Estimating the effect-size of gene dosage on cognitive ability across the coding genome

AUTHORS:
Guillaume Huguet[1,2]**; Catherine Schramm[1,2,3]**; Elise Douard[1,2]; Tamer Petra[1,2], Antoine Main[2,4]; Pauline Monin[2,5]; Jade England[1,2]; Khadije Jizi[1,2]; Thomas Renne[2,6]; Myriam Poirier[1,2]; Sabrina Nowak[1,2]; Charles-Olivier Martin[1,2]; Nadine Younis[1,2]; Inga Sophia Knoth[1,2]; Martineau Jean-Louis[1,2]; Zohra Saci[1,2]; Maude Auger[1,2]; Frédérique Tihy[1,2]; Géraldine Mathonnet[1,2]; Catalina Maftei[1,2]; France Léveillé[1,2]; David Porteous[7,8,9], Gail Davies[7], Paul Redmond[7], Sarah E. Harris[7], W. David Hill[7], Emmanuelle Lemyre[1,2]; Gunter Schumann[10]; Thomas Bourgeron[11,12,13]; Zdenka Pausova[14]; Tomas Paus[15,16,17]; Sherif Karama[18,19,20]; Sarah Lippe[2,21]; Ian J. Deary[7]; Laura Almasy[22]; Aurélie Labbe[4]; David Glahn[23]; Celia M.T. Greenwood [3,24]; Sébastien Jacquemont[1,2]

** Shared first authorship

1 Department of Pediatrics, Université de Montréal, Montreal, Quebec, Canada
2 Center Hospitalier Universitaire Sainte-Justine Research Center, Montreal, Quebec, Canada
3 Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada
4 Département de Sciences de la Décision, HEC Montreal, Montreal, Quebec, Canada
5 Human Genetics and Cognitive Functions, University Paris Diderot, Sorbonne Paris Cité, Paris, France
6 Universite de Rouen Normandie, UFR des Sciences et Techniques, Rouen, France
7 Lothian Birth Cohorts group, Department of Psychology, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Edinburgh, EH8 9JZ, UK.
8 Medical Genetics Section, Centre for Genomic & Experimental Medicine, MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, UK.
9 Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.
10 Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, England
11 Department of Neurosciences, Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France
12 Centre National de la Recherche Scientifique Genes, Synapses and Cognition Laboratory, Institut Pasteur, Paris, France
13 Human Genetics and Cognitive Functions, University Paris Diderot, Sorbonne Paris Cité, Paris, France
14 The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada
15 Rotman Research Institute, Baycrest, Toronto, Ontario, Canada
16 Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, Canada
17 Child Mind Institute, New York, New York
18 Montreal Neurological Institute, McGill University, Montreal, QC, Canada.
19 McConnell Brain Imaging Center, McGill University, Montreal, QC, Canada
20 Douglas Mental Health University Institute, Montreal, QC, Canada
21 Psychology, Université de Montréal, Montreal, QC, Canada.
22 Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA
23 Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; Olin Neuropsychiatric Research Center, Institute of Living, Hartford Hospital, Hartford, CT, USA.
24 Gerald Bronfman Department of Oncology, Departments of Epidemiology, Biostatistics & Occupational Health and Human Genetics, McGill University, Montreal, Quebec, Canada

Corresponding authors:

Guillaume Huguet
Sainte Justine University Hospital
3175 chemin de la Côte-Sainte-Catherine,
Montréal, QC H3T 1C5
guillaumeaf.huguet@gmail.com

Sébastien Jacquemont
Sainte Justine University Hospital
3175 chemin de la Côte-Sainte-Catherine,
Montréal, QC H3T 1C5
sebastien.jacquemont@umontreal.ca

ABSTRACT

Rare genomic Copy Number Variants (CNVs) are major contributors to neurodevelopmental disorder. The vast majority of pathogenic CNVs reported back to patients are ultra-rare and their quantitative effects on traits such as intelligence are undocumented.

Here, we identified all CNVs ≥ 50 kilobase in 24,092 individuals from unselected and autism cohorts. We developed statistical models to estimate the effect-size of CNVs on intelligence based on their coding and non-coding characteristics.

Measures of intolerance to haploinsufficiency best explained the effect of any deletion or duplication on general intelligence. There was no heterogeneity across unselected and autism cohorts. Validation was performed using an intraclass concordance and showed that model estimates of general intelligence were 78% accurate with mean effect-sizes previously published for 47 CNVs.

Inheritance data on 27,766 CNVs showed that deletions and duplications with the same large effect-size on intelligence occur *de novo* at the same frequency.

Our first outline for the effect sizes of all coding genes on intelligence suggests that around 10,000 genes affect this trait.

Introduction

Copy Number Variants (CNVs) are deletions or duplications larger than 1000 base pairs. The contribution of CNVs to the etiology of intellectual disability (ID)[1–3], autism[4–6] and schizophrenia[6–8] is well established. CNVs studies have mainly been conducted using a case-control approach testing association with a categorical diagnosis[1,9,10]. As a result, 1) only the most recurrent CNVs are documented and 2) the interpretation of CNVs in research and medical diagnostics remains essentially binary: benign or pathogenic (contributing to mental illness). Their characteristics such as size, genes they encompass and *de novo* status are also used to classify them as benign or pathogenic[11]. The routine implementation of Chromosomal Micro-Arrays (CMAs) as a first-tier diagnostic test currently identifies "pathogenic" CNVs in 10 to 15 % of children with neurodevelopmental disorders (NDD)[12]. A binary interpretation is however of limited use in the clinic where patients present a broad spectrum of cognitive symptoms ranging from severe ID to mild learning disabilities, or autism with IQ above population mean. The quantitative effects of CNVs are rarely documented even for commonly available cognitive dimensions such as general intelligence. For the most recurrent CNVs, this data may be available but it originates from patients ascertained in the clinic leading to potentially gross overestimation of effect-sizes. One study in an unselected population from Iceland showed that in aggregate, 12 recurrent CNVs associated with schizophrenia decrease IQ by 15 points (1 z-score)[13]. A second study in UK Biobank (UKBB) showed that 24 out of 33 recurrent CNVs were associated with reduced performance on at least one cognitive test. Schizophrenia-associated CNVs were associated with larger impairments[13,14]. However, the vast majority of pathogenic CNVs reported back to patients are ultra-rare (non-recurrent) and there is no data to quantify their effect-size on cognitive function.

There is significant genetic correlation between intelligence and psychiatric disorders and impairments in intelligence represent a major referral criterion to the NDD clinic. The

heritability of general intelligence is estimated at around 50 to 80%[15], with the former estimate representing all age groups, and the latter referring to adults only. The heritability of variants in linkage disequilibrium with common SNPs is estimated to be around 22.7%, with variants in poor linkage disequilibrium with SNPs, including rare CNVs, explaining 31.3% of the phenotypic variation in intelligence[16]. Two recent GWAS, has identified over 200 loci associated with intelligence and education[17,18] and these loci potentially implicate 1000 candidate genes. The latter were largely non-overlapping with genes previously linked to ID or other NDD[17]. Contrary to the majority of non-coding SNPs studied in GWAS, there is no ambiguity in the molecular interpretation of a fully deleted or duplicated gene, which invariably decreases or increases transcription respectively. Therefore, CNVs represent a powerful tool to map the effect-sizes of genes (altered by gene dosage) on human traits.

We propose a framework to estimate and predict the effect on intelligence of any CNVs including undocumented ones. We hypothesize that the effect-size of CNVs can be statistically estimated based on coding and non-coding characteristics of the affected genomic region.

We have previously shown that linear models[19] using the sum of the "probability of being loss-of-function intolerant" (pLI) scores[20] of all genes included in a deletion can predict their effect-size on intelligence quotient (IQ) with a concordance of 75% with empirical measures. Our initial study, performed in 2 unselected populations, was not sufficiently powered to measure the effects size of duplications and investigate the effect of alteration of individual genes. It was also unknown whether results would generalize to patients with NDD referred to the clinic. The pLI score (ranging from 1 to 0) is a measure of a gene's intolerance to haploinsufficiency based on lower-than-expected rates of protein-loss-of-function (pLoF) variants in the general population. Genes with high pLI scores (≥0.9) are intolerant to pLoF, whereas genes with low pLI scores (≤0.1) are tolerant or there may be insufficient data to

assess their tolerance. However, the pLI is mainly a binary variable and new continuous measures such as the LOEUF[21] (Loss-of-function Observed/Expected Upper bound Fraction) may better reflect the full spectrum of intolerance to pLoF. LOEUF range from 0 to 2, and values below 0.35 are suggestive of intolerance.

Our present aims are 1) to replicate effect sizes estimates of deletions on general intelligence, in a large dataset including unselected and NDD populations, 2) to estimate the effect size on general intelligence of genomic duplications, 3) to estimate effect sizes of all individual protein-coding genes on intelligence (using categories of constraint scores), and 4) to investigate the quantitative relationship between effect size on general intelligence and *de novo* events.

We identified and annotated CNVs in 24,092 individuals from five general populations and two autism cohorts. We then scored CNVs using genetic annotations to identify variables that contribute the most to variation in general intelligence. We also investigated the inheritance of 27,766 CNVs in unselected populations, autism, and clinical samples. After model validation, we implemented an online tool to help clinicians and researchers estimate the effect size of any CNVs on general intelligence.

RESULTS

1) The effect-size of haploinsufficiency on measures of general intelligence is similar regardless of assessment and ascertainment.

We identified 18,860 autosomal deletions and 18,799 autosomal duplications larger than 50 kb in 24,092 individuals. General intelligence was assessed using different neurocognitive tests across cohorts: measures of non-verbal intelligence quotient (NVIQ) were available in five samples and general intelligence factor (g-factor)[22] was computed in three samples using cognitive tests, primarily assessing fluid non-verbal reasoning (Table1, Supplementary figure 1). The concordance between z-scored NVIQ and g-factor available for three cohorts ranged from 60 to 77% (Supplementary table 1).

We sought to replicate our previous estimates for the effect-size of deletions (measured by pLI) on general intelligence[19]. We computed the sum of pLI of all coding genes with all transcripts fully encompassed within CNVs. This sum was used as the main explanatory variable in a linear model estimating the effect-size of CNVs on general intelligence. The meta-analysis performed on 20,151 individuals from unselected populations showed that one deleted point of pLI decreases NVIQ or g-factor by 0.18 z-score (95% CI: -0.23 to -0.14, which is equivalent to 2.7 points of NVIQ, Figure 1a, Supplementary table 2).

Since genomic variants with large effects on general intelligence (e.g. deletions occurring *de novo*) are thought to be removed from the general population as a result of negative selective pressure. This may have led to an underestimation of the effect size of CNVs in unselected populations. To examine this possibility, we analyzed 3,941 individuals from two autism cohorts, which include individuals with ID and *de novo* deletions. The effect size on general intelligence of intolerance to pLoF measured by pLI was the same (-0.17 z-score, 95% CI: -0.22 to -0.12) and we did not observe heterogeneity across autism and unselected populations (Figure 1A, supplemental table 2). Of note, the effect size of pLI was essentially

identical for NVIQ and g-factor, which were both available in three cohorts (Supplementary table 3).

2) The effect size of duplications on general intelligence is 3-fold smaller than deletions

Our previous study[15] was unable to estimate the effect size of duplications on general intelligence, likely due to insufficient power. We fit a linear model using the sum of pLI of all coding genes encompassed in duplications per individual as the explanatory variable of general intelligence. The meta-analysis combining all samples shows a decrease of 0.05 z-score (95% CI: -0.07 to -0.04), of general intelligence per duplicated point of pLI which is equivalent to 0.75 points of general intelligence. Again, estimates are similar across autism and unselected populations. (Figure 1b, Supplemental table 2).

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

Given the homogeneity across cohorts, we pooled samples after adjusting for specific nuisance variables (including cognitive test and cohorts, Supplementary methods) to perform a mega-analysis of 24,092 individuals. We selected 13,001 deletions and 15,856 duplications that could potentially be detected by all arrays used in the mega-analysis (encompassing ≥10 probes for any array). These CNVs encompassed 36% ($N_{CNVs\ gene}$=6,315, $N_{Del.\ gene}$=2,282, $N_{Dup.\ gene}$=5,223) of the coding genome (Figure 2a, Supplementary Figure 2a). The mean effect of one point of pLI on general intelligence corresponded to an estimated decrease of 0.175 z-score (SE=0.016, P=1.45×10^{-28}) and 0.054 z-score (SE=0.009, P=1.90×10^{-9}) for deletions and duplications, respectively (Supplementary table 4). All analyses only accounted for genes with all transcripts fully encompassed in CNVs. Including different categories of

partially deleted or duplicated genes did not improve the model (Supplementary methods, Supplementary table 5 to 7).

Comparing 11 variables, pLI and 1/LOEUF best explained (based on AIC) the variance of general intelligence (Supplementary table 4). There was no effect of the interaction between age or sex and either constraint scores on general intelligence (Supplementary Table 8 to 11). Non-linear models including polynomial function of order 2 and a kernel method did not improve model fit (Supplementary Table 12 and 13), suggesting an additive effect of constraint scores. For the remainder of the study, we transitioned to using LOEUF because it is a continuous variable (the pLI is essentially binary) and is now recommended as the primary constraint score by gnomAD. Analyses using pLI are presented in supplemental results.

4) Genes encompassed in recurrent neuropsychiatric CNVs and non-recurrent CNVs have the same effect sizes on intelligence

We asked if the mean effect-size of 1/LOEUF is influenced by recurrent neuropsychiatric CNVs. Therefore, we removed 608 individuals carrying any of the 121 recurrent CNV previously associated with neuropsychiatric conditions[19]. This sensitivity analyses demonstrate that the effect size of 1/LOEUF on general intelligence remains the same and highly significant for deletions and duplications (Supplementary table 14). It has been posited that the deleteriousness of large psychiatric CNVs may be due to interactions between genes encompassed in CNVs. We therefore asked if the effect-size of 1/LOEUF is the same for CNVs encompassing small and large numbers of genes. We recomputed the linear model 6 times after incrementally excluding individuals with a total sum of 1/LOEUF ≥60, 40, 20, 10, 4 and 2.85 (2.85 corresponds to 1/0.35, the cut-off for intolerance to pLoF gnomAD) for deletions and duplications separately. The effect of each unit of 1/LOEUF remains similar

whether deletions encompass >10 or >60 points of 1/LOEUF (Figure 2b, results are similar with pLI Supplementary figure 2b).

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

Our ability to capture extreme effect size genes is likely limited by the properties of LOEUF. Intolerance can be robustly inferred only if the number of expected pLOF variants (size of the coding region) is large enough. To improve model accuracy for extreme and large effect size genes, we used a list of 256 ID-genes[2,23], previously identified with an excess of *de novo* mutations in NDD cohorts. We identified 66 CNVs encompassing at least one ID-gene in ASD cohorts (31 deletions and 35 duplications) and 60 in the general population (13 deletions and 47 duplications) (Figure 3, supplementary methods).

We recomputed the model by integrating 4 explanatory variables: the sum of 1/LOEUF for ID and non-ID-genes encompassed in deletions and duplications. The mean effect-size on intelligence of 1 point of 1/LOEUF for ID-genes was a decrease of 0.174 z-score in deletions (SE=0.035; $P= 9.2\times10^{-7}$) and 0.076 in duplications (SE=0.026; $P= 3.7\times10^{-3}$), 7 to 11-fold higher than the mean effect of 1/LOEUF for non-ID genes (Supplementary table 15). As an illustration, the mean effect on IQ of ID-genes intolerant to pLoF (LOEUF<0.35) was 20 (ranging from 8 to 70) for deletions and 9 (from 3 to 31) for duplications (Supplementary table 16). The effect of 1/LOEUF for non-ID-genes remained unchanged (Supplementary table 15). Conclusions are the same using pLI (Supplementary table 15 and 16, Supplementary figure 3).

6) Replicating model estimates in the neurodevelopmental disorder clinic

We asked if the estimates of our model can generalize to cases (282 individuals from 75 families) carrying large deleterious CNVs ascertained in the neurodevelopmental disorder clinic and rarely observed in the general population or in autism cohorts (Table 1). Using a kinship matrix, IQ measured in relatives with or without a CNV provided information to account for genetic and environmental background present in families. Results are very similar to those observed in the mega-analysis for deletions (estimate $_{1/LOEUF}$ = -0.024, P= 5.3×10^{-8}) and duplications (estimate $_{1/LOEUF}$ =-0.009, P=1.8×10^{-3}, Supplementary table 17). There were too few observations in this dataset to provide significant estimates for ID-genes.

7) Model validation using clinical series and the UK Biobank.

To investigate the performance of our model for non-pathogenic and pathogenic recurrent CNV we compared model estimates to published observations. We identified a total of 47 recurrent CNVs with sufficient data on general intelligence reported in clinical series and in the UKBB[17] (Supplementary table 18). When cognitive data was available from both clinical and the UKBB (13 of the 47 recurrent CNVs), we used the mean of both effect sizes (Figure 4). Concordance between the 1/LOEUF-ID-gene model and previously published measures was 0.78 for all deletions and duplications (95% CI, 0.66-0.86, P= 4.3×10^{-11}, Figure 4). Accuracy was similar for deletions (ICC=0.71 [0.5;0.84], P= 1.8×10^{-5}, Figure 4) and duplications (ICC=0.85 [0.7;0.93], P= 3×10^{-7}, Figure 4) as well as for small and extreme effect size CNVs including trisomy 21. The absence of any significant concordance between effect-sizes reported in UKBB and those reported in clinical case series highlighted the specific biases of these different ascertainment methods (Supplementary figure 4). pLI models provided the same level of concordance (Supplementary figure 5).

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

Because models including constraint scores explain equally well the effect sizes of deletions and duplications on intelligence, we investigated if the relationship between intelligence and *de novo* frequency is similar for both types of CNVs. We established inheritance for 26,437 CNVs in 6 cohorts (Supplemental Table 19). There was a strong relationship between 1/LOEUF for ID and non-ID genes and the frequency of *de novo* observations for deletions (P=1.9×10^{-65} and P <10^{-314} respectively) and duplications (P=4.6×10^{-24} and P=2.6×10^{-182} respectively, Figure 5a).

Deletions and duplications with the same impact on general intelligence show the same *de novo* frequency for large effect size CNVs (Figure 5a). CNVs with neutral estimated effects on general intelligence showed a much higher rate of *de novo* frequency in diagnostic databases (6.52 and 3.27% for deletions and duplications) compared to unselected populations (0.58 and 1.19%, Supplementary table 19). The concordance between the probability of occurring *de novo* estimated by the model (after removing recurrent CNVs) and *de novo* frequency reported in the DECIPHER database on 31 recurrent CNVs was 0.81 ([0.67-0.9]; P=8.2×10^{-8}) (Figure 5b, Supplementary table 20, Supplemental figure 6).

9) Estimating effect-sizes of individual genes using LOEUF

Although models relying on the sum of pLI or 1/LOEUF provide accurate estimates of CNV effects on general intelligence, it is unlikely that they can properly estimate the effects of individual genes. Since we were vastly underpowered to perform a gene-based GWAS, we first divided all genes in 4 categories: highly intolerant genes (LOEUF<0.2; n=980), moderately intolerant genes (0.2≤LOEUF<0.35 n=1,762), tolerant genes (0.35≤LOEUF<1; n=7,442) and highly tolerant genes (LOEUF≥1; n=8,267). The sum of genes in each category was used as four explanatory variables to explain general intelligence in the same linear

model. For deletions, the three first variables showed negative effects on general intelligence (Figure 6, Supplementary table 21). For duplications only moderately intolerant genes showed negative effects (Supplementary figure 7 and table 21).

We were underpowered to further subdivide these LOEUF categories, so as an exploratory analysis, we tested 38 overlapping sliding windows across LOEUF values in 38 linear models. Each model used 2 explanatory variables: number of genes within and outside the window (size = 0.15 LOEUF). Negative effects of deletions on general intelligence were observed for genes within 13 out of 38 windows and 2 showed positive effects (after corrected for FDR). For duplications, only 2 windows had negative effects (after corrected for FDR, Figure 6, Supplementary figure 7 and table 22).

DISCUSSION

Main findings

The mean effect size associated with deleting one point of pLI is robust across cohorts, clinical diagnoses, general intelligence assessments, a broad age range and sex. It is similar for CNVs previously associated with psychiatric disorders and undocumented CNVs. The same robustness of effect size is also observed for duplications, with an effect size which is 3-fold smaller than deletions. The linear sum of pLI or 1/LOEUF predicted the effect size of deletions and duplication on intelligence with over 78% of concordance with empirical measures. Using categories of LOEUF values, we provide the first estimates for the individual effect sizes of protein-coding genes. Results suggest that 1) the effect sizes of genes intolerant to pLoF increase sharply as their LOEUF values decrease and 2) over 7000 genes considered as tolerant to pLoF also impact cognitive ability with mild effect sizes.

Intolerance to altered gene dosage rather than just haploinsufficiency

We show that scores measuring intolerance to pLoF explain equally the effect-size of deletions and duplications on intelligence. pLI and LOEUF may therefore be considered as measures of intolerance to altered gene dosage, regardless of whether gene expression is increased or decreased. This is consistent with observation of structural variants in gnomAD[24]. The 3:1 effect-size ratio between deletions and duplications may help estimate the effect-size of a duplicated gene when only information on its pLoF effect-size is available or *vice versa*.

Model validation and ascertainment biases

Predictions from linear models using constraint scores show 78% of concordance with empirical data in estimation of the effect size of CNVs on IQ. Several CNVs show large discordances between model estimates and previous reports from the literature. This may be due to either 1) unidentified large effect size genes with unreliable LOEUF measures due to the small size of the protein coding region, and 2) ascertainment bias which is illustrated by the model's trend toward underestimation and overestimation when compared to clinical and UKBB data respectively (Figure 4). However, biases from clinically referred individuals can be adjusted for using intrafamilial controls to account for additional factors present in the family[25,26]. This is confirmed by our results using the Ste Justine family genetic cohort which demonstrated effect-size estimates consistent with the other cohorts of the mega-analysis. Our results suggest that the effect size of pathogenic CNVs are underestimated in the UKBB[25] while those of small CNVs are largely overestimated in clinical series. Of note, the maximum effect size measured in UKBB was 0.4 z-scores including pathogenic CNVs such as 16p11.2 deletions and CNVs containing an ID-gene (*WDFY4*) and genes highly intolerant to pLoF: *SNRNP200, SEMA4C, KANSL3, ARID5A*. On the other hand, mild effect size variants such as

the 16p13.11 duplications and 1q21.1 CNVs are likely overestimated in clinical series[27]. Currently, in the absence of unbiased CNV, cognitive and behavioral data collected in a systematic way[28], statistical models using a variety of disease and unselected cohorts are likely to better account for biases and provide more accurate estimates.

Individual effect of genes

We propose a first interpretation of the effect size of individual genes based on categories of LOEUF values and groups of genes with an excess of *de novo* variants in NDD (ID-genes). Deleting or duplicating an ID-gene (1.3% of the coding genome) leads to a mean decrease of 20 and 9 IQ points respectively. Other than this small group of ID-genes, the other genes intolerant to pLoF (15%) have effect sizes far below 1 SD (11 and 3 IQ points for deletions and duplications respectively). Given the sharp decrease in effect-size related to increasing LOEUF values, we speculate that it is unlikely that large numbers of additional genes with extreme effects on intelligence (when deleted or duplicated) remain to be identified.
Based on our observed relationship between IQ and *de-novo* frequency, as well as our model, we speculate that new candidate ID-genes will likely have moderate effect sizes (far below 1SD), which is important to acknowledge in the clinic or when designing animal models. Results based on categories of LOEUF values suggest that a large group of coding genes tolerant to pLoF (40%) also impact intelligence when deleted. A larger sample size will improve the characterization of these groups.

Potential clinical application

Models developed in this study provide a translation of gnomAD constraint scores into effect sizes on cognitive abilities. Model outputs are implemented in a prediction tool (https://cnvprediction.urca.ca/), which is designed to estimate the marginal or population-

average effect-size of any given CNV on general intelligence, not the cognitive ability of the individual who carries the CNV. If the cognitive deficits of an individual are concordant with the effect size of the CNV they carry, one may conclude that the CNV contributes substantially to those deficits. When discordant, the clinician may conclude that most of the contribution lies in additional factors which should be investigated. This tool can assist in the interpretation of undocumented CNVs.

The relationship between genetic fitness and cognitive abilities

The reasons underlying the tight relationship between general intelligence and epidemiological measures of intolerance to pLoF, is unclear. Behavioral interpretations are intuitive for severe ID but do not apply for much milder effects: altering gene dosage of one intolerant gene may decrease NVIQ by only a few points, but nevertheless increases *de novo* frequency. This relationship is further highlighted by the fact that deletions and duplications with the similar impact on intelligence occur *de novo* with similar frequencies. Larger samples are required to investigate whether some genes that are not under constraint may affect IQ. Variables unrelated to genetic fitness but relevant to brain function, such as the topology of gene expression in the brain[29], may shed light on such genes.

Limitations

Constraint scores are limited by the number and quality of pLoF variants observed in the gnomAD database and are unreliable for small coding genes. We were not able to observe CNVs encompassing genes with very large effects in our unselected and autism cohorts. Our estimates for ID genes may, therefore, reflect a less severe subgroup and model outputs should be interpreted with caution when CNVs encompass ID-genes with very low LOEUF values. LOEUF and pLI are epidemiological measures of genetic fitness in human

populations, without any consideration of gene function[20,21]. It is likely that some genes decrease fitness (eg. genes involved in fertility) without affecting general intelligence. Further study combining intolerance scores with gene ontology categories are required to investigate this question. Models were limited to genes fully encompassed in CNV because we were unable to identify effects of genes partially deleted or duplicated. This was likely due to inaccurate CNV coordinates related to array resolution as well as a lack of power. Models were trained on CNVs encompassing at least once, 36% of the coding genome. Projections suggest that 500K individuals from an unselected population would increase coverage to 78% (Supplementary figure 8). Heterogeneous measures of general intelligence across the different cohorts may have introduced noise, affecting our estimates. We were also underpowered to analyze CNVs on the X-chromosome harbouring only a few CNVs.

Finally, all models imply additive effects and massive datasets will be required to test if CNVs have different effects sizes depending on genetic or environmental factors. However, the fact that very large CNVs (such as trisomy 21) are accurately estimated by the model suggests that genetic interactions within these large genomic segments or even chromosomes cannot be readily observed.

Conclusions

The effect size of deletions or duplications on intelligence can be accurately estimated with additive models using genetic constraint scores. Accuracy across a broad range of CNVs suggests that the same principles of gene dosage apply to small benign CNVs as well as extreme CNVs such as Down syndrome. We provide a map of effect sizes at the individual gene level but to move beyond this rough outline, much larger sample sizes are required. Nonetheless, these results suggest that a large proportion (56%), if not the entire genome, influences cognitive abilities. One may therefore view the genetic contribution to cognitive

difference as an emergent property of the entire genome not restricted to a set of variants affecting a limited number of biological pathways.

Acknowledgments

Funding/Support: This research was enabled by support provided by Calcul Quebec (http://www.calculquebec.ca) and Compute Canada (http://www.computecanada.ca). Sebastien Jacquemont is a recipient of a Bursary Professor fellowship of the Swiss National Science Foundation, a Canada Research Chair in neurodevelopmental disorders, and a chair from the Jeanne et Jean Louis Levesque Foundation. Catherine Schramm is supported by an Institute for Data Valorization (IVADO) fellowship. Petra Tamer is supported by a Canadian Institute of Health Research (CIHR) Scholarship Program. Guillaume Huguet is supported by the Sainte-Justine Foundation, the Merit Scholarship Program for foreign students, and the Network of Applied Genetic Medicine fellowships. Thomas Bourgeron is a recipient of a chair of the Bettencourt-Schueler foundation. This work is supported by a grant from the Brain Canada Multi-Investigator initiative and CIHR grant 159734 (Sebastien Jacquemont, Celia Greenwood, Tomas Paus). The Canadian Institutes of Health Research and the Heart and Stroke Foundation of Canada fund the Saguenay Youth Study (SYS). SYS was funded by the Canadian Institutes of Health Research (Tomas Paus, Zdenka Pausova) and the Heart and Stroke Foundation of Canada (Zdenka Pausova). Funding for the project was provided by the Wellcome Trust. This work was also supported by an NIH award U01 MH119690 granted to Laura Almasy, Sebastien Jacquemont and David Glahn and U01 MH119739. The authors gratefully acknowledge the resources provided by the Autism Speaks MSSNG project and the Autism Genetic Resource Exchange Consortium, as well as the participating families. We are grateful to all the families who participated in the Simons Variation in Individuals Project (VIP) and the Simons VIP Consortium (data from Simons VIP are available through SFARI Base). We thank the coordinators and staff at the Simons VIP and SCC sites. We are grateful to all of the families at the participating SSC sites and the principal investigators (A. Beaudet,

M.D., R. Bernier, Ph.D., J. Constantino, M.D., E. Cook, M.D., E. Fombonne, M.D., D. Geschwind, M.D., Ph.D., R. Goin-Kochel, Ph.D., E. Hanson, Ph.D., D. Grice, M.D., A. Klin, Ph.D., D. Ledbetter, Ph.D., C. Lord, Ph.D., C. Martin, Ph.D., D. Martin, M.D., Ph.D., R. Maxim, M.D., J. Miles, M.D., Ph.D., O. Ousley, Ph.D., K. Pelphrey, Ph.D., B. Peterson, M.D., J. Piggot, M.D., C. Saulnier, Ph.D., M. State, M.D., Ph.D., W. Stone, Ph.D., J. Sutcliffe, Ph.D., C. Walsh, M.D., Ph.D., Z. Warren, Ph.D., and E. Wijsman, Ph.D.). We appreciate obtaining access to phenotypic data on SFARI base.

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

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

Tables and Figures

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**	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	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	HOE-12V	602	321 (53%)	49.5 (4.9)	g-factor, 12 cognitive measures‡	0 (1)
	LBC1936	610Kq	504	247 (49%)	70.0 (-)*	Moray House Test (and g-factor)	0.05 (0.96) ***
	CaG-GSA	GSA	2,074	1,094 (53%)	54.3 (7.6)	g-factor, Reasoning, Memory, Reaction time	-0.02 (1.03)
	CaG-Omni2.5	Omni2.5	515	281 (55%)	52.4 (8.6)		-0.10 (1.02)
	CaG (all)	GSA; Omni2.5	2,589	1,375 (53%)	53.9 (7.8)		-0.03 (1.03)
	G-Scot	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**	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	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	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)	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	WGS	1,379	275 (20%)	9.2 (4.4)	WISC-IV n=46; WASI-II n=338; Leiter n=372; Raven n=214; Standford Binet n=281; WPPSI n=128	-0.47 (1.58)
NDD** (n=282)	**Ste-Justine-probands**	Agilent 180 K array	75	29 (35%)	7.23 (4.46)	WISC-V n=25; WASI-II n=5; WPPSI-IV n=23; Leiter-R n=11; Mullen n=19	-1.34 (0.96)
	Ste-Justine-siblings		37	17 (46%)	10.06 (6.62)	WISC-V n=12; WASI-II n=9; WPPSI-IV n=11; Leiter-R n=2; Mullen n=3	-0.26 (1.06)
	Ste-Justine-parents		170	100 (59%)	37.72 (6.88)	WASI-II	-0.12 (1.13)

Table 1: Cohorts description

Cohorts include 24,092 individuals, including 14,874 unrelated individuals. SSC and CaG cohorts were broken down into sub-samples based on array technology (Supplementary methods). †63 and ‡ 12 cognitive measures were respectively used to compute the g-factor in SYS children and parents (Supplementary methods). NDD: neurodevelopmental disorders, SYS: Saguenay Youth Study, CaG: CARTaGEN, LBC: Lothian Birth Cohort, SSC: Simons Simplex Collection; n=number of individuals remaining for analysis after quality control. The

mean and Standard Deviation (SD) for g-factor slightly deviate from 0 and 1 in some cohorts since it was computed on all available data (before the exclusion of some individuals for poor quality array) and summarized here only for individuals included in the analyses. *All individuals from LBC1936 were assessed at 70 years old explaining the absence of SD computation. **The NDD cohort was used only in the replication analysis and was not included in meta- or mega-analyses. *** We displayed the Z-scored of IQ, because IQ was preferred to g-factor for all analyses, even if results were similar (Supplementary table 1 and 3).

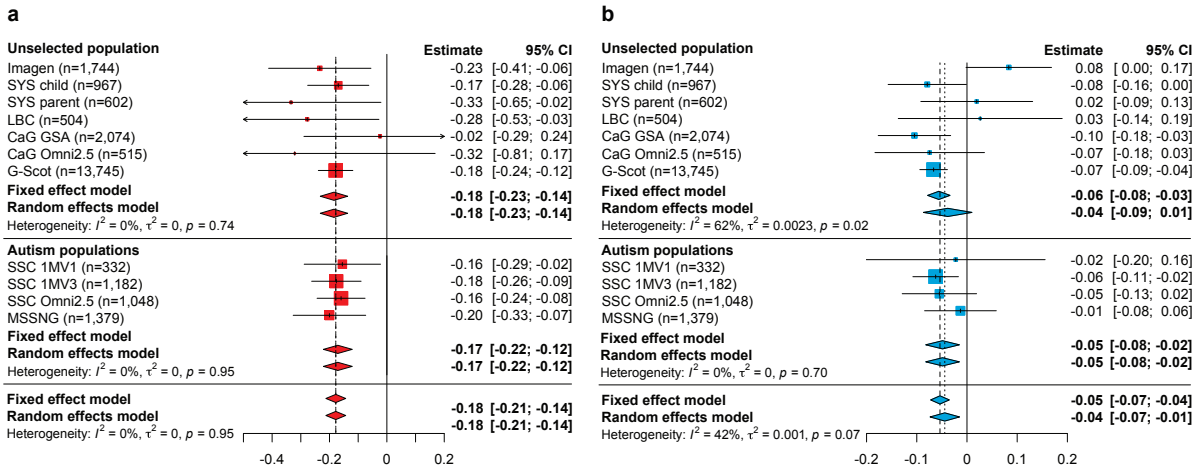

Figure 1: Effect of pLI on general intelligence measured for deletions and duplications. Meta-analysis estimating the effect of deletions **a.** and duplications **b.,** measured by sum of pLI, on general intelligence (Supplementary table 2). X-axis values represent z-scores of general intelligence. Deleting one point of pLI decreases the general intelligence by 0.18 z-scores (2.7 points of IQ). Duplicating one point of pLI decreases the general intelligence by 0.05 z-scores (0.75 points of IQ). The squares represent the effect size computed for each sample. Their size negatively correlated to variance. Diamonds represent the summary effect across cohorts. Their lengths correspond to the 95% confidence intervals of the mean effect size.

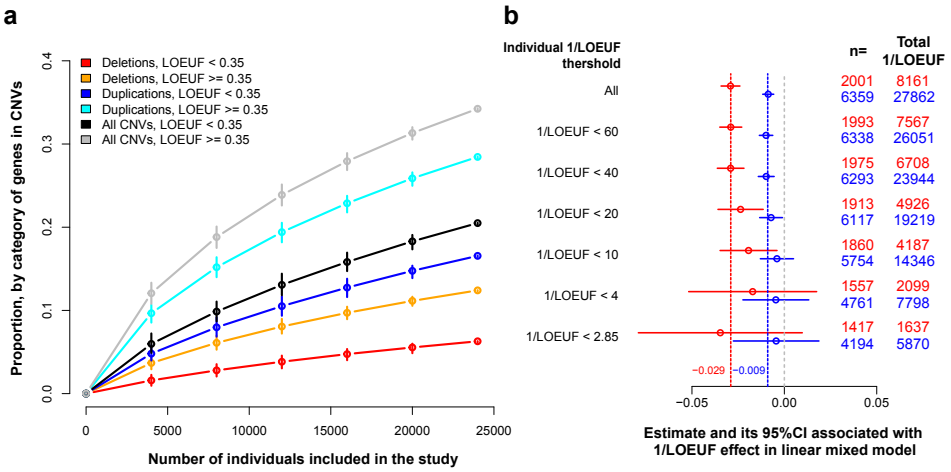

Figure 2: Sensitivity analyses for models based on 1/LOEUF score.

a. Estimated proportion of the coding genome within each category defined by LOEUF, encompassed in CNVs present in the mega-analysis according to sample size (randomly selected within the mega-analysis). **b.** Estimated effect of 1/LOEUF on general intelligence after removing individuals with a sum of 1/LOEUF larger than 60, 40, 20, 10, 4 and 2.85. n: number of individuals with a total sum of 1/LOEUF > 0.

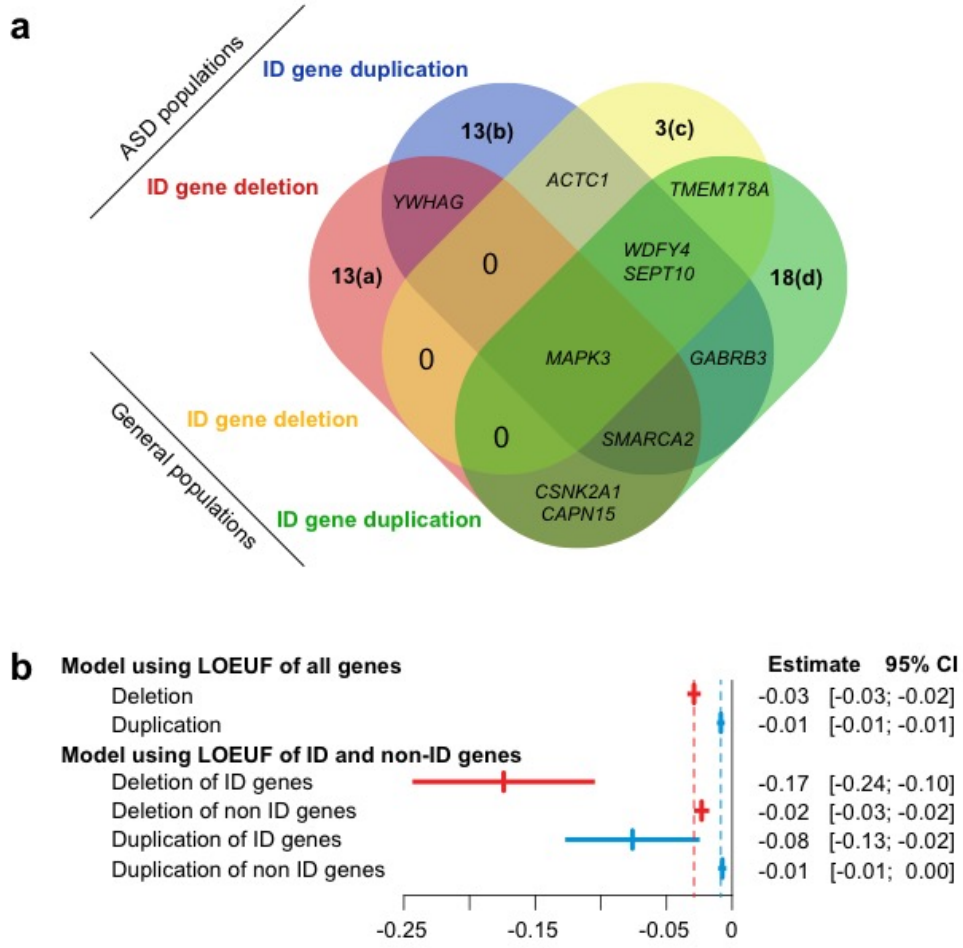

Figure 3: Effect size of intellectual disability (ID) genes on general intelligence.

a. Venn diagram of ID genes in ASD and in general population cohorts. Genes were previously defined as harboring an excess of *de novo* loss of function (bold) or missense mutations in neurodevelopmental cohorts: (a) *DYNC1H1,* ***PHF21A, SHANK3, TRA2B, FOXP1, SETD5,*** *NR4A2,* ***TCF7L2, SOX5, POU3F3, ARID1B, EBF3, HNRNPU***; (b) ***SET, ZBTB18, DLG4, CHAMP1, CNOT3****, U2AF2, HIST1H2AC, DNM1,* ***RAI1, CREBBP, HIST1H1E, ASXL1****, CABP7*; (c) *PRPF18, PPP2R1A, EEF1A2*; (d) *TRAF7, DEAF1, STC1,* ***MYT1L, BRPF1****, CBL,* ***SPAST, WDR87, NFE2L3, STARD9, TCF20, KMT2C, FAM200B, KDM5B, CHD2,*** *BTF3, ITPR1, HMGXB3*. **b.** Effect-size of 1/LOEUF on general intelligence estimated in a model using two explanatory variables (sum of 1/LOEUF of deleted and dupliacted genes) or 4 explanatory variables (sum of 1/LOEUF of ID genes and non-ID genes for deletions and duplication).

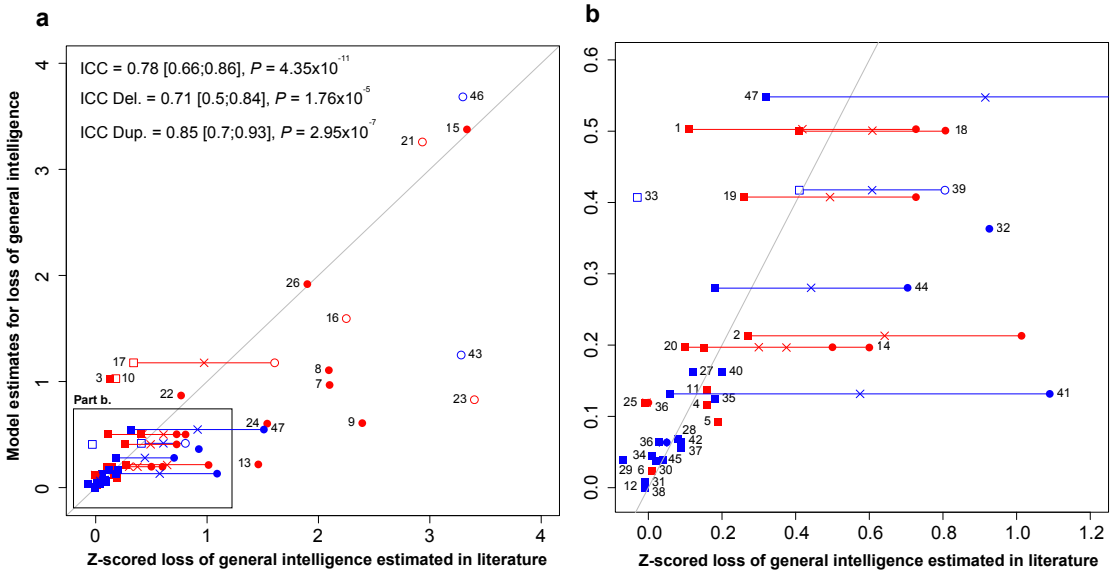

Figure 4: Concordance between model predictions and published observations for CNV effects on general intelligence.

a. and **b.** Concordance between model estimates (with 1/LOEUF and ID-genes) and literature of clinical data and UKBB reports for general intelligence loss observed in respectively 27 and 33 recurrent CNVs for a total of ascertained carriers of 47 recurrent CNVs (supplementary table 15). X- and Y-values: effect size of CNVs on z-scored general intelligence. **b.** Zoom of the rectangle drawn in the lower left section of panel **a**. We represented values from clinical data by a circle and those from UKBB data by a square. The cross represents the mean value of z-scored IQ loss for the 13 recurrent CNVs observed both in literature and in UKBB. Deletions are in red and duplications in blue. Empty circles or square are CNVs encompassing ID-genes. The model uses 2 explanatory variables (1/LOEUF of non-ID-genes and ID-genes). ICC indicates intraclass correlation coefficient (3, 1). Each point represents a recurrent CNV: (1) TAR Deletion; (2) 1q21.1 Deletion; (3) 2q11.2 Deletion; (4) 2q13 Deletion; (5) *NRXN1* Deletion; (6) 2q13 (*NPHP1*) Deletion; (7) 3q29 (*DLG1*) Deletion; (8) 7q11.23 (William-Beuren) Deletion; (9) 8p23.1 Deletion; (10) 10q11.21q11.23 Deletion; (11) 13q12.12 Deletion; (12) 13q12 (*CRYL1*) Deletion; (13)

15q13.3 (BP4-BP5) Deletion; (14) 15q11.2 Deletion; (15) 16p11.2-p12.2 Deletion; (16) 16p13.3 ATR-16 syndrome Deletion; (17) 16p11.2 Deletion; (18) 16p11.2 distal Deletion; (19) 16p13.11 Deletion; (20) 16p12.1 Deletion; (21) 17p11.2 (Smith-Magenis) Deletion; (22) 17q12 Deletion; (23) 17q21.31 Deletion; (24) NF1-microdeletion syndrome Deletion; (25) 17p12 (*HNPP*) Deletion; (26) 22q11.2 Deletion; (27) TAR Duplication; (28) 1q21.1 Duplication; (29) 2q21.1 Duplication; (30) 2q13 Duplication; (31) 2q13 (*NPHP1*) Duplication; (32) 7q11.23 Duplication; (33) 10q11.21q11.23 Duplication; (34) 13q12.12 Duplication; (35) 15q11q13 (BP3-BP4) Duplication; (36) 15q11.2 Duplication; (37) 15q13.3 Duplication; (38) 15q13.3 (*CHRNA7*) Duplication; (39) 16p11.2 Duplication; (40) 16p11.2 distal Duplication; (41) 16p13.11 Duplication; (42) 16p12.1 Duplication; (43) 17p11.2 Duplication; (44) 17q12 (*HNF1B*) Duplication; (45) 17p12 (*CMT1A*) Duplication; (46) Trisomic 21 Duplication; (47) 22q11.2 Duplication.

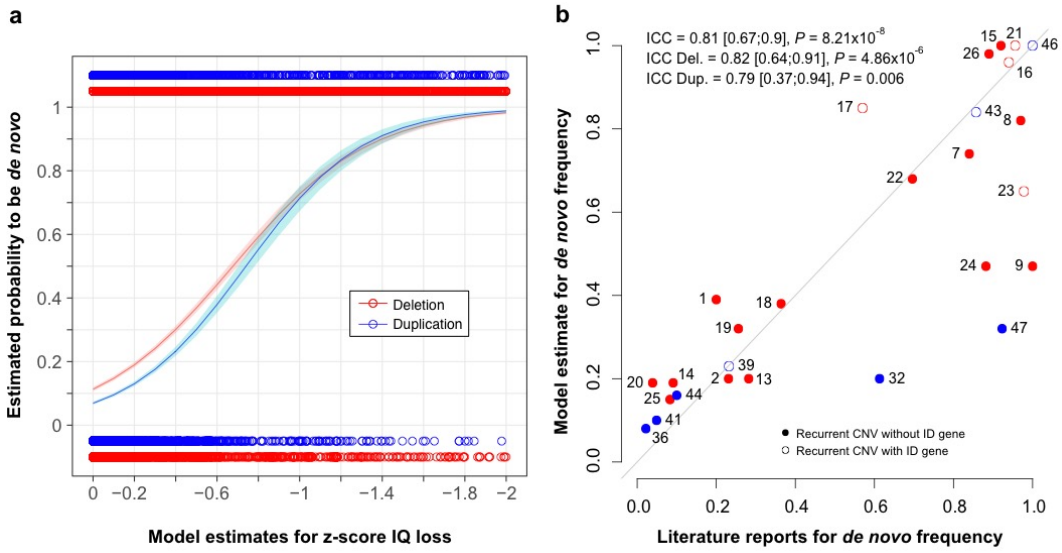

Figure 5: Concordance between estimates and literature reports for *de novo* frequency.

a. Probability of *de novo* estimated by our *de novo* model (Y-axis) according to the loss of IQ estimated by a model using 1/LOEUF for ID and non-ID genes as two explanatory variables (X-axis). The *de novo* model was fitted on 13,114 deletions (red) and 13,323 duplications (blue) with available inheritance information observed in DECIPHER, CHU Sainte-Justine, SSC, MSSNG, SYS and G-Scot. **b.** Concordance between *de novo* frequency observed in DECIPHER (X-axis) and the probability of being *de novo* estimated by models when excluding recurrent CNVs of the training dataset (Y-axis) 1/LOEUF for ID and non-ID genes as an explanatory variable for 27 recurrent CNVs. The first bisector represents the perfect concordance. ICC indicates intraclass correlation coefficient (3, 1). Each point corresponds to a known recurrent CNV: (1) TAR Deletion; (2) 1q21.1 Deletion; (7) 3q29 (*DLG1*) Deletion; (8) 7q11.23 (William-Beuren) Deletion; (9) 8p23.1 Deletion; (13) 15q13.3 (BP4-BP5) Deletion; (14) 15q11.2 Deletion; (15) 16p11.2-p12.2 Deletion; (16) 16p13.3 ATR-16 syndrome Deletion; (17) 16p11.2 Deletion; (18) 16p11.2 distal Deletion; (19) 16p13.11 Deletion; (20) 16p12.1 Deletion; (21) 17p11.2 (Smith-Magenis) Deletion; (22) 17q12 Deletion; (23) 17q21.31 Deletion; (24) NF1-microdeletion synd. Deletion; (25) 17p12 (*HNPP*) Deletion; (26) 22q11.2 Deletion; (32) 7q11.23 Duplication; (36) 15q11.2

Duplication; (39) 16p11.2 Duplication; (41) 16p13.11 Duplication; (43) 17p11.2 Duplication; (44) 17q12 (*HNF1B*) Duplication; (46) Trisomic 21 Duplication; (47) 22q11.2 Duplication.

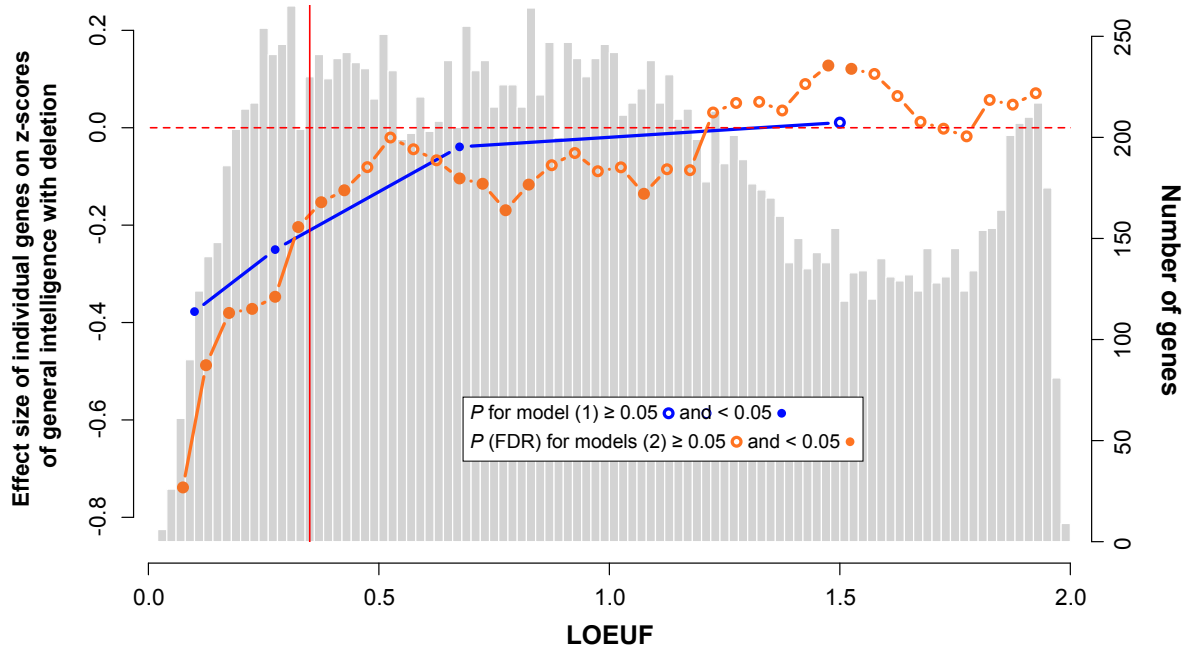

Figure 6: Effect size of individual genes included in deletion on z-scores measuring the general intelligence.

The light grey histogram represents the distribution of LOEUF values for 18,451 autosomal genes. The blue line represents the estimates for a gene in each of the 4 categories of LOEUF included in the model (Supplementary methods): highly intolerant genes (LOEUF <0.2, n=980), moderately intolerant genes (0.2≤LOEUF<0.35 n=1,762), tolerant genes (0.35≤LOEUF<1, n=7,442) and genes highly tolerant to pLoF (LOEUF≥1, n=8,267). The orange line represents the estimated effect size of 37 categories of genes based on their LOEUF values (sliding windows=0.15) in the model (Supplementary methods). Genes with a LOEUF below 0.35 (vertical red line) are considered to be intolerant to pLoF by gnomAD. Left Y-axis values: z-scored general intelligence (1 z-score is equivalent to 15 points of IQ) for deletion. Right Y-axis values: number of genes represented in the histogram.

REFERENCES

1. Coe, B. P. et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. 46, 1063–1071 (2014).

2. Coe, B. P. et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. Nat. Genet. 51, 106–116 (2019).

3. Wilfert, A. B., Sulovari, A., Turner, T. N., Coe, B. P. & Eichler, E. E. Recurrent de novo mutations in neurodevelopmental disorders: properties and clinical implications. Genome Med. 9, (2017).

4. Huguet, G., Ey, E. & Bourgeron, T. The genetic landscapes of autism spectrum disorders. Annu Rev Genomics Hum Genet 14, 191–213 (2013).

5. Pinto, D. et al. Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. Am. J. Hum. Genet. 94, 677–694 (2014).

6. Maillard, A. M. et al. The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. Mol Psychiatry 20, 140–7 (2015).

7. Sakai, M. et al. Assessment of copy number variations in the brain genome of schizophrenia patients. Mol. Cytogenet. 8, (2015).

8. Szatkiewicz, J. P. et al. Copy number variation in schizophrenia in Sweden. Mol. Psychiatry 19, 762–773 (2014).

9. CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. 49, 27–35 (2017).

10. Sanders, S. J. et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 87, 1215–1233 (2015).

11. Riggs, E. R. et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American

College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet. Med. 1–13 (2019) doi:10.1038/s41436-019-0686-8.

12. Miller, D. T. et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. Am. J. Hum. Genet. 86, 749–764 (2010).

13. Stefansson, H. et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 505, 361–366 (2014).

14. Kendall, K. M. et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. Br. J. Psychiatry 214, 297–304 (2019).

15. Posthuma, D., de Geus, E. J. C. & Boomsma, D. I. Perceptual Speed and IQ Are Associated Through Common Genetic Factors. Behav. Genet. 31, 593–602 (2001).

16. Hill, W. D. et al. Genomic analysis of family data reveals additional genetic effects on intelligence and personality. Mol. Psychiatry 23, 2347–2362 (2018).

17. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat. Genet. 50, 912–919 (2018).

18. Hill, W. D. et al. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. Mol. Psychiatry 24, 169–181 (2019).

19. Huguet, G. et al. Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. JAMA Psychiatry 75, 447–457 (2018).

20. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016).

21. Karczewski, K. J. et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv 531210 (2019) doi:10.1101/531210.

22. Deary, I. J. Intelligence. Annu. Rev. Psychol. 63, 453–482 (2011).

23. McRae, J. F. et al. Prevalence and architecture of de novo mutations in developmental disorders. Nature 542, 433–438 (2017).

24. An open resource of structural variation for medical and population genetics | bioRxiv. https://www.biorxiv.org/content/10.1101/578674v1.full.

25. D'Angelo, D. et al. Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. JAMA Psychiatry 73, 20–30 (2016).

26. Moreno-De-Luca, A. et al. The Role of Parental Cognitive, Behavioral, and Motor Profiles in Clinical Variability in Individuals With Chromosome 16p11.2 Deletions. JAMA Psychiatry 72, 119–126 (2015).

27. Bernier, R. et al. Clinical phenotype of the recurrent 1q21.1 copy-number variant. Genet. Med. 18, 341–349 (2016).

28. Sanders, S. J. et al. A framework for the investigation of rare genetic disorders in neuropsychiatry. Nat. Med. 25, 1477–1487 (2019).

29. Hawrylycz, M. et al. Canonical Genetic Signatures of the Adult Human Brain. Nat. Neurosci. 18, 1832–1844 (2015).