Applications of Next Generation DNA Sequencing in Newborn Screening

Anne Goodeve
Sheffield Diagnostic Genetics Service
10th July 2014
Outline

Why undertake genetic analysis?

Sanger sequencing

Next generation sequencing

NGS for NBS project plan
Why undertake genetic analysis?

Definitive disease diagnosis/exclusion

Prognosis and management

Determine inheritance and disease risk in family members
Why undertake genetic analysis?

- Autosomal dominant?
- Autosomal recessive?
- X-linked recessive?
Genetic analysis

Gene of interest

26 exons and flanking introns ~25bp

Examine sequence for point mutations

Examine sequence for large deletions & duplications
Current Sanger DNA sequencing workflow

1. Blood sample
2. Genomic DNA
3. PCR amplification
4. DNA sequence
5. Data analysis

Provides information on point mutations
Sanger DNA sequencing

PCR amplification

DNA sequencing
Sanger DNA sequencing
Sanger DNA sequencing

Follow by bioinformatic analysis to determine which sequence variants may be disease associated
Changes in DNA sequencing technology

Sanger sequencing $\sim 3 \times 10^4$ bases

Next generation sequencing $\sim 3 \times 10^9$ bases
Next generation DNA sequencing

Massively parallel DNA sequencing
Many patients samples can be analysed together
Whole exome/genome analysis possible using larger capacity instruments
Workflows

Sanger sequencing

1. Genomic DNA
2. PCR amplification
3. Sequence
4. Analysis

Next generation sequencing

1. Genomic DNA
   → Sheared long range PCR
   → Clonal amplification
   → Sequence
   → Analysis

   → Tiled small amplicons
   → Sheared genomic DNA

Sequencing from sheared genomic DNA

Sheared genomic DNA for genes of interest selected by probe hybridisation

Hybridisation probes

Genes of interest selected by hybridisation
Sequencing from sheared genomic DNA

Indexing DNA enables association of results with correct patient.

Indexed & selected sheared genomic DNA

Patient 1

Patient 2

Patient 3
Sequencing from sheared genomic DNA

Aligned sequencing data

Patient 1

Patient 2

Patient 3

Sequence variant
Sequencing from sheared genomic DNA

Sequence coverage of exons for gene of interest

Diagnostic standard sequence coverage $\geq 30$ x / nucleotide
## Sequence output format

<table>
<thead>
<tr>
<th>Variant type</th>
<th>Gene (with HGVS)</th>
<th>1st check Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NM_001018112:exon9:c.710-12A&gt;G)</td>
<td></td>
</tr>
<tr>
<td>splicing</td>
<td>(NM_000135:exon33:c.3067-23G&gt;A,NM_001286167:exon33:c.3067-23G&gt;A)</td>
<td>SNP on Poly List</td>
</tr>
<tr>
<td>splicing</td>
<td>(NM_000135:exon33:c.3067-4T&gt;C,NM_001286167:exon33:c.3067-4T&gt;C)</td>
<td>SNP on Poly List</td>
</tr>
<tr>
<td></td>
<td>c.3263C&gt;T:p.S1088F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c.3654A&gt;G:p.P1218P</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c.3807G&gt;C:p.L1269L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c.3982A&gt;G:p.T1328A</td>
<td></td>
</tr>
</tbody>
</table>

All sequence variants identified listed

Manual check required to determine which if any may be pathogenic
Variant filtering workflow

1. Variants from sequencing
   - 500 variants

2. Variants within genes of interest
   - 265 variants

3. Remove frequent polymorphisms
   - 20 variants

4. Assess pathogenicity
   - SIFT
   - SNPs&GO
   - PolyPhen

5. Candidate mutation(s)
   - 4 variants
   - 2 variants
Large deletion detected by NGS

Ratio of read depth at first base of each exon of FANCA for patient compared to 7 controls

Heterozygous for ex 12_31del

Exon No.
Next generation sequencers

Life Technologies
- Ion PGM
  - 200-400 bp reads
  - 40 Mb to 1.5 Gb
  - 8 hours
- Ion Proton
  - 200 bp reads
  - Up to 10 Gb

Illumina
- MiSeq
  - 2x 250 bp reads
  - 8.5 Gb
  - 35 hours

Illumina
- MiSeq
  - 2x 250 bp reads
  - 8.5 Gb
  - 35 hours

Roche
- GS Junior
  - 400 bp reads
  - 28 Mb
  - 10 hours
- GS Flex Titanium
  - 700 bp reads
  - 0.7 Gb

Oxford Nanopore
- MinION
  - Average read 5.4 kb
  - Released 2014
  - In beta testing
Impact of NGS on genetic testing

Cost

Little impact on single gene disorders

Significantly reduced for large genes and for multigene disorders

Turnaround times

Initially most services 8 - 12 weeks for all genes

Potential for significant reduction
Newborn screening in the UK

5 current disorders;

Phenylketonuria (PKU)
Congenital hypothyroidism (CHT)
Sickle cell disease (SCD)
Cystic fibrosis (CF)
Medium chain acyl co-A dehydrogenase deficiency (MCADD)
Five pilot NBS disorders

Maple syrup urine disease (MSUD)
Homocystinuria (pyridoxine unresponsive) (HCU)
Isovaleric acidaemia (IVA)
Glutaric aciduria type 1 (GA1)
Long-chain hydroxyl acyl-CoA dehydrogenase deficiency (LCHADD)
Health Innovations Challenge Fund aims

Provide novel diagnostic tests or procedures

Permit timely diagnosis of conditions where no test currently exists

Offer solutions that can be readily integrated into and deployed widely across UK healthcare systems and beyond
Maple syrup urine disease

Birth

Dried blood spot

Result MSUD +ve

Clinical intervention

Day

0

5

7
Do no harm

Dietary management
Very little natural protein
Dietary supplements
Clinical monitoring & management
Lifelong intervention
Newborn screening

Biochemical analysis

Adjunct genetic testing

DNA sequencing

Unaffected

Screen positive

Genetic analysis to reduce ambiguity

Report

No. individuals

Analyte level
Genotype:phenotype correlation in Cystic Fibrosis
Aim 1

Expand the utility of adjunct genetic testing

Remove ambiguity

Enhance understanding of genotype : phenotype correlation

For pilot scheme disorders & MCADD
Healthy controls

Screen positive

Clinically affected

Genotype : phenotype database
Aim 2
Next generation DNA sequencing from a dried blood spot

For disorders where there is no biochemical marker suitable for newborn screening

The Challenge

Dried blood spot
+ve result
Clinical intervention

Birth

Day

0

5

7
Aim 2

Utilise healthy control individuals’ DNA

Compare DNA extracted from venous blood with DNA extracted from dried blood spots

Aim to obtain same sequence quality from dried blood spot DNA as from venous blood

Use current screened disorders to trial the analysis
Project outcome

Genotype : phenotype correlation $\uparrow$

Ambiguity $\downarrow$

Performance $\uparrow$ UK and worldwide programmes

Dried blood spots $\rightarrow$ DNA sequence

Enhanced sequencing pipeline for other clinical pathways and healthcare systems
The team

Steve Hannigan
CEO Climb
Patient advocate

Jim Bonham
National newborn laboratory screening lead

Mark Sharrard
Metabolic Physician
Metabolic team lead

Diana Johnson
Clinical Geneticist
Patient & family management

Darren Grafham
Head of Lab Services
NGS & technical management

Ann Dalton
Director SDGS
Genetics, links to NBS

Anne Goodeve
Research Lead Scientist
Research strategy

Sheffield Diagnostic Genetics Service
Sheffield Children's NHS Foundation Trust