
CLINICAL ADVANCEMENT FORECASTING *

Authors TBD
Related Sciences

ABSTRACT

This study examines the extent to which the outcomes of clinical trials can be predicted based on longitudinal properties of drug targets and diseases alone. We find that this is possible by comparing the historical performance of model-based target-disease pair prioritization methods to common baselines. Our primary objective is to demonstrate that statistical learning can effectively optimize such methods with no loss of interpretability. For example, non-negative linear models can produce simple weighting schemes across various types of human, animal and cell model evidence (for targets, diseases and pairings of the two) to identify target-disease pairs that advance beyond phase 2 trials with an average relative risk that is 2x higher than Open Targets composite scores. Other key characteristics of this study include: 1) a comprehensive longitudinal treatment of evidence as well as how it relates to leakage and reverse causality in biomedical research, 2) trial and/or drug details are not used in order to enable ranking for undeveloped targets/diseases, 3) analysis of the space of currently undeveloped, tractable targets with the highest likelihood of clinical success, 3) no data is used outside of Open Targets to ease reproduction and/or deployment, and 4) our method requires no expert knowledge and can easily support the inclusion of more lines of evidence over time, making it easy to operationalize.

1 Introduction

It has been well established that drugs with human genetic evidence linking their respective targets to indications in clinical trials are more likely to succeed [1, 2, 3, 4, 5, 6, 7, 8]. This information has been used to devise many target and target-disease ranking algorithms based primarily on a synthesis of multiple genetic signals alone [9, 10, 11]. It is also possible to expand the breadth of this genetic support to more targets/diseases based on knowledge graphs, protein interactions and/or disease ontologies [12, 13, 14, 15, 16]. To our knowledge, all such expansion methods identify a larger space of opportunities at the expense of expected success rates. This is not a focus of this work as we aim, instead, to establish a framework for identifying targets/diseases with the very highest possible likelihood of success first. This is accomplished by integrating human clinical, genetic/genomic, transcriptomic and proteomic data as well as cell/animal model evidence, pathway information and basic literature metrics from Open Targets [10] in a simple statistical modeling framework. It can be contrasted with far more integrative methods that rely on neural and/or graph models over extensive knowledge graphs [17, 18, 19, 20], which are more complex and difficult to interpret. We believe a desirable middle ground between these approaches and those that aim to combine many orthogonal indicators of success through expert knowledge in heuristic systems [21, 10] would: 1) permit inclusion of many types of evidence from many sources, 2) be highly interpretable, 3) support expert judgement where necessary, and 4) not require manual ranking/weighting schemes.

**Citation:* TBD

A substantial challenge inherent to building such a system is the need to account for the longitudinal nature of knowledge discovery in biomedicine. This is vital because any method that optimizes for likely future clinical success based on historical clinical success may easily be biased by the non-random nature with which evidence is absent otherwise. We use only "temporalized" evidence, i.e. evidence for which timing of its emergence can be determined, and outcomes when training and evaluating our methods before ultimately applying them to present-day evidence with no restrictions on timing. We discuss motivations, prior research and our own analysis on how important this problem is for each source in Section 2.9.

Addressing such problems is common, but not ubiquitous, in studies that attempt to predict clinical trial outcomes using temporalized predictors [22, 23, 24]. The need for this is often clear in that setting where the inclusion of predictors like historical success rates for targets/diseases, trial sponsor track records, eventual patient enrollment, etc. constitute clear information leaks otherwise. This is discussed more in [22] which notes several studies that do not account for this problem, and presents "quasi-prospective" as well as true "prospective" results. The difference between the two is that the former reconstructs timelines for predictors and outcomes based on recorded event dates while the latter relies on frozen predictions that are never evaluated until years later. Nomenclature for these formulations is conflicting though, where this definition of a "quasi-prospective" design is deemed entirely prospective in some cases, e.g. [25]. We will refer to our design in this study as quasi-prospective since this definition is the best fit.

The prior works discussed so far can largely be categorized as either 1) target and target-disease prioritization methods evaluated based on how well they correlate with clinical trial success and 2) clinical trial outcome prediction models. Both are measured against the same outcomes and an important distinction between them lies in how the former methods are **not** directly optimized for those outcomes while the latter methods are. In this study, we attempt to bridge these methodologies by predicting clinical trial advancement for target-disease pairs based solely on information that would be present well in advance of any drug program or individual trial. We then calibrate these predictions to determine what thresholds are necessary to match the observed success rates from benchmarks for genetic support like OMIM [26], ClinVar [27] and GWAS. Finally, we examine how many present-day target-disease pairs are undeveloped (i.e. have never been in clinical trials), have a tractable target and are likely to see success rates matching or exceeding those calibrated benchmarks.

2 Results

In order to model clinical advancement for target-disease (TD) pairs, we first define "advancement" as progression beyond any particular trial phase across all drugs associated with any one TD pair as indicated by the presence of a later-stage trial. All results to follow consider only advancement beyond phase 2 due to limitations described in Section 4. This binary outcome is then predicted based on a list of features shown in Supplementary Table 1. Information for each of these features is only used when it was published before the year **prior** to the first phase 2 trial observed, with an exception for genetic evidence discussed in Section 2.9. A training dataset is then formed by including only TD pairs where this first phase 2 year is between 1990 and 2015. The evaluation dataset then consists of all TD pairs entering phase 2 between 2016 and 2022, with a 2 year offset from the present year (2024) to allow enough time for some trials to complete. While the average phase 2 trial duration may be as low as 2 years [28], other estimates would suggest half of them take longer than 2.9 years [29]. This means a substantial fraction of outcomes are censored, that this is an important parameter to test sensitivity to and that time itself is likely to be a crucial covariate in this formulation. The distribution of these outcomes, the number of associated targets/diseases and a variety of other statistics on this dataset are presented in Supplementary Figure 10.

2.1 Features

The predictors used consist of 27 target-disease pair features, 5 target-specific features and 1 disease-specific feature. These are listed in Supplementary Table 1. The target and disease specific features are chosen carefully such that they are either capable of being associated with years in which events supporting them occurred or result from large-scale, unbiased methods that do not favor well-studied or drugged targets/diseases. Examples of this include target-specific

tissue expression specificity scores computed from Human Protein Atlas [30] and LOEUF [31] scores from gnomAD. Simply put, our dataset combines scores from Open Targets for target-disease evidence and a select subset of target prioritisation [32] fields with almost no modifications, other than to add target and disease specific indicators of maximum trial phases reached and two extra genetic association features.

2.2 Models

We train a variety of models including constrained and unconstrained linear and tree models. The constrained variants of these models force effects of all features to increase monotonically. This is possible with no feature transformations because all scores in Open Targets are constructed such that higher scores are presumed to be advantageous. We also apply these models to our evaluation dataset using several feature ablations in order to assess the value of groups of related features.

In order to compare these models to an Open Targets composite score, we use an equally weighted sum of all scores except for those assigned lower weights in [33]. Scores from these sources are multiplied by the corresponding weight before being summed.

2.3 Performance

Figure 1 demonstrates how well our primary model in this study – a constrained, linear L2-regularized model referred to from here on as "RDG" (short for ridge regression) – ranks TD pairs by comparison to a composite score from Open Targets ("OTS"). The relative risk (RR) metric used in this comparison as well as all others where not specified otherwise is defined as:

$$\frac{P(\text{advancement} | \text{rank} \geq N)}{P(\text{advancement} | \text{rank} < N)}. \quad (1)$$

The third ranking method presented in Figure 1, "RDG-T", differs from the RDG model only in that it also uses time since the phase 2 transition as a predictive factor as well as all others. We observe that the use of this information almost universally improves standard performance metrics like ROC (receiver operating characteristic) and AP (average precision), however it adds little to no value in rankings beyond a level where substantial relative risk increases can be observed. In other words, it constitutes an effective but coarse mechanism for ranking TD pairs while lacking the high precision of other factors like genetic support. This is consistent with its nature as a necessary but not sufficient condition for success, and it is likely that much of the value it adds in the tail of lower rankings could be captured and enhanced if other early indicators for the many reasons trials fail [4] were also included (see more on this in Section 4). As a more practical concern, we refrain from focusing on RDG-T, or the similar GBM-T model, because neither is readily applicable to undeveloped TD pairs for which the time since phase 2 transition is not available. They are, however, a useful performance ceiling towards which future work could build.

A subset of therapeutic areas was used for this analysis and this is reflected in both the mean estimates and distributions of Figure 1. A full list of these can be seen in Supplementary Figure 8, which also shows the RR estimates used in any summary statistics across therapeutic areas. These were selected based on criteria described in Section 5. Lastly, Supplementary Figure 11 shows RR distributions by model with the significance of their differences.

The relative risk estimates from the RDG model were also compared to univariate relative risk scores in Figure 2. This clearly implicates genetic features as being the most precise predictors, albeit with a relatively low frequency compared to others. Similar comparisons for target and disease specific features can be seen in Supplementary Figure 12. These suggest that past clinical success for targets and diseases (independent of pairings) are likely to be highly predictive of advancement as well, while genetic constraint and expression specificity of targets also exhibit modest but significant effects.

In order to establish baseline levels of success and coverage across TD pairs, we refer to several of the univariate features in Figure 2 throughout this study. Specifically, the `omim` and `ot_genetics_portal` features correspond to

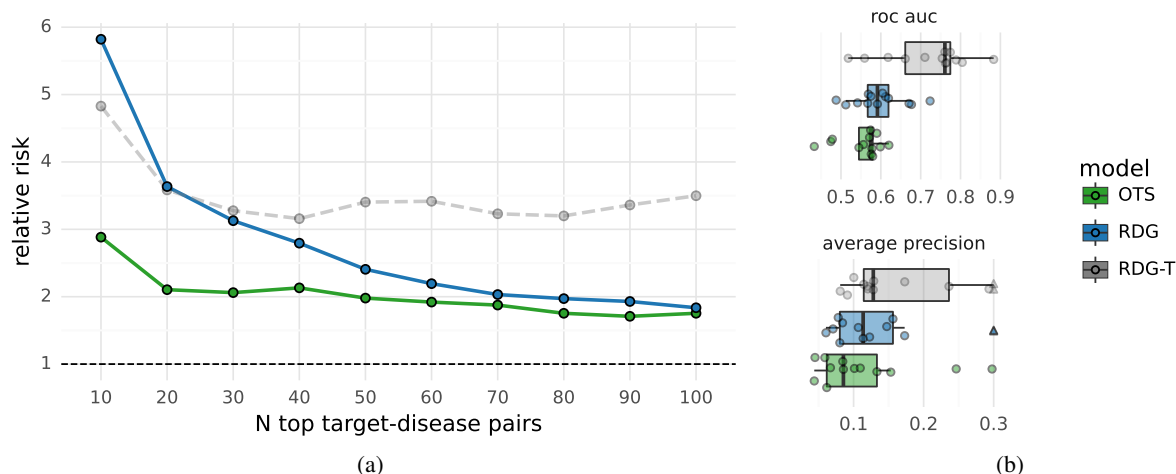


Figure 1: **Statistical learning optimizes expected target-disease pair advancement rates using longitudinal evidence.** (a) Equally-weighted average relative risk estimates across 13 therapeutic areas, by number of top rankings and 3 methods: RDG (ours), RDG-T (ours) and Open Targets composite scores. (b) Receiver operating characteristic (ROC) and average precision scores across the same 13 therapeutic areas with no limit on the number of rankings. See Supplementary Figure 8 for raw data underlying (a).

OMIM and GWAS baselines in the parlance of [2], [1] and [3]. We also use the `eva` feature as a baseline for ClinVar support, which has discriminative capabilities somewhere between these two. The extent to which the RDG model meets or exceeds these benchmarks in both relative risk for advancement and TD pair coverage is illustrated in Figure 3.

One objective of this study is to determine if any model, e.g. RDG, could sort TD pairs with genetic support such that at least some portion of that sorted list had a likelihood of advancement that consistently exceeds what is expected from any one source of genetic support alone. As a hypothetical example, we wanted to determine if the top 50 TD pairs out of 200 with GWAS support were more likely to advance than all 200 overall. We find that this goal is met and exceeded by the RDG model, which actually identifies more TD pairs than those that have either EVA or GWAS support alone at an expected rate of advancement exceeding that of the single source (respectively). This does not appear to be the case with the OMIM baseline, however the lack of examples in our evaluation dataset with OMIM support makes any determination difficult. See Section 2.8 for more on how these benchmarks are employed to contextualize opportunities among undeveloped TD pairs.

2.4 Models

The primary model for this analysis, RDG, is a constrained, L2-regularized linear regressor. Alternatives to this choice we explore include constrained and unconstrained gradient-boosted machine (GBM) models fit with the LightGBM [34] algorithm as well as unconstrained L2-regularized linear regressors (i.e. RDG with no constraints). The performance of these alternatives, as quantified by relative risk estimates among top predictions, is illustrated in Figure 4.

We conclude from these results that 1) linear models outperform, or are at least not worse, than tree models when time is not used as a predictor and 2) non-negativity constraints in linear models result in greater performance. This is consistent with our expectation that all Open Targets scores should have monotonically increasing effects on the likelihood of success, since they were designed with this intent.

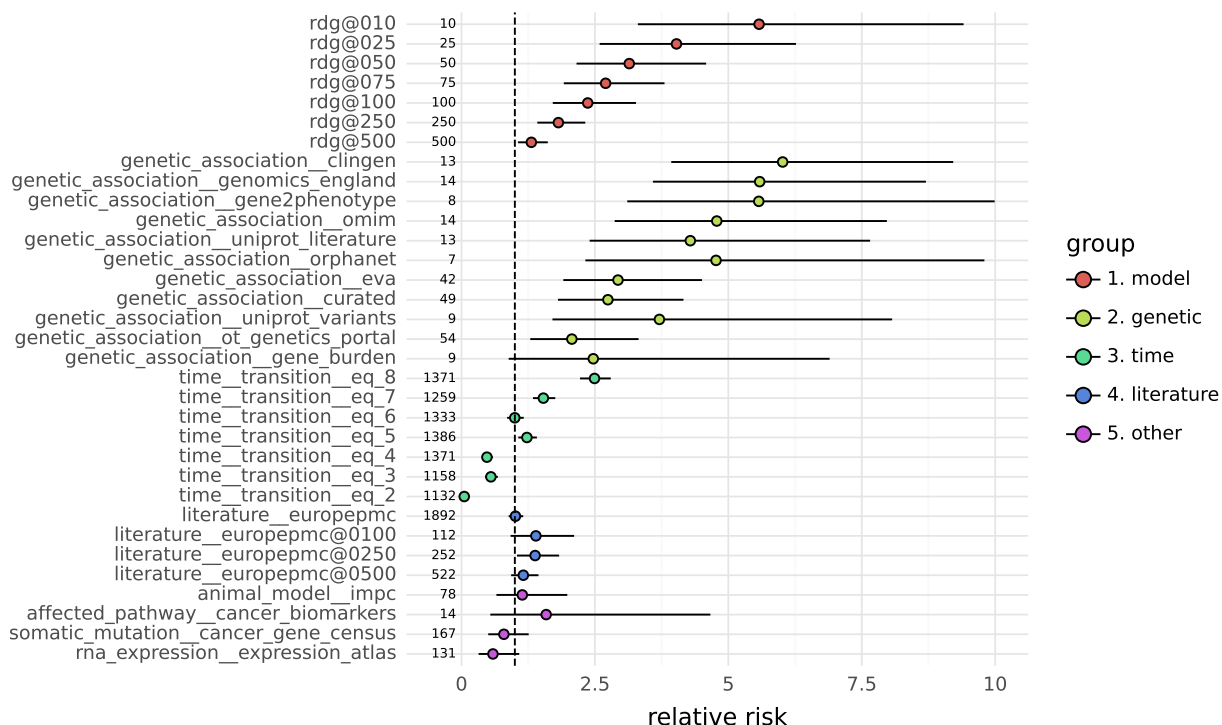


Figure 2: **Performance of individual features and predictive scores as measured by relative risk.** RDG model results denoted by `rdg@N` correspond to results for the `N` top TD pairs. The same convention is used for `literature` evidence and the `time_transition_eq_X` convention denotes RR estimates when the time since the phase 2 transition is equal to `X` years. All other features are assessed based on their existence. The counts along the origin indicate how many TD pairs were used to compute the RR numerator.

2.5 Sensitivity

In order to validate the stability of our findings in Section 2.3, we repeat this analysis across 18 different configurations listed in Supplementary Table 3. This includes 3 separate versions of Open Targets, 3 choices for the year defining the split between training and evaluation data and 2 choices for the length of the minimum advancement window (in years).

We find that the mean RR values from the RDG model consistently exceed the OTS model in all configurations among the very highest rankings ($N=10$) and also exceed the OTS model in all configurations except for 1 for N between 20 and 60. This data is shown in Supplementary Figure 15. The significance of these differences drops notably after $N=40$, which can be seen in the distribution of p-values from a Wilcoxon signed-rank test shown in Supplementary Figure 16.

2.6 Predictions

Top predictions from the RDG model over the evaluation dataset are shown in Supplementary Figure 9. This figure demonstrates that there are a significant number of TD pairs with genetic support from many sources and that the combined effects of non-genetic features also substantially influence these rankings.

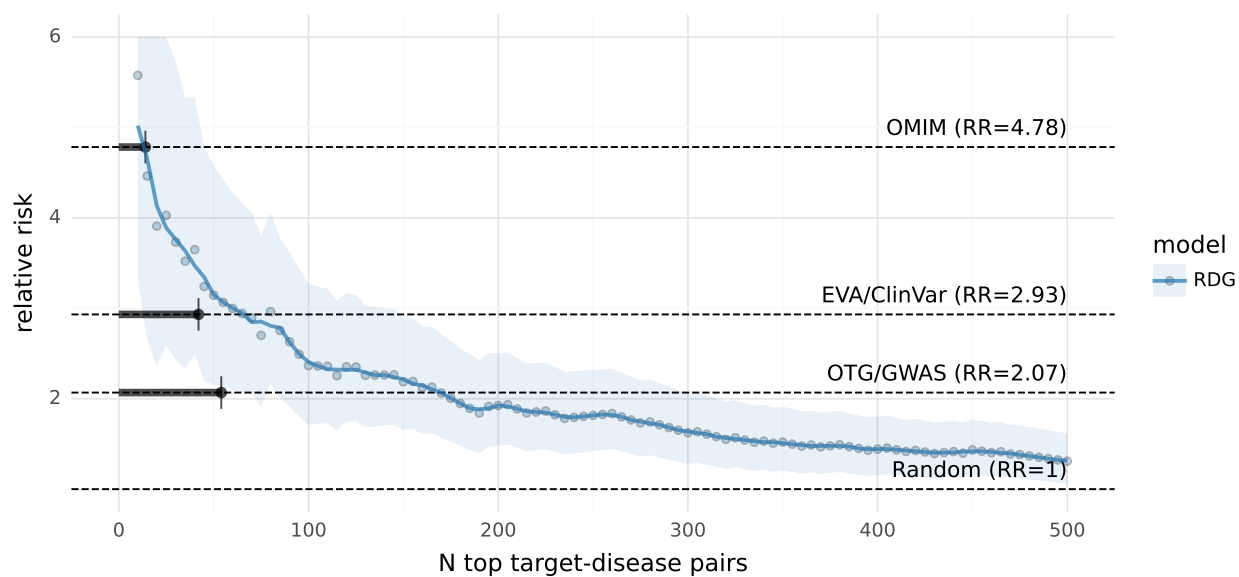


Figure 3: **RDG performance versus genetic support benchmarks across all therapeutic areas.** The RR estimates for each benchmark are based on the presence of the corresponding support across all TD pairs, and the number of pairs for which is present is represented by the horizontal bars extending horizontally from the y-axis. The bounds around the RDG RR estimates correspond to a 90% confidence interval.

| | | | | | | | | | |
|----------|------|------|--------|--------|--------|--------|--------|--------|-------------------------------------|
| RDG-T | 0.21 | 0.72 | 4.83 | 3.58 | 3.28 | 3.16 | 3.40 | 3.50 | feature ablations ([+] models) |
| RDG | 0.14 | 0.60 | 5.82 | 3.63 | 3.13 | 2.79 | 2.41 | 1.84 | |
| RDG-X | 0.12 | 0.57 | 4.27 | 2.75 | 2.71 | 2.75 | 2.38 | 1.94 | |
| OTS | 0.11 | 0.55 | 2.88 | 2.10 | 2.06 | 2.13 | 1.98 | 1.75 | |
| RDG[+] | 0.14 | 0.60 | 5.82 | 3.63 | 3.13 | 2.79 | 2.41 | 1.84 | model algorithms (core features) |
| RDG[+/-] | 0.14 | 0.59 | 5.04 | 3.53 | 3.13 | 2.53 | 2.33 | 1.86 | |
| GBM[+] | 0.13 | 0.54 | 4.09 | 3.34 | 2.89 | 2.63 | 2.54 | 1.97 | |
| GBM[+/-] | 0.13 | 0.58 | 4.44 | 3.29 | 2.64 | 2.41 | 2.24 | 1.70 | |
| OTS | 0.11 | 0.55 | 2.88 | 2.10 | 2.06 | 2.13 | 1.98 | 1.75 | |
| | AP | ROC | RR@010 | RR@020 | RR@030 | RR@040 | RR@050 | RR@100 | |

Figure 4: **Performance of ...**

2.7 Effects

The coefficients learned by the RDG model, and the average effects they have across the evaluation dataset, are shown in Figure 5. This model most highly prioritizes genetic signals that have the greatest coverage, i.e. associations from GWAS studies through the `ot_genetics_portal` feature and associations from any curated clinical genetics source, i.e. EVA, Orphanet, UniProt, Genomics England, ClinGen and gene2phenotype, via the `curated` feature. Notably, literature and target/disease specific clinical features also have substantial effects, followed by indicators of animal evidence and target genetic constraint / expression specificity. Any features not shown were deflated to have no effect, which

is possible in this model due to the non-negativity constraint. One such feature worth emphasizing is transcriptomic evidence from Expression Atlas. We found this somewhat surprising, but it is supported by arguments against transcript over/under expression as an indicator of genes that influence disease rather than the other way around [35].

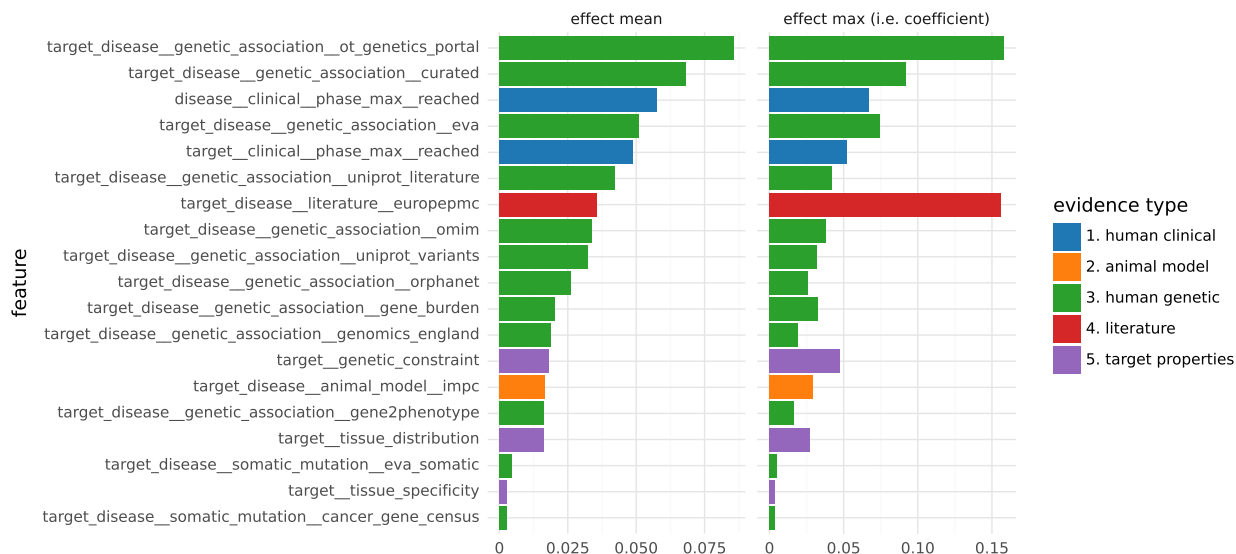


Figure 5: **RDG model feature effects.** The **effect max** values are equivalent to RDG model coefficients for the corresponding feature while the **effect mean** values indicate average values of the product between the coefficient and a particular feature value, when that feature is present.

It is worth noting that the discordance between the coefficients and the average feature effects of Figure 5 arises from both the frequency with which features exist and the distribution of their underlying scores. Scores for many clinical genetics features (e.g. OMIM, Genomics England, UniProt) are very frequently absent or close to 1. By comparison, scores for literature associations are typically far lower, even when limited only to cases where they exist, with a median value of 0.12 (mean=.23) in the evaluation data.

2.8 Opportunities

A common method for identifying druggable opportunities within a specific disease context involves first ranking TD pairs according to some prioritization methodology followed by filtering or reprioritizing those ranks based on knowledge of target tractability [36, 37, 11]. We use a similar approach to identify tractable targets associated with TD pairs that have yet to enter clinical trials. To aid in interpreting this approach, we also draw on the results of Figure 3. The data in this figure suggests thresholds for the RDG model that align to expected rates of advancement compared to several genetic support benchmarks. These thresholds are used to bucket undeveloped TD pairs before further bucketing them based on levels of tractability. The tractability buckets in Supplementary Table 2 provide **HIGH**, **MED**, and **LOW** confidence ratings for each type of tractability evidence based on the priorities suggested in [38].

Figure 6 shows the distribution of TD pair counts for select buckets across therapeutic areas as well as across the current maximum phase reached for any one pair. We find that there are ~2,400 small-molecule-enabled, ~1,400 antibody-enabled, and 14 PROTAC-enabled TD pairs with a probability of advancement that is nearly 3x other TD pairs based on the EVA threshold $RR=2.93$ in Figure 3. Top antibody-enabled pairs are shown in Supplementary Figure 14 along with their corresponding genetic and clinical support.

| | stage [threshold=EVA] | | | | | | threshold [stage=NONE] | | | tractability [stage=NONE, confidence=HIGH, threshold=EVA] | | | |
|----------------------------------------------|--------------------------|------|---------|---------|---------|---------|---------------------------|------|-------|-----------------------------------------------------------------|-----|----|------|
| | ALL | NONE | Phase 1 | Phase 2 | Phase 3 | Phase 4 | EVA | OMIM | OTG | AB | OC | PR | SM |
| therapeutic area | | | | | | | | | | | | | |
| ALL | 9724 | 8560 | 165 | 367 | 274 | 358 | 8560 | 679 | 20821 | 1456 | 655 | 14 | 2408 |
| genetic, familial or congenital disease | 4867 | 4638 | 32 | 73 | 58 | 66 | 4638 | 402 | 7641 | 399 | 271 | 12 | 993 |
| cancer or benign tumor | 2063 | 1531 | 122 | 209 | 125 | 76 | 1531 | 96 | 6506 | 430 | 124 | 0 | 639 |
| nervous system disease | 2093 | 1923 | 10 | 54 | 49 | 57 | 1923 | 166 | 3710 | 153 | 74 | 3 | 446 |
| musculoskeletal or connective tissue disease | 1597 | 1469 | 15 | 46 | 35 | 32 | 1469 | 158 | 2863 | 206 | 129 | 0 | 353 |
| gastrointestinal disease | 1149 | 995 | 22 | 48 | 30 | 54 | 995 | 82 | 3265 | 257 | 67 | 0 | 334 |
| immune system disease | 1294 | 1114 | 30 | 69 | 36 | 45 | 1114 | 107 | 2762 | 339 | 69 | 4 | 290 |
| nutritional or metabolic disease | 1491 | 1409 | 1 | 16 | 16 | 49 | 1409 | 132 | 2316 | 94 | 29 | 0 | 310 |
| endocrine system disease | 1071 | 933 | 17 | 42 | 26 | 53 | 933 | 99 | 2685 | 184 | 80 | 4 | 287 |
| cardiovascular disease | 956 | 834 | 7 | 25 | 32 | 58 | 834 | 81 | 1908 | 126 | 70 | 0 | 315 |
| psychiatric disorder | 820 | 735 | 2 | 22 | 16 | 45 | 735 | 73 | 1442 | 59 | 18 | 2 | 168 |
| disorder of visual system | 838 | 823 | 1 | 8 | 3 | 3 | 823 | 87 | 1289 | 40 | 24 | 0 | 131 |
| integumentary system disease | 733 | 646 | 14 | 35 | 22 | 16 | 646 | 51 | 1502 | 115 | 40 | 0 | 164 |
| hematologic disease | 702 | 589 | 23 | 41 | 31 | 18 | 589 | 45 | 1394 | 158 | 73 | 1 | 202 |
| respiratory or thoracic disease | 590 | 472 | 8 | 52 | 33 | 25 | 472 | 23 | 1552 | 177 | 45 | 0 | 134 |
| urinary system disease | 550 | 490 | 5 | 19 | 19 | 17 | 490 | 42 | 1164 | 94 | 47 | 0 | 121 |
| reproductive system or breast disease | 534 | 452 | 9 | 30 | 18 | 25 | 452 | 33 | 1240 | 69 | 22 | 6 | 128 |
| phenotype | 353 | 307 | 3 | 10 | 10 | 23 | 307 | 13 | 921 | 74 | 54 | 0 | 142 |
| pancreas disease | 326 | 278 | 7 | 9 | 7 | 25 | 278 | 37 | 770 | 65 | 24 | 0 | 96 |
| measurement | 7 | 7 | 0 | 0 | 0 | 0 | 7 | 1 | 69 | 1 | 2 | 0 | 3 |
| disorder of ear | 5 | 5 | 0 | 0 | 0 | 0 | 5 | 0 | 21 | 1 | 0 | 0 | 1 |

Figure 6: **Present-day target-disease pair counts by stage, likelihood of advancement and tractability.** The **stage** panel contains counts by maximum trial phase reached, the **threshold** panel contains counts of pairs with a RDG model score exceeding that of the associated benchmark for only undeveloped pairs, and the **tractability** panel shows pair frequencies among undeveloped pairs exceeding the EVA threshold that also have a HIGH tractability rating as defined in Supplementary Table 2. This corresponds to targets that have all been in clinical development already, except for the **OC** modality in which case it indicates that a target has been approved.

2.9 Inflation

Like most studies of this kind, we assume a "closed-world" [17] over the space of target-disease pairs and any evidence between them. This means that we do not differentiate between evidence that an association for any one pair truly does **not** exist (or is too weak to be relevant under the omnigenic model [39]), and the lack of any attempt to find that evidence in the first place. This also means that our estimate of the prognostic value for any one evidence source is subject to historical trends in biomedical research and the myriad ways that this research can be biased towards particular targets and diseases. We avoid attempting to comprehensively survey these biases in favor of offering an illustrative list of specific examples that are relevant in this study:

1. Mendelian randomization research is biased towards cardiovascular diseases as they have a disproportionate number of known, modifiable exposures [40]
2. Putative protein interactions that do not result from genome-scale or otherwise unbiased assays result in an overrepresentation of successful drug targets in resources like STRING [41], thereby inflating the success of network expansion methods over these databases to identify such targets [14].
3. Transcript expression studies run in late-stage clinical trials for a single indication, e.g. [42] linking SLE to IFN genes, are a degenerate indicator of advancement beyond earlier stage trials when the timing of this evidence is not accounted for.
4. Targets tested against more indications in clinical trials enrich for failures because the marginal cost of testing more indications decreases, but the evidence for these indications is often weaker [12].

5. Herding effects in pharma R&D pipelines around particular drug targets are becoming increasingly clear over time [43] and generate an excess of clinical evidence for those targets.

We also note that the skew in basic drug target research towards those that already have rich annotations and well characterized molecular function [44] as well as the disproportionate representation of particular target families in pharma R&D pipelines [45, 46] and the fact that literature is well known to be biased away from negative results in general [47] are all problematic.

While it is not possible to address all of these issues, we emphasize that there is a clear pattern across the examples in the list above in that they require **past** clinical successes and/or failures to arise in the first place. This suggests that accounting for when evidence first emerged would limit the extent of these problems. We do so in this study based solely on publication dates associated with any one piece of information linking target-disease pairs. This also offers a novel opportunity to attempt to quantify what kind of evidence suffers most from these biases. Figure 7 presents results for this based on a relative risk statistic defined as:

$$\frac{P(A|B)}{P(A|\neg B)} \quad (2)$$

where:

- A is the event that evidence for a TD pair arises after its first early-stage (phase 1 or 2) trial rather than before
- B is the event that a TD pair advances into late-stage trials (phase 3 or 4)

We refer to this as "inflation risk" so as not to confuse it with the relative risk statistic used in all other contexts, and it can be more simply described as the fraction of TD pairs for which evidence arises **after** the beginning of an ultimately successful early-stage trial divided by that same fraction for TD pairs that do not advance to late-stage trials. The intuition for this statistic is that it will be higher if successful trials lead to the generation of evidence of a particular type, and it should be 1 in cases where the emergence of evidence is independent of clinical success. We also measure this potential lack of independence through the more commonly used Fisher's exact test, e.g. [48], and both are presented in Figure 7.

We find that evidence from Reactome is the worst offender by this metric, implying that it often only arises for TD pairs after a certain level of clinical success has been attained. We also find that long-running aggregators/curators of published research often focused on individual diseases/phenotypes, like Expression Atlas, IMPC, CGS and Cancer Biomarkers exhibit this form of inflation as well.

Sources of genetic evidence appear to be much less inflated, or have too little data to reach significance. This is to be expected for GWAS evidence arising from genome-wide, phenome-wide biobank consortia, however much of historical GWAS evidence is not phenome-wide. More context on how much this is likely to matter comes from [49] in which it was estimated that as few as 6% of 500 FDA-approved targets for non-cancer drugs arose from programs highly motivated by pre-existing genetic support and that "the remaining 94% were probably identified using conventional pharmacology, biochemistry or molecular biology approaches". We then speculate that if the initiation of new drug programs was not historically motivated highly by the existence of genetic support, then the incentives for pursuing new genetic evidence based on clinical and commercial success are likely to be minimized. This, in conjunction with existing precedent [1, 2, 3, 4, 5, 7] and our inflation results, ultimately led us to the use of genetic evidence without temporalization. In other words, we do not treat genetic evidence as longitudinal features like all others associated with TD pairs. A breakdown of which features are treated in which manner is provided in Supplementary Table 1.

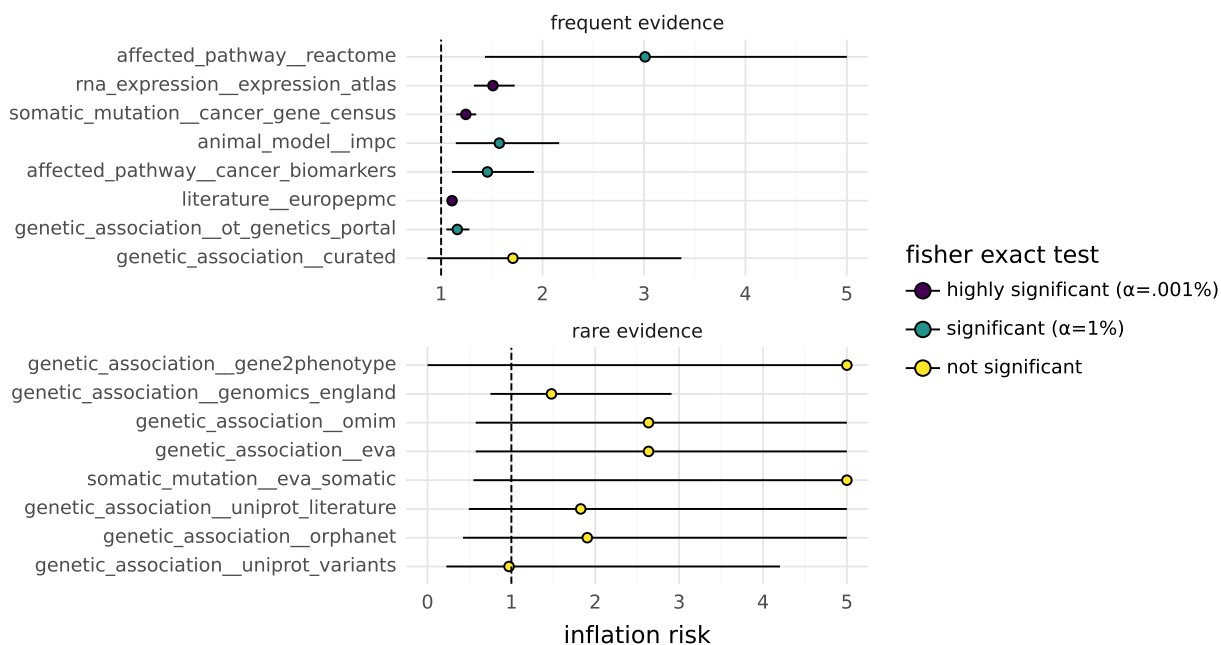


Figure 7: **Clinical success drives evidence discovery.**

3 Conclusion

We have demonstrated that simple machine learning methods applied to longitudinal biomedical evidence from many sources can be used to predict clinical outcomes for combinations of drug targets and diseases, without knowledge of molecular properties or trial design details. We have also shown that these methods are more precise in the extremes of their predictions than composite, heuristic scores like those from Open Targets. They also outperform such baselines by more comprehensive, traditional measures of classifier performance; however, we find this less compelling and easier to accomplish than improving performance among the upper tail of the opportunities implied by the very highest predictions. This framework would also support the addition of new lines of evidence over time well as it is designed to automatically determine the relevance of any new information without intervention. Lastly, we find that the space of present-day, undeveloped targets within a disease context that both exceed baseline levels of tractability and have a high predicted likelihood of clinical advancement is substantial. It is likely to grow quickly as well since the breadth of much of the underlying evidence is expanding rapidly [50, 51, 52].

4 Discussion

- Cover limitations with OT concerning temporalization for both evidence and drug approvals, and why only transitions from phase 2 are relevant for this work
- Discuss the possibility to do prospective evaluation with OT snapshots
- Discuss how typical classifier metrics are inadequate for this problem
- Talk about how the inclusion of time greatly improves performance and suggests it may be possible to fill in this proxy with predictors for more specific early indicators of other reasons trials fail
- We are focusing on prioritizing among targets/indications with genetic support, rather than expanding this space to find opportunities with weaker support

- "There is evidence that a 9.6% vs. 13.8% success rate for drugs from phase 1 trials to approval may mean a \$480 million difference in the median research and development cost required to bring a new drug to the market (Wouters, McKee, and Luyten 2020)." [53]
- From [12]: "It is important to bear in mind therefore that what we are measuring when looking at historical trial outcomes is not an unbiased measure of any given gene's true disease associations, but rather a view on how useful a given evidence source or analytical method has been for choosing drug targets based on current and historical drug discovery practices. Dramatic changes in these practices in the future could render some of our conclusions obsolete, though the fundamental observation that genetic association itself is retained in molecular networks will remain valid."
- Add select tractability and DepMap features?

5 Methods

- Discuss why all target prioritisation data fields are not used due to the potential leakage they may impose (e.g. target families, GO annotations, etc.)
- Describe how therapeutic areas were selected based on having at least 100 TD pairs with target-disease-specific evidence of any kind and with explicit omissions: ("biological_process", "pregnancy or perinatal disease", "injury, poisoning or other complication", "pregnancy or perinatal disease", "medical procedure", "infectious disease", "animal disease")
- Training/evaluation results are limited by temporalization while the present-day predictions are not
- OMIM is defined as EVA associations with publications
- The "genetic_association__curated" field is a union of all genetic association sources other than "gene_burden" and "ot_genetics_portal"
- Mention that the scores for a TD pairs in a year are the maximum score for that source, not the harmonic sum
- Mean imputation is used for target-specific features
- Mention specifics on RDG and GBM implementations, i.e. lightGBM and scikit-learn
- RDG is always trained on all features, but only applied to feature subsets where relevant

6 Supplementary Material

Table 1: Features used in modeling and analysis

| | feature | entity | kind |
|----|---------------------------------------------------------|----------------|----------|
| 1 | disease__clinical__phase_max__reached | disease | temporal |
| 2 | target__clinical__phase_max__reached | target | temporal |
| 3 | target__genetic_constraint | target | static |
| 4 | target__mouse_ko_score | target | static |
| 5 | target__tissue_distribution | target | static |
| 6 | target__tissue_specificity | target | static |
| 7 | target_disease__affected_pathway__cancer_biomarkers | target_disease | temporal |
| 8 | target_disease__affected_pathway__crispr | target_disease | temporal |
| 9 | target_disease__affected_pathway__crispr_screen | target_disease | temporal |
| 10 | target_disease__affected_pathway__progeny | target_disease | temporal |
| 11 | target_disease__affected_pathway__reactome | target_disease | temporal |
| 12 | target_disease__affected_pathway__slapenrich | target_disease | temporal |
| 13 | target_disease__affected_pathway__sysbio | target_disease | temporal |
| 14 | target_disease__animal_model__impc | target_disease | temporal |
| 15 | target_disease__genetic_association__clingen | target_disease | static |
| 16 | target_disease__genetic_association__curated | target_disease | static |
| 17 | target_disease__genetic_association__eva | target_disease | static |
| 18 | target_disease__genetic_association__gene2phenotype | target_disease | static |
| 19 | target_disease__genetic_association__gene_burden | target_disease | static |
| 20 | target_disease__genetic_association__genomics_england | target_disease | static |
| 21 | target_disease__genetic_association__omim | target_disease | static |
| 22 | target_disease__genetic_association__orphanet | target_disease | static |
| 23 | target_disease__genetic_association__ot_genetics_portal | target_disease | static |
| 24 | target_disease__genetic_association__uniprot_literature | target_disease | static |
| 25 | target_disease__genetic_association__uniprot_variants | target_disease | static |
| 26 | target_disease__known_drug__chembl | target_disease | temporal |
| 27 | target_disease__literature__europepmc | target_disease | temporal |
| 28 | target_disease__outcome__advanced | target_disease | temporal |
| 29 | target_disease__rna_expression__expression_atlas | target_disease | temporal |
| 30 | target_disease__somatic_mutation__cancer_gene_census | target_disease | temporal |
| 31 | target_disease__somatic_mutation__eva_somatic | target_disease | temporal |
| 32 | target_disease__somatic_mutation__intogen | target_disease | temporal |
| 33 | target_disease__time__transition | target_disease | temporal |

Table 2: Tractability bucket assignments

| | evidence | modality | confidence |
|----|-------------------------|----------|------------|
| 1 | Phase 1 Clinical | OC | LOW |
| 2 | Advanced Clinical | OC | MED |
| 3 | Approved Drug | OC | HIGH |
| 4 | GO CC med conf | AB | LOW |
| 5 | Human Protein Atlas loc | AB | LOW |
| 6 | UniProt SigP or TMHMM | AB | LOW |
| 7 | UniProt loc med conf | AB | LOW |
| 8 | GO CC high conf | AB | MED |
| 9 | UniProt loc high conf | AB | MED |
| 10 | Advanced Clinical | AB | HIGH |
| 11 | Approved Drug | AB | HIGH |
| 12 | Phase 1 Clinical | AB | HIGH |
| 13 | Database Ubiquitination | PR | LOW |
| 14 | Half-life Data | PR | LOW |
| 15 | Small Molecule Binder | PR | LOW |
| 16 | Literature | PR | MED |
| 17 | UniProt Ubiquitination | PR | MED |
| 18 | Advanced Clinical | PR | HIGH |
| 19 | Phase 1 Clinical | PR | HIGH |
| 20 | Druggable Family | SM | LOW |
| 21 | High-Quality Pocket | SM | LOW |
| 22 | Med-Quality Pocket | SM | LOW |
| 23 | High-Quality Ligand | SM | MED |
| 24 | Structure with Ligand | SM | MED |
| 25 | Advanced Clinical | SM | HIGH |
| 26 | Approved Drug | SM | HIGH |
| 27 | Phase 1 Clinical | SM | HIGH |

Table 3: Configurations for sensitivity analysis

| | open_targets_version | max_training_year | min_time_to_advancement_years |
|----|----------------------|-------------------|-------------------------------|
| 1 | 23.09 | 2017 | 4 |
| 2 | 23.12 | 2017 | 2 |
| 3 | 23.09 | 2015 | 2 |
| 4 | 23.12 | 2015 | 2 |
| 5 | 23.09 | 2017 | 2 |
| 6 | 23.06 | 2015 | 4 |
| 7 | 23.06 | 2017 | 2 |
| 8 | 23.06 | 2013 | 2 |
| 9 | 23.06 | 2015 | 2 |
| 10 | 23.06 | 2017 | 4 |
| 11 | 23.12 | 2013 | 4 |
| 12 | 23.09 | 2015 | 4 |
| 13 | 23.09 | 2013 | 2 |
| 14 | 23.12 | 2013 | 2 |
| 15 | 23.09 | 2013 | 4 |
| 16 | 23.12 | 2017 | 4 |
| 17 | 23.12 | 2015 | 4 |
| 18 | 23.06 | 2013 | 4 |

| | | OMIM | EVA/ClinVar | OTG/GWAS | OTS@010 | OTS@020 | OTS@030 | OTS@050 | OTS@100 | RDC@010 | RDC@020 | RDC@030 | RDC@050 | RDC@100 |
|----------------------------------------------|---------|-------|-------------|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| therapeutic_area | n_pairs | | | | | | | | | | | | | |
| average | | 9.65 | 5.78 | 1.95 | 2.88 | 2.10 | 2.06 | 1.98 | 1.75 | 5.82 | 3.63 | 3.13 | 2.41 | 1.84 |
| all | 9010 | 4.78 | 2.93 | 2.07 | 6.70 | 3.91 | 3.35 | 2.69 | 2.13 | 5.58 | 3.91 | 3.73 | 3.14 | 2.37 |
| cancer or benign tumor | 4013 | | 3.69 | 0.00 | 2.58 | 3.91 | 5.31 | 3.65 | 2.68 | 5.20 | 5.00 | 4.40 | 4.28 | 3.26 |
| genetic, familial or congenital disease | 2035 | 3.43 | 2.37 | 0.68 | 3.08 | 1.53 | 1.02 | 0.81 | 0.91 | 3.08 | 2.58 | 1.71 | 1.23 | 0.80 |
| nervous system disease | 1534 | 3.64 | 1.95 | 0.00 | 2.54 | 1.15 | 1.70 | 1.54 | 2.14 | 2.54 | 1.26 | 2.59 | 2.08 | 1.56 |
| gastrointestinal disease | 1260 | | 0.00 | 3.48 | 5.68 | 4.33 | 2.68 | 3.17 | 2.82 | 8.72 | 5.90 | 3.90 | 4.34 | 2.44 |
| immune system disease | 1225 | 21.09 | 21.44 | 2.88 | 6.39 | 4.30 | 4.43 | 2.61 | 1.99 | 11.05 | 5.48 | 4.43 | 3.10 | 2.53 |
| hematologic disease | 1023 | 26.89 | 27.59 | 0.00 | 5.48 | 2.71 | 1.79 | 2.05 | 1.68 | 8.44 | 4.18 | 3.78 | 2.22 | 1.64 |
| endocrine system disease | 949 | 0.00 | 0.00 | 11.26 | 0.00 | 0.00 | 2.12 | 2.88 | 2.21 | 10.83 | 5.36 | 4.90 | 2.88 | 1.77 |
| reproductive system or breast disease | 948 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.68 | 1.72 | 8.79 | 4.35 | 3.95 | 2.99 | 2.51 |
| musculoskeletal or connective tissue disease | 757 | 8.82 | 6.67 | 2.68 | 4.74 | 3.62 | 2.73 | 2.48 | 2.18 | 6.02 | 5.00 | 4.76 | 3.09 | 2.80 |
| respiratory or thoracic disease | 644 | | 0.00 | 0.00 | 4.37 | 2.74 | 1.80 | 1.97 | 1.52 | 3.22 | 2.15 | 2.19 | 1.51 | 1.83 |
| cardiovascular disease | 617 | 3.65 | 1.66 | 2.09 | 1.83 | 1.66 | 1.48 | 0.86 | 0.87 | 2.21 | 1.85 | 1.35 | 0.93 | 0.68 |
| integumentary system disease | 563 | | | 0.00 | 0.00 | 0.00 | 0.82 | 1.43 | 1.49 | 4.37 | 2.94 | 1.92 | 1.76 | 1.12 |
| urinary system disease | 561 | | 3.99 | 2.29 | 0.79 | 1.40 | 0.92 | 0.61 | 0.59 | 1.19 | 1.19 | 0.78 | 0.86 | 0.93 |

Figure 8: **Relative risk scores by method, benchmark and therapeutic area.** The **average** therapeutic area indicates mean values across all others except for **all**, which is an ungrouped estimate across all diseases regardless of therapeutic area.

| target_symbol | disease_name | advanced | prediction | target_disease_genetic_association_ot_genetics_portal | target_disease_literature_europepmc | target_disease_genetic_association_curated | target_disease_genetic_association_eva | disease_clinical_phase_max_reached | target_disease_genetic_association_uniprot_literature | target_clinical_phase_max_reached | target_disease_genetic_association_omim | target_disease_genetic_association_uniprot_variants | target_disease_genetic_association_gene_burden | target_disease_genetic_association_orphanet | target_disease_genetic_association_genomics_england | target_disease_animal_model_imp | target_genetic_constraint | target_disease_genetic_association_gene2phenotype | target_tissue_distribution | target_disease_somatic_mutation_eva_somatic | target_disease_somatic_mutation_cancer_gene_census | target_tissue_specificity |
|---------------|------------------------------------------|----------|------------|-------------------------------------------------------|-------------------------------------|--------------------------------------------|----------------------------------------|------------------------------------|-------------------------------------------------------|-----------------------------------|-----------------------------------------|-----------------------------------------------------|------------------------------------------------|---------------------------------------------|-----------------------------------------------------|---------------------------------|---------------------------|---------------------------------------------------|----------------------------|---------------------------------------------|----------------------------------------------------|---------------------------|
| CFB | age-related macular degeneration | False | 0.604 | 0.117 | 0.115 | 0.092 | 0.067 | 0.067 | 0.042 | 0.039 | 0.034 | | | | | | 0.013 | | 0.014 | | | 0.004 |
| SNCA | Parkinson disease | False | 0.545 | 0.153 | 0.103 | 0.092 | 0.068 | 0.067 | | 0.013 | 0.035 | | | | | | 0.010 | | | | | 0.003 |
| HTT | Huntington disease | True | 0.545 | | 0.156 | 0.092 | 0.074 | 0.067 | 0.042 | 0.013 | 0.038 | | | 0.026 | 0.020 | 0.014 | 0.002 | | | | | |
| FGFR3 | Achondroplasia | False | 0.538 | | 0.053 | 0.092 | 0.071 | 0.051 | 0.042 | 0.052 | 0.036 | 0.032 | | 0.026 | 0.020 | 0.016 | 0.015 | 0.016 | 0.014 | | | 0.003 |
| MYH7 | hypertrophic cardiomyopathy | True | 0.518 | | 0.058 | 0.092 | 0.072 | 0.067 | 0.042 | 0.026 | 0.036 | 0.032 | 0.026 | | 0.020 | | 0.014 | 0.016 | 0.014 | | | 0.003 |
| CXCR4 | WHIM syndrome | True | 0.486 | | 0.089 | 0.092 | 0.071 | | 0.042 | 0.052 | 0.036 | | | 0.026 | 0.020 | 0.012 | 0.028 | 0.016 | | | | 0.003 |
| AKT1 | Proteus syndrome | False | 0.455 | | 0.114 | 0.092 | 0.068 | 0.017 | 0.042 | 0.039 | | 0.032 | | | 0.020 | | 0.006 | 0.016 | | 0.005 | 0.004 | |
| MAPT | progressive supranuclear palsy | False | 0.448 | 0.139 | 0.028 | 0.084 | 0.068 | 0.051 | | 0.013 | 0.034 | | | | | | 0.014 | | 0.014 | | | 0.003 |
| MYH6 | hypertrophic cardiomyopathy | True | 0.443 | | 0.016 | 0.092 | 0.068 | 0.067 | 0.042 | 0.026 | 0.012 | 0.032 | | | 0.020 | 0.021 | 0.022 | | 0.020 | | | 0.004 |
| MYL3 | hypertrophic cardiomyopathy | True | 0.442 | | 0.016 | 0.092 | 0.071 | 0.067 | 0.042 | 0.026 | 0.034 | 0.032 | | | 0.020 | | 0.023 | 0.016 | | | | 0.003 |
| IL6 | asthma | False | 0.427 | 0.129 | 0.156 | | | 0.067 | | 0.039 | | | | | | | 0.019 | | 0.014 | | | 0.003 |
| FCGR2B | systemic lupus erythematosus | False | 0.426 | 0.035 | 0.154 | 0.092 | | 0.067 | 0.042 | | | | | | | 0.016 | 0.016 | | | | | 0.004 |
| IL33 | asthma | False | 0.416 | 0.145 | 0.087 | | | 0.067 | | 0.026 | | | 0.032 | | | | 0.042 | | 0.014 | | | 0.003 |
| LPA | cardiovascular disease | True | 0.403 | 0.156 | 0.117 | | | 0.067 | | | | | | | | | 0.039 | 0.020 | | | | 0.004 |
| MAPT | Alzheimer disease | False | 0.399 | | 0.156 | 0.066 | 0.054 | 0.067 | | 0.026 | | | | | | | 0.014 | 0.014 | | | | 0.003 |
| IL2RA | systemic lupus erythematosus | True | 0.380 | 0.122 | 0.106 | | | 0.067 | | 0.052 | | | | | | | 0.016 | 0.014 | | | | 0.003 |
| CD40 | systemic lupus erythematosus | False | 0.378 | 0.106 | 0.156 | | | 0.067 | | 0.026 | | | | | | | 0.010 | | 0.014 | | | |
| PIK3CA | megalencephaly-capillary... | False | 0.377 | | 0.023 | 0.092 | 0.071 | | 0.042 | 0.039 | 0.036 | 0.032 | | | 0.020 | | 0.001 | 0.016 | | 0.005 | | |
| | breast carcinoma | False | 0.376 | | 0.140 | 0.092 | | 0.067 | | 0.039 | | 0.032 | | | | | 0.001 | | | | 0.004 | |
| KRAS | lung cancer | False | 0.372 | | 0.027 | 0.092 | 0.068 | 0.067 | | 0.039 | | | | | 0.020 | 0.026 | 0.033 | | | | | |
| MYL2 | hypertrophic cardiomyopathy | True | 0.372 | | 0.020 | 0.092 | 0.071 | 0.067 | | 0.026 | | | | | | 0.020 | 0.042 | 0.016 | 0.014 | | | 0.003 |
| GRIN2B | West syndrome | False | 0.371 | | | 0.092 | 0.068 | 0.067 | | 0.052 | 0.034 | | | 0.026 | | | 0.000 | | 0.027 | | | 0.004 |
| MYH7 | dilated cardiomyopathy | False | 0.369 | | 0.034 | 0.092 | 0.072 | 0.067 | | 0.039 | | | | | | 0.017 | 0.014 | 0.016 | 0.014 | | | 0.003 |
| TNFSF13 | IGA glomerulonephritis | True | 0.358 | 0.096 | 0.156 | | | 0.067 | | 0.026 | | | | | | | 0.010 | | | | | 0.003 |
| ACE | stroke | False | 0.358 | | 0.059 | 0.092 | | 0.067 | 0.042 | 0.052 | | | | | | | 0.029 | 0.014 | | | | 0.003 |
| MAPT | Classical progressive supranuclear palsy | False | 0.358 | | 0.016 | 0.092 | 0.068 | 0.017 | 0.042 | 0.026 | 0.034 | 0.032 | | | | | 0.014 | 0.014 | | | | 0.003 |
| JAK2 | colitis | False | 0.352 | 0.117 | 0.108 | | | 0.067 | | 0.052 | | | | | | | 0.007 | | | | | |
| MC4R | obesity due to melanocortin 4... | False | 0.346 | | | 0.092 | 0.071 | | | 0.039 | 0.036 | | | 0.026 | | 0.018 | 0.040 | 0.020 | | | | 0.003 |
| MYL2 | dilated cardiomyopathy | False | 0.340 | | 0.039 | 0.064 | 0.052 | 0.067 | | 0.039 | | | | | | 0.019 | 0.042 | 0.014 | | | | 0.003 |
| WT1 | acute myeloid leukemia | True | 0.340 | | 0.156 | 0.029 | 0.024 | 0.067 | | 0.039 | | | | | | | 0.004 | | 0.014 | | 0.004 | 0.003 |

Figure 9: **Top RDG model evaluation dataset predictions.** Feature contributions are shown as the product of their underlying values and the RDG coefficients. The **advanced** field indicates whether the associated TD pair advanced beyond phase 2 as of 2024.

| | split | | | evaluation | | | | | | | | training | | | |
|----------------------------------------------|-----------|---------|----------|------------|------------|---------|-------------|-----------|---------|----------|----------|------------|---------|-------------|-----------|
| | statistic | balance | max_year | min_year | n_diseases | n_pairs | n_pairs_wév | n_targets | balance | max_year | min_year | n_diseases | n_pairs | n_pairs_wév | n_targets |
| therapeutic_area_name | | | | | | | | | | | | | | | |
| all | 9.01% | 2022 | 2016 | 1075 | 9010 | 2062 | 1063 | 19.76% | 2015 | 1990 | 1420 | 25398 | 3737 | 1226 | |
| cancer or benign tumor | 3.89% | 2022 | 2016 | 400 | 4013 | 1148 | 656 | 15.69% | 2015 | 1990 | 478 | 15303 | 2267 | 712 | |
| genetic, familial or congenital disease | 9.83% | 2022 | 2016 | 203 | 2035 | 378 | 548 | 23.12% | 2015 | 1990 | 286 | 4745 | 680 | 823 | |
| nervous system disease | 3.98% | 2022 | 2016 | 200 | 1534 | 271 | 557 | 21.21% | 2015 | 1990 | 278 | 4112 | 575 | 759 | |
| gastrointestinal disease | 3.65% | 2022 | 2016 | 136 | 1260 | 360 | 418 | 17.80% | 2015 | 1995 | 174 | 4061 | 572 | 603 | |
| immune system disease | 4.90% | 2022 | 2016 | 136 | 1225 | 347 | 461 | 16.72% | 2015 | 1992 | 167 | 3361 | 668 | 650 | |
| hematologic disease | 3.81% | 2022 | 2016 | 111 | 1023 | 262 | 389 | 19.13% | 2015 | 1990 | 139 | 3742 | 476 | 553 | |
| endocrine system disease | 3.06% | 2022 | 2016 | 118 | 949 | 289 | 376 | 17.65% | 2015 | 1990 | 140 | 3372 | 541 | 542 | |
| reproductive system or breast disease | 3.69% | 2022 | 2016 | 95 | 948 | 197 | 368 | 16.91% | 2015 | 1990 | 102 | 3294 | 331 | 486 | |
| musculoskeletal or connective tissue disease | 8.85% | 2022 | 2016 | 109 | 757 | 195 | 369 | 23.20% | 2015 | 1991 | 168 | 2190 | 389 | 667 | |
| respiratory or thoracic disease | 9.63% | 2022 | 2016 | 73 | 644 | 174 | 373 | 21.76% | 2015 | 1993 | 81 | 1769 | 318 | 609 | |
| cardiovascular disease | 27.71% | 2022 | 2016 | 88 | 617 | 121 | 285 | 28.67% | 2015 | 1993 | 116 | 1444 | 241 | 553 | |
| phenotype | 6.28% | 2022 | 2016 | 113 | 605 | 76 | 322 | 29.99% | 2015 | 1994 | 178 | 1594 | 121 | 497 | |
| integumentary system disease | 7.28% | 2022 | 2016 | 85 | 563 | 166 | 293 | 18.29% | 2015 | 1990 | 98 | 1624 | 268 | 513 | |
| urinary system disease | 25.31% | 2022 | 2016 | 62 | 561 | 134 | 360 | 21.32% | 2015 | 1992 | 73 | 1440 | 183 | 412 | |
| infectious disease | 6.24% | 2022 | 2016 | 51 | 529 | 115 | 343 | 19.05% | 2015 | 1991 | 72 | 803 | 76 | 345 | |
| psychiatric disorder | 3.25% | 2022 | 2016 | 65 | 523 | 96 | 257 | 29.79% | 2015 | 1990 | 78 | 1232 | 211 | 327 | |
| disorder of visual system | 5.24% | 2022 | 2016 | 57 | 382 | 47 | 275 | 12.44% | 2015 | 1990 | 67 | 611 | 71 | 292 | |
| nutritional or metabolic disease | 5.16% | 2022 | 2016 | 55 | 310 | 63 | 269 | 45.41% | 2015 | 1992 | 80 | 969 | 169 | 498 | |
| biological_process | 28.84% | 2022 | 2016 | 14 | 215 | 18 | 157 | 6.13% | 2015 | 1999 | 12 | 163 | 21 | 152 | |
| pancreas disease | 1.40% | 2022 | 2016 | 17 | 214 | 89 | 184 | 29.53% | 2015 | 1995 | 26 | 823 | 152 | 391 | |
| injury, poisoning or other complication | 0.00% | 2022 | 2016 | 19 | 123 | 32 | 117 | 15.04% | 2015 | 1997 | 27 | 359 | 29 | 209 | |
| pregnancy or perinatal disease | 4.17% | 2022 | 2016 | 8 | 24 | 3 | 23 | 78.85% | 2015 | 1996 | 15 | 104 | 4 | 99 | |

Figure 10: Training and evaluation dataset summary statistics.

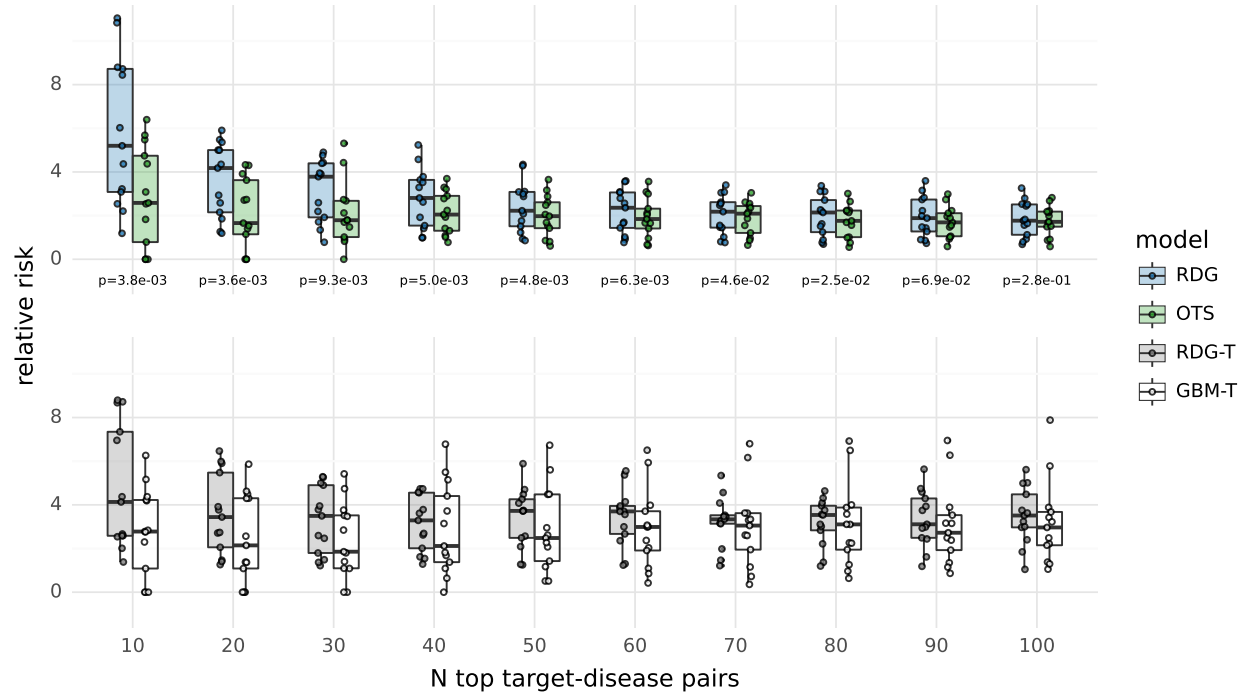


Figure 11: **Relative risk distributions across select therapeutic areas.** P-values are computed from a one-sided Wilcoxon signed-rank test with the alternative that the RDG model RR averages across therapeutic areas exceed OTS averages.

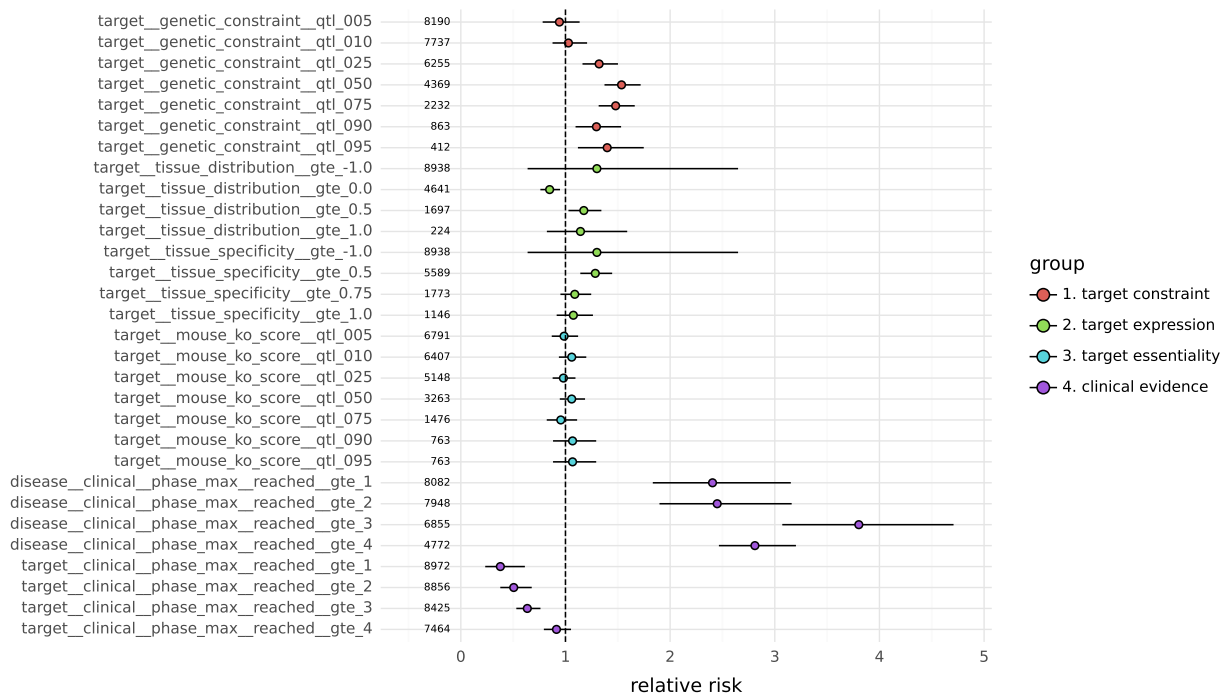


Figure 12: **Relative risk scores for target/disease features.** The features ending with **qtl_Q** denote binary indicators constructed from cases where the feature meets or exceeds quantile **Q** of its distribution. The features ending with **gte_X** denote indicators for when the feature meets or exceeds a specific value **X**.

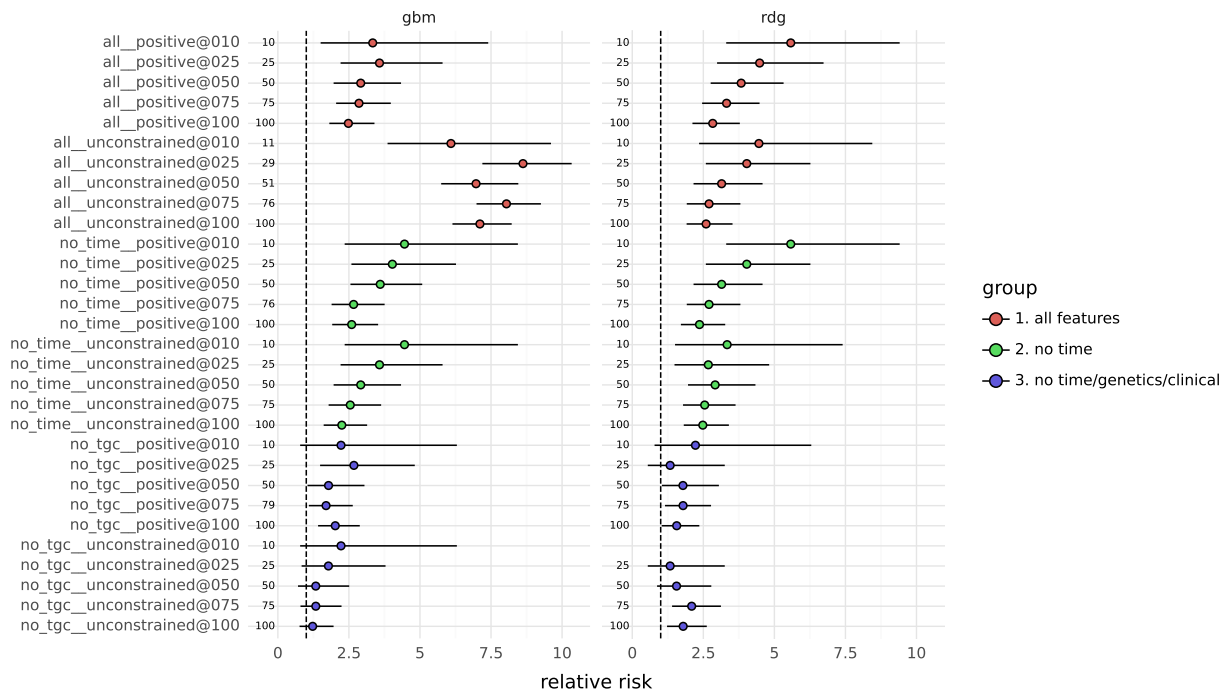


Figure 13: **Performance by algorithm, constraint type and feature grouping.**

| target_symbol | disease_name | prediction | target_clinical_phase_max_reached | disease_clinical_phase_max_reached | target_disease_genetic_association_curated | target_disease_genetic_association_eva | target_disease_genetic_association_of_genetics_portal | target_disease_genetic_association_omim | target_disease_genetic_association_genomics_england | target_disease_genetic_association_uniprot_literature | target_disease_genetic_association_uniprot_variants | target_disease_genetic_association_orphanet | target_disease_genetic_association_clingen | target_disease_genetic_association_gene2phenotype | target_disease_genetic_association_gene_burden |
|---------------|-------------------------------------------|------------|-----------------------------------|------------------------------------|--------------------------------------------|----------------------------------------|-------------------------------------------------------|-----------------------------------------|-----------------------------------------------------|-------------------------------------------------------|-----------------------------------------------------|---------------------------------------------|--------------------------------------------|---------------------------------------------------|------------------------------------------------|
| SNCA | Lewy body dementia | 0.669 | 2 | 4 | 1.0 | 0.9 | 0.7 | 0.9 | 1.0 | 1.0 | 1.0 | | | | |
| | systemic lupus erythematosus | 0.614 | 4 | 4 | 1.0 | 0.9 | 0.7 | | | 1.0 | | | | | |
| CTLA4 | Hashimoto's thyroiditis | 0.596 | 4 | 4 | 0.9 | 0.9 | 0.8 | | | | | | | | |
| FGFR2 | Apert syndrome | 0.577 | 4 | | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | |
| TNFSF4 | systemic lupus erythematosus | 0.574 | 2 | 4 | 1.0 | | 0.8 | | | 1.0 | | | | | |
| MAPT | Pick disease | 0.568 | 2 | 4 | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | | | | |
| VWF | von Willebrand disease 1 | 0.534 | 3 | | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | | | | |
| CTLA4 | celiac disease | 0.533 | 4 | 3 | 1.0 | 0.9 | 0.7 | 0.5 | 1.0 | 1.0 | | | | | |
| CSF1R | Hereditary diffuse leukoencephalopath... | 0.529 | 4 | | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | | | |
| CTLA4 | type 1 diabetes mellitus | 0.513 | 4 | 4 | 0.9 | 0.9 | 0.8 | | | | | | | | |
| VWF | von Willebrand disease 2 | 0.513 | 3 | | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | | | | |
| F2 | Congenital factor II deficiency | 0.504 | 4 | 1 | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | | | |
| HLA-DRB1 | rheumatoid arthritis | 0.500 | 2 | 4 | 1.0 | | 0.4 | | | 1.0 | | | | | |
| FGFR3 | hypochondroplasia | 0.490 | 4 | 2 | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | | 1.0 | 1.0 | 1.0 | |
| F11 | deep vein thrombosis | 0.489 | 4 | 4 | | | 0.9 | | | | | | | | |
| CTLA4 | autoimmune disease | 0.484 | 4 | 4 | 1.0 | | 0.8 | | 1.0 | | | | | | |
| FGFR2 | Crouzon syndrome | 0.482 | 4 | | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | |
| FGFR3 | campptodactyly-tall stature-scoliosis-... | 0.473 | 4 | | 1.0 | 0.9 | | 0.9 | 1.0 | | 1.0 | 1.0 | 0.5 | | |
| FGF23 | autosomal dominant hypophosphatemic r... | 0.472 | 4 | | 1.0 | 0.9 | | 0.9 | 1.0 | | 1.0 | 1.0 | | | |
| IL33 | allergic rhinitis | 0.471 | 3 | 4 | | | 0.7 | | | | | | | | 1.0 |
| IGF2 | type 2 diabetes mellitus | 0.466 | 2 | 4 | 0.9 | 0.9 | 0.6 | | | | | | | | |
| ITGB2 | Leukocyte adhesion deficiency type I | 0.464 | 4 | 1 | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | | | |
| FGFR3 | thanatophoric dysplasia type 1 | 0.463 | 4 | | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | |
| IL12B | inflammatory bowel disease | 0.458 | 4 | 4 | | | 0.8 | | | | | | | | |
| RPS19 | Diamond-Blackfan anemia | 0.458 | 4 | 2 | 1.0 | 0.9 | | | 1.0 | | | 1.0 | 1.0 | | |
| IL33 | infection | 0.457 | 3 | 4 | | | 0.8 | | | | | | | | |
| TACSTD2 | gelatinous drop-like corneal dystrophy | 0.457 | 4 | | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | | 1.0 | | 1.0 | |
| TNFSF11 | primary biliary cirrhosis | 0.456 | 4 | 4 | | | 0.8 | | | | | | | | |
| FGFR3 | Muenke syndrome | 0.456 | 4 | | 1.0 | 0.9 | | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | |
| TSLP | Nasal Cavity Polyp | 0.455 | 4 | 4 | | | 0.7 | | | | | | | | |

Figure 14: **Top ranked undeveloped, tractable target-disease pairs.** The highest scoring TD pairs per the RDG model that have not entered clinical trials despite having a target that has been in trials of an antibody-based drug. All values shown other than **prediction** are raw feature values, unweighted by RDG coefficients.

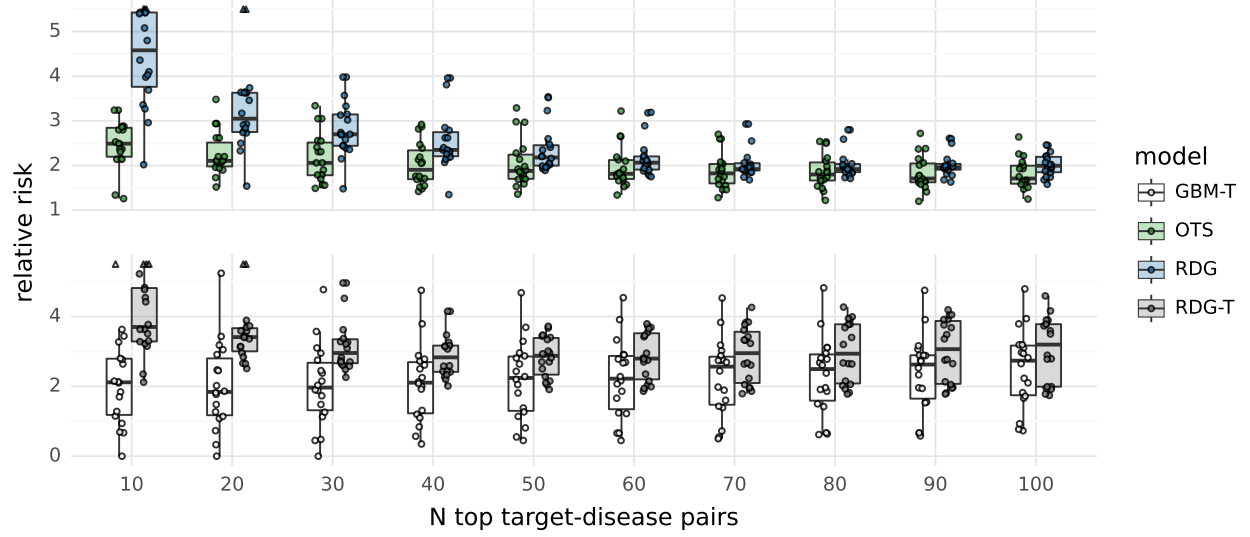


Figure 15: **Relative risk distributions across configurations in sensitivity analysis.** The distribution of the mean RR values displayed for a single configuration in Figure 1 is shown here across 18 configurations.

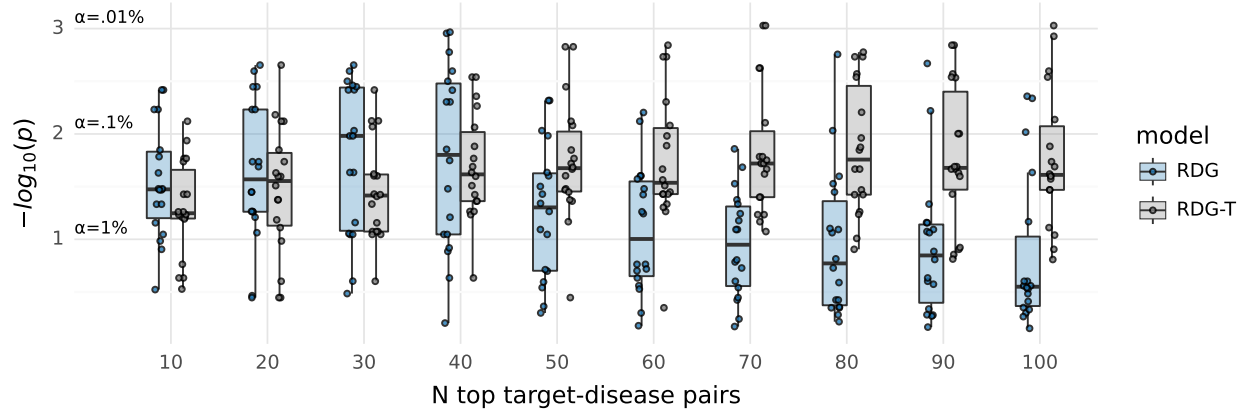


Figure 16: **P-value distributions across configurations in sensitivity analysis.** The distribution of the p-values displayed for a single configuration in Supplementary Figure 11 is shown here across 18 configurations.

References

- [1] Matthew R Nelson, Hannah Tipney, Jeffery L Painter, Judong Shen, Paola Nicoletti, Yufeng Shen, Aris Floratos, Pak Chung Sham, Mulin Jun Li, Junwen Wang, Lon R Cardon, John C Whittaker, and Philippe Sanseau. The support of human genetic evidence for approved drug indications. *Nat. Genet.*, 47(8):856–860, August 2015.
- [2] Emily A King, J Wade Davis, and Jacob F Degner. 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(12):e1008489, December 2019.
- [3] Eric Vallabh Minikel, Jeffery L Painter, Coco Chengliang Dong, and Matthew R. Nelson. Refining the impact of genetic evidence on clinical success. *medRxiv*, 2023.
- [4] Olesya Razuvayevskaya, Irene Lopez, Ian Dunham, and David Ochoa. Why clinical trials stop: The role of genetics. *medRxiv*, 2023.
- [5] Ben Kinnersley, Amit Sud, Elizabeth A Coker, Joseph E Tym, Patrizio Di Micco, Bissan Al-Lazikani, and Richard S Houlston. Leveraging human genetics to guide cancer drug development. *JCO clinical cancer informatics*, 2:1—11, December 2018.
- [6] David Cook, Dearn Brown, Robert Alexander, Ruth March, Paul Morgan, Gemma Satterthwaite, and Menelas N Pangalos. Lessons learned from the fate of astrazeneca’s drug pipeline: a five-dimensional framework. *Nature reviews. Drug discovery*, 13(6):419—431, June 2014.
- [7] David Ochoa, Mohd Karim, Maya Ghoussaini, David G Hulcoop, Ellen M McDonagh, and Ian Dunham. Human genetics evidence supports two-thirds of the 2021 fda-approved drugs. *Nature reviews. Drug discovery*, 21(8):551, August 2022.
- [8] Maya Ghoussaini, Matthew R Nelson, and Ian Dunham. Future prospects for human genetics and genomics in drug discovery. *Current opinion in structural biology*, 80:102568, June 2023.
- [9] Áine Duffy, Ben Omega Petrazzini, David Stein, Joshua K Park, Iain S Forrest, Kyle Gibson, Ha My Vy, Robert Chen, Carla Márquez-Luna, Matthew Mort, Marie Verbanck, Avner Schlessinger, Yuval Itan, David N Cooper, Ghislain Rocheleau, Daniel M Jordan, and Ron Do. Development of a human genetics-guided priority score for 19,365 genes and 399 drug indications. *Nature genetics*, 56(1):51—59, January 2024.
- [10] Gautier Koscielny, Peter An, Denise Carvalho-Silva, Jennifer A Cham, Luca Fumis, Rippa Gasparyan, Samiul Hasan, Nikiforos Karamanis, Michael Maguire, Eliseo Papa, Andrea Pierleoni, Miguel Pignatelli, Theo Platt, Francis Rowland, Priyanka Wankar, A Patrícia Bento, Tony Burdett, Antonio Fabregat, Simon Forbes, Anna Gaulton, Cristina Yenyxe Gonzalez, Henning Hermjakob, Anne Hersey, Steven Jupe, Şenay Kafkas, Maria Keays, Catherine Leroy, Francisco-Javier Lopez, Maria Paula Magarinos, James Malone, Johanna McEntyre, Alfonso Munoz-Pomer Fuentes, Claire O’Donovan, Irene Papatheodorou, Helen Parkinson, Barbara Palka, Justin Paschall, Robert Petryszak, Naruemon Pratanwanich, Sirarat Sarntivijal, Gary Saunders, Konstantinos Sidiropoulos, Thomas Smith, Zbyslaw Sondka, Oliver Stegle, Y Amy Tang, Edward Turner, Brendan Vaughan, Olga Vrousou, Xavier Watkins, Maria-Jesus Martin, Philippe Sanseau, Jessica Vamathevan, Ewan Birney, Jeffrey Barrett, and Ian Dunham. Open targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.*, 45(D1):D985–D994, January 2017.
- [11] Hai Fang, ULTRA-DD Consortium, Hans De Wolf, Bogdan Knezevic, Katie L Burnham, Julie Osgood, Anna Sanniti, Alicia Lledó Lara, Silva Kasela, Stephane De Cesco, Jörg K Wegner, Lahiru Handunnetthi, Fiona E McCann, Liye Chen, Takuya Sekine, Paul E Brennan, Brian D Marsden, David Damerell, Chris A O’Callaghan, Chas Bountra, Paul Bowness, Yvonne Sundström, Lili Milani, Louise Berg, Hinrich W Göhlmann, Pieter J Peeters, Benjamin P Fairfax, Michael Sundström, and Julian C Knight. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nature genetics*, 51(7):1082—1091, July 2019.
- [12] Aidan MacNamara, Nikolina Nakic, Ali Amin Al Olama, Cong Guo, Karsten B Sieber, Mark R Hurle, and Alex Gutteridge. Network and pathway expansion of genetic disease associations identifies successful drug targets. *Scientific reports*, 10(1):20970, December 2020.

- [13] Chaohui Bao, Hengru Wang, and Hai Fang. Genomic evidence supports the recognition of endometriosis as an inflammatory systemic disease and reveals disease-specific therapeutic potentials of targeting neutrophil degranulation. *Front. Immunol.*, 13:758440, March 2022.
- [14] Marie C Sadler, Chiara Auwerx, Patrick Deelen, and Zoltán Kutalik. Multi-layered genetic approaches to identify approved drug targets. *Cell Genom.*, 3(7):100341, July 2023.
- [15] Inigo Barrio-Hernandez and Pedro Beltrao. Network analysis of genome-wide association studies for drug target prioritisation. *Current opinion in chemical biology*, 71:102206, December 2022.
- [16] Inigo Barrio-Hernandez, Jeremy Schwartzentruber, Anjali Shrivastava, Noemi Del-Toro, Asier Gonzalez, Qian Zhang, Edward Mountjoy, Daniel Suveges, David Ochoa, Maya Ghoussaini, Glyn Bradley, Henning Hermjakob, Sandra Orchard, Ian Dunham, Carl A Anderson, Pablo Porras, and Pedro Beltrao. Network expansion of genetic associations defines a pleiotropy map of human cell biology. *Nature genetics*, 55(3):389—398, March 2023.
- [17] Saeed Paliwal, Alex de Giorgio, Daniel Neil, Jean-Baptiste Michel, and Alix Mb Lacoste. Preclinical validation of therapeutic targets predicted by tensor factorization on heterogeneous graphs. *Sci. Rep.*, 10(1):18250, October 2020.
- [18] Camilo Ruiz, Marinka Zitnik, and Jure Leskovec. Identification of disease treatment mechanisms through the multiscale interactome. *Nature communications*, 12(1):1796, March 2021.
- [19] Srivamshi Pittala, William Koehler, Jonathan Deans, Daniel Salinas, Martin Bringmann, Katharina Sophia Volz, and Berk Kapicioglu. Relation-weighted link prediction for disease gene identification, 2020.
- [20] Ozlem Muslu, Charles Tapley Hoyt, Mauricio Lacerda, Martin Hofmann-Apitius, and Holger Frohlich. Guiltytargets: Prioritization of novel therapeutic targets with network representation learning. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(1):491—500, 2022.
- [21] Petrina Kamyra, Ivan V Ozerov, Frank W Pun, Kyle Tretina, Tatyana Fokina, Shan Chen, Vladimir Naumov, Xi Long, Sha Lin, Mikhail Korzinkin, Daniil Polykovskiy, Alex Aliper, Feng Ren, and Alex Zhavoronkov. Pandaomics: An ai-driven platform for therapeutic target and biomarker discovery. *Journal of chemical information and modeling*, February 2024.
- [22] Alex Aliper, Roman Kudrin, Daniil Polykovskiy, Petrina Kamyra, Elena Tutubalina, Shan Chen, Feng Ren, and Alex Zhavoronkov. Prediction of clinical trials outcomes based on target choice and clinical trial design with multi-modal artificial intelligence. *Clinical pharmacology and therapeutics*, 114(5):972—980, November 2023.
- [23] Kien Wei Siah, Nicholas W Kelley, Steffen Ballerstedt, Björn Holzhauer, Tianmeng Lyu, David Mettler, Sophie Sun, Simon Wandel, Yang Zhong, Bin Zhou, Shifeng Pan, Yingyao Zhou, and Andrew W Lo. Predicting drug approvals: The novartis data science and artificial intelligence challenge. *Patterns (New York, N.Y.)*, 2(8):100312, August 2021.
- [24] Andrew W. Lo, Kien Wei Siah, and Chi Heem Wong. Machine Learning With Statistical Imputation for Predicting Drug Approvals. *Harvard Data Science Review*, 1(1), jul 1 2019. <https://hdsr.mitpress.mit.edu/pub/ct67j043>.
- [25] David Narganes-Carlón, Daniel J Crowther, and Ewan R Pearson. A publication-wide association study (pwas), historical language models to prioritise novel therapeutic drug targets. *Scientific reports*, 13(1):8366, May 2023.
- [26] Ada Hamosh, Alan F Scott, Joanna S Amberger, Carol A Bocchini, and Victor A McKusick. Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders. *Nucleic acids research*, 33(Database issue):D514—7, January 2005.
- [27] Melissa J Landrum, Jennifer M Lee, George R Riley, Wonhee Jang, Wendy S Rubinstein, Deanna M Church, and Donna R Maglott. Clinvar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research*, 42(Database issue):D980—5, January 2014.
- [28] FDA. Step 3: Clinical research — fda.gov. <https://www.fda.gov/patients/drug-development-process/step-3-clinical-research>, Apr 2018.

- [29] Chi Heem Wong, Kien Wei Siah, and Andrew W Lo. Estimation of clinical trial success rates and related parameters. *Biostatistics (Oxford, England)*, 20(2):273—286, April 2019.
- [30] Mathias Uhlén, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szigartyo, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per-Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M Schwenk, Marica Hamsten, Kalle von Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar von Heijne, Jens Nielsen, and Fredrik Pontén. Proteomics. tissue-based map of the human proteome. *Science (New York, N.Y.)*, 347(6220):1260419, January 2015.
- [31] Konrad J Karczewski, Laurent C Francioli, Grace Tiao, Beryl B Cummings, Jessica Alföldi, Qingbo Wang, Ryan L Collins, Kristen M Laricchia, Andrea Ganna, Daniel P Birnbaum, Laura D Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A Watts, Daniel Rhodes, Moriel Singer-Berk, Eleina M England, Eleanor G Seaby, Jack A Kosmicki, Raymond K Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X Chong, Kaitlin E Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H O’Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S Ware, Christopher Vittal, Irina M Armean, Louis Bergelson, Kristian Cibulskis, Kristen M Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto, Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade, Michael E Talkowski, Genome Aggregation Database Consortium, Benjamin M Neale, Mark J Daly, and Daniel G MacArthur. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434—443, May 2020.
- [32] Open targets platform 23.12 has been released! <https://blog.opentargets.org/open-targets-platform-23-12-release/#target-prioritisation>, Nov 2023.
- [33] Target - disease associations | Open Targets Platform Documentation. <https://platform-docs.opentargets.org/associations#data-source-weights>, 2023.
- [34] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [35] Eleonora Porcu, Marie C Sadler, Kaido Lepik, Chiara Auwerx, Andrew R Wood, Antoine Weihs, Maroun S Bou Sleiman, Diogo M Ribeiro, Stefania Bandinelli, Toshiko Tanaka, Matthias Nauck, Uwe Völker, Olivier Delaneau, Andres Metspalu, Alexander Teumer, Timothy Frayling, Federico A Santoni, Alexandre Reymond, and Zoltán Kutalik. Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nature communications*, 12(1):5647, September 2021.
- [36] Chris Finan, Anna Gaulton, Felix A Kruger, R Thomas Lumbers, Tina Shah, Jorgen Engmann, Luana Galver, Ryan Kelley, Anneli Karlsson, Rita Santos, John P Overington, Aroon D Hingorani, and Juan P Casas. The druggable genome and support for target identification and validation in drug development. *Science translational medicine*, 9(383):eaag1166, March 2017.
- [37] Chaohui Bao, Hengru Wang, and Hai Fang. Genomic evidence supports the recognition of endometriosis as an inflammatory systemic disease and reveals disease-specific therapeutic potentials of targeting neutrophil degranulation. *Frontiers in immunology*, 13:758440, 2022.
- [38] Open Targets Tractability Pipeline (version 2). https://github.com/chembl/tractability_pipeline_v2, Jul 2023.
- [39] Evan A Boyle, Yang I Li, and Jonathan K Pritchard. An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169(7):1177—1186, June 2017.

- [40] Stephen Burgess, Amy M Mason, Andrew J Grant, Eric A W Slob, Apostolos Gkatzionis, Verena Zuber, Ashish Patel, Haodong Tian, Cunhao Liu, William G Haynes, G Kees Hovingh, Lotte Bjerre Knudsen, John C Whittaker, and Dipender Gill. Using genetic association data to guide drug discovery and development: Review of methods and applications. *American journal of human genetics*, 110(2):195—214, February 2023.
- [41] Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L Gable, Tao Fang, Nadezhda T Doncheva, Sampo Pyysalo, Peer Bork, Lars J Jensen, and Christian von Mering. The string database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research*, 51(D1):D638—D646, January 2023.
- [42] Robert W Hoffman, Joan T Merrill, Marta M E Alarcón-Riquelme, Michelle Petri, Ernst R Dow, Eric Nantz, Laura K Nisenbaum, Krista M Schroeder, Wendy J Komocsar, Narayanan B Perumal, Matthew D Linnik, David C Airey, Yushi Liu, Guilherme V Rocha, and Richard E Higgs. Gene expression and pharmacodynamic changes in 1,760 systemic lupus erythematosus patients from two phase iii trials of baf blockades with tabalumab. *Arthritis and rheumatology (Hoboken, N.J.)*, 69(3):643—654, March 2017.
- [43] Christian Fougner, Julie Cannon, Lydia The, Jeffrey F Smith, and Olivier Leclerc. Herding in the drug development pipeline. *Nature reviews. Drug discovery*, 22(8):617—618, August 2023.
- [44] Winston A Haynes, Aurelie Tomczak, and Purvesh Khatri. Gene annotation bias impedes biomedical research. *Scientific reports*, 8(1):1362, January 2018.
- [45] Rita Santos, Oleg Ursu, Anna Gaulton, A Patrícia Bento, Ramesh S Donadi, Cristian G Bologa, Anneli Karlsson, Bissan Al-Lazikani, Anne Hersey, Tudor I Oprea, and John P Overington. A comprehensive map of molecular drug targets. *Nature reviews. Drug discovery*, 16(1):19—34, January 2017.
- [46] Chris Finan, Anna Gaulton, Felix Kruger, Tom Lumbers, Tina Shah, Jorgen Engmann, Luana Galver, Ryan Kelly, Anneli Karlsson, Rita Santos, John Overington, Aroon Hingorani, and Juan Pablo Casas. The druggable genome and support for target identification and validation in drug development, 2016.
- [47] Arielle Marks-Anglin and Yong Chen. A historical review of publication bias. *Research synthesis methods*, 11(6):725—742, November 2020.
- [48] Luca Abatangelo, Rosalia Maglietta, Angela Distaso, Annarita D’Addabbo, Teresa Maria Creanza, Sayan Mukherjee, and Nicola Ancona. Comparative study of gene set enrichment methods. *BMC bioinformatics*, 10:275, September 2009.
- [49] Katerina Trajanoska, Claude Bhérer, Daniel Taliun, Sirui Zhou, J Brent Richards, and Vincent Mooser. From target discovery to clinical drug development with human genetics. *Nature*, 620(7975):737—745, August 2023.
- [50] Ruth J F Loos. 15 years of genome-wide association studies and no signs of slowing down. *Nature communications*, 11(1):5900, November 2020.
- [51] Abdel Abdellaoui, Loic Yengo, Karin J H Verweij, and Peter M Visscher. 15 years of gwas discovery: Realizing the promise. *American journal of human genetics*, 110(2):179—194, February 2023.
- [52] Michael J Bamshad, Deborah A Nickerson, and Jessica X Chong. Mendelian gene discovery: Fast and furious with no end in sight. *American journal of human genetics*, 105(3):448—455, September 2019.
- [53] Catherine S Storm, Demis A Kia, Mona M Almramhi, Sara Bandres-Ciga, Chris Finan, International Parkinson’s Disease Genomics Consortium (IPDGC), Aroon D Hingorani, and Nicholas W Wood. Finding genetically-supported drug targets for parkinson’s disease using mendelian randomization of the druggable genome. *Nature communications*, 12(1):7342, December 2021.