Int. J. Adv. Eng. Pure Sci. 2026, 38(1): <210-230>
DOI: 10.7240/jeps.1834505

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

De-TarDis: Decoy-Target Discovery Database Integrating Off-Targets, Promiscuity, Adverse/Side-Effects, and Screening Panels via Reciprocal Rank Fusion for Safety Assessment

De-TarDis: Off-Targetlar, Promisküöz Hedefler, Advers/Yan Etkiler ve Güvenlik Panellerini Reciprocal Rank Fusion ile Bütünleştiren Güvenlik Odaklı Hedef Veritabanı

Sadettin Yavuz UĞURLU [1,2]

[1]*Novexus Ltd, 07058, Antalya, Turkey*

[2]*Akdeniz University, Faculty of Engineering, Department of Materials Science and Engineering, Antalya, Türkiye*

Abstract

Computational screening in drug discovery typically concentrates on a single "intended" target; yet as projects approach the clinic, unexpected liabilities—off-target binding, promiscuous (multi-target) proteins, targets implicated in adverse and side effects, and those monitored in safety panels (e.g., Eurofins SafetyScreen™ tiers 1–3)—often drive failure. In order to address such a challenge, this study addresses the lack of a unified resource that enables early, collective checks against such risky targets. We surveyed the literature and public databases, compiling 49 lists organized into five groups: (1) off-target, (2) promiscuous target, (3) adverse-effect target, (4) side-effect target, and (5) safety-check target. Target identifiers were standardized to UniProtKB, and each list was internally ranked using volume and study-specific scores. Reciprocal Rank Fusion (RRF) has been applied to merge these heterogeneous rankings into a single, robust ordering—RRF rewards targets that rank highly across multiple sources, elevating consistently implicated proteins to the top. The resulting resource, "De-TarDis" (Decoy-Target Discovery Database), yields a consolidated "avoid-these-targets" list for computational campaigns. It can be used directly during hit-to-lead and lead-optimization to flag compounds likely to bind safety-relevant proteins, thereby reducing late-stage, ADMET-driven surprises.

Keywords: Target safety, Off-target proteins, Promiscuous targets, Adverse/side-effect–target associations, Reciprocal Rank Fusion (RRF), Integrated database for drug discovery

Öz

İlaç keşfi sürecinde hesaplamalı taramalar genellikle yalnızca "istenen" hedef protein üzerine odaklanır; oysa klinik aşamaya yaklaşıldıkça beklenmeyen hedef-dışı bağlanmaları, çoklu bağlanım gösteren proteinler, advers ve yan etkiyle ilişkilendirilmiş ve güvenlik panellerinde (örn. Eurofins SafetyScreen™ tier 1–3) olan hedefler başarısızlık nedeni hâline gelebilmektedir. Bu çalışma, erken aşamada bu riskli hedefleri topluca kontrol etmeyi kolaylaştıracak bütüncül bir kaynak eksikliğini gidermeyi amaçlamaktadır. Bunun için literatür ve kamuya açık veri tabanları tarandıktan sonra, beş ana grupta (1) hedef-dışı, (2) çoklu bağlanım, (3) advers etki, (4) yan etki ve (5) güvenlik kontrolü ilaç hedefler olmak üzere 49 ayrı liste derlendi, hedef tanımlayıcıları (UniProtKB) standartlaştırıldı ve her liste kendi içinde hacim ve çalışma-özgül skorlarına göre sıralandı. Ardından, Reciprocal Rank Fusion (RRF), bir hedefin farklı kaynaklardaki yüksek sıralarını ödüllendirerek çoklu çalışmada tutarlı görünen hedefleri en üste taşıyan basit ama sağlam bir sıralama tekniği uygulandı. Sonuçta "De-TarDis" (Decoy-Target Discovery Database) adı verilen ve hesaplamalı taramalarda "kaçınılacak hedef" olarak kullanılmaya uygun birleşik bir liste elde edilmiştir. Bu liste, öncü ilaç adayı bileşik belirleme aşaması ve öncü ilaç adayı bileşik optimizasyonu aşamalarında, aday bileşiğin istenmeyen güvenlik ilişkili proteinlere bağlanma olasılığını taramak için doğrudan kullanılabilir; böylece ileri aşamadaki ADMET kaynaklı sürprizlerin azaltılmasına katkı sağlar.

Anahtar Kelimeler: Hedef güvenliği, Hedef-dışı proteinler, Promisküöz hedefler, Yan etki / advers etki ilişkisi, Karşılıklı Sıra Birleştirme (Reciprocal Rank Fusion), İlaç keşfi için birleşik veri tabanı

I. INTRODUCTION

Structure-based drug discovery (SBDD), ligand-based methodologies, and contemporary AI-guided screening share a common constraint: they perform optimally when the target dimension is precisely specified [1-4]. Docking, pharmacophore modeling, and binding-site–oriented virtual screening presume clear knowledge of the protein to be modulated and, crucially, the appropriate structural or sequence representation to deploy [5-7]. When the target is well characterized, scoring functions are more interpretable and hit triage more economical; when it is not, even high-quality ligands may be prioritized incorrectly. In other words, strengthening the target layer provides a direct, leverageable route to strengthening the screening layer. To mitigate the "missing or incomplete target" layer, a range of theoretical and hybrid pipelines has emerged. For example, AlphaFold [8, 9] and related

Corresponding Author: SADETTİN YAVUZ UĞURLU, **Tel** +90 544 386 7919, **E-mail:** syavuzugurlu@akdeniz.edu.tr

Submitted: 03.12.2025, **Revised:** 25.01.2026, **Accepted:** 11.02.2026

structure-prediction servers now enable access to high-quality 3D models for proteins that previously lacked structures, while sequence-based annotation (e.g., Pfam, domain-level mapping, target–pathway links) supplies auxiliary information that docking engines or ML filters can exploit to prioritize targets even in the absence of crystallographic data [10-12]. Such a shift has produced a setting in which a screening campaign can be routinely enriched not only with additional ligands but also with additional targets, such that an identical small-molecule set is evaluated against a more informative biological context. In turn, this dynamic has motivated the systematic construction of target-centric libraries.

Target-centric repositories, such as the Therapeutic Target Database [13], DrugBank target sets [14-16], or indication-specific catalogs, primarily enumerate "intended" or disease-relevant proteins, namely those to be modulated [17-20]. However, late-stage failures [21-23], idiosyncratic toxicities [24-27], and post-marketing withdrawals are frequently driven by the opposite class: undesired or non-aimed targets that bind candidates with consequential affinity [28, 29]. For example, off-target GPCRs, hERG-like ion channels, highly promiscuous kinases, and targets previously implicated in adverse drug reaction reports exemplify the proteins medicinal chemists would prefer to flag before advancing a series [33-43]. The absence of a single, standardized compendium of such safety-relevant targets forces each group to reconstruct overlapping lists—typically with gaps and inconsistent nomenclature.

In this study, we propose "De-TarDis" (Decoy-Target Discovery Database), a standardized, rank-ordered collection of safety-relevant targets designed to be plugged directly into virtual screening, selectivity filtering, or hit-to-lead risk assessment. After systematically searching the literature and public collections, 49 individual lists have been gathered [44-64]. The 49 lists have been organized under five evidence-based groups: (1) off-target [44-45], (2) promiscuous target [46-48], (3) adverse effect–target [49-53], (4) side effect–target [47, 51, 54, 55, 56], and (5) safety check–target (Eurofins-like panels) [57-64]. Because these sources differ in size, scoring scheme, and curation depth, all target identifiers were harmonized to UniProtKB, and Reciprocal Rank Fusion (RRF) was applied to yield a single [65-67], robust ordering that rewards targets consistently reported across independent studies. In turn, De-TarDis eliminates unstandardized entries and strengthens cross-study concordance, enabling evaluation of designed ligands against a consolidated set of "do-not-hit" proteins. De-TarDis thereby facilitates earlier anticipation of selectivity liabilities and mitigates the risk of costly attrition in late discovery stages. Operationally, De-TarDis is intended to be used at three concrete decision points in a discovery pipeline: (i) early hit triage, where newly designed or screened compounds are filtered against a compact top-ranked subset to flag likely 'do-not-hit' liabilities; (ii) selectivity planning, where the same ranked catalog is used to assemble an application-specific counter-screen panel (e.g., a broader head of the list) and to prioritize follow-up profiling; and (iii) hit-to-lead risk assessment, where rank- or score-based thresholds are applied to transparently justify target-avoidance decisions and to document why certain liabilities were considered. In each use case, the practical step is the same: choose an operating cut-off (top-m targets or score threshold), map the compound's known/predicted target set to UniProtKB, and check whether any intersections fall within the chosen De-TarDis operating region; the rank position then provides an interpretable "evidence-weighted" priority for mitigation or de-risking. The De-TarDis dataset is freely available for academic use at: https://github.com/yauz3/De-TarDis

II. METHODS

The construction of "De-TarDis" (Decoy-Target Discovery Database) was implemented as a stepwise pipeline to transform heterogeneous "do-not-hit" signals into a single, rank-ordered resource suitable for direct use in computational screening. Given that underlying studies differ in purpose (off-target panels, promiscuity summaries, ADR/side-effect mining, commercial safety screens), granularity (protein, domain, UniProt, gene), and scoring style (counts, p-values, effect factors, plain lists), the workflow was partitioned into four sequential stages: (i) Source Acquisition and Study-Specific Normalization: compilation of relevant sources and application of dataset-specific preprocessing to recover each study's intended ranking; (ii) Identifier Harmonization to UniProtKB: standardization of target identifiers and mapping of all entries to UniProtKB to eliminate duplicates arising from synonyms, isoforms, or database-specific IDs; (iii) Within-Source Evidence Scoring and Ranking: derivation of a single numeric score and within-dataset rank for each target using available evidence variables (e.g., record counts, side-effect counts, rate-like fields, transformed p-values); and (iv) Cross-Source Integration via Reciprocal Rank Fusion (RRF): aggregation across datasets to obtain a robust global order that prioritizes targets repeatedly flagged across independent sources. The following subsections detail these four stages.

2.1. Source Acquisition and Study-Specific Normalization

A coherent "do-not-hit" corpus was assembled to capture safety-relevant signals that are typically dispersed across assays, reports, and repositories. Candidate sources were screened for biomedical scope (human or clinically transferrable orthologs), evidential credibility (biochemical/cell-based affinity or functional readouts, curated clinical signals), and

metadata sufficiency (stable identifiers and extractable endpoints). Study-specific rules were then defined to preserve each source's native ranking intent while removing presentational artifacts.

A systematic review of the literature and public resources yielded five evidence-based groups: (i) off-target (undesired binding to non-intended proteins), (ii) promiscuous target (broad multi-target interaction behavior across ligands/assays), (iii) adverse-effect target (clinical or post-marketing ADR-linked), (iv) side-effect target (phenotype-linked associations with broader observational provenance), and (v) safety-check target (standardized tiered screening panels) (Table 1 and Figure 1).

Table 1. Overview of data sources by evidence group. Rows list the five groups ((i) off-target, (ii) promiscuous targets, (iii) adverse-effect target, (iv) side-effect target, and (v) safety-check target) plus a total row. Columns report Source number (count of distinct sources curated for each group), Protein number (unique UniProtKB targets captured per group after removing overlaps within groups), and References (bibliographic/source identifiers underlying each group).

Groups	Source number	Protein number	References
Off-target	2	4985	[44, 45]
Promiscuous targets	3	6746	[46-48]
Adverse-effect target	5	1625	[49-53]
Side-effect target	5	1001	[47, 51, 54, 55, 56]
Safety-check target	34	222	[57-64]
Total (group-wise sum; not the union; overlaps across groups are double-counted)	**49**	**14579**	**[44-64]**

In total, 49 lists were extracted from the literature for standardization, and a dedicated promiscuity group was constructed from the entire BindingDB corpus, using cross-ligand, cross-assay interaction breadth as the defining signal for multi-target behavior (Table 1). These groups differ in evidentiary granularity, endpoint type (affinity/functional vs. observational), and curation status, and therefore require distinct normalization strategies.

2.2. Identifier Harmonization to UniProtKB

Heterogeneous identifiers—gene symbols, protein names, domain labels, database-specific IDs, and isoform/accession variants—were reconciled to a single UniProtKB anchor to enable deduplication, cross-source linking, and rank integration. The following five subsections outline source-specific treatments prior to fusion: (i) collection and normalization of off-target datasets for RRF compatibility; (ii) rate- and record-based ranking of promiscuous targets; (iii) conversion of adverse effect–target associations into a composite score; (iv) ordering of side-effect tables using volume plus model-derived scores; and (v) re-ranking of safety-screen panels by assay coverage (Figure 1).

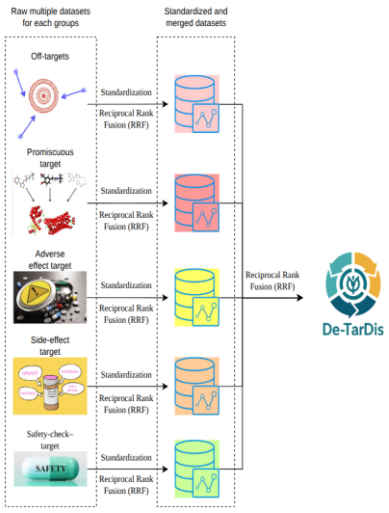

Figure 1. Schematic of the data-integration pipeline. Left: five input groups— (i) off-targets, (ii) promiscuous targets, (iii) adverse-effect targets, (iv) side-effect targets, and (v) safety-check targets—each originating from multiple raw datasets. Middle: for each group, datasets are standardized, ranked, and combined using Reciprocal Rank Fusion (RRF) to yield a group-level merged dataset. Right: the five group-level outputs are subsequently integrated via a final RRF step to produce the unified De-TarDis database.

2.2.1. Collection and normalization of off-target datasets for RRF

Two complementary off-target resources [44-45] were curated and transformed into rank-compatible inputs for Reciprocal Rank Fusion (RRF). (i) In the dataset of Rao et al. [44], target–compound associations were prioritized by a probability-oriented metric rather than raw counts: all pairs were first ordered by offtarget_rate_unique, which captures the fraction of unique compounds exhibiting off-target activity and thus normalizes for unequal screening depth; ties on offtarget_rate_unique were broken by n_records to privilege associations supported by larger observation counts. The resulting sequence was then linearly numbered to yield a strict 1…N rank suitable for direct input to RRF. (ii) For Off-targetP ML (the open-source off-target panel framework) [45], targets were ordered by model-derived confidence rather than frequency: entries were first sorted in descending Pseudo_Score_Mean, which summarizes predictive-model agreement/strength for each protein and more directly reflects the likelihood of a true off-target association; ties on Pseudo_Score_Mean were resolved using molecule_number to favor targets supported by more compounds. As above, the finalized sequence was linearly numbered to produce a strict 1…N rank for RRF. Such a two-source normalization ensures that probability/confidence signals drive ordering while record volume remains an evidence-preserving tie-breaker.

2.2.2. Rate- and record-based ranking of promiscuous targets

Three complementary sources were integrated to represent target-level binding promiscuity: (i) a list-style literature collection (ProtMiscuity) focused on catalytic promiscuity [46]; (ii) a second literature source providing a list without intra-list scores [47]; and (iii) a PDBBind-derived dataset that provides a quantitative, rankable backbone from crystallographic affinity evidence [48]. For the PDBBind arm, affinities were normalized to molar units and converted to pAffinity ($-\log_{10}$ M); ligands were labeled active per target at pAffinity ≥ 6.0 (≈ 1 μM, configurable) [68]. Uniqueness/diversity was controlled by counting unique ligands and optionally collapsing chemotypes via InChIKey-first-block or Bemis–Murcko scaffolds [69] to avoid analog inflation. The primary signal was the promiscuity_rate (n_active_unique / n_total_unique), augmented by two auxiliaries—mean pAffinity over actives (potency proxy) and the number of diverse active chemotypes (diversity proxy). These were combined into a composite PromiscuityScore = 0.6·rate + 0.3·diversity_norm + 0.1·potency_sigmoid, with ties broken by n_active_unique, then n_total_unique, and all targets mapped to UniProtKB for cross-source fusion. Together, this yields a data-driven ranking that can cross-validate literature-reported promiscuous proteins while furnishing a stable, occurrence-normalized backbone for integration via RRF.

A PDBBind-derived promiscuity signal can be influenced by structural coverage bias because certain proteins are disproportionately represented in the PDB. To mitigate such an effect, De-TarDis performs identifier harmonization to UniProtKB at the earliest stage and computes/aggregates promiscuity at the protein (UniProtKB accession) level, rather than at the structure, chain, or PDB-entry level. Consequently, an overrepresented protein does not generate multiple distinct targets in the final resource; instead, all evidence is consolidated under a single UniProtKB identifier. This design prevents "structure-rich" targets from inflating apparent coverage through duplicate target instances and ensures that downstream integration reflects target-level evidence rather than PDB-entry abundance.

The two literature datasets [46, 47] that enumerate promiscuous targets without quantitative priority (no activity counts, hit rates, or confidence values) were retained as existence signals. To integrate them into RRF without overpowering score-based evidence, every entry from each list-only source was assigned the median rank of the PDBBind-derived table. Such a design preserves the claim "reported as promiscuous elsewhere" and allows such targets to move up when corroborated by other sources, while avoiding the unrealistic assumption that all unscored entries deserve top ranks by default.

2.2.3. Conversion of adverse effect–target associations into a composite score

Adverse-effect evidence entered De-TarDis from five sources [49-53] with differing granularity and statistics; for each source, identifiers were first harmonized to UniProtKB, then a single monotonic priority per target was produced and linearly numbered for RRF. (1) ADReCS-Target (database) [49]: an occurrence-driven composite was used—score = (Number of records) × (Number of unique ADR terms) × (Number of unique drugs)—so targets recurring across many reports, terms, and drugs rank higher. (2) Soldatos et al. [50]: volume and disproportionality were combined—score = (Average number of records) × (Average number of unique reaction terms) × (Average number of reported cases for the target–reaction) × (Average PRR) × [−log10(Average P-value)] — thereby promoting targets both frequently observed and statistically enriched. (3) Zhao et al. [51]: the original published rank was adopted as the within-source priority after domain→UniProtKB mapping. (4) ChEMBL-derived adverse-effect library [52]: enrichment magnitude native to the study was used—score = (mean enrichment factor, EF)—and sorted in descending order; other descriptive fields were retained only for interpretation. (5) TARDIS-derived table [53]: recurrence and statistical strength were blended—score

= (Number of records) × (Number of unique side effects) × (Number of unique drugs) × (Mean −log10 P-value) × (Mean −log10 Q-value)—so repeatedly observed, statistically strong associations rise to the top; each source's priorities were then converted to strict ranks (1…N) with deterministic tie-breakers and passed to the RRF stage. By preserving both the diversity (unique adverse-effect counts) and volume of evidence within each source and then merging the resulting priority lists via RRF, this scheme enhances the robustness and generalizability of the downstream safety rankings.

2.2.4. Ordering of side effect–target tables using volume plus model-derived scores

Five side-effect–oriented resources [47, 51, 54, 55, 56] were standardized into rank-compatible inputs for De-TarDis, with a guiding rule to prefer directly reported, model-derived weights where available and otherwise use transparent volume proxies to preserve evidential breadth. (i) *Mizutani et al.* [54] entries were ordered by the source's own mean_weight (descending), retaining the study's statistical intent and yielding a clean, monotonic ranking. (ii) The table in the study of *Kuhn et al.* [55] was prioritized by n_side_effects (descending), a conservative proxy for how broadly a target is implicated across distinct side-effect terms. (iii) For the study of *Iwata et al.* [51], signals reported at the domain level were first mapped to UniProtKB proteins; within the resulting protein-level table, ordering followed the provided domain-association strength or, when only counts were available, the corresponding count field (descending), thereby preserving the original ranking logic while acknowledging granularity differences. (iv) The data resource of *Liu et al.* [47] was ranked by n_side_effects (descending), reflecting the extent of associations inferred by canonical component analysis without over-interpreting model internals. (v) *GESSE* [56], included as supportive predictive evidence, was ordered by n_records (descending) to reflect observation breadth without overriding empirically weighted studies. After within-source ordering, each table was linearly numbered (strict 1…N) with deterministic tie-breakers, enabling cross-dataset fusion via RRF. Collectively, these choices ensure that each side-effect resource contributes a priority signal that is faithful to its original statistical intent while still being comparable across datasets. By favoring empirically weighted scores where available and carefully defined volume-based proxies otherwise, the procedure preserves both the diversity and breadth of side-effect evidence. Converting these priorities into strict ranks and fusing them via RRF thus yields a robust, generalizable side-effect–centric view of target safety for downstream De-TarDis applications.

2.2.5. Re-ranking of safety-screen panels by assay coverage

Safety-panel evidence (Eurofins SafetyScreen–style (https://emea.eurofinsdiscovery.com/solution/tier-1-tier-2-and-tier-3-safety-panels) was standardized into a single, rank-compatible input for De-TarDis using a coverage-first strategy across 34 lists—19 from Tier 1, 13 from Tier 2, and 2 from Tier 3 [57-64]. After mapping all targets to UniProtKB and collapsing duplicates within each panel, assay coverage per target was recalculated as a tier-weighted sum of distinct assays so that deeper profiling contributes more strongly: (i) within each tier, the number of unique assays or panels hitting the target was counted (binding and functional entries were de-duplicated per assay format); (ii) a monotonically increasing tier weight was applied (Tier 1 < Tier 2 < Tier 3) and the tier-specific coverages were summed to produce a single coverage score; (iii) ties were resolved by assay modality (functional evidence preferred over binding-only), then by total panel count within the same tier, and finally by record volume where available. The resulting coverage-ordered list was linearly numbered (strict 1…N) to feed the cross-dataset RRF stage, ensuring that targets repeatedly profiled—especially across higher-tier panels—receive higher priority in the integrated ranking.

Transparency and reproducibility considerations were also addressed for this safety-panel component. The tiered SafetyScreen™ concept and panel offerings (Tier 1–3), including panel names and tier rationales, are publicly described by the provider, and panel composition is commonly disclosed at the level of target lists in publicly accessible brochures and web materials. Importantly, De-TarDis does not redistribute proprietary assay results or any paid screening outputs; instead, the integration uses only the safety-panel target identities and coverage-style metadata required to construct a rank-compatible "safety-priority" list, following UniProtKB standardization. Consequently, this component can be reproduced without purchasing commercial screening services by (i) using the publicly available panel composition information (or an institution's own licensed panel target list, if applicable) and (ii) running the provided mapping and re-ranking procedure. In practice, De-TarDis is intended to support early-stage de-risking by highlighting safety-relevant targets upfront, potentially reducing the need for broad, repeated commercial profiling and thereby saving both time and experimental resources during candidate prioritization.

To further harmonize heterogeneous safety-check sources into a robust "safety-priority" order, the merged table was re-ranked with a transparent rule set that favors breadth, depth, and pharmacological relevance. Specifically, targets were sorted by:

1. **Files/panels coverage**: the number of independent files/panels reporting the target (n_files_present), descending.
2. **Tier breadth**: the number of distinct tiers in which the target appears (tiers_present parsed as Tier-1/2/3), descending.
3. **Assay modality weight**: Functional assays were up-weighted over cell-based and binding formats using fixed weights: Functional = 3, Cell-based = 2, Binding = 1 (column Assay Sub Type).
4. **Safety-relevant family weight**: Pharmacologically sensitive classes were emphasized using fixed weights: GPCR = 3, Ion channel = 3, Transporter = 2, Nuclear receptor = 2, others = 1 (column Family).

The sorted list was then assigned a deterministic safety_rank (1…N). This re-ranking preserves the original coverage logic while injecting consistent preferences for higher-information assays and safety-critical target families, producing a stable, rank-compatible safety-panel list for the subsequent cross-group RRF integration.

2.3. Within-Source Evidence Scoring and Ranking

Within each evidence group, all constituent sources were first consolidated into a single within-group ranking using the source-specific scoring rules described above, followed by linear numbering (strict 1…N) with deterministic tie-breakers (Figure 1). Such a within-design brings several advantages: (i) it neutralizes heterogeneity in scale and format (counts, rates, model-derived weights) by translating each source into a common ordinal space; (ii) it prevents source-count bias, so groups represented by many lists (e.g., safety panels spanning multiple tiers) do not dominate groups with fewer, denser datasets; (iii) it reduces sensitivity to idiosyncratic scoring choices within any single study by allowing consensus to emerge at the group level; and (iv) it improves robustness by pooling partially overlapping evidence so that repeatedly implicated targets rise while one-off, low-signal entries recede. Therefore, within-group consolidation curbs outlier influence, preserves group-level consensus, and yields a stable, rank-compatible list for each evidence class—thereby increasing both the robustness and efficiency of De-TarDis. Such preparation also enables the subsequent cross-group integration via Reciprocal Rank Fusion (RRF).

To aggregate heterogeneous evidence across multiple sources, Reciprocal Rank Fusion (RRF) is employed as a rank-based integration strategy. For a given target t, let L denote the set of ranked lists (one list per evidence source), and let rank_l(t) be the 1-based rank position of t in list l ∈ L. The fused score is computed as:

$$RRF(t) = \sum_{l \in L} \frac{1}{k_{RFF} + rank_{l(t)}} \quad \textbf{(1)}$$

If a target t does not appear in a particular source list l, that source is treated as providing no explicit support for t and contributes zero to the fused score. In practice, the corresponding term is omitted from the summation; equivalently, rank_l(t) can be considered infinity (rank_l(t) = ∞), so that 1 / (k_{RFF} + ∞) = 0. This convention ensures that the fusion score is driven only by observed rankings (explicit evidence) and avoids assigning artificial placeholder ranks or unintended penalties to targets missing from a given source.

The damping constant k_{RRF} controls how strongly top-ranked positions dominate the fused score. Smaller k makes Eq. 1 more top-heavy: very early ranks receive substantially larger weight, so a single extremely high rank in one list can disproportionately influence the overall score. In contrast, larger k_{RFF} smooths rank contributions by reducing the marginal advantage of being extremely highly ranked in any single list, thereby increasing robustness to rank noise and limiting single-source dominance. Importantly, the original RRF study reports that k was fixed during a pilot investigation and that the choice of k_{RRF} was "near-optimal" around the selected value while "not critical," with only small changes in effectiveness across a broad range of k values. This is further supported by the pilot results (Table 1 in the original RRF paper), where performance varies only slightly across k_{RFF} (including values around 50–100), indicating that the method is relatively insensitive to k within this practical operating region [65]. In line with this established usage, many RRF implementations and subsequent applications adopt k = 60 as a practical default damping constant, reinforcing that the selected damping regime is standard rather than ad hoc [65, 67, 70].

In this study, a fixed damping constant k = 50 is used for all experiments to balance emphasis on high-confidence top ranks with robustness to rank fluctuations across heterogeneous sources. The k = 50 damping reduces marginal gains from extremely high positions in any single list and instead rewards targets that achieve consistently strong ranks across multiple independent evidence types, guarding against single-source dominance and promoting cross-source consensus [65, 67, 70].

Finally, De-TarDis is designed to be straightforward to update and extend. Because the public GitHub repository provides the raw inputs and end-to-end Python pipeline (including data preparation and integration scripts), the resource can be refreshed with new database releases, augmented with additional evidence sources, or re-generated under revised inclusion criteria in a transparent and reproducible manner [65, 67, 70].

2.4. Cross-Source Integration via Reciprocal Rank Fusion (RRF)

After producing a single, rank-compatible list within each evidence group, the lists are integrated across groups using Reciprocal Rank Fusion (RRF) with a fixed damping constant k_{RRF}=50 (Eq 1). Such a setting attenuates the influence of any one exceptionally high rank in a single list, moderates differences in list length and density, and rewards targets that achieve consistently excellent (not just singularly best) positions across multiple evidence sources. Such cross-source integration brings several advantages: (i) it amplifies concordant signals across heterogeneous evidence types (off-target, promiscuity, adverse/side-effect, safety-panel coverage), pushing upward targets repeatedly ranked high; (ii) it guards against single-source bias, since a strong position in one list cannot dominate without corroboration; (iii) it reflects the practical reality that, beyond the "intended" target(s), unintended binding can arise across many pathways—combining evidence classes yields a truer picture of liability; and (iv) it improves screening efficiency by concentrating attention on reproducibly unsafe proteins, saving time and budget in virtual and experimental campaigns and lowering the risk of late-stage attrition. In effect, cross-group RRF with k=50 converts many heterogeneous "do-not-hit" cues into one consensus order, enabling De-TarDis to prioritize unintended targets most relevant to early risk detection and cost-effective decision-making.

III. RESULT AND DISCUSSION

De-TarDis is a standardized, rank-ordered "do-not-hit" resource designed to surface plausible unintended targets early in drug discovery so that selectivity can be assessed before costly downstream studies. The database was assembled by surveying the literature and public repositories and consolidating 49 independent lists plus one PDBBind-derived promiscuity table into a unified framework centered on UniProtKB identifiers. Its contributions are threefold: (i) standardization—heterogeneous sources (counts, rates, model weights, panel coverage) are harmonized to a common target space with deterministic, within-source ranks; (ii) robust integration—cross-source Reciprocal Rank Fusion (RRF) (Eq. 1) promotes targets repeatedly implicated across independent evidence types while damping single-list artifacts; and (iii) practical usability—the output is a single, rank-compatible list that can be plugged directly into docking filters, ligand-based screening, or portfolio-level risk scans to triage selectivity liabilities.

Two complementary analyses are reported: 3.1 Global characteristics and cross-source overlap, which quantifies the scale of De-TarDis, summarizes rank distributions, and measures agreement among evidence classes; and 3.2 Group-wise safety profiles of target collections, which maps families or indication-oriented sets onto the De-TarDis ranks to reveal clusters of recurrent liabilities and the relative contributions of each evidence group.

3.1. Global characteristics and cross-source overlap

The integrated De-TarDis target set is strongly human-centric but multi-species after identifier harmonization to UniProtKB and deduplication of synonyms/isoforms (Figure 2). From the pie chart, Homo sapiens accounts for roughly 38–39% of unique targets; common preclinical species together make up most of the remainder: Mus musculus (~13%), Rattus norvegicus (~11%), Bos taurus (~4–5%), Sus scrofa (~2–3%), Gallus gallus (~2%), and Danio rerio (~1–2%). Well-used genetic or microbial models also appear with smaller shares, including Saccharomyces cerevisiae (~1–1.5%), Drosophila melanogaster (~1–1.5%), Caenorhabditis elegans (~1–1.5%), Xenopus laevis (~2%), Escherichia coli (~2%), and others (Figure 2).

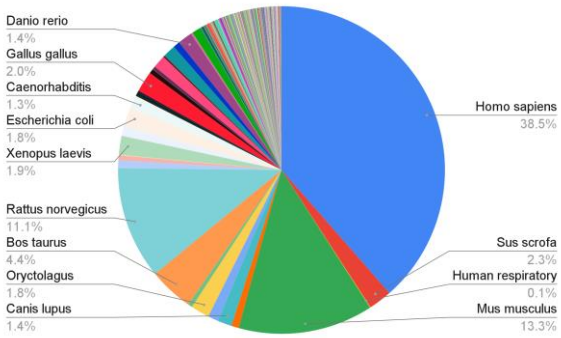

Figure 2. Species distribution of De-TarDis targets. A pie chart summarizing the organism (UniProtKB "organism" field) associated with each unique target aggregated across all evidence sources. Each slice corresponds to one species; call-out labels display the species name and its relative share of targets (percentage of the total). Major species (e.g., human, mouse, rat, pig, zebrafish, yeast, fruit fly, chicken, etc.) are individually annotated, while numerous lower-frequency species are shown as smaller slices. Percentages are computed as (number of targets mapped to a species) ÷ (total targets), after identifier harmonization to UniProtKB.

Beyond ranking, De-TarDis retains the organism annotation for every target as a first-class attribute. This adds value in several ways:

- **Translational alignment:** users can quickly focus on human targets when early safety flags arise or trace cross-species concordance.
- **Assay selection & reagent planning:** species tags guide which recombinant proteins, cell lines, or animal models are most appropriate for ortholog testing and follow-up validation.

- **Interpretability of cross-source signals:** evidence originating from non-human datasets (e.g., yeast genetic screens, fly/roundworm phenotypes) can be separated or down-weighted in decision making, while conserved targets across species can be prioritized for mechanistic probing.
- **Regulatory and safety context:** many historical safety panels are human; organism labels make these provenance differences explicit and help avoid species-mismatch pitfalls.

The small slices reflects the breadth of legacy datasets and specialized panels (microbial and amphibian proteins, less common mammalian models) (Figure 2). Although targets were collapsed to UniProtKB accessions, residual biases remain—the over-representation of human and major rodent species. Such a bias is expected given the dominance of human-centric pharmacology resources and rodent preclinical testing; the explicit organism field allows users to filter, stratify, or re-weight accordingly. Consequently, a user arrange and pick the "not-to-hit" targets by filtering organism, which increase the practicality of the database.

Figure 3A shows the number of targets with a non-null rank in each evidence group after harmonization to UniProtKB: (1) off-target, 4,985; (2) promiscuous target, 6,746; (3) adverse-effect target, 1,625; (4) side-effect target, 1,001; and (5) safety-check target, 222. Although Table 1 reports larger counts when each group is tallied by summing its constituent sublists, the figure clarifies why the final, integrated set is smaller—many targets appear in multiple evidence groups. Specifically, the group-wise files contain 14,579 total entries when counts are summed across evidence groups; this "raw tally" intentionally double-counts targets that appear in more than one group. After harmonization to UniProtKB identifiers and cross-group deduplication, these entries collapse to 10,736 unique targets, i.e., the union of all groups in De-TarDis. This gap between 14,579 (sum across groups) and 10,736 (union) reflects genuine cross-source overlap and therefore evidence convergence: a substantial fraction of targets is supported by multiple independent evidence groups rather than a single source, consistent with cross-validated prioritization for safety-relevant targets.

The largest bar in Figure 3A corresponds to the promiscuous target group (6,746). Unlike the other groups, which draw on curated safety or phenotypic resources, promiscuity had to be constructed from structure-centric evidence (e.g., PDBBind) to obtain broad coverage; this design choice naturally increases its list size relative to off-target pharmacology (4,985) or clinical/phenotypic sources such as adverse-effect (1,625) and side-effect (1,001). Because of these coverage differences, treating the entire De-TarDis set as "do-not-hit" decoys will be unnecessarily overly broad for practical usage, especially for methods like higher-fidelity modeling (e.g., docking, MD, or DFT-level studies).

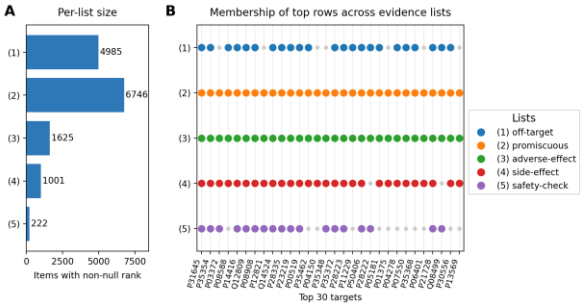

Figure 3. Overview of evidence-list coverage and per-target membership in De-TarDis. (A) Per-list size: bar lengths show the number of targets with non-null ranks in each evidence group, ordered as (1) off-target, (2) promiscuous target, (3) adverse-effect target, (4) side-effect target, and (5) safety-check target. (B) Membership of the first 30 targets in the CSV: each vertical guide corresponds to one UniProt accession on the x-axis; rows (top→bottom) again denote lists (1)–(5). Colored dots indicate that a target appears (has a non-null rank) in that list; small gray dots mark non-membership. The legend repeats the (1)–(5) labels used on the y-axis.

In practical usage of De-TarDis, Figure 3B illustrates this cross-source convergence for the first 30 UniProt accessions: colored dots mark membership in each list and reveal frequent multi-row co-occurrence across (1)–(4), with sparser coverage by (5). The overlap patterns in Figure 3B suggest a pragmatic path: prioritize targets that recur across independent evidence types with the relatively top target in five groups. Consequently, even a small panel of the top ~30 targets already captures substantial cross-source agreement, making it a compact, actionable screen for early liability checks.

As for the comprehensive investigation of overlap in ranks, the stacked bars (Figure 4) show that overlap across evidence sources is strong among the very top-ranked targets. For each threshold (k = 10, 20, 30, 40, 50), most selected targets are supported by four or five independent datasets, and only a few appear in just three (virtually no targets are single-dataset singletons at these ranks). As k increases from 10 to 50, the absolute number of multiply supported targets rises accordingly, indicating that high ranks are enriched for cross-source agreement (Figure 4). Operationally, this pattern supports using the top-k set as a robust, ready-made "not-to-hit" panel: even at k ≈ 50, the majority of entries carry convergent evidence from multiple lists, which is preferable for screening, modeling, or rule-based filtering.

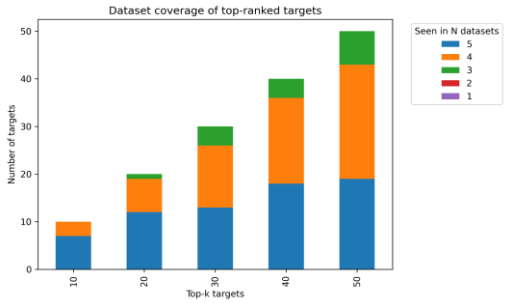

Figure 4. Dataset coverage of top-ranked targets (k = 10–50). Stacked bar chart showing, for each top-k threshold (10, 20, 30, 40, 50), how many of the k highest-ranked targets are present in exactly N independent evidence datasets (color legend at right, N = 1–5). Within each bar, colored segments sum to k.

The RRF score profile (Figure 5) exhibits a classic "heavy head, long tail." The leading target attains an RRF score of ~0.05, reflecting strong, multi-list support concentrated at extremely high ranks, whereas scores decay rapidly and approach ~0.00 across most of the list. Within the first few hundred ranks, the curve drops by an order of magnitude, and by the low-thousands rank, scores are already in the 10^{-3}–10^{-4} range. Such a motif indicates that only a relatively small subset of top-ranked targets accumulates substantial cross-evidence signal, while the bulk of entries in the tail are supported weakly—often by a single list or even a single source—yielding scores effectively near zero. Practically, the distribution supports using rank- or score-based cutoffs (e.g., focusing on the high-score head) when prioritizing "not-to-hit" targets for downstream analyses or prospective screening.

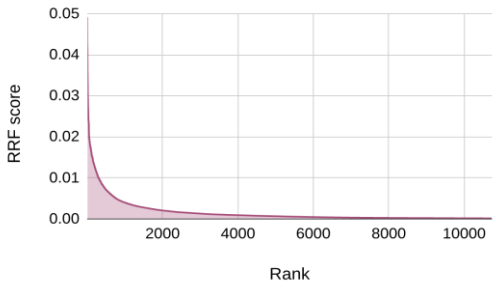

Figure 5. Reciprocal Rank Fusion (RRF) score distribution for the integrated De-TarDis ranking. Line/area plot showing RRF score (y-axis) as a function of global rank (x-axis) for all ranked targets (sorted by decreasing RRF). The curve is shaded to emphasize the score profile; axes are labeled "RRF score" and "Rank." Scores were computed by summing reciprocal-rank contributions across the five evidence groups and then plotted for the full De-TarDis list.

In conclusion, the species profile, cross-list membership, top-k coverage analyses, and the RRF score distribution indicate that De-TarDis is a systematically harmonized, evidence-weighted catalog of "not-to-hit" targets. After UniProtKB mapping and cross-group deduplication, the integrated resource comprises 10,736 unique targets (union across all evidence groups), while the larger raw tally arises from overlapping targets counted across multiple groups. The resulting collection retains a human-centric yet multi-species footprint, preserved as an explicit metadata field to support filtering and study design. Overlap across the five evidence groups is substantial at the head of the ranking—precisely where the stacked bars and dot matrix indicate frequent multi-dataset agreement—and then tapers into a long tail of single-source entries, reflected by the monotonic decay of RRF scores from ~0.05 to ~0.00. Such a well-designed structure makes De-TarDis practical at multiple operating points: compact, high-confidence panels (e.g., top 30–50 for quick screens or top 300–500 for modeling) capture convergent liabilities across sources, while the full set offers breadth for exhaustive safety sweeps. In short, De-TarDis couples rigorous cross-source integration with transparent organism tags and a rank/score framework, providing a well-structured, ready-to-use foundation for early risk triage, decoy selection, and safety-aware decision-making.

3.2. Group-wise safety profiles of target collections

Safety evidence in De-TarDis is heterogeneous by design: some signals arise from direct pharmacology (off-targets), others from structure-driven multi-ligand behavior (promiscuity), and still others from clinical or phenotypic readouts (adverse effects, side effects) or standardized safety screens (core panels). Analyzing these sources group-wise is therefore essential—each group encodes a distinct mechanism by which a target can become a "not-to-hit," and the operational actions that follow (e.g., assay choice, model prioritization, decoy selection) differ accordingly. De-TarDis supports both integrated, cross-source ranking and focused views when a project needs to privilege one evidence stream (for example, emphasizing side-effect concordance in late optimization or off-target pharmacology during hit triage). These focused views let users slice De-TarDis to match the decision at hand—zooming in on a single group for mechanism-specific questions or cross-checking patterns across groups when broader risk convergence is required. Therefore, the focused views have been investigated under five subsections: **(i)** Safety-relevant signals emerging from off-target sources, **(ii)** Prioritization of promiscuous targets in a safety context, **(iii)** High-risk targets highlighted by adverse-effect data, **(iv)** Targets recovered at the intersection of side-effect sources, and **(v)** Targets repeatedly present in core safety panels.

A) Safety-relevant signals emerging from off-target sources

Figure 6 shows how two off-target sources were scored before fusion, with clear separation and informative

spread in both. In the Pseudo Score Mean panel, the mean is ≈0.80, and the notched median sits slightly below the mean (Left boxplot, Figure 6), indicating a modest right skew; the interquartile range is comparatively broad, with numerous high-value outliers extending toward 1.0. In the off-target-rate (unique) panel (Right boxplot, Figure 6), the mean is ≈0.70, the notched median is close to the mean, and the Interquartile Range (IQR) is tighter, yielding a more compact distribution but still with a pronounced right tail. Such an contrast between boxplots—greater dispersion in the model-derived composite versus a tighter, rate-based distribution—ensures that De-TarDis ingests monotonic, well-stratified lists rather than near-ties. The broader IQR and richer tail of the Pseudo Score Mean (Left boxplot, Figure 6) ranking provide fine-grained internal ordering, while the off-target-rate ranking contributes a conservative, occurrence-normalized signal. Together, these properties stabilize within-source ranks and, after RRF, preferentially elevate proteins that are (i) consistently high across sources and (ii) not merely extreme in one dataset, improving the robustness of the integrated "do-not-hit" prioritization.

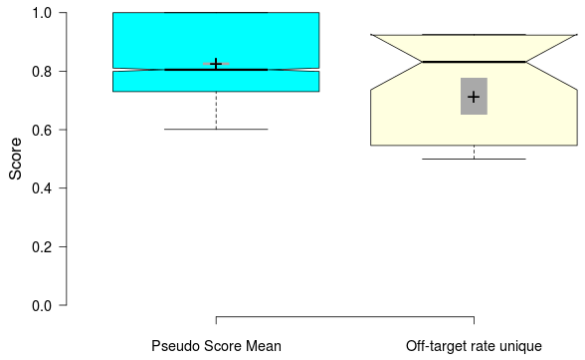

Figure 6. Demonstration of the internal scoring schemes used to order individual data sources before applying Reciprocal Rank Fusion (RRF). The boxplots display the score ranges for two representative sources: the dataset ranked by Pseudo Score Mean (left) and the off-target dataset ranked by unique off-target rate (right). Each box summarizes the distribution of per-target scores within its source, indicating how targets were linearly ordered prior to cross-dataset integration.

The fused list peaks at ~0.020 RRF and then drops rapidly (Figure 7): scores fall below ~0.005 by roughly rank 300–400, below ~0.001 by ~1,000, and approach ~0.0002 across the far tail. Compared with the cross-group ranking (whose head reaches ~0.050), this within-group peak is less than half as large, indicating weaker multi-source concordance inside this group—e.g., many targets are present in only one source or, when present in others, sit at much lower positions, which diminishes their reciprocal-rank contributions. Such a behavior (Figure 7) is expected for RRF: strong agreement concentrates near the very top, while disagreement or single-source evidence is relegated to the tail, yielding a stable, interpretable gradient that can be thresholded (e.g., by score or by top-k) for practical selection.

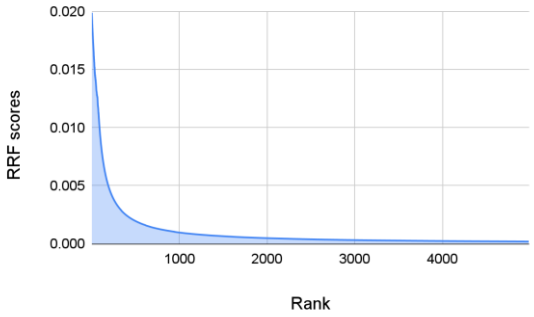

Figure 7. Rank–score profile of the off-target group after applying Reciprocal Rank Fusion (RRF). The curve shows RRF scores plotted against the final target ranks, illustrating the expected steep decay: a small set of targets accumulates higher fused evidence at the top of the list, while lower-ranked entries receive progressively smaller RRF contributions. This representation documents the monotonic ordering produced by RRF for subsequent safety-oriented analyses.

B) Prioritization of promiscuous targets in a safety context

The promiscuity-score distribution (Figure 8) shows ample spread and a high central tendency, supporting its use for prioritization in a safety context. The mean is ~0.60—clearly above the 0.50 midline—while the notched median appears slightly higher (≈0.65) (Figure 8), indicating that a majority of targets exhibit above-moderate promiscuity. The interquartile range spans roughly the mid-0.3s to high-0.8s, with whiskers extending from near 0 to ~1, which confirms both diversity (useful for ranking) and the presence of a well-defined "high-promiscuity" tail (Figure 8). Practically, this separation enables simple, thresholdable rules (e.g., prioritize ≥0.6–0.7) to surface candidates with broader binding/liability potential for early exclusion or intensified triage. The wide dispersion also suggests heterogeneous evidence sources (e.g., assay density, structure coverage) are contributing complementary signals rather than collapsing to ties, which is desirable for downstream fusion and for constructing compact "not-to-hit" panels.

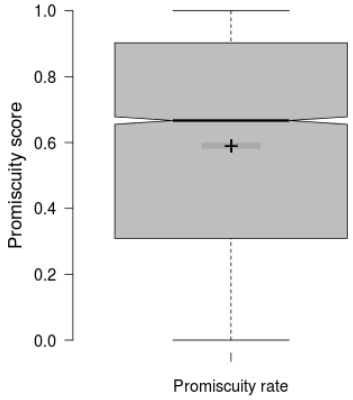

Figure 8. Distribution of PDBBind-derived target promiscuity. Notched box plot summarizing the target-level PromiscuityScore (y-axis, normalized 0–1) computed from the PDBBind dataset. The central horizontal line is the median; the notch approximates the 95% confidence interval around the median; the box spans the interquartile range (Q1–Q3); whiskers extend to 1.5×IQR; individual outliers (if any) lie beyond the whiskers. The gray "+" marks the mean. The x-axis label indicates the underlying promiscuity rate used in the score construction.

The promiscuity-specific RRF profile (Figure 9) exhibits a steep head-tail pattern with a maximum fused score of ~0.02, comparable to the within-group peak observed for the off-target set and roughly half of the ≈0.05 head value seen in the cross-group (all-evidence) ranking. This attenuation is expected: two of the three promiscuity sources do not provide per-target ranks or calibrated scores, so targets shared across sources inherit a single summary statistic (the PDBbind-derived mean) (Figure 9), which compresses between-source contrast and moderates reciprocal-rank contributions. The resulting curve implies only modest concordance among promiscuity sources at the top—high enough to distinguish a small set of consistently promiscuous proteins, but lower than the cross-group consensus where heterogeneous evidence (off-target, adverse/side-effect, safety-panel) accumulates. Notably, using the median in place of the mean yields a near-identical head magnitude, indicating that the fused scores are driven by the presence/absence pattern across sources rather than by a few extreme measurements, and underscoring the robustness—but also the conservative dynamic range—of the promiscuity aggregation.

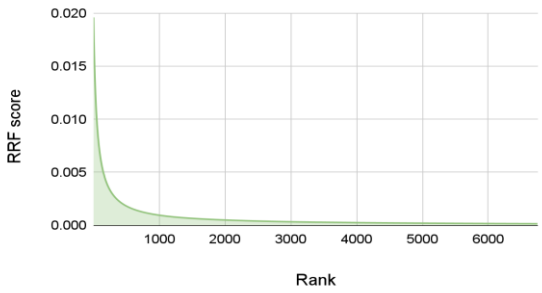

Figure 9. Reciprocal Rank Fusion (RRF) score distribution for the promiscuity list compiled by merging three sources. The area/line plot shows the RRF score (y-axis) assigned to each target as a function of its final rank (x-axis), with higher-ranked targets on the left. The curve is shaded under the line to visualize density across ranks; axes are labeled "RRF scores" and "Rank."

C) High-risk targets highlighted by adverse effect data

After min–max scaling, the five adverse-effect sources retain distinct score geometries that reflect their underlying evidence regimes (Figure 10). Source 3 (*Li et al.)* shows the highest central tendency and the most compact, informative spread (median ≈0.18–0.20; IQR roughly 0.12–0.25), consistent with a modeled/domain-to-protein signal that distributes weight across many targets rather than a few extreme hits. In contrast, Source 1 (ADReCS-Target) and Source 5 (TARDIS-derived) exhibit heavy right tails: medians near 0.04–0.06 with numerous outliers extending to ~1.0, a pattern expected from occurrence/enrichment style counts where a relatively small subset of proteins repeatedly co-appear with ADR terms. Sources 2 (Huang et al.) and 4 (ChEMBL adverse-effect library) are markedly compressed toward zero (medians ≲0.02), indicating sparse or conservative scoring that flags only a thin layer of high-scoring proteins above a broad low-signal background. Moreover, for integration, these differences are desirable: Tail-heavy sources (1, 5) supply discriminative "spikes" that help elevate repeatedly implicated proteins, while Source 3 contributes breadth and stabilizes within-source ranks through a wider IQR; the two compressed sources (2, 4) act as stringent filters that corroborate only the strongest candidates. Because De-TarDis ranks within each source before fusion, this heterogeneity does not bias the aggregate; instead, Reciprocal Rank Fusion (Eq. 1) exploits it to reward targets that are consistently high across diverse scoring philosophies (model-based, occurrence-driven, and enrichment-derived), thereby enriching for adverse-effect liabilities that are both recurrent and cross-validated.

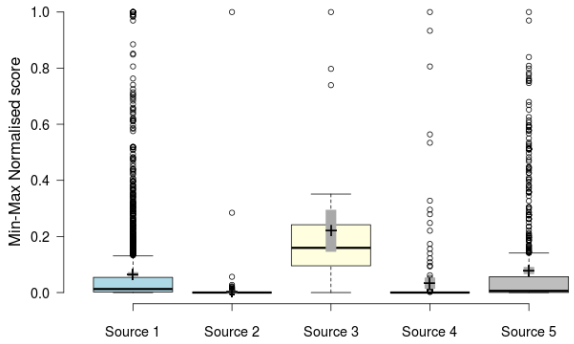

Figure 10. Min–max–normalized adverse effect evidence scores across five sources. Box-and-whisker plots summarize the per-target score distributions after column-wise min–max normalization (value = (x−minf_0)/(maxf_0−minf_0), applied within each source; constants mapped to 0). For each box, the central line is the median, the box spans the interquartile range (IQR), whiskers extend to 1.5×IQR, open circles mark points beyond the whiskers (outliers), and the "+" indicates the mean. Sources for the Adverse Effects: Source 1 = ADReCS-Target [49], Source 2 = *Soldatos et al.* [50], Source 3 = *Iwata et al.* [51], Source 4 = *Hindle et al.* [52], Source 5 = TARDIS-derived ADR–targets [53].

The adverse-effect fusion exhibits the sharpest "head" among all groups: the top RRF score is ≈0.064 (Figure 11)—higher than the within-group peaks for off-target (≈0.020) and promiscuity (≈0.020), and even above the previously noted cross-group peak (~0.050). This elevated head implies stronger internal concordance among adverse-effect sources, with fewer conflicts than in the other groups. Also, the curve then declines rapidly—≈0.02 by ~rank 100, ≈0.01 by ~250, and ≈0.005 by ~500—before tapering to ≈0.001 near ~1300 and approaching zero toward ~1600 (Figure 11). Taken together, the profile indicates that adverse-effect evidence is highly concentrated in a relatively small subset of proteins at the top (where multiple datasets align), whereas the long tail consists of targets with weak, sparse support.

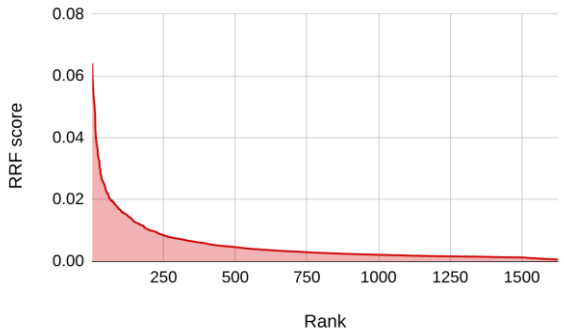

Figure 11. Adverse-effect evidence, RRF score vs. rank (five sources [49-53]). Line/area plot showing per-target Reciprocal Rank Fusion (RRF) score (y-axis) as a function of the target's position in the adverse-effect-oriented ranking (x-axis). Scores were obtained by summing 1/(k+rank) across five adverse-effect sources (fixed damping for each UniProt target); higher curves indicate larger fused scores. The red line traces the ordered RRF scores, and the shaded region highlights the cumulative area under the curve. Tick labels on the x-axis correspond to rank positions; the y-axis units are fused RRF scores.

D) Targets recovered at the intersection of side effect sources

Across the five side-effect sources [54-58], score distributions are heterogeneous and strongly skewed, indicating complementary information for fusion (Figure 12). Sources 1–3 [54-56] show significantly low medians near the floor after min–max normalization, with tight IQRs and long upper tails populated by numerous outliers—consistent with sparse but sharp signals for a limited subset of targets. Source 4 (CCA-based) [57] shifts upward with a wider IQR and a mean above the median, suggesting broader—but still right-skewed—coverage. Source 5 (GESSE) [56] exhibits the highest central tendency: both mean and median lie well above the other sources, and the IQR is comparatively large, implying a denser distribution of informative scores rather than isolated spikes (Figure 12). In general, these shapes justify treating the five resources as complementary layers: min-max scaling renders them comparable, while the mix of "spiky" (Sources 1–3) [54-56], moderate (Source 4) [47], and high-density (Source 5) [58] profiles provides variance that a rank-fusion step can exploit to elevate targets repeatedly implicated across independent side-effect lines of evidence (Figure 12). Consequently, such a high heterogeneity underpins De-TarDis's design as a comprehensive, rank-fusion–ready "not-to-hit" target resource for practical liability screening.

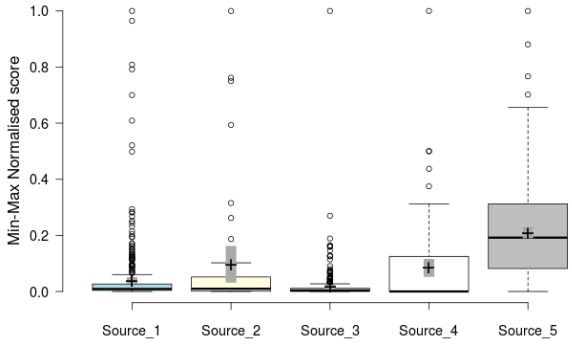

Figure 12. Side-effect evidence distributions across five sources after min–max normalization. Box and whisker plots show per-target scores from five side-effect–oriented resources, each normalized within-source to the 0,1 range using x′=(x−min(x))/(max(x)−min(x)) before plotting. Source labels correspond to: Source 1—Kyoto target-

effect dataset (Mizutani et al.) [54]; Source 2—Systematic identification of proteins that elicit drug side effects [55]; Source 3—Inferring protein domains associated with drug side effects based on drug-target interaction network [56]; Source 4—Relating Essential Proteins to Drug Side Effects using Canonical Component Analysis [47]; Source 5—GESSE (Predicting Drug Side Effects from Drug-Target Relationships) [56]. Boxes denote the interquartile range (IQR), the horizontal line is the median, the "+" marks the mean, whiskers extend to 1.5×IQR, and open circles indicate values outside the whisker range.

The side-effect fusion (RRF score distribution, Figure 13) exhibits a strong but slightly lower head than the adverse effect and a clearer plateau than off-target or promiscuity. The leading RRF score is ≈0.048 (just shy of the overall cross-group peak of ~0.050 and below the adverse-effect head at ~0.064), indicating substantial multi-source concordance among the very top side-effect targets, yet not as concentrated as in the adverse-effect case. The curve then decays smoothly: ≈0.02 by the ~50th rank, ~0.01 by ~200, ~0.004 by ~600, and ~0.002 near the 1,000th target (Figure 13). Relative to the off-target and promiscuity profiles (both peaking around ~0.020), the side-effect profile maintains higher fused scores across the top few hundred ranks, consistent with broader agreement among its contributing sources. The longer "shoulder" suggests that the informative side-effect signal persists beyond the very top ranks, making the top ~200–400 side-effect targets a practical band for downstream liability checks, while the tail primarily captures weaker, single-source indications.

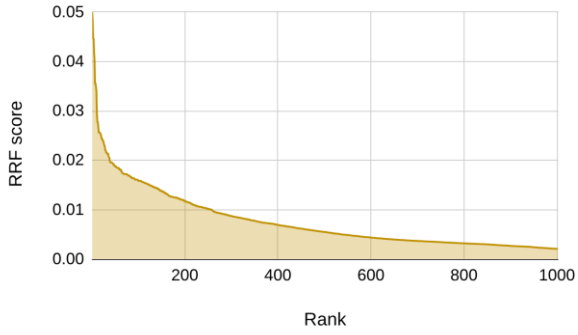

Figure 13. Side-effect RRF score profile. Line/area plot of Reciprocal Rank Fusion (RRF) scores versus target rank for the side-effect evidence set after cross-source integration (k = 50). Targets are sorted by decreasing RRF score; the x-axis shows the rank index and the y-axis shows the aggregated RRF score per UniProt target. The shaded region visualizes the score curve, where each point corresponds to one target derived from the five side-effect sources.

E) Targets repeatedly present in core safety panels

The step profile (Figure 14) indicates strong cross-tier recurrence among the highest-ranked safety-panel targets. In the head of the list (≈ranks 1–200), many entries register in ≥2 tiers, with intermittent tier-class = 3 spikes that reflect targets present across all Tier-1/2/3 panels (Figure 14). From ~200 to ~330 the curve stabilizes at tier-class = 2, showing that most targets are repeatedly profiled across two independent tiers. A pronounced drop to tier-class = 1 occurs around ~350, after which the tail remains largely single-tier up to ~450—consistent with items that are present but not broadly replicated across panels. Interpreted operationally, the plot supports simple cutoffs: the top ~50–100 provide the densest multi-tier consensus, ~100–330 maintain robust two-tier coverage suitable for expanded screening, and beyond ~350 the marginal benefit shifts toward breadth rather than replicated evidence.

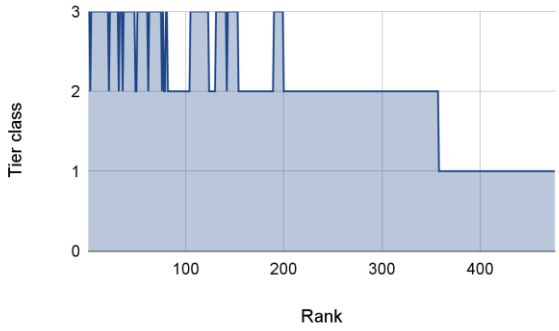

Figure 14. Step plot profile of targets repeatedly present across core safety-panel tiers. The step plot shows how many panel tiers (Tier-1/2/3; y-axis values 1–3) include each ranked target on the x-axis, while the filled area highlights cumulative coverage. Ranking was computed from the merged safety-panel table by sorting (descending) on (i) number of files/panels reporting the target (n_files_present), (ii) count of tiers in which the target appears (tiers_present), (iii) assay subtype weight (Functional = 3, Cell-based = 2, Binding = 1), and (iv) target-family weight (GPCR/Ion channel = 3; Transporter/Nuclear receptor = 2; others = 1); ties were broken alphabetically by target name.

In summary for the group-wise analysis, each evidence family contributes a distinct liability signal and characteristic score geometry, and their fused rank–score profiles show complementary strengths. Off-target and promiscuity lists display modest heads (≈0.02) with long tails, indicating heterogeneous, partly single-source support; side-effect yields a stronger and longer shoulder (head ≈0.048), and adverse-effect exhibits the sharpest head (≈0.064), reflecting tighter internal concordance. Notably, in four of five families the within-group peak RRF is lower than the cross-group peak (~0.050), underscoring that De-TarDis's integration improves robustness by reconciling conflicts across sources and promoting proteins repeatedly flagged by independent lines of evidence to the very top of the global ranking. Operationally, this validates two complementary usages: (i) focused, mechanism-specific triage within a single group when the question is narrow, and (ii) cross-group prioritization when seeking high-confidence "not-to-hit" targets supported by convergent data.

IV. CASE STUDY: PRACTICAL USAGE OF DE-TARDIS

To demonstrate the practical utility of De-TarDis in an applied safety context, seven withdrawn drugs were randomly selected from the literature and mapped to their reported primary molecular targets at the UniProtKB level (Table 2). The case study is intentionally simple and transparent: rather than introducing new experimental measurements, it tests whether De-TarDis—constructed by fusing heterogeneous "do-not-hit" evidence streams—can recover historically established withdrawal-linked targets within its integrated ranking. Importantly, for each of the seven examples, the literature-reported target(s) are already present in De-TarDis, indicating that the database captures many clinically consequential liabilities without requiring any drug-specific tuning.

Table 2. Case study illustrating practical use of De-TarDis using seven withdrawn drugs. For each drug, the primary UniProtKB target(s) reported in the literature are listed together with the corresponding De-TarDis global rank and the percentile-style ratio (%), computed as (rank / total number of De-TarDis targets) × 100, and the supporting literature reference.

Withdrawal drug	Target	De-TarDis Rank	Ratio (%)	Ref
Rofecoxib (Vioxx)	P35354	2	0.019	[71]
Benoxaprofen (Opren)	P35354; P23219	2; 11	0.019; 0.102	[72]
Pergolide	P14416	5	0.047	[73]
Terfenadine	P35367	41	0.382	[74]
Cerivastatin (Baycol/Lipobay)	P04035	47	0.438	[75]
Troglitazone	P37231	137	1.276	[76]
Rimonabant	P21554	239	2.226	[77]

A second, more stringent aspect of the case study concerns prioritization rather than mere presence. As shown in Table 2, the withdrawal-associated targets cluster near the head of the fused ranking, falling within approximately the top 3% of all UniProtKB entries contained in De-TarDis (as reflected by the Ratio (%) metric derived from the rank position). This concentration is consistent with the design goal of De-TarDis: targets repeatedly supported by off-target, promiscuity, adverse-effect/side-effect, and safety-panel signals should accumulate higher reciprocal-rank mass and therefore appear earlier in the final list. From a user perspective, this behavior is operationally valuable because a relatively small "top slice" of De-TarDis already contains many targets that have been implicated in severe post-marketing outcomes, providing an actionable starting point for early de-risking and hypothesis generation.

Such a case study also mirrors the intended user workflow: a drug (or a hit compound) is first mapped to a UniProtKB-standardized target set, after which De-TarDis is queried to retrieve (i) whether the target appears in the database and (ii) its global rank/percentile as an evidence-weighted proxy for safety liability priority. The same lookup step can be applied prospectively to new hits by substituting literature-reported targets with predicted/off-target candidates from docking, chemogenomics, or panel profiling.

The case study also serves as a qualitative validation of the Reciprocal Rank Fusion damping regime adopted in De-TarDis. While the fused ordering is, by construction, influenced by the damping constant k_{RRF}, Table 2 illustrates that the central signal here is robust: the identified withdrawal-linked targets remain strongly prioritized under the current default k_{RFF}=50. Conceptually, smaller k_{RFF} values amplify head dominance by granting greater relative weight to top-ranked positions within each contributing list, whereas larger values attenuate this effect and distribute contribution more evenly across ranks. In practical terms, modest re-parameterization may improve one or two individual drug–target positions but can simultaneously reduce others, implying that overall utility is expected to remain stable rather than exhibiting dramatic rank reshuffling—consistent with the broader sensitivity behavior of RRF reported in the information-retrieval literature and in our own stability notes elsewhere in the manuscript [67].

Beyond the fusion hyperparameter, Table 2 highlights a more consequential downstream decision: the operational decoy cut-off selected by the user (e.g., screening the top-10, top-50, or top-100 targets). Because De-TarDis integrates evidence across multiple heterogeneous sources, restricting analysis to a very small top set will inevitably exclude additional plausible liabilities that are supported by fewer lists or lower within-list ranks but remain safety-relevant. Conversely, screening a larger fraction of De-TarDis reduces sensitivity to minor rank perturbations (including those induced by k_{RFF}), because the dominant cross-group overlap and recurrent targets are captured regardless of small ordering differences.

In summary, this case study provides an interpretable, literature-grounded demonstration that De-TarDis (i) contains the canonical targets of several well-known withdrawn drugs and (ii) prioritizes these targets near the top of its integrated ranking (Table 2). While this analysis is not intended as prospective validation, it shows that De-TarDis behaves as a practical "safety-first" navigation layer: it consolidates heterogeneous liability signals into a unified, UniProtKB-standardized list that can be readily used to guide early-stage target avoidance, prioritize follow-up pharmacology, and structure decoy-based screening strategies in a reproducible manner.

V. LIMITATIONS AND FUTURE DIRECTIONS

De-TarDis is an integrative first release that fuses 49 rank lists into a single, harmonized "decoy-target" catalog. This version already enables substantial cross-source comparison and provides a unified database for safety-relevant target prioritization. Nevertheless, coverage is not intended to be exhaustive at this stage. Future versions can be expanded in two complementary ways: (i) increasing the number of integrated lists (e.g., adding additional off-target, safety pharmacology, ADR, and side-effect resources as they become available), and (ii) broadening the target-space definition by incorporating new safety-relevant target groups and panel types. Accordingly, De-TarDis should be viewed as a scalable framework whose list count and target-group scope can be systematically enlarged to improve comprehensiveness.

Regarding rank fusion, we currently use a fixed Reciprocal Rank Fusion damping parameter of k_{RRF} = 50. Sensitivity results indicate that moderate changes in k_{RRF} produce negligible differences in effectiveness. Specifically, performance is essentially stable across commonly used settings: k_{RRF} = 50 (MAP ≈ 0.2144), k_{RRF} = 60 (MAP ≈ 0.2145), and k_{RRF} = 100 (MAP ≈ 0.2142). In contrast, extreme values reduce effectiveness (k_{RRF} = 0, MAP ≈ 0.2072; k_{RRF} = 500, MAP ≈ 0.2098) [65]. This stability implies that, in practical downstream use, the dominant factor is typically not fine-tuning k_{RRF}, but deciding how many top-ranked UniProtKB targets to operationalize as decoys (e.g., selecting the top-5, top-50, or top-100). Using only a small top slice will inevitably omit many plausible decoys, whereas screening a large fraction of De-TarDis reduces the importance of exact ordering because most cross-group overlapping targets are included regardless of minor rank shifts.

Conceptually, k_{RRF} controls how strongly the fusion emphasizes top-ranked positions: smaller k_{RRF} increases the influence of top entries within each list, while larger k_{RRF} dampens head dominance and spreads contributions more evenly across ranks. The optimal setting is therefore application-dependent (e.g., aggressive early hazard identification versus broader exploratory screening). To support transparency and user-specific optimization, De-TarDis provides the curated inputs and end-to-end Python scripts so users can (i) vary k_{RRF} and (ii) choose an application-specific decoy cut-off without changing the underlying data model.

De-TarDis is intentionally target-centered: promiscuity is modeled at the protein (UniProtKB) level by aggregating multi-source evidence into a unified safety-oriented target ranking. This design provides a consistent "do-not-hit" map for early de-risking, but it does not explicitly capture the ligand-centered dimension of promiscuity—i.e., the tendency of particular chemical structures (or chemotypes) to engage multiple targets in a context-dependent manner. A complementary, ligand-centered analysis would be especially valuable for assessing polypharmacology-prone scaffolds and for distinguishing whether multi-target binding is driven by a small set of chemical liabilities versus broad target susceptibility. Integrating De-TarDis with a ligand-centered promiscuity resource (e.g., scaffold/chemotype-level multi-target propensity or promiscuity indices derived from chemogenomics activity matrices) is therefore a clear future direction that would increase practical utility by enabling joint target–ligand risk profiling.

Overall, De-TarDis should be viewed as a scalable, transparent framework rather than a closed final catalog: its coverage can be expanded by incorporating additional evidence lists and safety-relevant target families, while its ranking behavior remains robust to moderate variations in the RRF damping parameter. In downstream workflows, the most consequential operating choice is typically the decoy cut-off (how many top-ranked targets to screen), whereas fine tuning of k_{RRF} has a secondary effect within the stable regime. Finally, complementing the current protein-level safety map with a ligand-centered promiscuity layer represents a practical path to a more holistic and realistic safety model, supporting better decision-making for both target avoidance and chemotype-level liability management.

VI. CONCLUSION

De-TarDis consolidates heterogeneous safety evidence into a single, rank-ordered "do-not-hit" resource spanning five evidence families—off-target, promiscuity, adverse-effect, side-effect, and safety-panel targets—compiled from 49 input lists after harmonization to UniProtKB. The cross-source integration with Reciprocal Rank Fusion (RRF) yields a consensus ranking that rewards repeated high placements across independent studies. Quantitatively, the integrated list contains 10,736 unique targets after cross-group deduplication (union across evidence groups). Its rank–score profile exhibits a pronounced head and a long tail, consistent with strong multi-dataset agreement among top-ranked entries and weaker, often single-source support among lower-ranked targets. The difference between the summed group-wise tallies and the unique union thus provides a direct, interpretable indicator of evidence overlap across heterogeneous safety-related sources. Group-wise profiles display complementary geometry: off-target and promiscuity heads peak near ≈0.02, side-effect rises to ≈0.048, and adverse-effect shows the sharpest head at ≈0.064—together explaining why cross-group RRF concentrates high-confidence liabilities at the very top. These findings, alongside species annotations retained for every entry, establish De-TarDis as a standardized, evidence-weighted catalog that can be filtered and thresholded to match different operating points (e.g., compact panels or broader sweeps).

Practically, De-TarDis is designed to drop directly into discovery workflows. In early triage, teams can screen a compact top-k (e.g., ~30–50) to capture liabilities supported by several independent sources; for modeling-intensive selectivity analyses, a broader head (e.g., ~300–500) balances coverage with cross-evidence agreement. The resource's within-source normalization and cross-group RRF reduce single-study bias, promote convergent signals, and make operating thresholds transparent (rank or score). Because each group reflects a distinct mechanism—direct pharmacology, multi-ligand behavior, clinical/phenotypic implication, or standardized safety coverage—users can either focus on a single group for mechanism-specific questions or rely on the global RRF consensus when seeking robust, convergent "not-to-hit" targets for decision-making.

Accordingly, De-TarDis provides a usage framework that is both threshold-driven and reproducible: the same ranked resource supports a compact operating mode (small top-k for rapid triage) and a broad operating mode (larger top-k for comprehensive counter-screening), while preserving traceability to the underlying evidence families. This makes the decision logic explicit—targets are avoided or prioritized for follow-up not by ad hoc selection, but by a UniProtKB-standardized, cross-source rank consensus that can be re-parameterized (cut-off choice) to match different project risk tolerances and screening budgets.

ACKNOWLEDGEMENTS

We thank the maintainers and curators of the public resources and literature that underpin this work, including UniProtKB for identifier standardization; BindingDB and PDBbind for structure-centric interaction evidence; adverse/side-effect data providers aggregated from the literature; and safety-panel references modeled on Eurofins SafetyScreen™ tiers. Their efforts made possible the systematic sourcing, harmonization, and integration that define De-TarDis. We are also grateful to colleagues and reviewers whose feedback improved the normalization rules, fusion settings, and visual analytics presented here.

SUPPLEMENTARY INFORMATION

Across the mid-range cutoffs (top-75 to top-175) (Figure S1), the stacked bars show a steady, near-linear growth in total targets while the composition gradually

diversifies: the five-dataset (dark blue) and four-dataset (orange) layers dominate at smaller k, but as k increases, the three-, two-, and one-dataset layers (green/red/purple) contribute progressively more of the mass (Figure S1). In other words, the core of the ranking (≈top 75–100) is enriched for targets corroborated by four or five independent evidence lists, whereas expanding the window to 125–175 introduces a larger fraction of targets supported by only one to three lists—useful for breadth, but with weaker cross-dataset agreement.

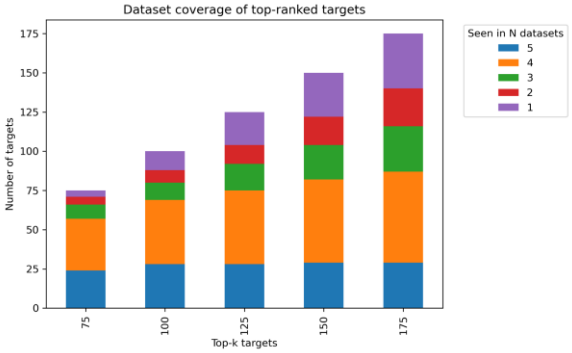

Figure S1. Dataset coverage of top-ranked targets (k = 75–175). Stacked bar chart for top-k thresholds 75, 100, 125, 150, and 175. Bars are partitioned by the number of datasets in which each target appears (N = 1–5, colors as indicated). Segment heights within a bar sum to k.

Across top-k thresholds of 200→400 (Figure S2), the stacked bars show a clear shift from multi-source to single-source coverage: as k grows, the portion confirmed in ≥4 datasets (blue+orange) rises only modestly, while the segments seen in 3, 2, or just 1 dataset (green, red, purple) expand steadily—especially the single-dataset segment, which becomes the largest slice by k≈350–400 (Figure S2). In practice, this pattern indicates a depth-to-breadth transition: the highest-ranked few hundred targets are repeatedly supported across evidence lists, but enlarging the cut-off primarily adds targets that are present in fewer sources. Such a structure helps pick operating points: use smaller k for high-confidence, cross-validated "not-to-hit" sets, and larger k when broader—but thinner—coverage is acceptable.

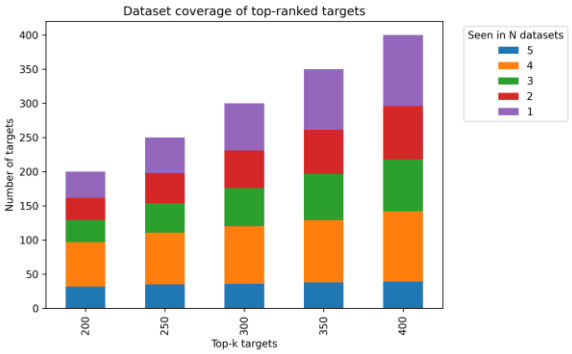

Figure S2. Dataset coverage of top-ranked targets (k = 200–400). Stacked bar chart for thresholds 200, 250, 300, 350, and 400. For each k, the bar is decomposed into counts of targets observed in exactly N datasets (N = 1–5). The total height of each bar equals k.

The stacked bars, Figure S3, summarize how widely each prefix of the ranked list is supported across the five evidence groups. Coverage is high among the very top ranks, but as k approaches 900, the composition shifts toward single-source memberships (purple), indicating that lower-ranked items are predominantly supported only by one dataset. This pattern also explains a key property of De-TarDis: although the integrated resource contains 10,736 unique targets after cross-group deduplication, the low-ranked tail (i.e., targets beyond the top-N displayed in Figure S3) is dominated by entries contributed primarily by the structure-centric promiscuity build (PDBBind-derived) with comparatively limited cross-group reinforcement. In other words, the head of the ranking is enriched for targets supported by multiple evidence streams, whereas the extreme tail more often reflects single-group support. Only a small fraction of those tails also carry deep, late off-target evidence, and when present, their ranks sit near the bottom of the off-target list. Illustrative tail just over the 9000th entries include O00337 (4712 in promiscuity rank), P11412 (4713 in promiscuity rank), P41512 (4713 in off-target rank), and P30189 (4714 in off-target rank)—all typical of single-dataset support in the extreme tail. Practically, this stratification suggests that decoy panels for routine screening can be drawn from the head of the list (e.g., top 300–500), where multi-dataset agreement is common, while the deepest ranks are best reserved for broad "long-tail" analyses or when maximal coverage is desired.

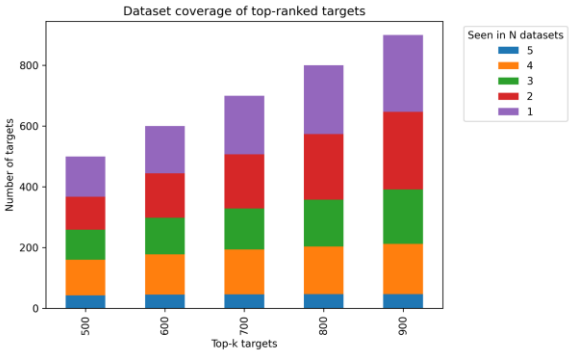

Figure S3. Dataset coverage of top-ranked targets (k = 500–900). Stacked bar chart for thresholds 500, 600, 700, 800, and 900. Colored segments report, for the top-k targets, the number present in exactly N datasets (N = 1–5; legend at right). Segment totals per bar equal k.

REFERENCES

[1] Prajapati, A. (2025). COMPUTER-AIDED DRUG DISCOVERY: TRANSFORMING THE LANDSCAPE OF PHARMACEUTICAL RESEARCH. Frontline Chemistry Nexus, 1(1), 57-91.

[2] Khurshid, B., Khurshid, S., Rafique, A. M., & Javaid, M. (2025). AI for Drug Discovery: From Algorithms to Medicines. Journal of Pharma and Biomedics, 3(2), 171-184.

[3] dos Santos Nascimento, I. J., & de Moura, R. O. (2023). Ligand and structure-based drug design (LBDD and SBDD): Promising approaches to discover new drugs. In Applied Computer-Aided Drug Design: Models and Methods (pp. 1-32). Bentham Science Publishers.

[4] Saini, M., Mehra, N., Kumar, G., Paul, R., & Kovács, B. (2025). Molecular and structure-based drug design: From theory to practice. In Advances in Pharmacology (Vol. 103, pp. 121-138). Academic Press.

[5] Lionta, E., Spyrou, G., K Vassilatis, D., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. Current topics in medicinal chemistry, 14(16), 1923-1938.

[6] Horvath, D. (2010). Pharmacophore-based virtual screening. Chemoinformatics and computational chemical biology, 261-298.

[7] Taft, C. A. (2014). Current state-of-the-art for virtual screening and docking methods. In New Developments in Medicinal Chemistry: Volume 2 (pp. 3-169). Bentham Science Publishers.

[8] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Tunyasuvunakool, K., ... & Hassabis, D. (2020). AlphaFold 2. Fourteenth Critical Assessment of Techniques for Protein Structure Prediction, 13.

[9] Ruff, K. M., & Pappu, R. V. (2021). AlphaFold and implications for intrinsically disordered proteins. Journal of molecular biology, 433(20), 167208.

[10] Sanchez Marin, M. (2025). Large-scale protein structure prediction methods for enhanced annotation (Master's thesis).

[11] Wuyun, Q., Chen, Y., Shen, Y., Cao, Y., Hu, G., Cui, W., ... & Zheng, W. (2024). Recent progress of protein tertiary structure prediction. Molecules, 29(4), 832.

[12] Abdin, O. (2024). Novel deep learning methods for the modelling and design of peptides and small proteins (Doctoral dissertation, University of Toronto (Canada)).

[13] Chen, X., Ji, Z. L., & Chen, Y. Z. (2002). TTD: therapeutic target database. Nucleic acids research, 30(1), 412-415.

[14] Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., ... & Hassanali, M. (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research, 36(suppl_1), D901-D906.

[15] Wishart, David S., et al. "DrugBank 5.0: a major update to the DrugBank database for 2018." Nucleic acids research 46.D1 (2018): D1074-D1082.

[16] Knox, Craig, et al. "DrugBank 6.0: the DrugBank knowledgebase for 2024." Nucleic acids research 52.D1 (2024): D1265-D1275.

[17] Koscielny, G., An, P., Carvalho-Silva, D., Cham, J. A., Fumis, L., Gasparyan, R., ... & Dunham, I. (2017). Open Targets: a platform for therapeutic target identification and validation. Nucleic acids research, 45(D1), D985-D994.

[18] Gao, Z., Li, H., Zhang, H., Liu, X., Kang, L., Luo, X., ... & Jiang, H. (2008). PDTD: a web-accessible protein database for drug target identification. BMC bioinformatics, 9(1), 104.

[19] Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., ... & Jiang, H. (2010). PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. Nucleic acids research, 38(suppl_2), W609-W614.

[20] Tanoli, Z., Alam, Z., Vähä-Koskela, M., Ravikumar, B., Malyutina, A., Jaiswal, A., ... & Aittokallio, T. (2018). Drug Target Commons 2.0: a community platform for systematic analysis of drug–target interaction profiles. Database, 2018, bay083.

[21] Girotra, K., Terwisch, C., & Ulrich, K. T. (2005). Managing the risk of development failures: A study of late-stage failures in the pharmaceutical industry. The Wharton School, University of Pennsylvania, January.

[22] Sun, A., & Benet, L. Z. (2020). Late-stage failures of monoclonal antibody drugs: a retrospective case study analysis. Pharmacology, 105(3-4), 145-163.

[23] Parasrampuria, D. A., Benet, L. Z., & Sharma, A. (2018). Why drugs fail in late stages of development: case study analyses from the last decade and recommendations. The AAPS journal, 20(3), 46.

[24] Kaplowitz, N. (2005). Idiosyncratic drug hepatotoxicity. Nature reviews Drug discovery, 4(6), 489-499.

[25] Ulrich, Roger G. "Idiosyncratic toxicity: a convergence of risk factors." Annu. Rev. Med. 58.1 (2007): 17-34.

[26] Spławiński, J., Kuźniar, J., Filipiak, K., & Zieliński, W. (2006). Evaluation of drug toxicity in clinical trials. Science and Engineering Ethics, 12(1), 139-145.

[27] Guengerich, F. P. (2011). Mechanisms of drug toxicity and relevance to pharmaceutical development. Drug metabolism and pharmacokinetics, 26(1), 3-14.

[28] Onakpoya, I. J., Heneghan, C. J., & Aronson, J. K. (2016). Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. BMC medicine, 14(1), 10.

[29] Rao, N., & Powar, R. (2023). Post-marketing drug withdrawals: a review. Pharmaceutical Chemistry Journal, 57(7), 1138-1146.

[30] Ballesteros, J., & Ransom, J. (2006). In-target versus off-target allosteric modulators of GPCRs. Drug Discovery Today: Therapeutic Strategies, 3(4), 445–450.

[31] Raschi, E., Vasina, V., Poluzzi, E., & De Ponti, F. (2008). The hERG K+ channel: target and antitarget strategies in drug development. Pharmacological Research, 57(3), 181–195.

[32] Mahé, J., de Campaigno, E. P., Chené, A.-L., Montastruc, J.-L., Despas, F., & Jolliet, P. (2018). Pleural adverse drugs reactions and protein kinase inhibitors: Identification of suspicious targets by disproportionality analysis from VigiBase. British Journal of Clinical Pharmacology, 84(10), 2373–2383.

[33] Brunetti, L., Francavilla, F., Leopoldo, M., & Lacivita, E. (2024). Allosteric modulators of serotonin receptors: a medicinal chemistry survey. Pharmaceuticals, 17(6), 695.

[34] Hilfiger, L., Zhao, Q., Kerspern, D., Inquimbert, P., Andry, V., Goumon, Y., ... & Charlet, A. (2020). A nonpeptide oxytocin receptor agonist for a durable relief of inflammatory pain. Scientific Reports, 10(1), 3017.

[35] Lagunas-Rangel, F. A., Liepinsh, E., Fredriksson, R., Alsehli, A. M., Williams, M. J., Dambrova, M., Jönsson, J., & Schiöth, H. B. (2024). Off-target effects of statins: molecular mechanisms, side effects and the emerging role of kinases. British Journal of Pharmacology, 181(20), 3799–3818.

[36] Chen, L., He, Y., Wang, X., Ge, J., & Li, H. (2021). Ventricular voltage-gated ion channels: Detection, characteristics, mechanisms, and drug safety evaluation. Clinical and Translational Medicine, 11(10), e530.

[37] Atabakhsh, E., & Schild-Poulter, C. (2012). RanBPM is an inhibitor of ERK signaling. PLoS One, 7(10), e47803.

[38] Benton, Jennifer. Chemoenzymatic applications and catalytic mechanism of 7-epi-zingiberene synthase. Diss. Cardiff University, 2022.

[39] Onakpoya, I. J., Heneghan, C. J., & Aronson, J. K. (2016). Worldwide withdrawal of medicinal products because of adverse drug reactions: a systematic review and analysis. Critical reviews in toxicology, 46(6), 477-489.

[40] Lexchin, J. (2005). Drug withdrawals from the Canadian market for safety reasons, 1963–2004. Cmaj, 172(6), 765-767.

[41] Craveiro, N. S., Lopes, B. S., Tomás, L., & Almeida, S. F. (2020). Drug withdrawal due to safety: a review of the data supporting withdrawal decision. Current drug safety, 15(1), 4-12.

[42] Ji, Z. L., Han, L. Y., Yap, C. W., Sun, L. Z., Chen, X., & Chen, Y. Z. (2003). Drug Adverse Reaction Target Database (DART): Proteins related to adverse drug reactions. Drug Safety, 26(10), 685–690.

[43] Mohseni, S., Tabatabaei-Malazy, O., Peimani, M., Ejtahed, H. S., Khodaeian, M., Nazeri, E., ... & Larijani, B. (2021). Withdrawal reasons of randomized controlled trials on type 2 diabetes: a systematic review. DARU Journal of Pharmaceutical Sciences, 29(1), 39-50.

[44] Rao, M. S., Gupta, R., Liguori, M. J., Hu, M., Huang, X., Mantena, S. R., ... & Van Vleet, T. R. (2019). Novel computational approach to predict off-target interactions for small molecules. Frontiers in big data, 2, 25.

[45] Naga, D., Muster, W., Musvasva, E., & Ecker, G. F. (2022). Off-targetP ML: an open source machine learning framework for off-target panel safety assessment of small molecules. Journal of Cheminformatics, 14(1), 27.

[46] Velez Rueda, A. J., Palopoli, N., Zacarías, M., Sommese, L. M., & Parisi, G. (2019). ProtMiscuity: a database of promiscuous proteins. Database, 2019, baz103.

[47] Liu, T., & Altman, R. B. (2015). Relating essential proteins to drug side-effects using canonical component analysis: a structure-based approach. Journal of chemical information and modeling, 55(7), 1483-1494.

[48] Wang, R., Fang, X., Lu, Y., Yang, C. Y., & Wang, S. (2005). The PDBbind database: methodologies and updates. Journal of medicinal chemistry, 48(12), 4111-4119.

[49] Huang, L. H., He, Q. S., Liu, K., Cheng, J., Zhong, M. D., Chen, L. S., ... & Ji, Z. L. (2018). ADReCS-Target: target profiles for aiding drug safety research and application. Nucleic acids research, 46(D1), D911-D917.

[50] Soldatos, T. G., Taglang, G., & Jackson, D. B. (2018). In silico profiling of clinical phenotypes for human targets using adverse event data. High-throughput, 7(4), 37.

[51] Iwata, H., Mizutani, S., Tabei, Y., Kotera, M., Goto, S., & Yamanishi, Y. (2013). Inferring protein domains associated with drug side effects based on drug-target interaction network. BMC systems biology, 7(Suppl 6), S18.

[52] Hindle, S. J., Munji, R. N., Dolghih, E., Gaskins, G., Orng, S., Ishimoto, H., ... & Bainton, R. J. (2017). Evolutionarily conserved roles for blood-brain barrier xenobiotic transporters in endogenous steroid partitioning and behavior. Cell reports, 21(5), 1304-1316.

[53] Galletti, C., Bota, P. M., Oliva, B., & Fernandez-Fuentes, N. (2021). Mining drug–target and drug–adverse drug reaction databases to identify target–adverse drug reaction relationships. Database, 2021, baab068.

[54] Mizutani, S., Pauwels, E., Stoven, V., Goto, S., & Yamanishi, Y. (2012). Relating drug–protein interaction network with drug side effects. Bioinformatics, 28(18), i522-i528.

[55] Kuhn, M., Al Banchaabouchi, M., Campillos, M., Jensen, L. J., Gross, C., Gavin, A. C., & Bork, P. (2013). Systematic identification of proteins that elicit drug side effects. Molecular systems biology, 9(1), 663.

[56] Pérez-Nueno, V. I., Souchet, M., Karaboga, A. S., & Ritchie, D. W. (2015). GESSE: predicting drug side effects from drug–target relationships. Journal of Chemical Information and Modeling, 55(9), 1804-1823.

[57] Berg, E. L. (2019). Human cell-based in vitro phenotypic profiling for drug safety-related attrition. Frontiers in Big Data, 2, 47.

[58] Liu, J., Gui, Y., Rao, J., Sun, J., Wang, G., Ren, Q., Qu, N., Niu, B., Chen, Z., Sheng, X., *et al.* (2024). In silico off-target profiling for enhanced drug safety assessment. *Acta Pharmaceutica Sinica B*, 14(7), 2927–2941.

[59] Brial, F., Puel, G., Gonzalez, L., Russick, J., Auld, D., Lathrop, M., ... & Gauguier, D. (2024). Stimulation of insulin secretion induced by low 4-cresol dose involves the RPS6KA3 signalling pathway. PloS one, 19(10), e0310370.

[60] Aluja, D., Delgado-Tomás, S., Barrabés, J. A., Miró-Casas, E., Ruiz-Meana, M., Rodríguez-Sinovas, A., ... & Inserte, J. (2024). Efficacy of a cysteine protease inhibitor compared with enalapril in murine heart failure models. Iscience, 27(10).

[61] Evans, R., Bolduc, P. N., Pfaffenbach, M., Gao, F., May-Dracka, T., Fang, T., ... & Peterson, E. A. (2024). The Discovery of 7-Isopropoxy-2-(1-methyl-2-oxabicyclo [2.1. 1] hexan-4-yl)-N-(6-methylpyrazolo [1, 5-a] pyrimidin-3-yl) imidazo [1, 2-a] pyrimidine-6-carboxamide (BIO-7488), a Potent, Selective, and CNS-Penetrant IRAK4 Inhibitor for the Treatment of Ischemic Stroke. Journal of Medicinal Chemistry, 67(6), 4676-4690.

[62] Baudry, M., Wang, Y., Bi, X., Luo, Y. L., Wang, Z., Kamal, Z., ... & Coulter, G. (2024). Identification and neuroprotective properties of NA-184, a calpain-2 inhibitor. Pharmacology Research & Perspectives, 12(2), e1181.

[63] Burnett, G. L., Yang, Y. C., Aggen, J. B., Pitzen, J., Gliedt, M. K., Semko, C. M., ... & Gill, A. L. (2022). Discovery of RMC-5552, a selective bi-steric inhibitor of mTORC1, for the treatment of mTORC1-activated tumors. Journal of medicinal chemistry, 66(1), 149-169.

[64] Santra, P., Ghosh, M., Ganguly, D., Basuchowdhuri, P., & Naskar, S. K. (2025). HF-RAG: Hierarchical Fusion-based RAG with Multiple Sources and Rankers. In Proceedings of the 34th ACM International Conference on Information and Knowledge Management (pp. 5202–5207). ACM.

[65] Cormack, G. V., Clarke, C. L., & Buettcher, S. (2009, July). Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval (pp. 758-759).

[66] Samuel, S., DeGenaro, D., Guallar-Blasco, J., Sanders, K., Eisape, S., Reddy, A., ... & Kriz, R. (2025, July). Mmmorrf: Multimodal multilingual modularized reciprocal rank fusion. In Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 4004-4009).

[67] Bruch, S., Gai, S., & Ingber, A. (2023). An analysis of fusion functions for hybrid retrieval. ACM Transactions on Information Systems, 42(1), 1-35.

[68] Sun, J., Jeliazkova, N., Chupakhin, V., Golib-Dzib, J. F., Engkvist, O., Carlsson, L., ... & Chen, H. (2017). ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics. Journal of cheminformatics, 9(1), 17.

[69] Dimova, D., Stumpfe, D., Hu, Y., & Bajorath, J. (2016). Analog series-based scaffolds: computational design and exploration of a new type of molecular scaffolds for medicinal chemistry. Future Science OA, 2(4), FSO149.

[70] Benham, R., & Culpepper, J. S. (2017). Risk-Reward Trade-offs in Rank Fusion. In Proceedings of the 22nd Australasian Document Computing Symposium (ADCS '17). ACM. https://doi.org/10.1145/3166072.3166084

[71] Bresalier, R. S., Sandler, R. S., Quan, H., Bolognese, J. A., Oxenius, B., Horgan, K., Lines, C., Riddell, R., Morton, D., Lanas, A., Konstam, M. A., & Baron, J. A.; Adenomatous Polyp Prevention on Vioxx (APPROVe) Trial Investigators. (2005). Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. New England Journal of Medicine, 352(11), 1092–1102.

[72] Taggart, H. M., & Alderdice, J. M. (1982). Fatal cholestatic jaundice in elderly patients taking benoxaprofen. BMJ (Clinical Research Edition), 284(6326), 1372–1374.

[73] Zanettini, R., Antonini, A., Gatto, G., Gentile, R., Tesei, S., & Pezzoli, G. (2007). Valvular heart disease and the use of dopamine agonists for Parkinson's disease. New England Journal of Medicine, 356(1), 39–46.

[74] Monahan, B. P., Ferguson, C. L., Killeavy, E. S., Lloyd, B. K., Troy, J., & Cantilena, L. R., Jr. (1990). Torsades de pointes occurring in association with terfenadine use. JAMA, 264(21), 2788–2790.

[75] Staffa, J. A., Chang, J., & Green, L. (2002). Cerivastatin and reports of fatal rhabdomyolysis. New England Journal of Medicine, 346(7), 539–540.

[76] Scheen, A. J. (2001). Thiazolidinediones and liver toxicity. *Drug Safety*, 24(11), 873–888.

[77] Chávez-Tapia, N. C., Hernández-Calleros, J., Tellez-Avila, F. I., et al. (2009). Systematic review and meta-analysis on the adverse events of rimonabant treatment: considerations for its potential use in hepatology. BMC Gastroenterology, 9, 75