

## Meeting report series

# Report of the 4<sup>th</sup> DSC WG Genome/Phenome teleconference

18 July 2014

### Organization

Organized by: IRDiRC Scientific Secretariat  
Teleconference

### Participants

Prof Anthony Brookes, Leicester, UK, chair  
Dr Kym Boycott, Ottawa, Canada  
Prof Han Brunner, Nijmegen, the Netherlands  
Dr Xavier Estivill, Barcelona, Spain  
Prof Olaf Riess, Tuebingen, Germany  
Dr George Patrinos, Patras, Greece  
Prof Ada Hamosh, Baltimore, USA

Dr Barbara Cagniard, IRDiRC Scientific Secretariat  
Dr Lilian Lau, IRDiRC Scientific Secretariat

### Apologies

Prof Jim Lupski, Houston, USA  
Prof Yves Moreau, Leuven, Belgium  
Dr Wei Wang, Shenzhen, China

### Agenda

1. General recap of WG discussions to date
2. Discussion on matchmaker status
3. Possible areas of focus going forward

## REPORT

### General recap of WG discussions to date

Among the many gaps in the field – standards, data sharing, discovery of data, databases to deposit into, use case, etc. – the WG defined 2 related priorities in previous discussions:

- ▶ Data availability
- ▶ Defining pathogenicity

This corresponds to what the Diagnostics Scientific Committee independently asked the WG to work on:

- ▶ An editorial on the priorities defined
- ▶ Matchmaker project
- ▶ List of databases to deposit information into
- ▶ List of resources to analyse the frequency of variant

An editorial on a range of related topics is being prepared for Nature Reviews Genetics.

A proposal for funding of activities related to ‘Matchmaker’ solutions was presented to the Executive Committee in June. As there is no common pot, discussion are still ongoing on how to finance this project.

Regarding the list of databases, several projects are aiming to or are working on it: RD-Connect, pending EU proposals, Bioinformatics WG of NIH Global Genomic Medicine Consortium, Orphanet, Global Alliance for Genomics and Health (GA4GH). Therefore it was suggested that the WG should review these lists as they emerge, to determine if any tools/resources are missing. A composite list could then be posted on several website (GA4GH, IRDiRC, etc.) for better diffusion.

The Diagnostics Scientific Committee, given their interest in establishing the frequency of human variants in different populations, recently created a new WG on Population Controls Variant Datasets to work on this topic.

Members of the Genome/Phenome WG should therefore think about concrete action for this WG in addition to the editorial to write and the Matchmaker project. As a reminder, the Diagnostics Scientific Committee intends to merge the WG on Sequencing and the WG on Ontologies and Rare Disease Prioritization (once their initial tasks are complete) into the WG on Genome/Phenome, probably in a year or so from now. The WG on Genome/Phenome will then use the tools developed by all the WGs to facilitate diagnostics and discovery of rare disease genes.

### Discussion on Matchmaker status

The idea of the Matchmaker project emerged in the course of the RD-SympathI symposium (April 2013), and a meeting held later the same year in Boston at the ASHG conference actually launched the project. This project - called ‘Matchmaker-Exchange’ - is an example of an open collaboration to address core

needs of the community, and it is facilitated by GA4GH (organizing teleconference and minutes) and deemed to be driven and supported equally IRDiRC. Several members of the Genome/Phenome WG are members of the project. A white paper is in preparation.

Members of the WG agreed that the goals of IRDiRC and GA4GH are similar in this area; the efforts of both WGs should thus be fully integrated to avoid duplication or competition:

- ▶ Members of this WG will be invited to the API WG teleconference
- ▶ Minutes of the API WG will be circulated to this WG
- ▶ This WG will review the white paper prepared by the API WG when available
- ▶ Both WGs should thus discussed a possible distribution of work

As a pilot for the matchmaker concept, API development is ongoing to allow the resources of GeneMatcher (USA) and PhenomeCentral (Canada) to communicate for data discovery. The first version of the API is based on a 'data submission' approach: i.e., a match 'requestor' must provide patient data, after which the query principles via which similar patients are identified is decided by the database receiving the request (e.g., gene-based matching, degree of phenotype similarity, mutation types, etc). The plan is to further develop this API (version 2) to enable a 'query based' approach, wherein the requestor is able to specify and control the exact basis for patient matching, to obtain more specific/customised hits. As resources connected through the Matchmaker Exchange will be different (e.g., rare genes only, variants of genes, alternative phenotype ontologies), it seems appropriate to develop a query system to limit the diversity of hits returned from different databases .

Numerous questions remain unsolved:

- ▶ Who to allow doing the search inquiry?
- ▶ How to authenticate the enquirer?
- ▶ How to ensure that the enquirer has authorisation?
- ▶ Who should be contacted when a match is found?
- ▶ Should data requestor and data depositor always provide their name?
- ▶ Should matchmaker requests and results always be logged?
- ▶ What ethics, consent terms and legislation should apply?

Decipher and Gem.app are two other resources interested in joining the MatchMaker project. Ideally, other resources (diagnostic laboratories, research institutes, EU consortia, etc.) should also be integrated in the future.

Next steps in the development of version 2 of the API

- ▶ Face-to-face meeting of the API WG in San Diego in October (Tentative date: Tuesday 10/21 from 6:30-8:30pm).
- ▶ 2-day workshop in 2015 (if funding from IRDiRC is obtained).

## Possible areas of focus going forward

### Metadata requirements and standardization

Members of the WG agreed that there is a big gap around metadata standardization, and that minimal metadata requirements should be developed (e.g., who can access data for what purposes, who should be acknowledged, what methods were used for data generation, data quality metrics, etc.) to be able to start connecting and using data more effectively.

Groups interested in the topic (e.g., NeurOmics, IRDiRC, BiomedBridges, pending EU funding proposals) should work together to define some minimum metadata standards. To start, a few workshops would be sufficient to establish the core issues and develop some details – perhaps funded by IRDiRC?

### Minimum content open data conventions

Work in GEN2PHEN previously raised the question of ‘what is data’ vs ‘what is a link to data’ with regards to patient information. For example, EBI (working only with Open data for its main services) and others contend that the HGVS name of a mutation is actually nothing more than an effective link/citation with so little informational content that it could and should always be made freely and openly available. Indeed, it has even been suggested that HGVS names plus headline disease/phenotype terms should be made openly available, to facilitate widespread data discovery and aggregate correlation of genotype-phenotype frequencies and relationships, without patient identification risks.

However, others contend that there will be ethical concerns with this approach, as even such minimal level information used as links still involves the use of patient data.

Members of the WG agreed that this topic should be further discussed.

### Other possible topics

- ▶ Listing and access to analytical tools: ongoing by Global Alliance for Genomics and Health
- ▶ Accessing/discovering Genome/Phenome data from IRDiRC members: what projects do members have?
- ▶ How to share knowledge without exposing data: aggregate information, graphical display, etc.
- ▶ Issues related to Whole Genome Sequencing compared to Whole Exome Sequencing: validity of software, etc.
- ▶ Pathogenicity classes and related matters (expressivity, penetrance)

### Main deliverables

- ▶ Further consideration of what topics to address and to prioritize, to be discussed at the next teleconference, to give recommendations to the DSC
- ▶ Organize a teleconference for mid-September