RD-CONNECT: DATA SHARING AND ANALYSIS FOR RARE DISEASE RESEARCH

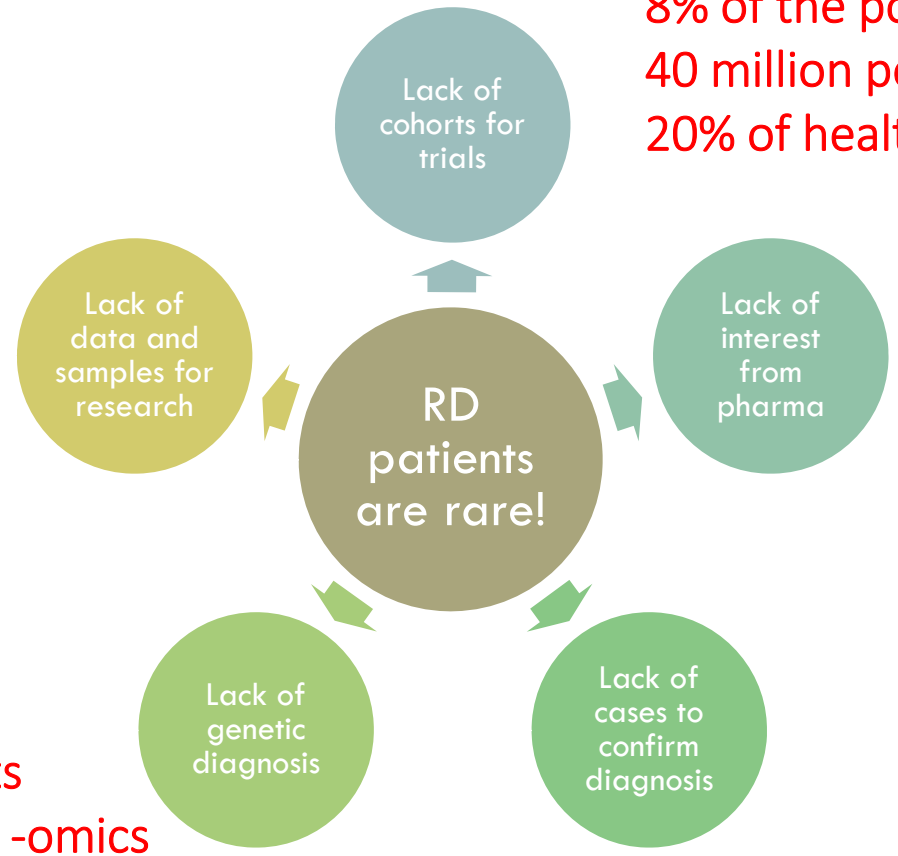
ANDREAS ROOS

INFRAFRONTIER/ IMPC STAKEHOLDER MEETING

14th November 2017
Athens, Royal Olympic Hotel

RARE DISEASES OFTEN LACK DIAGNOSIS AND TREATMENT, RESULTING IN DISABILITY AND LOSS OF QUALITY OF LIFE

8% of the population
40 million people in the EU
20% of health care spending



70% monogenetic
Pathways and targets
Genomics and other -omics
Genetic and advanced therapies

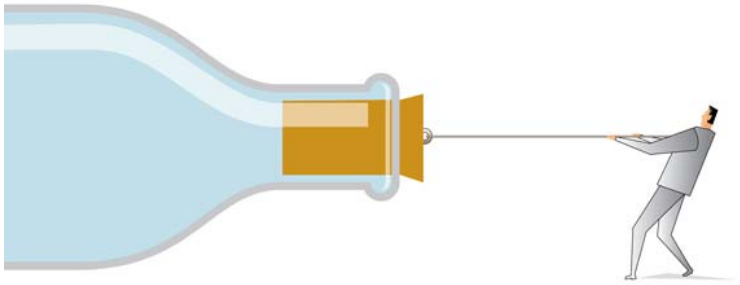
MANY RARE DISEASE BOTTLENECKS ARE CROSS-CUTTING

…across diseases and across research domains

A lot of them come down to data…

…Not just scarcity of data, but lack of options to reuse existing data

- Privacy protection issues, particularly across borders
- Lack of infrastructure for data sharing
- Lack of standards and interoperability
- Reluctance to share unpublished data
- Lack of capacity to analyse large amounts of data
- Challenges of linking different datasets in different places



SHARING: BENEFITS

Overcoming the “rare disease problem”

- Cohort size
- Powering trials
- Finding confirmatory cases

Reducing costs

Reducing duplication of effort

Facilitating validation of results

Enabling engagement with experts and the patient community

RD-CONNECT: INFRASTRUCTURE FOR RARE DISEASE DATA AND SAMPLES

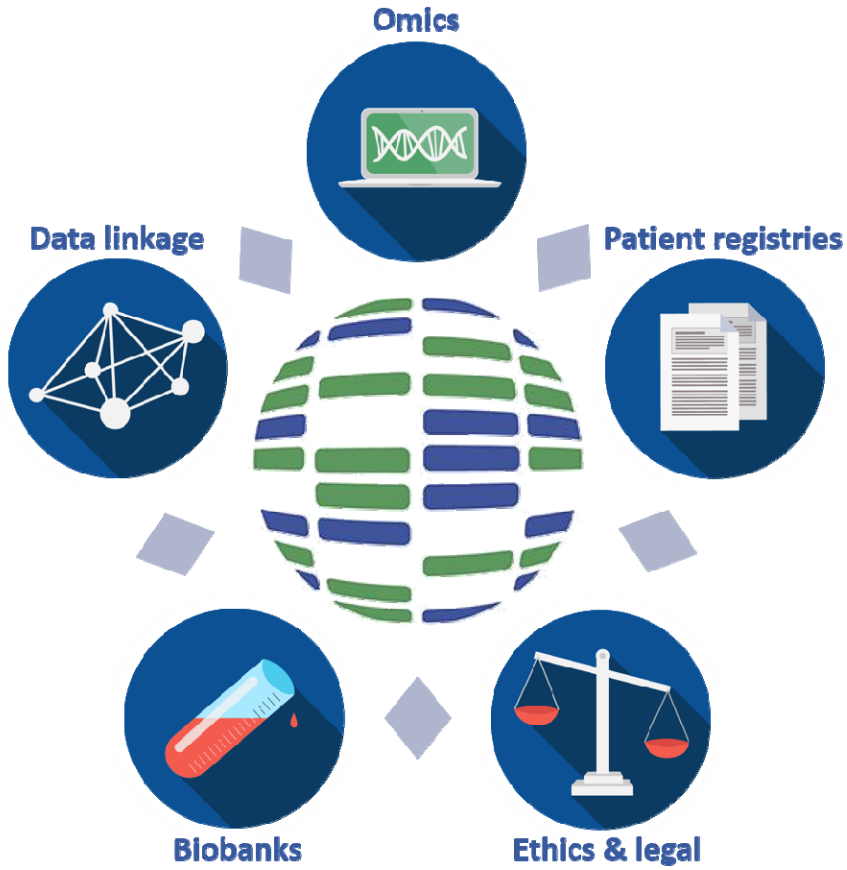
RD-Connect has created resources for use by the rare disease research community

Genome-phenome analysis platform – diagnostics and gene discovery on human data (thousands of individuals with rare disease)

Biosample catalogue – access and request the samples and cell lines you need to do your research

Registry and biobank finder – locate the resources that contain the data and samples you need

Data linkage – access expertise to make data FAIR (Findable, Accessible, Interoperable, Reusable)



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

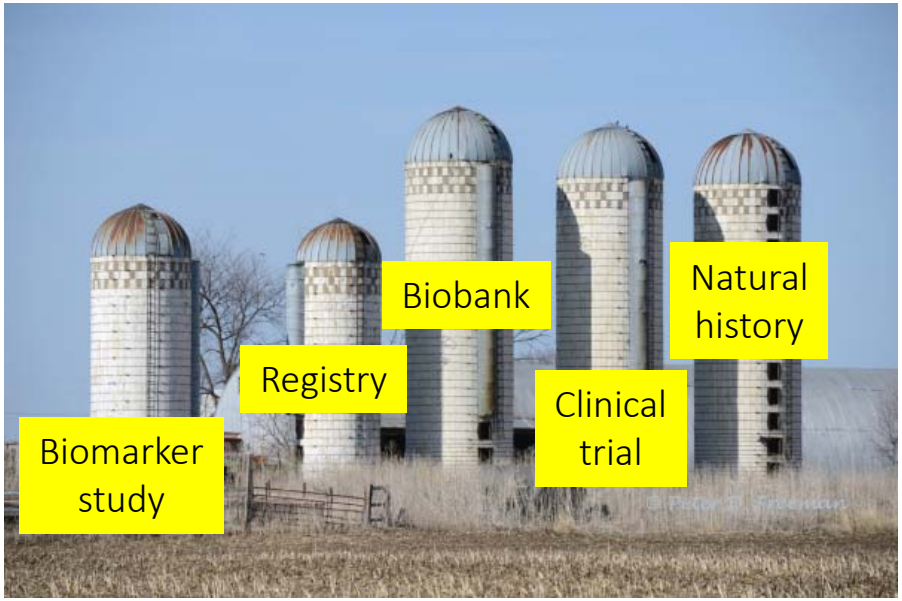
Standardized phenotypic data collection

Searchable catalogue of biosamples

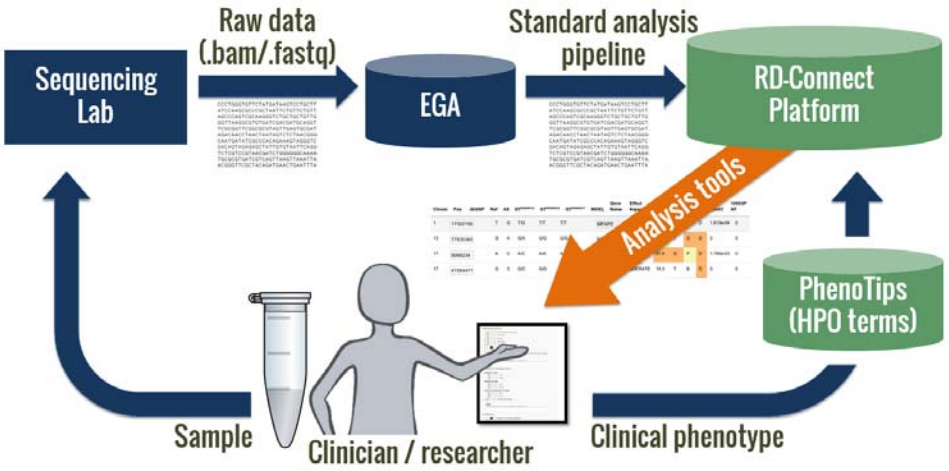
Data linkage across resources

Overcoming Silos

Data sharing for research and better data analysis



Omics data, clinical data and biosamples from individual with RD



Sample is findable in the Sample Catalogue

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

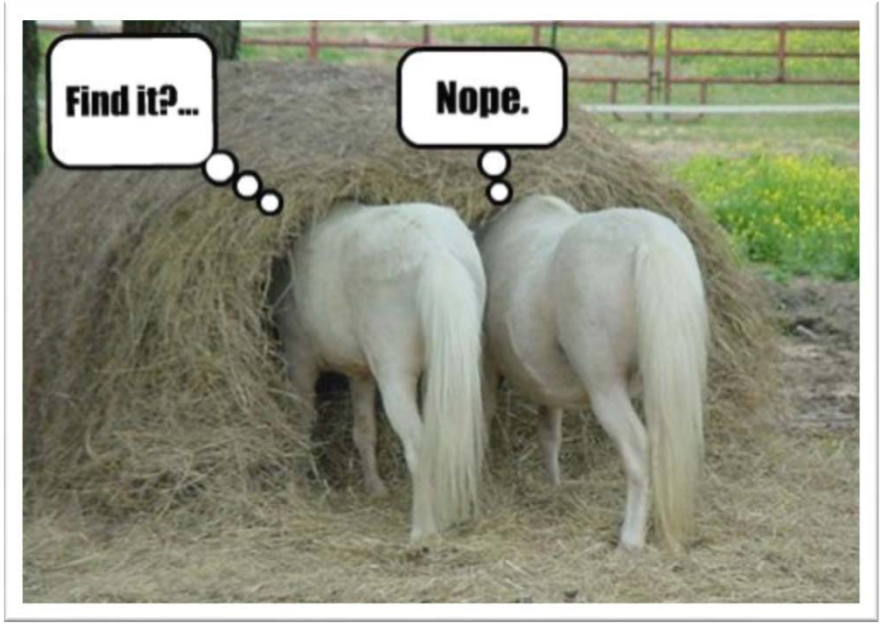
Disease-causing variant can be identified using the genomics analysis platform

THE DIAGNOSTIC AND GENE DISCOVERY CHALLENGE

Interpretation of DNA variants: how do I find the pathogenic mutation?

Exome sequencing →

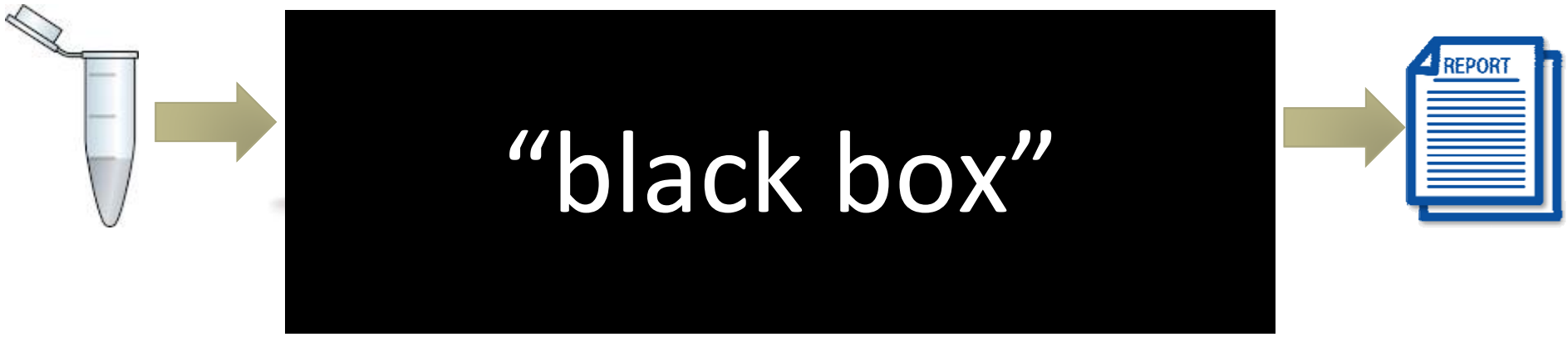
25,000- 50,000 variants ←→ 1 pathogenic mutation



INTERPRETATION IS STILL DIFFICULT

Molecular diagnostics in NGS era

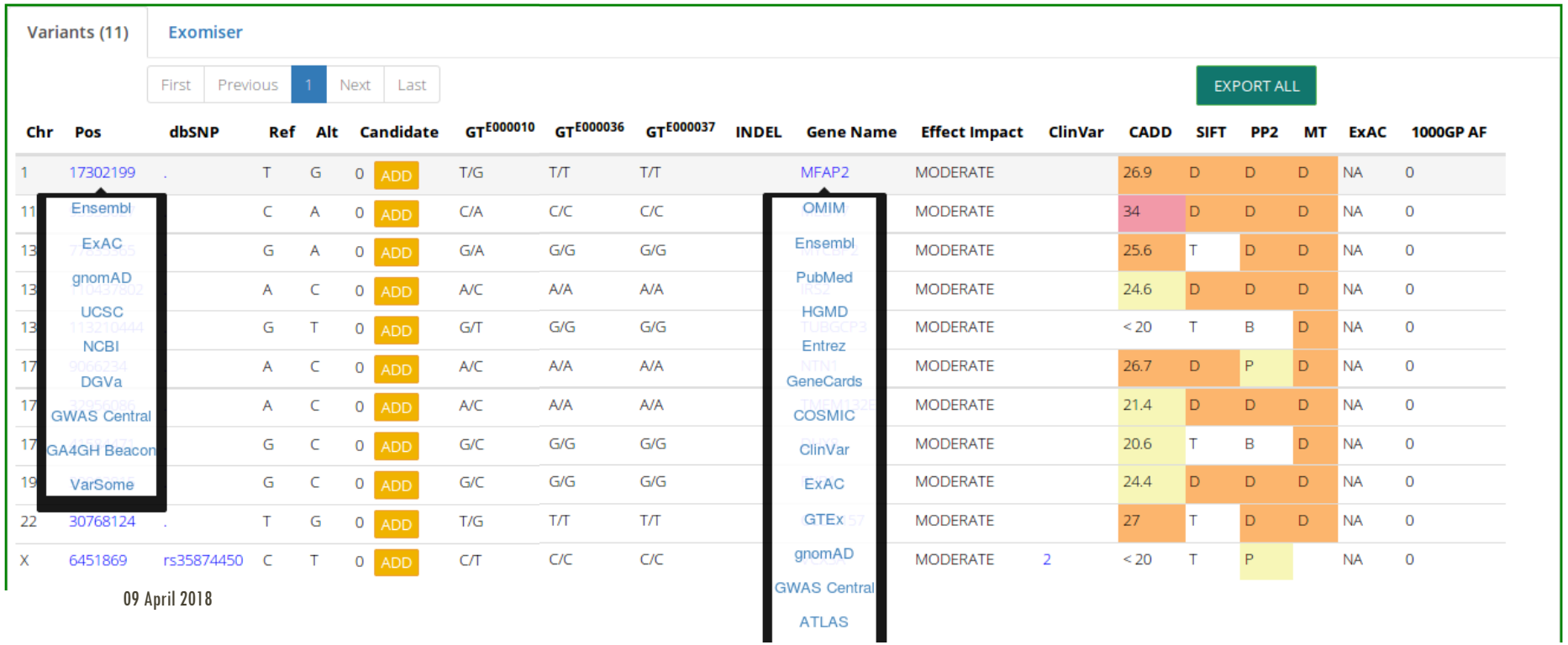
Sample in → Diagnosis out?



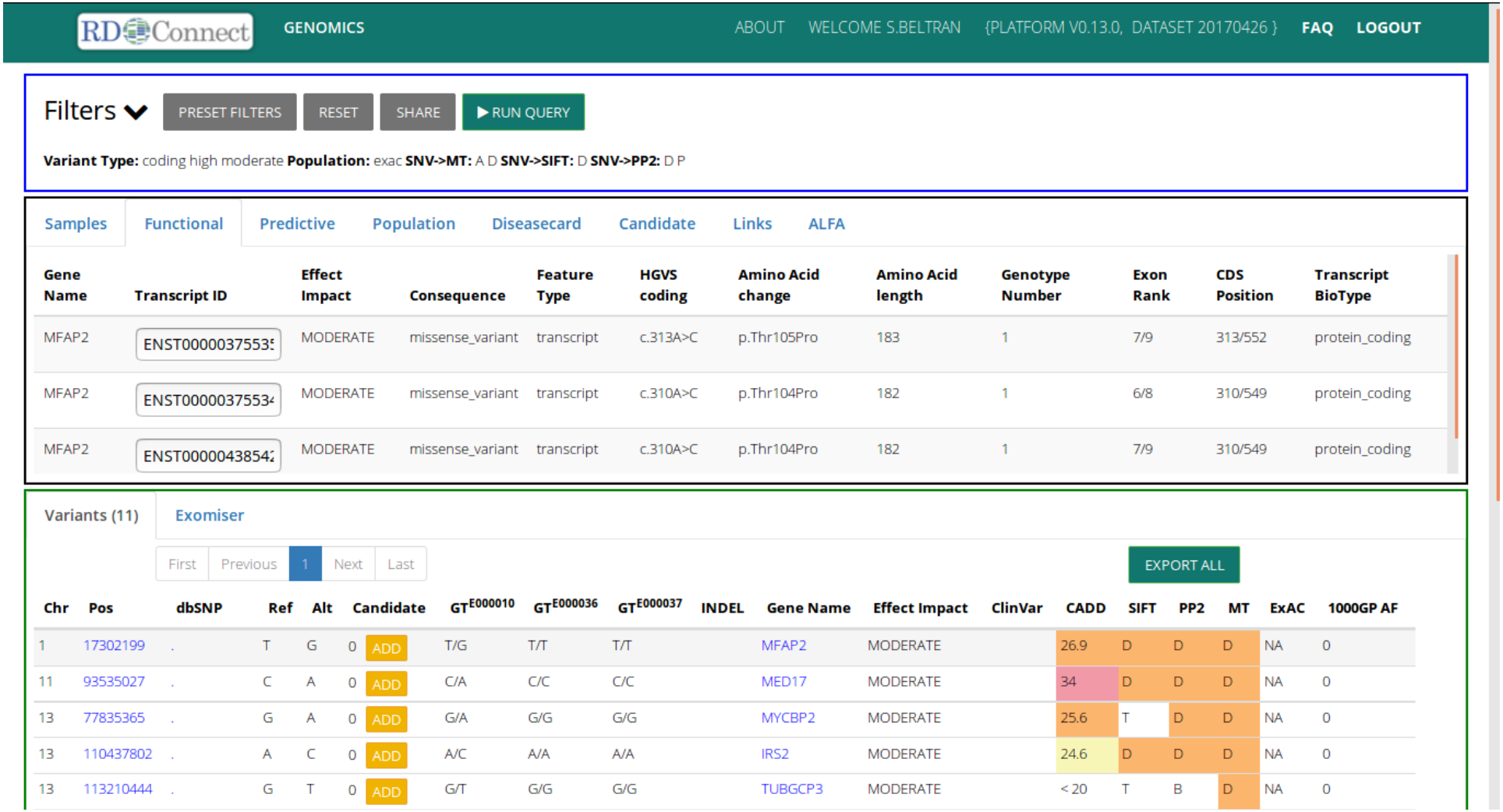
RD-CONNECT GENOME-PHENOME ANALYSIS PLATFORM

Analyse your own undiagnosed patients

Search for other patients with a related genotype or phenotype



ANALYSE INDIVIDUALS/FAMILIES FOR CAUSATIVE VARIANTS



RD Connect GENOMICS ABOUT WELCOME S.BELTRAN {PLATFORM V0.13.0, DATASET 20170426 } FAQ LOGOUT

Filters PRESET FILTERS RESET SHARE ▶ RUN QUERY

Variant Type: coding high moderate **Population:** exac **SNV->MT:** A D **SNV->SIFT:** D **SNV->PP2:** D P

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

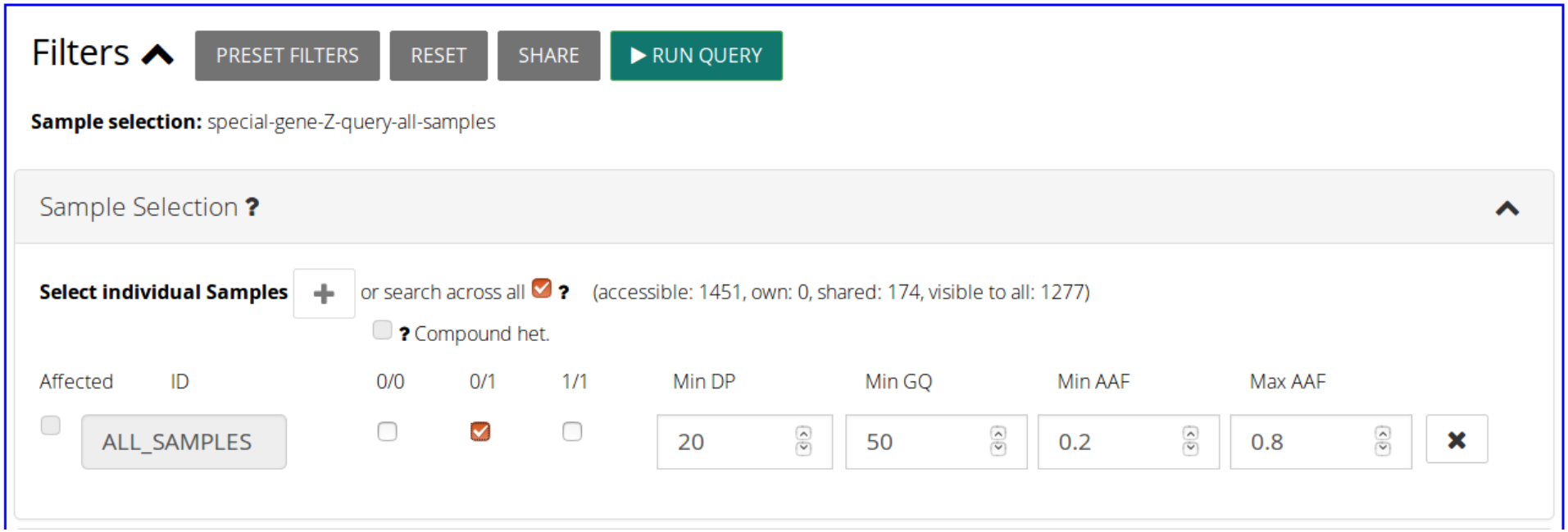
Gene Name	Transcript ID	Effect Impact	Consequence	Feature Type	HGVS coding	Amino Acid change	Amino Acid length	Genotype Number	Exon Rank	CDS Position	Transcript BioType
MFAP2	ENST0000037553	MODERATE	missense_variant	transcript	c.313A>C	p.Thr105Pro	183	1	7/9	313/552	protein_coding
MFAP2	ENST0000037553	MODERATE	missense_variant	transcript	c.310A>C	p.Thr104Pro	182	1	6/8	310/549	protein_coding
MFAP2	ENST0000043854	MODERATE	missense_variant	transcript	c.310A>C	p.Thr104Pro	182	1	7/9	310/549	protein_coding

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	93535027	.	C	A	0	ADD	C/A	C/C	C/C		MED17	MODERATE		34	D	D	D	NA	0
13	77835365	.	G	A	0	ADD	G/A	G/G	G/G		MYCBP2	MODERATE		25.6	T	D	D	NA	0
13	110437802	.	A	C	0	ADD	A/C	A/A	A/A		IRS2	MODERATE		24.6	D	D	D	NA	0
13	113210444	.	G	T	0	ADD	G/T	G/G	G/G		TUBGCP3	MODERATE		< 20	T	B	D	NA	0

SEARCH ACROSS ALL SAMPLES FOR VARIANTS IN SPECIFIC GENES



BIOSAMPLE SHARING

(1) Cataloguing and registration of rare disease biobanks

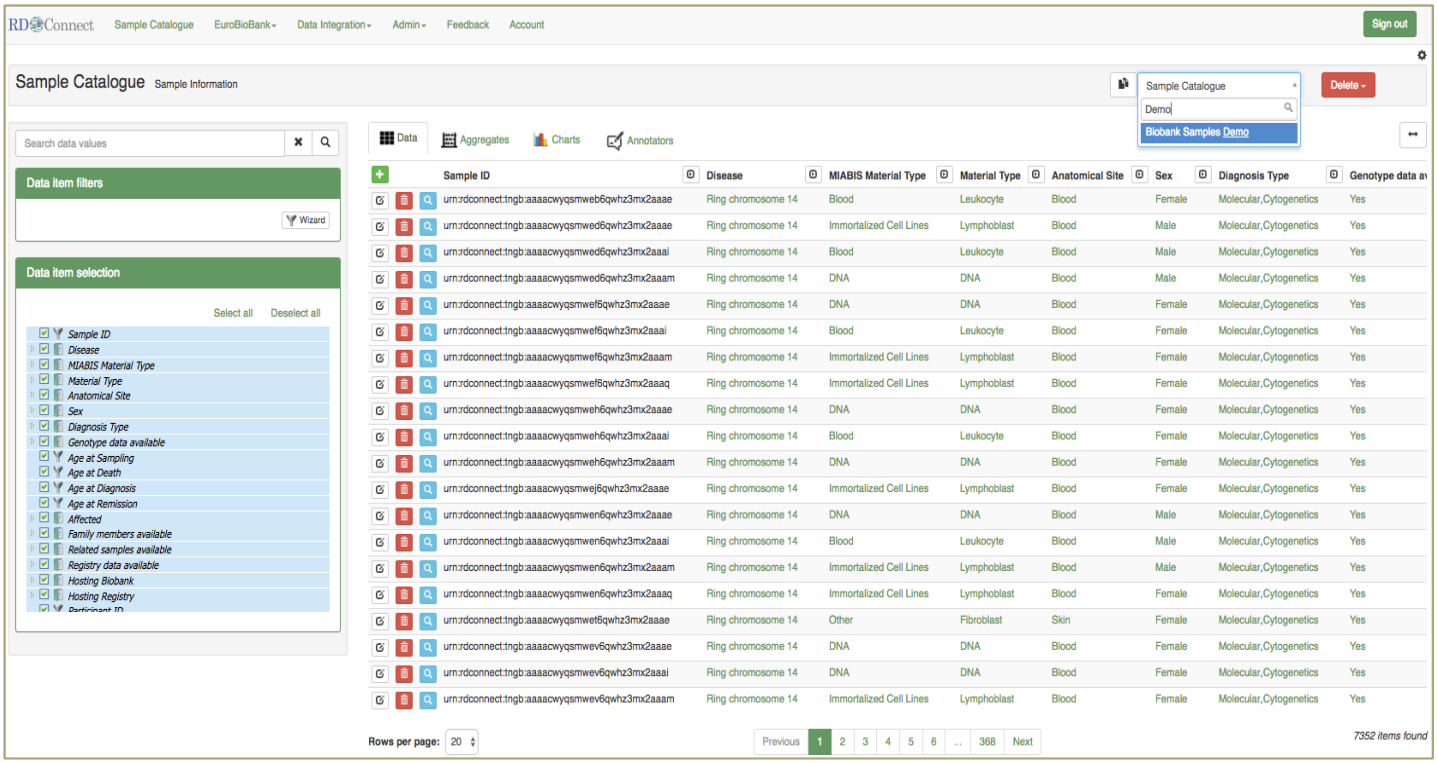
- Biobanks can sign up and give details of their biobank in an "ID card"
- Allows biobanks to participate in RD-Connect infrastructure and research
- Standardised assessment procedure, MTAs etc.

(2) Sharing sample-level data in a common database

- Not just sample numbers but drill-down right to individual samples
  - Researchers can find the samples they need for their research
  - Allows data from omics experiments to be traced back to the sample it came from for further research

SAMPLE-LEVEL CATALOGUE

The RD-Connect Sample Catalogue contains information on available samples across participating biobanks



SCIENTISTS ARE INVITED TO….

Deposit WES/WGS/panel raw data from sequencing projects for integration into the RD-Connect platform

Request an account to look at genomic data in the platform

Use the RD-Connect platform for gene discovery

Let our developers know if you want the functionality of the platform improved/adapted for your research

Add a registry/biobank to the RD-Connect catalogue (ID card)

Participate in multi-omics user groups, co-develop the functionality that you are interested in (use cases)

Make sure future projects are fit for sharing (consent)

Use the RD-Connect impact and dissemination channels

CLINICIANS ARE INVITED TO…

Use appropriate consent for NGS and data sharing

Use standardized ontologies to describe a patient's phenotype

Collect samples from patients in a standardized way and deposit them in appropriate biobanks

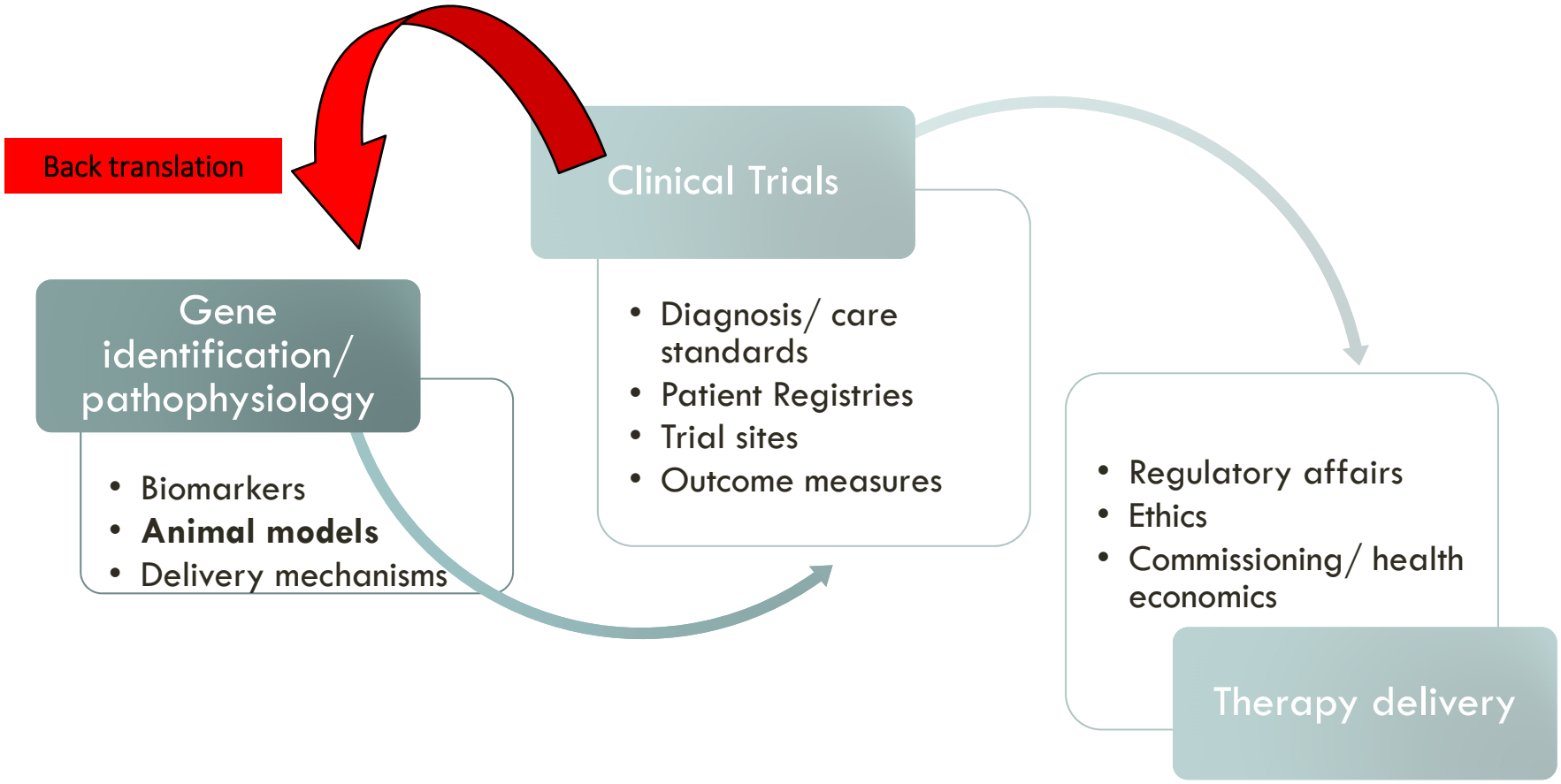
Get involved with the interpretation of genome/exome/panel results (genome rounds)

Feed back results to your patients and point them towards lay-friendly information, encourage their participation

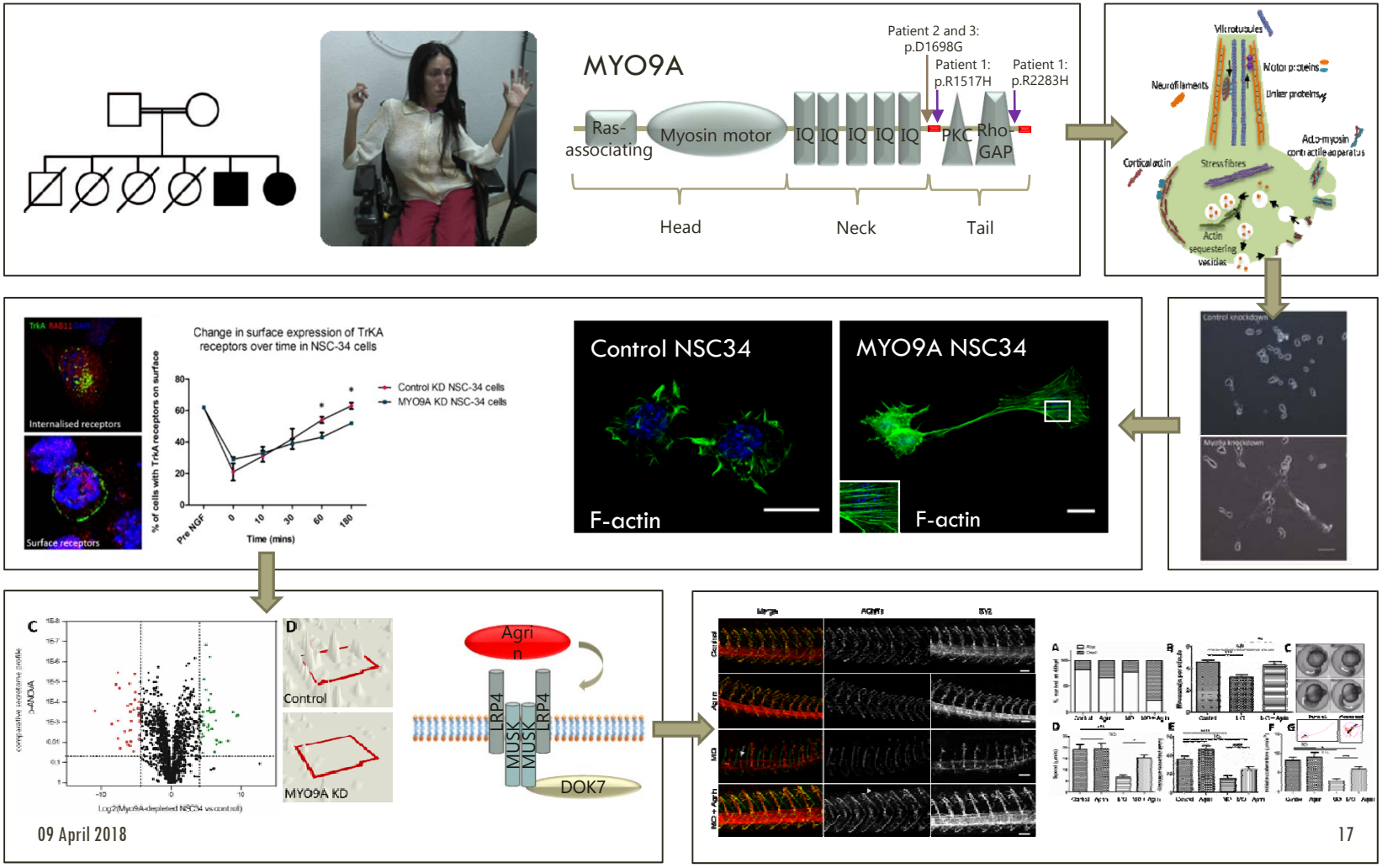
Include patients in other research (registries, cohorts, natural history, clinical trials, etc)

Use the RD-Connect impact and dissemination channels

ADDRESSING THE TRANSLATIONAL PATHWAY



FROM GENE TO POTENTIAL THERAPY



DEFINITION OF TREATABLE UNITS

Science. 2014 Sep 19;345(6203):1505-8. doi: 10.1126/science.1250744.

Neuromuscular disease. DOK7 gene therapy benefits mouse models of diseases characterized by defects in the neuromuscular junction.

Arimura S[1], Okada T[2], Tezuka T[1], Chiyo T[2], Kasahara Y[2], Yoshimura T[3], Motomura M[4], Yoshida N[5], Beeson D[6], Takeda S[2], Yamanashi Y[7].

FASEB J. 2016 Jun;30(6):2382-99. doi: 10.1096/fj.201500162. Epub 2016 Mar 18.

Neuromuscular junction immaturity and muscle atrophy are hallmarks of the CoIQ-

ACKNOWLEDGEMENTS/COLLABORATIONS

CNAG team in Barcelona led by Sergi Beltran

RD-Connect team (led by Hanns Lochmüller) and all RD-Connect partners