

Introduction to high throughput sequencing

Rita Holdhus
December 2011

microarray.no



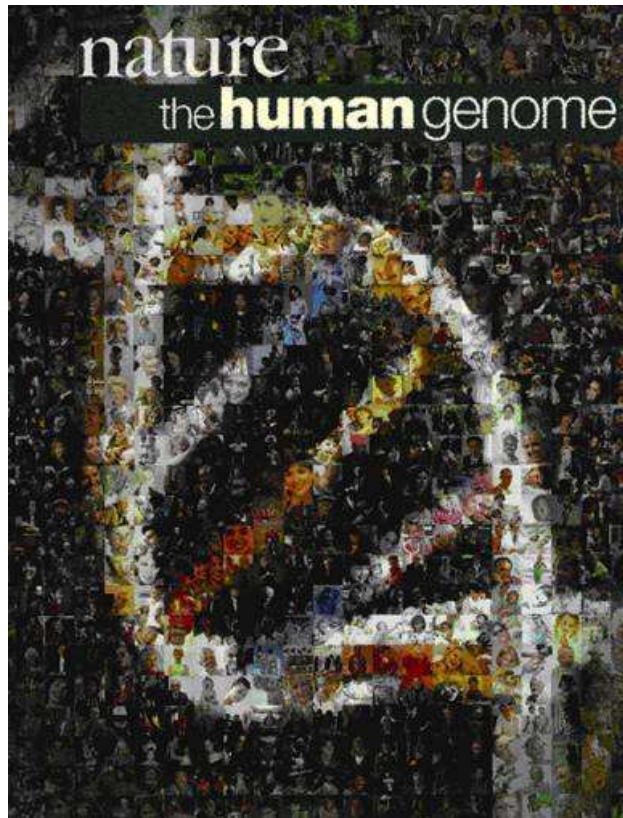
Topics

- Introduction
- Technology
- Applications

microarray.no



Human genome project



**Public HGP
1990-2003
approx. 3 Billion dollars**



Celera Genomics

microarray.no



1 000 000 000

Single
molecule?



Year

Human Genome Project

Stratton MR et al, Nature 2009

www.uib.no



microarray.no

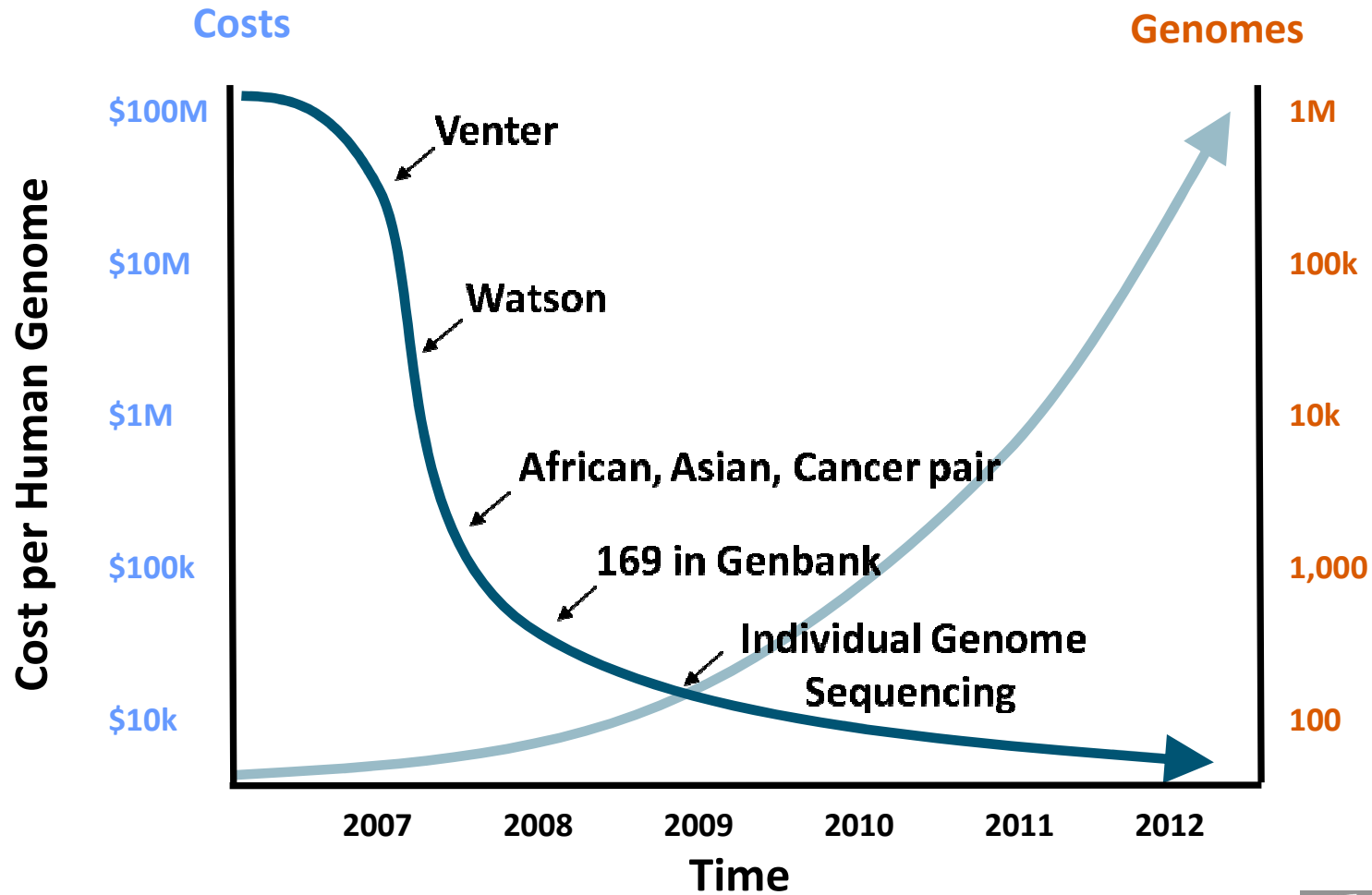
Sanger vs Next-Generation Sequencing

Sanger	Next Generation Sequencing
• Read length: ~750 bp	• Read length: 25 – 500 bp
• Microliter volumes	• Picoliter volumes
• Capacity: 96-384 capillaries	• Highly parallelized
• Expensive per base	• Cheap per base
• More accurate	• More error prone
• Some bias in amplification	• Bias free amplification

microarray.no



Cost per genome



Technology

Clonal cluster
sequencing

- **Solexa (Illumina)**
 - Sequencing by synthesis
- **454 (Roche)**
 - Pyrosequencing
- **SOLiD (Applied Biosystems)**
 - Sequencing by ligation
- **Single molecule sequencing**
 - Pacific Biosciences
- **Non optical**
 - Ion Torrent
- ++++++



microarray.no



Systems

	IlluminaHiSeq2000	SOLiD5500xl	Roche 454 FLX
Sequencing method	Sequencing by synthesis	Sequencing by ligation	Pyrosequencing
Amplification method	Isothermal bridge amplification	Emulsion PCR	Emulsion PCR
Read lengths	Up to 100bp	Up to 75bp	Up to 700bp
Output (Gb/run)	600 Gb	180 Gb ¹ / 300 Gb ²	400 Mb
Run time	1-8 days	1-7 days	10 - 23 hours
Flow cells/chips/plates	2 flow cells	2 flow chips	1 picotiter plate

microarray.no

¹Output using microbeads

²Output using nanobeads



Sequencing a genome of 432 Mb

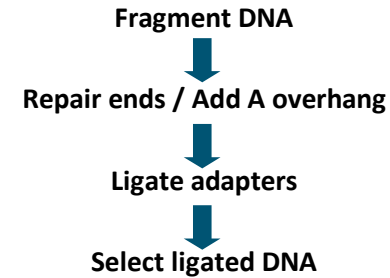
Platform	ABI3730xl Genome Analyzer	Roche (454) FLX	Illumina Genome Analyzer	ABI SOLiD	Helicos Heliscope
Sequencing Speed	0.03-0.07 Mb/h	13 Mb/h	25 Mb/h	21–28 Mb/h	83 Mb/h
Time to sequence (days)	2185.7	11.8	6.1	5.5	1.8

microarray.no



Sequencing workflow

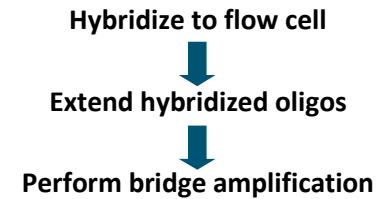
1 Library preparation



2 Automated Cluster Generation



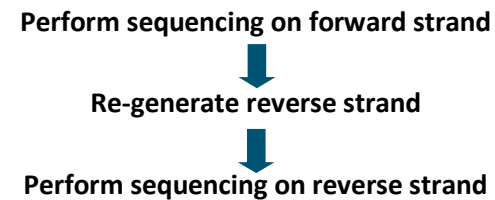
1-8 samples



3 Sequencing



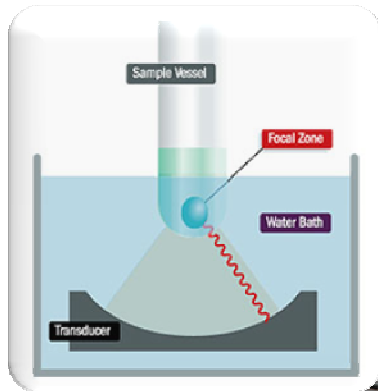
1-16 samples



microarray.no

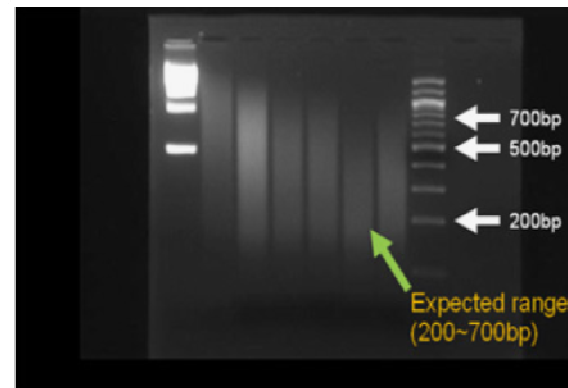
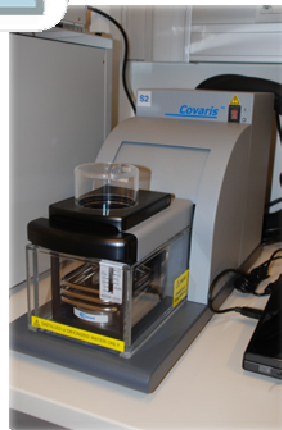


DNA fragmentation



COVARIS Adaptive Focused Acoustics

- Acoustic energy wave that converges and focuses to a small-localized area
- Shearing of DNA, RNA, Chromatin, +++



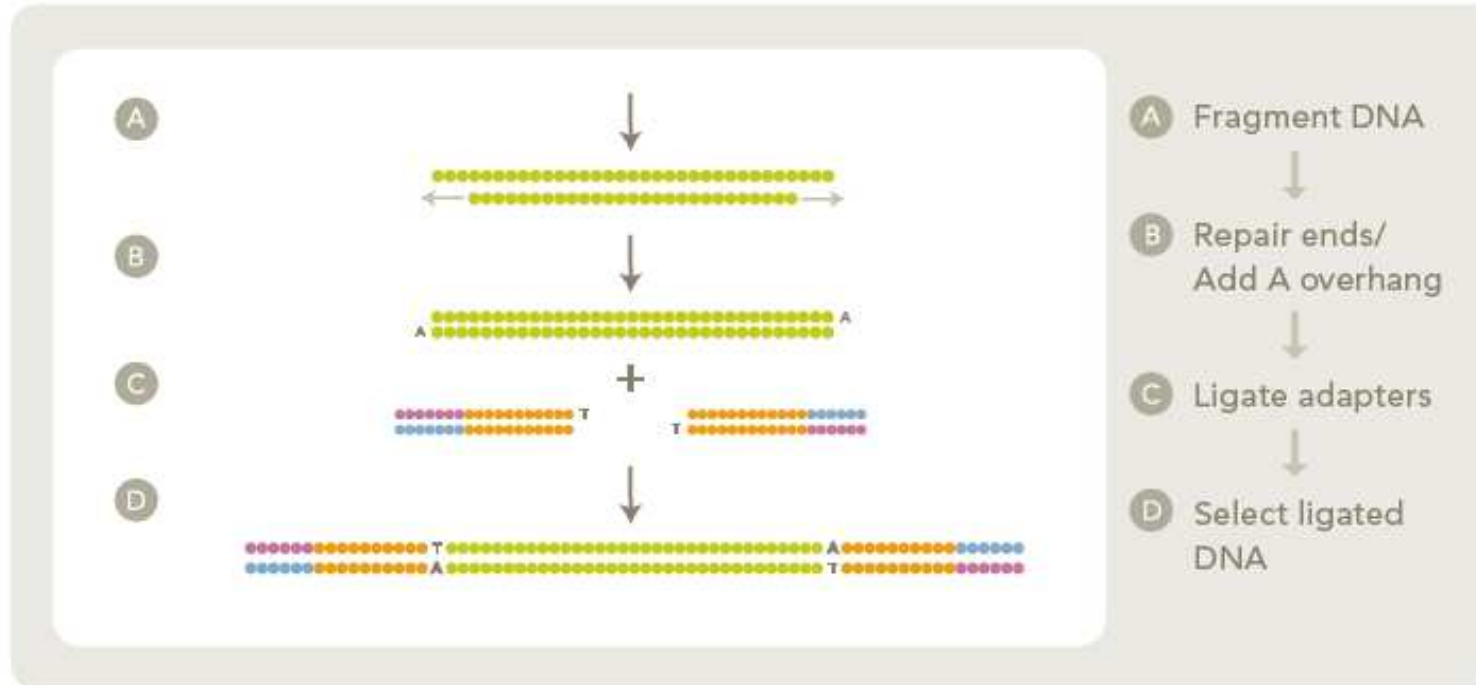
Illumina Sequencing by synthesis



microarray.no



Library prep

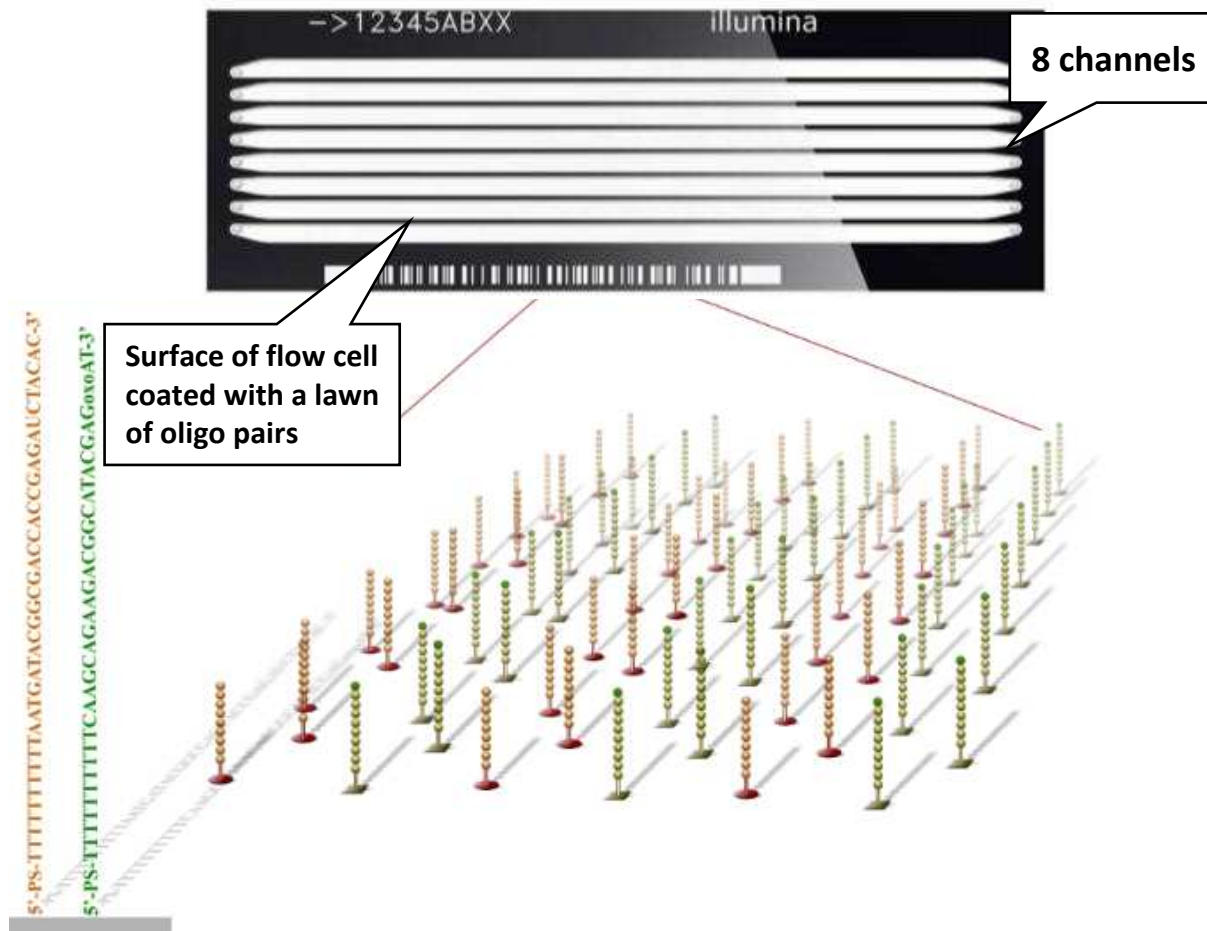


microarray.no

→ Library QC: Real-Time assay and Qubit quantification!



Flowcell



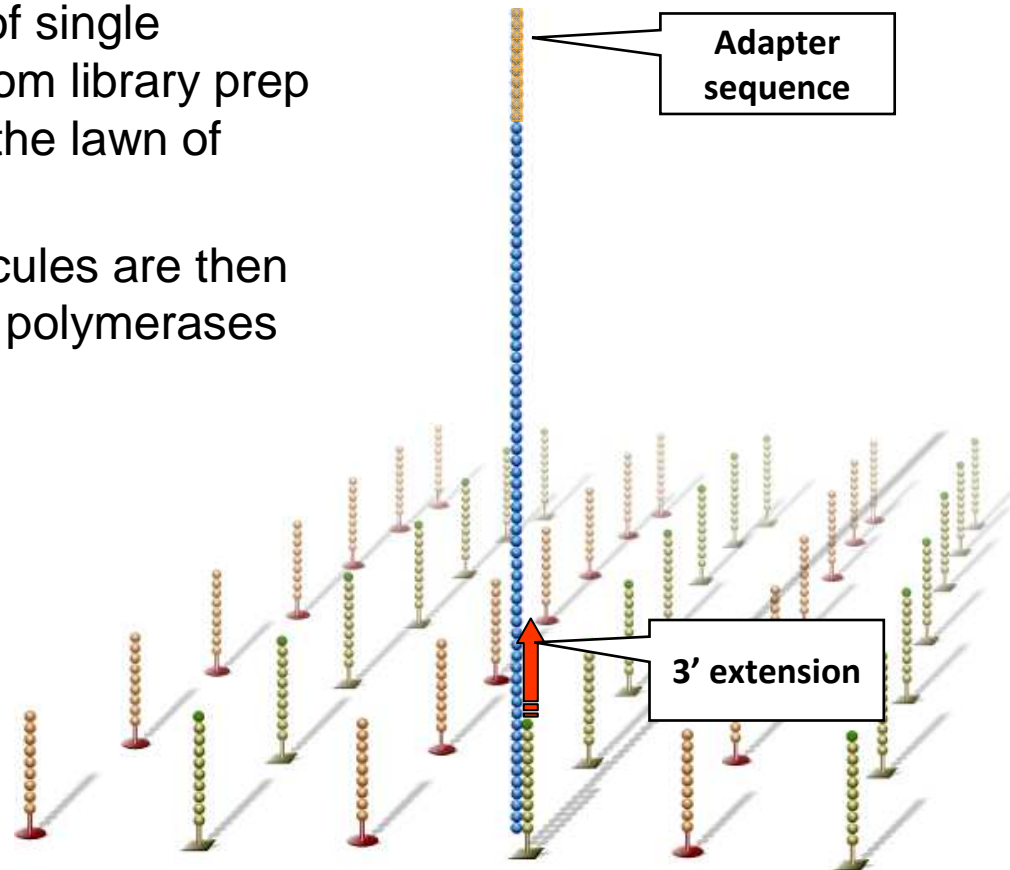
microarray.no



Clustering

Hybridize fragment and extend

- Thousands of single molecules from library prep hybridize to the lawn of primers
- Bound molecules are then extended by polymerases



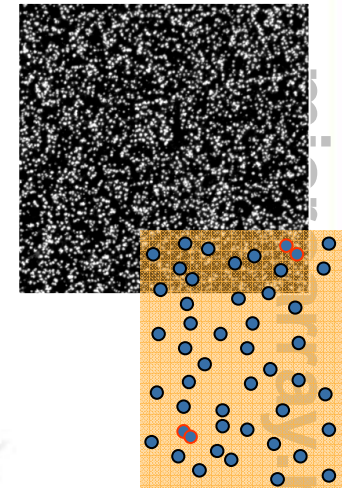
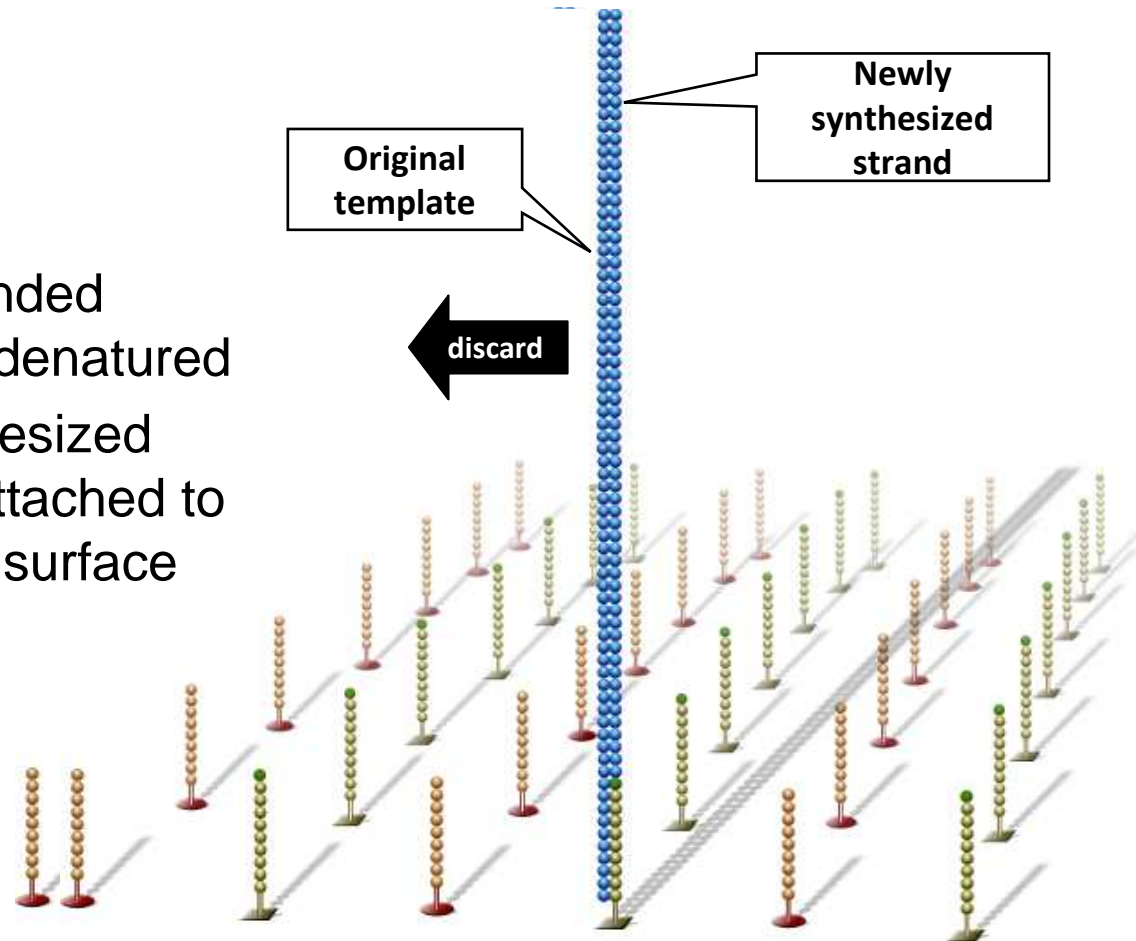
microarray.no



Clustering

Denature double-stranded DNA

- Double-stranded molecule is denatured
- Newly synthesized covalently attached to the flow cell surface



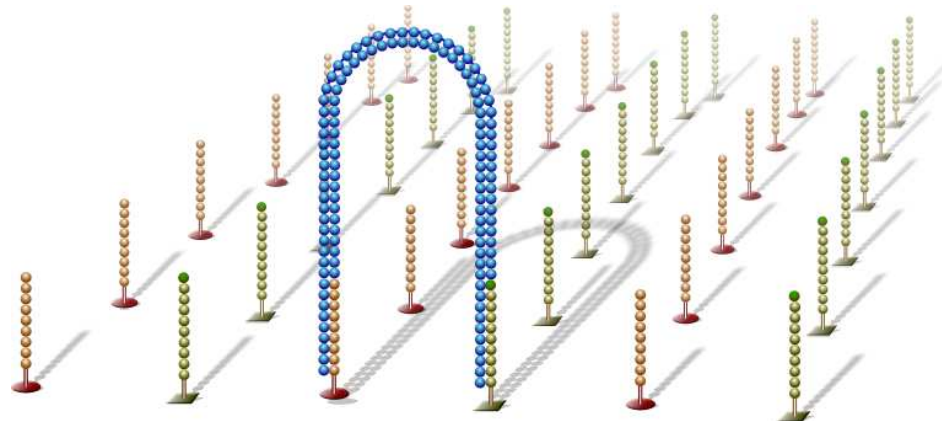
Single molecules bound to flow cell in a random pattern



Clustering

Bridge amplification

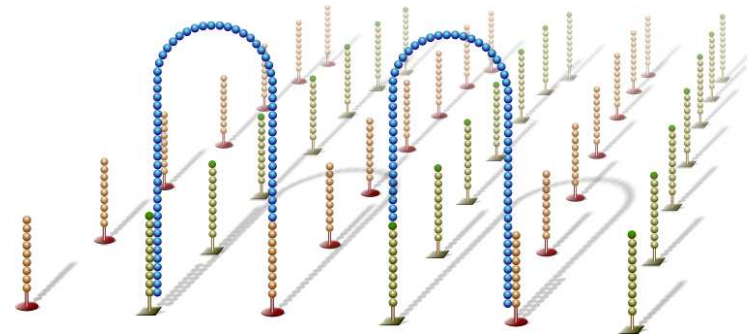
- Single-strand flips over to hybridize to adjacent primers to form a bridge
- Hybridized primer is extended by polymerases
- Double-stranded bridge is formed



Clustering

Denature double-stranded DNA

- Double-stranded bridge is denatured
- Result: Two copies of covalently bound single-stranded templates
- Single-strands flip over to hybridize to adjacent primers to form bridges
- Hybridized primer is extended by polymerase
- Process repeated 30 times



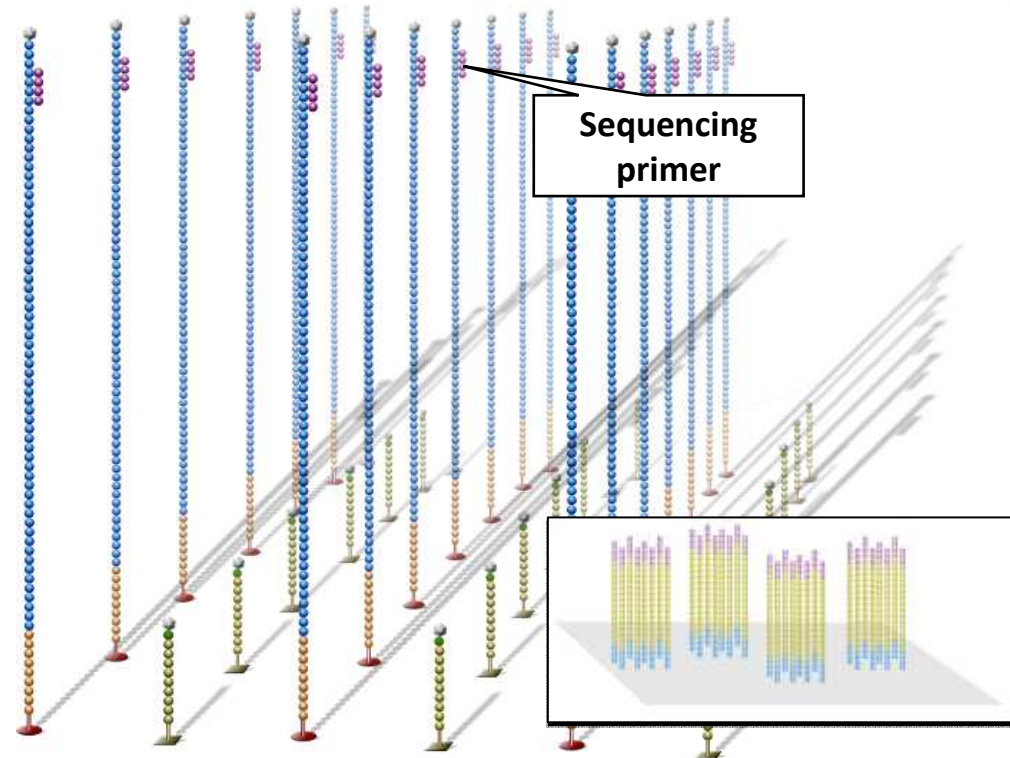
microarray.no



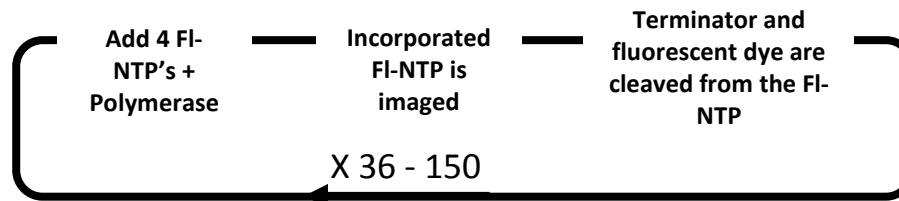
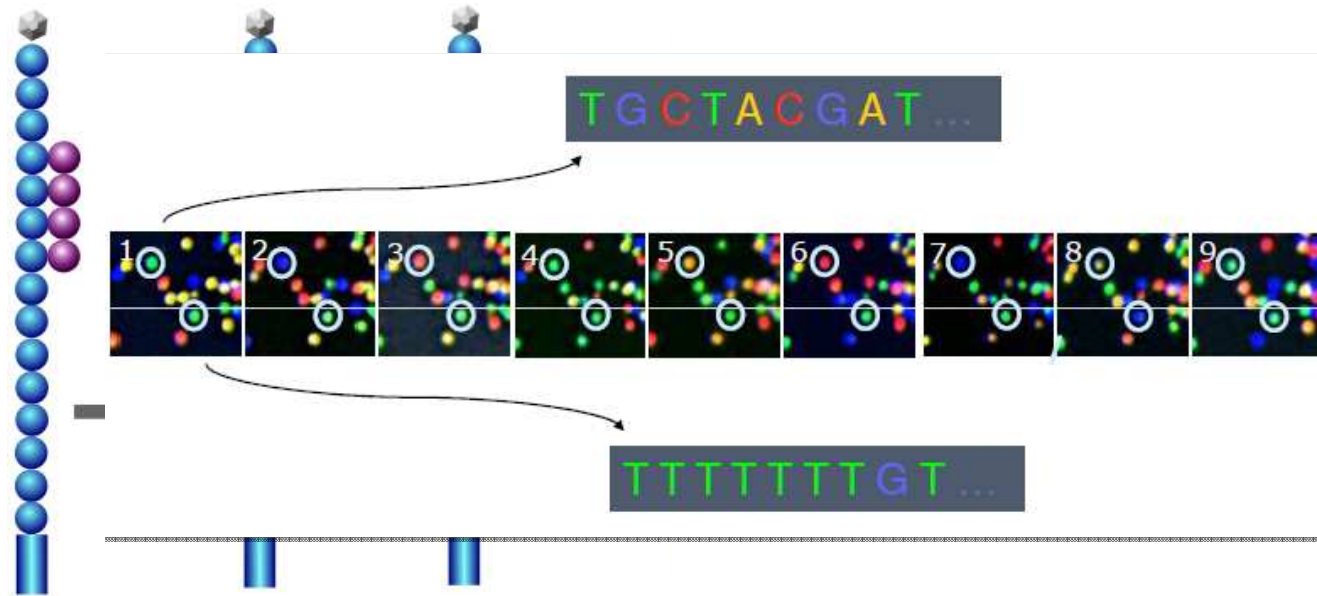
Clustering

Preparing for sequencing

- dsDNA bridges denatured
- Reverse strands cleaved and washed away
- ...leaving a cluster with forward strands only
- Free 3' ends are blocked to prevent unwanted DNA priming
- Sequencing primer is hybridized to adapter sequence



Sequencing by synthesis



microarray.no



SOLiD Sequencing by ligation



Library prep:

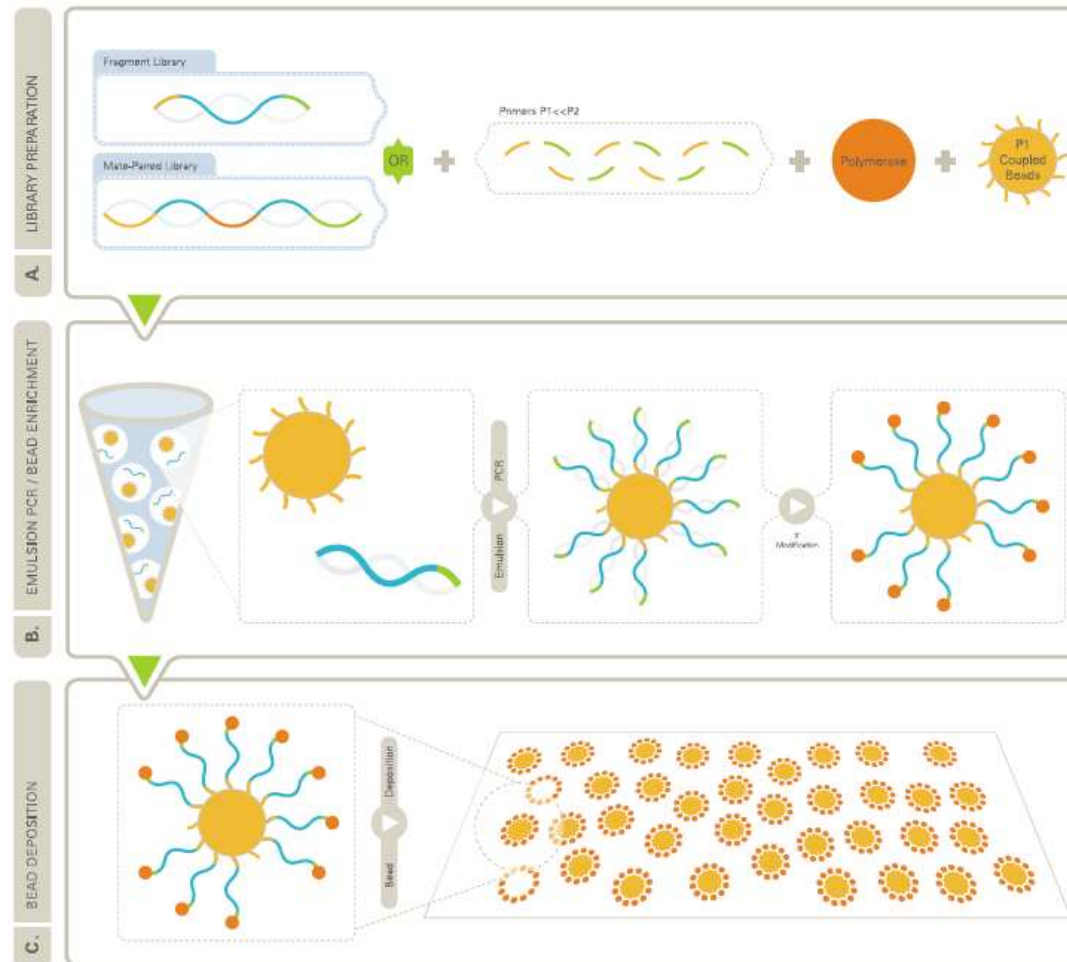
Adaptors are ligated onto the fragmented DNA

Emulsion PCR:

Template is amplified during emulsion PCR and 3' end modified

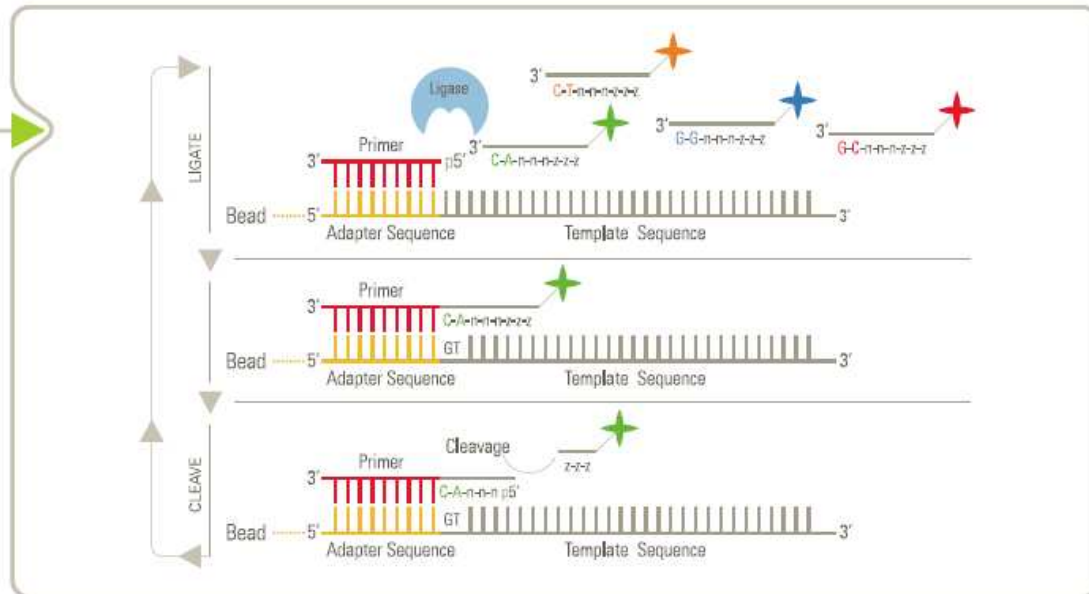
Bead deposition:

Beads are deposited and covalently attached to the Flow Chip



microarray.no





Primer Round	Primer	Read Position																																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
1	Universal seq primer (n) 3'	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
2	Universal seq primer (n-1) 3'	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
3	Universal seq primer (n-2) 3'	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
4	Universal seq primer (n-3) 3'	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
5	Universal seq primer (n-4) 3'	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

• Indicates positions of interrogation Ligation Cycle 1 2 3 4 5 6

D. SEQUENCING BY LIGATION / DATA ANALYSIS

Sequencing by ligation:

- Primer hybridize to adapter sequence
- 4 fluorescently labeled probes compete for ligation, interrogating every 1st and 2nd base in each ligation reaction
- Multiple cycles of ligations, detection and cleavage

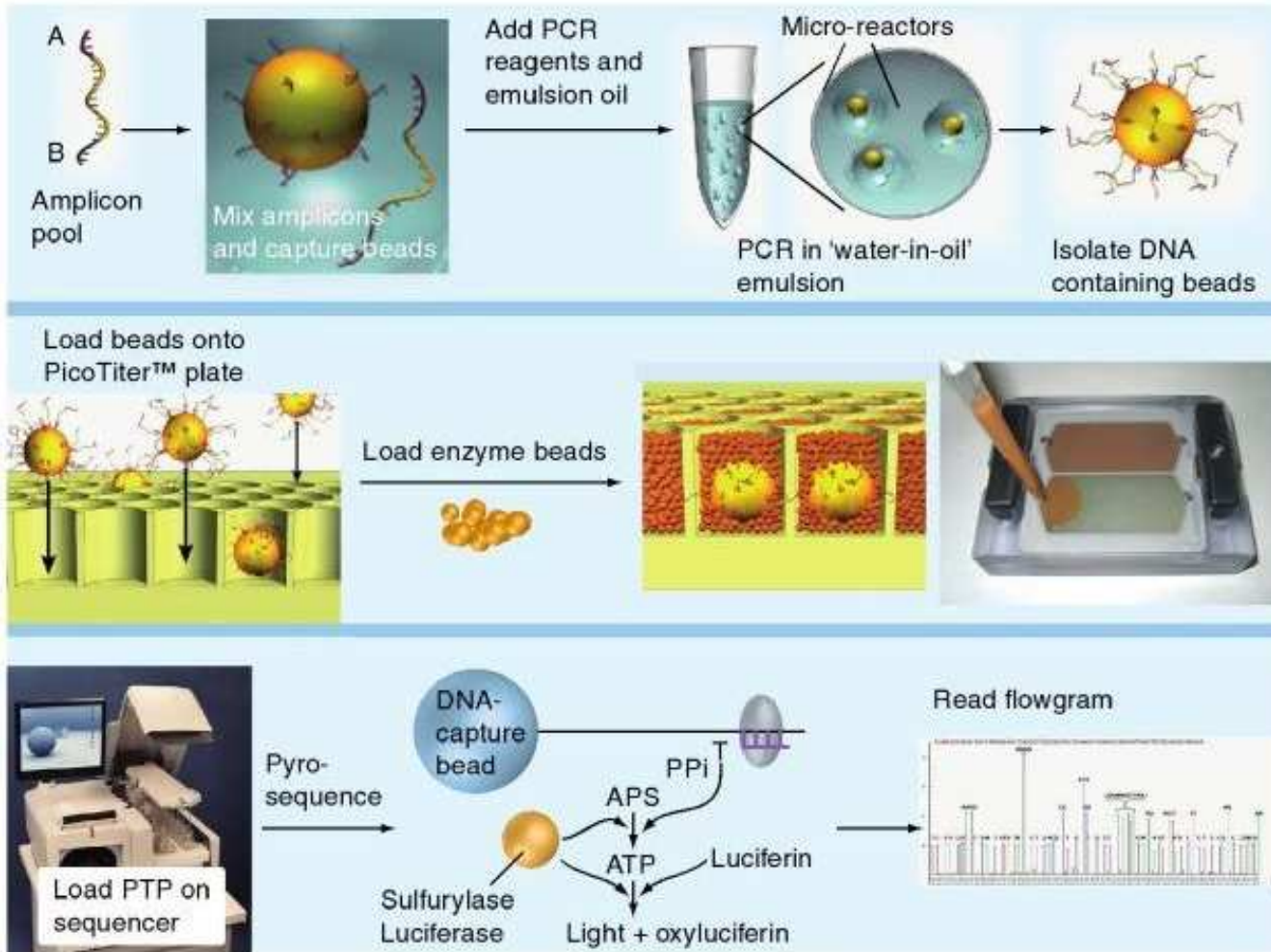
E. DUAL INTERROGATION OF EACH BASE

Primer reset:

- Template is reset with a n-1 primer
- Five rounds of primer sets are needed to complete a template



Pyrosequencing



microarray.no



Different types of libraries

Single- end read: Sequencing a linear fragment from one end

- Counting reads for gene expression
- Harder to align to the reference genome
- Not recommended for SNP calling

Paired- end read: Sequencing a linear fragment from both end
Sequencing larger genomes (de-novo sequencing)

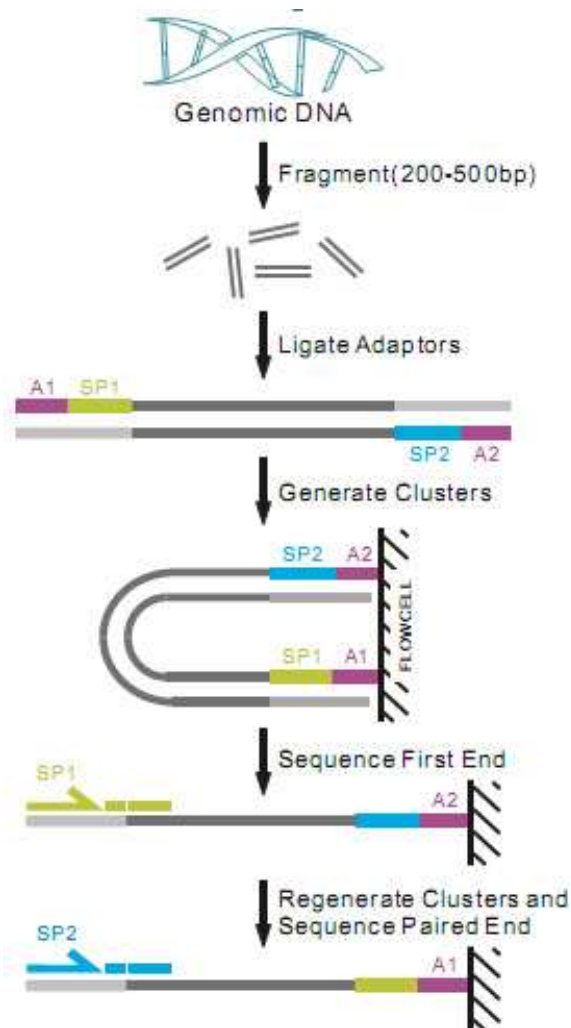
- Makes aligning to reference genome easier
- Easier to discover structural (insertions, deletions, CNVs, inversions and translocations) variation in the genome

Mate- pair libraries: Circular DNA molecules

- Large DNA fragments (1.5 – 6 kb)
- Powerful method for finding large structural events (insertions, deletions, CNVs, inversions and translocations) in the genome
- Sequencing larger genomes (de novo sequencing)



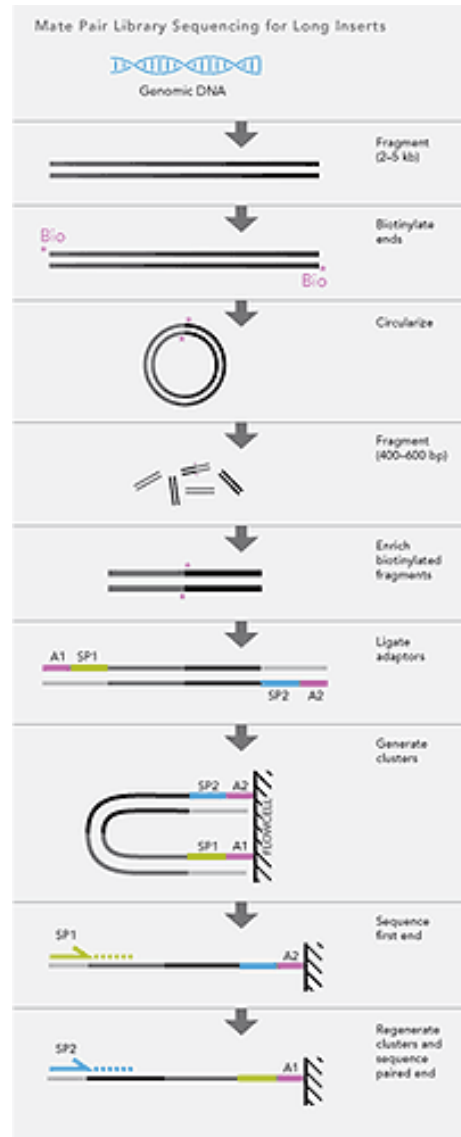
Pair-end sequencing



microarray.no



Mate-pair library



microarray.no

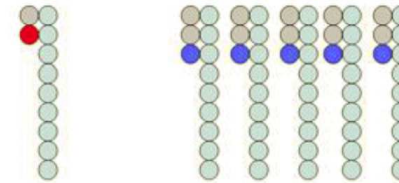


Technological challenges

Phasing/prephasing

Inefficiency in chemistry leads to some clusters lead/lag in incorporation of nucleotides

Phasing



Prephasing



Dye crosstalk

Overlap between dye emission spectra causes A to appear as C and G to appear as T

Solution - PhiX control lane

Dedicated lane used for sequencing PhiX in order to estimate correction parameters for phasing/prephasing, dye crosstalk etc.

microarray.no

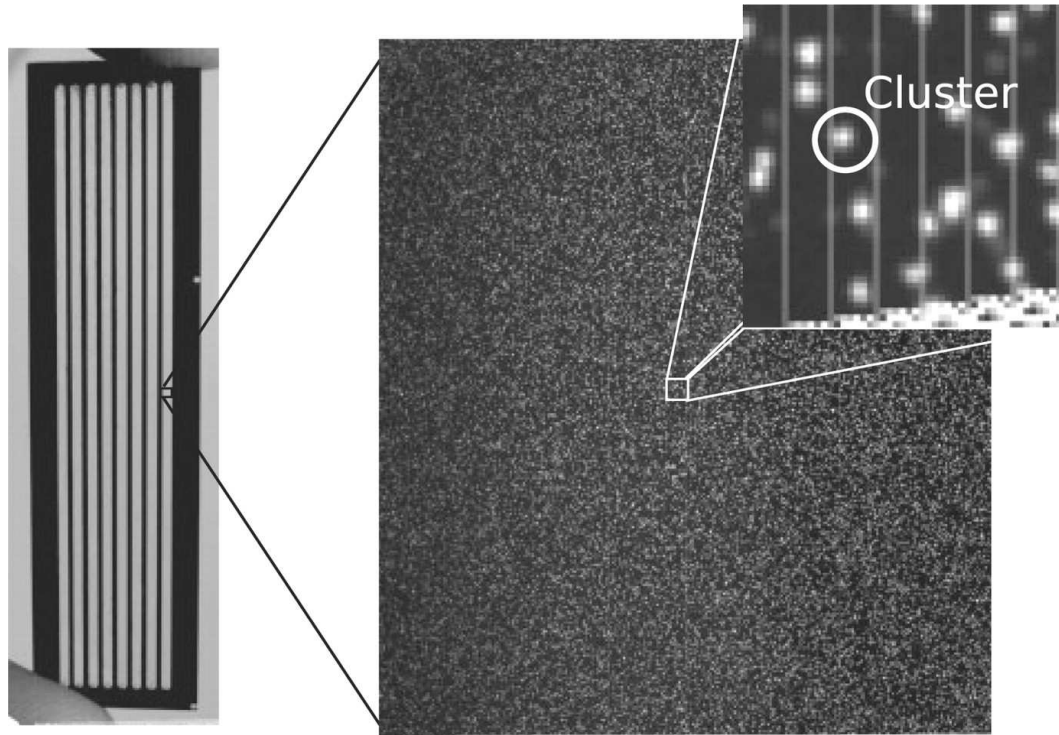


Too close/bright clusters

Too close clusters will look like one cluster, and too bright clusters will camouflage neighbouring clusters

Solution – Purity Filter

Algorithm that removes mixed clusters



microarray.no



Acknowledgements/References

- Tanks to Leonardo Meza-Zepeda (NMC-UiO) for some of the slides

References

- Metzker ML. Sequencing technologies – the next generation. Nature Reviews Genetics 11, 31-46 (2010)
- Zhou X et al. The next-generation sequencing technology and application. Protein Cell 1(6):520-536 (2010)
- Mardis E.R. A decade's perspective on DNA sequencing technology. Nature 470, 198-203 (2011)

