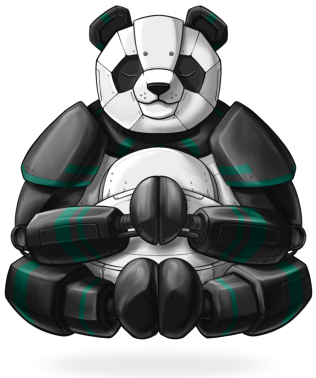
pandaomics@insilico.com

DATA DRIVEN **TARGET ID**



Problem

The search for a novel target starts from studying a vast body of literature, a full set of scientific articles published by the scientific community. The needle in the haystack here is to find any new study that highlights novel mechanisms of disease progression connected with proteins that were not yet or became recently under the spotlight of scientific investigation. Once a set of novel disease-correlated or even disease-driving genes is collected, it needs to be evaluated across multiple databases to collect the most comprehensive overview of gene and protein features and answer the most critical question: "Are any of them druggable?"

Approach

A typical target identification workflow in Pandomics includes two principal steps:

- **Selecting an appropriate AI trained model** of a gene to disease connection. Twenty distinct AI algorithms explore gene behavior extracted from OMICs dataset and text data. Once activated, the selected models bring at the top those genes enriched with the required evidence.
- **Setting up filters** that remove proteins that are not accessible for either small molecules or biologics, potentially unsafe candidates or well-known genes to focus only on novel hypotheses. The selection of robust healthy tissue expression thresholds will discard proteins with potential side effects due to increased baseline expression in many tissues.

pandaOmics

Pandaomics.com is a platform designed to shrink the most time-consuming search and classification evidence part to mere seconds. A set of artificial intelligence scoring approaches utilizes evidence connecting a gene to a disease extracted from two primary sources:

OMICs datasets

– a collection of gene expression changes and gene variants indicated across publically available datasets. It represents a gene-disease connection as a fact but not exactly validated in a focused, published experiment.

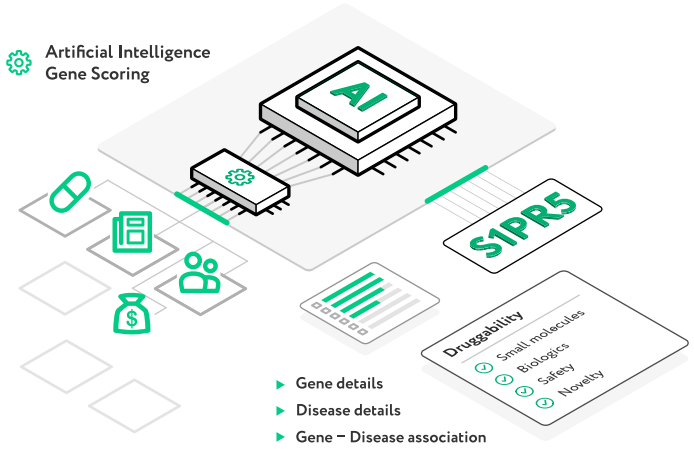
Text datasets

– a full set of textual data including scientific publications, grants, patents and more that state a connection between a gene and disease. The set contains a more validated data source of a gene-to-disease connection.

Using the **full spectrum of various scores** in Pandomics, one can follow different target identification strategies, including the one leaning toward novelty, as well as the one admiring more molecular evidence published. An equilibrated approach seems to be a promising **strategy to find target candidates for Crohn's disease**.

pandaomics@insilico.com

APPLICATION FOR **CROHN'S DISEASE**



Crohn's Disease (CD) is a subtype of Inflammatory Bowel Disease (IBD), and it can affect different sections of the gastrointestinal tract. Although the exact etiology of the disorder is not well understood, several components of risk and progression factors have been identified, including genetic factors, autoimmune mechanisms, and pathogenic infection. Independently from the exact causative mechanisms of this condition, each case shows severe inflammation in different parts of the gastrointestinal tract, and new therapies targeting different components of immune response and inflammation are being developed to cure the disorder, including TNF, PPAR-gamma and Interleukins. Unfortunately, not all CD patients are responsive to the standard targeted therapies requiring scientists to propose novel hypotheses for treatment.

For the case of CD novel target identification, the following optimal strategy is applied. OMICs datasets are a frequently underestimated source of molecular evidence, which can bring actionable hypotheses for novel targets. In several clicks, a user can:

1. Activate all OMICs-based hypothesis generation models (gene candidates demonstrate a strong connection to disease based on expression, gene variants, interactome, pathway, and other molecular levels),
2. Select proteins accessible for small molecules,
3. Set the maximum number of publications that study the gene

For the case demonstration all targets with drug candidates in clinical phases 2 and 3 were not filtered out to demonstrate the power of OMICs-based scores.

RESULTS

The system proposes several dozen target candidates indicating **S1PR5**, Sphingosine 1-phosphate receptor 5, ranked first. S1PR5 is a member of the sphingosine 1-phosphate receptor family (consisting of five proteins S1PR1, S1PR2, S1PR3, S1PR4, S1PR5) involved in immune-modulation and directly involved in suppression of innate immune responses from T cells. A non-obvious choice of target proposal is illustrated below.

So far, several compounds are targeting different family members as agonists and antagonists, both selective and nonselective. One of them, Ozanimod, is an agonist of S1PR1 and S1PR5 and was successfully launched against multiple sclerosis by Celgene (now part of Bristol-Myers Squibb) in March 2020.

In a healthy state, S1PR5 is expressed in the nervous system and blood cells representing a viable strategy for other inflammatory modulation therapies, which is reflected in repurposing efforts targeting Ulcerative colitis and Crohn's disease in Phase III of clinical trials.

Despite the apparent progress in clinical trials, it still leaves an opportunity for selective compound development targeting just the S1PR5 protein. Choosing an appropriate target-hunting strategy will identify repositioning candidates as well as entirely novel targets in just a few clicks