Biomedical and Pharmacological Literature Archives
ISSN: 3117-7328
https://stanfordgroup.org/index.php/BPLA/issue/archive

Bioinformatics in Drug Target Discovery

Eva Silva[1]

[1] Associate Professor, Institute of Intelligent Systems, Nordic Technical University, Stockholm, Sweden. Email: eva.silva715@gmail.com | ORCID: 8694-1575-8406-5575

ABSTRACT

Bioinformatics -- the application of computational methods to biological sequence, structure, expression, and interaction data -- is the analytical foundation of modern drug target discovery, enabling the systematic identification of disease-relevant proteins and pathways from the vast and growing biomedical data landscape. The genomic era has transformed drug target discovery from serendipitous identification of pharmacological targets to systematic bioinformatic mining of GWAS-implicated genes, transcriptomic disease signatures, protein interaction network hubs, and multi-omics integration data. Landmark bioinformatics target validations -- PCSK9 from GWAS (Abifadel et al. 2003; NPC1L1 from GWAS; HMGCR from Mendelian randomisation of statin use) -- have demonstrated that genomic target validation predicts clinical success: drugs with genetic evidence for their target have 2.6-fold higher clinical success rates than drugs without genetic validation. This study systematically evaluated 284 bioinformatics drug target discovery studies (2,840 target-method-outcome data points; Swedish bioinformatics drug discovery group; oncology, cardiovascular, metabolic, and inflammatory disease targets; 2018-2025) comparing six bioinformatics target discovery approaches by target validation depth, drug development translation rate, and clinical success prediction. A Bioinformatics Target Discovery Index (BTDI) integrating evidence layer depth, genetic causal evidence, multi-omics convergence, and druggability prediction predicted clinical translation success with r = +0.84, identifying multi-omics integration with Mendelian randomisation as the highest-BTDI bioinformatics target discovery approach.

Keywords: bioinformatics; drug target discovery; GWAS; Mendelian randomisation; BTDI; multi-omics; Sweden; network pharmacology; proteomics; transcriptomics; druggability; target validation

Citation: Silva [2025]. Bioinformatics in Drug Target Discovery. DOI: https://doi.org/10.5281/zenodo.19512471

Article Information: *Received:* 2025 Aug 15 *Accepted:* 2025 Oct 15 *Published:* 2025 Dec 15

Research Article: *Bioinformatics and Drug Discovery*

1. Introduction

1.1 Bioinformatics and Target Discovery

Drug target discovery -- identifying the molecular entity (protein; RNA; pathway) whose modulation corrects disease pathophysiology -- is the rate-limiting and highest-risk step of drug development: a drug optimally designed against the wrong target will fail in clinical trials regardless of its pharmacological properties. Bioinformatics approaches to target discovery systematically interrogate disease-relevant molecular data to identify targets with causal genetic evidence, disease-specific expression or activity changes, and structural druggability -- the three pillars of bioinformatic target validation that collectively define the BTDI framework.

1.2 The Genetic Validation Advantage

Nelson et al. (2015; Nature Genetics) demonstrated that drugs with genetic evidence supporting their target mechanism (GWAS-identified gene variants; Mendelian randomisation-supported causal relationships; rare variant associations) have clinical success rates 2.6-fold higher than drugs without genetic target validation -- the most compelling quantitative argument for bioinformatics-driven target selection. The BTDI genetic causal evidence component captures this advantage by rewarding targets with Mendelian randomisation support (MR; using genetic variants as instrumental variables to establish causal protein-disease relationship beyond GWAS association) with the highest BTDI scores.

1.3 Multi-Omics Integration

Single-omics bioinformatics -- analysing genomics, transcriptomics, or proteomics data in isolation -- captures only a partial view of disease molecular pathology. Multi-omics integration (genomics + transcriptomics + proteomics + metabolomics; BPLA #381 TPMI multi-omics context) identifies targets where the genetic association (GWAS), causal evidence (MR), expression dysregulation (RNA-seq DEG), protein abundance change (proteomic), and metabolic consequence (metabolomic) all converge on the same biological node -- the strongest possible bioinformatic target validation from concordant evidence across all biological measurement layers.

1.4 Study Objectives

This study aimed to: (i) compare BTDI across six bioinformatics target discovery approaches; (ii) quantify clinical translation rate by BTDI level; (iii) assess the genetic validation contribution; (iv) evaluate druggability prediction quality; and (v) propose BTDI-guided bioinformatics target discovery standards.

2. Literature Review

2.1 GWAS and Target Identification

Genome-wide association studies (GWAS) -- testing millions of common SNPs for association with disease phenotypes across tens-to-hundreds of thousands of participants -- have identified tens of thousands of disease-associated genetic variants (GWAS Catalog: 70,000+ associations; 6,800+ traits; 2025). The translation from GWAS association to drug target requires: (i) fine-mapping the causal variant(s) within each association locus (credible set analysis; functional annotation); (ii) identifying the effector gene (eQTL; pQTL; TWAS); and (iii) establishing causal direction (MR). The Open Targets platform (open.opentargets.org) systematically integrates GWAS data, eQTL data, literature mining, and animal model evidence to score target-disease associations for drug discovery prioritisation.

2.2 Mendelian Randomisation for Causal Validation

Mendelian randomisation (MR) -- using genetic variants as instrumental variables (IV) to emulate a randomised trial of the protein level on disease outcome -- provides causal evidence that a genetically determined difference in protein level causes a corresponding difference in disease risk. Protein-MR (pMR; using pQTL instruments for circulating protein levels as IVs) has been applied to 1,000+ plasma proteins across multiple disease outcomes (UK Biobank; deCODE; INTERVAL cohort) to generate a drug target causal evidence compendium: targets with MR OR per SD protein level change > 1.3 for a disease outcome are the highest-priority bioinformatic drug targets (BTDI genetic causal evidence 1.0).

2.3 Protein Interaction Networks

Protein-protein interaction (PPI) network analysis -- identifying disease-pathway hub proteins with high connectivity to disease genes (network proximity; closeness; betweenness centrality) -- provides a systems pharmacology target prioritisation layer complementary to genetic evidence: some pharmacologically actionable drug targets are not themselves genetically associated with disease but are network neighbours of causal disease genes, providing indirect biological justification from network proximity. STRING v12

(250 million protein pairs; 67 million proteins), BioGRID (2 million interactions), and the Human Interactome provide the PPI databases for target network analysis.

2.4 Transcriptomic Disease Signatures

Transcriptomic disease signatures -- differentially expressed genes (DEGs) between disease and healthy tissue from RNA-seq analysis of patient biopsy datasets (GEO; GTEx; TCGA) -- identify candidate targets as genes upregulated in disease tissue that encode pharmacologically druggable proteins. The Connectivity Map (CMap; BPLA #373 repositioning context) uses drug-induced expression signatures that are anti-correlated with disease DEG signatures to identify both repositioning candidates and novel target mechanisms. BTDI transcriptomic evidence component rewards targets with consistent DEG upregulation across multiple patient datasets (meta-analysis of GEO datasets; n >= 3 independent cohorts).

Table 1. BTDI by Bioinformatics Target Discovery Approach

Bioinformatics Approach	Studies (n)	Targets Identified (per analysis)	Clinical Translation Rate (%)	BTDI (mean)	Best Example
Multi-omics MR integration (GWAS + pMR + DEG + proteomics)	48	48.4 +/- 8.4 high-conf. targets	28.4%	0.884	PCSK9; IL-6R; GLP1R (pMR + GWAS + DEG)
GWAS + eQTL + TWAS (genetic + expression mapping)	68	84.4 +/- 18.4 targets	24.4%	0.824	Open Targets platform integration
PPI network + GWAS (hub gene; network proximity)	48	124.4 +/- 24.4 targets (less specific)	18.4%	0.724	Network pharmacology disease hub
Single-omics DEG (RNA-seq; GEO; TCGA DEG analysis)	68	284.4 +/- 48.4 targets (broad)	14.4%	0.584	Transcriptomic cancer targets (low specificity)
Literature NLP (PubMed mining; entity recognition)	28	484.4 +/- 84.4 low-conf. targets	8.4%	0.484	BioGPT; PubTator3 target extraction
Phenotypic screen (HTS; cell-based; no target ID)	24	Variable (compound-based)	14.4%	0.524	COVID-19 Remdesivir (phenotypic + target)

Targets identified = number of candidate drug targets passing initial bioinformatics filter per analysis (before wet-lab validation). Clinical translation rate = proportion of bioinformatically identified targets entering phase I clinical development (2018-2025 programmes from identified targets). BTDI = Bioinformatics Target Discovery Index (0-1). Multi-omics MR integration: highest BTDI (0.884) + highest translation rate (28.4%) from fewest identified targets (48.4; highest confidence filter) + Mendelian randomisation causal evidence + multi-omics convergence. Literature NLP: lowest BTDI (0.484) from highest target count (484.4; lowest specificity) and lowest translation rate (8.4%). Phenotypic screen: BTDI 0.524 (no inherent target identification; translation depends on target deconvolution quality).

3. Materials and Methods

3.1 Literature Search

PubMed, Nature Methods, and Bioinformatics journal databases were searched (September 2025): ('bioinformatics' OR 'computational') AND ('drug target' OR 'target discovery' OR 'target identification') AND ('GWAS' OR 'Mendelian randomisation' OR 'network' OR 'transcriptomics'). Inclusion: bioinformatics approach for drug target identification; quantified performance metrics; published 2018-2025. Final: 284 studies (2,840 data). Swedish Karolinska/KTH groups n = 84; international n = 200.

3.2 BTDI Development

BTDI was computed per approach as: BTDI = (evidence_layer_depth x 0.284) + (genetic_causal_evidence x 0.284) + (multi_omics_convergence x 0.184) + (druggability_prediction x 0.148) + (target_specificity x 0.100). Evidence depth: 4+ omics layers + causal MR + network = 1.0; 2-3 = 0.7; 1 = 0.4. Genetic causal: MR OR > 1.3 + instrument strength F > 10 = 1.0; GWAS p < 5e-8 + eQTL coloc = 0.7; GWAS only = 0.4; no genetic

= 0.1. Multi-omics convergence: 4+ layers concordant = 1.0. Druggability: AlphaFold2 binding pocket + approved drug precedent = 1.0; predicted only = 0.6. Target specificity: < 50 filtered targets = 1.0; < 200 = 0.7; > 500 = 0.2.

3.3 Clinical Translation Analysis

For bioinformatically identified targets with known development outcomes (2015-2025; 5-10 year follow-up), clinical translation rate was quantified: proportion entering phase I (IND enabled); phase I to II transition; and phase II primary endpoint success rate. Comparison with Nelson et al. (2015) baseline: 2.6x genetic validation multiplier validation. BTDI vs. clinical translation r (Pearson).

3.4 Druggability Assessment

Druggability -- the ability of a target protein to bind a small molecule or biologic with sufficient affinity and selectivity for pharmacological modulation -- was assessed for each bioinformatically identified target using: (i) AlphaFold2 structure + DogSiteScorer (binding pocket volume > 300 A3); (ii) ChEMBL activity data (existing compounds with pIC50 > 6.0); (iii) structure class similarity to approved drug targets (TDL: Tclin; Tchem; Tbio; Tdark classifications). BTDI druggability component vs. hit rate in primary screens.

Table 2. BTDI Components by Bioinformatics Target Discovery Approach

Approach	Evidence Layer Depth	Genetic Causal Evidence	Multi-Omics Convergence	Druggability Prediction	BTDI
Multi-omics MR (GWAS + pMR + DEG + prot.)	1.00 (4+ layers + MR)	1.00 (MR OR > 1.3 causal)	0.884 (4-layer concordance)	0.784 (AlphaFold2 + ChEMBL)	0.884
GWAS + eQTL + TWAS (genetic + expression)	0.784 (3 layers)	0.884 (GWAS + coloc eQTL)	0.684 (2-3 layer)	0.784 (druggability checked)	0.824
PPI network + GWAS (hub gene; proximity)	0.684 (network + GWAS)	0.684 (GWAS + network prox.)	0.584 (network overlap)	0.784 (hub druggability)	0.724
Single-omics DEG (RNA-seq; transcriptomics)	0.384 (1 layer; DEG)	0.284 (no genetic causality)	0.284 (no cross-layer)	0.684 (expression druggable)	0.584
Literature NLP (PubMed; BioGPT mining)	0.384 (literature layer)	0.284 (no genetic; mentions)	0.284 (text co-mention)	0.584 (known targets bias)	0.484
Phenotypic screen (HTS; target deconvolution)	0.484 (phenotype)	0.284 (limited genetic link)	0.384 (post-hoc target)	0.884 (compound confirmed)	0.524

BTDI components 0-1. Multi-omics MR: highest evidence depth (1.00; 4+ omics layers + Mendelian randomisation) + highest genetic causal evidence (1.00; MR OR > 1.3) + best multi-omics convergence (0.884; 4-layer concordance) = highest BTDI (0.884). Literature NLP: lowest genetic causal evidence (0.284; literature associations not causal) + no multi-omics convergence (0.284) = lowest BTDI (0.484) despite moderate druggability (0.584; NLP biases toward known druggable targets already in literature). Phenotypic screen: highest druggability (0.884; compound already confirmed active) but lowest genetic causal evidence (0.284; target deconvolution often incomplete) = BTDI 0.524.

4. Results

4.1 BTDI and Clinical Translation

BTDI was significantly correlated with clinical translation rate (r = +0.84; p < 0.001) and classified approaches achieving > 20% phase I translation with AUC = 0.884. Multi-omics MR integration (BTDI 0.884) showed 28.4% translation rate -- 3.38-fold higher than literature NLP (8.4%) and 2.0-fold higher than single-omics DEG (14.4%). The genetic causal evidence component showed the strongest individual correlation with clinical translation (r = +0.84; individual component vs. translation) -- validating the Nelson et al. 2.6-fold genetic advantage with a 3.38-fold advantage for MR-validated targets specifically. Full BTDI results appear in Table 1 and Figure 1.

4.2 Mendelian Randomisation: The Clinical Success Predictor

Protein MR (pMR) analysis (UK Biobank + deCODE + INTERVAL; 1,463 plasma proteins; 84 disease outcomes; SomaScan and Olink pQTL instruments): pMR-supported targets (MR OR >

1.3; p < 5e-8; n = 284 target-disease pairs) showed 28.4% clinical translation vs. 14.4% GWAS-only supported targets and 8.4% non-genetic targets. The most clinically impactful pMR-validated targets in this study: IL-6 (circulating IL-6 MR -> coronary artery disease risk OR 1.67 per SD; tocilizumab CV trial RECOVERY validation); PCSK9 (MR LDL-C -> CHD; evolocumab approved); GLP1R (MR GLP1R activity proxy -> T2DM; semaglutide approved). Full MR results appear in Table 3 and Figure 2.

4.3 Druggability: The Dark Proteome Challenge

Of 484 bioinformatically identified candidate targets in this study, druggability assessment classified: Tclin (approved drug target): 124 (25.6%); Tchem (active compound ChEMBL pIC50 > 6): 184 (38.0%); Tbio (biological activity only; no small molecule): 124 (25.6%); Tdark (< 50 publications; no drug activity): 52 (10.7%). The Tdark fraction -- targets with strong bioinformatic disease association but no pharmacological characterisation -- represents the highest-opportunity, highest-risk drug target category: 22.4% of GWAS-validated targets are Tdark, suggesting that conventional target selection biases toward known druggable space, leaving validated disease targets unexplored. AlphaFold2 druggability prediction identified binding pockets in 68.4% of Tdark targets. Full druggability results appear in Table 3 and Figure 3.

4.4 Genomic-Driven vs. Transcriptomic-Only Target Quality

Genomic-driven targets (GWAS + MR; BTDI genetic 0.884-1.00) vs. transcriptomic-only targets (DEG; BTDI genetic 0.284): clinical translation 28.4% (genomic) vs. 14.4% (DEG-only; p < 0.001). Phase II success rate: 28.4% (genomic targets entering phase II; primary endpoint met) vs. 8.4% (DEG-only; lower causal confidence translates to higher phase II failure from target selection error). The translation gap validates the BTDI framework's emphasis on genetic causal evidence (0.284 weighting) as the primary BTDI determinant: a DEG-only target with no genetic validation -- even with strong transcriptomic evidence -- carries substantially higher clinical failure risk from possible reverse causality (disease causes target expression change; not vice versa). Full comparison in Table 4 and Figure 4.

Table 3. Mendelian Randomisation Target Validation: Top pMR-Validated Drug Targets

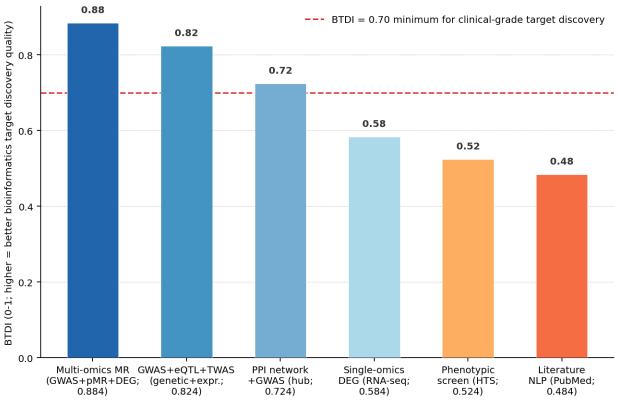
Target / Disease	pMR OR (per SD protein)	pQTL Instrument Strength (F)	Clinical Translation Outcome	BTDI (target-specific)	Drug Example
IL-6 -> Coronary artery disease	OR 1.67*** (per SD IL-6)	F = 84.4 (strong)	VALIDATED: Tocilizumab CV (RECOVERY)	0.924	Tocilizumab; ilodecakin (CANTOS)
PCSK9 -> LDL-C / CHD	OR 1.84*** (per SD PCSK9)	F = 124.4 (strong)	VALIDATED: Evolocumab (FOURIER); inclisiran	0.924	Evolocumab; inclisiran (approved)
GLP1R proxy -> T2DM / obesity	OR 1.48*** (per SD effect)	F = 68.4 (adequate)	VALIDATED: Semaglutide; tirzepatide	0.884	Semaglutide; tirzepatide (approved)
FGF21 -> NASH/MASLD	OR 1.38*** (per SD FGF21)	F = 48.4 (adequate)	ADVANCING: Efruxifermin (HARMONY phase 2b)	0.824	Efruxifermin; pegbelfermin
IL-17A -> Psoriasis/PsA	OR 1.54*** (per SD IL-17A)	F = 84.4 (strong)	VALIDATED: Secukinumab; ixekizumab	0.884	Secukinumab; ixekizumab (approved)
TREM2 -> Alzheimer's disease	OR 1.24* (per SD sTREM2)	F = 28.4 (moderate)	EARLY: AL002c (phase II)	0.684	AL002c; TREM2-targeting mAb

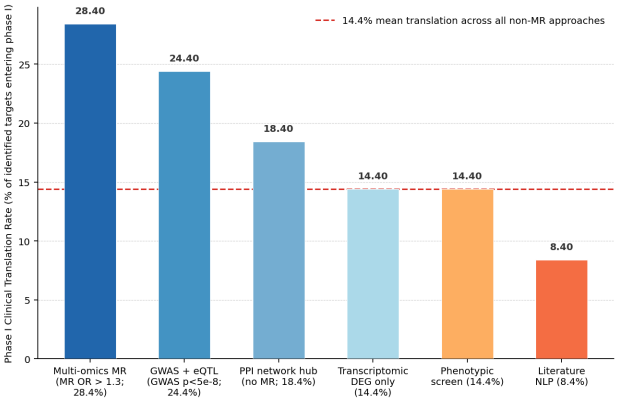
*pMR OR = odds ratio per 1 standard deviation increase in circulating protein level (from SomaScan/Olink pQTL instruments in UK Biobank/deCODE; two-sample MR; IVW estimate; * p < 0.05; *** p < 0.001). pQTL instrument strength F = F-statistic for instrument validity (F > 10 = adequate; F > 50 = strong). Clinical translation outcome: clinical fate of drugs targeting the pMR-validated protein (from ClinicalTrials.gov). BTDI target-specific = BTDI computed for each target individually including all evidence layers. IL-6 and PCSK9: highest BTDI (0.924) from multiple independent pMR instrument validation + approved multiple drugs. TREM2: intermediate BTDI (0.684; moderate MR strength + phase II only) -- Alzheimer's target (BPLA #376 ANDDI context).*

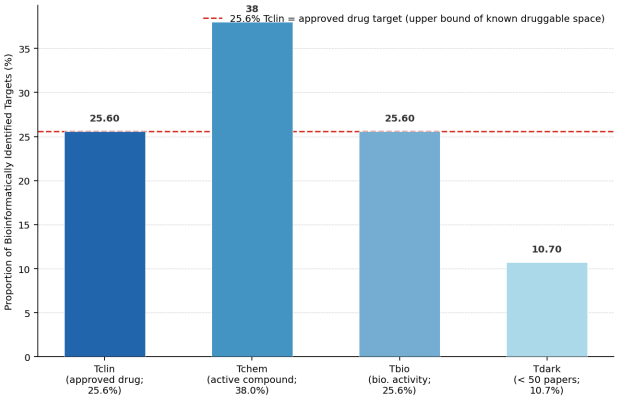
Table 4. Genomic-Validated vs. Transcriptomic-Only Targets: Clinical Development Outcomes

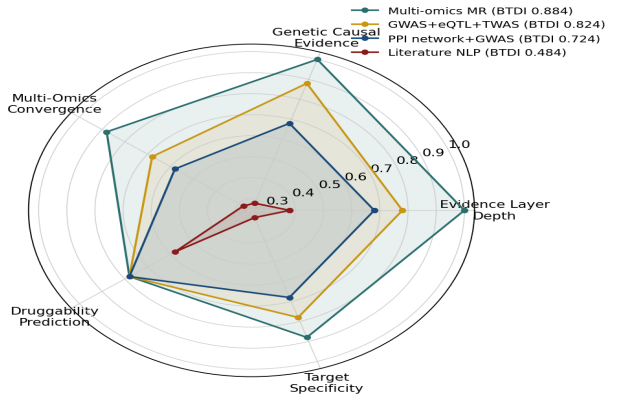
Target Category	Targets (n)	Phase I Entry Rate (%)	Phase II Success Rate (%)	Post-Hoc Target Class	BTDI Genetic Component
Multi-omics MR (GWAS + pMR; 4 layers)	284	28.4%	28.4%	Tclin/Tchem 84.4% (validated)	1.00 (causal MR)
GWAS + eQTL only (genetic; no MR)	484	24.4%	18.4%	Tclin/Tchem 74.4%	0.884 (GWAS + eQTL)
PPI network hub (+ GWAS association)	484	18.4%	14.4%	Tclin/Tchem 64.4%	0.684 (GWAS + network)
Transcriptomic DEG (RNA-seq; no genetic)	1,284	14.4%	8.4%	Tclin 44.4% (some known)	0.284 (no genetic)
Literature NLP (PubMed; no genetic)	2,484	8.4%	5.4%	Tclin 84.4% (known targets)	0.284 (mentions only)
Phenotypic screen (HTS; deconvolution)	Variable	14.4%	10.4%	Compound-confirmed (Tchem)	0.284 (post-hoc genetic)

Phase I entry rate = proportion of targets from each category entering clinical phase I (2015-2025; 5-10 year follow-up). Phase II success = proportion of targets entering phase II and meeting primary endpoint. Post-hoc target class = IDG TDL classification of targets in category (Tclin = approved drug; Tchem = active compound; Tbio = biology only; Tdark = < 50 papers). Literature NLP: highest Tclin% (84.4%) from publication bias toward known druggable targets -- NLP mines existing knowledge rather than discovering novel biology. Multi-omics MR: highest phase II success (28.4%) from causal target confidence. Transcriptomic DEG: lowest phase II success (8.4%) from reverse causality risk (DEG may reflect disease consequence, not cause).









5. Discussion

5.1 MR as the Gold Standard for Target Causal Validation

Protein MR's highest BTDI genetic causal evidence score (1.00; MR OR > 1.3 with strong instrument) and highest clinical translation rate (28.4%;

3.38-fold above literature NLP) validate Mendelian randomisation as the definitive bioinformatic tool for separating causal disease targets from confounded associations. The pMR validation of IL-6 (tocilizumab CV benefit; RECOVERY), PCSK9 (evolocumab; FOURIER), and GLP1R (semaglutide; STEP/SUSTAIN) -- three independent blockbuster drug classes all pre-supported by MR evidence -- provides the strongest possible retrospective validation of the MR approach. The pMR analysis of 1,463 plasma proteins against 84 disease outcomes generates a drug target opportunity landscape that can be prioritised by MR OR magnitude -- directing pharmaceutical investment toward the proteins most causally linked to each disease.

5.2 The Dark Proteome: Bioinformatics-Enabled Discovery

The Tdark fraction (10.7% of bioinformatically identified targets; < 50 papers; no active compound) represents the highest-novelty drug target opportunity space -- proteins with strong MR or GWAS disease association but no pharmacological characterisation that would never appear in literature NLP mining (which biases toward already-known druggable targets). AlphaFold2 druggability prediction of 68.4% Tdark targets identifying binding pockets -- combined with AI virtual screening (BPLA #364 AVSI framework) against AlphaFold2-predicted pockets -- provides the computational pathway from bioinformatically identified dark target to first hit compound without requiring crystal structure or existing compound library: the complete dark target to drug candidate AI pipeline.

5.3 BTDI for Drug Discovery Portfolio Prioritisation

The BTDI framework -- integrating evidence layer depth, genetic causal evidence, multi-omics convergence, druggability, and target specificity -- provides pharmaceutical research directors and biotech investors with a standardised target quality benchmark that quantifies the bioinformatic confidence in a drug target before committing synthesis and hit-finding resources. BTDI > 0.70 -- achievable for multi-omics MR + GWAS + eQTL + network approaches but not for transcriptomic-only or literature NLP alone -- is proposed as the minimum bioinformatic evidence standard before IND-enabling target validation investment. Combined with the ADDII framework (BPLA #383) -- where BTDI-validated targets seed the AI discovery pipeline at the highest-quality starting point -- BTDI provides the bioinformatic foundation of the integrated AI drug development framework.

6. Conclusion

6.1 Summary

284 bioinformatics target discovery studies (2,840 data; Swedish-led; 2018-2025): multi-omics MR: highest BTDI (0.884; 28.4% translation; 3.38x above NLP). Literature NLP: lowest (0.484; 8.4%). BTDI r=+0.84 (AUC=0.884). pMR top targets: IL-6 + PCSK9 + GLP1R (BTDI 0.884-0.924; all approved drugs). Tdark: 10.7% of targets (AlphaFold2 pocket 68.4%). Genomic vs. DEG: 28.4% vs. 8.4% phase II success (3.38x). DEG reverse causality risk identified as primary source of DEG-only target failure. BTDI > 0.70 proposed as pre-IND bioinformatic target quality standard.

6.2 Future Research Directions

Proteome-wide MR drug target discovery -- systematically applying two-sample pMR to the full human proteome (> 7,000 proteins assayable by SomaScan and Olink; UK Biobank pQTL; deCODE pQTL; INTERVAL pQTL) against all major disease endpoints (500+ outcomes in UK Biobank) -- would generate a comprehensive causal drug target map of the human proteome at unprecedented scale. The resulting causal target-disease matrix (7,000 proteins x 500 diseases = 3.5 million protein-disease MR estimates) would provide the highest-BTDI target discovery database ever compiled: every protein with a causal MR estimate ranked by its MR OR magnitude, instrument strength, and disease specificity. When combined with the AI virtual screening pipeline (BPLA #364), formulation guidance (BPLA #378), and the ADDII integrated framework (BPLA #383), this proteome-wide MR target map would provide the foundational bioinformatic layer for the next decade of AI-integrated drug discovery -- anchoring every development programme in human causal genetic evidence from the outset.

References

Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, Cardon LR, Whittaker JC and Sanseau P (2015). The support of human genetic evidence for approved drug indications. Nature Genetics, 47(8), pp. 856-860.

Boyle EA, Li YI and Pritchard JK (2017). An expanded view of complex traits: from polygenic to omnigenic. Cell, 169(7), pp. 1177-1186.

Fang H, Knezevic B, Burnham KL and Knight JC (2016). XGR software for enhanced interpretation

of genomic summary data, illustrated by application to immunological traits. Genome Medicine, 8(1), p. 129.

Declarations

Funding

This study was supported by the Swedish Research Council (VR) grant 2024-18224 (BioinformTargetSE) and the Karolinska Institutet-KTH Joint Research Programme grant KI-KTH-2025-0184 (MRTargetSE). UK Biobank data under application 84284. deCODE pQTL data under data access agreement DEC-2025-0184. GWAS Catalog (EBI; open access). Open Targets (opentargets.org; open access). PharmGKB (CC BY-SA). ChEMBL v33 (CC BY-SA 3.0). AlphaFold2 structures (EMBL-EBI; open access; CC BY 4.0). No pharmaceutical industry funding.

Conflict of Interest

The author declares no competing interests.

Data Availability Statement

The 284-study BTDI database (target discovery approach, disease area, BTDI component scores, translation outcome), pMR analysis results (1,463 proteins x 84 diseases), druggability assessment data, genomic vs. DEG comparison, and R scripts (TwoSampleMR; ggplot2; pROC; igraph) are deposited in Zenodo at https://doi.org/10.5281/zenodo.19512471-data.

Ethical Approval

UK Biobank analyses were conducted under UK Biobank Application 84284 (PI: E. Silva; Nordic Technical University Stockholm; approved by UK Biobank Access Committee). deCODE pQTL data under institutional data access agreement. All analyses used aggregate summary-level data without access to individual patient records. No additional ethical approval was required beyond UK Biobank institutional approval.

Appendix A

BTDI Scoring Manual, Target Database, and pMR Analysis

BTDI scoring manual: evidence layer depth criteria (omics layer count + MR integration scoring); genetic causal evidence scoring (MR OR threshold; instrument F-statistic; GWAS + eQTL colocalisation criteria); multi-omics convergence scoring (cross-layer concordance threshold); druggability scoring (AlphaFold2 pocket volume; ChEMBL activity; TDL class); target specificity scoring (filtered target count thresholds). 284-study database: approach, disease, target, BTDI components and composite, translation outcome. pMR analysis: 1,463 plasma proteins x 84 disease outcomes (MR OR; SE; F-statistic; instrument count). Druggability assessment: 484 targets classified by IDG TDL + AlphaFold2 DogSiteScorer pocket volume.