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 in that sample. Then, SLAPenrich quantifies the divergence from expectation of the total number of samples with genomic alterations in a pathway, using a Poisson binomial distribution. The SLAPenrich evidence used for the associations between targets and cancer is based on Reactome-pathway gene-sets only, diverging from the original analysis (10).

PROGENy (11) is a method to infer pathway activity from gene expression. The data includes 11 pathways, namely EGFR (epidermal growth factor receptor), MAPK (mitogen-activated protein kinase), PI3K (phosphoinositide 3-kinase), VEGF (vascular endothelial growth factor), JAK-STAT (janus kinase-signal transducer and activator of transcription), TGFb (transforming growth factor beta), TNFa (tumor necrosis factor alpha), NFkB (nuclear factor kappa-light-chain-enhancer of activated B cells), Hypoxia, p53-signaling and DNA damage response, and cell death via apoptosis (Trail). In contrast to other pathway methods, PROGENy is based on the gene expression signature downstream of the pathway, rather than on the expression of the pathway components, generating signatures that represent a consensus over many different conditions. Contrary to SLAPenrich, which implicates pathways but not their activation or inhibition, PROGENy scores indicate whether a pathway is activated or inhibited if they are higher or lower between conditions. Both tools can hence be used in combination to infer driver pathways hit by mutations, and then link those to their downstream effects.

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

Updated data sources for disease and target associations

In addition to new data sources, since our first publication we have also had updates on the original data sources through our frequent cycles of data release. Our data ingest and processing pipeline is now run every two months for the integration of the most up-to-date information from our data providers. Each release has a YY.MM timestamp and may contain increased (and updated) coverage of the evidence described in our original paper (3) as well as new data sources (see above section). Furthermore, some of the original data sources have had their scope expanded or amended through ongoing projects within Open Targets. These correspond to the (i) inclusion of targeted genotyping arrays from the GWAS Catalogue (12); (ii) addition of the new tier 2 cancer genes (13) from the Cancer Gene Census (14); (iii) coverage of trinucleotide repeat data from ClinVar (15) available in the European Variation Archive (https://www.ebi.ac.uk/eva/?Home) and (iv) withdrawal of differential expression studies reported on human cell lines with no disease as study factor from Expression Atlas (16). In the following sections, we describe the updates in the original data sources used for our target-disease associations in more detail.

Targeted, non genome-wide genotyping arrays

We have started a collaboration with the NHGRI-EBI GWAS Catalogue (12) for the inclusion of non genome-wide arrays including the Immunochip (17) for immunogenetics, and Metabochip (18) for metabolic diseases. So far, this has enabled us to include 823 SNP-trait associations for 120 independent associations curated from 55 publications. The inclusion of both Immunochip and Metabochip arrays increases the availability of germline variants (or SNPs) that are associated with autoimmune, inflammatory and metabolic diseases. These causal genetic variants in the Open Targets Platform will help uncover strong candidate genes for those diseases, prioritise target-disease associations and explore pleiotropy, if the candidate genes are associated with more than one of the diseases for which the array was designed.

Tiered cancer gene census

In our initial paper (3), we described somatic mutation evidence from the Cancer Gene Census (14) used to support target-cancer associations. This census has recently introduced new criteria to assess the level of evidence that supports a gene as a driver gene in cancer, which leads to the concept of a tier system (13). Genes in tier 1 must have: (i) evidence of activity that may drive or suppress cancer; (ii) evidence of mutations, detected in cancer that change the activity of the protein and promote oncogenic transformation and (iii) evidence that the somatic mutation patterns in cancer samples are typical of tumour suppressor genes (e.g. inactivating mutations) or of oncogenes (e.g. missense mutations). Although tier 2 genes are strongly associated with a role in cancer, they have less evidence than their tier 1 counterparts. For both tier 1 and 2 genes, Poisson tests are carried out to assess whether somatic mutations in a gene occur more frequently than in other genes in the same disease, and whether a gene is mutated significantly more frequently in a given disease when compared to all other diseases. All mutations detected in the Cancer Gene Census genes have a base association score of 0.5, which is modified by applying the following rules: (i) for tier 1 genes, a significant result (FDR < 0.025) for either of each of the two Poisson tests adds 0.25 to the score; (ii) for tier 2 genes, the score will be 0.5; (iii) if only one sample is mutated, 0.25 is subtracted from the score and (iv) if tier 1 genes are known to drive cancer only through fusions, all mutation types except fusions get a score of 0.5. Fusions will then be scored according to the rules above.

Trinucleotide repeat expansions and new clinical significance terms

We import trinucleotide repeats from ClinVar (15) that are stored in the European Variation Archive (https://www.ebi.ac.uk/eva/?Home) for the genetic associations between triplet repeat expansion disorders, e.g. Huntington's Disease and Fragile X syndrome, and their possible drug targets, such as Huntingtin and Synaptic functional regulator FMR1 proteins. In order to incorporate trinucleotide repeats, a new consequence term, trinucleotide expansion (SO_0002165) (https://www.targetvalidation.org/variants) was defined by ECO (6). Moreover, besides mutations described as 'pathogenic', which have been included since the first release of the Open Targets Platform, we have

Table 1. Sources and evidence counts used for target-disease associations in the Open Targets Platform

Data source*	Data type	Evidence Count**
Genomics England PanelAPP (v2.2.0)	Genetic associations	15 289
PheWAS catalogue (Sep-2017)	Genetic associations	47 302
GWAS catalogue (July 2018)	Genetic associations	101 511 (*32 363*)
UniProt (July 2018)	Genetic associations	26 640 (*21 870*)
UniProt literature (July 2018)	Genetic associations	4494
European Variation Archive∧ (July 2018)	Genetic associations	73 805 (*28 050*)
Gene2Phenotype (May 2017)	Genetic associations	1604 (*975*)
UniProt (July 2018)	Somatic mutations	282
Cancer Gene Census (COSMIC v85)	Somatic mutations	55 963 (*23 440*)
IntOGen (December 2014)	Somatic mutations	2371 (*2377*)
European Variation Archive∧ (July 2018)	Somatic mutations	7624 (*456*)
ChEMBL (v24)	Drugs	410 436 (*120 520*)
Reactome (v65)	Affected pathways	9735 (*6143*)
PROGENy (April 2018)	Affected pathways	308
SLAPenrich (August 2017)	Affected pathways	89 661
Expression Atlas (February 2018)	Expression	288 273 (*529 084*)
Europe PMC (July 2018)	Text mining	4 906 527 (*3 678 967*)
PhenoDigm (November 2017)	Animal model	465 887 (*395 331*)

*Database version (or date) in parentheses.

**As per 18.08 release of the Open Targets Platform. Parentheses show the number (in italics) of evidence count reported previously (3). Note, the reduction in the number of evidence from Expression Atlas (see main text for explanation).

∧Containing ClinVar data from May 2017.

Detailed target-disease association counts can be found in the Supplementary Table.

Data sources in bold are new data, whereas the remaining sources have been described in our first publication and shown here are updates from the previous report.

begun incorporating mutation data with other terms of clinical significance from ClinVar, namely 'protective', 'association', 'risk factor', 'affects' and 'drug response'. The full description of these terms are available elsewhere (https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/).

Human cell lines

We have removed studies from Expression Atlas where experiments were carried out in human cell lines where the disease was not a factor in the study (e.g. cell lines derived from cancers used in other studies). This has lead to a 34% reduction in the number of evidence strings for the assessment of differential expression of drug targets, and therefore removal of false-positive associations in the RNA expression data type.

In summary, new data sources, quality assessment and further refinements to the original set of data sources have increased the scope of our target-disease associations. A summary of the latest set of data sources and count of evidence from each source are provided in Table 1. The statistics for our releases (current and previous ones) can be found in our Release Notes (https://www.targetvalidation.org/release-notes).

New annotations and visualisations for targets and diseases

Besides providing target-disease associations, the Open Targets Platform integrates comprehensive annotation of individual human targets and diseases on dedicated pages to support target prioritisation. The target profile page contains information at the gene and protein levels, whereas the disease profile page displays disease annotations, such as phenotypes and ontology classification. Note that our targets can be both protein coding genes and non-coding RNA genes, such as HOTAIR and MIR23A non-coding genes.

Since our first publication, target and disease annotations have been enhanced with new visualisations and data. A new plugin architecture for the pages enables widgets and/or data to be added more easily, allowing increased content flexibility and faster loading time for a better user experience. We have also changed the order of the annotations displayed in the target profile page based on usage statistics from anonymised web traffic logs, now displaying the more relevant information first, at the top of the page. In the following sections, we provide details on these new target annotations, in addition to updates on the visualisation of expression data and scientific literature.

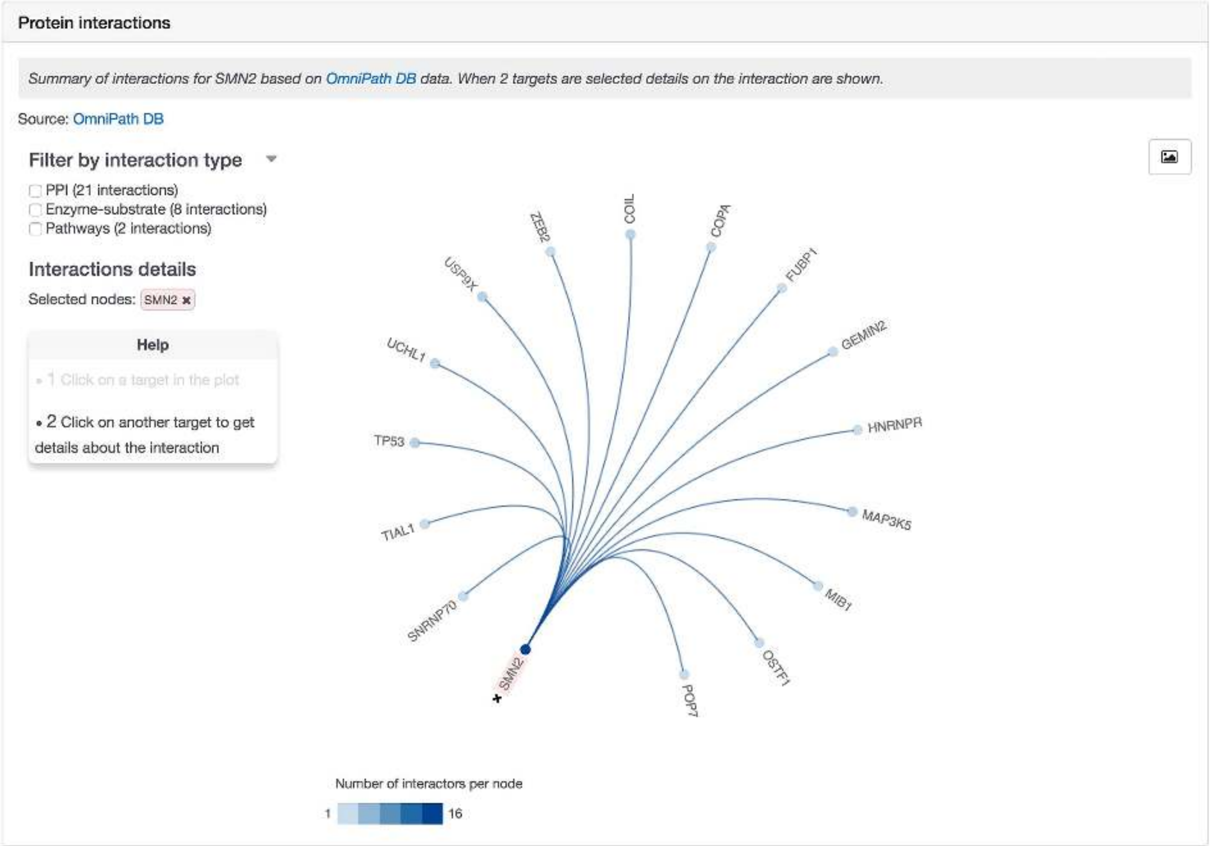
Target enabling packages and chemical probes

A Target Enabling Package (TEPs) is a collection of reagents, protocols and data for rapid exploration and characterization of proteins (potential drug target candidates) with genetic linkage to key disease areas (19). All 16 TEPs, currently available from the Structural Genomics Consortium portal, can be accessed from the relevant target profile pages in the Open Targets Platform.

We also link to a set of the 215 high-quality chemical probes (20,21) available for 188 different targets, giving access to reagents and assays to aid *in vitro* and/or *in vivo* investigation of phenotype and mechanism of a target. An additional set of potential chemical probes for 2300 human targets from Probe Miner (https://probeminer.icr.ac.uk/#/) is also available in the Open Targets Platform.

Protein–protein interactions

We provide a summary of direct protein interactions with the selected target to explore interactome information and facilitate drug target prioritisation. Currently, we display a



Protein interactions

Summary of interactions for SMN2 based on OmniPath DB data. When 2 targets are selected details on the interaction are shown.

Source: OmniPath DB

Filter by interaction type

□ PPI (21 interactions)
□ Enzyme-substrate (8 interactions)
□ Pathways (2 interactions)

Interactions details

Selected nodes: SMN2 ✖

Help

- 1 Click on a target in the plot
- 2 Click on another target to get details about the interaction

Figure 1. Interactive visualisation of protein–protein interactions in dedicated target profile pages.

summary of protein interaction data from OmniPath (22). This data can be filtered by enzyme–substrate interactions, protein–protein interactions or pathways (Figure 1).

Mouse phenotypes

We list the annotated strain-specific phenotypes and the allelic composition of every laboratory mouse with a knockout gene curated by and available in the Mouse Genome Informatics database (23).

Cancer hallmarks

We summarise information on cancer hallmarks (24,25), which are curated by COSMIC (26) and integrated in the Cancer Gene Census (14). These essential alterations in cell physiology that can dictate malignant growth can be found in the profile pages of a target implicated in cancer. We list the hallmarks (e.g. invasion and metastasis, change of cellular energetics) for a given target as either being promoted or suppressed. We currently have cancer hallmarks for 251 targets from the Cancer Gene Census.

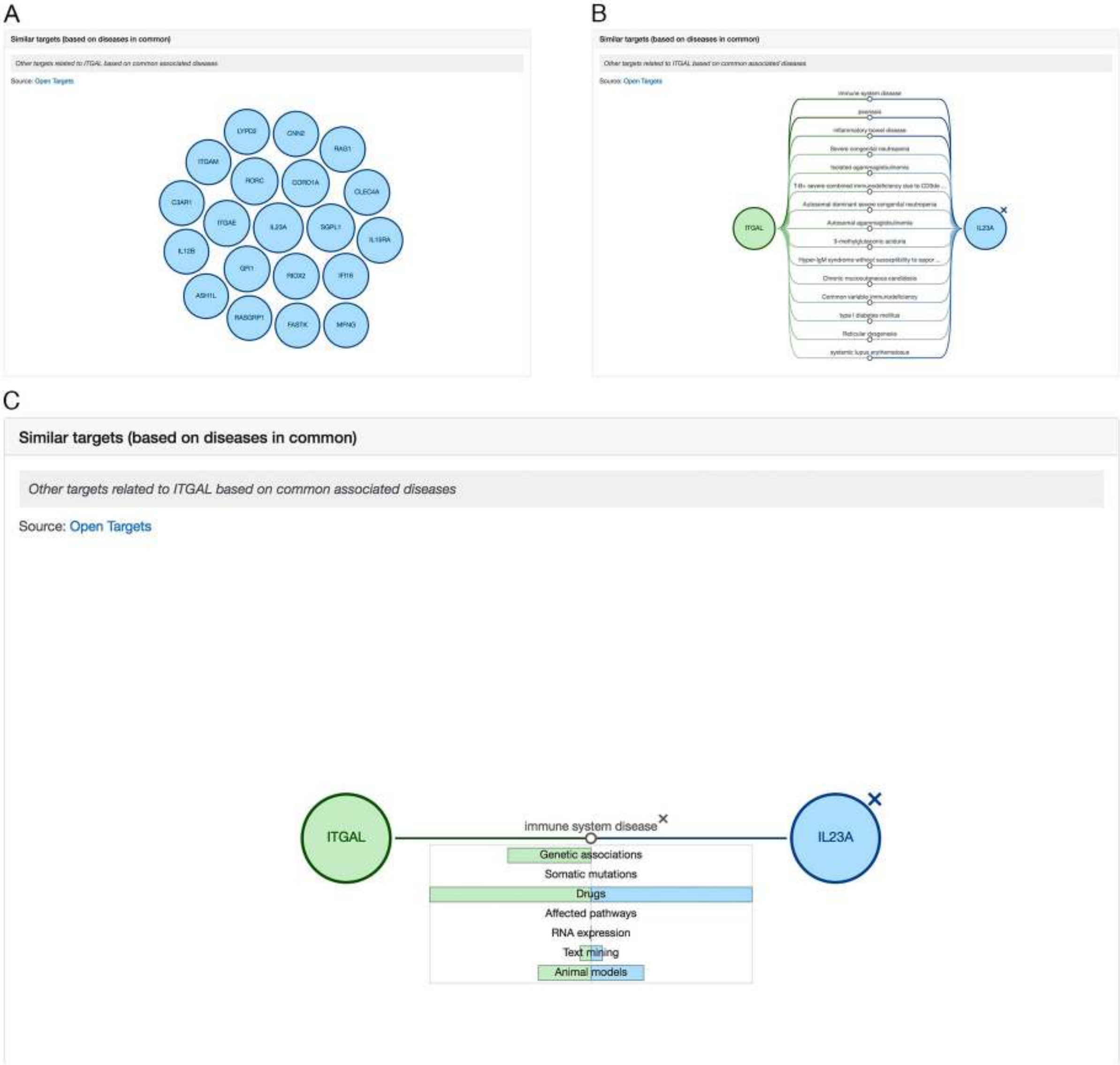
Cancer biomarkers

We incorporate a collection of genomic biomarkers of drug responses (sensitivity, resistance and toxicity) and their level of clinical significance from the Cancer Biomarkers database (27).

Similar targets and their diseases

We have developed a new feature to show suggested targets that are similar to any target of choice. This is based on a network analysis of shared diseases obtained from our set of target-disease associations as a bipartite graph, with targets and diseases as vertices. In order to reduce noise in the data, we consider only disease-association pairs with at least three evidence supporting the edge, and whose overall association score is greater than 0.1. We calculate a target relationship score based on the ratio of shared diseases between the targets to the total number of diseases for both targets.

Our relationship scoring method (i) outputs a closer distance between two targets sharing a rare disease than two targets sharing diseases that are data rich, such as cancer and (ii) considers shared targets that are specifically linked to fewer diseases more relevant than targets that are commonly linked to many types of diseases. In order for this process to remain computationally feasible, given the billions of possible target–target (and disease–disease) combinations, we have implemented an efficient, high performance computation strategy, which uses (i) an heuristic estimation from aggregate statistics allowing us to skip the computation of the distance for pairs that are below the cut-off and (ii) LSH (locality-sensitive hashing) (28–30) to calculate target relatedness, retaining only the most confident relationships. The resulting set of similar targets are displayed as an interactive visualisation (Figure 2A), which



Similar targets (based on diseases in common)

Other targets related to ITGAL based on common associated diseases

Source: Open Targets

A
LYPD3
CNN2
RAG1
ITGAM
RORC
CORO1A
CLEC4A
C3AR1
ITGAE
IL23A
SGPL1
IL15RA
IL12B
GPI1
RIOX2
IFI16
ASH1L
RASGRP1
FASTK
MPNG
B
ITGAL
IL23A
immune system disease
psoriasis
inflammatory bowel disease
Severe congenital neutropenia
Isolated agammaglobulinemia
T-B+ severe combined immunodeficiency due to CD3de ...
Autosomal dominant severe congenital neutropenia
Autosomal agammaglobulinemia
3-methylglutaconic aciduria
Hyper-IgM syndrome without susceptibility to oppor ...
Chronic mucocutaneous candidiasis
Common variable immunodeficiency
type I diabetes mellitus
Reticular dysgenesis
systemic lupus erythematosus
C
ITGAL
immune system disease
IL23A
Genetic associations
Somatic mutations
Drugs
Affected pathways
RNA expression
Text mining
Animal models

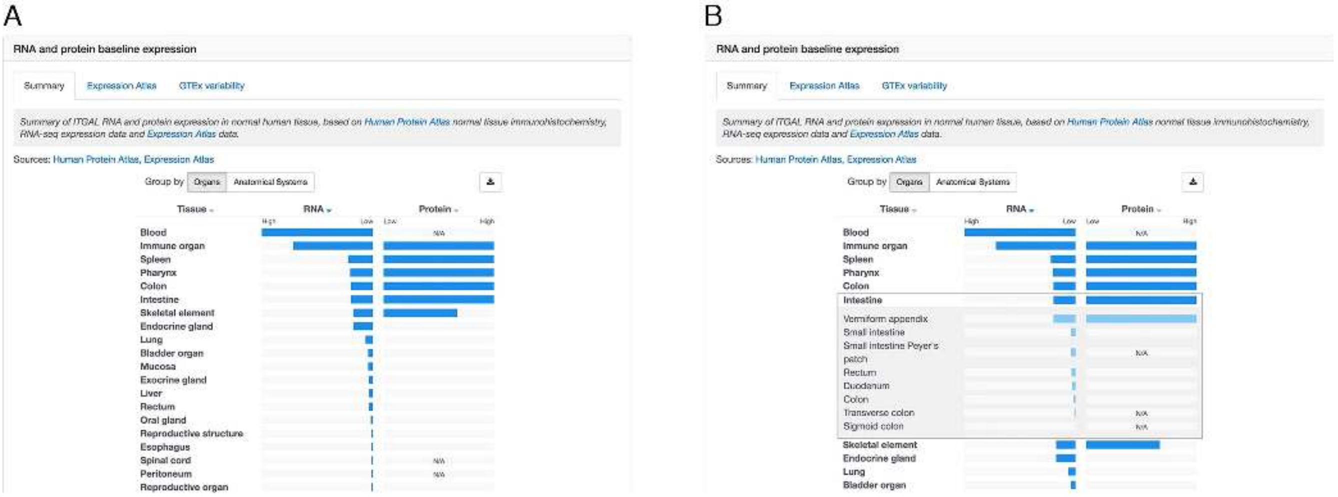
Figure 2. Similar targets are displayed as an interactive visualisation (**A**) in the target profile page. By selecting a target, the view gets updated to show the diseases shared between any two targets (**B**). Clicking on any of the shared diseases reveals the underlying evidence (e.g. Genetic associations, Drugs, Text mining, Animal models) that supports the association between a disease and its two selected targets.

summarises the top ranking shared diseases for any two targets up to 20 (Figure 2B), and presents the underlying evidence (e.g. Genetic associations, Affected pathways, Drugs) supporting the association (Figure 2C).

The relationship scoring procedure is also applied for diseases sharing the same targets and the relations can be visualised on the disease profile page.

RNA and protein baseline expression

When trying to identify a new target, users working in drug discovery often want to understand the expression of the target across human tissues and cells. We have enhanced both the data and the visualisation for baseline expression by combining both RNA (16) and protein (31) expression data under a single section entitled 'RNA and protein baseline expression'. Within this section, there are three tabs, 'Summary', 'Expression Atlas' and 'GTEx variability'. The first tab shows RNA and protein expression data side by side for a quicker comparison (Figure 3A). This can be especially useful for targets that show different levels of RNA and protein expression, e.g. low expression at the RNA level, but high expression at the protein level. The expression data can be visualised grouped either by 'Organs' (by default) or 'Anatomical Systems'. For either option, users can click on the name of a tissue, e.g. 'Intestine', and see a detailed breakdown of expression in different parts of the tissue/organ, such as 'Vermiform appendix' and 'Duodenum' (Figure 3B). Two other displays of expression data are also available in additional tabs: an interactive heat map



Summary of ITGAL RNA and protein expression in normal human tissue, based on Human Protein Atlas normal tissue immunohistochemistry, RNA-seq expression data and Expression Atlas data.

Sources: Human Protein Atlas, Expression Atlas

Summary of ITGAL RNA and protein expression in normal human tissue, based on Human Protein Atlas normal tissue immunohistochemistry, RNA-seq expression data and Expression Atlas data.

Sources: Human Protein Atlas, Expression Atlas

Figure 3. RNA and protein expression data are displayed side by side for easy comparison of target expression levels in healthy human tissues. (**A**) Each horizontal bar representing a tissue, e.g. Intestine, can be expanded to provide a detailed breakdown of expression in different parts of the tissue/organ, such as 'Vermiform appendix' and 'Duodenum' (**B**).

from Expression Atlas (16) and a box plot to visualise gene expression variability in GTEx data (32) (Figure 4).

Bibliography

We have improved how the 'Bibliography' section is displayed on both target and disease profile pages. We have moved to a bespoke navigation and visualisation of scientific papers from Europe PMC (33) using 'chips' (Figure 5A). These are created from automated topic identification and entity recognition using LINK, the Open Targets LIterature coNcept Knowledgebase (https://link.opentargets.io). LINK extracts key entities from PubMed abstracts (34) using a precompiled set of dictionaries to recognise genes and diseases from the Open Targets Platform, phenotypes from the Human Phenotype Ontology (8), drugs in clinical trials or on the market from ChEMBL (35), and relevant MESH headings, such as anatomy, diagnostics and locations. We analyse each title and abstract with spaCY (36), and extract key concepts and semantic relations in the form of subject-predicate-object triples.

A drop-down menu is also available in the 'Bibliography' section to filter the publications according to key entities, i.e. concepts (e.g. loss of heterozygosity), genes, diseases, drugs, journals and authors. Both drop down menu and 'chips' allow for interactive filtering of abstracts (Figure 5B) selected in the target and disease profile pages of the Open Targets Platform.

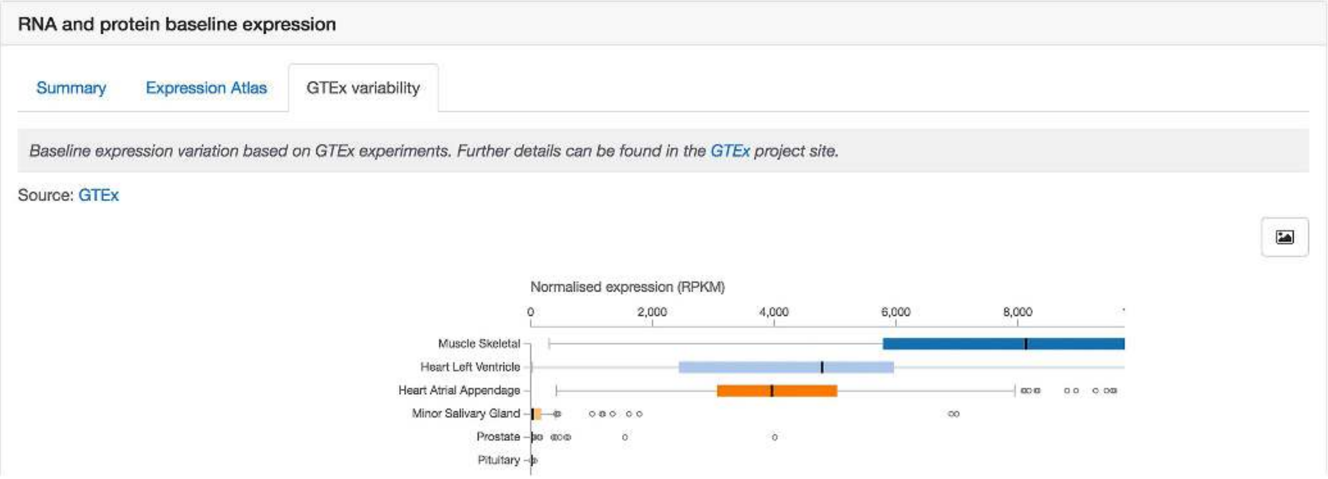
New filtering options in the user interface and URL sharing

We have introduced filters on the GUI to allow for targets to be selected based solely on their properties and facilitate prioritisation of the most promising (best) targets for downstream analysis. The properties available for filtering are 'Target class', e.g. enzyme, surface antigen, as defined by ChEMBL (35) and 'RNA tissue specificity', e.g. to restrict targets that are expressed preferentially in the selected cell or tissue (e.g. brain) compared to other tissues, based on Expression Atlas data (16). We have also implemented 'Your target list', where users can upload their own list (in .CSV or .TXT) of targets (either as Ensembl gene IDs, HGNC symbols, UniProt IDs or synonyms) to restrict the associations table to the user's targets only.

Other changes to the interface include URL sharing for specific views and pages, such as the bubbles view ('?view=t:bubbles' in the URL) and the evidence page based on a specific type of data e.g. Genetic associations ('view=sec:genetic_association' in the URL).

Alternative ways to access data

The user interface of the Open Targets Platform allows searches for an entry to be carried out on a one-by-one case basis only: one disease or one target, for example. However, we have use cases that start from a list of targets, rather than a single target. For bulk searching, we have launched a batch search at https://www.targetvalidation.org/batch-search, an easy-to-use and interactive web tool that takes a list of up to 200 targets, identified by HGNC symbols, UniProt or Ensembl IDs, or gene/protein synonyms, and uploaded as .TXT or .CSV. In addition to upload, users can also paste their targets into the box available in the entry page. The batch search will return (i) diseases associated with the list of targets ranked by significance using a hypergeometric distribution; (ii) pathways enriched in the set of targets ranked by the probability of finding a pathway that is associated with and specific to the target list; (iii) gene ontology terms enriched among the targets, ranked following a hypergeometric distribution; (iv) drugs that are known to modulate the targets in the list and (v) a visualisation of protein interactions, showing the interactions between the targets in the batch list. From the batch search results, links to other pages in the Open Targets Platform are available for further exploration of pathway and drug summaries, and/or associations and evidence pages. A detailed tutorial on the batch



Baseline expression variation based on GTEx experiments. Further details can be found in the GTEx project site.

Source: GTEx

Figure 4. An additional visualisation to summarise expression data is also available, depicting gene expression variability in GTEx data.

search tool is available in our documentation page (https://docs.targetvalidation.org/getting-started/batch-search).

The Open Targets REST API provides extra flexibility for data retrieval and filtering options for larger queries, which are not supported by the batch search. The REST API has been updated to version 3, changed its base URL to https://api.opentargets.io/v3/platform and no longer requires an API key for fair usage. More details on the REST API can be found in our documentation page (https://api.opentargets.io/v3/platform/docs/swagger-ui), an interactive page where users can execute the REST API calls and check the response before using these calls in workflows for automated analysis.

New name, new homepage and improved search functionality

We have renamed the portal from 'Target Validation Platform' to 'Open Targets Platform' in line with the rebranding of the overall partnership (http://www.opentargets.org). The homepage has been completely redesigned to highlight the main entry point for our users, the search function. In addition to target and disease names (or symbols and their synonyms), we allow searching for phenotypes, orthologous genes (e.g. from mouse, rat, fruit fly, worm) and drugs. If searching for a drug, the Platform accepts chemical (e.g. acetylsalicylic acid), generic (aspirin) and brand (Durlaza) names, and will return a list of targets and diseases that have associations where the drug is involved. When searching for drug names, these get matched to the 'drugs.evidence data' field, which contains words, such as Acetylsalicylic acid, Acetylsalicylic Acid, acetylsalic acid, Salicylic Acid Acetate, Acetylsalic Acid, Durlaza.

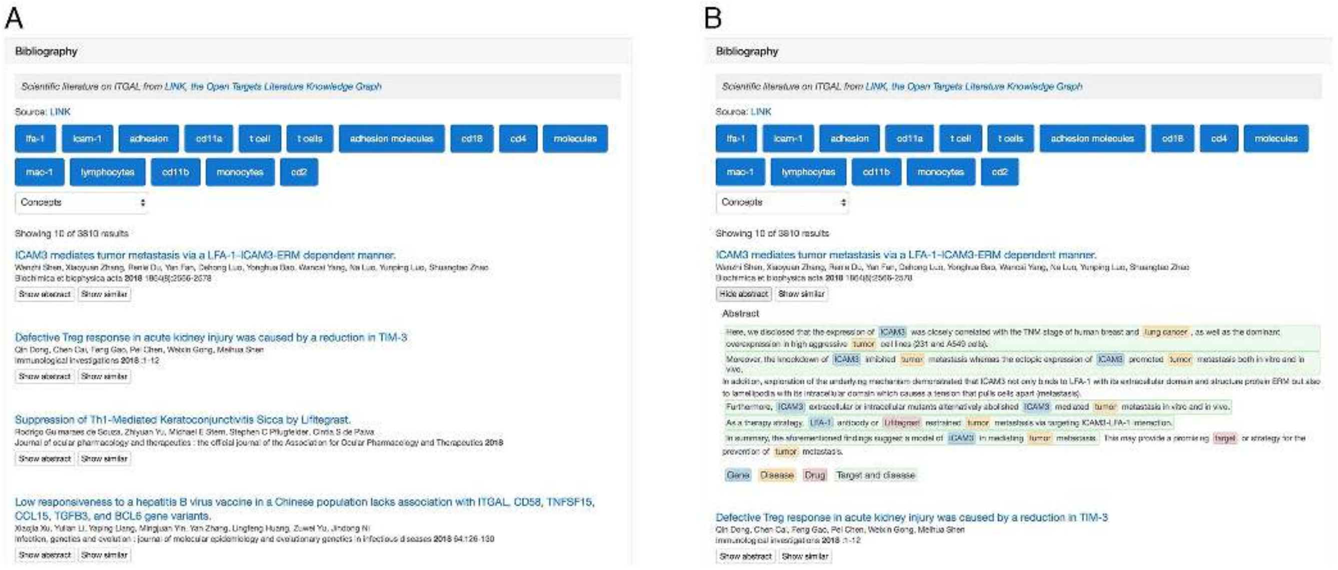
The new homepage displays the main statistics of the latest release of the Open Targets Platform, namely the number of targets, diseases, associations and data sources that provide evidence for the target-disease associations. It also provides links to other modes of data access (see above), directs users to free tutorials and documentation pages, and includes feeds from Open Targets social media channels, such as the Blog, Twitter and Facebook.

Outreach, training and user support

The GUI https://www.targetvalidation.org is the main portal to access data from the Open Targets Platform. It had 685 visits and 2369 unique page views from 45 countries in the week prior to the submission of our first publication, i.e. between 12 August 2016 and 19 August 2016. We have observed an increase to 1290 visits and 5411 unique page views from 62 countries for the same eight-day period in 2018 (12 August to 19 August). We offer free hands-on workshops on the Open Targets Platform, both face-to-face and as live webinars (Carvalho-Silva *et al.* DOI: 10.1371/journal.pcbi.1006419). Our recorded webinars, online tutorials and short demos are available on the Open Targets YouTube channel (https://www.youtube.com/channel/UCLMrondxbT0DIGx5nGOSYOQ/featured), which features 12 videos. Our user community can follow our news and upcoming developments on Twitter (twitter.com/targetvalidate), Facebook (https://www.facebook.com/OpenTargets/), LinkedIn (https://www.linkedin.com/company/open-targets/), and by subscribing to our monthly newsletters (http://bit.ly/Open-Targets-News). We also have a blog (http://blog.opentargets.org) that featured 17 posts over the last 12 months and is mirrored on Medium (https://medium.com/opentargets). Direct support via email is available through support@targetvalidation.org.

CONCLUSIONS

The Open Targets Platform is part of an increasing effort on the integration of public resources to assist target identification and prioritisation for drug discovery. Although many of these resources focus on interactions between drug compounds and their targets, fewer databases explore the evidence available that links a target with a disease. Related resources to the Open Targets Platform, such as DisGeNET (37,38) and Pharos (39,40), have been compared and their complementarities and differences highlighted in our previous publication (3). Recently, Zhang *et al* (1) have provided a more comprehensive comparative analysis.



Bibliography

Scientific literature on ITGAL from LINK, the Open Targets Literature Knowledge Graph

Source: LINK

Showing 10 of 3810 results

ICAM3 mediates tumor metastasis via a LFA-1-ICAM3-ERM dependent manner.
Wenzhi Shen, Xiaoyuan Zhang, Renle Du, Yan Fan, Dehong Luo, Yonghua Bao, Wancai Yang, Na Luo, Yunping Luo, Shuangtao Zhao
Biochimica et biophysica acta 2018 1864(8):2566-2578

Defective Treg response in acute kidney injury was caused by a reduction in TIM-3
Qin Dong, Chen Cai, Feng Gao, Pei Chen, Weixin Gong, Meihua Shen
Immunological investigations 2018 :1-12

Suppression of Th1-Mediated Keratoconjunctivitis Sicca by Lifitegrast.
Rodrigo Guimaraes de Souza, Zhiyuan Yu, Michael E Stern, Stephen C Pflugfelder, Cintia S de Paiva
Journal of ocular pharmacology and therapeutics : the official journal of the Association for Ocular Pharmacology and Therapeutics 2018

Low responsiveness to a hepatitis B virus vaccine in a Chinese population lacks association with ITGAL, CD58, TNFSF15, CCL15, TGFB3, and BCL6 gene variants.
Xiaojia Xu, Yulan Li, Yaping Liang, Mingjuan Yin, Yan Zhang, Lingfeng Huang, Zuwei Yu, Jindong Ni
Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 2018 64:126-130

Bibliography

Scientific literature on ITGAL from LINK, the Open Targets Literature Knowledge Graph

Source: LINK

Showing 10 of 3810 results

ICAM3 mediates tumor metastasis via a LFA-1-ICAM3-ERM dependent manner.
Wenzhi Shen, Xiaoyuan Zhang, Renle Du, Yan Fan, Dehong Luo, Yonghua Bao, Wancai Yang, Na Luo, Yunping Luo, Shuangtao Zhao
Biochimica et biophysica acta 2018 1864(8):2566-2578

Abstract

Here, we disclosed that the expression of ICAM3 was closely correlated with the TNM stage of human breast and lung cancer , as well as the dominant overexpression in high aggressive tumor cell lines (231 and A549 cells).
Moreover, the knockdown of ICAM3 inhibited tumor metastasis whereas the ectopic expression of ICAM3 promoted tumor metastasis both in vitro and in vivo.
In addition, exploration of the underlying mechanism demonstrated that ICAM3 not only binds to LFA-1 with its extracellular domain and structure protein ERM but also to lamellipodia with its intracellular domain which causes a tension that pulls cells apart (metastasis).
Furthermore, ICAM3 extracellular or intracellular mutants alternatively abolished ICAM3 mediated tumor metastasis in vitro and in vivo.
As a therapy strategy, LFA-1 antibody or Lifitegrast restrained tumor metastasis via targeting ICAM3-LFA-1 interaction.
In summary, the aforementioned findings suggest a model of ICAM3 in mediating tumor metastasis. This may provide a promising target or strategy for the prevention of tumor metastasis.

Defective Treg response in acute kidney injury was caused by a reduction in TIM-3
Qin Dong, Chen Cai, Feng Gao, Pei Chen, Weixin Gong, Meihua Shen
Immunological investigations 2018 :1-12

Figure 5. The new visualisation in the 'Bibliography' section of target and disease profile pages. Both titles (**A**) and abstracts (**B**) are available and can be filtered by selecting one of the 'chips' at the top of the table. A drop-down menu is also available to allow selection of publications according to the available (biological) concepts, genes, diseases, drugs, journals and authors.

Following its launch in December 2015, the Open Targets Platform has been expanding its scope, increasing its data coverage, reaching out to user communities worldwide, and growing its user base, all carried out with a constant focus on usability and user design to maintain its easy-to-use and interactive features. We currently integrate over six billion evidence from 18 publicly available data sources and compute almost three billion associations between 21 149 human genes and 10 101 disease and phenotypes. The upcoming months will see the integration of new data from the Open Targets experimental programme including synthetic lethality data from CRISPR/Cas9 knockout screens in cancer cell lines, as well as the release of Open Targets Genetics (https://genetics.opentargets.org/), a new resource that combines GWAS and functional genomics data to prioritise likely causal variants at disease-associated loci. In summary, we will sustain and build on our efforts to date, and continue to provide the Open Targets Platform to facilitate drug target identification and prioritisation, and ultimately increase the odds of success in drug discovery.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank the members of the Open Targets consortium for stimulating discussions on the development of the Open Targets Platform, our data sources for providing evidence for our target-disease associations and the databases that provide us with annotations for our targets and diseases. We thank all our users, especially those who have taken the time to contact us with suggestions and other helpful comments. We greatly appreciate and acknowledge all who took part in our usability sessions of views and tools that were under development. We also thank Holly Foster for providing extensive administrative support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

FUNDING

Open Targets. Funding for open access charge: Open Targets.

Conflict of interest statement. None declared.

REFERENCES

1. Zhang,W., Zhang,H., Yang,H., Li,M., Xie,Z. and Li,W. (2018) Computational resources associating diseases with genotypes, phenotypes and exposures. *Brief. Bioinform.*, doi:10.1093/bib/bby071.
2. Cook,D., Brown,D., Alexander,R., March,R., Morgan,P., Satterthwaite,G. and Pangalos,M.N. (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.*, **13**, 419–431.
3. Koscielny,G., An,P., Carvalho-Silva,D., Cham,J.A., Fumis,L., Gasparyan,R., Hasan,S., Karamanis,N., Maguire,M., Papa,E. *et al.* (2017) Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.*, **45**, D985–D994.
4. Caulfield,M., Davies,J., Dennys,M., Elbahy,L., Fowler,T., Hill,S., Hubbard,T., Jostins,L., Maltby,N., Mahon-Pearson,J. *et al.* (2017) The 100,000 genomes project protocol. doi:10.6084/m9.figshare.4530893.v2.
5. Malone,J., Holloway,E., Adamusiak,T., Kapushesky,M., Zheng,J., Kolesnikov,N., Zhukova,A., Brazma,A. and Parkinson,H. (2010) Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics*, **26**, 1112–1118.
6. Chibucos,M.C., Mungall,C.J., Balakrishnan,R., Christie,K.R., Huntley,R.P., White,O., Blake,J.A., Lewis,S.E. and Giglio,M. (2014) Standardized description of scientific evidence using the Evidence Ontology (ECO). *Database*, **2014**, bau075.
7. Karamanis,N., Carvalho-Silva,D., Cham,J.A., Fumis,L., Hasan,S., Hulcoop,D., Koscielny,G., Maguire,M., Newell,W., Ong,C. *et al.* (2018) Designing an intuitive web application for drug discovery scientists. *Drug Discov Today*, **23**, 1169–1174.
8. Köhler,S., Vasilevsky,N.A., Engelstad,M., Foster,E., McMurry,J., Aymé,S., Baynam,G., Bello,S.M., Boerkoel,C.F., Boycott,K.M. *et al.*

(2017) The human phenotype ontology in 2017. *Nucleic Acids Res.*, **45**, D865–D876.

9. Denny,J.C., Bastarache,L., Ritchie,M.D., Carroll,R.J., Zink,R., Mosley,J.D., Field,J.R., Pulley,J.M., Ramirez,A.H., Bowton,E. *et al.* (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.*, **31**, 1102–1110.
10. Iorio,F., Garcia-Alonso,L., Brammeld,J.S., Martincorena,I., Wille,D.R., McDermott,U. and Saez-Rodriguez,J. (2018) Pathway-based dissection of the genomic heterogeneity of cancer hallmarks' acquisition with SLAPenrich. *Sci. Rep.*, **8**, 6713–6728.
11. Schubert,M., Klinger,B., Klünemann,M., Sieber,A., Uhlitz,F., Sauer,S., Garnett,M.J., Blüthgen,N. and Saez-Rodriguez,J. (2018) Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.*, **9**, 20–30.
12. MacArthur,J., Bowler,E., Cerezo,M., Gil,L., Hall,P., Hastings,E., Junkins,H., McMahon,A., Milano,A., Morales,J. *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.
13. Sondka,Z., Bamford,S., Cole,C.G., Ward,S.A., Dunham,I. and Forbes,S.A. (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer*, **18**, 696–705.
14. Futreal,P.A., Andrew Futreal,P., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
15. Landrum,M.J., Lee,J.M., Benson,M., Brown,G.R., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Jang,W. *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
16. Papatheodorou,I., Fonseca,N.A., Keays,M., Tang,Y.A., Barrera,E., Bazant,W., Burke,M., Füllgrabe,A., Fuentes,A.M.-P., George,N. *et al.* (2018) Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.*, **46**, D246–D251.
17. Cortes,A. and Brown,M.A. (2011) Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.*, **13**, 101–103.
18. Voight,B.F., Kang,H.M., Ding,J., Palmer,C.D., Sidore,C., Chines,P.S., Burtt,N.P., Fuchsberger,C., Li,Y., Erdmann,J. *et al.* (2012) The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.*, **8**, e1002793.
19. Bradley,A.R., Echalier,A., Fairhead,M., Strain-Damerell,C., Brennan,P., Bullock,A.N., Burgess-Brown,N.A., Carpenter,E.P., Gileadi,O., Marsden,B.D. *et al.* (2017) The SGC beyond structural genomics: redefining the role of 3D structures by coupling genomic stratification with fragment-based discovery. *Essays Biochem.*, **61**, 495–503.
20. Arrowsmith,C.H., Audia,J.E., Austin,C., Baell,J., Bennett,J., Blagg,J., Bountra,C., Brennan,P.E., Brown,P.J., Bunnage,M.E. *et al.* (2015) The promise and peril of chemical probes. *Nat. Chem. Biol.*, **11**, 536–541.
21. Antolin,A.A., Tym,J.E., Komianou,A., Collins,I., Workman,P. and Al-Lazikani,B. (2018) Objective, Quantitative, Data-DrivenAssessment of Chemical Probes. *Cell Chem. Biol.*, **25**, 194–205.
22. Türei,D., Korcsmáros,T. and Saez-Rodriguez,J. (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods*, **13**, 966–967.
23. Smith,C.L., Blake,J.A., Kadin,J.A., Richardson,J.E., Bult,C.J. and the Mouse Genome Database Group (2017) Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res.*, **46**, D836–D842.
24. Hanahan,D. and Weinberg,R.A. (2000) The hallmarks of cancer. *Cell*, **100**, 57–70.
25. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
26. Forbes,S.A., Beare,D., Boutselakis,H., Bamford,S., Bindal,N., Tate,J., Cole,C.G., Ward,S., Dawson,E., Ponting,L. *et al.* (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.*, **45**, D777–D783.
27. Tamborero,D., Rubio-Perez,C., Deu-Pons,J., Schroeder,M.P., Vivancos,A., Rovira,A., Tusquets,I., Albanell,J., Rodon,J., Tabernero,J. *et al.* (2018) Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.*, **10**, 25–32.
28. Indyk,P., Motwani,R., Raghavan,P. and Vempala,S. (1997) Locality-preserving hashing in multidimensional spaces. In: *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing - STOC '97*. ACM Press, NY, pp. 618–625.
29. Samet,H. (2006) *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, Burlington.
30. Chin,A. (1994) Locality-preserving hash functions for general purpose parallel computation. *Algorithmica*, **12**, 170–181.
31. Uhlén,M., Björling,E., Agaton,C., Szigyarto,C.A.-K., Amini,B., Andersen,E., Andersson,A.-C., Angelidou,P., Asplund,A., Asplund,C. *et al.* (2005) A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Mol. Cell. Proteomics*, **4**, 1920–1932.
32. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, tatistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
33. Levchenko,M., Gou,Y., Graef,F., Hamelers,A., Huang,Z., Ide-Smith,M., Iyer,A., Kilian,O., Katuri,J., Kim,J.-H. *et al.* (2018) Europe PMC in 2017. *Nucleic Acids Res.*, **46**, D1254–D1260.
34. NCBI Resource Coordinators (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **46**, D8–D13.
35. Gaulton,A., Hersey,A., Nowotka,M., Bento,A.P., Chambers,J., Mendez,D., Mutowo,P., Atkinson,F., Bellis,L.J., Cibrián-Uhalte,E. *et al.* (2017) The ChEMBL database in 2017. *Nucleic Acids Res.*, **45**, D945–D954.
36. Honnibal,M. and Johnson,M. (2015) An Improved Non-monotonic Transition System for Dependency Parsing. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, pp. 1373–1378.
37. Pinero,J., Queralt-Rosinach,N., Bravo,A., Deu-Pons,J., Bauer-Mehren,A., Baron,M., Sanz,F. and Furlong,L.I. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, **2015**, bav028.
38. Piñero,J., Bravo,À., Queralt-Rosinach,N., Gutiérrez-Sacristán,A., Deu-Pons,J., Centeno,E., García-García,J., Sanz,F. and Furlong,L.I. (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.*, **45**, D833–D839.
39. Nguyen,D.-T., Mathias,S., Bologa,C., Brunak,S., Fernandez,N., Gaulton,A., Hersey,A., Holmes,J., Jensen,L.J., Karlsson,A. *et al.* (2017) Pharos: collating protein information to shed light on the druggable genome. *Nucleic Acids Res.*, **45**, D995–D1002.
40. Oprea,T.I., Bologa,C.G., Brunak,S., Campbell,A., Gan,G.N., Gaulton,A., Gomez,S.M., Guha,R., Hersey,A., Holmes,J. *et al.* (2018) Unexplored therapeutic opportunities in the human genome. *Nat. Rev. Drug Discov.*, **17**, 317–332.