

Open Targets:

An Innovative Public-Private Partnership

<https://www.opentargets.org/>

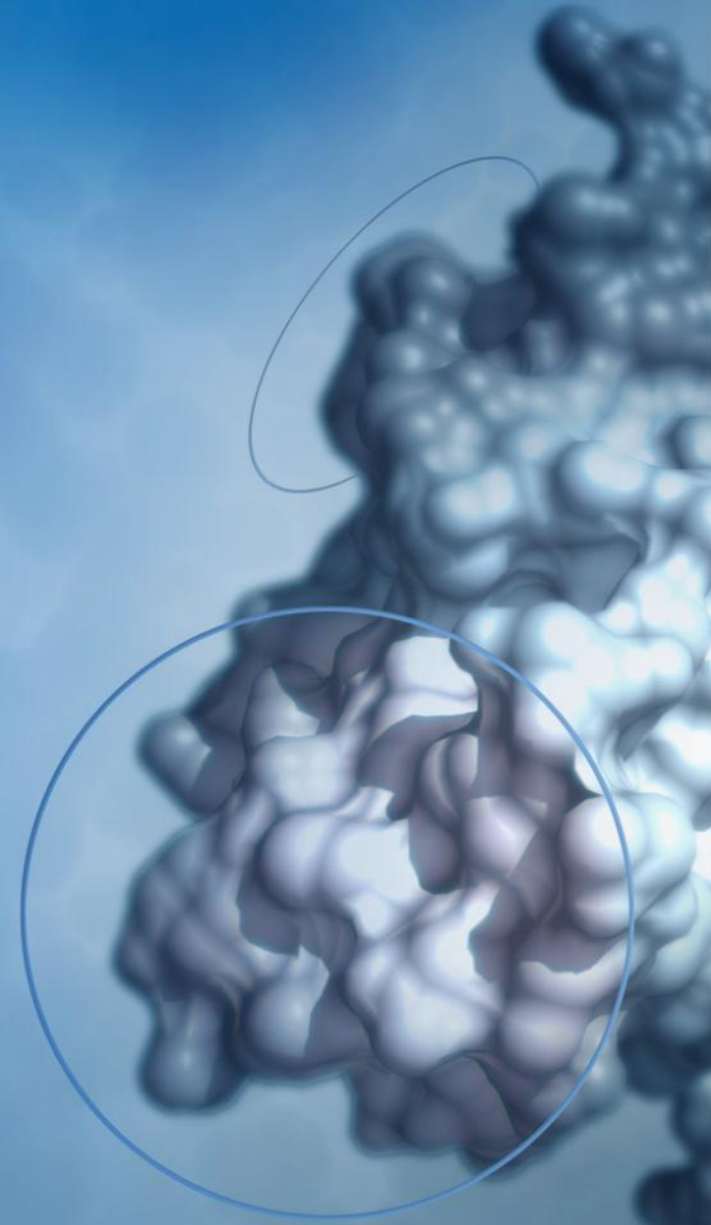
Philippe Sanseau

GSK, Target Sciences

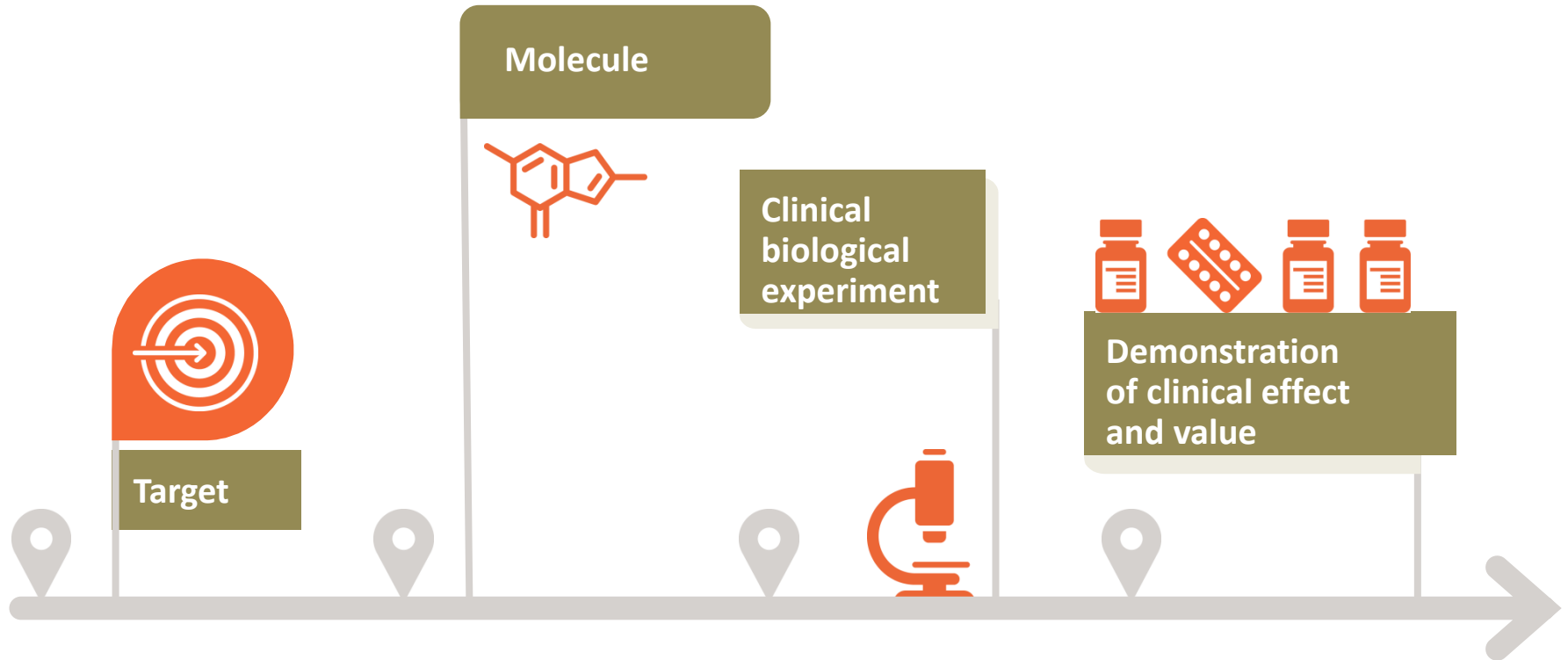
20th June 2017



Open Targets



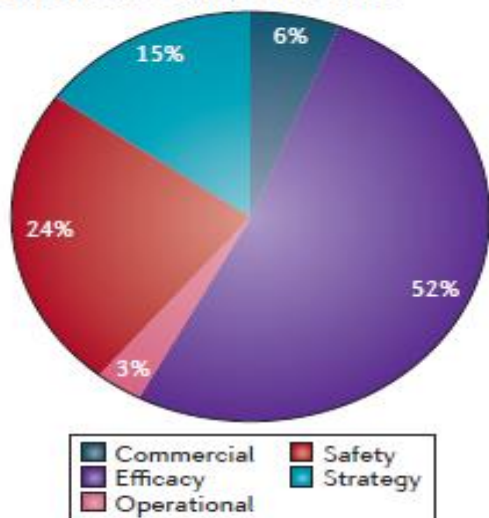
There are four big decisions in making medicines



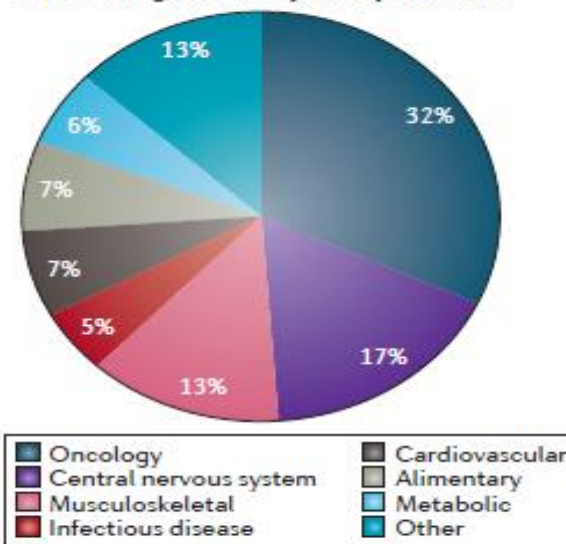
.....and it costs over \$1bn and more than a decade

Clinical failures are costly

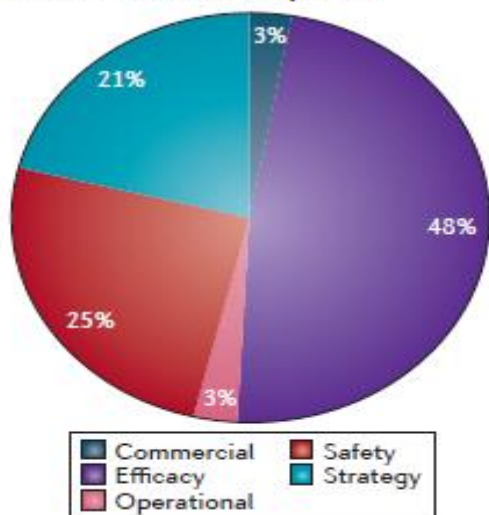
a Reason for failure 2013–2015



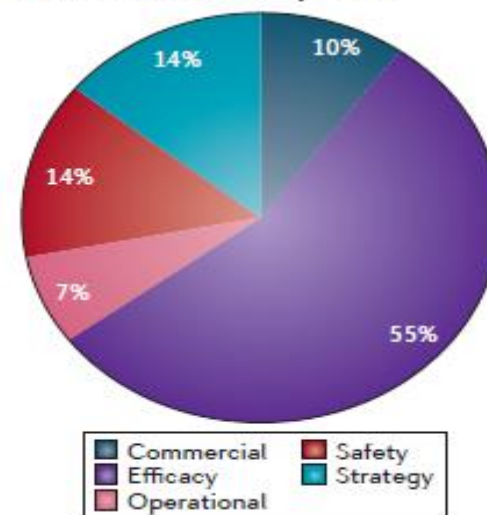
b Percentage failure by therapeutic area



c Reason for failure in phase II



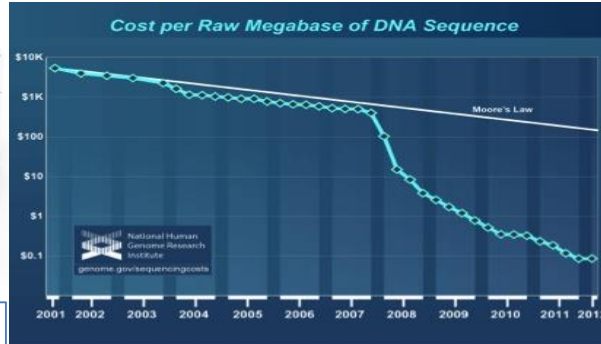
d Reason for failure in phase III



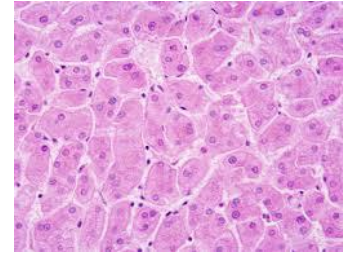
The right time?



Technology



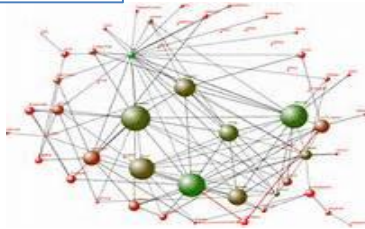
Gene editing



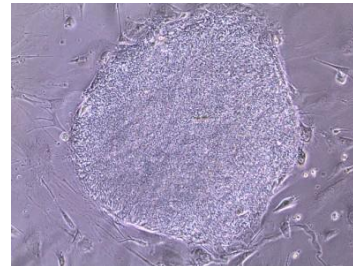
Human tissues



Genetic associations



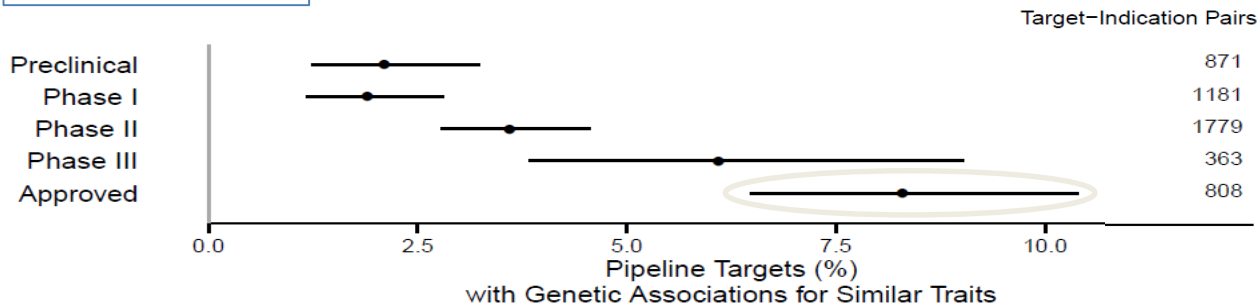
Informatics



iPS cells



ENCODE data & genomics technologies



Nelson, Nat Genet, 2015

The Vision



A public-private initiative to transform drug discovery by enabling the systematic identification and prioritisation of targets



*Professor Dame
Janet Thornton
former Director, EMBL-EBI*




*Patrick Vallance, President
Pharmaceuticals R&D
GlaxoSmithKline*



*Professor Sir
Mike Stratton
Director, Sanger Institute*

Core Principles

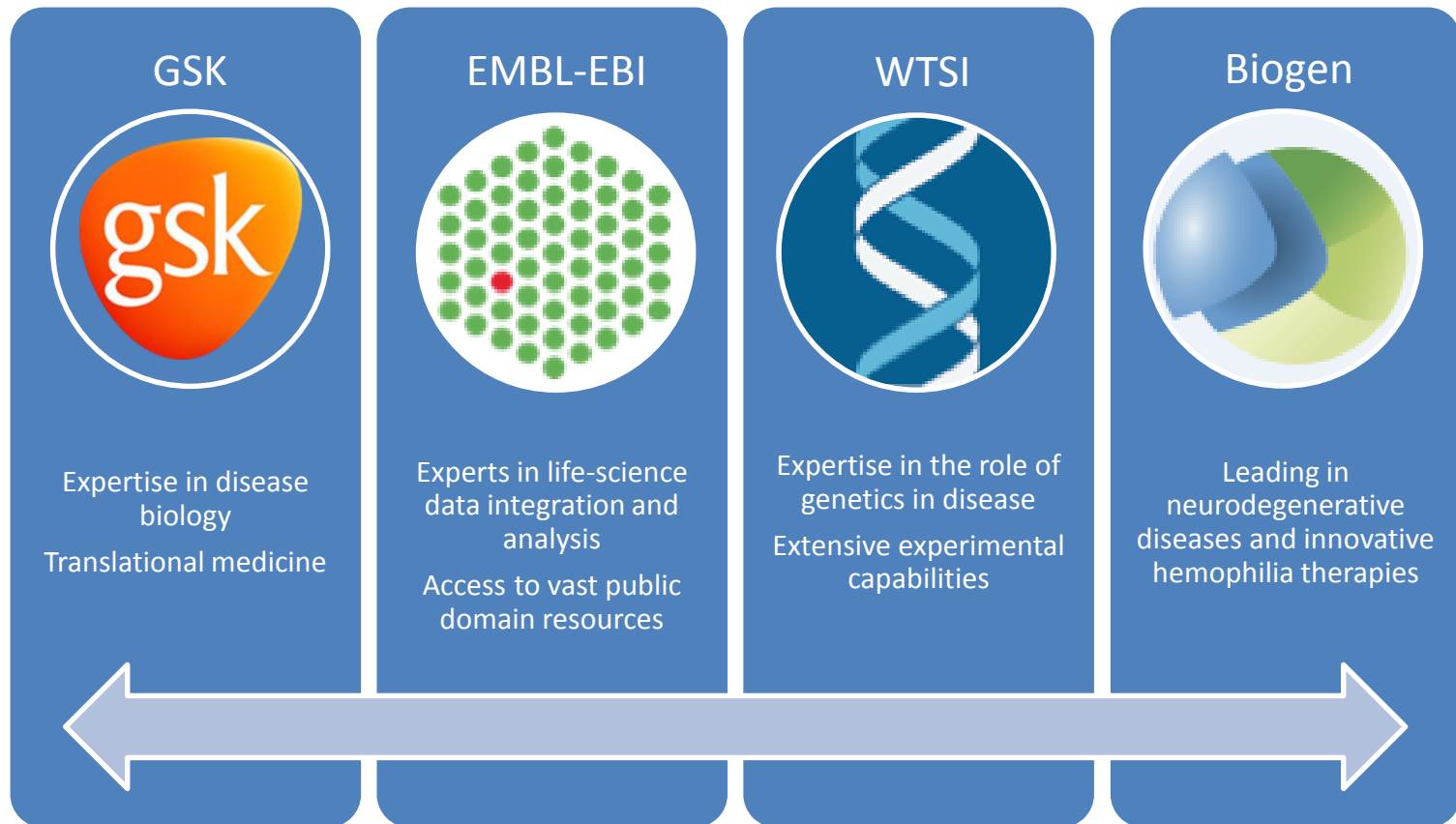


We are focused on **pre-competitive research** that will enable the systematic identification and prioritisation of targets.

We are committed to **rapid publication** and making data, methods and results **publically available** as soon as possible.

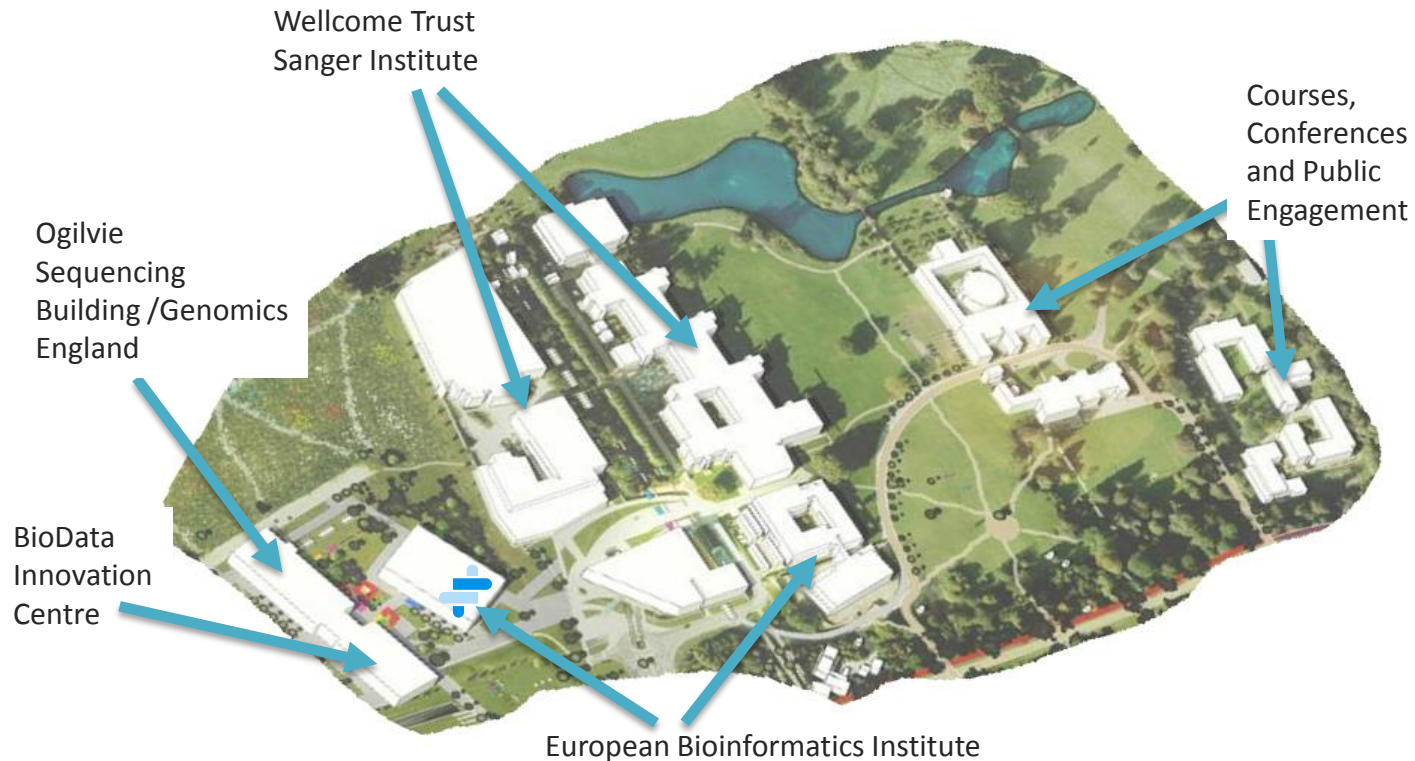
We believe in non-exclusive **partnerships** that foster the free exchange of ideas and expertise.

The Partners



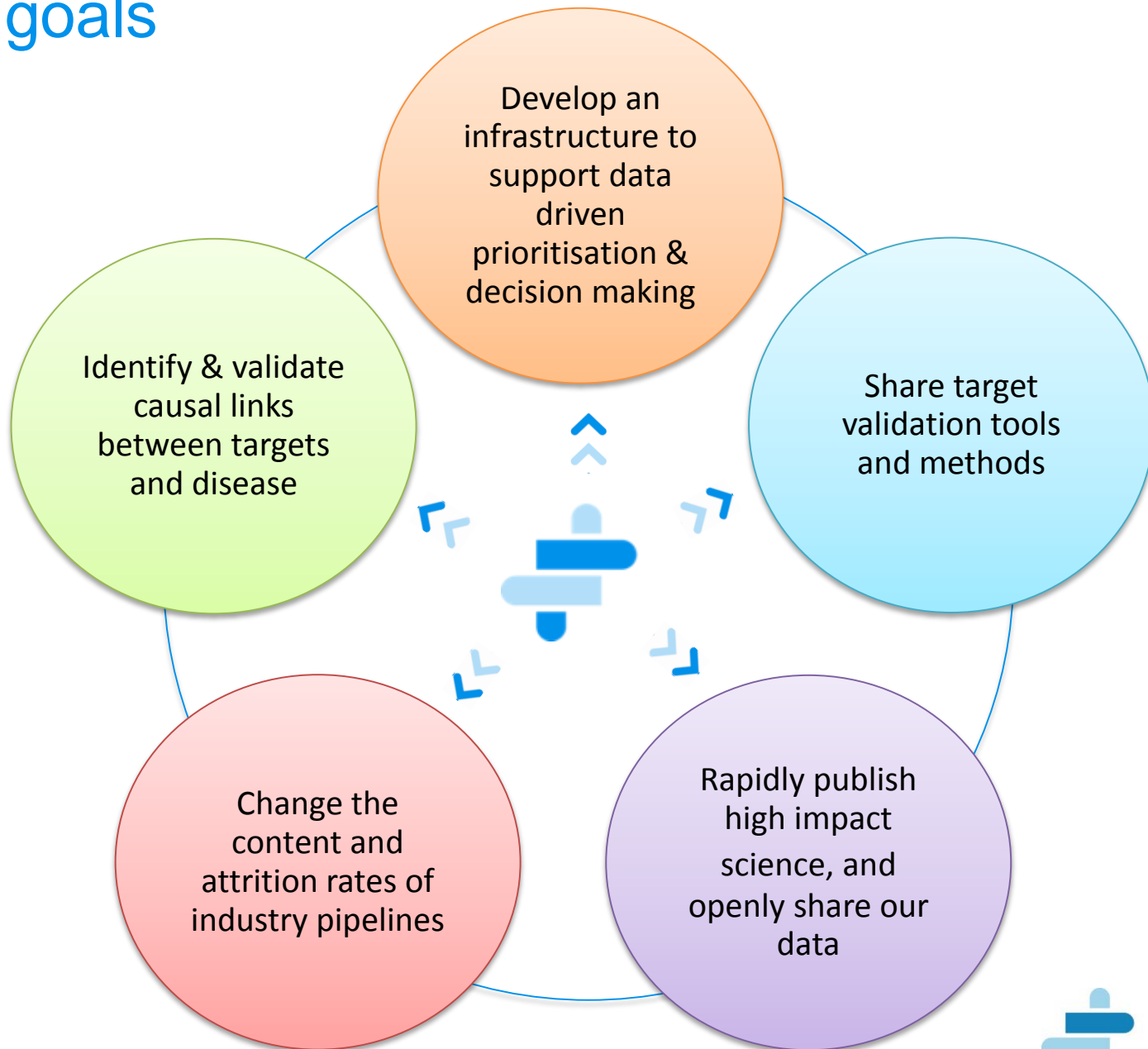
The partners shared the idea that target validation could be improved but that one institute could not necessarily do it alone.

Based on the Wellcome Genome Campus




- The Campus - one of the world's largest genomic research centres (home to over 2000 genomic scientists; hosts of major international conferences; situated within one of the world's top innovation districts (Cambridge, UK))
- Location enables close collaboration and leverages existing campus expertise and synergies - integrating scientists specialising in genomics, drug discovery and bioinformatics in one centre to focus on major challenges in target validation.

Our goals

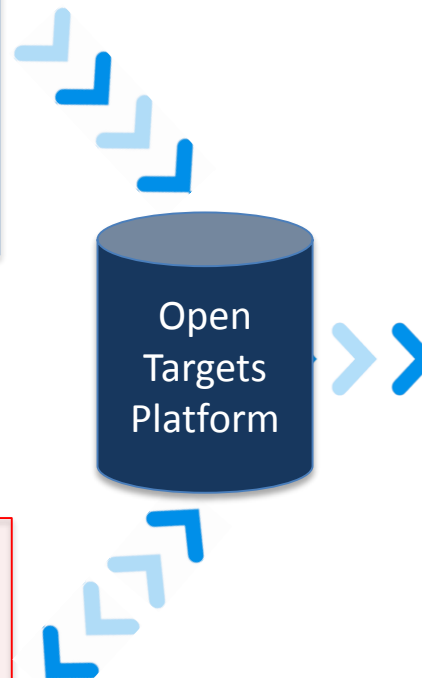


Target Validation Knowledge Cycle

Public Databases and Pipelines



ChEMBL UniProt
Europe PMC REACTOME e!
IntAct intOGen Integrative Onco Genomics G2P



www.targetvalidation.org

Open Targets About Help API Downloads Blog Search for a target or disease (e.g. BRAF or BRCA1)

384 diseases associated with BRAF
View BRAF profile

Filter by

Data types

- Genetic associations (15)
- Genetic interactions (14)
- Drugs (3)
- Affected pathways (2)
- Web resources (2)
- Taxonomy (2)
- Anter nodes (1)

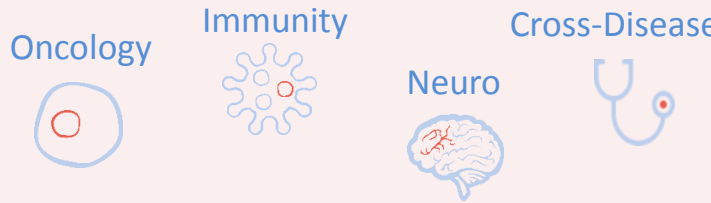
Therapeutic areas

- Neoplasms (12)
- Genetic disorder (8)
- Skin disease (7)
- Endocrine system disease (2)
- Immunological system disease (2)
- Nervous system disease (2)
- Digestive system disease (2)
- Cardiovascular disease (2)
- Eye disease (2)
- Reproductive system disease (2)
- Phenotype (2)
- Respiratory system disease (2)
- Immune system disease (1)
- Diarrhoeal system disease (1)
- Mutability disease (1)

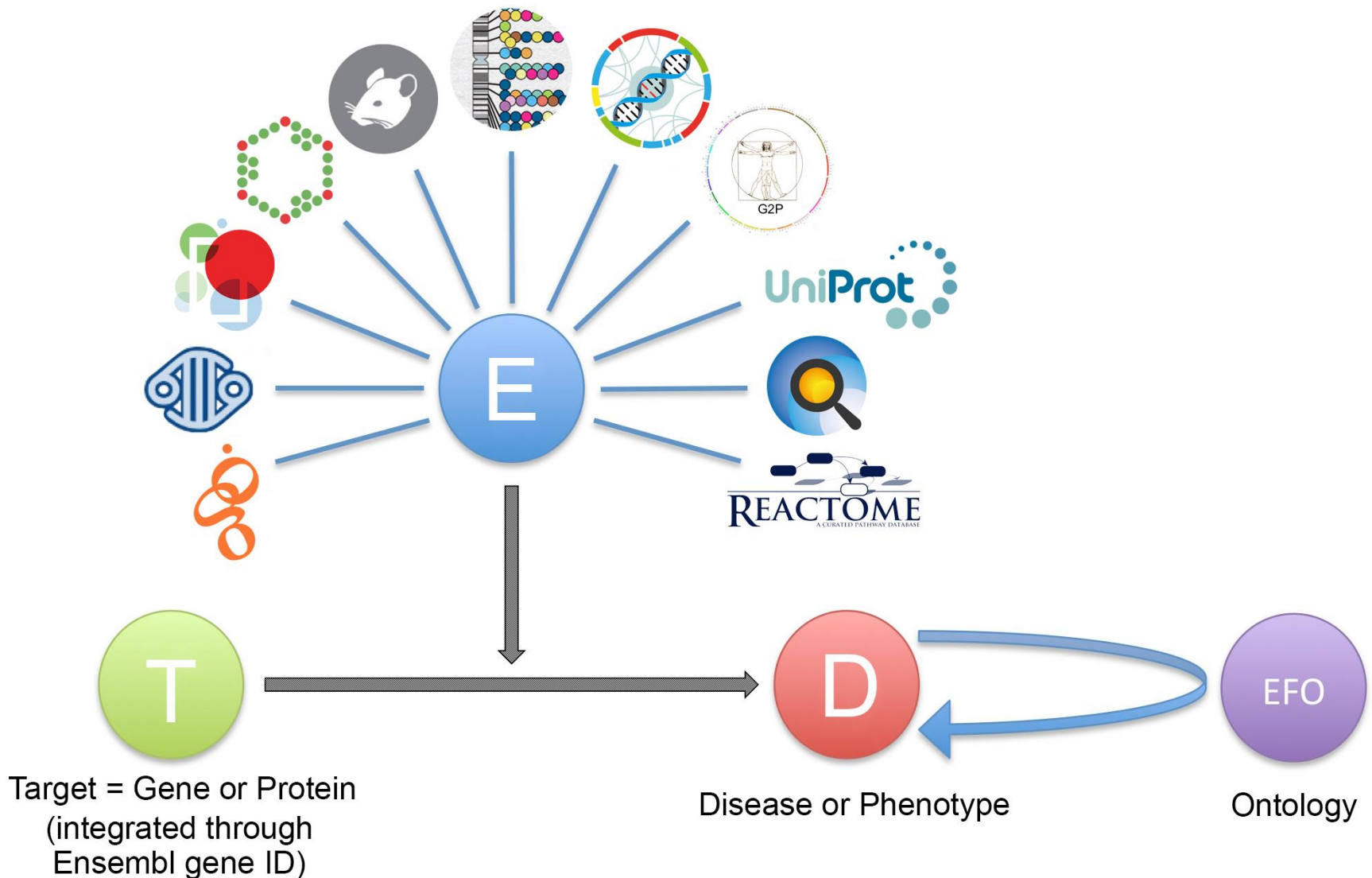


New experimental data
Physiologically relevant and at scale

Oncology Immunity Cross-Disease
Neuro



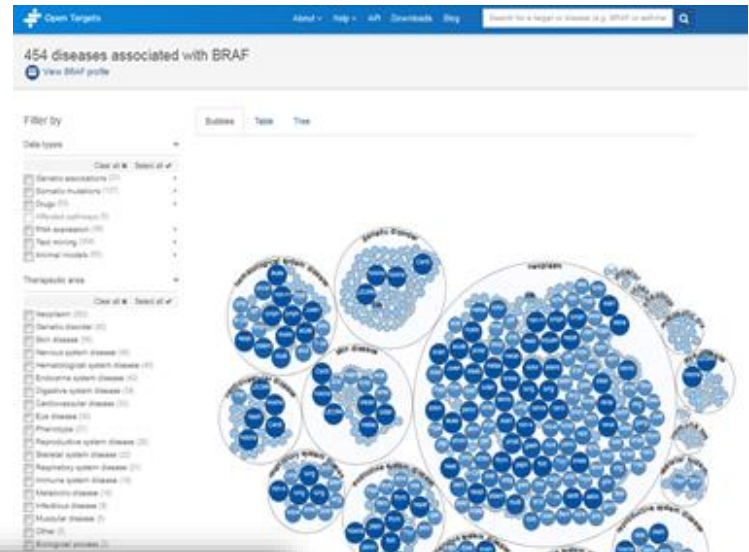
The Evidence Model



Using a User Experience Approach



Putting the User first



not clear what these are (circled)

"type of line describes connection. This one is solid, so there is a connection"

DISEASE → Phenotypes

Network of genes — gene classes? Associations

CONFIDENCE → [genetic] interaction for a human

abstract ... gene entry

more info. (expand)

both search parameters "active" at once

doesn't make sense for this to be a network of other genes ...

CONFUSED (not had to leave early)

anything that was in the "Filters" should be represented here

"Is it really useful to have microarray data when you've already zoomed in on a disease?"

"would try to download data and analyse it myself"

TP2

Epigenetic data → expression data (connection)

"Network"

Colour — where different data sources have relationships

Short List. links to literature

or ... overlapping bits could relate to COPD. other areas are where this gene is involved in other things

positive evidence from lots of data

raw data (papers usually from one source only)

• Prominently targeted (plasma data)

• Clinical trials

DATA SOURCE X: Reference of gene to disease

DATA SOURCE Y: GWAS data (individual SNPs) or phenotypic data

DATA SOURCE Z: Gene Expression data (comparisons) smok./cancer

Summary of the relationship of link to original study

Three sources of data: GWAS, omics, literature

Summary of a phenotypic link to original GWAS study

2 -1 +1 +2

cOPD smoker ex-smoker

Facilitating decision-making

H) Can I look at multiple targets at a time?

A) Which targets are associated with a disease?

B) What evidence supports this target-disease association?

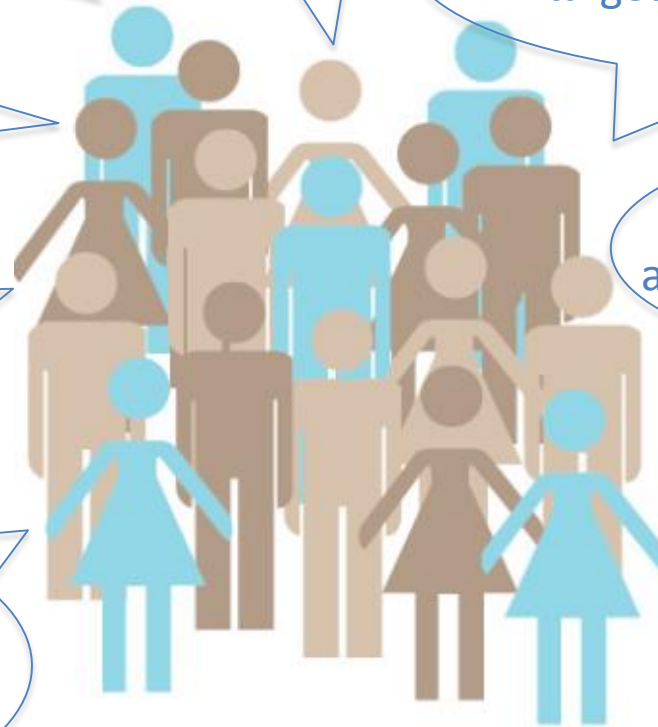
G) What else can I find out about my disease?

C) Are there drugs for this association in clinical trials?

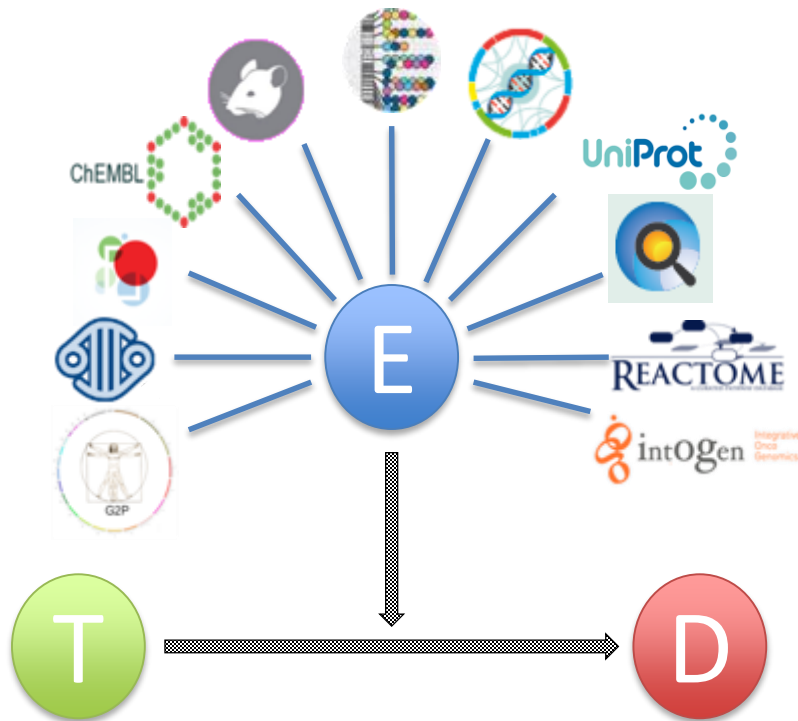
F) What else can I find out about my drug target?

E) Can I find the association focusing on two (or more) different therapeutic areas?

D) For a given target, are there other diseases associated with it?



Current status



Releases

Dec 2015
 Apr 2016
 Aug 2016
 Dec 2016
 Mar 2017

Evidence Type	Source	Target-Disease Association Objects
Marketed drugs	ChEMBL	120,520
Affected pathways	Reactome	6,143
Somatic mutations	Cancer Gene Census	23,440
Proteins with disease roles	Uniprot	21,870
Genetic for rare diseases	Gene2Phenotype	975
Rare genetic mutations	European Variation Archive	28,050
Complex disease	GWAS Catalog	32,363
Cancer driver mutations	IntOGen	2,377
Mouse models	Phenodigm	395,622
RNA expression	ArrayAtlas	529,084
Text mining	EuropePMC	3,678,967

The Open Targets Platform



Which other diseases are associated with PDE4D?

Diseases associated with PDE4D

Found 125 diseases | [View PDE4D profile](#)

[View more information about PDE4D](#)

Filter by

Datatypes

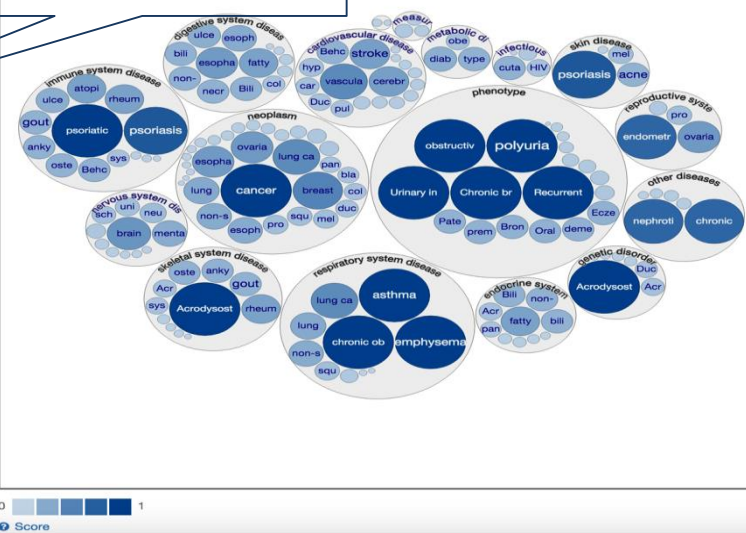
- Clear all Select all
- Genetic associations (19)
 - Somatic mutations (0)
 - Drugs (55)
 - Affected pathways (0)
 - RNA expression (33)
 - Text mining (57)
 - Animal models (1)

Therapeutic Areas

- Neoplasm (31)
- Phenotype (19)
- Cardiovascular disease (17)
- Digestive system disease (13)
- Immune system disease (13)
- Nervous system disease (11)
- Skeletal system disease (11)
- Respiratory system disease (10)
- Endocrine system disease (10)
- Genetic disorder (7)
- Other diseases (6)
- Reproductive system disease (4)
- Skin disease (4)
- Infectious disease (4)
- Metabolic disease (3)
- Measurement (3)
- Eye disease (2)

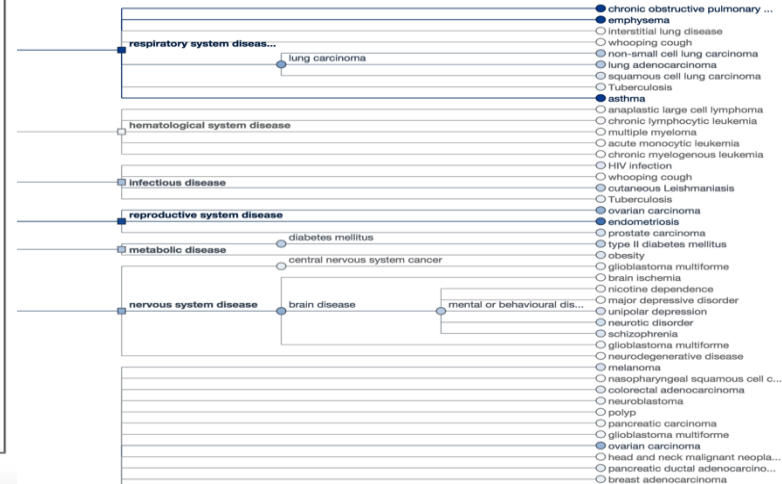
Bubbles Table Tree

Filter by therapeutic area



View diseases grouped in therapeutic areas or organised in a tree

Table Tree



Which targets are associated with breast carcinoma?

Filter by evidence type: e.g. 237 targets are supported by somatic mutations

Filter by pathway type and particular pathways

Targets associated with breast carcinoma
Found 17422 targets | [View disease profile](#)

Filter by

Datatypes

Clear all ✕ Select all ✓

- Genetic associations (179)
- Somatic mutations (237)
- Drugs (181)
- Affected pathways (3)
- RNA expression (16k)
- Text mining (6k)
- Animal models (7)

Pathway type

Clear all ✕ Select all ✓

- Signal Transduction (2k)
- Metabolism (1k)
- Immune System (1k)
- Gene Expression (1k)
- Metabolism of proteins (683)
- Developmental Biology (647)
- Disease (635)
- Cell Cycle (502)
- Transmembrane transport of s... (478)
- Hemostasis (449)
- Cellular responses to stress (317)
- Organelle biogenesis and mal... (277)
- Extracellular matrix organization (247)
- Vesicle-mediated transport (210)
- Neuronal System (203)

Showing 1 to 50 of 17,422 targets
Search:

Filter by target symbol

Order by evidence type and evidence strength

	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	
ESR1									estrogen receptor 1
AKT1									v-akt murine thymoma viral onc...
ERBB2									erb-b2 receptor tyrosine kinase 2
TERT									telomerase reverse transcriptase
EGFR									epidermal growth factor receptor
BRIP1									BRCA1 interacting protein C-ter...
BRCA1									breast cancer 1, early onset
CDK4									cyclin-dependent kinase 4
BRCA2									breast cancer 2, early onset
PIK3CA									phosphatidylinositol-4,5-bispho...
CDK6									cyclin-dependent kinase 6
RET									ret proto-oncogene
NBN									nibrin
KIT									v-kit Hardy-Zuckerman 4 feline ...
EBF1									early B-cell factor 1
FGFR2									fibroblast growth factor receptor 2
CHEK2									checkpoint kinase 2
KLF4									Kruppel-like factor 4 (gut)
HIST1H3B									histone cluster 1, H3b
CDH11									cadherin 11, type 2, OB-cadheri...
GATA3									GATA binding protein 3
POLE									polymerase (DNA directed), epsi...
ERBB3									erb-b2 receptor tyrosine kinase 3

A comprehensive and intuitive resource

3076 targets associated with asthma

[View disease profile](#)

Filter by

Data types

- Clear all x Select all v
- Genetic associations (330)
 - Somatic mutations (0)
 - Drugs (72)
 - Affected pathways (0)
 - RNA expression (1k)
 - Text mining (2k)
 - Animal models (1)

Pathway types

- Clear all x Select all v
- Signal Transduction (614)
 - Immune System (581)
 - Metabolism (397)
 - Disease (234)
 - Hemostasis (212)
 - Developmental Biology (199)
 - Metabolism of proteins (197)
 - Gene Expression (193)
 - Transmembrane transport o... (112)
 - Extracellular matrix organiz... (103)
 - Cellular responses to stress (82)
 - Cell Cycle (63)
 - Vesicle-mediated transport (57)
 - Programmed Cell Death (52)
 - Neuronal System (47)
 - Cell-Cell communication (41)
 - DNA Repair (36)
 - Muscle contraction (34)
 - Organelle biogenesis and ... (32)
 - Chromatin organization (25)

Showing 1 to 50 of 3,076 targets

Search:

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining
PDE4D							
ORMDL3							
PTGDR							
PTGS2							
ALOX5							
IL5							
CHI3L1							
ADRB2							
ADAM33							
NR3C1							
NPSR1							
PLA2G7							
PTGS1							
VDR							
CYSLTR1							
MUC7							
DPP10							
IRAK3							
HRH1							
ADRB1							
CHRM3							
DENND1B							
PDE4B							
ADRA1B							
IGHE							
ADRB3							
ADORA2B							
ADORA2A							
PDE4A							
CHRNB2							
ADRA1A							
ADRA1D							
ADORA3							
PDE3A							
PDE4C							
ATP4A							
ATP4B							
PDE3B							
CHRNA4							
ADORA2C							
ADORA2A							
ADORA2B							
ADORA1							
PTGDR2							
GSDMB							
HMSCR							
GSDMA							

Drug Discovery Head

Nick is a business decision maker. He is concerned about the attrition rate prior to Phase III & believes correct target selection could be a key factor in addressing this. He would like to have more concrete target-disease evidence to include in his 3-year business plan; more evidence means less risk and increased confidence in target-disease associations.

Nick



Pulls lots of information together in an intuitive interface!

Laura



Drug Discovery Manager

Laura is interested in finding the best target for a specific disease area. She is concerned about the high failure rate in identifying targets for her team's disease area. Her team aims to investigate gaps & risks in knowledge about a particular target. Laura needs up-to-date information on several targets. She currently uses PubMed, some public databases and other in-house analysis pipelines and tools for investigating targets.

Powerful resource, clear links and easy to use without training, especially for a non-bioinformatician!

Derek



Investigator

Derek explores target space to find out what is already known about a target's role in his disease area of interest. Derek works in a matrix (multidisciplinary) team with scientists from genetics, computational biology, disease experts and clinical. He likes to work on a case-by-case examples of specific targets. It is not a large-scale approach, but rather detailed and focussed. Typically he accesses information by PubMed and uses Excel spreadsheets to analyse the data he collects. Derek occasionally uses bioinformatics database websites such as OMIM, UniProt, ChEMBL and Orphanet.

Lots of data in one place so saves time going to different websites. It's easy to go into more depth if needed as the data is all there!



Open Targets

Platform Development Acknowledgements



GWAS Catalog



Atlas



ChEMBL



Web production



Phenodigm



COSMIC
Catalogue of somatic mutations in cancer



WEB Development



EBI
Embassy
Cloud

EBI
Database
Admin



Statistical Integration
Oliver Stegle,
Naruemon (Ploy)
Pratanwanich, Verena
Zuber



Jenny Cham,
Francis Rowland
Bren Vaughan

Core platform development team



Andrea
Pierleoni



ChuangKee
Ong



Denise
Carvalho-Silva



Eliseo
Papa



Gareth
Peat



Gautier
Koscielny



Ian
Dunham



Luca
Fumis



Miguel
Carmona



Miguel
Pignatelli



Naveen
Kumar



Niki
Karamanis



Priyanka
Wankar



William
Newell



Open Targets

Jeff Barrett, Ewan
Birney, Philippe
Sanseau, Jessica
Vamathevan, David
Hulcoop, Holly Foster

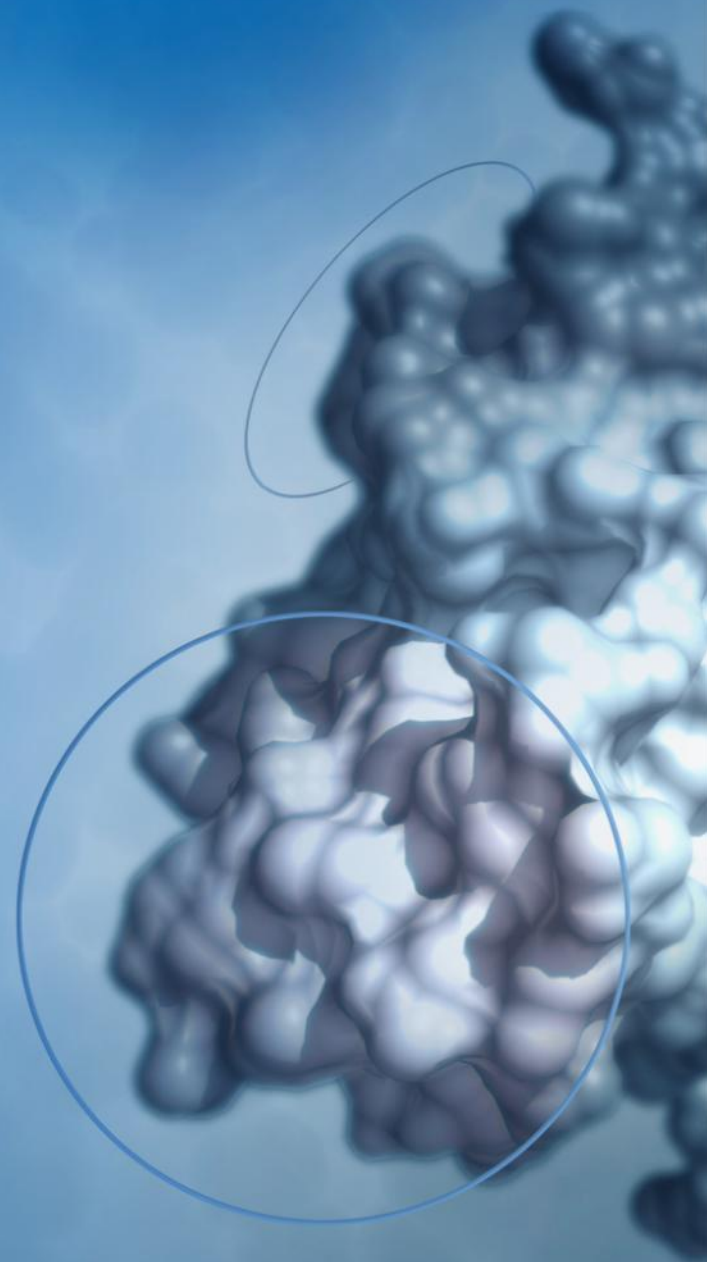


100 GSK beta
testers



Open Targets

Thank you



Open Targets

Nucleic Acids Research Database 2017

Open targets: a platform for therapeutic target identification and validation, NAR Database 2017, In Press

Open targets: a platform for therapeutic target identification and validation

Gautier **Koscielny**^{1,2,1}, Peter **An**^{1,3}, Denise **Carvalho-Silva**^{1,4}, Jennifer A. **Cham**^{1,4}, Alfonso **Munoz-Pomer Fuentes**^{1,4}, Luca **Fumis**^{1,4}, Rippa **Gasparyan**^{1,3}, Samiul **Hasan**^{1,2}, Nikiforos **Karamanis**^{1,4}, Michael **Maguire**^{1,4}, Eliseo **Papa**^{1,2}, Andrea **Pierleoni**^{1,4}, Miguel **Pignatelli**^{1,4}, Theo **Platt**^{1,3}, Francis **Rowland**^{1,4}, Priyanka **Wankar**^{1,3}, A. Patricia **Bento**^{1,4}, Tony **Burdett**^{1,4}, Antonio **Fabregat**^{1,4}, Simon **Forbes**^{1,5}, Anna **Gaulton**^{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}, Claire **O'Donovan**^{1,4}, Irene **Papathodorou**^{1,4}, Helen **Parkinson**^{1,4}, Barbara **Palka**^{1,4}, Justin **Paschall**^{1,4}, Robert **Petryszak**^{1,4}, Naruemon **Pratanwanich**^{1,4}, Sirarat **Santivijjal**^{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}, Cristina Yenyxe **Gonzalez**^{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**^{1,4,7}

¹Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, ²GSK, Medicines Research Center, Gunnels Wood Road, Stevenage, SG1 2NY, UK, ³Biogen, Cambridge, MA 02142, USA, ⁴European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, ⁵Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK and ⁶National 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

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 Plat-

form 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

¹To 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, Deodar Road, Cambridge, CB21 4QR, UK.

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



Comprehensive, robust data integration



www.targetvalidation.org

Open Targets About Help API Downloads Blog Search for a target or disease (e.g. BRAF or asthma)

118 diseases associated with ITGA4
View ITGA4 profile

Filter by Bubbles Table Tree

Data types

- Genetic associations (5)
- Somatic mutations (0)
- Drugs (14)
- Affected pathways (0)
- RNA expression (4)
- Text mining (7)
- Animal models (0)

Therapeutic areas

- Neoplasm (45)
- Immune system disease (35)
- Genetic disorder (15)
- Hematological system disease (14)
- Digestive system disease (13)
- Respiratory system disease (12)
- Nervous system disease (10)
- Infectious disease (8)
- Reproductive system disease (7)
- Skin disease (7)
- Phenotype (8)
- Endocrine system disease (4)
- Measurement (4)
- Eye disease (4)
- Metabolic disease (3)
- Skeletal system disease (3)
- Cardiovascular disease (1)
- Biological process (1)
- Other (1)

United data model, pathways, statistical integration