D1056–D1065 *Nucleic Acids Research, 2019, Vol. 47, Database issue* *Published online 20 November 2018*
doi: 10.1093/nar/gky1133

Open Targets Platform: new developments and updates two years on

Denise Carvalho-Silva[1,2,*], Andrea Pierleoni[1,2], Miguel Pignatelli[1,2], ChuangKee Ong[1,2], Luca Fumis[1,2], Nikiforos Karamanis[1,2], Miguel Carmona[1,2], Adam Faulconbridge[1,2], Andrew Hercules[1,2], Elaine McAuley[1,2], Alfredo Miranda[1,2], Gareth Peat[1,2], Michaela Spitzer[1,2], Jeffrey Barrett[2,3], David G. Hulcoop[2,4], Eliseo Papa[2,5], Gautier Koscielny[2,4] and Ian Dunham[1,2,*]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK, [2]Open Targets, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK, [3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK, [4]GSK, Medicines Research Center, Gunnels Wood Road, Stevenage, SG1 2NY, UK and [5]Biogen, Cambridge, MA 02142, USA

Received September 14, 2018; Revised October 22, 2018; Editorial Decision October 23, 2018; Accepted October 26, 2018

ABSTRACT

The Open Targets Platform integrates evidence from genetics, genomics, transcriptomics, drugs, animal models and scientific literature to score and rank target-disease associations for drug target identification. The associations are displayed in an intuitive user interface (https://www.targetvalidation.org), and are available through a REST-API (https://api.opentargets.io/v3/platform/docs/swagger-ui) and a bulk download (https://www.targetvalidation.org/downloads/data). In addition to target-disease associations, we also aggregate and display data at the target and disease levels to aid target prioritisation. Since our first publication two years ago, we have made eight releases, added new data sources for target-disease associations, started including causal genetic variants from non genome-wide targeted arrays, added new target and disease annotations, launched new visualisations and improved existing ones and released a new web tool for batch search of up to 200 targets. We have a new URL for the Open Targets Platform REST-API, new REST endpoints and also removed the need for authorisation for API fair use. Here, we present the latest developments of the Open Targets Platform, expanding the evidence and target-disease associations with new and improved data sources, refining data quality, enhancing website usability, and increasing our user base with our training workshops, user support, social media and bioinformatics forum engagement.

INTRODUCTION

Drug discovery is a long and costly endeavour characterized by high failure rates. Failure often occurs at the later stages of the drug discovery pipeline and the reasons for the low success are largely twofold: lack of safety and/or lack of efficacy. This reflects insufficient understanding of the role of the chosen target in disease, and the consequences of modulating it with a drug. Over the last several years, there has been an increase in the number of biological and chemical databases available for better understanding of drug targets (1). These databases can be used to assist with target identification, one of the most important stages in drug discovery (2).

The Open Targets Platform (https://www.targetvalidation.org) is a freely available resource for the integration of genetics, omics and chemical data to aid systematic drug target identification and prioritisation. The Open Targets Platform capitalises on publicly available databases to create a virtuous cycle where we add value to the original data by computing, scoring and ranking integrated target-disease associations (3), linking these associations back to the underlying evidence and its provenance.

We have expanded the Platform to include data from more projects and initiatives in translational research and medicinal chemistry, such as Genomics England (4), the Structural Genomics Consortium (https://www.thesgc.org) and the Institute of Cancer Research (https://www.icr.ac.

*To whom correspondence should be addressed. Tel: +44 1223 492 697; Fax: +44 1223 494 468; Email: denise@ebi.ac.uk
Correspondence may also be addressed to Ian Dunham. Tel: +44 1223 492 636; Fax: +44 1223 494 468; Email: dunham@ebi.ac.uk

uk), and continue to adhere and contribute to international naming standards and ontologies through our ongoing collaboration with the Experimental Factor Ontology (EFO) (5) and the Evidence & Conclusion Ontology (ECO) (6).

Although the main port of access for all our data is a graphical user interface (GUI) designed for bench scientists working in early drug discovery (7), we have observed an uptake in the use of our REST-API and data downloads. Moreover, due to the availability of the Open Targets Platform snapshots, our database can now be re-created by other parties. Here, we describe the developments since our first publication, focusing on the new data sources for target-disease associations, new target and disease annotations for target prioritisation, and new intuitive visualisations designed with ongoing focus on usability.

NEW DEVELOPMENTS AND PROGRESS

More data with continuing emphasis on user experience

A key factor for drug target identification and prioritisation is the causal association of the target with a disease. We compute an association score based on genetics, genomics, transcriptomics, drug information, animal models and scientific literature evidence. Our scoring framework has been described in our previous publication (3). Briefly, the computation is carried out at four different levels to give rise to evidence scores, data source scores, data type scores, and an overall association score. To compute the evidence score, we take into account specific factors that affect the strength of the evidence used for the target-disease associations (See table 2 in (3). In order to obtain a score for data sources and data types, we use the harmonic sum to aggregate individual evidence and data source scores, respectively. Our overall association score is the result of the aggregation of all data sources using the harmonic sum (Supplementary Figure S1).

Since our first publication, we have continued to explore new datasets to be included as evidence for new target-disease associations or refinement of existing ones. Our criteria to consider new data sources are: (i) relevance (can the data be used to associate targets with diseases? Does it suggest a causal link between a target and a disease? Does it enable prioritisation decision by target properties?); (ii) ease of integration (does the data use an ontology? Are the targets provided as either UniProt ID or Ensembl gene IDs? How much term mapping will be required? Is there a score or threshold that can be used to rank the data points?); (iii) accessibility (is the data publicly available, free and easy to access through an API or downloads?) and (iv) sustainability (is the data source likely to be maintained over the long term? Is the data frequently updated?). Once we select new data sources, they are combined into broader data types: Genetic associations, Somatic mutations, Drugs, Affected pathways, RNA expression, Animal models and Text mining.

In addition to including new data sources since our first publication (3), we have carried out further quality assessment of our transcriptomics evidence and expanded the scope and coverage of many of our original data sources.

New data sources for target-disease associations

We have incorporated four new data sources to enhance our evidence: Genomics England PanelApp and the PheWAS catalogue (within the data type Genetic associations) and SLAPenrich and PROGENy (for the data type Affected pathways).

Genetic associations

Since our previous publication, we have added two new data sources as evidence for Genetic associations between targets and Mendelian and more common diseases: Genomics England PanelAPP and the PheWAS catalog, respectively. With these new data sources, we have been able to identify new associations (e.g. between SERPING1 and Immunodeficiency due to an early component of complement deficiency based on evidence from the Genomics England PanelAPP, or between MC1R in hyperlipidemia based on evidence from the PheWAS catalog) or added further support to previously identified associations (e.g. between KCNE3 in Brugada syndrome based on evidence from the Genomics England PanelAPP, or NOD2 in Crohn's disease based on evidence from the PheWAS catalog).

We have included the Genomics England PanelApp Green genes (version 1+ panels) (4) along with their (mainly) rare, Mendelian diseases or phenotypes, providing these can be mapped to an ontology, such as EFO (5), Orphanet (http://www.orpha.net) or Human Phenotype Ontology (HP) (8). We use the PanelApp WebServices (https://panelapp.genomicsengland.co.uk/#!Webservices) to obtain the associations and Ontoma (https://pypi.org/project/ontoma/) for the automatic mapping of diseases and phenotypes.

The Genomics England Green genes are curated and crowdsourced by experts; hence the target-disease associations that are supported by this evidence in our Platform have the highest score of 1.

For common and complex diseases, we have added genetic evidence from PheWAS (9), scored following our methodology for GWAS evidence (3) but scaled according to the maximum number of cases (8800) and the *P*-value range (0.05 and 1e-25) of the PheWAS data.

Details on the scoring of these new data sources for Genetic associations are described in our help documentation (https://docs.targetvalidation.org/getting-started/scoring).

Affected pathways

Besides the new data sources for Genetic associations, we have also included two new data sources for Affected pathways, more specifically in cancer. The new data sources, SLAPenrich (10) and PROGENy (11), have mostly highlighted new associations, such as EGFR in squamous cell lung carcinoma based on PROGENy and PTEN in prostate adenocarcinoma based on SLAPenrich.

SLAPenrich identifies pathways that harbour genomic alterations, more frequently than expected by chance, across a population of cancer samples from their somatic mutation profiles. The alteration status of a pathway is determined by the collective status of its genes: a pathway is altered in a sample if at least one of its genes is somatically mutated

Published online 8 December 2016 *Nucleic Acids Research, 2017, Vol. 45, Database issue* ***D985–D994***
doi: 10.1093/nar/gkw1055

Open Targets: a platform for therapeutic target identification and validation

Gautier Koscielny[1,2,*], Peter An[1,3], Denise Carvalho-Silva[1,4], Jennifer A. Cham[1,4], Luca Fumis[1,4], Rippa Gasparyan[1,3], Samiul Hasan[1,2], Nikiforos Karamanis[1,4], Michael Maguire[1,4], Eliseo Papa[1,3], Andrea Pierleoni[1,4], Miguel Pignatelli[1,4], Theo Platt[1,3], Francis Rowland[1,4], Priyanka Wankar[1,3], A. Patrícia Bento[1,4], Tony Burdett[1,4], Antonio Fabregat[1,4], Simon Forbes[1,5], Anna Gaulton[1,4], Cristina Yenyxe Gonzalez[1,4], Henning Hermjakob[1,4,6], Anne Hersey[1,4], Steven Jupe[1,4], Şenay Kafkas[1,4], Maria Keays[1,4], Catherine Leroy[1,4], Francisco-Javier Lopez[1,4], Maria Paula Magarinos[1,4], James Malone[1,4], Johanna McEntyre[1,4], Alfonso Munoz-Pomer Fuentes[1,4], Claire O'Donovan[1,4], Irene Papatheodorou[1,4], Helen Parkinson[1,4], Barbara Palka[1,4], Justin Paschall[1,4], Robert Petryszak[1,4], Naruemon Pratanwanich[1,4], Sirarat Sarntivijal[1,4], Gary Saunders[1,4], Konstantinos Sidiropoulos[1,4], Thomas Smith[1,4], Zbyslaw Sondka[1,5], Oliver Stegle[1,4], Y. Amy Tang[1,4], Edward Turner[1,4], Brendan Vaughan[1,4], Olga Vrousgou[1,4], Xavier Watkins[1,4], Maria-Jesus Martin[1,4], Philippe Sanseau[1,2], Jessica Vamathevan[4], Ewan Birney[1,4], Jeffrey Barrett[1,4,5] and Ian Dunham[1,4,*]

[1]Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, [2]GSK, Medicines Research Center, Gunnels Wood Road, Stevenage, SG1 2NY, UK, [3]Biogen, Cambridge, MA 02142, USA, [4]European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, [5]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK and [6]National Center for Protein Research, No. 38, Life Science Park Road, Changping District, 102206 Beijing, China

Received August 19, 2016; Revised October 19, 2016; Editorial Decision October 20, 2016; Accepted November 03, 2016

ABSTRACT

We have designed and developed a data integration and visualization platform that provides evidence about the association of known and potential drug targets with diseases. The platform is designed to support identification and prioritization of biological targets for follow-up. Each drug target is linked to a disease using integrated genome-wide data from a broad range of data sources. The platform provides either a target-centric workflow to identify diseases that may be associated with a specific target, or a disease-centric workflow to identify targets that may be associated with a specific disease. Users can easily transition between these target- and disease-centric workflows. The Open Targets Validation Platform is accessible at https://www.targetvalidation.org.

INTRODUCTION

The fundamental tenet of pharmacology is that a drug (small molecule or biological) can be identified that specifically interacts with a target molecule (usually a protein) to modulate a physiological process and thus alter the course of a disease (1,2). The pharmaceutical industry has developed powerful approaches to discover and optimize drug molecules that affect the function of a target. There are also complex strategies in practice to deal with drug efficacy, dosing and safety issues that accompany getting a drug into humans and finally to market. However, analysis of progress through development pipelines has highlighted that lack of efficacy is a major cause of failure, particularly in the later, more expensive, clinical stages (3,4). The implication is that

*To whom correspondence should be addressed. Tel: +44 1223 494 650; Fax: +44 1223 494 468; Email: gautier.x.koscielny@gsk.com
Correspondence may also be addressed to Ian Dunham. Tel: +44 1223 492 636; Fax: +44 1223 494 468; Email: dunham@ebi.ac.uk
Present address: Francis Rowland, Sigma Consulting Solutions, The Officers' Mess Business Centre, Royston Road, Duxford Cambridgeshire, CB22 4QR, UK.