Unraveling the Link between CNVs, General Cognition, and Individual Neuroimaging Deviation Scores from a Reference Cohort

Charlotte Fraza[1,2], Ida E. Sønderby[5,6,7], Rune Boen[5,6,], Christian F. Beckmann[1,2,4], Andre F. Marquand[1,2,3]
[1] Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, the Netherlands
[2] Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, the Netherlands
[3] Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London, London, UK
[4] Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK
[5]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway
[6]Norwegian Centre for Mental Disorders Research, Division of Mental Health and Addiction, Oslo University Hospital, University of Oslo, Oslo, Norway
[7]KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway

Corresponding Author - Charlotte Fraza, Charlotte.fraza@donders.ru.nl

Keywords – Copy Number Variations (CNVs), Neurodevelopmental Disorders, Normative Modeling, Brain Structure.

Abstract
Background: Copy number variations (CNVs) are genetic variants that can have substantial influence on neurodevelopment, neuropsychiatric traits and morphometric brain changes, yet their impact at the individual level remains unknown. Common case-control approaches for analyzing CNVs often overlook individual variation between carriers and exclude rarer variants, due to their limited sample size. This study aims to map individualized brain deviation scores in individuals with pathogenic CNVs.
Methods: We used normative modeling to map neuroimaging features from several large neuroimaging datasets and applied these models to understand the neurobiological profile of CNV carriers in UK Biobank**.** We highlighted the 1q21.1 distal deletion and duplication, as an example of our normative modeling-CNV approach. Next, we counted the number of extreme deviations for each participant from the mean and centiles of variation from population reference norms, giving us a combined risk score per participant per imaging modality.
Results: We show a high degree of heterogeneity between pathogenic CNV carriers in their implicated brain regions. For example, the cerebellum, brainstem and the pallidum show large negative deviations for certain 1q21.1 duplication carriers. For certain 1q21.1 deletion CNV carriers the caudate and accumbens show notable positive deviations. Finally, we show that negative deviations from these models are correlated to cognitive function.
Conclusions: This study marks a starting point in comprehending the impact of pathogenic

CNVs on brain phenotypes, underscoring the intricacies of these genetic variations at the individual level and provides a means to studying the effects of rare CNVs in carrier individuals.

1. Introduction

Copy number variations (CNVs) are genetic variants that can have substantial influence on neurodevelopment, neuropsychiatric traits and morphometric brain changes[1,2]. Some CNVs emerge as significant genetic risk factors for a spectrum of neurodevelopmental and other psychiatric disorders[3–7]. Specifically, previous studies have shown that certain rare recurrent CNVs increase the risk for schizophrenia[5,8–12], attention deficit hyperactivity disorder (ADHD)[13], autism spectrum disorder (ASD)[14,15] and links between a decreased intelligence quotient (IQ) score and pathogenic CNVs have been established[16]. Despite their known importance, the effects of the majority of individual CNVs on the brain and behavior remain largely unknown. CNVs are typically studied at the group level, which may mask considerable inter-individual variation in their phenotypic presentation. Furthermore, the group-based approach is only feasible for the more common CNVs, limiting our ability to map the effects of rarer CNVs with high penetrance or large effect size. Understanding the effects of specific CNVs on both brain organization and cognition at the individual level is crucial for mapping the landscape of their impact on mental disorders.

CNVs often have pleiotropic effects, influencing multiple downstream processes simultaneously[17–19]. The diversity in genetic CNV-mediated effects is heightened by interactions with both the rest of the genome and environmental factors[20]. This renders each CNV's impact unique to the individual, complicating the task of unraveling their contributions to overall mental health. Conventional imaging and behavioral CNV studies often adopt a cases vs. controls framework, which has led to tremendous insights. However, in order to grasp the full spectrum of CNV effects, which manifest in a quite heterogeneous manner, we need to move beyond group-level distinctions[21].

Understanding the individual-level impact of CNVs on the brain and behavior has posed a challenge, yet the emergence of large-scale normative models may offer a solution. By employing this approach, z-scores can be calculated for each individual across various neuroimaging modalities, quantifying individual variations against the mean and centiles of population reference norms. Normative models thereby shift focus from group-level to individual-level inferences[22–25] and allow us to quantify atypical developmental trajectories[26,27]. Normative modeling has proven its efficacy in correlating individual behavioral phenotypes with deviations from reference cohorts, spanning disorders like schizophrenia, autism, and ADHD[28–30] and mapping disease progression in Alzheimer's disease[31]. Using normative models, we can map brain deviation scores for individuals with a specific pathogenic CNV. Afterward, we can use these pathogenic CNV-brain phenotypic deviation scores and correlate them with behavioral traits, uncovering hidden facets that group analyses could potentially miss. Importantly, owing to its focus on individual differences, this approach is not limited to common CNVs; it extends to rarer variants, opening doors to tailored research for smaller populations of rare genetic variants.

The normative modeling approach is specifically designed to give insight into individual brain deviation scores and their relationship to mental health. Mental health disorders such as major depression and schizophrenia lie on a gradient of severity and can be seen as the extreme values on continuous dimensions[32]. This notion aligns with normative modeling, as it positions individuals with large brain deviation scores at the

fringes of the normative spectrum[33]. In this study, our primary objective is to delve into the individualized impacts of CNVs on brain structure and behavior. By adopting a normative modeling approach, we take a step towards crafting personalized risk profiles that facilitate cross-individual comparisons. These profiles prove insightful for individuals with similar genetic mutations that manifest in comparable behavioral effects and share diverse neuroimaging fingerprints. We hypothesize that: (i) individuals with a CNV related to cognitive deficits or neurodevelopmental disorders, will have larger deviation scores compared to a reference model across many brain areas and (ii) that the patterns of deviation across brain regions will be highly variable across these CNV carriers.

Methods and Materials

From Pathogenic CNVs to Brain Structures: A Normative Modeling Approach

An overview of our analytic workflow is presented in Fig. 1. In brief, to discern whether certain neuroimaging modalities exhibited more atypicalities than others, we first employed normative modeling to analyze Image Derived Phenotypes (IDPs) from the UK Biobank study[34], covering diverse functional, structural, and diffusion tensor imaging measures. Afterwards, to create an exploration of the effect of pathogenic CNVs on the brain, we established a voxel-based morphometric variation model using Jacobian determinant images derived from the non-linear image registration to the MNI152 space. To create the voxel wise Jacobian normative model, we pooled a large dataset from publicly available repositories, leveraging Jacobian determinant images from non-linear image registration, specifically via the anatomical processing tools found in FSL (details available in the supplement). Jacobians, in their essence, provide a spatially precise measure of voxel-based morphometric differences, capturing the extent of volumetric adjustments—either expansion or contraction—needed to align each sample with the registration template for each voxel. These determinants provide a more informative dataset compared to other derived measures[35] and describe aggregate differences, avoiding the partly arbitrary distinction between grey and white matter. Moreover, it is well-established that specific CNVs influence intracranial volume (ICV)[36,37]. Consequently, we anticipated that this influence would manifest as either an increase or decrease in the necessary volumetric adjustments for individuals with these CNVs, which would be reflected in their Jacobians.

Every normative model constructed factored in covariates such as age and sex, in addition to fixed effects for site, the mathematical details are described in supplement. To ensure accurate modeling of non-linear and non-Gaussian effects, we employed a warped Bayesian linear regression (BLR) model[38].

We focussed on 92 CNVs proposed to be pathogenic (henceforth 'pathogenic'), and their reciprocal CNVs[39–41]. These can be requested from https://biobank.ndph.ox.ac.uk/ukb/app.cgi?id=14421. We highlighted the 1q21.1 distal deletion and duplication, as an example of our normative modeling-CNV approach. This CNV has shown moderate to strong effects on cognition[4,40], a dose-response per copy number for head circumference[42], with microcephaly in deletion carriers and macrocephaly in duplication carriers, and has been associated with global cortical surface structure alterations[36]. Furthermore, individuals with a 1q21.1 deletion and duplication show an increased risk for several neurodevelopmental disorders[3,4,42–44]. While our study is primarily designed to unravel individualized pathogenic CNV effects, we also explored the potential for aggregating subjects with similar CNVs to uncover converging neuroimaging alterations across pathogenic CNVs, opening avenues for comparison with traditional case-control

studies. Finally, we linked the found brain deviation scores to cognition by testing a correlation between the negative deviation scores and general cognitive ability[45].

Fig. 1A visually outlines our study workflow i. We first mapped participants with pathogenic CNVs in UK Biobank and placed them in our test set of normative models. ii. We created the normative models for each individual IDP or voxel, taking into account several covariates, such as age and sex. iii. We counted the number of extreme deviations for each participant and mapped where participants with a pathogenic CNV lay on this distribution, giving us a combined risk score per participant per imaging modality. iv. We correlated the total number of extreme z-scores with a measure of general cognitive ability. In Fig. 1B, we show an overview of the datasets used, encompassing IDPs from the UK biobank and Jacobian data from seven sites, details can be found in the supplement. In total, we used 44,456 participants in the IDP study, and in the Jacobian-voxel-based study 19,620 participants who had underwent and passed visual quality control. There were 375 individuals with pathogenic CNVs in the quality controlled Jacobian neuroimaging dataset. In Fig. 3A and 4A, the final pathogenic CNV sample used for the Jacobian normative models study is shown.

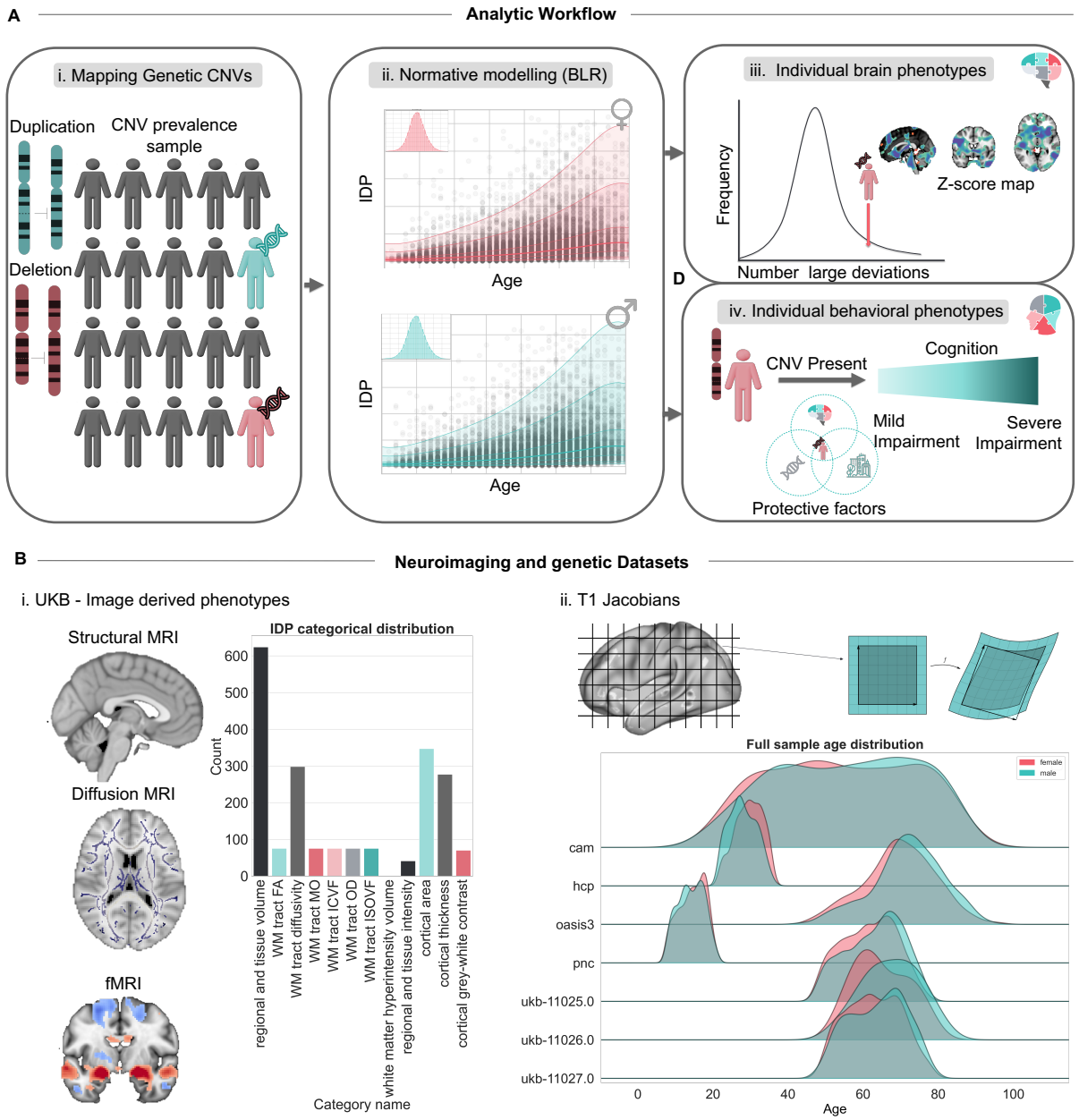


Fig. 1 | Overview of study design and data resources. A. Schematic overview of the study workflow and hypothesis. First, we quantify the number of participants with pathogenic CNVs, previously linked to neurodevelopmental and psychiatric disorders. Then we create normative models for the IDPs and the voxel-wise Jacobians. Afterward, we calculate the number of large deviation scores (|Z|>2). We plot the number of large deviation scores for individuals with pathogenic CNVs compared to the rest of the population. Finally, we correlate the extreme brain deviation scores and general cognitive ability. CNV, Copy Number Variant; IDP, image derived phenotype; BLR, Bayesian Linear Regression. **B. Overview of the neuroimaging datasets used in this study.** i. Distribution of the IDPs present in the UK biobank dataset, derived from functional, structural, and diffusion tensor imaging. ii. Distributions of the data from seven sites used in the Jacobian normative model, split by sex. In total, for the IDP study we use 44,456 participants, and for the Jacobian-voxel-based study we use 19,620 visually quality controlled participants.

Results

Initially, we fitted a normative model to the general IDPs derived from the structural, functional and diffusion measures from the UK biobank. Specifically, we were interested in the differences between participants who had pathogenic CNVs in comparison to those without. Detailed results from this IDP normative model can be explored in the supplementary Fig. 1, 2 and 3. A notable observation was that structural measures appeared most informative when we explored individual deviation scores. In supplementary Fig. 2 and 3, we can see that the structural measures gave a larger spread of the number of extreme deviations compared to other measures, which indicates a larger variation in individual differences.

Recognizing the significance of structural measures, we mapped a voxel-based morphometric variation model to characterize these differences at a finer scale, see Fig. 2. The resulting model accounted for as much as 52% of the variation in morphometric changes. Validation of the models entailed the application of several criteria. Among these were the kurtosis and skewness of the resulting z-score distribution, which measure the effectiveness of the warping function in capturing the non-linearity and non-Gaussianity of the data. In brief, all voxels show relatively low skew (i.e. $|skew| < 1$) and acceptable excess kurtosis (<5).

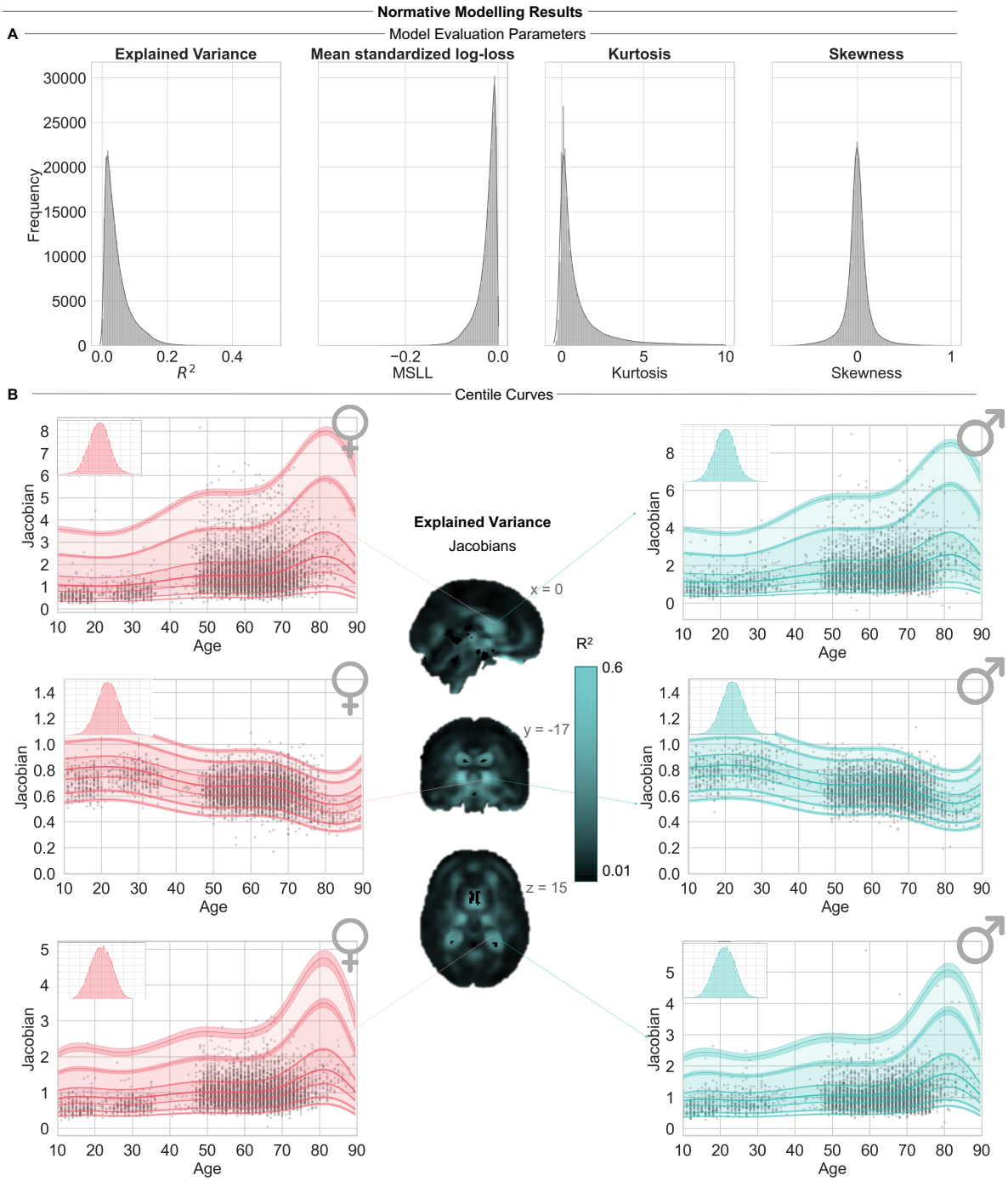


Fig. 2 | Overview of normative modeling results. A. Performance metrics for the test set. Both skew and kurtosis serve as indicators of the model's accuracy in estimating shape via warped Bayesian Linear Regression. **B**. Depiction of varied normative trajectories across distinct voxels, showing in the corner the histogram of z-scores, with the accompanying whole brain explained variance (R^2) map, split based on sex.

Individual risk profiles - combining pathogenic CNVs and brain imaging deviation scores

Next, we aimed to characterize the individual variations in brain deviation scores associated with specific pathogenic CNVs. To accomplish this, we counted the number of extreme deviations (|Z|>2) for each individual. An overview of the average counts of negative and positive deviation scores across all pathogenic CNVs is presented in supplementary Fig. 4.

Afterward, we constructed individualized risk profiles for 1q21.1 distal deletion (1q21.1del) and duplication (1q21.1dup). We chose this CNV to highlight our method, as it is relatively well-characterized with regard to brain structural differences[21], has shown impairments in cognition and has been associated with neurodevelopmental disorders[42]. Fig. 3 and 4 depict the number of positive and negative deviation scores for individuals with a deletion or duplication, respectively, in comparison to participants without pathogenic CNVs. Additionally, these figures showcase deviation score brain maps for individuals with a pathogenic CNV, pinpointing regions exhibiting pronounced positive or negative deviations when compared to a reference cohort, showcasing a high degree of variability of volumetric alterations across carriers. The maps show areas with marked negative or positive Jacobian signals or localized volumetric alterations. The interpretation of positive and negative deviations can vary depending on the mean of the normative model. For example, when the deviation is positive compared to the mean, it indicates more expansion and a relatively small volume in that voxel compared to the norm. We can see from panel B that positive deviations, i.e. more volume expansions than predicted, were more present for participants with a 1q21.1del, and negative deviations, i.e. more volume contractions than predicted by the model, were more prevalent for participants with a 1q21.1dup. For the detailed brain maps of all the participants with a 1q21.1 distal deletion or duplication, see supplementary Fig. 5 and 6. The most implicated brain areas, calculated using the mean deviation score per ROI, are also summarized in word clouds.

Then, we conducted a joint analysis for the 1q21.1del and 1q21.1dup groups, respectively, to examine the common distribution of deviation scores across all subjects. We would expect that there is convergence in deviation scores amongst the individual subjects in certain brain regions[46], especially those that can be found with traditional cases vs. control paradigms, but also we expect divergence amongst the subjects in deviation scores in other brain regions, as we know that not every participant with a pathogenic CNV is implicated in their behavioral phenotype or cognition. Thus, a participant can have a pathogenic CNV and still be within the bounds of standard cognitive performance. The analysis revealed prominent deviations in specific brain regions, see Fig 5. Notably, for 1q21.1dup, substantial negative deviations manifested in the occipital cortex, while for 1q21.1del, pronounced positive deviations were observed in the cerebellum and thalamus. Moreover, we also show the brain regions implicated in 16p13.11 deletion and duplication, which is a CNV of which less is known currently, showcasing the versatility of our approach, which can be applied to diverse rare pathogenic CNVs.

Relationship brain deviation scores and cognitive deficits

Fig. 6 outlines our hypothesis concerning the impact of pathogenic CNVs on cognition, proposing that certain CNVs might contribute to cognitive impairment in the absence of protective factors. We plotted the fluid intelligence scores among participants with pathogenic CNVs and participants without pathogenic CNVs in Fig. 6A. To analyze the relationship between large deviation scores and cognition, we generated a general cognitive ability score for each participant by calculating the first principal component from various cognitive tests within the UK Biobank dataset, see the supplementary methods for details. We examined the Pearson correlation between the total count of extreme positive deviation scores ($Z>2$) and the general cognitive ability score ($r = 0.0080$, $p = 0.6639$) and the extreme negative deviation scores ($Z <-2$) and the general cognitive ability score ($r = -0.04$, $p = 0.03$), see Fig. 6. This correlation indicated that a higher number of extreme negative

deviations, indicating more volume contractions compared to the mean of the population, was significantly associated with a lower general cognitive ability score.

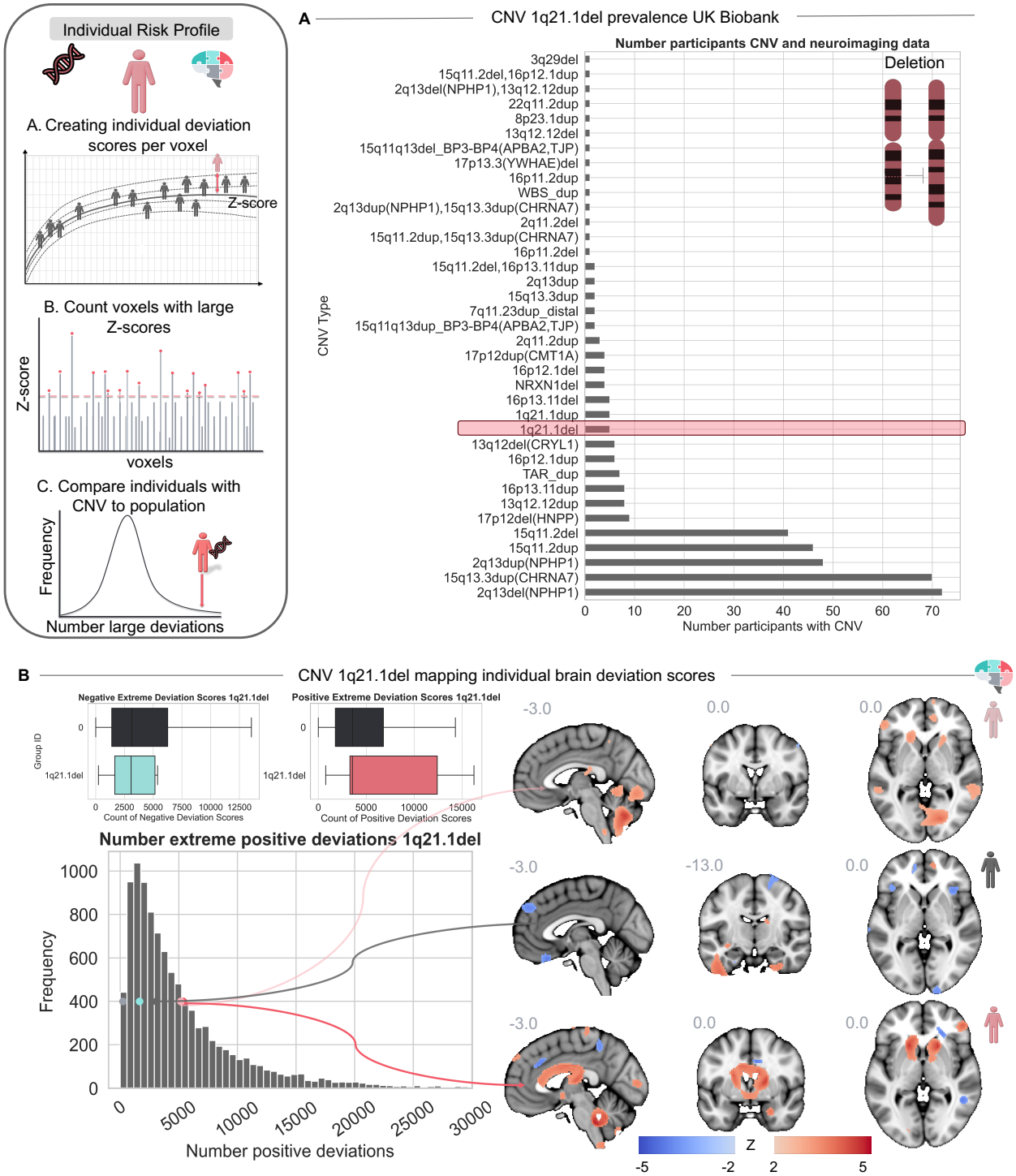


Fig. 3 | Individual Risk Profiles 1q21.1 deletion. A. The prevalence of pathogenic CNV carriers including 1q21.1del in the UK Biobank neuroimaging dataset used in Jacobian analysis. **B.** Counts of extreme positive and negative deviation scores (|Z|>2) among participants with a 1q21.1 deletion in contrast to participants without a pathogenic CNV. Left: Dots show each individual 1q21.1 deletion carrier's position in the distribution. Right: Displaying the profile of three selected 1q21.1 distal deletion CNV carriers.

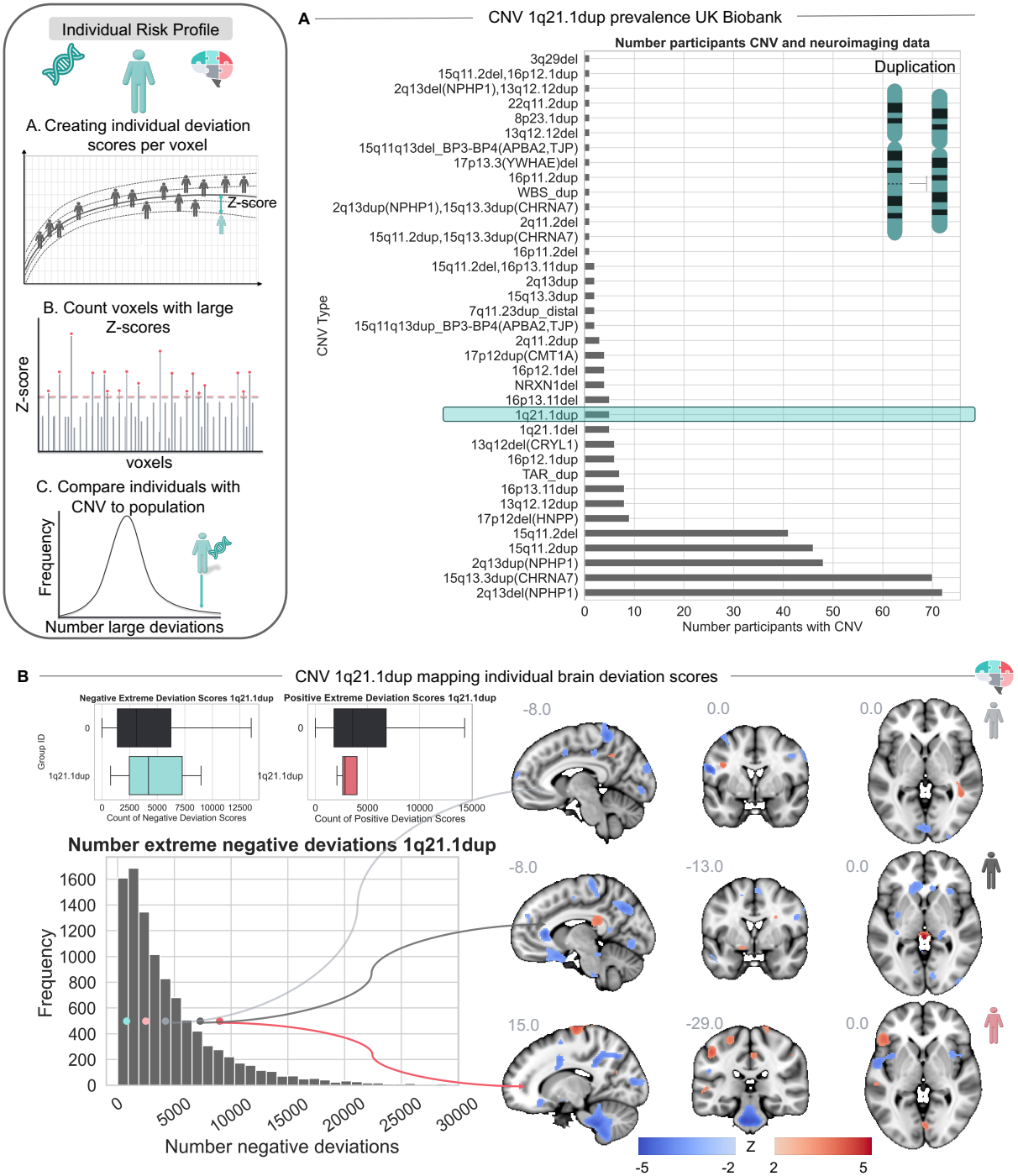


Fig. 4 | Individual Risk Profiles 1q21.1 duplication. **A**. The prevalence of pathogenic CNV carriers including 1q21.1dup in the UK Biobank neuroimaging dataset used in Jacobian analysis.. **B.** The counts of extreme positive and negative deviation scores (|Z|>2) among participants with a 1q21.1 duplication in contrast to participants without a pathogenic CNV. Left: Dots show each individual 1q21.1 duplication carrier's position in the distribution. Right: Displaying the profile of three selected 1q21.1 distal duplication CNV carriers.

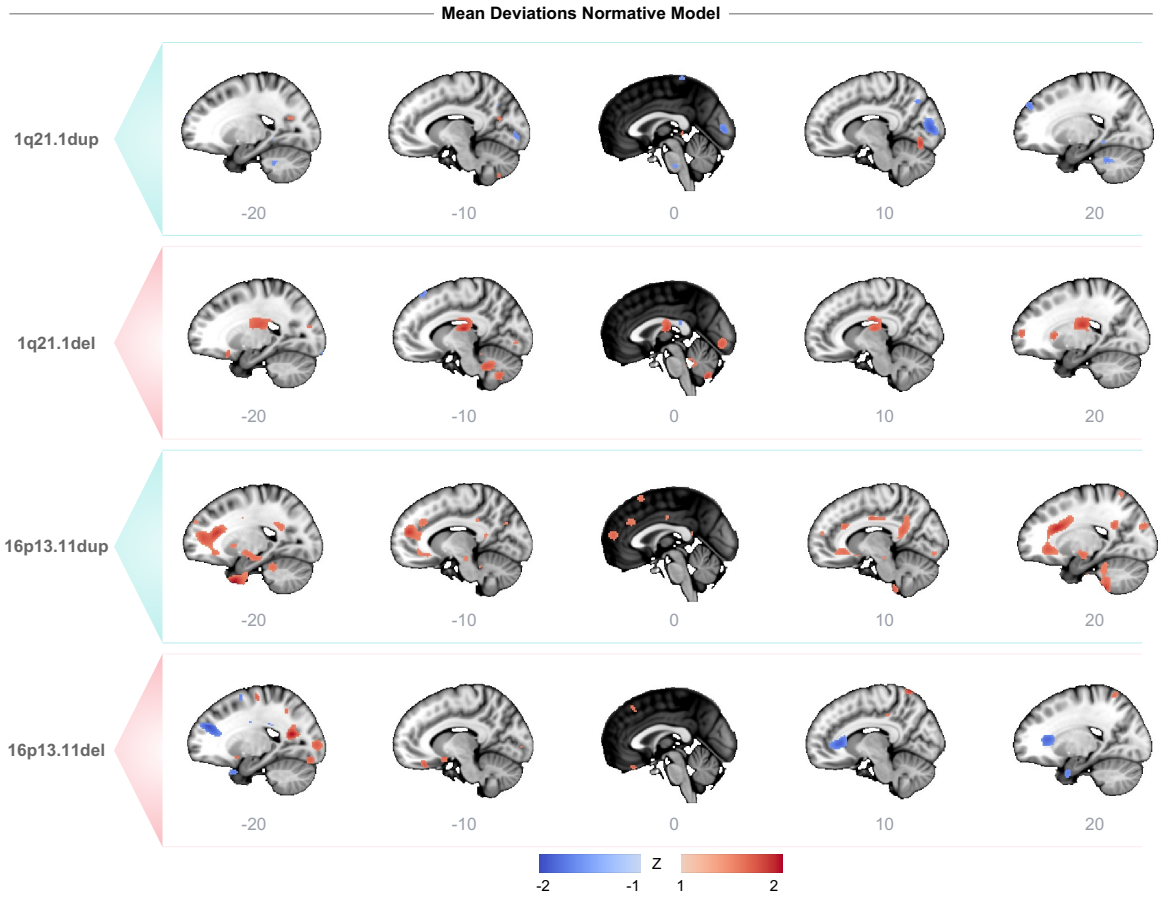


Fig. 5 | Convergence of Extreme Positive or Negative Deviation Scores. The overlapping brain deviation score maps for participants with the 1q21.1 distal or 16p13.11 CNVs. On the x-axis, we show different sagittal slices with steps of 10.

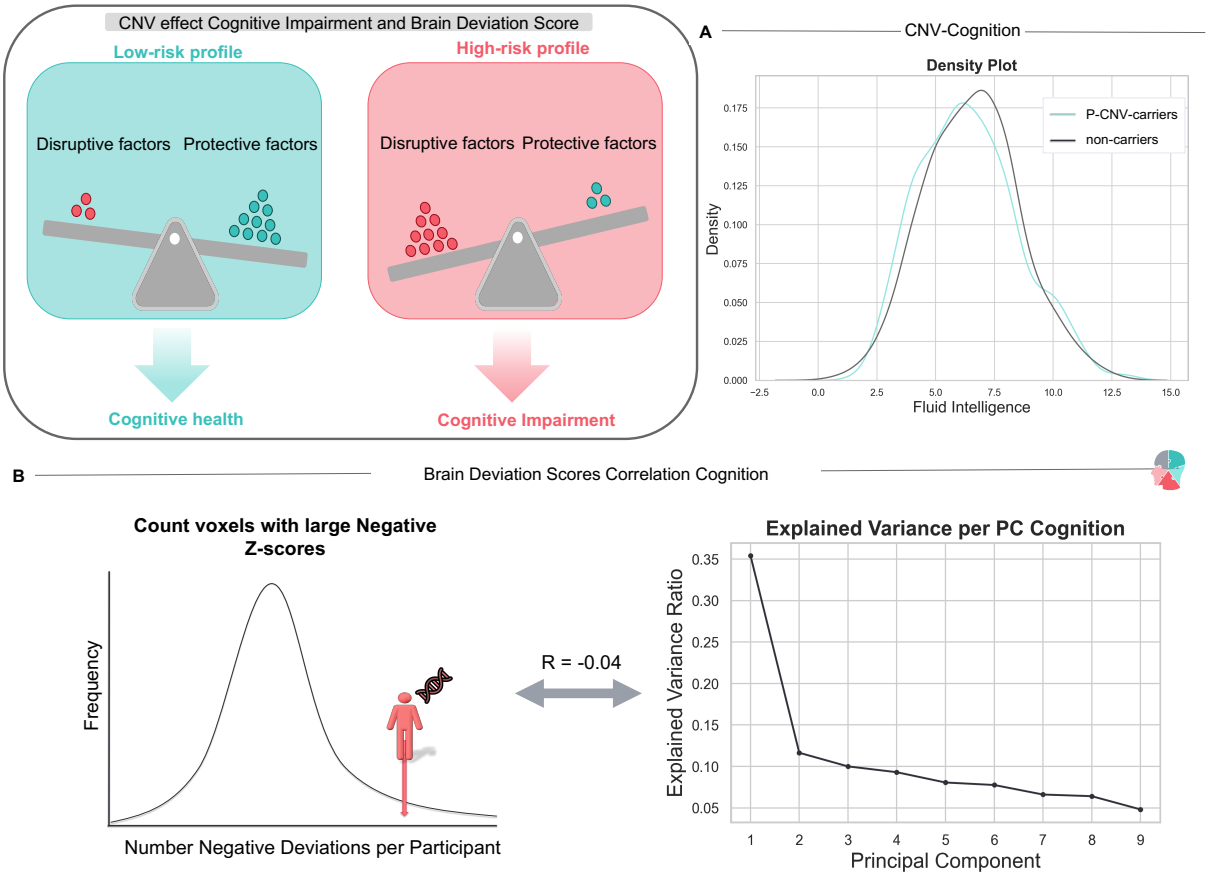


Fig. 6 | Pathogenic CNVs Impact on Cognitive Function. Illustration of how protective and disruptive factors, which may include (pathogenic) CNVs, might lead to cognitive impairment. **A.** Variations in fluid intelligence scores among participants with pathogenic CNVs (P-CNVs) and without pathogenic CNVs (non-carriers). **B**. Depicting the found Pearson correlation between the number of large negative deviation scores (Z<-2) and the explained variance of the first principal component of cognition, which is usually called general cognitive ability.

Discussion

Individualized risk profiles through normative models - The aim of this study was to charachterize individualized risk profiles for participants with pathogenic CNVs related to neurodevelopmental and psychiatric disorders. In this endeavor, we are the first to concurrently map individualized brain deviation scores in individuals carrying pathogenic CNVs. Our study took advantage of the UK Biobank, enabling us to analyze multiple pathogenic CNVs concurrently. We employed a normative modeling approach to generate individual difference maps for both IDPs and detailed Jacobian measures indicating volumetric changes at the voxel level and the level of the individual. Initially, we established normative reference models from IDPs to identify the neuroimaging modality most influenced by the pathogenic CNVs. Our findings highlighted that structural measures appeared more informative. Following this, we utilized a whole-brain Jacobian model to map deviation scores with voxel-wise precision.

In the introduction, we posed two main hypotheses: (i) individuals with a CNV related to cognitive deficits or neurodevelopmental disorders will have larger deviation scores compared to a reference model in certain brain areas and (ii) that the patterns of

deviation across brain regions will be highly variable across these pathogenic CNV carriers. To confirm these hypotheses, we delved deeper by analyzing the deviations from the reference model's normative trajectories, by binning the deviation scores. We counted the number of extreme volume contractions (Z<-2) and expansions (Z>2), as compared to the mean and variance of the reference model. This methodology supplied us with an individualized map of extreme deviation scores. For each participant with a pathogenic CNV, this allowed us to pinpoint the brain regions showcasing the most pronounced deviations. Expanding on this, we counted the total positive and negative deviation counts, and juxtaposed it against the broader population's risk scores. As we expected, certain participants with a pathogenic CNV displayed an elevated number of positive or negative deviations compared to the general population, while others demonstrated deviation scores that aligned more with the 'norm'.

Finally, we examined how the deviation scores are associated to general cognitive ability. We chose cognitive ability because it is consistently linked to different mental conditions[47], making it a useful gauge of mental health, and allowing us to establish a biologically meaningful connection between brain deviation scores and behavior. Our analysis revealed a significant correlation between the count of negative deviations and general cognitive ability.

Using normative models to understand how pathogenic CNVs affect the brain hasmimportant benefits. It helps uncover hidden insights that group analyses might miss. Importantly, this approach focuses on individual differences, so it's not limited to common CNVs but can be applied on rarer variations, making it possible to do research for smaller groups with uncommon genetic variants. For each person with a common or rare pathogenic CNV, we can identify the specific brain areas where they differ significantly from the norm. This allows us to see which brain regions are affected in each case, going beyond studies that need large sample sizes and can only look at more common CNVs.

Individualized risk profiles for 1q21.1 distal CNV

We generated individualized risk profiles for both carriers of the 1q21.1 distal deletion and duplication from the deviation scores. This CNV exhibits a remarkably diverse range of traits[48,49], which aligns with our proposition of an individualized approach. Notably, 1q21.1 CNV is associated with several, different neurodevelopmental disorders[2,4]. For participants with a 1q21.1 deletion or duplication, we counted instances of extreme positive and negative deviation scores, contrasting these frequencies with the broader population. Subsequently, we generated brain maps alongside these deviation scores to highlight the regions of the brain where deviations were most prominent. We aggregated the deviation scores across participants with duplications and separately for those with deletions. This allowed us to identify brain regions where the effects converged across individuals and we can then subsequently compare it with previous literature that use case-control setups.

For the Jacobian normative model, we quantify individual deviations from the mean volumetric change for a specific voxel. In this context, negative deviations refer to instances where certain brain regions show more volume contractions relative to the mean value of the Jacobian normative model. Put simply, these deviations might indicate that a specific brain region had a larger volume originally than what the normative model predicts for a typical voxel. The Jacobian image subsequently corrects for these differences. Positive deviations refer to cases where certain brain regions exhibit more volume expansions relative to the mean value of the Jacobian normative model. These deviations could

represent an original lower brain volume in certain voxels compared to the predicted or typical voxel. Interestingly, individuals with a 1q21.1 duplication showed more negative deviations, thus these individuals on average have more volume contractions than expected by the model, reflecting larger intracranial volume and macrocephaly[36] in 1q21.1 distal duplication carriers. In contrast, 1q21.1 deletion carriers showed more positive deviations, which indicates that these participants have relatively more voxels with a lower brain volume compared to the mean of the population, which might reflect their smaller intracranial volume and microcephaly[36]. In other words, we identified more positive deviations scores in deletion carriers and more negative deviation score in duplication carriers, reflecting previous findings of dosage effects on the brain of the 1q21.1 distal carriers.

Previous literature has also revealed various effects associated with the 1q21.1 distal CNV on the brain, including positive dosage effects on ICV and total cortical surface area, particularly in the frontal and cingulate cortices, and negative dosage effects on caudate and hippocampal volumes[36]. Another study found higher intraindividual variability in brain structure in 1q21.1 distal CNV carriers, with distinct regional effects on cortical surface area and thickness. Additionally, 1q21.1 distal deletion carriers exhibited reduced global cortical surface area, impacting primarily frontal and association cortices[21]. Moreover, this CNV is linked to a high prevalence of micro- and macrocephaly in deletion and duplication carriers, respectively[48,49]. From our results we can see that the dosage effects of the 1q21.1 distal carriers remains the same. Meaning that a duplication of the copy number of the 1q21.1 region is associated with more negative deviations in the Jacobian, which indicate an increase in volume of certain brain structural features and the 1q21.1 deletion is associated with more positive deviations in the Jacobian, indicating a relative decrease in volume of certain brain regions.

In general, our study shows different regions that are implicated for different participants, for example the cerebellum, brainstem and the pallidum show large negative deviations for certain 1q21.1dup carriers. For certain 1q21.1del CNV carriers the caudate and accumbens show significant positive deviations. A recent multivariate analysis of eight CNVs revealed that the cingulate gyrus, insula, supplementary motor cortex, and cerebellum were the top regions contributing to shared alterations across the CNVs[46]. This overlaps with our findings that hightlight the cerebellum in several 1q21.1 distal carriers. Likewise, in our study, for certain 1q21.1del CNV carriers the caudate and accumbens show significant positive deviations, also overlapping with previous findings[36]. The reason the implicated regions we identify might differ slightly from previous studies could be that we focused on individual variations from the norm of the population, rather than comparing the average differences between cases and controls. Moreover, although there is some overlap between studies, our study group might represent a set of CNV carriers with a somewhat different profile of those studied previously.

Towards Personalized Psychiatry: Individualized Risk Profiles and Beyond - When interpreting the normative model outputs, a common pitfall is to default to a case-control thinking paradigm. This interpretation often categorizes individuals into groups, emphasizing group patterns or seeking group effects instead of individual-level results. While brain deviation score maps can be superimposed to uncover commonalities among subjects with identical pathogenic CNVs, it is not a necessity. In our pursuit of understanding pathogenic CNVs and their effects on mental disorders, it is essential to recognize that we

cannot solely rely on aggregated group level data. While grouping subjects that exhibit similar behavioral phenotypes or possess the same CNV can provide insights into convergence points, this approach overlooks the diversity in the effects of pathogenic CNVs on brain structures and behaviors. For instance, those with a CNV linked to cognitive deficits might range from typical cognitive functioning to severe impairment[39,40,50]. Arguably, the starting point should be individual patient risk profiles, including all the known risk factors. Once we curate these profiles against a reference population, we can start to understand the implications of pathogenic CNVs at an individual level. In this paper, we made a starting point towards understanding the individual implications of pathogenic CNVs on several brain phenotypes.

Adopting an "individual patient first" approach reshapes our perspective on psychiatry, emphasizing that no two mental disorders are truly identical. Traditional psychiatric diagnoses have been classified into separate mental disorders, each presumed to have distinct origins and symptomologies. However, we have come to realize that such boundaries are far from clear-cut. Most patients with one diagnosis often have one or more comorbid conditions[22]. This wide-ranging clinical manifestation, coupled with multifactorial etiological risk factors and comorbidities, underscores that most psychiatric disorders do not correspond to single disease entities. Although initiatives like the Research Domain Criteria (RDoC)[51] have aimed to transition to more dimensional approaches, current categorizations like the DSM-V persist, even amidst challenges such as coexisting conditions and the complexity of categorizing patients.

Addressing Complexities in Genetic Studies

One of the primary challenges in human genetic research is constructing a cohort that delves into rare variants while accounting for multiple genetic and environmental factors. Our study acknowledges that while investigating CNVs, there may be additional influencing factors not yet fully characterized, for instance the remaining genome. Furthermore, to create a comprehensive cognitive health risk profile, it is essential to consider the dynamic interactions between genetics, the brain, and the environment. Research indicates that environmental factors can lead to molecular “scars", impacting brain function over time and contributing to disorders like schizophrenia[20]. Epigenetic pathways can offer a valuable perspective for studying these factors. While initial efforts to understand brain-environment-genetic interactions have started[52], an important question remains what the role of CNVs are in this complex picture.

Finally, fully mapping the effects of pathogenic CNVs demands a temporal perspective. Different pathogenic CNVs, when combined with environmental influences, could affect individuals differently across various life stages. Ongoing efforts to gather extensive longitudinal samples for developmental periods promise to enrich our normative models significantly[26]. This expansion will demonstrate critical time windows, shedding light on the complex interplay between genetics, environment, and brain function.

In conclusion, addressing these complexities is crucial for advancing our understanding of genetic influences on mental health, requiring a multi-dimensional approach that integrates genetics, environment, and developmental aspects. The present study marks an initial step toward unraveling the impacts of pathogenic CNVs on the brain at an individualized level.

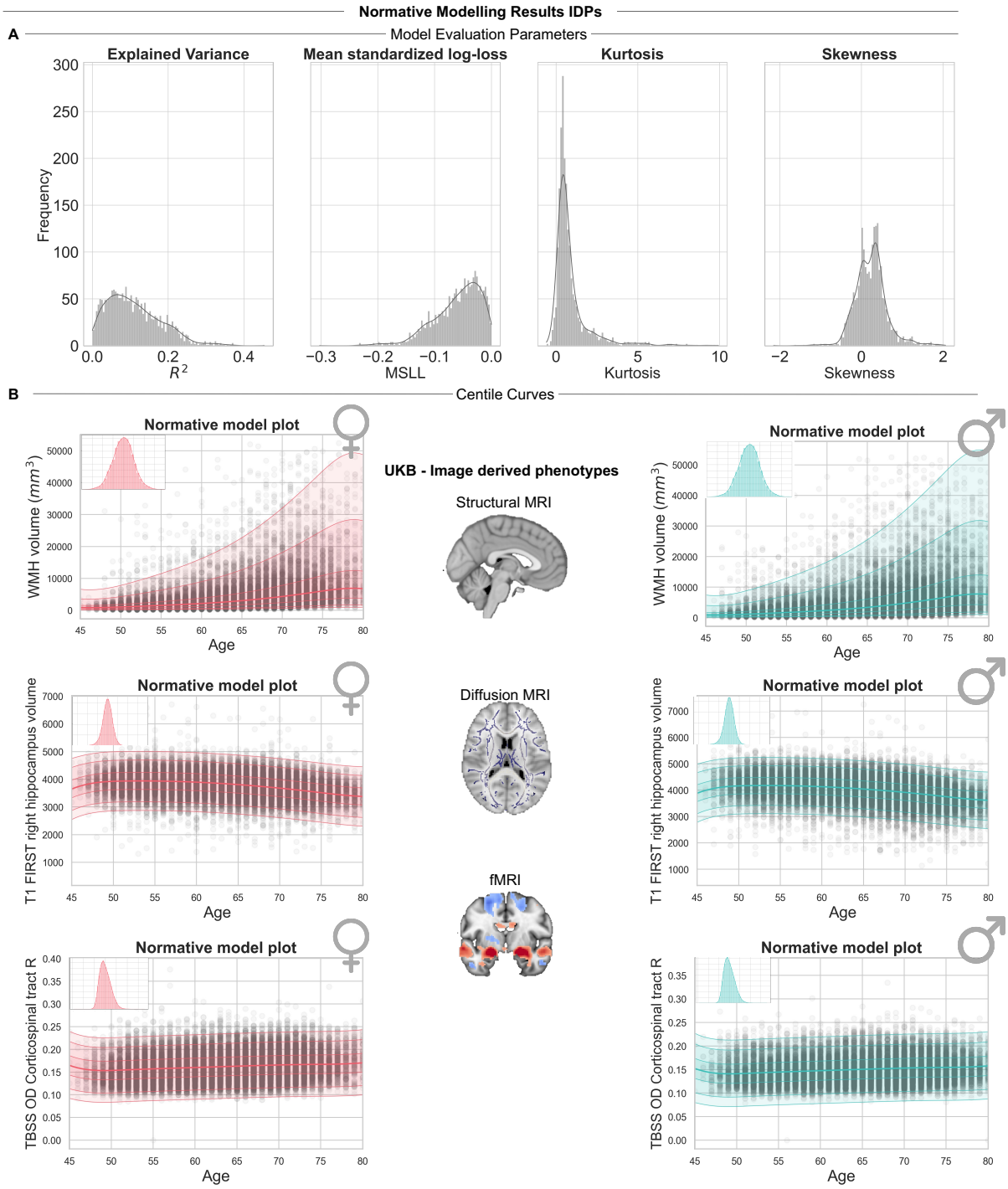
Acknowledgements This research was supported by grants from the European Research Council (ERC, grant "MENTALPRECISION" 10100118) and the Dutch Organisation for Scientific Research (VIDI grant 016.156.415). This research has been conducted using the UK Biobank resource under application number 23668. Rune Bøen and Ida Sønderby are supported by the Research Council of Norway (#223273) and South-Eastern Norway Regional Health Authority (#2020060). In addition, Ida Sønderby is supported by the European Union's Horizon2020 Research and Innovation Programme (CoMorMent project; Grant #847776) and Kristian Gerhard Jebsen Stiftelsen (SKGJ-MED-021).

Conflict of interest The authors declare that they have no conflict of interest.

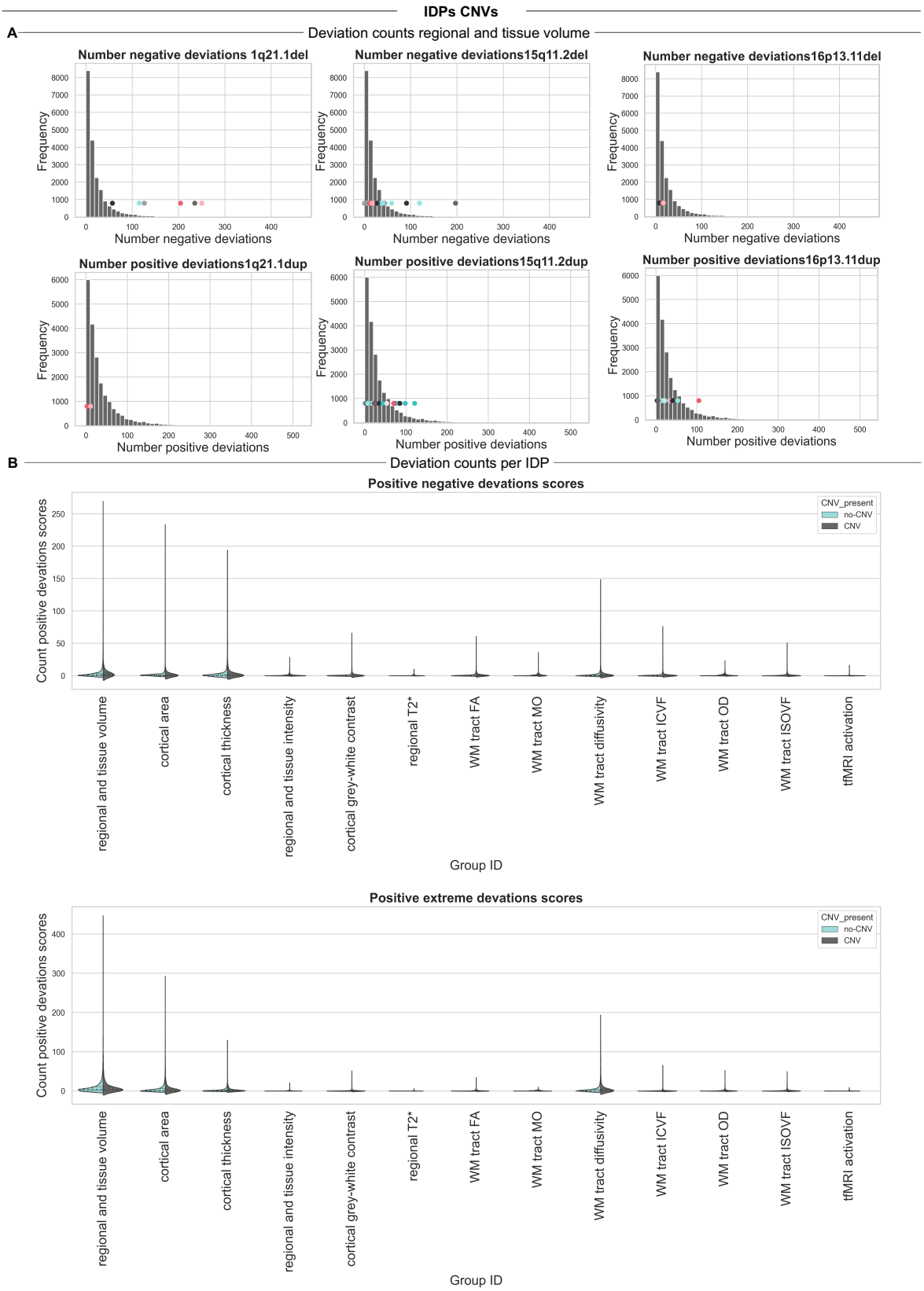
Supplemental Figures and tables

Supplemental Table S1 | Demographics across sites for the Jacobian model.

Site	N	Sex (F%/M%)	Mean age (SD)	Age range
Cam-CAN	656	50.6/49.4	54.93 (18.60)	18-89
HCP	1112	54.5/45.5	28.80 (3.70)	22-37
OASIS 3	2144	56.8/43.2	70.60 (9.52)	43-97
PNC	1296	51.9/48.1	14.37 (3.45)	8-21
UKB-11025.0	12133	52.2/47.8	62.34 (7.47)	44-80
UKB-11026.0	46	54.3/45.7	66.07 (7.34)	53-80
UKB-11027.0	2233	54.6/45.4	63.19 (7.44)	47-80

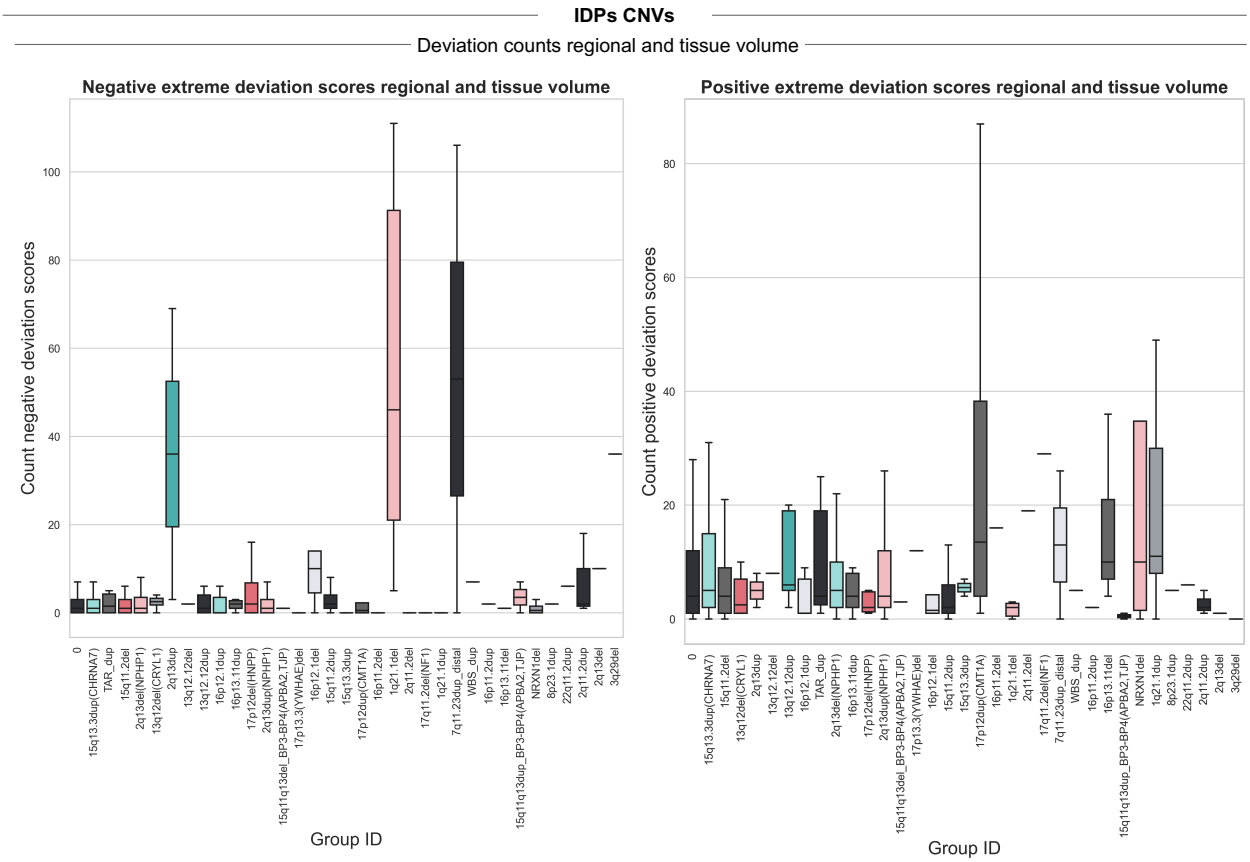


Supplemental Fig. 1 | Overview, normative modeling results IDPs. A. Performance metrics on the IDP normative models for the test set. Both skew and kurtosis serve as indicators of the model's accuracy in estimating shape via warped Bayesian Linear Regression; ideal values approach zero. **B.** Depiction of varied normative trajectories across distinct IDPs.

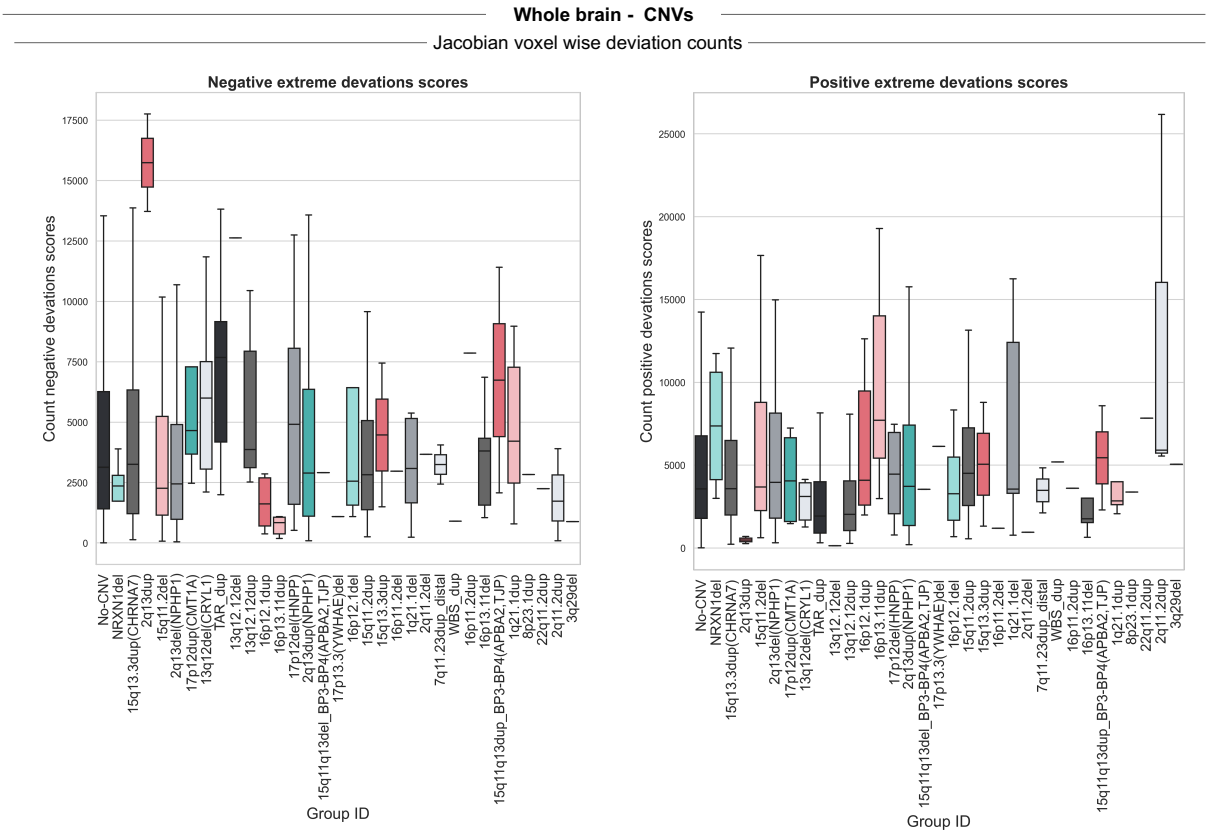


Supplemental Fig. 2 | Deviation counts per IDP category. A. Examples of extreme negative deviation counts for the regional and tissue volume IDPs highlighting participants with 1q21.1del, 15q11.2del, 16p13.11del and showing the extreme positive deviation counts for participants with CNV 1q21.1dup, 15q11.2dup, 16p13.11dup. **B**. Showing the positive and

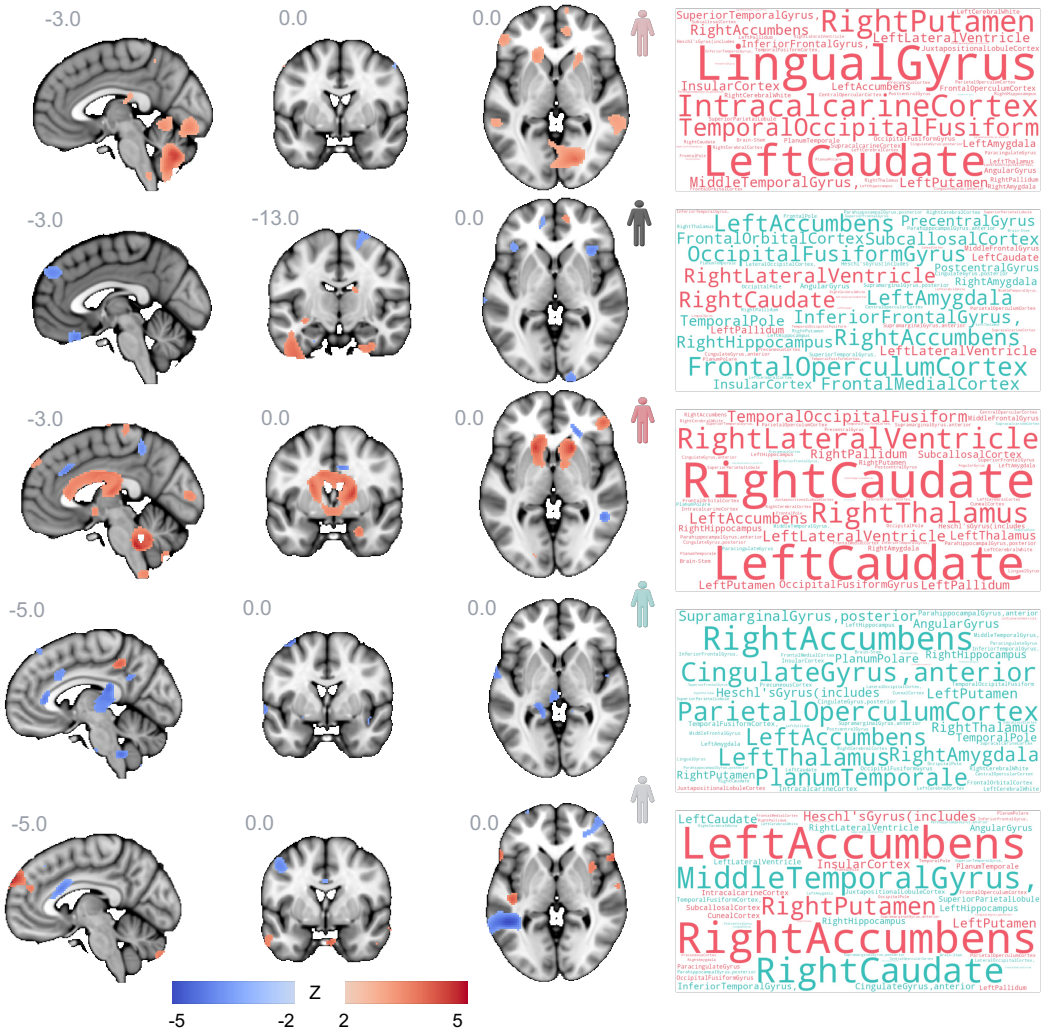
negative deviation counts for different IDP categories, comparing participants with or without a CNV.



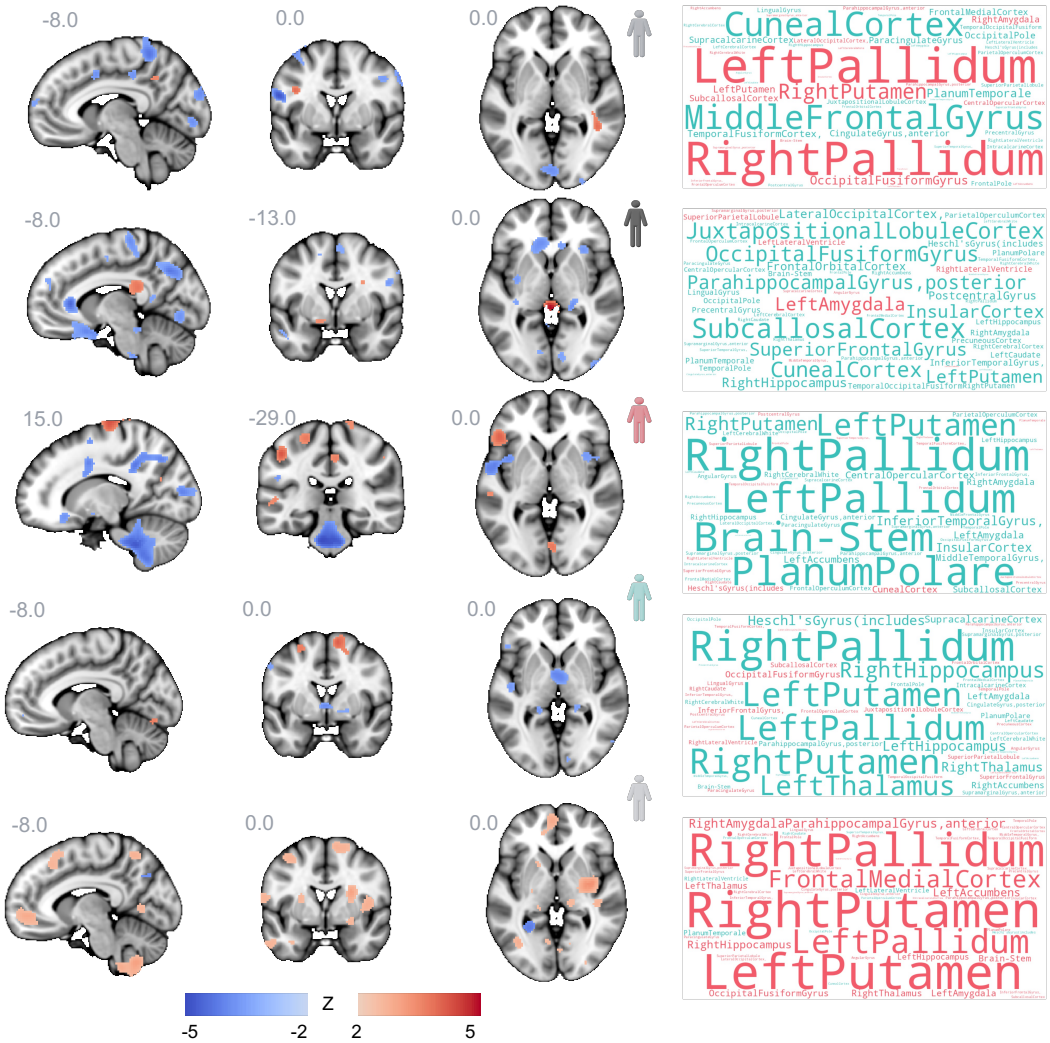
Supplemental Fig. 3 | Overview, deviation counts IDPs per CNV. Showing the positive and negative deviation counts from the IDP normative models for different pathogenic CNVs.



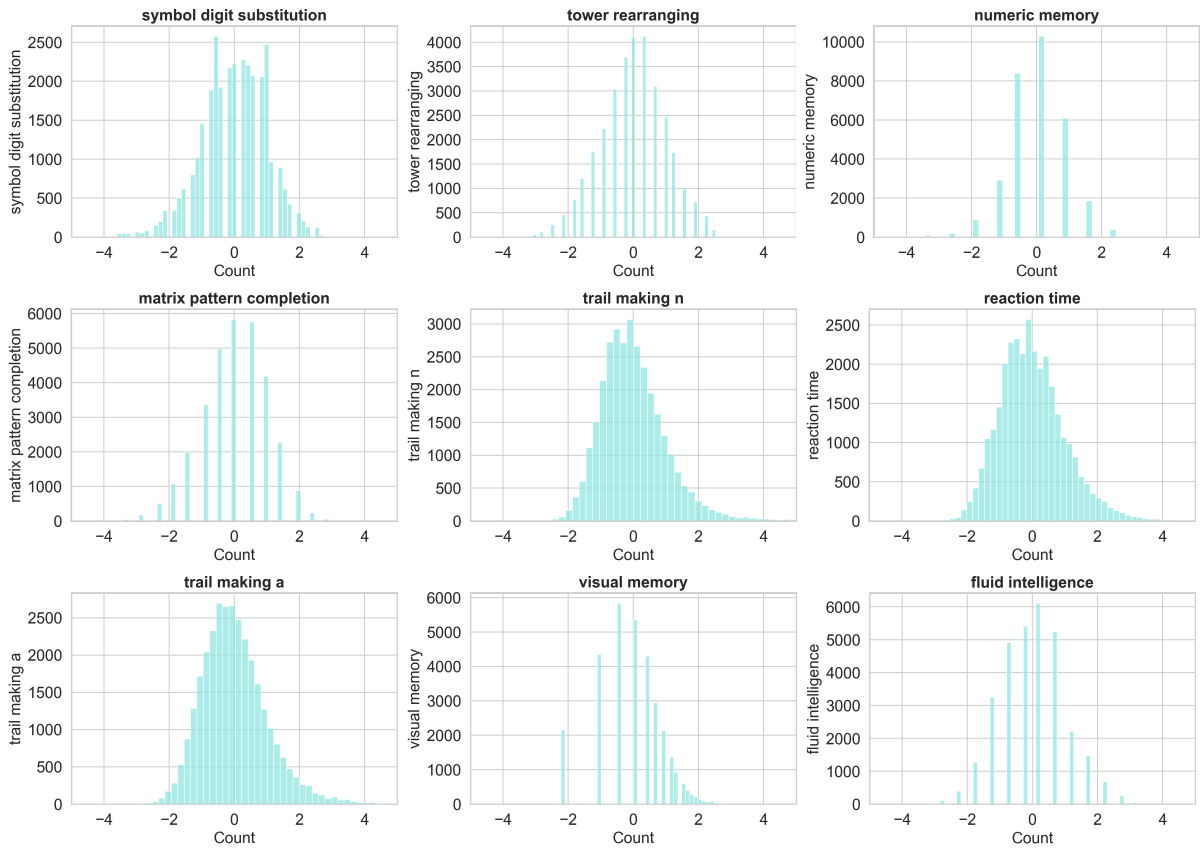
Supplemental Fig. 4 | Overview, deviation counts voxel wise Jacobian model per CNV. Showing the positive and negative deviation counts from the voxel wise Jacobian normative models for different pathogenic CNVs.



Supplemental Fig. 5 | Individual brain map and deviation scores. Showing the individual positive and negative deviation scores for all five participants with a 1q21.1 distal deletion, and the accompanying word clouds, indicating the mean z-value for each brain area, using the Harvard cortical and subcortical atlas.



Supplemental Fig. 6 | Individual brain map and deviation scores. Showing the individual positive and negative deviation scores for all five participants with a 1q21.1 distal duplication, and the accompanying word clouds, indicating the mean z-value for each brain area, using the Harvard cortical and subcortical atlas.



Supplemental Fig. 7 | Overview cognitive tests UK Biobank.

Reference cohort assembly

Data - Image derived phenotypes

In this study, two distinct types of analyses were performed using different datasets. For the first analysis, a dataset containing the image derived phenotypes (IDPs) was used. In alignment with prior research, these IDPs underwent preprocessing using FUNPACK[53], an automated toolkit for normalization, parsing and cleaning, developed at the Wellcome Centre for Integrative Neuroimaging. The set of IDPs comprises three primary imaging modalities: structural, functional, and diffusion MR data. Encompassed within these IDPs are diverse metrics, ranging from global measurements like total brain volume to more intricate assessments such as interregional connectivity within the brain. In total the data of 44,456 participants from the UK biobank and 2084 IDPs were used in the analysis.

Data - Whole brain voxel-wise Jacobian model

The subsequent analysis involved constructing a whole-brain voxel wise model, which entailed combining the T1-weighted anatomical images from seven distinct sites to form a comprehensive normative sample. A detailed breakdown of each site's demographics, including factors such as site, gender, and age, can be found in Supplementary Table 1. All the data employed in this study were sourced from publicly accessible repositories: Cam-CAN[54], HCP[55], Oasis[56], PNC[57], and UK Biobank[34]. For the UK Biobank, exclusively data from the initial imaging visit were included. Visual inspection was performed to ensure data quality[24], resulting in a total of 19,620 participants with MRI scans passing the quality control assessment. The data were preprocessed with a registration to a standard spatial reference using FSL's FLIRT and FNIRT, for further detail see[58]. Within the framework of a

voxel-based model, we opted for the Jacobian determinants. Jacobian determinant images derived from the non-linear image registration to the MNI152 space indicate the degree of local volume expansion or contraction in various regions of the brain when compared to a standard reference brain template (MNI152 space). These images are used in neuroimaging to analyze and visualize how brain structures change in size or shape in relation to the chosen reference template. A Jacobian determinant measures the local expansion or contraction factor at each voxel in the brain. Values higher than one indicate volume expansion in that region, while values below one indicate volume contraction. The magnitude of the value can also provide information about the degree of change. This approach removes the need for arbitrary distinctions between white and gray matter and has exhibited robust correlations with specific demographic variables[35].

CNVs UK Biobank

We identified CNVs based on the returned dataset from Crawford et al.[41]. For details on the preprocessing and calling pipeline and quality control of the CNVs see[39,40]. Briefly their quality control included included genotypic call rate <0.96, waviness factor of <−0.03 and >0.03, >30 CNVs per person and log R ratio s.d. of >0.35. Among the participants with QCed neuroimaging , 375 individuals.

Normative model formulation

We employed a Bayesian linear regression model (BLR) with likelihood warping for the normative analysis. For an in-depth explanation of the mathematical framework, refer to[38]. Python version 3.8 and PCNtoolkit version 0.28 were used for all statistical analyses. Our dataset was divided into a 50-50 train-test split, with the test set including most participants with pathogenic CNVs. Covariates encompassed age, binary gender, and binary site ID within the covariance matrix. Morphometric variation models were estimated using Jacobian determinants from non-linear registration, utilizing the BLR algorithm from PCNtoolkit. This algorithm employed Sinarcsinh likelihood warping and Powell optimization. To concisely present the method: we take $\boldsymbol{Y} = (y_{nd}) \in \mathbb{R}^{N\times D}$ with y_{nd} the d-th neuroimaging variable, voxel or IDP, of the n-th subject. We collected the covariates into one matrix $\boldsymbol{X} = (x_{nm}) \in \mathbb{R}^{N\times M}$, where x_{nm} is the m-th covariate of the n-th subject. To keep the notation concise we will concentrate on one variable, labelled d and drop the subscript. Thus, for every IDP we denote $\boldsymbol{y} = (y_1, \dots, y_N)^T$ and take the set of independent variables $\boldsymbol{x}_n = (x_{n1}, \dots, x_{nM})^T$. For every subject we specified the model as follows:

$$\varphi(y_n) = \boldsymbol{w}^T \phi(x_n) + \epsilon$$

Where, $\mathbf{w}^{\mathrm{T}}$ is the estimated vector of weights, $\phi(\mathbf{x})$ a basis expansion of the covariate vector x_n. In our case, a cubic B-spline basis expansion with 5 evenly spaced knots was chosen. Empirically, this was enough to capture the curvature in space caused by the age covariate. $\epsilon_s = \mathcal{N}(0, \beta^{-1})$ is a Gaussian noise distribution with mean zero and noise precision term β (the reciprocal of the variance). The likelihood warping φ, a sinarcsinh function, accommodates non-Gaussianity:

$$\varphi(y_n, \boldsymbol{\gamma})_{Sinarcsinh} = \sinh\left(b * \operatorname{arcsinh}(y_n) - a\right)$$

With $\boldsymbol{\gamma} = (a, b)$ the identified parameters for the warping function. We captured the site variation using a fixed-effects model, according to [21,39]. Powell's conjugate direction method minimized the negative log-likelihood during optimization. Subsequently, z-scores in the warped space were calculated as:

$$Z_{nd} = \frac{y_{nd} - \hat{y}_{nd}}{\sqrt{\sigma_d^2 + (\sigma_*^2)_d}}$$

Where, y_{nd} is the true response, $\hat{y}_{nd}$ is the predicted mean, σ_d^2 is the estimated noise variance (reflecting uncertainty in the data), and $(\sigma_*^2)_d$ is the variance attributed to modeling uncertainty, for the full derivations see[63]. Model evaluation employed criteria including R^2, mean squared log-loss (MSLL), skewness, and kurtosis, providing insights into central tendency, variance, warping function performance, and overall fit.

Cognitive phenotype prediction

We used the cognitive data extracted from the UK Biobank dataset to explore associations between cognitive attributes and deviations from the normative model. These phenotypes come from nine cognitive tests available in the UK Biobank, as shown in Supplementary Fig. 7. These tests were conducted via a touchscreen questionnaire and encompassed dimensions such as numerical memory, reaction time, fluid intelligence, visual memory, prospective memory, executive function, declarative memory, and non-verbal reasoning[60]. Further details about these cognitive assessments in the UK Biobank can be found in[61]. To reduce the complexity of cognitive tests while retaining essential information, we employed principal component analysis (PCA) on the cognitive measurements. This approach helps capture a latent factor often associated with overall cognitive ability or the 'g-factor'[45]. By relating this general cognitive ability to extreme z-deviations ($|Z|>2$) using Pearson's correlation, we investigated potential associations between cognitive performance and deviations from the norm.

References

1. Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**, 1223–1241 (2012).
2. Sanders, S. J. *et al.* A framework for the investigation of rare genetic disorders in neuropsychiatry. *Nat Med* **25**, 1477–1487 (2019).
3. Gudmundsson, O. O. *et al.* Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. *Transl Psychiatry* **9**, 1–9 (2019).
4. 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).

5. Kirov, G. *et al.* The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry* **75**, 378–385 (2014).
6. Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nat Genet* **43**, 838–846 (2011).
7. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).
8. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
9. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
10. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
11. Bearden, C. E. & Forsyth, J. K. The many roads to psychosis: recent advances in understanding risk and mechanisms. *F1000Res* **7**, F1000 Faculty Rev-1883 (2018).
12. Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022).
13. Lionel, A. C. *et al.* Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci Transl Med* **3**, 95ra75 (2011).
14. Moreau, C. A. *et al.* Mutations associated with neuropsychiatric conditions delineate functional brain connectivity dimensions contributing to autism and schizophrenia. *Nat Commun* **11**, 5272 (2020).
15. Douard, E. *et al.* Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome. *Am J Psychiatry* **178**, 87–98 (2021).
16. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505**, 361–366 (2014).

17. Lee, P. H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).
18. Andrews, T. *et al.* Gene networks underlying convergent and pleiotropic phenotypes in a large and systematically-phenotyped cohort with heterogeneous developmental disorders. *PLoS Genet* **11**, e1005012 (2015).
19. Moreau, C. *et al.* The general impact of haploinsufficiency on brain connectivity underlies the pleiotropic effect of neuropsychiatric CNVs. 2020.03.18.20038505 Preprint at https://doi.org/10.1101/2020.03.18.20038505 (2020).
20. Richetto, J. & Meyer, U. Epigenetic Modifications in Schizophrenia and Related Disorders: Molecular Scars of Environmental Exposures and Source of Phenotypic Variability. *Biological Psychiatry* **89**, 215–226 (2021).
21. Boen, R. *et al.* Beyond the Global Brain Differences: Intra-individual Variability Differences in 1q21.1 Distal and 15q11.2 BP1-BP2 Deletion Carriers. *Biol Psychiatry* S0006-3223(23)01530–5 (2023) doi:10.1016/j.biopsych.2023.08.018.
22. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol Psychiatry* **80**, 552–561 (2016).
23. Marquand, A. F. *et al.* Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry* **24**, 1415–1424 (2019).
24. Rutherford, S. *et al.* Charting brain growth and aging at high spatial precision. *eLife* **11**, e72904 (2022).
25. Rutherford, S. *et al. Evidence for Embracing Normative Modeling*. http://biorxiv.org/lookup/doi/10.1101/2022.11.14.516460 (2022) doi:10.1101/2022.11.14.516460.

26. Bučková, B. R. *et al.* Using normative models pre-trained on cross-sectional data to evaluate longitudinal changes in neuroimaging data. 2023.06.09.544217 Preprint at https://doi.org/10.1101/2023.06.09.544217 (2023).
27. Bethlehem, R. a. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
28. Wolfers, T. *et al.* Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry* **75**, 1146–1155 (2018).
29. Wolfers, T. *et al.* Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol Med* **50**, 314–323 (2020).
30. Zabihi, M. *et al.* Fractionating autism based on neuroanatomical normative modeling. *Transl Psychiatry* **10**, 1–10 (2020).
31. Pinaya, W. H. L. *et al.* Using normative modelling to detect disease progression in mild cognitive impairment and Alzheimer's disease in a cross-sectional multi-cohort study. *Sci Rep* **11**, 15746 (2021).
32. Owen, M. J. & O'Donovan, M. C. Schizophrenia and the neurodevelopmental continuum:evidence from genomics. *World Psychiatry* **16**, 227–235 (2017).
33. Fraza, C., Zabihi, M., Beckmann, C. F. & Marquand, A. F. The Extremes of Normative Modelling. 2022.08.23.505049 Preprint at https://doi.org/10.1101/2022.08.23.505049 (2022).
34. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400–424 (2018).
35. Monté-Rubio, G. C., Falcón, C., Pomarol-Clotet, E. & Ashburner, J. A comparison of various MRI feature types for characterizing whole brain anatomical differences using linear pattern recognition methods. *NeuroImage* **178**, 753–768 (2018).

36. Sønderby, I. E. *et al.* 1q21.1 distal copy number variants are associated with cerebral and cognitive alterations in humans. *Transl Psychiatry* **11**, 1–16 (2021).

37. Modenato, C. *et al.* Lessons Learned From Neuroimaging Studies of Copy Number Variants: A Systematic Review. *Biological Psychiatry* **90**, 596–610 (2021).

38. Fraza, C. J., Dinga, R., Beckmann, C. F. & Marquand, A. F. Warped Bayesian linear regression for normative modelling of big data. *Neuroimage* **245**, 118715 (2021).

39. Kendall, K. M. *et al.* Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biological Psychiatry* **82**, 103–110 (2017).

40. Kendall, K. M. *et al.* Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *The British Journal of Psychiatry* **214**, 297–304 (2019).

41. Crawford, K. *et al.* Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *J Med Genet* **56**, 131–138 (2019).

42. Clinical phenotype of the recurrent 1q21.1 copy-number variant - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S1098360021043306?via%3Dihub.

43. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).

44. Green, E. K. *et al.* Copy number variation in bipolar disorder. *Mol Psychiatry* **21**, 89–93 (2016).

45. Sripada, C., Angstadt, M., Rutherford, S., Taxali, A. & Shedden, K. Toward a “treadmill test” for cognition: Improved prediction of general cognitive ability from the task activated brain. *Human Brain Mapping* **41**, (2020).

46. Modenato, C. *et al.* Effects of eight neuropsychiatric copy number variants on human brain structure. *Transl Psychiatry* **11**, 399 (2021).

47. Trivedi, J. K. Cognitive deficits in psychiatric disorders: Current status. *Indian Journal of Psychiatry* **48**, 10–20 (2006).

48. Brunetti-Pierri, N. *et al.* Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* **40**, 1466–1471 (2008).

49. Bernier, R. *et al.* Clinical phenotype of the recurrent 1q21.1 copy-number variant. *Genet Med* **18**, 341–349 (2016).

50. 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).

51. Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders | American Journal of Psychiatry. https://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.2010.09091379?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.

52. Xu, J. *et al.* Effects of urban living environments on mental health in adults. *Nat Med* 1–12 (2023) doi:10.1038/s41591-023-02365-w.

53. McCarthy, P. Funpack. (2020) doi:10.5281/zenodo.3761702.

54. Taylor, J. R. *et al.* The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage* **144**, 262–269 (2017).

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

56. Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C. & Buckner, R. L. Open Access Series of Imaging Studies (OASIS): Longitudinal MRI Data in Nondemented and Demented Older Adults. *J Cogn Neurosci* **22**, 2677–2684 (2010).

57. Satterthwaite, T. D. *et al.* Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *Neuroimage* **86**, 544–553 (2014).

58. Holz, N. E. *et al.* A stable and replicable neural signature of lifespan adversity in the adult brain. *Nat Neurosci* 1–10 (2023) doi:10.1038/s41593-023-01410-8.

59. Kia, S. M. *et al.* Federated Multi-Site Normative Modeling using Hierarchical Bayesian Regression. *bioRxiv* 2021.05.28.446120 (2021) doi:10.1101/2021.05.28.446120.

60. Fawns-Ritchie, C. & Deary, I. J. Reliability and validity of the UK Biobank cognitive tests. *PLOS ONE* **15**, e0231627 (2020).

61. Lyall, D. M. *et al.* Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLOS ONE* **11**, e0154222 (2016).