

Neurodevelopmental copy number variants increase risk of internalising and cardiometabolic multimorbidity: findings from UK Biobank

Ioanna K Katzourou¹, LINC consortium, Inês Barroso², Julie Clayton³, Golam Khandaker⁴, Daniel Stow⁵, Nicolas Timpson⁴, Ruby Tsang⁴, Jack Underwood¹, Megan Wood⁶, George Kirov¹, James Walters¹, Michael J Owen^{1,7}, Peter Holmans¹, Marianne B M van den Bree^{1,7}

¹Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK

²Medical School, University of Exeter, UK

³Centre for Academic Primary Care, University of Bristol, UK

⁴Bristol Medical School, University of Bristol, UK

⁵Wolfson Institute for Population Health, Queen Mary University of London, UK

⁶School of Psychology, University of Leeds, UK

⁷Neuroscience and Mental Health Innovation Institute, Cardiff University, UK

Abstract

Background

Internalising and cardiometabolic multimorbidity (ICM-MM) represents a major clinical challenge, negatively impacting life expectancy and quality of life and resulting in considerable healthcare costs. Individuals with a copy number variant associated with increased risk of neurodevelopmental conditions (ND-CNV) are more likely to develop mental or physical ill health, however the effects on ICM-MM remain poorly understood.

Methods

We used data from UK Biobank (ND-CNV N= 7,549, 1.62%). ICM-MM was identified using electronic healthcare records. It was defined as combinations of any internalising condition (either depression, anxiety or somatoform disorder) with each of five cardiometabolic conditions (hypertension, dyslipidemia, obesity, type 2 diabetes and chronic kidney disease). We also studied whether ICM-MM risk in those with ND-CNV differed by sex, presence of a deletion versus a duplication and we established associations of dosage-sensitive genes within

ND-CNVs with ICM-MM. Finally, we explored the interaction between presence of ND-CNV and polygenic risk scores (PRSs) of internalising and cardiometabolic traits on ICM-MM risk.

Results

The presence of ND-CNV was associated with ICM-MM (OR range 1.21-1.57). Female participants with ND-CNV were more likely to have any internalising condition and T2D, and those with a deletion any internalising condition and obesity. The number of deleted haploinsufficient genes, but not duplicated triplosensitive genes, was associated with ICM-MM. No significant interactions between ND-CNVs and PRSs were found.

Conclusions

This is the first study to report that ND-CNVs increase the likelihood of ICM-MM, and we find evidence of sex differences and stronger effects for deletions. Increased clinical awareness can help ameliorate this risk.

Background

Multimorbidity, also referred to as multiple long-term conditions (MLTC), indicates the presence of two or more chronic health conditions in the same individual. Multimorbidity represents a major public health concern, with at least 50 million people affected in the European Union alone(1). An estimated 25% of the population of high-income countries is living with two or more conditions, and rates are rapidly increasing in low- and middle-income countries(1,2). Multimorbidity is associated with high personal as well as societal healthcare costs(2). Moreover, multimorbidity is difficult to manage, resulting in high risk of failure of care and putting considerable strain on healthcare systems(1–3).

Cardiometabolic conditions, such as hypertension, obesity and type 2 diabetes (T2D), and internalising conditions, such as depression and anxiety, are highly prevalent, affecting millions of people worldwide and resulting in increased disability and functional decline, poor quality of life and premature mortality(4,5). The co-occurrence of internalising and cardiometabolic conditions represents the most common type of physical and mental health multimorbidity in older adults(3). While both cardiometabolic and internalising conditions pose considerable public health challenges, their presence in combination is particularly burdensome for the individuals affected (4,5). Internalising disorders are associated with an increased risk of subsequent cardiometabolic disorders(6–8) as well as cardiovascular morbidity and premature mortality(9,10). The reverse is also the case, with those with cardiometabolic conditions experiencing more depression and anxiety(7,10–12). Importantly, the lower life expectancy associated with depression(13) is partially attributable to comorbid cardiovascular disease(14). These findings highlight the importance of uncovering the risk factors contributing to internalising and cardiometabolic multimorbidity (ICM-MM).

While the exact mechanisms leading to the development of ICM-MM are unknown, shared genetic risk factors may play a considerable role(15,16). Polygenic risk scores (PRS) of cardiometabolic disorders have been found to be associated with depression(17,18), while high genetic overlap has also been found between depression and T2D(19). Most published studies focus on common genetic variation, while the effect of rare genetic variants on

multimorbidity remains mostly unexplored. Copy number variants (CNVs) are structural alterations in chromosomes involving the deletion or duplication of a section of varying length(20). A range of CNVs increase the risk of neurodevelopmental conditions (NDCs), such as intellectual disability, autism spectrum disorder, attention deficit hyperactivity disorder and schizophrenia(21–24). These are referred to as ND-CNVs(25). In addition to NDCs, these CNVs can also increase the risk of internalising conditions(26) as well as a range of physical health conditions and traits(27), including hypertension, diabetes and body mass index (BMI). There is increasing evidence of increased risk of cardiovascular morbidity in individuals with some of these variants(28–30). As ND-CNVs are often diagnosed in childhood, increased understanding of the development of multimorbidity in individuals with ND-CNVs presents opportunities for better understanding the development of disease development and potentially also provide important insights into the biological mechanisms of multimorbidity.

In this study, we explored the association between ICM-MM and ND-CNVs(21) in UK Biobank (UKBB)(31).

The objectives of this study were to:

- Determine the association of ND-CNVs, loss (deletion) and gain (duplication) of chromosomal material and dosage-sensitive genes in ND-CNVs with ICM-MM;
- Investigate sex differences in the effect of ND-CNVs on ICM-MM;
- Establish if the effect of common variation (PRSs of internalising and cardiometabolic traits) on ICM-MM differs in individuals with versus without a ND-CNV.

Methods

Data source

UKBB is a prospective cohort of over 500,000 individuals living in the United Kingdom(31). UKBB received ethical approval from the North West - Haydock Research Ethics Committee (reference 16/NW/0274). Participants provided electronic signed consent at recruitment. This study was conducted under application number 79704.

Participants

Participants aged between 40 and 69 years old were recruited into UKBB between 2006 and 2010(31). Sociodemographic, lifestyle and medical history information were collected using touchscreen questionnaires. Physical and functional measurements, biochemical assays and genome-wide genotyping were collected at a baseline assessment. Linkage to the National Health Service provided data on deaths, cancer diagnoses, hospital inpatient/outpatient episodes and primary care records. Details of the UKBB study design are provided elsewhere(31). Out of the total number of UKBB participants (502,411), 459,483 (91.47%) had available CNV call data and were included in this study.

Phenotyping

This study took place under the Lifespan Multimorbidity Research collaborative (LINC) (<https://www.cardiff.ac.uk/lifespan-multimorbidity-research-collaborative>). LINC seeks to understand the development of internalising and cardiometabolic multimorbidity (ICM-MM) over the life course. The conditions included in LINC's definition of ICM-MM were selected following discussions with primary and secondary care doctors, the LINC patient and public involvement team, as well as the LINC team of researchers and clinicians.

Three internalising conditions were included: depression, anxiety and somatoform disorder. Five cardiometabolic conditions were included: hypertension, dyslipidemia, obesity, T2D and chronic kidney disease (CKD). The cardiometabolic conditions included in the definition are restricted to those that generally occur earlier in life than established cardiovascular disease, allowing the opportunity to study the increase in risk over an extended period.

Individuals with each of the conditions of interest were identified through linked electronic healthcare records (EHR; primary care records and hospital episode statistics (HES)) using established clinical codelists (see Supplementary Material). We examined the pairwise combinations of aggregated internalising conditions, defined as the presence of one or more of

the three internalising conditions listed above (referred to as ‘any internalising condition’ from here onwards), and each of the five cardiometabolic conditions. We also constructed an outcome variable - any ICM-MM - defined as the presence of one or more of the three internalising conditions and one or more of the five cardiometabolic conditions.

CNV calling

The process of the calling of the NDD-CNVs in UKBB is described in detail elsewhere(32). Briefly, calling was performed using PennCNV-Affy 1.0.3 protocols(33) on the UKBB genotype array data. Samples were excluded if they carried 30 or more CNVs, had a waviness factor greater than 0.03 or less than -0.03; a single-nucleotide polymorphism call rate lower than 96%; or log R ratio SD higher than 0.35. CNVs were excluded if they were covered by fewer than 20 probes; had a density coverage of less than 1 probe per 20,000 base pairs; or a confidence score lower than 10. This resulted in 459,483 individuals with available CNV call data.

Statistical analyses

All statistical analyses were performed in R(34).

Aggregated CNVs

We focused on a set of 54 CNVs that show strong evidence of increasing the risk of developing a neurodevelopmental disorder (ND-CNVs)(21). We assessed the association of the presence of any ND-CNV (aggregated ND-CNV; binary variable) with each of the conditions of interest individually as well as pairwise combinations of any internalising conditions and each cardiometabolic condition and any ICM-MM (binary variables). We used logistic regression of each of these phenotypes on aggregated ND-CNV, adjusting for age at baseline, sex and Townsend deprivation index (as a measure of socioeconomic status) and the first five genetic principal components to account for population stratification.

Sensitivity analyses

While HES data are available for the whole UKBB cohort (N = 502,390), primary care records are available for ~40% of the participants (N = 229,951). A sensitivity analysis was performed including only participants for whom both HES and primary care records were available (N= 229,951).

Deletions in the 16p11.2 region have been previously associated with class III obesity (formerly known as morbid obesity)(35–37). Obesity is one of the five cardiometabolic conditions we examined, as well as a known risk factor for the remaining four. In order to determine if the overall effect of ND-CNV on ICM-MM was driven by these specific ND-CNVs, we repeated the regression analysis after removing individuals with a proximal or distal deletion of 16p11.2 (N= 185). We also repeated the regressions adjusting for BMI at baseline (excluding obesity as an outcome) to account for possible effects of the ND-CNVs on body mass.

Post-hoc analyses

To assess if there are any sex differences in the associations described above, we also conducted the above regression including an interaction term between ND-CNV and sex. Moreover, we also assessed the association of the phenotypes with the type of NDD-CNV, e.g. loss (deletion) or gain (duplication). For this analysis, genotype was coded as a three-level variable (no ND-CNV, deletion or duplication). Logistic regressions of each of the phenotypes on genotype were performed, adjusting for age, sex Townsend deprivation index and the first five genetic principal components. Individuals with both a deletion and a duplication were excluded from this analysis (N= 12).

We next examined if dosage-sensitive genes within ND-CNVs are associated with ICM-MM. The coordinates for each ND-CNV were used to map which genes were affected in each individual. Out of the genes affected by a ND-CNV in each individual, we identified those that are haploinsufficient (deletion intolerant) or triplosensitive (duplication intolerant), based on a dosage sensitivity analysis conducted by Collins *et al.* (38) and calculated the number of dosage sensitive genes affected in each individual. For this analysis, deletions and duplications were

analysed separately. Thus, we regressed the phenotypes of interest (e.g. pairwise combinations of any internalising conditions with each cardiometabolic condition and any ICM-MM) on the number of haploinsufficient genes included in each deletion or the number of triplosensitive genes included in each duplication, adjusting for age, sex, Townsend deprivation index, the total number of genes included in each ND-CNV as a proxy of overall biological effect of each ND-CNV, and the first five genetic principal components to account for population stratification.

Interaction between CNVs and common genetic variation

To assess whether the association of common genetic variation and the risk of multimorbidity differs in individuals with and without a ND-CNV, the evidence for interaction between the presence of a ND-CNV and PRSs of our internalising and cardiometabolic conditions of interest was assessed. PRSs for major depressive disorder (MDD)(39), anxiety(40), low density cholesterol (LDL; as a proxy for dyslipidemia)(41), body mass index (BMI; as a proxy of obesity)(42), systolic blood pressure (SBP; as a proxy for hypertension)(43), T2D(44) and CKD(45) were computed. The genome-wide association studies (GWAS) used for the generation of these PRSs were selected because they were the latest and largest studies with publicly available summary statistics that excluded UKBB. They are described in Supplementary Table 1.

PRS-CS(46) was used for PRS calculation. PRS-CS is a Bayesian algorithm that can infer posterior effect sizes of SNPs via continuous shrinkage(46), therefore avoiding the need for linkage disequilibrium pruning and p-value thresholding. The inferred posterior effect sizes were used for PRS generation on PLINK 2.0(47). In order to produce PRSs that are on the same scale across individuals from different ancestries, we adjusted them for ancestral differences in mean and variance using the 1000Genomes dataset as reference, as described by Khan et al.(48). The ancestry adjustment is described in detail elsewhere(49).

Logistic regression analyses were performed as described previously, including the main effects of each ancestry-adjusted PRS, ND-CNV and the interaction term PRS*ND-CNV.

Individual CNVs

We assessed the association of the combination of aggregated internalising conditions with each cardiometabolic condition and also any ICM-MM with each of the ND-CNVs individually. The number of individuals with each of the ND-CNVs are shown in Supplementary Table 2. To ensure the analysis was statistically viable, CNVs that were observed fewer than 5 times in the total sample were excluded, resulting in 33 CNVs included in this analysis. The associations were assessed using logistic regression, adjusting for age at baseline, sex, Townsend deprivation index and the first five genetic principal components.

Results

Associations of aggregated ND-CNVs with ICM-MM

There were 7,546 individuals (1.64%) with a ND-CNV in UKBB. The demographic characteristics of the individuals with and without a ND-CNV are shown in Supplementary Table 3. Individuals with a higher Townsend deprivation index were more likely to have a ND-CNV (OR = 1.55, p-value = 1.19×10^{-34}), meaning individuals with a ND-CNV were more likely to live in a more deprived environment than those without. The number of individuals with and without a ND-CNV that have a diagnosis of the eight conditions of interest is shown in Supplementary Table 4.

14.2% of individuals with a ND-CNV had ICM-MM, compared to 11.5% of individuals without a ND-CNV. The number of individuals with and without a ND-CNV that have ICM-MM is shown in Table 1.

	ND-CNV (N=7,546) present		ND-CNV (N=451,937) absent	
	Cases	Frequency (%)	Cases	Frequency (%)
<i>Any internalising and T2D</i>	291	3.9	10,717	2.4
<i>Any internalising and obesity</i>	397	5.3	15,580	3.5
<i>Any internalising and hypertension</i>	764	10.1	37,688	8.3
<i>Any internalising and dyslipidemia</i>	564	7.5	26,578	5.9
<i>Any internalising and CKD</i>	193	2.6	8,085	1.8
<i>Any ICM-MM</i>	1,073	14.2	51,919	11.5

Table 1. Counts and frequency for each of the ICM-MM phenotypes of interest in individuals with and without a ND-CNV.

All combinations of ICM-MM were significantly associated with the presence of a ND-CNV after Bonferroni correction for multiple testing (Bonferroni p-value threshold = 8.33×10^{-3}). The OR for any ICM-MM was similar to that for any internalising condition or any cardiometabolic condition, however the ORs for certain ICM-MM combinations (any internalising conditions and

obesity, any internalising condition and T2D, any internalising condition and CKD) were increased over any internalising condition or any cardiometabolic condition separately. The associations between the presence of a ND-CNV and the different combinations of aggregated internalising conditions with each cardiometabolic condition and ICM-MM are shown in Figure 1 and Supplementary Table 5.

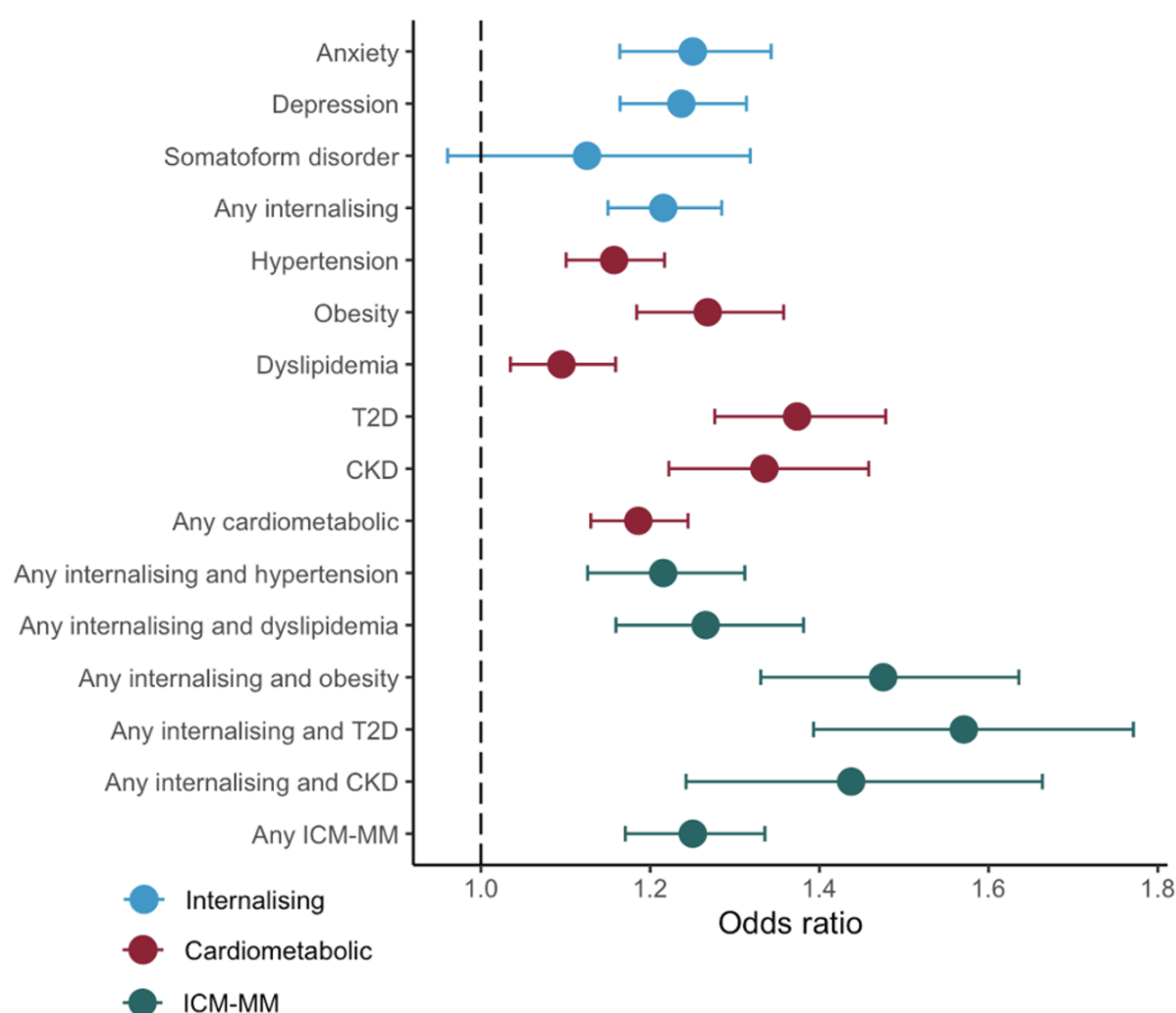


Figure 1. Association of ND-CNVs with individual conditions and multimorbidity.

The associations remained significant when adjusting for BMI (Supplementary Table 6). Restricting the analysis to individuals with both primary care and HES data (N= 229,951) did not substantively alter our findings (Supplementary Figure 1 and Supplementary Table 7), while

the ORs were higher and p-values lower when including the full UKBB sample. Excluding individuals with proximal and distal deletions in 16p11.2 (N = 185), a known risk factor for early onset morbid obesity, also did not substantively alter the findings (Supplementary Figure 2 and Supplementary Table 8).

We examined whether the associations between having any ND-CNV and the phenotypic presentations of interest differed by sex. There was evidence of interaction between aggregated ND-CNV and sex in association with any cardiometabolic condition and the combination of any internalising condition and T2D (Figure 2 and Supplementary Table 9). Supplementary Figure 3 shows that the increased risk of T2D or hypertension associated with a ND-CNV is higher in female than in male participants.

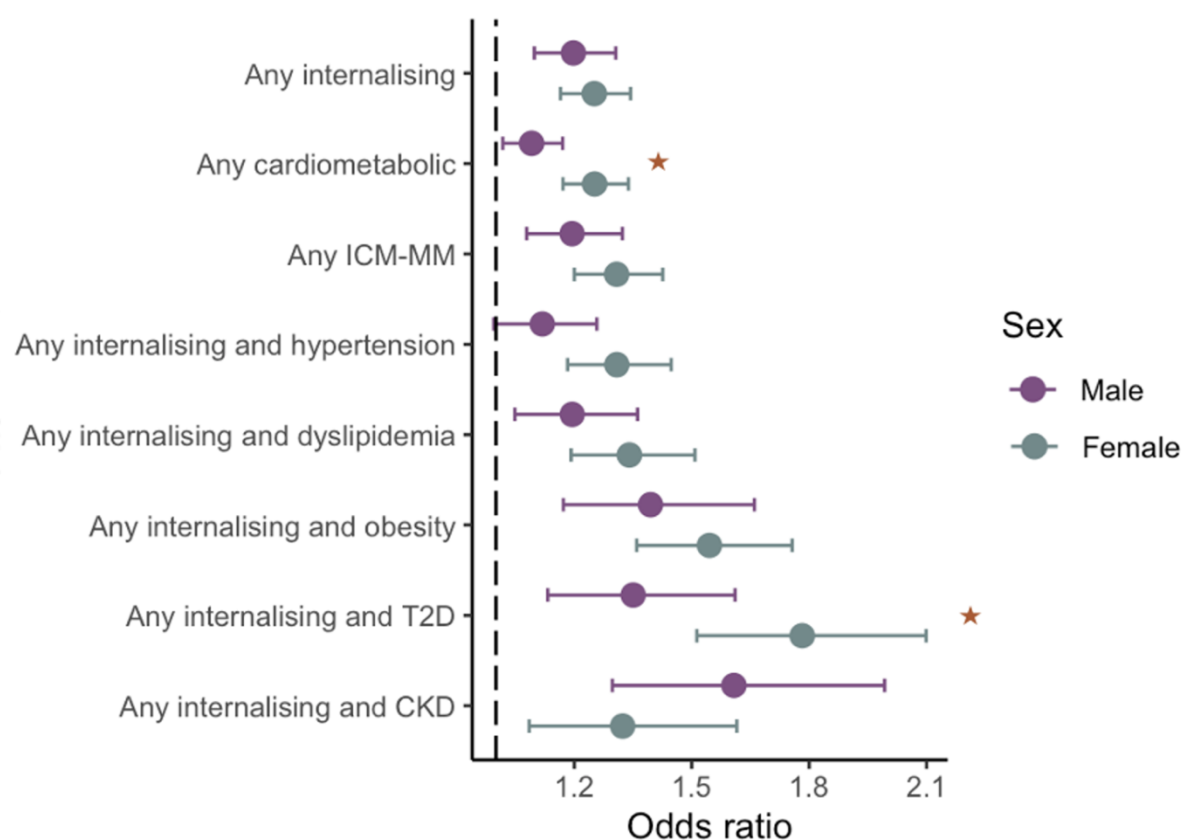


Figure 2. Association of ND-CNVs with multimorbidity for male and female sex. Stars indicates a significant interaction between the presence of ND-CNV and sex (p-value < 0.05).

We examined whether the associations between the outcomes of interest differed for individuals with loss (deletion) or gain (duplication) of chromosomal material. Having a deletion increased the risk of having internalising condition and obesity over having a duplication (Figure 3 and Supplementary Table 10). When considering individual conditions, having a deletion was associated with greater the odds of having obesity and T2D over having a duplication while having a duplication significantly was associated with greater odds of having dyslipidemia over having a duplication, as seen in Supplementary Figure 4.

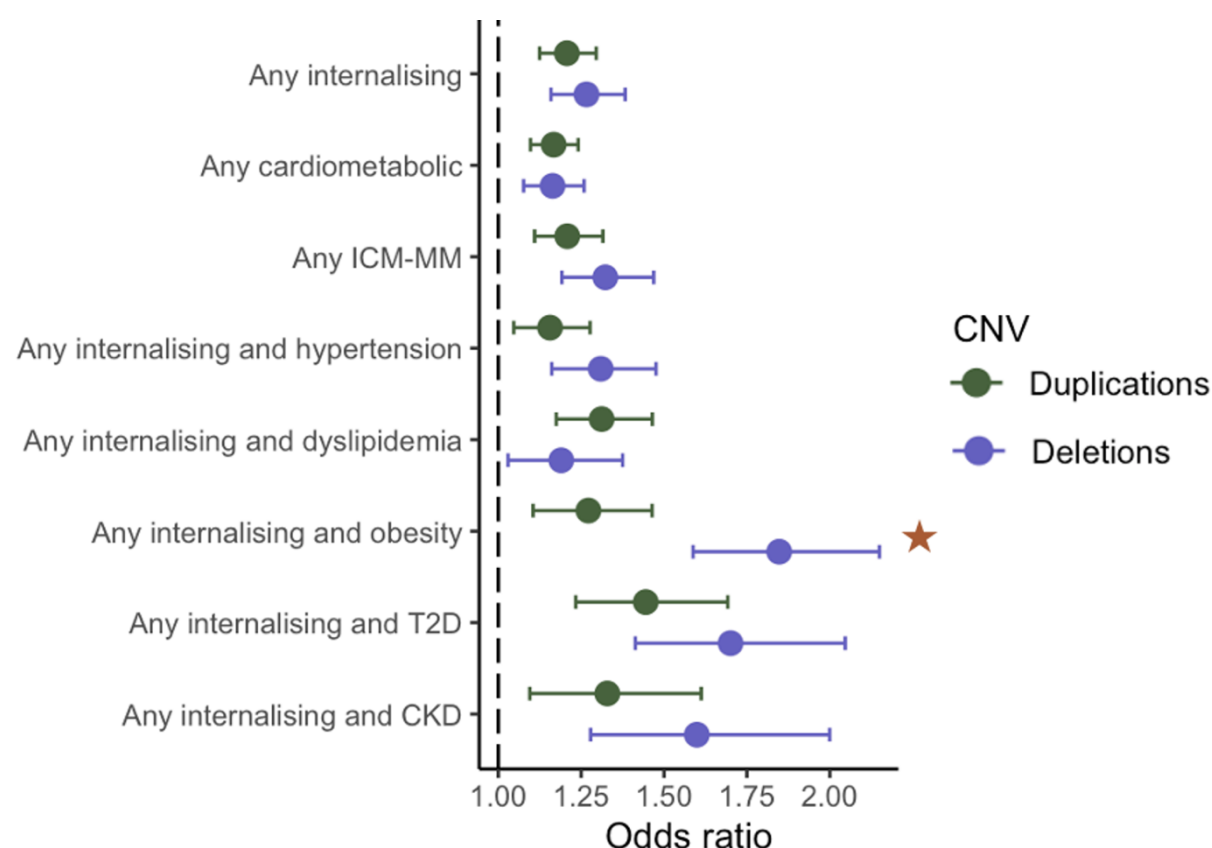


Figure 3. Association of duplications and deletions with multimorbidity. Stars indicate a suggested difference between deletions and duplications (p -value < 0.05).

The number of haploinsufficient genes within a deletion was significantly associated with all outcomes. The number of triplosensitive genes within a duplication did not show the same pattern (Supplementary Figure 5). The total number of genes affected by a deletion or a duplication was significantly associated with all outcomes (Supplementary Table 11).

Interaction between aggregated ND-CNVs and common genetic variation

The evidence for association between ND-CNV and combinations of ICM-MM remained significant when adjusting for each PRS (and vice versa). However, interactions between PRS and aggregated ND-CNV were not significant after Bonferroni correction for multiple testing, with three interactions reaching nominal significance ($p\text{-value} < 0.05$; any internalising condition and obesity and any internalising condition and CKD for anxiety PRS, and any internalising condition and CKD for LDL PRS), as seen in Supplementary Table 12. PRSs of MDD ($OR = 1.001$, $p\text{-value} = 1.34 \times 10^{-4}$), anxiety ($OR = 1.001$, $p\text{-value} = 2.40 \times 10^{-6}$) and CKD ($OR = 1.001$, $p\text{-value} = 2.54 \times 10^{-3}$) were associated with the presence of a ND-CNV.

Associations of individual ND-CNVs with internalising and cardiometabolic multimorbidity

For each of the 33 ND-CNVs that were present more than 5 times (Supplementary Table 1) in individuals in UKBB, we investigated their association with the pairwise combinations of any internalising condition with each cardiometabolic condition, as well as any ICM-MM (Supplementary Figure 6 and Supplementary Table 13). After Bonferroni correction for multiple testing, 9 of the 33 ND-CNVs showed evidence for association with ICM-MM ($p\text{-value} < 2.52 \times 10^{-4}$). Several deletions on the p arm of chromosome 16 were associated with multiple ICM-MM outcomes, while 22q11.2 duplication, 15q24 duplication and 15q13.3 deletion were associated with any internalising condition and obesity. All significant associations found indicated an increased risk of ICM-MM.

Discussion

The impact of ND-CNVs on risk of multimorbidity between mental and physical health conditions has received little research attention to date. The aim of this study was to investigate the association of ND-CNVs with the presence of ICM-MM in a population cohort. Our findings indicate that individuals with ND-CNVs are more likely to experience ICM-MM, with particularly

increased odds for specific combinations of internalising and cardiometabolic conditions such as any internalising condition and T2D, any internalising condition and CKD and any internalising condition and obesity. We also found evidence of differential associations with ICM-MM in individuals with ND-CNVs depending on sex and whether they carry a deletion or duplication, and that genes intolerant to deletion (haploinsufficient genes) within ND-CNVs were more likely to contribute to ICM-MM risk than genes intolerant to duplication (triplosensitive genes). Finally, we found no evidence of interactions between ND-CNV status and PRSs of the internalising and cardiometabolic conditions under study, indicating that common genetic variation does not modify risk of ICM-MM in individuals with a ND-CNV.

We found that having any ND-CNV was associated with 25% increased likelihood of ICM-MM, with the frequency of any ICM-MM in individuals with a ND-CNV being 14.2%, compared to 11.5% in individuals without a ND-CNV. Given the way UKBB participants were ascertained(50), this is likely to be a lower bound estimate, as both the outcomes of interest and ND-CNVs are associated with premature mortality and the older age of participants could introduce a survival bias. While ND-CNVs have mostly been studied in relation to NDCs(51–53) and other psychiatric conditions(25,26,54), they have also recently been linked to multiple physical health outcomes(55,56), including obesity(35,57), renal disease(58) and diabetes(59). Whilst much of the literature linking ND-CNVs with physical health outcomes has focused on clinical populations and rare and severe syndromic phenotypes, we studied a population cohort that is not enriched for more severe manifestations of ND-CNVs. Whilst unknown, it is also likely that many of the individuals with ND-CNVs in UKBB are unaware of their genotype. We explored common, earlier onset cardiometabolic conditions that are on the causal pathway to cardiovascular morbidity and mortality, and we found that ND-CNVs were significantly associated with all the cardiometabolic conditions we examined, as well as ICM-MM. Our results suggest that ND-CNVs are associated with common conditions and ICM-MM even in the general population, while it is important to keep in mind that these outcomes are likely to be more severe in clinically ascertained carriers. Associations between ND-CNVs and ICM-MM remained significant when removing ND-CNVs that are known risk factors for morbid obesity or when correcting for BMI, suggesting that in this dataset the associations are unlikely to be driven by the effect of ND-CNVs on body mass.

In addition to our composite phenotype of any ICM-MM, we also assessed the impact of ND-CNVs on pairwise combinations of any internalising condition and each of the five cardiometabolic conditions. Our rationale was that while the internalising disorders are highly phenotypically and genetically correlated(49,60,61), the cardiometabolic disorders are more heterogeneous, therefore the associations with ND-CNVs may differ between the pairwise ICM-MM combinations. Our findings confirm this assumption. While the odds for any ICM-MM conferred by ND-CNVs was quite similar to those for any internalising condition or any cardiometabolic condition alone, the odds for some of the combinations (e.g. any internalising condition and obesity, any internalising condition and T2D, any internalising condition and CKD) was higher than those for any internalising condition, any cardiometabolic condition or any ICM-MM. This indicates that individuals with ND-CNVs have higher odds of developing certain combinations of internalising and cardiometabolic conditions than others and could benefit from closer monitoring for early indications of particular conditions.

Interestingly, we observed that the effect of the presence of a ND-CNV on the odds of any cardiometabolic condition as well as any internalising condition and T2D appeared to be higher in female than in male participants. While there was no strong evidence for the remaining outcomes, there was a pattern for larger effects for female than for male participants for all ICM-MM combinations apart from any internalising condition and CKD. Previous studies on sex differences in the effect of ND-CNVs on NDCs and other psychiatric disorders have produced conflicting results(26,62–65), with some studies reporting higher rates of depression in female adults with ND-CNVs(26) and of anxiety and depression in female children with with a diagnosed NDC and an ND-CNV(65), whilst others found no evidence of sex differences in ND-CNV burden for anxiety and depression(62). While further studies are required to understand the underlying mechanisms behind possible sex differences, our findings highlight the importance of considering sex as a factor when designing tailored health monitoring strategies for individuals with ND-CNVs.

Moreover, we found suggestive evidence of differences between the effect of deletions and duplications on ICM-MM, with deletions being associated with higher odds of having any internalising condition and obesity over duplications. Deletions in the 16p11.2 region are known risk factors for class III obesity(35–37), and it is likely that these regions contribute to this

finding, however the association between ND-CNVs and any internalising condition and obesity remained significant when deletions in this region were removed, suggesting that other ND-CNVs also predispose to this type of multimorbidity. In order to further explore the way in which ND-CNVs are involved in the pathogenesis of multimorbidity, we examined the association of dosage-sensitive genes within a ND-CNV with ICM-MM. Haploinsufficient genes (those intolerant to deletion) within ND-CNVs were associated with higher odds of most combinations of ICM-MM, whereas triplosensitive genes (those intolerant to duplication) did not show evidence of strong effect. The total number of genes within both deletions and duplications was significantly associated with most combinations of ICM-MM, however the effect size was small. Our findings suggest that, for the phenotypes we studied, in individuals with deletions the loss of haploinsufficient genes is pathogenic, however the same might not be the case for duplications, where the total number of genes hit or other parameters, like the total length or the penetrance of the variant, might be more important than the number of triplosensitive genes duplicated. Importantly, for individuals with deletions, the number of haploinsufficient genes can be used to identify individuals that might be at risk of developing a more severe phenotype.

We explored the possibility of interaction between the presence of a ND-CNV and PRSs of our internalising and cardiometabolic traits of interest on risk of ICM-MM. The absence of detectable interaction between ND-CNVs and the PRSs suggest that effect of ND-CNVs on the risk of ICM-MM is not modified by the effect of common genetic variation. This is the first study to explore the interaction of common variation with ND-CNVs on the risk of multimorbidity. Although UKBB is one of the largest cohorts internationally available for the study of our aims, ND-CNVs are rare (they were present in 1.6% of participants) and it is likely that we lacked sufficient statistical power to detect possible interactions, however our results are in agreement with previous studies exploring the effect of CNVs or other rare variants and common variation on depression phenotypes in cohorts of similar or smaller size (66–68).

When examining the ND-CNVs individually, we found nine with evidence for association with at least one type of ICM-MM. Four of these were deletions on the p arm of chromosome 16 (16p13.11 deletion, 16p11.2 deletion, 16p12.1 deletion and 16p12.1 distal deletion). 16p11.2 deletion, 16p12.1 deletion and 16p13.11 deletion were associated with all combinations of ICM-MM, while 16p12.1 distal deletion was associated with any internalising condition and

T2D and any internalising condition and obesity. These results are consistent with previous research finding that deletions in this region are associated with psychiatric disorders and traits(25,51), obesity(35–37), T2D(36) and kidney disease biomarkers(36). 22q11.2 duplication was associated with any internalising condition and obesity. This variant is considered to be associated with a variable and mild phenotype compared to deletions in the same region(69). Our results highlight the considerable heterogeneity between ND-CNVs. Where increasing sizes of population-based cohorts allow the opportunity, each CNV should be studied in isolation to better understand its specific contribution to disease risk and provide valuable insight into the biological underpinning of ICM-MM.

Our findings give evidence of association between ND-CNVs and the co-occurrence of common internalising and cardiometabolic conditions; a multimorbid presentation that can lead to substantive functional impairment, a reduced quality of life and premature mortality. This finding has several clinical implications. Firstly, it underscores the need for healthcare systems to adopt a more holistic approach to multimorbidity, particularly in relation to individuals with rare genetic variants. Current healthcare service provision tends to look at each diagnosis in isolation, but it is evident from our findings that this approach may not be suited for individuals with ND-CNVs, who are at risk from multiple conditions throughout their lifetime, many of which need to be managed by different medical specialties. It is imperative that healthcare systems transition to a more comprehensive multidisciplinary model of care, particularly for individuals with known pathogenic genetic variants. Secondly, our results suggest that the impacts of ND-CNVs extend beyond the phenotypes they are traditionally associated with (e.g. congenital abnormalities, intellectual disability, autism, schizophrenia) to include common morbidities and multimorbidity seen in older age people in the general population. Recent breakthroughs in sequencing technology allow for identification of an increasing range of rare genetic variants by medical genomics services, and children with NDCs will often be screened for ND-CNVs. ND-CNVs are often inherited and reduced penetrance may mean that seemingly unaffected relatives might carry an ND-CNV that predisposes them to ICM-MM, and could also stand to benefit from screening.

There are some limitations to our study. UKBB is one of the largest publicly available population cohorts, and was designed as a prospective study of middle and older age(31), with median age at recruitment being 57 years. It is therefore well-suited to this study as the age range of the participants allowed us to study conditions that tend to appear later in life. However, it is also susceptible to survival bias. As multimorbidity and CNV burden are associated with premature mortality(56,70), it is likely that individuals with the most severe outcomes are underrepresented, which would lead to effective reduction multimorbidity frequency and effect sizes compared to the general population. Moreover, the participants of UKBB have been found to have higher socioeconomic status, report higher overall health and be less ethnically diverse than the average UK population(71), which could also lead to reduced population-based effect sizes. Furthermore, the generalisability of our findings remains uncertain, particularly given the predominance of individuals of European descent in the UKBB. When assessing the interaction between ND-CNVs and PRSs, we found that some of the PRSs were associated with ND-CNV status (MDD, anxiety and CKD PRS). These associations could introduce a collider bias in the analysis, whereby adjusting for PRS would alter the ND-CNV association with the outcomes in a manner that resembles an interaction. However, we did not find evidence of interaction between ND-CNVs and any of the PRSs in our analysis. Finally, the variants we examined tend to be low in frequency, some appearing less than 100 instances in half a million of UKBB participants. We aggregated ND-CNVs to have increased statistical power, which facilitated discovering associations that would otherwise be missed. This approach provides valuable insights into the common effects of the variants but does not allow clarification of ND-CNV-specific relationships.

In conclusion, this study demonstrates that, on a population level, individuals with ND-CNVs have higher odds of developing common and preventable forms of ICM-MM. As the field of medical genetics continues to expand, understanding how rare genetic variants contribute to multimorbidity will be essential for improving the health outcomes of affected individuals.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 79704.

This work was funded by the Tackling Multimorbidity at Scale Strategic Priorities Fund programme (MR/W014416/1) delivered by the Medical Research Council and the National Institute for Health Research in partnership with the Economic and Social Research Council and in collaboration with the Engineering and Physical Sciences Research Council. IKK is supported by this grant.

Full list of LINC members:

Marianne B. M. van den Bree, George Kirov, Michael J. Owen, James T. R. Walters, Peter A. Holmans, Jane Lynch, Ioanna K. Katzourou, Jack F. J. Underwood (Cardiff University, UK)
 David A. van Heel, Sarah Finer, Daniel Stow (Queen Mary University of London, UK)
 Golam M. Khandaker, Nicholas J. Timpson, John A. A. MacLeod, Julie P. Clayton, Ruby S. M. Tsang, Jane Sprackman, Shahid Khan (University of Bristol, UK)
 Inês Barroso, Rupert A. Payne (University of Exeter, UK)
 Mark Mon-Williams, Megan L. Wood, Nabila Ali (University of Leeds, UK)
 Hilary C. Martin (Wellcome Sanger Institute, UK)
 Thomas Werge, Andrés Ingason, Morteza Vaez, Lam O. Huang (Institute of Biological Psychiatry, Denmark)

We thank the members of the LINC study public advisory group for their contribution.

Funding

This work was funded by the Tackling Multimorbidity at Scale Strategic Priorities Fund programme (MR/W014416/1) delivered by the Medical Research Council and the National Institute for Health Research in partnership with the Economic and Social Research Council and in collaboration with the Engineering and Physical Sciences Research Council.

Author contribution

Study conceptualisation and design: IKK, MvdB, PH, GK, MJO, JW, LINC. Analytical consultation and interpretation: IKK, MvdB, PH, GK, MJO, JW, AI, RT, DS, IB. UKBB data curation: IKK. Genetic data preparation: IKK. Supervision: MvdB, PH, GK, MJO, PH, JW. All listed authors critically edited the manuscript.

Data availability

Relevant data is available from UK Biobank subject to standard procedures (www.ukbiobank.ac.uk).

Declaration of interest

Michael J Owen is receiving research grants from Takeda Pharmaceuticals and Akvivia Health. The remaining authors declare no competing interests.

References

1. The Academy of Medical Sciences. Multimorbidity: A Priority for Global Health Research. [Internet]. 2018. Available from: <https://acmedsci.ac.uk/file-download/82222577>
2. World Health Organization. Multimorbidity [Internet]. Geneva: World Health Organization; 2016 [cited 2025 Jan 16]. 28 p. (Technical Series on Safer Primary Care). Available from: <https://iris.who.int/handle/10665/252275>
3. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet Lond Engl*. 2012 Jul 7;380(9836):37–43.
4. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet Lond Engl*. 2013 Nov 9;382(9904):1575–86.
5. GBD 2013 DALYs and HALE Collaborators, Murray CJL, Barber RM, Foreman KJ, Abbasoglu Ozgoren A, Abd-Allah F, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. *Lancet Lond Engl*. 2015 Nov 28;386(10009):2145–91.
6. Mezuk B, Eaton WW, Albrecht S, Golden SH. Depression and type 2 diabetes over the lifespan: a meta-analysis. *Diabetes Care*. 2008 Dec;31(12):2383–90.
7. Cohen BE, Edmondson D, Kronish IM. State of the Art Review: Depression, Stress, Anxiety, and Cardiovascular Disease. *Am J Hypertens*. 2015 Nov;28(11):1295–302.
8. Yu M, Zhang X, Lu F, Fang L. Depression and Risk for Diabetes: A Meta-Analysis. *Can J Diabetes*. 2015 Aug;39(4):266–72.

9. Nicholson A, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. *Eur Heart J*. 2006 Dec 1;27(23):2763–74.
10. Dickens C. Depression in people with coronary heart disease: prognostic significance and mechanisms. *Curr Cardiol Rep*. 2015 Oct;17(10):83.
11. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care*. 2001 Jun;24(6):1069–78.
12. Fulton S, Décarie-Spain L, Fioramonti X, Guiard B, Nakajima S. The menace of obesity to depression and anxiety prevalence. *Trends Endocrinol Metab TEM*. 2022 Jan;33(1):18–35.
13. Laursen TM, Musliner KL, Benros ME, Vestergaard M, Munk-Olsen T. Mortality and life expectancy in persons with severe unipolar depression. *J Affect Disord*. 2016 Mar 15;193:203–7.
14. May HT, Horne BD, Knight S, Knowlton KU, Bair TL, Lappé DL, et al. The association of depression at any time to the risk of death following coronary artery disease diagnosis. *Eur Heart J Qual Care Clin Outcomes*. 2017 Oct 1;3(4):296–302.
15. Dong G, Feng J, Sun F, Chen J, Zhao XM. A global overview of genetically interpretable multimorbidities among common diseases in the UK Biobank. *Genome Med*. 2021 Jul 5;13:110.
16. Amare AT, Schubert KO, Klingler-Hoffmann M, Cohen-Woods S, Baune BT. The genetic overlap between mood disorders and cardiometabolic diseases: a systematic review of genome wide and candidate gene studies. *Transl Psychiatry*. 2017 Jan;7(1):e1007.
17. Wong BCF, Chau CKL, Ao FK, Mo CH, Wong SY, Wong YH, et al. Differential associations of depression-related phenotypes with cardiometabolic risks:

Polygenic analyses and exploring shared genetic variants and pathways. *Depress Anxiety*. 2019 Apr;36(4):330–44.

18. Hagenaars SP, Coleman JRI, Choi SW, Gaspar H, Adams MJ, Howard DM, et al. Genetic comorbidity between major depression and cardio-metabolic traits, stratified by age at onset of major depression. *Am J Med Genet B Neuropsychiatr Genet*. 2020;183(6):309–30.
19. Baranova A, Liu D, Chandhoke V, Cao H, Zhang F. Unraveling the genetic links between depression and type 2 diabetes. *Prog Neuropsychopharmacol Biol Psychiatry*. 2025 Jan 19;137:111258.
20. Lee C, Scherer SW. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med*. 2010 Mar 9;12:e8.
21. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet*. 2014 Oct;46(10):1063–71.
22. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011 Aug 14;43(9):838–46.
23. Rees E, Walters JTR, Georgieva L, Isles AR, Chambert KD, Richards AL, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry*. 2014 Feb;204(2):108–14.
24. Truty R, Paul J, Kennemer M, Lincoln SE, Olivares E, Nussbaum RL, et al. Prevalence and properties of intragenic copy-number variation in Mendelian disease genes. *Genet Med Off J Am Coll Med Genet*. 2019 Jan;21(1):114–23.
25. Chawner SJRA, Owen MJ, Holmans P, Raymond FL, Skuse D, Hall J, et al. Genotype–phenotype associations in children with copy number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): a case-control cohort study. *Lancet Psychiatry*. 2019 Jun 1;6(6):493–505.

26. Kendall KM, Rees E, Bracher-Smith M, Legge S, Riglin L, Zammit S, et al. Association of Rare Copy Number Variants With Risk of Depression. *JAMA Psychiatry*. 2019 Aug;76(8):818–25.
27. Crawford K, Bracher-Smith M, Owen D, Kendall KM, Rees E, Pardiñas AF, et al. Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *J Med Genet*. 2019 Mar 1;56(3):131–8.
28. Voll SL, Boot E, Butcher NJ, Cooper S, Heung T, Chow EWC, et al. Obesity in adults with 22q11.2 deletion syndrome. *Genet Med Off J Am Coll Med Genet*. 2017 Feb;19(2):204–8.
29. Van L, Heung T, Malecki SL, Fenn C, Tyrer A, Sanches M, et al. 22q11.2 microdeletion and increased risk for type 2 diabetes. *EClinicalMedicine*. 2020 Sep 10;26:100528.
30. Perrone L, Marzuillo P, Grandone A, del Giudice EM. Chromosome 16p11.2 deletions: another piece in the genetic puzzle of childhood obesity. *Ital J Pediatr*. 2010 Jun 11;36:43.
31. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018 Oct;562(7726):203–9.
32. Kendall KM, Rees E, Escott-Price V, Einon M, Thomas R, Hewitt J, et al. Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biol Psychiatry*. 2017 Jul 15;82(2):103–10.
33. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007 Jan 11;17(11):1665–74.
34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2022.

35. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*. 2010 Feb;463(7281):666–70.
36. Hanssen R, Auwerx C, Jöeloo M, Sadler MC, Henning E, Keogh J, et al. Chromosomal deletions on 16p11.2 encompassing *SH2B1* are associated with accelerated metabolic disease. *Cell Rep Med*. 2023 Aug 15;4(8):101155.
37. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*. 2010 Feb;463(7281):671–5.
38. RL C, Jt G, E P, M L, R B, C L, et al. A cross-disorder dosage sensitivity map of the human genome. *Cell [Internet]*. 2022 Apr 8 [cited 2025 Jan 14];185(16). Available from: <https://pubmed.ncbi.nlm.nih.gov/35917817/>
39. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018 May;50(5):668–81.
40. Meier SM, Trontti K, Purves KL, Als TD, Grove J, Laine M, et al. Genetic Variants Associated With Anxiety and Stress-Related Disorders: A Genome-Wide Association Study and Mouse-Model Study. *JAMA Psychiatry*. 2019 Sep 1;76(9):924–32.
41. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and Refinement of Loci Associated with Lipid Levels. *Nat Genet*. 2013 Nov;45(11):1274.
42. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015 Feb;518(7538):197–206.
43. Keaton JM, Kamali Z, Xie T, Vaez A, Williams A, Goleva SB, et al. Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat Genet*. 2024 May;56(5):778–91.

44. A M, D T, M T, Nr R, Jm T, Nw R, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet [Internet]. 2018 Nov [cited 2024 Sep 18];50(11). Available from: <https://pubmed.ncbi.nlm.nih.gov/30297969/>
45. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun [Internet]. 2016 [cited 2024 Sep 18];7. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4735748/>
46. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019 Apr 16;10(1):1776.
47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.
48. Khan A, Turchin MC, Patki A, Srinivasasainagendra V, Shang N, Nadukuru R, et al. Genome-wide polygenic score to predict chronic kidney disease across ancestries. Nat Med. 2022 Jul;28(7):1412–20.
49. Katzourou IK, Consortium L, Barroso I, Bengner L, Ingason A, Stow D, et al. Contributions of common and rare genetic variation to different measures of mood and anxiety disorder in the UK Biobank. BJPsych Open. 2025 May;11(3):e97.
50. Alten S van, Domingue BW, Faul J, Galama T, Marees AT. Reweighting UK Biobank corrects for pervasive selection bias due to volunteering. Int J Epidemiol. 2024 May 7;53(3):dyae054.
51. Niarchou M, Chawner SJRA, Doherty JL, Maillard AM, Jacquemont S, Chung WK, et al. Psychiatric disorders in children with 16p11.2 deletion and duplication. Transl Psychiatry. 2019 Jan 16;9(1):8.

52. Linden SC, Watson CJ, Smith J, Chawner SJRA, Lancaster TM, Evans F, et al. The psychiatric phenotypes of 1q21 distal deletion and duplication. *Transl Psychiatry*. 2021 Feb 4;11(1):105.
53. Schneider M, Debbané M, Bassett AS, Chow EWC, Fung WLA, van den Bree MBM, et al. Psychiatric Disorders From Childhood to Adulthood in 22q11.2 Deletion Syndrome: Results From the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *Am J Psychiatry*. 2014 Jun;171(6):627–39.
54. Adams RL, Baird A, Smith J, Williams N, van den Bree MBM, Linden DEJ, et al. Psychopathology in adults with copy number variants. *Psychol Med*. 53(7):3142–9.
55. Crawford K, Bracher-Smith M, Owen D, Kendall KM, Rees E, Pardiñas AF, et al. Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *J Med Genet*. 2019 Mar;56(3):131–8.
56. Auwerx C, Lepamets M, Sadler MC, Patxot M, Stojanov M, Baud D, et al. The individual and global impact of copy-number variants on complex human traits. *Am J Hum Genet*. 2022 Apr 7;109(4):647–68.
57. Stahel P, Nahmias A, Sud SK, Lee SJ, Pucci A, Yousseif A, et al. Evaluation of the Genetic Association Between Adult Obesity and Neuropsychiatric Disease. *Diabetes*. 2019 Sep 10;68(12):2235–46.
58. Cw W, Ty L, C W, S S, B Z, L S, et al. Copy Number Variation Analysis Facilitates Identification of Genetic Causation in Patients with Congenital Anomalies of the Kidney and Urinary Tract. *Eur Urol Open Sci [Internet]*. 2022 Jan 9 [cited 2025 Jan 21];44. Available from: <https://pubmed.ncbi.nlm.nih.gov/36185583/>
59. Berberich AJ, Huot C, Cao H, McIntyre AD, Robinson JF, Wang J, et al. Copy Number Variation in GCK in Patients With Maturity-Onset Diabetes of the Young. *J Clin Endocrinol Metab*. 2019 Mar 26;104(8):3428–36.

60. Mei L, Gao Y, Chen M, Zhang X, Yue W, Zhang D, et al. Overlapping common genetic architecture between major depressive disorders and anxiety and stress-related disorders. *Prog Neuropsychopharmacol Biol Psychiatry*. 2022 Mar 8;113:110450.
61. Kessler RC, Gruber M, Hettema JM, Hwang I, Sampson N, Yonkers KA. Co-morbid major depression and generalized anxiety disorders in the National Comorbidity Survey follow-up. *Psychol Med*. 2008 Mar;38(3):365–74.
62. Martin J, Asjadi K, Hubbard L, Kendall K, Pardiñas AF, Jermy B, et al. Examining sex differences in neurodevelopmental and psychiatric genetic risk in anxiety and depression. *PLoS ONE*. 2021 Sep 2;16(9):e0248254.
63. Polyak A, Rosenfeld JA, Girirajan S. An assessment of sex bias in neurodevelopmental disorders. *Genome Med*. 2015 Aug 27;7(1):94.
64. Jung B, Ahn K, Justice C, Norman L, Price J, Sudre G, et al. Rare copy number variants in males and females with childhood attention-deficit/hyperactivity disorder. *Mol Psychiatry*. 2023 Mar;28(3):1240–7.
65. Martin J, Tammimies K, Karlsson R, Lu Y, Larsson H, Lichtenstein P, et al. Copy number variation and neuropsychiatric problems in females and males in the general population. *Am J Med Genet*. 2019 Sep;180(6):341–50.
66. Mollon J, Schultz LM, Huguet G, Knowles EEM, Mathias SR, Rodrigue A, et al. Impact of Copy Number Variants and Polygenic Risk Scores on Psychopathology in the UK Biobank. *Biol Psychiatry*. 2023 Oct 1;94(7):591–600.
67. Vaez M, Montalbano S, Waples R, Krebs MD, Hellberg KLG, Gådin J, et al. Evaluating the Joint Effects of Recurrent Copy Number Variants and Polygenic Scores on the Risk of Psychiatric Disorders in the iPSYCH2015 Case-Cohort Sample [Internet]. *medRxiv*; 2024 [cited 2025 Feb 11]. p. 2024.09.23.24314234. Available from: <https://www.medrxiv.org/content/10.1101/2024.09.23.24314234v3>

68. Tian R, Ge T, Kweon H, Rocha DB, Lam M, Liu JZ, et al. Whole-exome sequencing in UK Biobank reveals rare genetic architecture for depression. *Nat Commun*. 2024 Feb 26;15:1755.
69. Zamariolli M, Auwerx C, Sadler MC, Graaf A van der, Lepik K, Schoeler T, et al. The impact of 22q11.2 copy-number variants on human traits in the general population. *Am J Hum Genet*. 2023 Jan 26;110(2):300.
70. Chowdhury SR, Das DC, Sunna TC, Beyene J, Hossain A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. *eClinicalMedicine* [Internet]. 2023 Mar 1 [cited 2025 Jan 21];57. Available from: [https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370\(23\)00037-8/fulltext](https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(23)00037-8/fulltext)
71. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017 Nov 1;186(9):1026–34.

