IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Open Targets Association Score: Evidence & Sensitivity Guide

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-11-09 · open targets association score · open targets platform · harmonic sum · datasource weights · direct vs indirect associations · target prioritisation · target validation · score sensitivity analysis · open targets 26.09

Original article: https://intuitionlabs.ai/articles/open-targets-association-scores-evidence-sensitivity



In this article

1. Executive Summary
2. Introduction and Background
3. The Score Hierarchy and Direct Versus Indirect Associations
4. Source Weights and a Transparent Harmonic-Sum Calculation
5. Release Provenance and the 26.09 Example
6. Sensitivity Review: Leave-One-Source-Out and Alternative Weights
7. Evidence-to-Experiment Matrix
8. Versioned Export and Scientific Sign-Off Checklist
9. Reported Scale Across Releases
10. Implications and Future Directions
11. Frequently Asked Questions (FAQs)
12. Conclusion

Executive Summary

An **Open Targets association score** is a number between 0 and 1 that ranks how much aggregated evidence links one target (a gene or protein) to one disease or phenotype in the Open Targets Platform. The project's documentation describes the scores as "a heuristic based on the availability of data" (platform-docs.opentargets.org) and states that they "should not be interpreted as a confidence score for the target-disease association" (platform-docs.opentargets.org). For a team deciding whether to fund validation experiments, that sentence frames this report: the score supports ranking and triage, not a statement of probability, causality or safety.

The calculation has four levels. Evidence items carry source-specific scores, a **harmonic sum** (each sorted score divided by its squared rank) aggregates them per data source, and the result is divided by a maximum theoretical sum that the Platform approximates as **1.644** using a vector of 1,000 ones (community.opentargets.org). The overall score is a weighted harmonic sum of data source scores (pmc.ncbi.nlm.nih.gov). Default weights are **0.2** for Europe PMC, Expression Atlas and IMPC, **0.5** for Cancer Biomarkers and OTAR Projects, and **1** for other sources (platform-docs.opentargets.org). The documentation example of evidence scored 1, 0.9 and 0.8 gives 1 + 0.9/4 + 0.8/9 = 1.3139, and dividing by the 1,000-term normaliser (1.64393) yields **0.799** (computed for this report; a documentation example, not a real target score).

Direct associations use evidence annotated to the exact disease term, while indirect associations add evidence annotated to ontology descendants (pmc.ncbi.nlm.nih.gov). Rankings therefore depend on the release, the ontology and the weights. As of November 2026 the API landing page reports data version 26.09 (api.platform.opentargets.org) and API version 26.9.0 (api.platform.opentargets.org). Release **26.09** added **1.24 million** molecular QTL (quantitative trait locus) credible sets (blog.opentargets.org) and a de-duplication step that coalesces **1,706** duplicate disease terms (blog.opentargets.org). These are provider-reported figures, not independently counted, and rank movements they cause reflect changed data and ontology, not improved biological validity.

Genetic support is informative at portfolio level. One analysis estimates a **2.6 times** higher probability of clinical success for drug mechanisms with genetic support (www.nature.com), yet the same paper says its metric does not address whether a particular disease-associated gene will yield a successful drug (www.nature.com). This report therefore sets out a sensitivity protocol (leave-one-source-out, alternative weights, direct versus indirect mode), an evidence-to-experiment matrix and a versioned-export checklist. This article does not report a measured target shortlist, rank change or sensitivity benchmark: the sensitivity table uses clearly labelled synthetic inputs, and the code and documentation cited are described as published, not re-run.

Introduction and Background

Target-validation programs usually start from a ranked list. Someone queries the Open Targets Platform for a disease, exports the top targets, and a committee decides which few deserve laboratory work. The ranking behind that decision comes from a scoring framework with several adjustable and release-dependent parts. This report treats the association score as a measurement instrument and documents what it measures, how it is built, how sensitive a shortlist is to its parameters, and what provenance a team needs to repeat a ranking later.

The 2017 architecture paper describes an association score from 0 to 1 derived for each piece of evidence (pmc.ncbi.nlm.nih.gov), and the 2019 paper describes the framework as evidence, data source, data type and overall levels, with harmonic sums used for aggregation (pmc.ncbi.nlm.nih.gov). The 2021 paper adds that the overall score is a weighted harmonic sum (pmc.ncbi.nlm.nih.gov).

The scope is deliberately narrow. This is not a platform overview or a review of AI target discovery. It answers the questions a discovery scientist or data engineer is likely to ask: how scores are calculated, what the harmonic sum does, what the weights are, how direct and indirect associations differ, what changed in recent releases, and how to test whether a shortlist survives a change in weighting. The report is written as a durable reference rather than a trend piece. Disclosure: IntuitionLabs, the publisher, is an advisory firm and not a vendor of any scoring platform.

Two cautions apply. Documentation and source-code comments are cited as published, and differences between them are flagged for verification against a same-release export. Where this report computes numbers, it labels them as computed from the documented formula, separate from provider-reported figures.

The Score Hierarchy and Direct Versus Indirect Associations

Open Targets defines evidence as events pinpointing a target as a potential causal gene or protein for a disease, with all evidence mapped to an Ensembl gene identifier (platform-docs.opentargets.org). Individual resources are data sources, and these are grouped into larger categories called data types. Each data source has its own scoring logic, so an evidence score of 0.5 from two sources can mean very different things: a GWAS (genome-wide association study) record is scored from a machine-learning gene prioritisation whose predictions below 0.05 are filtered out (platform-docs.opentargets.org), while a literature record is scored from text co-occurrence statistics normalised between 0 and 1 (platform-docs.opentargets.org). *Table 1* below summarises the four score levels and the inferences each does and does not support.

Score level	Input	Normaliser	Weights	Interpretation	Unsupported inference
Evidence	One record from a source, mapped to an Ensembl gene and an EFO (Experimental Factor Ontology) term	None; each source defines its own 0 to 1 scoring	None	Strength of one record on that source's own scale	Comparing raw evidence scores across sources as if on one scale
Data source	Sorted vector of evidence scores for one source and pair	Maximum theoretical harmonic sum, approximately 1.644 (community.opentargets.org)	None applied at this level	Consensus view of one source's evidence	Treating a high score as replication across independent studies
Data type	Vector of data source scores in the type, for example genetic association	Vector of as many ones as data sources in the type, so single-source types are not penalised (platform-docs.opentargets.org)	Documentation says weighted by data source weights; see the code note below	Strength within a category of evidence	Reading a type score as independent of how many sources the type contains
Overall	Data source scores, regardless of data type	Fixed constant of about 1.644 (community.opentargets.org)	Data source weights	Ranking aid for targets or diseases	Probability of approval, confidence or safety

Two features of the normalisation deserve emphasis. At the data type level, a type with one source is scaled by 1.0/1^2, which equals 1, while a type with three sources is scaled by 1 + 1/4 + 1/9, which the documentation gives as **1.36**. At the overall level, Open Targets staff say the normalising value is "not dynamically updated with the weights" (community.opentargets.org). The consequence, derived here from the documented formula, is that a target whose only evidence is one data source at weight 1 and score 1.0 receives an overall score of about 1/1.644 = **0.61**, and a source at weight 0.2 can contribute at most about 0.12 on its own. Overall scores near 1 require several strong sources; that is a property of the formula, not a statement about biology.

Evidence granularity also differs by source, which matters because the harmonic sum rewards repetition with sharply diminishing returns. Some sources score one record per underlying object, such as a ClinVar record, while others aggregate before scoring, such as a single UniProt item that combines all supporting publications. A reviewer who compares evidence counts across sources is comparing unlike units.

Direct Versus Indirect Disease Associations

The Platform separates **direct associations** (evidence annotated to the two specific terms) from **indirect associations**, which add evidence annotated to the disease's ontology descendants. The documentation example is NOD2 and inflammatory bowel disease, where the indirect association selects evidence for the disease "or any of its ontology descendants" including Crohn's disease (platform-docs.opentargets.org). The 2021 paper describes the mechanism as propagating evidence from child nodes to parent nodes (pmc.ncbi.nlm.nih.gov), and the 2017 paper notes that it derives new associations that may have no direct evidence (pmc.ncbi.nlm.nih.gov).

Where each mode appears matters for reproducibility. According to the documentation, the page listing diseases for a target shows direct evidence only, whereas the page listing targets for a disease includes both direct and indirect evidence, and the same calculations apply with only the included evidence differing (platform-docs.opentargets.org). A shortlist exported from the disease view is therefore by default an indirect ranking, and a table assembled from target pages is direct. Mixing the two silently is a common source of non-reproducible comparisons.

One data type is excluded from propagation by design: RNA expression evidence is not propagated in the ontology (platform-docs.opentargets.org). The pipeline source repository (main branch) explodes each evidence record to the disease and its ancestors (raw.githubusercontent.com), propagates a datasource by default when it is not listed in configuration (raw.githubusercontent.com), and shows one `toPropagate: false` setting (raw.githubusercontent.com). The API exposes the same concept as a per-datasource `propagate` setting (api.platform.opentargets.org) and an `enableIndirect` argument that uses the disease ontology to retrieve descendants (api.platform.opentargets.org).

Propagation also affects sensitivity. A broad parent term accumulates evidence from many descendants, so its scores reflect how many descendant diseases carry evidence as well as the strength of any one. The 2021 paper reports that adopting a new EFO version in release 19.11 increased the number of associations by **90%** (pmc.ncbi.nlm.nih.gov), which shows how far counts can move with ontology structure alone. Comparisons across releases should record the EFO version in use.

Source Weights and a Transparent Harmonic-Sum Calculation

Documented default weights

The documentation says evidence is weighted by a factor intended to calibrate the relevance of each data source relative to others (platform-docs.opentargets.org). The defaults are 0.2 for Europe PMC, Expression Atlas and IMPC (International Mouse Phenotyping Consortium), 0.5 for OTAR Projects (partner preview only) and Cancer Biomarkers, and 1 for all others. The 2017 paper gives the general rationale: GWAS association scores are upweighted and text mining scores downweighted (pmc.ncbi.nlm.nih.gov). The 2025 paper adds that, in the interactive view, evidence with expected weaker causal links is down-weighted by default (pmc.ncbi.nlm.nih.gov). The pipeline configuration lists Europe PMC at the same 0.2 default (raw.githubusercontent.com); it also assigns 0.5 to three further sources (encore, ot_crispr and ot_crispr_validation), which this report reads as corresponding to the "OTAR Projects" row but could not confirm.

Users can change the weights. The Associations on the Fly page offers user control over the weighting of each data source (platform-docs.opentargets.org), the 2025 paper confirms the same (pmc.ncbi.nlm.nih.gov), and the GraphQL schema describes weights as between 0 and 1 (api.platform.opentargets.org).

Worked calculation

The following uses the documentation's own example, which is not a real target score. For a vector of evidence scored 1, 0.9 and 0.8:

- **Step 1, sort and index:** scores are sorted in descending order and assigned positions 1, 2 and 3 (platform-docs.opentargets.org).
- **Step 2, harmonic sum:** 1.0/1^2 + 0.9/2^2 + 0.8/3^2 = 1 + 0.225 + 0.0889 = 1.3139 (computed for this report).
- **Step 3, finite normaliser:** the sum of 1/i^2 for i = 1 to 1000 is 1.643935 (computed for this report), which the Platform rounds to 1.644 (platform-docs.opentargets.org).
- **Step 4, scale:** 1.3139 / 1.6439 = **0.799**; using 1.644 gives the same value to three decimals.

The exact infinite sum is pi squared over six, approximately 1.644934 (dlmf.nist.gov), so a 1,000-term vector slightly understates it. Substituting the exact constant moves this example from 0.7992 to 0.7988, which would not change a ranking but shows why a reproduction should state its constant. Open Targets staff gave a real-data illustration for a genetics score: two L2G (locus-to-gene) values summed harmonically to 0.7647525, and dividing by 1.644 gave 0.4651779 (community.opentargets.org).

Points to verify against an export

Manual calculation is a check, not a replacement for the pipeline. Several implementation details appear in the current repository rather than the documentation, and a reproduction should confirm them against a same-release export:

- **Top-50 cap:** a code comment says the harmonic sum uses only the highest 50 scores (raw.githubusercontent.com); the documentation does not mention a cap.
- **Normaliser constant:** the same module defines the theoretical maximum of the harmonic sum for score normalisation as a class constant (raw.githubusercontent.com).
- **Data type weighting:** the documentation says data type scores use data source scores weighted by data source weights, whereas a comment in the association module says scores are not weighted by datasource weight in the data type aggregation (raw.githubusercontent.com), and that overall aggregation is weighted (raw.githubusercontent.com).
- **Unlisted sources:** documentation and staff say sources outside the weights table fall into "Others" with weight 1, while the association module joins to an explicit weight table with an inner join (raw.githubusercontent.com). Whether any source falls outside the table in a given release should be checked.
- **Repository location:** the pipeline now lives in a repository described as made up of four parts (github.com), and the older Scala ETL repository is archived (github.com), so older community answers may point to superseded code.

None of these items is a defect claim. They are reasons to validate a hand-built calculation against exported association tables before relying on it.

Release Provenance and the 26.09 Example

Scores are a function of data, ontology and pipeline version, so each ranking needs a release identifier. As of November 2026 the API landing page lists data version 26.09 (api.platform.opentargets.org) and API version 26.9.0 (api.platform.opentargets.org), and the release notes date 26.09 to 24 September 2026, though other

project pages may give a slightly different day, so cite the release as September 2026 (platform-docs.opentargets.org). The GitHub releases page for the platform repository states that there aren't any releases there (github.com), so release notes should be read from the documentation site and project blog, and a third-party copy of release notes should not be assumed current.

What changed in 26.09

The blog reports 1.24 million additional molecular QTL credible sets (blog.opentargets.org), including a single-cell eQTL (expression QTL) dataset contributing over 70,900 credible sets in 42 cell types and contexts (blog.opentargets.org). It also reports a GWAS Catalog update adding 14,690 studies and 44,859 credible sets from 143 publications (blog.opentargets.org), relaxed quality-control thresholds so exome and whole-genome studies contribute signals (blog.opentargets.org), and two new gene-burden sources contributing over 10,000 evidence records (blog.opentargets.org). For ontology, a cross-ontology de-duplication step was added (blog.opentargets.org), coalescing 1,706 duplicate terms and consolidating over 3 million evidence records into a single term (blog.opentargets.org); the release notes list obesity, hypertension and stroke as examples (platform-docs.opentargets.org). These are provider-reported figures that this report did not independently count.

The effect on a shortlist is mixed and cannot be inferred from headline numbers. New molecular QTL credible sets can feed colocalisation and L2G scoring for genetic evidence, which the Platform computes at credible-set level, while de-duplicated disease terms change which evidence attaches to which term and so both direct and indirect scores. A rank change after 26.09 shows that data or ontology changed; it does not show the new rank is more biologically valid.

Earlier releases with scoring relevance

Earlier releases show the same pattern. The 25.12 post hid measurement terms from the default target associations view (blog.opentargets.org), replaced SuSiE credible sets with PICS for multi-ancestry GWAS (blog.opentargets.org), and split the evidence dataset into per-datasource datasets (blog.opentargets.org). The 26.03 post reports removing around 300 disease-phenotype duplicates (blog.opentargets.org) and a fix to scoring differences between the API and the web interface (blog.opentargets.org), a reminder to record which interface and settings produced a table. The 26.06 post describes a mass replacement of EFO disease terms with Mondo terms (blog.opentargets.org) and a literature pipeline revision. For that revision the release notes say about 1.8 million evidence records were added and around 600,000 direct associations removed (platform-docs.opentargets.org), while the blog reports that 60,000 direct associations are no longer supported by literature evidence (blog.opentargets.org). The two project pages differ by a factor of ten, and the pages do not say whether they count the same thing (for example gross versus net removals). Neither figure should be used as an effect size; analyses depending on that change should recompute it from exports. The L2G documentation also advises expecting some variation in prediction scores with each release.

Ontology provenance

EFO describes itself as the core ontology for Open Targets (www.ebi.ac.uk). Its repository states that existing EFO terms kept their EFO identifiers (github.com), that new terms imported from MONDO kept their MONDO identifiers (github.com), that deprecated terms carry an owl:Deprecated annotation (github.com), and that the

MONDO import is pinned per update, with a recent commit pinning the mirror to a 2026-11-03 release (github.com). The latest EFO version is attached to each release on its repository page (github.com). A versioned analysis should store the EFO release alongside the Platform release, and store disease identifiers rather than labels.

Sensitivity Review: Leave-One-Source-Out and Alternative Weights

The sources reviewed for this article include no peer-reviewed or preprint analysis of the ranking robustness of Open Targets association scores. The 2021 paper states that the lack of appropriate gold standards across therapeutic areas limits the effectiveness of benchmarks (pmc.ncbi.nlm.nih.gov) and identifies the need to appropriately weight and benchmark scored evidence (pmc.ncbi.nlm.nih.gov). A team therefore has to design its own checks. This report proposes four perturbations on a fixed release, applied to exported data source scores:

- **Leave-one-source-out:** recompute the overall score after removing each data source in turn, and record the rank of every shortlisted pair.
- **Alternative weights:** set all weights to 1, set literature to 0, and up-weight the sources the program trusts most, staying within the API's 0 to 1 range (api.platform.opentargets.org).
- **Direct versus indirect mode:** repeat the ranking with direct-only evidence and with propagated evidence.
- **Release comparison:** repeat the baseline on the next release, recording the EFO version, and separate rank movements caused by identifier changes from those caused by new evidence.

For each pair, record: target and disease identifiers, fixed release, direct or indirect mode, number of contributing sources, top contributing source, baseline rank, and rank under each declared perturbation. *Table 2* shows the format with entirely synthetic inputs. It is a **(Hypothetical Example)**: the three targets and their data source scores are invented, the calculation applies the documented overall formula (weighted scores sorted descending, divided by squared position, divided by 1.6439), and no real target was measured.

Scenario (synthetic data)	T1 score (rank)	T2 score (rank)	T3 score (rank)	Reading
Baseline, documented default weights	0.157 (3)	0.331 (2)	0.665 (1)	T3 leads because one source scores 1.0
All weights set to 1	0.612 (2)	0.347 (3)	0.718 (1)	T1 and T2 swap: literature-heavy evidence rises when its discount is removed
Drop the single pathway source	0.157 (3)	0.331 (1)	0.192 (2)	T3 falls from first to second: its lead rested on one source
Drop the GWAS source	0.121 (3)	0.222 (2)	0.648 (1)	Order unchanged, but T2 loses a third of its score
Europe PMC weight set to 0	0.145 (3)	0.327 (2)	0.663 (1)	Order unchanged: literature contributes little at default weight

Synthetic inputs: T1 has Europe PMC 0.80, Expression Atlas 0.55, IMPC 0.50 and GWAS 0.20; T2 has Europe PMC 0.30, GWAS 0.45 and ClinVar 0.35; T3 has Europe PMC 0.45, IMPC 0.40, GWAS 0.25, ChEMBL 0.20 and a pathway source 1.0.

The interpretation is methodological, and three reading rules follow from it:

- **Single-source dependence:** if a rank changes when one source is removed, record that source as the decisive evidence and open its records first.
- **Weight-sensitive pairs:** pairs that swap when documented discounts are lifted need a written reason why the discounted source is credible or not.
- **Stable pairs:** pairs that hold position across every perturbation are the candidates for funding, subject to the separate checks in the next section.

In this invented case the headline target is fragile to one source, and two lower-ranked targets exchange places when a documented discount is lifted. Neither result says which target is biologically better; they show which conclusions depend on a modelling choice. A shortlist whose top candidates hold their positions across all perturbations, with several independent sources each, is more defensible for funding than one whose lead depends on a single record. Because the normaliser is not rescaled when weights change (community.opentargets.org), down-weighting a source also lowers absolute scores, so absolute thresholds should not be carried between weighting schemes.

Evidence-to-Experiment Matrix

Several decision-relevant questions sit outside the association score. The Target Prioritisation view groups target properties into four sections (platform-docs.opentargets.org), and the 2025 paper notes that none of those factors alone would discard a therapeutic hypothesis (pmc.ncbi.nlm.nih.gov). *Table 3* maps each evidence family to what it can support, what it cannot establish, and which experiment type usually addresses the gap. The experiment column is analyst guidance for discussion by the program's scientists, not a protocol and not an Open Targets requirement.

Evidence family	What the source score captures	What it does not establish	Typical follow-up experiment	Record before sign-off
Genetic association (GWAS credible sets, burden, ClinVar)	Gene prioritisation at credible sets; L2G predictions below 0.05 are filtered out	Direction of effect, tissue, ancestry transfer; 86.3% of genomics studies were in individuals of European descent as of June 2021 (pmc.ncbi.nlm.nih.gov)	Variant-to-function assays; knock-down or knock-in in disease-relevant cells	Study and credible set IDs, L2G, ancestry, effect direction
Literature (Europe PMC)	Text co-occurrence, normalised 0 to 1	Causality; research attention reflects generic gene properties rather than disease relevance (pmc.ncbi.nlm.nih.gov)	Manual curation of primary papers; orthogonal perturbation data	PMIDs read, publication types, independence of papers

Evidence family	What the source score captures	What it does not establish	Typical follow-up experiment	Record before sign-off
Pathways and curated resources	Manually curated items, for example in Reactome	Disease-specific causality or effect size	Pathway perturbation and rescue experiments	Reaction IDs, curator source, whether one record drives the score
Animal models (IMPC)	Mouse-human phenotype concordance	Human relevance, dosage, compensation	Conditional or humanised model; cross-species cell assays	Allele, phenotype terms, concordance measure
Clinical precedence	Clinical stage of drugs acting on the target	Efficacy in a new indication	Review clinical reports; indication-similarity analysis	Drug, phase, stop reason, indication match
RNA expression	Differential expression, not propagated in the ontology	Causality or druggability	Perturbation and cell-type-resolved expression	Contrast IDs, tissue, direction of change
Tractability, safety, mechanism (outside the score)	Separate Platform sections	Anything about association strength	Modality feasibility, target-safety review, selectivity panels	Modality, known liabilities, mechanism, direction of effect

The matrix should be read with portfolio-level genetics literature. An analysis of 29,476 target-indication pairs (www.nature.com) found genetic support for 2,166 (www.nature.com) and estimated 2.6 times higher probability of success for supported mechanisms (www.nature.com). The same work reports that the estimate improves with confidence in the causal gene but is largely unaffected by effect size, allele frequency or year of discovery (www.nature.com), that its analyses are naive to direction of genetic effect (www.nature.com), and that GWAS support is somewhat less predictive than Mendelian support (www.nature.com). An earlier analysis reported the share of mechanisms with direct genetic support rising from 2.0% at the preclinical stage to 8.2% among approved drugs (www.nature.com), and a replication found an approval increase above two-fold only when causal genes are clear (journals.plos.org). These are population-level findings; none converts one overall score into a probability for a specific program.

Versioned Export and Scientific Sign-Off Checklist

A reproducible ranking is a dataset plus a log. The release-triggered process below keeps source and ontology versions pinned and records every table update.

Data access choices

Platform datasets are distributed as Parquet only (platform-docs.opentargets.org), and download paths changed after release 25.03. The **26.09** FTP output directory lists association datasets for data source, data type and overall levels, each in direct and indirect form (ftp.ebi.ac.uk). Names have moved over time: the 25.03 directory used a differently named indirect overall dataset (ftp.ebi.ac.uk), so scripts should list the

directory rather than hard-code a name. The BigQuery public dataset exposes the direct overall table as `open-targets-prod.platform.association_overall_direct` (platform-docs.opentargets.org), with the first 1 TB of queries per month free. Open Targets staff report that **all previous releases** remain available on the FTP server and in a Google Cloud bucket (community.opentargets.org), with a complete collection reported to start at the earliest releases (community.opentargets.org), but no formal retention policy was found, so teams should archive their own copy of every release used for a decision. The overall dataset schema describes the score as the current disease-target association score of evidence from any datasource (raw.githubusercontent.com), and the release notes record a schema that embeds yearly scores, so field lists in older scripts should be re-checked each release.

Checklist

- **Pin the release:** record the Platform data version (26.09 as of November 2026), API version, and download location and date.
- **Pin the ontology:** record the EFO version and store disease identifiers, not labels.
- **Pin the mode:** state direct or indirect for every table, and the interface or API settings used.
- **Pin the weights:** store the full weight table, including any change from documented defaults.
- **Reproduce one score:** recalculate at least one pair from exported evidence, note which normaliser was used, and compare with the export.
- **Run the four perturbations:** leave-one-source-out, alternative weights, direct versus indirect, next-release comparison.
- **Open the evidence records:** retain source IDs for every record behind a shortlisted pair.
- **Assess separately:** independence of evidence, ancestry, direction of effect, tractability and safety, none of which the overall score encodes.
- **Write the update log:** on each release, list data, ontology and schema changes that could move ranks.
- **Sign off:** a named scientist records the decision and the limits of inference.

Reported Scale Across Releases

The quantitative record below is drawn from the project's papers and release notes. Each figure is provider-reported, and the counts come from different releases, so they cannot be read as a trend.

- **2017 paper, release 1.2:** associations covered 30,591 targets and 9,425 diseases and phenotypes (pmc.ncbi.nlm.nih.gov).
- **2019 paper:** almost three billion associations between 21,149 human genes and 10,101 diseases and phenotypes (pmc.ncbi.nlm.nih.gov), from over six billion evidence in 18 public data sources (pmc.ncbi.nlm.nih.gov).
- **2021 paper:** 10,154,924 pieces of evidence mapped and curated (pmc.ncbi.nlm.nih.gov), with an overall score for 7,282,832 target-disease associations (pmc.ncbi.nlm.nih.gov).
- **2025 interactive-view paper:** 7.8 million target-disease associations, a figure the paper does not tie explicitly to a release (its introduction cites release 24.12 for different counts), so it is not comparable with the release-specific counts below (www.ncbi.nlm.nih.gov).
- **2025 database paper:** evidence curated in 23 independent public sources (pmc.ncbi.nlm.nih.gov).

- **Release 26.06:** 42,394,639 evidence strings and 17,199,165 target-disease associations (platform-docs.opentargets.org).

Two observations follow. The papers do not state a common counting definition (the 2019 count is almost three billion associations, against 17.2 million in the 26.06 notes), so any trend claim needs same-definition counts computed from exports. And the number of data sources has moved from 18 in 2019 to 23 in the 2025 paper, so source composition changes between releases.

External benchmarks give context for how much weight genetic evidence can carry. In the 2024 analysis, only 1.1% of genetically supported gene-indication relationships had been explored clinically (www.nature.com), and the type 2 diabetes example shows that associations separating clinical-stage targets from launched ones do not mean a large fraction of associations will be fruitful (www.nature.com). For Open Targets Genetics associations, that study found the estimate sensitive to confidence in variant-to-gene mapping (www.nature.com), which is the quantity L2G estimates. A replication found that targets with a genetic association to a dissimilar indication were less likely to be approved than pairs with no known association (journals.plos.org). On ancestry, polygenic risk scores were reported to be many-fold more accurate in individuals of European ancestry than in other ancestries (pmc.ncbi.nlm.nih.gov). On research bias, most human gene research concentrates on approximately 2,000 of the 19,000 genes (pmc.ncbi.nlm.nih.gov), which bears directly on the literature data type. The broader framework for assessing targets, such as the five-R scheme (right target, patient, tissue, safety and commercial potential) (pubmed.ncbi.nlm.nih.gov), goes beyond any association score, and regulators are reported to be increasingly requesting data on activity against targets with recognised adverse effect relationships (www.nature.com).

Implications and Future Directions

For discovery leaders, the main implication is procedural. A shortlist should be approved only with its provenance record: release, ontology version, mode, weights and a perturbation table. Funding a validation experiment on an overall score alone invites the misreading the documentation warns against. The documentation also notes that under-studied diseases are unlikely to produce high-scoring targets because evidence is scarce (platform-docs.opentargets.org), and that some sources rely on predictions and should be considered with caution, so a low score is not evidence of low promise and a high score is not evidence of validity.

For scientific data engineers, scoring is an upstream dependency that changes. Release notes record schema changes, dataset splits and renamed fields, and release-over-release comparison needs identifier mapping as well as score comparison. Automating the checklist (a job triggered by a new release that downloads pinned datasets, recomputes the perturbation table and writes the update log) is a modest engineering task, though the code-versus-documentation differences listed above mean validation against a same-release export must be part of it. Implementation details depend on each organisation's data stack and quality system, and the program's own scientists and statisticians should review any ranking before it is used to commit funds.

Open questions remain. No published robustness study of Open Targets scores was found, the weights have not been benchmarked against a gold standard (the 2021 paper notes none exists across therapeutic areas), and the interaction between ontology de-duplication and indirect scores has not been measured here. The

project's 2023 paper describes the Platform as released five times a year (pmc.ncbi.nlm.nih.gov), so a refresh process should be designed to run repeatedly. A public benchmark of rank stability across releases would be useful, but none exists in the sources reviewed.

Frequently Asked Questions (FAQs)

What is an Open Targets platform association score?

It is a 0 to 1 value summarising aggregated evidence for one target-disease pair, described by the documentation as a heuristic useful for ranking rather than a confidence score. Scores exist at the data source, data type and overall levels.

How are Open Targets association scores calculated?

Evidence is sorted, divided by squared rank, summed and normalised by a maximum theoretical sum of approximately 1.644. The documentation example for evidence scored 1, 0.9 and 0.8 gives about 0.799 (computed here).

What is the harmonic sum behind the overall score?

It is the same aggregation applied to data source scores weighted by data source weights, regardless of data type; the 2021 paper calls it a weighted harmonic sum.

What are the data source weights?

Defaults are 0.2 for Europe PMC, Expression Atlas and IMPC, 0.5 for OTAR Projects and Cancer Biomarkers, and 1 for others; users can change them in the API and the web interface.

What changed in recent release notes that could affect scores?

Release 26.09 added molecular QTL credible sets and de-duplicated disease terms (see the 26.09 section above), and 26.06 replaced many EFO disease identifiers with Mondo identifiers. The release notes do not tie these changes to a rank change for any target.

How are direct and indirect associations different?

Direct associations use evidence annotated to the two specific terms, while indirect ones add evidence from ontology descendants.

What evidence data sources does the platform use?

The 2025 paper describes curation from 23 independent public sources, grouped into data types, with source-specific scoring.

How should teams use Open Targets for target prioritisation in drug discovery?

Combine the association score with separate tractability, safety and precedence views, and apply the sensitivity protocol above.

How do you run a score sensitivity analysis?

Use a fixed release and vary one thing at a time: drop each source, change weights, switch direct and indirect mode, and compare releases. The synthetic Table 2 shows the log format.

Conclusion

The Open Targets association score is a documented, adjustable ranking heuristic. Its construction is simple enough to reproduce by hand for a documentation example, which gives about 0.799 for evidence scored 1, 0.9 and 0.8, and complex enough that a real reproduction needs a pinned release, a pinned ontology, a declared direct or indirect mode, and the exact weight table. The project itself states the limit: it is not a confidence score.

For a funding decision, stability is the more useful question than magnitude. Does the target hold its position when each source is removed, when literature is discounted, when propagation is switched off, and when the next release arrives? Does the evidence come from several independent sources whose records have been opened, and have direction of effect, ancestry, tractability and safety been assessed outside the score? Where the answers are yes, a validation experiment is easier to justify; where a rank depends on one record or one weighting choice, that is a prompt for more evidence before spending. This article does not report a measured target shortlist, rank change or sensitivity benchmark, so the sensitivity table is a template filled with synthetic numbers, and each team should generate its own from same-release exports.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://platform-docs.opentargets.org/associations
2. https://community.opentargets.org/t/how-is-the-ot-genetics-score-listed-in-the-ot-platform-calculated/1399
3. https://pmc.ncbi.nlm.nih.gov/articles/PMC7779013
4. https://api.platform.opentargets.org
5. https://blog.opentargets.org/open-targets-platform-26-09-has-been-released
6. https://www.nature.com/articles/s41586-024-07316-0
7. https://pmc.ncbi.nlm.nih.gov/articles/PMC5210543
8. https://pmc.ncbi.nlm.nih.gov/articles/PMC6324073
9. https://platform-docs.opentargets.org/evidence
10. https://platform-docs.opentargets.org/gentropy/locus-to-gene-l2g
11. https://community.opentargets.org/t/how-to-calculate-the-overall-association-score/1260
12. https://community.opentargets.org/t/overall-association-score-with-scored-weights-of-0/1577
13. https://raw.githubusercontent.com/opentargets/pipeline/main/pts/src/pts/pyspark/associations_utils/evidence.py
14. https://raw.githubusercontent.com/opentargets/pipeline/main/pts/config.yaml
15. https://api.platform.opentargets.org/api/v4/graphql/schema
16. https://pmc.ncbi.nlm.nih.gov/articles/PMC11701534

17. https://platform-docs.opentargets.org/web-interface/associations-on-the-fly
18. https://dlmf.nist.gov/25.6
19. https://raw.githubusercontent.com/opentargets/pipeline/main/pts/src/pts/pyspark/associations_utils/dataset.py
20. https://raw.githubusercontent.com/opentargets/pipeline/main/pts/src/pts/pyspark/associations_utils/association.py
21. https://github.com/opentargets/pipeline
22. https://github.com/opentargets/platform-etl-backend
23. https://platform-docs.opentargets.org/release-notes
24. https://github.com/opentargets/platform/releases
25. https://blog.opentargets.org/open-targets-platform-25-12-has-been-released
26. https://blog.opentargets.org/open-targets-platform-26-03-has-been-released
27. https://blog.opentargets.org/open-targets-platform-26-06-has-been-released
28. https://www.ebi.ac.uk/ols4/ontologies/efo
29. https://github.com/EBISPOT/efo
30. https://platform-docs.opentargets.org/web-interface/target-prioritisation
31. https://pmc.ncbi.nlm.nih.gov/articles/PMC7614889
32. https://pmc.ncbi.nlm.nih.gov/articles/6143198
33. https://www.nature.com/articles/ng.3314
34. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008489
35. https://platform-docs.opentargets.org/data-access/datasets
36. https://ftp.ebi.ac.uk/pub/databases/opentargets/platform/26.09/output
37. https://ftp.ebi.ac.uk/pub/databases/opentargets/platform/25.03/output
38. https://platform-docs.opentargets.org/data-access/google-bigquery
39. https://community.opentargets.org/t/querying-older-open-targets-platform-versions-need-data-from-24-09-for-neurodegenerative-disease/1936
40. https://community.opentargets.org/t/older-versions-of-opentargets-data/1541
41. https://raw.githubusercontent.com/opentargets/pipeline/main/croissant/src/ot_croissant/assets/recordset/association_overall_direct.json
42. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11968318
43. https://pmc.ncbi.nlm.nih.gov/articles/PMC6563838
44. https://pubmed.ncbi.nlm.nih.gov/24833294
45. https://www.nature.com/articles/s41573-024-00942-3
46. https://pmc.ncbi.nlm.nih.gov/articles/PMC9825572

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.