

**DNA technology: a quick guide to its  
history and evolution**  
Tom Burr, Source BioScience

## What does DNA technology mean?

PCR?

Genetically  
Modified  
Organisms?

Data  
storage?

Molecular  
biology?

DNA/RNA  
analysis?

DNA  
computing?

Cloning?

Gene  
therapy?

Synthetic  
biology?

## What DNA technology means to Source BioScience

PCR?

Genetically  
Modified  
Organisms?

Data  
storage?

Molecular  
biology?

**DNA/RNA  
analysis**

DNA  
computing?

Cloning?

Gene  
therapy?

Synthetic  
biology?

## Who cares about DNA analysis?

Physiology or Medicine 1959: Severo Ochoa and Arthur Kornberg  
(*DNA polymerase*)

Physiology or Medicine 1962: Francis Harry Compton Crick, James Dewey Watson  
and Maurice Hugh Frederick Wilkins (*DNA structure*)

Physiology or Medicine 1968: Robert W. Holley, Har Gobind Khorana and Marshall  
W. Nirenberg (*Genetic code*)

Chemistry 1980: Walter Gilbert and Frederick Sanger (*DNA sequencing*)

Chemistry 1993: Kary B. Mullis (*PCR*)



## Who uses DNA analysis?

- Forensics
- Paternity
- Diagnostics/screening
- Biomarkers
- Drug targets
- Authentication/typing
- Safety/testing
- Combinatorial biology & rational design
- Evolutionary biology
- .....

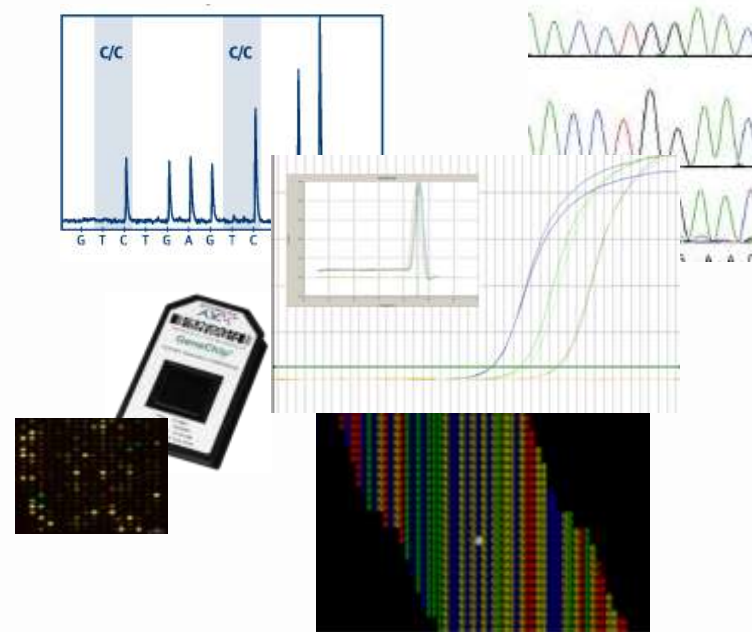
## A variety of approaches

Capillary-/Pyro-sequencing

Real-time PCR

Microarrays

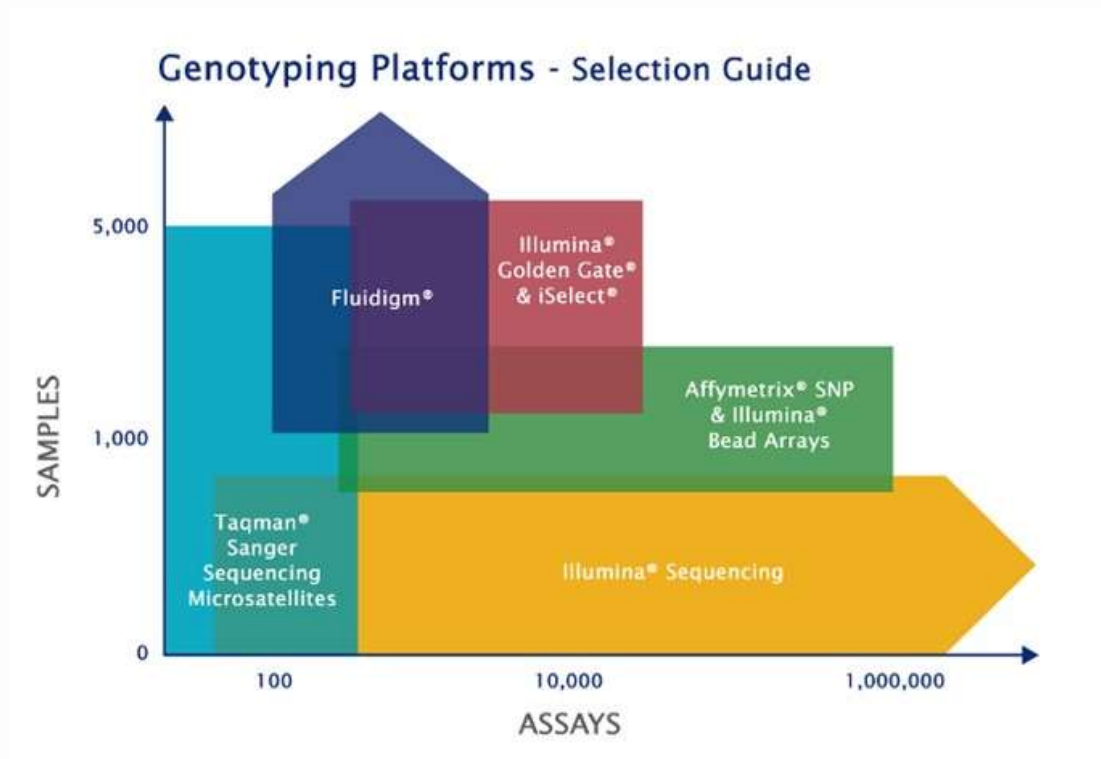
Next Generation Sequencing



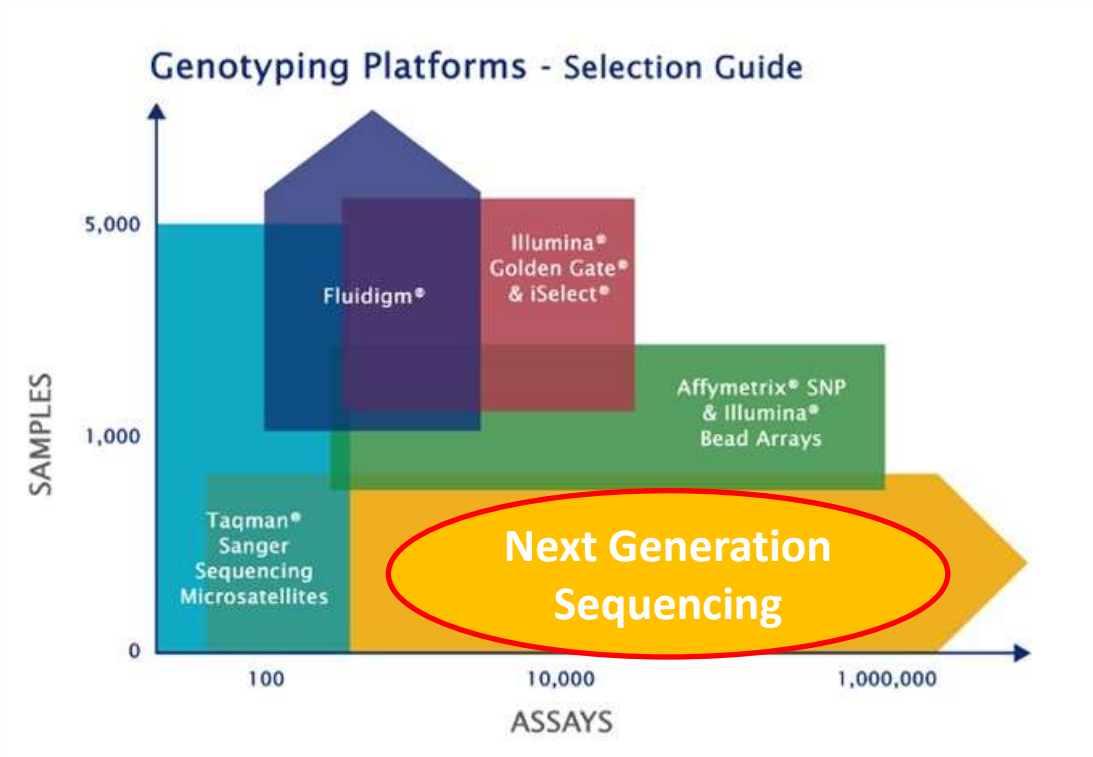
From snapshot to bigger picture, from known to unknown.



## A plethora of platforms



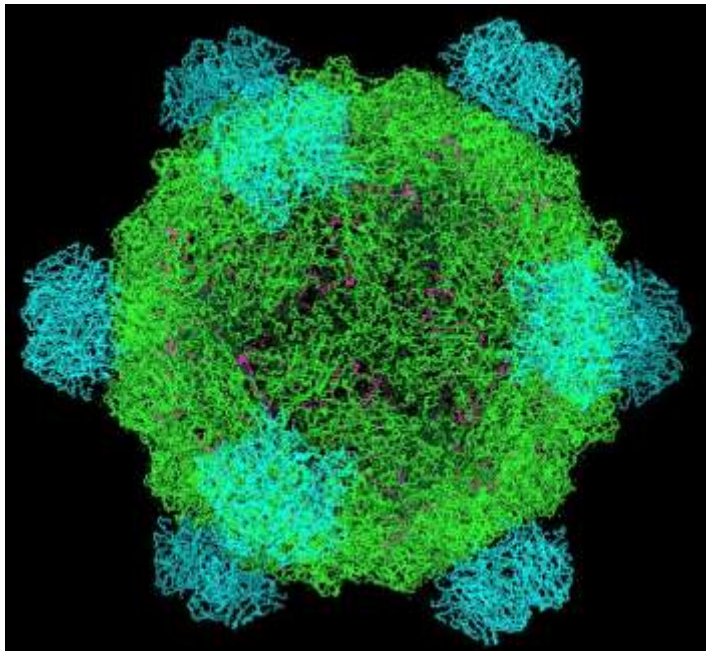
## One size fits all?





# DNA sequencing – a historical perspective

First genome (1977),  $\phi$ -X174: 5,386 bases.



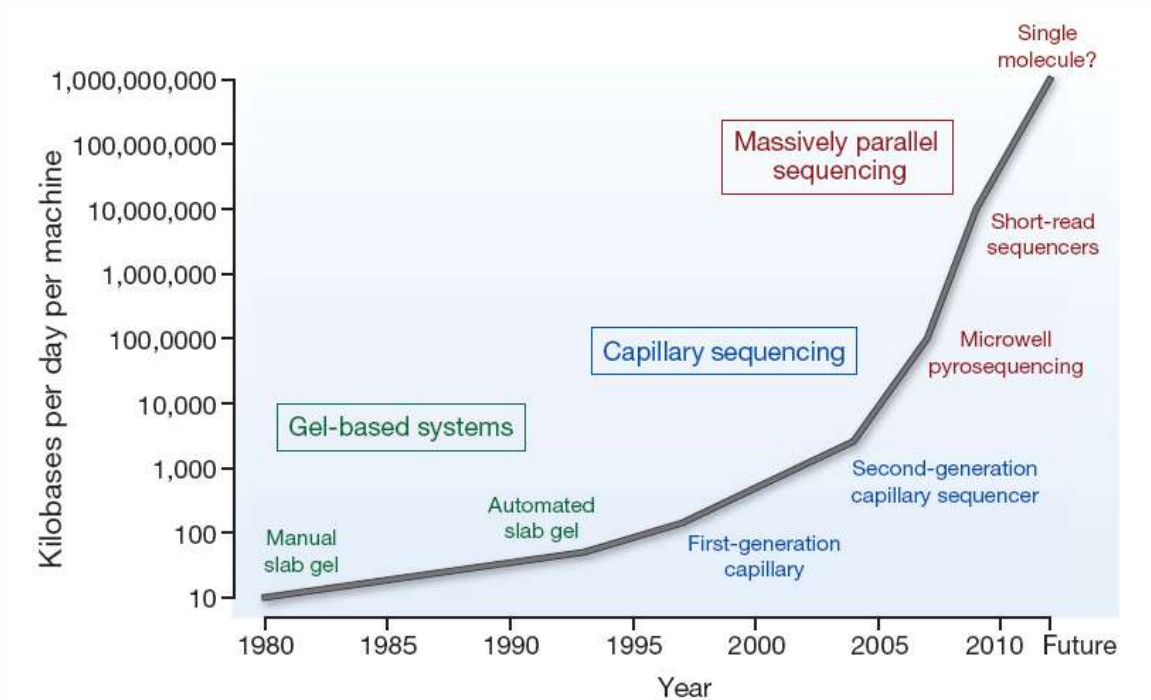
Fdardel / CC-BY-SA-3.0

```

agctcgtgagacttccaggacccgcaccaggctgtgggtttctcagataactgg
gcccctgagctcaggaggcctccacctctgctctggtaaaagtccatggaaacga
aaagaatggatttactgctcttcgcttgaagaagtacaaatgctcattaatgcta
tgcagaaaatcttagagtgctccatctgtctggagttgatcaaggaaacctgtccca
caaatgtgaccacatatttgcacatttggcctgctgaaactctcaaccagaa
aaggccctccacagtgctcctttagtgaagatgataaccacaaggagcctcaag
aaagtacagatctgctcaactgtgaaagctatfgaaatcatttgtgctttc
agctgacacaggttggagtagcaaacagctataatftgcaaaaggaaata
actctccgaaatcacaagaagaagtctctatccaaagtatgggctacagaa
accgtgcaaaagactctacagagtgaaacccgaaatccttccctgacagaaacaa
gtctcagtgccaaatctcaacctggaaactgtgaaactctgaggaacaaagcagc
ggatcaaacctcaaaagacgtctgtctacattgaaatgggctctgattctctgaa
atacctgttaataaggcaacttattgctgttgggagatcaagaattgttacaatca
cccccaaggaaacagggatgaactcagtttggattctgcaaaaaggctgctgtg
aatttctgagagcggatgtaacaaactctgaaactctcaaccagtaataatgatt
tgaacacccctgagaagcgtgcagctgagagggcatccagaagatcagggtagtt
ctgttccaaactgcatgtgggacatctgggcaaacctctcaagcagctcatfac
agctgagaaacagcagtttattactcactaaagacagaatgaaatgaaagagctg
aattctgtaataaaagcaaacagcctggcttagcaaggagcccaactaacagatggg
ctggagtaggaacaatgttaatgtagggggactcccggcagagaaagaggttag
atctgaaatgctgatcccctgtgtgaggaagaaatggaaataagcagaactgcaat
gctcagagaaatcctagagatctcgaagatgttccctggataaacactaaatagcagca
ttcagaaggttaatgagtggtttccagaagtgatgaactgttaggttctgtagact
cacaatgagggagctgcaatcaaaagcacaagtagctgagtagttaggacttctaa
atgaggtgaaatgaaatctcgggtctccagagaaatagacttacggccagtagctc
ctcaatgaggtttaaatagttaaaagtagaaaggttccctccaaatcagtagagtag
atattgaaacaaatattgggaaacactatcggaaagggcagagcctcccact
taagcagtagtaactgaaatctaatataggagcatttgttactgagccacagataa
taacagagcgtcccctcaaaatcaaatcaagcgtaaagggagactacatcagggc
ttcaatctgaggttttatcaagaaagcagatttggcagttcaaaagactcctgaa
tgataaatcagggaaactaaaccaagggacagaaatggctcaagttagaataatctctc
atagtggtcagtagaatcaaaccaaggtgattctatcagaatgagaaatctctc
accocataagatcactcgaaaaagaaatctgcttcaaaacgaagctgaaacctataa
ggcagtagtaagcaatctggaactcgaatcaaatctcaacaatcaaaagcactca
aaagaaatggcttggggggaggtctctcaccaggaatctcaagcgttgaactag
tagtctgtgaaatcaagccaccatattgactgaattgcaaatgtagttagtt
ctgctgtaaaagataaaagaaaaaaagtaaaccaaatgcaagtaaaagcagca

```

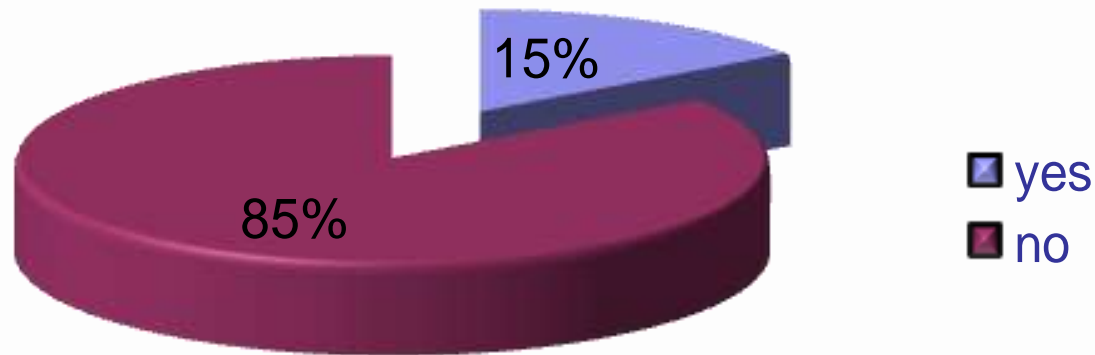
# Progress in DNA sequencing



Stratton et al. Nature Vol 458 April 2009

## The 'Next Generation' in Sequencing

**Q:** "Are you familiar with the new generation of emerging high throughput sequencing technologies eg, 454 Genome Sequencer 20, Agencourt SOLiD and Solexa 1G Genome Analyzer?"



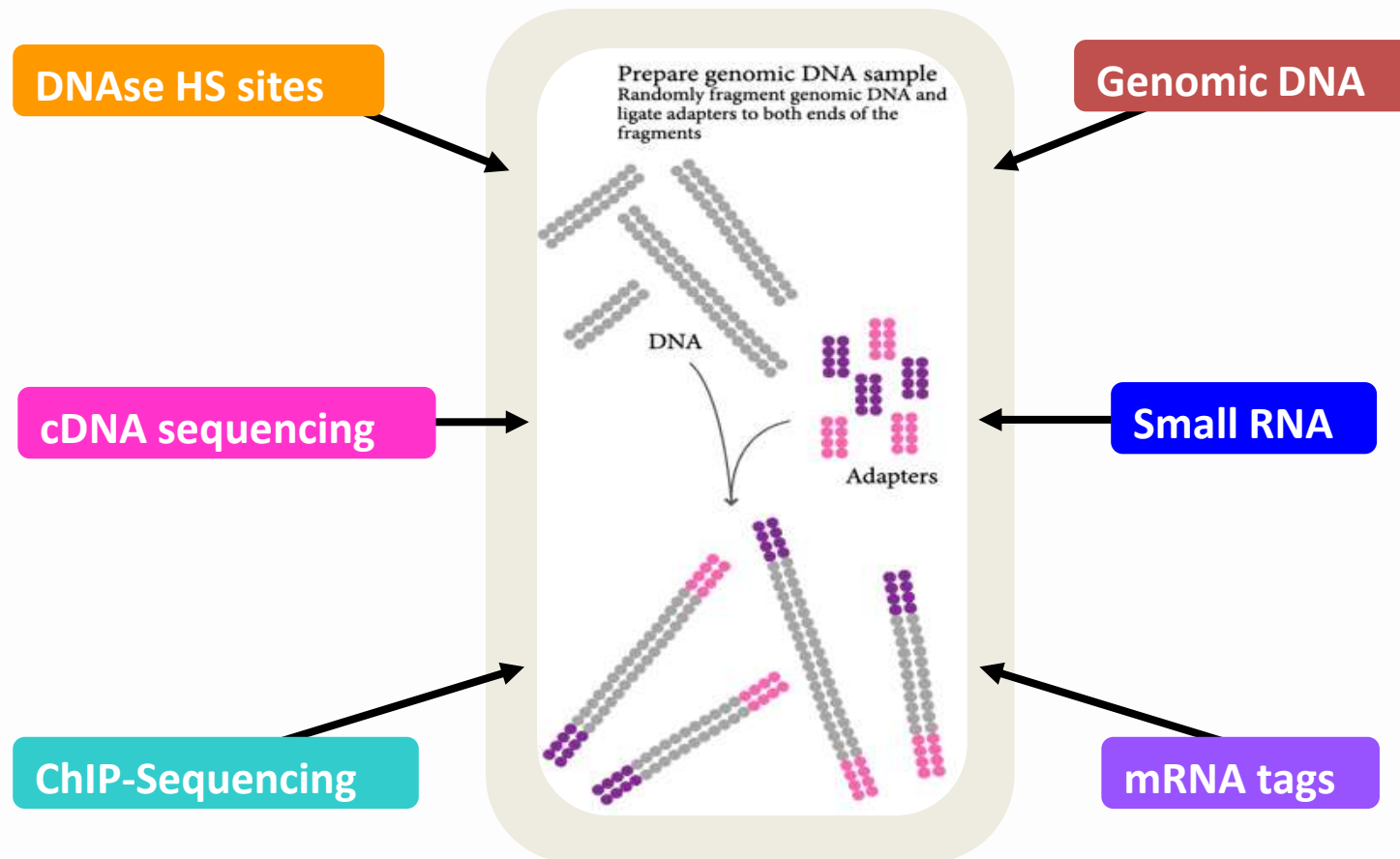
(April 2007)

## Why sequence?

Sequencing of DNA, RNA and small RNA can reveal:

- SNPs
- somatic mutations
- copy number variations
- DNA methylation (methyl-seq/MeDIP-seq)
- genomic structural re-arrangements
- DNase hypersensitivity
- Protein:DNA interactions (ChIP-seq)
- Protein:RNA interactions (RIP-seq)
- DNA:DNA interactions (3C-seq)
- mRNA structural variations (RNA-seq)
- mRNA expression levels
- small RNA (miRNA) discovery and expression levels
- RNA changes (RNA editing)

## Sequencing applications





# Whole genome sequencing

- Approaching point where whole genomes will be within reach.
- Currently, sequencing costs still relatively high.

## ARTICLES

nature  
genetics

### Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing

Yan J Cao<sup>1</sup>, Philip J Stephens<sup>1,2</sup>, Eric D Pleasance<sup>1</sup>, Sarah O'Meara<sup>1</sup>, Hong Li<sup>1</sup>, Thomas Szallasi<sup>1,3</sup>, Clare Harly<sup>1</sup>, Ian W Toussie<sup>1</sup>, Andrew Menzies<sup>1</sup>, Richard A Quail<sup>1</sup>, Anthony Cox<sup>1</sup>, Clive Brown<sup>1</sup>, Graham R Bignell<sup>1</sup>, Michael R Stratton<sup>1,4</sup>

## ARTICLES

### A comprehensive catalogue of somatic mutations from a human cancer genome

Eric D Pleasance<sup>1,2</sup>, R Kiro Chubbam<sup>1,3</sup>, Philip J Stephens<sup>1</sup>, David J McLeod<sup>1</sup>, Sean J Humphrey<sup>1</sup>, Charles Greenway<sup>1</sup>, Grant M Karali<sup>1</sup>, Meng-Lay Lin<sup>1</sup>, Genovita R Ostriner<sup>1</sup>, Graham R Bignell<sup>1</sup>, Ian W Toussie<sup>1</sup>, Mark J Sauer<sup>1</sup>, David Hoadley<sup>1</sup>, Adam Butler<sup>1</sup>, Richard J Carter<sup>1</sup>, Lisa Chen<sup>1</sup>, Anthony J Cox<sup>1</sup>, Sarah Lobb<sup>1</sup>, Paul J Baker-Gambell<sup>1</sup>, Heidi A Gorman<sup>1</sup>, Robert Tarone James<sup>1</sup>, Mingming Ji<sup>1</sup>, Jyoti Kishore<sup>1</sup>, Co Laura J Mudd<sup>1</sup>, Daniel Ning<sup>1</sup>, Tom Ross<sup>1</sup>, On B S Loken Sjölander<sup>1</sup>, Ian Tatters<sup>1</sup>, David Williamson David R Bentley<sup>1</sup>, P Andrew Futreal<sup>1</sup> & Michael R

123456789

## LETTERS

### Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes

Gilles L Delguez<sup>1</sup>, Kyle Fung<sup>1</sup>, Chris Grotzer<sup>1</sup>, Lisa Chen<sup>1</sup>, Graham Bignell<sup>1</sup>, Adam Butler<sup>1</sup>, Helen Davies<sup>1</sup>, Sarah Edkins<sup>1</sup>, Claire Harly<sup>1</sup>, Gill Lohner<sup>1</sup>, Ian Tatters<sup>1</sup>, Jenny Andrews<sup>1</sup>, Svitlana Borkova<sup>1</sup>, Dave Bovee<sup>1</sup>, Gemma Burns<sup>1</sup>, Peter J Campbell<sup>1</sup>, Simon Farber<sup>1</sup>, Mingming Ji<sup>1</sup>, Daniel Jones<sup>1</sup>, Henry Khatib<sup>1</sup>, Chen Xia Kuo<sup>1</sup>, King Wai Lau<sup>1</sup>, Catherine Laroy<sup>1</sup>, Meng-Lay Lin<sup>1</sup>, David J McLeod<sup>1</sup>, Mark Middleton<sup>1</sup>, Simon Maguire<sup>1</sup>, Kirstin McLay<sup>1</sup>, Andrew Menzies<sup>1</sup>, Tatiana Mironenko<sup>1</sup>, Lee Muldering<sup>1</sup>, Laura Mudd<sup>1</sup>, Sarah O'Meara<sup>1</sup>, Erin Pleasance<sup>1</sup>, Arjunan Rajalingam<sup>1</sup>, Rebecca Shepherd<sup>1</sup>, Raffaella Smith<sup>1</sup>, Lucy Stubbings<sup>1</sup>, Philip Stephens<sup>1</sup>, Gerard Tang<sup>1</sup>, Patrick S. Tarczy<sup>1</sup>, Kelly Tunstall<sup>1</sup>, Karl J. Dijkema<sup>1</sup>, Suk Kwan Khoo<sup>1</sup>, David Paterson<sup>1</sup>, Bill Wongkam<sup>1</sup>, John Azzari<sup>1</sup>, Richard J. Katschek<sup>1</sup>, Bin Tian Teo<sup>1,2</sup>, Michael R. Stratton<sup>1,2</sup> & P. Andrew Futreal<sup>1</sup>

Clear cell renal cell carcinoma (ccRCC) is the most common form of adult kidney cancer, characterized by the presence of inactivating mutations in the VHL gene in most cases<sup>1</sup>, and by subsequent somatic mutations in tumour suppressor genes. To determine further the genetics of ccRCC, we have sequenced 101 cases through 5,248 protein-coding genes. Here we report the identification of inactivating mutations in two genes standing out from the rest: histone modifications—SETD3, a histone H3 lysine 36 methyltransferase, and JARID1C (also known as KMT6C), a histone H3 lysine 9

measures, some of which may be implicated in cancer development are ubiquitous and low-frequency and have low (to greater) impact than those high-risk of clonal expansion. By investigating read pairs that did not align correctly we inferred 100,000 structural variants and 103 somatic deletions and somatic rearrangements were markedly different segments outside these regions, notably including tandem duplications, including two low-interval tandem duplications. Greater variants were preferentially on the reads denoting the stability of structural variants, cancer genes, driving the presence of a new class of

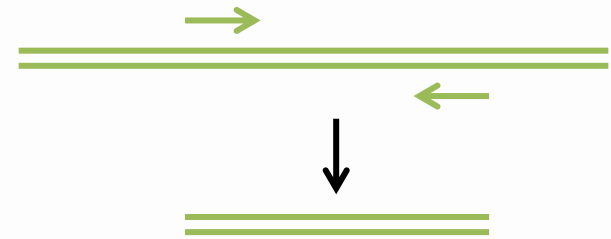
123456789



## In the meantime: two principle methods for targeting regions of interest

### 1. Single-gene sequencing:

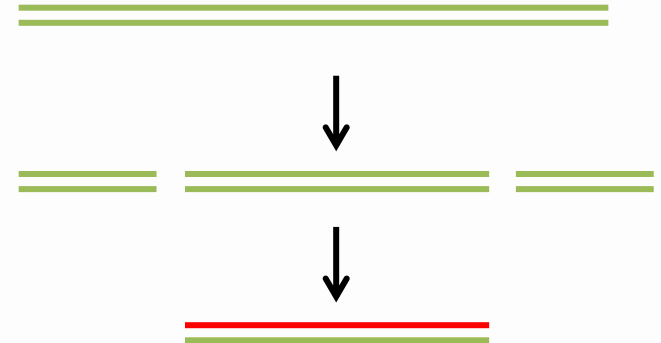
Gene-specific PCR (amplicon sequencing)



## In the meantime: two principle methods for targeting regions of interest

### 1. Single-gene sequencing:

Gene-specific PCR (amplicon sequencing)



### 2. 'Sequence capture' methods for larger regions:



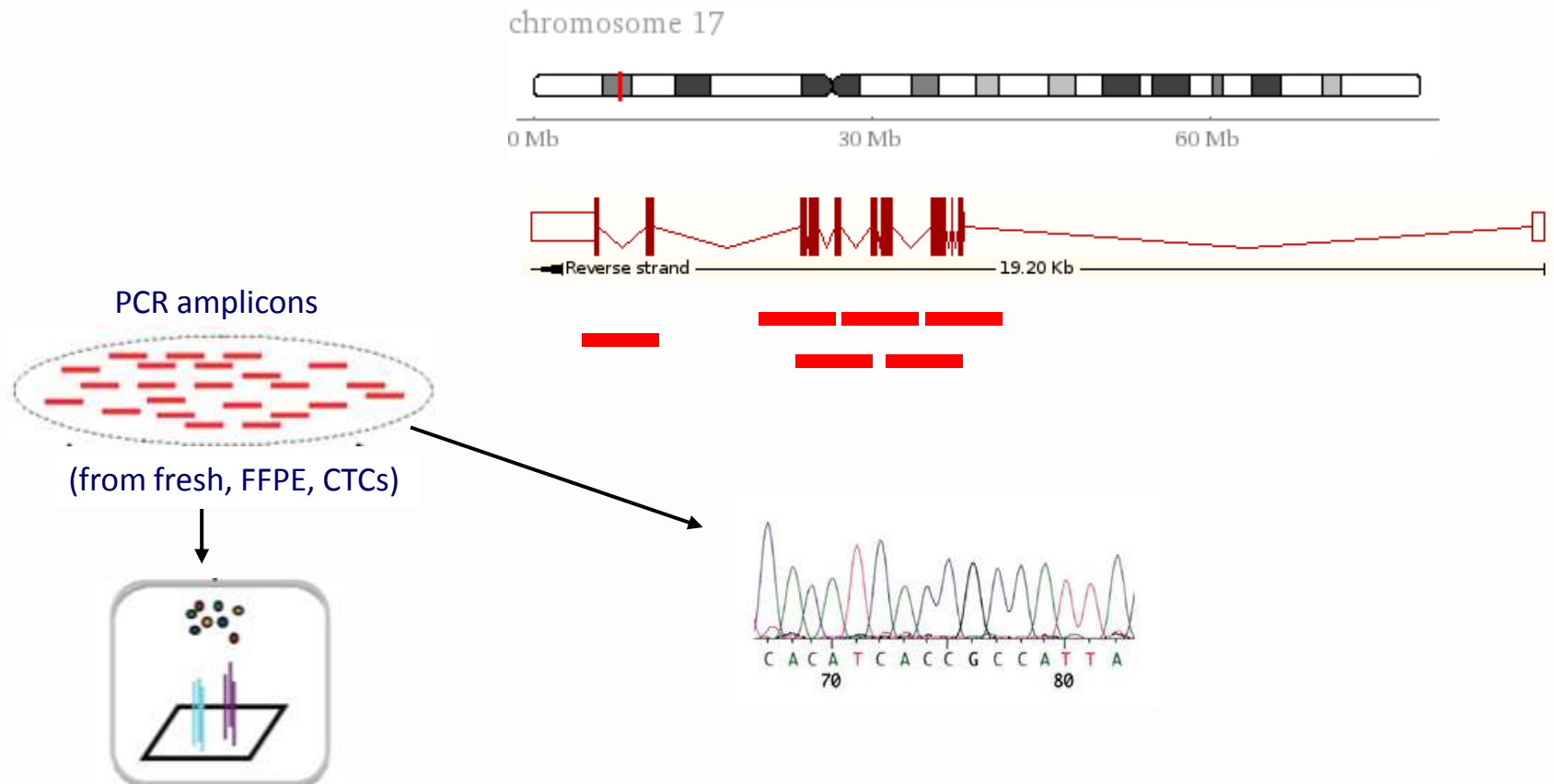
e.g. 'Exome' sequencing:  $\sim 50 \times 10^6$  bp including all known protein -coding regions.

## Next Generation Sequencing @ SBS

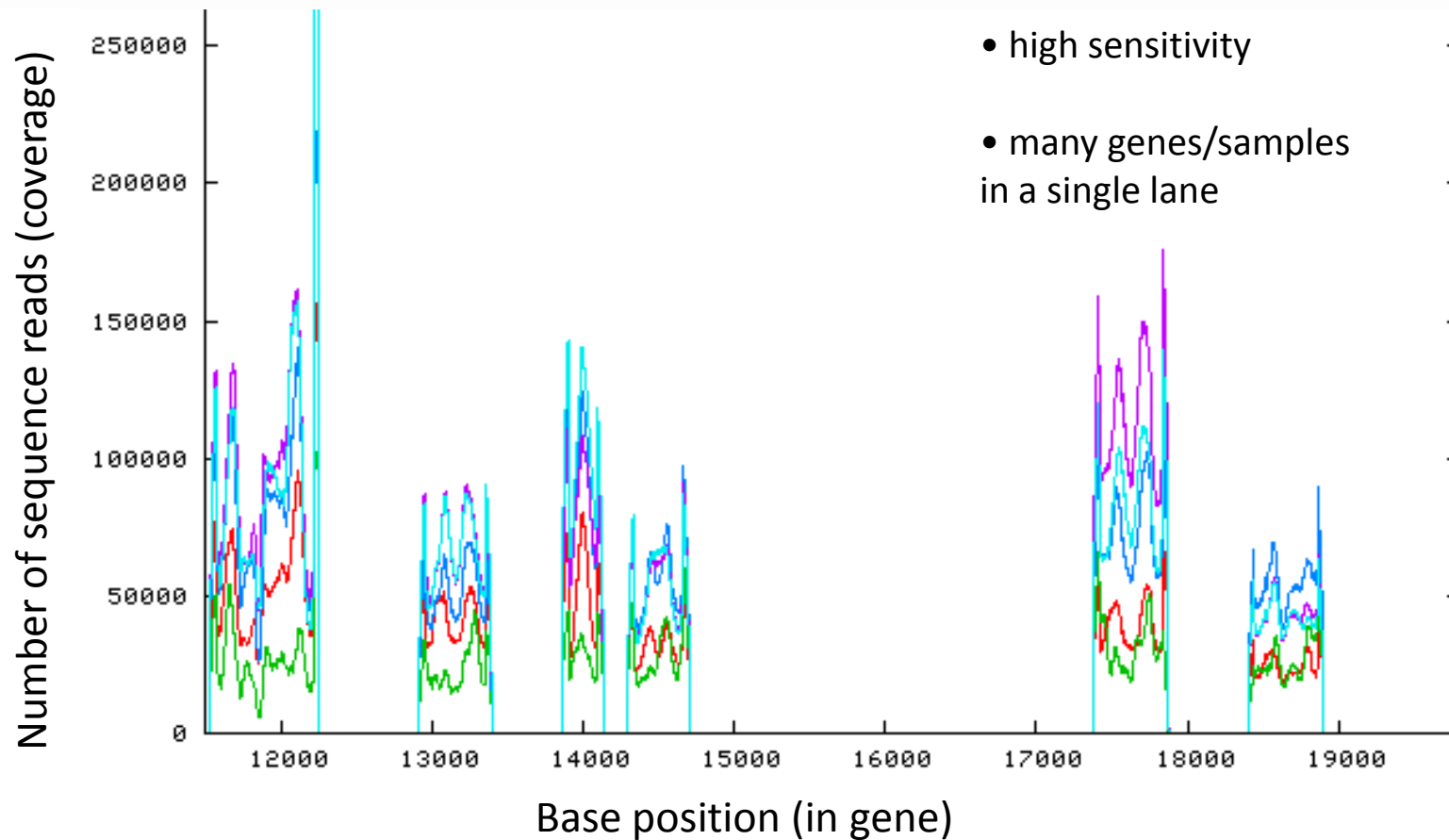
1. Single-gene sequencing: 'deep' sequencing

2. 'Sequence capture' methods for larger regions: *personalised medicine*

# Deep sequencing



## Deep sequencing



# The future of personalised medicine?

**Personalised Medicine:** “A medical model using molecular profiling technologies for tailoring the right therapeutic strategy for the right person at the right time.”

## CRUK Stratified Medicines Programme

Aimed at demonstrating utility of molecular diagnostics in healthcare

- **Phase I:** (two year pilot programme)
  - 9000 samples, multiple cancers (breast, bowel, lung, prostate, ovary, melanoma), multiple tests
- **Phase II:**
  - to extend into general practice, **to integrate innovative testing platforms**



## The future of personalised medicine?

Technology Strategy Board  
Driving Innovation

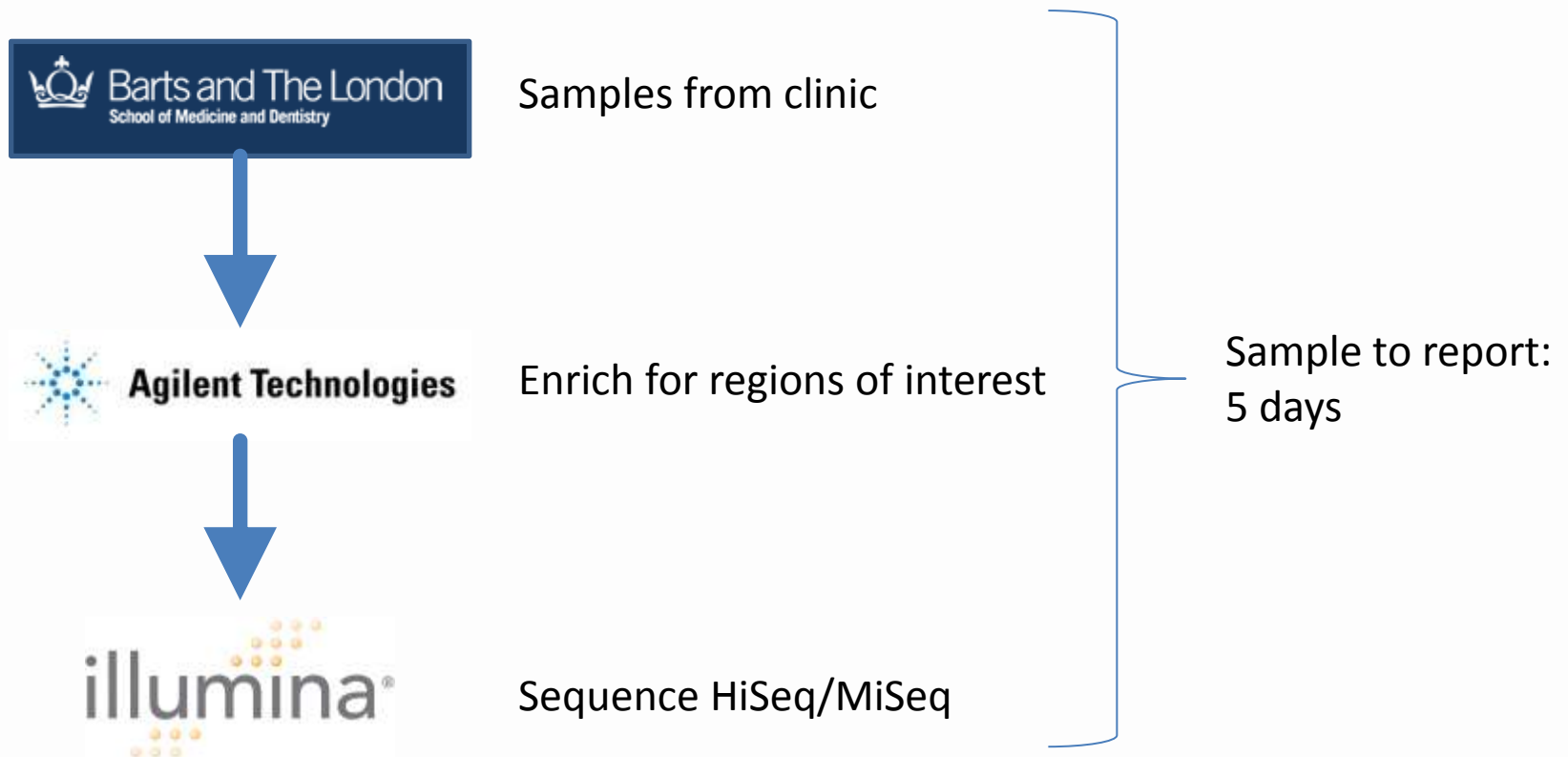


Stratified medicines programme:  
Tumour profiling and data  
capture to improve cancer care

COMPETITION FOR COLLABORATIVE R&D FUNDING  
JANUARY 2011



## The future of personalised medicine?



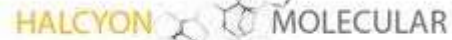
## The sequencing race ... no sign of slowing down



\$10 million  
100 'medical-grade' genomes



Complete  
genomics



HALCYON MOLECULAR



PACIFIC  
BIOSCIENCES™



picoseq



Helicos  
BioSciences Corporation



ion torrent



LIGHTSPEED  
GENOMICS™



VISIGEN  
BIOTECHNOLOGIES, INC



by life technologies™



Oxford  
NANOPORE  
Technologies®



IBM



Intelligent  
Bio  
Systems

