

Next Generation Sequencing: Overview and Challenges

Nina Norgren, NBIS
Göteborg, May 2019

Slides adapted from:

Olga Vinnere Pettersson, PhD

National Genomics Infrastructure hosted by ScilifeLab,

Uppsala Node (UGC)



SciLifeLab

SciLifeLab

TECHNOLOGIES & SERVICES ▼

RESEARCH ▼

EDUCATION ▼

COLLABORATION ▼

Find more information and search for what you need on the page for Technologies & Services

What is the difference between national and regional facilities?

Search for Technologies & Services

National facilities

Affinity Proteomics

Biobank Profiling
Cell Profiling
Fluorescence Tissue Profiling
PLA Proteomics
Protein and Peptide Arrays
Tissue Profiling

Bioimaging

Advanced Light Microscopy
Fluorescence Correlation Spectroscopy

Bioinformatics

Bioinformatics Compute and Storage (UPPNEX)
Bioinformatics Long-term Support (WABI)
Bioinformatics Short-term Support and Infrastructure (BILS)

Chemical Biology Consortium Sweden

Laboratories for Chemical Biology Umeå (LCBU)
The Laboratories for Chemical Biology at Karolinska Institutet (LCBK1)
Uppsala Drug Optimization and Pharmaceutical Profiling (UDOPP)

Clinical Diagnostics

Clinical Biomarkers
Clinical Genomics
Clinical Sequencing

Drug Discovery and Development

ADME (Absorption Distribution, Metabolism Excretion) of Therapeutics (UDOPP)
Biochemical and Cellular Screening
Biophysical Screening and Characterization
Human Antibody Therapeutics
In Vitro and Systems Pharmacology
Medicinal Chemistry – Hit2Lead
Medicinal Chemistry – Lead Identification
Protein Expression and Characterization

Functional Genomics

Karolinska High Throughput Center (KHTC)

National Genomics Infrastructure

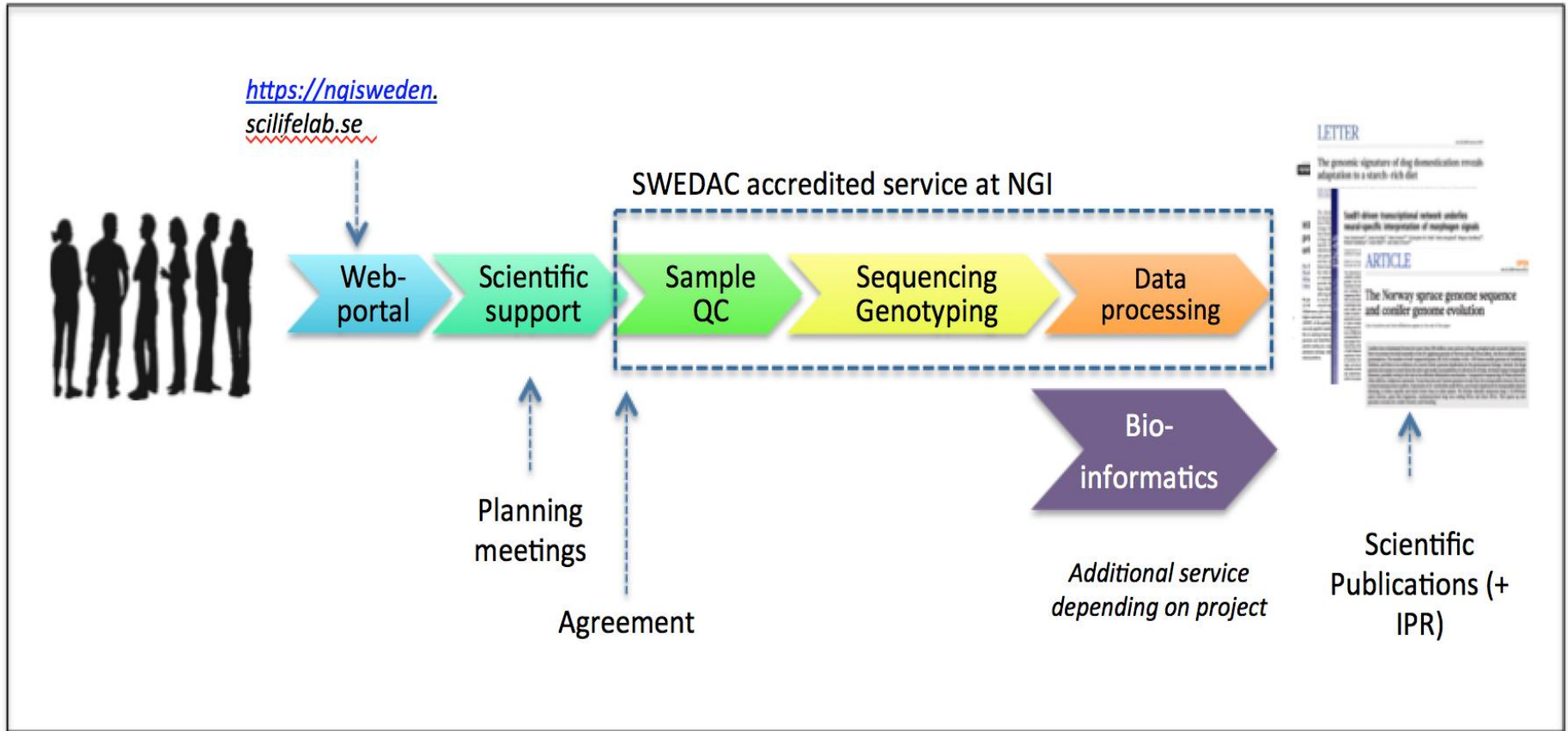
NGI Stockholm (Genomics Applications)
NGI Stockholm (Genomics Production)
NGI Uppsala (SNP&SEQ Technology Platform)
NGI Uppsala (Uppsala Genome Center)

Structural Biology

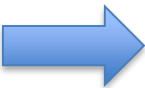
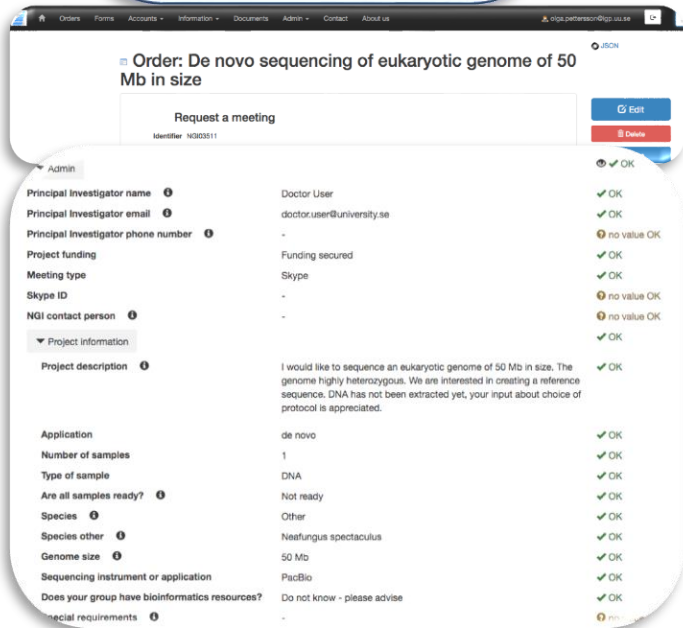
Protein Science Facility



Project handling at NGI



How does a project go? Project request



Short History of NGS

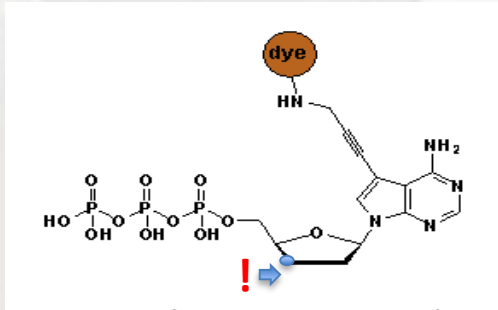
Once upon a time...

- Fredrik Sanger and Alan Coulson
Chain Termination Sequencing (1977)

Nobel prize 1980

Principle:

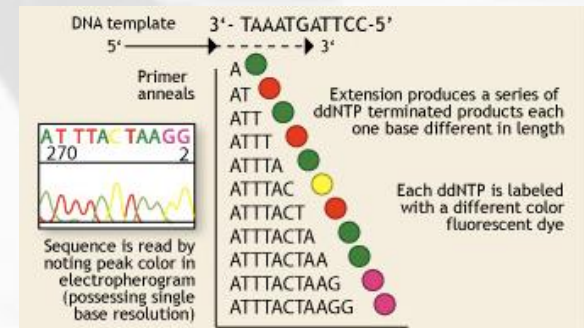
SYNTHESIS of DNA is randomly **TERMINATED** at different points
Separation of fragments that are 1 nucleotide different in size



Lack of OH-group at 3' position of deoxyribose

1 molecule sequenced at a time = 1 read

Capillary sequencer: 384 reads per run



2006: NGS was born

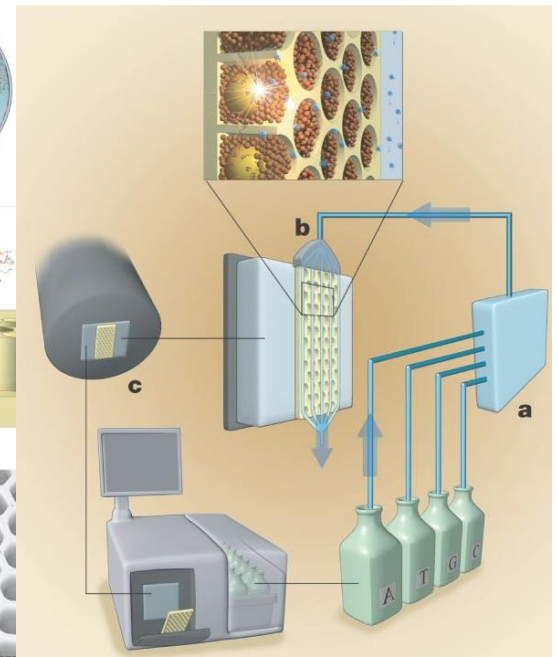
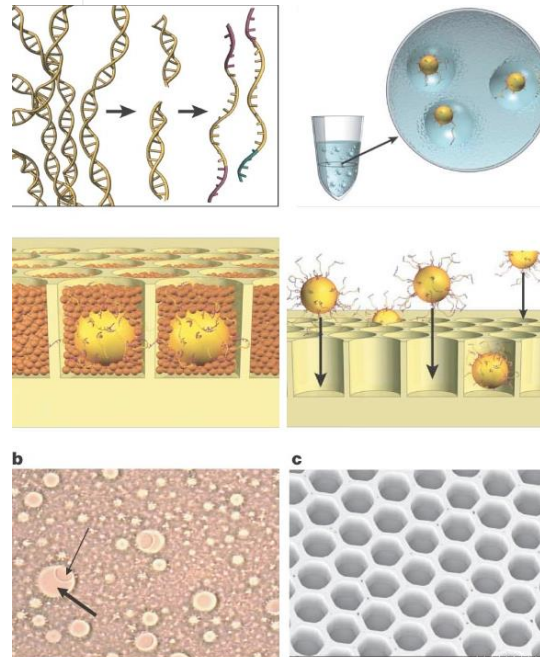


Thousands of molecules sequenced in parallel

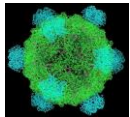
1 mln reads sequenced per run



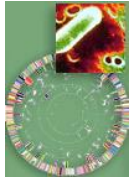
Roche 454 GS FLX



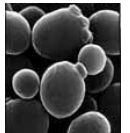
Since the beginning of Genomics:



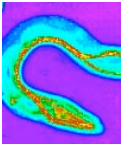
First genome: virus ϕ X 174 - 5 368 bp (1977)



First organism: *Haemophilus influenzae* - 1.5 Mb (1995)



First eukaryote: *Saccharomyces cerevisiae* - 12.4 Mb (1996)



First multicellular organism: *Cenorhabditis elegans* - 100 MB (1998-2002)



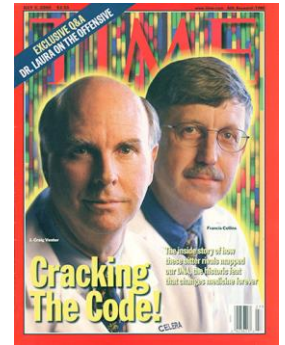
First plant: *Arabidopsis thaliana* - 157 Mb (2000)

... prices go down

Human genome sequencing:

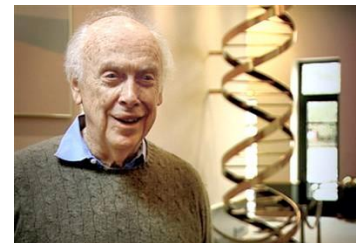
2004: Genome of Craig Wenter costs 70 mln \$

- Sanger's sequencing



2007: Genome of James Watson costs 2 mln \$

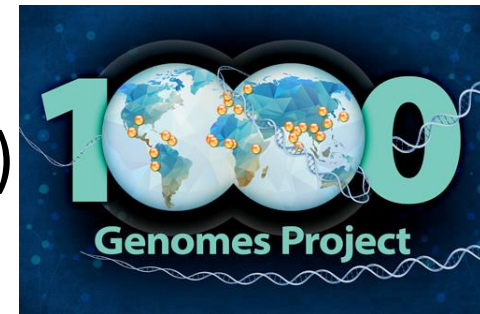
- 454 pyrosequencing



2014: Ultimate goal: 1000 \$ / individual

2016: Illumina Xten: Almost there! (1200 \$)

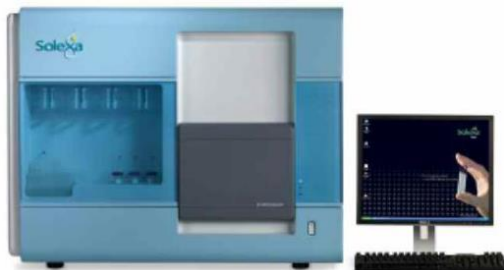
2017: NovaSeq: "Hold my beer..." (100 \$)



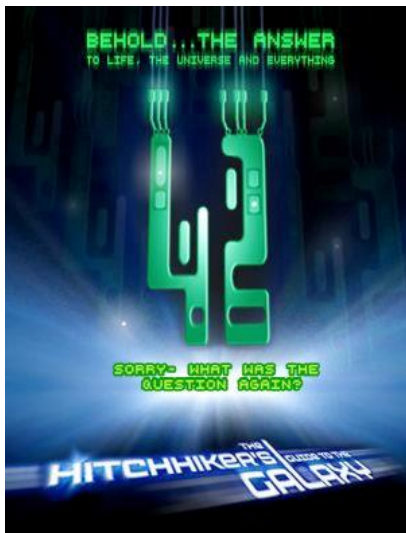


... paradigm changes

- From single genes to complete genomes
- From single transcripts to whole transcriptomes
- From single organisms to complex metagenomic pools
- From model organisms to the species you are studying
- Personal genome = personalized medicine



... scientific value diminishes



Science 5 September 1997:
Vol. 277 no. 5331 pp. 1453-1462
DOI: 10.1126/science.277.5331.1453

IF 31.6

[< Prev](#) | [Table of Contents](#) | [Next >](#)

ARTICLES

The Complete Genome Sequence of *Escherichia coli* K-12

Frederick R. Blattner^{*}, Guy Plunkett III^{*}, Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau and Ying Shao

Journal of Biotechnology
Article in Press, Corrected Proof - Note to users

IF 2.9



doi:10.1016/j.jbiotec.2010.12.018 | [How to Cite or Link Using DOI](#)

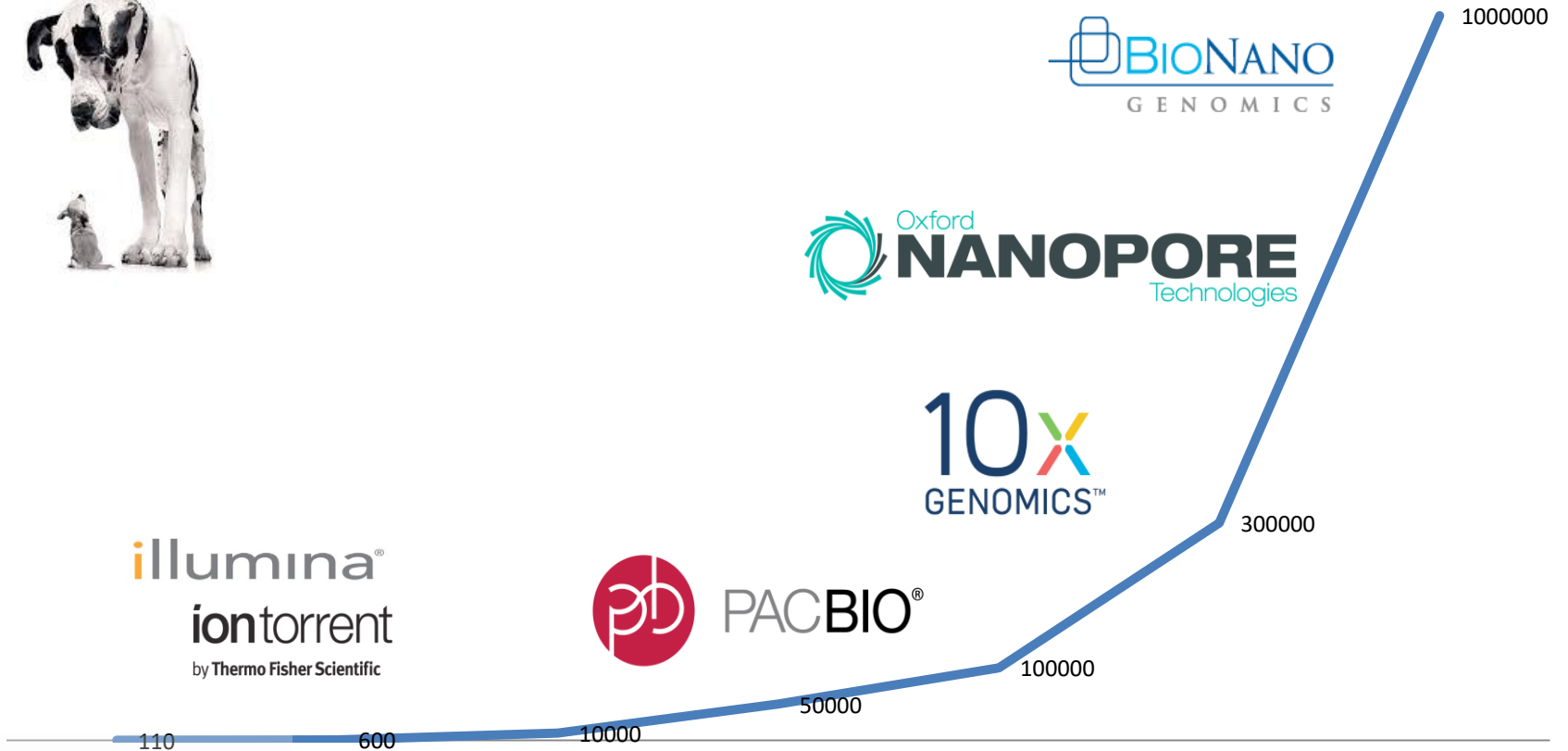
[Permissions & Reprints](#)

The complete genome sequence of the dominant *Sinorhizobium meliloti* field isolate SM11 extends the *S. meliloti* pan-genome

Susanne Schneider-Bekel^a, Daniel Wibberg^a, Thomas Bekel^b, Jochen Blom^b, Burkhard Linke^b, Helko Neuweger^b, Michael Stiens^{a, c}, Frank-Jörg Vorhölter^a, Stefan Weidner^a, Alexander Goesmann^b, Alfred Pühler^a and Andreas Schlüter^a, , 

Current Technologies

Read length



Illumina

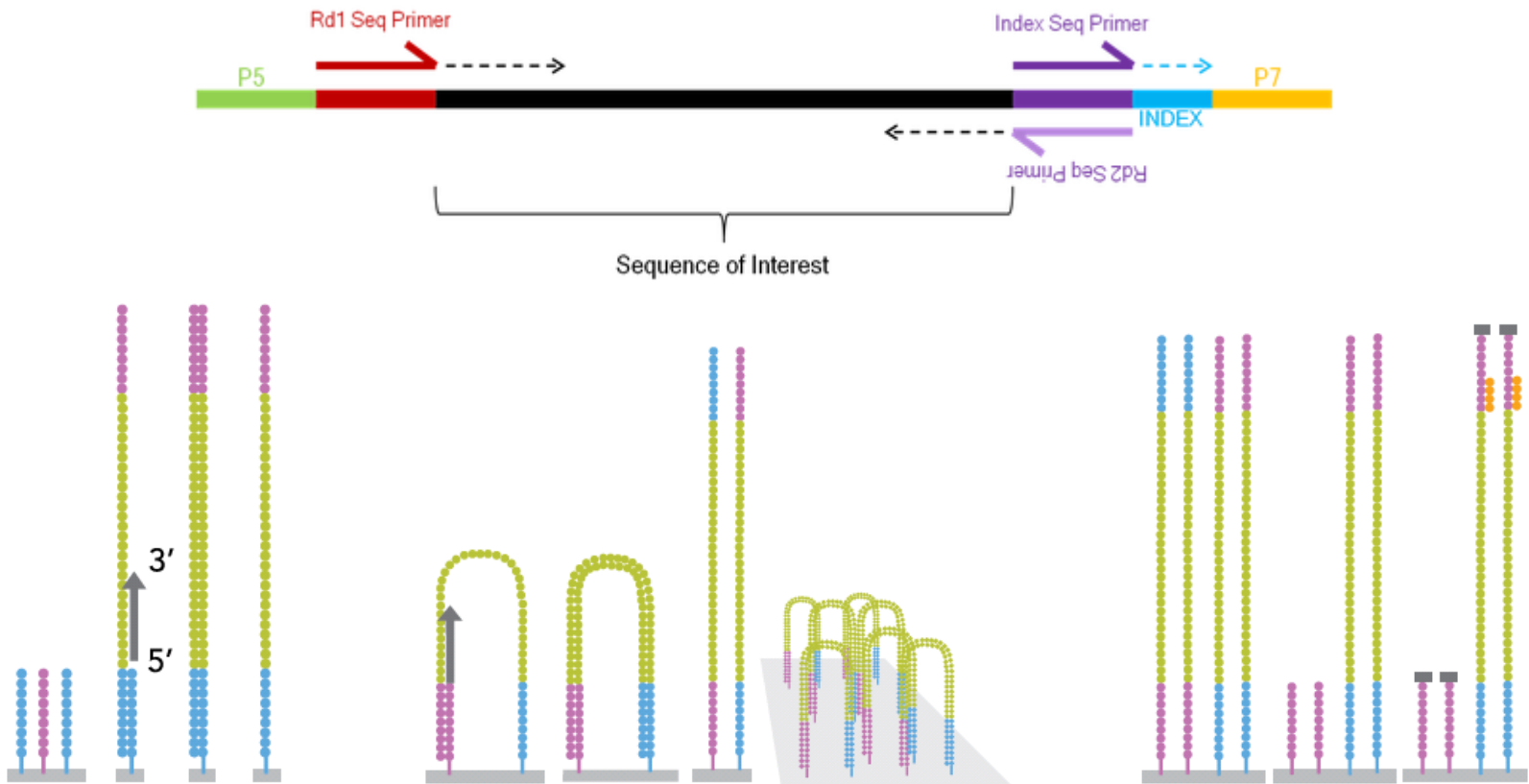
Instrument	Yield and run time	Read Length	Error rate	Error type
HiSeq2500	120 Gb – 600 Gb 27h or standard run	110x110 (250x250)	0.1%	Subst
MiSeq	540 Mb – 15 Gb (4 – 48 hours)	up to 350x350	0.1%	Subst
HiSeqXten	800 Gb - 1.8 Tb (3 days)	150x150	“	“
NovaSeq 6000	250 Gb – 3 Tb	150x150	“	“

Main applications

- Whole genome, exome and targeted reseq
- Transcriptome analyses
- Methylome and ChIPSeq
- Rapid targeted resequencing (MiSeq)
- Human genome seq (Xten)



Illumina: bridge amplification

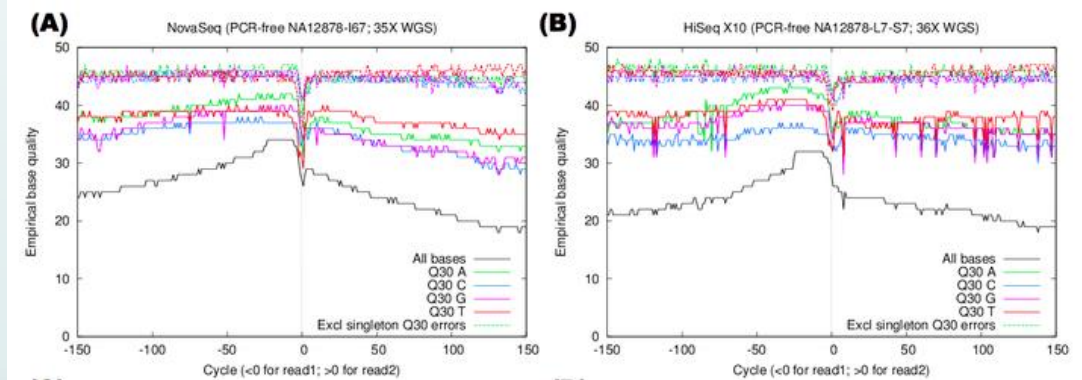


<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

NovaSeq 6000



- NGI has five instruments
- Flexible and scalable using multiple flow cell types
- Quick and easy operation using RFID labeled reagent cassettes
- Onboard clustering and automatic washing minimises hands on time during runs
- 2 color chemistry T=Green
C=Red
A=Green/Red
G=no signal

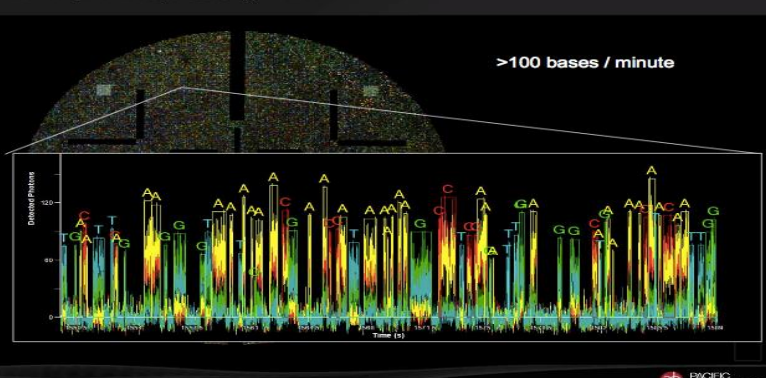


PacBio

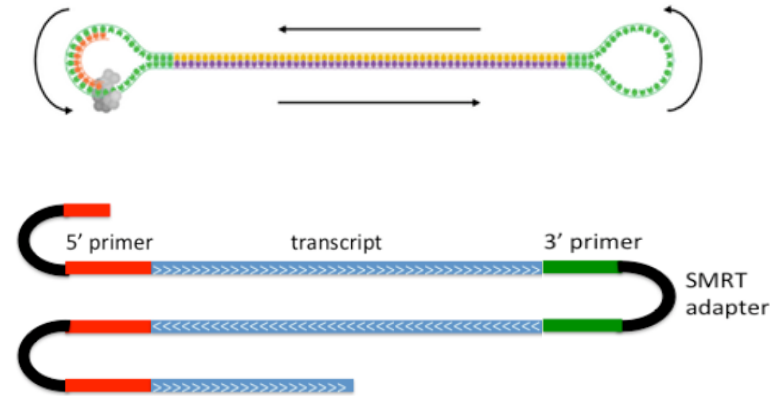
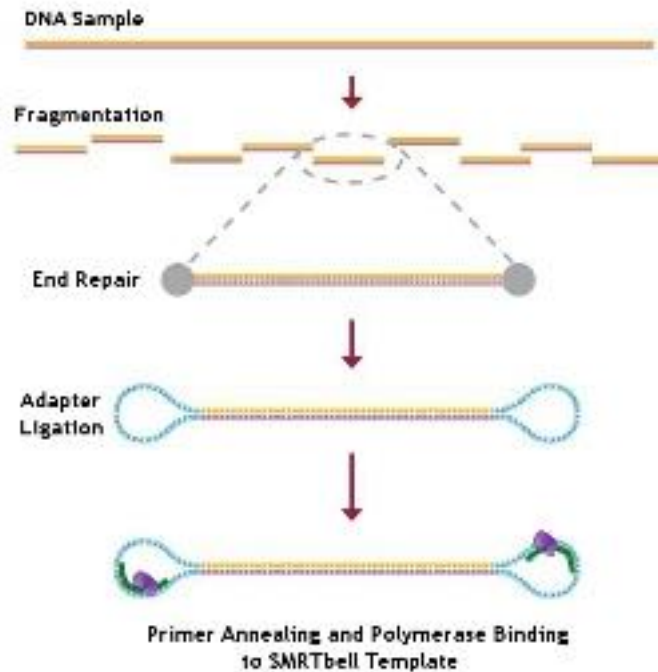
Instrument	Yield/cell and run time	Read Length	Error rate	Error type
RSII	250 Mb – 1.8 Gb 30 - 600 min	250 bp – 60 kb (78 kb)	15 % (single pass) 0.0001% (circular consensus)	Indels, random
SEQUEL	2-14 Gb 30-2400 min	250 bp – 80 kb (160 kb)	as RSII	Indels, random

Single-Molecule, Real-Time DNA sequencing

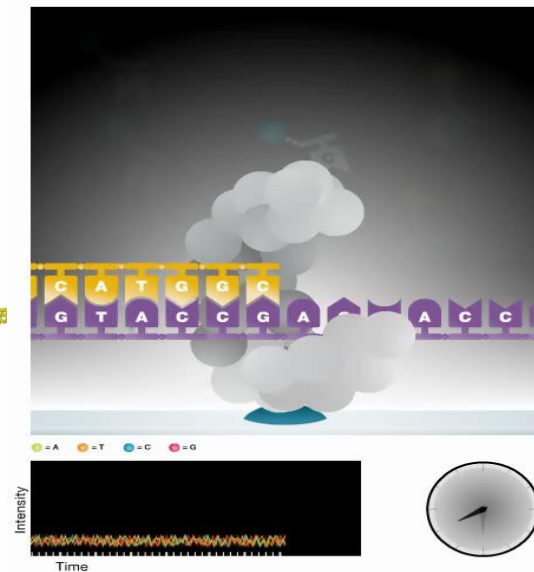
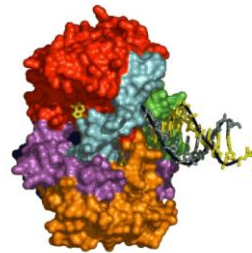
Example Sequencing Run



PacBio: SMRT - technology



SMRT =
Single Molecule Real Time



SMRT sequencing: common misconceptions

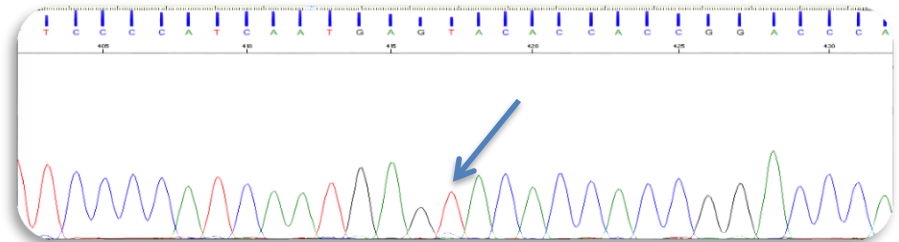
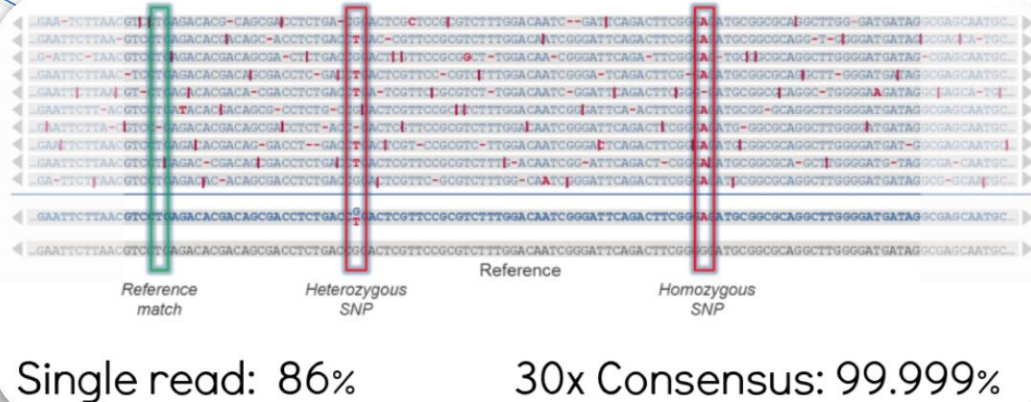
High error rate?

Irrelevant, because errors are random

Depending on coverage

Examples:

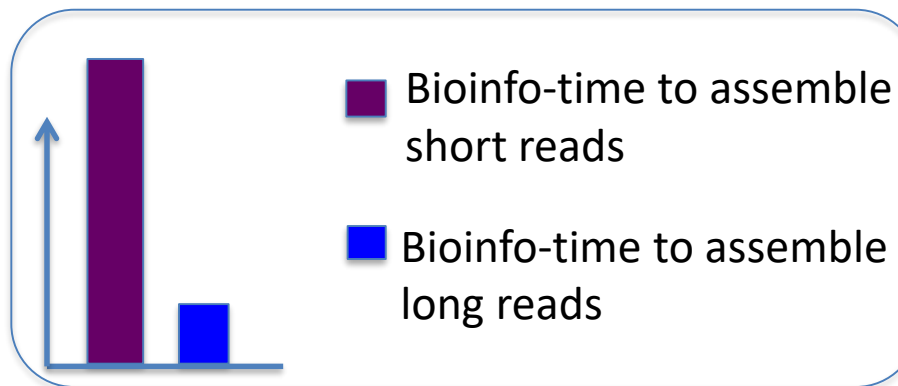
- 8 Mb genome, 8 SNPs detected
- 65 kb construct: 100% correct sequence
- Detection of low frequency mutations



High price?

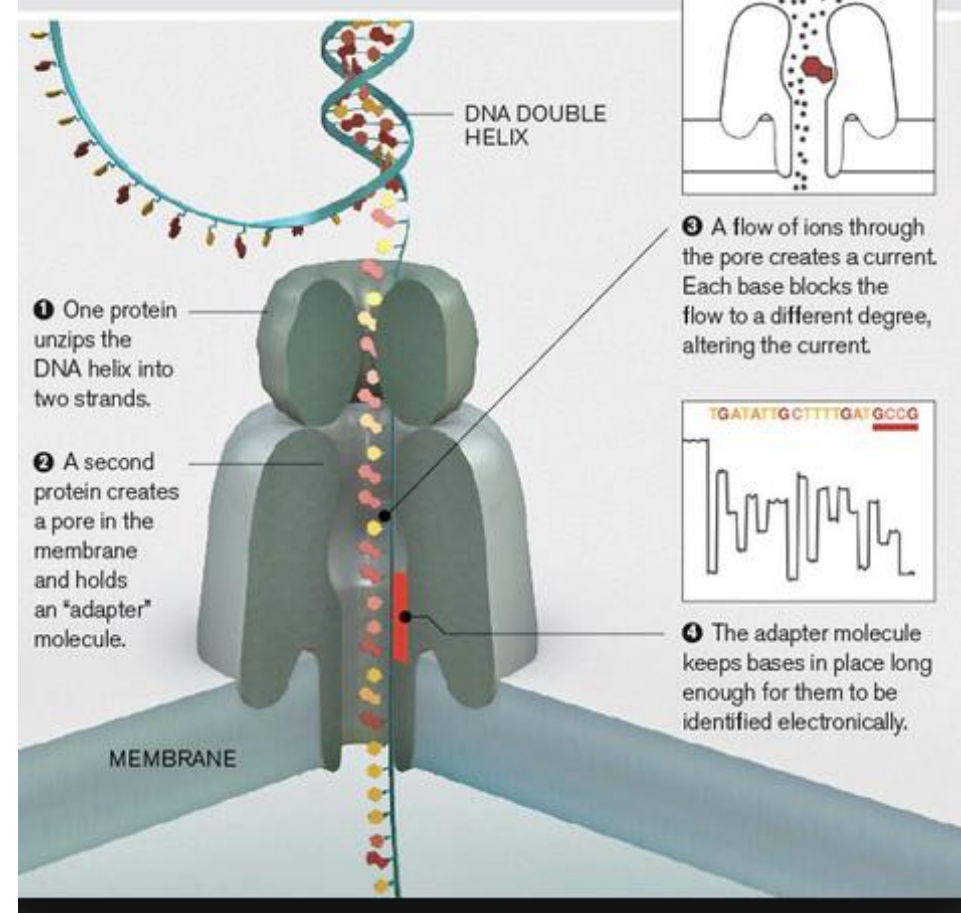
Not for small genomes

Better assembly quality
Single-molecule reads without PCR-bias



Oxford Nanopore

Flow Cells run in parallel	Yield - run time
MinION (1)	1 – 10 Gb / cell
GridION (5)	5 – 50 Gb / 5 cells
PromethION (12 - 24 - 48)	20 – 100 Gb / cell

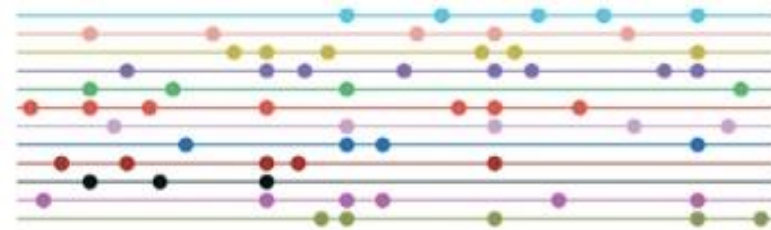
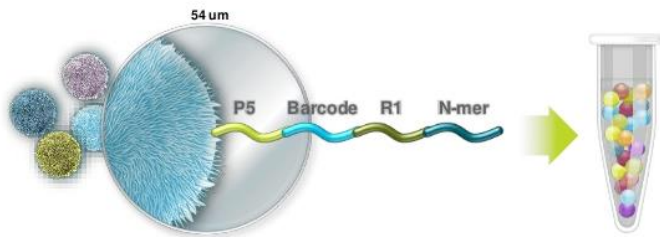
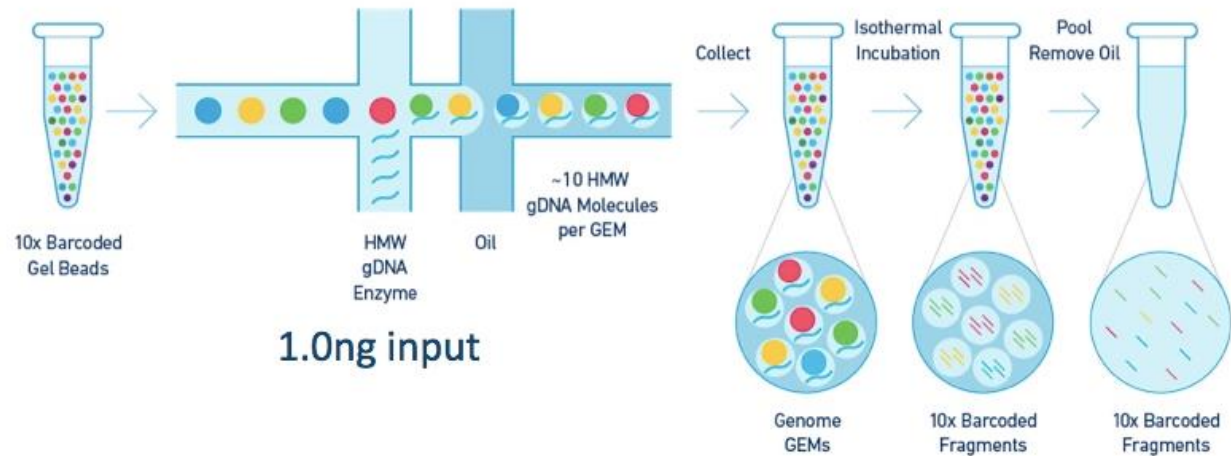


Reads up to 6-8 Gb
 10-15% error rate
 Life time 5 days

Longest reads:
 beyond 1 Mb



10x Genomics (Chromium)



Fragment length: 50 kb – 100+ Kb

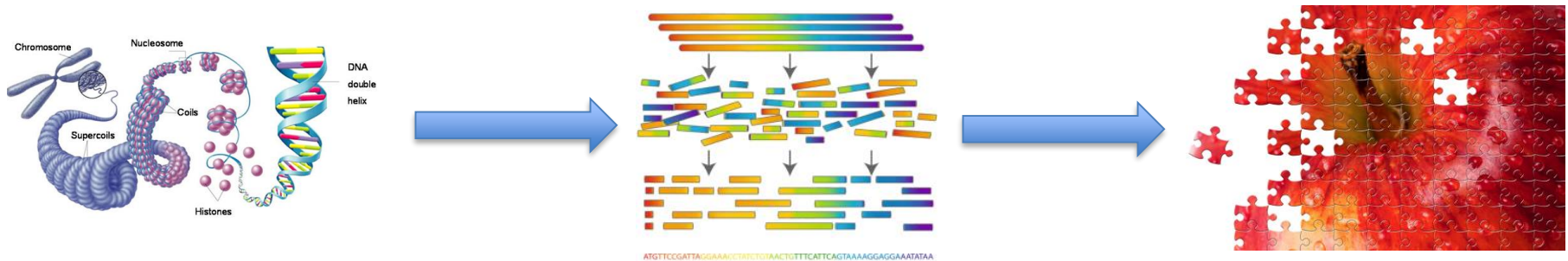
NGS Applications

NGS/MPS applications

- Whole genome sequencing:
 - De novo sequencing
 - Re-sequencing
- Transcriptome sequencing:
 - mRNA-seq
 - miRNA
 - Isoform discovery
- Target re-sequencing
 - Exome
 - Large portions of a genome
 - Gene panels
 - Amplicons

Whole genome sequencing: de novo

De novo: used to assemble a genome without previous reference



Conventional strategy (Golden Standard):

Illumina 50x sequencing on HiSeqX or NovaSeq, several insert sizes (+ Mate Pairs)

Current recommendation* (Platinum genome):

100x PacBio (ONT) only + Hi-C (coverage depends on heterozygosity)

Plus RNA-seq data for annotation

* 2019-02-05

De novo – do it with long reads!



Science AAAS

Home News Journals Topics Careers

Science Science Advances Science Immunology Science Robotics Science Signaling Science Translational Medicine

SHARE RESEARCH ARTICLE

Long-read sequence assembly of the gorilla genome

David Gordon^{1,2,*}, John Huddleston^{1,2,*}, Mark J. P. Chaisson^{1*}, Christopher M. Hill^{1*}, Zev N. Kronenberg^{1*}, Katherine ...
 * See all authors and affiliations

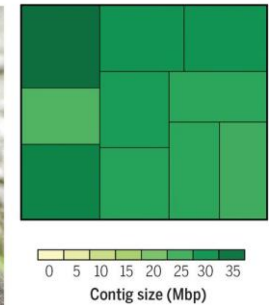
Science 01 Apr 2016:
 Vol. 352, Issue 6281, aae0344
 DOI: 10.1126/science.aae0344

PRE Peer Reviewed

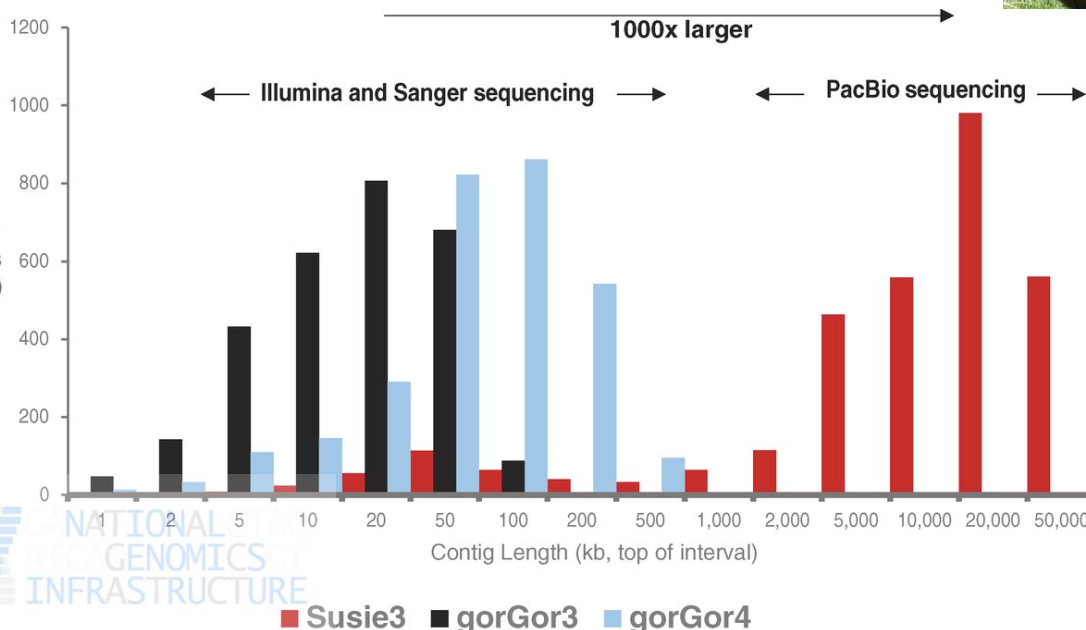
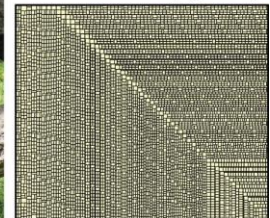
A Susie, reference sample



B Long-read assembly (Susie3)

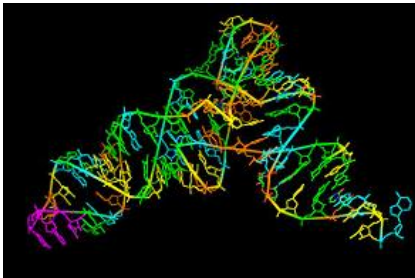


C Short-read assembly (gorGor3)



Beware: up to 80% of novel structural variants can be missing from short-read data.

⇒ **Sequence fewer genomes, but with long reads**



Transcriptome sequencing (RNA-seq)



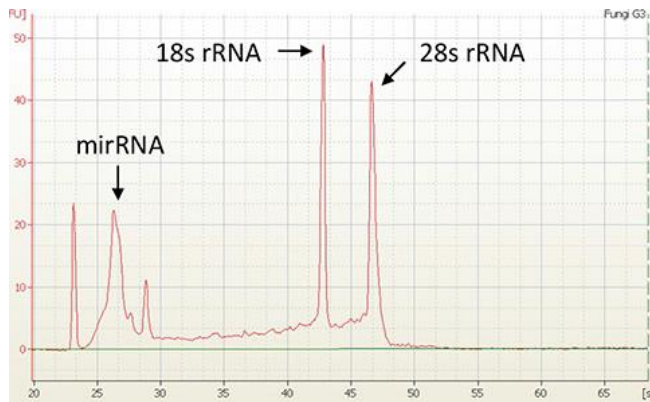
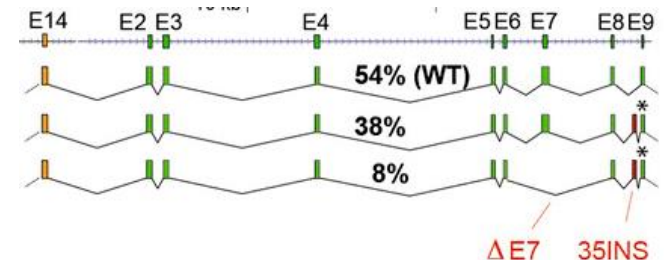
mRNA

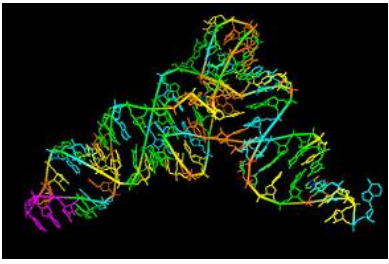
- **Dif.ex.**
- Annotation

Non-codingRNA miRNA

- Transcriptional regulation

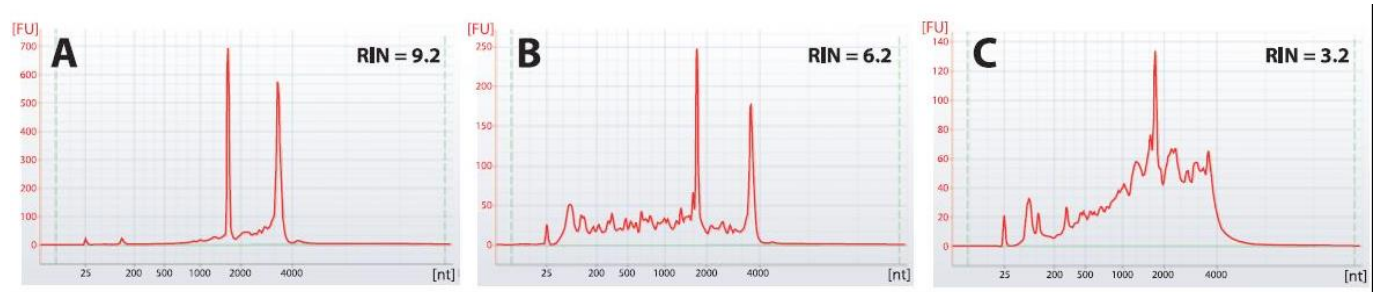
Splice isoforms



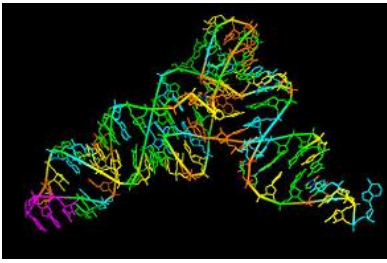


RNA-seq experimental setup

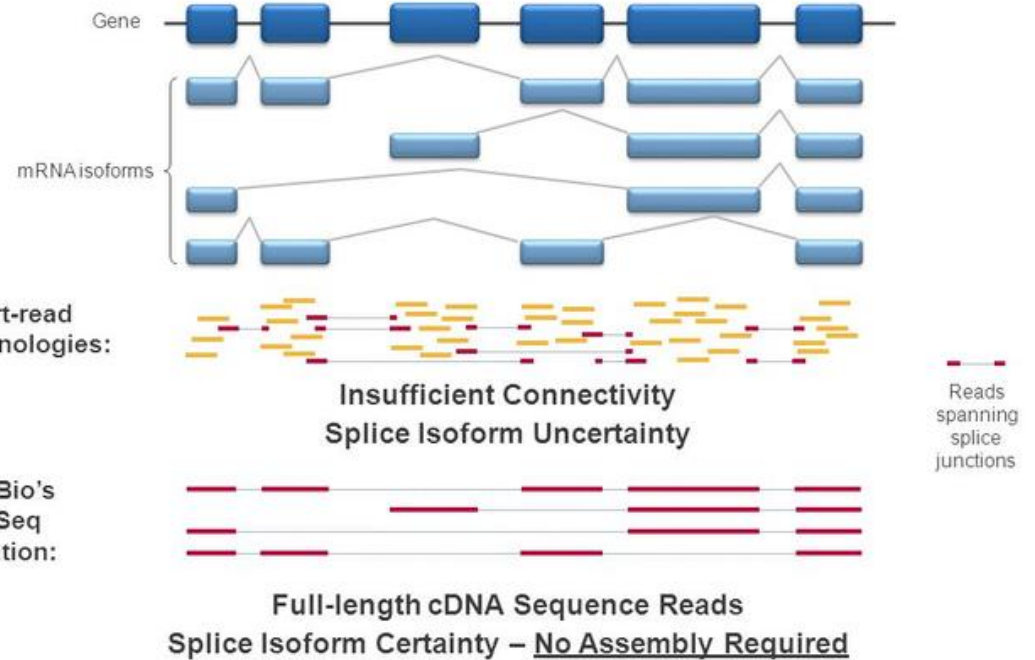
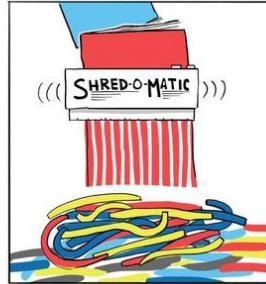
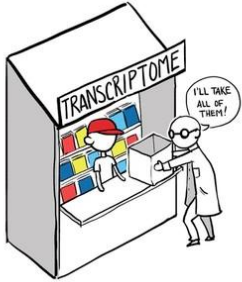
- mRNA only: any kit
- mRNA **and** miRNA: only specialized kits
- Always use DNase!
- RIN value above 8.



- CONTROL vs experimental conditions
- Biological replicates: 4 strongly recommended



RNA-seq with long reads



NATURE METHODS | NEWS AND VIEWS



Genomics: the state of the art in RNA-seq analysis

Ian Korf

Nature Methods 10, 1165–1166 (2013) | doi:10.1038/nmeth.2735

Published online 26 November 2013

PacBio Iso-seq: full-length transcriptome seq

Coming soon: direct RNA-seq on ONT

Main types of equipment & applications



Illumina HiSeq
NextSeq, HiSeqX10, MiSeq,
MiniSeq, NovaSeq

Short paired reads
HIGH throughput

Human WGS
Re-sequencing 30x
mRNA and miRNA
De novo transcriptome
Exome
ChIP-seq
Short amplicons
Methylation



Ion S5 XL

Short single-end reads
FAST throughput

mRNA and miRNA
Exome
ChIP-seq
Short amplicons
Gene panels
Clinical samples



PacBio RSII
SEQUEL

Ultra-long reads
FAST throughput

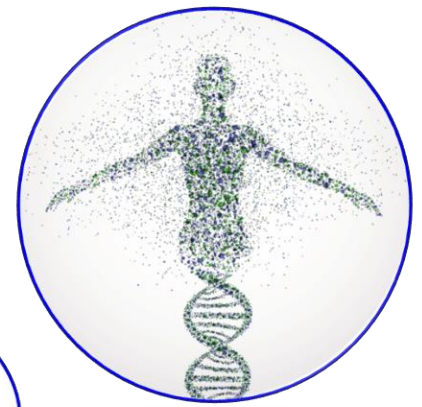
Long amplicons
Re-sequencing
De novo sequencing
Novel isoform discovery
Fusion transcript analysis
Resolving haplotypes
Clinical samples

BIG DATA

2025 projection: data storage needs

1 petabyte = 10^{15} bytes

1 exabyte = 10^{18} bytes



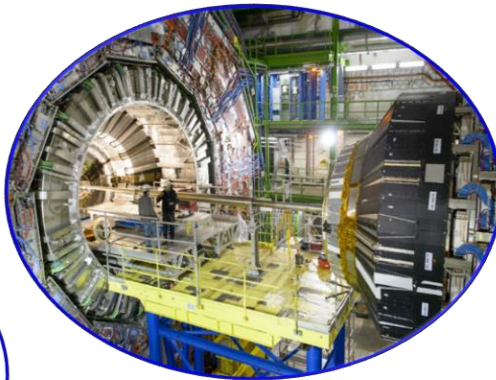
2-40 exabytes/year



1-2 exabytes/year



1 exabyte/year



Large Hadron Collider
42 petabytes/year



1-17 petabytes/year



Thanks for listening! Questions?

support@ngisweden.se



SciLifeLab