nature portfolio

Peer Review File

Development of a Genetic Priority Score to Predict Drug Side Effects Using Human Genetic Evidence

Corresponding Author: Professor Ron Do

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)
This study extends a Genetic Priority Score score from a recent publication to a Side Effect Genetic Priority Score (SE-GPS). The in silico tool is designed to predict adverse drug side effects for 19,422 protein-coding genes and 470 side effects using comprehensive genetic data from single variants to GWAS. High SE-GPS scores (top 0.01%-0.40%) were strongly associated with increased risk of drug side effects, particularly severe ones like those linked to boxed warnings or toxicity-related drug withdrawals, with fold increases of up to 45.5. The method seems to be technically sound, nevertheless the manuscript appears somewhat underdeveloped in certain sections and resembles more of an analysis plan than a manuscript. E.g. The results section contains elements which could be moved into the method section. This would allow for more space to focus on presenting further results. I have several comments aligned with this direction:

Abstract: The abstract remains too vague and presents the results in a partly unclear manner. A revision is needed to provide more specific and concrete information. E.g. The phrase “five lines of human genetic evidential support” feels a bit abstract. Is the second-to-last sentence a validation as drugs with known side effects have a high score?

Line 77: The five genetic features should be provided in the introduction and together with all data sources in Figure 1.

Caption to Figure 1 is not finished. “severity using a crowdsourced severity score (include ref?).. Sh”

Line 87-92, revise for clarity. Repeated description of the methods, but unclear due to the use of different terms. The section “Construction of the drug genetic dataset” reads more like a methods description than a results section and should be moved/revised.

Line 99, are they removed twice? Compare to line 92/93

Line 115/166: which subset? What are the 5 genetic features? Clinical variants, coding variants and GWAS phenotypes, GWAS and? The reason why Locus2Gene24 and eQTL phenotype were not merged is missing. Polygenic risk scores (i.e. The Polygenic Score Catalog) are missing.

Line 119, where does the 15 phecode categories come from? The number should be described as a result in the section before.

Line 120-126. It is unclear why the finding that "phenotype terms overlap with the drug indication" (e.g., [provide 1-2 specific examples]) is included under the section "Association of genetic features with drug side effects." It seems more relevant to the first result or method section. Clarifying this connection would help improve the flow and focus of the section. Furthermore, this side effect filter should also be used in the analysis showed in Supp. Fig. 3.

What is the rational to report the top 0.40%, 0.05% and 0.01% of the SE-GPS (equivalent to scores 184 greater than 0.9, 1.5 and 2.1)? The 0.6 score shows also an OR>2.

Line 142-144, The reader (may) have some difficulty to follow this finding, as Extended Data Figure 2 contains four subplots. Furthermore, it is unclear why the clinical variant feature was split according to the number of data sources in this section,

but this approach was not applied in the earlier results sections, such as in Figure 2 and Supplementary Figure 3. The authors have to provide additional guidance on how to navigate the figure e.g. provide more detailed results in the manuscript.

Figure caption 3, vertical red dashed line should be explained

Line 163-166: What is the phecode “Genetic”? Is this side effect correlated with its drug indication? If so, why does this Phecode not have a higher odds ratio (OR), given that there should be a clear genetic background supporting the association? Should be discussed.

Line 166: It is unclear which finding "these differences" is referring to. Is the high OR of infectious disease, etc highlighting the impact of genetics or non-genetic factors, or both? Revise.

Line 170. FDA box warnings drug results should be added i.e. to a plot like 3a and b, stratified by "phenotype terms overlap with the drug indication” or "phenotype terms not overlapping with the drug indication”

Line 180 Reference is missing. For this analysis "phenotype terms overlap with the drug indication” were removed before the analysis?

Line 181 Is this the rational? ”Thus, we next considered if extremes of the SE-GPS were more likely to yield a drug side effect” as “drug toxicity is a result of on-target or mechanism-based toxicity, where the toxic effect is a result of a response directly related to the therapeutic effect” on the drug target? Should be added here again to give some guidance.

Line 185 Instead of acknowledging this finding, I would suggest framing it as something more in line with expectations since otherwise many drugs were not anymore on the market.

Line 193-201: What is the difference to the FDA drug box warning analysis in lines 170-178? This analysis should also be stratified by "phenotype terms overlap with the drug indication” or "phenotype terms not overlapping with the drug indication”

Line 211 Reference is missing

Extended Data Fig.4 Caption is not finished, ore (include ref?); at which level of significance?)

Line 228-229 and line 234-236: Unclear statements, please revise.

Data and app should also be provided using a zenodo link or similar.

In the statistical analysis section, the p-value threshold and the method used for p-value adjustment are not specified.

Comments specific to the Shiny App:
• Enable column search
• No gene data available (seems to be fixed)
• Gene select input with live search
• select input with live search
• More helping text or a video tutorial should be provided. i.e. helping text for the table columns, particularly where abbreviations are used. Which cut off should be used? Provide some helping text, ect.

Reviewer #2

(Remarks to the Author)
The authors have embarked on an ambitious project collating many data sets to address an important question about prediction of adverse reactions to new drugs. I think I like the concept of the paper, but I am left very unclear how I would use. My comments are below, sometimes I step through to just make a logically flow of thinking. As I wrote my comments my understanding increased, hence it is possible that some of my comments I have self-answered. In general, I think I had to work too hard to understand what is being done.

1. Goal
a) First, I am left feeling about how to use the results of the study. The goal of the study is to “help inform the likelihood of an on-target side effect” (Discussion second sentence).
So, for a general audience, I think this could be explained with a couple more sentences (i.e., on-target vs off-target - obvious to the authors I realise).
b) I find it ambiguous exactly how one show use the results and the online tool. Can this be stepped through. Is it that when a new drug is being developed that one should identify the gene target of the drug and look this up in the database?
Can examples, be given of how it can be used in practice.
c) My understanding is that only a small proportion of genes are actually druggable.
Of the genes that have results (I am unclear how many genes have results) what proportion are considered druggable.
d) I am unclear what obvious (i.e. known) results are missed.
Can some examples be given.

e) Mendelian Randomisation has been proposed to identify on-target adverse effects. See doi: 10.1093/ije/dyx207 "Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities" which has been cited 174 times
Can you compare your results to theirs (or papers that have used this approach) either directly, or at least in discussion to explain how approaches differ.

2. Methods
I have spent a very long time on this. I added line numbers to the word file provided.
a. Figure 1 should include more steps to describe the process. Include the structure of the files generated at each stage. Currently it is not a very helpful summary. In Step 3 the description is "v applying this method to 19,422 genes and 470 phenotypes to identify targets" please state the total number of gene-phenotype rows and give an example where a gene has 2 phenotype codes, and the same phenotype has 2 genes.
b. The description in the main text is difficult to map to the more detailed methods. The numbers at lines 100-102 (987 drugs, 733 genes, 348 unique drug indications and 417 unique side effects, whereas the OnSIDES dataset consisted of 806 drugs, 697 genes, 349 unique drug indications and 396 unique side effects) are not mentioned in the Methods. Since at line 324 it says "1,037 drugs, 748 genes, 321 unique drug indications and 385 unique side effects mapped to phecode integers." For OpenTargets and for Onsides line 334 does not match. Presumably because of another QC step, but its adds to making it a very confusing read.
c. At line 105. State that both the drug indications and the side effects were mapped to phecodes to make a set of gene-phenotypes.
d. Line 411 I think you need to add a section "Combined gene data set" which states (to my understanding) that the previously listed genetic databases are combined to give a matrix of rows gene-phenotype 19,422 protein coding genes that have at least 1 yes/no entry for clinical variants. Say how many rows in total.
e. Line 106" We integrated both side effect datasets with human genetic evidence at the gene-phenotype level using nine data sources". "Integrated" is ambiguous. Give more detail so it is clear what the file is that is used for analysis.
f. Line 113. Suggest 'A detailed description of each data source is provided in the Methods and an overview of these gene-phenotype observations across 19,422 genes and 470 phenotypes is shown in Supplementary Fig. 2 ' is updated to 'A detailed description of each data source is provided in the Methods leading to an analysable data matrix where the entries are zero or one of x gene-phenotype rows comprising 19,422 protein coding genes, 470 phenotypes, z phecodes and x columns. The columns comprise…. An overview of gene-phenotype observations is shown in Supplementary Fig. 2 '.
g. Line 429 "Generation of the integrated drug–genetic dataset" section. As in comments above, explicitly state the inputs and output of this step
h. It is not clear why Supp Fig 2 summarises as phenotypes rather than phecodes. From Supp Fig 2 it is clear most phenotypes are unique, whereas phecodes presumably combine similar phenotypes. Should an additional supp Figure be made that gives the same info for phecodes.
i. Line 117. I found I had to work too hard to understand this section "Association of genetic features with drug side effects" please step the reader through more.
So the y-variable is yes/no for side effects for a gene (385 separate analyses or combined into one). How many genes in the analysis? By the time I get to line 238 I realise not all genes. An equation would help avoid ambiguity as then the authors would define the regression more specifically. The equation includes 15 phecodes as covariates. It is not explained what these 15 phecodes are – are they phecode categories, nor why they are selected, nor why the analysis is biased if these are not included. Please provide more information.
j. Line 129. In the section "Construction of the SE-GPS". I have read this many times and cannot understand what has been done. What is the y-variable in the analysis. OK I see in extended data Figure 1, y= side effect, and in page 20 methods. I guess this is implied in statement "mixed-effect regression model of the five genetic features with drug side effects", but I hope you can see that "with drug side effects" is ambiguous to the reader.
k. Line 130-131. The first 2 lines, implies information from the total data section (c. above analysis) is brought into the cross-validation analysis. This feels uncomfortable for the potential for data leakage/overfitting. The whole pipeline of analyses should be conducted in the discovery 80% and applied to the left-out analysis. This is important, and may lead to reduced consistency in Extended Data figure 1.
l. Line 140-141 "We used the cross-validated test with the highest OR and extracted the coefficients from this mixed-effect model to calculate the SE-GPS in the OnSIDES dataset (Extended Data Table 1)." Extended Data Table 1, is for Open Targets relating to Extended data Figure 1. The OnSides Table is Extended Data Table 2. In addition provide a Figure like Extended Data Figure 1 for OnSIDES.
m. Page 20 SE-GPS_Oti calculation. The first section (ie. up to ". For each of the five folds ..") should include an equation, with each term defined which will be easier to follow than a description in words alone. This section has the y variable as drug-side effect, so the regression estimates are log(OR) of genetic feature to drug side effect.
n. The second section of SE-GPS, giving the equation for SE-GPS_Oti. This now makes a value for a gene-phenotype. My haziness maps back to part a above. Explicitly define for j=1, n. Provide the beta-weights in a table
o. Page 21 "In addition, we also applied these weights to the 19,422 protein-coding genes and 470 phenotype pairs" the equation is for SE-GPS_Oti, where i=gene-phenotype, hence the wording 19,422 genes and 470 phenotype pairs is hard to understand, do you mean pairs made up of 19422 genes and 470 phenotypes ie 19422*470, or only a subset – I guess only a much smaller, subset, but when not stated explicitly it is hard for me to know if I am following or not…..In the legend of extended data Figure 2 "Open Target dataset for 1,150,086 SE-GPS across n = 9,188 gene-phenotype combinations and c) OnSIDES dataset for 1,104,048 SE-GPS across n = 7,713 gene-phenotype combinations" this is confusing as SE-GPS is defined for "I" = gene-phenotype, but looks like there is an SE-GPS for each feature. So the y-axis percentile is based on the SE-GPS for 9,188 gene-phenotype pairs, but then you take a subset of the j=1,n features to make the plots per row?
p. I would like to see a plot of the distribution of SE-GPS_Oti, it must be very skewed. It could be added to Extended Data Fig 2. The a. and c x -axis starts at 98th percentile. It looks like only e-QTL contribute to percentiles less than 98th percentile in

which case, it gives me an uneasy feeling about the statistic used. The eQTL only contribute when they have another feature. It is hard to interpret the per sd results at the top of page 7, without understanding of the meaning of one unit in distributional terms.
q. Extended Data Fig 2 a and c look similar but this simply reflect that most of the weight for the side effects is associated with clinical variants. Clinical variants 1, 2 and 3 are not explicity defined, only by first, second, third in methods. Add definition to the legend. You could show the distribution of the beta weights per feature label?
r. Extended Data Figure 2 legend "The sample size (n) and" sample size is ambiguous, this is the number of observation per features.
3. Results
a. This explanation "These differences may be attributed to several factors, including differences between on- and off-target effects in each category, variations in the side effects reported between the clinical trial phase and post-marketing and the filtering of side effects in our dataset with a frequency greater than 5%." On page 8 makes it very hard to evaluate if the results generated are useful.
b. Page 9 focus on top < 1%. It is not clear to me if anything new is discovered.
c. Page 11 "Finally, we extended both methods to 19,422 protein-coding genes and 470 phenotypes" I thought it had already been applied to this. Make clear earlier on the genes and phenotypes being used at each stage. So this section is where something new is discovered, but the examples given go back to what is known. The section about IL2RA is interesting and should go in the discussion.
4. Code. I was surprised not to receive the code to review. I looked instead at https://zenodo.org/records/10095684 from the published paper. This code starts with OT_drugdataset split into CV sets. The code to generate this file from downloaded data and the primary data were not provided, nor the final dataset are provided. I found this disappointing. The files extracted from OT and Onsides do not seem big, and especially the derived files used for the regression should be supplied.

Other comments
1. Page 5. Correlation of 0.7 increasing to 0.74. It is ambiguous what is being correlated. Explain in more detail?

2. Page 5. The 987 drugs from OpenTargets and 806 drugs from Onsides, provide venn diagrams of overlaps of drugs and genes and side effects and other logical comparators.

3. Methods.Page 14 Exclusion of 58 comm side effects – can these be listed in a supp Table.
Phecodes that lacked genetic evidence. What is the N. Can these side effects be listed.

4. Page 14 "Finally, we removed any phecodes that lacked genetic evidence" Both drug-indications and Side effects are mapped to phecodes. Would be helpful to distinguish between, SE-phecode and DI-phecodes.
Not clear what you mean. Each phecodes is linked to a DI-gene, but the gene had no entries from the 5 features.

In general, The use of the word "phenotype" is confused and confusing.
On Page 5 "We integrated both side effect datasets with human genetic evidence at the gene-phenotype level", I think here phenotype is side effect.
In the next line "GWAS phenotype" could be "GWAS trait"
Later in the section is eQTL phenotype suggest trait-eQTL that maps to a phecode-eQTL
Page 16 disease associated phecodes

5. Page 15 "We subsetted both drug datasets to drugs that either had a box warning or had been withdrawn due to toxicity risk. In Open Targets, these side effects are annotated as toxicity classes, which we then mapped to phecode categories as follows" Is this supposed to be where the 15 phecode categories comes from – fewer than 15 here

6. Removal of drugs from OpenTargets that were present in Onsides – explain why.
Is it because the databases actually extract info from the same primary sources.
What is a definition of a drug in this case? Some drugs are very similar and could be combined?

7. Page 19: "The clinical variant category was derived from genetic association data from ClinVar19, HGMD20 and OMIM21, which we consolidated into a single feature recorded as the number of overlapping entries." In the original paper these were separate; explain why now consolidated.
8. In the original paper, pQTLs were used as well as eQTLs; explain why these are not usesd

9. Clinical Variants. "We applied a more stringent filtering approach than previously.." this means compared to ref 13 or 15… explain why

10. StatisticaL Analysis: "We calculated the side effect ratio of reporting frequency (RRF) as detailed in equation 6 from Paccanaro et al4. " Please repeat it here, so save readers looking it up.

11. Extended Data 4. Since this is a subset of N=967 genes selected from Extended Data Figure 1. Add number of genes to legend of extended Data Figure 1, to allow comparison.

12. Online Tool:

a. The first gene listed in “Gene examples” GRIN2A – gives no data, which doesn’t seem ideal – even though GRIN2A is listed in Table 2. Other genes in Table 2 are not in the database.
b. The number of genes and phenotypes on the front page differs from the numbers on the summary page (= number in Fig 1). The numbers to which the algorithm was applied is different to the numbers with entries in the search. Eg, The front page says 466 phenotypes but the phenotype search page only has ~120.
c. In the Summary info, an explanation should be given why many entries have no information in the ONSIDES or OpenTargets side effects columnAllow download on the files behind the search.
d. It seems strange that there is no information about drugs. Make a link from gene name to drugdatabase?

13. Figure 2. Improve the legend so a reader can understand as standalone. Write so it is clear what the OR is for.

14. Figure 3. Make it clear that 80% of OpenTargets and OnSides is used to generate the SE-GPS, and that this Figure is results for application of the remaining 20%

Reviewer #3

(Remarks to the Author)
Duffy et al have submitted a well-written, interesting manuscript which describes their genetic priority score for predicting drug side effects using human genetic evidence. There is arguably an overlap in concept with their 2024 Nature Genetics publication, but I feel this study successfully builds on their prior paper and focuses on a novel, more specific topic: side effect prediction. My recommendation is to accept with minor edits, including inclusion of the code used in this work (manuscript currently says this will be made available upon publication). I have not been able to review or run any of the code.
I can confirm I was able to access all other files pertinent to the review and was able to access the online web tool.

I would like the authors to address the following points:
Line 92: ‘we removed common side effects observed in greater than 5% of drugs’. As there is undoubtedly value in predicting common side effects such as nausea, I would like the authors to expand on this point further. Is their approach not suited to predicting common side effects?
Line 103: what is the overlap in drug side effects, compounds etc reported in OpenTargets and OnSIDES? I do not believe these resources will be completely distinct, and so a detailed characterisation of the differences between these datasets is required before assigning them to be training and validation datasets.
Line 127: ‘We retained this side effect filter..’ Please add a brief comment to the Discussion section regarding how this may limit the utility of the approach for predicting side effects of a similar phenotype to the disease
Line 224: ‘In OnSIDES, we observed similar enrichments for the positive SE-GPS DOE..’ Does this present any caveats for application of the approach? Please expand slightly.
Line 252 and elsewhere: ‘mendelian’ requires a capital ‘M’
Line 282: remove comma before ‘therefore’
Line 340: a table would be much clearer for illustrating the mappings between toxicity classes and phecode categories
Line 409: a double comma
Line 728: ‘at which level of significance?’ appears to be a comment for the authors, but needs to be addressed!
Line 758: ‘ (include ref?)..’ – comment needs to be addressed and reference included!
Line 761: ‘at 5% or multiple testing corrected?’ Please address
Line 806: ‘(at which level of significance?)’ – this is a good question, please answer
Line 822: ‘(at which level of significance?)’ – what they said
Line 836: ‘(at which level of significance?)’ – same again
The web tool is responsive and has clear utility, but more detailed documentation such as worked examples or tutorials would be extremely beneficial.

I look forward to seeing the revised manuscript.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have responded to all of my remarks and have addressed them satisfactorily.

(Remarks on code availability)
There is a README file included with the code that appears to provide instructions for installation and running the application. However, there is no explanation on how to retrieve the data underlying the analysis. If this is a requirement from the journal, such information should be added to ensure full reproducibility.

Reviewer #2

(Remarks to the Author)
The authors have made an impressive response document. I find the paper much clearer the added tables and Figures are helpful.

I have 2 points which I feel should be addressed.
1) Comment 23 of reviewer 1 asked about correction for multiple testing. The authors state that they use $p< 0.05$ as significant. I agree with reviewer 1 that the p-value threshold for significance should be stated (at line 926 it says “P-value > 0.05”, which should be “P-value < 0.05”. Furthermore, there should be a threshold for significance that accounts for multiple testing.
2) I find Supp Fig 8 (p32 of rebuttal document) illuminating. I believe the authors need to be upfront. If an SE-GPS > 1.6 is observed then this is useful information, but that the vast, vast majority of the time the SE-GPS is uninformative. This should feature in the abstract, Figure1, Discussion.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have addressed my concerns in the response to reviewer comments and I am satisfied by their additions to the manuscript. I have been able to run the code provided and have reproduced their results, but I suggest the authors add additional comments to the code as it was not always clear to me what each chunk was doing and why this was necessary. Overall the manuscript is improved.

(Remarks on code availability)
See above

REVIEWER COMMENTS

Authors reply to all Reviewers: We thank the Reviewers' for their helpful comments. We believe that the manuscript has improved considerably as a result of their suggestions and feedback. Notably, as the Reviewers' suggests, we have made several additions to the manuscript, including: 1) we integrated L2G and eQTL phenotype as one 'GWAS loci' feature; 2) instead of examining the association between the continuous SE-GPS and drug side effect, and looking at the unit increase, we analyzed a binarized version of the SE-GPS instead; 3) we have included druggable gene annotations as well as a druggability score (DrugnomeAI) when providing the SE-GPS across all protein-coding genes; 4) we have updated the web application to improve usability, including additional filtering options, the ability to download results and clearer user instructions; 5) we included two additional tables to further explore how the SE-GPS can predict both severe adverse events and identify potential side effects for undrugged targets; and 6) we have provided the full analysis code and the final datasets here: https://zenodo.org/records/15334136?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImY1ZWY3ZmVmLWM0YmEtNGM1YS1iZjkxLTgxYWJlMGM2OGM1MCIsImRhdGEiOnt9LCJyYW5kb20iOiJiZTI1ZTJjMjRkNGMxOWM5MmQ2OWY3ZjhmN2UwMjFiYSJ9.dGGnnoLIUVaXQ4VKndr8c2vayJRAC05b0caR7gakbUrncTP9fg1hnSVzP132KFKErSKBuRKhc32ONJJT8Wfbrw and https://github.com/rondolab/SE-GPS.

Furthermore, we made four additional changes to the analysis which were: 1) we have kept phecodes that map to the Neoplasms category to allows us to include oncology-related phenotypes that are reported as side effects from non-oncology drugs (we still remove oncology drugs using the Anatomical Therapeutic Chemical (ATC) classification); 2) we removed the 40 categorical traits from Genebass, following this suggestion by Dr Shicheng Guo[1] on the Open Targets community feedback page**;** 3) we have updated the genetic evidence data using the very recent release of Open Targets Platform v25.03 (March 2025) and 4) we have moved the extended data and tables to supplementary material.

Following these updates, we note that the main findings and conclusions remain the same.

Reviewer #1 (Remarks to the Author):

This study extends a Genetic Priority Score from a recent publication to a Side Effect Genetic Priority Score (SE-GPS). The in silico tool is designed to predict adverse drug side effects for 19,422 protein-coding genes and 470 side effects using comprehensive genetic data from single variants to GWAS. High SE-GPS scores (top 0.01%-0.40%) were strongly associated with increased risk of drug side effects, particularly severe ones like those linked to boxed warnings or toxicity-related drug withdrawals, with fold increases of up to 45.5. The method seems to be technically sound, nevertheless the manuscript appears somewhat underdeveloped in certain sections and resembles more of an analysis plan than a manuscript. E.g. The results section contains elements which could be moved into the method section. This would allow for more space to focus on presenting further results. I have several comments aligned with this direction:

Authors reply: We thank the Reviewer for their comments and we are grateful for their critical feedback. We have addressed their comments below and have updated the manuscript accordingly.

Abstract: The abstract remains too vague and presents the results in a partly unclear manner. A revision is needed to provide more specific and concrete information. E.g. The phrase "five lines of human genetic evidential support" feels a bit abstract. Is the second-to-last sentence a validation as drugs with

known side effects have a high score?

Authors reply: We appreciate the reviewers feedback and have revised the abstract to explicitly mention the lines of genetic evidence as well as provide a clearer summary of the construction of the score in Open Targets, OnSIDES before applying this to all genes. We also modified the second-to-last sentence to also include drug targets supported by a SE-GPS with no current clinical trial evidence, to highlight utility of the SE-GPS tool.

Manuscript changes: We have revised the Abstract:

Many drug failures in clinical trials are due to inadequate safety profiles. A genetic tool that predicts side effects offers a valuable approach to prioritizing safer drug targets. We developed an in-silico side effect genetic priority score (SE-GPS) that leverages human genetic evidence across clinical variants, coding variants, single variants and genome-wide association trait loci to inform the likelihood of side effect occurrence for a given drug target. We construct the SE-GPS in the Open Target dataset using post-marketing side effect data, externally test it in OnSIDES using side effects reported from drug labels and then generate SE-GPS for 19,422 protein coding genes and 502 phecodes. We observe that restricting to at least two lines of genetic evidence conferred a 2.3- and 2.5-fold increased risk in side effects in Open Targets and OnSIDES respectively. When restricting to drugs with boxed warnings or drugs withdrawn due to risk of toxicity, this enrichment increased to 2.5- and 5.2-fold. Finally, we highlight drugs with side effects as well as targets with no current clinical trial evidence, that are supported by a high SE-GPS, to demonstrate utility of the SE-GPS tool. To consider drug mechanism, we incorporated the direction of genetic effect into a directional version of the score called the SE-GPS-DOE and make all predictions publicly available in a web portal (https://rstudio-connect.hpc.mssm.edu/sideeffect-geneticpriorityscore/).

1. Line 77: The five genetic features should be provided in the introduction and together with all data sources in Figure 1.

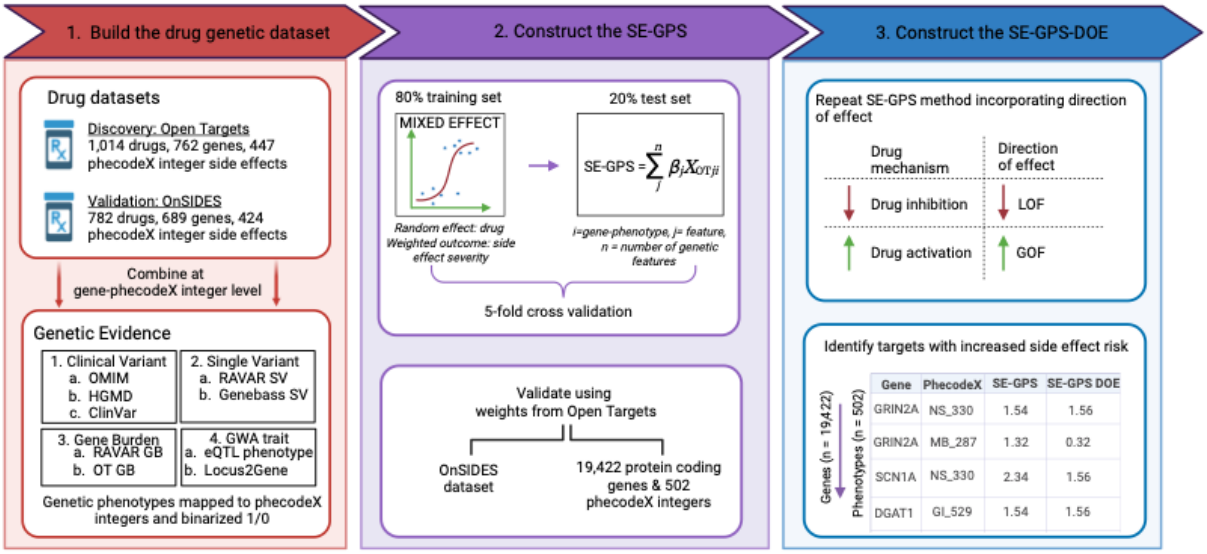
Authors reply: We have now provided the nine data sources that contribute to the four genetic features in the Introduction. Additionally, in response to a later comment, we have now combined L2G and eQTL into a single feature and are now therefore using four distinct genetic features rather than five. We have updated Figure 1 to include these data sources.

Manuscript changes: We have added the following text to the introduction:

'These features include: 1) clinical variant evidence from ClinVar, HGMD and OMIM, consolidated into a single feature quantified as the number of overlapping entries; 2) single coding variants encompassing pLOF and missense single variants curated from Genebass and RAVAR; 3) Gene burden tests from Open Targets and RAVAR and 4) genome-wide association (GWA) loci, represented by two separate features: Locus2Gene and eQTL phenotype.'

We have revised Figure 1:

Fig. 1. Schematic of steps to build the SE-GPS to assess side effect risk.



A workflow of the data sources and steps to construct the SE-GPS and SE-GPS-DOE as outlined in this analysis. The SE-GPS and SE-GPS-DOE were created in the Open target dataset (discovery), validated in OnSIDES and then generated for 19,422 genes and 502 phecodeX integers, for which 15,139 genes linked to 499 phenotypes had support from at least one genetic feature and directional evidence (n= 146,011). SE, side effect, OT, Open Targets; SE-GPS, side effect genetic priority score; SE-GPS-DOE, side effect genetic priority score with direction of effect. Created with BioRender.com.

2. Caption to Figure 1 is not finished. "**Supplementary Fig. 1**). using a crowdsourced cscore (include ref?).. Sh"

Authors reply: We thank the reviewer for pointing out this clerical error. We have updated the caption to include this reference for Figure 1 and Extended Figure 1 (now Supplementary Fig. 6)

Manuscript changes: We have revised the following Figure caption for Supplementary Fig .6 (previously Extended Data Fig. 1).

The Open Target dataset (n=1,014 drugs, 762 genes and 447 phenotypes) was split into 80% training and 20% test sets of non-overlapping groups of unique gene-phenotype pairs in five-fold cross-validation. A mixed effect regression was run for each cross-validation training set with drug side effect as the outcome, the four genetic features and 16 phecode categories as the predictor variables and the drug as the random effect variable. The side effect outcome was weighted by severity using a crowdsourced severity score[2]. Shown is a forest plot of beta coefficients with 95% CIs from the four genetic features included in each cross-validated model. Each cross-validated sample is color labeled and filled circles indicate a beta coefficient with a significant *P*-value > 0.05 and the 95% CIs are defined as error bars. The red dashed line represents the null beta coefficient (β = 0). CI, confidence interval.

3. Line 87-92, revise for clarity. Repeated description of the methods, but unclear due to the use of different terms. The section "Construction of the drug genetic dataset" reads more like a methods description than a results section and should be moved/revised.

Authors reply: We appreciate the reviewer's feedback and have revised this section for clarity. We have moved the quality control steps to the methods section (removal of common side effects and exclusion of oncology drugs) and ensured the terminology is the same between the methods and the results to avoid confusion.

Manuscript changes: We have revised the "Construction of the drug genetic dataset paragraph" for clarity and consistent terminology between the methods and the results.

Please see Line 106-130 for paragraph starting: 'We utilized two datasets that report side effect data: Open Targets[3] as our discovery dataset, which compiles post-marketing surveillance data from the FDA Adverse Event Reporting System (FAERS)[4], and OnSIDES as our validation dataset[5], which extracts adverse drug reactions from drug labels reported during clinical trials.'

4. Line 99, are they removed twice? Compare to line 92/93

Authors reply: No, common side effects greater than 5% are only removed once. We realize our wording was unclear and have revised this section for clarity. The filtering details have been moved to the methods and we have updated **Supplementary Fig.1** to show the ratio of reporting frequency before QC only to avoid confusion since both plots show a similar strong correlation (0.7 compared to 0.74).

Manuscript changes: We have revised the following text in the Results section:

'To measure the frequency of reported side effects and compare differences in side effect reporting across clinical trial data (OnSIDES) and post-marketing data (Open Targets), we plotted the ratio of reporting frequency (RRF), calculated as the normalized count of drugs associated with a given side effect from Paccanaro et al. **Supplementary Fig. 1** shows the RRF of each side effect in Open Targets, correlated against the side effect data in OnSIDES. Both datasets indicate that most reported side effects are drug-specific, with similar reported frequency (r=0.7).'

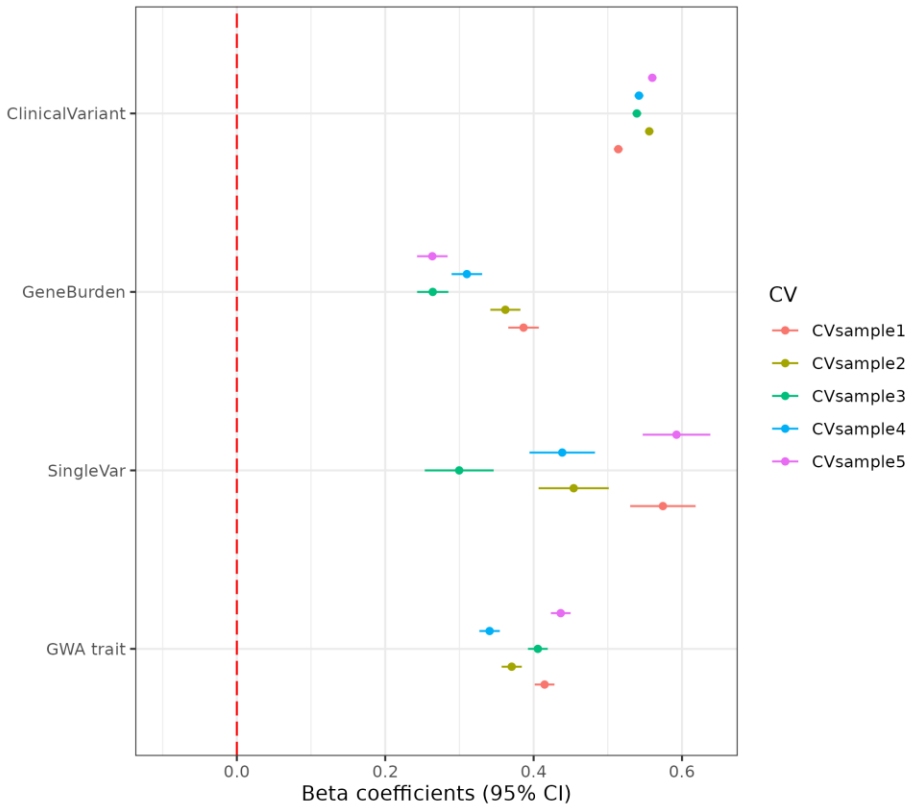
5. Line 115/166: which subset? What are the 5 genetic features? Clinical variants, coding variants and GWAS phenotypes, GWAS and? The reason why Locus2Gene24 and eQTL phenotype were not merged is missing. Polygenic risk scores (i.e. The Polygenic Score Catalog) are missing.

Authors reply: We now provide clarity on the description of the genetic features used in our method. Furthermore, we have decided to merge L2G and eQTL, especially since the p-values from eQTL phenotype were non-significant in each cross validated mixed model. As a result, we have combined L2G and eQTL into a single feature, 'GWAS trait', binarized to indicate the presence or absence of either data source, and have used this feature in all analyses (previously Extended Fig 1, now Supplementary Fig. 6). We note that following this change, all beta estimates in the mixed model are now significant. We did not include polygenic risk scores because we already include common variant association loci in our approach and also because polygenic risk scores are typically calculated across many variants and genes across the genome. However, we recognize the potential of PRS for risk stratification in drug discovery and have included this in our discussion.

Manuscript changes: We have added the following text to the Results:
'Using gene-phecode pairs as the common identifier, we combined both side effect datasets with the nine human genetic data sources at the gene-phecode level, consolidating these into four genetic features to use for analysis: Clinical Variant, Single Variant, Gene Burden, and GWA Trait to reflect the

different types of genetic support. These features were constructed as follows: the Clinical Variant feature was derived from genetic association data from ClinVar[6], HGMD[7] and OMIM[8], consolidated into a single feature recorded as the number of overlapping entries. The Single Variant feature comprised pLOF single variants curated from Genebass[9] and RAVAR[10], while the Gene Burden feature consisted of gene burden tests curated from Open Targets[3] and RAVAR. Lastly, the GWA trait feature consisted of genes identified from genome-wide association significant variants identified using Locus2Gene[11] and eQTL phenotype[12]. For Single Variant, Gene Burden and GWA trait, we binarized the features based on the presence or absence data from either source. A detailed description of each data source is provided in the Methods and an overview of these gene-phecode observations across 19,422 genes and 502 phecodes is shown in **Supplementary Fig. 2'**

Manuscript changes: We have revised Supplementary Fig. 6



Manuscript changes: We have added the following sentences to the Discussion:
'Additional methods includes Mendelian Randomization, which offers several advantages, particularly the ability to infer causality rather than associations[13] and polygenic risk scores, which offer the opportunity to stratify patients in clinical trials according to disease risk[14].'

6. Line 119, where does the 15 phecode categories come from? The number should be described as a result in the section before.

Authors reply: PhecodeX includes 18 categories in total, aligning with ICD chapters that reflect broad organ systems[15]. We exclude phecodes classified under the Pregnancy and Neonatal category. Initially, we also excluded the Neoplasms category due to the intrinsic cytotoxicity and different acceptable side

effect profiles of oncology drugs. However, upon further consideration, while we still remove oncology drugs using the Anatomical Therapeutic Chemical (ATC) classification, we have not removed the Neoplasms category. This allows us to include oncology-related phenotypes that are reported as side effects from non-oncology drugs. For example, a rare side effect reported from the drug Pioglitazone, used to treat type 2 diabetes, is Bladder cancer[16].

Manuscript changes: We have added the following text to the results to the 'Construction of the drug genetic dataset'

'To construct the drug datasets, we mapped the side effect and drug indication data to phecodeX integer terms across 16 phecode categories, similar to the GPS, and outline additional quality control steps in the **Methods'**

Manuscript changes: We have added the following text to the Methods

'We excluded the phecode categories Neonatal and Pregnancy, resulting in 16 remaining PhecodeX categories.'

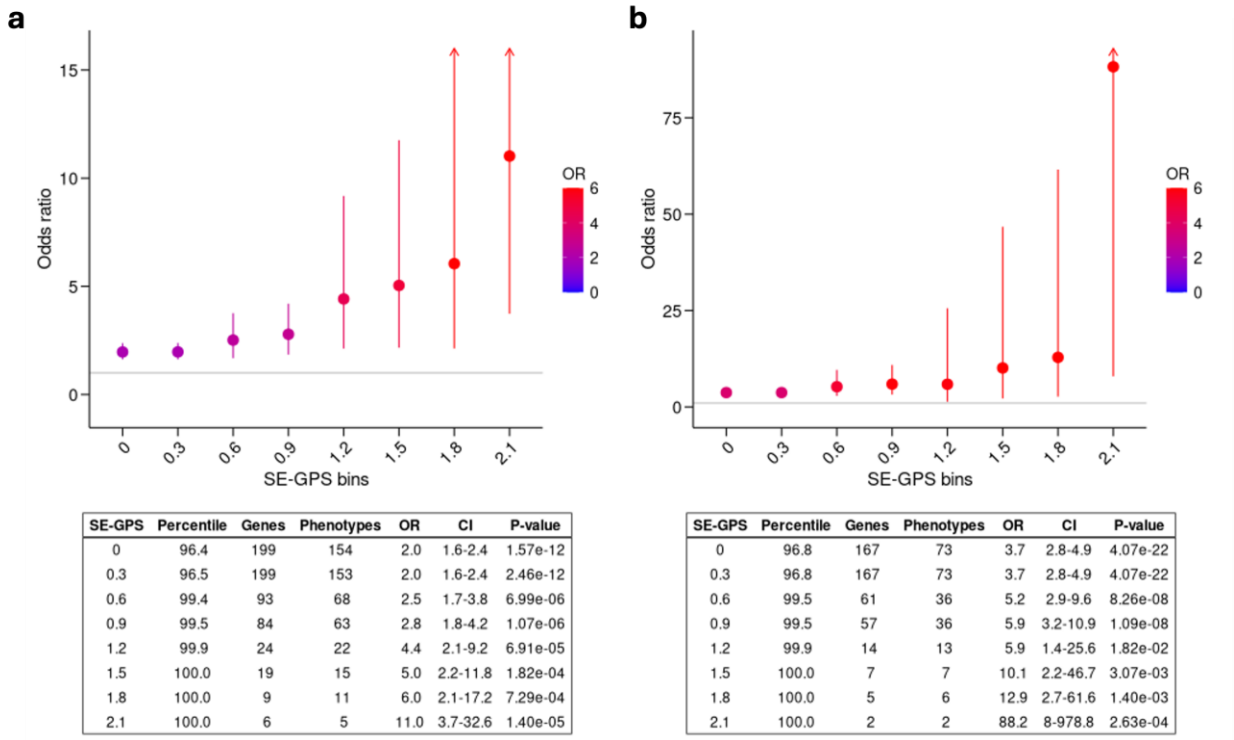
7. Line 120-126. It is unclear why the finding that "phenotype terms overlap with the drug indication" (e.g., [provide 1-2 specific examples]) is included under the section "Association of genetic features with drug side effects." It seems more relevant to the first result or method section. Clarifying this connection would help improve the flow and focus of the section. Furthermore, this side effect filter should also be used in the analysis showed in Supp. Fig. 3.

Authors reply: We appreciate the reviewers feedback and agree that this filter makes more sense in the 'construction of the drug-genetic dataset' section. We have used this side effect filter throughout as our side effect outcome (i.e. removal of phenotype terms that matched a drug indication) and thus to avoid confusion we have removed Supplementary Fig. 1, which previously included overlapping terms and changed Figure 5 to evaluate the association of the SE-GPS with severe drug side effects removing overlapping terms with the drug indication. We include hypothyroidism as an example of a drug indication and side effect reported by levothyroxine sodium.

Manuscript changes: We have added the following text to the results to the 'Construction of the drug genetic dataset'.

'We observe that a proportion of side effects in Open Targets and OnSIDES shared their phecode terms with the drug indication (9.29 % and 11.46%, respectively). This overlap is likely due to several reasons, including side effects that result from an exaggerated pharmacological response directly related to the drug's therapeutic effect, misclassification of disease symptoms as side effects, and issues with data reporting. For example, the drug levothyroxine sodium reports hypothyroidism as both an indication and side effect. To ensure this overlap did not drive our genetic enrichment analyses, we excluded those side effects where the drug was approved for an indication that shared the same phecode term. We retained this side effect filter, i.e. removal of phecode terms that matched a drug indication, for subsequent analyses, as in previous studies

Manuscript changes: We have revised Figure 5:



SE-GPS	Percentile	Genes	Phenotypes	OR	CI	P-value
0	96.4	199	154	2.0	1.6-2.4	1.57e-12
0.3	96.5	199	153	2.0	1.6-2.4	2.46e-12
0.6	99.4	93	68	2.5	1.7-3.8	6.99e-06
0.9	99.5	84	63	2.8	1.8-4.2	1.07e-06
1.2	99.9	24	22	4.4	2.1-9.2	6.91e-05
1.5	100.0	19	15	5.0	2.2-11.8	1.82e-04
1.8	100.0	9	11	6.0	2.1-17.2	7.29e-04
2.1	100.0	6	5	11.0	3.7-32.6	1.40e-05

SE-GPS	Percentile	Genes	Phenotypes	OR	CI	P-value
0	96.8	167	73	3.7	2.8-4.9	4.07e-22
0.3	96.8	167	73	3.7	2.8-4.9	4.07e-22
0.6	99.5	61	36	5.2	2.9-9.6	8.26e-08
0.9	99.5	57	36	5.9	3.2-10.9	1.09e-08
1.2	99.9	14	13	5.9	1.4-25.6	1.82e-02
1.5	100.0	7	7	10.1	2.2-46.7	3.07e-03
1.8	100.0	5	6	12.9	2.7-61.6	1.40e-03
2.1	100.0	2	2	88.2	8-978.8	2.63e-04

The Open Target and OnSIDES datasets were restricted to drugs with a boxed warning or drugs withdrawn due to toxicity risk and the side effect phecodes matching the toxicity class. The association of increasing SE-GPSs with these severe drug side effects was investigated by binning the boxed warning dataset into 0.3 increments of the SE-GPS and comparing SE-GPS greater or equal to each increment with SE-GPS equal to zero. A logistic regression model was performed for each increment bin with drug side effect as the outcome variable and the SE-GPS bin as the predictor variable, adjusting for phecode categories as covariates. ORs with 95% CIs are defined in the forest plot as circles and error bars, with filled circles indicating an OR with a significant P-value > 0.05. Panel a displays results for Open Targets (n = 69,290 independent drug– gene–phenotype combinations) and panel b displays results for OnSIDES (n = 30,652 independent drug– gene–phenotype combinations). The grey vertical line represents the null odds ratio (OR=1). CI, confidence interval; OR, odds ratio; SE-GPS, side-effect genetic priority score.

8. What is the rational to report the top 0.40%, 0.05% and 0.01% of the SE-GPS (equivalent to scores 184 greater than 0.9, 1.5 and 2.1)? The 0.6 score shows also an OR>2.

Authors reply: We initially selected 0.9 as our cutoff as this corresponded to genetic evidence from at least two features. However, following our re-run > 0.6 now corresponds to greater than two features as shown in Supplementary Fig. 7 and Supplementary Fig. 9 for both Open Targets and OnSIDES. As a result, we use this as our cutoff instead. Additionally, we provide all gene-phecode observations with scores greater than 0 on our website, to allow users to explore different cutoff thresholds. We have clarified this threshold in the results.

Manuscript changes: We added the following sentence to the results

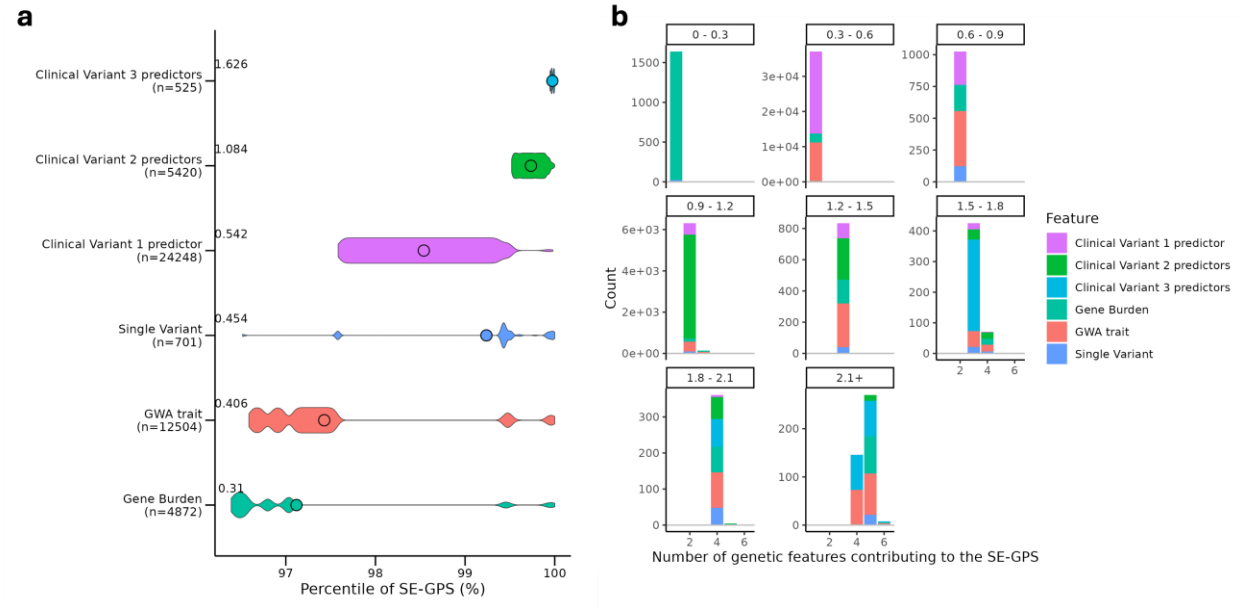
We selected a cutoff greater than 0.6 as our initial threshold to define a high SE-GPS, corresponding to evidence from at least two genetic features (**Supplementary Figure. 7**), which reflects an OR > 2.3 and corresponds to 365 genes and 254 phecodes.

10. Line 142-144, The reader (may) have some difficulty to follow this finding, as Extended Data Figure 2 contains four subplots. Furthermore, it is unclear why the clinical variant feature was split according to the number of data sources in this section, but this approach was not applied in the earlier results sections, such as in Figure 2 and Supplementary Figure 3. The authors have to provide additional guidance on how to navigate the figure e.g. provide more detailed results in the manuscript.

Authors reply: We have split Extended Data Figure 2 into two figures: **Supplementary Figure. 7**, now shows the contribution of each genetic feature to the SE-GPS in the Open Target dataset, **Supplementary Figure. 9** displays the SE-GPS across OnSIDES. As the Clinical Variant feature was coded as the number of overlapping entries (0/1/2 or 3), we wanted to show the enrichment of observations that had 2 and 3 features which is why we split the observation in this figure only. For Figure 2 on the other hand, the Clinical Variant feature is treated as a continuous variable rather than a categorical variable, and thus Figure 2 reflects the side effect risk with each unit increase in the Clinical variant predictor

Manuscript changes: We have revised the following Supplementary Fig. 7 and Supplementary Fig. 9

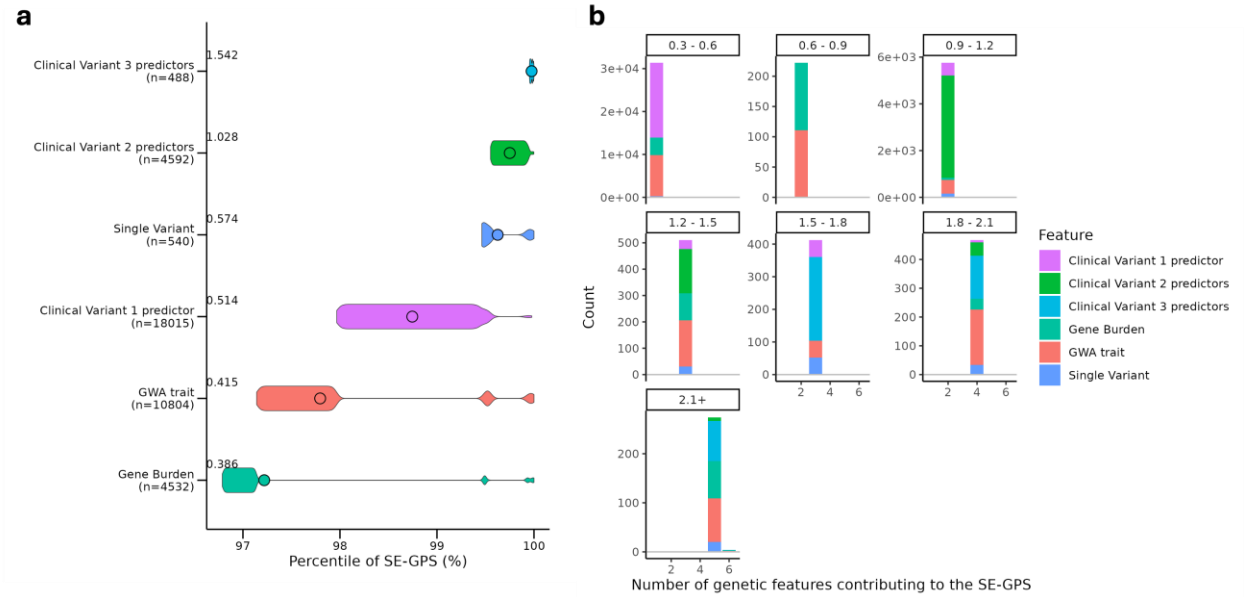
Supplementary Fig. 7 Contribution of each genetic feature to the SE-GPS in Open Targets.



In panel a) violin plots show the distribution of each genetic feature that collectively sums to form the SE-GPS in the Open Target dataset for 1,273,056 SE-GPS across n = 34,0614 gene-phecode combinations. The x-axis represents the percentile of the SE-GPS, starting at 97% to show non-zero scores only, while the y-axis separates the scores across each contributing genetic feature. The width of each violin plot represents the density of the genetic feature at each percentile, with the mean percentile marked as a circle. The total sample size of gene-phecodeX integer observations for each feature (n) and the mean weight from the five cross-validated samples is recorded under each feature on the y-axis, ordered by

increasing value of these weights across the six features. The clinical variant feature was split according to the number of data sources (1, 2 or 3) for each gene phenotype observation. In panels b) bar plots show the contribution of the genetic features to the SE-GPS at 0.3 increment bins in the Open Target dataset. On the x-axis of each bar plot is the number of genetic features contributing to each score, colored by each feature present. The y-axis shows the count for each feature. In both plots, we demonstrate that as the SE-GPS increases, the number of features contributing to the score increases. SE-GPS, side-effect genetic priority score.

Supplementary Figure. 8 Contribution of each genetic feature to the SE-GPS in OnSIDES.



In panel a) violin plots show the distribution of each genetic feature that collectively sums to form the SE-GPS in the OnSIDES dataset for 1,161,760 SE-GPS across n = 29,2136 gene-phecode combinations. The x-axis represents the percentile of the SE-GPS, starting at 97% to show non-zero scores only, while the y-axis separates the scores across each contributing genetic feature. The width of each violin plot represents the density of the genetic feature at each percentile, with the mean percentile marked as a circle. The total sample size of gene-phecodeX integer observations for each feature (n=) and the mean weight from the five cross-validated samples is recorded under each feature on the y-axis, ordered by increasing value of these weights across the six features. The clinical variant feature was split according to the number of data sources (1, 2 or 3) for each gene phenotype observation. In panels b) bar plots show the contribution of the genetic features to the SE-GPS at 0.3 increment bins in the OnSIDES dataset. On the x-axis of each bar plot is the number of genetic features contributing to each score, colored by each feature present. The y-axis shows the count for each feature. In both plots, we demonstrate that as the SE-GPS increases, the number of features contributing to the score increases. SE-GPS, side-effect genetic priority score.

11. Figure caption 3, vertical red dashed line should be explained

Authors reply: We have updated the caption of Fig. 3 as well as the captions of Fig. 2, and Fig. 6, Supplementary Fig. 6 and Supplementary Figure. 10 to clarify that the red dashed line represents the

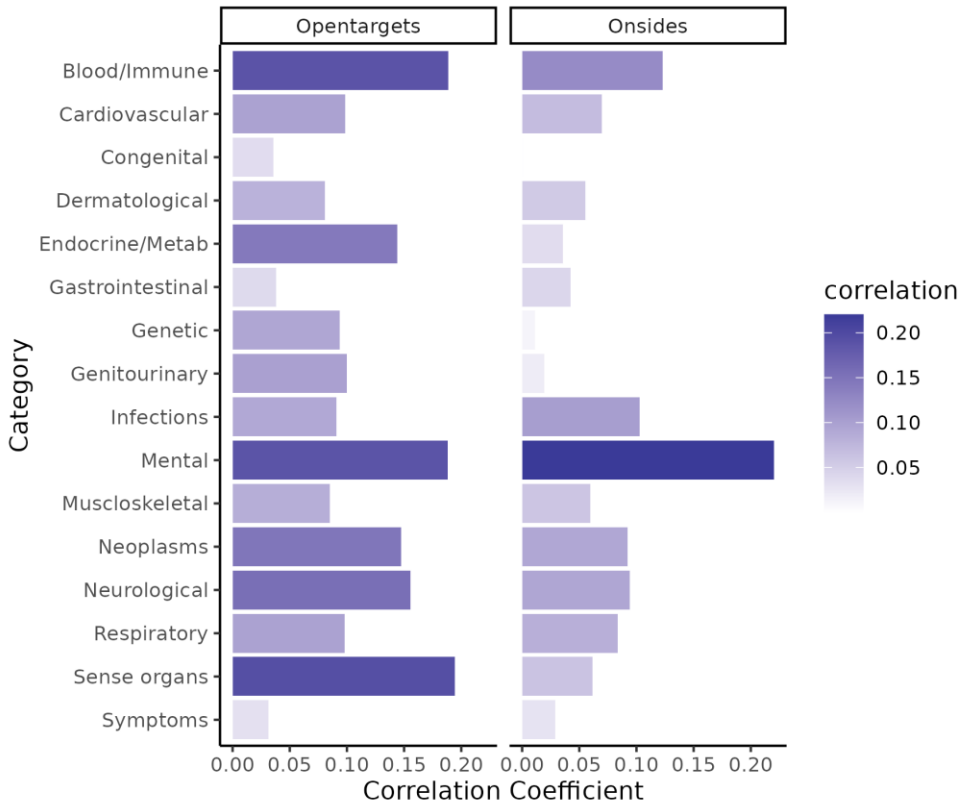
null odds ratio (OR=1). Furthermore, we have also updated the captions of Fig. 2, Fig. 5, Supplementary Fig. 16 and Supplementary Fig. 17 to clarify that the grey vertical line represents the null odds ratio (OR=1). Lastly, for Supplementary Fig. 6 and 13 we clarify that the red dashed line represents the null beta coefficient (β = 0).

Manuscript changes: We have included the following sentence in the figure captions:

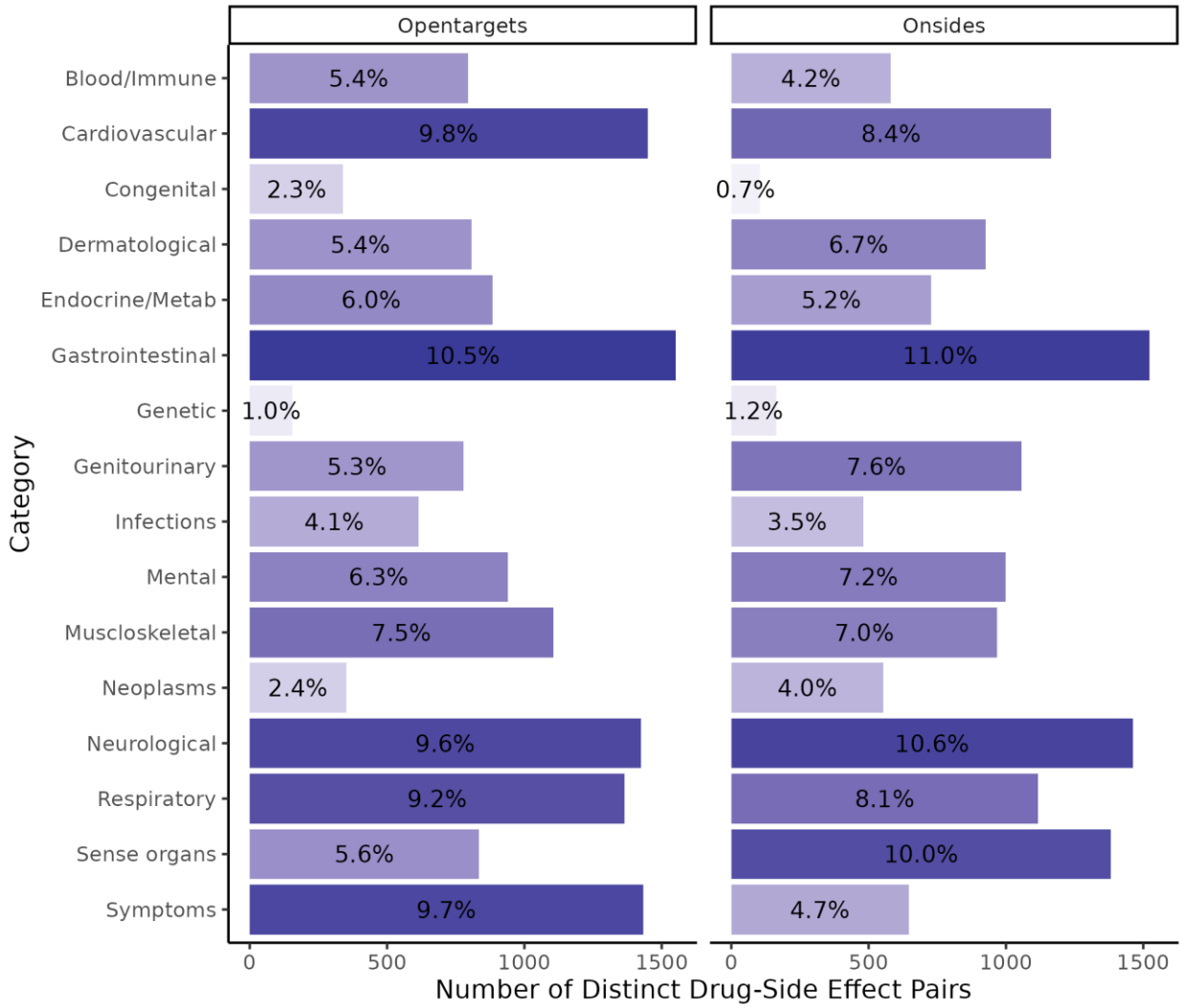
'The red dashed line represents the null odds ratio (OR=1).'

12. Line 163-166: What is the phecode "Genetic"? Is this side effect correlated with its drug indication? If so, why does this Phecode not have a higher odds ratio (OR), given that there should be a clear genetic background supporting the association? Should be discussed.

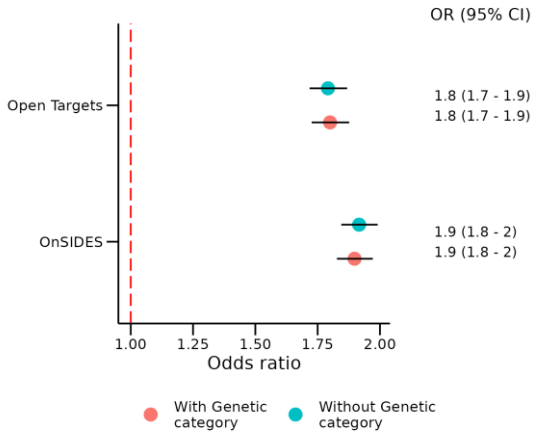
Authors reply: The "Genetic" Phecode category includes 112 phecodeX terms collapsed to 20 phecode integer terms, and side effects mapped to this category include: 'factor i deficiency' mapped to phecode GE_971, pseudoporphyria mapped to phecodeX GE_966l myelodysplastic syndrome mapped to GE_960 and polycythaemia mapped to GE_981. They are not highly correlated with drug indications, likely due to the fact that only a small proportion of side effects map to this category.



As shown in the histogram below (**Supplementary Fig. 3)**, only 1.0% of the total number drug-side effect pairs in Open Targets map to the Genetic category and in OnSIDES this is 1.2%.

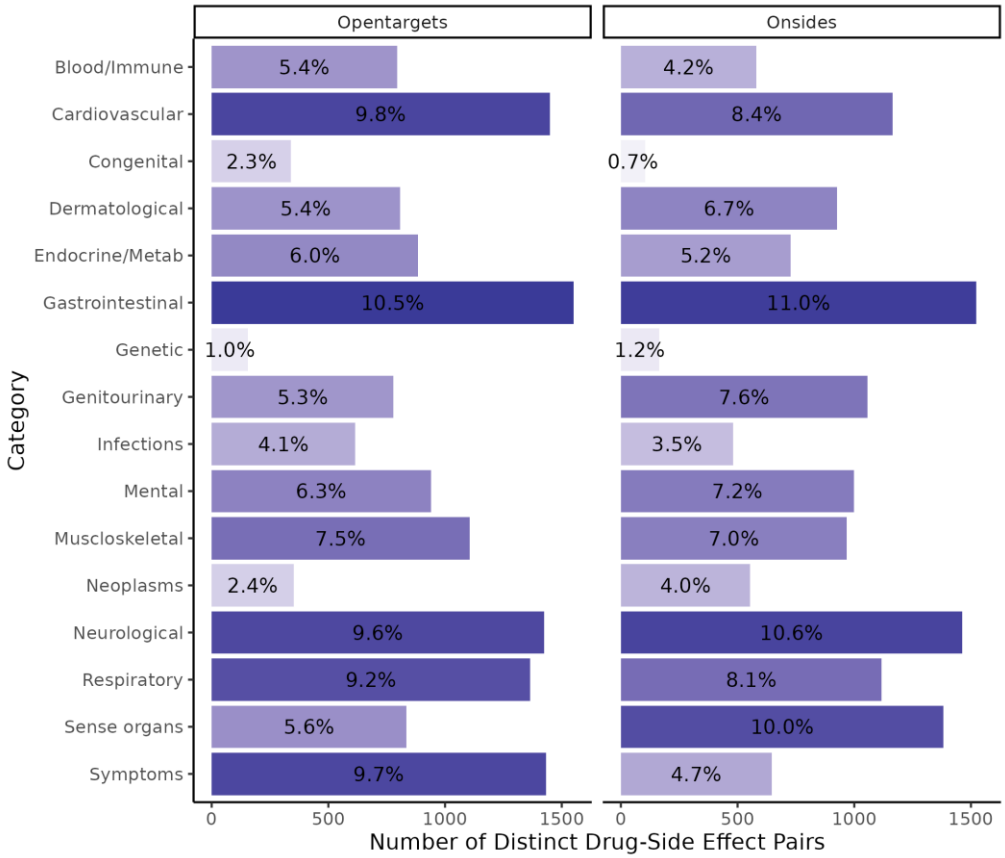


We compared the overall SE-GPS with and without the Genetic category in Open Targets and OnSIDES and found no difference between the odds ratios.
Therefore, we include this category in our analysis despite its low coverage of side effects because, as the reviewer noted, there is a clear genetic background supporting the association with a gene target. These phenotypes are important for understanding the full implications of modulating a gene target when we apply the method to all genes and all targets.



Manuscript changes: We have included Supplementary Fig. 3:

Supplementary Fig. 3 Distribution of side effect pairs by phecodeX category



Bar plots showing the number of distinct drug side effect pairs grouped by phecodeX category.

13. Line 166: It is unclear which finding "these differences" is referring to. Is the high OR of infectious disease, etc highlighting the impact of genetics or non-genetic factors, or both?

Authors reply: We have now clarified this text and have revised accordingly. Specifically, we aimed to highlight differences between side effects reported in onsides, which are side effects reported during clinical trials and Open targets, which are side effects reported from post marketing.

Manuscript changes: We added the following text to the results
'We observed significant variability in the odds ratio, highlighting that the impact of genetics is more pronounced in certain side effect categories than others. Infectious disease-related SEs had large odds ratios in both Open Targets and OnSIDES whereas congenital-related SEs were not significant in either dataset. Furthermore, the degree of enrichment differs between categories when comparing side effects reported in Open Targets and OnSIDES, potentially reflecting differences in side effect reporting between clinical trials and post-marketing surveillance'

14. Line 170. FDA box warnings drug results should be added i.e. to a plot like 3a and b, stratified by "phenotype terms overlap with the drug indication" or "phenotype terms not overlapping with the drug indication"

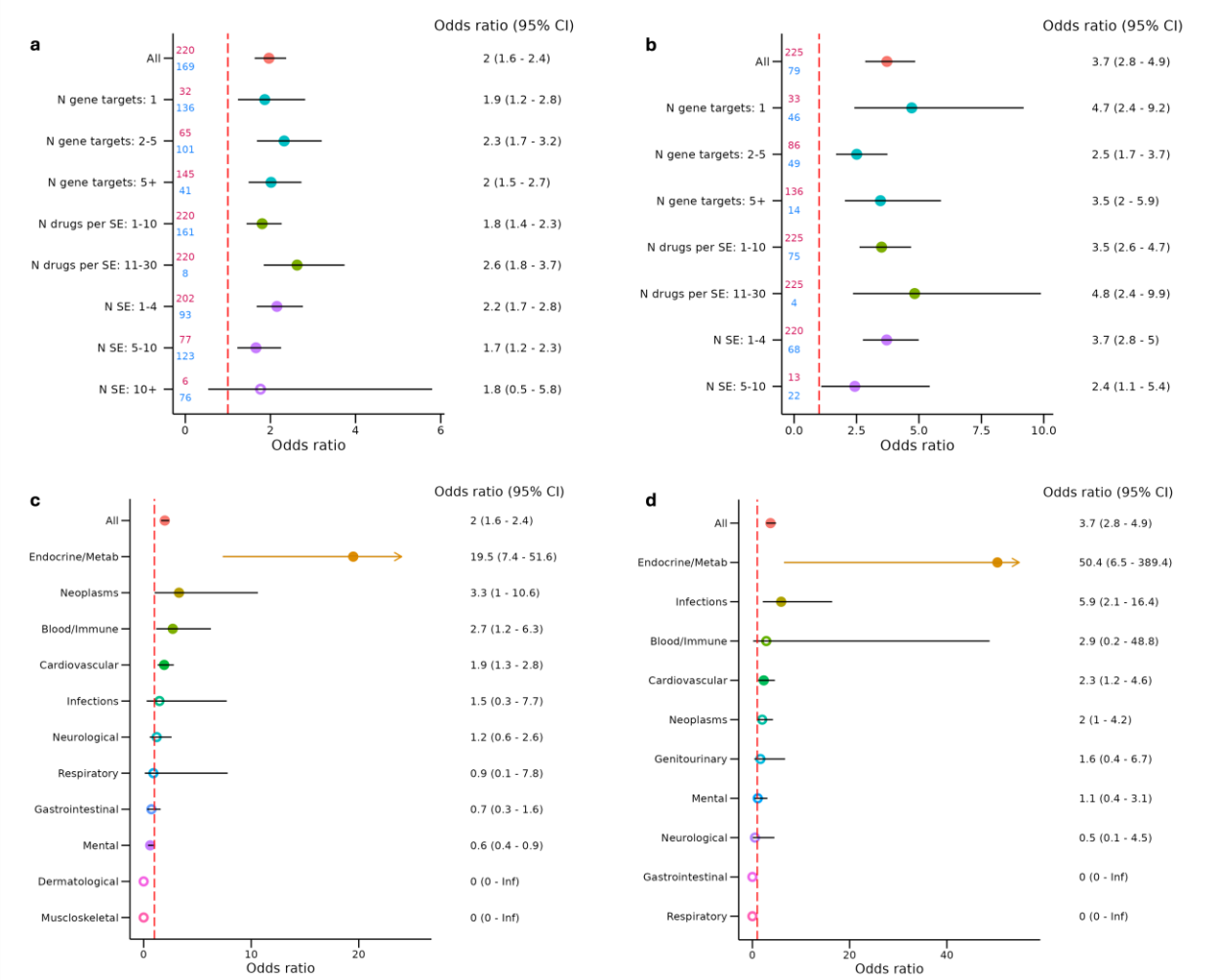
Authors reply: Similar to plot 3a and 3b we have stratified the FDA box warnings drug results and included these results as **Supplementary Figure. 10**. We see similar patterns where drug-specific side effects had stronger enrichments, however the confidence intervals where much larger and overlapped. When stratified by phecode category, 'Endocrine related side effects exhibited the largest enrichments in both datasets however also with very large confidence intervals.

Manuscript changes: We added the following text to the results

'We further explored these enrichments stratified by drug grouping and disease category (**Supplementary Figure. 10**) however note much larger overlapping confidence intervals due to lower observations.'

Manuscript changes: We added Supplementary Fig. 10

Supplementary Fig. 10 Association of the SE-GPS with drug side effects in the severe Open Target and OnSIDES datasets by drug side effect groupings.



A) *Forest plot showing ORs with 95% CI for the association between the presence of a SE-GPS > 0 (binarized as 1) and drug side effects, adjusted for 16 phecode categories using logistic regression. This was performed across the severe Open Target dataset (n = 69,290 independent drug– gene–phenotype combinations) with the OR colored in red and stratified by the number of gene targets per drug (1, 2-5, 5+; blue), the number of side effects per drug (1-4, 5-10, 10+; purple) and the number of drugs per side effect (1-10,11-30, 30+; green). For each feature, unique genes (red) and unique phenotypes (blue) are recorded on the y-axis.* ***B)*** *Replication analysis of A) using the OnSIDES severe dataset (n = 30, 652 independent drug– gene–phenotype combinations).* ***C)*** *Forest plot showing ORs with 95% CI for the association between the presence of a SE-GPS > 0 (binarized as 1) and drug side effects, stratified by phenotype category in the severe Open Target dataset.* ***D)*** *Replication analysis of C) using the severe OnSIDES dataset. Filled circles indicate an OR with a significant P-value. The red dashed line represents the null odds ratio (OR=1). CI, confidence interval; N, number; OR, odds ratio, SE, side effect.*

15. Line 180 Reference is missing. For this analysis "phenotype terms overlap with the drug indication" were removed before the analysis?

Authors reply: We thank the reviewer for pointing this out and have included the reference. Yes, for all analyses we removed phenotype terms which overlap with the drug indication.

16. Line 181 Is this the rational? ”Thus, we next considered if extremes of the SE-GPS were more likely to yield a drug side effect” as “drug toxicity is a result of on-target or mechanism-based toxicity, where the toxic effect is a result of a response directly related to the therapeutic effect” on the drug target? Should be added here again to give some guidance.

Authors reply: The rationale behind looking at the extremes of the SE-GPS is to assess whether increasing number and strength of genetic evidence, reflected by higher scores, have a greater enrichment with drug side effects. We have clarified this in the text.

Manuscript changes: We have revised the following text in the results:

‘We previously observed that at increased increments of the GPS there was an increased likelihood of a gene being a successful drug target. Thus, by applying score thresholds, we next considered whether higher SE-GPS had a greater side effect risk’

17. Line 185 Instead of acknowledging this finding, I would suggest framing it as something more in line with expectations since otherwise many drugs were not anymore on the market.

Authors reply: The reviewer makes a great point. We have removed this sentence and revised to summarize our findings at the end of this section.

Manuscript changes: We have revised the following text in the results:

‘Nonetheless, despite not observing increased enrichment across higher thresholds in OnSIDES, we observed in both datasets that incorporating evidence from at least two lines of genetic evidence can identify a subset of targets with a greater likelihood of side-effect risk.’

18. Line 193-201: What is the difference to the FDA drug box warning analysis in lines 170-178? This analysis should also be stratified by "phenotype terms overlap with the drug indication” or "phenotype terms not overlapping with the drug indication”

Authors reply: In lines 170-178, we examine the overall association of the SE-GPS with drug side effects in the severe drug dataset and we observe an OR of 2.0 in Open Targets and an OR of 3.7 in OnSIDES. In Line 193-201 we investigate the association of the SE-GPS with drug side effects in the same dataset, but stratify the score by 0.3 increments and note that by implementing score cutoffs, we observe an increase enrichment.

Previously, the outcome of this analysis was side effects, and we included side effects with terms overlapping the drug indication. For consistency however, we have now re-performed the analysis after removing phenotype terms that overlapped with the drug indication and have removed the analysis where phenotype terms overlap with the drug indication.

Manuscript changes: We have revised the following section:

‘Within this restricted set of drugs, we observed a significant increase in OR of 2.0 (95% CI =1.6–2.4, $P < 1.6 \times 10^{-12}$) in Open Targets and an OR of 3.7 in OnSIDES (95% CI =2.8–4.9, $P < 4.1 \times 10^{-22}$)’

19. Line 211 Reference is missing

Authors reply: We thank the reviewer for pointing this out and have included the reference.

Manuscript changes: We have included the following reference to the Results
'We used LoGoFunc[17] and estimates of effect from quantitative trait loci (QTL) to infer the direction of the associated genetic effect as described previously[18].'

20. Extended Data Fig.4 Caption is not finished, ore (include ref?); at which level of significance?)

Authors reply: We apologize for the typo. We have updated the caption of Extended Data Fig.4 (now Supplementary Fig. 6) to include the severity score reference as well as include significant *P*-value > 0.05.

Manuscript changes: We have added the following text to the figure caption for Supplementary Fig. 6 (previously Extended Data Fig.4)

'… The side effect outcome was weighted by severity using a crowdsourced severity score[2]… Each cross-validated sample is color labeled and filled circles indicate a beta coefficient with a significant *P*-value > 0.05 and the 95% CIs are defined as error bars.'

21. Line 228-229 and line 234-236: Unclear statements, please revise.

Authors reply: We have clarified this text.

Manuscript changes: We have revised the following text in the results section:

'We next considered whether applying threshold cutoffs for SE-GPS-DOE, was associated with a greater side effect risk'.
'Due to the fewer observations and the fact that LOF and GOF directional predictions are based on inference, we suggest using the SE-GPS-DOE as a complementary score to the SE-GPS'.

22.Data and app should also be provided using a zenodo link or similar.

Authors reply: We have included the Open Targets dataset and OnSIDES dataset as well as the code to re-create the SE-GPS and SE-GPS-DOE and all analyses here: https://zenodo.org/records/15334136?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImY1ZWY3ZmVmLWM0YmEtNGM1YS1iZjkxLTgxYWJlMGM2OGM1MCIsImRhdGEiOnt9LCJyYW5kb20iOiJiZTI1ZTJjMjRkNGMxOWM5MmQ2OWY3ZjhmN2UwMjFiYSJ9.dGGnnoLIUVaXQ4VKndr8c2vayJRAC05b0caR7gakbUrncTP9fg1hnSVzP132KFKErSKBuRKhc32ONJJT8Wfbrw and https://github.com/rondolab/SE-GPS. We have included the link to the shiny website (https://rstudio-connect.hpc.mssm.edu/sideeffect-geneticpriorityscore/) on the zenodo page as well.

Manuscript changes: We have included the following sentence under code availability

'Analytic code to create the SE-GPS and SE-GPS-DOE is available at https://zenodo.org/records/15334136?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImY1ZWY3Zm

VmLWM0YmEtNGM1YS1iZjkxLTgxYWJlMGM2OGM1MCIsImRhdGEiOnt9LCJyYW5kb20iOiJiZTI1ZTJjMjRkNGMxOWM5MmQ2OWY3ZjhmN2UwMjFiYSJ9.dGGnnoLIUVaXQ4VKndr8c2vayJRAC05b0caR7gakbUrncTP9fg1hnSVzP132KFKErSKBuRKhc32ONJJT8Wfbrw and https://github.com/rondolab/SE-GPS.'

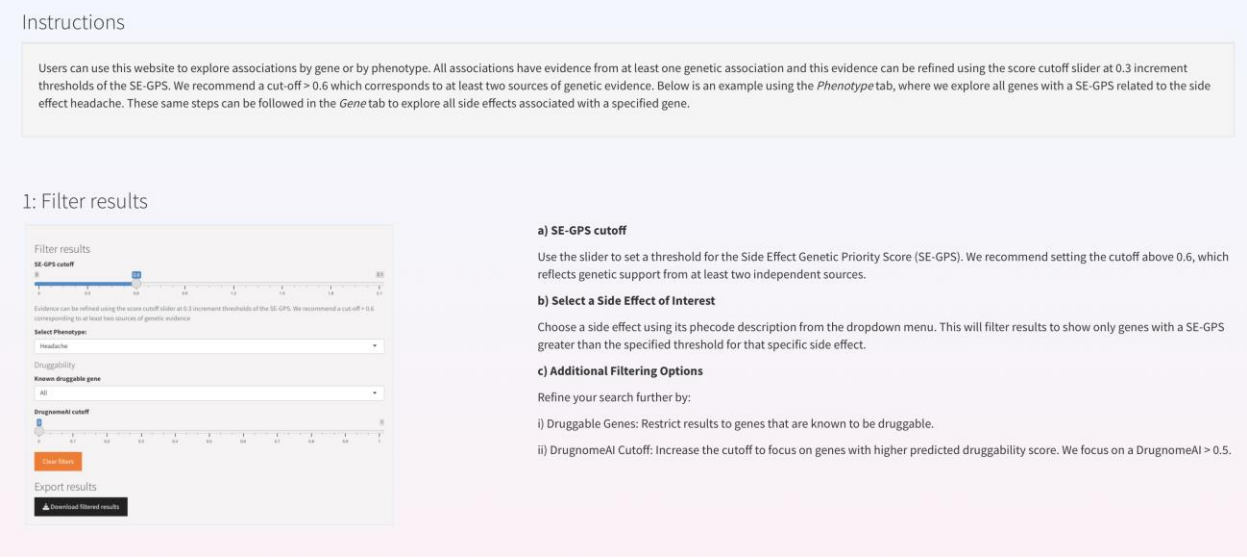
23. In the statistical analysis section, the p-value threshold and the method used for p-value adjustment are not specified.

Authors reply: We used $p < 0.05$ as the threshold for statistical significance. No multiple testing correction was applied. We have updated the figure captions to reflect this.

24. Comments specific to the Shiny App:
• Enable column search
• No gene data available (seems to be fixed)
• Gene select input with live search
• select input with live search
• More helping text or a video tutorial should be provided. i.e. helping text for the table columns, particularly where abbreviations are used. Which cut off should be used? Provide some helping text, ect.

Authors reply: We thank the reviewer for these suggestions. We have updated the Shiny app to enable column search and live selection of the gene target. Furthermore, we have included text under the SE-GPS cutoff to indicate a suggested cutoff and have expanded the instructions text on the about tab. Finally, we include an Instructions tab to walk through the steps to using this resource with headache as an example. Below is a screenshot of the instructions page, which can be further scrolled down to show the results and evidence.

Website changes: We have included an additional tab Instructions.



Reviewer #2 (Remarks to the Author):

The authors have embarked on an ambitious project collating many data sets to address an important question about prediction of adverse reactions to new drugs. I think I like the concept of the paper, but I am left very unclear how I would use. My comments are below, sometimes I step through to just make a logically flow of thinking. As I wrote my comments my understanding increased, hence it is possible that some of my comments I have self-answered. In general, I think I had to work too hard to understand what is being done.

Authors reply: We thank the Reviewer for their positive comments and greatly appreciate their suggestions and feedback. We have addressed their comments below.

1. Goal
a) First, I am left feeling about how to use the results of the study. The goal of the study is to "help inform the likelihood of an on-target side effect" (Discussion second sentence).
So, for a general audience, I think this could be explained with a couple more sentences (i.e., on-target vs off-target - obvious to the authors I realise).

Authors reply: We agree with the reviewer. We have clarified that most later-stage side effects are linked to the drug's action at the primary biological target and also mention that on-target side effects is the focus of this study.

Manuscript changes: We have revised the following text in the Introduction

'A considerable proportion of these later-stage side effects are linked to the drug's action at the primary biological target ('on-target') rather than secondary targets ('off-target'), underscoring the inherent challenges in drug discovery, including the limited time frame and sample size of clinical trials and poor translation from animal to human studies.'

b) I find it ambiguous exactly how one show use the results and the online tool. Can this be stepped through. Is it that when a new drug is being developed that one should identify the gene target of the drug and look this up in the database?

Authors reply: Our intention is for the tool to be used as source of evidence during the target discovery stage to help minimize the occurrence of later stage on-target side effects. By applying this framework to all 19,422 gene targets, we provide evidence for both known drug targets and targets that have not yet been drugged. We have further annotated which targets are currently druggable and have incorporated a druggability probability score to identify targets with a higher druggability likelihood that currently lack clinical trial evidence. This enables the user to filter to targets where the SE-GPS framework is arguably more relevant. Thus, as the author indicates, when a new drug is being developed this framework can be used to provide a overview of possible on-target side effects that could result.

To better illustrate the use of this framework, we have revised the manuscript to include the addition of the druggability score, DrugnomeAI > 0.5, and have included examples of current undrugged targets with high predicted druggability (DrugnomeAI > 0.5), to showcase how the SE-GPS can be applied at target discovery to help identify potential on-target side effects for targets with no prior clinical trial evidence **(Table 3).**

Lastly, we have also included additional examples of severe side effects highlighted by Carss et al[20]., which were accurately captured by our score **(Table 2).**

We note that this tool should only be used as a starting point however and recommend the addition of complementary genetic methods to further capture associations not captured by the SE-GPS.

Manuscript changes: We have included the following text in the Results section.

Second, we evaluated the performance of the SE-GPS using two examples of well-known targets discussed by Carss et al., where genetic evidence has previously provided strong support for the observation of severe side effects for drugs that led to clinical trial failure **(Table 2).** The SE-GPS provides strong support for gastrointestinal side effects following inhibition of *DGAT1* and neurological disorders from inhibition of *SPR*. Despite the side effects associated with *DGAT1*, it remains an attractive target for many autoimmune, metabolic and oncological diseases. Therefore, recognizing the possible gastrointestinal side effects of *DGAT1* can enable appropriate monitoring, risk assessment and the development of more selective inhibitors. Third, we highlight examples of current undrugged targets with high predicted druggability (DrugnomeAI > 0.5), showcasing how the SE-GPS can be applied at target discovery to help identify potential on-target side effects for targets with no prior clinical trial evidence **(Table 3).**

Manuscript changes: We have included the following Tables.

Table 2. Examples of failed clinical trial targets supported by the SE-GPS and SE-GPS-DOE.

Prioritized examples	Genetic Evidence	Direction of effect evidence
Gene: *DGAT1* Phenotype (code): Symptoms involving digestive system (529) Side effect: Gastrointestinal side effects, including diarrhoea[21] SE-GPS:1.54 SE-GPS-DOE:1.56 Mechanism: Inhibitor	ClinVar: Congenital diarrhea 7 with exudative enteropathy HGMD: failure to thrive recurrent fractures nephrocalcinosis and chronic diarrhoea; protein-losing enteropathy early-onset; congenital diarrhoeal disorder; chronic diarrhoea delayed-onset OMIM: diarrhea 7, protein-losing enteropathy type, 615863 (3)	ClinVar: LOF HGMD: LOF OMIM: LOF
Gene: *SPR* Phenotype (code): Extrapyramidal and movement disorders (324) Side effect: Neurological effects[22] SE-GPS:1.03 SE-GPS-DOE:1.04 Mechanism: Inhibitor	HGMD: intellectual disability; tetrahydrobiopterin deficiency; dystonia OMIM: dystonia, dopa-responsive, due to sepiapterin reductase deficiency, 612716 (3)	HGMD: LOF; OMIM: LOF

SE-GPS, side effect genetic priority score; SE-GPS DOE, side effect genetic priority score with direction of effect, OMIM, Online Mendelian Inheritance in Man; HGMD, Human Gene Mutation Database

Table 3: Examples of possible side effects for undrugged targets using the SE-GPS and SE-GPS-DOE.

Prioritized examples	Genetic Evidence	Direction of effect evidence
Gene: *GJB2* Phenotype (code): Hearing impairment (396) SE-GPS:2.34 SE-GPS-DOE:1.89 Suggested Indication: GJB2-targeted cancer immunotherapy[23] Mechanism: Inhibitor	ClinVar: Autosomal dominant nonsyndromic hearing loss 3A; Hearing loss, autosomal recessive; Hearing impairment HGMD: deafness nonsyndromic sensorineural; sensorineural hearing loss; deafness autosomal recessive 1; deafness nonsyndromic; deafness; hearing loss non-syndromic; hearing impairment nonsyndromic; hearing impairment; deafness autosomal dominant 3; hearing loss; ichthyosis follicularis sensorineural hearing loss and punctate palmoplantar keratoderma; vohwinkel syndrome; hearing loss non-syndromic autosomal recessive; ichthyosiform erythroderma corneal involvement & deafness; deafness and palmoplantar hyperkeratosis; deafness and palmoplantar keratoderma; knuckle pads leukonychia sensorineural deafness; knuckle pads hyperkeratosis and deafness; sensorineural hearing loss & leukonychia; keratitis-ichthyosis-deafness syndrome; hearing impairment postlingual; hearing loss non-syndromic autosomal dominant; hearing impairment bilateral sensorineural OMIM: bart-pumphrey syndrome, 149200 (3); deafness, autosomal dominant 3a, 601544 (3); deafness, autosomal recessive 1a, 220290 (3); hystrix-like ichthyosis with deafness, 602540 (3); keratitis-ichthyosis-deafness syndrome, 148210 (3); keratoderma, palmoplantar, with deafness, 148350 (3); vohwinkel syndrome, 124500 (3) RaVAR GB: hearing loss; sensorineural hearing loss L2G: hearing loss; age-related hearing impairment; sensorineural hearing loss	ClinVar: LOF HGMD: LOF OMIM: LOF RaVAR Gene Burden: LOF
Gene: *SLC13A5* Phenotype (code): Epilepsy, recurrent seizures, convulsions (330) SE-GPS:1.54 SE-GPS-DOE:1.60 Suggested Indication: Kidney disease[24] Mechanism: Inhibitor	ClinVar: Developmental and epileptic encephalopathy, 25 HGMD: epileptic encephalopathy early infantile; epileptic encephalopathy; west syndrome & severe psychomotor development retardation; developmental and epileptic encephalopathy; epileptic encephalopathy early-onset; epilepsy early-onset; encephalopathy; epilepsy; kohlschütter-tönz syndrome; global developmental delay epilepsy chronic kidney disease; paediatric movement disorder OMIM: developmental and epileptic encephalopathy 25, with amelogenesis imperfecta, 615905 (3)	ClinVar: LOF HGMD: LOF OMIM: LOF

Gene: *GJA1* Phenotype (code): Abnormal intraocular pressure (375) SE-GPS:1.83 SE-GPS-DOE:1.36 Suggested Indication: Alzheimer's[25] Mechanism: Inhibitor	HGMD: microcornea and glaucoma; open angle glaucoma and microcornea OMIM: oculodentodigital dysplasia, 164200 (3) RaVAR GB: glaucoma L2G: open-angle glaucoma	HGMD: LOF OMIM: LOF RaVAR Gene Burden: LOF
Gene: *ORAI1* Phenotype (code): Immunodeficiencies (179) SE-GPS:1.03 SE-GPS-DOE:1.04 Suggested Indication: Duchenne muscular dystrophy[26] Mechanism: Inhibitor	HGMD: immunodeficiency combined; severe combined immune deficiency syndrome; immunodeficiency muscular hypotonia & anhidrotic ectodermal dysplasia; severe combined immune deficiency syndrome and residual t-cell function OMIM: immunodeficiency 9, 612782 (3)	HGMD: LOF OMIM: LOF

SE-GPS, side effect genetic priority score; SE-GPS DOE, side effect genetic priority score with direction of effect, OMIM, Online Mendelian Inheritance in Man; HGMD, Human Gene Mutation Database

c) My understanding is that only a small proportion of genes are actually druggable.
Of the genes that have results (I am unclear how many genes have results) what proportion are considered druggable.

Authors reply: The reviewer is correct that only a small proportion of genes are actually druggable. We have therefore now identified the proportion of genes with genetic evidence that are druggable (3,818 out of 15,139) and have compared the SE-GPS within this subset of druggable genes to non-druggable genes (Mann-Whitney test, $P < 5.9 \times 10^{-214}$).

We apply this method to all 19,422 genes, enabling this tool to be used for targets that are currently undrugged but with advancements in gene editing and RNA therapeutics, may be drugged in the future. However, despite technological advancements, not all of these undrugged targets are likely to elicit a therapeutic effect. Therefore, we have incorporated the druggability prediction score, DrugnomeAI [27], to identify targets with a higher likelihood of being drugged (**Table 3** above), and thus more relevant for predicting side effects.
We have included both of these druggable annotations to our website.

Manuscript changes: We have included the following text to the Results.
Furthermore, given that only a small fraction of protein-coding genes are currently considered druggable, we assessed the proportion of targets with genetic evidence classified as druggable genes. Out of 15,139 genes, 3,818 genes were identified as druggable with significantly higher SE-GPS compared to non-druggable genes (Mann-Whitney test, $P < 5.9 \times 10^{-214}$).

d) I am unclear what obvious (i.e. known) results are missed.
Can some examples be given.

Authors reply: We agree with the Reviewer that it's possible that some known results could be missed by our SE-GPS tool. Due to our use of phecodeX terminology and collapsing phecodes by integer terms, it is possible that some genetic evidence phenotypes and side effect terms did not always map correctly, potentially leading to weaker or absent support. Furthermore, when providing evidence for side effect risk, these terms are more general and thus require further refinement. One example of a known side effect missed by our score includes the example from Walker et al[13]., as highlighted by the reviewer below: migraine risk associated with CASR. This example highlights the importance of considering multiple data sources and using different statistical methods to capture as much of the genetic landscape as possible. We have added this in our discussion.

Manuscript changes: We have added the following sentences to our discussion

'Additional methods include Mendelian Randomization, which offers several advantages, particularly the ability to infer causality rather than associations'

e) Mendelian Randomisation has been proposed to identify on-target adverse effects. See doi: 10.1093/ije/dyx207 "Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities" which has been cited 174 times
Can you compare your results to theirs (or papers that have used this approach) either directly, or at least in discussion to explain how approaches differ.

Authors reply: We thank the reviewer for highlighting this study and, as mentioned above, have added this work to our discussion. In this study, the authors demonstrate how Mendelian Randomization can be used to identify unintended drug effects and repurposing opportunities. They highlight two examples: the association of increased diabetes risk (higher body weight, waist circumference, plasma insulin concentration and plasma glucose concentration) with HMGCR inhibition and the potential use of CASR inhibitors to treat migraine risk using GWAS.

In our results, we report a SE-GPS of 0.41 for diabetic risk from HMGCR, which falls just below our proposed cut-off of 0.6. This cut-off is based on incorporating at least two lines of genetic evidence, and thus as this association was captured solely through GWAS data, it emphasizes a limitation of overlooking single lines of evidence. We have added this to our discussion, highlighting that our suggested cut-off should only be used as a starting point. For the other result, we were not able to confirm migraine risk with CASR using the SE-GPS due to missing genetic evidence.

Manuscript changes: We have added the following sentences to our discussion

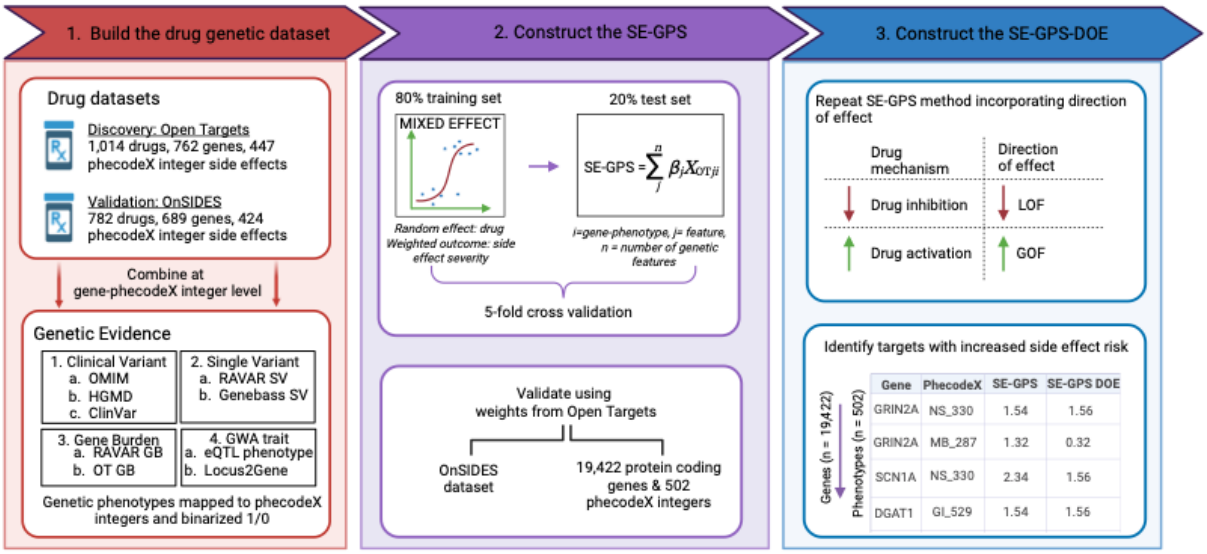
'The overlap between targets with predicted side effect risk and known drug targets for similar drug indications emphasizes the importance of integrating all aspects of genetic evidence and disease biology when selecting a potential drug target to ensure that it is both effective and safe. Although we prioritize associations supported by multiple lines of genetic evidence by suggesting a cut-off of 0.6, this approach may overlook signals captured by a single line of evidence. For example, prior work by Walker et al. highlighted the association between *HMGCR* inhibition and increased diabetes risk, for which we observed a SE-GPS of 0.41, captured exclusively through GWAS evidence. Thus, we provide this cut-off and framework as a starting point and recommend the addition of complementary genetic methods to further strengthen this evidence and capture genetic associations not currently included in the SE-GPS.

One such example is the incorporation of somatic variant data from tumor tissues, included by Minikel et al., who similarly demonstrated that side effects with human genetic support are 2.0 times more likely to occur. Additional methods includes Mendelian Randomization, which offers several advantages, particularly the ability to infer causality rather than associations and polygenic risk scores, which offer the opportunity to stratify patients in clinical trials according to disease risk'

2. Methods
I have spent a very long time on this. I added line numbers to the word file provided.
a. Figure 1 should include more steps to describe the process. Include the structure of the files generated at each stage. Currently it is not a very helpful summary. In Step 3 the description is "v applying this method to 19,422 genes and 470 phenotypes to identify targets" please state the total number of gene-phenotype rows and give an example where a gene has 2 phenotype codes, and the same phenotype has 2 genes.

Authors reply: We thank the reviewer for carefully going over this section. We agree that Figure 1 was not descriptive enough. We have expanded this figure to provide a more comprehensive summary of the data and steps to create the SE-GPS. We expand on the nine data sources used and also included the construction of the SE-GPS-DOE in this schematic. In the examples listed in the table, we show two examples of phenotypes associated with *GRIN2A* and two genes associated with the phenotype (NS_330). In the figure caption, we clarify that we construct the score for 19,422 genes and 502 phecodeX integers, for which 15,139 genes linked to 499 phenotypes had support from at least one genetic feature and directional evidence (n=146,011).

Manuscript changes: We have updated Figure 1 and the corresponding figure caption.



A workflow of the data sources and steps to construct the SE-GPS and SE-GPS-DOE as outlined in this analysis. The SE-GPS and SE-GPS-DOE were created in the Open target dataset (discovery), validated in OnSIDES and then generated for 19,422 genes and 502 phecodeX integers, for which 15,139 genes linked to 499 phenotypes had support from at least one genetic feature and directional evidence

(n=146,011). SE, side effect, OT, Open Targets; SE-GPS, side effect genetic priority score; SE-GPS-DOE, side effect genetic priority score with direction of effect. Created with BioRender.com.

b. The description in the main text is difficult to map to the more detailed methods. The numbers at lines 100-102 (987 drugs, 733 genes, 348 unique drug indications and 417 unique side effects, whereas the OnSIDES dataset consisted of 806 drugs, 697 genes, 349 unique drug indications and 396 unique side effects) are not mentioned in the Methods. Since at line 324 it says "1,037 drugs, 748 genes, 321 unique drug indications and 385 unique side effects mapped to phecode integers." For OpenTargets and for Onsides line 334 does not match. Presumably because of another QC step, but its adds to making it a very confusing read.

Authors reply: We thank the reviewer for highlighting this issue and agree that the mismatch in numbers is confusing. We have updated these numbers in both the results and methods section which are now consistent.

Manuscript changes: We have revised the following sentence in the Results section.

Following quality control, the Open Target dataset comprised 1,014 drugs, 762 genes, 362 unique drug indications and 447 unique side effects, whereas the OnSIDES dataset consisted of 782 drugs, 689 genes, 366 unique drug indications and 424 unique side effects.

c. At line 105. State that both the drug indications and the side effects were mapped to phecodes to make a set of gene-phenotypes.

Authors reply: We thank the reviewer for this comment and have clarified this in the text.

Manuscript changes: We added the following sentence to the Results

'To construct the drug datasets, we mapped the side effects and drug indication data to phecodeX integer terms across 16 phecode categories, similar to the GPS, and outlined additional quality control steps in the **Methods.'**

d. Line 411 I think you need to add a section "Combined gene data set" which states (to my understanding) that the previously listed genetic databases are combined to give a matrix of rows gene-phenotype 19,422 protein coding genes that have at least 1 yes/no entry for clinical variants. Say how many rows in total.

Authors reply: We thank the reviewer for their suggestion. We expanded the 'Generation of the integrated drug–genetic dataset' section to include the total number of rows for Open Targets and Onsides and then created an additional section 'Generation of the integrated gene–phenotype dataset across 19,422 protein-coding genes and 470 phecode pairs' to describe the creation of the matrix for the 19,422 genes.

Manuscript changes: We have included the following sentences and section to the Methods.

'Each drug-gene (n = 2,848) pair is repeated for 447 side effect phecode integers giving a total of 1,273,056 rows. We formatted the OnSIDES validation dataset similarly where each drug-gene (n = 2,740) pair is repeated for 424 side effect phecode integers giving a total of n= 1,161,760 rows.'

'Generation of the integrated gene–phecodeX integer dataset across 19,422 protein-coding genes and 502 phecodeX integer pairs.

Similar to generating the drug-genetic datasets, we integrated the nine data sources described above for all 19,422 protein-coding genes for 502 unique phecodes. This resulted in a matrix of 9,749,844 gene-phecode pairs, for which 17,214 genes and 502 phecodes had support from at least one genetic feature. We integrated the DrugnomeAI probability score[27] and druggable genes were defined using the following sources: drugbank[28], chembl[29] and two published supplementary tables which list druggable genes[30,31]'

e. Line 106" We integrated both side effect datasets with human genetic evidence at the gene-phenotype level using nine data sources". "Integrated" is ambiguous. Give more detail so it is clear what the file is that is used for analysis.

Authors reply: We have clarified the text to describe how we combined the drug data with the genetic evidence to create our overall datasets for analysis.

Manuscript Changes: We have added the following text to the Results section:

'Using gene-phecode pairs as the common identifier, we combined both side effect datasets with nine human genetic data sources at the gene-phecode level, consolidating these into four genetic features to use for analysis…'

f. Line 113. Suggest 'A detailed description of each data source is provided in the Methods and an overview of these gene-phenotype observations across 19,422 genes and 470 phenotypes is shown in Supplementary Fig. 2 ' is updated to 'A detailed description of each data source is provided in the Methods leading to an analyzable data matrix where the entries are zero or one of x gene-phenotype rows comprising 19,422 protein coding genes, 470 phenotypes, z phecodes and x columns. The columns comprise…. An overview of gene-phenotype observations is shown in Supplementary Fig. 2 '.

Authors reply: We thank the reviewer for their suggestion and have incorporated the proposed text with the following adjustments.

Manuscript changes: We have included the following sentences in the Results section.

'A detailed description of each data source is provided in the **Methods**, resulting in an analyzable data matrix in which the Clinical Variant predictor is encoded as 0, 1, 2 or 3 and all other predictors are binary (0 or 1), across 9,749,844 gene–phecode pairs comprising 19,422 protein-coding genes, 502 phecode terms and 16 phecode categories**.** An overview of these gene-phecode observations across is shown in **Supplementary Fig. 2'**

g. Line 429 "Generation of the integrated drug–genetic dataset" section. As in comments above, explicitly state the inputs and output of this step

Authors reply: We have included the total number of drug-gene pairs, number of phecodes and total rows for the Open Target and Onside datasets.

Manuscript changes: We have included the following sentences in the Methods section.

'Each drug-gene (n = 2,848) pair is repeated for 447 side effect phecode integers giving a total of 1,273,056 rows. We formatted the OnSIDES validation dataset similarly where each drug-gene (n = 2,740) pair is repeated for 424 side effect phecode integers giving a total of n= 1,161,760 rows'

h. It is not clear why Supp Fig 2 summarises as phenotypes rather than phecodes. From Supp Fig 2 it is clear most phenotypes are unique, whereas phecodes presumably combine similar phenotypes. Should an additional supp Figure be made that gives the same info for phecodes.

Authors reply: We apologize for the confusion. We had used the term phenotype to reflect phecode integer codes and thus Supplementary Fig. 2 is at the phecode level. We have changed the use of phenotype to phecode to address this confusion.

i. Line 117. I found I had to work too hard to understand this section "Association of genetic features with drug side effects" please step the reader through more.
So the y-variable is yes/no for side effects for a gene (385 separate analyses or combined into one). How many genes in the analysis? By the time I get to line 238 I realise not all genes. An equation would help avoid ambiguity as then the authors would define the regression more specifically. The equation includes 15 phecodes as covariates. It is not explained what these 15 phecodes are – are they phecode categories, nor why they are selected, nor why the analysis is biased if these are not included. Please provide more information.

Authors reply: We apologize for the lack of clarity in this section. In the univariate regression analysis, we assessed the association between each genetic feature and drug side effects across the entire dataset in the one model. For Open Targets, *n* = 1,273,056 independent drug–gene–phenotype combinations and for OnSIDES *n* = 1,161,760 independent drug– gene–phenotype combinations. In the univariate model, the y variable represents drug side effects, binarized as 1 or 0 (reflecting presence or absence). This was tested against each genetic feature, with the 16 phecodeX categories included as covariates in each model. To provide further clarification, in Figure 2, we have updated the y-axis to display the proportion of unique genes with genetic evidence and an observed side effect over the total number of unique genes with genetic evidence. Further, we report the proportion of unique phecodeX integers with genetic evidence and an observed side effect over the total number of unique phecodeX integers with genetic evidence. We have included this equation in the methods.

$$P(SE) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 Feature + \beta_1 Category + \sum_{j=1}^{k} \gamma_j . Category_j)}}$$

P(SE) represents the probability of the side effect outcome, β_1Feature is the effect of the genetic feature of interest and Category$_j$ represent the k disease categories included as covariates to account for confounding across side effect classes.

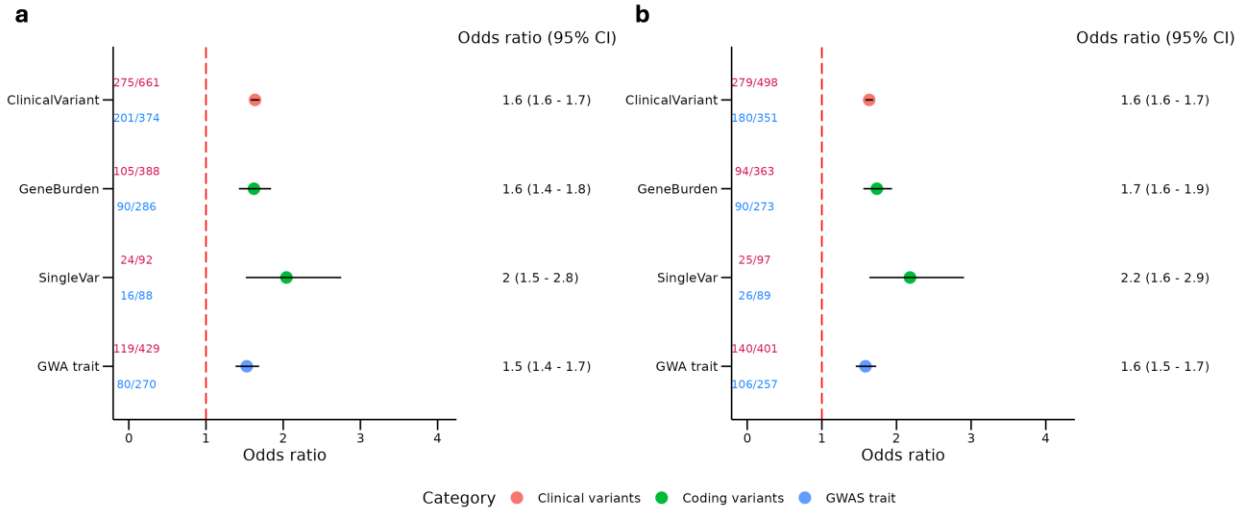
The PhecodeX categories, as defined by Shuey et al.[32], group the phecode terms into broader categories similar to ICD chapters. Of the original 18 categories, we excluded neonatal, and pregnancy-related

categories, leaving 16 phecode categories. These were included as covariates to account for potential confounding effects, ensuring that the observed associations are not driven by differences between the disease categories. We've included Supplementary Fig. 3 to show the breakdown of the number of unique drug-side effect pairs across each category, noting significant differences between the number of unique drug side effects between categories. Furthermore, we've stratified the datasets by phecode category and again perform a univariate analysis of each feature with drug side effect. We observe significant variability within each disease category, in particular for the SingleVar feature, which has the smaller number of total observations (Supplementary Fig. 4 and Supplementary Fig. 5).

Manuscript changes: We have included the following sentences in the Results section:

'We performed univariate associations assessing the enrichment of the four genetic features with the drug side effects outcome in the Open Target and OnSIDES datasets. Given the variation in the number of unique drug-side effect pairs across the PhecodeX categories (Supplementary Fig. 3), and to account for disease heterogeneity, we adjusted for the 16 phecodeX categories as covariates. We observed significant associations of each feature in both datasets (Fig. 2). Furthermore, we examined the association between each genetic feature and drug side effects within each disease category, which revealed variability in the strength of enrichments across categories and between genetic features (Supplementary Fig. 4; Supplementary Fig. 5). Notably, the single variant feature had a lower number of observations overall, and thus when stratified by category, this resulted in much wider confidence intervals.'

Furthermore, we have updated Fig. 2 and included Supplementary Fig.3, Supplementary Fig. 4 and Supplementary Fig .5.



We have included the following sentences in the Methods section:
We tested the association of each genetic feature, the SE-GPS and the SE-GPS-DOE with drug side effects in a univariate logistic regression model with drug side effects as the outcome using the glm function, adjusting for the 16 phecode categories as covariates. This equation is as follows:

$$P(SE) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 Feature + \beta_1 Category + \sum_{j=1}^{k} \gamma_j . Category_j)}}$$

where *P(SE)* represents the probability of the side effect outcome, β1Feature is the effect of the genetic feature of interest and Categoryj represent the k disease categories included as covariates to account for confounding across side effect classes.

j. Line 129. In the section "Construction of the SE-GPS". I have read this many times and cannot understand what has been done. What is the y-variable in the analysis. OK I see in extended data Figure 1, y= side effect, and in page 20 methods. I guess this is implied in statement "mixed-effect regression model of the five genetic features with drug side effects", but I hope you can see that "with drug side effects" is ambiguous to the reader.

Authors reply: We appreciate the reviewer's feedback and acknowledge that our description of the analysis could be clearer. We have now explicitly stated that the y variable in our mixed effect model is drug side effects. We have revised the text to make this clear.

Manuscript changes: We have updated the following text in the Results section
'We next constructed the SE-GPS based on the cumulative effects of the four genetic features. Specifically, we used 80% of the Open Target dataset as the training set and applied a multivariable mixed-effect regression model of the association of the four genetic features with drug side effects as the outcome to obtain
the effect sizes from the association of each feature to use as weights in the score'

k. Line 130-131. The first 2 lines, implies information from the total data section (c. above analysis) is brought into the cross-validation analysis. This feels uncomfortable for the potential for data leakage/overfitting. The whole pipeline of analyses should be conducted in the discovery 80% and applied to the left-out analysis. This is important, and may lead to reduced consistency in Extended Data figure 1.

Authors reply: We thank the reviewer for highlighting this point. Our description in the initial version had been unclear, leading to confusion. To clarify, the mixed-effect regression used to estimate the beta coefficients for the four genetic features was conducted within the 80% training set. These estimated coefficients were then used to compute the SE-GPS score in the remaining 20% test set, which was held out entirely from model training. This process was repeated in a five-fold cross-validation framework, ensuring that each 20% test set was evaluated independently and had no overlap with the corresponding 80% training set. As the reviewer notes, this design prevents information leakage and overfitting.

l. Line 140-141 "We used the cross-validated test with the highest OR and extracted the coefficients from this mixed-effect model to calculate the SE-GPS in the OnSIDES dataset (Extended Data Table 1)." Extended Data Table 1, is for Open Targets relating to Extended data Figure 1. The OnSides Table is Extended Data Table 2. In addition provide a Figure like Extended Data Figure 1 for OnSIDES.

Authors reply: We apologize for the confusion. We realize that an additional table displaying the effect sizes corresponding to Extended data Figure 1 should have been included and have now added this table as **Table S1.**

Additionally, we have clarified how the highest OR was determined from the cross-validated tests. Specifically, to identify the fold from which we obtained effect sizes to apply as weights to OnSIDES, we performed a logistic regression model with drug side effects as the outcome, the SE-GPS as the predictor, and the 16 phecodeX categories included as covariates. The OR from this model was previously **Extended Data Table 1**, and now **Table S2.**
Extended Data Figure 1 (now Supplementary Fig. 6) reflects the effect sizes from the mixed effect regression model and thus does not apply to OnSIDES.

Manuscript changes: We have included the following text to the results section:

'We applied a mixed-effect regression model (using the lme4 R package, version 1.1-35.1) within a fivefold cross-validation framework, and extracted the association coefficients as weights for each genetic feature contributing to the score as detailed in equation (1)**(Table S1)**.'

'Within each cross-validated test set, we assessed the association between the SE-GPS and drug side effects using a logistic regression model, with drug side effect as the outcome, the SE-GPS as the predictor, and the 16 phecodeX categories included as covariates **(Table S2)**. We used cross-validated test 1, which had the highest OR and applied the coefficients from this mixed-effect model to further validate the SE-GPS in the OnSIDES dataset'

We have included Table S1.

Table S1: Beta estimates from mixed effect model of five genetic features on drug side effect in Open Targets five cross validated training sets

CV	Predictor	beta	95% CI	P-value
CVsample1	ClinicalVariant	0.51	0.51 - 0.52	0.00E+00
CVsample1	GWA trait	0.41	0.4 - 0.43	0.00E+00
CVsample1	GeneBurden	0.39	0.37 - 0.41	4.56E-297
CVsample1	SingleVar	0.57	0.53 - 0.62	9.59E-144
CVsample2	ClinicalVariant	0.56	0.55 - 0.56	0.00E+00
CVsample2	GWA trait	0.37	0.36 - 0.38	0.00E+00
CVsample2	GeneBurden	0.36	0.34 - 0.38	1.41E-268
CVsample2	SingleVar	0.45	0.41 - 0.5	6.89E-79
CVsample3	ClinicalVariant	0.54	0.53 - 0.54	0.00E+00
CVsample3	GWA trait	0.41	0.39 - 0.42	0.00E+00
CVsample3	GeneBurden	0.26	0.24 - 0.29	1.62E-133
CVsample3	SingleVar	0.3	0.25 - 0.35	1.03E-36
CVsample4	ClinicalVariant	0.54	0.54 - 0.55	0.00E+00
CVsample4	GWA trait	0.34	0.33 - 0.35	0.00E+00
CVsample4	GeneBurden	0.31	0.29 - 0.33	3.02E-192
CVsample4	SingleVar	0.44	0.39 - 0.48	1.56E-84
CVsample5	ClinicalVariant	0.56	0.55 - 0.57	0.00E+00
CVsample5	GWA trait	0.44	0.42 - 0.45	0.00E+00
CVsample5	GeneBurden	0.26	0.24 - 0.28	4.86E-139

CVsample5	SingleVar	0.59	0.55 - 0.64	2.94E-143

Abbreviations: CV, cross-validated; CI, confidence interval

m. Page 20 SE-GPS_Oti calculation. The first section (ie. up to ". For each of the five folds ..") should include an equation, with each term defined which will be easier to follow than a description in words alone. This section has the y variable as drug-side effect, so the regression estimates are log(OR) of genetic feature to drug side effect.

Authors reply: We have included the mixed effect equation as equation (1).

Manuscript changes: We have included equation (1) in the Methods section

$$logit\big(P(SE)\big) = \beta_0 + \beta_1.CV + \beta_2.GB + \beta_3.GW + \beta_4.SV + \beta_5.Category + (1\,|Drug)$$

where *P(*SE) represents the probability of the outcome, and βi are the fixed effect coefficients for the covariates: Clinical Variant (CV), Gene Burden (GB), GWA trait (GW) and Single Variant (SV), and the 16 PhecodeX categories included as covariates. A random intercept was included for drugs.

n. The second section of SE-GPS, giving the equation for SE-GPS_Oti. This now makes a value for a gene-phenotype. My haziness maps back to part a above. Explicitly define for j=1, n. Provide the beta-weights in a table

Authors reply: In equation 2, j corresponds to each (j) genetic feature. We have now included the beta weights in **Table S1** for each of the five cross-validated folds.

Manuscript changes: We have included Table S1 as included above.

o. Page 21 "In addition, we also applied these weights to the 19,422 protein-coding genes and 470 phenotype pairs" the equation is for SE-GPS_Oti, where i=gene-phenotype, hence the wording 19,422 genes and 470 phenotype pairs is hard to understand, do you mean pairs made up of 19422 genes and 470 phenotypes ie 19422*470, or only a subset – I guess only a much smaller, subset, but when not stated explicitly it is hard for me to know if I am following or not…..In the legend of extended data Figure 2 "Open Target dataset for 1,150,086 SE-GPS across n = 9,188 gene-phenotype combinations and c) OnSIDES dataset for 1,104,048 SE-GPS across n = 7,713 gene-phenotype combinations" this is confusing as SE-GPS is defined for "I" = gene-phenotype, but looks like there is an SE-GPS for each feature. So the y-axis percentile is based on the SE-GPS for 9,188 gene-phenotype pairs, but then you take a subset of the j=1,n features to make the plots per row?

Authors reply: We appreciate the reviewers feedback and recognize that this explanation was unclear. As expanded in point l), we used the beta coefficients from the cross-validated test set that yielded the maximum odds ratio. Using these four beta coefficients, we constructed the SE-GPS for 19,422 and 502 phecodeX integer pairs as the sum of the weighted effect sizes across the four genetic features. Across this matrix, 17,214 genes and 502 phecodeX integers had support from at least one genetic feature, resulting in a genes-phecodeX matrix of 168,920 rows with a SE-GPS >0.

For Extended data Figure 2 (now Supplementary Fig. 7), we wanted to see the contributions of each individual feature that collectively form each SE-GPS in the Open Target dataset and in the OnSIDES dataset (now Supplementary Fig. 9). Lower scores on the left tail of the X axis are driven by the presence of a single feature whereas higher scores towards the right tail result from the cumulative effect of multiple contributing features. We have clarified this in the legend to make it clearer that the SE-GPS is for each gene-phenotype and not each feature.

Manuscript changes: We have made the following changes to the caption of Supplementary Fig. 7.
In panel a) violin plots show the distribution of each genetic feature that collectively sums to form the SE-GPS in the Open Target dataset for 1,273,056 SE-GPS across n = 34,0614 gene-phenotype combinations. The x-axis represents the percentile of the SE-GPS, starting at 97% to show non-zero scores only, while the y-axis separates the scores across each contributing genetic feature. The width of each violin plot represents the density of the genetic feature at each percentile, with the mean percentile marked as a circle. The total sample size of gene-phecodeX integer observations for each feature (n) and the mean weight from the five cross-validated samples is recorded under each feature on the y-axis, ordered by increasing value of these weights across the six features. The clinical variant feature was split according to the number of data sources (1, 2 or 3) for each gene phenotype observation. In panels b) bar plots show the contribution of the genetic features to the SE-GPS at 0.3 increment bins in the Open Target dataset. On the x-axis of each bar plot is the number of genetic features contributing to each score, colored by each feature present. The y-axis shows the count for each feature. In both plots, we demonstrate that as the SE-GPS increases, the number of features contributing to the score increases. SE-GPS, side-effect genetic priority score

p. I would like to see a plot of the distribution of SE-GPS_Oti, it must be very skewed. It could be added to Extended Data Fig 2. The a. and c x -axis starts at 98th percentile. It looks like only e-QTL contribute to percentiles less than 98th percentile in which case, it gives me an uneasy feeling about the statistic used. The eQTL only contribute when they have another feature. It is hard to interpret the per sd results at the top of page 7, without understanding of the meaning of one unit in distributional terms.

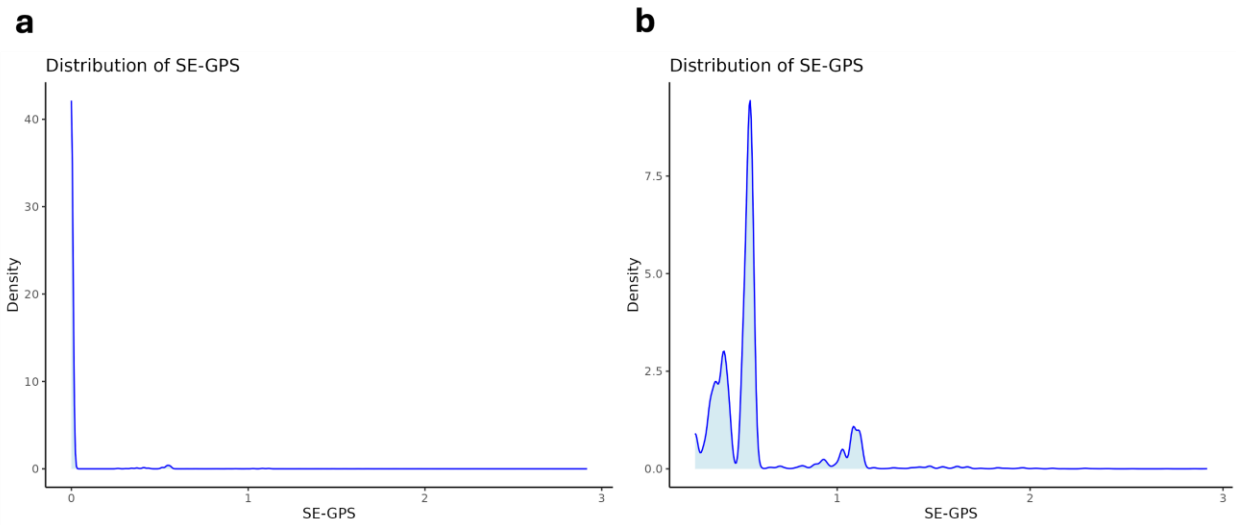
Authors reply: The reviewer is correct that the distribution of the SE-GPS is skewed, with 97% of gene-phenotype observations having a score equal to 0. This is expected and due to the nature of the SE-GPS approach where we only consider genetic evidence from various sources (such as genetic association with significant P-value threshold). We have included the distribution of all SE-GPS as well as the distribution when restricting to non-zero scores as Supplementary Fig. 8. We agree with the reviewer that without defining one unit the results are hard to interpret and thus have reframed our results accordingly for Figure 3 and Figure 6. Instead of analyzing the association of the continuous SE-GPS with drug side effects, we now present the association of any SE-GPS above zero (binarized as 1) to illustrate the overall relationship between genetic evidence and drug side effects more clearly.

Manuscript changes: We have added the following text to the results section:

'We note that across the five-training test splits, only ~ 3% of all gene-phecode pairs had a SE-GPS greater than zero (**Supplementary Fig. 7; Supplementary Fig. 8**)'

Manuscript changes: We have included the following Figure as Supplementary Fig. 8

Supplementary Fig. 8. Density distribution of the SE-GPS across the Open Target dataset

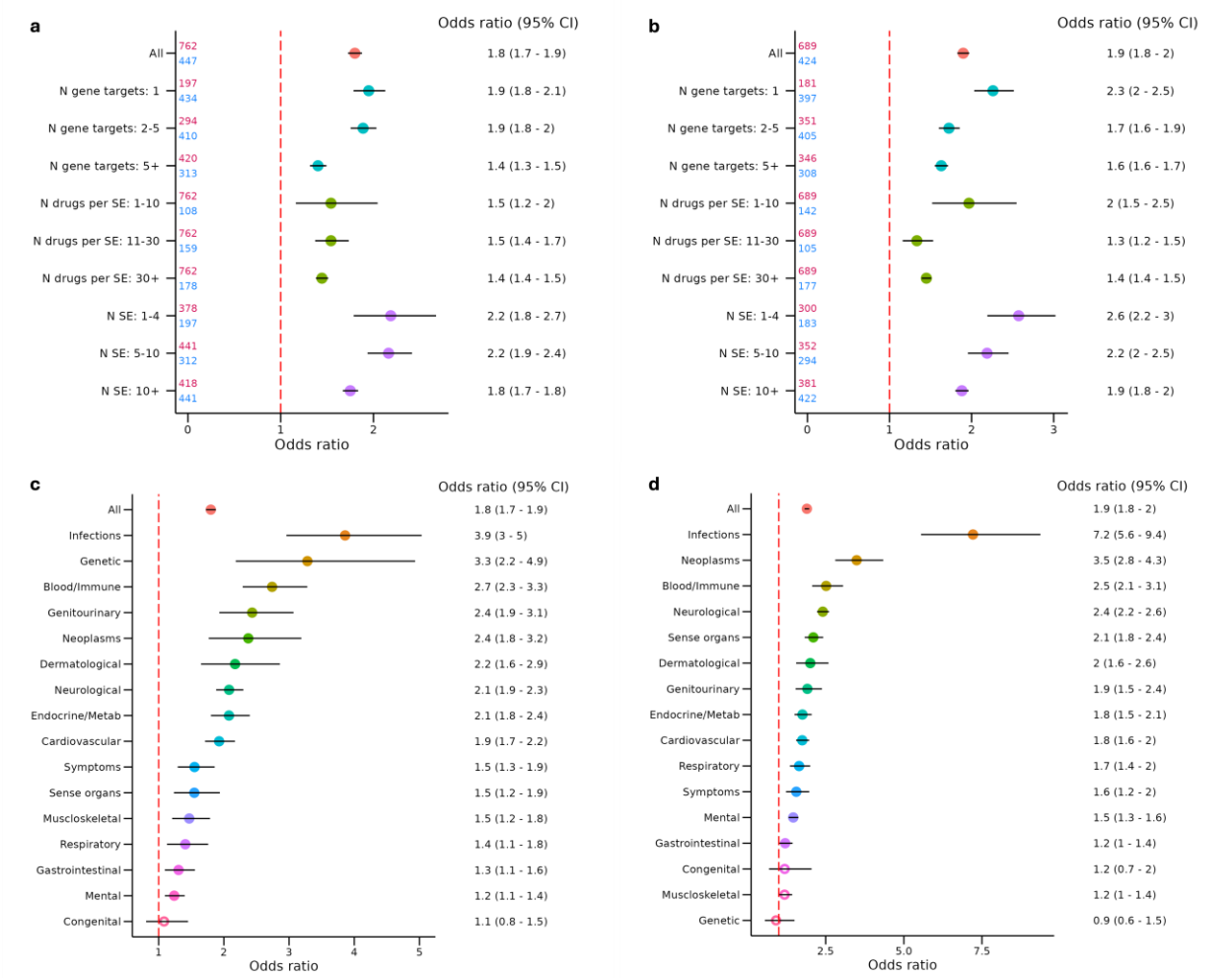


The left panel shows the full distribution of SE-GPS, while the right panel displays the distribution restricted to gene–phecode pairs with SE-GPS greater than zero.

Manuscript changes: We revised the following sentences in the Results.

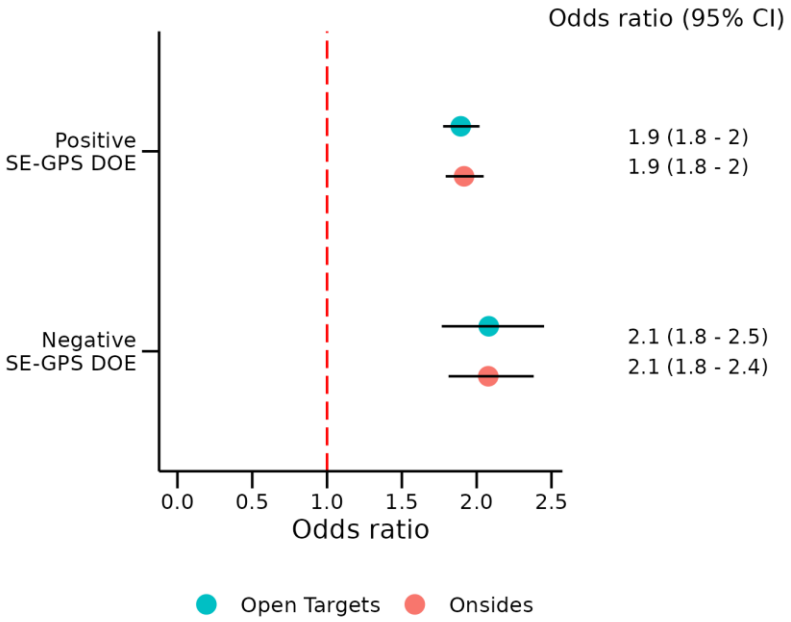
'We first evaluated the overall association of the SE-GPS with drug side effects. We observed a 1.8-fold increase risk in drug side effects (95% confidence interval (CI) =1.7–1.9, P < 4.1 x 10-167) in Open Targets and 1.9 -fold (95% CI = 1.8–2.0, P < 2.8 x 10-240) in OnSIDES.'

Manuscript changes: We have revised Figure 3 and Figure 6



A) Forest plot showing ORs with 95% CI for the association between the presence of a SE-GPS > 0 (binarized as 1) and drug side effects, adjusted for 16 phecode categories using logistic regression. This was performed across the full Open Target dataset (*n* = 1,273,056 independent drug– gene–phenotype combinations) with the OR colored in red and stratified by the number of gene targets per drug (1, 2-5, 5+; blue), the number of side effects per drug (1-4, 5-10, 10+; purple) and the number of drugs per side effect (1-10,11-30, 30+; green). For each feature, unique genes (red) and unique phenotypes (blue) are recorded on the y-axis. **B)** Replication analysis of A) using the OnSIDES dataset (*n* = 1,161,760 independent drug– gene–phenotype combinations). **C)** Forest plot showing ORs with 95% CI for the association between the presence of a SE-GPS > 0 (binarized as 1) and drug side effects, stratified by phenotype category in the Open Target dataset. **D)** Replication analysis of C) using the OnSIDES dataset. There is a break in the X-axis to include the large OR observed in the infections category and the full 95% CI is not shown here. Filled circles indicate an OR with a significant *P*-value. The red dashed line represents the null odds ratio (OR=1). CI, confidence interval; N, number; OR, odds ratio, SE, side effect.

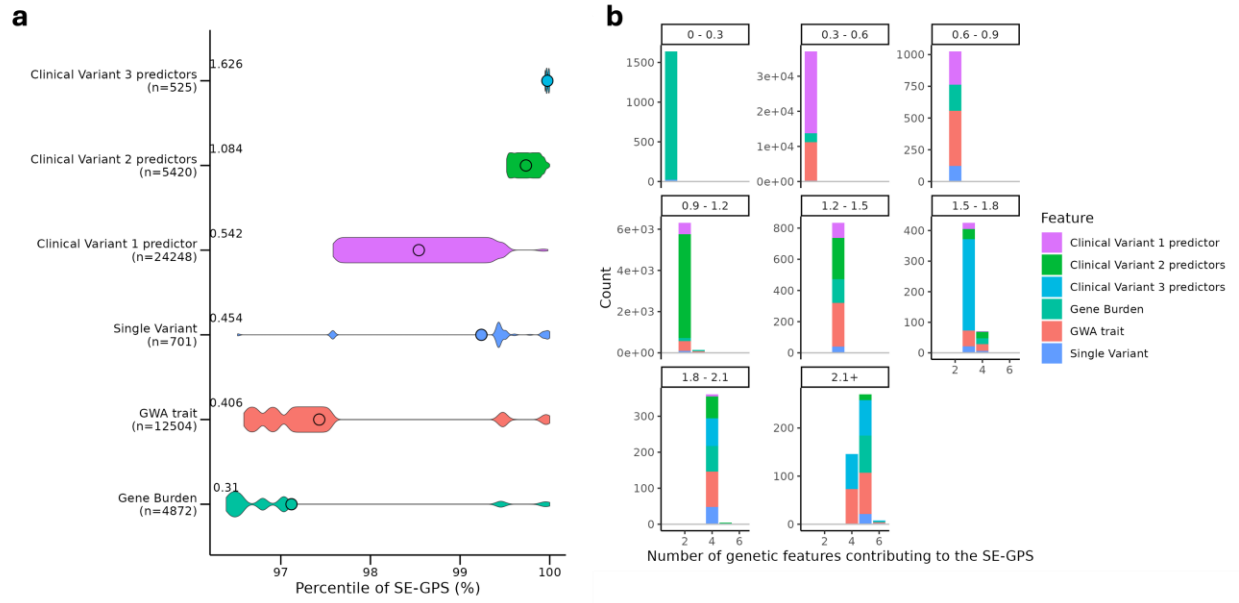
Fig. 3 Association of the SE-GPS-DOE with drug side effects in the Open Target and OnSIDES datasets.



Association between positive SE-GPS-DOE and drug side effects among inhibitor drugs in the Open Target (n= 849,282) and OnSIDES (n= 602,364) datasets, and between negative SE-GPS-DOE and drug side effects among activator drugs in the Open Target (n= 284,679) and OnSIDES (n= 341,668) datasets. In logistic regression models, non-zero SE-GPS-DOE values were binarized and compared to scores of 0, adjusting for 16 phecode categories. The red dashed line represents the null odds ratio (OR=1). CI, confidence interval; OR, odds ratio; SE-GPS-DOE, side-effect genetic priority score with direction of effect.

q. Extended Data Fig 2 a and c look similar but this simply reflect that most of the weight for the side effects is associated with clinical variants. Clinical variants 1, 2 and 3 are not explicity defined, only by first, second, third in methods. Add definition to the legend. You could show the distribution of the beta weights per feature label?

Authors reply: We have defined the Clinical Variant predictor more clearly in Extended Figure 2, referring to Clinical Variant 1 predictor, Clinical Variant 2 predictors, and Clinical Variant 3 predictors, reflecting the number of observations across OMIM, HGMD and ClinVAR (0,1, 2or 3).



'…The clinical variant feature was split according to the number of data sources (1, 2 or 3) for each gene phenotype observation.'

r. Extended Data Figure 2 legend "The sample size (n) and" sample size is ambiguous, this is the number of observation per features.

Authors reply: We have clarified the caption for Extended Data Figure 2 (now **Supplementary Fig. 7 and Supplementary Fig. 9)** to make this clearer. As the reviewer suggests, the sample size (n=) recorded under each feature on the y-axis is the number of gene-phecodeX integer observations for each feature in the Open Target dataset.

Manuscript changes: We have made the following changes to the caption of Supplementary Figure 7 and 9

'…The total sample size of gene-phecodeX integer observations for each feature (n) and the mean weight from the five cross-validated samples is recorded under each feature on the y-axis, ordered by increasing value of these weights across the six features…'

3. Results

a. This explanation "These differences may be attributed to several factors, including differences between on- and off-target effects in each category, variations in the side effects reported between the clinical trial phase and post-marketing and the filtering of side effects in our dataset with a frequency greater than 5%." On page 8 makes it very hard to evaluate if the results generated are useful.

Authors reply: We thank the reviewer for their feedback, which is a comment similar to Reviewer 1, comment 13. Specifically, from this analysis, we aimed to emphasize that the impact of genetic evidence on side effect prediction varies across disease categories, and when comparing category evidence between datasets, the degree of enrichment differs, potentially highlighting differences between side

effects reported during clinical trials (OnSIDES) and side effects reported during post-marketing (Open Targets).

Manuscript changes: We have revised the following text in the Results section:
'We observed significant variability in the odds ratio, highlighting that the impact of genetics is more pronounced in certain side effect categories than others. Infectious disease-related SEs had large odds ratios in both Open Targets and OnSIDES whereas congenital-related SEs were not significant in either dataset. Furthermore, the degree of enrichment differs between categories when comparing side effects reported in Open Targets and OnSIDES, potentially reflecting differences in side effect reporting between clinical trials and post-marketing surveillance.'

b. Page 9 focus on top < 1%. It is not clear to me if anything new is discovered.

Authors reply: Only a small proportion of gene-phecode pairs in Open Targets and OnSIDES have supporting genetic evidence, resulting in a small percentage of SE-GPS greater than 0. Furthermore, we focus on an even smaller subset of scores with at least two lines of evidence. Despite this small percentage however, this still corresponds to 365 genes and 254 phenotypes, with an OR > 2.3 in Open Targets and 344 genes and 233 phenotypes, with an OR > 2.5 in OnSIDES. By demonstrating the predictive value of genetic evidence in identifying known drug side effects, this provides confidence in applying this method to 19,422 protein-coding genes to provide evidence for side effect risk for undrugged targets without current clinical evidence.

c. Page 11 "Finally, we extended both methods to 19,422 protein-coding genes and 470 phenotypes" I thought it had already been applied to this. Make clear earlier on the genes and phenotypes being used at each stage. So this section is where something new is discovered, but the examples given go back to what is known. The section about IL2RA is interesting and should go in the discussion.

Authors reply: We apologize for the confusion and have now stated the total number of gene-phenotypes that were not included in either the Open Target or OnSIDES dataset.
In addition to including known targets, we have expanded our examples further. In Table 3, we've highlighted examples of current undrugged targets to showcase how the SE-GPS can be applied at target discovery to identify unknown side effects. We've incorporated the DrugnomeAI score, to focus on targets with a higher predicted druggability.
We agree that *IL2RA* is an important example and highlights how a side effect can point to a drug repurposing opportunity. We have included *IL2RA* in our discussion and thank the reviewer for this point.

Manuscript changes: We have added the following sentence to the Results

Finally, we extended both methods to 19,422 protein-coding genes and 502 phecodes, of which 18,427-genes and 45 phecodes were not included in either the Open Target or OnSIDE dataset.

Manuscript changes: We have added the following sentence to the Discussion

'Furthermore, by incorporating the direction of genetic effect, we demonstrate the relevance of each genetic score to the direction of the therapeutic hypothesis, distinguishing between targets inhibited and targets activated. This clarification can separate a side effect from a drug repurposing opportunity, as illustrated by *IL2RA.'*

Manuscript changes: We have included Table 3.

Table 3: Examples of possible side effects for undrugged targets using the SE-GPS and SE-GPS-DOE.

Prioritized examples	Genetic Evidence	Direction of effect evidence
Gene: *GJB2* Phenotype (code): Hearing impairment (396) SE-GPS:2.34 SE-GPS-DOE:1.89 Suggested Indication: GJB2-targeted cancer immunotherapy[23] Mechanism: Inhibitor	ClinVar: Autosomal dominant nonsyndromic hearing loss 3A; Hearing loss, autosomal recessive; Hearing impairment HGMD: deafness nonsyndromic sensorineural; sensorineural hearing loss; deafness autosomal recessive 1; deafness nonsyndromic; deafness; hearing loss non-syndromic; hearing impairment nonsyndromic; hearing impairment; deafness autosomal dominant 3; hearing loss; ichthyosis follicularis sensorineural hearing loss and punctate palmoplantar keratoderma; vohwinkel syndrome; hearing loss non-syndromic autosomal recessive; ichthyosiform erythroderma corneal involvement & deafness; deafness and palmoplantar hyperkeratosis; deafness and palmoplantar keratoderma; knuckle pads leukonychia sensorineural deafness; knuckle pads hyperkeratosis and deafness; sensorineural hearing loss & leukonychia; keratitis-ichthyosis-deafness syndrome; hearing impairment postlingual; hearing loss non-syndromic autosomal dominant; hearing impairment bilateral sensorineural OMIM: bart-pumphrey syndrome, 149200 (3); deafness, autosomal dominant 3a, 601544 (3); deafness, autosomal recessive 1a, 220290 (3); hystrix-like ichthyosis with deafness, 602540 (3); keratitis-ichthyosis-deafness syndrome, 148210 (3); keratoderma, palmoplantar, with deafness, 148350 (3); vohwinkel syndrome, 124500 (3) RaVAR GB: hearing loss; sensorineural hearing loss L2G: hearing loss; age-related hearing impairment; sensorineural hearing loss	ClinVar: LOF HGMD: LOF OMIM: LOF RaVAR Gene Burden: LOF
Gene: *SLC13A5* Phenotype (code): Epilepsy, recurrent seizures, convulsions (330) SE-GPS:1.54 SE-GPS-DOE:1.60	ClinVar: Developmental and epileptic encephalopathy, 25 HGMD: epileptic encephalopathy early infantile; epileptic encephalopathy; west syndrome & severe psychomotor development retardation; developmental and epileptic encephalopathy; epileptic encephalopathy early-onset; epilepsy early-onset; encephalopathy; epilepsy; kohlschütter-tönz	ClinVar: LOF HGMD: LOF OMIM: LOF

Suggested Indication: Kidney disease[24] Mechanism: Inhibitor	syndrome; global developmental delay epilepsy chronic kidney disease; paediatric movement disorder OMIM: developmental and epileptic encephalopathy 25, with amelogenesis imperfecta, 615905 (3)	
Gene: *GJA1* Phenotype (code): Abnormal intraocular pressure (375) SE-GPS:1.83 SE-GPS-DOE:1.36 Suggested Indication: Alzheimer's[25] Mechanism: Inhibitor	HGMD: microcornea and glaucoma; open angle glaucoma and microcornea OMIM: oculodentodigital dysplasia, 164200 (3) RaVAR GB: glaucoma L2G: open-angle glaucoma	HGMD: LOF OMIM: LOF RaVAR Gene Burden: LOF
Gene: *ORAI1* Phenotype (code): Immunodeficiencies (179) SE-GPS:1.03 SE-GPS-DOE:1.04 Suggested Indication: Duchenne muscular dystrophy[26] Mechanism: Inhibitor	HGMD: immunodeficiency combined; severe combined immune deficiency syndrome; immunodeficiency muscular hypotonia & anhidrotic ectodermal dysplasia; severe combined immune deficiency syndrome and residual t-cell function OMIM: immunodeficiency 9, 612782 (3)	HGMD: LOF OMIM: LOF

SE-GPS, side effect genetic priority score; SE-GPS DOE, side effect genetic priority score with direction of effect, OMIM, Online Mendelian Inheritance in Man; HGMD, Human Gene Mutation Database

4. Code. I was surprised not to receive the code to review. I looked instead at https://zenodo.org/records/10095684 from the published paper. This code starts with OT_drugdataset split into CV sets. The code to generate this file from downloaded data and the primary data were not provided, nor the final dataset are provided. I found this disappointing. The files extracted from OT and Onsides do not seem big, and especially the derived files used for the regression should be supplied.

Authors reply:
We apologize for not providing the code initially and have included the final datasets for Open Targets, OnSIDES and the entire integrated dataset (19,422 genes- and 503 phecodex integer pairs). We now include all analysis code to create the SE-GPS and subsequent analyses under the zenodo link: https://zenodo.org/records/15334136?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImY1ZWY3ZmVmLWM0YmEtNGM1YS1iZjkxLTgxYWJlMGM2OGM1MCIsImRhdGEiOnt9LCJyYW5kb20iOiJiZTI1ZTJjMjRkNGMxOWM5MmQ2OWY3ZjhmN2UwMjFiYSJ9.dGGnnoLIUVaXQ4VKndr8c2vayJRAC05b0caR7gakbUrncTP9fg1hnSVzP132KFKErSKBuRKhc32ONJJT8Wfbrw and https://github.com/rondolab/SE-GPS.

Other comments
1. Page 5. Correlation of 0.7 increasing to 0.74. It is ambiguous what is being correlated. Explain in more detail?

Authors reply: We have clarified the ratio of reporting frequency (RRF) further in the results and method text to indicate that we are comparing the frequency of each side effect in Open Targets, against the frequency of that side effect reported in OnSIDES. From **Supplementary Fig. 1,** we see that most side effects have low RRF values, indicating that these are drug specific. We further indicate which side effects have a frequency greater than >5% and are thus removed from our dataset.
We initially performed this correlation for the side effect data pre and post QC filtering but have since removed the post filtering as felt this added confusion.

Manuscript changes: We have made the following changes to the results section:

To measure the frequency of reported side effects and compare differences in side effect reporting across clinical trial data (OnSIDES) and post-marketing data (Open Targets), we plotted the ratio of reporting frequency (RRF), calculated as the normalized count of drugs associated with a given side effect from Paccanaro et al. **Supplementary Fig. 1** shows the RRF of each side effect in Open Targets, correlated against the side effect data in OnSIDES. Both datasets indicate that most reported side effects are drug-specific, with similar reported frequency (r=0.7).

Manuscript changes: We have made the following changes to the Methods section:

We calculated the side effect ratio of reporting frequency (RRF) as detailed in equation 6 from Paccanaro et al[33]. Specifically, for each side effect, the side effect ratio of reporting frequency (RRF) represents a normalized count of the number of associated drugs. This equation is as follows:

$$RRF(j) = \frac{\sum_i^n X_{ij}}{Z}$$

where Xij represents the entry in row i, column j of the matrix X, n represents the total number of drugs and Z is the maximum number of associations for the side effects.

2. Page 5. The 987 drugs from OpenTargets and 806 drugs from Onsides, provide venn diagrams of overlaps of drugs and genes and side effects and other logical comparators.

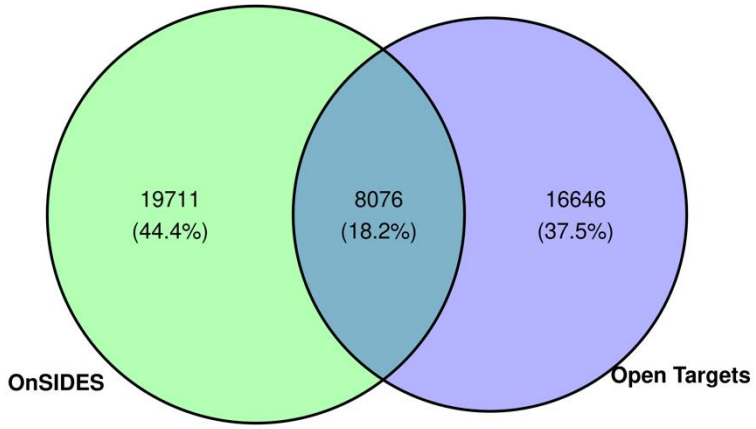
Authors reply: We appreciate the reviewers suggestion regarding the use of Venn diagrams to illustrate the overlap between the datasets. As we found a large percentage of drugs (85.97%) in OnSIDES also present in Open Targets we removed any overlapping drugs during QC from the Open Target dataset to ensure independence.
While there are consequently no overlapping drugs, we provide a venn-diagram to highlight the number of overlapping gene-phecode pairs with a side effect between the two final datasets. We note in the Venn diagram that 18.2% of gene-phecode side effect pairs are shared between Open Targets and OnSIDES, while the remaining pairs were unique to either Open Targets or OnSIDES. We have included this is **Supplementary Fig. 18.**

Manuscript changes: We have made the following changes to the results section:

We removed any drugs from our Open Target dataset that were also present in the replication dataset OnSIDES, however we note that 18.2% of gene-side effect pairs are still found in both datasets (**Supplementary Fig. 18).**

Manuscript changes: We have included Supplementary Fig. 18:

Supplementary Fig. 18 Overlap of gene-phecode side effect pairs in Open Targets and OnSIDES



Venn diagram showing the number of unique and shared gene-phecode side effect pairs between OnSIDES and Open Targets

3. Methods.Page 14 Exclusion of 58 comm side effects – can these be listed in a supp Table. Phecodes that lacked genetic evidence. What is the N. Can these side effects be listed.

Authors reply: We have included the 58 common side effects from Open Targets and the 278 common side effects from OnSIDES in **Table S6** and **Table S10**. Furthermore, we have included the 46 and 52 phecode integer terms that we removed that lacked genetic evidence as **Table S7** and **Table S11** from Open Targets and Onsides respectively.

We have added Table S6, Table S7, Table S10 and Table S11.

4. Page 14 "Finally, we removed any phecodes that lacked genetic evidence" Both drug-indications and Side effects are mapped to phecodes. Would be helpful to distinguish between, SE-phecode and DI-phecodes.
Not clear what you mean. Each phecodes is linked to a DI-gene, but the gene had no entries from the 5 features.

Authors reply: We have now included the mapped side effect terms to phecodes in Tables S8 and S12 for both Open Targets and OnSIDES and the mapped drug indications to phecodes in Tables S9 and S13. In Open Targets, 362 phecodes overlapped between side effect and drug indication terms, while 85 phecodes were unique to side effects. Similarly, in OnSIDES, 366 phecodes were shared, with 58 phecodes unique to side effects.

For certain side effects that were mapped to a corresponding phecode, there were no genetic associations observed for that phecode across the four genetic features. As a result, these side effects and corresponding phecodes were excluded from the dataset. For example, the side effect cardiogenic shock which mapped to phecode SS_810 was observed across 89 drugs, but our genetic features had no

associations for this phecode. We thus excluded this side effect from our list. These phecodes are listed in **Table S7 and Table S11**.

Manuscript changes: We have added Tables S7, S8, S9, S11, S12 and S13.

Manuscript changes: We have made the following changes to the Methods section:

'We list the side effects and drug indications mapped to phecode integers in Tables S8 and S9'

.'We list the side effects and drug indications mapped to phecode integers in Tables S12 and S13'

In general, The use of the word "phenotype" is confused and confusing.
On Page 5 "We integrated both side effect datasets with human genetic evidence at the gene-phenotype level", I think here phenotype is side effect.

Authors reply: We apologize for the confusion and have used the term phenotype to reflect phecode integer codes. We have changed phenotype to phecode to address this confusion.

In the next line "GWAS phenotype" could be "GWAS trait"

Authors reply: We thank you for the suggestion and have referred to GWAS phenotypes as GWAS traits.

Later in the section is eQTL phenotype suggest trait-eQTL that maps to a phecode-eQTL
Page 16 disease associated phecodes

Authors reply: We kept the term eQTL phenotype to remain consistent with our previous study.

5. Page 15 "We subsetted both drug datasets to drugs that either had a box warning or had been withdrawn due to toxicity risk. In Open Targets, these side effects are annotated as toxicity classes, which we then mapped to phecode categories as follows" Is this supposed to be where the 15 phecode categories comes from – fewer than 15 here

Authors reply: No, in Open targets, drug warnings are reported as toxicity classes, which we have included as **Table S14.** Of the original PhecodeX 18 categories, defined by Shuey et al.[32], we excluded neonatal, and pregnancy-related categories, leaving 16 phecodeX categories in our analysis. Of these, 12 mapped to toxicity classes as indicated in the table below.

Manuscript changes: We have added Table S14
Table S14: Table of toxicity class mappings to Phecode categories.

Toxicity class	Phecode category
Carcinogenicity	Neoplasms
Cardiotoxicity	Cardiovascular
Dermatological toxicity	Dermatological
Gastrointestinal toxicity	Gastrointestinal

Hematological toxicity	Blood/immune
Hepatotoxicity	Gastrointestinal
Immune system toxicity	Blood/immune
Infectious disease	Infections
Metabolic toxicity	Endocrine/metabolic
Musculoskeletal toxicity	Musculoskeletal
Nephrotoxicity	Genitourinary
Neurotoxicity	Neurological
Psychiatric toxicity	Mental
Respiratory toxicity	Respiratory
Vascular toxicity	Cardiovascular

6. Removal of drugs from OpenTargets that were present in Onsides – explain why.
Is it because the databases actually extract info from the same primary sources.
What is a definition of a drug in this case? Some drugs are very similar and could be combined?

Authors reply: We found a large percentage of drugs in OnSIDES that were also present in Open Targets (85.97%). While these datasets source extract the side effect data from different primary sources, Open Targets uses FAERS where OnSIDES uses drug label using different methods, there still an overlap between side effects reported. As a result, we removed the overlapping drugs from the Open Target dataset to ensure independence, which enabled us to use OnSIDES as an independent validation set to evaluate SE-GPS.

7. Page 19: "The clinical variant category was derived from genetic association data from ClinVar19, HGMD20 and OMIM21, which we consolidated into a single feature recorded as the number of overlapping entries." In the original paper these were separate; explain why now consolidated.

Authors reply: We chose to consolidate these features because a substantial portion of observations in ClinVar, in particular, overlapped with those in OMIM and HGMD (Supplementary Fig. 2). In a mixed-effects model, this overlap diluted the individual contribution of each feature, leading to a loss of signal. To address this, we combined the features to better capture their collective effect.

8. In the original paper, pQTLs were used as well as eQTLs; explain why these are not usesd

Authors reply: We choose not to include pQTLs in this study based on results from a univariate analysis (shown in Table below), where the pQTL feature was not significantly associated with the drug side effects in either the Open Targets or OnSIDES datasets when we excluded side effects where the drug was approved for an indication that shared the same phecode term.

However, as part of updating the locus-to-gene (L2G) scores using the latest Open Targets release (v25.03), we have now integrated the corresponding colocalization results from overlapping molecular

QTL datasets to infer directionality. We thus included colocalization results from overlapping protein QTLs (pQTLs), as well as expression QTLs (eQTLs), and splice QTLs (sQTLs). However, we note that the number of available pQTL studies is substantially smaller than for eQTLs and sQTLs, which may limit the overall contribution of pQTLs in the current analysis.

Outcome	OR	CI	P.value
Main indication	4.136	3.6-4.8	$1.55x10^{-80}$
Side effect	1.243	1.1-1.5	0.012
Side effect (exclude overlapping MI)	0.947	0.8-1.2	0.592
Main indication	3.827	3.3-4.4	$2.10\ x10^{-84}$
Side effect	1.222	1-1.5	0.042
Side effect (exclude overlapping MI)	1.079	0.9-1.3	0.468

9. Clinical Variants. "We applied a more stringent filtering approach than previously.." this means compared to ref 13 or 15…explain why

Authors reply: We choose to apply stringent thresholds to reduce the number of false positives in our dataset. Previously our filtering included: 'First, evidence was filtered on clinical significance terms, which are as follows: 'likely pathogenic,' 'association,' 'confers sensitivity,' 'drug response' and 'pathogenic'. Second, we filtered based on the confidence of the submission assigned as follows: 'criteria provided, single submitter'; 'criteria provided, conflicting interpretations'; 'criteria provided, multiple submitters, no conflicts'; 'reviewed by expert panel' and 'practice guideline'.
We have reduced this evidence to 'This evidence was filtered on clinical significance terms: likely pathogenic and pathogenic, and based on the confidence of the submission: criteria provided, multiple submitters, no conflicts, reviewed by expert panel and practice guidelines'.

10. Statistical Analysis: "We calculated the side effect ratio of reporting frequency (RRF) as detailed in equation 6 from Paccanaro et al4. " Please repeat it here, so save readers looking it up.

Authors reply: We have added equation 6 from Paccanaro et al to the method section.

Manuscript changes: We have added the following sentences to the Methods section:

We calculated the side effect ratio of reporting frequency (RRF) as detailed in equation 6 from Paccanaro et al. Specifically, for each side effect, the side effect ratio of reporting frequency (RRF) represents a normalized count of the number of associated drugs. This equation is as follows:

$$RRF(j) = \frac{\sum_i^n X_{ij}}{Z}$$

where Xij represents the entry in row i, column j of the matrix X, n represents the total number of drugs and Z is the maximum number of associations for the side effects.

11. Extended Data 4. Since this is a subset of N=967 genes selected from Extended Data Figure 1. Add number of genes to legend of extended Data Figure 1, to allow comparison.

Authors reply: We thank the reviewer for highlighting this and have added the total number of drugs, genes and phenotypes to the legend of Supplementary Fig. 6 and Supplementary Fig. 13 (previously Extended Data figure 1 and Extended data figure 4.)

Manuscript change: We have updated the legend of Supplementary Fig. 6 (previously Extended Data Figure 1):

'The Open Target dataset (n=1,014 drugs, 762 genes and 447 phenotypes) was split into 80% training and 20% test sets of non-overlapping groups of unique gene-phenotype pairs in five-fold cross-validation.'

Manuscript change: We have updated the legend of Supplementary Fig. 13 (previously Extended Data Figure 4):

'The Open Targets dataset was restricted to drugs classified as inhibitor or activator (n=922 drugs, 733 genes and 447 phenotypes). The dataset was split into 80% training and 20% test sets of non-overlapping groups of unique gene-phenotype pairs in five-fold cross-validation.'

12. Online Tool:

a. The first gene listed in "Gene examples" GRIN2A – gives no data, which doesn't seem ideal – even though GRIN2A is listed in Table 2. Other genes in Table 2 are not in the database.

Authors reply: We recognize that using a text box to search for Gene targets made this tool less clear. To improve its function, we have replaced it with a dropdown input, including a live search for easier navigation. In addition, we have adjusted the SE-GPS threshold to 0.6 and thus a greater number of gene-phecodes should be reported, including evidence for GRIN2A and other genes reported in Table 2. This threshold can be adjusted, allowing the user to modify the stringency of evidence required.

b. The number of genes and phenotypes on the front page differs from the numbers on the summary page (= number in Fig 1). The numbers to which the algorithm was applied is different to the numbers with entries in the search. Eg, The front page says 466 phenotypes but the phenotype search page only has ~120.

Authors reply: On the front page we report the number of genes and phenotypes with a SE-GPS >0 whereas for figure 1 we report the total number of genes and phenotypes that we applied this method to. We have clarified this further in the About tab on the shiny application. The number of phenotypes on the search page changes depending on the applied SE-GPS cutoff. When the SE-GPS cutoff > 0, then there are 499 phecodes in the drop-down list.

Website change:
'…This was applied to 19,422 protein-coding genes and 502 phenotypes and we obtained SE-GPS >0 for 15,139 genes and 499 Phenotypes.'

c. In the Summary info, an explanation should be given why many entries have no information in the ONSIDES or OpenTargets side effects column.Allow download on the files behind the search.

Authors reply: We thank the reviewer for this observation and have clarified in the About section that only a proportion of the total gene-phecode pairs have a corresponding side effect reported in either Open Targets or OnSIDE, resulting in many entries with no information. Furthermore, we have included a 'download filtered results' tab to allow users to download the evidence table behind their search query.

Website changes:

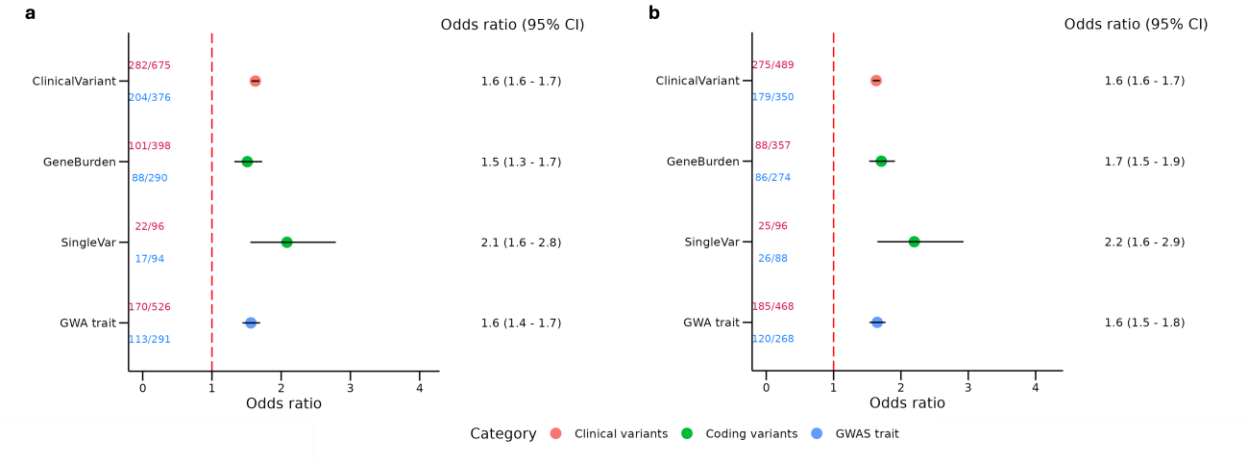
d. It seems strange that there is no information about drugs. Make a link from gene name to drugdatabase?

Authors reply: We chose not to include drug names because, in many cases, a gene target was associated with a large number of drugs. For example, DRD2 was reported as the target for over 70 different drugs. Including all of these would have made the data presentation less clear.

13. Figure 2. Improve the legend so a reader can understand as standalone. Write so it is clear what the OR is for.

Authors reply: We have updated OR with odds ratio for Figure 2.

Manuscript changes: We have revised Figure 2 below:



14. Figure 3. Make it clear that 80% of OpenTargets and OnSides is used to generate the SE-GPS, and that this Figure is results for application of the remaining 20%

Authors reply: When we calculate the SE-GPS in Open Targets we calculate this for the five test folds and then combine these five folds so that the score is calculated in 100% of the data. For OnSIDES we calculate the scores in the 100% of the data using the coefficients from Open Target. We apologize for not making this clear and have clarified this in the methods.

Manuscript changes: We have updated the Methods section to include the following sentence:

'We combine the five test folds for downstream analyses.'

Reviewer #3 (Remarks to the Author):

Duffy et al have submitted a well-written, interesting manuscript which describes their genetic priority score for predicting drug side effects using human genetic evidence. There is arguably an overlap in concept with their 2024 Nature Genetics publication, but I feel this study successfully builds on their prior paper and focuses on a novel, more specific topic: side effect prediction. My recommendation is to accept with minor edits, including inclusion of the code used in this work (manuscript currently says this will be made available upon publication). I have not been able to review or run any of the code.
I can confirm I was able to access all other files pertinent to the review and was able to access the online web tool.

Authors reply: We thank the Reviewer for their positive comments and greatly appreciate their suggestions and feedback. We have addressed their comments below.

I would like the authors to address the following points:
Line 92: 'we removed common side effects observed in greater than 5% of drugs'. As there is undoubtedly value in predicting common side effects such as nausea, I would like the authors to expand on this point further. Is their approach not suited to predicting common side effects?

Authors reply: While we agree with the reviewer that there is huge value in predicting common side effects, we removed them following the approach by Nguyen et al. [34], which highlighted that common side effects may be less likely to reflect target-mediated mechanisms, and instead a consequence of off-target or systemic effects. We list the common side effects that we removed from Open Targets and OnSIDES in **Table S6 and Table S8**, and show for example, that the top side effects such as nausea is reported in 512 drugs in Open Targets and 1577 drugs in OnSIDES, making it non-specific.

Manuscript changes: We edited the following sentence in the Methods
'as these side effects are less likely to reflect target-mediated mechanisms, and instead a likely consequence of off-target or systemic effects[34]'

Line 103: what is the overlap in drug side effects, compounds etc reported in OpenTargets and OnSIDES? I do not believe these resources will be completely distinct, and so a detailed characterisation of the differences between these datasets is required before assigning them to be training and validation datasets.

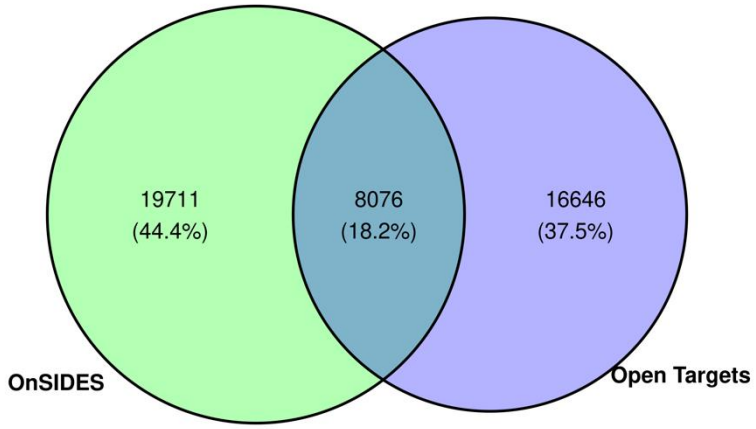
Authors reply: We found a large percentage of drugs in OnSIDES that were also present in Open Targets (85.97%) and thus removed these overlapping drugs from our Open Targets dataset during QC. While there are consequently no overlapping drugs, the reviewer is correct that these resources are still not completely distinct. We provide a venn-diagram (Supplementary Fig. 18) to highlight the number of overlapping gene-phecode pairs with a side effect between the two final datasets. We note in the Venn diagram that 18.2% of gene-phecode side effect pairs are shared between Open Targets and OnSIDES, while the remaining pairs were unique to either Open Targets or OnSIDES.

Manuscript changes: We have made the following changes to the results section:

We removed any drugs from our Open Target dataset that were also present in the replication dataset OnSIDES, however we note that 18.2% of gene-side effect pairs are still found in both datasets (**Supplementary Fig. 18).**

Manuscript changes: We have included Supplementary Fig. 18:

Supplementary Fig. 18 Overlap of gene-phecode side effect pairs in Open Targets and OnSIDES



Venn diagram showing the number of unique and shared gene-phecode side effect pairs between OnSIDES and Open Targets

Line 127: 'We retained this side effect filter..' Please add a brief comment to the Discussion section regarding how this may limit the utility of the approach for predicting side effects of a similar phenotype to the disease

Authors reply: We found that implementing the side effect filter (removal of phenotype terms that matched a drug indication) reduced the effect sizes for each predictor. However, when applying this method to all 19,422 genes, we generated a SE-GPS for all possible gene – phecode combinations, regardless of if the gene-phecode is a side effect or drug indication. Thus the phenotype headache, which we would have removed due to its prevalence, still has a corresponding SE-GPS score under the phecode NS_331, with a SE-GPS > 0.6 for *FHL5, COL4A1, CACNA1A, SCN1A, and ATP1A2* to list a few examples.

As a result, however, this means that there is a large overlap between targets with predicted side effect risk and a drug indication. We have already included this as a comment in the discussion.

'The overlap between targets with predicted side effect risk and known drug targets for similar drug indications emphasizes the importance of integrating all aspects of genetic evidence and disease biology when selecting a potential drug target to ensure that it is both effective and safe'

Line 224: 'In OnSIDES, we observed similar enrichments for the positive SE-GPS DOE..' Does this present any caveats for application of the approach? Please expand slightly.

Authors reply: We observed similar results in both the discovery OpenTargets dataset and the validation OnSIDES dataset with respect to the association of the SE-GPS-DOE with side effect risk. We view these consistent enrichments – despite differences between side effect reporting in both OpenTargets and OnSIDES - as strong support for the robustness of our tool in predicting SE risk across different drug side effect datasets.

Line 252 and elsewhere: ‘mendelian’ requires a capital ‘M’

Authors reply: We have made this edit and thank the reviewer for spotting this error.

Line 282: remove comma before ‘therefore’
Authors reply: We have made this edit and thank the reviewer for spotting this error.

Line 340: a table would be much clearer for illustrating the mappings between toxicity classes and phecode categories

Authors reply: We agree with the reviewer and thank them for this suggestion. We have included **Table S14** to detail the mappings between toxicity classes and phecode categories.

Table S14: Table of toxicity class mappings to Phecode categories.

Toxicity class	Phecode category
Carcinogenicity	Neoplasms
Cardiotoxicity	Cardiovascular
Dermatological toxicity	Dermatological
Gastrointestinal toxicity	Gastrointestinal
Hematological toxicity	Blood/immune
Hepatotoxicity	Gastrointestinal
Immune system toxicity	Blood/immune
Infectious disease	Infections
Metabolic toxicity	Endocrine/metabolic
Musculoskeletal toxicity	Musculoskeletal
Nephrotoxicity	Genitourinary
Neurotoxicity	Neurological
Psychiatric toxicity	Mental
Respiratory toxicity	Respiratory
Vascular toxicity	Cardiovascular

Line 409: a double comma

Authors reply: We have made this edit and thank the reviewer for spotting this error.

Line 728: ‘at which level of significance?’ appears to be a comment for the authors, but needs to be addressed!

Authors reply: We have included significant *P*-value > 0.05 and thank the reviewer for spotting this comment.

Manuscript changes:
'Each cross-validated sample is color labeled and filled circles indicate a beta coefficient with a significant P-value > 0.05 and the 95% CIs are defined as error bars'

Line 7TAB: ' (include ref?)..' – comment needs to be addressed and reference included!

Authors reply: We have updated the caption of Extended Data Fig.4 to include the severity score reference as well as include significant *P*-value > 0.05.

Manuscript changes:
'… The side effect outcome was weighted by severity using a crowdsourced severity score[2]… Each cross-validated sample is color labeled and filled circles indicate a beta coefficient with a significant *P*-value > 0.05 and the 95% CIs are defined as error bars.'

Line 761: 'at 5% or multiple testing corrected?' Please address

Authors reply: We have included significant *P*-value > 0.05 and thank the reviewer for spotting this comment.

Line 806: '(at which level of significance?)' – this is a good question, please answer
Authors reply: We have included significant *P*-value > 0.05 and thank the reviewer for spotting this comment.

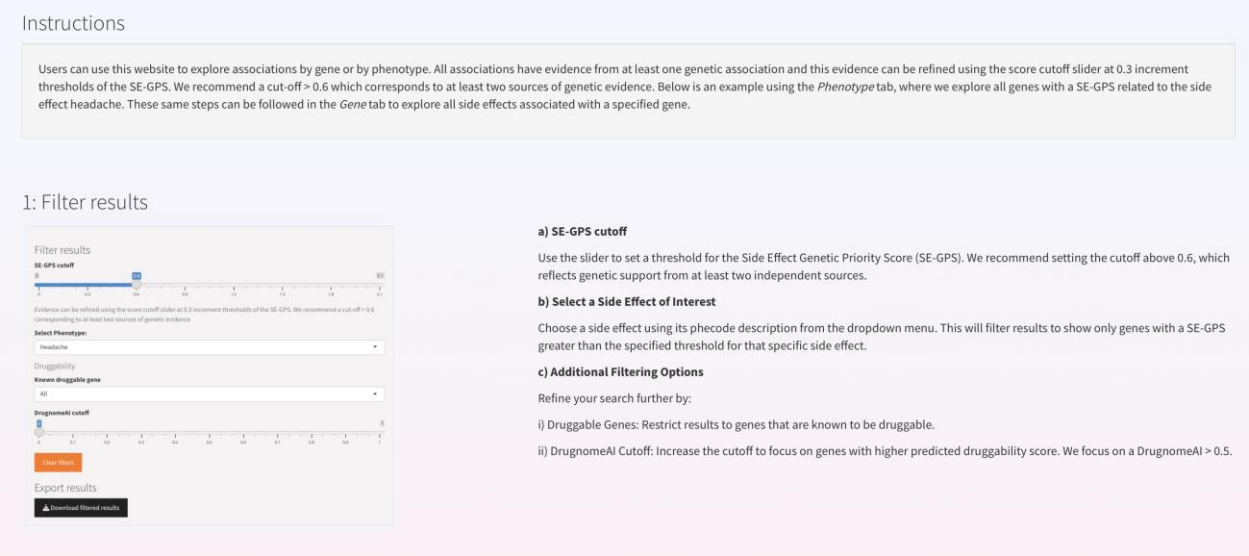
Line 822: '(at which level of significance?)' – what they said
Authors reply: We have included significant *P*-value > 0.05 and thank the reviewer for spotting this comment.

Line 836: '(at which level of significance?)' – same again
Authors reply: We have included significant *P*-value > 0.05 and thank the reviewer for spotting this comment.

The web tool is responsive and has clear utility, but more detailed documentation such as worked examples or tutorials would be extremely beneficial.

Authors reply: We agree that an example would provide clearer utility and have expanded the Instructions tab in the About page to detail how to use this tool with headache as a side effect example. Below is a screenshot of the instructions page, which can be further scrolled down to show the results and evidence.

Website changes: We have included an additional tab Instructions.



I look forward to seeing the revised manuscript.

References:

1. Shicheng, G. https://community.opentargets.org/t/remove-all-category-based-analysis-of-genebass-from-opentargets-platform/1393.
2. Gottlieb, A., Hoehndorf, R., Dumontier, M. & Altman, R.B. Ranking adverse drug reactions with crowdsourcing. *J. Med. Internet Res.* **17**, e80 (2015).
3. Buniello, A. *et al.* Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Research* **53**, D1467-D1475 (2024).
4. Khaleel, M.A., Khan, A.H., Ghadzi, S.M.S., Adnan, A.S. & Abdallah, Q.M. A Standardized Dataset of a Spontaneous Adverse Event Reporting System. *Healthcare (Basel)* **10**(2022).
5. Tanaka, Y. *et al.* OnSIDES (ON-label SIDE effectS resource) Database : Extracting Adverse Drug Events from Drug Labels using Natural Language Processing Models. *medRxiv*, 2024.03.22.24304724 (2024).
6. Landrum, M.J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062-D1067 (2017).
7. Stenson, P.D. *et al.* The Human Gene Mutation Database (HGMD(®)): optimizing its use in a clinical diagnostic or research setting. *Hum. Genet.* **139**, 1197-1207 (2020).
8. Hamosh, A. *et al.* Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **30**, 52-5 (2002).
9. Karczewski, K.J. *et al.* Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genom* **2**, 100168 (2022).
10. Cao, C. *et al.* RAVAR: a curated repository for rare variant-trait associations. *Nucleic Acids Res.* **52**, D990-D997 (2024).
11. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527-1533 (2021).
12. Karczewski, K.J. *et al.* Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. *medRxiv*, 2024.03.13.24303864 (2024).

13. Walker, V.M., Davey Smith, G., Davies, N.M. & Martin, R.M. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. *Int J Epidemiol* **46**, 2078-2089 (2017).
14. Fahed, A.C., Philippakis, A.A. & Khera, A.V. The potential of polygenic scores to improve cost and efficiency of clinical trials. *Nat Commun* **13**, 2922 (2022).
15. Shuey, M.M. *et al.* Next-generation phenotyping: introducing phecodeX for enhanced discovery research in medical phenomics. *Bioinformatics* **39**, btad655 (2023).
16. Lewis, J.D. *et al.* Risk of bladder cancer among diabetic patients treated with pioglitazone: interim report of a longitudinal cohort study. *Diabetes Care* **34**, 916-22 (2011).
17. Stein, D. *et al.* Genome-wide prediction of pathogenic gain- and loss-of-function variants from ensemble learning of a diverse feature set. *bioRxiv*, 2022.06.08.495288 (2022).
18. Duffy, Á. *et al.* Development of a human genetics-guided priority score for 19,365 genes and 399 drug indications. *Nat. Genet.* **56**, 51-59 (2024).
19. Minikel, E.V. & Nelson, M.R. Human genetic evidence enriched for side effects of approved drugs. *PLoS Genet* **21**, e1011638 (2025).
20. Carss, K.J. *et al.* Using human genetics to improve safety assessment of therapeutics. *Nat. Rev. Drug Discov.* **22**, 145-162 (2023).
21. Denison, H. *et al.* Diacylglycerol acyltransferase 1 inhibition with AZD7687 alters lipid handling and hormone secretion in the gut with intolerable side effects: a randomized clinical trial. *Diabetes, Obesity and Metabolism* **16**, 334-343 (2014).
22. Booth, B. Painful Truth: The Successful Failure Of A Biotech Startup. *Forbes* (17 November 2017) https://www.forbes.com/sites/brucebooth/2017/.
23. Jia, Y. *et al.* Pan-cancer analysis of the prognostic and immunological role of GJB2: a potential target for survival and immunotherapy. *Front Oncol* **13**, 1110207 (2023).
24. Gill, D. *et al.* The citrate transporter SLC13A5 as a therapeutic target for kidney disease: evidence from Mendelian randomization to inform drug development. *BMC Medicine* **21**, 504 (2023).
25. Kajiwara, Y. *et al.* GJA1 (connexin43) is a key regulator of Alzheimer's disease pathogenesis. *Acta Neuropathologica Communications* **6**, 144 (2018).
26. Cheng, A.J., von Walden, F. & Lanner, J.T. Orai1 as a potential "fits-all approach" therapeutic target for the treatment of DMD. *Journal of General Physiology* **155**, e202213224 (2023).
27. Raies, A. *et al.* DrugnomeAI is an ensemble machine-learning framework for predicting druggability of candidate drug targets. *Communications Biology* **5**, 1291 (2022).
28. Wishart, D.S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074-D1082 (2018).
29. Zdrazil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res* **52**, D1180-D1192 (2024).
30. Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* **9**(2017).
31. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19-34 (2017).
32. Shuey, M.M. *et al.* Next-generation phenotyping: introducing phecodeX for enhanced discovery research in medical phenomics. *Bioinformatics* **39**(2023).
33. Galeano, D. & Paccanaro, A. Machine learning prediction of side effects for drugs in clinical trials. *Cell Rep Methods* **2**, 100358 (2022).
34. Nguyen, P.A., Born, D.A., Deaton, A.M., Nioi, P. & Ward, L.D. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat. Commun.* **10**, 1579 (2019).

Reviewer #1 (Remarks to the Author):

The authors have responded to all of my remarks and have addressed them satisfactorily.

Authors reply: We thank the reviewer for taking the time to go through our response.

Reviewer #1 (Remarks on code availability):

There is a README file included with the code that appears to provide instructions for installation and running the application. However, there is no explanation on how to retrieve the data underlying the analysis. If this is a requirement from the journal, such information should be added to ensure full reproducibility.

Authors reply: We appreciate the reviewer's feedback. We have revised the README file to include clear instructions on downloading the data necessary to rerun all scripts in the GitHub repository.

'**Data Access:**
The data required to run the scripts can be downloaded as a compressed folder (Data.zip) from the following Zenodo link: https://zenodo.org/records/15334136
After downloading, please extract the contents of Data.zip into the same directory as the scripts folder. This will create a Data/ folder that is accessed by the pipeline.'

Reviewer #2 (Remarks to the Author):

The authors have made an impressive response document. I find the paper much clearer the added tables and Figures are helpful.

Authors reply: We thank the reviewer for taking the time to go through our response and appreciate their feedback.

I have 2 points which I feel should be addressed.
1) Comment 23 of reviewer 1 asked about correction for multiple testing. The authors state that they use p< 0.05 as significant. I agree with reviewer 1 that the p-value threshold for significance should be stated (at line 926 it says "P-value > 0.05", which should be "P-value < 0.05". Furthermore, there should be a threshold for significance that accounts for multiple testing.

Authors reply: We have updated the p-value threshold to 'P-value < 0.05' and thank them for pointing out this error. Furthermore, for Figures 4, 5 and Supplementary Figures 16 and 17, we have now applied a Bonferroni correction for multiple testing, using a significance threshold of $P < 0.05 / 8$.

Manuscript changes: We have revised the following text in Figures 4, 5 and Supplementary Figures 16 and 17.

'ORs with 95% CIs are defined in the forest plot as circles and error bars, with filled circles indicating an OR with a significant *P*-value < 0.05 after correcting for multiple testing.'

2) I find Supp Fig 8 (p32 of rebuttal document) illuminating. I believe the authors need to be upfront. If

an SE-GPS > 1.6 is observed then this is useful information, but that the vast, vast majority of the time the SE-GPS is uninformative. This should feature in the abstract, Figure1, Discussion.

Authors reply: We agree with the reviewer that the SE-GPS is informative for only a small proportion of gene–phecode pairs. In response, we have revised the text in the abstract, results and discussion to make clearer the percentage of the dataset with a SE-GPS > 0.

Manuscript changes: We have revised the following text in the Abstract

'We construct the SE-GPS in the Open Target dataset using post-marketing side effect data, externally test it in OnSIDES using side effects reported from drug labels and then generate SE-GPS for 19,422 protein coding genes and 502 phecodes, of which 1.7% had a SE-GPS > 0.'

Manuscript changes: We have revised the following text in the results

Line 146: 'Across this matrix, 1.7% of gene-phecode pairs had a least one source of genetic evidence.'

Line 184: 'In Open Targets and OnSIDES, 3.5 % and 3.2% of gene-phecode pairs had a SE-GPS > 0.'

Manuscript changes: We added the following paragraph to the discussion

Line 364: 'Although only 3.6% and 3.2% of gene-phecode pairs in Open Targets and OnSIDES have a SE-GPS > 0, this still reflects 11,620 and 9,416 gene-phecode pairs associated with a 1.8- and 1.9-fold increased risk of side effects. This subset of gene-phecodes with SE-GPS > 0 highlights how genetic evidence can point to biologically relevant mechanisms underlying on-target adverse effects and provides a starting point for deeper phenotypic profiling. We expect this proportion to increase as GWA and rare variant evidence continues to expand. However, the absence of a SE-GPS should not be interpreted as evidence for the absence of a side effect, and additional functional and experimental validation is required when modulating a particular target.'

Reviewer #3 (Remarks to the Author):

The authors have addressed my concerns in the response to reviewer comments and I am satisfied by their additions to the manuscript. I have been able to run the code provided and have reproduced their results, but I suggest the authors add additional comments to the code as it was not always clear to me what each chunk was doing and why this was necessary. Overall the manuscript is improved.

Reviewer #3 (Remarks on code availability):

Authors reply: We thank the reviewer for taking the time to go through our response. We have updated the GitHub repository with additional comments throughout each script to enhance clarity. Furthemore, to help streamline the code and improve readability, we have created a separate script

‘Plot_functions.R’, which consolidates plotting functions that were previously repeated across the analysis code.