WP5: GENOME-PHENOME ANALYSIS PLATFORM UPDATE

Ivo Gut and Sergi Beltran
RD-Connect Platform 2017-2018 highlights

**Genome-Phenome Analysis Platform**
- 3003 Genomes & Exomes
- 436 Users from 170 Groups

**Sample Catalogue**
- 7352 Biosamples
- Covering 90 Diseases

**Registry & Biobank Finder**
- 381 Patient Registries and Biobanks
- Covering 1441 Diseases

✅ **EJP-RD proposal**: part of Pillar 2 Virtual Platform, increase capabilities, promote usage by ERNs, add new functionalities, multi-omics, connect to LifeScience/ELIXIR AAI, participation in PPRL, security and GDPR compliance assessment, FAIR assessment, provide documentation and guidelines.
RD-Connect Platform
(platform.rd-connect.eu)

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

Welcome to the central platform for access to data submitted by RD-Connect’s partner projects. The online Genome-Phenome Analysis Platform is now open for submissions from all users. Our automated registration system will come online shortly, but if you would like to access the interface now please email platform@rd-connect.eu and we will contact you to request the information we need to set you up on the system.

Get started today
A paradigm for best practice in data sharing in genomic studies?
1: Clinical researchers submit patient IDAT from Biobanks (context 1) and RD-connect (context 2) to EUPID which will store, encrypt and pseudonymise the data. A PSN ID is returned to both instances.

2: Clinical researchers enter the PSN of the patient in PhenoTips and an automatic verification into the EUPID-biobank-RD-connect linked database is performed.

3: If matching PSNs are detected within RD-connect or between Biobanks and RD-connect, an alert will be send to PhenoTips.
# RD-Connect GPAP interface

## Filters
- **Variant Type:** coding (high/LG, moderate/LG)
- **Population:** exac SNV->MT: A, D, SNV->SIFT: D, SNV->PP2: D, P

<table>
<thead>
<tr>
<th>Samples</th>
<th>Functional</th>
<th>Predictive</th>
<th>Population</th>
<th>Diseasecard</th>
<th>Candidate</th>
<th>Links</th>
<th>ALFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene Name</td>
<td>Transcript ID</td>
<td>Effect Impact</td>
<td>Consequence</td>
<td>Feature Type</td>
<td>HGVS coding</td>
<td>Amino Acid change</td>
<td>Amino Acid length</td>
</tr>
<tr>
<td>MFAP2</td>
<td>ENST00000375533</td>
<td>MODERATE</td>
<td>missense_variant</td>
<td>transcript</td>
<td>c.313A&gt;C</td>
<td>p.Thr105Pro</td>
<td>183</td>
</tr>
<tr>
<td>MFAP2</td>
<td>ENST00000375533</td>
<td>MODERATE</td>
<td>missense_variant</td>
<td>transcript</td>
<td>c.310A&gt;C</td>
<td>p.Thr104Pro</td>
<td>182</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>
</tr>
</tbody>
</table>

## Variants (11)

<table>
<thead>
<tr>
<th>Chr</th>
<th>Pos</th>
<th>Ref</th>
<th>Alt</th>
<th>Candidate</th>
<th>GT E000010</th>
<th>GT E000036</th>
<th>GT E000037</th>
<th>INDEL</th>
<th>Gene Name</th>
<th>Effect Impact</th>
<th>ClinVar</th>
<th>CADD</th>
<th>SIFT</th>
<th>PP2</th>
<th>MT</th>
<th>ExAC</th>
<th>1000GP AF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17302199</td>
<td>T</td>
<td>G</td>
<td>0</td>
<td>ADD</td>
<td>T/G</td>
<td>T/T</td>
<td>T/T</td>
<td>MFAP2</td>
<td>MODERATE</td>
<td>25.9</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>93530527</td>
<td>C</td>
<td>A</td>
<td>0</td>
<td>ADD</td>
<td>C/A</td>
<td>G/C</td>
<td>G/C</td>
<td>MEDI7</td>
<td>MODERATE</td>
<td>34.3</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>78835365</td>
<td>G</td>
<td>A</td>
<td>0</td>
<td>ADD</td>
<td>G/A</td>
<td>G/G</td>
<td>G/G</td>
<td>MYCBP2</td>
<td>MODERATE</td>
<td>25.6</td>
<td>T</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>110437802</td>
<td>A</td>
<td>C</td>
<td>0</td>
<td>ADD</td>
<td>A/C</td>
<td>A/A</td>
<td>A/A</td>
<td>IRS2</td>
<td>MODERATE</td>
<td>24.6</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>113210444</td>
<td>G</td>
<td>T</td>
<td>0</td>
<td>ADD</td>
<td>G/T</td>
<td>G/G</td>
<td>G/G</td>
<td>TUBGCP3</td>
<td>MODERATE</td>
<td>&lt; 20</td>
<td>T</td>
<td>B</td>
<td>D</td>
<td>D</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>9066234</td>
<td>A</td>
<td>C</td>
<td>0</td>
<td>ADD</td>
<td>A/C</td>
<td>A/A</td>
<td>A/A</td>
<td>NTN1</td>
<td>MODERATE</td>
<td>28.7</td>
<td>D</td>
<td>P</td>
<td>D</td>
<td>D</td>
<td>NA</td>
<td>0</td>
</tr>
</tbody>
</table>
RD-Connect GPAP functionalities

- **Standard filters and annotations** (variant impact on protein, frequency in control populations, frequency within database, *in silico* predictors, gene, genomic position, etc.)
- **Filtering by genes of interest** (predefined/custom lists, associated to OMIM disease, HPO symptoms or Reactome)
- Filter by variants annotated in **ClinVar database**
- Filter to regions with observed long **Runs of homozygosity** (RoH)
- Direct link to **multiple external resources** (Ensembl, EXaC, gnomAD, HGMD, UCSC, NCBI, DiseaseCards, ALFA, etc.)
- **Phenotype driven strategy for variant prioritization** (Exomiser)
- **Anonymized data sharing** through Matchmaker exchange and Beacon
3,003 samples fully processed (42%)

7 GUI releases since last annual meeting
- PhenoTips tab added + generate gene lists based on HPO terms
- Pathways tab added + generate gene list by pathway
- Filter to user-defined co-ordinates or ranges (via upload)
- Displaying RD-Connect variant frequencies + can use as filter

MMEv1.0 (internal & external matchmaking—PhenomeCentral)

User management enhancements in RD-Connect's CAS

Data submission interface released
- Bulk metadata upload
- Owners can view who has been accessing their samples
ELIXIR Implementation Study to visualize data deposited at the EGA in real-time

1. File submission to the EGA & ID-mapping
2. Request
   - File ID/s
   - Chr. coordinates
3. Reply
   - BAM/CRAM slice
   - Data representation
4. Rendering*

*Broad’s IGV screenshot

Secure and encrypted connections
ELIXIR Implementation Study: Integration of ELIXIR-IIB in ELIXIR Rare Diseases activities

MINT / MENTHA (protein interaction databases)

VHLdb: A database of von Hippel-Lindau protein interactors and mutations

GALT-PROTEIN-DB 2.0

Human Mitochondrial DataBase

eDGAR disease-gene Association database

REDIportal An ATLAS of A-to-I RNA editing events in human
BENCHMARKING AND STANDARDIZING GENOMICS PIPELINE

- Working in collaboration with the RD communities to implement the adequate quality reporting standards

- Working in close collaboration with ELIXIR-EXCELERATE community benchmarking (WP2, OpenEbench)

- ELIXIR IS: "Development of Architecture for Software Containers at ELIXIR and its use by EXCELERATE use-case communities"
### ELIXIR bio.tools including RD-Connect tools/resources

<table>
<thead>
<tr>
<th><strong>RD-Connect Genome-Phenome Analysis Platform</strong></th>
<th><strong>RD-Connect Sample Catalogue</strong></th>
<th><strong>PhenoTips</strong></th>
<th><strong>Variant Analysis and Filtration Tool (VarAFT)</strong></th>
<th><strong>Online Mendelian inheritance in Man (OMIM)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>An online tool for diagnosis and gene discovery in rare disease research. The platform features allow identifying disease-causing mutations in rare disease</td>
<td>This catalogue is intended to facilitate the discovery of samples and samples data from Rare Diseases biobanks. It also provides information about sample data.</td>
<td>Open source software tool for collecting and analyzing phenotypic information for patients with genetic disorders.</td>
<td>VarAFT (Variant Annotation and Filter Tool) is a standalone and multiplatform (Windows, Mac, Unix) tool for the annotation and filtering of DNA variants.</td>
<td>Catalog of human genes and genetic disorders with specific test, reference.</td>
</tr>
<tr>
<td>Sequence analysis</td>
<td>Deposition</td>
<td>Pathway or network analysis</td>
<td>SNP annotation</td>
<td>Genetic variation analysis</td>
</tr>
<tr>
<td>Data retrieval</td>
<td>Data retrieval</td>
<td>Phylogenetic tree analysis</td>
<td>Service invocation</td>
<td>Service invocation</td>
</tr>
<tr>
<td>Genetic variation analysis</td>
<td></td>
<td>Gene functional annotation</td>
<td>Filtering</td>
<td>Genetic variation analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Variant classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**API will enable displaying list of tools in RD-Connect website**
Contributors

WP1: Coordination
Hanns Lochmüller / Ivo Gut
(CRG / Newcastle)

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)

CNAG
I. Gut
S. Beltran
D. Piscia
S. Laurie
A. Papak.
I. Martinez
R. Tonda
J.R. Trotta
L. Matalonga
D. Picó
C. Luengo
F. Camacho

BSC
A. Valencia
S. Capella
J.M. Fernández
M. Bossio

AMU (Marseille)
C. Béroud
D. Salgado
J.P. Desvignes

ISCIII
M. Posada
E. López

UCMG
M. Swertz
D. van Eckenvort

Interactive BioSoftware
A. Blavier
S. Lair

U. Aveiro
J.L. Oliveira
P. Lopes
P. Sernaleda

LUMC
P.B. t’Hoen
M. Roos
M. Thompson
R. Raliyaperumal
B. Mons

U. of Patras
G. Patrinos

EGA
T. Keane
D. Spalding
J. Paschall
J. Almeida-King
J. Rambla

Murdoch U.
M. Bellgard

Newcastle U.
H. Lochmüller
R. Thompson
A. Topf
I. Zaharieva

U. Of Toronto
M. Brudno
M. Girdea
S. Dumitriu
O. Buske