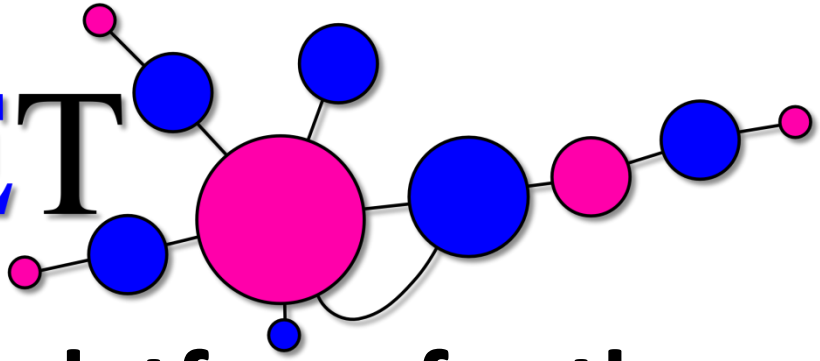
DisGeNET

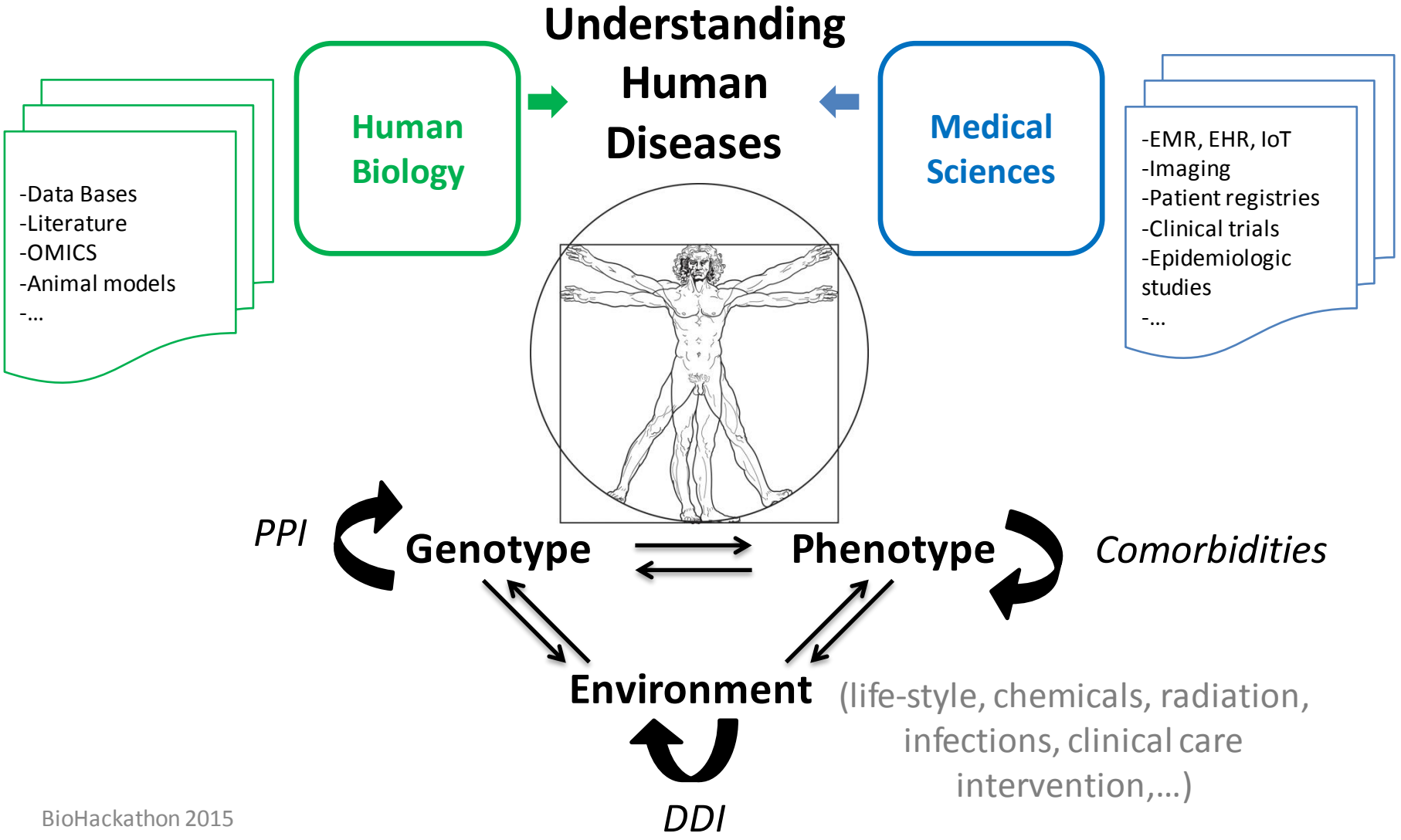


A discovery platform for the dynamical exploration of human diseases and their genes

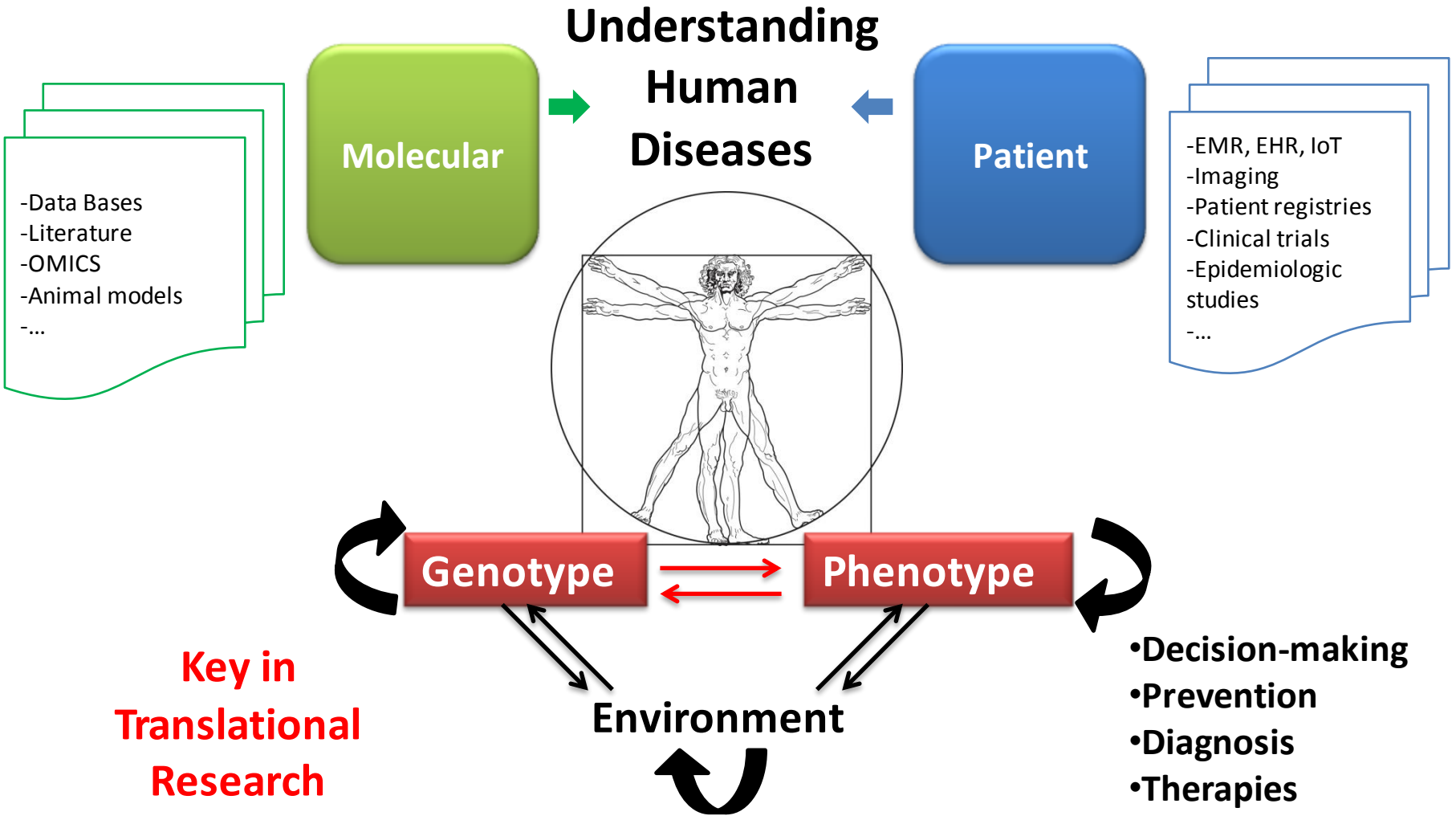
Núria Queralt Rosinach
Integrative Biomedical Informatics Group (IBI)
Research Programme on Biomedical Informatics (GRIB)
Hospital del Mar Research Institute (IMIM)
Pompeu Fabra University (UPF)
Barcelona

GRIB RESEARCH PROGRAMME ON BIOMEDICAL INFORMATICS
Hospital del Mar
Parc de Salut MAR Barcelona
upf. Universitat Pompeu Fabra Barcelona

Big Questions 4 Big Data



Translational Research



Access to Gene-Disease Associations

Mental retardation - ? - SOX3

SOX3

OMIM Online Mendelian Inheritance in Man — OMIM:300123; OMIM:312000

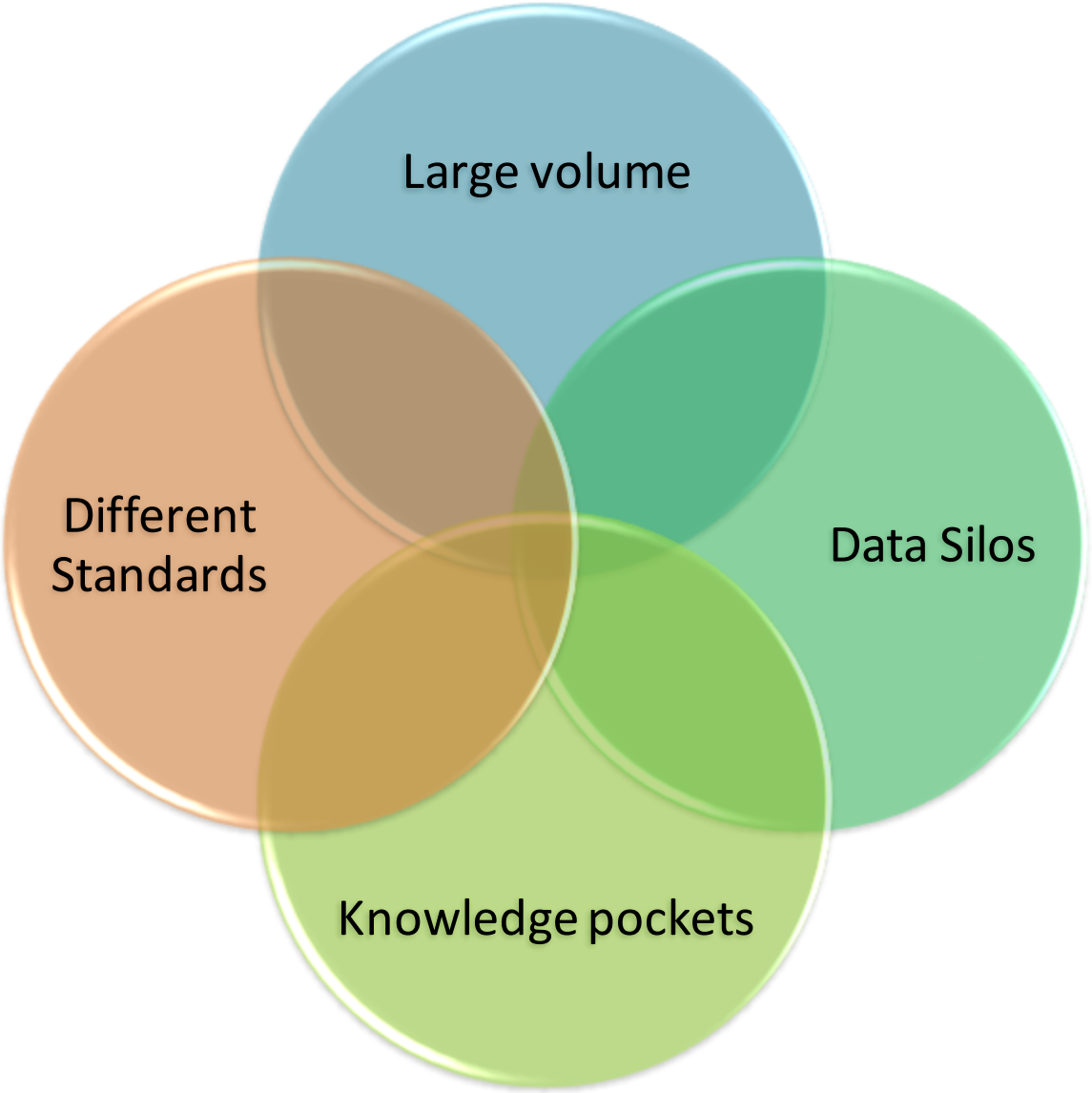
orphanet — ORPHA393; ORPHA90695; ORPHA3157; ORPHA79495; ORPHA67045

UniProt — Mental Retardation; Panhypopituitarism; 46,XX sex reversal 3

TATATCT ACCTCAC ClinVar — No Data

ctd™ — MESH:C538613; MESH:C538613

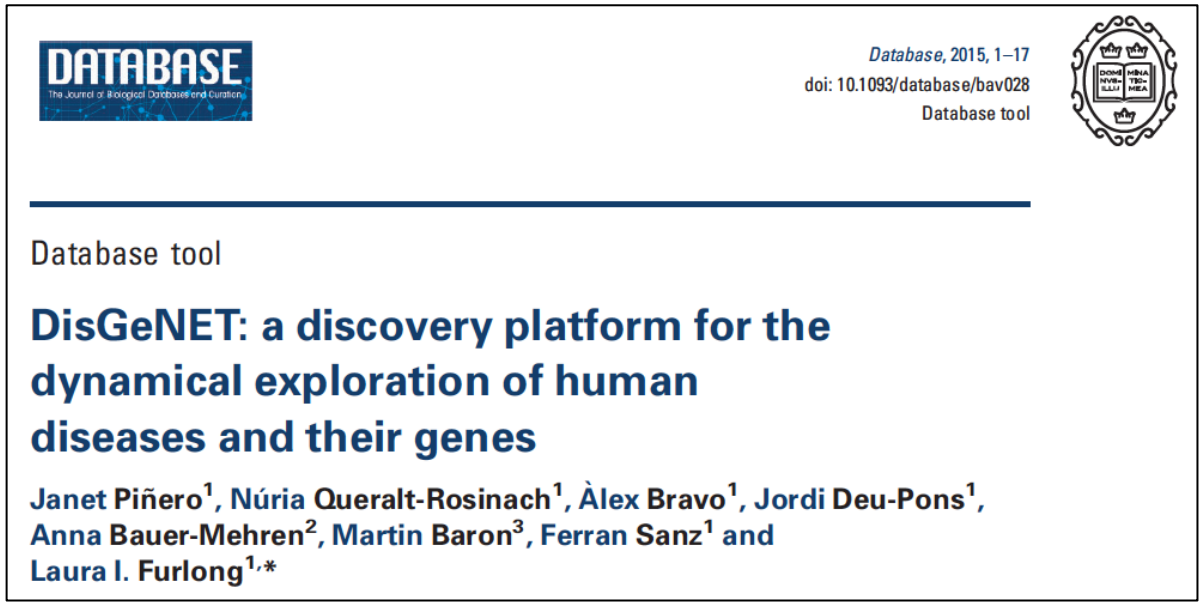
Access to Gene-Disease Associations



DisGeNET

http://www.disgenet.org/

- **Knowledge platform** on human **gene-disease associations (GDAs)**
- Integrates information from expert-**curated** databases and from the **literature** (text mining)
- All disease areas
- Supporting **evidence**



DATABASE
The Journal of Biological Databases and Curation

Database, 2015, 1–17
doi: 10.1093/database/bav028
Database tool

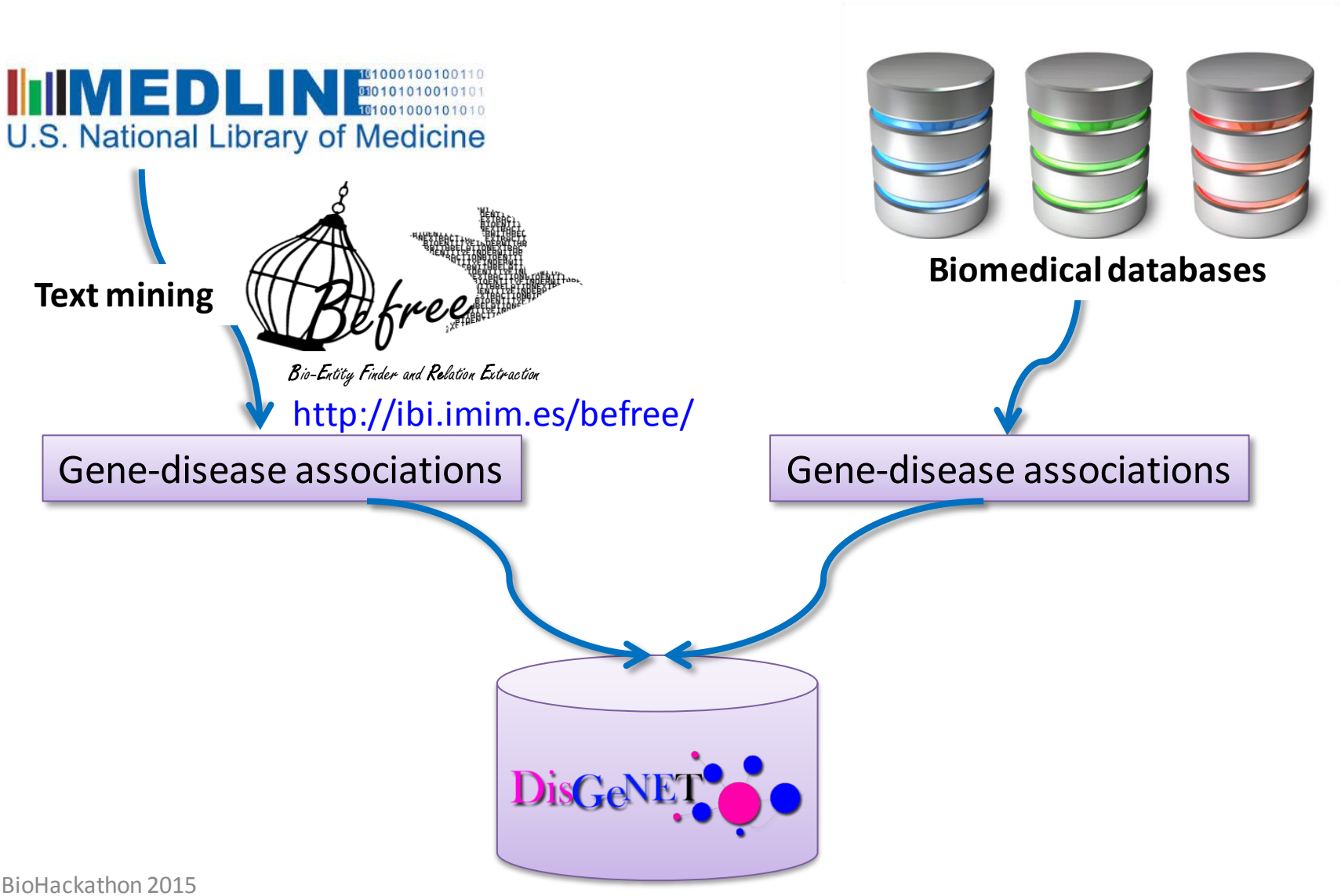
Database tool

DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes

Janet Piñero[1], Núria Queralt-Rosinach[1], Àlex Bravo[1], Jordi Deu-Pons[1], Anna Bauer-Mehren[2], Martin Baron[3], Ferran Sanz[1] and Laura I. Furlong[1,*]

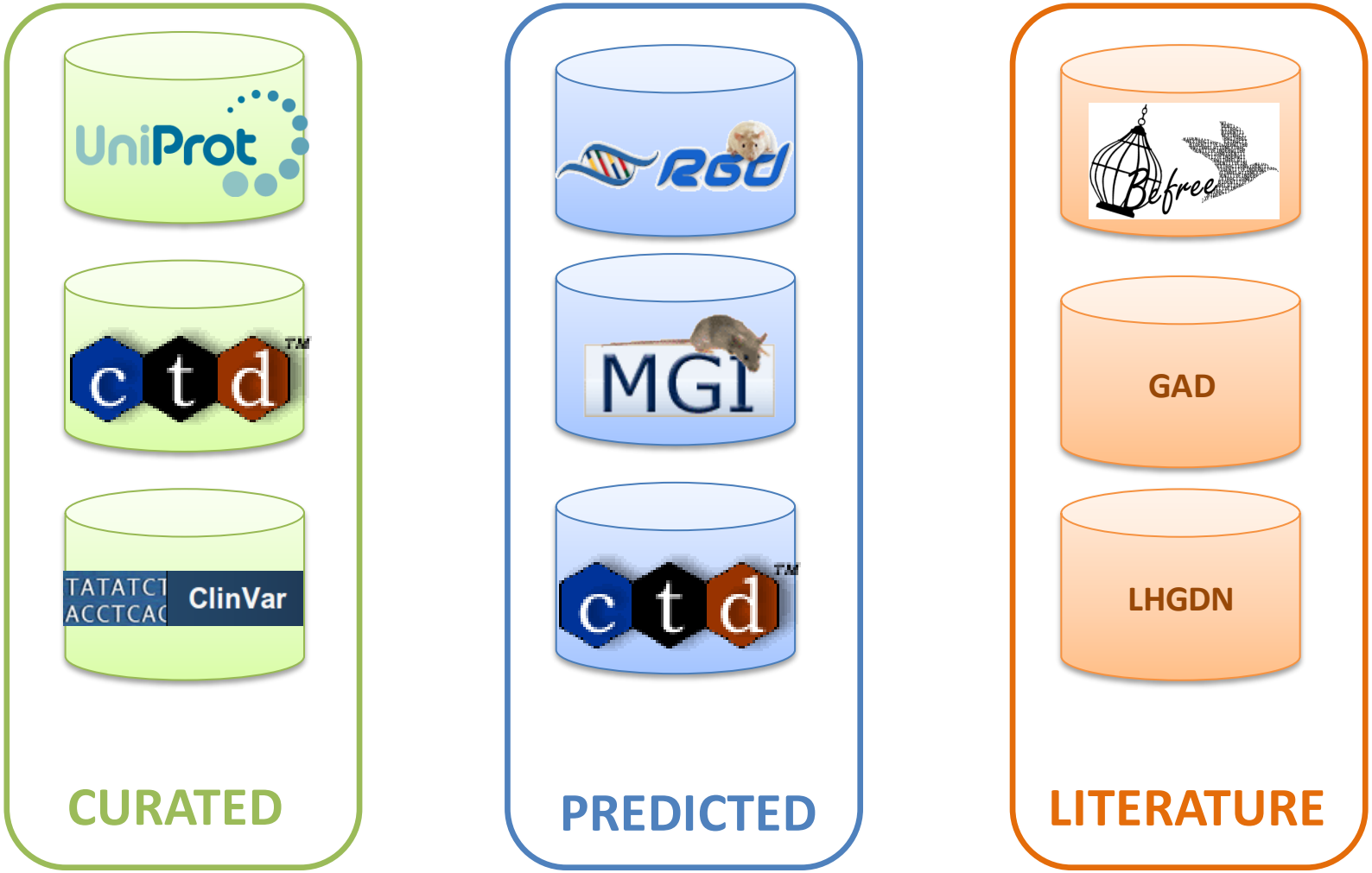
•Piñero *et al*. **DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes**. Database (2015) Vol. 2015: article ID bav028, (2015)

DisGeNET Implementation



DisGeNET Sources

DisGeNET v3.0



DisGeNET Statistics (May 15th, 2015)

DisGeNET v3.0

Source	Genes	Diseases	Associations
Curated	7,878	6,761	26,522
Predicted	2,557	2,003	9,536
Literature	16,298	11,374	408,175
All	**17,181**	**14,619**	**429,111**

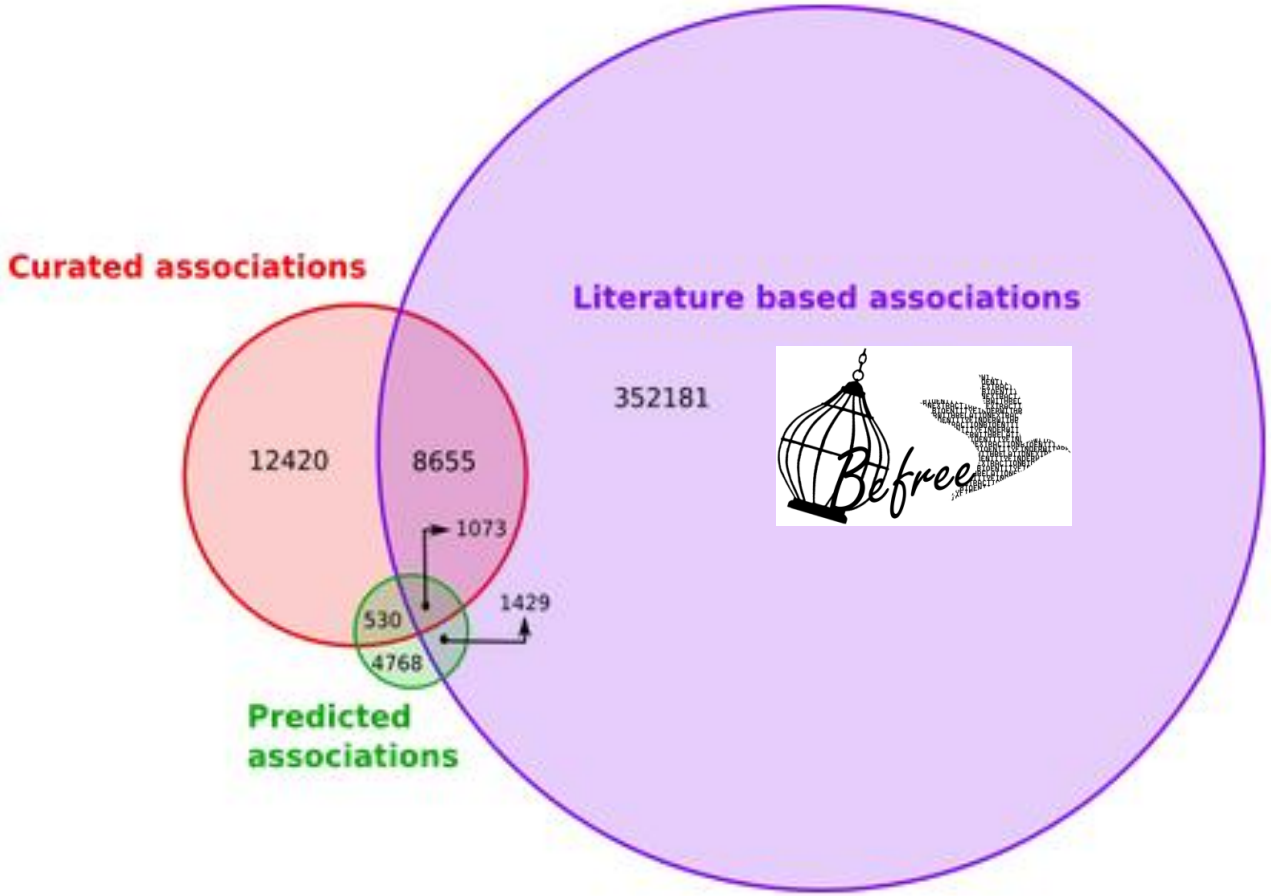
82 %

Large volume of information unlocked by text mining the literature

Text Mining

DisGeNET v3.0

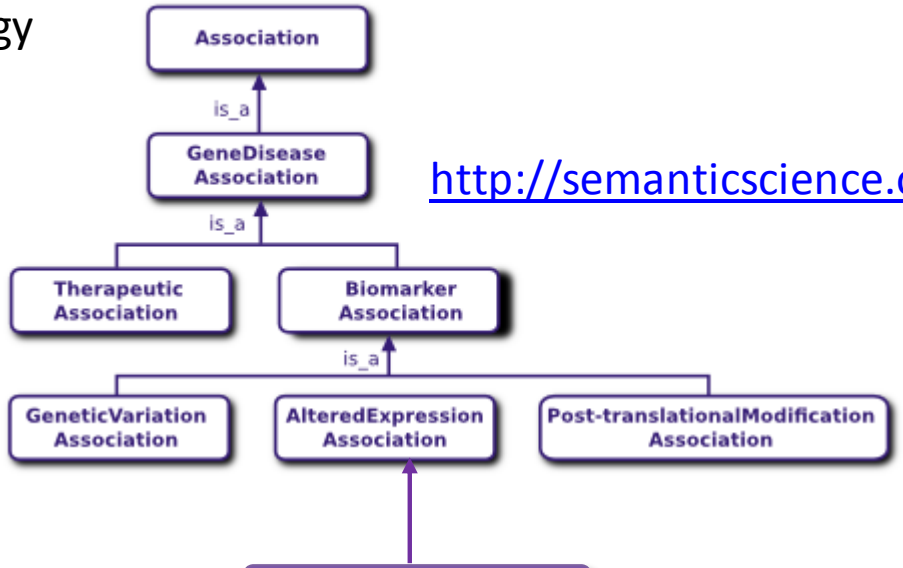
Little overlap between text mined GDAs and curated GDAs in DBs



Need
of
biocuration pipelines

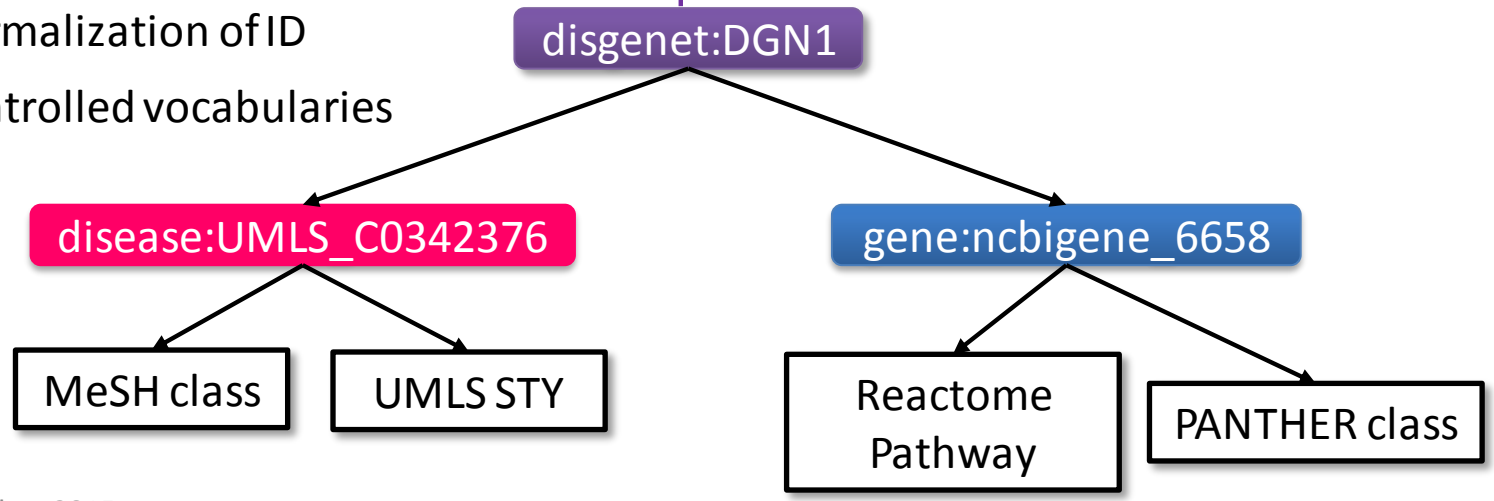
DisGeNET Standardization

- DisGeNET ontology (type of relation)



http://semanticscience.org/ontology/sio.owl

- Normalization of ID
- Controlled vocabularies



Data Integration

SOX3
NCBI:6658

Panhypopituitarism
UMLS:C0342376

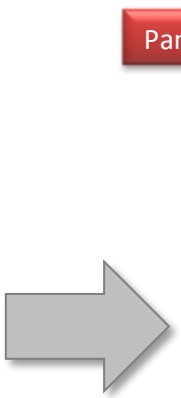
UniProt ✓

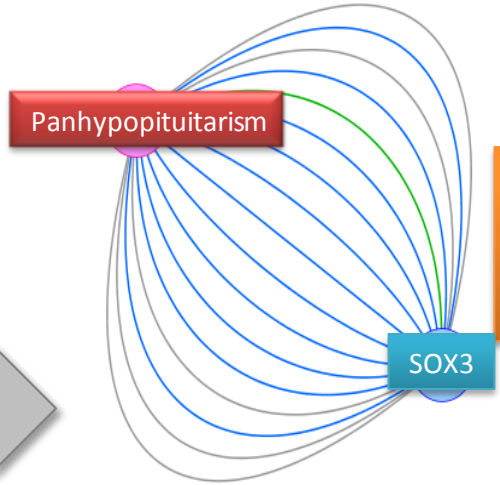
ctd ✓

MGI CELEBRATING 25 YEARS X

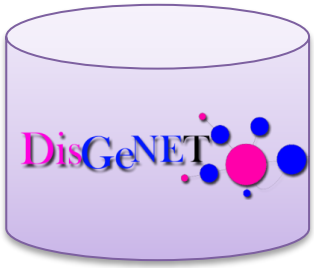
RGD X

MEDLINE U.S. National Library of Medicine X





Edge=EVIDENCE
(Source, PMID,
type of relation)

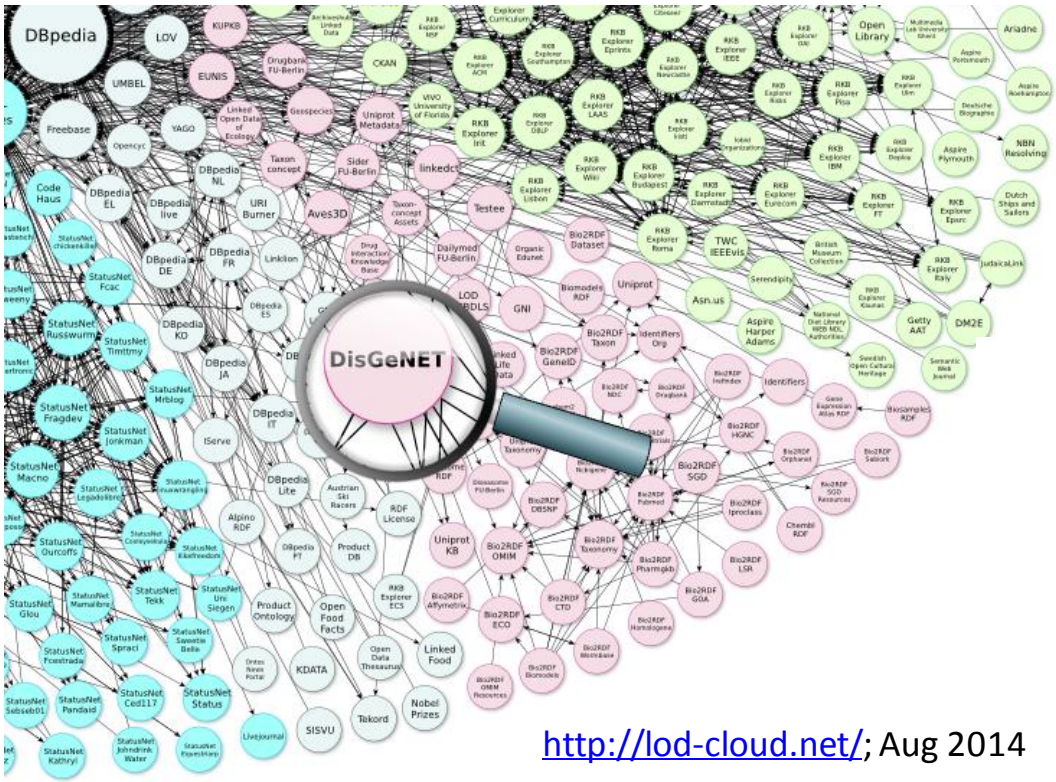


Score
SNP

DisGeNET as Linked Data

- **RDF** and **trusty nanopublications**
  - **URIs:** RDF providers or identifiers.org
  - **SIO**
  - Use of standards (**11 ontologies** in NCBO)

BioPortal



http://lod-cloud.net/; Aug 2014

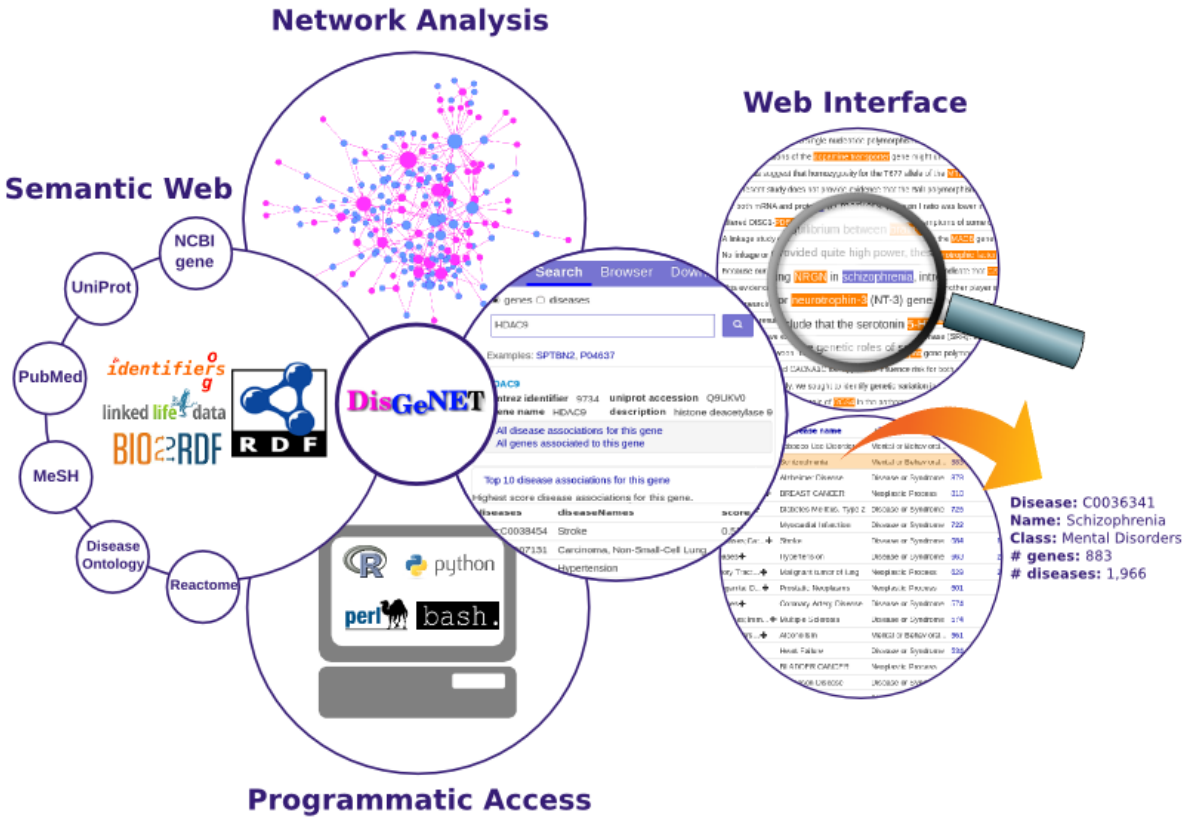
- Metadata description (W3C HCLS)
- Interlinking
  - Bio2RDF
  - linked life data
- Access
  - **Download** Data Dump
  - **SPARQL** Endpoint
  - **Faceted** Browser
  - Open PHACTS Discovery Platform
- **Open** license
- DataHub
- Software
  - D2RQ
  - Virtuoso Universal Server
  - OPENLINK SOFTWARE Making Technology Work For You®

Disease Annotation in DisGeNET

- X-ref to other disease terminologies:
  - MeSH
  - OMIM
  - DO (Human Disease Ontology)
  - Orphanet
  - NCI
  - **ICD9CM**
  - HPO (Human Phenotype Ontology)
- Phenotype annotation from

Interoperability

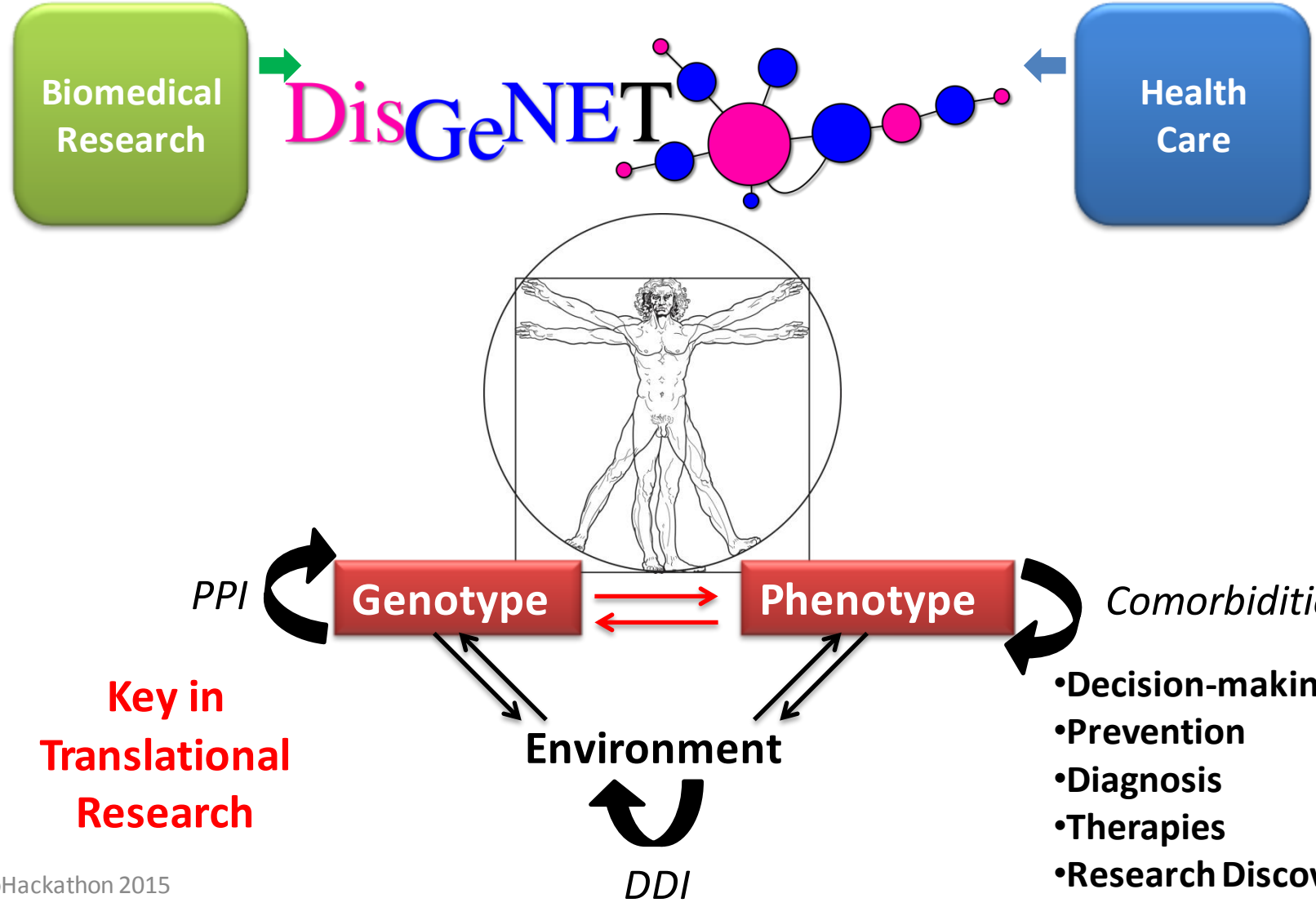
Tools for exploration



Usage stats (May2014-May 2015):

- 7 695 **users**, 15539 sessions (4:39 min/session)
- 16 130 **downloads** (database, Cytoscape plugin, RDF/Nanopubs)
- DisGeNET **used in 20+** publications, **cited in +60** articles

Understanding Human Diseases



Acknowledgments

IBI Group

Alba Gutiérrez-Sacristán

Àlex Bravo

Janet Piñero

Núria Queralt Rosinach

Alexia Giannoula

Miguel A. Mayer

Laura I. Furlong

Ferran Sanz

Special thanks

Christine Chichester

Michel Dumontier

Tobias Kuhn

Mark Thompson

Jesse Van Dam

Open PHACTS collaborators

and

DisGeNET users!!!

Especially

BioHackathon 2015 in Nagasaki

Organizers

Toshiaki Katayama
Shin Kawano
Shuichi Kawashima
Jin-Dong Kim
Yuji Kohara
Mari Minowa
Hiroyuki Mishima
Yuki Moriya
Toshihisa Takagi
Toshiaki Tokimatsu
Hongyan Wu
Atsuko Yamaguchi
Yasunori Yamamoto

NBDC National Bioscience Database Center

DBCLS Database Center for Life Science

Atomic Bomb Disease Institute Nagasaki University

Thanks for your attention!
Questions are welcome!