

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/157200/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Lin, Jhih-Rong, Zhao, Yingjie, Jabalameli, M Reza, Nguyen, Nha, Mitra, Joydeep, International 22q11.2DS Brain and Behavior Consortium*, Swillen, Ann, Vorstman, Jacob, Chow, Eva, van den Bree, Marianne, Emanuel, Beverly, Vermeesch, Joris, Owen, Michael, Williams, Nigel, Bassett, Anne, McDonald-McGinn, Donna, Gur, Raquel, Bearden, Carrie, Morrow, Bernice, Lachman, Herbert and Zhang, Zhengdong 2023. Rare coding variants as risk modifiers of the 22q11.2 deletion implicate postnatal cortical development in syndromic schizophrenia. *Molecular Psychiatry* 28, pp. 2071-2080. 10.1038/s41380-023-02009-y

Publishers page: <https://doi.org/10.1038/s41380-023-02009-y>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



1 **Rare coding variants as risk modifiers of the 22q11.2 deletion**
2 **implicate postnatal cortical development in syndromic**
3 **schizophrenia**

4
5
6
7 Jih-Rong Lin¹, Yingjie Zhao¹, M. Reza Jabalameli¹, Nha Nguyen¹, Joydeep Mitra¹,
8 International 22q11.2DS Brain and Behavior Consortium*, Ann Swillen², Jacob A. S.
9 Vorstman³, Eva W. C. Chow³, Marianne van den Bree⁴, Beverly S. Emanuel⁵, Joris R.
10 Vermeesch², Michael J. Owen⁴, Nigel M. Williams⁴, Anne S. Bassett³, Donna M.
11 McDonald-McGinn⁵, Raquel E. Gur⁶, Carrie E. Bearden⁷, Bernice E. Morrow¹, Herbert M.
12 Lachman¹, Zhengdong D. Zhang^{1,§}

- 13
14
15
16 1. Department of Genetics, Albert Einstein College of Medicine, NY, USA
17 2. Department of Human Genetics, KU Leuven, Leuven, Belgium
18 3. Department of Psychiatry, University of Toronto, Toronto, ON, Canada
19 4. MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological
20 Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff,
21 UK
22 5. Division of Human Genetics and 22q and You Center, the Children’s Hospital of
23 Philadelphia, Philadelphia, PA, USA
24 6. Department of Psychiatry and Lifespan Brain Institute, Penn Medicine—CHOP,
25 University of Pennsylvania, Philadelphia, PA, USA
26 7. Departments of Psychiatry and Biobehavioral Sciences and Psychology, Semel
27 Institute for Neuroscience and Human Behavior, University of California, Los Angeles,
28 CA, USA

29
30 * A full list of members and their affiliations appears in the supplement.

31 § Corresponding author (E-mail: zhengdong.zhang@einsteinmed.edu)

32
33
34
35 Keywords: 22q11.2 deletion, schizophrenia, rare variants, disease risk modification

39 **ABSTRACT**

40

41 22q11.2 deletion is one of the strongest known genetic risk factors for schizophrenia.
42 Recent whole-genome sequencing of schizophrenia cases and controls with this deletion
43 provided an unprecedented opportunity to identify risk modifying genetic variants and
44 investigate their contribution to the pathogenesis of schizophrenia in 22q11.2 deletion
45 syndrome. Here, we apply a novel analytic framework that integrates gene network and
46 phenotype data to investigate the aggregate effects of rare coding variants and identified
47 modifier genes in this etiologically homogenous cohort (223 schizophrenia cases and 233
48 controls of European descent). Our analyses revealed significant additive genetic
49 components of rare nonsynonymous variants in 110 modifier genes (adjusted $P = 9.4E-04$)
50 that overall accounted for 4.6% of the variance in schizophrenia status in this cohort, of
51 which 4.0% was independent of the common polygenic risk for schizophrenia. The
52 modifier genes affected by rare coding variants were enriched with genes involved in
53 synaptic function and developmental disorders. Spatiotemporal transcriptomic analyses
54 identified an enrichment of coexpression between modifier and 22q11.2 genes in cortical
55 brain regions from late infancy to young adulthood. Corresponding gene coexpression
56 modules are enriched with brain-specific protein-protein interactions of *SLC25A1*, *COMT*,
57 and *PI4KA* in the 22q11.2 deletion region. Overall, our study highlights the contribution
58 of rare coding variants to the SCZ risk. They not only complement common variants in
59 disease genetics but also pinpoint brain regions and developmental stages critical to the
60 etiology of syndromic schizophrenia.

61

62

63 **INTRODUCTION**

64

65 22q11.2 deletion syndrome (22q11.2DS) is a severe developmental disorder, caused
66 predominantly by a 3-Mb genomic deletion (in ~85% of cases)^{1,2}. It is one of the most
67 common chromosomal abnormalities (~1 in 4000 live births, ~1 in 1000 fetuses)^{3,4}.
68 Neuropsychiatric dysfunction is a prominent feature among its heterogeneous phenotypic
69 presentations: About one in four individuals with 22q11.2DS develops schizophrenia
70 (SCZ), usually in late adolescence or early adulthood⁵⁻⁷. 22q11.2DS has been widely used
71 as a model to study SCZ due to its highly increased disease risk and clinical features
72 compatible with the more common, idiopathic form⁷⁻¹⁰. Thus, identifying genetic risk
73 factors for SCZ in addition to 22q11.2 deletion in those with 22q11.2DS has high clinical
74 and scientific value^{9,11}.

75

76 In addition to 22q11.2 deletion and common variants associated with SCZ, increasing
77 evidence shows that many rare variants may act as the 'second hits' for SCZ in 22q11.2DS⁷
78 and thus contribute to its variable expressivity^{12,13}. Identifying rare coding variants separate
79 from the common polygenic risk can complement risk prediction of 22q11.2DS-associated
80 SCZ^{9,11} and pinpoint genes that play a role in its pathophysiology. Analyzing such modifier
81 genes may reveal important biological processes in the development of syndromic SCZ
82 and their connection to disease risk genes in the 22q11.2 deletion regions.

83

84 To uncover genetic variation that contributes to the high prevalence of SCZ in 22q11.2DS,
85 the International 22q11.2DS Brain and Behavior Consortium (IBBC) assembled and
86 whole-genome sequenced (WGS) a cohort of 519 22q11.2DS patients with and without
87 SCZ, providing an indispensable resource to identify modifier rare coding variants that
88 change the disease risk of the 22q11.2 deletion in syndromic SCZ⁹. Although the cohort is
89 by far the largest to date, the conventional rare variant association tests failed to identify
90 significant associations with SCZ for any gene or pathway after multiple test correction,
91 likely due to insufficient statistical power from the moderate sample size. Therefore, it is
92 imperative to use a different approach.

93
94 We developed an integrated method – the integrated gene signal processing (IGSP) – that
95 can substantially increase the power to identify risk genes implicated by rare coding
96 variants in case-control sequencing studies using gene networks and mouse knockout
97 phenotypes¹⁴. Importantly, IGSP follows a 'discovery-driven' strategy to score risk genes
98 without relying on prior disease-specific knowledge to avoid scoring bias (**Supplementary**
99 **note**). In this study, with a new analytic framework that incorporates IGSP, we investigated
100 modifier rare coding variants associated with SCZ in 22q11.2DS with three interconnected
101 aims to (1) identify the risk component of rare coding variants in each 22q11.2DS patient
102 and its contribution to SCZ, (2) uncover modifier genes and related biological processes
103 that play important roles in SCZ risk modification in 22q11.2DS, and (3) elucidate the
104 functional connection between modifier genes and 22q11.2 deletion in brain development.

105 106 107 **RESULTS**

108 109 **Rare coding variants contribute to the genetic risk of SCZ in 22q11.2DS**

110
111 Our study cohort consisted of 223 SCZ cases and 233 controls of European ancestry in the
112 sequenced cohort of 22q11.2DS patients (**Supplementary Table S2**). We identified
113 173,752 rare coding variants (170,919 SNPs and 2,833 indels) with alternative allele
114 frequency (AAF) < 0.01 across 18,828 coding genes in the study cohort. This includes
115 61,900 synonymous, 101,944 missense, and 4,220 loss-of-function (stop gain and frame
116 shift) rare variants. There was no significant difference in the exome-wide minor allele
117 counts of either rare coding variants or rare nonsynonymous variants between SCZ cases
118 and controls ($P = 0.857$ and $P = 0.2$, respectively) (**Supplementary Fig. S1**).

119
120 In this study, we used an integrated approach with IGSP¹⁴ to address the issue of
121 insufficient statistical power of the previous study⁹, which found no significant rare variant
122 association with SCZ in the same 22q11.2DS patient cohort. We developed IGSP to
123 improve the power for prioritizing risk genes implicated by rare variants by integrating
124 gene-level rare variant association signals with gene functional network and gene knock-
125 out phenotypes. IGSP scores genes based on both rare variant association signals of genes
126 and their joint functional and phenotypic relationship (**Supplementary note**). According
127 to simulations, the IGSP score outperforms the rare variant gene association signal by 2 to
128 3 times¹⁴ in prioritizing disease risk genes.

130 Using IGSP to leverage its improved power of prioritizing risk genes, we first defined a
131 modifier risk score (MRS) to quantify aggregate effects of rare coding variants in risk genes
132 at the individual level (see **Methods**). We then evaluated how much SCZ risk that MRS
133 could account for in our study cohort with a new bootstrapping pipeline. Briefly, we
134 randomly divided our study cohort into 500 discovery and corresponding target subcohorts.
135 For each discovery subcohort, we ran IGSP to score the risk of genes implicated by rare
136 variants. Next, using top-scoring genes, we calculated the MRS of each subject in the
137 corresponding target subcohort. Finally, we averaged the MRS ranks across all target
138 subcohorts for each subject in the study cohort (**Fig. 1**). Our simulation showed that the
139 average MRS rank can effectively quantify individuals' relative SCZ risk from modifier
140 rare coding variants. It approaches the optimal MRS (after standardization) when the
141 prioritization power of risk genes increases (**Supplementary note** and **Fig. S2**).

142
143 The results of our bootstrapping analysis (**Fig. 1**) showed that SCZ cases had a significantly
144 higher MRS than controls (**Fig. 2a**): rare nonsynonymous variants in 110 putative modifier
145 genes prioritized by IGSP with a full integration of both gene network and phenotypes
146 explained 4.6% of the variance in SCZ status in the study cohort (i.e., Nagelkerke's $R^2 =$
147 4.6%, 99% CI: 3.7%-5.6%, adjusted $P = 9.4E-04$). On average, subjects with a higher MRS
148 ($> 50\%$) in a target cohort have an odds ratio (OR) of ~ 1.3 to develop SCZ (**Fig. 2b**). We
149 identified a smaller aggregate genetic signal from rare coding variants in 40 putative
150 modifier genes prioritized by IGSP with only a network integration (Nagelkerke's $R^2 =$
151 3.2%, 99% CI: 2.4%-4.2%, adjusted $P = 9.6E-03$) (**Supplementary Fig. S3**). In contrast,
152 we did not observe a clear aggregate effect of rare coding variants (Nagelkerke's $R^2 = 1.0\%$,
153 99% CI: 0.5%-1.5%, adjusted $P = 0.64$) in the top 90 genes prioritized by genetic
154 association signals alone (**Supplementary Fig. S4**). Overall, these results suggest that the
155 integration of the gene network and mouse knockout phenotypes with genetic association
156 signals significantly improved modifier gene prioritization.

157
158 To better understand the nature of rare coding variants as SCZ risk modifiers, we
159 aggregated them in different ways (**Fig. 2c**). The results strongly suggested that modifier
160 rare coding variants in the study cohort can either increase or decrease SCZ risk, as the risk
161 contribution of the modifier rare coding variants would not have been identified without
162 accounting for the direction derived from the burden test. Aggregation of rare variants
163 without normalizing their effects at the gene level substantially diminished our ability to
164 identify their contribution to disease risk, likely due to bias toward large genes or ones with
165 a high density of rare coding variants. The results also showed that including the variant
166 weights based on their predicted functionality (quantified as CADD scores¹⁵) improved the
167 identification of the risk contribution from rare coding variants, supporting the hypothesis
168 that functional rare variants in SCZ risk genes outside the 22q11.2 deletion region can
169 modify the risk of developing SCZ among 22q11.2DS patients. This hypothesis was further
170 supported by the observation that no risk contribution from rare synonymous variants was
171 detected (**Fig. 2d**). We further examined individual types of rare nonsynonymous variants
172 and observed an association between SCZ and the risk component from missense ($P =$
173 0.009) or loss-of-function ($P = 0.028$) rare variants (**Fig. 2d** and **Supplementary Fig. S5**).
174 The agreement in risk direction between missense and loss-of-function variants suggests
175 that most functional modifier rare variants contribute to the SCZ risk by impairing the

176 function of carrier genes (**Supplementary Fig. S5**). Of 173,752 rare variants in our study
177 cohort, a very small proportion – 2,312 variants (1.3%) – have AAF > 0.01 among non-
178 Finnish Europeans in gnomAD, a large population reference panel. After excluding them
179 from MRS calculation, we observed a small increase in the identified risk contribution
180 (Nagelkerke’s $R^2 = 4.8\%$, 99% CI: 4.1%-5.9%), likely due to specific ancestries and
181 heterogeneous effects of some excluded rare variants.

182

183 Although common variants associated with idiopathic SCZ also contribute to syndromic
184 SCZ in 22q11.2DS⁹, a finding that we replicated in our analysis (Nagelkerke’s $R^2 = 4.9\%$,
185 adjusted $P = 5.0E-04$, **Supplementary Fig. S6**), it is not clear how common and rare
186 variants together contribute to the overall genetic risk for developing SCZ in 22q11.2DS.
187 To answer this question, for all individuals in the study cohort we first calculated polygenic
188 risk scores (PRS) for idiopathic SCZ using PRSice-2 with the P -value threshold that
189 yielded the most robust result. Risk scores for modifier rare coding variants were obtained
190 by combining MRS across 500 target subcohorts (i.e., the average MRS rank) as described
191 in **Fig. 1**. We confirmed that the risk components from common and rare variants were
192 highly orthogonal to each other: 4.0% out of 4.6% and 4.3% out of 4.9% of the variance in
193 the SCZ status was explained exclusively by MRS and SCZ PRS, respectively (See
194 **Methods**). With the orthogonal nature of two different types of risk scores, we showed that
195 their combination can explain SCZ expression (Nagelkerke’s $R^2 = 8.9\%$) substantially
196 better than MRS or SCZ PRS alone (**Fig. 3**).

197

198 **Rare variants implicate genes that modify risk of SCZ in 22q11.2DS**

199

200 To better understand the contribution of rare variants to the development of SCZ in
201 22q11.2DS, we used the full integration of IGSP to identify disease risk modifying genes
202 (i.e., modifier genes hereafter) by scoring genes outside 22q11.2 deletion region for their
203 connection with SCZ based on rare nonsynonymous variants in them in the full study
204 cohort. Our statistical framework showed that rare nonsynonymous variants in the top 110
205 IGSP-scored genes (with a full integration) of a random subcohort ($n = 406$, 89% of the
206 full study cohort) can best explain the SCZ status of the remaining subjects ($n = 50$) in the
207 study cohort (**Fig. 2a**). Therefore, we selected the 110 top-scoring genes of the full study
208 cohort (**Supplementary Table S3**) as putative modifier genes and analyzed their relevant
209 clinical support, pathway enrichment, regulatory elements, and cross-ethnicity. Of these
210 110 modifier genes, 54 and 56 had a higher weighted burden of rare nonsynonymous
211 variants among SCZ cases and controls (and thus increase or decrease SCZ risk in
212 22q11.2DS), respectively. These candidate modifiers have been annotated for
213 neurodevelopmental disorders in the ClinVar database (as of July 31, 2021)
214 (**Supplementary Table S4**).

215

216 To explore the biology involved in the disease risk modulation of modifier genes, we
217 analyzed gene-set enrichment among them with 10 preselected gene sets relevant to SCZ
218 etiology (see **Methods**). We identified a significant enrichment of synaptic genes ($P =$
219 $8.29E-09$) and developmental disorder genes ($P = 9.66E-04$) (**Fig. 4a** and **Supplementary**
220 **Fig. S7**). Notably, although the 22q11.2 deletion rarely occurs, we also detected an
221 enrichment of loss-of-function intolerant genes ($P = 9.36E-06$) and missense constrained

222 genes ($P = 3.19E-05$). Using a hypothesis-free approach, we further investigated the
223 enrichment of modifier genes in gene sets for different biological processes and found that
224 gene sets with the most significant enrichment were highly relevant to neurological (e.g.,
225 neurogenesis and differentiation) and developmental biological processes (e.g., neuron
226 development and heart development) (**Fig. 4b**). While our gene-set enrichment analysis
227 revealed biological connections between modifier genes and SCZ, further examination
228 identified known SCZ genes among them with more specific pathological roles in synaptic
229 function (e.g., *TNIK*¹⁶ and *NRXNI*¹⁷) or calcium channel activity (e.g. *CACNA1C*¹⁸). In
230 addition, some modifier genes (e.g., *BDNF* and *HIF1A*)¹⁹⁻²¹ are associated with other
231 mental disorders, such as bipolar disorder and major depression disorder, that share disease
232 pathogenesis with SCZ²².

233

234 To ascertain the role of modifier genes in modulating SCZ risk, we examined their
235 expression (on the exon level) during brain development and its connection to the SCZ
236 association signals of rare variants. Using BrainSpan RNA-seq data, we identified 1,318
237 exons in 92 (83.6%) of the 110 modifier genes with general expression in developing brains
238 (See **Methods**). Association tests (SKAT-O) on 2,353 rare variants in those exons
239 confirmed a significant association with SCZ ($P = 1.44E-09$). Moreover, 1,018 exons of
240 the 110 modifier genes with specific expression in developing brains, and 1,477 rare
241 variants in them showed less evidence of association with SCZ ($P = 0.0015$).

242

243 Given the WGS data, we were interested in not only rare but low-frequency ($1\% \leq \text{AAF} <$
244 5%) noncoding variants in the regulatory elements of modifier genes. Such noncoding
245 variants are important for two reasons. First, they may constitute additional risk
246 components not covered by SCZ PRS (common variants) and MRS (rare coding variants).
247 Second, their SCZ association can confirm the involvement of the identified modifier genes
248 in developing 22q11.2DS-associated SCZ. Using SKAT-O, we tested SCZ association of
249 rare variants aggregated in four types of regulatory regions of those 110 modifier genes
250 (see **Methods**) and identified a significant SCZ association of transcriptional regulator
251 binding sites in neurons (neuron-TRBS) ($P = 1.87E-04$). Next, we added low-frequency
252 variants to the tests and identified a significant SCZ association of enhancers ($P = 4.69E-$
253 04) (**Supplementary Table S5**).

254

255 In addition to 22q11.2DS patients of European ancestry, the WGS data generated by IBBC
256 included patients of other minority ethnicities (**Supplementary Table S2**), among which
257 the Hispanic subcohort was the largest ($n = 31$; 19 SCZ cases and 12 controls). We
258 investigated whether the modifier genes identified among 22q11.2DS patients of European
259 ancestry (i.e., our study cohort) were also enriched with the modifier variants among
260 patients of other ethnicities. We first aggregated rare nonsynonymous variants in the 110
261 modifier genes in the Hispanic subcohort and tested the association with SCZ for this set
262 of variants. We did not identify a significant association (SKAT-O, $P = 0.094$). Moreover,
263 for the 110 modifier genes, our association test (SKAT-O) identified a significant SCZ
264 association of rare noncoding variants in the neuron-TRBS ($P = 0.03$) and of rare and low-
265 frequency noncoding variants in the brain enhancer regions ($P = 0.047$), respectively, in
266 this Hispanic subcohort, despite its small sample size.

267

268 Modifier genes are potential therapeutic targets for SCZ in 22q11.2DS. To investigate
269 whether they are involved in common biological processes with genes affected by SCZ
270 drugs, we analyzed the expression of modifier genes and that of genes most differentially
271 expressed in cells treated with SCZ drugs (see **Methods**). We selected 3 FDA-approved
272 antipsychotics – i.e., Haloperidol, Clozapine, and Quetiapine – and identified a significant
273 enrichment of coexpression between modifier genes and drug-induced top 10 differentially
274 expressed genes ($P = 8.4E-04$, $P = 9.4E-04$, and $P = 8.0E-04$, respectively). The
275 significance of the enrichment was not sensitive to different numbers of top differentially
276 expressed genes.

277

278 **Both modifier and 22q11.2 genes are involved in brain development**

279

280 22q11.2 deletion is the strongest known molecular genetic risk factor for SCZ²³. We
281 hypothesize that the modifier genes and 22q11.2 deletion share convergent pathological
282 mechanisms in brain development. We conducted contextualized analysis using brain
283 transcriptomic data to assess this hypothesis. First, using data from PsychENCODE²⁴ we
284 constructed gene coexpression networks corresponding to different spatiotemporal
285 combinations of regions and stages during brain development (**Supplementary Table S6**,
286 **S7**, and **S8**), following a previously developed approach²⁵. A sliding window was used to
287 combine three consecutive time periods into a time frame, so there was an overlap of
288 samples between two successive time frames. First, we assessed the enrichment of
289 coexpression among the modifier genes in the spatiotemporal combinations of brain
290 development to test the hypothesis that modifier genes are likely involved in the same
291 biological process in a spatiotemporal combination (see **Methods**). Our results showed that
292 modifier genes tended to be coexpressed in the cortical region from the neonatal period to
293 young adulthood, the limbic system from late infancy to young adulthood, and mediodorsal
294 nucleus of the thalamus and cerebellar cortex from late fetal stage to early childhood. (**Fig.**
295 **5a** and **Supplementary Table S6** and **S7**). Next, we tested whether modifier genes tended
296 to be coexpressed with genes in the 22q11.2 deletion region (hereafter 22q11.2 genes) in
297 any spatiotemporal combinations of brain development (see **Methods**). We identified
298 enriched coexpression between modifier and 22q11.2 genes with spatiotemporal
299 combinations localized to the cortical region from late infancy to young adulthood (**Fig.**
300 **5b**); five convergent spatiotemporal combinations were implicated by both modifier genes
301 and connection between modifier and 22q11.2 genes: P7-9/R1, P8-10/R1, P8-10/R2, P9-
302 11/R1, and P10-11/R1. This suggests that modifier and 22q11.2 genes likely disrupt
303 intersecting biological processes at these convergent points of postnatal cortical
304 development.

305

306 To uncover potential intersecting biological processes disrupted by modifier genes and
307 22q11.2 deletion, we carried out the weighted gene co-expression network analysis
308 (WGCNA) to identify gene coexpression modules in the aforementioned five convergent
309 spatiotemporal combinations of brain development, all of which have at least 49 samples
310 (**Supplementary Table S8**). Among them, P8-10/R2 did not produce a soft thresholding-
311 based scale free topology model fit > 0.8 (**Supplementary Fig. S8**) and thus was excluded
312 from the following module analyses. 10 ~ 18 gene coexpression modules (each with more
313 than 100 genes) were identified at each combination (see **Methods**) (**Supplementary Fig.**

314 **S9** and **Table S9-12**). We then searched for modules that may be affected by both modifier
315 and 22q11.2 genes. For the deletion, we assembled a gene set (i.e., 22q-SB-PPI genes)
316 including both 22q11.2 genes and their direct interaction partners in a spatiotemporal brain
317 protein-protein interaction network (SB-PPI) (see **Methods**) since 22q11.2 genes have
318 been shown to form a brain-developmental PPI network that may affect SCZ-associated
319 modules²⁶. As we assessed whether 22q11.2 deletion affects a module by testing its
320 enrichment of 22q-SB-PPI genes, we identified significant concurrent enrichment of both
321 modifier and 22q-SB-PPI genes in modules of three out of four spatiotemporal
322 combinations: M3_{P7-9/R1}, M4_{P9-11/R1}, and M1_{P10-11/R1} (**Fig. 5c-f**). M3_{P7-9/R1} corresponded to
323 the period from late infancy to late childhood, while M4_{P9-11/R1} and M1_{P10-11/R1}
324 corresponded to the period from middle childhood to young adulthood and the period from
325 adolescence to young adulthood, respectively. The three modules were from the same brain
326 region (R1) and highly overlapped, sharing 638 genes among them (**Supplementary Fig.**
327 **S10**). They were associated with nervous system development, cell migration and
328 angiogenesis (**Supplementary Fig. S11**) and were enriched with cell-marker genes for
329 neurons and astrocytes. To investigate specific 22q11.2 genes that contributed to the
330 observed enrichment, we tested the enrichment of SB-PPI using individual 22q11.2 genes
331 as the seed gene in these three modules. We identified significant enrichment of SB-PPI
332 for *SLC25A1* in all of them, *COMT* in M4_{P9-11/R1} and M1_{P10-11/R1}, *SEPT5* and *CRKL* in
333 M3_{P7-9/R1}, and *PI4KA* and *CLDN5* in M1_{P10-11/R1} (**Fig. 5f**).

334

335

336 **DISCUSSION**

337

338 Using an integrated approach to gene prioritization and a risk-scoring framework based on
339 bootstrapping, we analyzed the WGS-based genotype data of a 22q11.2DS cohort to
340 identify rare variants that modify syndromic SCZ risk. Our modifier risk scoring uncovered
341 potential rare coding variant-based genetic risk to develop SCZ in 22q11.2DS at the
342 individual patient level that were largely independent of common polygenic risk for
343 idiopathic SCZ. The implicated modifier genes were enriched with genes involved in
344 neurodevelopment and synaptic functions. Recent genetic studies of rare and common
345 variants for idiopathic SCZ suggest the convergent disease origin especially in synaptic
346 biology²⁷⁻²⁹ and highlight the importance of using different approaches. Our results shed
347 new light on genetic modifiers of SCZ in 22q11.2DS and show that synapse genes harbor
348 genetic variants modifying SCZ risk caused by 22q11.2 deletion, suggesting a convergent
349 disease origin between idiopathic and syndromic SCZ. In addition, our results provide not
350 only insights to improve risk prediction for 22q11.2DS-associated SCZ but genetic support
351 of drug targets for treatment in this cohort. In addition to rare coding variants, we identified
352 SCZ associations of rare and low-frequency variants in enhancer regions of modifier genes
353 active in the prefrontal cortex (PFC), suggesting that their gene regulatory activities in the
354 PFC are involved in the development of SCZ. Finally, our gene expression analysis
355 revealed a concurrent enrichment of modifier and 22q11.2-connected genes in gene
356 coexpression modules localized to the cortical region from late infancy to young adulthood.
357 This result suggested that modifiers are involved in biological pathways for the postnatal
358 cortical development perturbed by 22q11.2 deletion.

359

360 Earlier studies of 22q11.2 genes have provided important biological insights into the
361 potential impact of 22q11.2 deletion on brain development. For example, 22q11.2 genes
362 were found to be enriched with spatiotemporal PPIs during childhood that may be driven
363 by the pathologies of associated brain disorders³⁰. Also, SCZ-associated
364 neurodevelopmental modules implicated by common risk variants in idiopathic SCZ were
365 found to contain many genes in the brain developmental PPI network of 22q11.2 genes,
366 especially *SEPT5*, *PI4KA*, and *SNAP29*²⁶. In our study, we used a conceptionally different,
367 bottom-up approach to first find gene coexpression modules associated with SCZ in
368 22q11.2DS in brain development implicated by modifier genes and then investigated the
369 connection between those modules and 22q11.2 genes. Notably, we uncovered such gene
370 coexpression modules in childhood that were enriched with 22q11.2-connected genes. In
371 addition to *SEPT5* and *PI4KA*, our analysis suggested that the enrichment was driven by
372 other genes especially *COMT* and *SLC25A1*. *COMT* is one of the most widely studied
373 22q11.2 genes for SCZ due to its role in degradation of dopamine^{31,32}. *SLC25A1* is a
374 mitochondrial protein whose interactome was recently found to participate in synaptic
375 function and was altered in SCZ patients with 22q11.2DS^{33,34}. Overall, our study confirms
376 results from previous ones and provides additional insights specific to SCZ in 22q11.2DS.
377

378 Previous studies of rare coding variants in idiopathic SCZ were limited to extremely rare
379 and highly deleterious variants due to high genetic heterogeneity and selective pressure in
380 this disorder^{28,35}. We were able to demonstrate aggregate effects of rare coding variants on
381 SCZ in this 22q11.2DS cohort, despite the relatively moderate sample size, likely for three
382 reasons. First, our analytic framework incorporates a gene network and phenotypes
383 (through IGSP) to improve prioritization of risk genes. Second, the cohort shares the same
384 disease-causing copy number variant – i.e., 22q11.2 deletion – and thus likely also shares
385 certain modifier risk variants. Third, SCZ risk variants among carriers of 22q11.2 deletion
386 tend to have larger effect sizes according to the liability model. Modifier genes by nature
387 are candidate drug targets to treat 22q11.2DS-associated SCZ. Of note, modifier genes that
388 we identified include targets of antipsychotic drugs (e.g., *DRD1*, *DRD2*, and *CACNA1C*)
389 and, especially, ones that have not been well-studied (e.g., *C3* and *OPRK1*). This highlights
390 the potential of studying modifier rare coding variants in 22q11.2DS to identify therapeutic
391 targets for SCZ in the general population.
392

393 The recent SCHEMA study implicated ultra-rare coding variants (URVs) in 10 genes as
394 conferring substantial risk for SCZ (odds ratios of 3–50, $P < 2.14E-06$) and 32 genes at a
395 false discovery rate of $<5\%$ ²⁸. Among the latter, *NR3C2* also appeared as a modifier risk
396 gene in our study. Two reasons may account for the lack of more substantial overlap. First,
397 we tested SCZ association based on the weighted burden of all nonsynonymous rare
398 variants in a gene, not limited to ultra-rare coding variants. Second, many ultra-rare coding
399 variants with large effect sizes may be too rare to be observed in a study cohort with
400 hundreds of subjects. Nevertheless, those genes implicated by ultra-rare coding variants
401 are also highly relevant to synapse functions, suggesting a convergence of disease
402 mechanisms at the pathway level regardless whether the disease is part of a syndrome.
403

404 Having shown that modifier genes are enriched in modules associated with nervous system
405 development and cell migration, we considered whether the mouse phenotype data

406 integration could potentially add scoring bias. As per the design of the IGSP algorithm,
407 genes with certain phenotypes in human tend to have higher phenotype scores only if
408 stronger gene association signals are enriched among genes whose mouse orthologs have
409 annotations of relevant phenotypes. Therefore, although certain types of mouse knockout
410 phenotypes (e.g., of the nervous system) are better studied and thus probably more
411 complete than others, such a research variation does not introduce scoring bias in IGSP
412 since the distribution of gene association signals is independent of the degree of
413 completeness of mouse phenotype annotation. On the other hand, the power of our
414 integrative statistical framework will be reduced when the disease and related phenotypes
415 are less studied in mouse models. While the moderate sample size is no doubt a limiting
416 factor of this study, this problem is being actively addressed by the Consortium through
417 continuously recruiting more 22q11.2DS patients.

418
419 In addition to 22q11.2 deletion and SCZ-associated common variants, modifier rare
420 variants also contribute to the disease risk of syndromic SCZ. Although their aggregate
421 effects may not be directly significant to idiopathic SCZ, given the genetic heterogeneity
422 of the disease in general, the identification and analysis of modifier rare variants is
423 complementary to that of highly pathogenic ultra-rare variants in SCZ of the general
424 population. For example, our study shows how rare variants across genes important in brain
425 development collectively contribute to substantial risk in syndromic SCZ, which shares
426 similar clinical features with idiopathic SCZ. This sheds light on important biological
427 processes in the etiology of the disease and highlights the value of studying SCZ using
428 22q11.2DS as a model.

430 431 **METHODS**

432 433 **22q11.2DS cohort and rare variants**

434
435 The original WGS data consisted of 519 unrelated 22q11.2DS patients recruited across 22
436 international sites and underwent stringent measures of quality control⁹ (**Supplementary**
437 **note**). The study complies with all relevant ethical regulations and was approved by local
438 institutional research ethics boards and have informed consent from all studied subjects.
439 Among 519 22q11.2DS patients, there were 259 SCZ cases who have been diagnosed at
440 any age by a stringent case consensus procedure and 260 controls who had no history of
441 any psychotic illness when assessed at age ≥ 25 years. To maximize the sample size with
442 a homogeneous genetic background, our study cohort was the subcohort of European
443 ancestry that included 223 SCZ cases and 233 controls (**Supplementary note**). The largest
444 subcohort of minorities, consisting of 31 subjects of Hispanic ancestry, was used to
445 investigate whether modifier genes identified in the study cohort were enriched with
446 modifier rare variants. WGS was performed with the Illumina pipeline. Sequence
447 alignment was carried out with PEMapper to map WGS reads to the human genome build
448 hg38. Variants were called with PEPcaller⁹ (**Supplementary note**). Rare variants were
449 defined as variants with AAF $< 1\%$ in the study cohort. Rare variants with a missing
450 genotype rate > 0.01 in the study cohort were excluded from the analyses. This study

451 focused on autosomal rare variants outside the 22q11.2 deletion region; variants in the
452 22q11.2 deletion region were not analyzed.

453

454 **Variant annotation method**

455

456 We identified coding variants using CADD annotation (v.1.6, 'CodingTranscript'). In our
457 study, different types of coding variants were defined by the Ensembl Variant Effect
458 Predictor (VEP) as part of CADD: nonsynonymous variants are coding variants not
459 annotated as 'synonymous' variants, while loss-of-function variants are either 'stop_gained'
460 or 'frameshift' variants.

461

462 **Modifier rare coding variants**

463

464 The statistical framework that we used to identify and analyze modifier rare coding variants
465 included sample bootstrapping, risk gene prioritization based on rare coding variants (the
466 IGSP method), modifier risk calculation and evaluation to examine the aggregate effects
467 of rare coding variants on SCZ risk (**Fig. 1**).

468

469 First, we bootstrapped the study cohort to randomly generate 500 target subcohorts and
470 their paring discovery subcohorts. In each iteration, we randomly selected $n_t/2$ cases and
471 $n_t/2$ controls from the study cohort to create the target subcohort; the remaining $223-n_t/2$
472 cases and $233-n_t/2$ controls in the study cohort constituted its corresponding discovery
473 subcohort. n_t determines not only the sample size of target subcohorts and thus the rank
474 resolution (see details below) but also the sample size of the corresponding discovery
475 subcohorts and thus the power for prioritizing risk genes. In this study, we set $n_t = 50$ and
476 showed that our result was not sensitive to selection of n_t within a reasonable range
477 (**Supplementary Fig. 12**). Our sample bootstrapping procedure ensured that every
478 individual in the study cohort was included in at least one target subcohort. Using each
479 discovery subcohort, we applied IGSP to prioritize genes for risk modification based on
480 the SCZ association of rare coding variants weighted by orthogonal information contained
481 in the gene network and mouse phenotypes. Specifically, we collected gene association
482 signals of each protein-coding gene outside the 22q11.2 deletion region by applying the
483 weighted burden test (using the R package SKAT) to rare nonsynonymous variants with
484 their CADD scores (v.1.6). The test used the SCZ status as the phenotype and included as
485 covariates sex and the top 10 principal components to account for the subpopulation
486 structure, which were obtained using PLINK (v.1.9) based on common variants (MAF >
487 0.05). The output association P -values were then used as input to IGSP to score and rank
488 genes.

489

490 Next, we defined a modifier risk score for each individual in the target subcohorts based
491 on rare nonsynonymous variants in the top n putative modifier genes:

492

$$\sum_{i=1}^n \left(\frac{d_i}{v_i} \sum_{j=1}^{v_i} c_{ij} \cdot a_{ij} \right) \quad (1)$$

493

494 in which d_i is the risk direction of gene i , taking on the value of 1 or -1 if the direction of
 495 variant burden is on SCZ cases or controls, respectively, v_i the number of rare
 496 nonsynonymous variants in gene i , c_{ij} and a_{ij} the CADD score and the number of alternative
 497 alleles of rare nonsynonymous variant j in gene i , respectively. For each putative modifier
 498 gene (one of the top n genes scored by IGSP), the equation calculates the weighted sum of
 499 alternative alleles for rare nonsynonymous variants in the gene weighted by the
 500 corresponding CADD scores. To avoid scoring bias to large genes and genes harboring a
 501 high density of rare coding variants, we performed a normalization process in which the
 502 weighted sum was divided by the number of rare nonsynonymous variants for the same
 503 gene observed in the corresponding discovery subcohort.

504
 505 Finally, we evaluated whether the MRS explained the variance in SCZ status in the target
 506 subcohorts. Since MRS across different target subcohorts may involve different top scoring
 507 genes and thus are not directly comparable, we ranked subjects in each target subcohort in
 508 ascending order of their MRS and summarized each subject's modifier risk by taking the
 509 average of his MRS ranks over target subcohorts. The relationship between the summarized
 510 MRS and SCZ status was evaluated using logistic regression including as covariates sex
 511 and the top 10 principal components for correcting population substructure. Subjects may
 512 appear in different target subcohorts, which could introduce uncertainty to average MRS
 513 rank. We used another bootstrapping procedure to evaluate rare variant SCZ risk
 514 characterized by the average MRS rank. Briefly, in each bootstrap replicate, we calculated
 515 the variance of SCZ status explained by the average MRS rank from 20 random
 516 observations of each subject (i.e., 20 random target subcohorts including the subject)
 517 (**Supplementary Fig. S13**). The estimated variance of SCZ status explained by the average
 518 MRS rank and 99% confidence interval are derived from 201 replicates (median and the
 519 range after removing the first and last estimates after sorting, respectively).

520
 521 We started the process with the MRS based on the top 10 genes and repeated the process
 522 by including the next 10 genes until the top 500 genes were examined. Overall, 50 tests are
 523 carried out to test relationship between MRS and SCZ status; however, the tests are
 524 correlated because the tested genes are highly overlapped. Therefore, we used the minimal
 525 P -value test³⁶ to calculate the corrected P -value of the most significant association in 50
 526 tests. First, we calculated e , the effective number of independent tests as :

$$e = M - \sum_{i=1}^M [I(\lambda_i > 1)(\lambda_i - 1)], \quad (2)$$

528
 529 in which M equals the number of tests (50), λ_s are the eigenvalues of the $M \times M$ correlation
 530 matrix of the P -values of M tests, and I is an indicator function. The P -value correlation
 531 matrix can be calculated based on the Pearson correlation coefficient between the vectors
 532 of P -values from the 201 bootstrap replicates. Next, given e and the lowest P -value among
 533 M tests (P_{\min}), the corrected P -value can be calculated as:

$$P = 1 - (1 - P_{\min})^e. \quad (3)$$

535

536 **Common polygenic risk for idiopathic SCZ**

537

538 We calculated common polygenic risk of idiopathic SCZ for individuals with 22q11.2DS
539 based on the PRS analysis using PRSice-2^{37,38}. We first downloaded the summary statistics
540 of the idiopathic SCZ GWAS²⁷ from the Psychiatric Genomics Consortium (PGC). We
541 selected common SNPs in our 22q11.2DS cohort (internal MAF > 5%) and carried out LD
542 clumping if they were within 250 kbps and $R^2 > 0.1$. Next, we used 19 *P*-value thresholds
543 – 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.01, 1E-3, 1E-4, 1E-5, 1E-6, 1E-7, 1E-8, 1E-
544 9, and 1E-10 – to select SNPs for PRS scoring after clumping. Sex and the top 10 principal
545 components for population substructure correction were used as covariates in PRSice-2.

546

547 **Cross-analysis of SCZ risk from common and rare coding variants**

548

549 We analyzed SCZ status using four logistic regression models with (1) SCZ PRS and
550 average MRS rank as predictor variables, and sex and top 10 principal components for
551 population substructures as covariates, (2) SCZ PRS and covariates, (3) average MRS rank
552 and covariates, and (4) covariates only. The independent risk component exclusively from
553 common and rare coding variants, respectively, can be evaluated across these models. The
554 combined genetic risk scores that characterize the overall genetic risk from both common
555 and rare coding variants for SCZ in 22q11.2DS can be derived by applying coefficients
556 from the logistic regression model (1).

557

558 **Gene-set enrichment analysis**

559

560 We systematically investigated gene sets enriched with 110 modifier genes using both
561 hypothesis-driven and hypothesis-free approaches. In the hypothesis-driven approach, we
562 investigated the enrichment of the modifier genes in 10 gene sets that were commonly used
563 in SCZ studies: synaptic genes, FMRP target, postsynaptic density from human neocortex
564 (hPSD), neurotransmitter system, presynaptic genes, calcium channel activity,
565 developmental disorder, loss-of-function intolerant genes, and missense constrained genes
566 (**Supplementary Table S13**). The first eight gene sets are relevant to SCZ biology and the
567 last two were enriched with rare variants in idiopathic SCZ³⁵. In the hypothesis-free
568 approach, we investigated the enrichment of modifier genes in 7,481 gene sets annotated
569 with gene ontology (GO) terms for different biological processes from the Molecular
570 Signatures Database (v.7.4)^{39,40}. For each gene set, we considered only the subset of genes
571 that was scored by IGSP. We used Fisher's exact test to assess the significance of
572 enrichment of modifier genes for a gene set, using 8,028 genes scored by IGSP (the full
573 integration requires the phenotype annotation of mouse gene knockouts, which cover only
574 half of human gene homologs) as the background. Bonferroni multiple test correction was
575 used to correct *P*-values for 10 and 7,481 tests performed in the two approaches,
576 respectively.

577

578 To investigate whether a gene coexpression module was enriched with modifier genes, we
579 used Fisher's exact test and considered only genes in modules scored by IGSP (8,028 genes
580 as the background). To investigate whether a gene coexpression module was enriched with

581 22q-SB-PPI genes or cell-marker genes⁴¹, we used Fisher's exact test with all 21,196
582 coding genes as the background.

583

584 **Exon expression of modifier genes**

585

586 We used RNA-seq gene expression data at the exon level from BrainSpan
587 (<https://www.brainspan.org/static/download.html>)⁴² to identify regions (exons) of modifier
588 genes with general and specific expression during brain development, defined as exons
589 with median RPKM ≥ 1 and < 1 , respectively, across all samples cataloged in BrainSpan
590 regardless of brain regions or age.

591

592 **Regulatory regions of modifier genes**

593

594 We investigated noncoding rare variants in four different types of regulatory regions of
595 modifier genes: promoters, brain-active enhancers, transcriptional regulator binding sites
596 in neurons (neuron-TRBS), and DNase I hypersensitive sites in neural tissues (neural-
597 DHS). We defined a gene's promoter as the genomic region within 500 bp of its
598 representative transcription start site based on FANTOM CAGE data⁴³ and obtained its
599 brain-active enhancer(s) (in the PFC) from the gene regulatory network generated by
600 PsychENCODE⁴⁴. We collected neuron-TRBS and neural-DHS as potential regulatory
601 elements for a gene within its 50-kb upstream and 50-kb downstream regions from ReMap
602 database (Homo sapiens; nonredundant peaks)⁴⁵ and an online repository
603 (<https://zenodo.org/record/3838751#.Y7yD3i-B2-x>)⁴⁶, respectively.

604

605 **SCZ association for a set of variants**

606

607 We used the weighted burden test to derive rare variant associations at the gene level as
608 input to IGSP because the direction of rare variant burden was required to calculate MRS.
609 For other analyses that aimed to simply test SCZ association for a set of variants, we used
610 weighted SKAT-O⁴⁷ (with CADD scores as the variant weights). When testing associations
611 in the study cohort ($n = 456$) and discovery subcohorts ($n = 406$), we used sex and the top
612 10 principal components for the subpopulation structure correction as covariates. When
613 testing associations in the Hispanic subcohort ($n = 31$), we used sex and the top 4 principal
614 components as covariates.

615

616 **Modifier genes and drug-induced differentially expressed genes**

617

618 We tested whether modifier genes tend to be coexpressed with differentially expressed
619 genes induced by SCZ drugs. Top differentially expressed genes were identified using
620 drug-induced gene expression signatures from the Expanded CMap LINCS Resource 2020
621 (<https://clue.io/data/CMap2020#LINCS2020>)⁴⁸. Briefly, for an antipsychotic drug, we
622 identified in CMap n genes that most frequently appear in the list of top n most
623 overexpressed and underexpressed genes, respectively, across different treatments of the
624 drug. For this coexpression analysis, we focused on brain tissues between middle childhood
625 and young adulthood. We considered two gene to be coexpressed if there is significant
626 coexpression between them in any one of the 12 spatiotemporal combinations of four brain

627 regions and three stages of brain development (**Supplementary Tables S6, S7, and S8**)
628 (see the next subsection). We assessed the degree of coexpression using a permutation test
629 with 100,000 iterations of randomization, in each of which the degree of coexpression
630 between 110 random IGSP-scored genes and drug-induced top differentially expressed
631 genes was calculated to construct the null distribution.

632

633 **Transcriptional analysis in brain development**

634

635 Gene expression data were downloaded from PsychENCODE
636 (<http://development.psychencode.org/>). Four brain regions and 11 overlapping stages of
637 brain development defined in a previous study²⁵ were used to construct 44 spatiotemporal
638 combinations (**Supplementary Tables S6, S7 and S8**). For each spatiotemporal
639 combination, we measured high-confidence coexpression (1 or 0) based on a stringent
640 threshold of Pearson correlation coefficient (the absolute value ≥ 0.85). We measured the
641 degree of coexpression among modifier genes based on the sum of high-confidence
642 coexpression between each pair of modifier genes. To assess the degree of coexpression
643 among modifier genes in a spatiotemporal combination, we performed a permutation test
644 with 100,000 iterations of randomization, in each of which the degree of coexpression
645 among 110 random IGSP-scored genes was calculated to construct the null distribution.
646 We measured the degree of coexpression with 22q11.2 genes for each IGSP-scored gene
647 based on the number of high-confidence coexpression between the gene and each of 46
648 22q11.2 genes⁴⁹. To investigate whether modifier genes tended to be coexpressed with
649 22q11.2 genes, we performed logistic regression to regress the status of modifier genes
650 (i.e., 110 modifier genes and the remaining 7,918 IGSP-scored genes as non-modifier
651 genes) on the degree of coexpression with 22q11.2 genes. Bonferroni correction was
652 applied to 44 spatiotemporal combinations. To identify gene coexpression modules in a
653 spatiotemporal combination, we performed the weighted gene co-expression network
654 analysis (WGCNA), which uses a soft-thresholding method to better detect gene modules.
655 First, to determine a soft thresholding power, we run “pickSoftThreshold” function to
656 obtain the first soft thresholding power (starting from 2 to 15) of which the corresponding
657 scale-free topology model fit $R^2 > 0.8$ without considering sign of coexpression
658 (networktype = 'unsigned'). Next, we transformed the adjacency matrix of a coexpression
659 network to topological overlap matrix (TOM; unsigned 'networkType' and 'TOMType')
660 using the selected soft thresholding power and clustered genes based on the corresponding
661 dissimilarity as the distance measure. The gene clustering was based on a hierarchical
662 clustering function 'flashClust' (the 'average' method). Last, we determined gene modules
663 by running 'cutreeDynamic' function (the 'tree' method) with a minimum module size =
664 100 genes. Genes in a coexpression network not classified into any modules (i.e., grey
665 genes) were excluded from our module analyses.

666

667 **Spatiotemporal brain-PPI network of 22q11.2 genes**

668

669 We defined a spatiotemporal brain protein-protein interaction network of 22q11.2 (22q-
670 SB-PPI) genes for a spatiotemporal combination as genes that have spatiotemporal brain
671 protein-protein interactions with at least one 22q11.2 gene. We determined a
672 spatiotemporal brain protein-protein interaction between two genes if they satisfied two

673 conditions: first, their protein products physically interacted according to MIST (v.5.0,
674 Homo sapiens)⁵⁰; second, both genes are transcribed in a specific spatiotemporal
675 combination and have evidence of coexpression at transcriptional level (i.e., the absolute
676 value of Pearson correlation coefficient ≥ 0.7).

677
678

679 **DATA AVAILABILITY**

680

681 This study is a secondary data analysis of the whole genome sequencing data of SCZ in
682 22q11.2DS generated by the International 22q11.2DS Brain and Behavior Consortium
683 (IBBC). All summary statistics for the SCZ association of rare coding variants in the IBBC
684 22q11.2DS cohort are available at <http://zdzlab.einsteinmed.org/1/sz-22q.html>. All
685 predicted modifier genes are available in the supplementary material. Due to privacy
686 concerns for our research participants, individual-level genetic data from the IBBC study
687 of SCZ in 22q11.2DS are not publicly available; however, access to anonymized data can
688 be requested from a qualified academic investigator to the IBBC Executive Committee,
689 providing the data transfer is approved by the Institutional Review Board and regulated by
690 a material transfer agreement.

691
692

693 **CODE AVAILABILITY**

694

695 IGSP is publicly available at <https://zenodo.org/record/1034362#.X-JWQNgzY2w>. All
696 other software used in our analyses is open source and described in Methods.

697
698

699 **ACKNOWLEDGEMENTS**

700

701 This work was supported by funding to ZDZ from NIH Grants U01 MH101720, R01
702 AG057909, U19 AG056278, R01 AG061521, R01 AG057706, RF1 AG057341, R01
703 AG061155, P01 AG047200, P01 AG017242, P30 AG038072, and a Career Scientist
704 Award from the Irma T. Hirschl Trust.

705

706 **COMPETING INTERESTS**

707

708 All authors declare no competing interests.

709

710

711 **REFERENCES**

712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755

1. Edlmann, L., Pandita, R.K. & Morrow, B.E. Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am J Hum Genet* **64**, 1076-86 (1999).
2. Shaikh, T.H. *et al.* Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet* **9**, 489-501 (2000).
3. Tezenas Du Montcel, S., Mendizabai, H., Ayme, S., Levy, A. & Philip, N. Prevalence of 22q11 microdeletion. *J Med Genet* **33**, 719 (1996).
4. Botto, L.D. *et al.* A population-based study of the 22q11.2 deletion: phenotype, incidence, and contribution to major birth defects in the population. *Pediatrics* **112**, 101-7 (2003).
5. Murphy, K.C., Jones, L.A. & Owen, M.J. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry* **56**, 940-5 (1999).
6. Fung, W.L. *et al.* Elevated prevalence of generalized anxiety disorder in adults with 22q11.2 deletion syndrome. *Am J Psychiatry* **167**, 998 (2010).
7. Bassett, A.S. *et al.* Rare Genome-Wide Copy Number Variation and Expression of Schizophrenia in 22q11.2 Deletion Syndrome. *Am J Psychiatry* **174**, 1054-1063 (2017).
8. Gur, R.E. *et al.* A neurogenetic model for the study of schizophrenia spectrum disorders: the International 22q11.2 Deletion Syndrome Brain Behavior Consortium. *Mol Psychiatry* **22**, 1664-1672 (2017).
9. Cleyne, I. *et al.* Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion. *Mol Psychiatry* (2020).
10. Sumitomo, A. *et al.* A mouse model of 22q11.2 deletions: Molecular and behavioral signatures of Parkinson's disease and schizophrenia. *Sci Adv* **4**, eaar6637 (2018).
11. Davies, R.W. *et al.* Using common genetic variation to examine phenotypic expression and risk prediction in 22q11.2 deletion syndrome. *Nat Med* **26**, 1912-1918 (2020).
12. Merico, D. *et al.* Whole-Genome Sequencing Suggests Schizophrenia Risk Mechanisms in Humans with 22q11.2 Deletion Syndrome. *G3 (Bethesda)* **5**, 2453-61 (2015).
13. Michaelovsky, E. *et al.* Risk gene-set and pathways in 22q11.2 deletion-related schizophrenia: a genealogical molecular approach. *Transl Psychiatry* **9**, 15 (2019).
14. Lin, J.R., Zhang, Q., Cai, Y., Morrow, B.E. & Zhang, Z.D. Integrated rare variant-based risk gene prioritization in disease case-control sequencing studies. *PLoS Genet* **13**, e1007142 (2017).
15. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894 (2019).
16. Wang, Q. *et al.* The psychiatric disease risk factors DISC1 and TNIK interact to regulate synapse composition and function. *Mol Psychiatry* **16**, 1006-23 (2011).
17. Kirov, G. *et al.* Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophr Bull* **35**, 851-4 (2009).

- 756 18. Bhat, S. *et al.* CACNA1C (Cav1.2) in the pathophysiology of psychiatric disease.
757 *Prog Neurobiol* **99**, 1-14 (2012).
- 758 19. Fernandes, B.S. *et al.* Peripheral brain-derived neurotrophic factor (BDNF) as a
759 biomarker in bipolar disorder: a meta-analysis of 52 studies. *BMC Med* **13**, 289
760 (2015).
- 761 20. Kishi, T., Yoshimura, R., Ikuta, T. & Iwata, N. Brain-Derived Neurotrophic Factor
762 and Major Depressive Disorder: Evidence from Meta-Analyses. *Front Psychiatry*
763 **8**, 308 (2017).
- 764 21. Shibata, T. *et al.* The alteration of hypoxia inducible factor-1 (HIF-1) and its target
765 genes in mood disorder patients. *Prog Neuropsychopharmacol Biol Psychiatry* **43**,
766 222-9 (2013).
- 767 22. Brosch, K. *et al.* Reduced hippocampal gray matter volume is a common feature of
768 patients with major depression, bipolar disorder, and schizophrenia spectrum
769 disorders. *Mol Psychiatry* **27**, 4234-4243 (2022).
- 770 23. Bassett, A.S. *et al.* Clinically detectable copy number variations in a Canadian
771 catchment population of schizophrenia. *J Psychiatr Res* **44**, 1005-9 (2010).
- 772 24. Li, M. *et al.* Integrative functional genomic analysis of human brain development
773 and neuropsychiatric risks. *Science* **362**(2018).
- 774 25. Willsey, A.J. *et al.* Coexpression networks implicate human midfetal deep cortical
775 projection neurons in the pathogenesis of autism. *Cell* **155**, 997-1007 (2013).
- 776 26. Forsyth, J.K. *et al.* Synaptic and Gene Regulatory Mechanisms in Schizophrenia,
777 Autism, and 22q11.2 Copy Number Variant-Mediated Risk for Neuropsychiatric
778 Disorders. *Biol Psychiatry* **87**, 150-163 (2020).
- 779 27. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology
780 in schizophrenia. *Nature* **604**, 502-508 (2022).
- 781 28. Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for
782 schizophrenia. *Nature* **604**, 509-516 (2022).
- 783 29. Iyegbe, C.O. & O'Reilly, P.F. Genetic origins of schizophrenia find common
784 ground. *Nature* **604**, 433-435 (2022).
- 785 30. Lin, G.N. *et al.* Spatiotemporal 16p11.2 protein network implicates cortical late
786 mid-fetal brain development and KCTD13-Cul3-RhoA pathway in psychiatric
787 diseases. *Neuron* **85**, 742-54 (2015).
- 788 31. Egan, M.F. *et al.* Effect of COMT Val108/158 Met genotype on frontal lobe
789 function and risk for schizophrenia. *Proc Natl Acad Sci U S A* **98**, 6917-22 (2001).
- 790 32. Williams, H.J., Owen, M.J. & O'Donovan, M.C. Is COMT a susceptibility gene for
791 schizophrenia? *Schizophr Bull* **33**, 635-41 (2007).
- 792 33. Gokhale, A. *et al.* Systems Analysis of the 22q11.2 Microdeletion Syndrome
793 Converges on a Mitochondrial Interactome Necessary for Synapse Function and
794 Behavior. *J Neurosci* **39**, 3561-3581 (2019).
- 795 34. Li, J. *et al.* Mitochondrial deficits in human iPSC-derived neurons from patients
796 with 22q11.2 deletion syndrome and schizophrenia. *Transl Psychiatry* **9**, 302
797 (2019).
- 798 35. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among
799 4,877 individuals with schizophrenia. *Nat Neurosci* **19**, 1433-1441 (2016).
- 800 36. Flannick, J. *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440
801 controls. *Nature* **570**, 71-76 (2019).

- 802 37. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score software.
803 *Bioinformatics* **31**, 1466-8 (2015).
- 804 38. Choi, S.W. & O'Reilly, P.F. PRSice-2: Polygenic Risk Score software for biobank-
805 scale data. *Gigascience* **8**(2019).
- 806 39. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach
807 for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**,
808 15545-50 (2005).
- 809 40. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics*
810 **27**, 1739-40 (2011).
- 811 41. Zhang, X. *et al.* CellMarker: a manually curated resource of cell markers in human
812 and mouse. *Nucleic Acids Res* **47**, D721-D728 (2019).
- 813 42. BrainSpan Atlas of the Developing Human Brain [Internet]. Available at
814 <http://brainspan.org>. (2011).
- 815 43. Consortium, F. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**,
816 462-70 (2014).
- 817 44. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model
818 for the human brain. *Science* **362**(2018).
- 819 45. Hammal, F., de Langen, P., Bergon, A., Lopez, F. & Ballester, B. ReMap 2022: a
820 database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from
821 an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res*
822 **50**, D316-D325 (2022).
- 823 46. Meuleman, W. *et al.* Index and biological spectrum of human DNase I
824 hypersensitive sites. *Nature* **584**, 244-251 (2020).
- 825 47. Lee, S. *et al.* Optimal unified approach for rare-variant association testing with
826 application to small-sample case-control whole-exome sequencing studies. *Am J*
827 *Hum Genet* **91**, 224-37 (2012).
- 828 48. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and
829 the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e17 (2017).
- 830 49. Zinkstok, J.R. *et al.* Neurobiological perspective of 22q11.2 deletion syndrome.
831 *Lancet Psychiatry* **6**, 951-960 (2019).
- 832 50. Hu, Y. *et al.* Molecular Interaction Search Tool (MIST): an integrated resource for
833 mining gene and protein interaction data. *Nucleic Acids Res* **46**, D567-D574 (2018).
834
835

836 **FIGURE LEGEND**

837

838 **Figure 1. Statistical framework to examine the aggregate effects of modifier rare**
839 **coding variants on SCZ.** We first randomly split the study cohort of SCZ in 22q11.2DS
840 into a target subcohort with 25 SCZ cases and 25 controls and a discovery subcohort with
841 the remaining subjects. We then calculated the relative SCZ risk of subjects in a target
842 subcohort based on their rare coding variants in genes prioritized in the corresponding
843 discovery subcohort (MRS, Equation 1). We repeated the calculation 500 times and
844 evaluated MRS for SCZ across subcohorts based on the average MRS rank.

845

846 **Figure 2. SCZ risk from rare coding variants in 22q11.2DS. a.** SCZ status and MRS.
847 The plot shows the fraction of variance in SCZ status in target subcohorts explained by the
848 MRS (y-axis) based on rare nonsynonymous variants in the top n genes prioritized by IGSP
849 scoring with a full integration (using gene network and phenotypes) in discovery
850 subcohorts (x-axis). D(+/-) denotes that a higher MRS corresponds to a higher/lower risk
851 of SCZ. The vertical bars denote a 99% confidence interval. **b.** Odds ratio. To directly
852 evaluate whether MRS quantifies the SCZ risk, for each of 500 target subcohorts, we
853 calculated odds ratio of SCZ patients using top 50% MRS (based on the top 110 genes in
854 the discovery subcohort) as the exposure. The x- and y-axis show the number of observed
855 rare nonsynonymous variants and the number of affected genes, respectively, in a target
856 subcohort. **c.** SCZ status vs. MRS calculated in four different ways. S1 is for using Equation
857 1, while S2, S3, and S4 denote using Equation 1 without c_{ij} for considering the predicted
858 functionality of rare coding variants (i.e., CADD-score weighting), without v_i for
859 normalizing variant effects at the gene level, and without d_i for considering risk direction,
860 respectively. **d.** SCZ status vs. MRS calculated for different types of rare coding variants.

861

862 **Figure 3. Composition of genetic risk of SCZ in the study cohort.** The horizontal bars
863 show min-max normalized PRS, average MRS rank, and combined risk scores for 456
864 22q11.2DS patients with and without SCZ (i.e., cases and controls). The combined risk
865 scores were calculated by applying coefficients of PRS and average MRS rank obtained
866 from a logistic regression model (with the SCZ status as outcome and sex and top 10
867 principal components for population substructure correction as covariates). The same SCZ
868 cases are connected by pink lines. Three SCZ cases with PRS lower than the median but
869 with high MRS are highlighted by red connecting lines.

870

871 **Figure 4. Gene-set enrichment analysis of modifier genes. a.** Ten gene sets known to be
872 highly relevant to SCZ. **b.** Gene sets for GO terms of biological processes. Of the 7,481
873 such gene sets, only 10 with the most significant enrichment are shown. Nominal
874 enrichment P -values calculated by Fisher's exact tests are indicated. The error bars
875 represent a 95% confidence interval. The ones in red denote significant enrichment
876 (adjusted $P < 0.05$) after Bonferroni correction.

877

878 **Figure 5. Transcriptomic analyses of modifier and 22q11.2 genes in brain**
879 **development. a.** Enrichment of coexpression among modifier genes in brain development.
880 In each spatiotemporal combination, we tested whether modifier genes tended to be
881 coexpressed using permutation tests. **b.** Enrichment of coexpression between modifier and

882 22q11.2 genes in brain development. In each spatiotemporal combination, we tested
883 whether modifier genes tended to be coexpressed with 22q11.2 genes using logistic
884 regression. **c-f.** Concurrent enrichment of modifier and 22q-SB-PPI genes in modules of
885 P7-9/R1, P8-10/R1, P9-11/R1, and P10-11/R1. The heatmap shows the enrichment of cell
886 markers among module genes for different cell types (only relevant cell types with at least
887 50 marker genes cataloged in CellMarker⁴¹ were considered). N: Neuron. A: Astrocyte. M:
888 Macrophage. O: Oligodendrocyte. **g.** Contribution of specific 22q11.2 genes to the
889 enrichment of 22q-SB-PPI genes in M3_{P7-9/R1}, M4_{P9-11/R1}, and M1_{P10-11/R1}. One, two, and
890 three asterisks denote Bonferroni adjusted $P < 0.05$, < 0.01 and < 0.001 for 138 tests (46
891 22q11.2 genes and 3 modules), respectively.
892