

# Chemical biology resources at EMBL-EBI, 2016-2024

**Andrew Leach**

**Formerly**

Head of Chemical Biology and  
Head of Industry Partnerships



None of the following would be possible without an amazing team



# EBI's chemical biology resources



2004

ChEBI  
~60K structures



GLOBAL  
CORE  
BIODATA  
RESOURCE



2009

ChEMBL  
>2m structures



GLOBAL  
CORE  
BIODATA  
RESOURCE



2013

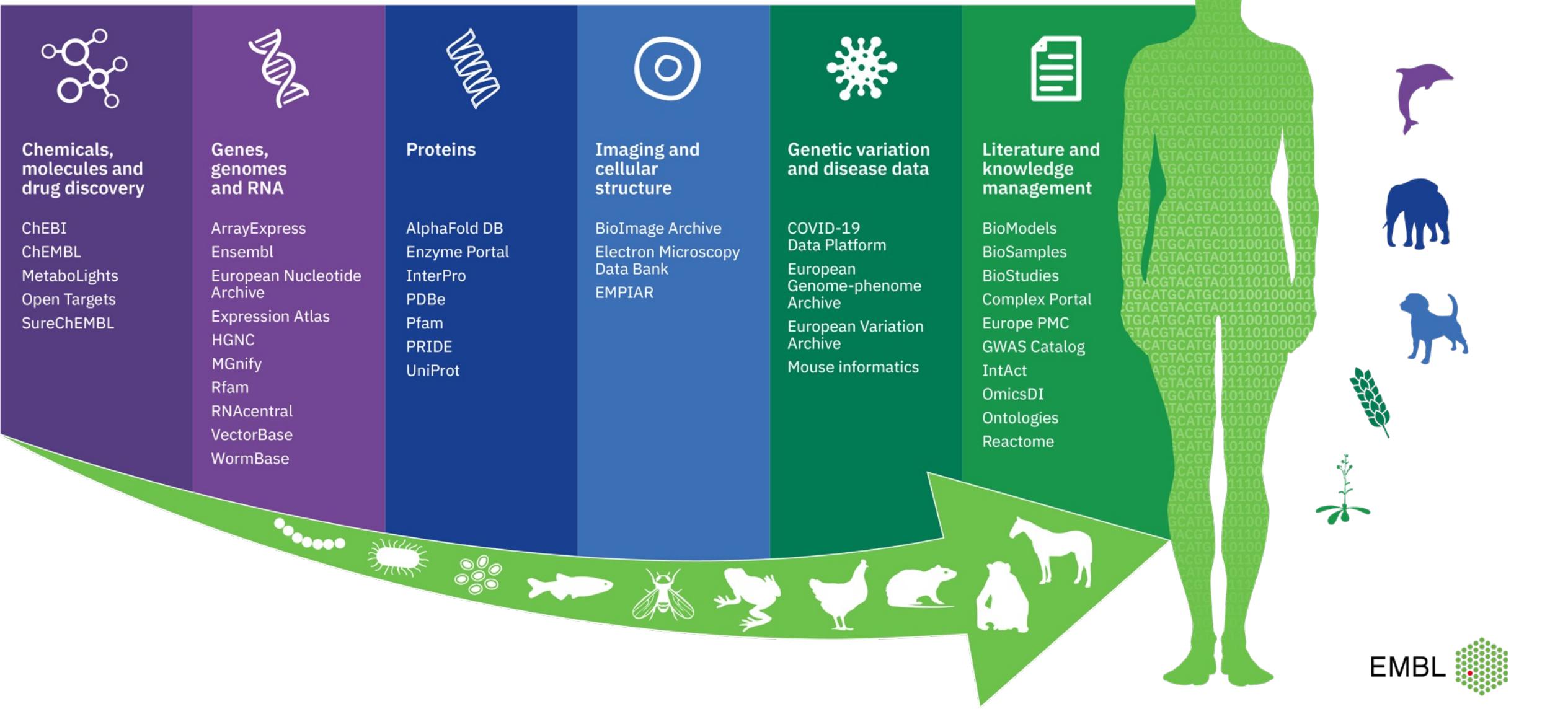
UniChem  
Links ~40 resources  
>175m structures



2014

SureChEMBL  
>20m structures

# Multiple EBI resources include or reference small molecules



# A few “headline” achievements

>50 Team members

£10m in new funding

Significant growth in data resources

>40 publications

GBC and Elixir accreditation for ChEMBL & ChEBI

Major software overhaul and rewrite

Two service reviews, one research review

Open Targets impact incl. tractability and safety

Multiple collaborations

Survived 1 pandemic

# Software infrastructure: legacy

	UniChem	ChEMBL	ChEBI	SureChEMBL
Frontend	Vanilla JS, no framework used 	BACKBONEJS 	Vanilla JS, no framework used 	Vanilla JS, no framework used 
Backend	 			  
Database and search engine		elastic  ORACLE PostgreSQL	Solr 	  Solr PostgreSQL

# Software infrastructure: future

	UniChem	ChEMBL	ChEBI	SureChEMBL
Frontend	• Sharing expertise and components between projects			
Backend	• Sharing expertise and components between projects			
Database and search engine	• Sharing expertise and components between projects	 elastic	 PostgreSQL	 Solr  PostgreSQL

# New interfaces for all our resources

This screenshot shows the ChEMBL Targets search results. At the top, there's a survey prompt: "Do data resources managed by EMBL-EBI and our collaborators make a difference to your work? Please take 10 minutes to fill in our annual user survey, and help us make the case for why sustaining open data resources is critical for life sciences research." Below this, the search results are displayed under the heading "Targets". The results table includes columns for ID, Name, UniProt Accessions, Type, and Selectivity Group. There are also filters on the left for various taxonomic levels and protein types.

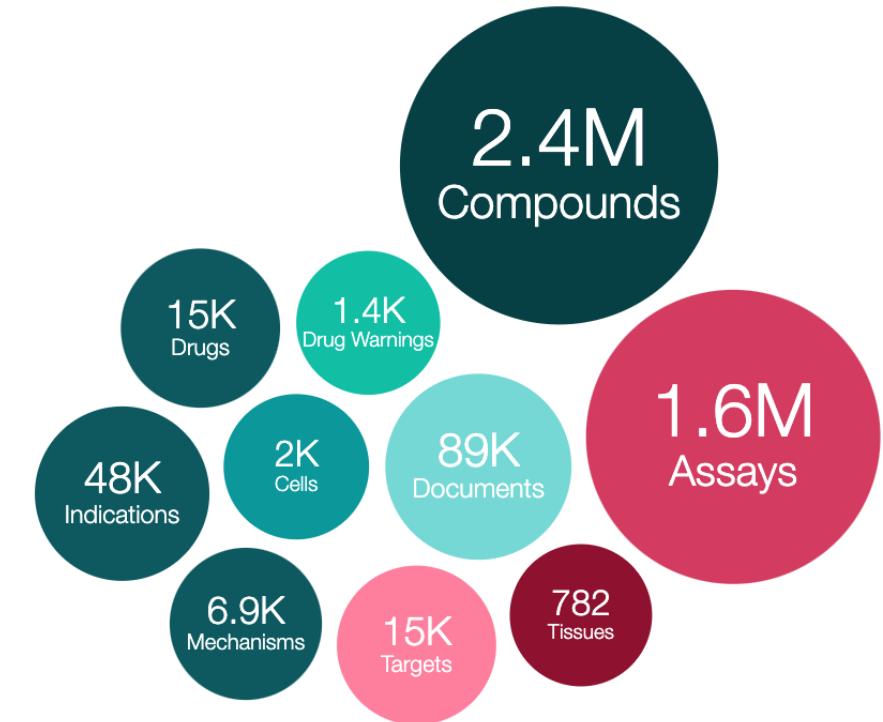
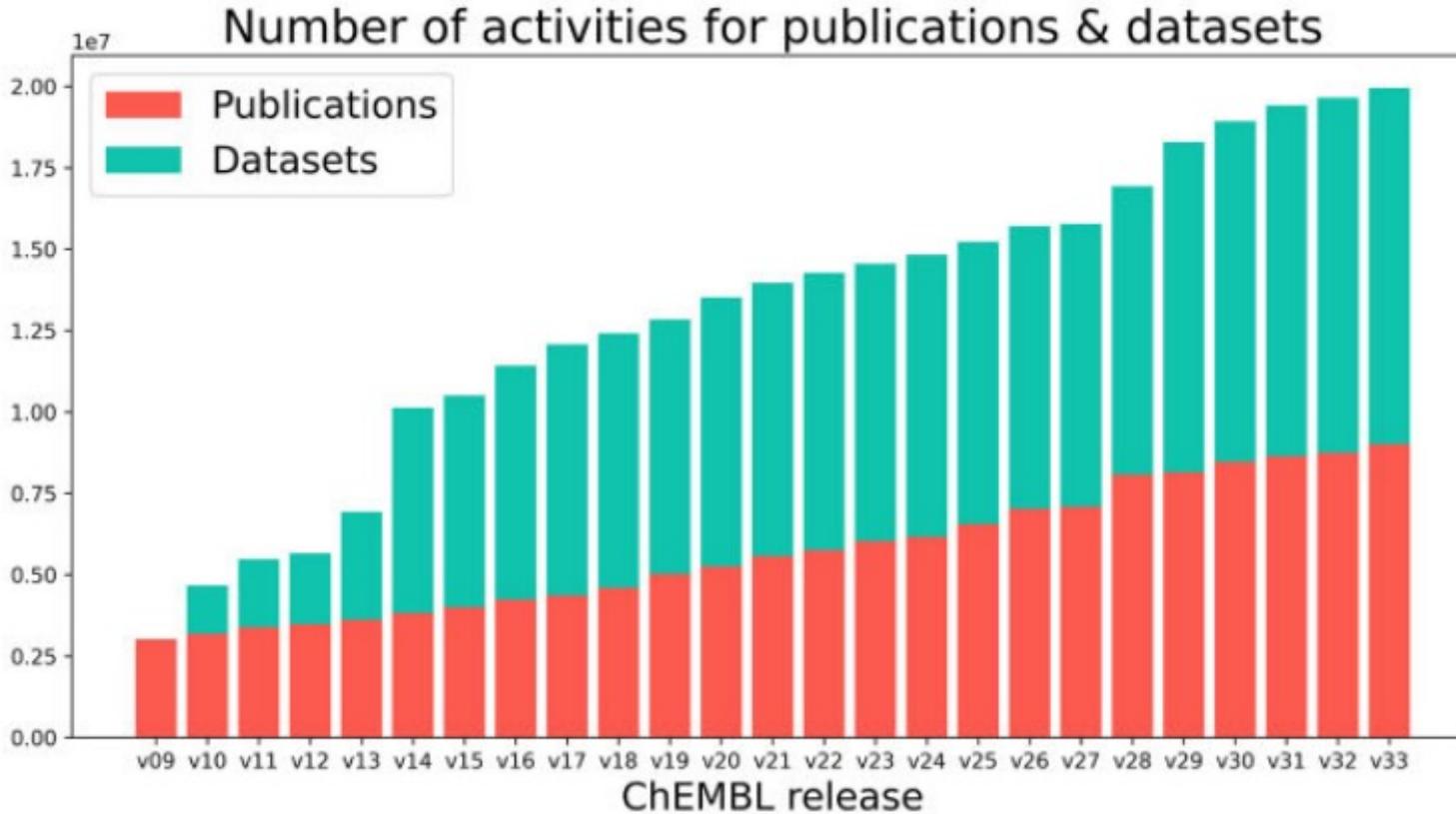
This screenshot shows the SureChEMBL Beta search interface. It features a search bar with the query "diabet\*", a "SEARCH" button, and three toggle options: "All chemically annotated authorities", "Specify dates", and "Structure search". Above the search bar, there's a survey prompt: "Do data resources managed by EMBL-EBI and our collaborators make a difference to your work? Please take 10 minutes to fill in our annual user survey, and help us make the case for why sustaining open data resources is critical for life sciences research." Below the search bar, there are links for "Search", "Downloads", "Wiki", and "Contact Us".

This screenshot shows the ChEBI homepage. The header includes links for "EMBL-EBI | Chemical Biology | ChEBI", "Advanced Search", "Submit", "Downloads", "Tools", "Docs", "About", and "Contact". The main content area features a search bar with placeholder text "Enter your search term" and a "Search" button. Below the search bar, there are examples of searches: "iron\*", "InChI=1S/CH4O/c1-2/h2H,1H3, caffeine | Advanced Search". The page is divided into several sections: "Advanced Search" (with a magnifying glass icon), "Submit" (with a green gear icon), "Downloads" (with a download icon), "Tools" (with a wrench and screwdriver icon), "Documentation" (with a document icon), and "About ChEBI" (with a gear icon). The background features a blue and green abstract molecular structure.

Coming soon!

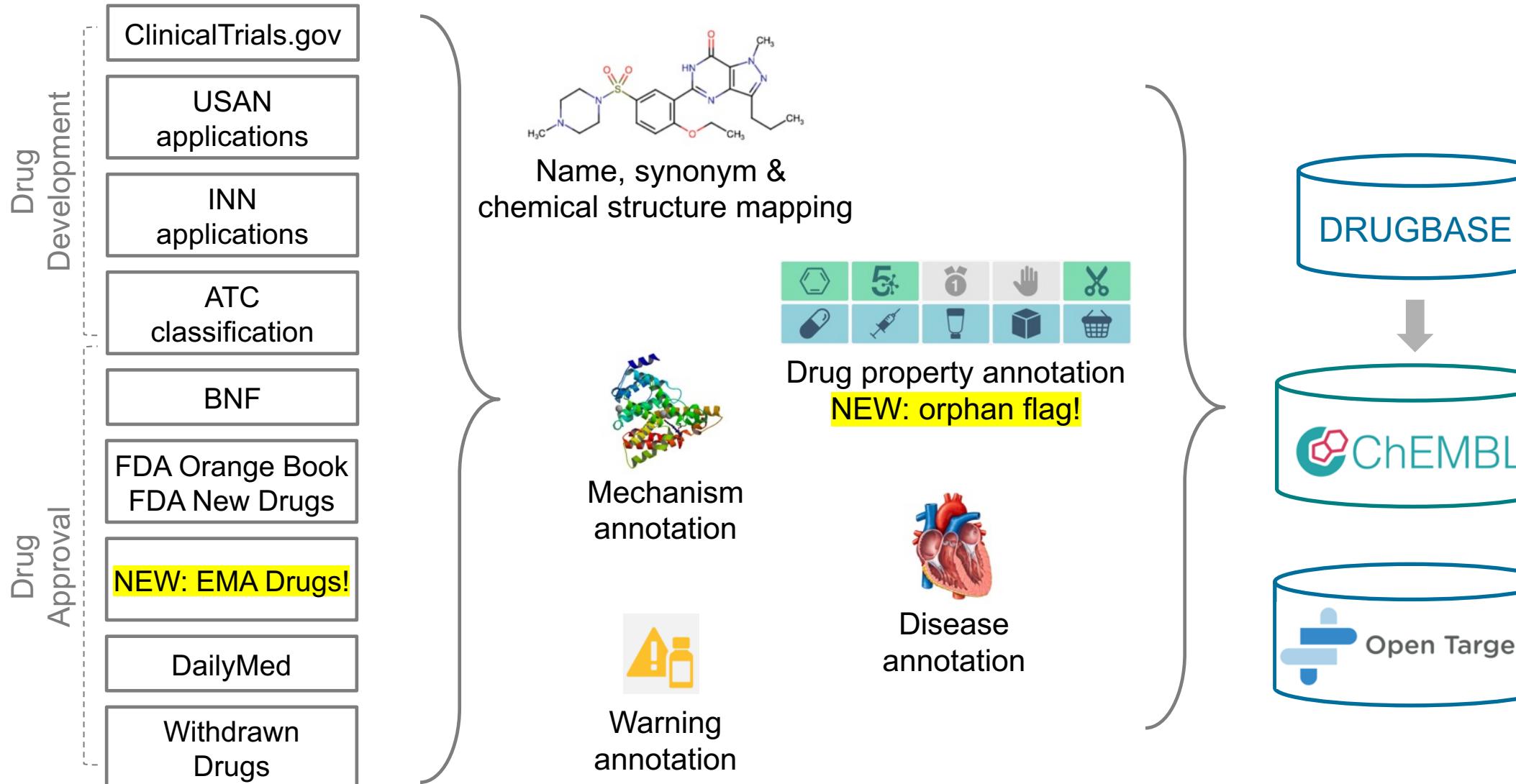
We have also stopped and retired some things

# Continued growth in database size and content: ChEMBL



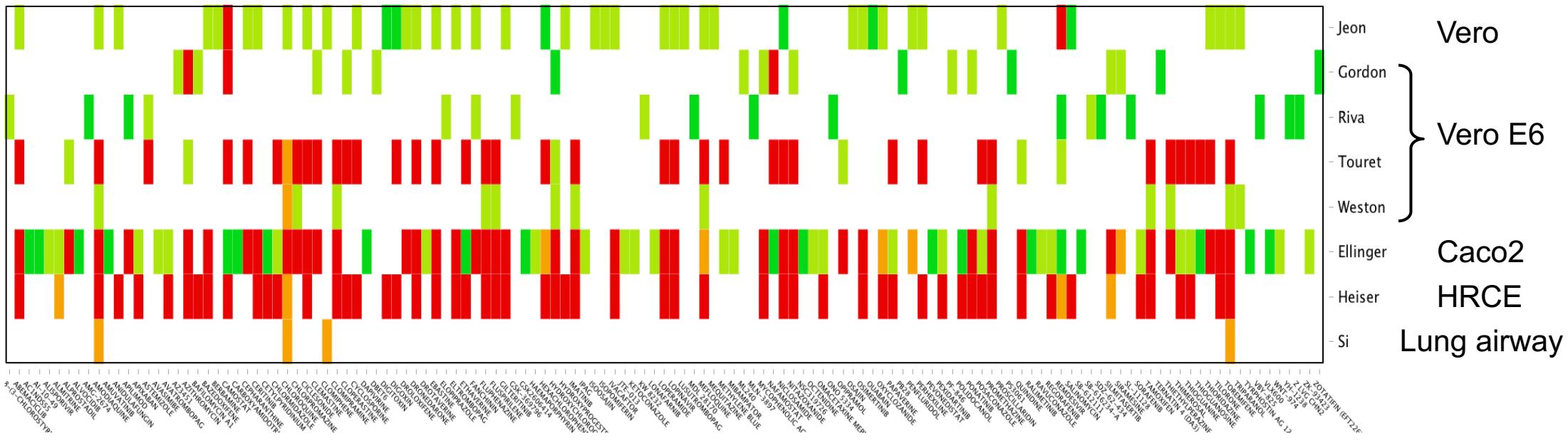
Data depositions are very welcome!

# ChEMBL Drug and Clinical Candidate data: a unique resource!

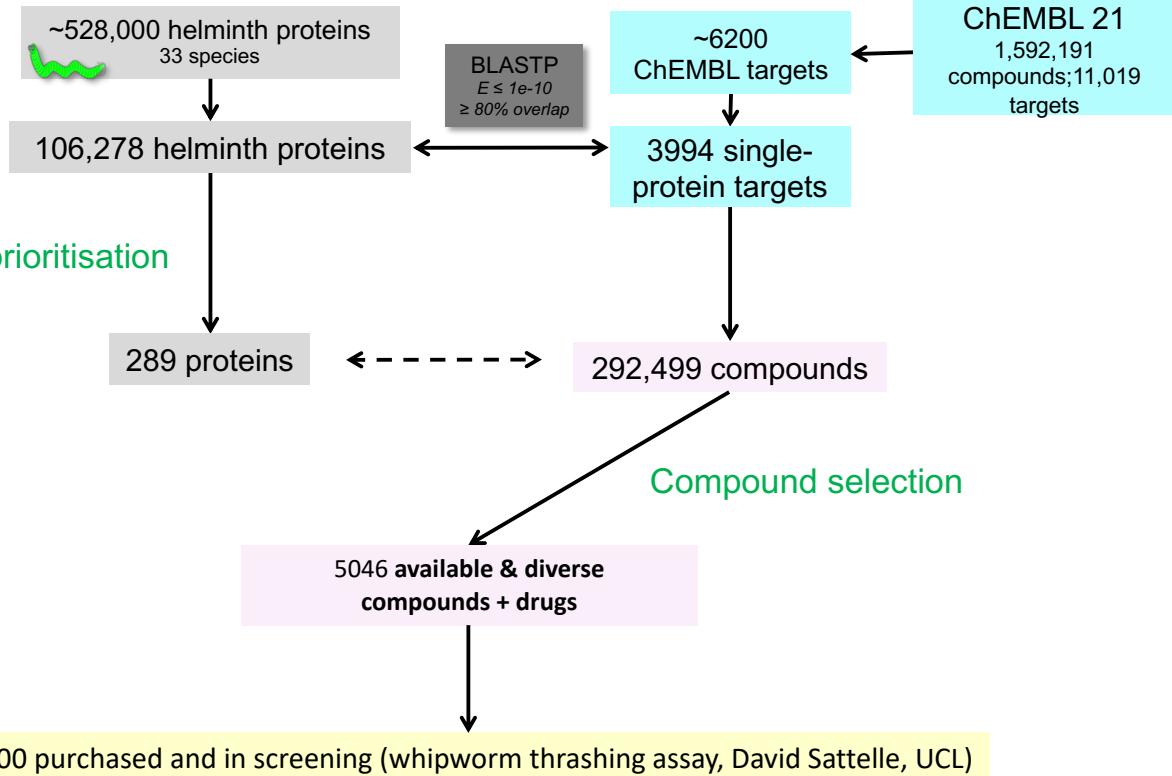
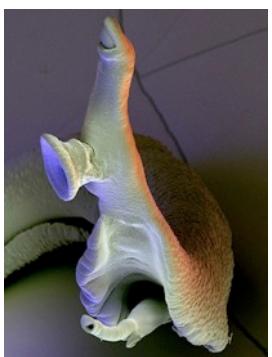
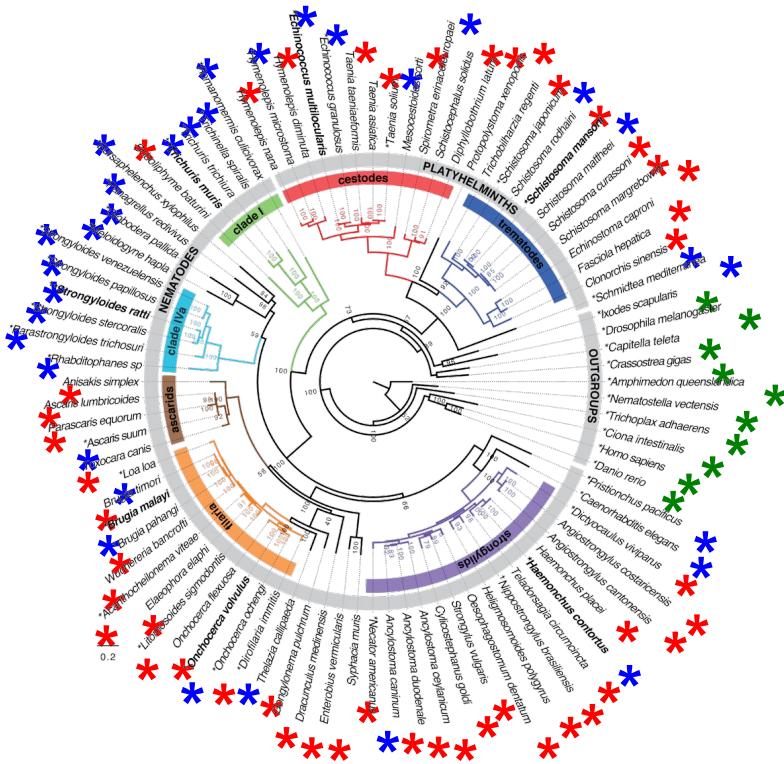


# ChEMBL Covid “special release”

- Focus on large-scale “drug repurposing” screens from recognised labs
  - Compounds should inhibit viral activity without killing cells
  - Rapid curation and release
  - Only 26 compounds active in more than one data set



# Helminth chemogenomics analysis and compound selection



**nature genetics**

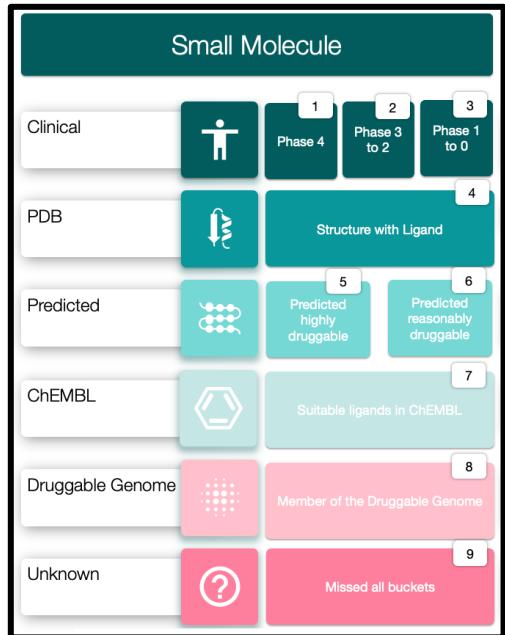
ARTICLES  
<https://doi.org/10.1038/s41588-018-0262-1>  
OPEN

## Comparative genomics of the major parasitic worms

International Helminth Genomes Consortium\*

Parasitic nematodes (roundworms) and platyhelminths (flatworms) cause debilitating chronic infections of humans and animals, decimate crop production and are a major impediment to socioeconomic development. Here we report a broad comparative study of 81 genomes of parasitic and non-parasitic worms. We have identified gene family births and hundreds of expanded gene families at key nodes in the phylogeny that are relevant to parasitism. Examples include gene families that modulate host immune responses, enable parasite migration through host tissues or allow the parasite to feed. We reveal extensive lineage-specific differences in core metabolism and protein families historically targeted for drug development. From an in silico screen, we have identified and prioritized new potential drug targets and compounds for testing. This comparative genomics resource provides a much-needed boost for the research community to understand and combat parasitic worms.

# Tractability assessment: a key factor in target selection



Summary of tractability information for PSEN1 for small molecule and antibody modalities.

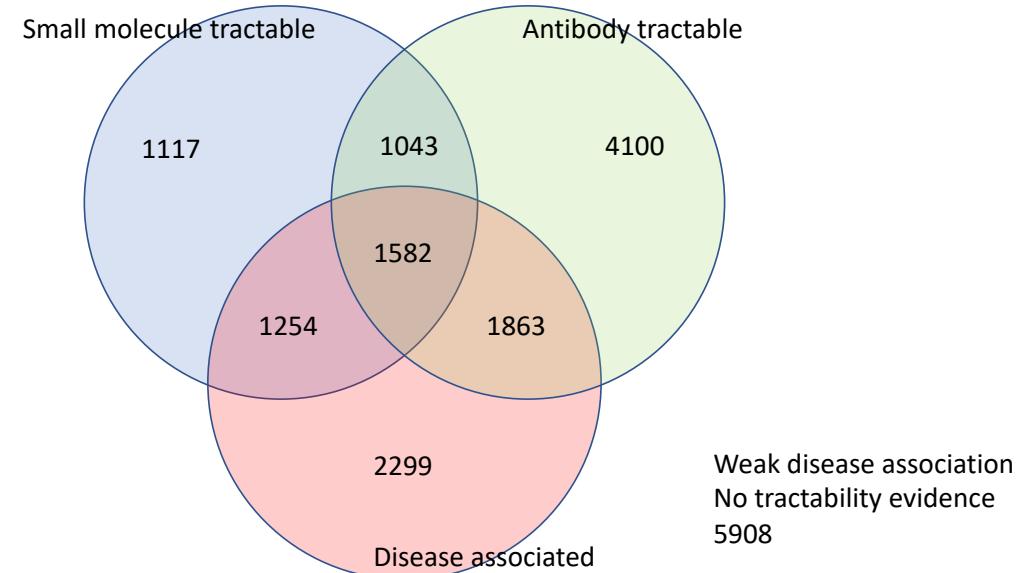
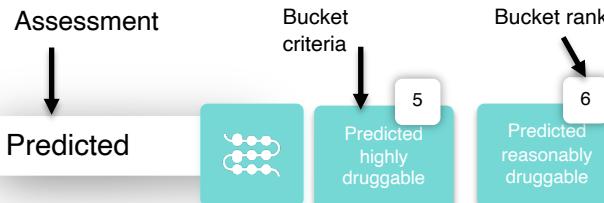
Source: Open Targets

**Small molecule**

Clinical precedence	Discovery precedence	Predicted tractable
Phase 4	Phase 0 or 1	PDB targets with ligands
Phase 2 or 3	Active compounds in ChEMBL	DrugEability score > 0.7
		DrugEability score 0 to 0.7
		Druggable genome

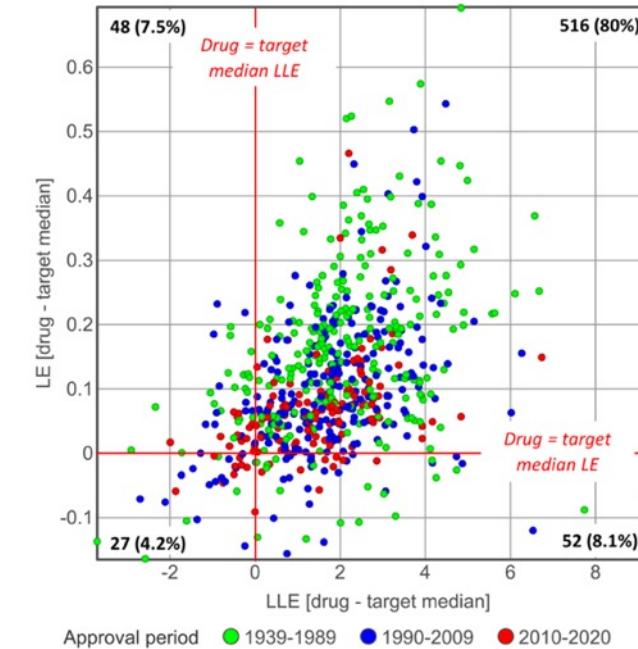
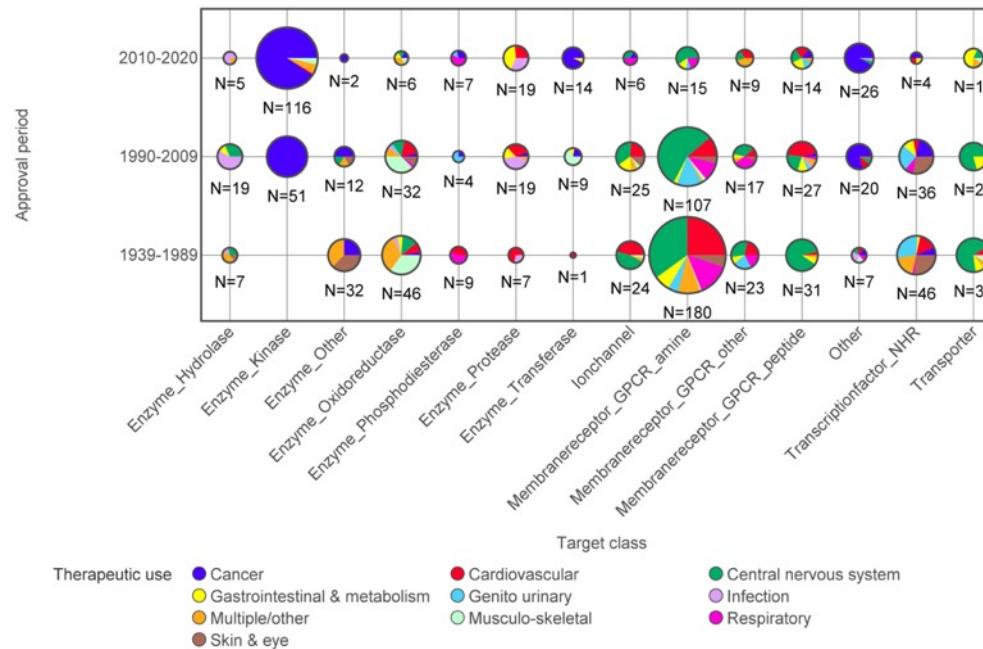
**Antibody**

Clinical precedence	Predicted tractable - high confidence	Predicted tractable - medium to low confidence
Phase 4	Phase 2 or 3	Phase 0 or 1
	UniProt location - high confidence	UniProt location - low or unknown confidence
	GO cell component - high confidence	UniProt predicted signal peptide or transmembrane region
		GO cell component - medium confidence
		Human Protein Atlas - high confidence

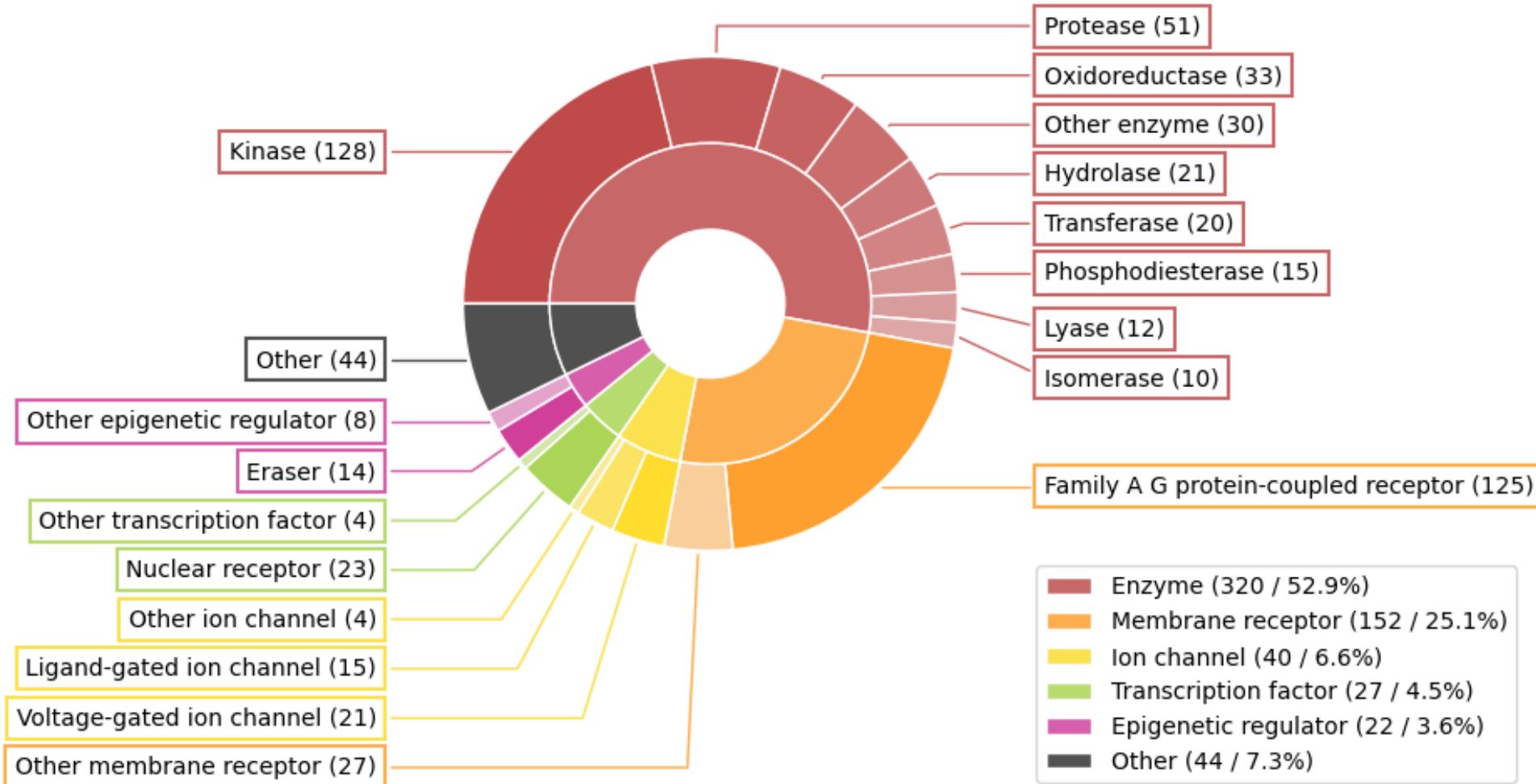


# ChEMBL data differentiates drugs from “bioactive molecules”

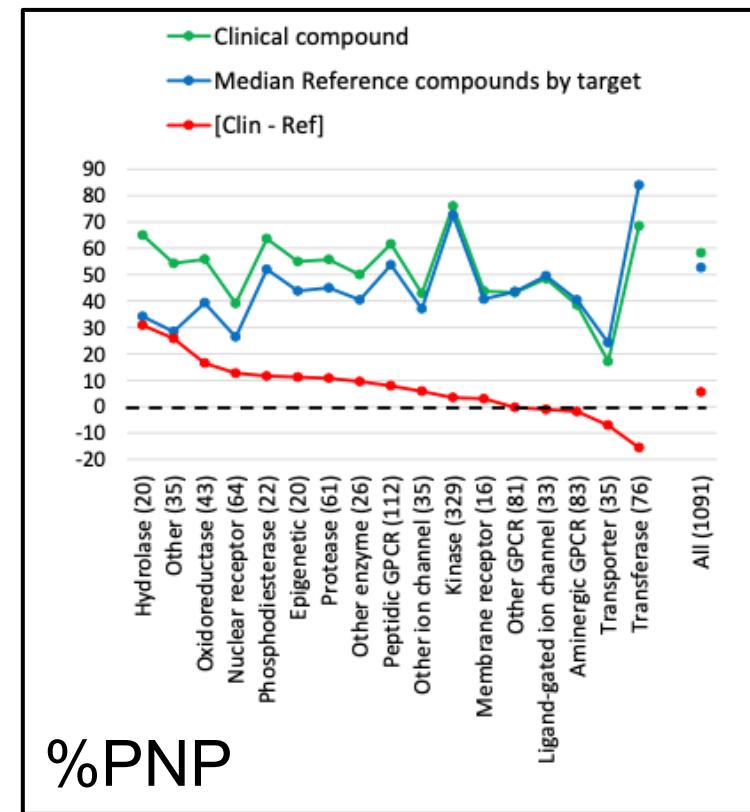
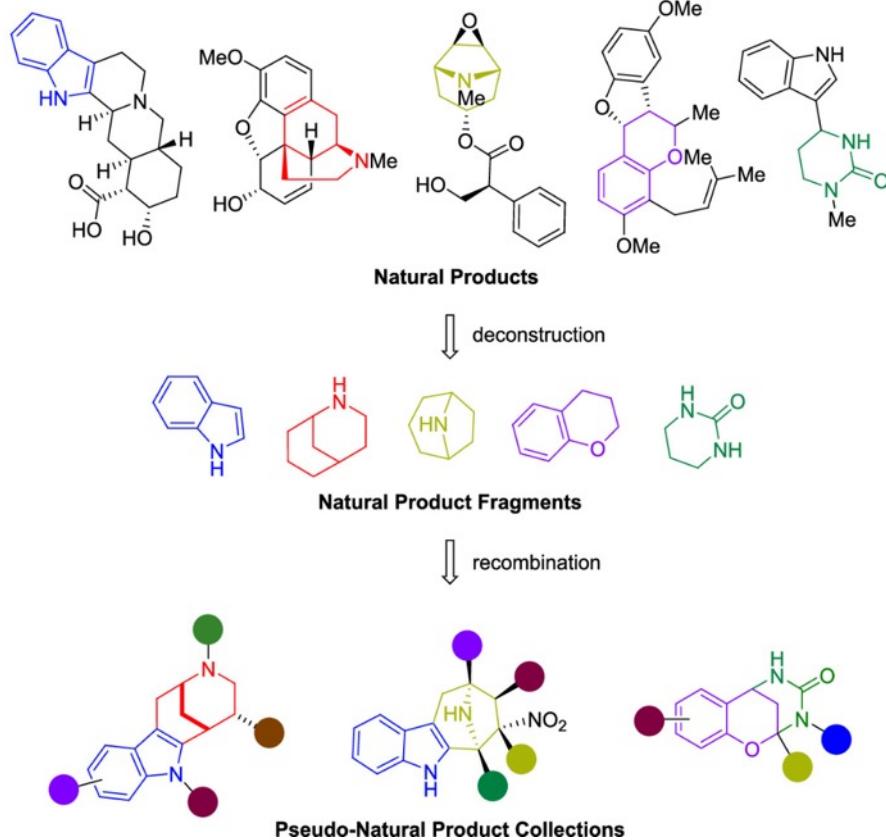
- 643 drugs against 271 targets
- 360K comparator compounds
- All with quantitative activity data and calculated properties
- Drugs continue to be more efficient than their comparator compounds



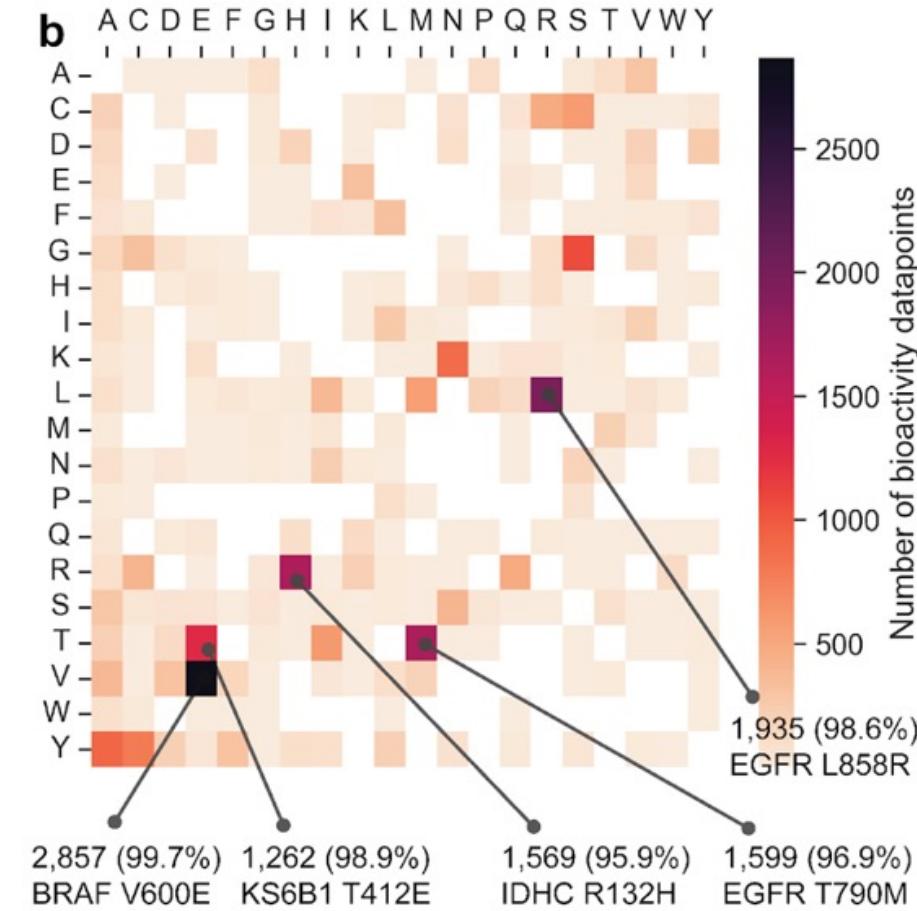
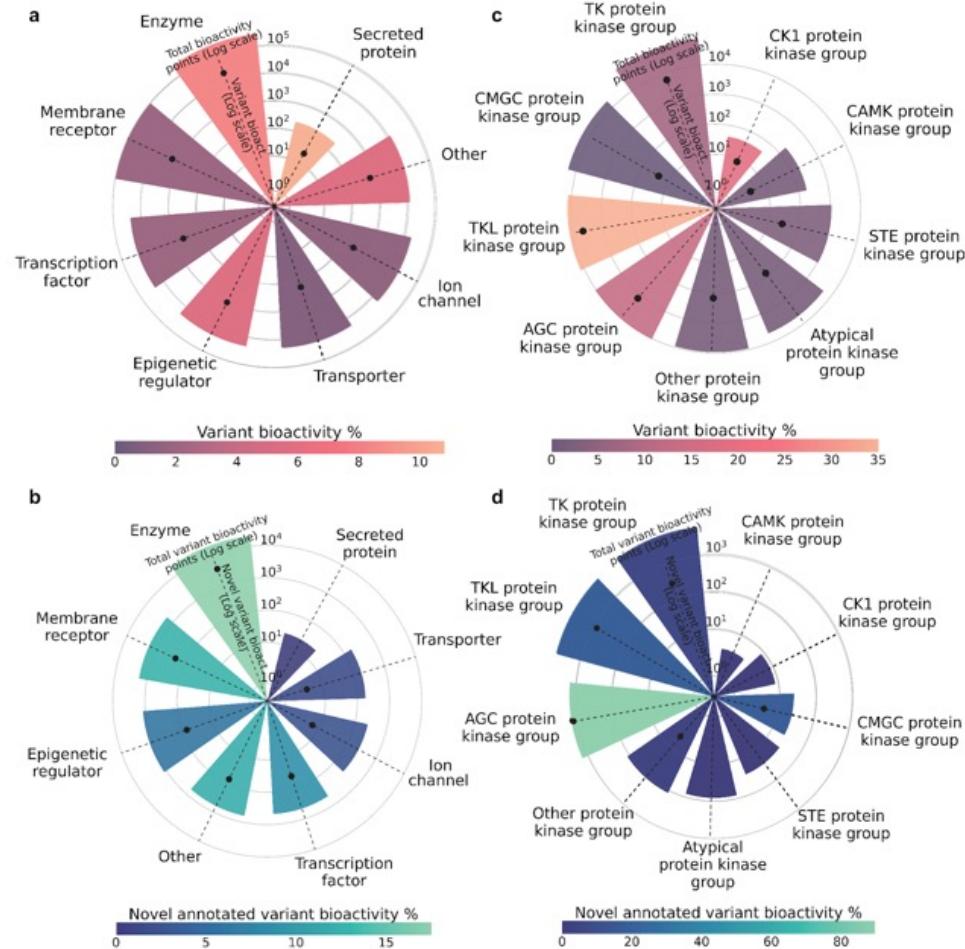
# Robust & reliable datasets for community use



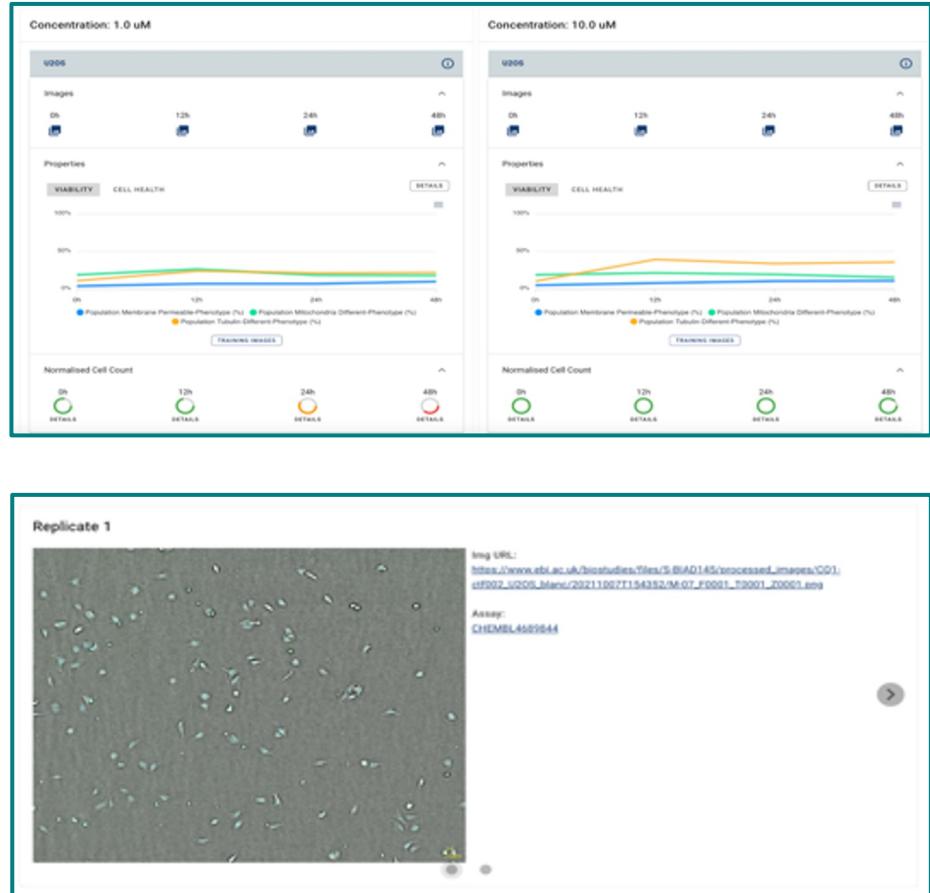
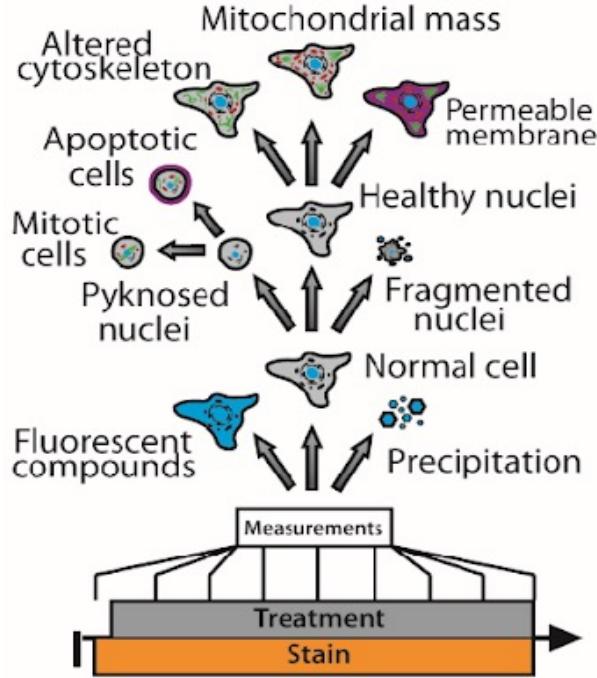
# “Natural selection” in small molecule drug discovery



# Protein variants & bioactivity data (Gerard's talk)



# ChEMBL underpins multiple external collaborations (Brian's talk)



Key data in ChEMBL,  
also visible via  
EUbOPEN gateway



Plate images per time  
point & replicate.  
Hosted by the BioImage  
Archive.



Open Targets



IDG  
ILLUMINATING the  
DRUGGABLE GENOME



transQST



EUbOPEN  
Enabling & Unlocking Biology in the OPEN



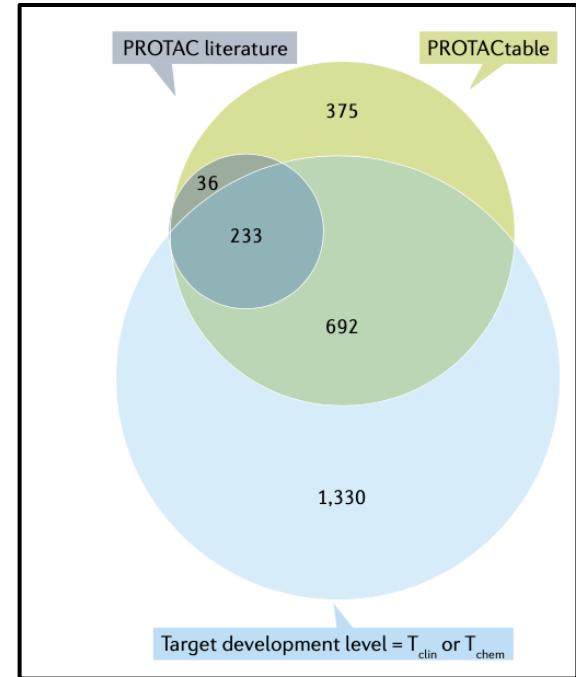
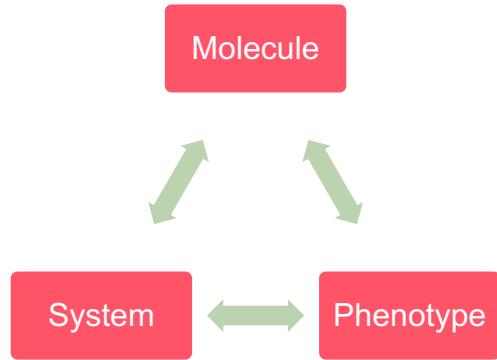
eTRANSAFE

# Open Targets is a key strategic partner

**DW Drug Warnings**  
Manually curated withdrawn and black box warnings for **ROSIGLITAZONE**. Source: ChEMBL

Search

Warning type	Adverse event	ChEMBL warning class	Country / region
Withdrawn	cardiotoxicity?	cardiotoxicity	Armenia; Guatemala; Egypt
Withdrawn	edema?	cardiotoxicity	European Union
Withdrawn	heart failure?	cardiotoxicity	European Union
Black Box Warning	N/A	cardiotoxicity	United States



**TR Tractability**  
Target tractability assessment for **NOD2**. Source: Open Targets.

Target	Association Score $\leq$	Target in clinic	Membrane protein	Secreted protein	Lipid binder	Small molecule binder	Predicted pocket	Mouse ortholog identity	Chemical probes	Genetic constraint	Mouse models	Gene essentiality	Known adverse events	Cancer driver gene	Paralogs	Tissue specificity	Tissue distribution
NOD2	■	●	○	○	●	●	●	●	●	●	●	●	●	●	●	●	●

**Small molecule**

- Approved Drug
- Advanced Clinical
- Phase 1 Clinical
- Structure with Ligand
- High-Quality Ligand
- High-Quality Pocket
- Med-Quality Pocket
- Druggable Family

**Antibody**

- Approved Drug
- Advanced Clinical
- Phase 1 Clinical
- Human Protein Atlas loc

**PROTAC**

- Approved Drug
- Advanced Clinical
- Phase 1 Clinical
- UniProt loc high conf
- GO CC high conf
- UniProt loc med conf
- UniProt SigP or TMHMM
- GO CC med conf
- Human Protein Atlas loc

**Other modalities**

- Approved Drug
- Advanced Clinical
- Phase 1 Clinical
- Literature
- UniProt Ubiquitination
- Database Ubiquitination
- Half-life Data
- Small Molecule Binder

**IL12B**

**ITGA4**

**TNF**

**NR3C1**

Associations per page: 50 ▾ 1-50 of 5870 < >

Priority indicator: Unfavorable (red) Favourable (green)

**2000 articles reporting probes-targets**

**800 articles with a disease in the same sentence**

**Case A:**  
PMID: 16721373  
*"Growth inhibition and differentiation of human **breast cancer** cells by the **PAFR** antagonist **WEB-2086**"*

**1200 articles without a clear disease in sentence**

**Case B:**  
PMID: 17476331  
*"The **PIK1** inhibitor **BI2536** and siRNA against NEDD1 act synergistically to reduce **cell viability**"*

*Not specific enough!*

# What will the next 15 years bring?



For more meaningful insights, see the talks, view the posters, meet the team!

# My next chapter: LifeArc, a self-financing charity addressing unmet patient need

## Our Translational Challenges

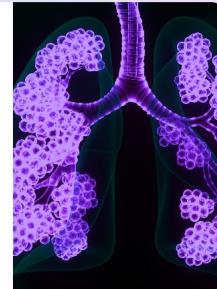
### Motor Neuron Disease

Our vision is a world where motor neuron disease is preventable and treatable.

[Read more](#)

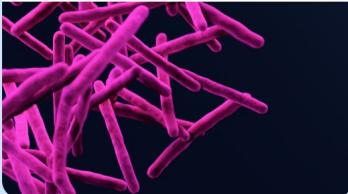
### Chronic Respiratory Infection

Our vision is improved quality of life for those living with cystic fibrosis and bronchiectasis.

[Read more](#)

### Global Health

Our vision is a world with affordable and accessible solutions to better understand, treat, and prevent infectious diseases.

[Read more](#)

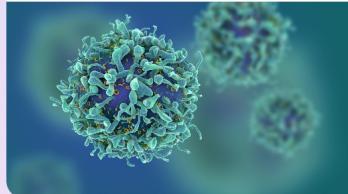
### Rare Disease

Our vision is an ecosystem working together to get discoveries and treatments to rare disease patients faster.

[Read more](#)

### Childhood Cancer

Our vision is to drive life-changing innovations for children with cancer.

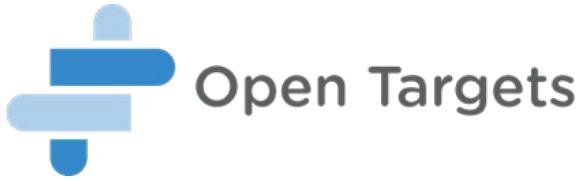
[Read more](#)

16 May 2024

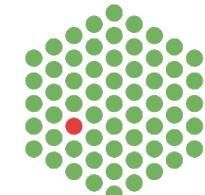
**Britain's largest health research programme, Our Future Health, forms new £10m partnership with LifeArc**

[News releases](#)

# Chemical Biology resources: funders & collaborators



Biotechnology and  
Biological Sciences  
Research Council



# Above all, a big thank you to all colleagues past & present

