

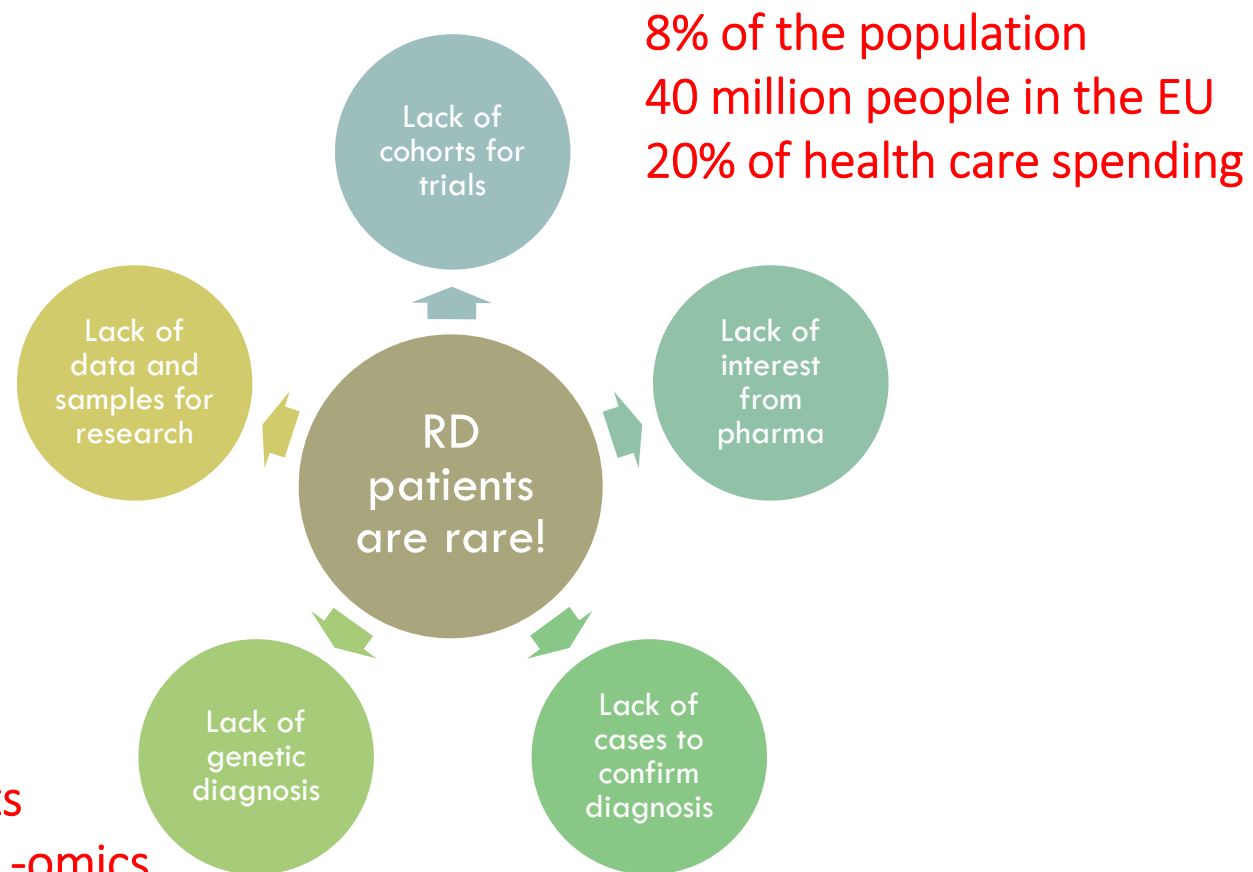
# RD-CONNECT: DATA SHARING AND ANALYSIS FOR RARE DISEASE RESEARCH

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



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

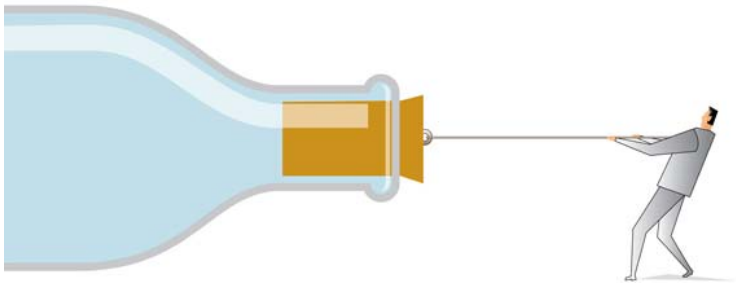
# MANY RARE DISEASE BOTTLENECKS ARE CROSS-CUTTING

...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- Reluctance to share unpublished data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

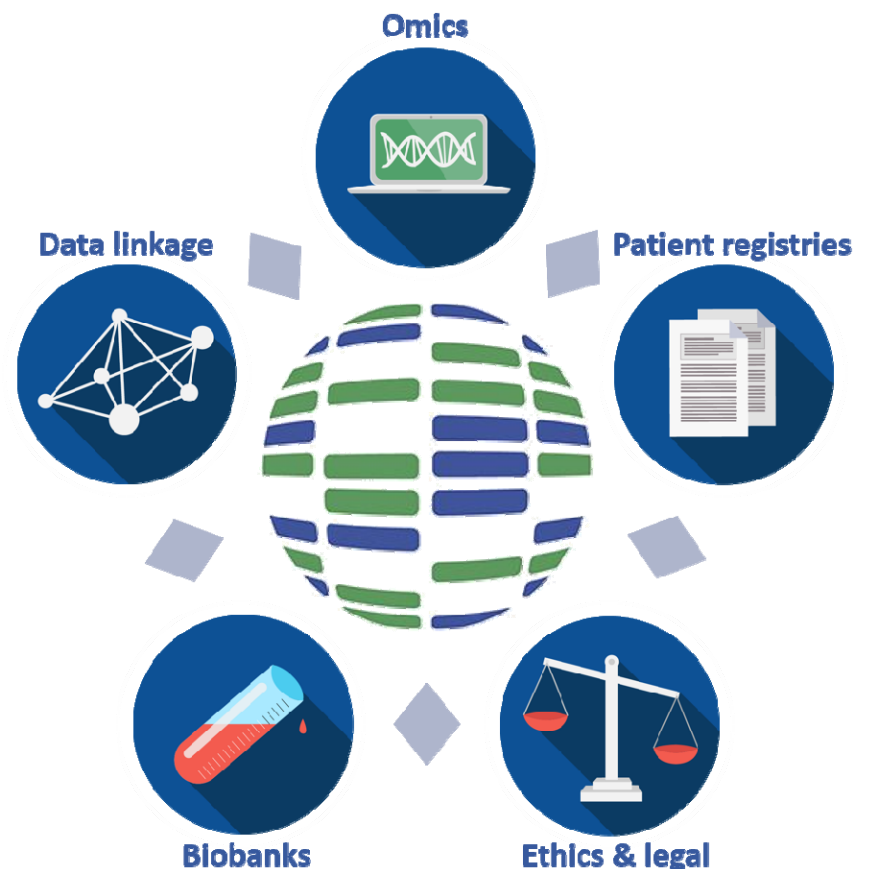
**RD-Connect has created resources for use by the rare disease research community**

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)





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

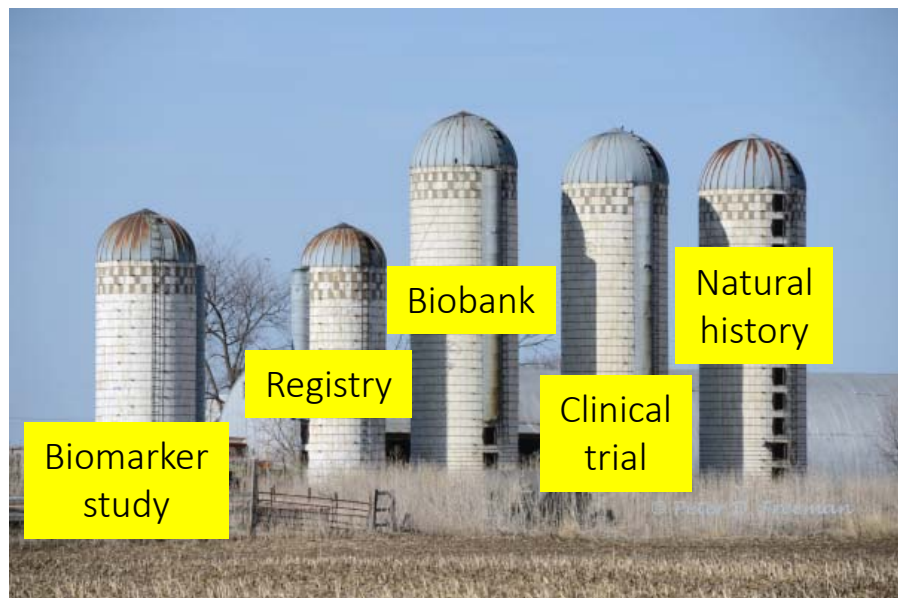
Standardized phenotypic data collection

Searchable catalogue of biosamples

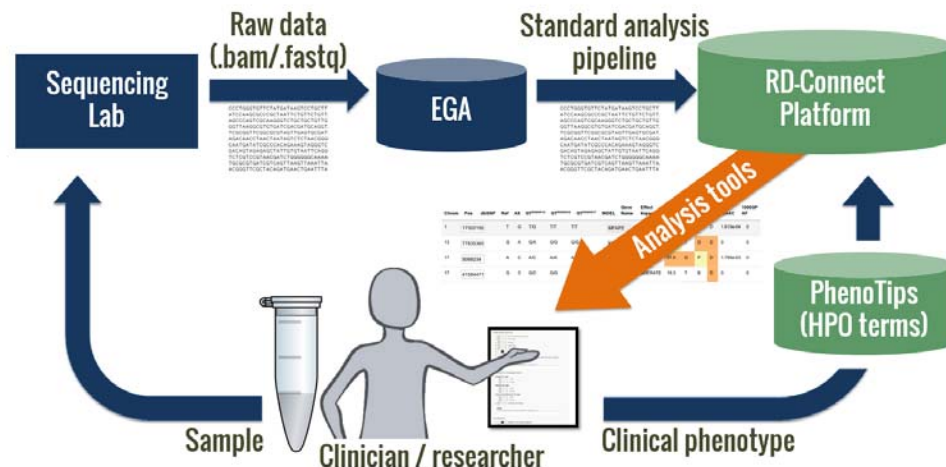
Data linkage across resources

## 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

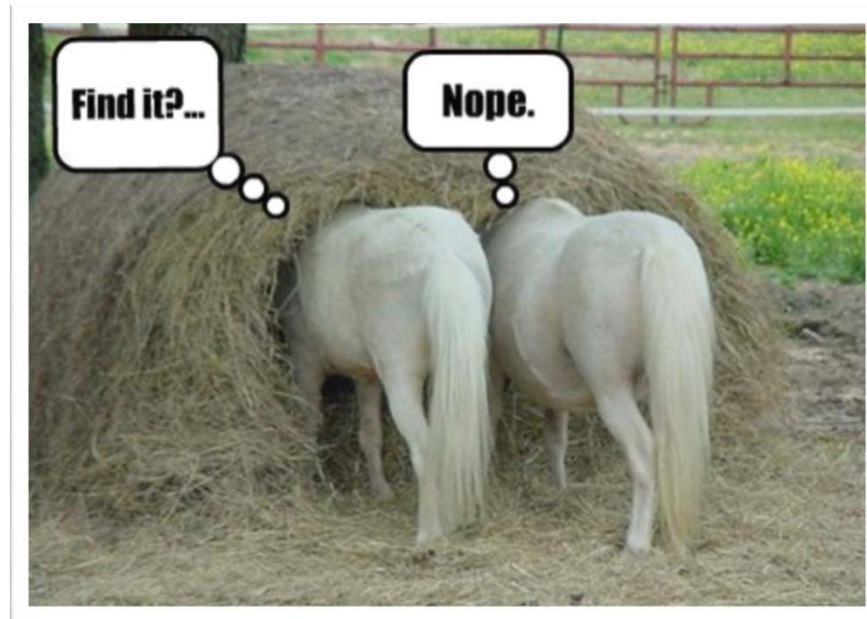
Disease-causing variant can be identified using the genomics analysis platform

# THE DIAGNOSTIC AND GENE DISCOVERY CHALLENGE

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

Exome sequencing →

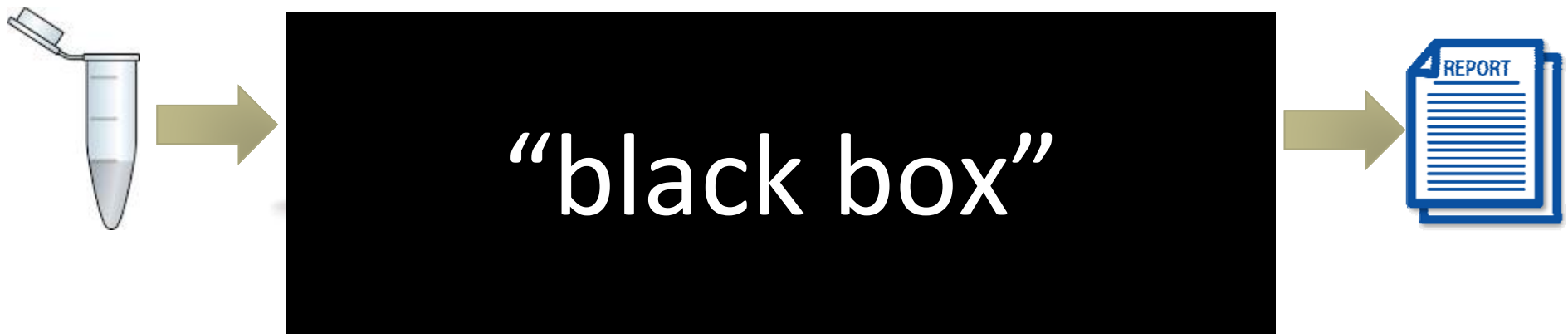
25,000- 50,000 variants  $\longleftrightarrow$  1 pathogenic mutation



# 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

Variants (11)																		Exomiser	
<div>FirstPrevious1NextLast</div>																		EXPORT ALL	
Chr	Pos	dbSNP	Ref	Alt	Candidate	GT <sup>E000010</sup>	GT <sup>E000036</sup>	GT <sup>E000037</sup>	INDEL	Gene Name	Effect Impact	ClinVar	CADD	SIFT	PP2	MT	ExAC	1000GP AF	
1	17302199	.	T	G	0ADD	T/G	T/T	T/T		MFAP2	MODERATE		26.9	D	D	D	NA	0	
11	Ensembl		C	A	0ADD	C/A	C/C	C/C		OMIM	MODERATE		34	D	D	D	NA	0	
13	ExAC		G	A	0ADD	G/A	G/G	G/G		Ensembl	MODERATE		25.6	T		D	NA	0	
13	gnomAD		A	C	0ADD	A/C	A/A	A/A		PubMed	MODERATE		24.6	D		D	NA	0	
13	UCSC		G	T	0ADD	G/T	G/G	G/G		HGMD	MODERATE		< 20	T	B	D	NA	0	
17	NCBI		A	C	0ADD	A/C	A/A	A/A		TUBGCP3	MODERATE		26.7	D	P	D	NA	0	
17	DGVa		A	C	0ADD	A/C	A/A	A/A		NTN1	MODERATE		21.4	D		D	NA	0	
17	GWAS Central		G	C	0ADD	G/C	G/G	G/G		GeneCards	MODERATE		20.6	T	B	D	NA	0	
17	GA4GH Beacon		G	C	0ADD	G/C	G/G	G/G		COSMIC	MODERATE		24.4	D		D	NA	0	
19	VarSome		G	C	0ADD	G/C	G/G	G/G		ClinVar	MODERATE		27	T		D	NA	0	
22	30768124	.	T	G	0ADD	T/G	T/T	T/T		ExAC	MODERATE		< 20	T	P		NA	0	
X	6451869	rs35874450	C	T	0ADD	C/T	C/C	C/C		gnomAD	MODERATE	2	< 20	T			NA	0	

09 April 2018

09 April 2018

# ANALYSE INDIVIDUALS/FAMILIES FOR CAUSATIVE VARIANTS



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 <sup>E000010</sup>	GT <sup>E000036</sup>	GT <sup>E000037</sup>	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

Filters 

PRESET FILTERS


RESET

SHARE

 RUN QUERY

**Sample selection:** special-gene-Z-query-all-samples

Sample Selection ? 

**Select individual Samples**  or search across all ☒ ? (accessible: 1451, own: 0, shared: 174, visible to all: 1277)

☐ ? Compound het.

Affected

ID

0/0

0/1

1/1

Min DP

Min GQ

Min AAF

Max AAF

☐

ALL\_SAMPLES

☐☒☐

20

50

0.2

0.8



# BIO SAMPLE 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

The screenshot displays the RD-Connect Sample Catalogue interface. The top navigation bar includes links for Sample Catalogue, EuroBioBank, Data Integration, Admin, Feedback, and Account, along with a Sign out button. The main header shows 'Sample Catalogue' and 'Sample Information' tabs, with a search bar and a 'Delete' button. A dropdown menu is open, showing 'Sample Catalogue', 'Demo', and 'Biobank Samples Demo'. The left sidebar contains 'Data item filters' and 'Data item selection' sections. The main table lists samples with columns for Sample ID, Disease, MIABIS Material Type, Material Type, Anatomical Site, Sex, Diagnosis Type, and Genotype data. The table shows 20 rows of data, with a 'Rows per page' dropdown set to 20. The bottom right corner indicates '7352 items found'.

Sample ID	Disease	MIABIS Material Type	Material Type	Anatomical Site	Sex	Diagnosis Type	Genotype data
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	Blood	Leukocyte	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	Blood	Leukocyte	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aaam	Ring chromosome 14	DNA	DNA	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	Blood	Leukocyte	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aaam	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aaaq	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aaam	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	DNA	DNA	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	Blood	Leukocyte	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aaam	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aaaq	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	Other	Fibroblast	Skin	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aae	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwysmweb6qwhz3mx2aaam	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes

# 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

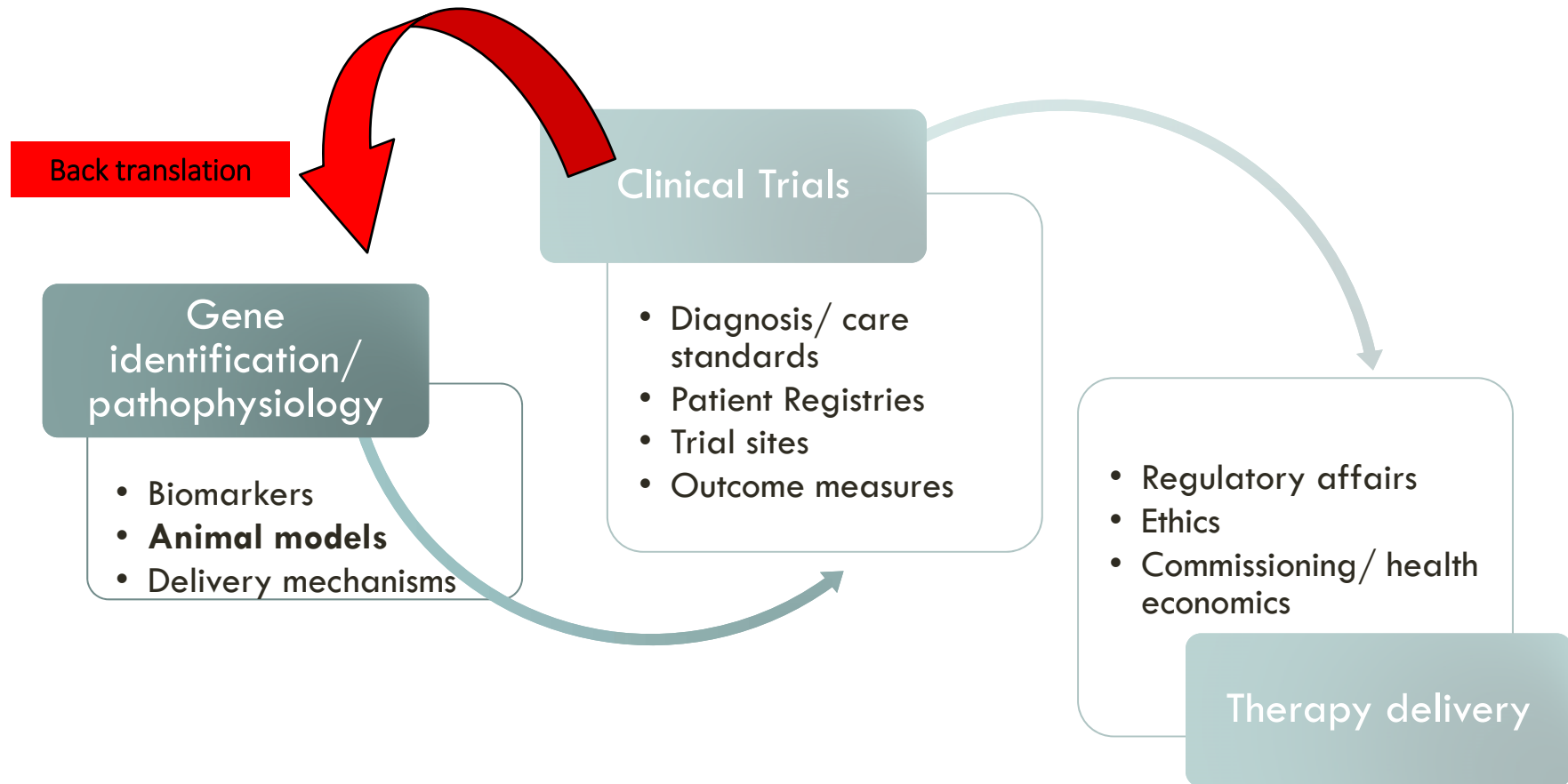
Get involved with the interpretation of genome/exome/panel results (genome rounds)

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)

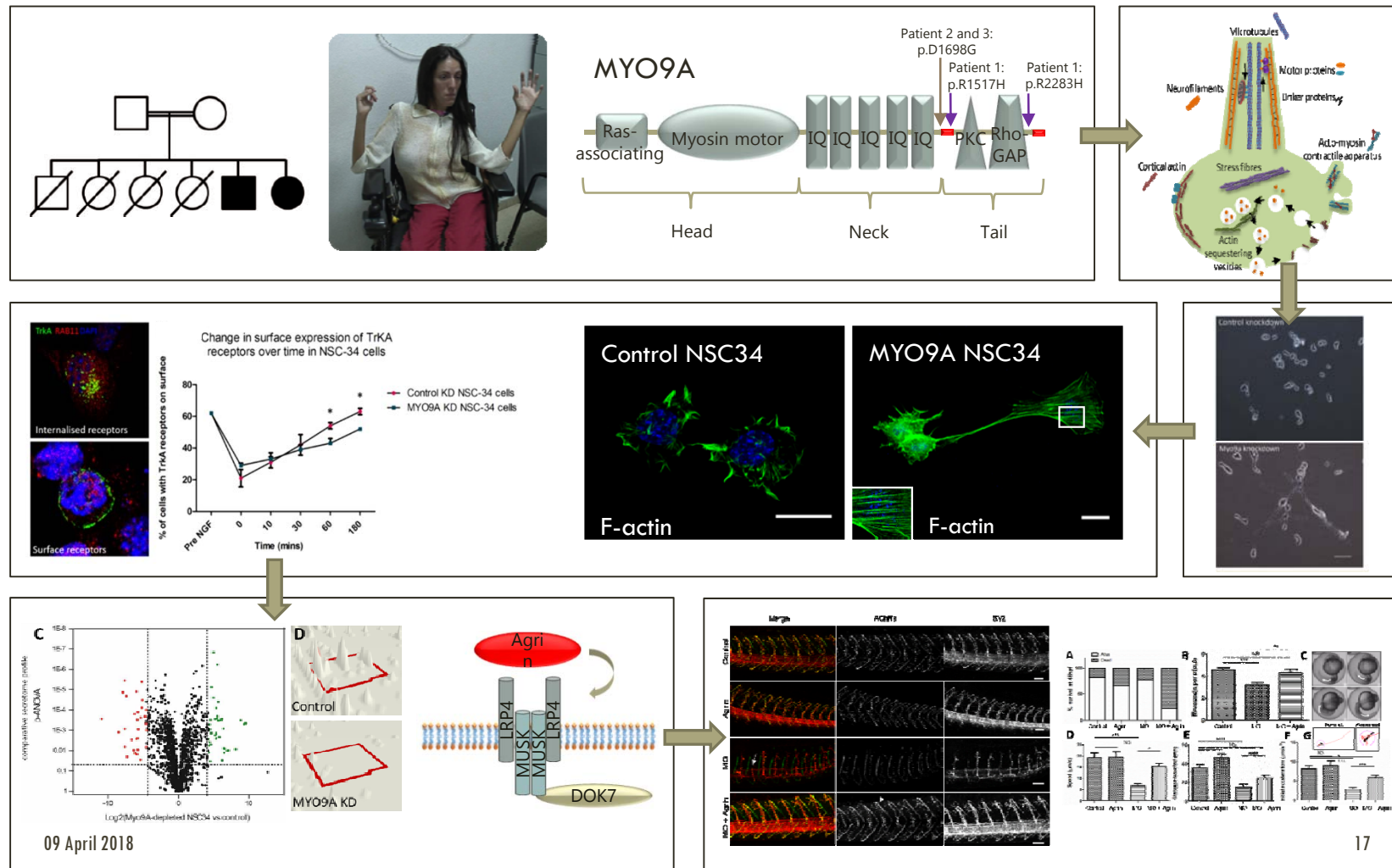
Use the RD-Connect impact and dissemination channels

# 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<sup>1</sup>, Okada T<sup>2</sup>, Tezuka T<sup>1</sup>, Chiyo T<sup>2</sup>, Kasahara Y<sup>2</sup>, Yoshimura T<sup>3</sup>, Motomura M<sup>4</sup>, Yoshida N<sup>5</sup>, Beeson D<sup>6</sup>, Takeda S<sup>2</sup>, Yamanashi Y<sup>7</sup>.

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 ColQ-**

Gene: Dok7



Name	docking protein 7	<a href="#">Login to register interest</a>
Synonyms	<ul style="list-style-type: none"><li>Dok-7</li><li>EF-12</li><li>Oit5</li><li>A930013K19Rik</li></ul>	<a href="#">Order</a>
Status	<div>ES Cells</div> <div>Mice tm1</div> <div>Mice tm1.1</div> <div>phenotype data available</div>	
Links	<a href="#">MGI:3584043</a> <a href="#">Ensembl Gene</a> <a href="#">Ensembl Orthologs</a> <a href="#">ENU(12)</a>	

# ACKNOWLEDGEMENTS/COLLABORATIONS



CNAG team in Barcelona led  
by Sergi Beltran



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

