Year 5 Summary

Overview

RD-Connect is a global research and infrastructure resource for rare diseases (RD). Set up to overcome the siloing, fragmentation and inaccessibility of datasets from different projects, it links omics data with phenotypic data and information in registries and biobanks at both an individual-patient and whole-cohort level to enable researchers to analyse their own data and gain a complete view of their disease and patient population of interest. Data shared through RD-Connect is accessible beyond the usual institutional and national boundaries and researchers across the world can benefit from the opportunity to work with others with an interest in the same field, relate human phenotypes to a particular gene or pathway of interest, pool data to create larger cohorts, find confirmatory cases, and access samples for further study.

The project's objectives are to develop:

- an integrated platform to host and analyse data from RD omics research projects
- clinical bioinformatics tools for analysis and integration of molecular and clinical data to discover new disease genes, pathways and therapeutic targets in RD
- common standards and data elements for RD patient registries
- common standards and a sample-level catalogue for RD biobanks
- best ethical practices and recommendations for a regulatory framework for linking medical and personal data related to RD

RD-Connect was launched on 1 November 2012 and this report focuses on the project's fifth year of operation, from November 2016 to October 2017.

Year 5 highlights

This year has seen extensive development in the functionality of the RD-Connect tools and services, consolidation of RD-Connect's reputation in the European rare disease community (as evidenced by invitations for RD-Connect resources to play a central role in major new initiatives), and substantial progress towards future sustainability of the RD-Connect assets, which are now mature resources for RD researchers to use in their everyday practice. At the project's annual meeting in Berlin, the external Scientific Advisory Board commended the project partners for having built a rare disease community integrating scientists from different fields, clinicians and patients. This also reflects the comments from project partners during sustainability discussions: that RD-Connect is more than the sum of its parts and that its ethos of collaboration and data sharing has significantly contributed to greater community recognition of the value of data sharing and reuse.

The successful partnership with NeurOmics (coordinated by O Riess, University of Tübingen) and EURenOmics (coordinated by F Schaefer, University of Heidelberg) continued for the fifth and final year of the partner projects. This included the transfer of a large number of NGS datasets and associated phenotypic profiles into the RD-Connect platform and the European Genome-Phenome Archive (EGA), support for gene discovery and other -omics based research, joint teaching and training activities, and a final joint annual meeting with an outreach day that the three projects held together in Berlin in May 2017. The achievements of this highly productive collaboration over five years are summarised in a joint publication (Lochmüller, Badowska et al, EJHG 2018, in press) and have also resulted in fruitful new scientific links between the partners, including successful grant applications and collaborations within the framework of the new European Reference Networks (ERNs). The development of close relationships with several ERNs is another highlight of the year: these important international initiatives will have a major impact on the RD landscape in Europe and many of them are looking to RD-Connect to provide not only the infrastructure for omics data sharing and analysis but also the expertise in data stewardship including ontologies and FAIR data principles.

In addition to ERN-RND and ERKNet, the ERNs for rare neurological and kidney disorders which are led by the NeurOmics and EURenOmics coordinating institutions and thus become natural extensions of the close collaborations developed there, the ERNs for rare bone, eye, endocrine, haematological and immunodeficiency disorders have collaborated with RD-Connect in 2017. The coming year sees the launch of Solve-RD, a major new Horizon 2020 project in which the RD-Connect platform will play a central role, and here the ERNs for neurological disorders, neuromuscular disorders, intellectual disability and genetic tumour risk syndromes will submit data to RD-Connect. This is an excellent example of the way in which the value of European funding can be maximised: the reuse of infrastructure developed in one project by its inclusion in new initiatives enables the development of a virtuous circle where previous investment is capitalised on. This is also the case in the upcoming E-Rare call in 2018, which mandates the use of infrastructure like RD-Connect for omics sharing and analysis, and must also be a key part of the infrastructural component of the new European Joint Programme for rare disease.

RD-Connect operates within the context of the International Rare Diseases Research Consortium (IRDiRC) as one of the EU's flagship projects under this initiative. The therapeutic side of the initial IRDiRC goals – 200 new therapies by 2020 – was reached sooner than envisaged, at the end of 2016 (PMID 28796411), while the diagnostic side – diagnosis of most rare diseases by 2020 – was acknowledged to be more challenging than foreseen. This led to a broad consultation on objectives and strategy for rare disease research for the next decade, and RD-Connect partners have been major contributors to this process. The new objectives for 2017 were developed by the IRDiRC scientific and executive committees (members include RD-Connect partners H Dawkins, L Monaco and D Taruscio and RD-Connect coordinator H Lochmüller as chair of the Interdisciplinary Science Committee) and were discussed by the wider community during the IRDiRC conference in Paris in February 2017. A publication (PMID 28796445) formally set out the new vision of enabling all people living with a rare disease to receive an accurate diagnosis, care, and available therapy within one year of coming to medical attention, and RD-Connect's resources are expected to be major contributors to achievement of this goal. Three additional IRDiRC publications were led by RD-Connect coordinator H Lochmüller on IRDiRC's key assets, achievements and impact, including policies and guidelines, recognized resources, diagnosis and data sharing (PMID 29158551, 27782107, 28475856).

Integrated platform

The integrated platform developed by RD-Connect is hosted by the Centro Nacional de Análisis Genómico (CNAG-CRG) in Barcelona and brings together omics data and clinical data from participating projects with tools and services to analyse this data online. The central portal available at http://platform.rd-connect.eu provides access to the genome-phenome analysis platform including the genomic analysis interface and the PhenoTips database that stores human phenotype ontology-coded phenotypic profiles (PMID: 27899602) for the individuals whose genomic data is accessible in the system, and the catalogue of biobanks and patient registries.

Genome-phenome analysis platform

The RD-Connect Genome-Phenome Analysis Platform under the leadership of I Gut and S Beltran is one of the project's flagship tools, and has developed into a uniquely powerful and user-friendly resource that enables the contributing researchers themselves to analyse the data they submit. To date, 2078 whole exomes, 322 gene panels, and 74 whole genome datasets have been made accessible through the platform, with additional datasets becoming available after expiry of the embargo period that gives the submitting centres priority access to their own data. While the majority of datasets are related to rare neuromuscular and rare neurodegenerative diseases, 2017 saw an increasing number of submissions from new projects and partners for leukodystrophies, brain malformations, intellectual disability, kidney, immunological, metabolic and mitochondrial diseases, and from undiagnosed patient projects, which will become available for sharing

and matching in 2018. Importantly, several new projects have chosen RD-Connect as their primary analysis platform. After a pilot submission in 2017, an international consortium focusing exclusively on titinopathies will submit up to 5000 NGS datasets to RD-Connect in order to explore the range of pathogenic and benign variation in the *TTN* gene. This demonstrates the value to researchers of being able to explore large numbers of individual-level genomic datasets linked with phenotype even when only a single gene is in focus. Through the new Solve-RD project to be launched in January 2018, up to 19,000 undiagnosed exomes from European Reference Networks are earmarked for deposition and re-analysis through RD-Connect.

The power the analysis interface puts in the hands of the end-user is its strongest feature, and its functionality and user-friendliness has been further improved this year through integration of new features such as predefined analysis filters, updated population frequencies and customised gene lists for filtering, real-time design of virtual gene panels from HPO and OMIM, the deployment of Exomiser, integration of Human Splicing Finder, addition of ClinVar information and filters by runs of homozygosity. Work on multi-omics targets the implementation of RNAseq analysis and visualisation, integrated with DNA sequencing and phenotyping data on an individual sample level in the RD-Connect platform which will become available to users in 2018.

This year has seen the first large use case of the RD-Connect genome-phenome analysis platform to solve cases which have not previously been investigated in a similar platform, specifically 900 exomes that have been sequenced as part of the BBMRI-LPC call, which started to become available on the platform in March 2017. Preliminary analysis of approximately half of these cases by the clinical genomics specialist at CNAG has resulted in molecular diagnoses of between 33-50% of cases (pending confirmation from collaborators), which is in the range expected for studies of this nature. This proves that the platform is functioning well, and indeed some of these cases are already being written up into papers for submission.

Biobanks and patient registries

Patient registry and biobanking activities were led by D Taruscio (Rome) and L Monaco (Milan) and focused on integration, quality assessment and interoperability. To promote visibility, the online catalogue of biobanks and registries previously known as ID-Cards was formally named the RD-Connect Registry & Biobank Finder. This resource currently contains detailed descriptions of 360 RD registries and this year each registry has been categorised according to its alignment with the 24 newly established ERNs. The integration of EuroBioBank continued with more biobanks participating in the Finder.

The RD-Connect/EuroBioBank sample catalogue has been further developed with links to the RD-Connect genomics platform and the Registry & Biobank Finder. Work continued towards the deployment of the Sample Catalogue on the RD-Connect server hosted by CNAG, leading to the release of a Docker version of the RD-Connect catalogue software. Upgrades of the catalogue software also took place to ensure its capability to support planned integration of BBMRI-ERIC Negotiator for sample workflow. The mapping of the first set of biobanks was completed and data upload is currently underway. The RD-Connect Panel for Biobank Assessment is now operational and has successfully completed an evaluation leading to integration in the RD-Connect EuroBioBank Network. This new addition brought the number of members to 25 RD biobanks. During this reporting period, RD-Connect EuroBioBank also successfully disseminated DNA samples to omics projects and provided banking services in the 2016 BBMRI-LPC WES Call, handling altogether nearly 900 DNA samples.

The data linkage plan developed in collaboration with ELIXIR continued to work with registries and biobanks to promote and support Findable, Accessible, Interoperable and Reusable (FAIR) data as promoted by M Roos and P 't Hoen (Leiden University). Progress is being made towards an architecture for registry software providers to become FAIR data generating tools, allowing registries to implement the FAIR Data Point API and expose data in a linkable format. Several conferences, workshops, courses and tutorial have been delivered,

and training events and hands-on data workshops have been aligned with ELIXIR, in particular through its specific work package on rare diseases led by RD-Connect partners as part of its EXCELERATE project. Extensive advice and support has been provided to ERNs wishing to collect clinical data in a FAIR manner, including participation in workshops to update and extend the human phenotype ontology and Orphanet nomenclature (ERN-EYE) and support in developing FAIR registries (ERN-EYE, ERN-BOND, ENDOERN).

Bioinformatics tools

In addition to the central resources offered through the platform interface, RD-Connect partners have under the leadership of C Beroud (Marseille) developed a number of bioinformatics tools to assist researchers in omics analysis and therapeutic target identification (PMID 27599893). Tools for the annotation and analysis of sequence variants, particularly ALFA and VarAFT, have been further developed and new versions released. New bioinformatic tools to simplify the design of new therapeutic molecules able to induce trans-splicing, and to evaluate the impact of therapeutic modifications on gene expression have been developed, and tools for pharmacogenomic analysis and facial phenotypes to better define disease classification were further developed for rare disease applications. In parallel, all partners have promoted the RD-Connect bioinformatics tools through training, industry activities and scientific publications.

Ethical, legal and social issues and patient involvement

Ethical, legal and social issues in genomics and biobanking were explored by RD-Connect partners under the leadership of M Hansson (Uppsala). The newly developed and approved code of practice, including access rules and user adherence agreement, were deployed in the RD-Connect platform and are in everyday use. RD-Connect partners contribute to the development of the Code of Conduct for the new EC General Data Protection Regulation (GDPR), led by BBMRI ERIC (JE Litton, PMID: 28128265). This new European regulation comes into effect in May 2018 and thanks to active input from RD-Connect partners and others in the RD community it will be possible for RD research to comply with the regulation. As the regulation is implemented into national law across the member states, the situation will continue to be monitored to ensure that data sharing and access for research is not hindered and to update RD-Connect resources as needed for continued compliance.

RD-Connect partner EURORDIS (V Bros-Facer) leads the input of patient representatives for all RD-Connect activities, with members of the Patient Advisory Committee (PAC) actively engaged in the scientific work as well as capacity building and dissemination of the project's outputs to the wider rare disease patient community. In addition, EURORDIS supports the European Patient Advocacy Groups involved in the different ERNs in order to ensure a direct link and communication with the relevant European and international projects, networks and consortia including RD-Connect.

Impact, outreach and extending collaborations

The key tasks for this fifth year were to consolidate RD-Connect's position in the community as the leading resource for access to biosamples and patient registries and for analysis of genome-phenome datasets for diagnosis and gene discovery, to plan for and secure funding sources to maintain assets beyond the end of the original FP7 grant, and to increase collaborations with new data submitters to increase the flow of data into the platform. Significant progress has been made towards securing the sustainability of key assets, particularly the RD-Connect genome-phenome analysis platform in Barcelona through an important role in the Solve-RD project (EU Horizon 2020, coordinated by Olaf Riess, University of Tübingen) and through the inclusion in pillar 2, in close collaboration with Orphanet (directed by Ana Rath, Paris), of the future European Joint Cofund for Rare Diseases. Collaborations with the new European Reference Networks (inaugurated in March 2017) were strengthened (ERNs EURO-NMD, RND and ERK-NET) and newly developed (ERNs EYE, GENTURIS, ITHAKA, RITA and ENDO).

RD-Connect has been presented at over 70 scientific meetings in 2017 and the project's own dedicated annual meetings attract over 100 people each year. The RD-Connect website has 20,000 visitors annually from 146 countries and the newsletter is sent out monthly to 1331 subscribers. The new YouTube explanatory video is available in 7 languages and has had close to 2,000 views within 6 months. The genomephenome analysis platform is used by several hundred researchers in the rare disease community and currently contains nearly 3000 individual geno:pheno datasets (December 2017) with 7,000 expected by the end of 2018 and over 20,000 expected in the coming years through the Solve-RD project and other collaborations. The platform and project have been referenced or acknowledged in 140 peer-reviewed publications. The platform has collaborations with 18 sequencing projects to date and has begun discussions with all 24 European Reference Networks to host their research sequencing data for analysis.

An important development for RD-Connect in 2017 has been the preparation of a new funding mechanism for rare disease research in Europe, the European Joint Programme for Rare Diseases (EJP-RD). This is a cofunding initiative between the European Commission and the member states, and it includes not only provision for open calls for specific rare disease research projects modelled on the current E-Rare programme, but also infrastructural-type funding for coordinated access to rare disease data and resources. Rather than having each new research project reinvent the wheel in terms of data management, it is necessary to provide a new infrastructure at European scale that can support RD researchers in standardising, structuring and opening up their own datasets for reuse and accessing data generated by others. It intends to build on Europe's existing bespoke services for RD data and biosamples (Orphanet, RD-Connect, EuroBioBank) and the concepts of European Open Science Cloud to create a sustainable, FAIR and open infrastructure for RD data in Europe that will transform the RD research community's ability to analyse, reuse and exploit the rich reserves of data generated by all projects and all stakeholders. The goal is to provide a user-friendly interface to integrated services, including real-time online access to integrated omics and phenotypic data of a large and growing number of RD patients for data mining and discovery and access to RD biosamples associated with high-quality, searchable metadata for research purposes. Thanks to the active participation of RD-Connect partners at every stage in the planning of this new initiative, RD-Connect resources are well positioned for inclusion, while at the same time through the presence of RD-Connect the EJP-RD benefits from a proven solution for omics data analysis and data and sample sharing, thus substantially de-risking the initiative. Further progress on this is expected prior to the submission deadline of March 2018.

In its fifth year of operation, RD-Connect has matured into a highly functional, solid and trusted resource and a key infrastructure for rare disease research in Europe which is increasingly used by the research and clinical community across disease areas. Efforts in the final year of the FP7 funding will aim to develop further functionality and user-friendliness, exponential scale-up of datasets and capacity, and full integration in future research initiatives and infrastructures, while maintaining the RD-Connect ethos and community.