

Peer Review Information

Journal: Nature Genetics

Manuscript Title: Global impact of unproductive splicing on human gene expression

Corresponding author name(s): Dr Yang (I) Li

Reviewer Comments & Decisions:

Decision Letter, initial version:
--

26th Feb 2024

Dear Dr Li,

Your Article, "Global impact of unproductive splicing on human gene expression levels and traits" has now been seen by 2 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, including with the chief editor, with a view to identifying key priorities that should be addressed in revision and sometimes overruling referee requests that are deemed beyond the scope of the current study. In this case, we ask you to address Reviewers' comments in full. We hope that you will find the prioritized set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

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 www.springernature.com/orcid.

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

Sincerely,
Chiara

Chiara Anania, PhD
Associate Editor
Nature Genetics
<https://orcid.org/0000-0003-1549-4157>

Referee expertise:

Referee #2: genomics

Referee #3: signed review

Reviewers' Comments:

Reviewer #2:

Remarks to the Author:

This is a very interesting, thought-provoking and thorough study that presents multiple lines of evidence that unproductive splicing accounts for around 15% of transcripts from the average human gene, exceeding 50% for many long genes expressed at low levels. This level of unproductive transcription, estimated by nascent RNA sequencing, is substantially higher than previous estimates and evidence is presented that common genetic variants quite frequently impact gene expression and clinical traits via altering the production of these unproductive mRNAs. The results also have a potentially important translational impact: the existence of multiple cryptic splice sites in the vast majority of human genes implies that there are abundant targets that can be manipulated by splice-switching molecules for therapeutic effects. I think these results and ideas will be of very high interest to the readership of Nature Genetics.

Suggestions and comments

Many premature stop codons do not actually trigger NMD. It would be good that the authors take into account the more complete rules of NMD that have now been worked out (Lindeboom Nat Gen 2019, 2016) to better annotate with transcripts do - and which do not - strongly trigger NMD and how this impacts their conclusions.

Fig. 1E is confusing. how is the quintile binning used?

Reviewer #3:

Remarks to the Author:

Fair et al. examines the extensive impact of unproductive splicing on gene expression levels and complex traits, highlighting that alternative splicing (AS) significantly influences gene expression by producing unproductive transcripts targeted for degradation via nonsense-mediated decay (NMD). It presents findings from population-scale genomic data analysis, revealing a threefold increase in unproductive splicing compared to previous estimates, with 15% of transcript molecules from protein-coding genes being unproductive. The study also explores genetic variation across cell lines, showing that AS-related loci associated with GWAS traits often involve NMD-induced expression level differences rather than differences in protein isoform usage. Utilizing the splice-switching drug risdiplam, the study demonstrates that a substantial portion of drug-induced isoforms are targeted by NMD, suggesting that aberrant splicing predominantly affects expression levels. This challenges the

traditional view of AS's role in proteome diversification, proposing that its impact through NMD-induced gene expression changes is more significant than previously understood. The manuscript is clearly and logically written and the figures are of high quality. Overall, this study presents an important advance in our understanding of the role of AS in complex traits, although several questions need to be addressed and clarified:

1. The statistics about symmetric exons is confusing. The authors found that <40% of exons in the naRNA are symmetric (60% unsymmetric), and this number increases to 55% in steady-state RNA (45% unsymmetric). This means that 15% of exons (all of which are non-symmetric) are degraded by NMD. Does this observation suggest that the remaining 45% of unsymmetric exons are not targets of NMD? What percentage of these unsymmetric exons lead to normal isoforms (isoforms without premature stop codons)? As some of these unsymmetric exons necessarily lead to premature stop codons, why do they escape NMD?

Also, the author should consider toning down this statement that "enrichment for frame-preserving exons in steady-state RNA is largely a product of NMD surveillance" because, if the reported statistics is correct, only 15% of unsymmetric exons are targeted by NMD, compared to the 45% that is not.

2. Why is 1.5% - 2.5% of splicing events assumed to be unproductive? The authors need to explain the reasoning behind this assumption. Otherwise, it makes the estimate of 11% - 18% look like curve-fitting to the LRS data.

3. The log2 fold enrichment of colocalized SNP in genomic annotation is a key observation in this study and is more definitive than figure 3c. However, the results of this figure raise some questions: 1) the author observed in figure 3c that APA accounts for a mere 5%, yet the colocalized SNPs are most strongly enriched in APA regions. Isn't this a bit self-contradictory? 2) only splice acceptors are enriched, and splice donors are not. What would cause this discrepancy? Are splice acceptors the primary determinants of nonsymmetric exons?

In addition, instead of using enrichment, why not directly quantify the percentage of colocalized SNPs (along with their LD proxies for the lack of causal annotation) falling into each genomic annotation? That would be a more direct and definitive observation than figure 3c.

4. The D2HGDH example, as the authors pointed out, seems to be a case of multiple independent causal variants modulating the expression of the same gene. The HyPrColoc method used by this paper considers a single causal variant, thus unable to detect the colocalization between u-sQTL and eQTL. Many colocalization software (the latest Coloc for example) is able to model multiple causal variants. Although this D2HGDH example is a valid example, a more intriguing question would be after multiple causal variants have been taken into consideration, are there still u-sQTLs that don't colocalize with eQTL? If so, why?

Minor comments:

- The manuscript does not have line numbers, making it a bit difficult to comment on specific lines.
- Mostgenes -> most genes
- Figure 1 caption: as heatmap -> as a heatmap
- Protein coding gene -> protein-coding gene
- The author might consider citing [PMID: 31043754] when using colocalization scatter plots.
- In addition to the ALS and COVID-19 example, [PMID: 31123710] is another experimentally

validated example of NMD-led eQTL affecting complex diseases.

Sincerely,
Boxiang Liu

Author Rebuttal to Initial comments

Reviewer #3:

Remarks to the Author:

Fair et al. examines the extensive impact of unproductive splicing on gene expression levels and complex traits, highlighting that alternative splicing (AS) significantly influences gene expression by producing unproductive transcripts targeted for degradation via nonsense-mediated decay (NMD). It presents findings from population-scale genomic data analysis, revealing a threefold increase in unproductive splicing compared to previous estimates, with 15% of transcript molecules from protein-coding genes being unproductive. The study also explores genetic variation across cell lines, showing that AS-related loci associated with GWAS traits often involve NMD-induced expression level differences rather than differences in protein isoform usage. Utilizing the splice-switching drug risdiplam, the study demonstrates that a substantial portion of drug-induced isoforms are targeted by NMD, suggesting that aberrant splicing predominantly affects expression levels. This challenges the traditional view of AS's role in proteome diversification, proposing that its impact through NMD-induced gene expression changes is more significant than previously understood. The manuscript is clearly and logically written and the figures are of high quality. Overall, this study presents an important advance in our understanding of the role of AS in complex traits, although several questions need to be addressed and clarified:

1. The statistics about symmetric exons is confusing. The authors found that <40% of exons in the naRNA are symmetric (60% unsymmetric), and this number increases to 55% in steady-state RNA (45% unsymmetric). This means that 15% of exons (all of which are non-symmetric) are degraded by NMD. Does this observation suggest that the remaining 45% of unsymmetric exons are not targets of NMD? What percentage of these unsymmetric exons lead to normal isoforms (isoforms without premature stop codons)? As some of these unsymmetric exons necessarily lead to premature stop codons, why do they escape NMD?

Also, the author should consider toning down this statement that “enrichment for frame-preserving exons in steady-state RNA is largely a product of NMD surveillance” because, if the reported statistics is correct, only 15% of unsymmetric exons are targeted by NMD, compared to the 45% that is not.

Thank you for the feedback. Upon re-examination, we agree that this section was confusing for readers. We have revised this section about symmetric exons (**Line#196**) and included **a new supplemental figure (Figure S7)** which we hope will clarify our findings. To more directly address the reviewer's questions we provide the following explanation of our revisions:

Firstly, we wish to be clear that we are thinking about this analysis about symmetric exons in qualitative way, and we do not mean to quantitatively conclude that 15% of exons are degraded by NMD from this analysis (though, in other sections of the manuscript, we do suggest that ~15% of transcript molecules are degraded by NMD, as evidenced by other analyses). For this analyses about symmetry, the conclusion we wished to highlight is that the apparent enrichment for symmetric (frame-preserving exons) amongst alternatively spliced exons (previously documented in (Merkin et al. 2012; Mudge et al. 2011; Barbosa-Morais et al. 2012; Long, Rosenberg, and Gilbert 1995; Bénitère, Necsulea, and Duret 2022)) is a product of NMD-surveillance that degrades frame-shifting exons after splicing, rather than evidence that only frame-preserving exons are preferentially alternatively spliced by the spliceosome (described in Figure S7A). Indeed, while we can recapitulate the enrichment for symmetric exons amongst alternatively spliced exons (defined as $PSI \sim 50\%$) in steady state-RNA, this enrichment is largely absent in naRNA or after NMD knockdown (Figure 1D).

We want to emphasize that our analysis does speak to whether transcripts containing non-symmetric exons will be degraded or not. In fact, many non-symmetric exons are constitutively spliced, highly-used, and do not induce NMD. To further highlight this point, we expanded the x-axes in Figure 1D into FigureS7B, which we believe requires such in-depth explanation that we chose to include it is a supplement rather than outright replacing Figure1D: FigureS7B highlights that constitutive exons ($PSI \sim 100\%$) are not strongly enriched for symmetric exons. This is to be expected, because constitutive exons need not to be symmetric to maintain reading frame as they are (by definition) always spliced-in. In other words, asymmetric exons are not expected to induce a PTC if they are constitutive. However, we note that the distinction between “constitutive” and “alternative” can be blurry, and can change depending on whether we are only considering a single cell type (as we are in this analysis) versus exons that may be tissue/cell-type AS-switch exons. Thus, if some AS exons ($PSI \sim 50\%$) are misclassified as constitutive ($PSI \sim 100\%$) or unspliced ($PSI \sim 0\%$), the level of enrichment for symmetry amongst AS exons would be reduced. For these reasons, we caution against quantitative interpretation of the difference in this enrichments (eg, the reviewer's confusion that 60% asymmetric exons in naRNA, minus 45% asymmetric exons in steady-state RNA, equals 15% of asymmetric exons not targeted by NMD). **In conclusion, we decided to rewrite this section and we added FigureS7A, which we believe clarifies our thinking while avoiding over-interpretation of our analysis about exon symmetry.** For convenience, we pasted the revised section with new supplemental figure below:

We further describe the abundance of unproductive splicing revealed by nascent RNA in a qualitative manner by analyzing exons whose length is divisible by three (“symmetric exons”), and thus are frame-preserving whether the exon is skipped or included. Previous studies using steady-state RNA-seq data found that highly-included and conserved alternatively spliced exons (“cassette exons”) are biased towards being symmetric^{1–4,59}, suggesting that there is selective pressure to maintain coding frame for highly-included cassette exons (**Figure S7A**). Indeed, ~55% of highly-included cassette exons are symmetric compared to only ~35% of rarely-used cassette exons (Percent spliced-in, PSI < 1%). However, we find that in naRNA the fraction of symmetric exons is low, under 40%, even for the most highly-included cassette exons (**Figure 1D**, **Figure S7B**). We find similar results in steady-state RNA after knockdown of NMD factors (**Figure 1D**, **Figure S7B**). These observations show that in the absence of NMD, there is no bias for alternatively spliced exons to be symmetric. We conclude that the enrichment for frame-preserving cassette exons in steady-state RNA is largely a product of NMD surveillance rather than a direct consequence of natural selection for high usage of symmetric cassette exons (**Figure S7B**). These findings question the widely-held belief that the main function of AS is protein diversification.

2. Why is 1.5% - 2.5% of splicing events assumed to be unproductive? The authors need to explain the reasoning behind this assumption. Otherwise, it makes the estimate of 11% - 18% look like curve-fitting to the LRS data.

Thank you for the suggestion and we agree that these numbers were not properly introduced in the previous version of our manuscript. We have now clarified where 1.5%-2.5% comes from (two highlighted sections starting at **line#271 and line#281**). We also will summarize that here: While in short read data we find that (across naRNA samples, a median of) 2.3% of splice junctions are unproductive junctions, which under the simple binomial model would yield ~17% of unproductive full-length transcripts for transcripts with 8 junctions. In long read data, rather than 2.3% of unproductive junctions, we observe a range from 1.8% to 2.3% of unproductive junctions, which we (unnecessarily, but conservatively) round to 1.5-2.5%, to yield a binomial expectation of 11-18% of unproductive full length transcripts. We believe these simple binomial models with simple numbers are faithful to the data while also best communicating the general idea that we wish to convey: that a relatively low mis-splicing rate can compound across multi-intronic genes. Whether we use a per-intron mis-splicing rate of 2.3%, or 1.5-2.5%, 1.8%-2.3% to illustrate the theoretical (binomial model) per-transcript mis-splicing

rate, we still arrive at the conclusion that ~15% of transcripts from a typical gene are unproductive, which is a large increase from previous estimates based on steady-state RNA (Pickrell 2010).

Furthermore, we note that we expanded this analysis full-length reads from long-read data (**new Figure S17C**, see point 1) to again conclude that previous estimates of unproductive splicing using steady-state RNA are hugely downwardly biased.

3. The log2 fold enrichment of colocalized SNP in genomic annotation is a key observation in this study and is more definitive than figure 3c. However, the results of this figure raise some questions: 1) the author observed in figure 3c that APA accounts for a mere 5%, yet the colocalized SNPs are most strongly enriched in APA regions. Isn't this a bit self-contradictory? 2) only splice acceptors are enriched, and splice donors are not. What would cause this discrepancy? Are splice acceptors the primary determinants of nonsymmetric exons?

In addition, instead of using enrichment, why not directly quantify the percentage of colocalized SNPs (along with their LD proxies for the lack of causal annotation) falling into each genomic annotation? That would be a more direct and definitive observation than figure 3c.

Thank you for this suggestion. The apparent contradiction in question 1) is explained by differences in the number of splice sites and APA sites in each gene. A typical gene may have >10 splice sites, and only one or a few APA sites. Thus, while splice sites may explain more eQTLs, eQTL signals may be similarly “enriched” for splice sites as APA sites (when enrichment is defined by the fraction of eQTL SNPs in a feature set divided by the number or nt length of the feature set). To highlight this, we performed an analysis that is similar to the reviewer’s suggestion of quantifying the percentage of colocalized SNPs in each annotation (**New supplemental figure panel, Figure S15D**). That is, we considered the same set of post-transcriptional eQTLs and transcriptional eQTLs that we used for enrichment analysis (Figure S15C) and calculated the relative proportion of fine-mapped posterior inclusion probability in each feature set (considering splice site-related and APA site feature sets). **In this analysis it becomes clear that splice site regions account for more post-transcriptional eQTL signal than APA sites.**

As to the reviewer’s second question (why are splice acceptors significantly enriched but not splice donors), we note that we are testing differences from the null (no enrichment) and we have not made explicit comparisons between splice acceptors and donors. That is to say, from our analysis we cannot conclude whether splice donor variants contribute more to unproductive sQTLs than splice acceptor variants. We note that of the 999 u-sQTLs we identify (Fig. 3F), we see similar numbers of alternative acceptors as alternative donors (358 of the u-sQTL junctions are alternative donors, 370 are alternative acceptors, 214 are alternative acceptor/donor combinations (with annotated productive

splice sites, just in new pairings), and 57 use both new splice donor and splice acceptor. Thus we believe there is unlikely a strong difference between the true abundance of u-sQTLs due to variations in splice donors versus splice acceptors. We hope to further investigate the relative contribution of cryptic splice donors versus cryptic splice acceptors in future studies.

4. The D2HGDH example, as the authors pointed out, seems to be a case of multiple independent causal variants modulating the expression of the same gene. The HyPrColoc method used by this paper considers a single causal variant, thus unable to detect the colocalization between u-sQTL and eQTL. Many colocalization software (the latest Coloc for example) is able to model multiple causal variants. Although this D2HGDH example is a valid example, a more intriguing question would be after multiple causal variants have been taken into consideration, are there still u-sQTLs that don't colocalize with eQTL? If so, why?

Thank you for the suggestion. Indeed we have considered using colocalization softwares that consider multiple variants (eg, the latest Coloc, which under the hood uses SuSiE, Wallace 2021 *Plos Gen*), however, we found empirically that this approach likely gains considerable power only when applied to datasets with much larger sample sizes than ours. Supporting this view, we could only identify secondary eQTLs with SuSiE for about one quarter of eGenes, which deterred us from choosing this approach. Thus, rather than using colocalization, we based much of our analysis on simply considering the marginal eQTL signal at each top sQTL SNP (Figures 3E, 3F). And indeed, addressing the more intriguing question that the reviewer raises, **we do see that there are some u-sQTLs with no apparent eQTL signal (or eQTL signal with an opposite direction effect size than what is expected). We consider the following explanations:**

1. We expect the eQTL effect will be related to the sQTL effect size measured in the more relevant, unstandardized units. That is, u-sQTLs with a small absolute change in PSI (which may be a large fold change in PSI, e.g., an unproductive junction spliced at PSI=1% in the reference allele that changes to PSI=2% in the alternative allele) will still have small effects on expression that may be undetectable at our sample size. We added a new supplemental figure (Figure S17) to explore this, in which we plot the correlation of sQTL and eQTL effect sizes in unstandardized units (deltaPSI, and expression log2FC, respectively). The general trend in spearman correlation coefficients is similar to Figure 3F. A similar analysis based on risdiplam-induced unproductive splicing (Figure S25B) again finds a correlation between deltaPSI and log2FC, suggesting **there is a correlation between splicing effect sizes and expression effects such that smaller effect on**

splicing tend to have smaller effects on expression, but the correlation is weak and we speculate other biological factors are at play. **We discuss this on line#444.**

2. **Differential efficiencies of NMD.** As described above (in our response to reviewer 1's first point), it is known that some PTC-bearing transcripts are targeted for NMD more strongly than others. For example, transcripts with a PTC in a non-terminal exon of normal size tend to elicit NMD stronger than those with a PTC in a particularly long (>407nt) non-terminal exon (Lindeboom et al 2019). We also see evidence of this (Figure S13C), but our ability to distinguish the splice junctions that create the different types of NMD-targeted transcripts is limited to those that we can observe in their full-transcript context from long read sequencing data. We attempted to further explore how these differential NMD efficiency rules apply to the splice junctions we identified with u-sQTLs, (that is, to recreate Figure 3F stratified by the 7 types of splice junctions described in Figure S13), but because only a small number of junctions could be classified this way (ranging from 1-50 u-sQTL junctions in each category), we could not detect any significant differences NMD efficiency from u-sQTLs and decided not to include that under-powered analysis in the revised version. Nonetheless, **our analysis in Figure S12C (described above in our response to reviewer 1's first point) is consistent with differential efficiencies of NMD for different unproductive splice junctions.**
3. **Other unknown biological reasons remain.** In the revised discussion (line #700) we speculate that transcriptional adaptation, a recently described phenomena wherein NMD-degraded transcript fragments travel back to the nucleus to activate transcription of the same gene and paralogs, may also be at play. In this case, one might expect the effects of u-sQTL effects of NMD may be canceled-out with transcriptional adaptation effects, yielding weaker expression changes. Indeed, we note that the NUDT14 u-sQTL that we highlight (Figure S22) actually has an apparent increase in expression detected in naRNA for the allele that increases unproductive splicing, consistent with transcriptional adaptation. Exploring whether transcriptional adaptation may be more generally at play among these u-sQTLs is an interesting topic for future studies.

Minor comments:

- The manuscript does not have line numbers, making it a bit difficult to comment on specific lines.

We have now added the line numbers, which we use to reference responses to your and the other reviewer's comments.

- Mostgenes -> most genes

- *Figure 1 caption: as heatmap -> as a heatmap*

- *Protein coding gene -> protein-coding gene*

Thank you for spotting the three typos above, they have now been corrected.

- *The author might consider citing [PMID: 31043754] when using colocalization scatter plots.*

Thank you for this suggestion, we have added this reference in our revised manuscript.

- *In addition to the ALS and COVID-19 example, [PMID: 31123710] is another experimentally validated example of NMD-led eQTL affecting complex diseases.*

Great reference. We have now added this example and reference to our revised manuscript.

Sincerely,

Boxiang Liu

Decision Letter, first revision:

22nd Apr 2024

Dear Dr. Li,

Thank you for submitting your revised manuscript "Global impact of unproductive splicing on human gene expression levels and traits" (NG-A64266R). 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.

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTeX)-- we can not proceed with PDFs at this stage.

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.

Congratulations!

Best wishes,
Chiara

Chiara Anania, PhD
Associate Editor
Nature Genetics
<https://orcid.org/0000-0003-1549-4157>

Reviewer #2 (Remarks to the Author):

The authors have addressed my concerns.

Reviewer #3 (Remarks to the Author):

I appreciate the effort by the authors for this revision. The revised manuscript now fully addresses my original concerns.

A few points to note:

- In your final print, there is little need to round to 1.5 – 2.5%. Using the original estimate (1.8 – 2.3%) works just as well.
- I am very happy to see that Figure S15D clearly demonstrates the importance of splice donors and acceptors over APA sites.

I wish you all the best with your outstanding study. Great work!!

Sincerely,
Boxiang Liu

Final Decision Letter:

16th Jul 2024

Dear Dr. Li,

I am delighted to say that your manuscript "Global impact of unproductive splicing on human gene expression" 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.

You may wish to make your media relations office aware of your accepted publication, in case they consider it appropriate to organize some internal or external publicity. Once your paper has been scheduled you will receive an email confirming the publication details. This is normally 3-4 working days in advance of publication. If you need additional notice of the date and time of publication, please let the production team know when you receive the proof of your article to ensure there is sufficient time to coordinate. Further information on our embargo policies can be found here: <https://www.nature.com/authors/policies/embargo.html>

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"](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.

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 strongly recommend that you upload the step-by-step protocols used in this manuscript to protocols.io. protocols.io is an open online resource that allows researchers to share their detailed experimental know-how. All uploaded protocols are made freely available and are assigned DOIs for ease of citation. Protocols can be linked to any publications in which they are used and will be linked to from your article. You can also establish a dedicated workspace to collect all your lab Protocols. By uploading your Protocols to protocols.io, you are enabling researchers to more readily reproduce or adapt the methodology you use, as well as increasing the visibility of your protocols and papers. Upload your Protocols at <https://protocols.io>. Further information can be found at <https://www.protocols.io/help/publish-articles>.

Thank you.

Sincerely,
Chiara

Chiara Anania, PhD
Associate Editor
Nature Genetics
<https://orcid.org/0000-0003-1549-4157>