MEDICAMENTOS INNOVADORES
Plataforma Tecnológica Española

Análisis masivos de datos y reutilización de datos para la investigación biomédica: experiencias concretas

Ferran Sanz

La Plataforma Tecnológica Española Medicamentos Innovadores cuenta con apoyo del Ministerio de Economía, Industria y Competitividad

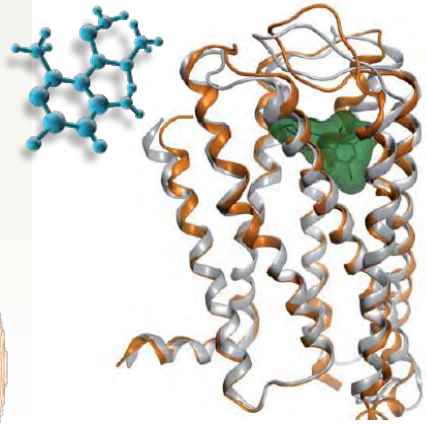
GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD

Biomedical Big Data

Social media

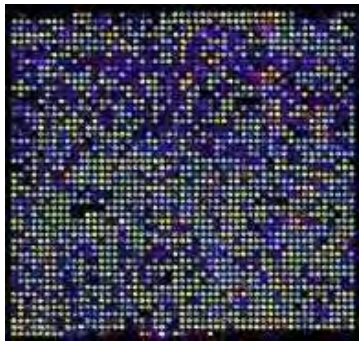
Biomedical literature

Clinical data (EHR)



Drugs & other chemicals

'omics information



Ejemplos de proyectos de *Big Data*

Ejemplos de proyectos IMI relacionados con el *Big Data*:

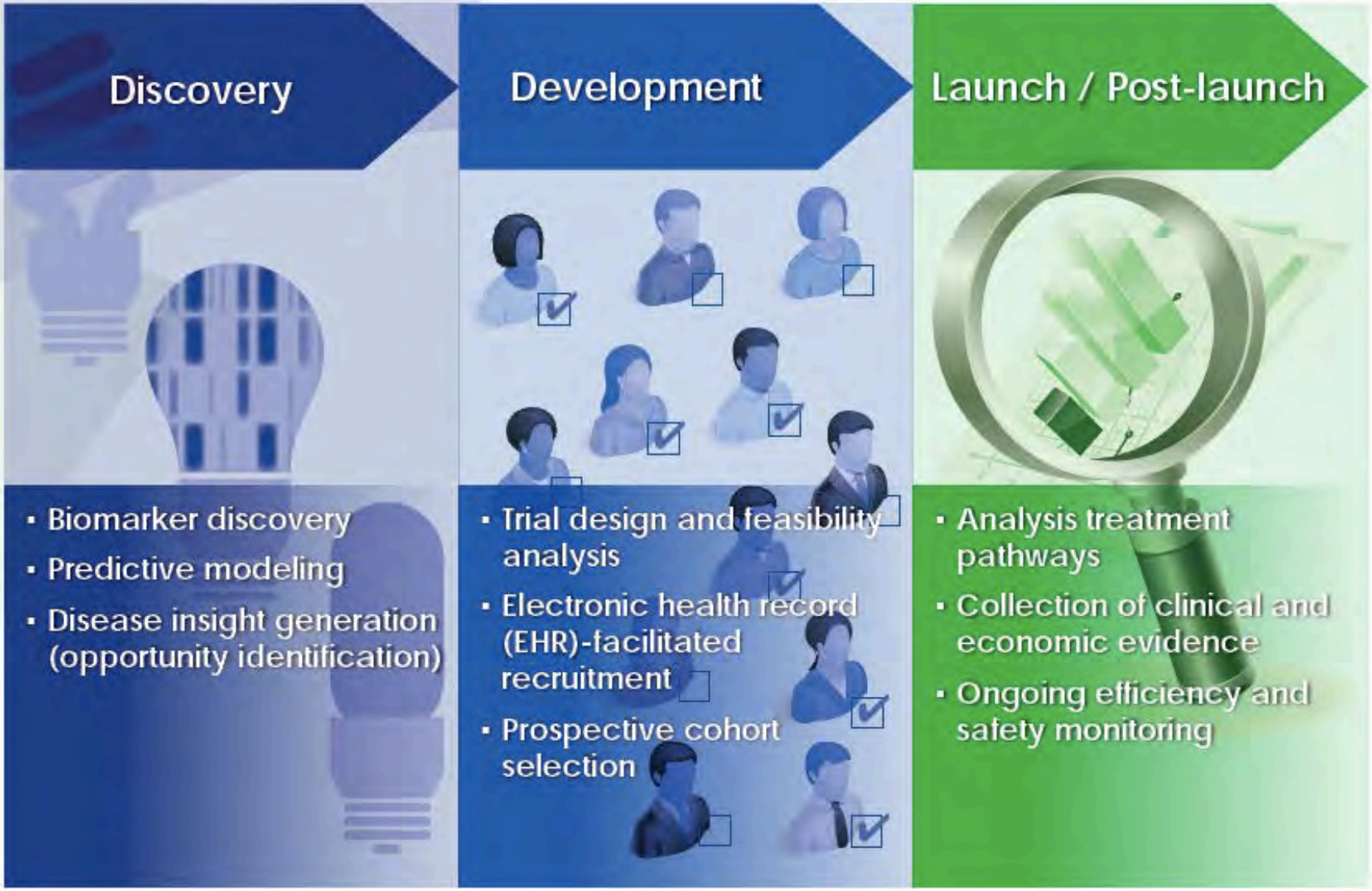
- **EMIF:** Plataforma para la (re)utilización de datos clínicos en la investigación biomédica.
- **Open PHACTS:** Descubrimiento de conocimiento en el I+D farmacéutico mediante la aplicación de la tecnología de la web semántica.
- **eTOX:** Predicción de la toxicidad de medicamentos mediante análisis integrado de casuística acumulada en la industria farmacéutica (continuidad y extensión a través de **eTRANSAFE**)

Otros proyectos europeos (H2020) sobre el tema:

- **EU-ADR:** Detección de efectos adversos mediante análisis masivo de historias clínicas y sustanciación bioinformática.
- **MedBioinformatics:** Aplicaciones bioinformáticas para la investigación traslacional y la práctica clínica.
- **ELIXIR-EXCELERATE:** Infraestructura Europea de datos en ciencias de la vida.

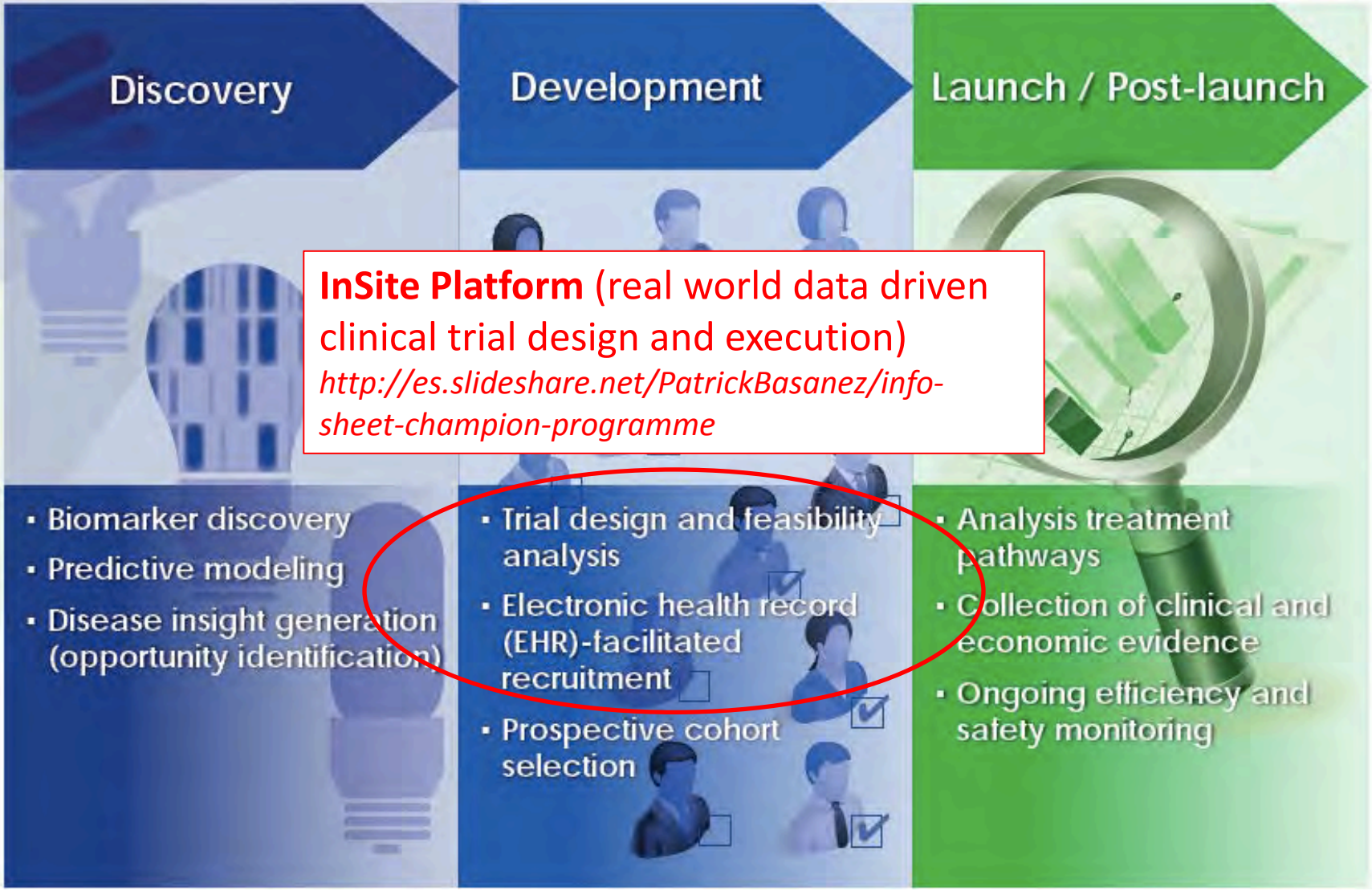
Why EMIF is Needed

Potential Applications for Real World Data in Drug Development



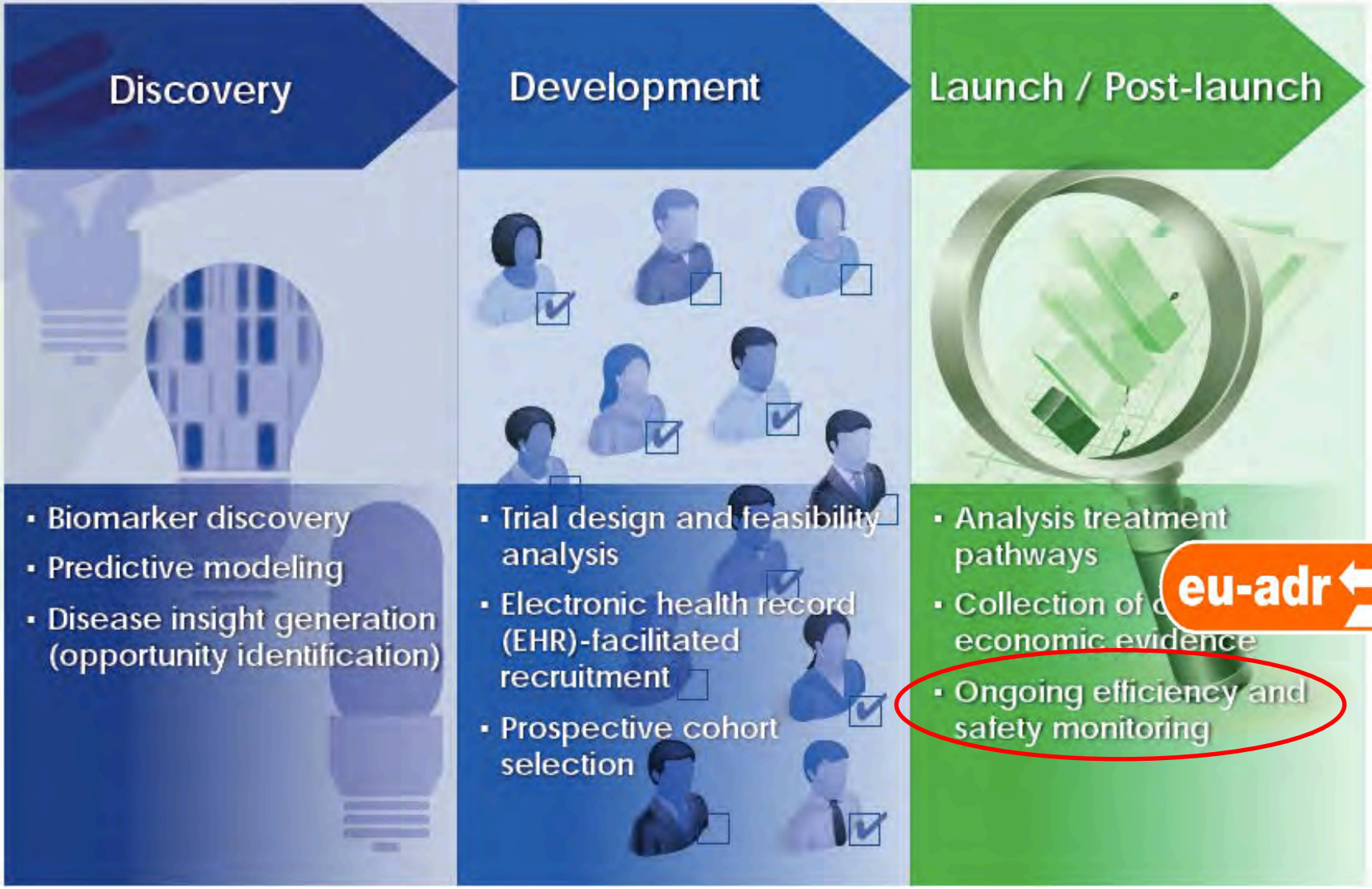
Why EMIF is Needed

Potential Applications for Real World Data in Drug Development

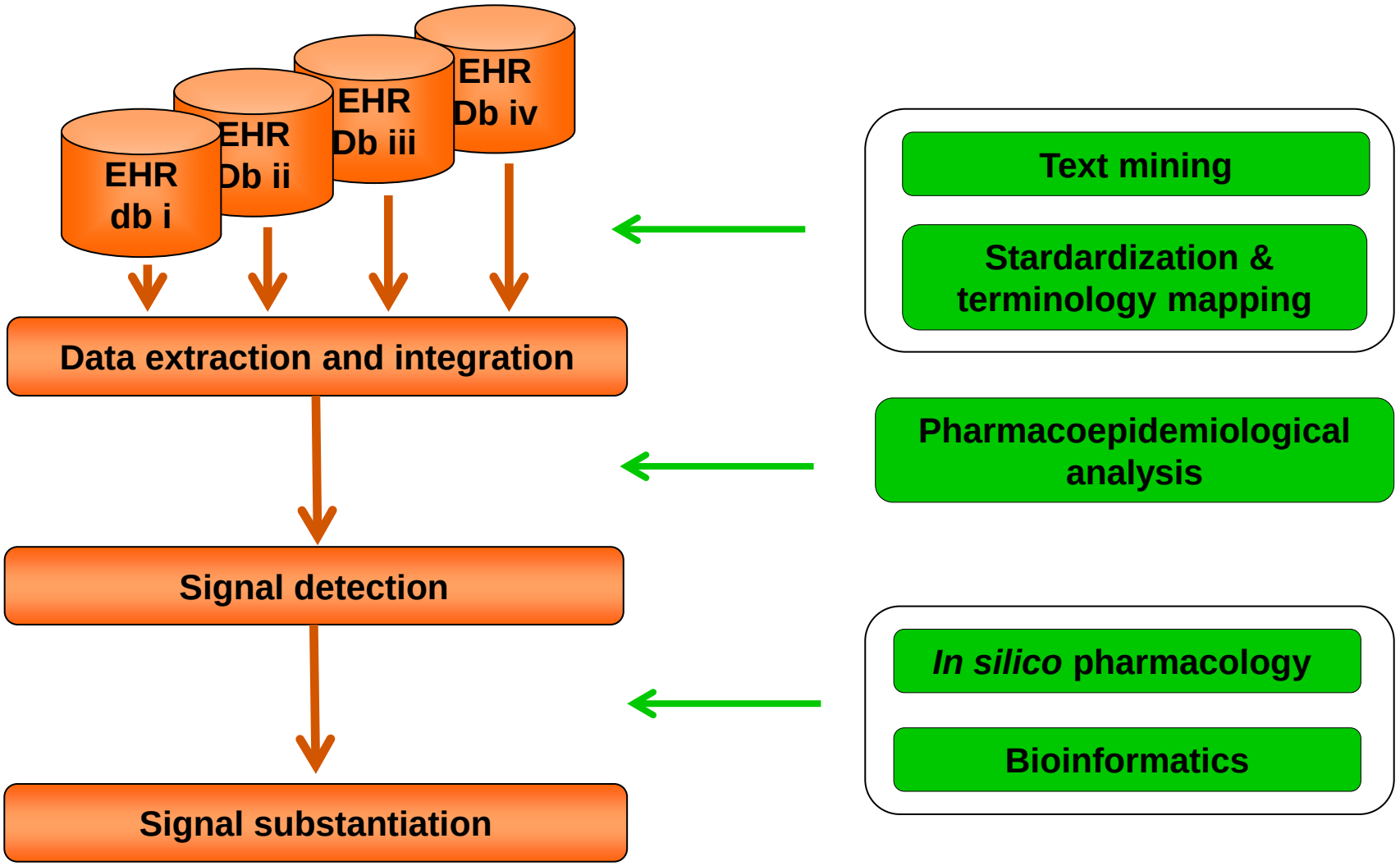


Why EMIF is Needed

Potential Applications for Real World Data in Drug Development



Exploitation of the Biomedical Big Data in pharmacovigilance



OPEN ACCESS Freely available online

PLOS ONE

Drug-Induced Acute Myocardial Infarction: Identifying 'Prime Suspects' from Electronic Healthcare Records-Based Surveillance System

Preciosa M. Coloma[1]*, Martijn J. Schuemie[1], Gianluca Trifirò[1,2], Laura Furlong[3], Erik van Mulligen[1], Anna Bauer-Mehren[3], Paul Avillach[4,5], Jan Kors[1], Ferran Sanz[3], Jordi Mestres[3], José Luis Oliveira[6], Scott Boyer[7], Ernst Ahlberg Helgee[7], Mariam Molokhia[8], Justin Matthews[9], David Prieto-Merino[9], Rosa Gini[10], Ron Herings[11], Giampiero Mazzaglia[12], Gino Picelli[13], Lorenza Scotti[14], Lars Pedersen[15], Johan van der Lei[1], Miriam Sturkenboom[1,16], on behalf of the EU-ADR consortium

Workflows for signal substantiation

OPEN ACCESS Freely available online

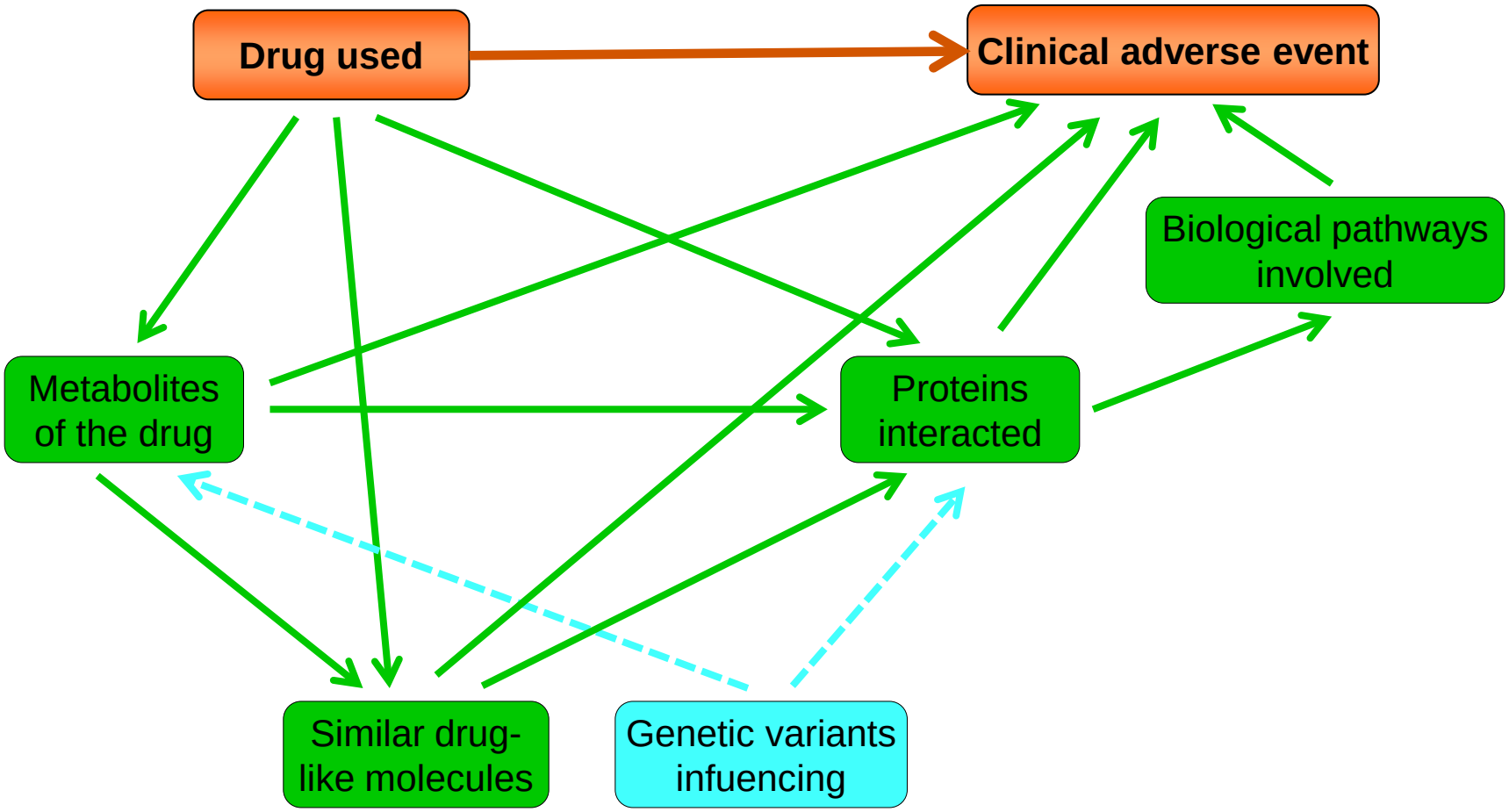
PLoS COMPUTATIONAL BIOLOGY

Automatic Filtering and Substantiation of Drug Safety Signals

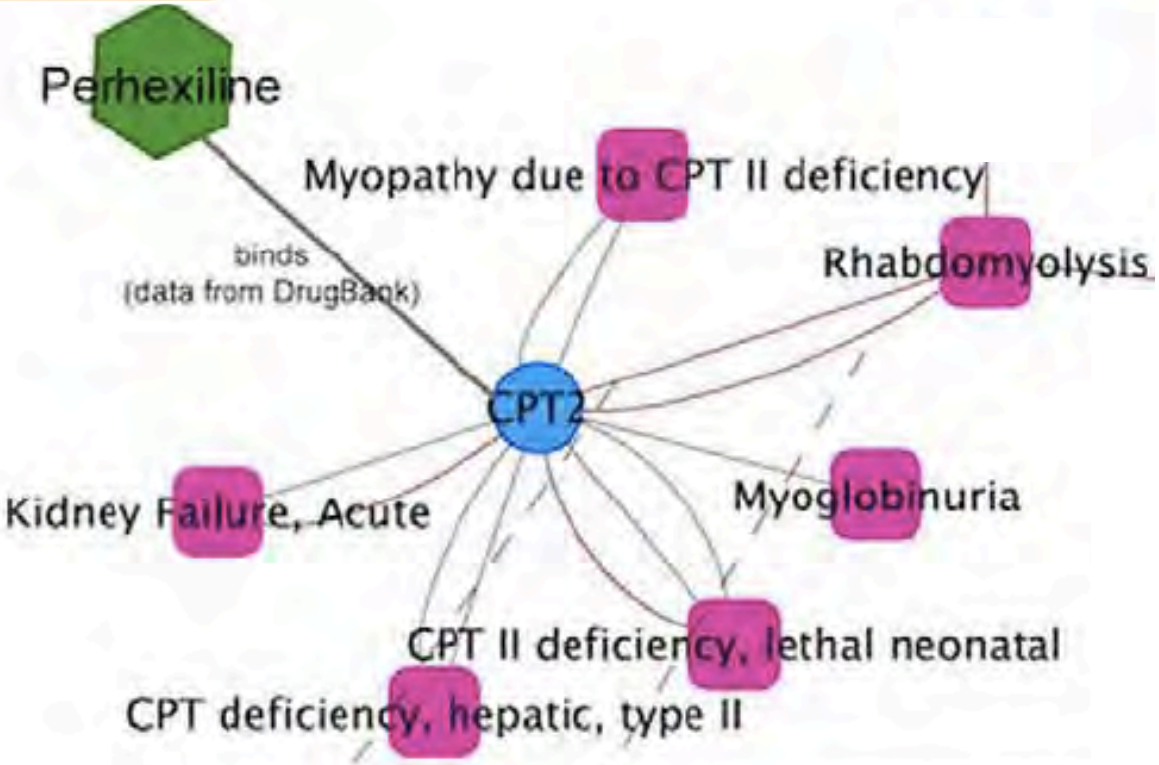
Anna Bauer-Mehren[1], Erik M. van Mullingen[2], Paul Avillach[3,4], María del Carmen Carrascosa[1], Ricard Garcia-Serna[1], Janet Piñero[1], Bharat Singh[2], Pedro Lopes[5], José L. Oliveira[5], Gayo Diallo[3], Ernst Ahlberg Helgee[6], Scott Boyer[6], Jordi Mestres[1], Ferran Sanz[1], Jan A. Kors[2], Laura I. Furlong[1]*

1 Research Programme on Biomedical Informatics (GRIB), IMIM-Hospital del Mar Research Institute, DCEX, Universitat Pompeu Fabra, Barcelona, Spain, 2 Erasmus University Medical Center, Rotterdam, The Netherlands, 3 LESIM-ISPED, Université de Bordeaux, Bordeaux, France, 4 LERTIM, EA 3283, Faculté de Médecine, Université de Aix-Marseille, Marseille, France, 5 DETI/IEETA, Universidade de Aveiro, Aveiro, Portugal, 6 AstraZeneca, Mölndal, Sweden

Workflows for signal substantiation



Knowledge discovery by information linkage



We report a Japanese adult form of CPT II deficiency associated with a homozygous F383Y mutation causing myalgia and rhabdomyolysis.

A novel variant of CPT II was found in a patient with rhabdomyolysis & acute renal failure: a deletion of cytosine & thymine at codon 408, resulting in a stop signal at 420, & an Arg631Cys mutation/frame shift at 408 has never been described before.

From Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F, Furlong LI. PLoS One 2011; 6(6): e20284

- Un recurso integral de información sobre las asociaciones entre genes y enfermedades humanos.

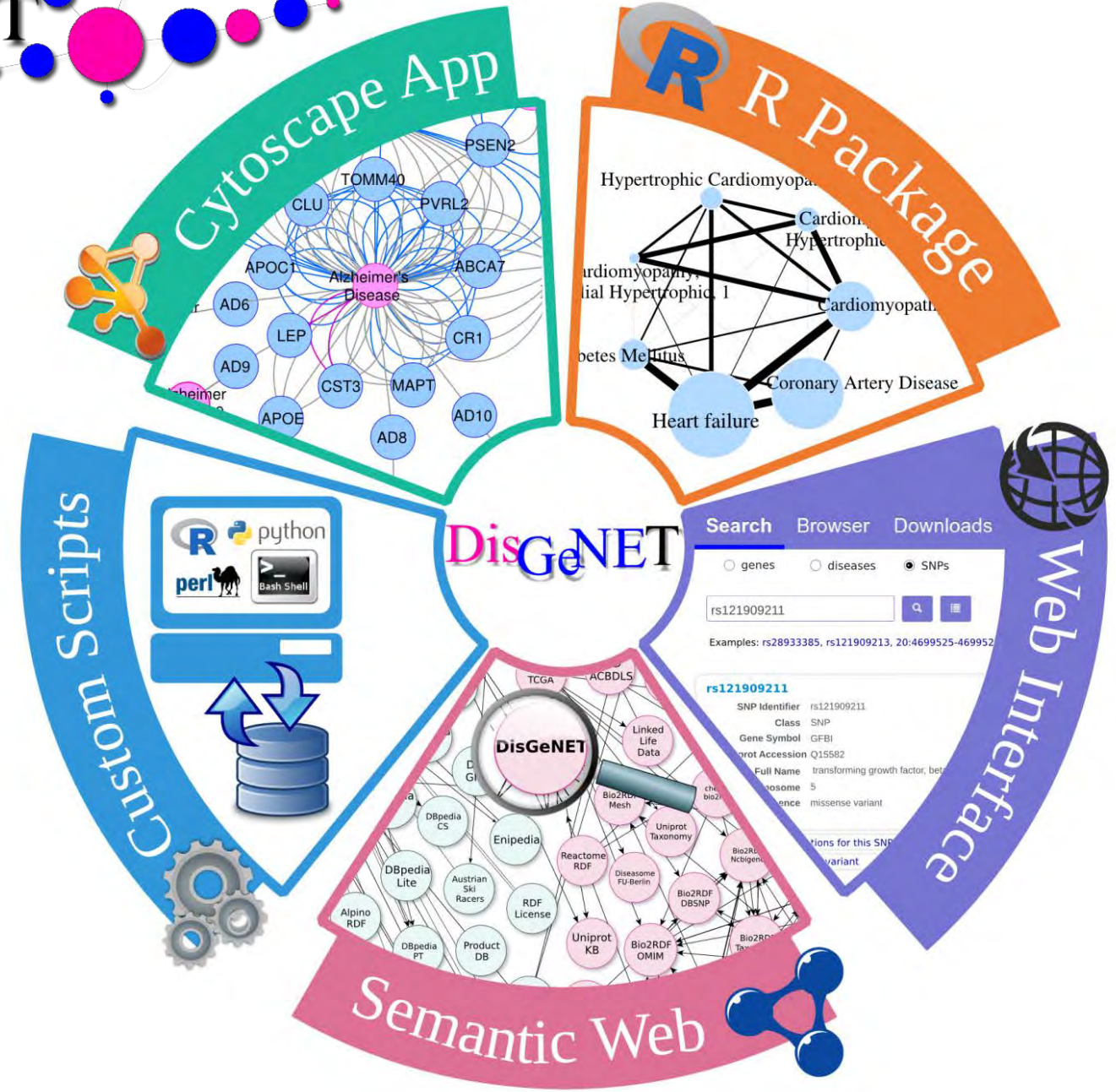
- Integra información de bases de datos disponibles públicamente y de la literatura científica mediante la minería de textos.

- DisGeNET v4.0 (Junio 2016) contiene 429.036 asociaciones que involucran 17.381 genes y 15,093 enfermedades y fenotipos

- Disponible libremente en: http://www.disgenet.org

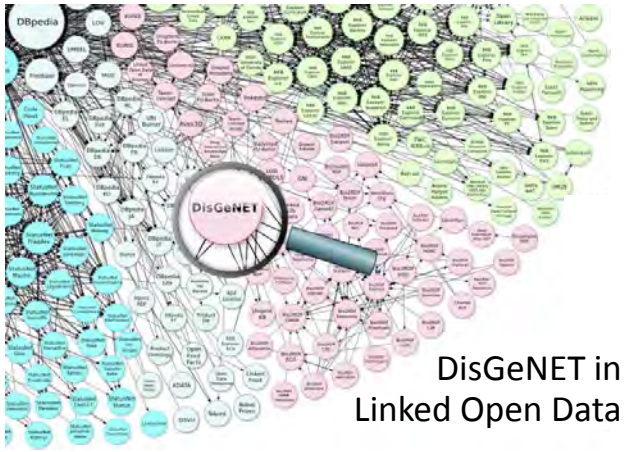
DisGeNET version 4.0: Top scoring genes for Major Depression

Gene	Number of diseases	DisGeNET score	DSI	DPI	Number of PMIDs	Number of SNPs
SLC6A4	374	0.236	0.411	0.852	157	5
TPH2	89	0.211	0.548	0.667	26	1
HTR2A	222	0.155	0.463	0.778	45	17
PCLO	20	0.130	0.696	0.333	12	5
CRHR1	118	0.127	0.531	0.778	11	11
CYP2D6	316	0.127	0.428	0.852	11	2
FKBP5	78	0.126	0.563	0.814	16	1
SP4	16	0.125	0.739	0.296	3	1
GRM7	32	0.123	0.666	0.444	5	1
GNAI3	7	0.122	0.812	0.296	2	1

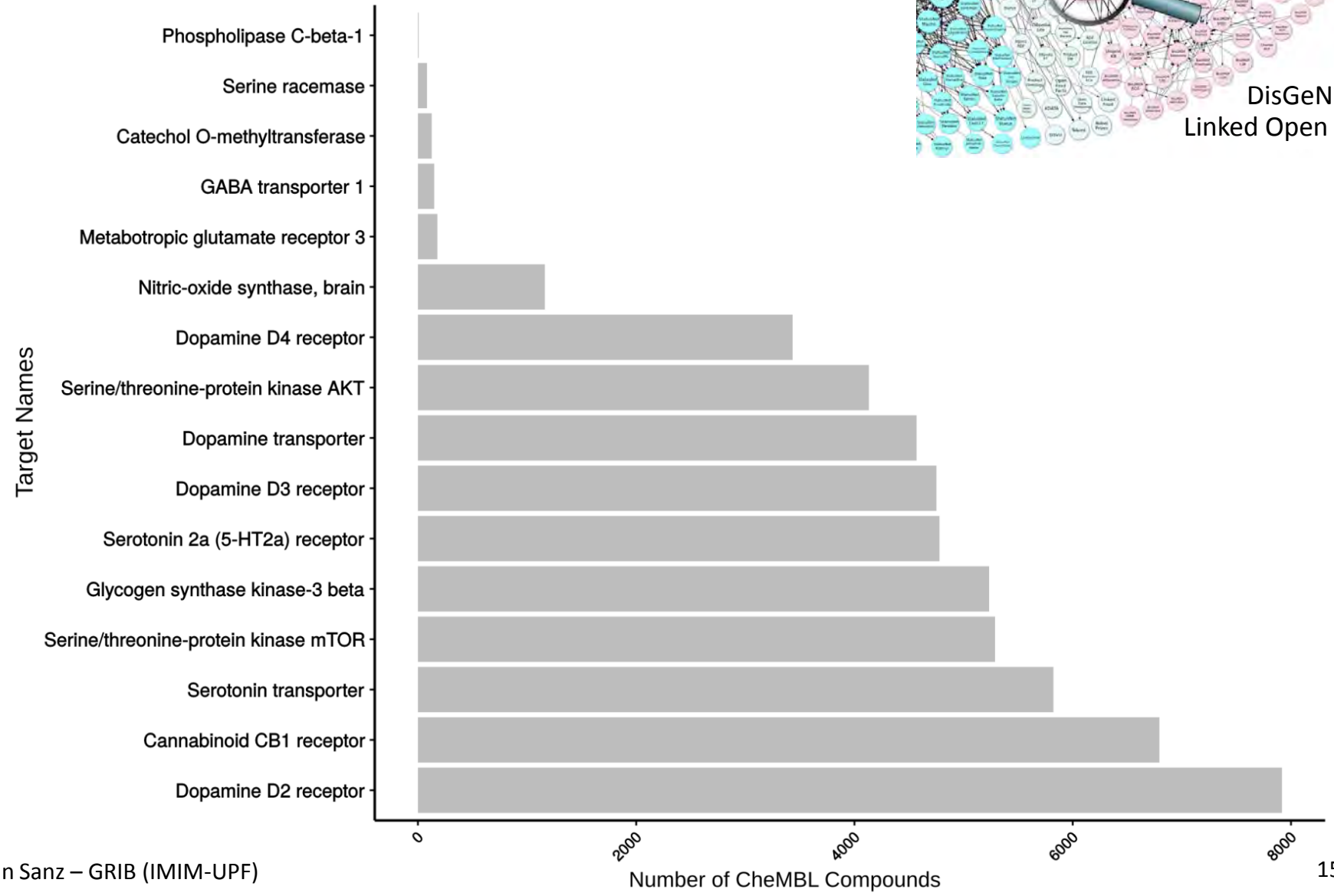

DisGeNET
Cytoscape App
PSEN2
TOMM40
CLU
PVRL2
APOC1
Alzheimer's Disease
ABCA7
AD6
LEP
CR1
AD9
CST3
MAPT
APOE
AD10
AD8
R Package
Hypertrophic Cardiomyopa
Cardiomyopath
Coronary Artery Disease
Heart failure
Custom Scripts
python
perl
Bash Shell
DisGeNET
Web Interface
Search
Browser
Downloads
genes
diseases
SNPs
rs121909211
Examples: rs28933385, rs121909213, 20:4699525-469952
rs121909211
SNP Identifier rs121909211
Class SNP
Gene Symbol GFBI
Full Name transforming growth factor, bet
missense variant
Semantic Web
TCGA
ACBDLS
DisGeNET
Linked Life Data
Bio2RDF Mesh
Uniprot Taxonomy
DBpedia CS
Enipedia
Reactome RDF
Diseasome FU-Berlin
DBpedia Lite
Austrian Ski Racers
RDF License
Bio2RDF DBSNP
Alpino RDF
DBpedia PT
Product DB
Uniprot KB
Bio2RDF OMIM

DisGeNET

Which schizophrenia genes are targeted by chemicals?



DisGeNET in
Linked Open Data



¡Muchas gracias por su atención!

http://grib.upf.edu