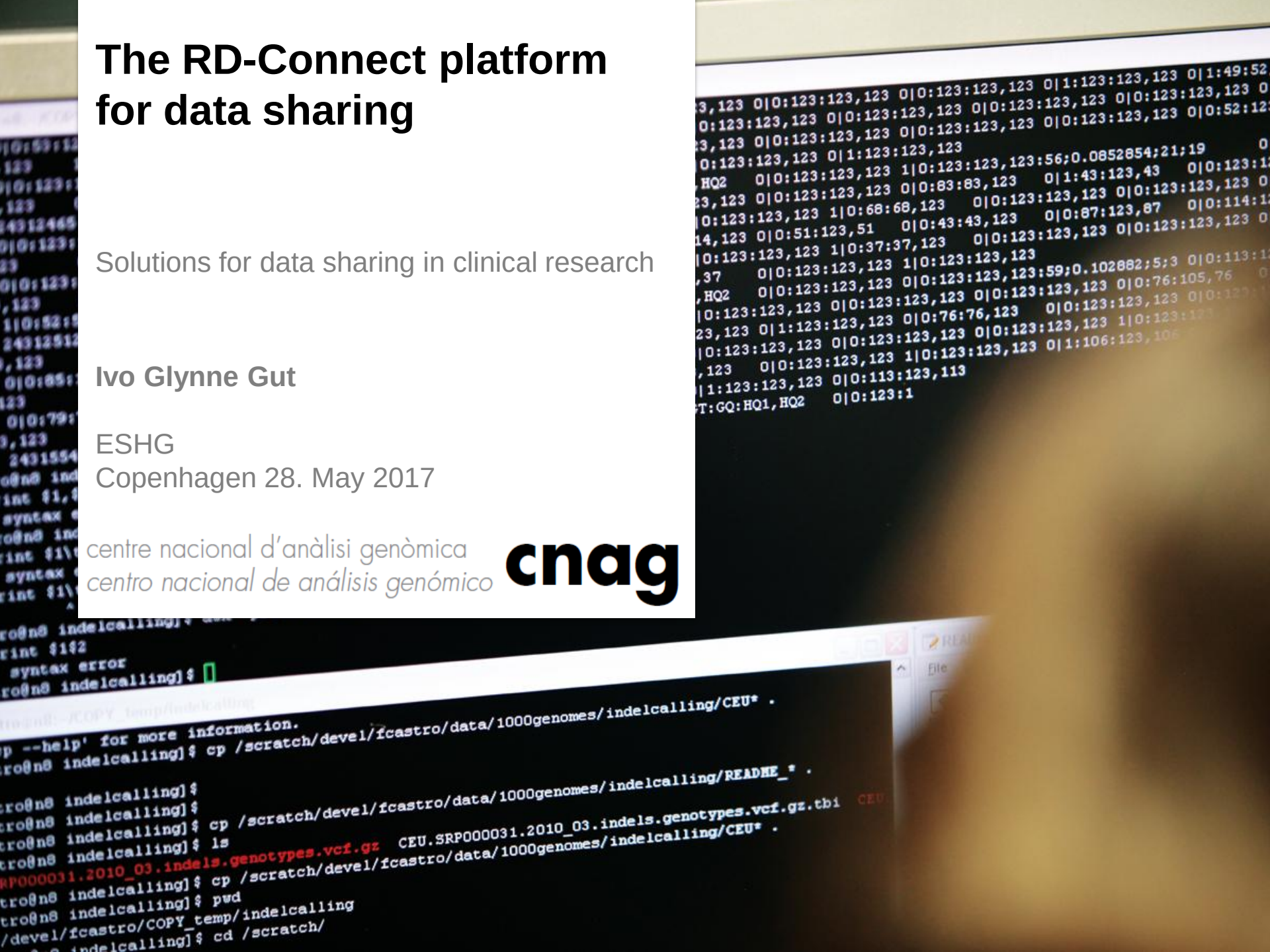
The RD-Connect platform for data sharing

Solutions for data sharing in clinical research

Ivo Glynne Gut

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico
cnag

ESHG
Copenhagen 28. May 2017



6-year project funded by EU 7th Framework Programme

An integrated platform connecting –omics data, clinical information, registries, and biobanks for rare disease research

- Contributing to **IRDiRC objectives** of delivering **200 new therapies** for rare diseases, and **means to diagnose most rare diseases** by 2020
- Creating a central system for **reprocessing**, **storage**, **analysis** and **sharing** of -*omics* data
  - Including integration of **phenotypic** and **biosample** data, and development of **new bioinformatic tools** to aid detailed analysis

Introduction to Rare diseases: data fragmentation

RD Connect

Infrastructure for data sharing in rare disease research

Flagship IRDiRC project implementing IRDiRC policies and guidelines on data sharing

EU 7th Framework Programme, 12M EUR, 6 years

Genomic analysis and gene discovery

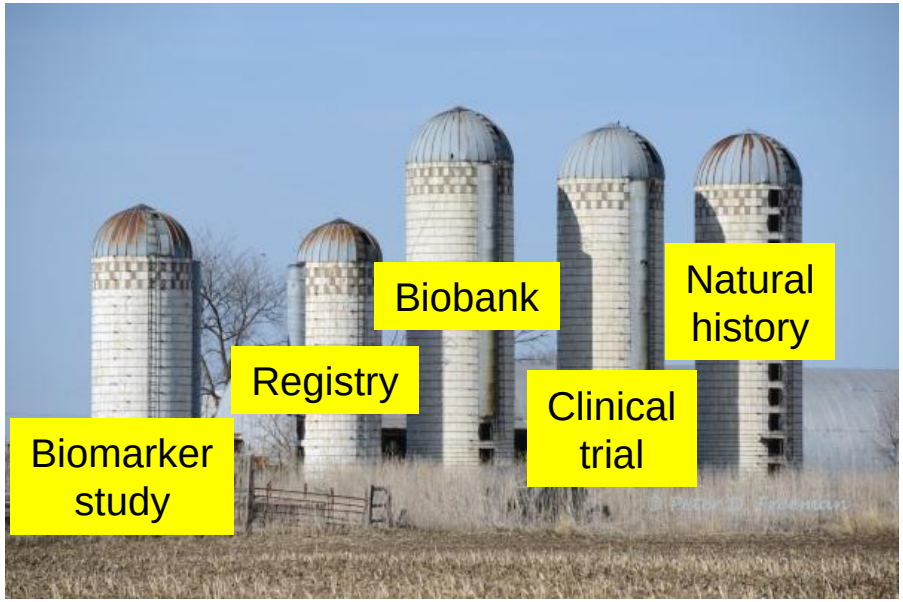
Standardized phenotypic data collection

Searchable catalogue of biosamples

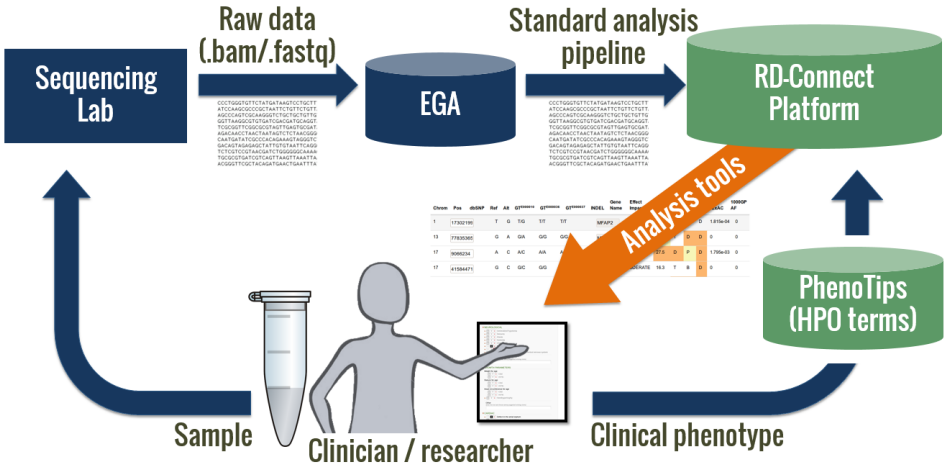
Data linkage across resources

Overcoming Silos

Data sharing for research and better data analysis



Omics data, clinical data and biosamples from individual with RD



Disease-causing variant can be identified using the **genomics analysis platform**

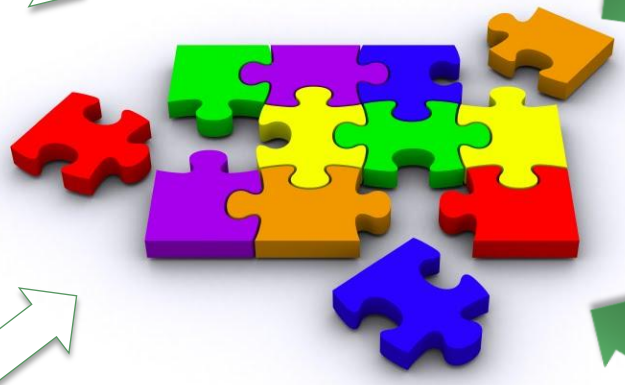
Sample is findable in the **Sample Catalogue**

Registry data in the **ID-Cards directory** of registries and biobanks

Data in the RD-Connect Platform

Genomic data
WES, WGS,
gene panels

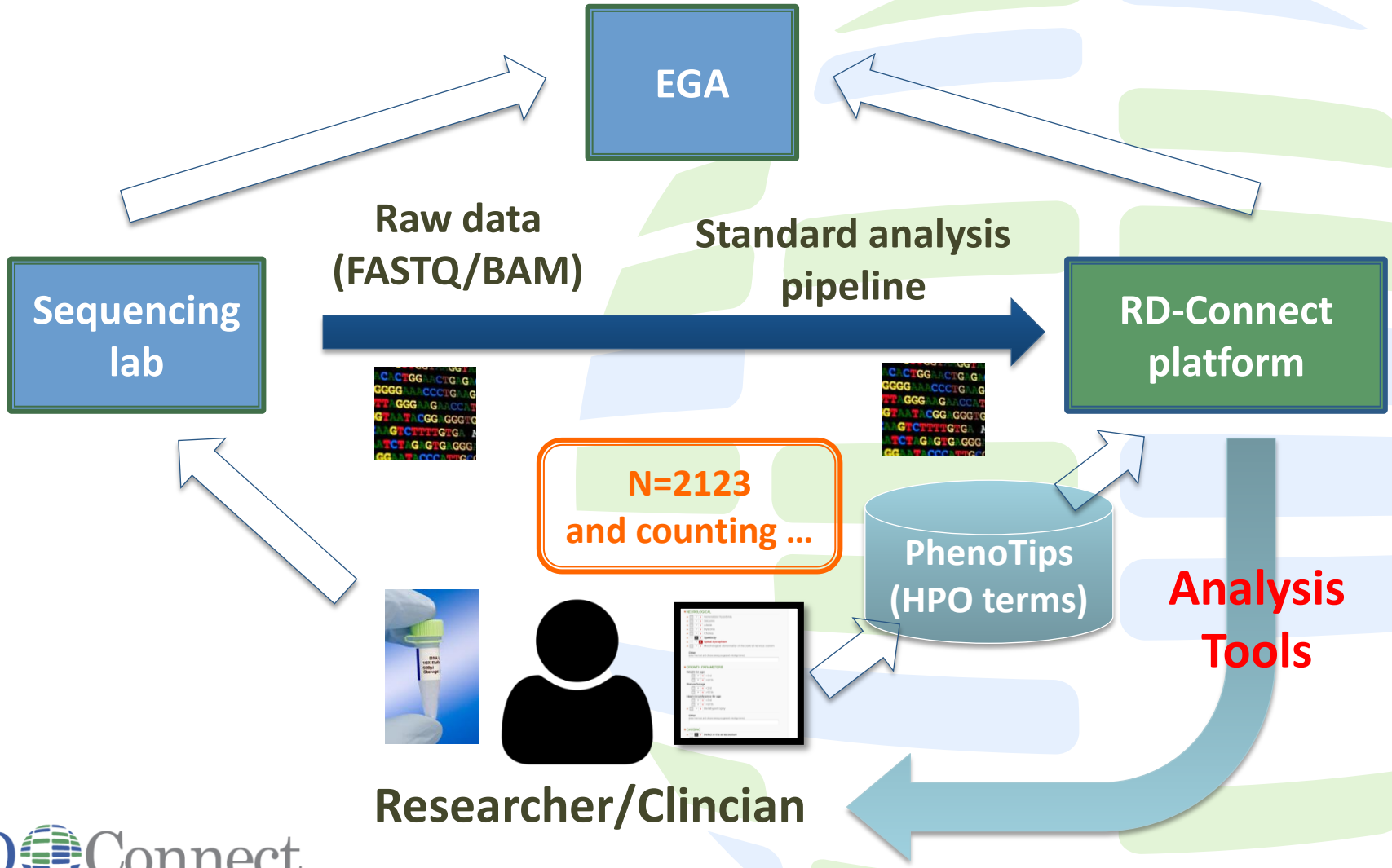
Clinical data
phenotypic data,
patient registries



Sample data
biobanks

Other omics data
transcriptomics,
metabolomics,
proteomics …

Geno:pheno data flow in RD-Connect



Data Submission Workflow

RD Connect SUBMISSION PLATFORM ABOUT WELCOME FAQ LOGOUT

Submission Data to RDconnect

See instructions

1- Submit Participant Set

2- Submit Experiment Set

3- Upload data

Manage your RDconnect data

Participant

Experiment

Uploaded files

Submit participants (linked to ID-Cards patient registries)

RD Connect SUBMISSION PLATFORM

Submit Participant/s to RDconnect

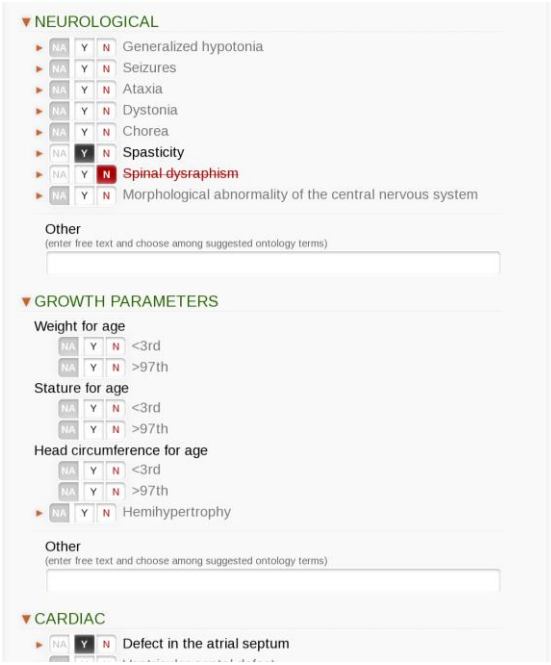
See instructions

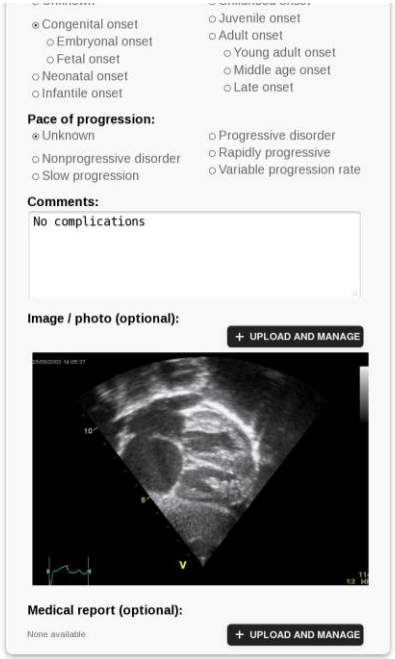
+ | +10 | +100 | VALIDATE AND CONTINUE EDITING | SAVE

Group assigned to the entries: cnag

PhenoTips ID *	Patient Registry	Matchmaker exchange	Sex *	Study	Mode of inheritance	Consanguin
P0000700 *Phenotips_ID already exists in the database*	× NBS Connect(ID=11509) ×	☐	*This field is required.*			
P0000701	× European patient registry on autoimmur × RDCRN Contact registry(ID=14682) ×	☑	*male*	*CMD/CM*	*unknown*	*unknown*
		☐				

RD-Connect:PhenoTips Instance



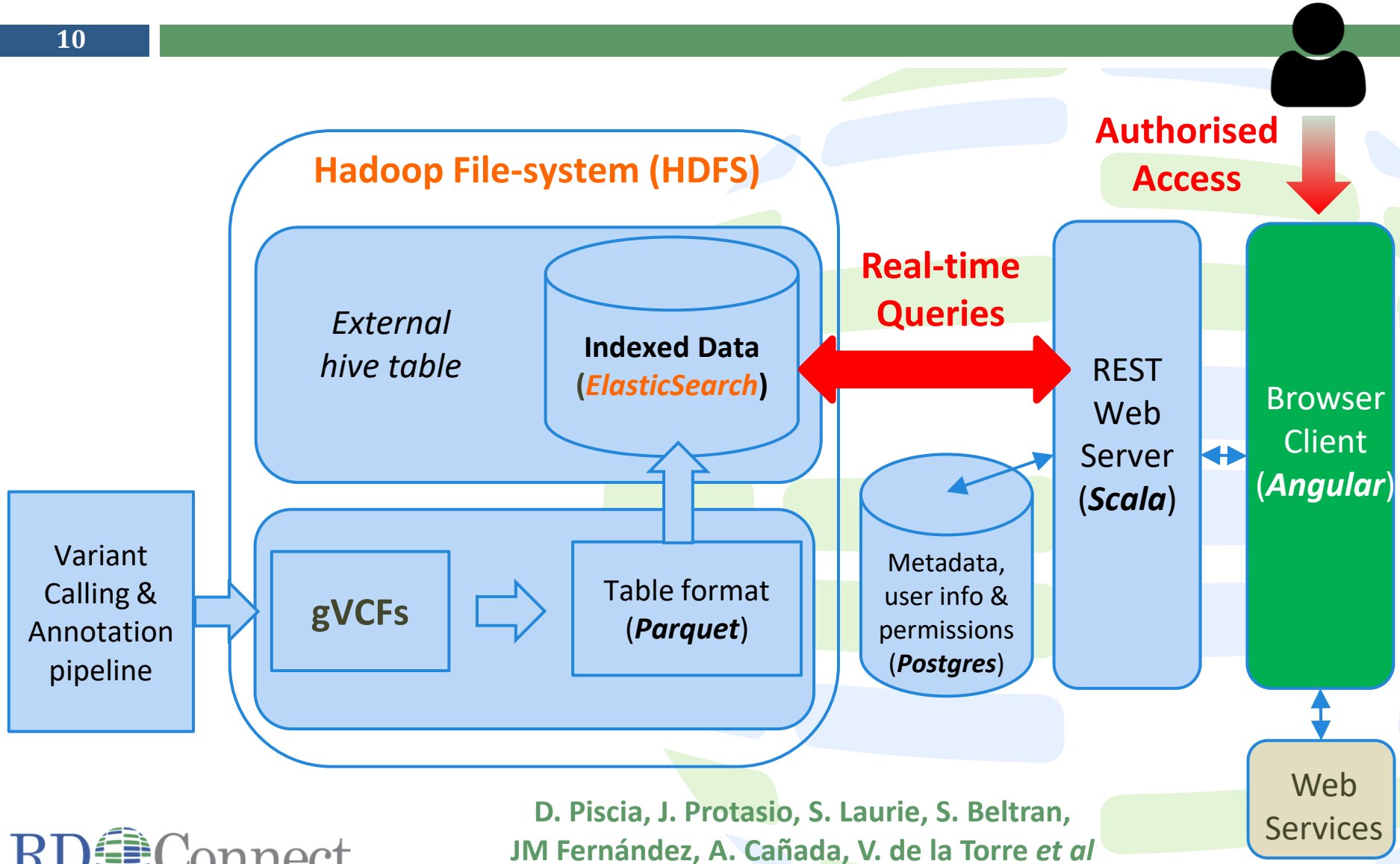


Deep phenotyping in PhenoTips (Brudno *et al.*) achieved using the Human Phenotype Ontology (HPO – Robinson, Köhler *et al.*)

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. variant prioritization, MME)

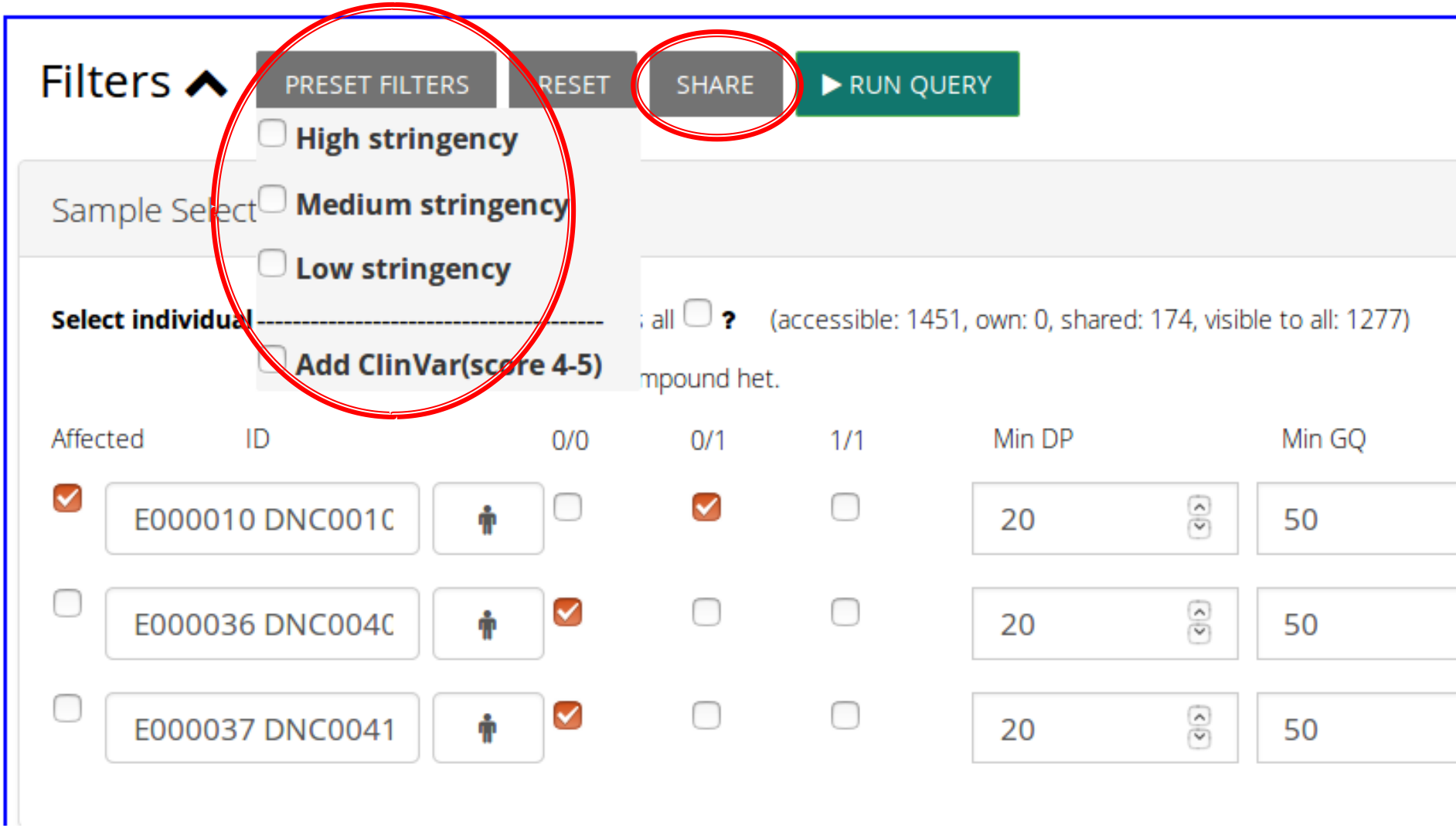
Genomics platform architecture



D. Piscia, J. Protasio, S. Laurie, S. Beltran, JM Fernández, A. Cañada, V. de la Torre *et al*

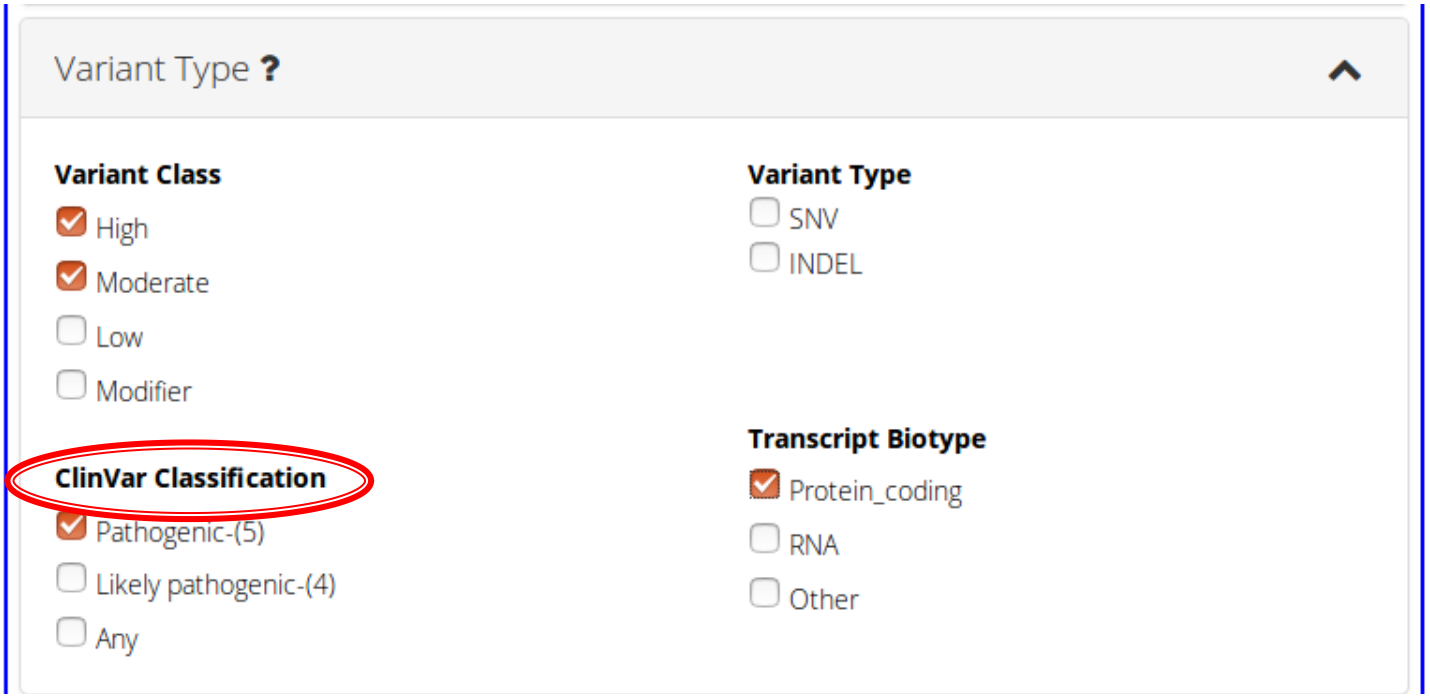
RD Connect

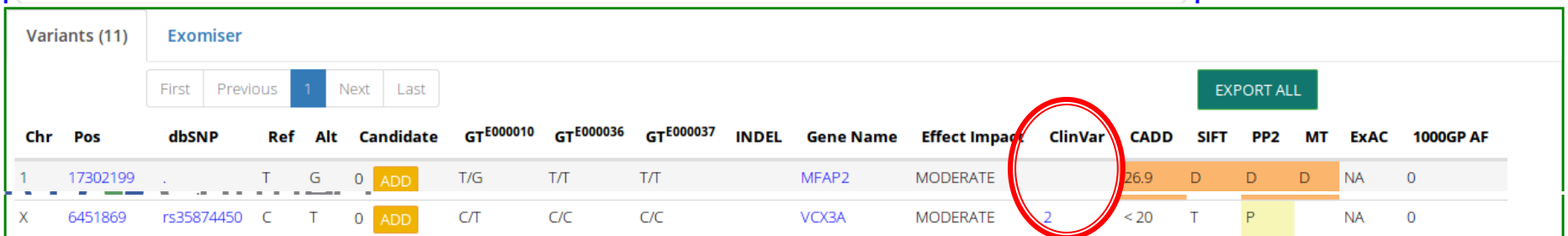
Select samples, set filters and share queries



Filters: variant type

- ClinVar can be used for filtering, and ClinVar categories are shown
- Started conversations to explore integration of HGMD





Filters: lists of genes

Genes, Disorders and Phenotypes

Gene Name(s)

e.g.:OTUD3

Select a predefined gene list

ACMG Medically Actionable Genes (n=59)
BabySeq Class A and B Genes (n=889)
Digenic gene list, Feb2017 (n=136)
Imprinted-confirmed list, Feb2017 (n=80)
Imprinted-all, Feb2017 (n=253)
Medically Interpretable Genome (n=5,419)
Muscle Gene Table, July2016 (n=416)
Muscle Gene Table, Sep2015 (n=403)
Mitocarta 2.0 (n=1158)
Tubingen HSP Version 6 (n=140)
Tubingen SCA Version 8 (n=183)

Added more lists of genes

OMIM and HPO related genes accessed through OMIM and PhenoTips APIs

Search OMIM

#254300 MYASTHENIC SYNDROME,

Genes linked to 254300 : DOK7, C4orf25, CMS10

Search HPO

HP:0009053 Distal lower limb musc

Genes linked to HP:0009053 : SLC46A1, IFT172, DCAF8, ANTXR2, GLE1, RPS6KA3, FAM134B, ZFYVE26, TBK1, TFG, AP1S2, SLC12A6, PRKACA, SIL1, MCCC2, CPT1A, ENTPD1, MUSK, PHKA1, CPT1C, SOX10, HSPG2, SLC5A7, AFG3L2, AR, TRNL1, WDR81, BIN1, PRKAR1A, SUCLG1, WDPCP, TMEM126B, DNA2, PFN1, GNE, MTMR2, DHH, PIK3R2, C9ORF72, C12ORF65, TTPA, NEU1, COX3, COX2, COX1, PIP5K1C, C5ORF42, ABCA1, BBS2, BBS1, EGR2, DMPK, INSR, SLC33A1, PYCR2, BBIP1, ERLIN2, ACADSB, SETX, FAM111B, GJB1, WNK1, NDUFAF4, TDP1, NDUFAF5, MKS1, NDUFAF2, OGDH, CD36, SPRTN, SFXN4, NDUFAF3, ATM, ALDH18A1,

Results (integrated data and links)

Samples | Functional | Predictive | Population | Diseasecard | Candidate | Links | ALFA

Gene Name	Transcript ID	Effect Impact	Consequence	Feature Type	HGVS coding	Amino Acid change	Amino Acid length	Genotype Number	Exon Rank	CDS Position	Transcript BioType
MFAP2	ENST0000037553	MODERATE	missense_variant	transcript	c.313A>C	p.Thr105Pro	183	1	7/9	313/552	protein_coding
MFAP2	ENS 5534	MODERATE	missense_variant	transcript	c.310A>C	p.Thr104Pro	182	1	6/8	310/549	protein_coding
MFAP2	ENS 542	MODERATE	missense_variant	transcript	c.310A>C	p.Thr104Pro	182	1	7/9	310/549	protein_coding

ExAC
Ensembl
HSF

Variants (11) | Exomiser

First | Previous | 1 | Next | Last

EXPORT ALL

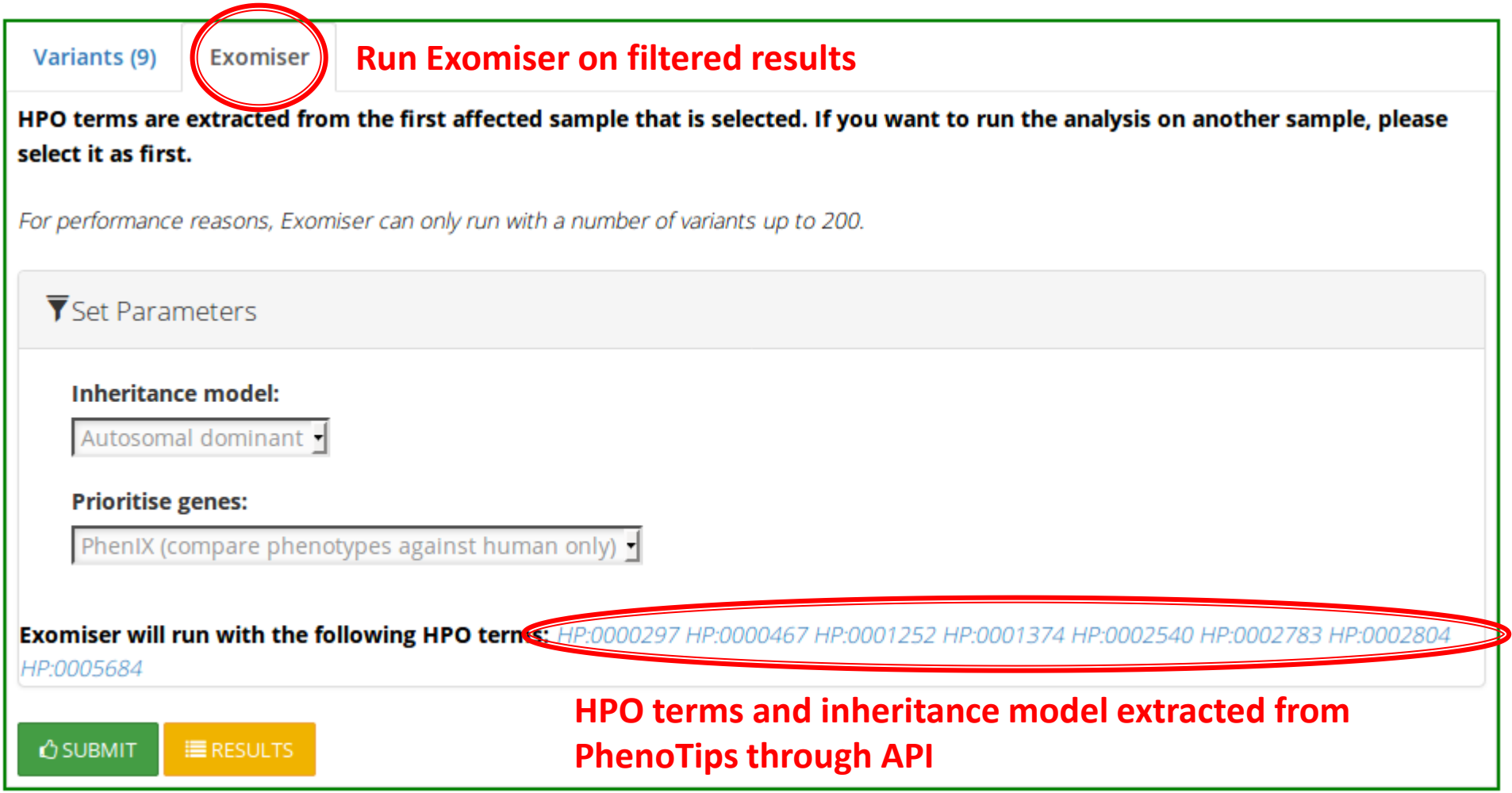
Chr	Pos	dbSNP	Ref	Alt	Candidate		$GT^{E000010}$	$GT^{E000036}$	$GT^{E000037}$	INDEL	Gene Name	Effect Impact	ClinVar	CADD	SIFT	PP2	MT	ExAC	1000GP AF
1	17302199	.	T	G	0	ADD	T/G	T/T	T/T		MFAP2	MODERATE		26.9	D	D	D	NA	0
11			C	A	0	ADD	C/A	C/C	C/C			MODERATE		34	D	D	D	NA	0
13			G	A	0	ADD	G/A	G/G	G/G			MODERATE		25.6	T	D	D	NA	0
13			A	C	0	ADD	A/C	A/A	A/A			MODERATE		24.6	D	D	D	NA	0
13			G	T	0	ADD	G/T	G/G	G/G			MODERATE		< 20	T	B	D	NA	0
17			A	C	0	ADD	A/C	A/A	A/A			MODERATE		26.7	D	P	D	NA	0
17			A	C	0	ADD	A/C	A/A	A/A			MODERATE		21.4	D	D	D	NA	0
1			G	C	0	ADD	G/C	G/G	G/G			MODERATE		20.6	T	B	D	NA	0
1			G	C	0	ADD	G/C	G/G	G/G			MODERATE		24.4	D	D	D	NA	0
22	30768124	.	T	G	0	ADD	T/G	T/T	T/T			MODERATE		27	T	D	D	NA	0
X	6451869	rs35874450	C	T	0	ADD	C/T	C/C	C/C			MODERATE	2	< 20	T	P		NA	0

Ensembl
ExAC
gnomAD
UCSC
NCBI
DGVa
GWAS Central
GA4GH Beacon
VarSome

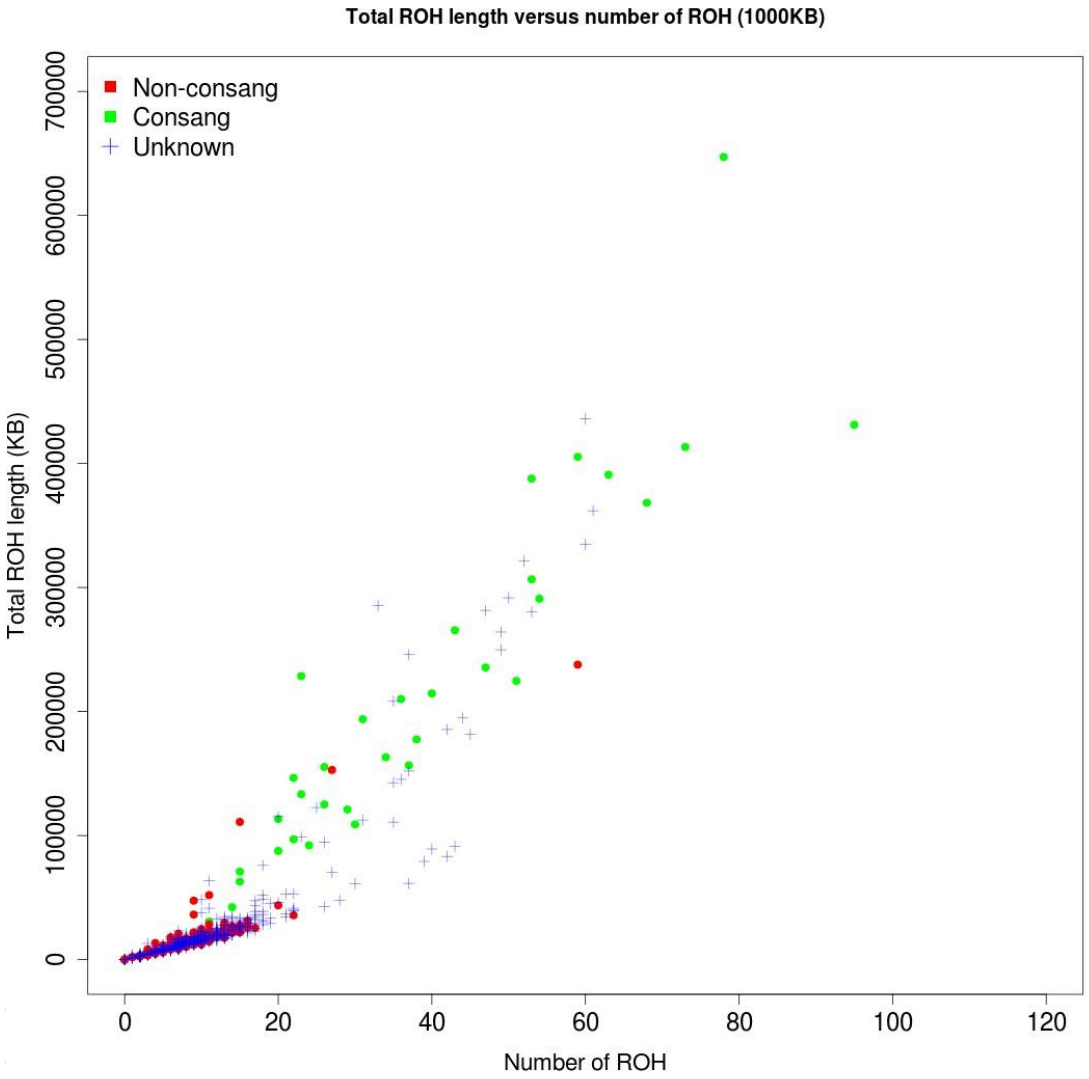
OMIM
Ensembl
PubMed
HGMD
Entrez
GeneCards
COSMIC
ClinVar
ExAC
GTEx
gnomAD
GWAS Central
ATLAS
WikiPathways
Open PHACTS

RD Connect

Exomiser for prioritising variant lists



Runs of homozygosity

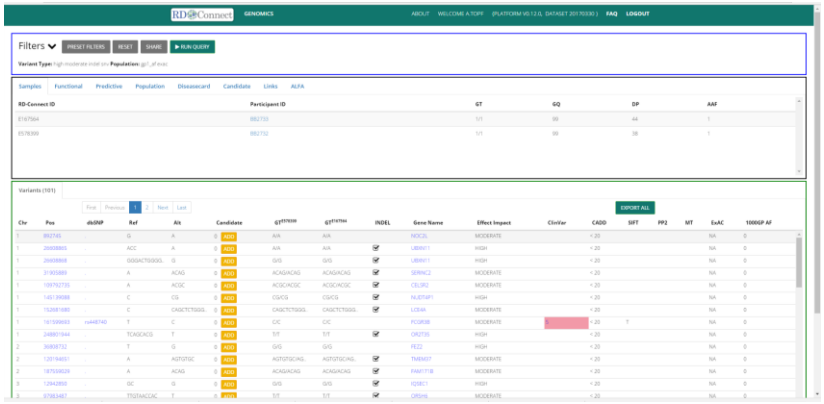


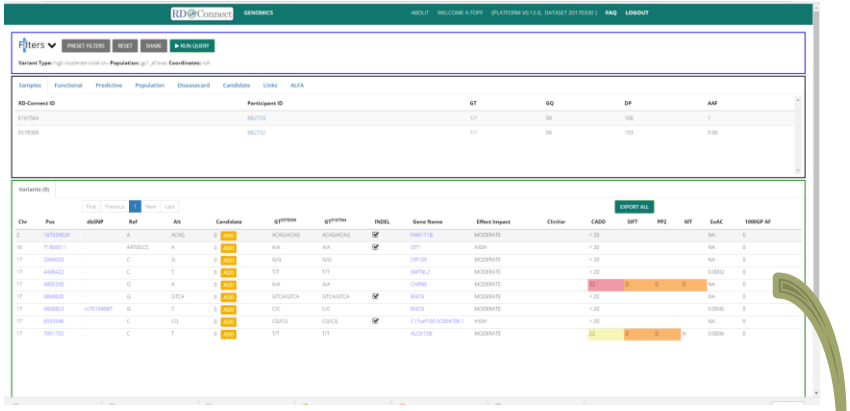
- Allows identification of consanguineous cases even when not identified as such
- Narrows down regions to focus gene discovery search
- NeurOmics PIs with interesting candidates in these regions have been contacted – feedback welcome

Homozygosity filter

Standard filters – 101 variants

Homozygous blocks – 9 variants





17	4805260	G	A	A/A	A/A	CHRNE	MODERATE	32	D	D	D	NA	0

BBMRI-LPC Whole Exome Sequencing Call for RD (2016)

Goal:
to promote the utilization of cutting-edge next-generation sequencing technology for the identification of novel causative variants and genes and to molecularly diagnose rare disease patients. BBMRI-LPC also wants to promote biobanking for rare diseases, the use of rare diseases biobanks and responsible data sharing.

To sequence and analyse:
900 exomes in 17 coordinated projects.

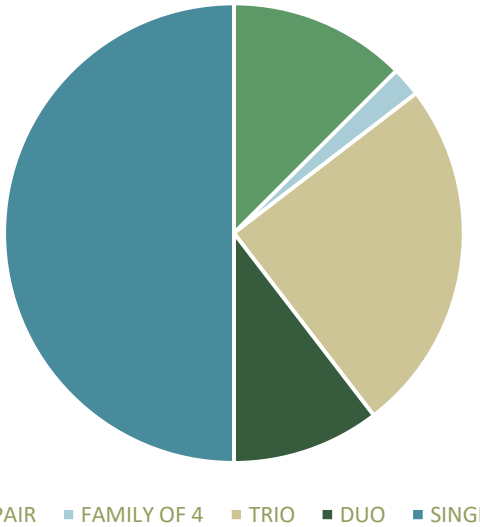
Sequencing and analysis carried out at the CNAG-CRG and the Wellcome Trust Sanger Institute (WTSI).

Project results are being released through RD-Connect, where researchers can analyse their data

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico
CRG
Centre for Genomic Regulation

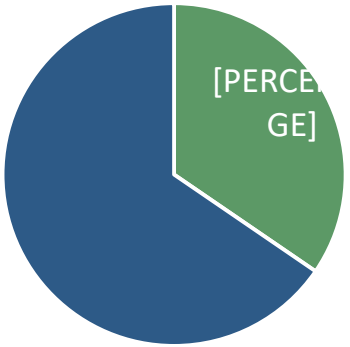
BBMRI-LPC 1st subcohort (Newcastle/Munich)

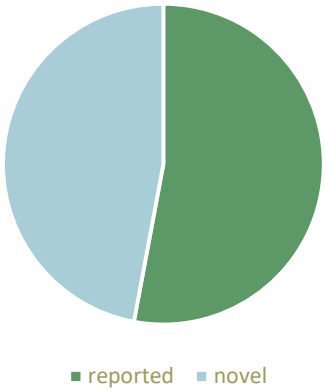
- Congenital myasthenic syndrome (majority), myopathy and neuropathy cases
- Investigators (Senderek, Lochmüller) signed adherence forms with RD-Connect; data on RD-Connect platform
- Samples deposited in Biobank (EuroBioBank/RDC catalogue)
- Sequencing (WES) with CNAG
- Clinical data uploaded on PhenoTips
- 87 samples: 47 kindred; 55 affected

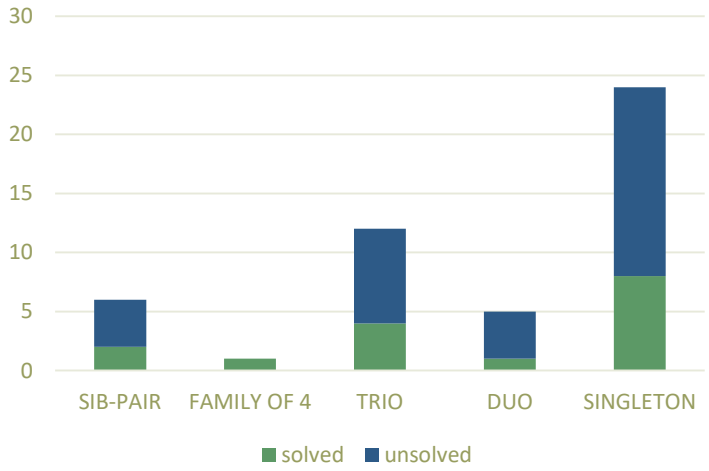


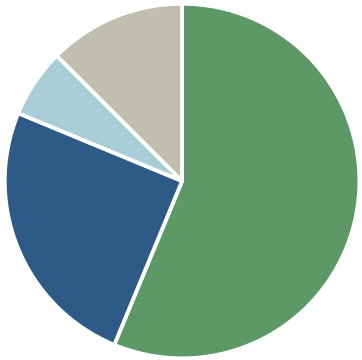
Detection rate

- Data analysed on the RD-Connect genomic platform
- 16 families solved so far
- 9 novel variants identified









homozygous heterozygous compound heterozygous de novo

Project	Phenotype	PI	Total N	Status
NeurOmics	NMD/NDD	Multiple	?1000	680 WES available; 65 WGS available pending HW upgrade; further upload by deCODE required
EURenOmics	Kidney	Multiple	?1000	453 panel/26 WES received; processing pending
Neptune	Kidney	Sampson	450	Pending receipt
CNAG RD	Multiple RD	Multiple	300	Pending consent
Neurogenetics	Neurogenetic/ mitochondrial	Horvath: Newcastle	336 WES	Received; processing pending
SeqNMD	NMD	Newcastle	169 WES	122 available; 47 failed – reprocessing pending
MYO-SEQ	LGMD	Straub: Newcastle	1000 WES (all index cases)	27 solved cases available; additional solved pending; unsolved pending PI consent
Titinopathies	NMD/ titinopathy	Udd/Hackman/ Savarese	(I) 76 panel / 15 WES / 12 WGS (II) 2000 WES/WGS	(I) Received; processing underway (II) To be submitted
Rare Immuno	Rare immunodeficiencies	Hambleton: Newcastle	160 WES	Received 38; processing pending
BBMRI-LPC	17 projects / multiple RD	Multiple	900 WES	About half are now complete
Consequitur	Consanguineous neurogenetic	Lochmüller: Newcastle	500 WES	New project; sequencing pending
SERBORDISinn	Multiple RD	Pavlovic: Belgrade	9 WES	Available
Sayer Group	Kidney	Sayer: Newcastle	30 WES	Patient consent pending
ISCIII	Multiple RD	Posada: Madrid	23 WES	Available
Italian NMD	NMD	Nigro	tbc	Negotiation underway
Telethon UDP	Syndromic undiagnosed	Multiple	tbc	Negotiation underway

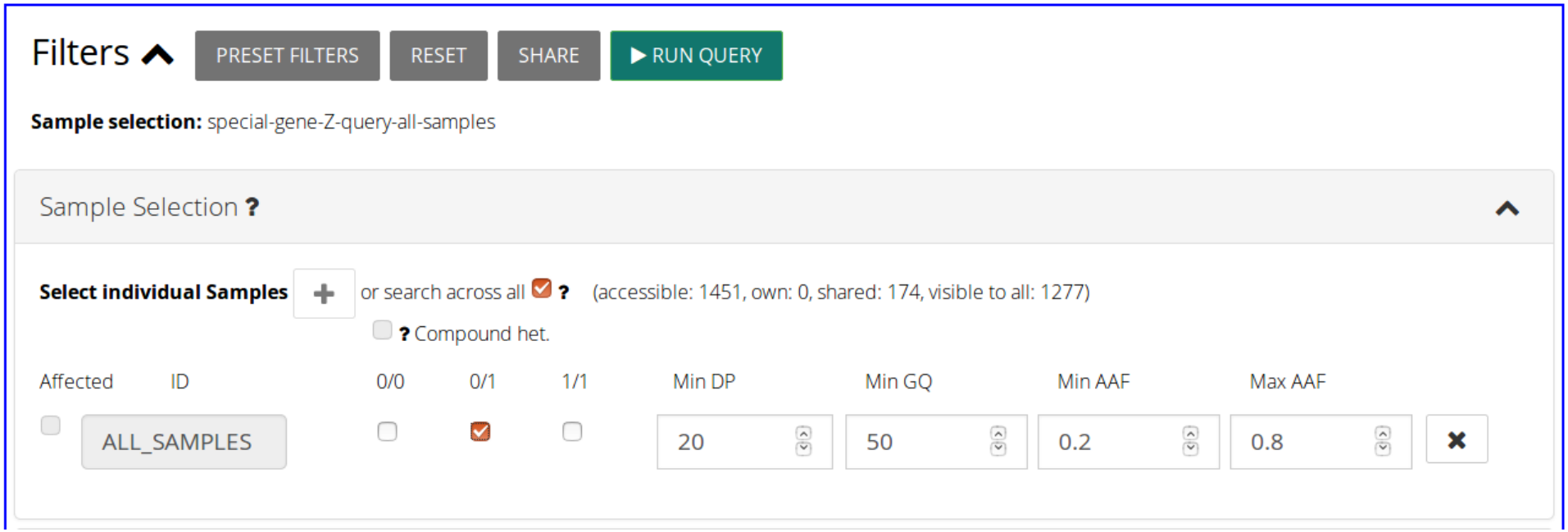
RD-Connect : Data sharing

RD-Connect enables:

- Full data sharing and analysis within RD-Connect for authorized users
- Partial data sharing outside RD-Connect, in accordance with ethical and legal limitations

Search across samples (per gene/s) with all filters

- Find all cases with a candidate mutation in a gene of interest



RD-Connect : Data sharing

RD-Connect enables:

- Full data sharing and analysis within RD-Connect for authorized users
- Partial data sharing outside RD-Connect, in accordance with ethical and legal limitations

Data Sharing: GA4GH Beacon

Posted: May 29, 2015

Beacon Project

Being implemented on the website of the world's top genomic organizations to test the willingness of international sites to share genetic data.

Beacon

About this Project

The **Beacon project** is a project to test the willingness of international sites to share genetic data in the simplest of all technical contexts. It is defined as a simple public web service that any institution can implement as a service. The service is designed merely to accept a query of the form "Do you have any genomes with an 'A' at position 100,735 on chromosome 3" (or similar data) and responds with one of "Yes" or "No." A site offering this service is called a "beacon". This open web service is designed to be technically simple, easy to implement, and to not return privacy violating information.

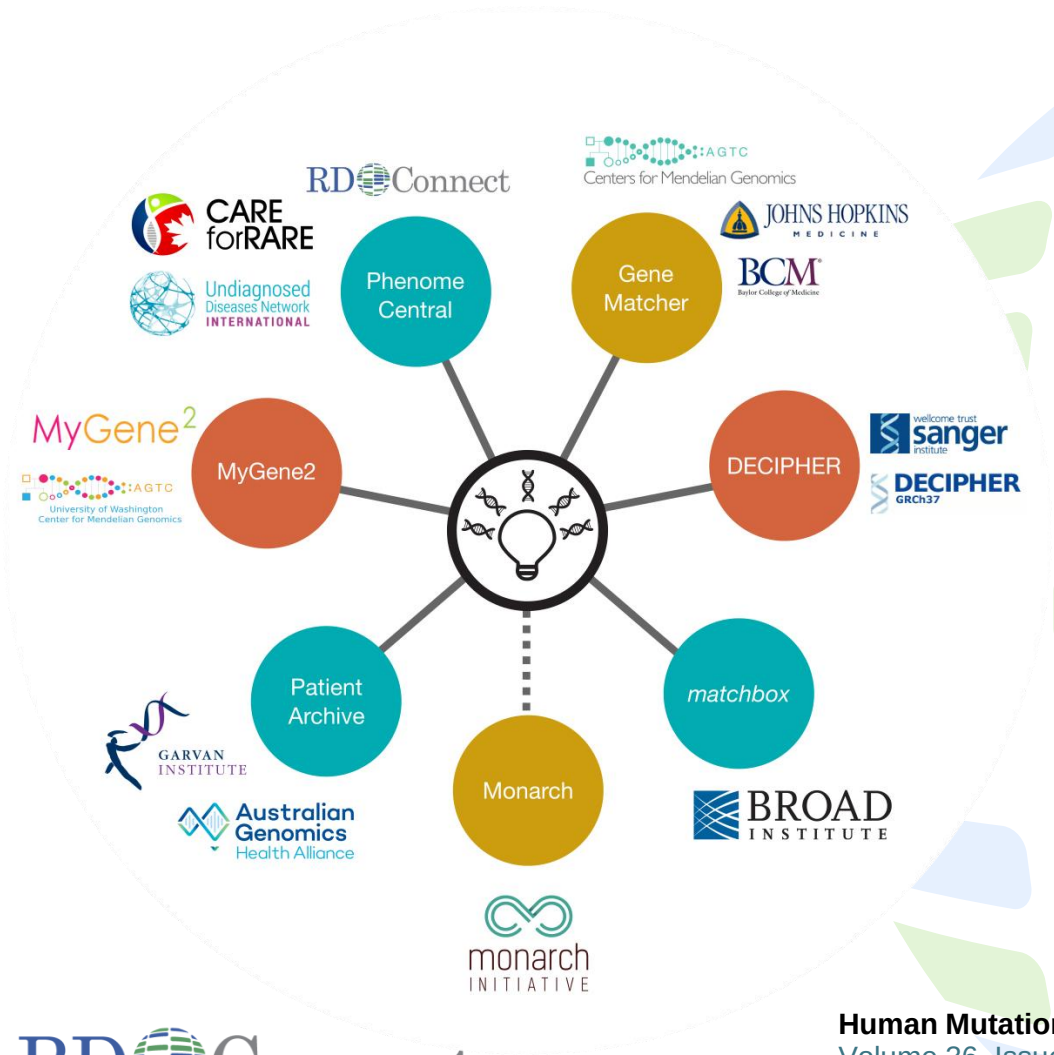
For current Beacons, and a short guide about how to light a Beacon, please visit: Beacon Network »

Question:
Have you seen this **variant** in **any** sample in your database?

Answer:
Yes / No

https://genomicsandhealth.org/work-products-demonstration-projects

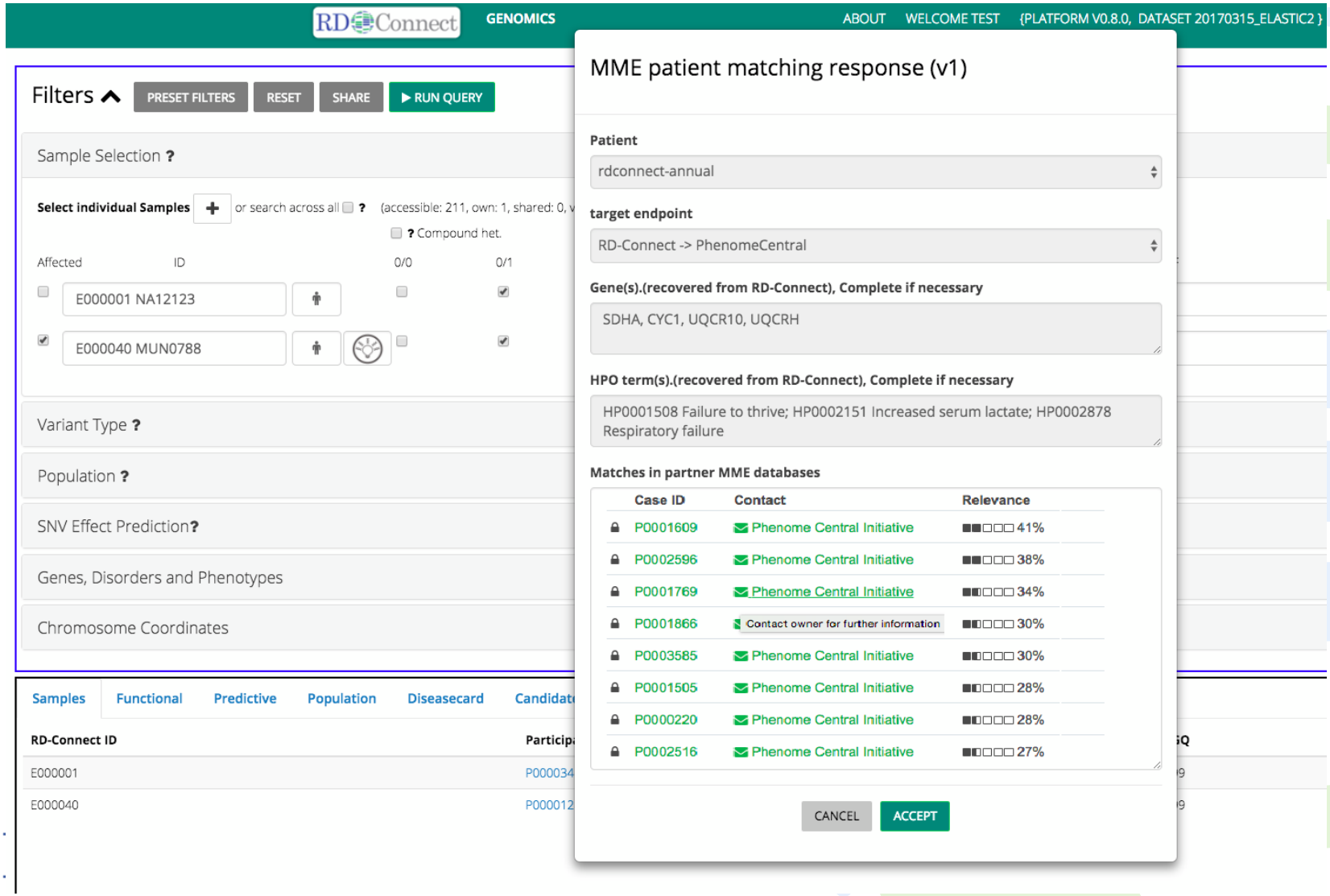
MatchMaker Exchange (MME, IRDiRC, GA4GH)



Question: Do you have a patient with similar phenotype and genotype as mine?

Human Mutation
Volume 36, Issue 10, pages 915-921, 17 SEP 2015 DOI: 10.1002/humu.22858
http://onlinelibrary.wiley.com/doi/10.1002/humu.22858/full#humu22858-fig-0001

MatchMaker Exchange in RD-Connect (in pre-production)



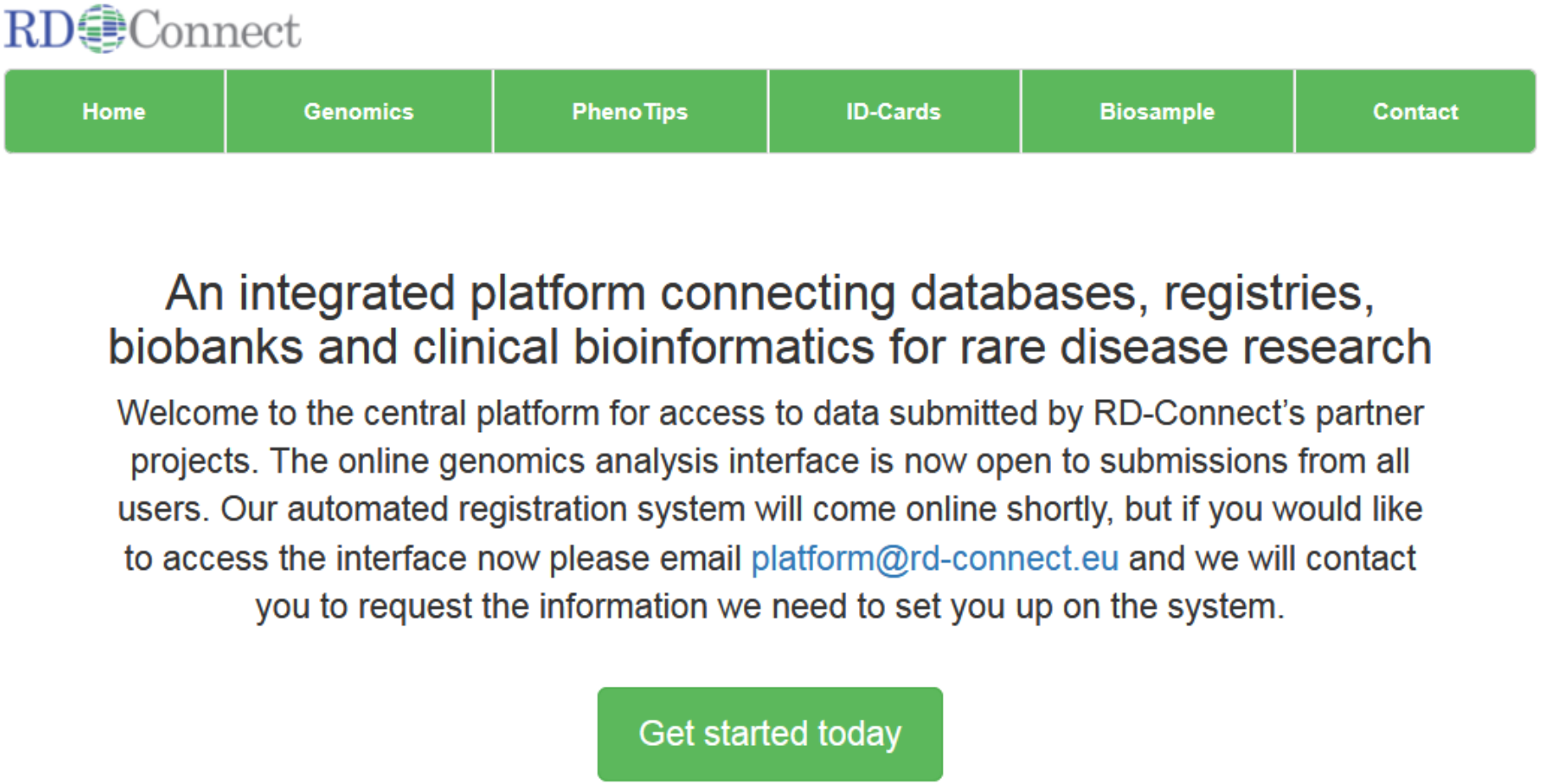
Ethical and Legal Issues

- Database registered in the Agencia Española de Protección de Datos

- To submit and/or to access donor data, a Code of Practice and and Adherence Agreement must be signed. Documents were approved by Comité Ètic d'Investigació Clínica del Parc de Salut Mar in 2015.

- Activity of the users is logged.

platform.rd-connect.eu



Contributors

RD-Connect

WP1: Coordination
Hanns Lochmüller
(Newcastle and TREAT-NMD)

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)

excelerate

WP2: Benchmarking
Soren Brunak & Alfonso Valencia

WP5: Elixir Interoperability Backbone
Barend Mons, Carole Goble, Helen Parkinson

WP8: Rare Diseases Use Case
Ivo Gut & Marco Roos

WP9: Human Data Use Case
Jordi Rambla & Helen Parkinson

WP11: Training Platform
Chris Ponting & Patricia Palagi

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG Centre for Genomic Regulation

S. Beltran
D. Piscia
S. Laurie
J. Protasio
A. Papakonstantinou
I. Martinez
R. Tonda
J.R. Trotta
L. Matalonga
M. Gut
J. Whalley
I. Gut

Generalitat de Catalunya
Departament d'Economia i Coneixement

Generalitat de Catalunya
Departament de Salut

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Instituto de Salud Carlos III

elixir

INB

EUROMICS

NEUROMICS

SEVENTH FRAMEWORK PROGRAMME

IRDiRC
INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM