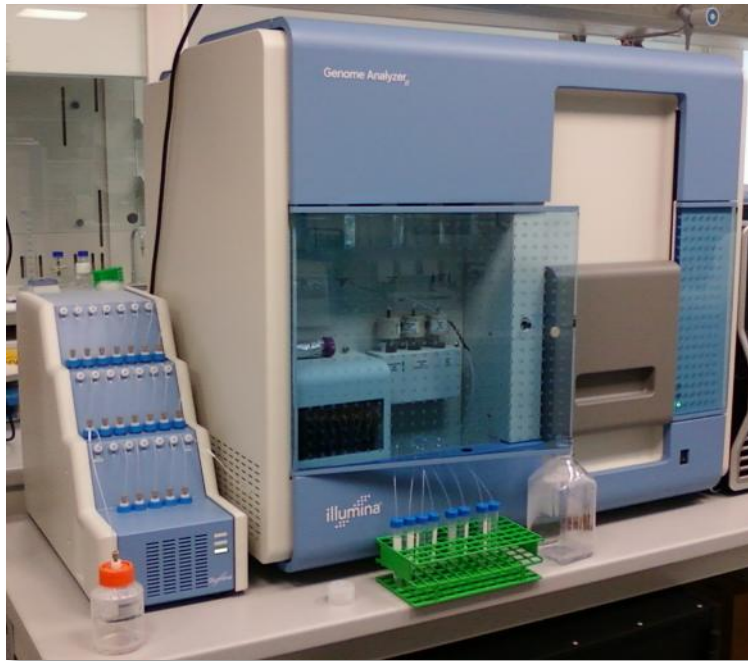


# Illumina Genome Analyzer IIx for high throughput sequencing



**Cosentino Cristian, *PhD***

Genomics and Bioinformatics unit  
Filarete Foundation – Milan (IT)  
*cosentia@gmail.com*

# Summary

1

Towards NGS sequencing

2

NGS with Illumina GAIIx

Genome Analyzer IIx

3

Data management

4

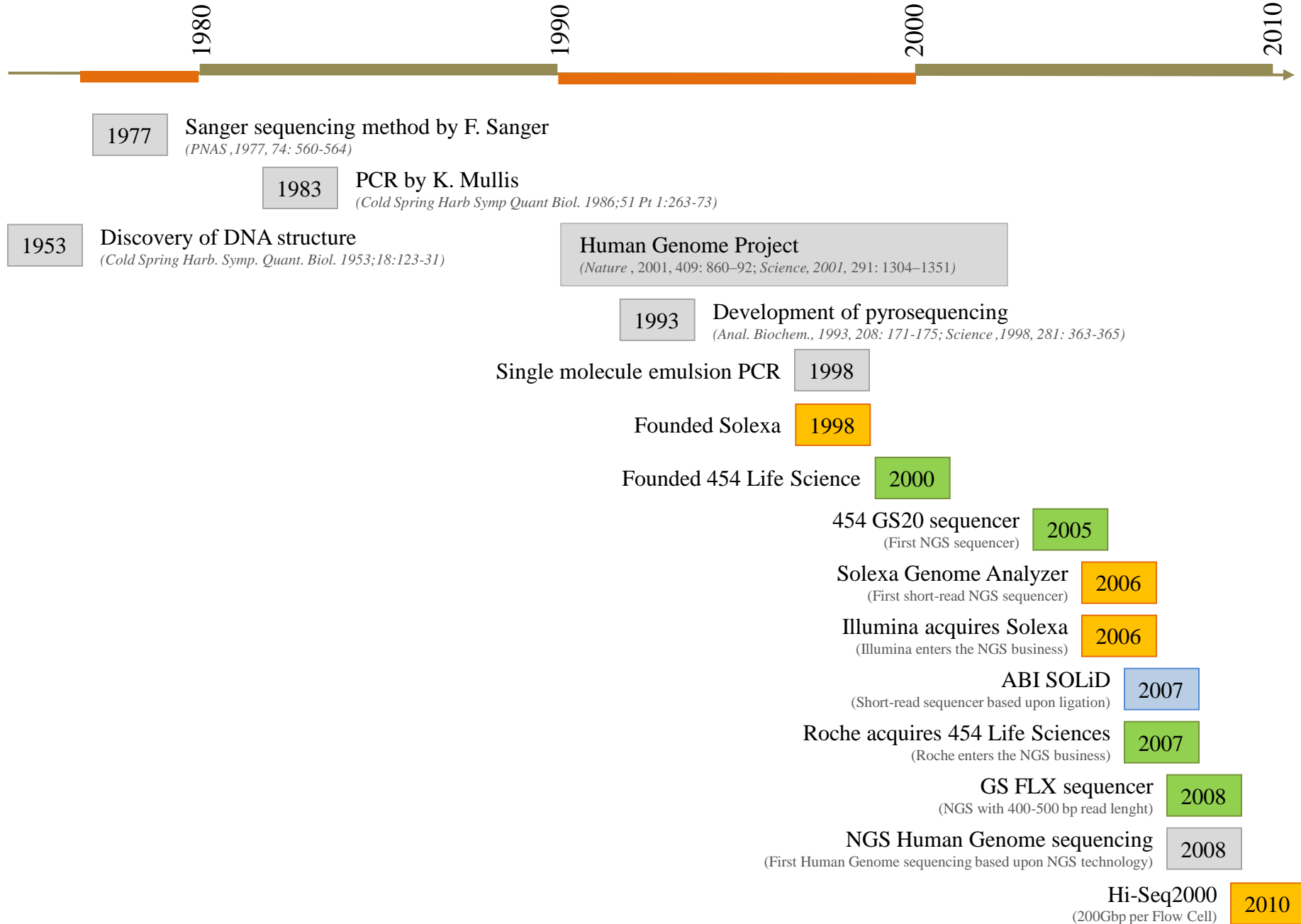
Target enrichment

# Summary

- 1 Towards NGS sequencing
- 2 NGS with Illumina GAIIx
- 3 Data management
- 4 Target enrichment

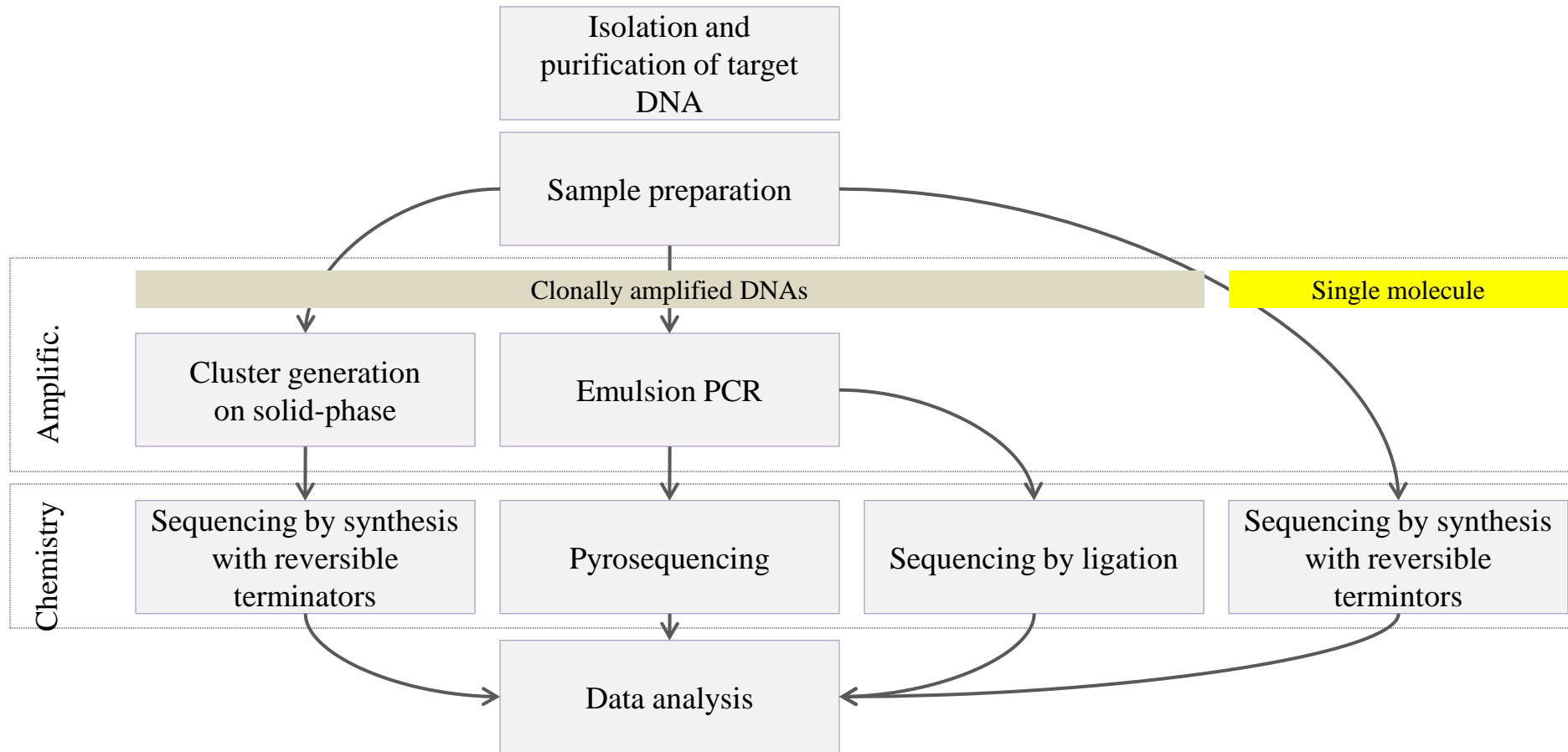
# Approaching to NGS

1



# Next-generation sequencing platforms

1



Illumina GAII



Roche 454



ABI SOLiD



Helicos HeliScope

# Summary

- 1 Towards NGS sequencing
- 2 NGS with Illumina GAIIx
- 3 Data management
- 4 Target enrichment

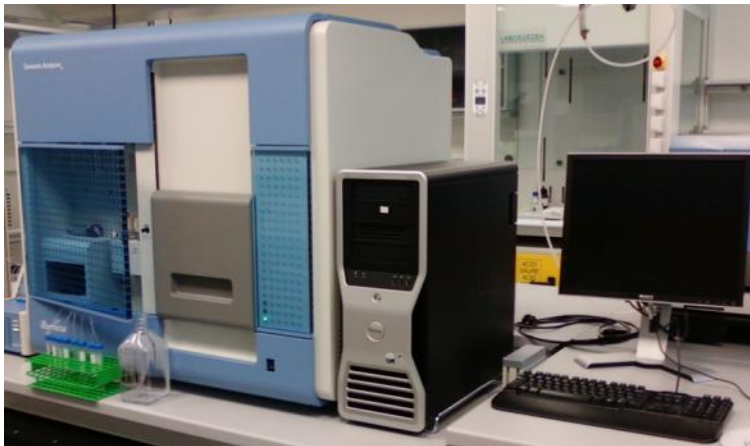
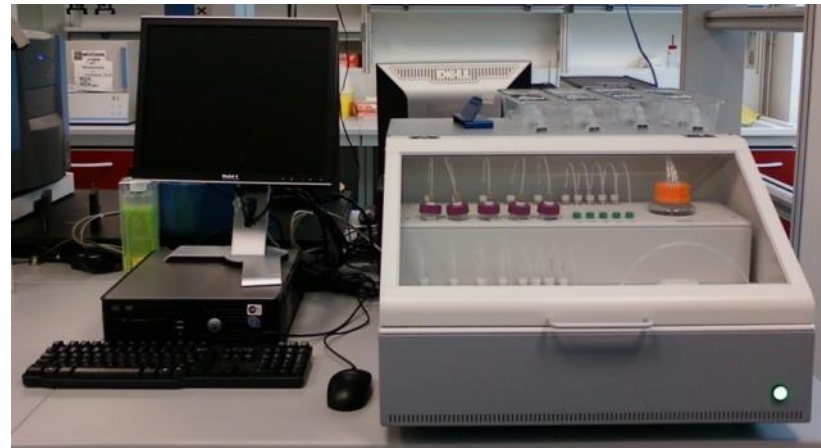
Bioanalyzer 2100



Flow Cell



Cluster station



Genome Analyzer IIx



Paired-end module



Linux server

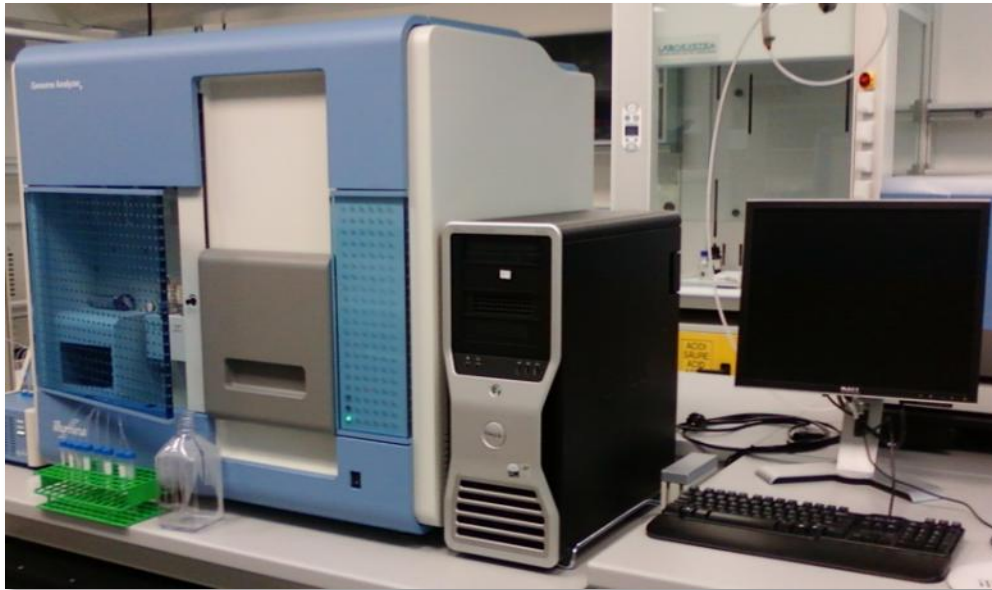
*de novo* sequencing (whole-genome)

*re*-sequencing (whole-genome or targeted)

RNA-seq

smallRNA-seq

CHiP-seq



Single-read

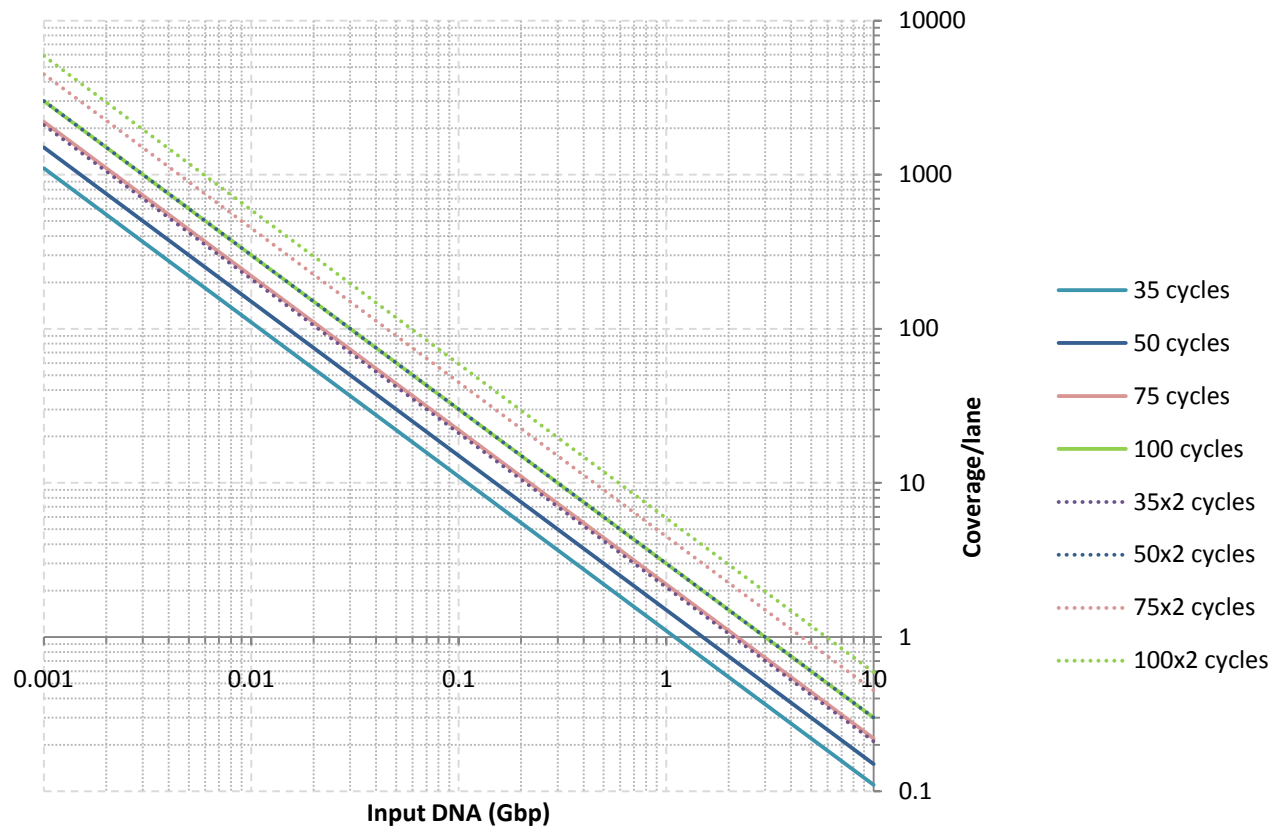
Paired-end

Multiplexing



Parameter	Performance
Amplification	Bridge-PCR on solid-phase
Chemistry	SBS with reversible terminators
Cost	2 \$/Mbp

Advantages	Disadvantages
<ul style="list-style-type: none"><li>•Most widely used platform (&gt; 90 science/nature publication)</li><li>•Sample preparation automatable</li><li>•SBS, real-time analysis and base calling are performed simultaneously to the run</li><li>•Automated cluster generation procedure</li></ul>	<ul style="list-style-type: none"><li>•Low multiplexing capability</li><li>•Substitution errors</li></ul>

Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

READ LENGTH	RUN TIME (DAYS)	CLUSTERS PASSING FILTER	OUTPUT (GB)	THROUGHPUT (GB/DAY)	BASE CALLS WITH Q $\geq$ 30	RAW READ ACCURACY	% PERFECT READS
1 $\times$ 35 bp	~2	225–250 million	8.0–9.0	~4.0–4.5	75–90%	$\geq$ 99%	$\geq$ 90%
2 $\times$ 35 bp	~4	225–250 million	16.0–18.0	~4.0–4.5	75–90%	$\geq$ 99%	$\geq$ 90%
2 $\times$ 50 bp	~5	225–250 million	22.5–25.0	~4.5–5.0	75–90%	$\geq$ 99%	$\geq$ 85%
2 $\times$ 75 bp	~7.5	225–250 million	34.0–38.0	~4.5–5.0	70–85%	$\geq$ 98.5%	$\geq$ 80%
2 $\times$ 100 bp	~9.5	225–250 million	45.0–50.0	~4.75–5.25	$\geq$ 70%	$\geq$ 98%	$\geq$ 70%

**SAMPLES**

**Throughput:** eight channels per flow cell, up to 12 samples per channel using Illumina Multiplexing Reagents

**Input requirement:** 0.1–1.0  $\mu$ g (single- and paired-end reads), 10  $\mu$ g (Mate Pair reads)

**Genomic DNA sample prep:** Three hours hands-on, six hours total for single or paired-end libraries



Sample  
preparation

Workbench

Clusters  
amplification

Cluster Station

Sequencing by  
synthesis

Genome Analyzer

Analysis pipeline

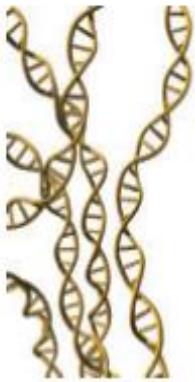
Linux Server

Sample  
preparation

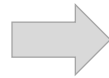
Clusters  
amplification

Sequencing by  
synthesis

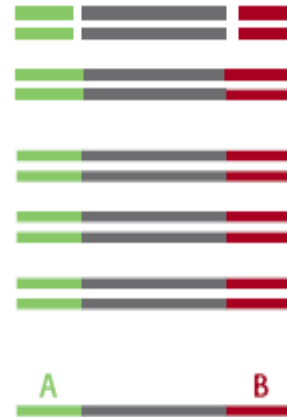
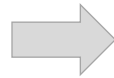
Analysis  
pipeline



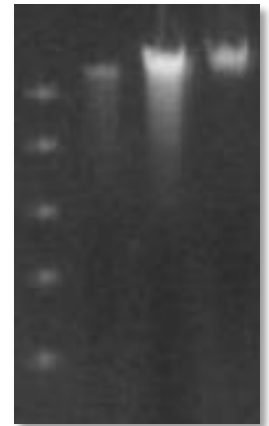
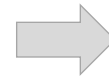
gDNA



Fragmented  
DNA



Adaptor-  
ligated DNA



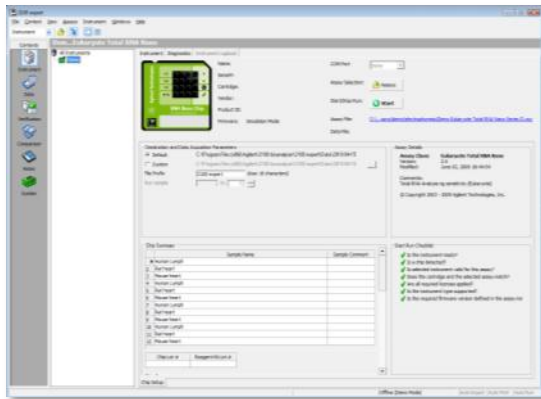
Gel purification

Sample  
preparation

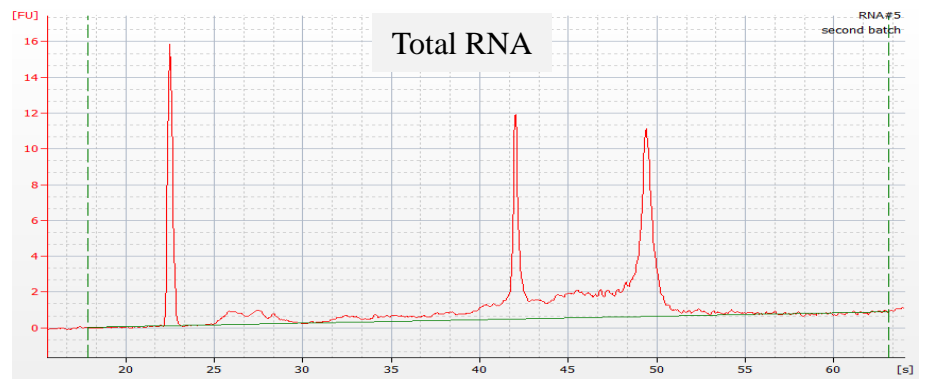
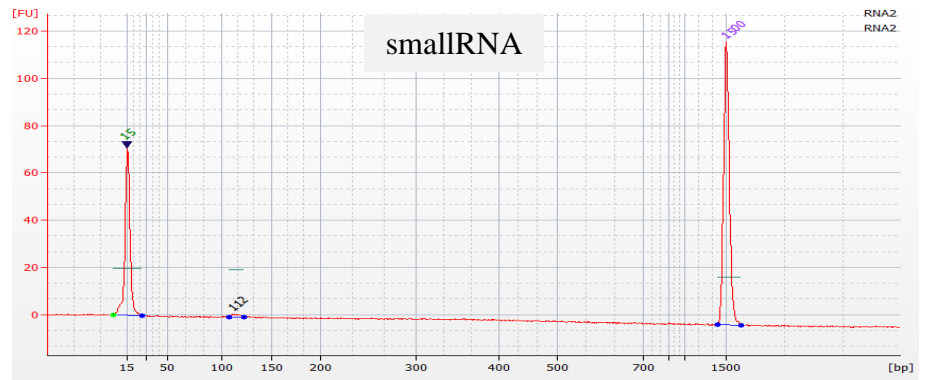
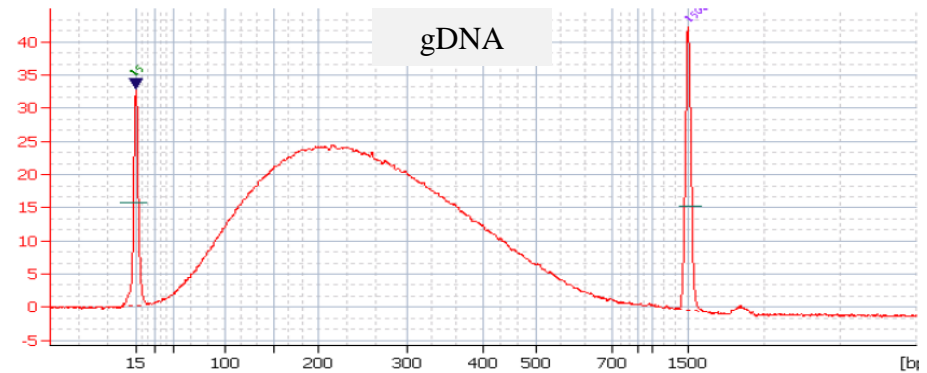
Clusters  
amplification

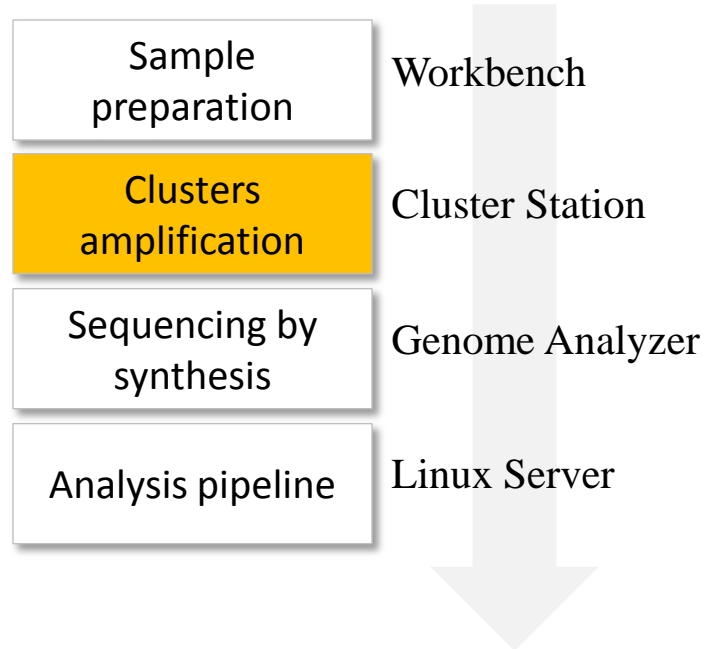
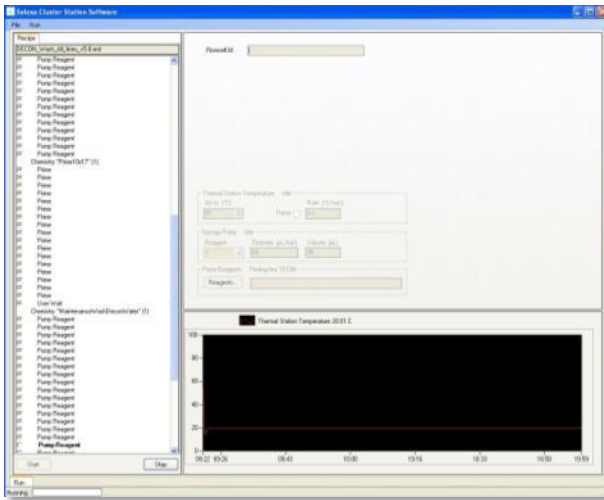
Sequencing by  
synthesis

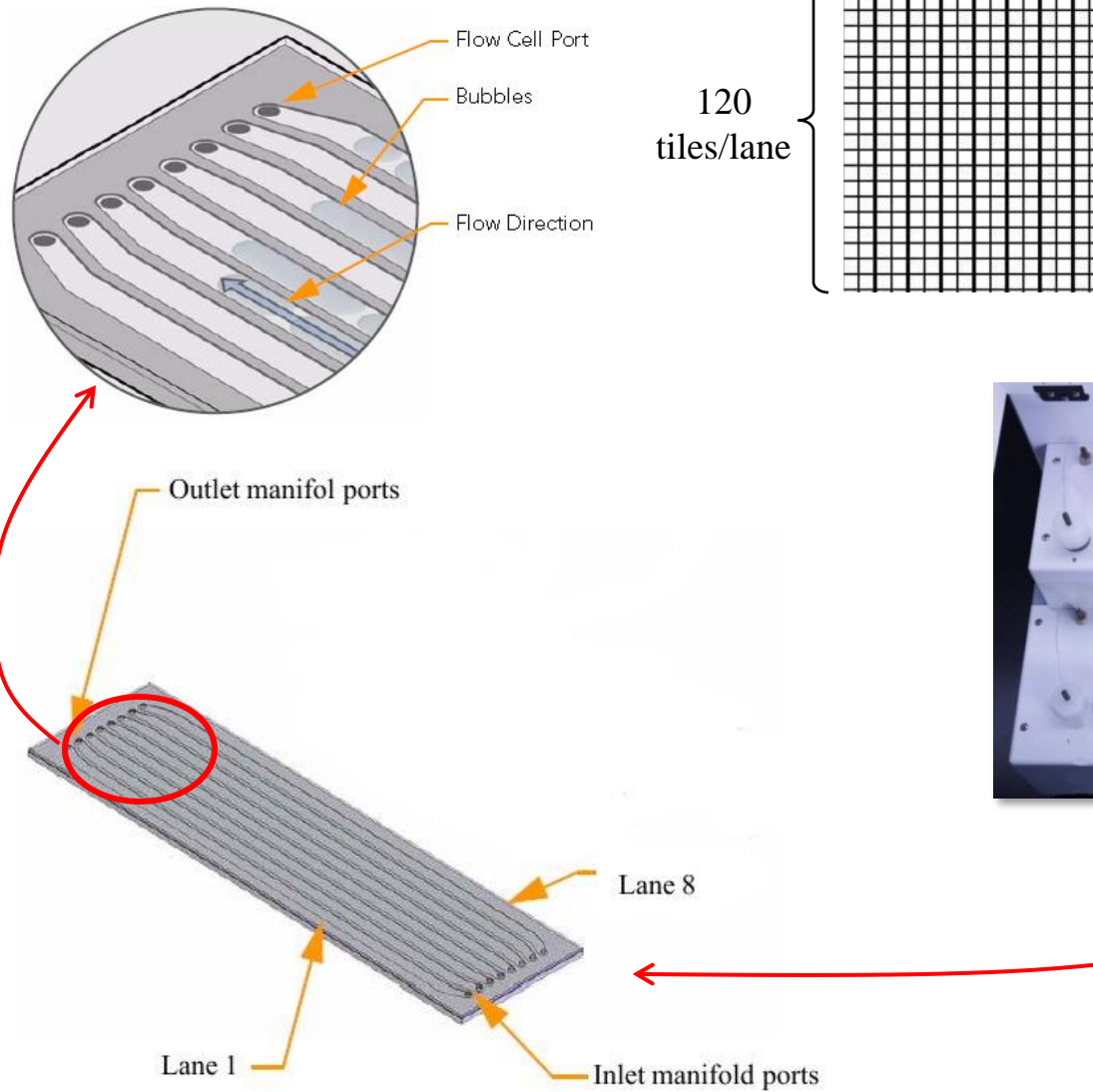
Analysis  
pipeline

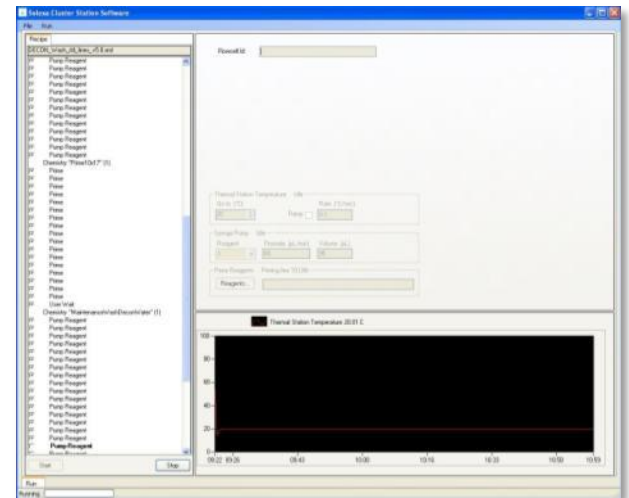
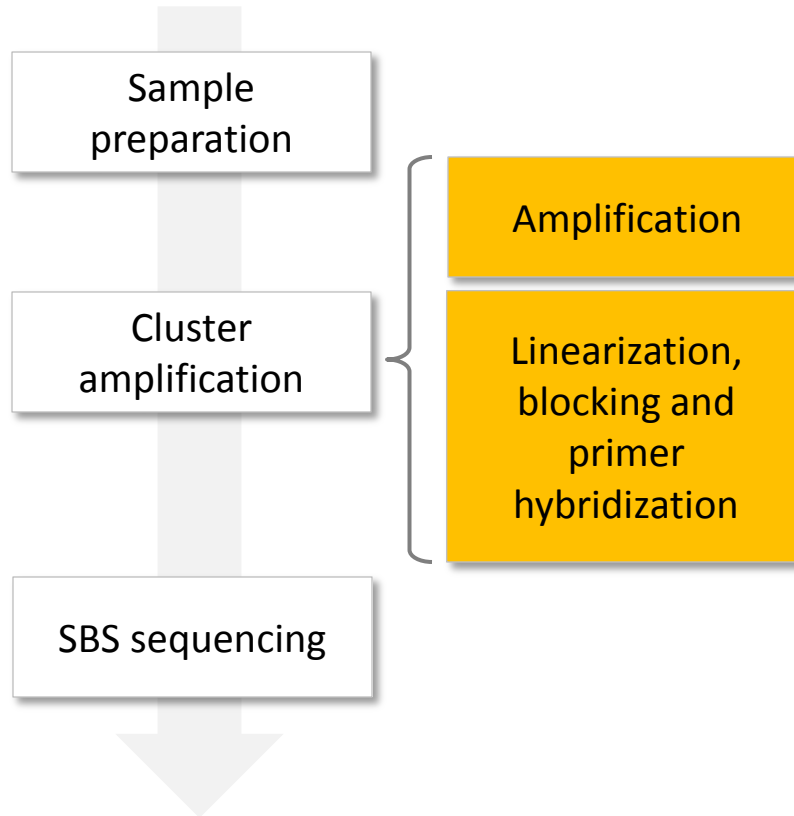


Bioanalyzer 2100

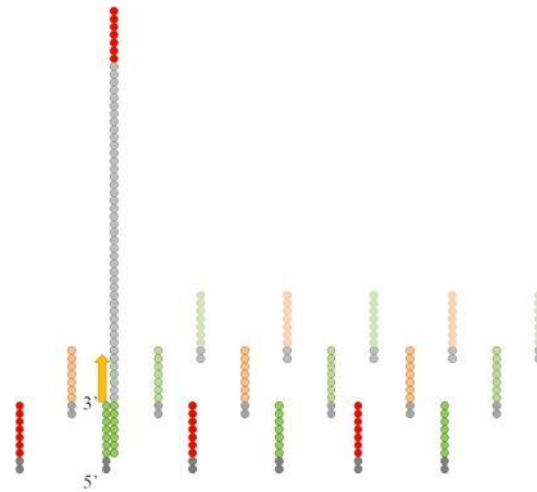




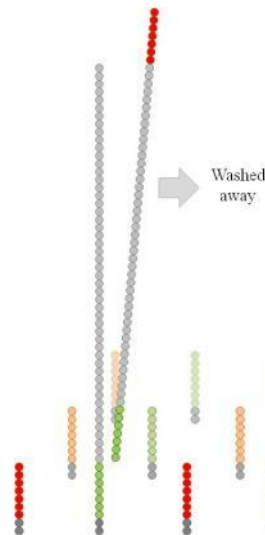
Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline



Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

Hybridize adapter-ligated forward  
fragment and extend

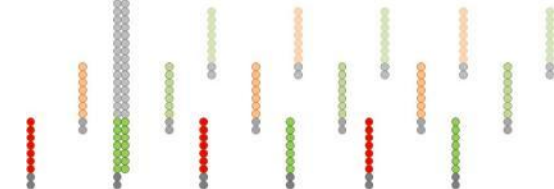


Washed  
away

Denature dsDNA and wash original forward template;  
reverse template stays covalently attached to the array



Extension is completed

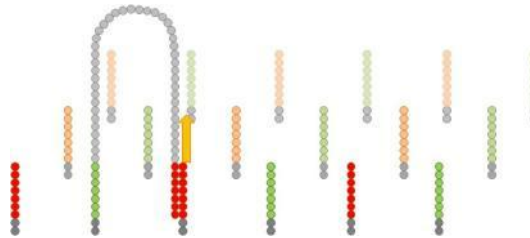


Sample  
preparation

Clusters  
amplification

Sequencing by  
synthesis

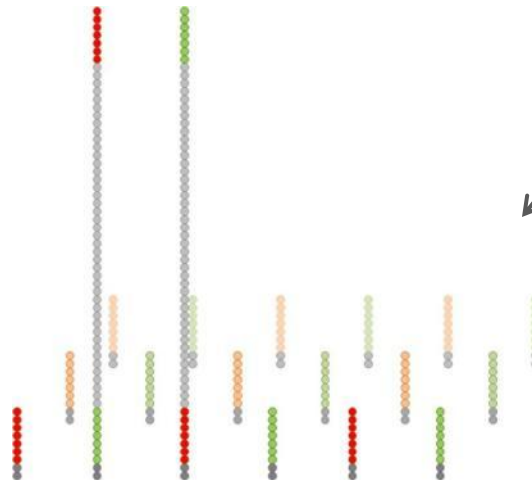
Analysis  
pipeline



Bridge amplification of the reverse fragment



Double-strand bridge is formed



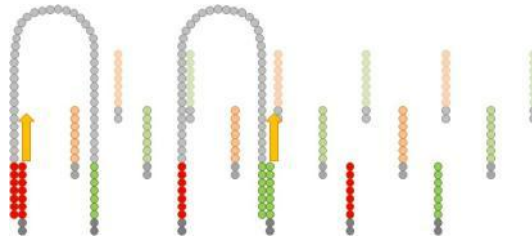
Double strand bridge is denatured and reverse as well as forward fragments are covalently attached to the array

Sample  
preparation

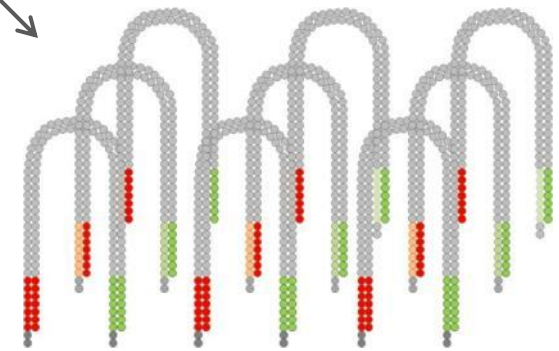
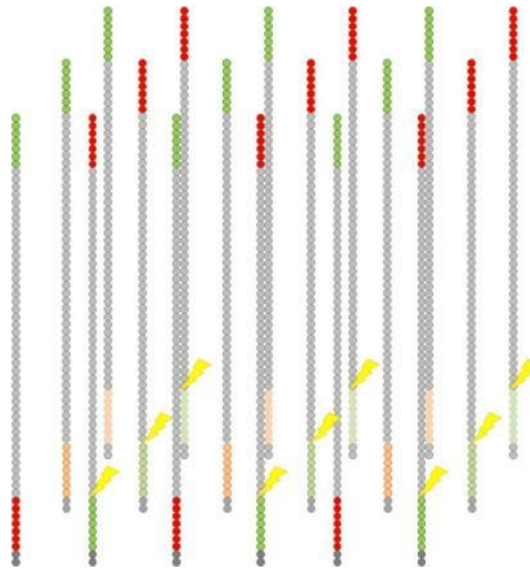
Clusters  
amplification

Sequencing by  
synthesis

Analysis  
pipeline



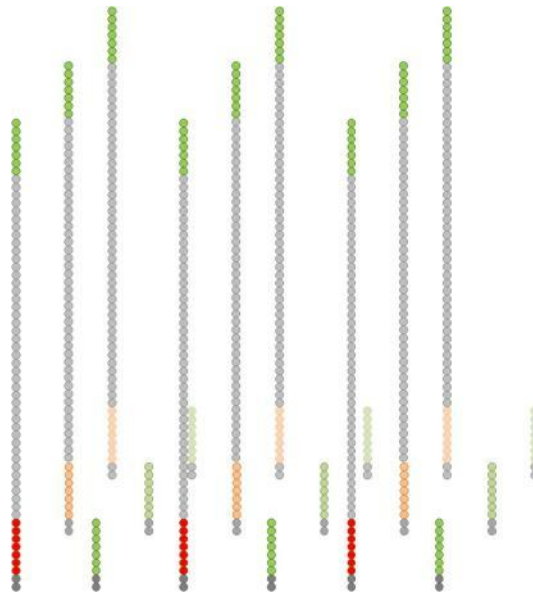
Bridge amplification is repeated to enlarge the cluster



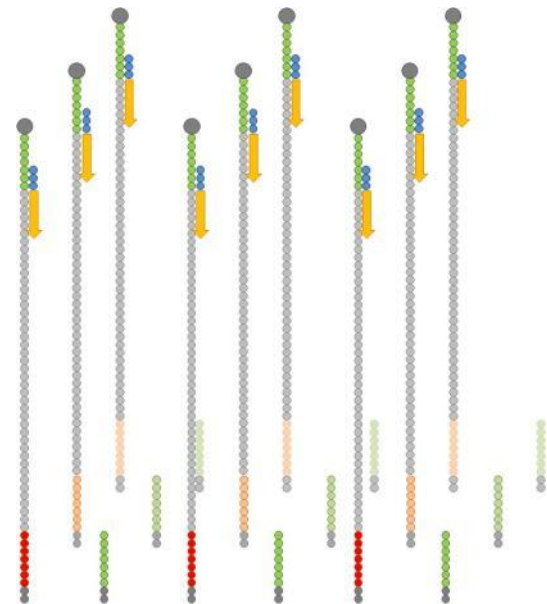
Double-strand bridges are denatured



Reverse strands fragments are cleaved and washed away

Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

Cluster with forward strands only,  
covalently attached to the array



Sequencing primers start the  
SBS process

[illegible]

## Sample preparation

## Workbench

## Clusters amplification

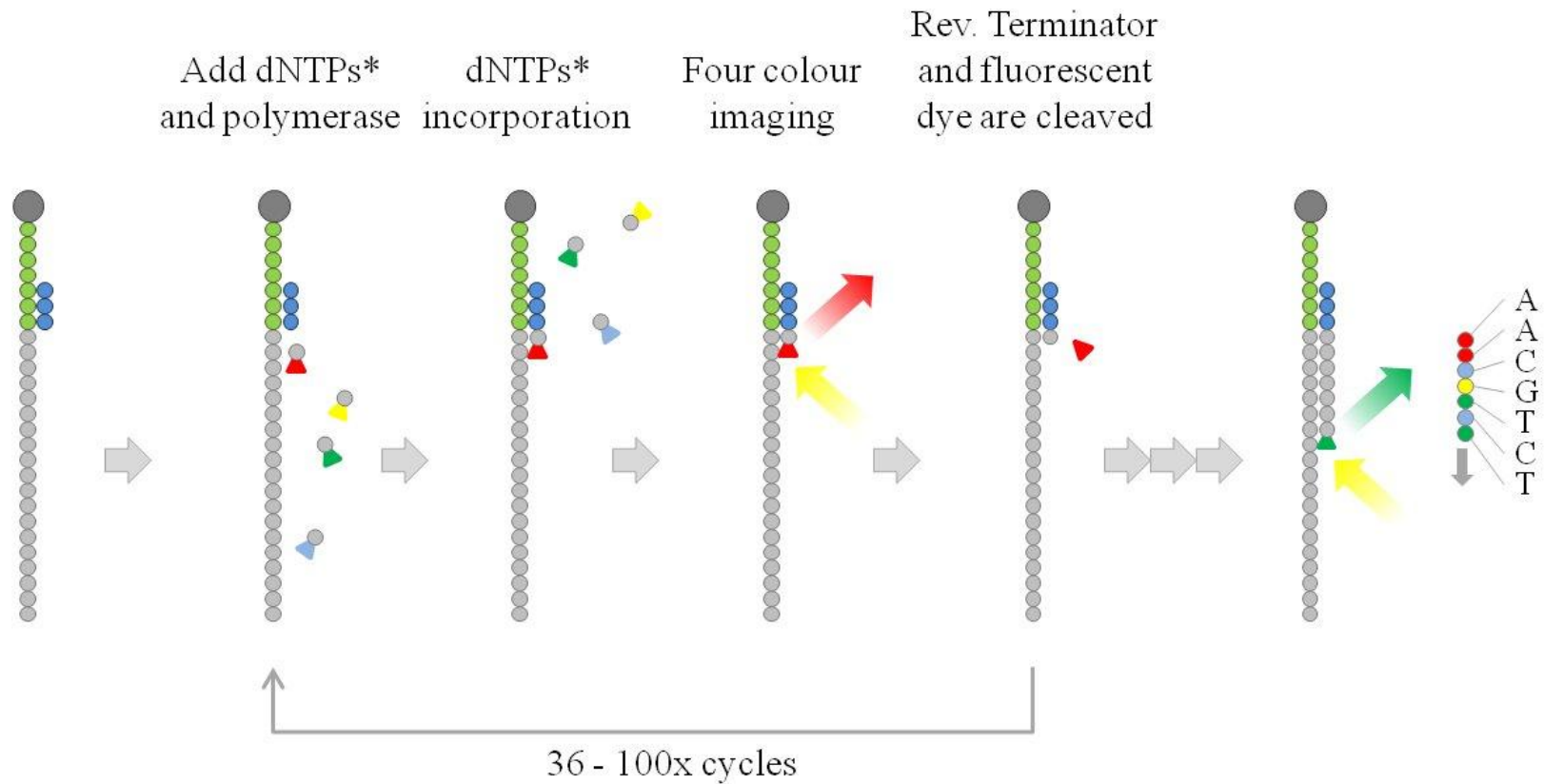
## Cluster Station

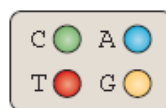
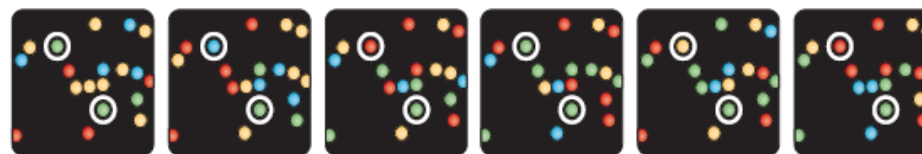
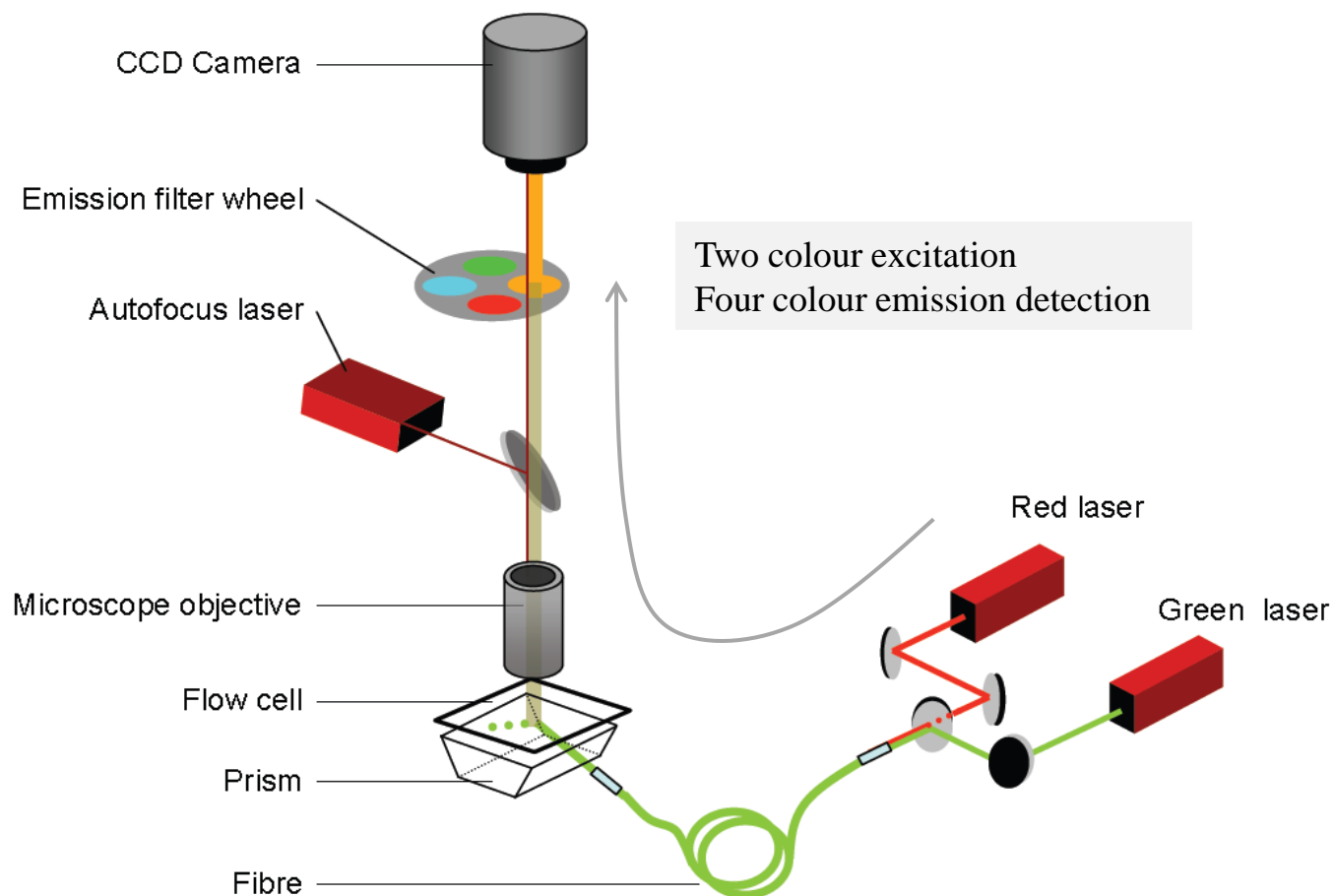
## Sequencing by synthesis

# Genome Analyzer

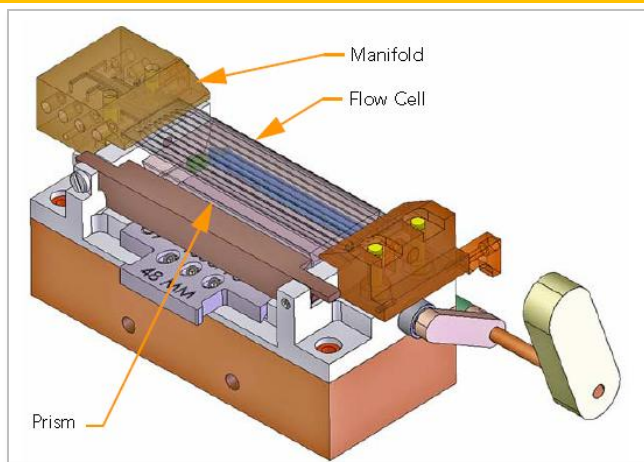
## Analysis pipeline

# Linux Server

Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

Top: CATCGT  
Bottom: CCCCCC

Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipelineCluster amplified  
FlowCell

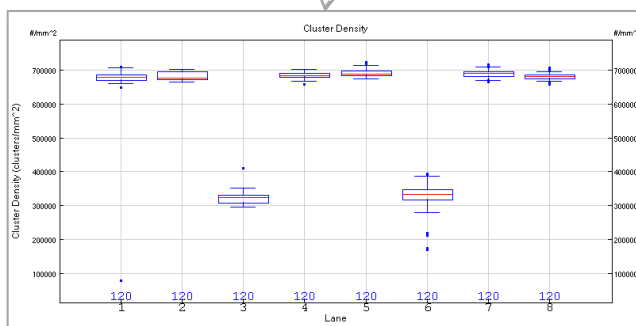
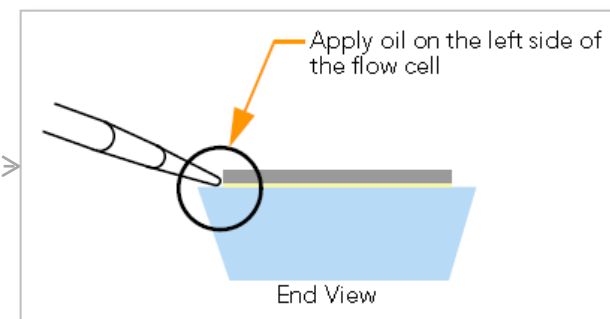
Install prism

Install flow-cell

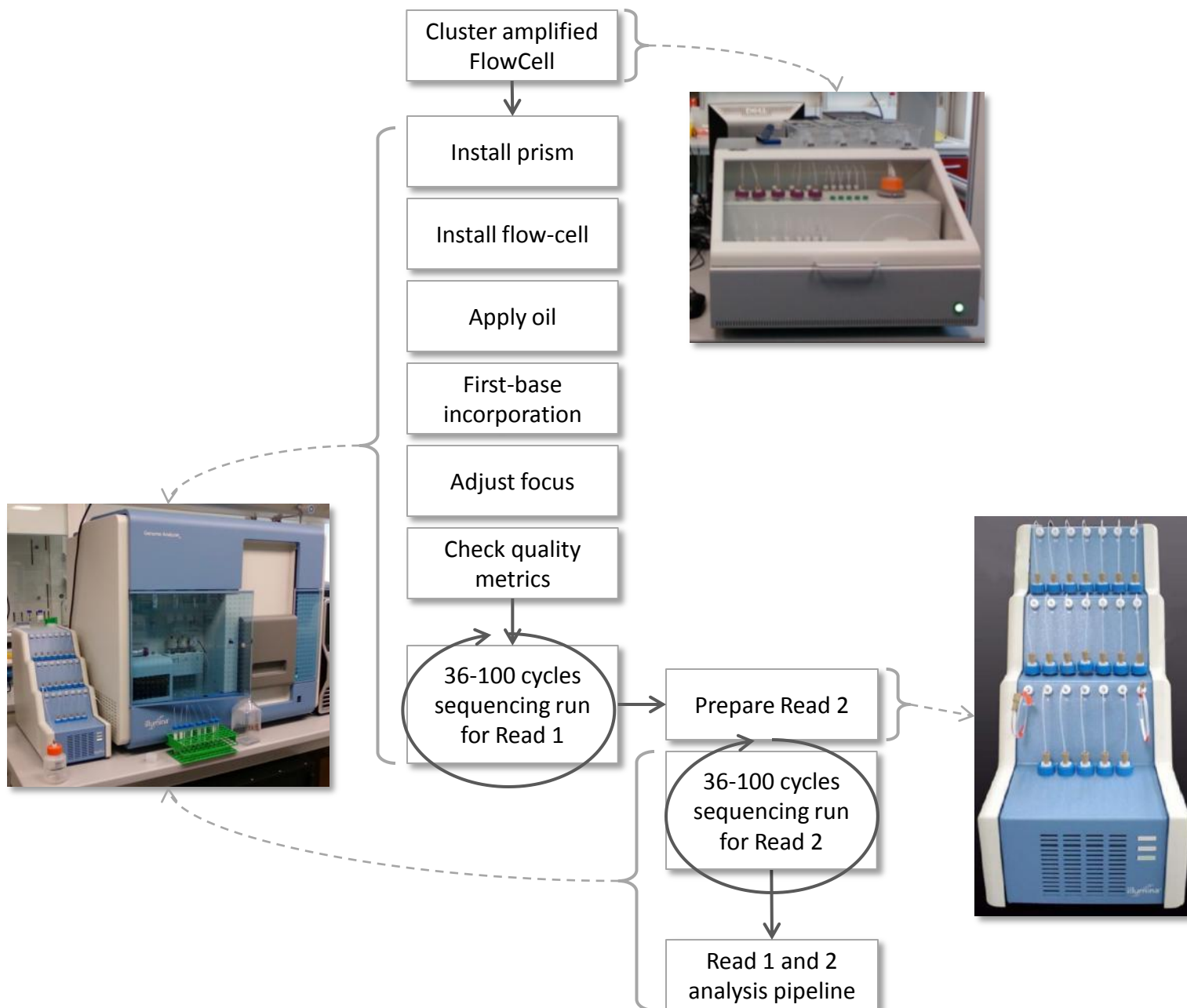
Apply oil

First-base  
incorporation

Adjust focus

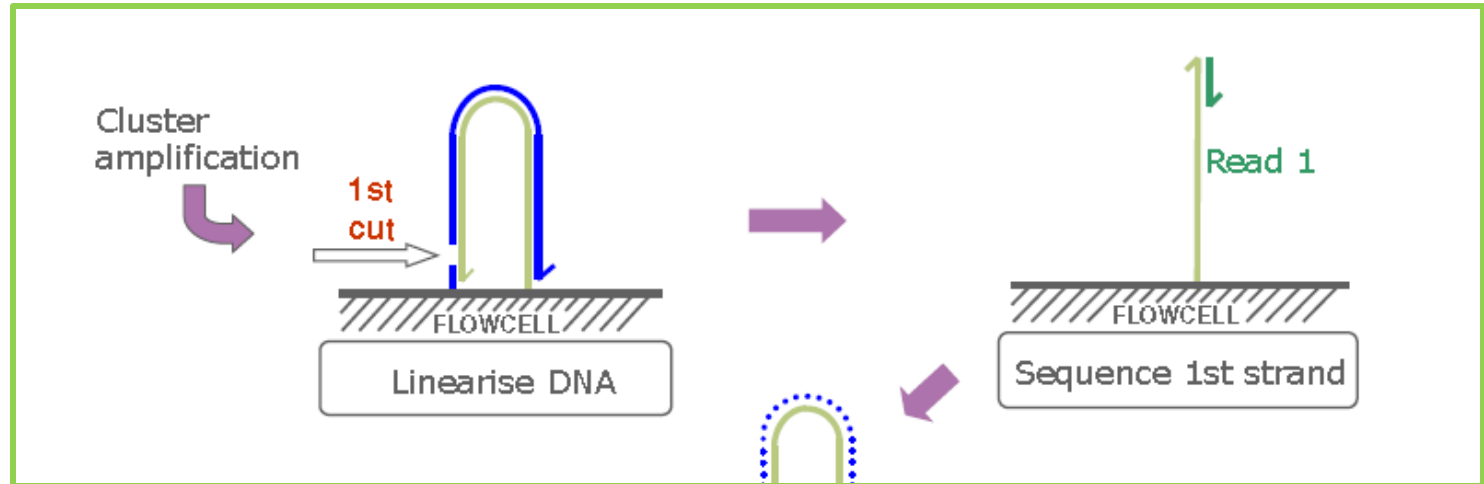
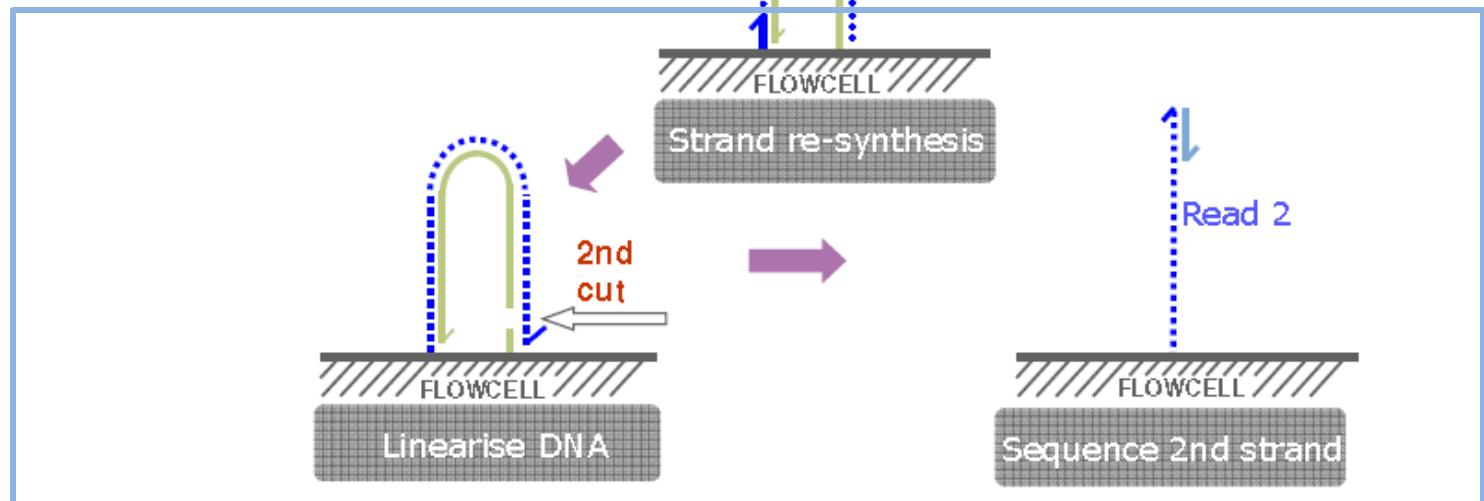
Check quality  
metrics36-100 cycles  
sequencing run  
for Read 1Read 1 analysis  
pipeline



Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

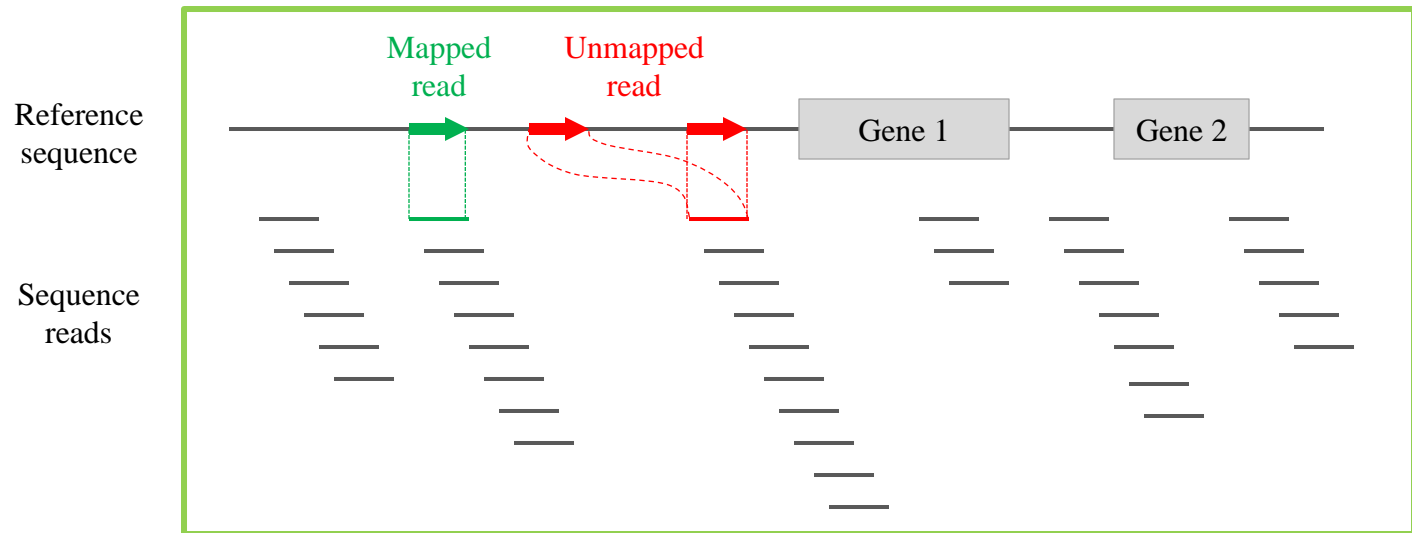
Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

Paired-end sequencing works into GA and uses chemicals from PE module to perform cluster amplification of the reverse strand

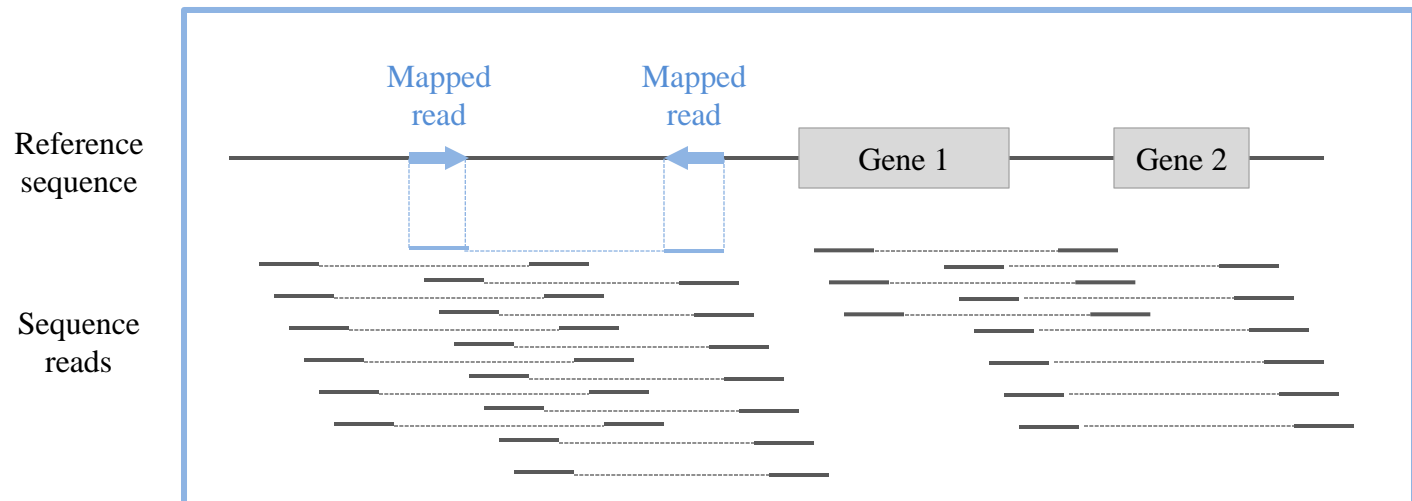
Single-read  
(CS)Paired-end  
(PEM)

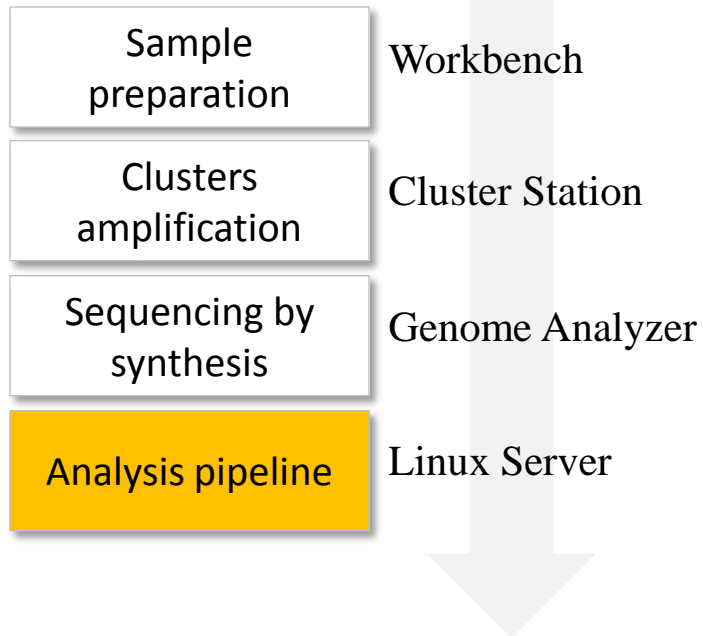
Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

## Single-read (read 1)



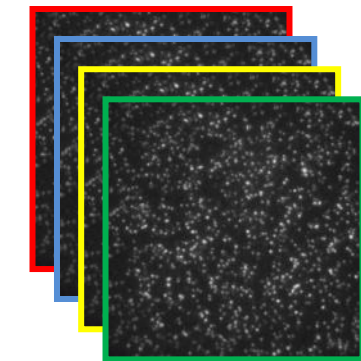
## Paired-end (read 1 &amp; read 2)





Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

Image files

Firecrest  
*From image  
to intensity*

Intensity files

File	Lane	Tile	x	y	Intensity
s_1_0007_int.txt	7	135	563	168.9	347.7
s_1_0007_int.txt	7	180	621	231.3	341.9
s_1_0007_int.txt	7	245	626	218.4	356.8
s_1_0007_int.txt	7	241	509	187.7	382.7
s_1_0007_int.txt	7	214	595	173.5	372.1
s_1_0007_int.txt	7	155	544	172.2	339.5
s_1_0007_int.txt	7	301	507	353.8	472.1
s_1_0007_int.txt	7	175	606	210.4	331.4
s_1_0007_int.txt	7	242	522	267.9	511.0
s_1_0007_int.txt	7	196	522	220.2	455.9
s_1_0007_int.txt	7	237	612	167.0	457.7
s_1_0007_int.txt	7	160	528	172.6	400.7
s_1_0007_int.txt	7	164	543	205.7	385.0
s_1_0007_int.txt	7	179	581	207.2	372.9
s_1_0007_int.txt	7	226	623	218.3	400.6
s_1_0007_int.txt	7	139	583	241.0	358.9
s_1_0007_int.txt	7	220	618	225.1	496.8
s_1_0007_int.txt	7	360	507	194.0	339.0
s_1_0007_int.txt	7	334	512	249.0	590.6
s_1_0007_int.txt	7	125	517	218.7	345.4
s_1_0007_int.txt	7	343	541	183.5	375.9
s_1_0007_int.txt	7	241	608	208.6	361.2
s_1_0007_int.txt	7	176	520	226.3	338.6
s_1_0007_int.txt	7	371	592	298.6	566.4
s_1_0007_int.txt	7	271	508	175.8	391.5
s_1_0007_int.txt	7	195	520	236.4	388.5
s_1_0007_int.txt	7	301	592	181.8	378.8
s_1_0007_int.txt	7	248	548	197.7	525.1
s_1_0007_int.txt	7	743	570	709.7	702.0

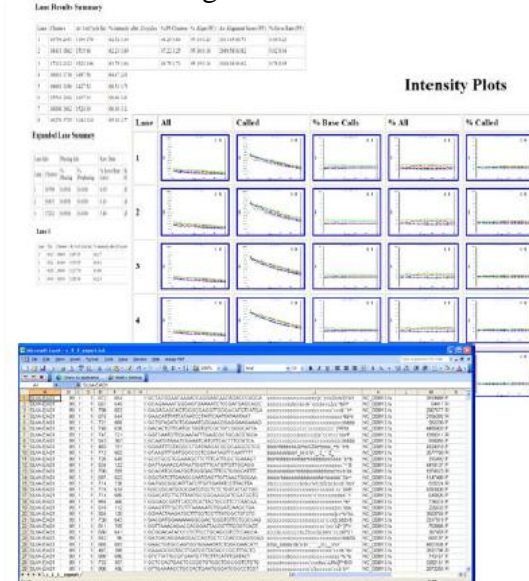
Bustard  
*From intensity  
to reads*

Base calls files

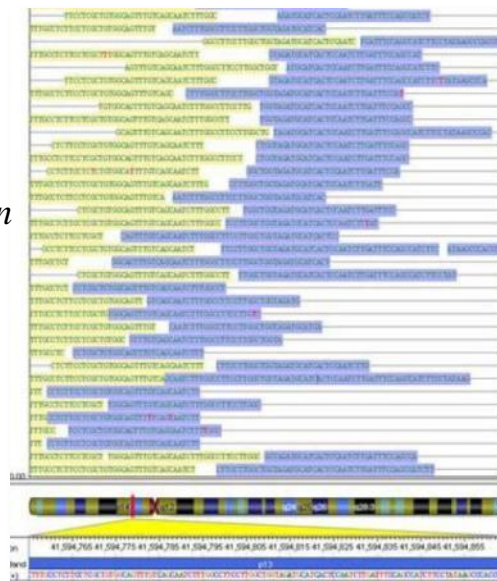
File	Lane	Tile	x	y	Base Call
s_1_0007_seq.txt	7	135	563	168.9	TTTGAACAGCATTTTATGATGACGAC
s_1_0007_seq.txt	7	180	621	231.3	TGTTTTTTTTTTTTTTTGGAGACAG
s_1_0007_seq.txt	7	245	626	218.4	TTTGAACAGTTTTTTTGGTGTGAGGC
s_1_0007_seq.txt	7	241	509	187.7	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	214	595	173.5	TACAAATCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	155	544	172.2	TTATCTGCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	301	507	353.8	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	175	606	210.4	TTGGAATCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	242	522	267.9	TACATATATATATATATATATATATAT
s_1_0007_seq.txt	7	196	522	220.2	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	237	612	167.0	TTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	160	528	172.6	TATCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	164	543	205.7	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	179	581	207.2	TTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	226	623	218.3	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	139	583	241.0	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	220	618	225.1	TTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	360	507	194.0	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	334	512	249.0	TTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	125	517	218.7	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	343	541	183.5	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	241	608	208.6	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	176	520	226.3	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	371	592	298.6	TTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	271	508	175.8	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	195	520	236.4	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	301	592	181.8	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	248	548	197.7	TCTCTCTCTCTCTCTCTCTCTCTCTCT
s_1_0007_seq.txt	7	743	570	709.7	TCTCTCTCTCTCTCTCTCTCTCTCTCT

Gerald/ELAND  
*Alignment to genome*

Alignment files



Assembly

GenomeStudio  
*Data visualization*

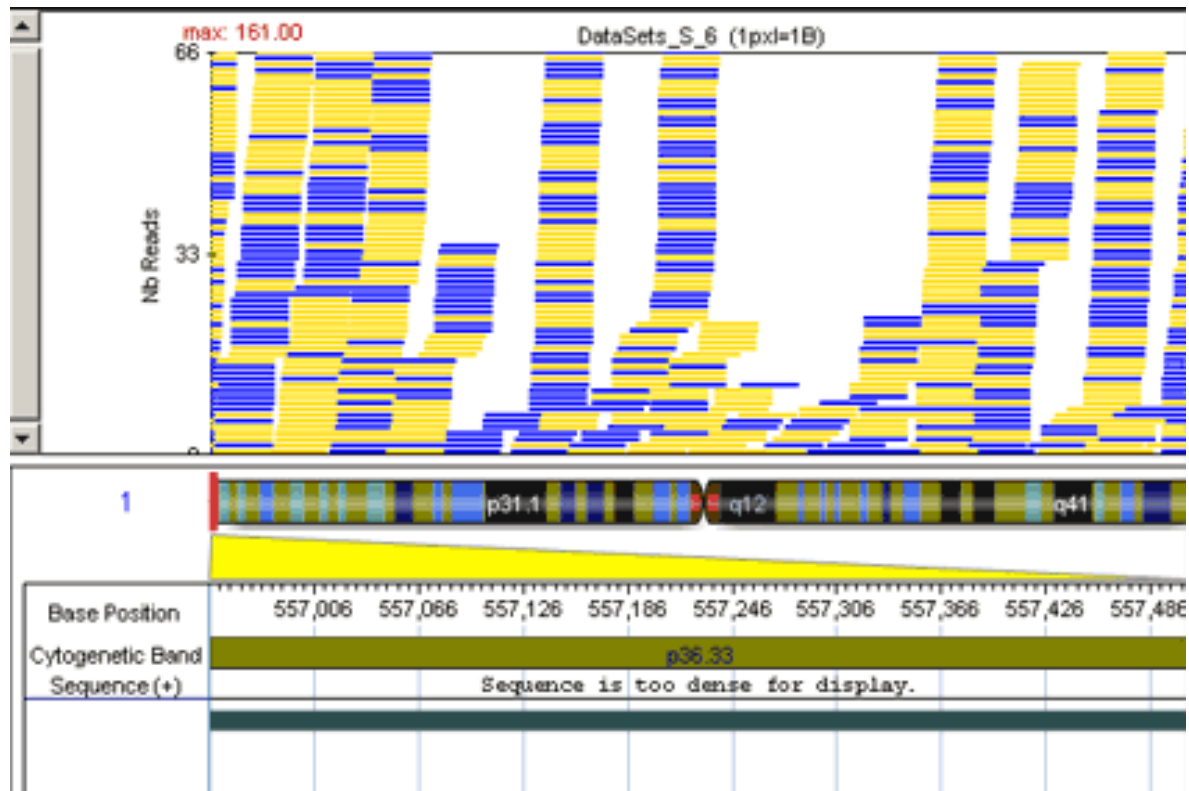
CASAVA

*Consensus assembly*Sequence  
ANALYSIS



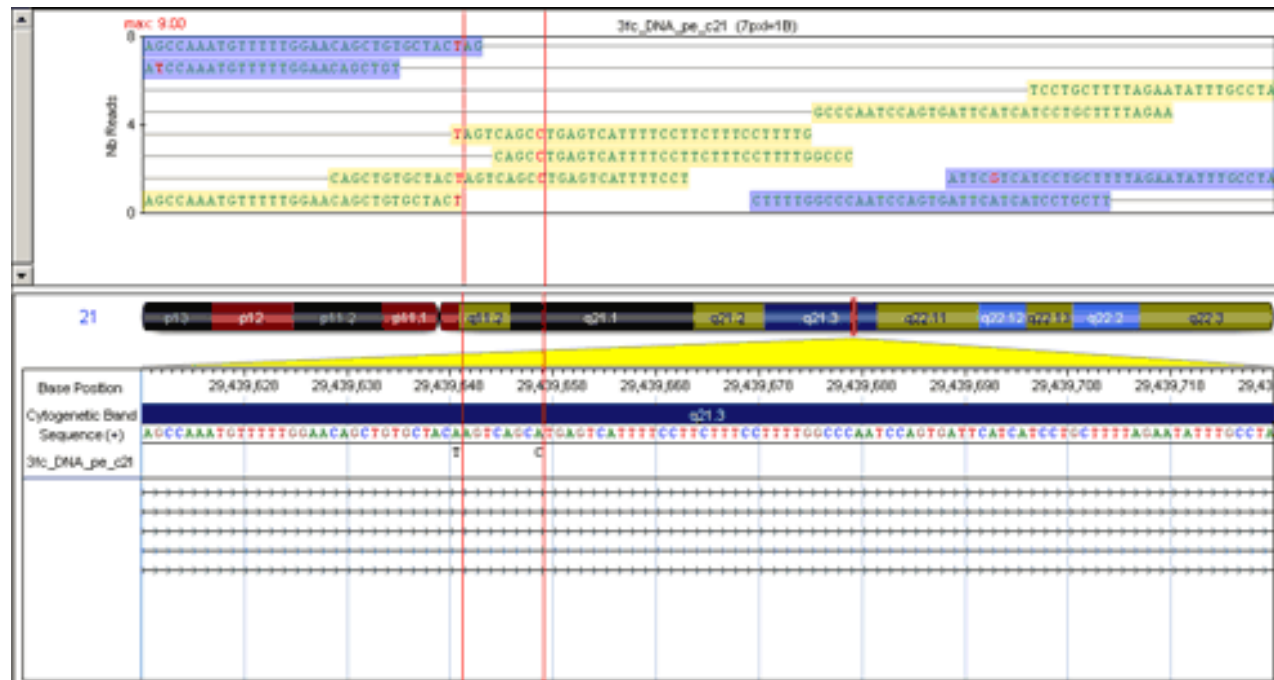
Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

## RNA sequencing



Sample  
preparationClusters  
amplificationSequencing by  
synthesisAnalysis  
pipeline

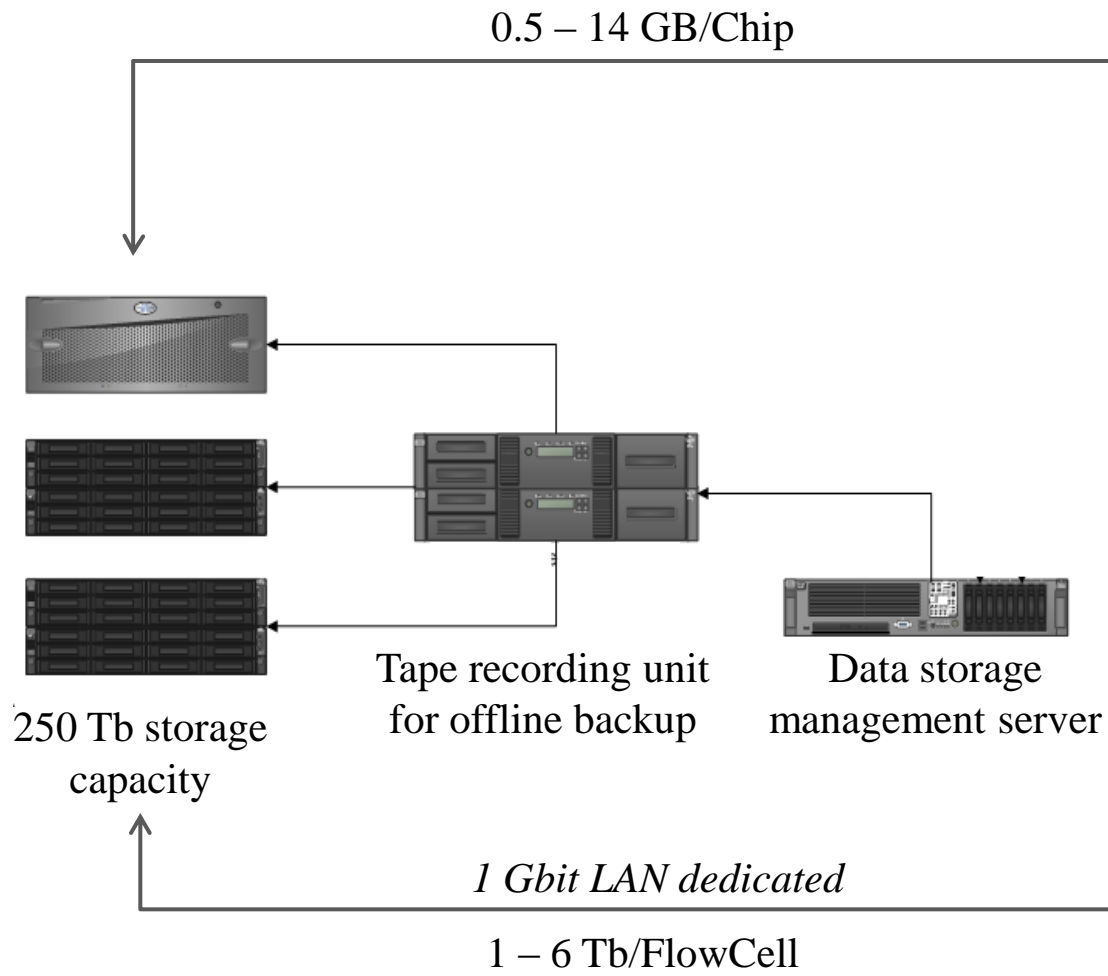
## DNA sequencing



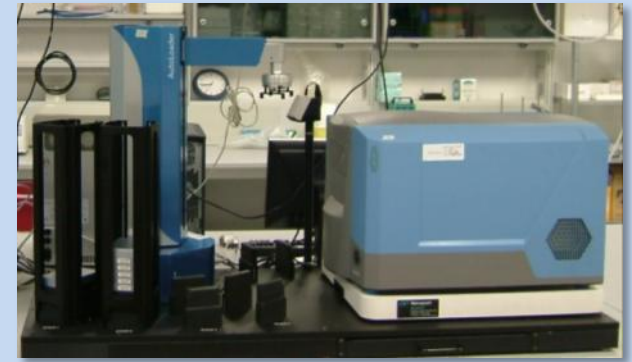


# Summary

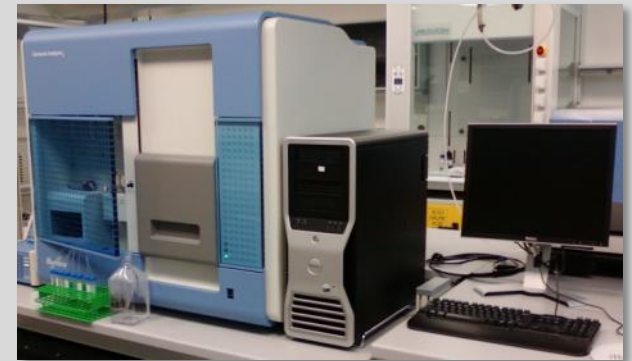
- 1 Towards NGS sequencing
- 2 NGS with Illumina GAIIx
- 3 Data management
- 4 Target enrichment

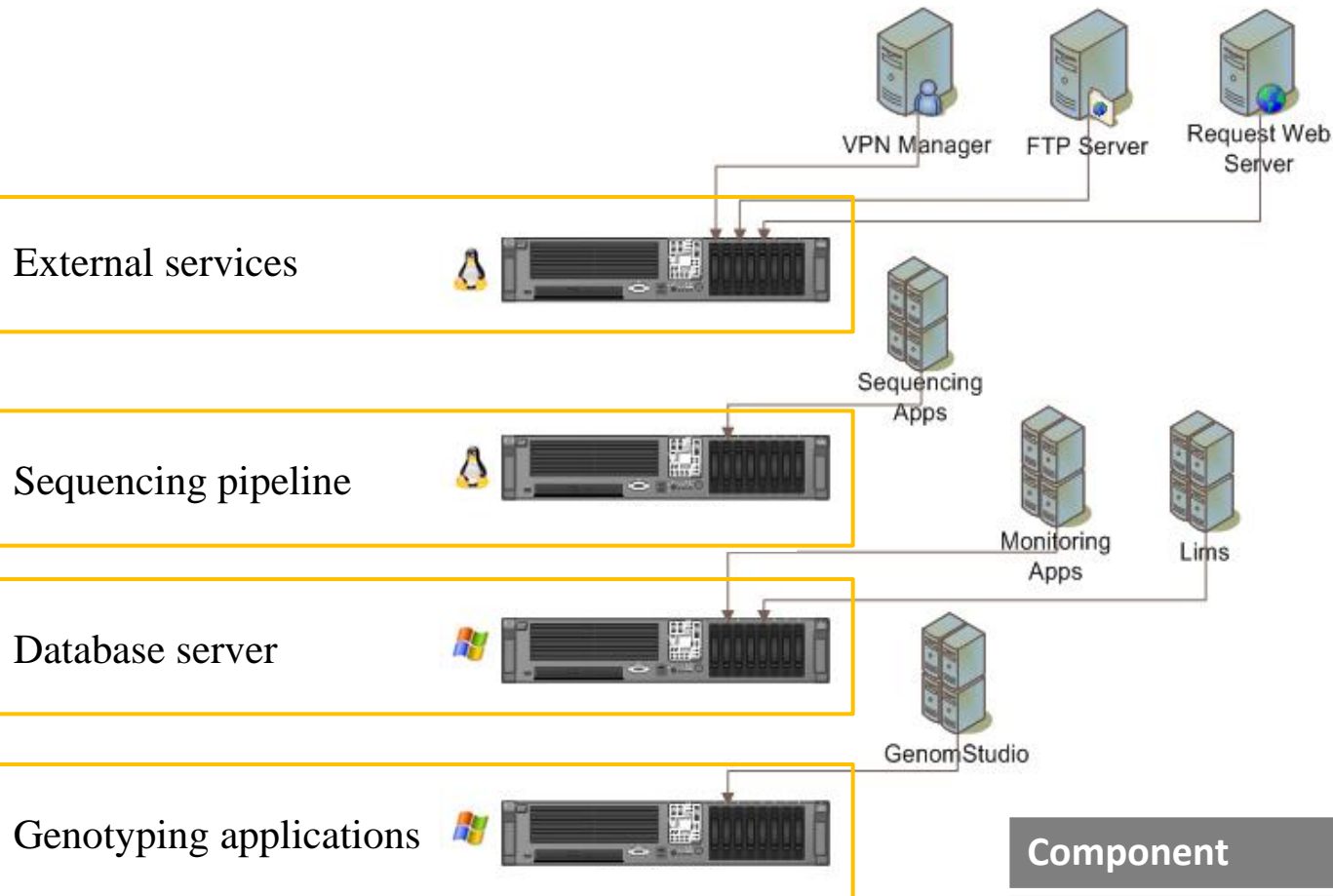


Genotyping units



Sequencing unit

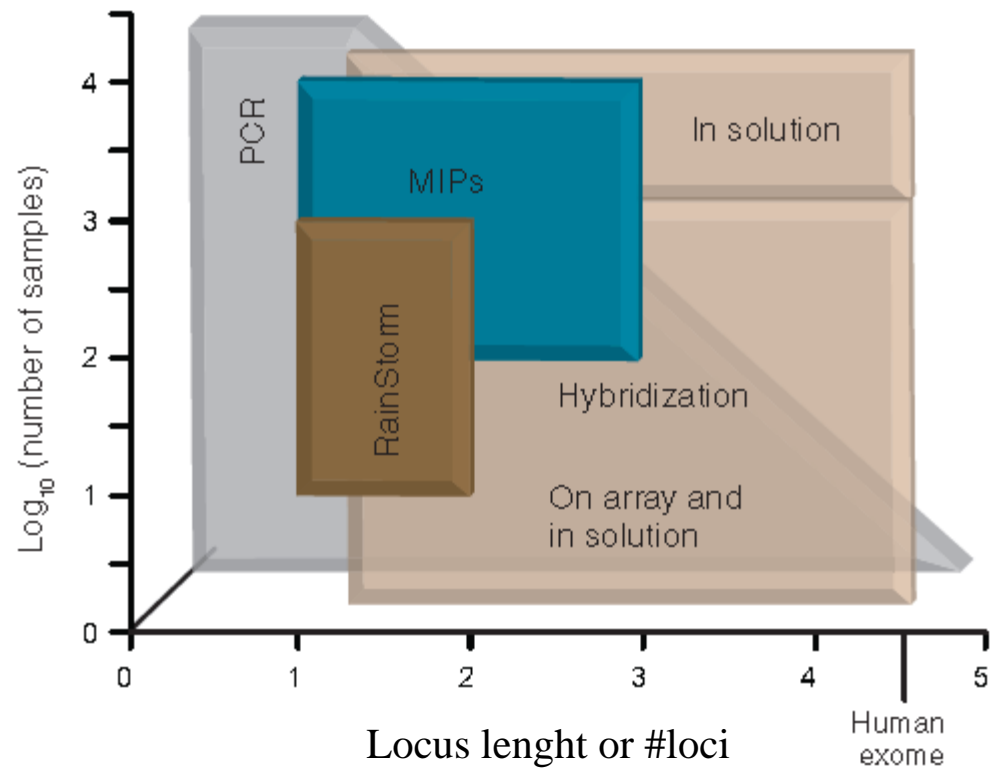




Component	Availability
CPU	40 core equivalent
RAM	1 Tb
Storage on-line (HD)	250 Tb
Storage off-line (Tape)	80 Tb normal 160 Tb compressed

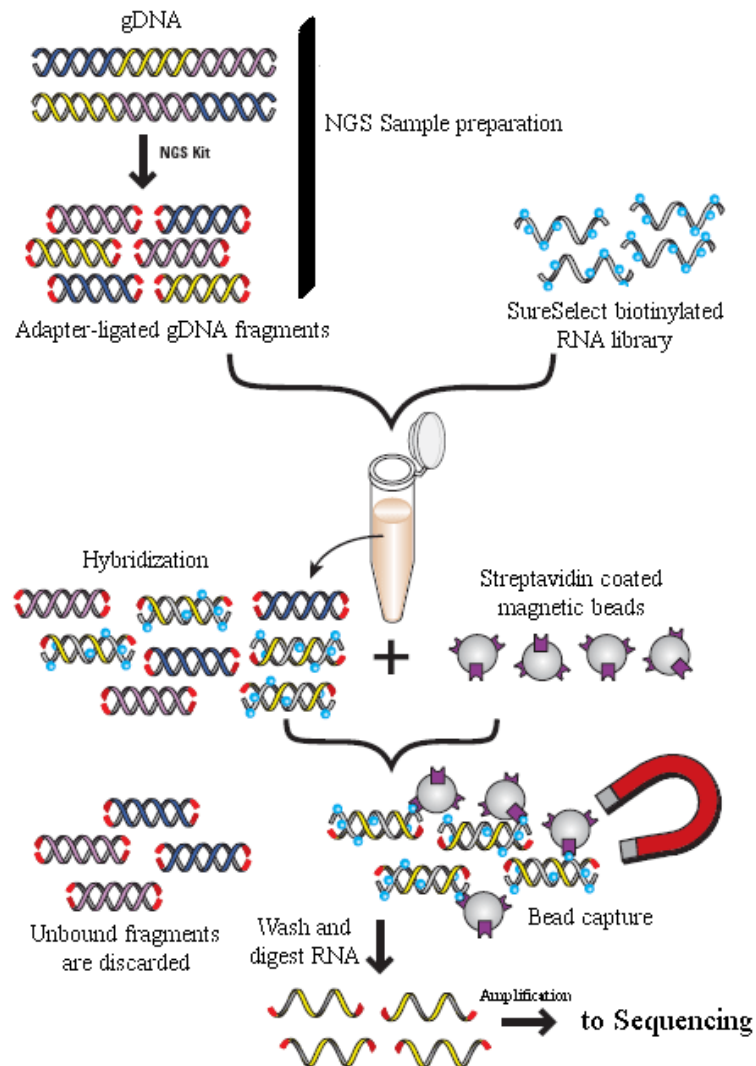
# Summary

- 1 Towards NGS sequencing
- 2 NGS with Illumina GAIIx
- 3 Data management
- 4 Target enrichment



*Nature Methods*, 2010, 7: 111-118

## Agilent SureSelect Solution-phase capture with streptavidin-coated magnetic beads



Reported 60-80% of capture efficiency

# The end