RD Connect

WHAT ARE OTHER EUROPEAN NETWORKS OFFERING? WHAT IS THE BENEFIT OF SHARING DATA AND SAMPLES THROUGH EXISTING STRUCTURES E.G. RD-CONNECT?

IRDiRC INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

SEVENTH FRAMEWORK PROGRAMME

Rachel Thompson • Newcastle University

Q: What are other European networks offering?

A: quite a lot!

RD Connect

elixir

European Reference Networks

TREAT-NMD
Neuromuscular Network

ECRIN
EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK

BBMRI-ERIC
Biobanking and BioMolecular resources Research Infrastructure

CORBEL
SHARED SERVICES FOR LIFE-SCIENCE

MYO-MRI
MRI & MRS IN NEUROMUSCULAR DISEASE

EPIRARE
European Platform for Rare Disease Registries

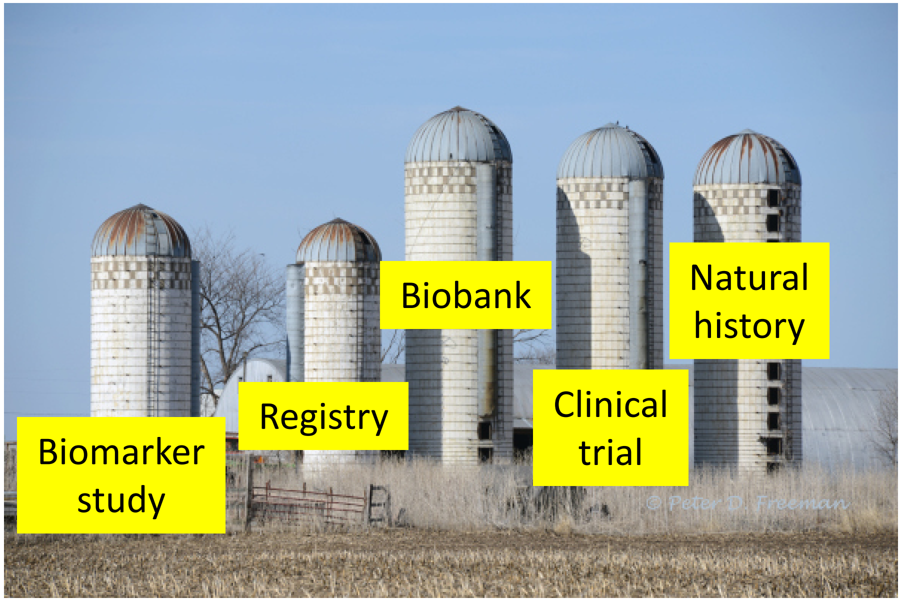
EuroBioBank

Global Alliance
for Genomics & Health

The guiding principle

Data sharing for research and better data analysis

- □ Gene and modifier discovery
- □ Samples for further research
- □ Genotype-phenotype correlation
- □ Patient recruitment
- □ Global natural history comparisons
- □ Biomarkers, therapeutic targets…



Overcoming silos!

RD-Connect: Infrastructure for RD data sharing

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

Overarching objectives:

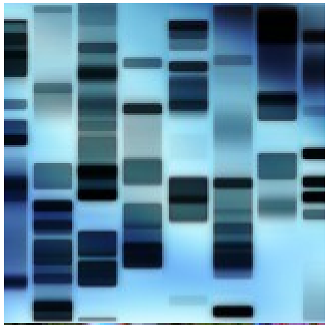
- □ Contribution to the IRDiRC objectives of delivering 200 new therapies for rare diseases and means to diagnose most rare diseases by the year 2020
- □ Development of an integrated, quality-assured and comprehensive platform in which complete clinical profiles are combined with -omics data and sample availability for rare disease research, in particular IRDiRC-funded research.

RD-Connect's main aims

- Creation of central system and repository for **reprocessing**, **storing** and **analysing** omics data
  - Raw data hosted at European Genome-phenome Archive (EGA)
  - Raw data reprocessed through standard analysis pipeline for consistency
  - Reprocessed data accessible via Barcelona platform with user-friendly online analysis interface
- Integration of phenotypic data
- Integration of biosample data
- Development of new bioinformatic tools
- Ethical and legal considerations for data sharing
- Patient input
- Outreach and impact: interaction with rare disease community

Sharing: What?

- □ Raw data from all types of studies
- □ Genomic data
- □ Phenotypic data
- □ Natural history data
- □ Clinical trial data
- □ Biosamples (blood, DNA, tissue samples, cell lines…)
- □ Linked data and samples
- □ Access to patients
- □ …



Sharing: Barriers

- **General**
  - Privacy protection issues: “do I have the patient’s permission?”
  - Lack of infrastructure: “I want to share data but where do I put it?”
  - Lack of standards and interoperability
- **Academia**
  - Culture of protecting research results: “someone else might scoop my publication!”
  - Lack of incentives for sharing
- **Industry**
  - IP issues/competition (when pharma is asked to share its own data)
  - Concerns over data quality, regulatory compliance (when pharma wants to reuse data from academia)

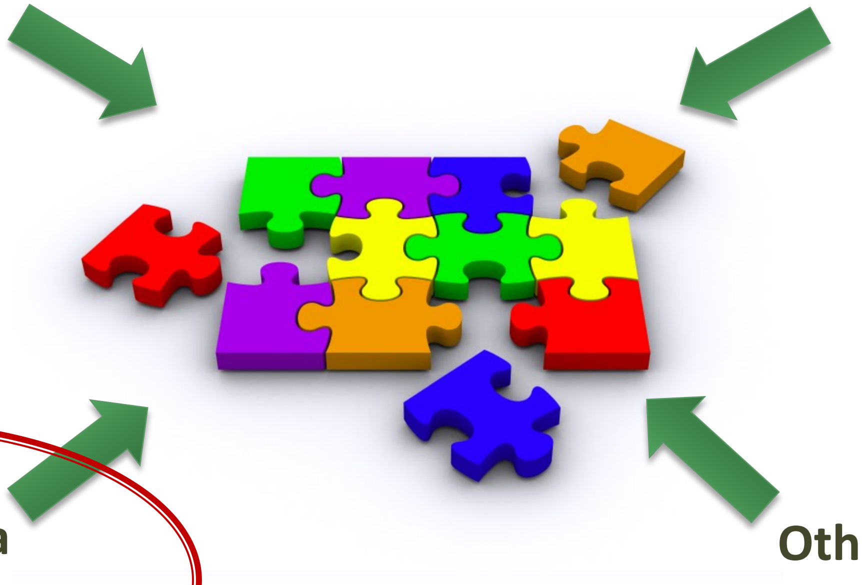
Sharing: Benefits

- 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

Data integration in RD-Connect

Sample data
(biobank databases)

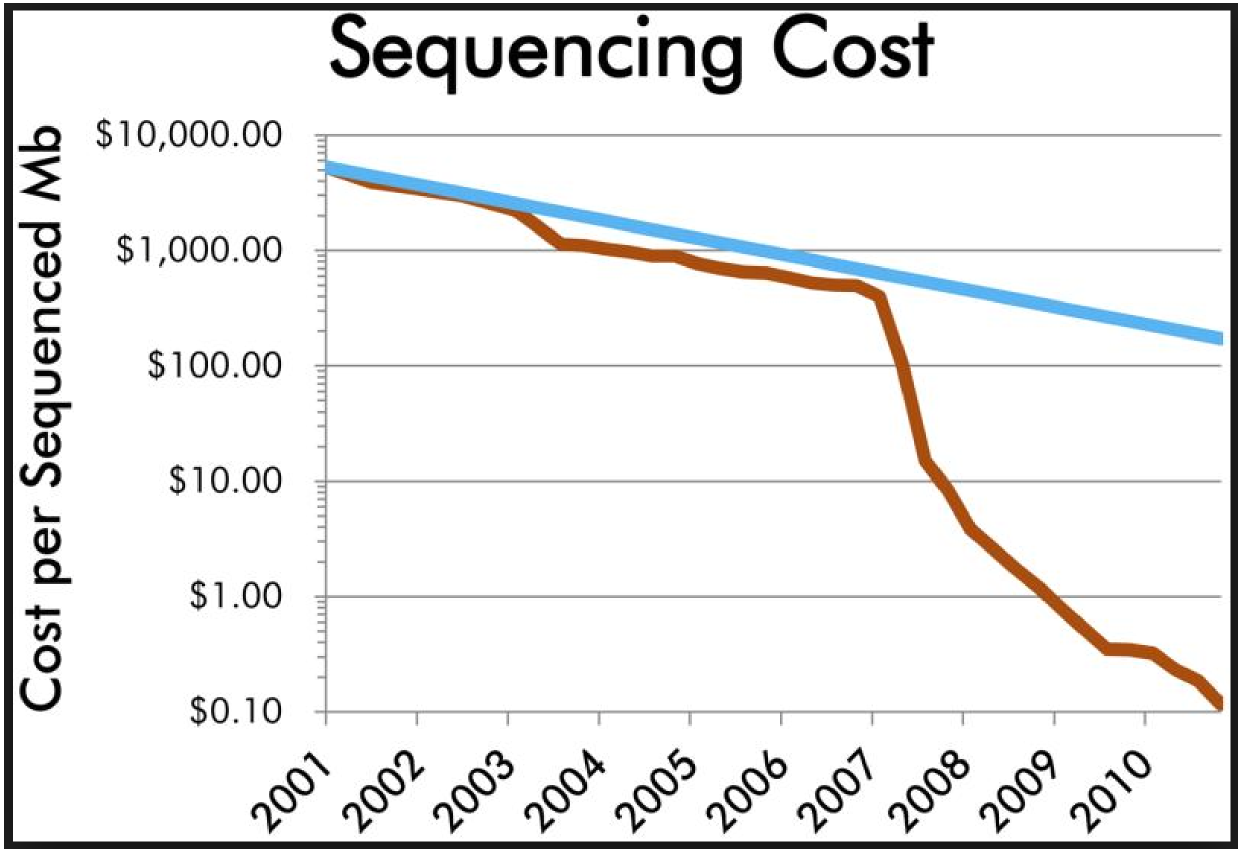
Clinical data
(registries, and phenotypic databases)



Genomic data
(WES, WGS)

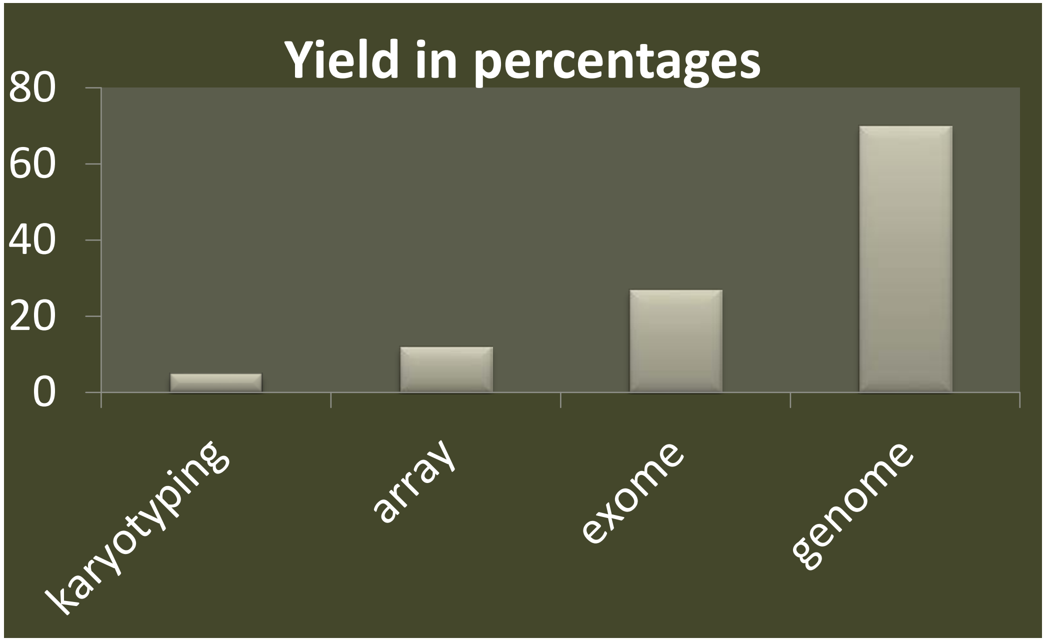
Other omics data
(transcriptomics, metabolomics, proteomics …)

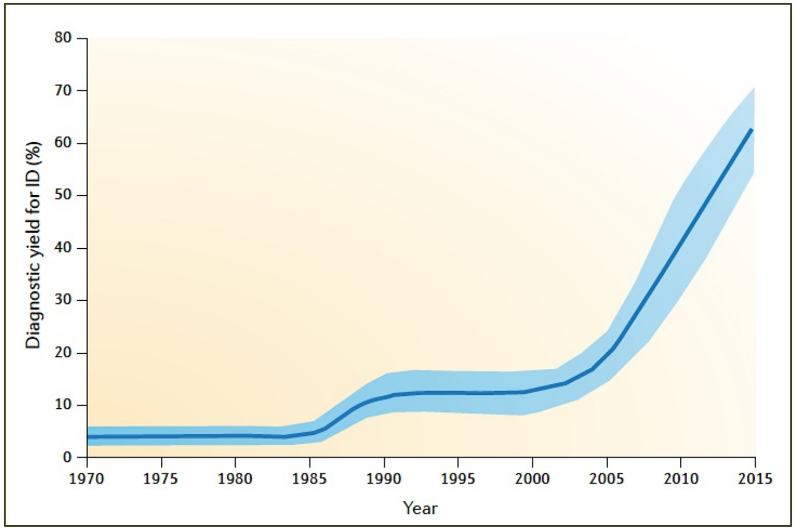
NGS is becoming affordable



Number of new genes discovered is increasing

- Example: intellectual disability



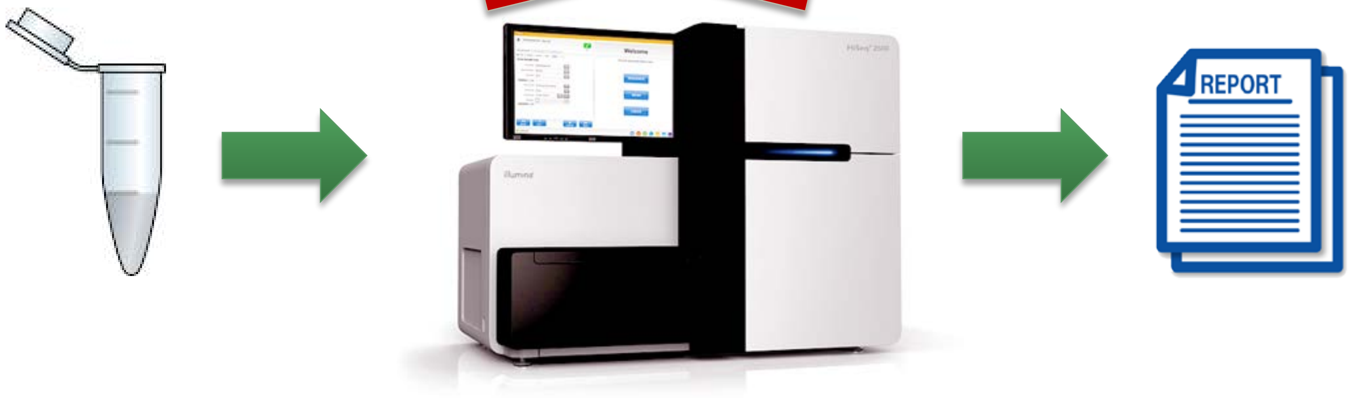


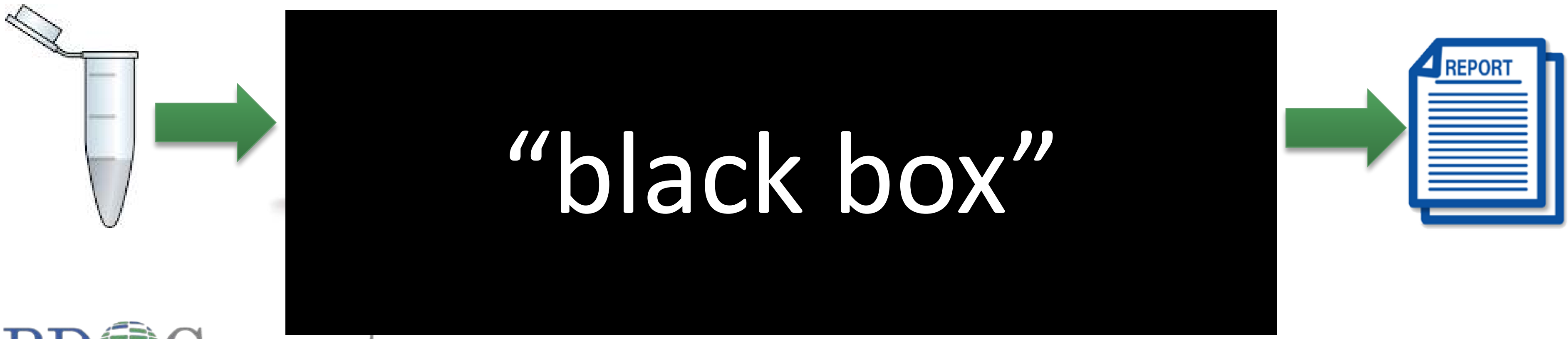
Vissers et al., Nature Rev Genet 2016

But: interpretation is still difficult

Molecular diagnostics in NGS era

Sample ~~in → Di~~agnosis out?





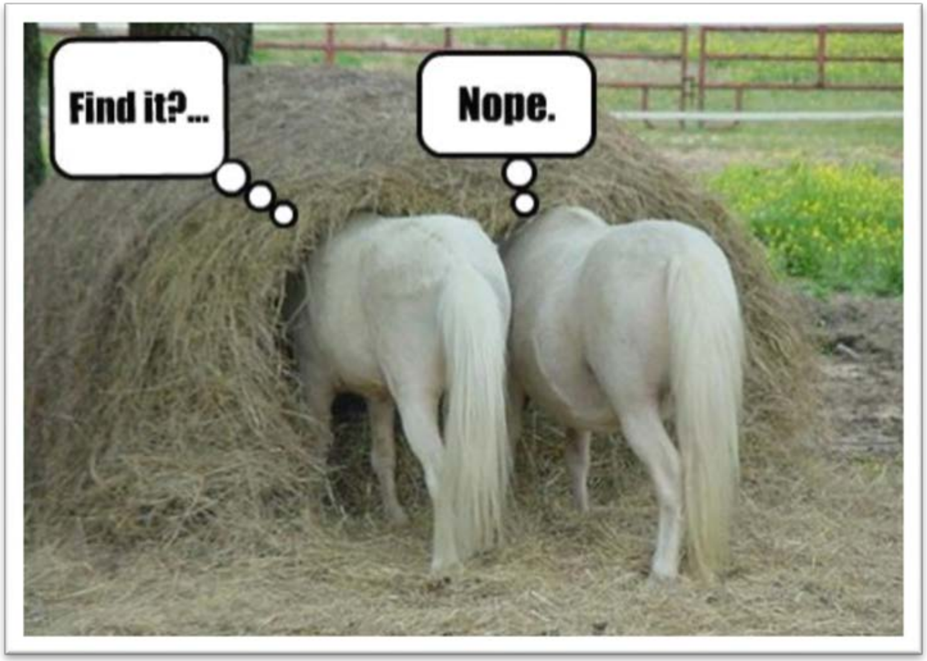
“black box”

The challenge

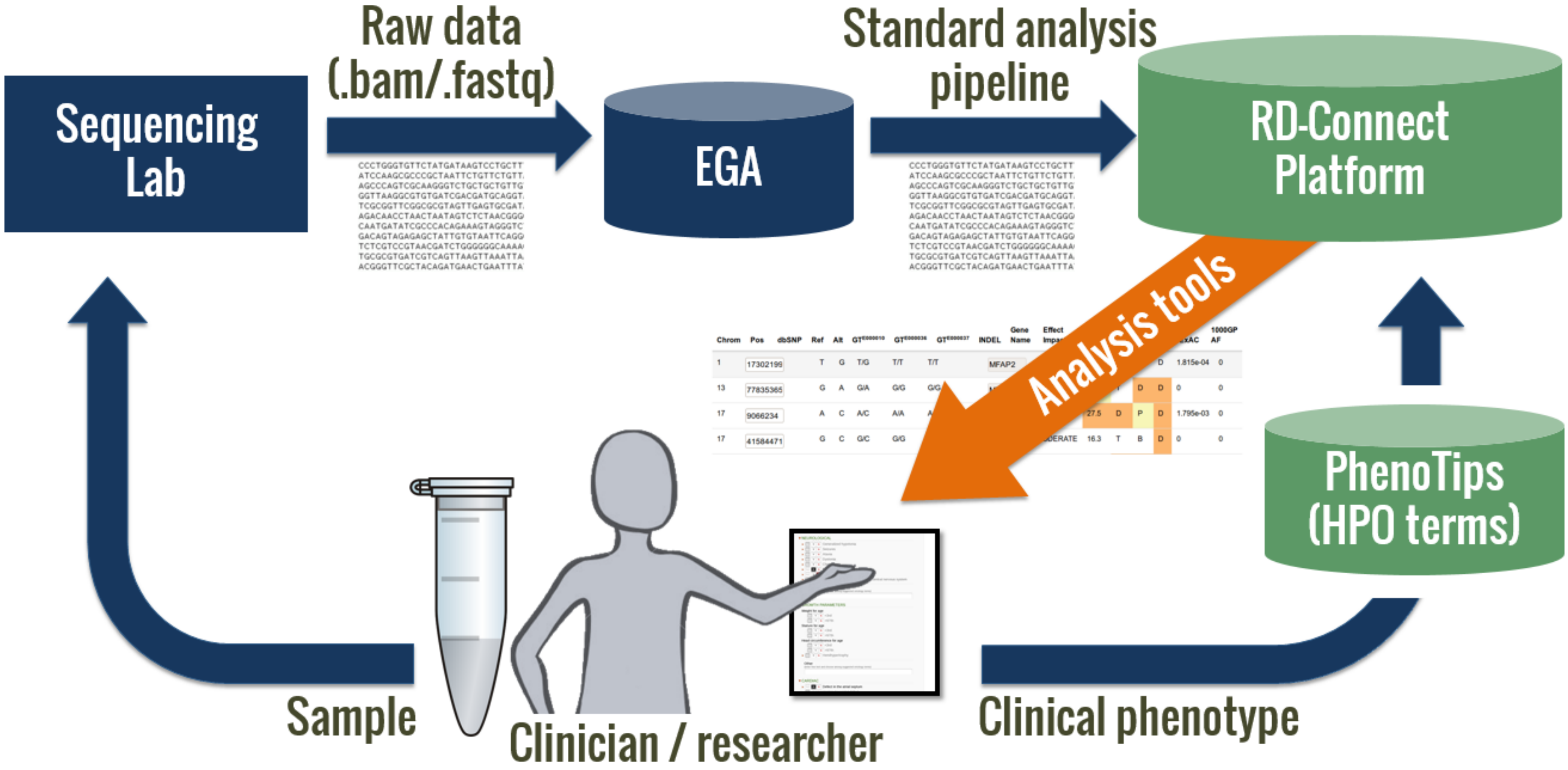
Interpretation of DNA variants: how do I find the pathogenic mutation?

Exome sequencing →

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

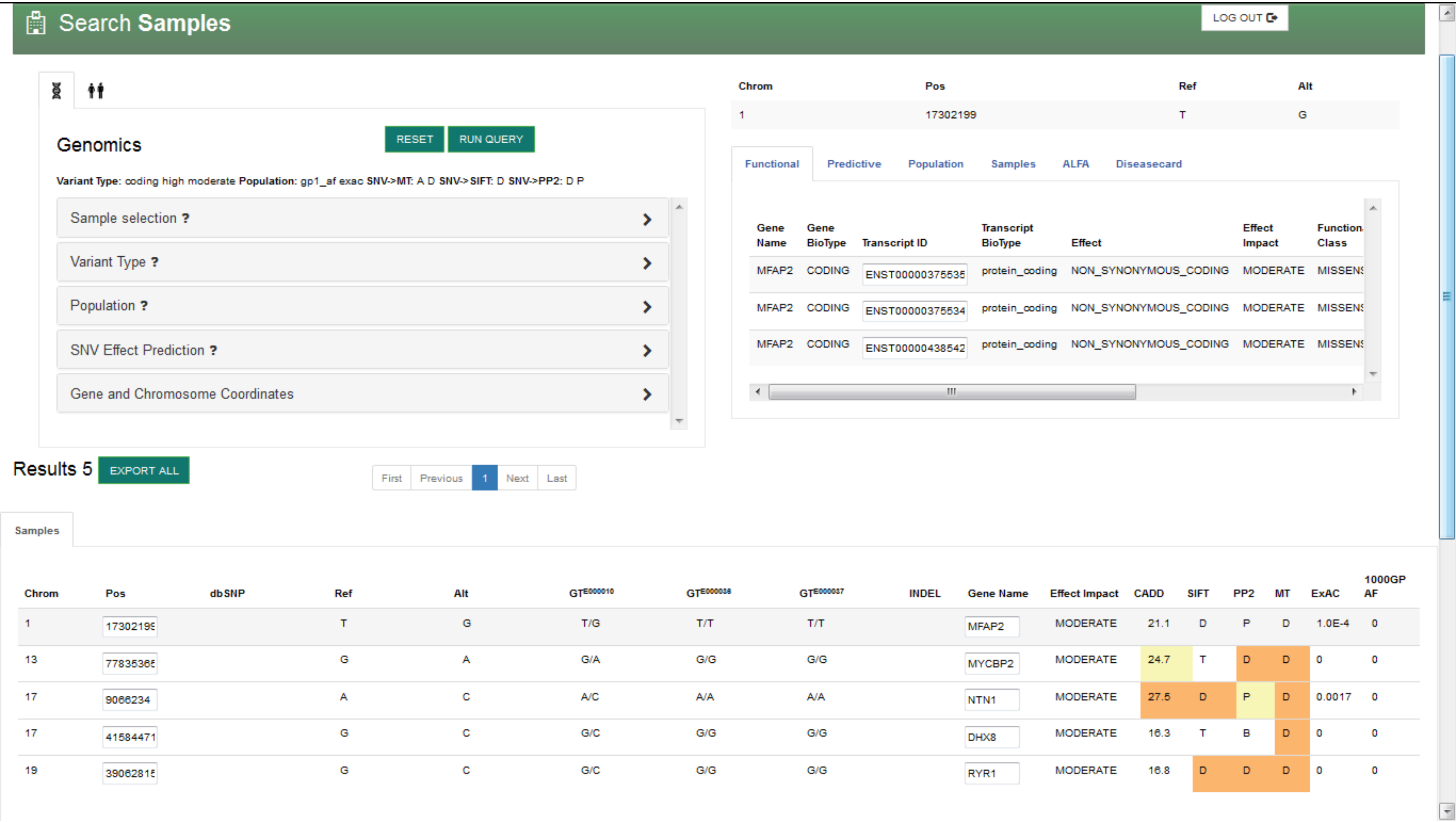


Genomic data flow in RD-Connect


Sequencing Lab
Raw data (.bam/.fastq)
CCCTGGGTGTTCTATGATAAGTCCTGCTT
ATCCAAGCGCCCGCTAATTCTGTTCTGTT
AGCCCAGTCGCAAGGGTCTGCTGCTGTTG
GGTTAAGGCGTGTGATCGACGATGCAGGT
TCGCGGTTCGGCGCGTAGTTGAGTGCGAT
AGACAACCTAACTAATAGTCTCTAACGGG
CAATGATATCGCCCACAGAAAGTAGGGTC
GACAGTAGAGAGCTATTGTGTAATTCAGG
TCTCGTCCGTAACGATCTGGGGGGCAAAA
TGCGCGTGATCGTCAGTTAAGTTAAATTA
ACGGGTTCGCTACAGATGAACTGAATTTA
EGA
Standard analysis pipeline
CCCTGGGTGTTCTATGATAAGTCCTGCTT
ATCCAAGCGCCCGCTAATTCTGTTCTGTT
AGCCCAGTCGCAAGGGTCTGCTGCTGTTG
GGTTAAGGCGTGTGATCGACGATGCAGGT
TCGCGGTTCGGCGCGTAGTTGAGTGCGAT
AGACAACCTAACTAATAGTCTCTAACGGG
CAATGATATCGCCCACAGAAAGTAGGGTC
GACAGTAGAGAGCTATTGTGTAATTCAGG
TCTCGTCCGTAACGATCTGGGGGGCAAAA
TGCGCGTGATCGTCAGTTAAGTTAAATTA
ACGGGTTCGCTACAGATGAACTGAATTTA
RD-Connect Platform
Analysis tools
PhenoTips (HPO terms)
Sample
Clinician / researcher
Clinical phenotype

Chrom	Pos	dbSNP	Ref	Alt	GT^{E000010}	GT^{E000036}	GT^{E000037}	INDEL	Gene Name	Effect Impa	xAC	1000GP AF
1	17302199		T	G	T/G	T/T	T/T		MFAP2		1.815e-04	0
13	77835365		G	A	G/A	G/G	G/G				0	0
17	9066234		A	C	A/C	A/A	A				1.795e-03	0
17	41584471		G	C	G/C	G/G				DERATE	0	0

RD-Connect genomic analysis platform



Genomics

Variant Type: coding high moderate **Population:** gp1_af exac **SNV->MT:** A D **SNV->SIFT:** D **SNV->PP2:** D P

- Sample selection **?**
- Variant Type **?**
- Population **?**
- SNV Effect Prediction **?**
- Gene and Chromosome Coordinates

Chrom	Pos	Ref	Alt
1	17302199	T	G

Functional | Predictive | Population | Samples | ALFA | Diseasecard

Gene Name	Gene BioType	Transcript ID	Transcript BioType	Effect	Effect Impact	Function Class
MFAP2	CODING	ENST00000375535	protein_coding	NON_SYNONYMOUS_CODING	MODERATE	MISSENS
MFAP2	CODING	ENST00000375534	protein_coding	NON_SYNONYMOUS_CODING	MODERATE	MISSENS
MFAP2	CODING	ENST00000438542	protein_coding	NON_SYNONYMOUS_CODING	MODERATE	MISSENS

Results 5

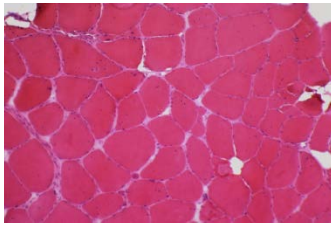
Samples

Chrom	Pos	dbSNP	Ref	Alt	$GT^{E000010}$	$GT^{E000038}$	$GT^{E000037}$	INDEL	Gene Name	Effect Impact	CADD	SIFT	PP2	MT	ExAC	1000GP AF
1	17302199		T	G	T/G	T/T	T/T		MFAP2	MODERATE	21.1	D	P	D	1.0E-4	0
13	77835365		G	A	G/A	G/G	G/G		MYCBP2	MODERATE	24.7	T	D	D	0	0
17	9066234		A	C	A/C	A/A	A/A		NTN1	MODERATE	27.5	D	P	D	0.0017	0
17	41584471		G	C	G/C	G/G	G/G		DHX8	MODERATE	16.3	T	B	D	0	0
19	39062815		G	C	G/C	G/G	G/G		RYR1	MODERATE	16.8	D	D	D	0	0

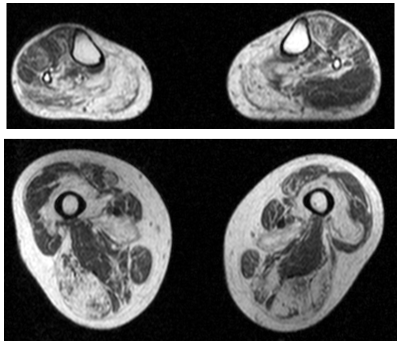
Exome sequencing and data sharing: new congenital myopathy gene

Newcastle case

- □ Childhood onset
- □ Proximal muscle weakness, mainly lower limbs
- □ Slow progression
- □ CK: normal or mildly elevated
- □ Muscle biopsy: internal nuclei, fibre splitting and fibre type 1 predominance
- □ Pattern resembling DNM2 patients



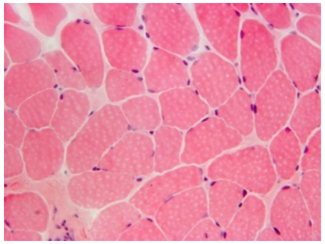
51 years



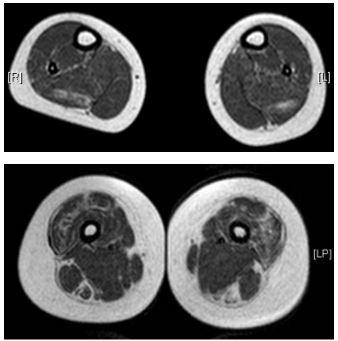
65 years

London case

- □ Antenatal onset with reduced foetal movement
- □ Proximal muscle weakness, mainly lower limbs
- □ Axial weakness
- □ Joint laxity of hands and ankles
- □ Slow improvement
- □ Muscle biopsy: minicores, central cores and some internal nuclei



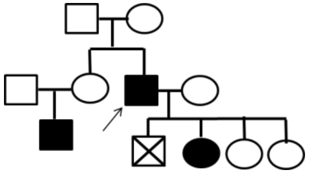
4 years



4 years

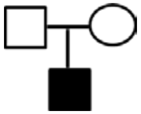
Exome sequencing and data sharing: new congenital myopathy gene

Newcastle case



- □ Stop gain
- □ novel (absent from 62k)
- □ chrX:153049846 G>A; p.Trp415Ter

London case



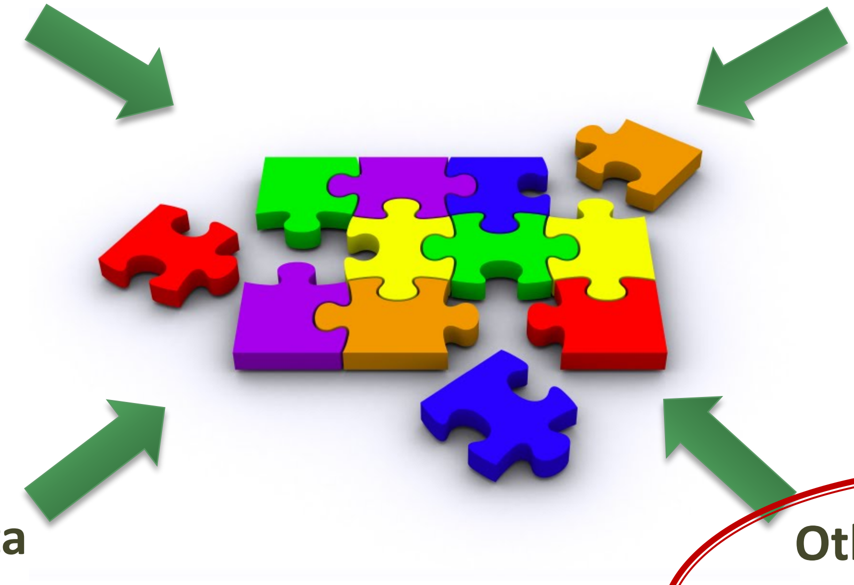
- □ Essential Splice Site
- □ novel (absent from 62k)
- □ chrX:153050629 G>A

SRPK3

Ana Töpf

- □ Serine/arginine protein kinase
- □ Muscle specific, regulated by myocyte enhancer factor 2 (MEF2)
- □ Known to regulate mRNA splicing and nuclear lamina proteins
- □ KO mice develop centronuclear myopathy (Nakagawa et al 2005)
- □ Preliminary data in zebrafish morpholino knockdown shows slow movement and muscle disorganization (unpublished)
- □ Four new mutations found (manuscript in preparation)

Data integration in RD-Connect



Sample data
(biobank databases)

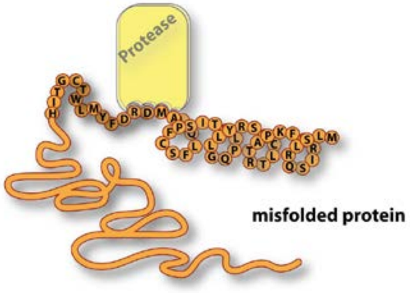
Clinical data
(registries, and phenotypic databases)

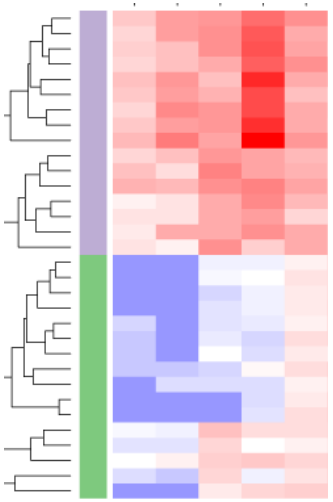
Genomic data
(WES, WGS)

Other omics data
(transcriptomics, metabolomics, proteomics ...)

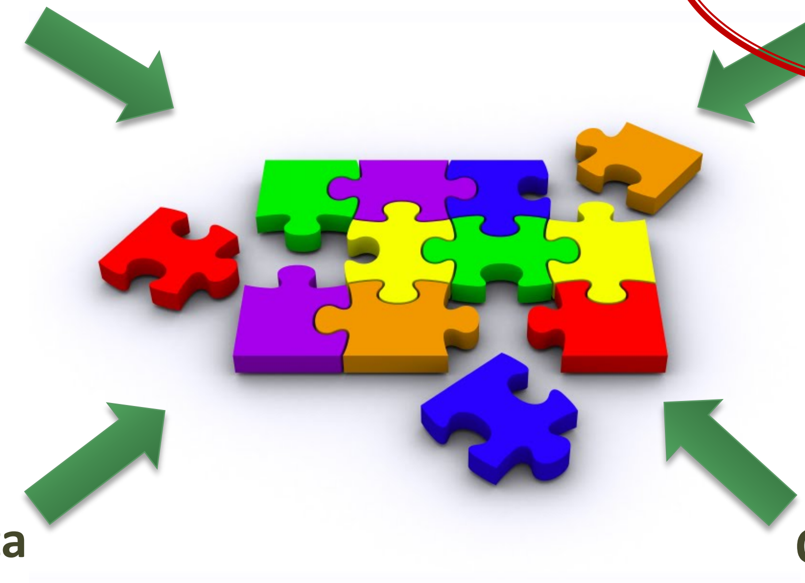
Other omics – work in progress

- □ Integration of other omics data types – transcriptomic, proteomic, lipidomic, metabolomic profiles – is a work in progress
- □ Challenges with standardization of data done on different machines/from different centres
- □ Need to work out the multi-omics research questions that people want to answer
- □ Integration on a per-patient level to allow comparison across data types





Data integration in RD-Connect



Sample data
(biobank databases)

Clinical data
(registries, and phenotypic databases)

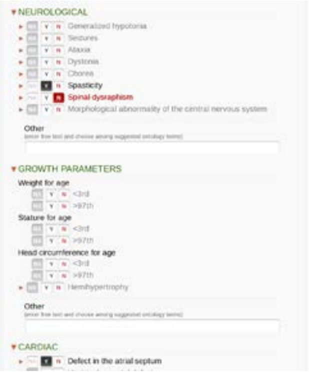
Genomic data
(WES, WGS)

Other omics data
(transcriptomics, metabolomics, proteomics …)

Clinical and phenotypic data

- □ Phenotype is **more important than ever** in the context of clinical outcome measures and next-generation sequencing analysis
- □ Requires transformation into a “computable” form
- □ Requires linkage from different sources (multiple registries, phenotypic databases…)

➡ **FAIR DATA**



▾ NEUROLOGICAL

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

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

▾ GROWTH PARAMETERS

Weight for age
- Y N <3rd
- Y N >97th

Stature for age
- Y N <3rd
- Y N >97th

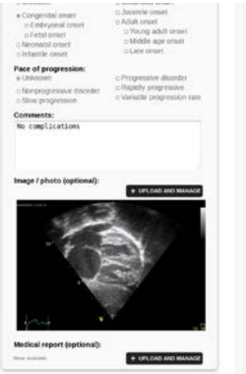
Head circumference for age
- Y N <3rd
- Y N >97th

- ▸ Y N Hemihypertrophy

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

▾ CARDIAC

- ▸ Y N Defect in the atrial septum



● 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):

Medical report (optional):
None available

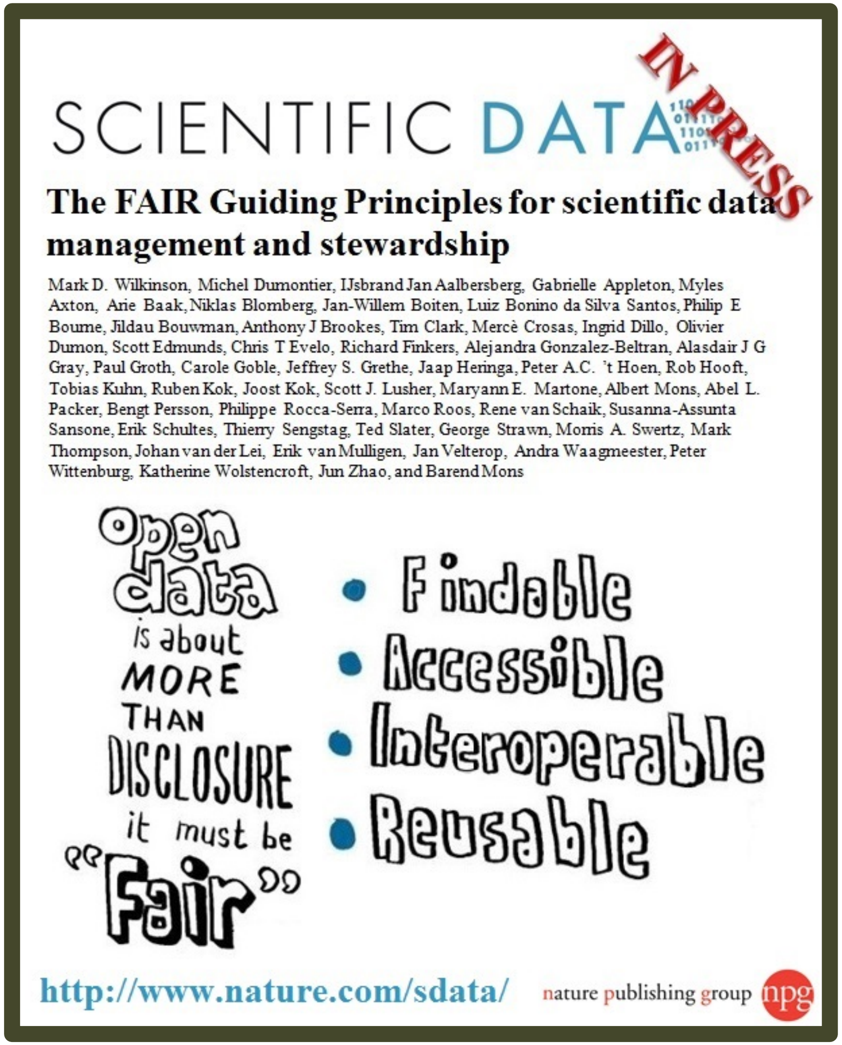
What is FAIR data?

Findable - (meta)data is uniquely and persistently identifiable. Should have basic machine readable descriptive metadata.

Accessible - data is reachable and accessible by humans and machines using standard formats and protocols.

Interoperable - (meta)data is machine readable and annotated with resolvable vocabularies/ontologies.

Reusable - (meta)data is sufficiently well-described to allow (semi)automated integration with other compatible data sources.



IN PRESS

SCIENTIFIC DATA

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

Open data is about MORE THAN DISCLOSURE it must be "Fair"

- Findable
- Accessible
- Interoperable
- Reusable

http://www.nature.com/sdata/ nature publishing group npg

Common data elements

- Attempts to standardize elements collected in patient registries – an ongoing challenge!

PROJECT OVERVIEW CDE SEARCH CRF SEARCH FORM BUILDER CONTACT

NINDS Common Data Elements

Harmonizing Information. Streamlining Research.

EUCERD Joint Action

MINIMUM DATA SET

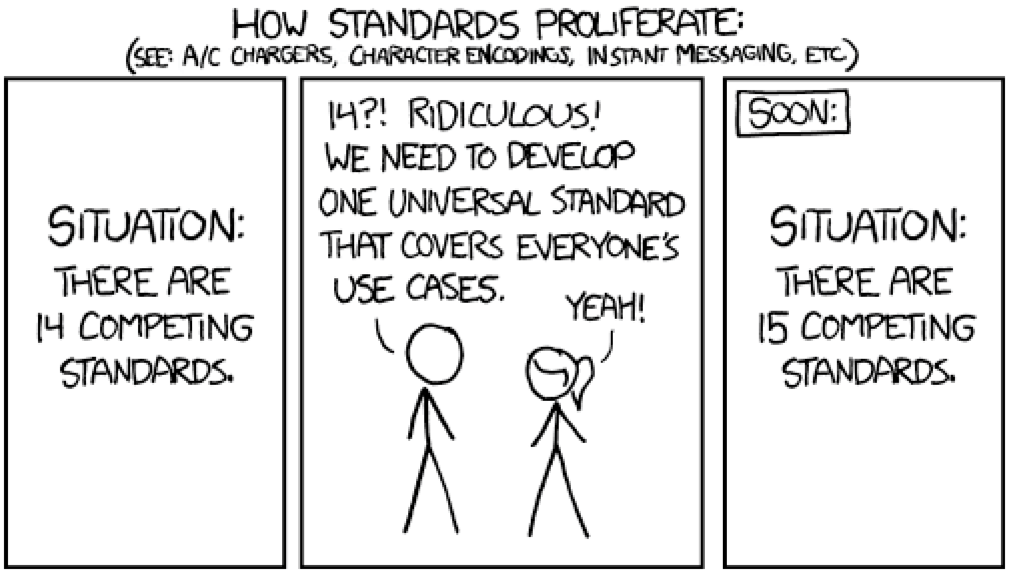
FOR RARE DISEASE REGISTRIES

PARENT

cross-border PAtient REgistries iNiTiative

The EPIRARE proposal of a set of indicators and common data elements for the European platform for rare disease registration

Domenica Taruscio, Emanuela Mollo, Sabina Gainotti, Manuel Posada de la Paz, Fabrizio Bianchi, and Luciano Vittozzi


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

Ontologies

Consensus on most useful ontologies in rare disease:

- □ Human Phenotype Ontology (HPO)
  - For phenotypic descriptions (observations)
- □ Orphanet Rare Disease Ontology (ORDO)
  - For "naming" a disease

Advantages of ontology use:

- □ Computers understand them
- □ Tree structure (if x is true then everything above x is also true)
- □ Allows computational analysis and matchmaking approaches

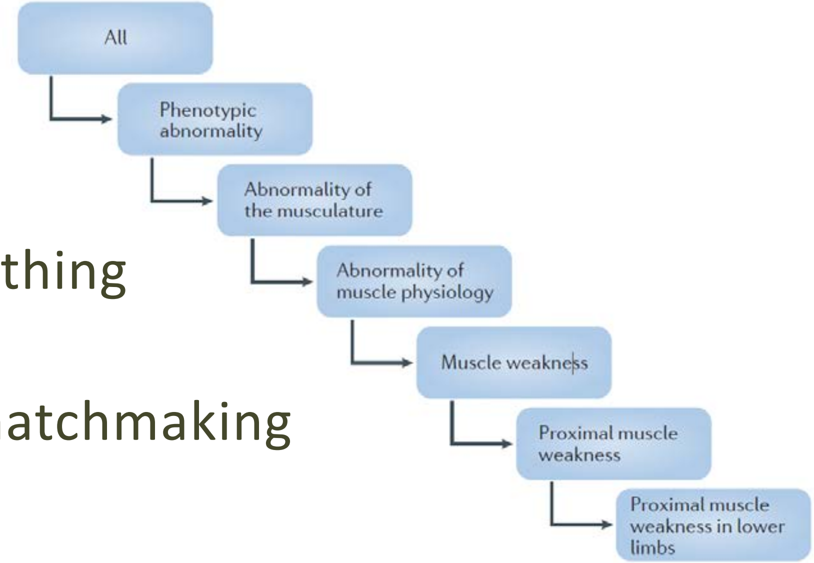
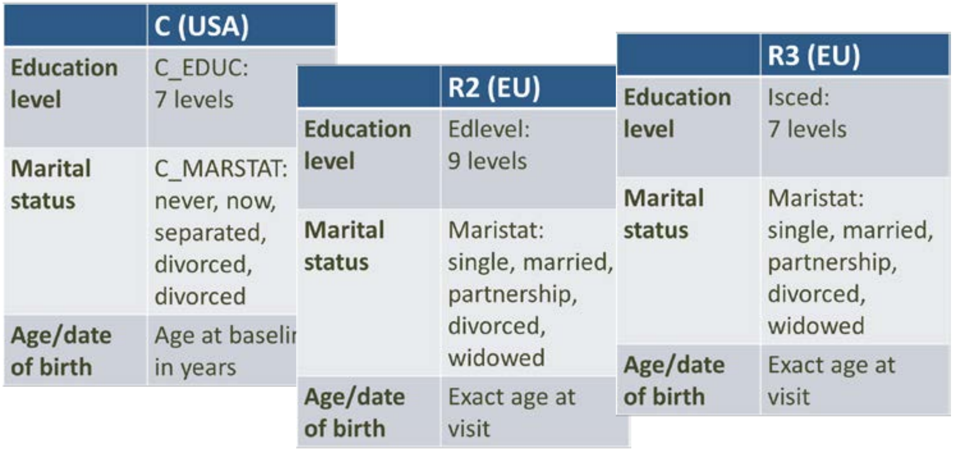

All
Phenotypic abnormality
Abnormality of the musculature
Abnormality of muscle physiology
Muscle weakness
Proximal muscle weakness
Proximal muscle weakness in lower limbs

Figure 4 | **Example of a limb-girdle phenotype hierarchy from the Human Phenotype Ontology (HPO).**

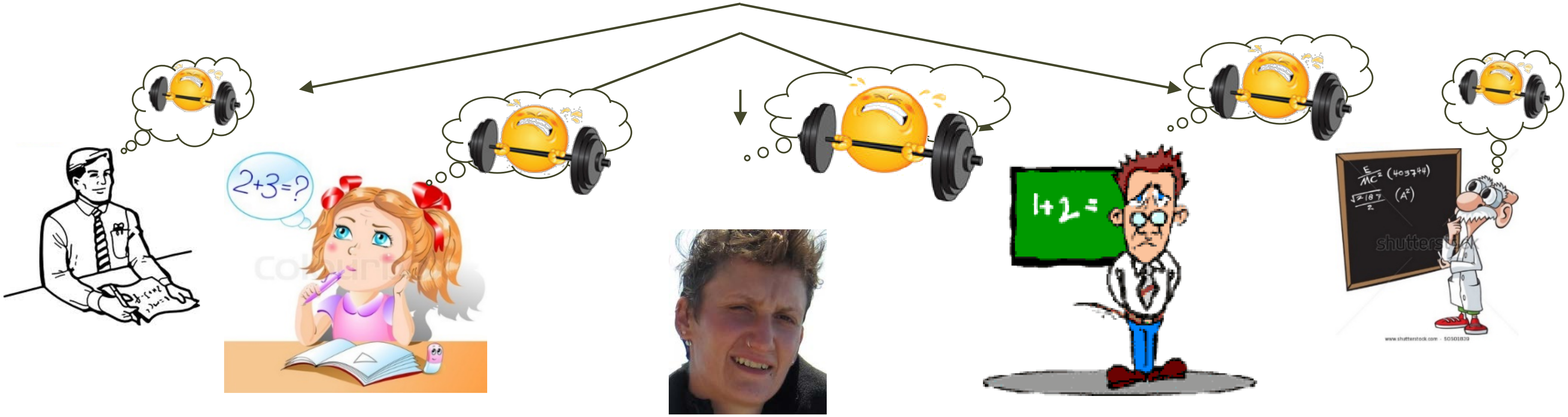
When data is not prepared for cross-resource analysis



	C (USA)
Education level	C_EDUC: 7 levels
Marital status	C_MARSTAT: never, now, separated, divorced, divorced
Age/date of birth	Age at baselir in years

	R2 (EU)
Education level	Edlevel: 9 levels
Marital status	Maristat: single, married, partnership, divorced, widowed
Age/date of birth	Exact age at visit

	R3 (EU)
Education level	Isced: 7 levels
Marital status	Maristat: single, married, partnership, divorced, widowed
Age/date of birth	Exact age at visit



How much time do researchers spend on preparing data for integration

When data is **not** linkable at the source

Back of the envelope calculation

- □ 6 months per data set
- □ Reuse: 5x on average, 6x5=30 Months
- □ For every RD: 6000x(6x5) = 180000 M

How much time do researchers spend on preparing data for integration

When data is linkable at the source

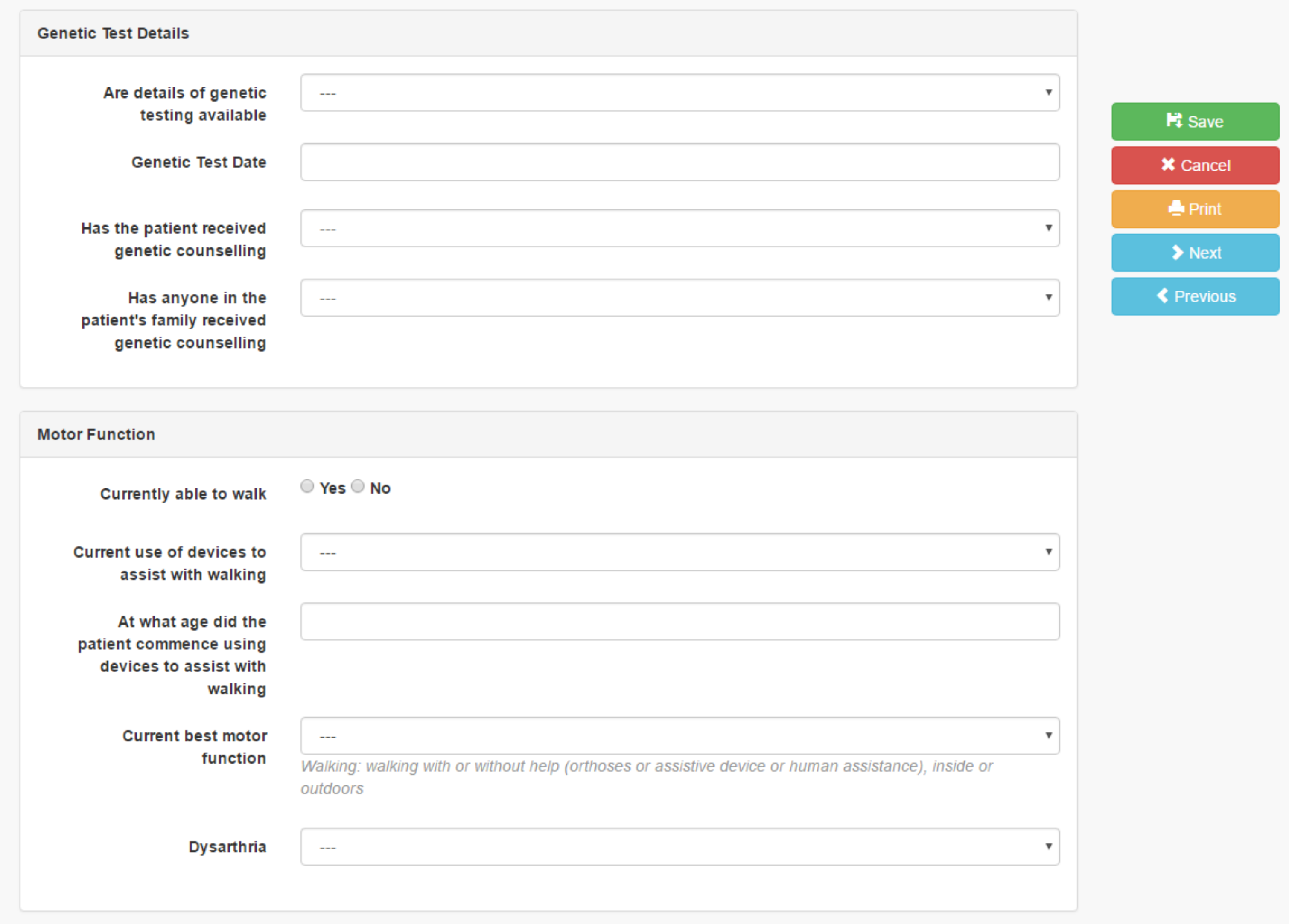
- □ 6 months once
- □ Reuse: 5x on average, +1M, 1x5=5 M (30)
- □ For every RD: 6000x(6+1x5) = 66000 M

(180000)

How much time do researchers spend on preparing data for integration

- □ Benefits for cross-resource analysis
- □ 66% efficiency gain (more time for research)
- □ Researchers can start analysing 6x faster

Rare Disease Registry Framework



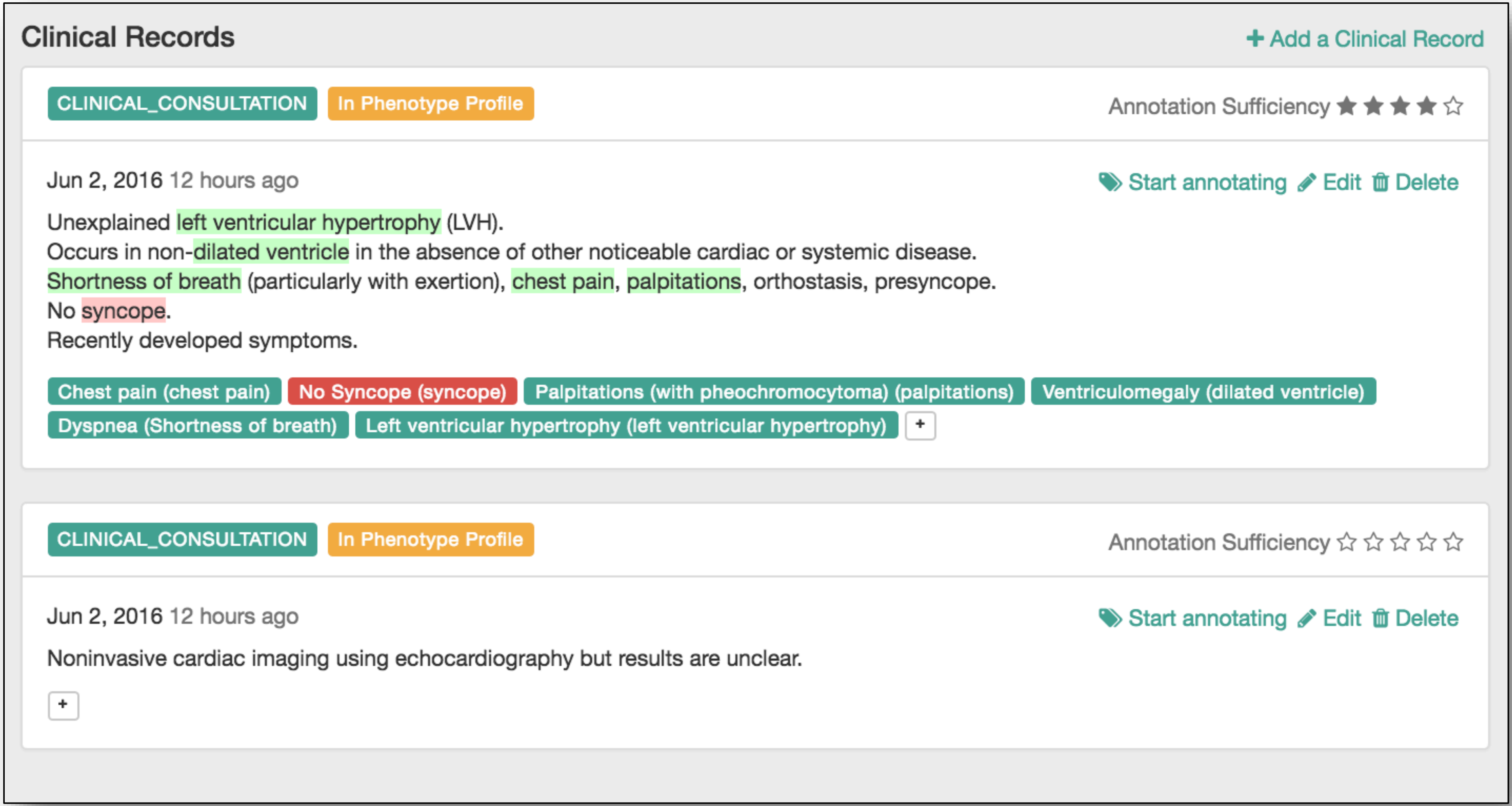
Genetic Test Details

Are details of genetic testing available	---
Genetic Test Date	
Has the patient received genetic counselling	---
Has anyone in the patient's family received genetic counselling	---

Motor Function

Currently able to walk	○ Yes ○ No
Current use of devices to assist with walking	---
At what age did the patient commence using devices to assist with walking	
Current best motor function	--- *Walking: walking with or without help (orthoses or assistive device or human assistance), inside or outdoors*
Dysarthria	---

Patient Archive (HPO)



Clinical Records

Jun 2, 2016 12 hours ago

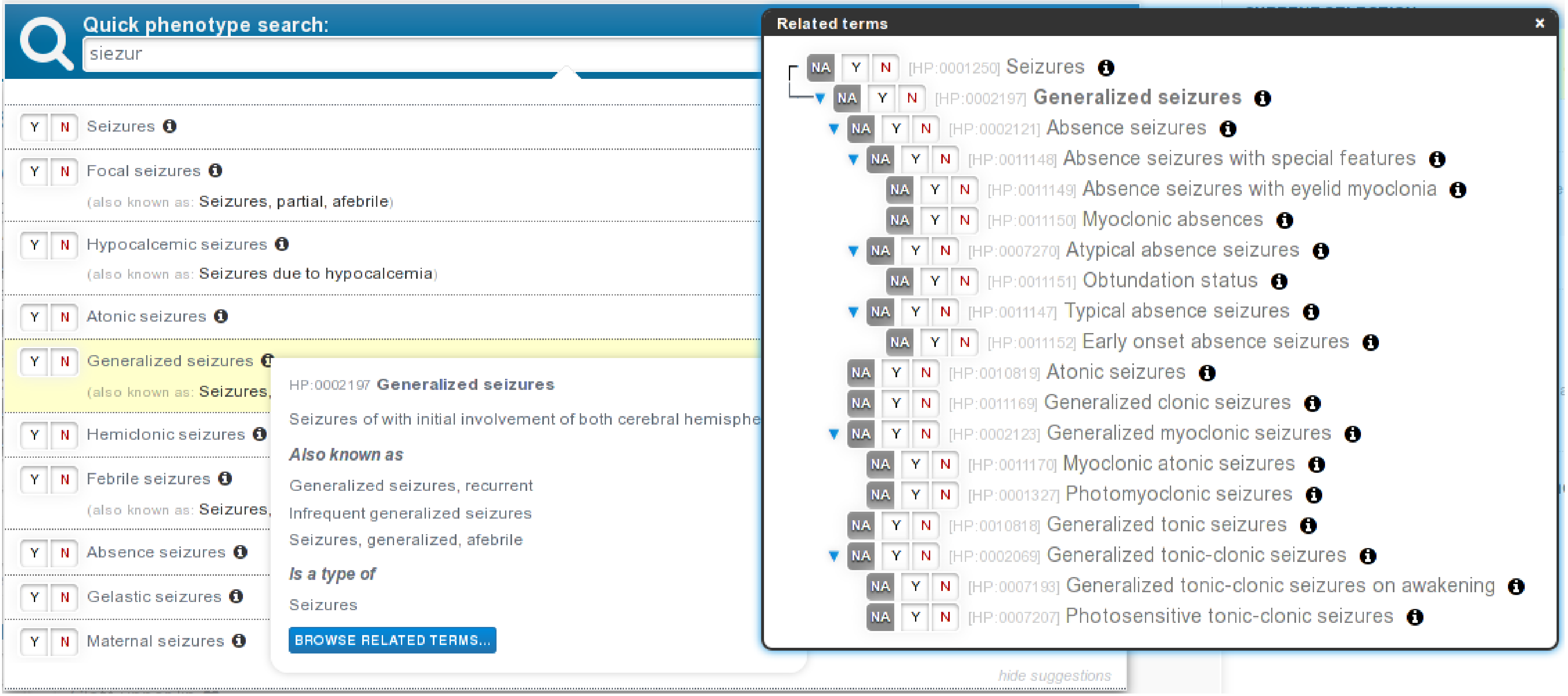
Unexplained left ventricular hypertrophy (LVH).
Occurs in non-dilated ventricle in the absence of other noticeable cardiac or systemic disease.
Shortness of breath (particularly with exertion), chest pain, palpitations, orthostasis, presyncope.
No syncope.
Recently developed symptoms.

Chest pain (chest pain) | No Syncope (syncope) | Palpitations (with pheochromocytoma) (palpitations) | Ventriculomegaly (dilated ventricle) | Dyspnea (Shortness of breath) | Left ventricular hypertrophy (left ventricular hypertrophy) | +

Jun 2, 2016 12 hours ago

Noninvasive cardiac imaging using echocardiography but results are unclear.

PhenoTips (HPO)



Quick phenotype search:

siezur

- Y N Seizures
- Y N Focal seizures
  (also known as: Seizures, partial, afebrile)
- Y N Hypocalcemic seizures
  (also known as: Seizures due to hypocalcemia)
- Y N Atonic seizures
- Y N Generalized seizures
  (also known as: Seizures,
- Y N Hemiclonic seizures
- Y N Febrile seizures
  (also known as: Seizures,
- Y N Absence seizures
- Y N Gelastic seizures
- Y N Maternal seizures

HP:0002197 **Generalized seizures**

Seizures of with initial involvement of both cerebral hemisphe

Also known as

Generalized seizures, recurrent

Infrequent generalized seizures

Seizures, generalized, afebrile

Is a type of

Seizures

BROWSE RELATED TERMS...

hide suggestions

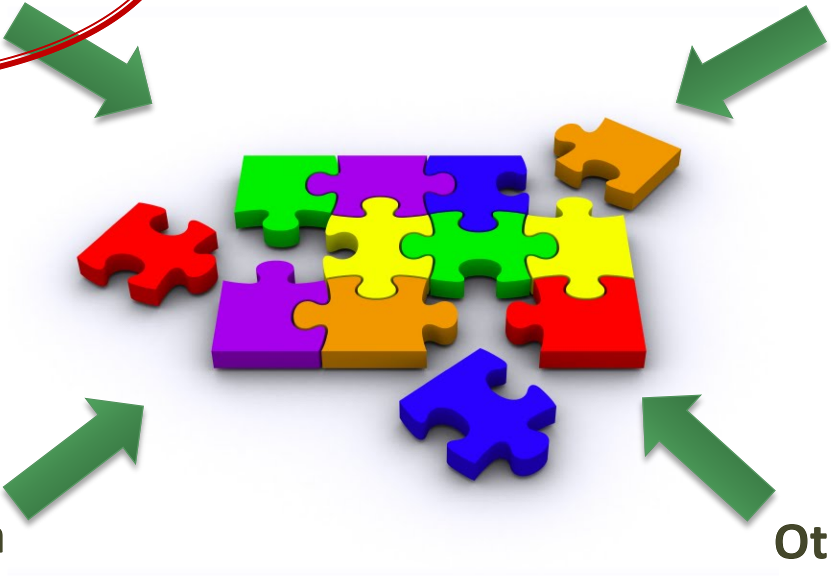
Related terms

- NA Y N [HP:0001250] Seizures
  - NA Y N [HP:0002197] **Generalized seizures**
    - NA Y N [HP:0002121] Absence seizures
      - NA Y N [HP:0011148] Absence seizures with special features
        - NA Y N [HP:0011149] Absence seizures with eyelid myoclonia
        - NA Y N [HP:0011150] Myoclonic absences
      - NA Y N [HP:0007270] Atypical absence seizures
        - NA Y N [HP:0011151] Obtundation status
      - NA Y N [HP:0011147] Typical absence seizures
        - NA Y N [HP:0011152] Early onset absence seizures
    - NA Y N [HP:0010819] Atonic seizures
    - NA Y N [HP:0011169] Generalized clonic seizures
    - NA Y N [HP:0002123] Generalized myoclonic seizures
      - NA Y N [HP:0011170] Myoclonic atonic seizures
      - NA Y N [HP:0001327] Photomyoclonic seizures
    - NA Y N [HP:0010818] Generalized tonic seizures
    - NA Y N [HP:0002069] Generalized tonic-clonic seizures
      - NA Y N [HP:0007193] Generalized tonic-clonic seizures on awakening
      - NA Y N [HP:0007207] Photosensitive tonic-clonic seizures

Data integration

Sample data
(biobank databases)

Clinical data
(registries, and phenotypic databases)



Genomic data
(WES, WGS)

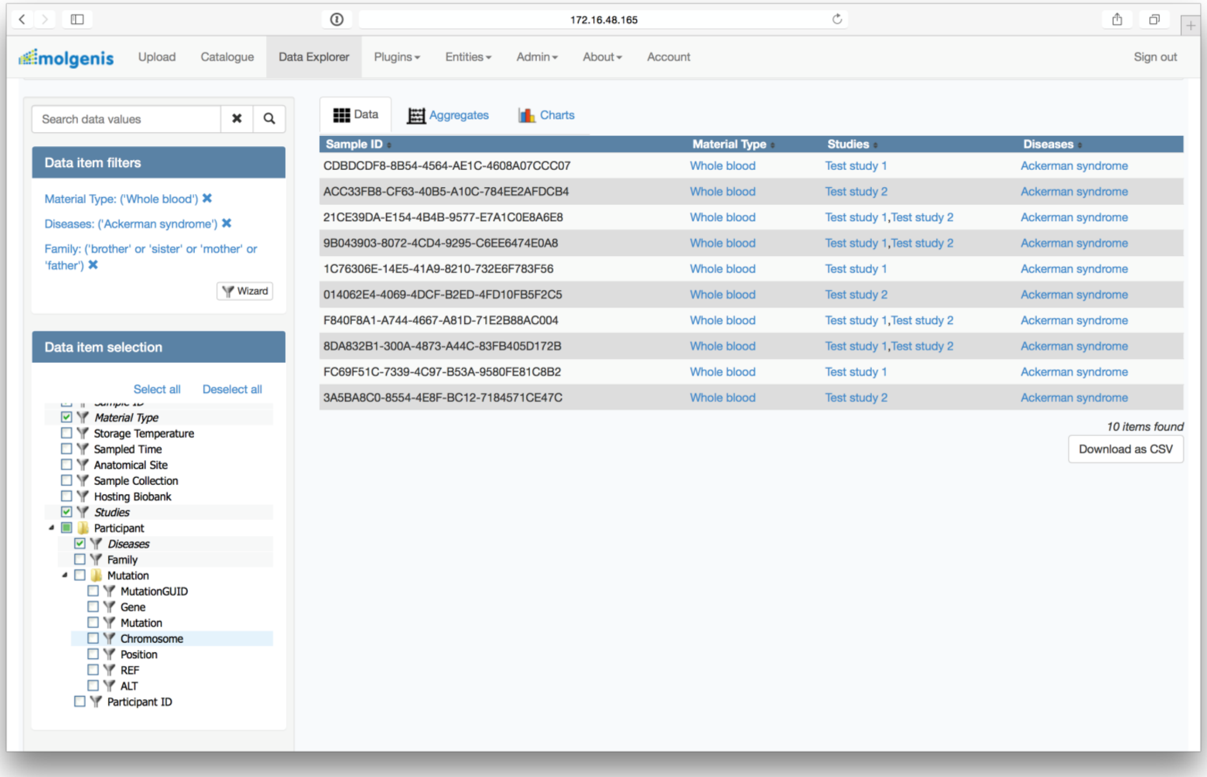
Other omics data
(transcriptomics, metabolomics, proteomics …)

Biosample data

- (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

Biosample database

- RD-Connect biosample database contains sample-level data from all participating biobanks



Data item filters

Material Type: ('Whole blood') ✖

Diseases: ('Ackerman syndrome') ✖

Family: ('brother' or 'sister' or 'mother' or 'father') ✖

Wizard

Data item selection

Select all Deselect all

- ☑ *Material Type*
- ☐ Storage Temperature
- ☐ Sampled Time
- ☐ Anatomical Site
- ☐ Sample Collection
- ☐ Hosting Biobank
- ☑ *Studies*
- ■ Participant
  - ☑ *Diseases*
  - ☐ Family
  - ☐ Mutation
    - ☐ MutationGUID
    - ☐ Gene
    - ☐ Mutation
    - ☐ Chromosome
    - ☐ Position
    - ☐ REF
    - ☐ ALT
  - ☐ Participant ID

Sample ID	Material Type	Studies	Diseases
CDBDCDF8-8B54-4564-AE1C-4608A07CCC07	Whole blood	Test study 1	Ackerman syndrome
ACC33FB8-CF63-40B5-A10C-784EE2AFDCB4	Whole blood	Test study 2	Ackerman syndrome
21CE39DA-E154-4B4B-9577-E7A1C0E8A6E8	Whole blood	Test study 1,Test study 2	Ackerman syndrome
9B043903-8072-4CD4-9295-C6EE6474E0A8	Whole blood	Test study 1,Test study 2	Ackerman syndrome
1C76306E-14E5-41A9-8210-732E6F783F56	Whole blood	Test study 1	Ackerman syndrome
014062E4-4069-4DCF-B2ED-4FD10FB5F2C5	Whole blood	Test study 2	Ackerman syndrome
F840F8A1-A744-4667-A81D-71E2B88AC004	Whole blood	Test study 1,Test study 2	Ackerman syndrome
8DA832B1-300A-4873-A44C-83FB405D172B	Whole blood	Test study 1,Test study 2	Ackerman syndrome
FC69F51C-7339-4C97-B53A-9580FE81C8B2	Whole blood	Test study 1	Ackerman syndrome
3A5BA8C0-8554-4E8F-BC12-7184571CE47C	Whole blood	Test study 2	Ackerman syndrome

10 items found

Download as CSV

Patient (research participant) identifier

- Platform cannot store personally identifiable information (for obvious privacy reasons)
- How do we track that different data items (biosample, natural history data, exome sequence) all come from the same patient?
  - Assign an identifier (e.g. EURenOmics case: HEIDELBERG1234)
    - Advantage: Simple
    - Limitation: requires a central point (e.g. clinician) who knows the link between the patient and the identifier for all datasets
  - Generate an identifier from personally identifiable information
    - Advantage: the same patient will always have the same ID even if clinician A (who stored the biosample) doesn't know that clinician B uploaded an exome for the same patient
    - Limitation: requires consensus on a set of PII sufficient for generating a unique identifier – may be hard to do retrospectively if this info was not available

Existing systems for identifier

US NIH GUID

- □ Originally used by National Database for Autism Research; concept now extended to several other NIH projects, with plans for a RD GUID
- □ Based on a standardised set of PII (including first, middle and surname as on birth certificate, date of birth, city of birth as on birth certificate)
- □ Participant PII is entered into a Java webservice client application, which generates a one-way hash
- □ Hash is sent to central NIH server, which returns a GUID for that participant

Plan moving forward for identifier

- □ At least in the interim, RD-Connect will establish an ID system for European RD projects contributing data to RD-Connect (partner projects can assign all patients an RD-ID)
- □ BUT use the SAME set of PII used in NIH (and Huntington) systems (interoperable)
- □ Continue to enable linkage of data in the platform by other mechanisms (e.g. manually generated ID) where it is not possible to generate an ID due to lack of PII (legacy data)
- □ Contribute to the task force jointly set up by IRDiRC and GA4GH and implement its output when ready

Questions/feedback

Coordinator:

Hanns Lochmüller – hanns.lochmuller@ncl.ac.uk

Data helpdesk – personalised support!

John Dawson – john.dawson@ncl.ac.uk

Technical questions:

Sergi Beltran – sergi.beltran@cnag.crg.eu

Research questions:

Rachel Thompson – rachel.thompson@ncl.ac.uk

All other questions:

Emma Heslop – emma.heslop@ncl.ac.uk

THE John Walton MUSCULAR DYSTROPHY RESEARCH CENTRE

Newcastle University

The Newcastle upon Tyne Hospitals NHS Foundation Trust

TREAT-NMD Neuromuscular Network

MRC Centre for Neuromuscular Diseases

@ConnectRD
@treat_nmd
@bushbykate

RD Connect

NeurOmics

EURenOmics

rachel.thompson@ncl.ac.uk
hanns.lochmuller@ncl.ac.uk

#Brexit – thanks for all the support!