



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genome wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability

Citation for published version:

Huguet, G, Schramm, C, Douard, E, Tamer, P, Main, A, Monin, P, England, J, Jizi, K, Renné, T, Poirier, M, Nowak, S, Martin, C-O, Younis, N, Knoth, IS, Jean-louis, M, Saci, Z, Auger, M, Tihy, F, Mathonnet, G, Maftai, C, Léveillé, F, Porteous, DJ, Davies, G, Redmond, P, Harris, S, Hill, WD, Lemyre, E, Schumann, G, Bourgeron, T, Pausova, Z, Paus, T, Karama, S, Lippe, S, Deary, I, Almasy, L, Labbe, A, Glahn, DC, M.T Greenwood, C & Jacquemont, S 2021, 'Genome wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability', *Molecular Psychiatry*.
<https://doi.org/10.1038/s41380-020-00985-z>

Digital Object Identifier (DOI):

[10.1038/s41380-020-00985-z](https://doi.org/10.1038/s41380-020-00985-z)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Molecular Psychiatry

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



1 **Genome wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes**
2 **modulate cognitive ability**

3
4 **Single sentence summary:** CNVs' effect-sizes on intelligence are predicted using measures of
5 intolerance to haploinsufficiency and are distributed across half of the coding genes.

6 **AUTHORS:**

7 Guillaume Huguet^{1,2***}; Catherine Schramm^{1,2,3***}; Elise Douard^{1,2}; Petra Tamer^{1,2}, Antoine
8 Main^{2,4}; Pauline Monin^{2,5}; Jade England^{1,2}; Khadije Jizi^{1,2}; Thomas Renne^{2,6}; Myriam Poirier^{1,2};
9 Sabrina Nowak^{1,2}; Charles-Olivier Martin^{1,2}; Nadine Younis^{1,2}; Inga Sophia Knoth^{1,2}; Martineau
10 Jean-Louis^{1,2}; Zohra Saci^{1,2}; Maude Auger^{1,2}; Frédérique Tihy^{1,2}; Géraldine Mathonnet^{1,2};
11 Catalina Maftai^{1,2}; France Léveillé^{1,2}; David Porteous^{7,8,9}, Gail Davies⁷, Paul Redmond⁷, Sarah E.
12 Harris⁷, W. David Hill⁷, Emmanuelle Lemyre^{1,2}; Gunter Schumann¹⁰; Thomas Bourgeron^{11,12,13};
13 Zdenka Pausova¹⁴; Tomas Paus^{15,16,17}; Sherif Karama^{18,19,20}; Sarah Lippe^{2,21}; Ian J. Deary⁷; Laura
14 Almasy²²; Aurélie Labbe⁴; David Glahn²³; Celia M.T. Greenwood^{3,24}; Sébastien Jacquemont^{1,2}

15
16 ** Shared first authorship

17
18 1 Department of Pediatrics, Université de Montréal, Montreal, Quebec, Canada

19 2 Center Hospitalier Universitaire Sainte-Justine Research Center, Montreal, Quebec, Canada

20 3 Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

21 4 Département de Sciences de la Décision, HEC Montreal, Montreal, Quebec, Canada

22 5 Human Genetics and Cognitive Functions, University Paris Diderot, Sorbonne Paris Cité, Paris, France

23 6 Université de Rouen Normandie, UFR des Sciences et Techniques, Rouen, France

24 7 Lothian Birth Cohorts group, Department of Psychology, School of Philosophy, Psychology and Language Sciences, The University of
25 Edinburgh, Edinburgh, EH8 9JZ, UK.

26 8 Medical Genetics Section, Centre for Genomic & Experimental Medicine, MRC Institute of Genetics & Molecular Medicine, University of
27 Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, UK.

28 9 Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.

29 10 Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, England

30 11 Department of Neurosciences, Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France

31 12 Centre National de la Recherche Scientifique Genes, Synapses and Cognition Laboratory, Institut Pasteur, Paris, France

32 13 Human Genetics and Cognitive Functions, University Paris Diderot, Sorbonne Paris Cité, Paris, France

33 14 The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada

34 15 Rotman Research Institute, Baycrest, Toronto, Ontario, Canada

35 16 Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, Canada

36 17 Child Mind Institute, New York, New York

37 18 Montreal Neurological Institute, McGill University, Montreal, QC, Canada.

38 19 McConnell Brain Imaging Center, McGill University, Montreal, QC, Canada

39 20 Douglas Mental Health University Institute, Montreal, QC, Canada

40 21 Psychology, Université de Montréal, Montreal, QC, Canada.

41 22 Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA

42 23 Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; Olin Neuropsychiatric Research Center, Institute of
43 Living, Hartford Hospital, Hartford, CT, USA.

44 24 Gerald Bronfman Department of Oncology, Departments of Epidemiology, Biostatistics & Occupational Health and Human Genetics, McGill
45 University, Montreal, Quebec, Canada

46
47 **Corresponding authors:**

48
49 Guillaume Huguet

50 Sainte Justine University Hospital

51 3175 chemin de la Côte-Sainte-Catherine,

52 Montréal, QC H3T 1C5

53 guillaumeaf.huguet@gmail.com

54
55 Sébastien Jacquemont

56 Sainte Justine University Hospital

57 3175 chemin de la Côte-Sainte-Catherine,

58 Montréal, QC H3T 1C5

59 sebastien.jacquemont@umontreal.ca

60 **ABSTRACT**

61 Genomic Copy Number Variants (CNVs) are routinely identified and reported back to patients
62 with neuropsychiatric disorders, but their quantitative effects on essential traits such as cognitive
63 ability are poorly documented. We have recently shown that the effect-size of deletions on
64 cognitive ability can be statistically predicted using measures of intolerance to
65 haploinsufficiency. However, the effect-sizes of duplications remain unknown. It is also
66 unknown if the effect of multigenic CNVs are driven by a few genes intolerant to
67 haploinsufficiency or distributed across tolerant genes as well.

68 Here, we identified all CNVs >50 kilobases in 24,092 individuals from unselected and autism
69 cohorts with assessments of general intelligence. Statistical models used measures of intolerance
70 to haploinsufficiency of genes included in CNVs to predict their effect-size on intelligence.
71 Intolerant genes decrease general intelligence by 0.8 and 2.6 points of IQ when duplicated or
72 deleted, respectively. Effect-sizes showed no heterogeneity across cohorts. Validation analyses
73 demonstrated that models could predict CNV effect-sizes with 78% accuracy. Data on the
74 inheritance of 27,766 CNVs showed that deletions and duplications with the same effect-size on
75 intelligence occur *de novo* at the same frequency.

76 We estimated that around 10,000 intolerant and tolerant genes negatively affect intelligence when
77 deleted, and less than 2% have large effect-sizes. Genes encompassed in CNVs were not enriched
78 in any GOterms but gene regulation and brain expression were GOterms overrepresented in the
79 intolerant subgroup. Such pervasive effects on cognition may be related to emergent properties of
80 the genome not restricted to a limited number of biological pathways.

81

82 **Introduction**

83

84 Copy Number Variants (CNVs) are deletions or duplications larger than 1000 base pairs. The
85 contribution of CNVs to the etiology of intellectual disability (ID)[1–3], autism[4–6] and
86 schizophrenia[6–8] is well established. The interpretation of CNVs in research and medical
87 diagnostics remains essentially binary: benign or pathogenic (contributing to mental illness)[9].
88 The routine implementation of Chromosomal Micro-Arrays (CMAs) as a first-tier diagnostic test
89 identifies “pathogenic” CNVs in 10 to 15 % of children with neurodevelopmental disorders
90 (NDD)[10]. A binary interpretation is however of limited use because patients present a broad
91 spectrum of cognitive symptoms ranging from severe ID to learning disabilities. The quantitative
92 effects of CNVs are poorly documented even for important traits such as general intelligence. It
93 may be available for the most frequently recurrent CNVs but data is often collected in patients
94 ascertained in the clinic with a bias towards severely affected individuals, leading to potentially
95 gross overestimation of effect size. Only two studies have been conducted in unselected
96 populations [11, 12] showing reduced performance on cognitive test for 24 recurrent CNVs.
97 However, recurrent CNVs only represent a very small fraction of the total amount of ultra-rare
98 CNVs identified in the neurodevelopmental disorder clinic as well as in the general population.
99

100 Intelligence is a major trait assessed in the developmental pediatric and psychiatric clinic. There
101 is a significant genetic correlation between intelligence and psychiatric disorders and cognitive
102 impairments represent a major referral criterion to the NDD clinic. The heritability of general
103 intelligence is estimated at around 50 to 80% [13]. The heritability of variants in linkage
104 disequilibrium with common SNPs is estimated to be around 22.7%, with variants in poor linkage
105 disequilibrium with SNPs, including rare CNVs, explaining 31.3% of the phenotypic variation in
106 intelligence[14]. Two recent GWAS, have identified over 200 loci associated with intelligence
107 and education[15, 16] , potentially implicating 1000 genes. The latter were largely non-

108 overlapping with genes previously linked to ID[15]. Contrary to SNPs, there is no ambiguity in
109 the molecular interpretation of a fully deleted or duplicated gene, which invariably decreases or
110 increases transcription respectively. Therefore, CNVs represent a powerful tool to map the effect-
111 sizes of genes (altered by gene dosage) on human traits.

112 We have previously proposed a framework to estimate and predict the effect-size on intelligence
113 of CNVs. We showed that linear models[17] using the sum of the “probability of being loss-of-
114 function intolerant” (pLI) scores[18] of all genes included in a deletion can predict their effect-
115 size on intelligence quotient (IQ) with 75% accuracy. Our initial study was underpowered to
116 measure the effect-size of duplications. It is also unknown if only a limited number of intolerant
117 genes or a large proportion of genes within CNVs are driving effects on cognitive abilities. More
118 broadly, the number of genes modulating general intelligence remains unknown. The pLI used in
119 our earlier model, ranges from 0 to 1 but has a bimodal distribution and is essentially a
120 categorical variable classifying genes as intolerant (>0.9) or tolerant (≤ 0.9) to protein-loss-of-
121 function (pLoF) [18]. Continuous measures such as the LOEUF[19] (Loss-of-function
122 Observed/Expected Upper bound Fraction) were recently introduced to reflect the full spectrum
123 of intolerance to pLoF. LOEUF range from 0 to 2, and values below 0.35 are suggestive of
124 intolerance.

125 Our present aims were 1) to test the robustness of effect-size estimates for CNVs across
126 unselected and NDD populations, 2) to establish the effect-size on general intelligence of
127 genomic duplications, 3) to investigate the quantitative relationship between effect-size on
128 general intelligence and the frequency of *de novo* events, and 4) to estimate individual effect-
129 sizes for all protein-coding genes that are intolerant as well as tolerant to pLoF.

130 We identified CNVs in 24,092 individuals from five general populations, two autism cohorts and
131 one neurodevelopmental cohort. Measures of intolerance to pLoF were used as variables to
132 estimate the effect of CNVs and individual genes on general intelligence. Validation procedures

133 using cognitive data on CNVs from 47 published reports and the UKBB demonstrated a near
134 80% accuracy of model estimated. We implemented an online tool to help clinicians and
135 researchers estimate the effect-size of any CNVs on general intelligence.

136 **Materials and Methods**

137 **1. Cohorts**

138 We included five cohorts from the general population, two autism cohorts and one familial cohort
139 with at least one CNV-carrier child recruited for a neurodevelopmental disorder (Table1). Studies
140 for each cohort were reviewed by local institutional review boards. Parents/guardians and adult
141 participants gave written informed consent and minors gave assent.

142 **2. Measures of general intelligence**

143 General intelligence was assessed using the neurocognitive tests detailed in table 1. Measures of
144 non-verbal intelligence quotient (NVIQ) were available in five cohorts and general intelligence
145 factor (g-factor)[20] was computed in four cohorts, based on cognitive tests, primarily assessing
146 fluid non-verbal reasoning (Table1, Supplementary Fig. 1). Intelligence measures were
147 normalized using z-score transformations to render them comparable. The concordance between
148 z-scored NVIQ and g-factor available for three cohorts ranged from 60 to 77% (Supplementary
149 Table 1).

150 **3. Genetic information**

151 **CNV calling and filtering**

152 For all SNP array data, we called CNVs with PennCNV and QuantiSNP using previously
153 published methods [17]. For the MSSNG dataset[21], we used CNVs called on whole genome
154 sequencing by Trost *et al.* [22].

155 CNV filtering steps were previously published (Supplemental material). For the mega-analysis,
156 we applied an additional filtering criterion, selecting CNVs encompassing at least 10 probes for
157 all array technologies used across all cohorts.

158 The Sainte-Justine CNV-family cohort included participants on the basis of one pathogenic CNV
159 identified in the diagnostic cytogenetic laboratory using an Agilent 180K array.

160 Annotation of CNVs

161 We annotated the CNVs using Gencode V19 (hg19) with ENSEMBL
162 (<https://grch37.ensembl.org/index.html>). Genes with all transcripts fully encompassed in CNVs
163 were annotated using 12 variables present in previous article[17]. Non-coding regions were
164 annotated with the number of expression quantitative trait loci (eQTLs) regulating genes
165 expressed in the brain[23]. CNV scores were derived by summing all scores of genes within
166 CNVs.[17]. Also, we used a list of 256 ID-genes[2, 24], previously identified with an excess of
167 *de-novo* mutations in NDD cohorts.

168

169 4. Statistical analyses

170 Modelling the effect of CNVs on intelligence

171 General intelligence was adjusted within each cohort for age and sex when required ($Z_{adj\ Intell.}$;
172 see supplemental material and Supplementary Fig. 2 and 3). To estimate the effect of CNVs on
173 general intelligence, we fit the model developed by Huguet et al. [17] where the sum of pLI (or
174 any of the 10 other scores) for all genes encompassed in deletions or duplications, respectively, is
175 the variable used to predict the adjusted Z -score of general intelligence:

$$176 \quad \text{Model for deletion } (\mathcal{M}1_{DEL}): Z_{adj\ Intell.} \sim \beta_{0,DEL} + \beta_{1,DEL} \times \sum_{gene} pLI$$

177 where $\beta_{0,DEL}, \beta_{1,DEL}$ are the regression coefficients. The same model was applied to duplications.

178 First, models $\mathcal{M}1_{DEL}$ and $\mathcal{M}1_{DUP}$ were fitted independently and adjusted for each cohort and
179 results were used in the meta-analyses. Second, in the mega-analysis, $\mathcal{M}1_{DEL}$ and $\mathcal{M}1_{DUP}$ were
180 fitted after pooling all samples and adjusting on the type of cognitive measure and cohort.

181 To take into account ID-genes that have a greater impact on intelligence, we used a model
182 including 4 predictive variables ($\mathcal{M}2$):

$$\begin{aligned}
183 \quad Z_{adj\ intell.} &\sim \beta_0 + \beta_1 \times \sum_{ID\ gene\ in\ deletion} \frac{1}{LOEUF} + \beta_2 \times \sum_{ID\ gene\ in\ duplication} \frac{1}{LOEUF} + \beta_3 \\
184 \quad &\times \sum_{non-ID\ gene\ in\ deletion} \frac{1}{LOEUF} + \beta_4 \times \sum_{non-ID\ gene\ in\ duplication} \frac{1}{LOEUF}
\end{aligned}$$

185 where $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 are the regression coefficients.

186 The variance explained by deletions and duplications (measured by pLI) was computed using
187 partial R^2 in the full dataset as well as the subgroup (n=14,874) of unrelated individuals.

188 Sensitivity analyses

189 We tested non-linearity of the effect of haploinsufficiency scores on general intelligence by using
190 polynomial regression model and by exploring a smooth function of the effect of
191 haploinsufficiency scores using a Gaussian kernel regression method ([https://cran.r-](https://cran.r-project.org/web/packages/KSPM/index.html)
192 [project.org/web/packages/KSPM/index.html](https://cran.r-project.org/web/packages/KSPM/index.html)) flexible enough to account for various types of
193 effects (Supplementary material).

194 Model Validation

195 To validate our models, we computed the concordance between model predictions and loss of IQ
196 measured for 47 recurrent CNVs obtained in previous publications (supplementary material). The
197 concordance was computed using the intraclass coefficient correlation of type (3,1) ($ICC_{(3,1)}$)
198 [25].

199 Modelling the probability to be de novo

200 We performed logistic regressions to estimate the probability of a CNV being *de novo* ($P_{de\ novo}$)
201 as a function of the haploinsufficiency scores:

202
203 Model for deletions ($\mathcal{M}_{3_{DEL}}$):

$$204 \quad \text{logit}(P_{de\ novo}) \sim \beta_{0,DEL} + \beta_{1,DEL} \times Z_{adj\ intell.estimated\ by\ \mathcal{M}_2\ deletion.}$$

205 where $\beta_{0,DEL}$, $\beta_{1,DEL}$ are the regression coefficients. The same model was applied to duplications
206 ($\mathcal{M}3_{DUP}$)

207 For these analyses, we added two clinical populations (Decipher, decipher.sanger.ac.uk/) and the
208 cytogenetic database of Sainte-Justine Hospital, where genetic data could be compared between
209 the child and their parents, and applied the same filtering as for the previous CNV selection
210 leading to a total of 26,437 CNVs. (Supplementary Table 2). The binary outcome variable was
211 the type of transmission (1=*de novo*, 0=inherited).

212 To validate these models, we computed the concordance between model estimates and percentage
213 of *de novo* variants computed with Decipher for 27 recurrent CNVs.

214 Estimating the effect-size of individual genes based on LOEUF values

215 We used 4 categories of LOEUF values to estimate the effect-size of genes classified as highly
216 intolerant (LOEUF <0.2, n=980), moderately intolerant (0.2≤LOEUF<0.35 n=1,762), tolerant
217 (0.35≤LOEUF<1, n=7,442), and highly tolerant to haploinsufficiency (LOEUF≥1, n=8,267). For
218 deletions, model 4 is as follow:

219 ($\mathcal{M}4_{del}$):

$$220 Z_{adj\ intell.} \sim \beta_0 + \beta_1 \times \sum (\text{highly intolerant genes } i) + \beta_2 \times \sum (\text{moderately intolerant genes } i) \\ 221 + \beta_3 \times \sum (\text{tolerant genes } i) + \beta_4 \times \sum (\text{highly tolerant genes } i)$$

222 where $\beta_{0,CNV\ type}$, $\beta_{1,CNV\ type}$, $\beta_{2,CNV\ type}$, $\beta_{3,CNV\ type}$ and $\beta_{4,CNV\ type}$ are the regression
223 coefficients. The same model was applied for duplications.

224 To explore smaller categories of LOEUF values, we slid a window of size 0.15 LOEUF units, in
225 increments of 0.05 units thereby creating 38 categories across the range of LOEUF values. We
226 performed 38 linear models:

227 ($\mathcal{M}5_{del}$):

$$228 Z_{adj\ intell.} \sim \beta_{0,CNV\ type} + \beta_{1,CNV\ type} \times \sum (\text{genes } i \text{ inside the window})$$

229

$$+ \beta_{2,CNV\ type} \times \sum (genes\ i\ outside\ the\ window)$$

230 where $\beta_{0,CNV\ type}$, $\beta_{1,CNV\ type}$ and $\beta_{2,CNV\ type}$ are the regression coefficients.

231 The same models were performed for duplications. Estimates were corrected for multiple testing
232 (38 tests) using FDR.

233 GOterms Enrichment

234 For the GOterms enrichment for the tolerant and intolerant genes with all a genome and CNVs
235 between unselected, ASD and both populations, we used DAVID release 6.8[26] (<https://david->
236 [d.ncifcrf.gov](https://david-ncifcrf.gov)). We kept the defaults parameters and save only the terms with Bonferroni
237 corrected p-values <0.05. We then passed the list to REVIGO[27] (<http://revigo.irb.hr/>) to
238 summarize and group the redundant GO.

239

240

241 RESULTS

242 1) Deletions and duplications have a 3:1 effect-size ratio on general intelligence

243 We first sought to replicate our previous estimates for the effect-size of deletions on
244 general intelligence computed using pLI [17]. We performed a meta-analysis on 20,151
245 individuals from 5 unselected populations (Table 1, Supplementary Fig. 1) showing that the
246 deletion of one point of pLI decreases NVIQ or g-factor by 0.18 z-score (95% CI: -0.23 to -0.14,
247 equivalent to 2.7 points of NVIQ, Fig. 1a, Supplementary Table 3). For duplications, we
248 performed a meta-analysis using the same unselected populations. It shows that duplicating one
249 point of pLI decreases NVIQ or g-factor by 0.04 z-score (95% CI: -0.09 to -0.01), which is
250 equivalent to 0.75 points of IQ. Of notes, our previous study [17] was unable to estimate effect-
251 sizes of duplications on general intelligence, likely due to sample size. There was no
252 heterogeneity across cohorts. Sensitivity analyses showed that methods used for cognitive
253 assessments did not influence these results (Fig. 1, Supplementary Table 4).

254

255 2) The effect-size of CNVs on general intelligence is not influenced by ascertainment.

256 Since genomic variants with large effects on general intelligence are thought to be removed
257 from the general population as a result of negative selective pressure, this may have led to an
258 underestimation of the effect-size of CNVs in unselected populations. To examine this
259 possibility, we analyzed 3,941 individuals (Table 1, Supplementary Fig. 1) from two autism
260 cohorts, which include individuals with ID and *de novo* CNVs. Effect-sizes of pLI on general
261 intelligence were similar in males and females with autism, and the same than those observed in
262 unselected populations for deletions and duplications (Supplementary table 5 and 6). We did not
263 observe any heterogeneity across cohorts (Fig. 1, Supplementary Table 3). Finally, we asked if
264 effect-sizes of pLI were the same in large CNVs rarely observed in the general population or in
265 autism cohorts. We tested 226 CNV carriers and 325 intrafamilial controls from 132 families

266 ascertained in the clinic (Table 1). Effect-sizes of pLI on IQ were very similar with a decrease of
267 0.147 z-score, 95% CI: -0.18 to -0.11 ($P=1.1\times 10^{-15}$) in deletions and 0.069 z-score, 95% CI: -0.1
268 to -0.04 ($P=8.7\times 10^{-6}$) in duplications (Supplementary Table 7).

269

270 3) Mega-analysis suggests additive effects of constraint scores on general intelligence

271 We pooled samples after adjusting for variables including cognitive test and cohorts to perform a
272 mega-analysis of 24,092 individuals carrying 13,001 deletions and 15,856 duplications
273 encompassing 36% of the coding genome (Fig. 1b, Supplementary Fig. 4a). The effect-size of
274 pLI was unchanged, decreasing general intelligence by 0.175 z-score ($SE=0.016$, $P=1.25\times 10^{-28}$)
275 and 0.054 z-score ($SE=0.009$, $P=1.90\times 10^{-9}$) for deletions and duplications, respectively
276 (Supplementary Table 8). The partial R^2 shows that deletions and duplications measured by pLI
277 explain respectively 0.5% and 0.1% of the total variance of intelligence in the complete dataset;
278 in line with the fact that large effect-size CNVs are rare in the general population.

279 Among 11 variables, the 2 main constraint scores (pLI and 1/LOEUF) best explained (based on
280 AIC) the variance of general intelligence (Supplementary Table 8). For the remainder of the
281 study, we transitioned to using LOEUF because it is a continuous variable (the pLI is essentially
282 binary) and is now recommended as the primary constraint score by gnomAD. Analyses using
283 pLI are presented in supplemental results.

284 There was no interaction between constraint scores and age or sex (Supplementary Table 5, 6, 9
285 and 10). Non-linear models did not improve model fit (Supplementary Table 11 to 12),
286 suggesting an additive effect of constraint scores.

287

288 4) The effect-size of 1/LOEUF on intelligence is the same in recurrent neuropsychiatric
289 CNVs and non-recurrent CNVs

290 We show that removing 608 individuals carrying any of the 121 recurrent CNV previously
291 associated with neuropsychiatric conditions[17] does not influence the effect-size of 1/LOEUF
292 on general intelligence (Supplementary Table 13). It has been posited that the deleteriousness of
293 large psychiatric CNVs may be due to interactions between genes encompassed in CNVs. We
294 therefore asked if the effect-size of 1/LOEUF is the same for CNVs encompassing small and
295 large numbers of genes. We recomputed the linear model 6 times after incrementally excluding
296 individuals with a total sum of 1/LOEUF $\geq 60, 40, 20, 10, 4$ and 2.85 for deletions and
297 duplications separately. Effect-sizes remain similar whether deletions encompass >10 or >60
298 points of 1/LOEUF (Fig. 1d, Supplementary Fig. 4b).

299 5) Gene dosage of 1% of coding genes shows extreme effect-size on general intelligence.

300 Our ability to estimate large effect sizes is likely hampered by the explanatory variable
301 (1/LOEUF) used in the model because there is only a 60-fold difference between the smallest and
302 largest value. To improve model accuracy for large effect-size genes, we used a list of 256 ID-
303 genes[2, 24], previously identified with an excess of *de novo* mutations in NDD cohorts. We
304 identified 126 CNVs encompassing at least one ID-gene (Fig. 2).

305 We recomputed the model by integrating 4 explanatory variables: the sum of 1/LOEUF for ID
306 and non-ID-genes encompassed in deletions and duplications. The effect-size on intelligence of
307 1/LOEUF for ID-genes was 7 to 11-fold higher than the effect-size of non-ID genes which
308 remained unchanged (Supplementary Table 14, 15 and Fig. 5). The mean effect of ID-genes
309 intolerant to pLoF (LOEUF <0.35) was a decrease of 20 points of IQ for deletions and 9 points for
310 duplications (Supplementary Table 15).

311

312 6) Model explains nearly 80% of the effect-size of CNVs.

313 As a validation procedure, we compared model estimates to published observations for 47
314 recurrent CNVs reported in clinical series and in the UKBB[11] (Supplementary Table 16 and
315 17). When cognitive data was available from both clinical and the UKBB (n=13), we used the
316 mean of both effect-sizes. Concordance between model estimates and previously published
317 measures was 0.78 for all CNVs (95% CI, 0.66-0.86, $P=4.3\times 10^{-11}$, Fig. 3). Accuracy was similar
318 for deletions (ICC=0.71 [0.5;0.84], $P=1.8\times 10^{-5}$) and duplications (ICC=0.85 [0.7;0.93], $P=3\times 10^{-7}$)
319 as well as for small and large CNVs including trisomy 21 (Fig. 3a and 3b, Supplementary Fig.
320 6 and 7).

321

322 7) CNVs with the same impact on intelligence have the same *de novo* frequency.

323 Because measures of intolerance to haploinsufficiency explain equally well the effect-sizes of
324 deletions and duplications on intelligence, we investigated the relationship between effects on
325 intelligence and *de novo* frequency for deletions and duplications. We established inheritance for
326 26,437 CNVs in 6 cohorts (Supplementary Table 2). There was a strong relationship between
327 effects on general intelligence estimated by the model and the frequency of *de novo* observations
328 for deletions ($P=1.9\times 10^{-65}$) and duplications ($P=4.6\times 10^{-24}$, Fig. 3c).

329 Deletions and duplications with the same impact on general intelligence show similar *de novo*
330 frequency CNVs (Fig. 3c).

331 The concordance between the probability of occurring *de novo* estimated by the model (after
332 removing recurrent CNVs) and *de novo* frequency reported in the DECIPHER database on 31
333 recurrent CNVs was 0.81 ([0.67-0.9]; $P=8.2\times 10^{-8}$) (Fig. 3d, Supplementary Table 18 and Fig. 8).

334

335 8) Estimating effect-sizes of individual genes using LOEUF

336 Since we were underpowered to perform a gene-based GWAS, we first divided all genes in 4
337 categories: highly intolerant genes ($LOEUF < 0.2$; $n=980$), moderately intolerant genes
338 ($0.2 \leq LOEUF < 0.35$ $n=1,762$), tolerant genes ($0.35 \leq LOEUF < 1$; $n=7,442$) and highly tolerant
339 genes ($LOEUF \geq 1$; $n=8,267$). This dichotomization of LOEUF values also allowed to test whether
340 the previous linear models were driven by subgroups of genes. The sum of genes in each category
341 was used as four explanatory variables to explain general intelligence in the same linear model.
342 For deletions, highly, moderately intolerant and tolerant genes showed negative effects on
343 general intelligence (Fig. 4a, Supplementary Table 19). For duplications only moderately
344 intolerant genes showed negative effects (Supplementary Fig. 9 and Table 19).
345 We were underpowered to further subdivide these LOEUF categories, so we tested 38
346 overlapping LOEUF categories in 38 linear models. Each model used 2 explanatory variables:
347 number of genes within and outside the LOEUF category ($size = 0.15 LOEUF$). For
348 haploinsufficiency, negative effects on general intelligence were observed for genes within 13
349 categories across intolerant and tolerant LOEUF values. For duplications, only 2 categories had
350 negative effects (Fig. 4a, Supplementary Fig.9 and Table 20).

351

352 9) Most biological functions affect cognition.

353 The 6,114 different genes encompassed in the CNVs of our dataset did not show any GOterm
354 enrichment except for olfactory related terms (Supplementary Tables 21). We asked if intolerant
355 ($LOEUF < 0.35$) and tolerant genes ($0.35 < LOEUF < 1$), which negatively affect IQ in the analysis
356 above were enriched in GOterms. All intolerant and tolerant genes genome-wide, were enriched
357 in 365 and 30 GOterms respectively (Fig. 4b, Supplementary Tables 22, 23). The largest group of
358 GOterms enriched in intolerant genes represented gene regulation (RNA polymerase II
359 transcription factor activity, chromatin organization; Supplementary Fig. 10), cell death

360 regulation and neuronal function (dendrite and synapse). Among 23 tissues overrepresented in
361 intolerant genes, adult brain and epithelium showed the strongest enrichment (Supplementary
362 Table 22). Top enriched pathways included those in cancer, focal adhesion, Wnt signaling and
363 MAPK (Supplementary Table 22). For tolerant genes, milder enrichments included translation
364 (tRNA) and cytoskeletal structure. Among the 7 significant tissues adult brain showed the
365 strongest enrichment (Fig. 4b, Supplementary Table 23 and Fig. 11). The 2,862 intolerant and
366 tolerant genes encompassed in the CNVs of our dataset showed the same GOterm distribution
367 observed above for the full intolerant and tolerant coding genome. Genes encompassed in CNVs
368 were therefore represented well all molecular functions observed for each LOEUF group at the
369 genome-wide level (Supplementary Table 24).

370

371 **DISCUSSION**

372 Deletions and duplications have effect-sizes on cognitive ability that are robust across cohorts,
373 clinical diagnoses, and general intelligence assessments. The effect-size ratio on cognitive ability
374 of deletions to duplications is 3:1. The linear sum of pLI or 1/LOEUF predicted the effect-size on
375 intelligence of deletions and duplications with equal accuracy (78%). Using categories of LOEUF
376 values, we provide the first estimates for the individual effect-sizes of protein-coding genes,
377 suggesting that half of the coding genome affects intelligence. The 2,862 genes encompassed in
378 CNVs of our dataset show the same GOterm distribution observed in the intolerant and tolerant
379 coding genome.

380

381 [Model validation and ascertainment biases](#)

382 Models show 78% concordance with effect-size of CNVs on IQ from previous literature reports.
383 Estimates are discordant for several CNVs, which may be due to either 1) unidentified large
384 effect-size genes with unreliable LOEUF measures due to the small size of the protein coding
385 region, and 2) ascertainment bias. However, biases from clinically referred individuals can be
386 adjusted for using intrafamilial controls [28, 29]. This is confirmed by effect-sizes using the Ste
387 Justine family genetic cohort. Also, our results suggest that the effect-size of pathogenic CNVs
388 are underestimated in the UKBB[28] while those of small CNVs are largely overestimated in
389 clinical series. The maximum effect size measured in UKBB was only 0.4 z-score including
390 pathogenic CNVs such as 16p11.2, 2q11.2 deletions and 10q11.21-q11.23 deletion containing an
391 ID-gene (*WDFY4*). On the other hand, the effect size of variants such as the 16p13.11
392 duplications and 1q21.1 CNVs are likely overestimated in clinical series[30]. Therefore,
393 statistical models using a variety of disease and unselected cohorts are likely to provide the most
394 accurate estimates. Surprisingly, an autism diagnosis is not associated with a different impact of
395 CNVs on cognitive ability. A recent study characterizes this finding showing that CNVs similarly

396 decrease IQ in autism and in unselected populations but are nevertheless more frequent in autism
397 than in controls with same intelligence[31].

398

399 Individual effect-sizes of genes, and go their GOterm enrichments

400 Our study is based on CNVs encompassing intolerant and tolerant genes with the same GOterm
401 distribution observed in those LOEUF categories genome-wide. Only one percent of coding
402 genes with the highest intolerance to pLoF has large effects on cognitive ability (20 and 9 IQ
403 points for deletions and duplications of ID genes). The rest of the intolerant genes (15% of coding
404 genes) have moderate to mild effect-sizes. The group of all intolerant genes is enriched in many
405 GOterms including brain expression and gene regulation as previously reported for this group[2,
406 32]. Genes considered tolerant to pLoF ($0.35 < \text{LOEUF} < 1$; 40% of coding genes) impact
407 intelligence with small effect-size and are only mildly enriched in GOterms. This is reminiscent
408 of GWAS results for schizophrenia showing that most GOterms contribute to it's heritability
409 [33].

410

411 Potential clinical application

412 Models developed in this study provide a translation of gnomAD constraint scores into cognitive
413 effect-sizes. Model outputs are implemented in a prediction tool (<https://cnvprediction.urca.ca/>),
414 which is designed to estimate the population-average effect-size of any given CNV on general
415 intelligence, not the cognitive ability of the individual who carries the CNV. If the cognitive
416 deficits of an individual are concordant with the effect-size of the CNV they carry, one may
417 conclude that the CNV contributes substantially to those deficits. When discordant (ie. The ob-
418 served IQ drop is ≥ 15 points (1SD) larger than the model estimate), the clinician may conclude
419 that a substantial proportion of the contribution lies in additional factors which should be
420 investigated, such as additional genetic variants and perinatal adverse events (e.g. neonatal

421 hypoxic ischemic injury, seizure disorders etc). If IQ cannot be reliably measured (ie. ≤ 4 years or
422 in the case of severe behavioral disorders), the cognitive impact of the CNV predicted by the
423 model may allow to anticipate the need for potential interventions. Overall, the output of this tool
424 can help interpret CNVs in the clinic, but estimates should be interpreted with caution. The model
425 can provide an estimate for the effect size on intelligence of individual genes when deleted.
426 Therefore, one may use this information to estimate the effect size on intelligence of any SNV
427 resulting in a loss of function. However, larger datasets are required to refine the estimates for
428 individual gene.

429

430 [The relationship between genetic fitness and cognitive abilities](#)

431 The reasons underlying the tight relationship between general intelligence and epidemiological
432 measures of intolerance to pLoF, is unclear. This relationship is further highlighted by the fact
433 that deletions and duplications with the similar impact on intelligence occur *de novo* with similar
434 frequencies. Behavioral interpretations are intuitive for severe ID but do not apply for CNVs with
435 much milder effects. In other words, individuals with moderate or severe ID have limited
436 offspring due to behavioral deficits but it is unclear how small changes in intelligence may lead
437 to behavioral issues resulting in decreased fitness. Our results also suggest that genes considered
438 as “tolerant” with LOEUF < 1 affect cognitive abilities and are likely under “mild constraint”.
439 Larger samples are required to better characterize the effect of this broad category of “mildly
440 intolerant” genes on cognitive ability.

441

442 [Limitations](#)

443 The model relies on constraint scores (LOEUF or pLI), which are epidemiological measures of
444 genetic fitness in human populations, without any consideration of gene function[18, 19]. It is
445 likely that some genes decrease fitness (eg. genes involved in fertility) without affecting general

446 intelligence. Further studies combining intolerance scores with functional categories are required
447 to investigate this question. While LOEUF was designed to measure intolerance to loss of
448 function, we used it to assess both deletions and duplications. However, our results and a recent
449 report suggest that it also measures the intolerance to increased gene expression [34]. Noise in
450 the model may be related to unreliable constraint scores computed for small genes with a limited
451 number of pLoF variants observed in the gnomAD database. Bias in the model may be
452 introduced by ID genes observed in our dataset. Indeed, they may reflect a less severe subgroup
453 and model outputs should be interpreted with caution when CNVs encompass ID-genes. Another
454 potential bias is related to the fact that models were trained on CNVs encompassing 36% of the
455 coding genome. Projections suggest that 500K individuals from an unselected population would
456 cover 78% (Supplementary Fig. 12).

457 Finally, all models imply additive effects and massive datasets would be required to test for gene-
458 gene and gene-environment interactions. However, the fact that very large CNVs (such as trisomy
459 21) are accurately estimated by the model suggests that genetic interactions within large genomic
460 segments or even chromosomes cannot be readily observed. There is long standing discordance
461 between observations made at the microscopic and macroscopic level. Indeed, molecular studies
462 provide unequivocal evidence that gene-gene interactions are common but quantitative genetic
463 theory suggests that contributions from non-additive effects to phenotypic variation in the
464 population are small. Reconciling these two observations, polygenic models assume that
465 interactions are the rule rather than the exception. Interactions are, in fact, accounted for in the
466 additive models[35]. For example, LOEUF values are correlated with the number of protein-
467 protein interactions[19] and our results also show that the intolerant genes are enriched in
468 GOterms linked to “gene regulation”. In other words, the level of interactions for a given gene is
469 directly related to its “individual” effect size on intelligence (ie. chromatin remodelers have a
470 very broad interaction network, low LOEUF values and high effect sizes on intelligence).

471

472 **Conclusions**

473 The effect-size of deletions or duplications on intelligence can be accurately estimated with
474 additive models using constraint scores. The same relationship between gene dosage and
475 cognition apply to small benign CNVs as well as extreme CNVs such as Down syndrome. We
476 provide a map of effect-sizes at the individual gene level but to move beyond this rough outline,
477 much larger sample sizes are required. Nonetheless, these results suggest that a large proportion
478 (56%) of the coding genome covering all molecular functions influences cognitive abilities. One
479 may therefore view the genetic contribution to cognitive difference as an emergent property of
480 the entire genome not restricted to a limited number of biological pathways.

481 **REFERENCES**

482

- 483 1. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P,
484 et al. Refining analyses of copy number variation identifies specific genes associated with
485 developmental delay. *Nat Genet.* 2014;46:1063–1071.
- 486 2. Coe BP, Stessman HAF, Sulovari A, Geisheker MR, Bakken TE, Lake AM, et al.
487 Neurodevelopmental disease genes implicated by de novo mutation and copy number variation
488 morbidity. *Nat Genet.* 2019;51:106–116.
- 489 3. Wilfert AB, Sulovari A, Turner TN, Coe BP, Eichler EE. Recurrent de novo mutations in
490 neurodevelopmental disorders: properties and clinical implications. *Genome Med.* 2017;9.
- 491 4. Huguet G, Ey E, Bourgeron T. The genetic landscapes of autism spectrum disorders.
492 *Annu Rev Genomics Hum Genet.* 2013;14:191–213.
- 493 5. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of
494 Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. *Am J Hum Genet.*
495 2014;94:677–694.
- 496 6. Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S, et al. The
497 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Mol*
498 *Psychiatry.* 2015;20:140–147.
- 499 7. Sakai M, Watanabe Y, Someya T, Araki K, Shibuya M, Niizato K, et al. Assessment of
500 copy number variations in the brain genome of schizophrenia patients. *Mol Cytogenet.* 2015;8.
- 501 8. Szatkiewicz JP, O’Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM, et al. Copy
502 number variation in schizophrenia in Sweden. *Mol Psychiatry.* 2014;19:762–773.
- 503 9. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical
504 standards for the interpretation and reporting of constitutional copy-number variants: a joint
505 consensus recommendation of the American College of Medical Genetics and Genomics
506 (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med.* 2019:1–13.
- 507 10. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al.
508 Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for
509 Individuals with Developmental Disabilities or Congenital Anomalies. *Am J Hum Genet.*
510 2010;86:749–764.
- 511 11. Kendall KM, Bracher-Smith M, Fitzpatrick H, Lynham A, Rees E, Escott-Price V, et al.
512 Cognitive performance and functional outcomes of carriers of pathogenic copy number variants:
513 analysis of the UK Biobank. *Br J Psychiatry.* 2019;214:297–304.
- 514 12. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K,
515 Arnarsdottir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in
516 controls. *Nature.* 2014;505:361–366.
- 517 13. Posthuma D, de Geus EJC, Boomsma DI. Perceptual Speed and IQ Are Associated
518 Through Common Genetic Factors. *Behav Genet.* 2001;31:593–602.
- 519 14. Hill WD, Arslan RC, Xia C, Luciano M, Amador C, Navarro P, et al. Genomic analysis of
520 family data reveals additional genetic effects on intelligence and personality. *Mol Psychiatry.*
521 2018;23:2347–2362.
- 522 15. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-
523 wide association meta-analysis in 269,867 individuals identifies new genetic and functional links
524 to intelligence. *Nat Genet.* 2018;50:912–919.
- 525 16. Hill WD, Marioni RE, Maghziyan O, Ritchie SJ, Hagenaars SP, McIntosh AM, et al. A
526 combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis
527 and myelination in intelligence. *Mol Psychiatry.* 2019;24:169–181.
- 528 17. Huguet G, Schramm C, Douard E, Jiang L, Labbe A, Tihy F, et al. Measuring and
529 Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-

530 Based Samples. *JAMA Psychiatry*. 2018;75:447–457.

531 18. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of
532 protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291.

533 19. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation
534 across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance
535 across human protein-coding genes. *BioRxiv*. 2019:531210.

536 20. Deary IJ. *Intelligence*. *Annu Rev Psychol*. 2011;63:453–482.

537 21. Yuen RKC, Merico D, Bookman M, Howe JL, Thiruvahindrapuram B, Patel RV, et al.
538 Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum
539 disorder. *Nat Neurosci*. 2017;20:602–611.

540 22. Trost B, Walker S, Wang Z, Thiruvahindrapuram B, MacDonald JR, Sung WWL, et al. A
541 Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from
542 Whole-Genome Sequence Data. *Am J Hum Genet*. 2018;102:142–155.

543 23. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic
544 variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci*.
545 2014;17:1418–1428.

546 24. McRae JF, Clayton S, Fitzgerald TW, Kaplanis J, Prigmore E, Rajan D, et al. Prevalence
547 and architecture of de novo mutations in developmental disorders. *Nature*. 2017;542:433–438.

548 25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol*
549 *Bull*. 1979;86:420–428.

550 26. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene
551 lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44–57.

552 27. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long
553 Lists of Gene Ontology Terms. *PLOS ONE*. 2011;6:e21800.

554 28. D’Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L, et al. Defining
555 the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities.
556 *JAMA Psychiatry*. 2016;73:20–30.

557 29. Moreno-De-Luca A, Evans DW, Boomer KB, Hanson E, Bernier R, Goin-Kochel RP, et
558 al. The Role of Parental Cognitive, Behavioral, and Motor Profiles in Clinical Variability in
559 Individuals With Chromosome 16p11.2 Deletions. *JAMA Psychiatry*. 2015;72:119–126.

560 30. Bernier R, Steinman KJ, Reilly B, Wallace AS, Sherr EH, Pojman N, et al. Clinical
561 phenotype of the recurrent 1q21.1 copy-number variant. *Genet Med*. 2016;18:341–349.

562 31. Douard E, Zeribi A, Schramm C, Tamer P, Loum MA, Nowak S, et al. Effect Sizes of
563 Deletions and Duplications on Autism Risk Across the Genome. *Am J Psychiatry*.
564 2020:appi.ajp.2020.19080834.

565 32. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, et al. Large-
566 Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the
567 Neurobiology of Autism. *Cell*. 2020;180:568-584.e23.

568 33. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to
569 omnigenic. *Cell*. 2017;169:1177–1186.

570 34. An open resource of structural variation for medical and population genetics | bioRxiv.
571 <https://www.biorxiv.org/content/10.1101/578674v1.full>. Accessed 31 December 2019.

572 35. Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common Disease Is More
573 Complex Than Implied by the Core Gene Omnigenic Model. *Cell*. 2018;173:1573–1580.

574 36. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C, et al. The
575 IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology.
576 *Mol Psychiatry*. 2010;15:1128–1139.

577 37. Pausova Z, Paus T, Abrahamowicz M, Bernard M, Gaudet D, Leonard G, et al. Cohort
578 Profile: The Saguenay Youth Study (SYS). *Int J Epidemiol*. 2017;46:e19.

579 38. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort Profile: The Lothian Birth Cohorts of 1921

580 and 1936. *Int J Epidemiol.* 2012;41:1576–1584.

581 39. Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet J-P, Knoppers B, et al. Cohort
582 profile of the CARTaGENE study: Quebec’s population-based biobank for public health and
583 personalized genomics. *Int J Epidemiol.* 2013;42:1285–1299.

584 40. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort
585 Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its
586 participants and their potential for genetic research on health and illness. *Int J Epidemiol.*
587 2013;42:689–700.

588 41. Fischbach GD, Lord C. The Simons Simplex Collection: A Resource for Identification of
589 Autism Genetic Risk Factors. *Neuron.* 2010;68:192–195.

590

591

592

593

594 **Acknowledgments**

595

596 **Conflict of interest:** The authors declare that they have no conflict of interest.

597

598 **Funding/Support:** This research was enabled by support provided by Calcul Quebec
599 (<http://www.calculquebec.ca>) and Compute Canada (<http://www.computecanada.ca>).

600 Sebastien Jacquemont is a recipient of a Bursary Professor fellowship of the Swiss National
601 Science Foundation, a Canada Research Chair in neurodevelopmental disorders, and a chair from
602 the Jeanne et Jean Louis Levesque Foundation. Catherine Schramm is supported by an Institute
603 for Data Valorization (IVADO) fellowship. Petra Tamer is supported by a Canadian Institute of
604 Health Research (CIHR) Scholarship Program. Guillaume Huguet is supported by the Sainte-
605 Justine Foundation, the Merit Scholarship Program for foreign students, and the Network of
606 Applied Genetic Medicine fellowships. Thomas Bourgeron is a recipient of a chair of the
607 Bettencourt-Schueler foundation. This work is supported by a grant from the Brain Canada
608 Multi-Investigator initiative and CIHR grant 159734 (Sebastien Jacquemont, Celia Greenwood,
609 Tomas Paus). The Canadian Institutes of Health Research and the Heart and Stroke Foundation of
610 Canada fund the Saguenay Youth Study (SYS). SYS was funded by the Canadian Institutes of
611 Health Research (Tomas Paus, Zdenka Pausova) and the Heart and Stroke Foundation of Canada
612 (Zdenka Pausova). Funding for the project was provided by the Wellcome Trust. This work was
613 also supported by an NIH award U01 MH119690 granted to Laura Almasy, Sebastien
614 Jacquemont and David Glahn and U01 MH119739. The authors wish to acknowledge the
615 resources of MSSNG (www.mss.ng), Autism Speaks and The Centre for Applied Genomics at
616 The Hospital for Sick Children, Toronto, Canada. We also thank the participating families for
617 their time and contributions to this database, as well as the generosity of the donors who
618 supported this program. We are grateful to all the families who participated in the Simons

619 Variation in Individuals Project (VIP) and the Simons VIP Consortium (data from Simons VIP
620 are available through SFARI Base). We thank the coordinators and staff at the Simons VIP and
621 SCC sites. We are grateful to all of the families at the participating SSC sites and the principal
622 investigators (A. Beaudet, M.D., R. Bernier, Ph.D., J. Constantino, M.D., E. Cook, M.D., E.
623 Fombonne, M.D., D. Geschwind, M.D., Ph.D., R. Goin-Kochel, Ph.D., E. Hanson, Ph.D., D.
624 Grice, M.D., A. Klin, Ph.D., D. Ledbetter, Ph.D., C. Lord, Ph.D., C. Martin, Ph.D., D. Martin,
625 M.D., Ph.D., R. Maxim, M.D., J. Miles, M.D., Ph.D., O. Ousley, Ph.D., K. Pelphrey, Ph.D., B.
626 Peterson, M.D., J. Piggot, M.D., C. Saulnier, Ph.D., M. State, M.D., Ph.D., W. Stone, Ph.D., J.
627 Sutcliffe, Ph.D., C. Walsh, M.D., Ph.D., Z. Warren, Ph.D., and E. Wijsman, Ph.D.). We
628 appreciate obtaining access to phenotypic data on SFARI base.

629
630 **Additional Contributions:** Julien Buratti (Institute Pasteur), and Vincent Frouin, Ph.D.
631 (Neurospin), acquired data for IMAGEN. Manon Bernard, BSc (database architect, The Hospital
632 for Sick Children), and Helene Simard, MA, and her team of research assistants (Cégep de
633 Jonquière) acquired data for the Saguenay Youth Study. Antoine Main, M.Sc. (UHC Sainte-
634 Justine Research Center, HEC Montreal), Lionel Lemogo, M.Sc. (UHC Sainte-Justine Research
635 Center), and Claudine Passo, Pg.D. (UHC Sainte-Justine Research Center), provided
636 bioinformatical support. Maude Auger, Pg.D.; and Kristian Agbogba, B.Sc. (UHC Sainte-Justine
637 Research Center), provided website development. Dr. Paus is the Tanenbaum Chair in Population
638 Neuroscience at the Rotman Research Institute, University of Toronto, and the Dr. John and
639 Consuela Phelan Scholar at Child Mind Institute, New York.

640
641 **Role of the Funder/Sponsor:** The funder had no role in the design and conduct of the study;
642 collection, management, analysis, or interpretation of the data; preparation, review, or approval of
643 the manuscript; or decision to submit the manuscript for publication.

Ascertainment	Cohort	Array type	n=	Females, n (%)	Age in years Mean (SD)	Type of intelligence measures	Z-scored intelligence measure Mean (SD)
Unselected (n=20,151)	IMAGEN [36]	610Kq; 660Wq	1,744	891 (51%)	14.4 (0.4)	WISC-IV (and g-factor, similarities score, vocabulary score, block design score, matrix reasoning score)	0.44 (0.98) ***
	SYS children[37]	610Kq; HOE-12V	967	505 (52%)	15.0 (1.8)	WISC-III (and g-factor using 63 cognitive measures†)	0.30 (0.87) ***
	SYS parents[37]	HOE-12V	602	321 (53%)	49.5 (4.9)	g-factor, 12 cognitive measures‡	0 (1)
	LBC1936[38]	610Kq	504	247 (49%)	70.0 (-)*	Moray House Test (and g-factor)	0.05 (0.96) ***
	CaG-GSA[39]	GSA	2,074	1,094 (53%)	54.3 (7.6)	g-factor, Reasoning, Memory, Reaction time	-0.02 (1.03)
	CaG-Omni2.5[39]	Omni2.5	515	281 (55%)	52.4 (8.6)		-0.10 (1.02)
	CaG (all)[39]	GSA; Omni2.5	2,589	1,375 (53%)	53.9 (7.8)		-0.03 (1.03)
G-Scot[40]	610Kq	13,745	8,101 (59%)	46.7 (15.0)	g-factor, Logical Memory, Digit Symbol, Verbal fluency, Mill Hill Vocabulary	0.00 (0.99)	
Autism (n=3,941)	SSC-1Mv1[41]	1Mv1	332	44 (13%)	9.5 (3.2)	WISC-IV n=19; DAS-II E-Y n=96; DAS-II S-A n=179; Mullen n=12; WASI-I n=26	-0.55 (1.59)
	SSC-1Mv3[41]	1Mv3	1,182	157 (13%)	8.8 (3.5)	WISC-IV n=16; DAS-II E-Y n=531; DAS-II S-A n=539; Mullen n=77; WASI-I n=19	-0.98 (1.66)
	SSC-Omni2.5[41]	Omni2.5	1,048	140 (13%)	9.2 (3.7)	WISC-IV n=10; DAS-II E-Y n=403; DAS-II S-A n=494; Mullen n=124; WASI-I n=17	-1.25 (1.87)
	SSC (all)[41]	1Mv1; 1Mv3; Omni2.5	2,562	341 (13%)	9.03 (3.6)	WISC-IV n=45; DAS-II E-Y n=1,030; DAS-II S-A n=1,212; Mullen n=213; WASI-I n=62	-1.03 (1.75)
	MSSNG[21]	WGS	1,379	275 (20%)	9.2 (4.4)	WISC-IV n=46; WASI-II n=338; Leiter n=372; Raven n=214; Stanford Binet n=281; WPPSI n=128	-0.47 (1.58)
NDD** (n=551)	Ste-Justine-probands	Agilent 180 K array	132	52 (39%)	7.23 (5.46)	WISC-V n=36; WASI-II n=8; WPPSI-IV n=38; Leiter-R n=18; Mullen n=32	-1.31 (1.02)
	Ste-Justine-siblings		87	44 (50%)	7.75 (5.72)	WISC-V n=28; WASI-II n=13; WPPSI-IV n=31; Leiter-R n=3; Mullen n=12	-0.29 (0.98)
	Ste-Justine-parents		310	180 (58%)	37.80 (7.13)	WASI-II	-0.10 (1.16)
	Ste-Justine-other members		22	12 (54%)	43 (21.27)	WASI-II	-0.04 (1.32)

646
647 **Table 1. Cohort descriptions**

648 Cohorts include 24,092 individuals, including 14,874 unrelated individuals. SSC and CaG
649 cohorts were broken down into sub-samples based on array technology (Supplementary
650 methods). †63 and ‡ 12 cognitive measures were respectively used to compute the g-factor in
651 SYS children and parents (Supplementary methods). NDD: neurodevelopmental disorders, SYS:
652 Saguenay Youth Study, CaG: CARTaGEN, LBC: Lothian Birth Cohort, SSC: Simons Simplex

653 Collection; n=number of individuals remaining for analysis after quality control. The mean and
654 Standard Deviation (SD) for g-factor slightly deviate from 0 and 1 in some cohorts since they
655 were computed on all available data (before the exclusion of some individuals for poor quality
656 array) and summarized here only for individuals included in the analyses. *All individuals from
657 LBC1936 were assessed at 70 years old explaining the absence of SD computation. **The NDD
658 cohort was used only in the replication analysis and was not included in meta- or mega-analyses.
659 *** We displayed the Z-scores of IQ, because IQ was preferred to g-factor for all analyses, even
660 if results were similar (Supplementary Table 3 and 7).
661

662 **Fig. 1. Effect of intolerant score on general intelligence measured for deletions and**
663 **duplications.**
664 Meta-analysis estimating the effect of deletions **a.** and duplications **b.**, measured by sum of pLI,
665 on general intelligence (Supplementary Table 3). X-axis values represent z-scores of general
666 intelligence. Deleting one point of pLI decreases the general intelligence by 0.18 z-scores (2.7
667 points of IQ). Duplicating one point of pLI decreases the general intelligence by 0.05 z-scores
668 (0.75 points of IQ). The squares represent the effect-size computed for each sample. Their size
669 negatively correlated to variance. Diamonds represent the summary effect across cohorts. Their
670 lengths correspond to the 95% confidence intervals of the mean effect-size. **c.** Estimated
671 proportion of the coding genome within each category defined by LOEUF, encompassed in
672 CNVs present in the mega-analysis according to sample size (randomly selected within the mega-
673 analysis). We observed $N_{CNVs\ gene}=6,315$ with $N_{Del.\ gene}=2,282$ and $N_{Dup.\ gene}=5,223$). **d.** Estimated
674 effect of $1/LOEUF$ on general intelligence after removing individuals with a sum of $1/LOEUF$
675 larger than 60, 40, 20, 10, 4 and 2.85 (2.85 corresponds to $1/0.35$, the cut-off for intolerance to
676 pLoF gnomAD). n: number of individuals with a total sum of $1/LOEUF > 0$.

677

678

679 **Fig. 2. Effect-size of intellectual disability (ID) genes on general intelligence.**

680 **a.** Venn diagram of ID genes in ASD and in general population cohorts. We identified 66 CNVs
681 encompassing at least one ID-gene in ASD cohorts (31 deletions and 35 duplications) and 60 in
682 the general population (13 deletions and 47 duplications) (Supplementary methods). Genes were
683 previously defined as harboring an excess of *de novo* loss of function (bold) or missense
684 mutations in neurodevelopmental cohorts: (a) *DYNC1H1*, ***PHF21A***, ***SHANK3***, ***TRA2B***, ***FOXP1***,
685 ***SETD5***, *NR4A2*, *TCF7L2*, ***SOX5***, ***POU3F3***, ***ARID1B***, ***EBF3***, *HNRNPU*; (b) *SET*, ***ZBTB18***,
686 ***DLG4***, ***CHAMP1***, ***CNOT3***, *U2AF2*, *HIST1H2AC*, *DNM1*, ***RAI1***, ***CREBBP***, ***HIST1H1E***,
687 ***ASXL1***, *CABP7*; (c) *PRPF18*, *PPP2R1A*, *EEF1A2*; (d) *TRAF7*, *DEAF1*, *STC1*, ***MYT1L***, ***BRPF1***,
688 ***CBL***, ***SPAST***, ***WDR87***, ***NFE2L3***, ***STARD9***, ***TCF20***, ***KMT2C***, ***FAM200B***, ***KDM5B***, ***CHD2***,
689 ***BTF3***, ***ITPR1***, ***HMGXB3***. **b.** Effect-size of 1/LOEUF on general intelligence estimated in a model
690 using two explanatory variables (sum of 1/LOEUF of deleted and duplicated genes) or 4
691 explanatory variables (sum of 1/LOEUF of ID genes and non-ID genes for deletions and
692 duplication).

693

694

695 **Fig. 3. Concordance between model predictions and published observations for CNV effects**
696 **on general intelligence and for *de novo* frequency.**

697 **a. and b.** Concordance between model estimates (with 1/LOEUF and ID-genes) and literature of
698 clinical data and UKBB reports for general intelligence loss observed in respectively 27 and 33
699 recurrent CNVs for a total of ascertained carriers of 47 recurrent CNVs (Supplementary Table
700 17). X- and Y-values: effect size of CNVs on z-scored general intelligence. **b.** Zoom of the
701 rectangle drawn in the lower left section of panel **a.** We represented values from clinical data by a
702 circle and those from UKBB data by a square. The cross represents the mean value of z-scored
703 IQ loss for the 13 recurrent CNVs observed both in literature and in UKBB. **c. and d.** The model
704 uses 2 explanatory variables (1/LOEUF of non-ID-genes and ID-genes). **c.** Probability of *de novo*
705 estimated by our *de novo* model (Y-axis) according to the loss of IQ estimated by a model using
706 1/LOEUF for ID and non-ID genes as two explanatory variables (X-axis). The *de novo* model
707 was fitted on 13,114 deletions (red) and 13,323 duplications (blue) with available inheritance
708 information observed in DECIPHER, CHU Sainte-Justine, SSC, MSSNG, SYS and G-Scot. **d.**
709 Concordance between *de novo* frequency observed in DECIPHER (X-axis) and the probability of
710 being *de novo* estimated by models when excluding recurrent CNVs of the training dataset (Y-
711 axis) 1/LOEUF for ID and non-ID genes as an explanatory variable for 27 recurrent CNVs. The
712 first bisector represents the perfect concordance. Deletions are in red and duplications in blue.
713 Empty circles or square are CNVs encompassing ID-genes. ICC indicates intraclass correlation
714 coefficient (3, 1). Each point represents a recurrent CNV: (1) TAR Deletion; (2) 1q21.1 Deletion;
715 (3) 2q11.2 Deletion; (4) 2q13 Deletion; (5) *NRXN1* Deletion; (6) 2q13 (*NPHPI*) Deletion; (7)
716 3q29 (*DLG1*) Deletion; (8) 7q11.23 (William-Beuren) Deletion; (9) 8p23.1 Deletion; (10)
717 10q11.21q11.23 Deletion; (11) 13q12.12 Deletion; (12) 13q12 (*CRYLI*) Deletion; (13) 15q13.3
718 (BP4-BP5) Deletion; (14) 15q11.2 Deletion; (15) 16p11.2-p12.2 Deletion; (16) 16p13.3 ATR-16
719 syndrome Deletion; (17) 16p11.2 Deletion; (18) 16p11.2 distal Deletion; (19) 16p13.11 Deletion;

720 (20) 16p12.1 Deletion; (21) 17p11.2 (Smith-Magenis) Deletion; (22) 17q12 Deletion; (23)
721 17q21.31 Deletion; (24) NF1-microdeletion syndrome Deletion; (25) 17p12 (*HNPP*) Deletion;
722 (26) 22q11.2 Deletion; (27) TAR Duplication; (28) 1q21.1 Duplication; (29) 2q21.1 Duplication;
723 (30) 2q13 Duplication; (31) 2q13 (*NPHPI*) Duplication; (32) 7q11.23 Duplication; (33)
724 10q11.21q11.23 Duplication; (34) 13q12.12 Duplication; (35) 15q11q13 (BP3-BP4) Duplication;
725 (36) 15q11.2 Duplication; (37) 15q13.3 Duplication; (38) 15q13.3 (*CHRNA7*) Duplication; (39)
726 16p11.2 Duplication; (40) 16p11.2 distal Duplication; (41) 16p13.11 Duplication; (42) 16p12.1
727 Duplication; (43) 17p11.2 Duplication; (44) 17q12 (*HNF1B*) Duplication; (45) 17p12 (*CMT1A*)
728 Duplication; (46) Trisomic 21 Duplication; (47) 22q11.2 Duplication.
729

730 **Fig. 4. Effect-size on general intelligence of individual genes encompassed in CNVs and**
731 **their GOterms enrichment.**

732 **a.** , the light grey histogram represents the distribution of LOEUF values for 18,451 autosomal
733 genes. The blue line represents the estimates for a gene in each of the 4 categories of LOEUF
734 included in the model (Supplementary methods): highly intolerant genes ($LOEUF < 0.2$, $n=980$),
735 moderately intolerant genes ($0.2 \leq LOEUF < 0.35$, $n=1,762$), tolerant genes ($0.35 \leq LOEUF < 1$,
736 $n=7,442$) and genes highly tolerant to pLoF ($LOEUF \geq 1$, $n=8,267$). The orange line represents the
737 estimated effect-size of 37 categories of genes based on their LOEUF values (sliding
738 windows=0.15) in the model (Supplementary methods). Genes with a LOEUF below 0.35
739 (vertical red line) are considered to be intolerant to pLoF by gnomAD. Left Y-axis values: z-
740 scored general intelligence (1 z-score is equivalent to 15 points of IQ) for deletion. Right Y-axis
741 values: number of genes represented in the histogram. **b.** each point represents a GOterm for
742 which enrichment was observed for all intolerant ($n=2,742$) or tolerant genes ($n=7,442$), for all
743 intolerant ($n=609$) or tolerant genes ($n=2,251$) encompassed in CNVs when compared to the
744 whole coding genome (Bonferroni). X-axis: % of genes included in the GOterm genome-wide;
745 Y-axis: % of genes included in the GOterm for all intolerant ($0 < LOEUF < 0.35$) and tolerant genes
746 ($0.35 \leq LOEUF < 1$).