RD Connect

MULTI-OMICS ANALYSES IN RD-CONNECT

IRDiRC INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

SEVENTH FRAMEWORK PROGRAMME

Queries (to be) enabled by RD-Connect

- *Show me all **patients** with a **variant** in gene X who also have **phenotype** descriptor Y*
- *Which **biobanks** contain blood **samples** from **patients** with **disease** X and phenotype Y*

- *Are there **biobank samples** available for individuals with **disease** X and genetic **variant** Y?*

What's next?

- Inclusion of other –omics data:
  - D4.19 (M60): Development of Integrative -Omics Analysis Suite

- **A: disease-centered questions**
- **B: patient-centered questions**

A. Disease-centered questions

- *Is the involvement of signal transduction **pathway** X in rare **disease** Y supported by both **genomic and proteomic data**?*
- *Do **molecular profiles** of **rare diseases** X and Y converge at the level of **protein** (and potential **drug target**) Z?*

B. Patient-centered questions

- *Which serum **proteins** distinguish slow from fast **disease progression** in patients with the same **genetic variant** X?*
- *Is there **proteomic** evidence for translation of **genetic variant** X (proteogenomics)?*
- *Which **genetic variants and metabolites** can be used to **predict the response to drug** X in patients with rare **disease** Y?*

Different requirements

- A. requires aggregated data
- B. requires patient-level data

- Implications for data storage
  - Where (inside, outside RD-Connect database), what (processed, unprocessed), how (formats)
- Implications for query structure and tooling

Different tooling requirements

- A. Connection of molecular features and signatures from different datasets across diseases and –omics levels
- B. Construction of subgroups of patients and/or prepare multi-level –omics data for the user in convenient formats to be used in subsequent statistical analyses

We may need to choose

- Resources are limited
- In particular for ‘joint’ RD-Connect efforts

- Pros and Cons???

Pros and cons A:

- Pros:
  - Most multi-omics experiments are case-control designs and do not have extensive phenotype information for individual samples
  - Automated or user of molecular signatures (sets of (differentially expressed) genes, proteins, metabolites, possibly with some test statistic) by users is not too difficult to accommodate
  - Best suited to find commonalities between different rare disorders
  - Most suited for the discovery of druggable pathways and targets for rare diseases (IRDiRC 2020 goals)
- Cons:
  - Does not easily fit in the current database structure, which is focused on individual-level data
  - No clear connection to biobank and registry components of the RD-Connect platform

Pros and cons B:

- Pros:
  - This set-up may prepare for personalized drug treatment, the medicine of the future
  - Knowledge of *e.g.* transcript or protein expression associated with particular variants may be useful in establishing gene function and plausibility of variant pathogenicity
  - Capturing individual-level data also allows linkage back to the original biosample to enable further studies on the same samples
  - Hosting data in this form is more closely aligned with current database structure, which is also individual-centered
  - The set-up is unique: there is no (?) other platform where (deep) phenotype information is integrated with individual-level cross-omics analyses
- Cons:
  - The number of multi-omics datasets with (deep) individual-level phenotype information is (still) small
  - This type of use cases may be difficult to answer without a properly designed study and dedicated data mining and statistical tests

Implications for RD-Connect platform

- Integration with other RD-Connect resources (phenotype database, biobanks, registries, genomics platform)
- Interaction with EBI resources
- Role for user-operated data processing pipelines

Role of EBI repositories

- EGA: WES/WGS and RNA-seq
- PRIDE: Proteomics
- Metabolights: Metabolomics

- Sustainable resources for primary and processed data and metadata (!) preservation
- Working together on:
  - Submission workflows
  - Identifier mapping issues

Standardized analysis pipelines

- WES/WGS: in place
- RNA-seq: in progress
- Proteomics: no plans yet. Discussion with PeptideShaker[1]
- Metabolomics: no plans yet.
- Proteomics and Metabolomics commercial data generators in B-projects have their own data processing pipelines (more or less standardized per project)

[1] Vaudel et al. *Nature Biotechnology* **33**, 22–24 (2015)

Multi-omics task force

- Joint RD-Connect / Neuromics / EuRenOmics taskforce
- Involvement from academic and commercial partners
- Intention to collaborate on shared infrastructural challenges and the analysis of specific use cases

Multi-omics analyses

- Use cases identified based on available multi-omics data (2016)
  - Proteomics, metabolomics and miRNA urinary biomarker profiles of steroid-resistant nephrotic syndrome (SRNS) (Schaeffer, EURenomics)
  - Longitudinal Metabolomics (serum, urine) and Transcriptomics (blood, muscle) in mouse models for DMD (Spitali, Neuromics)

- Worked on in 2015
  - Finding relevant biological connections between transcriptomics and metabolomics blood biomarker profiles of presymptomatic and symptomatic Huntington's patients (van Roon, Neuromics)

Acknowledgements

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

SEVENTH FRAMEWORK
PROGRAMME

EUReni Omics

NeuroOmics

Rachel Thompson
Sergi Beltran
Marco Roos
Kristina Hettne