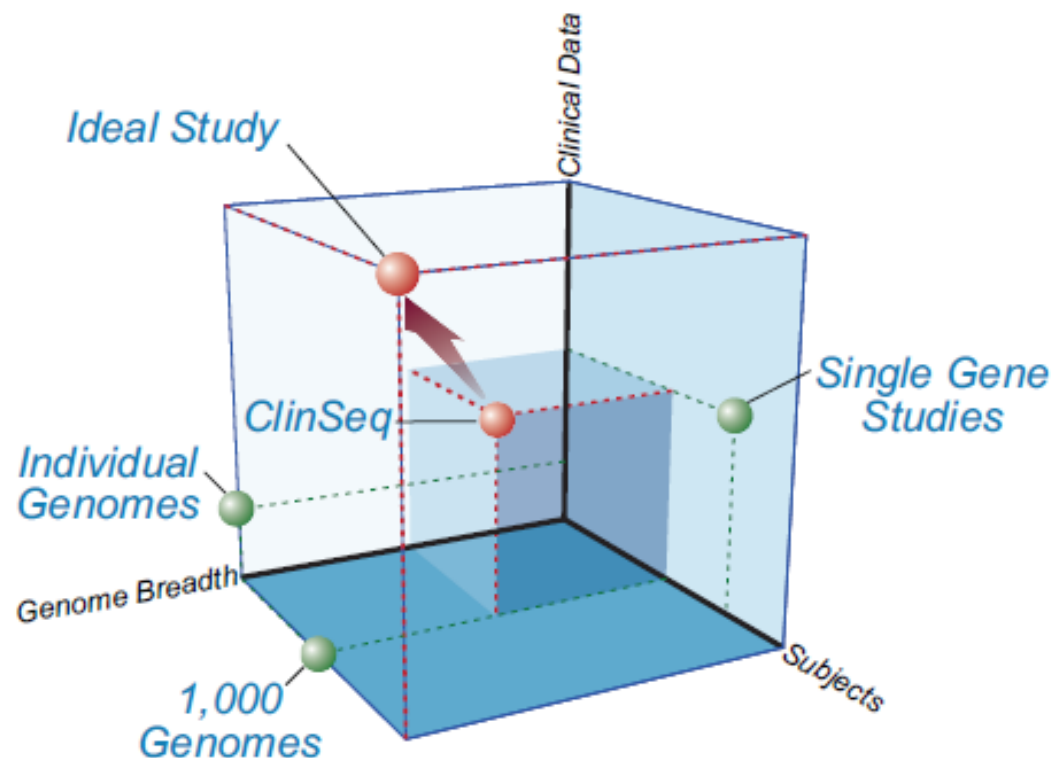


Clinical sequencing

Three key dimensions :

- Genome Breadth: The fraction of the genome that is interrogated
- Subjects: The number of participants used in a study
- Clinical Data: The amount of clinical data associated with the individuals



Biesecker et al 2009, Genome Research 19:1665-74

The Simplest Sequencing Process

1 Library prep (~ 6 hrs)



1000's M
DNA!

Fragment DNA
↓
Repair ends / Add A overhang
↓
Ligate adapters
↓
Select ligated DNA

2 Automated Cluster Generation (~ 5 hrs)



1-8 samples

Hybridize to flow cell
↓
Extend hybridized oligos
↓
Perform bridge amplification

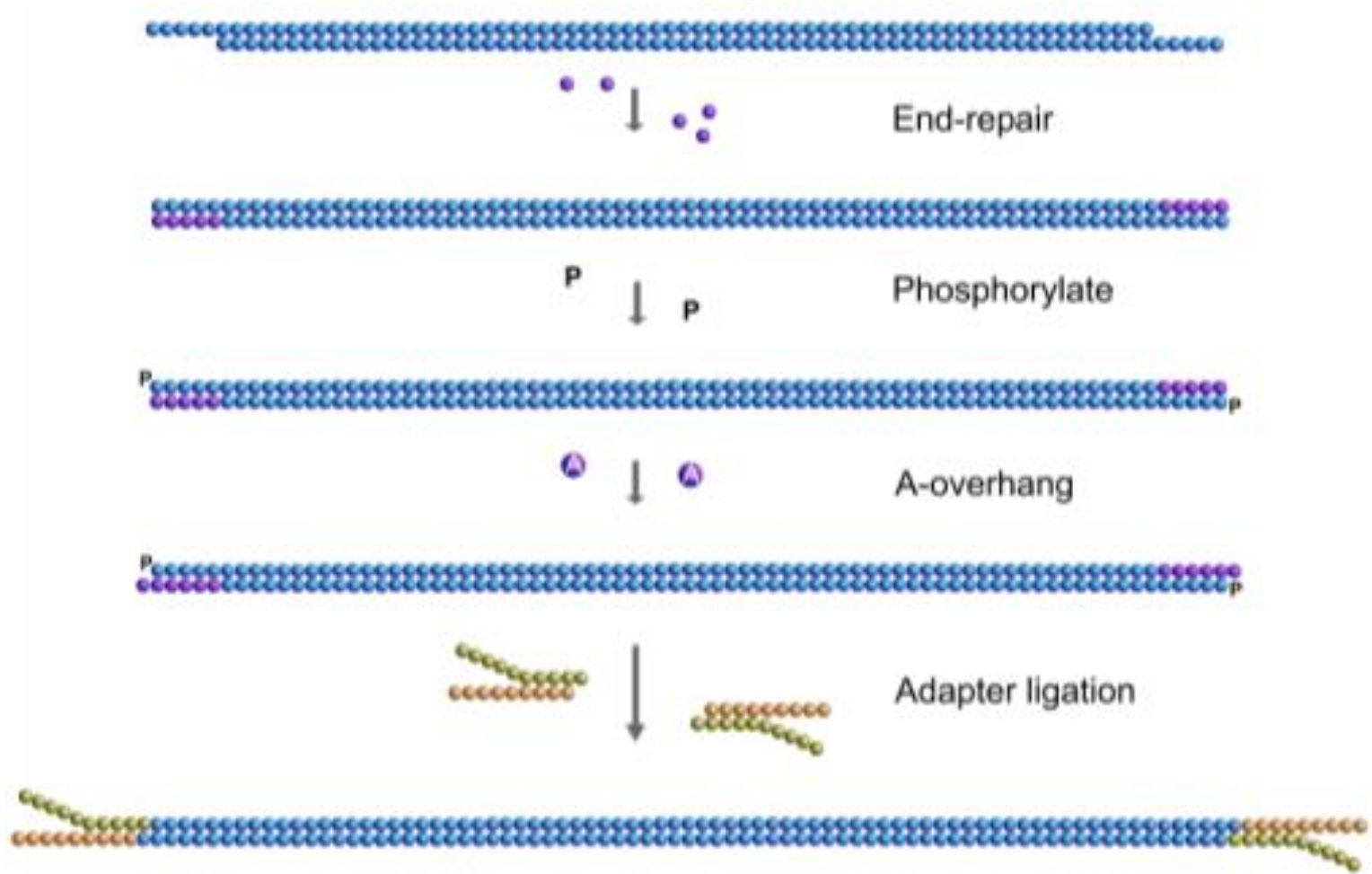
3 Sequencing (~ 1-8 days)



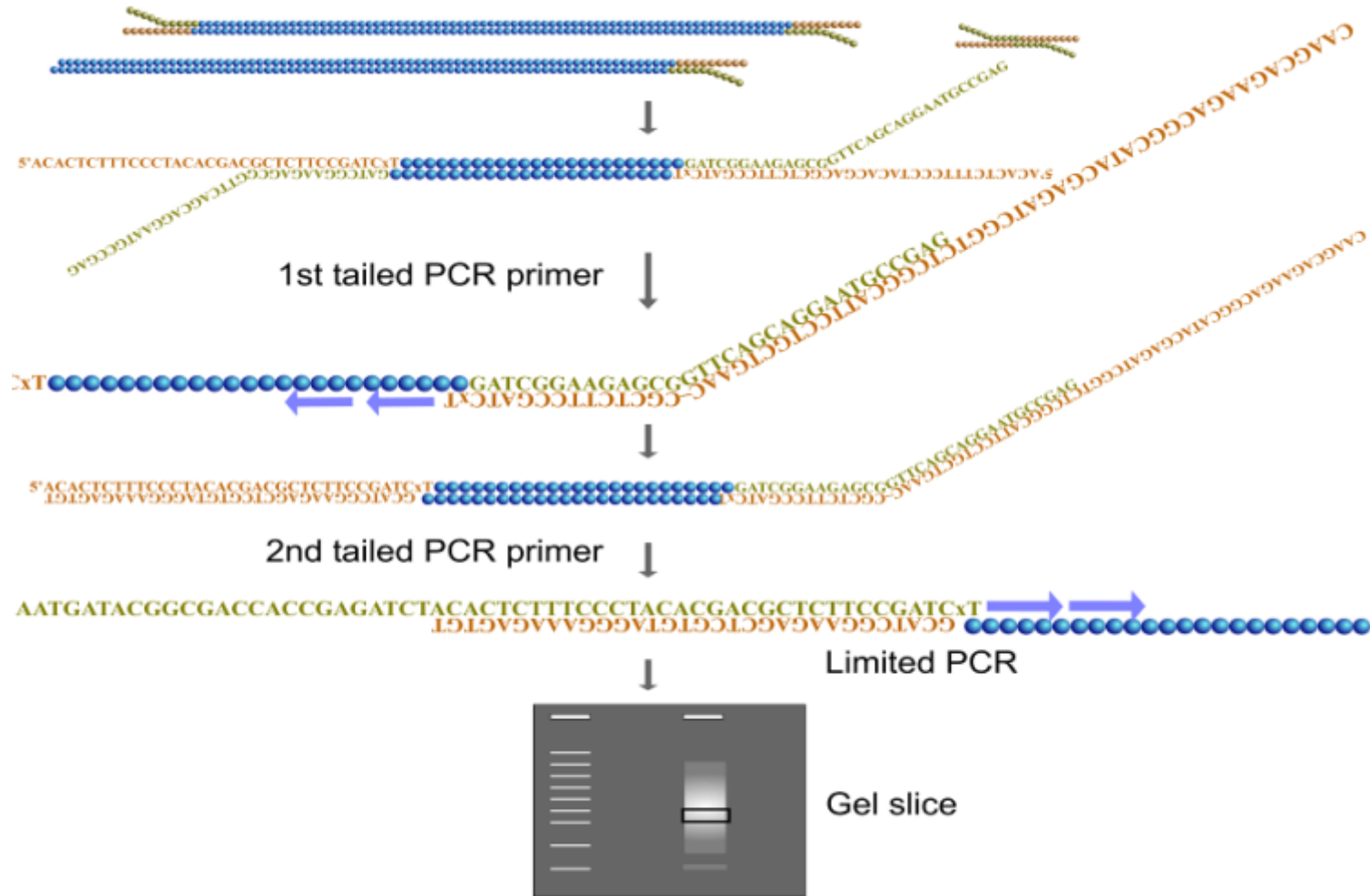
1-16 samples

Perform sequencing on forward strand
↓
Re-generate reverse strand
↓
Perform sequencing on reverse strand

Preparing libraries



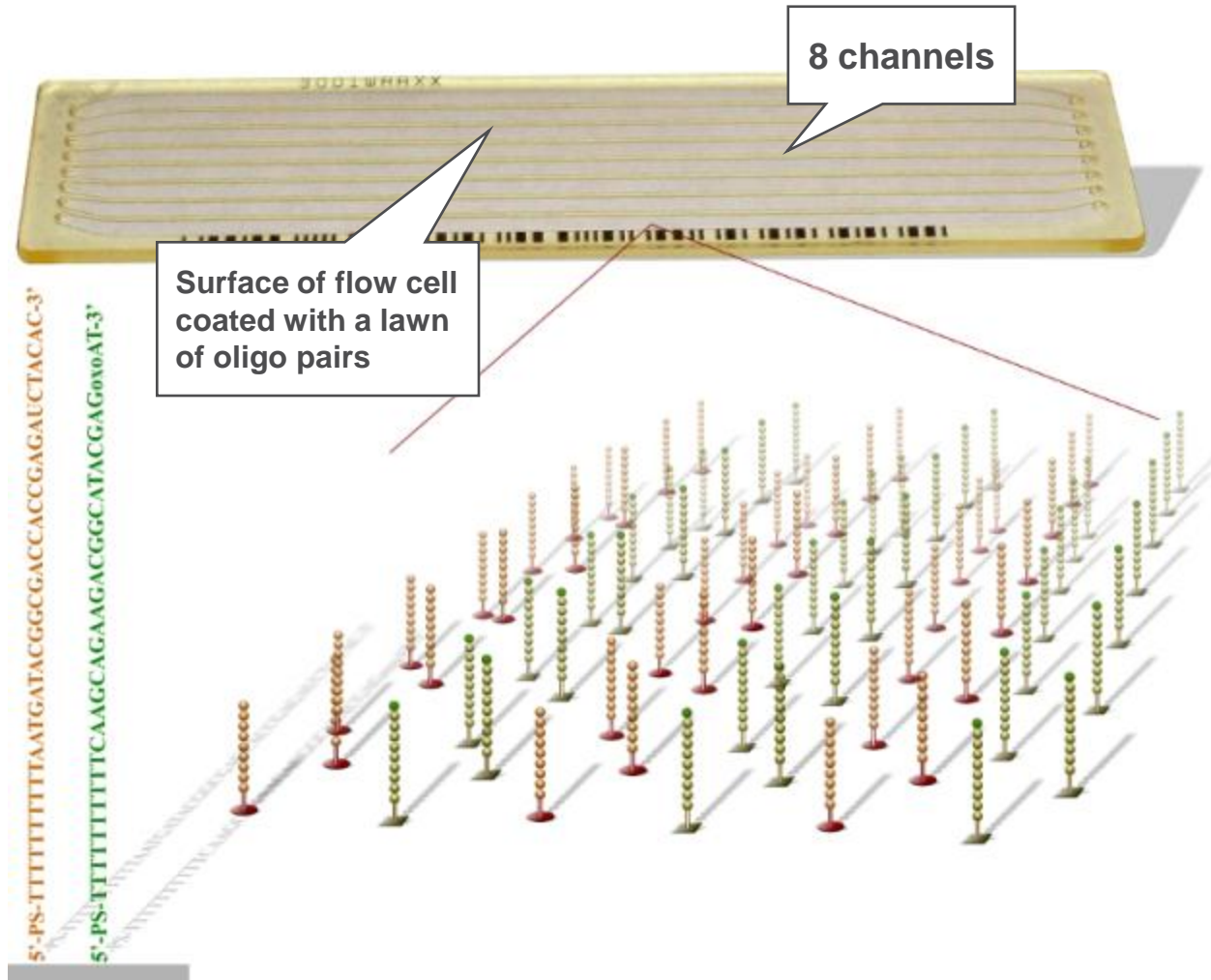
Purifying Libraries



PCR free modifications

- Short Y adapter is replaced with **a longer adapter**
- Adapter primer dimers after ligation are removed by **SPRI** or Sephadex beads
- Library containing DNA fragments (ligated, partially ligated and non-ligated) is introduced into flow cells
- Bridge amplification is performed on the library
 - Non-ligated DNA products do not bind to the flow cell
 - Partially ligated products bind but do not amplify
 - Ligated products bind and bridge amplify
- Cluster size is dependent on sequence content
- Algorithms detect all clusters with equal efficiency

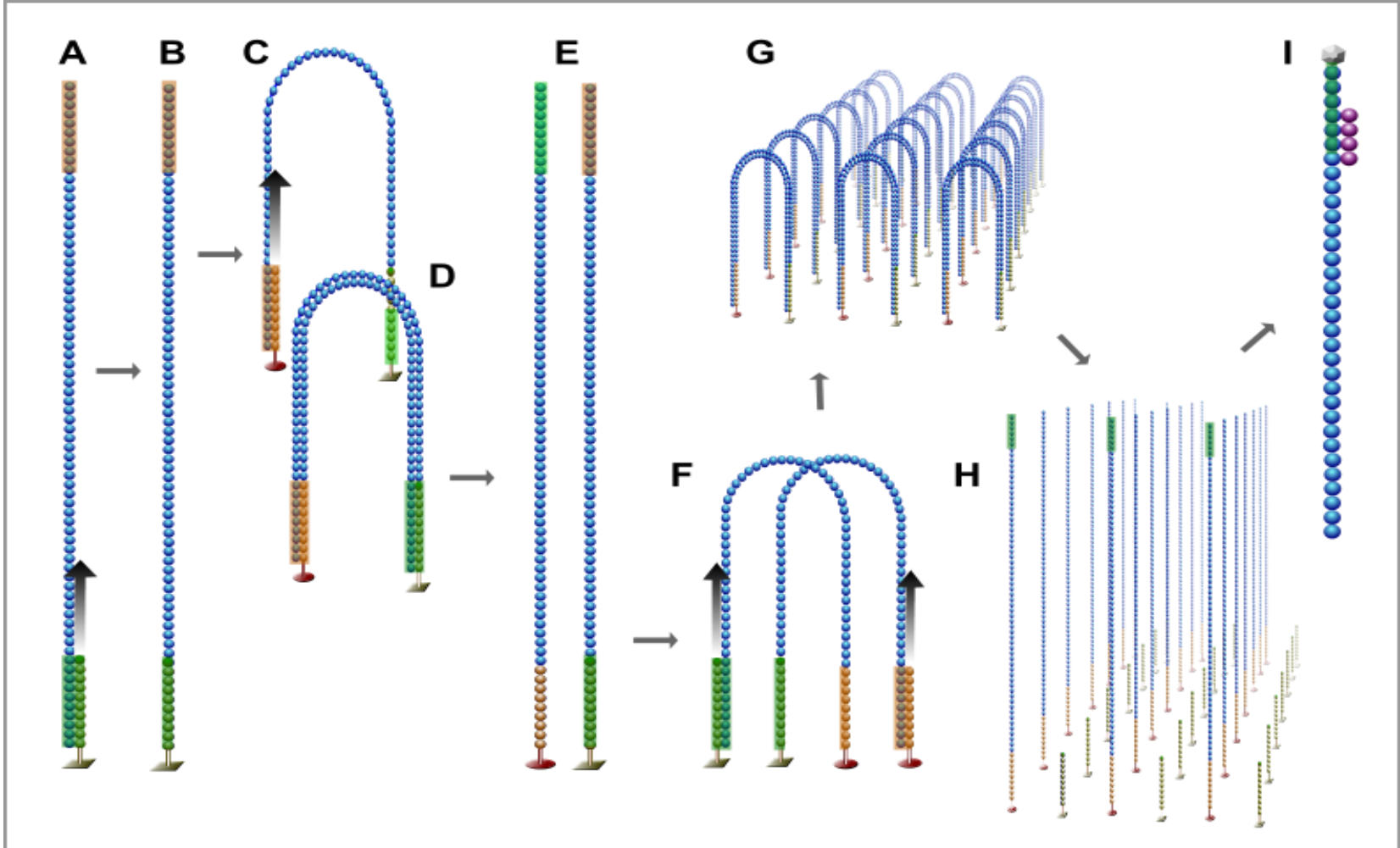
Flow Cell



Simplified workflow

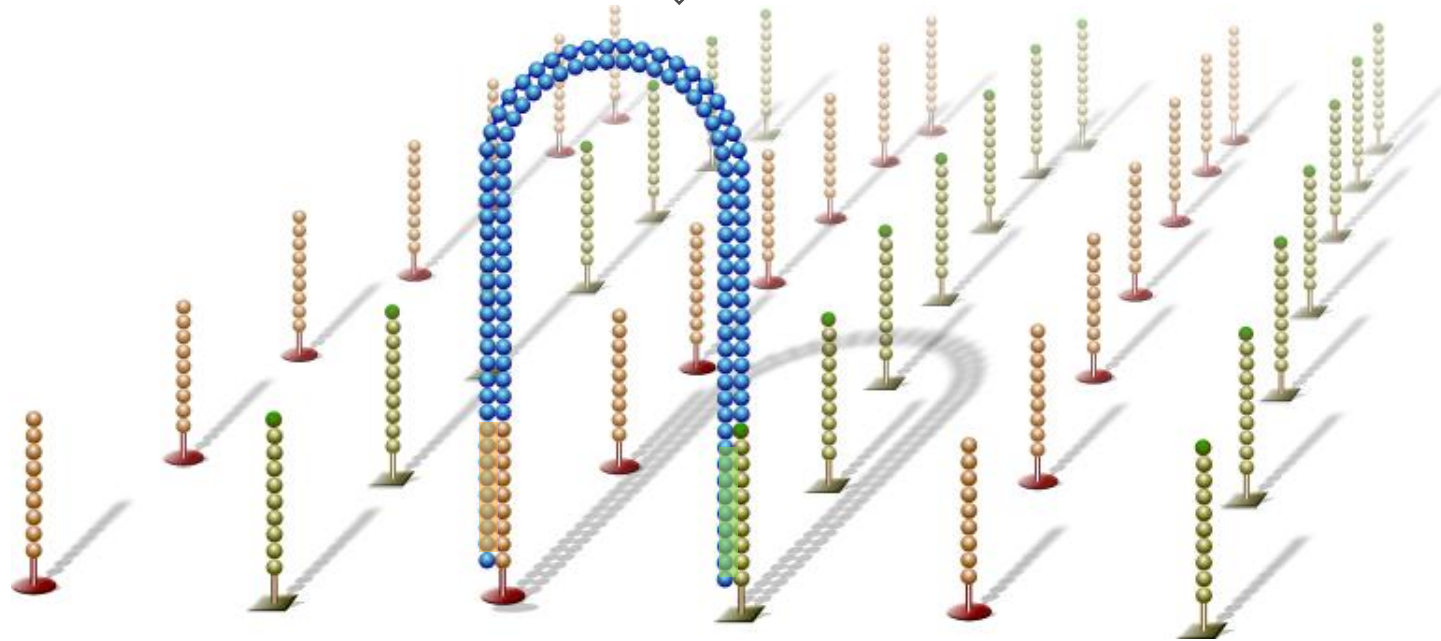
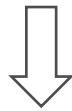
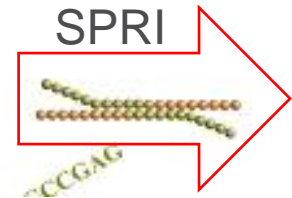
- Clusters in a contained environment (no need for clean rooms)
- Sequencing performed in the flow cell on the clusters

Cluster generation



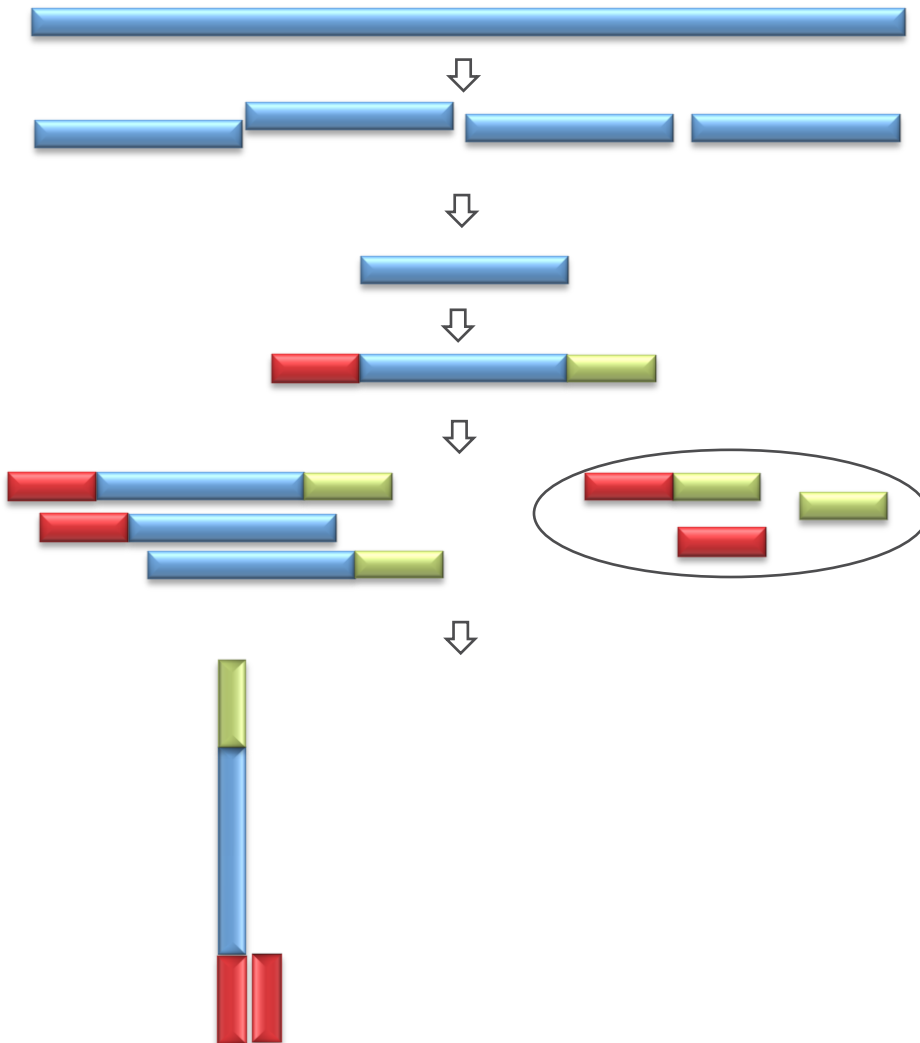
Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

PCR free workflow



Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

Reverse transcription in flow cell



1. RNA
2. Fragment
3. Repair
4. Ligate RNA adapters
5. Remove free adapters and adapter-adapter dimers
6. Introduce into flow cells
7. Reverse transcribe
8. Bridge amplify

The Simplest Sequencing Process

1 Library prep (~ 6 hrs)



1000's M
DNA!

Fragment DNA
↓
Repair ends / Add A overhang
↓
Ligate adapters
↓
Select ligated DNA

2 Automated Cluster Generation (~ 5 hrs)



1-8 samples

Hybridize to flow cell
↓
Extend hybridized oligos
↓
Perform bridge amplification

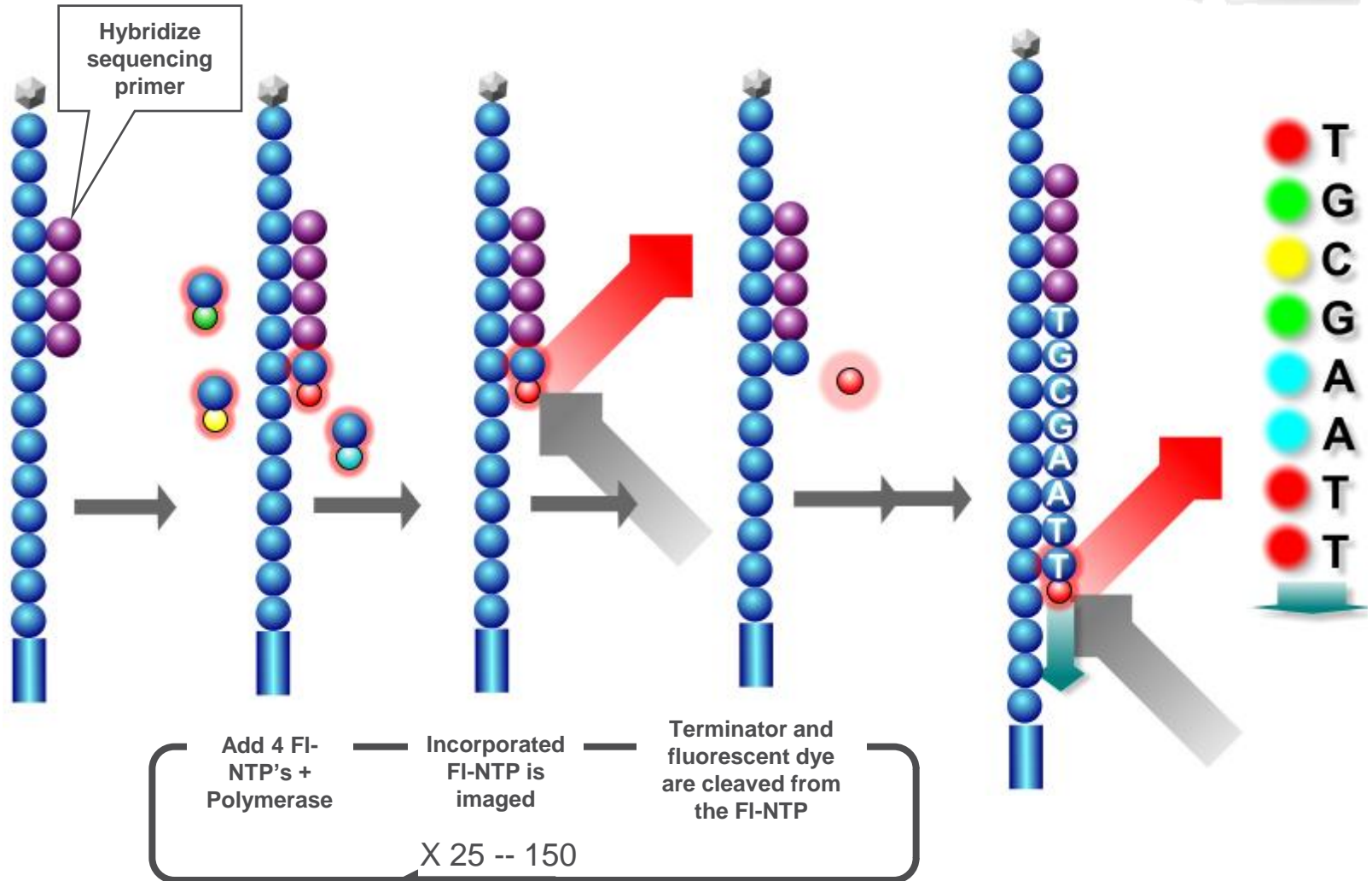
3 Sequencing (~ 1-8 days)



1-16 samples

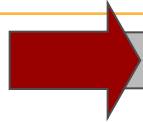
Perform sequencing on forward strand
↓
Re-generate reverse strand
↓
Perform sequencing on reverse strand

Sequencing Forward Strand



Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

Sequencing with Paired Ends



Reference

This is really the best way to do sequencing

Single-reads

This is

...

is really

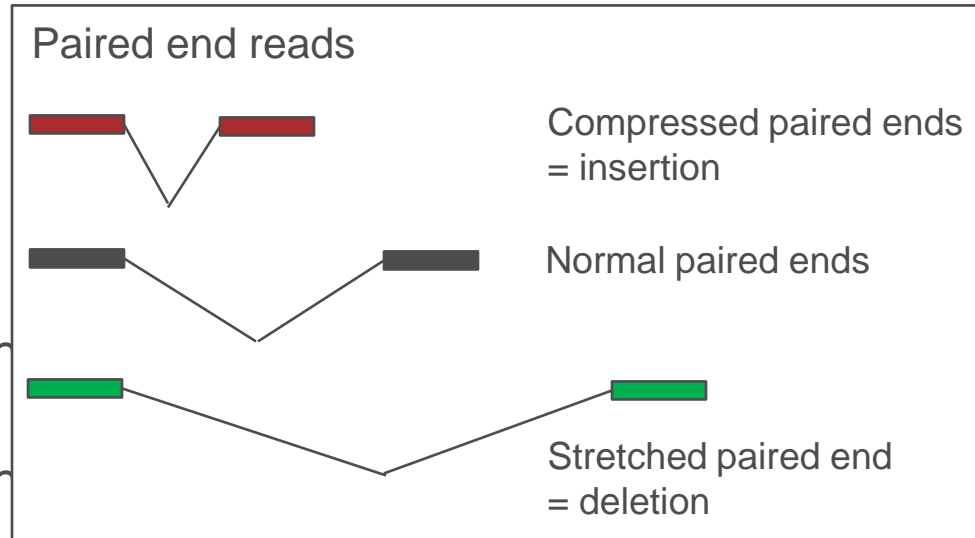
...

really th

...

th

...



sequencing

Paired-reads

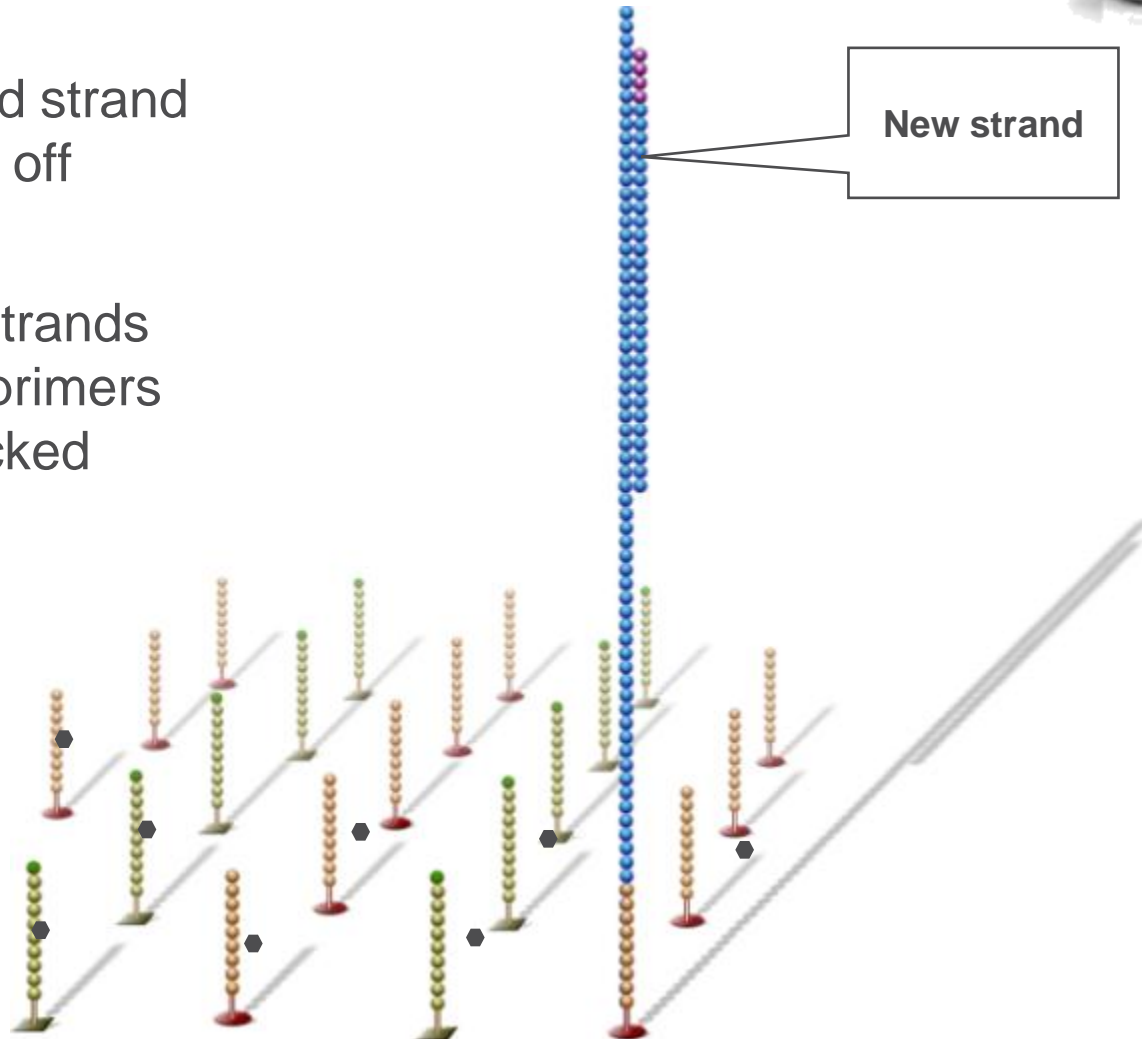
This is (-----26 characters-----) sequencing

Assembly becomes easier!!

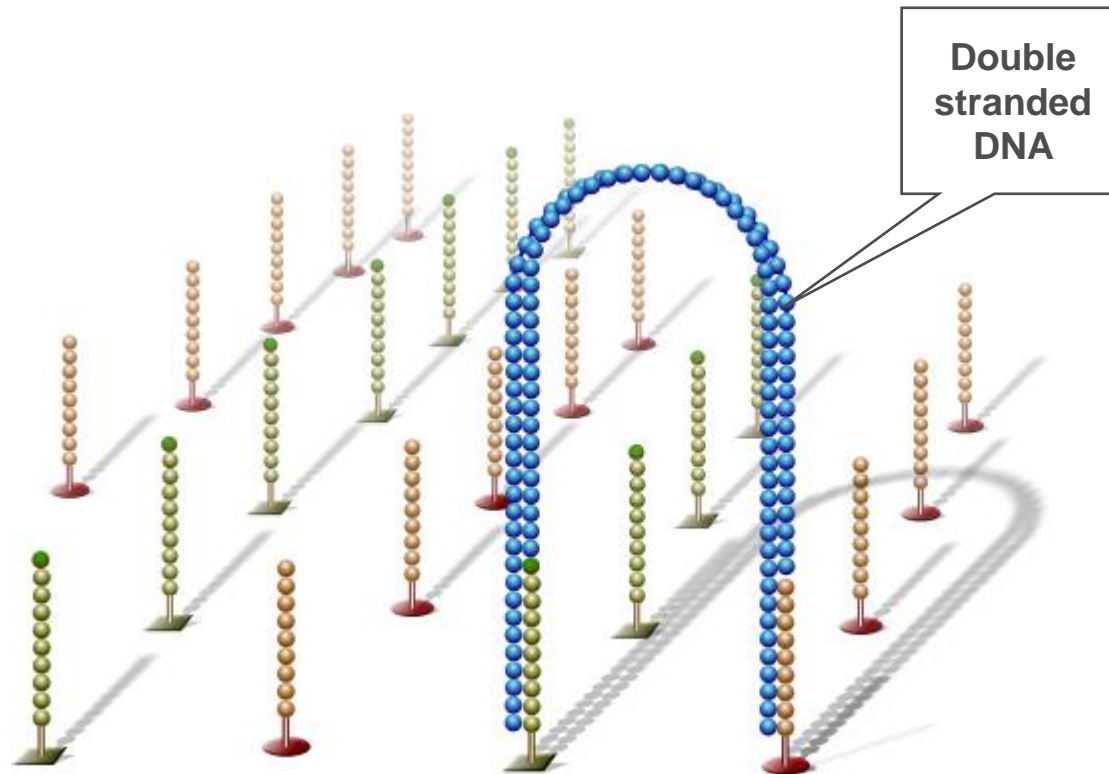
Paired End Sequencing



- Sequenced strand is stripped off
- 3'-ends of template strands and lawn primers are unblocked



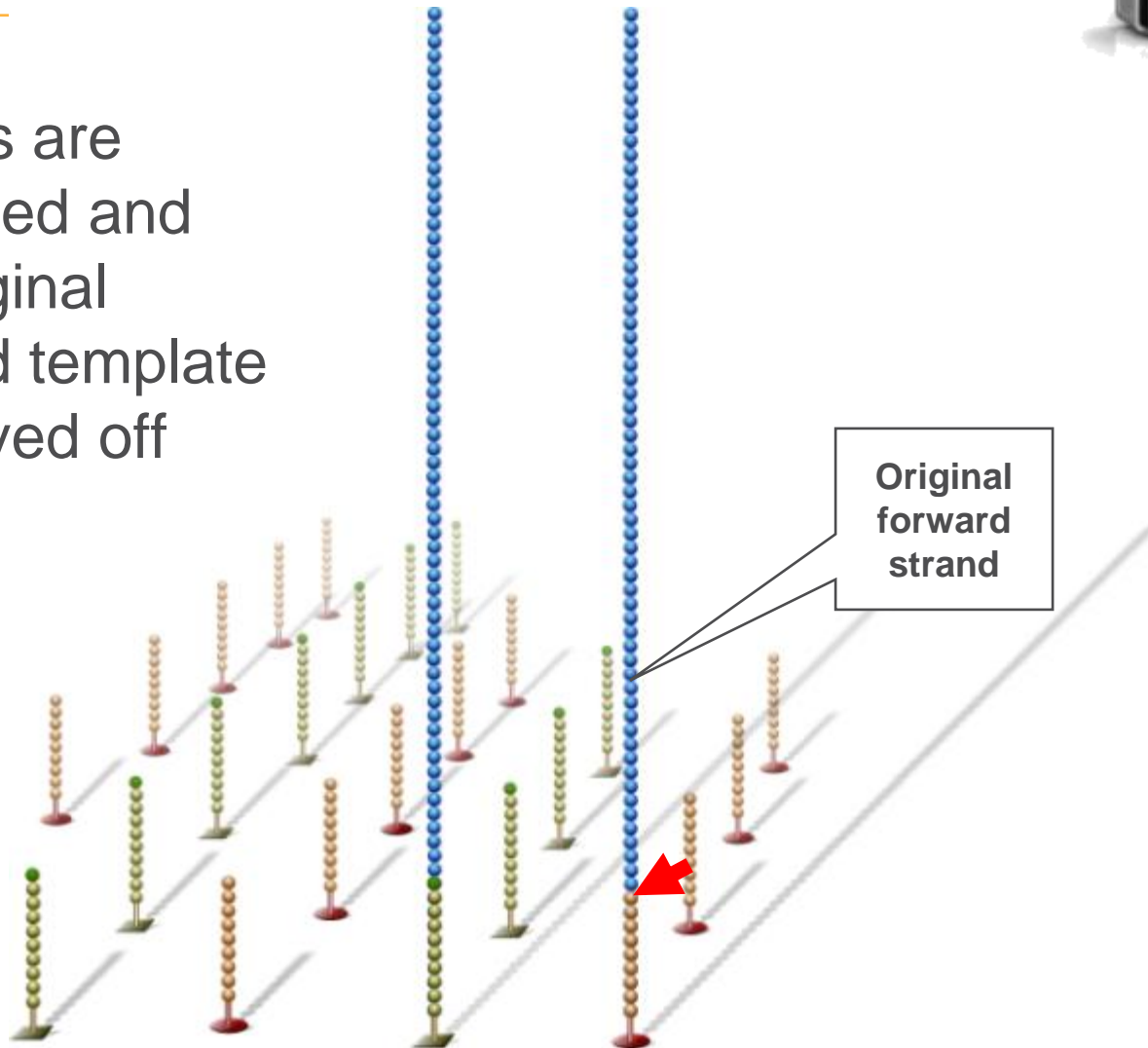
Paired End Sequencing



Paired End Sequencing



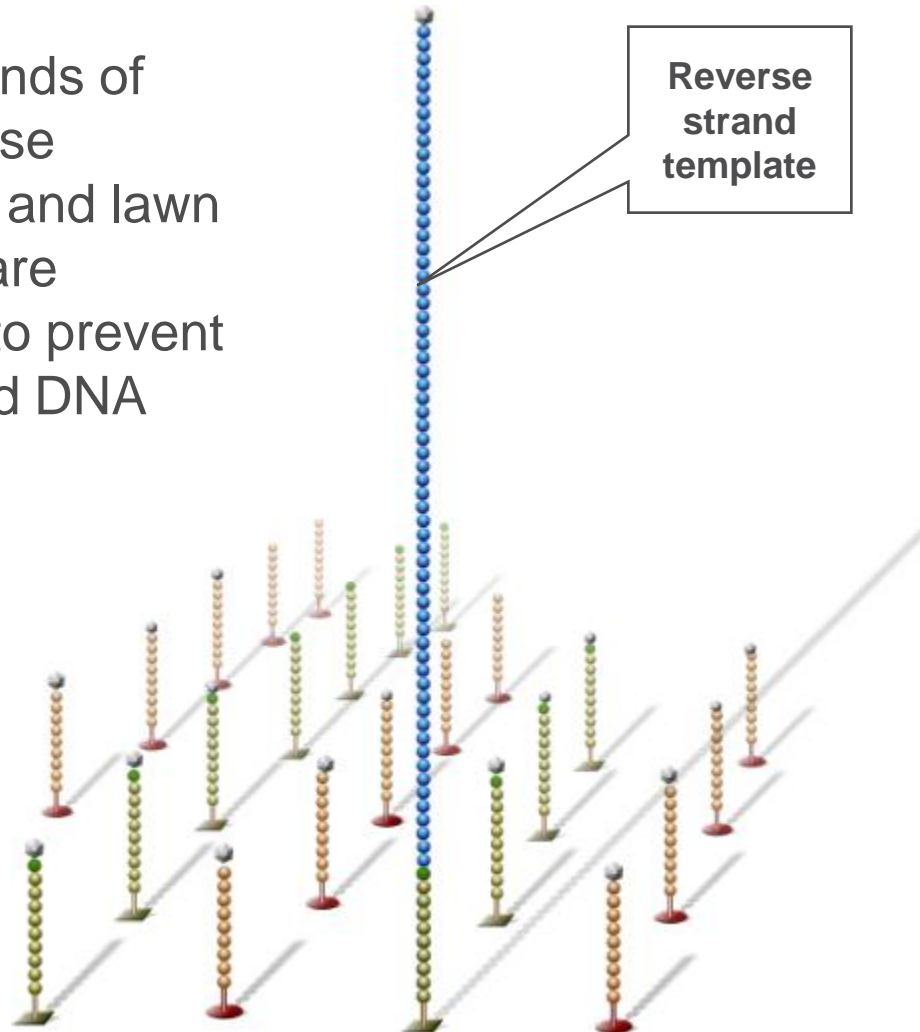
- Bridges are linearized and the original forward template is cleaved off



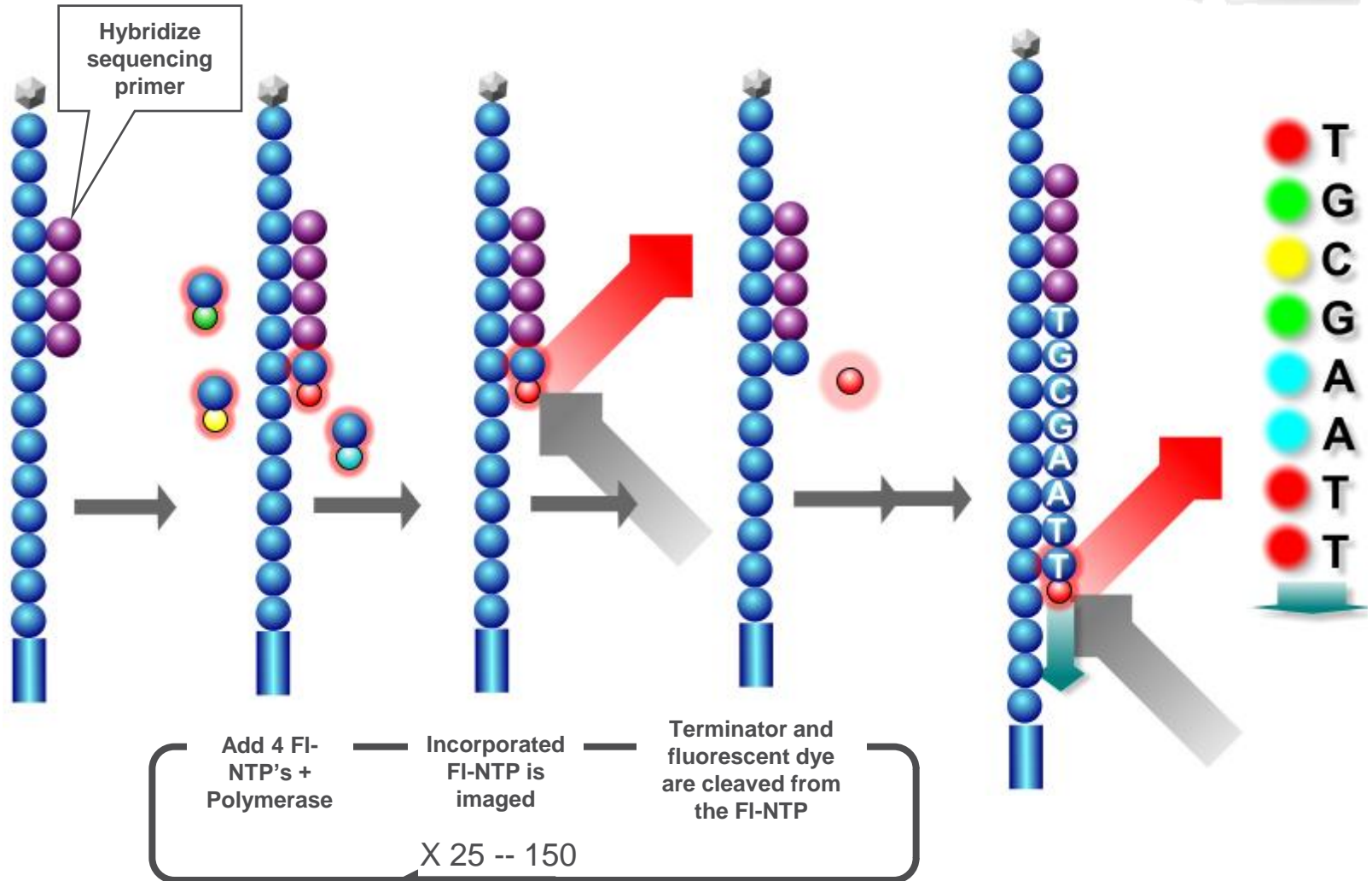
Paired End Sequencing



- Free 3' ends of the reverse template and lawn primers are blocked to prevent unwanted DNA priming

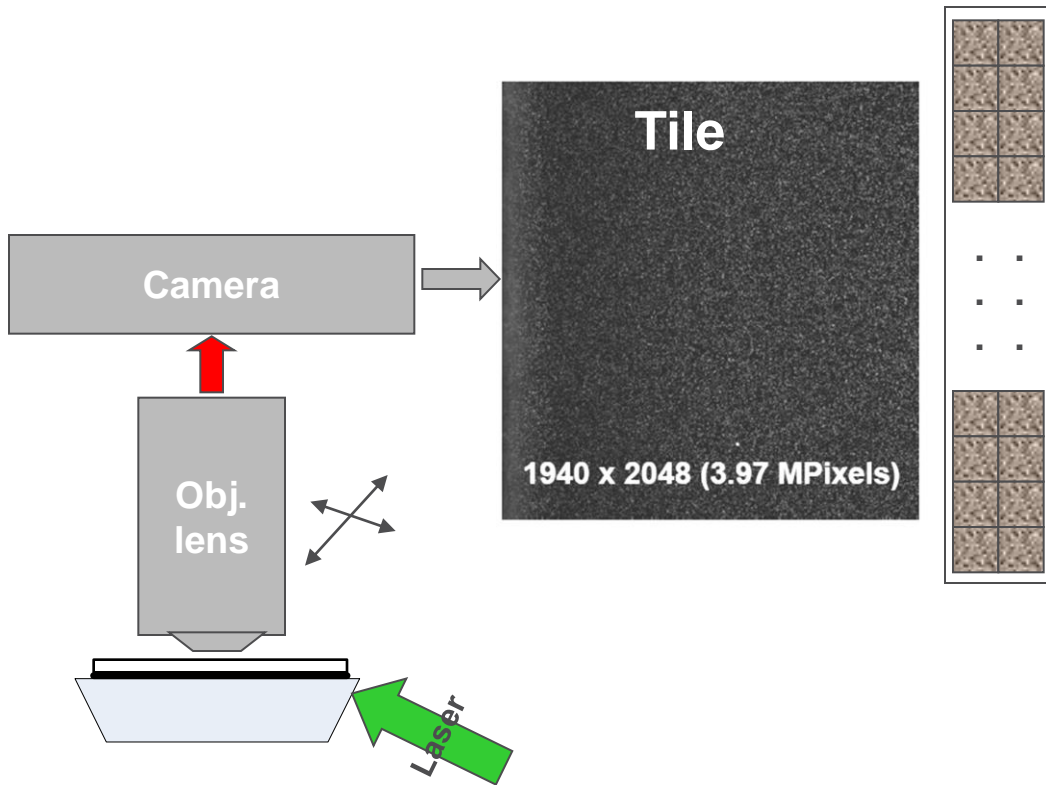


Sequencing Reverse Strand



Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

Genome Analyzer imaging



3-4.5 TB/run

640,000 images x 7 MB/image

75-100 x 2 bases

4 images/base

8 channels/flow cell

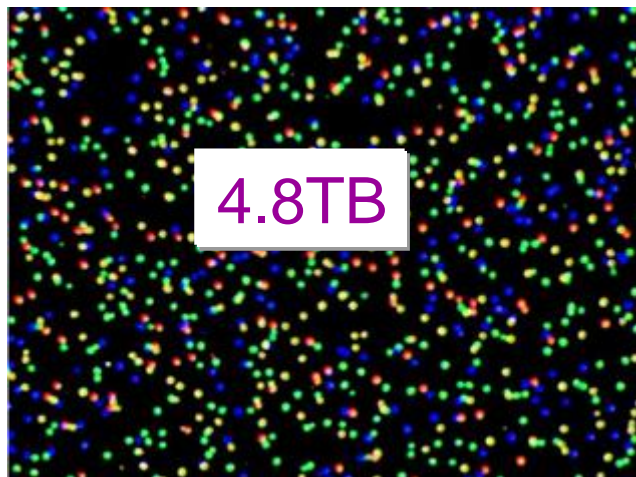
2 columns/channel

55 tiles/column

Data Analysis



images



intensities

Lane	Tile	X	Y	Cycle 1 - A C G T				Cycle 2 - A C G T			
5	12	924	1560	493.1	388.9	3626.7	2359.4	185.6	122.3	360.4	307.8
5	12	773	395	85.5	113.0	2327.5	1158.0	156.3	166.9	113.5	909.6
5	12	165	786	1243.8	741.1	45.8	67.4	318.4	692.6	48.3	41.7
5	12	598	690	250GB				3.6	505.7	1919.1	959.3
5	12	1107	1207					8.6	230.5	815.1	512.1
5	12	1074	466					38.4	41.8	64.9	1102.9
5	12	887	356	743.1	486.4	42.2	305.0	230.3	603.6	-63.1	-20.1
5	12	642	1769	63.2	54.3	861.7	595.7	81.5	86.0	54.9	385.4
5	12	599	314	845.5	533.2	45.2	581.0	260.9	560.9	13.0	78.4
5	12	839	1103	372.0	812.6	16.7	70.5	59.4	69.4	35.4	1394.9
5	12	347	1792	343.8	706.9	108.4	638.5	73.2	43.9	121.6	1882.2
5	12	807	1114	63.9	63.8	828.3	1369.0	1074.4	714.3	-39.9	29.4

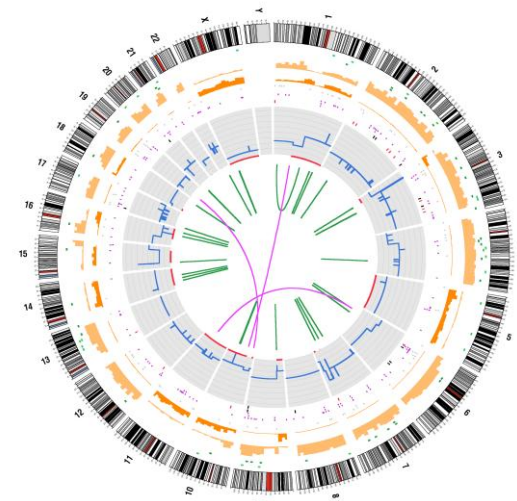
base calls

```
s_1_0007_seq.txt - WordPad
File Edit View Insert Format Help
GAAC AAGCATAT
TTTTTTTTTTTT
GATCATGTTTTC
CCTGCCTCAGCC
214 595 TACAAAATCCCTGCC
155 544 TTATCTGCATCCGGT
301 507 TCCCTGCTTATTGAC
175 606 TTGGAATCGGGGTTA
242 522 TAAC TAATATACAGG
196 522 TGTACACAGGAGGAA
237 612 TTGCTGCAAGCTCAG
160 528 TCTGATTTTTACACA
164 543 TCTCAGAGAAACGTG
```

Detecting clusters
 Measuring the color
 for each cluster
 ... for every cycle

Data Analysis

A simple, familiar workflow



**HiSeq CONTROL
SOFTWARE**

Base calls

CASAVA

**Alignments,
variations, builds**

VISUALIZATION

**GenomeStudio, or
favorite browser**

Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

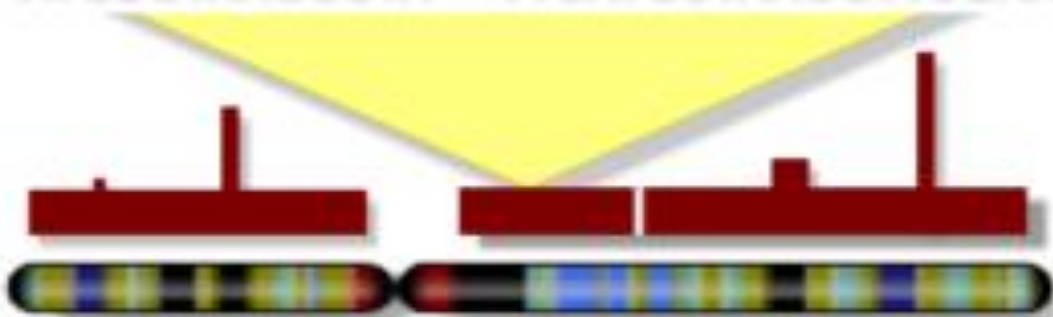
Alignment to reference

```
TGCGTAAGGCTAGGTTCATGCTAAGGTTTCGAA
A GCGTAAGGCTAGGTTCATGCTAAGGTTTCGAA
AT CGTAAGGCTAGGTTCATGCTAAGGTTTCGAA
ATG GTAAGGCTAGGTTCATGCTAAGGTTTCGAA
ATGC TAAGGCTAGGTTCATGCTAAGGTTTCGAA
ATGCG AAGGCTAGGTTCATGCTAAGGTTTCGAA
ATGCGT AGGCTAGGTTCATGCTAAGGTTTCGAA
ATGCGTA GCTAGGTTCATGCTAAGGTTTCGAA
ATGCGTAA CTAGGGTTCATGCTAAGGTTTCGAA
```

Reads

ATGCGTAAGGCTA - - TTCATGCTAAGGTTTCGAA

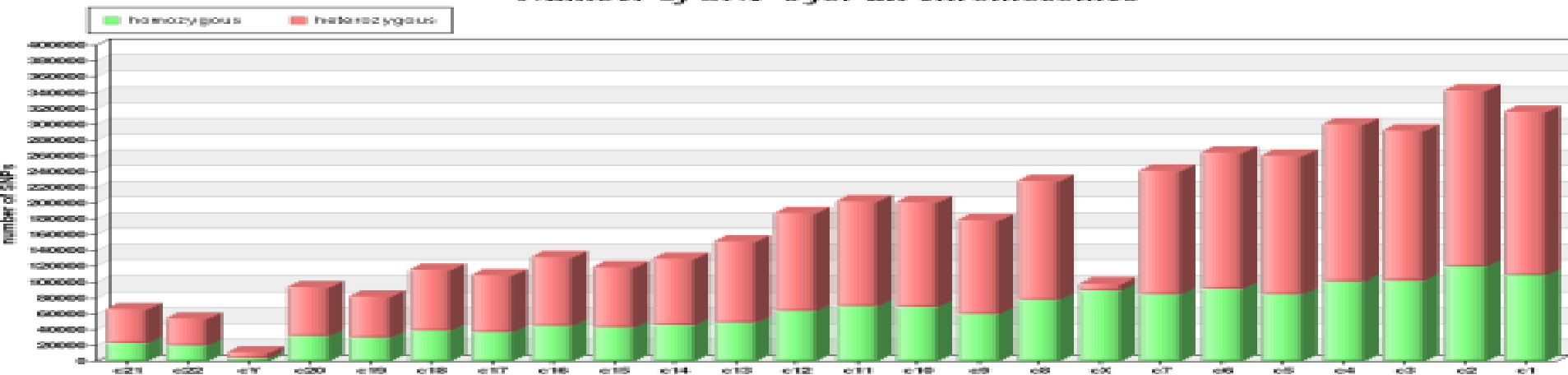
Aligned to reference



Read depth

CASAVA – software to determine variation

Number of SNP's for all chromosomes



	A	B	C	D	E	F	G	H	I	J	K	L
1	#position	A	C	G	T	modified_call	total	used	score		reference	type
2	18260646	0	34	0	0	C	36	34	112.79		T	SNP_diff
3	18261869	0	0	13	0	G	13	13	41.19		T	SNP_diff
4	18262422	0	0	9	26	TG	40	35	82.78:28.20		T	SNP_het1
5	18262476	12	0	0	37	TA	54	49	104.98:36.99		T	SNP_het1
6	18262564	0	31	0	16	CT	50	47	94.33:53.20		C	SNP_het1
7	18263563	0	0	36	0	G	42	36	128.39		C	SNP_diff
8	18264404	0	0	28	0	G	30	28	83.48		T	SNP_diff
9	18264677	30	0	0	0	A	31	30	96.79		C	SNP_diff

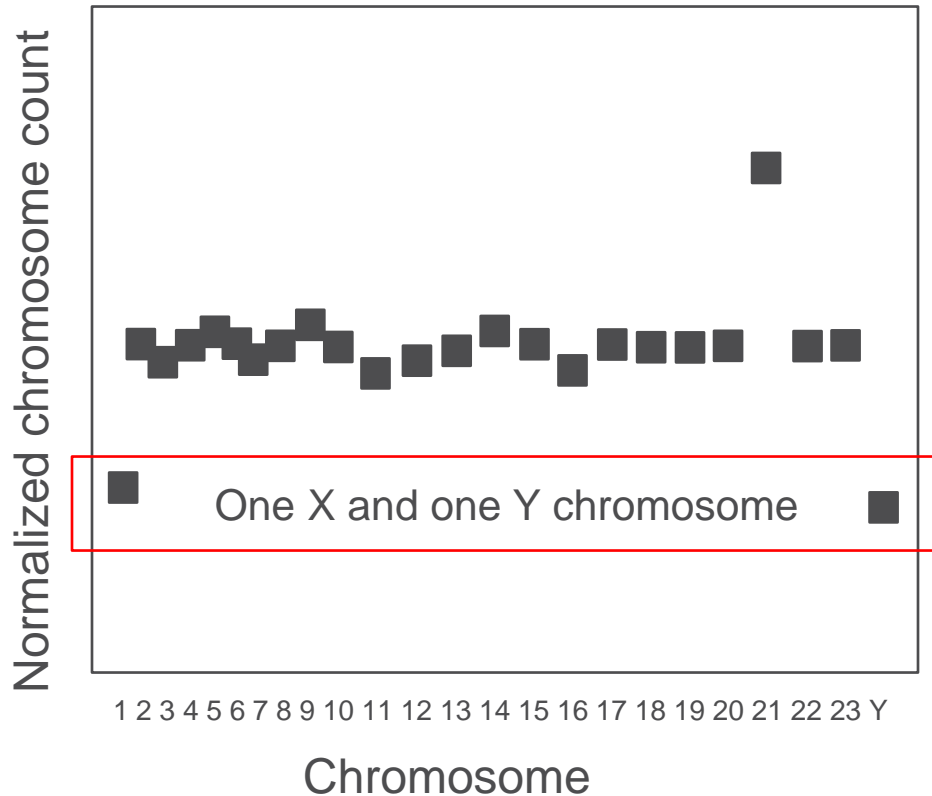
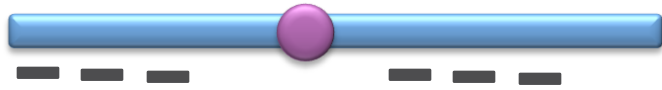
Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

Sequencing genomes on HiSeq 2000

	Target size	Sequencing depth	No. of samples/ FC	Cost/sample (Seq only, 2x100)	Cost/sample (Cluster + Seq, 2x100)
Aneuploidy	3 Gb	0.3 x	~ 100	€ 50	€ 85
CNV	3 Gb	1-3 x	~ 10-30	€ 170-500	€ 280-850
GWAS	3 Mb	30(-50) x	~ 1000	€ 5	€ 9
Exome	30 Mb	30(-50) x	~ 100	€ 50	€ 85
SNV discovery	3 Gb	6 x	~ 5	€ 1000	€ 850
SNV/SNP validation	3 Gb	30x	~ 1	€ 5000	€ 8500

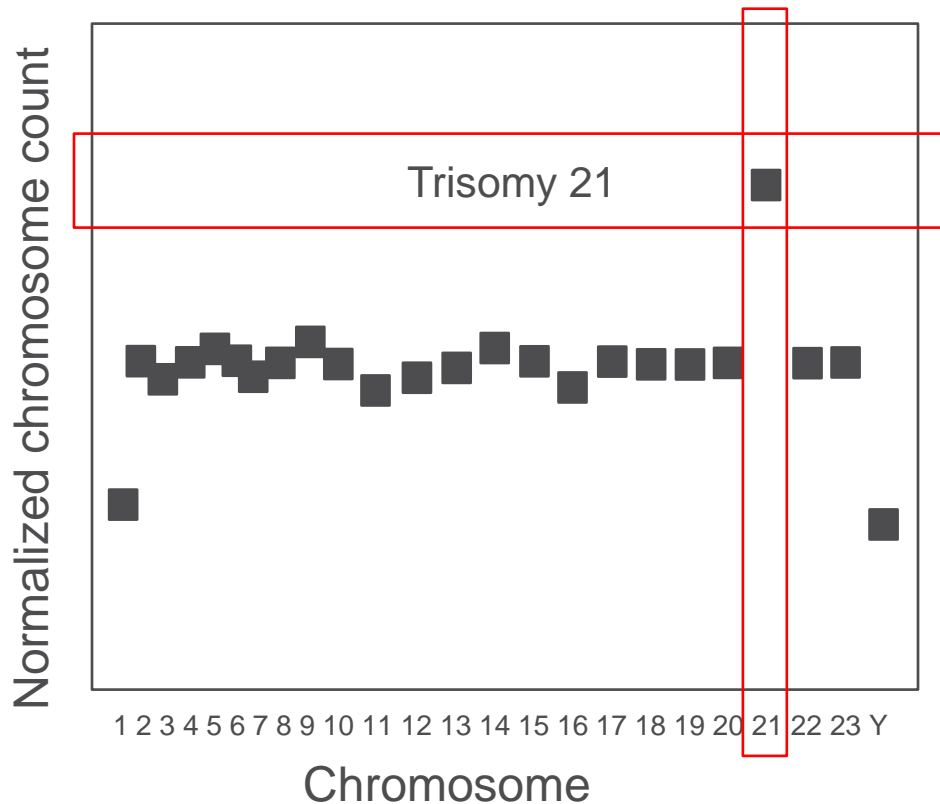
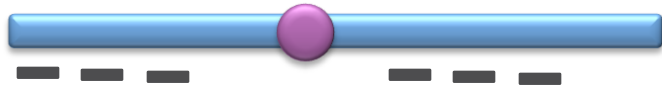
Excludes cost of sample prep.

Sequencing Hu genomes at 0.1-0.3x (€ 100/sample)



1. 75 bp reads at 0.1 x human genome coverage
2. Reads map at approx. 1 read every 1 kb
3. Add reads for each chromosome
4. Divide total reads with chromosome length
5. Determine chromosome count

Sequencing Hu genomes at 0.1-0.3x (€ 100/sample)



1. 75 bp reads at 0.1 x human genome coverage
2. Reads map at approx. 1 read every 1 kb
3. Add reads for each chromosome
4. Divide total reads with chromosome length
5. Determine chromosome count

Prenatal aneuploidy by low depth sequencing

PNAS

Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma

Rossa W. K. Chiu^{a,b}, K. C. Allen Chan^{a,b}, Yuan Gao^{c,d}, Virginia Y. M. Lau^{a,b}, Wenli Zheng^{a,b}, Tak Y. Leung^a, Chris H. F. Foo^e, Bin Xie^e, Nancy B. Y. Tsui^{a,b}, Fiona M. F. Lun^{a,b}, Benny C. Y. Zee^f, Tze K. Lau^a, Charles R. Cantor^{g,1}, and Y. M. Dennis Lo^{a,b,1}

^aCentre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, Departments of ^bChemical Pathology and ^cObstetrics and Gynaecology, and ^dCentre for Clinical Trials, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China; ^eCenter for the Study of Biological Complexity and ^fDepartment of Computer Science, Virginia Commonwealth University, Richmond, VA 23284; and ^gSequenom, Inc., San Diego, CA 92121

—“Sequencing is the clear way to do non-invasive prenatal testing. ... existing noninvasive Down syndrome tests are not very informative and provide variable results depending on the ethnicity of those taking the test.”



PNAS

Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood

H. Christina Fan^{*}, Yair J. Blumenfeld^d, Usha Chitkara^a, Louanne Hudgins^a, and Stephen R. Quake^{a,b}

^{*}Department of Bioengineering, Stanford University and Howard Hughes Medical Institute, 318 Campus Drive, Clark Center, Room E300, Stanford, CA 94305; ^dDivision of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Stanford University, 300 Pasteur Drive, Room HH333, Stanford, CA 94305; and ^fDivision of Medical Genetics, Department of Pediatrics, Stanford University, 300 Pasteur Drive, Stanford, CA 94305

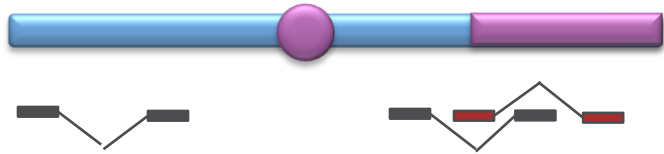
Non-invasive prenatal diagnosis by single molecule counting technologies

Rossa W.K. Chiu^{1,2}, Charles R. Cantor³ and Y.M. Dennis Lo^{1,2}



Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

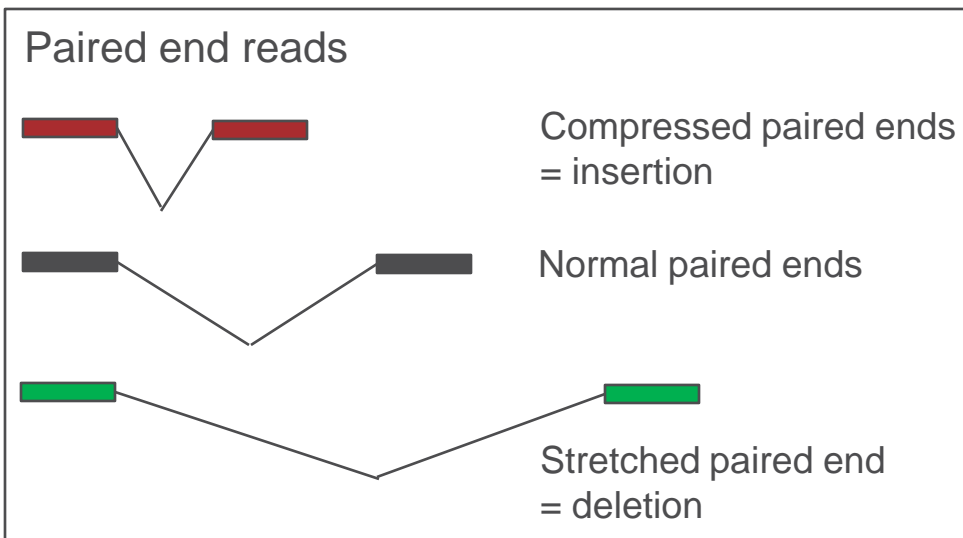
Sequencing whole genomes with paired-ends



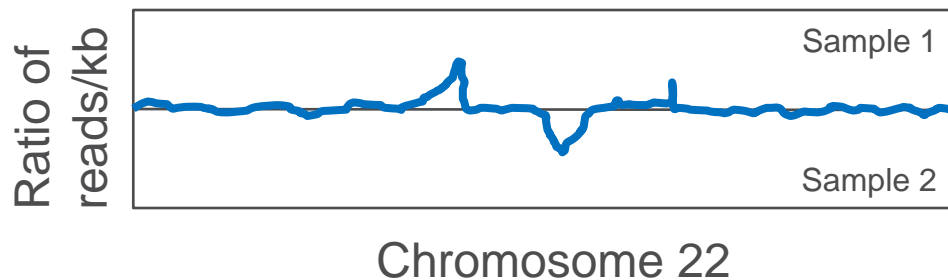
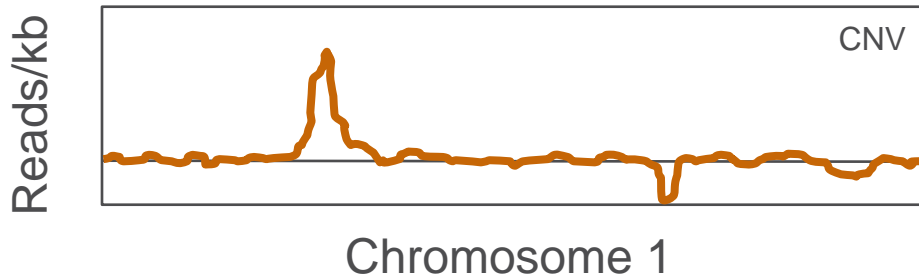
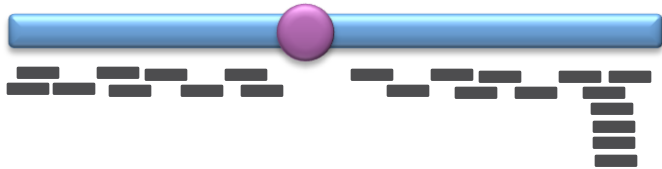
Short-insert or long-insert paired end reads provide more information on structural variation

GATCGGTTGCGATTCGG ATCGGTGGGACTGGG

↑
Read spanning a translocation



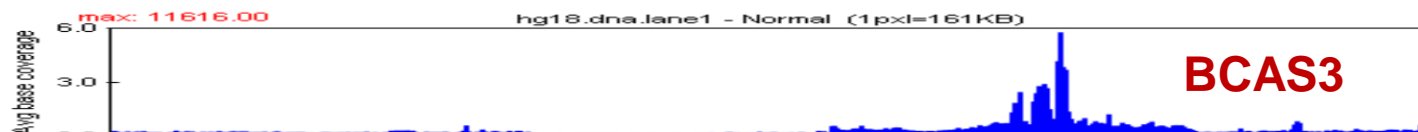
Sequencing Hu genomes at 1-3x (€ 850/sample)



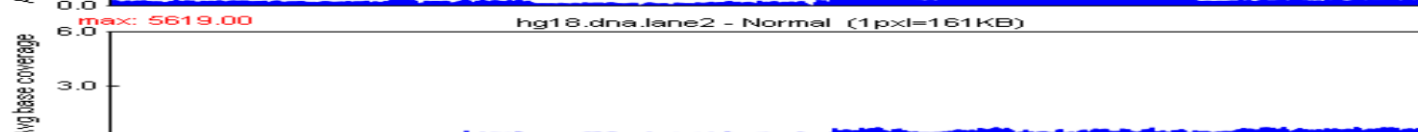
1. 75 bp reads at 1 x human genome coverage
2. Reads map at approx. 1 read every 100 bp
3. Average reads per 1 kb region
4. Ratio of avg reads for Sample 1 and 2
5. Plot average read ratio across chromosomes
6. Determine copy number variations

CNVs in cancer cell lines with 1x sequencing depth

MCF7



ZR-75-1



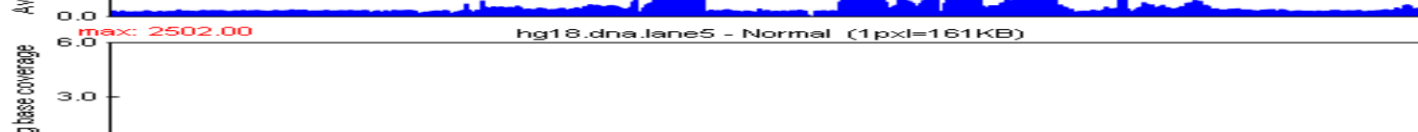
T47D



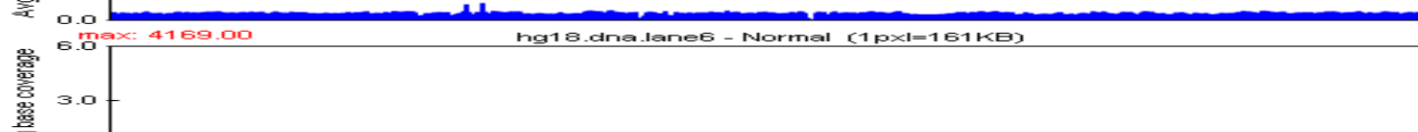
BT474



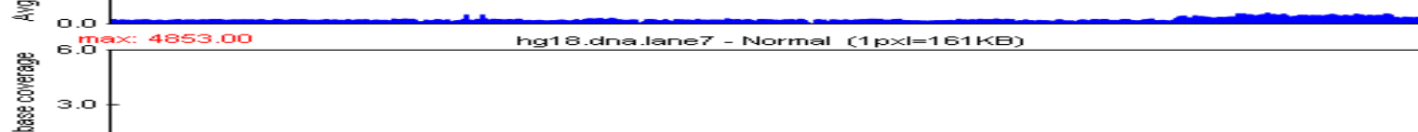
MDA-MB-231



MDA-MB-468



BT20

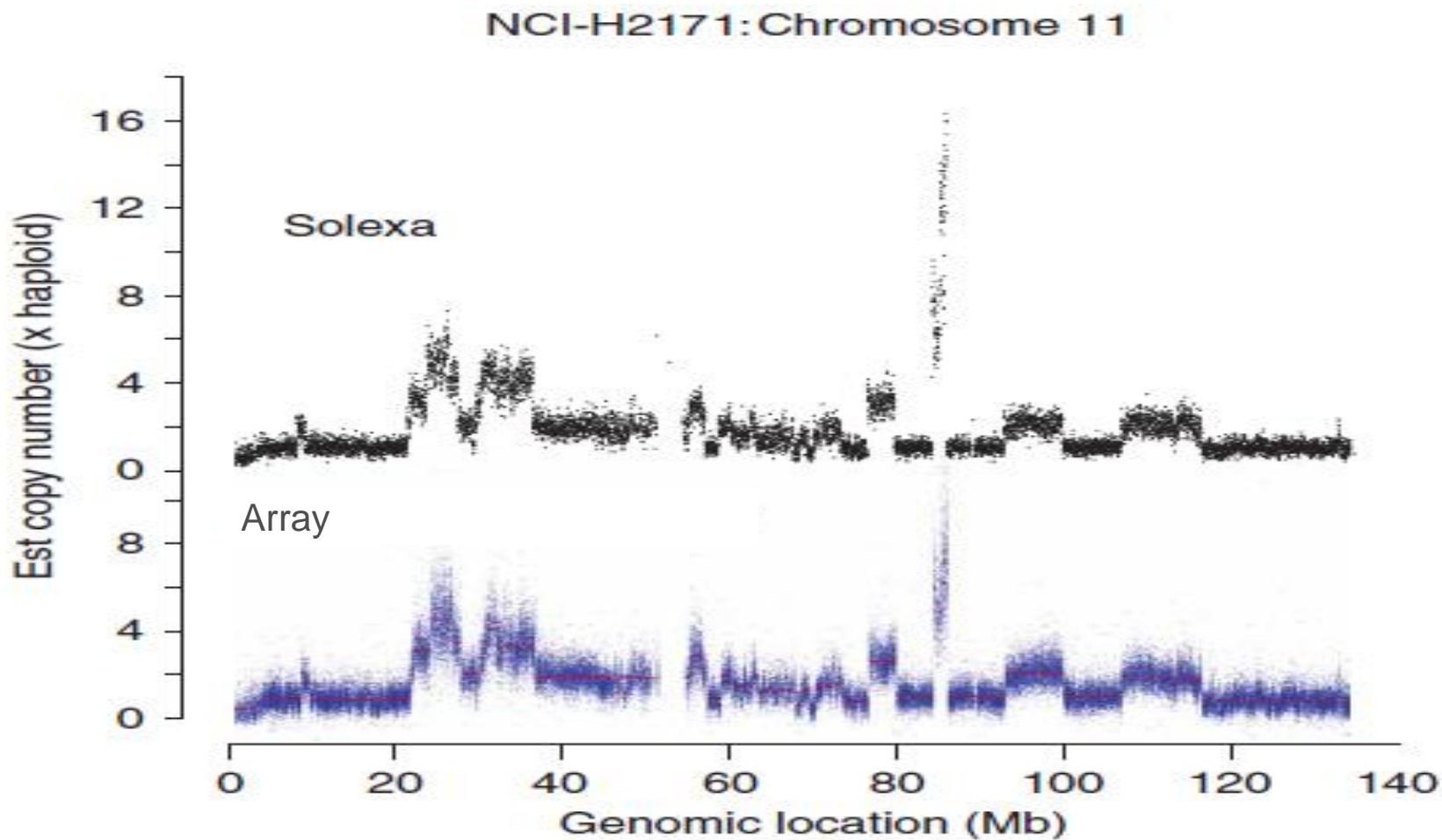


MCF10A



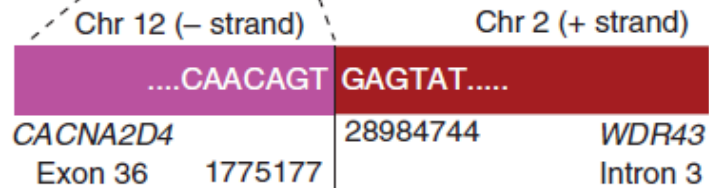
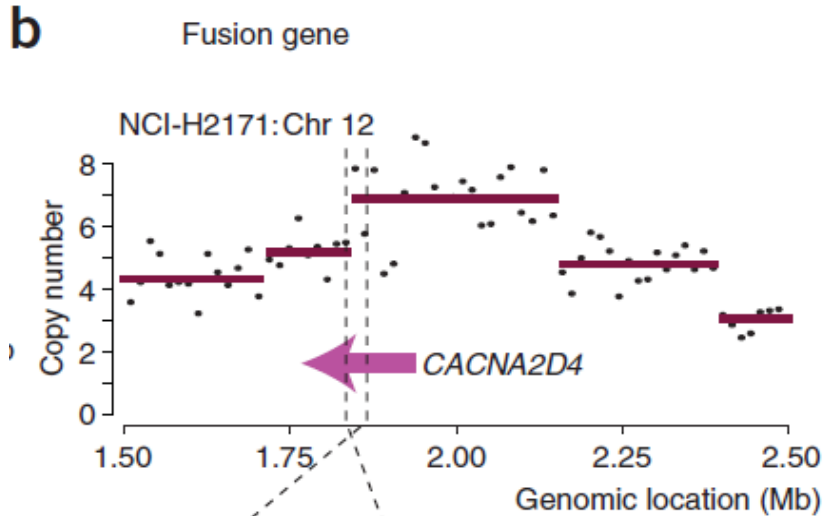
Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

CNV by medium depth sequencing



Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

CNV and fusion gene

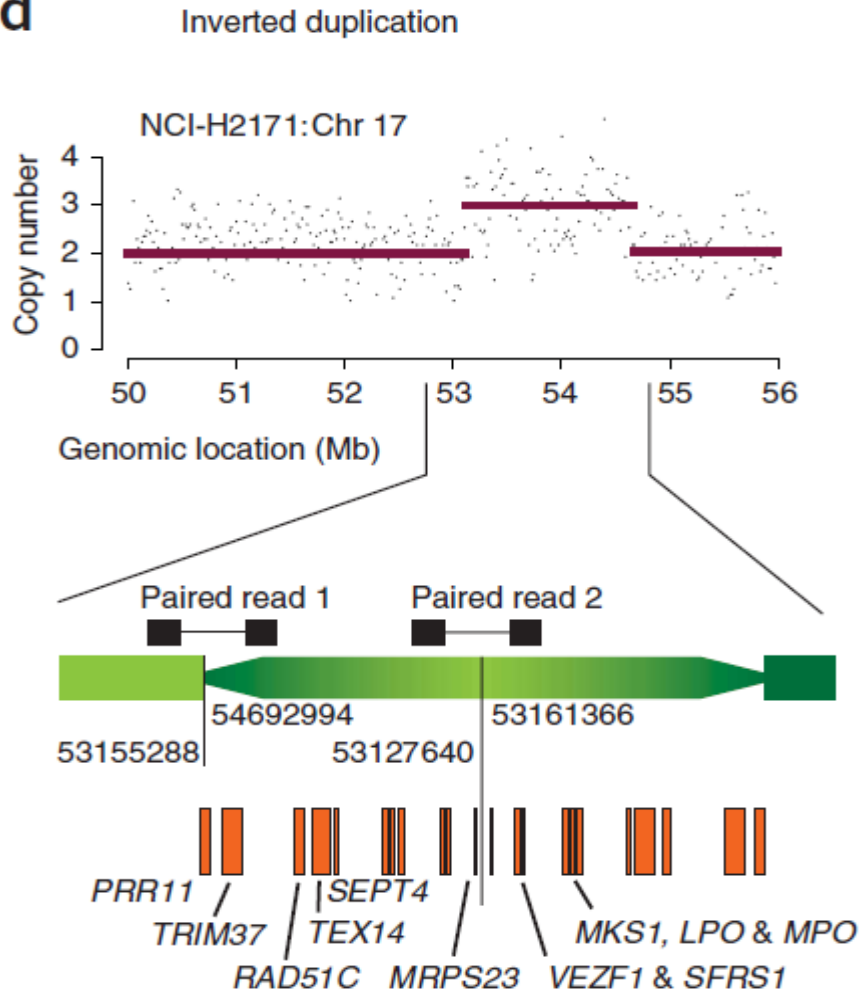


- 5' of CACNA2D4 is amplified
- Paired-end reads show break in exon 36 of CACNA2D4 fusing into intron 3 of WDR43
- Resulting in a fusion transcript with a shortened exon 36 from CACNA2D4.

Campbell 2008

CNV and inversion

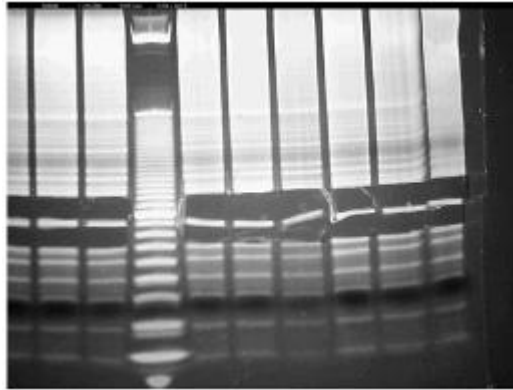
d



- An inverted duplication in chromosome 17 by localized increase in copy number.
- Two paired-end reads spanned both inverted breakpoints.

Campbell 2008

SNPs by pooled genome sequencing



DNA digestion, fractionation, and size selection



SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries

Curtis P Van Tassel¹, Timothy P L Smith², Lakshmi K Matukumalli^{1,3}, Jeremy F Taylor⁴, Robert D Schnabel⁴, Cynthia Taylor Lawley⁵, Christian D Haudenschild², Stephen S Moore⁶, Wesley C Warren⁷ & Tad S Sonstegard¹



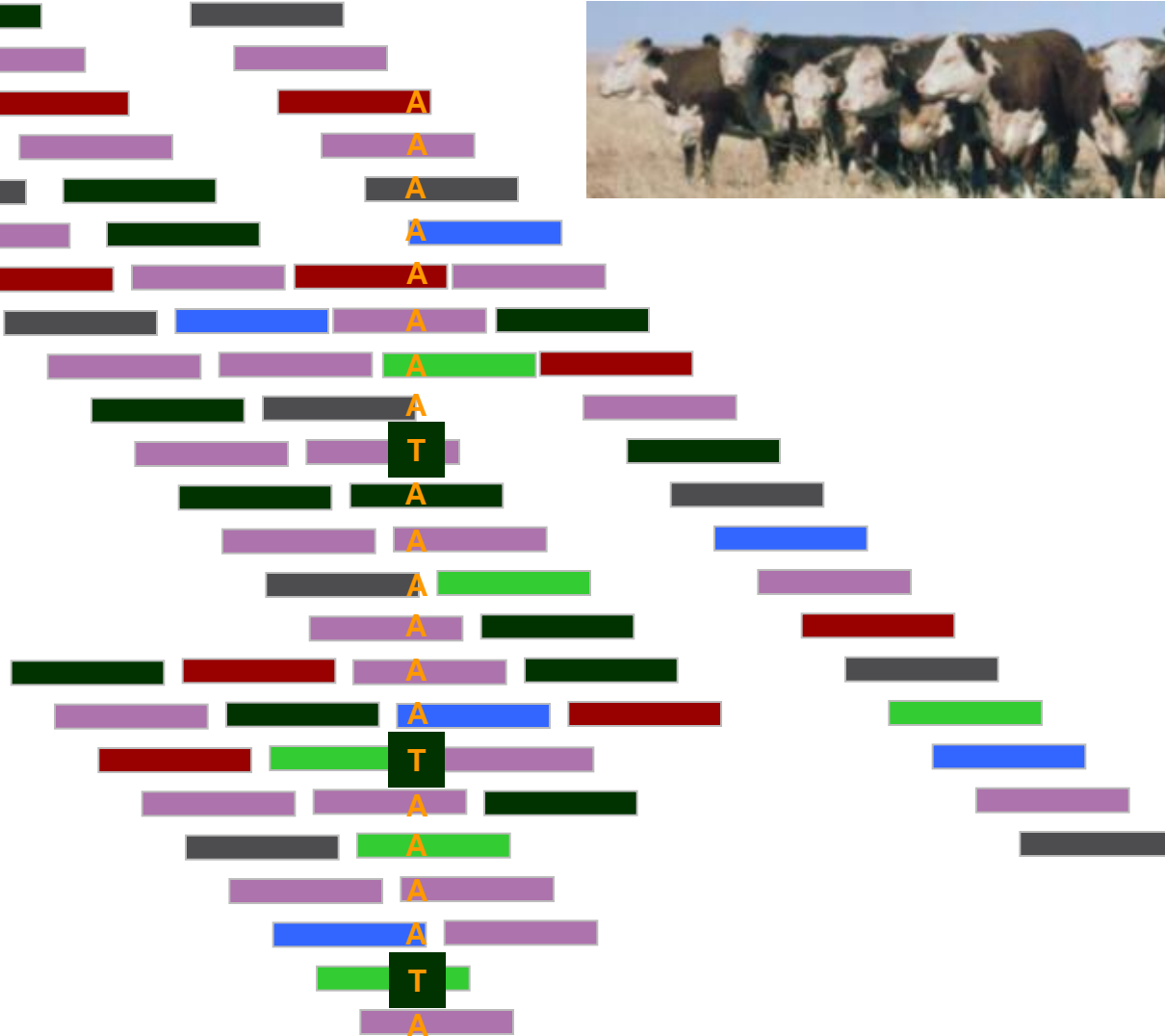
- Bovine SNP m
- Following sto
- High-content
- A yeast genom
- Resolving RNA

Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

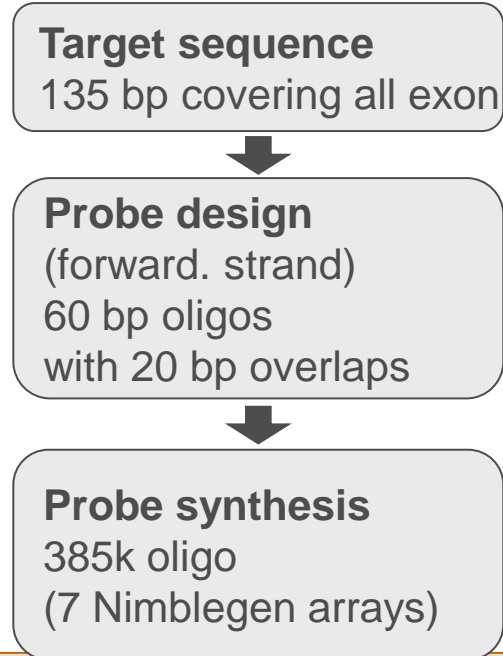
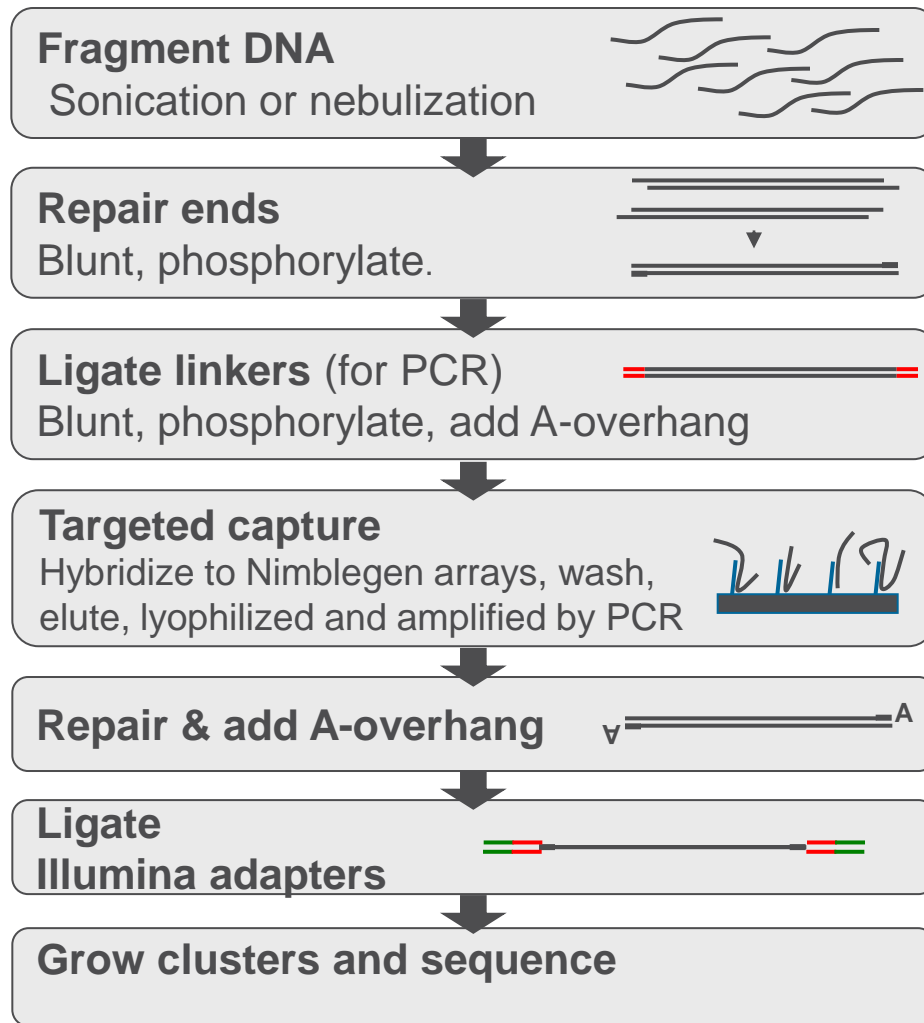
SNP Frequencies from Reduced Representation Libraries



Each colour represents a read from a different genome. Base frequencies will indicate SNPs.



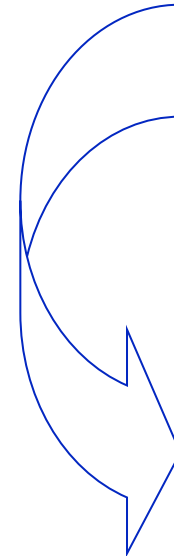
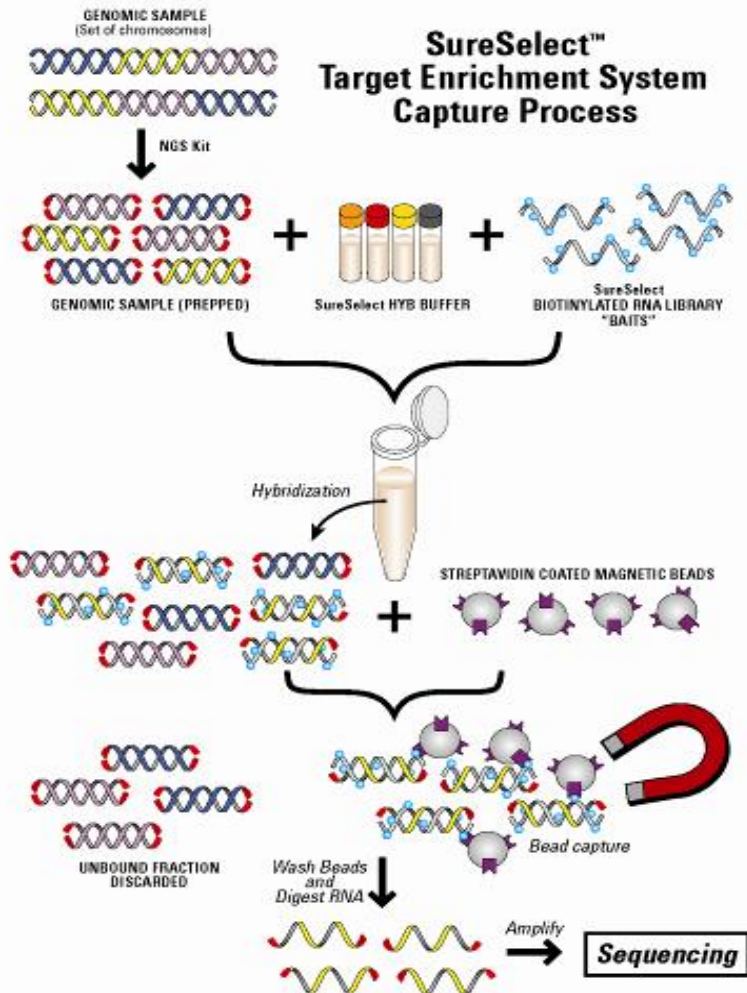
Targeted sequencing by Solid Phase Capture of all Exons



Sequencing a fraction of the genome



Agilent Technologies SureSelect™ Target Enrichment System



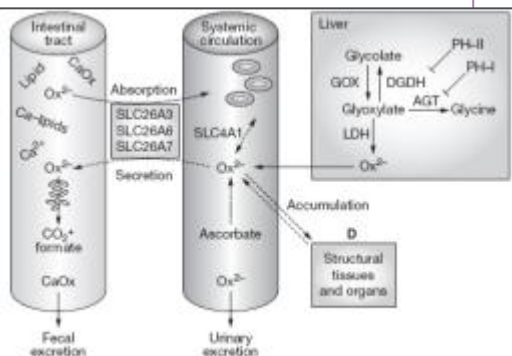
Sequence only

- Exome
- Cardiac genes
- Diabetes genes

Exome capture in diagnosis

● Unanticipated genetic diagnosis

- congenital chloride diarrhea with a suspected diagnosis of Bartter syndrome, a renal salt-wasting disease.
- Homozygous in *SLC26A3* (known congenital chloride diarrhea locus).
- 5 additional patients suspected to have Bartter syndrome had mutations in *SLC26A3*.



A

Reference	P L N I E V P K I S L H S L I L D F S A V S F L D V S S V R G L K
GIT 264-1	P L N I E V P K I S L H S L I L N F S A V S F L D V S S V R G L K
Sense	5'-CCTCTCAACATTGAGGTCCCAAAATCAGCCCTCCACAGCCTCATTCTCGACTTTTCAGCAGTGTCTTCTTTGATGTTTCTTCAGTGAGGGGCCCTTAAA-3'
Antisense	3'-GGAGAGTTGTAACCTCCAGGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGCTGAAAAGTCGTCACAGGAAAGAACTACAAAAGAGTCACTCCCGGAATTT-5'

3'-GGAGCGTTGTAACCTCCAGGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGTT-5'

3'-AACTCCAGGGTTTTTCGTCGGAGGGGTCGGAGTAAGAGTTGAAAA-5'

5'-ctccaggggttttagtcggaggtgctcggagtaagagttgaaaagtcgtca-3'

3'-CCAGGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCACA-5'

5'-gggggttttagtcggaggtgctcggagtaagagttgaaaagtcgtcacagga-3'

3'-TTTTGGTGGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCACAGGAAAG-5'

3'-TTTGTAGTCGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAA-5'

3'-GTCGGAGCGTCCGAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAACTAC-5'

5'-cggaggtgctcggagtaagagttgaaaagtcgtcacaggaagaactacaa-3'

3'-GGGGGGTCCGAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAACTACAAA-5'

5'-gaggtgctcggagtaagagttgaaaagtcgtcacaggaagaactacaaag-3'

3'-GGGTCGGAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAACTACAAAAGAG-5'

5'-tcggagtaagagttgaaaagtcgtcacaggaagaactacaaagaagtc-3'

3'-GAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAACTACAAAAGAGTCACTC-5'

5'-agagttgaaaagtcgtcacaggaagaactacaaagaagtcactccccgg-3'

3'-GTTGAAAAGTCGTCACAGGAAAGAACTACAAAAGAGTCACTCCCGGAAT-5'

Targeted sequencing

● Familial breast cancer

- TP53, BRCA1, and BRCA2 mutations in established tumour cell lines and DNA from patients with germline mutations. All of the known pathogenic mutations were identified ... clonal sequencing outperforms current diagnostic methods.

● Resistant tumors

- Mutations in MEK1, novel mechanisms of resistance, important clinical implications

● Cancer-related exome subset

● Joubert syndrome 2

- Neurological, psychomotor retardation. Mutation in the TMEM216 gene. Hetero- non-symptomatic.

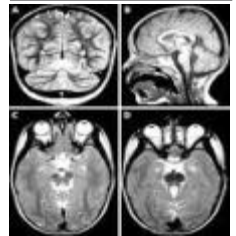
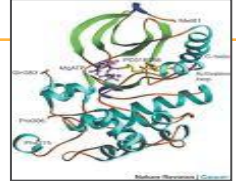
● Hereditary poikiloderma

- Homozygous A>C mismatch in intron 4 of C16orf57 gene. (unknown function)

● Freeman-Sheldon syndrome

- Autosomal dominant

● Neanderthal genome





De novo assemblies of microbes, BACs

- Mutations with enhanced resistance to killing by chicken heterophils, reflecting avian host adaptive evolution.

Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*

Bethan V. Lowder¹, Caltriona M. Guinane², Nour L. Ben Zakour³, Lucy A. Weinstock⁴, Andrew Conway-Morris⁵, Robyn A. Cartwright², A. John Simpson², Andrew Rambaut², Ulrich Nübel², and J. Ross Fitzgerald¹

¹The Roslin Institute and Centre for Infectious Diseases, Royal (Dick) School of Veterinary Studies, Institute of Evolutionary Biology, and Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh EH8 9JY, Scotland, United Kingdom; and ²Robert Koch Institut, 38855 Wernigerode, Germany

Edited by Richard P. Novick, New York University School of Medicine, New York, NY, and approved September 18, 2009
(Received for review August 14, 2009)

The impact of globalization on the emergence and spread of pathogens is an important veterinary and public health issue. *Staphylococcus aureus* is a notorious human pathogen associated with serious nosocomial and community-acquired infections. In addition, *S. aureus* is a major cause of animal diseases including skeletal infections of poultry, which are a large economic burden on the global broiler chicken industry. Here, we provide evidence that the majority of *S. aureus* isolates from broiler chickens are the descendants of a single human-to-poultry host jump that occurred approximately 38 years ago (range, 30 to 63 years ago) by a subtype of the worldwide human ST5 clonal lineage unique to Poland. In contrast to human subtypes of the ST5 radiation, which demonstrate strong geographic clustering, the poultry ST5 clade was distributed in different continents, consistent with wide dissemination via the global poultry industry distribution network. The poultry ST5 clade has undergone genetic diversification from its human progenitor strain by acquisition of novel mobile genetic elements from an avian-specific accessory gene pool, and by the inactivation of several proteins important for human disease pathogenesis. These genetic events have resulted in enhanced resistance to killing by chicken heterophils, reflecting avian host-adaptive evolution. Taken together, we have determined the evolutionary history of a major new animal pathogen that has undergone rapid avian host adaptation and international dis-

semination. The reasons for the emergence and subsequent increase in incidence of BCO among chickens are unknown.

Results and Discussion

The Majority of *S. aureus* Isolates from Poultry Belong to a Single Clonal Complex (CCS) Usually Associated with Humans. To examine the population genetics of *S. aureus* strains infecting farmed birds, we carried out multi-locus sequence typing (MLST) of 57 *S. aureus* isolates, including 48 isolates from healthy and diseased poultry, in 8 countries on 4 continents isolated in the past 54 years. In addition to 9 isolates from different species of reared game and wild birds [supporting information (SI) Table S1]. Remarkably, the majority of all avian isolates ($n = 35$; 61%) including 32 (67%) from broiler chickens, belonged to a single sequence type (ST), ST5, or its single locus variants (ST1342, ST1346, and ST1350), including isolates from all countries examined except Australia (Fig. 1). The clonal complex CCS (which includes related haplotypes differing at a small number of loci) is one of the most successful human-associated lineages of *S. aureus*, characterized by its global distribution and frequent emergence of methicillin-resistant strains (9). Of the non-CCS isolates identified, which included isolates from broiler chickens, reared bird species such as pheasant and partridge, and a wild buzzard, 11 belonged to an unrelated clonal complex, CC385,

emerging and dissemination of pathogens. Shifts in agricultural practice result in opportunities for pathogens to expand into new host species and to spread rapidly to new territories. For example, the epidemics of bovine spongiform encephalitis (1) and the foot and mouth disease epidemic (2) were caused by changing agricultural practices providing new opportunities for transmission, including the use of meat and bone meal in cattle feed, and the long-distance transport of livestock, respectively.

The broiler poultry industry has been transformed within the last 50 years from a market dominated by smallholder chicken farms to a multibillion dollar industry controlled by a handful of multinational companies who supply a limited number of breeding lines to a global market (3, 4). Infectious diseases of chicken flocks are a major economic burden on the industry. In particular, *Staphylococcus aureus* is associated with several infections of poultry including septic arthritis, subdermal abscesses (i.e., “bumble foot”), and gangrenous dermatitis (5). In the 1970s, a new form of *S. aureus* infection of broiler poultry known as bacterial chondronecrosis with osteomyelitis (BCO) was described (6). Since then, BCO has increased in frequency to become a leading cause of lameness in the broiler chicken

all being related or identical to ST5 previously identified among human *S. aureus* isolates (Fig. 1A and B).

The Poultry ST5 Clade Is the Result of a Single, Recent Human-To-Poultry Host Jump. Recently, a high-resolution analysis of the phylogenetic structure of the human ST5 clonal radiation was carried out by mutation discovery at 108 loci (46 kb), resulting in the identification of at least 14 distinct lineages within the ST5

Author contributions: B.V.L., A.J.S., U.N., and J.R.F. designed research; B.V.L., C.M.G., N.L.B.Z., L.A.W., A.C.M., R.A.C., A.R., and U.N. performed research; N.L.B.Z. contributed new reagents/analytic tools; B.V.L., C.M.G., N.L.B.Z., L.A.W., A.C.M., A.J.S., A.R., U.N., and J.R.F. analyzed data; and B.V.L. and J.R.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. CP801381, CP801382, CP801383, and CP801384) (2008 genome sequence, chromosome and plasmids, respectively) and AC200000000 (M81) Whole Genome Shotgun Project.

To whom correspondence should be addressed. E-mail: ross.fitzgerald@ed.ac.uk. This article contains supporting information online at www.pnas.org/cgi/content/full/060208106DCSupplemental.

Sequencing whole genomes

Accurate whole human genome sequencing using reversible terminator chemistry

A list of authors and their affiliations appears at the end of the paper

The diploid genome sequence of an Asian individual

Jun Wang^{1,2,3,4*}, Wei Wang^{1,2*}, Ruiqiang Li^{1,2,3,4*}, Yingzi Li^{1,2,3,4*}, Geng Tian^{1,2}, Laurie Goodman¹, Wei Fan¹, Junqing Zhang¹, Jun Li¹, Juanbin Zhang¹, Yiran Guo¹, Binxiao Feng¹, Heng Li^{1,5}, Yao Lu¹, Xiaodong Fang¹, Haiqing Liang¹, Zhenglin Du¹, Dong Li¹, Yiqing Zhao¹, Yujie Hu¹, Zhenzhen Yang¹, Hancheng Zheng¹, Ines Hellmann¹, Michael Izura¹, John Peo¹, Xin Yi^{1,2}, Jing Zhao¹, Jinjie Duan¹, Yan Zhou¹, Junjie Qin¹, Lisa Ma^{1,2}

DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome

Thomas J. Schumacher^{1,2,3,4}, Li Ding^{1,2,3,4}, Bob Fulton^{1,2,3,4}, Michael D. McLellan^{1,2,3,4}, Ken Chen^{1,2,3,4}, David Dooling^{1,2,3,4}, McGrath^{1,2,3,4}, Matthew Hickenbotham^{1,2,3,4}, Lisa Cook^{1,2,3,4}, Rachel Abbott^{1,2,3,4}, David E. Larson^{1,2,3,4}, Scott Smith^{1,2,3,4}, Amy Hawkins^{1,2,3,4}, Scott Abbott^{1,2,3,4}, Devin Locke^{1,2,3,4}, LaDeana W. Hillier^{1,2,3,4}, Vincent Magrini^{1,2,3,4}, Todd Wylie^{1,2,3,4}, Jarret Glasscock^{1,2,3,4}, Joshua Conyers^{1,2,3,4}



Craig Venter

\$20M+
Capillary electrophoresis

2006

Watson

\$2M
454

2007

3+ genomes

\$200K/genome
GA

2008

Personal genome

\$48K/genome
GAIIx

2009

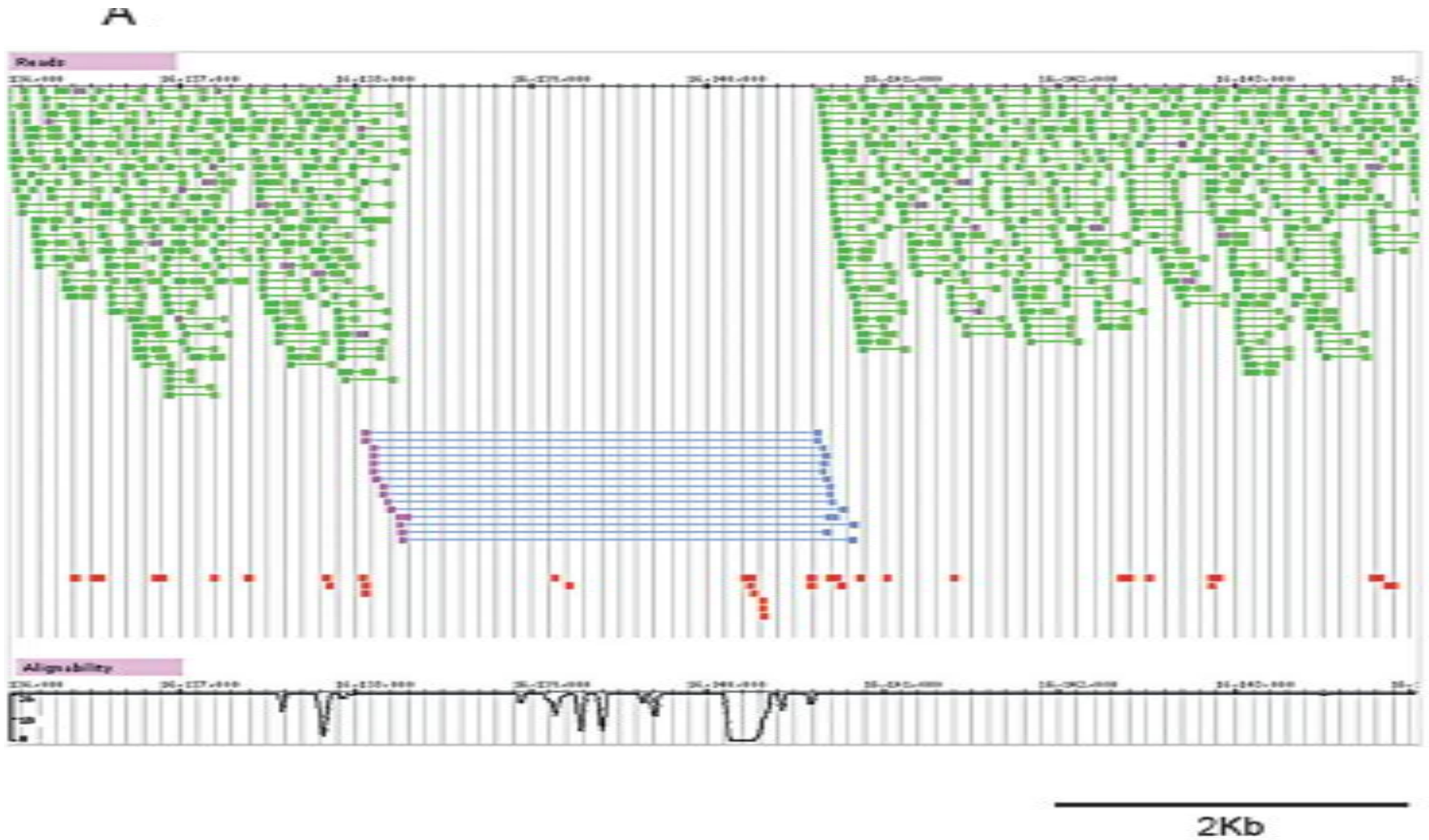
Human genome

\$10K/genome
HiSeq2000

2010

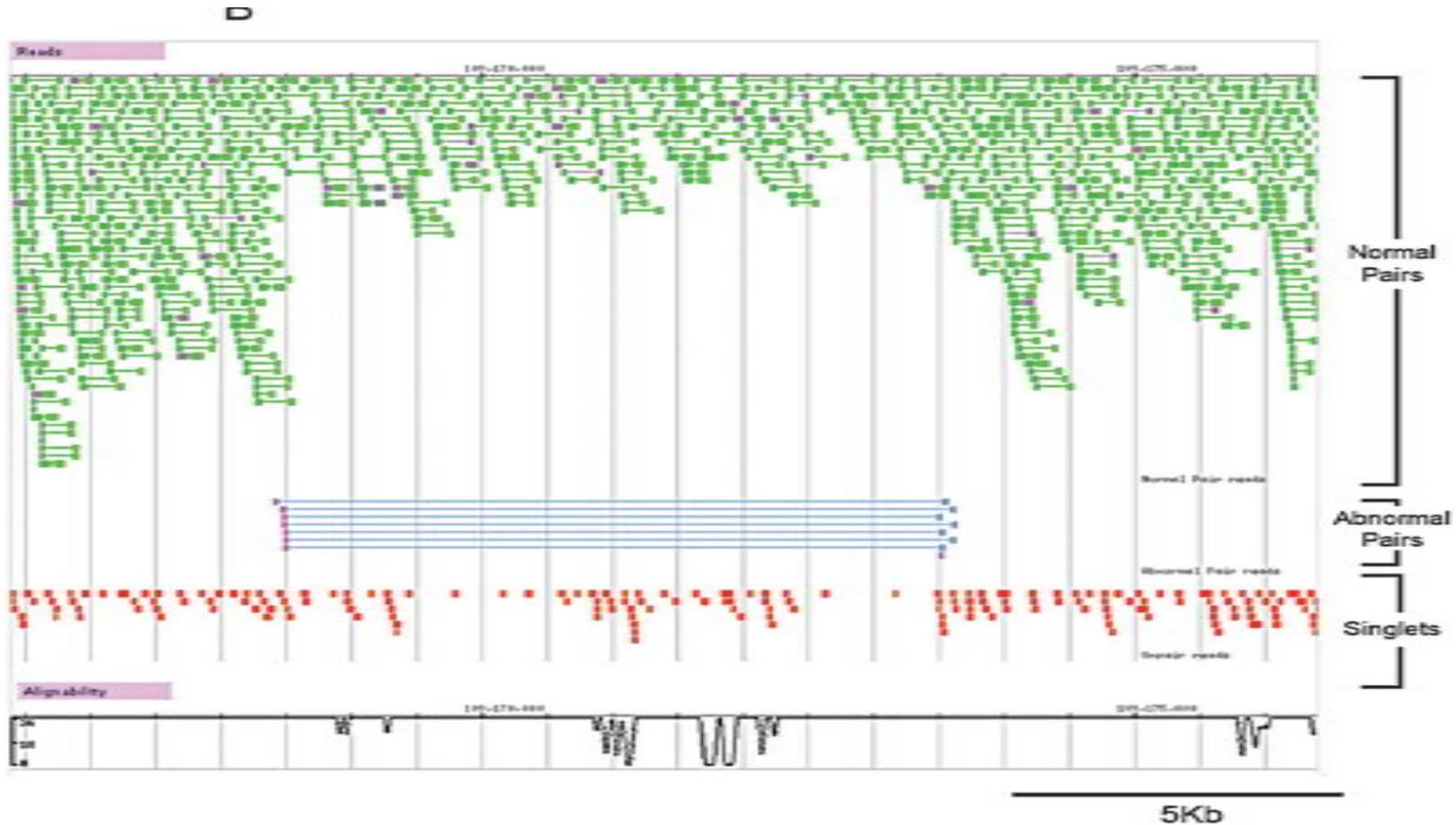
Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

Homozygous deletion by paired-end sequencing



Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

Heterozygous deletion by paired-end sequencing

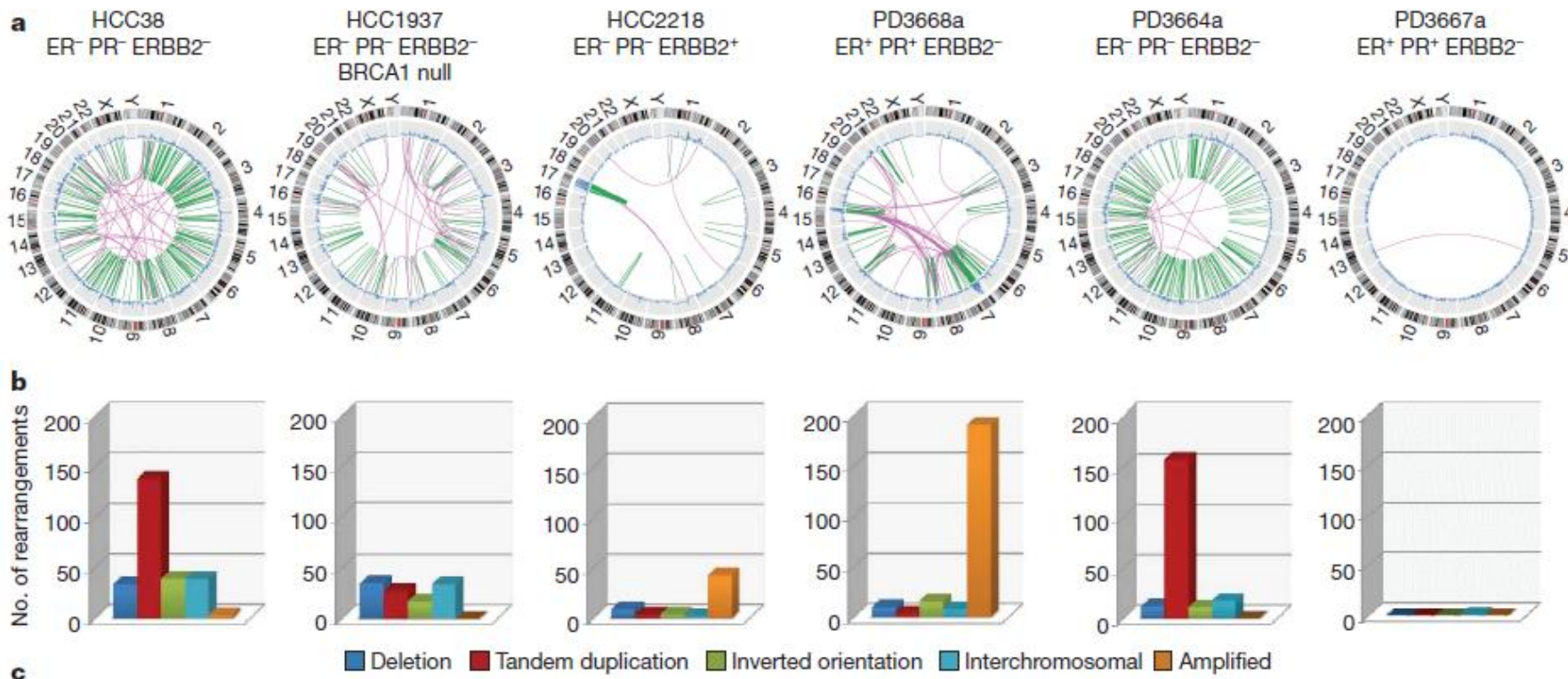


Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes



Complex landscapes of somatic rearrangement in human breast cancer genomes

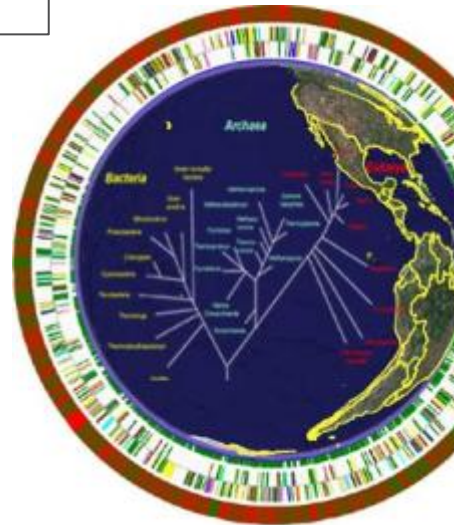
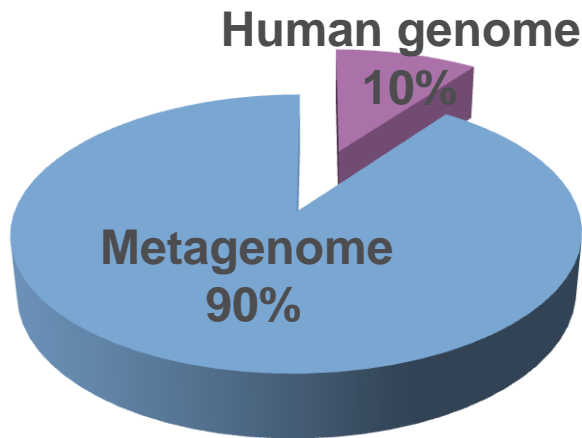
Philip J. Stephens¹, David J. McBride¹, Meng-Lay Lin¹, Ignacio Varela¹, Erin D. Pleasance¹, Jared T. Simpson¹, Lucy A. Stubbins¹, Catherine Lopus¹, Sarah Edkins¹, Laura J. Mudra¹, Chris D. Greenman¹, Mingming Li¹



Metagenomics

A human gut microbial gene catalogue established by metagenomic sequencing

Metagenomic study of the oral microbiota by Illumina high-throughput sequencing.



Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes

Summary

- The decrease in cost of sequencing has revolutionized genomics
- Aneuploidy at higher sensitivity and lower cost than any existing technology
- Copy number variations without any prior assumptions, with higher resolution and sensitivity and lower cost than CGH arrays
- Discovery of SNVs, indels, structural variation in either a fraction of the genome (by targeted sequencing) or in the whole genome at surprisingly low cost (from \$1,000 to \$10,000/sample)

Thank you



Clinical sequencing / Seq process / Seq methods / Costs / Seq at low depth / Medm depth / Targeted / Whole genomes / Metagenomes