**Summary of OpenTargets scoring system for genetic associations**

**Important Note from Open Targets:**

**Association scores are a heuristic based on the availability of data.** While scores are useful to rank lists of targets or diseases, they should not be interpreted as a confidence score for the target-disease association. **Under-studied diseases are unlikely to produce high-scoring targets due to the lack of available evidence.** In such diseases, a relatively low-scoring target might still be the top-ranked target and potentially a very interesting lead from a therapeutic standpoint.

Similarly, not all associations with available target-disease evidence should be considered legitimate target-disease associations. **Some of our data sources rely on predictions to assess the relationship between a target and a disease.** Thus, they should be considered with caution and always take their relative support into consideration.

A diagram of data source score

Description automatically generated

**Target-disease evidence**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **PanelApp** | **GeneBurden** | **ClinVar** | **UniProt** | **G2P** | **Clingen** | **Orphanet** |
| **0.0** |  |  |  |  |  |  |  |
| **0.1** |  |  |  |  |  |  |  |
| **0.2** |  | **P=1e-7** |  |  |  |  |  |
| **0.3** |  |  | **Uncertain** |  |  |  |  |
| **0.4** |  |  |  |  |  |  |  |
| **0.5** | **Amber** |  | **Risk factor** | **Medium** | **Moderate** | **Moderate** | **Not yet assessed** |
| **0.6** |  |  |  |  |  |  |  |
| **0.7** |  |  | **Likely pathogenic** |  |  |  |  |
| **0.8** |  |  |  |  |  |  |  |
| **0.9** |  |  | **Association** |  |  |  |  |
| **1.0** | **Green** | **P<1e-17** |  | **Strong** | **Strong** | **Strong** | **Assessed** |

**Data source weighting**

Data sources are weighted on their reliability. For all genetic association data sources, the weighting is 1.

**Genetic association calculation**

To calculate the **data source association score** for a vector of evidence scored 1, 0.9 and 0.8 the Platform will follow the next logic

*Step 1: Sorting/Indexing*

evidence with score=1.0 -> positional id=1

evidence with score=0.9 -> positional id=2

evidence with score=0.8 -> positional id=3

*Step 2: Harmonic sum calculation*

harmonic sum score = 1.0/1^2 + 0.9/2^2 + 0.8/3^2

*Step 3: Scaling*

max. theoretical harmonic sum score = 1.0/1^2 + 1.0/2^2 + 1.0/3^2 + 1.0/4^2 + ...

normalised harmonic sum score = harmonic sum score / max. theoretical harmonic sum

**Sources:**

Target-disease evidence: https://platform-docs.opentargets.org/evidence#gene-burden

Target-disease associations: [https://platform-docs.opentargets.org/associations#](https://platform-docs.opentargets.org/associations)

**In our data, how is the number of genes distributed in relation to genetic association risk score?**

Note that there are no genes with scores >1.0 – although this is the strongest score across all databases, it indicates **none of the genes have a perfect score**. May mean we need to be more lenient, and not simply pick a score based on the evidence classification. **Possibly 0.3 or 0.4?**

**A screen shot of a graph

Description automatically generated**