RD-Connect: an integrated infrastructure for data sharing and analysis in rare disease research

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], D. Badowska[10], V. Straub[10], M. Bellgard[11], J. Paschall[12], M. Roos[6], P.A.C. 't Hoen[6], A. Valencia[3,14], L. Monaco[15], CM. Wang[15], D. Salgado[4,5], C. Béroud[4,5,13], D. Taruscio[16], S. Gainotti[16], Y. Kodra[16], C. Carta[16], P. Torreri[16], 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, [15]Fondazione Telethon, Milan, Italy, [16]Istituto Superiore di Sanità, Rome, Italy

Summary

Around 300 million people worldwide are affected by one of the 6000+ known rare diseases (RD).

RD research faces specific challenges because patients, clinical expertise and research communities are scarce and fragmented. Data sharing between researchers is therefore crucial.

Next-generation sequencing (NGS) and genomics have opened up new possibilities for gene discovery and diagnosis, but many RD expert centres lack the bioinformatics expertise and computational support to take full advantage of the new genomic paradigm.

RD-Connect is an EU-funded infrastructure bringing together multiple data types used in rare disease research into a common resource for researchers and clinicians.

RD-Connect consists of three systems: Genome-Phenome Analysis Platform (*platform.rd-connect.eu*), Sample Catalogue (*samples.rd-connect.eu*) and Registry & Biobank Finder (*catalogue.rd-connect.eu*), which are open to any RD and available free of charge.

To register visit platform.rd-connect.eu

Genome-Phenome Analysis Platform 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. RD researchers or clinical centres worldwide can submit data, analyse own patients and compare with data submitted by others, and find patients in external databases with matching phenotype and candidate variant in the same gene.

Registry & Biobank Finder is a global directory of rare disease patient registries and biobanks, including the numbers of register cases for each disease.

Sample Catalogue allows browsing biosample collections of RD biobanks and find detailed information about individual biosamples.

RD-Connect ethical and legal experts provide guidance to researchers and work on optimal models for data sharing, while engagement of patient representatives at every level of the project work ensures patient-centred approach.

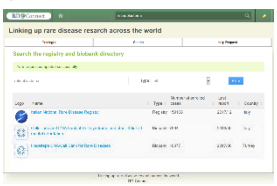
For overview, watch our explanatory video: https://youtu.be/i0C03vpGhDM

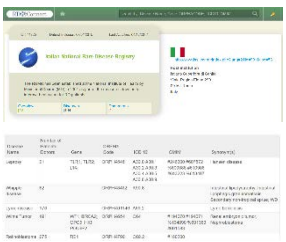
RD-Connect overview

A unique feature of RD-Connect is that it links into a common resource different types of data used in rare disease research: genomics and other omics, phenotypes, patient registries, biobanks and biosamples. 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.

Registry & Biobank Finder

Registry & Biobank Finder is an online directory of RD patient registries and biobanks. It lets researchers find the registries and biobanks that store data on their rare disease of interest. The system provides databases' contact data and the numbers of registered cases for each disease, regularly updated by database curators. Users can search the directory by disease name and its synonyms, ORPHA- or OMIM-codes and other keywords, e.g. country or registry manager. The system also provides access to documents such as databases' study protocols, case report forms, informed consent templates and data access agreements.



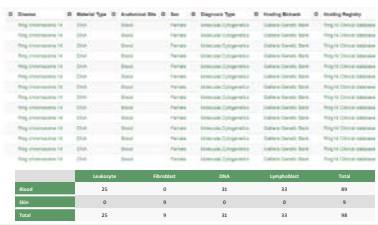


We also support creating new rare disease patient registries and help registry managers set their registry adapting the FAIR principles that allows linking up with other databases.

Sample Catalogue

Sample Catalogue allows browsing biosample collections stored in RD biobanks using powerful filtering functions. It provides detailed **information about individual biosamples**, including not only the disease, sample type, sex and availability of genetic and registry data, but also genotype, diagnosis type, and availability of longitudinal samples and samples from unaffected family members. Sample Catalogue includes now **over 25,000 samples** uploaded by the EuroBioBank Network. The work is ongoing to make the Sample Catalogue interconnected with the other RD-Connect tools, to enable queries such as finding biosamples from patients with a specific genetic variant.

Example query: DNA samples from females with Ring chromosome 14 syndrome.



	Leukocyte	Fibroblast	DNA	Lymphoblast	Total
Blood	25	0	31	33	89
Skin	0	9	0	0	9
Total	25	9	31	33	98

Analyse your data with RD-Connect

Genome-Phenome Analysis Platform is open for genomics and phenotype data submissions from any research project or centre engaged in rare disease research.

Basic researchers 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 Platform, support with data submission and additional advice e.g. with ensuring patient consents are suitable for data sharing. ***Contact platform@rd-connect.eu to register.***

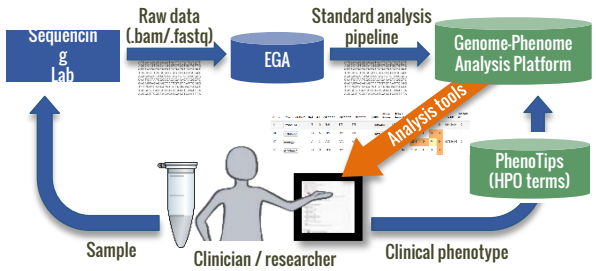
Patient registries and biobanks can sign up at *catalogue.rd-connect.eu* to publicize details of their resource, access templates, Material Transfer Agreements and more.

Biobanks are welcome to submit information about their sample to the Sample Catalogue. ***Please visit samples.rd-connect.eu.***

visit *platform.rd-connect.eu*

Genome-Phenome Analysis Platform

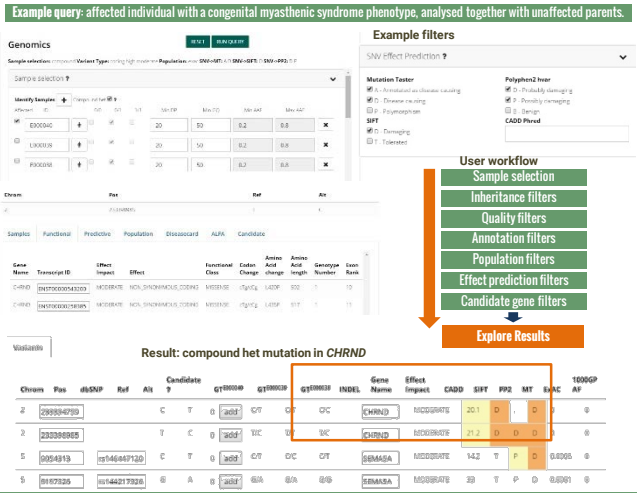
Genome-Phenome Analysis Platform is not only a data repository but also a full-featured genomic analysis interface with a particular focus on diagnosis and gene discovery. Raw sequencing data are stored for long-term archival at the **European Genome-phenome Archive (EGA)**. Before being made accessible in RD-Connect, data are reprocessed through a standard analysis pipeline to allow for comparisons of output from different sequencing providers. The Platform combines - at a per-patient level! - sequencing data with detailed phenotypic information stored in the **PhenoTips** system and standardized using the **Human Phenotype Ontology (HPO)**, OMIM and Orpha-codes. Other types of omics data, in particular transcriptomic, metabolomic and proteomic profiles, are in the process of being incorporated where available.



Registered and validated RD researchers and clinicians can quickly and easily analyse own data and compare them with the data submitted by others. The Platform is open for submissions and already contains thousands of datasets from several research projects. By 2019, it is expected to contain **30,000 samples**, including 19,000 from the European Reference Networks submitted by the Solve-RD project.

The **user-friendly** analysis interface enables combined analyses of affected individuals with family members and finding disease-causing variants in a single exome or genome using a range of filters and candidate gene lists. The integrated **Global Alliance MatchMaker Exchange** enables remote querying of external databases to find matching cases that may allow a diagnosis to be confirmed.

Example query: affected individual with a congenital myasthenic syndrome phenotype, analysed together with unaffected parents.



Patient and ELSI work

The genomic era brings new opportunities for people living with rare diseases, particularly those with no genetic diagnosis. It also brings new ethical, legal and social challenges: how do we maximize the benefits and opportunities for new diagnoses and knowledge while minimizing risks relating to sharing genomic data? RD-Connect's work on **Ethical, Legal and Social Issues (ELSI)** 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.

rd-connect.eu

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