# Clonal sequencing technologies

**Project leader: Chris Mattocks**
**chris.mattocks@salisbury.nhs.uk**

## Introduction

The new generation of sequencing technologies have the potential to profoundly impact genetic diagnostics. Novel methodologies in conjunction with massive parallelisation allow huge capacity at the same time reducing the per-base cost of sequencing. The availability of such high capacity facilitates a wide range of applications from mutation scanning to investigation of genome architecture. Although these technologies have been easily adopted into research, diagnostic implementation will not be so straightforward. The NGRL(Wessex) projects on 'next generation' sequencing are targeted at some of the key issues related to diagnostic implementation.

## Technologies

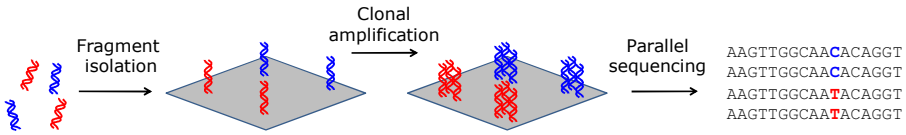
Whilst the methodologies vary in detail, they are broadly based on the same principle. After fragmentation of the DNA sample, the resulting fragments are spatially separated and sequenced in parallel generating millions to hundreds of millions of sequence reads varying in length depending on the technology used. To generate detectable signal, second generation technologies require the DNA fragments to be clonally amplified before the sequencing stage. A third generation of technologies, which permit direct sequencing of single DNA molecules, promise a further step change in capacity, speed and cost.

**Conventional Sanger sequencing:**

- Each sequencing reaction represents a single PCR product up to ~1Kb
- Analysis trace represents an averaged result from 000's molecules


**Second generation sequencing:**

- The entire input sample is sequenced
- Millions of small DNA fragments (50-500bp) are sequenced in parallel.
- Each sequence read represents a single starting DNA molecule




## Platforms

| Company | Platform | Run capacity (Gb) | Run Time | Throughput (Gb/day) | Cost (£/Mb)* |
|---|---|---|---|---|---|
| Applied Biosystems | 3730XL | ~0.0001 | 1 hr | 0.001 | ~3000.00 |
| Roche (454) | GS-FLX (Titanium) | 0.4 – 0.6 | 10 hrs | 1 | ~20 .00 (£10K/500Mb) |
| Illumina (Solexa) | Genome Analyzer (GAII) | 4.5 – 30 | 2 – 10 days | 2.3 – 3.5 | ≤1.60 (£8k/5000Mb) |
| Applied Biosystems | SOLiD 3 | 25 – 60 | 3.5 – 14 days | 2.5 – 3.5 | ≤1.00 (£10K/10000Mb) |
| Helicos | HeliScope | 21 – 28 | 8 days | 2.5 –3.5 | ≤0.90 (£18K/20000Mb) |

Capacities and run cost for the four main commercially available sequencing platforms with conventional capillary sequencing for comparison.
*Estimated from various sources (accessed September 2009) – consumable run cost only.

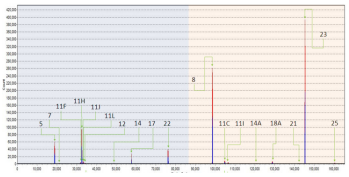
## Potential applications

Mutation scanning
Copy number analysis by counting single reads (array comparative hybridisation equivalent)
Genome architecture using paired end reads (including balanced variations)
Non-invasive prenatal diagnosis
Methylation/epigenetic analysis
Mitochondrial DNA analysis
Deep sequencing / tumour profiling
Expression analysis
cDNA (RNA) sequencing

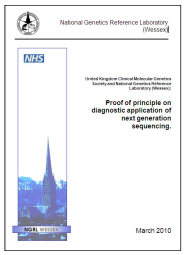
## NGRL projects

This is a field of rapidly developing technology – all work in this area is focused on developing (as far as possible) platform independent solutions.

**Targeting methodologies:** A key factor influencing effective diagnostic implementation of new sequencing technologies is the effective utilisation of capacity. Whilst array based targeting solutions are commercially available, these methods do not currently offer the flexibility or precision required for many diagnostic applications. We are working with a range of highly multiplexed, fluid phase targeting methodologies including circularisation, and concatemerisation of target molecules, that will be applicable to diagnostic mutation scanning.



Multiplex targeting of 18 BRCA1 & 2 exons by cirularisation

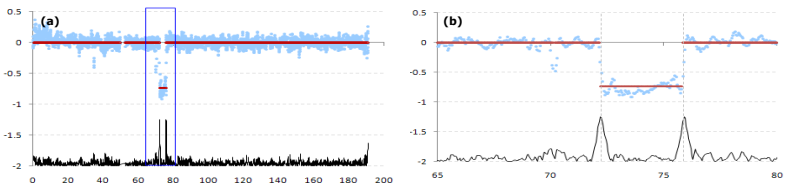



**CMGS Next generation sequencing evaluation:** NGRL(Wessex) are coordinating a proof of principle study on diagnostic application of next generation sequencing. The aims of this project are:

- To stimulate dialogue between suppliers of next generation sequencing technologies and the diagnostic community.
- To collect data on all aspects of sample processing and data analysis to enable development of suitable work-flows.
- To provide a background for collaborative development of diagnostically tailored applications.

Companies have been invited to propose a methodology for investigation of one hundred and forty samples with mutations in 7 different colorectal cancer genes. The most promising solutions will be taken forward to full analysis.

**CNV analysis:** The possibility of using sequencing to analyse CNVs and other structural variations is being investigated. Two general methodologies are applicable: direct counting of sequencing reads is directly equivalent to array CGH analysis, whereas paired end sequencing allows investigation of balanced anomalies.




SQ6 – Chromosome 4 deletion 72.2Mb – 75.9MB detected sequencing. x-axis = Mb on chr4, y-axis = log2(dosage quotient). (a) full chromosome view, (b) close-up

**If you would like to get involved in the CMGS evaluation, comment on our current projects or make suggestions for future work please contact Chris Mattocks at National Genetics Reference Laboratory (Wessex), Salisbury District Hospital, Odstock Road, Salisbury, SP2 8BJ. Email: chris.mattocks@salisbury.nhs.uk tel: 01722 429016, fax: 01722 338095**