npj | genomic medicine

Article

Published in partnership with CEGMR, King Abdulaziz University

https://doi.org/10.1038/s41525-024-00426-8

Polygenic scores stratify neurodevelopmental copy number variant carrier cognitive outcomes in the UK Biobank

Check for updates

Thomas J. Dinneen[1,2] ✉, Fiana Ní Ghrálaigh[3], Cathal Ormond[1], Elizabeth A. Heron[1], George Kirov[4], Lorna M. Lopez[3] & Louise Gallagher[1,2,5,6]

Rare copy-number variants associated with neurodevelopmental conditions (ND-CNVs) exhibit variable expressivity of clinical, physical, behavioural outcomes. Findings from clinically ascertained cohorts suggest this variability may be partly due to additional genetic variation. Here, we assessed the impact of polygenic scores (PGS) and rare variants on ND-CNV carrier fluid intelligence (FI) scores in the UK Biobank. Greater PGS for cognition (PS_{Cog}) and educational attainment (PS_{EA}) is associated with increased FI scores in all ND-CNVs (n = 1317), 15q11.2 del. (n = 543), and 16p13.11 dup. carriers (n = 275). No association of rare variants associated with intellectual disability, autism, or putatively loss-of-function, brain-expressed genes was found. Positive predictive values in the first deciles of PS_{cog} and PS_{EA} showed a two- to five-fold increase in the rate of low FI scores compared to baseline rates. These findings demonstrate that PGS can stratify ND-CNV carrier cognitive outcomes in a population-based cohort.

Copy number variants (CNVs) are a class of large structural variants (>50bp) that can encompass one or many genes[1]. CNVs in 49 genomic loci have been associated with neurodevelopmental disorders (NDDs), primarily intellectual disability/developmental delay (ID/DD) and autism, i.e. ND-CNVs[2,3] (Supplementary Table 4). Carriers often present with heterogenous clinical, physical, behavioural, and cognitive outcomes, including some carriers that are clinically unaffected. Phenotypic variability or incomplete penetrance is also observed in related individuals in the context of inherited ND-CNV in families[4]. Sex differences in clinical outcomes have been observed in ND-CNV carriers in clinical cohort studies. For example, females in families with autistic individuals tend to carry larger CNVs and are clinically unaffected compared to their male carrier relatives[5]. Deleterious CNVs (>400kb) are also more likely to be maternally inherited[6].

Significant heterogeneity of outcomes and complex inheritance patterns hamper clinical interpretation of ND-CNVs. This can be a barrier to providing appropriate counselling and creates uncertainty, and confusion for parents and caregivers[7]. It raises the question; can we further stratify ND-CNV carriers based on additional genetic factors to identify those predisposed to NDD outcome(s)?

Additional genetic variation may partially explain heterogenous outcomes in ND-CNV carriers[8]. Studies on 22q11.2 deletion (del.)[9,10], 16p12.1 del[11]., and other ND-CNVs[12,13] have demonstrated the effectiveness of utilising polygenic scores (PGS) and additional rare variants (≤1% minor allele frequency [MAF]) to stratify ND-CNV carrier risk to NDDs in clinically ascertained psychiatric cohorts. ND-CNV carriers in these clinical studies are likely enriched for morbidity due to clinical ascertainment. Similarly, participants in these cohorts are enriched for highly penetrant ND-CNVs, like 22q11.2 deletion, and contain smaller numbers of ND-CNVs with lower penetrance estimates[14]. These penetrance estimates are also not proportionate to ND-CNV size. To understand the effects of additional genetic variants in ND-CNV carriers on NDD-related outcomes more generally, further analyses are required in both family and population cohorts, with the latter capturing a greater range of heterogeneous outcomes. Large, well-phenotyped population biobanks, such as the UK Biobank (UKB)[15], enable

[1]Trinity College Dublin, Department of Psychiatry, School of Medicine, Trinity Centre for Health Sciences, St. James' Hospital, Dublin 8, Ireland. [2]The Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, 686 Bay St., Toronto, ON, M5G 0A4, Canada. [3]Department of Biology, Maynooth University, Maynooth, Co, Kildare, Ireland. [4]MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK. [5]Centre for Addiction and Mental Health, 80 Workman Way, Toronto, ON, M6J 1H4, Canada. [6]Department of Psychiatry, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, M5S 1A1, Canada. ✉e-mail: tom.dinneen@sickkids.ca

investigations of ND-CNVs with smaller penetrance estimates in carriers that were not clinically ascertained. ND-CNVs in the UKB have been associated with lower cognitive performance[16], depression[3], and other medical/physical phenotypes[17,18]. More recently, a combined effect of PGS and ND-CNVs on psychopathology in the UKB has been observed[19]. However, the effects of both rare and common additional genetic variants in ND-CNV carriers on cognitive outcomes have not been investigated.

Here we assess whether additional genetic variants in the background of ND-CNV carriers modify cognitive outcomes. We use fluid intelligence (FI) score as a proxy for cognitive outcomes given its strong reliability estimates in the UKB[20,21]. First, we investigated the relationship between additional common and rare genetic variants in ND-CNV carriers and FI scores to determine if there were observable modifying effects of additional variants in a population-based cohort. Second, we examined the accuracy of cognition PGS (PS_{cog}) and educational attainment PGS (PS_{EA}) prediction of lower FI scores in ND-CNV carriers. Third, we assessed whether ND-CNV carriers had greater PGS associated with higher cognitive performance and fewer rare additional genetic variants compared to non-carriers. We hypothesised that ND-CNV carriers have higher PGS associated with increased cognitive performance and have fewer variants associated with lower cognitive outcomes in comparison with non-ND-CNV carriers, i.e. non-carriers, as a protective effect[10]. This is due to UKB participants having higher than average cognitive performance and low ID/DD rates. Sex-effects associated with ND-CNV carrier additional variant load have been investigated in clinically ascertained cohorts[11] but remain to be tested in population-based cohorts. We hypothesised that male ND-CNV carriers would have more additional variants associated with higher cognitive outcomes and carry fewer additional variants associated with lower cognitive outcomes, in agreement with previous findings from clinically ascertained cohorts[11,22,23].

Results

Cohort description

After inclusion criteria and genetic QC filters were applied, 153,643 UKB participants were included in the PGS analyses (*n* ND-CNV = 1318), and 148,169 participants were included in the rare variant analyses (*n* ND-CNV = 1275) (Table 1.). ND-CNVs in 30 of 49 loci were carried by participants included in the analysis (Supplementary Table 5.). All 30 ND-CNVs were heterozygous and are associated with ID/DD. 18 are associated with autism, 13 with schizophrenia (SCZ), and nine with attention deficit hyperactivity disorder (ADHD). ND-CNV carriers showed, on average, lower FI scores than non-carriers with a medium estimate effect size ($P = <2.2 \times 10^{-16}$, *Cohen's d* = 0.42). There were no sex differences in ND-CNV carrier status or in ND-CNV carrier FI scores. Mean FI scores were lower in non-carrier females than males although the estimated effect size was negligible ($P = <2.2 \times 10^{-16}$, *Cohen's d* = 0.09). ND-CNV carriers had on average higher Townsend deprivation index (TDI) scores than non-carriers with an estimated small effect size ($P = <2.2 \times 10^{-16}$, *Cohen's d* = 0.21). There were no sex differences in TDI scores in ND-CNV carriers or non-carriers (Table 1.).

Relationship between additional variants and FI scores in ND-CNV carriers

Linear regression analyses were used to investigate the relationship between additional variants in the background of ND-CNV carriers and FI scores in all ND-CNV carriers, carriers of ND-CNV duplications, carriers of ND-CNV deletions, carriers of specific ND-CNV loci, and non-carriers, adjusting for covariates (Methods). All variance inflation factor (VIF) values were under 3 (Supplementary Tables 6–8, 12–14).

First, we assessed the effect of PGS for cognition (PS_{Cog}; 59,368 variants included), educational attainment (PS_{EA}; 42,361 variants included), and autism (PS_{ASD}; 59,368 variants included), on ND-CNV carrier FI scores. One ND-CNV carrier and 35 non-carriers were removed due to missing values for educational attainment and employment status. PS_{Cog} and PS_{EA} were significantly associated with FI scores in all ND-CNV carriers, in carriers of specific ND-CNV loci, and in non-carriers (Table 2., Fig. 1., and Supplementary Figures 3. and 4.). According to *Cohen's* f^2 estimates, PS_{Cog} had a medium global effect size and a small local effect size in all ND-CNV carriers and specific ND-CNV loci. PS_{EA} had an estimated small local effect size in all ND-CNV carriers and specific ND-CNV loci. PS_{ASD} was significantly associated with FI scores in non-carriers but showed no association with FI scores in ND-CNV carriers or specific ND-CNVs (Table 2. and Supplementary Fig. 5.). Analyses were also run in a subset of ND-CNV carriers, duplication carriers and deletion carriers excluding 15q11.2 del. and 16p13.11 dup. Two significant association were observed between PS_{EA} and FI scores in ND-CNV carriers (excluding 15q11.2 del. and 16p13.11 dup.) and in deletion carriers (excluding 15q11.2 del.) (Supplementary Table 9.). The model with covariates only had an adjusted R^2 of 0.15- 0.19 depending on the subset. Running the models without the educational attainment covariate resulted in reduced adjusted R^2 and increase beta estimates for PS_{Cog} and PS_{ASD} to the adjusted models in Table 2 (Supplementary Table 10.).

There were no significant associations between rare variant load (SFARI/DDD/LOF) and FI scores in the ND-CNV group or specific ND-CNV groups (Supplementary table 11. and Supplementary Figures 6–8). Rare variants in LOF loci, i.e. intolerant to loss-of-function (LoF), brain-expressed genes, were associated with lower FI scores in non-carriers, however, SFARI and DDD loci were not. Demographics of this subset are detailed in Supplementary Table 3 whereas rare variant counts per gene set are shown in Supplementary Table 6. There was no change in the association of rare variants with FI scores when the models were run without the educational attainment covariate (Supplementary Table 15).

Finally, we replicated previous reports of significantly negative effects of ND-CNV loci on FI scores in the UKB however, we also adjusted for the effect of background additional common and rare genetic variants (Supplementary Table 16.).

Table 1 | Demographics of included ND-CNV carriers and non-carriers who attempted the fluid intelligence (FI) test

Variable	All ND-CNVs	All Dup. carriers	All Del. carriers	15q11.2 del.	16p13.11 dup.	22q11.2 dup.	Non-carriers
n max.	1318	759	546	543	275	83	152,325
Sex (%male)	51	50	52	50	49	53	47
Age (years)	57.16 [8.35]	57.65 [8.36]	56.53 [8.33]	58.33 [8.35]	57.32 [8.14]	55.39 [8.31]	58.53 [8.28]
TDI score (z)	−1.01 [2.99]	−1.23 [2.89]	−0.69 [3.12]	−1.34 [2.77]	−0.98 [3.08]	−0.44 [2.87]	−1.58 [2.77]
FI score (z)	−0.41 [0.96]	−0.45 [0.99]	−0.38 [0.94]	−0.33 [0.92]	−0.3 [0.98]	−0.58 [0.79]	0.004 [0.1]
- *male*	−0.39 [0.98]	−0.44 [1.05]	−0.35 [0.93]	−0.27 [0.92]	−0.26 [1.04]	−0.54 [0.82]	−0.06 [1.03]
- *female*	−0.43 [0.94]	−0.46 [0.93]	−0.41 [0.94]	−0.4 [0.91]	−0.35 [0.91]	−0.62 [0.76]	−0.04 [0.97]

The mean [s.d] are presented for quantitative phenotypes. FI scores and TDI scores are presented as z-scores (Methods and Supplementary Information). See Supplementary Table 2. for an expanded description of the cohort and additional variables. See Supplementary Table 3. for a description of participants included in rare variant analyses. A full list of ND-CNVs included in the analysis can be found in Supplementary Table 4. *Dup.* Duplication; *Del.* Deletion.

Table 2 | Relationship between FI scores and PGS in ND-CNV carriers and non-carriers

ND-CNV Cohort	n	IV	β	Adj. R^2	*P*	Cohen's f^2 (Global)	Cohen's f^2 (Local)
All ND-CNV carriers	1,317	PS_Cog	0.13	0.18	4.92×10^{-7}*	0.23	0.02
		PS_EA	0.15	0.08	7.79×10^{-5}*	0.09	0.02
		PS_ASD	5.00×10^{-3}	0.17	0.85	-	-
All Del. Carriers	764	PS_Cog	0.12	0.19	1.44×10^{-4}*	0.25	0.02
		PS_EA	0.15	0.08	6.78×10^{-6}*	0.09	0.03
		PS_ASD	-4.70×10^{-3}	0.18	0.88	-	-
All Dup. Carriers	561	PS_Cog	0.12	0.19	2.16×10^{-3}*	0.23	0.02
		PS_EA	0.14	0.08	4.76×10^{-4}*	0.09	0.02
		PS_ASD	−0.10	0.18	0.80	-	-
15q11.2 del.	543	PS_Cog	0.17	0.18	1.80×10^{-5}*	0.23	0.03
		PS_EA	0.15	0.09	2.83×10^{-4}*	0.10	0.02
		PS_ASD	-1.10×10^{-3}	0.16	0.78	-	-
16p13.11 dup.	275	PS_Cog	0.11	0.20	0.04	0.24	0.01
		PS_EA	0.17	0.06	3.93×10^{-3}*	0.07	0.03
		PS_ASD	0.06	0.19	0.72	-	-
Non-carriers	152,290	PS_Cog	0.10	0.15	$< 2.00 \times 10^{-16}$*	0.18	0.01
		PS_EA	0.18	0.06	$< 2.00 \times 10^{-16}$*	0.07	0.04
		PS_ASD	1.30×10^{-3}	0.15	9.01×10^{-8}*	0.17	2.00×10^{-4}

All analyses were adjusted for covariates (Methods) and only covariates with significant associations were retained. Seven individuals carried both a deletion and duplication of which four carried both 15q11.2 del. and 16p13.11 dup. All ND-CNV carriers includes carriers of 15q11.2 del. and 16p13.11 dup as well as other loci (See Supplementary Table 5.). The standard β estimates are reported along with the adjusted (adj.) R^2 from multiple linear regression. Both the global and local *Cohen's* f^2 are reported for significant p-values. The local *Cohen's* f^2 is only calculated for IVs. Adj. R^2 = 0.15–0.19 for PS_{Cog} and PS_{ASD} models with only covariates for each subset. Adj. R^2 = 0.03–0.08 for PS_{EA} models with only covariates for each subset. * = *P*-values that remained significant after Bonferroni correction for 27 analyses (three PGS and nine sub-groups; $p < 1.85 \times 10^{-3}$). *IV* independent variable.

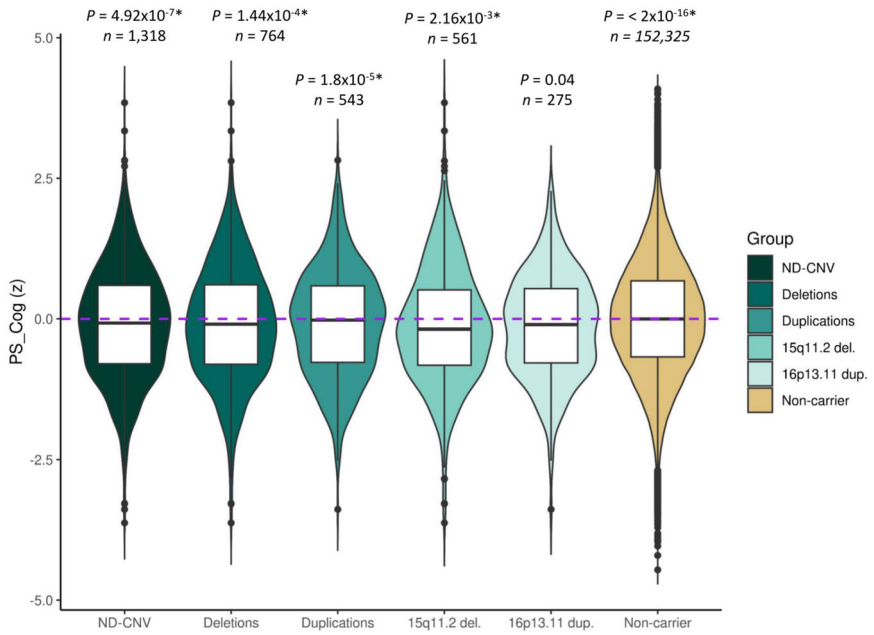

Fig. 1 | Distribution of cognition polygenic scores amongst ND-CNV carriers. The violin-boxplots show all ND-CNV carrier (*n* = 1318), 15q11.2 del. carriers (*n* = 543), 16p13.11 dup. carriers (*n* = 275) and non-carrier (*n* = 152,325) distributions of PS_Cog (z-score). It displays the minimum (bottom of violin plot), maximum (top of violin plot), median (black line), cohort mean (purple-dashed line) and interquartile range (black bars of boxplot) of PS_Cog. The *P*-values shown are derived from the linear regression analyses listed in Table 2. adjusted for standard covariates (methods) and not corrected for multiple testing. * Indicates *P*-values that remained significant after Bonferroni correction for 12 analyses.

Individual prediction of lower FI scores using polygenic scores

We calculated positive predictive values (PPVs) for ND-CNV carriers grouped based on deciles of PGS. FI scores were converted to a binary outcome by categorising participant scores using a cut-off of two standard deviations below the cohort mean. 7% of ND-CNV carriers, 6% of deletion carriers, 4% of duplication carriers, 4% of 15q11.2 del. carriers, 4% of 16p13.11 dup. carriers, and 3% of non-carriers had FI scores two standard-deviations below the cohort mean (Fig. 2.).

We observed that 13% of ND-CNV carriers in the first decile for PS_{Cog} had FI scores two standard deviations below the mean (PPV = 13%, *P* = 0.01) whereas 1% of individuals in the tenth decile of PS_{Cog}, had FI scores two standard deviations below the mean (PPV = 2%, *P* = 0.02). Thus, ND-CNV carriers in the first decile of PS_{Cog} have an almost two-fold increase in low FI scores compared to the pre-test probability for the group. Similar results are observed of duplication ND-CNVs, 15q11.2 del. and 16p13.11 dup. carriers, and non-carriers (Fig. 2a, b.; Supplementary Tables

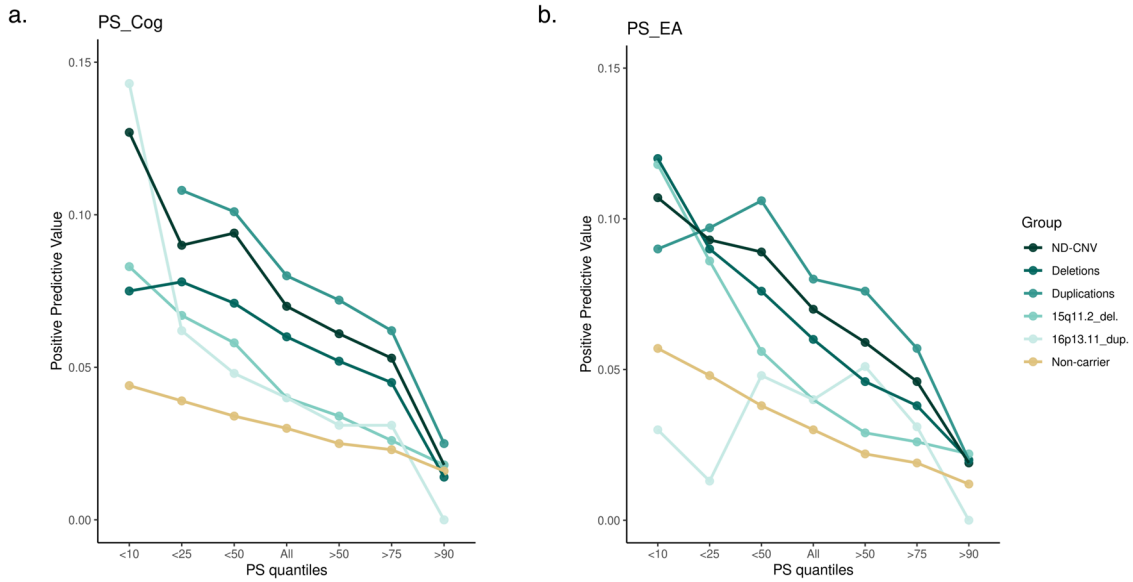

Fig. 2 | Individual prediction of low FI scores by deciles of PGS. PPVs (y-axis) for PS_Cog (**a**) and PS_EA (**b**) are shown here for deciles of both PGS. The proportion of low FI scores for each of the groups are as follows: ND-CNV = 0.07, Deletions = 0.06, Duplication = 0.08, 15q11.2 del. = 0.04, 16p13.11 dup. = 0.04, and non-carrier = 0.04. 95% confidence intervals are given in Supplementary Tables 17 and 18. '*All*' *indicates the baseline probability for each group.*

17. and 18.). Interestingly, carriers of duplication ND-CNVs in the first decile of PS_{Cog} had an almost five-fold increase in low FI scores compared to the pre-test probability for the group (PPV = 0.19; $P = 3.38 \times 10^{-3}$) (Fig. 2a. and Supplementary Table 18.). 11% of ND-CNV carriers in the lowest decile for PS_EA had FI scores two standard deviations below the mean but were not significant (PPV = 11%, $P = 0.08$). In contrast 11% of 15q11.2 del. carriers in the lowest decile for PS_EA had FI scores two standard deviations below the mean (PPV = 11%, $P = 5.71 \times 10^{-3}$) and were significant. 2% of individuals in the highest decile of PS_{EA} had FI scores two standard deviations below the mean (PPV = 2%, $P = 0.03$) (Fig. 2b. and Supplementary Table 18.).

Additional variant load differences in ND-CNV carriers

In line with our hypotheses, we tested differences in the genetic load of common variants that are associated with higher cognitive outcomes between carriers of ND-CNVs compared with non-carriers. Next, we compared the genetic load of rare variants associated with lower cognitive ability between carriers of ND-CNVs compared with non-carriers.

Statistically significant differences in common variant load were assessed using two sample, two-sided t-tests. We observed six significant differences in PS_{Cog} and PS_{EA} in all ND-CNV carriers and 15q11.2 del. carriers, however, only one remained significant following Bonferroni correction for 45 tests. Distributions of the PGS in each sub-group of ND-CNV carrier and non-carrier are shown in Supplementary Figs. 9–11. Contrary to our hypothesis, mean PS_{EA} was significantly reduced in the ND-CNV carriers ($\mu = -0.09$) compared to non-carriers ($\mu = -1 \times 10^{-4}$; $P = 4.41 \times 10^{-4}$) (Fig. 3. and Supplementary Table 19.). However, the effect size was negligible (*Cohen's d* = 0.01). For rare variants, Wilcoxon Rank Sum Tests showed no significant differences in the mean SFARI and DDD variants between ND-CNV carriers and non-carriers. Similarly, there were no significant differences in the mean number of LOF variants in brain expressed genes between ND-CNV carriers and non-carriers (Supplementary Table 20.). Counts of rare variants per gene-set (DDD, SFARI, LOF) in each sub-group of ND-CNV carrier and non-carrier are shown in Supplementary Figs. 12–14.

Finally, we investigated whether there were sex differences in additional common and rare variant load in the all ND-CNV, 15q11.2 del., 16p13 dup., 22q11.2 dup., 1q21. dup. carrier and non-carrier groups. Contrary to our hypotheses that males would carry more common variants associated with higher cognitive outcomes and fewer rare, deleterious variants compared with females, no sex differences in PS_{Cog} and PS_{EA} were observed in any ND-CNV carrier group. Male non-carriers have lower PS_{EA} ($\mu = -0.01$) than females ($\mu = 0.01$) although the effect size was negligible ($P = 1.51 \times 10^{-4}$, *Cohen's d* = 0.02) (Supplementary Table 19). Similarly, no sex differences were found in rare variant load in ND-CNV carriers. Female non-carriers had a greater LOF rare variant load (median [range] = 0 [0, 5]) than male non-carriers (median [range] = 0 [0,5]; $P = 1.05 \times 10^{-11}$). However, the estimated effect size was negligible ($r = 0.02$) (Supplementary Table 20.).

Discussion

Here we report the first investigations into the effect of additional genetic variants on ND-CNV carrier cognitive outcomes in a population-based biobank. Analyses within ND-CNV carriers indicated that PS_{Cog} and PS_{EA} are significantly associated with higher FI scores in ND-CNV carriers with small to medium global effect size estimates, explaining ~18-19% and ~8% of the variance respectively when adjusted for covariates. We also report a novel finding of the effect of PS_{Cog} and PS_{EA} on FI scores in carriers of 15q11.2 del. and 16p13.11 dup. The PS_{Cog} was associated with a positive and medium effect on FI scores in both groups, explaining 20% and 21% of the variance respectively when adjusted for covariates. The PS_{EA} was associated with a positive and small effect on FI scores in both groups, explaining 9% and 6% of the variance respectively when adjusted for covariates. Furthermore, we replicated previous results that demonstrated that ND-CNVs carried by participants in the UKB are associated with lower FI scores[16,24], even when controlled for both common and rare additional genetic variants (Supplementary Table 16). Indeed, controlling for additional genetic variants does significantly increase the variance explained in FI scores by multiple-linear regression models.

The only previous investigation into the effect of additional common genetic variants in the background of carriers of these ND-CNV loci reported an association between PGS for SCZ and SCZ diagnosis but did not investigate cognitive outcomes[13]. This previous study demonstrated that for ND-CNV loci with lower penetrance estimates, such as 15q11.2 del. and 16p13.11 dup., PGS has a stronger effect on SCZ outcome. Given a small number of carriers of other ND-CNVs with higher penetrance estimates in the UK Biobank compared with 15q11.2 del and 16p13.11 dup, we were unable to investigate differences in additional variant load between ND-

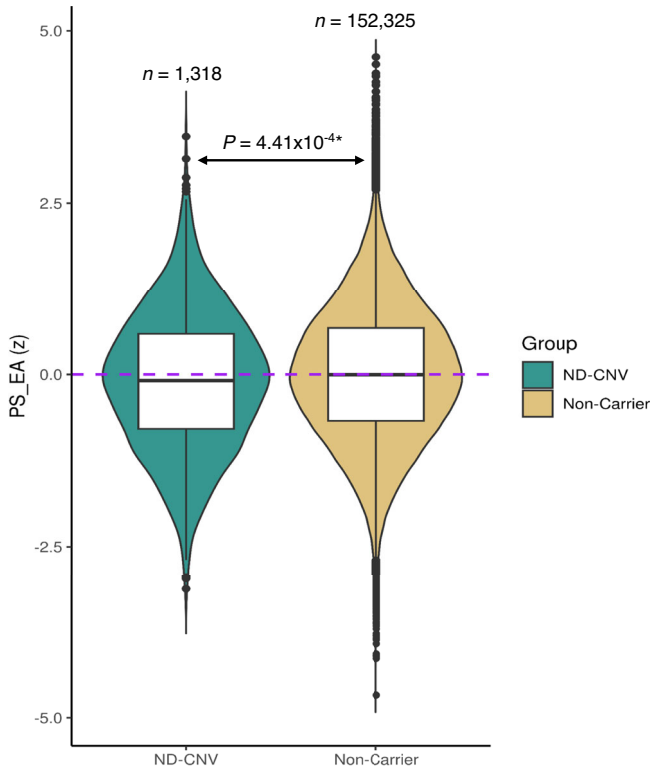

Fig. 3 | Distribution of PS_EA in all ND-CNV carriers vs non-carriers. The violin-boxplot shows the all ND-CNV carrier ($n = 1318$) and non-carrier ($n = 152,325$) groups distributions of PS_EA (z-score). It includes summaries of minimum (bottom of violin plot), maximum (top of violin plot), group median (black line), cohort mean (purple-dashed line) and interquartile range (black bars of boxplot) of PS_EA. The *p*-value shown is derived from the two-sample, two-sided *t*-test for PS_EA (Methods). * = passed Bonferroni correction for 36 tests. See Supplementary Table 19. for full *t*-test results.

CNV carrier groups stratified by penetrance. This was expected, given that this is a population-based cohort, and the previously identified ND-CNV loci with the highest frequencies in the UKB have low penetrance estimates[25,26].

Importantly, the covariates in our models (including educational attainment) explained ~15–19% of the variance in FI scores depending on the group observed. Educational attainment and TDI score were highly significant in all models. Omitting educational attainment as a covariate led to reductions in adjusted R^2 and increases in beta estimates for PS_{Cog}, PS_{ASD}, and rare variant linear regression models (Supplementary Tables 10, 15). Thus, the educational attainment covariate might be capturing environmental factors not captured by TDI score e.g. the number of years in education[27], quality of teaching[28]. This finding highlights modifiable environmental factors, such as education and socio-economic status, may make a stronger contribution to differences in cognitive outcomes compared with additional genetic factors in ND-CNV carriers in population cohorts.

The autism PGS has been associated with higher cognitive outcomes in the general population[29]. We observed an association between autism PGS and non-carrier FI scores but not with ND-CNV carrier FI scores. This lack of association in ND-CNV carriers may be due to a lack of power to detect small effects of the autism PGS or that the autism PGS was derived from less powered GWAS summary statistics (n significant loci = 5)[30]. Similarly, there may be a larger contribution from rare variants in participants with autism in the original GWAS sample given the inclusion of individuals with autism and ID. This could have led to a lower estimation of common variant effects.

We also calculated PPVs to assess the utility of PGS for cognition and educational attainment in predicting lower cognitive outcomes in ND-CNV carriers. This allows us to understand the effectiveness of stratification using PGS given PPVs are dependent on both prevalence and strength of association[10]. We observed that in all ND-CNV carriers, 15q11.2 del., and 16p13.11 dup. carriers in the lowest decile of PS_{Cog}, 13–19% had low FI scores (defined as ≥ 2 s.d below the mean) whereas only 1% in the highest decile of PS_{Cog} had a low FI score. This represents a two- to five-fold increase in probability for a low FI score compared with the pre-test probability for the respective ND-CNV groups. The most dramatic increase was observed in duplication ND-CNV carriers, indicating PGS may have a stronger influence in duplication carriers compared to deletion carriers. Similarly, for all ND-CNV carriers in the lowest decile of PS_{EA}, 11% had low FI scores vs 2% in the highest decile of PS_{EA}. However, the low frequency of individuals with ID/DD in the UKB means that low PS_{Cog} or PS_{EA} have potentially reduced predictive ability of low FI scores in population cohorts compared with clinically-ascertained cohorts, such as 22q11.2 DS[10].

No modifying effect was observed from an ND-CNV carrier's load of rare SNVs and indels in loci associated with autism and ID/DD on FI scores. We also did not observe any effect of intolerant to LoF SNVs and indels in brain expressed genes on ND-CNV carrier FI scores. This might be due to the low frequency of individuals in the cohort carrying multiple large effect rare variants in NDD-associated loci. This is to be expected in a population-based cohort. In fact, this result has also been observed in clinically ascertained cohorts. A previous study did not find significance in a 22q11.2 DS cohort when restricting to LOF, rare variants in SCZ-associated genetic loci[9]. Another study found only one significant association which was between the load of LOF rare variants carried by 16p12.1 del. carriers and a strong family history of severe clinical features[11]. A previous study in the UKB did observe a significant association of the load of rare protein truncating variants and missense variants in 599 DDG2P genes and lower FI scores[31]. However, this previous study did not adjust for the effects of socio-economic status, educational attainment, or employment status which have strong associations with FI scores in the UK Biobank[24]. Furthermore, we used a more stringent filtering pipeline and included 372 DDG2P genes with an association with ID/DD (Supplementary Figure 2.). The finding that rare variants in LOF loci is associated with low FI scores in non-carriers of ND-CNV replicates a previous finding, however, educational attainment was not adjusted for in this study[32]. In all, population cohorts may not be effective for investigating the modifying effect of rare variants on ND-CNV carrier NDD/NDD-related outcomes due to their expected low prevalence[32,33].

We found no evidence to support the hypotheses that ND-CNV carriers carry proportionately more additional variants associated with higher cognitive outcomes. In fact, ND-CNV carriers as a group had lower PS_{EA} compared to non-carriers although the estimated effect size was negligible. We expected ND-CNV carriers to have higher PGS than non-carriers within this population-based sample given the baseline negative effect of carrying an ND-CNV on FI scores compared to non-carriers. This finding is interesting as PGS for educational attainment may also be reflective of environmental factors. Although the UKB does not contain parental genetic data to investigate the inheritance of genetic factors, it can be speculated that the majority of the ND-CNVs in the UKB are inherited due to their low effect sizes[33]. PGS for educational attainment is also inherited and therefore, paired with the hypothesis that most ND-CNVs are inherited, there may be a small, assortative mating bias present in the UKB that has influenced this result. Indeed, a previous investigation into gene-environment effects across geographic regions in the UKB showed a strong geographic clustering for both educational attainment and cognition polygenic scores, suggesting assortative mating effects[34]. However, this conclusion cannot be validated in the UKB due to the absence of parental data. Moreover, the negligible effect size estimate suggests there may be no meaningful difference. To further investigate this hypothesis, data from a large, family-based cohort would be necessary. Similarly, our expectation that carriers of ND-CNVs carry fewer rare variants associated with lower cognitive outcomes compared to non-carriers was not supported. This may point to the differences in common variant load between clinically-ascertained and population-based biobanks, the former being more likely to have been ascertained based on

neurodevelopmental disorders, carrying more genetic risk factors, and more frequently observed ND-CNVs with higher penetrance estimates[9,10,12,13].

ND-CNV carriers did not show sex differences in additional variant load. This finding differs from studies of sex-differences in additional variant load in clinical cohorts[11,22,23]. The difference here may be related to lack of power to detect these differences in ND-CNV carriers. Similarly, we do not observe a sex bias in additional variant load. Importantly, UKB is a mostly healthy and well-educated cohort[15], thus, sex-differences in variant load might not be observable here given the under-representation of individuals at the lower end of the IQ scale.

Finally, there are a few limitations in this study. Firstly, the lack of an IQ measure in the UKB means cognitive outcomes can only be partially measured. While FI is a reliable proxy for IQ[20,21] there are some drawbacks to this measure. Most notably FI measures only a proportion of an individual's overall general intelligence or "*g*"[35] and does not measure fluid intelligence but rather verbal-numeric reasoning. We considered deriving a *g*-factor[35] from available UKB cognitive measures following a previously described method[21], however, we concluded this would make the interpretation of the results more ambiguous. We came to this conclusion as it requires the imputation of cognitive test results for many individuals across a range of cognitive tests with missing data and the arbitrary nature of defining the cognitive tests used to create the *g*-factor. An accurate *g*-factor requires multiple tests across multiple domains of cognitive ability which is not available in the UKB. In contrast, the UKB FI test is a defined measure of verbal-numeric reasoning that does not require participant results to be imputed and has strong correlations with other tests in the UKB[20,21]. Secondly, the under-representation of highly penetrant ND-CNVs meant we could not compare the effect of additional variant load between large and small effect ND-CNVs. However, this cohort does provide the opportunity to study largely non-affected carriers of ND-CNVs with lower penetrance estimates which is often not the case in clinically ascertained cohorts. Registry recruited cohorts such as iPsych[36] might lend themselves better to investigating the effect of additional variants between large and small effect ND-CNVs[37].

In conclusion, PGS for cognition and educational attainment show a strong, positive relationship with ND-CNV carrier cognitive outcomes. PGS for cognition also contributes to the cognitive outcomes of 15q11.2 del. and 16p13.11 dup. carriers. No protective effect of cognition or educational attainment PGS was observed in ND-CNV carriers and contrary to our hypothesis, ND-CNV carriers had lower educational attainment PGS than non-carriers. Rare additional variants do not modify ND-CNV carrier or non-carrier cognitive outcomes in this cohort, and no reduction of rare additional variant was observed in ND-CNV carriers in comparison with non-carriers. No sex effects were observed in ND-CNV carrier additional variant load despite male non-carriers having higher educational attainment PGS. These results show that stratification of ND-CNV carriers using PGS is possible in population-based cohorts, demonstrating the potential clinical utility of PGS and agreeing with previous results from clinically-ascertained cohorts.

Methods

The UK Biobank

The UKB is a database with both genetic and deep phenotypic information on over 500,000 participants that were recruited between 2006–2010. They were aged 40–69 years at time of recruitment and 54% are female. A full description of the UKB is detailed by Bycroft et al[15]. UKB data were released to Trinity College Dublin under approved project no. 48915. This study was performed in compliance with relevant ethical regulations and in accordance with the ethical standards of the Declaration of Helsinki. The UKB is approved by the North-West Multi-centre Research Ethics Committee (https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics). All participants provided their informed, explicit consent to take part in the UK Biobank project (https://www.ukbiobank.ac.uk/media/t22hbo35/consent-form.pdf). Participants were free to withdraw consent at any time and were withdrawn from all subsequent data analysis conducted in this paper.

Inclusion criteria for the present study are that a participant must be of British descent according to the UKB principal component analysis (UKB Field 22006), had attempted the fluid intelligence test, and had complete information for genetic sex, age, Townsend deprivation index (TDI) scores, and assessment centre location. Exclusion criteria included a diagnosis of a neurodegenerative conditions, and brain traumas (see Supplementary Table 1 for demographic information on excluded participants). Participants who had withdrawn consent from the UKB were also excluded from this analysis. Excluded samples are described in Supplementary Table 1.

Cognitive measures

The UKB fluid intelligence (FI) test scores were used as a proxy for cognitive outcomes in this analysis (UKB Field 20016). The FI test is a verbal-numeric test comprised of 13 questions and participants were given two minutes to complete the test. Despite its name, the FI test does not measure fluid intelligence (*g*F) directly, which is generally derived through multiple tests in different cognitive domains. The FI test in the UKB only measures the verbal-numeric reasoning cognitive domain. An ordinal score of 0–13 was given to participants who attempted the test. Where there were multiple attempts, the first FI test attempt score was included in the analysis. FI scores were converted to z-scores by normalising to the cohort mean. A full description of the FI test can be found in the *Supplementary Information*.

Genome-wide imputation data

Blood samples from ~450,000 UKB participants were genotyped using the Affymetrix UKB Axiom Array (825,927 markers) and ~50,000 UKB participants were genotyped using the UK BiLEVE Array (807,411 markers – 95% overlap with UKB Axiom Array). Genome-wide imputation was performed on 487,442 samples using IMPUTE2, yielding 97,059,328 variants. UKB genotype calling, quality control and genotype imputation method has been described in full previously[15]. The imputed genotype BGEN files were processed using PLINK 2.0[38] guided by the recommended RICOPILI QC pipeline for GWAS (https://sites.google.com/a/broadinstitute.org/ricopili/preimputation-qc). Briefly, variants were first passed through an imputation filter ≥0.8, then sample QC filters were applied for checking sex, removing samples with heterozygosity +/− 0.2, removing samples related at greater than or equal to the third degree (PI_HAT ≥ 0.125), and removing samples of non-British inferred ancestry using the UKB generated principal components. SNP QC filters were applied to exclude SNPs with genotype missingness ≥10%, sample missingness ≥10%, and MAFs ≤1%. Hardy-Weinberg equilibrium filters were applied to ND-CNV carriers and non-carriers SNPs separately. Carrier SNPs were included if they had a HWE ≥ 1×10^{-6} and non-carrier SNPs were included if they had a HWE ≥1×10^{-10}. This SNP QC was repeated, and no LD-pruning was applied as LD-clumping would be used when estimating polygenic scores.

Whole-exome sequence data

Whole exome sequencing (WES) was performed on 469,835 samples using the Illumina NovaSeq 6000 platform. BWA-MEM as used to map WES reads to the hg38 reference build of the human genome. WES variant calling and quality control have been described previously[39]. We performed additional QC and variant control using our own in-house pipeline. QC and variant filtration of the exome data was conducted through UKB-DNAnexus Research Analysis Platform (RAP). VCF files were subsetted to UKB participants who attempted the UKB FI touchscreen questionnaire and UKB participant samples that passed sample QC (sex-check, inferred -ancestry, relatedness) for imputed genotype data using BCFtools v1.15.1[40]. SNVs and Indels with a DP<10 and GQ<20 were marked as missing. Variants missing in >10% of genotypes and heterozygous variants with an allelic balance <20 were removed.

CNV calling

CNV calls for the UKB cohort were obtained via UKB returns field 1701[17]. CNVs were called across 92 putatively pathogenic CNVs[2] in 47 genomic loci using PennCNV on the raw intensity CEL files. The 92 CNVs were identified as pathogenic based on two independent studies which showed associations with genomic disorders, congenital malformations, NDDs (ID/DD and ASD) and other clinical phenotypes. The methods for CNV calling are fully described in Crawford et al[17]. A subset of 49 CNVs from this list are primarily associated with ID/DD but share associations with other NDDs such as autism, SCZ, and ADHD. These CNVs were selected based on substantial evidence for associations with and penetrance for NDDs. Only carriers of the smaller subset of 49 ND-CNVs are classed as ND-CNV carriers for this analysis. Carriers of CNVs with no penetrance information for NDDs or substantial evidence for NDD associations, other pathogenic CNVs not associated with NDDs, or large CNVs >500Mb were not analysed as they were considered not pathological regarding cognitive outcomes, thus, were included in the non-carrier group. Non-carriers were defined as non-ND-CNV carriers. ND-CNVs carried by participants included in the analysis are described in Supplementary Table 4.

Polygenic score estimation

Using GWAS summary statistics from large cohort studies of ASD[30], cognition[41] (excluding the UKB) and educational attainment (EA)[42], polygenic scores were generated using PRSice2 v2.3.0[43]. A clumping and thresholding (C+T) method was applied. Variants which passed quality control were clumped over a 250Kb window. To determine the optimum threshold, several *p*-value thresholds were calculated (*1 x 10^{-8}, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0*). A *p*-value threshold of 0.1 was selected for cognition (PS_{Cog}) and ASD PGS (PS_{ASD}) whereas a threshold of 0.05 was selected for EA PGS (PS_{EA}) (Supplementary Fig. 1.).

Rare variant annotation and filtering

Rare SNVs and indels from WES quality controlled VCF files were annotated using VEP v107[44]. A filtering pipeline was developed based on recent NDD analyses[32,45]. Synonymous and intronic variants flagged as LoF were not included in this analysis. Variants exclusive to nonsense-mediated decay transcripts were also excluded. Only variants with an allele frequency of ≤1% in the gnomAD non-Finnish European (NFE)[46] dataset and the UKB were included for analysis.

Three gene-sets were chosen to filter variants based on their association with lower cognitive outcomes and deleteriousness – autism-associated genes (https://gene.sfari.org/), ID/DD-associated genes (https://www.deciphergenomics.org/), and intolerant to LoF brain-expressed genes (https://www.proteinatlas.org/). Autism and ID/DD gene-sets were chosen given their associations with lower cognitive performance[33]. The intolerant to LoF, brain expressed gene-set was chosen as a comparative and less restrictive gene-set[11]. For a gene to be included in the autism gene-set, they must have a SFARI gene score of 1 (High-confidence) [*n* genes = 213]. To be included in the ID/DD gene-set, a gene must be classed as both definitive and associated with ID/DD (HPO term- HP:0001249) in the Developmental Disorders Genotype-to-Phenotype database (DDG2P) [*n* genes = 372]. To be included in the intolerant to LoF brain-expressed gene-set, a gene must have a normalised expression (nTPM) score ≥1 in any of 13 brain tissues according to a consensus nTPM scores from the Human Protein Atlas (HPA) and GTeX. Genes with a probability of LoF (pLI) score[47] of >0.9 or a LoF observed/expected upper-bound fraction (LOEUF) score[46] <0.37 were included in the gene-set [*n* genes = 2875].

CADD v1.6[48] scores were assigned to all variants. Non-missense variants for SFARI and DD gene-sets with a CADD score ≥20 were included in the analysis. For the LOF gene-set, non-missense variants with a CADD score ≥30 were included in the analysis. Where variants overlapped multiple genes and/or transcripts, the minimum CADD score was taken. For missense variants, MPC[49] scores were assigned via dbNSFP v.4.3a. Only missense variants with MPC ≥2 were included for analysis (Supplementary Figure 2).

Linear regression analysis

To assess the relationship between FI scores and additional variants, multiple linear regression analyses were used. Linear regression models were used similar to previous analyses[16,24] as FI, although an ordinal score, is normally distributed in the UKB. PGS generated for cognition[41], educational attainment[42], and autism[30] using PRSice2[43].were normalised by conversion to z-scores and were used as the independent variables. The effects of rare variants characterised as putatively loss of function variants in genes associated with ID/DD (DDD), ASD (SFARI), and in intolerant to LoF, brain-expressed genes (LOF) on FI were also tested. Counts of DDD, SFARI and LOF rare variants were converted to z-scores by normalising to the cohort mean and used as independent variables. Analyses were adjusted for age at time of FI assessment, sex, TDI score, employment status, educational attainment, assessment centre location, genotype batch and array, and the first 20 ancestry principal components calculated by the UK Biobank. These covariates were chosen based on previous UKB cognitive test analyses[16] and associations to cognitive outcomes. A further description of these covariates can be found in the *Supplementary Information*. PS_{EA} models did not include the educational attainment covariate to isolate the genetic effects captured by PS_{EA}, avoiding potential confounding adjustments to its biological component captured by educational attainment. Multi-collinearity was tested using a VIF of <3 and diagnostic plots were used to assess the linear model assumptions (Supplementary Tables 6–8, 12–14; Supplementary Fig. 15). A backwards elimination approach was taken to remove any non-significant covariates from the model. The adjusted R^2 values are reported. Global and local effect size estimates were calculated for linear regressions using *Cohen's f*[2][50]. Tests were corrected using Bonferroni correction. Regression analyses were carried out in all ND-CNV carriers, carriers of deletion ND-CNVs, carriers of duplication ND-CNVs, and in the top two specific ND-CNV loci (15q11.2 del. and 16p13.11 dup.) with the greatest number of carriers, and non-carriers. To assess the effect of correcting for educational attainment in the linear regression analyses, analyses were re-run without the education attainment covariate (see Supplementary Tables 10 and 15.).

Additional variant load differences

Two sample, two-sided t-tests were performed to investigate additional common variant load differences both within and between ND-CNV carriers and non-carriers. For this, PGS were normalised to the cohort mean. *Cohen's d* effect size estimates were calculated for t-tests with significant findings. Non-parametric, Wilcoxon rank sum tests were performed to investigate additional rare variant load differences both within and between ND-CNV carriers and non-carriers. A non-parametric test was selected due to non-normal distributions of rare variant load. Effect size estimates (*r*) for significant Wilcoxon rank sum test findings were calculated by dividing the Z statistic by the square root of the sample size. Sex differences in the additional variant load were also investigated within both groups. Power analysis was performed using G*Power 3.1[51] which indicated that a minimum of 32 individuals is needed to detect medium effect sizes at a power of 0.8 for *t*-tests. A minimum of 74 individuals in each group is needed to detect medium effect sizes at a power of 0.8 for Wilcoxon rank sum tests.

Positive predictive value calculation

Positive predictive values (PPVs) were used to measure how accurately different quantiles PGS predicted higher FI scores. PPVs were calculated for significantly correlated PGS in all ND-CNV carriers, 15q11.2 del. carriers, 16p13.11 dup. carriers and non-carriers. PGS were split into quantiles and FI scores were split using a cut-off of two standard deviations below the cohort mean. PPVs were calculated using formula (1).

$$PPV = \left[\frac{TP}{(TP + FP)}\right] \times 100 \quad (1)$$

where *TP* = true positive i.e. the number of individuals in a PGS quantile with an FI score two standard deviations below the cohort mean; *FP* = false

positive i.e. the number of individuals in a PGS quantile with an FI score above two standard deviations below the cohort mean. *P*-values were estimated using the binomial probability.

Data availability

The data that support the findings of this study are available from the UK Biobank. However, restrictions apply to the availability of these data, which were used under licence for the study. Data are available through the UK Biobank upon application request.

Code availability

https://github.com/TomODuinnin/UKBscripts.

Received: 3 January 2024; Accepted: 4 September 2024;
Published online: 28 September 2024

References

1. Zarrei, M. et al. A copy number variation map of the human genome. *Nat. Rev. Genet.* **16**, 172–183 (2015).
2. Coe, B. P. et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* **46**, 1063–1071 (2014).
3. Kendall, K. M. et al. Association of Rare Copy Number Variants with Risk of Depression. *JAMA Psychiatry* **76**, 818–825 (2019).
4. Woodbury-Smith, M. et al. Variable phenotype expression in a family segregating microdeletions of the NRXN1 and MBD5 autism spectrum disorder susceptibility genes. *NPJ Genom. Med* **2**, 1–8 (2017).
5. Desachy, G. et al. Increased female autosomal burden of rare copy number variants in human populations and in autism families. *Mol. Psychiatry* **20**, 170–175 (2015).
6. Jacquemont, S. et al. A higher mutational burden in females supports a 'female protective model' in neurodevelopmental disorders. *Am. J. Hum. Genet* **94**, 415–425 (2014).
7. Fitzgerald, J. et al. 'More than a box of puzzles': Understanding the parental experience of having a child with a rare genetic condition". *Eur. J. Med. Genet* **64**, 104164 (2021).
8. Dinneen, T. J. et al. How does genetic variation modify ND-CNV phenotypes? *Trends Genet* **38**, 140–151 (2022).
9. Cleynen, I. et al. Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion. *Mol. Psychiatry* **26**, 1–15 (2020).
10. Davies, R. W. et al. Using common genetic variation to examine phenotypic expression and risk prediction in 22q11.2 deletion syndrome. *Nat. Med* **26**, 1912–1918 (2020).
11. Pizzo, L. et al. Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. *Genet Med* **21**, 816–825 (2019).
12. Tansey, K. E. et al. Common alleles contribute to schizophrenia in CNV carriers. *Mol. Psychiatry* **21**, 1085–1089 (2016).
13. Bergen, S. E. et al. Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. *Am. J. Psychiatry* **176**, 29–35 (2019).
14. Moreno-De-Luca, A. et al. Developmental brain dysfunction: Revival and expansion of old concepts based on new genetic evidence. *Lancet Neurol.* **12**, 406–414 (2013).
15. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
16. Kendall, K. M. et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: Analysis of the UK Biobank. *Br. J. Psychiatry* **214**, 297–304 (2019).
17. Crawford, K. et al. Medical consequences of pathogenic CNVs in adults: Analysis of the UK Biobank. *J. Med. Genet* **56**, 131–138 (2019).
18. Owen, D. et al. Effects of pathogenic CNVs on physical traits in participants of the UK Biobank. *BMC Genomics* **19**, 1–9 (2018).
19. Mollon, J. et al. Impact of Copy Number Variants and Polygenic Risk Scores on Psychopathology in the UK Biobank. *Biol. Psychiatry* **94**, 591–600 (2023).
20. Fawns-Ritchie, C. & Deary, I. J. Reliability and validity of the UK Biobank cognitive tests. *PLoS One* **15**, e0231627 (2020).
21. Lyall, D. M. et al. Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLoS One* **11**, e0154222 (2016).
22. Satterstrom, F. K. et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568–584 (2020).
23. Wigdor, E. M. et al. The female protective effect against autism spectrum disorder. *Cell Genomics* **2**, 100134 (2022).
24. Kendall, K. M. et al. Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biol. Psychiatry* **82**, 103–110 (2017).
25. Kirov, G. et al. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol. Psychiatry* **75**, 378–385 (2014).
26. Rosenfeld, J. A. et al. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genet Med* **15**, 478–481 (2013).
27. Lövdén, M. et al. Education and Cognitive Functioning Across the Life Span. *Psychol. Sci. Public Interest* **21**, 6–41 (2020).
28. Atlay, C. et al. Instructional quality and achievement inequality: How effective is teaching in closing the social achievement gap? *Learn Instr.* **63**, 101211 (2019).
29. Clarke, T. K. et al. Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Mol. Psychiatry* **21**, 419–425 (2015).
30. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet* **51**, 431–444 (2019).
31. Kingdom, R. et al. Genetic modifiers of rare variants in monogenic developmental disorder loci. *Nat. Genet* **56**, 861–868 (2024).
32. Rolland, T. et al. Phenotypic effects of genetic variants associated with autism. *Nat. Med* **29**, 1671–1680 (2023).
33. Zhou, X. et al. Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nat. Genet* **54**, 1305–1319 (2022).
34. Abdellaoui, A. et al. Gene–environment correlations across geographic regions affect genome-wide association studies. *Nat. Genet* **54**, 1345–1354 (2022).
35. Spearman, C. 'General Intelligence,' Objectively Determined and Measured. *Am. J. Psychol.* **15**, 201 (1904).
36. Pedersen, C. B. et al. The iPSYCH2012 case-cohort sample: New directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
37. Calle Sánchez, X. et al. Comparing Copy Number Variations in a Danish Case Cohort of Individuals with Psychiatric Disorders. *JAMA Psychiatry* **79**, 59–69 (2022).
38. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742–015 (2015).
39. Krasheninina, O. et al. Open-source mapping and variant calling for large-scale NGS data from original base-quality scores. bioRxiv https://doi.org/10.1101/2020.12.15.356360 (2020).
40. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, 1–4 (2021).
41. Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 1–16 (2018).
42. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
43. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8**, (2019).
44. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 1–14 (2016).

45. Antaki, D. et al. A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. *Nat. Genet* **54**, 1284–1292 (2022).
46. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans, Genome Aggregation Database Consortium. *Nature* **581**, 434–443 (2020).
47. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
48. Rentzsch, P. et al. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886–D894 (2019).
49. Samocha, K. E. et al. Regional missense constraint improves variant deleteriousness prediction. https://doi.org/10.1101/148353 (2017).
50. Selya, A. S. et al. A practical guide to calculating Cohen's f^2, a measure of local effect size, from PROC MIXED. *Front Psychol.* **3**, 111 (2012).
51. Faul, F. et al. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res Methods* **39**, 175–191 (2007).

Acknowledgements

This research has been conducted using the UK Biobank Resource under application no. 48915. It was funded by the Innovative Medicines Initiative 2 Joint Undertaking (AIMS-2-TRIALs) under grant agreement no. 777394. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of the manuscript. We would like to acknowledge Dr Gail Davies from the University of Edinburgh who provided GWAS summary statistics for cognition.

Author contributions

T.J.D., L.G. and L.M.L. contributed to the design of this study. T.J.D, F.N.G, and C.O contributed to the bioinformatic and statistical analysis. E.H. and G.K. reviewed the manuscript and provided feedback that was incorporated into the final manuscript. All authors approved of the submission of this manuscript.

Competing Interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-024-00426-8.

Correspondence and requests for materials should be addressed to Thomas J. Dinneen.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024