# Diagnostic application of new sequencing technologies

Chris Mattocks
National Genetics Reference Laboratory (Wessex)

# Why new sequencing technologies?

- Increased speed of current tests
- Increased capacity for current tests
- Reduced cost
- Increased scope of current tests
- Development of new test areas

# Platforms

| Company | Platform | Pre-amplification | Basis of sequencing | Capacity |
|---|---|---|---|---|
| Roche | GS-FLX | emPCR | Pyrosequencing | $2x10^8$bp/run |
| Applied Biosystems | SOLiD | emPCR | Extension by ligation | $3x10^9$bp/run |
| Illumina (Solexa) | Genome Analyzer | bridge amplification | Reversable termination | $4x10^9$bp/run |
| Helicos (tSMS) | Heliscope | none | Reversable termination | $2x10^9$bp/day |
| Pacific Biosciences | | none | Single molecule real time sequencing (SMRT) | $>1x10^4$bp/sec |
| Visigen | | none | Real time FRET baseb identification | $>1x10^4$bp/sec |
| ZSG | | for labelling | Direct visualisation by TEM | ? |
| Various | Nanopore sequencing | none | Real time electronic base identification | $>1x10^4$bp/sec |
| Various others – GE Healthcare, Complete genomics, BioNanoMatrix / Agilent, IntellegentBioSystems, NABsys, Reveo …. | | | | |

# Potential applications

- Parallel sequencing amenable to a wide range of applications including:

Mutation scanning

RNA analysis

Genome architecture / structural analysis

Methylation analysis

Expression analysis

Tumour profiling / deep sequencing

Quantitative analysis / CNVs

# Factors for consideration

- Length of fragment sequenced
  - GS-FLX ~300bp
  - SOLiD / Genome Analyzer 25-30bp
- Methodology
  - GS-FLX: pyrosequencing not good for homopolymer regions
- Data / IT infrastructure
  - Amount of data generated and how analysed / stored
- Analysis
  - SOLiD: 2 base encoding very accurate (may reduce required depth)
- Amount of starting material
- Workflow
  - Breakdown between prep time / machine time

# Potential diagnostic strategies

- Whole genome sequencing - Sequence entire genome - analyse regions of interest

  - ☑ Simple / standard methodology for all referrals
  - ☑ Pre-emptive sequencing
  - ☑ Cost

  - ☒ Data volume
  - ☒ Who does the work?
  - ☒ How data handled / interpreted?
  - ☒ Incidental findings?
  - ☒ Unused data?
  - ☒ Not currently feasible

# Potential diagnostic strategies

- Universal targeting - Target particular region[s] of clinical interest - analyse regions relevant for patient

  - ☑ Capacity wastage limited
  - ☑ Currently feasible?

  - ☒ Choice of regions?
  - ☒ Utilisation of capacity?
  - ☒ Lack of flexibility
  - ☒ Shelf life?

# Potential diagnostic strategies

- Customised targeting - Specific targeting for each patient

  - ☑ Standardised methodology
  - ☑ No capacity wastage
  - ☑ Flexibility
  - ☑ Currently feasible?

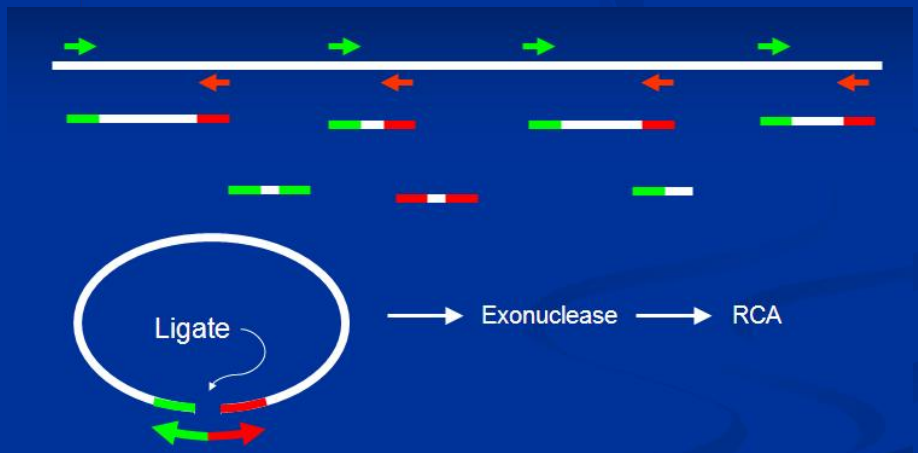
  - ☒ Developing methodology
  - ☒ Process control
  - ☒ Utilisation of capacity?
  - ☒ Shelf life?

# Targeting – PCR based

- Roche FLX
  - ~4000 PCRs (400,000 amplicons / 100 required depth)
  - ~40 x 96 well plates
- Illumina Genome analyser / AB SOLiD
  - ~48000 PCRs (~600 samples x 80 fragments)
  - ~480 plates
- All PCRs require:
  - Purification
  - Quantitation and normalisation
  - Mixing

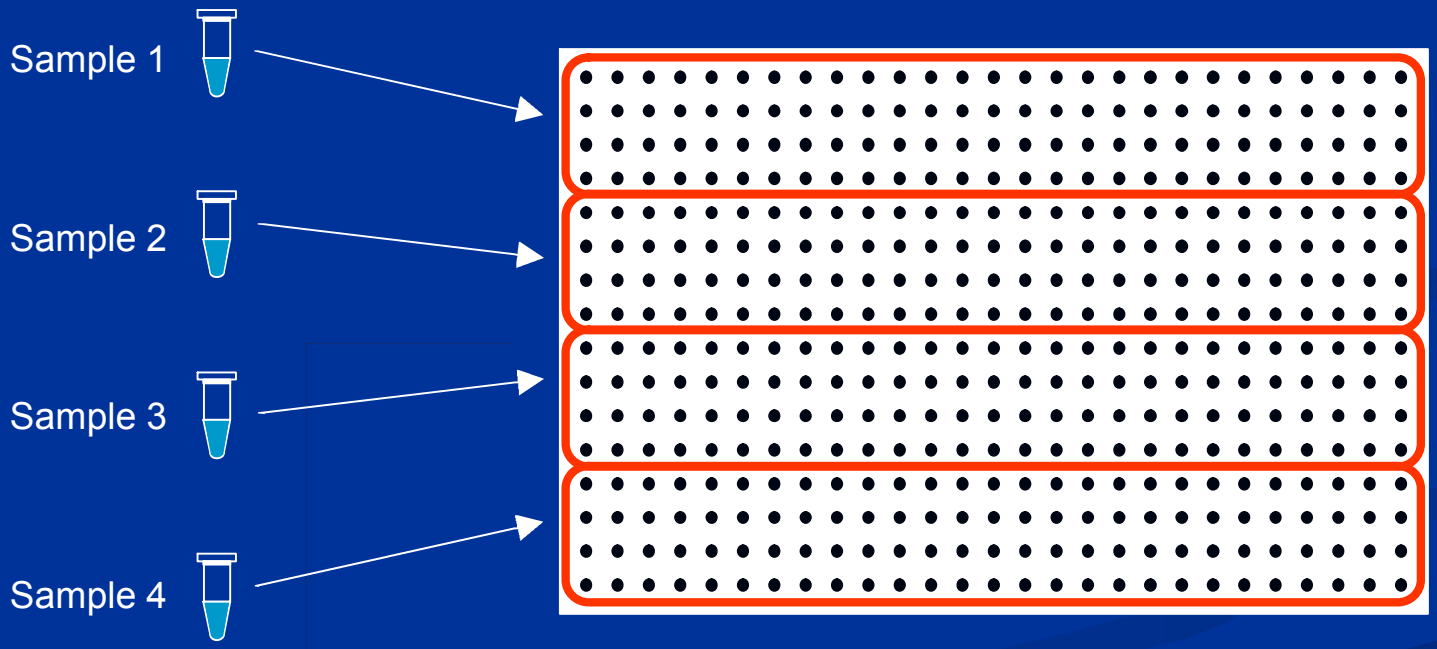
# Targeting - Alternatives

- Array based
  - Using standard arrays (Nimblegen)
- Liquid phase (probe based)
  - Circularisation
  - RNA probes (Agilent)
  - Fully flexible targeting

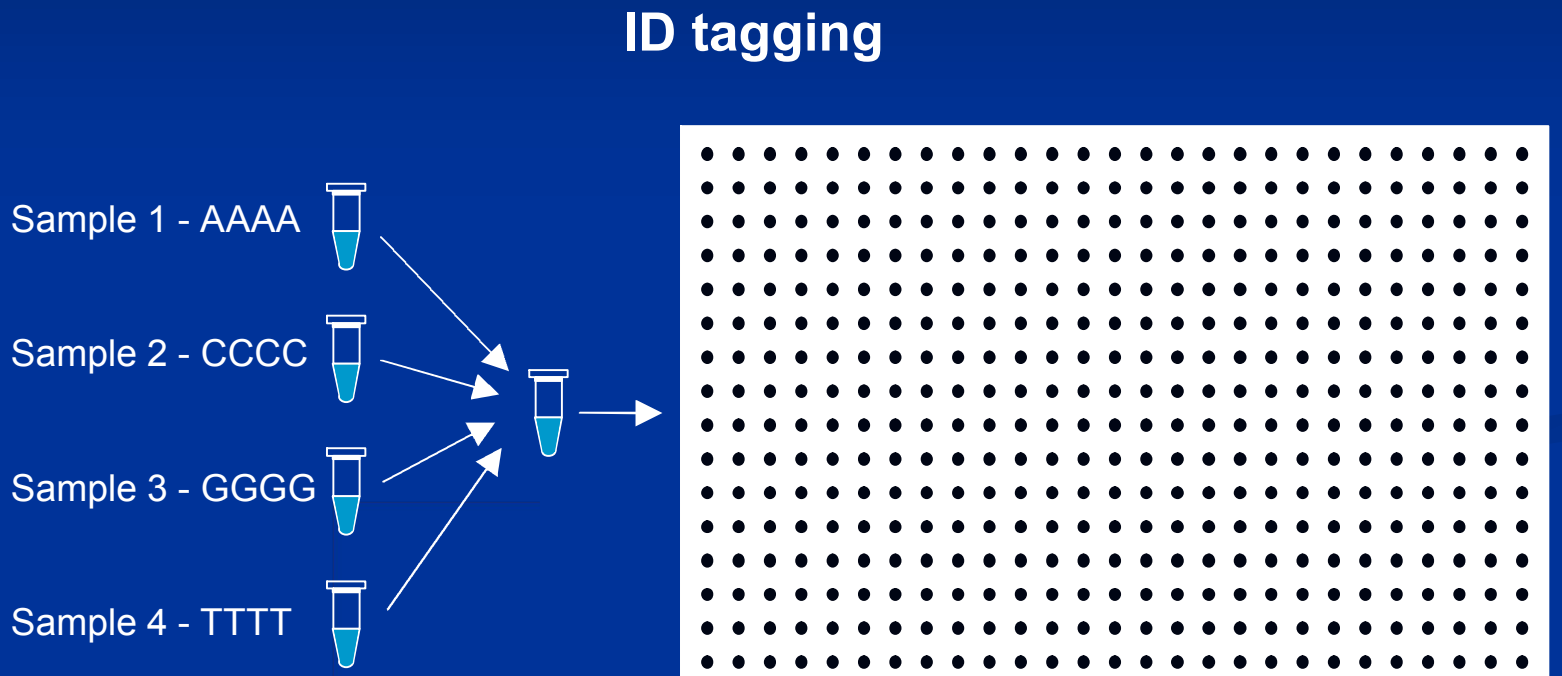


# Analysing Multiple Samples

## Physical separation



# Analysing Multiple Samples



# Cost benefits

Screening 120 amplicons 250-300bp (30Kb)
Required coverage 50x per allele
*(Approximately equivalent to BRCA screen)*

**Roche GS-FLX - $1x10^8$ bp/run**

- 33 samples /run
- ~£60 / sample (run cost)

**Illumina Genome Analyser**

- 600 samples / run
- ~£15 / sample (run cost)

# Utilisation of capacity

- NHS centralised technical facility
- Farm technical work out to commercial lab
- Small inter-laboratory collaborations
- Local research and inter-disciplinary collaborations
- Go for lower capacity technology
- Increase work portfolio

# Summary

- Analysing multiple patient samples in parallel
  - Physical separation
  - ID tagging
  - Pooling strategies
- Targeting specific regions of interest
  - PCR
  - Arrays
  - Probe based circularisation
- Rationalising amplification process
  - High level multiplexing
  - Capturing specific fragments
- Data handling Capacity and skills

# Conclusions

- New generation sequencing technologies promise capacities and TP several orders of magnitude greater than current capabilities

But..

- These are new technologies and there are significant issues to be resolved
- Even current capacities will be difficult to use for current diagnostic applications
- Technology is evolving rapidly
  - Consideration of longer term requirements
  - Development of platform independent sample prep

# The future

- Single molecule sequencing
  - Eliminates phasing problems
  - More quantitative
  - Not without issues
- Very long reads (10,000s bases)
  - Reduces analysis problems (assembly)
  - Resolution of repeats
  - Simplified structural analysis
- Real time detection at incorporation
  - Very fast