

## Meeting report series

# Report of the 2<sup>nd</sup> DSC WG Genome/Phenome teleconference

August 29, 2013

### Organization

Organized by: IRDiRC Scientific Secretariat  
Teleconference

### Participants

Prof Anthony Brookes, Leicester, UK, chair  
Dr Kym Boycott, Ottawa, Canada  
Dr Ya Gao, Shenzhen, China  
Prof Ada Hamosh, Baltimore, USA  
Prof Jim Lupski, Houston, USA  
Prof Olaf Riess, Tuebingen, Germany  
Dr Ralph Schuster, Bonn, Germany  
Dr Jianguo Zhang, Shenzhen, China

Dr Barbara Cagniard, IRDiRC Scientific Secretariat  
Dr Sophie Höhn, IRDiRC Scientific Secretariat

### Apologies

Prof Han Brunner, Nijmegen, the Netherlands  
Dr Xavier Estivill, Barcelona, Spain  
Dr Wei Wang, Shenzhen, China  
Dr George Patrinos, Patras, Greece

### Agenda

1. Welcome
2. Review of the topics previously discussed
3. Suggested action plan for the proposal to Global Alliance
4. Discussion on possible actions to “free the data”
5. Other topics

## REPORT

### Review of the topics previously discussed during the first teleconference

During the first teleconference, the following subjects were discussed:

- ▶ Data need to be freed for maximal sharing and progress
- ▶ IRDiRC should contribute to the Global Alliance initiative in terms of rare disease related data
- ▶ Key problems of data sharing and data using in the Genome/Phenome field

The next steps of this WG would be to:

- ▶ Make an approach to the Global Alliance initiative in order for IRDiRC to take a leading role in the rare disease aspect
- ▶ Write a proposal explaining why IRDiRC should take a leading role in the rare disease field, what IRDiRC could do and what would be the issues
- ▶ Present that proposal during the next Executive Committee that will take place in Miami on September 23-24<sup>th</sup>

### Suggested action plan for the proposal to Global Alliance initiative

The action plan to be presented to the Executive Committee should:

- ▶ Present a summary of current initiatives in the field and explain how IRDiRC is connected to those initiatives
- ▶ Present IRDiRC actions and achievements in rare disease field
- ▶ Present the gaps in terms of standards, tools, and strategies

However, the working group is aware that rare diseases may be seen by the Global Alliance initiative as a very important aspect, and that they may want to personally manage this topic.

IRDiRC should also present in their action plan to Global Alliance suggestions for the relative roles and relationships between at least these issues:

- ▶ Patient registries, diagnostic lab data and research lab data
- ▶ Collaboration of the big initiatives for research data
- ▶ Universal patient identifiers
- ▶ Data quality and completeness challenge
- ▶ Capture and granularity of clinical and phenotype data
- ▶ Data discovery and knowledge sharing, as important adjuncts to data sharing

## Discussion on possible actions to help “free the data”

### Patient source data

- ▶ Emphasize patient engagement and guidance in the construction of patients’ registries

The patient’s interest in their own disease and in rare diseases is significant, and their contribution is going to be key.

There are several ways to try to catch their input:

- ▶ Patients support groups or action groups
- ▶ Capturing data from the patient into a data ecosystem
  - Construction of software, tools, modules, components for the patients
  - Construction of a federated layer of patients’ registries in the data ecosystem

How should IRDiRC balance, inter-related and enable the creation of an effective federated layer of patient registries with one or more central systems for searching/or mirroring (some of) this content?

### Data sharing issues

The two big problems with data sharing are:

- ▶ Whether the person ‘owning’ the data wishes to share it
- ▶ Whether the person ‘owning’ the data is allowed to share it legally

Valuable phenotype and genotype data related to patients reside in the healthcare system. The data sharing problem can be much more challenging in that case, in that each different hospital, in each different country, in each different disease, may place different constraints on whether patients are allowed to share their data, whether the hospital wants to cooperate, whether the phenotype can be gathered in that situation. Accessing these data from diagnostics laboratories is difficult, not least because they often do not even have access to the phenotype data in digitised form.

So work needs to be done to encourage and facilitate more detailed digital recording of these data in clinical care situations, and the establishment of policies and procedures for connecting or feeding this into the broader RD data ecosystem.

A parallel strategy for accessing whatever data are digitised at any time point would be to engage the assistance of the patient in freeing this information, by having them organise the release of their electronic healthcare record data. We should work closely with several other projects already making great progress in this area (e.g., PatientsKnowBest).

Additionally, patients may wish to enter data about themselves directly into the data ecosystem – an approach which might raise questions over the veracity of the data, but which cuts out the clinical system and hence negates the problematic diversity of hospital data sharing policy. Patients are

often legally and emotionally prepared to share their data. How could this WG help such patients to get their data organized? IRDiRC should assist patient groups with constructing registries federated with the ecosystem, and encourage them to enter their own detailed phenotype data into these registries.

A key area of standardisation required to underpin all the above, relates to metadata content. This is a complex area to develop and standardise, and so urgent effort needs to be into this question, to inform and allow data to flow or be discovered within the desired database and registry ecosystem.

What can IRDiRC realistically do to move this forward?

- ▶ Convince the clinical groups to more effectively gather and make available RD genotype and especially phenotype data, by sharing and/or discovery modes
- ▶ Get clinical groups to routinely enlist their RD patients as research subjects and share data through the research mechanism

### **Initiatives' collaboration**

IRDiRC should try to bring together the big initiatives:

- ▶ DECIPHER initiative (UK)
- ▶ CARE for RARE system (Canada)
- ▶ RD-Connect (European Union)
- ▶ BGI (China)
- ▶ Centers for Mendelian Genomics (USA)

Labs are increasingly collaborating at the national level, to get data from clinical settings under control. Funding is also often focussed on national level initiatives. Major national groups therefore need to come together to start defining the standards and the real commonalities, minimal data format, etc. Those partners are already involved in IRDiRC. IRDiRC could get them to work together as one.

However, with increased connectivity of data, issues of dataset duplication and adjoining different datasets about each patient, will need to be addressed.

### **Universal global patient identification**

Each individual patient needs to be resolved in the rare disease data ecosystem. Therefore, a universal global patient identification is needed. A few proposals about how to achieve this have been floated, but there is a lack of agreement and a lack of momentum to move forward. Finding a universal global patient ID solution should be one of the WG key objectives.

### **Data quality and completeness issues**

Hospital databases have to collect the right data in a locally mandated standardized ways. This is especially true for phenotype data. A campaign should be started to improve how and how much data are collected.

## The data discovery approach

In a number of countries, diagnostic laboratories wish to expose patients' data to other diagnostics laboratories within their country. When this happens they tend to share those data behind a firewall system, not openly and not internationally.

To maximise the utility of these data, they could be made 'discoverable' in a platform system whereby each of these laboratories can search across all of the other laboratories for certain phenotypes and/or mutations of interest, without revealing the actual data themselves. Once hits are located, direct approaches to the source lab could be made to request deeper access to the data.

This provides a powerful, fully open way to free data which are not consented for any kind of sharing. The data remain private at all times.

Data discovery approaches, and related knowledge sharing (achieved by facilitating views into aggregation level representations of data organised for discovery purposes) offer new and complementary approaches to data sharing.

## Main deliverables

- ▶ Write one or two pages about how IRDiRC could become a leadership in rare diseases in the Global Alliance initiative and present it during the next Executive Committee meeting that will be held in Miami on September 23-24<sup>th</sup>
- ▶ Present a summary of current initiatives in the field and explain how IRDiRC is connected to those initiatives
- ▶ Present IRDiRC actions and achievements in the rare disease field
- ▶ Present the gaps in terms of standards, tools, and strategies