

Standards and Interoperability within the RD-Connect project

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

cnag



```
ro@n8 indelcalling]$ cp /scratch/devel/fcastro/data/1000genomes/indelcalling/CEU* .
ro@n8 indelcalling]$ ls
CP000031.2010_03.indels.genotypes.vcf.gz CEU.SRP000031.2010_03.indels.genotypes.vcf.gz.tbi
ro@n8 indelcalling]$ pwd
/devel/fcastro/COPY_temp/indelcalling
ro@n8 indelcalling]$ cd /scratch/
```

```
O|O:123:123,123 O|O:123:123,123 O|1:123:123,123 O|1:49:52
123,123 O|O:123:123,123 O|O:123:123,123 O|O:123:123,123 O
O|O:123:123,123 O|O:123:123,123 O|O:123:123,123 O|O:52:12
123,123 O|1:123:123,123
O|O:123:123,123 1|O:123:123,123:56:0.0852854:21:19 O
O|O:123:123,123 O|O:83:83,123 O|1:43:123,43 O|O:123:1
123,123 1|O:68:68,123 O|O:123:123,123 O|O:123:123,123 O
O|O:51:123,51 O|O:43:43,123 O|O:87:123,87 O|O:114:1
123,123 1|O:37:37,123 O|O:123:123,123 O|O:123:123,123 O
O|O:123:123,123 1|O:123:123,123
O|O:123:123,123 O|O:123:123,123:59:0.102882:5:3 O|O:113:1
123,123 O|O:123:123,123 O|O:123:123,123 O|O:76:105,76 O
O|1:123:123,123 O|O:76:76,123 O|O:123:123,123 O|O:123:1
123,123 O|O:123:123,123 O|O:123:123,123 1|O:123:123
O|O:123:123,123 1|O:123:123,123 O|1:106:123,106
123,123 O|O:113:123,113
Q1,HQ2 O|O:123:1
```



RD-Connect Infrastructure for Rare Disease Research

2

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)

3



An integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research

Welcome to the central platform for access to data submitted by RD-Connect's partner projects. The online Genome-Phenome Analysis Platform is now open for submissions from all users.

This site acts as a dashboard for the component parts of RD-Connect. Here you will find links to the tools and associated guides that have been developed to help you with your rare disease research.

If you would like more information on the RD-Connect project, please visit the [RD-Connect Website](#).

Registration

To access the RD-Connect online Genome-Phenome Analysis Platform, you need to be a validated user. We are happy to accept registrations from users who wish to explore the data as well as users who want to submit their genomics data.

To register, please click the button below where we'll request some information to validate your credentials. If you have any questions, please email us at help@rd-connect.eu to request access to the analysis interface.

[Registration Form »](#)

Genome-Phenome Analysis Platform (GPAP)

Data from sequencing experiments submitted by participating research projects is processed by a standard pipeline and made available for online analysis through a user-friendly interface to authorised users.

If you want to submit data or have any questions, please email us at platform@rd-connect.eu.

[Login »](#)

[Video tutorial »](#)

[FAQ »](#)

[Release notes »](#)

[Try out the GPAP »](#)

(Fake data)

PhenoTips

Detailed phenotypic information for donors with omics data in the system is available in the platform through PhenoTips.

[Access PhenoTips »](#)

[PhenoTips Guide »](#)

Registry & Biobank Finder

The Registry & Biobank Finder is an online catalogue listing biobanks and patient registries connected to RD-Connect.

[Access to the Registry & Biobank Finder »](#)

[Propose new registries/biobanks »](#)

[Join EuroBioBank »](#)

[Registry & Biobank Finder video tutorial »](#)

Biosample Catalogue

The Sample Catalogue is an online sample-level database enabling researchers to search for rare disease biosamples and contact the EuroBioBank biobank holding the samples.

[Search for samples »](#)

[Access the sample catalogue »](#)

[About EuroBioBank »](#)

Bioinformatics Tools

Bioinformatics tools and related resources developed within or made available through the NeuroOmics, EUREnOmics and RD-Connect consortia.

[List of tools »](#)

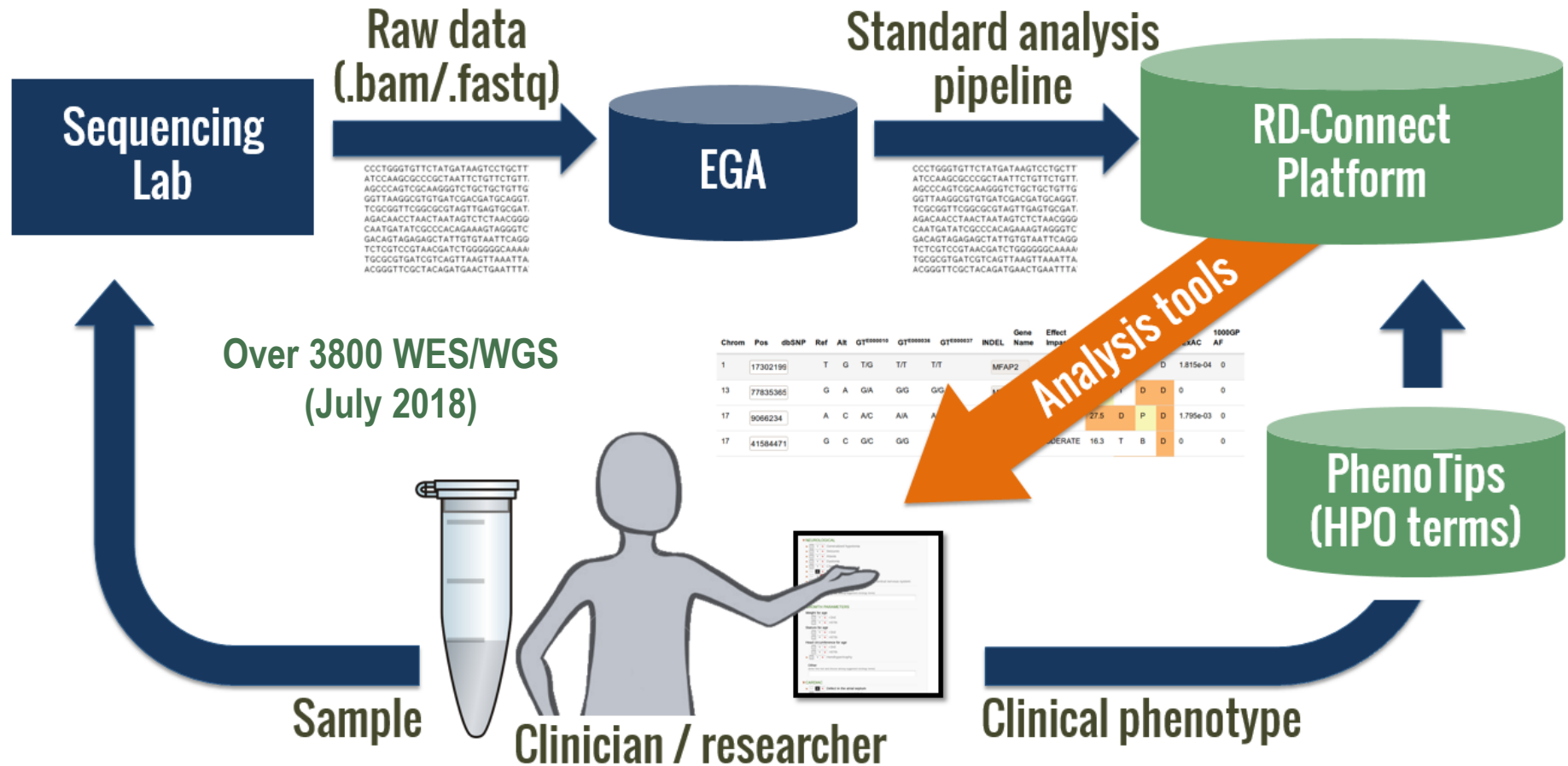


RD-Connect Genome-Phenome Analysis Platform (GPAP)



Recognised
Resource

4



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



RD-Connect uses PhenoTips for collecting phenotypic data

5

▼ NEUROLOGICAL

- Generalized hypotonia
- Seizures
- Ataxia
- Dystonia
- Chorea
- Spasticity
- Spinal dysraphism**
- Morphological abnormality of the central nervous system

Other
(enter free text and choose among suggested ontology terms)

▼ GROWTH PARAMETERS

Weight for age

Stature for age

Head circumference for age

Hemihypertrophy

Other
(enter free text and choose among suggested ontology terms)

▼ CARDIAC

- Defect in the atrial septum

Onset

- Congenital onset
- Embryonal onset
- Fetal onset
- Neonatal onset
- Infantile onset
- Juvenile onset
- Adult onset
- Young adult onset
- Middle age onset
- Late onset

Pace of progression:

- Unknown
- Nonprogressive disorder
- Slow progression
- Progressive disorder
- Rapidly progressive
- Variable progression rate

Comments:

No complications

Image / photo (optional):

+ UPLOAD AND MANAGE

Medical report (optional):

None available

+ UPLOAD AND MANAGE

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)

Phenotype

- Basics
- Diagnosis
- Clinical symptoms

Data standards and ontologies for FAIR compliance and interoperability

Phenotypic data

- User interface
- Phenopackets
- API



Clinical symptoms:

Human Phenotype Ontology*

Clinical diagnosis:

Orphanet Rare Disease Ontology (ORDO)*

Molecular diagnosis:

Online Mendelian Inheritance in Man*

Genomic data

- FastQ/BAM/VCF
- User interface
- API



Gene nomenclature:

HUGO gene nomenclature (HGNC)

Variant nomenclature:

Human genome variation society (HGVS)



Variant annotation:

Ensembl (VEP)


Standard analysis pipeline: Laurie et al., 2016



RD-Connect Register Biobank Finder

7


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

RD Connect   

Linking up rare disease research across the world

Catalogue Search Help Request

General User Reporting System

Type: 

Name	Type	Number of Cases	Data Access Committe	Request data	Number of access
European HD REGISTRY	Registry	13000	yes	http://rd-connect.eu	50
DM Scope (Myotonic dystrophy patient registry in France)	Registry	4536	not specified	http://rd-connect.eu	32
Cell line and DNA Biobank from patients affected by Genetic Diseases	Biobank	4328	not specified	http://rd-connect.eu	33
Biobank of the Institute of Rare Diseases Research/Institute of Health Carlos III (IIER-ISCI)	Biobank	4317	not specified	http://rd-connect.eu	68
Italian cystic fibrosis patient registry	Registry	4159	not specified	http://rd-connect.eu	10
Marfan Syndrome and Associated Pathologies registries	Registry	3946	not specified	http://rd-connect.eu	14



RD-Connect Sample Catalogue

8

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

RD-Connect Sample Catalogue Register Biobank RD-Connect Project Sign in

Sample Information Sample Information

Search data values x Q

Search criteria Set up

Data Aggregates

Group by Disease x Material Type

Distinct Select ...

	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
Behçet 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

195.169.23.114

Data standards and ontologies for FAIR compliance and interoperability

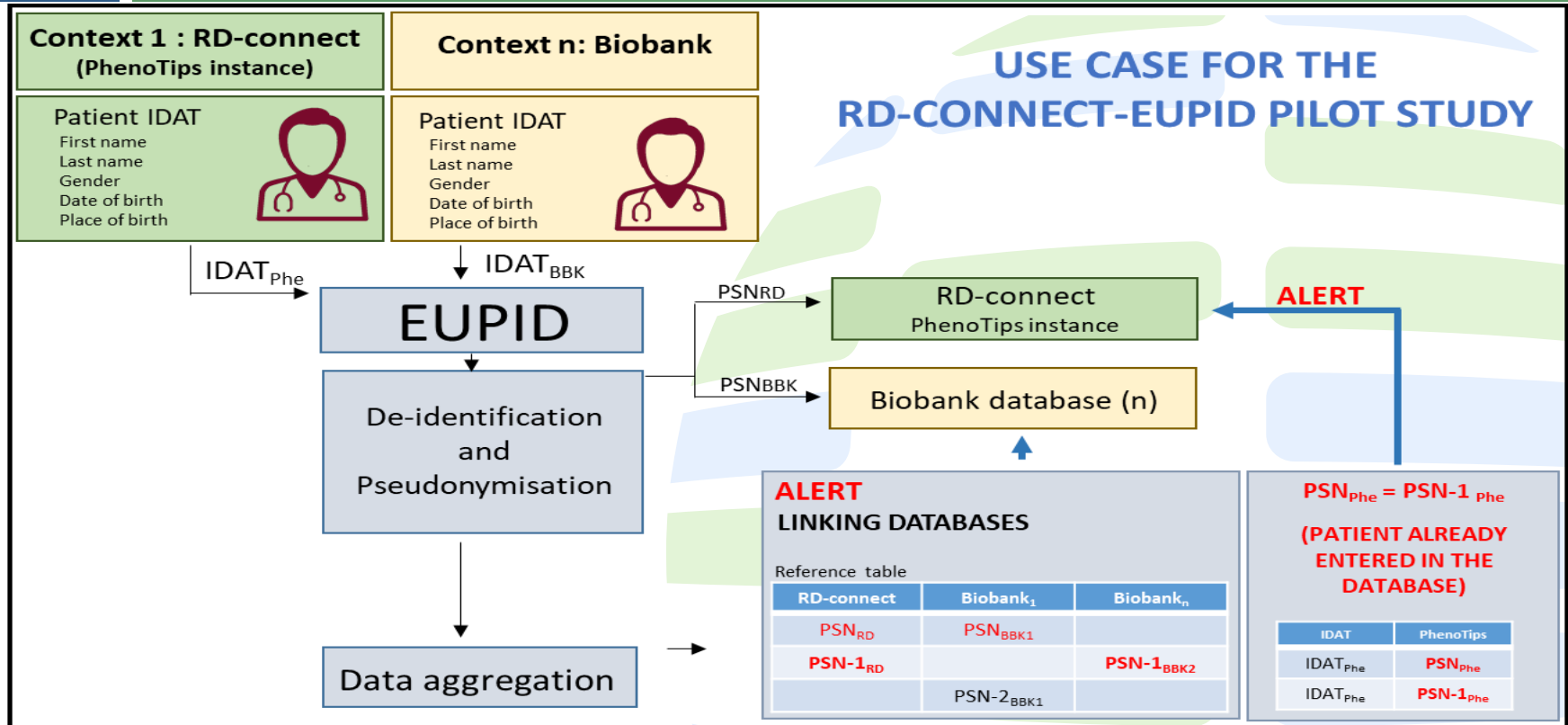
The use of data standards and ontologies enables the RD-Connect platform to be interoperable and interconnected (API) to other platforms and resources:

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




RD-Connect – EUPID pilot study

10



- Integrate pseudoanonymised data from different contexts
- Detect duplicate entries

Filters 

PRESET FILTERS


RESET


SHARE

RUN QUERY

Genes: gene-name

Sample Selection ?

Select individual Samples 

or search across all 

(accessible: 211, own: 1, shared: 0, v

Affected


ID

0/0

0/1

☐

E000001 NA12123





☐

☒

☒

E000040 MUN0788





☐

☒

Gene Name(s)

SDHA, CYC1, UQCR10, UQCRH

Select a predefined gene list

Upload comma separated list of HGNC identifiers

Seleccionar archivo

Ningún archivo seleccionado

MME patient matching request (v1)

Researcher username

rdconnect-annual

formal querier

RD-Connect

Date

26/04/2017

Patient (Phenotips id.)

P0000128

target endpoint

RD-Connect -> PhenomeCentral

Gene(s).(recovered from RD-Connect), Complete if necessary

SDHA, CYC1, UQCR10, UQCRH

HPO term(s).(recovered from RD-Connect), Complete if necessary

HP0001508 Failure to thrive; HP0002151 Increased serum lactate; HP0002878 Respiratory failure

CANCEL

SUBMIT

Candidate genes

- TAG
- PhenoTips

HPO terms

- PhenoTips



Matchmaker Exchange: initiate contact

12

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.

Acceptar



Data Sharing: GA4GH Beacon








13

GRCh37 ▾ 13 : 32954208 A>T Search

Response	All None
<input checked="" type="checkbox"/> Found	8
<input checked="" type="checkbox"/> Not Found	43
<input type="checkbox"/> Error	8

Organization All None

- ☐ AMPLab, University of C...
- ☐ BGI
- ☐ BioReference Laboratories
- ☒ Broad Institute
- ☐ Centre for Genomic Regu...
- ☒ CNAG
- ☐ Curoverse
- ☐ DNASTack
- ☒ EMBL European Bioinfor...
- ☐ Global Alliance for Geno...
- ☐ Google
- ☒ Institute for Systems Biol...
- ☒ Mike Lin
- ☒ National Center for Biote...

	EBI - 1000 Genomes Project, ... EMBL European Bioinformatics Institute	Not Found
	ExAC Broad Institute	Not Found
	ICGC - Cancer Projects Ontario Institute for Cancer Research	Not Found
	Kaviar Institute for Systems Biology	Found
	NHLBI Exome Sequence Proj... National Center for Biotechnology Information	Not Found
	RD-Connect CNAG	Not Found
		Not Found

Question:

Does any sample in
your database have
variant V?

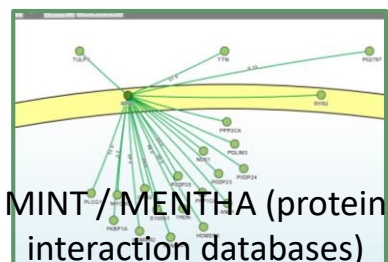
Answer:

Yes / No



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

14



Filters

PRESET FILTERS

RESET

SHARE

RUN QUERY

Variant Type: high moderate

Population: exac

Genes: gene-list

Samples

Functional

Predictive

Population

Pathways

Protein Interaction

Diseasecard

Candidate

Links

ALFA

Visualizer

Browser

RYR1

RYR1

Protein A (Gene name)

Protein B (Gene name)

Protein B: UniProt id

Type of interaction

Mentha score

RYR1

TRDN

Q13061

direct interaction

0.643

RYR1

FKBP1A

P62942

direct interaction

0.702

RYR1

HOMER1

Q96YM7

direct interaction

0.623

Phenotype

Analysis status

Variant (3)

Exoniser

First

Previous

Next

Last

EXPORT ALL

Chr

Pos

dbSNP

Ref

Alt

Candidate

G1000010

INDEL

Gene Name

Effect Impact

ClinVar

CADD

SIFT

PP2

MT

EXAC

1000GPAF

Internal Freq

12

32994056

rs147240502

A

C

0

Thc

A/C

PKP2

MODERATE

255/302

25.6

D

D

D

D

0.0047

0.0027

0.002009

16

15844046

rs111404182

G

A

0

Thc

G/A

MYH11

MODERATE

0/0

33

D

D

D

D

0.0005

0.0004

0.001211

19

39062815

G

C

0

Thc

G/C

RYR1

MODERATE

24.4

D

D

D

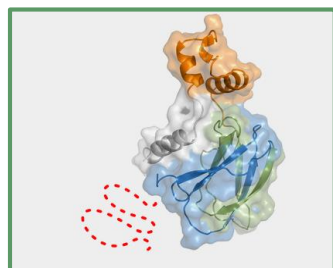
NA

0

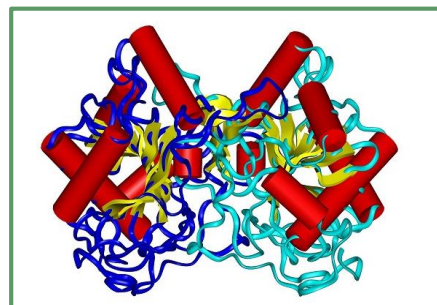
0.000192

eDGAR

disease-gene
Association database



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



GALT-PROTEIN-DB 2.0



REDportal

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



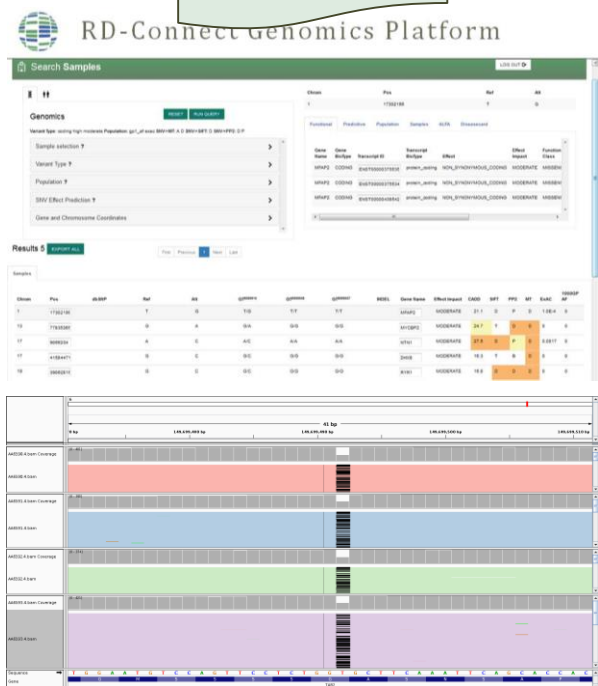
Human Mitochondrial DataBase



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

15

RD-Connect
BAM/CRAM

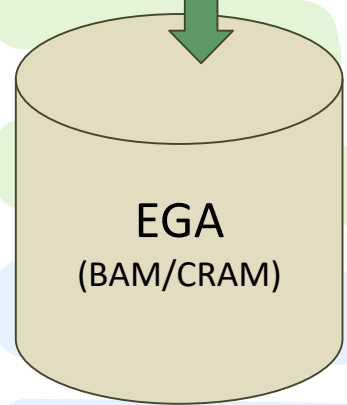


4 Rendering*

1 File submission to the EGA & ID-mapping

2 Request { File ID/s
Chr. coordinates

3 Reply { BAM/CRAM slice
or
Data representation



HTSget API

Secure and encrypted connections

*Broad's IGV screenshot



SolveRD RD3 database

16

- 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

17

- 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

18

- 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)

19

- ❑ <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

20

- 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 through 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



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 Bérout
(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)



S. Beltran	M. Ingham
S. Laurie	E. Casals
D. Piscia	R. Alcántara
L. Matalonga	
A. Papakon.	I. Gut
D. Picó	M. Gut
R. Tonda	M. Bayès
J.R. Trotta	B. Fusté
G. Parra	CNAG lab
C. Luengo	CNAG admin
J. Camps	
I. Martinez	
D. Ovelheiro	
J. Protasio	



BBMRI – Large Prospective Cohorts

