Technological Advances in Next Generation Sequencing

Joris Veltman, PhD
Department of Human Genetics
Radboud University Nijmegen Medical Centre
Nijmegen, The Netherlands
j.veltman@gen.umcn.nl
Challenges in genetic diagnostics

- Clinically diagnosing rare genetic disease is an art, ordering the right genetic test is difficult
- Single gene tests are laborious to set-up & expensive
- Diseases can be caused by different types of genetic variation, requiring different tests
- Genetic cause 1000s of rare diseases is unknown
- Common diseases are genetically heterogeneous & their genetic causes are largely unknown

No genetic diagnosis for majority of diseases  Role of genetics in medicine is limited

Need for simple, cheap & effective genetic diagnosis
Next generation sequencing: Simple, cheap & effective?

DNA from blood

Genome sequence with all variation

Important:
- Accuracy
- Speed
- Price
DNA sequencing becomes cheaper fast!

Sequencing output increases dramatically

Output per instrument run

<table>
<thead>
<tr>
<th>Year</th>
<th>Output (kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>10^3</td>
</tr>
<tr>
<td>2002</td>
<td>10^3</td>
</tr>
<tr>
<td>2003</td>
<td>10^6</td>
</tr>
<tr>
<td>2004</td>
<td>10^8</td>
</tr>
<tr>
<td>2005</td>
<td>10^10</td>
</tr>
<tr>
<td>2006</td>
<td>10^12</td>
</tr>
<tr>
<td>2007</td>
<td>10^14</td>
</tr>
</tbody>
</table>

Platforms

- ABI 3730xl capillary sequencer
- 454 GS-20 pyrosequencer
- Solexa/Illumina sequence analyser
- ABI SOLiD sequencer
- Roche/454 Titanium, Illumina GAII
- Illumina GAIIx, SOLiD 3.0
- Illumina Hi-Seq 2000

Event Timeline

- 2001: Draft human genome
- 2002: HapMap Project begins
- 2003: ENCODE Project begins
- 2004: 1,000 Genomes, Human Microbiome projects begin
- 2005: ENCODE Project pilot publications
- 2006: First tumour: normal genome publication
- 2007: Watson genome publication
- 2008: 1,000 Genomes pilot and HapMap3 publications
- 2009: Human genetic syndromes publications
Next generation sequencing equipment

- SOLiD 4
- Ion Torrent
- Pacific biosciences
- HiSeq
- 454
- Complete Genomics
### And there is much more to come....

1. **Illumina**  
   SbP Fluorescent read length (2x110bp)

2. **AB-SOLiD**  
   SbL Longest ligation reads (2x 70bp)

3. **Complete Genomics**  
   SbL $2000 genome, colony grid (7x 10bp)

4. **Polonator**  
   SbL/P Open-source, $170k device, (2x 30bp)

5. **Roche 454**  
   SbP Long reads (900bp)

6. **Helicos**  
   SbP – SM High parallelism & quantitation (2x 30bp)

7. **Ion Torrent**  
   SbP $50K, small device, (100bp)

8. **Pacific Bio**  
   SbP – SM Long reads (>2kb)  
   **Available**

9. **Oxford Nanopore**  
   Pore-protein – SM small device  
   **Under development**

10. **GnuBio**  
    SbP picoliter droplets

11. **Halcyon**  
    EM – SM Long reads

12. **Visigen/StarLite**  
    SbP – SM Qdot-Pol-dNTP FRET

13. **Bionanomatrix**  
    SbP – SM Fluorescent mapping

14. **Intelligent Bio**  
    SbP hexagonal grid

15. **Nabsys**  
    Pore-SbH – SM small device

16. **IBM**  
    Pore Si – SM small device

17. **Genizon BioSci**  
    SbH in situ sequencing

18. **LightSpeed**  
    SbL 16x density, 10x speed

19. **ZS Genetics**  
    EM – SM Iodine labels  
    **Modified from George Church**

20. **Electronic BioSci**  
    Pore-protein – SM

21. **GE Global**  
    SbP – SM
Important characteristics of a technology

- Accuracy of sequence
- Ease of mapping/assembling sequence
- Sequence throughput / coverage

- Different applications
- Robust/straightforward laboratory protocols
- Ease of interpretation
- Costs per nucleotide
Exome sequencing

‘Exome’ (all exons of a genome)

~1% of the human genome

‘All’ coding sequences of a human genome (>180,000 exons), sequenced and analyzed in one experiment
Exome sequencing – exome-wide mutation detection

In liquid exome enrichment:
Agilent’s SureSelect

Next generation sequencing:
Life technologies 5500XL

16 exomes / system / week
Mapping and annotation of exome sequencing reads
Prioritization of variants causing genetic disease

- Number of coding variants:
  \(~ 20,000\)

- Private* variants affecting the protein sequence:
  \(~ 200\)

*Not detected in dbSNP or other control data. Careful here for recessive disease!

How to identify the 1 causative variant?
NGS-based disease gene identification strategies

Linkage based strategy

Homozygosity based strategy

Double-hit based strategy

Overlap based strategy

De novo based strategy

Candidate based strategy

Gilissen et al. EJHG 2012
Example: Detection of recurrent mutations in a clinically well-defined rare disease

Schinzel-Giedion Syndrome
Dominant sporadic disease

<table>
<thead>
<tr>
<th>Neurodevelopmental anomalies</th>
<th>Our study</th>
<th>Literature: 46 cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>6 f/7 m</td>
<td></td>
</tr>
<tr>
<td>Developmental delay</td>
<td>11/11</td>
<td>39/39</td>
</tr>
<tr>
<td>Seizures</td>
<td>12/13</td>
<td>32/35</td>
</tr>
<tr>
<td>Vision impairment</td>
<td>8/9</td>
<td>11/12</td>
</tr>
<tr>
<td>Hearing impairment</td>
<td>8/9</td>
<td>8/11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Craniofacial features</th>
<th>Our study</th>
<th>Literature: 46 cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large anterior fontanelle</td>
<td>10/12</td>
<td></td>
</tr>
<tr>
<td>Prominent forehead</td>
<td>12/13</td>
<td>43/43</td>
</tr>
<tr>
<td>Mid-face retraction</td>
<td>13/13</td>
<td>46/46</td>
</tr>
<tr>
<td>Hypertelorism</td>
<td>12/13</td>
<td></td>
</tr>
<tr>
<td>Short, upturned nose</td>
<td>13/13</td>
<td>40/42</td>
</tr>
<tr>
<td>Low-set ears</td>
<td>12/13</td>
<td>37/39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structural anomalies</th>
<th>Our study</th>
<th>Literature: 46 cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genital</td>
<td>13/13</td>
<td>35/38</td>
</tr>
<tr>
<td>Hydronephrosis or vesicoureteral reflux</td>
<td>13/13</td>
<td>42/45</td>
</tr>
<tr>
<td>Cardiac defect</td>
<td>7/13</td>
<td>20/35</td>
</tr>
<tr>
<td>Characteristic skeletal malformations</td>
<td>11/11</td>
<td>12/29</td>
</tr>
<tr>
<td>Choanal stenosis</td>
<td>4/13</td>
<td></td>
</tr>
</tbody>
</table>

## Disease gene identification: Filter & combine

<table>
<thead>
<tr>
<th>Variants</th>
<th>Patient 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total called</td>
<td>22,916</td>
</tr>
<tr>
<td>Exonic + SpliceSites(SS)</td>
<td>12,196</td>
</tr>
<tr>
<td>Non-synonymous (NS) + SS</td>
<td>5,556</td>
</tr>
<tr>
<td>Novel (not in dbSNP130)</td>
<td>405</td>
</tr>
<tr>
<td>Novel (not in in-house db)</td>
<td>299</td>
</tr>
</tbody>
</table>

Human Genetics Nijmegen
...and filter more!

<table>
<thead>
<tr>
<th>Variants</th>
<th>Patient 1</th>
<th>Patient 2</th>
<th>Patient 3</th>
<th>Patient 4</th>
<th>Mean</th>
<th>Genes with variants in all samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total called</td>
<td>22,916</td>
<td>22,602</td>
<td>22,152</td>
<td>19,528</td>
<td>21,800</td>
<td>4,735</td>
</tr>
<tr>
<td>Exonic + SpliceSites(SS)</td>
<td>12,196</td>
<td>12,255</td>
<td>11,796</td>
<td>10,498</td>
<td>11,686</td>
<td>3,331</td>
</tr>
<tr>
<td>Non-synonymous (NS) + SS</td>
<td>5,556</td>
<td>5,618</td>
<td>5,427</td>
<td>4,802</td>
<td>5,351</td>
<td>1,634</td>
</tr>
<tr>
<td>Novel (not in dbSNP130)</td>
<td>405</td>
<td>401</td>
<td>390</td>
<td>387</td>
<td>396</td>
<td>35</td>
</tr>
<tr>
<td>Novel (not in in-house db)</td>
<td>299</td>
<td>289</td>
<td>275</td>
<td>288</td>
<td>288</td>
<td>12</td>
</tr>
<tr>
<td>Novel (not in ~100 exomes)</td>
<td>180</td>
<td>186</td>
<td>154</td>
<td>172</td>
<td>173</td>
<td>1</td>
</tr>
</tbody>
</table>
De novo mutations in SETBP1
Conclusions from this study

- **SETBP1** mutations in 12/13 SGS patients (+20 new cases)
- Never observed in controls
- **All de novo**, all clustering to 11bp (4AAs)
- Overlapping CNVs cause different phenotypes
- Hypothesis: **gain-of-function** or dominant negative effect
- First dominant disease solved by exome sequencing
- Clinical collection and detailed phenotyping crucial

Hoischen et al. Nature Genetics 2010
What is the success rate of exome sequencing in disease gene identification?

- **Unbiased retrospective analysis**
  - Collect 39,424 pathogenic mutations in HGMD
  - Study coverage of mutated nucleotides in 50 exomes
  Results: ~80% of mutations likely to be detected
  - 6% of mutated nucleotides not covered
  - 14% covered <10-fold on average

- **Biased own experience**
  - 24 unsolved rare Mendelian disorders
  - Novel genes in 14 (58%), known disease gene mutations in 3 (13%)
  - Missed mutations in at least 1 disease because of lack of coverage:
    - **MLL2** for Kabuki syndrome

*Gilissen et al. EJHG 2012*
Exome sequencing revolutionizes rare disease gene identification!

Key:
- Notable publication
- Recessive disease gene identification by WES
- Dominant disease gene identification by WES

Gilissen et al Genome Biology 2011
Lessons learned

- Success rate of exome sequencing 70-80%, determined largely by the quality of clinical collection & sequencing
- Sporadic diseases have become amenable to genetic disease research, no need for families!
- De novo mutations; Important cause of sporadic disease
- Pathogenic mutations have disruptive effects:
  - Nonsense mutations (ASXL1)
  - Affect specific domains of a gene (SETBP1)
  - Missense mutations affect evolutionary conserved nucleotides (PhyloP>2.5)
Next step: Exome sequencing in diagnostics!

- We launched diagnostic exome sequencing in Sept. 2011
- Pilot study of 500 samples for 5 diseases:

<table>
<thead>
<tr>
<th>Disease</th>
<th>1st approach</th>
<th>2nd approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intellectual disability</td>
<td>Genomewide <em>de novo</em> mutations (trio-sequencing)</td>
<td>-</td>
</tr>
<tr>
<td>Movement disorders</td>
<td>Known gene list (in-silico enrichment)</td>
<td>Genomewide search for relevant candidates (no timeline)</td>
</tr>
<tr>
<td>Blindness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deafness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolic disorders</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In addition: Diagnostic exome sequencing in patients with undiagnosed disease!
Diagnostic workflow

Blindness, deafness, movement disorders, Mitochondrial disorders, hereditary cancer

Intellectual disability

DNA-sample

Exome sequencing

filter known genes

Confirmation diagnosis?

yes

no

Report

Exome analysis

Candidate gene

Candidate ID gene

Novel gene

“de novo” analysis

negative

Report

Known ID gene

Known ID gene

Report

Trio-analysis

Exome sequencing

Report
Clinical interpretation of exome data
Many challenges!

- How to distinguish pathogenic from benign variation?
- Clinical counselling complex for the many unique genetic findings: How sure are we that they cause disease?
- Best evidence for causality: Second patient with mutation in same gene!
- Additional challenge: “Incidental findings”
- Need for practical guidelines for clinical interpretation!
- Evaluate clinical utility of exome sequencing diagnostics
Important: Diagnostic NGS software
All individuals must agree with the entire procedure.

All individuals must understand the possibility of unsought for findings and agree to be informed.

Unsought for findings will be assessed by independent expert committee.
Conclusions

Next generation sequencing revolution:

Rare disease gene identification made easy.
Genetic causes of most rare diseases will be resolved soon!

From paradox to paradigm:

De novo mutations are a common cause of reproductively lethal disorders, can be detected by trio-based sequencing.

From research to diagnostics:

Exome sequencing to be used as generic test in diagnostics.
Improvements in next generation sequencing technology are of great importance for diagnostic applications.
ALL COLLABORATORS WORLDWIDE
ALL PATIENTS & PARENTS!

Han Brunner!
Next generation genome diagnostics in pediatric neurology; Improving patient care and management

- 100 consecutive patients with undiagnosed disease entering the pediatric neurology department of our UMC
- Perform conventional diagnosis and exome sequencing in separated arms of the project
- Compare diagnostic yield after one year
- Compare costs
- Compare patient/family experience