Diagnostic tools for all rare diseases by 2020 – International Rare Disease Research Consortium (IRDiRC)

Gert Matthijs, member of the Diagnostics Scientific Committee
May 28, 2016
Objectives of IRDiRC by 2020

1. 200 new therapies for rare diseases

2. Means to diagnose most rare diseases
IRDiRC Consortium

Formally launched in 2012

(43) Funding organizations from:
• Asia & Middle East
• Australasia
• Europe
• North America

Present commitment exceeds $2B worldwide
IRDiRC – Basic Principles

- Cooperation at international level to stimulate, better coordinate & maximize output of rare disease research efforts around the world
- Teams up public and private organizations investing in rare diseases research
- Research funders with relevant programs >US$10 million over a 5-year period can join & work together
- Each organization funds research in its own way
- Funded projects adhere to a common framework
Governance Structure

- 1 representative per funding body or group of funders (accumulative funding)
- Representatives of umbrella organizations of patient advocacy groups
- Chairs of the Scientific Committees
- Coordinator of the Scientific Secretariat

~15 members with balanced representation of scientists, patients, industry, etc.

Scientific Secretariat

Executive Committee

- Diagnostics
- Interdisciplinary
- Therapies

Scientific Committees

Task Forces

Patient-related/relevant outcome measures (2015)

Small population clinical trials (2015)

Matchmaker Exchange (2015)

Other topics

Nominated experts, on *ad hoc* basis
IRDiRC’s DSC committee

- Diagnostic scientific committee (DSC)
  - 14 members
  - Chair: Dr. Kym Boycott
Number of rare diseases

Cumulative number of new rare diseases by month since 2010

Source: Orphanet Data
## IRDiRC Commentary

### Table 1. Factors Contributing to Bottlenecks in the Gene Discovery Pipeline

<table>
<thead>
<tr>
<th>Category</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical data</strong></td>
<td>Ultra-rare genetic diseases</td>
</tr>
<tr>
<td></td>
<td>Non-specific clinical presentations; e.g. developmental delay, hypotonia</td>
</tr>
<tr>
<td></td>
<td>Lack of natural history information</td>
</tr>
<tr>
<td></td>
<td>Imprecise or lack of standard phenotyping</td>
</tr>
<tr>
<td><strong>Genomic data analyses</strong></td>
<td>Lack of standardized and optimized tools for informatics pipeline</td>
</tr>
<tr>
<td></td>
<td>Lack of control datasets; population specific</td>
</tr>
<tr>
<td></td>
<td>Structural variation and copy number variants not captured well by WES</td>
</tr>
<tr>
<td></td>
<td>Challenges to interpreting noncoding variation</td>
</tr>
<tr>
<td><strong>Genetic validation</strong></td>
<td>N-of-1, lack of infrastructure for sharing clinical and genomic data for unsolved patients</td>
</tr>
<tr>
<td><strong>Functional validation</strong></td>
<td>Lack of standardized and moderate-throughput analyses of variant impact</td>
</tr>
<tr>
<td></td>
<td>Lack of biological insight into the function of many human genes</td>
</tr>
<tr>
<td><strong>Disease mechanisms</strong></td>
<td>Other mechanisms including tissue-specific mosaicism, methylation, di- or oligo-genic inheritance</td>
</tr>
</tbody>
</table>
Diagnostic challenges

- Promote the use of available tools for analysis, data collection and variant interpretation
- Improve phenotyping
- Find the patients
- Solve the ‘unsolved’
  - Gene discovery
  - Other mechanisms of disease
- Promote quality of the diagnostics
- Transfer NGS to the health care system
Matchmaker Exchange

- Provides data sharing tools between clinical geneticists to match unsolved genome/exome sequence cases

- Ensures optimal collaboration between all projects contributing to the interpretation of variants and of matching phenotypes and variants

- Joint IRDiRC-GA4GH collaboration
Two-sided Hypothesis Based Matching

“I am looking for samples with LoF mutations in GeneX”

“I am looking for samples with LoF mutations in GeneX”

Courtesy: Dr Kym Boycott
Currently Connected MME Services

Courtesy: Dr Kym Boycott
THE BIRTHDAY PARADOX

If you survey a random group of just 23 people there is a 50% chance two of them will have the same birthday.

Courtesy: Dr Kym Boycott
65 undiagnosed cases with correct leads before first match is made; first several matches made quickly

Much larger number of cases are needed to discover a significant fraction of all undiagnosed diseases

50,000-250,000 cases will be required to identify about 2000 disease genes
Matchmaker Exchange
Genomic discovery through the exchange of phenotypic & genotypic profiles

Special Issue
Guest Editors: Kym Boycott, Ada Hamosh, and Heidi Rehm

Matchmaker Exchange
Multiple disconnected projects

- Disease and Variants
- Model Organisms
- Variant and Disease
- Disease and VCFs
- Gene
- Gene Matcher
- Gene and Phenotype (HPO)
- Variants
- Variants and Phenotype
- Gene and Phenotype (HPO)
- Variome
- VCFs and Phenotype (HPO)
- VCFs and Phenotype (HPO)
- Disease and VCFs
- Phenotype
- Diseases Program
- Phenome Central
- Phenome Central
- GEM.app
- Genome Connect
- GEM.app
- LOVD
- Variome
- Gene Yenta
“IRDiRC Recommended”

- Label highlighting tools, standards, platforms and guidelines which contribute directly to IRDiRC objectives

- Identification of key resources for research communities to accelerate clinical translation
“IRDiRC Recommended” Resources

- International Charter of Principles for sharing Bio-Specimens and Data
- Orphanet
- PhenomeCentral
- Orphanet Rare Disease Ontology (ORDO)
- DECIPHER
- OMIM
- GA4GH Framework for Responsible Sharing
- HPO
- ICHPT
- TREAT-NMD Patient Registries
- TREAT-NMD Standard Operating Procedures
Diagnostic challenges

- Promote the use of available tools for analysis, data collection and variant interpretation
- Improve phenotyping
- Find the patients
- Solve the ‘unsolved’
  - Gene discovery
  - Other mechanisms of disease
- Promote quality of the diagnostics
- Transfer NGS to the health care system
Guidelines for diagnostic next-generation sequencing

Gert Matthijs*,1,8, Erika Souche1,8, Mariëlle Alders2, Anniek Corveleyn1, Sebastian Eck3, Ilse Feenstra4, Valérie Race1, Erik Sistermans5, Marc Sturm6, Marjan Weiss5, Helger Yntema4, Egbert Bakker7, Hans Scheffer4 and Peter Bauer6

We present, on behalf of EuroGentest and the European Society of Human Genetics, guidelines for the evaluation and validation of next-generation sequencing (NGS) applications for the diagnosis of genetic disorders. The work was performed by a group of laboratory geneticists and bioinformaticians, and discussed with clinical geneticists, industry and patients’ representatives, and other stakeholders in the field of human genetics. The statements that were written during the elaboration of the guidelines are presented here. The background document and full guidelines are available as supplementary material. They include many examples to assist the laboratories in the implementation of NGS and accreditation of this service. The work and ideas presented by others in guidelines that have emerged elsewhere in the course of the past few years were also considered and are acknowledged in the full text. Interestingly, a few new insights that have not been cited before have emerged during the preparation of the guidelines. The most important new feature is the presentation of a ‘rating system’ for NGS-based diagnostic tests. The guidelines and statements have been applauded by the genetic diagnostic community, and thus seem to be valuable for the harmonization and quality assurance of NGS diagnostics in Europe.

Quality of the reports

Patient identification
Surname XXXX
First name XXXX
Gender X
EADnr 0000000
Birth date 00/00/0000

Sampling time 00:00:0000 00:00
Sample type blood
Sample number XXXX
GC code GC000000 (A,GC000000)
Project code MD1-MT283-Gempanel-20

Subpanel HCM ⊆ gene panel HCM_LQT (v1.0)

Annex 1 General
The general characteristics of the subset of the capture panel, the subpanel HCM, is shown.

List of the 75 genes available with a minimum of 100% of the data.

Method
For each sample, DNA Sample Preparation遭遇 samples were pooled and enriched fragments were sequenced. This was

Annex 2 Quality parameters
Gene panel HCM_LQT applied to sample GC000000
- The data satisfy the proposed quality criteria.
- 99.14% of the gene panel HCM_LQT is genotyped.
- Total number of variants HCM_LQT: 3782

Subpanel HCM applied to sample GC000000
- 99.99% of subpanel HCM is genotyped; the 26 genes transcriptome of subpanel HCM are completely genotyped unless it is indicated otherwise (e.g., \( \text{\#} = 75\%\):)

- (partially) genotyped by Sanger

Annex 3 List of variants
The variants are assigned to one of the following classes: (1) benign, (2) likely benign, (3) variants of unknown significance, (4) likely pathogenic, or (5) pathogenic. Variants of class (1) and (2) are not included below.

(Likely) pathogenic mutation(s) associated with the phenotypes

<table>
<thead>
<tr>
<th>Gene</th>
<th>4DNA</th>
<th>Genomic position</th>
<th>Protein</th>
<th>Classification</th>
<th>Zygosity</th>
<th>Heredity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYHDCPS</td>
<td>NSQ01636361</td>
<td>16q21.32-33:563</td>
<td>p.Thr2346Arg</td>
<td>likely pathogenic</td>
<td>heterozygote</td>
<td>ADO</td>
</tr>
</tbody>
</table>

There is no variant of unknown significance that could be associated with the phenotype.
Variant classification

CLASS 1
- Neutral

CLASS 2
- Likely neutral

CLASS 3
- Effect unknown

CLASS 4
- Likely deleterious

CLASS 5
- Deleterious

LR <0.1%
>1% controls
in trans with pathogenic variant

LR 0.1-4.9%
>1% ethnic group (founder)
synonymous substitution without effect on mRNA

LR 5-94.9%
Insufficient evidence
Silent variation with effect on splicing
Missense variation with effect on splicing
In-frame indel

LR 95-99%
IVS+/1 in frame
IVS+/-2 in frame

LR >99%
Nonsense mutations in any exon except the last
IVS+/-1
IVS+/-2
Large genomic deletions
Deep sequencing reveals 50 novel genes for recessive cognitive disorders


Mutations affecting housekeeping genes

In the LARP7 gene, we found a frameshift mutation in a family with intellectual disability and microcephaly. LARP7 is a negative transcriptional regulator of polymerase II genes, acting by means of the 7SK RNP system. Within the 7SK RNP complex, the positive transcription elongation factor b (P-TEFb) is sequestered in an inactive form, preventing RNA polymerase II phosphorylation and subsequent transcriptional elongation. Hitherto, no disease association has been reported for LARP7.

Presumably causative homozygous mutations were also found in KDM5A and KDM6B. These genes encode histone demethylases that specifically demethylate histone H3 at lysine 4 and lysine 27, respectively, and they both have a central role in the histone code. We have in humans, adenosine kinase deficiency should lead to intellectual disability, whereas in the mouse, overexpression of Adk causes neurological symptoms, and Adk deficiency gives rise to early lethal liver steatosis. Nothing is known yet about the function of the CI2orf57 gene, apart from its apparent overlap with ATNI (see UCSC Genome Browser, NCBI36/hg18). CAG trinucleotide expansion in the ATNI gene is the cause of dentatorubral pallidoluysian atrophy (DRPLA), another syndromic form of intellectual disability. A comprehensive list of families with single, probably disease-causing mutations is shown in Table 2.

Despite exhaustive validation of our data and stringent filtering against all known neutral and pathogenic sequence variants (see Supplementary Information and Supplementary Tables 3–6), it is still possible that not all of these changes will turn out to be causative. Particularly for the numerous missense mutations observed, functional studies will be required to rule out rare polymorphisms that are unrelated to intellectual disability. In a previous study, 1% of the protein-truncating mutations on the X chromosome were found to be unrelated to disease, and in our study, 12 observed inactivating mutations did not co-segregate with intellectual disability (see Supplementary Table 4). However, we believe that the vast majority of the changes presented here as probably pathogenic will be confirmed, even if they have been observed only once, because most of the proteins encoded by these novel candidate genes interact with the
One-sentence summary: Patients with genetic disease of unknown causes can be rapidly diagnosed by bioinformatic analysis of disease-associated DNA sequences and phenotype.

Editor’s Summary:
Efficient Diagnosis of Genetic Disease

We know which genes are mutated in almost 3000 inherited human diseases and have good descriptions of how these mutations affect the human phenotype. Rapid sequencing of these genes in a group of 40 patients’ physiological characteristics and phenotype knowledge enabled the authors to acquire only about 2 hours of a geneticists’ time, enabling a fast straightforward process by which inherited disease in certain people...
Variant reference set

1. List of unique variable positions
2. VCF with reference calls
   - Sample 1
   - Sample 2
   - Sample n
3. BAM
   - Sample 1
   - Sample 2
   - Sample n
4. VCF with reference calls
   - Sample 1
   - Sample 2
   - Sample n
5. Summary population VCF with
   - Allele Count (AC)
   - Allele Number (AN)
   - Genotype Count (GTC)
Task Forces to Consider

▶ Clinical Data Sharing for Diagnosis

❖ Strategies for Diagnostic Approach for Patients with Rare Diseases
  – What patients should be reimbursed – patient indication

❖ Public health system integration
  – Health economic evaluations
  – Clinical outcomes

❖ Clinical sharing of genome-wide data for secondary use
  – Core data elements for secondary use of data and justification
  – Strategies for sharing
  – Interface, Consent
  – Patient-driven sharing
  – Multi-stakeholder (Diagnostic Laboratories, Patients, Clinicians, Payers)
Governance, Policies and Guidelines

Available on www.irdirc.org