RD-CONNECT: an integrated platform connecting registries, biobanks and clinical bioinformatics for rare disease research
FP7 Collaborative project HEALTH.2012.2.1.1-1-C
Grant agreement number: 305444
Start date: 1st November 2012
Duration: 72 months
Coordinator: Prof Hanns LOCHMULLER, University of Newcastle upon Tyne

RD-Connect Partners

Participant organisation name	PI	Country
University of Newcastle upon Tyne (UNEW)	Hanns Lochmüller, Kate Bushby	UK
Centro Nacional de Analisis Genómico (CNAG)	Ivo Gut	ES
Aix-Marseille University Medical School (AMU)	Christophe Béroud	FR
Istituto Superiore di Sanità (ISS)	Domenica Taruscio	IT
Uppsala University (UU)	Mats G. Hansson	SE
Leiden University Medical Center (LUMC)	Peter-Bram 't Hoen	NL
Centro Nacional de Investigaciones Oncológicas (CNIO)	Alfonso Valencia	ES
Fondazione Telethon (FTELE)	Lucia Monaco	IT
University of Aveiro (UAVR)	Jose Luis Oliveira	PT
Karolinska Institutet (KI)	Jan-Eric Litton	SE
University of Patras (UPAT)	George P. Patrinos	EL
European Organisation for Rare Diseases (EURORDIS)	Yann LeCam, Anna Kole	FR
Interactive Biosoftware (IBS)	André Blavier	FR
Finovatis (Finovatis)	David Koubi	FR
Instituto de Salud Carlos III (ISCIII)	Manuel Posada	ES
Innolyst, Inc., Patient Crossroads (PC)	Kyle Brown	US
Medical University of Graz (MUG)	Kurt Zatloukal	AT
University Paris Diderot – Paris 7 (UPD-7)	Odile Boespflug-Tanguy	FR
University of Malta (UoM)	Alex Felice	MT
Fondation maladies rares (Fondation MR)	Nicolas Lévy	FR
University Hospital of Ulm (UULM)	Michael Orth	DE
University of Zürich (UZH)	Nenad Blau	CH
Charles University (CUNI)	Milan Macek	CZ
National Institutes of Health – Office of Rare Diseases Research (NIH-ORDR)	Steve Groft, Yaffa Rubinstein	US
Murdoch University (MURDOCH)	Matthew Bellgard	AU
Office of Population Health Genomics (OPHG)	Hugh Dawkins	AU
European Bioinformatics Institute (EMBL-EBI)	Paul Flicek	UK

Publishable Summary

Period 1 (1Nov 2012 – 31 Oct 2013)

Overview

Following negotiations with the European Commission and the signing of a grant agreement, RD-Connect became operational on 1 November 2012. The overarching objectives for the 6-year funding period are to develop an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research and to contribute to the International Rare Disease Research Consortium (IRDiRC) objectives of delivering 200 new therapies for rare diseases and means to diagnose most rare diseases by the year 2020. For this reason, close cooperation with two omics research projects funded at the same time – Neuromics (coordinated by Olaf Riess, Tübingen) and EURenOmics (coordinated by Franz Schaefer Heidelberg) – was envisaged.

In its first year of operation, RD-Connect has successfully achieved its objectives for the period and has begun to establish its position as an important part of the global rare disease research infrastructure. Owing to the need to integrate with existing initiatives, the primary focus of the year has been on ensuring that RD-Connect activities are fully aligned with the needs of the associated projects that will submit data to the system, and on developing interoperability with related tools and projects operating in the same area. Of particular note in this regard, strong collaborations with EURenOmics and Neuromics have ensured that RD-Connect is developing in a direction that enables it to be of utility to the data-generating projects, while the incorporation of new associated partners involved in related work aims to avoid duplication of effort and enable RD-Connect activities to be fully embedded in the wider rare disease research field.

Project launch and IRDiRC integration

In order to increase the profile of the IRDiRC and promote the first major EU-funded projects falling under its auspices, an official press event was held in Barcelona on 24 January 2013 to formally launch RD-Connect, Neuromics, EURenOmics and Support-IRDiRC, projects representing a total of 38M EUR of funding for rare disease research. The subsequent joint kick-off meeting for RD-Connect, Neuromics and EURenOmics was crucial to foster early collaborations between the three projects, and the 2014 annual meeting in Heidelberg is likewise being held jointly. As strongly encouraged by RD-Connect's Scientific Advisory Board, chaired by Bartha Knoppers, close links with the coordinators of EURenOmics and Neuromics have been established to ensure integration of activities. RD-Connect coordinator Hanns Lochmüller chairs the IRDiRC Interdisciplinary Scientific Committee and a number of RD-Connect partners are members of IRDiRC committees and working groups, which has helped ensure harmonisation with IRDiRC activities. RD-Connect partners made important contributions to developing the IRDiRC “policies and guidelines”, which were adopted by the IRDiRC Executive Committee in 2013 and published online, and are providing input into the IRDiRC's “gap analysis” and “roadmap”, which are to be finalized in 2014.

Outreach: extending collaborations

To improve integration and reduce duplication, extensive efforts have been made during the first year of the project to connect with other groups engaged in related activities who were not part of the original RD-Connect consortium, and the response to this outreach has been highly encouraging. New collaborators include the Human Phenotype Ontology (Peter Robinson, Berlin), the PhenoTips and PhenomeCentral software tools (Mike Brudno, Toronto), the DECIPHER project (Helen Firth and Matt Hurles, Sanger), the Canadian Care4Rare IRDiRC project (Kym Boycott, Ottawa), the Genomics Coordination Center at Groningen (Morris Swertz, Groningen), the US ClinVar and ICCG initiatives (Heidi Rehm, Boston), and the Global Alliance for sharing of genomic and clinical data. Several of these groups have been offered associated partner status to formally recognise their collaboration with RD-Connect.

Integrated platform

RD-Connect is working towards an integrated platform in which data from omics experiments is combined with detailed phenotypic data and information on biomaterial availability and made accessible online through an interactive interface incorporating a range of bioinformatics tools. This work is led by Ivo Gut from the Centro Nacional de Analisis Genómico (CNAG) in Barcelona and key partners Alfonso Valencia (Centro Nacional de Investigaciones Oncológicas, Madrid), Paul Flicek and Justin Paschall (European Bioinformatics Institute (EBI), Hinxton) and Rachel Thompson (Newcastle University). Work in Year 1 has focused on putting the foundations of this platform in place, including ensuring interoperability with other systems and meeting the requirements of EURenOmics, Neuromics and additional IRDiRC projects generating omics data that will be linked with RD-Connect.

The cross-project “jamboree” held in July 2013 was a key milestone in this regard. “Use cases” provided by the associated projects formed the basis for discussions on the tools and resources that investigators require to facilitate their omics research, and cross-project working groups were created to take activities forward jointly. Concretely, this has enabled agreement to be reached on data sharing with Neuromics and EURenOmics, with both projects committing to archiving their data in the European Genome-phenome Archive (EGA) at the EBI and to permitting its further processing by RD-Connect and its ultimate release to the research community under an agreed timescale managed by the projects themselves. The first data transfer is envisaged to take place in January 2014.

Extensive engagement with ontology developers and the associated projects has ensured that the submitted omics data will be accompanied by standardised phenotypic descriptions using the Human Phenotype Ontology and that the RD-Connect system will be compatible with other clinical data collections using ontologies that have agreed to harmonise their terms through the International Consortium for Human Phenotype Terminologies (ICHPT), which will also enable data submitted to RD-Connect to participate in “matchmaking” services that allow researchers looking for similar/confirmatory cases to find others with a matching case.

The platform infrastructure and data repository have been designed and development using MongoDB has been initiated on a Virtual Machine server installed at CNAG. A first version of the alignment and analysis pipeline has been developed and an initial proposal of variant calling data

fields that should be stored in the database has been issued. Nanopublication schemes for gene-disease associations and for transcription start site annotation have been designed and improved and the modelling of genetic variant descriptions (including the development of sequence ontologies that can deal with different reference sequences) has been improved. A comprehensive project website (www.rd-connect.eu) has been developed and an electronic newsletter launched, and as a commitment to data sharing, open standards and interoperability, the RD-Connect consortium has signed up to the Global Alliance for Sharing of Genomic and Clinical Data.

Bioinformatics tools

Various suites of clinical bioinformatics tools to extract knowledge from high throughput experiments, clinical databases and biobanks are being developed within RD-Connect. This work is led by Christophe Béroud from Aix-Marseille University and key partners Peter-Bram 't Hoen and Marco Roos (Leiden University Medical Center), Jose Luis Oliveira and Pedro Lopes (University of Aveiro), Matthew Bellgard (Murdoch University, Perth, Australia) and André Blavier (Interactive Biosoftware, Rouen). As part of the development of a DNA Variant Analysis and Prioritisation Suite, the new version of UMD-Predictor®, based on the hg19 version of human genome and on Ensembl v71 gene annotations, has been developed into one optimised PostgreSQL database for all chromosomes. A website is now available with a user-friendly interface. Concomitantly, the new version of Human Splicing Finder (HSF) v3.0 will soon be released to allow predictions of the impact of mutations on splicing signals. In the development of an Integrative omics Analysis Suite, partners have started a showcase for integration of -omics data in the Huntington's Disease (HD) field. Currently this includes a strategy for network-based integration of model organism and human data, integration of microarray data from human HD and control brains, and NGS-based transcriptomics data from blood from HD patients and controls. As part of the development of a Clinico-Genomic Knowledge Discovery Suite, a pilot project has begun on the semantic integration of clinical (phenotype) and genomic (genotype) information from the Australian Skeletome and DMD registries. In parallel, much has been done on the enrichment of Diseasecard, an online rare diseases research portal, which is built on the COEUS Semantic Web application framework and makes an API available for developers.

Patient registries

Within the patient registries area, activities within the first year have primarily centred on identification and scoping of the full range of European and international registries with a rare disease focus and utility for research. This work is led by Domenica Taruscio from the Istituto Superiore di Sanità in Rome and key partners Christophe Béroud (Aix-Marseille University) and Manuel Posada (Instituto de Salud Carlos III). An extensive mapping exercise carried out jointly with the biobanking work package has resulted in a list of registries and biobanks that will now be surveyed to establish their research focus and invited to participate in RD-Connect activities. The detailed list will be transformed into an online database searchable by disease name, Orphacode, ICD10 number and OMIM code. Participating registries will have the opportunity to provide their own data describing their resource.

As the project progresses, RD-Connect will develop best practice recommendations for registries, and in Year 1 a Core Implementation Group (CIG) of advanced registries has been constituted to assist with this process. The CIG registries will both provide their own examples of best practice that

can be disseminated to newer registries, and also advise on the feasibility of implementing RD-Connect recommendations (e.g. relating to ontologies or patient identifiers), in some cases piloting implementation of recommendations in order to ensure that implementation is feasible and not too onerous.

In order to be able to link records from the same individual across multiple resources, RD-Connect plans to implement a globally unique identifier or GUID. Progress has been made in evaluating the various options already in use in other projects, including the NIH GRDR and Huntington's Disease systems, and a decision will be taken in 2014 based on practical and ethical considerations. However, the GUID will not be a prerequisite for entry of data into the system, in order to ensure that no data from contributing projects has to be excluded.

Biobanks

The goal of RD-Connect's biobank work is to improve access to rare disease biomaterial samples by developing a comprehensive and searchable online catalogue that enables human RD biomaterials and their linked data to be made and available to the scientific community, and to improve interoperability between existing biomaterial resources through the development of standards for sample collections and recommendations for the operational workflow of a biobank. This work is led by Lucia Monaco from Fondazione Telethon in Milan and key partners Jan-Eric Litton (Karolinska Institutet, Stockholm) and Kurt Zatloukal (Medical University of Graz). Within Year 1, a mapping exercise has been launched to ensure outreach to all biobanks holding biomaterials related to rare disease, and contact has been made in particular with biobanks associated with Neuromics and EURenOmics. Progress has been made towards development of the biobank catalogue database structure and on activities related to biobanking standards through gaining an overview of major existing standards and guidelines for biobanking and developing "Minimal Information Standards" for sample collections. The incorporation of linked data strategies will improve the opportunities for integrating common data elements across biobanks, registries and other resources and making them computer-accessible and searchable. A workflow describing the entire process of sample preparation, storage and distribution has been proposed and is planned to become part of the assessment process for new biobanks wishing to become associated with RD-Connect.

Ethical, legal and social issues

Proactively engaging with the ethical issues raised by omics experiments and patient data sharing is crucial to ensure RD-Connect has the opportunity to influence policy and promote expedient ethical frameworks for rare disease research. This work is led by Mats Hansson from the University of Uppsala and key partners Anna Kole (Eurordis) and Pauline McCormack (Newcastle University). Within Year 1 examples of such proactive engagement include evaluating recent proposals from the European Parliament's Committee of Civil Liberties, Justice and Home Affairs that will have severe repercussions for medical research using biobanks and registries. Three publications in international peer-reviewed journals have been produced on this topic. Further achievements include the drafting of a charter with principles and template for sharing and access to data, and the launch of a cross-project Rare Disease Patient and Ethics Council (RD-PEC) including participants from Neuromics and EURenOmics as a significant collaborative tool to address issues arising throughout the three projects.

With regard to patient and stakeholder engagement, work has begun on investigating participation strategies that might be employed to create mechanisms to allow individual patients and patient groups to influence the project's day to day activities. In order to structure patient input in project activities, 16 patient representatives have been invited to become members of a Patient Advisory Council (PAC), which has begun to participate in several aspects of the project's activities that ethical, legal and social questions. An interactive workshop on patient participation held during the January kick-off meeting resulted in a list of priorities to address, while a stakeholder conference held in Brussels in October aimed to ascertain stakeholder views and provide guidance on regulatory hurdles impeding rare disease research. This conference provided a basis for further work on the identification of regulatory hurdles in research.

Impact

In its first year of operation, RD-Connect was presented at 93 national and international conferences and workshops, including the IRDiRC conference in Dublin, had 10,000 visitors to the website from 122 countries, 351 subscribers to the newsletter, and was cited or acknowledged in 9 peer-reviewed publications, including Nature Reviews Genetics, The Lancet Oncology and the European Journal of Human Genetics. RD-Connect has generated interest with industry stakeholders e.g. Genzyme, and has led to several grant applications with both national and international funders.