nature portfolio

Peer Review Information

Journal: Nature Genetics
Manuscript Title: Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention deficit hyperactivity disorder
Corresponding author name(s): Dr Ditte Demontis

Reviewer Comments & Decisions:

Decision Letter, initial version:

13th Sep 2021

Dear Dr Demontis,

Your Article, "Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention deficit hyperactivity disorder" has now been seen by 3 referees. You will see from their comments below that while they find your work of interest, some important points are raised. We are interested in the possibility of publishing your study in Nature Genetics, but would like to consider your response to these concerns in the form of a revised manuscript before we make a final decision on publication.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, with a view to identifying key priorities that should be addressed in revision. In this case, all referees have identified aspects of the analyses and the presentation that need to be improved. We particularly ask that you address the technical issues, perform additional analyses, and adjust some of the claims to address all referees' points as thoroughly as possible. We hope that you will find the prioritized set of referee points to be useful when revising your study.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file. At this stage we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

*1) Include a "Response to referees" document detailing, point-by-point, how you addressed each

referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions, available
here.
Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf
It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.
A revised checklist is essential for re-review of the paper.

Please be aware of our guidelines on digital image standards.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within three to six months. If you cannot send it within this time, please let us know.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Sincerely,

Wei

Wei Li, PhD
Senior Editor
Nature Genetics
New York, NY 10004, USA
www.nature.com/ng

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
This study seems to be well executed. It is a GWAS of ADHD subtypes (childhood, persistent, adult). While few genome-wide significant loci were identified, informative genetic overlaps were calculated between various phenotypes. Fig. 2 is the most important. An exome-sequencing analysis showed that rare protein-truncating variants are more enriched in childhood and persistent ADHD, less so in late-diagnosed (although it is not clear to me what the dependent variable was in this analysis).

Although professionally done, this study seems to add very little beyond recent papers (Demontis et al., 2019; Satterstrom et al., 2019; Rovira et al., 2020) and therefore I would rate its significance below what is required for Nature Genetics. But this is a matter for the editor to decide.

The writing in the paper is sometimes awkward or malformed, particularly in the Discussion and Methods. For example, the word "however" is hardly ever used grammatically. Perhaps the journal can be counted on to clean up these things, but nevertheless the authors may wish to go over the manuscript carefully.

A general comment is that many of the PGS analyses do not seem to add anything beyond the genetic correlations. In many cases the authors should consider dropping the redundant showing of one phenotype's PGS significantly predicting another phenotype with which it is genetically correlated.

Here are some more specific comments.

abstract: Much of this is confusing and possibly circular sounding to someone who has not read the paper. What is a "higher polygenic score (PGS) of ADHD risk variants"? Since three GWAS of ADHD subtypes have been performed, there are three possibilities. A higher average in one subgroup of cases might just mean that the PGS was constructed from the GWAS of that group. It seems absurd to write out the truism that "a later diagnosis of ADHD could be due in part to genetic factors." "beta" will suggest regression coefficient to most readers, but its meaning in this context will be obscure (see below).

p. 4: "a recent study found individuals with late-diagnosed ADHD to have a burden of common ADHD risk alleles at a level comparable to individuals without ADHD" Statements like this should always make it clear which type of cases were used in the GWAS used to construct the PGS.

p. 7: I do not understand how the PGS was constructed. I can see how 5-fold cross validation can produce a prediction R^2 that is not inflated by overfitting, but not how it can lead to PGS weights for prediction of the phenotype in subsets of the training sample itself. Perhaps the problem lies in my own ignorance, but regardless I would like the procedure to be explained better in the Methods. Also, I cannot believe that the "SE" in the parentheses can really mean "standard error." If so, then none of the PGS means is significantly greater than zero! Perhaps it should instead be "SD" for "standard deviation."

p. 11: "We only identified significant enrichment in rPTVs in childhood and persistent ADHD in highly constrained genes (Pchildhood_ADHD=2.41xa10-11, Ppersistent_ADHD = 1.90x10-3)." This sentence is confusing. Is it merely giving P-values corresponding to estimates and standard errors given earlier in the paragraph?

p. 13: "The same pattern was observed in the PGS analyses where we found a higher PGS for inattention in childhood and persistent ADHD than in late-diagnosed ADHD, however pair-wise comparisons only revealed nominally significant higher PGS for hyperactivity in childhood ADHD compared to late-diagnosed ADHD." I do not understand what this run-on sentence means.

p. 18: There are some minor differences between this GWAS of ADHD and the one reported in 2019. If I am understanding correctly, the previous one used dosages and this one uses best-guess discrete genotypes. I am sure this makes no practical difference, but I am wondering why the change? Perhaps relatedly, the previous GWAS had more than 8 million SNPs and this one fewer than 6 million. What accounts for the decrease?

p. 21: "as no sample overlap and population stratification was present." Besides being ungrammatical, this seems unjustified. How can you be sure there is no population stratification? If the authors want to constrain the intercept in order to reduce the standard error, they should say instead that they are *assuming* the absence of stratification.

p. 23: I regrettably do not understand how the rare-variant enrichment analysis was performed. This should be explained more explicitly, perhaps with an equation.

p. 24: "new P-value threshold=0.02" Where does this number come from? Shouldn't 0.05 divided by the number of tests (I think 9) be roughly .005?

Demontis, D. et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature Genetics, 51, 63-75.

Satterstrom, F.K. et al. (2019). Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. Nature Neuroscience, 22, 1961-1965.

Rovira, P. et al. (2020). Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacology, 45, 1617-1626.

Reviewer #2:
Remarks to the Author:
Summary of key results.

The manuscript investigates the genetic architecture of childhood ADHD, persistent ADHD and late-diagnosed ADHD, finding different GWAS hits and an overall pattern of results that suggests that the genetic aetiology of these ADHD subgroups is somewhat different.

Originality and significance.

Age-dependent genetic effects are a really interesting topic of research that has been investigated much more in twin studies than in DNA studies, mainly because of limitations in samples. The case of ADHD is particularly interesting given the growing literature on late-onset ADHD in a disorder that had been exclusively conceptualized as a child-onset disorder. Using the I-PSYCH sample, the authors make an important contribution to remedy this gap in the literature.

Data

Can the authors briefly present ICD10 diagnosis code F90. Are symptoms/criteria applied differently in childhood or adulthood, which would be similar to "methodological differences of assessment between children and adults" that the authors mentioned in the introduction.

Is the prevalence of ICD10 diagnoses in the Danish registries known? Are the prevalence figures used for GCTA analyses based on this information? No reference is provided for those prevalence figures.

Can a flowchart be used to present the sample. Were cases of other diseases excluded from controls?

Statistics

In the abstract, maybe provide the Betas (as provided in ST5) rather than Mean PGS. I assume beta are standardized and can be read as, e.g., 0.41SD over controls, which is more interpretable than mean PGS?

Genetic overlap with ADHD symptoms in the general population: "Inattention and hyperactivity were highly correlated with both childhood ADHD …. showed a considerably lower correlation with late-diagnosed". To say that they are considerably lower, the correlations should be directly compared rather than compared to a correlation of one (e.g. test for difference between rginattention 0.86 in childhood, to rg inattention 0.57 in adulthood).

How do the authors explain the very high correlation with "age at mother´s death" (more than for

education)?

"The results reported here suggest that the association of FOXP2 with ADHD is driven to a larger extent by late-diagnosed ADHD (OR=1.11) than childhood ADHD (OR=1.05)." Is the difference significant?

Any justification why 10PCAs, not 8 or 100? Just usage?

From what I understand from the PGS analyses using cross validation, the training datasets were split into ADHD versus not without distinction between early and late-onset. And, based on this overall GWAS, the mean differences in polygenic scores by group were then analysed in the target sample. Is it a problem for this analysis that the overall sample is a mixture of early, persistent,and late groups with unequal sample sizes (i.e the resulting PGS reflecting more one group than another)?

In supplementary Figure 1, caption (C) should be persistent rather than late diagnosed?

Clarity, context and intepretation: abstract, introduction, conclusions

The first paragraph of the abstract, with the sentence: "ADHD is a neurodevelopment disorder with onset in childhood" seems to imply that ADHD is intrinsically "childhood-onset" and that late onset is simply a delayed diagnosis. The author chose the term "late-diagnosed" rather than "late-onset" that is commonly used. Does this mean the authors think that late-onset is just an artefact of measurement, and if it is the case, how does it make sense to look for specific genetic markers of a disease that is not fundamentally different? Their own results suggest that the genetic aetiology is somewhat different (e.g. link with depression, less burden of rare variants). Maybe it is time to stop using the by-the-book definition of ADHD as a "childhood neurodevelopmental disorder" and use parallel terms like "childhood-onset" and "adult-onset" to give the adult onset a fair shot. In the rest of the introduction, the authors, consistent with findings from twin studies, suggest that late-onset ADHD may genetically differ from childhood-onset. If the genetic aetiology is different, this would support the fact that they are to some extent different entities with similar phenotypic manifestations (a bit like breast cancer), rather than simply the same disease with a delayed diagnosis. This is not major, but maybe the authors could first justify their choice of terms; second, more explicitly set out the alternatives in the introduction, e.g. (i) "late'-diagnosed" with largely similar (genetic) aetiology with some chance or transient environmental factors explaining delay in diagnosis; (ii) or different aetiology which means that a parallel term to childhood onset, i.e. late-onset may be more appropriate?

The meta-analyses of inattention and hyperactivity dimensions are not published. The authors mention in the method section that meta-analyses were done on a sample of children and adolescents. Thus, the fact that they correlate with childhood ADHD is not particularly surprising and does not tell us much regarding the correlation between late diagnosis and ADHD dimensions, but simply tells us that childhood diagnosis correlates with childhood dimensions. The correlation with dimensions might be as high in adulthood, just with an adult gwas of dimensions. I am not sure that the conclusion "so late-diagnosed individuals genetically are less predisposed to be inattentive and hyperactive and thus their ADHD could be unnoticed until later in life. " makes sense. In addition to

the technical problem mentioned above, this would imply that adults are diagnosed with ADHD more independently of their actual symptoms?

What was the average age of the different groups when the data was extracted? If part of the childhood onset ADHD sample was younger than age 18, then it seems logical that adolescent/adult onset disorders were more frequent in persistent and late-onset, rather than reflecting necessarily different comorbidities. Later in the manuscript the authors say "The observed differences in comorbidity patterns with other psychiatric disorders could reflect age differences among the groups, but they could also be influenced by differences in the genetic architecture." But they do no address directly that question. The discussion says: "The comorbidity pattern for other psychiatric disorders differed among the groups, and our results suggest that the higher comorbidity of adult onset psychiatric disorders among individuals with persistent and late-diagnosed ADHD is not only due to those individuals being older, but also due to a higher genetic risk". Maybe this argument should be developed and clarified. Which "other psychiatric disorders"? How does that demonstrate that results are not only due to age? Can a sensitivity analysis be done restricting childhood ADHD group to those reaching adulthood and not being diagnosed again (e.g. so that age difference between late onset and childhood onset is null), or is data for this not available.

Results say "while the genetic correlation between childhood ADHD and late-diagnosed ADHD was moderate (rg=0.64, SE=0.03), but not significantly lower than the other two correlations". Discussion say while "The genetic correlation of childhood ADHD with persistent ADHD was high (rg=0.82) … while the genetic correlation of childhood ADHD with late-diagnosed ADHD was considerably lower (rg=0.65)." The correlation went from 0.64 to 0.65 (which is it?) and, most importantly, from not significantly lower to considerably lower. Phrasing chosen in discussion does not seem to be a faithful representation of results. Not everything is about significance testing but a bit more caution may be required.

To what extent do the genetic correlation findings (e.g. higher rg for childhood adhd with autism and higher late onset adhd with depression or alcohol disorder) simply reflects the phenotypic comorbidity (e.g. 27% of late diagnosed ADHD have MDD) and hence are tautological? Genetic patterns follow very closely phenotypic patterns of comorbidity in Sup Table 2.

Can it be said that results are "population-based" when this is a case-control study? I guess it avoids some caveats of classical case control studies (i.e. selection bias of those attending recruiting clinics) but even this is not obvious as those receiving ICD diagnoses may also be also selected (e.g. unequal access to mental health care in the general population).

Reviewer #3:
Remarks to the Author:
Overall, I found this a well-executed study on an important topic. The comparisons within subgroups

of ADHD patients allow for interesting insights into this disease. A few questions/concerns below:
- Last sentence in the abstract needs to be proofread.
- The authors mention diagnostic criteria using age thresholds of 7 or 12 years old. Did the authors consider separating out a younger group as potentially having a larger/different genetic component? Would an ~ pre-/post-puberty stratification be relevant for ADHD?
- Why is there a requirement for a child to "have a known mother"? Did that eliminate many participants?
- Did the authors consider using a control group that did not have individuals with conditions with overlapping genetic influences (like autism) instead of just all non-ADHD?
- It seems possible based on definitions used that some of the childhood ADHD would go on to be persistent ADHD patients over time. Is it feasible with this dataset to define a set of individuals who had a childhood diagnosis but that have follow-up data for some amount of time without an adult diagnosis? Perhaps that would result in too few individuals, but, in theory, that might be a better comparison group for the persistent ADHD group for some of the research questions.
- Still having a little trouble understanding how the PGS score was constructed. So used a leave one out approach 5 times, getting a PGS score for all individuals, and then these were standardized. So, was it not the same set of variants underlying?
- Ln 505: Please define rg at first mention (this instance is missing the subscripted "g").
- Ln 508-9, 512-13: Sentence needs to be proofread.
- Spell out "ACMG".
- Ln 149: It would be useful to provide the relevant general population prevalence for the comorbid conditions so that we can easily see how inflated the presence of autism and depression are among these ADHD patients.
- Ln 194: I find it unclear the conclusion that we are supposed to draw from these results. Please restate.
- Ln 225: In the methods "jackknife" is one word, while here it is 2.
- Spell out "OCD" (I know that it is in the figure caption but should also be defined at the first mention in the text).
- "low constrained genes" sounds a bit awkward. "Less constrained genes" or "tolerant genes"?
- Ln 282: "The chromosome 18 index variant is located [near? In?] DCC, a gene recently linked to the general liability
- Ln 314: Please proofread for the incomplete sentence.
- Ln 345: Please mention again the 5 selected phenotypes in the opening of this paragraph.
- Ln 821: "Horizontal bars" instead of "horizontal bares" (also should be corrected in Ln 831).
- Figure 2 legend refers to "AOF" for mother's age at first birth, while the figure has "AOB".
- Figure 2 legend: Please find some way to incorporate the significance for comparisons between groups in this figure rather than sending the reader to the supplement.
- In supplemental tables, please do not list a P-value as 0. If the analysis tool has a lower limit of significant digits that it will display before putting 0, please find that value and list as "p<10-200" or whatever value that would be.
- Supplemental Table 1: It is unclear which comparisons the P-values represent.
- Supplemental Table 2: "Disorder" is misspelled on line 11.

Author Rebuttal to Initial comments

Response to reviewer comments

We thank the editor for giving us the opportunity to revise our manuscript and the reviewers for their valuable time and useful contributions that have improved the manuscript. Below you can find our point-by-point response to the reviewer comments. We hope that our answers and the actions we have taken to improve the manuscript will clear any doubts.

Reviewer #1:

Remarks to the Author:

This study seems to be well executed. It is a GWAS of ADHD subtypes (childhood, persistent, adult). While few genome-wide significant loci were identified, informative genetic overlaps were calculated between various phenotypes. Fig. 2 is the most important. An exome-sequencing analysis showed that rare protein-truncating variants are more enriched in childhood and persistent ADHD, less so in late-diagnosed (although it is not clear to me what the dependent variable was in this analysis).

Comment #1
Although professionally done, this study seems to add very little beyond recent papers (Demontis et al., 2019; Satterstrom et al., 2019; Rovira et al., 2020) and therefore I would rate its significance below what is required for Nature Genetics. But this is a matter for the editor to decide.

Response
We thank the reviewer for their overall positive comments on our work.

However, we disagree with their view that our paper adds little over the previous work. The reviewer expressed this view in reference to three earlier papers - Demontis et al., 2019, Satterstrom et al., 2019, Rovira et al., 2020.

In our previous work (Demontis et al (2019) and Satterstrom et al (2019)) we reported on, respectively, common and rare variant contributions to ADHD risk in the general population. In the current manuscript, we report how these common and rare variant contributions differ with respect to age of diagnosis. This knowledge is novel and also important both from a clinical and biological perspective.

Our findings suggest that late-diagnosed ADHD is genetically different from persistent ADHD, since we find a lower burden of rare ADHD risk variants in late-diagnosed ADHD compared to those with persistent ADHD. Additionally, the lower burden of variants associated with ADHD symptoms in late diagnosed ADHD compared to both childhood and persistent ADHD suggests that individuals diagnosed with ADHD in adulthood are genetically less predisposed to ADHD symptoms. Finally, our analyses demonstrate that depending on age of onset, the genetic overlap with other phenotypes differs, which

informs about comorbidity risks such as increased risk of comorbid autism in childhood ADHD and increased risk of depression among late-diagnosed ADHD individuals (due to a higher proportion of shared genetics with depression for late-diagnosed ADHD compared to childhood ADHD).

Although Rovira et al (2020) has studied the genetic differences between childhood and adult ADHD, there are important differences between Rovira et al. and our current study.

1. The adult ADHD group that we studied in Rovira et al. included both those who had an ADHD diagnosis during both childhood and adulthood (persistent ADHD) and those who had an ADHD diagnosis only during adulthood (late-diagnosed ADHD). In that study we did not have information to separate the two groups whereas here we studied the genetic contributions to persistent and late-diagnosed ADHD separately. We identify here some striking differences between persistent and late-diagnosed ADHD (e.g., ADHD-PGS and rare variant burden differences), which support that it's important to study the genetic contributions to disease risk separately for the two groups.

2. Rovira et al. analyzed only common variants whereas we looked at both common and rare variants in the present study. Rare variants contribute to ADHD liability significantly, and as much as they contribute to ASD liability (Satterstrom et al. 2019). Hence, when studying the genetic differences across ADHD subtypes, it's important to study both common and rare variants. Also, to the best of our knowledge no one has looked into the rare variant burden differences across ADHD subtypes before.

3. Our sample size, is larger compared to Rovira et al. (providing increased analytic power):
   Childhood ADHD, current study: 14,878 cases vs Roviera et al. 10,617 cases.
   Persistent/late-diagnosed ADHD, current study: 1,473 persistent cases and 7,188 late-diagnosed cases vs Roviera et al. 6,532 persistent+late-diagnosed cases.

Comment #2

The writing in the paper is sometimes awkward or malformed, particularly in the Discussion and Methods. For example, the word "however" is hardly ever used grammatically. Perhaps the journal can be counted on to clean up these things, but nevertheless the authors may wish to go over the manuscript carefully.

Response

We have sent our manuscript out for professional proofreading in British English (proof reading edits are not highlighted in the revised submitted manuscript). We have uploaded a certificate stating this.

Comment #3

A general comment is that many of the PGS analyses do not seem to add anything beyond the genetic correlations. In many cases the authors should consider dropping the redundant showing of one phenotype's PGS significantly predicting another phenotype with which it is genetically correlated.

Response

LD score regression (LDSC) requires only GWAS summary statistics and is a great (and fast) method to estimate the shared genetic component between two phenotypes. However, LDSC requires that both the set of summary statistics come from well-powered GWASs that adequately capture the polygenic signal (i.e. deviation in the distribution of test statistics away from the null in the qq-plot). Our GWAS of persistent ADHD is not well-powered and it is therefore important to demonstrate consistency of the findings using another method. Thus, in our opinion the PGS analyses strengthen our study even though these analyses overall reach the same conclusion as observed in the genetic correlation analyses. Additionally, PGS analyses offer more power particularly when the training GWASs have huge sample sizes. This increased power enables testing genetic associations even when the target sample is smaller. For instance, we were only able to detect pair-wise significant differences in the genetic correlations of childhood ADHD and late-diagnosed ADHD with other phenotypes while, in the PGS analyses, we were also able to detect significant differences between childhood ADHD and persistent ADHD (besides the differences between childhood and late-diagnosed ADHD).

Comment #4

Abstract: Much of this is confusing and possibly circular sounding to someone who has not read the paper. What is a "higher polygenic score (PGS) of ADHD risk variants"? Since three GWAS of ADHD subtypes have been performed, there are three possibilities. A higher average in one subgroup of cases might just mean that the PGS was constructed from the GWAS of that group.

Response

We do not have large training GWASs to generate PGSs specific for childhood, persistent and late-diagnosed ADHD. Instead, we have generated a PGS that represents the general liability to ADHD (please see response to comment #8 addressing questions related to the method). The cohort analyzed in this study is population-based since it contains all individuals born in Denmark between 1981 and 2008 diagnosed with ADHD before or in 2016. This means that the PGS is based on weights representing association with the general liability to ADHD in the Danish population. We have therefore added the information that the PGS represents the general liability to ADHD to the abstract and result section (page 8), and added the following description to the method section page 23:

"Because our ADHD cohort is population-based, including all individuals with ADHD born in Denmark between 1981 and 2008 and diagnosed before or in 2016, the generated PGS represents the general liability to ADHD (because the cross-validation approach is based on all population-based cases)."

We understand the point raised by the reviewer since the PGS is based on a training that includes a higher proportion of childhood ADHD individuals than persistent and late-diagnosed ADHD, which potentially could result in a better prediction of childhood ADHD than the other two groups. However, we observe the opposite: a higher mean ADHD-PGS and higher beta in the adult groups (both in persistent and late-

diagnosed groups) compared to childhood ADHD, which reinforces our finding of higher load of common ADHD risk variants in persistent ADHD.

Comment #5

It seems absurd to write out the truism that "a later diagnosis of ADHD could be due in part to genetic factors."

Response

We are not sure how this comment from the reviewer should be understood, so we try to answer the comment in the two ways that might have been the intention:

1) By "truism" the reviewer means that it is a truism that late-diagnosed ADHD is partly caused by genetic factors. If this was the intention with the comment we fully agree, this would be a truism. The intention was to state that genetics influence the point in time (i.e. older age) that the first diagnosis is given. We have re-written the sentence so this should be clear now. The sentence in the abstract now reads:

   "… and that individuals diagnosed later in life with ADHD have a different genetic architecture than childhood ADHD."

2) By "truism" the reviewer means that it is a truism that genetics influence age at first diagnosis. If this was the intention with the comment, we do not agree with the reviewer that this is a truism. It seems logical that genetics could influence age at first diagnosis. However, age at first diagnosis could also be caused by non-genetic factors such as a more supportive home or a less demanding environment during childhood that may lead to later contact to the healthcare system. Here we find a lower genetic correlation of late-diagnosed ADHD with variants associated with ADHD symptoms and lower burden of rare-variants in late-diagnosed ADHD, suggesting some differences in the genetic architecture of late-diagnosed ADHD compared to childhood and persistent ADHD. To our knowledge there is no empirical evidence demonstrating the impact of genetics on late-diagnosed of ADHD derived from a large genetic study (a previous small study has found no increased burden of ADHD-PGS in late-diagnosed ADHD (N=98) compared to controls).
   It is only recently studies of complex disorders have evaluated the role of genetics on age at diagnosis, and this mainly for adult-onset disorders (e.g. diabetes 2 and cardiovascular disorders). Very little is known about this in ADHD and, in general, this has only been evaluated for a few complex disorders with onset in childhood (e.g. allergy[1]). So, to conclude, we do not think that it is a truism that individuals which are diagnosed later in life have a different genetic architecture than individuals diagnosed earlier in life.

Comment #6

"beta" will suggest regression coefficient to most readers, but its meaning in this context will be obscure (see below).

Response
We gave the mean ADHD-PGS values for the three groups in line with what was reported in the abstract of the paper by Agnew-Blais et al. 2021 on PGS in ADHD sub-groups. But we agree with the reviewer that a beta value would make more sense (which also takes various covariates into consideration that are not accounted for by the mean values). We have therefore replaced the mean PGS values for the three subtypes with the beta values from logistic regression on the case-control status. Please see response to comment #8, where we explain the method, which hopefully will make it clear that the beta not is "obscure".

Comment #7
p. 4: "a recent study found individuals with late-diagnosed ADHD to have a burden of common ADHD risk alleles at a level comparable to individuals without ADHD" Statements like this should always make it clear which type of cases were used in the GWAS used to construct the PGS.

Response
The PRS in the paper by Agnew-Blais et al. 2021, was generated using SNP weights derived from the largest GWAS meta-analysis of ADHD (Demontis et al. Nature Genetics, 2019). We have added this information to the text on page 4:

"…ound that individuals with late-diagnosed ADHD have a burden of common ADHD risk alleles at a level comparable to individuals without ADHD[2] when analysing polygenic scores (PGSs) based on variant weights from the latest ADHD genome-wide association study (GWAS) meta-analysis[3]"

Comment #8
p. 7: I do not understand how the PGS was constructed. I can see how 5-fold cross validation can produce a prediction R^2 that is not inflated by overfitting, but not how it can lead to PGS weights for prediction of the phenotype in subsets of the training sample itself. Perhaps the problem lies in my own ignorance, but regardless I would like the procedure to be explained better in the Methods.

Response
There are no independent ADHD GWAS with sufficient power, therefore, we used the "leave-one out" method which we have used previously (e.g. Demontis et al. Nature Genetics 2019; Grove et al. Nature Genetics 2019).
After re-reading the text, we agree with the reviewer, that more information about how the scores were generated is needed. We have added more information about this, by adding the following to the method section, 23:

"The scores from the five target groups were then pooled at each threshold and tested for association with general ADHD (i.e. all cases vs controls) and the P-value threshold with the scores explaining the

maximum variance (estimated by Nagelkerke's R^2) in the target data of general ADHD (i.e. all ADHD cases vs controls) was used to test for differences in the ADHD-PGS load across ADHD sub-groups ($P_{threshold}<0.1$)."

Additionally, some of the confusion might be related to what the score actually represents. As described above (response to comment #4), the PGS represents the general liability to ADHD risk, which we have stressed in the method section on page 23:

"Because our ADHD cohort is population-based, including all individuals with ADHD born in Denmark between 1981 and 2008 and diagnosed before or in 2016, the generated PGS represents the general liability to ADHD (because the cross-validation approach is based on all population-based cases).**"**

Please also notice than when generating the PRSs the target sample is independent of the training data. We have added this to page 22: **"**In short, the sample was split into five groups, aiming for approximately equal numbers of ADHD cases and controls within each group. We then conducted a GWAS using four out of five groups to derive effect sizes with respect to ADHD risk. These effect sizes were then used to estimate the PGS for the remaining target group. Thus, the training data were independent of the target data. This procedure was repeated five times until PGSs were estimated in all target groups."

Comment #9

Also, I cannot believe that the "SE" in the parentheses can really mean "standard error." If so, then none of the PGS means is significantly greater than zero! Perhaps it should instead be "SD" for "standard deviation."

Response

We thank the reviewer for noticing this error. The values in parentheses are standard deviations and this has been corrected in Supplementary Table 5.

Comment #10

p. 11: "We only identified significant enrichment in rPTVs in childhood and persistent ADHD in highly constrained genes (Pchildhood_ADHD=2.41xa10-11, Ppersistent_ADHD = 1.90x10-3)." This sentence is confusing. Is it merely giving P-values corresponding to estimates and standard errors given earlier in the paragraph?

Response

We have improved the sentence. Additionally, we have added a sentence commenting on the differences in rPTV load between in persistent ADHD and late-diagnosed ADHD. The sentence on page 12 now reads:

"When compared with controls, the load of rPTVs in highly constrained genes was comparable and significantly increased in childhood and persistent ADHD (childhood ADHD beta=0.13, SE=0.02,

P=2.41×10^{-11}; persistent ADHD beta=0.12, SE=0.04, P=1.90×10^{-3}) but lower in late-diagnosed ADHD (beta=0.06, SE=0.03, P=0.02). For *de novo* highly constrained genes, the load of rPTVs was also comparable for childhood ADHD (beta=0.1, SE=0.02, P=2.13×10^{-7}) and persistent ADHD (beta=0.09, SE=0.04, P=0.02), but lower in late-diagnosed ADHD (beta=0.03, SE=0.03, P=0.28) (Figure 3 and Supplementary Table 11). It is worth noticing that even though the sample size for persistent ADHD was less than half of the sample size of late-diagnosed ADHD, we were able to detect a significantly increased load of rPTVs in highly constrained genes in persistent ADHD but not in late-diagnosed ADHD (Supplementary Table 12), implying that rPTVs are less enriched in late-diagnosed ADHD than in persistent ADHD."

Comment #11

p. 13: "The same pattern was observed in the PGS analyses where we found a higher PGS for inattention in childhood and persistent ADHD than in late-diagnosed ADHD, however pair-wise comparisons only revealed nominally significant higher PGS for hyperactivity in childhood ADHD compared to late-diagnosed ADHD." I do not understand what this run-on sentence means.

Response

We have re-written the sentence on page 14, which now reads:

"A lower burden of ADHD-symptom-associated variants in late-diagnosed ADHD was also supported by the PGS analyses, where we found a higher PGS for inattention and hyperactivity in childhood and persistent ADHD than in late-diagnosed ADHD when compared to controls"

Comment #12

p. 18: There are some minor differences between this GWAS of ADHD and the one reported in 2019. If I am understanding correctly, the previous one used dosages and this one uses best-guess discrete genotypes. I am sure this makes no practical difference, but I am wondering why the change? Perhaps relatedly, the previous GWAS had more than 8 million SNPs and this one fewer than 6 million. What accounts for the decrease?

Response

The GWAS published in 2019 was only based on the first wave of iPSYCH data (iPSYCH1) which included genotypes of ~20,000 ADHD cases and ~25,500 population-based controls. In this study we have also included the second wave of yet unpublished ADHD iPSYCH data (iPSYCH2) including genotypes of additional ~10,500 ADHD cases and ~18,000 controls (numbers are before QC).

We have reprocessed all the iPSYCH genotypes (both iPSYCH1 and iPSYCH2), starting with a stringent QC and subsequently imputed genotypes separately for iPSYCH1 and iPSYCH2 using the Haplotype Reference Consortium data as reference panel (described on page 20 and 21). The data in the previous GWAS was imputed using 1000 Genome phase 3 as reference. Subsequently we merged best guess genotypes for iPSYCH1 and iPSYCH2 and only high-quality variants imputed in both were included (INFO>0.80, MAF > 0.01 and missing rate <1%). Since iPSYCH1 and iPSYCH2 samples have been genotyped using two different SNP-arrays (the PsychChip and the Global screening array, respectively),

some variants were lost in the merging step. The difference in imputation reference panels and our requirement of a variant being imputed with high quality in both iPSYCH1 and iPSYCH2 creates the difference in number of variants included in the previous GWAS meta-analysis and this study.
We used best-guess genotypes instead of dosages because merging of dosage data across two samples is computationally challenging and there are no inbuilt options for this in Plink that can accomplish this task. Hence, we used best-guess genotypes with the assumption that it would not have a significant impact on our results to analyze best-guess genotypes.. E.g. in the paper by Choi and O´Reilly (GigaScience, 2019)[4] they found that when doing polygenic risk score analyses moving from best guess genotypes to dosage data the variance explained was only affected at the third decimal or less (an increase in R^2 of 0.0006 to 0.001 depending on the phenotype).

Comment #13
*p. 21: "as no sample overlap and population stratification was present." Besides being ungrammatical, this seems unjustified. How can you be sure there is no population stratification? If the authors want to constrain the intercept in order to reduce the standard error, they should say instead that they are *assuming* the absence of stratification.*

Response
We thank the reviewer for the suggestion on how to phrase this in a better way. The sentence, page 25, now reads:

"The intercept was restricted to 1, when calculating r_g with ADHD symptoms as no sample overlap and no population stratification were assumed."

Comment #14
p. 23: I regrettably do not understand how the rare-variant enrichment analysis was performed. This should be explained more explicitly, perhaps with an equation.

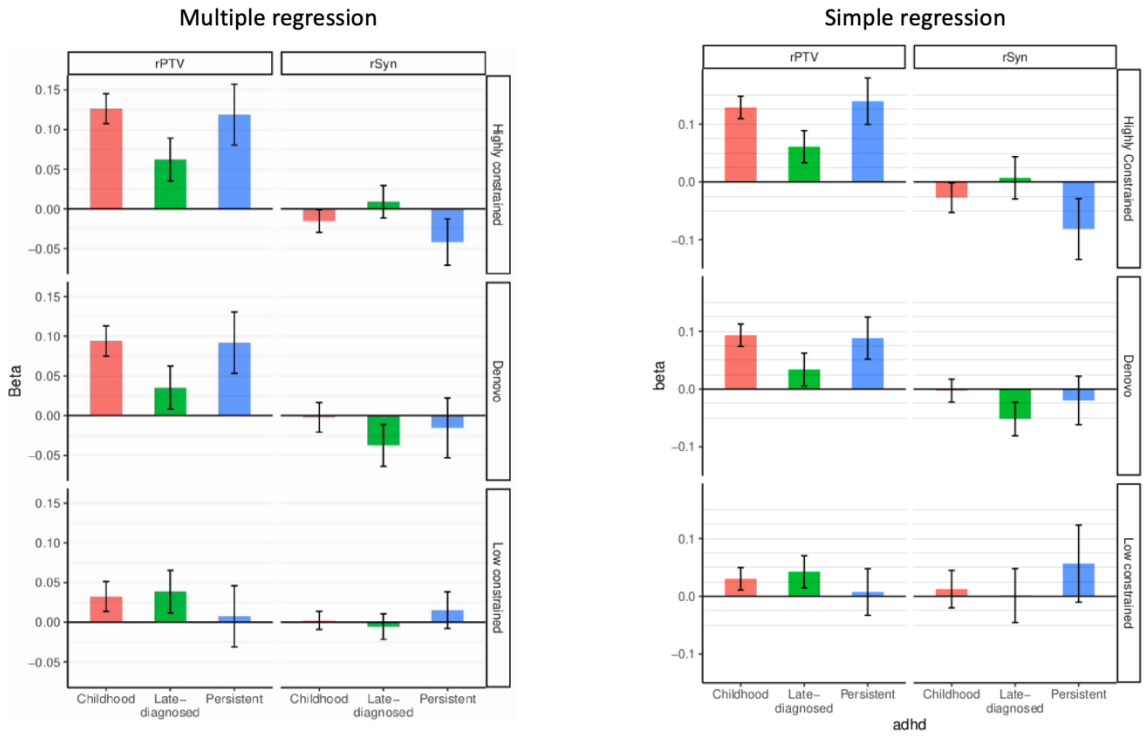
Response
We apologize for not explaining this clearly. Now we have added the following lines to the method section on page 27/28:

"The outcome (or dependent) variable was rPTV count and the predictor (or independent) variables were ADHD given as a categorical variable with multiple factors: childhood, persistent, late-diagnosed and controls (with controls as the reference factor) and relevant covariates:

rPTV counts ~ ADHD (controls | childhood | persistent | late-diagnosed) + covariate 1 + covariate 2 + . . . + covariate N).

ADHD was coded as the independent variable rather than dependent variable in order to make pairwise comparisons between ADHD subtypes in the same analysis."

We coded ADHD as the independent variable rather than dependent variable because it allowed us to make pair-wise comparisons between ADHD subtypes and obtain P-values and effect sizes in one analysis. In order to evaluate the potential impact of ADHD being coded as the independent variable rather than the dependent, we also performed simple regression where individual ADHD subtypes were analyzed separately with each ADHD subtype coded as the dependent variable and rPTV counts as independent (i.e., ADHD (case|control) ~ rPTVs + covariate 1 + . . . + covariate N). The effect size patterns from this analysis were similar to that from our multiple regression analysis as shown in the figure below. We have not included these results in the manuscript, but they could be added to the Supplementary Information if the reviewer (or editor) thinks this would be informative to the reader.



Comment #15

p. 24: "new P-value threshold=0.02" Where does this number come from? Shouldn't 0.05 divided by the number of tests (I think 9) be roughly .005?

Response

We thank the reviewer for noticing this and we are sorry for the confusion. The reviewer is right that the P-value threshold should be 0.006 (0.05/9) and this has now been corrected in the text. The error does not affect the conclusions as all P-values significant in the previous version, remain significant.

Reviewer #2:

Remarks to the Author:

Summary of key results.

The manuscript investigates the genetic architecture of childhood ADHD, persistent ADHD and late-diagnosed ADHD, finding different GWAS hits and an overall pattern of results that suggests that the genetic aetiology of these ADHD subgroups is somewhat different.

Originality and significance.

Comment #1
Age-dependent genetic effects are a really interesting topic of research that has been investigated much more in twin studies than in DNA studies, mainly because of limitations in samples. The case of ADHD is particularly interesting given the growing literature on late-onset ADHD in a disorder that had been exclusively conceptualized as a child-onset disorder. Using the I-PSYCH sample, the authors make an important contribution to remedy this gap in the literature.

Response
We thank the reviewer for the positive comments and we appreciate that the reviewer sees the importance in investigating the role of genetics in age at first diagnosis in ADHD. We agree that there is indeed a knowledge gap regarding this, and that it is important to get a better understanding of late-diagnosed ADHD and whether this group differs biologically/genetically from individuals receiving the diagnosis in childhood.

Comment #2
Can the authors briefly present ICD10 diagnosis code F90. Are symptoms/criteria applied differently in childhood or adulthood, which would be similar to "methodological differences of assessment between children and adults" that the authors mentioned in the introduction.

Response
In this study, ADHD cases were identified via the Danish Psychiatric Central Research Register. Cases were defined as all individuals born in Denmark between 1981 and 2008 who have received a clinical diagnosis (F90.0) in a hospital (including in- and out-patient clinics). The diagnosis criteria are the same

for children and adults but we cannot rule out potential differences in ascertainment, so we have included a sentence regarding this in the discussion (page 17):

"We cannot rule out potential differences in ascertainment of children and adults. For instance, it could be speculated that late-diagnosed ADHD, to a larger extent than childhood ADHD, is diagnosed because patients display other symptoms such as comorbid depression or alcohol use disorder. However, the genetic correlations of late-diagnosed and persistent ADHD (which requires an ADHD diagnosis in childhood) with depression and alcohol use disorder are very similar, suggesting that such ascertainment bias between the two groups is limited."

We have also provided more details on the F90.0 diagnosis code in the method section (page 18):

"The ICD10 diagnosis code F90.0 describes a disorder characterized by early onset, usually in the first five years of life, with hyperactivity and decreased attention. The disorder is considered a childhood onset disorder and therefore, according to the current diagnosis criteria, individuals diagnosed with ADHD as adults should be able to describe ADHD symptoms in childhood retrospectively."

Comment #3

Is the prevalence of ICD10 diagnoses in the Danish registries known? Are the prevalence figures used for GCTA analyses based on this information? No reference is provided for those prevalence figures.

Response

The prevalence of ADHD in the Danish population has been investigated in a cohort consisting of all individuals born in Demark between 1995 and 2016 and diagnosed with ADHD (F90.0 and F98.8 (14.6% of the cases)) (Dalsgaard et al. JAMA Psychiatry 2019). By 18 years of age the cumulative incidence of ADHD in boys was 5.90% [95% CI, 5.76%-6.03%] and in girls 3.04% [95% CI, 2.97%-3.11%]. We have added this reference and one other relevant reference (Faraone et al. 2015) to the method section (page 22).

For the GCTA analysis, we used an arbitrary estimate based on the literature, but we do not know the precise prevalence estimates of childhood, late-diagnosed and persistent ADHD. So, we have now extended the heritability analysis to include a range of prevalence values spanning from 1% to 15%. As expected, the heritability estimates increased with increase in population prevalence. At all prevalence points, the h^2 differences among the sub-groups were maintained (see figure below which we have also added to the Supplementary Information).

We have added the following to the Methods section page 22:

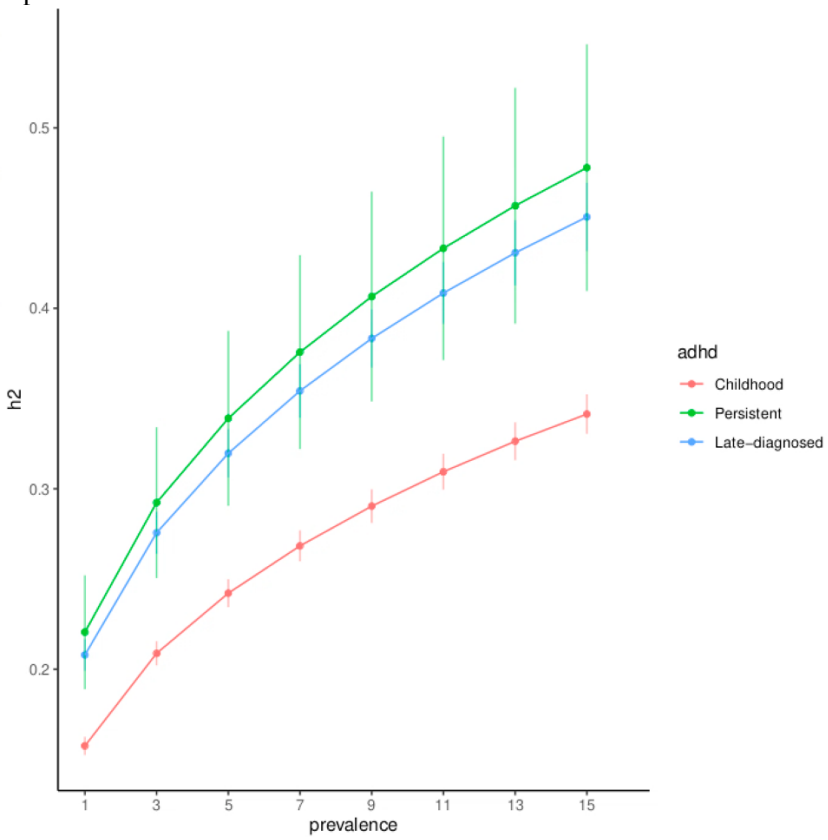
"Because the prevalence estimates of ADHD subtypes are not known precisely, we also estimated h^2_{snp} in the sub-groups over a range of population prevalence values spanning from 1% to 15%."

We have added the following to the Results section, page 7:

"Because the population prevalence of ADHD sub-types is not known precisely, we also estimated the heritability over a range of prevalence values. At all points persistent ADHD demonstrated the highest h^2_{SNP}, followed by late-diagnosed ADHD (Supplementary Figure 3 and Supplementary Table 5.B)."

We have added the figure below to the Supplementary Information:

Supplementary Figure 3. SNP-heritability (h2) estimates in childhood (N=14,878), persistent (N=1,473) and late-diagnosed ADHD (6,961 cases) compared to 38,303 controls using GCTA. Vertical bares represent standard errors.



Comment #4
Can a flowchart be used to present the sample.

Response

Thank you for the suggestion. We have now added a flowchart (Supplementary Figure 7) illustrating the sample and the genetic analyses performed**.**

Comment #5
Were cases of other diseases excluded from controls?

Response
Cases with other diseases were not excluded from the controls. We would like to remind the reviewers about the unique design of iPSYCH, which is a population-based case-cohort[5,6]. The study base for iPSYCH includes all singleton births to mothers who were living in Denmark between 1st of May 1981 and 31st of December 2008, where the child was alive and resided in Denmark at their one-year birthday (N=1,657,449). Control samples were selected randomly from the study base. The control group in iPSYCH1 includes 30,000* randomly selected individuals and in iPSYCH2 the control group includes 21,000* randomly selected individuals. This means that the controls are truly population-based, which is one of the strengths in the iPSYCH design. But as a consequence of the selection criteria, some individuals in the population-based cohort will have at least one of the focus disorders in iPSYCH (2,958* control individuals had one or more iPSYCH disorder) or other mental disorders. When analyzing ADHD, we excluded ADHD among the controls but no other disorders in order to be in line with the iPSYCH design.
*The numbers are before QC.

Comment #6
In the abstract, maybe provide the Betas (as provided in ST5) rather than Mean PGS. I assume beta are standardized and can be read as, e.g., 0.41SD over controls, which is more interpretable than mean PGS?

Response
We agree with the reviewer that betas would be more informative. We have now added betas instead of mean PGS values in the abstract.

Comment #7
Genetic overlap with ADHD symptoms in the general population: "Inattention and hyperactivity were highly correlated with both childhood ADHD …. Showed a considerably lower correlation with late-diagnosed". To say that they are considerably lower, the correlations should be directly compared rather than compared to a correlation of one (e.g. test for difference between rg inattention 0.86 in childhood, to rg inattention 0.57 in adulthood).

Response
We do test directly for difference in rg using jackknife estimated P-values for differences in the genetic correlations of ADHD sub-groups with ADHD symptoms (described in method section page 25, and results in Supplementary Table 6). We realized that this information is missing in the result section, and we thank the reviewer for noticing this. We have included a sentence regarding this on page 9:

"The genetic correlation of hyperactivity with late-diagnosed ADHD was significantly lower than the genetic correlation of hyperactivity with childhood ADHD (P=0.004)."

Comment #8
How do the authors explain the very high correlation with "age at mother´s death" (more than for education)?

Response
We agree with the reviewer that it would improve the manuscript to comment on the result regarding genetic correlations with "age at mother´s death". We have included a few sentences on this in the discussion, page 16:

"We also found a 1.6 times stronger negative genetic correlation of mother's age at death with persistent and late-diagnosed ADHD compared with childhood ADHD. The phenotype mother's age at death captures all variants that affect longevity, for example, variants associated with cardiovascular diseases[7]. Our results therefore suggest that ADHD in adulthood is more enriched in variants that decrease longevity compared with childhood ADHD."

Comment #9
The results reported here suggest that the association of FOXP2 with ADHD is driven to a larger extent by late-diagnosed ADHD (OR=1.11) than childhood ADHD (OR=1.05)." Is the difference significant?

Response
We have tested for significant differences in the effect size of the *FOXP2* locus, and have added the following to the result section, page 7:

"One genome-wide significant locus was identified for late-diagnosed ADHD on chromosome 7, located in *FOXP2* (Table 1 and Supplementary Figures 1.B and 2.E). The effect size was significantly higher in late-diagnosed ADHD (odds ratio [OR]=1.11, standard error [SE]=0.01) compared with childhood ADHD (OR=1.05, SE=0.02) (P=0.012)."

We have added the following to the method section, page 21:

"We tested whether the effect size of the genome-wide significant locus in late-diagnosed ADHD was significantly higher than the effect size observed for the other groups using a z-test and effect sizes from association analyses based on non-overlapping samples. The numbers of non-overlapping controls were: childhood controls, 24,443; persistent controls, 2,289; and late-diagnosed controls, 11,571."

We have updated the sentence in the discussion to reflect the new results, on page 13:

”The effect size was significantly higher in late-diagnosed ADHD compared with childhood ADHD, a finding that suggests the association of *FOXP2* with ADHD is driven to a greater extent by late-diagnosed ADHD than childhood ADHD.”

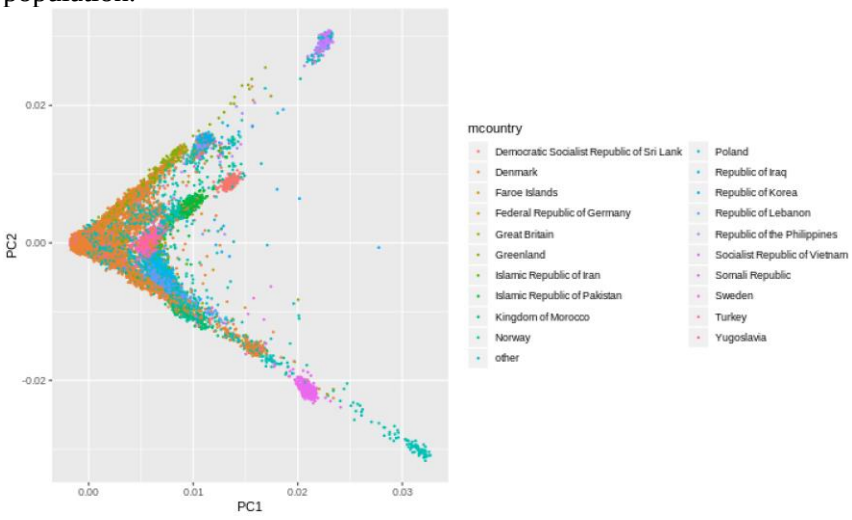
Comment #10

Any justification why 10PCAs, not 8 or 100? Just usage?

Response

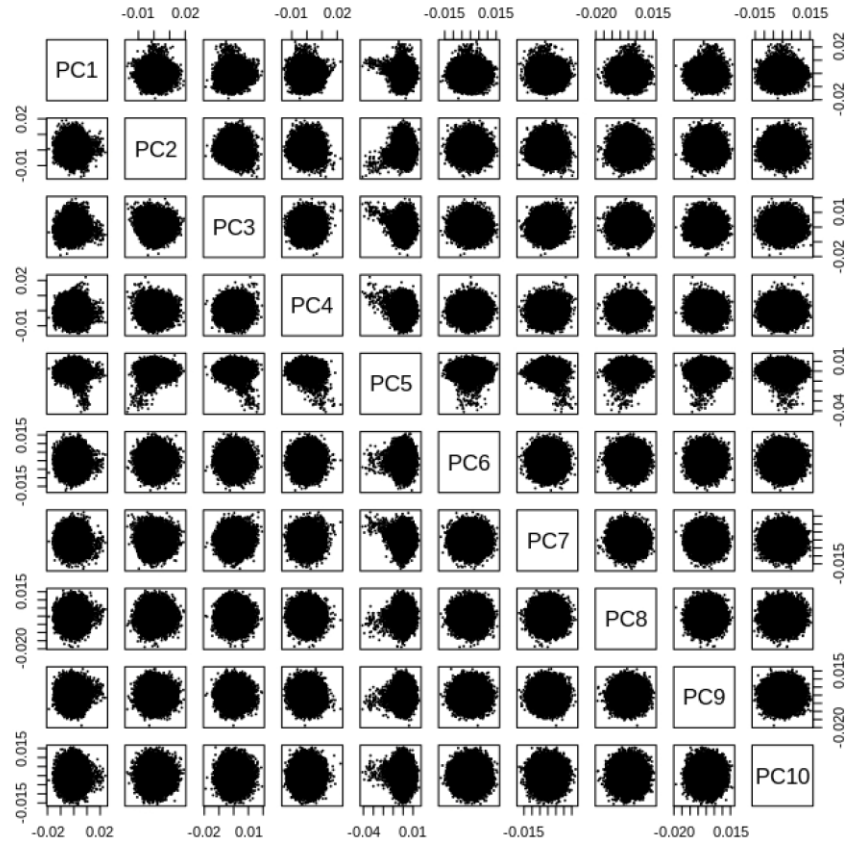
We followed standard GWAS procedures (similar to what we did in Demontis et al. Nature Genetics 2019; Grove et al. Nature Genetics 2019) to remove unrelated and non-European individuals to minimize confounding by relatedness or population stratification.

After removing related individuals, we performed PCA and removed genetic outliers in a stepwise manner, inspecting the PC scatter plots at every step, the details of which are described in the method section. Using the information about the parents’ birth country, we were also able to confirm that the structures captured by the first PCs reflected ancestries. The figure below shows a scatter plot of PC1 vs PC2 of all unrelated iPSYCH individuals. The individuals are color coded based on their parents’ birth countries (mother’s birth country is used in this plot). These structures reflect the ancestries in the Danish population.



By using a set of individuals whose birth place was Denmark at least for three generations as reference, we defined a multi-dimensional ellipsoid (the number of dimensions is decided based on how many PCs contribute to population structures) and then removed individuals located outside the ellipsoid. The PC scatter plots after removing genetic outliers, are displayed below. The number of PCs to include as covariates is an arbitrary choice with the tendency always being to overcorrect rather under-correct. Based on visual inspection of the pairwise scatter plots of the first 10 PCs (see figure below), the first 6 to

8 PCs capture the remaining structures. However, we included the first 10 PCs to ensure that we also account for any residual structures that might exists.



Comment #11

From what I understand from the PGS analyses using cross validation, the training datasets were split into ADHD versus not without distinction between early and late-onset. And, based on this overall GWAS, the mean differences in polygenic scores by group were then analysed in the target sample. Is it a problem for this analysis that the overall sample is a mixture of early, persistent,and late groups with unequal sample sizes (i.e the resulting PGS reflecting more one group than another)?

Response

Thank you for addressing this. Based on this comment and comments from reviewer#1, we have included additional information about what the ADHD PGS represents. The cohort analyzed in this study is

population-based representing all individuals born in Denmark between 1981 and 2008 diagnosed with ADHD before or in 2016. This means that the PGS is based on weights representing association with the general liability to ADHD in the Danish population. We have added this information to the method section, page 23:

"Because our ADHD cohort is population-based, including all individuals with ADHD born in Denmark between 1981 and 2008 and diagnosed before or in 2016, the generated PGS represents the general liability to ADHD (because the cross-validation approach is based on all population-based cases)."

We have also stressed in the abstract, result section (page 8) and discussion (page 14) that the PGS represents the general liability to ADHD. However, the point raised by the reviewer is relevant, since the PGS is based on a training that includes a considerably higher proportion of childhood ADHD individuals than persistent and late-diagnosed ADHD. This could potentially result in a better prediction of childhood ADHD than the other two groups. But despite this we do observe the opposite, a higher mean ADHD-PGS/higher beta in the adult groups (both in persistent and late-diagnosed groups) compared to childhood ADHD, which reinforces the validity of the results. This consideration has been added to the result section page 14:

"The analysed ADHD-PGSs represent the general liability to ADHD because they are derived from data representing all individuals with ADHD in the Danish population born between 1981 and 2008 (see online methods). In relation to this, it should be noted that the training data included a considerably higher proportion of childhood ADHD individuals than persistent and late-diagnosed ADHD, which potentially could result in a better prediction of childhood ADHD than the other two groups. But despite this we do observe the opposite, a higher ADHD-PGS in the adult groups (both in persistent and late-diagnosed groups) compared to childhood ADHD, which reinforces the validity of the results."

Comment #12

In supplementary Figure 1, caption (C) should be persistent rather than late diagnosed?

Response

We thank the reviewer for noticing the error. The caption of Supplementary Figure 1 C has been changed to persistent ADHD.

Comment #13

The first paragraph of the abstract, with the sentence: "ADHD is a neurodevelopment disorder with onset in childhood" seems to imply that ADHD is intrinsically "childhood-onset" and that late onset is simply a delayed diagnosis. The author chose the term "late-diagnosed" rather than "late-onset" that is commonly used. Does this mean the authors think that late-onset is just an artefact of measurement, and if it is the case, how does it make sense to look for specific genetic markers of a disease that is not fundamentally different? Their own results suggest that the genetic aetiology is somewhat different (e.g. link with depression, less burden of rare variants). Maybe it is time to stop using the by-the-book

definition of ADHD as a "childhood neurodevelopmental disorder" and use parallel terms like "childhood-onset" and "adult-onset" to give the adult onset a fair shot.

Response

We use information from the Danish Psychiatric Central Research Register where the diagnoses are given according to the ICD10 criteria, implying that the individuals in the late-diagnosed group should have had ADHD symptoms already as a child. However, this assessment may suffer from recall bias. Therefore, we cannot make any firm conclusions about the actual onset, which is why we have chosen the term late-diagnosed ADHD.

There is an ongoing debate among researchers and clinicians concerning the existence of a purely adult-onset ADHD subtype, which is one of the reasons we conducted this study. We wanted to evaluate, based on the data we have at hand, if there exist genetic differences between late-diagnosed ADHD and the other sub-groups. Based on our results, we can conclude that the genetic architecture of late-diagnosed ADHD differs from the other two groups and thus late-diagnosed ADHD is probably not an artefact of measurement. However, since we do not have age of onset information, it would be an overinterpretation of the results to make any conclusions regarding actual onset.

Comment #14

In the rest of the introduction, the authors, consistent with findings from twin studies, suggest that late-onset ADHD may genetically differ from childhood-onset. If the genetic aetiology is different, this would support the fact that they are to some extent different entities with similar phenotypic manifestations (a bit like breast cancer), rather than simply the same disease with a delayed diagnosis. This is not major, but maybe the authors could first justify their choice of terms; second, more explicitly set out the alternatives in the introduction, e.g. (i) "late'-diagnosed" with largely similar (genetic) aetiology with some chance or transient environmental factors explaining delay in diagnosis; (ii) or different aetiology which means that a parallel term to childhood onset, i.e. late-onset may be more appropriate?

Response

We thank the reviewer for the suggestion. We have added the following to the introduction, page 4:

"However, the disorder is often diagnosed in adolescence and can also be diagnosed in adult life, which according to the current diagnostic criteria is termed 'late-diagnosed ADHD'. Currently, there are insufficient data to clarify if ADHD diagnosed in adulthood has the same underlying causes as childhood ADHD or if late-diagnosed ADHD is a disorder having a somehow different aetiology than childhood ADHD resulting in a later diagnosis or even adult onset-ADHD[8]."

Comment #15

The meta-analyses of inattention and hyperactivity dimensions are not published. The authors mention in the method section that meta-analyses were done on a sample of children and adolescents. Thus, the fact that they correlate with childhood ADHD is not particularly surprising and does not tell us much regarding the correlation between late diagnosis and ADHD dimensions, but simply tells us that

childhood diagnosis correlates with childhood dimensions. The correlation with dimensions might be as high in adulthood, just with an adult gwas of dimensions. I am not sure that the conclusion "so late-diagnosed individuals genetically are less predisposed to be inattentive and hyperactive and thus their ADHD could be unnoticed until later in life. " makes sense.

Response

We do see a larger genetic overlap of ADHD symptoms/dimensions with childhood ADHD compared to late-diagnosed ADHD. However, this is unlikely due to age specific differences in the genetic signals, among children and adults since we observe the largest genetic overlap of ADHD symptoms with persistent ADHD. The age distribution in the persistent and late-diagnosed groups are similar, with all individuals being adults (older that 18 years of age). Hence, the likely explanation would be that the genetic burden associated with ADHD symptoms is relatively lower in late-diagnosed individuals compared to childhood and persistent ADHD, at least when it comes to variants associated with ADHD symptoms in childhood. We have added some lines regarding this to the discussion, page 14/15:

"We cannot exclude that different age distributions in the GWAS meta-analyses of ADHD symptoms and the present study could have influenced the results. However, the age distribution is similar in the persistent and late-diagnosed groups (all are above 18 years of age), indicating that the decreased genetic overlap with late-diagnosed ADHD is most likely not caused by age differences. Our results therefore suggest that a part of the explanation for a later diagnosis of ADHD could be genetic, with late-diagnosed individuals being less genetically predisposed to be inattentive and hyperactive, leaving their ADHD could be unnoticed until later in life"

Comment #16

In addition to the technical problem mentioned above, this would imply that adults are diagnosed with ADHD more independently of their actual symptoms?

Response

We do not have the data to evaluate if adults are diagnosed with ADHD more independently of their actual symptoms. Any statements regarding this would be speculation at this point, but it would surely be interesting to evaluate this in another study designed to evaluate this question.

Comment #17

What was the average age of the different groups when the data was extracted?

Response

The average age of the groups by the end of follow up are: childhood 17.2 (SD=4.95) yrs., persistent 23.9 (±3.13) yrs., late-diagnosed 28.8 (±3.47), controls 22.8 (±3.47) yrs. We have added this information to Supplementary Table 1.

Comment #18

If part of the childhood onset ADHD sample was younger than age 18, then it seems logical that adolescent/adult onset disorders were more frequent in persistent and late-onset, rather than reflecting necessarily different comorbidities. Later in the manuscript the authors say "The observed differences in comorbidity patterns with other psychiatric disorders could reflect age differences among the groups, but they could also be influenced by differences in the genetic architecture." But they do no address directly that question. The discussion says: "The comorbidity pattern for other psychiatric disorders differed among the groups, and our results suggest that the higher comorbidity of adult onset psychiatric disorders among individuals with persistent and late-diagnosed ADHD is not only due to those individuals being older, but also due to a higher genetic risk". Maybe this argument should be developed and clarified. Which "other psychiatric disorders"? How does that demonstrate that results are not only due to age?

Response

When reading the above comment from the reviewer, we think a further specification of the definition of the childhood group is needed, so it is clear that the group contains individuals both younger and older than 18 years of age. The definition has been specified in the result section (page 5) in the method section, page 19:

"1) childhood ADHD, defined as cases diagnosed with ADHD and less than 18 years of age in 2016 or cases older than 18 years by the end of follow-up who did not receive another ADHD diagnosis (N=15,338 before QC)"

Regarding the comorbidity patterns, the late onset disorders depression, schizophrenia and alcohol use disorder were more frequent among late-diagnosed and persistent ADHD compared to childhood ADHD. If this was purely due to differences in age between the groups, we would not see any differences in the genetic correlations. We have stressed this in the discussion by re-writing the sentence in the discussion that the reviewer has highlighted, on page 17:

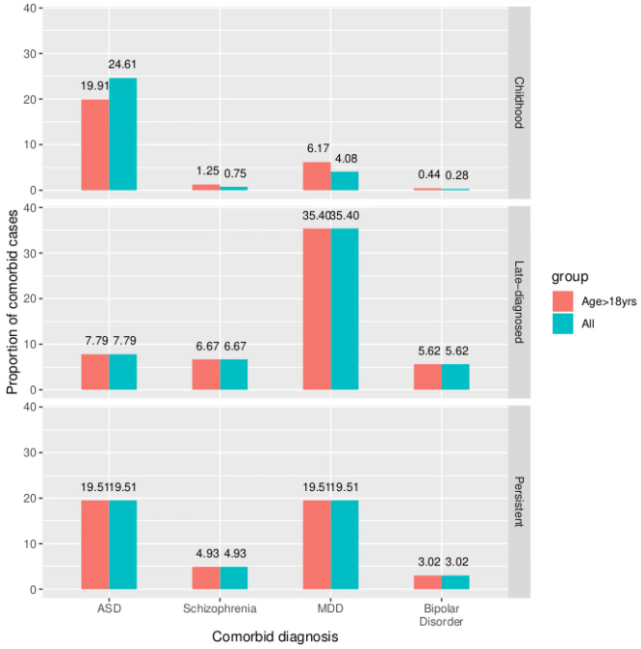
"The comorbidity of depression and alcohol use disorder was highest in the late-diagnosed group. If this observation was only due to age differences among groups, we would not expect the genetic correlations to differ, but we found a higher genetic overlap of these disorders with late-diagnosed ADHD compared with childhood ADHD. This suggests that the higher comorbidity among individuals with late-diagnosed ADHD is not only due to those individuals being older, but also due to a higher genetic risk."

Comment #19

Can a sensitivity analysis be done restricting childhood ADHD group to those reaching adulthood and not being diagnosed again (e.g. so that age difference between late onset and childhood onset is null), or is data for this not available.

Response

As per reviewer's suggestion, we evaluated the proportions of four major psychiatric comorbidities in all childhood ADHD individuals and only those older than 18 years (N= 6,214) at the end of follow-up (in 2016). Restricting to only those older than 18 years of age led to a slight drop in ASD comorbid cases and a slight rise in depression, schizophrenia and bipolar disorder comorbid cases in the childhood ADHD group as shown in the figure below. However, the overall results remained the same: ASD is more prevalent among the childhood group and other psychiatric disorders are more prevalent among persistent and late-diagnosed groups.



We have added a sensitivity analysis in order to evaluate if the differences in the genetic correlations with psychiatric disorders among the groups could be driven by the comorbid individuals (which could indicate problems with the diagnoses e.g. if a large proportion of individuals in the late-diagnosed group are wrongly diagnosed with ADHD when they in reality have depression). In order to investigate this, we performed a PGS analysis where we removed individuals diagnosed with the disorder of the PGS being evaluated (i.e. removing individuals with ADHD+MDD when testing for the association of MDD-PGS with ADHD in the sub-groups). Even after excluding the comorbid cases the results demonstrated the same pattern. That is, we observed higher ASD-PGS in childhood ADHD compared to the other groups and a higher PGS for schizophrenia, MDD and bipolar disorder in late-diagnosed compared to childhood ADHD. We have added this sensitivity analysis to the manuscript. In the method section, page 23:

"We evaluated if the differences in PGS load of four major psychiatric disorders (depression, schizophrenia, bipolar disorder and autism) primarily could be caused by comorbid individuals. In this

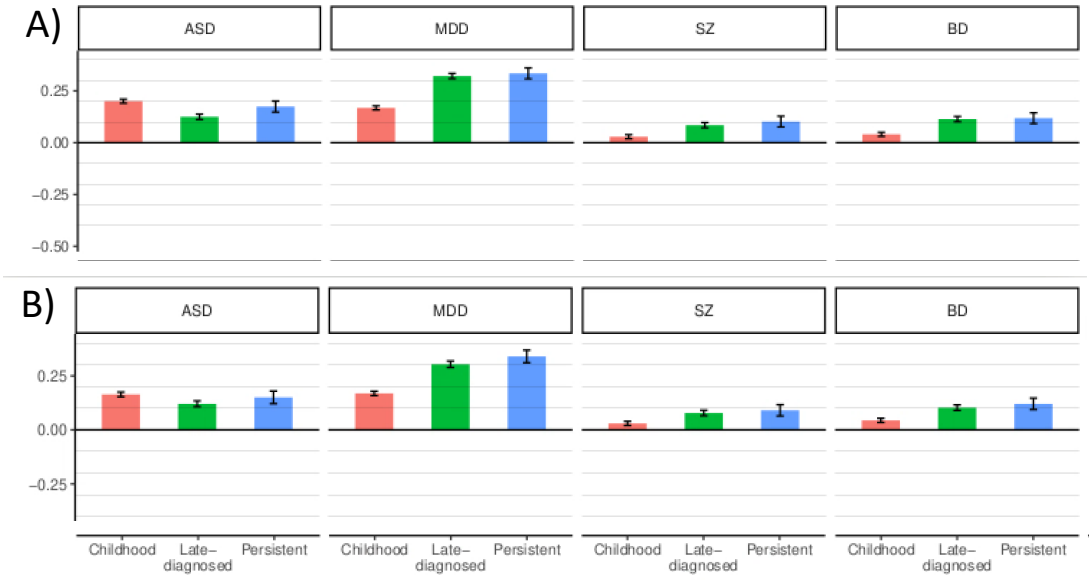
analysis we excluded all individuals in the target sample with the diagnosis of the PGS analyzed (e.g. all depression cases were excluded in the analysis of depression-PGS (sample sizes are given in Supplementary Table 9.B.)).”

We have added the following to the result section, page 11, and results to the Supplementary Tables 9.B and 10.B:

“Finally, we performed two PGS sensitivity analyses. First, we evaluated the PGS for autism, schizophrenia, bipolar disorder and depression in the three ADHD groups in which we excluded individuals with the disorder of the PGS being analysed. The results revealed the same patterns seen in the full sample, but the association of PGS-autism was only nominally significantly associated with childhood ADHD (Supplementary Figure 5 and Supplementary Tables 9.B and 10.B).”

We have added a figure demonstrating the results to the Supplementary Information (Supplementary Figure 5):

Supplementary Figure 5. Results from PGS analyses, demonstrating the association of PGS for autism (ASD), depression (MDD), schizophrenia (SZ), and bipolar disorder (BD) with childhood, late-diagnosed and persistent ADHD compared to controls in the full sample (A) and in samples where individuals are excluded if they are diagnosed with the disorder of the analyzed PGS (B). The beta from multiple-nominal regression is given on the y-axis. Vertical bares represent standard errors.



Comment #20

Results say "while the genetic correlation between childhood ADHD and late-diagnosed ADHD was moderate (rg=0.64, SE=0.03), but not significantly lower than the other two correlations". Discussion say while "The genetic correlation of childhood ADHD with persistent ADHD was high (rg=0.82) … while the genetic correlation of childhood ADHD with late-diagnosed ADHD was considerably lower (rg=0.65)." The correlation went from 0.64 to 0.65 (which is it?) and, most importantly, from not significantly lower to considerably lower. Phrasing chosen in discussion does not seem to be a faithful representation of results. Not everything is about significance testing but a bit more caution may be required.

Response

We would like to thank the reviewer for noticing this error, the correct value is 0.65 and it has now been corrected in the Results section. We have deleted "considerably" in the sentence highlighted by the reviewer, page 14, which now reads:

"The genetic correlation of childhood ADHD with persistent ADHD was high (r_g=0.82) and at the same level as reported previously (r_g=0.81)[9], while the genetic correlation of childhood ADHD with late-diagnosed ADHD was lower (r_g=0.65)."

Comment #21

To what extent do the genetic correlation findings (e.g. higher rg for childhood adhd with autism and higher late onset adhd with depression or alcohol disorder) simply reflects the phenotypic comorbidity (e.g. 27% of late diagnosed ADHD have MDD) and hence are tautological? Genetic patterns follow very closely phenotypic patterns of comorbidity in Sup Table 2.

Response

We addressed this point in the response to comment 19 above. Even after removing the comorbid cases, we still see the same pattern. Hence, the genetic differences observed in the full sample were not purely driven by a subgroup of comorbid cases.

Comment #22

Can it be said that results are "population-based" when this is a case-control study? I guess it avoids some caveats of classical case control studies (i.e. selection bias of those attending recruiting clinics) but even this is not obvious as those receiving ICD diagnoses may also be also selected (e.g. unequal access to mental health care in the general population).

Response

We use the term "population-based" because the study includes all individuals born in Denmark between 1981 and 2008 diagnosed with ADHD before or in 2016. The analyzed blood sample is taken at birth which allows us to avoid the biases arising from non-participation. For example, it allows us to include

severe cases that have died prematurely or are too affected to have the resources to participate in a research study. But we agree with the reviewer that a potential source of bias could be that some ADHD cases never receive a diagnosis because they do not come in contact with hospitals and outpatient clinics. This could potentially create a bias in the cohort that is difficult to detect. In Denmark such a bias will be low because all Danish citizens have equal access to the health care system that is free of charge, but we acknowledge that some cases might be missed. We still think the term "population-based" can be used in relation to the cohort (in line with the wording we use in other iPSYCH studies[3,6,10]). It may also be noted that our cohort do not suffer from sex-differential participation bias, unlike many other large studies (PMID 33888908).

In order to make it clear that we might miss cases we have added the following to the method section page 18:

"We acknowledge that we might miss cases that never come in contact with the health care system due to unequal access to mental health care. However, we think the bias due to this is minor, because Denmark is a small country with health care facilities distributed across the country and citizens have equal access to the health care system that is free of charge."

Reviewer #3:

Remarks to the Author:

Comment #1
Overall, I found this a well-executed study on an important topic. The comparisons within subgroups of ADHD patients allow for interesting insights into this disease. A few questions/concerns below:

Response
We thank the reviewer for the positive comment and the appreciation of our findings.

Comment #2
Last sentence in the abstract needs to be proofread.

Response
We agree with the reviewer that the sentence needs re-writing, it now reads:

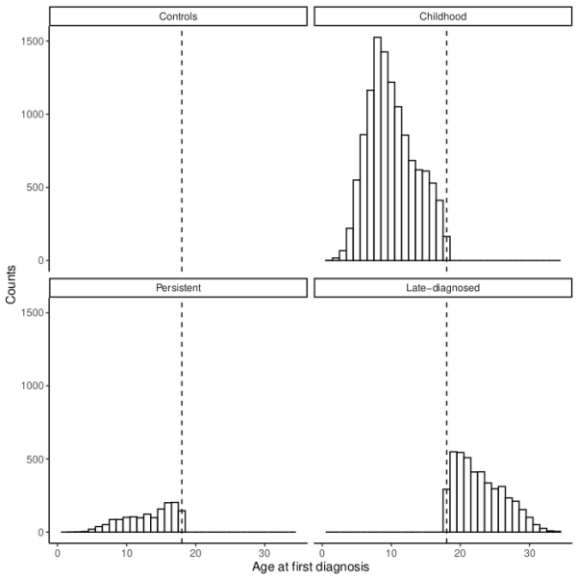
"Overall, our results suggest that genetic factors influence age at first ADHD diagnosis, persistence of ADHD and different comorbidity patterns in the sub-groups."

Comment #3
The authors mention diagnostic criteria using age thresholds of 7 or 12 years old. Did the authors consider separating out a younger group as potentially having a larger/different genetic component? Would an ~ pre-/post-puberty stratification be relevant for ADHD?

Response

It is an ongoing debate among scientists and clinicians if late-diagnosed ADHD (potentially with an adult onset) differs from childhood and persistent ADHD. The focus of this study was to evaluate potential differences in the genetic architecture of these three groups. We agree with the reviewer that stratification of ADHD cases in other ways surely could be interesting and relevant, and, with larger sample sizes, it could be interesting to evaluate if there are some particular genetic characteristics of the early diagnosed group (e.g. younger than 12 years).

However so far, our data suggest that this is not the case. We find a significantly increased ADHD-PGS in the persistent group compared to childhood ADHD, but when looking at the age of first diagnosis (see figure below), the age at first diagnosis in the persistent group is not skewed towards a diagnosis earlier than observed for the childhood ADHD.



Comment #4

Why is there a requirement for a child to "have a known mother"? Did that eliminate many participants?

Response

The requirement for a child to have a "known mother" was a part of the iPSYCH design[5,6], and not decided specifically for this study. An advantage with this requirement is that we can use family information in studies of individuals in the iPSYCH cohort because siblings are identified through the personal security number of the parents. We exclude practically no one by this requirement. The paper by

Pedersen et al. 2006[11] states that: "Among persons born in Denmark (N=7,131,608), the percentage of those with a link to a mother increased rapidly from 8.0% in 1950 to 98.7% in 1960 and 99.9% in 1970, remaining at that level (Figure 5). Similarly, the percentage of persons with a link to a father increased rapidly from 7.6% in 1950 to 95.7% in 1960 and 99.2% in 1970, remaining at that level". We have added this reference and the following sentence to the method section, page 18:

"..(99.9% of all individuals born in Denmark since 1970 have a known mother[11])".

Comment #5

Did the authors consider using a control group that did not have individuals with conditions with overlapping genetic influences (like autism) instead of just all non-ADHD?

Response

We did not consider to remove the individuals with other disorders among the controls. The iPSYCH control group contains a random sample of individuals from the entire Danish population. This unique design is one of the major strengths of iPSYCH[6]. Please also see response to reviewer#2, comment 5, where we have addressed this issue more detail.

Comment #6

It seems possible based on definitions used that some of the childhood ADHD would go on to be persistent ADHD patients over time. Is it feasible with this dataset to define a set of individuals who had a childhood diagnosis but that have follow-up data for some amount of time without an adult diagnosis? Perhaps that would result in too few individuals, but, in theory, that might be a better comparison group for the persistent ADHD group for some of the research questions.

Response

When reading the above comment (and comment #18 from reviewer #2), we think a further specification of the definition of the childhood group is needed, so it is clear that the group contains individuals both younger and older than 18 years of age. The definition has been specified in the result section (page 5) in the method section, page 19:

"(1) childhood ADHD, defined as cases diagnosed with ADHD and less than 18 years of age in 2016 or cases older than 18 years by the end of follow-up who did not receive another ADHD diagnosis (N=15,338 before QC)"

Additionally, we thank the reviewer for the very relevant comment. We agree that it would be relevant to evaluate the impact of age in the childhood group on the results. We have therefore included a PGS sensitivity analysis where we have split the childhood group into those being younger than 18 years of age and those being older than 18 years of age (by the end of follow-up). We have added the following to the method section, page 24:

"(2) We evaluated the potential genetic heterogeneity in the childhood group caused by age. This was done by splitting the individuals in the childhood group into two groups, those younger than 18 years of age (N=8,664) and those older than 18 years (N=6,214) of age by the end of follow-up. We then redid the PGS analyses including the two childhood ADHD groups, persistent and late-diagnosed ADHD."

We have added the following text to the result section, page 8, regarding ADHD-PGS:

"The results did not change in the sensitivity analysis when splitting the childhood group into those younger than 18 years and those older than 18 years of age by the end of follow-up (Supplementary Figure 4)."

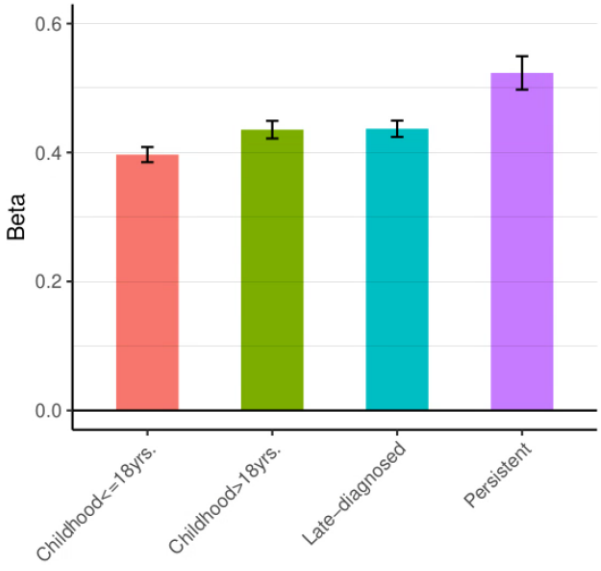
We have added the following to the result section, page 11, regarding PGS analyses of other phenotypes:

"Second, we redid the PGS analyses but this time with the childhood group split in two, namely younger than 18 years and older than 18 years of age by the end of follow-up. The PGS load in the two childhood ADHD groups were generally at the same level except for schizophrenia and bipolar disorder, for which the load was higher in the individuals older than 18 years of age (Supplementary Figure 6 and Supplementary Table 11)."

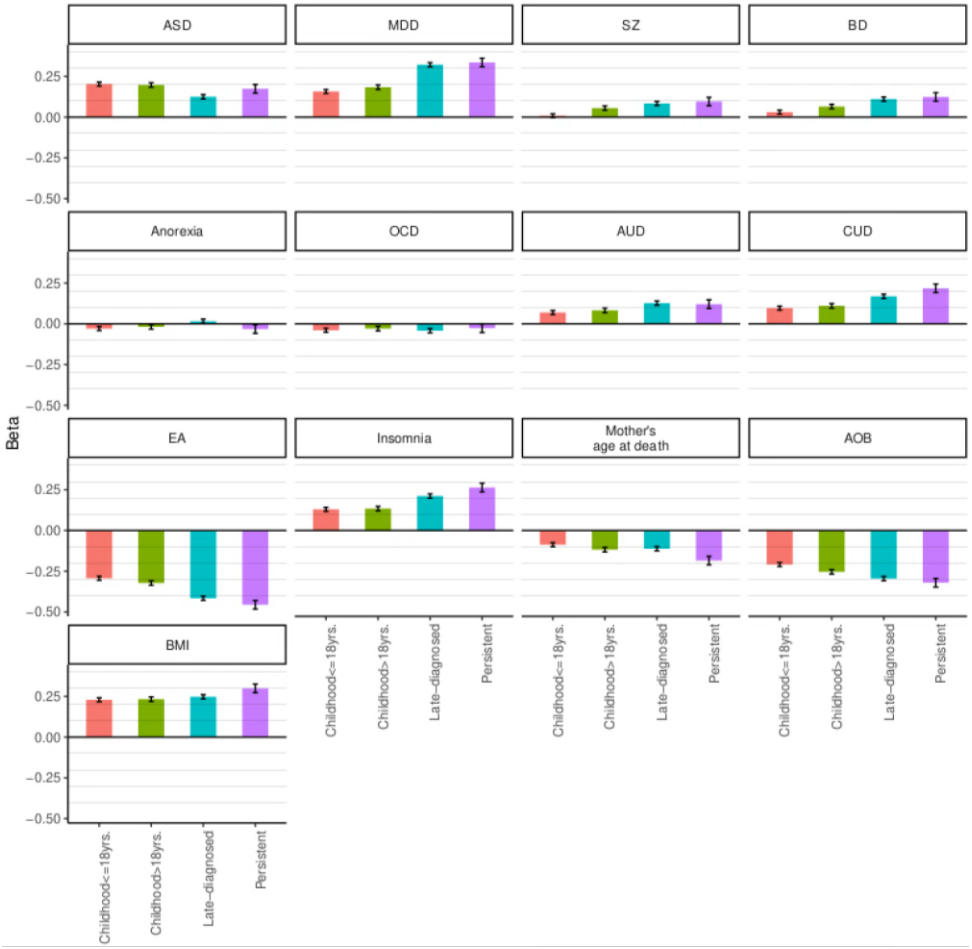
We have added a new Supplementary Table 11 with the results from the sensitivity analysis with the split childhood group.

We added the following two new figures to the Supplementary Information:

Supplementary Figure 4. ADHD-PGS load in childhood ADHD younger than 18 years of age (N=8,664), childhood ADHD older than 18 years of age (N= 6,214), Persistent ADHD (N=1,473), late diagnosed ADHD (N=6,961) compared to controls (N=38,303). The beta from multiple-nominal regression is given on the y-axis.



Supplementary Figure 6. Results from PGS analysis demonstrating the association of PGS with childhood ADHD individuals younger than 18 years of age (N=8,664), childhood ADHD individuals older than 18 years of age (N= 6,214), persistent ADHD (N=1,473) and late-diagnosed ADHD. (N=6,961). PGS for psychiatric disorders: autism spectrum disorder (ASD), major depressive disorder (MDD), schizophrenia (SZ), bipolar disorder (BD), anorexia, obsessive compulsive disorder (OCD), alcohol use disorder (AUD), cannabis use disorder (CUD). PGS for five phenotypes representing domains highly correlated with ADHD: educational attainment (EA), insomnia, mother´s age at death and age of first birth (AOB). On the y-axis is the beta from multi-nominal regression against controls (N=38,303), vertical bares represent standard errors (see also Supplementary Table 11).



Comment #7

Still having a little trouble understanding how the PGS score was constructed. So used a leave one out approach 5 times, getting a PGS score for all individuals, and then these were standardized. So, was it not the same set of variants underlying?

Response

We agree with the reviewer that the ADHD-PGS construction could be better described. We have added the following to the method section, page 23:

”The PGSs were standardised for each of the five target sample groups (subtracting the mean and dividing by the standard deviation). The scores from the five target groups were then pooled at each threshold and

tested for association with general ADHD (i.e. all cases vs controls) and the P-value threshold with the scores explaining most of the variance (estimated by Nagelkerke's R^2) in the target data of general ADHD (i.e. all ADHD cases vs controls) were used to test for differences in the ADHD-PGS load among ADHD sub-groups ($P_{threshold}<0.1$). Because our ADHD cohort is population-based, including all individuals with ADHD born in Denmark between 1981 and 2008 and diagnosed before or in 2016, the generated PGS represents the general liability to ADHD (because the cross-validation approach is based on all population-based cases)."

Comment #8
Ln 505: Please define rg at first mention (this instance is missing the subscripted "g").

Response
We have corrected this in manuscript.

Comment #9
Ln 508-9, 512-13: Sentence needs to be proofread.

Response
We thank the reviewer for noticing the errors, which are now corrected.

Comment #10
Spell out "ACMG".

Response
We have spelled out ACMG - "American College of Medical Genetics".

Comment #11
Ln 149: It would be useful to provide the relevant general population prevalence for the comorbid conditions so that we can easily see how inflated the presence of autism and depression are among these ADHD patients.

Response
The controls used in this study represent ~50.000 individuals randomly selected from the same birth cohort as the cases. The cohort comprises all singletons born in Denmark between May 1, 1981, and December 31, 2008, who have a known mother and were residents in Denmark on their first birthday and (N = 1,472,762). The prevalence of disorders among the controls represents therefore an estimate of the prevalence in a comparable population-based control group. We have added a column to Supplementary Table 2, with the information about the prevalence of the disorders in the population-based control group.

Comment #12
Ln 194: I find it unclear the conclusion that we are supposed to draw from these results. Please restate.

Response
We have elaborated the sentence, page 8, which now reads:

"However, the differences were non-significant and we cannot make any strong conclusions based on this small replication sample."

Comment #13
Ln 225: In the methods "jackknife" is one word, while here it is 2.

Response
We thank the reviewer for noticing the error, it has now been corrected.

Comment #14
Spell out "OCD" (I know that it is in the figure caption but should also be defined at the first mention in the text).

Response
OCD is now spelled out the first time it occurs.

Comment #15
"low constrained genes" sounds a bit awkward. "Less constrained genes" or "tolerant genes"?

Response
We have re-written the sentence on page 12, which now reads:

"(3) "less constrained genes", genes being relatively tolerant to loss of function mutations with a pLI < 0.1 (9,662 genes)".

We have re-written the sentence on page 27, which now reads:

"For comparison we also tested a set of 9,662 evolutionary less constrained genes, i.e. having a pLI score < 0.1."

Comment #16
Ln 282: "The chromosome 18 index variant is located [near? In?] DCC, a gene recently linked to the general liability

Response
We have added the missing information, the sentence now reads:

"The chromosome 18 index variant is located in *DCC*, a gene recently linked to the general liability"

Comment #17
Ln 314: Please proofread for the incomplete sentence.

Response
We have deleted the incomplete sentence.

Comment #18
Ln 345: Please mention again the 5 selected phenotypes in the opening of this paragraph.

Response
We have added information about the five selected phenotypes to the sentence, as suggested by the reviewer.

Comment #19
Ln 821: "Horizontal bars" instead of "horizontal bares" (also should be corrected in Ln 831).

Response
We thank the reviewer for noticing the errors, which are now corrected.

Comment #20
Figure 2 legend refers to "AOF" for mother's age at first birth, while the figure has "AOB".

Response
We thank the reviewer for noticing the error, which is now corrected.

Comment #21
Figure 2 legend: Please find some way to incorporate the significance for comparisons between groups in this figure rather than sending the reader to the supplement.

Response
We thank the reviewer for the suggestion, the requested information has now been incorporated into Figure 2.

Comment #22

In supplemental tables, please do not list a P-value as 0. If the analysis tool has a lower limit of

significant digits that it will display before putting 0, please find that value and list as "p<10-200" or whatever value that would be.

Response
We have added the lower p-value limits to Supplementary Table 2, 7 and 9.

Comment #23
Supplemental Table 1: It is unclear which comparisons the P-values represent.

Response
Thank you for noticing this, we have added the information so the comparisons should be clear now.

Comment #24
Supplemental Table 2: "Disorder" is misspelled on line 11.

Response
Thank you for noticing the error, which has been corrected.

##########

References

1 Ferreira, M. A. R. *et al.* Age-of-onset information helps identify 76 genetic variants associated with allergic disease. *PLoS genetics* **16**, e1008725, doi:10.1371/journal.pgen.1008725 (2020).
2 Agnew-Blais, J. C. *et al.* Polygenic Risk and the Course of Attention-Deficit/Hyperactivity Disorder From Childhood to Young Adulthood: Findings From a Nationally-Representative Cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*, doi:10.1016/j.jaac.2020.12.033 (2021).
3 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics* **51**, 63-75, doi:10.1038/s41588-018-0269-7 (2019).
4 Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, doi:10.1093/gigascience/giz082 (2019).
5 Bybjerg-Grauholm, J. *et al.* The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. *medRxiv* (2020).

6 Pedersen, C. B. *et al.* The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Molecular psychiatry*, doi:10.1038/mp.2017.196 (2017).
7 Pilling, L. C. *et al.* Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging (Albany NY)* **8**, 547-560, doi:10.18632/aging.100930 (2016).
8 Asherson, P. & Agnew-Blais, J. Annual Research Review: Does late-onset attention-deficit/hyperactivity disorder exist? *J Child Psychol Psychiatry* **60**, 333-352, doi:10.1111/jcpp.13020 (2019).
9 Rovira, P. *et al.* Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology* **45**, 1617–1626, doi:10.1038/s41386-020-0664-5 (2020).
10 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).
11 Pedersen, C. B., Gotzsche, H., Moller, J. O. & Mortensen, P. B. The Danish Civil Registration System. A cohort of eight million persons. *Danish Medical Bulletin* **53**, 441-449 (2006).

##

Decision Letter, first revision:

9th Mar 2022

Dear Dr Demontis,

Your Article, "Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention deficit hyperactivity disorder" has now been seen by 3 referees. You will see from their comments below that while they find your work of interest, some important points are raised. We are interested in the possibility of publishing your study in Nature Genetics, but would like to consider your response to these concerns in the form of a revised manuscript before we make a final decision on publication.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file. At this stage we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

*1) Include a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions, available
here.
Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf
It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.
A revised checklist is essential for re-review of the paper.

Please be aware of our guidelines on digital image standards.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit <a
href="http://www.springernature.com/orcid">www.springernature.com/orcid.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your

work.

Sincerely,
Wei

Wei Li, PhD
Senior Editor
Nature Genetics
New York, NY 10004, USA
www.nature.com/ng

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The authors have mostly addressed my concerns. They seem to have done a good job addressing the extensive comments of the other reviewers. Here are my latest comments.

abstract: "individuals diagnosed later in life with ADHD have a different genetic architecture than childhood ADHD" There is a lack of parallelism here, in that individuals are being compared to a disorder. A fix is to say "the genetic architecture of childhood ADHD differ from that of late-diagnosed ADHD".

p. 8: "The highest mean ADHD-PGS was found for persistent ADHD (mean=0.41, SE=0.95), followed by late-diagnosed ADHD (mean=0.27, SE=0.98) and then childhood ADHD (mean=0.26, SE=0.96) (Supplementary Table 5)." I raised this point last time but perhaps wasn't clear. I believe that "SE" here should be "SD."

p. 12: "It is worth noticing ..." I suggest deleting this sentence. It is hard to understand. Also, very often, the argument "I got a significant result with a smaller sample size, which means that the effect is especially big" is fallacious.

p. 16: "We also found a 1.6 times stronger negative genetic correlation of mother's age at death with persistent and late-diagnosed ADHD compared with childhood ADHD." This is a funny way to put it. Does it make sense to examine the ratio of two correlations? Perhaps the ratio of two Fisher's Z-transformed correlations would be more appropriate? I suggest omitting the attempt to quantify.

p. 25: "the intercept was restricted to 1" This must be a mistake, because the intercept of bivariate LD Score regression is equal to zero in the absence of stratification and sample overlap.

Reviewer #2:
Remarks to the Author:
The authors have responded to a number of my concerns. I think that caution is warranted in some places.

The additional information provided by the author regarding the diagnosis, suggests that it is unclear what the late-diagnosed group truly represents. As people in this group are supposed to be able to describe ADHD symptoms in childhood retrospectively, this group could include: 1. Persistent ADHD, where the diagnosis was missed in childhood; 2. Sub-threshold ADHD in childhood followed by meeting threshold in adulthood; 3. True adult-onset ADHD if the patients did not have ADHD in childhood and the symptoms they mention from that period are due to recall bias. And this group would not include all adult-onset ADHD (i.e. patients meeting threshold for ADHD as adults but not presenting or recalling childhood symptoms). Because the group is a mixture, it is unclear how different findings for this group should be expected to be in comparison with the persistent group, nor whether they reflect a truly developmental perspective (i.e. adult onset). Also, the childhood limited cases usually have a lower phenotypic load (i.e. trajectory studies show that they have a lower baseline as well as a sharper decrease compared to persistent). So the findings may largely reflect differences in severity between childhood limited versus (persistent + late-diagnosed). And the differences between the two later groups may simply reflect differences in diagnosis requirements (e.g. people with depression more likely to recall having childhood ADHD). Overall, this deserves a more in-depth discussion as a limitation of the study.

The authors state that the PGS represents the general liability to ADHD. It may be worth being slightly more cautious here: the childhood onset (whether limited or persistent) is probably over-represented, given the more recent awareness surrounding adult ADHD (likely underdiagnosed in particular in older cohorts). So this general liability is likely still skewed towards childhood ADHD.

"We acknowledge that we might have missed cases that never came into contact with the health care system due to unequal access to mental health care. However, we think the bias due to this is minor, because Denmark is a small country with health care facilities distributed across the country and citizens have equal access to the health care system that is free of charge." It is worth being more cautious here, the bias may be 'mitigated' rather than minor. Whereas free access may limit issues, possible biases do not only stem from the fact that people may be refused access, but also from the fact that some people may seek less help in the first place. Many other factors than free access may thus influence whether and when people seek medical help (including genetic liability to ADHD…).

Writing still needs polishing in places, e.g. the new sentence: "

Our results therefore suggest that a part of the explanation for a later diagnosis of ADHD could be genetic, with late-diagnosed individuals being less genetically predisposed to be inattentive and hyperactive, leaving their ADHD could be unnoticed until later in life."

Reviewer #3:
Remarks to the Author:
I am satisfied with the authors' responses to the comments. I would like to also commend them for the extensive additional analyses and writing undertaken to thoughtfully address the concerns of the reviewers. The manuscript is much clearer.

Author Rebuttal, first revision:

Response to reviewer comments

We thank the editor for giving us the opportunity to revise our manuscript and the reviewers for their valuable time and useful contributions that have improved the manuscript. Below you can find our point-by-point response to the reviewer comments. We hope that our answers and the actions we have taken to improve the manuscript will clear the last doubts.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:

Comment #1
The authors have mostly addressed my concerns. They seem to have done a good job addressing the extensive comments of the other reviewers. Here are my latest comments.

Response
We would like to thank the reviewer for acknowledging our effort to improve the manuscript.

Comment #2
abstract: "individuals diagnosed later in life with ADHD have a different genetic architecture than childhood ADHD" There is a lack of parallelism here, in that individuals are being compared to a disorder. A fix is to say "the genetic architecture of childhood ADHD differ from that of late-diagnosed ADHD".

Response
We thank the reviewer for suggesting a better wording, we have corrected the sentence accordingly.

Comment #3

p. 8: "The highest mean ADHD-PGS was found for persistent ADHD (mean=0.41, SE=0.95), followed by late-diagnosed ADHD (mean=0.27, SE=0.98) and then childhood ADHD (mean=0.26, SE=0.96) (Supplementary Table 5)." I raised this point last time but perhaps wasn't clear. I believe that "SE" here should be "SD."

Response

The reviewer is correct, it should be SD stated together with the mean ADHD-PGS. In the revision we corrected this in the Supplementary Tables, but somehow, we missed to get this right in the text. We are sorry for that. It should be correct now.

Comment #4

p. 12: "It is worth noticing ..." I suggest deleting this sentence. It is hard to understand. Also, very often, the argument "I got a significant result with a smaller sample size, which means that the effect is especially big" is fallacious.

Response

We have followed the reviewer´s recommendation and deleted the sentence on page 12.

Comment #5

p. 16: "We also found a 1.6 times stronger negative genetic correlation of mother's age at death with persistent and late-diagnosed ADHD compared with childhood ADHD." This is a funny way to put it. Does it make sense to examine the ratio of two correlations? Perhaps the ratio of two Fisher's Z-transformed correlations would be more appropriate? I suggest omitting the attempt to quantify.

Response

We have followed the reviewer´s suggestion and omitted the attempt to quantify. The sentence on page 16 now reads: "We also found a stronger negative genetic correlation of mother's age at death with persistent and late-diagnosed ADHD compared with childhood ADHD."

Comment #6

p. 25: "the intercept was restricted to 1" This must be a mistake, because the intercept of bivariate LD Score regression is equal to zero in the absence of stratification and sample overlap.

Response

We thank the reviewer for noticing this. We have corrected the sentence on page 25, so it states correctly what was done:

"No sample overlap and no population stratification were assumed when calculating r_g with ADHD symptoms and therefore the intercept was restricted by setting the single-trait intercepts to 1 and cross-trait intercepts to 0."

Reviewer #2:
Remarks to the Author:

Comment #1
The authors have responded to a number of my concerns. I think that caution is warranted in some places.

The additional information provided by the author regarding the diagnosis, suggests that it is unclear what the late-diagnosed group truly represents. As people in this group are supposed to be able to describe ADHD symptoms in childhood retrospectively, this group could include: 1. Persistent ADHD, where the diagnosis was missed in childhood; 2. Sub-threshold ADHD in childhood followed by meeting threshold in adulthood; 3. True adult-onset ADHD if the patients did not have ADHD in childhood and the symptoms they mention from that period are due to recall bias. And this group would not include all adult-onset ADHD (i.e. patients meeting threshold for ADHD as adults but not presenting or recalling childhood symptoms). Because the group is a mixture, it is unclear how different findings for this group should be expected to be in comparison with the persistent group, nor whether they reflect a truly developmental perspective (i.e. adult onset).

Response
One of the main aims of this study was to evaluate if late-diagnosed ADHD differ genetically from persistent ADHD, motivated by the ongoing discussion about why some individuals are diagnosed later in life, which could occur due to different reasons outlined by the reviewer. Thus, we had no prior expectations about what we would find, or how different the late-diagnosed group would be from the persistent group.
We agree with the reviewer that we do not know the reason why some individuals are diagnosed later in life, but we think that our results could contribute to explain a part of this. We observe a nominally significant lower ADHD-PRS in late-diagnosed ADHD compared to persistent ADHD and lower genetic correlation of late-diagnosed ADHD with ADHD symptoms compared to the genetic correlation observed for both childhood and persistent ADHD. These results suggest that a part of the explanation of a later diagnosis might be biological due to a lower load of ADHD symptoms associated variants in late-diagnosed ADHD compared to the persistent group, which has never been demonstrated previously.

If we instead had observed no difference between the persistent and late diagnosed groups in the genetic analyses it would still have been informative, since such an observation would suggest that it primarily would be non-genetic factors that explain why some are diagnosed later in life.

As pointed out by the reviewer some heterogeneity might exist in the late-diagnosed group and it would be very interesting to investigate this further in follow-up studies, e.g., in studies involving deep phenotyping of late-diagnosed individuals to obtain information about age at first symptoms.

We agree with the reviewer that we might miss true adult-onset ADHD if the patients did not have ADHD in childhood and therefore no recalling of symptoms in childhood. There is no diagnosis in ICD10 or DSMV for adult-onset ADHD, so we do not know how common it is to diagnose these individuals with ADHD even though symptoms were absent in childhood (if we assume adult-onset ADHD exist). We have added the following sentence to the method section, page 19 in order to address this issue: "We would like to note that some adults with ADHD might be missed if they were not able to recall having ADHD symptoms in childhood."

Comment #2
Also, the childhood limited cases usually have a lower phenotypic load (i.e. trajectory studies show that they have a lower baseline as well as a sharper decrease compared to persistent). So the findings may largely reflect differences in severity between childhood limited versus (persistent + late-diagnosed).

Response
We think the hypothesis suggested by the reviewer is very relevant, i.e., that differences in ADHD symptoms/severity could, in part, be influenced by the genetic differences we observe. When considering ADHD-PGS we identify a significant increased load of ADHD risk variants in persistent ADHD compared to both childhood and late-diagnosed ADHD. A previous study has shown that the ADHD-PGS is associated with higher number of ADHD symptoms at ages 5, 7, 10, and 12 in a smaller sample of 327 ADHD cases (PMID: 33440202), supporting the hypothesis suggested by the reviewer. Therefore, it could be hypothesized that the persistent group, besides having ADHD symptoms persisting into adulthood, also represents individuals, more likely to have severe ADHD.

We have included the following sentence in the discussion regarding this hypothesis on page 13: "…The latter study identified a significant association of ADHD-PGS with increased number of ADHD symptoms, and it could be speculated that individuals with persistent ADHD on average have more ADHD symptoms than individuals with childhood and late-diagnosed ADHD."

Comment #3

And the differences between the two later groups may simply reflect differences in diagnosis requirements (e.g. people with depression more likely to recall having childhood ADHD). Overall, this deserves a more in-depth discussion as a limitation of the study.

Response

We can follow the reviewer´s thoughts that individuals with depression might reflect more over the past (either voluntary or during the clinical assessment for depression) compared to individuals without depression resulting in an over representation of individuals with ADHD and depression in the late-diagnosed group compared to the persistent ADHD group, which potentially could result in an increased load of depression associated variants in late-diagnosed ADHD. However, this is not what we observe; the genetic correlation of persistent ADHD with depression (r_g=0.78) is higher than the genetic correlation of late-diagnosed ADHD with depression (r_g=0.69) and in the PGS analysis we found a similar polygenic risk for depression in late-diagnosed (beta=0.32) and persistent ADHD (beta=0.33). Our findings therefore support a similar polygenic risk load for depression in persistent and late-diagnosed ADHD. If there exist differences in diagnostic requirements due to a comorbid diagnosis of depression, this is not reflected in the genetic analyses.

In general, comorbid psychiatric diagnoses do not seem to have a strong impact on the genetic differences we observe between the groups. In the PGS sensitivity analyses where we exclude individuals diagnosed with the disorder of the analyzed PGS (Supplementary Figure 5; Supplementary Table 10.B), we observed the same pattern of genetic differences among the groups as when analyzing the full sample (with the note that the autism-PGS was only nominal significantly increased in childhood ADHD compared to late-diagnosed ADHD when excluding autism cases). When excluding individuals with depression, the depression-PGS was similar in late-diagnosed ADHD (beta=0.30) and persistent ADHD (beta=0.34)) which is written on page 11.

In order to address the point raised by the reviewer, we have elaborated the discussion about ascertainment bias on page 17 and added the following sentences: “Likewise, in the PGS sensitivity analyses, where individuals with the disorder of the PGS being analysed were excluded, the genetic differences among the groups demonstrated the same patterns as observed in the full sample. This further reinforces the conclusion that comorbid psychiatric disorders do not have a strong influence on the observed genetic differences among the ADHD sub-groups. So, even though we cannot exclude ascertainment biases, they do not seem to have a strong impact on the genetic differences”

Comment #4

The authors state that the PGS represents the general liability to ADHD. It may be worth being slightly more cautious here: the childhood onset (whether limited or persistent) is probably over-represented, given the more recent awareness surrounding adult ADHD (likely underdiagnosed in particular in older cohorts). So this general liability is likely still skewed towards childhood ADHD.

Response

We agree with the reviewer that some individuals in the cohort might be missed and that this, in particular might be individuals with late-diagnosed ADHD. We have therefore made it clear in the text that the PGS represents the general liability to ADHD with respect to diagnosed ADHD. We have added the following to the method section page 23: "We acknowledge that the (Danish) health care system might not capture all individuals with ADHD, and that late-diagnosed ADHD probably is more likely to be missed than childhood and persistent ADHD, because the disorder seems to be underdiagnosed among adults[66]. We would therefore like to stress that the PGS only reflects the general liability with respect to diagnosed ADHD."

Additionally, we have added "diagnosed" to the text (in the abstract and page 13 and 23), in order to be more precise about that the PGS represents general liability to diagnosed ADHD.

Comment #5

"We acknowledge that we might have missed cases that never came into contact with the health care system due to unequal access to mental health care. However, we think the bias due to this is minor, because Denmark is a small country with health care facilities distributed across the country and citizens have equal access to the health care system that is free of charge." It is worth being more cautious here, the bias may be 'mitigated' rather than minor. Whereas free access may limit issues, possible biases do not only stem from the fact that people may be refused access, but also from the fact that some people may seek less help in the first place. Many other factors than free access may thus influence whether and when people seek medical help (including genetic liability to ADHD…).

We thank the reviewer for bringing this up. We certainly agree on the point that other factors than access to health care can influence if an individual gets diagnosed or not. We have modified the sentence on page 18/19 to address this point: "We acknowledge that we might have missed cases that never came into contact with the health care system or cases that only briefly interacted with the system and never got diagnosed. This could be due to various reasons such as lack of resources (mentally or physically) to engage with the system, lack of persistency if the

diagnostic investigation took too long or unequal access to mental health care. Regarding the latter, we think Denmark, world-wide is one of the countries with least bias caused by unequal access to health care. Denmark is a small welfare state with health care facilities distributed across the country and citizens have equal access to the health care system that is free of charge."

Comment #6
Writing still needs polishing in places, e.g. the new sentence: "Our results therefore suggest that a part of the explanation for a later diagnosis of ADHD could be genetic, with late-diagnosed individuals being less genetically predisposed to be inattentive and hyperactive, leaving their ADHD could be unnoticed until later in life."

Response
We are sorry for the unclear sentence. We have corrected this, and the sentence now reads: "Our results therefore suggest that a part of the explanation for a later diagnosis of ADHD could be genetic, with late-diagnosed individuals being less genetically predisposed to be inattentive and hyperactive, leaving their ADHD unnoticed until later in life."

Additionally, we have done careful proofreading of the manuscript to correct remaining errors.

Reviewer #3:
Remarks to the Author:

Comment #1
I am satisfied with the authors' responses to the comments. I would like to also commend them for the extensive additional analyses and writing undertaken to thoughtfully address the concerns of the reviewers. The manuscript is much clearer.

Response
We thank the reviewer for the positive feedback, and we appreciate that the reviewer acknowledges our effort to improve the manuscript.

Decision Letter, second revision:

nature portfolio

Our ref: NG-A58150R1

16th May 2022

Dear Dr. Demontis,

Thank you for submitting your revised manuscript "Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention deficit hyperactivity disorder" (NG-A58150R1). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Genetics, pending minor revisions to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements soon. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics Please do not hesitate to contact me if you have any questions.

Sincerely,
Wei

Wei Li, PhD
Senior Editor
Nature Genetics
New York, NY 10004, USA
www.nature.com/ng

Reviewer #2 (Remarks to the Author):

I am satisfied with the authors' responses and the additional clarity provided as to the limitations of the study.

Author Rebuttal, second revision:

Reviewer #2 (Remarks to the Author):

I am satisfied with the authors' responses and the additional clarity provided as to the limitations of the study.

Response:

We are happy to hear that the remaining doubts have been cleared and we thank the reviewer for the valuable comments, that have improved our manuscript.

Final Decision Letter:

In reply please quote: NG-A58150R2 Demontis

22nd Jun 2022

Dear Dr. Demontis,

I am delighted to say that your manuscript "Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention deficit hyperactivity disorder" has been accepted for publication in an upcoming issue of Nature Genetics.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Genetics style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

You will not receive your proofs until the publishing agreement has been received through our system.

Due to the importance of these deadlines, we ask that you please let us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

Your paper will be published online after we receive your corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the Nature Press Office (press@nature.com) after sending your e-proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NG-A58150R2) and the name of the journal, which they will need when they contact our Press Office.

Before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Genetics. Our Press Office may contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Please note that Nature Genetics is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <a href="https://www.nature.com/nature-portfolio/editorial-policies/self-archiving-and-license-to-publish. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Please note that Nature Portfolio offers an immediate open access option only for papers that were first submitted after 1 January, 2021.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

If you have not already done so, we invite you to upload the step-by-step protocols used in this manuscript to the Protocols Exchange, part of our on-line web resource, natureprotocols.com. If you complete the upload by the time you receive your manuscript proofs, we can insert links in your article that lead directly to the protocol details. Your protocol will be made freely available upon publication of your paper. By participating in natureprotocols.com, you are enabling researchers to more readily reproduce or adapt the methodology you use. Natureprotocols.com is fully searchable, providing your protocols and paper with increased utility and visibility. Please submit your protocol to https://protocolexchange.researchsquare.com/. After entering your nature.com username and password you will need to enter your manuscript number (NG-A58150R2). Further information can be found at https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#protocols

Sincerely,
Wei

Wei Li, PhD
Senior Editor
Nature Genetics
New York, NY 10004, USA
www.nature.com/ng