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

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

- Lack of cohorts for trials
- Lack of data and samples for research
- Lack of interest from pharma
- Lack of genetic diagnosis
- Lack of cases to confirm diagnosis

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- Reluctance to share unpublished data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 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)
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?

“black box”
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

#### Filters

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

#### Gene Table

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Transcript ID</th>
<th>Effect Impact</th>
<th>Consequence</th>
<th>Feature Type</th>
<th>HGVS Coding</th>
<th>Amino Acid Change</th>
<th>Amino Acid Length</th>
<th>Genotype Number</th>
<th>Exon Rank</th>
<th>CDS Position</th>
<th>Transcript BioType</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFAP2</td>
<td>ENST0000037553</td>
<td>MODERATE</td>
<td>missense_variant</td>
<td>transcript</td>
<td>c.313A&gt;C</td>
<td>p.Thr105Pro</td>
<td>183</td>
<td>1</td>
<td>7/9</td>
<td>313/552</td>
<td>protein_coding</td>
</tr>
<tr>
<td>MFAP2</td>
<td>ENST0000037553</td>
<td>MODERATE</td>
<td>missense_variant</td>
<td>transcript</td>
<td>c.310A&gt;C</td>
<td>p.Thr104Pro</td>
<td>182</td>
<td>1</td>
<td>6/8</td>
<td>310/549</td>
<td>protein_coding</td>
</tr>
<tr>
<td>MFAP2</td>
<td>ENST0000043854</td>
<td>MODERATE</td>
<td>missense_variant</td>
<td>transcript</td>
<td>c.310A&gt;C</td>
<td>p.Thr104Pro</td>
<td>182</td>
<td>1</td>
<td>7/0</td>
<td>310/549</td>
<td>protein_coding</td>
</tr>
</tbody>
</table>

#### Variants (11)

| Chr | Pos  | Ref | Alt | Candidate | GT|000010 | GT|000036 | GT|000037 | 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   | D  | NA   | 0         |
| 11  | 93535027 | C   | A   | 0         | ADD | C/A   | C/C    | C/C    |       | MED17 | MODERATE     | 34      | D     | D    | D   | D  | NA   | 0         |
| 13  | 7785365 | G   | A   | 0         | ADD | G/A   | G/G    | G/G    |       | MYCBP2 | MODERATE     | 25.6    | T     | D    | D   | D  | NA   | 0         |
| 13  | 110437802 | A   | C   | 0         | ADD | A/C   | A/A    | A/A    |       | IR52  | MODERATE     | 24.6    | D     | D    | D   | D  | NA   | 0         |
| 13  | 113210444 | G   | T   | 0         | ADD | G/T   | G/G    | G/G    |       | TUBGCP3 | MODERATE     | < 20    | T     | B    | D   | D  | NA   | 0         |
SEARCH ACROSS ALL SAMPLES FOR VARIANTS IN SPECIFIC GENES

<table>
<thead>
<tr>
<th>Filters</th>
<th>PRESET FILTERS</th>
<th>RESET</th>
<th>SHARE</th>
<th>RUN QUERY</th>
</tr>
</thead>
</table>

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

Sample Selection?

Select individual Samples ✗ or search across all ✓?

- Affected: [ ]
- ID: [ ]
- 0/0: [ ]
- 0/1: [ ]
- 1/1: [ ]

- Compound het: [ ]

- All_SAMPLES: [ ]

- Min DP: 20
- Min GQ: 50
- Min AAF: 0.2
- Max AAF: 0.8

(accessible: 1451, own: 0, shared: 174, visible to all: 1277)
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

Gene identification/pathophysiology:
- Biomarkers
- Animal models
- Delivery mechanisms

Clinical Trials:
- Diagnosis/care standards
- Patient Registries
- Trial sites
- Outcome measures

Therapy delivery:
- Regulatory affairs
- Ethics
- Commissioning/health economics

Back translation
FROM GENE TO POTENTIAL THERAPY

LRP4, MUSK, Agrin, LRP4-MUSK, DOK7

Ras-associating Myosin motor IQ IQ IQ IQ IQ PKC Rho-GAP

Head Neck Tail

Patient 1: p.R1517H
Patient 1: p.R2283H
Patient 2 and 3: p.D1698G

Patient 1:
p.R2283H

Patient 2 and 3:
p.D1698G

F-actin F-actin

Control NSC34 MYO9A NSC34 F-actin

09 April 2018
DEFINITION OF TREATABLE UNITS


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


**Neuromuscular junction immaturity and muscle atrophy are hallmarks of the ColQ-

Gene: Dok7

<table>
<thead>
<tr>
<th>Name</th>
<th>docking protein 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonyms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dok-7</td>
</tr>
<tr>
<td></td>
<td>EF-12</td>
</tr>
<tr>
<td></td>
<td>Gt5</td>
</tr>
<tr>
<td></td>
<td>AS30013K19Rik</td>
</tr>
<tr>
<td>Status</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ES Cells</td>
</tr>
<tr>
<td></td>
<td>Mice tm1</td>
</tr>
<tr>
<td></td>
<td>Mice tm1.1</td>
</tr>
<tr>
<td></td>
<td>phenotype data available</td>
</tr>
<tr>
<td>Links</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MGI:3584043</td>
</tr>
<tr>
<td></td>
<td>Ensembl Gene</td>
</tr>
<tr>
<td></td>
<td>Ensembl Orthologs</td>
</tr>
<tr>
<td></td>
<td>ENU(12)</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS/COLLABORATIONS

CNAG team in Barcelona led by Sergi Beltran

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