nature portfolio

Peer Review File

Quantifying the regulatory potential of genetic variants via a hybrid sequence-oriented model with SVEN

Corresponding Author: Professor Ge Gao

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)
In this manuscript, Wang et al presented a machine learning model called SVEN to predict regulatory potential of genetic variants based on DNA sequence. SVEN is capable of quantifying tissue-specific gene expression alterations impacted by structural variants (SV). The authors demonstrated that SVEN can accurately predict genome tracks and gene expression levels across various tissues. Interpreting the functional consequences of SVs, particularly regarding tissue specificity, is a significant area of interest in this field. There are several major concerns regarding the evaluation of the model's performance. Additionally, the existence of several similar tools, such as Enformer, mitigates the novelty of this study. These comments are listed below as major and minor, to help improve the quality of the manuscript.

Major:

What is the novelty of SVEN compared to existing DNA sequence-based models such as Enformer, ExPecto, and Basenji2? This is an important issue to address so readers have a clear idea of the innovation of SVEN compared to the current state of the art. Also, what is the rationale of splitting the SVEN model into three components instead of directly predicting gene expression-related experiments such as CAGE as used in Enformer?

The author claims "SVEN showed consistently better performance across different tissues than ExPecto and Enformer". However, only 6 tissues were compared for Figure 2a. This result is susceptible to cherry-picking. A scatter plot with histograms showing distributions of correlation coefficients between SVEN vs Enformer and SVEN vs Expecto for all tissues can provide a more comprehensive comparison. It is fine to show expected and observed expression levels for 6 example tissues for Figure2b though.

In results, line 65-66, the authors wrote "Our evaluations suggested that these networks successfully learned the underlying regulatory codes from the inputs directly." Plotting predictions for only one case is not sufficient to demonstrate the model's capability of learning genome tracks accurately. Evaluation metrics such as correlations between predictions and experiments across bins should be used to quantify the performance.

The SV dataset used for evaluating impacts on gene expressions includes mainly deletions and insertions. Does the model performance differ across different SV types? How does the model perform for duplications, inversions and even translocations? Enformer can also make predictions for SV impacts on gene expression and should be benchmarked.

Is there a requirement for SV length/type when predicting SV-impacted gene expression for SVEN? What if an SV partially overlaps with the 131072bp window of gene TSS? For Figure 3a, is the performance different between different SV types?

When reporting the results, relying solely on a single descriptive statistic such as average correlations is insufficient. It is essential to include measures such as standard deviations and p-values when comparisons are being made.

It is not clear what the model is predicting and comparing with benchmark methods for the section of "SVEN improves small noncoding variant effects prediction" (line 131). The author used AUC for model performance evaluation, what is the outcome here? If the outcome is gene expression change, why and how to dichotomize it? The benchmark methods here have different outputs. For example, DeepSea is a sequence-based model predicting chromatin profiles. CADD is used to

predict SNP deleteriousness scores. What is the definition of small noncoding variants here?

For the result section “screening of SVs detected in large-scale population and clinical SVs”, SVs without disease-related phenotypes tend not to affect gene expressions, and pathogenic deletions are more likely to affect gene expression. This result should be quantified. In addition, why the authors only use pathogenic deletions from ClinVar? Comparisons can be made between pathogenic and benign SVs for different SV types, and perhaps some interesting biological insights can be inferred from such analysis.

When predicting SV effects on gene expressions, it would be interesting to see if tissue-specific predictions are different and whether the tissue with significant expression changes relates with phenotype. For example, COSMIC database has SVs for different cancer types and can be used to examine if tissue-specific predictions of SVEN are aligned with cancer tissue type.

The input sequence length of Enformer is 197k, which is larger than 131bp window of SVEN. This is advantageous of modeling distal gene regulatory effects since many regulatory variants can be located kilobases away from transcription start site. How would the authors address this?

Minor:

Define what is positive and negative variants, does it refer to variants altering gene expression levels? (Line 135-136) Or does it refer to those increasing/decreasing expression levels?

How are class-oriented genome tracks derived? Are they the average across different tissues and cell lines? It should be addressed in the method.

What are the definitions for small variants and large-scale SVs used in this study?

When referring figures to text, subfigure information should be added. For example, at line 114, it should likely be Extended Data Fig.8a.

(Remarks on code availability)
The code and software documentation can be improved. For example, what are the expected input and output, and how to interpret the results.

Right now I only see the following statement:

"This is a quick guide for usage, the full guideline is coming soon."

Reviewer #2

(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.

(Remarks on code availability)
The resource includes instructions for installing and running the application.

Reviewer #3

(Remarks to the Author)
In this manuscript, the authors introduce a novel computational method (SVEN) for predicting tissue-specific gene expression levels from the human genome sequence. They claim that SVEN can quantify the regulatory effects of both small and large-scale genetic variants in >350 tissues. The basic architecture of SVEN looks similar to previous models, such as Basenji2 and Enformer, but SVEN is a hybrid model with many additional processes. Although the manuscript is easily read and the topic is very important and interesting, I think that this work suffers from the following points:
1. Although the superiority of SVEN shown in Fig.2 and Fig.3d is impressive, its predictability for the effects of pathogenic SVs is questionable: among the top 5 deletions (from many samples; I could not find the exact number), only two were statistically significant; in addition, the values of relative expression in Fig.3b(right) look very small and the right graph does not seem to be the prediction result for the experimentally deleted sequence. Overall, I have an impression that the authors tend to present only favorable examples in the main Figures and avoid presenting more comprehensive results (i.e., cherry-picking).
2. As noted above, their model is complicated and even looks ad hoc. For example, the details of the feature-oriented model are not explained well (although we downloaded its code, this module was still unclear). As another example, the treatment of the mRNA-decay feature does not look systematic (in other words, does it really predict mRNA decay accurately?).
3. Moreover, each module's contribution is not tested/reported. An ablation study (or sort of) is necessary. For example, what happens if they apply the outputs of Enformer to their downstream steps (i.e., feature selection and the construction of gradient-boosting trees)?
4. Some results presented in Fig.4 are hard to interpret. Looking at Figs. 4a and 4b, is it reasonable to conclude that the

contribution of the mRNA decay is the largest? More explanations are needed. Similarly, despite the statistical significance notation, drawing any conclusions from Figs 4d, 4e, and 4f seems difficult. I suggest adding the information on corresponding TSS ranges to their x-axis; adding precise p-values might be helpful, too.
5. If I understand correctly, the prediction results do not cover deleted regions. However, there are no such gaps in Fig.3c. Also, the changes are more abundantly seen over the gene-body regions rather than the promoter regions, which may be strange.
6. More explanation/data on the 'SVEN-CellType-Match' should be given. Related to this perhaps, Fig.3a is the result for cell lines, though they describe that SVEN was trained with the data from 32 diverse genomes.

(Remarks on code availability)
(my co-reviewer checked the code and I included her comment in my comments)

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.

(Remarks on code availability)
The README file says "For reproducing results from our manuscript, please use Full mode", but the code only supports fast mode.
Since the code does not support full mode, it is not possible to reproduce the results.

Reviewer #5

(Remarks to the Author)
I read with great interest the manuscript by Wang et al. entitled "Quantifying the regulatory potential of genetic variants via a 1 hybrid sequence-oriented model". In this manuscript the authors implement a novel computational tool, SVEN, that can be used to predict tissue specific gene expression across many cell types/tissues and, importantly, quantify the potential transcriptional impact of structural variants, which in turn might help interpreting their pathogenicity.

Although the tool seems to have great potential, I have the overall feeling that its performance, especially with regard to pathogenic SVs, has not been properly tested. Furthermore, the authors seem to neglect the major and well-established role of distal regulatory elements, such as enhancers and insulators, in the controls of cell-type specific gene expression patterns as they "infer the gene expression levels from promoter-proximal sequences in a tissue-specific manner". My major concerns are the following ones:

- The authors claim that "SVEN employs a hybrid architecture to learn "regulatory codes" and infer the gene expression levels from promoter-proximal sequences in a tissue-specific manner". However, later in the text as well as in the methods it is stated that they consider 131 Kb around the TSS of genes, which clearly does not only include promoter-proximal sequences. Moreover, why did the authors use a 131 Kb window?.
- My main conceptual problem is that by focusing on promoter features, such as histone marks or TF that preferentially bind at promoters, it is not surprising that the authors can infer gene transcript levels as it is well known that many of those promoter features show very strong positive or negative correlations with gene expression. However, those promoter features are not necessarily causative of gene expression, particularly in the case of developmental and tissue-specific genes whose expression is largely dependent on distal regulatory elements such as enhancers. Moreover, the specificity of those distal elements is determined, at least to a large degree, by insulators and TAD boundaries.
Without learning the sequence features that are causative of gene expression changes, it is highly unlikely that SVEN can predict the transcriptional consequences of SVs acting through long-range mechanisms (e.g. enhancer adoption, enhancer disconnection).
- To test SVEN performance in predicting the effects of SVs, the authors perform global correlation analysis. However, most SVs should either have no effects on gene expression or significantly affect the expression of a few genes close to their breakpoints. Therefore, even if SVEN would not accurately predict the effects of the SVs, the global correlations might still look pretty OK. The authors should identify SVs that significantly change the expression of genes close to their breakpoints according to RNA-seq data and then evaluate SVEN performance for those genes affected by the SVs.
- In the context of SVs implicated in disease, it is becoming increasingly clear that SVs can act through long-range pathomechanisms (enhancer adoption, enhancer hijacking). In order to predict and interpret these long-range pathomechanisms it is essential to consider 3D chromatin organization (i.e. TAD maps) and tissue-specific enhancer maps. However, SVEN does not consider these important regulatory features but rather focuses on proximal-promoter sequences. The authors should collect information regarding SVs acting through long-range pathomechanisms (e.g. DECIPHER database, PMID: 36999617, PMID: 32973355) and evaluate SVEN performance for those SVs.
- As currently implemented, my impression is that SVEN can successfully predict the transcriptional (and pathogenic) impact of SVs that affect either promoter or coding sequences. However, these pathogenic mechanisms are relatively easy to predict and there are other tools that have already been implemented for those tasks. However, as stated above, what remains challenging is to predict the pathogenicity of SVs acting through long-range pathomechanisms. Therefore, the authors should perform a more thorough assessment of how SVEN compares with other tools (e.g. CADD-SV, STRVCTVRE, TADA, SVScore, POSTRE, etc.) on SVs whose coding, promoter or long-range pathomechanisms have been experimentally validated and that can be considered as positive controls. Similarly, the authors could also collect SVs

identified in healthy individuals and that can be thus considered as benign SVs (negative controls) and compared SVEN performance to that of other tools focusing on genes located close to the SVs breakpoints and whose expression is not affected by the SVs.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have generally addressed all my previous comments and concerns. It is a pity that the author of ExPecto didn't respond to your emails, yet the author of Enformer cannot located their original results any more; nevertheless, it is great to know that you were able to reproduce the published results on 218 tissues using the procedures previously described.

I appreciate your efforts in performing COSMIC analysis in response to the comments. I wonder if you can dig deeper into the three tissues (Adrenal gland, liver, prostate). The statistics looks very significant, and differ from other tissues a lot, so something else may be going on. Maybe manually examine some specific examples of the top pairs, or something similar, can help reveal what are the underlying causes for these three tissues that are drastically from others.

(Remarks on code availability)
The repository is updated to address my concern.

Reviewer #2

(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.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I believe the authors have satisfactorily replied to almost all my comments. Specifically, I still find it difficult to extract conclusions from Figure 3d-f as the differences are very subtle. Maybe using a different scale in the y-axis could help to highlight the difference in feature importance.

(Remarks on code availability)
My co-reviewer checked the code and the documentation and confirmed that the revision was done appropriately.

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.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)
I would like to thank the authors for addressing my previous comments and concerns. Although they have addressed some of my concerns, mostly those associated to the focus on promoter regions, the relevance and advanatges of the tool in comparison to other existing ones is still not fully resolved. As already mentioned in the abstract, the relevance of SVEN is ultimately link to the importance of SVs for human disease. In this regard, the comparison with other computational tools that can predict the pathogenicity of SVs indicates that SVEN is not superior in that task. Therefore, the authors might need to consider how to improve SVEN (e.g. add 3D data) in this regard as well as to peform a more exhaustive bookmarking of the tool in comparison to other existing tools including pathogenic and benign SVs as well as all major types of SVs (not only

CNVs).

(Remarks on code availability)

Responses to reviewers:

Reviewer #1 (Remarks to the Author):

In this manuscript, Wang et al presented a machine learning model called SVEN to predict regulatory potential of genetic variants based on DNA sequence. SVEN is capable of quantifying tissue-specific gene expression alterations impacted by structural variants (SV). The authors demonstrated that SVEN can accurately predict genome tracks and gene expression levels across various tissues. Interpreting the functional consequences of SVs, particularly regarding tissue specificity, is a significant area of interest in this field. There are several major concerns regarding the evaluation of the model's performance. Additionally, the existence of several similar tools, such as Enformer, mitigates the novelty of this study. These comments are listed below as major and minor, to help improve the quality of the manuscript.

Thanks for your encouraging and valuable comments! Based on comments, we have revised the manuscript accordingly.

Major comments:

1. What is the novelty of SVEN compared to existing DNA sequence-based models such as Enformer, ExPecto, and Basenji2? This is an important issue to address so readers have a clear idea of the innovation of SVEN compared to the current state of the art. Also, what is the rationale of splitting the SVEN model into three components instead of directly predicting gene expression-related experiments such as CAGE as used in Enformer?

Thanks for your comments! Eukaryotic tissue-specific gene expression regulation is based on complex gene expression networks, involves the coordination of multiple regulators and specific DNA elements and they act in combinatorial way [1]. Different from canonical "one-holistic-network-for-all" design, SVEN takes a hybrid architecture where it first learns distinct "regulatory codes" with multiple separated sub-networks, then makes the final inference by integrating over these various sub-models.

Benefiting from its unique design, SVEN shows consistently superior performance over canonical "one-holistic-network-for-all"-based Enformer in predicting tissue-specific

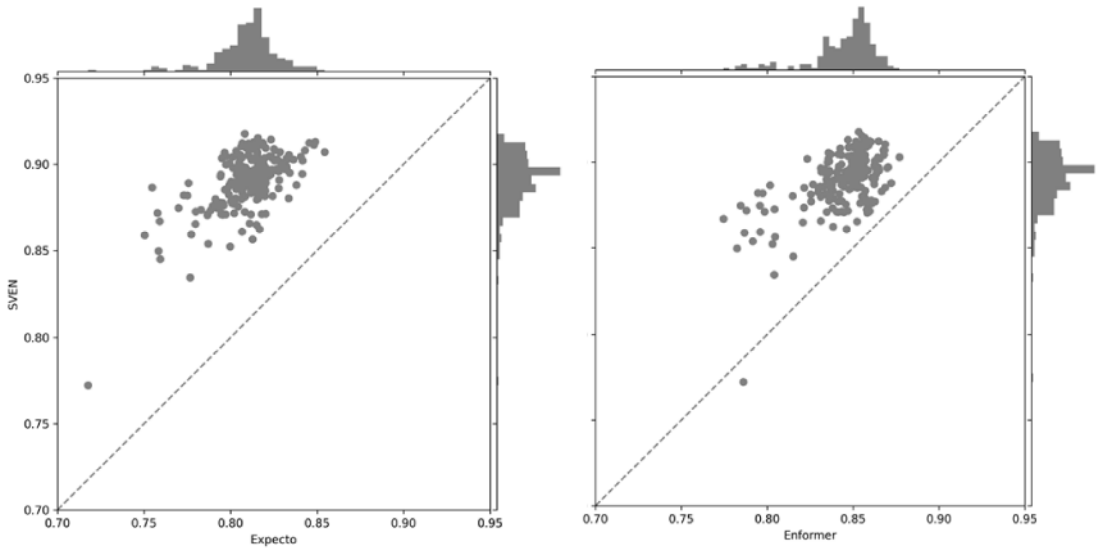
gene expression level (**Fig. 2a**) and assessing effects of SVs on gene expression (**Additional Table 1 and 3**), with 40% smaller model size (153M for SVEN largest model and 249M for Enformer).

The last but not the least, SVEN's modular design not only makes SVEN more interpretable, but also enables a customized model for users' particular tasks (we have provided the tutorial at https://github.com/gao-lab/SVEN?tab=readme-ov-file#customize-your-own-sven-model).

2. The author claims "SVEN showed consistently better performance across different tissues than ExPecto and Enformer". However, only 6 tissues were compared for Figure 2a. This result is susceptible to cherry-picking. A scatter plot with histograms showing distributions of correlation coefficients between SVEN vs Enformer and SVEN vs ExPecto for all tissues can provide a more comprehensive comparison. It is fine to show expected and observed expression levels for 6 example tissues for Figure2b though.

Thanks for the suggestion! The reason why we only compared 6 tissues in original **Fig 2a** is that ExPecto and Enformer only reported the detailed performance of these 6 tissues in their original publications. Unfortunately, when we tried to reach the authors, the author of ExPecto didn't response to our e-mail and the author of Enformer replied with "I can't seem to find the exact correlations any more. Please use the average performance in the paper as the key measure or reproduce the results".

Therefore, we tried to reproduce their results according to their papers and released codes on 218 tissues and cell lines (see Line 341–346, Page 13 for more details). The results well confirmed the superiority of SVEN (**Additional Fig. 1**, as **Fig 2a** in the revised manuscript).



Additional Fig. 1 SVEN improves tissue-specific gene expression prediction.

3. In results, line 65-66, the authors wrote "Our evaluations suggested that these networks successfully learned the underlying regulatory codes from the inputs directly." Plotting predictions for only one case is not sufficient to demonstrate the model's capability of learning genome tracks accurately. Evaluation metrics such as correlations between predictions and experiments across bins should be used to quantify the performance.

Thanks for the nice reminder! We have provided the average Pearson correlation between predicted and experimental signals (Line 66, Page 4) as well as detailed Pearson correlations as **Supplementary Table 1**.

4. The SV dataset used for evaluating impacts on gene expressions includes mainly deletions and insertions. Does the model performance differ across different SV types? How does the model perform for duplications, inversions and even translocations? Enformer can also make predictions for SV impacts on gene expression and should be benchmarked.

Thanks for the great heads-up! For a more comprehensive evaluation on assessing effects of SVs on gene expressions, we added a new SV dataset from the 1000 Genomes Project (https://doi.org/10.1101/2024.04.18.590093) for evaluation (Line 385–390, Page 15). The final dataset consisted of 94,366 SV-gene pairs, including 31,352

insertions, 2,109 mobile-element insertions, 36,450 deletions, 8,671 duplications, 88 inversion duplications and 15,696 complex SVs. We didn't find significant performance difference across distinct type of SVs, and SVEN shows a consistently superior performance over Enformer, especially with these unexpressed genes removed (**Additional Table 1**, as **Table 1** in the revised manuscript).

Additional Table 1 Model performance in assessing the effects of SVs on gene expression

SV type	**SV-gene pair number**	**SVEN**	**Enformer**
All	94,366 (61,811)	0.921 (0.907)	0.841 (0.786)
Insertion	31,352 (20,360)	0.921 (0.910)	0.841 (0.785)
MEI	2,109 (1,401)	0.928 (0.915)	0.869 (0.810)
Deletion	36,450 (24,186)	0.923 (0.905)	0.844 (0.785)
Duplication	8,671 (5,805)	0.923 (0.909)	0.839 (0.792)
Inversion duplication	88 (62)	0.923 (0.899)	0.846 (0.794)
Complex	15,696 (9,997)	0.914 (0.903)	0.831 (0.785)

(): results after removal of unexpressed genes.

5. Is there a requirement for SV length/type when predicting SV-impacted gene expression for SVEN? What if an SV partially overlaps with the 131072bp window of gene TSS? For Figure 3a, is the performance different between different SV types?

Thanks for your comments! As we have described in the manuscript (Line 368–373, Page 14), all SVs met the following requirements are taken into account when assessing effects on gene expressions: (1) the full sequences of both reference and alternative allele are available (as SVEN is a sequence-oriented model); (2) the SVs fall within the SVEN's context window (i.e., 131,072 bp centered on the TSS), regardless of their type/length. Consistently, empirical evaluation suggests SVEN is able to work robustly across different types of SV (**Additional Table 1**, as **Table 1** in the revised manuscript).

6. When reporting the results, relying solely on a single descriptive statistic such as average correlations is insufficient. It is essential to include measures such as standard deviations and p-values when comparisons are being made.

Thanks for the nice reminder! We have incorporated exact p-values as well as standard deviations when comparisons are being made.

7. It is not clear what the model is predicting and comparing with benchmark methods for the section of "SVEN improves small noncoding variant effects prediction" (line 131). The author used AUC for model performance evaluation, what is the outcome here? If the outcome is gene expression change, why and how to dichotomize it? The benchmark methods here have different outputs. For example, DeepSEA is a sequence-based model predicting chromatin profiles. CADD is used to predict SNP deleteriousness scores. What is the definition of small noncoding variants here?

Sorry for the potential confusing statement in the original manuscript. Briefly, in this benchmark, we first calculated predicted score S (Line 146–151, Page 6) of each variant, then compared with the experimental outcomes for 5,248,124 noncoding variants less than 50bp (i.e., "small noncoding variants") curated in existing database [2].

The outcome for **Figure 3d and 3e** is whether a particular variant is experiment-backed functional noncoding variant ("1") or not ("0"). Essentially, we took the raw scores reported by each tool (i.e. the functional significance score for DeepSEA [3] and the C scores for CADD [4]), plot ROC curve and calculate AUROC accordingly, aiming to assessing all tools within a uniform setup.

8. For the result section "screening of SVs detected in large-scale population and clinical SVs", SVs without disease-related phenotypes tend not to affect gene expressions, and pathogenic deletions are more likely to affect gene expression. This result should be quantified. In addition, why the authors only use pathogenic deletions from ClinVar? Comparisons can be made between pathogenic and benign SVs for different SV types, and perhaps some interesting biological insights can be inferred from such analysis.

Thanks for the nice reminder! We used SVs with clinical annotations ("pathogenic" and "benign") from dbVar (nstd102, 20231012) and selected deletions (11,509), duplications (573) and inversions (14) due to the lack of sequence information for other

types of SVs. Inversions were excluded owing to the limited number of SVs and finally we obtained 19,262 SV-gene pairs (pathogenic: 19,048; benign: 214) for following analysis (**Methods**).

We compared pathogenic deletions (18,620 SV-gene pairs) and duplications (428 SV-gene pairs) with reported benign deletions (175 SV-gene pairs) and duplications (39 SV-gene pairs), as well as deletions (84,881 SV-gene pairs) and duplications (18,777 SV-gene pairs) detected in large-scale population. While we found that pathogenic deletions are more likely to affect gene expression than benign and population ones (Wilcoxon rank-sum one-sided test P-value = 1.878×10^{-17} and 0, respectively), no statistically significant differences were detected for duplications (one-side Wilcoxon = 0.785 and 0.2117, respectively) which may be (partially) attributed to the relatively lower statistical power caused by limited number of duplications.

9. When predicting SV effects on gene expressions, it would be interesting to see if tissue-specific predictions are different and whether the tissue with significant expression changes relates with phenotype. For example, COSMIC database has SVs for different cancer types and can be used to examine if tissue-specific predictions of SVEN are aligned with cancer tissue type.

Thanks for the great suggestion! We downloaded SV set from COSMIC (v100, GRCh38) and processed with same procedure described in our manuscript (Line 366–381, Page 14) for following analysis. As SVEN requires reference and alternative allele sequences for SVs, we focus on deletions, duplications and inversions (deletions: 33,327; duplications: 2,892; inversions: 2,774). Then we matched SVEN's prediction with cancer tissue type and we got 36,078 SV-gene pairs. We found that SVs in adrenal gland, liver and prostate were more likely to cause tissue-specific changes (absolute $\log_2$ fold change > 1, one-sided Fisher's exact test P-value < 0.01, **Additional Table 2**).

Additional Table 2 Number of SV-gene pairs in matched tissues

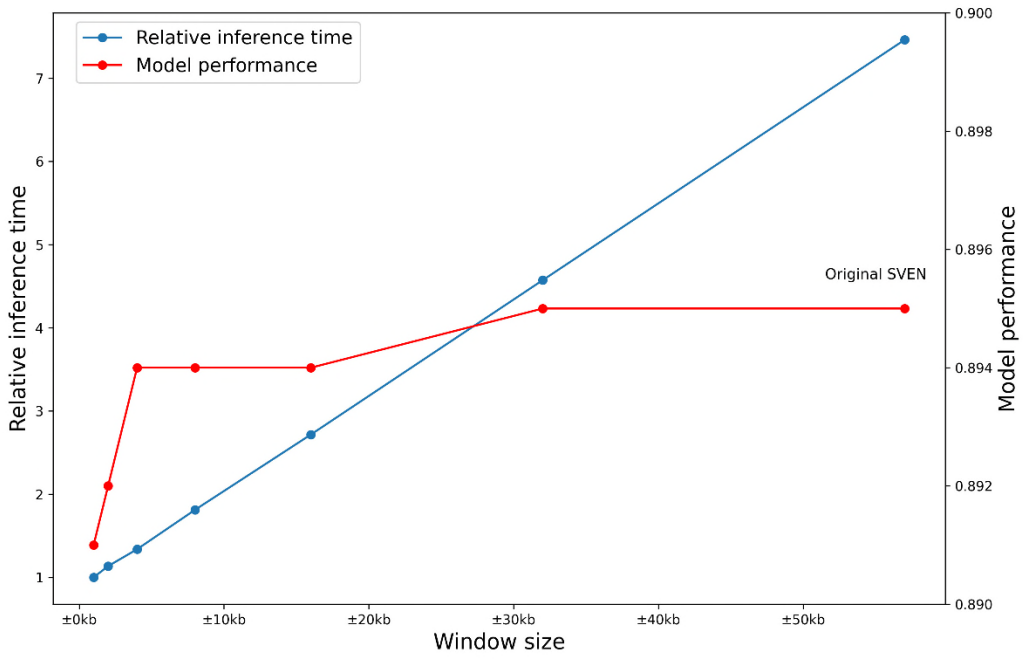
Tissue	**Number of SV-gene pairs**	**Number of SV-gene pairs (Absolute $\log_2$ fold change > 1)**	**Odds ratio**	**P-value**
Adrenal gland	200	33	2.857	8.089×10^{-7}
Breast	1,230	88	1.135	0.144

Liver	1,144	154	2.300	1.188×10^{-17}
Lung	614	39	0.986	0.558
Ovary	13,344	811	0.875	0.999
Pancreas	16,812	880	0.692	1.000
Prostate	2,734	314	2.071	3.958×10^{-26}

10. The input sequence length of Enformer is 197k, which is larger than 131bp window of SVEN. This is advantageous of modeling distal gene regulatory effects since many regulatory variants can be located kilobases away from transcription start site. How would the authors address this?

Thanks for your comments! We don't think that directly extend sequence length is an efficient way to model distal regulation. With much longer sequence as input, the improvement of model performance is marginal with much higher computation cost (**Additional Fig. 2**). Besides, given the model performance comparison between SVEN and Enformer, though SVEN used a much shorter sequence than Enformer, SVEN showed consistently better performance in tissue-specific gene transcription prediction than Enformer.

As we described in our manuscript (Line 214–218, Page 9), we have noticed that three-dimensional structure of human genome may help to model distal regulation better as well as expand the scope of applicable genetic variants in a more efficient way.



Additional Fig. 2 Performance and relative inference time of models with different window size of annotations (TSS ± 1/2/4/8/16/32/57 kb) on 53 GTEx tissues. Model performance was the mean Spearman correlation across 53 GTEx tissues (standard deviation was 0.021 for first two points and 0.022 for rest points). To estimate the compute costs of models with difference coverage of annotations, we used the inference time of the largest model in SVEN annotation module (153M) as approximate. Briefly, we first generated 2,048 random sequences with different length respectively, then predicted with annotation model (batch size = 16) for 5 rounds (2^{nd} – 6^{th} round) and calculate average inference time. Here, $Relative\ \ inference\ \ time = \frac{T}{T_{min}}$, where T was inference time and T_{min} was the minimum inference time of all.

Minor comments:

1. Define what is positive and negative variants, does it refer to variants altering gene expression levels? (Line 135-136) Or does it refer to those increasing/decreasing expression levels?

Thanks for the reminder! As we have mentioned above, we would call a variant "positive" as long as it is considered to be an expression-altering variant according to MPRA experiments. These have been incorporated into the revised manuscript (Line 144–145, Page 6).

2. How are class-oriented genome tracks derived? Are they the average across different tissues and cell lines? It should be addressed in the method.

Thanks for your heads-up! The annotation module of SVEN predicted 4,516 functional genomic features, including 1,896 TF binding features, 1,976 histone modification features, and 684 DNA accessibility features. The outputs of three class-oriented models were $N \times 896 \times 1896$, $N \times 896 \times 1976$ and $N \times 896 \times 684$ respectively (N: number of input sequences). For feature-oriented model, the output was $N \times 896 \times 1$. These have been incorporated into the revised manuscript (Line 259–262, Page 10).

3. What are the definitions for small variants and large-scale SVs used in this study?

As being described in the manuscript (Line 140–141, Page 6; Line 36, Page 3), we take variants smaller than 50bp as "small variants" and others as "large-scale SVs" according to literature [5].

4. When referring figures to text, subfigure information should be added. For example, at line 114, it should likely be Extended Data Fig.8a.

Thanks for your reminder! We have added subfigure information for all extended data figures.

Reviewer #1 (Remarks on code availability):

1. The code and software documentation can be improved. For example, what are the expected input and output, and how to interpret the results. Right now I only see the following statement: "This is a quick guide for usage, the full guideline is coming soon."

Thanks for your comments! We have improved the code with detailed usage guide (https://github.com/gao-lab/SVEN/blob/main/README.md) on the online GitHub repo.

Reviewer #2 (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.

Reviewer #2 (Remarks on code availability):

1. The resource includes instructions for installing and running the application.

Thanks for your comments! We have improved the code with detailed usage guide (https://github.com/gao-lab/SVEN/blob/main/README.md) on the online GitHub repo.

Reviewer #3 (Remarks to the Author):

In this manuscript, the authors introduce a novel computational method (SVEN) for predicting tissue-specific gene expression levels from the human genome sequence. They claim that SVEN can quantify the regulatory effects of both small and large-scale genetic variants in >350 tissues. The basic architecture of SVEN looks similar to previous models, such as Basenji2 and Enformer, but SVEN is a hybrid model with many additional processes. Although the manuscript is easily read and the topic is very important and interesting, I think that this work suffers from the following points:

Thanks for your encouraging and valuable comments! Based on comments, we have revised the manuscript accordingly.

1. Although the superiority of SVEN shown in Fig.2 and Fig.3d is impressive, its predictability for the effects of pathogenic SVs is questionable: among the top 5 deletions (from many samples; I could not find the exact number), only two were statistically significant; in addition, the values of relative expression in Fig.3b (right) look very small and the right graph does not seem to be the prediction result for the experimentally deleted sequence. Overall, I have an impression that the authors tend to present only favorable examples in the main Figures and avoid presenting more comprehensive results (i.e., cherry-picking).

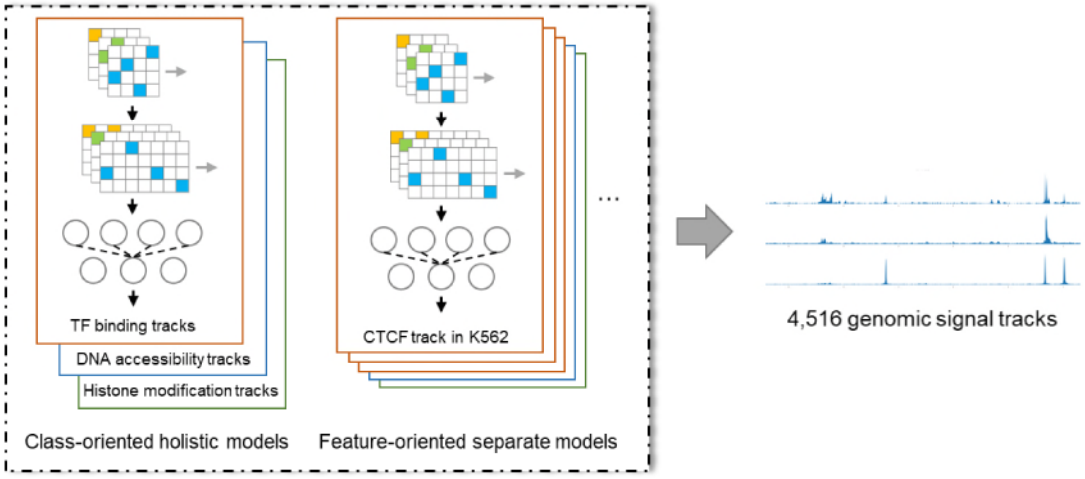
Thanks for your comments! As being described in the manuscript (Line 437–443 , Page 16), SVs from two datasets generated by *Ebert et al.* [6] and *Beyter et al.* [7] are taken as candidates for experimental validation. We filtered deletions with length less than 1 kb and got 61,114 SV-gene pairs. To evaluate the generalization ability of SVEN, we focused on deletions predicted to cause the upregulation of gene expression in four cell lines (HepG2, MCF-7, A375 and A549). We selected the top 5 deletions with available CRISPR target sites according to their effects on the target gene in the target cell line and validated these 5 deletions in A375 cell line (**Methods**). SVEN successfully predicted the direction of the impact on gene expression for 4 deletions, while 3 for Enformer and 2 for CNV Functional Annotation (**Supplementary Table 4**).

As what you pointed out correctly, the Fig.3b (right) was **not** the predicted, but the experimentally assessed result for these deletions. Meanwhile, these seemingly very small numbers are partly due to the fact that the *RPL41*, a highly expressed gene across all tissues (https://www.ncbi.nlm.nih.gov/gene/6171), was taken as reference gene initially. We have done normalization for better illustration.

2. As noted above, their model is complicated and even looks ad hoc. For example, the details of the feature-oriented model are not explained well (although we downloaded its code, this module was still unclear).

Sorry for our previous potentially ambiguous description. The basic idea here is to combine feature-oriented models (to learn the context-sensitive sequence-to-regulatory code, like the CTCF binding events in K562) and class-oriented models (to learn a more generalized rule, like TF bindings across different cell lines/tissues) for better utilizing available data. (**Additional Fig. 3**).

The detailed codes could be accessible at https://github.com/gao-lab/SVEN/blob/main/sven/models.py, with functions “sep_f1_model”/“sep_f2_model” for feature-oriented, and functions “acc_hol_model”/ “his_hol_model”/“tf_hol_model” for class-oriented models. We have incorporated a dedicated model-to-code page for easy code inspection (https://github.com/gao-lab/SVEN/blob/main/sven/README.md).



Additional Fig. 3 Hybrid architecture for SVEN annotation module.

As another example, the treatment of the mRNA-decay feature does not look systematic (in other words, does it really predict mRNA decay accurately?).

It has reported that the GC content and lengths of different functional regions (such as 5'UTR and 3'UTR) explained about 40% variability of mRNA half-lives [8–10]. The number of UTRs varies among difference genes, therefore we used four descriptive statistics (minimum, maximum, median, and mean values) of all 3'UTRs and 5'UTRs for the target genes to generate fixed-length mRNA decay features for each gene. Consistently, we found that removing mRNA decay features decreased model performance (Wilcoxon rank-sum one-sided test P-value = 5.353×10^{-11}, **Additional Table 3**), highlighting the contribution of mRNA decay features.

3. Moreover, each module's contribution is not tested/reported. An ablation study (or sort of) is necessary. For example, what happens if they apply the outputs of Enformer to their downstream steps (i.e., feature selection and the construction of gradient-boosting trees)?

Thanks for the nice reminder! To test each module's contribution of SVEN, we made several tests based on gene expression profiles across 218 tissues and cell lines used by ExPecto, Enformer and SVEN. We replaced SVEN annotation module with Enformer annotations (same 4,516 functional genomic features), removed mRNA decay features and replaced gradient-boosting trees with linear models respectively (**Additional Table 3** and **Supplementary Table 6**). As expected, we found that removing mRNA decay features and replacing gradient-boosting trees with linear models would decrease model

performance (Wilcoxon rank-sum one-sided test P-value = 5.353×10^{-11} and 8.332×10^{-39} respectively).

Of interest, we noticed that simply replacing SVEN annotation module with Enformer annotations could further increase performance, highlighting the extensibility and flexibility of current SVEN architecture.

Additional Table 3 Model performance on 218 tissues and cell lines

Model	Model performance
Original SVEN	0.890 ± 0.017
Replace annotation module with Enformer	0.900 ± 0.018
Remove mRNA decay features	0.881 ± 0.017
Replace gradient-boosting trees with linear models	0.870 ± 0.016

Values: Mean Spearman R ± standard deviation

4. Some results presented in Fig.4 are hard to interpret. Looking at Figs. 4a and 4b, is it reasonable to conclude that the contribution of the mRNA decay is the largest? More explanations are needed. Similarly, despite the statistical significance notation, drawing any conclusions from Figs 4d, 4e, and 4f seems difficult. I suggest adding the information on corresponding TSS ranges to their x-axis; adding precise p-values might be helpful, too.

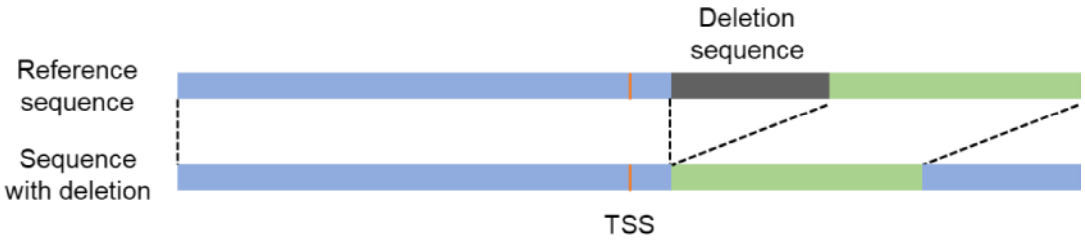
Thanks for the great reminder! While the mRNA decay features had the highest mean feature importance, the total contribution of TF binding, as a class, is the most significant one as we described in the manuscript accordingly. We have revised this part for better clarification (Line 165–166, Page 7).

For **Fig. 4d, 4e** and **4f**, we have added corresponding TSS ranges as well as precise P-values for more clear presentation.

5. If I understand correctly, the prediction results do not cover deleted regions. However, there are no such gaps in Fig.3c.

Thanks for your comments! As being described in the manuscript (Line 366–381, Page 14), we firstly predicted the SVEN functional annotation for the reference sequence and

the sequence with deletion (**Additional Fig. 4**), then transformed predicted annotations and incorporated mRNA decay features to trained gradient boosting tree model (here was A375 model) for predicting gene expression level of reference sequence and the sequence with deletion respectively.



Additional Fig. 4 Illustration of prediction of sequence by SVEN.

Also, the changes are more abundantly seen over the gene-body regions rather than the promoter regions, which may be strange.

Thanks for the comments. As being illustrated in **Fig. 3c**, most changes occur around the TSS of gene. Consistently, it has been well reported that H3K4me3 is typically located around the TSS and can extend downstream into gene body [11] and H3K27ac preferentially resides at promoters and enhancers [12] (also see the https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&position=chr11%3A49207162-49210352&hgsid=2354365847_yz1i5yjOOzBZh2Be78i6MZDOT6hY for a genomic view).

6. More explanation/data on the 'SVEN-CellType-Match' should be given. Related to this perhaps, Fig.3a is the result for cell lines, though they describe that SVEN was trained with the data from 32 diverse genomes.

Thanks for your heads-up! As being described in our manuscript (Line 145–151, Page 6), the predicted score S was used for variant effect evaluation, without any additional training involved. For "SVEN-CellType-Match", we extracted 265 and 483 cell line-matched features in the HepG2 and K562 cell lines, respectively, from 4,516 SVEN annotations to calculate predicted score S for evaluating variant effects.

Of note, we did not train SVEN with any variant dataset. Briefly, we first trained multiple regulatory-specific neural networks based on ENCODE datasets for TF

binding, histone modification and DNA accessibility. Then we trained a set of gradient-boosting tree models with transformed features selected based on a separated feature selection procedure. These models, along with gene expression profiles from GTEx and ENCODE, were used to infer gene expression in 365 tissues and cell lines (**Methods**).

For original Fig.3a (now is **Supplementary Table 3**), we used a large-scale SV dataset identified from 32 diverse human genomes for evaluating the ability of SVEN in predicting the effects of SVs on gene expression. As the paired RNA-seq data was generated from EBV-transformed lymphoblastoid cell lines, therefore, we used SVEN (GM12878 model) to predict the effects of these SVs on gene expression and compared the predicted values with the observed values (**Methods**).

We're sorry for the potentially confusing statements, and have polished this part in the revision.

Reviewer #3 (Remarks on code availability):

My co-reviewer checked the code and I included her comment in my comments.

Thanks for your comments! We have improved the code with detailed usage guide (https://github.com/gao-lab/SVEN/blob/main/README.md) on the online GitHub repo.

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.

Reviewer #4 (Remarks on code availability):

1. The README file says "For reproducing results from our manuscript, please use Full mode", but the code only supports fast mode. Since the code does not support full mode, it is not possible to reproduce the results.

Thanks for your comments! We have revised the repo with Full mode models (https://github.com/gao-lab/SVEN/tree/main/models/full_mode) as well as a detailed usage guide (https://github.com/gao-lab/SVEN/blob/main/README.md) incorporated.

Reviewer #5 (Remarks to the Author):

I read with great interest the manuscript by Wang et al. entitled "Quantifying the regulatory potential of genetic variants via a 1 hybrid sequence-oriented model". In this manuscript the authors implement a novel computational tool, SVEN, that can be used to predict tissue specific gene expression across many cell types/tissues and, importantly, quantify the potential transcriptional impact of structural variants, which in turn might help interpreting their pathogenicity.

Although the tool seems to have great potential, I have the overall feeling that its performance, especially with regard to pathogenic SVs, has not been properly tested. Furthermore, the authors seem to neglect the major and well-established role of distal regulatory elements, such as enhancers and insulators, in the controls of cell-type specific gene expression patterns as they "infer the gene expression levels from promoter-proximal sequences in a tissue-specific manner".

Thanks for your encouraging and valuable comments! Based on comments, we have revised the manuscript accordingly.

Major comments:

1. The authors claim that "SVEN employs a hybrid architecture to learn "regulatory codes" and infer the gene expression levels from promoter-proximal sequences in a tissue-specific manner". However, later in the text as well as in the methods it is stated that they consider 131 KB around the TSS of genes, which clearly does not only include promoter-proximal sequences. Moreover, why did the authors use a 131 KB window?

Sorry for our previous potentially ambiguous description. SVEN not only focused on promoter-proximal sequence but also considered 131 kb sequence around TSS. We have changed the word "promoter-proximal" to "TSS-centered" (Line 61–62, Page 4).

The current window size (131 kb) is mostly the balance between model performance and model size. Though longer sequence window could involve more distal regulation elements, the improvement of model performance may be marginal with much higher computation cost (**Additional Fig. 2**).

2. My main conceptual problem is that by focusing on promoter features, such as histone marks or TF that preferentially bind at promoters, it is not surprising that the authors can infer gene transcript levels as it is well known that many of those promoter features show very strong positive or negative correlations with gene expression. However, those promoter features are not necessarily causative of gene expression, particularly in the case of developmental and tissue-specific genes whose expression is largely dependent on distal regulatory elements such as enhancers. Moreover, the specificity of those distal elements is determined, at least to a large degree, by insulators and TAD boundaries. Without learning the sequence features that are causative of gene expression changes, it is highly unlikely that SVEN can predict the transcriptional consequences of SVs acting through long-range mechanisms (e.g. enhancer adoption, enhancer disconnection).

Thanks for the nice reminder! As being described above, SVEN takes not only promoter-proximal but also remote sequences into account. For example, the current SVEN annotation module effectively model CTCF binding events, an essential component for TAD [13], as well as H3K27ac, a well-known histone marker for active enhancer [14], across multiple tissues and cell lines based on experiment-backed data (**Supplementary Table 1**). The last but not the least, as being discussed in the manuscript, we're approaching to incorporate 3D genome structural architecture into the further version of SVEN (Line 214–218, Page 9).

3. To test SVEN performance in predicting the effects of SVs, the authors perform global correlation analysis. However, most SVs should either have no effects on gene expression or significantly affect the expression of a few genes close to their breakpoints. Therefore, even if SVEN would not accurately predict the effects of the SVs, the global correlations might still look pretty OK. The authors should identify SVs that significantly change the expression of genes close to their breakpoints according to RNA-seq data and then evaluate SVEN performance for those genes affected by the SVs.

Thanks for your comments! For a more comprehensive evaluation on assessing effects of SVs on gene expressions, we added a new SV dataset from the 1000 Genomes Project (https://doi.org/10.1101/2024.04.18.590093) for evaluation (Line 385–390, Page 15). The final dataset consisted of 94,366 SV-gene pairs, including 31,352 insertions, 2,109 mobile-element insertions, 36,450 deletions, 8,671 duplications, 88 inversion duplications and 15,696 complex SVs (**Additional Table 1**). To evaluate the model performance of SVEN on SVs which can significantly change the expression, we first calculate the $\log_2$ fold change of RNA-seq level between target samples and background reference (SVEN's training dataset). We selected genes with significantly change (absolute $\log_2$ fold change >1) and got 31,689 SV-gene pairs. Compared with Enformer, SVEN consistently can predict effects of SVs on gene expressions more accurately (Spearman R = 0.801 and 0.740 respectively).

4. In the context of SVs implicated in disease, it is becoming increasingly clear that SVs can act through long-range pathomechanisms (enhancer adoption, enhancer hijacking). In order to predict and interpret these long-range pathomechanisms it is essential to consider 3D chromatin organization (i.e. TAD maps) and tissue-specific enhancer maps. However, SVEN does not consider these important regulatory features but rather focuses on proximal-promoter sequences. The authors should collect information regarding SVs acting through long-range pathomechanisms (e.g. DECIPHER database, PMID: 36999617, PMID: 32973355) and evaluate SVEN performance for those SVs.

Thanks for the great suggestions! We have checked all 25 pathogenic SVs (for 17 deletions, 4 duplications, and 4 inversions) which are reported to cause diseases through long-range mechanisms from public resource [15], but found, unfortunately, none of them could be predicted by current SVEN due to the length (median length: 887 kb) or the distance to TSS (median distance: 960 kb).

As being described in out manuscript (Line 214–218, Page 9), we have noticed that there were still several paths for further improving the accuracy and scalability of SVEN, including combining the three-dimensional structure of human genome to model distal regulation better as well as expand the scope of applicable genetic variants. We're approaching to incorporate 3D genome structural architecture into the further version of SVEN.

5. As currently implemented, my impression is that SVEN can successfully predict the transcriptional (and pathogenic) impact of SVs that affect either promoter or coding sequences. However, these pathogenic mechanisms are relatively easy to predict and there are other tools that have already been implemented for those tasks. However, as stated above, what remains challenging is to predict the pathogenicity of SVs acting through long-range pathomechanisms. Therefore, the authors should perform a more thorough assessment of how SVEN compares with other tools (e.g. CADD-SV, STRVCTVRE, TADA, SVScore, POSTRE, etc.) on SVs whose coding, promoter or long-range pathomechanisms have been experimentally validated and that can be considered as positive controls. Similarly, the authors could also collect SVs identified in healthy individuals and that can be thus considered as benign SVs (negative controls) and compared SVEN performance to that of other tools focusing on genes located close to the SVs breakpoints and whose expression is not affected by the SVs.

Thanks for the great suggestion! While SVEN was designed to quantify the tissue-specific transcriptomic impacts of SVs, rather than to predict the pathogenicity of SVs directly, we have incorporated the additional evaluation by comparing with CADD-SV [16], StrVCTVRE [17], TADA [18], SVScore [19] and POSTRE [15] based on 135 curated pathogenic SVs (for 132 deletions and 3 duplications) from ClinVar (20231012)[1]. As being expected, we found that SVEN showed a relatively poorer performance than those pathogenicity-oriented algorithms (**Additional Table 4**), which, we believe, is partially due to SVEN's expression-oriented nature and the fact that gene expression alteration does not necessarily mean disease-causing[2] [20].

Additional Table 4 Model performance on predicting the pathogenicity of SVs

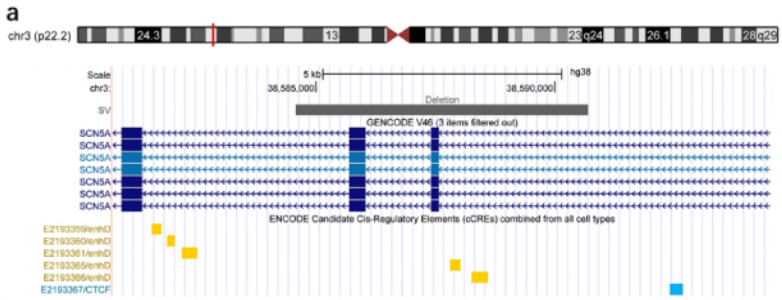
Tool	Model performance (AUROC)
SVEN	0.853
POSTRE	0.890

[1] Briefly, we selected positive set with SVs met: (1) deletions and duplications (supported by all tools); (2) clinical annotation was "pathogenic"; (3) review status had at least two gold stars ("criteria provided, multiple submitters, no conflicts", "reviewed by expert panel" or "practice guideline"); (4) disrupting promoters or exons. For negative controls, we used SVs from the dataset we mentioned above (generated by the 1000 Genomes Project) and selected 35,203 SVs (for 29,026 deletions and 6,117 duplications) tended to have no significant effects (absolute log2 fold change <1 for all SV-gene pairs) on gene expression (StrVCTVRE only supported SVs disrupting exons, so it was evaluated on 1,454 SVs rather than 35,203 SVs).

[2] As evidence, we found a SVEN expression-only metric (for each SV, we calculated the maximum absolute $\log_2$ fold change across all SV-gene pairs in 53 GTEx tissues as predicted output) would result in further inferior performance (AUPRC = 0.670).

TADA	0.931
CADD-SV	0.963
StrVCTVRE	0.967
SVScore	0.997

Meanwhile, we'd noted that the expression-oriented SVEN and these pathogenicity-oriented algorithms could be rather complementary, enabling a possibility to jointly identify pathogenic SVs which function through transcriptomic changes (e.g., **Additional Fig. 5**). We have incorporated these into the revised Discussion accordingly (Line 207–212, Page 9).



Tool	Output metrics	Final predicted
SVEN	181.393	benign
POSTRE	1.000	pathogenic
TADA	0.658	pathogenic
CADD-SV	7.545	pathogenic
StrVCTVRE	0.755	pathogenic
SVScore	50.400	pathogenic

Tool	Output metrics	Final predicted
SVEN	262.158	pathogenic
POSTRE	0.610	benign
TADA	0.388	benign
CADD-SV	3.037	benign
StrVCTVRE	0.023	benign
SVScore	19.865	benign

Additional Fig. 5 a. Visualization of a 6-kb benign deletion (chr3: 38584573-38590687) covering exon 15 and 16 of gene *SCN5A*. SVEN predicted this SV as benign and all other tools predicted as pathogenic. Only SVEN can predict this SV correctly. **b.** Visualization of a 210-bp benign deletion (chr12: 93404508-93404718) disrupting exon 5 of gene *NUDT4* and a potential enhancer. SVEN predicted this SV as pathogenic and all other tools predicted as benign. Only SVEN cannot predict this SV correctly.

References

[1] Reményi A, Schöler HR, Wilmanns M. Combinatorial control of gene expression. Nat Struct Mol Biol 2004;11:812–5.

[2] Wang Y, Shi F, Liang Y, Gao G. REVA as a well-curated database for human expression-modulating variants. Genomics, Proteomics & Bioinformatics 2021;19:590–601.

[3] Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning–based sequence model. Nat Methods 2015;12:931–4.
[4] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.
[5] Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. Nat Rev Genet 2020;21:171–89.
[6] Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science 2021;372:eabf7117.
[7] Beyter D, Ingimundardottir H, Oddsson A, Eggertsson HP, Bjornsson E, Jonsson H, et al. Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. Nat Genet 2021;53:779–86.
[8] Courel M, Clement Y, Bossevain C, Foretek D, Vidal CO, Yi Z, et al. GC content shapes mRNA storage and decay in human cells. Elife 2019;8.
[9] Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MSH. Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. DNA ResearchDNA Research 2008;16:45–58.
[10] Spies N, Burge CB, Bartel DP. 3′ UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. Genome Res 2013;23:2078–90.
[11] Beacon TH, Delcuve GP, López C, Nardocci G, Kovalchuk I, van Wijnen AJ, et al. The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. Clin Epigenetics 2021;13:138.
[12] Zhao W, Xu Y, Wang Y, Gao D, King J, Xu Y, et al. Investigating crosstalk between H3K27 acetylation and H3K4 trimethylation in CRISPR/dCas-based epigenome editing and gene activation. Sci Rep 2021;11:15912.
[13] Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell 2015;163:1611–27.
[14] Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 2010;107:21931–6.
[15] Sánchez-Gaya V, Rada-Iglesias A. POSTRE: a tool to predict the pathological effects of human structural variants. Nucleic Acids Res 2023.
[16] Kleinert P, Kircher M. A framework to score the effects of structural variants in health and disease. Genome Res 2022;32:766–77.
[17] Sharo AG, Hu Z, Sunyaev SR, Brenner SE. StrVCTVRE: a supervised learning method to predict the pathogenicity of human genome structural variants. The American Journal of Human Genetics 2022;109:195–209.
[18] Hertzberg J, Mundlos S, Vingron M, Gallone G. TADA—a machine learning tool for functional annotation-based prioritisation of pathogenic CNVs. Genome Biol 2022;23:67.

[19] Ganel L, Abel HJ, Hall IM. SVScore: an impact prediction tool for structural variation. Bioinformatics 2017;33:btw789.
[20] Corbett AH. Post-transcriptional regulation of gene expression and human disease. Curr Opin Cell Biol 2018;52:96–104.

Responses to reviewers:

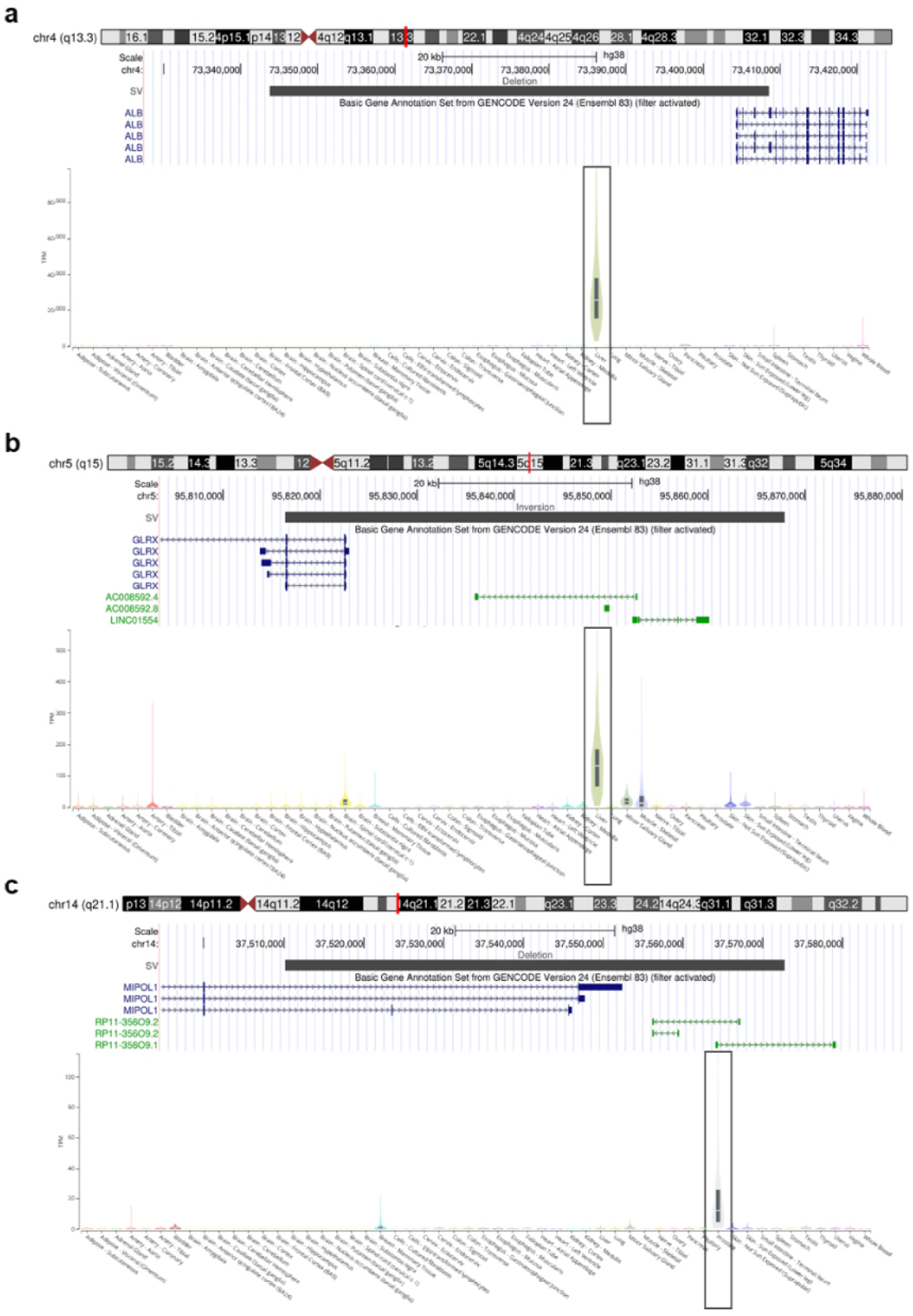
Reviewer #1 (Remarks to the Author):

The authors have generally addressed all my previous comments and concerns. It is a pity that the author of ExPecto didn't respond to your emails, yet the author of Enformer cannot located their original results anymore; nevertheless, it is great to know that you were able to reproduce the published results on 218 tissues using the procedures previously described.

Thanks for your encouraging comments!

I appreciate your efforts in performing COSMIC analysis in response to the comments. I wonder if you can dig deeper into the three tissues (Adrenal gland, liver, prostate). The statistics looks very significant, and differ from other tissues a lot, so something else may be going on. Maybe manually examine some specific examples of the top pairs, or something similar, can help reveal what are the underlying causes for these three tissues that are drastically from others.

Thanks for the nice suggestion! Manually examining top SV-gene pairs with significant changes (absolute $\log_2$ fold change > 1) suggests a potential contribution from tissue-specific genes for the observed interesting pattern. For example, gene *ALB* expresses strictly in liver (https://www.ncbi.nlm.nih.gov/gene/213); *LINC01554* is a long noncoding RNA (lncRNA) mainly expressed in liver (https://www.ncbi.nlm.nih.gov/gene/202299), and was reported to be associated with tumor invasion, tumor size, tumor staging, and shorter survival of hepatocellular carcinoma patients [1]; *RP11-356O9.1* is annotated as a lncRNA and predominantly expresses in the prostate [2] (**Additional Fig. 1**). We didn't find tissue-specific genes from top pairs in other tissues.



Additional Fig. 1 a. Visualization of a 64-kb liver-targeted deletion (chr4:73343769-73408563) which disrupts gene *ALB*. Gene expression level was obtained from GTEx Portal (release v10, https://www.gtexportal.org/). **b.** Visualization of another liver-targeted 51-kp inversion (chr5:95816394-95867929) covering gene *LINC01554*. **c.**

Visualization of a 62-kb prostate-targeted deletion (chr14:37510161-37572732) disrupting gene *RP11-356O9.1*. No significance detected in adrenal gland-targeted SV-pairs which could partly be attributed to its relatively limited number of SV-gene pairs (200 out of 36,078; 0.55% of all pairs)

Reviewer #1 (Remarks on code availability):

The repository is updated to address my concern.

Thanks for your comments!

Reviewer #2 (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.

Reviewer #3 (Remarks to the Author):

I believe the authors have satisfactorily replied to almost all my comments. Specifically, I still find it difficult to extract conclusions from Figure 4d-f as the differences are very subtle. Maybe using a different scale in the y-axis could help to highlight the difference in feature importance.

Thanks for your nice suggestions! We have revised the **Fig. 4d-f** for more clear presentation.

Reviewer #3 (Remarks on code availability):

My co-reviewer checked the code and the documentation and confirmed that the revision was done appropriately.

Thanks for your comments!

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.

Reviewer #5 (Remarks to the Author):

I would like to thank the authors for addressing my previous comments and concerns. Although they have addressed some of my concerns, mostly those associated to the focus on promoter regions, the relevance and advantages of the tool in comparison to other existing ones is still not fully resolved. As already mentioned in the abstract, the relevance of SVEN is ultimately link to the importance of SVs for human disease. In this regard, the comparison with other computational tools that can predict the pathogenicity of SVs indicates that SVEN is not superior in that task. Therefore, the authors might need to consider how to improve SVEN (e.g., add 3D data) in this regard as well as to perform a more exhaustive bookmarking of the tool in comparison to other existing tools including pathogenic and benign SVs as well as all major types of SVs (not only CNVs).

Thanks for the great comments! As being discussed in the manuscript, we have noticed that there were still several paths for further improving the accuracy and scalability of SVEN, including combining the three-dimensional structure of human genome to model distal regulation better as well as expand the scope of applicable genetic variants (Line 215–219, Page 9).

Meanwhile, we'd noted that SVEN was designed for quantifying the tissue-specific transcriptomic impacts of SVs, and the fact that gene expression alteration does not necessarily mean disease causing [3–5], we would expect that the SVEN will hardly surpass these pathogenicity-oriented algorithms in predicting the pathogenicity of SVs[1]. Instead, we'd suggest that SVEN and these algorithms could be rather complementary, enabling a possibility to jointly identify pathogenic SVs which function through transcriptomic changes. We have further revised the part of Discussion for better

[1] As evidence, we found that expression-oriented SVEN showed a relatively poorer performance than pathogenicity-oriented algorithms as showed in Additional Table 4 of previous response letter (and now as Supplementary Table 2 in this revision).

clarification (Line 207–213, Page 9).

References

[1] Zheng Y, Li L, Jia Y, Zhang B, Li J, Zhu Y, et al. LINC01554-Mediated Glucose Metabolism Reprogramming Suppresses Tumorigenicity in Hepatocellular Carcinoma via Downregulating PKM2 Expression and Inhibiting Akt/mTOR Signaling Pathway. Theranostics 2019;9:796–810.
[2] Schimmelpfennig C, Rade M, Füssel S, Löffler D, Blumert C, Bertram C, et al. Characterization and evaluation of gene fusions as a measure of genetic instability and disease prognosis in prostate cancer. BMC Cancer 2023;23:575.
[3] Corbett AH. Post-transcriptional regulation of gene expression and human disease. Curr Opin Cell Biol 2018;52:96–104.
[4] Sun BB, Kurki MI, Foley CN, Mechakra A, Chen C, Marshall E, et al. Genetic associations of protein-coding variants in human disease. Nature 2022;603:95–102.
[5] Wang S, Sun S. Translation dysregulation in neurodegenerative diseases: a focus on ALS. Mol Neurodegener 2023;18:58.