Potential for Diagnostic Application of New Sequencing Technologies

Chris Mattocks
National Genetics Reference Laboratory (Wessex)

Overview

- Technologies and platforms
- Applications
- Issues for diagnostic utility
- The future

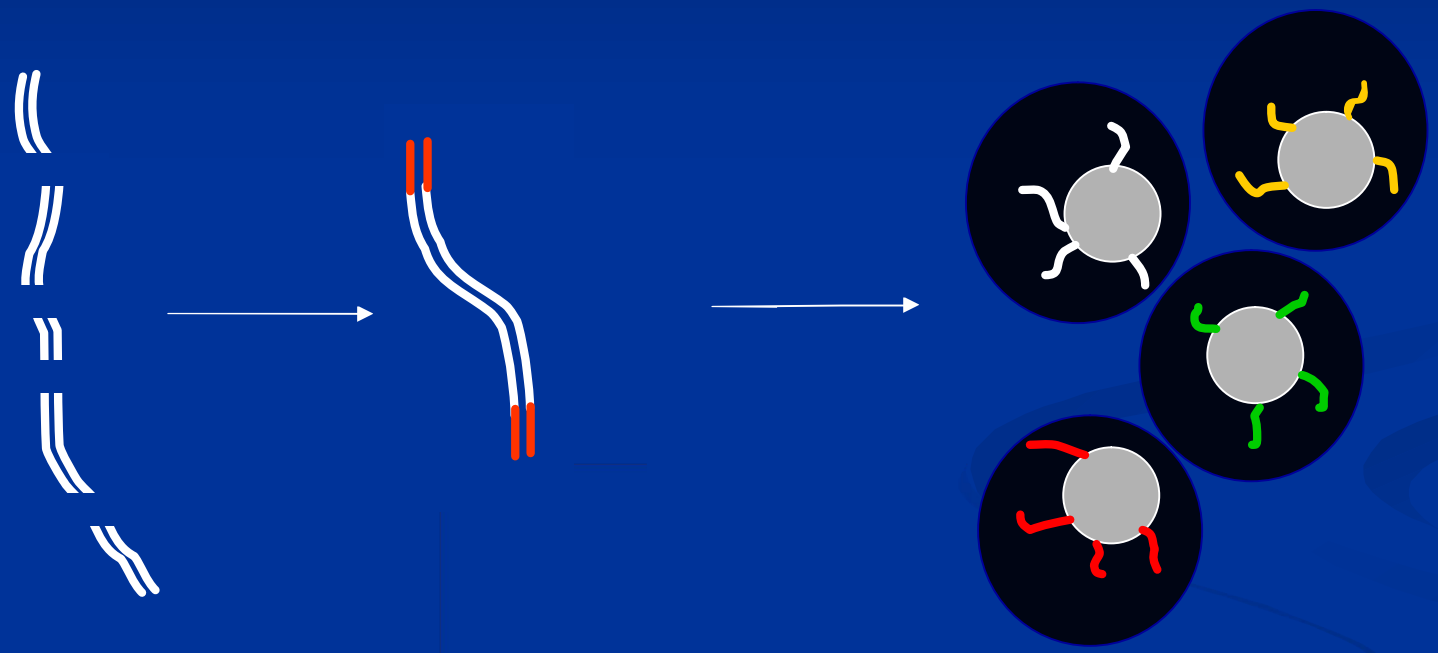
Why new sequencing technologies?

- 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 ….				

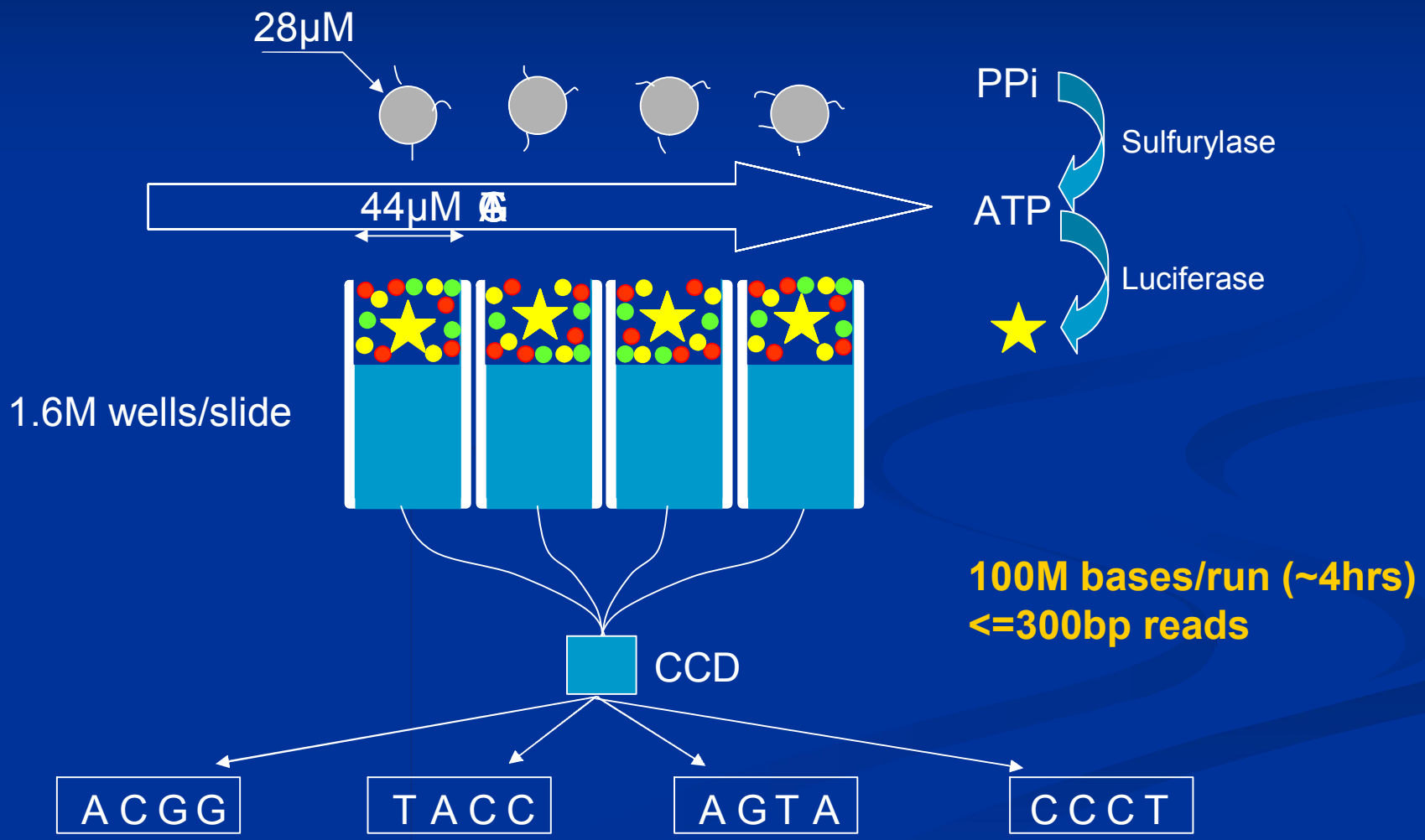
Emulsion PCR (emPCR)



Sequencing by extension

454 life sciences Corp. - Margulies et al Nature. 2005 Sep 15;437(7057):376-80

Roche GS20 / FLX

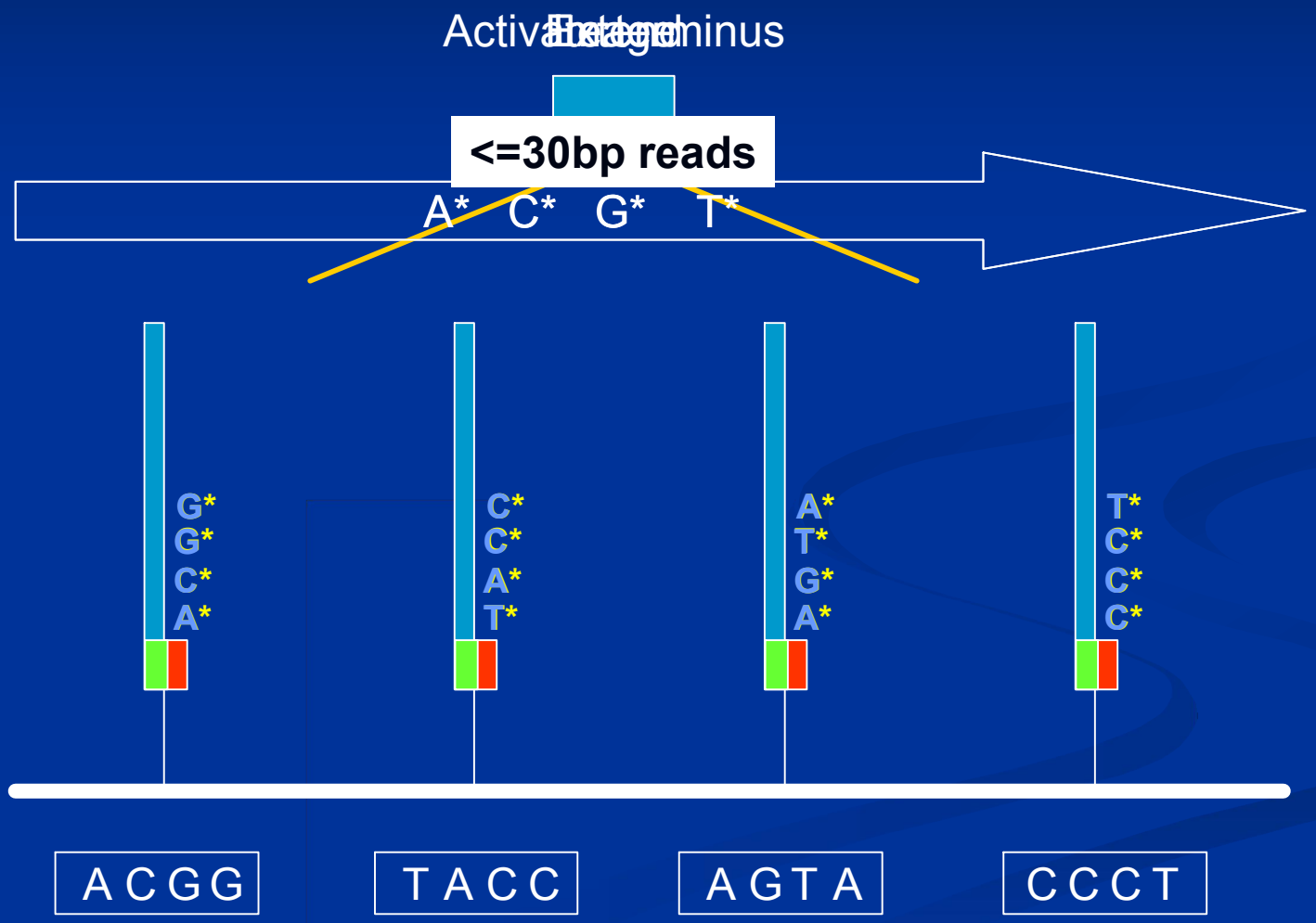


100M bases/run (~4hrs)
<=300bp reads

Sequencing by Synthesis (SBS)

Solexa (Illumina) - http://www.solexa.com

Helicos - http://www.helicosbio.com/ (VIRTUAL teminators)

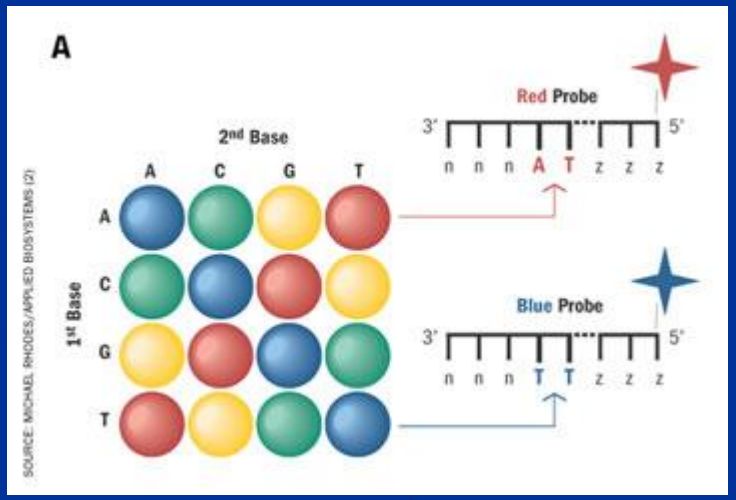


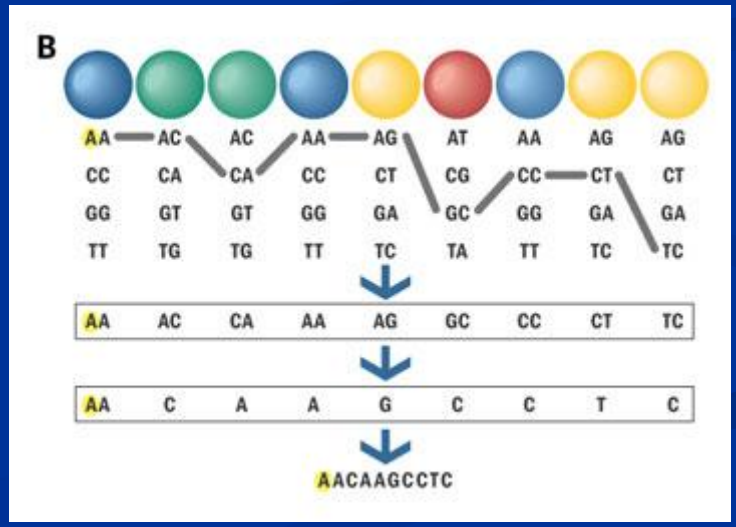
SOLiD

(Supported Oligonucleotide Ligation and Detection)

Applied Biosystems - solid.appliedbiosystems.com/

3B bases/run
<=30bp reads





Pros and cons

- 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 stored
- Analysis
  - SOLiD: 2 base encoding very accurate (may reduce required depth)
- Amount of starting material
- Workflow
  - Breakdown between prep time / machine time

Applications

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

Mutation scanning

Quantitative analysis / CNVs

Methylation analysis

RNA analysis

Expression analysis

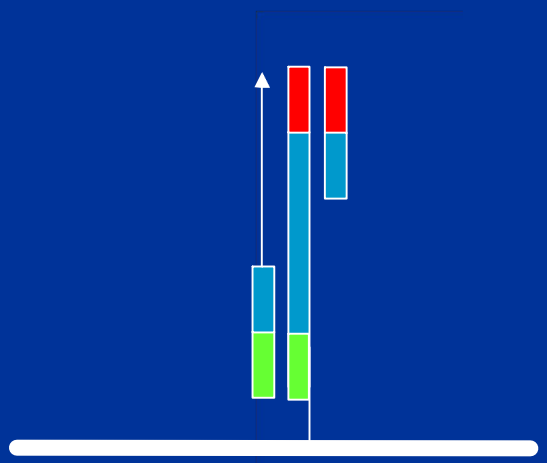
Tumour profiling / deep sequencing

Genome architecture / structural analysis

Structural analysis

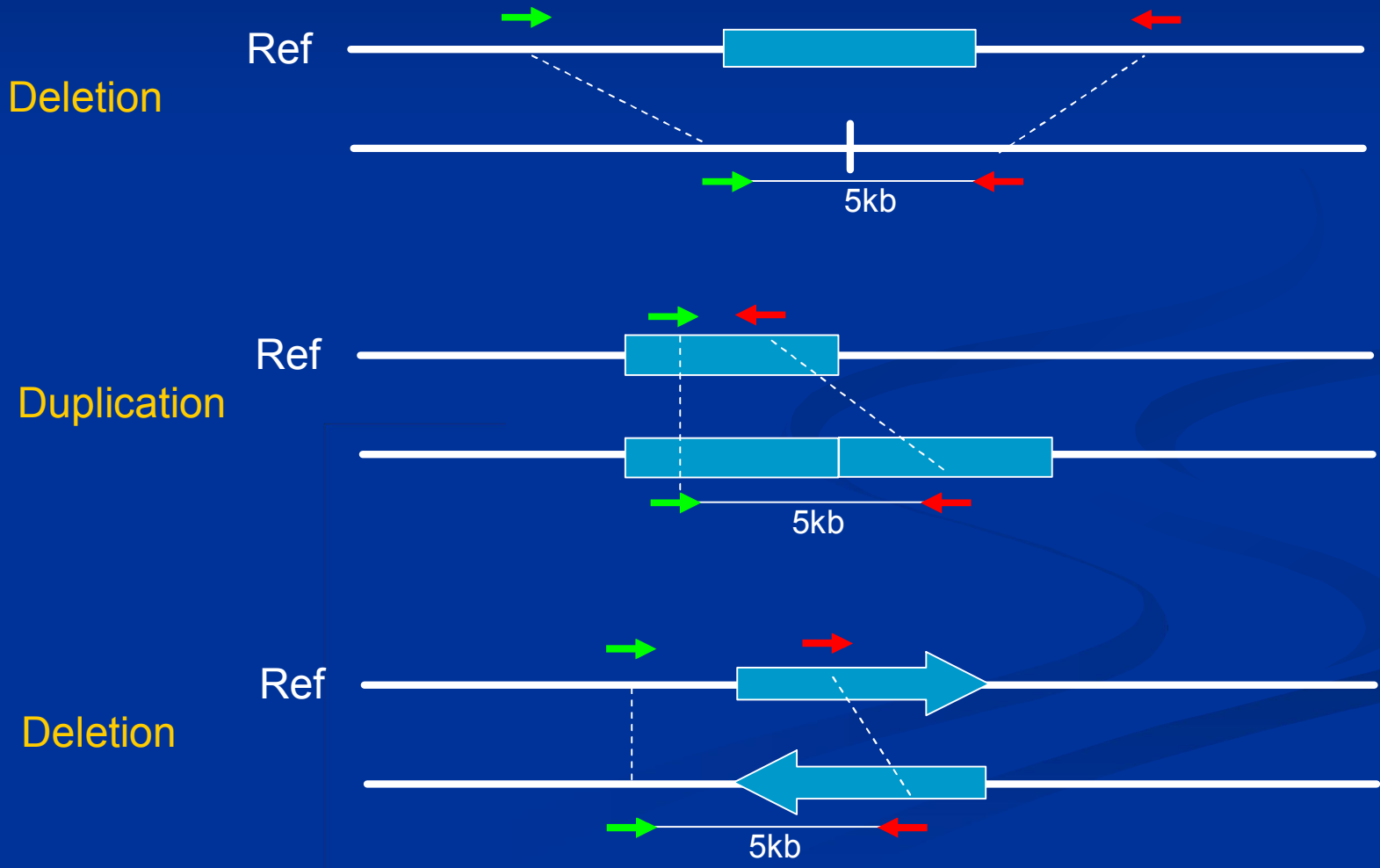
Paired end sequencing

- Fragment DNA
- Select a particular size fraction (e.g. 5000bp)
- Sequence the two ends
  - Sequence the same ‘spot’ twice from two opposing ends
  - Construct a paired end tag by circularisation



Structural analysis

- Map end reads back to reference genome



BRCA mutation scanning

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

- amplicons 250-300bp long = 400,000 sequence reads
- Coverage 50x per allele = 4,000 amplicon coverage
- ~120 amplicons per patient = 33 patients /run
- TAT ~ 1 week [sample prep?..analysis?]
- Run cost ~£2000 ≈ £60 / patient

Illumina Genome Analyser

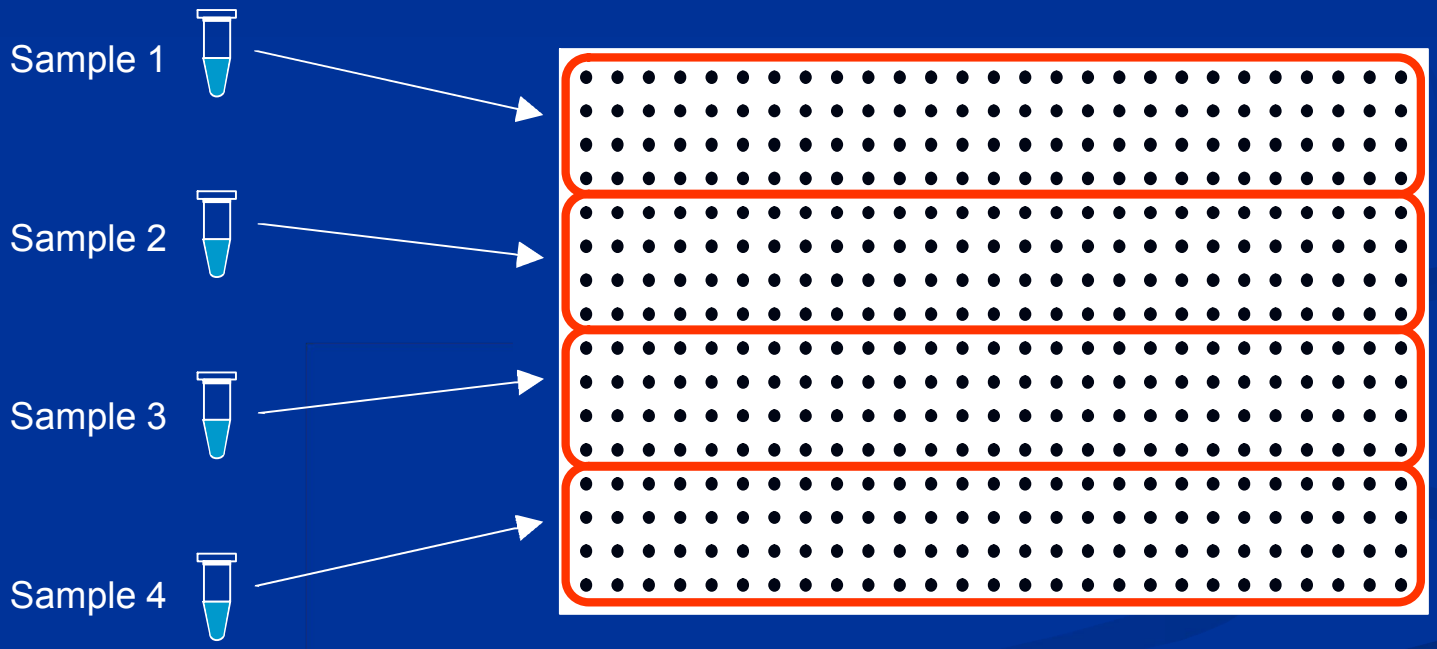
- 600 samples / run @ ~£15 / sample run cost

Diagnostic requirements

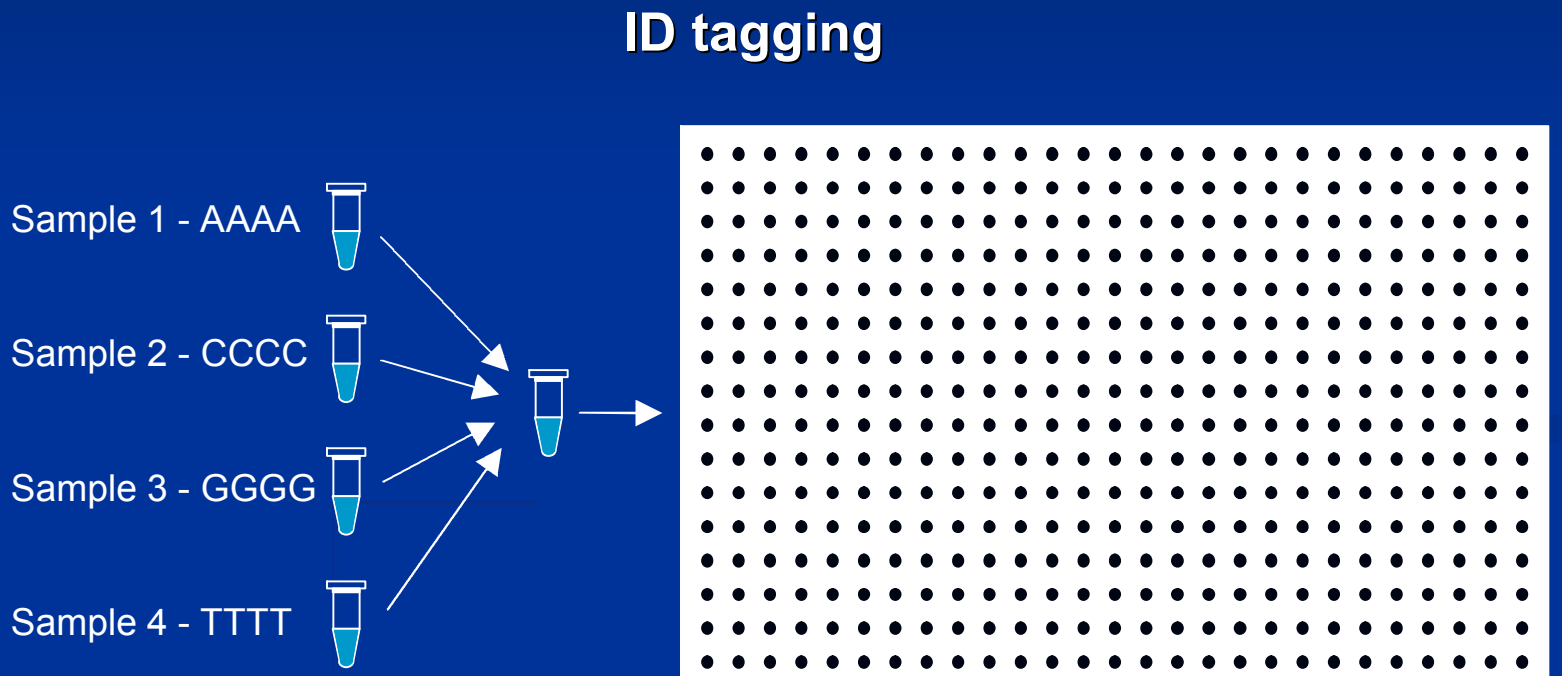
- Analysing multiple patient samples in parallel
  - Physical separation
  - ID tagging
  - Pooling strategies
- Targeting specific regions of interest
  - PCR
  - Arrays
  - Probe based circularisation
- Rationalising amplification process
  - High level multiplexing
  - Capturing specific fragments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

Analysing Multiple Samples

Physical separation



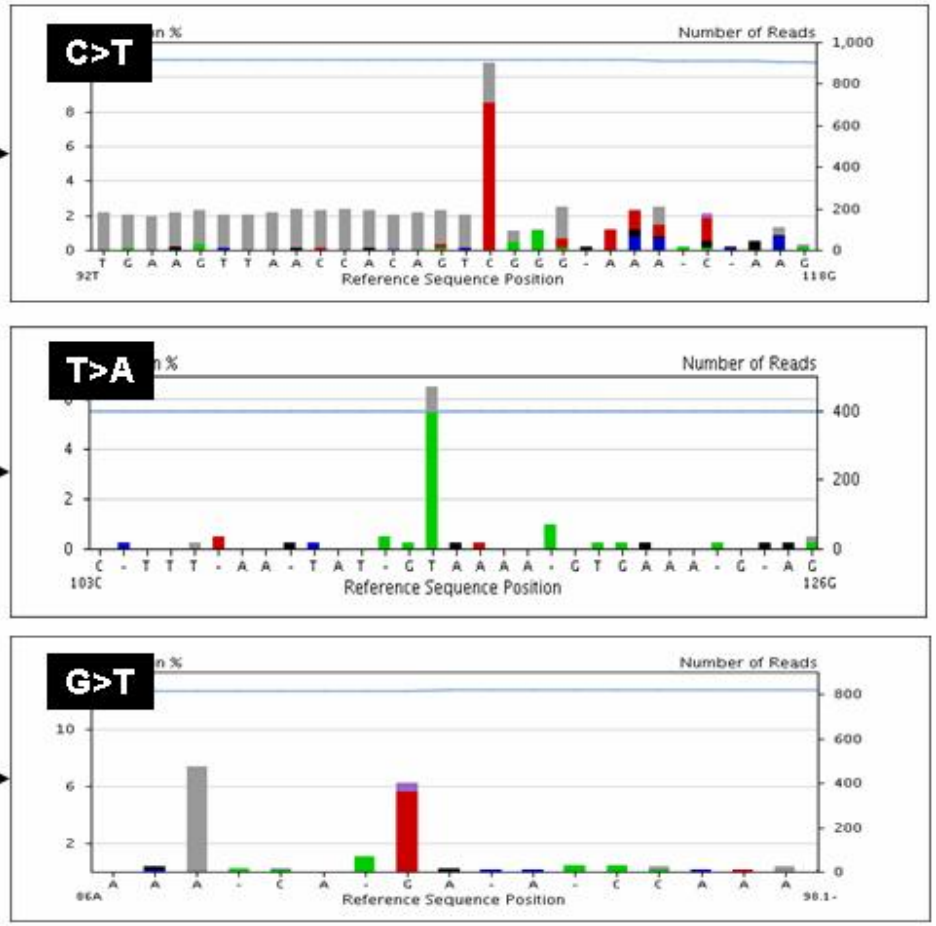
Analysing Multiple Samples



Pilot Study (CRUK)

Strategic pooling

sample	pool 1	2	3	4	5
01					X
02				X	
03			X		
04		X			
05	X				
06				X	X
07			X		X
08			X	X	
09		X			X
10		X		X	
11		X	X		
12	X				X
13	X			X	
14	X		X		
15	X	X			
16			X	X	X
17		X		X	X
18		X	X		X
19		X	X	X	
20	X			X	X
21	X		X		X
22	X	X		X	
23	X	X	X		
samples	9	10	10	10	10
Het	0.06	0.05	0.05	0.05	0.05

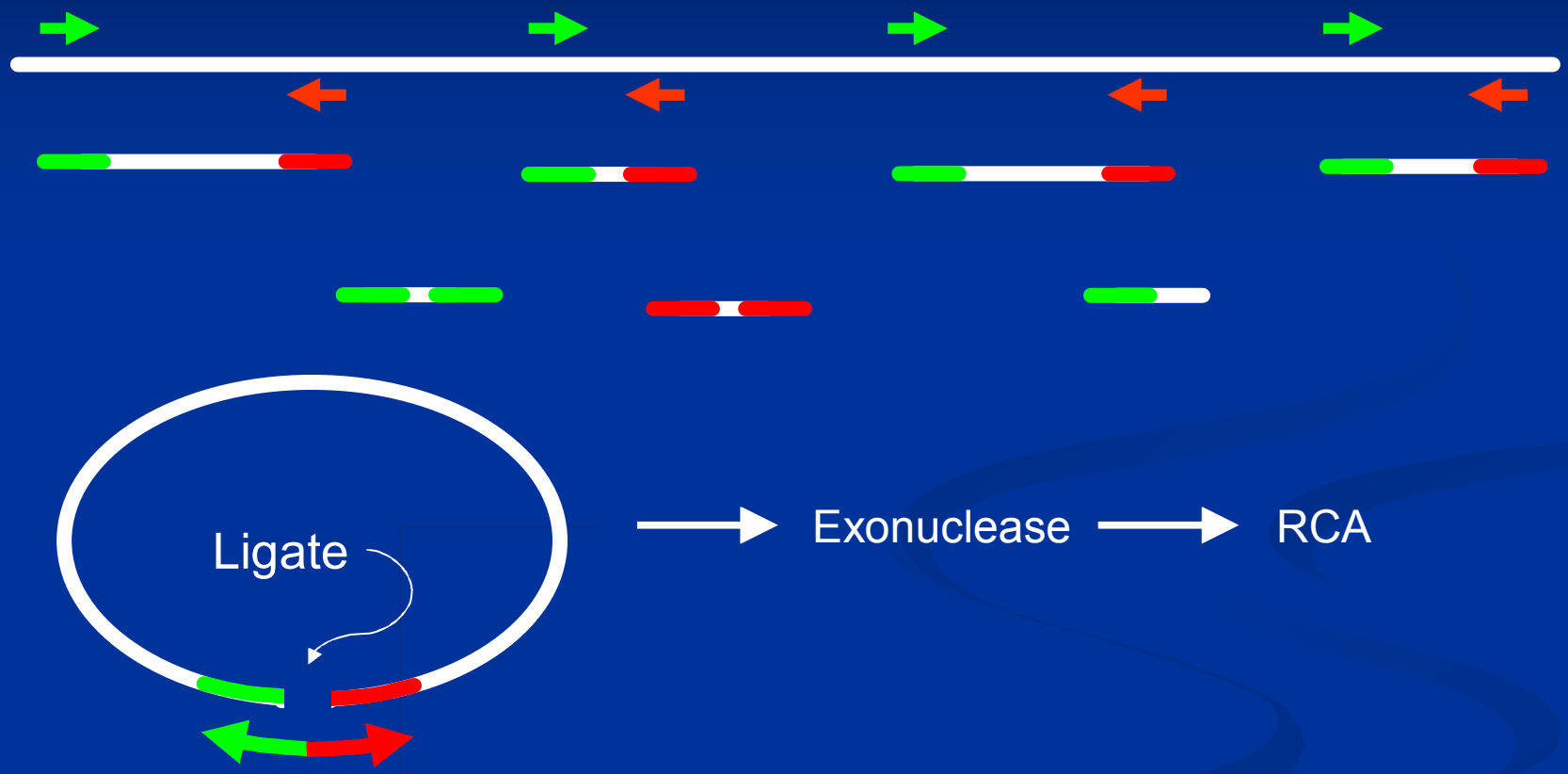


Assay targeting

PCR based sample preparation:

- 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

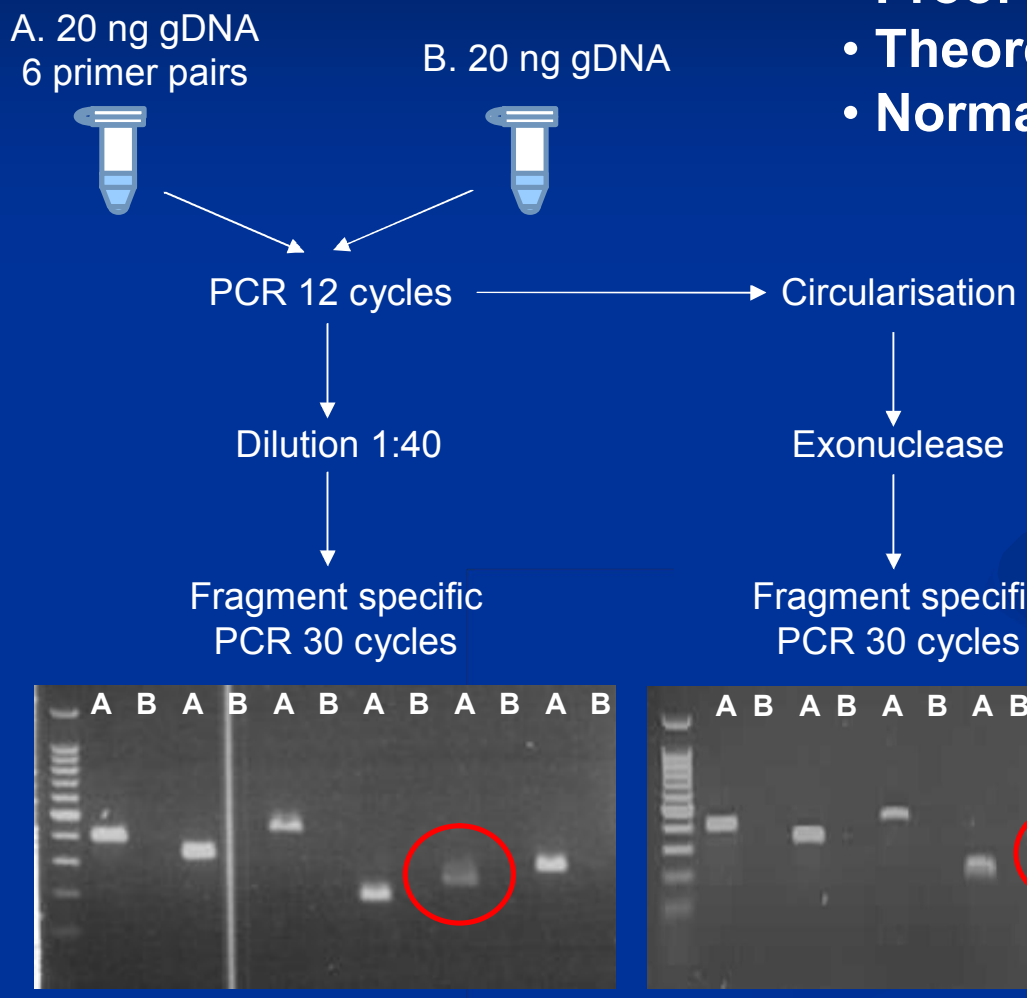
High Level Multiplex PCR



Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector
Simon Fredriksson,* Johan Banér, Fredrik Dahl, Angela Chu, Hanlee Ji, Katrina Welch, and Ronald W. Davis
Nucleic Acids Res. 2007 April; 35(7): e47

High Level Multiplex PCR

- **Proof of principle on 6 plex**
- **Theoretical plex level of 1000s**
- **Normalisation of product based on quantity of probe ?**

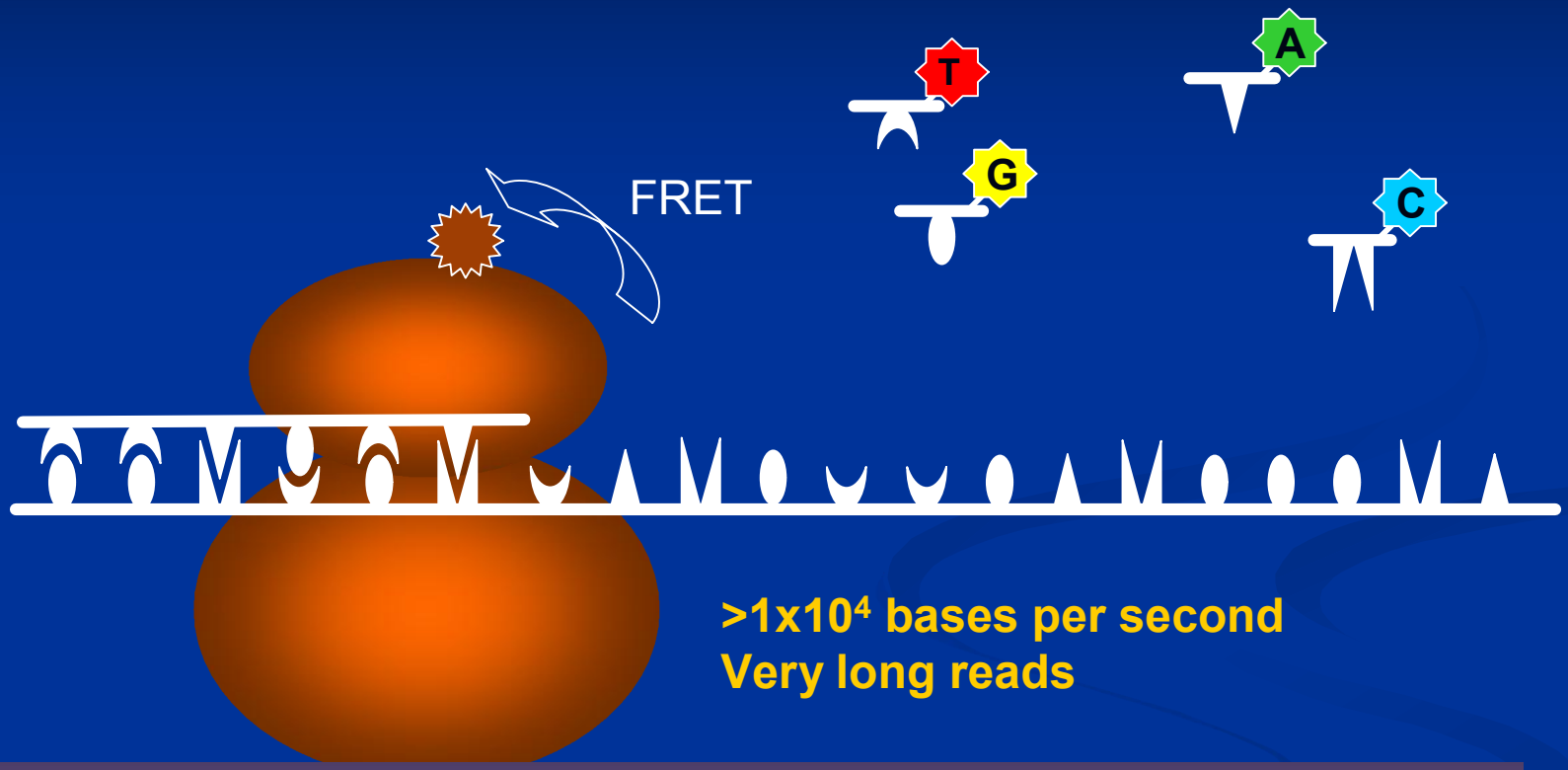


The future

- 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

Real time detection at incorporation

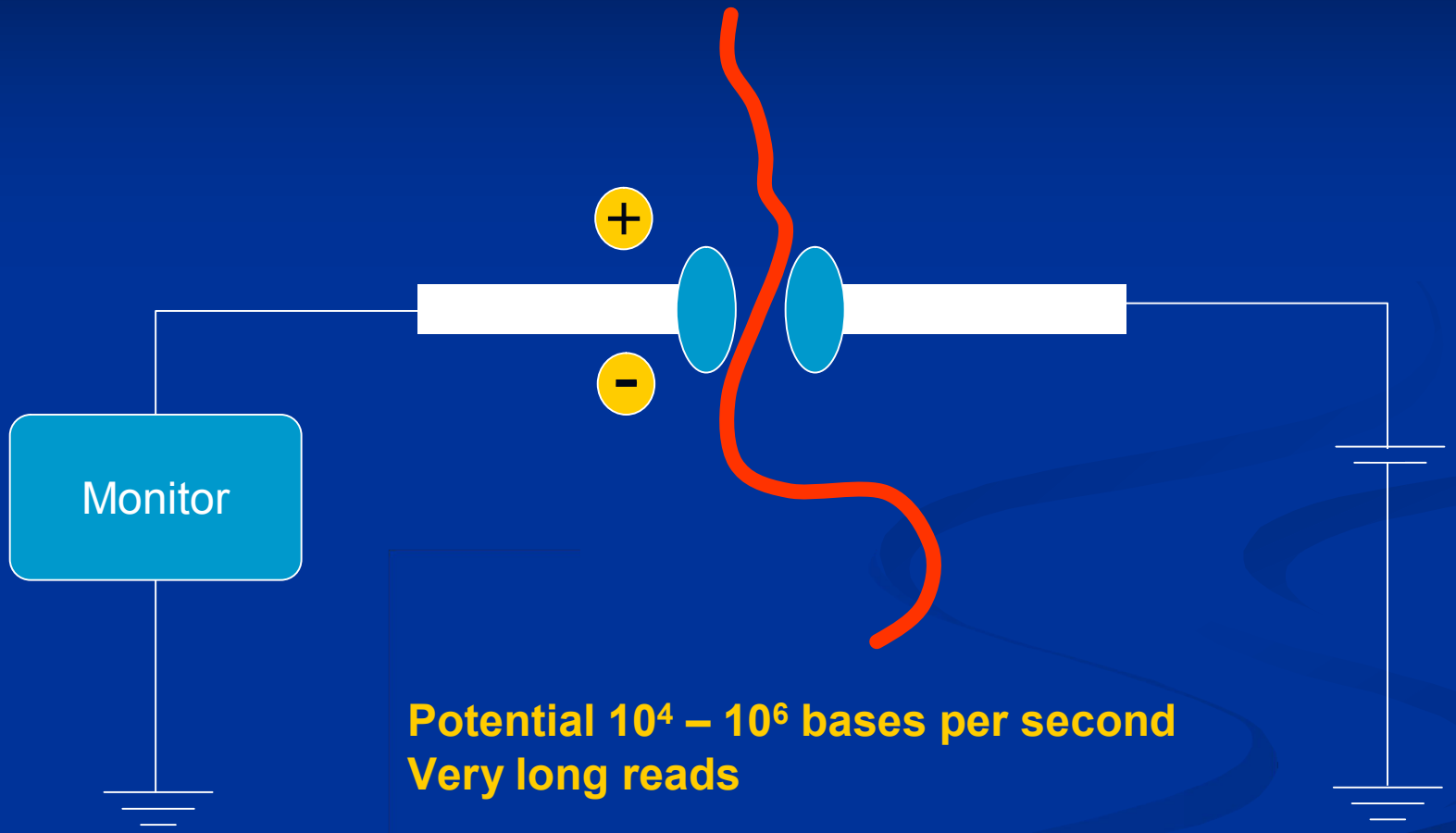
Visigen - http://visigenbio.com/
Pacific BioSciences – http:/pacificbiosciences.com



>1x10⁴ bases per second
Very long reads

Nanopore Sequencing

Harvard Nanopore group - http://www.mcb.harvard.edu/branton/index.htm



Potential 10^4 – 10^6 bases per second
Very long reads

Summary and 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

Summary and conclusions

- Diagnostics will require:
  - Effective targeting
  - Sample identification / separation
  - Effective methodologies to rationalise sample prep
  - Data handling capacity and skills
- 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

References

1. Fredriksson S, Banér J, Dahl F, Chu A, Ji H, Welch K, Davis RW.
   Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector.
   Nucleic Acids Res. 2007;35(7):e47. Epub 2007 Feb 22.

2. Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, LeProust EM, Peck BJ, Emig CJ, Dahl F, Gao Y, Church GM, Shendure J.
   Multiplex amplification of large sets of human exons.
   Nat Methods. 2007 Nov;4(11):931-6. Epub 2007 Oct 14.

3. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA.
   Direct selection of human genomic loci by microarray hybridization.
   Nat Methods. 2007 Nov;4(11):903-5. Epub 2007 Oct 14.

4. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME.
   Microarray-based genomic selection for high-throughput resequencing.
   Nat Methods. 2007 Nov;4(11):907-9. Epub 2007 Oct 14.

Web sites

- https://www.roche-applied-science.com/sis/sequencing/flx/index.jsp
- http://www.illumina.com/pages.ilmn?ID=204
- http://marketing.appliedbiosystems.com/mk/get/SOLID_KNOWLEDGE_LANDING
- http://www.helicosbio.com
- http://www.pacificbiosciences.com/index.php
- http://visigenbio.com/technology.html
- http://www.zsgenetics.com/thetech/reading/reading.html
- http://www.agilent.com/about/newsroom/presrel/2007/05nov-al07002.html
- http://www.bionanomatrix.com/index.html
- http://completegenomics.com/
- http://www.intelligentbiosystems.com/index%20mod%201.html
- http://www.nabsys.com/
- http://www.reveo.com/us/node/309

New Sequencing Technologies: Discussion

Chris Mattocks
National Genetics Reference Laboratory (Wessex)

How can capacity be effectively used?

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

What should we be investigating ?

- Specifically targeted for each patient
- One test all samples
  - is there a point where this is acceptable?

Economics

Practicality

Ethics

How and where should data be stored?

- Laboratory
- Clinician / Doctor
- Central NHS
- Personal
- Unused data should not be kept

Would you have your genome sequenced?