

# UC San Diego

## UC San Diego Previously Published Works

### Title

Chromosomal inversion polymorphisms shape human brain morphology.

### Permalink

<https://escholarship.org/uc/item/4nq2h44m>

### Journal

Cell Reports, 42(8)

### Authors

Makowski, Carolina

Zhang, Yanxiao

Qi, Anna

et al.

### Publication Date

2023-08-29

### DOI

10.1016/j.celrep.2023.112896

Peer reviewed



Published in final edited form as:

Cell Rep. 2023 August 29; 42(8): 112896. doi:10.1016/j.celrep.2023.112896.

## Chromosomal inversion polymorphisms shape human brain morphology

Hao Wang<sup>1</sup>, Carolina Makowski<sup>1</sup>, Yanxiao Zhang<sup>2,3,4</sup>, Anna Qi<sup>1</sup>, Tobias Kaufmann<sup>5,6</sup>, Olav B. Smeland<sup>6</sup>, Mark Fiecas<sup>7</sup>, Jian Yang<sup>3,4</sup>, Peter M. Visscher<sup>8</sup>, Chi-Hua Chen<sup>1,9,\*</sup>

<sup>1</sup>Center for Multimodal Imaging and Genetics, University of California San Diego, La Jolla, CA 92093, USA

<sup>2</sup>Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA

<sup>3</sup>School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China

<sup>4</sup>Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China

<sup>5</sup>Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, 72076 Tübingen, Germany

<sup>6</sup>Norwegian Centre for Mental Disorders Research, Oslo University Hospital and University of Oslo, 0450 Oslo, Norway

<sup>7</sup>Division of Biostatistics, University of Minnesota School of Public Health, Minneapolis, MN 55455, USA

<sup>8</sup>Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

<sup>9</sup>Lead contact

### SUMMARY

The impact of chromosomal inversions on human brain morphology remains underexplored. We studied 35 common inversions classified from genotypes of 33,018 adults with European ancestry. The inversions at 2p22.3, 16p11.2, and 17q21.31 reach genome-wide significance, followed by 8p23.1 and 6p21.33, in their association with cortical and subcortical morphology. The 17q21.31, 8p23.1, and 16p11.2 regions comprise the *LRRC37*, *OR7E*, and *NPIP* duplicated gene families. We find the 17q21.31 *MAPT* inversion region, known for harboring neurological risk, to be the most salient locus among common variants for shaping and patterning the cortex. Overall, we observe the inverted orientations decreasing brain size, with the exception that the 2p22.3 inversion is associated with increased subcortical volume and the 8p23.1 inversion is

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\*Correspondence: [chc101@ucsd.edu](mailto:chc101@ucsd.edu).

#### AUTHOR CONTRIBUTIONS

Conceptualization, H.W. and C.-H.C.; methodology, H.W., M.F., J.Y., P.M.V., and C.-H.C.; formal analysis and investigation, H.W. and C.-H.C.; resources, T.K. and O.B.S.; writing – original draft, H.W. and C.-H.C.; writing – review & editing, H.W., C.M., Y.Z., T.K., O.B.S., M.F., J.Y., P.M.V., and C.-H.C.; visualization, H.W. and A.Q.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

#### SUPPLEMENTAL INFORMATION

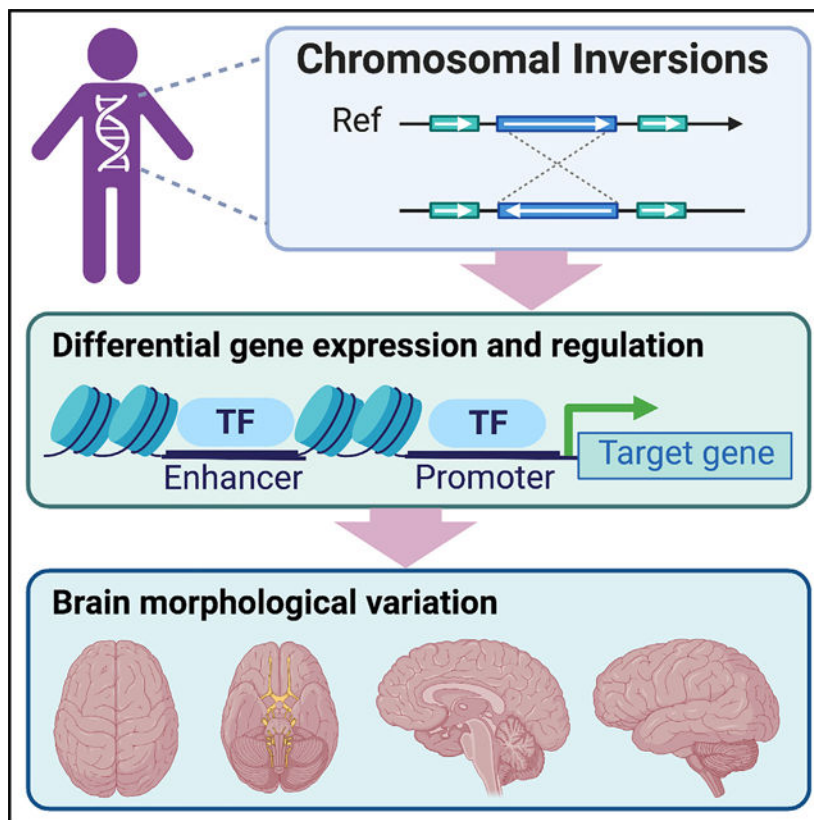
Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2023.112896>.

associated with increased motor cortex. These significant inversions are in the genomic hotspots of neuropsychiatric loci. Our findings are generalizable to 3,472 children and demonstrate inversions as essential genetic variation to understand human brain phenotypes.

## In brief

Wang et al. demonstrate that inversion polymorphisms are located within the genomic hotspots of neuropsychiatric loci and are linked to brain morphology. This study emphasizes that inversion polymorphisms, along with other genetic variants, play a significant role as contributors to the phenotypic diversity observed in brain development and disorders.

## Graphical Abstract



## INTRODUCTION

Polymorphic chromosomal inversions are common structural variations in which a chromosomal segment is in the reversed orientation.<sup>1,2</sup> Recent work identified 729 inversions in humans and an average of 117–156 inversions per human genome.<sup>3–5</sup> The inversions included in this study span from less than 1 Kbps to 4 Mbps, and their reported frequencies may range from about 6% to over 70% in various populations.<sup>6</sup> Significant geographic- and ethnicity-related variation in frequencies has been observed in several inversion regions, such as the 17q21.31 inversion (17q21.31-inv), which is most common in the European population (~20%), particularly in the Mediterranean region (~30%), but

is nearly absent in South and East Asia.<sup>7</sup> Nearly 60 inversions intersect with regions of microdeletion and microduplication syndromes.<sup>4,5</sup> These findings suggest that inversions constitute a substantial source of genetic variation for diseases and normal phenotypes. A small fraction of these inversions can be robustly characterized using genotype-array data through their impact on linkage disequilibrium (LD) patterns, leaving a detectable signature by principal-component analysis in inversion-calling algorithms and allowing for testing associations between inversions and phenotypes.<sup>8,9</sup> In this study, we use the term “inversion” to refer to an inversion region, an inverted variant/allele (haploid), or an inverted genotype (diploid).

Large inversions are often flanked by segmental duplications (SDs) (repeats with 90% sequence identity) or mobile elements (e.g., *LINE1* or *Alu*) at inversion breakpoints.<sup>3,10,11</sup> The hominid lineage has experienced a burst of SDs, which constitute ~7% of the human genome. SDs contribute disproportionately to genetic variation within and between ape species and play an important role on human-specific diseases and traits.<sup>12,13</sup> Human SDs are gene-rich repeats, and gene duplication is recognized as a source of evolutionary innovation.<sup>14</sup> The *LRRC37*, *NPIP* (nuclear pore interacting protein), and *OR7E* gene families located at 17q21.31-inv, 16p11.2-inv, and 8p23.1-inv, respectively, are among the 14 core duplicons previously identified to have expanded in the hominid lineage.<sup>15,16</sup> There is some evidence to suggest that these gene families are associated with increased brain size across primates,<sup>17</sup> although further replication is needed. Inversions suppress recombination leading to extended LD blocks and segregation of haplotype families, which provides a unique lens into the genomic impact of these complex structural variation regions on human phenotypic variation.

Although human and great ape species have ~98% sequence similarity, inversions are the most common chromosomal rearrangements that differentiate the genomes of the two species.<sup>18</sup> In addition to differences between species, inversions are also important for phenotypic variations in humans, and their impact on brain-related phenotypes remains underexplored.<sup>9,19</sup> Previous evidence showed an association between the 8p23.1-inv and neuroticism.<sup>9,20</sup> Without using inversion calling, genetic variants at 17q21-inv were uncovered by genome-wide association studies (GWASs) of cortical and subcortical structures<sup>21,22</sup> or by haplotype analysis of the *MAPT/tau* gene at 17q21.31 linked to risk of various neurodegenerative disorders.<sup>23–27</sup>

Compared to pathogenic copy-number variants (CNVs), common inversion polymorphisms are typically a balanced rearrangement, rather than a loss, of genomic materials and are more prevalent in the population. This facilitates the investigation of inversions within population-based databases, such as the UK Biobank (UKB) and Adolescent Brain Cognitive Development (ABCD) study. We previously observed genetic variants at the 8p23.1-inv and 17q21.31-inv associated with cortical morphology.<sup>28</sup> Here, we present a comprehensive association analysis on 17 common inversions implemented in an inversion-calling method, *scoreInvHap*,<sup>6</sup> to study their links to morphometrics derived from neuroimaging. Furthermore, the new high-throughput genotyping method characterizes ~45 common inversions, some of which can be tagged by single-nucleotide polymorphisms (SNPs).<sup>19</sup> We utilize these tag SNPs to expand the list of the inversions in this study. For

significant associations, we investigated expression patterns of genes at the inversion regions across developmental time points from the BrainSpan Atlas.<sup>29</sup> We further focus on the most significant association, namely 17q21.31-inv, to uncover its genetic impact to study how the 17q21.31-inv could lead to alterations of gene expression, methylation, and isoform splicing.<sup>30,31</sup>

## RESULTS

### SNPs at 17q21.31-inv play a key role in cortical morphology

In our previous work, we identified the salient latent factors underlying 24 regions of cortical surface area and thickness parcellated according to our genetically informed atlas (Table S1).<sup>32,33</sup> This was conducted with a two-factor genomic structural equation model (SEM), as informed by the highest cumulative variance explained and model fit indices (see genomic structural equation modeling in STAR Methods), using autosomal variants of 34,720 adults from the UKB cohort of European ancestry.<sup>28</sup> These genetic latent factors represented pleiotropic effects on morphometric variations across the cortex and pointed to 17q21.31-inv as a key contributor to individual size variations on cortical morphology. We showed a convergence of common SNPs and inversions where 1,627 SNPs mapped to 17q21.31-inv were found to be significantly associated with the above-mentioned latent factors (Figure 1). These findings prompted us to estimate the effect sizes of multiple inversions to elucidate their influence on the human brain.

### Inversion calling yields expected allele frequencies

To investigate the impact of inversion polymorphisms on cortical morphology, we included the variants of both autosomes and the X chromosome and performed inversion calling in our sample of 34,720 adults (Table S2). Out of the 21 inversions that can be classified by *scoreInvHap*, we excluded three that were not yet experimentally validated according to the Human Polymorphic Inversion database<sup>34</sup> or the Human Genome Structural Variation Consortium (HGSVC) (Table S3)<sup>4</sup> and one with low result quality (only 73.7% passed the two quality control [QC] steps) (Table S3; see STAR Methods). After removing 859 related subjects (genetic relatedness > 0.1), 33,861 participants were included in a multiple linear regression analysis (Table S3). The *scoreInvHap* method identifies individuals as having one of three different haplotype genotypes based on inverted (I) or non-I (NI) state, I/I, N/I, and N/N, and coded as 2, 1, and 0 by the dose of the I allele. All of our inversion genotypes showed good correspondence to the expected allele frequencies in the reference genome (Table S3). To further test the accuracy of our inversion calling, we classified the 17q21.31-inv status of 17 available HGSVC subjects by *scoreInvHap* and found that they were highly consistent with those reported by the HGSVC based on a multiplatform sequencing approach (Table S3). We also included 26 additional inversions with their genotypes called based on their reported good tag ( $LD\ r^2 > 0.9$ ) SNPs in the European population.<sup>19</sup> Eight inversions (HsInv0004 [1q31.3], HsInv0040 [2q22.1], HsInv0045 [21q21.3], HsInv0058 [6p21.33], HsInv1053 [7q11.22], HsInv 0092 [6q23.1], HsInv0501 [8p23.1], and HsInv0573 [17q21.31]) were also included in *scoreInvHap*, and the results from *scoreInvHap* and based on tag SNPs were highly consistent with accuracies of 86.69%–99.98% (Table S3).

### Conspicuous effect of the 17q21.31-inv on global brain morphometry

The strongest associations emerged with the 17q21.31-inv with decreased total surface area and modestly increased cortical thickness ( $p < 2.3 \times 10^{-3}$ , lowest  $p = 6.3 \times 10^{-65}$ , where  $p < 0.05/t_e$ , with  $t_e = 22$  being the effective number of independent phenotypes<sup>35,36</sup>). The effects of the 17q21.31-inv are relatively large compared to GWAS signals of common polygenic phenotypes (~0.01% variance explained),<sup>37</sup> where we found the largest association between this inversion and a decreasing anteromedial temporal area (0.86% variance explained). Furthermore, the 2p22.3-inv, 8p23.1-inv, and 16p11.2-inv were related to multiple cortical phenotypes, among which the 16p11.2-inv was also related to reduced total surface area ( $p = 2.1 \times 10^{-4}$ ; Figure 2; Table S4).

We further studied 20 subcortical and intracranial volumes based on the *aseg* atlas in FreeSurfer (Table S1).<sup>38</sup> Again, we observed strong associations between the 17q21.31-inv and reduced intracranial volume ( $p = 3.1 \times 10^{-26}$ ,  $t_e = 15$ ; Figure 2; Table S4). The same 2p22.3-inv, 6p21.33-inv, 8p23.1-inv, and 16p11.2-inv were linked to one or more subcortical volumes (lowest  $p = 2.2 \times 10^{-8}$ ,  $2.9 \times 10^{-3}$ ,  $4.2 \times 10^{-4}$ , and  $3.5 \times 10^{-15}$ , respectively). The effects of 16p11.2-inv on the decreases in caudate and putamen volume explained ~0.1%–0.2% of variance.

To assess how well our results generalize to an independent sample, we performed an additional analysis to generalize the above inversion association analysis to a relatively small cohort of 3,472 children from the ABCD study. We observed significant consistency between the results from the UKB discovery and ABCD generalization cohorts, with inversion effect correlation estimates of 0.82 ( $p = 1.4 \times 10^{-11}$ ) for cortical area phenotypes and 0.83 ( $p < 2.2 \times 10^{-16}$ ) for subcortical phenotypes, accounting for errors in estimated inversion effects, as well as sign concordance rates of 0.88 ( $p = 2.5 \times 10^{-7}$ ) for cortical area phenotypes and 0.79 ( $p = 1.1 \times 10^{-6}$ ) for subcortical phenotypes. In contrast, we found no significant beta correlations between UKB and ABCD for the cortical thickness phenotypes. Interestingly, while the 17q21.31-inv was associated with increased thickness in UKB, this association was absent in ABCD (Table S4; Figures S2 and S3). The 17q21.31-inv has been previously identified as a protective variant for neurodegenerative diseases, and it is possible that carriers of this inversion experience less thickness decline over their lifespan. However, the effects of inversions on thickness are not as strong as those on surface area, which may partially contribute to the observed lack of correlation. More in-depth studies are needed to understand how this genetic variant influences cortical thickness in different age groups.

### Autosomal and X chromosomal inversions associated with regional brain morphology

To further assess the impacts of inversions on regional measures, we repeated the above association analyses after adjusting for global measures by preresidualization (e.g., total surface area or mean cortical thickness). This analysis aimed to demonstrate the contributions of inversions to regional variations that have disproportionately larger effects relative to the entire brain. Thus, a positive association with frontal surface area would represent a larger increase relative to global expansion, as was the case for 8p23.1-inv (lowest  $p = 1.1 \times 10^{-7}$ ), or a smaller decrease relative to global reduction, as was the case for 17q21.31-inv (lowest  $p = 4.4 \times 10^{-11}$ ). Overall, significant associations of 2p22.3-

inv, 6p21.33-inv, 8p23.1-inv, 11q13.2-inv, 16p11.2-inv, 17q21.31-inv, and Xq13.2-inv were demonstrated with cortical regions ( $t_e = 22$ ; Figures 3, S1, and S3; Table S4), and 2q22.3-inv, 8p23.1-inv, 17q21.31-inv, and 16p11.2-inv were linked to subcortical structures ( $t_e = 15$ ; Figures 3, S1, and S3; Table S4).

### Distinct spatial gradients of inversion-morphology association patterns across the cortex

We observed distinct spatial distribution patterns of significant inversion-morphology associations across the cortex, with their effect sizes following an anterior-posterior (A-P), ventral-dorsal, or medial-lateral gradient. For example, the 17q21.31-inv was associated with graded reductions in surface area along the A-P axis, with the greatest reduction in occipital and anteromedial temporal surface areas, accompanied by modest graded increases in thickness, with the largest increase in the orbitofrontal cortex (Figure 2). The Xq13.2-inv was related to increased rostral but decreased dorsal cortical thickness after adjusting for global measures (Figure S1).

### Convergent effects of the inversions and SNPs at inversion regions

We performed conditional analysis using GCTA-COJO to evaluate the significant inversion regions and found the association signals of the inversions and the SNPs within these genomic regions to be highly overlapping. These significant inversions were well tagged by SNPs. One extreme case is 17q21.31-inv, where at least ~100 SNPs are found to be in perfect LD with the inversion and any of them can tag the inversion as indicated previously.<sup>7,39</sup> This is in line with our previous GWAS on cortical morphology in which we identified significant SNPs that mapped to genes located within the 17q21.31-inv, 8p23.1-inv, and 16p11.2-inv regions.<sup>28</sup> Thus, although inversions do not add more variance explained than those already discovered in GWASs, identifying inversion associations could aim at understanding the underlying genetic mechanisms of these GWAS signals and at designing more effective follow-up experiments given that inversions alter genomic structure and could disrupt regulatory architecture more so than individual SNPs.<sup>31,40</sup>

### Developmental changes in the expression of inversion-related genes

Next, we explored the importance of genes within inversions on brain structure through their expression patterns from prenatal life to adulthood (Figures 4 and S4). We obtained gene expression data from the BrainSpan Developmental Transcriptome in the form of reads per kilobase million (RPKM) values.<sup>41</sup> Notably, we observed substantially high mean expression levels for genes located in the 17q21.31-inv region (Figure 4A). Upon further analysis, we found that *MAPT* and *NSF* were primary drivers of this high expression (Figure 4B). These genes are known to play a critical role in maintaining normal synaptic function and have been linked to neurodegenerative disorders, autism, and anxiety.<sup>42–44</sup> Specifically, *MAPT* was highly expressed at the mid-late prenatal stage, which corresponds to the phases of neurogenesis, neuronal migration, and early synapto- and gliogenesis.<sup>45</sup>

### Altered gene expression and regulatory landscape with the 17q21.31-inv

Using differential expression (DE) analysis, we identified significant differences in gene expression levels between individuals who carry the 17q21.31-inv and those who do not.

Similar to prior work,<sup>46</sup> we also observed downregulated expression of genes for the *MAPT* H2 haplotype at the upstream breakpoint (e.g., a pseudogene: *LRRC37A4P*) together with upregulated expression of genes within and downstream of the inversion (Figure 5, Table S5). Our expression quantitative trait locus (eQTL) analysis further confirmed these findings (Figure 5, Table S5). In addition, we demonstrated that the inversion acts as a molecular QTL for numerous genes in the inversion region, such as *MAPT*, *KANSL1*, *CRHR1*, and *WNT3*, along with their corresponding antisense or long non-coding RNAs (lncRNAs) (e.g., *MAPT-IT1*, *KANSL1-AS1*, *CRHR1-IT1*). These non-coding RNAs are known to have regulatory functions.<sup>46,47</sup> Importantly, several of these genes have well-established roles in neurodevelopmental and physiological functions, highlighting the potential impact of the inversion on complex biological processes.<sup>46</sup> Our study identified several copies of *LRRC37A*, a family of core duplicon genes that have undergone expansion in primates and have acquired novel promoters, enabling their expression in multiple tissues, are among the genes regulated by the inversion. The exact function of these genes remains to be elucidated, but presumably, they play a role in cellular migration, chemotaxis, and immunity.<sup>48,49</sup> Similar to our findings, evidence from several studies indicates that 17q21.31-inv is significantly associated with the expression levels of *LRRC37A4P*, A2, and A1 in adult samples<sup>9,46,48,50</sup> and fetal brains.<sup>51,52</sup> In short, these findings suggest that the inversion has a significant impact on gene regulation. 17q.21.31-inv with balanced rearrangement does not drastically change gene dosage in the same way as do rare pathogenic CNVs but rather involves changes in a remarkably diverse range of gene regulatory activity with downstream impact on normative variation in brain structure and function. Note that there are common CNVs at the chromosomal 17 inversion breakpoints, which could partially explain the effects of the inversion.

## DISCUSSION

This study provides evidence of common inversions associated with brain morphometry. We found that the 2p22.3-inv, 16p11.2-inv, and 17q21.31-inv reached genome-wide significance ( $p < 5 \times 10^{-8}$ ). Eight additional inversions reached significance after the Bonferroni correction, among which 6p21.33 and 8p23.1 are nearly at genome-wide significance ( $p = 2.1 \times 10^{-6}$  and  $1.1 \times 10^{-7}$ , respectively). These eight inversions are 6p21.33, 8p23.1, 7p14.1, 6q15, 1q32.1, 4q22.1, 14q23.1, and Xq13.2. Furthermore, our findings were generalizable to 3,472 children from the ABCD study, indicating their replicability. We observed a distinct distribution pattern of significant inversion-morphology associations where strength of association varies across the brain with orderly spatial gradients along core developmental axes. Specifically, the 17q21.31-inv was associated with global surface area reductions, especially posteriorly, with the greatest reduction in occipital and anteromedial temporal areas, accompanied by a reversed pattern of modest increases in cortical thickness prominently in the orbitofrontal cortex. Many of the above-mentioned cytobands are hotspots of structural variations and also encompass pathogenic CNVs linked to neuropsychiatric conditions.<sup>53–55</sup> Inversions appear to have relatively smaller effects than pathogenic CNVs, perhaps because of their balanced rearrangement and thus high prevalence in the normal population.



The 17q21.31-inv region is one of the most dynamic and complex regions of chromosomal rearrangement in the human genome, spanning up to 1.08–1.49 Mbp.<sup>56</sup> It comprises two major haplotype families, the NI H1 and the I H2, which differ in orientation and gene copy numbers (also known as *MAPT* haplotypes).<sup>57</sup> We observed a conspicuous correlation of 17q21.31-inv status with multiple morphological phenotypes of the brain, indicating the inversion's crucial role in cerebral development and maturation. Previous GWASs of brain MRI phenotypes identified associated variants at the 17q21.31-inv region without determining the direction of inversion effects given the absence of inversion calling. However, the inversion effect can be determined because the 17q21.31-inv is perfectly tagged by SNPs. Interrogating the tag SNPs in summary data from previous studies revealed an association between the H2 haplotype and decreased global total surface area and subcortical and intracranial volumes,<sup>21,22,46,58,59</sup> which is consistent with the direction of effects we reported. Here, we present detailed analyses of individual brain regions and multiple inversions, revealing intriguing spatial gradations of inversion-associated brain maps. For instance, the H2 haplotype is linked to the greatest regional reductions in cortical area within the visual occipital lobe in conjunction with the greatest thickness increases in the orbitofrontal cortex, a region important for reward processing and decision-making.<sup>60</sup> A number of genes in this inversion are related to neurodevelopment such as *MAPT*, a gene highly expressed during the prenatal stage (Figure 4). The 17q21.31 I allele frequency varies significantly among populations.<sup>7</sup> This inversion was shown to be the most significant polymorphism for shaping brain morphology in European populations, but it is rare in Asian populations. Therefore, the results cannot simply be extrapolated across populations; thus, there is a need for genetic studies on other populations.

Compared with 17q21.31-inv, which has dramatic effects on both cortical and subcortical morphometry, we observed opposing effects of 2p22.3-inv on cortical thickness and subcortical volume that were negligible on surface area. The expression of local gene *RASGRP3* was found to be highest in the white matter, followed by various subcortical structures comparing all brain regions,<sup>61</sup> and highest in oligodendrocytes among all cell types.<sup>62</sup> This is in contrast to the main local genes of 17q21.31-inv, such as *MAPT*, the expression of which was demonstrated to be highest in the cerebral cortex and within neurons.<sup>61,62</sup>

In contrast to the 17q21.31-inv, we noted that the 8p23.1-inv was associated with increased cortical area and reduced thickness, especially frontal and superior temporal areas and perisylvian thickness comprising language-related cortical regions (Figure 2). The 8p23.1 region is the largest known inversion site in humans (~4.5 Mbp) and captures numerous associated signals of diseases and traits,<sup>63</sup> with the allele frequency of the inversion ranging from 60% in Africans to 20% in Asians.<sup>64</sup> The inversion is bordered by the *microcephalin* (*MCPHI*) gene, which is involved in the development of microcephaly and predominantly affects the frontal lobes.<sup>65</sup>

We also found notable associations between the 16p11.2-inv and brain morphometry, particularly the subcortical volumes. The 16p11.2-inv contains *NPIP*B duplicated genes. The *NPIP* gene family showed signatures of positive selection as one of the most rapidly evolving gene families.<sup>66</sup> Most chromosomal rearrangements are disease causing and are

under strong negative selection, but several human expanded gene families (core duplicons) have been shown to be positively selected, potentially harboring genes with selective advantages.<sup>17,67</sup> *NPIP* functions are unclear but may be immune related.<sup>17</sup> Prior research has shown 16p11.2 CNVs to be associated with global brain volumes (e.g., duplication carriers tend to have smaller volumes), including total intracranial and gray and white matter volumes.<sup>68–70</sup> Regional variations associated with 16p11.2 CNVs have also been reported in cortical (e.g., insula/visual cortex) and subcortical structures (e.g., basal ganglia, caudate, and hippocampus),<sup>68–70</sup> consistent with some of our findings with inversions in this region. These rare pathogenic CNVs show a larger effect than the effect of our inversion in the same 16p11.2 region. Interestingly, the distal CNVs and the inversion (also located distally at 16p11.2) showed the largest associations with reductions of basal ganglia volume among the brain structures investigated.<sup>71</sup>

The 17q21.31, 8p23.1, 16p11.2, and 6p21.33 regions associated with brain morphology here were also previously reported to be linked to multiple neuropsychiatric disorders. The 17q21.31-inv or its local genes have been associated with neurodegenerative disorders.<sup>24,25,72</sup> 8p23.1-inv was associated with increased risk of autism and neuroticism.<sup>20,73,74</sup> The *MSRA* gene located at 8p23.1 has been linked to bipolar disorder, schizophrenia, autism, and alcohol use disorder.<sup>75</sup> CNVs at 16p11.2 showed strong association with autism.<sup>76,77</sup> The 6p21.33-inv is located within the major histocompatibility complex, which has been linked to schizophrenia (e.g., *C4A*) and mood disorders.<sup>75,78</sup>

Our multiomics analysis for 17q21.31-inv showed intriguing patterns of inversion-associated transcriptomic and epigenomic changes. Similar to prior work,<sup>46</sup> we also observed downregulated expression of genes for the H2 haplotype at the upstream breakpoint (e.g., a pseudogene: *LRRC37A4P*) together with upregulated expression of genes within and downstream of the inversion. Here, we showed that the inversion is the molecular QTL for many different genes within the inversion region, including several significant genes underlying neurodevelopmental and physiological functions (e.g., *MAPT*, *KANSL1*, *CRHR1*, *WNT3*) and their corresponding antisense or lncRNAs (e.g., *MAPT-IT1*, *KANSL1-AS1*, *CRHR1-IT1*). These antisense or lncRNAs could regulate gene expression of their protein-coding genes and isoforms.<sup>46,47</sup> Further, we observed and replicated the significance and effect direction of several *LRRC37A* copies: core duplicon genes acquired novel promoters and uniquely expanded in primates with unclear gene function, presumably cellular migration, chemotaxis, and immunity.<sup>48,49</sup> *LRRC37A4P*, *A2*, and *A1* have been repeatedly shown to be associated with 17q21.31-inv<sup>9,46,48,50</sup> and have been demonstrated in fetal brains,<sup>51,52</sup> though with the cautioning that probes can bind to more than one target genes, especially using earlier array technology.<sup>50</sup> In short, the 17q.21.31-inv with balanced rearrangement does not drastically change gene dosage in the same way as do pathogenic CNVs but rather involves changes in a remarkably diverse range of gene regulatory activity, with downstream impact on normative variation in brain structure and function.

In conclusion, this study demonstrates that inversion polymorphisms are in the genomic hotspots of neuropsychiatric loci and are associated with brain morphology. In addition to other genetic variants, we underscore that inversion polymorphisms are important contributors to phenotypic diversity pertaining to brain development and disorders.

## Limitations of the study

A key limitation of the present study was ethnicity, owing to the inversion-calling tool *scoreInvHap* that relies on an LD reference panel built on the European population.<sup>6</sup> Also, *scoreInvHap* predictions are based on limited genotypes, leading to variable reliability of calling results for the inversions with more genotypes. More neuroimaging data are needed from non-European participants for the development of LD reference panels and inversion-calling algorithms that can be applied to other populations. In addition, the array-based approach that this study employed was only capable of detecting larger inversions of specific LD patterns. These issues may be addressed in the future by using other tools, especially novel sequencing-based methods.<sup>79</sup>

## STAR★METHODS

### RESOURCE AVAILABILITY

**Lead contact**—Further information and requests should be directed to and will be fulfilled by the lead contact, Chi-Hua Chen (chc101@ucsd.edu).

**Materials availability**—This study did not generate new unique reagents and physical samples. All computational results of this study are described below in the Data and code availability section.

#### Data and code availability

- The individual-level raw data used in this study can be obtained from UK Biobank (<https://www.ukbiobank.ac.uk/>) under accession number UKB: 27412 and Adolescent Brain Cognitive Development (ABCD) study (<https://abcdstudy.org>). Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (<https://www.ppmi-info.org/access-data-specimens/download-data/>). Please refer to [ppmi-info.org](https://ppmi-info.org) for up-to-date information on the study. We made use of publicly available software and tools.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

### EXPERIMENTAL MODEL AND SUBJECT PARTICIPANT DETAILS

**UKB:** For discovery, genomic, imaging and demographic data were extracted from the UK Biobank (UKB) population cohort, under accession number 27412.<sup>81–84</sup> The composition, set-up, and data gathering protocols of the UKB have been extensively described elsewhere.<sup>81–83</sup> Quality control (QC) of imaging and demographic data were detailed in our previous work.<sup>28</sup> In brief, individuals with ICD10 diagnosis of a neurological or mental disorder were excluded, as well as those with bad structural scan quality based on their age- and sex-adjusted Euler numbers.<sup>85</sup>

**ABCD:** We conducted an additional analysis to see whether results generalize to a neurodevelopmental cohort. Genomic and phenotype data of children were collected from the Adolescent Brain Cognitive Development (ABCD) study release 3.0 ([abcdstudy.org](http://abcdstudy.org)).<sup>86,87</sup> The database and QC steps were the same as described in our previous work.<sup>28</sup> In short, those with difficulties to communicate in English, comply with the protocol, or complete a baseline MRI scan were excluded.

In this study, we only included those of European ancestry, because our inversion calling tool *scoreInvHap* relies on its reference haplotypes built from European individuals in the 1000 Genomes (1KG) Project,<sup>6</sup> and there is a low prevalence of certain inversion genotypes (e.g., 17q21.31-inv) in non-European populations.<sup>39</sup> Our final sample included 34,720 participants from UKB (age range: 45.13–81.83 years, male-female ratio of 0.9), and 3,472 children from ABCD (age range: 8.92–11.00 years, male-female ratio of 1.1) (Table S2).

## METHOD DETAILS

**Genotype data**—We used UKB Version 3 release and ABCD release 3.0 of imputed genotype data and removed individuals with more than 10% missingness, as well as SNPs with more than 5% missingness, failing the Hardy-Weinberg equilibrium test at  $p = 1e-6$  or with minor allele frequencies (MAF) below 0.01.<sup>28</sup> In this study, we also added imputed X chromosomal data with the same QC steps.

**Inversion calling**—We included 18 out of the 21 human chromosomal inversions (Table S3) that can be studied by the development version of *scoreInvHap*, a tool to classify inversion status from individual SNP data,<sup>6</sup> based on a selection criterion that these inversions are classified as experimentally validated inversions in the Human Polymorphic Inversion Database (<http://invfestdb.uab.cat/>)<sup>34</sup> or present in the list of human inversions recently categorized from the Human Genome Structural Variation Consortium (HGSVC).<sup>4</sup>

The inversion regions of the target genotype data were imputed following the recommended *imputeInversion* pipeline to achieve desirable numbers of SNPs for dense genotype information within the inversion regions. Imputed genotype data were then analyzed by *scoreInvHap* implemented in R 4.0.2. A small percentage of subjects failed inversion calling and did not have inversion genotypes (<1%), and therefore were removed from subsequent analyses.

The built-in inversion genotype classifier was derived from 503 reference individuals with European ancestry in the 1KG. Classification is done for each inversion region separately. The classifier utilizes the frequency,  $f(s)$ , of an SNP genotype (0,1 or 2) present in a haplotype-genotype group (e.g., inversion-genotype groups [NN, NI, or II]), which was pre-calculated in the reference genomes. The *scoreInvHap* then classifies the haplotype-genotype group  $K$  of a new individual based on similarity scores ( $H_k$ ), calculated as the weighted sum of the above frequencies across all the SNPs within the inversion region ( $SNP_1 \dots SNP_L$ ), where the weights are constructed by the maximum LD between  $SNP_i$  and the haplotype groups ( $\rho_i^2$ ) and by the certainty (posterior probability) of the imputed  $SNP_i$

genotype ( $p_i(t)$ ). The inferred haplotype-genotype group of each individual is assigned to the one with the maximum similarity score among all the groups, denoting the biallelic inverted (I) or non-inverted (N) status as NN, NI or II (see details in Ruiz-Arenas et al., 2019).<sup>6</sup>

$$H_k = \frac{\sum_{i=1}^L \sum_{s_i=0,1,2} p_i(t) \cdot f_{ki}(s_i) \cdot \rho_i^2}{\sum_{i=1}^L \rho_i^2}$$

We expanded the list of inversions based on tag SNPs ( $LD r^2 > 0.9$ ) in the European population that were reported by Giner-Delgado and colleagues.<sup>19</sup> Genotypes of 18 additional inversions were called based on inverted (O2) alleles.

**X-linked dosage compensation**—For the Xq13.2-inv, we took account of X-linked dosage compensation (DC) due to random inactivation of one of the X chromosomes in female cells, to balance allele dosage differences in X-linked genes between sexes.<sup>88</sup> Owing to limited sample size and study power, we could not explicitly determine the presence of DC for all of the phenotypes. Therefore, we conducted association analyses on Xq13.2-inv with both full DC and no DC, coding the inversion genotypes in male subjects as {0,2} and {0,1}, respectively. The results were largely similar under the two extreme conditions from our sensitivity analysis.

**MRI data and atlases**—For UKB, T1-weighted MRI scans were collected from three scanning sites throughout the United Kingdom, all on identical Siemens Skyra 3T scanners.<sup>83</sup> The standard “recon-all -all” processing pipeline of Freesurfer v5.3 was applied to perform automated surface-based morphometry segmentation.<sup>38</sup> For ABCD, MRI scans were performed and harmonized across 21 sites and three scanner manufacturers (Siemens Prisma, General Electric 650, and Phillips).<sup>89</sup> Details on image processing were detailed in previous publications.<sup>90</sup>

For cortical phenotypes, we adopted two genetically-informed atlases, including 12 regions for surface area and 12 for cortical thickness, and 2 global measures of total surface area and mean thickness (Table S1). These atlases have been previously developed by our group, using a data-driven fuzzy clustering technique to identify parcels of the human cortex that are maximally genetically correlated based on the MRI scans of over 400 twins.<sup>32,33</sup> The subcortical structures were parcellated based on the widely used *aseg* atlas,<sup>38</sup> including volumes of 20 regions and one global measure of estimated intracranial volume (Table S1). We combined measures of each phenotype across both hemispheres, in view of largely bilateral symmetry of genetic patterning demonstrated in our previous work.<sup>28,91</sup>

## QUANTIFICATION AND STATISTICAL ANALYSIS

**Genomic structural equation modeling**—Previously we explored the joint genetic architecture underlying cortical phenotypes.<sup>28</sup> We performed genomic structural equation modeling (SEM) on the GWAS summary statistic data of the 12 area and 12 thickness phenotypes separately, which estimates the genetic covariance structure of complex traits. The two-stage analysis includes: 1) estimating the empirical genetic covariance matrix and its associated sampling covariance matrix, and 2) structural equation modeling with user-

defined parameters to minimize the discrepancy between the model-informed and empirical genetic covariance matrices.<sup>92</sup> We did not need to pre-residualize for global effects as they were reflected in the common latent factors.

We used exploratory factor analysis to determine the specifications in the models. Two-factor models were found to be optimal with both the highest cumulative variance explained (0.665 for areas and 0.733 for thicknesses) and lowest number of latent factors. The confirmatory factor analysis demonstrated good model fit for both the area ( $\chi^2(48) = 1033.55$ , Akaike Information Criterion (AIC) = 1093.55, comparative fit index (CFI) = 0.99, standardized root-mean-square residual (SRMR) = 0.028) and thickness ( $\chi^2(48) = 1477.29$ , AIC = 1541.29, CFI = 0.98, SRMR = 0.042) phenotypes, and outperformed the common factor models for area ( $\chi^2(54) = 1912.19$ , AIC = 1960.19, CFI = 0.98, SRMR = 0.041) and thickness ( $\chi^2(54) = 4603.33$ , AIC = 4651.33, CFI = 0.95, SRMR = 0.074). The analysis was computed by the R package *genomicSEM*.<sup>92</sup>

**Quality control for inversion calling**—The QC steps of the *ImputeInversion* pipeline after inversion calling involve two steps. We used the default QC parameters to evaluate the reliability of inversion inference for each individual, including the SNP call rates and differences of similarity scores. First, the SNP call rate reports the number of SNPs used in the computation. It reflects the SNP coverage within the inversion regions. *scoreInvHap* recommends at least 15 SNPs for each calling.<sup>6</sup> Second, a similarity score quantifies how closely the target SNP set resembles the reference genotype of a specific inversion. Classification is good when the highest score is close to 1 and the remaining scores are small, which indicates that the SNP set in the individual is almost identical to one of the reference genotypes and different from the rest, and can be reflected by the differences between the top and second highest similarity scores. The default threshold of difference score is 0.1.<sup>6</sup>

Desirable numbers of SNPs for inversion calling were achieved for all subjects after imputation by the *imputeInversion* pipeline for the first QC criterion. The principal reason for poor inversion calling was the second criterion of similarity scores, indicating that the genotype of samples was more different from the reference haplotype than expected, and/or there were more than one inversion genotypes that the samples resembled. Several inversions have more than two haplotypes<sup>6</sup> and tend to have lower differences in similarity scores, such as 3q26.1-inv and 7p11.2-inv. Based on the second QC criterion, we excluded 3q26.1-inv, of which less than 75% of calling results passed both QC steps (Table S3), resulting in 17 experimentally validated inversions with good calling quality for the subsequent association analyses.

**Inversion calling accuracy test**—To validate if inversion genotyping from our inversion calling yields the same genotyping as that from a multi-platform sequencing approach, we downloaded subjects from the International Genome Sample Resource (IGSR) built by the 1KG Project, particularly the data contributed by the HGSC.<sup>93</sup> We performed inversion calling to classify the 17q21.31-inv status of 17 subjects from the 1KG participants with Hi-C data, using the *scoreInvHap* pipeline. We compared the results with those from the HGSC data (<https://www.internationalgenome.org/data-portal/sample>),<sup>94</sup> which used

a multi-platform approach to generate a comprehensive and orthogonally validated set of inversions. They integrated Strand-seq, Bionano optical mapping, and Phased Assembly Variant (PAV)-based variant discovery and identified on average 117 inversions per sample and 316 non-redundant inversions across 32 reference samples from diverse human populations (Table S3).<sup>4</sup> Although the *scoreInvHap* pipeline is primarily for European data, all of the 17 tested subjects were concordant with the HGSVC inversion genotyping.

We also compared the inversion calling results of five inversions (HsInv0004 (1q31.3), HsInv0040 (2q22.1), HsInv0045 (21q21.3), HsInv0058 (6p21.33), HsInv1053 (7q11.22)) that were included in *scoreInvHap* and with perfect tagging SNPs reported in the European population,<sup>19</sup> and observed high consistencies between the two methods (accuracy range: 99.39–99.98%, Table S3).

**Identification of inversion-tag SNPs**—To identify SNPs correlated with the inversions, we conducted GWAS with the inversion genotypes as the outcome variables in the UKB data using fastGWA linear regression models.<sup>95</sup> The Manhattan plots demonstrated that the most significant hits were indeed at the inversion regions as expected (Figure S5). The most significant association signals range from  $p = 1.29e-303$  to  $p = 4.94e-324$  across the inversions. There were no GWAS significant loci (i.e.,  $p < 5e-8$ ) outside of the inversion regions, suggesting that there are no existing strong inter- or intra-chromosomal interactions between inversions and SNPs that are outside inversion regions. This set of analyses also confirms the 17q21.21-inv marker SNPs identified previously,<sup>7,39</sup> given that these marker SNPs were also among the top loci associated with the 17q21.31-inv.

**Conditional and joint analysis**—To determine whether the identified morphology-inversion associations were driven by inversions or the SNPs which the inversions were in LD with, we used conditional and joint (COJO) analysis implemented in Genome-wide Complex Trait Analysis (GCTA),<sup>96</sup> to select LD-independent SNPs associated with each inversion based on the inversion GWAS described above. Using a stepwise model selection procedure, four SNPs for 2p22.3-inv, ten for 8p21.3-inv, five for 16p11.2-inv and one for 17q21.31-inv were selected. The numbers of major alleles of the COJO-selected SNPs were included as covariates in linear models for each inversion-morphology association analysis where each brain phenotype is the dependent variable and each inversion is the independent variable. We observed that nearly all associations were no longer statistically significant ( $p < 5.68e-4$  for cortical or  $p < 8.33e-4$  for subcortical results), suggesting that the association signals between the COJO-selected SNPs and the inversions were overlapping and the COJO-selected SNPs tagged the inversions.

To evaluate the difference between the genetic effects of the COJO-selected SNPs and inversions, we computed R squared derived from linear regression models where the outcome was the brain phenotype of interest, and the independent variables were the inversions of interest or their COJO-selected SNPs. We observed mostly consistent variance explained by inversions and COJO-selected SNPs in the models, indicating their similar genetic effects.

**Inversion association analysis**—For UKB discovery, prior to inversion association analysis, we regressed out age, sex, scanner site, a proxy of scan quality (FreeSurfer’s Euler number),<sup>85</sup> and the first ten genetic principal components from each morphometric measure. We also controlled whether or not the individual had a brain diagnosis based on ICD10 diagnostic information collected by the UK Biobank. Similarly for ABCD replication, we regressed out age, sex and scanner site, and removed the imaging data that didn’t pass FreeSurfer QC because Euler numbers were not available. Subsequently, we applied a rank-based inverse normal transformation to the residuals of each measure, ensuring normally distributed input. For the regional measures, the corresponding global measures were also regressed out (i.e., total surface area, mean cortical thickness and intracranial volume for area, thickness and subcortical phenotypes, respectively). This was done to ensure that observed effects were specific to the region of interest, rather than the association with the global measures.

Linear multiple regression models were used to study the associations of inversion genotypes with brain morphometric measures and fluid intelligence scores. The effect of an inversion was considered as additive depending on the number of inverted alleles. We included all 17 inversions as independent variables for analysis of each phenotype.

We removed related individuals prior to the association testing. Using GCTA,<sup>96</sup> we calculated the pairwise genetic relationship matrix (GRM) based on genome-wide autosomal variants, and removed one related individual from pairs ( $N = 859$  for UKB, and  $N = 1,850$  for ABCD) with an estimated GRM greater than 0.1, which indicates relatedness closer than third cousins. We also removed subjects with failed inversion calls, resulting in 33,018 UKB and 3,472 ABCD participants.

To consider potential correlation between phenotypes, we applied matrix spectral decomposition (matSpD) to determine the effective number of independent phenotypes ( $t_e$ ),<sup>36</sup> using correlation matrices of cortical and subcortical measures. Statistical significance was then defined by Bonferroni correction for multiple comparisons ( $p < 0.05/t_e$ ).

**Generalization analyses**—We assessed the correspondence between the UKB and ABCD results by computing correlation coefficients between the beta estimates of all nominally significant ( $p < 0.05$ ) inversion identified in the inversion association analysis in UKB. We also used binomial tests to evaluate the consistency of effect directions between UKB and ABCD results by calculating beta sign concordance rate. P-values  $< 0.05$  indicate significant effect size correlations and consistency in the direction of associations between the discovery and replication results.<sup>28</sup> Further details of these approaches to assess replicability can be found elsewhere.<sup>28,97,98</sup> Note that although using the approaches to test replicability, we use the term ‘generalization’ rather than ‘replication’ between UKB and ABCD to present the results, given that the two cohorts are not in the same age range.<sup>28</sup>

**BrainSpan Developmental Transcriptome**—To explore developmental changes in expression of inversion-related genes, we listed genes at the studied inversion regions based on positions in the UCSC Genome Browser (<https://genome.ucsc.edu/>). The expression levels in RPKM (reads per kilobase of transcript per million reads mapped) of listed genes



were extracted from the BrainSpan Developmental Transcriptome,<sup>29</sup> and presented as the mean for each inversion at every developmental age through a heatmap.

**Gene expression analysis of the 17q21.31-inv**—To investigate whether inversion genotypes are associated with gene expression changes, genotype and transcriptomic data from blood samples of 187 healthy (age range: 30.60–83.70, mean 60.88 (SD: 11.35) years) participants were downloaded from the Parkinson’s Progressive Marker Initiative (PPMI).<sup>80</sup> Their 17q21.31-inv status was called following the *scoreInvHap* pipeline as described above. Homozygous inverted subjects (II) were combined with heterozygous subjects (NI) to form the inverted cohort. Differential expression (DE) analysis was performed by the *Seurat* package for R,<sup>99</sup> comparing the inverted and non-inverted cohorts (II and NI versus NN). Batch effects were removed by *BEER* script for R.<sup>100</sup> Significant DE genes were defined as FDR <0.05.

**Molecular QTL analysis of 17q21.31-inv in brain tissues**—To test whether the 17q21.31-inv is the eQTL or methylation quantitative trait loci (mQTL) accounting for gene expression changes, we used a tag SNP, rs1396862, to infer 17q21.31-inv status, and to query its eQTL and mQTL results provided in the data resources from summary-based mendelian randomization (SMR).<sup>101</sup> The default QTL p value threshold of 5e-8 was used. The brain eQTL and sQTL summary data has a sample size of 2865 brain cortex samples from 2443 individuals of European ancestry.<sup>102</sup> The brain mQTL summary data<sup>98</sup> has an estimated effective sample size of 1160 derived from a meta-analysis of ROSMAP,<sup>103</sup> Hannon et al.<sup>104</sup> and Jaffe et al.<sup>98,105</sup>

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

The authors would like to thank the research participants and staff involved in data collection of the UKB and ABCD study. This research was supported by the National Institutes of Health under R01MH118281 and R56AG061163. C.M. is supported by the Canadian Institutes of Health Research (CIHR) and the Kavli Institute for Brain and Mind (KIBM). T.K. is supported by the Research Council of Norway (#323961). P.M.V. and J.Y. acknowledge funding from the National Health and Medical Research Council (1113400) and the Australian Research Council (FT180100186 and FL180100072). J.Y. is supported by the Westlake Education Foundation and Tencent Foundation. PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson’s Research and funding partners. A list of the full names of all of the PPMI funding partners can be found at [www.ppmi-info.org/about-ppmi/who-we-are/study-sponsors](http://www.ppmi-info.org/about-ppmi/who-we-are/study-sponsors). The graphical abstract was created with BioRender ([biorender.com](http://biorender.com)).

## REFERENCES

1. Sturtevant AH (1921). A Case of Rearrangement of Genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 7, 235–237. [PubMed: 16576597]
2. Feuk L, Carson AR, and Scherer SW (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. [PubMed: 16418744]
3. Porubsky D, Höps W, Ashraf H, Hsieh P, Rodriguez-Martin B, Yilmaz F, Ebler J, Hallast P, Maggiolini FAM, Harvey WT, et al. Haplotype-resolved Inversion Landscape Reveals Hotspots of Mutational Recurrence Associated with Genomic Disorders. *10.1101/2021.12.20.472354*.

4. Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, et al. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* 372, eabf7117. 10.1126/science.abf7117. [PubMed: 33632895]
5. Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez OL, Guo L, Collins RL, et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* 10, 1784. [PubMed: 30992455]
6. Ruiz-Arenas C, Cáceres A, López-Sánchez M, Tolosana I, PérezJurado L, and González JR (2019). scoreInvHap: Inversion genotyping for genome-wide association studies. *PLoS Genet.* 15, e1008203. [PubMed: 31269027]
7. Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Mehdi SQ, Kajuna SLB, Barta C, Kungulilo S, Karoma NJ, Lu R-B, et al. (2010). The distribution and most recent common ancestor of the 17q21 inversion in humans. *Am. J. Hum. Genet.* 86, 161–171. [PubMed: 20116045]
8. Escaramís G, Docampo E, and Rabionet R (2015). A decade of structural variants: description, history and methods to detect structural variation. *Brief. Funct. Genomics* 14, 305–314. [PubMed: 25877305]
9. González JR, Ruiz-Arenas C, Cáceres A, Morán I, López-Sánchez M, Alonso L, Tolosana I, Guindo-Martínez M, Mercader JM, Esko T, et al. (2020). Polymorphic Inversions Underlie the Shared Genetic Susceptibility of Obesity-Related Diseases. *Am. J. Hum. Genet.* 106, 846–858. [PubMed: 32470372]
10. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, and Eichler EE (2002). Recent Segmental Duplications in the Human Genome. *Science* 297, 1003–1007. 10.1126/science.1072047. [PubMed: 12169732]
11. Ji X, and Zhao S (2008). DA and Xiao-two giant and composite LTR-retrotransposon-like elements identified in the human genome. *Genomics* 91, 249–258. [PubMed: 18083327]
12. Marques-Bonet T, and Eichler EE (2009). The evolution of human segmental duplications and the core duplicon hypothesis. *Cold Spring Harb. Symp. Quant. Biol.* 74, 355–362. [PubMed: 19717539]
13. Vollger MR, Guitart X, Dishuck PC, Mercuri L, Harvey WT, Gershman A, Diekhans M, Sulovari A, Munson KM, Lewis AP, et al. (2022). Segmental duplications and their variation in a complete human genome. *Science* 376, eabj6965. [PubMed: 35357917]
14. Lynch M, and Katju V (2004). The altered evolutionary trajectories of gene duplicates. *Trends Genet.* 20, 544–549. [PubMed: 15475113]
15. Jiang Z, Tang H, Ventura M, Cardone MF, Marques-Bonet T, She X, Pevzner PA, and Eichler EE (2007). Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nat. Genet.* 39, 1361–1368. [PubMed: 17922013]
16. Dennis MY, and Eichler EE (2016). Human adaptation and evolution by segmental duplication. *Curr. Opin. Genet. Dev.* 41, 44–52. [PubMed: 27584858]
17. Bekpen C, and Tautz D (2019). Human core duplicon gene families: game changers or game players? *Brief. Funct. Genomics* 18, 402–411. 10.1093/bfpg/elz016. [PubMed: 31529038]
18. Maggolini FAM, Sanders AD, Shew CJ, Sulovari A, Mao Y, Puig M, Catacchio CR, Dellino M, Palmisano D, Mercuri L, et al. (2020). Single-cell strand sequencing of a macaque genome reveals multiple nested inversions and breakpoint reuse during primate evolution. *Genome Res.* 30, 1680–1693. [PubMed: 33093070]
19. Giner-Delgado C, Villatoro S, Lerga-Jaso J, Gayà-Vidal M, Oliva M, Castellano D, Pantano L, Bitarello BD, Izquierdo D, Noguera I, et al. (2019). Evolutionary and functional impact of common polymorphic inversions in the human genome. *Nat. Commun.* 10, 4222. [PubMed: 31530810]
20. Okbay A, Baselmans BML, De Neve J-E, Turley P, Nivard MG, Fontana MA, Meddens SFW, Linnér RK, Rietveld CA, Derringer J, et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* 48, 624–633. [PubMed: 27089181]

21. Ikram MA, Fornage M, Smith AV, Seshadri S, Schmidt R, Debette S, Vrooman HA, Sigurdsson S, Ropele S, Taal HR, et al. (2012). Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat. Genet.* 44, 539–544. [PubMed: 22504418]
22. Grasby KL, Jahanshad N, Painter JN, Colodro-Conde L, Bralten J, Hibar DP, Lind PA, Pizzagalli F, Ching CRK, McMahon MAB, et al. (2020). The genetic architecture of the human cerebral cortex. *Science* 367, eaay6690. 10.1126/science.aay6690. [PubMed: 32193296]
23. Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees AJ, Fung HC, Duckworth J, Leung D, Gibson A, Morris CM, et al. (2005). The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. *Hum. Mol. Genet.* 14, 2399–2404. [PubMed: 16000317]
24. Laws SM, Friedrich P, Diehl-Schmid J, Muller J, Eisele T, Bäuml J, Förstl H, Kurz A, and Riemenschneider M (2007). Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease. *Mol. Psychiatr.* 12, 510–517.
25. Zabetian CP, Hutter CM, Factor SA, Nutt JG, Higgins DS, Griffith A, Roberts JW, Leis BC, Kay DM, Yearout D, et al. (2007). Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease. *Ann. Neurol.* 62, 137–144. [PubMed: 17514749]
26. Heckman MG, Kasanuki K, Brennan RR, Labbé C, Vargas ER, Soto AI, Murray ME, Koga S, Dickson DW, and Ross OA (2019). Association of MAPT H1 subhaplotypes with neuropathology of lewy body disease. *Mov. Disord.* 34, 1325–1332. [PubMed: 31234228]
27. Pittman AM, Myers AJ, Abou-Sleiman P, Fung HC, Kaleem M, Marlowe L, Duckworth J, Leung D, Williams D, Kilford L, et al. (2005). Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J. Med. Genet.* 42, 837–846. [PubMed: 15792962]
28. Makowski C, van der Meer D, Dong W, Wang H, Wu Y, Zou J, Liu C, Rosenthal SB, Hagler DJ Jr., Fan CC, et al. (2022). Discovery of genomic loci of the human cerebral cortex using genetically informed brain atlases. *Science* 375, 522–528. [PubMed: 35113692]
29. Tebbenkamp ATN, Willsey AJ, State MW, and Sestan N (2014). The developmental transcriptome of the human brain: implications for neurodevelopmental disorders. *Curr. Opin. Neurol.* 27, 149–156. [PubMed: 24565942]
30. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025. [PubMed: 25959774]
31. Carvalho CMB, and Lupski JR (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17, 224–238. [PubMed: 26924765]
32. Chen C-H, Gutierrez ED, Thompson W, Panizzon MS, Jernigan TL, Eyler LT, Fennema-Notestine C, Jak AJ, Neale MC, Franz CE, et al. (2012). Hierarchical genetic organization of human cortical surface area. *Science* 335, 1634–1636. [PubMed: 22461613]
33. Chen C-H, Fiecas M, Gutiérrez ED, Panizzon MS, Eyler LT, Vuoksima E, Thompson WK, Fennema-Notestine C, Hagler DJ Jr., Jernigan TL, et al. (2013). Genetic topography of brain morphology. *Proc. Natl. Acad. Sci. USA* 110, 17089–17094. [PubMed: 24082094]
34. Martínez-Fundichely A, Casillas S, Egea R, Ràmia M, Barbadilla A, Pantano L, Puig M, and Cáceres M (2014). InvFEST, a database integrating information of polymorphic inversions in the human genome. *Nucleic Acids Res.* 42, D1027–D1032. [PubMed: 24253300]
35. Li J, and Ji L (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95, 221–227. [PubMed: 16077740]
36. Li M-X, Yeung JMY, Cherny SS, and Sham PC (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131, 747–756. [PubMed: 22143225]
37. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, and Visscher PM; GIANT Consortium (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* 27, 3641–3649. 10.1093/hmg/ddy271. [PubMed: 30124842]
38. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. [PubMed: 11832223]

39. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, et al. (2005). A common inversion under selection in Europeans. *Nat. Genet.* 37, 129–137. [PubMed: 15654335]
40. Spielmann M, Lupiáñez DG, and Mundlos S (2018). Structural variation in the 3D genome. *Nat. Rev. Genet.* 19, 453–467. 10.1038/s41576-018-0007-0. [PubMed: 29692413]
41. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AMM, Pletikos M, Meyer KA, Sedmak G, et al. (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483–489. [PubMed: 22031440]
42. Xie M-J, Iwata K, Ishikawa Y, Nomura Y, Tani T, Murata K, Fukazawa Y, and Matsuzaki H (2021). Autistic-Like Behavior and Impairment of Serotonin Transporter and AMPA Receptor Trafficking in N-Ethylmaleimide Sensitive Factor Gene-Deficient Mice. *Front. Genet.* 12, 748627. 10.3389/fgene.2021.748627. [PubMed: 34745222]
43. Su X, Li W, Lv L, Li X, Yang J, Luo X-J, and Liu J (2021). Transcriptome-Wide Association Study Provides Insights Into the Genetic Component of Gene Expression in Anxiety. *Front. Genet.* 12, 740134. [PubMed: 34650599]
44. Zimmer-Bensch G, and Zempel H (2021). DNA Methylation in Genetic and Sporadic Forms of Neurodegeneration: Lessons from Alzheimer’s, Related Tauopathies and Genetic Tauopathies. *Cells* 10, 3064. 10.3390/cells10113064. [PubMed: 34831288]
45. Marín O (2016). Developmental timing and critical windows for the treatment of psychiatric disorders. *Nat. Med.* 22, 1229–1238. [PubMed: 27783067]
46. Campoy E, Puig M, Yakymenko I, Lerga-Jaso J, and Cáceres M (2022). Genomic architecture and functional effects of potential human inversion supergenes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 377, 20210209. [PubMed: 35694745]
47. Canzio D, Nwakeze CL, Horta A, Rajkumar SM, Coffey EL, Duffy EE, Duffié R, Monahan K, O’Keeffe S, Simon MD, et al. (2019). Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin  $\alpha$  Promoter Choice. *Cell* 177, 639–653.e15. [PubMed: 30955885]
48. Bowles KR, Pugh DA, Liu Y, Renton AE, Bandres-Ciga S, GanOr Z, Heutink P, Siitonen A, Bertelsen S, Cherry JD, et al. 17q21.31 Sub-haplotypes Underlying H1-Associated Risk for Parkinson’s Disease Are Associated with LRRC37A/2 Expression in Astrocytes. 10.1101/860668.
49. Bekpen C, Tastekin I, Siswara P, Akdis CA, and Eichler EE (2012). Primate segmental duplication creates novel promoters for the LRRC37 gene family within the 17q21.31 inversion polymorphism region. *Genome Res.* 22, 1050–1058. [PubMed: 22419166]
50. de Jong S, Chepelev I, Janson E, Strengman E, van den Berg LH, Veldink JH, and Ophoff RA (2012). Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. *BMC Genom.* 13, 458.
51. O’Brien HE, Hannon E, Hill MJ, Toste CC, Robertson MJ, Morgan JE, McLaughlin G, Lewis CM, Schalkwyk LC, Hall LS, et al. (2018). Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* 19, 194. [PubMed: 30419947]
52. Walker RL, Ramaswami G, Hartl C, Mancuso N, Gandal MJ, de la Torre-Ubieta L, Pasaniuc B, Stein JL, and Geschwind DH (2020). Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* 181, 745. [PubMed: 32359439]
53. Niarchou M, Moore TM, Tang SX, Calkins ME, McDonald-McGuinn DM, Zackai EH, Emanuel BS, Gur RC, Gur RE, Hanson E, et al. (2019). Psychiatric disorders in children with 16p11.2 deletion and duplication. *J. Psychiatr. Res.* 118, 8. [PubMed: 31445317]
54. Cervera-Carles L, Pagonabarraga J, Pascual-Sedano B, Pastor P, Campolongo A, Fortea J, Blesa R, Alcolea D, Morenas-Rodríguez E, Sala I, et al. (2016). Copy number variation analysis of the 17q21.31 region and its role in neurodegenerative diseases. *Am. J. Med. Genet. B Neuropsychiatr. Genet* 171B, 175–180. [PubMed: 26453547]
55. Mohajeri K, Cantsilieris S, Huddleston J, Nelson BJ, Coe BP, Campbell CD, Baker C, Harshman L, Munson KM, Kronenberg ZN, et al. (2016). Interchromosomal core duplicons drive both evolutionary instability and disease susceptibility of the Chromosome 8p23.1 region. *Genome Res.* 26, 1453–1467. [PubMed: 27803192]

56. Steinberg KM, Antonacci F, Sudmant PH, Kidd JM, Campbell CD, Vives L, Malig M, Scheinfeldt L, Beggs W, Ibrahim M, et al. (2012). Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat. Genet.* 44, 872–880. [PubMed: 22751100]
57. Zody MC, Jiang Z, Fung H-C, Antonacci F, Hillier LW, Cardone MF, Graves TA, Kidd JM, Cheng Z, Abouelleil A, et al. (2008). Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nat. Genet.* 40, 1076–1083. [PubMed: 19165922]
58. Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivières S, Jahanshad N, Toro R, Wittfeld K, Abramovic L, Andersson M, et al. (2015). Common genetic variants influence human subcortical brain structures. *Nature* 520, 224–229. [PubMed: 25607358]
59. Elvsåshagen T, Shadrin A, Frei O, van der Meer D, Bahrami S, Kumar VJ, Smeland O, Westlye LT, Andreassen OA, and Kaufmann T (2021). The genetic architecture of the human thalamus and its overlap with ten common brain disorders. *Nat. Commun.* 12, 2909. [PubMed: 34006833]
60. Bechara A, Damasio H, and Damasio AR (2000). Emotion, Decision Making and the Orbitofrontal Cortex. *Cerebr. Cortex* 10, 295–307. 10.1093/cercor/10.3.295.
61. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419. [PubMed: 25613900]
62. Mahdessian D, Cesnik AJ, Gnann C, Danielsson F, Stenström L, Arif M, Zhang C, Le T, Johansson F, Schutten R, et al. (2021). Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* 590, 649–654. [PubMed: 33627808]
63. Lo M-T, Hinds DA, Tung JY, Franz C, Fan C-C, Wang Y, Smeland OB, Schork A, Holland D, Kauppi K, et al. (2017). Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* 49, 152–156. 10.1038/ng.3736. [PubMed: 27918536]
64. Salm MPA, Horswell SD, Hutchison CE, Speedy HE, Yang X, Liang L, Schadt EE, Cookson WO, Wierzbicki AS, Naoumova RP, and Shoulders CC (2012). The origin, global distribution, and functional impact of the human 8p23 inversion polymorphism. *Genome Res.* 22, 1144–1153. [PubMed: 22399572]
65. Britten RJ (2010). Transposable element insertions have strongly affected human evolution. *Proc. Natl. Acad. Sci. USA* 107, 19945–19948. [PubMed: 21041622]
66. Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, and Eichler EE (2001). Positive selection of a gene family during the emergence of humans and African apes. *Nature* 413, 514–519. [PubMed: 11586358]
67. Dennis MY, Harshman L, Nelson BJ, Penn O, Cantsilieris S, Huddleston J, Antonacci F, Penewit K, Denman L, Raja A, et al. (2017). The evolution and population diversity of human-specific segmental duplications. *Nat. Ecol. Evol.* 1, 69. [PubMed: 28580430]
68. Qureshi AY, Mueller S, Snyder AZ, Mukherjee P, Berman JI, Roberts TPL, Nagarajan SS, Spiro JE, Chung WK, Sherr EH, et al. (2014). Opposing brain differences in 16p11.2 deletion and duplication carriers. *J. Neurosci.* 34, 11199–11211. [PubMed: 25143601]
69. Cárdenas-de-la-Parra A, Martin-Brevet S, Moreau C, Rodriguez-Herreros B, Fonov VS, Maillard AM, Zürcher NR, 16p11.2 European Consortium; Hadjikhani N, Beckmann JS, et al. (2019). Developmental trajectories of neuroanatomical alterations associated with the 16p11.2 Copy Number Variations. *Neuroimage* 203, 116155. [PubMed: 31494251]
70. Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S, Dukart J, Ferrari C, Conus P, Männik K, et al. (2015). The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Mol. Psychiatr.* 20, 140–147.
71. Sønderby IE, Gústafsson Ó, Doan NT, Hibar DP, Martin-Brevet S, Abdellaoui A, Ames D, Amunts K, Andersson M, Armstrong NJ, et al. (2020). Dose response of the 16p11.2 distal copy number variant on intracranial volume and basal ganglia. *Mol. Psychiatr.* 25, 584–602.
72. Ming C, Wang M, Wang Q, Neff R, Wang E, Shen Q, Reddy JS, Wang X, Allen M, Ertekin-Taner N, et al. (2022). Whole genome sequencing-based copy number variations reveal novel pathways and targets in Alzheimer’s disease. *Alzheimers Dement.* 18, 1846–1867. [PubMed: 34918867]

73. Sakamoto Y, Shimoyama S, Furukawa T, Adachi M, Takahashi M, Mikami T, Kuribayashi M, Osato A, Tsushima D, Saito M, et al. (2021). Copy number variations in Japanese children with autism spectrum disorder. *Psychiatr. Genet.* 31, 79–87. [PubMed: 33591083]
74. Glancy M, Barnicoat A, Vijeratnam R, de Souza S, Gilmore J, Huang S, Maloney VK, Thomas NS, Bunyan DJ, Jackson A, and Barber JCK (2009). Transmitted duplication of 8p23.1–8p23.2 associated with speech delay, autism and learning difficulties. *Eur. J. Hum. Genet.* 17, 37–43. [PubMed: 18716609]
75. Mullins N, Forstner AJ, O’Connell KS, Coombes B, Coleman JRI, Qiao Z, Als TD, Bigdeli TB, Børte S, Bryois J, et al. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* 53, 817–829. [PubMed: 34002096]
76. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, et al. (2007). Strong association of de novo copy number mutations with autism. *Science* 316, 445–449. [PubMed: 17363630]
77. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MAR, Green T, et al. (2008). Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* 358, 667–675. [PubMed: 18184952]
78. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, Tooley K, Presumey J, Baum M, Van Doren V, et al. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature* 530, 177–183. [PubMed: 26814963]
79. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, and Kamatani Y (2019). Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol.* 20, 117. [PubMed: 31159850]
80. Simuni T, Caspell-Garcia C, Coffey C, Lasch S, Tanner C, and Marek K; PPMI Investigators (2016). How stable are Parkinson’s disease subtypes in de novo patients: Analysis of the PPMI cohort? *Parkinsonism Relat. Disorders* 28, 62–67.
81. Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, Marchini J, and Smith SM (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 562, 210–216. 10.1038/s41586-018-0571-7. [PubMed: 30305740]
82. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12, e1001779. [PubMed: 25826379]
83. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, Bartsch AJ, Jbabdi S, Sotiropoulos SN, Andersson JLR, et al. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19, 1523–1536. [PubMed: 27643430]
84. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O’Connell J, et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209. [PubMed: 30305743]
85. Dale AM, Fischl B, and Sereno MI (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194. [PubMed: 9931268]
86. Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, Pérez-Stable EJ, Riley WT, Bloch MH, Conway K, et al. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev. Cogn. Neurosci.* 32, 4–7. [PubMed: 29051027]
87. Uban KA, Horton MK, Jacobus J, Heyser C, Thompson WK, Tapert SF, Madden PAF, and Sowell ER; Adolescent Brain Cognitive Development Study (2018). Biospecimens and the ABCD study: Rationale, methods of collection, measurement and early data. *Dev. Cogn. Neurosci* 32, 97–106. [PubMed: 29606560]
88. Sidorenko J, Kassam I, Kemper KE, Zeng J, Lloyd-Jones LR, Montgomery GW, Gibson G, Metspalu A, Esko T, Yang J, et al. (2019). The effect of X-linked dosage compensation on complex trait variation. *Nat. Commun.* 10, 3009. [PubMed: 31285442]
89. Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, Soules ME, Teslovich T, Dellarco DV, Garavan H, et al. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev. Cogn. Neurosci* 32, 43–54. [PubMed: 29567376]

90. Hagler DJ Jr., Hatton S, Cornejo MD, Makowski C, Fair DA, Dick AS, Sutherland MT, Casey BJ, Barch DM, Harms MP, et al. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage* 202, 116091. [PubMed: 31415884]
91. Chen C-H, Panizzon MS, Eyler LT, Jernigan TL, Thompson W, Fennema-Notestine C, Jak AJ, Neale MC, Franz CE, Hamza S, et al. (2011). Genetic influences on cortical regionalization in the human brain. *Neuron* 72, 537–544. [PubMed: 22099457]
92. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, Ip HF, Marioni RE, McIntosh AM, Deary IJ, et al. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Human Behav.* 3, 513–525. [PubMed: 30962613]
93. Clarke L, Fairley S, Zheng-Bradley X, Streeter I, Perry E, Lowy E, Tassé AM, and Flicek P (2017). The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res.* 45, D854–D859. [PubMed: 27638885]
94. 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, and Abecasis GR (2015). A global reference for human genetic variation. *Nature* 526, 68–74. [PubMed: 26432245]
95. Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, and Yang J (2019). A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* 51, 1749–1755. [PubMed: 31768069]
96. Yang J, Lee SH, Goddard ME, and Visscher PM (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. [PubMed: 21167468]
97. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, Graff M, Eliassen AU, Jiang Y, Raghavan S, et al. (2022). A saturated map of common genetic variants associated with human height. *Nature* 610, 704–712. [PubMed: 36224396]
98. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, Zhu Z, Kemper K, Yengo L, Zheng Z, et al. (2018). Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* 9, 2282. [PubMed: 29891976]
99. Satija R, Farrell JA, Gennert D, Schier AF, and Regev A (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Bio-technol.* 33, 495–502.
100. Zhang F, Wu Y, and Tian W (2019). A novel approach to remove the batch effect of single-cell data. *Cell Discov.* 5, 46. [PubMed: 31636959]
101. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, and Yang J (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487. [PubMed: 27019110]
102. Qi T, Wu Y, Fang H, Zhang F, Liu S, Zeng J, and Yang J (2022). Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat. Genet.* 54, 1355–1363. [PubMed: 35982161]
103. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, Xu J, Yu L, Gaiteri C, Bennett DA, et al. (2017). An xQTL map integrates the genetic architecture of the human brain’s transcriptome and epigenome. *Nat. Neurosci.* 20, 1418–1426. 10.1038/nn.4632. [PubMed: 28869584]
104. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, Troakes C, Turecki G, O’Donovan MC, Schalkwyk LC, et al. (2016). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* 19, 48–54. [PubMed: 26619357]
105. Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, and Kleinman JE (2016). Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* 19, 40–47. Stand-alone Supplementary Tables (Excel file). [PubMed: 26619358]

### Highlights

- Chromosomal inversion polymorphisms are associated with human brain morphology
- Inversions are in the genomic hotspots of neuropsychiatric loci
- The 17q21.31 inversion is linked to altered regulation and expression of its local genes

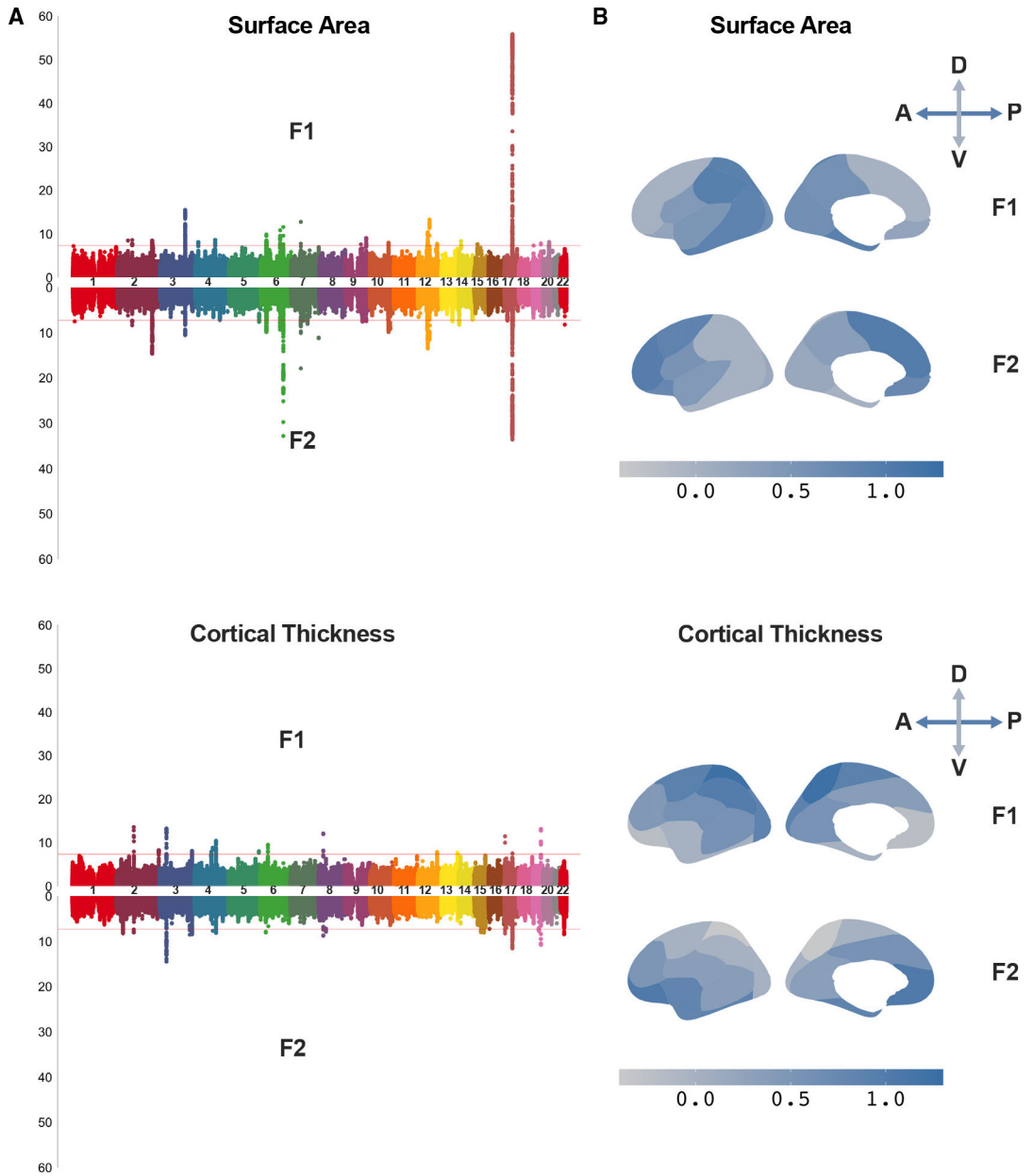
Author Manuscript

Author Manuscript

Author Manuscript

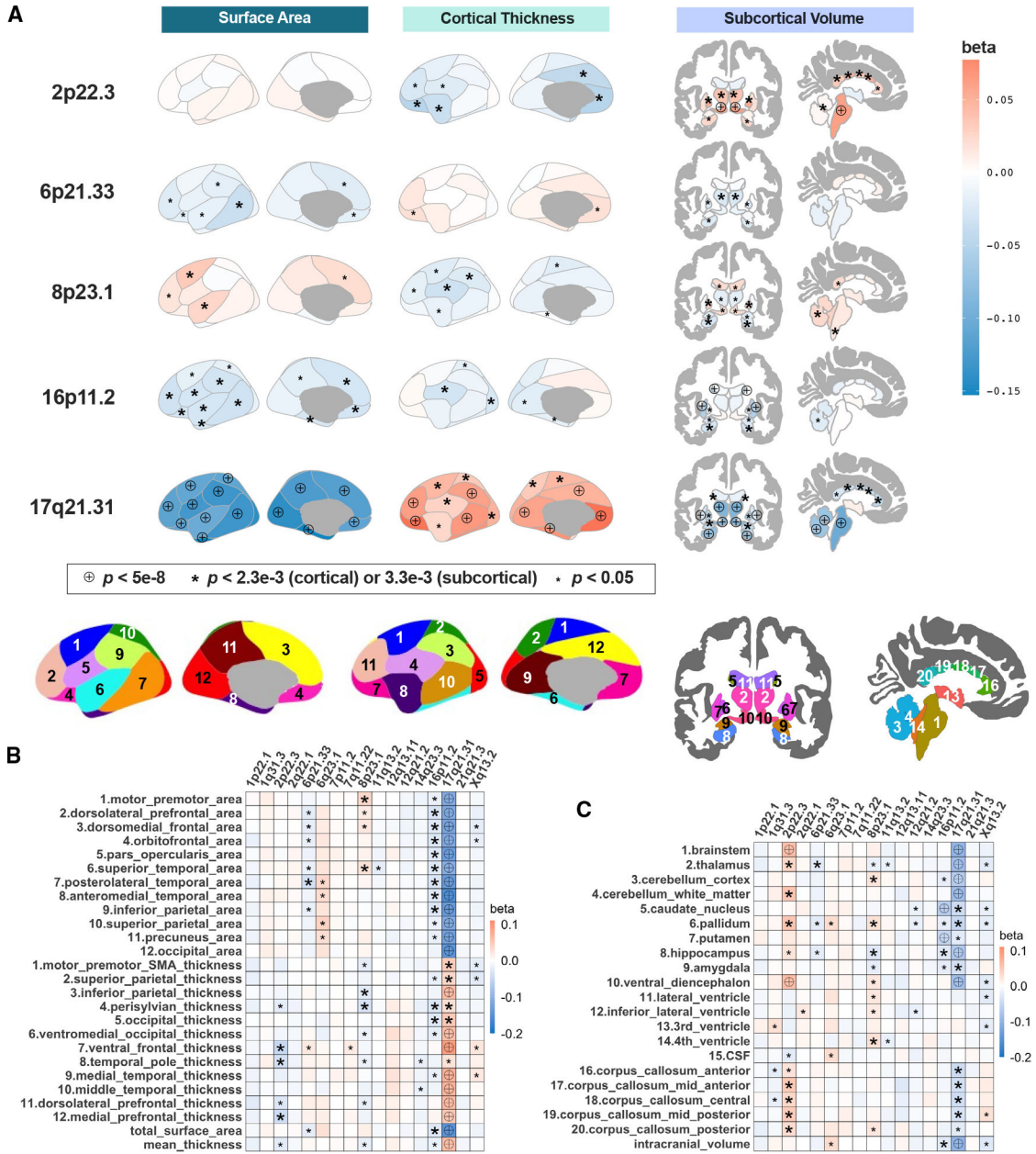
Author Manuscript





**Figure 1. Genetic latent factors underlying cortical morphology point to inversion polymorphisms**  
 (A) Miami plots of SNP association with each salient latent factor of the two-factor models estimated by genomic SEM for area and thickness separately. Genetic latent factors (F1 and F2) represent major axes of genetic variation underlying all cortical regions. The strongest association signals between the latent factors and the genetic variants reside in the 17q21.31 inversion region for area ( $p < 10^{-56}$ ), whereas there are more widespread effects across the genome such as from chromosomes 3 and 17 for thickness. Genome-wide significance line represents  $p = 5 \times 10^{-8}$ .

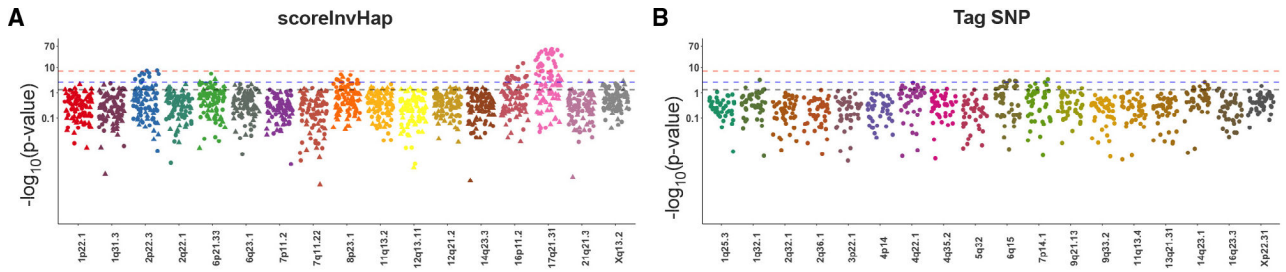
(B) Brain maps show standardized effects of each latent factor on each brain region, adapted from Makowski et al.<sup>28</sup> The two latent factors recapitulated the anterior-posterior (A-P) and dorsal-ventral (D-V) gradations of cortical patterning for area and thickness, respectively.



**Figure 2. Genetic effects of inversions on cortical and subcortical structures**

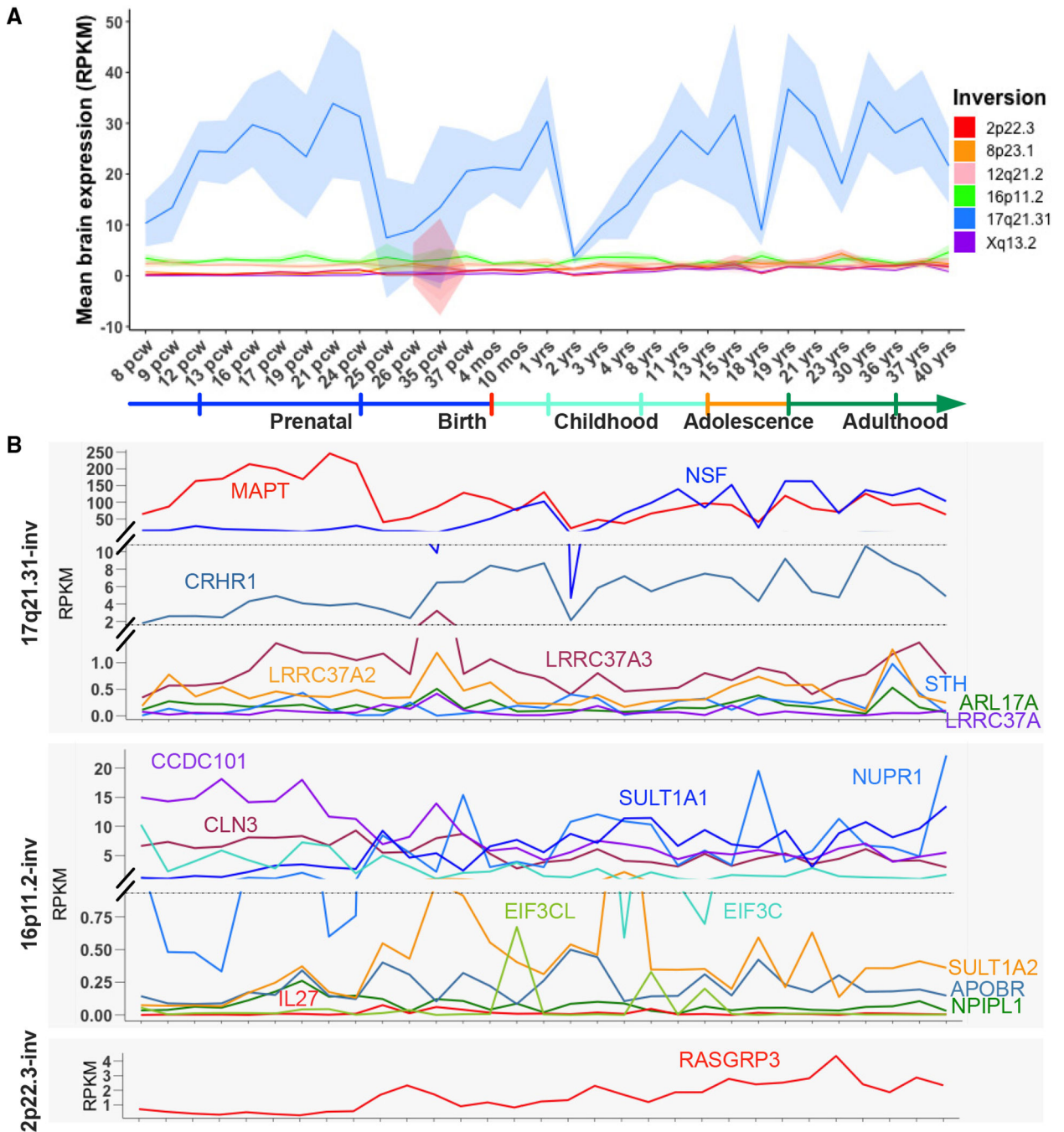
(A) Brain maps highlighting the spatial patterns of genetic associations with morphometric measures for five significant inversions. The bottom row depicts the brain atlases. Numbering of regions follows labels shown on the heatmaps in the bottom panels. (B–C) Heatmaps of associations colored according to the beta coefficients of the regression models for all comparisons (left: cortical; right: subcortical structures). Labels denote nominal (small asterisks), Bonferroni-corrected (big asterisks), and genome-wide significance (encircled crosses).

See Figures 1 and 2 and Table S4 for inversion association results with regional brain morphology when global brain size is taken into account.



**Figure 3. Manhattan plots visualize the association results for each individual inversion-morphology pair using both *scoreInvHap* and tag SNP approaches**

The dotted lines indicate the significant levels, ranging from genome-wide ( $p < 5 \times 10^{-8}$ ), Bonferroni-corrected ( $p < 2.3 \times 10^{-3}$ ), to nominal significance ( $p < 0.05$ ). For the inversions identified through *scoreInvHap* (A), several significant inversion-morphology pairs exceeded the Bonferroni-corrected significance level. These significant pairs also showed good generalizability between the UKB and ABCD cohorts, as indicated by the symbols used (circles for UKB and triangles for ABCD). Specifically, three inversions (2p22.3, 16p11.2, and 17q21.31) reached genome-wide significance in the UKB cohort. Two additional inversions (6p21.33 and 8p23.1) reached nearly genome-wide significance. See Figure S3 and Table S4 for results after adjusting for global measures.



**Figure 4. Expression profiles of genes located within the inversion regions across developmental time points**

(A) Mean expression levels of genes located within the inversion regions are plotted across time for the inversions that showed significant association with brain morphology. Shaded parts represent 95% confidence intervals.

(B) The line graphs show the expression of individual genes within the 2p22.3 inversion (2p22.3-inv), 16p11.2-inv, and 17q21.31-inv, the three inversions that were identified to be associated with brain morphology with genome-wide significance ( $p < 5 \times 10^{-8}$ ). Because these genes are expressed at very different ranges, we insert breaks on the y axis to give the

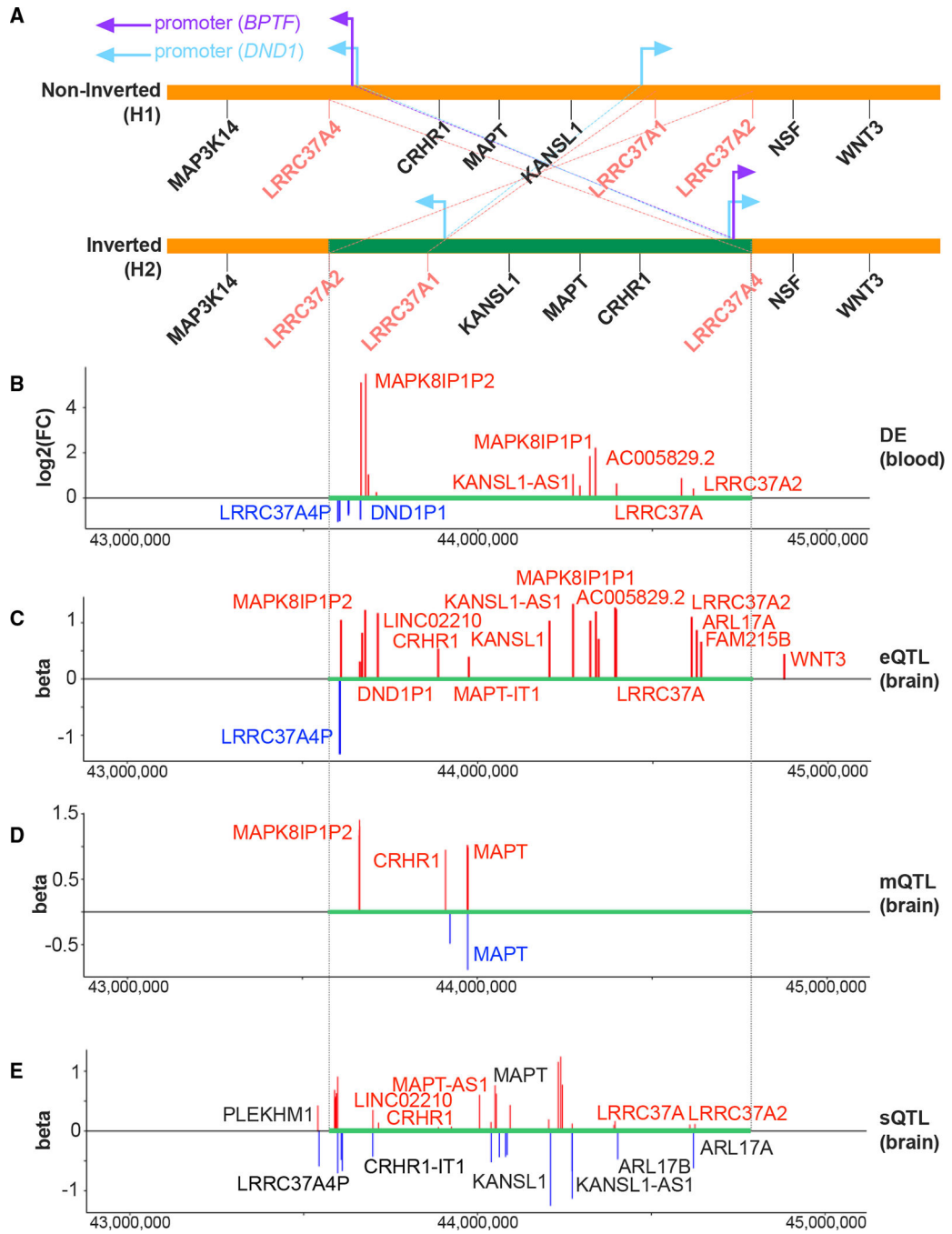
full range of expression levels of all the genes. Each line shows a different gene, and each point shows a different developmental time point. Heatmaps for the 8p23.1-inv, 12q21.2-inv, and Xq13.2-inv are provided in Figure S4. We did not plot the remaining inversions from *scoreInvHap* since they do not overlap with any genes. Only genes present in the BrainSpan Atlas are displayed. pcw, post-conception week(s); yo, year(s) old.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 5. Gene regulatory landscape of the 17q21.31-inv region**

Plots are aligned based on the base pair position.

(A) Schematic plot of the 17q21.31-inv region (in green), adapted from Bekpen et al.<sup>49</sup> The arrows indicate transcription directions of acquired novel promoters from *BPTF* and *DND1*. Overlying blue bands illustrate regions of segmental duplications (copy-number polymorphisms [CNPs]), with numbers denoting the lengths.

(B) Differential expression (DE) of genes comparing inversion carriers with non-carriers. The genes denoted in blue or red represent lower or higher expression for the inversion carriers, respectively.

(C) The eQTL analysis reveals that the 17q21.31-inv locus is a genetic variant that affects the expression of its target genes that show differences in gene expression between the inverted and non-inverted alleles.

(D) The methylation QTL (mQTL) analysis shows that the 17q21.31-inv locus is a genetic variant that affects the methylation of its target genes. There are multiple CpG sites for a given gene.

(E) The splicing QTL (sQTL) analysis demonstrates that the 17q21.31-inv is associated with isoform-level transcriptional abundance of its target genes. The genes indicated in black exhibit both higher and lower abundance of different isoforms.

See also Table S5.



## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
UKB demographic, clinical, genotype and neuroimaging data	UK Biobank version 3	<a href="https://www.ukbiobank.ac.uk/">https://www.ukbiobank.ac.uk/</a>
ABCD demographic, genotype and neuroimaging data	ABCD data repository	<a href="https://nda.nih.gov/abcd/">https://nda.nih.gov/abcd/</a>
BrainSpan developmental brain transcriptomic data	BrainSpan (Tebbenkamp et al.) <sup>29</sup>	<a href="http://brainspan.org/static/download.html">http://brainspan.org/static/download.html</a>
Human inversion data	Human Polymorphic Inversion Database (Martinez-Fundichely et al.) <sup>34</sup>	<a href="http://invfestdb.uab.cat/">http://invfestdb.uab.cat/</a>
Human inversion data	Human Genome Structural Variation Consortium (Ebert et al.) <sup>4</sup>	<a href="https://www.internationalgenome.org/human-genome-structural-variation-consortium">https://www.internationalgenome.org/human-genome-structural-variation-consortium</a>
PPMI DNA and RNA sequencing data	Parkinson's Progression Markers Initiative (PPMI) database (Simuni et al.) <sup>80</sup>	<a href="https://www.ppmi-info.org/access-data-specimens/download-data/">https://www.ppmi-info.org/access-data-specimens/download-data/</a>
Software and algorithms		
R 4.0.2	R	<a href="https://cran.r-project.org/">https://cran.r-project.org/</a>
<i>scoreInvHap</i> development version	Bioconductor	<a href="https://bioconductor.org/packages/release/bioc/html/scoreInvHap.html">https://bioconductor.org/packages/release/bioc/html/scoreInvHap.html</a>
imputeInversion scripts	GitHub	<a href="https://github.com/isglobal-brge/imputeInversion">https://github.com/isglobal-brge/imputeInversion</a>
<i>seurat</i> version 4.1.0	GitHub	<a href="https://github.com/satijalab/seurat">https://github.com/satijalab/seurat</a>
GCTA version 1.93.2 beta Linux	Yang Lab	<a href="https://yanglab.westlake.edu.cn/software/gcta/">https://yanglab.westlake.edu.cn/software/gcta/</a>
SMR version 1.03	Yang Lab	<a href="https://yanglab.westlake.edu.cn/software/smr">https://yanglab.westlake.edu.cn/software/smr</a>
Genome table browser	UCSC Genome Browser	<a href="https://genome.ucsc.edu/cgi-bin/hgTables">https://genome.ucsc.edu/cgi-bin/hgTables</a>