

# Assortative mating and parental genetic relatedness drive the pathogenicity of variably expressive variants

Corrine Smolen,<sup>1,2</sup> Matthew Jensen,<sup>1,2</sup> Lisa Dyer,<sup>3</sup> Lucilla Pizzo,<sup>1</sup> Anastasia Tyryshkina,<sup>1,4</sup> Deepro Banerjee,<sup>1,2</sup> Laura Rohan,<sup>1</sup> Emily Huber,<sup>1</sup> Laila El Khattabi,<sup>5</sup> Paolo Prontera,<sup>6</sup> Jean-Hubert Caberg,<sup>7</sup> Anke Van Dijck,<sup>8</sup> Charles Schwartz,<sup>9</sup> Laurence Faivre,<sup>10,11</sup> Patrick Callier,<sup>10,11</sup> Anne-Laure Mosca-Boidron,<sup>12</sup> Mathilde Lefebvre,<sup>11</sup> Kate Pope,<sup>13</sup> Penny Snell,<sup>13</sup> Paul J. Lockhart,<sup>13,14</sup> Lucia Castiglia,<sup>15</sup> Ornella Galesi,<sup>15</sup> Emanuela Avola,<sup>15</sup> Teresa Mattina,<sup>16</sup> Marco Fichera,<sup>15,16</sup> Giuseppa Maria Luana Mandarà,<sup>17</sup> Maria Grazia Bruccheri,<sup>15</sup> Olivier Pichon,<sup>18</sup> Cedric Le Caignec,<sup>19,20</sup> Radka Stoeva,<sup>21</sup> Silvestre Cuiat,<sup>18</sup> Sandra Mercier,<sup>18</sup> Claire Bénéteau,<sup>18</sup> Sophie Blesson,<sup>22</sup> Ashley Nordsletten,<sup>23</sup> Dominique Martin-Coignard,<sup>22</sup> Erik Sistermans,<sup>24</sup> R. Frank Kooy,<sup>8</sup> David J. Amor,<sup>14</sup> Corrado Romano,<sup>16,17</sup> Bertrand Isidor,<sup>18</sup> Jane Juusola,<sup>3</sup> and Santhosh Girirajan\*<sup>1,2,4,25</sup>

1. Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA.
2. Bioinformatics and Genomics Graduate program, Pennsylvania State University, University Park, PA 16802, USA.
3. GeneDx, Gaithersburg, MD 20877, USA.
4. Neuroscience Graduate program, Pennsylvania State University, University Park, PA 16802
5. Assistance Publique–Hôpitaux de Paris, Department of Medical Genetics, Armand Trousseau and Pitié-Salpêtrière Hospitals, Paris, France.
6. Medical Genetics Unit, Hospital "Santa Maria della Misericordia", Perugia, Italy.
7. Centre Hospitalier Universitaire de Liège. Domaine Universitaire du Sart Tilman, Liège, Belgium.

- 25 8. Department of Medical Genetics, University and University Hospital Antwerp, Antwerp,  
26 Belgium.
- 27 9. Greenwood Genetic Center, Greenwood, SC 29646, USA.
- 28 10. Centre de Genetique et Centre de Référence Anomalies du développement et syndromes  
29 malformatifs, Hôpital d'Enfants, CHU Dijon, Dijon, France.
- 30 11. GAD INSERM UMR1231, FHU TRANSLAD, Université de Bourgogne Franche Comté, Dijon,  
31 France
- 32 12. Laboratoire de Genetique Chromosomique et Moleculaire, CHU Dijon, France.
- 33 13. Department of Paediatrics, University of Melbourne, Melbourne, Australia.
- 34 14. Bruce Lefroy Center, Murdoch Children's Research Institute, Melbourne, Australia
- 35 15. Research Unit of Rare Diseases and Neurodevelopmental Disorders, Oasi Research Institute-  
36 IRCCS, 94018 Troina, Italy
- 37 16. Medical Genetics, Department of Biomedical and Biotechnological Sciences, University of  
38 Catania, 95123 Catania, Italy.
- 39 17. Medical Genetics, ASP Ragusa, Ragusa, Italy.
- 40 18. CHU Nantes, Department of Medical Genetics, Nantes, France.
- 41 19. CHU Toulouse, Department of Medical Genetics, Toulouse, France
- 42 20. ToNIC, Toulouse Neuro Imaging, Center, Inserm, UPS, Université de Toulouse, Toulouse,  
43 France
- 44 21. Service de Cytogenetique, CHU de Le Mans, Le Mans, France.
- 45 22. Department of Genetics, Bretonneau University Hospital, Tours, France.
- 46 23. Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA.
- 47 24. Department of Clinical Genetics, Amsterdam UMC, Amsterdam, The Netherlands.
- 48 25. Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA.

49 \*Correspondence: [sxg47@psu.edu](mailto:sxg47@psu.edu)

## 50 **ABSTRACT**

51 We examined more than 38,000 spouse pairs from four neurodevelopmental disease cohorts and the UK  
52 Biobank to identify phenotypic and genetic patterns in parents associated with neurodevelopmental  
53 disease risk in children. We identified correlations between six phenotypes in parents and children,  
54 including correlations of clinical diagnoses such as obsessive-compulsive disorder ( $R=0.31-0.49$ ,  
55  $p<0.001$ ), and two measures of sub-clinical autism features in parents affecting several autism severity  
56 measures in children, such as bi-parental mean Social Responsiveness Scale (SRS) scores affecting  
57 proband SRS scores (regression coefficient=0.11,  $p=0.003$ ). We further describe patterns of phenotypic  
58 and genetic similarity between spouses, where spouses show both within- and cross- disorder  
59 correlations for seven neurological and psychiatric phenotypes, including a within-disorder correlation  
60 for depression ( $R=0.25-0.72$ ,  $p<0.001$ ) and a cross-disorder correlation between schizophrenia and  
61 personality disorder ( $R=0.20-0.57$ ,  $p<0.001$ ). Further, these spouses with similar phenotypes were  
62 significantly correlated for rare variant burden ( $R=0.07-0.57$ ,  $p<0.0001$ ). We propose that assortative  
63 mating on these features may drive the increases in genetic risk over generations and the appearance of  
64 “genetic anticipation” associated with many variably expressive variants. We further identified parental  
65 relatedness as a risk factor for neurodevelopmental disorders through its inverse correlations with  
66 burden and pathogenicity of rare variants and propose that parental relatedness drives disease risk by  
67 increasing genome-wide homozygosity in children ( $R=0.09-0.30$ ,  $p<0.001$ ). Our results highlight the  
68 utility of assessing parent phenotypes and genotypes in predicting features in children carrying variably  
69 expressive variants and counseling families carrying these variants.

## 70 INTRODUCTION

71 Rare genomic variants, including copy-number variants (CNVs) and single nucleotide variants (SNVs),  
72 contribute to a significant proportion of individuals with neurodevelopmental and psychiatric disorders<sup>1-</sup>  
73 <sup>7</sup>. Many of these variants are often inherited and exhibit variable expressivity and phenotypic  
74 heterogeneity, in contrast to syndromic variants with invariable clinical presentation<sup>8</sup>. For example,  
75 phenotypic heterogeneity has been observed among carriers of rare recurrent SNVs associated with  
76 neurodevelopmental disorders<sup>9</sup>, such as the candidate schizophrenia gene *NRXN1*, which also  
77 contributes to intellectual disability/developmental delay (ID/DD), autism, and congenital  
78 malformations with varying degrees of penetrance<sup>10</sup>. This variable expressivity and phenotypic  
79 heterogeneity can be attributed to a multi-hit model, where a primary variant sensitizes the genome for  
80 neuropsychiatric features and genetic interactions of the primary variant with other (“secondary”)  
81 variants in the genetic background determine the ultimate phenotypic outcome<sup>11-14</sup>. In fact, pathogenic  
82 rare CNVs on chromosome 1q21.1, 15q13.3, and 16p12.1 tend to be transmitted from a relatively mildly  
83 affected parent, with the increase in phenotypic severity corresponding to an increase in the number (i.e.  
84 burden) of secondary variants over generations<sup>11,14,15</sup>. These secondary variants are often inherited from  
85 the non-carrier parent and have been shown to act synergistically with the primary pathogenic  
86 variant<sup>14,16</sup>. This phenomenon of complex inheritance forms the basis of many neurodevelopmental  
87 disorders and complicates predicting the phenotypic trajectory of individuals within families with these  
88 disorders.

89         These complexities necessitate examining other factors, such as parental phenotype and  
90 genotype, in affected probands that may elucidate their ultimate phenotypic trajectory. Previous reports  
91 have shown that family history is a risk factor for neurodevelopmental disorders, with increased within  
92 and cross-disorder risk observed among relatives of individuals with psychiatric disease<sup>17-19</sup>. This  
93 increase in liability towards multiple diagnostic outcomes may be due to the overlapping genetic

94 etiologies of neurodevelopmental disorders<sup>20</sup> and is particularly relevant for understanding mechanisms  
95 of variable expressivity of rare variants. Here, we sought to broaden our understanding of the effect of  
96 various phenotypic and genetic factors in parents on neurodevelopmental disease outcomes in children.  
97 To that end, we examined more than 38,000 spouse pairs and parent pairs of children with  
98 neurodevelopmental disorders to identify risk factors that affect disease trajectories in affected children,  
99 including clinical and subclinical neurological and psychiatric features and patterns of assortative mating  
100 on those features, as well as genetic similarity between parents. We further investigated the implications  
101 of these patterns across generations and identified potential mechanisms by which they contribute to  
102 disease risk. Our findings highlight assortative mating and parental kinship as previously unexplored  
103 risk factors for neurodevelopmental and neuropsychiatric disease.

## 104 **SUBJECTS AND METHODS**

### 105 **Cohort Information**

106 We assessed parental factors affecting neurodevelopmental disease in two North American autism  
107 cohorts: the Simons Simplex Collection (SSC)<sup>21</sup> and the Simons Foundation Powering Autism Research  
108 (SPARK)<sup>22</sup>. We analyzed genetic and/or phenotypic data from 2,517 trios from SSC and 7,175 trios  
109 from SPARK. We also analyzed genetic and/or phenotypic data from an additional 67 parent pairs in  
110 SSC and 1,570 parent pairs in SPARK for families where the relevant proband data were not available.

111 We also collected genomic and/or phenotypic data for another neurodevelopmental disease  
112 cohort of 203 samples from 43 trios, 4 larger families, and 21 spouse pairs carrying the 16p12.1  
113 deletion. Probands in these families were ascertained for developmental delays or other  
114 neurodevelopmental phenotypes and were identified as 16p12.1 deletion carriers through clinical  
115 diagnostic tests. Families provided informed consent to give blood samples and phenotypic data under  
116 Pennsylvania State University Institutional Review Board approved protocol #STUDY00000278, and  
117 deidentified data from medical reports were collected under approved protocol #STUDY00017269.

118 We further assessed assortative mating and spousal relatedness in 26,610 partner pairs from the  
119 UK Biobank<sup>23</sup>, identified using Data Return 1872<sup>24</sup> (Application ID: 45023). UK Biobank is a database  
120 containing genetic, lifestyle, and health information from 500,000 volunteer participants from the UK.  
121 Additionally, we assessed profiles of genetic relatedness in parent pairs of children ascertained for  
122 neurodevelopmental disorders carrying rare copy number variants (CNVs) from another clinical cohort  
123 evaluated at GeneDx. This cohort consists of families with affected children referred by clinicians to  
124 GeneDx for genetic testing and is distinct from the 16p12.1 deletion cohort. Families were subsequently  
125 selected for those carrying pathogenic CNVs (see *CNV calling*), resulting in 357 parent pairs from  
126 families carrying 21 distinct disease-associated CNVs (**Supp. Table 1**). The study of this cohort was  
127 conducted under an IRB protocol approved by the Western Institutional Review Board (IRB #STUDY

128 1169768, WIRB Pro Number 20162523), which states that this research meets the requirements for a  
129 waiver of consent. We note that the terms “spouses” and “partners” are used interchangeably, while the  
130 term “parents” refers to “spouses” or “partners” in the context of affected children in this manuscript.

131

### 132 **DNA Extraction and Sequencing**

133 DNA extraction and sequencing of samples within the SPARK, SSC, and UK Biobank cohorts were  
134 performed by the relevant biobank. For directly recruited participants from the 16p12.1 deletion cohort,  
135 DNA was extracted from peripheral blood using the QIAamp DNA Blood Maxi extraction kit (Cat. No:  
136 51104, Qiagen), while clinical collaborators sent DNA samples directly to the Girirajan laboratory. SNP  
137 arrays for 36 samples were generated as previously described<sup>14</sup>, while 92 additional samples were run on  
138 Illumina OmniExpress 24 v 1.1 microarrays at the Northwest Genomics Center at the University of  
139 Washington (Seattle, WA, USA). Illumina HiSeq X was used to generate 150-bp paired-end whole  
140 genome sequence (WGS) data at  $\geq 30X$  coverage for 118 individuals by Macrogen Labs (Rockville, MD,  
141 USA). DNA extraction and sequencing from the GeneDx cohort were performed as previously  
142 described<sup>25</sup>.

143

### 144 **Variant calling**

145 *CNV calling*: CNVs in the GeneDx cohort were first called from exome sequencing data and confirmed  
146 via microarray, as previously described<sup>25</sup>. Disease-associated CNVs were identified from any overlap  
147 with previously reported breakpoints for pathogenic CNVs<sup>15</sup>.

148 *SNV calling*: Single nucleotide variants (SNVs) and small indels were identified in 16p12.1 deletion  
149 families using the GATK Best Practices pipeline. After removing duplicates with PicardTools<sup>26</sup>, we  
150 performed base-pair quality score recalibration and then used GATK HaplotypeCaller v.3.8<sup>27</sup> to call  
151 variants. Calls from all samples were merged to call variants jointly using GATK GenotypeGVCFs

152 v.4.0.11<sup>27</sup>, followed by variant quality score recalibration. Variants were filtered for those with an allele  
153 balance between 0.25 and 0.75 or >0.9 and a read depth of  $\geq 8$ .

154 Whole genome (for SSC) and whole exome (for SPARK) variant call files (VCFs) were  
155 downloaded from SFARI Base (<https://base.sfari.org>). Whole exome VCFs were obtained from the UK  
156 Biobank using Data Field 23157. VCFs from all three cohorts were left-normalized and multi-allelic  
157 records were split with BCFtools<sup>28</sup>. Calls were filtered using the same criteria as above. Variants from  
158 the 16p12.1 deletion cohort, SSC, and SPARK were additionally filtered for a quality  $\geq 50$  and a ratio of  
159 quality to alternative allele depth  $\geq 1.5$ .

160 For all four cohorts, variant frequency in gnomAD<sup>29</sup> was annotated using vcfanno<sup>30</sup> and variants  
161 were filtered for those with gnomAD exome and genome frequency  $\leq 0.1\%$ . Variants in the SSC,  
162 SPARK, and UK Biobank cohorts were filtered for an intracohort frequency  $\leq 0.1\%$ . To achieve a  
163 similar frequency threshold with a smaller sample size, variants in the 16p12.1 deletion cohort were  
164 filtered for those that were private within related individuals. ANNOVAR<sup>31</sup> was used to assign variants  
165 to genes from GENCODE v19 (16p12.1 deletion cohort) or GENCODE v38 (SSC, SPARK, and UK  
166 Biobank)<sup>32</sup>, and variants in all cohorts were filtered for loss-of-function (LOF), missense, or splice  
167 variants in protein-coding genes. LOF variants were those annotated as 'stopgain', 'stoploss',  
168 'frameshift\_deletion', 'frameshift\_insertion', or 'frameshift\_substitution' by ANNOVAR, and missense  
169 variants were those annotated as 'nonsynonymous\_SNV', 'nonframeshift\_deletion',  
170 'nonframeshift\_insertion', or 'nonframeshift\_substitution'. CADD Phred-like scores<sup>33</sup> were annotated  
171 using vcfanno as a deleteriousness metric and missense and splice variants were filtered for those with a  
172 CADD Phred score  $\geq 25$ . For probands within SSC, variants were additionally annotated with SFARI  
173 Gene Tier of the affected gene<sup>34</sup>.

174 *Polygenic risk score calculations:* We obtained autism GWAS summary statistics from Grove and  
175 colleagues<sup>35</sup>. Duplicate and ambiguous single nucleotide polymorphisms (SNPs) were removed, and the  
176 remaining SNPs were filtered for imputation INFO scores  $> 0.8$ . We then used PLINK<sup>36</sup> to perform an



177 initial round of quality control by filtering SNPs with minor allele frequency  $<0.05$ , Hardy-Weinberg  
178 equilibrium test p-value  $<1.0 \times 10^{-5}$ , and genotype rate  $<0.05$ , and removed samples missing  $>1\%$  of  
179 genotypes or having a Mendelian error rate  $>5\%$ . We then used the HRC-1000 Genomes Imputation  
180 toolkit (<https://www.well.ox.ac.uk/~wrayner/tools>) to process PLINK files into individual chromosomes  
181 for imputation, and used VcfCooker (<https://genome.sph.umich.edu/wiki/VcfCooker>) to convert PLINK  
182 files to VCF files. SNPs were imputed using the TOPMed v.r2 imputation server using Eagle v2.4 for  
183 phasing<sup>37</sup>. After imputation, VCF files were converted back to PLINK format, and SNPs were again  
184 filtered using the criteria from the initial quality control. SNPs were additionally filtered to remove  
185 samples with  $\geq 3$ SD of the mean heterozygosity and strand-flipped to match the GWAS summary  
186 statistics. We filtered the GWAS summary statistics for SNPs present in the HapMap3 dataset<sup>38</sup>, and  
187 used 1000 Genomes datasets to calculate linkage disequilibrium matrices for the SNPs. After regressing  
188 betas of GWAS SNPs according to linkage disequilibrium, we used the LDpred2-auto model<sup>39</sup> to  
189 calculate autism PRS for probands in the SSC cohort.

190

## 191 **Phenotype Assessment**

192 We assessed self-reported psychiatric phenotypes in probands and parents in the SPARK cohort and 23  
193 quantitative autism and cognitive metrics in probands (**Supp. Table 2**) and two quantitative autism  
194 metrics in parents (Social Responsiveness Scale (SRS)<sup>40</sup> and Broad Autism Phenotype Questionnaire  
195 (BAPQ)<sup>41</sup>) from the SSC cohort downloaded from SFARI base (<https://base.sfari.org>). We further  
196 assessed neurological and psychiatric phenotypes in 57 parent pairs from 16p12.1 deletion families.  
197 Information was collected on each adult family member either through self-reports using a standardized  
198 clinical questionnaire or clinician reports. Curated data was generated using a union of clinical  
199 diagnoses, self-reported phenotypes, and interpreted responses to clinical questionnaires. We also  
200 assessed phenotypic correlations in the UK Biobank using self-reported diagnoses of psychiatric  
201 disorders from Data Field 20544.

202 Linear regression was used to assess the predictive power of parental SRS and BAPQ scores and  
203 various genetic factors, including autism PRS and rare CNVs and SNVs, towards child quantitative  
204 phenotypes, while tetrachoric correlations were used to assess phenotypic correlations between  
205 psychiatric diagnoses in parents and children and between spouse pairs. To minimize the effects of  
206 comorbidities and isolate the specific correlations of interest, pairs were removed from cross-disorder  
207 tetrachoric correlations if both members of the pair had either diagnosis of interest. For example, when  
208 examining the relationship between the female member of the spouse pair having anxiety and their  
209 partner having depression, pairs where the male partner had anxiety or the female partner had depression  
210 were excluded from the analysis.

211

## 212 **Kinship Calculations**

213 In the 16p12.1 deletion, SSC, SPARK, and UK Biobank cohorts, we followed the same initial SNP  
214 quality control steps outlined above for PRS calculations. Additionally, SNPs within long-range linkage  
215 disequilibrium regions<sup>42</sup> were removed, as were samples with  $\pm 2$  standard deviations of the mean cohort  
216 heterozygosity rate. For SPARK, SSC, and the 16p12.1 cohort samples, we next assessed ancestry using  
217 fastStructure<sup>43</sup> to cluster samples (k=5-6) against individuals with known ancestry from the HapMap3  
218 project<sup>38</sup>. Kinship analysis in these cohorts was restricted to pairs with self-reported or calculated  
219 European ancestry. UK Biobank samples were filtered for pairs of European ancestry using Data Field  
220 22006.

221 KING<sup>44</sup> was used to calculate kinship coefficients for 44 parent pairs from the 16p12.1 cohort,  
222 1,996 parent pairs from the SSC cohort, 4,211 parent pairs from the SPARK cohort, 6,315 partner pairs  
223 from the UK Biobank cohort, and 357 parent pairs from the GeneDx cohort. In the GeneDx cohort, 11  
224 parent pairs with a kinship coefficient less than -0.1 were removed from further analysis, as this highly  
225 negative value may indicate that these samples come from different ancestral populations.

226

## 227 **Runs of Homozygosity**

228 To identify runs of homozygosity (ROH), we applied the same initial SNP quality control filters  
229 described above for PRS calculations on 3,714 probands ascertained for autism in the SPARK cohort  
230 and 1,981 probands from the SSC cohort. ROHs were calculated using the PLINK *homozyg* function<sup>36</sup>.  
231 We allowed a maximum of one heterozygous and five missing SNPs in each ROH, and runs were  
232 restricted to those at least 1 Mbp in length.

233

## 234 **Statistical Analysis**

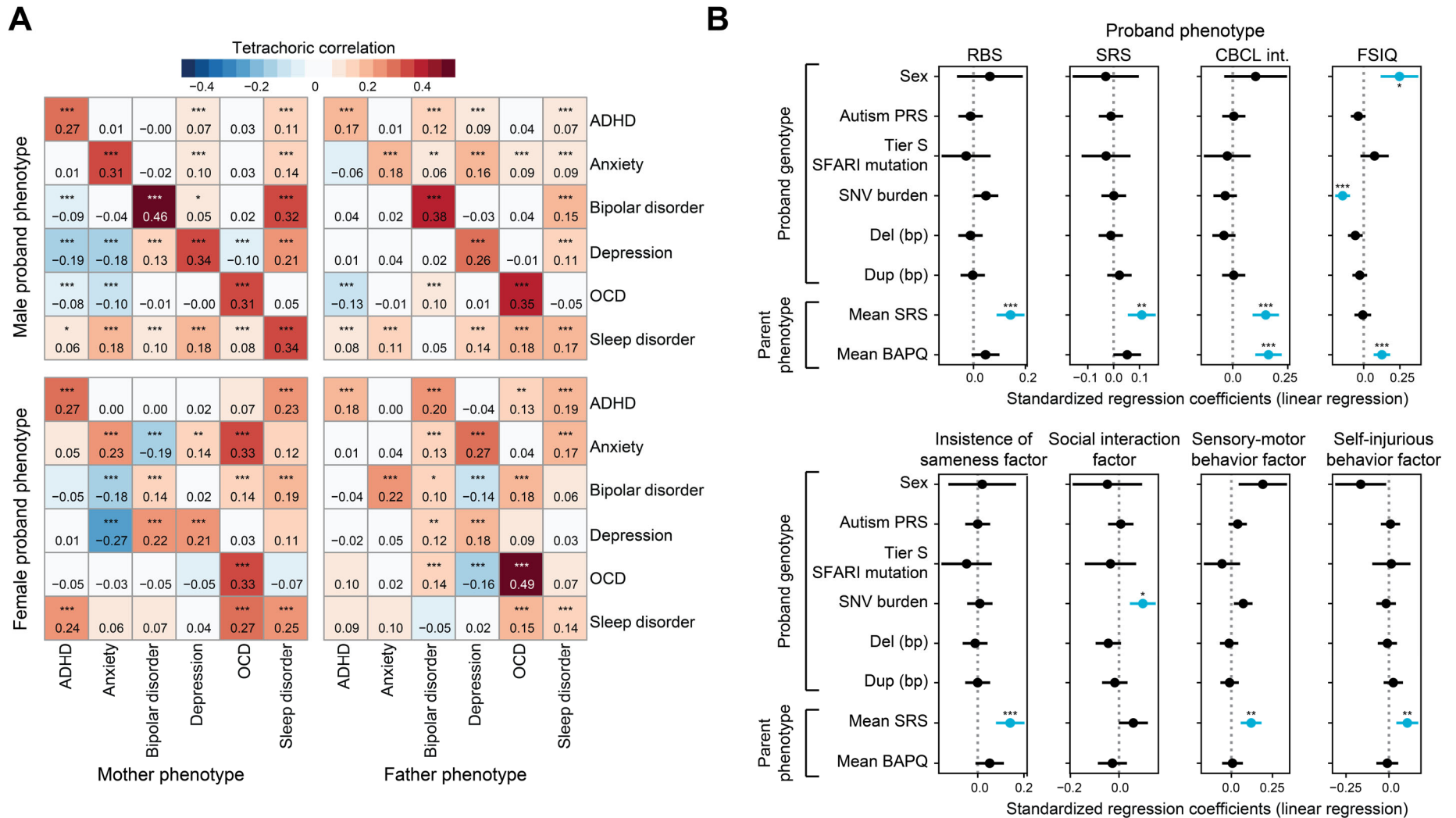
235 Tetrachoric correlations were performed using the *correlation* package<sup>45</sup> in R. Linear regression was  
236 performed using *statsmodels* in Python, with proband age at the time the Autism Diagnostic Observation  
237 Schedule (ADOS) was administered included as a covariate in the models. Pearson correlations and t-  
238 tests were performed using the *pearsonr* and *ttest\_ind* functions in the *scipy stats* (v. 1.7.1) package<sup>46</sup> in  
239 Python (v. 3.9.7), respectively.

## 240 RESULTS

### 241 Parental phenotypes predict child phenotypes

242 To understand the relevance of parental phenotypes in neurodevelopmental disorders, we examined the  
243 correlation between six parent and child psychiatric phenotypes in the Simons SPARK autism cohort  
244 (**Fig. 1A, Table S4A**). We observed significant positive within-disorder correlations between mothers  
245 and fathers and male and female probands for most phenotypes (Pearson correlation,  $R=0.1-0.49$ ,  
246 Bonferroni-corrected  $p<0.05$ ) (**Fig. 1A, Table S4A**). We also observed several cross-disorder  
247 correlations, including those common to both sexes, such as the correlation between depression in  
248 parents and anxiety in probands ( $R=0.10-0.27$ ,  $p<0.01$ ), and those which are unique to one sex, such as  
249 the correlation between OCD in fathers and ADHD in female probands ( $R=0.13$ ,  $p=0.003$ ), suggesting  
250 parental and sex-specific correlations for certain pairs of features (**Fig. 1A, Table S4A**). These  
251 correlations recapitulate previous patterns of within- and cross- disorder correlations observed in studies  
252 of both psychiatric disease cohorts and cohorts from the general population<sup>17,47,48</sup>.

253 To assess the relationship of parent and child phenotypes outside the context of a family history  
254 of psychiatric disorders, we used linear models to assess the relationship between 23 quantitative  
255 measures of autism and cognitive features (**Table S2**) in probands from the simplex autism cohort SSC  
256 and two measures of autism features in their unaffected parents, Social Responsiveness Scale (SRS)<sup>40</sup>  
257 and Broad Autism Phenotype Questionnaire (BAPQ)<sup>41</sup> scores (**Fig. 1B, Table S4B**). Several measures  
258 of autism features in probands were significantly influenced by parental SRS scores, including  
259 Repetitive Behaviors Scale (RBS)<sup>49</sup> ( $\beta=0.14$ , Bonferroni-corrected  $p=1.95\times 10^{-6}$ ), SRS ( $\beta=0.11$ ,  
260  $p=0.002$ ), and Child Behavior Checklist (CBCL)<sup>50</sup> internalizing ( $\beta=0.15$ ,  $p=1.11\times 10^{-5}$ ) scores of  
261 children with autism were positively associated with biparental mean SRS scores (**Fig. 1B, Table S4B**).  
262 Biparental mean SRS and/or BAPQ scores were in fact stronger predictors of autism severity in affected  
263 children than any individual group of rare or common genetic variants for several measures (**Fig. 1B,**



**Figure 1. Parental phenotypes predict clinical outcomes in children.**

(A) Tetrachoric correlations between self-reported psychiatric disorders in mothers and fathers and male and female probands within the SPARK autism cohort ( $n=592-4,757$  for each correlation). (B) Regression coefficients from linear models incorporating effects of both genetic factors and parental phenotypes towards quantitative proband autism and cognitive phenotypes ( $n=1,628-2,168$  for each model). RBS: Repetitive Behaviors Scale, SRS: Social Responsiveness Scale, CBCL int: Child Behavior Checklist 6-18 internalizing, FSIQ: Full-scale IQ. Selected models shown here, full data in Table S4. Blue dots and lines indicate significance after Bonferroni-correction.

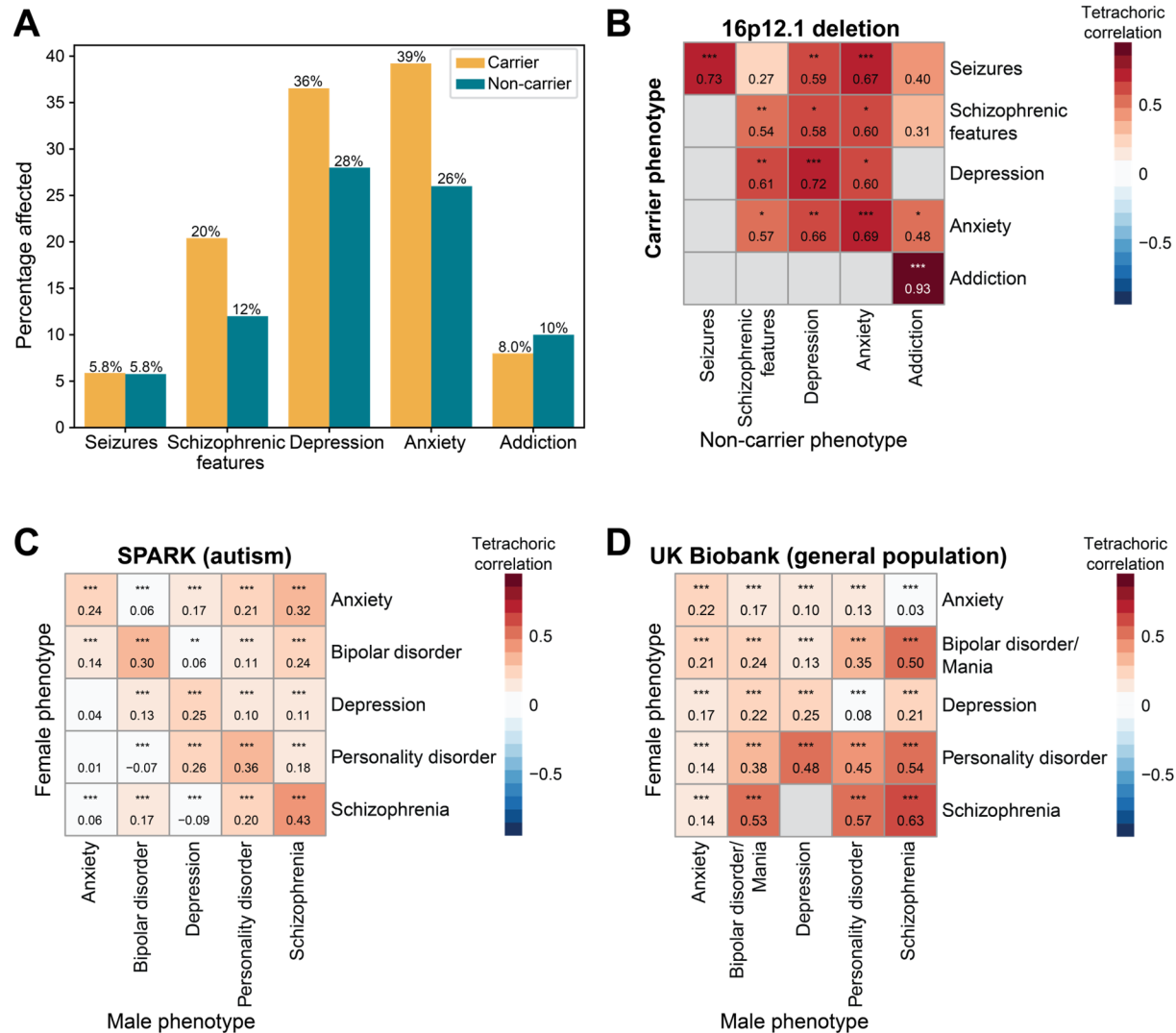
\* Bonferroni-corrected  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ .

264 **Table S4B**). These data highlight the potential utility of assessing parent phenotypes to predict child  
265 phenotypes. Even in simplex families, where parents are largely unaffected, child phenotypes can be  
266 predicted by parent phenotypes, suggesting some sub-clinical level of genetic liability in the parents of  
267 affected children which compounds within a single generation to surpass a disease threshold. This also  
268 indicates the presence of multiple genetic factors which individually are not sufficient to cause overt  
269 phenotypes in parents but may additively or synergistically interact in offspring to cause more overt  
270 phenotypes in subsequent generations.

271

### 272 **Phenotypic assortment among partner pairs**

273 Given the relevance of parental phenotypes towards predicting phenotypes in affected children, we  
274 hypothesized that assortative mating may be a relevant risk factor for neurodevelopmental disease.  
275 Assortative mating has been well described for neurological and psychiatric phenotypes<sup>24,51–58</sup> and has  
276 previously been shown to result in increasing disease liability over generations using national registrar  
277 data from Sweden and Denmark<sup>17,47,59</sup>. This is akin to patterns of increasing disease liability observed in  
278 neurodevelopmental disease cohorts, particularly in families carrying variably expressive variants<sup>11,14,15</sup>.  
279 Using the 16p12.1 deletion as a paradigm for these disorders, we first examined the frequency of self-  
280 reported neurological and psychiatric features in adult family members of children with the 16p12.1  
281 deletion ascertained for neurodevelopmental phenotypes. Adult deletion carriers in our cohort  
282 manifested several neurological and psychiatric phenotypes, including seizures (5.8%), schizophrenic  
283 features (20%), depression (36%) and anxiety (39%) symptoms, and addiction phenotypes (8.0%) (**Fig.**  
284 **2A**). In line with expectations of assortative mating, their non-carrier partners also manifested these  
285 features at similar frequencies (seizure: 5.8%, schizophrenic features: 12%, depression: 28%, anxiety:  
286 26%, addiction: 10%) (**Fig. 2A**). As these data are indicative of assortative mating, we further assessed  
287 spousal assortment based on these phenotypes and were able to identify several significant within- and  
288 cross-disorder correlations (**Fig. 2B, Table S4C**). For example, 16p12.1 deletion carriers with anxiety



**Figure 2. Patterns of assortative mating in multiple cohorts.**

(A) Percentage of carrier and non-carrier spouses in a cohort of 16p12.1 deletion carrier families with five neurological and psychiatric disorders (n=49-52). (B-D) Tetrachoric correlations of neurological and psychiatric phenotypes between (B) carrier and non-carrier spouses in the 16p12.1 deletion cohort (n=28-50), (C) mothers and fathers in the SPARK autism cohort (n=3,956-6,142), and (D) female and male spousal partners in the UK Biobank (n=19,468-22,953). Grey boxes indicate insufficient sample size for correlation. \* Bonferroni-corrected  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ .

289 were likely to have partners with anxiety ( $R=0.69$ , Bonferroni-corrected  $p=8.14\times 10^{-7}$ ) or depression  
290 ( $R=0.66$ ,  $p=0.001$ ) (**Fig. 2B, Table S4C**). These data confirm that the well-documented patterns of  
291 assortative mating for psychiatric phenotypes are typically prevalent in families carrying variably  
292 expressive variants with a high load of genetic liability for neurodevelopmental disorders. We also  
293 compared patterns of phenotypic assortment in parents of children in an affected cohort and the general  
294 population by analyzing five self-reported psychiatric phenotypes in the SPARK autism cohort and the  
295 UK Biobank (**Fig. 2C-D, Table S4D**). Although these phenotypes occur much more frequently in the  
296 SPARK cohort (**Supp. Table 3**), positive within- and cross- disorder correlations were observed in both  
297 of these cohorts (**Fig. 2C-D, Table S4D**). For example, women in both cohorts with bipolar disorder or  
298 mania were more likely to have partners with bipolar disorder (SPARK:  $R=0.30$ , Bonferroni-corrected  
299  $p=7.69\times 10^{-129}$ ; UKB:  $R=0.24$ ,  $p=3.53\times 10^{-304}$ ) or schizophrenia (SPARK:  $R=0.24$ ,  $p=3.90\times 10^{-80}$ ; UKB:  
300  $R=0.50$ ,  $p=0$ ) (**Fig 2C-D**). This confirms that spousal assortment on psychiatric phenotypes occurs both  
301 in individuals ascertained for disease and those from the general population.

302

### 303 **Genetic consequences of assortative mating**

304 Partner assortment will result in positive correlations between their risk alleles, which underly the  
305 phenotypes driving assortment<sup>60</sup>. While genetic risk of assortative mating is often measured using  
306 association studies of common variants<sup>24,60-62</sup>, we hypothesized that rare deleterious variants also  
307 correlate among partners, given the links between rare variants and psychiatric disorders on which  
308 partners tend to assort<sup>7,63,64</sup>. To confirm this pattern in the assessed cohorts, we correlated the burden of  
309 rare deleterious coding single nucleotide variants (SNVs) between partners in each cohort (**Fig. 3A**). We  
310 observed that the burden of rare variants positively correlated between partners in the 16p12.1 deletion  
311 cohort (Pearson's  $R=0.46$ ,  $p=2.79\times 10^{-4}$ ), SPARK ( $R=0.52$ ,  $p<10^{-12}$ ), Simons Simplex Collection (SSC,  
312  $R=0.57$ ,  $p<10^{-12}$ ), and the UK Biobank ( $R=0.07$ ,  $p=6.80\times 10^{-9}$ ) cohorts (**Fig. 3A**), confirming that partner



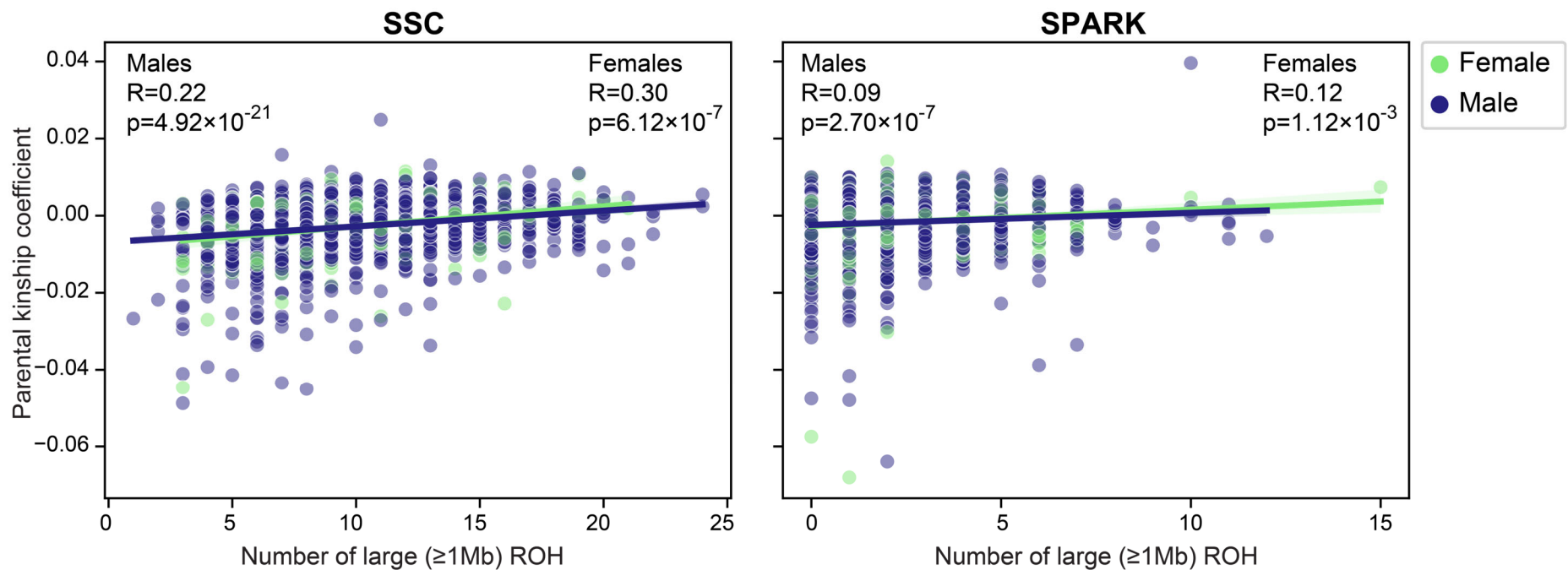


313 assortment based on related phenotypes drives correlations in genetic risk. Interestingly, the correlation  
314 was much weaker in the UK Biobank ( $R=0.07$ ) than the disease cohorts ( $R\geq 0.46$ ), possibly due to the  
315 lower frequency of psychiatric phenotypes (**Supp. Table 3**) and lower burden of genetic risk in this  
316 cohort. The ultimate outcome of this correlation in burden is an increase in the variance of genetic  
317 liability over generations<sup>65</sup>, where affected individuals will have more affected children and unaffected  
318 or mildly affected individuals will have more mildly affected children (**Fig. 3B**), mimicking patterns  
319 observed in families carrying variably expressive variants. For example, 16p12.1 deletion carrier  
320 families typically show increases in clinical severity over generations, from unaffected or mildly  
321 affected grandparents to parents with psychiatric features and then to children with neurodevelopmental  
322 disorders (**Fig. 3C**) and this increase in phenotypic severity corresponds to an increase in genetic  
323 burden<sup>11,14,16</sup> (**Supp. Figure 1**). Taken together, this suggests that assortative mating drives the increases  
324 in genetic liability that result in the severe phenotypes observed in neurodevelopmental disease cohorts,  
325 particularly in carriers of variably expressive variants.

326

### 327 **Parental relatedness as a risk factor for neurodevelopmental disease**

328 Another parental factor which can increase disease risk is the level of consanguinity or genetic  
329 relatedness of parents, through increasing both the frequency of autosomal recessive conditions and  
330 genome-wide homozygosity<sup>66,67</sup>. Long stretches of homozygosity have been associated with increased  
331 risk for several neurodevelopmental disorders<sup>68,69</sup>. Although the parents assessed in our cohorts are not  
332 consanguineous, we hypothesized that increases in parental relatedness lead to increases in genome-  
333 wide homozygosity and increased risk of neurodevelopmental disorders. To assess this, we examined  
334 large ( $\geq 1$  Mbp) runs of homozygosity (ROH) in 3,714 and 1,981 children of the parent pairs analyzed in  
335 the SPARK and SSC cohorts, respectively. We found a significant correlation between parental  
336 relatedness (as measured by kinship coefficients) and the number of large ROH in both male and female

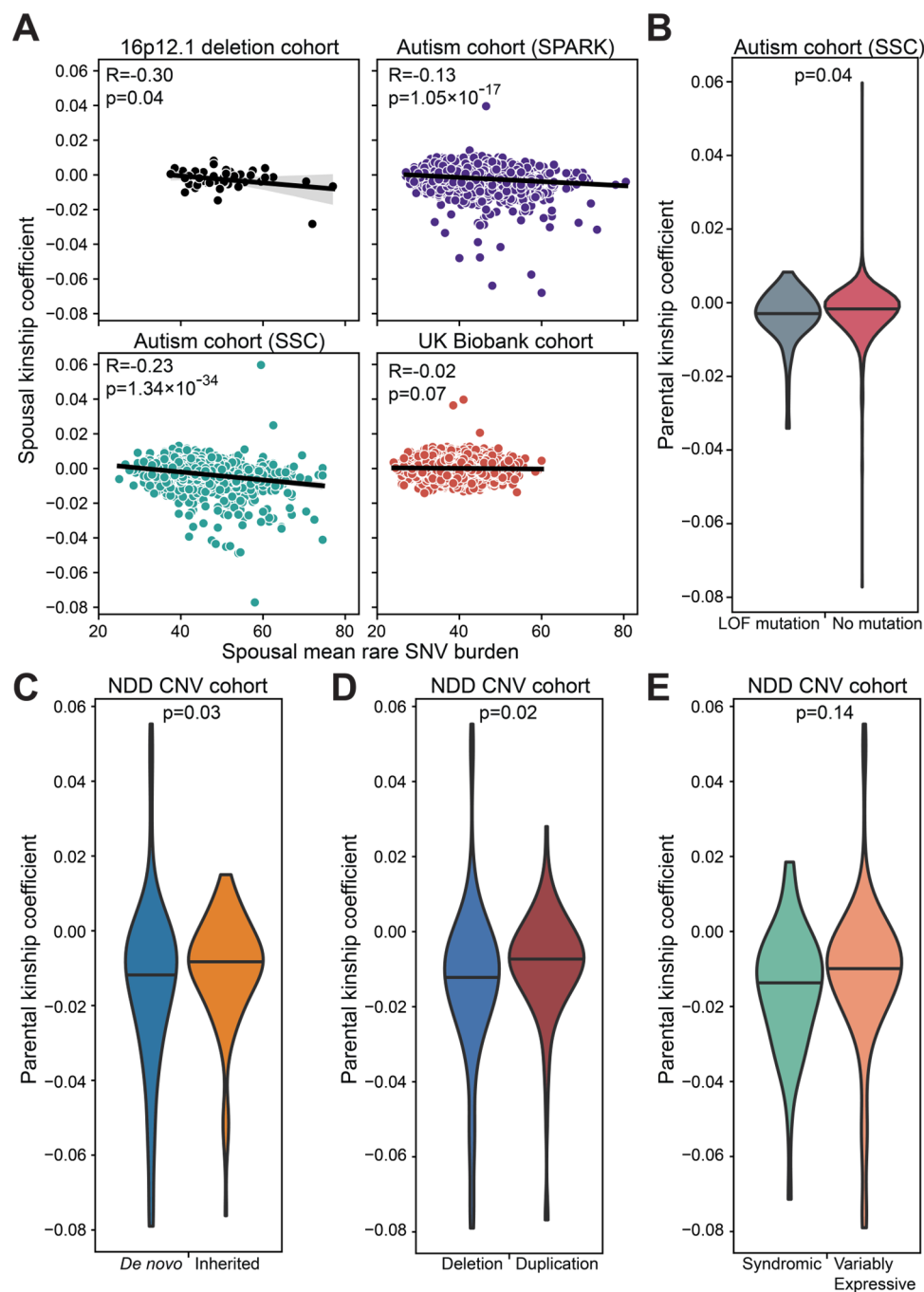


**Figure 4. Parental relatedness correlates with homozygosity in children.**

Pearson correlations of parental relatedness (measured as kinship coefficients) and the number of large (≥1 Mb) runs of homozygosity in male and female probands in the SSC (male n=1,711, female n=270) and SPARK (male n=2,956, female n=758) autism cohorts.

337 probands from both SSC (male  $R=0.22$ ,  $p=4.92\times 10^{-21}$ , female  $R=0.30$ ,  $p=6.12\times 10^{-7}$ ) and SPARK (male  
338  $R=0.09$ ,  $p=2.70\times 10^{-7}$ , female  $R=0.12$ ,  $p=1.12\times 10^{-3}$ ) (**Fig. 4**), indicating that increased spousal  
339 relatedness could lead to increased disease risk by unmasking recessive alleles within those intervals.

340 We then asked how this parental relatedness could interact with other neurodevelopmental  
341 disease factors. We calculated kinship coefficients as a measure of relatedness between partner pairs  
342 from all four cohorts and found negative correlations between spousal kinship and the average spousal  
343 rare variant burden in all three neurodevelopmental disease cohorts (16p12.1 deletion:  $R=-0.30$ ,  $p=0.04$ ;  
344 SPARK:  $R=-0.13$ ,  $p=1.05\times 10^{-17}$ , SSC:  $R=-0.23$ ,  $p=1.34\times 10^{-23}$ ) (**Fig. 5A**). However, no significant  
345 correlation was observed for partners in the UK Biobank ( $R=-0.02$ ,  $p=0.07$ ) (**Fig. 5A**). We then  
346 examined parental relatedness in the context of severe damaging mutations in children ascertained for  
347 autism in SSC. We found that kinship coefficients of 102 parent pairs of children with loss-of-function  
348 mutations in canonical autism-associated genes (Tier S SFARI genes)<sup>34</sup> were significantly lower than  
349 kinship coefficients in 1,251 parent pairs of affected children without damaging mutations in autism-  
350 associated genes (one-tailed t-test  $p=0.041$ ) (**Fig. 5B**). We also examined kinship coefficients of 346  
351 parent pairs from an independent cohort of children ascertained for neurodevelopmental disorders  
352 carrying 21 different pathogenic CNVs (**Supp. Table 2**). We found that parental kinship had an inverse  
353 relationship with severity as well as penetrance of the primary variant in children. Kinship coefficients  
354 were significantly higher among parents of children with inherited CNVs compared to parents of  
355 children with *de novo* CNVs (one-tailed t-test,  $p=0.03$ ) (**Fig. 5C**). Similarly, parents of children with  
356 duplications, which are typically associated with incomplete penetrance<sup>70</sup> and less severe features<sup>71</sup> than  
357 deletions, had higher kinship coefficients compared to parents of children with deletions (one-tailed t-  
358 test,  $p=0.02$ ) (**Fig. 5D**). Although not statistically significant, we also identified a trend of higher kinship  
359 values in families with variably expressive CNVs compared to syndromic CNVs (one-tailed t-test,  
360  $p=0.14$ ) (**Fig. 5E**).



**Figure 5. Parental relatedness is a risk factor for neurodevelopmental disorders.**

(A) Pearson correlations of the average burden of rare, likely deleterious variants and relatedness (measured as kinship coefficients) in spouse pairs from the 16p12.1 deletion ( $n=44$ ), SPARK ( $n=4,095$ ) and SSC ( $n=1,813$ ) autism, and UK Biobank ( $n=6,315$ ) cohorts. (B) Comparison of parental relatedness among SSC probands with loss-of-function mutations in Tier S SFARI genes ( $n=102$ ) and without mutations in SFARI genes ( $n=1,243$ ). Lines indicate median. One tailed t-test. (C-E) Comparisons of parental relatedness among probands ascertained for neurodevelopmental disease with (C) *de novo* ( $n=129$ ) and inherited ( $n=116$ ) CNVs, (D) deletions ( $n=249$ ) and duplications ( $n=97$ ), and (E) syndromic ( $n=47$ ) and variably expressive ( $n=299$ ) CNVs. Lines indicate median. One tailed t-test.

361           These patterns are in accordance with a liability threshold model (LTM), which posits that an  
362 individual will manifest disease once the sum of various independent risk factors exceeds a certain  
363 threshold<sup>72</sup>. An inverse relationship between all independent risk factors contributing to disease is  
364 implied by the LTM, and has been observed previously when examining additive effects of rare and  
365 common variants towards neurodevelopmental disease<sup>1,73,74</sup>. Thus, the negative relationship between  
366 parental kinship and rare SNV burden or presence of highly pathogenic mutations supports parental  
367 relatedness as a risk factor for variably expressive pathogenic variants, likely through increased levels of  
368 genome-wide homozygosity. This is a particularly compelling hypothesis considering the lack of  
369 correlation between spousal kinship and mean spousal rare variant burden seen in the UK Biobank pairs  
370 (**Fig. 5A**). Unlike the other cohorts, the pairs in the UK Biobank were not ascertained for children with  
371 neurodevelopmental disease and thus may not exist within the same range of liability threshold as the  
372 other cohorts.

## 373 **DISCUSSION**

374 Assortative mating has previously been hypothesized to explain the high degree of heritability for  
375 complex neurodevelopmental and psychiatric disorders<sup>54,75</sup>. However, the implications of assortative  
376 mating have not been explored in the context of rare variably expressive pathogenic variants. Here, we  
377 confirm that the well-documented patterns of spousal assortment on neuropsychiatric features<sup>24,51–58</sup>  
378 hold true for families carrying rare disease-associated variants. It is possible that the observed  
379 phenotypic similarity in partners may be caused by mechanisms outside of assortative mating, including  
380 partner convergence<sup>76</sup> or, in disease cohorts, caregiver burden on parents of affected children<sup>77</sup>.  
381 However, we find it unlikely that these mechanisms could account for all observed partner similarity for  
382 three key reasons. Firstly, previous studies have indicated that convergence has only a modest effect on  
383 spousal similarity<sup>76,78</sup>. Further, of the 27 participants in the 16p12.1 deletion cohort who indicated the  
384 age at depression symptom onset, 70% (19/27) began having symptoms before the birth of their first  
385 child. Thus, caregiver burden is an unlikely cause of depression in these individuals and, although these  
386 data are limited, we expect similar trends for other disorders. Finally, if spousal phenotypic similarity  
387 were the result of convergence or caregiver burden, we would not expect to see a correlation in genetic  
388 risk between partners. However, as we do see a strong correlation in rare variant burden, we can  
389 conclude that active assortment accounts for much of the observed phenotypic similarity between  
390 partners.

391 Partner assortment on these phenotypes drives increases in genetic risk alleles across generations  
392 in affected individuals<sup>75</sup>, similar to the observed patterns of increasing secondary variant burden that are  
393 well documented in cohorts of variably expressive CNVs<sup>11,14</sup> and are concordant with a “two-hit”  
394 mechanism of pathogenicity. Under assortative mating, individuals with similar psychiatric phenotypes  
395 but distinct underlying genetic etiologies may preferentially mate and have a child who carries genetic  
396 risks from both parents, leading to a compounding of genetic burden over generations. Increased disease

397 liability in children resulting from assortative mating has been previously shown<sup>17,59</sup>, and the  
398 implications for this finding in the context of variably expressive rare variants has been previously  
399 hypothesized<sup>79</sup>. Here, our analysis provides more evidence to support the relationship between parental  
400 assortment on psychiatric phenotypes and compounding genetic burden towards disease severity.  
401 Unique combinations of rare variants arising from recent mutations have been identified as a major  
402 contributor to disease<sup>80,81</sup> and assortative mating may drive accumulation of those new combinations.  
403 This observation has important implications on the trajectory of individuals within a family carrying a  
404 variably expressive variant and provides insights into the genetic “anticipation” observed in these  
405 families.

406 We further highlight the role of spousal relatedness towards complex disorders, showing that  
407 spousal kinship acts as a risk factor for variably expressive variants. Long stretches of homozygosity  
408 have previously been implicated in neurodevelopmental disorders<sup>68,69</sup> and we show here that parental  
409 relatedness increases genome-wide homozygosity and, likely through this mechanism, acts as a risk  
410 factor for neurodevelopmental disease. Increases in spousal kinship are often due to population  
411 stratification<sup>76</sup>, which has heretofore been an underexplored risk factor for neurodevelopmental  
412 disorders and may represent an exciting new avenue for further research.

413 While these results have interesting implications for our understanding of disease pathogenicity,  
414 our study has some limitations. We restricted our analysis of phenotypic assortment to only a subset of  
415 distinct phenotypes, and while we attempted to control for comorbidities that could inflate correlation  
416 estimates, it is likely that some cross-disorder contamination remains. Further, all cohorts assessed here  
417 were primarily of European descent and further research is needed to understand how assortment and  
418 partner similarity may vary across racial and ethnic groups, particularly for disease-associated rare  
419 variants. This may be particularly relevant in the context of isolated populations or those practicing  
420 endogamy, who, due to their unique recent genetic history, may have a higher load of both parental



421 relatedness and rare variant burden<sup>82,83</sup>. Further research is required to examine the genetic architecture  
422 in these populations and how these changes in architecture may impact disease risk.

423 In summary, we provide evidence for assortative mating on neuropsychiatric disorders,  
424 particularly in parents of children with neurodevelopmental phenotypes. We show patterns of rare  
425 variant burden similarities between spouses and highlight compounding of rare variant burden over  
426 generations resulting from phenotypic assortment. We also highlight increased spousal kinship as a risk  
427 factor for neurodevelopmental disorders through the unmasking of recessive alleles. Overall, our results  
428 highlight the importance of assessing parental history of neuropsychiatric disease and parental  
429 relatedness for more accurate assessment of neuropsychiatric disease risk in children and counseling of  
430 families carrying variably expressive variants.

431

## 432 **ACKNOWLEDGEMENTS**

433 This work was supported by grants T32GM102057 to Miss Corrine Smolen and NS122398 and  
434 GM121907 to Dr. Santhosh Girirajan from the National Institutes of Health. Dr. Charles Schwartz was  
435 supported, in part, by a grant from the South Carolina Department of Disabilities and Special Needs. Dr.  
436 Corrado Romano was supported by the Italian Ministry of Health-Ricerca Corrente 2022.

437

438 We are grateful to the families in the 16p12.1 deletion cohort who participated in this study. We are  
439 grateful to all the families in SPARK, the SPARK clinical sites and SPARK staff. We are also grateful  
440 to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal  
441 investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-  
442 Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles,  
443 O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z.  
444 Warren, E. Wijsman). We appreciate gaining access to genetic and phenotypic data on SFARI Base.

445 Approved researchers can obtain the SSC and SPARK datasets described in this study by applying at  
446 <https://base.sfari.org>. This research has been conducted using data from UK Biobank, a major  
447 biomedical database. More information about UK Biobank is available at <https://www.ukbiobank.ac.uk/>.

448

## 449 **DECLARATIONS OF INTERESTS**

450 Dr. Santhosh Girirajan reports grants from the National Institutes of Health outside the submitted work.  
451 Drs. Lisa Dyer and Jane Juusola are employees of GeneDx, LLC. All other authors report no financial  
452 relationships with commercial interests.

453

## 454 **DATA AND CODE AVAILABILITY**

455 Microarray and genome sequencing data from participants in the 16p12.1 deletion cohort who consented  
456 to having their information shared anonymously will be made available at NCBI dbGaP study accession  
457 phs00245. Phenotype and genetic data from the SPARK and SSC autism and UK Biobank cohorts  
458 through the cohorts' data portals: SFARI base (<https://base.sfari.org>) and UK Biobank database  
459 (<https://www.ukbiobank.ac.uk>), respectively. All code for analysis will be available at GitHub  
460 ([https://github.com/girirajanlab/Assortative\\_mating\\_project](https://github.com/girirajanlab/Assortative_mating_project)).

461 **REFERENCES**

- 462 1. Antaki, D. *et al.* A phenotypic spectrum of autism is attributable to the combined effects of rare  
463 variants, polygenic risk and sex. *Nat Genet* (2022) doi:10.1038/s41588-022-01064-5.
- 464 2. Havdahl, A. *et al.* Genetic contributions to autism spectrum disorder. *Psychol Med* **51**, 2260–2273  
465 (2021).
- 466 3. Karayiorgou, M. *et al.* Schizophrenia susceptibility associated with interstitial deletions of  
467 chromosome 22q11. *Proc Natl Acad Sci U S A* **92**, 7612–7616 (1995).
- 468 4. Kvarnung, M. & Nordgren, A. Intellectual Disability & Rare Disorders: A Diagnostic Challenge.  
469 *Adv Exp Med Biol* **1031**, 39–54 (2017).
- 470 5. Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide  
471 study of 41,321 subjects. *Nat Genet* **49**, 27–35 (2017).
- 472 6. Rein, B. & Yan, Z. 16p11.2 Copy Number Variations and Neurodevelopmental Disorders. *Trends*  
473 *Neurosci* **43**, 886–901 (2020).
- 474 7. Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*  
475 **604**, 509–516 (2022).
- 476 8. Jensen, M. & Girirajan, S. An interaction-based model for neuropsychiatric features of copy-number  
477 variants. *PLoS Genet* **15**, e1007879 (2019).
- 478 9. Moreno-De-Luca, A. *et al.* Developmental brain dysfunction: revival and expansion of old concepts  
479 based on new genetic evidence. *Lancet Neurol* **12**, 406–414 (2013).
- 480 10. Schaaf, C. P. *et al.* Phenotypic spectrum and genotype-phenotype correlations of NRXN1 exon  
481 deletions. *Eur J Hum Genet* **20**, 1240–1247 (2012).
- 482 11. Girirajan, S. *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe  
483 developmental delay. *Nat Genet* **42**, 203–209 (2010).
- 484 12. Girirajan, S. & Eichler, E. E. Phenotypic variability and genetic susceptibility to genomic disorders.  
485 *Hum Mol Genet* **19**, R176-187 (2010).

- 486 13. Leblond, C. S. *et al.* Genetic and functional analyses of SHANK2 mutations suggest a multiple hit  
487 model of autism spectrum disorders. *PLoS Genet* **8**, e1002521 (2012).
- 488 14. Pizzo, L. *et al.* Rare variants in the genetic background modulate cognitive and developmental  
489 phenotypes in individuals carrying disease-associated variants. *Genet Med* **21**, 816–825 (2019).
- 490 15. Girirajan, S. *et al.* Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N*  
491 *Engl J Med* **367**, 1321–1331 (2012).
- 492 16. Jensen, M. *et al.* Combinatorial patterns of gene expression changes contribute to variable  
493 expressivity of the developmental delay-associated 16p12.1 deletion. *Genome Med* **13**, 163 (2021).
- 494 17. Nordsletten, A. E. *et al.* Evaluating the Impact of Nonrandom Mating: Psychiatric Outcomes Among  
495 the Offspring of Pairs Diagnosed With Schizophrenia and Bipolar Disorder. *Biol Psychiatry* **87**,  
496 253–262 (2020).
- 497 18. Rasic, D., Hajek, T., Alda, M. & Uher, R. Risk of mental illness in offspring of parents with  
498 schizophrenia, bipolar disorder, and major depressive disorder: a meta-analysis of family high-risk  
499 studies. *Schizophr Bull* **40**, 28–38 (2014).
- 500 19. Sullivan, P. F. *et al.* Family history of schizophrenia and bipolar disorder as risk factors for autism.  
501 *Arch Gen Psychiatry* **69**, 1099–1103 (2012).
- 502 20. Jensen, M. & Girirajan, S. Mapping a shared genetic basis for neurodevelopmental disorders.  
503 *Genome Med* **9**, 109 (2017).
- 504 21. Fischbach, G. D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism  
505 genetic risk factors. *Neuron* **68**, 192–195 (2010).
- 506 22. SPARK Consortium. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research.  
507 *Neuron* **97**, 488–493 (2018).
- 508 23. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**,  
509 203–209 (2018).

- 510 24. Howe, L. J. *et al.* Genetic evidence for assortative mating on alcohol consumption in the UK  
511 Biobank. *Nat Commun* **10**, 5039 (2019).
- 512 25. Retterer, K. *et al.* Assessing copy number from exome sequencing and exome array CGH based on  
513 CNV spectrum in a large clinical cohort. *Genet Med* **17**, 623–629 (2015).
- 514 26. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079  
515 (2009).
- 516 27. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-  
517 generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
- 518 28. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
- 519 29. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291  
520 (2016).
- 521 30. Pedersen, B. S., Layer, R. M. & Quinlan, A. R. Vcfanno: fast, flexible annotation of genetic  
522 variants. *Genome Biol* **17**, 118 (2016).
- 523 31. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from  
524 high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
- 525 32. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res* **49**, D916–D923 (2021).
- 526 33. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the  
527 deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886–D894  
528 (2019).
- 529 34. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism  
530 spectrum disorders (ASDs). *Mol Autism* **4**, 36 (2013).
- 531 35. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat*  
532 *Genet* **51**, 431–444 (2019).
- 533 36. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage  
534 analyses. *Am J Hum Genet* **81**, 559–575 (2007).

- 535 37. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287  
536 (2016).
- 537 38. International HapMap 3 Consortium *et al.* Integrating common and rare genetic variation in diverse  
538 human populations. *Nature* **467**, 52–58 (2010).
- 539 39. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–  
540 5431 (2021).
- 541 40. Constantino, J. N. & Gruber, C. P. *Social responsiveness scale second edition (SRS-2): Manual*.  
542 (Western Psychological Services (WPS), 2012).
- 543 41. Hurley, R. S. E., Losh, M., Parlier, M., Reznick, J. S. & Piven, J. The broad autism phenotype  
544 questionnaire. *J Autism Dev Disord* **37**, 1679–1690 (2007).
- 545 42. Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am J Hum*  
546 *Genet* **83**, 132–135; author reply 135-139 (2008).
- 547 43. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population  
548 structure in large SNP data sets. *Genetics* **197**, 573–589 (2014).
- 549 44. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies.  
550 *Bioinformatics* **26**, 2867–2873 (2010).
- 551 45. Makowski, D., Ben-Shachar, M., Patil, I. & Lüdecke, D. Methods and Algorithms for Correlation  
552 Analysis in R. *JOSS* **5**, 2306 (2020).
- 553 46. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature*  
554 *Methods* **17**, 261–272 (2020).
- 555 47. Dean, K. *et al.* Full spectrum of psychiatric outcomes among offspring with parental history of  
556 mental disorder. *Arch Gen Psychiatry* **67**, 822–829 (2010).
- 557 48. Stapp, E. K., Mendelson, T., Merikangas, K. R. & Wilcox, H. C. Parental bipolar disorder, family  
558 environment, and offspring psychiatric disorders: A systematic review. *J Affect Disord* **268**, 69–81  
559 (2020).

- 560 49. Bodfish, J. W., Symons, F. J., Parker, D. E. & Lewis, M. H. Repetitive behavior scale—revised.  
561 *Journal of Autism and Developmental Disorders* (1999).
- 562 50. Achenbach, T. M. The Child Behavior Checklist and related instruments. in *The use of*  
563 *psychological testing for treatment planning and outcomes assessment, 2nd ed.* 429–466 (Lawrence  
564 Erlbaum Associates Publishers, 1999).
- 565 51. Agrawal, A. *et al.* Assortative mating for cigarette smoking and for alcohol consumption in female  
566 Australian twins and their spouses. *Behav Genet* **36**, 553–566 (2006).
- 567 52. Grant, J. D. *et al.* Spousal concordance for alcohol dependence: evidence for assortative mating or  
568 spousal interaction effects? *Alcohol Clin Exp Res* **31**, 717–728 (2007).
- 569 53. Greve, A. N. *et al.* A Nationwide Cohort Study of Nonrandom Mating in Schizophrenia and Bipolar  
570 Disorder. *Schizophr Bull* **47**, 1342–1350 (2021).
- 571 54. Nordsletten, A. E. *et al.* Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric  
572 Disorders. *JAMA Psychiatry* **73**, 354–361 (2016).
- 573 55. Plomin, R. & Deary, I. J. Genetics and intelligence differences: five special findings. *Mol Psychiatry*  
574 **20**, 98–108 (2015).
- 575 56. Boomsma, D. I. *et al.* Genetic epidemiology of attention deficit hyperactivity disorder (ADHD  
576 index) in adults. *PLoS One* **5**, e10621 (2010).
- 577 57. Krueger, R. F., Moffitt, T. E., Caspi, A., Bleske, A. & Silva, P. A. Assortative mating for antisocial  
578 behavior: developmental and methodological implications. *Behav Genet* **28**, 173–186 (1998).
- 579 58. Mathews, C. A. & Reus, V. I. Assortative mating in the affective disorders: a systematic review and  
580 meta-analysis. *Compr Psychiatry* **42**, 257–262 (2001).
- 581 59. Gottesman, I. I., Laursen, T. M., Bertelsen, A. & Mortensen, P. B. Severe mental disorders in  
582 offspring with 2 psychiatrically ill parents. *Arch Gen Psychiatry* **67**, 252–257 (2010).
- 583 60. Yengo, L. *et al.* Imprint of assortative mating on the human genome. *Nat Hum Behav* **2**, 948–954  
584 (2018).

- 585 61. Robinson, M. R. *et al.* Genetic evidence of assortative mating in humans. *Nat Hum Behav* **1**, 0016  
586 (2017).
- 587 62. Torvik, F. A. *et al.* Modeling assortative mating and genetic similarities between partners, siblings,  
588 and in-laws. *Nat Commun* **13**, 1108 (2022).
- 589 63. Cheng, S. *et al.* Exome-wide screening identifies novel rare risk variants for major depression  
590 disorder. *Mol Psychiatry* **27**, 3069–3074 (2022).
- 591 64. Smoller, J. W. The Genetics of Stress-Related Disorders: PTSD, Depression, and Anxiety Disorders.  
592 *Neuropsychopharmacology* **41**, 297–319 (2016).
- 593 65. Fisher, R. A. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance.  
594 *Earth and Environmental Science Transactions of The Royal Society of Edinburgh* **52**, 399–433  
595 (1919).
- 596 66. Alkuraya, F. S. Homozygosity mapping: one more tool in the clinical geneticist’s toolbox. *Genet*  
597 *Med* **12**, 236–239 (2010).
- 598 67. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity:  
599 windows into population history and trait architecture. *Nat Rev Genet* **19**, 220–234 (2018).
- 600 68. Gamsiz, E. D. *et al.* Intellectual disability is associated with increased runs of homozygosity in  
601 simplex autism. *Am J Hum Genet* **93**, 103–109 (2013).
- 602 69. Lencz, T. *et al.* Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc*  
603 *Natl Acad Sci U S A* **104**, 19942–19947 (2007).
- 604 70. Rosenfeld, J. A., Coe, B. P., Eichler, E. E., Cuckle, H. & Shaffer, L. G. Estimates of penetrance for  
605 recurrent pathogenic copy-number variations. *Genet Med* **15**, 478–481 (2013).
- 606 71. Douard, E. *et al.* Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome.  
607 *AJP* **178**, 87–98 (2021).
- 608 72. Falconer, D. S. Inheritance of liability to certain diseases estimated from incidence among relatives.  
609 *Ann. Hum. Genet.* **29**, 51–76 (1965).



- 610 73. Bergen, S. E. *et al.* Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk  
611 for Schizophrenia. *Am J Psychiatry* **176**, 29–35 (2019).
- 612 74. Martin, J., O’Donovan, M. C., Thapar, A., Langley, K. & Williams, N. The relative contribution of  
613 common and rare genetic variants to ADHD. *Transl Psychiatry* **5**, e506 (2015).
- 614 75. Plomin, R., Krapohl, E. & O’Reilly, P. F. Assortative Mating-A Missing Piece in the Jigsaw of  
615 Psychiatric Genetics. *JAMA Psychiatry* **73**, 323–324 (2016).
- 616 76. Versluys, T. M. M., Flintham, E. O., Mas-Sandoval, A. & Savolainen, V. Why do we pick similar  
617 mates, or do we? *Biol Lett* **17**, 20210463 (2021).
- 618 77. Turnage, D. & Conner, N. Quality of life of parents of children with Autism Spectrum Disorder: An  
619 integrative literature review. *J Spec Pediatr Nurs* **27**, e12391 (2022).
- 620 78. Tambs, K. & Moum, T. No Large Convergence during Marriage for Health, Lifestyle, and  
621 Personality in a Large Sample of Norwegian Spouses. *Journal of Marriage and the Family* **54**, 957  
622 (1992).
- 623 79. Costain, G. *et al.* Adult neuropsychiatric expression and familial segregation of 2q13 duplications.  
624 *Am J Med Genet B Neuropsychiatr Genet* **165B**, 337–344 (2014).
- 625 80. Lupski, J. R., Belmont, J. W., Boerwinkle, E. & Gibbs, R. A. Clan genomics and the complex  
626 architecture of human disease. *Cell* **147**, 32–43 (2011).
- 627 81. Pounraja, V. & Girirajan, S. A general framework for identifying oligogenic combinations of rare  
628 variants in complex disorders. *Genome Res* **32**, 904–915 (2022).
- 629 82. Bittles, A. H. Endogamy, consanguinity and community disease profiles. *Community Genet* **8**, 17–  
630 20 (2005).
- 631 83. Halachev, M. *et al.* Increased ultra-rare variant load in an isolated Scottish population impacts  
632 exonic and regulatory regions. *PLoS Genet* **15**, e1008480 (2019).

633  
634