

High Throughput Sequencing applications

microarray.no

Rita Holdhus
December 2011



Topics

- Different sequencing applications
- Bioinformatic challenges
- The future technology
- Personalized genomics
- Ethical challenges

microarray.no



Why use HTS

- View the whole transcriptome
 - Distinguish strand specific expression patterns
 - Detect SNPs at a low coverage
 - Assess globally DNA-protein binding interactions
 - Discover novel transcripts and splice variations
 - Characterize structural rearrangements
-
- Large number of low- cost reads makes the HTS technology useful for many applications



Applications

- De novo sequencing
- Metagenomics
- Epigenetics
 - DNA methylation
 - CHIP-Seq
- Whole genome/targeted resequencing
 - Targeted region specific sequencing
 - Exome sequencing
- RNA sequencing
 - Whole transcriptome

microarray.no



De novo sequencing

- Initial generation of genomic sequence of new species
- Detailed genetic analysis only possible after de novo sequencing has been performed
- Long reads is most useful and paired end reads are essential
- Challenge to assemble new genomes

Table 3 Statistics of eukaryotic genome sequencing projects (data tabulated on March 1, 2010)

organism group	complete	assembly	in progress	sum
animals	4	137	146	287
plants	3	23	85	111
fungi	10	120	93	223
protists	6	49	64	119
total	23	329	388	740





Grape genome

Sanger sequencing and pyrosequencing used to generate a consensus sequence of the grape genome (Velasco et al., 2007)



Giant panda genome

Using high throughput sequencing alone researchers have sequenced 94% of the giant panda genome (Li R. *Nature* 463, 311-317, 2010)

microarray.no



Metagenomics

- Genomic analysis of microorganisms from microbial communities

Global ocean sampling expedition

Craig Venter's expedition around the world collecting metagenomic samples. Pilot project in the Sargasso Sea found DNA from nearly 2000 different species, including 148 bacteria never seen before

Human Microbiome Project (HMP)

Mapping microbial communities associated with different parts in the human body and to analyze the role of these microbes in human health and disease.



microarray.no

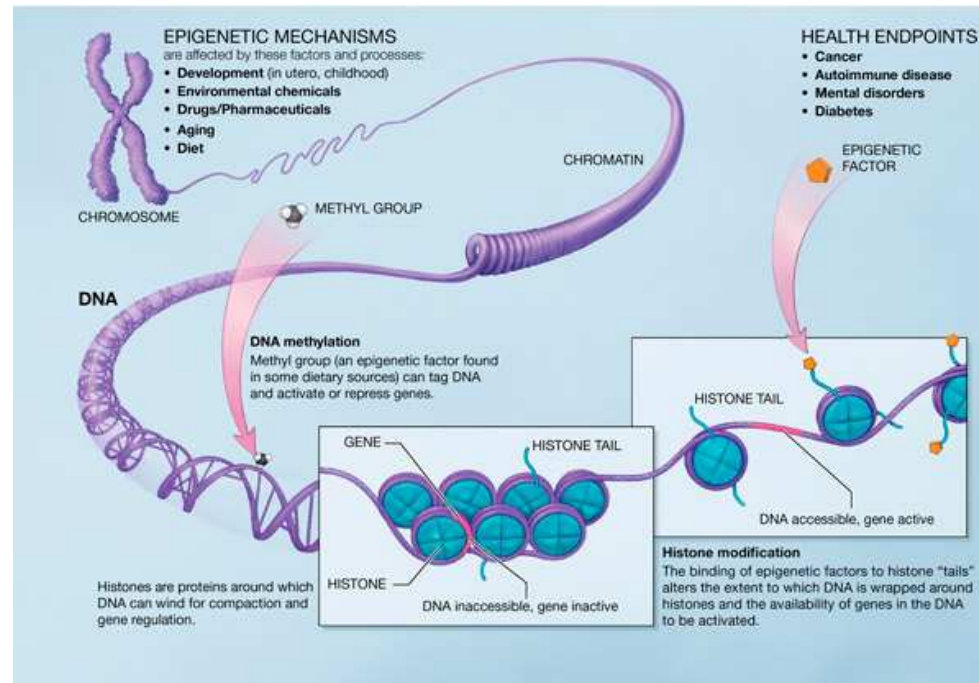


Epigenetics

- Epigenetics is the study of heritable gene regulation that does not involve the DNA sequence itself but modifications and higher-order structures
- HTS has been applied in several epigenomic areas such as DNA methylation, posttranslational modifications of histones and binding of transcription factors

DNA methylation

chIP-Seq

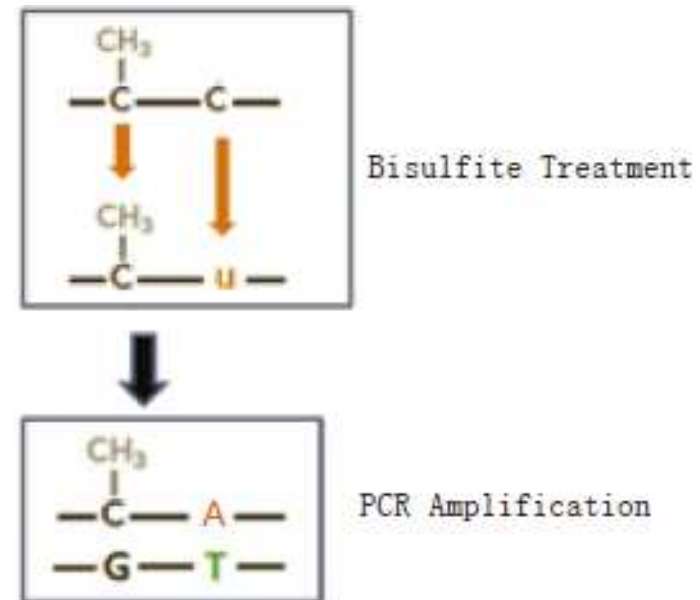


microarray.no



Genome-wide methylation Bisulfite sequencing

- Single- base resolution can be achieved by sodium bisulfite treatment of genomic DNA
- Converts cytosines, but not methylcytosines, to uracil
- Methylcytosine will be sequenced as cytosine, and unmethylated cytosine as thymine.



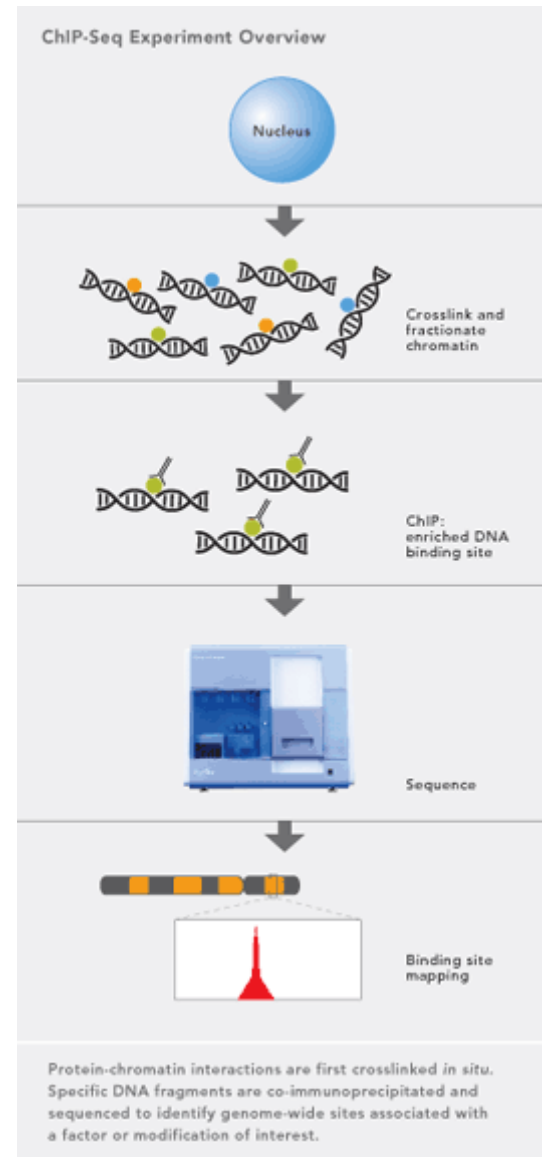
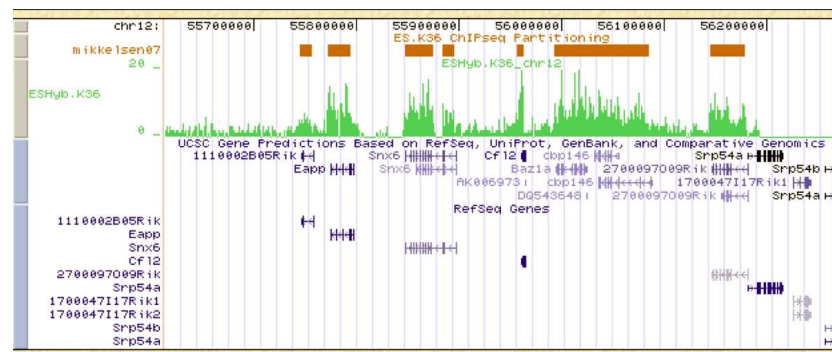
microarray.no

chIP-Seq

A technique for genome-wide profiling of DNA binding proteins, histone modifications and nucleosomes

Cromatin ImmunoPrecipitation (IP)

Enrich specific DNA-protein complexes using an antibody against the protein of interest.



microarray.no



Whole- genome resequencing

- Sequencing of the entire genome of a already sequenced specie

Table 5 Sequencing statistics of six individual human genomes

platform	individual	No. of reads (millions)	read length (bases)	read coverage	genome coverage (%)	SNPs (millions)	No. of runs	estimated cost (US\$)	references
Sanger	J. Craig Venter	31.9	800	7.5 ×	N/A	3.21	>340,000	70,000,000	Levy et al., 2007
Roche 454	James D. Watson	93.2	250	7.4 ×	95	3.32	234	1,000,000	Wheeler et al., 2008
SOLiD	James R. Lupski	238	35	29.6 ×	99.8	3.42	3	75,000	Lupski et al., 2010
Illumina Solexa	Yoruba male (NA18507)	3681	35	40.6 ×	99.9	4	40	250,000	Pushkarev 2009
	Han Chinese male (YH)	2950	35	36 ×	99.9	3.07	35	500,000	Wang et al., 2008
	Korean male (SJK)	1647	35, 74	29.0 ×	99.9	3.44	15	250,000	Ahn et al., 2009
	Korean male (AK1)	1910	36, 88, 106	27.8 ×	99.8	3.45	30	200,000	Kim et al., 2009
Helicos	Stephen R. Quake	2725	32	28 ×	90	2.81	4	48,000	Pushkarev et al., 2009



Targeted resequencing



Exome sequencing

Many potential disease-causing variants lie within the exonic regions of the genome.

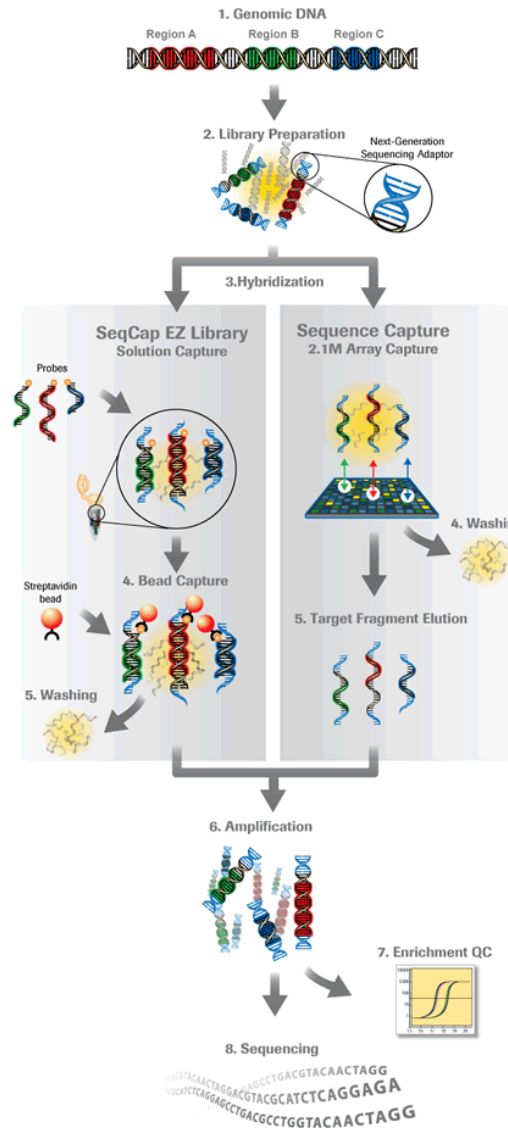
Allow us to sequence more persons with a higher sequence depth per individual

Limitations

Capture probes can only target known exons

Efficiency of capture probes varies

Some probes fail to capture target



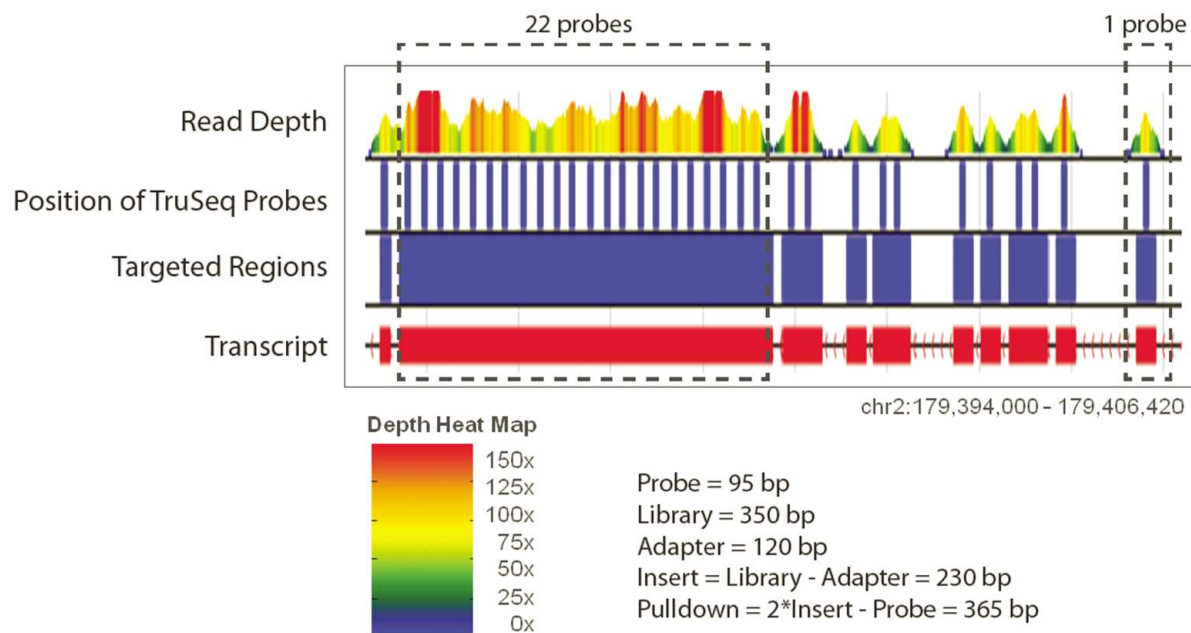
microarray.no



Exome Sequencing coverage

Some regions are difficult to sequence due to repetitive sequences, high GC-content etc.

→ Variable sequence coverage in targeted regions

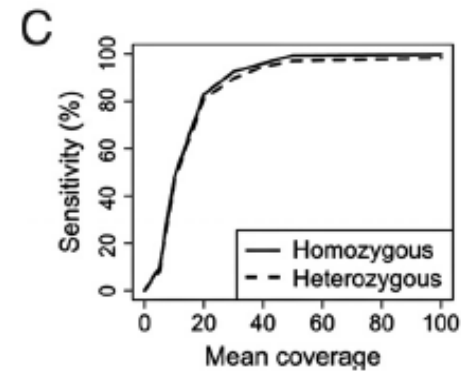


Exome Sequencing coverage

% Bases Covered	Mean Normalized Coverage	Mean Sequencing Coverage Required (at 10x desired coverage)
90	0.2	50x
75	0.4	25x
50	0.8	12.5x

Sensitivity to detect variants

Increases steeply from 5X to 20X, then more gradually and platous at around 50X

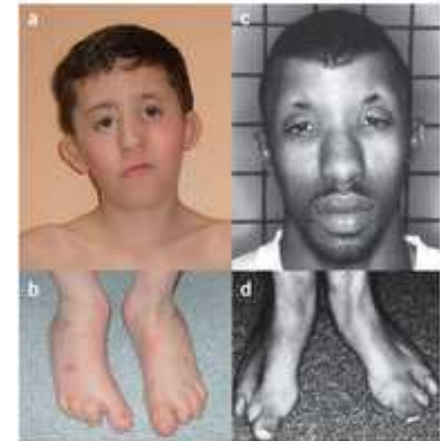


Choi M. et al. PNAS (2009)



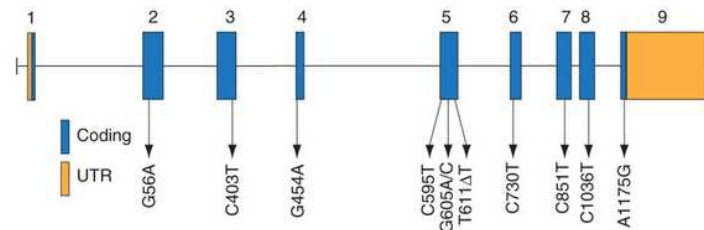
Exome sequencing Miller syndrome

- Exome sequencing were used to discover the gene that causes the very rare mendelian disorder, Miller syndrome
- Four affected individuals were exome sequenced



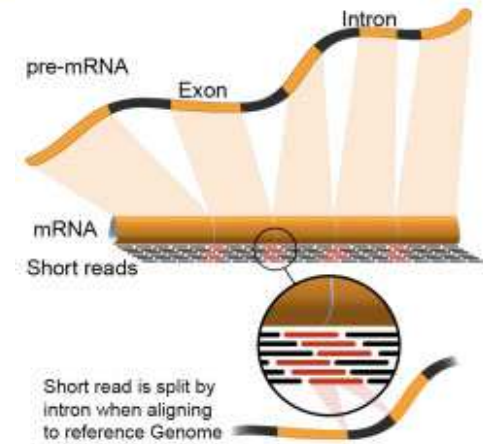
microarray.no

→ DHODH (encodes a key enzyme in the de novo pyrimidine biosynthesis pathway) discovered to be the disease causing gene



RNA sequencing

- Quantification of the transcriptome
- Characterization of RNA transcripts
 - Transcription Start Sites (TSS)
 - Strand-specific measurements
 - Gene fusion detection
 - small RNA characterization
 - Alternative splicing detection



microarray.no



Quantifying the transcriptome

- Sequencing the whole transcriptome in order to quantify expressed RNA molecules
- Not limited on quantifying transcripts from known genomic sequence
- Detect novel transcripts
- Higher dynamic range than microarrays (no upper limit for quantification)
- Very low background signals

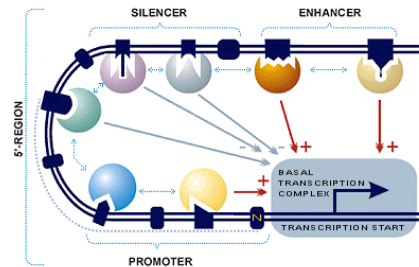
microarray.no



Characterization of RNA transcripts

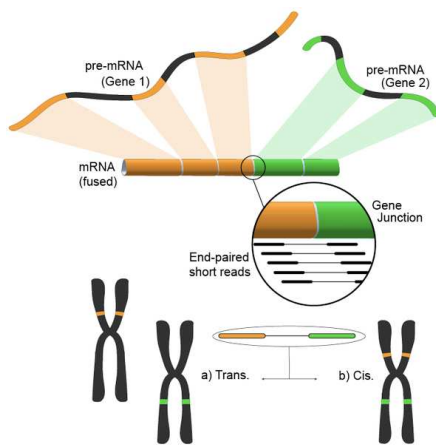
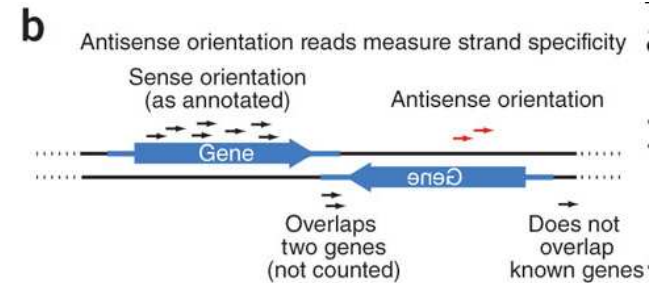
Transcription Start Sites (TSS)

Identify promoter regions that regulates gene expression of each transcript



Strand-specific RNA-seq

Identify antisense transcripts with potential regulatory roles



Gene fusion

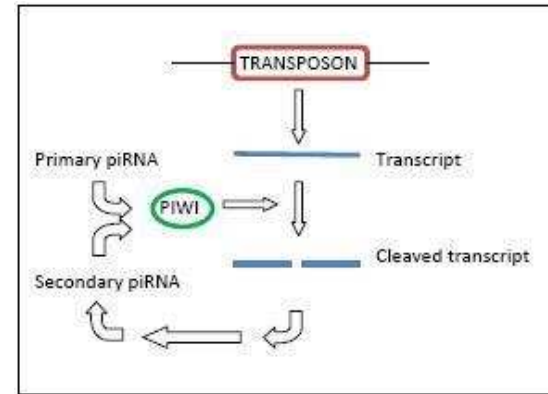
Reads mapping to exon-exon junction where the exons come from different genes

Characterization of RNA transcripts

small RNA

Identify novel small RNAs and profiling of small RNA genes.

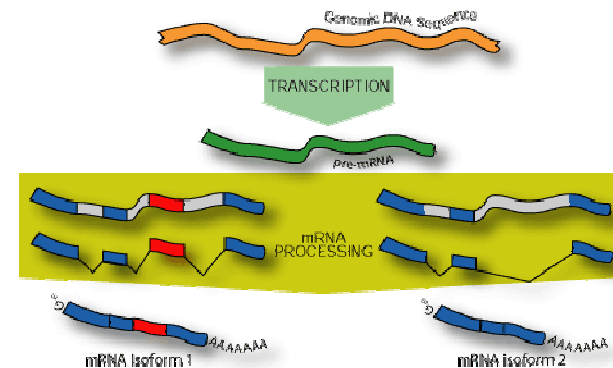
No absolute quantification of small RNAs
(Linsen et al. Nature Methods, 2009)



Alternative splicing

Using long reads or pair-end reads it is possible to characterize alternative splicing patterns in genes

A recent study found more than 95% of human multi-exon genes had alternative spliceforms



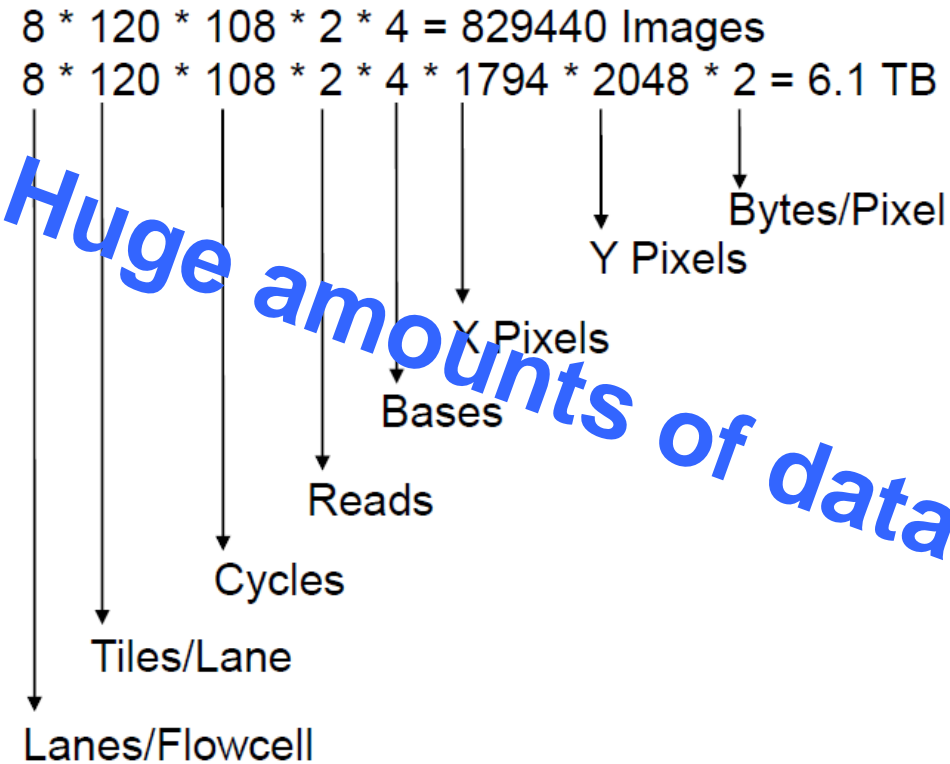
Sequencing vs. Array

Sequencing	Microarray
Higher dynamic range	Limited to signal intensity
Need no reference sequence	Need reference sequence
Possible to multiplex samples	No multiplexing
Still expensive	Cheaper

microarray.no



Bioinformatic challenges



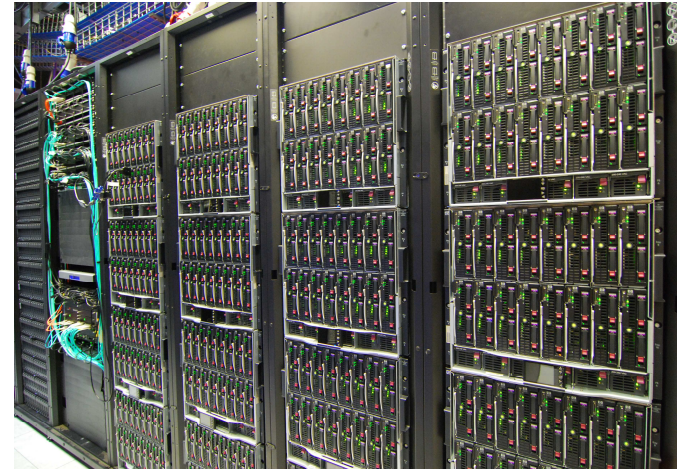
microarray.no



Bioinformatic challenges

Data

- Very large files
- Need large amount of data power (CPUs)
- Storage
- Data security/privacy issues (human samples)



microarray.no

Analysis tools

- Many tools freely available, but steep learning curve
- Programs for "the biologist" exists, but not as good as open source programs?

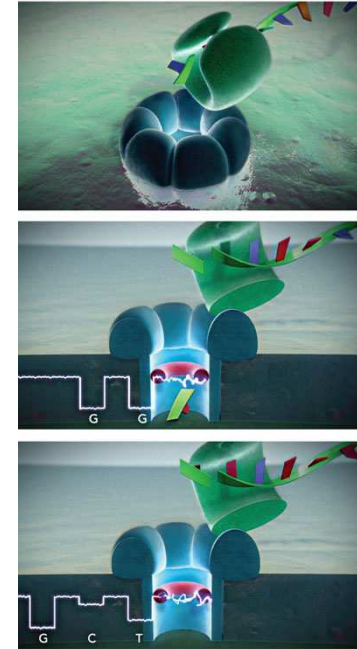


"Future"

Third-generation sequencing technology

The technology is already here!

Sequencing single-molecules in hours

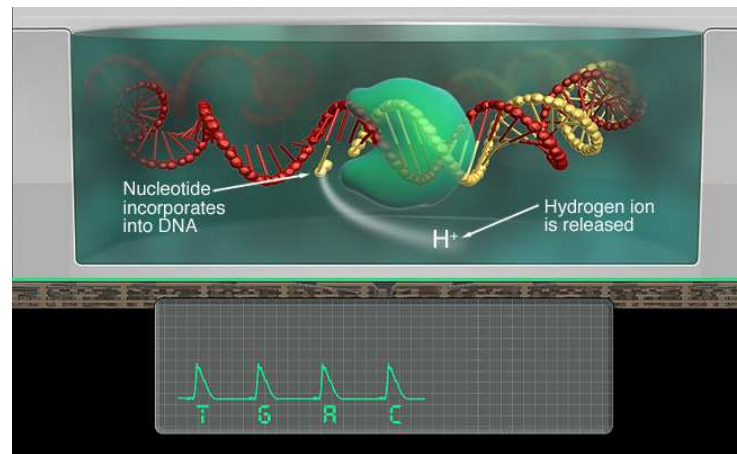


Oxford Nanopore

microarray.no

Sequencing in diagnostics

Sequence custom gene/exon panels in just hours




Personalized genomics



"your genes are a road-map to better health"



Start filling in the gaps with your DNA

 "Because I had given my doctor information from 23andme, he got to a diagnosis much faster. 23andme saved my life." Kirk C.

\$99* **Our new low price for all!**
Was \$199

Order Now »

*Requires a 1-year commitment to the Personal Genome Service® at \$9/mo. **Order for \$399** without commitment.

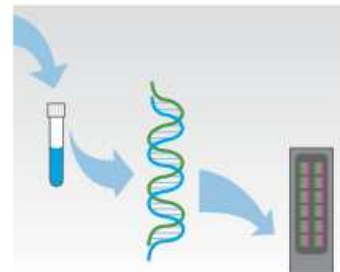
microarray.no



1. Order a kit from our [online store](#).



2. Register your kit, spit into the tube, and send it to the lab.



3. Our CLIA-certified lab analyzes your DNA in 6-8 weeks.



4. Log in and start exploring your genome.



Ethical challenges



Biotechnology regulation in Norway

Sequencing of all genes is considered a predictive procedure since analysis of such data can give unintended findings and information about future diseases for the patient

Predictive procedures are regulated by the "bioteknologiloven"

This law demands written consent and genetic counseling before, during and after the predictive gene testing

This situation raises several challenges!

microarray.no

