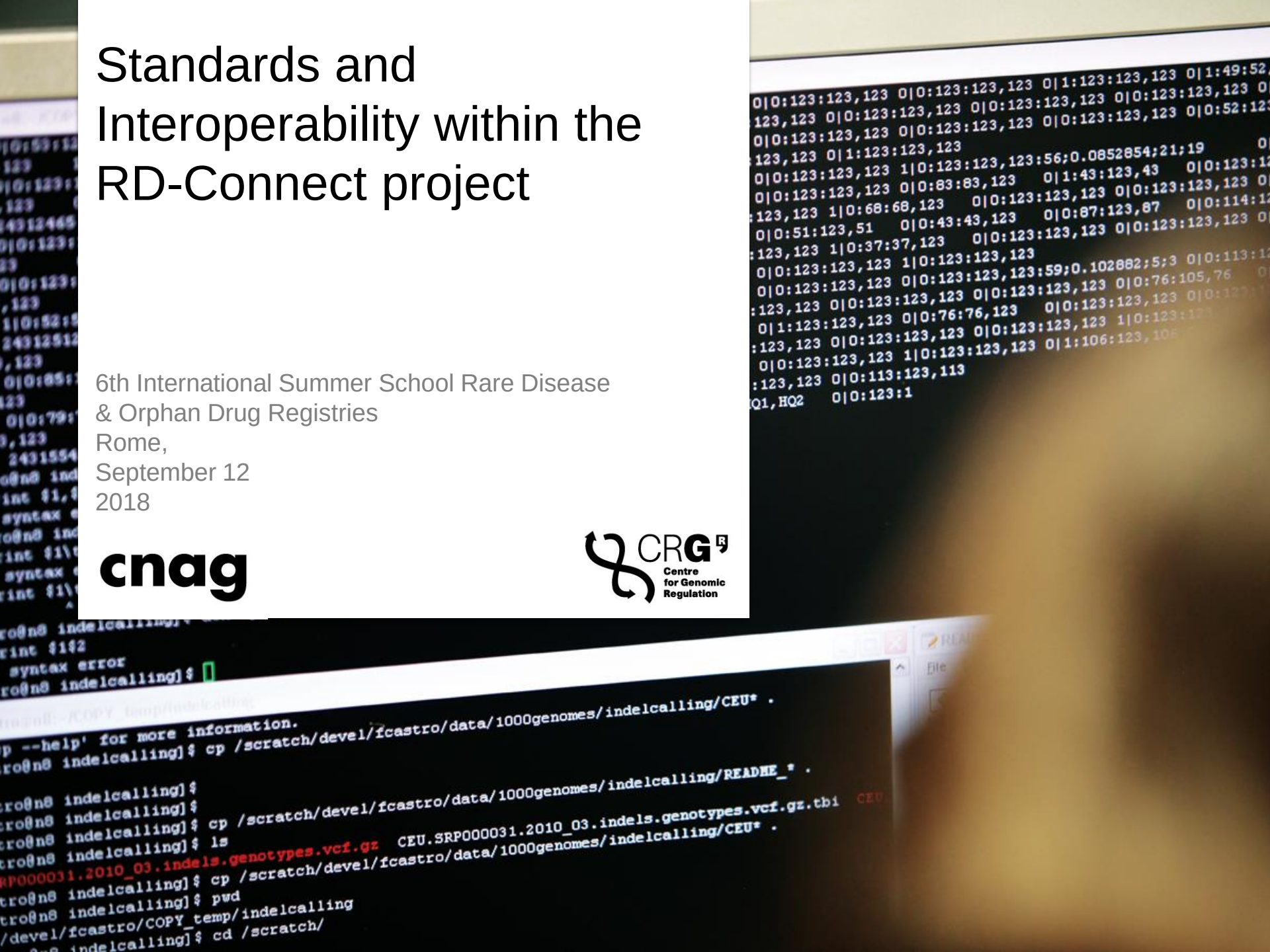
Standards and Interoperability within the RD-Connect project

6th International Summer School Rare Disease & Orphan Drug Registries
Rome,
September 12
2018

cnag

CRG Centre for Genomic Regulation



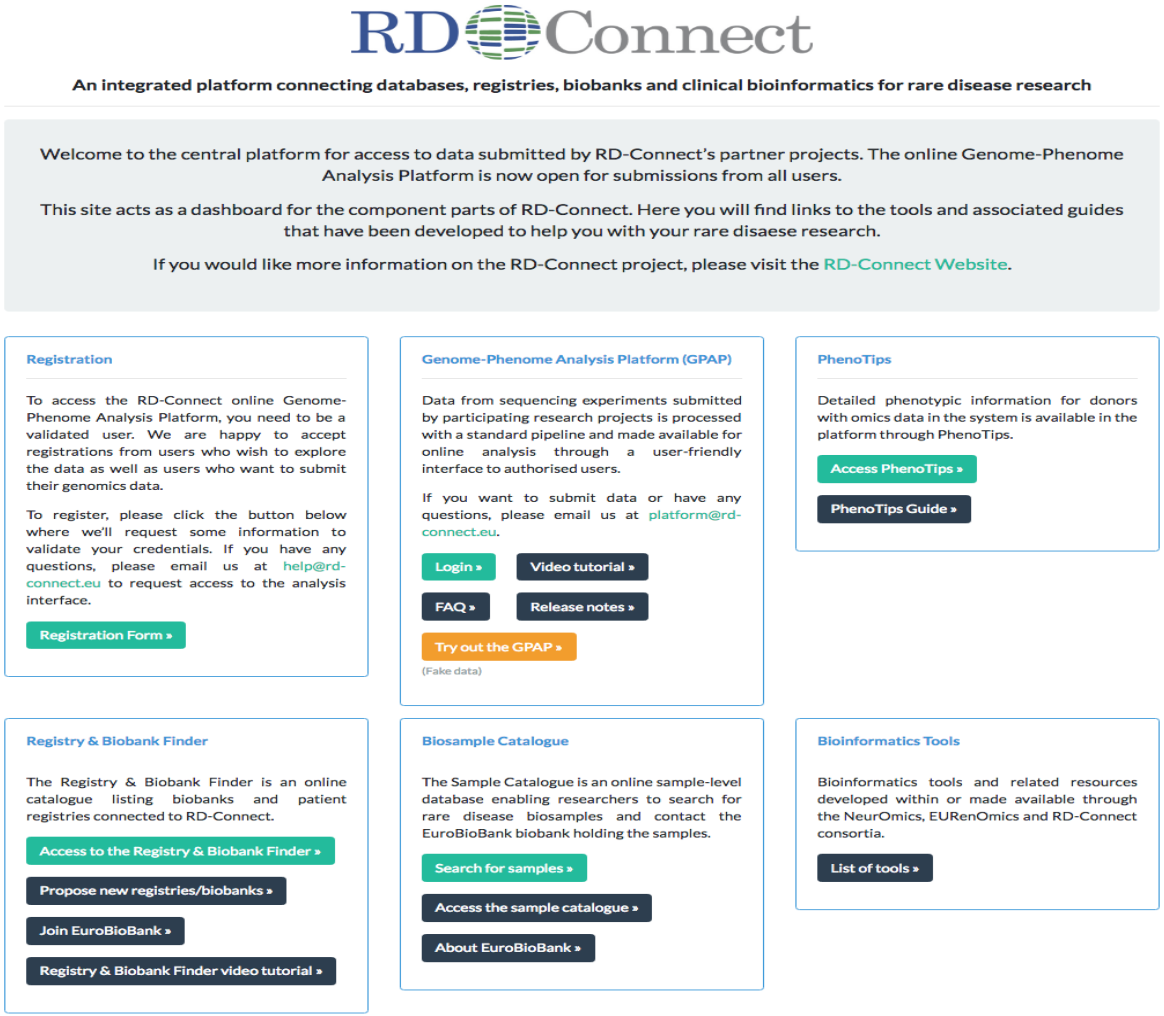
RD-Connect Infrastructure for Rare Disease Research

6-year project funded by EU 7th Framework Programme

An integrated platform connecting **genomic data** with **clinical information**, **registries**, and **biobanks** for **rare disease research**

- Overall objective: To contribute to **IRDiRC objectives** of delivering **200 new therapies** for rare diseases, and **means to diagnose most rare diseases** by 2020
- To create a central system for the **reprocessing**, **storing**, **analysis** and **sharing** of -*omics* data
  - Including integration of **phenotypic** and **biosample** data
  - Development of **new bioinformatic tools** to aid detailed analysis
- Ethical and legal considerations for sharing of sensitive data

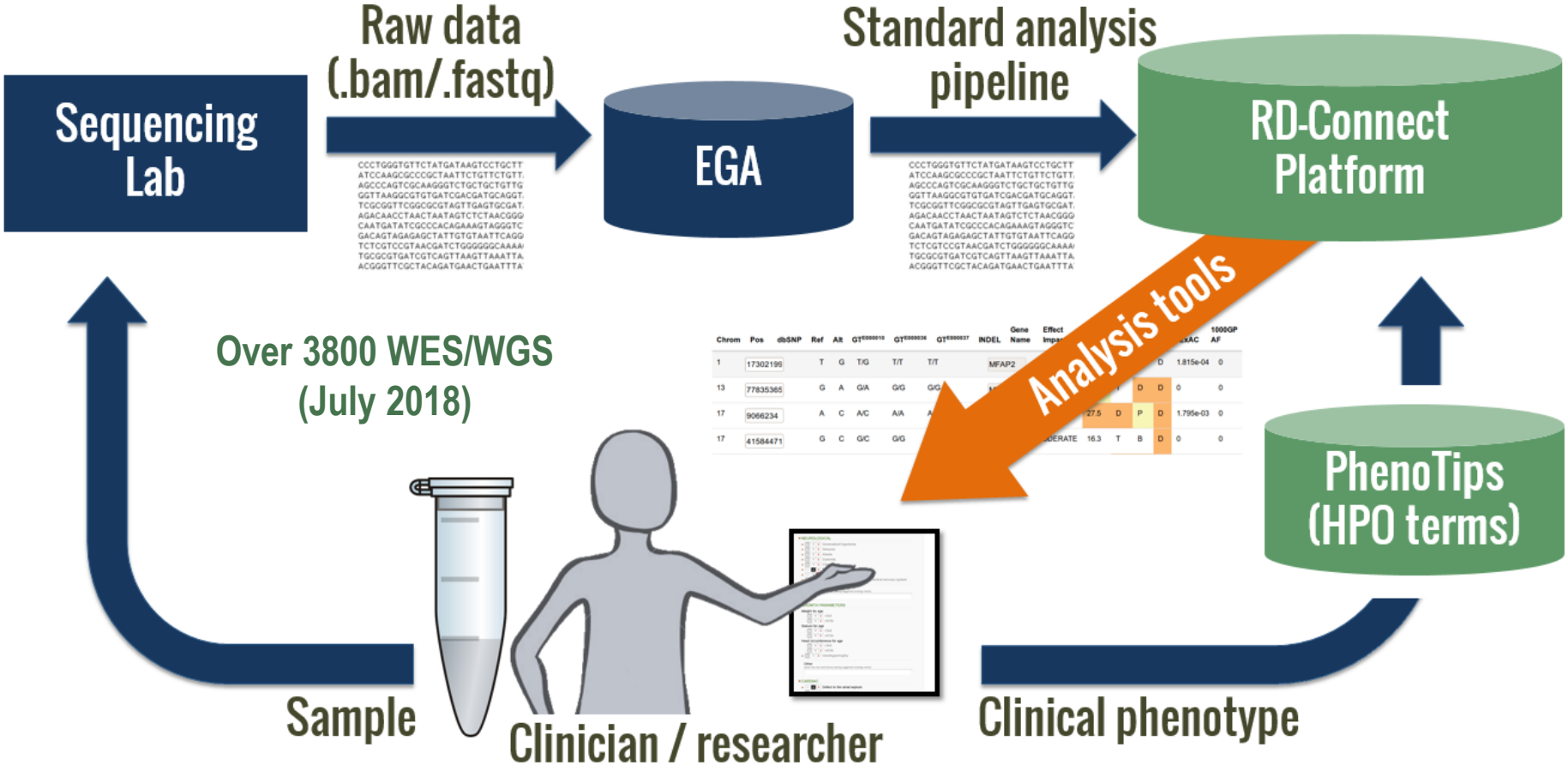
RD-Connect Platform (platform.rd-connect.eu)



RD Connect

RD-Connect Genome-Phenome Analysis Platform (GPAP)

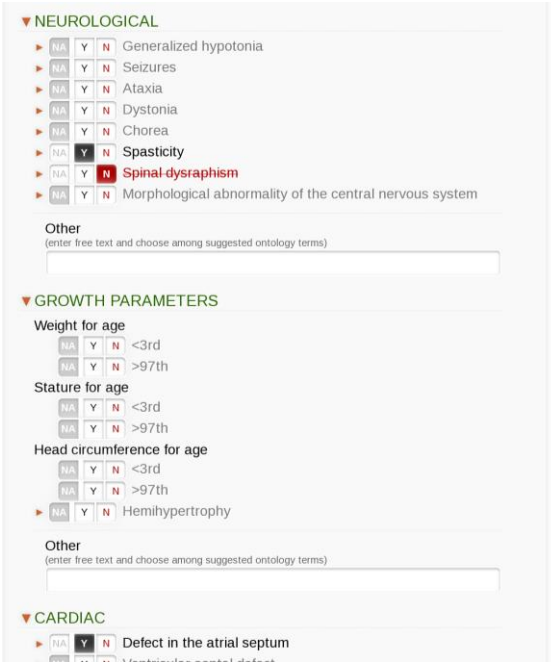
IRDiRC INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM
Recognised Resource

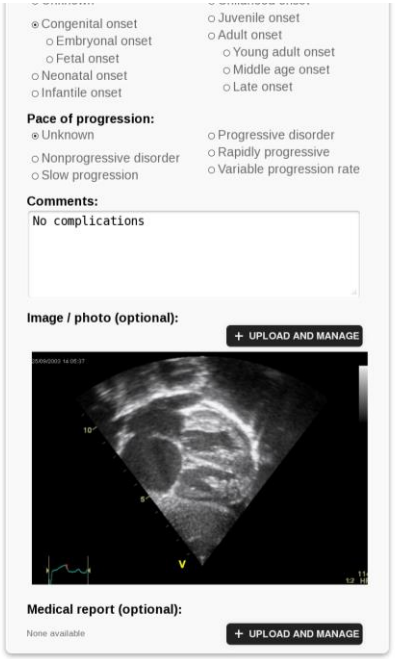


Ethics Committee Approval, GDPR Compliance, Code of Conduct, Data Access Committee, Security audited, User Activity logged

RD Connect

RD-Connect uses PhenoTips for collecting phenotypic data





Deep phenotyping in PhenoTips (Brudno *et al.*) achieved using the Human Phenotype Ontology (HPO – Robinson, Köhler *et al.*)

Diseases classified using the Orphanet Rare Disease Ontology and OMIM identifiers

Information from PhenoTips can be sent directly to other tools within the platform (e.g. Exomiser, MME)

RD Connect

Data standards and ontologies for FAIR compliance and interoperability

Phenotypic data

- User interface
- Phenopackets
- API

Genomic data

- FastQ/BAM/VCF
- User interface
- API

Clinical symptoms:
Human Phenotype Ontology*

Clinical diagnosis:
Orphanet Rare Disease Ontology (ORDO)*

Molecular diagnosis:
Online Mendelian Inheritance in Man*

Gene nomenclature:
HUGO gene nomenclature (HGNC)

Variant nomenclature:
Human genome variation society (HGVS)

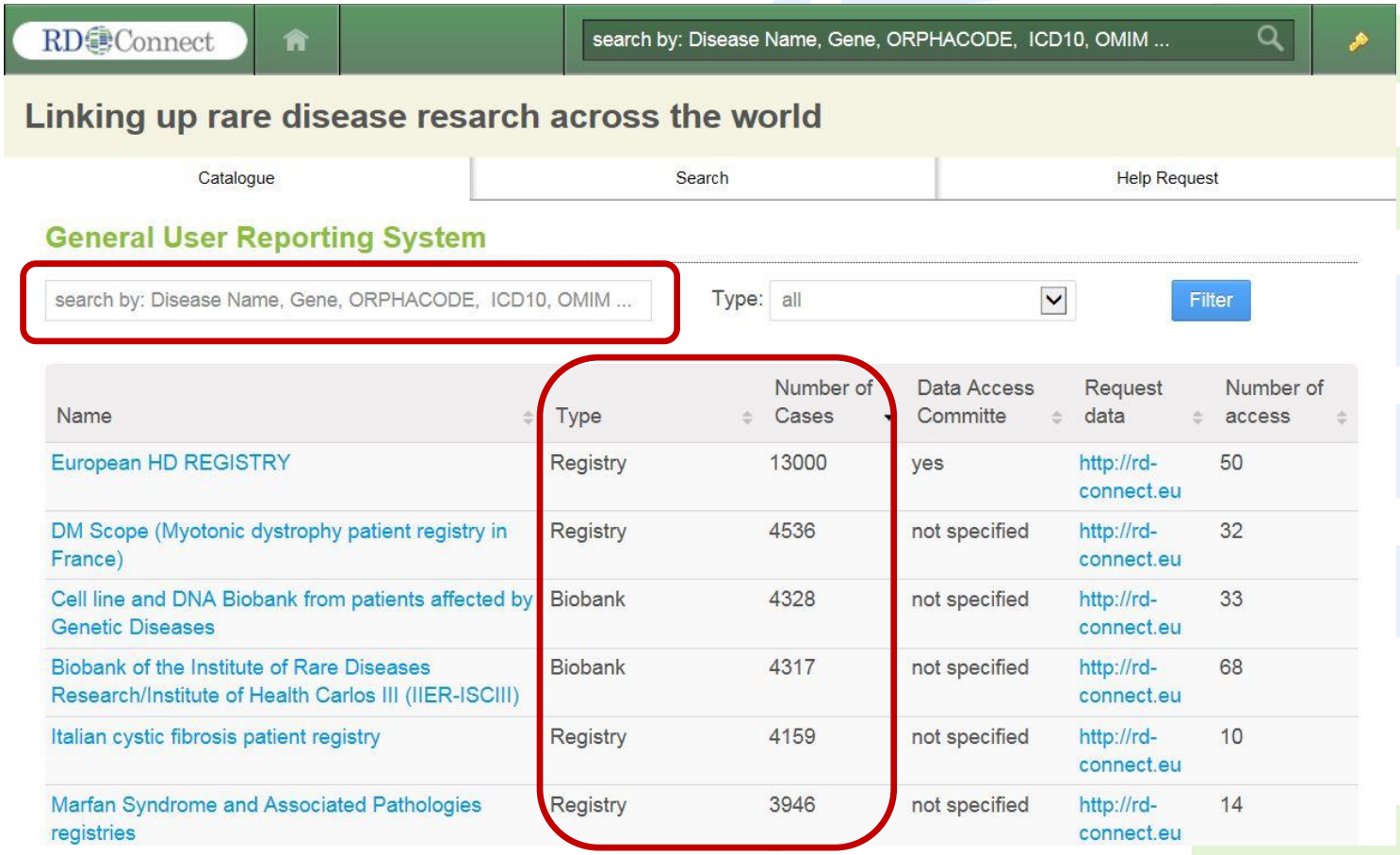
Variant annotation:
Ensembl (VEP)

Standard analysis pipeline: Laurie et al., 2016

IRDiRC INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM
IRDiRC INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM
SEVENTH FRAMEWORK PROGRAMME

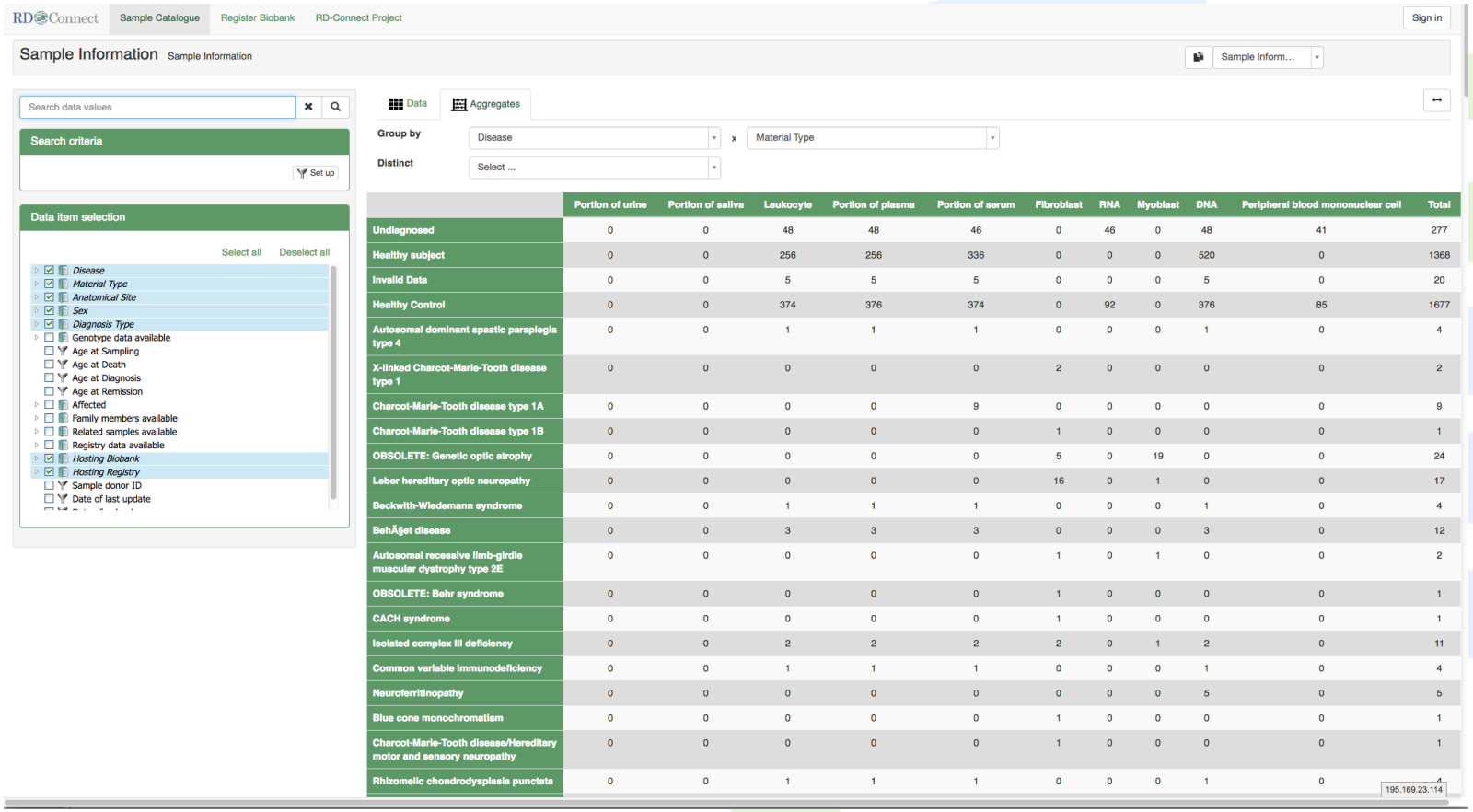
RD-Connect Register Biobank Finder

http://catalogue.rd-connect.eu/



RD-Connect Sample Catalogue

https://samples.rd-connect.eu/



	Portion of urine	Portion of saliva	Leukocyte	Portion of plasma	Portion of serum	Fibroblast	RNA	Myoblast	DNA	Peripheral blood mononuclear cell	Total
Undiagnosed	0	0	48	48	46	0	46	0	48	41	277
Healthy subject	0	0	256	256	336	0	0	0	520	0	1368
Invalid Data	0	0	5	5	5	0	0	0	5	0	20
Healthy Control	0	0	374	376	374	0	92	0	376	85	1677
Autosomal dominant spastic paraplegia type 4	0	0	1	1	1	0	0	0	1	0	4
X-linked Charcot-Marie-Tooth disease type 1	0	0	0	0	0	2	0	0	0	0	2
Charcot-Marie-Tooth disease type 1A	0	0	0	0	9	0	0	0	0	0	9
Charcot-Marie-Tooth disease type 1B	0	0	0	0	0	1	0	0	0	0	1
OBSOLETE: Genetic optic atrophy	0	0	0	0	0	5	0	19	0	0	24
Leber hereditary optic neuropathy	0	0	0	0	0	16	0	1	0	0	17
Beckwith-Wiedemann syndrome	0	0	1	1	1	0	0	0	1	0	4
BehÃ§et disease	0	0	3	3	3	0	0	0	3	0	12
Autosomal recessive limb-girdle muscular dystrophy type 2E	0	0	0	0	0	1	0	1	0	0	2
OBSOLETE: Behr syndrome	0	0	0	0	0	1	0	0	0	0	1
CACH syndrome	0	0	0	0	0	1	0	0	0	0	1
Isolated complex III deficiency	0	0	2	2	2	2	0	1	2	0	11
Common variable immunodeficiency	0	0	1	1	1	0	0	0	1	0	4
Neuroferritinopathy	0	0	0	0	0	0	0	0	5	0	5
Blue cone monochromatism	0	0	0	0	0	1	0	0	0	0	1
Charcot-Marie-Tooth disease/Hereditary motor and sensory neuropathy	0	0	0	0	0	1	0	0	0	0	1
Rhizomelic chondrodysplasia punctata	0	0	1	1	1	0	0	0	1	0	4

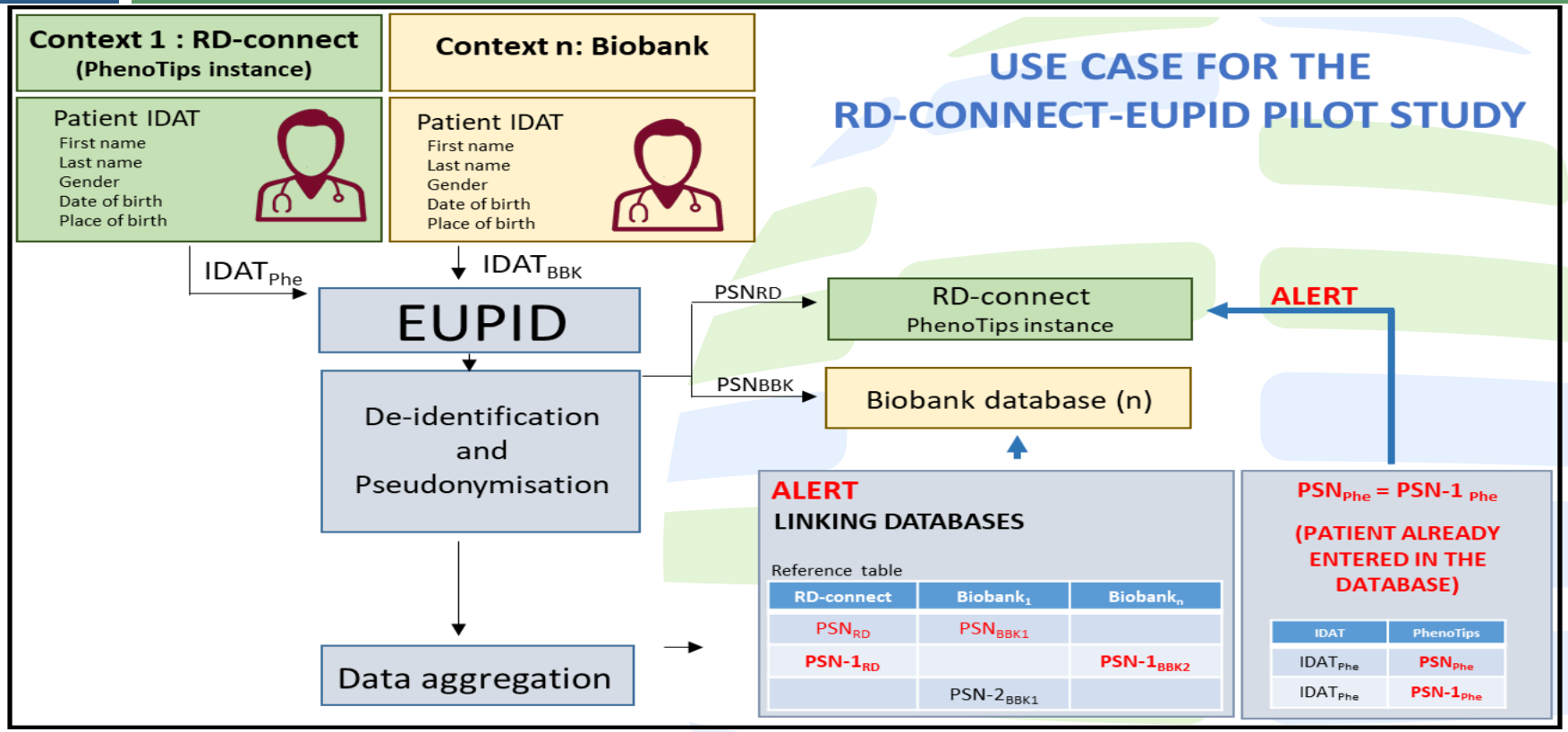
RD Connect

Data standards and ontologies for FAIR compliance and interoperability

The use of data standards and ontologies enables the RD-Connect platform to be interoperable and interconnected (API) to other platforms and resources:

Infrastructures /networks	PhenomeCentral Matchmaker Exchange GA4GH beacon
Ontologies	OMIM ORDO HPO
Tools	Human Splicing Finder MINT / Mentha EDGAR HmtDB

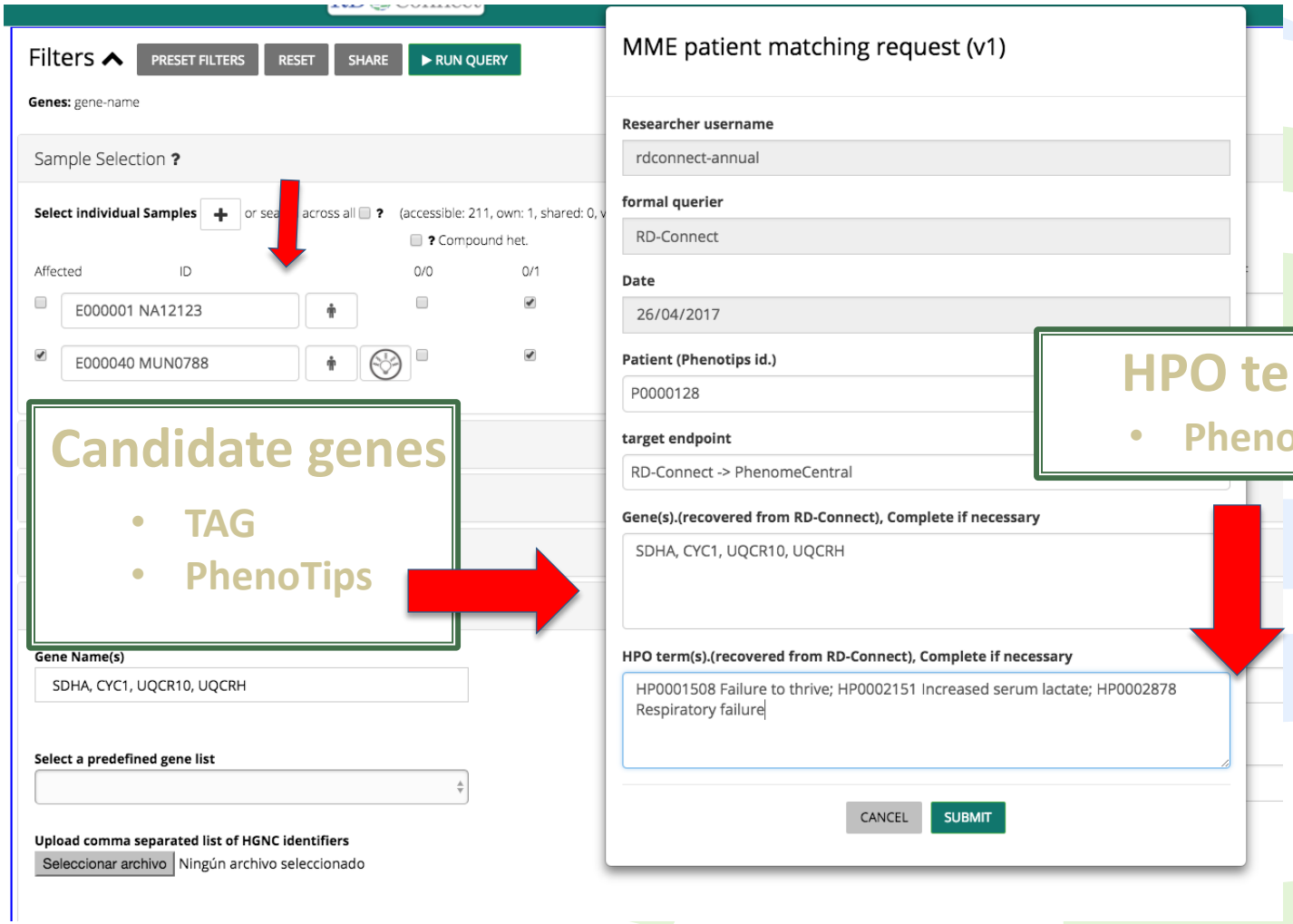
RD-Connect – EUPID pilot study



- Integrate pseudoanonymised data from different contexts
- Detect duplicate entries

RD Connect

Matchmaker exchange



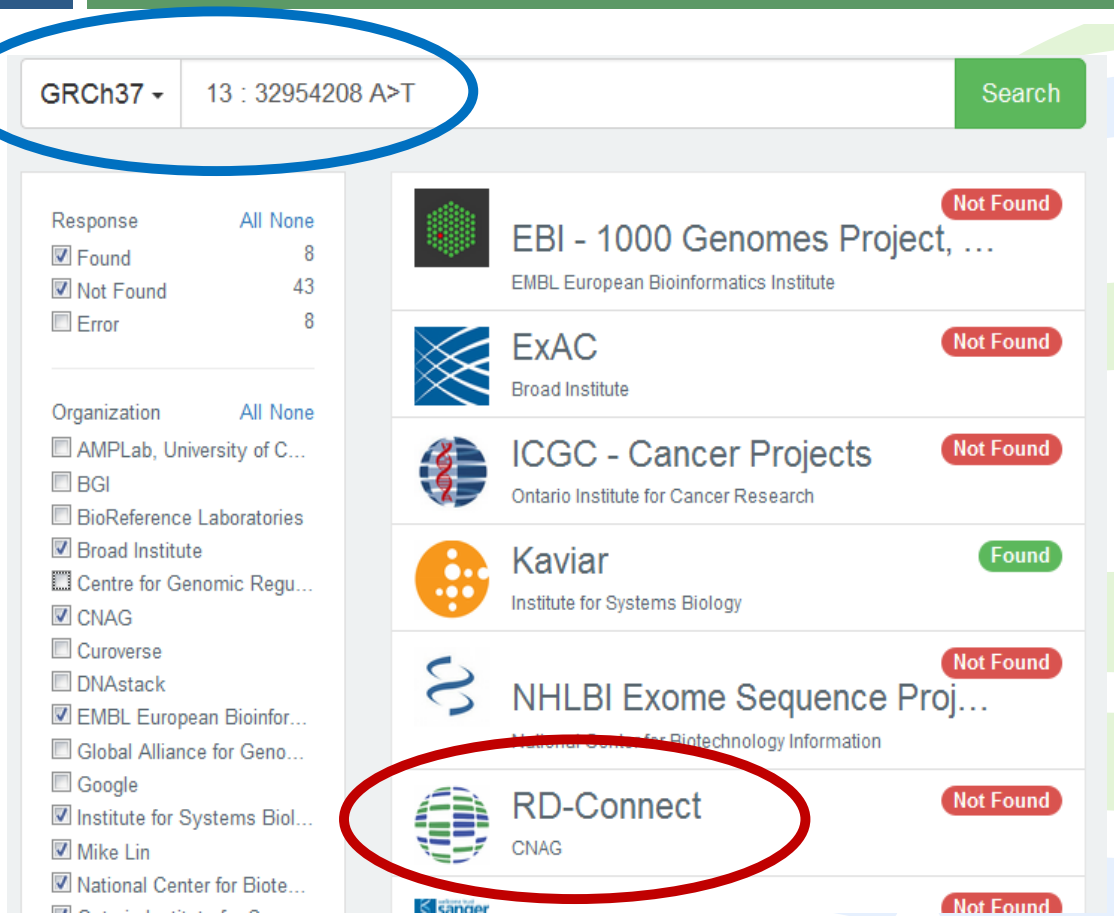
Matchmaker Exchange: initiate contact

score	patient	Contact	Phenotype	Genomics
0.544	P0003777	Initiate contact	more...	more...
0.543	P0003471	Initiate contact	more...	more...
0.530	P0003470	Initiate contact	more...	more...
0.518	P0001699	Initiate contact	more...	more...
0.518	P0001484	Initiate contact	more...	more...
0.512	P0003185	Initiate contact	more...	more...

An email will be sent to yourself and the submitter of the matched patient record in order to initiate the contact. The email will contain the matched gene and HPO terms and the patient ID from RD-Connect and the other platform. A copy of this email will also be sent to platform@rd-connect.eu to keep track of the matches initiated through MatchMaker Exchange.

Aceptar

Data Sharing: GA4GH Beacon

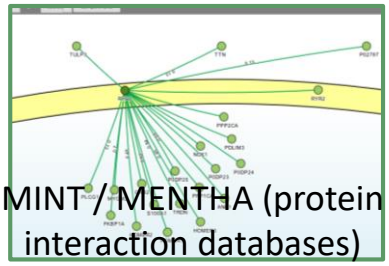


Question:
Does any sample in your database have **variant V**?

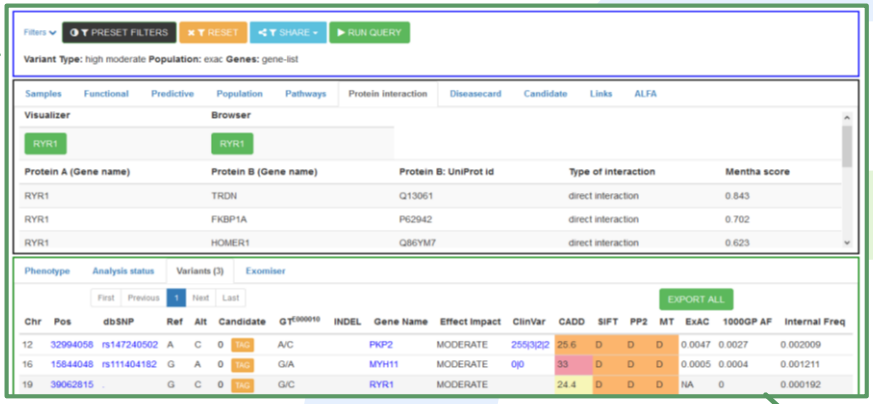
Answer:
Yes / No

https://beacon-network.org/#/

ELIXIR Implementation Study: Integration of ELIXIR-IIB in ELIXIR Rare Diseases activities

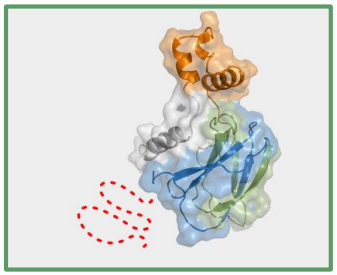


MINT / MENTHA (protein interaction databases)

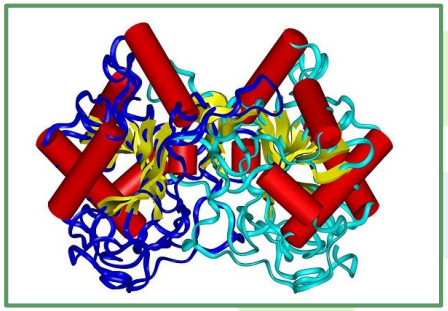


eDGAR

disease-gene Association database



VHLdb: A database of von Hippel-Lindau protein interactors and mutations



GALT-PROTEIN-DB 2.0

REDIportal

An ATLAS of A-to-I RNA editing events in human

HmtDB Human Mitochondrial DataBase

Human Mitochondrial DataBase

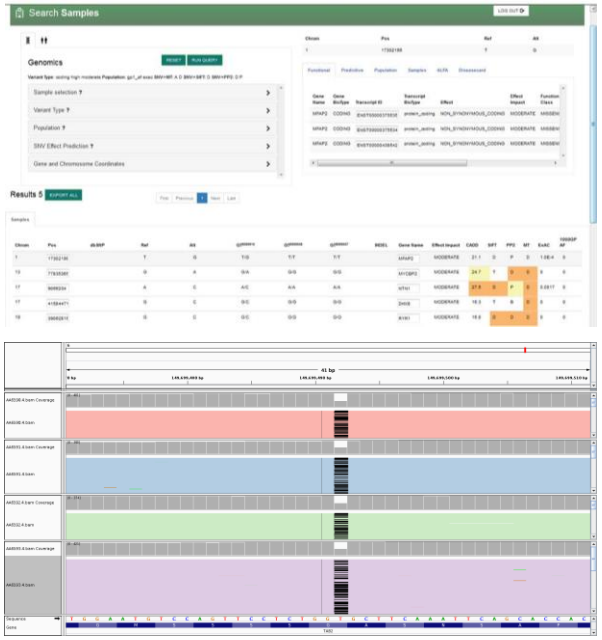
RD Connect

elixir

ELIXIR Implementation Study to visualize data deposited at the EGA in real-time

RD-Connect
BAM/CRAM

RD-Connect Genomics Platform



1 File submission to the EGA & ID-mapping

2 Request
- File ID/s
- Chr. coordinates

EGA
(BAM/CRAM)

3 Reply
- BAM/CRAM slice
- or
- Data representation

HTSget API

4 Rendering*

Secure and encrypted connections

*Broad's IGV screenshot

RD Connect

Adoptable by other projects and data types

elixir

SolveRD RD3 database

- Create a federated database with the information of all data available within SolveRD data repositories
  - RD-Connect GPAP
  - RD-Connect Sample Catalogue
  - EGA
  - …
- Map metadata between repositories and against metadata “standards”
  - Bioschemas
  - Biocaddie DATS v2.2
  - GA4GH

PhenoPackets

- PhenoPackets is an open standard for representing and sharing detailed descriptions of phenotypic abnormalities and characteristics of individual patients, organisms, diseases, and publications.
- Within RD-Connect we create PhenoPackets from PhenoTips records to share phenotypic information between SolveRD members
- Information contained in a PhenoPacket
  - Hashed ID of the participant
  - Gender
  - HPO terms
  - OMIM/ORDO diagnosis
  - Gene (Candidate/Solved/Rejected/..)
  - Inheritance

GA4GH Search API

- Evolution of MME version 1 case
- Specifications: https://github.com/ga4gh/mme-apis/blob/master/search-api.md
- RD-Connect GPAP is in the process of implementing
- Ask if another dataset has a participant with:
  - A variant in a gene of specific region of genome
  - The participant has a specific phenotype or set of phenotypes (HPOs)

OpenAPI (not only genomics standard)

- https://www.openapis.org/about
- Rewriting some of the GPAP APIs to adhere to OpenAPI specifications
- The OpenAPI Specification is a community-driven open specification within the OpenAPI Initiative, a Linux Foundation Collaborative Project.
- The OpenAPI Specification (OAS) defines a standard, programming language-agnostic, interface description for REST APIs, which allows both humans and computers to discover and understand the capabilities of a service without requiring access to source code, additional documentation, or inspection of network traffic. When properly defined via OpenAPI, a consumer can understand and interact with the remote service with a minimal amount of implementation logic.

RD-Connect and GDPR

- A new Code of Conduct (RD-Connect GPAP) has been explicitly prepared for GDPR . It clearly states how the data will be used and for which purposes. (Important aspects : not identifiable data, GDPR makes some exception for research data,..)
- Setting a Data Access Committee and Ethics Committee Approval
- A new user has to accept the Code of Conduct in order to get trough the registration process
- All the platform users have been contacted to agree with the new terms and the user is disabled until he/she accepts the new conditions
- We enhanced security measures for preventing data breach (encryption, security audit, penetration test,etc..)

Acknowledgements

RD-Connect

WP1: Coordination
Ivo Gut / Hanns Lochmüller
(CNAG-CRG)

WP2: Patient registries
Domenica Taruscio (ISS and EPIRARE)

WP3: Biobanks
Lucia Monaco
(Fondaz. Telethon & EuroBioBank)

WP4: Bioinformatics
Christophe Béroud
(INSERM Marseille)

WP5: Unified platform
Ivo Gut (CNAG Barcelona)

WP6 Ethical/legal/social
Mats Hansson (Uppsala)

WP7: Impact/Innovation
Kate Bushby
(Newcastle and EUCERD/ EJARD)

cnag CRG Centre for Genomic Regulation

S. Beltran
S. Laurie
D. Piscia
L. Matalonga
A. Papakon.
D. Picó
R. Tonda
J.R. Trotta
G. Parra
C. Luengo
J. Camps
I. Martinez
D. Ovelleiro
J. Protasio

M. Ingham
E. Casals
R. Alcántara

I. Gut
M. Gut
M. Bayès
B. Fusté
CNAG lab
CNAG admin

IRDiRC International Rare Diseases Research Consortium

bbmri-lpc
BBMRI – Large Prospective Cohorts

Generalitat de Catalunya Departament d'Economia i Coneixement

Generalitat de Catalunya Departament de Salut

PERIS 2016 2020

SolveRD

elixir

excelerate

INB

ISC Instituto de Salud Carlos III

Gobierno de España Ministerio de Economía y Competitividad

Seventh Framework Programme