1	High-effect gene-coding variants impact cognition, mental well-being, and neighborhood
2	safety substrates in brain morphology
3	
4	Jakub Kopal ^{1,2} , Guillaume Huguet ³ , Justin Marotta ^{1,2} , Shambhavi Aggarwal ^{1,2} , Nicole Osayande ^{1,2} ,
5	Kuldeep Kumar ³ , Zohra Saci ³ , Martineau Jean-Louis ³ , Xiaogian J Chai ⁴ , Tian Ge ⁵ , B.T. Thomas Yeo ^{6,7,8} ,
6	Paul M Thompson ⁹ . Carrie E Bearden ¹⁰ . Ole A. Andreassen ^{11,12} . Sébastien Jacquemont ^{3,13} .
7	Danilo Bzdok ^{1,2,14, *}
8	
9	¹ Department of Biomedical Engineering, Faculty of Medicine, McGill University, Montreal, Ouebec, Canada
10	² Mila - Quebec Artificial Intelligence Institute. Montréal. Quebec. Canada
11	³ Centre de recherche CHU Sainte-Justine, Montréal, Quebec, Canada
12	⁴ Department of Neurology and Neurosurgery, McGill University, USA
13	⁵ Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General
14	Hospital, Boston, Massachusetts, United States of America
15	⁶ Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore.
16	⁷ Centre for Sleep & Cognition & Centre for Translational Magnetic Resonance Research, Yong Loo Lin School of
17	Medicine, Singapore, Singapore.
18	⁸ N.1 Institute for Health & Institute for Digital Medicine, National University of Singapore, Singapore,
19	Singapore.
20	⁹ Imaging Genetics Center, Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC,
21	Los Angeles, CA, USA
22	¹⁰ Semel Institute for Neuroscience and Human Behavior, Departments of Psychiatry and Biobehavioral
23	Sciences and Psychology, UCLA, Los Angeles, CA, USA
24	¹¹ NORMENT, Division of Mental Health and Addiction, Oslo University Hospital and University of Oslo, Oslo,
25	Norway
26	¹² KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway
27	¹³ Department of Pediatrics, University of Montréal, Montréal, Quebec, Canada
28	¹⁴ TheNeuro - Montreal Neurological Institute (MNI), McConnell Brain Imaging Centre, Faculty of Medicine,
29	McGill University, Montreal, Quebec, Canada
30	

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

31 Corresponding author

32 * Danilo Bzdok, <u>danilo.bzdok@mcgill.ca</u>

33

34 Abstract

35 Our genetic makeup, together with environmental and social influences, shape our brain's 36 development. Yet, the imaging genetics field has struggled to integrate all these modalities to 37 investigate the interplay between genetic blueprint, environment, human health, daily living skills and 38 outcomes. Hence, we interrogated the Adolescent Brain Cognitive Development (ABCD) cohort to outline the effects of rare high-effect genetic variants on brain architecture and corresponding 39 40 implications on cognitive, behavioral, psychosocial, and socioeconomic traits. Specifically, we 41 designed a holistic pattern-learning algorithm that quantitatively dissects the impacts of copy number 42 variations (CNVs) on brain structure and 962 behavioral variables spanning 20 categories in 7,657 43 adolescents. Our results reveal associations between genetic alterations, higher-order brain networks, 44 and specific parameters of the family well-being (increased parental and child stress, anxiety and 45 depression) or neighborhood dynamics (decreased safety); effects extending beyond the impairment 46 of cognitive ability or language capacity, dominantly reported in the CNV literature. Our investigation 47 thus spotlights a far-reaching interplay between genetic variation and subjective life quality in 48 adolescents and their families.

49 Introduction

50 Our genes collectively provide the blueprint of our biological makeup upon which the dynamic 51 interplay of environmental influences, social interactions, and developmental experiences unfolds. 52 The genetic architecture contributes directly and indirectly to the wiring of brain circuits and provides the foundation of behavior repertoire manifestation^{1,2}. By understanding genetic underpinnings, we 53 54 can unravel the origins of individual differences in cognitive processes and behaviors, which in turn 55 offers insights into both adaptive capacities and developmental vulnerabilities³. Identifying biological 56 determinants behind brain organization and behavioral differentiation necessitates an integrative 57 approach that cuts across an array of disciplines. Nevertheless. neuroimaging genetics, psychiatric 58 genetics, and environmental factor studies have been conducted in isolated silos. Without the 59 integration of these disciplines, the full extent of genetic influences on human health will remain 60 concealed.

61

62 The genetic underpinnings have been traditionally studied through genome-wide association 63 studies (GWAS). However, GWAS have been restricted to common variants which mainly reside in 64 non-coding regions and exert only small effects on phenotypes⁴. Compared to incumbent single 65 nucleotide polymorphism analyses in GWAS, protein-coding copy number variations (CNVs) represent 66 rare and consequential genome-wide perturbations invariably leading to a large decrease or increase 67 in gene expression - a currently under-exploited genetic phenomenon. This class of genetic variation 68 is defined as either a deletion or duplication of sequences of nucleotides more than 1,000 base pairs long^{5,6}. Moreover, given recently accumulating evidence, many protein-coding CNVs are now being 69 understood to exert body-wide implications^{7,8} as well as increase the risk of neurodevelopmental 70 71 disorders, including intellectual disability, autism spectrum disorder or schizophrenia⁹⁻¹¹. Since 72 protein-coding CNVs are cumulatively relatively frequent in the population and have the potential for 73 substantial effects on a given phenotype, they represent a potent imaging-genetics tool, which we will 74 here deploy for interrogating the effects of genetic modifications on brain physicality and behavioral 75 differentiation in adolescents. 76

77 Adolescence is a critical juncture where the seeds of mental health and well-being are sown. 78 During this period, brain circuits and behavioral tendencies undergo dynamic changes shaped by genetic factors, environmental influences, and their interactions^{12,13}. Importantly, adolescence is also 79 80 a life stage during which symptoms of numerous psychiatric disorders become apparent¹⁴. Recent 81 findings underscore the necessity of adopting a multidimensional and interdisciplinary approach that 82 cuts across sociology, psychology, and biology, conventionally studied in isolation. Such a holistic 83 perspective is essential for a more nuanced understanding of the intricate interplay of genetic, socioeconomic, and environmental factors influencing healthy children's development¹⁵. By integrating 84 85 information from cognitive assessments, genetic information, and socio-environmental measures, we 86 can identify potential risk factors as well as unveil protective elements contributing to resilience in individuals navigating the complexities of adolescence^{16,17}. Consequently, the analysis of CNVs in 87 88 adolescents is uniquely positioned to carve out important interactions between our genetic heritage, 89 the environmental milieu, and the intricacies of cognitive and social development, laying the 90 groundwork for a more comprehensive approach to mental health.

91

In the present study, we leveraged under-studied rare genetic alterations (genome-wideCNVs) with strong downstream effects. We interrogated the Adolescent Brain Cognitive Development

94 (ABCD) study¹⁸, which represents one of the largest collections of brain images and genetic profiles 95 from over 10,000 children aged 9 to 11 years at baseline. Moreover, these adolescents are 96 prospectively deeply phenotyped by means of an extensive battery of cognitive, behavioral, clinical, 97 psychosocial, and socioeconomic measures. Benefiting from this comprehensive multimodal data, we 98 investigated the effects of a genomic deletion or duplication on patterns of individual subjects' brain 99 architecture linked with cognitive, behavioral, psychosocial, and socioeconomic measures in a single 100 multivariate analysis. Specifically, we first probed the carefully curated data from 7,657 children for 101 the presence of CNVs. We then deployed multivariate pattern learning algorithms in the children 102 without any CNV to estimate modes of population covariation between brain architecture, 103 represented by 148 regional atlas volumes and ~1000 behavioral variables spanning 20 rich categories. 104 Finally, we quantified the effects of deletions and duplications on the revealed canonical modes. The 105 robustness of derived modes and CNV-induced differences were assessed by cross-validation and permutation testing^{15,19}. This multidimensional and doubly multivariate framework revealing the 106 107 multi-faceted relationships between genes, brain architecture, and behavior paves the way for 108 innovation in neuroscience, genetics, and personalized medicine.

110 Results

111 Genome-wide mutations alter patterns of brain and behavior

112 We used a pattern-learning approach to analyze the impact of genome-wide CNVs in the ABCD 113 cohort by means of its uniquely deep phenotype profiling. To this end, in the group of 7,657 children 114 that passed genetic and MRI quality control, we first identified 514 children carrying at least one 115 genomic deletion and 1,472 children carrying at least one duplication that fully encompassed at least 116 one gene. The 136 children that carried both a deletion and duplication were not analyzed in this study 117 due to the limited sample size. The remaining 73% (5,535) of the children did not carry any protein-118 coding CNV larger than 50kb across the genome (Fig. 1A). These participant groups (deletions, 119 duplications, controls) showed similar proportions of sex (percentage of females: 44-48%) and 120 distributions of age (Fig. 1A).

121 Next, we zoomed in on the CNVs that we localized in the children's genetic profiles (Fig. 1B). 122 Almost 60% of deletions encompassed a single complete gene. Duplications generally encompassed 123 more affected genes than deletions, although a single-gene duplication was the most common (~30% 124 of cases). Besides the genetic profiling, the ABCD resource provides brain and behavior measurements 125 for each participant: brain measurements were represented by 148 regional brain volumes defined 126 according to the Desikan-Killiany standard atlas. Behavioral measures drew across 962 different 127 phenotypes spanning 20 categories for in-depth follow-up analyses.

128 In order to investigate how genetic mutations impact brain and behavior, we first established 129 the link between measurements of brain architecture and behavior using a single holistic multivariate 130 model. Specifically, we brought to bear a partial least squares (PLS) model that maximizes the 131 covariation between the weighted set (linear combination) of sociodemographics, family well-being, 132 physical characteristics, or behavioral measures and a weighted set (linear combination) of brain 133 structure measures (Fig. 1C). The PLS model parameters were initially estimated in participants 134 without any CNV, as a reference group, to reveal the modes of covariation that reflect the general 135 population. The subject-wise expressions of each brain-behavior covariation mode (i.e., canonical 136 variables) will henceforth be called scores. In other words, these scores are calculated as a linear 137 combination (weighted sum) of the original variables with PLS weights. Each identified PLS mode can 138 thus be characterized by a set of brain and behavior scores for all subjects and for each PLS mode. 139 Using a robust protocol for cross-validation and empirical permutation testing¹⁵, we identified three 140 significant PLS modes (Sup. Fig. 1). The revealed major sources of covariation in adolescents captured 141 the ways in which brain features are intertwined with early life events, mental well-being, or 142 environment.

143 In the next step, we wished to evaluate if carrying a coding CNV led to statistically significant 144 shifts in the observed brain and behavior patterns. To that end, we devised a bootstrap validation 145 scheme that compares PLS scores between controls and CNV carriers (Sup. Fig. 2). In this CNV-control 146 comparison, deletion and duplication carriers are pitted against control participants not used to derive 147 PLS parameters in order to prevent overfitting (cf. Methods for details). The comparison was based 148 on separately testing the difference in the average behavior scores and the average brain scores. We 149 were thus able to assess mode expression differences separately for each CNV status (Fig. 1C). We 150 found deletions to impact behavior scores in all three identified modes (p-value_{mode1} = 10^{-4} , p-151 value_{mode2} = 0.014, p-value_{mode3} = 0.041). By contrast, duplications affected only the first behavioral mode (p-value_{mode1} = 10^{-4} , p-value_{mode2} = 0.151, p-value_{mode3} = 0.336). Furthermore, there was a 152 153 significant shift in brain scores for the first two modes in duplications (p-valuemode1 = 10⁻⁴, p-valuemode2 154 = 0.027, p-value_{mode3} = 0.097) and the second mode in deletions (p-value_{mode1} = 0.395, p-value_{mode2} =

0.043, p-value_{mode3} = 0.052). A sensitivity analysis demonstrated that the obtained differences were
not driven by the presence of recurrent CNVs, such as 16p11.2 or 22q11.2 (Sup. Fig. 3). These results
revealed that carrying a CNV significantly impacts the expression of patterns linking brain architecture
and diverse aspects of cognitive, psychosocial, and socioeconomic measures in our ABCD sample.
Therefore, genetic factors shape individual differences in brain-behavior correspondences in
adolescents.

161

162 Canonical modes bind patterns of whole-brain architecture with real-life functioning, mental well-

163 being and environment

164 After identifying robust deviations of brain-behavior patterns in CNV carriers at population 165 scale, we examined each revealed mode in more detail. The dominant mode portrayed the ties 166 between large-scale brain networks with sociodemographics and cognition. Specifically, we first re-167 expressed the difference in PLS scores between controls and CNV carriers using Cohen's d measure to 168 provide a standardized measure of CNV-carriership effect size. The dominant PLS mode was 169 characteristic of significantly altered behavior scores, with the shift being more prominent in CNV 170 deletions (CNV - controls Cohen's $d_{DEL} = -0.16$, $d_{DUP} = -0.11$) (Fig. 2A). To find which phenotypes play a 171 prominent role in the first mode, we calculated brain and behavior loadings. Our version of these 172 loadings was here obtained by Pearson's correlation between a respective PLS score and the original 173 measurement. As an example, each brain loading indexes the linear association strength between 174 brain region measurements and brain scores in our reference group. Among the strongest brain 175 effects, we observed the medial orbital sulcus (average of left and right hemisphere Pearson's r = 176 0.25), a part of the frontal lobe which may be involved in various cognitive functions, including 177 decision-making, emotional processing, and social cognition ²⁰. Since duplication carriers displayed 178 higher brain scores compared to controls and since medial orbital sulcus was associated with positive 179 loading (higher volume = higher score), this result pointed to increased volume in this region for 180 duplication carriers. Other strong negative loadings included the middle occipital sulcus, subcallosal 181 area, superior occipital gyrus, or right lingual gyrus (Fig. 2B). We subsequently submitted these 182 loadings to a bootstrap test to determine if they were significantly different from zero (cf. Methods). 183 This test was based on 1,000 PLS model instances built on a randomly perturbed version of our ABCD 184 participants created by sampling a participant cohort of the same sample size (with replacement). We 185 further computed the relative number of significant region effects (i.e., loadings) among each set of 186 regions belonging to one of the seven large-scale networks according to Schaefer-Yeo definitions. We 187 observed that at least 50% of the loadings in each of the seven networks were significant, highlighting 188 the brain-wide pattern of this first mode.

189 Furthermore, we inspected a broad portfolio of behavior characteristics interlocked with the 190 above-described brain-level effects. To that end, we calculated behavior loadings similarly to brain 191 loadings. The strongest loadings included family income (Pearson's r = 0.70), poverty index (Pearson's 192 r = -0.68), parental education (Pearson's r = 0.60), measures of cognitive performance (Pearson's r = -193 0.57), and also screen time or sleep duration (Pearson's r = 0.46) (Fig. 2C). To obtain a synopsis of the 194 dominant behavioral profile, we averaged absolute behavior loadings in each of the 20 categories. 195 Demographics, cognitive, and socioeconomic categories had the strongest average loadings (average 196 absolute Pearson's r > 0.22). Since CNV carriers displayed higher expression compared to controls for 197 this mode characterized by negative loading for measures of cognitive performance (lower 198 performance = higher score), these results thus point to decreased cognitive abilities and real-life

functioning, especially in deletion carriers. Collectively, the dominant canonical mode highlighted thecrosslinks between most of the brain networks and assessments of cognition and demographics.

201

202 The second PLS mode spotlighted opposing gene dosage effects on the brain structure that 203 we identified to tie into family history of mental health. Specifically, we observed significant opposing 204 brain average expressions for both deletions and duplication (Fig. 3A), which might reflect the 205 mirroring effect on brain architecture previously reported for CNVs at specific genomic loci²¹. Only 206 behavior scores in deletion carriers passed the significance testing. According to our obtained brain 207 loadings, the mirroring brain scores were mainly driven by the lingual gyrus (across-hemisphere 208 average Pearson's r = -0.38), followed by supramarginal, precentral, or lateral superior temporal gyri 209 (Fig. 3B). Despite being part of distinct brain networks, these regions were previously associated with 210 neural mechanisms supporting complex cognitive tasks, especially those involving semantic 211 processing or executive functions^{22,23}. Collectively, following the conducted bootstrap significance 212 test, most of the significant regions belonged to dorsal attention, somatomotor, and frontoparietal 213 networks (more than 67% significant regions in each network). The interactions and coordinated 214 activity of these networks are known to be essential for the efficient integration and execution of 215 complex cognitive and motor tasks²⁴.

216 The prominent deviations in behavior scores in deletion carriers can be explained by elevated 217 assessments of mental well-being as revealed by behavior loadings. Specifically, phenotypes from the 218 Child Behavior Checklist (CBCL) and the Adult Self Report (ASR) dominated the set of relevant behavior 219 loadings (Fig. 3C). Particularly, the total scores of CBCL (Pearson's r = 0.80) and ASR (Pearson's r = 0.77) 220 emerged as the two strongest loadings. They were followed by measures of both parental and child 221 anxiety, stress, depression as well as child sleep disorders. Indeed, when averaged across categories, 222 the sleep category joined child behavior and parental questionnaires as the most prominent (average 223 Pearson's r = 0.22). The combination of flagged phenotypes from both children and adult assessments 224 suggests that the second mode captures a comprehensive view of the well-being intricately tied to the 225 family system. In addition, it points towards potential dynastic effects, i.e., the impacts of (inherited) 226 genetic variants on family environments. Collectively, the second canonical mode proposed decreased 227 familial mental well-being as a prominent marker of deletion carriers.

228

229 In the third and last canonical mode, we observed the relationship of the default mode and 230 frontoparietal networks with environmental measures. Despite the mirrored effects on brain 231 structure, the only significant shift was found for behavior scores in deletion carriers (Fig. 4A). The 232 third mode was characterized by a strong contribution of the insula (average Pearson's r = 0.33) as 233 well as middle temporal and lateral superior temporal gyri (Fig. 4B). The majority of significant regions 234 based on bootstrap testing was part of the default mode network or the frontoparietal network (23% 235 and 20% of regions, respectively). These two networks belong to the multimodal end of the unimodal-236 to-multimodal characterization of large-scale brain networks. Prior research suggests that the 237 relevance of these two networks and underlying regions could imply their crucial roles in cognitive 238 flexibility and the integration of thought processes necessary for problem-solving and decision-239 making²⁵.

Examining the behavior profile in the third canonical mode highlighted variables associated with the environment (Fig. 4C). Concretely, phenotypes related to the neighborhood, such as crime reports (Pearson's r = 0.63), drug possession, violent crimes, adult offense, and feelings of safety emerged as strongly associated. These phenotypes reflect social and community dynamics, which

244 might affect the overall quality of life for individuals within that context. In summary, the third 245 significant mode revealed an alteration in how environmental differences link to higher-order 246 networks in adolescent deletion carriers.

- 247
- 248

Decoding brain and behavioral patterns: Population, genetic and temporal perspectives

249 After describing how genetic mutations influence the expression of behavior patterns, we 250 explored how characteristics of genes encompassed in CNVs shaped behavior scores. In other words. 251 for each individual with a deletion or duplication, we scored the genes inside a CNV using a total of 252 seven complementary descriptions, including the average temporal expression, number of genes 253 preferentially expressed in the brain, number of genes associated with autism spectrum disorder, 254 schizophrenia, or broader portfolio of disorders, and a functional intolerance score: loss-of-function 255 observed/expected upper bound fraction (LOEUF). LOEUF score provides insights into whether "Loss 256 of function" variants in a given gene are under negative selection pressure in a population. It reflects 257 the gene's functional impact by indicating how likely the gene will lead to disruptive changes when it 258 is inactivated by a genetic variant. We then performed an exploratory analysis using Pearson's 259 correlation between behavior scores and the quantitative descriptions of CNVs occurring in the 260 genome (Fig. 5A). For deletions, the strongest observed association was with gene brain expression 261 (Pearson's r = -0.10, $p_{FDR} = 0.13$). For duplications, the strongest observed association was with 262 temporal gene expression (Pearson's r = -0.11, $p_{FDR} = 0.0006$). This result suggested deteriorating 263 impact of altering dosage in genes expressed early during the human development. Other strongly 264 associated description was the sum of LOEUF but, due to the limited number of CNV carriers, only the 265 association with genetic temporal profile reached significance after applying False Discovery Rate 266 (FDR) correction to control for multiple comparisons. Nevertheless, the reported associations can 267 serve as valuable pointers for further research. In summary, our findings underscore the intricate 268 relationship between genetic characteristics and behavioral outcomes, highlighting the importance of 269 considering both genetic and temporal dimensions in understanding the etiology of behavioral 270 patterns and susceptibility to disorders.

271

272 We then explored whether modes of population stratification, that is, specific 273 sociodemographically defined groups, also influence the derived patterns. In other words, we 274 quantified if ethnicity, sex, age, or genetic background are linked with the shifts in behavior scores. As 275 a concrete example of this sensitivity analysis, we stratified participant-wise behavior scores for the 276 first mode by reported ethnicity (Fig. 5B). Using one-way ANOVA, we assessed whether there were 277 significant differences in scores as a function of these diverse ethnic categories. Notably, the findings 278 revealed that the scores did not exhibit a statistically significant difference among ethnicities (F-279 statistic = 0.89, p-value = 0.56). We then extended this post-hoc analysis to other modes of population 280 covariation and other metrics of population stratification. Namely, we quantified the difference in 281 scores between males and females using a two-sample t-test and as a function of the 21 recruitment 282 sites using one-way ANOVA. Moreover, we probed the association of scores with age and the ancestry 283 structure of the cohort measured using the first ten principal components of genotyping data. We 284 collected all p-values and applied FDR correction to control for multiple comparisons across the 285 totality of 52 performed tests. None of the performed tests revealed significant association (Fig. 5C). 286 This comprehensive examination provided valuable insights into the potential universality of the 287 observed scores among modes of population stratification, underscoring the importance of 288 considering the generalizability aspect in the broader context of the study's implications²⁶.

289

290 As the final step, we extended our analyses by examining longitudinal changes in brain 291 structure between controls and CNV carriers at the two-years-after imaging time point. Investigating 292 the trajectory of brain development over time can provide insights into whether individuals with CNVs 293 exhibit distinct patterns of structural change. We benefited from the availability of 3,654 brain scans 294 measured two years after the first visit (48% of participants passing genetic quality control). We 295 observed a high correlation between regional volumes acquired at these two time points (Sup. Fig. 3). 296 We used a single PLS model that subsequently evaluated brain measurement from the baseline and 297 2-year follow-up (cf. Methods). Put differently, a single PLS model provided a holistic summary of brain 298 maturation by calculating a brain score for each participant in each visit across the three modes. As 299 an example, similar to brain structure measurements, we observed a strong link between PLS scores 300 in the dominant mode between the baseline and follow-up measurements in controls (Pearson's r =301 0.90) (Fig. 5D). Furthermore, we used a bootstrap mean test (cf. Methods) to evaluate temporal 302 change in brain scores in controls and CNV carriers across the three canonical modes (Fig. 5E). We 303 found a significant decrease in the mean brain score only for the first canonical mode in controls (p-304 value_{FDR} = 0.001). Nevertheless, both CNV groups displayed similar brain maturation patterns 305 compared to controls. Given the observed similarity in brain structure developmental patterns of CNV 306 carriers and controls, further exploration of earlier stages of life may provide valuable insights into 307 distinctions in neurodevelopmental processes.

308 Discussion

309 In the current study, we carefully examined the ramifications of carrying an exonic CNV on 310 brain organization and behavior. To this end, we designed an analytic framework based on a single 311 holistic pattern-learning algorithm that can quantitatively dissect the impact of genetic mutations on 312 a multimodal measurement that cut across disciplines to untangle the complex genes-brain-behavior 313 interplay. This multivariate model uncovered three significant modes of brain-behavior covariation. 314 The first mode connected volumetric differences in more than 50% of brain regions with measures of 315 cognition and demographics. The second mode linked dorsal attention, somatomotor, and 316 frontoparietal networks with mental health measures. Finally, the third mode highlighted associations 317 between the higher-order networks and environmental factors. We then drew a detailed picture of 318 how carrying a genomic deletion or duplication influences the expression of these comprehensive 319 brain and behavior patterns. Specifically, both classes of CNVs were linked to adverse impacts on 320 family well-being, as seen in the adverse effects on cognitive functioning, mental health, and 321 socioeconomic outcomes. Our results also highlight the similar ramifications for cognition and 322 behavior associated with deletions and duplication despite their distinct effects on brain anatomy, 323 corroborating some of our earlier CNV-imaging studies on the UK Biobank^{8,27}. Finally, we 324 demonstrated that different genomic characteristics drive these shifts in behavior differentiation and 325 brain maturation. 326

327 Our results call for broadening the scope of genetic analyses in human health beyond what is 328 traditionally considered. However, the analyses of genetic influences have long been dominated by univariate frameworks^{28,29}. These standard regression approaches model one outcome variable at a 329 330 time and thus focus on individual variables independently while neglecting the complex relationships 331 and synergies that exist among genes, brain, and behavior. In other words, univariate approaches 332 struggle with high-dimensional data and are prone to the "curse of dimensionality", making it 333 challenging to capture the joint influence of multiple variables³⁰. Scientists have long been waiting for 334 multimodal datasets with deep phenotypic profiling of each participant. However, the availability of 335 such resources prompts a change in our analytic toolkit³¹. PLS addresses several limitations of mass 336 univariate approaches by embracing a doubly-multivariate strategy, providing a more nuanced and 337 integrated perspective on the relationships between thousands of measures of brain architecture and behavior in the general population^{15,19}. Prior research showed that genes are an important contributor 338 339 to the interindividual variability of thus uncovered latent patterns³². Building on the heritability of the 340 latent patterns, we showcased that their expression is further shaped by the presence of genome-341 wide protein-coding mutations, demonstrating their effects on social, familial, and environmental 342 factors.

343

344 The consequences for various aspects of human health and well-being often go unnoticed 345 because analyses of genomic deletions and duplications most commonly focus on intellectual disability and developmental delay^{33,34}. Developmental delay phenotypes, especially language and 346 347 motor disorders, are the earliest symptoms for which children are clinically referred for assessments 348 and genetic testing³⁵. Recent results showcased potential lifelong implications represented by 349 diminished academic qualifications, occupation or household income for a small set of schizophreniaassociated CNVs³⁶. As an extension, our results demonstrated that the genome-wide presence of any 350 351 coding CNV could lead to impaired real-life functioning, represented here by cognitive performance, 352 income, education, screen time, or sleep duration. These characteristics played a driving role in our

dominant mode of population covariation. A similar dominant mode characterized by cognitive measures as well as screen time was identified in the HCP population resources¹⁹. The stronger influence of deletions on the dominant mode compared to duplications is concordant with the more pronounced effect of deletions on cognitive ability observed in clinical studies³⁷. Therefore, the validity of the PLS-derived dimensions is corroborated by recapitulating key findings from previous clinical studies.

359 The power of our doubly-multivariate approach allowed us to reveal additional consequences 360 beyond just the dominant population mode, which are at risk of staying hidden in classical analyses³⁸. 361 Concretely, our second mode highlighted impoverished familial mental well-being as a prominent 362 marker of deletion carriers. Notably, the presence of phenotypes from both child and parental 363 questionnaires demonstrates how well-being is closely tied to the family system as a whole. It has 364 been estimated that over 99% of CNVs are inherited, whereas others arise de novo³⁹. Therefore, in 365 addition to influencing offspring phenotype through genetic inheritance, the parental genotype can 366 indirectly influence offspring phenotype through its expression in the parental phenotype⁴⁰. In other 367 words, carriers of the gene-dosage alterations may show certain characteristics similar to the parent's 368 condition. Where this occurs, offspring may be subject to both phenotype-associated CNV and 369 phenotype-associated environments from parents. These dynastic effects, i.e. "inheritance" of the 370 environment in addition to genotype, might explain how CNVs associated with difficult behavior in the 371 parent's generation create unfavorable environments, which might inflate behavioral problems in 372 children. In conclusion, the CNVs we studied in adolescents are likely to have been passed down from 373 either parent, which points toward influences on the overall family system. The multigenerational 374 impact where genetic and also environmental legacies contribute to the behavioral outcomes 375 highlights the complex interplay between inherited genetic variations and the environments shaped 376 by parental phenotypes.

377 Finally, our analyses also spotlighted genomic deletions as shifting the expression of the mode 378 linking the environment and higher-order brain networks. Interestingly, the higher association cortex, 379 especially the default mode network, was suggested to be more "life wired" - resulting from 380 differences in the circumstances and contexts in which people grow up and everyday life 381 experiences⁴¹. In other words, the deeper layers of the neural processing hierarchy, such as the default 382 mode network, allow for greater environmental influence and plasticity, as demonstrated by 383 prolonged maturation and slower myelination compared to sensory/motor circuits in human 384 primates^{42,43}. Our finding adds more evidence for the adaptive and dynamic nature of the recently evolved parts of the human brain, emphasizing the significant role of genetic and environmental 385 386 interplay in shaping neural development and function. Importantly, the environmental milieu, here 387 represented by measures of crimes, drug possession, or violence, is related to health through psychological, physiological, and behavioral pathways⁴⁴. Prior research documented chronic health 388 389 conditions to be more prevalent in low-income neighborhoods, including those affecting infants (low 390 birth weight), children (asthma), and adults (cardiovascular health)⁴⁵. Specifically, low socioeconomic 391 status neighborhoods⁴⁶ and neighborhoods perceived as unsafe⁴⁷ displayed elevated physiological 392 risk, which includes indicators of inflammation and neuroendocrine and cardiovascular functioning. 393 Our study thus implies that genetic underpinnings that govern neurodevelopmental trajectories may 394 potentially magnify the impact of environmental stressors on long-term health disparities.

395

The effects of protein-coding alterations are not limited to a single brain circuit, such as the default mode network. Instead, we painted a detailed picture of effects on brain patterns spanning

398 the entire cortex. In concordance, a previous study identified brain-wide patterns of regional 399 alterations that robustly differentiate controls from carriers of clinical CNVs⁸. Here, we broaden the 400 incumbent analysis scope of a few selected CNVs towards any coding CNV present in the genome. We 401 observed opposite effects of deletions and duplications, recapitulating the mirroring effects observed 402 in clinical studies²¹. As a primary example, the lingual gyrus played a dominant role in two altered brain 403 patterns. The effects of CNVs on this region have been documented for carriers of 16p11.2 CNVs⁴⁸. 404 Similarly, as prominent examples from the frontal lobe, we observed significant contributions of the 405 middle and superior frontal gyrus, which have been shown to be impacted by 1q21.1⁴⁹, resp. 15q11.2⁵⁰ alteration. Impairments of lingual and frontal gyri have been associated with anxiety-depression 406 407 severity⁵¹ or attention control deficit⁵² – phenotypes often present in CNV carriers⁵³. Though these 408 examples are region-specific, our research indicates far-reaching implications across entire networks. 409 While the identified whole-brain patterns represent a general trend in each class of genetic mutation, 410 the specific alterations pertinent to a specific CNV (e.g., 22q11.2 or 16p11.2) are further molded by 411 the attributes of genes that are affected by a given CNV. According to analyses, the final brain and 412 behavior profile is shaped by various attributes, including the number of deleted/duplicated genes, 413 their tolerance to being mutated, or the temporal expression profile of affected genes. This may be 414 part of the reason why prior research found brain patterns associated with deletions at 22q11.2 loci 415 to strongly resemble deletions at 15q11.2 loci while being different from 16p11.2 deletions⁸. In other 416 words, the here-revealed brain modes represent the bedrock of the global CNV effect upon which the 417 specific CNV effects unfold guided by the properties of encompassed genes, such as temporal and 418 spatial expression profiles.

419

420 In conclusion, we developed a multi-level pattern-learning framework to investigate the 421 effects of genome-wide protein-coding mutations on brain organization and behavior. This approach 422 offers a comprehensive view of the multifaceted impact of rare genetic variations, surpassing 423 important limitations of traditional univariate frameworks. We revealed that both genomic deletions 424 and duplications can lead to a toll on family well-being through associations with increased parental 425 and child stress, anxiety and depression, as well as neighborhood violence. This demonstrates that the 426 consequences of genetic mutations are far more extensive than just cognitive functioning, playing a 427 significant role in shaping socioeconomic and environmental aspects of life experiences. Future 428 research building on our multidisciplinary approach can better appreciate the complexity of the 429 relationship between genetic determinants and human health.

431 Methods

432 ABCD population data source

433 Brain imaging, behavioral, clinical, and genetic data in this study were obtained from the 434 Adolescent Brain Cognitive Development Study (ABCD; https://abcdstudy.org/), representing the 435 most extensive biomedical child development study of its kind. The ABCD Study acquired data from 436 11,877 children aged 9–10 years (mean age = 9.49 years) from 21 sites across the United States (48% 437 girls; 57% Caucasian, 15% African American, 20% Hispanic, 8% other)⁵⁴. We leveraged baseline 438 measurements from ABCD Annual curated release 4.0 (https://data-archive.nimh.nih.gov/abcd). Data 439 Release 4.0 contains baseline data on the entire participant cohort as well as early longitudinal data, 440 including 2-year follow-up neuroimaging data (second brain imaging timepoint). All protocols for ABCD 441 are approved by either a central or site-specific institutional review board committee⁵⁵. Caregivers 442 have provided written, informed consent and children provide verbal assent to all research 443 protocols⁵⁶. Additional information about the ABCD Study can be found in⁵⁷. This dataset is 444 administered by the National Institutes of Mental Health Data Archive and is freely available to all 445 qualified researchers upon submission of an access request. All relevant instructions to obtain the 446 data can be found at https://nda.nih.gov/abcd/request-access.

447

448 Genetic annotation and CNV calling

449 Our study is built on the identification of exonic CNVs in the ABCD study sample. The 450 genotyping protocol for the ABCD sample has been described previously⁵⁸. In addition to the QC 451 provided by ABCD, we performed several additional steps to ensure high quality of the genetic data. 452 Using PLINK v1.9⁵⁹, we removed SNP variants with a missing rate > 5% as well as SNPs with a Hardy-453 Weinberg equilibrium exact test p-value < 0.0001. We only considered arrays with call rate \geq 99 %, log 454 R ratio SD < 0.35, B allele frequency SD < 0.08, the absolute value of wave factor < 0.05 and the count 455 of all unfiltered CNV per sample \leq 10. We also removed subjects with > 5 % missing rate (n = 73) and, 456 all individuals with duplicated data (n = 419, high degree of identity-by-descent, PI HAT > 0.8) or with 457 discordant phenotypic and genetic information regarding sex (n = 0).

458 The identification of CNVs using SNP array (GRCh37/hg19) data followed previously published 459 methods^{34,60}. **CNVs** were called using the pipeline described at 460 https://github.com/labjacquemont/MIND-GENESPARALLELCNV. In short, we computed (PFB)-files 461 (Human Genome Build NCBI37/hg19) based on 500 best arrays in ABCD, and we used GC (content)-462 model files (https://kentinformatics.com and https://github.com/ucscGenomeBrowser/kent.git). 463 Only autosomal CNVs detected by both PennCNV⁶¹ and QuantiSNP⁶² were used, to minimize the 464 number of potential false positives. All identified CNVs met stringent quality control criteria: 465 confidence score \geq 30 (for at least one of the two detection algorithms), size \geq 50 kb, unambiguous 466 type (deletion or duplication), overlap with segmental duplicates, and HLA regions or centromeric 467 regions < 50 %. Finally, all carriers (n = 1) of a structural variant \geq 10Mb, a mosaic CNV or a 468 chromosome anomaly (aneuploidy or sexual chromosome anomaly) were removed. For the final set 469 of participants, we calculated the first 10 genetic principal components (PCs) using the --pca function 470 from PLINK v2.3⁶³.

471 All identified CNVs were annotated using Gencode V19 (hg19) with ENSEMBL 472 (https://grch37.ensembl.org/index.html). In this study, we only used exonic CNVs that fully 473 encompassed at least one gene. In addition to the number of encompassed genes, each CNV was 474 further annotated with seven other previously used scores. Specifically, we used an annotation 475 quantifying the tolerance to protein-loss-of-function of each gene: Loss-of-function

476 Observed/Expected Upper bound Fraction -LOEUF⁶⁴. Each CNV was then characterized by the sum of 477 LOEUF of encompassed genes. Furthermore, CNVs were described using average temporal 478 expression⁶⁵ and average peak epoch. Gene-wise temporal expression was calculated as the 479 developmental trajectory that the gene follows based on trajectory analysis (gene-specific trajectory 480 coding: 'Rising' = 1, 'Non-transitional' = 0, 'Falling' = -1). The peak epoch corresponds to an epoch of 481 highest expression, where epochs correspond to the developmental period defined previously⁶⁶. Each 482 CNV was also characterized by the number of genes, which expressions in the brain were labeled as 483 'High' or 'Elevated' according to the GTEx resource (https://www.gtexportal.org). Finally, we 484 quantified how many genes in each CNV were previously associated with autism spectrum disorder 485 (risk genes from ref.⁶⁷, schizophrenia (risk genes from ref.⁶⁸, and any disorder by either rare or common variation⁶⁵. The similarity of the seven annotations is summarized in Sup. Fig. 5. 486

487 As part of our sensitivity analyses, we compiled a list of 85 CNVs previously proposed to be 488 pathogenic^{10,33,69–71} (sum of 1/LOEUF for each gene encompassed in CNV \ge 6 or inclusion in ClinGen 489 resource⁷²). Regional coordinates are available in ref. ³⁴. CNV was defined as recurrent if it overlapped 490 \ge 40% with one of the 85 CNVs and/or included the key genes of corresponding region (see details for 491 each recurrent CNV in Supplementary Table 1).

492

493 Deep phenotyping using behavioral, cognitive, and socio-demographic data

494 We analyzed a rich battery of cognitive, socio-demographic and environmental data partially 495 reported in prior research¹⁵. Tabulated data of 1,319 phenotypes measured at baseline from 11,879 496 participants were imported and processed using Python. In line with previous research¹⁵, we used 497 robust z-scores for the preprocessing of each phenotype. The robust z-scores are derived by 498 calculating each phenotype's absolute deviation from the median absolute deviation $(MAD)^{73}$. In other 499 words, the resulting score indicates how many standard deviations each value deviates from the 500 median, with robustness to outliers. Subsequently, we removed values with a z-score > 4 (4 \times MAD). 501 We then excluded phenotypes with less than 75% retained values before excluding subjects with less 502 than 75 % retained values across the retained phenotypes. The remaining subjects (n = 11,867) were 503 considered for further analysis. Finally, all data included in the study were visually checked by the 504 same researcher (JK). The complete list of 962 phenotypes from 20 predefined categories included in 505 the analysis is available in Table S1. As the last step, for the purpose of data analysis, missing values 506 were imputed using the KNNImputer function (n neighbors = 5, weights = "uniform") from the scikit-507 learn package. All derived phenotypic measures were then adjusted for variation that can be explained 508 by age and sex.

509

510 MRI imaging-derived phenotypes

511 Our data sample included expert-curated brain-imaging phenotypes of grey matter 512 morphology (T1w MRI). The images were acquired across 21 sites in the United States with 513 harmonized imaging protocols for GE, Philips, and Siemens scanners⁷⁴. We used baseline structural 514 T1-weighted tabulated MRI data from ABCD curated release 4.0. We only included participants who 515 passed quality assurance using the recommended QC parameters (n = 11,723) described in the ABCD 516 4.0 Imaging Instruments Release Notes. ABCD preprocessing and QC steps are described in detail in 517 the methodological reference for the ABCD study⁷⁴.

Tabulated brain-imaging phenotypes were guided by the topographical brain region definitions based on the Destrieux parcellation atlas⁷⁵. This feature-generation step provides neurobiologically interpretable measures of gray matter volume in 148 regions. For each included regional volume, we calculated the MAD for each brain region and removed values with MAD > 4^{15} . Subjects with less than 90% of regional volume retained in any region were excluded from the analysis. The remaining subjects (n = 11,723) were included for further analysis. Finally, interindividual variations in the volumes that could be explained by nuisance variables outside primary interest were adjusted for by regressing out: age, sex, total brain volume, and scanning site.

526 We used the BrainStat toolbox⁷⁶ to contextualize obtained patterns with respect to large-scale 527 brain networks defined by Schaefer-Yeo definition⁷⁷. Specifically, we mapped the brain loadings sizes 528 from the 148 regions to fsaverage5 vertices, which we then averaged across the seven resting state 529 brain networks.

530 To analyze temporal changes in brain structure, we also acquired structural T1-weighted 531 tabulated MRI data during the follow-up two years after the first MRI recording. Brain imaging data 532 from this second time point underwent the same cleaning steps as the baseline brain imaging data. In 533 total, follow-up brain measurements were available for 5,663 subjects.

534

535 Multivariate pattern analysis protocol

536 Our analysis aimed to reveal dominant modes of covariation between brain and behavior in 537 the general population. Such a holistic perspective on complex brain-behavior relationships might be 538 challenging to discern through univariate analyses. We leveraged 7,657 participants with both brain 539 and behavior measurements that passed quality control. As a first data cleaning step, each brain and 540 behavior measurement was normalized (z-scored) and submitted (separately for brain and behavior 541 measurements) to principal component analysis (PCA) in order to increase the robustness of 542 subsequent steps by avoiding potential issues with rank deficiency and fitting to noise. Based on a 543 thorough examination, we extracted the first 100 PCA components for both the brain and behavior 544 measurements (Sup. Fig. 6). It is noteworthy that although we used regional volumes as measures of 545 brain structure, comparable results, especially for behavioral loadings, were obtained with regional 546 thickness and area (Sup. Fig. 7).

547 After initial cleaning steps, we used measurements from the participants not carrying any CNV 548 to reveal modes of covariation in the general population. Therefore, the first set of cleaned input 549 variables comprised regional brain volumes (5,535 × 100 matrix) and the second input variable set was 550 constructed from behavior measurement (5,535 \times 100 matrix). We elected canonical partial least 551 squares analysis (PLS) as a multivariate pattern learning approach that is ideally suited to delineate 552 relationships between two high-dimensional sets of variables by identifying latent structures that 553 maximize the covariance between them⁷⁸. In other words, PLS involves finding canonical modes that 554 maximize covariations between the weighted set (linear combination) of behavior measures and a 555 weighted set (linear combination) of brain structure measures. These PLS modes are identified by 556 solving the generalized eigenvalue problem of the brain-behavior cross-covariance matrix. The 557 subject-wise latent variables (i.e., linear combinations of original measures) are called brain and 558 behavior scores, respectively, throughout the manuscript. The ensuing set of PLS k orthogonal modes 559 of variation were uncorrelated with each other and naturally ordered from the most to least 560 important. Therefore, the first and strongest mode explained the largest fraction of covariance 561 between brain and behavior measurements. Each ensuing mode captured a fraction of brain-behavior 562 covariation not explained by one of the k-1 other modes. Notably, PLS draws similarities with 563 canonical correlation analysis (CCA). However, CCA can be prone to instability⁷⁹. Nevertheless, our 564 obtained PLS solutions strongly resemble those obtained with CCA (Sup. Fig. 8). 565

566 **Contribution of original phenotypes to latent variables**

567 In order to quantify the contribution of each phenotype (e.g., regional volume or behavioral 568 measure) to the construction of the latent structures, we computed PLS loading. These loadings are 569 obtained as Pearson's correlation between a respective PLS score and the original measurement 570 across subjects. The loadings thus indicate the strength and direction of the relationship between the 571 original phenotype and the identified canonical components. Stronger loading values signify greater 572 importance in contributing to the latent structures, offering insights into which variables drive the 573 covariation patterns between brain architecture and behavior.

As a test for significance, above-chance phenotype contribution, we embraced a bootstrapping resampling strategy for the PLS model. In the first phase, a randomly perturbed version of the dataset was created by sampling controls with the same sample size with replacement. We repeated the bootstrap resampling procedure with 1,000 iterations. In so doing, we obtained different realizations of the entire analysis workflow and ensuing PLS model estimates. Concretely, the bootstrapping algorithm resulted in 1,000 instances of the trained PLS models used to obtain 1,000 sets of associated PLS coefficients.

581 Statistically salient coefficients had a distribution of 1,000 PLS coefficients significantly 582 different from 0. Specifically, they were robustly different from zero if their two-sided confidence 583 interval according to the 2.5/97.5% bootstrap-derived distribution did not include zero coefficient 584 value, indicating the absence of an effect.

585

586 **Optimal number of PLS dimensions in the general population**

587 Each identified PLS mode was entered into statistical significance tests of robustness 588 consistent with an established combination of cross-validation and permutation testing¹⁵ (Sup. Fig. 1). 589 Initially, controls are split into 10 folds, where nine folds of participants are used as a train set, and 590 one fold is used as a test set. The controls in the training set are used to estimate the parameters of 591 all subsequent tools. In the first step, each brain or behavior measurement is z-scored column-wise 592 across all controls in the training set. PCA then separately reduces the dimension of brain and behavior 593 measurements to 100 features. In the next step, the behavior and brain measures are used as input 594 variables to estimate a single multivariate canonical PLS model, where the output of the model is a set 595 of scores (latent variables). This PLS model can also be characterized by weights (projection matrices 596 used to transform input variables). PLS weights are back-projected using the PCA model to obtain 597 brain and behavior weights in the original non-reduced ambient space.

In the next step, brain and behavior scores are computed for controls from the test set. Specifically, z-scoring followed by PCA dimensionality reduction is applied with parameters learned using the training set. The resulting preprocessed measurements are multiplied by PLS weights to obtain PLS scores for test-sample controls. Finally, the covariance between brain and behavior scores is calculated for each canonical mode. We took the average of these canonical covariances across the 10 folds. This procedure is repeated 100 times with a random fold split of controls to obtain a distribution of out-of-sample covariances for each PLS mode.

To assess the significance of the resulting PLS modes, we ran 1,000 iterations of the same 10fold cross-validation procedure described above, where the order of participants of the brain measurements was randomly permuted in each iteration. In contrast to the unpermuted dataset, we collected covariances for the training rather than the testing subjects to account for overfitting by the PLS. In other words, using covariance from the permuted train set, and not the test set, represents a

- 610 more stringent criterion. Finally, P values for each of the PLS modes were calculated as a percentage
- of cases when permuted covariance was greater than the mean cross-validated covariance.

613 Group differences in the expression of brain and behavior patterns

614 One of the main objectives of our analyses was to investigate the effect of CNV carriership on 615 the revealed modes of brain-behavior covariation. To that end, we developed a pipeline quantifying 616 the differences in brain and behavior scores for the identified significant PLS mode (Sup. Fig. 1). 617 Initially, participants without any CNV are split into a training set (80%) and a test set (20%). As 618 described above, brain and behavior scores are computed for controls from the test set using 619 parameters learned from the training set. In addition, brain and behavior scores are also calculated 620 for CNV carriers (either deletion or duplication). Specifically, z-scoring followed by PCA dimensionality 621 reduction is applied with parameters learned using the training set of controls. The resulting 622 preprocessed measurements are multiplied by PLS weights to obtain PLS scores for CNV carriers. 623 Finally, the differences in both brain and behavior scores between out-of-sample controls and CNV 624 carriers are calculated for each canonical mode. This procedure is repeated 1,000 times with a random 625 80:20 split of controls to obtain a distribution of PLS score differences. Finally, P values for each of the 626 PLS modes were calculated as a percentage of cases when the difference between mean scores of 627 CNV carriers and mean scores of out-of-sample controls was greater than zero (resp. lower for modes 628 with negative mean expression).

629

630 Temporal shift in brain pattern expressions

631 In total, 3,654 participants passed the quality control of genetic data and had brain recordings 632 measured at the baseline and 2-year follow-up. For these subjects, we compared brain scores from a 633 single PLS model that subsequently evaluated brain baseline and follow-up measurements. In the first 634 step, the parameters of a PLS model were estimated using baseline measurements of brain and 635 behavior. Using these brain loadings, we then calculated a brain score for each participant in each visit 636 across the significant PLS modes. Furthermore, we used a bootstrap mean test to evaluate if there 637 was a significant temporal change in brain scores in controls and CNV carriers. In this non-parametric 638 approach to statistical inference, 1,000 bootstrap samples are drawn with replacements from the 639 brain scores. The mean is then calculated for each of these bootstrap samples, creating a distribution 640 of sample means for both sets of brain scores. Finally, there is a significant difference in PLS scores 641 between baseline and follow-up if the two-sided confidence interval according to the 2.5/97.5% 642 distribution of 1,000 differences does not include zero.

- 643644 Effect size of CNV carriership
- 645 We used Cohen's d to quantify the effect size of the CNVs on revealed PLS modes. For a given 646 mode and separately for brain and behavior, Cohen's d is calculated as:
- 647 $d = \frac{x_1 x_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}},$

648 where $\underline{x_1}$ corresponds to the mean PLS score across CNV carriers, $\underline{x_2}$ corresponds to the mean PLS 649 score across controls. Similarly, s1 and s2 correspond to standard deviations of PLS scores of CNV 650 carriers and controls.

652 Data availability

The data supporting the findings of this study are available from the Adolescent Brain Cognitive Development (ABCD) dataset. The ABCD dataset is a publicly available resource accessible through the National Institute of Mental Health Data Archive (NDA). The specific data used in this study can be located within the ABCD dataset under X.

657

658 Code availability

The processing scripts and custom analysis software used in this work are available in a publicly accessible GitHub repository, along with examples of key visualizations in the paper: https://github.com/dblabs-mcgill-mila/CNV-covariation.

663 664	Re 1.	ferences Thompson, P. M. <i>et al.</i> Genetic influences on brain structure. <i>Nat Neurosci</i> 4 , 1253–
665		1258 (2001).
666	2.	Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B.
667		Genetic architecture: the shape of the genetic contribution to human traits and disease.
668		Nat Rev Genet 19 , 110–124 (2018).
669	3.	Claussnitzer, M. et al. A brief history of human disease genetics. Nature 577, 179–189
670		(2020).
671	4.	Manolio, T. A. et al. Finding the missing heritability of complex diseases. Nature 461,
672		747–753 (2009).
673	5.	Conrad, D. F. et al. Origins and functional impact of copy number variation in the human
674		genome. <i>Nature</i> 464 , 704–712 (2010).
675	6.	Freeman, J. L. et al. Copy number variation: New insights in genome diversity. Genome
676		<i>Res.</i> 16 , 949–961 (2006).
677	7.	Auwerx, C. et al. The individual and global impact of copy-number variants on complex
678		human traits. Am J Hum Genet S0002-9297(22)00061–1 (2022)
679		doi:10.1016/j.ajhg.2022.02.010.
680	8.	Kopal, J. et al. Rare CNVs and phenome-wide profiling highlight brain structural
681		divergence and phenotypical convergence. Nat Hum Behav 1–17 (2023)
682		doi:10.1038/s41562-023-01541-9.
683	9.	Chawner, S. J. R. A. et al. Genotype-phenotype associations in children with copy
684		number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): a
685		case-control cohort study. Lancet Psychiatry 6, 493–505 (2019).
686	10	. Marshall, C. R. et al. Contribution of copy number variants to schizophrenia from a
687		genome-wide study of 41,321 subjects. Nat Genet 49, 27–35 (2017).
688	11	. Sanders, S. J. et al. A framework for the investigation of rare genetic disorders in
689		neuropsychiatry. Nat Med 25, 1477–1487 (2019).

- 12. Fuhrmann, D., Knoll, L. J. & Blakemore, S.-J. Adolescence as a Sensitive Period of
- Brain Development. *Trends Cogn Sci* **19**, 558–566 (2015).
- 13. Tost, H., Champagne, F. A. & Meyer-Lindenberg, A. Environmental influence in the
- brain, human welfare and mental health. *Nat Neurosci* **18**, 1421–1431 (2015).
- 694 14. Paus, T., Keshavan, M. & Giedd, J. N. Why do many psychiatric disorders emerge
 695 during adolescence? *Nat Rev Neurosci* 9, 947–957 (2008).
- 15. Alnæs, D., Kaufmann, T., Marquand, A. F., Smith, S. M. & Westlye, L. T. Patterns of
- 697 sociocognitive stratification and perinatal risk in the child brain. *Proceedings of the*
- 698 National Academy of Sciences **117**, 12419–12427 (2020).
- 16. Dahl, R. E., Allen, N. B., Wilbrecht, L. & Suleiman, A. B. Importance of investing in
- adolescence from a developmental science perspective. *Nature* **554**, 441–450 (2018).
- 701 17. Millan, M. J. *et al.* Altering the course of schizophrenia: progress and perspectives. *Nat* 702 *Rev Drug Discov* 15, 485–515 (2016).
- 18. Casey, B. J. *et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging
- acquisition across 21 sites. *Developmental Cognitive Neuroscience* **32**, 43–54 (2018).
- 19. Smith, S. M. *et al.* A positive-negative mode of population covariation links brain
- connectivity, demographics and behavior. *Nature Neuroscience* **18**, 1565–1567 (2015).
- 20. Stalnaker, T. A., Cooch, N. K. & Schoenbaum, G. What the orbitofrontal cortex does not
 do. *Nat Neurosci* 18, 620–627 (2015).
- 709 21. Modenato, C. *et al.* Effects of eight neuropsychiatric copy number variants on human
 710 brain structure. *Transl Psychiatry* **11**, 1–10 (2021).
- 711 22. Hartwigsen, G. *et al.* Phonological decisions require both the left and right supramarginal
- 712 gyri. *Proc Natl Acad Sci U S A* **107**, 16494–16499 (2010).
- 23. Mechelli, A., Humphreys, G. W., Mayall, K., Olson, A. & Price, C. J. Differential effects of
- word length and visual contrast in the fusiform and lingual gyri during reading. *Proc Biol Sci* 267, 1909–1913 (2000).
- 716 24. Vossel, S., Geng, J. J. & Fink, G. R. Dorsal and Ventral Attention Systems.
- 717 *Neuroscientist* **20**, 150–159 (2014).

- 718 25. Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. The brain's default network:
- anatomy, function, and relevance to disease. *Ann N Y Acad Sci* **1124**, 1–38 (2008).
- 720 26. Kopal, J., Uddin, L. Q. & Bzdok, D. The end game: respecting major sources of
- 721 population diversity. *Nat Methods* **20**, 1122–1128 (2023).
- 722 27. Kopal, J. et al. Using rare genetic mutations to revisit structural brain asymmetry. bioRxiv
- 723 2023.04.17.537199 (2023) doi:10.1101/2023.04.17.537199.
- 724 28. Bzdok, D. Classical Statistics and Statistical Learning in Imaging Neuroscience. *Front.* 725 *Neurosci.* **0**, (2017).
- 726 29. Smith, S. M. & Nichols, T. E. Statistical Challenges in 'Big Data' Human Neuroimaging.
 727 *Neuron* 97, 263–268 (2018).
- 30. Bzdok, D., Nichols, T. E. & Smith, S. M. Towards Algorithmic Analytics for Large-scale
 Datasets. *Nat Mach Intell* 1, 296–306 (2019).
- 31. Bzdok, D. & Ioannidis, J. P. A. Exploration, Inference, and Prediction in Neuroscience
 and Biomedicine. *Trends in Neurosciences* 42, 251–262 (2019).
- 32. Nicolaisen-Sobesky, E. *et al.* A cross-cohort replicable and heritable latent dimension
- 733 linking behaviour to multi-featured brain structure. *Commun Biol* **5**, 1–11 (2022).
- 734 33. Cooper, G. M. *et al.* A Copy Number Variation Morbidity Map of Developmental Delay.
 735 *Nat Genet* 43, 838–846 (2011).
- 736 34. Huguet, G. et al. Measuring and Estimating the Effect Sizes of Copy Number Variants on
- 737 General Intelligence in Community-Based Samples. *JAMA Psychiatry* **75**, 447–457
- 738 (2018).
- 35. Kim, S. H. *et al.* Language characterization in 16p11.2 deletion and duplication
- syndromes. *Am J Med Genet B Neuropsychiatr Genet* **183**, 380–391 (2020).
- 741 36. Kendall, K. M. et al. Cognitive performance and functional outcomes of carriers of
- pathogenic copy number variants: analysis of the UK Biobank. Br J Psychiatry 214, 297–
- 743 304 (2019).

- 744 37. Mollon, J., Almasy, L., Jacquemont, S. & Glahn, D. C. The contribution of copy number
- variants to psychiatric symptoms and cognitive ability. *Mol Psychiatry* 1–14 (2023)
- 746 doi:10.1038/s41380-023-01978-4.
- 38. Wang, H.-T. *et al.* Finding the needle in a high-dimensional haystack: Canonical
- correlation analysis for neuroscientists. *Neuroimage* **216**, 116745 (2020).
- 39. McCarroll, S. A. et al. Integrated detection and population-genetic analysis of SNPs and
- 750 copy number variation. *Nat Genet* **40**, 1166–1174 (2008).
- 40. Morris, T. T., Davies, N. M., Hemani, G. & Smith, G. D. Population phenomena inflate
- genetic associations of complex social traits. *Sci Adv* **6**, eaay0328 (2020).
- 41. Benkarim, O. et al. Population heterogeneity in clinical cohorts affects the predictive
- accuracy of brain imaging. *PLOS Biology* **20**, e3001627 (2022).
- 42. Glasser, M. F. & Van Essen, D. C. Mapping Human Cortical Areas In Vivo Based on
- Myelin Content as Revealed by T1- and T2-Weighted MRI. *J Neurosci* **31**, 11597–11616
 (2011).
- 43. Gogtay, N. et al. Dynamic mapping of human cortical development during childhood
- through early adulthood. *Proc Natl Acad Sci U S A* **101**, 8174–8179 (2004).
- 44. Schulz, A. & Northridge, M. E. Social determinants of health: implications for
- r61 environmental health promotion. *Health Educ Behav* **31**, 455–471 (2004).
- 45. Pickett, K. E. & Pearl, M. Multilevel analyses of neighbourhood socioeconomic context
- and health outcomes: a critical review. *J Epidemiol Community Health* 55, 111–122
 (2001).
- 46. Robinette, J. W., Charles, S. T. & Gruenewald, T. L. Neighborhood Socioeconomic
- 766 Status and Health: A Longitudinal Analysis. *J Community Health* **42**, 865–871 (2017).
- 767 47. Robinette, J. W., Piazza, J. R. & Stawski, R. S. Neighborhood safety concerns and daily
 768 well-being: A national diary study. *Wellbeing, space and society* 2, (2021).
- 48. Martin-Brevet, S. et al. Quantifying the Effects of 16p11.2 Copy Number Variants on
- 770 Brain Structure: A Multisite Genetic-First Study. *Biological Psychiatry* 84, 253–264
- 771 (2018).

- 49. Modenato, C. et al. Lessons Learned From Neuroimaging Studies of Copy Number
- 773 Variants: A Systematic Review. *Biological Psychiatry* **90**, 596–610 (2021).
- 50. Ulfarsson, M. O. et al. 15q11.2 CNV affects cognitive, structural and functional correlates
- of dyslexia and dyscalculia. *Transl Psychiatry* **7**, e1109 (2017).
- 51. Couvy-Duchesne, B. et al. Lingual Gyrus Surface Area Is Associated with Anxiety-
- 777 Depression Severity in Young Adults: A Genetic Clustering Approach. *eNeuro* 5,
- 778 ENEURO.0153-17.2017 (2018).
- 52. Japee, S., Holiday, K., Satyshur, M. D., Mukai, I. & Ungerleider, L. G. A role of right
- middle frontal gyrus in reorienting of attention: a case study. *Front Syst Neurosci* 9, 23
- 781 (2015).
- 53. Alexander-Bloch, A. et al. Copy Number Variant Risk Scores Associated With Cognition,
- 783 Psychopathology, and Brain Structure in Youths in the Philadelphia Neurodevelopmental
- 784 Cohort. JAMA Psychiatry **79**, 699–709 (2022).
- 54. Volkow, N. D. *et al.* The conception of the ABCD study: From substance use to a broad
 NIH collaboration. *Dev Cogn Neurosci* 32, 4–7 (2018).
- 55. Auchter, A. M. et al. A description of the ABCD organizational structure and
- 788 communication framework. *Dev Cogn Neurosci* **32**, 8–15 (2018).
- 56. Paul, S. E. et al. Associations Between Prenatal Cannabis Exposure and Childhood
- 790 Outcomes: Results From the ABCD Study. *JAMA Psychiatry* **78**, 64–76 (2021).
- 57. Garavan, H. *et al.* Recruiting the ABCD sample: Design considerations and procedures.
 Dev Cogn Neurosci 32, 16–22 (2018).
- 58. Fan, C. C., Loughnan, R., Wilson, S., Hewitt, J. K., & ABCD Genetic Working Group.
- 794 Genotype Data and Derived Genetic Instruments of Adolescent Brain Cognitive
- 795 Development Study® for Better Understanding of Human Brain Development. *Behav*
- 796 *Genet* **53**, 159–168 (2023).
- 59. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer
- 798 datasets. *Gigascience* **4**, 7 (2015).

- 60. Huguet, G. et al. Genome-wide analysis of gene dosage in 24,092 individuals estimates
- that 10,000 genes modulate cognitive ability. *Mol Psychiatry* **26**, 2663–2676 (2021).
- 801 61. Wang, K. et al. PennCNV: an integrated hidden Markov model designed for high-
- resolution copy number variation detection in whole-genome SNP genotyping data.
- 803 *Genome Res* **17**, 1665–1674 (2007).
- 62. Colella, S. et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and
- accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res*35, 2013–2025 (2007).
- 807 63. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population808 Based Linkage Analyses. *Am J Hum Genet* **81**, 559–575 (2007).
- 809 64. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in
- 810 141,456 humans. *Nature* **581**, 434–443 (2020).
- 811 65. Werling, D. M. et al. Whole-Genome and RNA Sequencing Reveal Variation and
- 812 Transcriptomic Coordination in the Developing Human Prefrontal Cortex. *Cell Reports*813 **31**, 107489 (2020).
- 814 66. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* 478, 483–
 815 489 (2011).
- 816 67. Satterstrom, F. K. et al. Large-Scale Exome Sequencing Study Implicates Both
- 817 Developmental and Functional Changes in the Neurobiology of Autism. *Cell* 180, 568818 584.e23 (2020).
- 819 68. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological
- insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
- 69. Coe, B. P. *et al.* Refining analyses of copy number variation identifies specific genes
 associated with developmental delay. *Nat Genet* 46, 1063–1071 (2014).
- 70. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in
 controls. *Nature* 505, 361–366 (2014).
- 825 71. Moreno-De-Luca, D. et al. Using large clinical data sets to infer pathogenicity for rare
- 826 copy number variants in autism cohorts. *Mol Psychiatry* **18**, 1090–1095 (2013).

- 827 72. Welcome to ClinGen. https://clinicalgenome.org/.
- 828 73. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: Do not use
- standard deviation around the mean, use absolute deviation around the median. *Journal*
- of Experimental Social Psychology **49**, 764–766 (2013).
- 831 74. Hagler, D. J. et al. Image processing and analysis methods for the Adolescent Brain
- 832 Cognitive Development Study. *Neuroimage* **202**, 116091 (2019).
- 833 75. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical
- gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).
- 835 76. Larivière, S. et al. BrainStat: A toolbox for brain-wide statistics and multimodal feature
- associations. *NeuroImage* **266**, 119807 (2023).
- 837 77. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic
- 838 Functional Connectivity MRI. Cereb Cortex 28, 3095–3114 (2018).
- 78. Wegelin, J. A Survey of Partial Least Squares (PLS) Methods, with Emphasis on the
 Two-Block Case. *Technical report* (2000).
- 79. Mihalik, A. et al. Canonical Correlation Analysis and Partial Least Squares for Identifying
- 842 Brain–Behavior Associations: A Tutorial and a Comparative Study. *Biological Psychiatry:*
- 843 Cognitive Neuroscience and Neuroimaging 7, 1055–1067 (2022).

845 Acknowledgements

846 DB was supported by the Brain Canada Foundation, through the Canada Brain Research Fund, 847 with the financial support of Health Canada, National Institutes of Health (NIH R01 AG068563A, NIH 848 R01 DA053301-01A1, NIH R01 MH129858-01A1), the Canadian Institute of Health Research (CIHR 849 438531, CIHR 470425), the Healthy Brains Healthy Lives initiative (Canada First Research Excellence 850 fund), the IVADO R³AI initiative (Canada First Research Excellence fund), and by the CIFAR Artificial 851 Intelligence Chairs program (Canada Institute for Advanced Research). This research was also supported 852 Calcul (http://www.calculquebec.ca) by Quebec and Compute Canada 853 (http://www.computecanada.ca), the Brain Canada Multi-Investigator initiative, the Canadian Institutes of Health Research, CIHR_400528, The Institute of Data Valorization (IVADO) through the 854 855 Canada First Research Excellence Fund, Healthy Brains for Healthy Lives through the Canada First 856 Research Excellence Fund. SJ is a recipient of a Canada Research Chair in neurodevelopmental 857 disorders and a chair from the Jeanne et Jean Louis Levesque Foundation. KK was supported by The 858 Institute of Data Valorization (IVADO) Postdoctoral Fellowship program through the Canada First 859 Research Excellence Fund. BTTY is supposed by the NUS Yong Loo Lin School of Medicine 860 (NUHSRO/2020/124/TMR/LOA), the Singapore National Medical Research Council (NMRC) LCG 861 (OFLCG19May-0035), NMRC CTG-IIT (CTGIIT23jan-0001), NMRC STaR (STaR20nov-0003), Singapore 862 Ministry of Health (MOH) Centre Grant (CG21APR1009), the Temasek Foundation (TF2223-IMH-01), 863 and the United States National Institutes of Health (R01MH120080 & R01MH133334). The funders 864 had no role in study design, data collection and analysis, the decision to publish or the preparation of 865 the manuscript.

- 866
- 867 868 Author Cont

8 Author Contributions Statement

DB and JK designed the study, analyzed brain and behavioral data, and drafted the manuscript.
SJ, GH, ZS, and MJL called CNVs. DB and SJ contributed to the interpretation of the results and in the
editing of the manuscript. All authors provided feedback on the manuscript. DB led data analytics.

873

874 Competing Interests Statement

DB is a shareholder and advisory board member at MindState Design Labs, USA. OAA is a
consultant to Cortechs.ai. PT obtained grant support from Biogen, Inc., for research unrelated to this
manuscript.

879 Figure captions

880 Figure 1

881 Genome-wide CNVs impact brain-behavior relationships across distinct modes of population 882 covariation

883 A) Genome-wide CNV identification in the ABCD population cohort. We investigated 7,657 children 884 from the ABCD database. In total, 5,535 children do not carry any protein-coding CNV, 514 carry a 885 deletion and 1,472 carry a duplication fully encompassing one or more genes. 136 participants carried 886 both deletion and duplication. The ratio of males and females is similar in every group (left plot, inner 887 circle). The age of participants is similar across different CNV groups (right plot). B) CNV characteristics. 888 The most common kind of CNV is a deletion/duplication encompassing a few genes. On average, 889 duplications contain more genes than deletions in the ABCD cohort. C) Partial least squares model 890 links the brain with behavior in one holistic model. We estimate a multivariate relationship structure 891 among 148 brain atlas volume measures and ~1000 behavior measures spanning 20 categories based 892 on measurements from children without any CNV. The canonical scores represent the latent variable 893 expressions calculated from linear combinations of the original brain and linear combinations of 894 behavior measurements that maximize the covariance between the two sets of variables. k is the 895 number of phenotypes per category. D) CNV status impacts individual expression strengths of brain 896 and behavior patterns. We compared the average brain and behavior scores for CNV carriers with 897 control participants not used to derive model parameters (i.e., controls not seen by the model during 898 training). Stars denote significant differences based on cross-validation testing (cf. Methods). These 899 results reveal that carrying a CNV significantly impacts canonical scores across different modes of 900 brain-behavior covariation, emphasizing the utility of a multivariate holistic framework that cuts 901 across single disciplines.

903 Figure 2

902

904 The leading population mode tracks decreased cognitive functioning in CNV carriers

905 A) CNVs significantly impact revealed dominant behavior pattern. Cohen's d values of canonical scores 906 calculated between controls and separately carriers of deletions and duplications are plotted for the 907 first canonical mode. Both deletion and duplication carriers show a significant shift in behavior scores, 908 with a stronger effect for deletions. Only duplications display significantly affected brain scores. 909 Significant differences are represented by solid dots. B) Brain region correlates reveal a whole-brain 910 pattern. Brain loadings are calculated as the correlation between brain scores and 148 regional brain 911 volumes. The bar plot depicts 20 regions with the strongest brain loadings, including a 95% confidence 912 interval based on the bootstrap resampling. G denotes gyrus, and S denotes sulcus. The radial bar 913 chart shows the percentage of significant brain loadings in each of the seven large-scale networks 914 based on the bootstrap significance test. C) Behavior correlates highlight real-life functioning. 915 Behavior loadings are calculated as the correlation between behavior scores and ~1000 behavior 916 measures. The strongest behavior loadings include a variety of cognitive scores (i.e., summaries of 917 tasks focused on language and vocabulary comprehension, working memory, abstract reasoning, or 918 problem-solving), parental education, or family income. On average, the demographic, 919 socioeconomic, and cognitive categories display the strongest coefficients. In summary, the first 920 canonical mode highlights the connection between most of the brain networks and assessments of 921 cognition and demographics.

923 Figure 3

924 The second population mode spotlights shift in brain scores associated with mental wellbeing

925 A) Canonical scores untangle gene dosage-induced differences and similarities. Deletions and 926 duplications lead to mirrored significant brain scores. On the other hand, only deletions display 927 significantly affected behavior scores. Significant differences are represented by solid points. B) Three 928 large-scale brain networks dominate brain loadings. The lingual gyrus (G=gyrus, S=sulcus) displays the 929 strongest loading of all regions. Summarized by large-scale networks, the highest percentage of 930 significant brain loadings is present for the dorsal attention, somatomotor, and frontoparietal 931 networks. C) Behavior loadings stress the importance of mental well-being. The strongest behavior 932 loadings include parental and child-reported levels of problems, stress, anxiety, and depression. 933 Therefore, this mode was dominated by mental well-being phenotypes mainly from parental behavior, 934 child questionnaires, and sleep categories based on the average absolute loadings. Collectively, the 935 second canonical mode proposes decreased mental well-being as a significant marker of deletion 936 carriers.

- 937
- 938 Figure 4

939 The third population mode links higher order networks to environment measures

940 A) Deletions, unlike duplications, shape behavior scores. In the third canonical mode, the only 941 significant effect is for behavior scores in deletion carriers. Significant differences are represented by 942 solid points. B) Higher-order networks play a prominent role in the third canonical mode. The anterior 943 insula, followed by the middle temporal gyrus (G=gyrus, S=sulcus), are regions most strongly 944 associated with brain scores. Furthermore, the default mode network, along with the frontoparietal 945 network, displays the highest percentage of significant regions based on the bootstrap significant test. 946 C) Environmental variables characterize behavior loadings. Measures of neighborhood violence, safety 947 and crime are among the strongest loadings. These environment-associated phenotypes come 948 primarily from the socioeconomic category. The third significant mode illustrates how deletions shift 949 the expression of the mode linking the environment and higher-order networks. 950

951 Figure 5

952 Population modes are driven by temporal gene characteristics rather than sociodemographic 953 factors

954 A) Behavior scores driven by temporal and spatial expression. Genes encompassed in each CNV were 955 characterized based on seven different metrics. Summaries across all the deleted/duplicated genes 956 for each CNV carrier were used to provide subject-specific annotations. Across the three different 957 modes, behavior scores are most strongly associated with temporal expression, brain expression, and 958 LOEUF scores. The three modes are labeled based on their dominant phenotype. * denotes significant 959 association after FDR correction. B) Behavior scores not explained by ethnicity. As an example, 960 behavior scores of the first canonical model are plotted for all subjects separated by participant 961 ethnicity. The rain cloud plot combines a scatter plot, a box plot (whiskers equal to 1.5 times the 962 interquartile range), and a violin plot. There is no significant difference in mean score expression 963 across the 16 ethnicities. C) Canonical modes not reflective of sociodemographic factors. In addition 964 to ethnicity, behavior scores across the three analyzed modes do not significantly differ based on sex, 965 age, site, and the first ten genetic components capturing major patterns of genetic variation. D) 966 Canonical modes capture brain maturation. A single PLS model evaluated brain measurement from 967 the baseline and 2-year follow-up measurements. The star denotes significant change based on a 968 bootstrap mean test. Both measurements are highly correlated. E) Similar cortical aging across all 3 969 groups. As an extension, we examine the difference between baseline and 2-year follow-up 970 measurements for all CNV groups and all three canonical modes. The arrow direction symbolizes the 971 direction of change. Significant changes are represented by solid markers and a star. Both CNV groups 972 display similar patterns of brain aging compared to controls, with particular change significant for the

973 first canonical mode.









