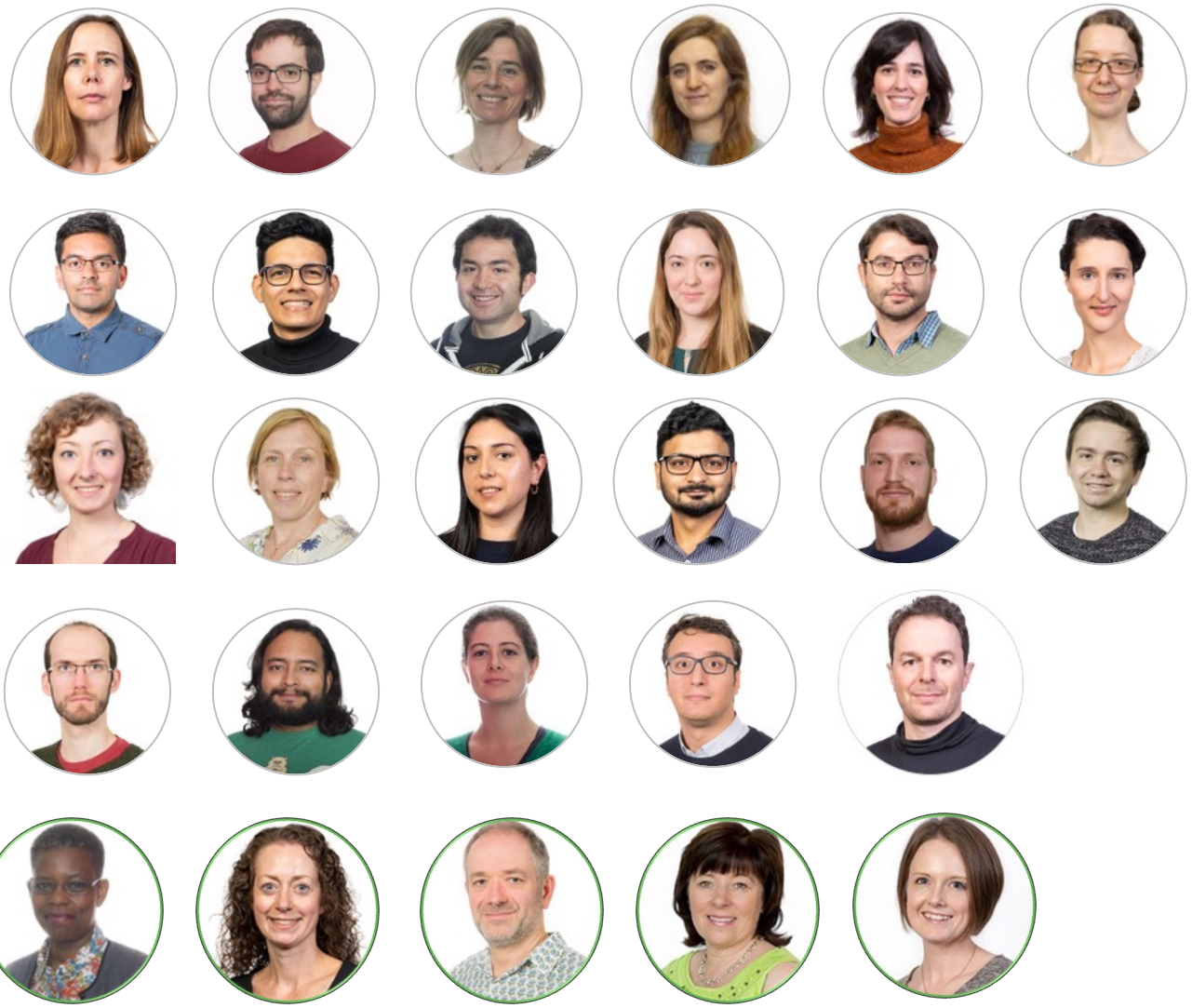
Chemical biology resources at EMBL-EBI, 2016-2024

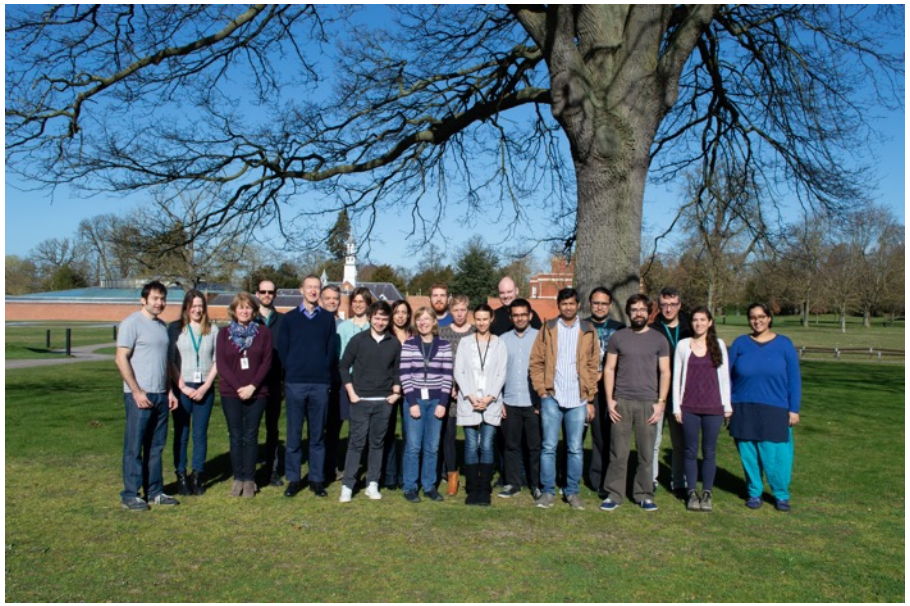
Andrew Leach

Formerly
Head of Chemical Biology and
Head of Industry Partnerships

EMBL-EBI

None of the following would be possible without an amazing team





EBI’s chemical biology resources

ChEBI
~60K structures

2004

ChEMBL
>2m structures

2009

UniChem
Links ~40 resources
>175m structures

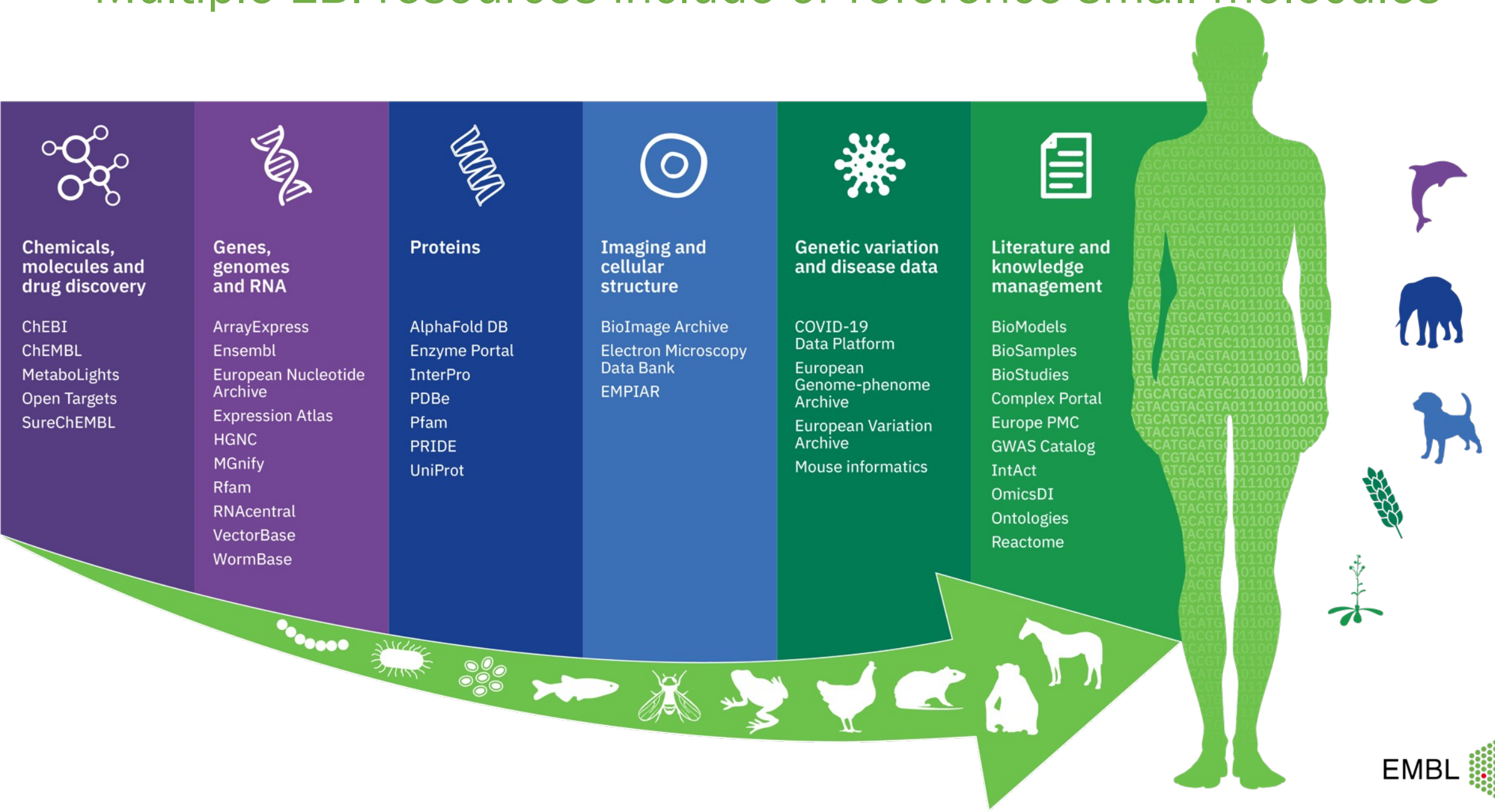
2013

SureChEMBL
>20m structures

2014

EMBL

Multiple EBI resources include or reference small molecules



EMBL

A few “headline” achievements

>50 Team members

£10m in new funding

Significant growth in data resources

Major software overhaul and rewrite

GBC and Elixir accreditation for ChEMBL & ChEBI

>40 publications

Two service reviews, one research review

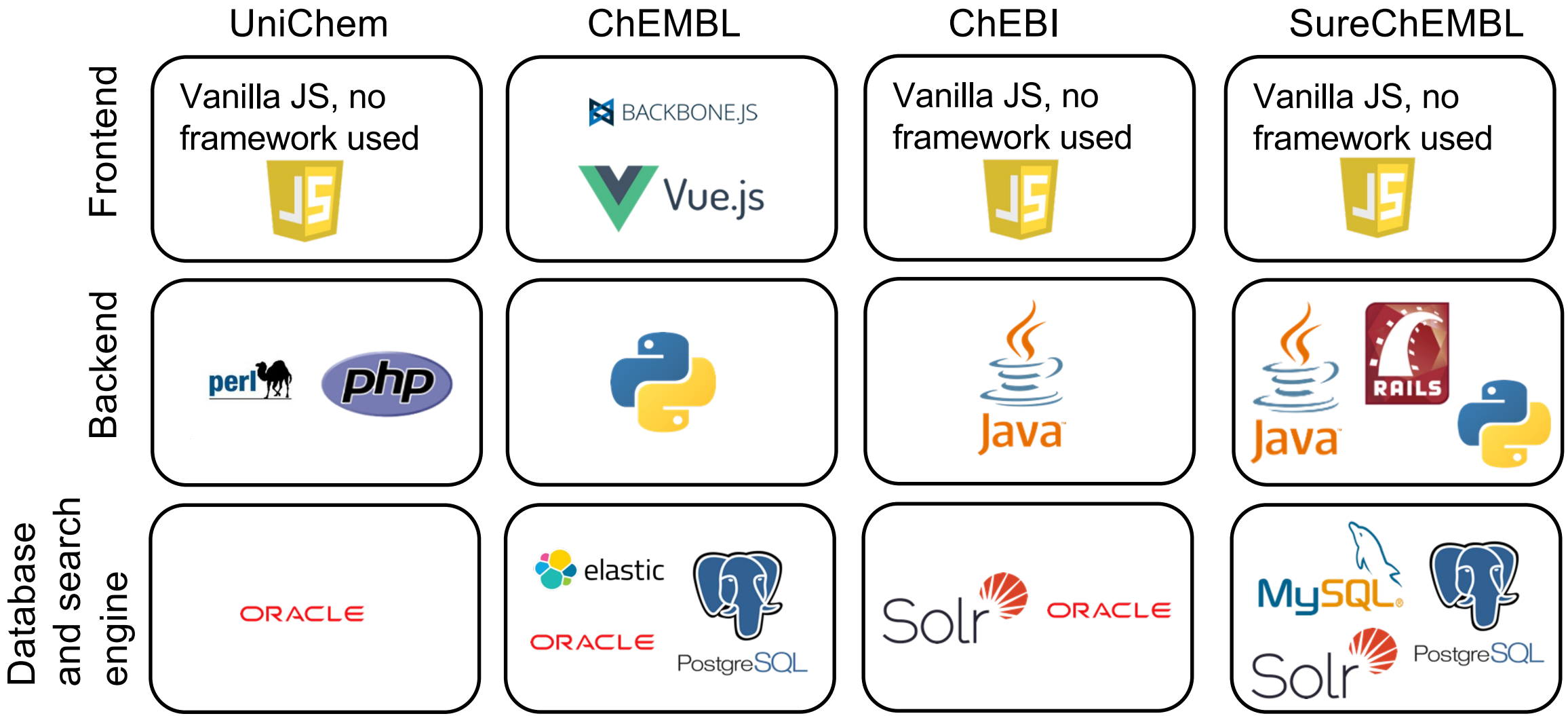
Open Targets impact incl. tractability and safety

Multiple collaborations

Survived 1 pandemic

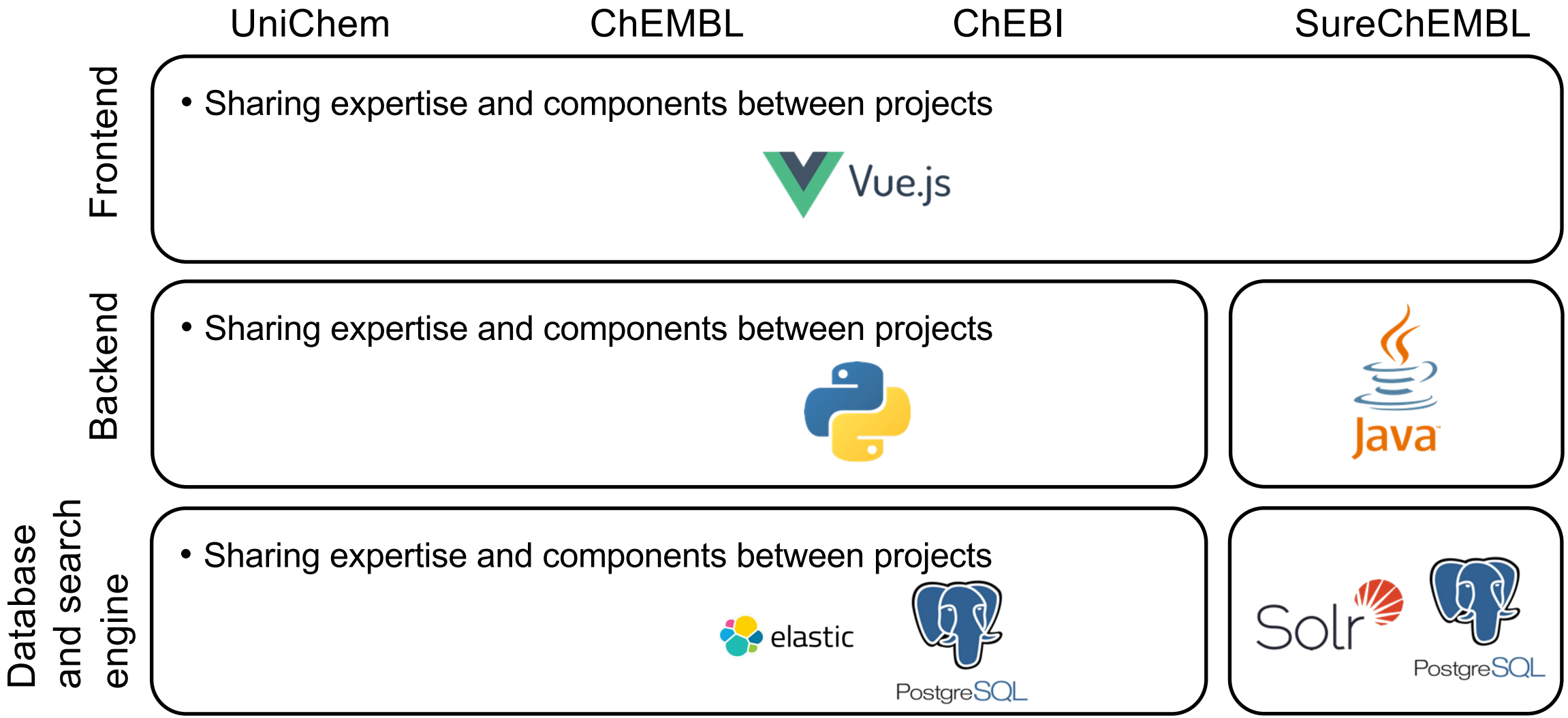
EMBL

Software infrastructure: legacy



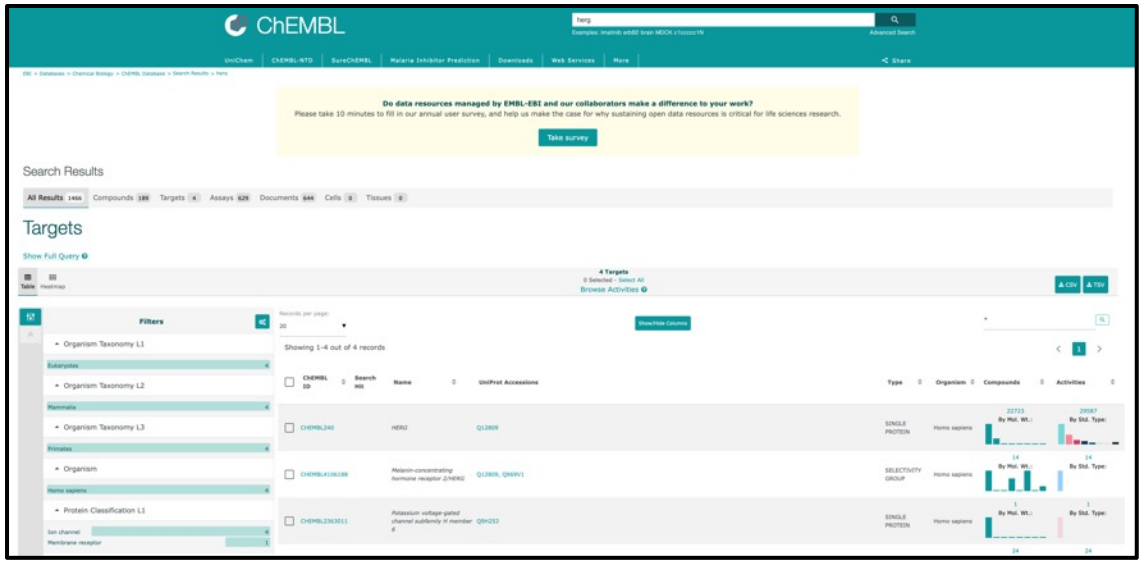
EMBL

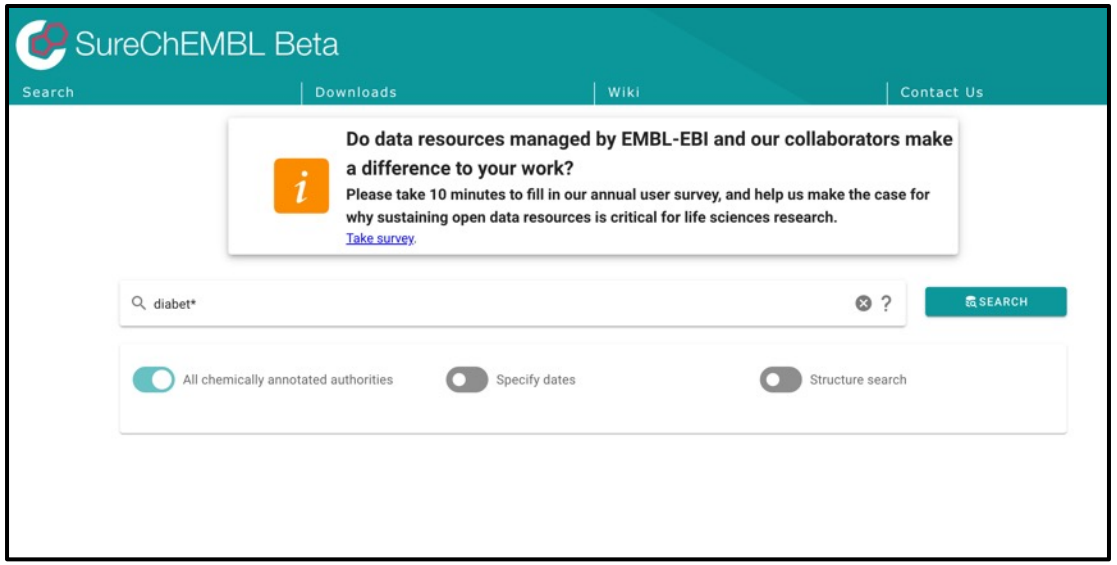
Software infrastructure: future



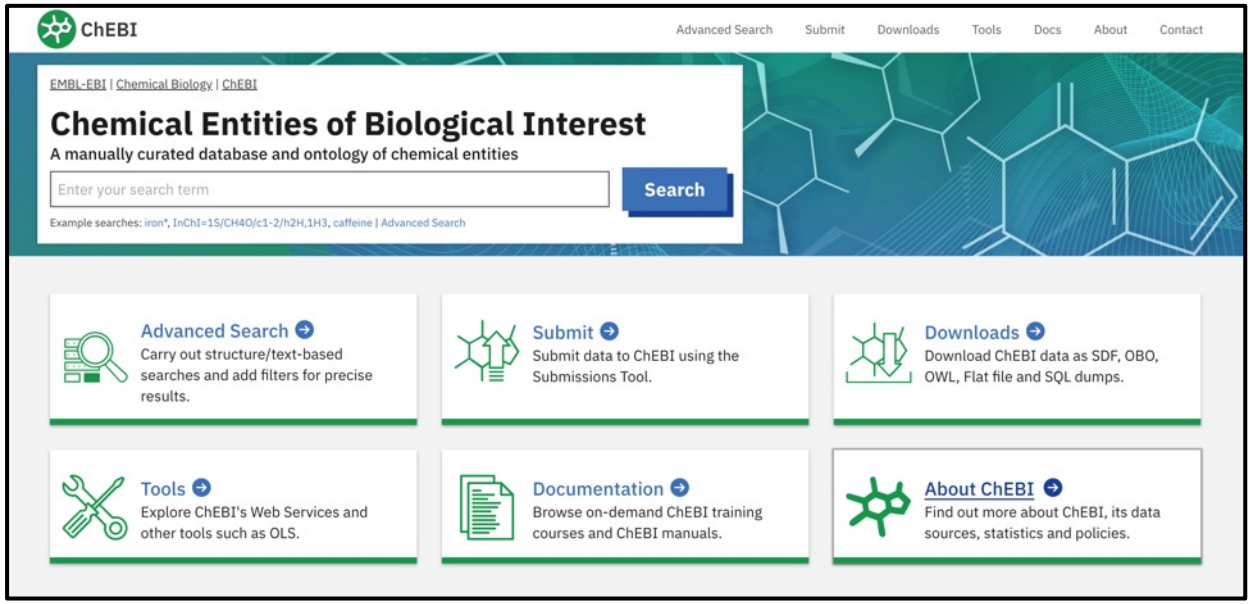
EMBL

New interfaces for all our resources





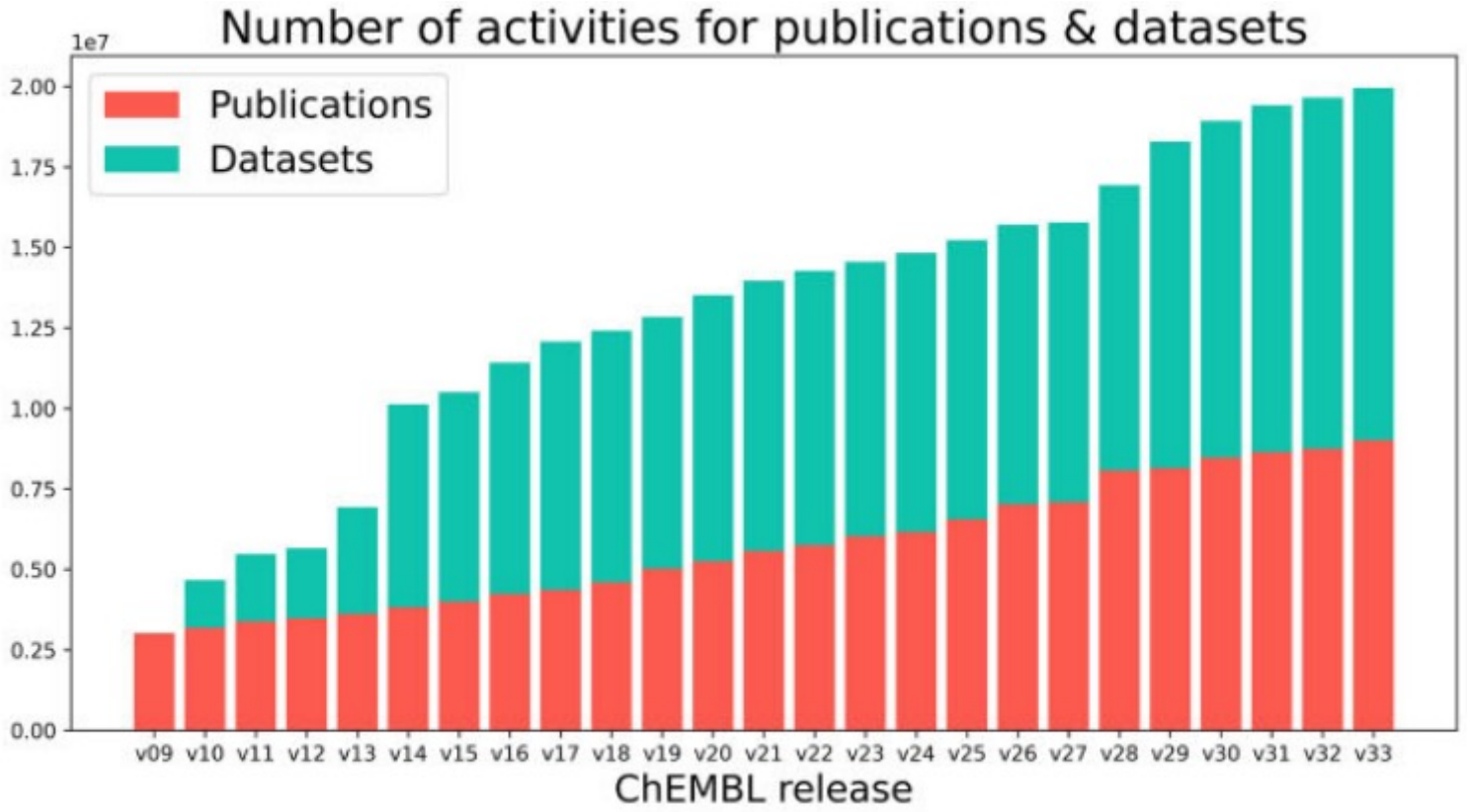
Coming soon!

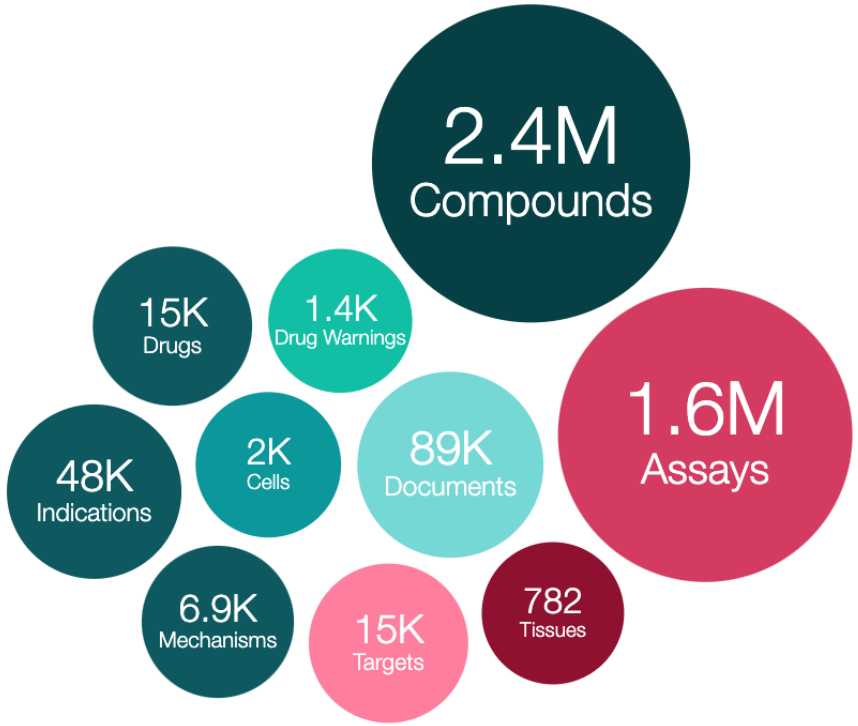


EMBL

We have also stopped and retired some things

Continued growth in database size and content: ChEMBL





Data depositions are very welcome!

Zdrazil et al NAR 2024 52:D1180-D1192. doi: 10.1093/nar/gkad1004. PMID: 37933841; PMCID: PMC10767899.

EMBL

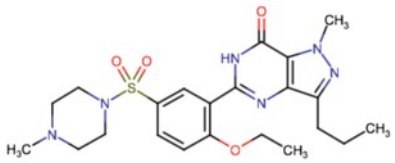
ChEMBL Drug and Clinical Candidate data: a unique resource!

Drug Development

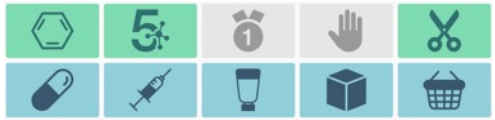
- ClinicalTrials.gov
- USAN applications
- INN applications
- ATC classification

Drug Approval

- BNF
- FDA Orange Book
  FDA New Drugs
- NEW: EMA Drugs!
- DailyMed
- Withdrawn Drugs

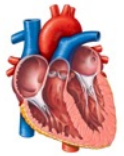


Name, synonym & chemical structure mapping



Drug property annotation
NEW: orphan flag!

Mechanism annotation



Disease annotation

Warning annotation

DRUGBASE

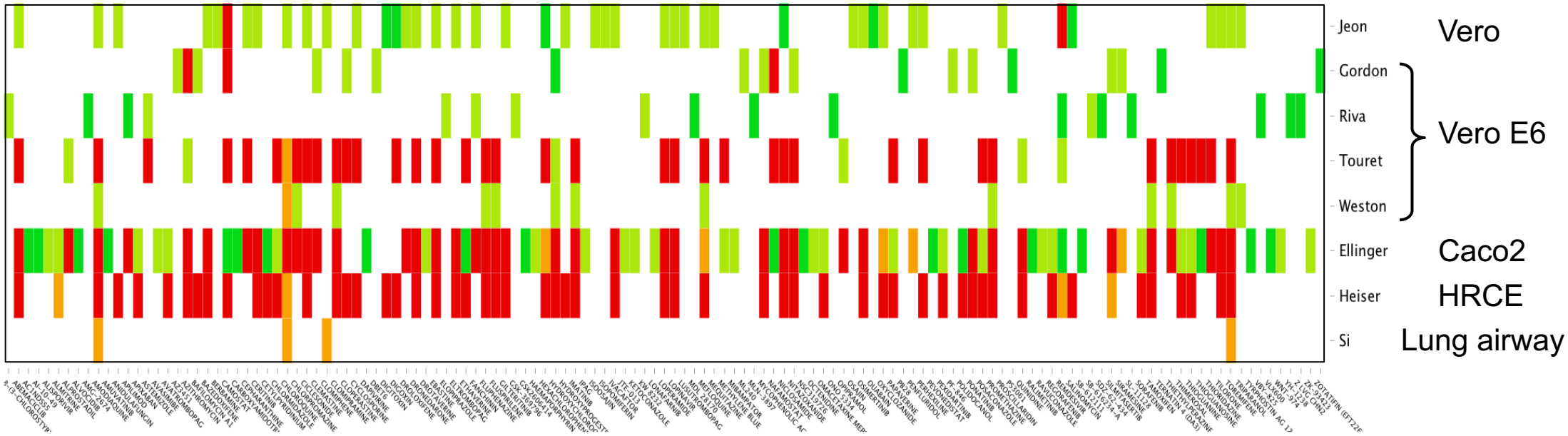
ChEMBL

Open Targets

EMBL

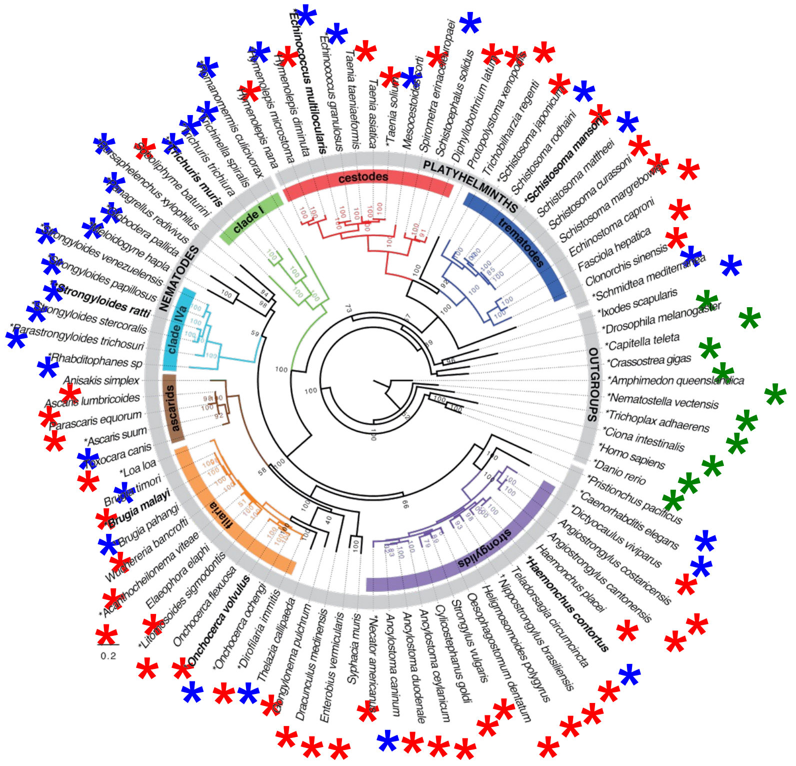
ChEMBL Covid “special release”

- 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



EMBL

Helminth chemogenomics analysis and compound selection



*33 published Helminth sequences; image by James Cotton & Bhavana Harsha

~528,000 helminth proteins
33 species

106,278 helminth proteins

BLASTP
E ≤ 1e-10
≥ 80% overlap

ChEMBL 21
1,592,191 compounds;11,019 targets

~6200 ChEMBL targets

3994 single-protein targets

Target prioritisation

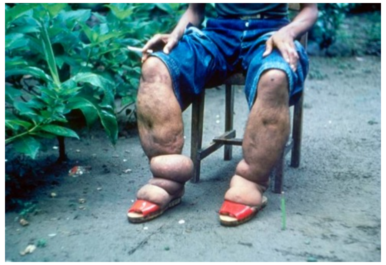
289 proteins

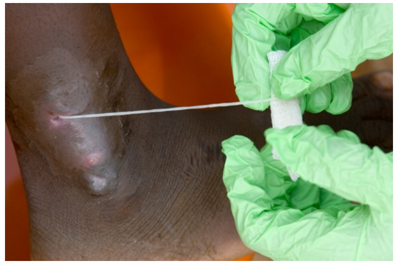
292,499 compounds

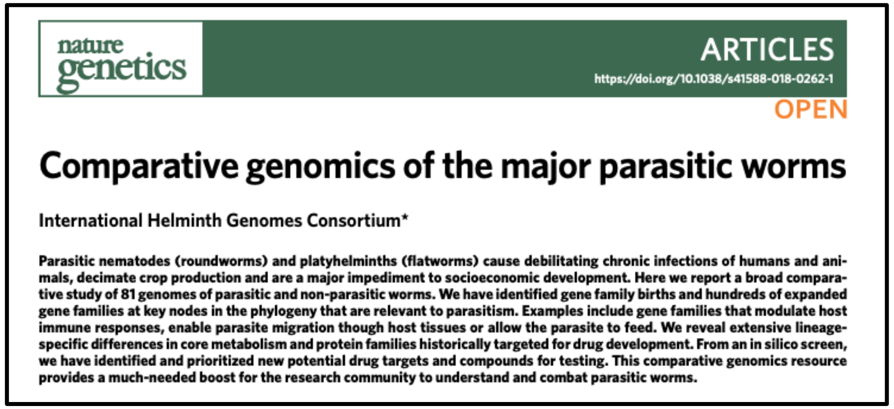
Compound selection

5046 available & diverse compounds + drugs

~400 purchased and in screening (whipworm thrashing assay, David Sattelle, UCL)







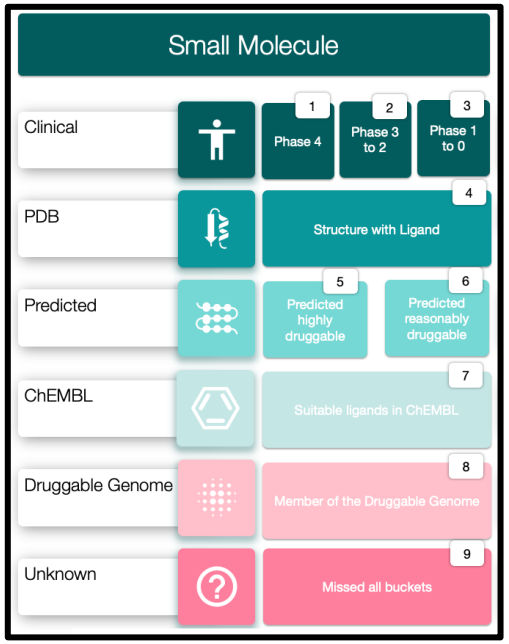
nature genetics ARTICLES
https://doi.org/10.1038/s41588-018-0262-1
OPEN

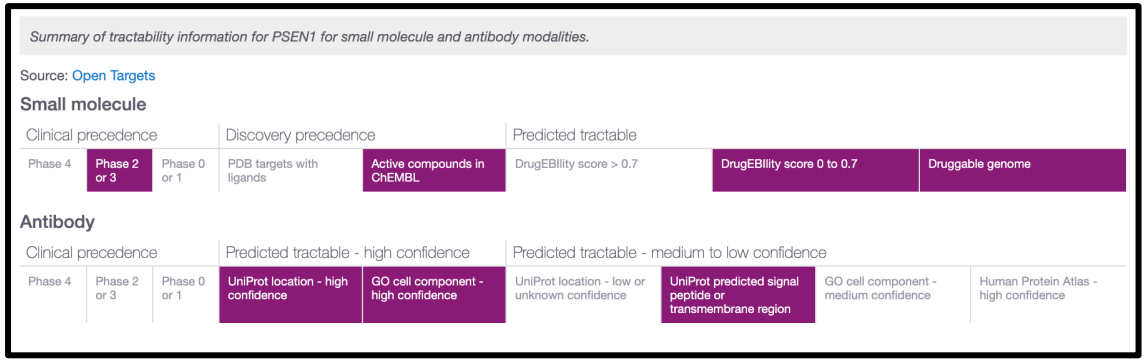
Comparative genomics of the major parasitic worms

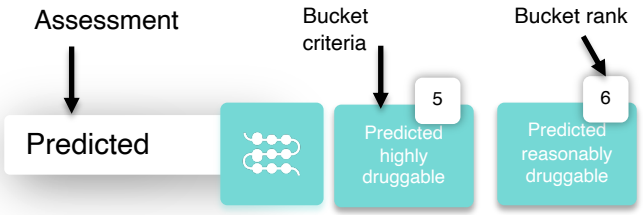
International Helminth Genomes Consortium*

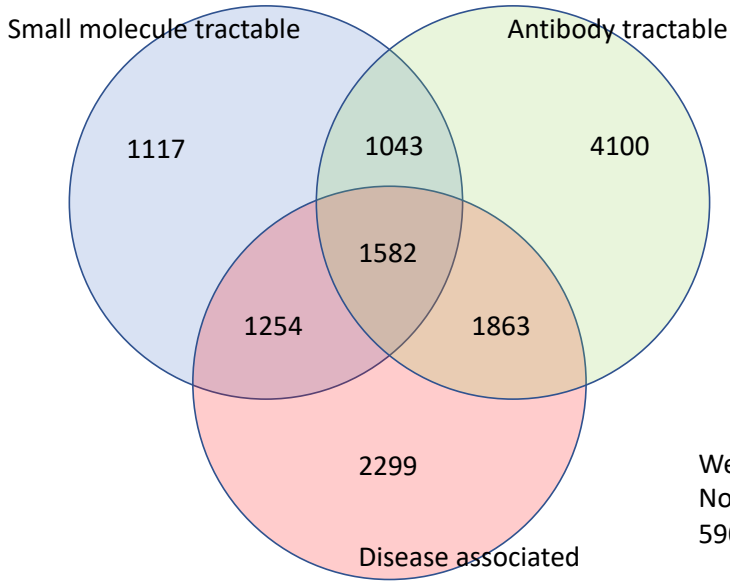
Parasitic nematodes (roundworms) and platyhelminths (flatworms) cause debilitating chronic infections of humans and animals, decimate crop production and are a major impediment to socioeconomic development. Here we report a broad comparative study of 81 genomes of parasitic and non-parasitic worms. We 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 though 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









Weak disease association
No tractability evidence
5908

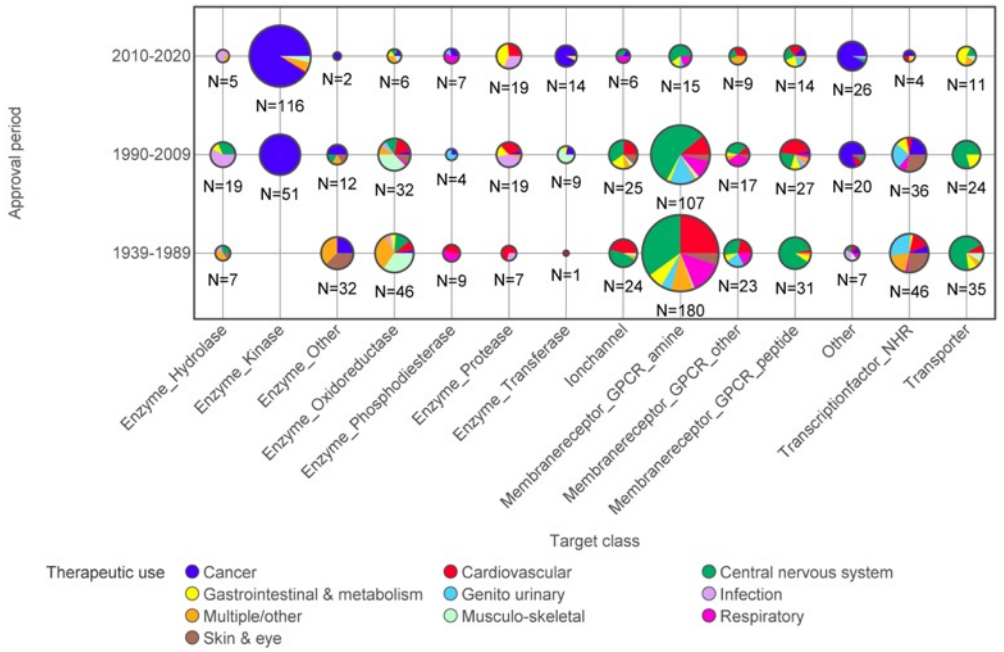
Open Targets

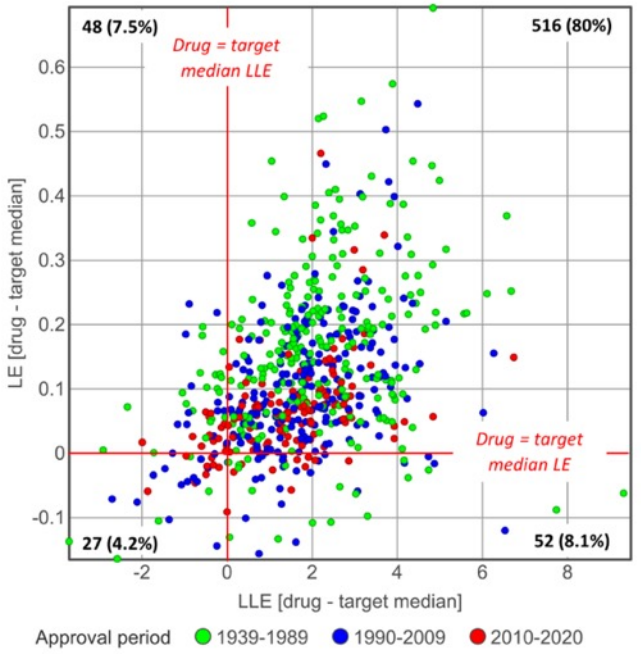
Schneider et al. Nat Rev Drug Discov 20, 789–797 (2021). https://doi.org/10.1038/s41573-021-00245-x

EMBL

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

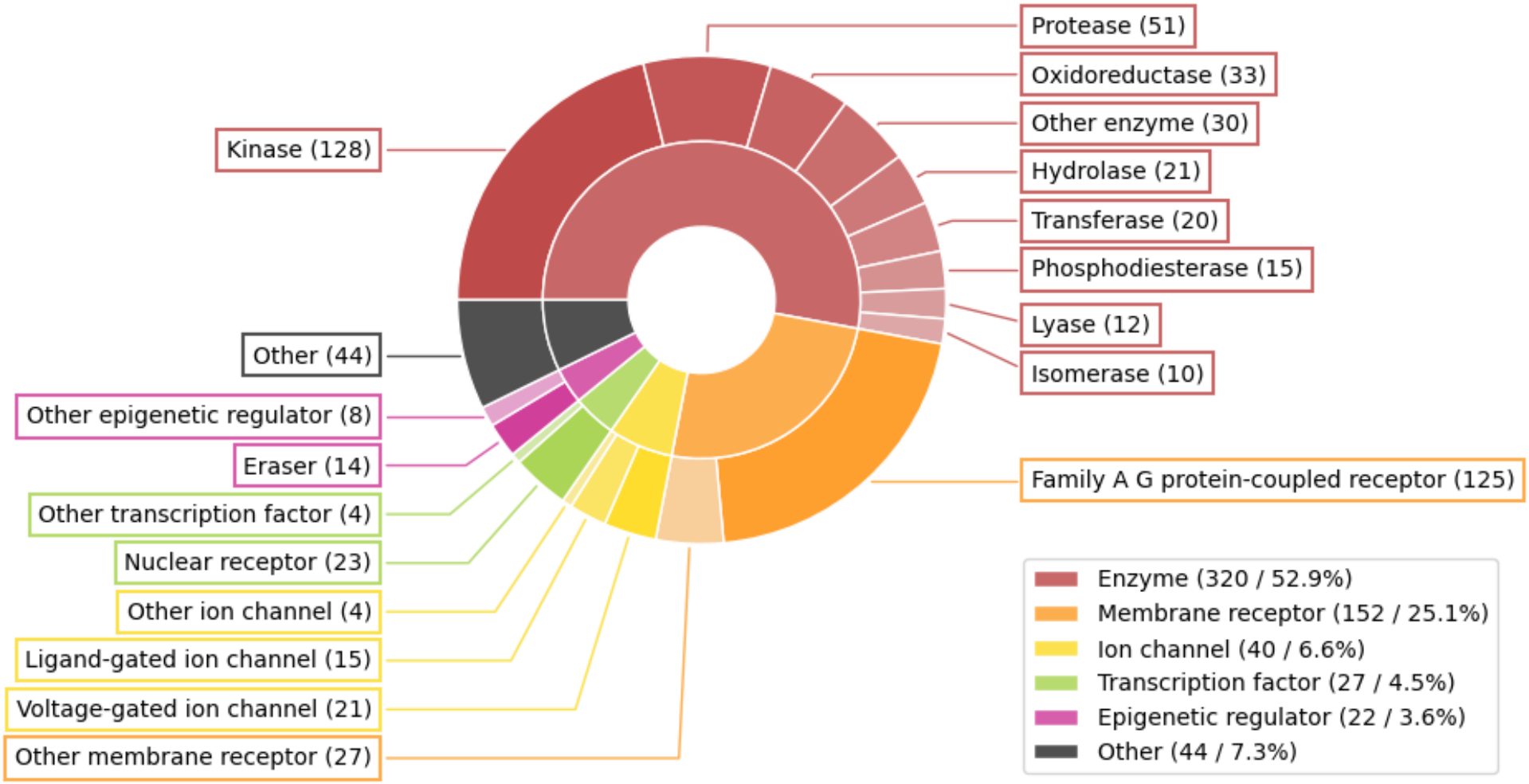




Leeson et al J Med Chem 2021 64, 7210-7230

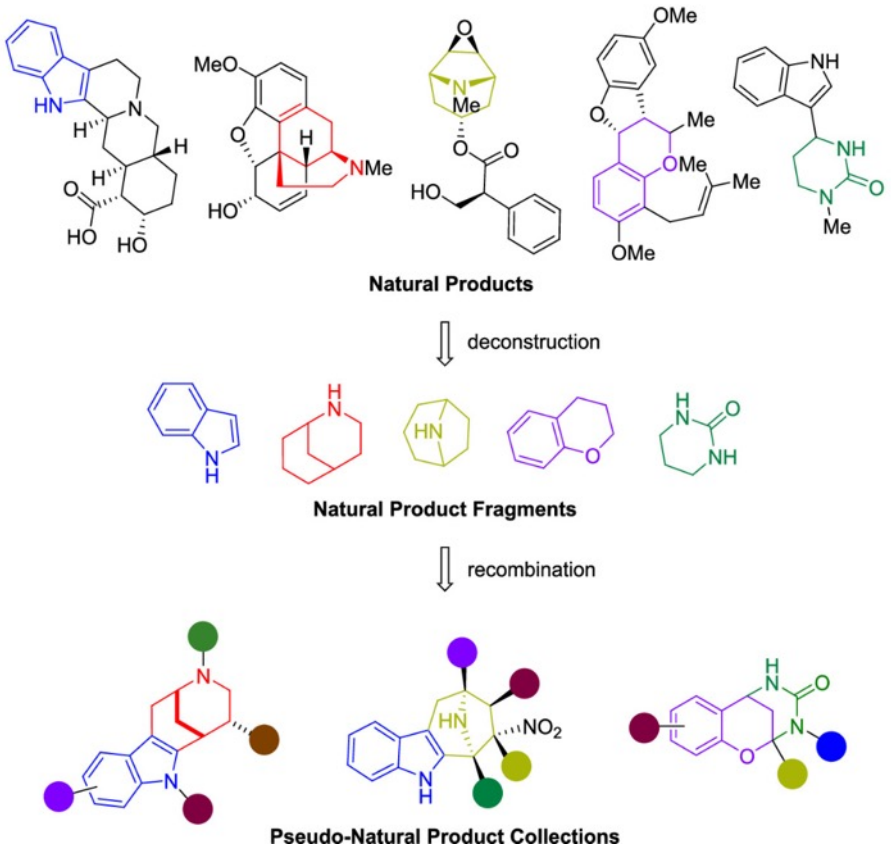
EMBL

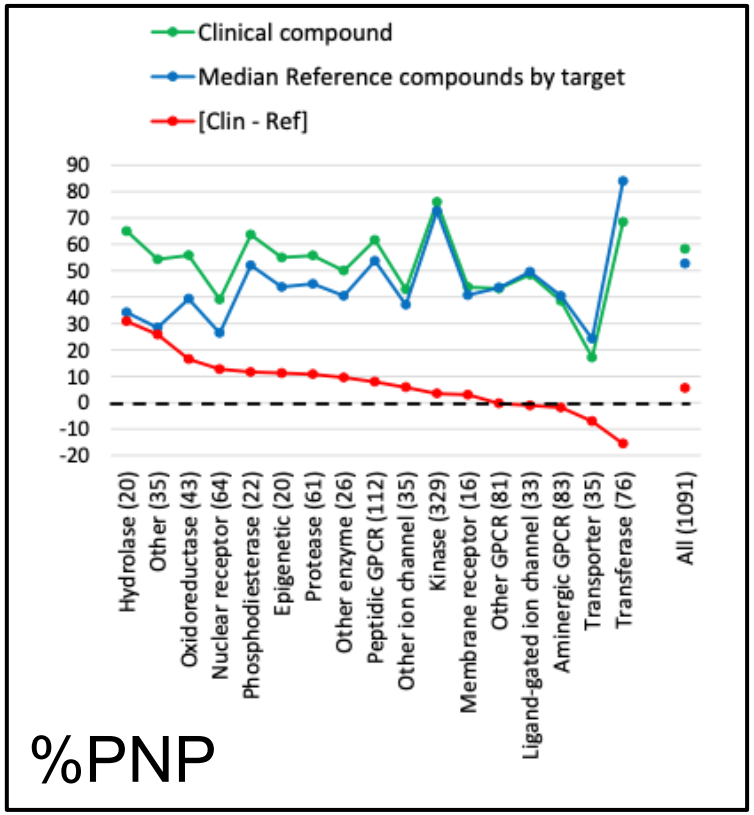
Robust & reliable datasets for community use



Heinzke, Zdrazil et al Scientific Data (in press) and ChemRxiv. (doi:10.26434/chemrxiv-2024-vj70m)

“Natural selection” in small molecule drug discovery

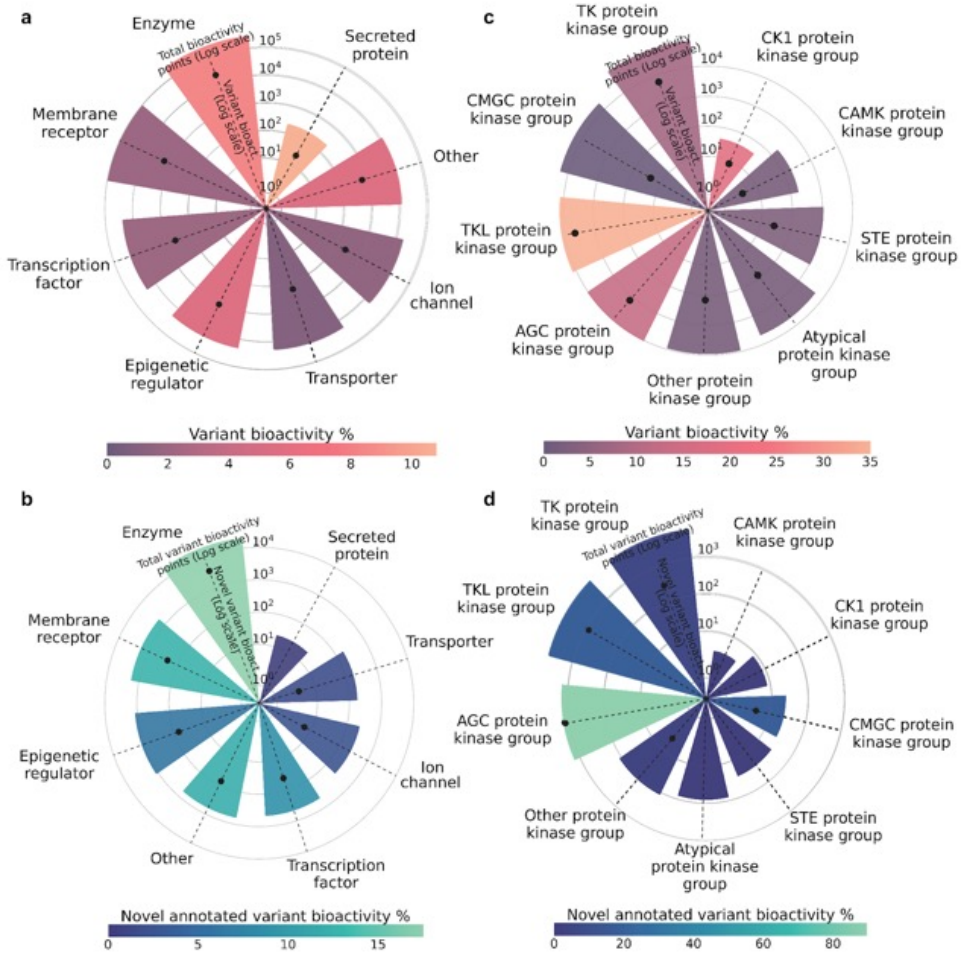


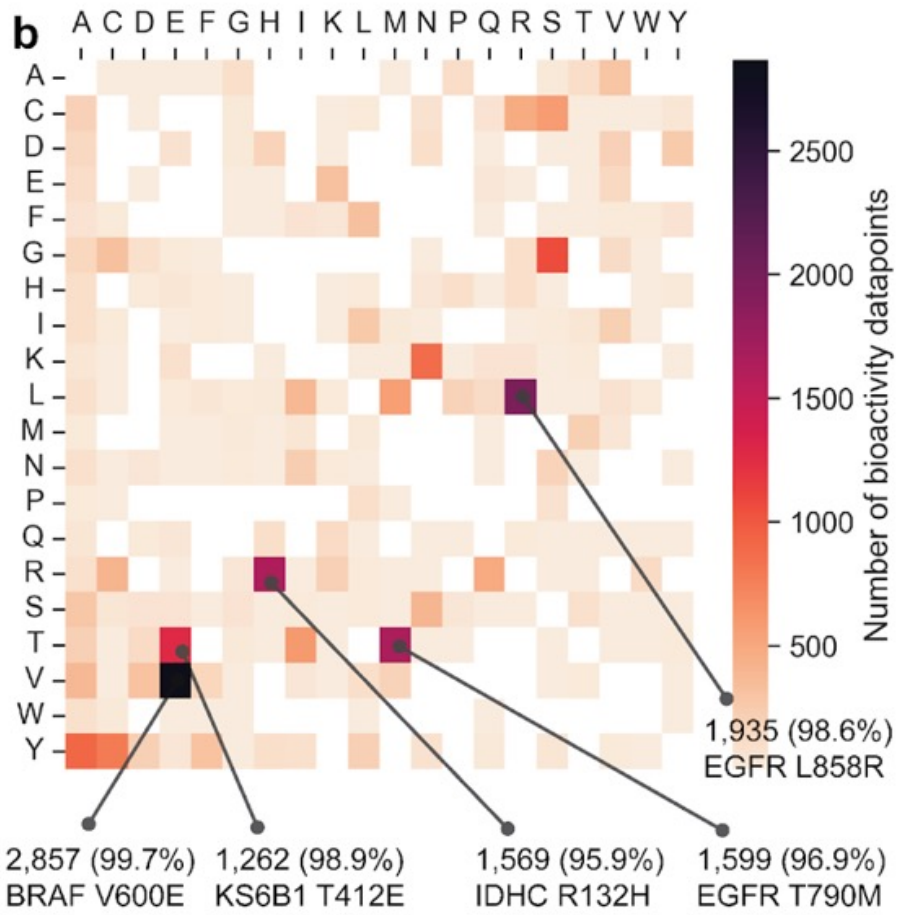


Heinzke et al J. Med. Chem. 2024, 67, 11226–11241

EMBL

Protein variants & bioactivity data (Gerard's talk)

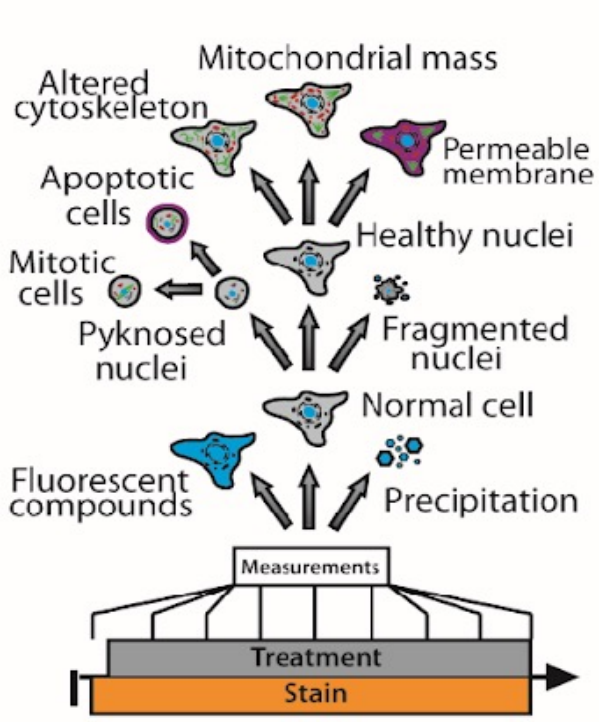




González et al, ChemRxiv. 2024; doi:10.26434/chemrxiv-2024-kxlgm

EMBL

ChEMBL underpins multiple external collaborations (Brian's talk)



Key data in ChEMBL, also visible via EUbOPEN gateway

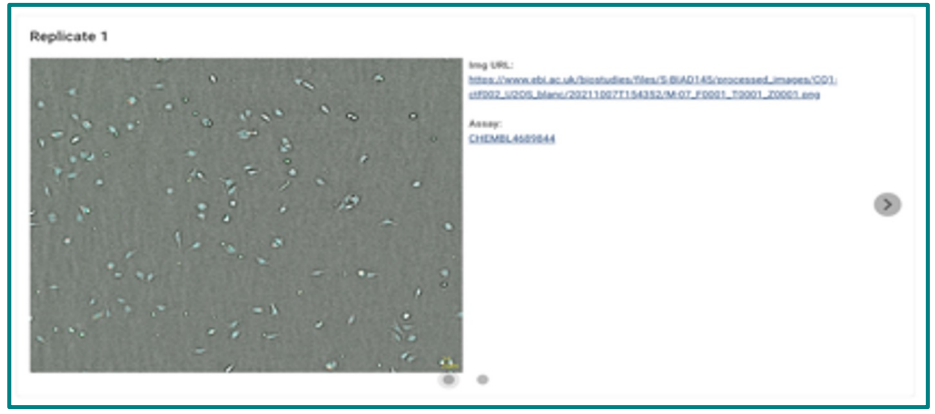


Plate images per time point & replicate. Hosted by the BioImage Archive.

Open Targets

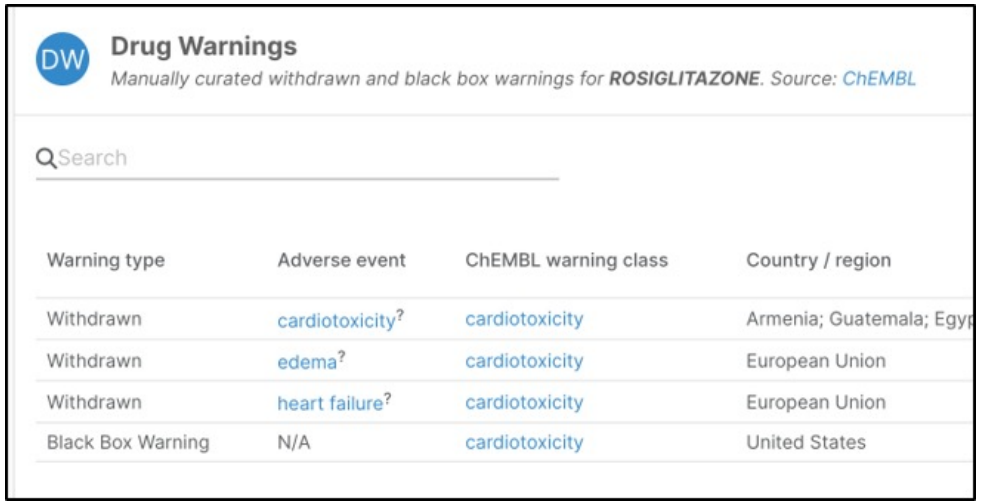
IDG Illuminating the Druggable Genome

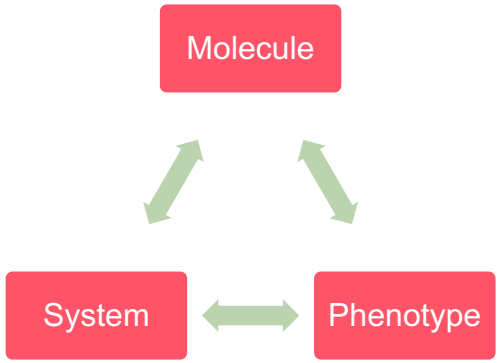
transQST

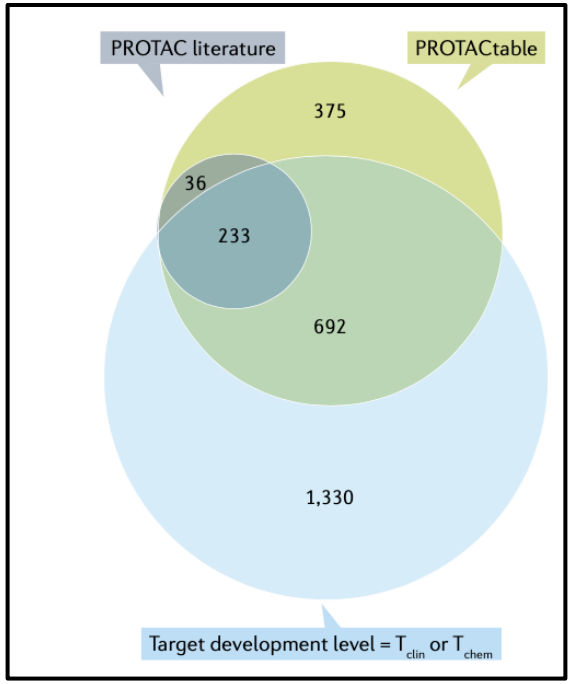
EUbOPEN Enabling & Unlocking Biology in the OPEN

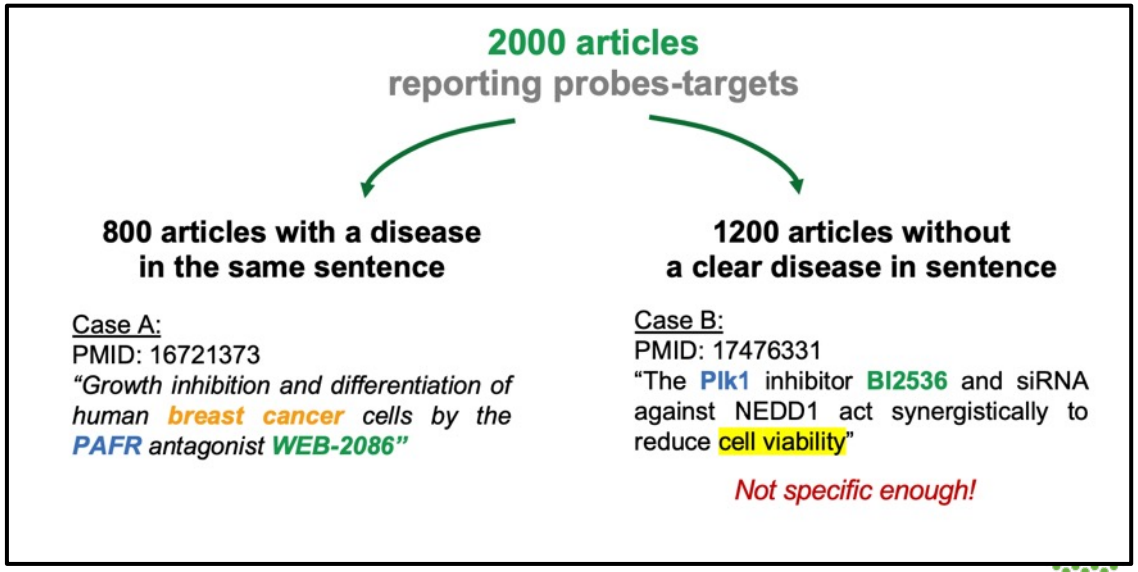
eTRANSAFE

Open Targets is a key strategic partner







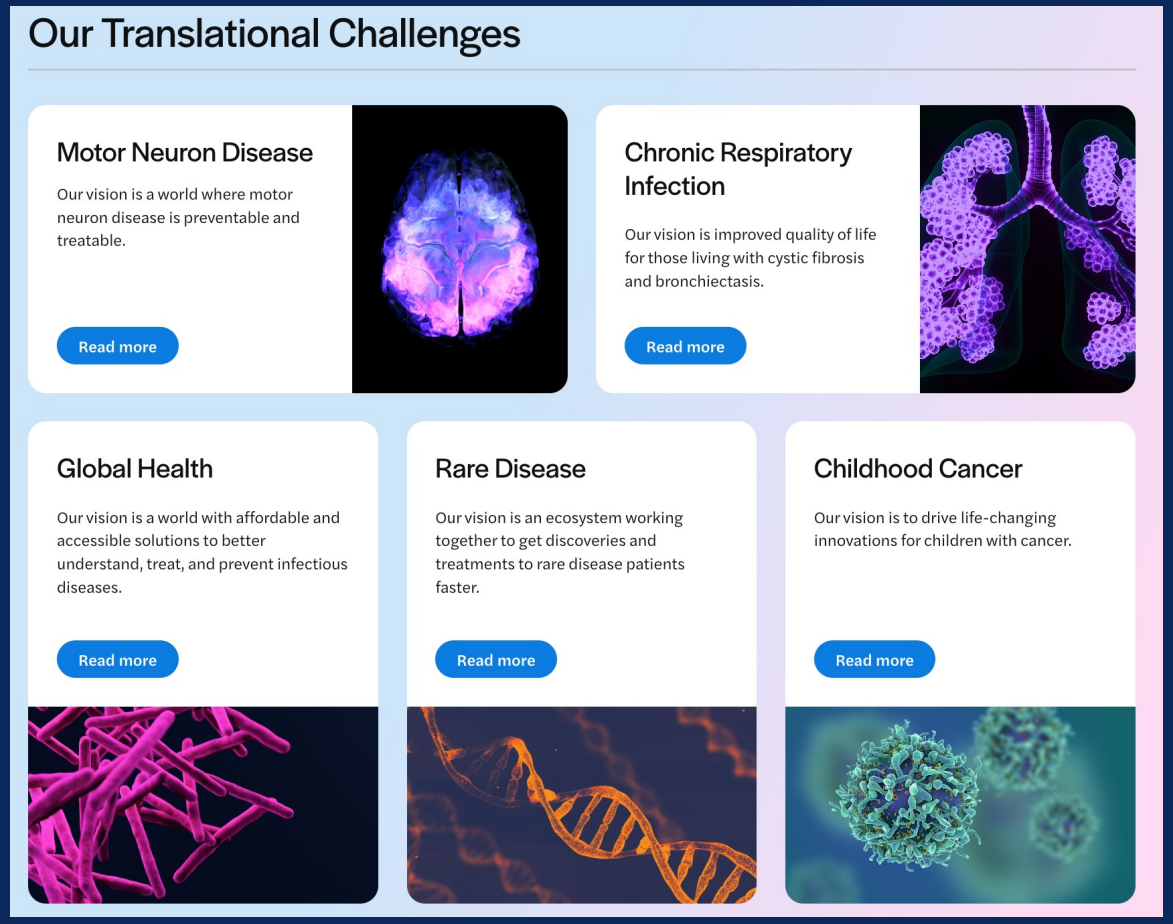


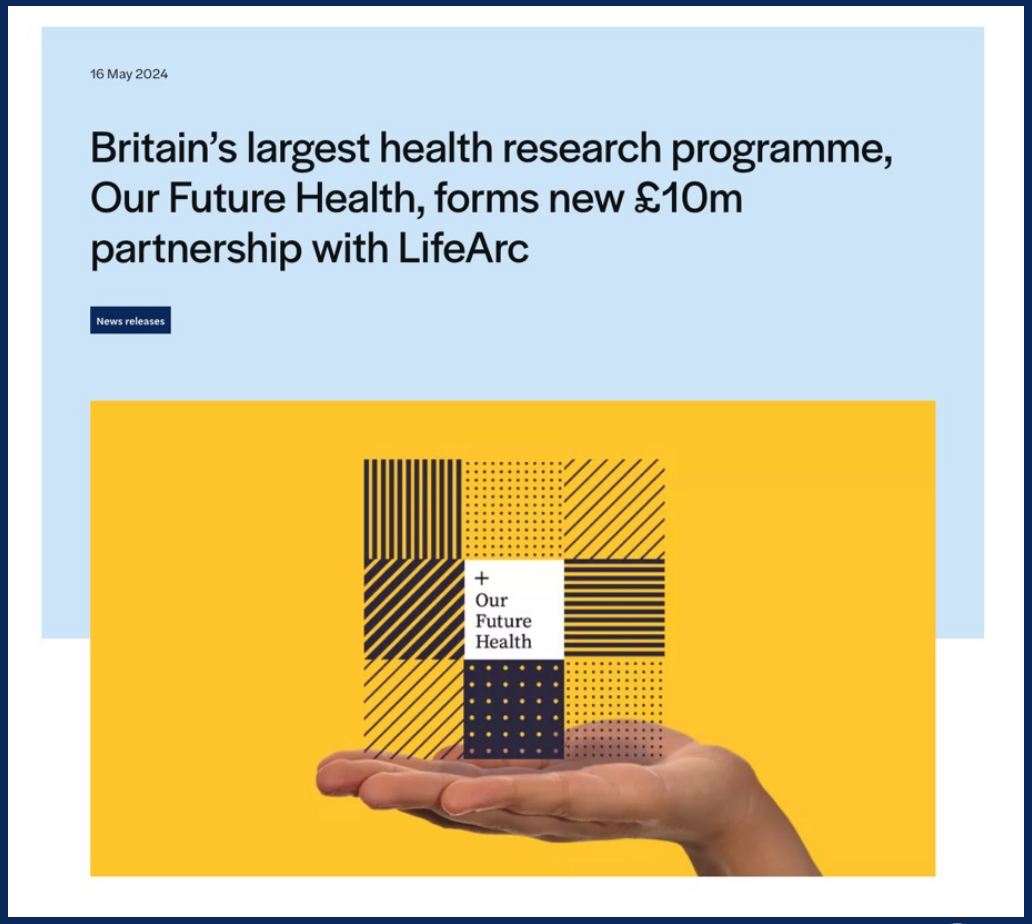
What will the next 15 years bring?

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

EMBL

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





Chemical Biology resources: funders & collaborators

NIH

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