Title: Mirror effect of genomic deletions and duplications on cognitive ability across the human cerebral cortex

Authors:

Kuldeep Kumar[1]*, Sayeh Kazem[1]*, Guillaume Huguet[1], Thomas Renne[1], Worrawat Engchuan[2,3], Martineau Jean-Louis[1], Jakub Kopal[4], Zohra Saci[1], Omar Shanta[5], Bhooma Thiruvahindrapuram[3], Jeffrey R. MacDonald[3], Josephine Mollon[6], Laura M. Schultz[7], Emma E.M. Knowles[6], David Porteous[8,9,10], Gail Davies[8], Paul Redmond[8], Sarah E. Harris[8], Simon R. Cox[8], Gunter Schumann[11], Zdenka Pausova[12,13,14], Celia M. T. Greenwood[15,16], Tomas Paus[1,17], Stephen W. Scherer[2,3,18], Laura Almasy[7,19], Jonathan Sebat[5], David C Glahn[6,20], Guillaume Dumas[1,21]**, Sébastien Jacquemont[1,22]**.

Affiliations:

1 Centre de recherche CHU Sainte-Justine and University of Montreal, Montreal, QC, Canada

2 The Hospital for Sick Children, Genetics and Genome Biology, Toronto, ON, Canada

3 The Hospital for Sick Children, The Centre for Applied Genomics, Toronto, ON, Canada

4 University of Oslo, NORMENT, Oslo, Norway

5 University of California, San Diego, Department of Psychiatry, Department of Cellular & Molecular Medicine, Beyster Center of Psychiatric Genomics, San Diego, CA, USA

6 Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

7 Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

8 Lothian Birth Cohorts, Department of Psychology, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Edinburgh, UK

9 Medical Genetics Section, Centre for Genomic & Experimental Medicine, MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK

10 Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK

11 Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK

12 Research Institute of the Hospital for Sick Children, Toronto, ON, Canada

13 Departments of Physiology and Nutritional Sciences, University of Toronto, Toronto, ON, Canada

14 ECOGENE-21, Chicoutimi, QC, Canada

15 Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada

16 Departments of Oncology, and Epidemiology, Biostatistics and Occupational Health, and Division of Cancer Epidemiology, McGill University, Montreal, QC, Canada

17 Department of Psychiatry and Addictology, Department of Neuroscience, Faculty of Medicine, University of Montreal, Montreal, QC, Canada

18 McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

19 Lifespan Brain Institute, Children's Hospital of Philadelphia and Penn Medicine, PA, USA

20 Boston Children's Hospital, Tommy Fuss Center for Neuropsychiatric Disease Research, Boston, MA, USA

21 Mila – Quebec Artificial Intelligence Institute, Montreal, QC, Canada

22 Department of Pediatrics, University of Montreal, Montreal, QC, Canada

* authors contributed equally

** Shared senior authorship

Abstract

Regulation of gene expression shapes the interaction between brain networks which in-turn supports psychological processes such as cognitive ability. How changes in level of gene expression across the cerebral cortex influence cognitive ability remains unknown. Here, we tackle this by leveraging genomic deletions and duplications - copy number variants (CNVs) that fully encompass one or more genes expressed in the human cortex - which lead to large effects on gene-expression levels. We assigned genes to 180 regions of the human cerebral cortex based on their preferential expression across the cortex computed using data from the Allen Human Brain Atlas. We aggregated CNVs in cortical regions, and ran a burden association analysis to compute the mean effect size of genes on general cognitive ability for each of the 180 regions. When affected by CNVs, most of the regional gene-sets were associated with lower cognitive ability. The spatial patterns of effect sizes across the cortex were correlated negatively between deletions and duplications. The largest effect sizes for deletions and duplications were observed for gene-sets with high expression in sensorimotor and association regions, respectively. These two opposing patterns of effect sizes were not influenced by intolerance to loss of function, demonstrating orthogonality to dosage-sensitivity scores. The same mirror patterns were also observed after stratifying genes based on cell types and developmental epochs markers. These results suggest that the effect size of gene dosage on cognitive ability follows a cortical gradient. The same brain region and corresponding gene-set may show different effects on cognition depending on whether variants increase or decrease transcription. The latter has major implications for the association of brain networks with phenotypes

.

Introduction

Psychological processes such as cognitive ability arise from large-scale brain networks [1–4], which integrate a diverse range of microscale functions influenced by genetic variations. Recent studies have linked spatial patterns of gene expression in healthy individuals[5] to structural and functional macroscale brain networks. However, despite significant progress, the role of variations in gene expression across the cerebral cortex in underpinning cognitive ability remains unknown. Specifically, in the absence of large transcriptomic datasets of brain tissue with corresponding phenotypes, a direct association between variation in transcription levels and cognition remains out of reach.

To circumvent this issue, we take advantage of two opportunities. First, over 10% [6,7] of the general population carries a genomic deletion or duplication fully encompassing one or more coding genes. These deletions and duplications are associated with large decreases and increases (> 1.5 to 2.5 standard deviations)[8], respectively, of the expression of genes they encompass[9]. Second, functional genomic resources such as the AHBA provide information on the spatial distribution of gene expression, allowing us to identify brain regions where genes are preferentially expressed.

Previous studies have estimated that a large proportion of the coding genome is associated with lower cognitive abilities when deleted or duplicated [6,7,10]. Most of our current knowledge of gene functions involved in cognitive abilities stem from studies investigating ontologies at the cellular, subcellular and molecular level, i.e., genes associated with intellectual disability are enriched in chromatin remodelling, gene regulation and neuronal communication processes[11,12]. However, whether the effects of decrease and increase in transcription (deleted and duplicated genes) on cognitive ability can be explained by their role in large scale networks across the cortex has not been investigated. We postulate that information on spatial profiles

of gene expression across the human cerebral cortex will provide novel insights into the effect sizes on cognitive ability of genomic deletions and duplications.

Results

Deleted and duplicated genes show specific spatial profiles of effect sizes across the cerebral cortex.

To understand the relationship between where genes are expressed in cortical networks, and their impact on cognitive ability, we first defined gene-sets preferentially expressed in 180 regions of the human cerebral cortex based on the Glasser atlas [13]. To do so, we used normalized spatial profiles of gene expression from the Allen Human Brain Atlas[14]. A total of 13,678 genes with detectable expression in the cerebral cortex were assigned to cortical regions if their normalized value was higher than 0.5 z-score. As a result we obtained 180 gene sets corresponding to 180 cortical regions.

We then identified all CNVs > 50 kilobases in 258,513 individuals from an aggregated dataset combining 6 unselected populations with assessments of general cognitive ability (Table 1). 8% of individuals carried at least one deletion/duplication fully encompassing one or more genes assigned to cortical regions. Among the 13,678 genes, 4,405 (32.2%) were deleted and 8,601 (62.9%) duplicated at least once (**Fig. 1a**). Given that deletion/duplication may occur multiple times, we observed a total of 28,108 and 77,695 deleted and duplicated genes, respectively (**Fig. 1b-c, Extended Data Fig. 2a**). The count of deleted and duplicated genes in each cortical region show the same spatial distribution across the cerebral cortex (r=0.97, p-spin<0.001, **Fig. 1g-h, Supplementary Table 6-8**), demonstrating an absence of bias of the distribution between deletions and duplications across the cortex.

Because the majority of CNVs affecting general cognitive ability are too rare to investigate at the variant- or gene-level, we performed a brain-based GWAS by i) aggregating all CNVs affecting gene-sets assigned to cortical regions and ii) performing 180 linear models to estimate the mean effect on general cognitive ability of genes assigned to each of the 180 cortical regions. Deletions and duplications were analyzed separately (**Fig. 1i-j**). To identify potential patterns, we projected color-scaled effect sizes on the Glasser parcellation (**Fig. 1k)**. For deletions, 128 out of 180 cortical regions showed significant (Bonferroni) negative impacts on general cognitive ability, with a maximum effect size of Cohen's d= -0.12. For duplication, 112 regions also showed significant negative impacts (approximately 2-fold smaller: max Cohen's d = -0.05, **Supplementary Table 2**). For deletions, the largest effect sizes were observed in sensorimotor regions (Cohen's d= -0.12) and were significantly higher than the mean effect of all genes expressed in the cerebral cortex (Cohen's d= -0.06). For duplications, the largest effect sizes were observed in association regions (Cohen's d= -0.06) and were also significantly higher than the mean effect of genes expressed in the cerebral cortex (Cohen's d= -0.02) (**Supplementary Table 2-3**). Regions with the larget effects in deletions were those with the smallest effects in duplications and *vice versa* (**Fig. 1l**). A group of regions including those clustered around the dorsolateral prefrontal cortex (DLPFC) were significantly altered in both deletions and duplications. These patterns of effect sizes across the cerebral cortex were robust and highly correlated between discovery, replication and full datasets as well as between UKBB and the other unselected populations (**Extended Data Fig. 4a-l, Supplementary Table 5**).

Deletion and duplication show mirror effects on cognitive ability across the human cerebral cortex

Previous studies have observed negative impacts of deletions and duplications on cognitive ability characterised by the inverted U shape effect of the average cortically expressed gene. Since it has been challenging to identify genes with preferential sensitivity to either deletions or duplications, we therefore systematically compared the effect sizes of both CNVs across all cortical regions. We observed a negative spatial correlation between effect sizes of deletions and duplications (r = -0.3, p-spin=1.3E-3), which was replicated (r= -0.52, p-spin<1E-4, **Extended Data Fig. 4m, Supplementary Table 2,4**) using an independent gene expression resource[15]. This negative spatial correlation could also be observed (r=-0.59, p=0.017, **Extended Data Fig. 4o**) using 16 cortical gene-sets defined by a co-expression analysis from an independent group and study[16], showing clustering of sensorimotor and association region enriched modules for deletions and duplications, respectively.

Sensitivity analyses showed that associations between deletion and duplication patterns and cortical hierarchies were largely unchanged in i) data-subsets; ii) restricting the analysis to specific cognitive assessment tools, iii) removing genes with lower differential stability scores (DS) (**Extended Data Fig. 3-5, Supplementary Table 2,4**).

Deletion and duplication effects on cognitive ability follow a unimodal transmodal cortical axis

To provide a basis for the structural and function interpretation of these CNV effect size maps, we systematically annotated them using 20 reference neuromaps[17] (**Fig. 2a,b**). The 20 neural maps provide molecular, microstructural, electrophysiological, developmental and functional characterization of human cerebral cortex. Deletion and duplication effect size maps showed

mild to moderate positive and negative correlations respectively with the cortical hierarchies following a unimodal-transmodal (sensorimotor-association) axes: cortical-thickness, PC1 gene expression, functional gradient, T1w/T2w, PC1 NeuroSynth, and high gamma power. In addition, mild to moderate spatial correlations were also observed for some maps that showed significant correlations with either deletion or duplication effect size maps of cognitive ability. These mild to moderate spatial correlations suggest a novel uni-modal/transmodal cortical hierarchy reflecting gene dosage effects on cognitive ability.

To annotate the spatial patterns of effect sizes of deletions and duplications, we used resting-state contextualization to map effect sizes to 17 functional connectivity cortical networks (**Fig. 2d**). The somatomotor, and visual networks showed the largest effect sizes, for deletions, while temporal-parietal, and limbic networks demonstrated the smallest effects. For duplications, the largest effect sizes were observed in the language, frontoparietal, default, and orbito-affective networks; and the smallest effects were observed for visual, posterior-multimodal, and dorsal-attention networks.

We asked if the cortical gradient of effect sizes observed for deletion and duplication would be influenced by the level of regional specificity of the "cortical" genes. To do so, we focussed on the most uni- and trans-modal regions using a previously established stratification [18] (**Fig. 2c**). We then computed the 180 gene-sets for three increasing levels of "regional specificity" (and decreasing gene-set size) by thresholding normalized expression > 0.5, >0.8 and >1.0 z-scores. Using gene-sets with increasing regional specificity, the effect sizes of deletions and duplications increased for sensorimotor (p-value = 0.02 and 0.05) and association regions (p-value= 0.02, and 0.09) respectively and the same cortical pattern was observed across all 3 levels of specificity. Sample size (power) did not allow to further increase regional specificity (**Extended Data Fig. 3c, Supplementary Table 1-2)**.

Intolerance to haploinsufficiency does not influence mirror effects of deletions and duplications across the cerebral cortex

Measures of intolerance to loss of function (LoF) remain to date the best predictor of effect sizes of deletions and duplications on cognitive abilities and risk for neurodevelopmental disorders [6,10,19]. We therefore tested if the findings presented above could be explained by measures of intolerance to LoF such as LOEUF[20], pHS, pTS[21]. We first considered all genes with detectable expression in the cerebral cortex and showed that their LOEUF values were very weakly correlated to their spatial expression (measured by the first dimension of the spatial distribution of gene expression across the cortex) (**Fig. 3a-b)**.

We then asked if LOEUF influences the spatial pattern of effect sizes on general cognitive ability of deletion and duplication. To do so, we estimated the effects on cognition of 180 cortical gene list as performed above after stratifying genes into 3 groups: intolerant (LOEUF<0.35), mildly intolerant (0.35<LOEUF<1.0), and tolerant (LOEUF>1.0). As expected, increasing intolerance to LoF was associated with increasing effect sizes (**Fig. 3d-f**). The spatial patterns of effect sizes on cognitive ability remained similar and significantly anticorrelated between deletions and duplications even for the group of tolerant genes (-0.38< r <-0.71, p-spin<1.3E-3, **Fig. 3e**). The differential effects of sensorimotor and association cortical regions were observed in all LOEUF groups but appeared higher for the intolerant group (**Fig. 3f, Supplementary Table 3**). Using alternative dosage sensitivity scores (pHS, pTS) provided similar results (**Extended Data Fig. 6c-f, Supplementary Table 3**).

Finally we focussed on a group of 256 genes previously associated with Intellectual disability, and characterized by extreme intolerance to LoF and effect sizes on cognitive abilities [6,22]. We found their expression to be enriched in unimodal regions as compared to all-genes (Kolmogorov-Smirnov test p-value=7e-4). This was driven by genes sensitive to likely-gene-

disruptive (LGD) mutations (Kolmogorov-Smirnov test p-value=2.5e-3) when compared with those sensitive to missense mutations (Kolmogorov-Smirnov test p-value=3e-1) (Supplemental Figure).

Gene dosage effects across the unimodal/transmodal axis are not explained by specific cell types or developmental periods.

To understand if the dissociation between deletion and duplication effects across the cerebral cortex were driven by genes ubiquitously or preferentially expressed in the human brain we stratified genes into 3 non-overlapping groups[15,23,24] : i) brain elevated genes; ii) other tissue elevated and iii) ubiquitous. The largest effect sizes and dissociation between unimodal/transmodal effects on general cognitive ability were observed for the brain-elevated gene set (p-value = 5e-8); however, significant (p= 9e-5) dissociation was also observed for genes with expression elevated in other tissues (**Fig. 4a-b**, **Supplementary Table 3**). This suggests that the mechanisms involved in these cortical patterns of effect sizes apply to a broad range of genes including those not preferentially involved in brain function and structure.

The temporal sequence of cortical refinement tends to progress in a spatially ordered manner [2,25]. Sensory and motor regions are predominantly modified during the first decade of life, in accordance with the early development of vision, audition, and motor control. In contrast, transmodal association regions display a more protracted developmental course, remaining comparatively immature throughout childhood and adolescence [2]. Because cortical patterns of effect sizes on general cognitive ability observed for deletions and duplications were estimated using gene expression from adult brain tissue, we asked if results could be explained by temporal dynamics in gene expression. We computed again in the same dataset the effect sizes on general cognitive ability of gene sets in 180 cortical regions and further stratified genes into 3 non-overlapping groups based on whether peak expression occurs during either one of the 3

developmental epochs previously defined by critical transitions in gene expression[26]. Epoch 1 ending around late-fetal transition begins around 10 post-conception weeks, Epoch 2 ending at 6 postnatal months (by then, the most dramatic changes are completed), and epoch 3 reaching into adulthood **(Fig. 4c).** The largest effects and largest uni-/trans-modal dissociation for deletion and duplication are observed for the late fetal period (epoch2) and epoch3, respectively **(Fig. 4d)**. However, genes preferentially expressed during childhood and adolescence also demonstrated significant dissociation for deletions and anticorrelation between deletion and duplication effect sizes **(Fig. 4d, Supplementary Table 2-4)**. Stratifying genes based on their temporal expression trajectories into falling, non-transitional, and rising genes showed negative correlation between deletion and duplication effect size maps for non-transitional and rising genes (**Extended Data Fig. 6a, Supplementary Table 2,4**).

Since bulk tissue expression data combine expression of underlying cell-constituents, we then stratified our spatial analysis using 6 adult cell-type marker gene sets including Astrocytes, Oligodendrocyte precursor cells (OPCs), Oligodendrocytes, Microglial, Excitatory neurons, and Inhibitory neurons [15,27]. Except for OPCs, effect sizes of gene sets stratified by brain cell types were larger than those observed for all genes (**Supplementary Table 2**). The dissociation between unimodal and transmodal effect sizes was larger for neuronal cells (excitatory and inhibitory) than for glial cells and astrocytes for deletions (**Supplementary Table 3**).

The negative correlation between deletion/duplication effect sizes were also stronger for neuronal cell types and weaker for OPCs and astrocytes (**Supplementary Table 4**). This again demonstrate that effects of gene dosage on cognitive ability are distributed according to large scale brain networks and these effects encompass and override the gene functions at the microscopic -cellular- level.

These results could not be explained by any difference in the spatial distribution of expression across gene sets. In other words the spatial expression of the different gene sets was similar. (**Extended Data Fig. 2e, Supplementary Table 6-8**).

Finally, we asked if molecular and sub-cellular biological functions (as detailed by GO-terms) may explain the pattern of CNV effects across the cerebral cortex. We first sought to provide the level of enrichment of each GO-term across the cortex. We performed enrichment analyses for 1399 GO-terms across 180 cortical regions. We focussed on the 72/1399 GO-terms which showed significant enrichments in at least 5 cortical regions (FDR across 1399 GO-terms x 180 cortical regions). We then mapped on the cortex the enrichment values of these 72 GO-terms and compared them with the pattern of effect size on general cognitive ability of deletions and duplications. None of the GO-terms showed a relationship with the pattern of effect size of duplications on general cognitive ability. Only 21 GO-terms showed spatial similarities with the deletion-cognitive ability pattern (r values ranging from 0.17 to 0.28 ; **Fig. 5b**). These included telencephalon development, neural projection, synaptic and plasma membrane terms **(Fig. 5b)**.

Discussion

This is the first study to leverage CNVs in a general population and to investigate the effect of inter-individual genetic variation across the cerebral cortex on general cognitive ability. Taken together, our findings demonstrate that the effect size of deleted and duplicated genes on general cognitive ability can be explained by their spatial expression across the macroscale organization of the human cerebral cortex. Our findings identified a novel functional axis that dissociates the effect of gene dosage on general cognitive ability and shows a mirror (negative correlation) pattern between deletion and duplication effect size maps. Previous studies have shown an inverted U-shape for the effects of deletions and duplications, where both show

negative effects on cognitive abilities. Interestingly, we found that the largest effects for deletions are observed in sensorimotor regions, which show smaller or near 0 effects for duplications, and *vice versa* for association regions for duplications.

Brain-based genome wide association study: We propose a new method to estimate the aggregate association on the basis of all CNVs affecting gene dosage of all genes preferentially expressed in a given brain region. A growing number of studies have characterised how spatial transcriptional specialization shapes cortical function [23]. Our findings may therefore be interpreted as an indirect approach to interrogate the association between molecular changes in neuroanatomical regions and phenotypic traits.

Effects of gene dosage on cognitive ability depend on where and when genes are expressed along the sensory-association cortical gradient. We show that the effect sizes of deletions and duplications are respectively positively and negatively correlated to the Sensory-Association axis. The latter provides a low-dimensional and powerful way to appreciate the immense complexity of the cerebral cortex but it does not capture the entire spectrum of cortical variability. Deletions show the largest effects in sensory motor regions, which are the first to develop, around epoch 2 (late-fetal transition), which is consistent with the fact that their effects were larger for genes with peak expression during epoch 2. These findings are reminiscent of GWAS findings that have shown that the expression of genes contributing to the shared genetic architecture across psychiatric disorders peaks during mid-gestation, suggesting that this period is critical for many conditions[28–31]. Similar enrichments of peak expression in mid-gestation were observed for genes harbouring an excess of de novo rare variants in ASD and intellectual disability[26,32]. On the other hand, duplications show the largest effects in association regions, The protracted development of such regions is consistent with the fact that duplication effects were larger for genes with peak expression in epoch 3 (late childhood, adolescence and adulthood).

Opposing alterations in gene dosage may lead to decreases in general cognitive ability through different cortical mechanisms: Although this study does not analyze any neuroimaging data, it uses brain parcellations previously defined by structural and functional MRI data. One of the intriguing characteristics of the associations between inter-individual variability in MRI-derived measures of brain structure/function and psychological phenotypes is their very small effect sizes. As a result, recent publications have highlighted the fact that Brain Wide Association Studies (BWAS) should use neuroimaging datasets with at least thousands of individuals [33].

Results from our study may provide some insights into the very small effect sizes of BWAS of cognitive ability[33]. Indeed, we demonstrate that the same cortical region (e.g. motor region) may have either the largest or the smallest effect on general cognitive ability depending on the class of genetic variant affecting genes preferentially expressed in the region of interest. Given that in a large neuroimaging dataset, all classes of genomic variants will be distributed across individuals, the molecular mechanisms leading to interindividual changes in general cognitive ability will be extraordinarily heterogeneous. Results from any BWAS of psychological phenotypes will therefore represent the average effect of all molecular and environmental factors influencing the trait. Similarly, if we perform our Brain Based GWAS on CNVs (without distinguishing deletions and duplication), our analysis would result in a blured much smaller average effect.

Primary Sensory Motor regions as unrecognized key players in general cognitive ability: We show that in the context of clear haploinsufficiency, sensorimotor regions are those with the largest effects on general cognitive ability. Research has shown that the motor network is not only responsible for controlling movement but also plays a role in cognitive functions such as attention, working memory, and decision-making. Previous studies suggests that the

somatomotor network may act as a "neural workspace" that integrates information from different brain regions, allowing for the flexible use of cognitive resources. This idea is supported by studies that have found correlations between motor network activity and individual differences in general intelligence. A series of functional MRI studies have also shown that overconnectivity of the somatomotor networks is a common observation across psychiatric conditions, cognitive ability and CNVs previously associated with psychiatric conditions and cognition [34]. This is consistent with the fact that most rare loss of function deleterious variants, that increase the risk for psychiatric conditions are also associated with delayed gross motor milestones [19,35] and development coordination disorders[36].

Previous GWAS of general cognitive ability have mapped common variants to genes that are preferentially expressed in the cerebral cortex. Our results show that this association is underlined by complex patterns linked to specific two classes of variants (deletions versus duplications). There were up to 5-fold differences in effect sizes depending on where genes were expressed throughout the cerebral cortex. This might suggest that the focus of prior transcriptomic studies on DLPFC. While we highlight the DLPFC, which shows large effects for both deletions and duplication, we make the case for moving beyond the DLPF centric transcriptomic studies to investigate the whole cerebral cortex. In line with the importance of sensorimotor regions in cognitive processes, a recent whole cortex transcriptomic study has shown that the primary visual cortex was the region with the largest changes in expression levels in autism versus control *postmortem* brains.

While the deletion and duplication effect size maps correlate with previously established cortical organization hierarchies, including brain MRI derived anatomical and functional hierarchies, the largest correlations are less than -0.5, suggesting that only a proportion of variance is explained by the established maps, and a large proportion of these effect size maps are unique. This suggests a general organizing principal of the human brain, specifically,

cerebral cortex, where signals from different data sources, including our investigation of inter-individual genetic variation on cognitive ability, follow this broad sensorimotor-association axis. At the same time, it begs the question of investigating the whole brain dynamics. For example, the limbic system spans sub-cortical structures also, moreover, thalamus is connected to the whole cortex, and cerebellum has shown its gradient and behaves like a mini-brain.

Spatial effects across the cortex are not captured by known contributors to cognitive ability.

LOEUF: Constraint scores have previously been associated with neurodevelopmental and psychiatric disorders and are strong predictors of the effect size of genes on cognitive abilities. Our results demonstrate that constraint and cortical regional effects on cognition are orthogonal. As an example, for deletion, even tolerant genes expressed in sensorimotor regions have effect sizes as large as the intolerant genes expressed in association regions.

Similar cortical patterns of effect sizes on cognitive ability were observed whether genes showed elevated expression in brain, specific expression across tissues or even low expression in brain relative to other tissues. It suggests that the "spatial organizing principle" in the cerebral cortex regroups genes with multiple functions across the entire body and different levels of constraint. This also may explain why the same CNVs were found to impact whole body phenotypes [37,38].

Stratifying by cell-types showed dissociation of the deletion and duplication effect sizes between sensorimotor and association regions across 6 cell-types. While the duplication maps will need larger sample sizes for reliable estimation, the findings point towards the deconvolution of the effect size maps by constituent cell types. For astrocytes and microglia, we observe largest effect size magnitudes for sensorimotor regions for both deletions and duplication, suggesting different patterns across support cell-types, albeit warranting further investigation to establish robustly.

Limitations:

Age: Our data are comprised of a large proportion of older adults (**Table 1**), which may affect the interpretation of our results. Specifically, whether some of the effect-size estimates are related to i) developmental effects (as suggested by prominent effects of genes with late fetal expression) or ii) cognitive decline (as suggested by the the large effects driven by genes preferentially expressed in microglia or astrocytes previously linked to risk for alzheimer's) remains unknown. Nevertheless, the effect size maps presented in **Fig. 1** showed consistent patterns in the UK biobank compared to the remaining five control cohorts, as well as in the whole dataset compared to individuals below the age of 60. **Due to sparse cortical sampling in the AHBA** (203 ± 27 samples per donor; 1,220 total samples across 6 donors), some spatial interpolation was required to produce our parcellated (n=180, left hemisphere) gene expression maps. However, each gene was assigned to several contiguous brain regions covering several cortical samples in AHBA. **Less robust estimates for duplications:** One limitation of our study is the sample size, mainly for duplications. However, with larger sample size, we will be able to stratify genes further and have reliable maps for comparison.

Our findings were robust across demographic parameters, brain parcellation, and were replicated in independent gene expression datasets as well as gene-sets based on co-expression modules. Although we did assess the impact of ancestry by including PCs and running analysis for White-British only, we acknowledge that a thorough analysis for multiple ancestries is required.

In summary, our proposed framework highlights the translational value of linking functional genomic resources, for example, the postmortem-derived spatial profiles of gene expression, with relevant phenotypes. This framework provides estimates on the cerebral cortex, and the

correlation of these effect sizes maps can be seen as functional genetic correlations. This begs the question of whether it can be extended to other rare variants and provide a systematic comparison across phenotypes.

Methods

Ethics

The study protocols were approved by the institutional review board for each of the cohorts involved with recruitment. Informed consent and permission to share the data were obtained from all individuals, in compliance with the guidelines specified by the institutional review boards of the recruiting centres. This study was approved by the ethics board of CHU Sainte Justine Research Center, Montreal, Canada.

Participants

In a previous article by Kendall et al 2017, "The UK Biobank recruited ~500 000 individuals aged 40–69 years (54% female) between 2006 and 2010". Many phenotypic and cognitive measures were tested at UK Biobank assessment centres or online on participants, also added demographic, socioeconomic, and health data. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. And when the procedure involving human participants were approved by the North West Multi-Centre Ethics Committee (approval number 11/NW/0382, Data were released to Cardiff University under application number 14421). We used cognitive ability data described in previous publications [7,42].

Sample Processing and Genotyping

The DNA was extracted with blood. All details of the genotyping processes used are on the UK Biobank website. The samples were genotyped at the Affymetrix Research Laboratory, Santa Clara, CA on 96-well plates.

Samples were genotyped at Affymetrix Research Services Laboratory (Santa Clara, California, USA). Samples were genotyped on two very similar arrays with ~95% probe overlap between them. The first array did on ~50 000 on the UK BiLEVE Array (includes over 820,000 probes) and ~450 000 on the UK Biobank Axiom Array (includes over 807,000 probes). Quality control filters included genotypic call rate <0.95, |waviness factor| of <0.05, log R ratio s.d. of >0.35 and BAF s.d. of >0.08. From 488,377 people with genotypic data, 28,522 were excluded for failing these quality control filters and if they were carrying a large CNV (≥ 10Mb or chromosome anomaly). Also 187,468 individuals among final participants had Fluid intelligence measure (code: 20016). We used only biallelic and common markers for both array for our CNV calls (~750,000 markers).

CNV calling and quality controls

For all selected SNP array data, we called CNVs with PennCNV[39] and QuantiSNP[40] using previously published methods[7,10,41–43] (https://martineaujeanlouis.github.io/MIND-GENESPARALLELCNV/#/). CNVs were called using both algorithms with the following parameters: number of consecutive probes for CNV detection ≥3, CNV size ≥1Kb, confidence scores ≥15. CNVs detected by both algorithms were combined (CNVision[44]) to minimize the number of potential false discoveries. After this merging step, the CNV inheritance analysis algorithm (in-house algorithm), was applied to concatenate adjacent CNVs (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. CNV filtering steps were previously published. After

filtering array according to their quality, we applied filtering for CNVs. The CNVs with the following criteria were selected for analysis: confidence score ≥ 30 (with at least one of both detection algorithms), size ≥ 50 kb, unambiguous type (deletions or duplications). Every recurrent CNV were manually visualized by at least one individual.

Annotation of CNVs

We annotated the CNVs using Gencode V19 annotation (the reference release for hg19 Human genome release) with ENSEMBL gene name (https://grch37.ensembl.org/index.html). Briefly, we used bedtools suite (https://bedtools.readthedocs.io/en/latest/) to compare the different elements of the genes that overlap CNVs (UTRs, start and stop codons, exons and introns). Thus, for each CNV, we obtained its proportion of overlap with each gene as well as with each gene component (UTR, Coding exons, start and stop codons, introns).

Gene expression data

Allen Human Brain Atlas (AHBA) data: The AHBA gene data matrix [5] downloaded from (https://human.brain-map.org/static/download) was further processed using the abagen toolbox (version 0.1.3; https://github.com/rmarkello/abagen) [45] based on recommendations in [46,47]. The following text was generated directly from abagen, describing the exact methodologies employed during preprocessing:

First, microarray probes were reannotated using data provided by [46]; probes not matched to a valid Entrez ID were discarded. Next, probes were filtered based on their expression intensity relative to background noise [14], such that probes with intensity less than the background in >=50% of samples across donors were discarded. When multiple probes indexed the expression of the same gene, we selected and used the probe with the most consistent pattern of regional variation across donors (i.e., differential stability; [14]. The MNI coordinates of tissue samples

were updated to those generated via non-linear registration using the Advanced Normalization Tools (ANTs; https://github.com/chrisfilo/alleninf). Inter-subject variation was addressed by normalizing tissue sample expression values across genes using a robust sigmoid function[48].Normalized expression values were then rescaled to the unit interval. Finally, gene expression values were normalized across tissue samples using an identical procedure. Samples assigned to the same brain region were averaged separately for each donor and then across donors, yielding a regional expression matrix. After processing, a final 180 cortical regions x 15,633 gene matrix was used for subsequent analysis.
We restricted to genes that were detected in Brain (list from the Human Protein Atlas[15]), leading to a final matrix of 180 cortical regions x 13678 genes.

BrainSpan data: processed using ABData package in R [49]. Briefly, data was grouped into 5 stages, and mean expression for 11 cortical structures was computed.

Human Protein Atlas (HPA) data: processed using the pipeline proposed in the Human Protein Atlas [15]. We obtained 79 regions x 16,000 gene matrix of normalized expression for subsequent analysis.

Assigning genes to cortical brain regions

We assigned genes to cortical regions based on whether they were preferentially expressed (Z-score>0.5) in those regions (**Fig. 1c-e**). Hereinafter referred to as preferentially-expressed-genes (PExGs). Since PExGs were defined as Z-score>0.5, a gene was on average assigned to 52 out of 180 cortical regions.

Assigning genes to developmental epochs

To investigate if our findings were dependent on the specific development stage (age) of gene-expression data, we stratified the adult gene-expression data into three non-overlapping gene

sets with peak expression in different peak epochs. Briefly, genes were assigned to one of the three developmental epochs, based on BrainVar resource [26]: i) Epoch 1 ranges from 10 to 19 post-conception weeks (pcw); ii) Epoch 2: ranges between 19pcw and 6 postnatal months, and covers the late-fetal transition period, which has been hypothesised to be a critical window for many disorders [26]; and iii) Epoch 3: starts around 6 month and reaches all the way to adulthood.

Assigning genes to temporal trajectories

Genes were also stratified based on their temporal trajectories, based on BrainVar[26] resource into: i) falling genes; ii) non-transitional genes; and iii) rising genes.

Brain and other tissue specific gene sets

To investigate if our findings are driven by brain-elevated genes or can be reproduced irrespective of tissue specificity of gene expression, we used the tissue specific gene lists from the Human Protein Atlas (HPA) [15] version 22. The HPA performed a normalized transcriptome analysis of protein coding genes across all major organs and tissue types in the human body and compared mRNA levels. A gene is considered to have elevated expression based on either of the three conditions: i) Tissue enriched: At least four-fold higher mRNA level in a particular tissue compared to any other tissue.; ii) Group enriched: At least four-fold higher average mRNA level in a group of 2-5 tissues compared to any other tissue; and iii) Tissue enhanced: At least four-fold higher mRNA level in a particular tissue compared to the average level in all other tissues. We looked into genes that were brain-elevated (2685 genes), other-tissue elevated, and those that showed low-tissue specificity (https://www.proteinatlas.org/humanproteome/brain) [15]. Of note, we excluded the genes which were not detected in brain as established by the HPA.

Adult cell-type marker gene sets

Since bulk tissue expression data is an ensemble of underlying cell-constituents, we looked into deconvolving our findings using cell-type marker gene sets defined in the HPA for major brain cell-types [15]. The HPA used single cell RNA sequencing (scRNA-seq) data to cluster genes by their expression across major cell type. In this study, we focussed on 6 major cell types in Brain: Astrocytes, Oligodendrocyte precursor cells (OPCs), Oligodendrocytes, Microglial, Excitatory neurons, and Inhibitory neurons. A genome-wide classification of all the protein-coding genes with regard to specificity was performed as follows: (i) cell type enriched genes with at least fourfold higher expression levels in one cell type as compared with any other analyzed cell type; (ii) group enriched genes with enriched expression in a small number of cell types (2 to 10); and (iii) cell type enhanced genes with only moderately elevated expression.. We considered all genes that met any of these criteria for this study. The HPA confirmed the scRNA-seq profiles and cell type specificity using immunohistochemistry (IHC) on tissue microarrays.

Statistics:

Spin permutation testing:

Cortical maps of effect sizes were correlated by computing Pearson correlation. The statistical significance of such correlations was assessed using spin permutation [47,50]testing which accounts for spatial autocorrelation between brain regions. Since, the neighbouring brain regions are expected to be similar, this approach avoids inflated significance values and generates plausible null brain maps for null-distribution.

Shrinkage for correlation of deletion / duplication effect size maps:

We applied a shrinkage where cortical regions that were significant (multiple comparison correction) for either deletions or duplications were kept. This shrinkage ensured that cortical regions without significant estimates were not driving the spatial correlation.

Brain parcellations:

We used Glasser brain parcellation [13] for our main analysis. Glasser parcellation was obtained using multi-modal magnetic resonance images from the Human Connectome Project (HCP)[51], by delineating 180 areas per hemisphere bounded by sharp changes in cortical architecture, function, connectivity, and/or topography. In addition, we also investigated a functional cortical parcellation (Schaefer[52]) at different resolutions as well as the Desikan[53] parcellation.

Brain-maps using ggseg:

All brain-maps shown in this study were generated using ggseg package [54] in R.

Resting state contextualization:

We used BrainStat toolbox (https://brainstat.readthedocs.io/en/master/) for resting state contextualization. We mapped the effect sizes from the 180 cortical regions to 17 resting state brain networks from Yeo 17 parcellation [55] by first mapping the 180 effect sizes to fsaverage5 vertices, and then averaging the effect sizes for 17 brain networks.

Established Cortical organization hierarchies:

We used the previously published cortical organization hierarchies summarized in [2], covering reference maps for molecular, microstructural, electrophysiological, developmental and functional characterization of human cortex. Deletion and duplication effect size maps of cognitive ability were annotated using 20 neural maps [2,17], that included two microstructural

(cortical thickness, T1w/T2w), four metabolic (cerebral blood flow, cerebral blood volume, glucose metabolism, oxygen metabolism), three functional (functional gradient, PC1 NeuroSynth, intersubject functional variability), four expansion (evolutionary and developmental expansion, and allometric scaling NIH/PNC), six band-specific electrophysiological signal power (alpha, beta, delta, low gamma, high gamma, theta), and one spatial transcriptomic map (PC1 gene expression). To compute a spatial correlation between deletion / duplication effect sizes maps of cognitive ability and the 20 neural maps, we obtained the 20 neural maps in fsavaerge5 space from the neuromaps package and mapped them to 180 cortical regions of Glasser parcellation by averaging values across all the vertices within a parcel. Finally, a spin permutation[17] method was employed to assess the significance of the spatial correlation.

Enrichment in gene ontologies across cortical gene sets:

The enrichment analyses encompassed a total of 1399 GO terms[56] from three categories: Gene Ontology biological process (GO-BP) with 1057 terms, Gene Ontology cellular component (GO-CC) with 177 terms, and Gene Ontology molecular function (GO-MF) with 165 terms. We used GO-term gene-sets from the Molecular Signatures Database[56] (MSigDB, the msigdbr package in R, version 7.5.1), with gene-set size >50 and <700. The GO-term analyses encompassed three main steps. Firstly, gene ontology (GO) enrichments were computed for each of the 180 brain regions and their corresponding gene sets using the ranked gene set enrichment method (the fgsea[57] package in R). Secondly, enrichment scores were projected onto cortical maps to generate spatial profiles for each GO-term. Finally, spatial correlation analyses were conducted between the GO-term enrichment maps and the deletion and duplication effect size maps of cognitive ability. To assess the relationship between GO-term enrichment maps and the deletion effect size map, a spin permutation method was employed,

controlling for false discovery rate (FDR, q<0.05). We restricted the spatial correlation analyses to those GO-term enrichment maps with >5 cortical regions showing significant enrichments (FDR q<0.05, across 1399 GO-terms x 180 cortical regions). The polarity of deletion and duplication effect size maps were reversed before computing spatial correlation, so that largest effect sizes are positive and correlated with positive enrichment scores.

Author contributions

K.K., S.K., G.D., and S.J. designed the study, analyzed data, and drafted the manuscript. M.J.L., Z.S., and G.H. performed CNV calling and quality control. All authors contributed to result interpretation and in the editing of the manuscript.

Additional Information

Data and materials availability

UK Biobank data was downloaded under the application 40980, and may be accessed via their standard data access procedure (see http://www.ukbiobank.ac.uk/register-apply). UK Biobank CNVs were called using the pipeline developed in the Jacquemont Lab, as described at https://github.com/MartineauJeanLouis/MIND-GENESPARALLELCNV. The final CNV calls are available for download from the UK Biobank returned datasets (Return ID: 3104, https://biobank.ndph.ox.ac.uk/ukb/dset.cgi?id=3104). References to the processing pipeline and R package versions used for analysis are listed in methods.

Code availability

Code for generating all the Figures reported in main analysis, supplement material, can be found at: https://github.com/kkumar-iitkgp/Mirror_effect_del_dup_cortex

References

1. Bressler, S. L. & Menon, V. Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* **14**, 277–290 (2010).
2. Sydnor, V. J. *et al.* Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. *Neuron* Preprint at https://doi.org/10.1016/j.neuron.2021.06.016 (2021).
3. Menon, V. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn. Sci.* **15**, 483–506 (2011).
4. Hansen, J. Y. *et al.* Mapping gene transcription and neurocognition across human neocortex. *Nature Human Behaviour* **5**, 1240–1250 (2021).
5. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
6. Huguet, G. *et al.* Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability. *Mol. Psychiatry* (2021) doi:10.1038/s41380-020-00985-z.
7. Huguet, G. *et al.* Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA Psychiatry* **75**, 447–457 (2018).
8. Han, L. *et al.* Functional annotation of rare structural variation in the human brain. *Nat.*

Commun. **11**, 2990 (2020).

9. Jacquemont, S. *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* **478**, 97–102 (2011).
10. Alexander-Bloch, A. *et al.* Copy Number Variant Risk Scores Associated With Cognition, Psychopathology, and Brain Structure in Youths in the Philadelphia Neurodevelopmental Cohort. *JAMA Psychiatry* (2022) doi:10.1001/jamapsychiatry.2022.1017.
11. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
12. Vogel-Ciernia, A. & Wood, M. A. Neuron-specific chromatin remodeling: a missing link in epigenetic mechanisms underlying synaptic plasticity, memory, and intellectual disability disorders. *Neuropharmacology* **80**, 18–27 (2014).
13. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
14. Hawrylycz, M. *et al.* Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).
15. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, (2020).
16. Wagstyl, K. *et al.* Transcriptional Cartography Integrates Multiscale Biology of the Human Cortex. *bioRxiv* 2022.06.13.495984 (2022) doi:10.1101/2022.06.13.495984.
17. Markello, R. D. *et al.* neuromaps: structural and functional interpretation of brain maps. *Nat. Methods* **19**, 1472–1479 (2022).
18. Seguin, C., Razi, A. & Zalesky, A. Inferring neural signalling directionality from undirected structural connectomes. *Nat. Commun.* **10**, 4289 (2019).

19. Douard, E. *et al.* Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome. *Am. J. Psychiatry* **178**, 87–98 (2021).

20. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

21. Collins, R. L. *et al.* A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**, 3041–3055.e25 (2022).

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

23. Burt, J. B. *et al.* Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).

24. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014).

25. Pines, A. R. *et al.* Dissociable multi-scale patterns of development in personalized brain networks. *Nat. Commun.* **13**, 2647 (2022).

26. Werling, D. M. *et al.* Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic Coordination in the Developing Human Prefrontal Cortex. *Cell Rep.* **31**, 107489 (2020).

27. Karlsson, M. *et al.* A single-cell type transcriptomics map of human tissues. *Sci Adv* **7**, (2021).

28. Schork, A. J. *et al.* A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nat. Neurosci.* **22**, 353–361 (2019).

29. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgh.harvard.edu & Cross-Disorder Group of the Psychiatric Genomics

Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469–1482.e11 (2019).

30. Sey, N. Y. A. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.* **23**, 583–593 (2020).
31. Hu, B. & Won, H. Chromatin architecture provides a roadmap to improve our understanding of psychiatric disorders. *Neuropsychopharmacology* **46**, 234–235 (2021).
32. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568–584.e23 (2020).
33. Marek, S. *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nature* 1–7 (2022).
34. Kebets, V. *et al.* Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology. *Biol. Psychiatry* **86**, 779–791 (2019).
35. Bishop, S. L. *et al.* Identification of Developmental and Behavioral Markers Associated With Genetic Abnormalities in Autism Spectrum Disorder. *Am. J. Psychiatry* **174**, 576–585 (2017).
36. Cunningham, A. C. *et al.* Movement Disorder Phenotypes in Children With 22q11.2 Deletion Syndrome. *Mov. Disord.* **35**, 1272–1274 (2020).
37. Auwerx, C. *et al.* The individual and global impact of copy-number variants on complex human traits. *Am. J. Hum. Genet.* **109**, 647–668 (2022).
38. Kopal, J. *et al.* Rare CNVs and phenome-wide profiling highlight brain structural divergence and phenotypical convergence. *Nat Hum Behav* **7**, 1001–1017 (2023).
39. Wang, K. *et al.* 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 (2007).

40. Colella, S. *et al.* 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 (2007).
41. Mollon, J. *et al.* Impact of Copy Number Variants and Polygenic Risk Scores on Psychopathology in the UK Biobank. *Biol. Psychiatry* (2023) doi:10.1016/j.biopsych.2023.01.028.
42. Huguet, G. *et al.* Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes. *bioRxiv* 2024.04.16.589618 (2024) doi:10.1101/2024.04.16.589618.
43. Wainberg, M. *et al.* Deletion of loss-of-function-intolerant genes and risk of 5 psychiatric disorders. *JAMA Psychiatry* **79**, 78–81 (2022).
44. Sanders, S. J. *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).
45. Markello, R. D. *et al.* Standardizing workflows in imaging transcriptomics with the abagen toolbox. *Elife* **10**, (2021).
46. Arnatkeviciute, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353–367 (2019).
47. Arnatkeviciute, A., Markello, R. D., Fulcher, B. D., Misic, B. & Fornito, A. Towards Best Practices for Imaging Transcriptomics of the Human Brain. *Biol. Psychiatry* (2022) doi:10.1016/j.biopsych.2022.10.016.
48. Fulcher, B. D., Little, M. A. & Jones, N. S. Highly comparative time-series analysis: the empirical structure of time series and their methods. *J. R. Soc. Interface* **10**, 20130048 (2013).

49. Grote, S., Prüfer, K., Kelso, J. & Dannemann, M. ABAEnrichment: an R package to test for gene set expression enrichment in the adult and developing human brain. *Bioinformatics* **32**, 3201–3203 (2016).

50. Markello, R. D. & Misic, B. Comparing spatial null models for brain maps. *Neuroimage* **236**, 118052 (2021).

51. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).

52. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).

53. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).

54. Mowinckel, A. M. & Vidal-Piñeiro, D. Visualization of brain statistics with R packages ggseg and ggseg3d. *Adv. Methods Pract. Psychol. Sci.* **3**, 466–483 (2020).

55. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

56. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).

57. Sergushichev, A., Korotkevich, G., Sukhov, V. & Artyomov, M. N. Fast gene set enrichment analysis. *bioRxiv* (2019).

Figures

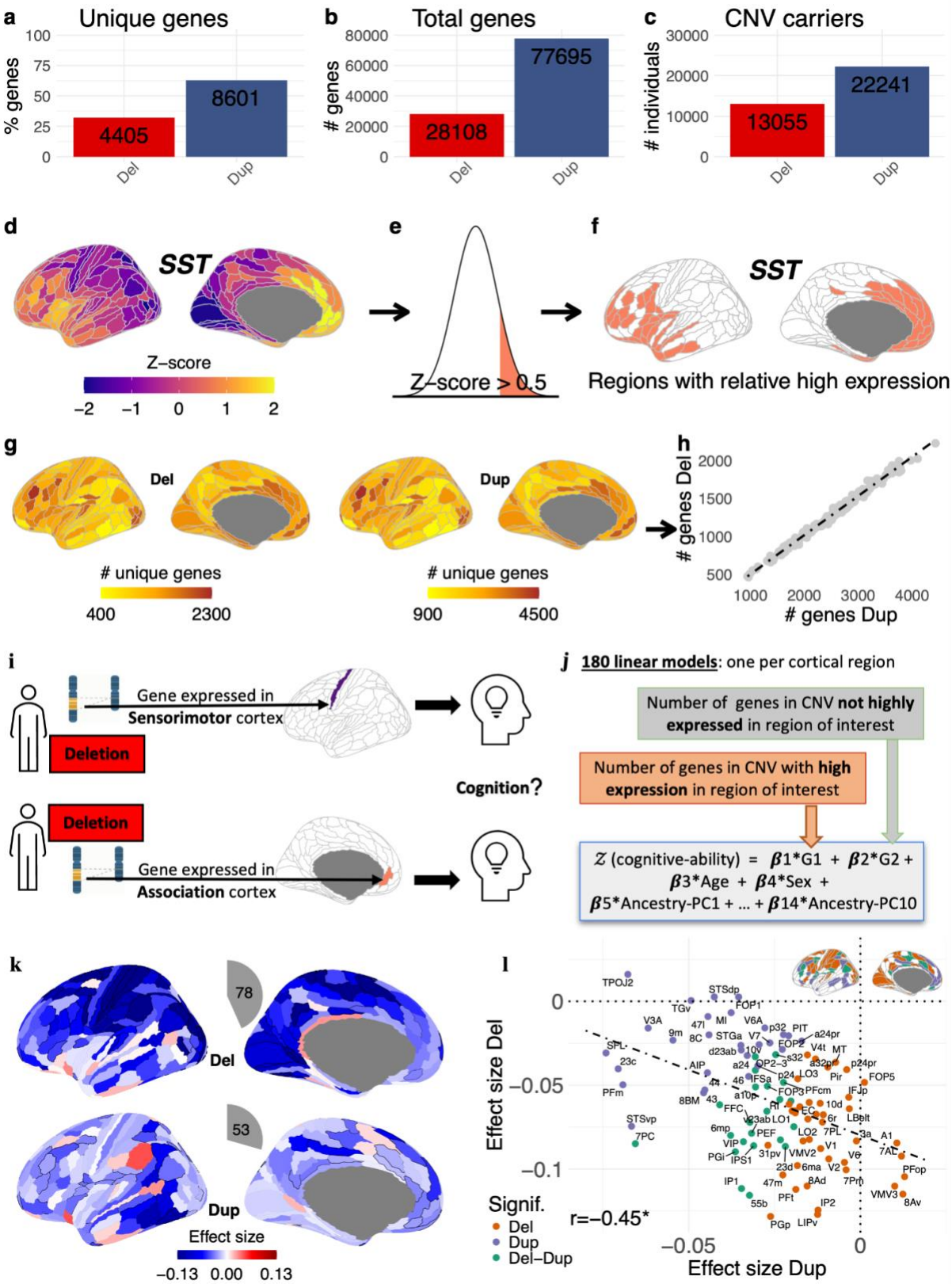

Fig. 1: Cortical maps of deletion and duplication effect sizes on cognitive ability.

a-c, Barplot showing the number (#) of unique genes, total genes (accounting for recurrence), and CNV carriers observed in deletions and duplications in our assembled data, covering

32.2% and 62.9% of 13678 genes with spatial expression data (AHBA) and have been detected in the brain (HPA). **d-f**, Method to identify genes with relative high expression across cortical regionsMethod to establish relative high expression. As an example, the expression of **d**, SST is z-scored across 180 cortical regions. We then **e**, thresholded the Z-score to **f**, identify the cortical regions with relative high expression of SST. This procedure was applied across all 15663 genes, resulting in 180 lists of relative high expression genes (RHExG) corresponding to each cortical region of the Glasser parcellation. **g**, The number of unique RHExGs per cortical region for deletions and duplications. Separate color bars highlight the difference in the range of values between deletions and duplications for both unique and total genes per cortical region. **h**, Scatterplot showing a high degree of similarity in the distribution of number of unique genes per cortical region between deletions and duplications. Each point represents a cortical region, with unique genes per cortical region for deletions and duplications on Y-axis and X-axis respectively. **i-j**, Summary of burden association analysis ran to estimate mean effect size of genes on cognitive ability for each of the 180 cortical regions. **j**, linear model for estimating the mean effect of genes with relative high expression in a cortical region, adjusting for genes within CNV with not-high expression in that cortical region. **h**, Effect sizes of genes in 180 cortical regions. We performed a burden analysis for each cortical region (n=180 linear models). The average effect size of genes with high expression in each region is color-coded according to the color scale (Z-scored cognitive ability) and mapped onto the cortex for deletions and duplications. 128 and 112 regions with black borders show significant effects for deletions and duplications, respectively (Bonferroni $p<0.05$). Grey pie charts show the proportion of regions with significant effects. **i**, Negative correlation between the effect sizes of deletions and duplications across the cortex (spin permutation p-value =2e-4). Each point is the mean effect size of genes in each of the cortical regions. X-axis: duplication; Y-axis: deletion effect sizes

respectively. Only significant effects are plotted, and corresponding significant regions are shown as a brain-plot in the top right corner.

Del: deletion; Dup: duplication; Signif.: significance; r: Pearson correlation; p: spin-permutation p-value.

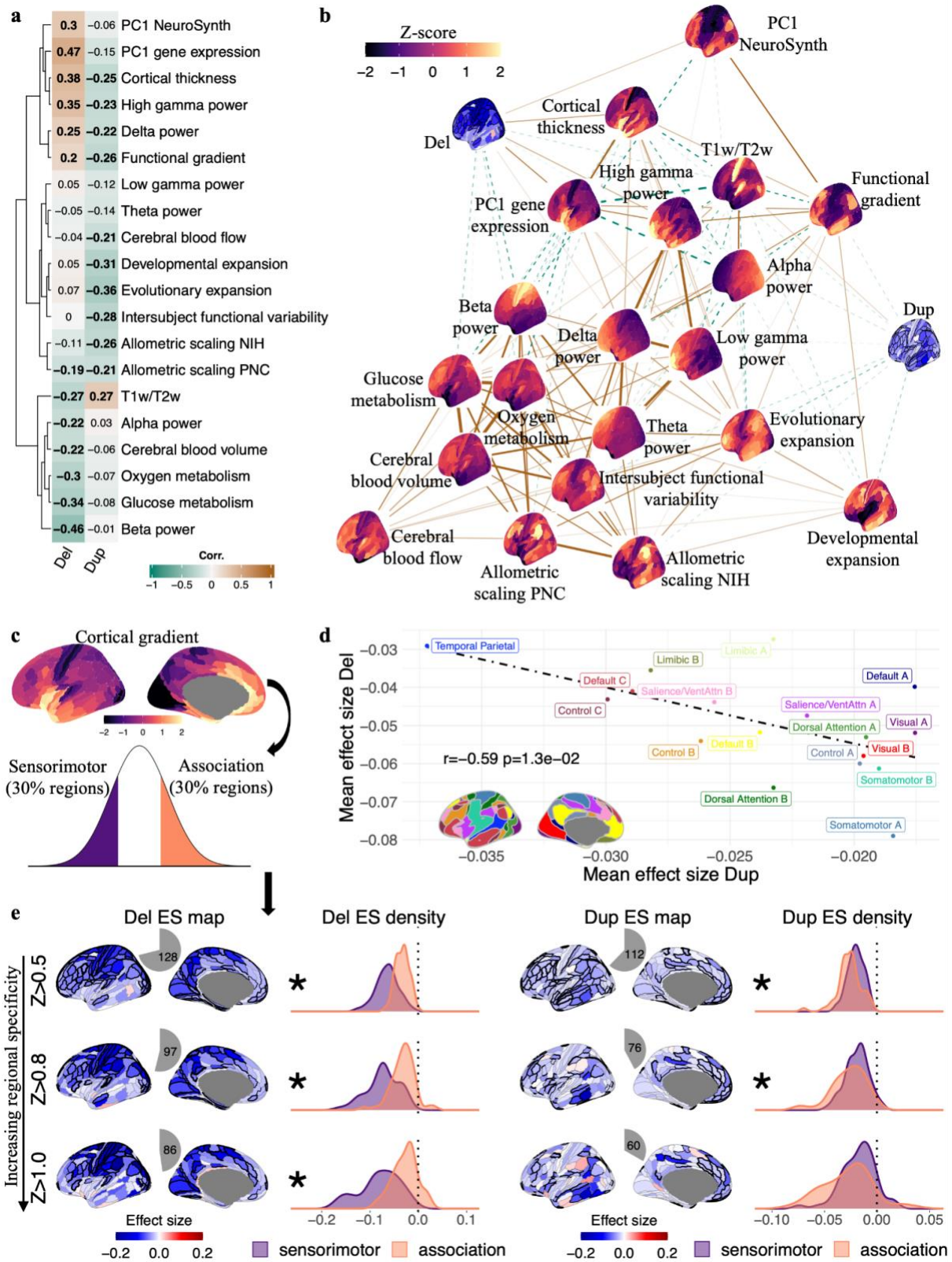

Fig. 2: Annotating deletion and duplication effect size maps.

a-b, Annotation of deletion and duplication effect size maps using 20 neural maps [2,17], including two microstructural (cortical thickness, T1w/T2w), four metabolic (cerebral blood

flow, cerebral blood volume, glucose metabolism, oxygen metabolism), three functional (functional gradient, PC1 NeuroSynth, intersubject functional variability), four expansion (evolutionay and developmental expansion, and allometric scaling NIH/PNC), six band-specific electrophysiological signal power (alpha, beta, delta, low gamma, high gamma, theta), and one spatial transcriptomic map (PC1 gene expression) **a,** Matrix shows the spatial correlation of CNV effect size maps (on cognitive ability) with 20 neuro-maps. The maps were obtained from neuromaps toolbox[17] and transformed to Glasser parcellation. Bold indicate significant spatial correlations (multiple comparison corrected (FDR) p-value < 0.05). **b**, Spring embedding representation of the correlation matrix, showing FDR significant edges. The deletion and duplication maps, and the 20 neuro-maps are shown as nodes. Positive and negative correlations are represented as solid and dotted lines. **c**, Dissociation of effect size across sensorimotor (unimodal) and association (transmodal) regions. Cortical regions were ranked based on previously established sensorimotor - association axis [2,18]. We examined the effects sizes of the top and bottom 30% of these regions. The sensorimotor and association cortical regions are colored in dark purple and light orange respectively. **d**, To characterize the spatial patterns of effect sizes of deletions and duplications on cognitive ability (Fig. 1), we used resting state contextualization to map effect sizes to 17 brain networks Yeo et al. (2011). Specifically, we mapped 180 cortical regions to the 17 brain networks and computed mean effect sizes for each of the 17 brain networks. Negative correlation between the mean effect sizes of deletions and duplications on cognitive ability across the 17 brain networks. The largest effect sizes for deletions and duplications are observed for sensorimotor and temporal parietal networks, respectively. The brain-map in bottom left outlines the 17 brain networks. We used open-source BrainStat toolbox for resting state contextualization. **e,** Effect sizes maps of cognitive ability and density plots comparing effect sizes of sensorimotor and association regions for deletions (left) and duplications (right) for 3 levels of regional specificity.

Increasing Z-score threshold (regional specificity parameter) leads to smaller regional gene-sets, which include genes showing higher regional specificity. * : represents significant difference between the sensorimotor and association effect sizes; Kolmogorov–Smirnov test p-value < 0.05.

Del: deletion; Dup: duplication; Signif.: significance; r: Pearson correlation; p: spin-permutation p-value.

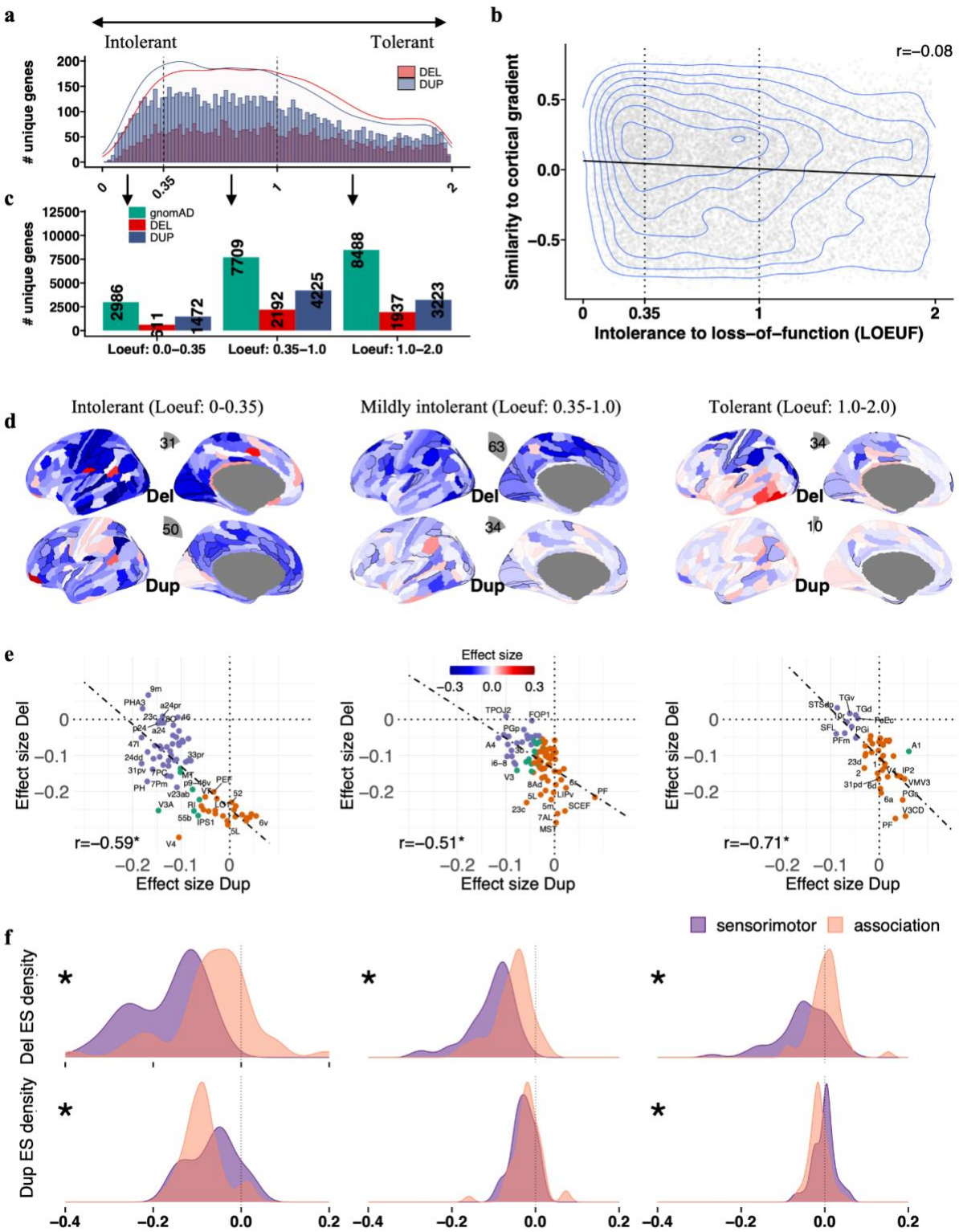

Fig. 3. Intolerance to loss of function and cortical maps of effect sizes.

a, Histogram showing the distribution of the number of genes deleted or duplicated at least once in our dataset, with respect to their intolerance to loss of function (LOEUF values). Deletions and duplications are in red and blue. X-axis : 100 bins of LOEUF values. Density

curves are added to compare the densities between deletion and duplications. **b,** Correlation between values of intolerance to Loss of Function (X-axis :LOEUF) and a measure of spatial expression (similarity of gene expression to cortical gradiant). Each of the dots represents a gene expressed in the cortex. The contours are 2 dimension density plots. **c,** genes stratified into three LOEUF categories: i) highly intolerant (Loeuf1: 0-0.35); ii) intolerant (Loeuf2: 0.35-1.0); and iii) tolerant (Loeuf3: 1.0-2.0). Barplot showing the number (#) of unique genes observed in deletions and duplications in our assembled data, for the three LOEUF categories. Vertical lines are added to highlight these categories.

d, Cognitive ability effect size maps of deletions and duplications for the three LOEUF categories. Regions with significance effect sizes have black edges (Bonferroni p-value < **0.05**). Grey edges are not-significant. Pie charts show the number and proportion of significant cortical regions. Color scale shows the mean effect size of a gene in that cortical region of interest. **e**, Negative correlation between the effect sizes of deletions and duplications across the cortex (spin permutation p-value < 2e-4). Each point is the mean effect size of genes in each of the cortical regions. **f**, Density plots comparing effect sizes of sensorimotor and association regions for deletions (top) and duplications (bottom) for the three intolerance to Loss of Function categories. * : represents significant difference between the sensorimotor and association effect sizes; Kolmogorov–Smirnov test p-value < 0.05.

DEL: deletions; DUP: duplications; ES: effect size; r: pearson correlation; Sig: significant.

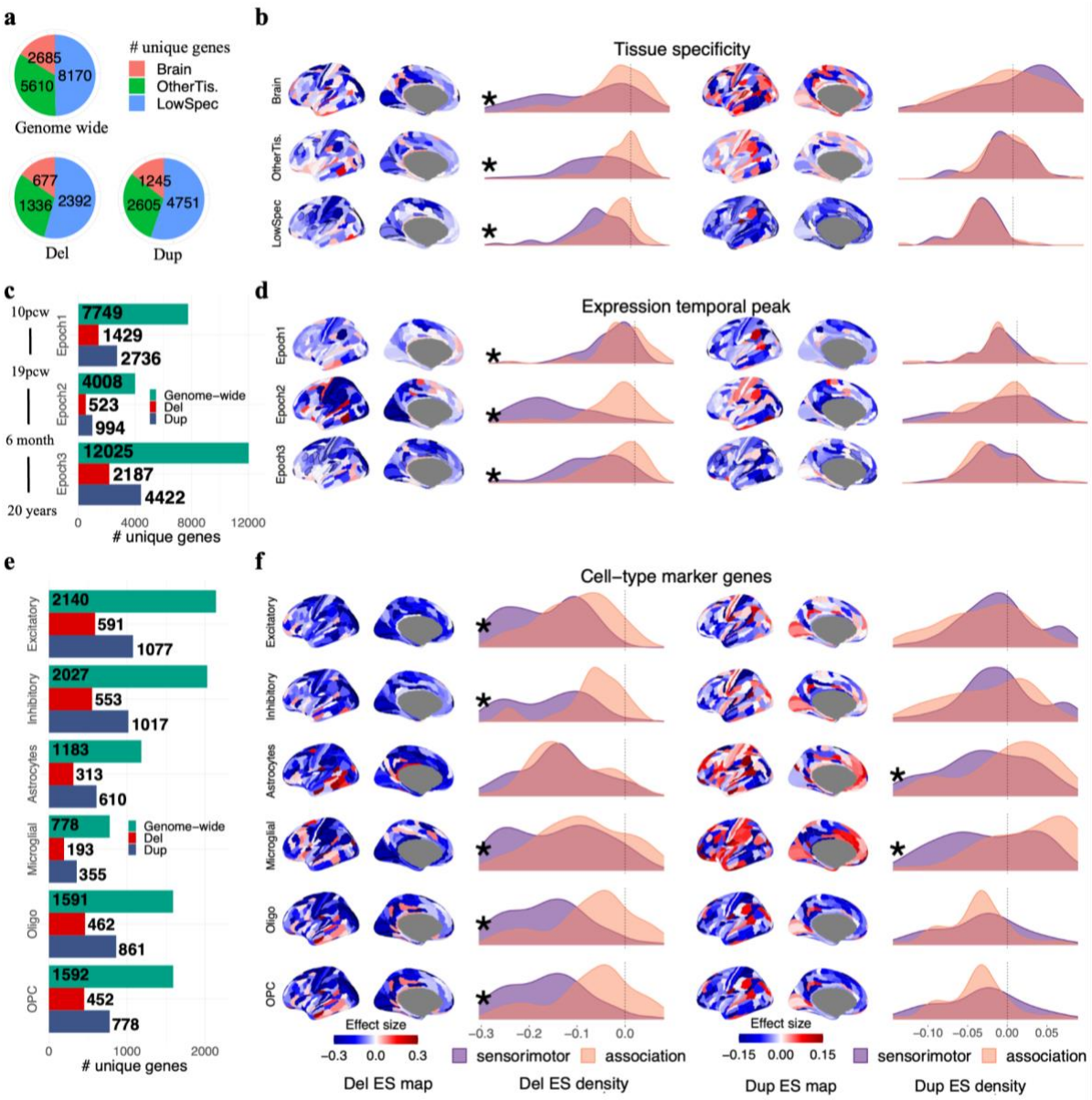

Fig. 4. Characterizing the impact of tissue-specificity, temporal dynamics of gene expression, and cell-types.

a, Pie charts showing the distribution of genes with elevated expresson in the brain (red), elevated expresson in non-brain tissues (green) and low tissues specificity (blue). Number (#) of unique genes observed in deletions and duplications in our assembled data, as well as genome-wide (reported in the Human Protein Atlas). **b**, Effect sizes maps of cognitive ability and density plots comparing effect sizes of sensorimotor and association regions for deletions (left) and duplications (right) for the three non-ovelapping categories of tissue specificity.

c, Barplot showing the number of genes with peak expression during epoch1 (10pcw to 19pcw), epoch2 (19pcw to 6 months after birth) and epoch3 (6 month onwards). Number (#) of unique genes observed in deletions (red) and duplications (blue) in our assembled data, as well as genome-wide (green, as reported in the BrainVar resource). **d**, Effect sizes maps of cognitive ability and density plots comparing effect sizes of sensorimotor and association regions for deletions (left) and duplications (right) for the three non-ovelapping categories of expression temporal peaks.

e, Barplot showing the number (#) of cell-type marker genes, across 6 categories, observed in deletions (red) and duplications (blue) in our assembled data, as well as genome-wide (green, reported in the Human Protein Atlas, HPA). **f**, Effect sizes maps of cognitive ability and density plots comparing effect sizes of sensorimotor and association regions for deletions (left) and duplications (right) for the six ovelapping cell-type marker categories.

* : represents significant difference between the sensorimotor and association effect sizes; Kolmogorov–Smirnov test p-value < 0.05. Del: deletions; Dup: duplications; ES: effect size; HPA: Human Protein Atlas; r: pearson correlation; Sig: significant.