The use of new sequencing technologies for genome analysis

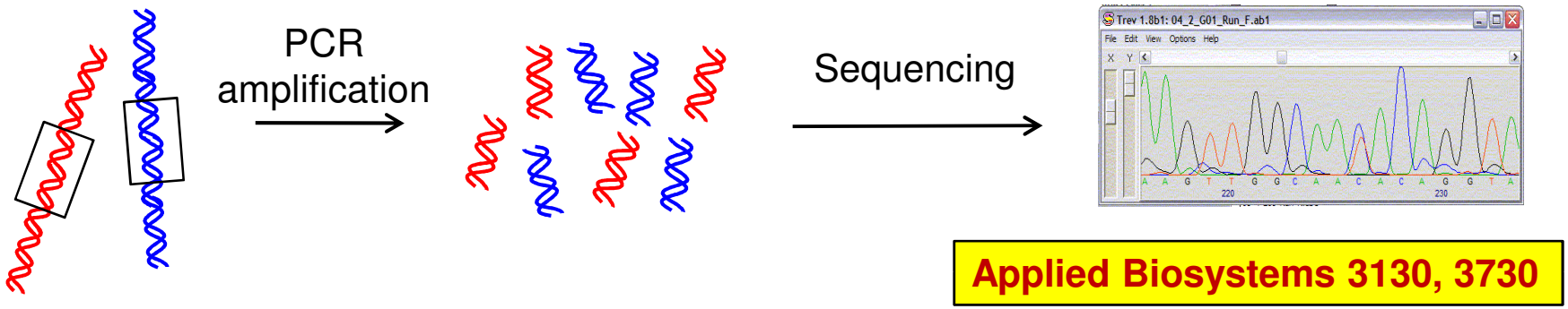
Chris Mattocks

National Genetics Reference Laboratory (Wessex)

Outline

- General principles of clonal sequencing
- Analysis principles
- Applications
- CNV analysis
- Genome architecture

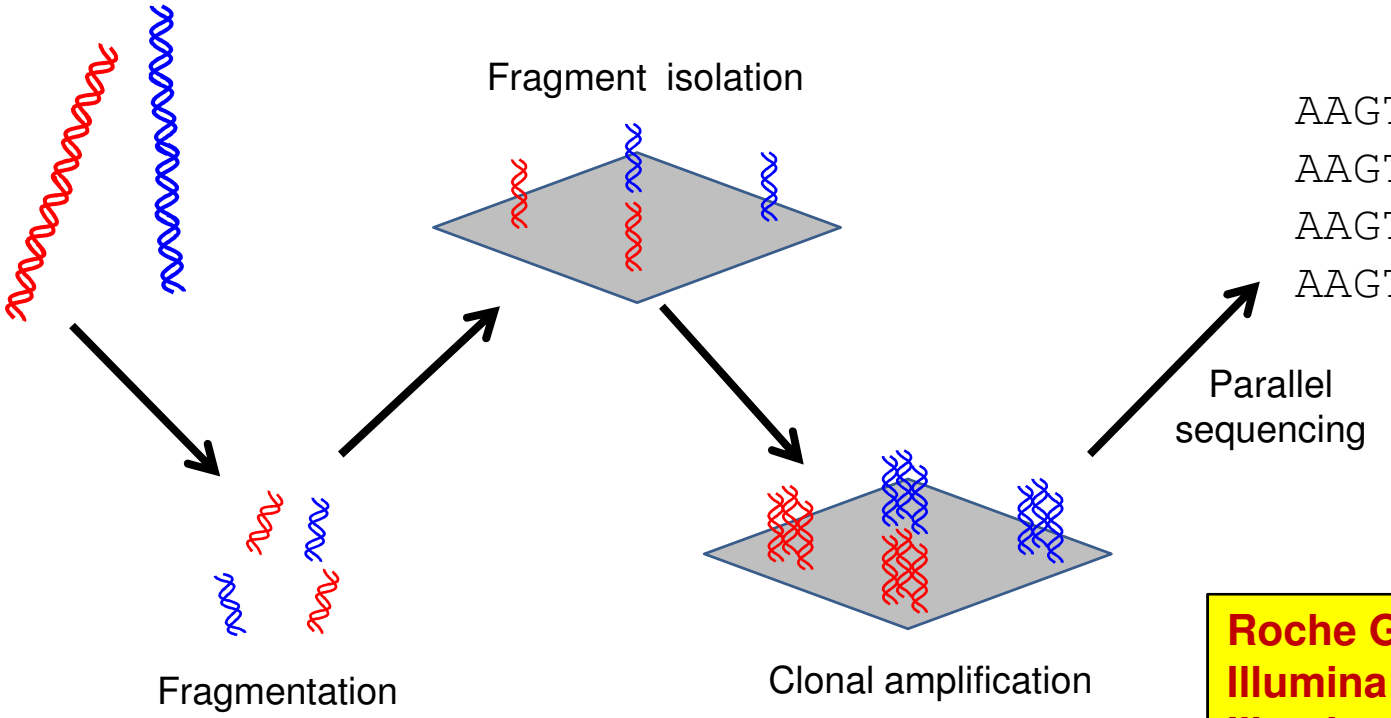
Sanger sequencing



Applied Biosystems 3130, 3730

- Inherently targeted approach
- Each sequencing reaction represents a single PCR product up to ~1Kb
- Analysis trace represents an averaged result from 000's molecules
- Capacity up to ~1Mb per day

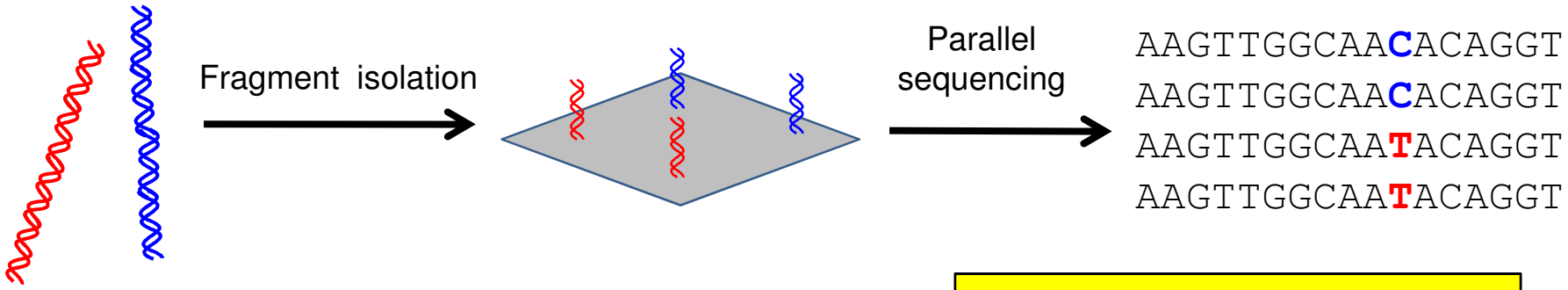
Second generation sequencing



Roche GS-FLX
Illumina Genome Analyser
Illumina HiSeq 2000
Applied Biosystems SOLiD4

- 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
- Capacity up to ~25Gb per day (~25,000x Sanger approach)

Third generation sequencing

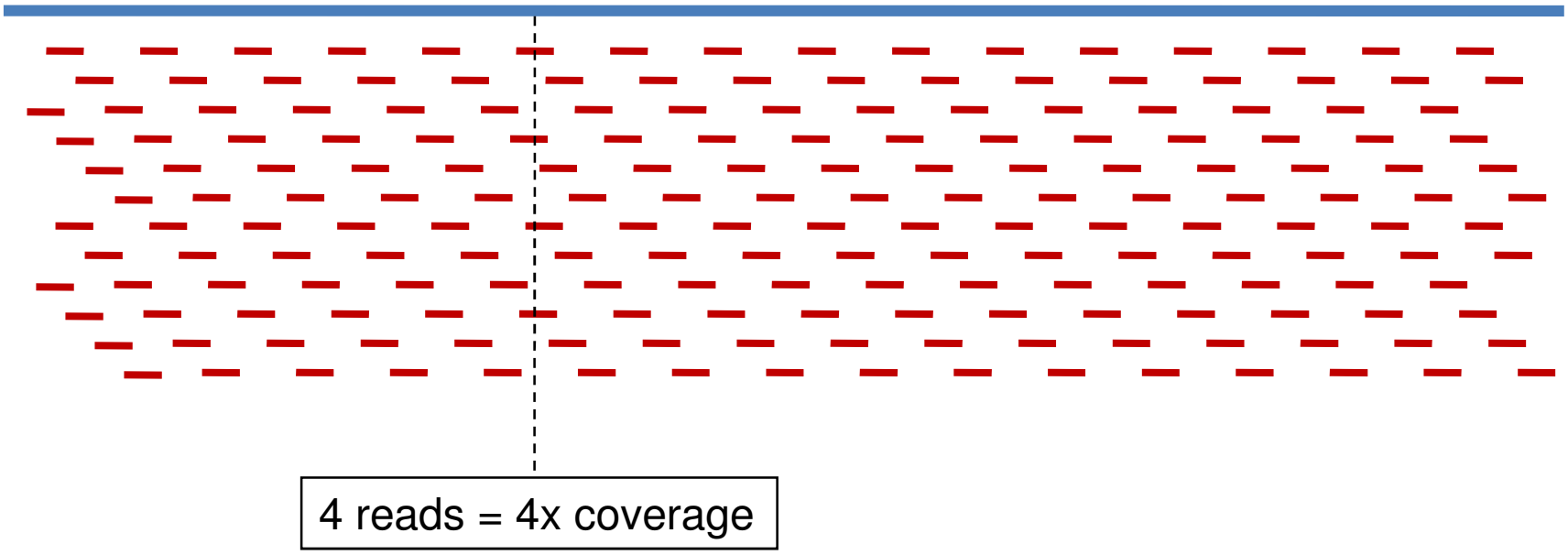


Helicos
Pacific Biosciences
Oxford Nanopore
Applied Biosystems (Visigen)
Ion Torrent

- The entire input sample is sequenced
- Potential for sequencing long DNA fragments (kb range)
- Each sequence read represents a single starting DNA molecule
- Potential capacity genome in 15mins

Analysis principles

1. Base calling
2. Mapping
    - One “spot” ≡ one sequence “read” ≡ one molecule



Analysis principles

3. Application specific analysis
   - One “spot” ≡ one sequence “read” ≡ one molecule

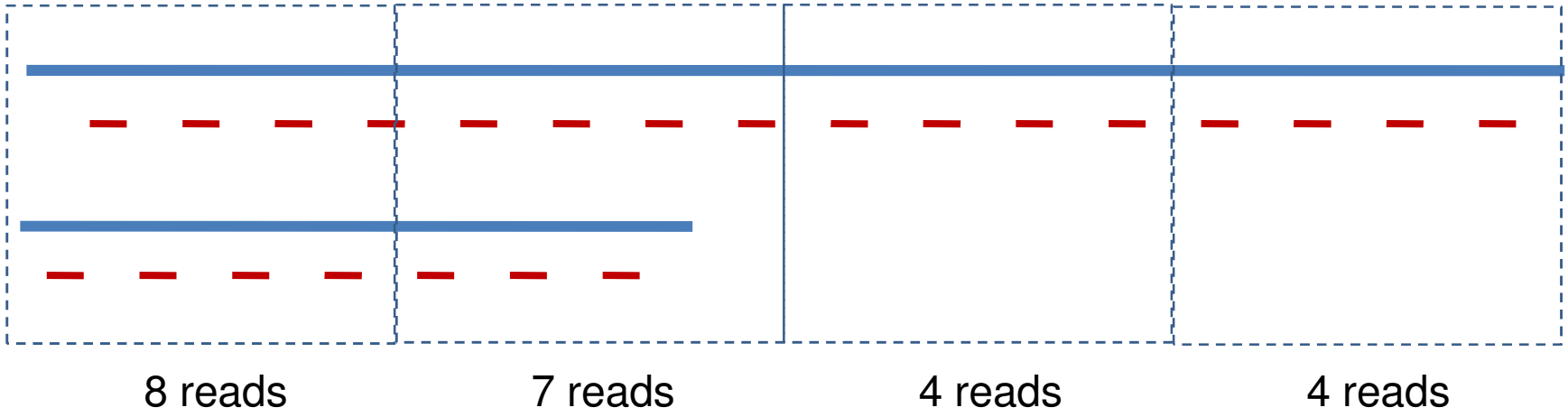
Types of analysis

Basis of analysis	Application
Sequence identity	standard molecular applications like mutation scanning
Read counting	copy number analysis
Relative location	paired end sequencing for analysis of genome architecture

Potential applications

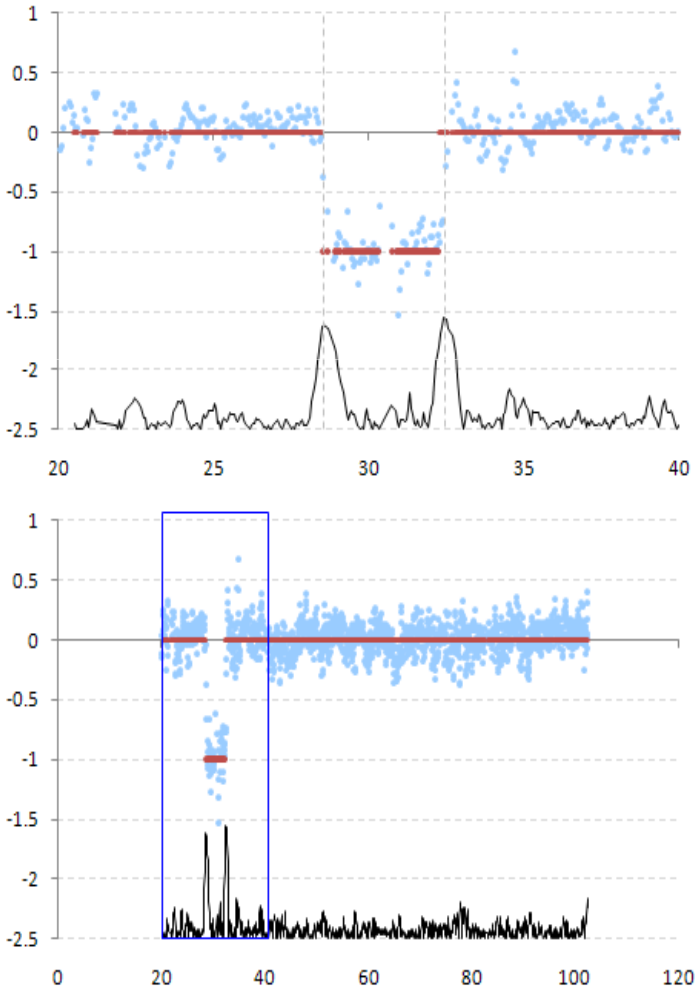
- Mutation scanning
- Methylation/epigenetic analysis
- Mitochondrial DNA analysis
- Deep sequencing / tumour profiling
- Expression analysis
- cDNA (RNA) sequencing
- Non-invasive prenatal diagnosis
- Copy number analysis by counting single reads (array comparative hybridisation equivalent)
- Genome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

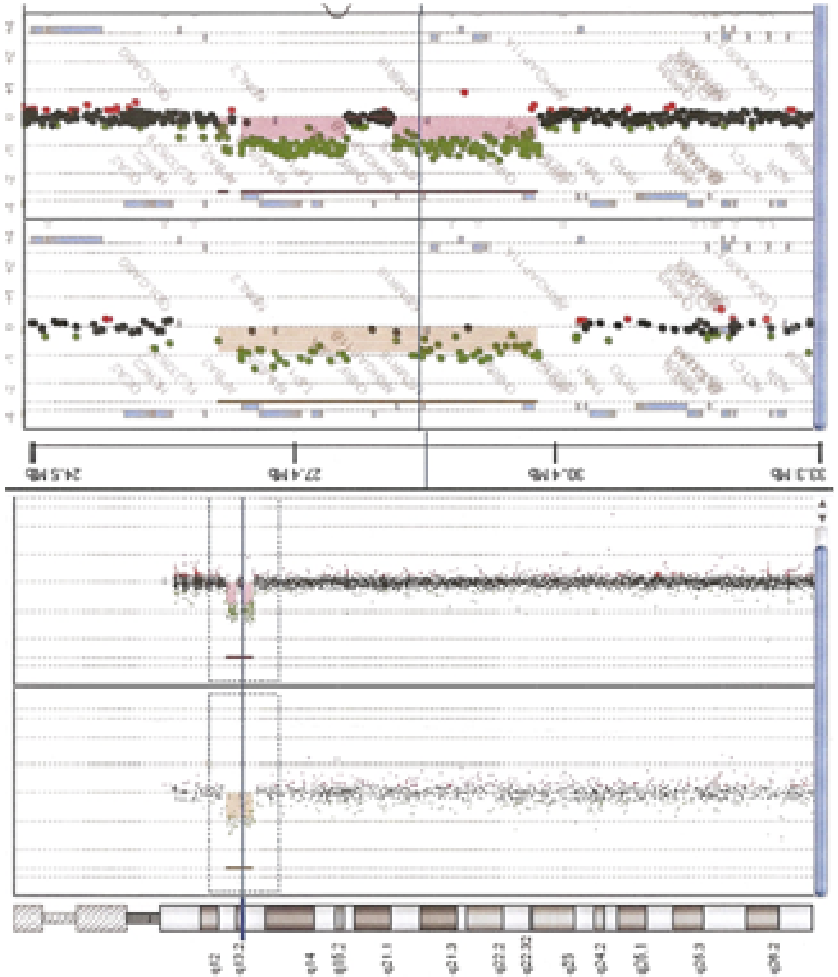
Copy number variation by read counting



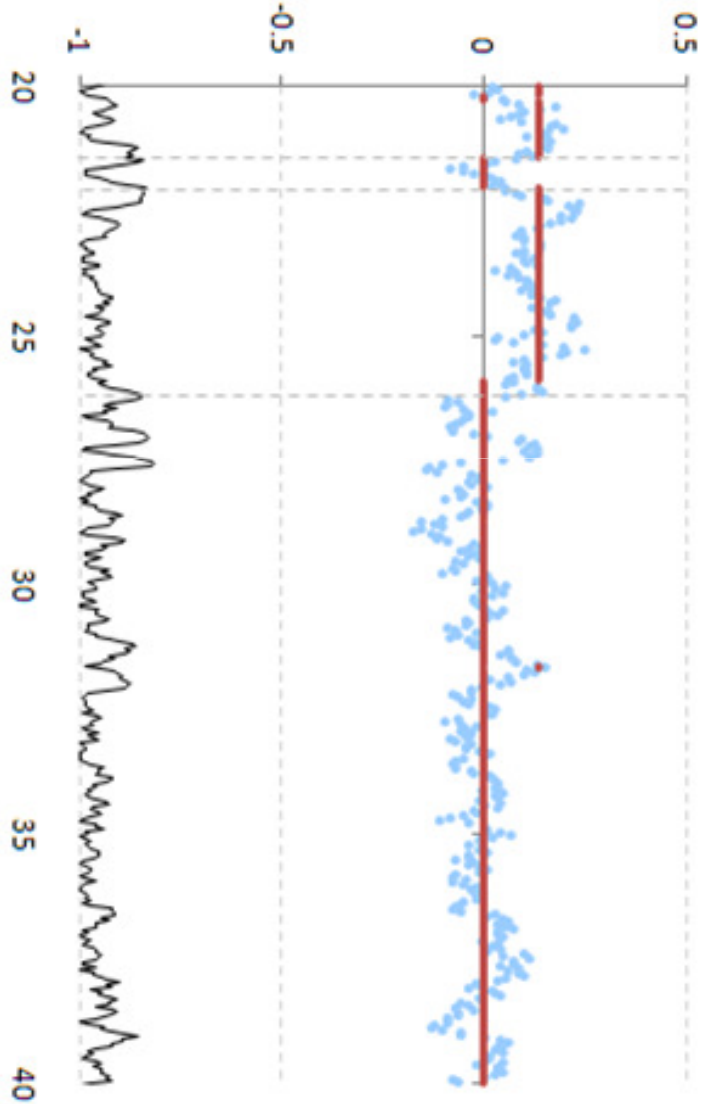
- Not necessary to have complete sequence coverage
- Can choose resolution by altering bin size
- ↑reads = ↑resolution ***but*** ↑ capcity usage and ↑ cost
- For given read depth: ↑bin size = ↑sensitivity ***but*** ↓size resolution

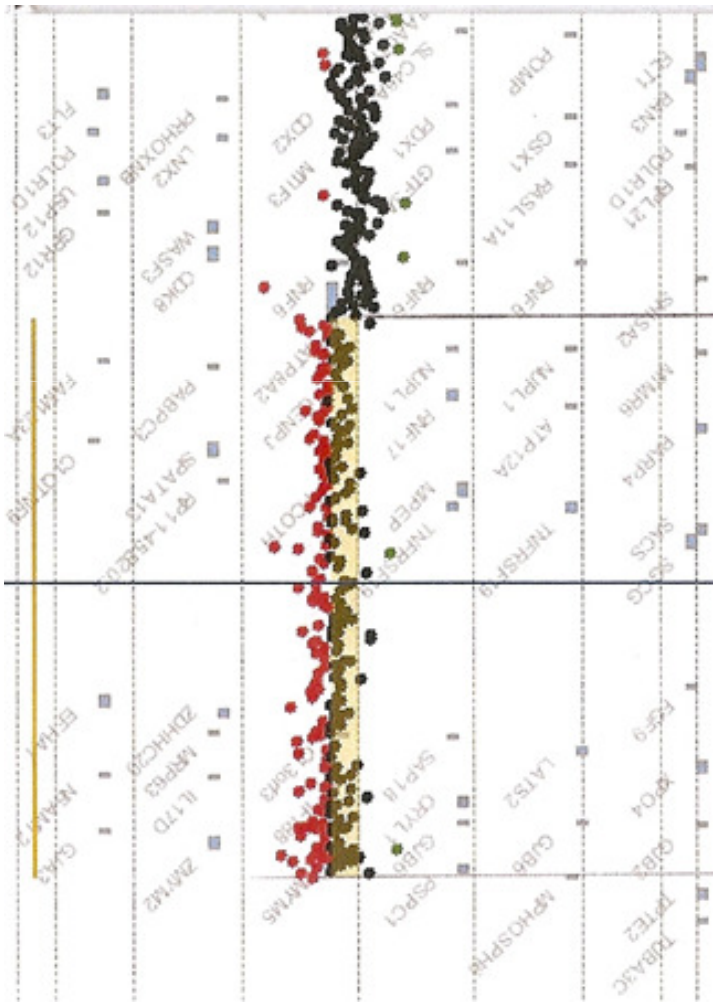
SQ10 - Chromosome 15 del 28.55-32.40Mb





Chromosome 13 dup 20.00-26.20Mb (Mosaic ~20%)



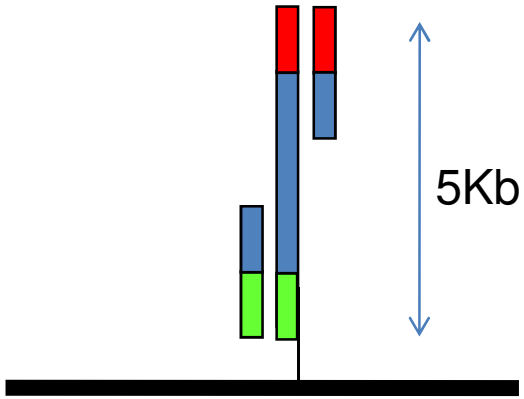


aCGH comparison

- 20 samples characterised on variety of aCGH platforms (Agilent 4x44K, 4x180, Affymetrix 27M)
- Pilot of 3 sample run
  - 8 variations detected (not blinded)
  - Size resolution ~400Kb
  - End resolution ~40Kb
  - Consumable cost ~£400

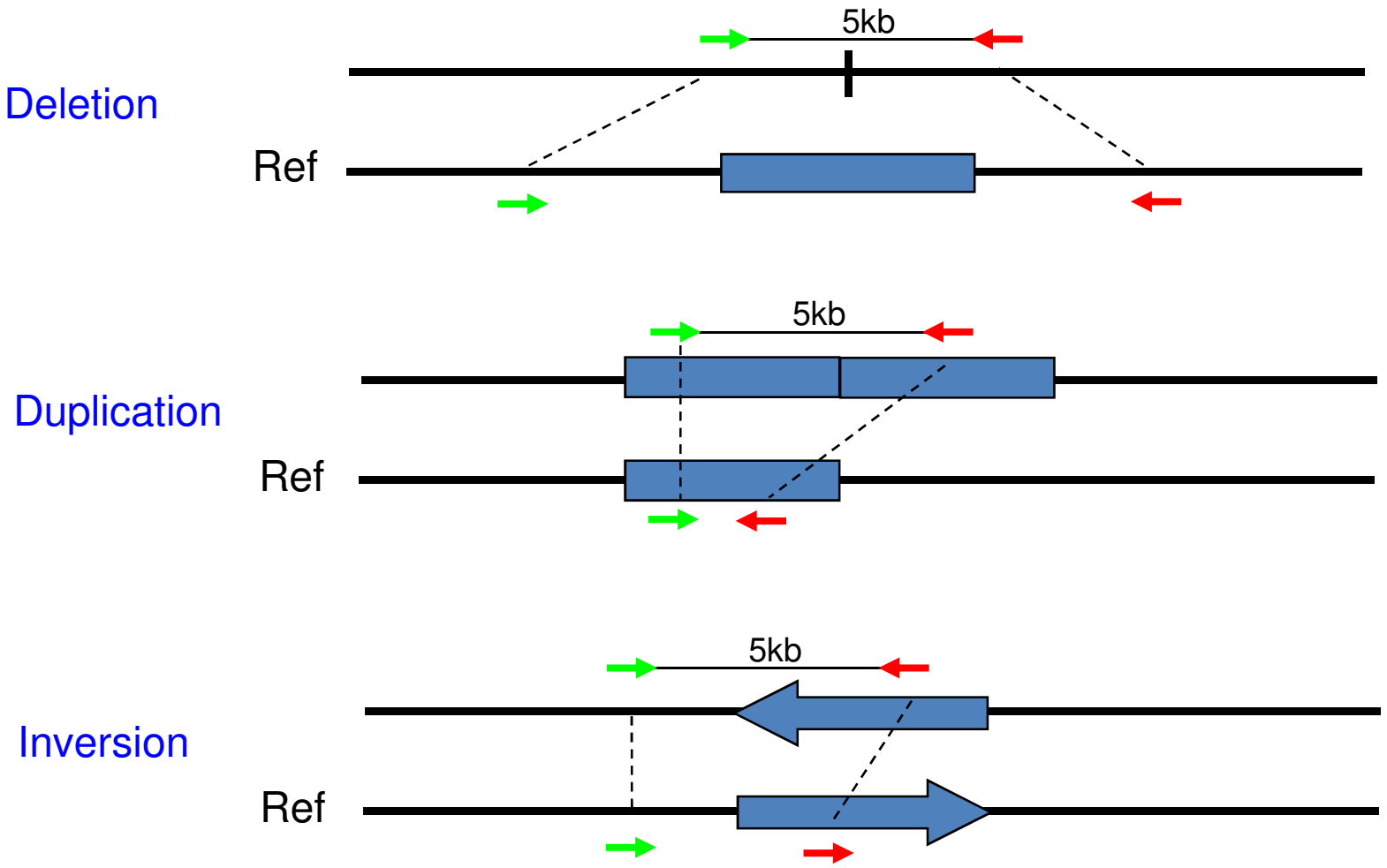
Paired end sequencing

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



Paired end sequencing

- Map end reads back to reference genome



Paired end sequencing

- Can be used to analyse genome architecture including balanced variation
- Can be used in conjuction with simple read counting to confirm results of CNV analysis and improve resolution
- Sequencing cost essentially x2

Summary and conclusions

- Clonal sequencing can be used to emulate aCGH by simple read counting
- Further work is needed to refine analysis and validate methods
- With current capacities consumable costs are roughly comparable
- Paired end sequencing may be useful for confirmation of read count data and investigate balanced variations
- Given the wide potential for application of clonal sequencing the may be there may be advantages in terms of process rationalisation

Acknowledgements

- NGRL (Wessex)
  - Shuwen Huang
  - John Crolla
- Leeds Institute of Molecular Medicine
  - Graham Taylor
  - Jo Morgan