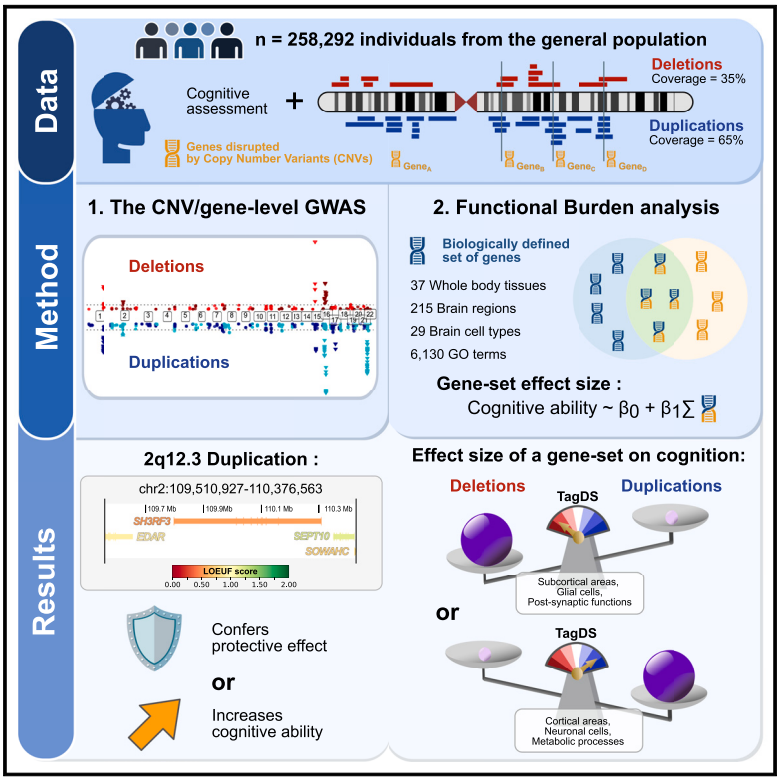
Cell Genomics

Article

Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

Graphical abstract



Highlights

- CNV-GWAS reveals the first positive impact on cognition for the 2q12.3 duplication
- The effects of deletions/duplications on cognitive ability are negatively correlated
- A new metric, tagDS, defines the gene-dosage-effect specificity of any set of genes
- Significant impact of genes expressed in non-brain tissues on cognitive ability

Authors

Guillaume Huguet, Thomas Renne, Cécile Poulain, ..., Laura Almasy, David C. Glahn, Sébastien Jacquemont

Correspondence

guillaumeaf.huguet@gmail.com (G.H.), sebastien.jacquemont@umontreal.ca (S.J.)

In brief

Copy-number variants are major contributors to neurodevelopmental disorders and are associated with lower cognition. Huguet et al. identified a duplication increasing cognitive ability. They highlighted that genes of many biological processes had unbalanced gene-dosage sensitivity toward deletions or duplications for both brain and non-brain functions.

Huguet et al., 2024, Cell Genomics *4*, 100721
December 11, 2024
https://doi.org/10.1016/j.xgen.2024.100721

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

Guillaume Huguet,[1,26,*] Thomas Renne,[1,2,26] Cécile Poulain,[1,2] Alma Dubuc,[3] Kuldeep Kumar,[1] Sayeh Kazem,[1,2] Worrawat Engchuan,[4,5] Omar Shanta,[6] Elise Douard,[1] Catherine Proulx,[1] Martineau Jean-Louis,[1] Zohra Saci,[1] Josephine Mollon,[7] Laura M. Schultz,[8] Emma E.M. Knowles,[9] Simon R. Cox,[10] David Porteous,[10,11,12] Gail Davies,[10] Paul Redmond,[10] Sarah E. Harris,[10] Gunter Schumann,[13] Guillaume Dumas,[1,14] Aurélie Labbe,[15] Zdenka Pausova,[16,17,18] Tomas Paus,[1,19] Stephen W. Scherer,[4,5,20] Jonathan Sebat,[21] Laura Almasy,[22] David C. Glahn,[23,24] and Sébastien Jacquemont[1,25,27,*]

[1]Centre Hospitalier Universitaire Sainte-Justine Research Center, Montreal, QC, Canada
[2]Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, QC, Canada
[3]École Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, Université de Lyon, Lyon, France
[4]The Hospital for Sick Children, Genetics and Genome Biology, Toronto, ON, Canada
[5]The Hospital for Sick Children, The Centre for Applied Genomics, Toronto, ON, Canada
[6]Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA
[7]Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[8]Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
[9]Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[10]Lothian Birth Cohorts, Department of Psychology, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Edinburgh EH8 9JZ, UK
[11]Medical Genetics Section, Centre for Genomic & Experimental Medicine, MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK
[12]Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
[13]Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK
[14]Mila – Quebec Artificial Intelligence Institute, Montréal, QC, Canada
[15]Département de Sciences de la Décision, HEC Montreal, Montreal, QC, Canada
[16]Research Institute of the Hospital for Sick Children, Toronto, ON, Canada
[17]Departments of Physiology and Nutritional Sciences, University of Toronto, Toronto, ON, Canada
[18]ECOGENE-21, Chicoutimi, QC, Canada
[19]Department of Psychiatry and Addictology, Department of Neuroscience, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada
[20]McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[21]University of California, San Diego, Department of Psychiatry, Department of Cellular & Molecular Medicine, Beyster Center of Psychiatric Genomics, San Diego, CA, USA
[22]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[23]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA
[24]Olin Neuropsychiatric Research Center, Institute of Living, Hartford Hospital, Hartford, CT, USA
[25]Department of Pediatrics, Université de Montréal, Montreal, QC, Canada
[26]These authors contributed equally
[27]Lead contact
*Correspondence: guillaumeaf.huguet@gmail.com (G.H.), sebastien.jacquemont@umontreal.ca (S.J.)
https://doi.org/10.1016/j.xgen.2024.100721

SUMMARY

Copy-number variants (CNVs) that increase the risk for neurodevelopmental disorders also affect cognitive ability. However, such CNVs remain challenging to study due to their scarcity, limiting our understanding of gene-dosage-sensitive biological processes linked to cognitive ability. We performed a genome-wide association study (GWAS) in 258,292 individuals, which identified—for the first time—a duplication at 2q12.3 associated with higher cognitive performance. We developed a functional-burden analysis, which tested the association between cognition and CNVs disrupting 6,502 gene sets biologically defined across tissues, cell types, and ontologies. Among those, 864 gene sets were associated with cognition, and effect sizes of deletion and duplication were negatively correlated. The latter suggested that functions across all biological processes were sensitive to either deletions (e.g., subcortical regions, postsynaptic) or duplications (e.g., cerebral cortex, presynaptic). Associations between non-brain tissues and cognition were driven partly by constrained genes, which may shed light on medical comorbidities in neurodevelopmental disorders.

INTRODUCTION

Copy-number variants (CNVs) are deletions or duplications larger than 1,000 base pairs.[1] CNVs are major contributors to risk for neurodevelopmental disorders (NDDs),[2] including intellectual disability (ID),[3–5] autism spectrum disorder (ASD),[6–8] and schizophrenia.[8–10] CNVs that increase the risk of psychiatric conditions also invariably affect cognitive abilities in individuals with or without a psychiatric diagnosis and regardless of ascertainment.[11–13] Such CNVs are often associated with multimorbidity in the clinic.[11–13] Whole-genome CNV detection is a first-tier diagnostic test routinely implemented in children referred to the clinic for NDDs.[14] Medical diagnostic laboratories attempt to classify CNVs as either benign or putative pathogenic, but beyond these categories, the effect sizes of CNVs on cognitive ability have been used to provide more nuanced information on the severity of a variant and to quantify the risk for NDDs. Indeed, cognitive ability remains one of the traits most commonly used in the pediatric clinic because it is predictive of the outcome and adaptive skills of children with neurodevelopmental symptoms.[15]

Due to statistical power, most studies have repeatedly analyzed a small set of the most frequently recurrent CNVs (population frequency > 1/10,000),[16–18] which collectively affect only approximately 2% of the coding genome.[19] As a result, our understanding of gene functions sensitive to gene dosage is highly biased. However, the vast majority of CNVs affecting neurodevelopmental and cognitive ability are ultra-rare (<1/10,000),[17] and associations have been established based on their size and gene content using burden analyses.[12,19–22] Such CNVs cover a large proportion of the coding genome and remain difficult to study individually with currently available sample sizes. Beyond CNVs, more generally, our understanding of gene-disrupting variants associated with cognitive ability and NDDs stems from approximately 200 genes disrupted by *de novo* variants.[4,23] Their functions are enriched in chromatin and transcription regulation, regulation of nervous system development, central nervous system neuron differentiation, and regulation of synapse structure and activity.[4,23] It is unclear, however, if these functions are most representative of cognitive ability or genetic constraint. In addition, previous studies reporting on the functional enrichment of ID- or NDD-associated genes have not stratified their findings based on classes of disrupting variants. It is, therefore, unknown whether specific biological functions and traits are preferentially sensitive to different classes of genomic variants (i.e., opposing gene dosage alterations such as deletions and duplications).

Knowledge gap: overall, it has been difficult to investigate the broad landscape of ultra-rare CNVs potentially involved in neurodevelopmental traits, such as cognitive ability. As a result, we have a limited understanding of the full range of gene-dosage-sensitive biological processes linked to cognitive ability. To circumvent the issue of power, research groups, including ours, have implemented alternative approaches aggregating rare variants disrupting genes with similar constraint scores in order to perform "constraint burden" association studies.[11,12,19–21,24] These burden analyses showed that genes with increasing intolerance to haploinsufficiency were associated with increasing effect sizes on cognitive ability and risk for psychiatric illnesses, such as ASD, schizophrenia, and bipolar disorder.[19,21] Similarly, studies have developed methods to aggregate common variants,[25] demonstrating that a robust association with a condition (e.g., ASD) can be established at the group level when individual single-nucleotide polymorphisms (SNPs) do not meet genome-wide criteria for association.

In this study, we aimed to investigate the full range of gene-dosage-sensitive biological processes linked to cognitive ability. To this end, we analyzed all CNVs >50 kb in 258,000 individuals across 6 cohorts from the general population. The CNV-level genome-wide association study (GWAS) identified the first CNV associated with higher cognitive ability. To further investigate CNVs too rare to be tested by the CNV-level GWAS, we performed functional-burden analyses. To do so, we aggregated all CNVs disrupting a group of genes assigned to a given biological function. Functional-burden associations were performed between cognitive ability and 6,502 gene sets assigned to biological functions at the tissue, cell type, and molecular levels. Functional-burden tests revealed that most functional gene sets were associated with cognitive ability when either deleted or duplicated, and only a few gene sets showed significant associations with cognition for both CNVs. As a result, we observed a negative correlation between the effect sizes of deletions and duplications across all functional gene sets, and this was not influenced by intolerance to haploinsufficiency. This suggests that the effects of most biological functions on cognitive ability are dependent on the type of gene dosage.

RESULTS

Gene dosage may be associated with higher cognitive ability

Among the 258,292 individuals from general population datasets, 15.6% carried at least one rare (allele frequency < 1%) autosomal CNV larger than 50 kb, fully encompassing one or more coding genes (hg19). Among all autosomal coding genes with loss-of-function observed/expected upper-bound fraction (LOEUF) values (n = 18,451), 71.8% were fully encompassed in one or more CNVs: 35% in deletions, 64.9% in duplications, and 28.1% in both deletions and duplications (Figures 1A–1C). Most of the genes encompassed in CNVs were contained in ultra-rare CNVs (<1/10,000) with fewer than 30 carriers (Figure 1C). We used a linear regression model (gene-level GWAS; cf. STAR Methods, statistical model 1) to test the association of general cognitive ability with 241 and 596 genes covered by at least 30 deletions or duplications, respectively (Figures 1D and 1E). We identified 6 deletions encompassing a total of 68 genes and 7 duplications encompassing a total of 122 genes with previously published negative effects (Table S1) that persisted when we conducted a meta-analysis across 9 sub-cohorts defined by cognitive assessments (Table 1; Figure 1E). We identified a novel association between a duplication at 2q12.3 and positive effects ($z = 0.434$, $p = 7.58 \times 10^{-3}$) on cognitive ability (Figures 1F and S1). This duplication, observed in 36 individuals, included 4 non-intolerant genes with an LOEUF $\geq$ 0.35 (*EDAR*, *SH3RF3*, *SEPT10*, *SOWAHC*) and was observed at a similar frequency (1–2 in 10,000) across cohorts (Fisher's

CellPress
OPEN ACCESS

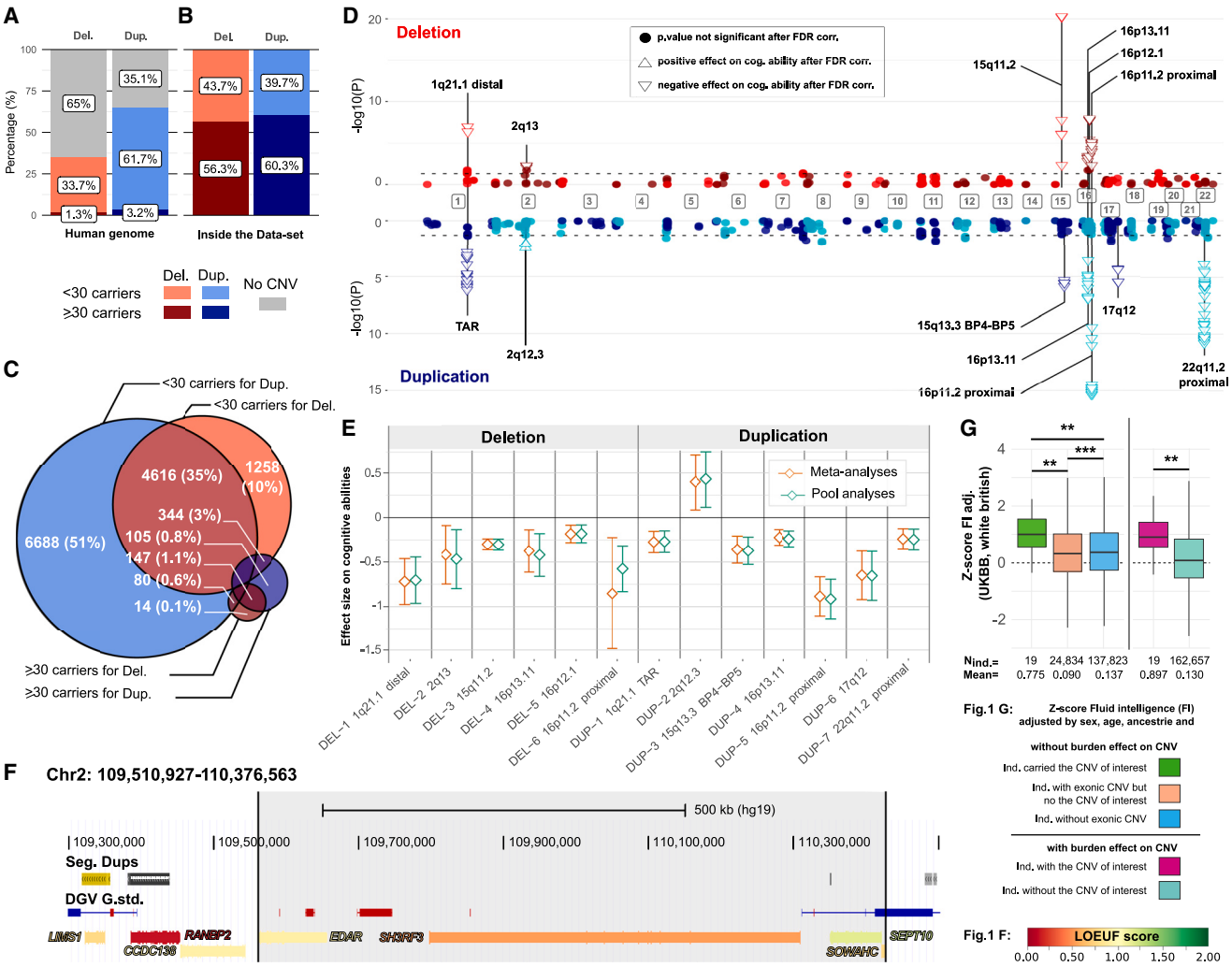


Figure 1. CNV-GWAS on general cognitive abilities at the gene level

(A) Proportion of genes deleted (red) or duplicated (blue) at least once in the general population pooled dataset among all genes in the human genome (hg19). Deleted or duplicated genes observed in less than 30 carriers (light color) and 30 or more (dark color), as well the proportion of genes not observed in any CNV (gray).

(B) The majority of deleted or duplicated genes were observed in more than 30 carriers.

(C) Venn diagram illustrating the overlap between gene content of ultra-rare and rare deletions and duplications, specifically the number of genes deleted and/or duplicated at least once in these CNVs from the pooled general population datasets.

(D) The Miami plot illustrates the −log10-transformed *p* value of the association with cognitive ability for each gene included in deletions (red) at the top, and duplications (blue) at the bottom, along the genome. Adjacent chromosomes are shown in alternating light and dark colors. Triangles represent significant genes after FDR correction, while circles represent non-significant genes. The direction of the triangle indicates the effect size. The dash line represents the nominal significant *p* value threshold.

(E) Data are represented as mean ± standard error for cognitive ability, green diamonds indicate pooled analyses (all cohorts regrouped), and orange diamonds represent meta-analyses (mean of effect sizes computed for each cohort separately). For meta-analyses, fixed-effect model values were chosen when the heterogeneity test was not significant ($p > 0.1$), and a random-effects model was employed when heterogeneity was significant. We displayed the values for the gene within the CNV that had the highest number of carriers (see also Table S1 and Figure S1).

(F) A specific duplication, chr2:109,510,927–110,376,563 (including *EDAR*, LOEUF = 0.91; *SH3RF3*, LOEUF = 0.53; *SEPT10*, LOEUF = 1.17; and *SOWAHC*, LOEUF = 0.77), exhibited a previously unobserved positive effect on cognitive ability in the CNV-GWAS. See also Figure S1.

(G) To further investigate this positive effect, we conducted a post hoc analysis using a two-sided t test (mean ± standard error for cognitive ability) on a homogeneous cohort with consistent technology, ancestry, and phenotype, aiming to eliminate biases. The t test revealed a significant difference between the two groups: (1) individual without CNV vs. individual carrying duplication 2q12.3, t = −3.08, degree of freedom (df) = 18.01, $p = 0.006$, (2) individual without CNV vs. individual carrying exonic CNVs without duplication 2q12.3, t = 6.96, df = 34314, $p = 3.57 \times 10^{-12}$, and (3) individual carrying duplication 2q12.3 vs. individual carrying exonic CNVs without duplication 2q12.3, t = 3.31, df = 18.03, $p = 0.004$. Our focus was specifically on individuals of White British ethnicity in the UK Biobank (UKBB) with adjusted fluid intelligence (FI). In the left part of the analysis (G), individuals were categorized into three groups: carriers of the CNV of interest (green), non-carriers of this specific CNV but carrying other exonic CNVs (light orange), and non-carriers of any exonic CNV (blue). The t tests were performed on

(legend continued on next page)

exact p value corrected for false discovery rate [p_{FDR}] > 0.05). Results were not related to ancestry, array platform, or cognitive assessment methods (Figure 1G). The positive effect remained significant when comparing 2q12.3 duplication carriers to individuals without any CNVs. The reciprocal deletion in this region showed a trend toward a negative effect on cognitive ability (z = −0.526, SD = 0.276, p = 0.058), but we were underpowered with only 12 carriers. Additionally, the gene-dosage model showed a positive effect (z = 0.415, SD = 0.138, $p = 2.65 \times 10^{-3}$) on cognitive ability per number of copies (1, 2, or 3) at this locus. In other words, this may represent the first locus with a mirror impact on cognitive ability.

A large proportion of intolerant and tolerant genes modulate cognitive ability

Even with the current sample size, CNVs observed in >30 individuals (and included in the gene/CNV-level GWAS above) cover only 3%–4% of coding genes. However, previous studies have shown that a much larger proportion of the coding genome is involved in cognitive ability.[12] To test the association of all rare CNVs with cognition, we used burden association methods. We created 38 overlapping gene categories by sliding a window (defined by a width of 0.15 LOEUF units) by 0.05 LOEUF units 37 times (Figure 2B; STAR Methods, statistical model 2). We added a 39th category of known ID-associated genes (defined by ClinGen; Table S2). We calculated 39 burden effect sizes using linear models. To estimate the mean effect size of a gene in a given category and prevent the inflation of effect size due to multigenic CNVs, we adjusted for genes within CNVs that were not included in the LOEUF category of interest (cf. STAR Methods, statistical model 2; Figure 2A). The 39 estimates provided by the meta-analysis across the 9 sub-cohorts were not different from those provided by aggregating these datasets (Figure 2B; Tables S3 and S4). Therefore, all subsequent analyses were performed on the aggregated dataset. The effects of deletions were, on average, 2.4-fold higher than duplications, and we observed a positive correlation between the effect sizes of deletions and duplications across LOEUF categories (Spearman's r = 0.5, $p_{permutation}$ = 0.02; Figure S2). Negative effects on cognitive abilities were observed in 8 and 11 non-tolerant categories (LOEUF < 1) for deletions and duplications, respectively. The more intolerant the LOEUF category, the more negative the effect size, with the ID gene set having the largest effects. Of note, 2 and 3 categories showed positive effects for deletions and duplications, respectively. In other words, the effect sizes of these categories were significantly higher than the average effect of gene categories used to adjust the model. Sensitivity analyses showed no biases related to ancestry, large multigenic CNVs, or low-quality control scores (Figures 2C and S3). Effect sizes of intolerant genes were higher when removing older age groups (≥60 or ≥70 years old; Figure 2C). Because the most intolerant CNVs are depleted in the general population, we included 3 ASD cohorts in a sensitivity analysis. This resulted in larger effects and smaller p values for highly intolerant LOEUF categories without changing the effects of other LOEUF categories ≥0.35 (Figure 2C).

Negative correlation between deletion and duplication effects on cognitive ability across brain regions

Previously published functional enrichment analyses[28,29] have focused on recurrent CNVs. We therefore developed a functional-burden test to systematically investigate gene functions that may underlie the pervasive association between CNVs (too rare to reach individual association) and cognitive ability. The functional burden aggregates all CNVs disrupting genes involved in a given biological process. It provides the average effect on the cognitive ability of genes assigned to a biological function and is computed separately for deletions and duplications.

We tested 215 gene sets assigned to 215 adult brain regions. To define gene sets, we first normalized (Z scored) the expression of each gene across all 215 regions. For each tissue, the corresponding gene set was defined based on relative overexpression by selecting all genes with a Z scored expression ≥ 1 in that tissue. Among the 215 regional gene sets, 91 and 94 (mostly non-overlapping) were associated with cognitive ability when deleted or duplicated, respectively, but only 25 of these gene sets impacted cognition when disrupted by both CNVs (cf. STAR Methods; Figure 3A). This suggests that genes assigned to brain regions affect cognitive ability when either deleted or duplicated.

These preferential effects were supported by the negative correlation observed between the effect sizes of deletions and duplications across all brain regions (Spearman's r = −0.43, $p_{permutation} = 9 \times 10^{-3}$; Figure 3B). Stratifying these brain gene sets into 3 independent LOEUF categories provided the same negative correlations (Figure 3C). Sensitivity analysis showed that the negative correlation was not due to unbalanced power between deletions and duplications or the relative expression threshold used to define gene sets (Figure S4). Previous publications have reported that the effect size of gene dosage on cognitive ability is U-shaped,[31] with the effects of deletions being 2- to 3-fold higher than those of duplications.[11,12] Studies, however, have not been able to test whether genes show preferential effects on cognitive ability when either deleted or duplicated. We developed the trait-associated gene dosage sensitivity score (tagDS) to test whether the deletion/duplication effect size ratio of a given gene set deviates from the null distribution (average ratio of 2.4 in our dataset; cf. STAR Methods). This normalized value reflects preferential sensitivity to deletions or duplications for a specific phenotype. Positive or negative tagDS depicted ratios of effect sizes between deletions and duplications biased toward deletions or duplications, respectively (cf. STAR Methods; Figure 3D). tagDS values indicated that cerebral cortex gene sets

FI adjusted for sex, 1–10 principal component for ancestry, and age. In the right part of the analysis (F), two groups were defined: carriers of the CNV of interest (dark pink) and non-carriers (light blue). The t tests were conducted on FI adjusted for sex, ancestry, age, and the burden of 1/LOEUF for deletions and duplications. For the duplication of chr2:109,510,927–110,376,563 observed in the CNV-GWAS (G), the carriers exhibited significantly higher cognitive ability measures compared to both other groups (two-sided t test: t = −3.76, df = 18.01, p = 0.001). Furthermore, when we weighted the cognitive ability by the burden of 1/LOEUF for deletions and duplications, a positive effect was also observed among carriers of the CNV of interest.

Table 1. Cohort descriptions

Unselected cohorts (n = 258,292)	N	Ancestry EUR (others)	Gender (F/M)	Age mean year, (±SD)	Cognitive ability assessments
CaG	2,589	2,472 (117)	1,375/1,214	53.943 (7.845)	g-factor
G-Scot	13,715	13,672 (43)	8,081/5,634	46.730 (14.996)	g-factor
IMAGEN	1,744	1,624 (120)	891/853	14.450 (0.366)	WISC-IV
LBC1936	503	500 (3)	246/257	69.825 (0.829)	Moray House Test[26]
SYS	1,565	1,561 (4)	824/742	28.177 (17.098)	WISC-III or g-factor
UKBB	73,882	71,364 (2,518)	39,317/34,565	60.022 (8.959)	g-factor[27]
UKBB	62,080	60,484 (1,596)	34,335/27,745	62.083 (7.663)	g-factor (online)
UKBB	88,441	80,427 (8,014)	47,789/40,652	58.139 (8.304)	FI
UKBB	13,773	13,458 (315)	8,2845,489	64.185 (7.685)	FI (online)

Analyses were performed (after quality control [QC]) in 258,292 individuals from 6 general population cohorts. SYS, Saguenay Youth Study; CaG, CARTaGENE; LBC1936, Lothian Birth Cohort 1936; N, number of individuals remaining for analysis after quality control. See also Figure S15 and Tables S6 and S7.

affected cognitive ability preferentially when duplicated, while the opposite was observed for non-cortical (subcortical and midbrain) gene sets and deletions (Figure 3A; Mann-Whitney $p_{permutation} = 1 \times 10^{-15}$). The same cortical/non-cortical gene dosage sensitivity was also observed when removing genes with low tissue specificity (Figure S4).[32]

At the microstructure and cell type levels (6 cortical layers, 7 adult, and 16 fetal brain cell types, using the same method described above based on normalized gene expression; cf. STAR Methods), we observed the same negative correlation ($r = -0.70$, $p_{permutation} < 1 \times 10^{-3}$; Figure S5). The largest effects for deletions and duplications were observed in gene sets assigned to fetal cell types. Deletions and duplications, respectively, showed preferential effects in non-neuronal (endothelial, glia) and neuronal (excitatory) cell types (Figure 3E).

Genes preferentially expressed in non-brain tissues also affect cognitive ability

There is a growing interest in whole-body health comorbidities among individuals with neurodevelopmental and psychiatric conditions, as well as CNVs affecting cognition.[18,33] We therefore asked if CNVs affecting genes preferentially expressed in non-brain tissues (not part of the nervous central system) were also associated with cognitive ability.

We used 37 gene sets defined by relative expression (same methods used for brain regions and cell types) in 37 whole-body tissues (12 brain and 25 non-brain tissues [$\geq$1 SD]; Figure 4A). Many non-brain gene sets showed effect sizes (Figure 4B) of similar magnitude to those observed for regional brain gene sets. This was not explained by the level of overlap between brain and non-brain gene sets (Figures 4A and 4B). We observe the same pattern of deletion-duplication negative correlation independently of the gene set definitions ($r = -0.64$, $p_{permutation} < 1 \times 10^{-3}$; Figures 4C and S6). To understand how gene set definitions influence these results, we first removed 8,194 genes with low-tissue specificity assigned to multiple gene sets. The resulting effect sizes were correlated with the initial estimates ($r = 0.57$; Figure S7). In fact, genes assigned to multiple tissues show higher intolerance (LOEUF) compared to tissue-specific genes ($p = 1 \times 10^{-11}$–3×10^{-161}; Figure S8). To further investigate the impact on results of gene set definitions, we tested 37 previously published gene sets assigned to 37 GTEx tissues computed by the top decile expression proportion (TDEP) method (proportional gene expression).[34] This method, which emphasizes specificity, excludes 5,454 genes, of which 1,586 and 696 are, respectively, moderately intolerant to haploinsufficiency (LOEUF = [0.35, 1[) and highly intolerant to haploinsufficiency (LOEUF < 0.35; Figure S8). Effect sizes were well correlated with our analysis, excluding LTS genes ($r = 0.76$), but TDEP gene sets were unable to detect any effect for deletions across all tissues (Figure S7).

The effects of deletion and duplication on cognitive ability are negatively correlated across all levels of biological observations

We asked if the deletion-duplication negative correlations observed for tissue-level gene sets were also present at the molecular and cellular component levels. We first investigated 293 synaptic gene ontologies (GOs) using SynGO.[35] We observed that postsynaptic genes showed the largest negative effects on cognitive ability when deleted, and in contrast, presynaptic genes showed the largest negative effects when duplicated (Figures 5A and S9). As a result, the effects of the 2 opposing CNVs were negatively correlated across SynGO terms ($r = -0.39$, $p_{permutation} = 1 \times 10^{-3}$; Figure S10).

We extended our analysis to 6,130 GO terms (and corresponding gene sets); 5.0% and 3.5% of the GO terms had an effect size on cognitive ability for deletions and duplications, respectively. A minority (0.7%) of GO terms showed significant effects for both. We observed again a deletion-duplication negative correlation across GO term effect sizes ($r = -0.54$, $p_{permutation} < 1 \times 10^{-3}$; Figures 5B, S11, and S12), which remained significant across three independent levels of LOEUF stratification (Figure 5C). We asked if tagDS was similar to pHI (probability of haploinsufficiency) and pTS (probability of triplosensitivity), 2 previously published metrics that are highly correlated with each other (0.78) and with LOEUF ($r = 0.90$ and 0.77, respectively). tagDS was unrelated to pHI and pTS scores[36] across GO terms

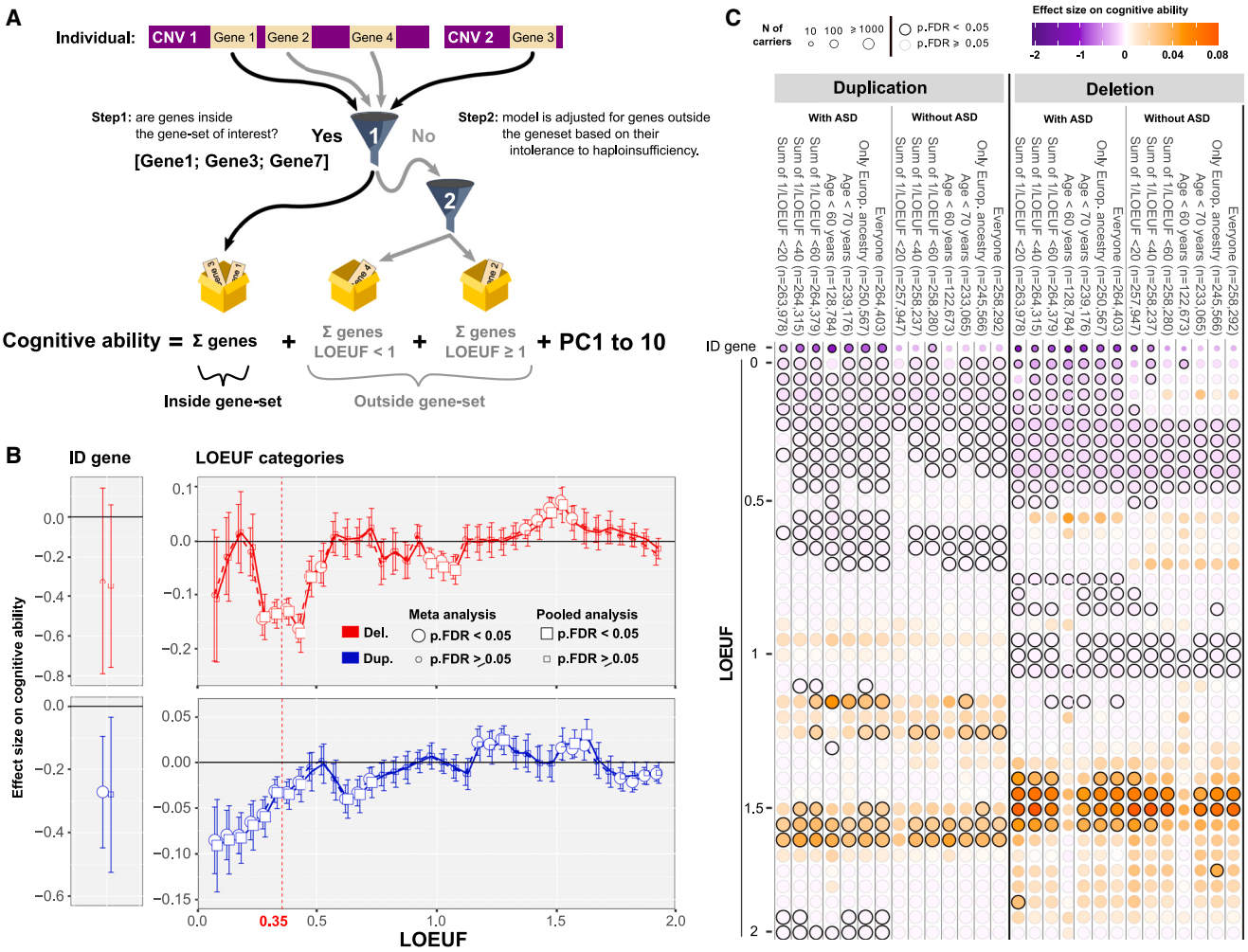


Figure 2. Effect sizes of autosomal coding genes on general cognitive abilities based on their LOEUF values

(A) The functional-burden test is a linear model estimating the mean effect size of all CNVs fully encompassing genes assigned to a biological function of interest. Because many CNVs are multigenic, the model is adjusted for genes included in a CNV but not assigned to the biological function of interest.

(B) Sliding window (STAR Methods, statistical model 2) estimating the mean effect size ± standard error on cognitive ability of deletions (top) and duplications (bottom) for 38 LOEUF categories (we slide a window of 0.15 LOEUF units in increments of 0.05 units, thereby creating 38 categories across the range of LOEUF values) and definitive ID genes curated by ClinGen. Estimates were computed using a meta-analysis (circles) as well as a pooled dataset (squares). The red dashed line defines intolerant genes (LOEUF < 0.35) (see also Tables S2, S3, and S4 and Figures S2 and S3).

(C) Heatmap showing the effect size (color scale) on cognitive ability of deletions and duplications across a range of sensitivity analyses removing non-Europeans, older participants (⩾60 or ⩾70 years old), large multigenic CNVs (those with a sum of 1/LOEUF >60, >40, and >20 corresponding to values of well-known recurrent CNVs: 22q11.2, 16p11.2, and TAR, respectively), as well as adding a neurodevelopmental dataset (autism spectrum disorder). All estimates were computed on the pooled dataset.

(Figure S13). Finally, several GO terms, such as neuronal, synaptic, and cardiac functions, showed preferential effects when deleted, while the opposite was observed for cellular response functions, transport, metabolic processes, and signaling pathways. Furthermore, "positive regulation" GO terms were more sensitive to deletions, while "negative regulation" terms showed preferential effects when duplicated (Figures 5D and S14).

DISCUSSION

In this large-scale CNV-GWAS on cognitive ability, we identified a duplication at 2q12.3 that is associated with higher cognitive ability. Although our sample size limited the discovery of new genome-wide signals at the variant level, we developed a functional-burden association test that allowed us to simultaneously test the contribution of all ultra-rare CNVs (covering 75% of the coding genome) and their function to cognitive ability. Constraint (LOEUF) and functional-burden analyses revealed that a substantial portion of the coding genome was associated with cognitive ability when deleted or duplicated. We also demonstrated that genes involved in a broad array of biological functions show preferential effects on cognitive ability when either deleted or duplicated. The latter was quantified by negative correlations between deletion and duplication effect sizes and by

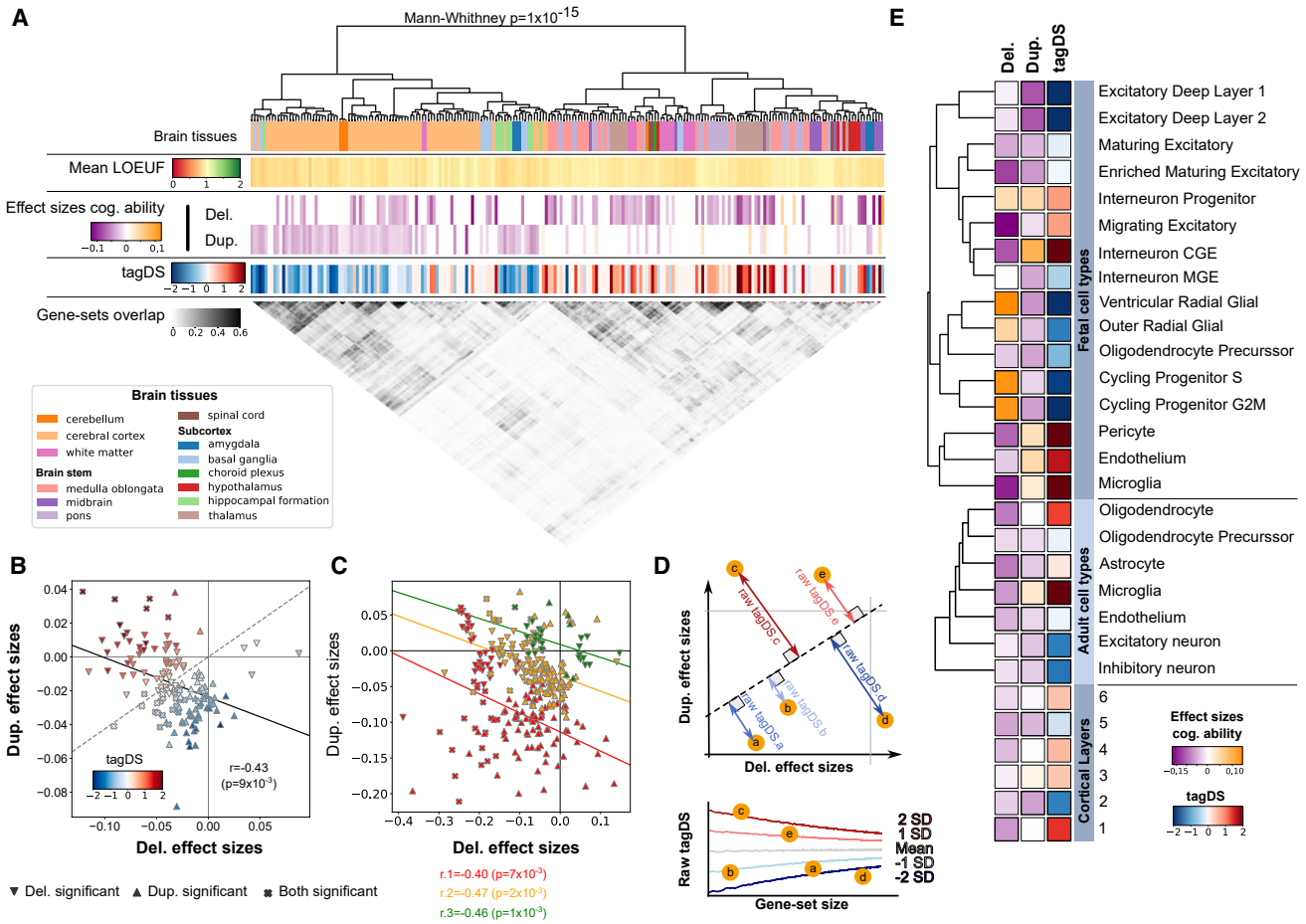


Figure 3. Effects on cognitive ability of genes assigned to brain regions and cell types

(A) Effect sizes on cognitive ability of gene sets assigned to 215 brain tissues/regions. Brain regions are color coded and clustered (first row, Ward's method[30]) based on the level of overlap (gray matrix) between their corresponding gene sets. The average LOEUF value for each gene sets is color coded in the second row. The mean effect sizes on the cognitive ability of genes assigned to each brain region are coded for deletions (third row) and duplications (fourth row). tagDS values are represented in the fifth row.

(B) Spearman correlation (black line) between the effect sizes of deletions and duplications across all gene sets with FDR significant effects on cognitive ability for either deletions (downward triangle), duplications (upward triangle), or both (cross). *p* values were obtained from permutations to account for the partial overlap between gene sets. Gene sets are color coded based on their tagDS. The dashed line represents the average exome-wide duplication/deletion effect size ratio (see also Figure S4).

(C) The same negative correlations between deletion and duplication were observed across 3 independent LOEUF groups: <0.35 (intolerant; red), [0.35, 1.0[(moderately intolerant; orange), and [1.0, 2.0] (tolerant; green).

(D) Raw tagDS is the Euclidean distance to the whole-genome ratio of effect sizes. tagDS is normalized following the null distribution of random gene sets of identical size.

(E) Effect size of deletions and duplications encompassing genes assigned to 6 cortical layers, 7 adult brain cell types, and 16 fetal brain cell types. Clustering was calculated on the level of overlap between cell type gene sets (Ward's method[30]). Purple and orange represent negative and positive effects on cognitive ability, respectively. Black edges indicate significant effects (see also Figure S5).

tagDS, a new normalized metric that assesses sensitivity to either deletions or duplications. We also show that genes assigned to non-brain tissues affected this "brain-centric" trait.

We identify, to our knowledge, the first CNV associated with higher cognitive ability. The 865 kb duplication (population frequency = ~1/7,200), which includes *EDAR*, *SH3RF3*, *SEPT10*, and *SOWAHC*, had not been previously associated with any trait or condition and showed a moderate effect size (z = 0.434, equivalent to 6.5 points of intelligence quotient [IQ]) on cognitive ability without significant heterogeneity across cohorts. Publications have identified associations between SNPs within this locus and 58 traits, including brain morphology,[37–39] schizophrenia,[40,41] Alzheimer's disease,[42] and neuroinflammatory biomarkers[43] (Table S5). An excess of *SEPT10 de novo* missense mutations have been reported in NDDs.[4] Given that the median age of our dataset is 60.7 years, it is possible that this duplication may be associated with a neuroprotective effect. We suspect that many more CNVs associated with higher cognitive ability will be identified in the future as sample sizes increase. Our functional-burden method identified gene sets with positive effects on cognitive

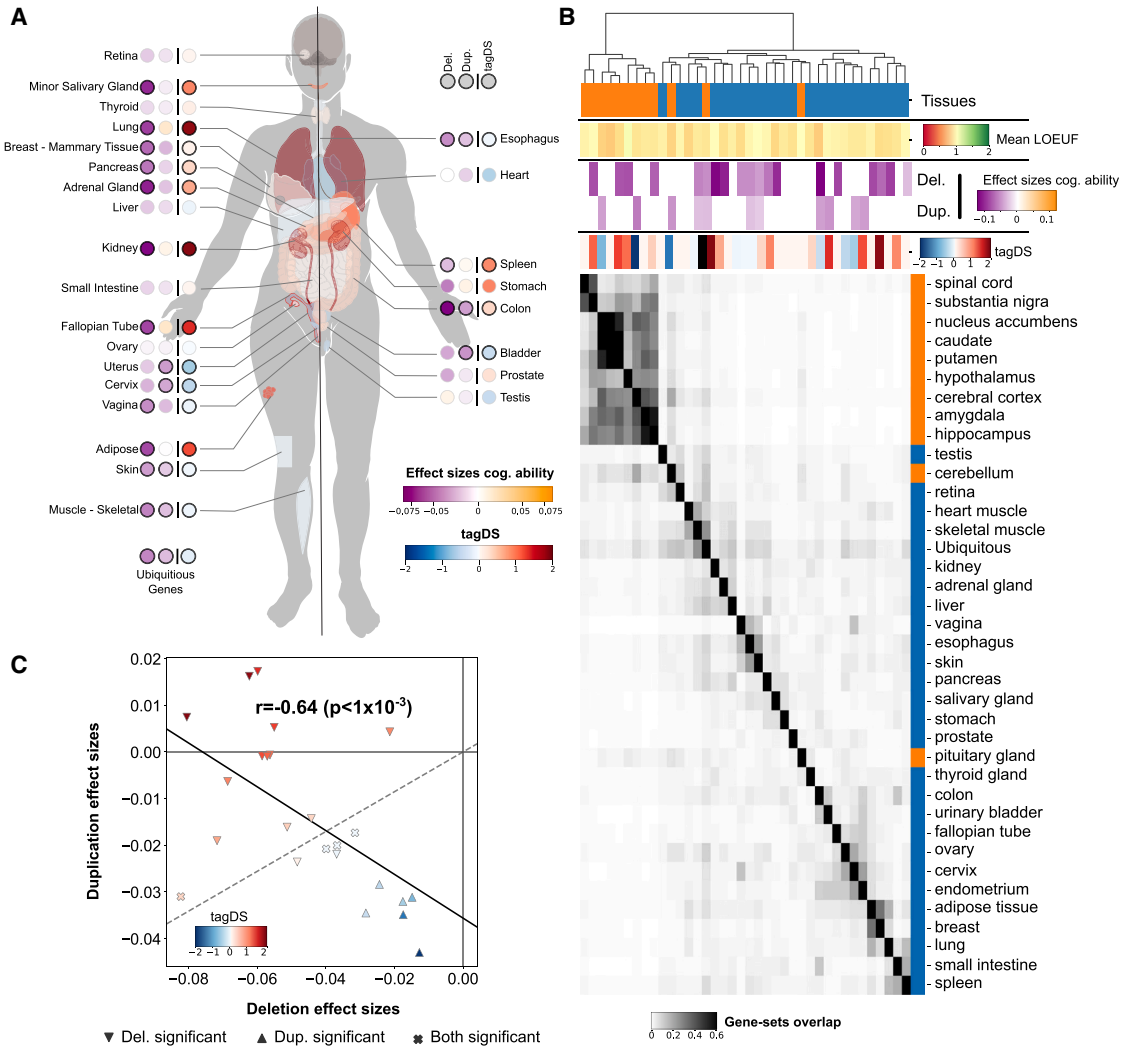


Figure 4. Effects on cognitive ability of CNVs affecting genes implicated in brain and non-brain tissues

(A) We defined 37 gene sets based on *Z* scored expression >1 SD. Expression of each gene was normalized across 37 tissues provided by GTEx. Gene sets were clustered (orange for brain tissues and blue for non-brain tissues) based on their overlap, which is shown in the grayscale matrix. High overlap was observed between brain gene sets (Ward's method[30]), and much lower overlap was present across non-brain tissues and between brain and non-brain tissues. The mean LOEUF of each gene set is color coded in the second row. Effect sizes on cognitive ability and tagDS across tissues are color coded in the third row as well as in the body map (B), adapted from GTEx. Genes with low tissue specificity were defined by the Human Protein Atlas.

(C) Spearman correlation (black line) between the effect sizes of deletions and duplications on cognitive ability. Downward and upward triangles and crosses represent significant effects for deletions, duplications, and both respectively. Gene sets are color coded based on their tagDS.

ability. Determining whether these gene sets truly increase cognitive ability or, instead, show smaller effects than the mean effect used to adjust for multigenic CNVs will require larger samples with data on CNVs disrupting single genes. Overall, the results suggest that gene dosage may be associated with a higher IQ, but most effects are masked by the multigenic nature of CNVs.

It has been challenging to evaluate haploinsufficiency and triplosensitivity. We show that tagDS for cognitive ability is orthogonal to genetic constraint, as well as previously published pHI and pTS measures. tagDS highlights sensitivity to either deletions or duplications across gene functions from macroscopic (cortical vs. non-cortical tissue) to microscopic (pre- vs. postsynaptic genes and positive vs. negative regulation) levels of observation.

Genetic covariance has almost exclusively been computed using common variants to investigate the genetic overlap between traits. While genetic covariance using rare variants is understudied due to a lack of statistical power, a recent study[44] aggregating rare variants at the gene level showed that the genetic correlation between protein loss-of-function and damaging missense variants associated with the same trait was, on average, 0.64 (with some correlations <0.5), implying that different classes of variants in the same genes may show different phenotypic effects.

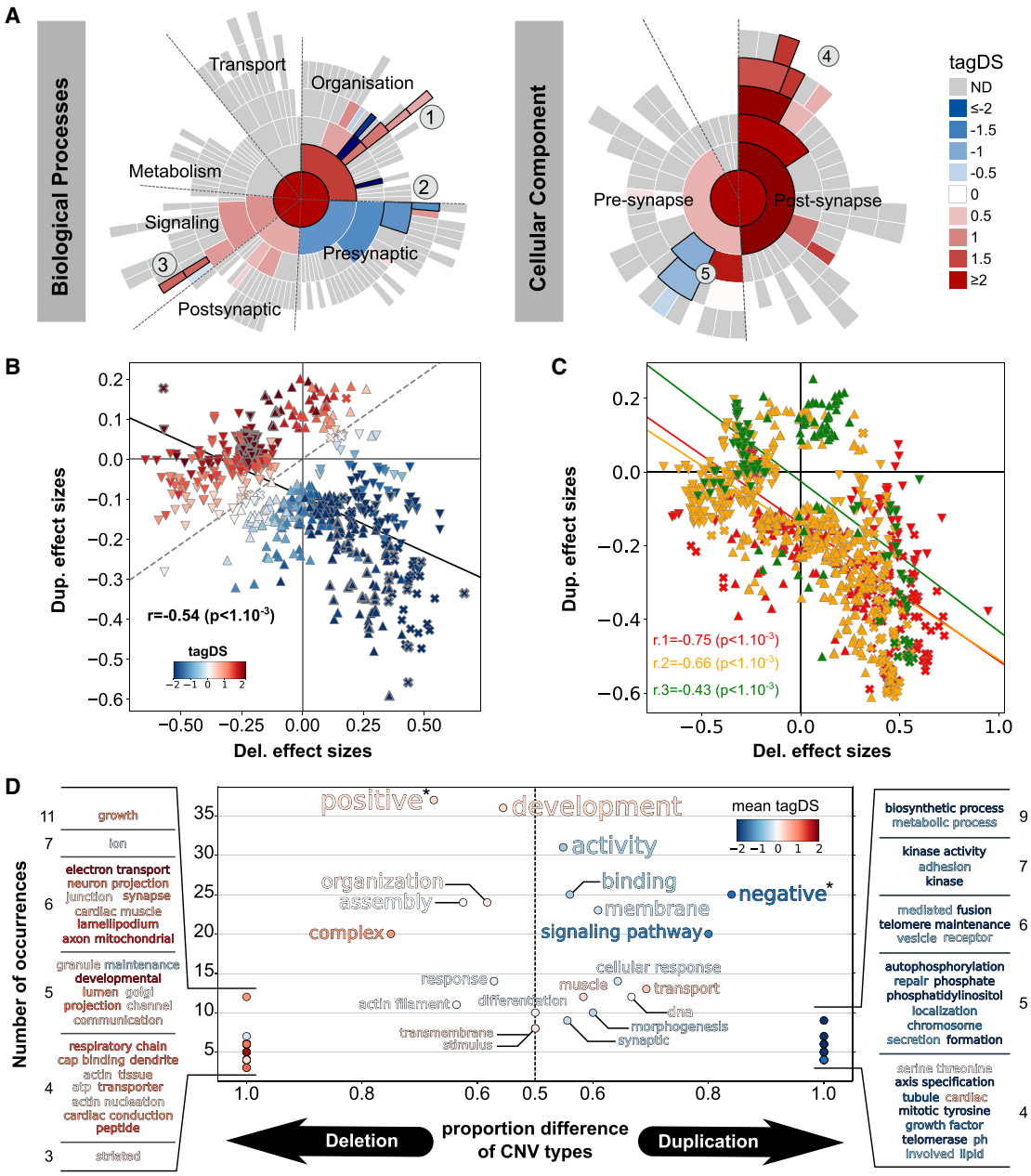


Figure 5. Effects on cognitive ability of gene sets based on GOs

(A) Effect sizes of synaptic molecular functions and cellular component gene sets as defined by SynGO[35] on cognitive ability (more details in Figure S9). Blue and red represent negative and positive cognitive ability tagDS, respectively. Ontologies with black edges indicate significant effects (FDR). The results are shown only for SynGO terms with more than 10 genes, observed at least 30 times in our dataset, and with a coverage greater than 20%. Note: (1) regulation of modification of postsynaptic actin cytoskeleton, (2) regulation of calcium-dependent activation of synaptic vesicle fusion, (3) presynaptic modulation of chemical synaptic transmission, (4) integral component of postsynaptic density membrane, and (5) synaptic vesicle membrane (see also Figures S9 and S10).

(B) There is a negative correlation (Spearman) between the effect sizes of deletions and duplication across 601 GO terms.

(C) The same deletion-duplication negative correlation was observed across 3 independent LOEUF groups (highly intolerant to haploinsufficiency <0.35: red, moderately intolerant to haploinsufficiency [0.35, 1.0[: orange, tolerant to haploinsufficiency [1.0, 2.0]: green).

(D) We adapted the word cloud package, which groups GO terms based on shared terminology. y axis: sum of associations of each word with significant deleted and duplicated GO terms. x axis: proportion of significant GO terms for a given CNV type used for the association. "Positive" and "negative" refer to "positive regulation" and "negative regulation," respectively (see also Figures S11, S12, and S14).

In our study, we show that two classes of variants with opposing molecular consequences have negatively correlated phenotypic effects. This negative correlation was observed regardless of whether CNVs were aggregated based on their function in tissues, cell types, or GO terms. This suggests that associating genes with traits or diseases is highly dependent on the class of genetic variants. Whether this negative correlation generalizes to other phenotypic traits is unknown.

There has been growing interest in the relationship between mental health and whole-body multi-morbidities. This is exemplified by the correlation between cognitive ability, medical conditions, such as coronary artery disease,[15,45] and longevity.[45,46] Recent studies also showed that poor physical health was more pronounced in neuropsychiatric illness than poor brain health.[33] In the current study, genes preferentially expressed in many non-brain organs show effects on cognition similar to those observed for brain tissue. The latter could not be explained by the level of overlap between brain and non-brain gene sets. However, our results suggest a trade-off of impact on cognitive ability between the intolerance to haploinsufficiency of genes and their tissue specificities. In other words, genes with lower tissue specificity and higher pleiotropy tend to have lower LOEUF values and therefore larger effect sizes on cognitive ability. Other interpretations include (1) gene-disrupting variants can alter non-brain organs, which in turn alter brain function due to suboptimal support, and (2) cognition is an embodied multi-organ trait includes both brain and non-brain organs. A whole-body contribution exists for other cognitive-modulating traits such as sleep (thought to be for and by the brain), which is also regulated by peripheral tissue.[47]

The main limitation of this study is the use of gene sets, which were defined either on the basis of well-established ontologies or using a "relative method" based on normalized expression values. In the latter approach, we chose thresholds that may have influenced our results. Multiple sensitivity analyses demonstrated that changing the threshold (and therefore the size of the gene set) did not influence our main findings. Expression profiles vary across space, cell types, and time for a given tissue. Our gene sets could not explore all of these aspects. Larger studies will be required to increase the granularity of these functional burdens on association tests.

In conclusion, our study demonstrated, for the first time, the positive effects of a CNV on cognitive abilities. We present a new approach to functionally aggregate rare and ultra-rare variants and uncover many gene functions that are preferentially sensitive to either deletions or duplications. Computing tagDS for other complex traits will help understand whether sensitivity to gene dosage is trait dependent.

RESOURCE AVAILABILITY

Lead contact

For additional information, as well as requests regarding resources, please direct your inquiries to the lead contact, Sébastien Jacquemont (sebastien.jacquemont@umontreal.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All general population data are available to other investigators online: IMAGEN: https://www.cataloguementalhealth.ac.uk, LBC: https://lothian-birth-cohorts.ed.ac.uk/, SYS (contact: T.P., tomas.paus@umontreal.ca), CaG: https://portal.canpath.ca/, Generation Scotland: https://www.ed.ac.uk/generation-scotland, and the UK Biobank: https://www.ukbiobank.ac.uk. All ASD population data are available to other investigators online: SSC: https://www.sfari.org/, SPARK: https://www.sfari.org/, and MSSNG: https://research.mss.ng/. All derived measures used in this study are available upon request (S.J., sebastien.jacquemont@umontreal.ca). The rest of the CNV carriers' data cannot be shared, as participants did not provide consent. Summary statistics and the gene sets used to compute them have been deposited on FigShare (see key resources table). All original scripts have been deposited and are publicly available as of the date of publication on GitHub repositories: (1) quality control and annotation of CNVs: https://martineaujeanlouis.github.io/MIND-GENESPARALLELCNV/, (2) CNV validation ("DigCNV"): https://github.com/labjacquemont/DigCNV, and (3) statistics and visualizations: https://github.com/labjacquemont/CNV_cognitive_ability.

ACKNOWLEDGMENTS

This research was enabled by support provided by Calcul Quebec (http://www.calculquebec.ca) and Compute Canada (http://www.computecanada.ca). S.J. is a recipient of a Canada Research Chair in neurodevelopmental disorders and a chair from the Jeanne et Jean Louis Levesque Foundation. This work is supported by a grant from the Brain Canada Multi-Investigator Initiative and CIHR grant 159734 (S.J., T.P.). 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 (T.P., Z.P.) and the Heart and Stroke Foundation of Canada (Z.P.). Funding for the project was provided by the Wellcome Trust. This work was also supported by NIH award U01 MH119690 granted to L.A., S.J., and D.C.G. and U01 MH119739. LBC1936 is supported by the Biotechnology and Biological Sciences Research Council and the Economic and Social Research Council (BB/W008793/1; which supports S.E.H.), Age UK (Disconnected Mind project), the Milton Damerel Trust, and the University of Edinburgh. S.R.C. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (221890/Z/20/Z). Genotyping was funded by the BBSRC (BB/F019394/1). This research has been conducted using the UK Biobank Resource under application no. 40980. The Cardiff Copy-Number Variant cohort (UKBB) was supported by the Wellcome Trust Strategic Award DEFINE and the National Center for Mental Health with funds from Health and Care Research Wales (code 100202/Z/12/Z). Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006) and is currently supported by the Wellcome Trust (216767/Z/19/Z). Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland, and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "Stratifying Resilience and Depression Longitudinally" (STRADL) (reference 104036/Z/14/Z). IMAGEN received support from the following sources: the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT-2007-037286); the Horizon 2020-funded ERC Advanced Grant "STRATIFY" (Brain network based stratification of reinforcement-related disorders) (695313); the Human Brain Project (HBP SGA 2, 785907, and HBP SGA 3, 945539); the Medical Research Council Grant "c-VEDA" (Consortium on Vulnerability to Externalizing Disorders and Addictions) (MR/N000390/1); the NIH (R01DA049238, A decentralized macro and micro gene-by-environment interaction analysis of substance use behavior and its brain biomarkers); the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London; the Bundesministerium für Bildung und Forschung (BMBF grants 01GS08152 and 01EV0711; Forschungsnetz AERIAL 01EE1406A and 01EE1406B; Forschungsnetz IMAC-Mind 01GL1745B); the Deutsche Forschungsgemeinschaft (DFG grants SM 80/7-2, SFB 940, TRR 265, and NE 1383/14-1); the Medical Research

Foundation and Medical Research Council (grants MR/R00465X/1 and MR/S020306/1); and the NIH-funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01). Further support was provided by grants from the ANR (ANR-12-SAMA-0004 and AAPG2019 – GeBra), the ERA-Net NEURON (AF12-NEUR0008-01 – WM2NA and ANR-18-NEUR00002-01 – ADORe), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris-Sud University IDEX 2012, the Fondation de l'Avenir (grant AP-RM-17-013), the Fédération pour la Recherche sur le Cerveau, Science Foundation Ireland (16/ERCD/3797), the NIH (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1), and NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence. The authors wish to acknowledge the resources of MSSNG (www.mss.ng), Autism Speaks, and The Center for Applied Genomics at The Hospital for Sick Children, Toronto, Canada. We also thank the participating families for their time and contributions to this database, as well as the generosity of the donors who supported this program. We thank the coordinators and staff at the SCC sites. We are grateful to all of the families at the participating SSC sites and the principal investigators (A. Beaudet, MD; R. Bernier, PhD; J. Constantino, MD; E. Cook, MD; E. Fombonne, MD; D. Geschwind, MD, PhD; R. Goin-Kochel, PhD; E. Hanson, PhD; D. Grice, MD; A. Klin, PhD; D. Ledbetter, PhD; C. Lord, PhD; C. Martin, PhD; D. Martin, MD, PhD; R. Maxim, MD; J. Miles, MD, PhD; O. Ousley, PhD; K. Pelphrey, PhD; B. Peterson, MD; J. Piggot, MD; C. Saulnier, PhD; M. State, MD, PhD; W. Stone, PhD; J. Sutcliffe, PhD; C. Walsh, MD, PhD; Z. Warren, PhD; and E. Wijsman, PhD). We appreciate obtaining access to phenotypic data on SFARI base. 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.

AUTHOR CONTRIBUTIONS

Conceptualization, G.H., T.R., and S.J.; data curation, G.H., M.J.-L., Z.S., and E.D.; formal analysis, G.H., T.R., C. Poulain, and A.D.; funding acquisition, L.A., D.C.G., and S.J.; investigation, G.H., T.R., C. Poulain, and A.D.; methodology, G.H., M.J.-L., Z.S., E.D., T.R., C. Poulain, C. Proulx, and A.D.; project administration, G.H. and S.J.; resources, G.H., M.J.-L., Z.S., T.R., and C. Poulain; software: G.H., M.J.-L., Z.S., E.D., T.R., C. Poulain, and C. Proulx; supervision: G.H. and S.J.; validation, G.H., T.R., and S.J.; visualization, G.H., T.R., and S.J.; writing – original draft, G.H., T.R., and S.J.; writing – review & editing, G.H., T.R., C. Poulain, A.D., K.K., S.K., W.E., O.S., E.D., C. Proulx, M.J.-L., Z.S., J.M., L.M.S., E.E.M.K., S.R.C., D.P., G.D., P.R., S.E.H., G.S., G.D., A.L., Z.P., T.P., S.W.S., J.S., L.A., D.C.G., and S.J.

DECLARATION OF INTERESTS

The authors declare that they have no conflicts of interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2024.100721.

Received: April 19, 2024
Revised: July 30, 2024
Accepted: November 13, 2024
Published: December 11, 2024

REFERENCES

1. Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. *7*, 85–97. https://doi.org/10.1038/nrg1767.
2. Zarrei, M., Burton, C.L., Engchuan, W., Higginbotham, E.J., Wei, J., Shaikh, S., Roslin, N.M., MacDonald, J.R., Pellecchia, G., Nalpathamkalam, T., et al. (2023). Gene copy number variation and pediatric mental health/neurodevelopment in a general population. Hum. Mol. Genet. *32*, 2411–2421. https://doi.org/10.1093/hmg/ddad074.
3. Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W.M., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E.L.M., et al. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. *46*, 1063–1071. https://doi.org/10.1038/ng.3092.
4. Coe, B.P., Stessman, H.A.F., Sulovari, A., Geisheker, M.R., Bakken, T.E., Lake, A.M., Dougherty, J.D., Lein, E.S., Hormozdiari, F., Bernier, R.A., and Eichler, E.E. (2019). Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. Nat. Genet. *51*, 106–116. https://doi.org/10.1038/s41588-018-0288-4.
5. Wilfert, A.B., Sulovari, A., Turner, T.N., Coe, B.P., and Eichler, E.E. (2017). Recurrent de novo mutations in neurodevelopmental disorders: properties and clinical implications. Genome Med. *9*, 101. https://doi.org/10.1186/s13073-017-0498-x.
6. Huguet, G., Ey, E., and Bourgeron, T. (2013). The genetic landscapes of autism spectrum disorders. Annu. Rev. Genomics Hum. Genet. *14*, 191–213. https://doi.org/10.1146/annurev-genom-091212-153431.
7. Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., et al. (2014). Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. Am. J. Hum. Genet. *94*, 677–694. https://doi.org/10.1016/j.ajhg.2014.03.018.
8. Maillard, A.M., Ruef, A., Pizzagalli, F., Migliavacca, E., Hippolyte, L., Adaszewski, S., Dukart, J., Ferrari, C., Conus, P., Männik, K., et al. (2015). The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. Mol. Psychiatry *20*, 140–147. https://doi.org/10.1038/mp.2014.145.
9. Sakai, M., Watanabe, Y., Someya, T., Araki, K., Shibuya, M., Niizato, K., Oshima, K., Kunii, Y., Yabe, H., Matsumoto, J., et al. (2015). Assessment of copy number variations in the brain genome of schizophrenia patients. Mol. Cytogenet. *8*, 46. https://doi.org/10.1186/s13039-015-0144-5.
10. Szatkiewicz, J.P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J.L., Neale, B.M., Fromer, M., Ruderfer, D., Akterin, S., Bergen, S.E., et al. (2014). Copy number variation in schizophrenia in Sweden. Mol. Psychiatry *19*, 762–773. https://doi.org/10.1038/mp.2014.40.

11. Huguet, G., Schramm, C., Douard, E., Jiang, L., Labbe, A., Tihy, F., Mathonnet, G., Nizard, S., Lemyre, E., Mathieu, A., et al. (2018). Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. JAMA Psychiatr. *75*, 447–457. https://doi.org/10.1001/jamapsychiatry.2018.0039.

12. Huguet, G., Schramm, C., Douard, E., Tamer, P., Main, A., Monin, P., England, J., Jizi, K., Renne, T., Poirier, M., et al. (2021). Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability. Mol. Psychiatry *26*, 2663–2676. https://doi.org/10.1038/s41380-020-00985-z.

13. Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Walters, G.B., Jonsdottir, G.A., Doyle, O.M., et al. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature *505*, 361–366. https://doi.org/10.1038/nature12818.

14. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). 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. https://doi.org/10.1016/j.ajhg.2010.04.006.

15. Deary, I.J. (2012). Intelligence. Annu. Rev. Psychol. *63*, 453–482. https://doi.org/10.1146/annurev-psych-120710-100353.

16. Mollon, J., Schultz, L.M., Huguet, G., Knowles, E.E.M., Mathias, S.R., Rodrigue, A., Alexander-Bloch, A., Saci, Z., Jean-Louis, M., Kumar, K., et al. (2023). Impact of Copy Number Variants and Polygenic Risk Scores on Psychopathology in the UK Biobank. Biological Psychiatry *0*. Biol. Psychiatry *94*, 591–600. https://doi.org/10.1016/j.biopsych.2023.01.028.

17. Kendall, K.M., Bracher-Smith, M., Fitzpatrick, H., Lynham, A., Rees, E., Escott-Price, V., Owen, M.J., O'Donovan, M.C., Walters, J.T.R., and Kirov, G. (2019). Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. Br. J. Psychiatry *214*, 297–304. https://doi.org/10.1192/bjp.2018.301.

18. Auwerx, C., Lepamets, M., Sadler, M.C., Patxot, M., Stojanov, M., Baud, D., Mägi, R., Estonian Biobank Research Team; Porcu, E., Reymond, A., and Kutalik, Z. (2022). The individual and global impact of copy-number variants on complex human traits. Am. J. Hum. Genet. *109*, 647–668. https://doi.org/10.1016/j.ajhg.2022.02.010.

19. Wainberg, M., Merico, D., Huguet, G., Zarrei, M., Jacquemont, S., Scherer, S.W., and Tripathy, S.J. (2022). Deletion of Loss-of-Function–Intolerant Genes and Risk of 5 Psychiatric Disorders. JAMA Psychiatr. *79*, 78–81. https://doi.org/10.1001/jamapsychiatry.2021.3211.

20. Alexander-Bloch, A., Huguet, G., Schultz, L.M., Huffnagle, N., Jacquemont, S., Seidlitz, J., Saci, Z., Moore, T.M., Bethlehem, R.A.I., Mollon, J., et al. (2022). Copy Number Variant Risk Scores Associated With Cognition, Psychopathology, and Brain Structure in Youths in the Philadelphia Neurodevelopmental Cohort. JAMA Psychiatr. *79*, 699–709. https://doi.org/10.1001/jamapsychiatry.2022.1017.

21. Douard, E., Zeribi, A., Schramm, C., Tamer, P., Loum, M.A., Nowak, S., Saci, Z., Lord, M.-P., Rodríguez-Herreros, B., Jean-Louis, M., et al. (2021). Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome. Am. J. Psychiatry *178*, 87–98. https://doi.org/10.1176/appi.ajp.2020.19080834.

22. CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. *49*, 27–35. https://doi.org/10.1038/ng.3725.

23. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell *180*, 568–584.e23. https://doi.org/10.1016/j.cell.2019.12.036.

24. Wilfert, A.B., Turner, T.N., Murali, S.C., Hsieh, P., Sulovari, A., Wang, T., Coe, B.P., Guo, H., Hoekzema, K., Bakken, T.E., et al. (2021). Recent ultra-rare inherited variants implicate novel autism candidate risk genes. Nat. Genet. *53*, 1125–1134. https://doi.org/10.1038/s41588-021-00899-8.

25. Weiner, D.J., Ling, E., Erdin, S., Tai, D.J.C., Yadav, R., Grove, J., Fu, J.M., Nadig, A., Carey, C.E., Baya, N., et al. (2022). Statistical and functional convergence of common and rare genetic influences on autism at chromosome 16p. Nat. Genet. *54*, 1630–1639. https://doi.org/10.1038/s41588-022-01203-y.

26. Harris, S.E., Marioni, R.E., Martin-Ruiz, C., Pattie, A., Gow, A.J., Cox, S.R., Corley, J., von Zglinicki, T., Starr, J.M., and Deary, I.J. (2016). Longitudinal telomere length shortening and cognitive and physical decline in later life: The Lothian Birth Cohorts 1936 and 1921. Mech. Ageing Dev. *154*, 43–48. https://doi.org/10.1016/j.mad.2016.02.004.

27. Lyall, D.M., Cullen, B., Allerhand, M., Smith, D.J., Mackay, D., Evans, J., Anderson, J., Fawns-Ritchie, C., McIntosh, A.M., Deary, I.J., and Pell, J.P. (2016). Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. PLoS One *11*, e0154222. https://doi.org/10.1371/journal.pone.0154222.

28. Tai, D.J.C., Razaz, P., Erdin, S., Gao, D., Wang, J., Nuttle, X., de Esch, C.E., Collins, R.L., Currall, B.B., O'Keefe, K., et al. (2022). Tissue- and cell-type-specific molecular and functional signatures of 16p11.2 reciprocal genomic disorder across mouse brain and human neuronal models. Am. J. Hum. Genet. *109*, 1789–1813. https://doi.org/10.1016/j.ajhg.2022.08.012.

29. Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F., et al. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes associated with the 16p11.2 CNV. Nature *485*, 363–367. https://doi.org/10.1038/nature11091.

30. Ward, J.H. (1963). Hierarchical Grouping to Optimize an Objective Function. J. Am. Stat. Assoc. *58*, 236–244. https://doi.org/10.1080/01621459.1963.10500845.

31. Auwerx, C., Jõeloo, M., Sadler, M.C., Tesio, N., Ojavee, S., Clark, C.J., Mägi, R., Estonian Biobank Research Team; Reymond, A., and Kutalik, Z. (2024). Rare copy-number variants as modulators of common disease susceptibility. Genome Med. *16*, 5. https://doi.org/10.1186/s13073-023-01265-5.

32. Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., et al. (2020). An atlas of the protein-coding genes in the human, pig, and mouse brain. Science *367*, eaay5947. https://doi.org/10.1126/science.aay5947.

33. Tian, Y.E., Di Biase, M.A., Mosley, P.E., Lupton, M.K., Xia, Y., Fripp, J., Breakspear, M., Cropley, V., and Zalesky, A. (2023). Evaluation of Brain-Body Health in Individuals With Common Neuropsychiatric Disorders. JAMA Psychiatr. *80*, 567–576. https://doi.org/10.1001/jamapsychiatry.2023.0791.

34. Bryois, J., Skene, N.G., Hansen, T.F., Kogelman, L.J.A., Watson, H.J., Liu, Z., Eating Disorders Working Group of the Psychiatric Genomics Consortium; International Headache Genetics Consortium; 23andMe Research Team; and Brueggeman, L., et al. (2020). Genetic Identification of Cell Types Underlying Brain Complex Traits Yields Insights Into the Etiology of Parkinson's Disease. Nat. Genet. *52*, 482–493. https://doi.org/10.1038/s41588-020-0610-9.

35. Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., et al. (2019). SynGO: an evidence-based, expert-curated knowledgebase for the synapse. Neuron *103*, 217–234.e4. https://doi.org/10.1016/j.neuron.2019.05.002.

36. Collins, R.L., Glessner, J.T., Porcu, E., Lepamets, M., Brandon, R., Lauricella, C., Han, L., Morley, T., Niestroj, L.-M., Ulirsch, J., et al. (2022). A cross-disorder dosage sensitivity map of the human genome. Cell *185*, 3041–3055.e25. https://doi.org/10.1016/j.cell.2022.06.036.

37. van der Meer, D., Kaufmann, T., Shadrin, A.A., Makowski, C., Frei, O., Roelfs, D., Monereo-Sánchez, J., Linden, D.E.J., Rokicki, J., Alnæs, D.,

et al. (2021). The genetic architecture of human cortical folding. Sci. Adv. *7*, eabj9446. https://doi.org/10.1126/sciadv.abj9446.

38. Shadrin, A.A., Kaufmann, T., van der Meer, D., Palmer, C.E., Makowski, C., Loughnan, R., Jernigan, T.L., Seibert, T.M., Hagler, D.J., Smeland, O.B., et al. (2021). Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. Neuroimage *244*, 118603. https://doi.org/10.1016/j.neuroimage.2021.118603.
39. van der Meer, D., Frei, O., Kaufmann, T., Shadrin, A.A., Devor, A., Smeland, O.B., Thompson, W.K., Fan, C.C., Holland, D., Westlye, L.T., et al. (2020). Understanding the genetic determinants of the brain with MOSTest. Nat. Commun. *11*, 3512. https://doi.org/10.1038/s41467-020-17368-1.
40. Wu, Y., Cao, H., Baranova, A., Huang, H., Li, S., Cai, L., Rao, S., Dai, M., Xie, M., Dou, Y., et al. (2020). Multi-trait analysis for genome-wide association study of five psychiatric disorders. Transl. Psychiatry *10*, 209–211. https://doi.org/10.1038/s41398-020-00902-6.
41. Genome-wide association study of schizophrenia in Ashkenazi Jews - Goes - 2015 - American Journal of Medical Genetics Part B: Neuropsychiatric Genetics - Wiley Online Library https://onlinelibrary.wiley.com/doi/10.1002/ajmg.b.32349.
42. Naj, A.C., Beecham, G.W., Martin, E.R., Gallins, P.J., Powell, E.H., Konidari, I., Whitehead, P.L., Cai, G., Haroutunian, V., Scott, W.K., et al. (2010). Dementia Revealed: Novel Chromosome 6 Locus for Late-Onset Alzheimer Disease Provides Genetic Evidence for Folate-Pathway Abnormalities. PLoS Genet. *6*, e1001130. https://doi.org/10.1371/journal.pgen.1001130.
43. Liu, C., and Yu, J. (2019). Genome-Wide Association Studies for Cerebrospinal Fluid Soluble TREM2 in Alzheimer's Disease. Front. Aging Neurosci. *11*, 297. https://doi.org/10.3389/fnagi.2019.00297.
44. Weiner, D.J., Nadig, A., Jagadeesh, K.A., Dey, K.K., Neale, B.M., Robinson, E.B., Karczewski, K.J., and O'Connor, L.J. (2023). Polygenic architecture of rare coding variation across 394,783 exomes. Nature *614*, 492–499. https://doi.org/10.1038/s41586-022-05684-z.
45. Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat. Genet. *50*, 912–919. https://doi.org/10.1038/s41588-018-0152-6.
46. Plomin, R., and von Stumm, S. (2018). The new genetics of intelligence. Nat. Rev. Genet. *19*, 148–159. https://doi.org/10.1038/nrg.2017.104.
47. Kawano, T., Kashiwagi, M., Kanuka, M., Chen, C.-K., Yasugaki, S., Hatori, S., Miyazaki, S., Tanaka, K., Fujita, H., Nakajima, T., et al. (2023). ER proteostasis regulators cell-non-autonomously control sleep. Cell Rep. *42*, 112267. https://doi.org/10.1016/j.celrep.2023.112267.
48. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. *12*, e1001779. https://doi.org/10.1371/journal.pmed.1001779.
49. Deary, I.J., Gow, A.J., Taylor, M.D., Corley, J., Brett, C., Wilson, V., Campbell, H., Whalley, L.J., Visscher, P.M., Porteous, D.J., and Starr, J.M. (2007). The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. BMC Geriatr. *7*, 28. https://doi.org/10.1186/1471-2318-7-28.
50. Pausova, Z., Paus, T., Abrahamowicz, M., Bernard, M., Gaudet, D., Leonard, G., Peron, M., Pike, G.B., Richer, L., Séguin, J.R., and Veillette, S. (2017). Cohort Profile: The Saguenay Youth Study (SYS). Int. J. Epidemiol. *46*, e19. https://doi.org/10.1093/ije/dyw023.
51. Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P.J., Dalley, J.W., Flor, H., Gallinat, J., et al. (2010). The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol. Psychiatry *15*, 1128–1139. https://doi.org/10.1038/mp.2010.4.
52. Awadalla, P., Boileau, C., Payette, Y., Idaghdour, Y., Goulet, J.-P., Knoppers, B., Hamet, P., and Laberge, C.; CARTaGENE Project (2013). Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. Int. J. Epidemiol. *42*, 1285–1299. https://doi.org/10.1093/ije/dys160.
53. Smith, B.H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S.M., Deary, I.J., MacIntyre, D.J., Campbell, H., McGilchrist, M., et al. (2013). Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int. J. Epidemiol. *42*, 689–700. https://doi.org/10.1093/ije/dys084.
54. C Yuen, R.K., Merico, D., Bookman, M., L Howe, J., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat. Neurosci. *20*, 602–611. https://doi.org/10.1038/nn.4524.
55. Fischbach, G.D., and Lord, C. (2010). The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. Neuron *68*, 192–195. https://doi.org/10.1016/j.neuron.2010.10.006.
56. Feliciano, P., Daniels, A.M., Snyder, L.G., Beaumont, A., Camba, A., Esler, A., Gulsrud, A.G., Mason, A., Gutierrez, A., Nicholson, A., et al. (2018). SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. Neuron *97*, 488–493. https://doi.org/10.1016/j.neuron.2018.01.015.
57. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
58. Martin, F.J., Amode, M.R., Aneja, A., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., et al. (2023). Ensembl 2023. Nucleic Acids Res. *51*, D933–D941. https://doi.org/10.1093/nar/gkac958.
59. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single-cell type transcriptomics map of human tissues. Sci. Adv. *7*, eabh2169. https://doi.org/10.1126/sciadv.abh2169.
60. Wagstyl, K., Adler, S., Seidlitz, J., Vandekar, S., Mallard, T.T., Dear, R., DeCasien, A.R., Satterthwaite, T.D., Liu, S., Vértes, P.E., et al. (2024). Transcriptional Cartography Integrates Multiscale Biology of the Human Cortex. Elife *12*, RP86933. https://doi.org/10.7554/eLife.86933.2.
61. Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., and Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. *35*, 2013–2025.
62. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., and Bucan, M. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. *17*, 1665–1674. https://doi.org/10.1101/gr.6861907.
63. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent de novo copy number variations (CNVs), including duplications of the 7q11.23 Williams-Beuren syndrome region, are strongly associated with autism. Neuron *70*, 863–885. https://doi.org/10.1016/j.neuron.2011.05.002.
64. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
65. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

66. Wain, L.V., Shrine, N., Miller, S., Jackson, V.E., Ntalla, I., Soler Artigas, M., Billington, C.K., Kheirallah, A.K., Allen, R., Cook, J.P., et al. (2015). Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir. Med. *3*, 769–781. https://doi.org/10.1016/S2213-2600(15)00283-0.

67. Deary, I.J., Cox, S.R., and Hill, W.D. (2022). Genetic variation, brain, and intelligence differences. Mol. Psychiatry *27*, 335–353. https://doi.org/10.1038/s41380-021-01027-y.

68. Hampshire, A., Highfield, R.R., Parkin, B.L., and Owen, A.M. (2012). Fractionating Human Intelligence. Neuron *76*, 1225–1237. https://doi.org/10.1016/j.neuron.2012.06.022.

69. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.

70. Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., and Zhang, K. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat. Biotechnol. *36*, 70–80. https://doi.org/10.1038/nbt.4038.

71. Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippani, M., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat. Neurosci. *24*, 425–436. https://doi.org/10.1038/s41593-020-00787-0.

72. Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A.G., Shi, X., Stein, J.L., Vuong, C.K., Nichterwitz, S., Gevorgian, M., Opland, C.K., et al. (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron *103*, 785–801.e8. https://doi.org/10.1016/j.neuron.2019.06.011.

73. Gene Ontology Consortium; Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., et al. (2023). The Gene Ontology knowledgebase in 2023. Genetics *224*, iyad031. https://doi.org/10.1093/genetics/iyad031.

74. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
UKBB raw data	Sudlow et al.[48]	https://www.ukbiobank.ac.uk/
Lothian Birth Cohort raw data	Deary et al.[49]	https://lothian-birth-cohorts.ed.ac.uk
Saguenay Youth Study raw data	Pausova et al.[50]	https://saguenay-youth-study.org/
Imagen raw data	Schumann et al.[51]	https://imagen.squarespace.com/
CartaGene	Awadalla et al.[52]	https://cartagene.qc.ca/
Generation Scotland raw data	Smith et al.[53]	https://genscot.igc.ed.ac.uk/welcome
MSSNG raw data	Yuen et al.[54]	https://research.mss.ng/
SSC raw data	Fischbach et al.[55]	https://www.sfari.org/
SPARK raw data	Feliciano et al.[56]	https://sparkforautism.org/
gnomAD v2	Karczewski et al.[57]	https://gnomad.broadinstitute.org/
Ensembl v109	Martin et al.[58]	https://www.ensembl.org/
Syngo Release 2021	Koopmans et al.[35]	https://www.syngoportal.org/
HPA v22	Sjöstedt et al.[32]	https://www.proteinatlas.org/
GTEx v8	Karlsson et al.[59]	https://gtexportal.org/home/
Brain cell types	Wagstyl et al.[60]	https://doi.org/10.7554/eLife.86933.2
Summary statistics data	This paper	10.6084/m9.figshare.27350322
Created gene-sets	This paper	10.6084/m9.figshare.27360612
Software and algorithms		
Pipeline for CNV quality control and annotation	Huguet et al.[12]	https://martineaujeanlouis.github.io/MIND-GENESPARALLELCNV/
Python version 3.10.2	Python Software Foundation	https://www.python.org; RRID:SCR_008394
R version 4.0.1	R Software	https://www.r-project.org; RRID:SCR_001905
QuantiSNP	Colella et al.[61]	https://github.com/cwcyau/quantisnp; RRID:SCR_013091
PennCNV	Wang et al.[62]	https://penncnv.openbioinformatics.org/en/latest/; RRID:SCR_002518
CNVision	Sander et al.[63]	https://www.softpedia.com/get/Science-CAD/CNVision.shtml
PLINK	Purcell et al.[64]	https://www.cog-genomics.org/plink/; RRID:SCR_001757
GENCODE	The GENCODE Project	https://www.gencodegenes.org/
BedTool	Quinlan et al.[65]	https://bedtools.readthedocs.io/en/latest/; RRID:SCR_006646
Analysis scripts	This paper	10.6084/m9.figshare.27328212
DigCNV	This paper	10.6084/m9.figshare.27328227

RESOURCE AVAILABILITY

We analyzed 258,292 individuals from six general population cohorts,[49–53,66] which can be further divided into 9 sub-cohorts based on cognitive assessment (Table 1). Three additional autism cohorts[54–56] were only used for sensitivity analyzes (Table S6, Figure S15). Each cohort received approval from their local institutional review boards. Parents/guardians and adult participants gave written informed consent, and minors gave assent.

General populations

In this study, we included five cohorts from the general population previously pooled and studied in Huguet et al. 2021.[12] In addition to these cohorts previously analyzed and studied, we added 238,176 individuals from the UK Biobank (UKBB) cohort (www.ukbiobank.ac.uk) after phenotypic and genotypic quality control. The UKBB consortium initially recruited ~500 000 individuals aged 40–69 years (54% female) between 2006 and 2010. Phenotypic and cognitive measures were tested at the UKBB assessment centers or online, and also included demographic, socioeconomic and health data.

Autism spectrum disorder cohorts

We also included two cohorts of children with autism spectrum disorder previously studied in Huguet et al., 2021.[12] In addition, we included 2,543 ASD probands with available IQ measures from the Simons Foundation Powering Autism Research (SPARK) database.[56]

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Measures of cognitive ability

General cognitive ability was measured by either non-verbal intelligence quotient (NVIQ or Moray House Test), FI (fluid intelligence questions), or general intelligence factor (g-factor).[15] Measures of cognitive ability were z-scored within each cohort based on sex and age (Table 1, Table S6 and S7). We used the exact same process and data as shown previously in Huguet et al., 2021.[12] The NVIQ or Moray House Test *Z* score has a mean of 100 and a standard deviation (SD) of 15. Since cognitive measures used in the computation of the g-factor are not the same between cohorts, the g-factor was computed and normalized separately within each cohort using the mean and SD computed on all available individuals. This was feasible since the g-factor was computed in general population cohorts only. Of note, FIs and g-factors were computed before excluding individuals due to array quality control, leading to means and SDs slightly different from 0 to 1 for the final subset of individuals included in our analyses. In UKBB, some individuals had multiple cognitive ability assessments. For those individuals we selected the most robust cognitive evaluations based on the following ranking (from the most to the least robust): 1) in-person g-factor, 2) online g-factor, 3) in-person FI 4) online FI.

Intelligence quotient

In the SPARK cohorts, adapted tests have been used and ranked. We computed the average IQ interval for each rank to establish a numerical value. To be able to compare the different cognitive measures, all IQs were z-scored based on a mean of 100 and a standard deviation (SD) of 15.

Fluid intelligence

In UKBB, the FI score was assessed both in person (N = 88,441, #20016) and online (N = 13,773, #20191). This score is derived from 13 questions, measuring the capacity to solve problems requiring logic and reasoning abilities, independent of acquired knowledge. Participants were allotted 2 min to complete as many questions as possible from the test. The FI obtained were transformed into a *Z* score using the mean of 6.07 and the SD of 2.15 for the subgroup assessed in person, and using the mean of 6.61 and the SD of 1.98 for the subgroup assessed online.

G-factor computation

The g-factor is an indirect measure of general intelligence, obtained by extracting the first unrotated principal component from principal component analysis (PCA) of different standardized cognitive measures. It is a robust measure of general cognitive ability that is not very sensitive to the exact subtests used to calculate it as long as they measure a wide range of cognitive abilities.[67] Since cognitive measures used in the computation of the g-factor are not the same between tests used (in person and online), the g-factor was computed and normalized separately within each test group (in person and online) using the mean and SD computed on all available individuals.

For SYS parents sample, we computed the g-factor based on 12 cognitive performances[50] assessed using the Cambridge brain sciences platform[68]: color-word remapping, spatial planning, self-ordered search, paired associates learning, digit span, spatial span, visuospatial working memory, interlocking polygons, feature match, odd one out, grammatical reasoning and spatial rotation. The observed variance for g-factor was 31.6%, the $\text{mean}_{\text{g-factor}} = -6.22 \times 10^{-12}$ and the $\text{SD}_{\text{g-factor}} = 1.95$, both were used to compute the *Z* score for this measure.

For SYS children, we computed the g-factor based on 63 cognitive measures[50]: dot location (visual/non-verbal memory), Newman's card sorting task (perseveration), self-ordered pointing task (working memory), grooved pegboard Test (fine motor skills), Children's Memory Scale (CMS) stories subtasks (auditory/verbal memory), Wechsler Intelligence Scale for Children III (WISC-III), Woodcock-Johnson III (Academic achievement), Stroop color-word test (interference), Ruff 2-&-7 selective attention test (selective attention), Verbal fluency (cognitive flexibility) and tapping. The observed variance for g-factor was 23.6%, the $\text{mean}_{\text{g-factor}} = 0.05$ and the $\text{SD}_{\text{g-factor}} = 3.80$, both were used to compute the *Z* score for this measure.

For CaG cohort, we computed the g-factor based on three cognitive tests: verbal and numeric reasoning (fluid intelligence), paired associates learning (episodic memory) and reaction time based on two-choice items. The observed variance for g-factor was 43.2%, the $\text{mean}_{\text{g-factor}} = -8.68 \times 10^{-16}$ and the $\text{SD}_{\text{g-factor}} = 1.08$, both were used to compute the *Z* score for this measure.

For G-Scot cohort, the g-factor was computed using four cognitive tests measuring processing speed, verbal declarative memory, executive functions and vocabulary. The observed variance for g-factor was 42.3%, the $\text{mean}_{\text{g-factor}} = -3.65 \times 10^{-16}$ and the $\text{SD}_{\text{g-factor}} = 1.3$, both were used to compute the *Z* score for this measure.

For UKBB, the g-factor was computed using four cognitive tasks assessed in person (N = 73,882) and online (N = 62,080): trail making test parts A and B (executive function), symbol digit substitution test (processing speed), paired associate learning test (verbal declarative memory) and picture vocabulary (crystallized ability) (Table S7). The observed variance were 31.8% and

43.7% for the g-factor in person and online respectively. The g-factors obtained were transformed into a *Z* score using the mean of 1.80e−15 and the SD of 1.26 for the subgroup assessed in person, and using the mean of −1.40e−15 and the SD of 1.48 for the subgroup assessed online.

METHOD DETAILS

Except for UKBB and SPARK, we used the same raw data as in the previous publications, Huguet et al.[11,12] The probes coordinates were updated from hg18 to hg19 using Illumina information and the liftover tool from the genome browser. UKBB used DNA extracted from blood and genotyped on two Affymetrix arrays (n = 50k on UK BiLEVE and n = 450k on UK Biobank Axiom)[17] with ∼95% probe overlap, using ∼750k common markers. SPARK used DNA extracted from saliva (OGD-500 kit, DNA Genotek) genotyped on Illumina GSA-24v1-0 array (654k SNP sites).

Genetic analysis on genotyping

For data processing and quality control, we employed PLINK[64] software, version 1.9. Each cohort was filtered to keep only autosomal SNPs with minor allele frequency (MAF) > 5%, probes providing genotypes that are not violating Hardy-Weinberg equilibrium (threshold <1 × 10-6) and probes with call rates >90%. Also, we used PLINK[64] to check for duplicated individuals, sex, and relationships for each participant with the same pipeline as previous work. We merged all genotyping data with PLINK. Finally ancestries (principal component [PC] 1 to 10) were determined with KING[69] (with 3,615 common SNPs, we used the same quality control as in the previous step), using the standard process defined on the website (https://www.kingrelatedness.com) and the 1000 Genomes as reference.

CNV calling

We applied the same methodology as in Huguet et al.[11,12] available online (https://martineaujeanlouis.github.io/MIND-GENESPARALLELCNV/) on the array data using PennCNV[62] and QuantiSNP[61] algorithms. The following parameters were used for both algorithms: number of consecutive probes for CNV detection ≥3, CNV size ≥1Kb, likelihood scores ≥15. CNVs detected by both algorithms were combined (CNVision[63]) to minimize the number of potential false discoveries. We defined all CNVs with less than 2 copies as deletions and all CNVs with more than 2 copies as duplications. After this merging step, an in-house algorithm based on CNV was applied to concatenate adjacent CNVs of the same type into one, according to the following criteria: a) gap between CNVs ≤150 kb; b) size of the CNVs ≥1000 bp; and c) number of probes ≥3.

Array filtering

After these steps, we remove from the analyses, all arrays for which a suspiciously high number of CNVs has been detected (≥50 for low resolution arrays [<1 million probes] and ≥200 for high resolution arrays [≥1 million probes]). For all cohorts, we used stringent quality-control criteria: call rate ≥95%; log R ratio-standard deviation <0.35; B allele frequency-standard deviation <0.08 and |waviness factor|<0.05. From a total of 488,377 people with genotypic data, 28,522 were excluded for failing only of these filters.

All individuals with duplicated data or with discordant phenotypic and genetic information about the sex were removed (N = 212). We did exclude CNVs ≥10Mb (a widely used threshold in the QC if CNVs[11,12,18]) because very large CNVs are rarely observed in general population cohorts and are almost always present as mosaics and/or somatic CNVs that can't be pooled with germline CNVs.

CNV filtering

After filtering the arrays according to their quality, we applied filtering for autosomal CNVs. The CNVs with the following criteria were selected for analyses: likelihood score ≥30 (for at least one of both detection algorithms), size ≥50 kb, unambiguous type (deletions and duplications) and overlap with segmental duplicates, HLA regions or centromeric regions <50%. To avoid frequency biases coming from the level of detection across technologies, we applied 3 criteria: 1) CNVs had to be covered by at least 10 probes across all array technologies used in the analyses; 2) CNVs with a frequency ≥1% in at least 1 cohort were removed from all cohorts; 3) CNVs with a coefficient of variance of frequency being part of the top 1% were removed (separated distribution of coefficients used according to how many cohorts included the CNV). For steps 2 and 3, CNVs were defined as similar if their sequences had a reciprocal overlap ≥50%. Every recurrent CNV was annotated (based on previously published methods[12]) and manually visualized (Log R and BAF-plots) by at least one CNV experts.

In addition, we applied an in-house algorithm based on a machine learning method to detect additional artifact CNVs (DigCNV, https://github.com/labjacquemont/DigCNV). This algorithm was based on the consensus of three machine learning methods (Random forest, bagging of KNN and SVM) and on 9 CNV characteristics (Array criteria: log R ratio-standard deviation, B allele frequency-standard deviation, wave frequency; Localization CNV criteria: % of CNV overlap with centromeric regions and with segmental duplications; CNV criteria: density of SNPs (numbers of SNPs/size of CNV), likelihood score/number of SNPs, % algorithms overlapping, percentage of shared sequence found by the both algorithms), type of CNV). This model was trained and tested respectively on 66% and 33% of 34,156 CNVs (31,746 true CNVs and 2,410 artifacts from 6 cohorts, excluding SPARK), This reference CNV set was manually inspected with Log R and BAF plots, by two CNV experts. DigCNV showed an AUC = 0.95, a sensitivity of

0.95 and a specificity of 0.85. This model was validated again on an additional naive dataset genotyped with another technology (GSA). We used 2,454 CNVs (1,936 true CNVs and 518 artifacts from SPARK cohort) and showed an AUC = 0.92, a sensitivity of 0.58 and a specificity of 0.97.

Annotation of CNVs

We annotated the CNVs using GENCODE V19 annotation (hg19) with Ensembl gene name (https://grch37.ensembl.org/index.html). We used bedtools suite to identify the different elements of the genes encompassed in CNVs.[65] CNV annotation was therefore defined by the sums of genes fully encompassed and being part of a biologically defined gene-set. These gene-sets were coming from the partition of the whole genome as defined in the following paragraphs.

LOEUF-based gene sets

Each coding gene was annotated using the Loss-of-function Observed/Expected Upper bound Fraction (LOEUF) score (gnomAD version 2.1.1),[57] which is available for 19,197 genes and ranges from 0.03 to 2, and values below 0.35 are suggestive of intolerance. The smaller the value is, the more the gene is intolerant to loss-of-function variants. We defined 38 overlapping gene-sets based on LOEUF values using a sliding window method (methodology as in Huguet et al.[11,12]). Each window was a 0.15 range of LOEUF values, and the sliding was 0.05.

Function-based gene sets

We defined 269 gene-sets based on relative gene expression (*Z* score >1SD) in 13 adult[70,71] and 16 fetal[72] brain cell types,[60] as well as bulk tissue from 215 brain regions (Human Protein Atlas, HPA v.22)[32] and 25 non-brain organs (GTEx v8,[32,59] Table S8). The expression values were normalized across all tissues for each gene. The same normalization was performed across cell types separately. As a sensitivity analysis, we defined the same gene-sets based on a previously published "Top Decile Expression Proportion" (TDEP)[34] method. The former and the latter methods favor relative and specific expression, respectively. Both methods exclude 1,370 and 5,369 genes that are not assigned to any tissue in GTEx. We also used 6,233 functional gene-sets based on 6,130 GOterms[73,74] (Ensembl v.109, April 2023), and 103 Synapse ontology terms (SynGO[35]). We were used with propagated annotations following Gene Ontology Consortium recommendations. Throughout this study, we only considered gene-sets meeting the following 3 criteria: i) those with more than 10 genes, ii) those disrupted by $\geq$ 30 CNV carriers, and iii) those with at least 20% of their genes affected by CNVs.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analyses were performed using R version 4.0.1 (http://www.R-project.org.), with "meta"(https://cran.r-project.org/web/packages/meta/index.html) and "metafor" (https://cran.r-project.org/web/packages/metafor/index.html) packages for meta-analyses. Python 3.10.2 (https://www.python.org) with "scipy 1.11.2" (https://pypi.org/project/scipy/), "statsmodels 0.13.5" (https://www.statsmodels.org) and "word-cloud 1.9.2" (https://amueller.github.io/word_cloud).

LOEUF and function-based burden associations

To estimate the effect on cognitive ability of gene-sets (and their corresponding biological functions or LOEUF categories), we adapted a previously published model.[12]

We performed a linear model for each of the 38 LOEUF gene categories and each of the 6,502 functional gene-sets. The outcome was cognitive ability measured in each individual. The explanatory variable was the sum of genes fully encompassed in a CNV for a gene-set of interest (Figure 2A). Since CNVs are multigenic, the effect size estimated for a given gene set may be inflated. Therefore, all models were adjusted for the total number of genes within the CNV but not members of the gene set of interest. These latter genes were categorized into three covariates: ID genes (only for Function-based gene-sets), genes with LOEUF<1, and genes with LOEUF$\geq$1. Other covariates included ancestry (10 PCs), age, and sex. Models were computed for deletions or duplications, separately. *p*-values were corrected for multiple testing (one for each biologic function) using FDR correction, separately for deletion and duplication.

Linear regression model

In our study, we used four distinct models to assess the average main effects of genes within specific categories of interest. In Model 1, the CNV-GWAS approach, we applied a linear model for each gene individually. Contrarily, in Models 2, 3, and 4, we used the gene-sets described before. These 3 models are also taking into account the genes encompassed in a CNV but not in the gene-set of interest as a covariate. We used a cut-off of 30 carriers to obtain a power of 85% to be able to detect CNVs with large effect size equivalent to Cohen's d = 0.7 (alpha = 0.005).

Model 1: For each gene, we implemented an individual linear model, considering a minimum of 30 carriers. The aim was to assess the average main effect of each gene (example for deletion).

$$Z_{cog.\ ability\ adj.} \sim \beta_{0,DEL} + \beta_{1,DEL} \times Z_{gene\ was\ deleted\ 1\ or\ not\ 0} + PC1\ to\ PC10$$

Model 2: Building on previously published work, we conducted 39 linear models to examine 38 overlapping LOEUF categories (using a sliding window with a size of 0.15 LOEUF and a step of 0.05 LOEUF), as well as a category comprising an ID gene list as defined by ClinGen (Table S2). Each model focused on the average main effect of a gene within the specified category, adjusting for the impact of other genes in the CNV with LOEUF values falling outside the window of interest (Figure 2A). We applied the same model for the ID gene-sets, as a replacement of the LOEUF window of interest.

$$Z_{cog.\,ability\,adj.} \sim \beta_0 + \beta_1 \times \sum (genes\ i\ inside\ the\ LOEUF\ window)$$
$$+ \beta_2 \times \sum (genes\ i\ with\ LOEUF < 1\ \&\ outside\ the\ LOEUF\ window)$$
$$+ \beta_3 \times \sum (genes\ i\ with\ LOEUF \geq 1\ \&\ outside\ the\ LOEUF\ window) + PC1\ to\ PC10$$

Model 3: We applied a linear model for each gene-set to estimate their effect sizes. These models evaluated the average main effects of genes within a gene-set, with adjustments for the influence of other genes in the CNV. Genes outside the gene-set, these were further subdivided into three categories: ID-gene, LOEUF <1, and LOEUF ≥1.

$$Z_{cog.\,ability\,adj.} \sim \beta_0 + \beta_1 \times \sum (genes\ i\ inside\ the\ gene\ set)$$
$$+ \beta_2 \times \sum (ID\ genes\ i < 1\ \&\ outside\ the\ gene\ set)$$
$$+ \beta_3 \times \sum (genes\ i\ with\ LOEUF < 1\ \&\ outside\ the\ gene\ set)$$
$$+ \beta_4 \times \sum (genes\ i\ with\ LOEUF \geq 1\ \&\ outside\ the\ gene\ set) + PC1\ to\ PC10$$

Model 4: Employing the same approach as in Model 3, we used a single linear model but divided the gene-set into three LOEUF categories: LOEUF <0.35, LOEUF in the range of [0.35, 1[, and LOEUF ≥1.

$$Z_{cog.\,ability\,adj.} \sim \beta_0 + \beta_1 \times \sum (genes\ i\ inside\ the\ gene\ set\ \&\ LOEUF < 0.35)$$
$$+ \beta_2 \times \sum (genes\ i\ inside\ the\ gene\ set\ \&\ 0.35 \leq LOEUF < 1)$$
$$+ \beta_3 \times \sum (genes\ i\ inside\ the\ gene\ set\ \&\ LOEUF \geq 1)$$
$$+ \beta_4 \times \sum (ID\ genes\ i < 1\ \&\ outside\ the\ gene\ set)$$
$$+ \beta_5 \times \sum (genes\ i\ with\ LOEUF < 1\ \&\ outside\ the\ gene\ set)$$
$$+ \beta_6 \times \sum (genes\ i\ with\ LOEUF \geq 1\ \&\ outside\ the\ gene\ set) + PC1\ to\ PC10$$

tagDS

Each gene-set is represented in two dimensions by their deletion and duplication effect sizes. The nominal tagDS is the Euclidean distance between the gene-set coordinates and the line of equation, $y = 2.4x$ which is the ratio effect size between duplications (x) and deletions (y) computed for a genome-wide gene-set. Because effect-sizes depends on the gene-set sizes, we normalized the nominal tagDS for each gene-set. Each nominal tagDS is then Z-scored based on the normal distribution of tagDS computing for 100 random gene-sets with the same number of genes. Finally, a tagDS of 0 suggests that the deletion/duplication effect-size ratio is equal to the expected ratio. A gene-set with a tagDS >2 indicates that its deletion/duplication effect-size ratio on cognitive ability is beyond 2 standard deviations of the null distribution (i.e., larger effect sizes are biased toward deletions).

Cell Genomics, Volume *4*

Supplemental information

Effects of gene dosage on cognitive ability:

A function-based association study

across brain and non-brain processes

Guillaume Huguet, Thomas Renne, Cécile Poulain, Alma Dubuc, Kuldeep Kumar, Sayeh Kazem, Worrawat Engchuan, Omar Shanta, Elise Douard, Catherine Proulx, Martineau Jean-Louis, Zohra Saci, Josephine Mollon, Laura M. Schultz, Emma E.M. Knowles, Simon R. Cox, David Porteous, Gail Davies, Paul Redmond, Sarah E. Harris, Gunter Schumann, Guillaume Dumas, Aurélie Labbe, Zdenka Pausova, Tomas Paus, Stephen W. Scherer, Jonathan Sebat, Laura Almasy, David C. Glahn, and Sébastien Jacquemont

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

List of supplementary tables and figures:

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

Supplementary tables

Cohort		Array type	N=	Ancestry		Gender		Age (year)		Z-scored intelligence measure (adj)			Cognitive ability assessments
				EUR	Others	F	M	Mean	SD	Mean	SD	Variables	
Unselected (n=258,292)	CaG	GSA	2074	1982	92	1094	980	54.317	7.601	0.107	0.973	Age, Age^2, sex, PC	g-factor, Reasoning, Memory, Reaction time
		Omni2.5	515	490	25	281	234	52.437	8.602	-0.009	0.956		
		GSA + Omni2.5	2589	2472	117	1375	1214	53.943	7.845	0.084	0.970		
	G-Scot	610Kq	13715	13672	43	8081	5634	46.730	14.996	0.050	0.974	Age, Age^2, sex, PC	g-factor, Logical Memory, Digit Symbol, Verbal fluency, Mill Hill Vocabulary
	Imagen	610Kq; 660Wq	1744	1624	120	891	853	14.450	0.366	0.441	0.977	PC	WISC-IV
	LBC1936	610Kq	503	500	3	246	257	69.825	0.829	0.047	0.946	PC	Moray House Test
	SYS children	610Kq	559	557	2	298	261	15.058	1.894	0.361	0.882	PC	WISC-III
		HOE-12V	408	408	0	207	201	14.906	1.760	0.212	0.848		
		610Kq + HOE-12V	967	965	2	505	462	14.994	1.839	0.298	0.871		
	SYS parents	HOE-12V	598	596	2	319	279	49.495	4.868	-0.021	0.934	Age, Age^2, sex, PC	g-factor, 12 cognitive measures‡
	UKBB	Affymetrix	73882	71364	2518	39317	34565	60.022	8.959	0.131	0.964	Age, Age^2, sex, PC	g-factor [47]
			62080	60484	1596	34335	27745	62.083	7.663	0.131	0.926		g-factor (online)
			88441	80427	8014	47789	40652	58.139	8.304	-0.035	0.961		FI
			13773	13458	315	8284	5489	64.185	7.685	-0.090	0.970		FI (online)
Autism (n=6,111)	SPARK	GSA	2543	1984	559	540	2003	12.359	6.190	-0.626	1.963	PC	IQ
	MSSNG	WGS	1007	768	239	202	805	9.503*	4.600*	-0.529	1.590	PC	IQ
	SSC	1Mv1	332	279	53	44	288	9.538	3.240	-0.602	1.558	PC	WISC-IV n=19; DAS-II E-Y n=96; DAS-II S-A n=179; Mullen n=12; WASI-I n=26
		1Mv3	1181	915	266	156	1025	8.769	3.523	-0.982	1.638		WISC-IV n=16; DAS-II E-Y n=530; DAS-II S-A n=539; Mullen n=77; WASI-I n=19
		Omni2.5	1048	786	262	140	908	9.160	3.712	-1.227	1.834		WISC-IV n=10; DAS-II E-Y n=403; DAS-II S-A n=494; Mullen n=124; WASI-I n=17
		1Mv1 + 1Mv3 + Omni2.5	2561	1980	581	340	2221	9.028	3.576	-1.033	1.722		WISC-IV n=45; DAS-II E-Y n=1,029; DAS-II S-A n=1,212; Mullen n=213; WASI-I n=62

Table S6. Cohort descriptions, Related to Table 1.

Cohorts include 264,403 individuals, including 258,292 general populations. †63 and ‡ 12 cognitive measures were respectively used to compute the g-factor in SYS children and parents (Huguet et al 2021). SYS: Saguenay Youth Study, CaG: CARTaGEN, LBC1936: Lothian Birth Cohort 1936, SSC: Simons Simplex Collection; n=number of individuals remaining for analysis after quality control. The mean and Standard Deviation (SD) for FI and g-factor slightly deviate from 0 and 1 in some cohorts since they were 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. * The MSSNG cohort gave participants years but for not all, for 280 age was missing.

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

Field name	Clinic value		Online value	
	code	Lyall et al 2016[47]	code	Gfact 5 Online
Fluid intelligence score	20016	use	20191	use
Trail making #2	6350	-	20157	use
Symbol digit substitution	23324	-	20159	use
Pairs matching	399	use	20132	use
Numeric memory	4282	use	20240	use
Prospective memory	20018	use	-	-
Mean time to correctly identify matches (Reaction time)	20023	use	-	-

Table S7. Descriptions of cognitive ability used in UKBB, Related to STAR Methods.

Name	HPA				GTEx				Cell types	SynGO	LOEUF catg.	GO-term
	SD≥0.5	SD≥1	SD≥1.5	SD≥2	SD≥0.5	SD≥1	SD≥1.5	SD≥2				
N lists	215	215	215	215	37	37	37	37	29	85	38	6,130
Mean list size	3,817	1,975	1,017	571	3,056	2,015	1,173	787	890	120	1,423	89
N unique coding gene	12,710	12,706	12,449	12,427	12,755	12,758	12,751	12,538	8,422	921	13,288	11,460

Table S8. Descriptions of Gene-sets, Related to STAR Methods.

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

Supplementary figures

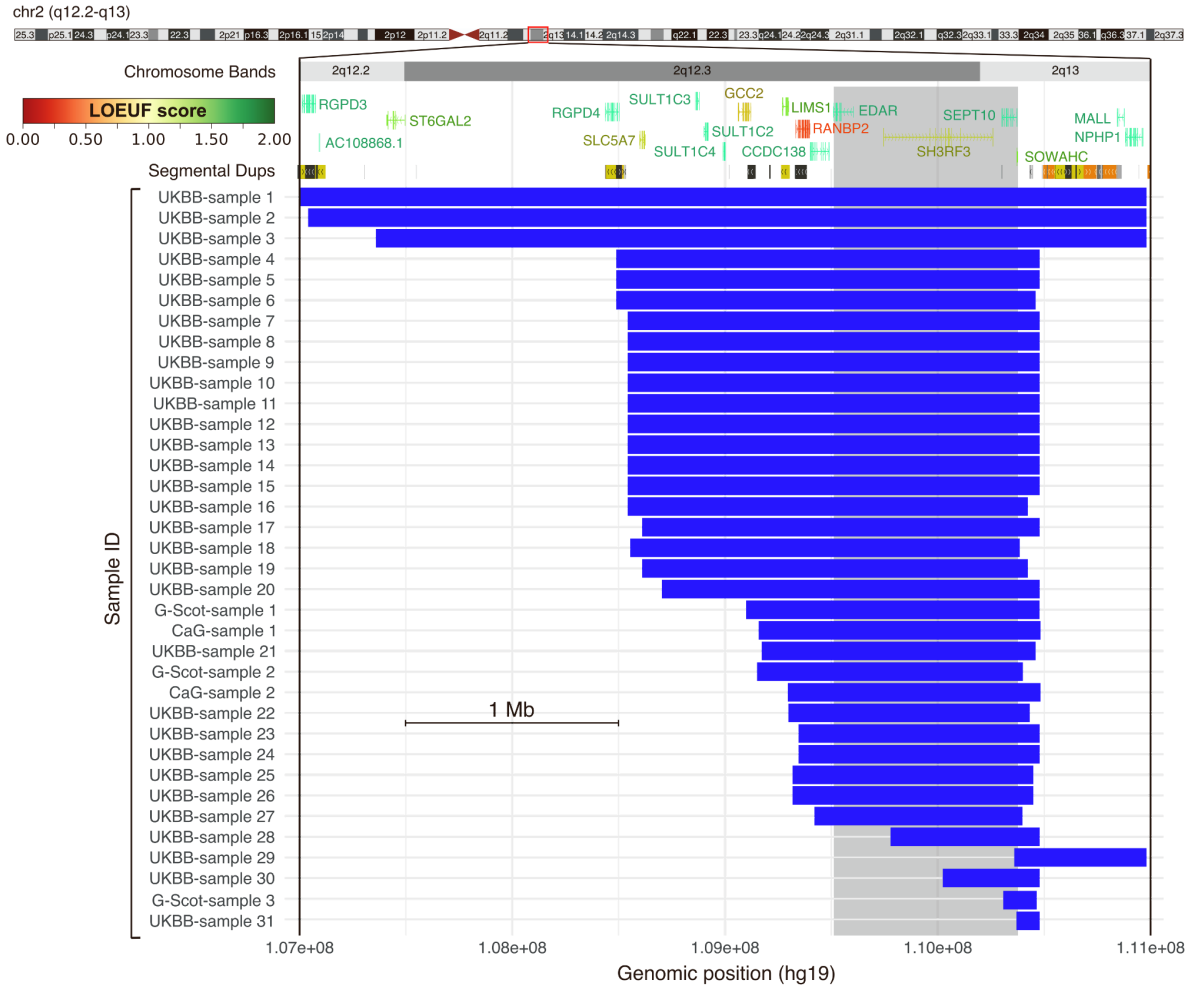


Figure S1: 36 duplications of the 2q12.3 observed in General Population, Related to Figure 1F.

This figure depicted genomic duplications (blue) occurring on chromosome 2 and overlapping with at least one of these genes: *EDAR, SH3RF3, SEPT10, SOWAHC*, in the general population. The x-axis repesented the genomic coordinates (hg19), and the y-axis listed the carriers' sample IDs. Three genomic annotations were included: chromosome bands; genes with their associated LOEUF scores; segmental duplications. LOEUF scores indicated the level of tolerance to loss-of-function for each gene, ranging from red (intolerant) to green (tolerant).

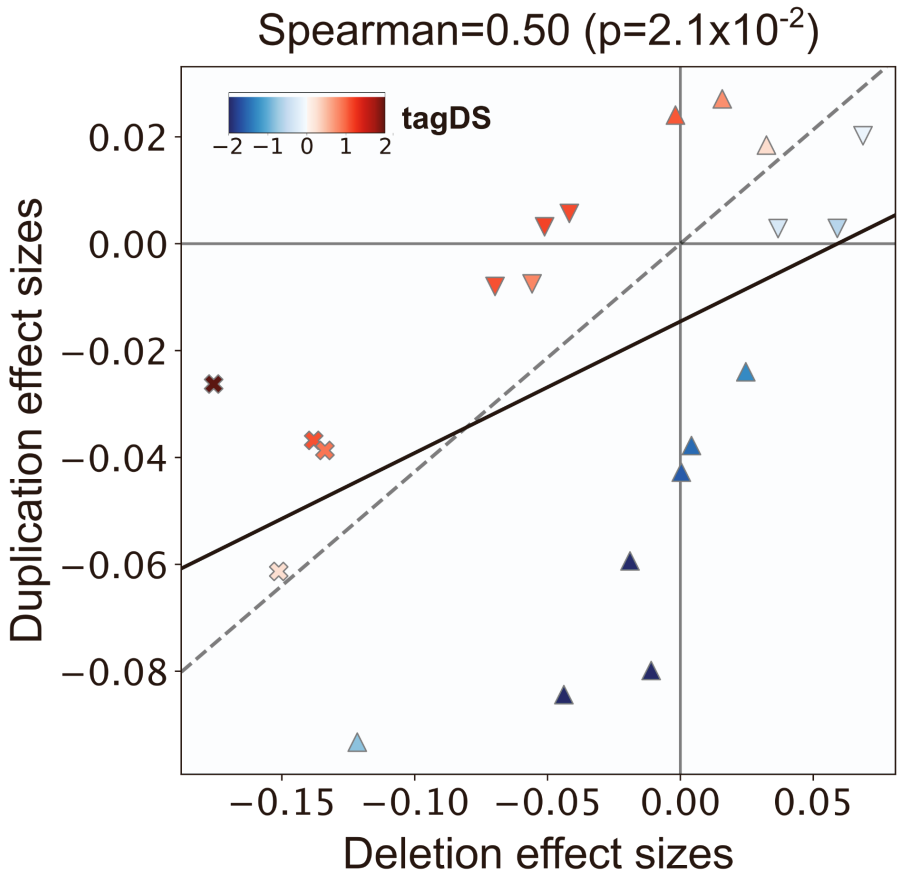


Figure S2: Correlation on cognitive ability of LOEUF categories, Related to Figure 2.

Spearman correlations (black line) between the effect sizes of deletions and duplications of gene-sets with FDR significant effects on cognitive ability for deletions (downward triangle), duplications (upward triangle), or both (cross). pvalue obtained from permutation test to account for the partial overlap between gene sets. Gene sets are color coded based on their tagDS. The dash line represents the average exome-wide duplication/deletion effect-sizes ratio

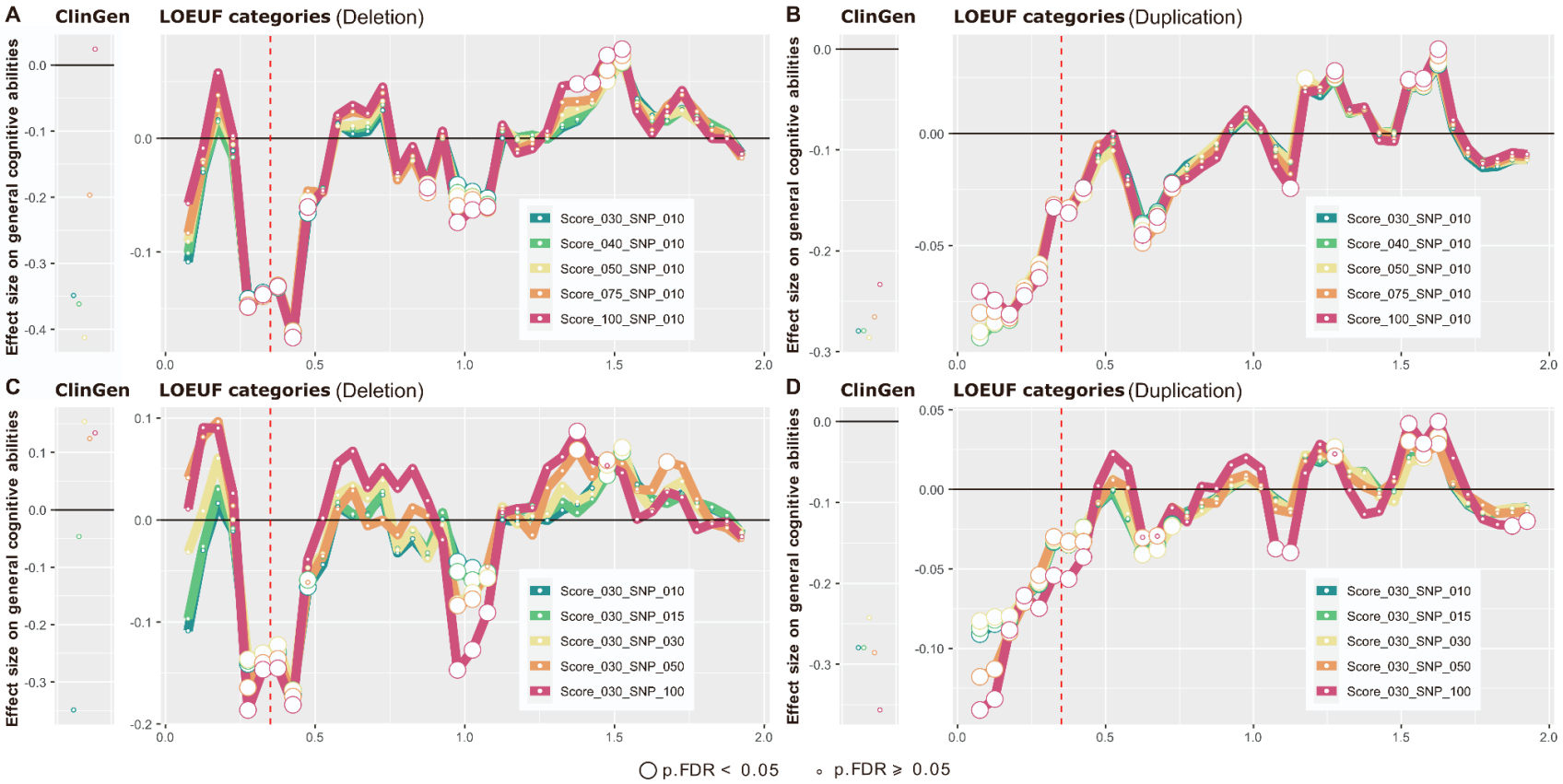


Figure S3: Effect sizes of autosomal coding genes on general cognitive abilities based on their LOEUF values, Related to Figure 2.

Sliding window estimating the effect size on cognitive ability of deletions (left) and duplications (right) for 38 LOEUF categories and definitive ID-genes curated by ClinGen (based on model 2). The line represents the estimated effect size of 38 categories of genes based on their LOEUF values in the model. Estimates were computed using a pooled dataset, large circles indicated significant pvalues adjusted by FDR and small one for non-significant. We ran sensitivity analyses based on different CNV cut-offs of quality controls with the likelihood score (Score ≥ 30, 40, 50, 75 and 100 and a fixed number of SNPs ≥10; for A and B) and numbers of SNPs inside CNV (SNPs ≥ 10, 15, 30, 50 and 100 and a fixed likelihood score ≥30; C and D).

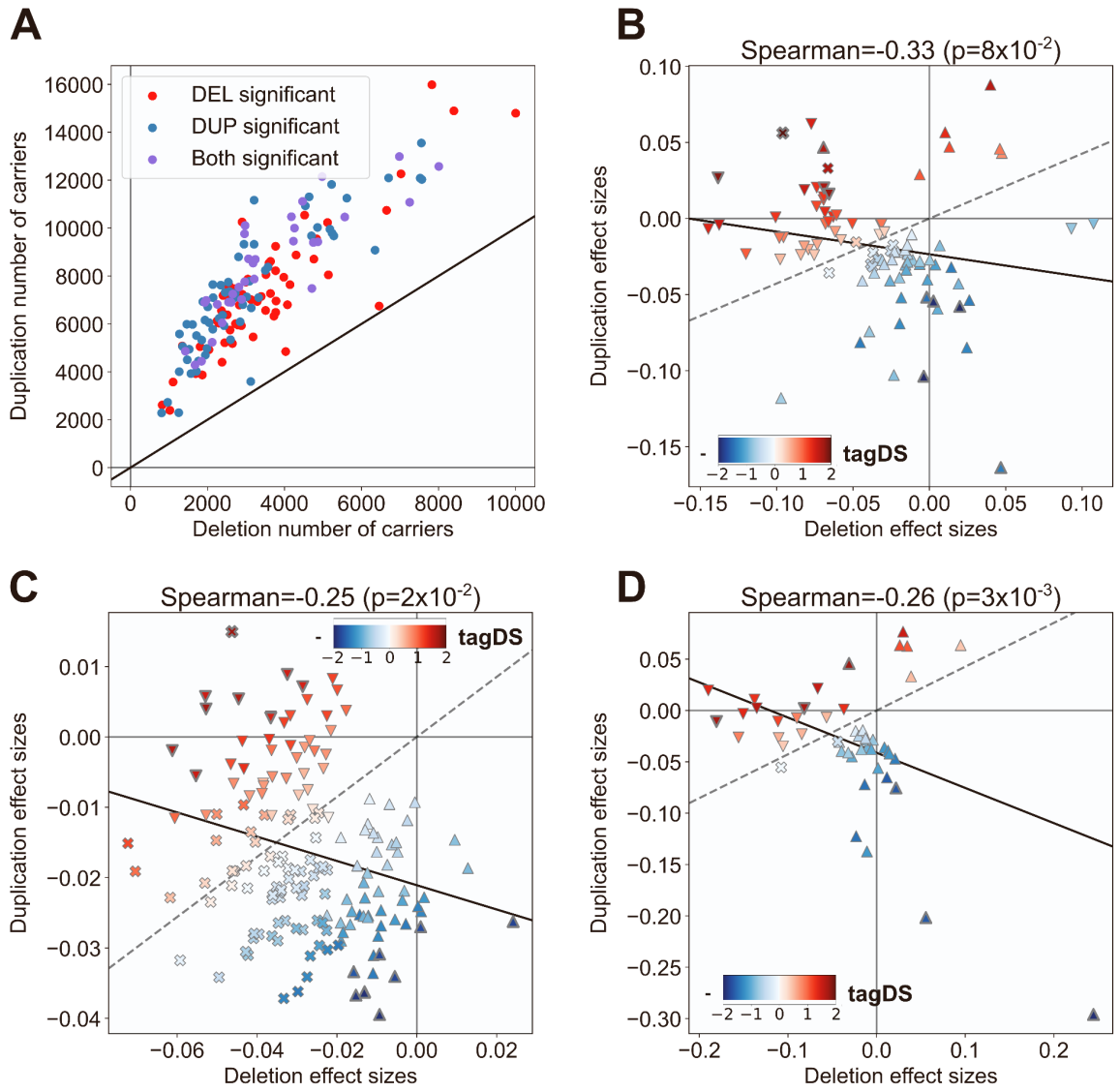


Figure S4: Sensitivity analysis on cognitive ability for multiple HPA gene expression thresholds, Related to Figure 3.

(A) Number of deletion and duplication carriers of genes for the 215 gene-sets analyzed in Figure 3. The black line represents the theoretical perfect concordance between Deletion and duplication carriers. Spearman correlations (black lines) between the effect sizes of deletions and duplications of tissue gene-sets with a normalized expression threshold >0.5SD (B), >1.5SD (C), and >2SD(D). FDR significant effects on cognitive ability for deletions (downward triangle), duplications (upward triangle), or both (cross). pvalues obtained from permutation tests to account for the partial overlap between gene sets. Gene sets are color coded based on their tagDS. The dash line represents the average exome-wide duplication/deletion effect-sizes ratio.

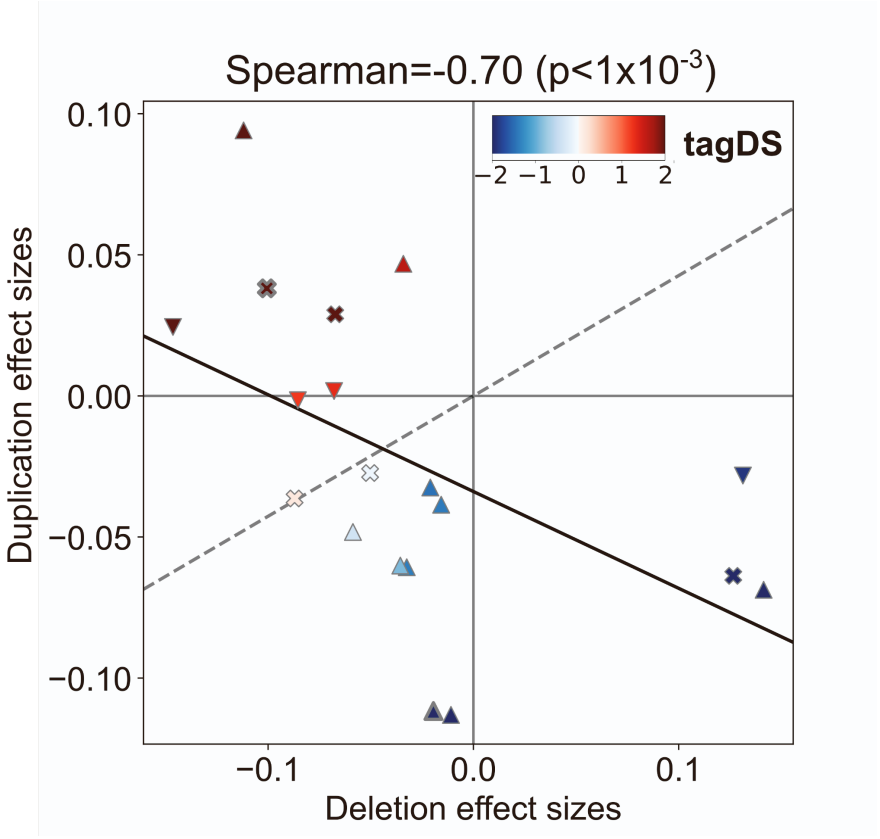


Figure S5: Correlation on cognitive ability of cell type, Related to Figure 3.

Spearman correlation (black line) between the effect sizes of deletions and duplications of gene-sets with FDR significant effects on cognitive ability for deletions (downward triangle), duplications (upward triangle), or both (cross). pvalues obtained from permutation tests to account for the partial overlap between gene sets. Gene sets are color coded based on their tagDS. The dash line represents the average exome-wide duplication/deletion effect-sizes ratio.

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

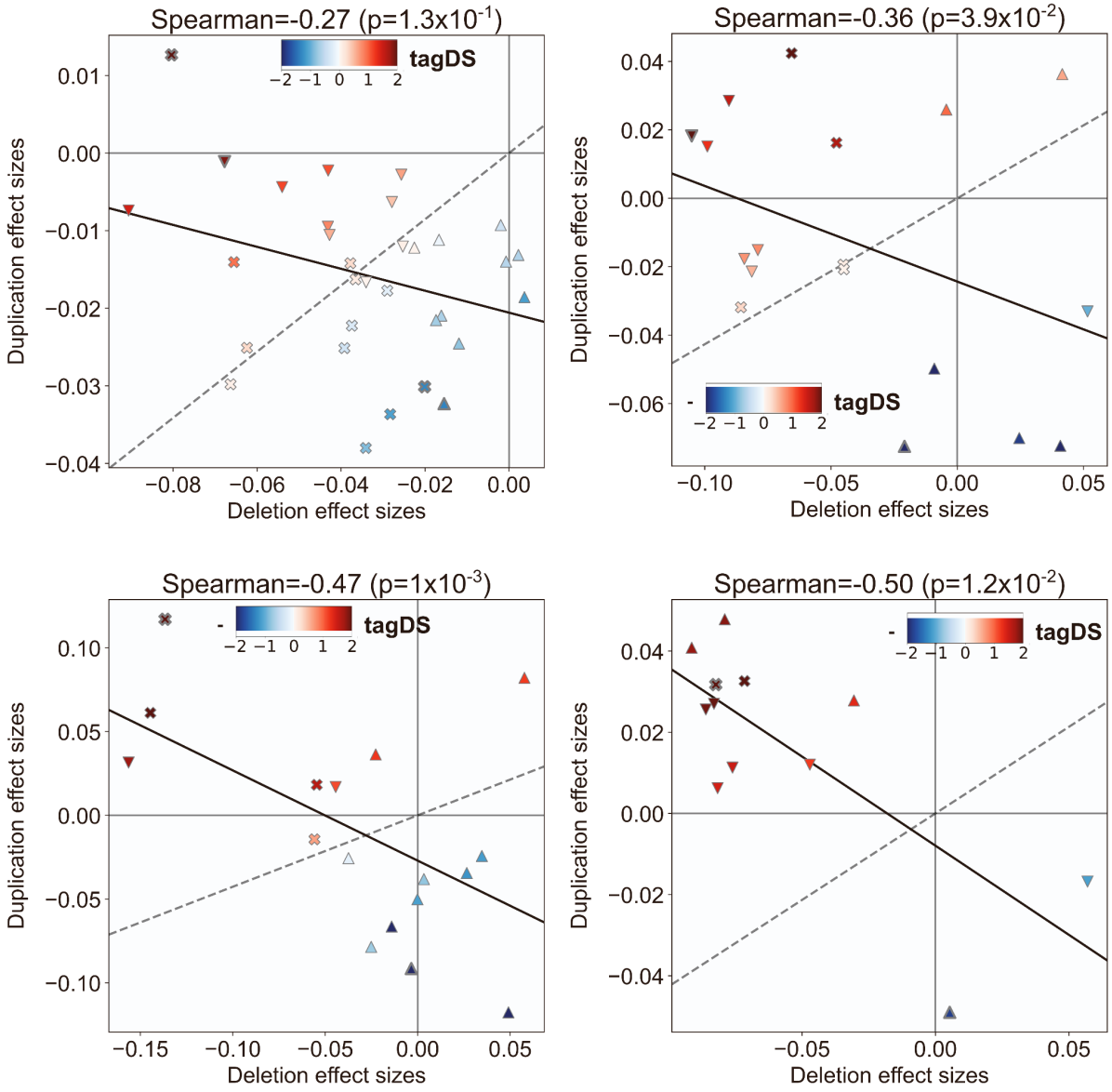


Figure S6: Correlations on cognitive ability of multiple GTEx gene specificity thresholds, Related to Figure 4.

Spearman correlations (black lines) between the effect sizes of deletions and duplications of tissue gene-sets with a normalized relative expression threshold >0.5SD (A), >1.5SD (B), >2SD(C) and >1SD without low-tissue-specificity genes(D). FDR significant effects on cognitive ability for deletions (downward triangle), duplications (upward triangle), or both (cross). pvalues obtained from permutation tests to account for the partial overlap between gene sets. Gene sets are color coded based on their tagDS. The dash line represents the average exome-wide duplication/deletion effect-sizes ratio.

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

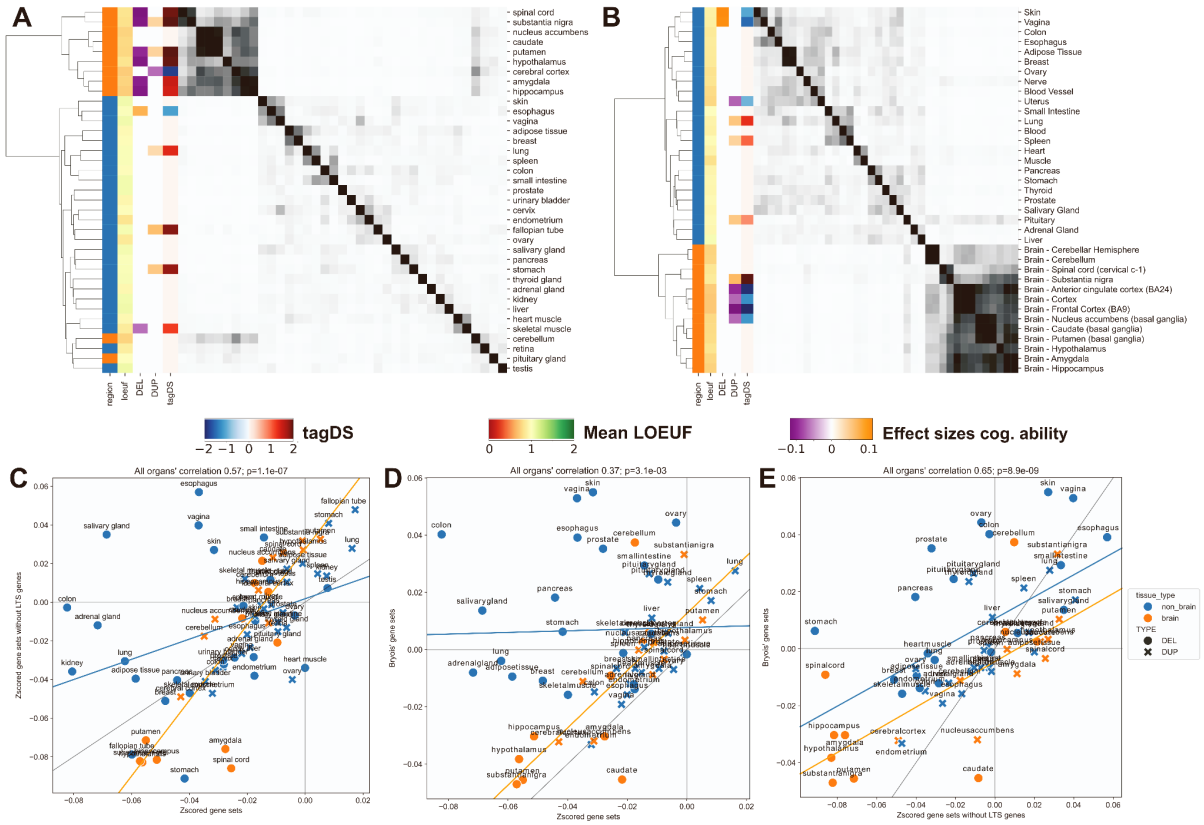


Figure S7: Comparisons of effect sizes for different gene-set associations, Related to Figure 4.

Clustering of gene-sets computed with z-score minus 8,194 low-tissue-specificity genes (A) and computed with TDEP (Bryois et al) (B). Orange represents brain tissues and blue non-brain tissues. Gene-set overlap matrix show high overlap between brain gene-sets and much lower overlap across non-brain tissues for both association methods. 2nd columns represent the average LOEUF score of the gene-set. 3rd and 4th columns represent the effect of gene-sets on cognitive ability when deleted and duplicated, respectively. The 5th column is the resulting tagDS. Correlation of effect-sizes for multiple gene-set definitions, z-scored gene-sets with all coding genes versus z-scored gene-sets without low-tissue-specificity genes (C), z-scored gene-sets with all coding genes versus TDEP gene-sets (D) and z-scored gene-sets without low-tissue-specificity genes versus TDEP gene-sets (E). Orange lines show the Spearman correlation between brain gene-sets (r=0.84 p=1e-6; r=0.82 p=9e-6; r=0.76 p=1e-4 for B, C and D respectively) and blue lines show correlation for non-brain gene-sets (r=0.47 p=4e-4; r=0.16 p=3e-1; r=0.58 p=4e-5 for B, C and D respectively). The grey line represents a theoretical perfect concordance between estimates. Pvalues are nominal pvalues without permutation tests.

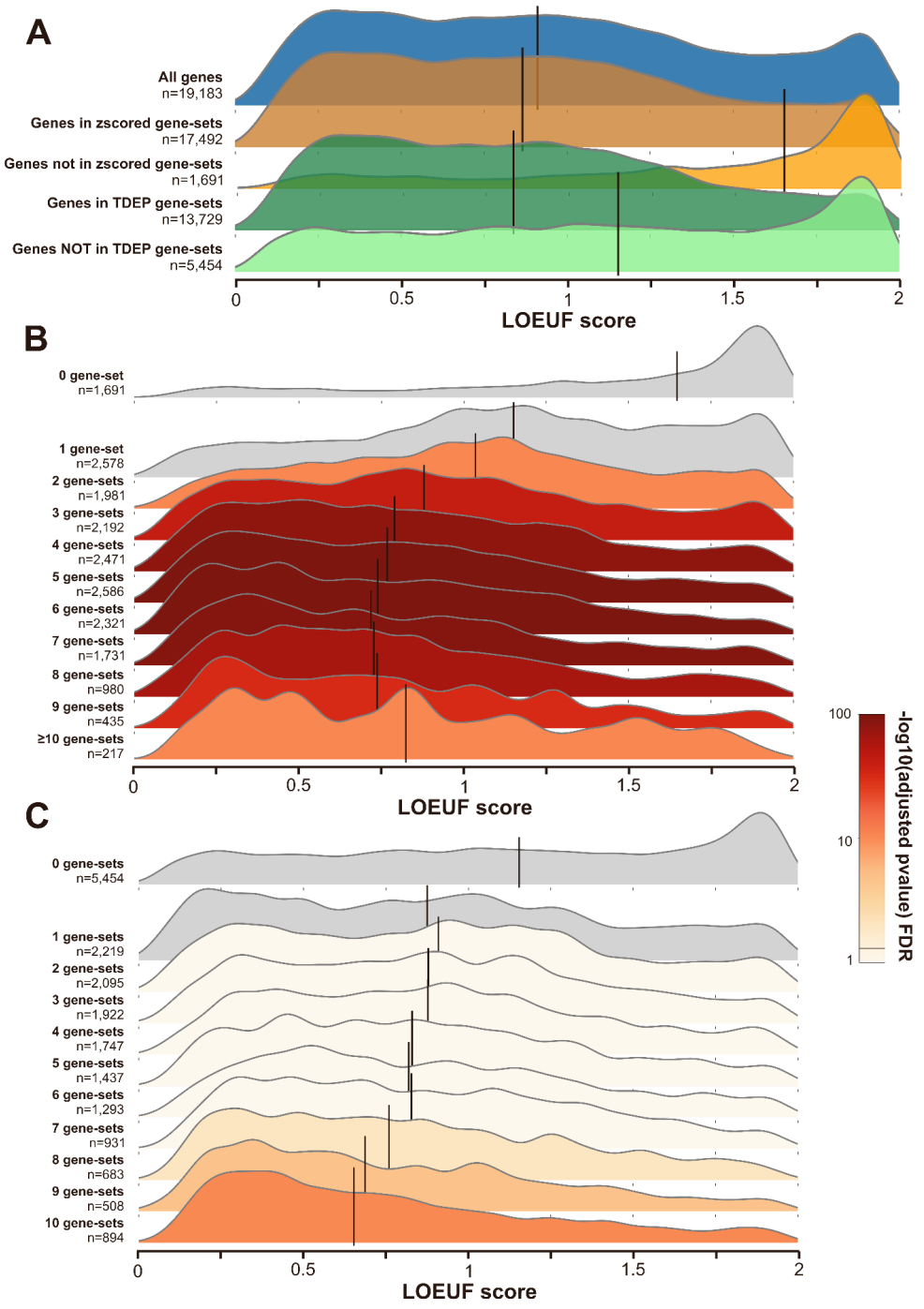


Figure S8: Ridgeplots representing the LOEUF distribution across multiple levels of specificity, Related to Figure 4.

(A) Distribution of LOEUF values for the whole coding exome, for genes assigned to at least one GTEx gene-set defined by z-score, for genes not assigned to any GTEx tissue (Z-score), for genes assigned to at least one GTEx tissue defined by TDEP and for genes not assigned to any GTEx gene-set (TDEP). Distribution of LOEUF for genes present in one or multiple gene-sets defined by Z-score (B) and TDEP (C) methods. The color represents the FDR-adjusted pvalue of the Mann-Whitney test between the distribution of LOEUF for specific genes (present in only one gene-set) and the distribution of interest. Black line represents the median LOEUF score for each category.

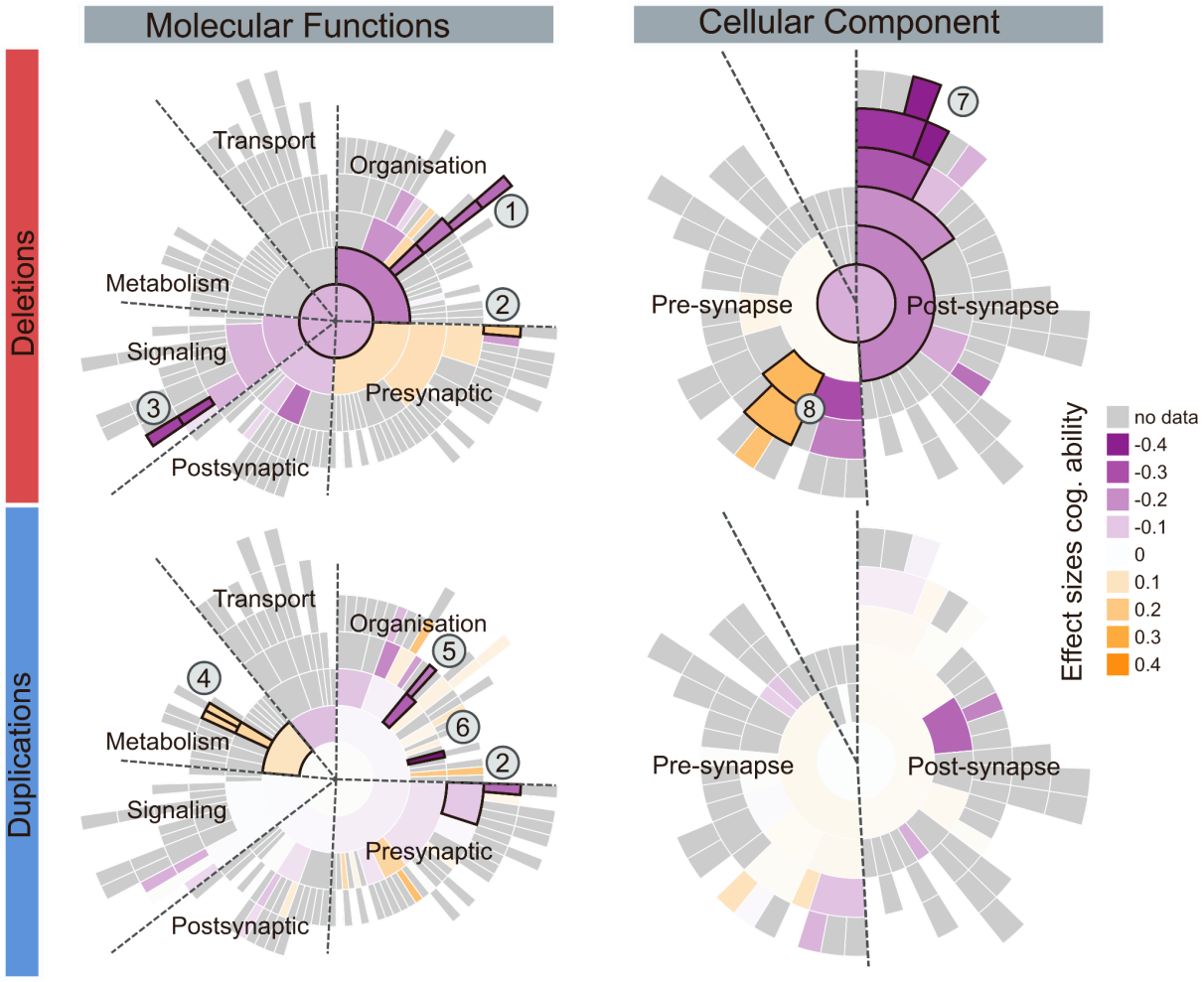


Figure S9: Effect sizes of Deletion and duplication on cognitive ability for SynGO gene-sets, Related to Figure 5.

Effect sizes of synaptic molecular functions and cellular component gene-sets as defined by SynGO[48] on cognitive ability. Purple and orange represent negative and positive effect size on cognitive ability, respectively. Ontologies with black edges indicate significant effects (FDR). The results are shown only for SynGO terms with more than 10 genes, observed at least 30 times in our dataset, and with a coverage greater than 20%. Note: 1) Regulation of modification of postsynaptic actin cytoskeleton, 2) Regulation of calcium-dependent activation of synaptic vesicle fusion, 3) Presynaptic modulation of chemical synaptic transmission, 4) translation at synapse, 5) regulation of postsynapse organization, 6) synapse adhesion between pre- and post-synapse, 7) Integral component of postsynaptic density membrane, 8) Synaptic vesicle membrane.

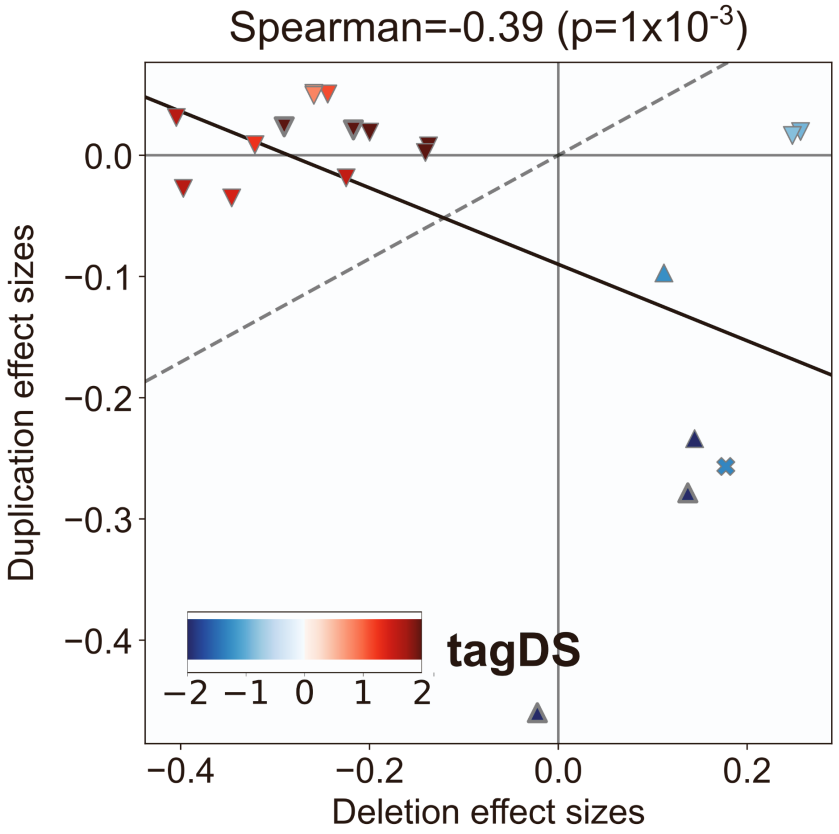


Figure S10: Correlation on cognitive ability of SynGO, Related to Figure 5.
Spearman correlations between the effect sizes of deletions and duplications of gene-sets with FDR significant effects on cognitive ability for deletions (downward triangle), duplications (upward triangle), or both (cross). pvalues obtained from permutations to account for the partial overlap between gene sets. Gene sets are color coded based on their tagDS.

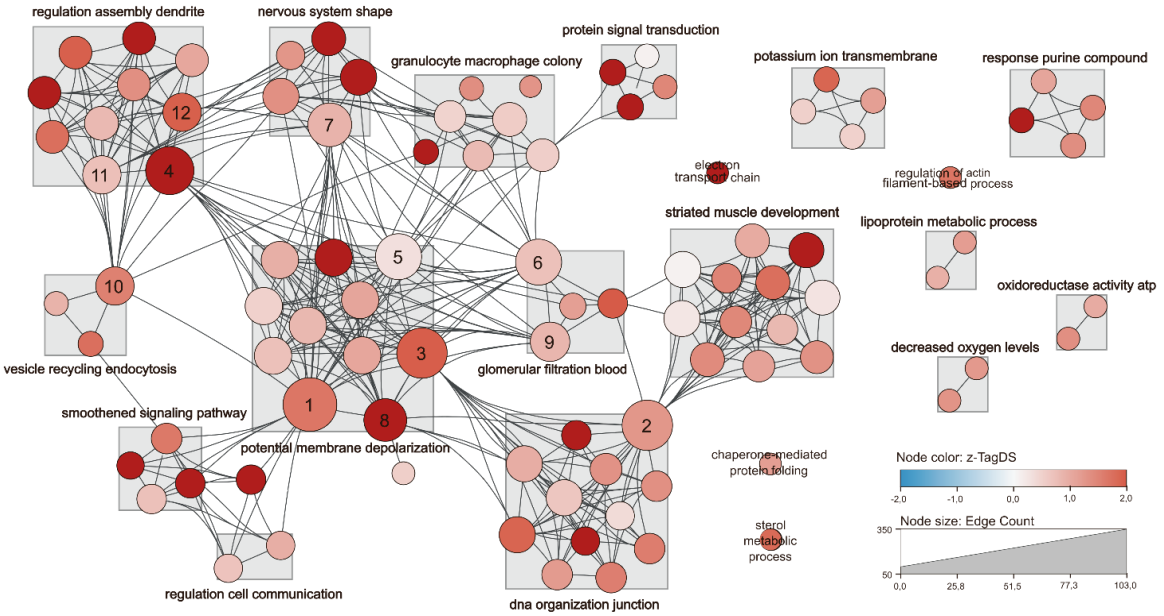


Figure S11: Network of GO-Terms associated with positive tagDS and negative impact on cognitive ability, Related to Figure 5.

This figure presents a Revigo-generated[49] network of GOterms based on GOterm associted with negative impact on cognitive ability and positive tagDS. Each node symbolizes a specific GOterm or GOterm cluster (clustering by standard Revigo criteria[49]). Node color denotes the tagDS value associated (red for positive tagDS). The node size of each node correlates with its edge count, reflecting the extent of its connectivity and relevance within the network. Larger nodes represent higher numbers of interactions with other GO-terms. Links between nodes depict the connection between GO-Terms. The bold text represents the supra-cluster defined by ReviGO[49], represented by a grey square. Nodes had at least 15 edge counts: 1) positive regulation of excitatory postsynaptic potential; 2) axonogenesis; 3) regulation of cellular component size; 4) regulation of synapse organization; 5) positive regulation of vasoconstriction; 6) regulation of glomerular filtration; 7) regulation of cell shape; 8) cell volume homeostasis; 9) negative regulation of blood pressure; 10) positive regulation of endocytosis; 11) positive regulation of dendrite development; 12) regulation of dendrite development.

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

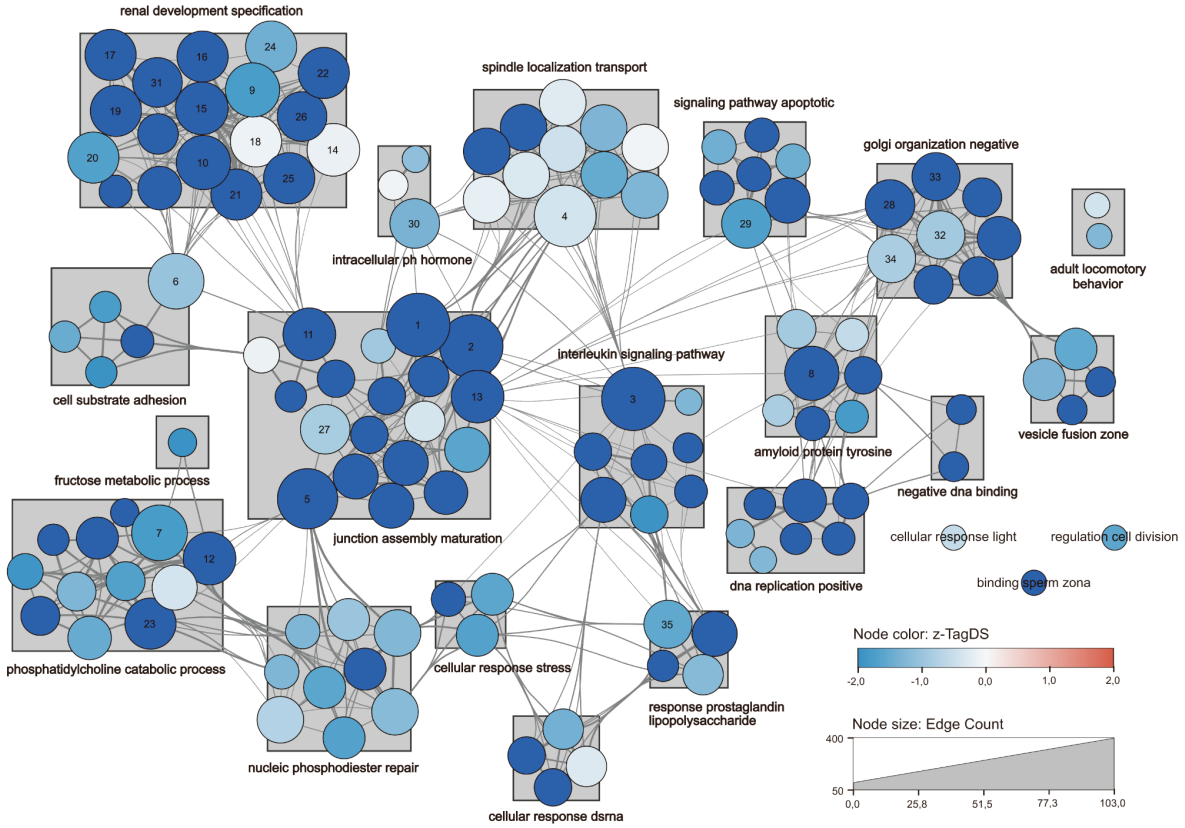


Figure S12: Network of GO-Terms associated with negative tagDS and negative impact on cognitive ability, Related to Figure 5.

This figure presents a Revigo-generated[49] network of GOterms based on GOterm associted with negative impact on cognitive ability and negative tagDS. Each node symbolizes a specific GOterm or GOterm cluster (clustering by standard Revigo criteria[49]). Node color denotes the tagDS value associated (blue for negative tagDS). The node size of each node correlates with its edge count, reflecting the extent of its connectivity and relevance within the network. Larger nodes represent higher numbers of interactions with other GO-terms. Links between nodes depict the connection between GO-Terms. The bold text represents the supra-cluster defined by ReviGO[49], represented by a grey square. Nodes had at least 15 edge counts: 1) vesicle fusion; 2) phagolysosome assembly; 3) signal release; 4) establishment of spindle localization; 5) telomere maintenance; 6) neuron migration; 7) NAD metabolic process; 8) negative regulation of translation; 9) cardiocyte differentiation; 10) cardiac neural crest cell development involved in heart development; 11) synapse maturation; 12) nucleoside phosphate biosynthetic process; 13) mitotic G2/M transition checkpoint; 14) muscle tissue development; 15) formation of primary germ layer; 16) trachea development; 17) somite development; 18) skeletal muscle organ development; 19) segmentation; 20) renal tubule development; 21) embryonic pattern specification; 22) cerebellar cortex development; 23) ceramide biosynthetic process; 24) blastocyst formation; 25) axis specification; 26) trachea morphogenesis; 27) nuclear division; 28) negative regulation of cell projection organization; 29) negative regulation of apoptotic signaling pathway; 30) hormone transport; 31) anterior/posterior axis specification; 32) negative regulation of supramolecular fiber organization; 33) negative regulation of neuron projection development; 34) negative regulation of cytoskeleton organization; 35) lipopolysaccharide-mediated signaling pathway.

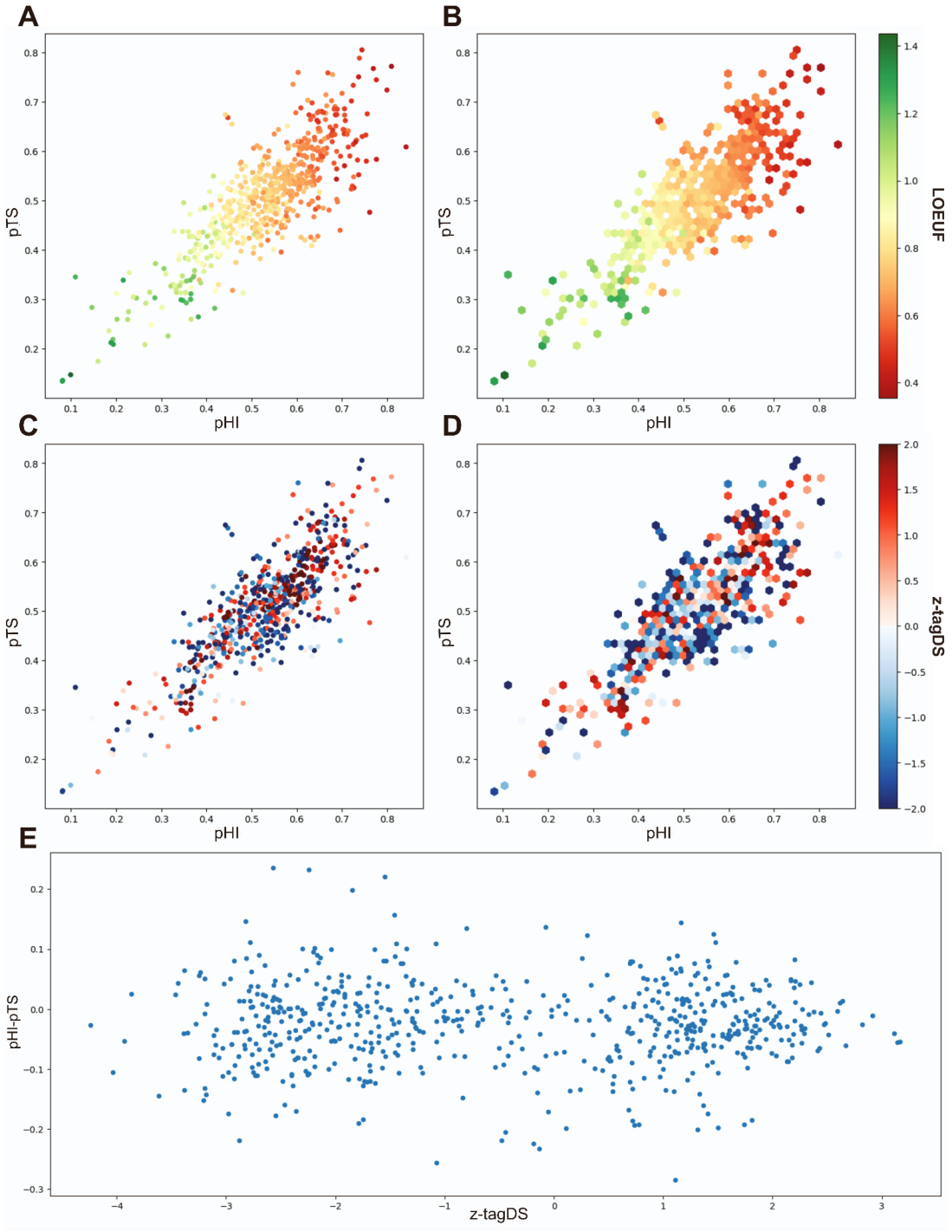


Figure S13: Comparison between Collins et al.'s hapoinsufficiency, triplosensitive scores and tagDS, Related to Figure 5.

Scatter (A) and hexbin (B) plots of the distribution of mean pHI and pTS of genes coming from the 645 significant GO-terms. The color-code represent the average LOEUF value of GO-terms. Hexbin plots visualize the average between overlapping points. Scatter (C) and hexbin (D) plots representing again the distribution of mean pHI and pTS of genes coming from the 645 significant GO-terms. The color-code represent the z-tagDS value extract from Figure 6. (E) Comparison between gene dosage specificity for cognitive ability as describe by z-tagDS and gene dosage specificity based on pHI, pTS difference.

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

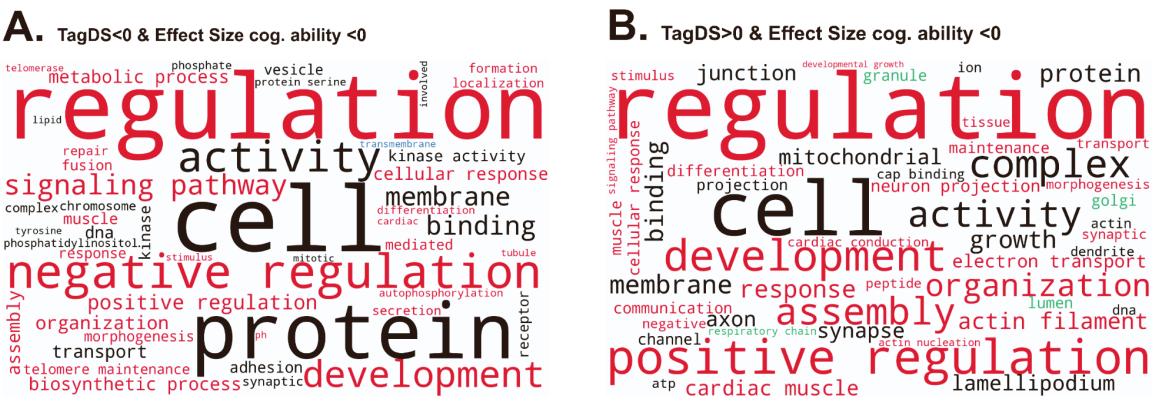


Figure S14: Word-cloud plots on effects on cognitive ability of gene clustered, based on Gene ontology annotation, Related to Figure 5.

Word-cloud plots (A) and (B) display the 50 most frequent words within GO-term names and having a negative effect on cognitive abilities, with negative (n GO-term = 279) and positive ($n_{GO\text{-}term}$ = 242) tagDS, respectively. Red, blue, and green words represent biological processes, molecular functions and cellular components terms, respectively. Black words are observed in multiple categories.

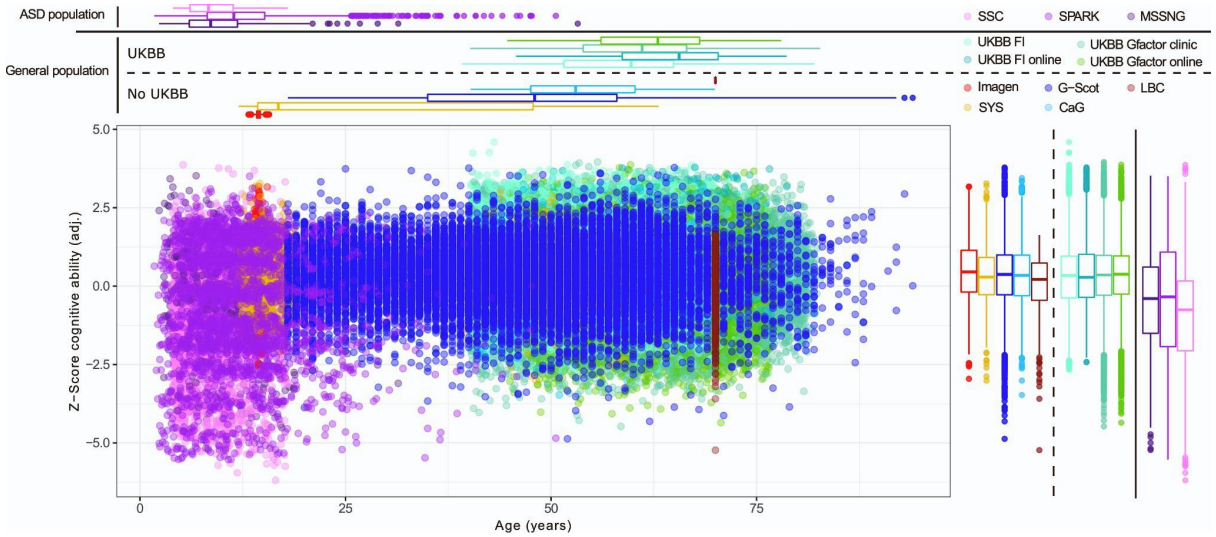


Figure S15: Description of cohorts in the dataset, Related to Table 1.

The figure showed the cognitive ability z-score adjusted (sex, age, PC1 to 10 and cognitive test) with age for each participant.

Supplementary data : Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

Supplementary reference :

1. Naj, A.C., Beecham, G.W., Martin, E.R., Gallins, P.J., Powell, E.H., Konidari, I., Whitehead, P.L., Cai, G., Haroutunian, V., Scott, W.K., et al. (2010). Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. PLoS Genet *6*, e1001130. https://doi.org/10.1371/journal.pgen.1001130.
2. Adhikari, K., Reales, G., Smith, A.J.P., Konka, E., Palmen, J., Quinto-Sanchez, M., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Fuentes, M., et al. (2015). A genome-wide association study identifies multiple loci for variation in human ear morphology. Nat Commun *6*, 7500. https://doi.org/10.1038/ncomms8500.
3. Goes, F.S., McGrath, J., Avramopoulos, D., Wolyniec, P., Pirooznia, M., Ruczinski, I., Nestadt, G., Kenny, E.E., Vacic, V., Peters, I., et al. (2015). Genome-wide association study of schizophrenia in Ashkenazi Jews. Am J Med Genet B Neuropsychiatr Genet *168*, 649–659. https://doi.org/10.1002/ajmg.b.32349.
4. Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J.-C., Al-Saadi, F., Johansson, J.A., Quinto-Sanchez, M., Acuña-Alonzo, V., et al. (2016). A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. Nat Commun *7*, 10815. https://doi.org/10.1038/ncomms10815.
5. Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet *48*, 709–717. https://doi.org/10.1038/ng.3570.
6. Adhikari, K., Fuentes-Guajardo, M., Quinto-Sánchez, M., Mendoza-Revilla, J., Camilo Chacón-Duque, J., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R.B., Pérez, G.M., et al. (2016). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat Commun *7*, 11616. https://doi.org/10.1038/ncomms11616.
7. Wu, S., Tan, J., Yang, Y., Peng, Q., Zhang, M., Li, J., Lu, D., Liu, Y., Lou, H., Feng, Q., et al. (2016). Genome-wide scans reveal variants at EDAR predominantly affecting hair straightness in Han Chinese and Uyghur populations. Hum Genet *135*, 1279–1286. https://doi.org/10.1007/s00439-016-1718-y.
8. Ahola-Olli, A.V., Würtz, P., Havulinna, A.S., Aalto, K., Pitkänen, N., Lehtimäki, T., Kähönen, M., Lyytikäinen, L.-P., Raitoharju, E., Seppälä, I., et al. (2017). Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. Am J Hum Genet *100*, 40–50. https://doi.org/10.1016/j.ajhg.2016.11.007.
9. Chesi, A., Mitchell, J.A., Kalkwarf, H.J., Bradfield, J.P., Lappe, J.M., Cousminer, D.L., Roy, S.M., McCormack, S.E., Gilsanz, V., Oberfield, S.E., et al. (2017). A Genomewide Association Study Identifies Two Sex-Specific Loci, at SPTB and IZUMO3, Influencing Pediatric Bone Mineral Density at Multiple Skeletal Sites. J Bone Miner Res *32*, 1274–1281. https://doi.org/10.1002/jbmr.3097.
10. Suhre, K., Arnold, M., Bhagwat, A.M., Cotton, R.J., Engelke, R., Raffler, J., Sarwath, H., Thareja, G., Wahl, A., DeLisle, R.K., et al. (2017). Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun *8*, 14357. https://doi.org/10.1038/ncomms14357.
11. Jia, P., Zhao, Z., Hulgan, T., Bush, W.S., Samuels, D.C., Bloss, C.S., Heaton, R.K., Ellis, R.J., Schork, N., Marra, C.M., et al. (2017). Genome-wide association study of HIV-associated neurocognitive disorder (HAND): A CHARTER group study. Am J Med Genet B Neuropsychiatr Genet *174*, 413–426. https://doi.org/10.1002/ajmg.b.32530.

12. Shaffer, J.R., Li, J., Lee, M.K., Roosenboom, J., Orlova, E., Adhikari, K., 23andMe Research Team, Gallo, C., Poletti, G., Schuler-Faccini, L., et al. (2017). Multiethnic GWAS Reveals Polygenic Architecture of Earlobe Attachment. Am J Hum Genet *101*, 913–924. https://doi.org/10.1016/j.ajhg.2017.10.001.
13. Haaland, Ø.A., Lie, R.T., Romanowska, J., Gjerdevik, M., Gjessing, H.K., and Jugessur, A. (2018). A Genome-Wide Search for Gene-Environment Effects in Isolated Cleft Lip with or without Cleft Palate Triads Points to an Interaction between Maternal Periconceptional Vitamin Use and Variants in ESRRG. Front Genet *9*, 60. https://doi.org/10.3389/fgene.2018.00060.
14. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun *9*, 2098. https://doi.org/10.1038/s41467-018-04362-x.
15. Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., et al. (2018). Genomic atlas of the human plasma proteome. Nature *558*, 73–79. https://doi.org/10.1038/s41586-018-0175-2.
16. Endo, C., Johnson, T.A., Morino, R., Nakazono, K., Kamitsuji, S., Akita, M., Kawajiri, M., Yamasaki, T., Kami, A., Hoshi, Y., et al. (2018). Genome-wide association study in Japanese females identifies fifteen novel skin-related trait associations. Sci Rep *8*, 8974. https://doi.org/10.1038/s41598-018-27145-2.
17. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet *50*, 1112–1121. https://doi.org/10.1038/s41588-018-0147-3.
18. Wyss, A.B., Sofer, T., Lee, M.K., Terzikhan, N., Nguyen, J.N., Lahousse, L., Latourelle, J.C., Smith, A.V., Bartz, T.M., Feitosa, M.F., et al. (2018). Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. Nat Commun *9*, 2976. https://doi.org/10.1038/s41467-018-05369-0.
19. Emilsson, V., Ilkov, M., Lamb, J.R., Finkel, N., Gudmundsson, E.F., Pitts, R., Hoover, H., Gudmundsdottir, V., Horman, S.R., Aspelund, T., et al. (2018). Co-regulatory networks of human serum proteins link genetics to disease. Science *361*, 769–773. https://doi.org/10.1126/science.aaq1327.
20. Yap, C.X., Sidorenko, J., Wu, Y., Kemper, K.E., Yang, J., Wray, N.R., Robinson, M.R., and Visscher, P.M. (2018). Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. Nat Commun *9*, 5407. https://doi.org/10.1038/s41467-018-07862-y.
21. Whitfield, J.B., Zhu, G., Madden, P.A.F., Montgomery, G.W., Heath, A.C., and Martin, N.G. (2019). Biomarker and Genomic Risk Factors for Liver Function Test Abnormality in Hazardous Drinkers. Alcohol Clin Exp Res *43*, 473–482. https://doi.org/10.1111/acer.13949.
22. Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet *104*, 65–75. https://doi.org/10.1016/j.ajhg.2018.11.008.
23. Morris, J.A., Kemp, J.P., Youlten, S.E., Laurent, L., Logan, J.G., Chai, R.C., Vulpescu, N.A., Forgetta, V., Kleinman, A., Mohanty, S.T., et al. (2019). An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet *51*, 258–266. https://doi.org/10.1038/s41588-018-0302-x.
24. Gialluisi, A., Andlauer, T.F.M., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., Ludwig, K.U., Czamara, D., St Pourcain, B., Brandler, W., et al. (2019). Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. Transl Psychiatry *9*, 77. https://doi.org/10.1038/s41398-019-0402-0.
25. Ivarsdottir, E.V., Benonisdottir, S., Thorleifsson, G., Sulem, P., Oddsson, A.,

Styrkarsdottir, U., Kristmundsdottir, S., Arnadottir, G.A., Thorgeirsson, G., Jonsdottir, I., et al. (2019). Sequence variation at ANAPC1 accounts for 24% of the variability in corneal endothelial cell density. Nat Commun *10*, 1284. https://doi.org/10.1038/s41467-019-09304-9.

26. Coltell, O., Sorlí, J.V., Asensio, E.M., Fernández-Carrión, R., Barragán, R., Ortega-Azorín, C., Estruch, R., González, J.I., Salas-Salvadó, J., Lamon-Fava, S., et al. (2019). Association between taste perception and adiposity in overweight or obese older subjects with metabolic syndrome and identification of novel taste-related genes. Am J Clin Nutr *109*, 1709–1723. https://doi.org/10.1093/ajcn/nqz038.
27. Akiyama, M., Ishigaki, K., Sakaue, S., Momozawa, Y., Horikoshi, M., Hirata, M., Matsuda, K., Ikegawa, S., Takahashi, A., Kanai, M., et al. (2019). Characterizing rare and low-frequency height-associated variants in the Japanese population. Nat Commun *10*, 4393. https://doi.org/10.1038/s41467-019-12276-5.
28. Zhu, Z., Guo, Y., Shi, H., Liu, C.-L., Panganiban, R.A., Chung, W., O'Connor, L.J., Himes, B.E., Gazal, S., Hasegawa, K., et al. (2020). Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. J Allergy Clin Immunol *145*, 537–549. https://doi.org/10.1016/j.jaci.2019.09.035.
29. Liu, C., and Yu, J. (2019). Genome-Wide Association Studies for Cerebrospinal Fluid Soluble TREM2 in Alzheimer's Disease. Front Aging Neurosci *11*, 297. https://doi.org/10.3389/fnagi.2019.00297.
30. Alic, L., Papac-Milicevic, N., Czamara, D., Rudnick, R.B., Ozsvar-Kozma, M., Hartmann, A., Gurbisz, M., Hoermann, G., Haslinger-Hutter, S., Zipfel, P.F., et al. (2020). A genome-wide association study identifies key modulators of complement factor H binding to malondialdehyde-epitopes. Proc Natl Acad Sci U S A *117*, 9942–9951. https://doi.org/10.1073/pnas.1913970117.
31. Qian, H., Kowalski, M.H., Kramer, H.J., Tao, R., Lash, J.P., Stilp, A.M., Cai, J., Li, Y., and Franceschini, N. (2020). Genome-Wide Association of Kidney Traits in Hispanics/Latinos Using Dense Imputed Whole-Genome Sequencing Data: The Hispanic Community Health Study/Study of Latinos. Circ Genom Precis Med *13*, e002891. https://doi.org/10.1161/CIRCGEN.119.002891.
32. Wu, Y., Cao, H., Baranova, A., Huang, H., Li, S., Cai, L., Rao, S., Dai, M., Xie, M., Dou, Y., et al. (2020). Multi-trait analysis for genome-wide association study of five psychiatric disorders. Transl Psychiatry *10*, 1–11. https://doi.org/10.1038/s41398-020-00902-6.
33. van der Meer, D., Frei, O., Kaufmann, T., Shadrin, A.A., Devor, A., Smeland, O.B., Thompson, W.K., Fan, C.C., Holland, D., Westlye, L.T., et al. (2020). Understanding the genetic determinants of the brain with MOSTest. Nat Commun *11*, 3512. https://doi.org/10.1038/s41467-020-17368-1.
34. Chen, M.-H., Raffield, L.M., Mousas, A., Sakaue, S., Huffman, J.E., Moscati, A., Trivedi, B., Jiang, T., Akbari, P., Vuckovic, D., et al. (2020). Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. Cell *182*, 1198-1213.e14. https://doi.org/10.1016/j.cell.2020.06.045.
35. Cuellar-Partida, G., Tung, J.Y., Eriksson, N., Albrecht, E., Aliev, F., Andreassen, O.A., Barroso, I., Beckmann, J.S., Boks, M.P., Boomsma, D.I., et al. (2021). Genome-wide association study identifies 48 common genetic variants associated with handedness. Nat Hum Behav *5*, 59–70. https://doi.org/10.1038/s41562-020-00956-y.
36. Tideman, J.W.L., Pärssinen, O., Haarman, A.E.G., Khawaja, A.P., Wedenoja, J., Williams, K.M., Biino, G., Ding, X., Kähönen, M., Lehtimäki, T., et al. (2021). Evaluation of Shared Genetic Susceptibility to High and Low Myopia and Hyperopia. JAMA Ophthalmol *139*, 601–609. https://doi.org/10.1001/jamaophthalmol.2021.0497.
37. Shelton, J.F., Shastri, A.J., Ye, C., Weldon, C.H., Filshtein-Sonmez, T., Coker, D., Symons, A., Esparza-Gordillo, J., 23andMe COVID-19 Team, Aslibekyan, S., et al. (2021). Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. Nat Genet *53*, 801–808.

https://doi.org/10.1038/s41588-021-00854-7.
38. Yao, Y., Chu, X., Ma, M., Ye, J., Wen, Y., Li, P., Cheng, B., Cheng, S., Zhang, L., Liu, L., et al. (2021). Evaluate the effects of serum urate level on bone mineral density: a genome-wide gene-environment interaction analysis in UK Biobank cohort. Endocrine *73*, 702–711. https://doi.org/10.1007/s12020-021-02760-8.
39. Cipriani, V., Tierney, A., Griffiths, J.R., Zuber, V., Sergouniotis, P.I., Yates, J.R.W., Moore, A.T., Bishop, P.N., Clark, S.J., and Unwin, R.D. (2021). Beyond factor H: The impact of genetic-risk variants for age-related macular degeneration on circulating factor-H-like 1 and factor-H-related protein concentrations. Am J Hum Genet *108*, 1385–1400. https://doi.org/10.1016/j.ajhg.2021.05.015.
40. Inoue, Y., Hasebe, Y., Igarashi, T., Kawagishi-Hotta, M., Okuno, R., Yamada, T., and Hasegawa, S. (2021). Search for genetic loci involved in the constitution and skin type of a Japanese women using a genome-wide association study. Exp Dermatol *30*, 1787–1793. https://doi.org/10.1111/exd.14430.
41. Shadrin, A.A., Kaufmann, T., van der Meer, D., Palmer, C.E., Makowski, C., Loughnan, R., Jernigan, T.L., Seibert, T.M., Hagler, D.J., Smeland, O.B., et al. (2021). Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. Neuroimage *244*, 118603. https://doi.org/10.1016/j.neuroimage.2021.118603.
42. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet *53*, 1415–1424. https://doi.org/10.1038/s41588-021-00931-x.
43. Jiang, L., Zheng, Z., Fang, H., and Yang, J. (2021). A generalized linear mixed model association tool for biobank-scale data. Nat Genet *53*, 1616–1621. https://doi.org/10.1038/s41588-021-00954-4.
44. van der Meer, D., Kaufmann, T., Shadrin, A.A., Makowski, C., Frei, O., Roelfs, D., Monereo-Sánchez, J., Linden, D.E.J., Rokicki, J., Alnæs, D., et al. (2021). The genetic architecture of human cortical folding. Science Advances *7*, eabj9446. https://doi.org/10.1126/sciadv.abj9446.
45. Gudjonsson, A., Gudmundsdottir, V., Axelsson, G.T., Gudmundsson, E.F., Jonsson, B.G., Launer, L.J., Lamb, J.R., Jennings, L.L., Aspelund, T., Emilsson, V., et al. (2022). A genome-wide association study of serum proteins reveals shared loci with common diseases. Nat Commun *13*, 480. https://doi.org/10.1038/s41467-021-27850-z.
46. Mitchell, B.L., Saklatvala, J.R., Dand, N., Hagenbeek, F.A., Li, X., Min, J.L., Thomas, L., Bartels, M., Jan Hottenga, J., Lupton, M.K., et al. (2022). Genome-wide association meta-analysis identifies 29 new acne susceptibility loci. Nat Commun *13*, 702. https://doi.org/10.1038/s41467-022-28252-5.
47. Lyall, D.M., Cullen, B., Allerhand, M., Smith, D.J., Mackay, D., Evans, J., Anderson, J., Fawns-Ritchie, C., McIntosh, A.M., Deary, I.J., et al. (2016). Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. PLOS ONE *11*, e0154222. https://doi.org/10.1371/journal.pone.0154222.
48. Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., et al. (2019). SynGO: an evidence-based, expert-curated knowledgebase for the synapse. Neuron *103*, 217-234.e4. https://doi.org/10.1016/j.neuron.2019.05.002.
49. Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLOS ONE *6*, e21800. https://doi.org/10.1371/journal.pone.0021800.