Rare Disease Research Roadmap: Navigating the bioinformatics and translational challenges for improved patient health outcomes

Matthew I. Bellgard, Mark W. Sleeman, Felix D. Guerrero, Sue Fletcher, Gareth Baynam, Jack Goldblatt, Yaffa Rubinstein, Callum Bell, Stephen Groft, Roberto Barrero, Alan H. Bittles, Stephen D. Wilton, Christopher E. Mason, Tarun Weeramanthri

Centre for Comparative Genomics, Murdoch University, Perth, Western Australia
National Centre for Genome Resources, Sante Fe, New Mexico, USA
Department of Physiology, Biochemistry & Molecular Biology, Monash University, Victoria, Australia
US Department of Agriculture, Agricultural Research Service, Knipling-Bushland, US Livestock Insects Research Laboratory, Kerrville, Texas, USA
Western Australian Neuroscience Research Institute & Centre for Neuromuscular and Neurological Disorders, University of Western Australia, Nedlands, Western Australia
Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, Western Australia
Office of Population Health Genomics, Public Health & Clinical Services Division, Department of Health, Western Australia
School of Paediatrics and Child Health, University of Western Australia, Perth, Western Australia
Institute for Immunology and Infectious Diseases, Murdoch University, Perth, Western Australia
Genetic Services & Familial Cancer Program of Western Australia, Subiaco, Western Australia
Office of Rare Disease Research, National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, Maryland, USA
Community Genetics, School of Medical Sciences, Edith Cowan University, Western Australia
Department of Physiology & Biophysics, Weill Cornell Medical College, New York, New York, USA
The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York USA
Feil Family Brain & Mind Research Institute, Weill Cornell College, New York, New York, USA
Public Health & Clinical Services Division, Western Australia Department of Health, Perth, Western Australia

Available online 19 August 2014

Abbreviations: CORD, Canadian Organisation for Rare Disease; EURORDIS, FP7, EU Framework 7 Programme; ICD, International Classification of Diseases; IRDiRC, International Rare Disease Research Consortium; NCATS, National Center for Advancing Translational Sciences; NORD, National Organisation for Rare Diseases

*Corresponding author at: Centre for Comparative Genomics, Murdoch University, Western Australia. Tel.: +61 8 9360 6088; fax: +61 8 9360 7238.
E-mail address: mbellgard@ccg.murdoch.edu.au (M.I. Bellgard).

http://dx.doi.org/10.1016/j.hlpt.2014.08.007
2211-8837/© 2014 Fellowship of Postgraduate Medicine. Published by Elsevier Ltd. All rights reserved.
KEYWORDS
Rare disease; Bioinformatics; Translational research

Abstract
Rare disease registries have now been recognized as a global priority for progress both in monitoring and documenting the natural course, and preventing and treating rare diseases. However, a disease registry is only one element of rare disease translational research. Here, we outline what we believe are ten key components in comprehensive rare disease translational research and describe critical relationships between them. These components are: (i) client-practitioner partnerships; (ii) disease registries; (iii) biobanks; (iv) genomics and other -omics platforms; (v) community-based and population-wide studies; (vi) bioinformatics and high performance computing; (vii) interactions with pharma to facilitate drug discovery; (viii) personalized treatments based on genotype-phenotype correlations; (ix) eHealth and a whole of life record; and (x) regulatory frameworks, particularly with regard to specimen and data sharing, and the return of results. Each component has its own inherent complexity, but if effectively integrated they will provide a comprehensive approach to the future management of rare diseases, and aid health care providers in delivering services to individuals affected with rare diseases. We demonstrate that navigation through the roadmap can provide relevant health stakeholders with a blueprint to understand the challenges and barriers which need to be overcome within and across the constituent components. The rare disease roadmap will assist decision-making at all health stakeholder levels and enable the seamless integration of new knowledge, standard operating procedures and the implementation of best practice.

© 2014 Fellowship of Postgraduate Medicine. Published by Elsevier Ltd. All rights reserved.

Introduction
Our understanding and appreciation of the complexity of the genetic basis of human disease, including rare diseases, is increasing rapidly. There are overlapping operational definitions of RDs and, for example, in Europe, an estimated prevalence of five or fewer affected persons in 10,000 is used. Using this definition, some 7-10% of people have one of the approximately 7000 rare diseases identified to date, 80% of which are genetic in origin [1].

The low incidence of each individual RD provides inherent challenges in ascertaining epidemiological data and recruiting sufficiently large cohorts for clinical trials or translational research. To address this shortcoming, there is for a need to establish global inclusive RD registries and biobanks [2-5] that facilitate the coordinated acquisition of fundamental disease information and research specimens to: (i) assess the health and economic impact of rare disorders individually or collectively; (ii) devise best practice disease management strategies; (iii) engage complementary national/international expertise; (iv) build sustainable academic, government and/or industry partnerships; and (v) develop large sustainable resources for translational research. Challenges and barriers that need to be overcome to facilitate these approaches include the international harmonization of informed consent, and specimen- and data-sharing practices.

In 2010, at the workshop, ‘Advancing Rare Disease Research: The Intersection of Patient Registries, Biospecimen Repositories, and Clinical Data’, the Office of Rare Disease Research (ORDR) at the National Center for Advancing Translational Sciences (NCATS)/NIH initiated a movement to create a global RD patient registry [3]. The workshop produced a set of recommendations, which are being implemented through the NIH/NCATS GRDRSM Program (The Global Rare Diseases Patient Registry Data Repository-GRDR - program - see https://grdr.ncats.nih.gov/).

More recently, EURORDIS, NORD and CORD issued a Joint Declaration of 10 Key Principles for Rare Disease Patient Registries [6], building on the European policy on rare disease. This initiative is an important strategic step in enabling the diagnosis and treatment of patients with an RD. Additional considerations that can complement the principles of the Declaration include how registries interact with orphan drug development [7]; personalized and therapeutic interventions [8]; RD diagnostics and novel phenotyping strategies [9]; population-wide association studies [10]; rapidly evolving integrated bioinformatics advances [11]; public policy [12]; and international standardization [13]. We argue that, to make substantial advances in the translation of RD research to clinical practice, all of these activities must coalesce to avoid duplication, and that key conceptual and methodological advances need to be effectively shared and harmonized to promote rapid adaption to the ever-changing RD landscape.

In this paper we identify ten key components necessary for successful RD translational research and development. The ten components are: (i) client-practitioner partnerships; (ii) disease registries; (iii) biobanks; (iv) genomics and other -omics platforms; (v) community-based and population-wide studies; (vi) bioinformatics and high performance computing; (vii) interactions with pharma to facilitate drug discovery; (viii) personalized treatments based on genotype-phenotype correlations; (ix) eHealth and a whole of life record; and (x) regulatory frameworks, particularly with regard to specimen- and data-sharing and the return of results. The implementation of each component is extremely complex and all have associated challenges (or noise, to use an engineering analysis) that are

technical, logistical, socio-political, ethical, legal, jurisdic-
tional and/or economic in nature, or workforce-related. For
instance, a poorly designed next generation DNA sequencing
experiment or incorrect sample extraction from a biobank
will introduce technical noise into the research on a given
RD [14].

Another example of such a barrier is the limited access to
particular RD patient samples from a single organization/
jurisdiction that might be overcome via access to large
internationally accessible biobanks for the same disease.
The scale of the barrier and its impact can vary within each
component and can be mitigated through international
harmonization of processes and our improved understanding
of biological systems, biotechnological advances, improved
partnerships, the introduction of internationally accepted
best practice, use of refined bioinformatics analysis method-
s, effective use of Internet technologies and/or high
performance computing. In turn, the adoption of these
strategies should facilitate translational research and ulti-
mately result in improved patient treatment protocols.

The challenge, and the opportunity we envisage, is to
understand how these different components interoperate
into a RD Roadmap and to identify and characterize the
potential interfaces within and across each component. We
discuss how to navigate through this RD Roadmap and
identify the challenges faced with each component, as well
as discussing the alternative paths that might be taken to
improve the care of individuals living with a RD.

Ten key components for translational RD
research

Ten components regarded as core to translational RD
research are outlined in the subsections below and summar-
ized in Figure 1, with each component considered under
three subheadings; Overview, Challenges and Ways Forward.

Patient-health professional partnerships

Overview

Individuals living with RDs engage with health professionals
in order to achieve a diagnosis and receive timely and
appropriate management. These interactions optimally
occur by marrying skill-sets to patient need in an environ-
ment of trust and open communication, which requires the
ability to describe the disease phenotype in a manner that is
captured in a disease- and patient-centric knowledge base.
This approach is, however, hampered by the lack of scalable
and robust RD knowledge management approaches and
systems that are suitable for health professionals working
under time-pressures in clinical settings. Such systems need
to mirror clinical workflows and be time-efficient, including
the requirement for the single entry and validation of
information, rather than multiple data entries by different
professionals.

Challenges

The major challenge with this component is often the lack of
a suitable database/registry that facilitates the capture of
and accessibility to pertinent patient information, and lack
of standardization among existing registries and databases to
facilitate data-sharing and the exchange of knowledge. A
situation of this type might arise for a number of reasons:
(a) limited health professional time available to extract or
enter relevant patient data; (b) delays in obtaining consent
to obtain personal and family information; (c) limited and
restrictive information management systems that either are
not user-friendly or are incapable of capturing all of the data
elements relevant to effective diagnosis and best practice;
d) inability to extract pertinent legacy information from
handwritten clinical notes or internal spreadsheets that

Figure 1 Rare disease roadmap: key components and relationships between them.
become part of de facto laboratory standards; and (e) system inability to capture a variety of data from established and evolving diagnostic technologies.

Ways forward

Mitigation strategies that are currently underway include: (a) the establishment of data elements that are common to all RDs [13]; (b) development of Internet-based laboratory information management systems that are scalable, secure and operate effectively in a distributed environment [4] [15]; (c) capture of deep phenotypic information in a structured (ontological) manner and strategies to build these structures dynamically [16]; (d) migration and curation of legacy clinical information systems and filing systems containing important RD patient information into new information management systems; and (e) interfaces that mirror clinical workflow to promote timely data capture [17].

If properly conducted, and concomitant with active research leading to iterative updates of common data elements and ontologies, knowledge management systems should have the capacity to dynamically and seamlessly incorporate these updates to improve patient care.

Disease registries

Overview

Rare disease registries represent a core element for effective RD translational research. Both the NIH workshop and the EURORDIS-NORD-CORD Joint Declaration state that patient RD Registries should have a wide geographical if not global scope. In addition, the Registries should focus on a disease or group of diseases, they should show interoperability between RD patients, use Common Data Elements, and be linked with biobank data [2,6]. Equally importantly, public-private partnerships also should be integrated into this effort, as the pharmaceutical industry is actively engaged in rare/orphan disease drug development [7]. International patient RD registries are perceived as critical, particularly to the pharmaceutical industry, in the provision of large, easily accessible cohorts of affected individuals. Thus there is an impetus for improved coordination between national- and regional-based registries so that they can more efficiently feed into international registries.

Challenges

Disease registries can suffer from deficient architectural design and inappropriate software technology choices. Integrating legacy systems or new registries must adhere to international standards for common data elements and ontologies, and their scalability. Numerous patient advocate organizations in each country drive research efforts on specific RDs, and the databases often have to vie for limited public and private funds. Unfortunately, older established registries frequently contain copious data that have not been curated/validated and so are likely to be discarded or remain unused. The key barrier is how these resources can dynamically capture knowledge about diseases and thus aid in patient management and translational research.

Ways forward

Registries capture data and thus become an information-rich resource for relevant stakeholders. The identification of common data elements and ontologies plays an important role in future-proofing accurate patient disease information for the range of RDs under investigation [18], and an RD Registry Checklist has been proposed [15]. There is evidence that pharmaceutical industries are recognizing efficiencies through working together [19], as well as interacting with academia and governments agencies [20]. International and national disease registries have become the vehicle within which to initiate private-public dialogues (e.g. http://www.euro-hd.net/html/registry, http://www.treat-nmd.eu/). National lead RD patient advocate groups are essential for communication with federal and state-based governments, rather than individual disease advocate groups presenting separately. To ensure sustainability the establishment of clinical disease registries must have government backing, be endorsed and driven by peak RD advocate groups, attract ongoing resourcing and professional and community engagement, and allow for curation.

Biobanks

Overview

A plethora of biobank samples are stored in freezers worldwide and, while many of these biobanks are recognized and accredited, others are more boutique in nature. Like the registries outlined above, the global push for robust, sustainable biorepositories is urgent and the practices involved have come under increased scrutiny. Internationally agreed standards are needed to ensure high fidelity in biosample extraction, preparation, long-term storage, and recordkeeping. Errors in any of these steps can destroy the utility of valuable samples and data, thus wasting time and resources [14].

Challenges and barriers

Biobanks can variously contain samples for genomic, epigenetic, transcriptomic, proteomic, or metabolomic studies. For any biobank, problems can arise due to human errors in sampling, specimen storage, flawed recordkeeping, and inappropriate use of samples.

A critical problem is the lack of internationally adopted standard operating procedures (SOPs) and so many organizations, both private and public, have their own internally designed SOPs. With increasing knowledge of the effects that sample extraction/preparation/environmental factors have on gene, protein, and metabolite expression, there is a need for well-designed and updated SOPs that reflect best practices in sample handling and storage using current technologies. Questions that arise include: (i) reliability of legacy biobanks; (ii) type of consent obtained at time of collection; and (iii) procedures for the return of results.
Other challenges include limiting access to particular biorepositories, limited residual sample amounts, intellectual property, or privacy/competition issues that can inhibit collaborative value-adding studies.

**Ways forward**

Consistent, internationally acceptable SOPs are important and there is a push to develop this approach world-wide [21]. International harmonization of these processes is vital, and technology is greatly improving our understanding of how samples should be managed in biobanks [14,22]. See [23] for a review on the current trends in biobanking for rare diseases.

**Genomics and other -omics platforms**

**Overview**

Quantum advances are continually being made in the genomics, transcriptomics, proteomics and metabolomics platform technologies, providing opportunities for significant advances in RD research. For example, whole genome sequencing of patients with an RD, compared to genomes from unaffected individuals, has facilitated the discovery of genetic changes underlying many RDs [24]. Integrating the various -omics platforms will make a significant contribution to improved personalized patient care [25].

**Challenges**

Developments in each of the -omics platforms are occurring rapidly and their applicability to RD research is evolving. But as more sequencing platforms have become available to RD researchers, problems with data interpretation have increased. Each sequencing assembly and annotation approach must be optimized, and to blend datasets arising from different approaches requires the development of new bioinformatics tools, e.g. for the detection of a disease-causing gene mutation [26].

In metabolomic studies, for instance, there are recognized issues associated with specific platforms, e.g. NMR versus mass spectrometry, that can have a dramatic impact on sensitivity versus specificity [27]. There will also be downstream noise for any comparative analysis between clinical laboratories treating patients with similar phenotypes or therapeutic tracking. Noise can be introduced when incorporating data generated from legacy sequencing technologies in the absence of descriptions of the experimental design. Hence, there are various ways in which noise can enter the system when working with samples targeted for genomic, transcriptomic, proteomic and metabolomic analysis.

**Ways forward**

Robust Laboratory Information Management Systems (LIMS) are essential for each of the platforms (e.g. an initiative for developing a LIMS for metabolomics studies has commenced in Australia, https://bitbucket.org/ccgmurdoch/mastr-ms/overview). Metadata documentation of experimental design and analysis audit trails must be captured to enable informed decision-making. Bioinformatics tools and systems are central to all of the -omics platforms, not only to provide solutions to data processing, quality control, and analysis but also for data integration.

**Population-wide studies**

**Overview**

Genome-wide association studies (GWAS) aim to both identify loci (genetic markers) associated with specific diseases and to locate candidate gene(s) that might become diagnostic and therapeutic targets. With the advances in next generation sequencing, it is now possible to cost-effectively screen large cohorts with large numbers (over a million) of markers [28].

**Challenges**

While GWAS aim to identify common variants, it is recognized that within common diseases rarer variants may have significant individual effects [10]. In an RD context, the challenge for GWAS is to acquire large cohorts which have the capacity to identify functional rare variants that have a minor allele frequency. There are numerous avenues for noise in population-wide association studies for RDs, and well-documented, inherent problems in GWAS studies include: cohort size, population structure and stratification, bioinformatics/biostatistical analysis [29], consideration of environmental factors [30], issues surrounding biobanks (outlined above) and epistasis (gene interaction/phenotype effects) [31].

**Ways forward**

Large patient cohorts are unlikely in an RD context unless approached on a global scale, including a more detailed understanding of population structure in RD [32]. Many rare diseases have a high incidence or are even unique to specific communities and sub-communities, having arisen as founder mutations in past generations [33]. A thorough working knowledge of population genetic structure and sub-structure is therefore paramount, particularly in the many global populations in which endogamous and consanguineous marriage is strongly favoured and rare diseases are significantly more likely to be expressed [34]. It is unfortunate that the genomic effects of population stratification often have been ignored, e.g. with a notable failure to account for caste differences in the design of case-control and cohort studies in India [35]. The 1000 Genomes Project will identify variants with a frequency of ≥ 1% [36,37] and identify rarer disease susceptible variants, and exome or whole genome studies on patients may help to circumvent the noise inherent in population-wide association studies.
Bioinformatics and high performance computing

Overview

Bioinformatics is involved in devising strategies for the management of data, its analysis and integration with tools that enable rapid scientific discovery, and informed decision-making. Bioinformatics issues relate to how patient data are captured and analysed (data standards, experimental design and analysis protocols), diseases are classified (ontologies), data are shared (interoperability, security, privacy), and are managed (appropriate design and implementation of information management systems). In an RD context, bioinformatics solutions must scale to work across laboratories and jurisdictional boundaries, be flexible in the transfer of analysis/ systems/tools from one RD to another, and be modular to integrate new knowledge.

Challenges

There are numerous avenues for noise in bioinformatics. Data currently captured for RDs are managed in many quite disparate formats; e.g. laboratory notebooks, spreadsheets and in-house databases containing important phenotypic and epidemiological data, and maintained independently by scientists, clinicians and health workers. They are also often located in geographically dispersed locations, such as offices, institutions, archival repositories and hospitals. This challenge of non-exchangeable data formats is not specific to RD research [38], and there is always the potential for groups and individuals to adopt non-scalable data management practices, to become dependent on restrictive computing resources for storing critical data, and to make poor choices in software development design [15]. In addition, there is a world-wide skill shortage for bioinformaticians to ensure that RD research is coordinated globally.

The lack of information systems' standards to implement a flexible and generic bioinformatics framework that can cope with the ever-evolving advances in technologies, algorithms, data formats and high performance cluster (HPC) infrastructure, while enabling personalized analysis strategies, may pose a challenge to most groups conducting research in the RD domain and other biomedical areas. To minimize false-negatives and maximize the sensitivity to identify sequence variants associated with human diseases, the bioinformatics framework should be modular, user-friendly and highly customizable to utilize the most comprehensive repertoire of computational workflows and tools [11,26].

Ways forward

From a strategic perspective there are a number of international initiatives attempting to address at least some of the bioinformatics bottlenecks in RD research. The European Union has endorsed a Framework 7 project, RD Connect, which is an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for RD research (http://rd-connect.eu). There are 27 full partners and 17 associate partners world-wide with guiding principles in security, open source, scalability, ongoing curation, data standards and shared knowledge inter-communicated in an open and collaborative environment. It is possible to make collective decisions on topics such as data management, the types of data that must be captured from each of the RD Roadmap components, data longevity, and where it should be stored and integrated. It is also possible to devise systems that enable non-computing experts to seamlessly access supercomputing, cloud computing and large-scale storage infrastructure world-wide for both storage and computation while addressing security concerns [11].

The Yabi web-based bioinformatics analytical system allows the design and deployment of a flexible and generic bioinformatics framework for specialized data analyses [11]. The modular design and architecture of Yabi is scalable, highly configurable, user-friendly, secure and open source, which facilitates seamless and transparent access to heterogeneous HPC environments. Yabi provides an analysis workflow environment that can create and reuse workflows as well as manage large amounts of both raw and processed data in a secure and flexible way across geographically distributed computing resources. These characteristics make Yabi an attractive system to globally coordinate bioinformatics efforts in RDs.

Drug discovery

Overview

The development of safe and effective therapeutics for the treatment of RDs has faced multiple challenges including, but not limited to, difficulties in attracting funding for research, recruiting and design for clinical trials, logistics for manufacture, and regulatory hurdles with national and international agencies. Acknowledging these issues, the US Orphan Drug Act was drafted in 1983 and in 1999 the EU adopted the Orphan Drug Regulation. The aim of the US legislation was to increase the investment of the pharmaceutical/biotechnology industry in the development of new diagnostics, therapeutics, and preventive practice. Prior to its enactment only ten products were approved to treat RDs in the United States. However, by 2008 over 2000 products had obtained orphan drug designation [39]. There have also been significant knock-on effects leading to increases in the number of clinical trials, which in turn has spurred innovation in novel drug technologies and led to the identification of new disease types. This has, however, not been without issues, such as the prohibitive costs for drugs/therapies for rare or orphan diseases.

Many pharmaceutical and biotechnology companies are now engaged with government, advocacy groups and academic researchers to promote RD research and product development. But disconnects still exist, primarily due to a lack of coordinated focus. Despite these shortcomings, optimism is high in this post-genomic era with respect to the development of innovative medical approaches, especially for some of the rare conditions that stem from single gene defects or processing errors.

Unfortunately, as with any disease translational research for RDs is a resource-intensive undertaking demanding both time and money. All parts of the RD community, whether
public or private, must embrace new models of engagement and research directions, such as the use of system-wide datasets, the interrogation of biobanks, advanced systems biology and computational technologies from the orphan disorders and clinical trials, with the aim of impacting even the rarest of diseases [40]. This can only happen through integrated transparent programs such as the rare disease roadmap, that share resources and infrastructure to make use of scarce funding, expertise, data, and biological samples.

Challenges

Basic research, mostly in academic institutions, has always been instrumental in identifying the molecular mechanisms underlying most RDs and it has accelerated the development of both diagnostics and therapeutics. Obviously the paucity of funding for such research limits the speed of advances in this field, but additional challenges exist in gaining access to the natural history of RDs, in the development of animal models of RDs, accessing human tissues through biobanks, and access to chemical compound libraries. In the past many of these resource-intensive research and development activities have been conducted by better resourced biotechnology and pharmaceutical companies, even if in isolation from the academic or RD community. The current challenge facing this sector however is the time, cost and risk involved in bringing any new drug to market. It represents a formidable obstacle for pharmaceutical developers and most are cutting their internal R&D efforts. While the issues are multifaceted, they do include increasing number of regulatory hurdles for safety assessments, risk management and post-approval research requirements.

For RDs this is even more devastating, and potential therapies have often languished because of low numbers of patients and difficulties around clinical trial designs and statistical analysis for efficacy studies. As the pharmaceutical industry rationalizes its operations into more integrated network models, we will have to deal with a disconnect that exists between the various stakeholders in RD, namely the pharmaceutical companies, academic institutions, patient groups, and regulatory bodies. The issues in question will include patients’ privacy concerns, intellectual property issues, and patent protection around databases and repositories.

Ways forward

Significant initiatives have been in place to stimulate companies to undertake drug development for RDs, such as research grants, research design support, FDA fee waivers, tax incentives and market exclusivity for orphan drugs, and some promising compounds have been brought to the point of clinical testing. However, as previously indicated, many pharmaceutical companies have been changing the manner in which they conduct research and development and are rationalizing their operations into more integrated network models that include all stakeholders, including academic research centers, patient groups and public-private partnerships [41].

The development of such consortia should allow more funding to drive innovation within particular disease-focus areas such as RDs, and it should leverage the capabilities and expertise of all stakeholders. In such arrangements it is important that registries and biorepositories have common standards, including protection for patients and research participants, and appropriate data-sharing arrangements. The joint development and sharing of chemical compound libraries and new advances in drug-targeting technologies will speed therapeutic development in the RD field. Alignment of objectives to deliver innovative new therapies to treat debilitating disease must be free of intellectual property issues, but retain privacy and avoid conflicts of interest with and between academic groups. Additional elements of an integrated strategy are outlined in the report by the Committee on Accelerating Rare Disease Research and Orphan Product Development [42].

Fulfilling the promise of personalized treatment

Overview

Personalized treatments are important for people with RDs in view of the heterogeneous causes of disease. All components in the RD Roadmap facilitate personalized research and therapies are developed that ultimately should lead to better health outcomes. For instance, -omics platforms generate personalized genomic data and GWAS can provide individualized genetic disease risk predictions that facilitate molecular therapies such as exon skipping [43], gene replacement, nonsense mutation read-through technologies, up-regulation of homologous genes, tailored for a specific gene mutation(s). Bioinformatics analysis strategies, tools, integrative analysis, quality of datasets, annotation, curation and evidence-gathering are pivotal [25] to these approaches. However, it must be acknowledged that personalized treatments are currently available for only a tiny minority of rare diseases.

Challenges

The major challenge is the difficulty in carrying out trials on small numbers of patients, recruiting and monitoring these patients scattered across large geographic distances and environments, and accounting for individual variation versus treatment effects, thereby strengthening the need for better knowledge and understanding of population genetic sub-structure. Adverse drug reactions are the fourth largest cause of death in the developed world [44], and drugs are metabolized by a number of pathways, some of which are highly polymorphic systems [45]. Different adverse drug reactions can be relatively rare events in the treatment of common diseases, and it will be much harder to link cause and effect when treating RDs, especially as polymorphisms can influence variation in the natural history of disease progression and so pose challenges in designing and tracking appropriate studies.
Ways forward

Well curated genotype and phenotype registries of individuals with RDs will facilitate the investigation of confounding factors in clinical trials, thus allowing tailoring of the drug to the individual and stratifying patient cohorts into likely responders, non-responders and those at risk of a serious adverse drug reaction. New therapies based upon the type of mutation, regardless of the specific gene, are being developed. It has been shown that read-through of nonsense mutations can be achieved with the aminoglycoside antibiotic gentamicin [46]. But the application is limited due to the serious side-effects associated with the drug, so extensive research efforts have been undertaken to develop safer read-through compounds, such as Ataluren [47]. However, controversy also surrounds the mechanism of action and efficacy of this drug [48].

Another form of personalized therapy aims to delay the progression of Duchenne muscular dystrophy (DMD) through targeted exon skipping, with antisense oligonucleotides which can specifically redirect pre-mRNA processing to restore an open reading frame. In this manner, the disease-causing protein truncating mutations are removed and the expression of a shorter but still functional protein is induced. Although not applicable to all genes, antisense oligomer-induced exon skipping may be relevant to genes encoding large structural products such as DMD [49].

The approach of studying the effects of therapies in patient cell models emphasizes the importance of quality biobanks of patient cells, i.e. cells that are relevant to the disease and appropriate for assessing potential therapies. Improved partnerships between researchers, clinicians and patient groups may lead to an improved shared understanding of the diagnosis, progression and treatment of disease. Such shared understanding needs to be reflected and standardised in clinical and disease coding systems.

Coding and classification of RD has been described as the major limitation for epidemiological surveillance and monitoring for RDs. The international reference for classification of diseases and conditions is the International Classification of Diseases (ICD), coordinated by the World Health Organisation. The current version of this classification system ICD-10, is currently used Australia-wide, in public and private health settings. This classification system provides common, comparable diagnostic coding. Importantly, it enables surveillance and monitoring through disease specific epidemiology and investigations of health care and patient outcomes.

Unfortunately, ICD-10 does not support comprehensive reporting of RDs, nor is there any other widely accepted RD coding scheme. Therefore, currently only about 3.5% of RDs are ICD-coded worldwide [50]. This is likely to result in significant under reporting of RD and limits the ability of the Australian health system to collect quality data for analyses of the: (i) individual and collective prevalence and incidence of RDs; (ii) impact and burden of RDs on individuals, carers, families and the healthcare system; (iii) quality of and equity of access to health care services; and (iv) impact of future changes in policy and service implementation on health and societal outcomes.

Importantly, the shortfalls in current coding are soon to be addressed through the next release of ICD codes, ICD-11 due for release in 2017. In anticipation of the adoption of ICD-11 Orphanet, an international consortium that provides a reference portal for RDs, has developed a new set of classifications for approximately 2500 RDs.

The International Rare Disease Research Consortium (IRDiRC) has been established to bring together researchers and organizations investing in rare disease research in order to achieve two main objectives by the year 2020, namely to deliver 200 new therapies for rare diseases and the means to diagnose most rare diseases (http://www.irdirc.org). In future, it may be feasible to examine the natural history of disease in each patient, and look for changes in the ‘slope’ following treatment in \( n=1 \) trials [51].

eHealth and a whole of life record

Overview

In Section 1, the need for enhanced patient/health professional partnerships through registries was discussed. These partnerships extend more broadly to engagement with other stakeholders including the biopharma industry. While it is important to characterize the differences between registries and eHealth personalized records, as part of the RD roadmap we also must understand the conditions under which these two sources of information can be unified. Ultimately, ongoing patient treatment and the capture of whole-of-life records must be interoperable with leading-edge translational RD research and development. Capturing the natural history of an RD across the patient population is another dimension that must be addressed.

Challenges

There are numerous challenges and barriers to changes in this component, ranging from the simple and accurate data capture of diagnosis and ongoing treatment, and the ability to revise legacy information with more informed disease characterizations, to integration of these data sets with national eHealth systems. Training of the next generation of health professionals who specialize in RD should include understanding standardized common data elements and best practice disease ontologies [52]. While information management systems can be developed, knowledge gained about a given RD must be transferrable to other rare conditions. Also, Internet-based software systems must be accessible in rural and remote areas where RDs may be expressed in relatively isolated populations, but the Internet bandwidth is reduced.

Ways forward

Information management systems must be developed to capture knowledge about specific RDs. This step will enable the generation of generalized rules for application to other rare conditions. Software applications that run on mobile devices and can be used by patients to track their own health need to be developed [53]. These solutions must be capable of linkage to eHealth records and utilize agreed standardized vocabularies for common data elements.
Policy and regulatory frameworks

Overview

The key to sustainability is to embed all of the above elements in an integrated and harmonized policy and regulatory framework. Such a framework will ideally be national in scope, international in perspective, and local, regional or community-based in its implementation. While prototype national RD protocols are very useful they only constitute a starting-point in improving the lives of patients with RDs, as each plan will need to consider resourcing, workforce, training, performance indicators and ongoing evaluation. Each country will also have critical differences in its national laws with respect to the protection of patient privacy in medical records, disclosure of information such as notifiable disease lists, data collection and storage, research and ethics, orphan drug development, and specimen storage, including biobanks.

Challenges

Progressing an RD roadmap requires consideration of jurisdiction-specific regulations and ethical, social and legal structures. Consideration must also be given to education of the public, community advocacy, professional training, the provision of incentives and subsidies, innovative financing (including from the private sector), standard-setting, and streamlining of approval processes where appropriate.

Ways forward

Respectful and meaningful national and international partnerships are the key to introducing and sustaining the change agenda. In democratic countries, consistent advocacy from affected patients and their families expressed through community organizations such as EURORDIS (www.eurordis.org) or rare voices Australia (http://www.rarevoices.org.au) is essential. These groups have a unique capacity to work with motivated individuals in the political arena, health care organizations, government departments, academia and the private sector, to address what are often difficult and complex problems.

In the absence of such advocacy, there are many other well-defined and numerically more pressing health concerns to occupy the attention of policy-makers. With such advocacy and partnerships, data can be assembled from a range of sources, e.g. population level prevalence; incidence and mortality data; quality of life surveys; hospital costing data; personal stories of impact and navigation through existing health care systems; evaluation of new models of care; and clinical trials. Together, they provide a more complete picture of the costs and benefits of moving away from the status quo [12]. A comprehensive picture of this nature is required by national and state governments before they will contemplate changes to the regulatory environment.

Discussion

The promise of personalized medicine, via new technologies and new -omics platforms, will only be fully realized through comprehensive system reform and integration, including population-level design elements. The RD Roadmap described in this paper shows that key components of this new paradigm begin and end with the individual living with an RD. There are challenges and opportunities in each of the components, but an even greater challenge in their linkage and integration. The lack of attention and resourcing given to RDs forces a level of analysis and new thinking that is often applicable to more common diseases. The ‘old’ technological and policy challenges relating to disease registries, biobanks, orphan drug development and e-health need addressing as much as the ‘new’ challenges relating to -omics and bioinformatics. Progress in these different areas will allow us to realize the potential of the new technologies for both rare and common diseases, but failure to develop an appropriate rare disease plan will slow the realization of benefits from personalized medicine.

Author statements

Funding
None.

Competing interests
None declared.

Ethical approval
Not required.

Acknowledgements

The authors received funding from the Australian National Health and Medical Research Council (APP634485, APP1055319) and EU FP7 Project (HEALTH.2012.2.1.1-C): RD Connect: An integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. F. Guerrero received funding from United States Department of Agriculture, Agriculture Research Service Project No. CRIS 6205-32000-031. USDA is an equal opportunity employer. The authors wish to acknowledge their involvement in the International Rare Disease Research Consortium. We wish to acknowledge Hugh Dawkins for his valuable feedback on this manuscript. M. Bellgard, initiated this research via a fellowship under the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agriculture Systems. This manuscript is dedicated to the memory of Professor Ernest F. Retzel (1949-2012).

References


