IQ). Genes are listed on the horizontal axis in order of chromosomal position. The -log10 p-values on the vertical axis are given a positive or negative direction based on the average direction of the single-tissue results.

S4 Fig: Test of enrichment of gene-based PheWAS results within carrier vs. non-carrier PheWAS results. The histograms show the distribution of ranks of the traits associated with individual 16p11.2 or 22q11.2 genes within the ranks of traits from the association analysis with carrier status. A kolmogorov-smirnov test was applied to determine whether the distribution of ranks of traits associated with individual CNV genes was different from the expected (i.e. uniform) distribution of ranks of traits associated with carrier status. The x-axis shows counts of traits from the gene-by-gene PheWAS. and the y-axis shows ranks. The gene-based PheWAS results would be considered enriched within carrier status PheWAS results if their ranks tended to skew towards one side of the distribution.

S1 Note: Supplementary Acknowledgements. Members of the Psychiatric Genomics Consortium contributing to this work.

S1 Table: List of genotyped discovery and replication cohorts used in the study. List of Pis from the Psychiatric Genomcis Consortium who have provided us with their data for analysis.

S2 Table: List of genes at or near 16p11.2 and 22q11.2. List of coding and non-coding genes in the CNV region, as well as flanking genes 200kb on either side. Genes for which PrediXcan models were available and the range of model qualities (R^2) are noted.

S3 Table: Results of MultiXcan and S-MultiXcan associations between CNV genes and autism, schizophrenia, bipolar disorder, BMI, and IQ. For autism, bipolar disorder, and schizophrenia, z-scores and p-values come from a METAL meta-analysis across PGC cohorts. For BMI and IQ, mean z-scores and p-values come direction from S-MultiXcan output.

S4 Table: Conditional analysis for independence of associations. Conditional analysis was performed on the PGC schizophrenia associations, as well as five BioVU clinical trait associations. Only pairs of genes which had correlated predicted expressions in at least one tissue were used for conditioning.

S5 Table: Replication of association results from the GWAS datasets and the clinical PheWAS. The replication criteria was $P < 0.1$ in an independent dataset (PGC, UK Biobank, or BioVU). See S2 Table for more information on replication datasets used.

S6 Table: Genes over-represented in CNV carriers. The four categories of CNV carrier – 16p11.2 duplication, 16p11.2 deletion, 22q11.2 duplication, 22q11.2 deletion – were tested separately. The results for all clinical traits tested are provided.

S7 Table: Top PheWAS associations of 16p11.2 and 22q11.2 genes. The top 15 associated traits for each gene, regardless of p-value, are shown. These represent the top 1% of

associations among all traits tested. Genes are listed in alphabetical order, with each trait's sample size and phecode (www.phewascatalog.org) noted.

S8 Table: Identifying 16p11.2 and 22q11.2 cases from electronic health records (EHR**).** Keyword searches across all documents within the Vanderbilt EHR were performed to identify individuals carrying 16p11.2 or 22q11.2 CNVs. Individuals with documents containing matching keywords were reviewed manually to confirm the presence of 16p11.2 or 22q11.2 CNV. Individuals were excluded from case groups if their records included a mention of additional CNVs. Individuals within the 16p11.2 case groups were also excluded if the size of the reported CNV was 200-250 kb. Individuals within the 22q11.2 case group were excluded if the size of the CNV was smaller than 500 kb or if there was a mention of "distal" when referring to the deletion or duplication. Confirmed case numbers are listed, with the non-genotyped counts in parentheses. Non-genotyped individuals were used for downstream phenome-wide analyses.

1. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, et al. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature. 2011;478: 97–102. doi:10.1038/nature10406
2. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet. 2009;41: 1223–1227. doi:10.1038/ng.474
3. Shinawi M, Liu P, Kang S-HL, Shen J, Belmont JW, Scott DA, et al. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet. 2010;47: 332–41. doi:10.1136/jmg.2009.073015
4. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, et al. Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet. 2007;17: 628–638. doi:10.1093/hmg/ddm376
5. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between Microdeletion and Microduplication at 16p11.2 and Autism. N Engl J Med. 2008;358: 667–675. doi:10.1056/NEJMoa075974
6. Bassett AS, Chow EWC, Husted J, Weksberg R, Caluseriu O, Webb GD, et al. Clinical features of 78 adults with 22q11 deletion syndrome. Am J Med Genet Part A. 2005;138A: 307–313. doi:10.1002/ajmg.a.30984
7. Campbell IM, Sheppard SE, Crowley TB, McGinn DE, Bailey A, McGinn MJ, et al. What is new with 22q? An update from the 22q and You Center at the Children's Hospital of Philadelphia. Am J Med Genet Part A. 2018;176: 2058–2069. doi:10.1002/ajmg.a.40637
8. Wentzel C, Fernström M, Öhrner Y, Annerén G, Thuresson A-C. Clinical variability of the 22q11.2 duplication syndrome. Eur J Med Genet. 2008;51: 501–510. doi:10.1016/J.EJMG.2008.07.005
9. Schneider M, Debbané M, Bassett AS, Chow EWC, Fung WLA, van den Bree MBM, et al. Psychiatric Disorders From Childhood to Adulthood in 22q11.2 Deletion Syndrome: Results From the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. Am J Psychiatry. 2014;171: 627–639. doi:10.1176/appi.ajp.2013.13070864
10. Voll SL, Boot E, Butcher NJ, Cooper S, Heung T, Chow EWC, et al. Obesity in adults with 22q11.2 deletion syndrome. Genet Med. 2017;19: 204–208. doi:10.1038/gim.2016.98
11. Carlson C, Papolos D, Pandita RK, Faedda GL, Veit S, Goldberg R, et al. Molecular analysis of velo-cardio-facial syndrome patients with psychiatric disorders. Am J Hum

Genet. 1997;60: 851–859.
12. Sahoo T, Theisen A, Rosenfeld JA, Lamb AN, Ravnan JB, Schultz RA, et al. Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. Genet Med. 2011;13: 868–880. doi:10.1097/GIM.0b013e3182217a06
13. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. Am J Hum Genet. 2008;84: 148–161. doi:10.1016/j.ajhg.2008.12.014
14. Bijlsma EK, Gijsbers ACJ, Schuurs-Hoeijmakers JHM, van Haeringen A, Fransen van de Putte DE, Anderlid B-M, et al. Extending the phenotype of recurrent rearrangements of 16p11.2: Deletions in mentally retarded patients without autism and in normal individuals. Eur J Med Genet. 2009;52: 77–87. doi:10.1016/j.ejmg.2009.03.006
15. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet. 2009;41: 1223–7. doi:10.1038/ng.474
16. Rees E, Kirov G, Sanders A, Walters JTR, Chambert KD, Shi J, et al. Evidence that duplications of 22q11.2 protect against schizophrenia. Mol Psychiatry. 2014;19: 37–40. doi:10.1038/mp.2013.156
17. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature. 2010;463: 671–675. doi:10.1038/nature08727
18. Smith ACM, McGavran L, Robinson J, Waldstein G, Macfarlane J, Zonona J, et al. Interstitial deletion of (17)(p11.2p11.2) in nine patients. Am J Med Genet. 1986;24: 393–414. doi:10.1002/ajmg.1320240303
19. Potocki L, Chen KS, Park SS, Osterholm DE, Withers MA, Kimonis V, et al. Molecular mechanism for duplication 17p11.2 - The homologous recombination reciprocal of the Smith-Magenis microdeletion. Nat Genet. 2000;24: 84–87. doi:10.1038/71743
20. Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. Mutations in RAI1 associated with Smith–Magenis syndrome. Nat Genet. 2003;33: 466–468. doi:10.1038/ng1126
21. Walz K, Paylor R, Yan J, Bi W, Lupski JR. Rai1 duplication causes physical and behavioral phenotypes in a mouse model of dup(17)(p11.2p11.2). J Clin Invest. 2006;116: 3035–3041. doi:10.1172/JCI28953
22. Williams JCP, Barratt-Boyes BG, Lowe JB. Supravalvular Aortic Stenosis. Circulation. 1961;24: 1311–1318. doi:10.1161/01.CIR.24.6.1311
23. Beuren AJ, Apitz J, Harmjanz D. Supravalvular Aortic Stenosis in Association with Mental Retardation and a Certain Facial Appearance. Circulation. 1962;26: 1235–1240. doi:10.1161/01.CIR.26.6.1235
24. Curran ME, Atkinson DL, Ewart AK, Morris CA, Leppert MF, Keating MT. The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. Cell. 1993;73: 159–168. doi:10.1016/0092-8674(93)90168-P
25. Ewart AK, Morris CA, Atkinson D, Jin W, Sternes K, Spallone P, et al. Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. Nat Genet. 1993;5: 11–16. doi:10.1038/ng0993-11
26. Koolen DA, Vissers LELM, Pfundt R, De Leeuw N, Knight SJL, Regan R, et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet. 2006;38: 999–1001. doi:10.1038/ng1853
27. Koolen DA, Sharp AJ, Hurst JA, Firth H V, Knight SJL, Goldenberg A, et al. Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. J Med Genet. 2008;45: 710–720. doi:10.1136/JMG.2008.058701
28. Koolen DA, Kramer JM, Neveling K, Nillesen WM, Moore-Barton HL, Elmslie F V, et al.

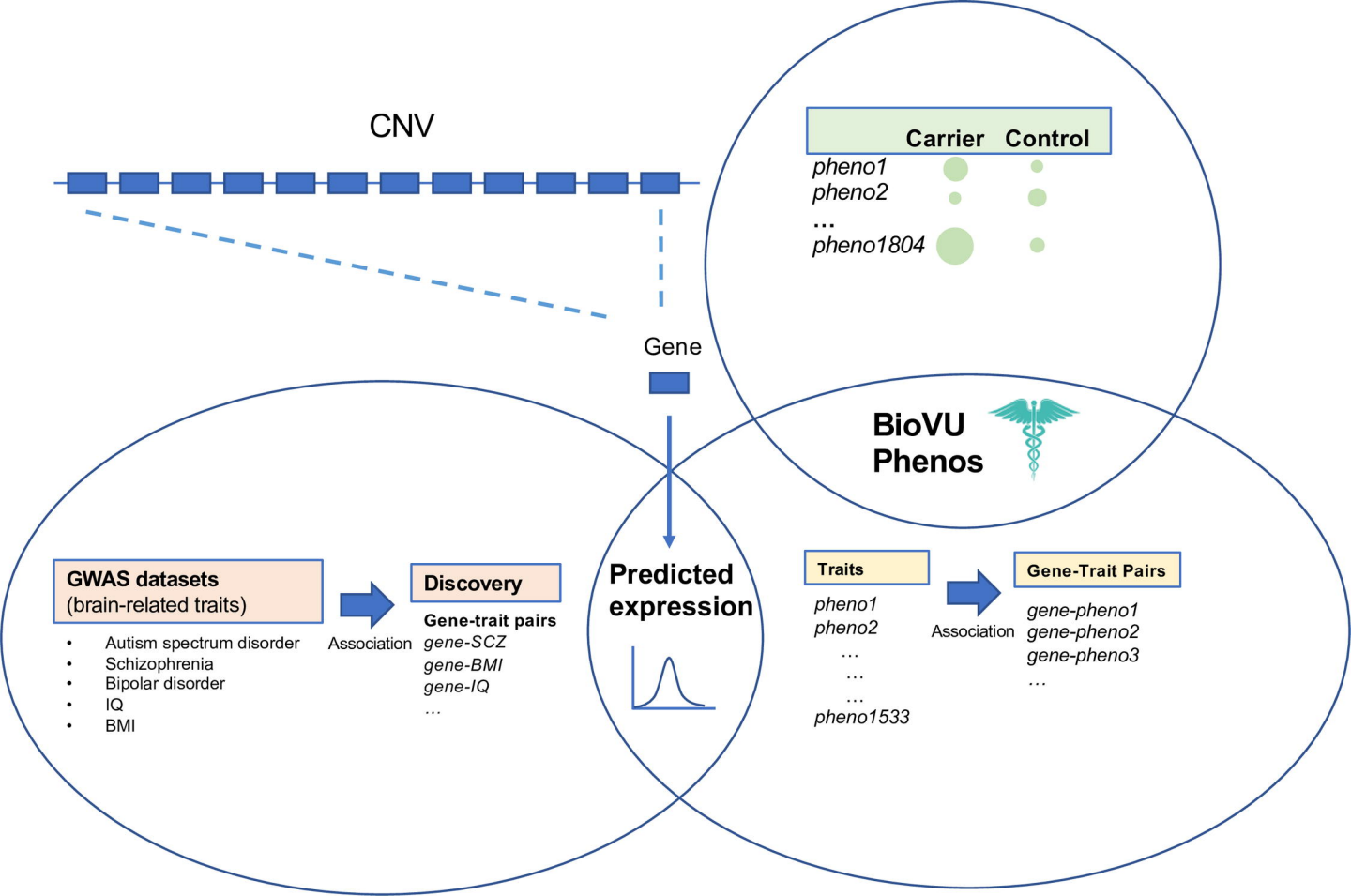
Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. Nat Genet. 2012;44: 639–641. doi:10.1038/ng.2262
29. Jerome LA, Papaioannou VE. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat Genet. 2001;27: 286–291. doi:10.1038/85845
30. Lindsay EA, Vitelli F, Su H, Morishima M, Huynh T, Pramparo T, et al. Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. Nature. 2001;410: 97–101. doi:10.1038/35065105
31. Merscher S, Funke B, Epstein JA, Heyer J, Puech A, Lu MM, et al. TBX1 Is Responsible for Cardiovascular Defects in Velo-Cardio-Facial/DiGeorge Syndrome. Cell. 2001;104: 619–629. doi:10.1016/S0092-8674(01)00247-1
32. Paylor R, Glaser B, Mupo A, Ataliotis P, Spencer C, Sobotka A, et al. Tbx1 haploinsufficiency is linked to behavioral disorders in mice and humans: implications for 22q11 deletion syndrome. Proc Natl Acad Sci U S A. 2006;103: 7729–34. doi:10.1073/pnas.0600206103
33. Ma G, Shi Y, Tang W, He Z, Huang K, Li Z, et al. An association study between the genetic polymorphisms within TBX1 and schizophrenia in the Chinese population. Neurosci Lett. 2007. doi:10.1016/j.neulet.2007.07.055
34. Clements CC, Wenger TL, Zoltowski AR, Bertollo JR, Miller JS, de Marchena AB, et al. Critical region within 22q11.2 linked to higher rate of autism spectrum disorder. Mol Autism. 2017;8: 58. doi:10.1186/s13229-017-0171-7
35. Crepel A, Steyaert J, De la Marche W, De Wolf V, Fryns J-P, Noens I, et al. Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. Am J Med Genet Part B Neuropsychiatr Genet. 2011;156: 243–245. doi:10.1002/ajmg.b.31163
36. Pucilowska J, Vithayathil J, Tavares EJ, Kelly C, Colleen Karlo J, Landreth GE. The 16p11.2 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain cytoarchitecture linked to the ERK MAPK pathway. J Neurosci. 2015;35: 3190–3200. doi:10.1523/JNEUROSCI.4864-13.2015
37. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, et al. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature. 2012;485: 363–7. doi:10.1038/nature11091
38. Blaker-Lee A, Gupta S, McCammon JM, De Rienzo G, Sive H. Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. Dis Model Mech. 2012;5.
39. Iyer J, Singh MD, Jensen M, Patel P, Pizzo L, Huber E, et al. Pervasive genetic interactions modulate neurodevelopmental defects of the autism-Associated 16p11.2 deletion in Drosophila melanogaster. Nat Commun. 2018;9: 1–19. doi:10.1038/s41467-018-04882-6
40. Paylor R, McIlwain KL, McAninch R, Nellis A, Yuva-Paylor LA, Baldini A, et al. Mice deleted for the DiGeorge/velocardiofacial syndrome region show abnormal sensorimotor gating and learning and memory impairments. Hum Mol Genet. 2001;10: 2645–2650. doi:10.1093/hmg/10.23.2645
41. Guna A, Butcher NJ, Bassett AS. Comparative mapping of the 22q11.2 deletion region and the potential of simple model organisms. J Neurodev Disord. 2015;7: 18. doi:10.1186/s11689-015-9113-x
42. McCammon JM, Blaker-Lee A, Chen X, Sive H. The 16p11.2 homologs fam57ba and doc2a generate certain brain and body phenotypes. Hum Mol Genet. 2017;26: 3699–3712. doi:10.1093/hmg/ddx255
43. Arbogast T, Ouagazzal A-M, Chevalier C, Kopanitsa M, Afinowi N, Migliavacca E, et al. Reciprocal Effects on Neurocognitive and Metabolic Phenotypes in Mouse Models of 16p11.2 Deletion and Duplication Syndromes. Barsh GS, editor. PLOS Genet. 2016;12:

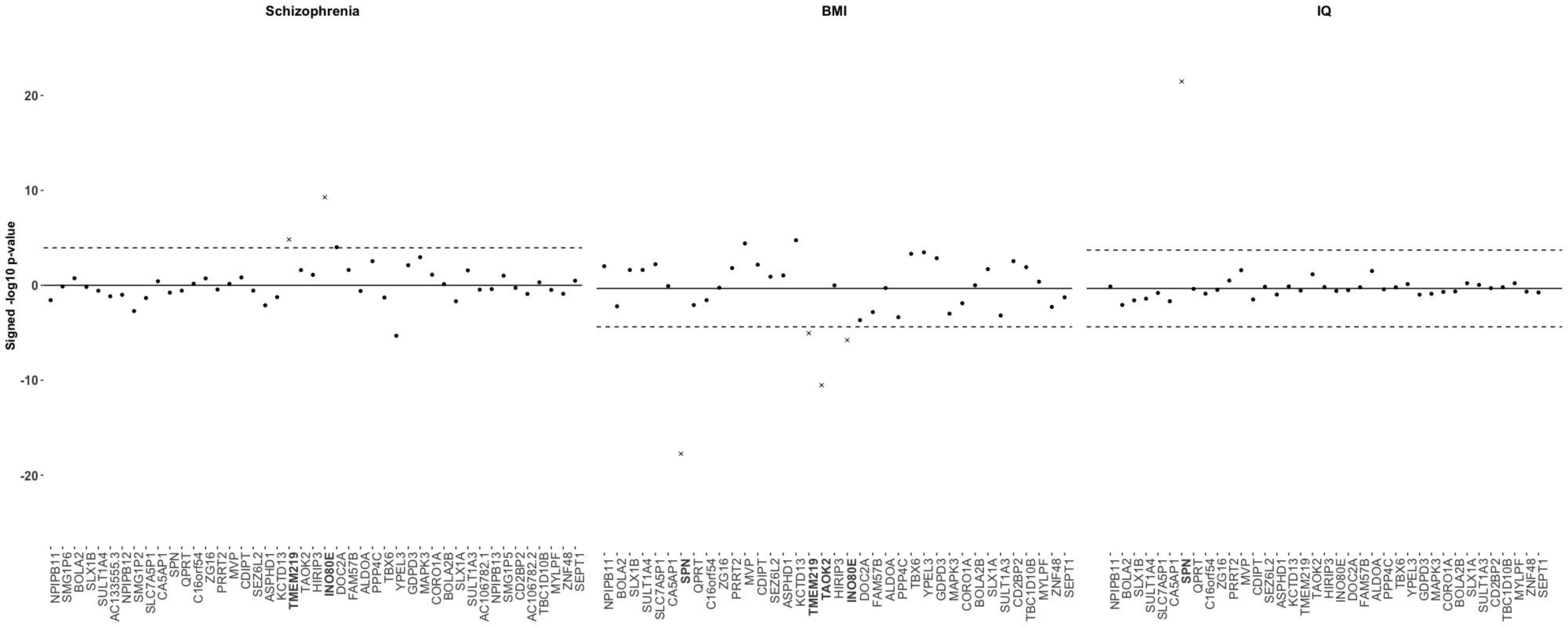
e1005709. doi:10.1371/journal.pgen.1005709
44. Ward TR, Zhang X, Leung LC, Zhou B, Muench K, Roth JG, et al. Genome-wide molecular effects of the neuropsychiatric 16p11 CNVs in an iPSC-to-iN neuronal model. bioRxiv. 2020; 2020.02.09.940965. doi:10.1101/2020.02.09.940965
45. Blumenthal I, Ragavendran A, Erdin S, Klei L, Sugathan A, Guide JR, et al. Transcriptional Consequences of 16p11.2 Deletion and Duplication in Mouse Cortex and Multiplex Autism Families. Am J Hum Genet. 2014;94: 870–883. doi:10.1016/j.ajhg.2014.05.004
46. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, et al. Genome-wide Transcriptome Profiling Reveals the Functional Impact of Rare De Novo and Recurrent CNVs in Autism Spectrum Disorders. Am J Hum Genet. 2012;91: 38–55. doi:10.1016/j.ajhg.2012.05.011
47. Zhang X, Zhang Y, Zhu X, Purmann C, Haney MS, Ward T, et al. Local and global chromatin interactions are altered by large genomic deletions associated with human brain development. Nat Commun. 2018;9: 5356. doi:10.1038/s41467-018-07766-x
48. Jalbrzikowski M, Lazaro MT, Gao F, Huang A, Chow C, Geschwind DH, et al. Transcriptome Profiling of Peripheral Blood in 22q11.2 Deletion Syndrome Reveals Functional Pathways Related to Psychosis and Autism Spectrum Disorder. van Amelsvoort T, editor. PLoS One. 2015;10: e0132542. doi:10.1371/journal.pone.0132542
49. Migliavacca E, Golzio C, Männik K, Blumenthal I, Oh EC, Harewood L, et al. A Potential Contributory Role for Ciliary Dysfunction in the 16p11.2 600 kb BP4-BP5 Pathology. Am J Hum Genet. 2015;96: 784–796. doi:10.1016/j.ajhg.2015.04.002
50. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. bioRxiv. 2019; 787903. doi:10.1101/787903
51. Aguet F, Ardlie KG, Cummings BB, Gelfand ET, Getz G, Hadley K, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550: 204–213. doi:10.1038/nature24277
52. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics. Nature Publishing Group; 2013. pp. 580–585. doi:10.1038/ng.2653
53. Ardlie KG, DeLuca DS, Segrè A V., Sullivan TJ, Young TR, Gelfand ET, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science (80-). 2015;348: 648–660. doi:10.1126/science.1262110
54. Freund MK, Burch K, Shi H, Mancuso N, Kichaev G, Garske KM, et al. Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. [cited 20 May 2018]. doi:10.1101/324558
55. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell. Elsevier; 2011. pp. 32–43. doi:10.1016/j.cell.2011.09.008
56. Blair DR, Lyttle CS, Mortensen JM, Bearden CF, Jensen AB, Khiabanian H, et al. A nondegenerate code of deleterious variants in mendelian loci contributes to complex disease risk. Cell. 2013;155: 70–80. doi:10.1016/j.cell.2013.08.030
57. Schizophrenia Working Group of the Psychiatric Genomics Consortium S. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511: 421–7. doi:10.1038/nature13595
58. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet. 2019;51: 793–803. doi:10.1038/s41588-019-0397-8
59. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51: 431–

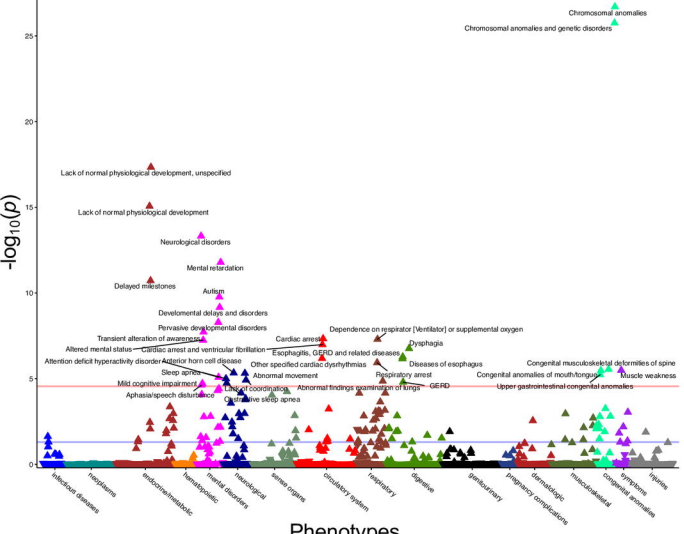
444. doi:10.1038/s41588-019-0344-8
60. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518: 197–206. doi:10.1038/nature14177
61. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018;50: 912–919. doi:10.1038/s41588-018-0152-6
62. Gamazon ER, Wheeler HE, Shah KP, Mozaffari S V., Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47: 1091–1098. doi:10.1038/ng.3367
63. Barbeira AN, Pividori MD, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. Plagnol V, editor. PLOS Genet. 2019;15: e1007889. doi:10.1371/journal.pgen.1007889
64. Roden D, Pulley J, Basford M, Bernard G, Clayton E, Balser J, et al. Development of a Large-Scale De-Identified DNA Biobank to Enable Personalized Medicine. Clin Pharmacol Ther. 2008;84: 362–369. doi:10.1038/clpt.2008.89
65. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008;84: 362–369. doi:10.1038/clpt.2008.89
66. Dantas AG, Santoro ML, Nunes N, de Mello CB, Pimenta LSE, Meloni VA, et al. Downregulation of genes outside the deleted region in individuals with 22q11.2 deletion syndrome. Hum Genet. 2019;138: 93–103. doi:10.1007/s00439-018-01967-6
67. UK Biobank — Neale lab. [cited 28 Mar 2020]. Available: http://www.nealelab.is/uk-biobank
68. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31: 1102–1111. doi:10.1038/nbt.2749
69. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010/03/24. 2010;26: 1205–1210. doi:10.1093/bioinformatics/btq126
70. Fernandez BA, Roberts W, Chung B, Weksberg R, Meyn S, Szatmari P, et al. Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. J Med Genet. 2010;47: 195–203. doi:10.1136/jmg.2009.069369
71. Karunanithi Z, Vestergaard EM, Lauridsen MH. Transposition of the great arteries - a phenotype associated with 16p11.2 duplications? World J Cardiol. 2017;9: 848–852. doi:10.4330/wjc.v9.i12.848
72. Umeki I, Niihori T, Abe T, Kanno S, Okamoto N, Mizuno S, et al. Delineation of LZTR1 mutation-positive patients with Noonan syndrome and identification of LZTR1 binding to RAF1–PPP1CB complexes. Hum Genet. 2018; 1–15. doi:10.1007/s00439-018-1951-7
73. Carmel M, Zarchi O, Michaelovsky E, Frisch A, Patya M, Green T, et al. Association of COMT and PRODH gene variants with intelligence quotient (IQ) and executive functions in 22q11.2DS subjects. J Psychiatr Res. 2014;56: 28–35. doi:10.1016/j.jpsychires.2014.04.019
74. Magini P, Smits DJ, Vandervore L, Schot R, Columbaro M, Kasteleijn E, et al. Loss of SMPD4 Causes a Developmental Disorder Characterized by Microcephaly and Congenital Arthrogryposis. Am J Hum Genet. 2019;105: 689–705. doi:10.1016/j.ajhg.2019.08.006

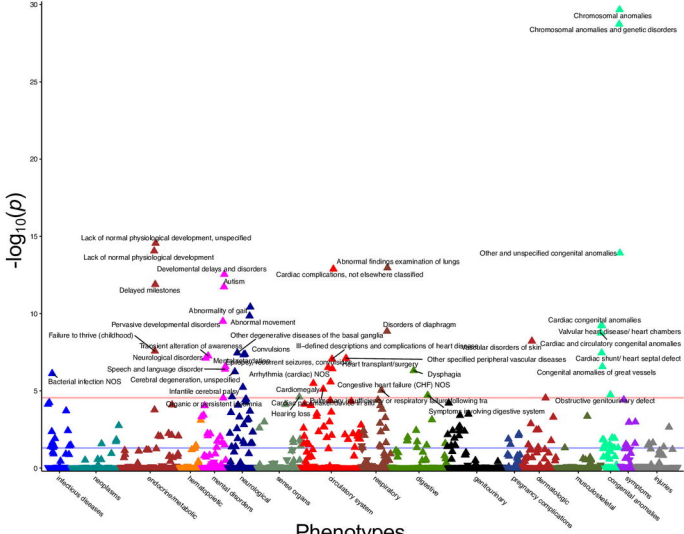
75. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE. Mouse Genome Database (MGD) 2019. Nucleic Acids Res. 2018;47: D801–D806. doi:10.1093/nar/gky1056
76. Powell CM, Miyakawa T. Schizophrenia-relevant behavioral testing in rodent models: a uniquely human disorder? Biol Psychiatry. 2006;59: 1198–207. doi:10.1016/j.biopsych.2006.05.008
77. Dobbyn A, Huckins LM, Boocock J, Sloofman LG, Glicksberg BS, Giambartolomei C, et al. Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. Am J Hum Genet. 2018;102: 1169–1184. doi:10.1016/J.AJHG.2018.04.011
78. Gusev A, Mancuso N, Won H, Kousi M, Finucane HK, Reshef Y, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat Genet. 2018;50: 538–548. doi:10.1038/s41588-018-0092-1
79. Pallant A, Eskenazi A, Mattei MG, Fournier REK, Carlsson SR, Fukuda M, et al. Characterization of cDNAs encoding human leukosialin and localization of the leukosialin gene to chromosome 16. Proc Natl Acad Sci U S A. 1989;86: 1328–1332. doi:10.1073/pnas.86.4.1328
80. Park JK, Rosenstein YJ, Remold-O'Donnell E, Bierer BE, Rosen FS, Burakoff SJ. Enhancement of T-cell activation by the CD43 molecule whose expression is defective in Wiskott-Aldrich syndrome. Nature. 1991;350: 706–709. doi:10.1038/350706a0
81. Marioni RE, Yang J, Dykiert D, Mõttus R, Campbell A, Davies G, et al. Assessing the genetic overlap between BMI and cognitive function. Mol Psychiatry. 2016;21: 1477–1482. doi:10.1038/mp.2015.205
82. Sabia S, Kivimaki M, Shipley MJ, Marmot MG, Singh-Manoux A. Body mass index over the adult life course and cognition in late midlife: the Whitehall II Cohort Study. Am J Clin Nutr. 2009;89: 601–607. doi:10.3945/ajcn.2008.26482
83. Ikeda M, Tanaka S, Saito T, Ozaki N, Kamatani Y, Iwata N. Re-evaluating classical body type theories: Genetic correlation between psychiatric disorders and body mass index. Psychol Med. 2018;48: 1745–1748. doi:10.1017/S0033291718000685
84. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47: 1236–1241. doi:10.1038/ng.3406
85. Mizuno A, Okada Y. Biological characterization of expression quantitative trait loci (eQTLs) showing tissue-specific opposite directional effects. Eur J Hum Genet. 2019. doi:10.1038/s41431-019-0468-4
86. Ben J, Jiang B, Wang D, Liu Q, Zhang Y, Qi Y, et al. Major vault protein suppresses obesity and atherosclerosis through inhibiting IKK–NF-κB signaling mediated inflammation. Nat Commun. 2019;10: 1801. doi:10.1038/s41467-019-09588-x
87. Lehuédé C, Li X, Dauvillier S, Vaysse C, Franchet C, Clement E, et al. Adipocytes promote breast cancer resistance to chemotherapy, a process amplified by obesity: Role of the major vault protein (MVP). Breast Cancer Res. 2019;21. doi:10.1186/s13058-018-1088-6
88. Gokalp H. Psychosocial Aspects of Hair Loss. Hair and Scalp Disorders. InTech; 2017. doi:10.5772/66156
89. Mercke Y, Sheng H, Khan T, Lippmann S. Hair loss in psychopharmacology. Annals of Clinical Psychiatry. Ann Clin Psychiatry; 2000. pp. 35–42. doi:10.1023/A:1009074926921
90. Curtis LA, Dolan TS, Cespedes RD. Acute urinary retention and urinary incontinence. Emergency Medicine Clinics of North America. W.B. Saunders; 2001. pp. 591–620. doi:10.1016/S0733-8627(05)70205-4
91. Kwiatkowski TJ, Bosco DA, LeClerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science (80-). 2009;323: 1205–1208. doi:10.1126/science.1166066

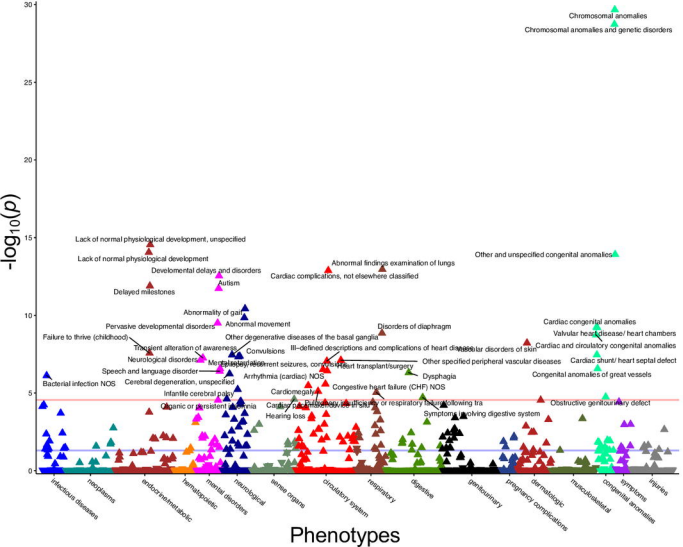
92. Tuttle R, Miller KR, Maiorano JN, Termuhlen PM, Gao Y, Berberich SJ. Novel senescence associated gene, YPEL3, is repressed by estrogen in ER+ mammary tumor cells and required for tamoxifen-induced cellular senescence. Int J Cancer. 2012;130: 2291–2299. doi:10.1002/ijc.26239
93. Shen Y, Chen X, Wang L, Guo J, Shen J, An Y, et al. Intra-family phenotypic heterogeneity of 16p11.2 deletion carriers in a three-generation Chinese family. Am J Med Genet Part B Neuropsychiatr Genet. 2011;156: 225–232. doi:10.1002/ajmg.b.31147
94. Ghebranious N, Giampietro PF, Wesbrook FP, Rezkalla SH. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. Am J Med Genet Part A. 2007;143: 1462–1471. doi:10.1002/ajmg.a.31837
95. Umeki I, Niihori T, Abe T, Kanno S ichiro, Okamoto N, Mizuno S, et al. Delineation of LZTR1 mutation-positive patients with Noonan syndrome and identification of LZTR1 binding to RAF1–PPP1CB complexes. Hum Genet. 2019;138: 21–35. doi:10.1007/s00439-018-1951-7
96. Forstner AJ, Degenhardt F, Schratt G, Nöthen MM. MicroRNAs as the cause of schizophrenia in 22q11.2 deletion carriers, and possible implications for idiopathic disease: a mini-review. Front Mol Neurosci. 2013;6: 47. doi:10.3389/fnmol.2013.00047
97. De la Morena MT, Eitson JL, Dozmorov IM, Belkaya S, Hoover AR, Anguiano E, et al. Signature MicroRNA expression patterns identified in humans with 22q11.2 deletion/DiGeorge syndrome. Clin Immunol. 2013;147: 11–22. doi:10.1016/j.clim.2013.01.011
98. Schizophrenia Working Group of the Psychiatric Genomics Consortium {fname}. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511: 421–7. doi:10.1038/nature13595
99. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018;9: 1825. doi:10.1038/s41467-018-03621-1
100. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550: 204–213. doi:10.1038/nature24277
101. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26: 2190–2191. doi:10.1093/bioinformatics/btq340
102. Colbran LL, Gamazon ER, Zhou D, Evans P, Cox NJ, Capra JA. Inferred divergent gene regulation in archaic hominins reveals potential phenotypic differences. Nat Ecol Evol. 2019;3: 1598–1606. doi:10.1038/s41559-019-0996-x
103. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48: 1279–1283. doi:10.1038/ng.3643
104. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48: 1284–1287. doi:10.1038/ng.3656
105. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics. 2014/04/14. 2014;30: 2375–2376. doi:10.1093/bioinformatics/btu197
106. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R . Bioinformatics. 2014;30: 2811–2812. doi:10.1093/bioinformatics/btu393

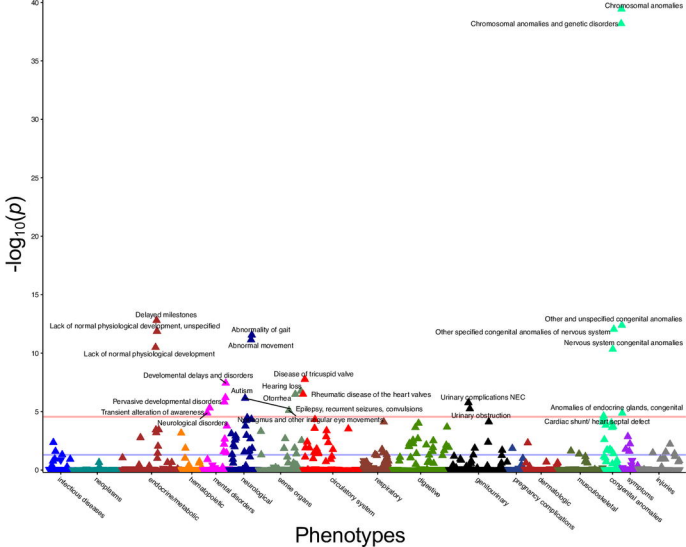

CNV
Gene
Carrier
Control
pheno1
pheno2
...
pheno1804
BioVU
Phenos
GWAS datasets
(brain-related traits)
Autism spectrum disorder
Schizophrenia
Bipolar disorder
IQ
BMI
Association
Discovery
Gene-trait pairs
gene-SCZ
gene-BMI
gene-IQ
...
Predicted
expression
Traits
pheno1
pheno2
...
...
...
pheno1533
Association
Gene-Trait Pairs
gene-pheno1
gene-pheno2
gene-pheno3
...

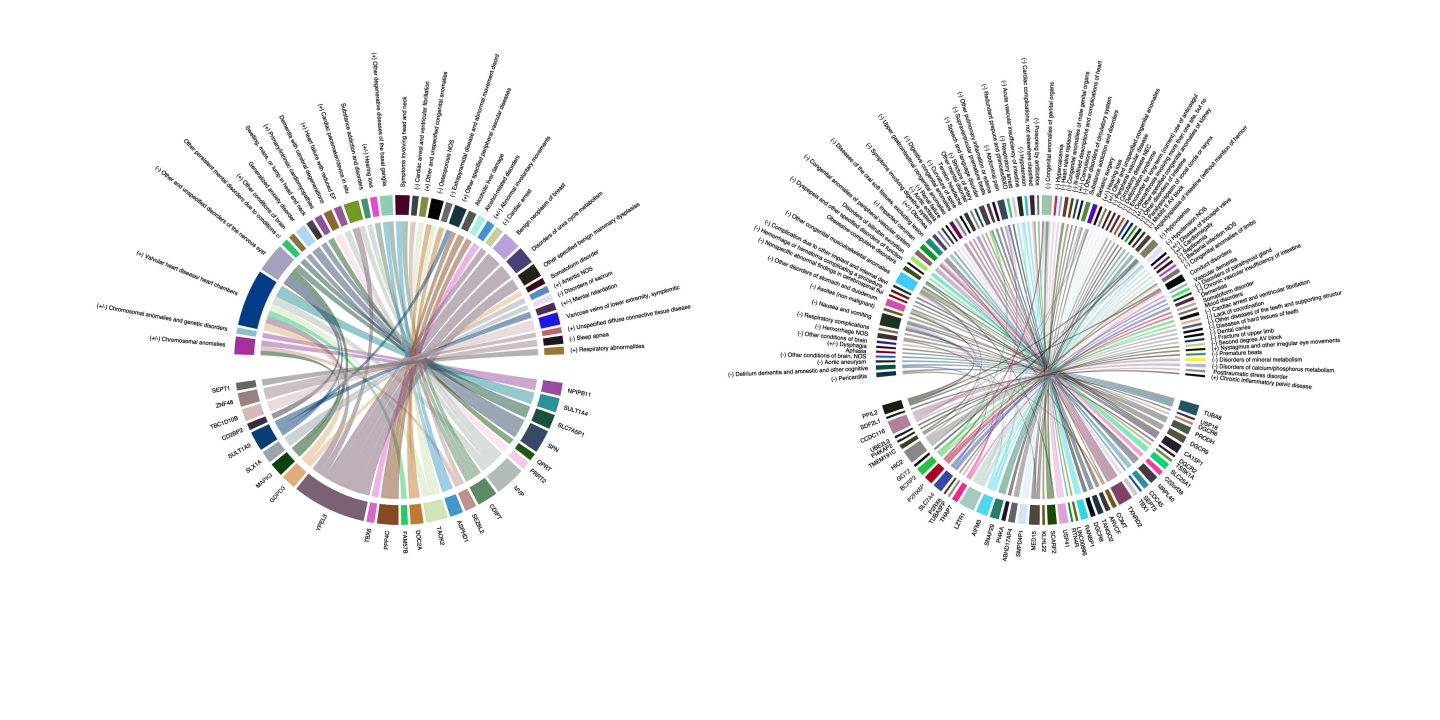

Schizophrenia
BMI
IQ
Signed -log10 p-value
20
10
0
-10
-20
NPIPB11
SMG1P6
BOLA2
SLX1B
SULT1A4
AC133555.3
NPIPB12
SMG1P2
SLC7A5P1
CA5AP1
SPN
QPRT
C16orf54
ZG16
PRRT2
MVP
CDIPT
SEZ6L2
ASPHD1
KCTD13
TMEM219
TAOK2
HIRIP3
INO80E
DOC2A
FAM57B
ALDOA
PPP4C
TBX6
YPEL3
GDPD3
MAPK3
CORO1A
BOLA2B
SLX1A
SULT1A3
AC106782.1
NPIPB13
SMG1P5
CD2BP2
AC106782.2
TBC1D10B
MYLPF
ZNF48
SEPT1


-log10(p)
Phenotypes
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
mental disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
pregnancy complications
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries
Chromosomal anomalies
Chromosomal anomalies and genetic disorders
Lack of normal physiological development, unspecified
Lack of normal physiological development
Neurological disorders
Mental retardation
Delayed milestones
Autism
Developmental delays and disorders
Pervasive developmental disorders
Transient alteration of awareness
Altered mental status
Cardiac arrest and ventricular fibrillation
Attention deficit hyperactivity disorder
Anterior horn cell disease
Sleep apnea
Mild cognitive impairment
Aphasia/speech disturbance
Other specified cardiac dysrhythmias
Abnormal movement
Lack of coordination
Obstructive sleep apnea
Cardiac arrest
Esophagitis, GERD and related diseases
Abnormal findings examination of lungs
Dependence on respirator [Ventilator] or supplemental oxygen
Dysphagia
Diseases of esophagus
Respiratory arrest
GERD
Congenital musculoskeletal deformities of spine
Congenital anomalies of mouth/tongue
Muscle weakness
Upper gastrointestinal congenital anomalies


-log10(p)
Phenotypes
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
mental disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
pregnancy complications
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries
Chromosomal anomalies
Chromosomal anomalies and genetic disorders
Lack of normal physiological development, unspecified
Lack of normal physiological development
Developmental delays and disorders
Autism
Delayed milestones
Abnormality of gait
Pervasive developmental disorders
Abnormal movement
Failure to thrive (childhood)
Transient alteration of awareness
Other degenerative diseases of the basal ganglia
Convulsions
Neurological disorders
Speech and language disorder
Cerebral degeneration, unspecified
Bacterial infection NOS
Cardiomegaly
Arrhythmia (cardiac) NOS
Congestive heart failure (CHF) NOS
Chronic or transient anemia
Hearing loss
Abnormal findings examination of lungs
Cardiac complications, not elsewhere classified
Other and unspecified congenital anomalies
Disorders of diaphragm
Ill-defined descriptions and complications of heart disease
Heart transplant/surgery
Dysphagia
Symptoms involving digestive system
Cardiac congenital anomalies
Valvular heart disease/ heart chambers
Cardiac and circulatory congenital anomalies
Other specified peripheral vascular diseases
Cardiac shunt/ heart septal defect
Congenital anomalies of great vessels
Obstructive genitourinary defect


-log10(p)
Phenotypes
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
mental disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
pregnancy complications
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries
Chromosomal anomalies
Chromosomal anomalies and genetic disorders
Lack of normal physiological development, unspecified
Lack of normal physiological development
Developmental delays and disorders
Autism
Delayed milestones
Abnormality of gait
Abnormal movement
Pervasive developmental disorders
Failure to thrive (childhood)
Other degenerative diseases of the basal ganglia
Transient alteration of awareness
Convulsions
Neurological disorders
Speech and language disorder
Cerebral degeneration, unspecified
Bacterial infection NOS
Infantile cerebral palsy
Organic or persistent amnesia
Cardiomegaly
Hearing loss
Abnormal findings examination of lungs
Cardiac complications, not elsewhere classified
Disorders of diaphragm
Ill-defined descriptions and complications of heart disease
Other specified peripheral vascular diseases
Heart transplant/surgery
Dysphagia
Congestive heart failure (CHF) NOS
Symptoms involving digestive system
Other and unspecified congenital anomalies
Cardiac congenital anomalies
Valvular heart disease/ heart chambers
Cardiac and circulatory congenital anomalies
Cardiac shunt/ heart septal defect
Congenital anomalies of great vessels
Obstructive genitourinary defect


-log10(p)
Phenotypes
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
mental disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
pregnancy complications
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries
Chromosomal anomalies
Chromosomal anomalies and genetic disorders
Delayed milestones
Lack of normal physiological development, unspecified
Abnormality of gait
Abnormal movement
Lack of normal physiological development
Developmental delays and disorders
Disease of tricuspid valve
Hearing loss
Autism
Otorrhea
Rheumatic disease of the heart valves
Pervasive developmental disorders
Epilepsy, recurrent seizures, convulsions
Transient alteration of awareness
Nystagmus and other involuntary eye movements
Neurological disorders
Urinary complications NEC
Urinary obstruction
Other and unspecified congenital anomalies
Other specified congenital anomalies of nervous system
Nervous system congenital anomalies
Anomalies of endocrine glands, congenital
Cardiac shunt/ heart septal defect


(+) Chronic degenerative disease of the lower gastrointestinal tract
Symptoms involving head and neck
(-) Cardiac arrest and ventricular fibrillation
(+) Other and unspecified congenital anomalies
(+) Osteoporosis NOS
(+) Endocrinology, nutrition and metabolic diseases
(+) Other specified neoplasms of lymphoid, hematopoietic and related tissue
Alcohol use disorder
(+) Alcohol-induced disorder
(+) Abnormal ventricular fibrillation
(-) Cardiac arrest
Benign neoplasm of breast
Disorders of urea cycle metabolism
Other specified benign mammary dysplasias
Somatoform disorder
(+) Arthritis NOS
(-) Disorders of sarcum
(+) Mental retardation
Varicose veins of lower extremity, symptomatic
(+) Unspecified diffuse connective tissue disease
(-) Sleep apnea
(+) Respiratory abnormalities
NPIPB11
SULT1A4
SLC7A5P1
SPN
OPRT
PRR15
KIF
CSRT
SCRE2
ASPD1
TAOK2
SCCDA
PARP8
DHAD
TBX6
TTRDL3
GUFCO
MAPKU
SLX4
SULT1A6
CD38P2
TBC1D19B
ZNF46
SEPT1
(+) Chromosomal anomalies
(+) Chromosomal anomalies and genetic disorders
(+) Vascular heart disease/ heart disorders
Other specified mental disorders due to conditions n.e.c.
(-) Other and unspecified disorders of the nervous syst
Sleep, wake, or arousal disorder
Generalized anxiety disorder
(+) Other condition of heart
(+) Pneumococcal pneumonia
Dermatitis and dermatoses NOS
(+) Heart failure with reduced EF
(+) Cardiac pneumoconiosis n.e.c.
Substance addiction and abusers
(+) Hearing loss
(+) Cardiac arrest and ventricular fibrillation
(-) Debrium dementia and amnestic and other cognitive
(-) Pericarditis
(-) Other conditions of brain
(+) Dysphagia
(-) Aortic aneurysm
(-) Other conditions of brain, NOS
(+) Respiratory complications
(-) Other conditions of brain
(-) Hemorrhage NOS
(+) Anemia (iron malignancy)
(-) Nausea and vomiting
(-) Other disorders of stomach and duodenum
(-) Hemorrhage in hemorrhagic cerebrovascular disorder
(-) Other congenital musculoskeletal anomalies
(-) Congestion due to other implant and internal devices
(-) Other, congenital musculoskeletal anomalies
Obstructive and restrictive lung disorders
(-) Congenital anomalies of peripheral nervous system
(-) Disorders of the anus, and rectum, excluding hernia
(-) Congenital anomalies of peripheral nervous system
(-) Disorders of the anus, and rectum, excluding hernia
(-) Congenital anomalies of peripheral nervous system
(-) Disorders and other specified disorders of the kidney
(-) Disorders and other specified disorders of the kidney
(-) Other, congenital musculoskeletal anomalies
(-) Congestion due to other implant and internal devices
(-) Hemorrhage in hemorrhagic cerebrovascular disorder
(-) Other disorders of stomach and duodenum
(-) Nausea and vomiting
(-) Respiratory complications
(-) Hemorrhage NOS
(+) Dysphagia
(-) Other conditions of brain, NOS
(-) Aortic aneurysm
(-) Other conditions of brain
(-) Pericarditis
(-) Debrium dementia and amnestic and other cognitive
PPL2
SDF2L2
CCDC116
DNAJB12
TMEM116
TMEM11C
HEX
GCK1
SLC22A1
PCDH20
SLC22A4
TUBA8
TMAP1
LTTH
APMA
SNAI2B
PHXA
ABHD17A
SMTNP1
MEC15
ACTN
ACAP2
ELF1
TMEM
DSG2B
DSG2B
TMEM207
TRAM2
MYH7
MYH6
DSG2
DSG2
CDC42
SLC4A1
SLC4A1
CDH2
CDH2
CDH2