Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples

Guillaume Huguet, PhD; Catherine Schramm, PhD; Elise Douard, MSc; Lai Jiang, PhD; Aurélie Labbe, PhD; Frédérique Tihy, PhD; Géraldine Mathonnet, PhD; Sonia Nizard, MD; Emmanuelle Lemyre, MD; Alexandre Mathieu, MSc; Jean-Baptiste Poline, PhD; Eva Loth, PhD; Roberto Toro, PhD; Gunter Schumann, PhD; Patricia Conrod, PhD; Zdenka Pausova, MD; Celia Greenwood, PhD; Tomas Paus, MD, PhD; Thomas Bourgeron, PhD; Sébastien Jacquemont, MD; for the IMAGEN Consortium

IMPORTANCE: Copy number variants (CNVs) classified as pathogenic are identified in 10% to 15% of patients referred for neurodevelopmental disorders. However, their effect sizes on cognitive traits measured as a continuum remain mostly unknown because most of them are too rare to be studied individually using association studies.

OBJECTIVE: To measure and estimate the effect sizes of recurrent and nonrecurrent CNVs on IQ.

DESIGN, SETTING, AND PARTICIPANTS: This study identified all CNVs that were 50 kilobases (kb) or larger in 2 general population cohorts (the IMAGEN project and the Saguenay Youth Study) with measures of IQ. Linear regressions, including functional annotations of genes included in CNVs, were used to identify features to explain their association with IQ. Validation was performed using intraclass correlation that compared IQ estimated by the model with empirical data.

MAIN OUTCOMES AND MEASURES: Performance IQ (PIQ), verbal IQ (VIQ), and frequency of de novo CNV events.

RESULTS: The study included 2090 European adolescents from the IMAGEN study and 1983 children and parents from the Saguenay Youth Study. Of these, genotyping was performed on 1804 individuals from IMAGEN and 977 adolescents, 445 mothers, and 448 fathers (484 families) from the Saguenay Youth Study. We observed 4928 autosomal CNVs larger than 50 kb across both cohorts. For rare deletions, size, number of genes, and exons affect IQ, and each deleted gene is associated with a mean (SE) decrease in PIQ of 0.67 (0.19) points ($P = 6 \times 10^{-4}$); this is not so for rare duplications and frequent CNVs. Among 10 functional annotations, haploinsufficiency scores best explain the association of any deletions with PIQ with a mean (SE) decrease of 2.74 (0.68) points per unit of the probability of being loss-of-function intolerant ($P = 8 \times 10^{-5}$). Results are consistent across cohorts and unaffected by sensitivity analyses removing pathogenic CNVs. There is a 0.75 concordance (95% CI, 0.39-0.91) between the effect size on IQ estimated by our model and IQ loss calculated in previous studies of 15 recurrent CNVs. There is a close association between effect size on IQ and the frequency at which deletions occur de novo (odds ratio, 0.86; 95% CI, 0.84-0.87; $P = 2.7 \times 10^{-88}$). There is a 0.76 concordance (95% CI, 0.41-0.91) between de novo frequency estimated by the model and calculated using data from the DECIPHER database.

CONCLUSIONS AND RELEVANCE: Models trained on nonpathogenic deletions in the general population reliably estimate the effect size of pathogenic deletions and suggest omnigenic associations of haploinsufficiency with IQ. This represents a new framework to study variants too rare to perform individual association studies and can help estimate the cognitive effect of undocumented deletions in the neurodevelopmental clinic.

Published online March 21, 2018.

© 2018 American Medical Association. All rights reserved.
Copy number variants (CNVs) contribute to a spectrum of neurodevelopmental disorders (NDDs) and psychiatric disorders, including intellectual disabilities (IDs), autism spectrum disorders, and schizophrenia. With the routine implementation of whole genome chromosomal microarrays in medical diagnostics, pathogenic CNVs (as defined by the American College of Medical Genetics) are identified in 10% to 15% of children referred for NDDs. Several recurrent CNVs have been individually associated with IDs, autism spectrum disorders, and schizophrenia. Beyond association with a psychiatric diagnosis, little is known about the effect size of CNVs on cognitive traits. A study performed in the general population of Iceland found that 26 psychiatric CNVs reduce IQ, in aggregate, by 15 points or 1 SD. With the use of the cognitive tests available in the UK Biobank, 54 loci were associated with decreased scores, ranging from 0.1 to 0.5 SD. However, most pathogenic CNVs reported back to patients are undocumented because they are ultrarare or even private to the patient or family. They cannot be investigated using individual association studies. Their associations with cognition and mechanisms by which they lead to neurodevelopmental symptoms remain unknown. These nonrecurrent CNVs have been studied in aggregate by size categories in a general population sample of 6819 individuals from Estonia. In this cohort, rare, large, and intermediate (>250 kilobases [kb]) deletions and large, rare duplications (>1 megabase [Mb]) were found in 10% of the population. In aggregate, these CNVs were associated with IDs and adversely affected educational achievement; cognitive measures were unavailable.

The aim of this study was to calibrate and validate models to measure and estimate effect sizes of nonrecurrent pathogenic CNVs on general intelligence measured by IQ. To achieve this, we estimated effect sizes of recent recurrent and nonrecurrent CNVs on IQ using 2 general population cohorts. We then scored CNVs using 10 functional annotations to identify variables that contribute the most to variation in IQ. This model, which was subsequently validated, will help clinicians and researchers estimate the association of pathogenic CNVs with IQ.

Methods

Cohorts

General Population Cohorts

We used 2 cohorts recruited from the general population: IMAGEN, including 2090 adolescents from Europe, and the Saguenay Youth Study (SYS), including 1983 individuals (1032 children, 951 parents, 486 families) from Quebec, Canada. All children completed tests of verbal IQ (VIQ) and performance IQ (PIQ) using the Wechsler Intelligence Scale for Children, Fourth Edition (subset) for IMAGEN and Wechsler Intelligence Scale for Children, Third Edition for SYS. Distribution of IQ scores are available in eTable 1 in the Supplement. The IMAGEN project had obtained ethical approval by the local ethics committees and written informed consent from all participants and their legal guardians. For SYS, the institutional review boards of all participating institutions approved all studies reported herein. For SYS and IMAGEN, the parents and adolescents provided written informed consent and assent, respectively. All data were deidentified.

Clinical Cohorts

We used the chromosomal microarray database from the cytogenetic laboratory of the pediatric hospital of Center Hospitalier Universitaire Sainte-Justine (CHU-SJ; Montreal, Canada), including 16586 individuals referred for NDDs, and the Simon simplex collection (SSC), including 2591 children with autism spectrum disorders and their family members.

Genotyping and CNV Detection

Genotyping technologies are detailed in the eMethods in the Supplement. A total of 1804 individuals from IMAGEN and 977 adolescents, 445 mothers, and 448 fathers (484 families) from SYS (Figure 1A) met stringent quality control criteria (call rate ≥99%, log 

R

 R ratio SD <0.35, B allele frequency SD <0.08, and wave factor <0.05). We computed relatedness separately in IMAGEN and SYS based on the identity by state using PLINK. The CNV detections from PennCNV and QuantiSNP were combined to minimize the number of potential false discoveries. We used standard filtering strategies detailed in the eMethods in the Supplement.

Annotation of CNVs

We annotated CNVs for size and number of genes using RefSeq genes (https://genome.ucsc.edu/), and genes were annotated using the probability of being loss-of-function intolerant (pLI), the residual variation intolerance score, the score rate for intolerance for deletions and duplications, the number of protein-protein interactions, and the differential stability score of regional patterns of gene expression in the brain. These 5 scores were transformed, and the score associated with a CNV is the sum of scores of genes with all isoforms fully contained in the CNV (complete genes) (eMethods in the Supplement). The CNVs were also annotated with 2 lists of genes, including postsynaptic density of the human cortex, genes regulated by the Fragile-X mental retardation protein, and...
and the number of expression quantitative trait loci regulating genes expressed in the brain\(^3\) (eMethods and eTable 2 in the Supplement).

**Statistical Analysis**

Only autosomal CNVs were analyzed, and 3 outliers were excluded (eMethods in the Supplement). \(P < .05\) indicates statistical significance, and all tests were 2-sided.

**Quantifying the Effect Size and Numbers of Genes and Exons of Autosomal CNVs on IQ**

We performed 3 multiple linear regressions to quantify the effect size of CNVs (model 1), number of genes (model 2), and number of exons (model 3) on PIQ and VIQ. For each model, the variable of interest was measured in 4 categories of CNVs according to frequency (rare or common) and type (deletion or duplication). Models 1 through 3 included adjustment for ancestry, sex, age, microarray technology, and intrafamilial relatedness (eMethods and eFigure 1 in Supplement).

**Characteristics of Autosomal Deletion Contents That Affect IQ**

We performed a stepwise variables selection procedure based on the Bayesian information criterion\(^3\) to investigate 10 variables that would best explain the association of deletions with PIQ and VIQ (eMethods in the Supplement). The best model is denoted as model 4 in the remainder of this article. Sensitivity analyses were performed for models 1 through 4 (eMethods in the Supplement).

We then examined whether model 4 could predict the association of IQ with 15 known recurrent CNVs by calculating the concordance between model prediction and empirically measured loss of IQ obtained from previous publications (eMethods, eFigures 2 and 3, and eTable 3 in the Supplement). The concordance was computed using the intraclass correlation (ICC\(_{3,1}\)).\(^{34}\)

**Association Between Effect Size of CNVs on IQ and De Novo Frequency**

Using data on inheritance from the CHU-SJ and SSC cohorts, we performed a logistic regression model (model 5) to establish the association between the probability at which CNVs occur de novo and their association with IQ predicted by model 4. We computed the ICC\(_{3,1}\) to evaluate the concordance between the probability for a CNV to be de novo predicted by model 5 and de novo frequency for the same 15 recurrent CNVs using data from the DECIPHER database (http://decipher.sanger.ac.uk) (eTable 3 in the Supplement).

**Results**

**Quantifying the Effect Size of CNVs, Number of Genes, and Number of Exons on IQ**

We observed 4928 autosomal CNVs larger than 50 kb across both cohorts (Figure 1B and Table 1). Rare CNVs of 250 kb or larger (\(n = 308\)) are mostly nonrecurrent (92.8%), and their frequencies, similar across both cohorts, are identical to a previously published study\(^16\) (eResults, eFigure 4, and eTables 4-6 in the Supplement). We examined variables recurrently associated with NDDs and psychiatric disorders,\(^7,8,16\) namely, CNV size (model 1), number of genes (model 2), and number of exons (model 3), and estimated their association with IQ for 4 CNV categories, namely, common and rare deletions and duplications. In all 3 models, only rare deletions had significant effects on IQ. The effect size (model 1) can be illustrated by a decrease of PIQ (mean [SE], 5.7 [2.0] points; \(P = 6 \times 10^{-4}\)) and VIQ (mean [SE], 3.6 [2.0] points; \(P = .03\)) for each deleted Mb (Table 2). These results are concordant with comparisons between carriers and noncarriers of rare CNVs stratified by size (eTable 7 in the Supplement). In model 2, each gene deleted by a rare CNV decreases PIQ by a mean (SE) of 0.67 (0.19) points \((P = 6 \times 10^{-4}\)) and VIQ by 0.72 (0.19) points \((P = 2 \times 10^{-4}\)).
model 3, each exon deleted by a rare CNV decreases PIQ by a mean (SE) of 0.07 (0.02) points (P = 2 × 10⁻⁵) and VIQ by 0.06 (0.02) points. For models 1 through 3, effects are similar in both cohorts separately. We found no measurable associations of common deletions or duplications with IQ (Table 2). The distributions of Akaike information criterion and Bayesian information criterion, obtained by fitting the models 1 through 3 on 1000 bootstrap samples of the pooled dataset each, show that gene and exon contents provide a better fit than size (eTable 8 in the Supplement). Applying models 1 through 3 on individuals with European ancestry shows similar results (eTable 9 in the Supplement).

Characteristics Underlying the Associations of Deletions With IQ
To understand factors that potentially drive the associations of deletions with IQ, we investigated 10 functional annotations in the subset of 1713 individuals carrying at least 1 autosomal deletion in the pooled data set. The stepwise variable selection procedure converges on model 4, including pLI alone (PIQ: effect = −2.69, bias corrected effect = −2.74, SE = 0.68, P = 8 × 10⁻⁵; VIQ: effect = −2.41; bias corrected effect = −2.52; SE = 0.71; P = 7 × 10⁻⁵). The associations of pLI estimated in IMAGEN and SYS separately are the same, and no differences are observed between the association with PIQ and VIQ (eTable 10 in the Supplement). In the bootstrap procedure, pLI is the most frequently selected covariate for PIQ (37.8%) and the second most frequently selected covariate for VIQ (23.5%) behind the residual variation intolerances score (28.1%) and is always preferred to size or number of genes. Model 4 relies on pLI score, and the distribution of associations with PIQ of 17102 individual genes shows that 33% of coding genes are predicted to affect PIQ by 1 point or more and 23% by 2 points or more. More than 968 genes (6%) have a maximum pLI of 1, with a corresponding effect size of −2.7 for PIQ and −2.5 for VIQ (Figure 2A and eFigure 5A in the Supplement), demonstrating that the model cannot estimate the association of 93 causal genes for IDs with very large or extreme associations with IQ. The variable selection procedure, performed after a principal component

### Table 1. Description of IMAGEN and Saguenay Youth Study Cohorts

<table>
<thead>
<tr>
<th>Variable</th>
<th>IMAGEN (610Kq and 660Wq)</th>
<th>Saguenay Youth Study (Children)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of samples (with PIQ or VIQ)</td>
<td>1744&lt;sup&gt;a&lt;/sup&gt;</td>
<td>559</td>
</tr>
<tr>
<td>Chromosomes 1-22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of deletions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1584</td>
<td>609</td>
</tr>
<tr>
<td>By sample</td>
<td>0.9 (0.9)</td>
<td>1.1 (1.0)</td>
</tr>
<tr>
<td>No. of duplications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1179</td>
<td>608</td>
</tr>
<tr>
<td>By sample</td>
<td>0.7 (0.8)</td>
<td>1.1 (1.1)</td>
</tr>
<tr>
<td>Chromosome X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of deletions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>308</td>
<td>185</td>
</tr>
<tr>
<td>By sample</td>
<td>0.2 (0.4)</td>
<td>0.3 (0.5)</td>
</tr>
<tr>
<td>No. of exonic deletions&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>519</td>
<td>264</td>
</tr>
<tr>
<td>By sample</td>
<td>0.3 (0.5)</td>
<td>0.5 (0.7)</td>
</tr>
<tr>
<td>No. of exonic duplications&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
<td>16</td>
</tr>
<tr>
<td>By sample</td>
<td>0.03 (0.2)</td>
<td>0.03 (0.2)</td>
</tr>
<tr>
<td>Age, mod&lt;sup&gt;d&lt;/sup&gt;</td>
<td>173.4 (4.4)</td>
<td>173.4 (22.7)</td>
</tr>
<tr>
<td>Male, No. (%)</td>
<td>853 (48.9)</td>
<td>261 (46.7)</td>
</tr>
<tr>
<td>PIQ</td>
<td>106.6 (14.8)</td>
<td>105.4 (13.2)</td>
</tr>
<tr>
<td>VIQ</td>
<td>110.0 (15.8)</td>
<td>103.7 (12.5)</td>
</tr>
</tbody>
</table>

Abbreviations: CNV, copy number variant; PIQ, performance intelligence quotient; VIQ, verbal intelligence quotient.

<sup>a</sup> Data are presented as mean (SD) unless otherwise indicated.

<sup>b</sup> One sample has PIQ but not VIQ.

<sup>c</sup> The term exonic refers to a CNV that includes at least 1 gene (only complete genes are considered).

<sup>d</sup> Age after imputation by mean of the cohort for the 376 samples with missing data in the IMAGEN cohort. The PIQ and VIQ means differ between the IMAGEN and SYS cohorts (2-tailed, unpaired t-test, P = 2 × 10⁻⁵).
Table 2. CNV Size and Numbers of Coding Protein Genes and Exons, Summed Across All CNVs Carried by the Same Individual

<table>
<thead>
<tr>
<th>CNV Category</th>
<th>CNV Size, Mb</th>
<th>No. of Carriers</th>
<th>PIQ Regression Coefficient, Mean (SE) [P Value]</th>
<th>No. of Genes in CNVs</th>
<th>No. of Exons of CNVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMAGEN (n = 1744)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rare deletions</td>
<td>391</td>
<td>-4.41 (2.01) [0.02]</td>
<td>-0.84 (2.14) [0.69]</td>
<td>69</td>
<td>-0.71 (0.34) [0.03]</td>
</tr>
<tr>
<td>Rare duplications</td>
<td>357</td>
<td>1.53 (1.61) [0.34]</td>
<td>1.81 (1.71) [0.29]</td>
<td>142</td>
<td>0.21 (0.2) [0.29]</td>
</tr>
<tr>
<td>Common deletions</td>
<td>845</td>
<td>-0.51 (3.91) [0.89]</td>
<td>1.18 (4.16) [0.77]</td>
<td>222</td>
<td>-0.62 (0.28) [0.02]</td>
</tr>
<tr>
<td>Common duplications</td>
<td>639</td>
<td>0.58 (1.47) [0.69]</td>
<td>-0.28 (1.57) [0.85]</td>
<td>344</td>
<td>0.24 (0.16) [0.13]</td>
</tr>
<tr>
<td>SYS (n = 967)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rare deletions</td>
<td>118</td>
<td>-8.7 (2.9) [0.002]</td>
<td>-9.29 (2.72) [7 × 10^-4]</td>
<td>25</td>
<td>-0.62 (0.22) [0.005]</td>
</tr>
<tr>
<td>Rare duplications</td>
<td>140</td>
<td>-0.83 (1.24) [0.50]</td>
<td>0.51 (1.16) [0.66]</td>
<td>67</td>
<td>-0.38 (0.18) [0.03]</td>
</tr>
<tr>
<td>Common deletions</td>
<td>622</td>
<td>0.52 (2.43) [0.83]</td>
<td>-0.84 (2.28) [0.71]</td>
<td>252</td>
<td>0.08 (0.24) [0.75]</td>
</tr>
<tr>
<td>Common duplications</td>
<td>577</td>
<td>2.18 (1.69) [0.19]</td>
<td>0.36 (1.59) [0.81]</td>
<td>313</td>
<td>-0.02 (0.19) [0.91]</td>
</tr>
<tr>
<td>Both (n = 2711)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rare deletions</td>
<td>509</td>
<td>-5.7 (1.64) [6 × 10^-4]</td>
<td>-3.63 (1.68) [0.03]</td>
<td>94</td>
<td>-0.67 (0.19) [6 × 10^-4]</td>
</tr>
<tr>
<td>Rare duplications</td>
<td>497</td>
<td>0.06 (1.95)</td>
<td>0.95 (1.01) [0.34]</td>
<td>209</td>
<td>-0.1 (0.13) [0.47]</td>
</tr>
<tr>
<td>Common deletions</td>
<td>1467</td>
<td>0.15 (2.12) [0.94]</td>
<td>-0.27 (2.13) [0.89]</td>
<td>474</td>
<td>-0.24 (0.19) [0.19]</td>
</tr>
<tr>
<td>Common duplications</td>
<td>1216</td>
<td>1.04 (1.12) [0.35]</td>
<td>-0.14 (1.15) [0.90]</td>
<td>657</td>
<td>0.16 (0.12) [0.18]</td>
</tr>
</tbody>
</table>

Abbreviations: CNV, copy number variant; Mb, megabase; PIQ, performance IQ; SYS, Saguenay Youth Study; and VIQ, verbal IQ.

* We used multiple regression analysis to study the estimated associations with PIQ and VIQ of the sum of autosomic CNV (≤50 kilobases [kb]) size and the sum of coding protein gene (for all isoforms) and exons for IMAGEN and SYS. Covariates included the 6 first principal components of genetic distance, sex, age, array technology, and familial relatedness. There was no interaction between sex and CNV categories, and this interaction term was subsequently removed. For size, the estimate of −5.7 translates into a loss of 5 points of PIQ per megabase of material included in rare deletions. For the number of genes, the estimate of −0.67 translates into a loss of 0.67 points of PIQ per gene included in a rare deletion (Bonferroni correction = 0.008). Results are unchanged if only individuals with more than 80% of European ancestry are included (Table II in the Supplement).
analysis does not provide a better fit than model 4 (eMethods, eResults, eFigure 7, eTables 12 and 13 in the Supplement). Of note, there is no interaction between sex and any of the variables tested in models 1 through 4.

Sensitivity Analyses
We examined whether a subgroup of CNVs biased or overly influenced the results. Sensitivity analyses show that effect sizes of rare deletions and pLI on IQ are unchanged even after re-
moving carriers with CNVs of 1 Mb or greater as well as recurrent CNVs previously associated with psychiatric NDDs. Transformed variables did not improve any of the models (eTables 14-16 in the Supplement). Additional sensitivity test results are detailed in eResults in the Supplement.

Model Validation
We compared IQ loss predicted by the model to IQ loss empirically measured in previous studies\(^\text{20,21}\) of 15 known recurrent CNVs without causal genes for IDs (eTable 3 and eFigure 6 in the Supplement). The concordance is 0.75 for PIQ (95% CI, 0.39-0.91; \(P = 5 \times 10^{-5}\)) and 0.72 for VIQ (95% CI, 0.35-0.90, \(P = 8 \times 10^{-4}\)) (Figure 2B and eFigure 5B in the Supplement). Widths of CIs are correlated with the effect size of the CNV, reflecting that CNVs with high pLI are rarely observed and are different from the distribution of CNVs observed in our general population cohorts. Of note, these results are similar whether we include or exclude, from the training data set, the 3 recurrent CNVs observed in 6 individuals in the pooled cohort (16p11.2 proximal BP4-BP5: 1 adolescent from IMAGEN; 16p12.1: 1 adolescent from IMAGEN and 2 sisters from SYS; 16p13.11: 2 adolescents from IMAGEN).

Implication for Medical Genetics
The widespread but small effect size of haploinsufficiency implies that pathogenic deletions could be found throughout the genome if the aggregate haploinsufficiency score is high enough to affect IQ. This finding is consistent with the fact that more than one-third (7429) of the coding genome is deleted by 1217 pathogenic autosomal deletions reported back to patients by the CHU-SJ (Figure 2C). Of note, genes included in pathogenic variants (n = 6799) or variants of unknown significance deletions (n = 1396) have a higher pLI than those included in benign deletions (n = 928) (Wilcoxon \(P = 4.8 \times 10^{-14}\) for genes included in pathogenic versus benign deletions and Wilcoxon \(P = 9.7 \times 10^{-6}\) for genes included in variant of unknown significance vs benign deletions) (Figure 2D and eFigure 5C in the Supplement).

As an illustration, we estimated the association with IQ of the aforementioned 1217 pathogenic deletions: the top quartile (25% of CNVs) is estimated to decrease IQ by more than 28 points, whereas the 2 middle quartiles decrease IQ between 28 and 4 points (eTable 17 in the Supplement). Of note, estimates for the lower quartile are smaller than a 4-point decrease in IQ, but most of the latter CNVs cannot be estimated properly because they disrupt a causal gene with large effects.

De Novo CNVs
In the neurodevelopmental clinic, de novo events are regarded as strong arguments in favor of pathogenicity, and CNV size was previously associated with de novo frequency.\(^\text{36}\) However, to our knowledge, the exact association between effect size on IQ and de novo frequency has not been studied. We examined inheritance of 2161 deletions 50 kb or larger from the CHU-SJ and SSC cohorts. The logistic regression model (model 5) suggests a tight association between effect size on IQ (estimated by model 4) and probability of being a de novo CNV (odds ratio, 0.86; 95% CI, 0.84-0.87; \(P = 2.7 \times 10^{-8}\)) (Figure 3A). Results are similar when recurrent CNVs are excluded. The concordance between the probability of occurring de novo estimated by model 5 and de novo frequency calculated using empirical data from the DECIPHER database on 15 recurrent deletions is 0.77 (95% CI, 0.43-0.91; \(P = 2.7 \times 10^{-4}\)) (https://decipher.sanger.ac.uk/) (Figure 3B). We also examined 1147 CNVs 100 kb or larger in 837 adolescents and their parents from the general population (SYS). Seventeen occurred de novo (1%); 6 deletions and 11 duplications), which is similar to frequencies previously reported in the general population (eFigure 8 and eTables 18 and 19 in the Supplement). Among the 6 de novo deletions, 3 have never been referenced in the database of genomic variants, suggesting class 1 de novo events\(^\text{37}\) (eFigure 9B, cases 1, 5, and 6, in the Supplement). Although model 4 predicts a large effect of −28.7 points for the deletion in case 5, the predicted effect for cases 1 and 6 is less than 5 points, suggesting that among class 1 de novo events\(^\text{37}\) effect sizes can be modest. The other CNVs (cases 2, 3, and 4) have general population frequencies greater than 0.01%, suggesting class 2 de novo events\(^\text{37}\) consistent with the small predicted associations with IQ.

Discussion
This study quantifies and predicts the effect size of deletions on IQ using data from the general population and clinical cohorts. Deletions are associated with a decrease in general intelligence, and our models suggest that the effect of haploinsufficiency can be reliably predicted for most pathogenic deletions. This approach provides a framework for studying the effect of CNVs that are too rare to study in individual association studies.

Our study suggests that haploinsufficiency of most of the coding genome potentially influences general intelligence, and one-third of the coding genome affects IQ by 0.67 points or more (mean effect of genes included in rare deletions). This finding is consistent with the omnigenic model\(^\text{37}\) of complex traits based on the observation that genome-wide association study association signals are spread across most of the genome, including variants near many genes without any obvious connection to disease.

This finding has important implications for the clinical interpretation and functional studies of CNVs. A dominant hypothesis that guides many studies is that a major gene(s) contributes to most of the neurodevelopmental symptoms observed in CNV carriers. Our study suggests an alternative hypothesis that the large effect size observed in pathogenic deletions may be polygenic in nature and attributable to the sum of small individual effects of each gene included in the deletion. This hypothesis could explain why causal genes or major drivers have been difficult to identify in most recurrent CNVs.\(^\text{38,39}\)

Intriguingly, our model predicts reasonably the effect size of the Smith-Magenis deletion without attributing a large effect size to the RAI1 gene (OMIM 607642) on IQ. Although RAI1 causes most of the dysmorphic and disruptive behavioral fea-
tudes of Smith-Magenis syndrome, its association with IQ may be smaller than expected. Of note, a recent study did not identify an excess of de novo mutations in RA11 in more than 7000 individuals with ID.

Large discordances between estimated and empirical estimates of IQ are of particular interest. For example, the model underestimated the effects of 15q13.3 and 3q29 deletions. This underestimation could be attributable to genes with large effect sizes, although none have been clearly identified in these CNVs by previous studies. Alternatively, the association of these 2 CNVs with IQ might be overestimated in the literature because carriers of these deletions are mostly referred to the clinic for behavioral or neurologic symptoms (eg, epilepsy). Although enrichment of the 17p12/hereditary neuropathy with liability to pressure palsies deletion was not previously reported in a neurodevelopmental cohort, our model predicts an IQ loss of 6 points. This finding is consistent with the enrichment observed in the CHU-SJ neurodevelopmental cohort (odds ratio, 3.25; 14 cases per 16586;  = .002). Previous studies reporting association with schizophrenia included individuals with neurodevelopmental disorders (odds ratio ranging from 1 to 5).

Our study quantifies the association between the effect size of deletions on IQ and the frequency at which they occur de novo. The probability of occurring de novo increases rapidly for deletions with small effect sizes on IQ (a few points), reaching a frequency of 100% for effect sizes of 30 points or greater. The model's prediction has a concordance of 0.75 with the de novo frequency of 15 recurrent CNVs calculated using empirical data. It is likely that many de novo deletions, which confer significant risk for NDDs, may lie on a continuum between class 1 and class 2 variants. In fact, most deletions that affect IQ have effect sizes of less than 30 points, are present in general population cohorts, and would be classified as class 2 de novo variants, which incorrectly reflects the risk they confer for NDDs.

Limitations

The predictive models presented in the study have several limitations. In particular, they are unable to attribute large effects to ID causal genes. This limitation is likely because calibration was performed in the general population (with too few cases of ID) and reliance on haploinsufficiency scores that were not intended to provide granularity among genes with large effects. Indeed, the model attributes a maximal effect of 2.74 points of PIQ loss, whereas causal genes for ID are associated with IQ loss between 40 and 60 points. On the other hand, it is likely that our model properly estimated small effect size because it was developed and calibrated in the general population based on a set of CNVs that contain genes with milder effects.

Conclusions

The association of deletions with IQ can be modeled using haploinsufficiency scores based on a linear and additive assump-
Measuring and Estimating Effect Sizes of Copy Number Variants on IQ

Study concept and design: accuracy of the data analysis. Responsibility for the integrity of the data and the had full access to all the data in the study and takes Jacquemont share last authorship. Dr Jacquemont share first authorship. Drs Bourgeron and Bourgeron, Jacquemont.

Douard, Loth, Conrod, Greenwood, Paus, Bourgeron, Jacquemont. Drs Bourgeron is supported by the Institut Pasteur, the University Paris Diderot, Centre National de la Recherche Scientifique, and the Bettencourt-Schueller Foundation. Dr 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. Dr Huguet is supported by the Sainte-Justine Foundation, the Merit scholarship program for foreign students, and the Network of Applied Genetic Medicine fellowships. Dr Schramm is supported by the Institute for Data Valorization fellowship. Dr Loth is supported by European Autism Interventions, which receives support from the Innovative Medicines Initiative Joint Undertaking under grant agreement 115300, the resources of which are composed of financial contributions from grant FP7/2007-2013 from the European Union’s Seventh Framework Programme, the European Federation of Pharmaceutical Industries and Associations companies’ kind contributions, and Autism Speaks. This work is supported by a grant from the Brain Canada Multi Investigator initiative (Dr Jacquemont). The Canadian Institutes of Health Research and the Heart and Stroke Foundation of Canada fund the Saguenay Youth Study. Funding for the project was provided by the Welcome Trust.

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

Group Members: The members of the IMAGEN Consortium are as follows: Tobias Banaschewski, MD, PhD, Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; Gregor Barker, PhD, Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, England; Arun L. W. Bokde, PhD, Discipline of Psychiatry, School of Medicine and Trinity College, Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland; Uli Bromberg Dipl-Psych, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany; Christian Büchel, MD, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany; Erinn Burke-Quinlan, PhD, Medical Research Council, Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, England; Herta Flor, PhD, Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, and Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany; Vincent Frouin PhD, Neurospin, CEA, Université Paris-Saclay, Gif-sur-Yvette, France; Hugh Garavan PhD, Departments of Psychiatry and Psychology, University of Vermont, Burlington; Penny Gowland, PhD, Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom; Andreas Heinz, MD, PhD, Department of Psychiatry and Psychotherapy, Charité, Universitätsmedizin Berlin, Berlin, Germany; Bend Ilttemann, PhD, Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, and Psychology, University of Hamburg, Hamburg, Germany; Sylviane Desrivieres PhD, Medical Research Council, Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, England; Sabina Millen, PhD, Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, and Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany; and all other individuals with IDs, are likely required to model the effects of duplications.
identifies specific genes associated with
Refining analyses of copy number variation
doi:10.1038/mp.2014.145


