Exome and genome sequencing in clinical practice

2nd IRDiRC conference - Shenzhen

Christian Gilissen Ph.D.
christian.gilissen@radboudumc.nl
07-11-2014
• Clinical genetics, diagnostics and research in one department
• Largest department of Human Genetics in the Netherlands
The challenge in diagnostics

- Single gene testing offered for more than 400 Mendelian diseases and more than 800 genes.

  ✔ Good diagnostic yield for diseases with few causative genes (e.g. Noonan syndrome).

  ❌ Poor diagnostic yield for genetically heterogeneous diseases. (e.g. Blindness, > 100 genes)

- **Solution**: Next generation sequencing?
Why exome sequencing?

<table>
<thead>
<tr>
<th>Sanger</th>
<th>Targeted</th>
<th>Exome</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Very accurate</td>
<td>• Optimization possible</td>
<td>• No bias for genes</td>
<td>• No bias in what you sequence</td>
</tr>
<tr>
<td>• Cheap per exon</td>
<td>• Low chance of incidental findings</td>
<td>• Standardized workflow</td>
<td>• Little technical biases</td>
</tr>
<tr>
<td>• High turn-around</td>
<td>• “Easy” analysis</td>
<td>• Re-use of performed exomes to interpret new ones</td>
<td>• Allows detection of SVs and SNVs in one experiment</td>
</tr>
<tr>
<td></td>
<td>• “Easy” interpretation</td>
<td>• Simple to add new genes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low diagnostic yield for genetically heterogeneous diseases</td>
<td>• Design and re-design required</td>
<td>• Sequencing bias</td>
<td>• Data analysis bottleneck</td>
</tr>
<tr>
<td></td>
<td>• Different designs for different disorders</td>
<td>• No non-coding regions</td>
<td>• Interpretation of non-coding variants</td>
</tr>
<tr>
<td></td>
<td>• Sufficient patients required</td>
<td>• Incidental findings</td>
<td>• Expensive, time-consuming</td>
</tr>
</tbody>
</table>

**Table:** Comparison of Sanger, Targeted, Exome, and Genome sequencing methods.
Current exome sequencing workflow

Out-sourced: BGI

- Enrichment (Agilent v4)
  - ~21,000 genes

Secondary Analysis

- Sequencing at BGI Copenhagen
  - Using Illumina 2x100bp, 75x median coverage

Report

- Read mapping with BWA
  - Variant calling with GATK

In-sourced: Nijmegen

- Quality control
- Sample mix-up check
- Variant annotation

Interpretation

- Gene package visualization
  - Standardized interpretation protocol
  - Independent interpretation by 2 people

Primary Analysis

- Enrichment (Agilent v4)
- ~21,000 genes

Report

- Validation by Sanger
  - [Segregation analysis and functional confirmation]
  - Report of results

±350 exome samples per month!
Diagnostic exome approaches

**Gene package approach**

*Most genes known*

- Variants in known genes
- **Variants patient**
- **Gene package**

**Pilot study:** 50 exomes for 5 disorders

Neveling *et al.* Hum mut. 2013

- Sequence the exome of the patient
- Look only at known genes

**Trio approach**

*Most genes unknown*

- **Variants mother**
- **Variants father**
- **Variants patient**

**Pilot study:** 100 trios for intellectual disability

De Ligt *et al.* NEJM, 2012

- Sequence the exome of father, mother and patient.
- Look for *de novo* mutations
Challenges in clinical exome sequencing
When it all works out...

- Prioritization of variants found by exome sequencing:

  - All variants: 43,906
  - Known genes: 404
  - Coding: 156
  - Protein affecting: 63
  - Rare dbSNP: 10
  - Rare in-house: 7
  - Recessive: 2
What can go wrong? Sample quality

- **Raw sequence and mapping statistics**
  - FastQC tool
  - Bedtools – coverage statistics
  - Per gene / exon target coverage

- **Variant statistics:**
  - Overlap dbSNP
  - Number of truncating mutations
  - Tr/Ti ratio
What can go wrong? Sample mix-up

- **Gender check:**
  - Calculate chrY/chrX target coverage ratio

- **SNP Test:**
  - 12 common SNPs tested separately by Sanger sequencing

- **Trio check:**
  - Compare high quality variant calls between child and parents
What can go wrong? No sequence coverage
What can go wrong: uncalled variants

O’Rawe et al. Genome med. 2013
## What can go wrong: incorrect interpretation

Bohring-Opitz syndrome is often fatal in early childhood.

http://evs.gs.washington.edu/EVS/
Better diagnostics by whole exome sequencing?

- Pilot for the gene panel approach
**Pilot study – gene package approach**

- **250 exomes**: 50 exomes for 5 genetically heterogeneous diseases

**Gene package design:**
- Only known genes are allowed, no candidate disease genes
- Gene lists must be up-to-date and is updated every ~3 months
- Created by team of experts from clinic, diagnostic and research division

<table>
<thead>
<tr>
<th></th>
<th>Number of genes (Sept. 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blindness</td>
<td>144</td>
</tr>
<tr>
<td>Deafness</td>
<td>98</td>
</tr>
<tr>
<td>Early onset colorectal cancer</td>
<td>115</td>
</tr>
<tr>
<td>Mitochondrial disorders</td>
<td>207</td>
</tr>
<tr>
<td>Movement disorders</td>
<td>152</td>
</tr>
</tbody>
</table>

Neveling et al. Hum mut. 2013
How to do 400 samples per month?

Variants and annotation

Quality control

Filtering
How to do 400 samples per month?

Variants and annotation

Quality control

Filtering

Patient DB

Variant DB
Diagnostic yield from exome sequencing

Neveling et al. Hum mut. 2013
Diagnostic yield from exome sequencing

Neveling et al. Hum mut. 2013
Current exome panels

Gene panels 2011

# genes in disease package

- Deafness: 99
- Early onset Cancer: 115
- Movement: 144
- Blindness: 145
- Mitochondrial: 219
Exome sequencing can be cost-efficient compared to Sanger when sequencing 3 genes or more.
Current exome panels

Novel gene panels 2014

# genes in disease package

- Deafness: 131
- Early onset Cancer: 115
- Movement: 183
- Blindness: 240
- Mitochondrial: 234
- ALS: 15
- Craniofacial: 64
- Immunological: 265
- Metabolic: 534
- Intellectual disability: 563
- Muscle: 116
- Kalman: 22
- Cardiomyopathy: 46
Better diagnostics by whole exome sequencing?

- Pilot for the trio approach
Pilot study – *de novo* approach

- **100 patients + 200 parents!**
  - Severe intellectual disability (IQ<50)
  - No etiological or syndromic diagnosis
  - Negative family history

- Patients have reached the end stage of conventional strategies
  - Targeted gene tests negative
  - Genomic array profile negative
Yield in 100 ID patients

<table>
<thead>
<tr>
<th>Positive diagnosis</th>
<th>June 2012</th>
<th>June 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All mutations</strong></td>
<td>16</td>
<td>29</td>
</tr>
<tr>
<td><strong>De novo mutations</strong></td>
<td>13</td>
<td>28</td>
</tr>
<tr>
<td>Autosomal dominant</td>
<td>10</td>
<td>23</td>
</tr>
<tr>
<td>X-linked</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Autosomal recessive</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Inherited mutations</strong></td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>X-linked</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Autosomal recessive</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Candidates</strong></td>
<td>19</td>
<td>11</td>
</tr>
</tbody>
</table>

Yield of ~30% in patients with severe ID

De Ligt et al. NEJM, 2012
Better diagnostics by whole genome sequencing?

- pilot study
Diagnosis in patients with severe intellectual disability (ID)

- Whole genome sequencing
- 50 trios at 80x coverage
- A *de novo* approach

±1,500 ID patients

- Single gene test: ~1-5%
- Genomic microarray: 12%\(^1\)
- Exome sequencing (n = 100): 27%\(^2\)
- Whole genome sequencing (n = 50): ??%

% of ID patients with a diagnosis: <40%

---

\(^1\) Vulto-van Silfhout, A. T. *et al* Hum mut 2013

\(^2\) de Ligt, J. *et al*. NEJM 2012
Can we identify *de novo* mutations?

- Validation rate: **38%**
  (80% for high confidence!)
- Coding *de novo* SNVs: **84**
- Comparison to WES SNVs: **12%**
Can we identify *de novo* mutations?

- Validation rate: **38%**
  (80% for high confidence!)
- Coding *de novo* SNVs: **84**
- Comparison to WES SNVs: **12%**

- Validation rate: **82%**
- Coding *de novo* SVs: **9**
Did we find more coding *de novo* mutations than other studies?

![Bar chart showing the number of identified coding *de novo* mutations per individual for different studies.](chart)

- **Controls**
- **Patients**

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of Identified Coding <em>de novo</em> Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xu et al. (22)</td>
<td>8</td>
</tr>
<tr>
<td>Sanders et al. (200)</td>
<td>125</td>
</tr>
<tr>
<td>Sanders et al. (238)</td>
<td>167</td>
</tr>
<tr>
<td>Xu et al. (53)</td>
<td>40</td>
</tr>
<tr>
<td>de Ligt et al. (100)</td>
<td>79</td>
</tr>
<tr>
<td>Gulsuner et al. (84)</td>
<td>67</td>
</tr>
<tr>
<td>lossifov et al. (343)</td>
<td>294</td>
</tr>
<tr>
<td>lossifov et al. (343)</td>
<td>305</td>
</tr>
<tr>
<td>Neale et al. (175)</td>
<td>167</td>
</tr>
<tr>
<td>Gulsuner et al. (105)</td>
<td>103</td>
</tr>
<tr>
<td>Rauch et al. (20)</td>
<td>23</td>
</tr>
<tr>
<td>Jiang et al. (32)</td>
<td>38</td>
</tr>
<tr>
<td>Epik. (264)</td>
<td>329</td>
</tr>
<tr>
<td>O’Roak et al. (189)</td>
<td>242</td>
</tr>
<tr>
<td>Rauch et al. (51)</td>
<td>85</td>
</tr>
<tr>
<td>This study (50)</td>
<td>84</td>
</tr>
</tbody>
</table>
Are these *de novo* SNVs related to ID?

- Significantly more *de novo* mutations in known/candidate ID genes
- Significantly more loss-of-function mutations ($P = 4.8 \times 10^{-6}$, $P = 0.02$)
De novo structural variants, example 1

- A patient with the clinical suspicion of Rett syndrome
- MECP2 gene tested by Sanger sequencing but no mutations identified

MECP2, exon 4 (~1000 bp)
De novo structural variants, example 1

- A patient with the clinical suspicion of Rett syndrome
- MECP2 gene tested by Sanger sequencing but no mutations identified

**MECP2, exon 4 (~1000 bp)**

```
t g a a c a a t g t c t t g c g c t c t c c c c c c c c c c c c c g g t g t t t c g c t t c c t g c c g g g g c g t t t g a t c a
```
Example 2: A *de novo* duplication on chromosome 4

- ~60 kb *de novo* duplication on chromosome 4
- Affecting the last 6 exons of the *TENM3* gene
- *TENM3* is associated with coloboma, and microphthalmia
Duplication from chr4 to chrX
Duplication from chr4 to chrX

Only possible with genome sequencing!
## Diagnostic yield genome sequencing

### Highly likely diagnosis

<table>
<thead>
<tr>
<th>SNV</th>
<th>SV</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBR1 (2x)</td>
<td>SHANK3</td>
</tr>
<tr>
<td>WDR45</td>
<td>VPS13B*</td>
</tr>
<tr>
<td>SMC1A</td>
<td>MECP2</td>
</tr>
<tr>
<td>SPTAN1</td>
<td>IQSEC2</td>
</tr>
<tr>
<td>RAI1</td>
<td>STAG1</td>
</tr>
<tr>
<td>MED13L</td>
<td>SMC1A</td>
</tr>
<tr>
<td>SATB2</td>
<td>16p11.2 microdel. syndrome</td>
</tr>
<tr>
<td>PPP2R5D</td>
<td>Multiple genes</td>
</tr>
<tr>
<td>KCNA1</td>
<td></td>
</tr>
<tr>
<td>SCN2A</td>
<td></td>
</tr>
<tr>
<td>POGZ</td>
<td></td>
</tr>
<tr>
<td>KANSL2</td>
<td></td>
</tr>
</tbody>
</table>

*Recessive variants. Known genes in bold.

1/50 cases diagnosed: 42%
Highly likely diagnosis\(^1\)

<table>
<thead>
<tr>
<th>SNV</th>
<th>SV</th>
<th>Possible diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBR1 (2x)</td>
<td>SHANK3</td>
<td>NGFR</td>
</tr>
<tr>
<td>WDR45</td>
<td>VPS13B*</td>
<td>GFPT2</td>
</tr>
<tr>
<td>SMC1A</td>
<td>MECP2</td>
<td>WWP2</td>
</tr>
<tr>
<td>SPTAN1</td>
<td>IQSEC2</td>
<td>ASUN</td>
</tr>
<tr>
<td>RAI1</td>
<td>STAG1</td>
<td>BRD3</td>
</tr>
<tr>
<td>MED13L</td>
<td>SMC1A</td>
<td>MAST1</td>
</tr>
<tr>
<td>SATB2</td>
<td>16p11.2 microdel. syndrome</td>
<td>APPL2</td>
</tr>
<tr>
<td>PPP2R5D</td>
<td>Multiple genes</td>
<td>NACC1</td>
</tr>
</tbody>
</table>

21/50 cases diagnosed: 42%

\(^1\) de Ligt et al. NEJM 2012.

*Recessive variants. Known genes in bold.
Diagnostic yield

- Single gene test: ~1-5%
- Genomic microarray: 11.6%
- Exome sequencing: 27%

% of ID patients with a diagnosis: <40%

Gilissen et al. Nature 2014
Diagnostic yield

1,500 ID patients

- Single gene test: ~1-5%
- Genomic microarray: 11.6%
- Exome sequencing: 27%
- Whole genome sequencing: 42%

62% of ID patients with a diagnosis

Gilissen et al. Nature 2014
Diagnostic yield

- Single gene test: ~1-5%
- Genomic microarray: 11.6%
- Exome sequencing: 27%
- Whole genome sequencing: 42%
- ±1,500 ID patients

% of ID patients with a diagnosis: 62%

- Majority is de novo!

- No diagnosis
- De novo SNVs
- De novo SVs
- Inherited

Gilissen et al. Nature 2014
Conclusions

1. Exome sequencing results in a higher diagnostic yield for genetically heterogeneous diseases than Sanger-based approaches.

2. Based on clinical diagnostic criteria we can provide a genetic diagnosis for the majority of severe ID cases by using WGS.

1. *De novo* coding mutations are the major cause of severe ID.
Discussion

Other advantages of genome sequencing:

• Comprehensive set of variants: novel genes/exons annotations are still being added

• Non-coding variants: *de novo* mutations introducing novel splice sites

• Mosaic variants: variants occurring in sub-population of cells

• Phasing: identifying whether two variants are on the same allele
One more coding *de novo* variant?
One more coding *de novo* variant?
One more coding *de novo* variant?

- Mutations in the original *RARb* transcript are a known cause for Microphthalmia.
- Validation confirmed: coding *de novo* mutation #85.
- Expression analysis: this specific exon is expressed in brain / fetal brain.
Diagnostic exome sequencing in 100 ID trios

Total dedicated gene tests: 233
Number of different genes: 56
Other tests performed: 100 SNP microarrays
79 metabolic screens
34 brain MRIs
4 cerebral CT scans

<table>
<thead>
<tr>
<th>Level of ID</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ &lt;30</td>
<td>62</td>
</tr>
<tr>
<td>IQ 30-50</td>
<td>38</td>
</tr>
<tr>
<td>IQ 50-70</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>47</td>
</tr>
<tr>
<td>Female</td>
<td>53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age groups</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10 yrs</td>
<td>37</td>
</tr>
<tr>
<td>10-20 yrs</td>
<td>41</td>
</tr>
<tr>
<td>&gt;20 yrs</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sibship size</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>≥5</td>
<td>2</td>
</tr>
<tr>
<td>unknown</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of major congenital anomalies</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>62</td>
</tr>
<tr>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>