

# UCLA

## UCLA Previously Published Works

### Title

Copy number variants and the tangential expansion of the cerebral cortex.

### Permalink

<https://escholarship.org/uc/item/4b45s6m7>

### Journal

Nature Communications, 16(1)

### Authors

Liao, Zhijie

Kumar, Kuldeep

Kopal, Jakub

et al.

### Publication Date

2025-02-17

### DOI

10.1038/s41467-025-56855-1

Peer reviewed













# Copy number variants and the tangential expansion of the cerebral cortex

Received: 29 March 2024

Accepted: 3 February 2025

Published online: 17 February 2025

 Check for updates


Zhijie Liao <sup>1,2</sup>, Kuldeep Kumar <sup>1</sup>, Jakub Kopal<sup>3,4</sup>, Guillaume Huguet <sup>1</sup>, Zohra Saci <sup>1</sup>, Martineau Jean-Louis <sup>1</sup>, Zdenka Pausova <sup>1,5,6</sup>, Igor Jurisica <sup>7,8,9</sup>, Carrie E. Bearden <sup>10</sup>, the IMAGEN Consortium\*, 16p11.2 European Consortium\*, Sebastien Jacquemont <sup>1,6,12</sup>  & Tomas Paus <sup>1,2,11,12</sup> 

The tangential expansion of the human cerebral cortex, indexed by its surface area (SA), occurs mainly during prenatal and early postnatal periods, and is influenced by genetic factors. Here we investigate the role of rare copy number variants (CNVs) in shaping SA, and the underlying mechanisms, by aggregating CNVs across the genome in community-based cohorts ( $N = 39,015$ ). We reveal that genome-wide CNV deletions and duplications are associated with smaller SA. Subsequent analyses with gene expression in fetal cortex suggest that CNVs influence SA by interrupting the proliferation of neural progenitor cells during fetal development. Notably, the deletion of genes with strong (but not weak) coexpression with neural progenitor genes is associated with smaller SA. Follow up analyses reveal similar mechanisms at play in three clinical CNVs, 1q21.1, 16p11.2 and 22q11.2. Together, this study of rare CNVs expands our knowledge about genetic architecture of human cerebral cortex.

The human cerebral cortex, essential for multiple perceptual, motor, and cognitive processes, is a highly folded sheath of tissue with about 1800 cm<sup>2</sup> of surface area<sup>1</sup>. Inter-individual variations in the cortical surface area (SA) contribute to the variations in cognitive abilities and vulnerability to mental illness<sup>2–4</sup>. The cerebral cortex expands dramatically during prenatal periods and the first two years of life, with subsequent age-related increases in SA being limited<sup>1,5–7</sup>. Thus, individual variations in SA observed in adulthood reflect mainly developmental events occurring during the prenatal and early postnatal periods. The cortical expansion during the prenatal period has been of great interest for decades. During fetal development, radial glia can self-renew and generate intermediate progenitor cells (IPCs), which

can self-amplify prior to generating neurons<sup>8–11</sup>. Based on a large body of work carried out in non-human primates, the radial unit hypothesis proposed that the proliferation of neural progenitor cells plays a central role in the tangential growth of cerebral cortex<sup>12</sup>. Increasing the proliferation of progenitor cells experimentally in the mouse embryo results in more neurons and a larger SA<sup>13</sup>.

Tangential expansion of the cerebral cortex is largely influenced by genetic variations, with about 90% heritability of SA as estimated in twin studies<sup>14</sup>. Genome-wide association studies (GWASs) have identified hundreds of loci related to SA of the adult cerebral cortex. These common genetic variants are enriched in regulatory activities in neural progenitor cells, which is consistent with the radial unit hypothesis<sup>15–17</sup>.

<sup>1</sup>Centre de Recherche du CHU Sainte-Justine, Montreal, QC, Canada. <sup>2</sup>Departments of Psychiatry and Addictology, University of Montreal, Montreal, QC, Canada. <sup>3</sup>Department of Biomedical Engineering, Faculty of Medicine, McGill University, Montreal, QC, Canada. <sup>4</sup>Mila - Quebec Artificial Intelligence Institute, Montréal, QC, Canada. <sup>5</sup>The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada. <sup>6</sup>Department of Pediatrics, University of Montreal, Montreal, QC, Canada. <sup>7</sup>Osteoarthritis Research Program, Division of Orthopedic Surgery, Schroeder Arthritis Institute, and Data Science Discovery Centre for Chronic Diseases, Krembil Research Institute, University Health Network, Toronto, ON, Canada. <sup>8</sup>Departments of Medical Biophysics and Computer Science, and the Faculty of Dentistry, University of Toronto, Toronto, ON, Canada. <sup>9</sup>Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovakia. <sup>10</sup>Semel Institute for Neuroscience and Human Behavior, Departments of Psychiatry and Biobehavioral Sciences and Psychology, University of California-Los Angeles, Los Angeles, CA, USA. <sup>11</sup>Department of Neuroscience, University of Montreal, Montreal, QC, Canada. <sup>12</sup>These authors jointly supervised this work: Sebastien Jacquemont, Tomas Paus. \*Lists of authors and their affiliations appear at the end of the paper.  e-mail: [sebastien.jacquemont@umontreal.ca](mailto:sebastien.jacquemont@umontreal.ca); [tpausresearch@gmail.com](mailto:tpausresearch@gmail.com)

Yet, most of these genetic variants reside in non-coding regions, posing a challenge regarding whether and how these variants influence gene expression and biological processes. In contrast, rare copy number variants (CNVs) with protein-coding gene(s) fully deleted or duplicated lead to substantial decrease and increase, respectively, in gene expression in the brain<sup>18</sup>. Thus, investigating the effects of CNVs on SA can provide valuable insights into the intricate relationship between gene expression and cortical expansion. Previous research on the association between CNVs and SA has focused on a few specific genomic loci, such as 1q21.1, 16p11.2, and 22q11.2, which confer high risk for psychiatric disorders such as autism and schizophrenia<sup>19–23</sup>. These studies have reported large effects of these CNVs on SA but identifying the contributing genes within these multigenic variants remains challenging. Beyond these specific CNVs, the majority of CNVs are very rare (<1/5000), and, as such, their individual effects on the brain are difficult to assess with currently available sample sizes. A common strategy to overcome this limitation has been to aggregate rare CNVs across the genome based on gene content and characteristics (e.g., intolerance to loss-of-function) to test association with phenotypes such as cognitive abilities and autism<sup>24–27</sup>.

In this work, we investigated the relationship between CNVs and SA by pooling data from three community-based cohorts (IMAGEN, Saguenay Youth Study, and the UK biobank). By aggregating - across the genome - all CNVs ( $\geq 50$  kilobases) based on genes' intolerance to loss-of-function, we find that genome-wide deletions and duplications are associated with smaller SA. By utilizing virtual ontogeny and expression Quantitative Trait Loci (eQTLs) from the fetal cortex, we then test the explanatory power of the radial unit hypothesis vis-à-vis the development of the human cerebral cortex. Lastly, we apply similar approaches to gain insights into the developmental mechanisms underlying the effects of three specific CNVs related to neurodevelopmental disorders, namely 1q21.1, 16p11.2, and 22q11.2, on cortical expansion. Our findings suggest that CNVs influence SA by interrupting the expression of genes involved in neural proliferation in the fetal period.

## Results

### CNVs and surface area

Across the genome, all CNVs that were  $\geq 50$  kilobases and fully encompassed at least one protein-coding gene were included in this study ("Methods"). To estimate the association between genome-wide CNVs and cortical SA, we conducted mega-analyses using data from three community-based cohorts: the IMAGEN study, Saguenay Youth Study (SYS), and the UK Biobank (UKBB), including a total of 39,015 individuals with both magnetic resonance imaging (MRI) of their brains and genome-wide assessment of CNVs (Fig. 1A). In this dataset, we identified 2767 unique protein-coding CNVs, of which 93% were present in fewer than five individuals (Supplementary Data 1). Out of the 39,015 individuals, 2733 (1423 females) and 3859 (2011 females) individuals carried, respectively, at least one deletion or one duplication fully encompassing one or more protein-coding genes. Thus, there were 6309 CNV carriers and 32,706 CNV non-carriers. Across the genome, 5822 genes were either deleted (2084 genes) and/or duplicated (4836 genes) at least once (Fig. 1A and Supplementary Data 2).

Given the rarity of most CNVs with protein-coding genes, testing their individual effects on SA was not feasible with the currently available sample size. Instead, recent studies have aggregated rare CNVs based on general characteristics of involved genes, such as intolerance to loss-of-function, to explore the association of CNV burden with cognitive abilities and autism risk<sup>24–27</sup>. Here, we used a well-established measure of genes' intolerance to loss-of-function, namely, the inverse loss-of-function observed/expected upper-bound fraction (1/LOEUF), to estimate the total CNV burden for each individual<sup>28</sup>. Thus, a higher burden score indicates more deletions or duplications of "intolerant" genes in an individual (Supplementary

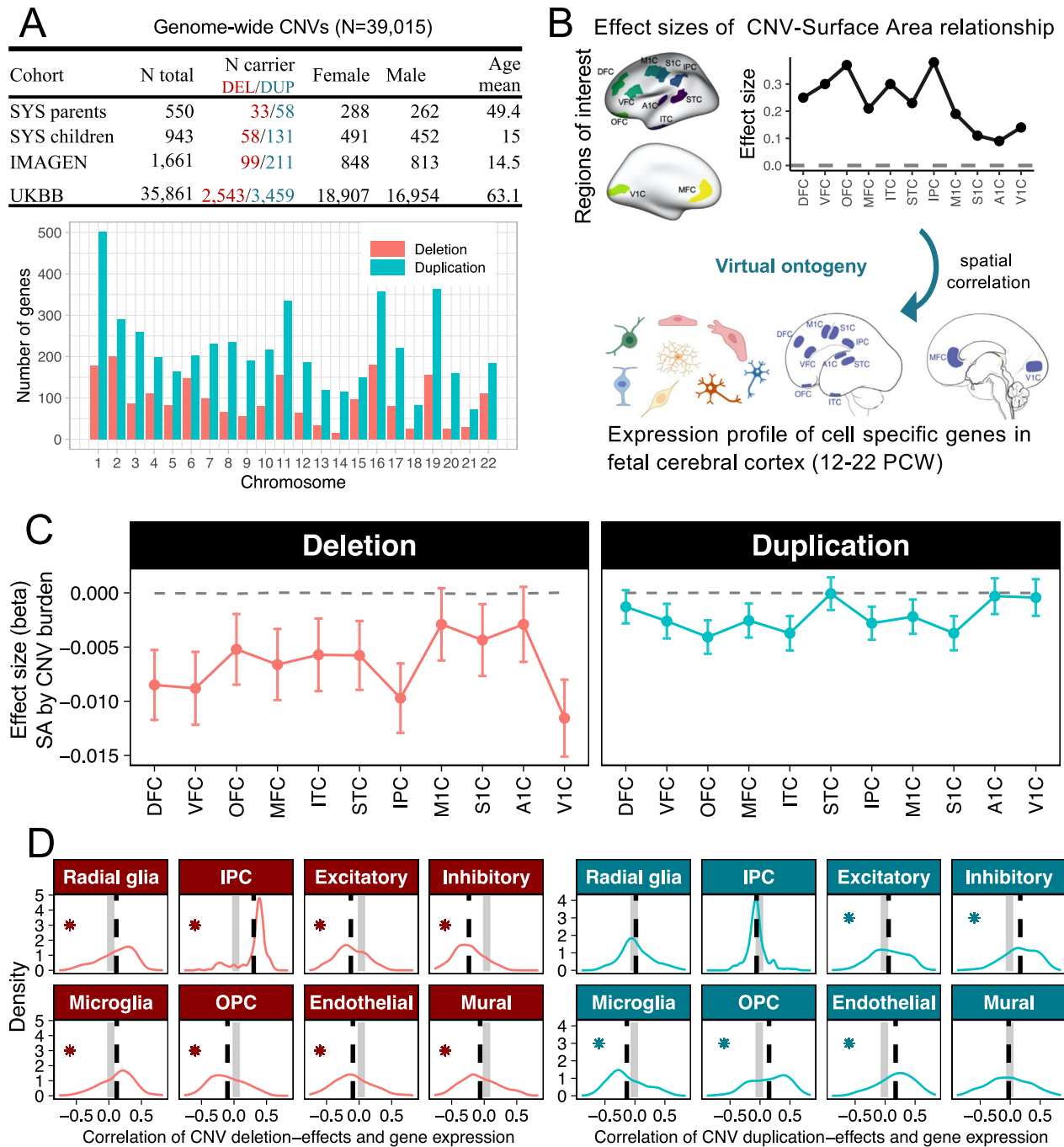
Fig. 1). We then quantified the effect sizes of the CNV burden on normalized SA (Z-scored across participants) of 11 cortical regions using linear mixed-effect models. These 11 regions of interest were selected based on the availability of gene-expression data obtained during the prenatal period, which were used in subsequent analysis (Supplementary Fig. 2). In addition, for each CNV with 20 or more carriers (35 CNVs), we provide the effect sizes on SA in Supplementary Fig. 3. We note that these recurrent CNVs have small effects on phenotypes, likely because they do not include genes that are intolerant to loss-of-function<sup>25,26,29</sup>. Estimating such small effects reliably would require substantially larger sample sizes.

We found that both CNV deletions and duplications were associated with a smaller SA across the 11 regions (Fig. 1C). On average, across the 11 regions, a one-point increase in the burden score for deletions and duplications was associated, respectively, with SA decreases of 0.0065 SD ( $p=1.75E-05$ ) and 0.0022 SD ( $p=5.81E-04$ ; Supplementary Data 3). Similar effect sizes were observed when utilizing the total number of genes deleted or duplicated in an individual as a simple measure of the CNV burden (Supplementary Fig. 4). The greater detrimental effects of deletions compared with duplications on SA aligns with previous research on cognitive abilities, where the effect of deletions is three times that of duplications<sup>24</sup>. Such difference may be attributed to larger changes in gene expression for CNV deletions compared with duplications<sup>18</sup>.

### Virtual ontogeny for CNV-SA relationships

To investigate the neurodevelopmental mechanisms underlying the CNV-related variations in SA, we used a previously established method named "virtual ontogeny" (Fig. 1B). This method provides a cellular-level interpretation of the MRI-based phenotype by testing spatial correlations between inter-regional profiles of CNV effects on SA and cell-specific gene expressions in mid-gestation across the 11 cortical regions<sup>3</sup>. More generally, similar approaches that test spatial correlations between MRI matrices and molecular maps across the cerebral cortex are used increasingly to explore biological underpinnings of various MRI-based findings<sup>30</sup>. In virtual ontogeny, based on a single-cell RNA-sequencing study<sup>31</sup>, we selected 200 gene markers for each of 8 cell types present in the fetal cerebral cortex (6–22 gestational weeks): radial glia, intermediate progenitor cells (IPC), microglia, oligodendrocyte progenitor cells (OPC), excitatory and inhibitory neurons, mural cells and endothelial cells. The expression of these cell-specific genes in each of the 11 regions during mid-gestation (12–22 postconceptional weeks) was obtained from the PsychENCODE bulk RNA-sequencing dataset<sup>32</sup>. We then tested the spatial correlation between interregional profiles of CNV-SA effect size and the expression profile of genes for each of the eight cell types. To simplify the interpretation of virtual ontogeny, we used absolute values of the effect sizes such that higher values indicate larger CNV-related variations in SA.

The virtual ontogeny revealed that cortical regions with higher expression of genes specific to radial glia, IPC, and microglia during mid-gestation, including OFC, MFC, and VIC, showed stronger effects of genomic deletions on SA. On the other hand, regions with higher expression of genes specific to the majority of neuronal and glial cell types showed weaker effects of CNV deletions (Fig. 1D, Supplementary Fig. 5, and Supplementary Data 4). For duplications, the CNV-SA effect size profile showed no spatial correlations with progenitor cells (radial glia, IPCs), and opposite correlations with the differentiated cell types, as compared with deletions. These results were validated by four sensitivity analyses: (1) 100 and 300 gene markers specific for each of the eight cell types, (2) bootstrapped estimation for the distribution of the mean correlation, (3) gene-set enrichment analysis (Supplementary Fig. 6 and Supplementary Datas 4–6), and (4) gene markers specific to 16 cell types derived from another single-cell transcriptomic dataset for mid-gestational cerebral cortex<sup>33</sup> (Supplementary Fig. 7)."



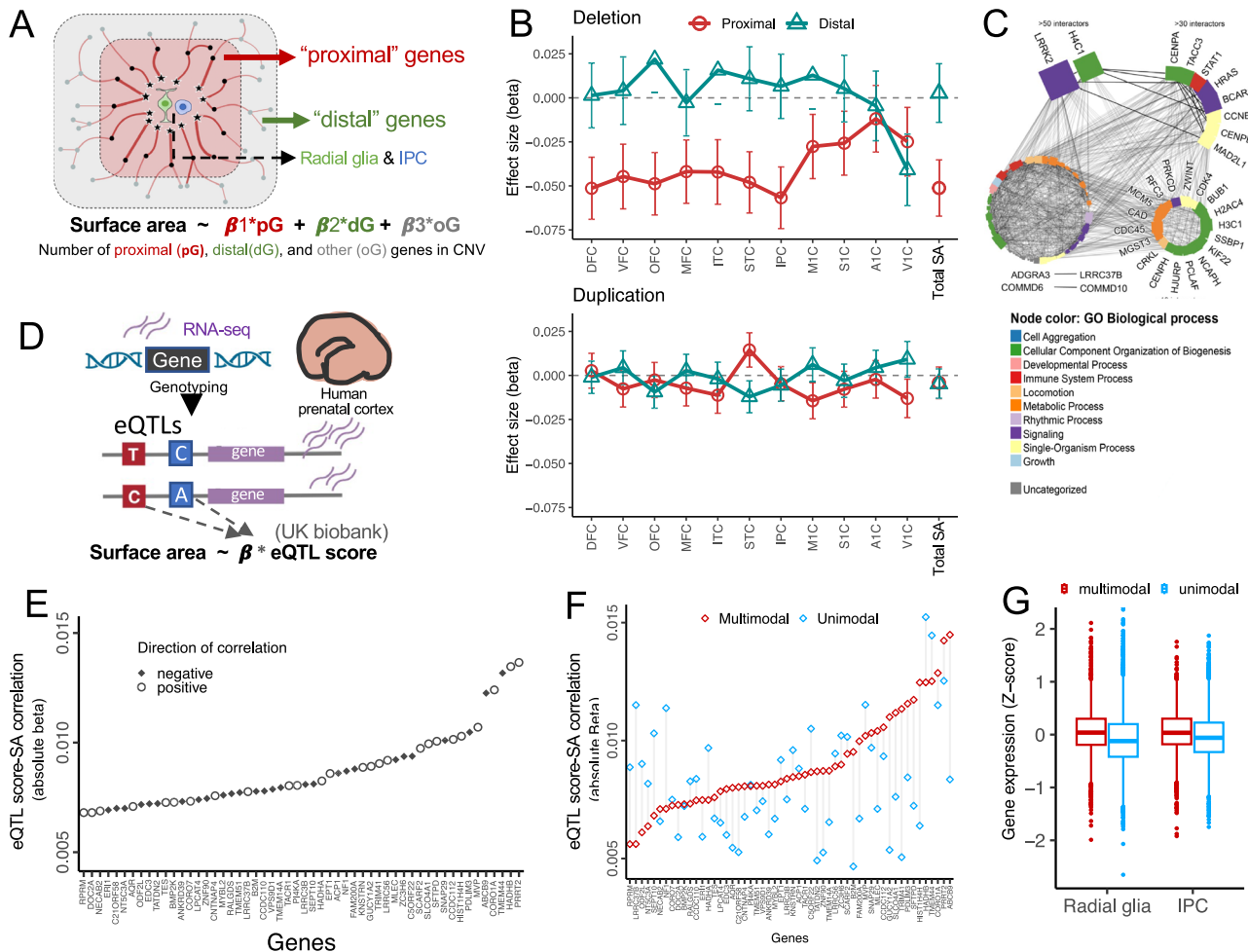
**Fig. 1 | Genome-wide CNVs and cortical expansion.** **A** Demographic of each cohort and the number of unique protein-coding genes being deleted and duplicated in the pooled dataset. **B** Workflow of the virtual ontogeny. We first estimated the effect sizes of genome-wide CNVs on cortical SA of 11 regions using data from community-based cohorts. Subsequently, “virtual ontogeny” analyses were conducted to investigate the underlying neurodevelopmental mechanisms by testing the relationships between the CNV-SA effect size profiles and the expression of cell-specific genes in the mid-gestational fetal cortex. **C** Estimated effect sizes of the burden of CNVs deletion/duplication on normalized SA (Z-score) in community-based cohorts ( $N = 39,015$ ). We used the sum of 1/LOEUF scores of genes within CNVs for each individual to quantify the burden of CNVs. The centered dots

represent the estimated effect size for each region. The error bars represent the standard error of the estimated effect size for each region. **D** Results of the virtual ontogeny. The dashed vertical line represents the mean biweight midcorrelation within each cell type. The gray box around zero represents 95% confidence intervals of the null distribution, which was derived by a permutation approach. The asterisk denotes  $p < 0.05$  (FDR). DFC, dorsal frontal cortex; VFC, ventral frontal cortex; OFC, orbitofrontal cortex; MFC, medial frontal cortex; ITC, inferior temporal cortex; STC, superior temporal cortex; IPC, inferior parietal cortex; MIC, primary motor cortex; S1C, primary somatosensory cortex; A1C, primary auditory cortex; V1C, primary visual cortex. SYS Saguenay Youth Study, UKBB United Kingdom BioBank.

**Proximal and distal genes in network of progenitor genes**

To investigate which of the deleted/duplicated genes contribute to the association between genome-wide CNV burden and SA, we were inspired by the omnigenic model, which proposes that a large number

of genes expressed in trait-relevant tissues contribute to the heritability of the trait through regulatory networks of core genes<sup>34</sup>. Accordingly, we hypothesized that genes that are highly co-expressed with neural progenitor genes (core) in the gene network have a



**Fig. 2 | Effects of genes on cortical surface area in relation to coexpression with neural progenitor genes.** **A** Illustration of defining proximal and distal genes. The proximal/distal categorization was based on coexpression levels with neural progenitor genes across the 11 cortical regions. “Proximal” denotes genes with high coexpression, while “distal” refers to genes with low coexpression. **B** The effect of proximal/distal gene deletions on SA ( $N = 39,015$ ). Deletions of proximal genes, but not distal genes, correlated with smaller SA. The centered symbols represent the estimated effect size for each region. Error bars represent the standard error of the estimated effect sizes for each region. **C** Physical protein-protein interactions among the proximal genes. Node size corresponds to the number of interacting

partners within the network. The names of genes with more than 19 interactions are shown. **D** Calculation of eQTL score for each proximal gene. For participants from the UKBB, we derived the eQTL score of a gene based on the gene’s eQTLs identified in the fetal cerebral cortex. **E** 54 proximal genes showed nominally significant correlations between their eQTL scores and SA. **F** Stronger relationships between eQTL scores and SA were found in the multimodal compared with the unimodal cortical regions. **G** The expression of neural progenitor genes in the multimodal and the unimodal cortices. There are 200 genes for radial glia and IPC, respectively. Centerline, median; box limits, upper and lower quartiles; whiskers, 1.5x inter-quartile range; points, outliers.

stronger influence on cortical expansion than those distal to this core set of genes. To test this hypothesis, we divided all deleted and duplicated genes in our dataset into two groups based on their levels of spatial coexpression with neural progenitor genes (radial glia and IPC genes) across the 11 regions in the fetal cerebral cortex. Genes with high coexpression were labeled “proximal”, while those with low coexpression were labeled “distal” (Fig. 2A). To ensure a balanced sample size for CNV carriers with proximal/distal genes affected, we used the median coexpression value as a threshold (Supplementary Fig. 9, Supplementary Data 2 and Supplementary Fig. 11 for results obtained with a sliding-window approach). We then tested the relationship between the number of deleted/duplicated proximal and distal genes and SA of the 11 regions, as well as the total SA, while controlling for the effects of the other group of genes and genes without expression data.

We found that, on average, the deletion of one proximal gene was associated with a 0.0512 SD reduction in total SA (equal to about 8 cm<sup>2</sup> in our dataset,  $p = 1.34E-03$ ), and a mean 0.0385 SD reduction across

the 11 regions ( $p = 3.42E-06$ , Fig. 2B). Deletions of the proximal genes had larger effects in “multimodal” regions, including DFC, VFC, OFC, MFC, ITC, STC and IPC which subserve complex perceptual and cognitive processes, than in “unimodal” regions (including MIC, SIC, AIC, VIC; Fig. 2B and Supplementary Data 7). Deletions of distal genes had no effect on SA (total SA:  $\beta = 0.0027$ ,  $p = 0.8695$ ; across 11 regions: mean  $\beta = 0.0027$ ,  $p = 0.593$ ; distal vs. proximal genes:  $p = 3.5E-04$ , Supplementary Data 7). An exception to this pattern was observed for VIC, where lower SA was associated with deletions of distal genes ( $\beta = -0.041$ ,  $p = 0.044$ ) but not of proximal genes ( $\beta = -0.025$ ,  $p = 0.201$ ). Note that VIC was the only cortical region with a significant effect of distal genes on SA (Fig. 2B).

Among the 579 deleted proximal genes, 448 had at least one physical protein-protein interaction with other proximal genes; the most interacting genes were involved in DNA replication and cell cycle, such as *H4CI* and *CENPA* (Fig. 2C). Pathway enrichment analysis based on these interacting proximal genes identified a number of biological processes, with the top pathways involving DNA replication and repair

( $q$  value =  $5.68E-06$ ), metabolism of co-factors and vitamins ( $q$  value =  $9.39E-06$ ), and cell growth and death ( $q$  value =  $1.12E-05$ , Supplementary Dataset 8). No pathway enrichment was found for the 479 interacting distal genes (Supplementary Fig. 10), suggesting a heterogeneous nature of this group of genes. We found no duplication-related effects on SA for either proximal or distal genes (Proximal: total SA:  $\beta = -0.0042$ ,  $p = 0.6406$ ; across the 11 regions: mean  $\beta = -0.0049$ ,  $p = 0.074$ . Distal: total SA:  $\beta = -0.0046$ ,  $p = 0.5799$ ; across the 11 regions: mean  $\beta = -0.0004$ ,  $p = 0.778$ ; vs. proximal genes:  $p = 0.321$ . Figure 2B and Supplementary Data 7). This is not surprising given that our virtual ontogeny analysis indicated no spatial correlation between the duplication-related variations in SA and expression of genes specific to proliferative cell types.

### Expression quantitative trait loci of genes and surface area

In CNV studies, elucidating the impact of individual genes within multigenic variants poses a challenge. Next, we aimed to answer two questions: (1) whether the prenatal expression of individual genes belonging to the above proximal group predicts cortical SA in the adult brain; (2) as motivated by the above results (Fig. 2B), whether the expression of these genes has stronger effects on the multimodal regions compared with the unimodal regions. To do so, in 30,083 CNV non-carriers from the UKBB, we computed an eQTL score as a proxy for each proximal gene's expression in the fetal cerebral cortex (Fig. 2D). We then assessed the correlation between each gene's eQTL score and SA across the 11 regions using a mixed-effect model. Among 543 proximal genes with eQTL scores, 54 showed nominally significant correlations between their eQTL scores and SA; 4/54 genes (*HADHB*, *PRRT2*, *TMEM44*, *CORO1A*) survived FDR correction (Fig. 2E and Supplementary Data 9). In a replication sample of 16,212 individuals from UKBB, two of the four genes showed significant correlation, in the same direction, between their eQTL scores and SA (*HADHB*:  $\beta = 0.012$ ,  $p = 0.007$ ; *CORO1A*:  $\beta = 0.01$ ,  $p = 0.022$ ; *PRRT2*:  $\beta = 0.006$ ,  $p = 0.154$ ; *TMEM44*:  $\beta = 0.003$ ,  $p = 0.508$ ). In addition, the effect sizes (betas) of the 54 genes in the replication dataset were correlated with those in the discovery dataset ( $r = 0.454$ ,  $p = 5.64E-04$ ; Supplementary Fig. 13).

To answer the second question, we compared the effect sizes of the relationships between the 54 genes' eQTL scores and SA of the multimodal and of the unimodal regions. The correlations between eQTL scores and SA were stronger in multimodal (vs. unimodal) regions (Fig. 2F. Paired  $t$  test:  $t = -2.23$ ,  $df = 53$ ,  $p = 0.030$ ). Furthermore, to explore the underlying reasons for such unimodal-multimodal differences, we compared the expression of neural progenitor genes in the multimodal and the unimodal regions during mid-gestation. During this developmental period, the multimodal regions showed higher expression of genes specific to radial glia and IPC compared with the unimodal regions (Fig. 2G. radial  $g$   $p = 2.90E-20$ ; IPC:  $p = 3.26E-05$ ). But note that we did not replicate the difference in the eQTL-SA relationships between unimodal and multimodal regions in the replication sample ( $t = 0.367$ ,  $p = 0.715$ ).

### Specific CNVs and surface area

Previous research has reported large effects on SA of specific CNVs, including 1q21.1, 16p11.2, and 22q11.2, known to increase the risk of neurodevelopmental disorders<sup>21</sup>. Yet the underlying mechanisms and contributing genes of these multigenic variants remain elusive. Next, we investigated neurodevelopmental mechanisms underlying these CNV-related variations in SA using the same methodology as for genome-wide CNVs. First, we estimated the effect size of specific CNVs, namely 1q21.1, 16p11.2, 22q11.2, on SA by comparing normalized SA of the same 11 regions between CNV carriers (deletion or duplication) and non-carriers from clinical and community cohorts (Supplementary Data 10). Compared with the non-carriers, deletions of 1q21.1 and 22q11.2 had smaller SA across the 11 regions (1q21.1:  $-1.04$  SD,

$p = 5.23E-07$ ; 22q11.2:  $-0.57$  SD,  $p = 1.40E-03$ ), while duplications of 1q21.1 and 22q11.2 showed larger SA (1q21.1:  $0.45$  SD,  $p = 5.14E-05$ ; 22q11.2:  $0.18$  SD,  $p = 8.26E-03$ . Figure 3A and Supplementary Data 3). In contrast, for 16p11.2, deletions displayed larger SA (vs. non-carriers:  $0.39$  SD,  $p = 2.59E-03$ ), while duplications showed smaller SA ( $-0.54$  SD,  $p = 8.77E-05$ ).

Second, using virtual ontogeny, we found that deletions of 1q21.1 and 22q11.2 and duplications of 16p11.2 and 22q11.2 showed results similar to genome-wide CNV deletion: regions with higher expression of progenitor genes (radial glia and IPC) in mid-gestation showed stronger effects of CNVs on SA (Fig. 3B, Supplementary Fig. 6, Supplementary Fig. 8 and Supplementary Datas 4–6). These findings suggest that the proliferation and differentiation of neural progenitor cells may be a common cellular mechanism underlying the CNV-SA relationships.

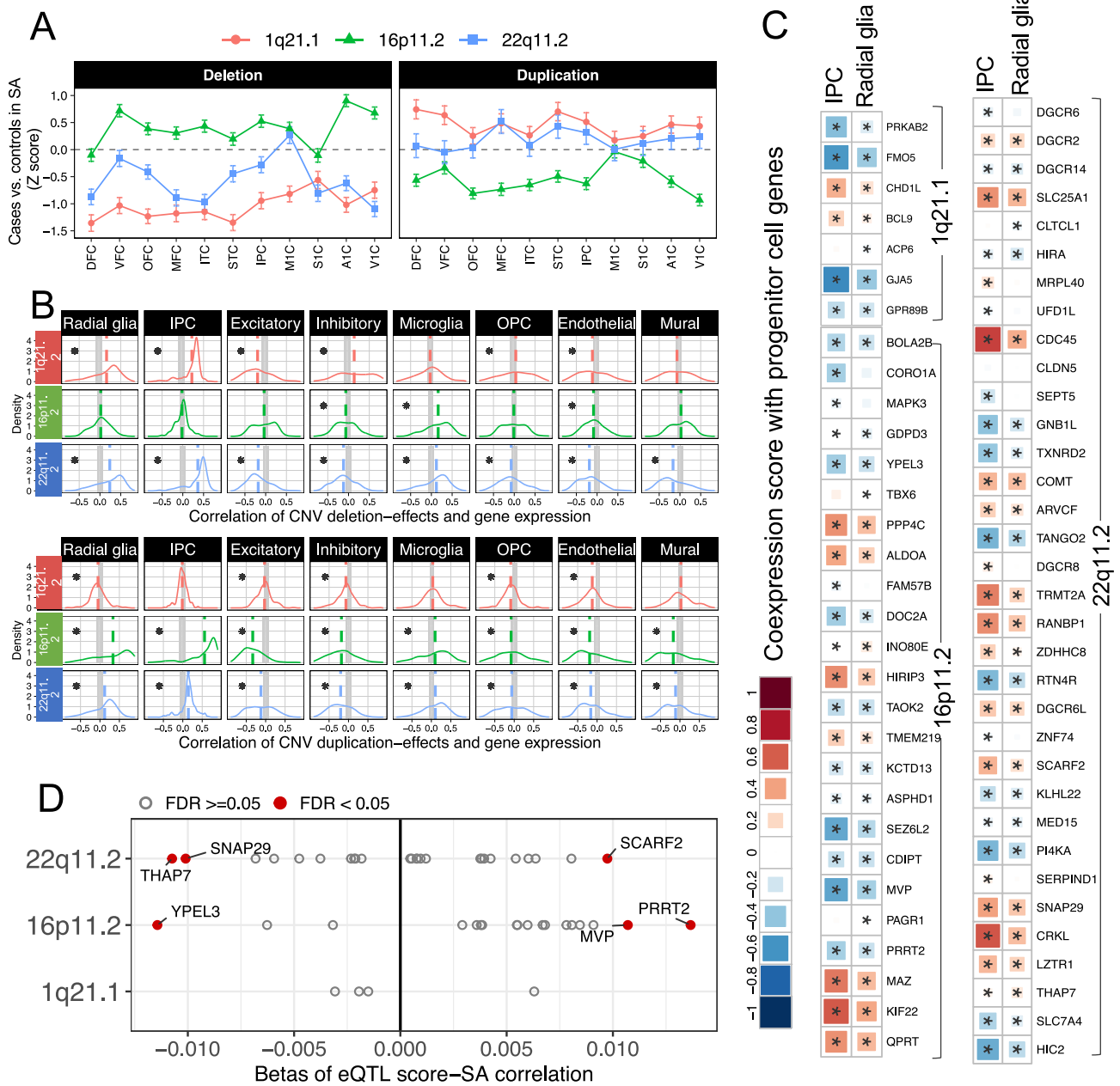
Third, we reasoned that genes coexpressed with the radial glia and IPC genes in the fetal cortex could be the main contributors to the effect of 1q21.1, 16p11.2, and 22q11.2 CNVs on the cortical expansion. Among the 65 genes (out of the 88 genes within the three genomic loci) with available prenatal expression data, 50 genes were spatially coexpressed with both radial glia and IPC genes across the 11 regions during the mid-gestation (6/7 genes in 1q21.1, 19/24 genes in 16p11.2, and 25/34 genes in 22q11.2. Fig. 3C). We subsequently tested whether the expression of these 50 genes in the fetal cortex predicts SA of the adult brain by testing the correlation between each gene's eQTL score and SA of the 11 regions in 30,083 CNV non-carriers from the UKBB. Among the 47 genes with eQTL scores, 6 showed significant correlations between their eQTL scores and the SA (*YPEL3*, *PRRT2*, *MVP*, *SCARF2*, *SNAP29*, *THAP7*. Figure 3D and Supplementary Data 11).

## Discussion

Leveraging current knowledge of developmental processes underlying cortical development, this genome-wide association study of rare CNVs provides novel insights into the genetic architecture of SA of the human cerebral cortex. We demonstrated that, in general populations, individuals with a higher number of deleted protein-coding genes have a smaller cortical SA. This effect was stronger (1) in cortical regions with higher mid-gestational expression of neural progenitor genes, and (2) for genes that were strongly co-expressed with neural progenitor genes. These findings were extended to common variants affecting transcription, showing that eQTLs of some of these co-expressed genes were associated with SA. Enrichment analyses of these co-expressed genes pointed to diverse pathways involved in cell proliferation, growth, and death, such as DNA replication/repair and metabolism of co-factors and vitamins. Altogether, our findings suggest a model wherein CNVs interfere with the tangential expansion of the human cerebral cortex by altering the expression of genes involved in neural proliferation in the fetal period.

Our findings align well with the radial unit hypothesis, which was formulated based on experimental work on non-human primates<sup>12</sup>. In the case of the human cerebral cortex, we showed that cortical regions with greater expression of proliferative cells (radial glia and IPCs) in the fetal cerebral cortex were more affected by CNVs. We also revealed that deletions of genes that had strong (rather than weak) coexpression with neural progenitor genes were associated with smaller SA. These effects were stronger in multimodal vs. unimodal cortical regions. The latter finding was corroborated by utilizing eQTLs in a large sample of CNV non-carriers, showing that the associations between eQTLs and SA were again stronger in the multimodal compared with unimodal regions.

In addition, as revealed by the enrichment analysis of the proximal genes within a protein-protein interactions network, these genes appear to be involved in crucial cell proliferation pathways, including regulation of the cell cycle, metabolism of co-factors, and vitamins



**Fig. 3 | Neurodevelopmental CNVs and cortical expansion.** (A) Case-control differences in normalized SA between CNV deletion/duplication at 1q21.1 (N deletion/duplication = 36/30), 16p11.2 (N deletion/duplication = 82/75), and 22q11.2 (N deletion/duplication = 69/19) and the control (CNV noncarriers,  $N = 347$ ). The error bars represent the standard error of estimated group differences for each region. (B) Virtual ontogeny through the lens of the correlation between CNV effects on SA and expression of cell-specific genes. The asterisk denotes a significant spatial

correlation between the group differences in SA and the expression of cell-specific genes ( $p < 0.05$  after FDR). (C) Coexpression scores between genes within the three CNVs and genes specific to radial glia and IPC. Coexpression scores represent the mean spatial correlations between prenatal expression of a CNV gene and 200 cell-specific genes and thus, range from  $-1$  to  $1$ . The color and size of squares indicate the magnitude of coexpression. (D) Effect size of correlations between CNV genes' eQTL score and SA of the II regions.

involved in DNA synthesis and repair. For example, the E2F-enabled inhibition of the pre-replication complex formation pathway is involved in regulating the initiation of DNA synthesis in the early cell cycle (G1/S-phase)<sup>35</sup>. The formyltetrahydrofolate biosynthesis pathway is crucial for the synthesis of tetrahydrofolic acid which acts as a cofactor essential for the synthesis or anabolism of amino acids and nucleic acids<sup>36</sup>. The Proliferating Cell Nuclear Antigen (PCNA)-Dependent Long Patch Base Excision Repair pathway is a key mechanism for removing damage-containing strands of DNA<sup>37</sup>. These results are consistent with previous reports indicating that common genetic variants

linked to adult SA were enriched in regulatory activities in neural progenitor cells during fetal development<sup>15,17</sup>. According to the radial unit hypothesis and related experimental work, the cortical surface area of a given region depends on the size of the neural progenitor population<sup>13</sup>. The self-renewing and amplification of radial glia and IPCs are critical for determining the size of the neuronal population: two radial glial cells, which can self-renew and produce self-amplifying IPCs, could generate more than 80 neurons following 8 rounds of cellular cycles<sup>11</sup>. Thus, subtle deviations or variations in the molecular processes underlying the division of progenitor cells may have a

profound impact on the tangential expansion of the cerebral cortex, especially in the multimodal regions, which have high expression of neural progenitor genes during midgestation and prolonged developmental trajectory<sup>38</sup>.

Furthermore, our findings suggest that genes not specific to neural progenitor cells may also influence cortical expansion, likely via intricate interactions within various biological networks. For example, most CNVs' effects on SA had associations with genes specific to endothelial and mural cells, components of developing cortical blood vessels, which are key regulators of neuronal proliferation<sup>39</sup>. In addition, microglia, an integral component of cortical proliferative zones, have been found to regulate the pool size of neural progenitor cells in the fetal cortex: deactivating or eliminating microglia in the fetal cerebral cortex increased the number of neural precursor cells<sup>40,41</sup>. In addition to the specific pathways described above, genes proximal to the neural progenitor genes were associated with broad biological pathways involving, for example, macromolecule metabolism and transport. The genes identified as key contributors to the effects of the three specific CNVs on SA are also involved in diverse biological processes with potential direct or indirect influences on neurogenesis, such as the regulation of DNA transcription (*THAP7*) cellular senescence (*YPEL3*), and cell adhesion (*SCARF2*).

The above observations align with the concept of omnigenic model, which suggests that complex traits are largely influenced by genes without direct links to the trait<sup>34</sup>. In other words, since the normal tangential growth of the cerebral cortex requires coordination among various cellular processes implemented by other systems, genetic variants of non-neuronal processes can have downstream effects on cortical expansion. For instance, perturbations in vascular development can lead to abnormal processes during cortical expansion, including premature neuronal differentiation and mispositioning growth of neural progenitor cells<sup>39,42–44</sup>.

It is important to note that the generalizability and interpretation of these findings are subject to certain limitations. Our study has focused on II cortical regions with available gene-expression data from the prenatal period. Although this limits the generalizability of the findings throughout the cerebral cortex, note that the II regions include a number of unimodal and multi-modal regions in all four cerebral lobes. Furthermore, we also included results with regard to the total surface area. Our main analyses were conducted using pooled data from adults (UKBB and SYS parents) and adolescents (IMAGEN, SYS). While the majority of the data were derived from adults, we believe that age at MRI has, at most, a minor influence on generalizing our findings. This is because the most significant cortical expansion occurs during the prenatal and early postnatal periods, reaching ~83% of adult size by 2 years of age<sup>45,46</sup>. Moreover, the eQTL scores used in our analyses were calculated based on the effects of genotypes on gene expression during fetal development. Lastly, the current study focused on testing the explanatory power of the radial unit hypothesis in the context of the CNVs, leveraging transcriptomic data from the mid-gestational fetal cortex. We acknowledge that other biological process, such as dendritic arborization and myelination during the first two years of life, may also contribute to the CNV-surface area relationship.

Taken together, this genome-wide association study of rare CNVs revealed a group of genes that can influence cortical expansion, providing mechanistic insights into the early development of the human cerebral cortex, and the potential impact of rare gene deletions/duplications on this process. Our findings highlight that deficits in neural progenitor proliferation could serve as a parsimonious theoretical framework to unify hypotheses concerning the molecular mechanisms underlying CNV-related brain abnormalities. We believe that this knowledge advances our understanding of how rare genetic variants may affect cognition and mental health.

## Methods

### Participants

We included data from three community-based cohorts: the IMAGEN study ( $n = 1661$ ), which recruited 2090 adolescents in four European countries; the Saguenay Youth Study (SYS,  $n = 1493$ ), which included both children and parents from 486 families in Quebec, Canada, and the UK Biobank (UKBB,  $n = 35,861$ ). Carriers of the three specific CNVs, namely 1q21.1, 16p11.2, and 22q11.2, were ascertained in several clinical cohorts and UKBB (Supplementary Data 10). These CNVs were selected because they are (1) the most frequently identified CNVs in the clinic and reported in our clinical cohorts, (2) known for their risk of neurodevelopmental disorders and detrimental effects on cognition, (3) have large effects on surface area<sup>21</sup>, and (4) relatively prevalent in the population<sup>19,22</sup>. As described in previous study<sup>21,47</sup>, carriers of these specific CNVs were selected based on following breakpoints according to the reference genome GRCh37/hg19: 16p11.2 proximal (BP4-5, 29.6–30.2MB), 1q21.1 distal (Class I, 146.4–147.5MB & II, 145.3–147.5MB), 22q11.2 proximal (BPA-D, 18.8–21.7MB). CNV carriers from clinical cohorts were referred to the genetic clinic either because of the presence of symptoms or because they were relatives of these symptomatic individuals. Control individuals from clinical cohorts do not carry any CNV at these genomic loci. Control individuals from UKBB were non-carriers matched for age and sex with CNV carriers from UKBB. Each cohort received ethical approvals from Local Research Ethics Committees.

### CNV detection and annotation in community-based cohorts

For the three community-based cohorts, we called CNVs using PennCNV<sup>48</sup> and QuantiSNP<sup>49</sup> with the same parameters as previously published: number of consecutive probes for CNV detection  $\geq 3$ , CNV size  $\geq 1$  Kb, and confidence scores  $\geq 15$ <sup>25</sup>. To minimize false discoveries, we merged CNVs detected by PennCNV and QuantiSNP with CNVision. We then applied a CNV inheritance analysis algorithm to concatenate adjacent CNVs of the same type, separated with a gap lower than 150 kb. Subsequently, we selected CNVs for analyses based on the following criteria: confidence score  $\geq 30$  with at least one detection algorithm, size  $\geq 50$  kb, and  $< 50\%$  overlap with segmental duplication. A detailed description of CNV calling is available in the Supplementary Methods. Note that the entire genome was searched without any pre-selection criteria beyond the three parameters reported above. No CNV regions were pre-selected based on gene characteristics, such as known neurodevelopmental effects or clinical significance.

Detected CNVs were then annotated using Gencode V19 (hg19) with ENSEMBL. We only considered genes with all transcripts (UTRs, start and stop codons, exons, and introns) fully encompassed within a CNV. Our dataset included a total of 5822 genes that were deleted or duplicated at least once. These genes were then annotated by the inverse loss-of-function observed/expected upper-bound fraction (1/LOEUF) using gnomAD, version 2.1.1.<sup>28</sup> The 1/LOEUF score is available for 5525 genes and ranges from 0.5 (tolerant to loss-of-function variants) to 27.0 (intolerant to loss-of-function variants). We calculated a CNV burden score by summing up 1/LOEUF scores of all genes encompassed in the deleted or duplicated CNVs for each individual. Individuals without CNVs or with no genes encompassed in any CNVs were assigned a score of 0. Sensitivity analysis with the number of genes deleted/duplicated as a CNV burden score was done (Supplementary Fig. 4).

### Regional surface area from magnetic resonance imaging data

All T1-weighted (T1w) MRI data (1.0–1.1 mm isotropic resolution) were processed with FreeSurfer. A detailed description of MRI acquisition and processing is available in Supplementary Methods and Supplementary Data 12. Briefly, T1w scans were processed through an automated cortical reconstruction pipeline in FreeSurfer ('recon-all').



Subsequently, the cortical reconstruction outputs were used to extract SA based on a parcellation scheme using FreeSurfer. We then derived SA for the 11 cortical regions that intersect with prenatal tissue sampling from the PsychENCODE Consortium. Further details are available in the Supplementary Methods.

### Estimating effects of CNVs on regional surface area

To estimate the effects of genome-wide CNV burden on SA in the general population, we pooled the IMAGEN, SYS, and UKBB data together and performed linear mixed-effect models with Restricted Maximum Likelihood. In the model, the CNV burden score was the independent variable of interest, and the SA (Z-score) of each of the 11 cortical regions was the dependent variable. Sex, age, scan site, and first 10 genetic principal components were included as fixed effects, and the family ID was the random effect to account for potential confounders. The effect size (beta) estimation with the model was done for deletion and duplication separately. Thus, we obtained one effect size (beta) profile (across 11 regions) for genome-wide CNV deletion and duplication, respectively. Subsequently, we tested the significance of the effect size (beta) profiles across the 11 regions using a permutation approach. Specifically, the permutation test included the following steps: (1) Permuting the SA of the 11 regions across individuals while preserving the inter-regional associations; (2) Calculating betas between the resampled SA and the CNV burden score to obtain a simulated effect-size profile using corresponding models; (3) Repeating steps 1 and 2 10,000 times and then averaging the simulated profiles to derive a null profile; (4) Using paired *t* test to test if the observed effect-size profile is different from the null profile.

### Virtual ontogeny

To investigate developmental mechanisms underlying the CNV-related variations on cortical expansion at the cellular level, we tested the spatial relationships between the profile of CNV-SA effect sizes and expression of cell-specific genes in the fetal cortex, as introduced in our previous work<sup>3</sup>. We first selected 200 genes with highest specificity for each of 8 cell types in the human cerebral cortex during the first and second trimesters based on single-cell RNA-sequencing data (5 donors, 6–22 gestational weeks)<sup>31</sup>. The specificity of a gene for a cell type was defined by the proportion of the gene's expression in that cell type relative to the total expression across all cell types<sup>50</sup>. Sensitivity analyses with the top 100 and 300 cell-specific genes, genes specific to 16 mid-gestational cortical cell types from another single-cell transcriptomic dataset<sup>33</sup> were also conducted (Supplementary Data 4 and Supplementary Methods). Prenatal expression levels of these cell-specific genes across 11 regions were obtained from the PsychENCODE bulk RNA sequencing dataset (14 donors, 10–22 post-conception weeks; <http://development.psychencode.org/>). We filtered out genes with low expression (defined as less than 0.5 RPKM mean expression) and adjusted for donor-specific covariates such as sex, hemisphere, RNA integrity number, ethnicity, and sequencing site. Then the median gene expression per cortical region across all 14 donors was used for the subsequent analysis. A detailed description of the processing of single-cell and bulk RNA sequencing data is available in the Supplementary Methods and previous publications<sup>3,32,50</sup>.

We then tested the spatial correlation between the inter-regional expression profiles of genes specific to a cell type and the profile of CNV-SA effects using a permutation test. Specifically, for a given cell type, we calculated the bicor between the cell-specific genes and CNV-SA effects across the 11 regions. Bicor is median-based rather than mean-based and, thus, is less sensitive to outliers than other correlation methods. The mean of the resulting 200 correlation coefficients was then tested for significance using a permutation approach. The null hypothesis postulates that the cell-specific genes have the same correlations with the CNV-SA effect profile as any random set of genes. Thus, in the permutation test, we randomly selected 200 genes not

specific to the cell type and calculated the mean of correlations between these genes' expression and the CNV-SA effect profile. This was repeated 10,000 times to generate a null distribution. The observed mean correlation coefficient for the cell type was compared against the null distribution using a two-sided test with FDR for the 8 cell types (Supplementary Methods).

In addition, we also performed two different sensitivity analyses: (1) bootstrapped estimation for the distribution of the mean correlation and (2) gene-set enrichment analysis. The bootstrap analysis aimed to obtain 95% confidence intervals for the mean correlation between the profile CNV-SA effects and the expression of cell-specific genes and tested if the distribution differed significantly from zero. For each cell type, we bootstrapped the 200 cell-specific genes as well as the 14 donors to account for the uncertainty related to the limited sample of donors. The gene-set enrichment analysis was done with R package "clusterprofile"<sup>51</sup> to test if the cell-specific genes have significant correlations with the CNV-SA effect profile than the other genes available in the bulk RNA dataset. Further details of these sensitivity analyses can be found in the Supplementary Methods.

### Proximal and distal genes in network of progenitor genes

We used prenatal expression data from PsychENCODE to estimate spatial coexpression (across the 11 cortical regions) between each of the deleted/duplicated genes within the genome-wide CNVs and genes specific to radial glia and IPC in the fetal cortical cortex. There were 3646 out of 5822 CNV genes with prenatal expression data. For each gene, we calculated the mean bicor between the expression profile of the gene and the expression profiles of 200 genes specific for radial glia and IPC, respectively. The coexpression with the neural progenitor cells was defined as the absolute value of the average coexpression between the gene and genes specific to radial glia and IPC. Thus, in this unsigned weighted network, a strong connection/link means that a gene has strong coexpression, either negative or positive, with neural progenitor cell genes in the midgestational fetal cortex. Note that genes showing strong coexpression, either negative or positive, had a larger effect size on SA compared with genes with weak coexpression (Supplementary Fig. 12).

We divided the 3646 genes into 2 groups based on the median of their coexpression values ( $= 0.193$ ). We termed the genes with higher and lower coexpression as "proximal" and "distal" genes, respectively. Then for each participant from the community cohorts, we obtained the counts of deleted or duplicated genes belonging to the proximal, distal, and other genes (genes without expression data). There were 843 and 1239 participants with at least one proximal and one distal gene deleted, respectively. A total of 5192 and 5112 participants had at least one proximal and one distal gene duplicated, respectively. We used linear models to estimate the effect size of proximal and distal genes on the SA of the 11 regions and the total SA of the whole cerebral cortex (Z-scored). Since many CNVs encompass more than one gene, we included the counts of proximal and distal genes as the independent variable of interest and the count of other genes as one of the confounders (sex, age, scan site, first 10 genetic principal components). This was done for deletion and duplication separately. The significance of the effect size (beta) profiles across the 11 regions for proximal and distal genes and differences between the two profiles was tested using *t* tests.

In addition to the above dichotomous approach, we also used a sliding window approach to obtain a nuanced picture of the relationship between genes' proximal/distal position to neural progenitors and their effects on cortical expansion. We first divided the 3646 genes deleted and/or duplicated in our dataset into 20 subgroups based on percentiles of their coexpression with neural progenitor genes. We then included the first four subgroups as the first "window" and moved the window forward by a step size of one subgroup. This resulted in 17 windows in total, with Window 1 having the lowest coexpression

(mean = 0.034) and Window 17 having the highest coexpression (mean = 0.428) with the neural progenitor genes. On average, there were 241 and 611 unique genes within a window deleted and duplicated at least once, respectively (Supplementary Fig. 9). We used this division approach to ensure that each window contained an equal number of genes, thus, making the number of participants with genes within a window deleted/duplicated comparable across windows, and having enough power to estimate effects of genes on SA. For a given window, we obtained the counts of deleted or duplicated genes belonging to the window, and outside the window for each participant from the community cohorts. We then estimated the effect of genes within the window on SA (11 regions and whole cerebral cortex), while controlling the effects of genes outside the window. On average, there were 479 and 1314 participants with at least one gene within a window deleted and duplicated, respectively (Supplementary Fig. 9).

### Protein-protein interaction and pathway enrichment analysis

Physical protein-protein interactions for the deleted proximal and distal genes were obtained from the Integrated Interactions Database (IID) version 2021-05 (<https://iid.ophid.utoronto.ca/>)<sup>52</sup>. The submitted gene lists (i.e., 579 deleted proximal and 633 deleted distal genes) were mapped to proteins, and their protein-protein interactions were derived using the default settings of IID. Individual protein interaction networks were then loaded into NAViGaTOR 48 ver. 3.0.17 for further annotation, analysis, and visualization<sup>53</sup>. There were 448 proximal and 479 distal genes that had at least one physical protein-protein interaction with other proximal and distal genes, respectively. Following, pathway enrichment analysis on these proximal/distal genes was conducted using pathDip, version 5.0.32.3<sup>54</sup>.

### Expression quantitative trait loci of genes and surface area

To investigate which of the proximal genes level of expression in the fetal cortex was associated with the SA, we used an eQTL score constructed from eQTLs of each proximal gene as the proxy for the gene's expression level in the fetal cerebral cortex. The largest eQTL study available for this purpose (201 mid-gestational human cortices) provides the summary statistics of eQTLs for 15,925 tested genes<sup>55</sup>. In this dataset, the 543 (out of 579) deleted proximal genes have at least one eQTL (nominal  $p$ -value < 0.01). Then, for each proximal gene, we calculated the eQTL score by summing up genotypes of each CNV non-carrier from UKBB, weighted by the corresponding genotype effect size (beta) derived from the eQTL study ( $p$ -value < 0.01). The calculation of eQTL scores was done using PRSice-2 with default clumping parameters (kb = 250,  $r^2$  = 0.1). A higher eQTL score of a gene indicates a genetic predisposition for higher expression levels of the gene in the fetal cerebral cortex. We then used linear mixed models to estimate the effect of these eQTL scores on the SA of each of the 11 cortical regions in the UKBB participants, with sex and the first 10 genetic principal components as covariates and participants as a random effect. Note that the SA and eQTL scores were normalized (Z-score) across individuals before performing the analyses. There were 54 genes that showed nominal significant correlations between their eQTL scores and SA. Following, we used the same mixed effect model to obtain the effect size (standardized beta) of correlations between the eQTL scores of these 54 genes and SA of unimodal and multimodal regions, respectively. Then paired t-test was used to compare the absolute value of effect sizes of eQTL-SA correlation between unimodal and multimodal regions.

To compare the expression of neural progenitor genes in the multimodal and the unimodal regions during midgestation, expression levels of radial glia and IPC genes for all midgestation donors ( $n = 14$ ) in PsychENCODE were unit-scaled following adjustment for sex, hemisphere, RIN, ethnicity, and sequencing site. Differences in expression between the clusters were evaluated using a linear mixed

effects model adjusting for the age of donors (in  $\log_2(\text{days})$ ), with random effects for donor id and gene id.

### Mechanism of specific CNVs and surface area

To estimate the effect size on SA of 1q21.1, 16p11.2, and 22q11.2 CNV deletions and duplications, we conducted a linear regression model for each CNV deletion and duplication, separately. In the model, the CNV status (1 for carrier, 0 for non-carrier) was the independent variable, and the normalized (Z-score) SA of each of the 11 cortical regions was the dependent variable. Age, sex, and data sources were included as covariates. Subsequently, we tested the significance of the effect size (beta) profiles across the 11 regions for each CNV using the same permutation approach as for the genome-wide CNVs with FDR across the six genetic variants (deletion/duplication  $\times$  3 CNVs). Virtual ontology analyses were conducted with the derived effect size profiles to neurodevelopmental mechanisms underlying these CNV-related variations in SA.

To identify potential key genes contributing to the CNV-related variations in SA, we first used prenatal expression data from PsychENCODE to identify CNV genes that were spatially co-expressed with the radial glia and IPC genes in the fetal cortex. Note that there are 88 genes within the three clinical CNVs: 8 in 1q21.1, 34 in 16p11.2, and 46 in 22q11.2. Out of these 88 genes, there were 65 genes (7 for 1q21.1, 24 for 16p11.2, and 34 for 22q11.2) with prenatal expression data in the 11 regions. For each of the 65 CNV genes, we calculated the bicor between the expression profile of the CNV gene and the expression profiles of 200 genes specific for radial glia and IPC, respectively. Then the mean of the resulting 200 bicor was tested for significance using the similar permutation approach as described above ("Virtual Ontogeny"). Thus, the mean of correlations between the expression of 200 randomly selected genes and the CNV gene's expression profile was calculated. This was repeated 5000 times to generate a null distribution. The observed mean correlation coefficient for the cell type (radial glia or IPC) was compared against the null distribution using a two-sided test with FDR for the 65 CNV genes. There were 50 genes spatially co-expressed with both radial-glia and IPC genes across the 11 regions during the midgestation. Following, we tested the correlations between the eQTL scores of 47 out of these 50 genes (3 genes had no eQTL scores) and the SA of 11 regions with FDR correction.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw data from population samples and clinical cohorts are protected and are not available due to data privacy laws. Access can be obtained upon application. All data from the UK Biobank are available to other investigators online ([ukbiobank.ac.uk](http://ukbiobank.ac.uk)). IMAGEN data are available upon request via: <https://www.imagen-project.org/>. Data from the Saguenay Youth Study are available upon request addressed to T.P. and Z.P. Data on 1q21.1, 16p11.2, and 22q11.2 from clinical cohorts are available upon request addressed to S.J. and C.E.B., and via <https://www.sfari.org/>. Open access resources used in the analysis are present in the links below: Single-cell RNA-seq data: <https://organoidreportcard.cells.ucsc.edu>. Bulk RNA sequencing PsychENCODE: <http://development.psychencode.org/>. Parcellation and cell-specific genes: <https://figshare.com/s/a6f4ec8a13a087b88a77>. Summary statistics of eQTLs in fetal cortex: <https://labs.dgsom.ucla.edu/geschwind/pages/eqtl-browser>. Source data are provided in this paper.

### Code availability

Script of brain parcellation fragmenter: [https://github.com/miykael/parcellation\\_fragmenter](https://github.com/miykael/parcellation_fragmenter). The processing scripts and custom analysis

software used in this work are available in a publicly accessible GitHub repository: <https://github.com/Zhijie31/CNV-SA>.

## References

- Bethlehem, R. A. I. et al. Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
- Patel, Y., Parker, N., Salum, G. A., Pausova, Z. & Paus, T. General psychopathology, cognition, and the cerebral cortex in 10-year-old children: Insights from the adolescent brain cognitive development study. *Front. Human Neurosci.* **15**, 781554 (2022).
- Patel, Y. et al. Virtual ontogeny of cortical growth preceding mental illness. *Biol. Psychiatry* **92**, 299–313 (2022).
- Schnack, H. G. et al. Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cereb. Cortex* **25**, 1608–1617 (2015).
- Cafiero, R., Brauer, J., Anwander, A. & Friederici, A. D. The concurrence of cortical surface area expansion and white matter myelination in human brain development. *Cereb. Cortex* **29**, 827–837 (2019).
- Hill, J. et al. Similar patterns of cortical expansion during human development and evolution. *Proc. Natl. Acad. Sci.* **107**, 13135–13140 (2010).
- Tamnes, C. K. et al. Development of the cerebral cortex across adolescence: A multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. *J. Neurosci.* **37**, 3402–3412 (2017).
- Fernández, V., Llinares-Benadero, C. & Borrell, V. Cerebral cortex expansion and folding: what have we learned? *EMBO J.* **35**, 1021–1044 (2016).
- Hansen, D. V., Lui, J. H., Parker, P. R. L. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
- Kriegstein, A., Noctor, S. & Martínez-Cerdeño, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nat. Rev. Neurosci.* **7**, 883–890 (2006).
- Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011).
- Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
- Nonaka-Kinoshita, M. et al. Regulation of cerebral cortex size and folding by expansion of basal progenitors. *EMBO J.* **32**, 1817–1828 (2013).
- Strike, L. T. et al. Genetic complexity of cortical structure: Differences in genetic and environmental factors influencing cortical surface area and thickness. *Cereb. Cortex* **29**, 952–962 (2019).
- Grasby, K. L. et al. The genetic architecture of the human cerebral cortex. *Science* **367**, eaay6690 (2020).
- Shadrin, A. A. et al. Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *NeuroImage* **244**, 118603 (2021).
- Shin, J. et al. Global and regional development of the human cerebral cortex: Molecular architecture and occupational aptitudes. *Cereb. Cortex* **30**, 4121–4139 (2020).
- Han, L. et al. Functional annotation of rare structural variation in the human brain. *Nat. Commun.* **11**, 2990 (2020).
- Kirov, G. et al. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol. Psychiatry* **75**, 378–385 (2014).
- Marshall, C. R. et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49**, 27–35 (2017).
- Modenato, C. et al. Effects of eight neuropsychiatric copy number variants on human brain structure. *Transl. Psychiatry* **11**, 399 (2021).
- Sønderby, I. E. et al. Effects of copy number variations on brain structure and risk for psychiatric illness: Large-scale studies from the ENIGMA working groups on CNVs. *Hum. Brain Mapp.* **43**, 300–328 (2022).
- Caseras, X. et al. Copy number variants increasing risk for schizophrenia: Shared and distinct effects on brain morphometry and cognitive performance. *Biol. Psychiatry Glob. Open Sci.* **3**, 902–911 (2023).
- Douard, E. et al. Effect sizes of deletions and duplications on autism risk across the genome. *Am. J. Psychiatry* **178**, 87–98 (2021).
- 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).
- Huguet, G. et al. Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability. *Mol. Psychiatry* **26**, 2663–2676 (2021).
- Männik, K. et al. Copy number variations and cognitive phenotypes in unselected populations. *JAMA* **313**, 2044–2054 (2015).
- Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
- 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* **79**, 699–709 (2022).
- Arnatkeviciute, A., Fulcher, B. D., Bellgrove, M. A. & Fornito, A. Imaging transcriptomics of brain disorders. *Biol. Psychiatry Glob. Open Sci.* **2**, 319–331 (2022).
- Bhaduri, A. et al. Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).
- Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
- Polioudakis, D. et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron* **103**, 785–801 (2019).
- Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
- Bosco, G., Du, W. & Orr-Weaver, T. L. DNA replication control through interaction of E2F-RB and the origin recognition complex. *Nat. Cell Biol.* **3**, 289–295 (2001).
- Ragsdale, S. W. Catalysis of methyl group transfers involving tetrahydrofolate and B(12). *Vitam. Horm.* **79**, 293–324 (2008).
- Gary, R., Kim, K., Cornelius, H. L., Park, M. S. & Matsumoto, Y. Proliferating cell nuclear antigen facilitates excision in long-patch base excision repair \*. *J. Biol. Chem.* **274**, 4354–4363 (1999).
- Kostović, I. et al. Perinatal and early postnatal reorganization of the subplate and related cellular compartments in the human cerebral wall as revealed by histological and MRI approaches. *Brain Struct. Funct.* **219**, 231–253 (2014).
- Paredes, I., Himmels, P. & de Almodóvar, C. R. Neurovascular communication during CNS development. *Dev. Cell* **45**, 10–32 (2018).
- Cunningham, C. L., Martínez-Cerdeño, V. & Noctor, S. C. Microglia regulate the number of neural precursor cells in the developing cerebral cortex. *J. Neurosci.* **33**, 4216–4233 (2013).
- Barger, N. et al. Microglia: An intrinsic component of the proliferative zones in the fetal Rhesus monkey (*Macaca mulatta*) cerebral cortex. *Cereb. Cortex* **29**, 2782–2796 (2019).
- Javaherian, A. & Kriegstein, A. A stem cell niche for intermediate progenitor cells of the embryonic cortex. *Cereb. Cortex* **19**, i70–i77 (2009).
- Lange, C. et al. Relief of hypoxia by angiogenesis promotes neural stem cell differentiation by targeting glycolysis. *EMBO J.* **35**, 924–941 (2016).
- Tata, M. et al. Regulation of embryonic neurogenesis by germinal zone vasculature. *Proc. Natl. Acad. Sci. USA* **113**, 13414–13419 (2016).

45. Lyall, A. E. et al. Dynamic development of regional cortical thickness and surface area in early childhood. *Cereb. Cortex* **25**, 2204–2212 (2015).
  46. Li, G. et al. Mapping region-specific longitudinal cortical surface expansion from birth to 2 years of age. *Cereb. Cortex* **23**, 2724–2733 (2013).
  47. Kopal, J. et al. Rare CNVs and phenome-wide profiling highlight brain structural divergence and phenotypical convergence. *Nat. Hum. Behav.* **7**, 1001–1017 (2023).
  48. 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).
  49. 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).
  50. Bryois, J. et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat. Genet.* **52**, 482–493 (2020).
  51. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
  52. Kotlyar, M., Pastrello, C., Sheahan, N. & Jurisica, I. Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic Acids Res.* **44**, D536–D541 (2016).
  53. Brown, K. R. et al. NAViGaTOR: Network analysis, visualization and Graphing Toronto. *Bioinformatics* **25**, 3327–3329 (2009).
  54. Rahmati, S., Abovsky, M., Pastrello, C. & Jurisica, I. pathDIP: an annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis. *Nucleic Acids Res.* **45**, D419–D426 (2017).
  55. Walker, R. L. et al. Genetic control of expression and splicing in developing human brain informs disease mechanisms. *Cell* **179**, 750–771 (2019).
- (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; the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797), U.S.A. (RO1 MH085772-01A1), and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centers of Excellence. SYS has been funded by the Canadian Institutes of Health Research and the Heart and Stroke Foundation of Canada. We thank all of the families participating at the Simons Searchlight sites, and Simons Searchlight Consortium (formerly the Simons VIP Consortium), and the 16p11.2 European Consortium. We appreciate obtaining access to brain-imaging and phenotypic data on SFARI Base. Approved researchers can obtain the Simons Searchlight population dataset described in this study by applying at <https://base.sfari.org>. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

### Author contributions

Z.L., S.J., and T.P. designed the study and drafted the paper. Z.L., K.K., J.K., G.H., Z.S., M.J., C.E.B., and I.J. analyzed neuroimaging and genetic data. T.P. and Z.P. designed and directed the Saguenay Youth Study. All authors provided feedback on the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56855-1>.

**Correspondence** and requests for materials should be addressed to Sebastien Jacquemont or Tomas Paus.

**Peer review information** *Nature Communications* thanks Mingfeng Li, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

### Acknowledgements

We thank Chunhui Hao for his helpful comments and inspiration. We also thank Steven Tilley, Yash Patel, and James Hengenius for their help in analyzing MRI data and gene coexpression. This research was enabled by support provided by Calcul Quebec and Compute Canada. This work is supported by a grant from the Brain Canada Multi-Investigator initiative and CIHR grant 159734 (S.J. and T.P.). The Cardiff CNV cohort 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). The CHUV cohort was supported by the SNF (Maillard Anne, Project, PMPDP3 171331). Data from the UCLA cohort provided by CEB (participants with 22q11.2 deletions or duplications and controls) was supported through grants from the NIH (U54EB020403), NIMH (R01MH085953, R01MH100900, R03MH105808), and the Simons Foundation (SFARI Explorer Award). IMAGEN received support from the European Union-funded FP6 Integrated Project IMAGEN (LSHM-CT- 2007-037286), the Horizon 2020 funded ERC Advanced Grant 'STRATIFY' (695313), Human Brain Project (HBP SGA 2, 785907, and HBP SGA 3, 945539), the Medical Research Council Grant 'c-VEDA' (MR/N000390/1), the NIH (R01DA049238), the NIHR Biomedical Research Center at South London and Maudsley NHS Foundation Trust and King's College London, the Bundesministerium für Bildung und Forschung (BMBF grants O1GS08152; O1EVO711; Forschungsnetz AERIAL O1EE1406A, O1EE1406B), the Deutsche Forschungsgemeinschaft (DFG grants SM 80/7-2, SFB 940, TRR 265, NE 1383/14-1), the Medical Research Foundation and Medical Research Council (grants MR/RO0465X/1 and MR/SO20306/1), the NIH funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01). Further support was provided by grants from: - the ANR (ANR-12-SAMA-0004, AAPG2019 - GeBra), the Eranet Neuron (AF12-NEURO008-01 - WM2NA; and ANR-18-NEURO0002-01 - ADORe), the Fondation de France

---

**the IMAGEN Consortium**

---

Tomas Paus  <sup>1,2,11,12</sup> 

**16p11.2 European Consortium**

---

Sebastien Jacquemont  <sup>1,6,12</sup> 