Cell Genomics Transparent Peer Review Record

Summary

Initial submission:	Received : 4/19/2024
	Scientific editor: Laura Zahn
First round of review:	Number of reviewers: 2 Revision invited : 6/25/2024 Revision received : 7/30/2024
Second round of review:	Number of reviewers: 2 Accepted : 11/13/2024
Data freely available:	Yes
Code freely available:	Yes

Referees' reports, first round of review

Reviewer #1: Huguet et al, present a large cohort in which they have performed an association study on CNVs associating them to cognitive ability. For the first time, they report a CNV gain with either a protective or advantageous effect on cognitive ability, as opposed to negative effects which have been previously reported. The main limitation of the study is that it is limited to a preselected set of genes, and not performed genome wide. Never the less, there is a significant number of samples (258k) which have been combined and the analysis moves beyond common CNV loci previously associated to NDD. After reading the manuscript I have the following comments,

Major

1. The authors have combined several datasets to be able to perform the GWAS on a number of different sets. The manuscript would be improved if descriptive statistics at the start of the results include more details. Ie what portion of the genome/ number of genes were finally included in the analysis. Likewise the frequency of CNVs, and the copy number range observed for the CNV regions. These detail are essentially present, but scattered through the text.
2. Figure 1C, can the authors explain the small overlap between the difference cohorts?
3. The analysis presented here, reports on the association of a single loci to cognitive burden, it is equally of interested to consider the genome wide CNV burden of individuals. Is this a covariate that can be incorporated.
4. Generally speaking the manuscript is written in a style which assumes the reader is already familiar with the analysis methods used. Readability for non-experts could be improved.
5. My under standing from the text is that 36 out of 258k individuals have the duplication of 2q12.3 (i.e. approximately 0.01% AF). It would be nice to see supplementary figure 2 in the main text so that readers can also see the effect size. Following on from this can the authors reflect on the effect of this gain in the discussion on the cognitive scores?
6. The effect size of del 16p12.2 proximal has a much broader range than the remainder of the loci, please expand the discussion and provide insights, is this a technical reason, or underlying biology.
7. At present the manuscript focuses on CNVs which result in a gene dosage effect on cognitive ability. Is it possible to negate the logic, and report on CNVs which are clearly not dosage sensitive?

Minor
1. Abstract, this ends very abruptly, and I would recommend adding a final concluding sentence.
2. Methods p5 Gene Annotation, please double check the GenCode version, this seems like a typo
3. Results
"Among 18,451 autosomal coding genes with LOEUF values, 35% and 64.9% were fully encompassed in one
or more deletion and duplication, respectively (75% across CNVs), with 40% observed in
both deletions and duplications (Figure 1A, B, C)."
The later part of this sentence is difficult to read. What does "respectively" refer to? likewise for the 40%?
3. Figure 2C, "older participants" is non-specific please provide details.
4. p9 Results, please check text "moderately (LOEUF [0.35, 1[)"
5. Please specify clearly which reference genome version is used, for example coordinates in mention in figure 1.
6. Figure 5 caption, please check text "moderately intolerant [0.35, 1.0[: orange"

Reviewer #2: The authors performed genome-wide association studies for the association between rare CNVs from in 258,292 individuals and cognitive ability. Likewise, they associated gene sets as defined by tissue and cell-type expression and gene ontologies affecting CNVs and cognitive ability.

This is a highly interesting manuscript that identifies a CNV that is associated with higher cognitive ability. This is to my knowledge the first identification of a CNV that increases (instead of decreases) cognitive ability and is thus a highly interesting result in its own right. In addition, the authors identify an interesting negative correlation between the deletions and duplication present in defined gene sets and their effect on cognitive ability which suggests that some genes (involved in the same biological entities) may have opposite effects based on dosage.
They also define a new trait-associated gene-dosage sensitivity score (tagDS), a normalized value reflecting whether a gene-set shows preferentially effects on cognitive ability when either deleted or duplicated.

The analyses seem well performed and well thought-through. Nevertheless, this reviewer sits with the feeling that the message becomes somewhat convoluted and drown in focus on numbers and complicated analyses and figures and less so on explanations. In addition, here and there, the language could have been more precise (some examples below). I do understand that the word counts set limitations for what can be explained but would still encourage it for broader readability.

E.g. some sentences in the introduction on background for and the development itself of tagDS would be helpful (instead of hiding the majority of the explanation in the (supplementary) methods) - also to provide a better red thread from the introduction to the discussion. A few sentences on the definition and use of gene sets might also be helpful.

Out of curiosity - it would be interesting to clarify whether there are any deletions of the 2q12.3 region in this population and if there is any reason to believe that may affect cognitive ability negatively or positively [not necessary to add to the manuscript].
Once these things have been adjusted (clarification of language, methods and terms), this manuscript is ready for publication.

Unclear in methods:
Page 5: 'We excluded from the analyses all individuals carrying a CNV ≥10Mb or a mosaic CNV. '
-unclear how mosaicism was evaluated… In addition, it would be good to specifify why the cut-off of >10Mb for large CNVs was made.

Unclear how overlap in UKB samples (e.g. between g-factor and fluid intelligence) were dealt with…
Also unclear how the overlapping CNVs were evaluated - were some of them confirmed by visual inspection of Log R & BAF-plots?

«We computed a linear regression model (gene-level GWAS, Supplementary method statistical model 1) on general cognitive ability for 241 and 596 genes covered by at least 30 deletions or duplications, respectively, in 258,292

individuals from the general population»
I did not find anywhere why the cut-off of 30 was chosen - please clarify.
Likewise, perhaps the sentence should be rewritten [e.g.. From 258,292 individuals, we identified 241 genes covered by at least 30 deletions and 596 covered by at least 30 duplications. For each gene, we performed a gene-level GWAS on cognitive ability].

Clarification needed in Supplementary methods:
PLINK - please write with capitel letters and provide the version - there are several versions out there.

Based on the methods, it is unclear whether the CNVs called for each study are comparable across these genotyping arrays. Would lack of calling some CNVs for some chips introduce bias?

"From a total of 488,377 people with genotypic data, 28,522 were excluded for failing all these filters."

All these filters - or at least one of these filters?

Better legends in supplementary tables - e.g. abbreviations are not written out.

Clarifications needed:
Abstract: "We identified a novel duplication at 2q12.3 associated with higher performance"
-perhaps more correct to say: "We identified a novel association of a duplication at 2q12.3 duplication with cognitive ability" [the duplication is not novel in itself, the association is…]

«The routine implementation of whole genome CNV detection, as a first-tier diagnostic test, identifies "pathogenic" CNVs in 10 to 15 % of children with neurodevelopmental disorders (NDD)14. "
This statement seems slightly incorrect given that the article cited include autism, developmental disabilities and multiple congenital anomalies,developmental delay. ADHD is also a neurodevelopmental disorder -

perhaps be specific instead of writing neurodevelopmental disorders in general.

Page 3 " But beyond the benign vs pathogenic categorical classification of genomic variants, their effect size on cognitive ability has been used to provide more nuanced information on the severity of a variant and to quantify the risk for NDD."
Putative pathogenic?

Page 3: 'Genetic fitness'
-what is that?

Page 4: "We then performed functional-burden associations with cognitive ability across 6,502 gene-sets corresponding to gene functions at the tissue, cell type, and molecular levels."
-what does that mean?

Page 4: 'that most biological functions have preferential effects on cognitive ability when either deleted or duplicated.'
-this sentence is unclear
Page 4: 'As a result, we observed a negative correlation between the effects of deletions and duplications across all levels of biological observation independently of intolerance to haploinsufficiency. '
Also unclear what this means…

Figure 1D - please write the chromosomal position of the duplication with the positive effect on the figure - it is easy to misunderstand and believe that it is a 2q13 duplication (and not the 2q12.3 duplication as noted in the text).

Page 7: «We identified 9 deletions encompassing a total of 69 genes and 10 duplications encompassing a total of 123 genes with previously published negative effects (Supplementary table 4) that persisted when we conducted a meta-analysis across 9 sub-cohorts defined by cognitive assessments (Table 1, Supplementary figure 2).»
-but looking at the figures (Supplementary figure 2 & Figure 1D) I only find 6 deletions and 7 duplications (not 9 & 10) - what is the explanation for this?

Figure 1E - it would have been good to show the overlap of all the 36 2q12.3

duplications with the region of interest (e.g. do they all have the same breakpoints or are they similar to NRXN1 with different breakpoints).
«We partitioned genes into 38 sets based on overlapping LOEUF categories»
-it is difficult to find (i.e. I did not when first reading) based on which criteria (i.e. LOEUF of <0.35, LOEUF 0-35-0.9 etc), these genes sets were made. Could you please clarify - perhaps refer to figure 2B earlier (since this seems to outline and is a bit more intuitive to understand).
Page 8: "We first focused on gene sets assigned to 215 adult brain regions.»
-unless one goes back to the methods, it is impossible to know what is meant here. Could a bit more detail be provided in the results (I understand there are word-restrictions) but this sentence on its own does not tell me anything.
Page 8: "This suggests that cognitive ability is preferentially affected by one CNV type, depending on the brain region.»
-would be easier to write if it read 'either deletion or duplciation' instead of 'by one CNV type'.
Page 8: "TagDS indicated that cerebral cortex gene-sets affected cognitive ability preferentially when duplicated, while th"
-again, if one had not read the methods, it is difficult to read. Please provide more explanation.
Figure 8: "asked if CNVs affecting genes assigned to non-brain tissues were also associated with cognitive ability»
-write 'not expressed in brain' ('non-brain tissues') instead of non-brain tissues alone to explain better.
Page 8: "Genes involved in non-brain tissues affect cognitive ability."
-perhaps write 'expressed in' in

Authors' response to the first round of review

Reviewer #1: Huguet et al, present a large cohort in which they have performed an association study on CNVs associating them to cognitive ability. For the first time, they report a CNV gain with either a protective or advantageous effect on cognitive ability, as opposed to negative effects which have been previously reported. The main limitation of the study is that it is limited to a preselected set of genes, and not performed genome wide. Never the less, there is a significant number of samples (258k) which have been combined and the analysis moves beyond common CNV loci previously associated to NDD. After reading the manuscript I have the following comments,

Major:

1. The authors have combined several datasets to be able to perform the GWAS on a number of different sets. The manuscript would be improved if descriptive statistics at the start of the results include more details. Ie what portion of the genome/ number of genes were finally included in the analysis. Likewise the frequency of CNVs, and the copy number range observed for the CNV regions. These detail are essentially present, but scattered through the text.

Response: We thank the reviewer for this comment and modified the first section of the results as follows to provide more descriptive statistics:

"Among the 258,292 individuals from general population datasets, 15.6% carried at least one rare (allele frequency <1%) autosomal CNV larger than 50Kb, fully encompassing one or more coding genes (hg19). Among all autosomal coding genes with LOEUF values (n=18,451), 71.8% were fully encompassed in one or more CNVs: 35% in deletions: 64.9% in duplication, and 28.1% in both deletions and duplications (Figure 1A, B, C). Most of the genes encompassed in CNVs were contained in ultra-rare CNVs (<1/10,000) with fewer than 30 carriers (Figure 1C). We used a linear regression model (gene-level GWAS, c.f. methods statistical model 1) to test the association of general cognitive ability with 241 and 596 genes covered by at least 30 deletions or duplications, respectively (Figure 1D, 1E). We identified 6 deletions encompassing a total of 68 genes and 7 duplications encompassing a total of 122 genes with previously published negative effects (Supplementary table 4) that persisted when we conducted a meta-analysis across 9 sub-cohorts defined by cognitive assessments (Table 1, Figure 1E). We identified a novel association between a duplication at 2q12.3, and positive effects (z=0.434, p=7.58x10-03) on cognitive ability (Figure 1F, Supplementary figure 2). This duplication observed in 36 individuals included 4 non-intolerant genes with a LOEUF≥0.35 (EDAR, SH3RF3, SEPT10, SOWAHC) and was observed at a similar frequency (1 to 2 in 10,000) across cohorts (Fisher's exact p-valueFDR >0.05). Results were not related to ancestry, array platform, or cognitive assessment methods (Figure 1G). The positive effect remained significant when comparing 2q12.3 duplication carriers to individuals without any CNVs. The reciprocal deletion in this region showed a trend towards a negative effect on cognitive ability (z = -0.526, p = 0.058) but we were underpowered with only 12 carriers. Additionally, the a gene-dosage model showed a positive effect of 0.415 z-scores (p=2.65x10-03) of cognitive ability per copy number of copies (1, 2 or 3) at this locus. In other words, this may possibly represent the first locus with a mirror impact on cognitive ability."

2. Figure 1C, can the authors explain the small overlap between the different

cohorts?

Response: We apologize for the confusion. Figure 1C does not represent the overlap between cohorts. Instead, it represents the overlap between the gene content of ultra-rare and rare deletions and duplications.

To improve clarity, we changed the legend:

"(C) Venn diagram illustrating the overlap between gene content of ultra-rare and rare deletions and duplications. Specifically, the number of genes deleted and-or duplicated at least once in these CNVs from the pooled general population datasets."

3. The analysis presented here, reports on the association of a single loci to cognitive burden, it is equally of interest to consider the genome wide CNV burden of individuals. Is this a covariate that can be incorporated.

Response: This is an interesting question and analysis that we have previously considered. However, among the 15% of individuals with a coding CNV, only 2% had a second coding CNV. Therefore, it is highly improbable that adjusting for an additional CNV would change our results.

4. Generally speaking the manuscript is written in a style which assumes the reader is already familiar with the analysis methods used. Readability for non-experts could be improved.

Response: We agree with the reviewer that the manuscript was technical. We therefore improved its readability for non-experts.

5. My under standing from the text is that 36 out of 258k individuals have the duplication of 2q12.3 (i.e. approximately 0.01% AF). It would be nice to see supplementary figure 2 in the main text so that readers can also see the effect size. Following on from this can the authors reflect on the effect of this gain in the discussion on the cognitive scores?

Response: We had initially included supplementary figure 2 in the main text but removed it due to space. The new figure is presented below:

We also added additional information on the effect size of this gain in the results and the discussion:

1- "We identified a novel association between a duplication at 2q12.3, and positive effects (z=0.434, p=7.58x10-03) on cognitive ability (Figure 1E)."

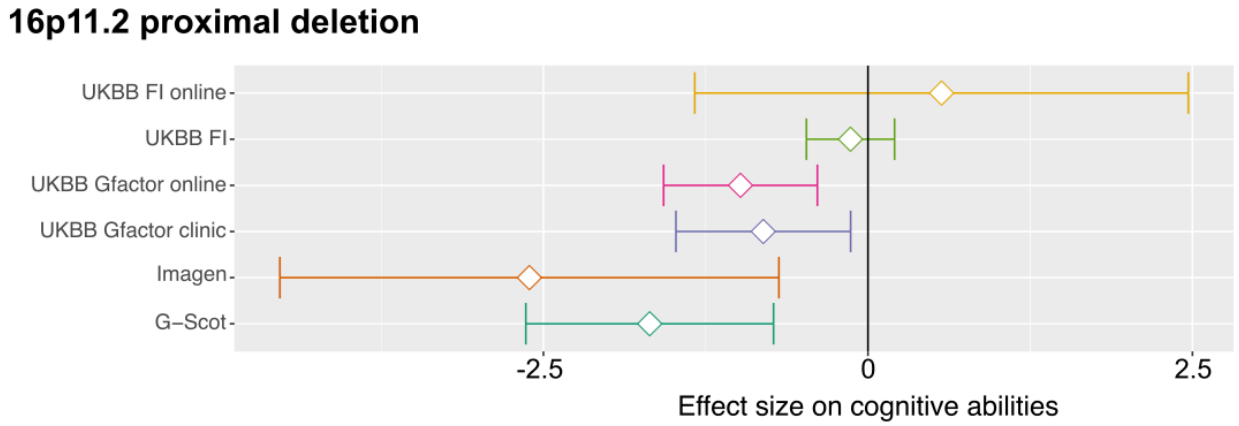
2- "effect size (z=0.434, equivalent to 6.5 points of IQ) on cognitive ability without significant heterogeneity across cohorts."

6. The effect size of del 16p12.2 proximal has a much broader range than the remainder of the loci, please expand the discussion and provide insights, is this a

technical reason, or underlying biology.

Response: The reviewer is correct to point out this confidence interval. This is due to the heterogeneity of effect sizes estimated for this CNV across the different cohorts, leading to a broader confidence interval provided by the meta-analysis. Specifically, this CNV had a smaller effect size in UKBB. This has also been previously reported in Kendall et al. 2019. We do not believe that there is a biological explanation for this result. In other words, it is unlikely that this CNV is more variable in its effect. One possible interpretation is that UKBB ascertained carriers of this CNV with minimal symptoms.

The heterogeneity for this CNV is now well visible in this figure (it can be added in the article):



7. At present the manuscript focuses on CNVs which result in a gene dosage effect on cognitive ability. Is it possible to negate the logic, and report on CNVs which are clearly not dosage sensitive?

Response: This is a very interesting proposal. However, absence of proof is not proof of absence. In other words, among the 241 and 596 genes covered by at least 30 deletions or duplications tested in this analysis, 173 or 474 did not show significant effects on cognitive ability. But this number may change drastically if a much larger sample size was available for analysis. Therfore, it is unclear which of those are truely not sensitive to gene dosage. Future larger studies will be able to answer this question.

Minor

1. Abstract, this ends very abruptly, and I would recommend adding a final concluding sentence.

Response : We agree with the reviewer and added a concluding sentence (highlighted in yellow) :

"…..We identified a novel duplication at 2q12.3 associated with higher performance

in cognitive ability. Functional burden tests performed on 6,502 gene sets identified 864 gene-sets associated with cognition. The effects-sizes of deletion and duplication were negatively correlated, suggesting that functions across all levels of biological observations were either sensitive to deletion (e.g. subcortical and postsynaptic functions.) or duplication (e.g. cerebral cortex and presynaptic functions). Associations between non-brain tissue and cognition, were driven in part by genes under constraint with lower tissue specificity. The latter may help investigate medical comorbidities observed in neurodevelopmental and psychiatric disorders."

2. Methods p5 Gene Annotation, please double check the GenCode version, this seems like a typo

Response : We double-checked the GENCODE version, and it is correct (v.19). We corrected the way GENCODE is written (all capitals).

3. Results

"Among 18,451 autosomal coding genes with LOEUF values, 35% and 64.9% were fully encompassed in one or more deletion and duplication, respectively (75% across CNVs), with 40% observed in both deletions and duplications (Figure 1A, B, C)."

The later part of this sentence is difficult to read. What does "respectively" refer to? likewise for the 40%?

Response: We agree with the reviewer and modified this section to improve clarity:

" Among all autosomal coding genes with LOEUF values (n=18,451), 71.8% were fully encompassed in one or more CNVs: 35% in deletions: 64.9% in duplication, and 28.1% in both deletions and duplications (Figure 1A, B, C). "

3. Figure 2C, "older participants" is non-specific please provide details.

Response: We agree with the reviewer and specified the criteria in the legend of Figure 2C : "older participants (≥60 or ≥70 years old)"

4. p9 Results, please check text "moderately (LOEUF [0.35, 1[)"

Response: We clarified this sentence by adding additional information on these 2 levels of intolerance to haploinsufficiency:

"moderately intolerant to haploinsufficiency(LOEUF = [0.35, 1[) and highly intolerant to haploinsufficiency (LOEUF<0.35; Supplementary figure 9)."

5. Please specify clearly which reference genome version is used, for example coordinates in mention in figure 1.

Response: We used hg19 throughout the manuscript. We added this information in the first paragraph of the results and Figure 1.

6. Figure 5 caption, please check text "moderately intolerant [0.35, 1.0[: orange"

Response: We clarified this sentence by adding additional information on these 2 levels of intolerance to haploinsufficiency:
"moderately intolerant to haploinsufficiency(LOEUF = [0.35, 1[) and highly intolerant to haploinsufficiency (LOEUF<0.35; Supplementary figure 9)."

Reviewer #2: The authors performed genome-wide association studies for the association between rare CNVs from in 258,292 individuals and cognitive ability. Likewise, they associated gene sets as defined by tissue and cell-type expression and gene ontologies affecting CNVs and cognitive ability. This is a highly interesting manuscript that identifies a CNV that is associated with higher cognitive ability. This is to my knowledge the first identification of a CNV that increases (instead of decreases) cognitive ability and is thus a highly interesting result in its own right. In addition, the authors identify an interesting negative correlation between the deletions and duplication present in defined gene sets and their effect on cognitive ability which suggests that some genes (involved in the same biological entities) may have opposite effects based on dosage.
They also define a new trait-associated gene-dosage sensitivity score (tagDS), a normalized value reflecting whether a gene-set shows preferentially effects on cognitive ability when either deleted or duplicated.
Response: We thank the reviewer for his/her kind words and his/her appreciation of our work.
The analyses seem well performed and well thought-through. Nevertheless, this reviewer sits with the feeling that the message becomes somewhat convoluted and drown in focus on numbers and complicated analyses and figures and less so on explanations. In addition, here and there, the language could have been more precise (some examples below). I do understand that the word counts set limitations for what can be explained but would still encourage it for broader readability.
Response: We have clarified the message and added a narrative thread throughout the results.
1) E.g. some sentences in the introduction on background for and the development itself of tagDS would be helpful (instead of hiding the majority of the explanation in the (supplementary) methods) - also to provide a better red thread from the introduction to the discussion. A few sentences on the definition and use of gene sets might also be helpful.
Response: We have clarified the message and added a narrative thread throughout the results: "Previous publications have reported that effect size of gene dosage on

cognitive ability is U-shaped28, with the effects of deletions being 2 to 3-fold higher than those of duplications.11,12 Studies, however, have not been able to test whether genes show preferential effects on cognitive ability when either deleted or duplicated. We developed the trait-associated gene dosage sensitivity score (tagDS), to test whether the deletion/duplication effect size ratio of a given gene-set deviates from the null distribution (average ratio of 2.4 in our dataset, c.f. methods). This normalized value reflects preferential sensitivity to deletions or duplications for a specific phenotype. Positive or negative tagDS depicted ratio of effect sizes between deletion and duplication biased toward deletions or duplications, respectively (c.f. methods, Figure 3D)."

2) Out of curiosity - it would be interesting to clarify whether there are any deletions of the 2q12.3 region in this population and if there is any reason to believe that may affect cognitive ability negatively or positively [not necessary to add to the manuscript].

Response: This is a very interesting comment that we previously explored, but unfortunately, there aren't enough carriers (n=8, below the n=30 threshold used for our analyses) to analyze the 2q12.3 deletion. We added the following sentence in the results : "....There were only 12 carriers (< 30 carriers threshold) of the reciprocal 2q12.3 deletion which did not provide enough power for any relevant analysis. …"

Once these things have been adjusted (clarification of language, methods and terms), this manuscript is ready for publication.

Response: We thank the reviewer for his/her appreciation of our work.

3) Unclear in methods:

Page 5: 'We excluded from the analyses all individuals carrying a CNV ≥10Mb or a mosaic CNVs. '

-unclear how mosaicism was evaluated… In addition, it would be good to specifify why the cut-off of >10Mb for large CNVs was made.

Response: We apologize for this confusing sentence. We did not systematically call mosaic CNVs in our dataset. We did exclude CNVs ≥ 10Mb (a widely used threshold in the QC if CNVs REF) because very large CNVs are rarely observed in general population cohorts and are almost always present as mosaics/somatic CNVs that can't be pooled with germline CNVs. This was indeed the case since we manually QCed all CNVs >10Mb.

We clarified this point in the methods : "We did exclude CNVs ≥ 10Mb (a widely used threshold in the QC if CNVs) because very large CNVs are rarely observed in general population cohorts and are almost always present as mosaics and/or

somatic CNVs that can't be pooled with germline CNVs."

4) Unclear how overlap in UKB samples (e.g. between g-factor and fluid intelligence) were dealt with…

Response: Thanks for your note. To deal with participants who had 2 measures of cognitive ability, we selected only one measure. We added the following sentence in the methods to clarify the selection.

Sentence: "In UKBB, some individuals had multiple cognitive ability assessments. For those individuals we selected the most robust cognitive evaluations based on the following ranking (from the most to the least robust) : 1) in-person g-factor, 2) online g-factor, 3) in-person FI 4) online FI."

5) Also unclear how the overlapping CNVs were evaluated - were some of them confirmed by visual inspection of Log R & BAF-plots?

Response: If we correctly understand the reviewer's question, it refers to the level of overlap between CNVs called by both PennCNV and quantiSNP. Our QC algorithm relies on this information and was trained as pointed by the reviewer on CNVs manually inspected using the Log R & BAF-plots.

6) In addition, recurrent CNVs were all manually inspected using those 2 metrics.

We modified two parts of the method as follows: ".... This model was trained and tested respectively on 66% and 33% of 34,156 CNVs (31,746 true CNVs and 2,410 artefacts from 6 cohorts, excluding SPARK), This reference CNV set was manually inspected with Log R and BAF plots, by two CNV experts. DigCNV showed an AUC = 0.95, a sensitivity of 0.95 and a specificity of 0.85. .."

For recurrent CNV, we added the following details : "Every recurrent CNV was annotated (based on previously published methods12) and manually visualized (Log R and BAF-plots)."

«We computed a linear regression model (gene-level GWAS, Supplementary method statistical model 1) on general cognitive ability for 241 and 596 genes covered by at least 30 deletions or duplications, respectively, in 258,292 individuals from the general population»

7) I did not find anywhere why the cut-off of 30 was chosen - please clarify. Likewise, perhaps the sentence should be rewritten [e.g.. From 258,292 individuals, we identified 241 genes covered by at least 30 deletions and 596 covered by at least 30 duplications. For each gene, we performed a gene-level GWAS on cognitive ability].

Response: We used a cut-off of 30 carriers to obtain a power of 85% to detect

CNVs with a large effect size equivalent to Cohen's d=0.7 (alpha=0.005). We clarified the rationale for this cut-off in the methods: "....We used a cut-off of 30 carriers to obtain a power of 85% to be able to detect CNVs with large effect size equivalent to Cohen's d=0.7 (alpha=0.005)."

8) Clarification needed in Supplementary methods:

PLINK - please write with capital letters and provide the version - there are several versions out there.

Response: PLINK is now in uppercase across the manuscript, and we added the version (version 1.9).

9) Based on the methods, it is unclear whether the CNVs called for each study are comparable across these genotyping arrays. Would lack of calling some CNVs for some chips introduce bias?

Response: We used a previously published method (ref) to exclude CNVs that could not be properly detected by all genotyping platforms used across this aggregated dataset. Only CNVs covered by >=10 probes for each technology were used for analysis. We added this information in the methods section : "....We used a previously published method (ref) to exclude CNVs that could not be properly detected by all genotyping platforms used across this aggregated dataset. Only CNVs covered by >=10 probes for each technology were used for analysis. ....."

10)"From a total of 488,377 people with genotypic data, 28,522 were excluded for failing all these filters." All these filters - or at least one of these filters?

Response: This was indeed unclear. This should read as: "..From a total of 488,377 people with genotypic data, 28,522 were excluded for failing only of these filters." We have corrected this in the methods.

11) Better legends in supplementary tables - e.g. abbreviations are not written out.

Response: We fully spelled out all of the abbreviations across all figures.

12)Clarifications needed:

Abstract: "We identified a novel duplication at 2q12.3 associated with higher performance" - perhaps more correct to say: "We identified a novel association of a duplication at 2q12.3 duplication with cognitive ability" [the duplication is not novel in itself, the association is…]

Response: We agree with this modification. It now reads as follows in the abstract: "We identified a novel association of a duplication at 2q12.3 with higher performance in cognitive ability."

13)«The routine implementation of whole genome CNV detection, as a first-tier diagnostic test, identifies "pathogenic" CNVs in 10 to 15 % of children with neurodevelopmental disorders (NDD)14. " This statement seems slightly incorrect given that the article cited include autism, developmental disabilities and multiple congenital anomalies, developmental delay. ADHD is also a neurodevelopmental disorder - perhaps be specific instead of writing neurodevelopmental disorders in general.

Response: We agree with the reviewer that CNV yield in the clinic is highly dependent on the clinical presentation. To avoid the complexity of linking the yield with multiple corresponding clinical presentations, we removed the exact percentage from the sentence. IT now reads as follows : "... Whole genome CNV detection is a first-tier diagnostic test routinely implemented in children referred to the clinic for neurodevelopmental disorders (NDD)14. "

14)Page 3 " But beyond the benign vs pathogenic categorical classification of genomic variants, their effect size on cognitive ability has been used to provide more nuanced information on the severity of a variant and to quantify the risk for NDD." Putative pathogenic?

Response: We agree with including the notion of "Putative". This was added in the introduction.

15)Page 3: 'Genetic fitness' - what is that?

Response: To avoid any confusion, we have changed the term "genetic fitness" to "genetic constraint", which is used across the manuscript.

16)Page 4: "We then performed functional-burden associations with cognitive ability across 6,502 gene-sets corresponding to gene functions at the tissue, cell type, and molecular levels." - what does that mean?

17)Page 4: 'that most biological functions have preferential effects on cognitive ability when either deleted or duplicated.' - this sentence is unclear

18)Page 4: 'As a result, we observed a negative correlation between the effects of deletions and duplications across all levels of biological observation independently of intolerance to haploinsufficiency. ' Also unclear what this means…

Response: We modified the entire section corresponding to comments 16, 17, and 18. We added more information on the functional burden test in the introduction: "To further investigate CNVs too rare to be tested by the CNV-level GWAS, we performed functional-burden analyses. To do so, we aggregated all CNVs disrupting a group of genes assigned to a given biological function. Functional-burden

associations were performed between cognitive ability and 6,502 gene-sets assigned to biological functions at the tissue, cell type, and molecular levels. Functional burden tests revealed that most functional gene sets were associated with cognitive ability when either deleted or duplicated, and only a few gene-sets showed significant associations with cognition for both CNVs. As a result, we observed a negative correlation between the effect sizes of deletions and duplications across all functional gene-sets and this was not influenced by intolerance to haploinsufficiency. This suggests that the effects of most biological functions on cognitive ability is dependent on the type of gene dosage."

19)Figure 1D - please write the chromosomal position of the duplication with the positive effect on the figure - it is easy to misunderstand and believe that it is a 2q13 duplication (and not the 2q12.3 duplication as noted in the text).

Response: We added more information in the figure 1.

20)Page 7: «We identified 9 deletions encompassing a total of 69 genes and 10 duplications encompassing a total of 123 genes with previously published negative effects (Supplementary table 4) that persisted when we conducted a meta-analysis across 9 sub-cohorts defined by cognitive assessments (Table 1, Supplementary figure 2).»

-but looking at the figures (Supplementary figure 2 & Figure 1D) I only find 6 deletions and 7 duplications (not 9 & 10) - what is the explanation for this?

Response: Thank you for highlighting this mistake. We effectively identified 6 deletions and 7 duplications. So we corrected this typo.

21)Figure 1E - it would have been good to show the overlap of all the 36 2q12.3 duplications with the region of interest (e.g. do they all have the same breakpoints or are they similar to NRXN1 with different breakpoints).

Response: The new figure below was added as supplemental figure 2.

22) «We partitioned genes into 38 sets based on overlapping LOEUF categories»

-it is difficult to find (i.e. I did not when first reading) based on which criteria (i.e. LOEUF of <0.35, LOEUF 0-35-0.9 etc), these genes sets were made. Could you please clarify - perhaps refer to figure 2B earlier (since this seems to outline and is a bit more intuitive to understand).

Response: We added further information on how we defined these gene sets :

"We created 38 overlapping gene categories by sliding a window (defined by a width of 0.15 LOEUF units) by 0.05 LOEUF units 37 times (figure 2B, methods, statistical model 2)"

23)Page 8: "We first focused on gene sets assigned to 215 adult brain regions.»

-unless one goes back to the methods, it is impossible to know what is meant here. Could a bit more detail be provided in the results (I understand there are word-restrictions) but this sentence on its own does not tell me anything.

Response: We agree with the reviewer and therefore added additional information on how we assigned the genes to brain regions in the results: “We tested 215 gene sets assigned to 215 adult brain regions. To define gene sets, we first normalized (z-scored) the expression of each gene across all 215 tissues. For each tissue, the corresponding gene set was defined based on relative over expression by selecting all genes with an expression z score >= 1 in that tissue. ”

24)Page 8: "This suggests that cognitive ability is preferentially affected by one CNV type, depending on the brain region.» - would be easier to write if it read 'either deletion or duplciation' instead of 'by one CNV type'.

Response: Thanks for the suggestion. We rewrote this part more explicitly: “This suggests that genes assigned to brain regions affect cognitive ability, when either deleted or duplicated.”

25)Page 8: "TagDS indicated that cerebral cortex gene-sets affected cognitive ability preferentially when duplicated, while th"-again, if one had not read the methods, it is difficult to read. Please provide more explanation.

Response: The reviewer pointed out with reason that most of the information to understand tagDS was hidden in the method. We now moved this to the main results to help with the narrative :

“Previous publications have reported that effect size of gene dosage on cognitive ability is U-shaped28, with the effects of deletions being 2 to 3-fold higher than those of duplications11,12, but studies have not been able to test whether genes show preferential effects on cognitive ability when either deleted or duplicated. We developed the trait-associated gene dosage sensitivity score (tagDS), to test whether the deletion/duplication effect size ratios of the 215 brain gene-sets deviate from the null distribution (average ratio of 2.4 in our dataset, c.f. methods). This normalized value reflects preferential sensitivity to deletions or duplications for a specific phenotype. Positive and negative tagDS depicted ratio of effect sizes between deletion and duplication biased toward deletions and duplications, respectively (c.f. methods, Figure 3D). ”

26) Page 8: "asked if CNVs affecting genes assigned to non-brain tissues were also associated with cognitive ability»

-write 'not expressed in brain' ('non-brain tissues') instead of non-brain tissues alone to explain better.

Response: We clarified in the results how these genes assigned to non-brain tissues were defined: “We, therefore, asked if CNVs affecting genes preferentially expressed in non-brain tissues (not part of the nervous central system) were also associated with cognitive ability. We used 37 gene-sets, defined by relative expression in 37 whole-body tissues(37 whole-body tissues(12 brain and 25 non-brain tissues (≥1SD, Figure 4A).”

27)Page 8: "Genes involved in non-brain tissues affect cognitive ability."
-perhaps write 'expressed in' in

Response: We clarified in the results how these genes assigned to non-brain tissues were defined:

Referees’ reports, second round of review

Reviewer #1: I am satisfied with the updates to the manuscript.

Reviewer #2: The authors did a thorough job of addressing the questions the reviewers raised. I have no more comments and recommend publication of this highly interesting manuscript that identifies a CNV that is associated with higher cognitive ability [also see previous review].

Authors’ response to the second round of review

none
