

Rare coding variants in CHRNA3 associate with reduced daily cigarette smoking across ancestries

Corresponding Author: Dr Veera M. Rajagopal

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Rare coding variants in CHRNA3 from diverse ancestries reduce the likelihood of heavy smoking

NCOMMS-24-80702-T

This is an important paper that provides convincing evidence that rare variants in the CHRNA3 gene, which encodes the $\beta 3$ nicotinic acetylcholine receptor (nAChR) subunit decreases number of cigarettes smoked per day individuals of Mexican ancestry. This observation compliments previous findings showing that common variants in the CHRNA3/A6 gene cluster are also associated with numbers of cigarettes smoked per day. The reported findings are convincing, the analyses appear appropriate, and the manuscript is generally well written. Data related to the consequences of the reported amino acid substitution (Glu284Gly) on the function of $\beta 3$ subunit-containing nAChRs would have strengthened the paper, particularly as $\beta 3$ is thought to largely regulate the assembly of $\alpha 6$ subunit-containing nAChRs. Nevertheless, the data are impactful.

1. Abstract: Please change “nicotine acetylcholine receptors” to nicotinic acetylcholine receptors”.
2. Abstract: Considering changing “smoking addiction” to “tobacco use disorder” or “nicotine addiction”.
3. Introduction: It is surprising the human genetics data lining allelic variation in the CHRNA3/A4/B4 gene locus is not mentioned in the introduction. This is, by far, the set of genes most reproducibly linked with number of cigarettes smoked per day and other indicators of tobacco use disorder.
4. Introduction: The $\alpha 9$ nicotinic acetylcholine subunit is NOT thought to be expressed in the human brain. Instead, this subunit is known for its expression in the cochlea and immune cells, where it plays a role in auditory function and inflammation. Likewise, the $\alpha 10$ subunit is primarily co-expressed in non-neuronal tissues, particularly in the cochlea and immune cells.
5. Results: On page 9, please change “adjacent to not far from” to “adjacent to or not far from”.

Reviewer #2

(Remarks to the Author)

The authors should be commended for investigating the role of rare variants in smoking liability across diverse populations. This paper provides new evidence for CHRNA3, and a deleterious missense variant (p.Glu284Gly) in a Mexican population, which could enable the discovery of new therapeutics. While the analyses seem robust, the lack of replication in another independent Mexican cohort beyond MCPs lowers the enthusiasm. Results are not replicated in other ancestries, albeit larger sample sizes, which is a striking finding that needs further attention in the discussion. Finally, associations seem to be trait-specific, yet the discussion does not allude as to which implications this may have for therapeutic potential; the argument “reward” vs “aversion” seems too speculative/unfounded, especially considering that the authors did not study any tobacco misuse phenotype. Overall, while the work is important and warrants publication, tempering down results, providing

additional interpretations to potential population-specific effects, and integrating the findings with misuse/disorder nicotine GWAS will improve the quality of the manuscript. Thank you for the opportunity to contribute to these efforts, with the modest comments below.

Introduction

1. The introduction focuses on smoking addiction yet the study omits misuse/disorder phenotypes; including relevant GWAS of nicotine dependence and tobacco use disorder (PMID: 38632388, PMID: 33144568) in the introduction, and throughout the manuscript, seems critical to better contextualize the findings.
2. Consider revisiting or further justifying the term “heaviness of smoking”. From the descriptions provided, it is not clear that the authors studied “heavy” smoking (see point 5 below).

Results

3. It would be helpful to see the distribution of cig per day across cohorts, for the full cohort as well as stratified by sex, in the main manuscript.
4. Replicating the findings in an independent Mexican genotype ancestry cohort would strengthen the findings, may another be available? Relatedly, I was curious why the authors did not include All of Us in any of the analyses (beyond looking up allele frequencies).
5. The lack of replication in UKB is striking (“Neither the gene-level nor variant-level testing yielded any significant rare variant associations with cig per day in the UKB, despite a 3.5-fold larger sample size compared to MCPS”). It seems that levels of consumption are substantially higher in UKB than MCPS (15 vs. 5 cig per day); could these discrepancies explain the lack of findings? Are there other demographic differences that could explain the lack of results (e.g., UKB represents generally older individuals; what’s the mean age in MCPS?)?
6. The lack of replication across ancestry groups also deserves further evaluation, particularly given the difficulty of analyzing admixed populations (and also perhaps highly related, with pedigrees?, but I am only speculating) like the ones used in this study. If it is ancestry-specific, the authors should elaborate. Relatedly, how confident are the authors that there is no residual population structure, or relatedness, biasing the analyses?
7. Findings do not replicate across smoking phenotypes. For example, “effect size and P value were modest in the former smokers compared to current smokers”, could also be due to the genetics of former smokers ability to quit smoking. Could the authors include other nicotine dependence or tobacco use disorder phenotypes, if available?
8. When introducing common variants, the authors mention that CHRNA3 is strongly implicated in the biology of smoking addiction but they omit comparing their results with misuse/disorder tobacco GWAS (PMID: 38632388, PMID: 33144568). When discussing GWAS findings, it is unclear why they did not compare findings to Saunders et al (albeit including other samples beyond UKB, but it is much more well-powered; PMID: 36477530).
9. Consider revisiting the definition of “heavy smoking”. If the comparison group is non-smokers, labeling the case group “heavy smokers” can be misleading.
10. Consider removing nominal associations (e.g., “rare variants in CHRNA2 were not significantly associated with cig per day, though a numerical reduction in cig per day was observed”) and instead discuss the lack of replication in the discussion. Similarly, consider tempering down the associations with “we found one missense variant, Ser280Leu, in VSIR, which showed the strongest association” until further replication.

Discussion

11. The arguments related to reward and aversion seem far-fetched and would be best omitted (e.g., “multiple lines of evidence point to $\beta 3$ ’s preferential involvement in aversion over reward pathways”; “Proving this is important as it would suggest that combined inhibition of $\beta 2$ and $\beta 3$ will achieve superior efficacy by targeting both reward and aversion neural mechanisms”).

If the goal is developing better targets for “smoking addiction”, how do the authors interpret the specificity of results for specific phenotypes? Also integrate with the broad spectrum of nicotine traits, including dependence and tobacco use disorder.

12. More elaboration on potential population-specific findings, and overall limitations of the study.

General comments

13. The manuscript is thorough but adding additional details in some areas may be beneficial. Here are some suggestions:
 - Define the cutoff for common vs. rare variants in the introduction.
 - Consider reporting the specific pvalue and MAF for the p.Glu284Gly variant in the results.
 - Define “partial LD or partial independent” (“partial” is too broad/ambiguous).
 - “All the association tests were adjusted for important confounders including age, sex, relatedness, genetic principal components, and any nearby common variant signal”. It is unclear how the authors adjusted for “nearby common variant signals”.
14. While the manuscript is overall well-written, there are a few passages that could be improved. Here are some suggestions:
 - Introduction, rephrase “Among all the human addiction traits, cigarette smoking is best understood in terms of its molecular mechanisms”, “studying their associations with smoking behavior can help understand the roles these subunits play in the addiction neural circuits”.
 - Results - rephrase “commonly referred” in “the top variant at this locus reported by Thorgeirsson et al. (ref19) in 2010 (rs1051730; $r^2=0.86$; $P=2e-69$) and the commonly referred CHRNA5 missense variant (rs16969968; $r^2=0.84$; $P=7e-46$)”, rephrase “ignored” in “early GWASs of smoking behavior have ignored the X chromosome”, the following sentence is too speculative (“may as well play a role”): “However, FAM163B (the closest gene to the new signal) encoding a protein of

unknown function expressed densely in the brain²² may as well play a role in smoking”.

- Discussion, rephrase “low hanging fruit” in “revealing one low-hanging fruit--the p.Glu284Gly missense variant in CHRNA3”; the following argument seems far-reaching “The downstream effects of this protection, for example, decreased risk of smoking-related diseases such as chronic obstructive pulmonary disease (COPD) and lung cancer and any associated health trade-offs will likely surface in the future follow-up studies based on the Mexican population.”

Reviewer #3

(Remarks to the Author)

The present manuscript reports results from an exome-wide association study of smoking primarily in the Mexico City Prospective study, with follow-up analysis conducted in other independent samples. While the significance/implications of the findings are somewhat narrow, I do believe that the findings are novel and important.

The paper is very well written and excellently organized. While rare-variant methods are not my area of expertise, I believe the follow-up analyses to the primary findings was well thought through, thus, I only have minor comments listed below.

- My understanding is that MCPS contains high levels of first-degree relatives and that regenie is not necessarily able to handle high levels of relatedness – is this correct? Or is the extent of relatedness less than I recall?
- Very minor but I believe at one point the authors report 944 carriers of the CHRNA3 missense variants, and at later point in the paper report 946 carriers.
- “Fine-scale ancestry...all the carriers were of Indigenous Mexican ancestry” – in the context of the MCPS sample, I assume all of most of the participants derive some portion of their genome from Indigenous Mexican ancestry (methods report that 66% of ancestry attributable to Indigenous Mexican ancestry in the sample). Does this statement mean the carriers have a significantly higher proportion of Indigenous Mexican ancestry than non-carriers?
- I’m intrigued by the evidence for non-additivity of the CHRNA3 missense variant given relatively little evidence for non-additivity more broadly. Could this be expanded on? The existence of interactive effects seems important for the goal of drug target identification.
- In the section ‘Association of other smoking phenotypes with CHRNA3’, I wonder if testing the association between current and former smokers would be relevant? It seems possible that it would be associated with an individuals’ likelihood of quitting smoking in addition to cigarettes per day – this would also seem to add to the promise of drug target development.
- Figure 3 shows the MAF of the CHRNA3 variant in All of Us American ancestry individuals – is there a reason this sample was not also incorporated in the paper given its size and WGS data available?

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have addressed my concerns. I have no further comments for their consideration. I commend them on a fine contribution.

Reviewer #2

(Remarks to the Author)

The authors have now addressed all my comments.

Reviewer #3

(Remarks to the Author)

I thank the authors for the detailed responses to my comments. All of my major concerns have been addressed. While I think that paper would be strengthened immensely with replication in the All of Us cohort, I understand that it may not yet be possible for researchers in private industry.

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer 1

This is an important paper that provides convincing evidence that rare variants in the CHRNAB3 gene, which encodes the $\beta 3$ nicotinic acetylcholine receptor (nAChR) subunit decreases number of cigarettes smoked per day individuals of Mexican ancestry. This observation compliments previous findings showing that common variants in the CHRNAB3/A6 gene cluster are also associated with numbers of cigarettes smoked per day. The reported findings are convincing, the analyses appear appropriate, and the manuscript is generally well written. Data related to the consequences of the reported amino acid substitution (Glu284Gly) on the function of $\beta 3$ subunit-containing nAChRs would have strengthen the paper, particularly as $\beta 3$ is thought to largely regulate the assembly of $\alpha 6$ subunit-containing nAChRs. Nevertheless, the data are impactful.

We thank the reviewer for taking the time to review our manuscript and providing valuable feedback.

1. Abstract: Please change “nicotine acetylcholine receptors” to nicotinic acetylcholine receptors”.

We have corrected the phrase to “*nicotinic acetylcholine receptors*”.

2. Abstract: Considering changing “smoking addiction” to “tobacco use disorder” or “nicotine addiction”.

We have changed the phrase to “*nicotine addiction*” as per the reviewer’s suggestion.

3. Introduction: It is surprising the human genetics data lining allelic variation in the CHRNA3/A4/B4 gene locus is not mentioned in the introduction. This is, by far, the set of genes most reproducibly linked with number of cigarettes smoked per day and other indicators of tobacco use disorder.

We have added a line in the Introduction (page 3, line 28-29) as follows: “*some of the strongest associations include loci 15q25.1 containing CHRNA5, CHRNA3 and CHRNB4 (ref¹⁴) , 8p11.21 containing CHRNB3 and CHRNA6 (ref¹⁵)*”

4. Introduction: The $\alpha 9$ nicotinic acetylcholine subunit is NOT thought to be expressed in the human brain. Instead, this subunit is known for its expression in the cochlea and immune cells, where it plays a role in auditory function and inflammation. Likewise, the $\alpha 10$ subunit is primarily co-expressed in non-neuronal tissues, particularly in the cochlea and immune cells.

We thank the reviewer for this correction. We have now revised the line (page 3, line 24) as follows: “*There are at least nine nAChR subunits ($\alpha 2-7$, $\beta 2-4$) well known to be expressed in the human brain*”.

5. Results: On page 9, please change “adjacent to not far from” to “adjacent to or not far from”.

We have removed the phrase “*not far from*”. The line reads now “*Four of these five new loci were located on chromosome 15 adjacent to the CHRNA5 cluster.*”

Reviewer 2

The authors should be commended for investigating the role of rare variants in smoking liability across diverse populations. This paper provides new evidence for CHRNA3, and a deleterious missense variant (p.Glu284Gly) in a Mexican population, which could enable the discovery of new therapeutics. While the analyses seem robust, the lack of replication in another independent Mexican cohort beyond MCPS lowers the enthusiasm. Results are not replicated in other ancestries, albeit larger sample sizes, which is a striking finding that needs further attention in the discussion. Finally, associations seem to be trait-specific, yet the discussion does not allude as to which implications this may have for therapeutic potential; the argument “reward” vs “aversion” seems too speculative/unfounded, especially considering that the authors did not study any tobacco misuse phenotype. Overall, while the work is important and warrants publication, tempering down results, providing additional interpretations to potential population-specific effects, and integrating the findings with misuse/disorder nicotine GWAS will improve the quality of the manuscript. Thank you for the opportunity to contribute to these efforts, with the modest comments below.

We thank the reviewers for taking the time review our manuscript and providing valuable feedback.

Introduction

1. The introduction focuses on smoking addiction yet the study omits misuse/disorder phenotypes; including relevant GWAS of nicotine dependence and tobacco use disorder (PMID: 38632388, PMID: 33144568) in the introduction, and throughout the manuscript, seems critical to better contextualize the findings.

We have revised the introduction to cite the studies that the reviewer recommended. Page 3, line 24: “*There are at least nine nAChR subunits ($\alpha 2-7$, $\beta 2-4$) known to be expressed in the human brain, most of which were linked to smoking behavior-related phenotypes such as ever smoking, heaviness of smoking, nicotine addiction and tobacco use disorder through common variant-based genome-wide association studies (GWAS)*”

2. Consider revisiting or further justifying the term “heaviness of smoking”. From the descriptions provided, it is not clear that the authors studied “heavy” smoking (see point 5 below).

We appreciate the reviewer’s point on the ambiguity of term “heavy smoking” and “heaviness of smoking” when referring to cig per day. In the abstract, we rephrased the line “strong protection from heavy smoking” to “significant reduction in daily cigarette consumption”. In the introduction, we rephrased “heaviness of smoking” with “smoking quantity” when describing the cig per day phenotype. Subsequently, throughout the manuscript we refer to this phenotype as “cig per day”.

We note that we have also studied heavy smoking as a binary phenotype, comparing against never smokers. Here, we have defined the heavy smoker and explicitly mentioned that we are comparing against never smokers as in the following lines from results section (page 7, line 26): *“We studied CHRNA3 associations with two binary phenotypes, namely ever vs. never smokers (cases are those who ever smoked regularly in their lifetime and controls are never smokers) and heavy vs. never smokers (cases are those who smoked 10 or more cigarettes per day) in the MCPH participants.”* Notably, this is the definition we used in our earlier work on the discovery of CHRNA2 rare variant association with smoking (Rajagopal et al. Nat Gen 2023). Further, to address reviewer’s concern, now we have rephrased every instance of “ever smoking” and “heavy smoking” in the article with “ever smoking (vs never)” and “heavy smoking (vs never)” to make it clear that we are comparing against never smokers.

As we have also shown that CHRNA3 rare variants reduce the daily cigarette consumption strongly as well as reduce the likelihood of heavy smoking (i.e., smoking more than 10 cig per day) in comparison to never smokers, we would prefer to retain the word “heavy smoking” in the manuscript title.

Results

3. It would be helpful to see the distribution of cig per day across cohorts, for the full cohort as well as stratified by sex, in the main manuscript.

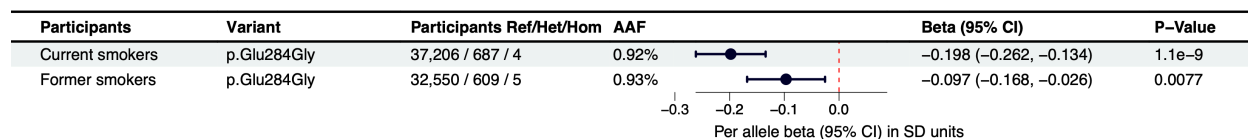
The statistics related to cig per day distribution were provided in the Supplementary Table 1. Now, we also include plots showing the cig per day distribution in males and females separately for current and former smokers in both the cohorts in Supplementary Fig 1a-d.

4. Replicating the findings in an independent Mexican genotype ancestry cohort would strengthen the findings, may another be available? Relatedly, I was curious why the authors did not include All of Us in any of the analyses (beyond looking up allele frequencies).

Our genetic discovery related to CHRNA3 comprise of two findings. The first one, the primary finding, is the protective association between an Indigenous Mexican ancestry-enriched missense variant (p.Glu284Gly) and cig per day. The second one is an overall protective association between deleterious variants in CHRNA3 and cig per day.

To replicate the first finding, like the reviewer suggested, we would need an independent Mexican cohort as the p.Glu284Gly appears to occur mainly in that population, however we do not have access to an independent cohort of Mexican ancestry with exome sequencing data and, to our knowledge, such a cohort does not exist currently. Like many other non-European populations, Mexican ancestry has been under-represented in genetic studies. We agree with the reviewer that the All of US cohort would be an alternative as it includes a substantial number of individuals of admixed American ancestries. However, we note that the All of Us data was only recently made available to industry researchers, and our company is still in the process of requesting access to this dataset. Hence, we don’t have access to All of Us data to include in this manuscript at this time.

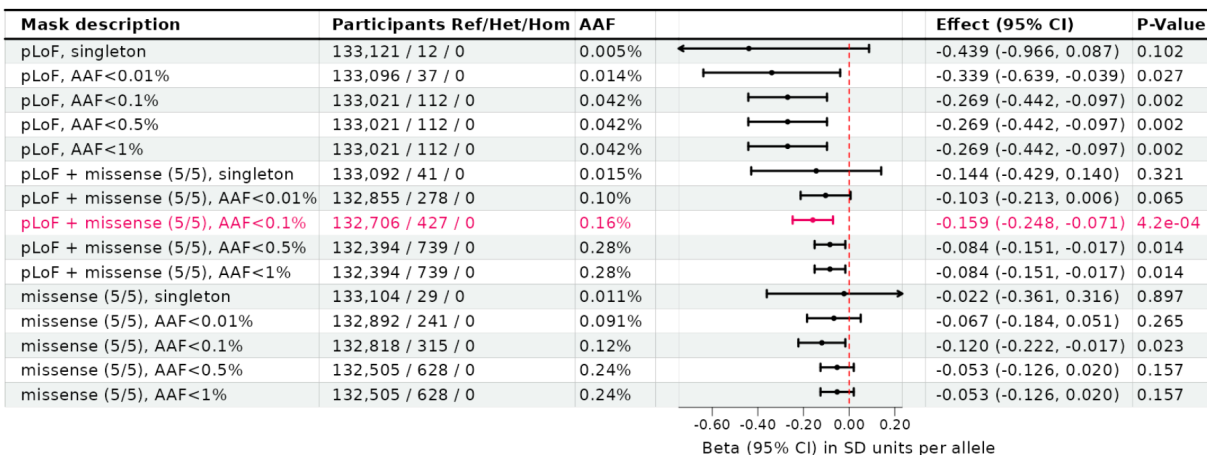
Nevertheless, we did replicate the finding in an independent sample from the MCPS cohort. We tested the association of p.Glu284Gly with cig per day in 33,162 individuals who were former smokers and reported on the number of cigarettes they smoked per day in the past. Please note that these individuals are independent of the 37,897 current smokers included in our primary analysis. As we reported in our manuscript (also, summarized in the below forest plot), we replicated the protective association between p.Glu284Gly with cig per day in former smokers at nominal significance (Effect = -0.10 SD; $P=5.4e-04$).



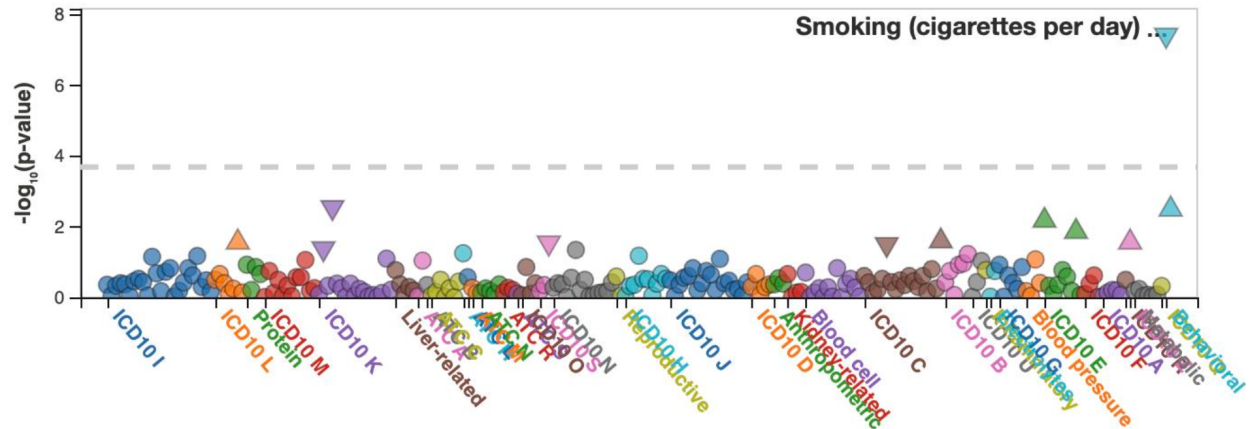
Regarding the second finding (protective association between rare deleterious variants in CHRN3 and cig per day), as reported in our manuscript, we replicated it in two independent cohorts: UK Biobank and Japan biobank.

In the UK Biobank, a rare-variant aggregate association analysis showed a nominally significant protective association between rare pLOFs and deleterious missense variants in CHRN3 in aggregate and cig per day (Effect = -0.16; $P=4.2e-04$, as shown in the forest plot below from Supplementary Fig. 8; highlighted row). Importantly, we also show a nominally significant association between pLOFs (AAF<1%) in aggregate and cig per day in the UK Biobank (Effect = -0.27; $P=0.002$).

Supplementary Fig. 8



In the Japan Biobank, we observe a genome-wide significant ($p<5e-8$) protective association between an East Asians-enriched pLOF (splice donor; c.52+1G>A) and cig per day (Effect = -0.19; $P=3.9e-8$). Importantly, cig per day was the top associated trait with this variant in the Japan Biobank as shown in the below PheWAS plot from the Japan biobank online repository.



To summarize, even though we did not have an independent Mexican ancestry cohort to replicate the genetic association between p.Glu284Gly and cig per day, we have replicated the variant's association in an independent sample from the MCPS cohort, and we have demonstrated additional evidence using an orthogonal class of variants in the UK Biobank and Japan biobank showing that deleterious variants in *CHRNA3* are associated with reduced cig per day (associations in UK Biobank and Japan Biobank driven by pLOFs while the association in MCPS driven by deleterious missense variants).

5. The lack of replication in UKB is striking (“Neither the gene-level nor variant-level testing yielded any significant rare variant associations with cig per day in the UKB, despite a 3.5-fold larger sample size compared to MCPS”). It seems that levels of consumption are substantially higher in UKB than MCPS (15 vs. 5 cig per day); could these discrepancies explain the lack of findings? Are there other demographic differences that could explain the lack of results (e.g., UKB represents generally older individuals; what’s the mean age in MCPS?)?

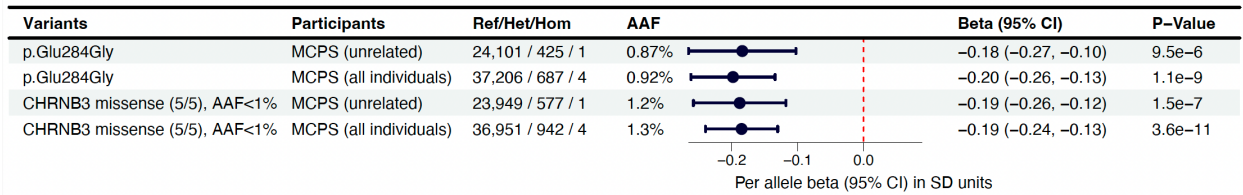
The quoted sentence by the reviewer (“Neither the gene-level ...”) was written in the context of gene discovery rather than replication. When comparing the exome-wide findings overall, we did not observe any exome-wide significant associations in the UK Biobank, despite a 3.5-fold larger sample size compared to MCPS. This observation was noted to emphasize the importance of studying non-European populations that are often enriched for certain rare deleterious variants (that are absent or extremely rare in the European populations) leading to novel gene discoveries. We note again that the CHRNA3 rare variant associations in aggregate with cig per day did replicate in the UK Biobank, albeit at nominal significance (as one would expect in the absence of drifted variants). The p.Glu284Gly variant itself is absent and therefore cannot be tested in the UK Biobank as the cohort is predominantly of individuals of British European ancestry. Hence, our perspective is not that there is “lack of replication in UKB”, rather an inability to check for replication, as this variant is not present at comparable levels.

The reviewer is correct about the levels of cigarette consumption being substantially higher in the UK Biobank compared to MCPS, a previously known difference believed to be primarily due to cultural factors. The mean age of the participants of the MCPS and UK Biobank is reported in Supplementary Table 1: 52.6 yr in the MCPS and 56.7 yr in the UK Biobank.

6. The lack of replication across ancestry groups also deserves further evaluation, particularly given the difficulty of analyzing admixed populations (and also perhaps highly related, with pedigrees?, but I am only speculating) like the ones used in this study. If it is ancestry-specific, the authors should elaborate. Relatedly, how confident are the authors that there is no residual population structure, or relatedness, biasing the analyses?

We again have a different perspective regarding a “lack of replication across ancestry groups”, for the reasons we have discussed in our responses to comments 5 and 6. We note again that we have demonstrated evidence across three ancestries (Indigenous Mexican, European and East Asian) that deleterious variants in CHRNA3 are associated with reduced cigarette consumption; only the variants contributing to associations were different in the three ancestries (p.Glu284Gly in Indigenous Mexican, rs147306385 (splice donor) in East Asian, and multiple pLOFs and deleterious missense variants in aggregate in the UK Biobank).

Regarding population structure and relatedness, our in-house statistical software, REGENIE, used for the association testing handles effectively both population structure, admixture, and relatedness (Mbatchou et al. Nat Gen 2021). We agree with the reviewer that the MCPS cohort involves a significantly larger number of related individuals. So, we tested the association of CHRNA3 rare variants with cig per day after excluding related individuals (up to 3rd degree). We observed a significant protective association between CHRNA3 rare variants and cig per day even when analyzing unrelated individuals, as summarized in the forest plot below (the relative drop in the statistical significance in unrelated sample is expected due to the drop in sample size). We added this analysis to the main manuscript (page 5, lines 19-22).



7. Findings do not replicate across smoking phenotypes. For example, “effect size and P value were modest in the former smokers compared to current smokers”, could also be due to the genetics of former smokers ability to quit smoking. Could the authors include other nicotine dependence or tobacco use disorder phenotypes, if available?

We did show a statistically significant association between CHRNA3 rare variants and cig per day in the former smoker (refer to our response to comment 4). Here the phenotype (cig per day) is same, only the study participants are different (current smokers vs former smokers). The reasons for the modest effect size in former smokers are not clear. We believe the difference could be due to multiple reasons, such as lack of accuracy in reporting (people tend to report number of cigarettes they smoke per day currently more accurately than the number of cigarettes they smoked previously due to recall bias), or environmental differences, for example,

people who have successfully quit smoking would have had reasons other than their genetics (such as doctors' advice) for reduced cigarette consumption thereby diluting the genetic signal.

We agree with the reviewer that it would be valuable to test the association with clinical phenotypes such as nicotine dependence or a diagnosis of tobacco use disorder. However, we do not have such phenotypes available for the MCPS participants. Hence, we could not study those phenotypes in this cohort.

8. When introducing common variants, the authors mention that CHRNA3 is strongly implicated in the biology of smoking addiction but they omit comparing their results with misuse/disorder tobacco GWAS (PMID: 38632388, PMID: 33144568). When discussing GWAS findings, it is unclear why they did not compare findings to Saunders et al (albeit including other samples beyond UKB, but it is much more well-powered; PMID: 36477530).

We have added the citation related to tobacco use disorder in the Introduction (refer to our response to comment 1). For common variants analysis, our primary comparison was with the UK Biobank rather than the published multi-ancestry, meta-analysis results by Saunders et al. because we wanted to emphasize common variant discoveries in the Indigenous Mexican ancestry in comparison to commonly studied European ancestry. However, when defining novel signals, we did compare with published large GWASs of smoking phenotypes including Saunders et al. For example, when reporting the common variant signals 15q25.1 and 9q34.2 we report that our signals were in LD with the signals reported previously by Saunders et al.

9. Consider revisiting the definition of "heavy smoking". If the comparison group is non-smokers, labeling the case group "heavy smokers" can be misleading.

We have addressed this in our response to the reviewers' comment 2.

10. Consider removing nominal associations (e.g., "rare variants in CHRNA2 were not significantly associated with cig per day, though a numerical reduction in cig per day was observed") and instead discuss the lack of replication in the discussion. Similarly, consider tempering down the associations with "we found one missense variant, Ser280Leu, in VSIR, which showed the strongest association" until further replication.

We have removed the sentence "*though a numerical reduction in cig per day was observed*". Regarding VSIR, we have rephrased the sentence to "*we found one missense variant, Ser280Leu, in VSIR, which showed a significant association with cig per day*".

Discussion

11. The arguments related to reward and aversion seem far-fetched and would be best omitted (e.g., "multiple lines of evidence point to β 3's preferential involvement in aversion over reward pathways"; "Proving this is important as it would suggest that combined inhibition of β 2 and β 3 will achieve superior efficacy by targeting both reward and aversion neural mechanisms"). If the goal is developing better targets for "smoking addiction", how do the authors interpret the

specificity of results for specific phenotypes? Also integrate with the broad spectrum of nicotine traits, including dependence and tobacco use disorder.

We appreciate the reviewer's concern about speculation regarding reward and aversion pathways. We have revised the reward and aversion pathway-related discussion in the manuscript (page 11, lines 1-25). We now focus only on our observations on the differential effects of *CHRNA3* rare genetic associations, established evidence (high expression in medial habenula, beta3 mouse knockout experiments), similarity to other nAChR literature (alpha4 mouse knockout experiments). We removed our interpretative statements such as "multiple lines of evidence point to beta3's preferential involvement in aversion over reward pathways". We have also removed speculative statements about combined inhibition of beta3 and beta3 inhibition.

12. More elaboration on potential population-specific findings, and overall limitations of the study.

We have added a limitation section in the discussion, including elaboration on population-specific findings.

Page 11, lines 33-44: *"Our study has several limitations. Although we replicated the findings in former smokers from the MCPS cohort and found overall consistent associations of CHRNA3 rare variants with cig per day across different ancestries, we lacked an independent Mexican ancestry cohort for direct replication of the p.Glu284Gly variant. Our phenotypic characterization focused primarily on cig per day rather than comprehensive clinical measures of nicotine dependence or tobacco use disorder, potentially limiting the direct assessment of clinical implications of the genetic findings. While the genetic associations were robust, functional characterization of CHRNA3 rare variants was not performed to precisely understand the molecular mechanisms of through which loss of beta3 subunit influences smoking behavior. Finally, population specific nature of the key variants (p.Glu284Gly in Indigenous Mexicans, c.52+1G>A in East Asians) limits replication of specific variant associations across ancestries. However, the consistent phenotypic impact of these different variants from diverse ancestries strengthens the overall conclusion that disrupting CHRNA3 function reduces cigarette consumption. This demonstrates that, although genetic variants diverge, underlying biology, and therefore the translational impact, often converge across ancestries."*

General comments

13. The manuscript is thorough but adding additional details in some areas may be beneficial. Here are some suggestions:

- Define the cutoff for common vs. rare variants in the introduction.
- Consider reporting the specific pvalue and MAF for the p.Glu284Gly variant in the results.
- Define "partial LD or partial independent" ("partial" is too broad/ambiguous).
- "All the association tests were adjusted for important confounders including age, sex, relatedness, genetic principal components, and any nearby common variant signal". It is unclear how the authors adjusted for "nearby common variant signals".

- We have defined the cut off for common ($MAF \geq 1\%$) and rare ($MAF < 1\%$) variants in the introduction (page 3, lines 27 and 38).
- Association statistics of p.Glu284Gly was reported in Figure 4 and Supplementary Table 5. We now also report in the results section: page 4, line 42 reads “Testing individual rare coding variants across the exome, we identified a missense variant, p.Glu284Gly (rs75384358), in CHRNA3 significantly associated with cig per day at exome-wide significance ($\beta = -0.19$; $P = 1.1 \times 10^{-9}$; Supplementary Fig. 3a).” We have reported MCPS-specific and Indigenous Mexican-specific MAF of p.Glu284Gly in results, page 5, lines 24-26: *“The MAF of p.Glu284Gly was 0.9% in the MCPS cohort. Fine-scale ancestry analysis showed that all the carriers were of Indigenous Mexican ancestry. The haplotype-resolved Indigenous Mexican-specific MAF of p.Glu284Gly was 1.4%”*.
- We have rephrased “partial LD” with the phrase “moderately correlated”. Further, revised the section on “partly independent” to bring more clarity as follows (page 7, lines 2-7): *“We studied this signal using conditional analyses and found that our common variant signal retains statistical significance after conditioning on the previously reported signal by Saunders et al. ($P = 1.5 \times 10^{-6}$). Similarly, when conditioning on our top variant, the Saunders et al. signal remains statistically significant in UKB ($P = 0.0004$). These results indicate that, despite a moderate correlation, the two associations at the CHRNA3 locus may represent independent signals (Supplementary Fig. 11).”*
- We have added a line in the methods section “Genetic association analysis”, describing how we adjusted for nearby common variant signals (page 15, lines 18-19): *“All rare variant association analyses were adjusted for nearby common variant signals by including all the fine-mapped index variant genotypes as covariates in the association runs.”*

14. While the manuscript is overall well-written, there are a few passages that could be improved. Here are some suggestions:

- Introduction, rephrase “Among all the human addiction traits, cigarette smoking is best understood in terms of its molecular mechanisms”, “studying their associations with smoking behavior can help understand the roles these subunits play in the addiction neural circuits”.
- Results - rephrase “commonly referred” in “the top variant at this locus reported by Thorgeirsson et al. (ref19) in 2010 (rs1051730; $r^2 = 0.86$; $P = 2 \times 10^{-69}$) and the commonly referred CHRNA5 missense variant (rs16969968; $r^2 = 0.84$; $P = 7 \times 10^{-46}$)”, rephrase “ignored” in “early GWASs of smoking behavior have ignored the X chromosome”, the following sentence is too speculative (“may as well play a role”): “However, FAM163B (the closest gene to the new signal) encoding a protein of unknown function expressed densely in the brain²² may as well play a role in smoking”.

We appreciate these suggestions by the reviewer. We have now rephrased the sentences as follows.

- “Among all the human addiction traits ...” to “Cigarette smoking has been extensively characterized at the molecular level compared to other addiction-related behaviors”.

- “studying their associations ...” to “investigating associations with smoking behavior provides insight into their functional roles within addiction-related neural pathways”
- “commonly referred” to “well established”
- “ignored” to “did not include analyses of”
- “may as well play a role in smoking” to “warrants further investigation for its potential involvement in smoking behavior”

- Discussion, rephrase “low hanging fruit” in “revealing one low-hanging fruit--the p.Glu284Gly missense variant in CHRNA3”; the following argument seems far-reaching “The downstream effects of this protection, for example, decreased risk of smoking-related diseases such as chronic obstructive pulmonary disease (COPD) and lung cancer and any associated health trade-offs will likely surface in the future follow-up studies based on the Mexican population.”

We have rephrased “low hanging fruit” to “notable discovery”. We have removed the line “The downstream effects Mexican population” as per reviewer’s suggestion.

Reviewer 3

The present manuscript reports results from an exome-wide association study of smoking primarily in the Mexico City Prospective study, with follow-up analysis conducted in other independent samples. While the significance/implications of the findings are somewhat narrow, I do believe that the findings are novel and important.

The paper is very well written and excellently organized. While rare-variant methods are not my area of expertise, I believe the follow-up analyses to the primary findings was well thought through, thus, I only have minor comments listed below.

We thank the reviewers for taking the time review our manuscript and providing valuable feedback.

- My understanding is that MCPS contains high levels of first-degree relatives and that regenie is not necessarily able to handle high levels of relatedness – is this correct? Or is the extent of relatedness less than I recall?

The reviewer is correct that the MCPS cohort is enriched with closely related individuals. We have addressed this point in our response to comment 6 from reviewer 2. Briefly, we have now added an analysis where we excluded individuals related up to 3rd degree and show that we still find a statistically significant protective association between CHRNA3 rare variants and cig per day.

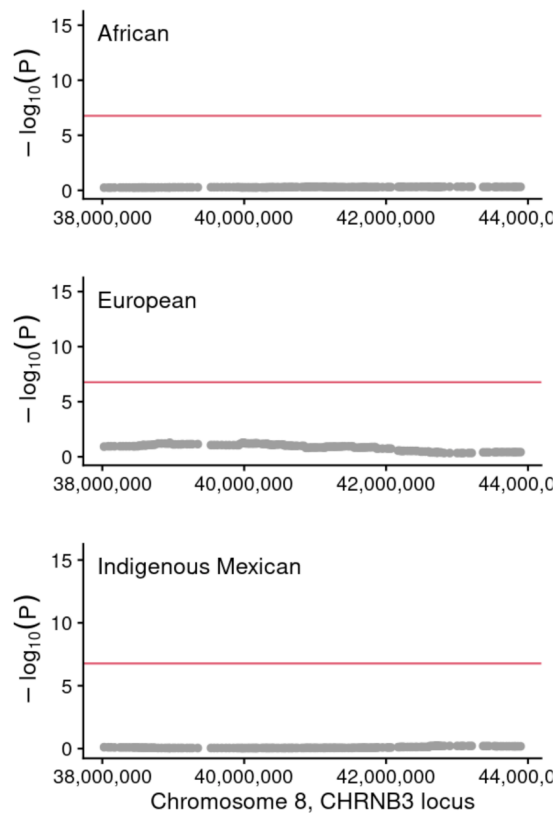
- Very minor but I believe at one point the authors report 944 carriers of the CHRNA3 missense variants, and at later point in the paper report 946 carriers.

We thank the reviewer for finding the error. The correct number is 944 carriers. We have corrected the text accordingly.

- “Fine-scale ancestry...all the carriers were of Indigenous Mexican ancestry” – in the context of the MCPS sample, I assume all or most of the participants derive some portion of their genome from Indigenous Mexican ancestry (methods report that 66% of ancestry attributable to Indigenous Mexican ancestry in the sample). Does this statement mean the carriers have a significantly higher proportion of Indigenous Mexican ancestry than non-carriers?

We appreciate this comment, and we have indeed tested for this. We specifically tested if the genomic location around p.Glu284Gly variant has more Indigenous Mexican ancestry than genome-wide average and found that was not the case. We have reported this in the manuscript in Supplementary 21 (shown below). The figure shows the P values for deviation of local ancestry proportions compared to genome-average in the CHRNA3 locus. We did not see a significant deviation. Hence, currently we don't see any evidence for selection at the CHRNA3 locus. While it is true that almost all the carriers were of Indigenous Mexican ancestry (predicted using genome-wide data), we don't see any evidence of selection at the locus itself.

Supplementary Fig. 21



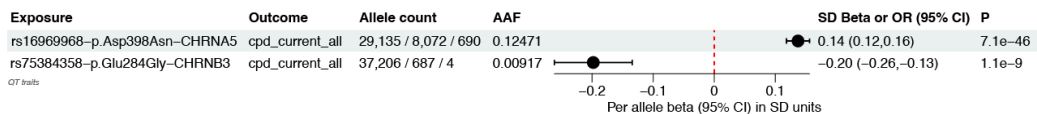
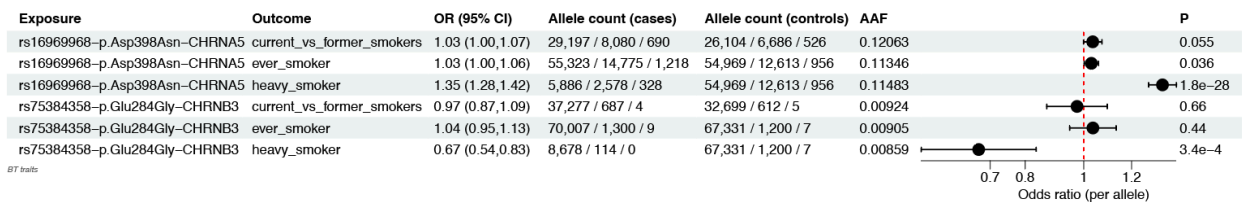
The figure displays the P values for deviation of local ancestry proportions (African, European and Indigenous Mexican ancestries) from the global ancestry proportion in genomic region surround CHRNA3. For more details on local ancestry estimation and derivation of P values, refer to [Ziyatdinov et al. Nature 2023](#)

• I'm intrigued by the evidence for non-additivity of the CHRNA5 missense variant given relatively little evidence for non-additivity more broadly. Could this be expanded on? The existence of interactive effects seems important for the goal of drug target identification.

We believe that this comment is regarding the heterozygous vs homozygous effect size where we discussed in the article that the homozygous effect size is far greater than what would be expected in an additive model. This is often seen when a deleterious variant has a loss of function consequence, in which case, homozygous genotype might have a much stronger effect—due to near complete inactivation of the gene—than the heterozygous effect. However, we do not have functional evidence that p.Glu284Gly is a loss of function variant currently. Hence, we merely noted this observation in the manuscript but refrained from speculating based on the observation.

• In the section ‘Association of other smoking phenotypes with CHRNA5’, I wonder if testing the association between current and former smokers would be relevant? It seems possible that it would be associated with an individuals’ likelihood of quitting smoking in addition to cigarettes per day – this would also seem to add to the promise of drug target development.

As per the reviewer’s advice we have now added included association of CHRNA5 rare variants with current vs former smokers in the MCPS in the manuscript (Supplementary Figure 14). We did not find a significant association with current vs former smokers. We note that this is line with what we see for CHRNA5 variant as well where the missense variant rs16969968 associates strongly with cig per day (and consequently, heavy smoker, see fig below) but has minimal association with ever smoking (vs never smoking) or current (vs former smoking). As we have discussed in the article, the CHRNA5 seems to be specifically influence the aversive effects of cigarette rather than rewarding effects.



• Figure 3 shows the MAF of the CHRNA5 variant in All of Us American ancestry individuals – is there a reason this sample was not also incorporated in the paper given its size and WGS data available?

We have addressed this point in our response to comment 4 from reviewer 2. At the time of this analysis, we did not have access to All of Us cohort, hence we couldn’t include it in our analysis.