Human Genetics (2023) 142:749–758
https://doi.org/10.1007/s00439-023-02548-y

ORIGINAL INVESTIGATION

Predicting ExWAS findings from GWAS data: a shorter path to causal genes

Kevin Y. H. Liang[1,2] · Yossi Farjoun[1,7,8,9] · Vincenzo Forgetta[1,7] · Yiheng Chen[1,3] · Satoshi Yoshiji[1,3,10,11] · Tianyuan Lu[1,2,7] · J. Brent Richards[1,2,3,4,5,6,7]

Received: 2 November 2022 / Accepted: 22 March 2023 / Published online: 2 April 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

GWAS has identified thousands of loci associated with disease, yet the causal genes within these loci remain largely unknown. Identifying these causal genes would enable deeper understanding of the disease and assist in genetics-based drug development. Exome-wide association studies (ExWAS) are more expensive but can pinpoint causal genes offering high-yield drug targets, yet suffer from a high false-negative rate. Several algorithms have been developed to prioritize genes at GWAS loci, such as the Effector Index (Ei), Locus-2-Gene (L2G), Polygenic Prioritization score (PoPs), and Activity-by-Contact score (ABC) and it is not known if these algorithms can predict ExWAS findings from GWAS data. However, if this were the case, thousands of associated GWAS loci could potentially be resolved to causal genes. Here, we quantified the performance of these algorithms by evaluating their ability to identify ExWAS significant genes for nine traits. We found that Ei, L2G, and PoPs can identify ExWAS significant genes with high areas under the precision recall curve (Ei: 0.52, L2G: 0.37, PoPs: 0.18, ABC: 0.14). Furthermore, we found that for every unit increase in the normalized scores, there was an associated 1.3–4.6-fold increase in the odds of a gene reaching exome-wide significance (Ei: 4.6, L2G: 2.5, PoPs: 2.1, ABC: 1.3). Overall, we found that Ei, L2G, and PoPs can anticipate ExWAS findings from widely available GWAS results. These techniques are therefore promising when well-powered ExWAS data are not readily available and can be used to anticipate ExWAS findings, allowing for prioritization of genes at GWAS loci.

✉ J. Brent Richards
brent.richards@mcgill.ca

1 Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, QC H3T 1E2, Canada

2 Quantitative Life Sciences Program, McGill University, Montréal, QC H3A 0G4, Canada

3 Department of Human Genetics, McGill University, Montréal, QC H3A 0G4, Canada

4 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC H3A 0G4, Canada

5 Department of Medicine, McGill University, Montréal, QC H3A 0G4, Canada

6 Department of Twin Research, King's College London, London, UK

7 5 Prime Sciences Incorporated, Montréal, Canada

8 Broad Institute, Cambridge, MA 02142, USA

9 Fulcrum Genomics LLC, Boulder, CO 80302, USA

10 Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

11 Japan Society for the Promotion of Science, Tokyo, Japan

Introduction

Drug development programs most often fail because of lack of efficacy (Paul et al. 2010; Seyhan 2019). This often occurs when the medicine targets a protein that is not involved in the etiology of the disease (Butcher 2003; Lindsay 2003; Seyhan 2019). One way to identify causal mechanisms of human disease is to use human genetics, since drug development programs supported by human genetic evidence have a two to four-fold increase in probability of success (Nelson et al. 2015; King et al. 2019; Ochoa et al. 2022). Genome-wide association studies (GWAS) have been fruitful in identifying the regions of the genome associated with susceptibility to disease. However, due to linkage disequilibrium (LD), loci that have significant variants can span multiple genes and therefore, GWAS results often do not clearly identify a specific causal gene at associated loci. Consequently, a major challenge in current human genetics studies is mapping GWAS loci to their causal genes.

This is a non-trivial task for three primary reasons. First, associated variants are often in LD with nearby variants

rendering resolution to a few causal variants difficult. Second, approximately 80–90% of all phenotype-associated variants are non-coding, and most have small effect sizes (Edwards et al. 2013; Mirza et al. 2014; Hrdlickova et al. 2014; Boyle et al. 2017). Thus, attributing their effects to specific genes is, even with experimental methods like CRISPR, challenging (Xu and Li 2020). Last, given the hundreds of loci associated with many common diseases, functional experimentation on a locus-by-locus basis is laborious and has rarely been achieved systematically.

One method to identify causal genes with less ambiguity is through exome-wide association studies (ExWAS) as this method relies upon coding genetic variants that may have high pathogenicity. As this technique focuses on coding variants, it provides direct evidence for causal gene inference. It can also provide information regarding the directionality and magnitude of the effects of genes on the phenotype of interest, which can be difficult to infer based on GWAS and gene prioritization metrics alone. However, a major limitation of ExWAS is that the statistical power is low due to the rarity of pathogenic coding mutations in general and consequently the false negative rate of ExWAS is high (Auer and Lettre 2015). This means that a much larger sample size is required to have sufficient variants across the exome to detect gene–trait associations. For example, an ExWAS for estimated bone mineral density (eBMD), a known polygenic trait, with close to 300,000 samples only identified 19 genes. Yet, many known causal genes were not identified amongst these 19 genes (unpublished data—accepted in principle). In comparison, GWAS with similar sample sizes have identified more than 500 independent GWAS loci (Kemp et al. 2017; Morris et al. 2019). This suggest that there are many known gene–trait associations missed in the ExWAS. This is likely due to the lack of statistical power and the lack of variants at some of the causal genes. Further, due to increased costs, ExWAS data is available often for a much smaller sample size than that in GWAS.

In comparison to ExWAS, GWAS results are much easier to obtain at large scale. For this reason, there is ongoing research into developing methods for improving the use of this abundant data-type for the purpose of drug target discovery. These methods aims to prioritize causal genes by incorporating LD information as well as easily obtainable genomic annotations with GWAS results (Weeks et al. 2020; Mountjoy et al. 2021; Forgetta et al. 2022). To date, the performance of these metrics when compared to the more direct (though more expensive) ExWAS results have not been evaluated systematically.

Here we evaluate the performance of four recently published gene prioritization metrics (Effector index [Ei] (Forgetta et al. 2022), Polygenic priority scores [PoPs] (Weeks et al. 2020), Locus-2-gene [L2G] (Ghoussaini et al. 2021), and Activity-by-contact [ABC] (Fulco et al. 2019)) in predicting ExWAS results in nine traits (calcium level, direct bilirubin level [Bilirubin], estimated bone mineral density [eBMD], standing height [Height], hypothyroidism, low density lipoprotein cholesterol level [LDL-C], red blood cell count [RBC], triglyceride level, and type 2 diabetes [T2D]). If ExWAS data can be predicted from GWAS data based on these metrics, it can potentially allow for the interrogation of causal genes at hundreds of GWAS loci and accelerate the delivery of medicines to the clinics.

Materials and methods

Trait selections

We chose nine traits for which there were large sample sizes in UK Biobank and also spanned a broad spectrum of disease pathology (Table S1). These included seven continuous traits (calcium level, direct bilirubin measurements, estimated bone mineral density, standing height, LDL-cholesterol level, red blood cell count, and triglyceride level) and two common disease traits (type 2 diabetes and hypothyroidism).

Sources of ExWAS results

Masks are defined as different variant inclusion criteria for burden testing (a list of all burden tests and masks are in Table S2). ExWAS results for all traits except eBMD were obtained from AZphewas (Wang et al. 2021) or GeneBass (Karczewski et al. 2022). For eBMD, ExWAS results were obtained from a disease specific ExWAS study (unpublished data – accepted in principle) (Table S1).The lowest ExWAS burden test p value across all ExWAS sources, burden tests and masks (or when specified, across a previously defined set of 'stringent masks' (Table S3)) were retained for each trait. Gene information was merged across datasets by matching gene names.

GWAS loci definition

GWAS results were obtained from publicly available resources or large-scale trait specific GWAS publications (Mahajan et al. 2018; Morris et al. 2019; Forgetta et al. 2022) (Table S4). GWAS loci were defined as previously described (Forgetta et al. 2022). Briefly, a set of independent SNPs were obtained through clumping using plink 1.9 (Purcell et al. 2007) using a window size of 250 kb, linkage disequilibrium threshold of 0.01 and a p value threshold of 5×10^{-8}. Neighbouring signals within 50 kb were merged. The resulting loci were then padded with 250Kbp on each

side resulting in loci with at least 500 kb in size. All genes were assigned to one of these loci if it overlaps the locus by more than 50% of its length. These genes and GWAS loci formed the basis of all locus-based analyses (i.e., AUPRC, AUROC, odds ratio, score distribution) throughout the study.

Sources of Ei scores

Ei scores for all traits were obtained from a recent publication (Forgetta et al. 2022). The goal of the Ei is to generate a score to quantify the probability of causality for each gene at each significant GWAS locus (Forgetta et al. 2022). Briefly, there were a total of twelve phenotypes used to develop the Ei model (T2D, LDL-C, height, calcium level, hypothyroidism, triglyceride level, eBMD, glucose level, red blood cell count, systolic blood pressure, diastolic blood pressure and direct bilirubin level). GWAS for each trait were fine-mapped then annotated to identify locus and gene-level features for use in the model. The Ei models were generated using a gradient boosting algorithm (XGBoost) with over 150 locus-, and variant-level annotations. Two measures were taken in the development of Ei to avoid potential overfitting (Forgetta et al. 2022). First, the Ei model for each trait was generated separately using a leave-one-out approach (e.g., the Ei model for T2D was generated by training the model on all traits, except for T2D). Second, to further avoid overfitting, the creators of Ei ensured that each gene contributed to the training of the models at most once.

Sources of L2G scores

L2G scores were obtained from their public online platform (https://genetics.opentargets.org/) (Table S5). L2G scores have been calculated for a number of GWASs published on GWAS catalog (https://www.ebi.ac.uk/gwas/). For this study, we chose L2G scores calculated based on GWASs that utilized UK Biobank data (Table S5). For each study, a list of significant variants and L2G scores for each associated gene was obtained. In cases where the same genes were implicated by multiple lead variants, the largest L2G score was retained. L2G is similar to Ei in that it also uses a gradient boosting algorithm (XGBoost) and assigns a value between zero and one; however, it differs from Ei in the features used to train the model (Carvalho-Silva et al. 2019; Ghoussaini et al. 2021; Mountjoy et al. 2021).

Sources of PoPs and ABC score

PoPs and ABC scores were obtained from a recent publication (Weeks et al. 2020) (Table S5). PoPs works under the assumption that causal genes share similar functional characteristics. As such, it aims to identify potential causal gene by analyzing a number of biological annotations ranging from gene expression data to biological pathways and protein–protein interaction information to identify the most probable causal gene (Weeks et al. 2020). ABC was developed to identify gene-enhancer relationship. As many GWAS significant variants are non-coding variants that influence the trait through interactions with gene regulatory elements, ABC aims to prioritize the most likely causal gene by identifying genes that are most likely to be regulated by enhancer elements linked to the lead fine-mapped variant (Fulco et al. 2019; Nasser et al. 2021). For each trait, the largest PoPs and ABC score for each gene was retained.

Evaluation of classification accuracy

Two metrics commonly used to evaluate the performance of binary classifiers are area under the receiver-operator curves (AUROC) and precision-recall curves (AUPRC). For each score, PRC and ROC were based on GWAS loci that contain at least one ExWAS significant gene. ExWAS significance status was defined in one of the following 2 ways:

1. Whether the lowest p values using variants from one of the previously defined set of 'stringent masks' is below the exome-wide significance threshold of 3.6×10^{-7}
2. Whether lowest p values using variants from any masks is below the exome-wide significance threshold of 3.6×10^{-7}

Since it is known that a large number of causal genes will be missed in ExWAS due to reduced power, we focused on genes that are located among GWAS loci with known ExWAS significant genes as to avoid the appearance of an elevated false-positive rate. We computed the AUROC and AUPRC for each of the nine traits separately and combined. This resulted in 34 tests when ExWAS significance was defined using variants from the set of stringent masks and 35 tests when ExWAS significance was defined using variants from all masks. A 99.9% confidence interval were reported which will maintain family-wise error rate of 0.05 (Table S6, S7). AUROC and AUPRC confidence intervals were determined using 2000 bootstrap replicates. Baseline for AUROC is that of an uninformative classifier (i.e., 0.50). Baseline for AUPRC is that of a random classifier, which corresponds to the proportion of ExWAS significant genes among all genes that have both an ExWAS burden test p value and the relevant gene prioritization score. As the baseline for AUPRC is based on the proportion of target genes among all relevant genes, it is specific for each trait and score.

Causal and drug target gene definitions

Causal and drug target genes (Table S8) were obtained from a prior study (Forgetta et al. 2022). Briefly, causal genes were identified by manual inspection of the Human Disease Ontology database (Schriml et al. 2019) for relevant ontological terms by clinician scientists. The relevant OMIM linkage information was then used to obtain a list of relevant genes associated with the diseases. Drug target genes were identified by first identifying a list of guideline-approved medications from UpToDate (https://www.uptodate.com/contents/search) by clinician scientists. This information is then linked to DrugBank (Wishart et al. 2018) to identify a list of drug targets.

Results

High scoring genes across all metrics are enriched in ExWAS significant genes

To quantify the performance of the four metrics, we chose nine phenotypes (T2D, Bilirubin, eBMD, height, hypothyroidism, LDL-C, RBC, calcium level, and triglyceride level) for which there were large sample sizes in UK Biobank, since the largest sets of ExWAS data are generally derived from this resource. Further, these diseases and traits represent metabolism, ageing, autoimmune disease, anthropometric, hematologic and lipid outcomes, providing a broad spectrum of pathogenic mechanisms for study. GWAS and ExWAS results were obtained from publicly available resources or recent publications (see Methods).

We first evaluated the difference in the distributions of the four prioritization scores (Ei, PoPs, L2G and ABC) between ExWAS significant and non-significant genes. ExWAS significant genes were defined as those having a p value below the exome-wide significance threshold of 3.6×10^{-7} for any of the predefined stringent set of variants using burden testing (Table S3 and S9). A burden test assesses whether a specific category of variants in a particular gene is more common among the cases than the controls (Curtis 2019). Different categories or combinations of categories (e.g., predicted loss of function (pLOF), pLOF + missense, synonymous, etc.), commonly referred to as "masks", are tested. It is also common to calculate different variations of rare variant association tests (i.e., SKAT, SKAT-O, standard gene-based burden tests) that differ in their assumptions of the effect of the variants on gene functions (Lee et al. 2012). This approach of gene-based collapsing test where the effects of multiple variants are aggregated is frequently done in ExWAS studies as sample sizes at the current biobank scale are often not large enough to reliably describe

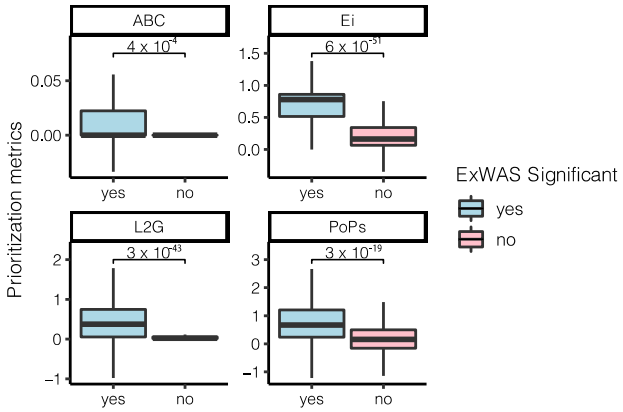


Fig. 1 Distributions of the untransformed prioritization metric between ExWAS significant and non-significant genes for Ei, ABC, L2G, and PoPs together with the p values for the differences between these groups are shown. The differences in score distribution between the two categories were determined using the Mann-Whitney test using all data points; however, outliers (defined as values outside the 1.5 * interquartile range) are removed for clarity of presentation. Only genes residing at GWAS loci which harboured ExWAS significant genes were considered. Only the lowest p value from burden tests based on one of the previously defined 'stringent masks' (Table S3) were considered for determining ExWAS significance status

the effect of individual rare coding variants (Curtis 2019). For all four metrics, the distribution of the untransformed values was higher in the ExWAS significant gene category than the non-significant category (Figs. 1, S1). The range of possible values, and therefore distributions, is different between the metrics. For example, Ei and L2G are bounded between 0 and 1, whereas the others are not. Despite the difference in value distributions, for each metric, ExWAS significant genes tended to have higher scores than ExWAS non-significant genes.

We then evaluated the difference in p value distribution between different gene prioritization metric percentiles. In all cases, the p values are significantly lower in the top percentiles compared to the bottom percentiles (Fig. 2). We also observed that the ExWAS p values became lower across increasing quintiles of the gene prioritization metrics (Fig. S2). Taken together, this suggest that the four gene prioritization metrics tend to assign higher scores to genes that tend to have lower ExWAS p values.

Gene prioritization metrics scores are positively correlated with ExWAS significance

Next, we estimated the odds that a gene will reach the ExWAS significance threshold per unit increase in each gene prioritization metric. To do so, we assessed only genes that resided at GWAS loci, which also harboured at least one ExWAS significant gene. To do this, we normalized each prioritization score (i.e., scaled to have unit

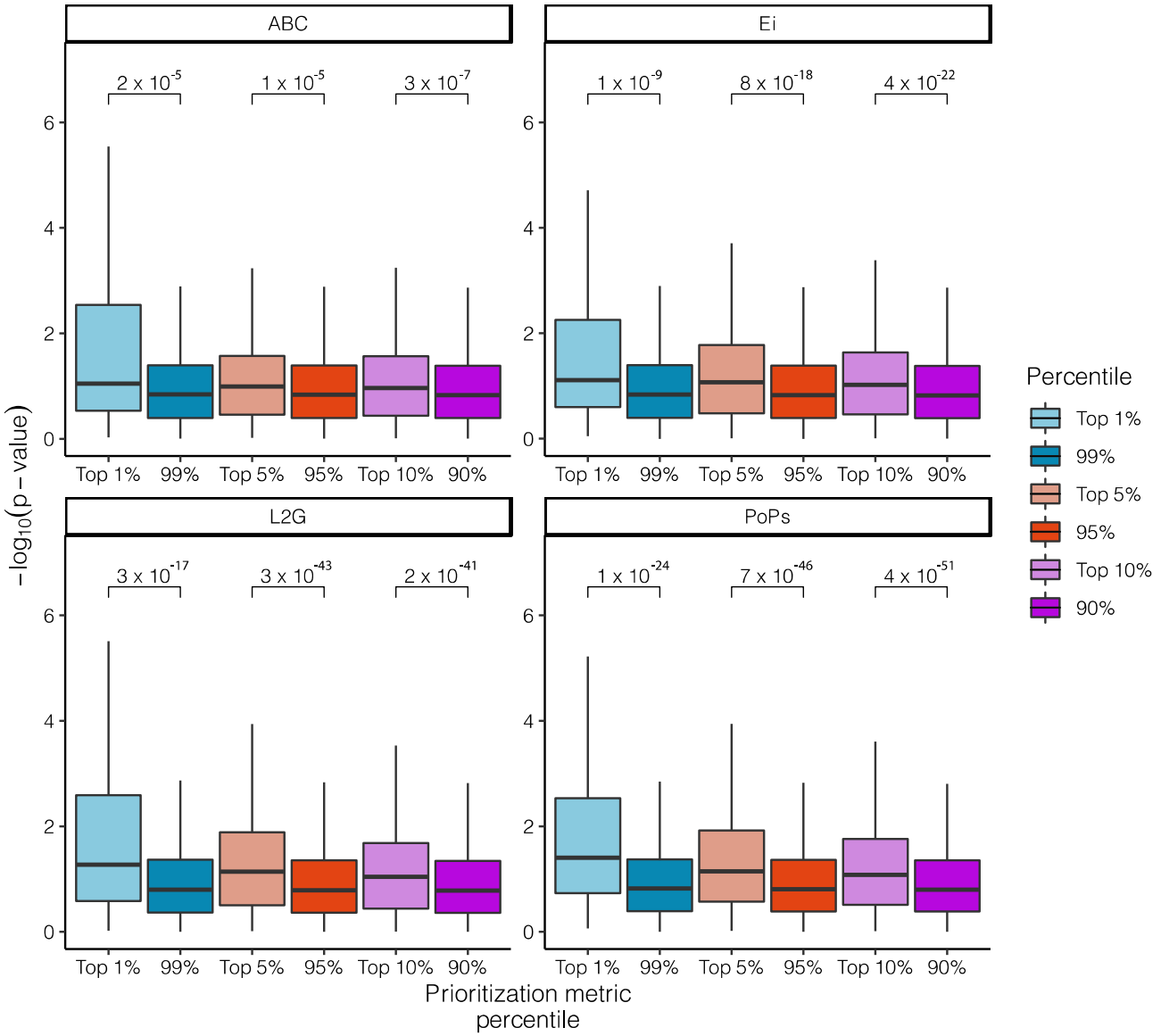


Fig. 2 Distribution of ExWAS burden test *p* values between the prioritization metric percentiles. The top 1% vs 99% (blue), 5% vs 95% (orange), and 10% vs 90% (purple) are shown. Difference in ExWAS *p* value distribution between the percentiles was determined using the Mann-Whitney test with all data points and the *p* values are shown. Outliers (defined as values outside the 1.5 * interquartile range) are omitted for clarity of presentation. Only *p* values from burden tests based on one of the previously defined 'stringent masks' (Table S3) are considered

variance and zero mean), and fitted a logistic regression model using each of the normalized prioritization scores as the predictor and ExWAS status as the outcome. Overall, we tested the associations of each score in each phenotype separately and all phenotypes combined resulting in 29 associations. Significance status was determined based on the 99.8% confidence intervals, which corresponds to the Bonferroni corrected type I error of 0.002 (i.e., 0.05/29 associations (Table S10)). Using this Bonferroni corrected confidence interval will maintain the family-wise error rate of 0.05 (Dunn 1961). We observed that all scores have odds ratio significantly higher than the null (Fig. 3). The largest odds was achieved by the Ei where a one unit increase in the scaled Ei score corresponds to a 4.6-fold (99.8% confidence interval (99.8% CI) 3.39–6.34) increase in the odds of a gene reaching exome-wide significance. The lowest odds (though still significantly greater than one) was for ABC where we observed an OR of 1.3 (95% CI 1.09–1.6).

The same trend was obtained when ExWAS significance status was determined using the lowest burden tests *p* values across all masks rather than just the tests with the most

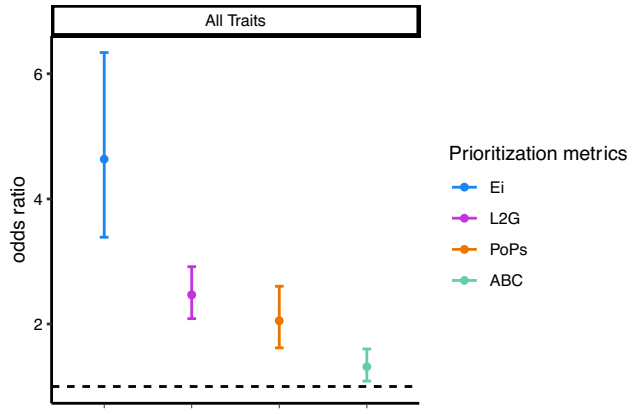
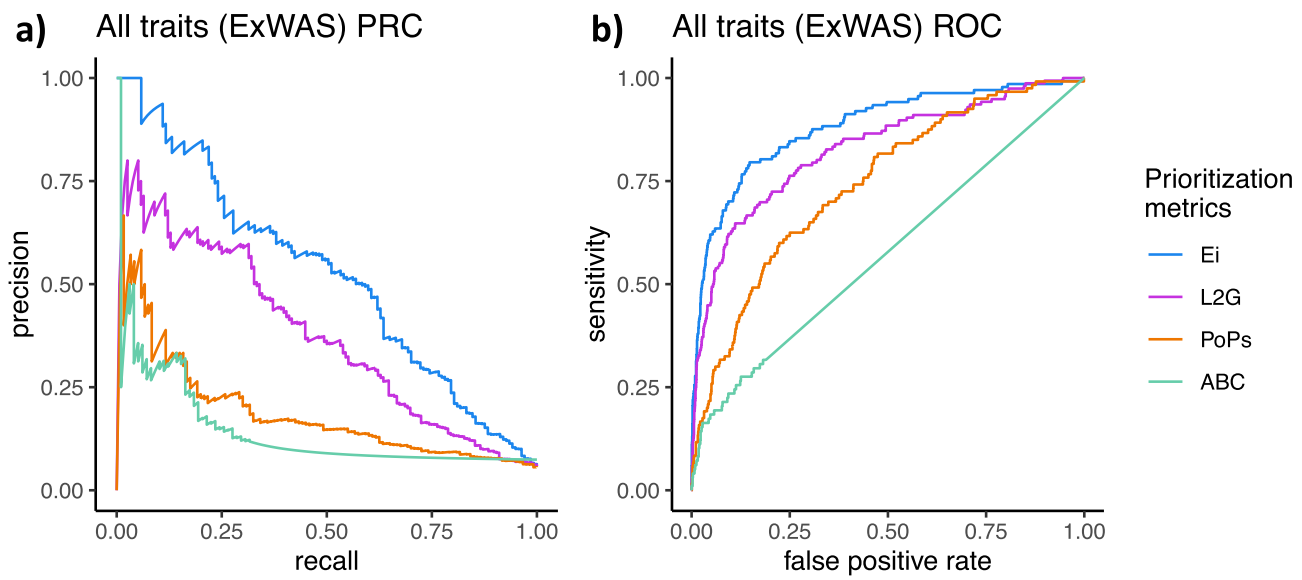


Fig. 3 Odds ratio obtained from logistic regression using ExWAS significant status as the outcome and each normalized prioritization score as a predictor. This was assessed for only genes found among GWAS loci which harboured ExWAS significant genes. Data were combined across all traits. The dotted line represents no effect (i.e., odds ratio of 1). 99.8% confidence interval (corresponding to a Bonferroni corrected type I error of 0.002 = 0.05/29 tests) is shown. Only *p* values from burden tests based on one of the previously defined 'stringent masks' (Table S3) are considered for determining ExWAS significance

stringent masks as defined above (Table S3, S11). Ei and L2G had the largest odds per unit increase in the scaled score followed by PoPs and ABC; however, the odds for all scores decreased (Table S11).

Gene prioritization metrics can accurately classify ExWAS significant genes

Next, we evaluated the ability of each metric to classify ExWAS significance status for each gene based on the same definition as above. We focused only on genes that were found at GWAS loci which also harboured an ExWAS significant gene. The ExWAS significant gene was denoted as the positive control and we calculated the area under the receiver operator curve (AUROC) and area under the precision recall curve (AUPRC) for each gene prioritization metric. Significance of AUROC and AUPRC was evaluated using the 99.9% confidence interval based on 2000 bootstrap replicates. The AUPRC was the highest for Ei followed by L2G, PoPs and ABC (Fig. 4A), with the exception of ABC, all were significantly higher than baseline based on the 99.9% confidence interval (Table S6). Similar trend is seen for AUROC. 99.9% CI of AUROC for all but ABC were



	Ei	L2G	ABC	PoPs
AUROC	**0.88 (0.83 - 0.94)**	0.83 (0.76 - 0.88)	0.57 (0.5 - 0.67)	0.74 (0.67 - 0.82)
AUPRC	**0.52 (0.37 - 0.64)**	0.37 (0.25 - 0.52)	0.14 (0.07 - 0.26)	0.18 (0.11 - 0.32)

Fig. 4 Performance curve for all four gene prioritization metrics. Precision–recall curve (PRC) **a** and receiver operator curve (ROC) **b** are shown for all four metrics. Area under the curve (AUC) for each metric is shown with the largest AUC in bold. 99.9% confidence intervals were evaluated using 2000 bootstrap replicates. Target genes were defined as genes with ExWAS burden test *p* value less than the Exome-wide significance threshold of 3.6×10^{-7}. Only *p* values from burden tests based on one of the previously defined 'stringent masks' (Table S3) are considered for determining ExWAS significance

above the baseline of 0.50, which corresponds to an AUC of an uninformative classifier (Fig. 4B). The best AUROCs were achieved using Ei and L2G (AUROC: 0.88 (99.9% CI 0.83–0.94), 0.83 (0.76–0.88), respectively). We note the AUPRC is more informative than the AUROC because there are many fewer positive control genes than other genes at the GWAS loci. The performance across all metrics decreased when ExWAS significance status was determined based on the lowest *p* value across all masks, much like what was observed above; however, relative performance of each metric remained similar (Fig. S3, Table S7).

We also noted that the ExWAS significant genes that are also known drug targets or are causal for the disease were almost always among the 10 highest scoring Ei and L2G genes within their corresponding GWAS locus (24/26 and 24/26, respectively) (Table S12). This suggest that it is possible to further rank ExWAS significant genes using Ei and L2G to obtain the most probable causal genes.

Discussion

Identifying drug targets with human genetic evidence will likely accelerate drug development programs. While ExWAS is able to precisely identify genes and sometimes provide direction of effect, it suffers a high false negative rate due to low statistical power. GWAS, on the other hand, has a high sensitivity to associate loci harbouring causal genes, but resolving associated loci to such genes is non-trivial. Here we have tested whether algorithms designed to prioritize genes at GWAS loci can predict ExWAS findings. Overall, these metrics perform reasonably well, but with a high degree of variability, and thus can be used to prioritize genes at GWAS loci that do not harbour an ExWAS significant gene, thus increasing the number of targets to accelerate genetics-based drug development.

All metrics have good performance with area under the receiver operator curve (AUROC) and area under the precision recall curve (AUPRC) reaching upwards of 0.88 and 0.52, respectively, for Ei. Across all traits, we have found that Ei, PoPs, and L2G have better performance than ABC in their ability to identify likely ExWAS hits. We found that all gene prioritization metrics can significantly enrich for ExWAS significant genes and are positively correlated with ExWAS *p* values where a 1 unit increase in the normalized scores were associated with a 1.3–4.6-fold increase in odds of the gene reaching exome-wide significance.

ABC was originally developed for identifying enhancer-gene relationships (Fulco et al. 2019) and thus may not be considered to be optimized to identify causal genes at GWAS loci. However, as it is well established that many non-coding variants can influence phenotypes through interactions with regulatory elements (Stranger et al. 2007; Nicolae et al. 2010; Zhang and Lupski 2015), it is expected that ABC is capable of identifying causal genes for which a gene-enhancer relationship exists, consistent with prior studies (Weeks et al. 2020; Nasser et al. 2021). PoPs, Ei and L2G were all developed with the goal of identifying likely causal genes and they do so by aggregating information from different sources. As a result, the difference in performance between these metrics are influenced, at least in part, by the features used in their model. For example, Ei used 154 features whereas L2G used 57. Ei focused more on locus and variant annotations such as GWAS *p* values, and variant functional annotations whereas L2G has a stronger focus on functional data such as pQTL, and eQTL colocalization results. Both metrics also shared several features such as distance to transcriptome start sites, number of genes within loci, and DNAse hypersensitivity sites. The difference between Ei and L2G scores is then likely due to the inclusion of data such as colocalization results involving pQTL and eQTL information. PoPs differs from both Ei and L2G in that it uses other omic measurements, such as protein and transcript levels to predict gene–trait associations defined by MAGMA (de Leeuw et al. 2015) results.

It is also worth noting that across all metrics evaluated, when the most relaxed ExWAS masks were used, performance decreased across all gene prioritization metrics. This may be because when such relaxed masks were used more variants are considered, this can result in noisier signals. Genes that do reach the significance threshold using these relaxed masks may potentially be less likely to be causal. This trade-off between signal and noise has previously been shown for ExWAS results where the burden tests that were based on masks that included additional filtering of missense variants through the use of missense intolerance score led to increased power (Wang et al. 2021).

We found that Ei and L2G consistently classify known causal genes and/or drug targets as one of the top 10 highest ranking genes within their respective loci. This suggest that it is possible to apply these GWAS-based metrics to ExWAS results and further rank ExWAS significant genes to prioritize the most likely causal gene. This is consistent with a recent study where it has shown that around 75% of rare coding variants identified via ExWAS are located within 1 megabase of GWAS lead variants and these two types of variants act independently on the trait (Backman et al. 2021). This underscores the benefit of a multifaceted approach to drug target identifications where multiple lines of evidence can be used to further reduce the search space for causal genes and identify the most probable candidate gene for downstream analysis.

This study has limitations. The performance of the metrics evaluated here are limited to the nine traits that were evaluated. PoPs had one trait missing (Bilirubin) while ABC

missed two (Bilirubin and T2D). As such, the evaluation for these two scores were limited to the remaining traits. However, the trend observed in performance remains the same when looking at individual trait analyses (Figs. S4–S6). In addition, the metrics evaluated in this study were not originally trained using ExWAS results as target genes, rather they were trained based on either heritability-based metrics, known drug targets, or known gene prioritization metrics such as distance-to-closest gene. As such, performance of all metrics will likely improve by re-training the model using ExWAS significance genes as target genes. Unlike L2G, PoPs, and ABC, the traits evaluated here were also used in the development of the Ei; however, the Ei models were trained using a leave-one-out approach to avoid overfitting. Although this approach would limit the scalability of Ei as individual models were required for each trait, it is an effective method for removing data overlap between training and testing data and should allow for an accurate assessment of performance. Due to the limited number of traits used in training the Ei models, another potential source of overfitting is genetic correlations between traits. However, LD score regression (Bulik-Sullivan et al. 2015) showed that most genetic correlations between traits are insignificant if not low (genetic correlation < 0.3) (Table S13). Furthermore, the performance of Ei for eBMD, which has low genetic correlations with all traits, is among one of the highest (Table S6, S7). Taken together, this suggests that the bias in the Ei models due to genetic correlations is low if any. Recently, another novel disease gene prioritization metric named combined SNP-to-gene (cS2G) was proposed (Gazal et al. 2022). cS2G prioritizes causal genes through the linear combination of individual SNP-to-gene strategies (e.g., eQTL *p* values, promoter information, etc.). However, as this method was evaluated using a different definition of precision and recall and the data are restricted to a specific subset of genes that satisfied a specific cS2G threshold (cS2G > 0.5) (Gazal et al. 2022), it cannot be easily compared to the other methods in this study (i.e., Ei, L2G, PoPs, and ABC) using the same evaluation metrics and is, therefore, not included.

It has been shown that the combination of different gene prioritization metrics that utilizes different features in their model in a consensus or linear combination approach allows for improved performance (Weeks et al. 2020; Gazal et al. 2022). Whether more sophisticated machine learning approaches can more effectively aggregate the information in all four metrics remains to be investigated once a larger set of traits with ExWAS, GWAS, Ei, PoPs, L2G and ABC results are available.

In conclusion, this study demonstrates that gene prioritization metrics based on GWAS results can identify ExWAS significant genes with reasonable accuracy highlighting another potential use of the abundance of GWAS data currently available. This suggests that for traits where well-powered ExWAS is not currently feasible, these gene prioritization metrics can serve as promising alternatives to identify genes that would have been found through a well-powered ExWAS. It is also possible to further rank ExWAS significant genes using these prioritization metrics to identify the most likely causal genes among all ExWAS significant genes. Furthermore, such methods could also be applied to diseases with existing ExWAS to help identify causal genes at GWAS loci where power was insufficient in the ExWAS. Because these methods and ExWAS aim to address the same issue (i.e., identify causal genes) using orthogonal data types, they can serve as an independent source of evidence allowing for more confident identification of causal genes to accelerate the delivery of medicines to the clinic.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-023-02548-y.

Author contributions Conception and design: KL and JBR. Data analyses: KL, YF, and VF. Manuscript writing: KL, YF, YC, SY, and JBR. Supervision: JBR. Interpretation of data: all authors. All authors were involved in the preparation and revision of the manuscript.

Funding The Richards research group is supported by the Canadian Institutes of Health Research (CIHR: 365825; 409511, 100558, 169303), the McGill Interdisciplinary Initiative in Infection and Immunity (MI4), the Lady Davis Institute of the Jewish General Hospital, the Jewish General Hospital Foundation, the Canadian Foundation for Innovation, the NIH Foundation, Cancer Research UK, Genome Québec, the Public Health Agency of Canada, McGill University, Cancer Research UK [grant umber C18281/A29019] and the Fonds de Recherche Québec Santé (FRQS). JBR is supported by a FRQS Mérite Clinical Research Scholarship. Support from Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the Welcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. These funding agencies had no role in the design, implementation or interpretation of this study.

Data availability Source code can be accessed through Github upon publication.

Declarations

Conflict of interest JBR's institution has received investigator-initiated grant funding from Eli Lilly, GlaxoSmithKline and Biogen for projects unrelated to this research. JBR is the CEO of 5 Prime Sciences (www.5primesciences.com), which provides research services for biotech, pharma and venture capital companies for projects unrelated to this research. VF, YF, and TL are employees of 5 Prime Sciences. Authors KYHL, YC, SY declares that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This article does not contain any studies with human participants.

References

Auer PL, Lettre G (2015) Rare variant association studies: considerations, challenges and opportunities. Genome Med 7:16. https://doi.org/10.1186/s13073-015-0138-2

Backman JD, Li AH, Marcketta A et al (2021) Exome sequencing and analysis of 454,787 UK Biobank participants. Nature. https://doi.org/10.1038/s41586-021-04103-z

Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. Cell 169:1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bulik-Sullivan BK, Loh P-R, Finucane HK et al (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295. https://doi.org/10.1038/ng.3211

Butcher SP (2003) Target discovery and validation in the post-genomic era. Neurochem Res 28:367–371. https://doi.org/10.1023/A:1022349805831

Carvalho-Silva D, Pierleoni A, Pignatelli M et al (2019) Open targets platform: new developments and updates two years on. Nucleic Acids Res 47:D1056–D1065. https://doi.org/10.1093/nar/gky1133

Curtis D (2019) A weighted burden test using logistic regression for integrated analysis of sequence variants, copy number variants and polygenic risk score. Eur J Hum Genet 27:114–124. https://doi.org/10.1038/s41431-018-0272-6

de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015) MAGMA: generalized gene-set analysis of GWAS data. PLOS Comput Biol 11:4219. https://doi.org/10.1371/journal.pcbi.1004219

Dunn OJ (1961) Multiple comparisons among means. J Am Stat Assoc 56:52–64

Edwards SL, Beesley J, French JD, Dunning AM (2013) Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet 93:779–797. https://doi.org/10.1016/j.ajhg.2013.10.012

Forgetta V, Jiang L, Vulpescu NA et al (2022) An effector index to predict target genes at GWAS loci. Hum Genet. https://doi.org/10.1007/s00439-022-02434-z

Fulco CP, Nasser J, Jones TR et al (2019) Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat Genet 51:1664–1669. https://doi.org/10.1038/s41588-019-0538-0

Gazal S, Weissbrod O, Hormozdiari F et al (2022) Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. Nat Genet. https://doi.org/10.1038/s41588-022-01087-y

Ghoussaini M, Mountjoy E, Carmona M et al (2021) Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res 49:D1311–D1320. https://doi.org/10.1093/nar/gkaa840

Hrdlickova B, de Almeida RC, Borek Z, Withoff S (2014) Genetic variation in the non-coding genome: involvement of micro-RNAs and long non-coding RNAs in disease. Biochim Biophys Acta BBA 1842:1910–1922. https://doi.org/10.1016/j.bbadis.2014.03.011

Karczewski KJ, Solomonson M, Chao KR et al (2022) Systematic single-variant and gene-based association testing of thousands of phenotypes in 426,370 UK Biobank exomes. Medrxiv. https://doi.org/10.1101/2021.06.19.21259117

Kemp JP, Morris JA, Medina-Gomez C et al (2017) Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. Nat Genet 49:1468–1475. https://doi.org/10.1038/ng.3949

King EA, Davis JW, Degner JF (2019) Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLOS Genet 15:e1008489. https://doi.org/10.1371/journal.pgen.1008489

Lee S, Emond MJ, Bamshad MJ et al (2012) Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 91:224–237. https://doi.org/10.1016/j.ajhg.2012.06.007

Lindsay MA (2003) Target discovery. Nat Rev Drug Discov 2:831–838. https://doi.org/10.1038/nrd1202

Mahajan A, Taliun D, Thurner M et al (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50:1505–1513. https://doi.org/10.1038/s41588-018-0241-6

Mirza AH, Kaur S, Brorsson CA, Pociot F (2014) Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. PLoS ONE 9:e105723. https://doi.org/10.1371/journal.pone.0105723

Morris JA, Kemp JP, Youlten SE et al (2019) An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet 51:258–266. https://doi.org/10.1038/s41588-018-0302-x

Mountjoy E, Schmidt EM, Carmona M et al (2021) An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat Genet 53:1527–1533. https://doi.org/10.1038/s41588-021-00945-5

Nasser J, Bergman DT, Fulco CP et al (2021) Genome-wide enhancer maps link risk variants to disease genes. Nature 593:238–243. https://doi.org/10.1038/s41586-021-03446-x

Nelson MR, Tipney H, Painter JL et al (2015) The support of human genetic evidence for approved drug indications. Nat Genet 47:856–860. https://doi.org/10.1038/ng.3314

Nicolae DL, Gamazon E, Zhang W et al (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 6:e1000888. https://doi.org/10.1371/journal.pgen.1000888

Ochoa D, Karim M, Ghoussaini M et al (2022) Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. Nat Rev Drug Discov. https://doi.org/10.1038/d41573-022-00120-3

Paul SM, Mytelka DS, Dunwiddie CT et al (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov 9:203–214. https://doi.org/10.1038/nrd3078

Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795

Schriml LM, Mitraka E, Munro J et al (2019) Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res. https://doi.org/10.1093/nar/gky1032

Seyhan AA (2019) Lost in translation: the valley of death across preclinical and clinical divide identification of problems and overcoming obstacles. Transl Med Commun. https://doi.org/10.1186/s41231-019-0050-7

Stranger BE, Nica AC, Forrest MS et al (2007) Population genomics of human gene expression. Nat Genet 39:1217–1224. https://doi.org/10.1038/ng2142

Wang Q, Dhindsa RS, Carss K et al (2021) Rare variant contribution to human disease in 281,104 UK Biobank exomes. Nature 597:527–532. https://doi.org/10.1038/s41586-021-03855-y

Weeks EM, Ulirsch JC, Cheng NY et al (2020) Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. MedRxiv. https://doi.org/10.1101/2020.09.08.20190561

Wishart DS, Feunang YD, Guo AC et al (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx1037

Xu Y, Li Z (2020) CRISPR-Cas systems: overview, innovations and applications in human disease research and gene therapy. Comput Struct Biotechnol J 18:2401–2415. https://doi.org/10.1016/j.csbj.2020.08.031

Zhang F, Lupski JR (2015) Non-coding genetic variants in human disease. Hum Mol Genet 24:R102–R110. https://doi.org/10.1093/hmg/ddv259

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.