Standardized analysis and sharing of genome-phenome data for neuromuscular and rare disease research through the RD-Connect platform

S. Beltran[1,2], D. Piscia[1,2], S. Laurie[1,2], J. Protasio[1,2], A. Papakonstantinou[1,2], A. Cañada[3,14], J.M. Fernández[3,14], M. Thompson[6], R. Kaliyaperumal[6], S. Lair[7], P. Sernadela[8], M. Girdea[9], M. Brudno[9], A. Blavier[7], R. Thompson[10], H. Lochmüller[10], **A. Roos**[10], V. Straub[10], M. Bellgard[11], J. Paschall[12], M. Roos[6], P.A.C. 't Hoen[6], A. Valencia[3,14], D. Salgado[4,5], C. Béroud[4,5,13], I. Gut[1,2] and the RD-Connect Consortium

[1]Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Spain, [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain, [3]Centro Nacional de Investigaciones Oncológicas (CNIO) , Madrid, Spain, [4]Aix-Marseille Université, Marseille, France, [5]Inserm, UMR_S 910, Marseille, France, [6]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, [7]Interactive Biosoftware, Rouen, France, [8]DETI/IEETA, University of Aveiro, Portugal, [9]Centre for Computational Medicine, Hospital for Sick Children and University of Toronto, Toronto, ON, Canada, [10]John Walton Muscular Dystrophy Research Centre, Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, UK, [11]Centre for Comparative Genomics, Murdoch University, Perth, Western Australia, [12]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, United Kingdom, [13]APHM, Hôpital TIMONE Enfants , Laboratoire de Génétique Moléculaire, Marseille, France, [14]Instituto Nacional de Bioinformática (INB), Spain

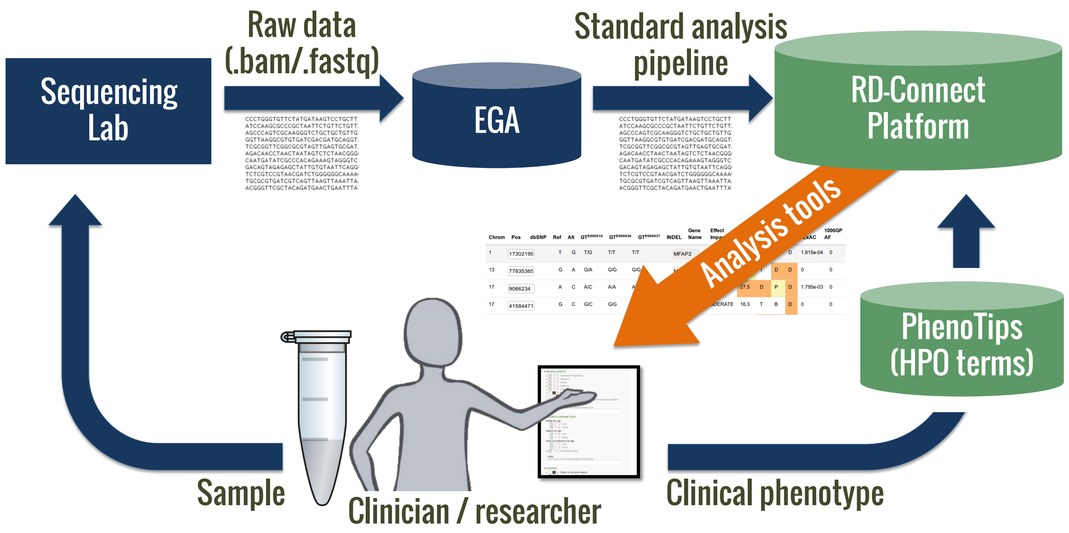
Summary

- Around 300 million people worldwide are affected by one of the 6000+ known rare diseases (RD).
- RD research faces specific challenges because patient groups, clinical expertise, and research communities are small in number and highly fragmented.
- Next-generation sequencing (NGS) and genomics research have opened up new possibilities for gene discovery for the many cases that remain undiagnosed after single-gene testing.
- Many RD expert centres lack the bioinformatics expertise and computational support to take full advantage of the new genomic paradigm.
- RD-Connect has developed a unique online platform that provides both a repository for RD research data and a user-friendly interface for NGS analysis that has much of the same functionality as commercial analysis suites.
- Research partners from any RD research project or clinical centre worldwide can submit data, analyse their own patients and compare with data submitted by other centres.
- The platform enables data sharing at various levels. **Global Alliance Beacon**: simple Y/N: does this variant exist in this cohort? **Matchmaker Exchange**: finding patients in external databases with matching phenotype and candidate variant in the same gene. **Internal sharing**: full access to shared datasets from other centres for further study.
- Although open to any RD, the platform is currently enriched for neuromuscular and neurodegenerative phenotypes, with several thousand datasets (exome, genome and gene panel) currently available to registered users.
- The platform also links to biosample data, other omics data and data from patient registries at a per-patient level, enabling combined queries across different resources and data types.
- Ethical and legal expertise is available to ensure data sharing and research meets appropriate consent and data protection provisions, and an active patient advisory council led by Eurordis, the European alliance of rare disease patient organizations, ensures explicit engagement of patients at every level.

To register visit platform.rd-connect.eu

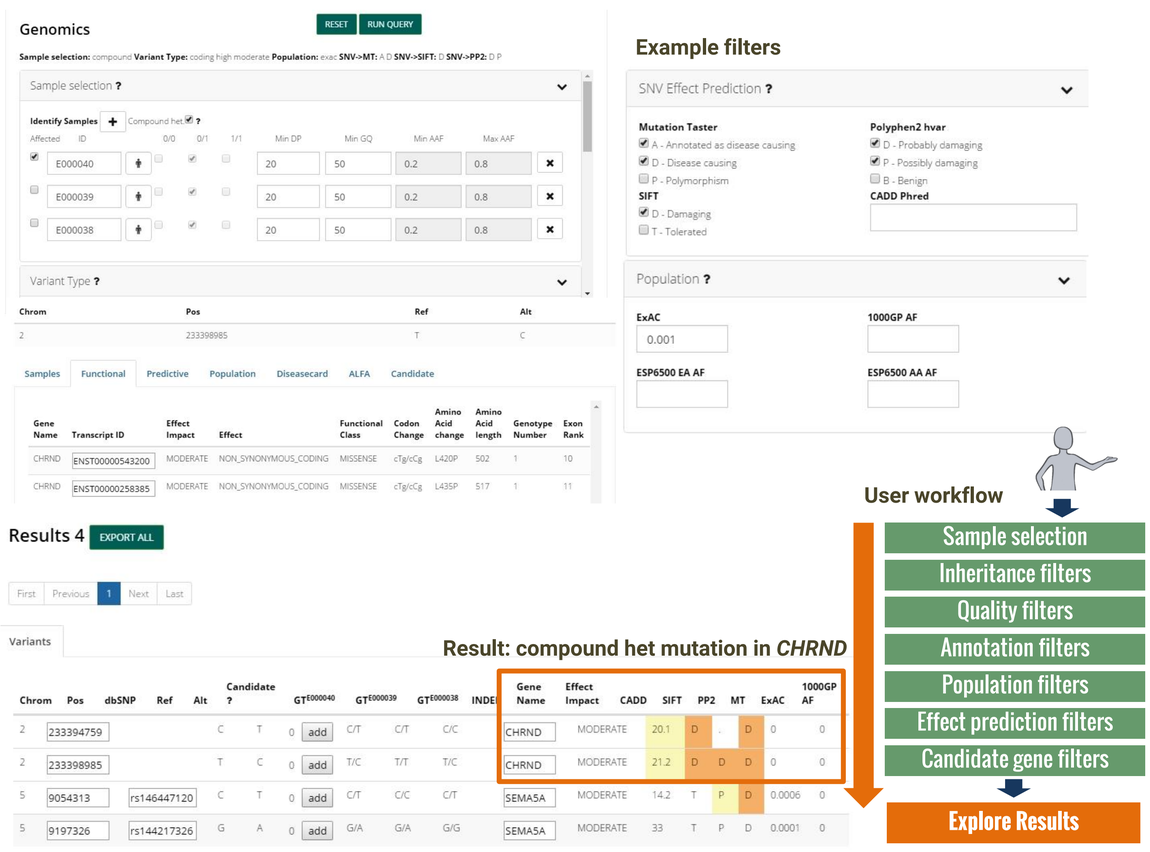
Genomic data workflow

Raw data (the output from the sequencing machine) is stored for long-term archival at the **European Genome-phenome Archive** (EGA). This raw data is reprocessed through a standard analysis pipeline before being made accessible in the RD-Connect platform. The standardization of variant calling and annotation helps output from different sequencing providers to become comparable. Within RD-Connect the sequencing data is combined with detailed phenotypic data standardized using the **Human Phenotype Ontology** (HPO) to allow combined genotype-phenotype analysis. Registered clinicians and researchers can analyse their own data and access data from other submitters.



Genomic analysis user interface

RD-Connect is not only a data repository but a full-featured genomic analysis platform with a particular focus on diagnosis and gene discovery. Validated RD researchers can upload their own data and perform combined analyses of index cases with family members, filtering the many thousand variants in a single exome or genome using a range of filters and candidate gene lists The underlying technologies are designed for big data, enabling real-time analysis at whole-genome scale. The screenshots here show an example query for an affected individual with a congenital myasthenic syndrome phenotype, analysed together with unaffected parents.

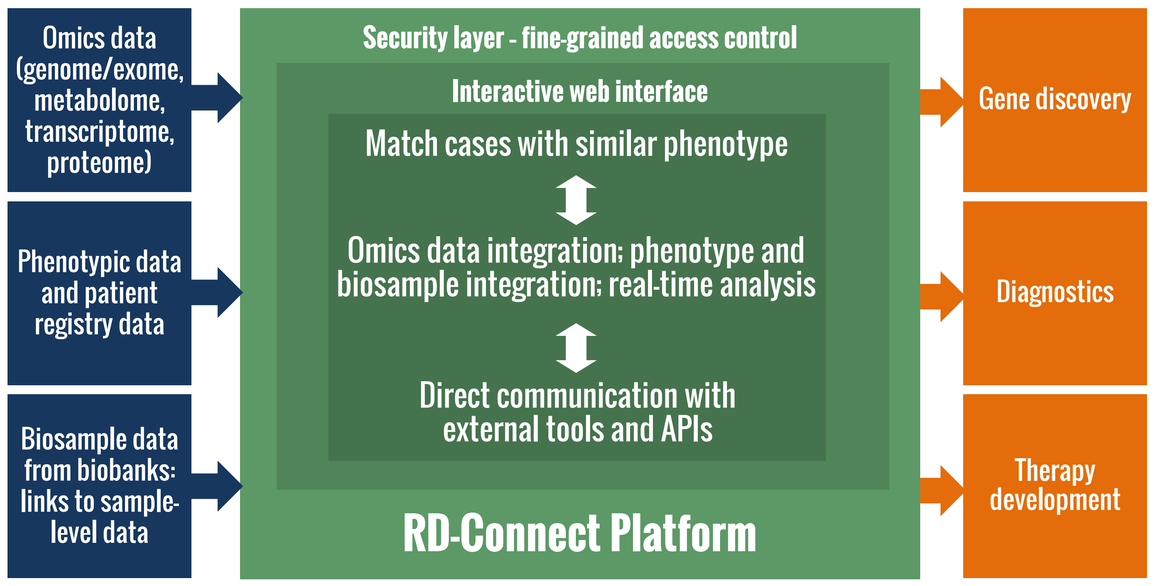


Patient and ELSI work

The genomic era brings many opportunities for people living with rare diseases, in particular those without a genetic diagnosis. It also brings new ethical, legal and social challenges: how do we maximize the benefits and opportunities for new diagnoses and new knowledge while minimizing risks relating to sharing genomic data? RD-Connect's ELSI work has resulted in the development of standardized procedures for informed consent, a charter of principles for data sharing, and work exploring patient attitudes to data sharing. The RD Patient and Ethics Council and Patient Advisory Committee enable in-depth patient engagement in all areas of the project's scope.

Platform overview

In the central RD-Connect platform, **genomic data** is united with **phenotypic data** at a per-patient level. Other types of omics data, in particular **transcriptomic**, **metabolomic** and **proteomic profiles**, are in the process of being incorporated where available. The corresponding clinical information from each individual is recorded in a connected PhenoTips instance, a software solution that simplifies the capture of clinical data using the Human Phenotype Ontology, and linked with OMIM and Orpha codes. The Global Alliance **MatchMaker Exchange** API is implemented to allow remote querying of external databases to find matching cases that may allow a diagnosis to be confirmed.
A unique feature of RD-Connect is its linkage with **biobanks** and **patient registries**. Through novel mechanisms for data linkage, remote resources whose data is Findable, Accessible, Interoperable and Reusable (FAIR) can be linked in at a per-patient or per-sample level, enabling researchers to find a cohort or sample of interest for further study.



Collaborating projects

Two partner projects funded at the same time as RD-Connect, NeurOmics and EURenOmics, were the first to submit data to the RD-Connect platform and helped to beta-test the system, suggesting new functionality. In summer 2016 the platform was opened up to external users and submissions are rapidly increasing. The data in the platform is currently highly enriched for neuromuscular and neurodegenerative phenotypes and provides an excellent opportunity for researchers engaged in study of these disorders to access whole-exome and whole-genome data and upload and analyse their own cohorts.
In addition to data-producing projects, RD-Connect collaborates closely with infrastructure partners such as ELIXIR, BBMRI, EuroBioBank and the Global Alliance for Genomics and Health in order to harmonize efforts and work together on shared goals.

NeurOmics (Tübingen, Riess) 1000 exomes and genomes, neuromuscular/neurodegenerative phenotypes

EURenOmics (Heidelberg, Schaefer) 1000 exomes and panels, rare kidney phenotypes

MYO-SEQ SEQUENCING GENETIC MUSCLE DISEASES (Newcastle, Straub) 1000 exomes (index cases), LGMD phenotype

Neurogenetics (Newcastle, Horvath): 350 exomes, neurogenetic and mitochondrial phenotypes

SeqNMD (Broad Institute, MacArthur): 500 exomes, LGMD phenotype

Other projects 500 exomes, various phenotypes

European Reference Networks

elixir EuroBioBank BBMRI-ERIC Biobanking and BioMolecular resources Research Infrastructure Global Alliance for Genomics & Health

Analyse your data with RD-Connect

- The RD-Connect platform is open for omics data submissions from any research project or centre engaged in rare disease research.
- Basic research groups can also sign up to access human data related to their research area.
- In all cases, the PI of the group must register and undergo a validation procedure and may then sign up group members.
- Registration provides access to the genomics platform, support with data submission and additional advice e.g. with ensuring patient consents are suitable for data sharing.
- Patient registries and biobanks can sign up at catalogue.rd-connect.eu to publicize details of their resource and access templates, MTAs and more.

Contact platform@rd-connect.eu to discuss your requirements.

www.rd-connect.eu

IRDiRC INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

The John Walton MUSCULAR DYSTROPHY RESEARCH CENTRE

The Newcastle upon Tyne Hospitals NHS Foundation Trust · Newcastle University

cnag centre nacional d'anàlisi genòmica *centro nacional de análisis genómico*

RD Connect

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 2012-305444