

1 Combinations of genes at the 16p11.2 and 22q11.2 CNVs contribute to neurobehavioral traits

2

3 Mikhail Vysotskiy<sup>1,2,3</sup>, Autism Working Group of the Psychiatric Genomics Consortium<sup>^</sup>,

4 Bipolar Disorder Working Group of the Psychiatric Genomics Consortium<sup>^</sup>, Schizophrenia

5 Working Group of the Psychiatric Genomics Consortium<sup>^</sup>, Lauren A Weiss<sup>1,2,3\*</sup>

6

7 <sup>1</sup>Institute for Human Genetics, University of California San Francisco, San Francisco, CA, 94143

8 USA

9

10 <sup>2</sup>Department of Psychiatry and Behavioral Sciences, University of California San Francisco, San

11 Francisco, CA, 94143 USA

12

13 <sup>3</sup>Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA,

14 94143, USA

15

16 \* Corresponding author, Lauren A. Weiss

17 513 Parnassus Ave., HSE 901

18 San Francisco CA 94143

19 Phone: 415-476-7650

20 Fax: 415-476-7970

21 [lauren.weiss@ucsf.edu](mailto:lauren.weiss@ucsf.edu)

22

23 <sup>^</sup> A list of consortia members is provided in Supplementary note 1

24 *Abstract*

25 The 16p11.2 and 22q11.2 copy number variants (CNVs) are associated with neurobehavioral  
26 traits including autism spectrum disorder (ASD), schizophrenia, bipolar disorder, obesity, and  
27 intellectual disability. Identifying specific genes contributing to each disorder and dissecting the  
28 architecture of CNV-trait association has been difficult, inspiring hypotheses of more complex  
29 models, such as the effects of pairs of genes. We generated pairwise expression imputation  
30 models for CNV genes and then applied these models to GWAS for: ASD, bipolar disorder,  
31 schizophrenia, BMI (obesity), and IQ (intellectual disability). We compared the trait variance  
32 explained by pairs with the variance explained with single genes and with traditional interaction  
33 models. We also modeled polygene region-wide effects using summed ranks across all genes in  
34 the region. In all CNV-trait pairs except for bipolar disorder at 22q11.2, pairwise effects explain  
35 more variance than single genes, which was specific to the CNV region for all 16p11.2 traits and  
36 ASD at 22q11.2. We identified individual genes over-represented in top pairs that did not show  
37 single-gene signal. We also found that BMI and IQ have a significant association with a  
38 regionwide score. Genetic architecture differs by trait and region, but 9/10 CNV-trait  
39 combinations showed evidence for multigene contribution, and for most of these, the importance  
40 of combinatorial models appeared unique to CNV regions. Our findings suggest that mechanistic  
41 insights for CNV pathology may require combinatorial models.

42

43

44

45

46

47 *Introduction*

48

49 Copy number variants (CNVs) at 16p11.2 and 22q11.2 contribute to neurobehavioral disorders  
50 including autism spectrum disorder (ASD), schizophrenia, bipolar disorder, intellectual disability,  
51 and obesity [1,2,11,3–10]. Specific gene-trait contributions at these regions have proven difficult  
52 to find. Single-gene fine-mapping approaches have been difficult due to a lack of highly-penetrant  
53 point mutations in these genes and inconsistent findings in animal models [12–15]. A potential  
54 reason for the lack of clear gene-phenotype relationships is that the architecture may be more  
55 complicated than single-gene contributions to each trait [16]. More complex models are good  
56 candidates for *in silico* analysis, as multiple hypotheses can be efficiently assessed in parallel.

57

58 Data in humans and mice suggest that the expression of 16p11.2 and 22q11.2 CNV genes is  
59 consistently upregulated/downregulated in duplication/deletion carriers [17–20]. From this  
60 observation, we can propose that gene expression dysregulation (and potential downstream protein  
61 expression) is likely to be a pathophysiological mechanism of CNV-associated traits. This implies  
62 that examination of the consequences of gene expression variation may uncover the genetic  
63 architecture of CNV-phenotype association. However, gene expression data for cases affected with  
64 neurobehavioral traits remains limited in availability and ambiguous with respect to causality.  
65 Instead, we can use expression-imputation methodology to use genetic data, available for a far  
66 greater number of (control) individuals, to determine gene expression under the assumption that  
67 genetic regulation is similar in cases and controls. This method allows us to analyze the  
68 architecture at a gene level (rather than individual SNPs) and because it is based on germline  
69 genetics, is not affected by potential confounding influences on gene expression such as age,

70 chronic illness, medication use, and circumstances of death and tissue preservation. eQTLs (in  
71 our case, SNPs used for expression prediction) are less likely to affect genes in a context-dependent  
72 manner, as eQTL-linked genes are less likely to be affected by enhancer activity compared to  
73 GWAS-linked genes [21]. Given that our regions of interest have trait associations via CNVs but  
74 very limited GWAS signal for the same traits, using eQTLs and expression prediction is likely to  
75 find additional information missed by GWAS analyses.

76

77 Previously, we used expression imputation to test whether individual genes at the 16p11.2 and  
78 22q11.2 CNV regions were contributing to our five traits of interest (ASD, schizophrenia, bipolar  
79 disorder, intellectual disability, and obesity) [22]. We found contributions of *INO80E* to  
80 schizophrenia and body mass index (BMI) and of *SPN* to BMI and IQ, both at 16p11.2. However,  
81 no individual genes were associated with 22q11.2 traits, despite using equally-powered genetic  
82 datasets. No genes at 16p11.2 were significantly associated with ASD or bipolar disorder using  
83 our experiment-wide threshold. These lack of findings in light of the overall success of our  
84 approach were disappointing given the high prevalence of traits such as ASD in 16p11.2 CNV  
85 carriers and schizophrenia in 22q11.2 deletion carriers. One explanation for lack of gene-trait  
86 association is that individual genes may not be independent contributors to these traits, rather the  
87 genetic architecture is combinatorial. Promisingly, it was found that several pairs of 16p11.2 genes  
88 in *Drosophila* showed evidence of stronger effects on eye phenotypes than individual genes, and  
89 double mutants of 16p11.2 genes in zebrafish led to hyperactivity and body size phenotypes  
90 [15,23]. Thus, we aimed to investigate combinatorial associations in our traits of interest in  
91 humans.

92

93 In a CNV carrier, all genes within the breakpoints are duplicated or deleted, typically with a similar  
94 increase/decrease of expression across all genes. In our previous study, we considered the level of  
95 expression of any individual gene, and its effect on relevant phenotypes in non-carriers. Here, we  
96 consider two additional models in non-carriers (Figure 1). First, as a feasible way to model  
97 multigene effects at specific pairs of genes, for each gene pair we look for trait association with  
98 expression increases or decreases across two genes. Second, we analyze association patterns when  
99 gene expression trends towards being upregulated or downregulated across the whole region as a  
100 way to capture effects of more than two genes.

101

## 102 *Methods*

103

### 104 Genes studied

105

106 We selected genes at the 16p11.2 and 22q11.2 CNV regions that fell into one of these annotation  
107 categories: protein-coding, lincRNA, pseudogene, antisense, miRNA. These were consistent with  
108 what was used for PrediXcan modeling previously, with miRNA included given the strong  
109 representation of miRNAs at 22q11.2 [24,25]. We included noncoding genes, as they have not  
110 received significant attention in studies of these regions, despite some evidence of miRNA  
111 contribution to 22q11.2 phenotypes. In addition, we considered flanking genes within 200kb of  
112 the region, as there is suggestive evidence of broader transcriptional effects in CNV carriers, and  
113 because we previously found evidence of flanking gene involvement in psychosis [22,27].  
114 Supplementary tables 1 and 2 contain single and pairwise CNV genes used in analysis.

115

116 Phenotypes and datasets

117

118 Imputed genotypes from the Psychiatric Genomics Consortium were used to study schizophrenia  
119 (wave 3 freeze), bipolar disorder (wave 2 freeze), and ASD (2019 freeze, used for analysis of  
120 variance explained only) [28–30]. An additional joint PGC-iPsych ASD summary statistic set  
121 was used to boost power for ASD analyses ([www.med.unc.edu/pgc/download-results/](http://www.med.unc.edu/pgc/download-results/)) [30].  
122 Summary statistics from the GIANT consortium (2015 freeze,  
123 [www.portals.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_files](http://www.portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files))  
124 were used to study BMI, and a VU-Amsterdam University cohort (wave 2 freeze,  
125 [www.ctg.cncr.nl/software/summary\\_statistics](http://www.ctg.cncr.nl/software/summary_statistics)) was used for IQ [31,32]. Individual-level IQ and  
126 BMI data from the UK Biobank were used for validating discoveries in individual-level data on  
127 phenotypes for which individual-level data were not available [33]. Cohorts and sample sizes are  
128 listed in Supplementary table 3.

129

130 Predicting the expression levels of individual 16p11.2 and 22q11.2 CNV genes

131

132 Analyses of single genes were performed using elastic net models from [www.predictdb.org](http://www.predictdb.org)  
133 trained on the GTEx version 8 data [34]. These prediction models were available for up to 42  
134 16p11.2 genes and up to 65 22q11.2 genes in at least one tissue. The elastic net approach was  
135 chosen for consistency with our pairwise model training approach. Gene expression for each  
136 CNV gene in each individual was predicted using the --predict option in PrediXcan, for each  
137 cohort [35]. Analyses on summary statistics did not require expression prediction.

138

139 Finding control gene sets

140

141 To create control gene sets to use in a permutation-based analysis, the 16p11.2 and 22q11.2  
142 regions were matched on three categories: (1) number of genes (exact), (2) length of the region  
143 (in bases, 80-120% of the region), (3) ratio of coding to non-coding genes (at least 80% that of  
144 the region to avoid picking up dense regions of noncoding genes). Gene sets that overlapped the  
145 distal 16p11.2 region or the Major Histocompatibility Complex (a known gene-dense major  
146 GWAS-identified locus for schizophrenia) were excluded [36]. Overall, we found 38 comparable  
147 regions to 16p11.2 and 28 to 22q11.2. The list of control regions can be found in Supplementary  
148 table 4.

149

150 Predicting the expression of pairs of 16p11.2 and 22q11.2 CNV genes

151

152 As a simple way to model pairwise gene expression, we took every pair of genes in each CNV  
153 and defined pairwise “joint genes” with expression equal to the inverse-normalized sum of the  
154 expressions of each gene in GTEx. Normalized GTEx gene expression sums were used as inputs  
155 to the PrediXcan elastic net model training pipeline  
156 ([www.github.com/hakyimlab/PredictDB\\_Pipeline\\_GTEx\\_v7](https://www.github.com/hakyimlab/PredictDB_Pipeline_GTEx_v7)), with covariates used for the  
157 GTEx v8 analyses downloaded from [www.gtexportal.org/home/datasets](https://www.gtexportal.org/home/datasets). Given that our goal was  
158 to evaluate the contribution of these pairwise genes to specific traits, rather than a general-use  
159 pairwise model training process, a high overlap between the SNPs in our models and the GWAS  
160 datasets was vital. For that reason, we chose to repeat the training process for each trait, leaving

161 only the SNPs in each GWAS dataset as inputs for model training. We repeated this model  
162 training process again on the control pairs of genes.

163

164 Association studies between predicted expression and traits

165

166 Individual level

167

168 Each PGC cohort was converted from PLINK to dosage format for PrediXcan input. Tissue-  
169 specific prediction models were applied to each tissue in each cohort. MultiXcan, a cross-tissue  
170 implementation of PrediXcan, was used to combine predicted expressions across tissues and  
171 perform association with trait [37]. Using the MultiXcan p-value and the average direction of  
172 effect of each gene across tissues, we used METAL to determine a per-gene result [38]. Both  
173 single gene and pairwise analyses were performed in the same way.

174

175 Summary level

176

177 The ‘MetaMany’ option in the MetaXcan package was applied to summary-level data using  
178 single-tissue prediction models to generate gene-level association results for each tissue [39]. S-  
179 MultiXcan, a cross-tissue implementation of PrediXcan for summary level data, was used to  
180 combine across tissues for tissue-wide association results [37]. Cross-tissue covariances were  
181 downloaded from PredictDB for single-gene models and generated from single tissue  
182 covariances for pairwise models using the covariance builder script available at



183 [www.github.com/hakyimlab/MetaXcan/blob/master/software/CovarianceBuilder.py](https://www.github.com/hakyimlab/MetaXcan/blob/master/software/CovarianceBuilder.py). Both  
184 single gene and pairwise analyses of summary statistics were performed in the same way.

185

#### 186 UK Biobank additional expression prediction

187 While the best-powered GWAS meta-analyses of BMI and IQ were available as summary  
188 statistics, certain analyses such as interaction models and percent variance explained require  
189 individual-level data. We obtained IQ and BMI measurements from the UK Biobank and took an  
190 average across visits for people with multiple measurements. Analysis was limited to individuals  
191 who reported their ethnicity as “white”, and included age, age-squared, and 40 principal  
192 components as covariates. A large number of principal components was used due to the  
193 correlation between the BMI phenotype and components in the PC 30-40 range. Expression  
194 imputation for single genes and pairs was performed with PrediXcan as described above.

195

#### 196 Significance thresholds for association studies

197

198 For all association studies, a permutation-based threshold was determined using the control gene  
199 sets. After association testing between control gene sets for a CNV and phenotype, the median of  
200 the 5<sup>th</sup> percentile of control sets was used as a 5% significance threshold for the true CNV  
201 region. As control genes are chosen independently of association with trait, using a median  
202 across all regions will counteract bias caused by any control gene set overlapping a strong  
203 GWAS peak for a trait.

204

#### 205 Estimating variance explained by pairwise models

206  
207 Variance in phenotype explained by imputed expression was measured as the  $R^2$  of the linear  
208 model between case-control status and imputed expression for all genes in the CNV.  
209 Specifically, the adjusted  $R^2$  was used, as using all pairs of genes involves a large number of  
210 variables. For every tissue-cohort pair,  $R^2$  values were calculated using all single genes, all  
211 pairwise genes, and interaction terms. The number of times a model (single, pairwise, or  
212 interaction) had the greatest  $R^2$  for a cohort-tissue pair was tallied. The same process was  
213 implemented for control gene sets. A chi-square test was performed to determine whether the  
214 proportion of pairwise models being “best” in a CNV region was different from the proportion in  
215 control regions. This approach required individual-level data, and as we used summary level data  
216 for ASD, IQ, and BMI, we performed it in PGC ASD individual-level data (without iPSYCH),  
217 and UK Biobank for IQ and BMI (each of which was treated as one single cohort).

218  
219 We acknowledge that previous attempts to solve the problem of variance explained by predicted  
220 expression were made by Liang et al [40]. We attempted this method and found extremely large  
221 estimates for variance explained. This inflation might be due to our relatively small (<5 MB)  
222 regions of interest with high SNP and predicted-expression correlation structure, as opposed to a  
223 predicted transcriptome-wide screen. The estimates provided by our approach, where the  
224 adjusted  $R^2$  rarely exceeds 0.05, are a more reasonable estimate of the effect of one small set of  
225 genes on a trait.

226

227 Testing a region-wide model

228 We estimated a region-wide score for each individual using their single-gene predicted  
229 expressions. First, we found the normalized rank of an individual for the expression of a gene;  
230 the median rank was used for genes expressed in multiple tissues. The sum of an individual's  
231 gene-specific rankings became the individual's region-wide score; these scores were converted  
232 to normalized (between 0 and 1) ranks. For quantitative traits, we quantified the relationship  
233 between score and phenotype as a Pearson correlation. For binary traits, we tested for a  
234 difference in score distribution between cases and controls (Kolmogorov-Smirnov test), as well  
235 as for a difference in score means between cases and controls (t-test).

236  
237 Initially, we attempted to approach region-wide association in the same way as pairwise  
238 association for schizophrenia. Region-wide sums of GTEx expressions of all CNV genes were  
239 used as inputs into elastic net models from GTEx SNPs, with the same covariates as before.  
240 After model quality filtering, models in only 5 tissues at 16p11.2 and 13 tissues at 22q11.2  
241 remained, all with  $R^2 < 0.1$ . As a result, we did not further pursue this method.

242

## 243 *Results*

244

### 245 Summary of individual gene results

246

247 We have updated our single-gene prioritization from our previous study using new models from  
248 GTEx version 8 and new data from schizophrenia PGC wave 3 (Table 1) [28,34]. With this  
249 enhancement, we find one 22q11.2 gene (*PPIL2*) significantly associated with schizophrenia at a  
250 permutation-based threshold (Table 1, Supplementary table 5). We note that the permutation-

251 based threshold is less conservative than the experiment-wide thresholds used in previous  
252 analysis [22]. However, we can identify five top genes at 22q11.2 associated with BMI (*YDJC*,  
253 *CCDC116*, *PPIL2*, *THAP7*, *UBE2L3*), primarily located outside the canonical CNV region (LCR  
254 D-E), three with bipolar disorder (*TMEM191B*, *TUBA8*, *PPIL2*), six with ASD (*CLTCLI*,  
255 *AC004471.10*, *UFD1L*, *DGCR14*, *CCDC188*, *DGCR9*), and two with IQ (*SEPT5*, *LINC00896*)  
256 (Table 1, Supplementary table 5). The top genes associated with ASD at 22q11.2 are located in  
257 the LCR A-B part of the variant, consistent with a previous study [41]. At 16p11.2, the majority  
258 of genes tested (30/38) show an association with BMI. We find that, after updating single-gene  
259 prediction models to GTEx v8, *SPN* is no longer a major driver of BMI and IQ, as the best  
260 predictive SNPs in the most up-to-date version of GTEx did not overlap with top *SPN* SNPs as  
261 before; however, new models for *SULT1A4* indicated this gene as a major contributor to both  
262 BMI and IQ (Table 1, Supplementary table 5). *INO80E* and *KCTD13* remained associated with  
263 BMI. We find that *INO80E* is a top association with bipolar disorder and ASD; this gene  
264 previously showed suggestive bipolar disorder association but did not meet experiment-wide  
265 significance criteria even with the updated models [22].

266

### 267 Predicting expression of pairs of 16p11.2 and 22q11.2 genes

268

269 We trained elastic net models for pairs of 16p11.2 and 22q11.2 genes (both coding and non-  
270 coding when possible) using dataset-specific SNP lists to maximize overlap. In general, the  
271 model quality (as measured by the performance  $R^2$ ) of pairwise models was in-between that of  
272 the two genes that it comprised, as expected (Supplementary table 2). In addition, we trained  
273 pairwise models for control gene regions (N=38 for 16p11.2 and N=28 for 22q11.2).

274

275 Pairwise prediction models explain more trait variance than single-gene or interaction models

276

277 To assess whether analyzing pairs of genes provided more information than individual genes, we  
278 calculated how much variance in CNV-associated traits was explained by predicted gene  
279 expression as the adjusted  $R^2$  of linear models of individual gene expression predictions,  
280 pairwise additive gene expression predictions, and pairwise interaction models. We calculated  
281 the proportion of tissue-cohort pairs for which pairwise gene expression was the best predictor.  
282 In all trait-region pairs, with the exception of bipolar disorder at 22q11.2, we found that the trait  
283 variance explained was greater for gene pairs proportionally more often than either single genes  
284 or interactions (Table 2). To confirm whether this phenomenon was CNV region-specific and not  
285 a polygenic property of the trait, we additionally performed this analysis for control gene sets.  
286 For all traits tested at 16p11.2, the proportion of pairwise models exceeding single or interaction  
287 was greater than that of control regions ( $P < 0.05$ ). At 22q11.2, the CNV region performed better  
288 than control regions in ASD ( $P = 1.3 \times 10^{-9}$ ) and BMI ( $P < 2.2 \times 10^{-16}$ ), but schizophrenia, bipolar  
289 disorder, and IQ showed similar proportion pairwise in control regions (Table 2).

290

291 Pairwise association signal is oligogenic

292

293 Using our pairwise models to perform association analysis, we found that there were 269  
294 16p11.2 and 278 22q11.2 pairs significantly associated with ASD, 204 16p11.2 and 132 22q11.2  
295 pairs associated with bipolar disorder, 695 16p11.2 and 129 22q11.2 pairs associated with  
296 schizophrenia, 74 16p11.2 and 30 22q11.2 pairs associated with IQ and 1,206 16p11.2 and 162

297 22q11.2 pairs associated with BMI (Table 1, Supplementary table 6, Supplementary table 7). The  
298 proportion of gene pairs exceeding the significance threshold was consistent with that of single  
299 genes, and in the cases where the proportions differed (such schizophrenia and IQ at 16p11.2,  
300 BMI and IQ at 22q11.2), the pairwise analysis had the lower proportion of significantly  
301 associated genes/pairs. We thus find that pairwise association signal is oligogenic, spread across  
302 many pairs rather than enrichment specific to top outlier results (Supplementary figure 1,  
303 Supplementary figure 2). Due both to the eQTL sharing between pairwise prediction models as  
304 well as to the sharing of genes across pairs, we are unable to use our approach to confidently  
305 identify specific candidate gene pairs; several pairs of potential interest are noted in the  
306 Discussion section.

307

### 308 Patterns of genes most represented in associated pairs differ by phenotype

309

310 We wanted to know whether the pairwise associations were primarily comprised of genes with  
311 independent association signal or indicated genes with uniquely combinatorial effects. The  
312 results were strikingly different across traits. In two cases – bipolar disorder at 16p11.2 and  
313 bipolar disorder at 22q11.2 – one gene stood out as a disproportionate contributor to pairs, but  
314 was also a top single gene association (*INO80E* and *PPIL2* respectively). No single gene  
315 contributed disproportionately to pairs for ASD, schizophrenia, or BMI at 16p11.2  
316 (Supplementary figure 3). At 22q11.2, however, ASD pairs disproportionately included five  
317 genes - *AC004471.10*, *CLTCL1*, and *CCDC118*, which were in the top ASD single genes, as well  
318 as *DGCR2* and *DGCR6* which we did not pick up as top single gene associations. For IQ at  
319 22q11.2, *COMT* was a gene that disproportionately appeared in pairs and was not a top single

320 gene (Supplementary figure 4). The remaining genes over-represented in pairs at both 16p11.2  
321 and 22q11.2 were primarily non-coding genes that did not have significant single-gene models,  
322 demonstrating potential regulatory effects of non-coding genes on CNV coding genes  
323 (Supplementary figure 3, Supplementary figure 4). Three of the common patterns – one single  
324 gene disproportionately represented, no genes disproportionately represented, and novel genes  
325 disproportionally represented – are illustrated in Figure 2.

326

### 327 Region-wide contributions of 16p11.2 and 22q11.2 CNVs to phenotype

328 After comparing the impacts of single genes and pairs of CNV genes on neurobehavioral traits,  
329 we wanted to test combinations greater than pairwise, but feasibility limited our combinatorial  
330 testing. Therefore, we considered a polygene region-wide model: whether the average deviation  
331 of the multigenic region contributes to a phenotype. We assigned a region-wide score to each  
332 individual and tested whether scores were significantly different between cases and controls or  
333 correlated with quantitative traits. We found that the region-wide score was positively correlated  
334 with BMI for 16p11.2 genes ( $P = 2.0 \times 10^{-11}$ ) and negatively correlated for 22q11.2 genes ( $P =$   
335  $0.0001$ ) (Figure 3). IQ was also negatively correlated with region-wide score for 16p11.2 genes  
336 ( $P = 8.7 \times 10^{-15}$ ) (Figure 3). None of the categorical traits showed a significant effect of region-  
337 wide contribution (Supplementary figure 5).

338

### 339 Discussion

340 Our study aimed to provide insight into the genetic architecture of the 16p11.2 and 22q11.2 copy  
341 number variants. We modeled the neurobehavioral trait consequences of pairs of genes expressed  
342 in the same direction, extending our previous single-gene analysis (Figure 4). Both 16p11.2 and

343 22q11.2 had pairs of genes associated with all tested phenotypes based on a permutation-based  
344 threshold, however, despite a larger number of genes tested, the count of associated genes was  
345 larger for 16p11.2 gene pairs. Still, we found that for nearly all traits tested, variance in  
346 phenotype was better explained by pairs of genes than by single genes or traditional interaction  
347 models. The only exception was bipolar disorder at 22q11.2, where single genes explain more  
348 variance. However, for schizophrenia, BMI, and IQ at 22q11.2 the pairwise model was not  
349 specific to the CNV regions but appeared to be a trait-based property extending to matched  
350 control regions. These findings suggest that the pairwise effects are different between regions.  
351 The advantage of summed pair models in control regions over single and interaction models –  
352 even when it was less pronounced than that of CNV regions - was somewhat surprising due to  
353 our hypothesis that CNV regions have the unique property of dysregulation of nearby genes in  
354 the same direction. However, perhaps regulatory landscape across many regions of the genome  
355 also biases towards expression dysregulation of physically colocalized genes in the same  
356 direction.

357

358 As we observed neither enrichment in the proportion of significant pairwise tests nor outlier top  
359 signal in the QQ plots, the pairwise contribution to explaining trait variance seems to be  
360 oligogenic across the region. However, in some cases we did observe outliers when examining  
361 the frequency of specific genes involved in top pairs. There was striking variation across traits  
362 and regions in terms of whether the top single genes were also the top contributors to pairs or  
363 novel genes were equally likely to contribute. A single gene was repeatedly contributing to top  
364 pairs for bipolar disorder at 16p11.2 (*INO80E*, 26% of top pairs) and schizophrenia at 22q11.2  
365 (*PPIL2*, 42% of top pairs). The individual association with these genes was not detected, but the



366 recurrent role of these genes in pairs suggests an important trait contribution. In contrast, for  
367 schizophrenia at 16p11.2 and ASD at 22q11.2, multiple top single genes participate  
368 disproportionately in top pairs. Intriguingly, although pairwise models show similar advantages  
369 for ASD at 16p11.2 and IQ at 22q11.2, genes across the region are more evenly represented in  
370 top pairs. Bipolar disorder at 22q11.2 (with single genes models most often explaining variance)  
371 showed association with flanking genes on either side, *TUBA8*, *TMEM191B*, and *PPIL2*; *PPIL2*  
372 appeared in most of the pairs, as well. Because we did not find overall support for a pairwise  
373 model for bipolar disorder at 22q11.2, this may simply reflect the independent association of  
374 *PPIL2*. Our finding of *PPIL2* as a bipolar disorder driving gene is supported by this gene's over-  
375 representation of rare protein truncating variants in the Bipolar Exome sequencing consortium  
376 data [42].

377  
378 Given that the pairwise signal tended to be oligogenic and that expression imputation of adjacent  
379 genes has high correlation, it is difficult to confirm the association of specific pairs of genes. For  
380 ASD at 16p11.2, the top 15 pairs include four with *FAM57B*. This gene was previously shown to  
381 have multiple within-region interactions in zebrafish [43]. Here, we find that the top pairwise  
382 contributions are with coding and non-coding genes in repetitive or flanking regions (*RP11-*  
383 *347C12.3*, *TBC1D10B*, *BOLA2B*, *NPIP12*). Studies of 16p11.2 CNV genes rarely include these  
384 flanking genes, but our data suggest that they may contribute to trait association. Notably, our  
385 expectation of expression dysregulation in the same direction would be less strong for flanking  
386 genes, so expanded testing of flanking regions may be worthwhile. The *FAM57B* and *DOC2A*  
387 pair, associated with hyperactivity, head size, and length in the zebrafish study, was in the top  
388 quarter of associated pairs for BMI and IQ. We note that McCammon *et al* specifically excluded

389 additive effects, while our study is based on genes contributing additively to pairs (which we find  
390 explains more variance than traditional interactions). For BMI at 16p11.2, the top ranked pair is  
391 *CDIPT* with *ALDOA*. It is notable that these two genes were not top-ranked individual genes for  
392 BMI, demonstrating the utility of our pairwise approach to prioritize pairs that might not be  
393 detected as individual genes. The top pair for IQ, *MVP* and *KCTD13*, on the other hand, includes  
394 one top IQ-associated gene (*MVP*) and one gene (*KCTD13*) not associated with IQ. This finding  
395 is similar to an observation in zebrafish, where the expressivity of head-size phenotypes driven  
396 by *KCTD13* overexpression was increased by additional overexpression of *MVP* [13]. For IQ at  
397 22q11.2, several top pairs contain *COMT* along with a non-coding gene. *COMT* is a gene with  
398 variants believed to affect multiple traits, including IQ [44], and whose expression is associated  
399 with IQ in the general population [45]. Our data provides a refined hypothesis that the  
400 relationship between *COMT* and IQ is dependent on additional non-coding genes at 22q11.2.  
401  
402 We also wondered whether there was a general contribution across many genes in the region. In  
403 our analyses, we found that there was a region-wide contribution to both BMI and IQ in both  
404 CNVs. The large number of 16p11.2 genes associated with BMI in both single and pairwise  
405 models was consistent with a region-wide signal. From previously established associations in  
406 CNV carriers, we would expect a negative correlation between increased expression and BMI for  
407 both 16p11.2 and 22q11.2 CNVs. However, we saw this only at 22q11.2 in the region-wide  
408 model. Previously, we found individual genes independently associated with both increases and  
409 decreases of BMI at 16p11.2 [22]. We hypothesized that there may be both BMI-increasing and  
410 BMI-decreasing genes in the 16p11.2 region due to our observation of association in both  
411 directions in single-gene models (and BMI decreases in syntenic deletion mice [12]), in which

412 case we might have been picking up more BMI-decreasing genes in our region-wide score.

413 However, one potential limitation of our cross-tissue expression prediction approach is that our  
414 results may not be driven by the biologically-relevant tissues and thus appear to be opposite in  
415 direction [46]. We also note that BMI and IQ are quantitative traits with high sample size, and  
416 so we may have had power limitations in other traits.

417  
418 Previously, we proposed that *INO80E* at 16p11.2 is a driver of schizophrenia and BMI, a finding  
419 that has been corroborated in similar analyses by others. However, we found that pairwise  
420 models explained more trait variance in both schizophrenia and BMI at 16p11.2, so it is possible  
421 that the pathophysiological contribution of *INO80E* will be better explained in combination with  
422 other genes than independently, a hypothesis that might be of interest for experimental design.  
423 Our pairwise findings also suggest that *INO80E* has an important contribution to at least two  
424 other traits. In bipolar disorder, *INO80E* is the top individual associated gene and is the most  
425 disproportionate contributor to pairs. In ASD, *INO80E* is a weakly associated top individual gene  
426 and is the most frequent (albeit not strongly disproportionate) contributor to significant pairs.  
427 This finding suggests that four traits may be influenced by the *INO80E* gene, and at least in the  
428 case of ASD, this gene works in combination with other genes. However, we have not found  
429 evidence of the involvement of *INO80E* in IQ, highlighting that the neurobehavioral phenotypes  
430 of 16p11.2 may be broader than the impact of this single gene, under the assumption that IQ in  
431 the general population is a good representation of the 16p11.2-mediated impact on intellectual  
432 ability.

433

434 There are a number of limitations in our approach to probing the architecture of 16p11.2 and  
435 22q11.2 CNVs using pairs of gene expression predictions and region-wide gene expression  
436 scores. There are numerous combinatorial models that have not been tested, and the true  
437 architecture of gene-trait pairs may lie anywhere in between what we can capture in simplified  
438 models. In fact, given the observation that the entire 16p chromosome arm is enriched for ASD  
439 risk signal and has high amount of chromosomal contact, the region itself, as we had defined it,  
440 could be insufficient [26,47].

441  
442 Another potential model that we have not tested is that only the extremes of the distribution –  
443 either in pairwise sums or region-wide scores – will impact a phenotype, and more modest  
444 increases and decreases in gene expression are buffered. For example, the BMI-16p11.2 panel in  
445 Figure 3 suggests a difference in the top and bottom decile compared to the BMI-score  
446 relationship in the intermediate deciles. Our study using all individuals has an advantage in  
447 statistical power if more typical gene expression levels are relevant to the trait, but a  
448 disadvantage given the potential noise that is introduced if only extreme expression deviation is  
449 relevant to uncommon traits such as schizophrenia, bipolar disorder, and ASD.

450  
451 A technical limitation of our study design is that available datasets are not always ideal for our  
452 approach. For BMI, IQ, and ASD, the best-powered datasets are summary statistics. We use the  
453 summary statistics for single and pairwise association testing, determining permutation-based  
454 significance cutoffs, and finding top individual genes that are represented in pairs. However, in  
455 order to measure variance explained and region-wide scoring, we use individual-level data. We  
456 have to consider heterogeneity across the cohorts as a caveat when comparing results. Still, for

457 both ASD and IQ the individual level data used is a subset of the full cohort comprising  
458 summary level statistics, minimizing the differences. Finally, our study is based on multiple  
459 tissues derived from adults, rather than more targeted analyses of the brain at early development.  
460 Similarly, when we decide which model explains more variance, we do not weight tissues  
461 differently (according to trait relevance, sample size, etc.). Despite the limitations, we may be  
462 detecting signal driven by a subset of the data; for example, ASD-donor cerebral organoids show  
463 cell-type specificity of *INO80E* to neuroepithelial cells during development, yet we detect a  
464 pairwise contribution in cross-tissue analysis [48].

465  
466 The 16p11.2 and 22q11.2 regions are highly penetrant for neurobehavioral traits, but require a  
467 better understanding of genetic architecture to indicate key biological pathways. By extending  
468 transcription imputation to study a simple summed model of pairwise gene expression, we  
469 uncover a consistent pattern of higher variance explained by gene pairs than either single genes  
470 or traditional interaction models and several traits showing region-wide association signal  
471 (Figure 4). Most of these patterns appear specific to CNV regions and did not appear to  
472 represent the genetic architecture in matched control regions. ASD, for which single gene  
473 approaches had small to no effect, shows pairwise association signal above that of controls at  
474 both 16p11.2 and 22q11.2. Having failed to dissect 22q11.2 with single-gene approaches, here  
475 we found least two 22q11.2 traits – BMI and IQ – that can be better modeled region-wide. Our  
476 study suggests that pathobiological insights might result from studying combinations of the  
477 genes in and near these CNVs, albeit with potentially differing genetic architecture across traits  
478 and regions.

479

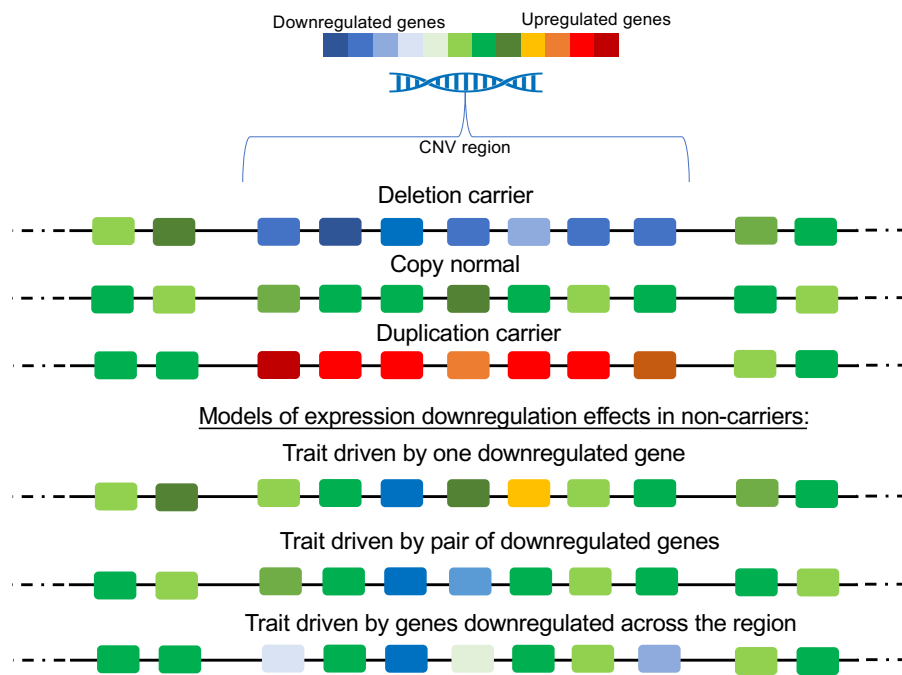
480

481

482

483 Figures

484



485

486 Figure 1: An overview of models of CNV pathogenicity due to gene expression.

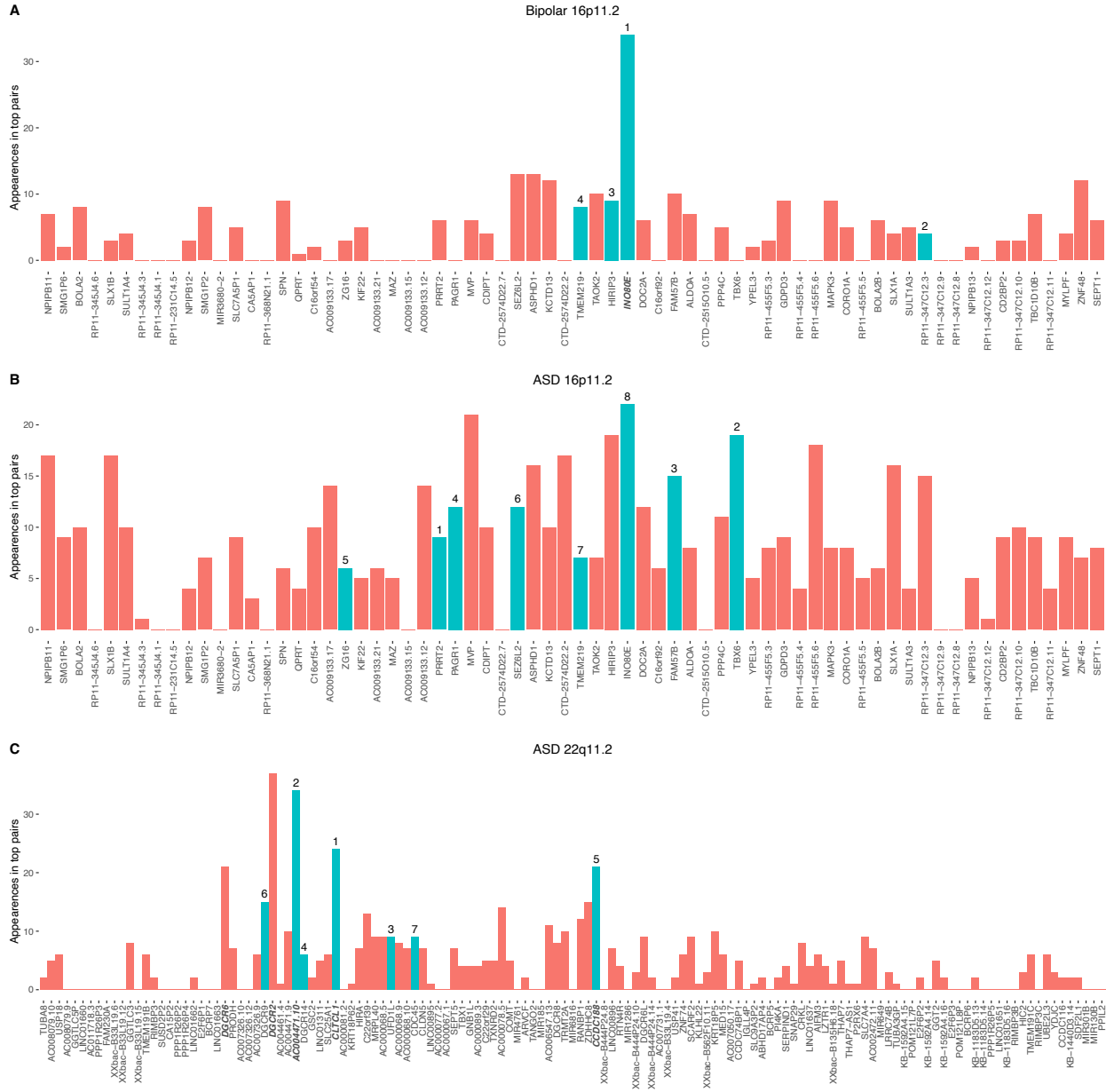
487 Rectangles represent individual genes in a chromosomal location. Warmer colors represent

488 increased mRNA expression. Cooler colors represent decreased mRNA expression. Greens

489 represent population average mRNA expression.

490 Top: Within a CNV region, deletion carriers have reduced expression across the majority of  
491 genes, duplication carriers have increased expression across the majority of genes, and copy  
492 normal individuals have “average” levels of expression across the majority of genes. These  
493 increases and decreases are specific to the CNV region experiencing increased or decreased  
494 DNA copies (potential positional effects on flanking genes not shown).

495 Bottom: Three models of how gene expression downregulation in a CNV region may influence a  
496 CNV-associated trait in non-carriers. In the first model, decreased expression of a single gene is  
497 sufficient. In the second model, a trait is impacted when two specific genes both have reduced  
498 expression. In the third model, the trait becomes more likely due to reduction of expression in  
499 many genes across the region. These three models are utilized in our study.



500

501

502 Figure 2: Three representative examples of patterns of top single genes contributing to

503 significant pairs.

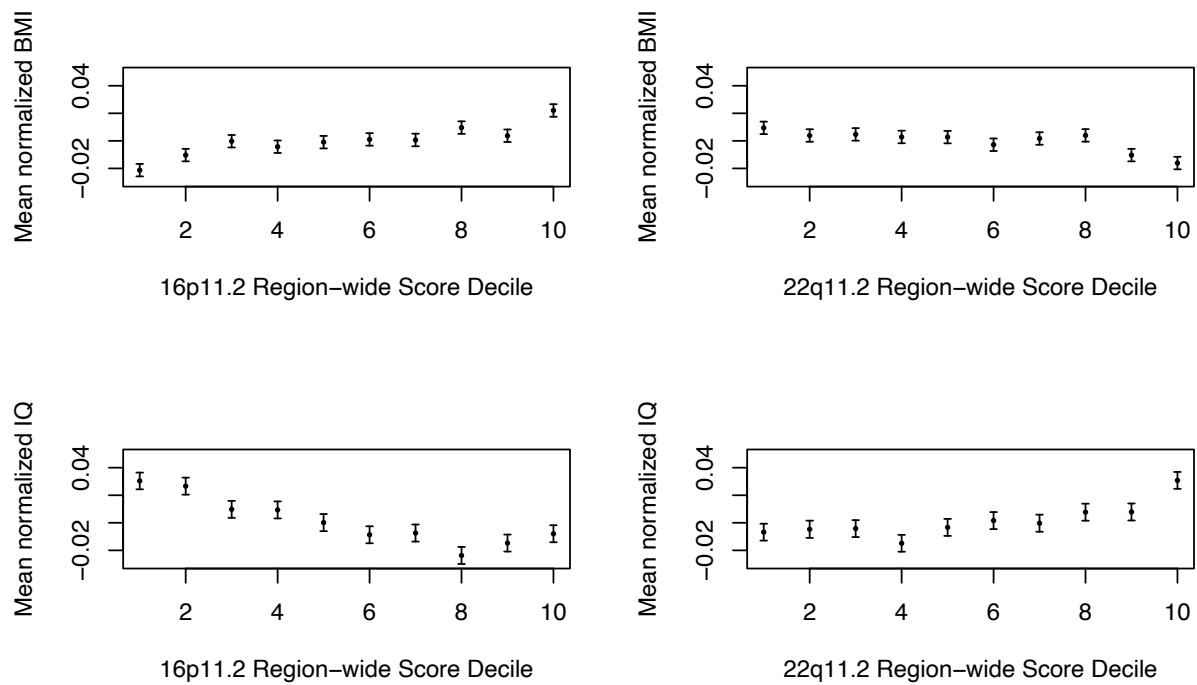
504 Y-axis: counts of the number of times each gene contributes to a significant pair (permutation  $P$ -

505 value < median of 5<sup>th</sup> percentiles of control region  $p$ -values). Bars in teal represent genes

506 significant (permutation  $P$ -value < median of 5<sup>th</sup> percentiles of control region  $p$ -values) in a



507 single gene model for the same trait, with rank indicated above the bar. Bars in salmon represent  
508 genes not significant in a single gene model. X-axis: genes in chromosomal order.  
509 Disproportionately represented genes (mean + 2.5 standard deviations) are bolded.  
510 (A) For bipolar disorder at 16p11.2, one gene, *INO80E* is disproportionately represented in pairs;  
511 this gene is also a top single gene.  
512 (B) For ASD at 16p11.2, genes significant in a single gene model are not disproportionately  
513 represented in significant pairs, with no disproportionate outliers evident.  
514 (C) For ASD at 22q11.2, there is a mixture of genes that disproportionately appear in pairs: both  
515 genes that were picked up by single gene analyses (like *CLTCLI*) as well as genes that were not  
516 (like *DGCR2*).  
517



519

520

521 Figure 3: IQ and BMI values are associated with region-wide score.

522 Region-wide scores across individuals were binned into deciles and the mean (dot) and standard

523 error (bars) of BMI and IQ values for each decile are plotted.

524

Region	Trait	Pairwise model	Novel paired genes	Region-wide model
16p11.2	ASD	Yes	No	No
	Bipolar Disorder	Yes	No	No
	Schizophrenia	Yes	No	No
	BMI	Yes	No	Yes
	IQ	Yes	Yes	Yes
22q11.2	ASD	Yes	Yes	No
	Bipolar Disorder	No	No	No
	Schizophrenia	Yes	Yes	No
	BMI	Yes	Yes	Yes
	IQ	Yes	Yes	Yes

Yes

No

525

526 Figure 4: Insights gained into CNV-trait pairs

527 For each CNV-trait pair, we specify whether pairwise models performed better than single gene

528 models (left column), whether genes represented disproportionately in significant pairs primarily

529 represented genes significant in a single gene model (middle column), and whether region-wide

530 association with a trait was significant. Yes: salmon, No: teal. Dotted fields in the first column

531 represent cases where the pairwise model advantage did not exceed that of control regions.

532 Dotted fields in the second column represent cases where they were no over-represented single

533 genes in pairs.

534

535

536

537

538

539 Tables

540

541 Table 1: Proportion of significantly associated (permutation  $P <$  median of 5<sup>th</sup> percentiles of

542 control region p-values) single genes (single) and pairwise gene sums (pairs) for each trait and

543 CNV.

	16p11.2		22q11.2	
Trait	<b>N single (%)</b>	<b>N pairs (%)</b>	<b>N single (%)</b>	<b>N pairs (%)</b>
ASD	8/42 (19%)	273/1542 (18%)	6/65 (9%)	282/3654 (8%)
Bipolar	5/37 (14%)	142/1536 (9%)	3/59 (5%)	137/3669 (4%)
Schizophrenia	21/37 (57%)	702/1543 (45%)	1/59 (2%)	129/4267 (3%)
BMI	31/38 (82%)	1212/1554 (78%)	5/52 (10%)	176/3229 (5%)
IQ	5/38 (13%)	74/1545 (5%)	2/65 (3%)	33/4052 (1%)

544

545

546

547

548

549

550

551

552 Table 2: Counts of the model estimated to explain most trait variance for each tissue-cohort pair.  
553 Best model is bolded in each case. *P*-value represents a chi-square test comparing the proportion  
554 of pairwise to non-pairwise counts between CNV regions and controls.  
555

		CNV Region	All Control Regions	
Region	Trait	single/interaction/pairwise (% pairwise)	single/interaction/pairwise (% pairwise)	region-specific P-value
16p11.2	ASD	205/169/ <b>243</b> (39%)	5891/ <b>7588</b> /6387 (32%)	0.00012
	Bipolar	359/390/ <b>721</b> (49%)	14806/19593/ <b>19631</b> (36%)	< 2.2x10 <sup>-16</sup>
	Schizophrenia	754/730/ <b>1554</b> (51%)	26589/37723/ <b>47784</b> (43%)	< 2.2x10 <sup>-16</sup>
	BMI	0/0/ <b>49</b> (100%)	48/159/ <b>1744</b> (89%)	0.016
	IQ	0/0/ <b>49</b> (100%)	98/232/ <b>1565</b> (83%)	0.0013
22q11.2	ASD	174/196/ <b>267</b> (42%)	4909/ <b>7016</b> /5313 (31%)	1.3x10 <sup>-9</sup>
	Bipolar	<b>536</b> /435/499 (34%)	11642/ <b>15167</b> /14381 (35%)	0.44
	Schizophrenia	871/816/ <b>1155</b> (41%)	19632/28147/ <b>35053</b> (42%)	0.07
	BMI	0/0/ <b>49</b> (100%)	17/68/ <b>1258</b> (94%)	0.069
	IQ	7/17/ <b>25</b> (51%)	20/92/ <b>1041</b> (90%)	<2.2x10 <sup>-16</sup>

556  
557 Legends of Supplements:  
558  
559 Supplementary figure 1: Pairwise signal at 16p11.2 is polygenic.  
560 Q-Q plots comparing PrediXcan association signal from single 16p11.2 genes (blue), pairs of  
561 16p11.2 genes (orange), single genes in control subsets (gray), and pairs of genes in control  
562 subsets (black).  
563  
564 Supplementary figure 2: Pairwise signal at 22q11.2 is polygenic.

565 Q-Q plots comparing PrediXcan association signal from single 22q11.2 genes (blue), pairs of  
566 22q11.2 genes (orange), single genes in control subsets (gray), and pairs of genes in control  
567 subsets (black).

568

569 Supplementary figure 3: Patterns of top single genes contributing to significant pairs at 16p11.2.

570 Y-axis: counts of the number of times each gene contributes to a significant pair (permutation  $P$ -  
571 value  $<$  median of 5<sup>th</sup> percentiles of control region  $p$ -values). Bars in teal represent genes  
572 significant (permutation  $P$ -value  $<$  median of 5<sup>th</sup> percentiles of control region  $p$ -values) in a  
573 single gene model for the same trait, with rank indicated above the bar. Bars in salmon represent  
574 genes not significant in a single gene model. X-axis: genes in chromosomal order.

575 Disproportionately represented genes (mean + 2.5 standard deviations) are bolded.

576

577 Supplementary figure 4: Patterns of top single genes contributing to significant pairs at 22q11.2.

578 Y-axis: counts of the number of times each gene contributes to a significant pair (permutation  $P$ -  
579 value  $<$  median of 5<sup>th</sup> percentiles of control region  $p$ -values). Bars in teal represent genes  
580 significant (permutation  $P$ -value  $<$  median of 5<sup>th</sup> percentiles of control region  $p$ -values) in a  
581 single gene model for the same trait, with rank indicated above the bar. Bars in salmon represent  
582 genes not significant in a single gene model. X-axis: genes in chromosomal order.

583 Disproportionately represented genes (mean + 2.5 standard deviations) are bolded.

584

585 Supplementary figure 5: Region-wide score association with ASD, bipolar disorder, and

586 schizophrenia.

587 Region-wide scores across individuals were binned into deciles and the mean (dot) and standard  
588 error (bars) of case-control ratios for each decile are plotted. Top: 16p11.2. Bottom: 22q11.2.

589 Left to right: ASD, Schizophrenia, Bipolar Disorder.

590

591 Supplementary note 1: Members of the Psychiatric Genomics Consortium contributing to this  
592 work.

593

594 Supplementary table 1: Pairwise and single predictive model qualities for 16p11.2 genes.

595 Four different model sets were created for SNP overlap (ASD and bipolar used the same PGC  
596 panel). For each pair of genes, the median of prediction qualities ( $R^2$ ) among tissues along with

597 the number of tissues for which predictive models are available (which can be zero) are noted.

598 These are compared to single-gene predictive model qualities and coding/noncoding genes are  
599 annotated.

600

601 Supplementary table 2: Pairwise and single predictive model qualities for 22q11.2 genes.

602 Four different model sets were created for SNP overlap (ASD and bipolar used the same PGC  
603 panel). For each pair of genes, the median of prediction qualities ( $R^2$ ) among tissues along with

604 the number of tissues for which predictive models are available (which can be zero) are noted.

605 These are compared to single-gene predictive model qualities and coding/noncoding genes are  
606 annotated.

607

608 Supplementary table 3: Cohorts used for analyses.

609 The specific cohorts from the Psychiatric Genomics Consortium that were used for this analysis  
610 are listed.

611

612 Supplementary table 4: Genomic regions used as controls for analyses.

613 Each region is annotated with genomic context (for example if it is dense with olfactory genes).

614 Regions that were matched to each CNV but were not used for analysis due to MHC or study

615 CNV overlap are labeled as such.

616

617 Supplementary table 5: Single gene associations with five neurobehavioral traits.

618 Genes are listed in order of MultiXcan/S-MultiXcan p-values and genes significant based on a

619 permutation-based threshold are highlighted.

620

621 Supplementary table 6: Pairwise associations with five neurobehavioral traits at 16p11.2.

622 Genes are listed in order of MultiXcan/S-MultiXcan p-values and genes significant based on a

623 permutation-based threshold are highlighted.

624

625 Supplementary table 7: Pairwise associations with five neurobehavioral traits at 22q11.2.

626 Genes are listed in order of MultiXcan/S-MultiXcan p-values and genes significant based on a

627 permutation-based threshold are highlighted.

628

629

630 Availability of data and materials



631 Individual-level genotypes for Psychiatric Genomics Consortium cohorts can be obtained by  
632 applying at [www.pgc.unc.edu/](http://www.pgc.unc.edu/) Summary level data from the PGC is at [https://pgc.unc.edu/for-](https://pgc.unc.edu/for-researchers/download-results/)  
633 [researchers/download-results/](https://pgc.unc.edu/for-researchers/download-results/) . Summary-level genetic datasets for BMI and IQ are available to  
634 freely download from GIANT BMI  
635 ([https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium)) and CNCR  
636 IQ ([https://ctg.cncr.nl/software/summary\\_statistics](https://ctg.cncr.nl/software/summary_statistics)). Individual-level UK Biobank data can be  
637 obtained by application at <https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>  
638 PrediXcan single-gene genome-wide models are available to download at [predictdb.org](http://predictdb.org). GTEx  
639 genotypes and phenotypes are requestable on dbGAP (phs000424.v8.p2). Summary statistics  
640 from association studies performed in this article are located in the supplement.

#### 641 Competing interests

642 The authors declare no competing interests.

#### 643 Funding

644 This work was supported by National Institute of Mental Health R01 MH107467 to LAW. The  
645 funding body had no role in study design, data collection and analysis, decision to publish, or  
646 preparation of the manuscript.

#### 647 Authors' contributions

648 MV performed all data analysis mentioned here. MV and LAW contributed to study design and  
649 manuscript writing/editing.

#### 650 Acknowledgements

651 We acknowledge Nancy J. Cox for her contribution to study conception and advice about  
652 methodology. Noah Zaitlen provided helpful ideas for testing the utility of pairwise models.

653

654 *References*

- 655 1. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, et al. Mirror  
656 extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature*.  
657 2011;478:97–102.
- 658 2. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al.  
659 Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*. 2009;41:1223–7.
- 660 3. Shinawi M, Liu P, Kang S-HL, Shen J, Belmont JW, Scott DA, et al. Recurrent reciprocal  
661 16p11.2 rearrangements associated with global developmental delay, behavioural problems,  
662 dysmorphism, epilepsy, and abnormal head size. *J Med Genet*. 2010;47:332–41.
- 663 4. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, et al. Recurrent  
664 16p11.2 microdeletions in autism. *Hum Mol Genet*. 2007;17:628–38.
- 665 5. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between  
666 Microdeletion and Microduplication at 16p11.2 and Autism. *N Engl J Med*. 2008;358:667–75.
- 667 6. Bassett AS, Chow EW. 22q11 deletion syndrome: a genetic subtype of schizophrenia. *Biol*  
668 *Psychiatry*. 1999;46:882–91.
- 669 7. Campbell IM, Sheppard SE, Crowley TB, McGinn DE, Bailey A, McGinn MJ, et al. What is  
670 new with 22q? An update from the 22q and You Center at the Children’s Hospital of  
671 Philadelphia. *Am J Med Genet Part A*. 2018;176:2058–69.
- 672 8. Wentzel C, Fernström M, Öhrner Y, Annerén G, Thuresson A-C. Clinical variability of the  
673 22q11.2 duplication syndrome. *Eur J Med Genet*. 2008;51:501–10.
- 674 9. Schneider M, Debbané M, Bassett AS, Chow EWC, Fung WLA, van den Bree MBM, et al.  
675 Psychiatric Disorders From Childhood to Adulthood in 22q11.2 Deletion Syndrome: Results  
676 From the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *Am J*

677 Psychiatry. 2014;171:627–39.

678 10. Voll SL, Boot E, Butcher NJ, Cooper S, Heung T, Chow EWC, et al. Obesity in adults with  
679 22q11.2 deletion syndrome. *Genet Med*. 2017;19:204–8.

680 11. Carlson C, Papolos D, Pandita RK, Faedda GL, Veit S, Goldberg R, et al. Molecular analysis  
681 of velo-cardio-facial syndrome patients with psychiatric disorders. *Am J Hum Genet*.  
682 1997;60:851–9.

683 12. Pucilowska J, Vithayathil J, Tavares EJ, Kelly C, Colleen Karlo J, Landreth GE. The 16p11.2  
684 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain  
685 cytoarchitecture linked to the ERK MAPK pathway. *J Neurosci*. 2015;35:3190–200.

686 13. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, et al. KCTD13 is a  
687 major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant.  
688 *Nature*. 2012;485:363–7.

689 14. Blaker-Lee A, Gupta S, McCammon JM, De Rienzo G, Sive H. Zebrafish homologs of genes  
690 within 16p11.2, a genomic region associated with brain disorders, are active during brain  
691 development, and include two deletion dosage sensor genes. *Dis Model Mech*. 2012;5.

692 15. Iyer J, Singh MD, Jensen M, Patel P, Pizzo L, Huber E, et al. Pervasive genetic interactions  
693 modulate neurodevelopmental defects of the autism-Associated 16p11.2 deletion in *Drosophila*  
694 *melanogaster*. *Nat Commun*. 2018;9:1–19.

695 16. Jensen M, Girirajan S. An interaction-based model for neuropsychiatric features of copy-  
696 number variants. *PLoS Genet*. 2019;15.

697 17. Ward TR, Zhang X, Leung LC, Zhou B, Muench K, Roth JG, et al. Genome-wide molecular  
698 effects of the neuropsychiatric 16p11 CNVs in an iPSC-to-iN neuronal model. *bioRxiv*.  
699 2020;2020.02.09.940965.

- 700 18. Blumenthal I, Ragavendran A, Erdin S, Klei L, Sugathan A, Guide JR, et al. Transcriptional  
701 Consequences of 16p11.2 Deletion and Duplication in Mouse Cortex and Multiplex Autism  
702 Families. *Am J Hum Genet.* 2014;94:870–83.
- 703 19. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, et al. Genome-wide  
704 Transcriptome Profiling Reveals the Functional Impact of Rare De Novo and Recurrent CNVs in  
705 Autism Spectrum Disorders. *Am J Hum Genet.* 2012;91:38–55.
- 706 20. Zhang X, Zhang Y, Zhu X, Purmann C, Haney MS, Ward T, et al. Local and global  
707 chromatin interactions are altered by large genomic deletions associated with human brain  
708 development. *Nat Commun.* 2018;9.
- 709 21. Mostafavi H, Spence JP, Naqvi S, Pritchard JK. Limited overlap of eQTLs and GWAS hits  
710 due to systematic differences in discovery. *bioRxiv.* 2022;
- 711 22. Vysotskiy M, Zhong X, Miller-Fleming TW, Zhou D, Cox NJ, Weiss LA. Integration of  
712 genetic, transcriptomic, and clinical data provides insight into 16p11.2 and 22q11.2 CNV genes.  
713 *Genome Med* 2021 131. 2021;13:1–26.
- 714 23. McCammon JM, Blaker-Lee A, Chen X, Sive H. The 16p11.2 homologs *fam57ba* and *doc2a*  
715 generate certain brain and body phenotypes. *Hum Mol Genet.* 2017;26:3699–712.
- 716 24. Forstner AJ, Degenhardt F, Schrott G, Nöthen MM. MicroRNAs as the cause of  
717 schizophrenia in 22q11.2 deletion carriers, and possible implications for idiopathic disease: a  
718 mini-review. *Front Mol Neurosci.* 2013;6:47.
- 719 25. Barbeira AN, Melia OJ, Liang Y, Bonazzola R, Wang G, Wheeler HE, et al. Fine-mapping  
720 and QTL tissue-sharing information improves the reliability of causal gene identification. *Genet*  
721 *Epidemiol.* 2020;44:854–67.
- 722 26. Weiner DJ, Ling E, Erdin S, Tai DJC, Yadav R, Grove J, et al. Statistical and functional

723 convergence of common and rare variant risk for autism spectrum disorders at chromosome 16p.  
724 medRxiv. 2022;13:2022.03.23.22272826.

725 27. Tai DJC, Razaz P, Erdin S, Gao D, Wang J, Nuttle X, et al. Tissue and cell-type specific  
726 molecular and functional signatures of 16p11.2 reciprocal genomic disorder across mouse brain  
727 and human neuronal models. bioRxiv. 2022;2022.05.12.491670.

728 28. Consortium TSWG of the PG, Ripke S, Walters JT, O'Donovan MC. Mapping genomic loci  
729 prioritises genes and implicates synaptic biology in schizophrenia. medRxiv.  
730 2020;2020.09.12.20192922.

731 29. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide  
732 association study identifies 30 loci associated with bipolar disorder. Nat Genet. 2019;51:793–  
733 803.

734 30. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of  
735 common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.

736 31. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body  
737 mass index yield new insights for obesity biology. Nature. 2015;518:197–206.

738 32. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, et al. Genome-wide  
739 association meta-analysis in 269,867 individuals identifies new genetic and functional links to  
740 intelligence. Nat Genet. 2018;50:912–9.

741 33. UK Biobank — Neale lab [Internet]. [cited 2020 Mar 28]. Available from:  
742 <http://www.nealelab.is/uk-biobank>

743 34. Barbeira AN, Bonazzola R, Gamazon ER, Liang Y, Park YS, Kim-Hellmuth S, et al.  
744 Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. Genome Biol.  
745 2021;22:1–24.

746 35. Gamazon ER, Wheeler HE, Shah KP, Mozaffari S V., Aquino-Michaels K, Carroll RJ, et al.  
747 A gene-based association method for mapping traits using reference transcriptome data. *Nat*  
748 *Genet.* 2015;47:1091–8.

749 36. Mokhtari R, Lachman HM. The Major Histocompatibility Complex (MHC) in  
750 Schizophrenia: A Review. *J Clin Cell Immunol.* 2016;7.

751 37. Barbeira AN, Pividori MD, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted  
752 transcriptome from multiple tissues improves association detection. Plagnol V, editor. *PLOS*  
753 *Genet.* 2019;15:e1007889.

754 38. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide  
755 association scans. *Bioinformatics.* 2010;26:2190–1.

756 39. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring  
757 the phenotypic consequences of tissue specific gene expression variation inferred from GWAS  
758 summary statistics. *Nat Commun.* 2018;9:1825.

759 40. Liang Y, Pividori M, Manichaikul A, Palmer AA, Cox NJ, Wheeler HE, et al. Polygenic  
760 transcriptome risk scores (PTRS) can improve portability of polygenic risk scores across  
761 ancestries. *Genome Biol.* 2022;23:1–18.

762 41. Clements CC, Wenger TL, Zoltowski AR, Bertollo JR, Miller JS, de Marchena AB, et al.  
763 Critical region within 22q11.2 linked to higher rate of autism spectrum disorder. *Mol Autism.*  
764 2017;8:58.

765 42. Palmer DS, Howrigan DP, Chapman SB, Adolfsson R, Bass N, Blackwood D, et al. Exome  
766 sequencing in bipolar disorder reveals shared risk gene AKAP11 with schizophrenia. *medRxiv.*  
767 2021;2021.03.09.21252930.

768 43. McCammon JM, Blaker-Lee A, Chen X, Sive H. The 16p11.2 homologs *fam57ba* and *doc2a*

769 generate certain brain and body phenotypes. *Hum Mol Genet.* 2017;26:3699–712.

770 44. Carmel M, Zarchi O, Michaelovsky E, Frisch A, Patya M, Green T, et al. Association of  
771 COMT and PRODH gene variants with intelligence quotient (IQ) and executive functions in  
772 22q11.2DS subjects. *J Psychiatr Res.* 2014;56:28–35.

773 45. Ni P, Liu M, Wang D, Tian Y, Zhao L, Wei J, et al. Association Analysis Between Catechol-  
774 O-Methyltransferase Expression and Cognitive Function in Patients with Schizophrenia, Bipolar  
775 Disorder, or Major Depression. *Neuropsychiatr Dis Treat.* 2021;17:567–74.

776 46. Mizuno A, Okada Y. Biological characterization of expression quantitative trait loci (eQTLs)  
777 showing tissue-specific opposite directional effects. *Eur J Hum Genet.* 2019;

778 47. Loviglio MN, Leleu M, Männik K, Passeggeri M, Giannuzzi G, van der Werf I, et al.  
779 Chromosomal contacts connect loci associated with autism, BMI and head circumference  
780 phenotypes. *Mol Psychiatry.* 2016;

781 48. Lim ET, Chan Y, Burns MJ, Guo X, Erdin S, Tai DJC, et al. Identifying cell type specific  
782 driver genes in autism-associated copy number loci from cerebral organoids. *bioRxiv.*  
783 2020;2020.11.15.375386.

784