# 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

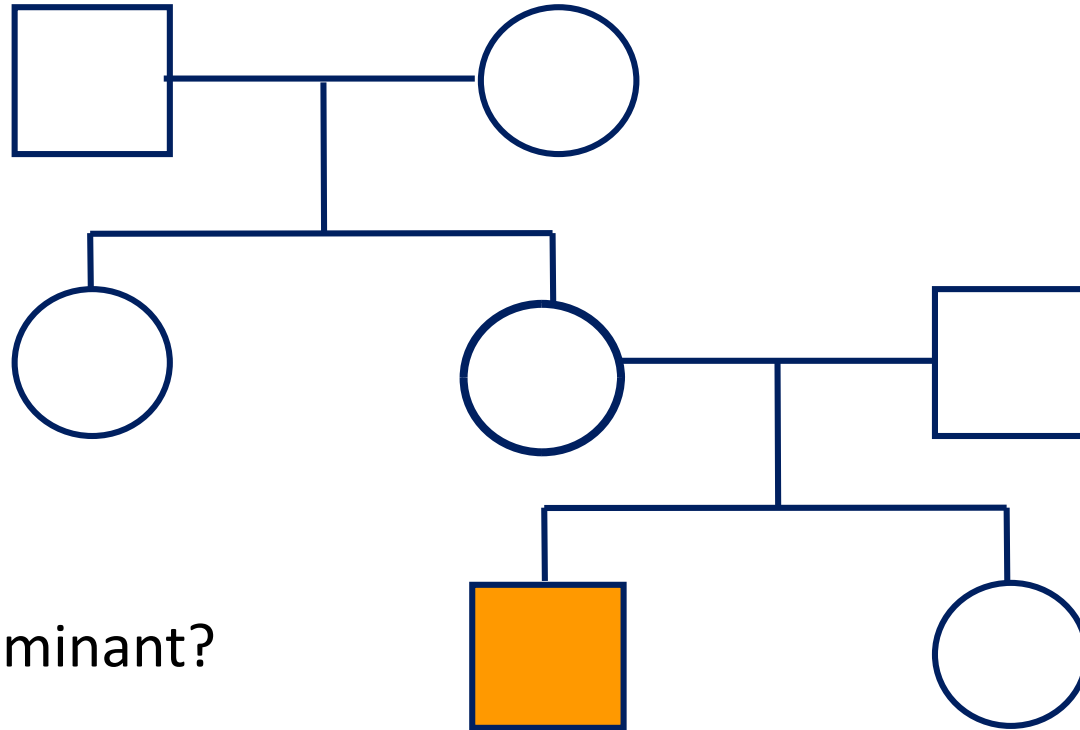Next generation sequencing

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**

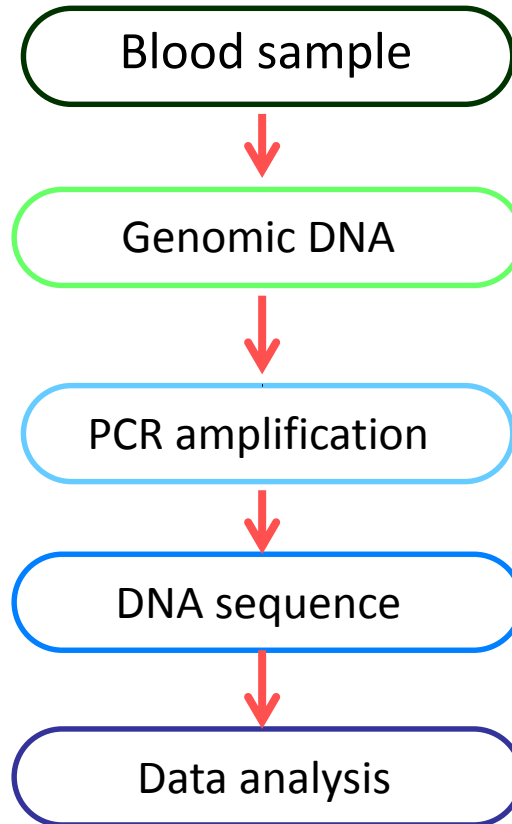| 0 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 180 kb |
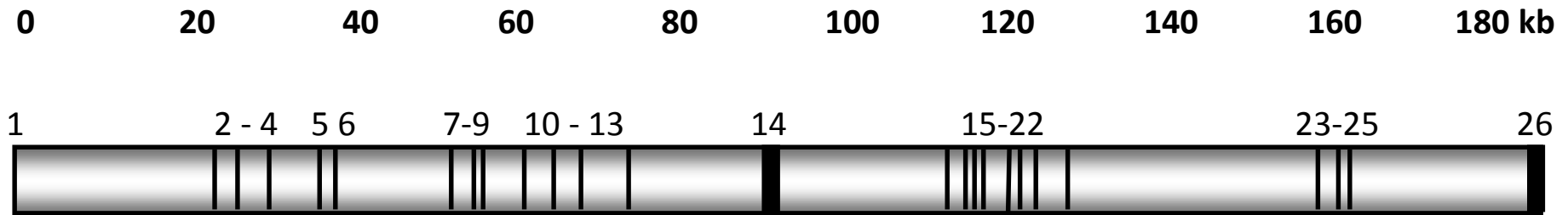


26 exons and flanking introns ~25bp

Examine sequence for point mutations

Examine sequence for large deletions & duplications

# Current Sanger DNA sequencing workflow

Blood sample

↓

Genomic DNA

↓

PCR amplification

↓

DNA sequence

↓

Data analysis

Provides information on point mutations

# Sanger DNA sequencing
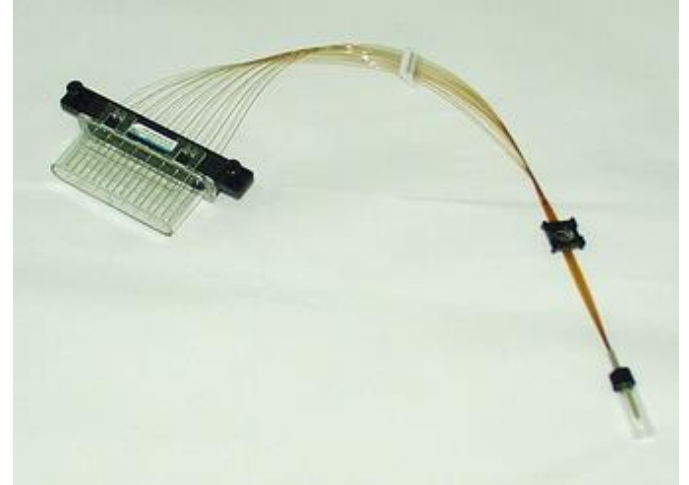


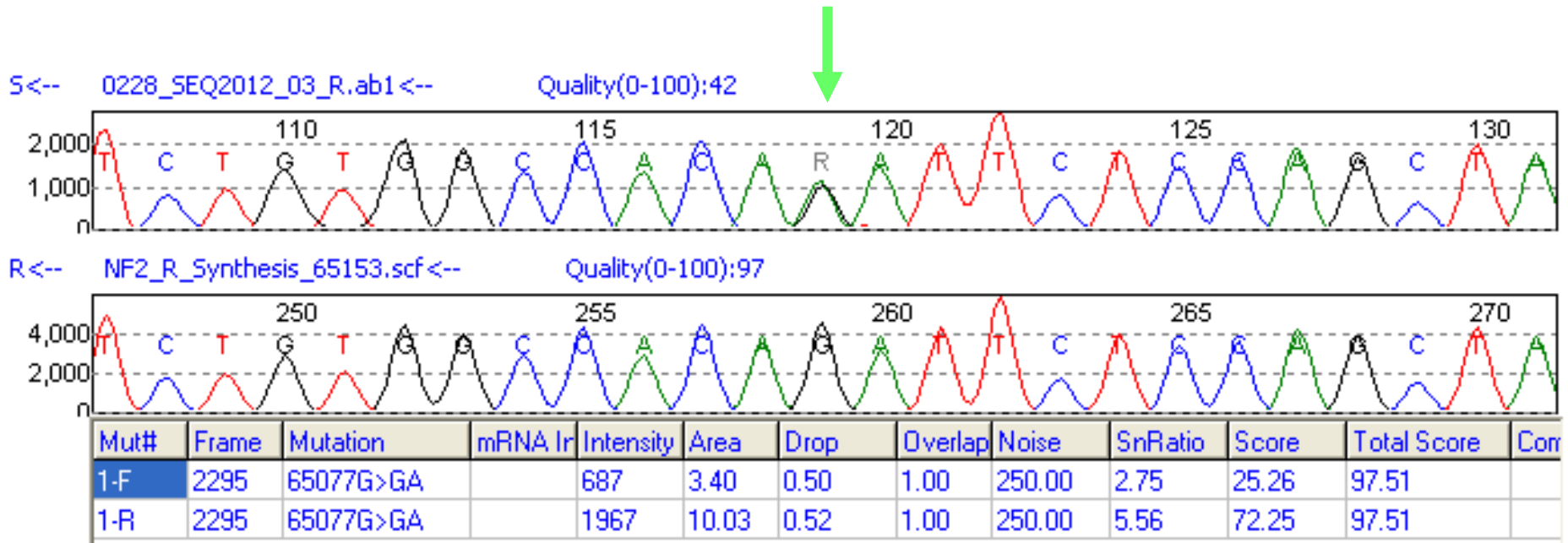Exon 1    Exon 2    Exon 3    Exon 4
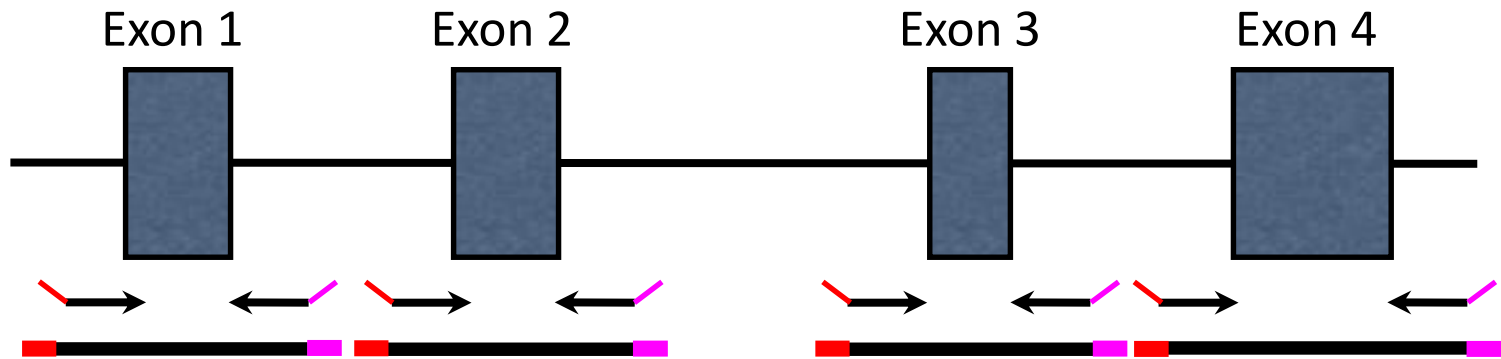
PCR amplification

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
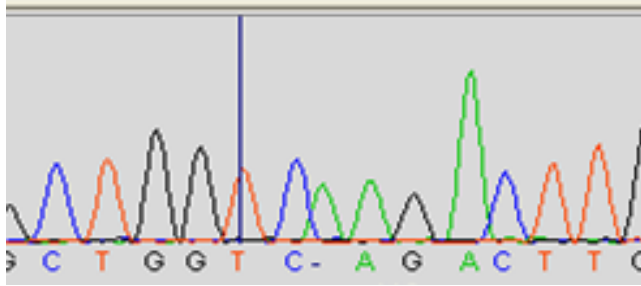


Sanger sequencing ~$3 \times 10^4$ bases

Next generation sequencing ~$3 \times 10^9$ bases

# Next generation DNA sequencing
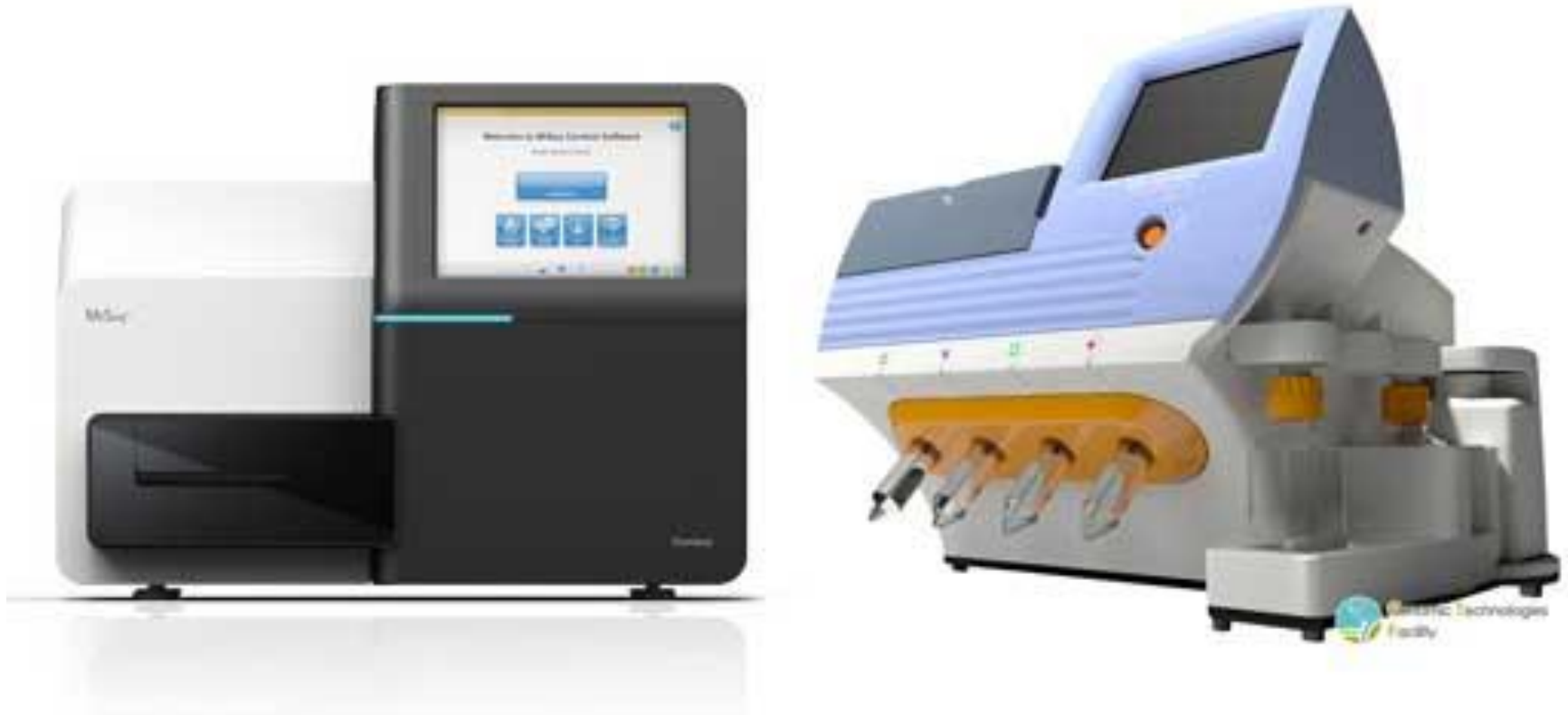


Massively parallel DNA sequencing
Many patients samples can be analysed together
Whole exome/genome analysis possible using larger
capacity instruments

# Workflows

## Sanger sequencing
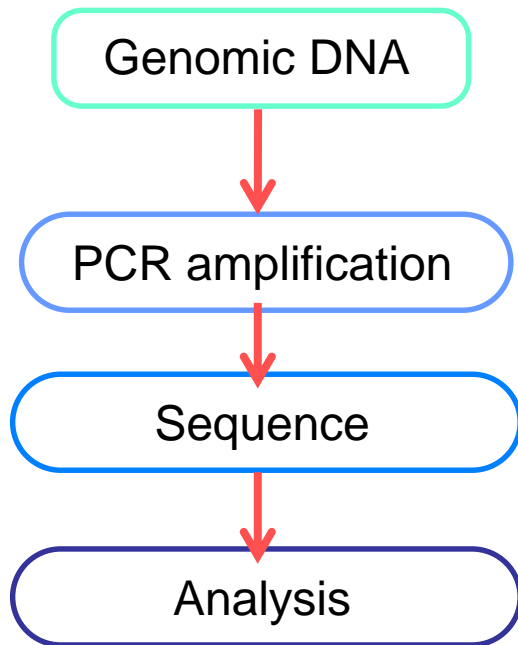
## Next generation sequencing

# Sequencing from sheared genomic DNA
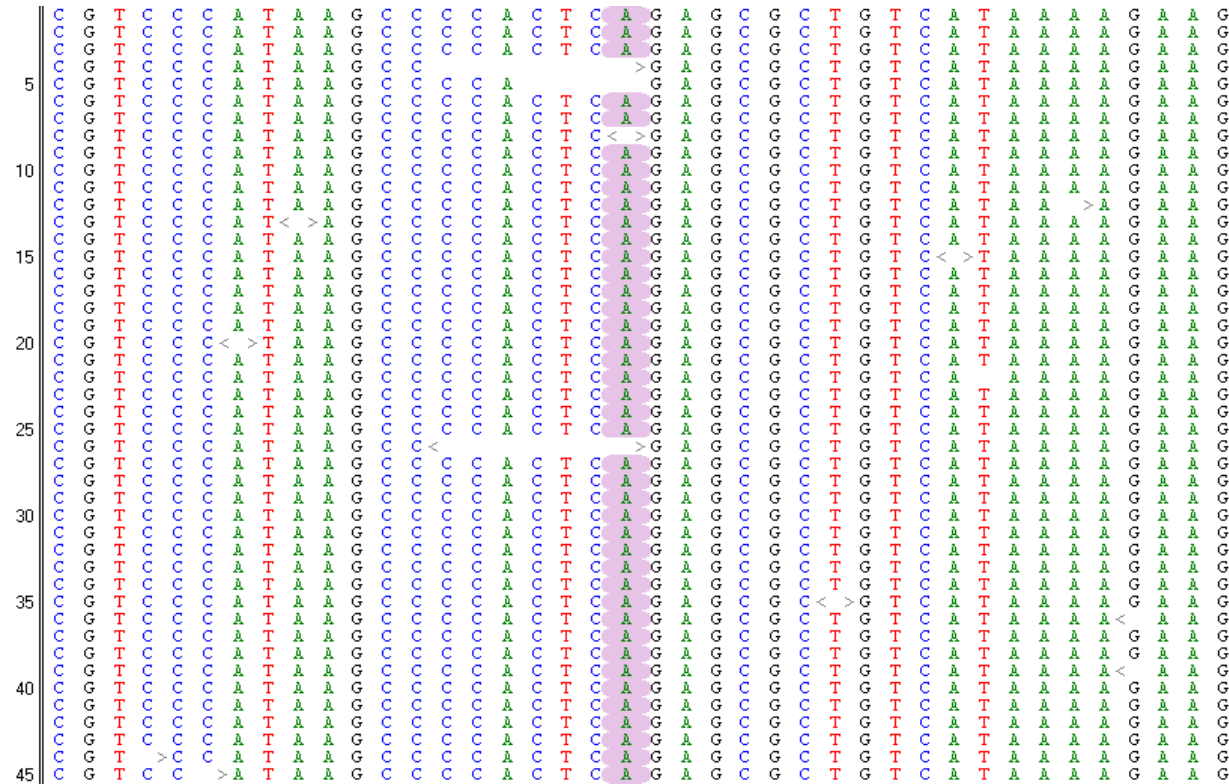
Gene A          Gene B          Gene C
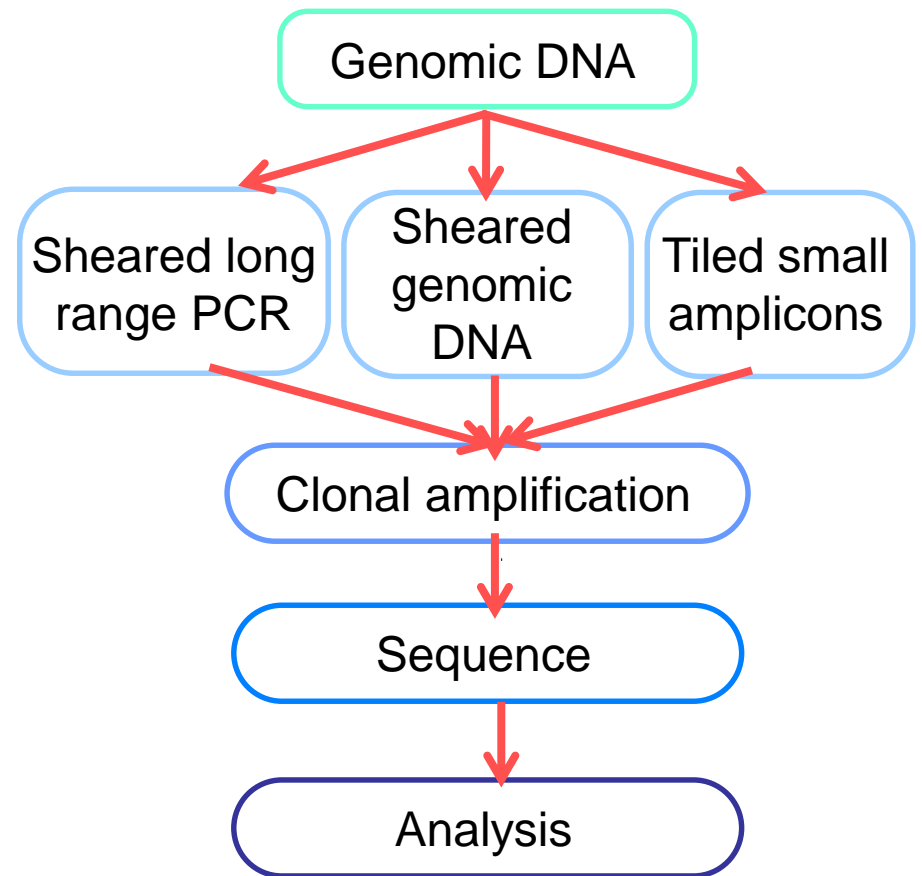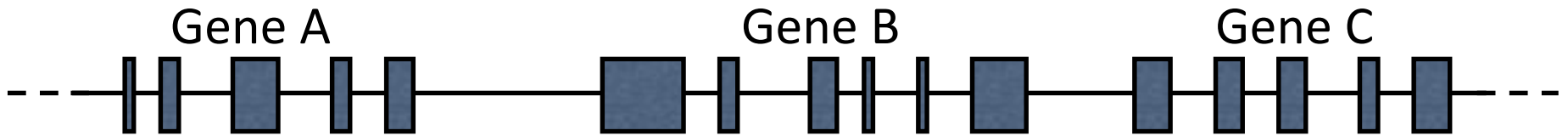
Sheared genomic DNA for genes of interest selected by probe hybridisation
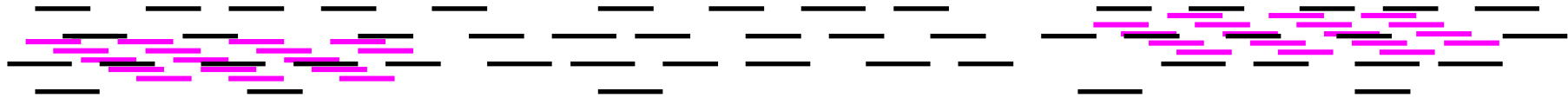
Hybridisation probes
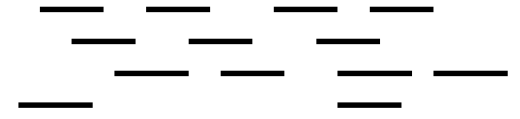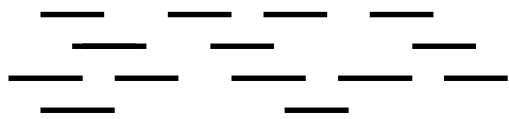
Genes of interest selected by hybridisation

# Sequencing from sheared genomic DNA

Indexing DNA enables association of results with correct patient

Indexed & selected
sheared genomic DNA

Patient 1

Patient 2
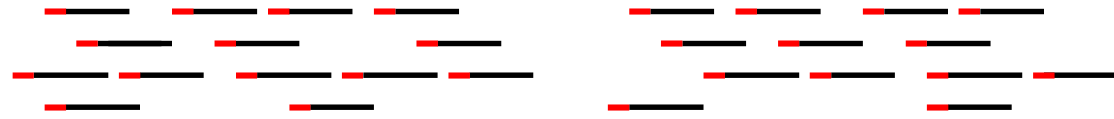
Patient 3

# Sequencing from sheared genomic DNA

A                  B                  C

Aligned sequencing data

Patient 1

Patient 2

Sequence variant

Patient 3
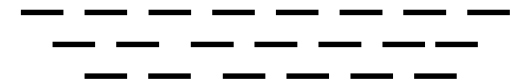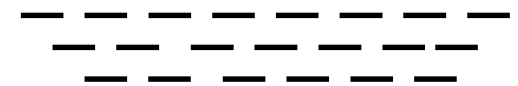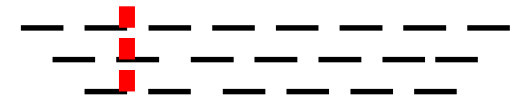
# Sequencing from sheared genomic DNA

Sequence coverage of exons for gene of interest
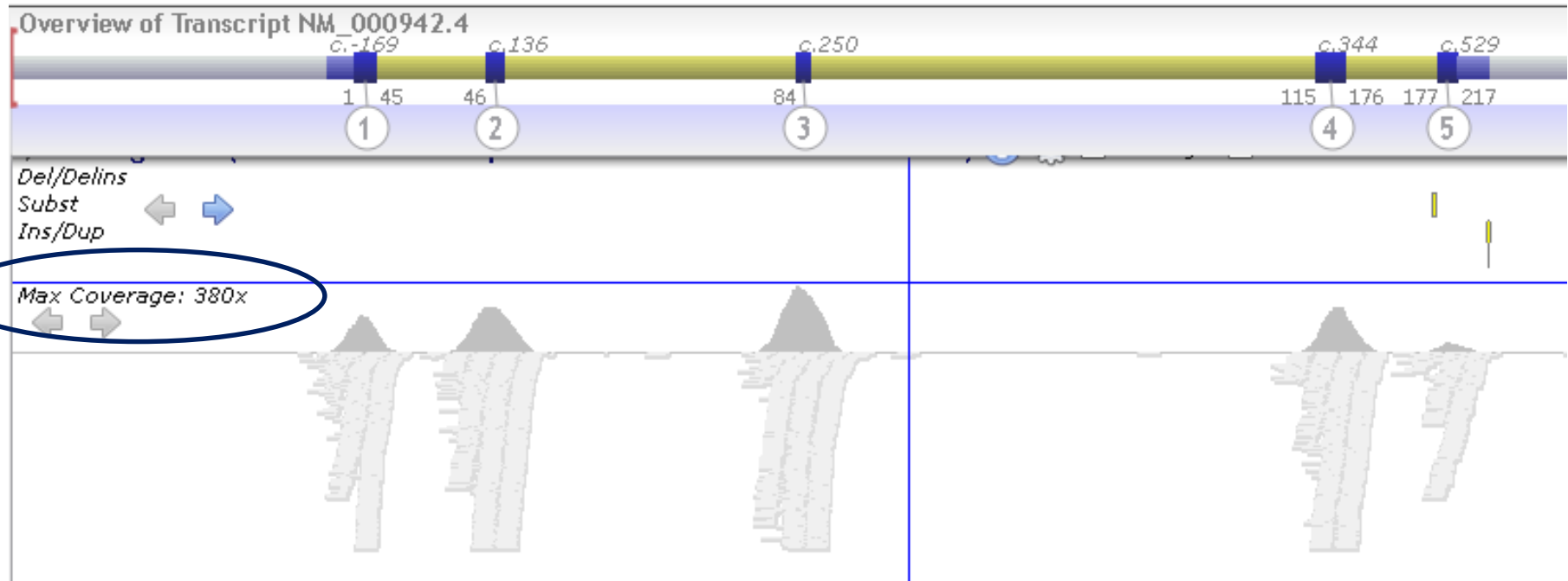


Diagnostic standard sequence coverage ≥30 x / nucleotide

# Sequence output format

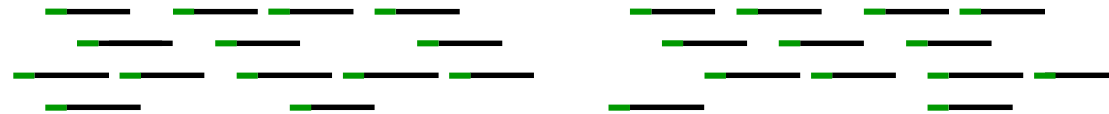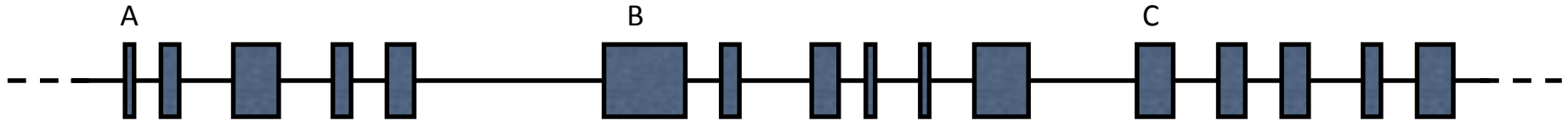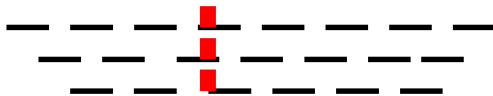| Variant type | Gene (with HGVS) | 1st check Comments |
|---|---|---|
| splicing | (NM_000135:exon9:c.710-12A>G,NM_001286167:exon9:c.710-12A>G, NM_001018112:exon9:c.710-12A>G) | SNP on Poly List |
| splicing | (NM_000135:exon12:c.894-8A>G,NM_001286167:exon12:c.894-8A>G) | SNP on Poly List |
| **splicing** | **(NM_000135:exon15:c.1226-2A>G,NM_001286167:exon15:c.1226-2A>G)** | **#:#** |
| splicing | (NM_000135:exon22:c.1900+24A>T,NM_001286167:exon22:c.1900+24A>T) | **Novel SNP placed on poly List** |
| splicing | (NM_000135:exon33:c.3067-23G>A,NM_001286167:exon33:c.3067-23G>A) | SNP on Poly List |
| splicing | (NM_000135:exon33:c.3067-4T>C,NM_001286167:exon33:c.3067-4T>C) | SNP on Poly List |
| nonsynonymous SNV | :NM_000135:exon33:c.3263C>T:p.S1088F,:NM_001286167:exon33: c.3263C>T:p.S1088F | SNP on Poly List |
| splicing | (NM_000135:exon34:c.3348+18A>G,NM_001286167:exon34:c.3348+18A>G) | SNP on Poly List |
| synonymous SNV | :NM_000135:exon37:c.3654A>G:p.P1218P,:NM_001286167:exon37: c.3654A>G:p.P1218P | SNP on Poly List |
| synonymous SNV | :NM_000135:exon38:c.3807G>C:p.L1269L,:NM_001286167:exon38: c.3807G>C:p.L1269L | SNP on Poly List |
| nonsynonymous SNV | :NM_000135:exon40:c.3982A>G:p.T1328A,:NM_001286167:exon40: c.3982A>G:p.T1328A | SNP on Poly List |

## All sequence variants identified listed

## Manual check required to determine which if any may be pathogenic

# Variant filtering workflow

| | |
|---|---|
| Variants from sequencing | 500 variants |
| Variants within genes of interest | 265 variants |
| Remove frequent polymorphisms | 20 variants |
| Assess pathogenicity — Variant Databases, SIFT, HSF, SNPs&GO, SpliceSiteFinder-Like, PolyPhen | 4 variants |
| Candidate mutation(s) | 2 variants |
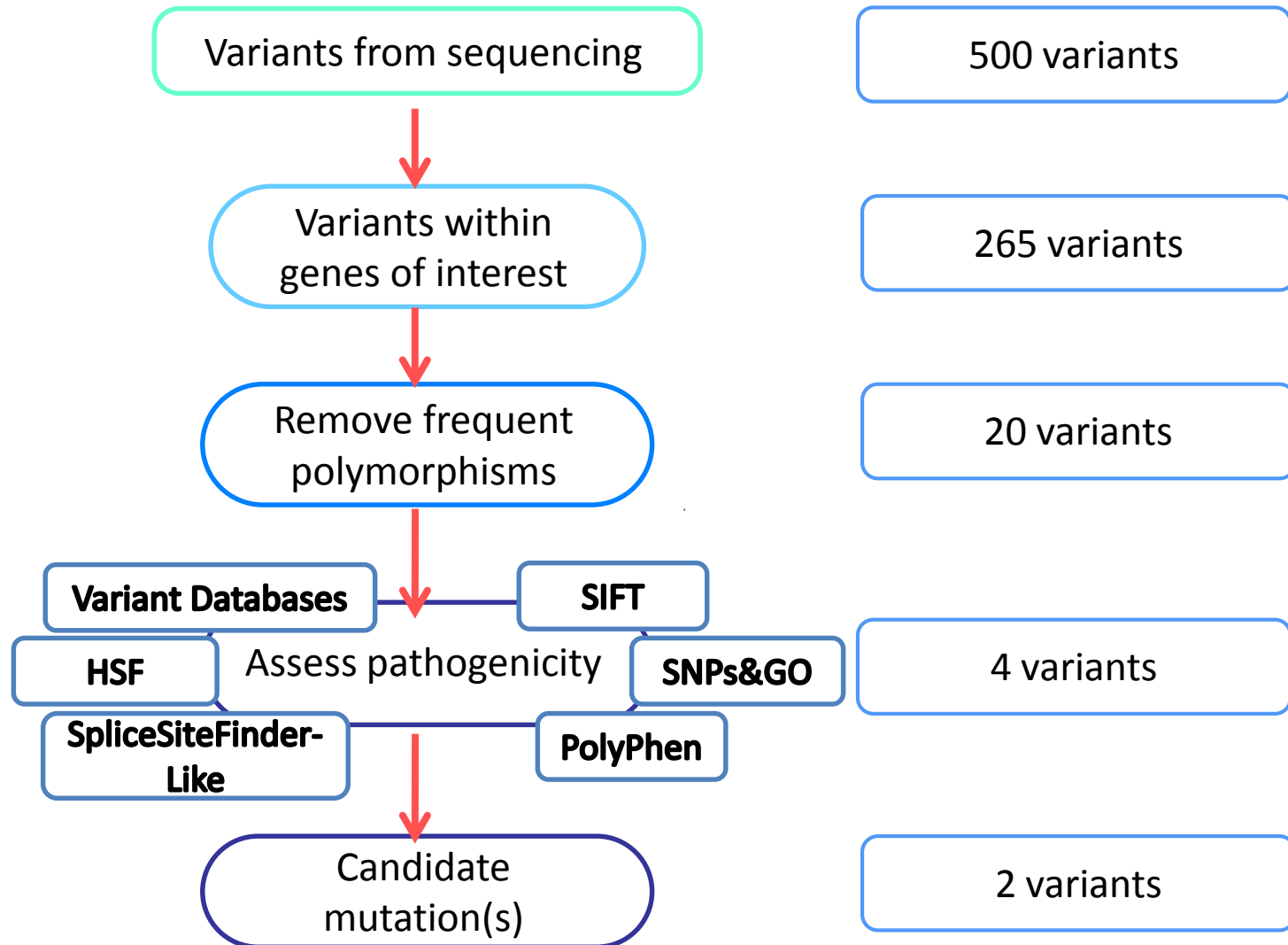
# Large deletion detected by NGS



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

Heterozygous for ex 12_31del

Read depth; patient / mean of 7 controls

Exon No.
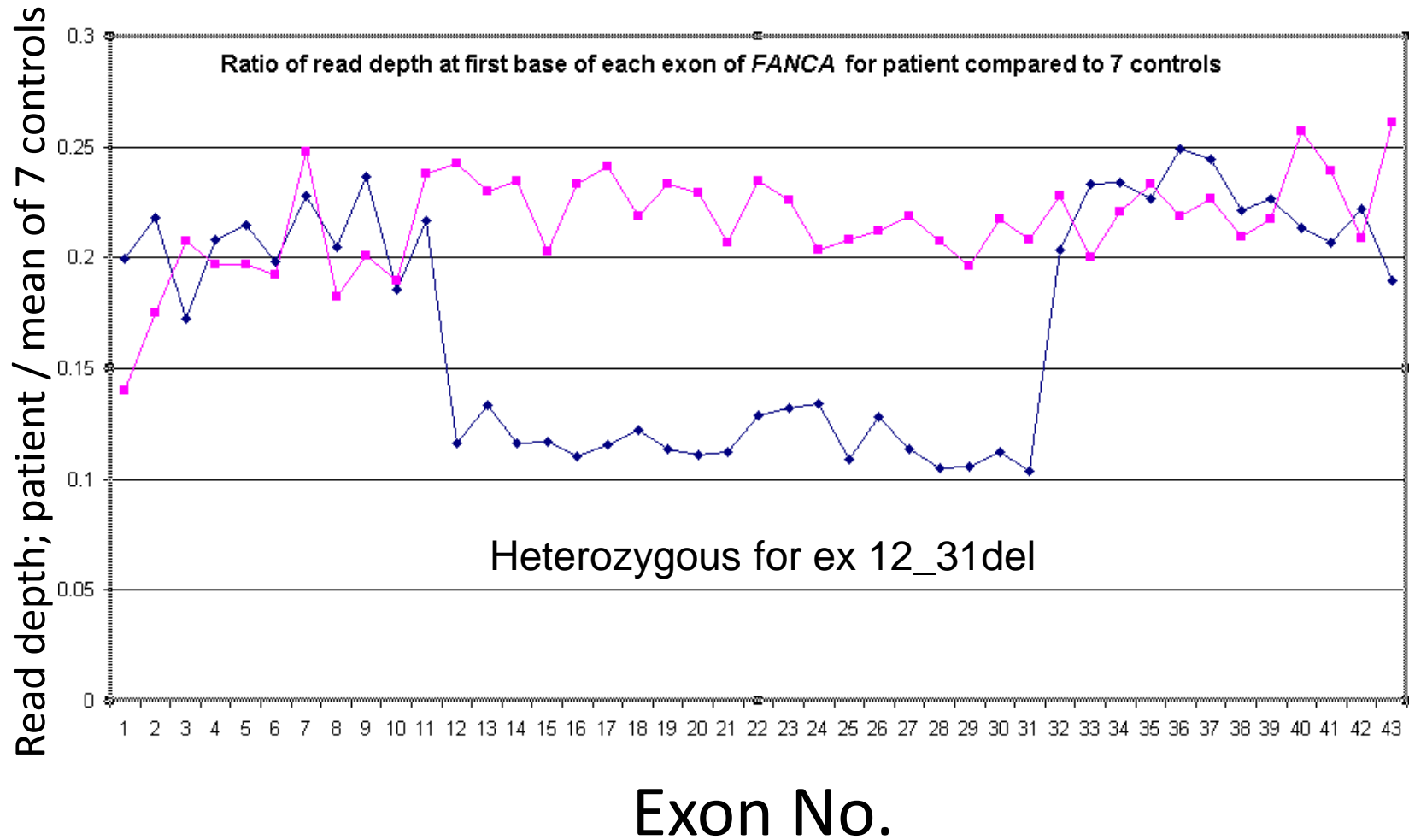
# Next generation sequencers



| Life Technologies | Illumina | Roche | Oxford Nanopore |
|---|---|---|---|
| Ion PGM | MiSeq | GS Junior | MinION |
| 200-400 bp reads | 2x 250 bp reads | 400 bp reads | Average read 5.4 kb |
| 40 Mb to 1.5 Gb | 8.5 Gb | 28 Mb | Released 2014 |
| 8 hours | 35 hours | 10 hours | In beta testing |
| Ion Proton | | GS Flex Titanium | |
| 200 bp reads | | 700 bp reads | |
| Up to 10 Gb | | 0.7 Gb | |

# 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)

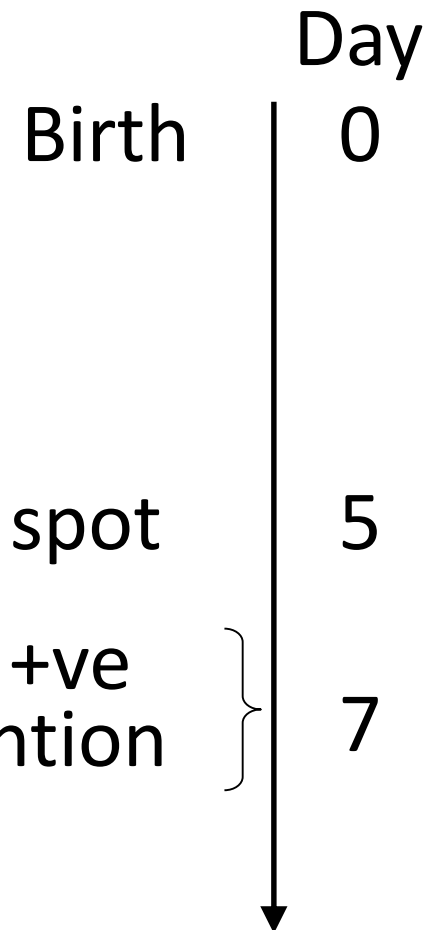# Health Innovations Challenge Fund aims

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



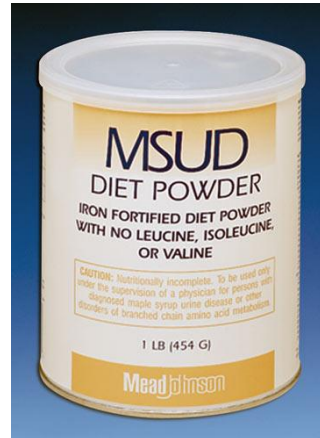| | Day |
|---|---|
| Birth | 0 |
| Dried blood spot | 5 |
| Result MSUD +ve Clinical intervention | 7 |

# Do no harm



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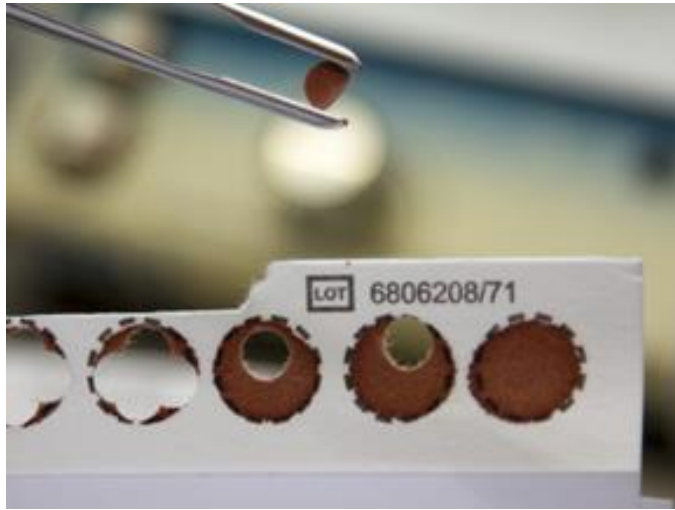  Dietary supplements
Clinical monitoring & management
Lifelong intervention



MSUD DIET POWDER
IRON FORTIFIED DIET POWDER WITH NO LEUCINE, ISOLEUCINE, OR VALINE
CAUTION: Nutritionally incomplete. To be used only under the supervision of a physician for persons with diagnosed maple syrup urine disease or other disorders of branched chain amino acid metabolism.
1 LB (454 G)
Mead Johnson
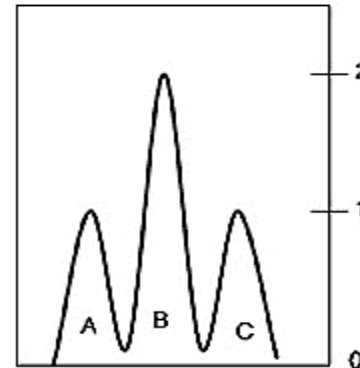
# Newborn screening



Biochemical analysis

Adjunct genetic testing

DNA sequencing

Report

Unaffected

Screen positive

No. individuals

Analyte level
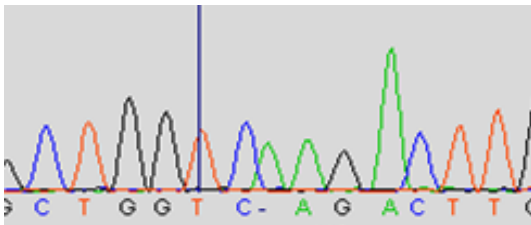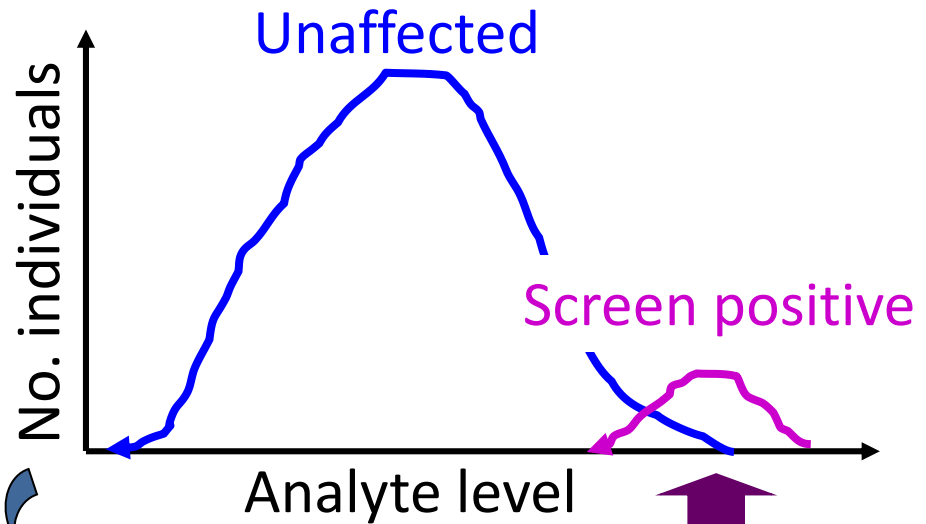
Genetic analysis to reduce ambiguity

F508del:F508del        F508del:R117H

- Lungs
- Pancreas
- Reproductive tract
- Shortened life span
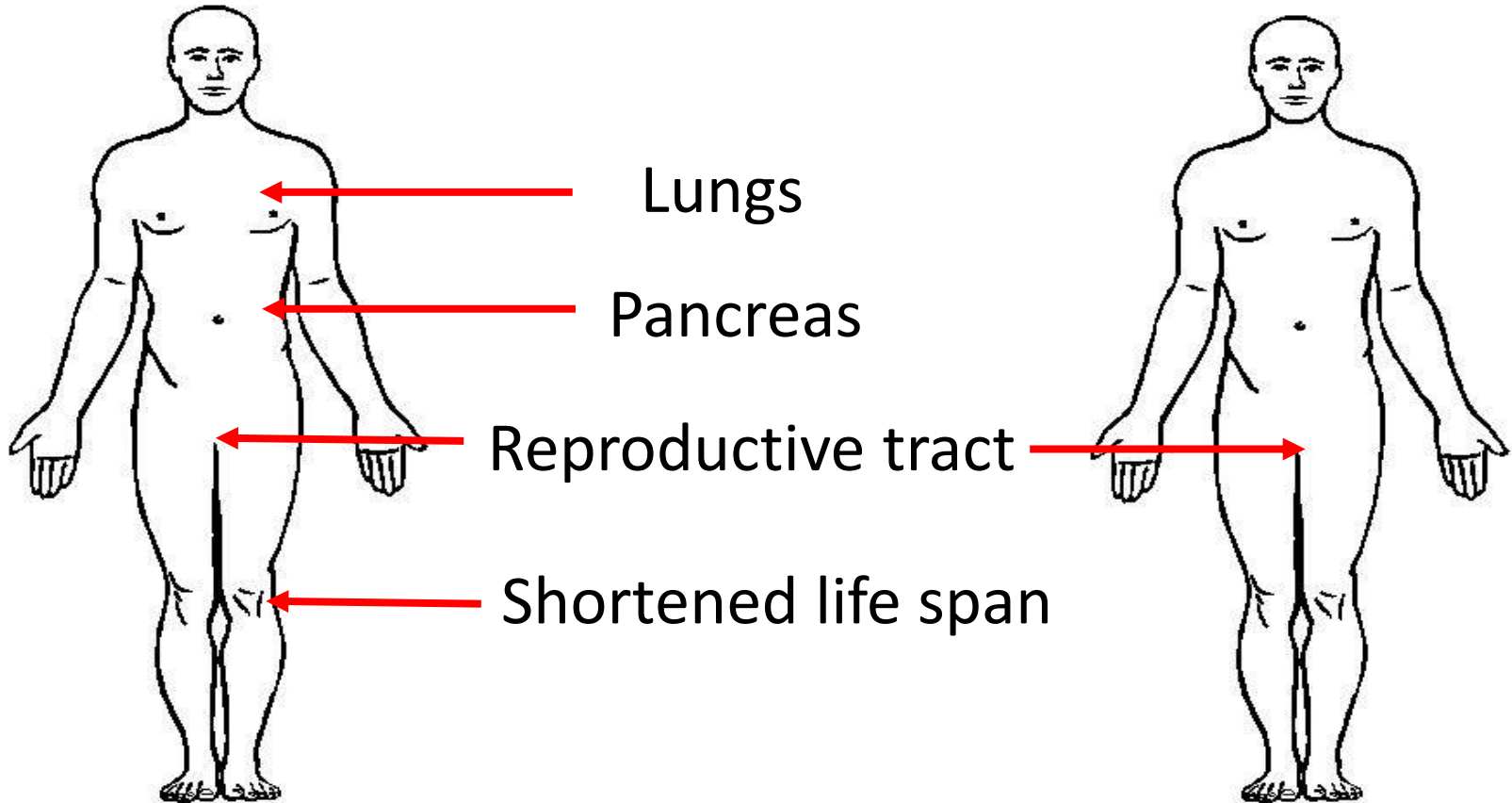
Genotype:phenotype correlation in Cystic Fibrosis

# 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

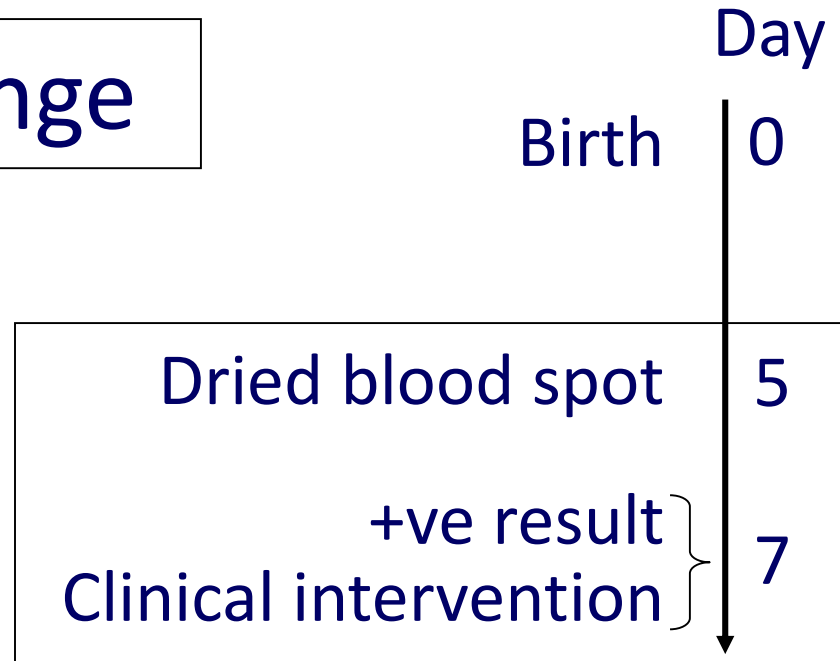Genotype : phenotype database

# 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

Day

Birth    0

Dried blood spot    5

+ve result
Clinical intervention    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

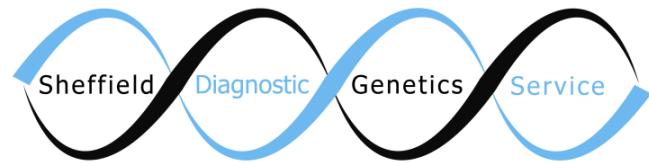Use current screened disorders to trial the analysis

# 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

Sheffield Diagnostic Genetics Service

Sheffield Children's NHS Foundation Trust

**Ann Dalton**
Director SDGS
Genetics, links to NBS

**Anne Goodeve**
Research Lead Scientist
Research strategy

**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