Impact of proteogenomic evidence on clinical success

Authors

Mohd Anisul Karim*[1,2], Abhay Hukku[3], Bruno Ariano[2], Emily Holzinger[3], Yakov Tsepilov[2], James Hayhurst[4], Annalisa Buniello[4], Ellen M. McDonagh[2,4], Stephane E Castel[1], Matthew R. Nelson[5,6], Joseph Maranville[3], Laura Yerges-Armstrong[1], Maya Ghoussaini*[2]

Affiliations

1. Variant Bio, Seattle, 98109, WA, USA
2. Open Targets, Wellcome Sanger Institute, Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK
3. Bristol-Myers Squibb, Cambridge, MA 02142, United States
4. Open Targets, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK
5. Deerfield Management Company LP, New York, NY, USA
6. Genscience LLC, New York, NY, USA

* To whom correspondence should be addressed:

Mohd Karim: mohd@variantbio.com

Maya Ghoussaini: mayaghoussainy@hotmail.com

Abstract

We assessed the impact of plasma protein quantitative trait loci (pQTL) on therapeutic hypotheses backed by human genetic evidence. We show that pQTL-supported target-indication pairs were 4.7 times more likely to advance from Phase I to launch, compared to a 2.6-fold increase observed with human genetic evidence lacking pQTL support. Moreover, pQTL-based enrichment was prominent in druggable protein families which had limited enrichment from human genetic evidence alone.

Main

Suboptimal therapeutic efficacy is a key reason for the failure of a drug for its selected disease indication. Several reports have shown that drug targets matched to disease indications via human genetic evidence have at least a two-fold higher probability of clinical success[1–3]. Genome-wide association studies (GWAS) of diseases are a major source of human genetic evidence but genome-wide association signals often span multiple genes, limiting the ability of GWAS alone to assign causal genes with confidence[4]. An approach to assign causal genes with higher confidence is to evaluate whether the disease-associated genetic signal colocalizes with genetic signals from intermediate molecular traits such as plasma proteins. Indeed the probability of clinical success of drug programs rises with increasing confidence that a particular gene is causal for disease[1]. Prior studies have mostly examined the enrichment of pQTLs in druggable targets to uncover new therapeutic opportunities[5,6]. A more recent study showed clinical success enrichment with cis-QTLs broadly (pQTL & eQTL)[7]. It was, however, unclear from this cis-QTL focused analysis whether there was added value of pQTLs to standard human genetic evidence in terms of higher clinical success.

Here, we attempt to expand on these studies including our own previous work[8] by integrating eight publicly available plasma proteogenomic datasets (**Supplementary Table 1**) with a broad collection of over 8000 complex traits from GWAS Catalog, Pan-UK Biobank, and FinnGen (**Supplementary Table 2**) and a recent large curated set of 29,476 drug target–indication (T–I) pairs from Citeline Pharmaprojects (Minikel et al). To assign genes to indications matching GWAS traits, Minikel et al derived a locus-to-gene (L2G) share which is the proportion of total L2G score assigned to a gene from among all possible locus-gene-trait triplets in Open Targets Genetics (version 8). The L2G score is derived from a trained supervised learning model that integrates multiple lines of evidence including genomic distance, chromatin interaction data, and functional annotations, especially cis-QTLs including cis-pQTLs from UKB-PPP[9], to predict causal genes at GWAS loci[4]. Because our analyses are not limited to cis-pQTLs from a single study, we investigated whether having broader pQTL-support can potentially complement L2G-based therapeutic hypotheses.

To ensure comparability of pQTL-supported genetic hypotheses with other sources of human genetic evidence, we used the same MeSH (MeSH similarity score ≥ 0.8) and L2G thresholds (L2G share ≥ 0.5) as Minikel et al in our main analysis, but varied the L2G thresholds by ±0.25 increments to evaluate the added value of pQTL-supported hypotheses. To generate pQTL-supported therapeutic hypotheses, we first performed a large-scale systematic MR - 47.2M tests - using genome-wide significant genetic instruments associated with plasma protein levels and ~8000 GWAS traits. We merged these significant MR target-trait associations with gene–indication pairs from Minikel et al's T-I dataset and retained pairs where our MR outcomes matched their known drug indications (MeSH similarity ≥ 0.8). For the enrichment analysis comparing pQTL-supported versus unsupported hypotheses, we further restricted to target-indication pairs with both MeSH similarity ≥ 0.8 and L2G share ≥ 0.5 within the background of proteomically measured genes, yielding 46 pQTL-supported target-indication pairs at Phase I used for relative success calculation (**Supplementary Figure S1)**.

In our enrichment analyses, we found that pQTL-supported target-indication pairs showed a substantial enrichment for clinical success (Phase I to Launch), with a relative success (RS) of 4.73 (95% CI: 3.51, 6.36); this is broadly comparable to OMIM-based evidence (RS: 3.67, 95% CI: 3.07, 4.37) (**Fig. 1a, Supplementary Table 4**). Notably, T–I pairs with both pQTL support and an L2G share of ≥0.75 achieved an RS of 5.65 (95% CI: 4.18, 7.62), more than double the RS for L2G ≥0.75 alone without pQTL support (RS of 2.60; 95% CI: 2.17, 3.13), suggesting added value of pQTLs to prioritise therapeutic hypotheses.

When we examined across MR-coloc subtypes, we found similar enrichment estimates with overlapping confidence intervals from mixed-, cis- or trans-MR, including a subset with cis-/trans-colocalization support (**Fig. 1a**). The similar RS across MR-coloc subtypes in our results and in previous studies motivated us to examine whether this pattern reflected overlapping evidence streams rather than independent predictive power. Indeed, most pQTL-supported launched T-I pairs were supported by two or more MR-coloc types, often combining cis, trans, and mixed MR with cis or trans colocalization, with intersection sizes up to four (**Fig. 1b, Supplementary Table 5**), highlighting the value of integrating evidence from multiple MR-coloc subtypes when prioritizing therapeutic hypotheses.

We also investigated whether enrichment was concentrated in certain druggable families using protein family annotations[10]. Under L2G alone, many well-known protein families like enzymes and kinases showed little or no enrichment, but when combined with pQTL support, they exhibited the largest RS increases (**Fig. 1c, Supplementary Table 6**). However, several medically important families (rhodopsin-class GPCRs, nuclear receptors, ion channels and many transporters) are poorly represented on current high-throughput Olink and Somascan panels, and in our data these classes have null enrichment; we suspect inadequate assay coverage (rather than absence of true biology) explains at least part of this gap. These results also imply that a

straightforward way to boost genetic target validation is to expand and diversify proteomic panels by increasing coverage of existing panels for these under-represented classes or by employing novel methods to quantify low-abundance proteins and protein isoforms[11,12].

We examined whether pQTL-supported enrichment varied by therapeutic area or clinical phase (**Supplementary Figures S2–S3 and Tables 7-15**). Although pQTL-supported target–indication pairs were unevenly distributed across therapeutic areas, enrichment estimates were broadly consistent across disease categories and phases, with some evidence of stronger enrichment for later-stage transitions. Notably, pQTL-supported target–indication pairs were not preferentially enriched in therapeutic areas with higher baseline success rates, arguing against confounding by disease category (**Supplementary Table 10-12**).

A by-product of this work is a browsable database of target–trait pairs supported at a liberal pQTL MR FDR-corrected threshold ($p < 0.05$), available at https://mk31.shinyapps.io/pqtl_mr_fdr05/. Given the liberal threshold, false positive associations are expected; we therefore provide extensive downstream annotations to support follow-up and prioritization (**Online Methods**).

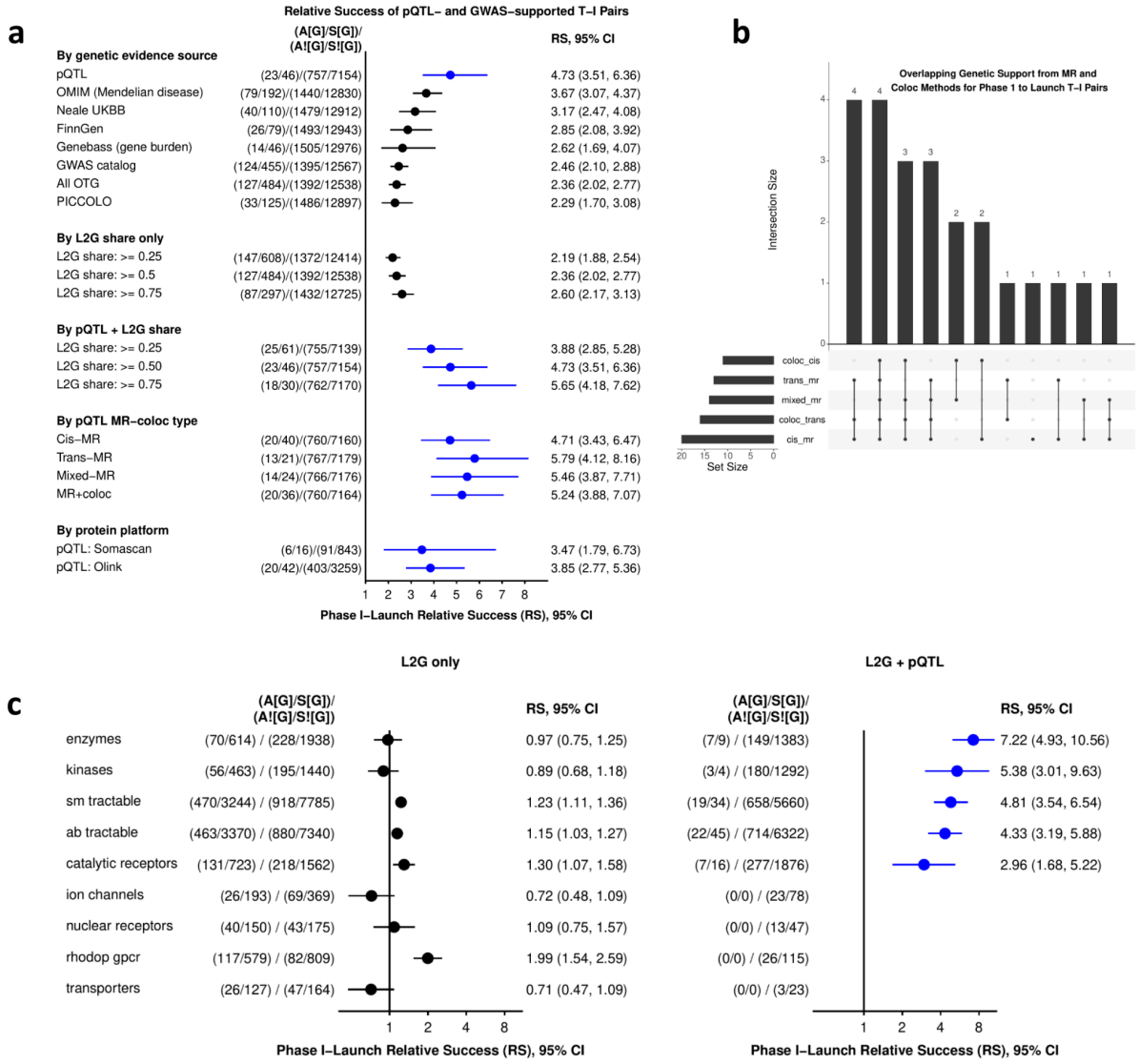
Using the combined evidence framework, we note several pQTL-based target-indication pairs, where the targets have a low or absent L2G score from Open Targets Genetics but are supported by other orthogonal sources of evidence (**Supplementary Table 16**). For example, genetically predicted higher tumor necrosis factor (TNF) was associated with increased risk of ankylosing spondylitis, consistent with the clinical efficacy of TNF inhibitors, yet this association is absent from Open Targets Genetics L2G predictions due to the exclusion of the complicated HLA/MHC region which is difficult to include in the post-GWAS analysis pipeline. Although this exclusion reduces confounding due to complex linkage disequilibrium, it can also remove informative associations from nearby genes such as *TNF*. We also observed non-HLA examples illustrating similar complementarity, for example *SOST* with osteoporosis, where MR evidence is supported by mechanistic data and approved therapies despite modest L2G scores in Open Targets Genetics[13,14], highlighting the importance of MR as a complementary line of evidence for revealing likely causal genes for disease association.

We also point out associations, especially trans-associations, where therapeutic directionality inference in the absence of other supporting data can be misleading. For example, we noted a trans-colocalizing signal 17_39871710_C_G - near *CSF3* (colony stimulating factor 3) gene - which was associated with higher plasma CSF3R (colony stimulating factor 3 receptor) levels and higher risk of neutropenia inconsistent with the function of CSF3R which typically increases granulocyte counts[15]. The CSF3R-neutropenia relationship was, however, consistent with CSF3R function when considering cis-acting variants (i.e. variants near the *CSF3R* gene). This highlights the need to consider therapeutic direction inferred from trans-acting variants in context

of their cis-acting counterparts since trans-signals may represent complex biological feedback mechanisms and, additionally, plasma protein levels may not reflect the cellular or tissue-specific context in which a target exerts its therapeutic effect.

There are several limitations to our work. First, the 23 pQTL-supported and launched T-I pairs supporting the current RS estimates are driven by only 12 targets, reflecting the strong dependence of proteogenomic validation on protein assay coverage; broader and more diverse protein panels are likely to expand this set. Second, some cis pQTL associations may be influenced by protein-altering variants that affect aptamer- or antibody-based measurements, potentially complicating direction-of-effect interpretation, although we annotate such variants to support cautious follow-up (**Supplementary Table 16**). Third, the underlying MR and colocalization analyses rely primarily on common variants from European-ancestry cohorts and on circulating plasma proteins, which may not reflect disease-relevant tissues and limit generalizability. Finally, we did not assess eQTL-based enrichment, as most available eQTL resources are limited to cis effects and are not directly comparable to the genome-wide pQTL framework used here.

Figure 1 - Enrichment analyses



a. Phase I–Launch relative success (RS) of target–indication (T–I) pairs, comparing pQTL-based evidence (blue) with non-pQTL evidence (black) by genetic evidence source, L2G share alone, L2G share plus pQTL evidence, MR-coloc type, and protein assay platform. pQTL-based RS used T–I pairs with L2G share ≥ 0.5 and Bonferroni-significant pQTL support ($p < 1.06 \times 10^{-9}$). Non-pQTL evidence sources and L2G definitions followed Minikel *et al.* (**b**) Upset plot showing clinically successful T–I pairs often have multiple MR-coloc support types (cis, trans, mixed). **c**. Phase I–Launch RS by protein family (Minikel classification), comparing L2G share ≥ 0.5 vs L2G share ≥ 0.5 + pQTL evidence; genes could belong to multiple families. A[G] = genetically supported T–I pairs that launched (among those entering Phase I); S[G] = genetically supported T–I pairs entering Phase I; A![G] = genetically unsupported T–I pairs that launched (among those entering Phase I); S![G] = genetically unsupported T–I pairs entering Phase I. Error bars: 95% CI.

Online Methods

We used eight publicly accessible proteogenomic datasets (**Supplementary Table 1)** for the pan-MR analyses using GSMR enabling selection of genetic instruments from across the genome; regional definitions were applied post-GSMR, i.e. cis-, defined as $\pm$1Mbp from transcription start site, trans-, or mixed-MR. The MR associations were expressed per standard deviation (SD) higher genetically predicted plasma protein concentrations. A subset of MR associations (FDR-corrected p-value < 0.05) also underwent genetic colocalization tests. Colocalization support was defined as H4 ≥ 0.8, recorded separately for cis and trans loci. We combined complex trait GWAS selected as outcomes from our previous release[8] with new traits from FinnGen (release 12), pan-UK Biobank, and GWAS Catalog, using the same criteria we used previously, i.e. - if they had at least one genome-wide significant ($p \leq 5 \times 10^{-8}$) locus, ensuring a systematic disease-agnostic trait selection strategy. This resulted in a total 8762 studies (**Supplementary Table 2)**. A list of studies that were removed due to their lack of medical relevance or redundant protein associations is provided in **Supplementary Table 3**. Details of genotyping protocols and QC of proteomic studies have been described previously in the respective studies' publications. Details of liftOver and harmonization of both proteomic and outcome studies are provided in our previous preprint[8].

For comparability of pQTL-based enrichment estimates with other sources of genetic evidence, we followed Minikel et al's approach. We linked targets to indications using the MeSH term based similarity matrix and kept target-indication (T-I) pairs with a similarity score ≥ 0.8. For associations from Open Targets Genetics (version 8), including the traits used in our pQTL MR analyses, we additionally required an L2G share of at least 0.5 (L2G share is the L2G score of a gene divided by the sum of L2G scores across all candidate genes for that trait–loci combination from Open Targets Genetics[16]). A target–indication (T-I) pair was considered pQTL-supported if the MR p-value passed a Bonferroni threshold ($p \leq 0.05 / 4.7\times10^{7}$, $\sim1.06\times10^{-9}$) and the pair satisfied the MeSH and L2G filters. A stricter "MR+coloc" group also required a cis- or trans-colocalization posterior probability $H_4 \geq 0.8$. For each T-I pair we selected the row with the highest clinical phase and, if needed, breaking ties by the MeSH similarity score. Using this set, we computed Phase I to launch relative success as RS = (A[G]/S[G]) / (A[!G]/S[!G]), where S is the number of Phase I entries and A the number of launches in the genetically supported (G) and unsupported (!G) groups. Confidence intervals for RS were obtained using the Katz log method, and where p-values are reported, they were calculated using Fisher's exact test. For pQTL-based analyses, we restricted the background set to Phase I T–I pairs whose target protein was measured on proteomic platforms of the respective pQTL studies (e.g. Olink). For other genetic evidence sources, the background consisted of Phase I T-I pairs that lacked that specific source of support as defined by Minikel et al.

To assess whether pQTL-based enrichment was confounded by the therapeutic area, we performed several sensitivity analyses. First, we tested whether pQTL evidence was

disproportionately distributed across therapeutic areas using a chi-square test (**Supplementary Table 7**). Second, we evaluated whether pQTL support was preferentially found in therapeutic areas with high baseline success rates using Spearman rank correlation between each therapeutic area's baseline success rate (among pQTL-unsupported targets) and the proportion of target-indication pairs with pQTL support (**Supplementary Table 10**). Third, we applied the Breslow-Day test for homogeneity of odds ratios across therapeutic area strata to assess whether the pQTL effect was consistent across disease categories (**Supplementary Table 13**). These tests were performed using two background universes: (1) all Phase I-entered target-indication pairs where the gene was measurable on proteomic platforms (Olink/SomaScan), and (2) the subset of these with existing genetic evidence from Open Targets Genetics, OMIM, PICCOLO, or Genebass (**Supplementary Tables 8-9, 11-12, 14-15**). As an additional sensitivity analysis, we also computed pQTL enrichment using the full Minikel et al. T-I universe without restricting to measured proteins (**Supplementary Figure S4**). Finally, we performed leave-one-therapeutic-area-out analyses to evaluate whether overall RS estimates were sensitive to removal of any single disease category (**Supplementary Tables 14-15**).

To assess whether cis-association signals were protein-altering variants (PAVs) or in linkage disequilibrium with PAVs, we used the TOP-LD tool[17]. Starting from Bonferroni-significant cis-MR index SNPs, we queried TOP-LD in Europeans ($r^2 \geq 0.6$, MAF ≥ 0.01) and extracted linked variants, their annotated genes, and Sequence Ontology (SO) consequences. We classified a variant as protein-altering if it carried a HIGH or MODERATE impact SO term[18]. For trans-associations, we recorded the nearest gene(s) for each trans index SNP from Open Targets. We also annotated replication across proteomic platforms and outcome datasets, the different combinations of MR and colocalization methods supporting each target-trait pair, and non-genetic evidence sources (for example, RNA expression, animal models, and literature) using Open Targets Platform[19] (https://platform.opentargets.org/, version 25.09.0) data types. Finally, we overlaid Pharmaprojects and ChEMBL to flag known drug targets, positive control T-I pairs (existing programmes for the same indication), and potential repositioning opportunities (known targets without programmes for the MR-inferred indication). Further implementation details, including data processing, harmonisation, and all thresholds used in enrichment and annotation analyses, are provided in the **Supplementary Methods.** All analyses and annotations, unless otherwise specified, were done in R version 4.3.3.

Acknowledgements

We would like to thank Eric Vallabh Minikel for reviewing preliminary analyses and figures related to this work.

Data availability

Summary data for both proteins and outcomes used for genetic analyses are publicly available from the GWAS catalog https://www.ebi.ac.uk/gwas/downloads/summary-statistics. Data on clinical precedence for approved and investigational drugs and representative target-indication pairs with L2G share, were obtained from the supplementary materials of Minikel et al. (2024).

MR and genetic colocalization results can be downloaded here:
https://doi.org/10.5281/zenodo.18451758

Open Targets Genetics L2G data version 8 is available from the EMBL-EBI FTP server:
https://ftp.ebi.ac.uk/pub/databases/opentargets/genetics/

Filtered results (MR FDR-corrected p-value < 0.05) can be browsed here:
https://mk31.shinyapps.io/pqtl_mr_fdr05/

Code availability

https://github.com/opentargets/mendelian-randomisation
https://github.com/mohdkarim/mrcoloc

Funding statement

MAK, BA, YT, JH, AB, EMM, and MG were funded by Open Targets. This research was funded in part by a Wellcome Trust [Grant number 206194]. For the purpose of Open Access, the authors have applied a CC-BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Author contributions

MAK, MG, and JM conceived the study. MAK, AH, and BA collaboratively performed all analyses. MG supervised the research. JH and AB worked on curation, integration and harmonisation of the datasets for the GWAS Catalog. MAK and MG drafted the manuscript. All authors provided valuable feedback and critical comments that informed the design, analyses and interpretation of the study and reviewed and approved the final manuscript.

Competing interests

MAK, SEC, and LYA are employees of Variant Bio. AH, EH, and JM are employees of Bristol-Myers Squibb. MG is an employee of Regeneron Pharmaceuticals. MRN is an employee of Deerfield and Genscience.

Ethics declarations

All institutions contributing cohorts to the proteomics and outcome studies received ethics approval from their respective research ethics review boards.

References

1. Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic evidence on clinical success. *Nature* **629**, 624–629 (2024).
2. King, E. A., Davis, J. W. & Degner, J. F. 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 (2019).
3. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nature Genetics* **47**, 856–860 (2015).
4. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet* **53**, 1527–1533 (2021).
5. Deng, Y.-T. *et al.* Atlas of the plasma proteome in health and disease in 53,026 adults. *Cell* **188**, 253–271.e7 (2025).
6. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* **52**, 1122–1131 (2020).
7. Ferolito, B. R. *et al.* Leveraging Large-Scale Biobanks for Therapeutic Target Discovery. *medRxiv* 2025.02.10.25321487 (2025) doi:10.1101/2025.02.10.25321487.
8. Karim, M. A. *et al.* Systematic disease-agnostic identification of therapeutically actionable targets using the genetics of human plasma proteins. *medRxiv* 2023.06.01.23290252 (2023) doi:10.1101/2023.06.01.23290252.
9. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, (2023).

10. Minikel, E. V. *et al.* Evaluating drug targets through human loss-of-function genetic variation. *Nature* **581**, (2020).

11. Lu, C., Bonini, A., Viel, J. H. & Maglia, G. Toward single-molecule protein sequencing using nanopores. *Nat Biotechnol* **43**, 312–322 (2025).

12. Motone, K. *et al.* Multi-pass, single-molecule nanopore reading of long protein strands. *Nature* **633**, 662–669 (2024).

13. Döring, Y. *et al.* CXCL12 Derived From Endothelial Cells Promotes Atherosclerosis to Drive Coronary Artery Disease. *Circulation* **139**, 1338–1340 (2019).

14. Aditya, S. & Rattan, A. Sclerostin Inhibition: A Novel Target for the Treatment of Postmenopausal Osteoporosis. *J Midlife Health* **12**, 267–275 (2021).

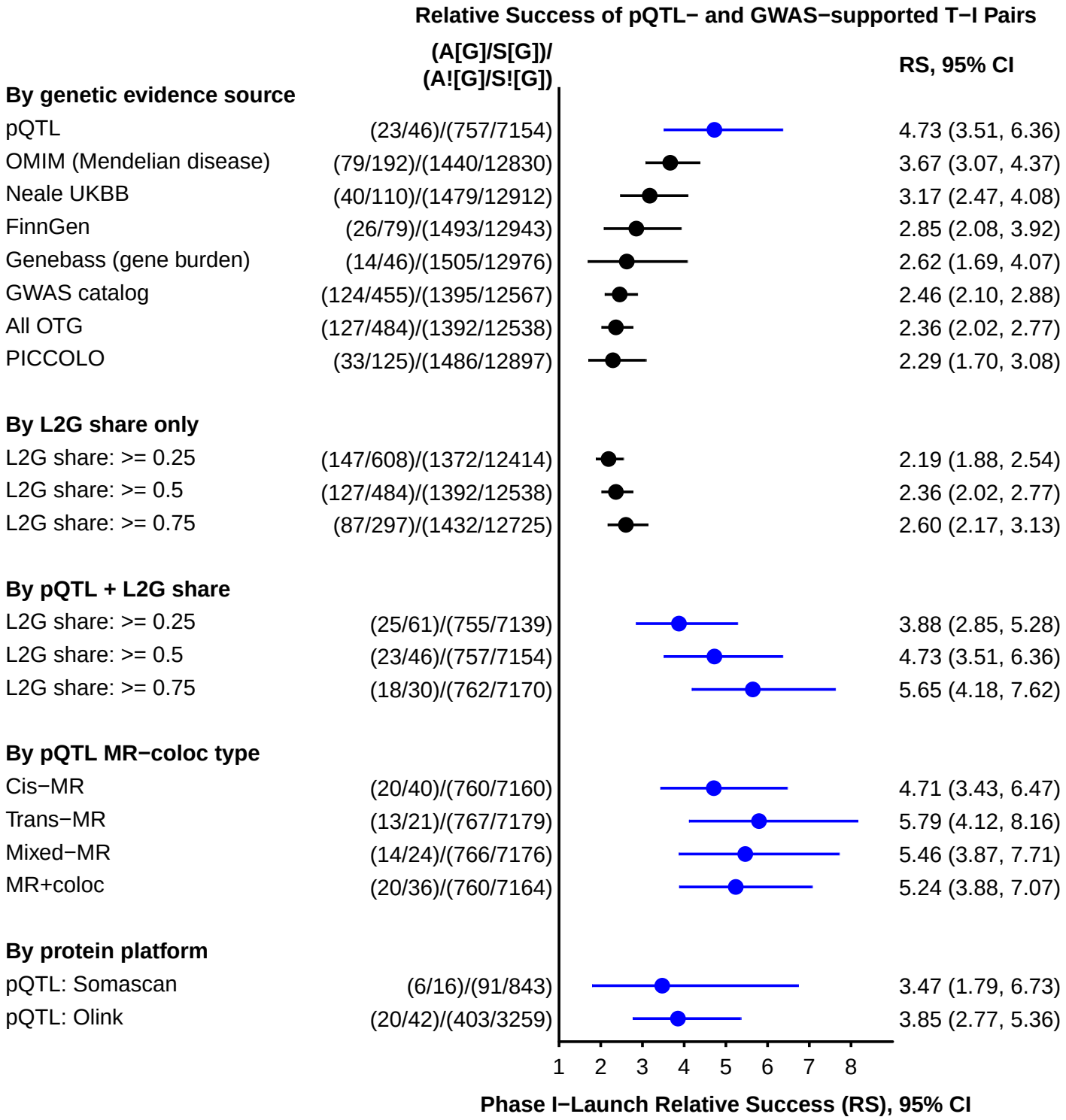
15. Dwivedi, P. & Greis, K. D. Granulocyte colony-stimulating factor receptor signaling in severe congenital neutropenia, chronic neutrophilic leukemia, and related malignancies. *Exp Hematol* **46**, 9–20 (2017).

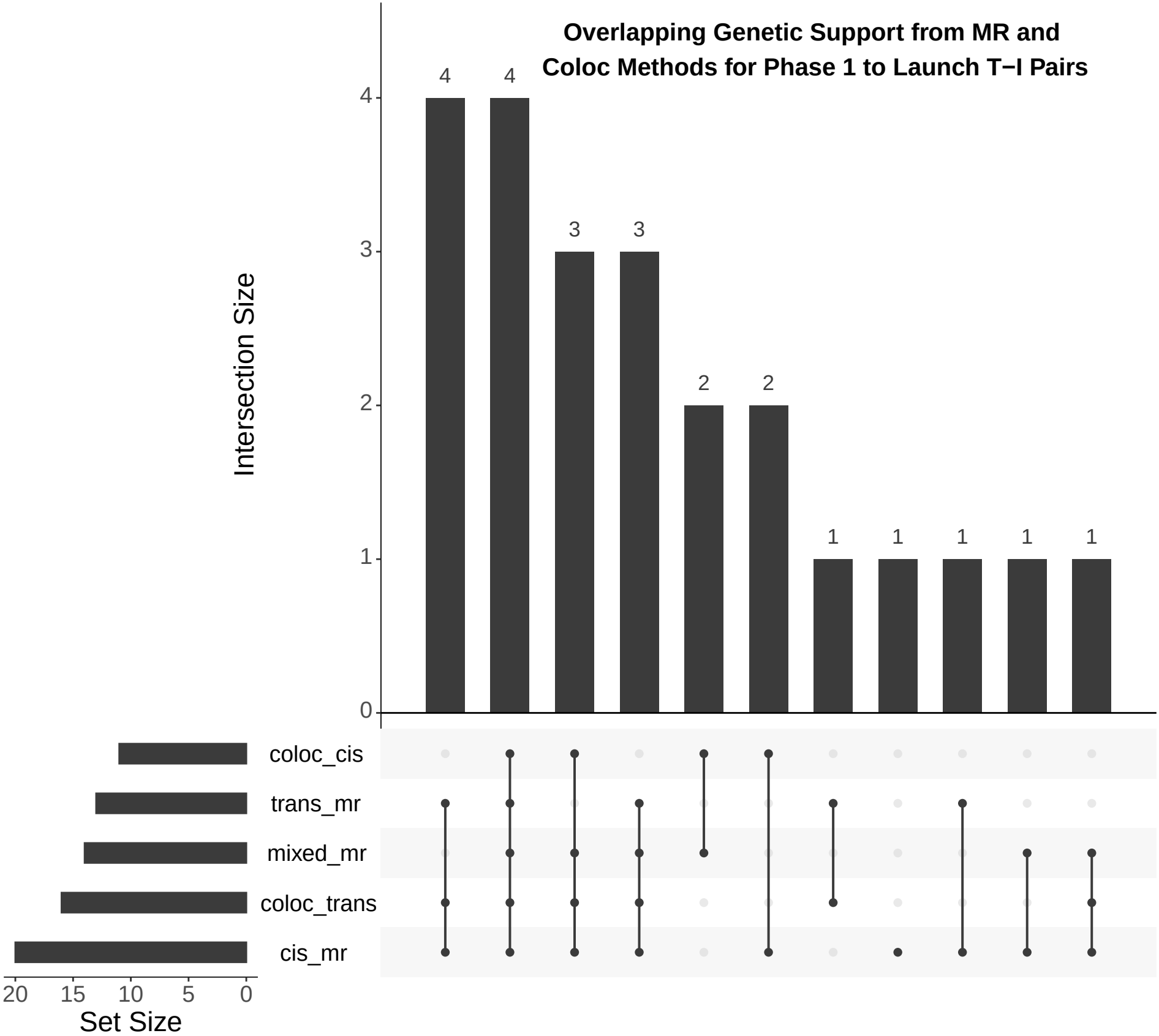
16. Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res* **49**, D1311–D1320 (2021).

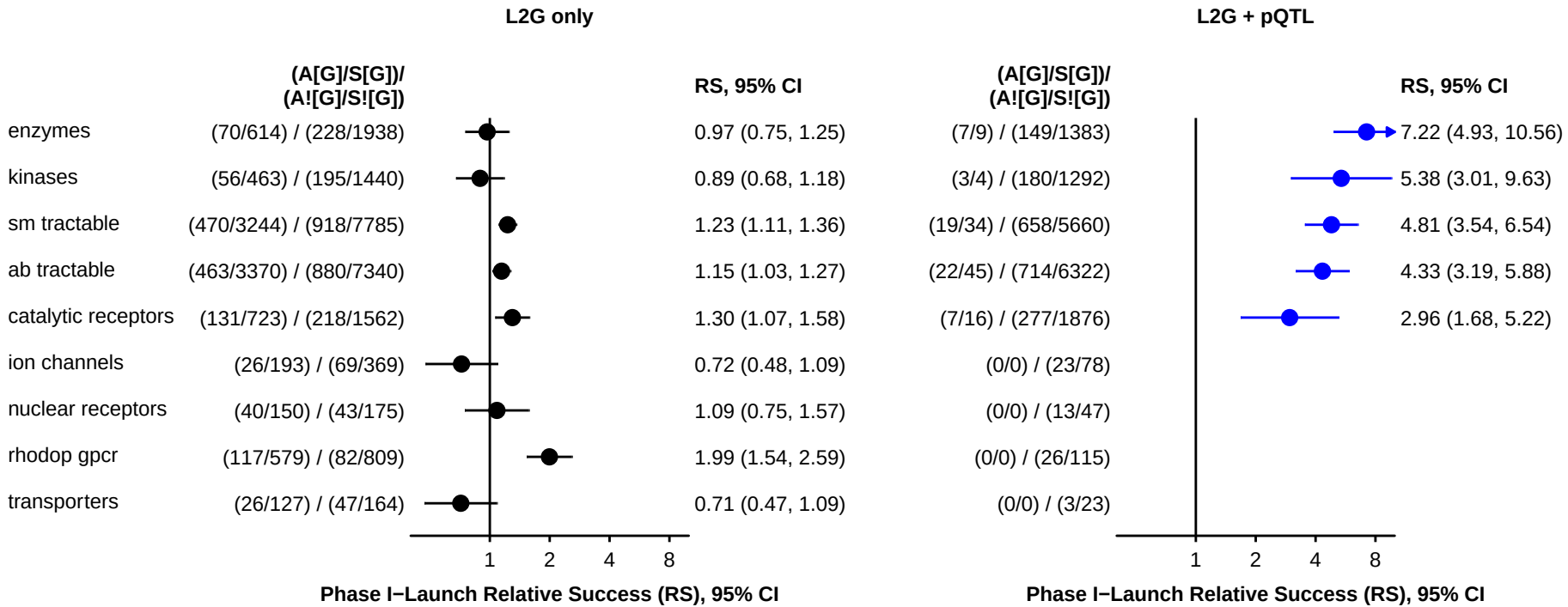
17. Huang, L. *et al.* TOP-LD: A tool to explore linkage disequilibrium with TOPMed whole-genome sequence data. *Am J Hum Genet* **109**, 1175–1181 (2022).

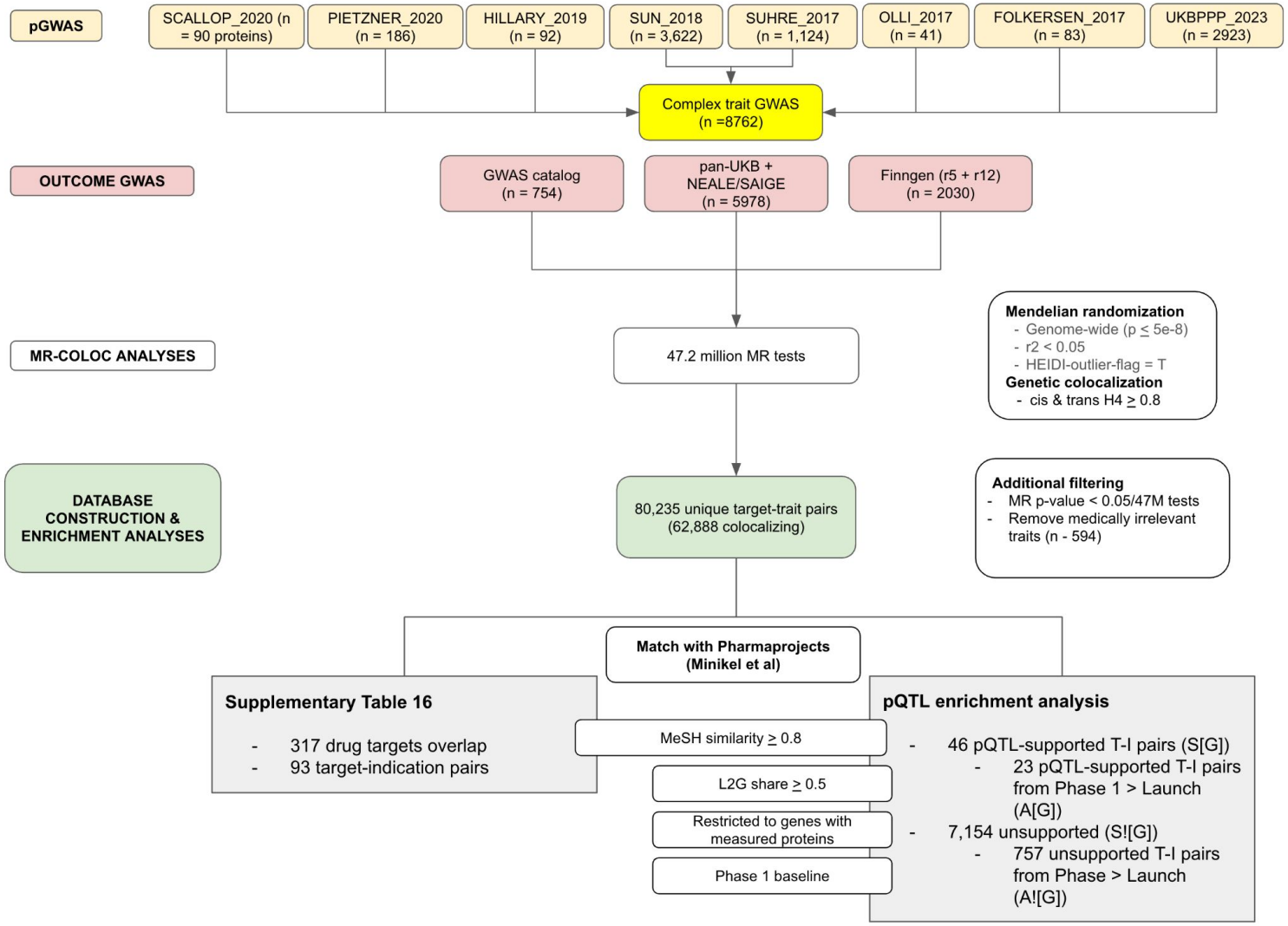
18. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).

19. Buniello, A. *et al.* Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res* **53**, D1467–D1475 (2025).

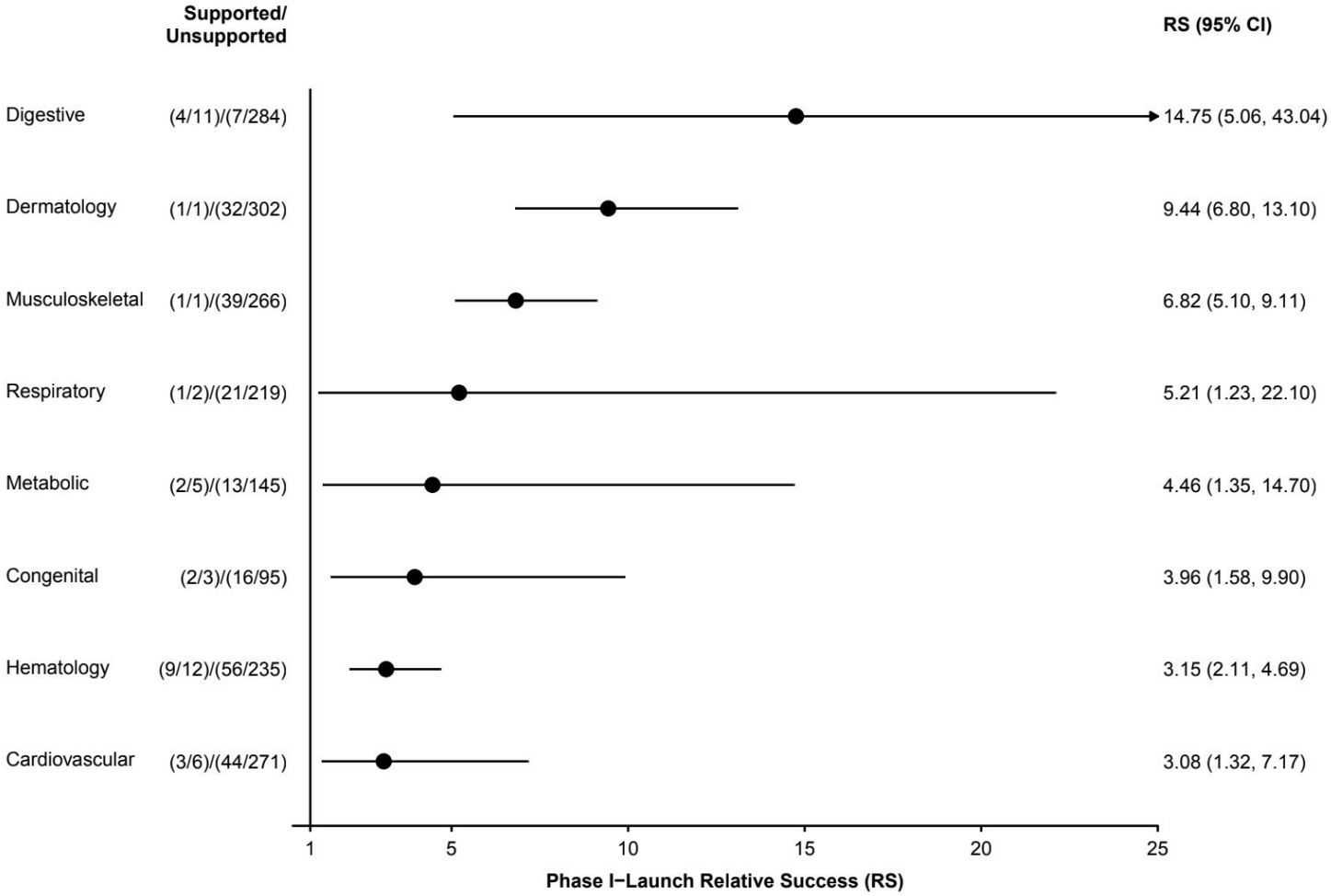

Relative Success of pQTL− and GWAS−supported T−I Pairs
(A[G]/S[G])/
(A![G]/S![G])
RS, 95% CI
By genetic evidence source
pQTL (23/46)/(757/7154) 4.73 (3.51, 6.36)
OMIM (Mendelian disease) (79/192)/(1440/12830) 3.67 (3.07, 4.37)
Neale UKBB (40/110)/(1479/12912) 3.17 (2.47, 4.08)
FinnGen (26/79)/(1493/12943) 2.85 (2.08, 3.92)
Genebass (gene burden) (14/46)/(1505/12976) 2.62 (1.69, 4.07)
GWAS catalog (124/455)/(1395/12567) 2.46 (2.10, 2.88)
All OTG (127/484)/(1392/12538) 2.36 (2.02, 2.77)
PICCOLO (33/125)/(1486/12897) 2.29 (1.70, 3.08)
By L2G share only
L2G share: >= 0.25 (147/608)/(1372/12414) 2.19 (1.88, 2.54)
L2G share: >= 0.5 (127/484)/(1392/12538) 2.36 (2.02, 2.77)
L2G share: >= 0.75 (87/297)/(1432/12725) 2.60 (2.17, 3.13)
By pQTL + L2G share
L2G share: >= 0.25 (25/61)/(755/7139) 3.88 (2.85, 5.28)
L2G share: >= 0.5 (23/46)/(757/7154) 4.73 (3.51, 6.36)
L2G share: >= 0.75 (18/30)/(762/7170) 5.65 (4.18, 7.62)
By pQTL MR−coloc type
Cis−MR (20/40)/(760/7160) 4.71 (3.43, 6.47)
Trans−MR (13/21)/(767/7179) 5.79 (4.12, 8.16)
Mixed−MR (14/24)/(766/7176) 5.46 (3.87, 7.71)
MR+coloc (20/36)/(760/7164) 5.24 (3.88, 7.07)
By protein platform
pQTL: Somascan (6/16)/(91/843) 3.47 (1.79, 6.73)
pQTL: Olink (20/42)/(403/3259) 3.85 (2.77, 5.36)
1 2 3 4 5 6 7 8
Phase I−Launch Relative Success (RS), 95% CI


Overlapping Genetic Support from MR and Coloc Methods for Phase 1 to Launch T−I Pairs
Intersection Size
4
4
3
3
2
2
1
1
1
1
1
0
1
2
3
4
coloc_cis
trans_mr
mixed_mr
coloc_trans
cis_mr
20
15
10
5
0
Set Size

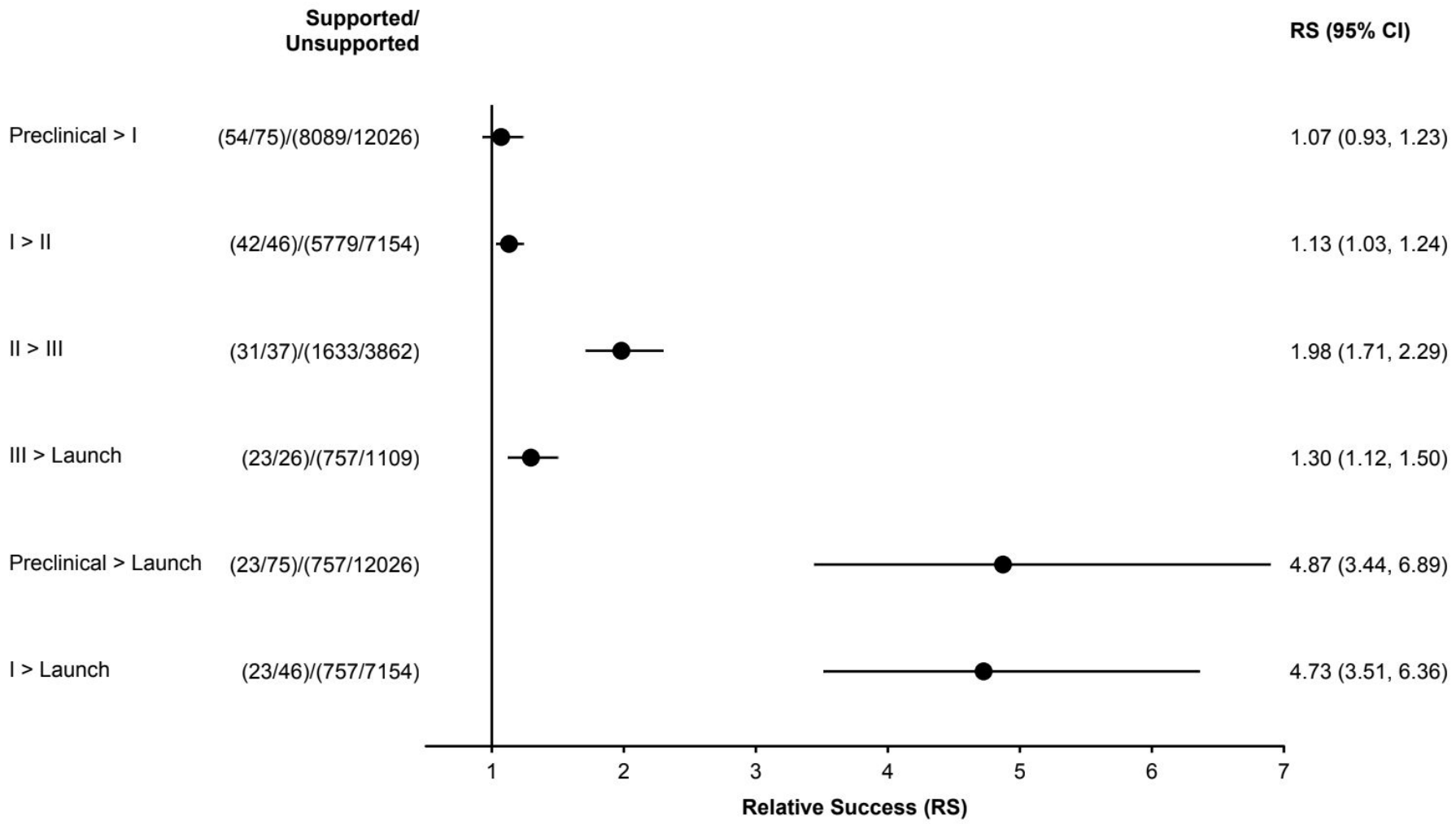

L2G only
(A[G]/S[G])/
(A![G]/S![G])
RS, 95% CI
enzymes
(70/614) / (228/1938)
0.97 (0.75, 1.25)
kinases
(56/463) / (195/1440)
0.89 (0.68, 1.18)
sm tractable
(470/3244) / (918/7785)
1.23 (1.11, 1.36)
ab tractable
(463/3370) / (880/7340)
1.15 (1.03, 1.27)
catalytic receptors
(131/723) / (218/1562)
1.30 (1.07, 1.58)
ion channels
(26/193) / (69/369)
0.72 (0.48, 1.09)
nuclear receptors
(40/150) / (43/175)
1.09 (0.75, 1.57)
rhodop gpcr
(117/579) / (82/809)
1.99 (1.54, 2.59)
transporters
(26/127) / (47/164)
0.71 (0.47, 1.09)
1
2
4
8
Phase I−Launch Relative Success (RS), 95% CI
L2G + pQTL
(A[G]/S[G])/
(A![G]/S![G])
RS, 95% CI
(7/9) / (149/1383)
7.22 (4.93, 10.56)
(3/4) / (180/1292)
5.38 (3.01, 9.63)
(19/34) / (658/5660)
4.81 (3.54, 6.54)
(22/45) / (714/6322)
4.33 (3.19, 5.88)
(7/16) / (277/1876)
2.96 (1.68, 5.22)
(0/0) / (23/78)
(0/0) / (13/47)
(0/0) / (26/115)
(0/0) / (3/23)
1
2
4
8
Phase I−Launch Relative Success (RS), 95% CI



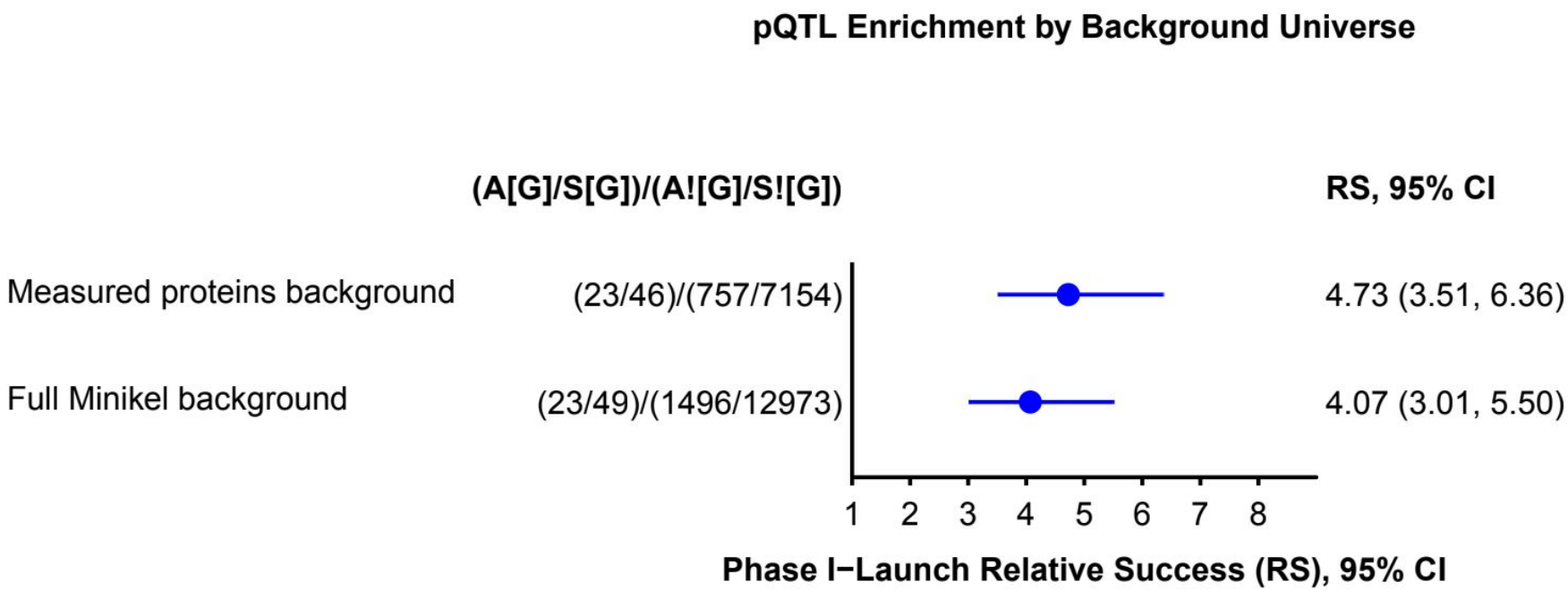
Supplementary Figure S1. Study design flowchart illustrating the two analytical paths: (1) systematic identification of pQTL-supported target-trait pairs through Mendelian randomization of eight proteogenomic datasets against ~8,000 complex trait GWAS, and (2) enrichment analysis comparing clinical success rates of pQTL-supported versus unsupported target-indication pairs using the Minikel et al. framework. Numbers at each step indicate the count of datasets, traits, associations, or T-I pairs retained after applying the indicated filters.



Supplementary Figure S2. Phase I-Launch relative success (RS) of pQTL-supported versus unsupported target-indication (T-I) pairs by therapeutic area. RS was computed within each therapeutic area using the Katz log method. Therapeutic areas were included if they had at least one pQTL-supported T-I pair entering Phase I and at least one pQTL-supported pair that reached launch (RS > 0). Nine therapeutic areas were excluded: six with no pQTL-supported T-I pairs at Phase I (endocrine, immune, infection, ophthalmology, other, psychiatry) and three with pQTL-supported pairs but no launches, yielding RS = 0 (neurology, oncology, signs/symptoms). The background was restricted to T-I pairs where the target was measured on at least one proteomic platform (Olink or SomaScan). Supported = A[G]/S[G] = launched/Phase I among pQTL-supported pairs; Unsupported = A[!G]/S[!G] = launched/Phase I among pQTL-unsupported pairs. Error bars: 95% CI (Katz log method).



Supplementary Figure S3. Relative success (RS) of pQTL-supported versus unsupported target-indication (T-I) pairs by clinical phase transition. RS was computed separately for each individual phase transition (Preclinical > I, I > II, II > III, III > Launch) and for cumulative transitions (Preclinical > Launch, I > Launch). The denominator for each transition includes only T-I pairs that entered the starting phase. The background was restricted to T-I pairs where the target was measured on at least one proteomic platform (Olink or SomaScan). Supported = A[G]/S[G] = succeeded/entered among pQTL-supported pairs; Unsupported = A[!G]/S[!G] = succeeded/entered among pQTL-unsupported pairs. Error bars: 95% CI (Katz log method).



Supplementary Figure S4. Sensitivity analysis comparing Phase I-Launch relative success (RS) of pQTL-supported T-I pairs under two background definitions: (1) measured proteins background, restricted to T-I pairs where the target was assayed on at least one proteomic platform used in the pQTL studies (Olink or SomaScan), matching the main analysis; and (2) full Minikel background, using all T-I pairs from Minikel et al. without restriction to measured proteins. pQTL-supported T-I pairs are identical in both analyses; only the unsupported comparator group changes. A[G]/S[G] = launched/Phase I among pQTL-supported pairs; A[!G]/S[!G] = launched/Phase I among pQTL-unsupported pairs. Error bars: 95% CI (Katz log method).