

# Disentangling mechanisms behind the pleiotropic effects of proximal 16p11.2 BP4-5 CNVs

Chiara Auwerx<sup>1,2,3,4,\*</sup>, Samuel Moix<sup>2,3,4</sup>, Zoltán Kutalik<sup>2,3,4,\*</sup> & Alexandre Reymond<sup>1,\*</sup>

<sup>1</sup> Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland

<sup>2</sup> Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

<sup>3</sup> Swiss Institute of Bioinformatics, Lausanne, Switzerland

<sup>4</sup> University Center for Primary Care and Public Health, Lausanne, Switzerland

\* Correspondence: [chiara.auwerx@unil.ch](mailto:chiara.auwerx@unil.ch); [zoltan.kutalik@unil.ch](mailto:zoltan.kutalik@unil.ch); [alexandre.reymond@unil.ch](mailto:alexandre.reymond@unil.ch)

## Abstract

Whereas 16p11.2 BP4-5 copy-number variants (CNVs) represent one of the most pleiotropic etiologies of genomic syndromes in both clinical and population cohorts, the mechanisms leading to such pleiotropy remain understudied. Identifying 73 deletion and 89 duplication carriers among unrelated white British UK Biobank participants, we performed a phenome-wide association study between the region's copy number and 117 complex traits and diseases, mimicking four dosage models. Forty-six phenotypes (39%) were affected by 16p11.2 BP4-5 CNVs, with the deletion-only, mirror, U-shape, and duplication-only models being the best fit for thirty, ten, four, and two phenotypes, respectively, aligning with the stronger deleteriousness of the deletion. Upon individually adjusting CNV effects for either body mass index (BMI), height, cognitive function, or socio-economic status as potential mediators, we found that sixteen testable deletion-driven associations (61%) – primarily with cardiovascular and metabolic traits – were BMI-dependent, with other mediators playing a more subtle role. Bidirectional Mendelian randomization supported that 13 out of these 16 associations (81%) were secondary consequences of the CNV's impact on BMI. For the 22 traits that remained significantly associated upon individual adjustment for mediators, matched-control analyses found that eleven phenotypes, including musculoskeletal traits, liver enzymes, fluid intelligence, platelet count, pulmonary capacity, pneumonia, and acute kidney injury, remained associated under strict Bonferroni correction, with eight additional nominally significant associations. These results paint a complex picture of 16p11.2 BP4-5's pleiotropic pattern that involves direct effects on multiple physiological systems and indirect co-morbidities consequential to the CNV's impact on BMI and cognition, acting through trait-specific dosage mechanisms.

34 **Keywords**

35 proximal 16p11.2 BP4-5 CNV; structural variants; genomic disorder; PheWAS; multi-system  
36 disorder; pleiotropy; mediation; obesity.

37

38 **Abbreviations**

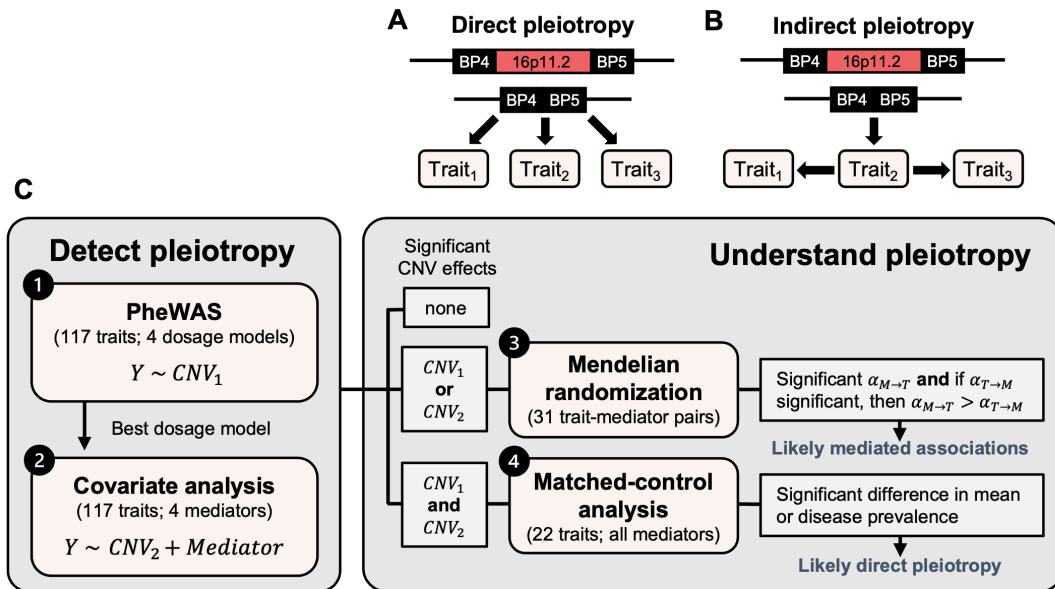
39 **ALP** = alkaline phosphatase; **ALT** = alanine aminotransferase; **AKI** = acute kidney injury; **AST** =  
40 aspartate aminotransferase; **BMI** = body mass index; **BP** = breakpoint; **CI** = confidence interval;  
41 **CKD** = chronic kidney disease; **CNV** = copy-number variants; **COPD** = chronic obstructive  
42 pulmonary disorder; **CRP** = C-reactive protein; **EA** = educational attainment; **FVC** = forced vital  
43 capacity; **GGT** = gamma-glutamyltransferase; **GWAS** = genome-wide association studies; **HbA1c**  
44 = glycated hemoglobin; **HDL** = high-density lipoprotein; **HTN** = essential hypertension; **ICD-10** =  
45 International Classification of Diseases, 10<sup>th</sup> Revision; **IGF-1** = insulin-like growth factor 1; **kb** =  
46 kilobase pair; **MR** = Mendelian randomization; **OA** = arthritis; **PheWAS** = phenome-wide  
47 association study; **SE** = standard error; **SES** = socioeconomic status; **SHBG** = sex hormone  
48 binding globulin; **SNP** = single nucleotide polymorphism; **T1D** = type 1 diabetes; **TDI** = Townsend  
49 deprivation index; **UKBB** = UK Biobank; **WHR** = waist-to-hip ratio; **WHRadjBMI** = WHR adjusted  
50 for BMI.

## 51 Introduction

52 Genomic disorders are caused by recurrent genomic rearrangements that lead to the gain  
53 (duplication) or loss (deletion) of large, multi-kilobase pair (kb) DNA fragments. The proximal  
54 16p11.2 rearrangement spans a region of ~600 kb between recurrent breakpoints (BP) 4 and 5  
55 and includes 27 unique protein-coding genes. Copy-number variants (CNVs) of the region  
56 represent one of the most common genomic disorders, with population prevalence estimates of  
57 1 in 3,000 and 1 in 2,800 for the deletion (MIM: 611913) and duplication (MIM: 614671),  
58 respectively (*in litt.*). Prevalence in clinical cohorts is about eight-fold higher, with a particularly  
59 strong enrichment in individuals ascertained for intellectual disability and developmental delay<sup>1-3</sup>  
60 or autism spectrum disorder<sup>4-7</sup>, the first phenotypes associated with the CNV. Other hallmark  
61 features include a negative dosage effect on body mass index (BMI)<sup>8-10</sup> and head  
62 circumference<sup>11,12</sup>, a predisposition for seizure disorders<sup>2,3,11,13</sup>, and a duplication-specific  
63 increased susceptibility to schizophrenia and other psychiatric conditions<sup>12,14-19</sup>. The recent  
64 establishment of large biobanks coupling genetic information to phenotypic data such as physical  
65 measurements, blood biomarkers, and electronic health records, has allowed to study the  
66 phenotypic expression of 16p11.2 BP4-5 rearrangements in individuals that are typically older and  
67 less severely affected than those recruited in pediatric clinical cohorts<sup>20-31</sup>. Results of these  
68 studies often converge onto similar pathophysiological processes than those highlighted by  
69 clinical studies but also report associations with biomarkers and common diseases that are  
70 typically overlooked or not assessed in clinical cohorts.

71  
72 If the pleiotropic nature of 16p11.2 BP4-5 rearrangements is now well-established, the  
73 mechanisms through which CNVs in the region affect such diversity of traits remain poorly studied.  
74 Under a model of direct (or horizontal) pleiotropy, the CNV causally impacts associated  
75 phenotypes through independent mechanisms (Figure 1A). Conversely, indirect (or vertical)  
76 pleiotropy implies that the CNV causally impacts a mediatory trait, which in turn causally impacts  
77 other traits that will appear as linked with the CNV in association studies (Figure 1B). These  
78 models are not mutually exclusive, and a fraction of the associations might result from direct  
79 effects while others might be secondary consequences. This question is particularly relevant  
80 given the BMI-modulating role of the 16p11.2 BP4-5 CNV<sup>8-10,20,21,24,25</sup> – which itself represents a  
81 strong risk factor for other diseases – and could therefore inform epidemiology of associated  
82 comorbidities and clinical practice. To address this knowledge gap, we re-analyzed two recent UK  
83 Biobank (UKBB) studies that assessed the impact of 16p11.2 BP4-5 rearrangements on 117

84 complex traits and common diseases<sup>21,22</sup> with the aims to i) determine the most likely dosage  
 85 mechanism for different traits and ii) estimate the fraction and nature of associations that are  
 86 mediated by primary changes in anthropometric measurements, cognitive ability, and socio-  
 87 economic status (Figure 1C).



**Figure 1. Study workflow.**

(A-B) Schematic representation of pleiotropy mechanisms. For illustration, the 16p11.2 BP4-5 deletion is depicted but the same concept applies to the duplication. (A) Direct (horizontal) pleiotropy: The CNV causally affects multiple traits – here Trait<sub>1</sub>, Trait<sub>2</sub>, and Trait<sub>3</sub> – through independent mechanisms. (B) Indirect (vertical) pleiotropy: The CNV causally impacts Trait<sub>2</sub>, which in turn causally affects Trait<sub>1</sub> and Trait<sub>3</sub>. The impact of the deletion on Trait<sub>1</sub> and Trait<sub>3</sub> is thus indirect and mediated by a shared mechanism, i.e., Trait<sub>2</sub>. (C) Overview of the study. The first two analyses aim at detecting and characterizing the pleiotropy of 16p11.2 BP4-5 CNVs through four distinct dosage models that estimate the effect of the CNV on the trait (Y) either (1) without or (2) with adjustment for one of four covariates that could potentially mediate the CNV-phenotype association. The second part of the study aims at understanding the mechanisms through which pleiotropy arises. (3) Bidirectional Mendelian randomization was used to investigate the causal relationship between trait-mediator pairs for which the significance of the CNV effect on the trait was affected by adjustment for the mediator. Support for mediation was claimed when the forward MR effect of the mediator on the trait ( $\alpha_{M \rightarrow T}$ ) is significant and larger than the reverse effect of the trait on the mediator ( $\alpha_{T \rightarrow M}$ ), providing the latter is significant. (4) For traits that showed a significant association with the CNV regardless of covariate adjustment, we performed a matched-control analysis that allowed us to adjust for all possible mediators simultaneously and detect likely genuine, direct pleiotropic associations. PheWAS = phenome-wide association study.

## 88 Materials and methods

### 89 1. Study material

#### 90 Software versions

91 Statistical analyses and graphs were generated with R v4.3.1. The Mendelian randomization  
 92 analysis pipeline uses the TwoSampleMR R package v0.5.7<sup>32</sup> and PLINK v1.9<sup>33</sup> and was  
 93 implemented in R v4.2.1.

94



95 *Cohort description & sample selection*

96 Analyses were carried out in the UKBB, a volunteer-based UK population cohort of about half a  
97 million individuals (54% females) aged 40-69 years at recruitment, who signed a broad informed  
98 consent form<sup>34</sup>. Available data include microarray genotype data acquired in GRCh37/hg19 from  
99 two similar arrays, as well as rich phenotypic data, including anthropometric measurements, vital  
100 signs, blood biomarker levels, life history and lifestyle questionnaire data, hospital-based  
101 International Classification of Diseases, 10<sup>th</sup> Revision (ICD-10) codes (up to September 2021),  
102 and self-reported conditions. Analyses conducted in this study focus on 331,522 unrelated  
103 individuals from the “white British” UKBB subset (54% females) that were filtered to exclude  
104 samples with abnormal CNV profiles and/or a report of blood malignancy. Filtering criteria to  
105 obtain this set are described elsewhere<sup>22</sup>.

106  
107 *CNV carrier identification*

108 CNV calls from a previous study were used<sup>21</sup>. Briefly, CNV calling was done based on the UKBB  
109 microarray data using standard PennCNV v1.0.5 settings<sup>35</sup>. Each call was attributed a quality  
110 score ranging from -1 (likely deletion) to 1 (likely duplication) reflecting the probability for the CNV  
111 to be a consensus call across three algorithms and thus a true positive<sup>36</sup>. 16p11.2 BP4-5 deletion  
112 and duplication carriers were identified as carrying a high-confidence CNV call (quality score < -  
113 0.5 for deletions; quality score > 0.5 for duplications) on chromosome 16 with start and end site  
114 within 29.4-29.8 Mb and 30.05-30.4 Mb, respectively. Individuals with a low-quality 16p11.2 BP4-  
115 5 CNV were excluded from copy-neutral controls. CNV genotype vectors were then encoded to  
116 allow the fitting of regression models according to four dosage mechanisms ([Table 1](#)).

**Table 1. Encoding of CNV carrier status for different dosage models.**

Numerical encoding of CNV genotypes for high confidence deletion carriers, copy-neutral individuals, and high confidence duplication carriers according to four dosage mechanisms of action. Individuals with a low-quality score CNV call are set as missing.

	<b>deletion (copy-number = 1)</b>	<b>copy-neutral (copy-number = 2)</b>	<b>duplication (copy-number = 3)</b>
<b>mirror</b>	-1	0	1
<b>U-shape</b>	1	0	1
<b>duplication-only</b>	NA	0	1
<b>deletion-only</b>	1	0	NA

117 *Phenotype selection*

118 We analyzed the same 117 phenotypes as defined in previous studies<sup>21,22</sup>. This includes 57  
119 quantitative traits<sup>21</sup> that were inverse normal transformed<sup>21</sup> before being corrected for sex (except  
120 for sex-specific traits), age (UKBB field identifier #21003), age<sup>2</sup>, genotyping batch, and principal

121 components 1-40. We further include 60 common diseases based on ICD-10 clinical diagnoses  
122 using a case-control definition procedure that excludes from controls individuals with a condition  
123 related to the one under investigation<sup>22</sup>.

124

#### 125 *Mediator selection*

126 We tested the role of four factors that could potentially mediate associations between 16p11.2  
127 BP4-5 CNVs and the assessed phenotypes:

- 128 • Body mass index (BMI): average over available instances of BMI (#21001).
- 129 • Educational attainment (EA): age at which full-time education was completed (#845).  
130 Values matching “prefer not to answer”, “never went to school”, and “do not know” were  
131 set as missing, and average over available instances was calculated. Individuals for which  
132 average age at which full-time education was completed was below 14 years or over 19  
133 years were set to 14 years and 19 years, respectively. Individuals reporting a “college or  
134 university degree” in their qualifications (#6138) were set to 19 years.
- 135 • Townsend deprivation index at recruitment (TDI; #22189).
- 136 • Height: average over available instances of standing height (#50).

137

#### 138 *GWAS summary statistics*

139 Mendelian randomization (MR) studies rely on publicly available genome-wide association  
140 studies (GWASs) summary statistics for both sexes and individuals of European ancestry. For  
141 mediators, summary statistics from Pan-UK Biobank with phenotype manifest updated on  
142 01/03/2023 (<https://pan.ukbb.broadinstitute.org/>)<sup>37</sup> were used for BMI, TDI, and height. For EA,  
143 summary statistics from a large meta-analysis were used (excluding 23andMe data;  
144 <http://www.thessgac.org/data>)<sup>38</sup>. For other phenotypes, summary statistics from the Neale group  
145 released on 07/2018 were used (<http://www.nealelab.is/uk-biobank>). These summary statistics  
146 were favored over those of large disease-specific consortia as summary statistics for binary traits  
147 were calculated through linear regression, allowing comparison of forward and reverse effects.  
148 For diseases, we used the closest possible match to our phenotype definition, i.e., phenotype  
149 code: E10 for “T1D” (type 1 diabetes); G47 for “sleep” (sleep apnea); I10 for “HTN\_essential”  
150 (essential hypertension); I35 for “valves” (cardiac valve disorders); I44 for “conduction” (cardiac  
151 conduction disorders); J45 for “asthma”; M19 for “OA” (arthrosis); N18 for “CKD” (chronic kidney  
152 disease); 20002\_1473 for “lipid” (lipidemias & lipoprotein disorders). Summary statistics for  
153 autosomal chromosomes were harmonized to the UK10K reference panel<sup>39</sup>. After excluding

154 palindromic single-nucleotide polymorphisms (SNPs) and adjusting strand-flipped SNPs, effect  
155 sizes were standardized to represent the square root of the explained variance.

156

## 157 **2. 16p11.2 BP4-5 association studies**

### 158 *Phenome-wide association study*

159 For the phenome-wide association study (PheWAS), regression analysis was performed to  
160 estimate the effect of the CNV genotype – encoded according to either of the four models in [Table](#)  
161 [1](#) – and the 117 selected phenotypes. For quantitative traits, linear regressions (`lm()` in R) were  
162 used and 95% confidence intervals (CI) were calculated as  $\beta \pm 1.96 \times \text{standard error (SE)}$ . For  
163 binary traits, Firth's bias-reduced penalized-likelihood logistic regression was used  
164 (`logistf(plconf = 2, maxit = 100, maxstep = 10)` from the `logistf` package  
165 v1.26.0 in R) to account for the fact that both CNV carriers and disease cases are rare. The same  
166 function also produces estimates for the 95% CIs. As disease diagnoses were defined as binary  
167 variables and could not be adjusted beforehand, sex (except for sex-specific traits), age,  
168 genotyping array, and principal components 1-40 were included as covariates. For each trait, the  
169 dosage model yielding the lowest p-value for the CNV effect was retained and effects were  
170 defined as strictly significant under Bonferroni correction criteria ( $p \leq 0.05/117 = 4.3 \times 10^{-4}$ ).

171

### 172 *Covariate analysis*

173 For all phenotype-mediator pairs (*Phenotype* & *Mediator selection*), including those involving  
174 phenotypes that did not significantly associate with the CNV status in our original PheWAS, we  
175 estimated the Pearson correlation (`cor(use = "pairwise.complete.obs")` in R), as well  
176 as the effect of the mediator on the phenotype in a linear/Firth regression model without  
177 covariates, as previously described (*Phenome-wide association study*). For pairs with Pearson  
178 correlation  $< 0.5$  and effect of the mediator on the trait  $p \leq 0.05/117 = 4.3 \times 10^{-4}$ , we estimated the  
179 effect of the CNV carrier status encoded according to the best PheWAS model. Regressions were  
180 implemented as previously described (*Phenome-wide association study*), adding the mediator as  
181 an additional covariate. Adjusted effects were defined as strictly significant when meeting  
182 Bonferroni correction criteria ( $p \leq 0.05/117 = 4.3 \times 10^{-4}$ ). We additionally compared effect  
183 estimates with ( $\beta_{adjusted}$ ) and without ( $\beta$ ) mediator adjustment based on a t-statistic

$$t = \frac{\beta - \beta_{adjusted}}{\sqrt{SE^2 + SE_{adjusted}^2}}$$

184 where SEs represent the standard errors of the effects. Two-sided p-values were calculated  
185 ( $2 * pnorm(-abs(t), mean = 0, sd = 1)$  in R). The difference in correlation between BMI-  
186 dependent and BMI-independent traits with BMI was assessed with a two-sided t-test.

187

### 188 *Mendelian randomization*

189 GWAS summary statistics were used to conduct bidirectional MR according to a previously  
190 published pipeline<sup>40,41</sup> for 31 mediator-trait pairs for which the CNV-trait association either gained  
191 or lost significance upon adjusting for that mediator. Concretely, the forward effect of the mediator  
192 (exposure) on the trait (outcome) and the reverse effect of the trait (exposure) on the mediator  
193 (outcome) were estimated. Harmonized SNPs significantly ( $p < 5 \times 10^{-8}$ ) associating with the  
194 exposure were clumped with PLINK v1.9 ( $p1 = 0.0001$ ,  $p2 = 0.01$ ,  $kb = 250$ , and  $r2$   
195  $= 0.01$ ) and retained as instrumental variables. Instrumental variables mapping to the extended  
196 HLA region (chr6:25,000,000–37,000,000; GRCh37/hg19) were excluded, as well as those with  
197 a difference in allele frequency ( $\geq 0.05$ ) between the outcome and exposure summary statistic.  
198 Steiger filtering was applied ( $Z \leq -1.96$ ) to ensure that the effect of the selected variants on the  
199 exposure was stronger than their effect on the outcome. Bidirectional inverse variance weighted  
200 MR analyses were carried out with the `TwoSampleMR` R package when at least two instrumental  
201 variables were available. MR effects were called significant under Bonferroni correction, when  $p$   
202  $\leq 0.05/62 = 8.1 \times 10^{-4}$ , to account for the 31 bidirectional tests performed.

203

### 204 *Matched-control analysis*

205 For each CNV carrier, we identified all copy-neutral unrelated individuals from the “white British”  
206 subset of UKBB participants that were matching based on sex (identical), age ( $\pm 2.5$  years), BMI  
207 ( $\pm 2.5$  years), TDI ( $\pm 2$ ), average household income before tax (#738) averaged over instances  
208 (identical category), and EA ( $\pm 1$  year). Fifty-eight deletion and sixty-one duplication carriers had  
209 no missing data and qualified for the matching procedure. The number of identified matching  
210 controls per carrier ranged from 1 to 918 and 12 to 1,590 for deletion and duplication carriers,  
211 respectively, with 49 deletion and 60 duplication carriers having at least 25 matching controls.  
212 When more than 25 matched controls were available, the ones used for the analysis were  
213 selected randomly (`sample_n()` in R), without replacement. For quantitative traits, we compared  
214 mean phenotypic values between deletion and duplication carriers and the respective control  
215 groups through a two-sided t-test. For binary traits, disease prevalence was compared between  
216 the same groups based on a two-sided Fisher test. Prevalence standard error was calculated as  
217

$$SE = \sqrt{\frac{q * (1 - q)}{n}}$$

218 where  $q$  is the disease prevalence, and  $n$  the sample size. Sample sizes vary between  
 219 phenotypes due to missing data. We define significant associations based on a Bonferroni  
 220 correction that accounts for the 22 traits of interest in this analysis ( $p \leq 0.05/22 = 2.3 \times 10^{-3}$ ), i.e.,  
 221 phenotypes that remained associated with the CNV under strict Bonferroni correction when  
 222 adjusting for BMI, height, EA, or TDI individually. We report all nominally significant ( $p < 0.05$ )  
 223 associations on figures.

224

225 In a related analysis aiming at assessing the consequences of losing samples for the matched-  
 226 control analysis, we used the same statistical framework to compare mean phenotypic value and  
 227 disease prevalence between deletion and duplication carriers that were included in the matched-  
 228 control analysis versus those that were not due to missing data or lack of sufficient controls.

229

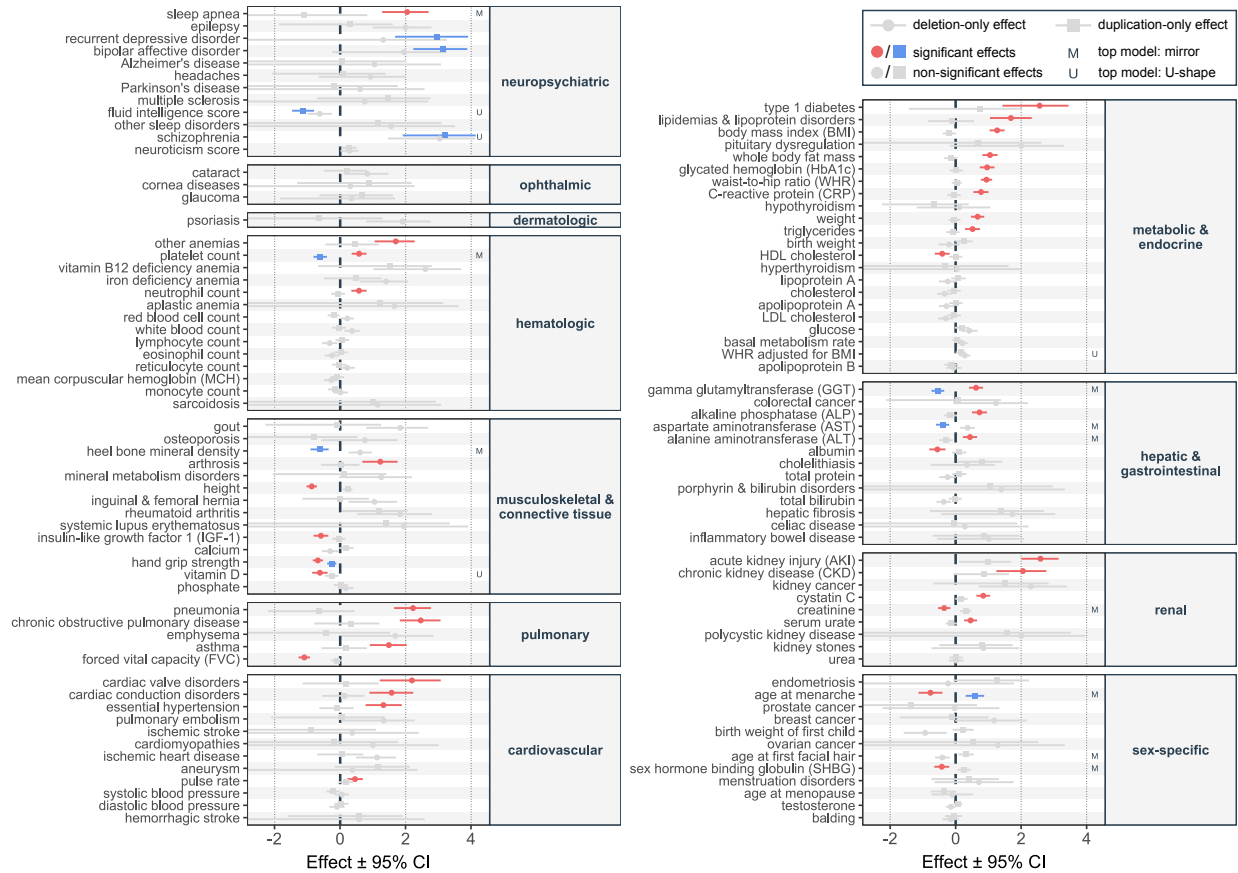
## 230 Results

231 Using previously published high confidence CNV calls for 331,522 unrelated, white British UKBB  
 232 participants<sup>21,22</sup>, we identified 73 and 89 individuals with a 16p11.2 BP4-5 (start: chr16:29.40-  
 233 29.80 Mb; end: chr16:30.05-30.40 Mb) deletion and duplication, respectively. CNV genotypes  
 234 were encoded to allow testing of four dosage mechanisms, namely a mirror model assessing the  
 235 additive impact of each additional copy, a U-shape model testing the same-direction impact of  
 236 any deviation from the copy-neutral state, and duplication- and deletion-only models that assess  
 237 the separate impact of duplications and deletions, respectively. Next, we evaluated the  
 238 association between an individual's CNV carrier status and 117 phenotypes – that comprise 57  
 239 quantitative variables including anthropometric measurements, vital signs, biomarker levels, life  
 240 history events, and 60 common diseases – while correcting for sex, age, age<sup>2</sup>, genotyping array,  
 241 and population stratification (Figure 2; Table S1).

### Table 2. Traits significantly associated with 16p11.2 BP4-5 CNVs.

Traits that are significantly ( $p \leq 0.05/117 = 4.3 \times 10^{-4}$ ) associated with 16p11.2 BP4-5 CNVs through at least one of the four assessed association models, following the ordering of Figure 2. (A) For quantitative traits, the mean value of the traits in copy-neutral individuals (controls) is provided along with the mean value and standard error (SE) among duplication and deletion carriers. The number of duplication and deletion carriers with available data is specified as N. Values are given in the indicated unit. (B) For binary disease traits, prevalence in percentage among copy-neutral individuals is provided along with prevalence and SE among duplication and deletion carriers. Diseased (case) and total (N) number of duplication and deletion carriers are indicated.

<b>CNV status</b>		<b>controls</b>		<b>deletion carriers</b>		<b>duplication carriers</b>
<b>A. Quantitative traits</b>						
	<b>unit</b>	<b>mean</b>	<b>N</b>	<b>mean ± SE</b>	<b>N</b>	<b>mean ± SE</b>
fluid intelligence score	points	6.24	26	5.15 ± 0.39	32	4.05 ± 0.37
platelet count	10 <sup>9</sup> cells/L	252.8	70	286.4 ± 7.5	85	219.4 ± 5.7
neutrophil count	10 <sup>9</sup> cells/L	4.25	70	5.00 ± 0.15	85	4.14 ± 0.15
heel bone mineral density	g/cm <sup>2</sup>	0.54	29	0.66 ± 0.04	47	0.46 ± 0.01
height	cm	168.8	73	163.3 ± 1.0	89	171.1 ± 1.1
insulin-like growth factor 1 (IGF-1)	nmol/L	21.4	71	18.9 ± 0.7	88	21.4 ± 0.6
hand grip strength	kg	30.7	73	26.7 ± 1.0	89	28.6 ± 1.2
vitamin D	nmol/L	49.8	68	36.0 ± 2.1	86	44.7 ± 2.2
forced vital capacity (FVC)	L	3.63	64	2.94 ± 0.11	72	3.64 ± 0.12
pulse rate	bpm	69.3	67	74.3 ± 1.5	81	71.3 ± 1.3
body mass index (BMI)	kg/m <sup>2</sup>	27.4	72	35.0 ± 0.9	89	26.3 ± 0.4
whole body fat mass	kg	24.9	67	35.5 ± 1.6	85	23.2 ± 0.9
glycated hemoglobin (HbA1c)	mmol/mol	36.0	70	43.3 ± 1.8	85	35.7 ± 0.5
waist-to-hip ratio (WHR)	-	0.87	72	0.97 ± 0.01	89	0.87 ± 0.01
C-reactive protein (CRP)	m/L	2.57	71	5.50 ± 0.76	88	2.03 ± 0.23
weight	kg	78.3	72	93.3 ± 2.7	89	77.1 ± 1.4
triglycerides	mmol/L	1.75	71	2.30 ± 0.14	87	1.60 ± 0.09
HDL cholesterol	mmol/L	1.46	61	1.26 ± 0.04	79	1.45 ± 0.04
WHR adjusted for BMI	-	0.00	71	0.03 ± 0.01	89	0.01 ± 0.01
gamma-glutamyltransferase (GGT)	U/L	37.3	71	64.1 ± 11.3	87	27.7 ± 2.8
alkaline phosphatase (ALP)	U/L	83.6	71	99.8 ± 3.4	88	79.8 ± 2.6
aspartate aminotransferase (AST)	U/L	26.2	70	35.8 ± 6.7	87	23.4 ± 0.7
alanine aminotransferase (ALT)	U/L	23.5	71	31.2 ± 2.1	88	20.8 ± 1.3
albumin	g/L	45.3	61	44.0 ± 0.4	79	45.6 ± 0.3
cystatin C	mg/L	0.91	71	1.04 ± 0.02	88	0.93 ± 0.02
creatinine	μmol/L	72.4	71	68.8 ± 1.6	87	76.9 ± 2.0
urate	μmol/L	309.4	70	355.2 ± 8.1	88	298.8 ± 9.1
age at menarche	years	12.9	27	11.7 ± 0.3	45	14.0 ± 0.4
relative age at first facial hair (group 1-3)	-	2.06	41	1.83 ± 0.06	38	2.26 ± 0.08
sex hormone binding globulin (SHBG)	nmol/L	51.9	61	38.4 ± 2.4	77	58.8 ± 3.4
<b>B. Disease Prevalence</b>						
		<b>prevalence</b>	<b>case/N</b>	<b>prevalence ± SE</b>	<b>case/N</b>	<b>prevalence ± SE</b>
sleep apnea		2.1%	9/60	15.0 ± 4.6%	0/70	0.0 ± 0.0%
recurrent depressive disorder		0.3%	0/49	0.0 ± 0.0%	3/68	4.4 ± 2.5%
bipolar affective disorder		0.4%	1/50	2.0 ± 2.0%	6/71	8.5 ± 3.3%
schizophrenia		0.2%	2/51	3.9 ± 2.7%	3/68	4.4 ± 2.5%
other anemias		5.5%	13/67	19.4 ± 4.8%	6/84	7.1 ± 2.8%
arthrosis (OA)		21.1%	23/62	37.1 ± 6.1%	15/74	20.3 ± 4.7%
pneumonia		5.9%	19/61	31.1 ± 5.9%	2/83	2.4 ± 1.7%
chronic obstructive pulmonary disease (COPD)		4.9%	17/55	30.9 ± 6.2%	4/76	5.3 ± 2.6%
asthma		12.1%	19/55	34.5 ± 6.4%	9/69	13.0 ± 4.1%
cardiac valve disorders		5.0%	7/33	21.2 ± 7.1%	3/54	5.6 ± 3.1%
cardiac conduction disorders		19.5%	18/44	40.9 ± 7.4%	13/64	20.3 ± 5%
essential hypertension (HTN)		35.3%	36/62	58.1 ± 6.3%	23/74	31.1 ± 5.4%
type 1 diabetes (T1D)		1.0%	4/42	9.5 ± 4.5%	1/75	1.3 ± 1.3%
lipidemias & lipoprotein disorders		22.0%	23/48	47.9 ± 7.2%	11/57	19.3 ± 5.2%
acute kidney injury (AKI)		4.7%	20/65	30.8 ± 5.7%	7/71	9.9 ± 3.5%
chronic kidney disease (CKD)		4.4%	9/54	16.7 ± 5.1%	6/70	8.6 ± 3.3%



**Figure 2. 16p11.2 BP4-5 phenome-wide association study.**

Effect sizes (beta; x-axis) with 95% confidence interval (CI) of the 16p11.2 BP4-5 deletion (circle) and duplication (square) on 117 complex traits and diseases, ordered by physiological system (y-axis). Effect sizes are in standard deviation units of the outcome (quantitative traits) or logarithms of the odds ratio of a logistic regression (disease traits). Deletion- and duplication-only effects that are significant under Bonferroni correction ( $p \leq 0.05/117 = 4.3 \times 10^{-4}$ ) are in blue and red, respectively, while non-significant effects are in grey. If the most significant among the four tested association models was the mirror or U-shape model, it is denoted with an “M” or “U”, respectively (right).

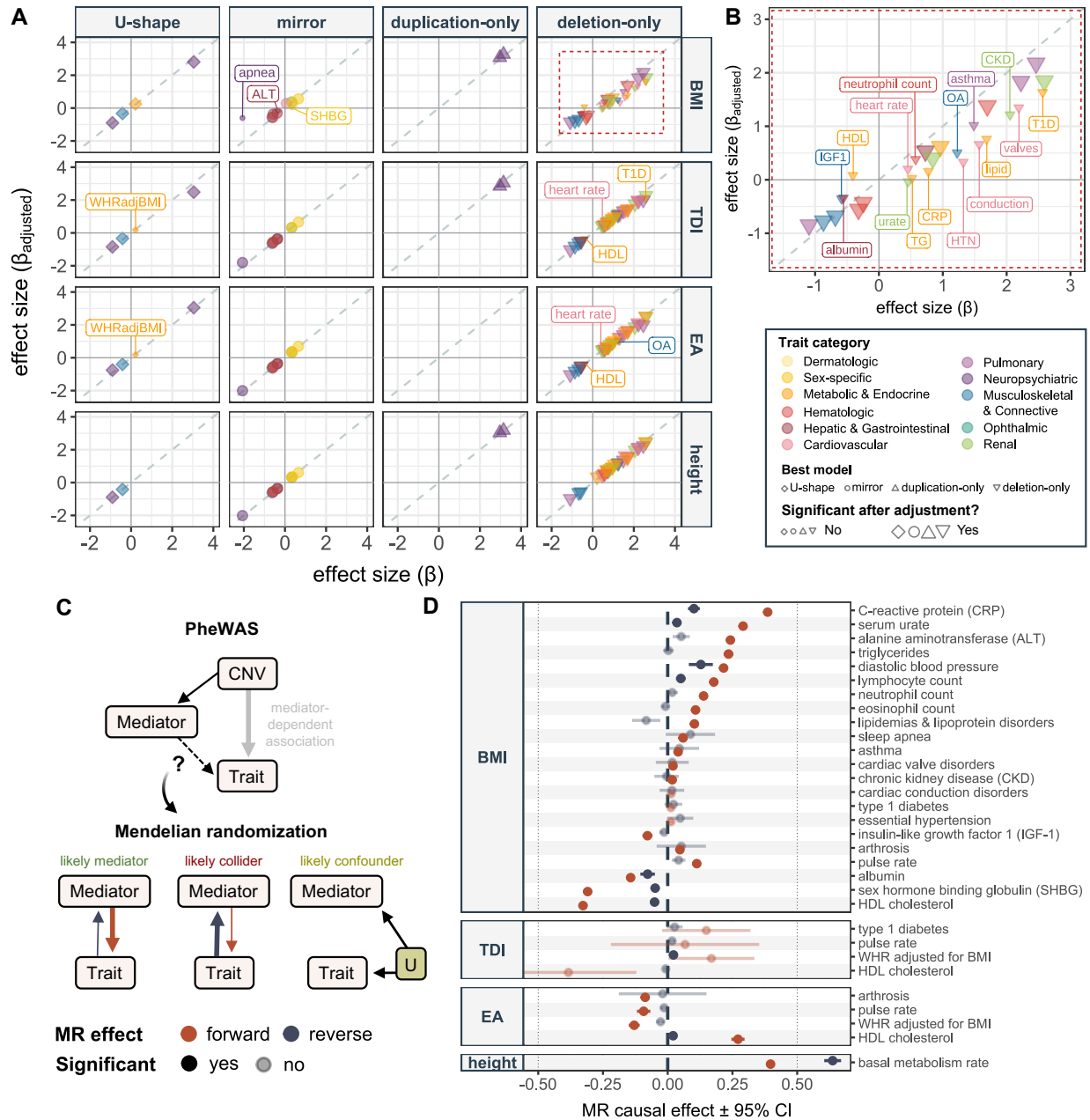
242 Overall, 46 (39%) traits, including 16 diseases, were associated with the CNV carrier status under  
 243 at least one association model (Bonferroni correction:  $p \leq 0.05/117 = 4.3 \times 10^{-4}$ ; [Table 2](#)), with an  
 244 additional 32 (27%) showing a trend for association (nominal significance:  $p \leq 0.05$ ). Specifically,  
 245 10 and 38 traits showed a significant association through the duplication-only and deletion-only  
 246 models, respectively, indicating a stronger propensity for pleiotropy and deleteriousness of the  
 247 deletion, compared to the duplication. Exceptions are recurrent depression and bipolar disorder,  
 248 the two only traits for which the duplication-only model yielded the most significant result. This is  
 249 in line with the duplication representing a strong susceptibility factor for psychiatric conditions<sup>12,14–</sup>  
 250 <sup>19</sup>. Similarly, the risk for schizophrenia was strongly increased by the duplication, even if our  
 251 analysis finds that the relation is better described by a U-shape model wherein the deletion also  
 252 tends to increase schizophrenia risk. Surprisingly, the CNV did not associate with neuroticism



253 score, despite the high genetic correlation between neuroticism and psychiatric conditions<sup>42</sup>.  
254 Three other traits, namely fluid intelligence, vitamin D, and waist-to-hip ratio adjusted for BMI  
255 (WHRadjBMI), were also most significantly associated through a U-shape effect, while grip  
256 strength was decreased in both deletion and duplication carriers, but more strongly in the former.  
257 Conversely, ten traits were most significantly associated through a mirror model, including multiple  
258 hepatic biomarkers, platelet count, and traits related to sexual characteristics such as puberty  
259 timing and sex hormone binding globulin (SHBG) levels. Finally, the deletion-only model was the  
260 most significant fit for 30 phenotypes, including mostly pulmonary, cardiovascular, metabolic, and  
261 renal traits.

262  
263 Having characterized the pleiotropic nature of 16p11.2 BP4-5 rearrangements, we next sought to  
264 establish whether some of these associations might be secondary to the CNV affecting core  
265 mediatory phenotypes, i.e., reflect indirect pleiotropy (Figure 1B). We focus on four traits that  
266 proxy hallmark features of the 16p11.2 BP4-5 rearrangement and have the potential to influence  
267 other associated traits: i) BMI, which characterizes the negative correlation between dosage and  
268 adiposity<sup>8–10,20,21,24,25</sup> and represents a major risk factor for many common diseases; ii) Height,  
269 which is reduced in deletion carriers<sup>16,20,21,24,36</sup> and can influence musculoskeletal phenotypes; iii)  
270 Educational attainment (EA) proxied by age at which an individual completed their education. This  
271 variable offers the advantage of being available for the near totality of the UKBB cohort while  
272 strongly correlating with fluid intelligence score that is limited to about half of its participants  
273 (Pearson correlation = 0.42), thereby reflecting the decreased cognitive function observed in both  
274 duplication and deletion carriers<sup>2,3,16,17,21,31,43</sup>; and iv) Townsend deprivation index (TDI) as a  
275 measure of socio-economic status (SES), which we expect to be reduced as a corollary of the  
276 health burden imposed by the CNV<sup>20</sup>. Of note, while TDI specifically aims at assessing SES, BMI,  
277 height, and EA also partly capture SES<sup>44</sup>. For the association between CNV and phenotype to be  
278 mediated by one of these factors, the mediator needs to significantly ( $p \leq 0.05/117 = 4.3 \times 10^{-4}$ )  
279 associate with the tested phenotype. Furthermore, phenotypes cannot be too correlated with the  
280 mediator (Pearson's correlation > 0.5), as in such situations distinguishing mediator and outcome  
281 would be particularly difficult. For all mediator-trait pairs that fulfilled these criteria, we tested the  
282 impact of adjusting the CNV-trait effect for mediatory factors by including them individually in the  
283 regression model yielding the most significant CNV-trait effect (Figure 3A; Table S2).





**Figure 3. Mediation of 16p11.2 BP4-5 pleiotropy through anthropometric traits and socio-economic factors.**

(A) Effects (beta) of 16p11.2 BP4-5 CNVs on traits with adjustment for potential mediators (y-axis) – i.e., body mass index (BMI), Townsend deprivation index (TDI), age at end of education (EA), and height (rows, right) – against those without adjustment (x-axis), stratified (columns, top) according to the best (i.e., most significant) association model (shape). Only associations that were significant prior to or become significant after adjustment are plotted. Traits are colored according to physiological system. Size reflects whether the effect is Bonferroni significant after adjusting for the potential mediator (large) or not (small). Traits losing significance upon adjustment are labeled. Grey dashed diagonal represents the identity line. (B) Enlargement of the area delimited by a red dashed rectangle in (A), showing the effect of BMI adjustment for deletion-driven association, using the same legend as in (A). (C) Schematic of the links between copy-number variant (CNV), potential mediators, and assessed traits. Covariate-adjusted phenome-wide association studies (PheWAS) identified CNV-trait associations that are dependent on either of the four tested factors (thick grey arrow) in (A). This scenario can be through mediation, collider bias, or confounding. We used Mendelian randomization (MR) to assess the genetically determined causal effect of the putative mediator on the trait (forward effect, red arrow) and of the trait on the mediator (reverse effect; dark blue arrow). MR effect arrows are proportional to

causal effect sizes. When the forward effect is larger than the reverse one, mediation is a likely scenario; when the reverse effect is larger, the putative mediator likely acts as a collider; absence of causal effects likely indicates presence of an unobserved confounder, U. Depending on the scenario, adjustment for the mediator in the regression analysis might (green) or might not (red) be appropriate, as reflected by the color of each scenario's title. **(D)** Bidirectional forward (red) and reverse (dark blue) MR effects with 95% confidence interval (CI; x-axis truncated on the right) of potential mediators (left y-axis) on traits (right y-axis) for all mediator-trait pairs that either gained or lost significance upon adjustment for the mediator. Non-significant effects ( $p > 0.05/62 = 8.1 \times 10^{-4}$ ) are semi-transparent. ALT = alanine aminotransferase; CKD = chronic kidney disease; CRP = C-reactive protein; HDL = high-density lipoprotein cholesterol; HTN = essential hypertension; IGF-1 = insulin-like growth factor 1; OA = arthritis; SHBG = sex hormone binding globulin; T1D = type 1 diabetes; TG = triglycerides; WHRadjBMI = waist-to-hip ratio adjusted for BMI.

284 Upon adjustment for BMI, TDI, EA, and height, nineteen, four, four, and zero CNV-trait  
285 associations fell below the significance cutoff ( $p \leq 0.05/117 = 4.3 \times 10^{-4}$ ), respectively. Comparing  
286 effect sizes, only the mirror association with sleep apnea was nominally significantly reduced upon  
287 adjustment for BMI ( $p = 0.04$ ). Remarkably, the association with basal metabolic rate (deletion-  
288 only) became significant upon adjustment for height, while the one with diastolic blood pressure  
289 (mirror), eosinophil count (deletion-only), and lymphocyte count (deletion-only) became so upon  
290 adjustment for BMI (Figure S1A), even though the change in effect size were not significant ( $p >$   
291  $0.45$ ). The impact of adjusting for BMI was most striking on deletion-driven associations, for which  
292 61% (16/26) of the associations fell below the significance threshold (Figure 3B). In line with  
293 expectations, BMI-dependent traits tended to have a stronger correlation with BMI than those that  
294 remained significant upon adjustment for BMI ( $p = 0.05$ ) (Figure S1B). Among the lost  
295 associations, we find nine out of the ten metabolic and cardiovascular traits associated with the  
296 deletion. These associations likely reflect secondary consequences of the propensity for obesity  
297 of deletion carriers as they include levels of serum lipid and the inflammation biomarker C-reactive  
298 protein (CRP), cardiac valve and conduction disorders, and hypertension. The effect of BMI on  
299 musculoskeletal, pulmonary, or renal traits is more balanced, with some associations, such as the  
300 ones with arthritis (OA), asthma, or urate and chronic kidney disease (CKD), appearing to be  
301 driven by BMI, while others, such as grip strength, chronic obstructive pulmonary disease  
302 (COPD), or cystatin C and acute kidney injury (AKI), remaining significant upon BMI adjustment.  
303 The mediating role of TDI and EA was much milder, as only four associations were lost upon  
304 adjustment for either variable – including a shared association with WHR adjusted for BMI, heart  
305 rate, and high-density lipoprotein (HDL) cholesterol – suggesting that TDI and EA capture partially  
306 overlapping mediatory processes. Surprisingly, associations with psychiatric disorders were not  
307 affected by EA, suggesting that cognition and psychiatric diseases are regulated by (at least  
308 partially) independent pathways. Finally, the observation that no associations were affected by  
309 adjusting for height confirms that the decrease in traits such as grip strength and forced vital  
310 capacity among deletion carriers is not driven solely by their short stature.

311

312 One caveat of our analysis is that it cannot distinguish whether changes in CNV-trait associations  
313 are indeed secondary effects of the mediator on the trait. At least three scenarios could result in  
314 the loss (or gain) of a CNV-trait effect upon covariate adjustment (Figure 3C). The first one is  
315 mediation, wherein the CNV affects the trait through the mediator, resulting in a dominant causal  
316 effect of the mediator on the trait. The second scenario is when the variable we adjusted for turns  
317 out to be a collider of the CNV and the trait, in which case we expect a dominant causal effect  
318 from the trait to the "mediator". Finally, data could be explained by an unobserved confounder that  
319 affects both the adjustment variable and the trait, in which case we do not expect any causal link  
320 between trait and mediator. Of note, in the latter scenario, we further distinguish between whether  
321 the CNV has an impact on the confounder, the "mediator", the trait, or a combination thereof.  
322 Importantly, adjusting for the mediator in the regression model is an appropriate solution to obtain  
323 meaningful direct CNV-trait effects (i.e., genuine direct pleiotropy) only in the i) mediator scenario  
324 or ii) the confounder scenario where the CNV has a direct effect on the trait, in which case  
325 adjustment for the mediator could result in a gain of power (Figure 3C). To identify cases where  
326 mediation is a likely scenario, we resorted to bidirectional Mendelian randomization (MR), a  
327 causal inference approach that allows to estimate the genetically determined causal effect of an  
328 exposure on an outcome (Figure 3D; Table S3). Firstly, we estimated the forward mediator-to-trait  
329 effect for all 31 mediator-trait pairs that either gained (N = 4; Figure S1A) or lost (N = 27; Figure  
330 3A-B) significance upon adjustment for the mediator. Except for the four TDI-dependent  
331 associations which had large confidence intervals due to the lack of good genetic instruments for  
332 TDI and the effect of BMI on hypertension, type 1 diabetes, and cardiac conduction disorders, all  
333 effects were significant ( $p \leq 0.05/62 = 8.1 \times 10^{-4}$ ), confirming that the mediators can causally  
334 influence the involved traits. Secondly, we estimated the reverse trait-to-mediator causal effects.  
335 Ten reverse effects were significant and thus represent mediator-trait pairs at risk for collider bias.  
336 Yet, for nine of them, the forward effect had a larger magnitude, making the mediator-to-trait  
337 causal path more likely. The only exception is the association between the deletion and basal  
338 metabolic rate that became significant upon adjustment for height and for which the reverse effect  
339 was stronger than the forward effect. This suggests that height could act as a collider and  
340 adjustment for it could bias estimates. Hence, we conclude based on the unadjusted effect that  
341 the association between the deletion and basal metabolic rate is non-significant. It is also worth  
342 noting that six out of seven associations lacking a significant forward effect also lacked a  
343 significant reverse effect, possibly indicating presence of an unobserved confounder. This is  
344 particularly likely for the BMI effect on hypertension, type 1 diabetes, and cardiac conduction

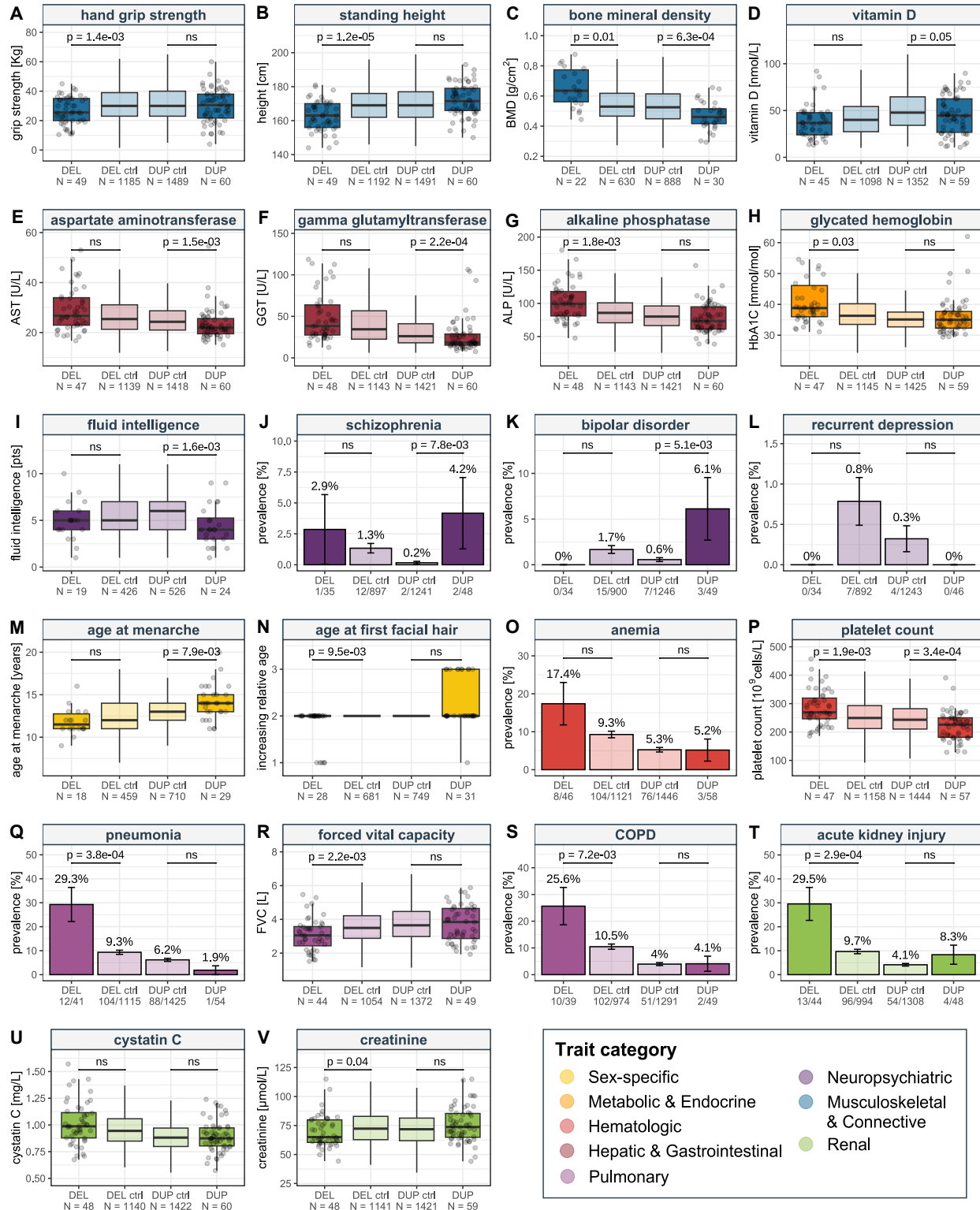
345 disorders, where estimates are close to null despite being well-instrumented ( $\geq 50$  instruments).  
346 Globally, these analyses support that a large fraction (74%) of the flagged associations are likely  
347 indirect consequences of the CNV's effect on our selected mediators.

348

349 Next, we focused on the 22 traits whose association with the CNV remained significant after  
350 adjusting for BMI, height, TDI, or EA. To confirm that these represent cases of genuine direct  
351 pleiotropy, we used a matched-control approach that offers the advantage of allowing adjustment  
352 for multiple mediatory variables at once but at the cost of losing some statistical power.  
353 Specifically, for each of the 58 deletion and 61 duplication carriers with sufficient data to carry out  
354 the matching, we identified individuals with matched age ( $\pm 2.5$  years), sex (identical), BMI ( $\pm 2.5$   
355  $\text{kg/m}^2$ ), TDI ( $\pm 2$ ), income class (identical), and EA ( $\pm 1$  year) among a pool of copy-neutral,  
356 unrelated, white British UKBB participants (Figure S2). For 49 deletion and 60 duplication carriers,  
357 at least 25 matched controls could be identified, and phenotype mean or disease prevalence  
358 between the two CNV groups and their respective controls were compared (Figure 4; Tables S4-  
359 5). Eleven traits (50%) retained a strictly significant effect ( $p \leq 0.05/22 = 2.3 \times 10^{-3}$ ), affecting six  
360 independent physiological systems: musculoskeletal, neuropsychiatric, pulmonary/immune,  
361 renal, hepatic, and hematological. Specifically, deletion carriers presented with decreased hand  
362 grip strength ( $p = 1.4 \times 10^{-3}$ ; Figure 4A), shorter stature ( $p = 1.2 \times 10^{-5}$ ; Figure 4B), increased  
363 alkaline phosphatase (ALP;  $p = 1.8 \times 10^{-3}$ ; Figure 4G), decreased forced vital capacity (FVC;  $p =$   
364  $2.2 \times 10^{-3}$ ; Figure 4R), and increased risk for pneumonia ( $p = 3.8 \times 10^{-4}$ ; Figure 4Q) and AKI ( $p =$   
365  $2.9 \times 10^{-4}$ ; Figure 4T). Duplication carriers showed decreased bone mineral density ( $p = 6.3 \times 10^{-}$   
366  $4$ ; Figure 4C), lower aspartate aminotransferase (AST;  $p = 1.5 \times 10^{-3}$ ; Figure 4E) and gamma-  
367 glutamyltransferase (GGT;  $p = 2.2 \times 10^{-4}$ ; Figure 4F) levels, and reduced fluid intelligence ( $p = 1.6$   
368  $\times 10^{-3}$ ; Figure 4I). Noteworthy is the strong mirror effect on platelet count (Figure 4P), with higher  
369 ( $p = 1.9 \times 10^{-3}$ ) and lower ( $p = 3.4 \times 10^{-4}$ ) counts observed in deletion and duplication carriers,  
370 respectively. Whereas for the other phenotypes the other CNV type did not meet the strict  
371 significance criteria, all effects, showed a trend for a mirror effect, except for fluid intelligence and  
372 AKI, which showed a U-shape trend. Besides reinforcing its long-established consequence on  
373 cognitive function, our results assert the role of the hepatic, musculoskeletal, and pulmonary  
374 systems in the 16p11.2 BP4-5 pathophysiology through mechanisms that are independent of the  
375 CNV's impact on anthropometric and socio-economic traits.

376

377



**Figure 4. 16p11.2 BP4-5 CNV carriers matched-control analyses.**

**(A-V)** Comparison between deletion (DEL) and duplication (DUP) carriers (darker shade) and their respective matched controls (DEL or DUP ctrl; lighter shade) for 22 traits that remained Bonferroni-significant after individually adjusting for body mass index (BMI), height, Townsend deprivation index (TDI), and age at end of education (EA) in Figure 3. For quantitative traits, data are represented as boxplots without outliers and data points for CNV carriers are shown as grey

dots. Sample size of each group is indicated as N. P-values of two-sided t-tests comparing CNV carriers to matched controls are indicated. For binary traits, bars represent disease prevalence in percentage and error bars represent the standard error. Number of cases and total sample size for each group is indicated. P-values of two-sided Fisher tests comparing CNV carriers to matched controls are indicated. "ns" indicates  $p > 0.05$ . Traits are colored according to physiological systems. COPD = chronic obstructive pulmonary disease.

378 Finally, we performed sensitivity analyses to validate the robustness of our conclusions. As a  
379 negative control, we performed the matched-control analysis for the 24 traits that were  
380 significantly associated with 16p11.2 BP4-5 CNVs in our PheWAS but whose association was  
381 dependent on adjustment for mediators or that could not be tested in the covariate analysis due  
382 to high trait-mediator correlation (Figure S3; Tables S4-5). In line with these associations being  
383 secondary consequences to the effect of the CNV on factors on which the matching was  
384 performed, only three traits had a nominally significant CNV association, and none survived  
385 Bonferroni correction. This strongly contrasts with our main matched-control analysis, where only  
386 three traits *lacked* a nominally significant effect: recurrent depression (Figure 4L), anemia (Figure  
387 4O), and cystatin C (Figure 4U). This absence of results could either be the result of a loss in  
388 statistical power resulting from CNV carrier subsampling or by these associations being driven by  
389 a combination of factors on which the matching was performed. The former could be exacerbated  
390 by the fact that CNV carriers with the more extreme phenotypes were less likely to have 25  
391 matched controls in the UKBB. To explore this hypothesis, we compared mean trait value or  
392 disease prevalence between the subset of CNV carriers used for the matched-control analysis  
393 and the one excluded due to missing data or lack of a sufficient number of matched controls  
394 (Figure S4; Tables S4-5). Except for recurrent depression and FVC, all comparisons were non-  
395 significant ( $p \geq 0.05$ ), indicating that subsampling does not strongly impact our results. For  
396 recurrent depression, the only three duplication carriers diagnosed with the disease were not  
397 included in the matched-control analysis ( $p = 0.03$ ; Figure S4L), indicating that the non-significant  
398 effect of the duplication on recurrent depression (Figure 4L) is likely caused by subsampling. For  
399 FVC, excluded deletion carriers exhibited a more pronounced phenotypic decrease than the ones  
400 retained for the matched-control analysis ( $p = 0.02$ ; Figure S4R), suggesting that an even more  
401 extreme difference would have been observed if these individuals had been included in the  
402 matched-control analysis (Figure 4R). Conversely, the role of the CNV on anemia risk and cystatin  
403 C is likely driven by the effect of the CNV on adiposity and socio-economic status.

404

## 405 Discussion

406 In this study, we perform a comprehensive PheWAS assessing the relation between 16p11.2 BP4-  
407 5 CNVs and 117 complex traits and diseases in the general population through four dosage



408 mechanisms of action. Our results confirm the extreme pleiotropy of 16p11.2 BP4-5  
409 rearrangements, with 46 traits associating with the CNV. In line with the more deleterious nature  
410 of the deletion, haploinsufficiency associated with 38 unique traits, while only 10 traits associated  
411 with the region's duplication. Further emphasizing how the same genetic region can affect  
412 different traits through different dosage mechanisms, we identify traits for which the loss and gain  
413 of a copy had an opposite (e.g., BMI or platelet count) or alternatively, a similar (e.g., grip strength  
414 or fluid intelligence) consequences on the phenotype. Besides assessing the role of dosage in  
415 pleiotropy, we also estimated the fraction of associations that are likely to be secondary to some  
416 hallmark features of the CNV and validated through bidirectional MR that mediation is a likely  
417 scenario. While height did not mediate any associations, sixteen (61%) of the deletion-driven  
418 associations were found to be BMI-dependent, thirteen of which (81%) received support from MR  
419 for a scenario wherein the association is consequential to an initial increase in BMI. Conversely,  
420 the role of EA and TDI was more subtle, with only five associations showing confounding by these  
421 factors. Importantly, some associations were found to be independent of all the tested mediators,  
422 suggesting genuine direct pleiotropy of the region on musculoskeletal, hepatic, metabolic,  
423 neuropsychiatric, reproductive, hematological, pulmonary, immune, and renal function.

424  
425 Our findings have far-reaching consequences for clinical practice and highlight knowledge gaps.  
426 First, our results show that increased BMI in deletion carriers drives numerous adult-onset  
427 comorbidities. Studies have shown that weight gain in 16p11.2 BP4-5 deletion carriers starts  
428 during early childhood, to rapidly progress to obesity<sup>9,16,45-47</sup>. This emphasizes the importance of  
429 following pediatric cases by a dedicated team of endocrinologists and nutritionists who can  
430 implement a weight control strategy at an early age to attenuate ensuing adult co-morbidities.  
431 Second, we show that some other traits are affected independently of the CNV's effect on BMI,  
432 cognition, and SES. Besides recapitulating well-established hallmark features, such as the CNV's  
433 negative impact on cognitive ability or the duplication-specific risk of bipolar disorder or  
434 depression, we also link the CNV with milder afflictions of systems that had previously been  
435 implicated in clinical cohorts. For instance, increased risk for AKI might be the consequence of  
436 subclinical structural defects of the kidney that could affect renal function in the long term,  
437 paralleling the predisposition of deletion carriers to congenital anomalies of the kidney and urinary  
438 tract<sup>48-50</sup>. Similarly, increased risk for pneumonia might reflect an impaired immune system that is  
439 exacerbated into a full-blown immunodeficiency in deletion carriers that also present with a loss-  
440 of-function variant in *CORO1A*<sup>51</sup> (MIM: 605000). Other traits that are affected through BMI-  
441 independent mechanisms, such as bone mineral density, platelet counts, pulmonary function, and

442 liver enzymes have not been linked with the CNV in clinical cohorts and future research should  
443 establish how often these traits are altered in carriers and which are the molecular mechanisms  
444 that mediate this pleiotropy. These could be explored by gene-to-trait mapping strategies such as  
445 rare variant gene burden tests<sup>52,53</sup>, as well as MR<sup>54</sup> or colocalization<sup>55</sup> that integrate association  
446 signals from common SNP-GWAS with transcriptomic and proteomic data to pinpoint genes linked  
447 with specific phenotypes. These data could also be leveraged to generate gene-by-trait  
448 association matrices whose clustering may reveal groups of traits with shared underlying genetic  
449 influences and for which CNV associations are more likely to disappear upon adjustment for one  
450 another. Thirdly, our results expose intriguing findings, casting light on questions that remain  
451 unanswered by the current study. For instance, the BMI-dependent association of the deletion  
452 with type 1 diabetes could be driven by misdiagnosing type 2 diabetes as type 1 due to early-  
453 onset diabetes following early-onset obesity. We also identify an association between the deletion  
454 and decreased creatinine levels. Creatinine levels are typically *elevated* in patients with renal  
455 dysfunction, as is the case for many deletion carriers. We speculate that these results could be  
456 the consequences of reduced hepatic function or muscle mass, both of which are present among  
457 deletion carriers. Similarly, it remains unclear whether elevated levels of ALP – for which levels of  
458 specific isoenzymes were not determined in UKBB – reflect hepatic, renal, or skeletal dysfunction.  
459 Validation of these hypotheses requires in-depth phenotypic characterization of carriers' medical  
460 records but will be crucial to better define the molecular pathophysiology of 16p11.2 BP4-5 CNV  
461 carriers and hopefully lead to actionable insights related to the management of the condition's co-  
462 morbidities.

463

464 Our study is not without limitations. First, by assessing a relatively homogenous cohort, our study  
465 likely misses pleiotropic consequences that are only expressed in certain genetic or  
466 environmental backgrounds, a phenomenon exacerbated by the relatively small absolute number  
467 of CNV carriers which hinders our statistical power. Future studies are needed to confirm trends  
468 that we observe at sub-significant level. Second, we decided to focus on only four covariates,  
469 which based on the literature, represent strong candidates to mediate indirect pleiotropic  
470 consequences of the region's rearrangement. While height and BMI can be measured with  
471 relatively high accuracy, EA and TDI only offer rough and imperfect proxies for complex  
472 characteristics such as cognitive function and SES, possibly explaining their weaker mediatory  
473 role. Other factors that we did not assess might mediate the relation between 16p11.2 BP4-5  
474 CNVs and some of the associated traits. Third, the conducted MR analysis comes with its own  
475 limitations, namely violation of the exclusion-restriction assumption via correlated pleiotropy,



476 which may have resulted in false positive mediator-to-trait causal effects<sup>56,57</sup>. Still, if both adjusted  
477 and unadjusted regression analyses show a significant CNV effect, we can convincingly suggest  
478 that independent pleiotropic mechanisms are at play. Finally, while our study brings us a step  
479 closer to understanding the pleiotropy of the region, it fails to provide molecular insights into  
480 mechanisms of pleiotropy, for which experimental approaches and leveraging of other mutational  
481 classes offer promising avenues.

482

483 In conclusion, our study provides a framework to start disentangling the complex pleiotropic  
484 patterns associated with genomic disorders. For 16p11.2 BP4-5, the latter appears to be a mixture  
485 of indirect effects mediated by the impact of the CNV on adiposity and cognition, and direct effects  
486 on a broad range of physiological systems. This suggests that independent molecular  
487 mechanisms are involved in translating dosage changes into the many co-morbidities linked to  
488 the genomic disorder.

489

## 490 **Data & Code availability**

491 UK Biobank data are available for registered users and were accessed through the application  
492 number 16389. Summary statistics used for Mendelian randomization studies are publicly  
493 available, as described in the Methods. All data produced by this study are available in  
494 [Supplemental Tables](#). Code will be made available upon publication at:  
495 [https://github.com/cauwerx/16p11.2\\_BP4-5\\_pleiotropy](https://github.com/cauwerx/16p11.2_BP4-5_pleiotropy).

496

## 497 **Author contributions**

498 CA performed all analyses, except for MR analyses conducted by SM; ZK supervised statistical  
499 analyses; CA, ZK, and AR interpreted the data; CA generated the figures and drafted the  
500 manuscript; ZK and AR made critical revisions; All authors approved the final manuscript.

501

## 502 **Acknowledgments**

503 We thank UKBB biobank participants for sharing their data. Computations were performed on the  
504 Urblauna server from the University of Lausanne. The study was funded by the Swiss National  
505 Science Foundation (31003A\_182632, AR; 310030\_189147, ZK), Horizon2020 Twinning projects  
506 (ePerMed 692145, AR), and the Department of Computational Biology (ZK) and the Center for  
507 Integrative Genomics (AR) from the University of Lausanne.

508

## 509 Declaration of interests

510 The authors have no conflicts of interest to declare.

511

## 512 References

- 513 1. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R.,  
514 Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. *Nat Genet* 43, 838–846.  
515 10.1038/ng.909.
- 516 2. Bijlsma, E.K., Gijbbers, A.C.J., Schuurs-Hoeijmakers, J.H.M., Van Haeringen, A., Fransen Van De Putte, D.E.,  
517 Anderlid, B.-M., Lundin, J., Lapunzina, P., Pé Rez Jurado, L.A., Delle Chiaie, B., et al. (2009). Extending the  
518 phenotype of recurrent rearrangements of 16p11.2: Deletions in mentally retarded patients without autism and in  
519 normal individuals. *Eur J Med Genet* 52, 77–87. 10.1016/j.ejmg.2009.03.006.
- 520 3. Rosenfeld, J.A., Coppinger, J., Bejjani, B.A., Girirajan, S., Eichler, E.E., Shaffer, L.G., and Ballif, B.C. (2010).  
521 Speech delays and behavioral problems are the predominant features in individuals with developmental delays  
522 and 16p11.2 microdeletions and microduplications. *J Neurodev Disord* 2, 26–38. 10.1007/S11689-009-9037-4.
- 523 4. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A.,  
524 Kendall, J., et al. (2007). Strong Association of De Novo Copy Number Mutations with Autism. *Science* (1979) 316,  
525 445–449. 10.1126/science.1138659.
- 526 5. Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira,  
527 M.A.R., Green, T., et al. (2008). Association between Microdeletion and Microduplication at 16p11.2 and Autism.  
528 *N Engl J Med* 358, 667–675. 10.1056/NEJMoa075974.
- 529 6. Kumar, R.A., Karamohamed, S., Sudi, J., Conrad, D.F., Brune, C., Badner, J.A., Gilliam, T.C., Nowak, N.J., Cook,  
530 E.H., Dobyns, W.B., et al. (2008). Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* 17, 628–638.  
531 10.1093/HMG/DDM376.
- 532 7. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren,  
533 Y., et al. (2008). Structural Variation of Chromosomes in Autism Spectrum Disorder. *Am J Hum Genet* 82, 477–  
534 488. 10.1016/j.ajhg.2007.12.009.
- 535 8. Jacquemont, S., Reymond, A., Zufferey, F., Harewood, L., Walters, R.G., Kutalik, Z., Martinet, D., Shen, Y.,  
536 Valsesia, A., Beckmann, N.D., et al. (2011). Mirror extreme BMI phenotypes associated with gene dosage at the  
537 chromosome 16p11.2 locus. *Nature* 478, 97–102. 10.1038/nature10406.
- 538 9. Walters, R.G., Jacquemont, S., Valsesia, A., De Smith, A.J., Martinet, D., Andersson, J., Falchi, M., Chen, F.,  
539 Andrieux, J., Lobbens, S., et al. (2010). A new highly penetrant form of obesity due to deletions on chromosome  
540 16p11.2. *Nature* 463, 671–675. 10.1038/nature08727.
- 541 10. Bochukova, E.G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., Saeed, S., Hamilton-Shield, J.,  
542 Clayton-Smith, J., O’Rahilly, S., et al. (2010). Large, rare chromosomal deletions associated with severe early-  
543 onset obesity. *Nature* 463, 666–670. 10.1038/nature08689.
- 544 11. Shinawi, M., Liu, P., Kang, S.H.L., Shen, J., Belmont, J.W., Scott, D.A., Probst, F.J., Craigen, W.J., Graham, B.H.,  
545 Pursley, A., et al. (2010). Recurrent reciprocal 16p11.2 rearrangements associated with global developmental  
546 delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet* 47, 332–341.  
547 10.1136/JMG.2009.073015.
- 548 12. McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J., Yoon, S., Perkins, D.O., Dickel, D.E.,  
549 Kusenda, M., Krastoshevsky, O., et al. (2009). Microduplications of 16p11.2 are associated with schizophrenia.  
550 *Nat Genet* 41, 1223–1227. 10.1038/ng.474.
- 551 13. Mefford, H.C., Muhle, H., Ostertag, P., von Spiczak, S., Buysse, K., Baker, C., Franke, A., Malafosse, A., Genton,  
552 P., Thomas, P., et al. (2010). Genome-Wide Copy Number Variation in Epilepsy: Novel Susceptibility Loci in  
553 Idiopathic Generalized and Focal Epilepsies. *PLoS Genet* 6, e1000962. 10.1371/JOURNAL.PGEN.1000962.
- 554 14. Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M.,  
555 Malhotra, D., Bhandari, A., et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental  
556 pathways in schizophrenia. *Science* (1979) 320, 539–543. 10.1126/science.1155174.
- 557 15. Hanson, E., Bernier, R., Porche, K., Jackson, F.I., Goin-Kochel, R.P., Green Snyder, L., Snow, A. V, Stevens  
558 Wallace, A., Campe, K.L., Zhang, Y., et al. (2015). The Cognitive and Behavioral Phenotype of the 16p11.2 Deletion  
559 in a Clinically Ascertained Population. *Biol Psychiatry* 77, 785–793. 10.1016/j.biopsych.2014.04.021.

- 560 16. Zufferey, F., Sherr, E.H., Beckmann, N.D., Hanson, E., Maillard, A.M., Hippolyte, L., Macé, A., Ferrari, C., Kutalik,  
561 Z., Andrieux, J., et al. (2012). A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and  
562 neuropsychiatric disorders. *J Med Genet* 49, 660–668. 10.1136/jmedgenet-2012-101203.
- 563 17. D'Angelo, D., Lebon, S., Chen, Q., Martin-Brevet, S., Snyder, L.A.G., Hippolyte, L., Hanson, E., Maillard, A.M.,  
564 Faucett, W.A., Macé, A., et al. (2016). Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and  
565 Medical Comorbidities. *JAMA Psychiatry* 73, 20–30. 10.1001/JAMAPSYCHIATRY.2015.2123.
- 566 18. Green Snyder, L.A., D'Angelo, D., Chen, Q., Bernier, R., Goin-Kochel, R.P., Wallace, A.S., Gerdtts, J., Kanne, S.,  
567 Berry, L., Blaskey, L., et al. (2016). Autism Spectrum Disorder, Developmental and Psychiatric Features in 16p11.2  
568 Duplication. *J Autism Dev Disord* 46, 2734–2748. 10.1007/S10803-016-2807-4.
- 569 19. Niarchou, M., Chawner, S.J.R.A., Doherty, J.L., Maillard, A.M., Jacquemont, S., Chung, W.K., Green-Snyder, L.A.,  
570 Bernier, R.A., Goin-Kochel, R.P., Hanson, E., et al. (2019). Psychiatric disorders in children with 16p11.2 deletion  
571 and duplication. *Transl Psychiatry* 9, 8. 10.1038/s41398-018-0339-8.
- 572 20. Aguirre, M., Rivas, M.A., and Priest, J. (2019). Phenome-wide Burden of Copy-Number Variation in the UK  
573 Biobank. *Am J Hum Genet* 105, 373–383. 10.1016/j.ajhg.2019.07.001.
- 574 21. Auwerx, C., Lepamets, M., Sadler, M.C., Patxot, M., Stojanov, M., Baud, D., Mägi, R., Porcu, E., Reymond, A., and  
575 Kutalik, Z. (2022). The individual and global impact of copy-number variants on complex human traits. *Am J Hum*  
576 *Genet* 109, 647–668. 10.1016/j.ajhg.2022.02.010.
- 577 22. Auwerx, C., Jöeloo, M., Sadler, M.C., Tesio, N., Ojavee, S., Clark, C.J., Mägi, R., Esko, T., Metspalu, A., Milani, L.,  
578 et al. (2024). Rare copy-number variants as modulators of common disease susceptibility. *Genome Med* 16, 5.  
579 10.1186/s13073-023-01265-5.
- 580 23. Crawford, K., Bracher-Smith, M., Owen, D., Kendall, K.M., Rees, E., Pardiñas, A.F., Einon, M., Escott-Price, V.,  
581 Walters, J.T.R., O'Donovan, M.C., et al. (2019). Medical consequences of pathogenic CNVs in adults: Analysis of  
582 the UK Biobank. *J Med Genet* 56, 131–138. 10.1136/jmedgenet-2018-105477.
- 583 24. Owen, D., Bracher-Smith, M., Kendall, K.M., Rees, E., Einon, M., Escott-Price, V., Owen, M.J., O'Donovan, M.C.,  
584 and Kirov, G. (2018). Effects of pathogenic CNVs on physical traits in participants of the UK Biobank. *BMC*  
585 *Genomics* 19, 1–9. 10.1186/s12864-018-5292-7.
- 586 25. Macé, A., Tuke, M.A., Deelen, P., Kristiansson, K., Mattsson, H., Nöukas, M., Sapkota, Y., Schick, U., Porcu, E.,  
587 Rüeger, S., et al. (2017). CNV-association meta-analysis in 191,161 European adults reveals new loci associated  
588 with anthropometric traits. *Nat Commun* 8, 744. 10.1038/s41467-017-00556-x.
- 589 26. Collins, R.L., Glessner, J.T., Porcu, E., Lepamets, M., Brandon, R., Lauricella, C., Han, L., Morley, T., Niestroj, L.M.,  
590 Ulirsch, J., et al. (2022). A cross-disorder dosage sensitivity map of the human genome. *Cell* 185, 3041-3055.e25.  
591 10.1016/j.cell.2022.06.036.
- 592 27. Hujoel, M.L.A., Sherman, M.A., Barton, A.R., Mukamel, R.E., Sankaran, V.G., Terao, C., and Loh, P.-R. (2022).  
593 Influences of rare copy-number variation on human complex traits. *Cell* 185, 4233–4248.  
594 10.1016/j.cell.2022.09.028.
- 595 28. Birnbaum, R., Mahjani, B., Loos, R.J.F., and Sharp, A.J. (2022). Clinical Characterization of Copy Number Variants  
596 Associated With Neurodevelopmental Disorders in a Large-scale Multiancestry Biobank. *JAMA Psychiatry* 79,  
597 250–259. 10.1001/JAMAPSYCHIATRY.2021.4080.
- 598 29. Martin, C.L., Wain, K.E., Oetjens, M.T., Tolwinski, K., Palen, E., Hare-Harris, A., Habegger, L., Maxwell, E.K., Reid,  
599 J.G., Walsh, L.K., et al. (2020). Identification of Neuropsychiatric Copy Number Variants in a Health Care System  
600 Population. *JAMA Psychiatry* 77, 1276–1285. 10.1001/jamapsychiatry.2020.2159.
- 601 30. Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman, G.R., Wainberg,  
602 M., Ollila, H.M., Kiiskinen, T., et al. (2021). Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat*  
603 *Genet* 53, 185–194. 10.1038/s41588-020-00757-z.
- 604 31. Männik, K., Mägi, R., Macé, A., Cole, B., Guyatt, A.L., Shihab, H.A., Maillard, A.M., Alavere, H., Kolk, A., Reigo, A.,  
605 et al. (2015). Copy number variations and cognitive phenotypes in unselected populations. *JAMA* 313, 2044–2054.  
606 10.1001/jama.2015.4845.
- 607 32. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J.,  
608 Langdon, R., et al. (2018). The MR-base platform supports systematic causal inference across the human  
609 phenome. *Elife* 7. 10.7554/ELIFE.34408.
- 610 33. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation  
611 PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 1–16. 10.1186/s13742-015-0047-8.

- 612 34. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O.,  
613 O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–  
614 209. 10.1038/s41586-018-0579-z.
- 615 35. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., and Bucan, M. (2007). PennCNV:  
616 An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome  
617 SNP genotyping data. *Genome Res* 17, 1665–1674. 10.1101/gr.6861907.
- 618 36. Macé, A., Tuke, M.A., Beckmann, J.S., Lin, L., Jacquemont, S., Weedon, M.N., Reymond, A., and Kutalik, Z.  
619 (2016). New quality measure for SNP array based CNV detection. *Bioinformatics* 32, 3298–3305.  
620 10.1093/bioinformatics/btw477.
- 621 37. Karczewski, K.J., Gupta, R., Kanai, M., Lu, W., Tsuo, K., Wang, Y., Walters, R.K., Turley, P., Callier, S., Baya, N.,  
622 et al. (2024). Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into  
623 ancestry-enriched effects. *medRxiv*. 10.1101/2024.03.13.24303864.
- 624 38. Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S.M., Sidorenko, J., Kweon, H., Goldman, G.,  
625 Gjorgjieva, T., et al. (2022). Polygenic prediction of educational attainment within and between families from  
626 genome-wide association analyses in 3 million individuals. *Nat Genet* 54, 437–449. 10.1038/s41588-022-01016-  
627 z.
- 628 39. Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R.B., Xu, C., Futema, M., Lawson,  
629 D., et al. (2015). The UK10K project identifies rare variants in health and disease. *Nature* 526, 82–90.  
630 10.1038/nature14962.
- 631 40. Moix, S., Sadler, M., Kutalik, Z., and Auwerx, C. (2024). Breaking down causes, consequences, and mediating  
632 effects of age-related telomere shortening on human health. *medRxiv*. 10.1101/2024.01.12.24301196.
- 633 41. Sadler, M.C., Auwerx, C., Lepik, K., Porcu, E., and Kutalik, Z. (2022). Quantifying the role of transcript levels in  
634 mediating DNA methylation effects on complex traits and diseases. *Nat Commun* 13, 1–14. 10.1038/s41467-022-  
635 35196-3.
- 636 42. Gale, C.R., Hagenaars, S.P., Davies, G., Hill, W.D., Liewald, D.C.M., Cullen, B., Penninx, B.W., Boomsma, D.I.,  
637 Pell, J., McIntosh, A.M., et al. (2016). Pleiotropy between neuroticism and physical and mental health: findings  
638 from 108 038 men and women in UK Biobank. *Transl Psychiatry* 6, e791–e791. 10.1038/tp.2016.56.
- 639 43. Kendall, K.M., Rees, E., Escott-Price, V., Einon, M., Thomas, R., Hewitt, J., O'Donovan, M.C., Owen, M.J., Walters,  
640 J.T.R., and Kirov, G. (2017). Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis  
641 of 152,000 UK Biobank Subjects. *Biol Psychiatry* 82, 103–110. 10.1016/j.biopsych.2016.08.014.
- 642 44. Tyrrell, J., Jones, S.E., Beaumont, R., Astley, C.M., Lovell, R., Yaghootkar, H., Tuke, M., Ruth, K.S., Freathy, R.M.,  
643 Hirschhorn, J.N., et al. (2016). Height, body mass index, and socioeconomic status: mendelian randomisation  
644 study in UK Biobank. *BMJ* 352, i582. 10.1136/bmj.i582.
- 645 45. Yu, Y., Zhu, H., Miller, D.T., Gusella, J.F., Platt, O.S., Wu, B.-L., and Shen, Y. (2011). Age- and gender-dependent  
646 obesity in individuals with 16p11.2 deletion. *J Genet Genomics* 389, 403–409. 10.1016/j.jgg.2011.08.003.
- 647 46. Gill, R., Chen, Q., D'Angelo, D., and Chung, W.K. (2014). Eating in the absence of hunger but not loss of control  
648 behaviors are associated with 16p11.2 deletions. *Obesity* 22, 2625–2631. 10.1002/OBY.20892.
- 649 47. Abawi, O., Wahab, R.J., Kleinendorst, L., Blankers, L.A., Brandsma, A.E., Van Rossum, E.F.C., Van Der Voorn, B.,  
650 Van Haelst, M.M., Gaillard, R., and Van Den Akker, E.L.T. (2023). Genetic Obesity Disorders: Body Mass Index  
651 Trajectories and Age of Onset of Obesity Compared with Children with Obesity from the General Population. *J*  
652 *Pediatr* 262, 113619. 10.1016/j.jpeds.2023.113619.
- 653 48. Verbitsky, M., Krithivasan, P., Batourina, E., Khan, A., Graham, S.E., Marasà, M., Kim, H., Lim, T.Y., Weng, P.L.,  
654 Sánchez-Rodríguez, E., et al. (2021). Copy Number Variant Analysis and Genome-wide Association Study Identify  
655 Loci with Large Effect for Vesicoureteral Reflux. *J Am Soc Nephrol* 32, 805–820. 10.1681/ASN.2020050681.
- 656 49. Verbitsky, M., Westland, R., Perez, A., Kiryluk, K., Liu, Q., Krithivasan, P., Mitrotti, A., Fasel, D.A., Batourina, E.,  
657 Sampson, M.G., et al. (2019). The copy number variation landscape of congenital anomalies of the kidney and  
658 urinary tract. *Nat Genet* 51, 117–127. 10.1038/s41588-018-0281-y.
- 659 50. Yang, N., Wu, N., Dong, S., Zhang, L., Zhao, Y., Chen, W., Du, R., Song, C., Ren, X., Liu, J., et al. (2020). Human  
660 and mouse studies establish TBX6 in Mendelian CAKUT and as a potential driver of kidney defects associated  
661 with the 16p11.2 microdeletion syndrome. *Kidney Int* 98, 1020–1030. 10.1016/j.kint.2020.04.045.
- 662 51. Shiow, L.R., Paris, K., Akana, M.C., Cyster, J.G., Sorensen, R.U., and Puck, J.M. (2009). Severe combined  
663 immunodeficiency (SCID) and attention deficit hyperactivity disorder (ADHD) associated with a coronin-1A  
664 mutation and a chromosome 16p11.2 deletion. *Clinical immunology : the official journal of the Clinical Immunology*  
665 *Society* 131, 24–30. 10.1016/j.clim.2008.11.002.

- 666 52. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D., Locke, A.E.,  
667 Balasubramanian, S., et al. (2021). Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*.  
668 10.1038/s41586-021-04103-z.
- 669 53. Karczewski, K.J., Solomonson, M., Chao, K.R., Goodrich, J.K., Tiao, G., Lu, W., Riley-Gillis, B.M., Tsai, E.A., Kim,  
670 H.I., Zheng, X., et al. (2022). Systematic single-variant and gene-based association testing of thousands of  
671 phenotypes in 394,841 UK Biobank exomes. *Cell genomics* 2, 100168. 10.1016/J.XGEN.2022.100168.
- 672 54. Porcu, E., Rüeger, S., Lepik, K., Agbessi, M., Ahsan, H., Alves, I., Andiappan, A., Arindrarto, W., Awadalla, P.,  
673 Battle, A., et al. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants  
674 of complex and clinical traits. *Nat Commun* 10, 1–12. 10.1038/s41467-019-10936-0.
- 675 55. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014).  
676 Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS*  
677 *Genet* 10, e1004383. 10.1371/JOURNAL.PGEN.1004383.
- 678 56. Morrison, J., Knoblach, N., Marcus, J.H., Stephens, M., and He, X. (2020). Mendelian randomization accounting  
679 for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet* 52, 740–747.  
680 10.1038/s41588-020-0631-4.
- 681 57. Darrous, L., Mounier, N., and Kutalik, Z. (2021). Simultaneous estimation of bi-directional causal effects and  
682 heritable confounding from GWAS summary statistics. *Nat Commun* 12, 1–15. 10.1038/s41467-021-26970-w.