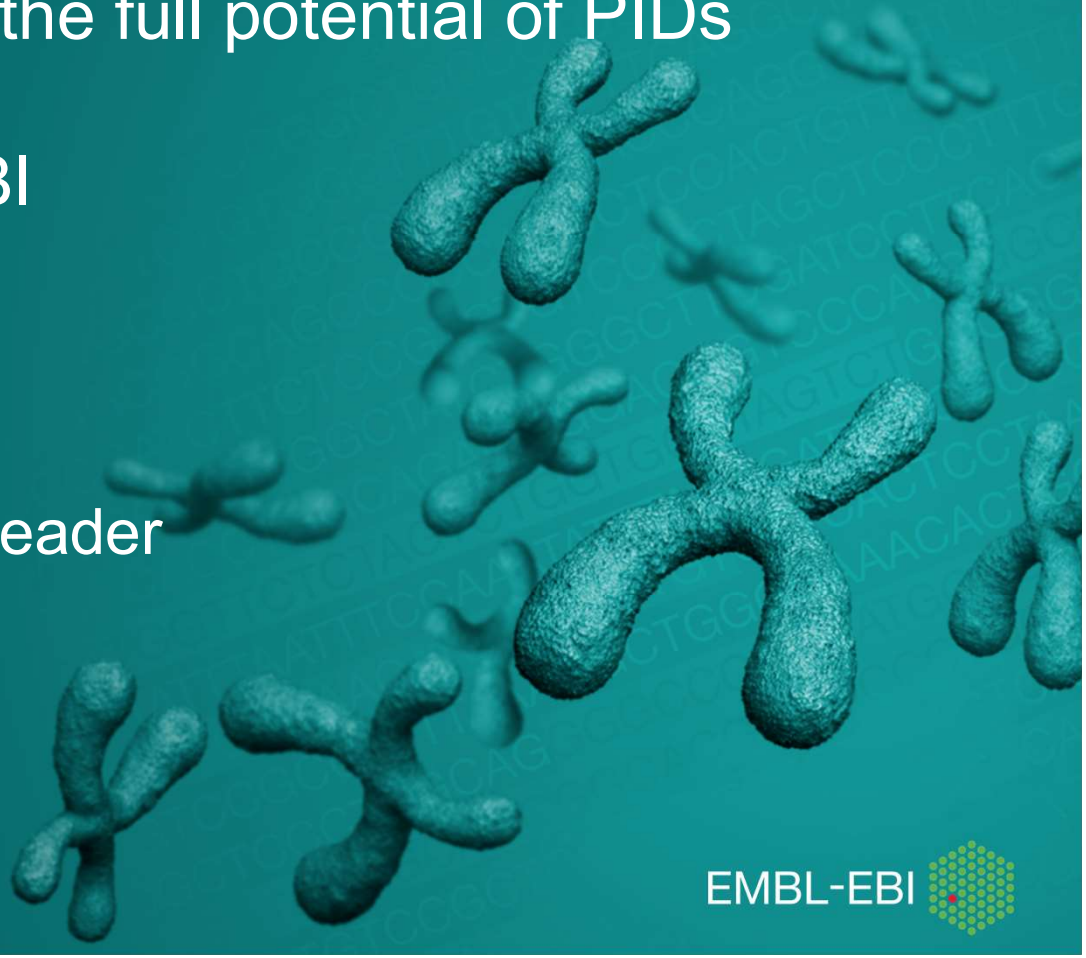
The European Bioinformatics Institute

On the way to realising the full potential of PIDs in BioSciences:
THOR work at EMBI-EBI

Robert Petryszak

Gene Expression Team Leader

EMBL-EBI



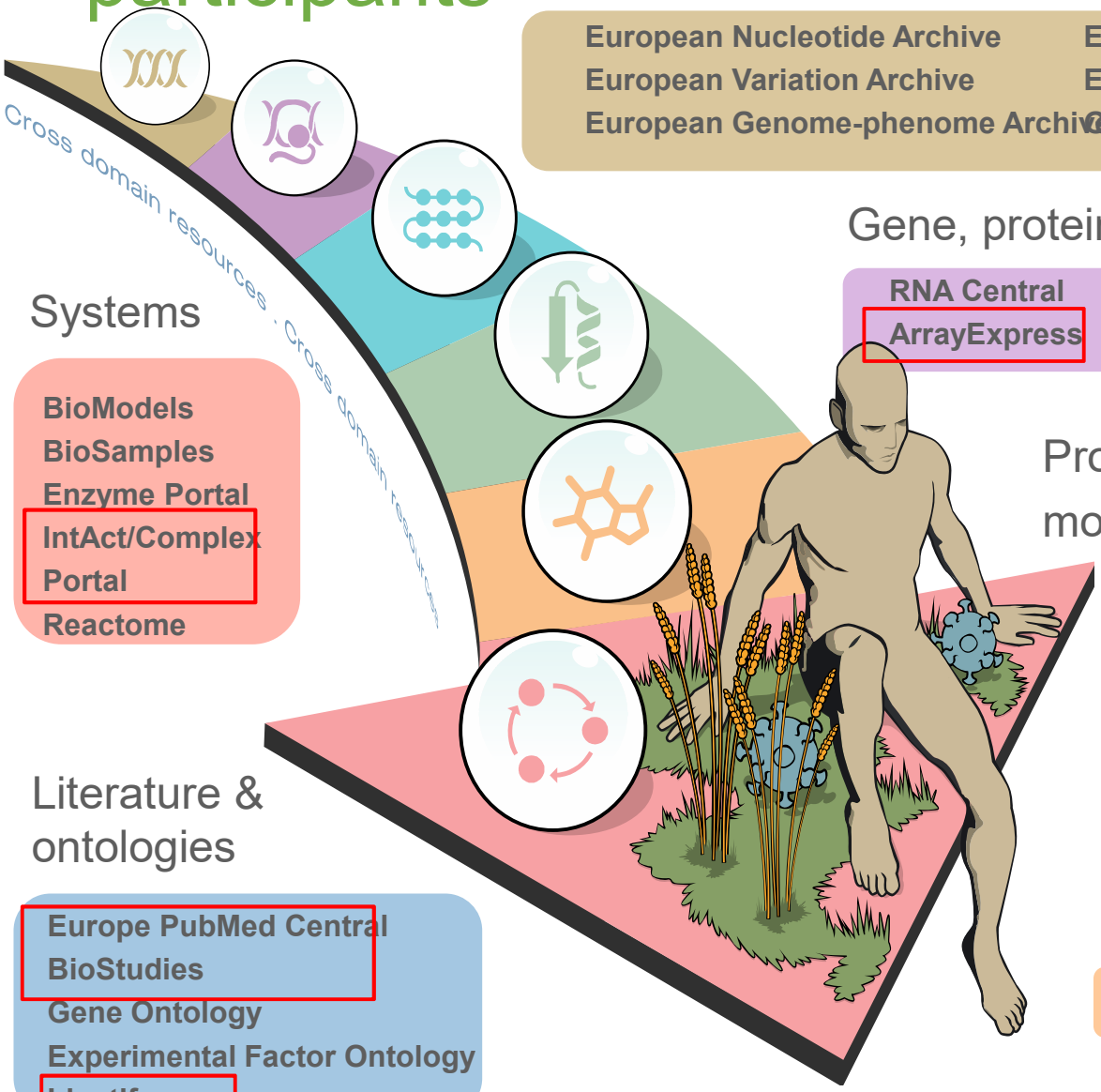
THOR

EMBL-EBI

What is EMBL-EBI?

- Europe's home for biological data services, research and training
- A trusted data provider for the life sciences – both academia and industry
- Part of the European Molecular Biology Laboratory, an intergovernmental research organisation
- International: 570 members of staff from 57 nations
- Home of the ELIXIR Technical hub

Data resources at EMBL-EBI – THOR participants



Genes, genomes & variation

European Nucleotide Archive
European Variation Archive
European Genome-phenome Archive
Ensembl
Ensembl Genomes
GWAS Catalog
Metagenomics portal

Gene, protein & metabolite expression

RNA Central
ArrayExpress
Expression Atlas
Metabolights
PRIDE

Protein sequences, families & motifs

InterPro
Pfam/Rfam
UniProt

Molecular structures

Protein Data Bank in Europe
Electron Microscopy Data Bank

Chemical biology

ChEMBL
SureChEMBL
ChEBI

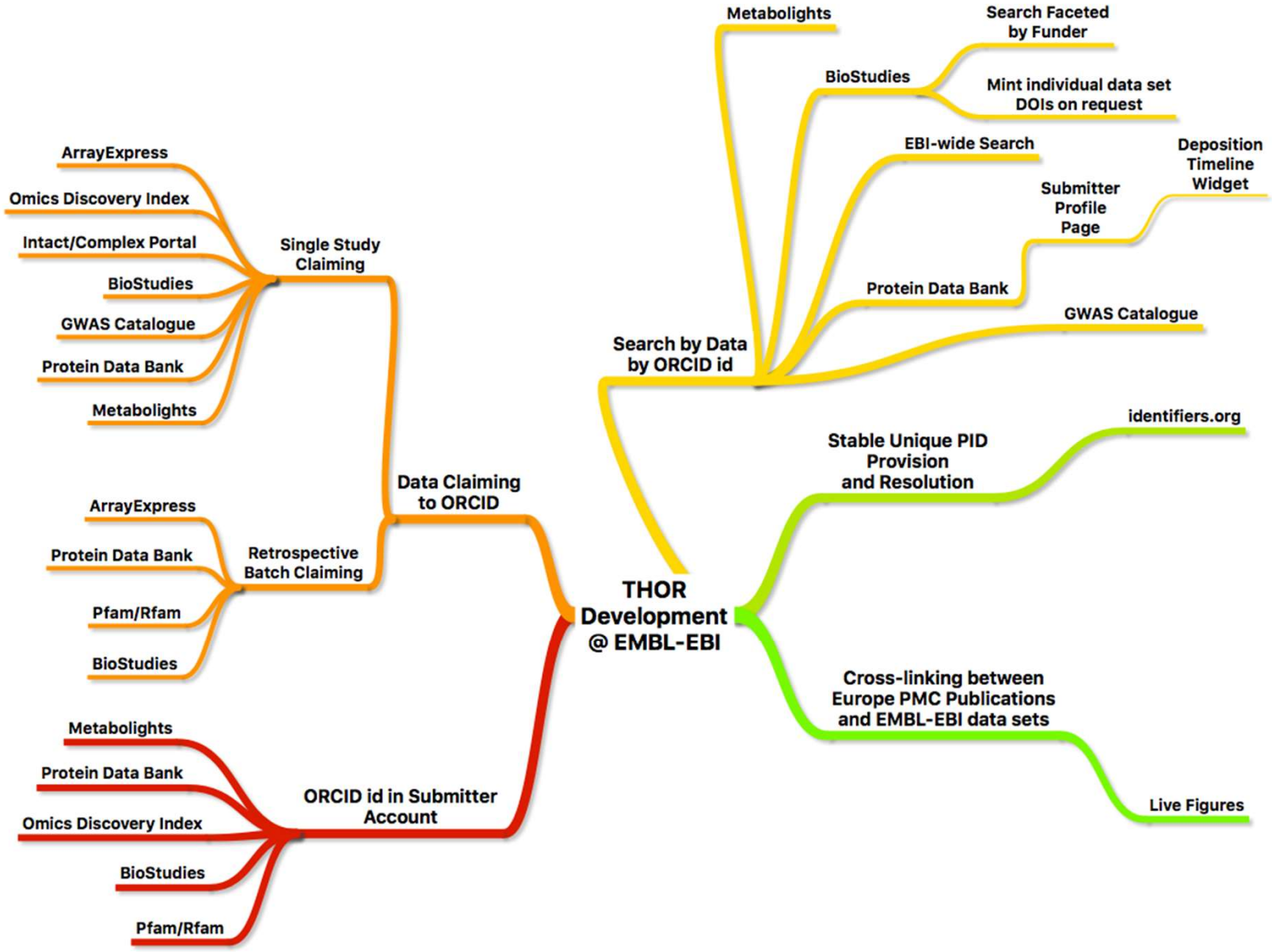
Systems

BioModels
BioSamples
Enzyme Portal
IntAct/Complex Portal
Reactome

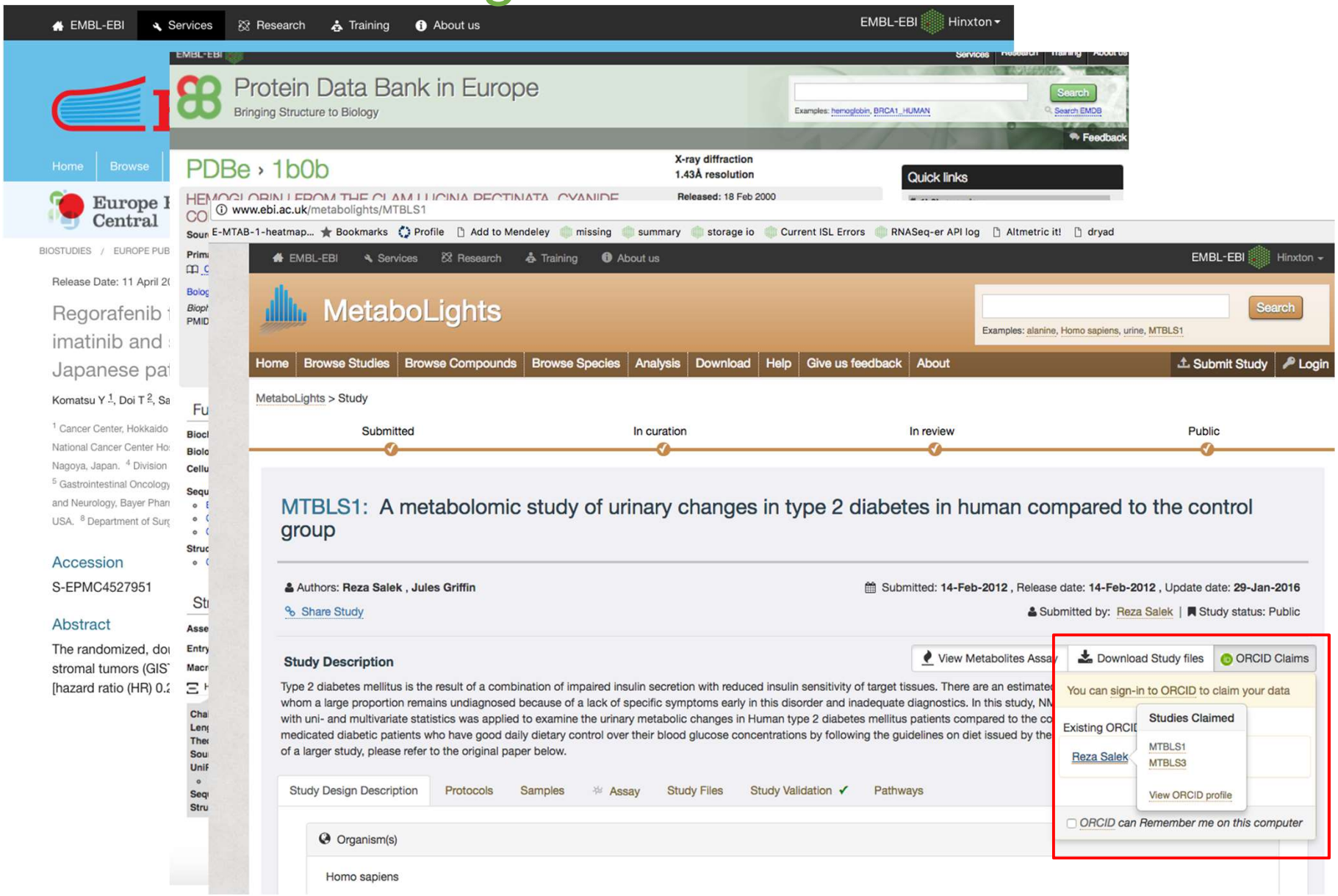
Literature & ontologies

Europe PubMed Central
BioStudies
Gene Ontology
Experimental Factor Ontology
Identifers.org

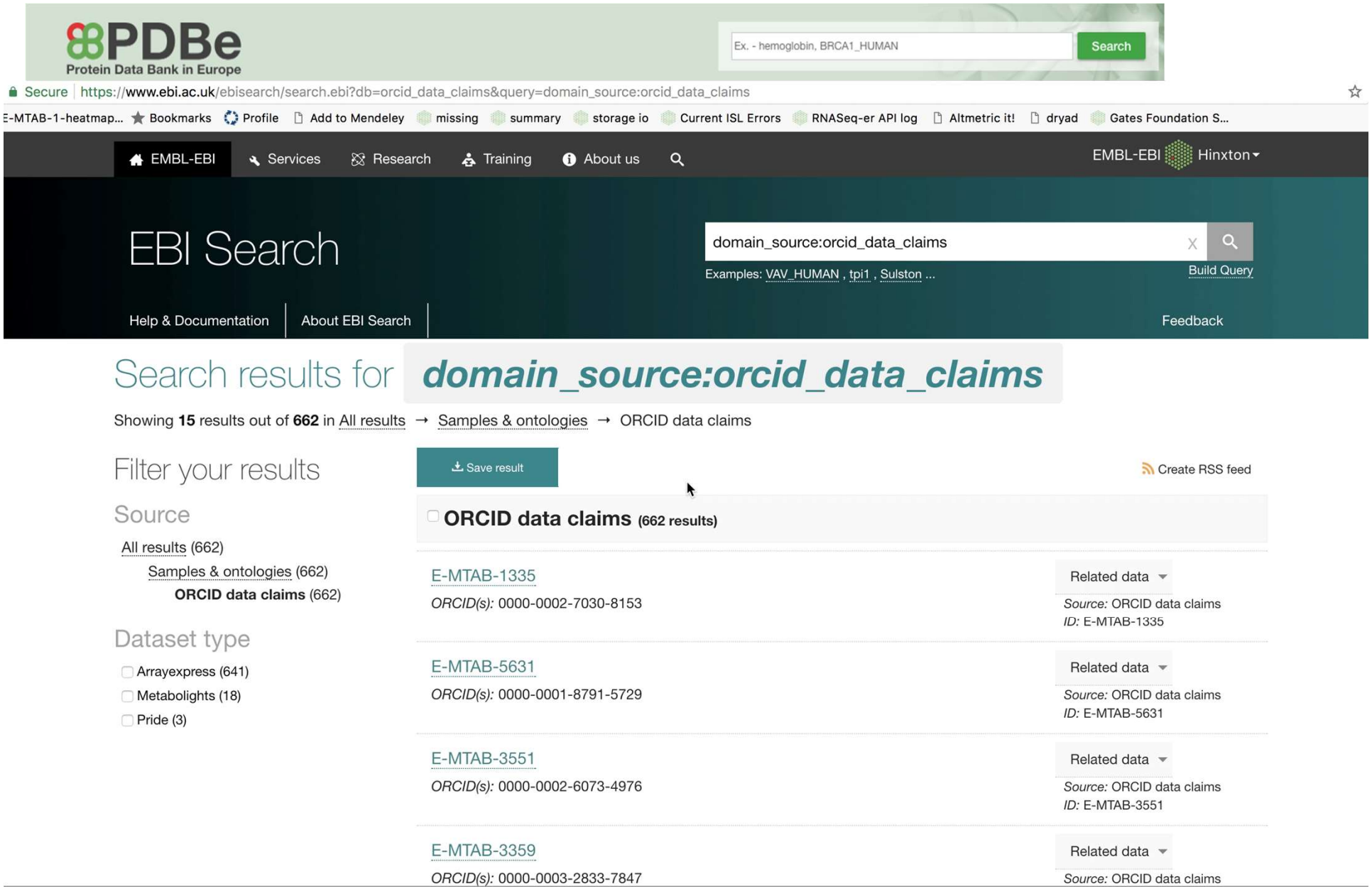
EMBL-EBI THOR Development – At a Glance



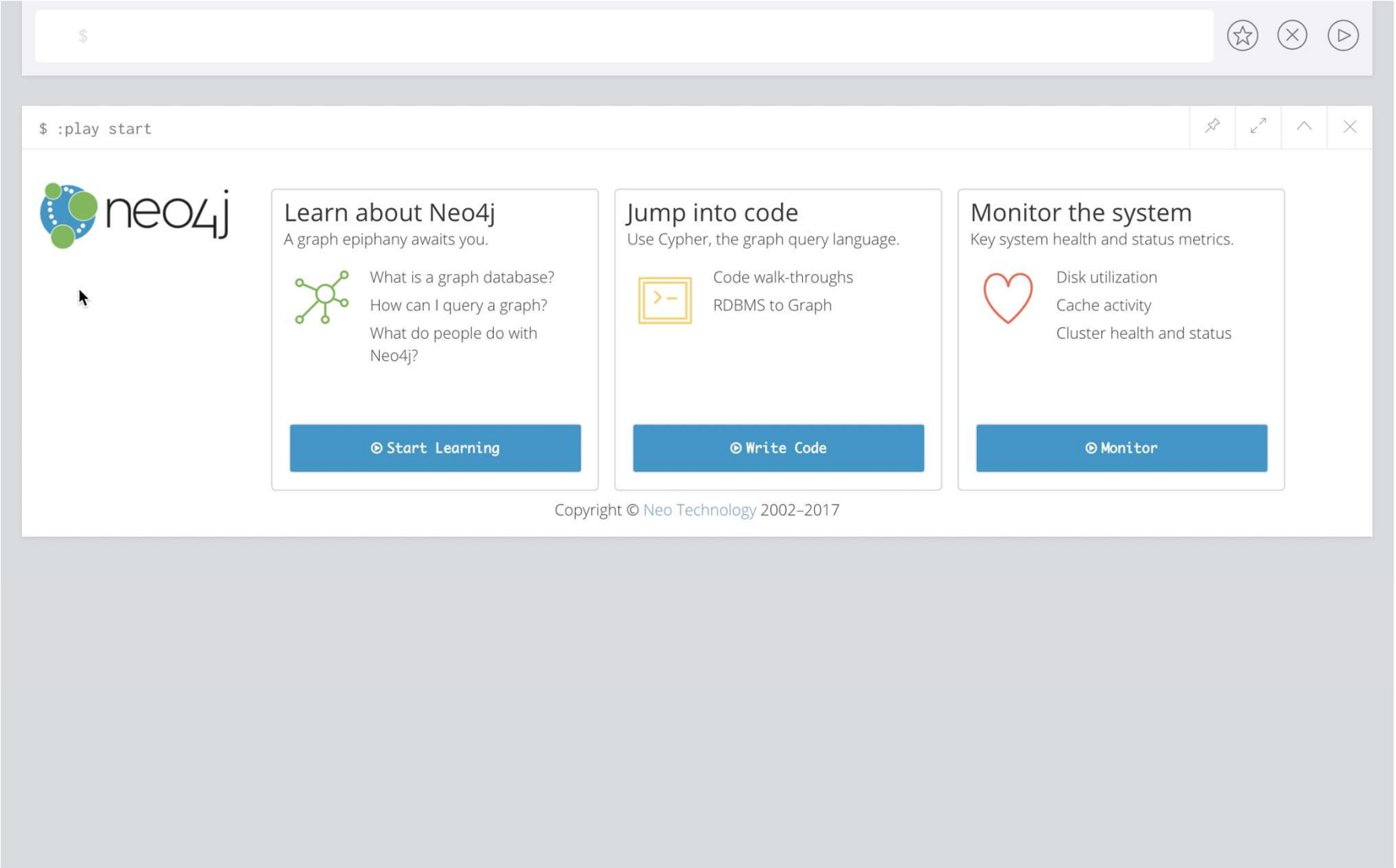
Claiming Individual Studies



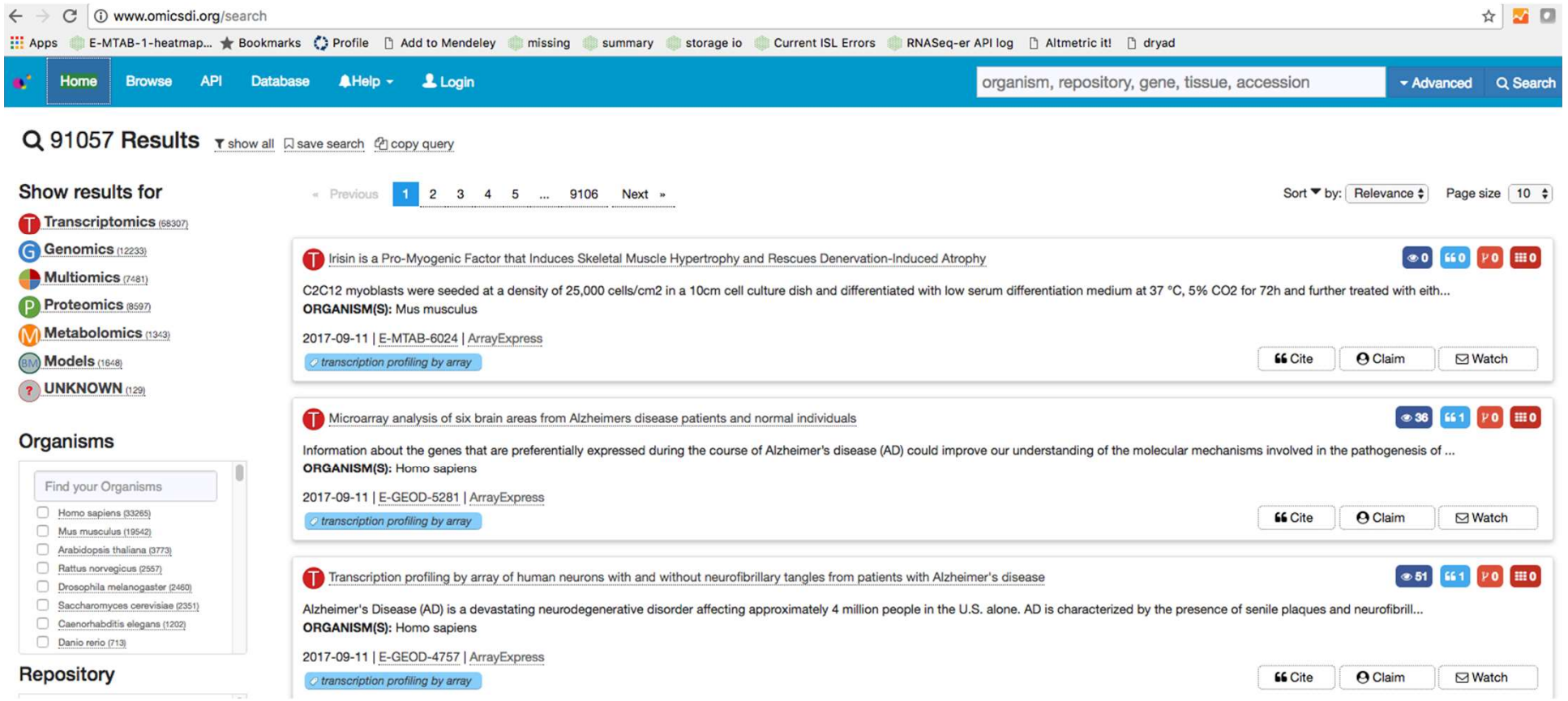
Batch Retrospective Claiming



Linked Research Visualisation Examples – Researchers – ArrayExpress studies -



Synergistic Efforts at EMBL-EBI
http://www.omicsdi.org/search



Synergistic Efforts at EMBL-EBI
http://www.omicsdi.org/search

MetaboLights
is a database for Metabolomics experiments and derived information.
286 datasets

MetabolomeExpress
A public place to process, interpret and share GC/MS metabolomics datasets.
58 datasets

GPMDB
The Global Proteome Machine Database was constructed to utilize the information obtained by GPM servers to aid in the difficult process of validating peptide MS/MS spectra as well as protein coverage patterns.
367 datasets

PAXDB
PaxDb contains estimated abundance values for a large number of proteins in several different species. Furthermore, you can find information about inter-species variation of protein abundances.
493 datasets

GNPS
The Global Natural Products Social Molecular Networking (GNPS) is a platform for providing an overview of the molecular features in mass spectrometry based metabolomics by comparing fragmentation patterns to identify chemical relationships.
572 datasets

LINCS
The Database contains all publicly available HMS LINCS datasets and information for each dataset about experimental reagents (small molecule perturbagens, cells, antibodies, and proteins) and experimental and data analysis protocols.
350 datasets

EGA
allows you to explore datasets from genomic studies, provided by a range of data providers.
Access to datasets must be approved by the specified Data Access Committee (DAC).
5950 datasets

PeptideAtlas
is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments.
2365 datasets

ArrayExpress
ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.
70411 datasets

BioModels Database
BioModels Database is a repository of computational models of biological processes. Models described from literature are manually curated and enriched with cross-references.
1648 datasets

JPOST Repository
jPOSTrepo (Japan ProteOme STandard Repository) is a new data repository of sharing MS raw/processed data.
57 datasets

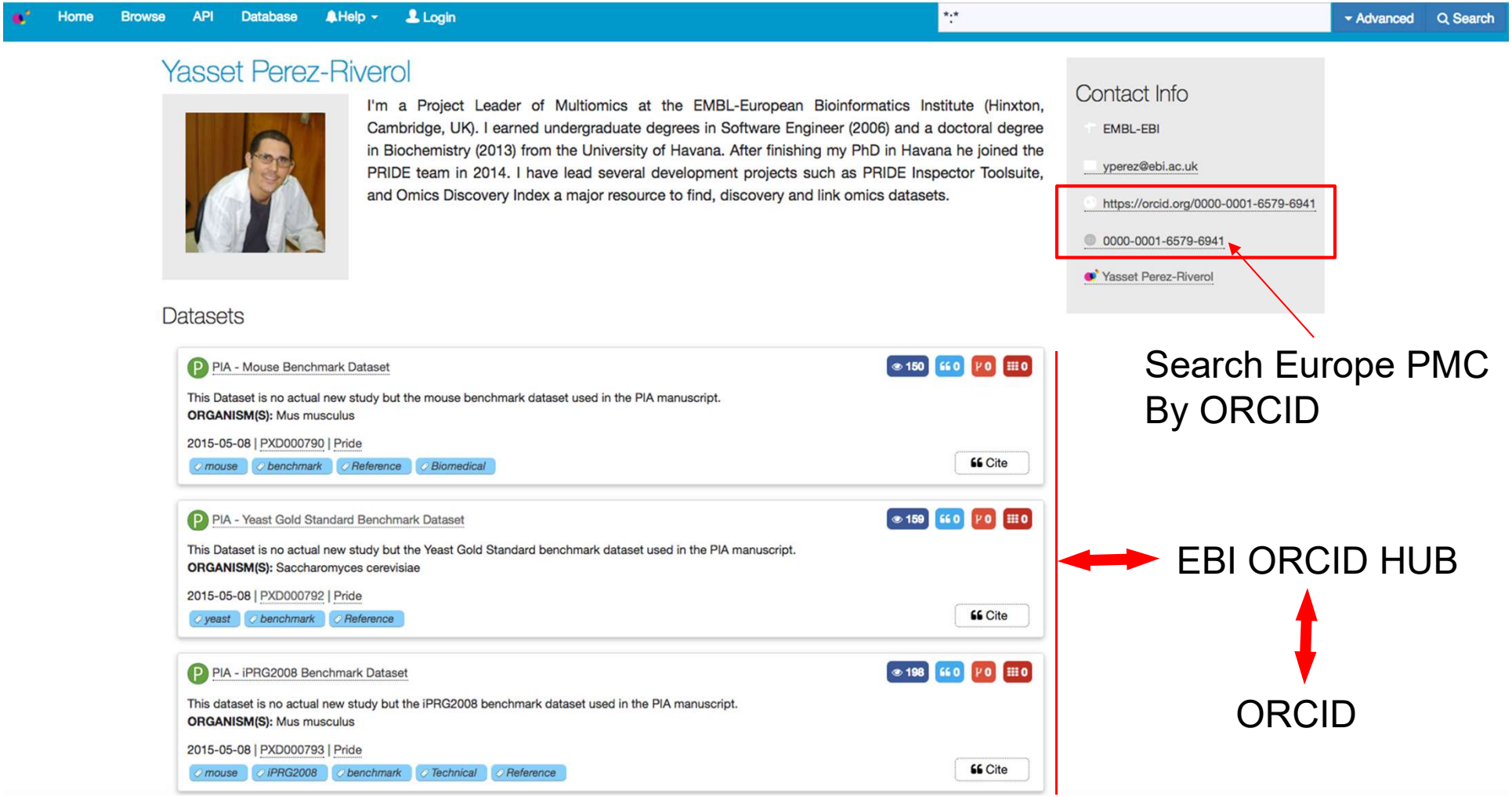
ExpressionAtlas
The Expression Atlas provides information on gene expression patterns under different biological conditions.
Gene expression data is re-analysed in-house to detect genes showing interesting baseline and differential expression patterns.
2913 datasets

Pride
is a centralized, standards compliant, public data repository for proteomics data, including protein and peptide identifications, post-translational modifications and supporting spectral evidence.
4128 datasets

MetabolomicsWorkbench
is a scalable and extensible informatics infrastructure which will serve as a national metabolomics resource.
425 datasets

Massive
is a community resource developed by the NIH-funded Center for Computational Mass Spectrometry to promote the global, free exchange of mass spectrometry data.
1034 datasets

Synergistic Efforts at EMBL-EBI
http://www.omicsdi.org/search

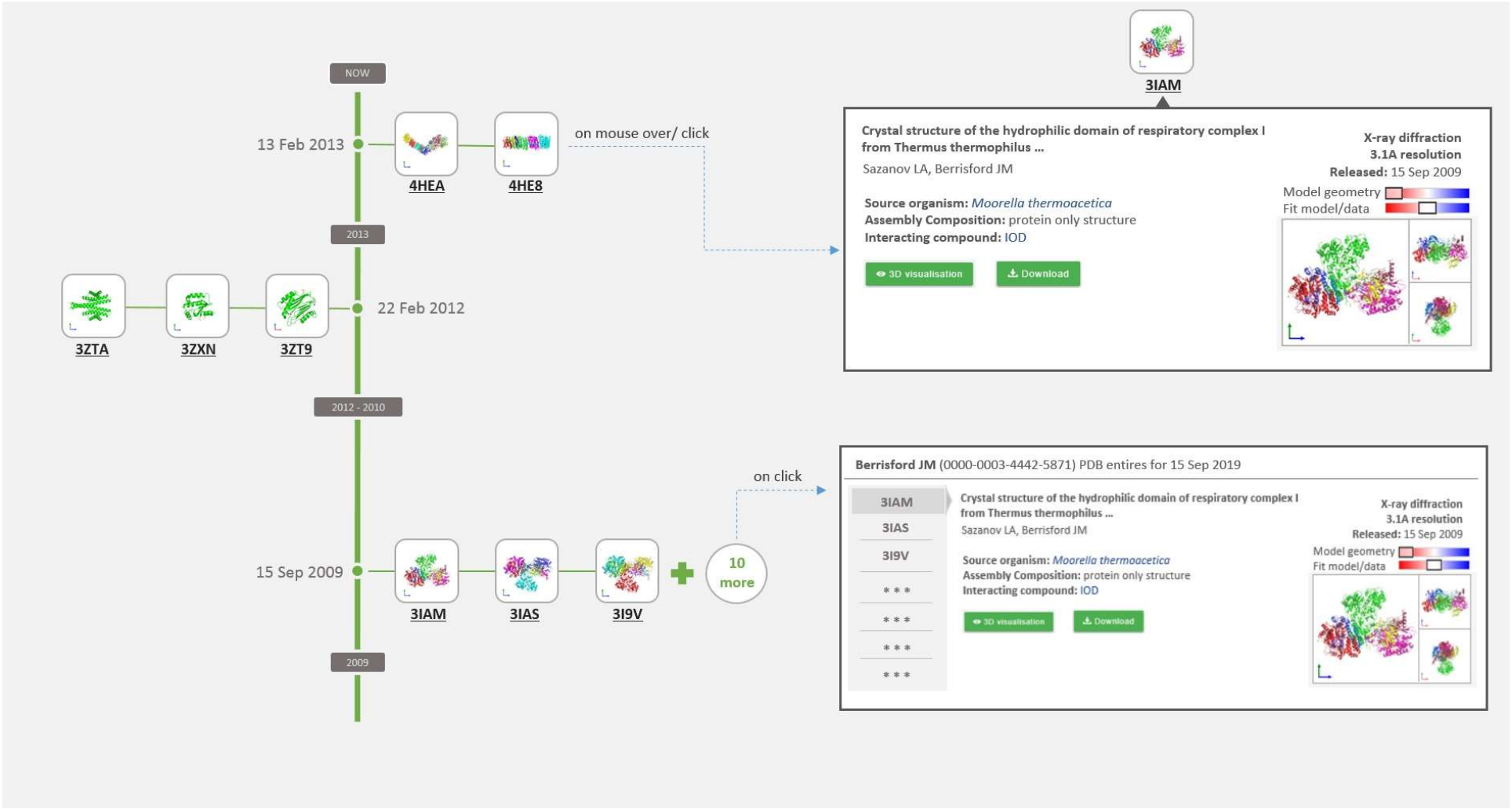


Search Europe PMC
By ORCID

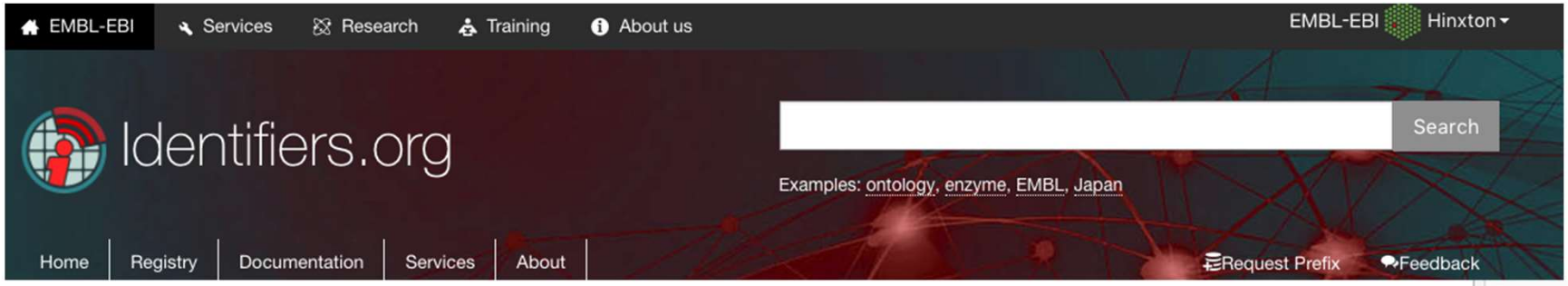
EBI ORCID HUB

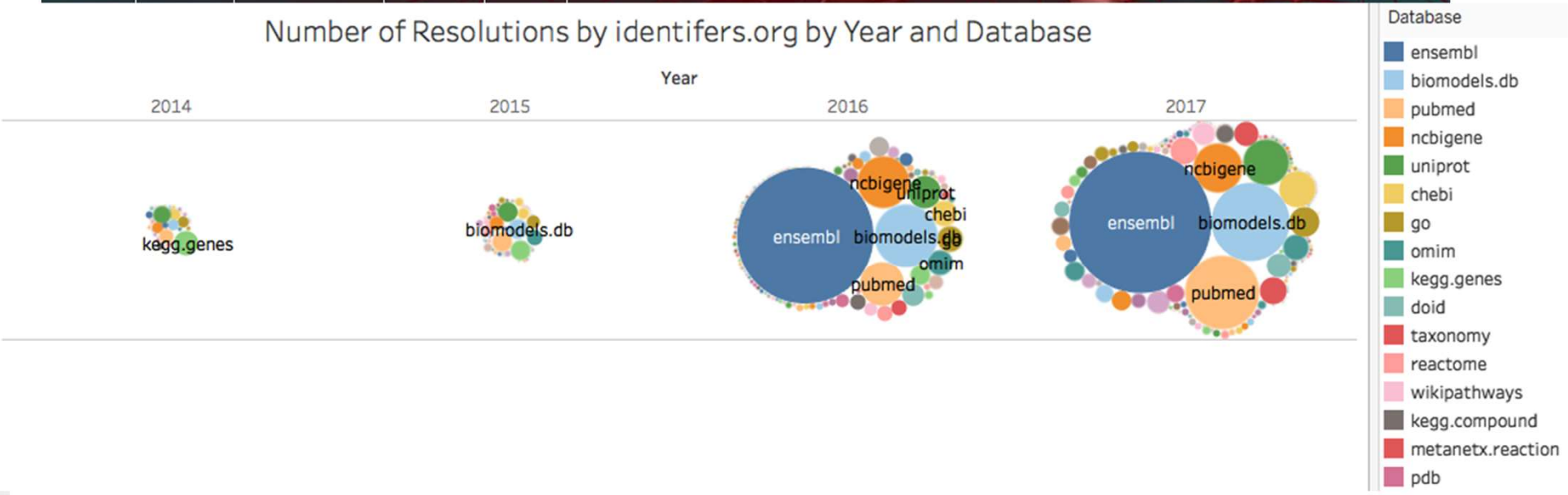
ORCID

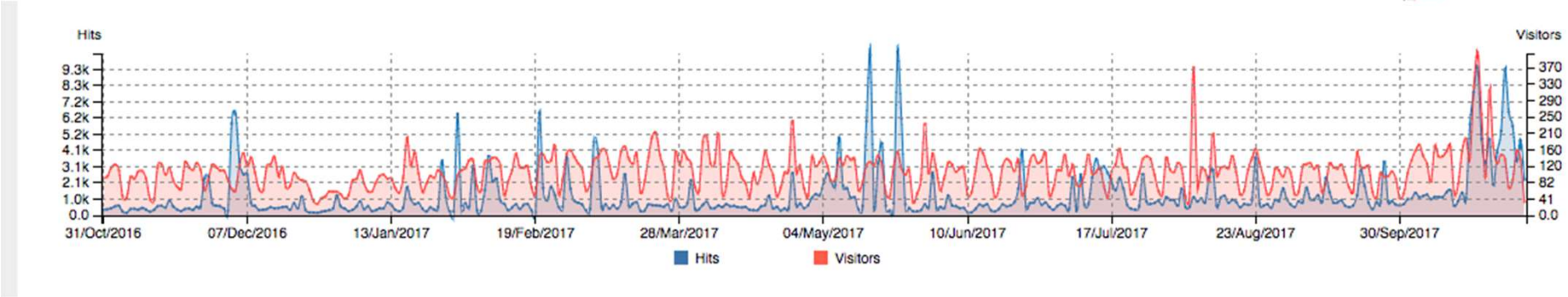
Protein Data Bank - Submitter Profile: Submission Timeline Widget



identifiers.org







Acknowledgements

- Jo McEntyre, Guilherme Mello, Florian Graef – Europe PMC
- Sarala Wilamaratne, Henning Hermjakob – identifiers.org
- EMBL-EBI teams implementing THOR functionality:
  - ArrayExpress, PDB,
  - PRIDE, OmicsDI,
  - Intact, BioStudies,
  - GWAS Catalogue, Pfam/Rfam,
  - EBI Search
- All my THOR collaborators