NTNU – Trondheim
Norwegian University of
Science and Technology

Realfagskonferansen 2022

Next Generation Sequencing Technology and Applications

From Sanger sequencing to the 100$ genome

Vidar Beisvåg
vidar.besivag@ntnu.no
Genomics Core Facility
Dept. of Clinical and Molecular Medicine
Mai 2022

HELSE MIDT-NORGE

NTNU

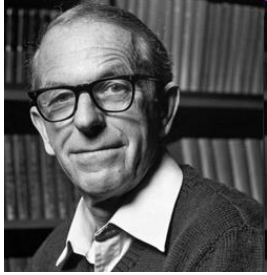
Outline

From Sanger sequencing to the 100$ genome

- History of sequencing
  - Sanger sequencing
  - Capillary electrophoresis sequencing
- Next Generation Sequencing (NGS), Second Generation Sequencing, High Throughput Sequencing (HTS) or Massive Parallel Sequencing
- NGS applications

HELSE MIDT-NORGE

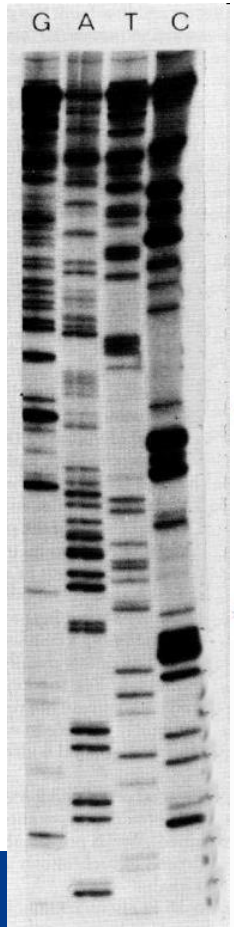
Early Days of Sequencing (1977 – 2004)

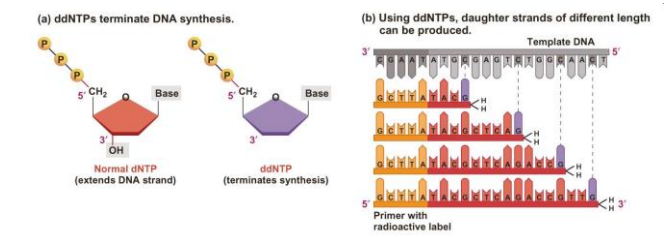


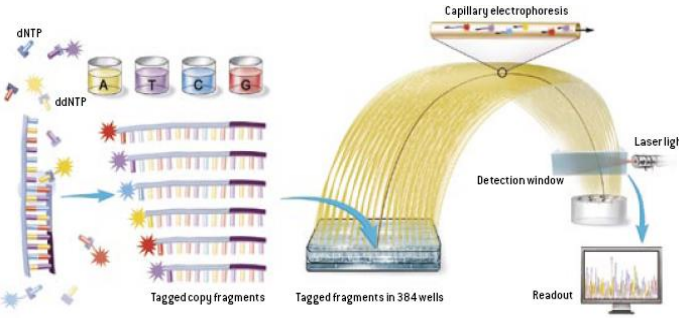
- Original method:
  - Used radio-labelled primers or dideoxynucleotides
  - Required four separate reactions to be separated by electrophoresis in four parallel lanes
  - Approximately 150- 250 bases read length

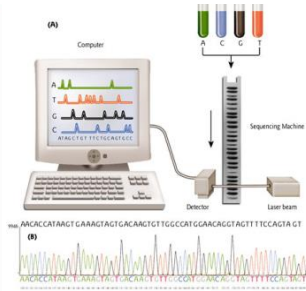
- Improved method:
  - Use of fluorescently labelled dideoxynucleotides
  - One lane electrophoresis in capillaries
  - Increased resolution and read length,now up to 1 000 bases

Sanger Sequencing









HELSE MIDT-NORGE

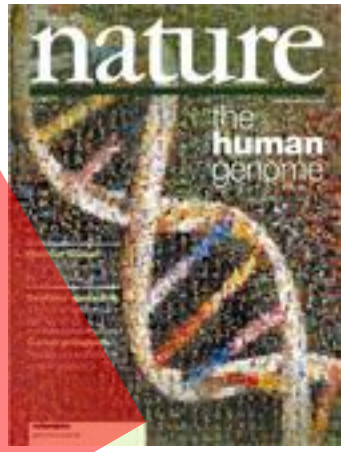
The Human Genome Project

- Initial sequencing and analysis of the human genome
  - International Human Genome Sequencing Consortium
    - Nature 409, 860 – 921 (**15 February 2001**)
- The sequence of the Human Genome
  - Celera Genomics (Venter et al.)
    - Science, Vol. 291, Issue 5507, 1304 – 1351 (**16 February 2001**)
- Finishing the euchromatic sequence of the human genome
  - International Human Genome Sequencing Consortium
    - Nature 431, 931 – 945 (**21 October 2004**)
- Last chromosome finally finished in **2006**…..

Time to finch: App. 13 years !!!

Costs: App. $2-3 billion (including development of technology)

"Raw" expenses estimated at $300 million



HELSE MIDT-NORGE

NTNU

Recent Achievements (by 2008)

- First full genome to be sequenced using high throughput sequencing technology
  - GS FLX /454 (Roche Diagnostics)

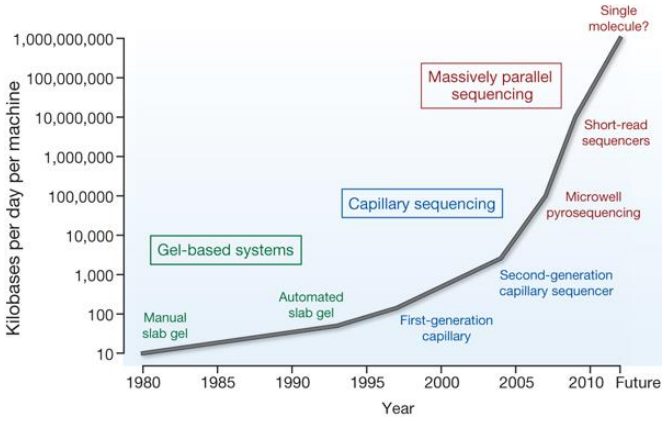
- James Watson`s genome sequenced at high speed (J. Wheeler et al. Nature 452, 872-876 (2008)
  - Four months job for a handful of scientist
  - Costs: less than $1.5 million

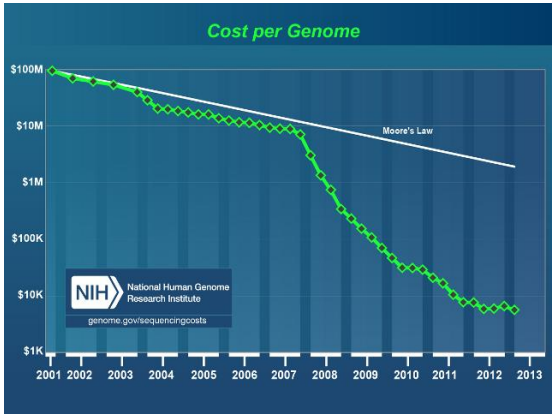
Goal: The $1 000 Genome
- Low cost
- High throughput
- High accuracy

Recent Achievements (by 2013)

- Sequencing of genomes in days is now a reality.
- MiSeq (Illumina) and Ion Torrent PGM (Life Technologies) are laser-printer sized and offer modest set-up and running costs.
- Each instrument can generate data required for e.g. sequencing a human genome/exome or make a draft of a bacterial genome in days, making them attractive for e.g. identifying and characterizing pathogens in the clinical setting.

Improved sequencing efficiency and reduced cost





NTNU

Recent achievements (by 2017) The 100$ Genome(?)

Illumina NovaSeq The next era in sequencing.

- NovaSeq offers scalable throughput and flexibility for virtually any sequencing method, genome, and scale of project
  - up to 10B clusters/reads (=> 20B PE reads) per flowcell/run => 750(25 av 30X) human genomes in two days.
- Although the instrument won't, now, achieve a $100 genome, it is believed that a new era has been launched.

Figure 1: The NovaSeq Series of Systems— Transforming sequencing by combining throughput, flexibility, and ease of use for virtually any method, genome, and scale.

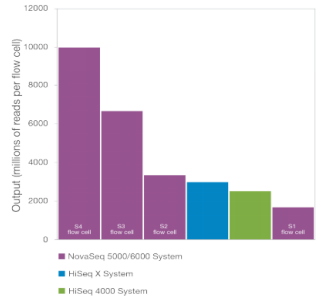


Figure 2: The NovaSeq Series Offers the Broadest Output Range— The NovaSeq Series generates from 167 Gb and 1.6 B reads to 3 Tb and 10 B reads of data in single flow cell mode. Note that when run in dual flow cell mode, output can be up to 6 Tb and 20 B reads. The tunable output makes the NovaSeq Series accessible for a wide range of applications. Note that S3 and S4 flow cells can only be used on the NovaSeq 6000 System.

Table 1: NovaSeq Series Flow Cell Specifications[a]

	NovaSeq 5000[b] and 6000 Systems		NovaSeq 6000 System	
Flow Cell Type	S1[b]	S2	S3[b]	S4[b]
Reads Passing Filter	up to 1.6 B	2.8-3.3 B	up to 6.6 B	up to 10 B
Ouput				
2 × 50 bp	up to 167 Gb	280-333 Gb	N/A[c]	N/A[c]
2 × 100 bp	up to 333 Gb	560-667 Gb	N/A[c]	N/A[c]
2 × 150 bp	up to 500 Gb	850-1000 Gb	up to 2000 Gb	up to 3000 Gb

a. Output and read number specifications based on a single flow cell using Illumina PhiX control library at supported cluster densities. NovaSeq 5000 and 6000 Systems can run 1 or 2 flow cells simultaneously.
b. NovaSeq 5000 System and NovaSeq Reagent Kits with S1, S3, or S4 flow cells have not been released; performance metrics are subject to change.
c. N/A: not applicable.

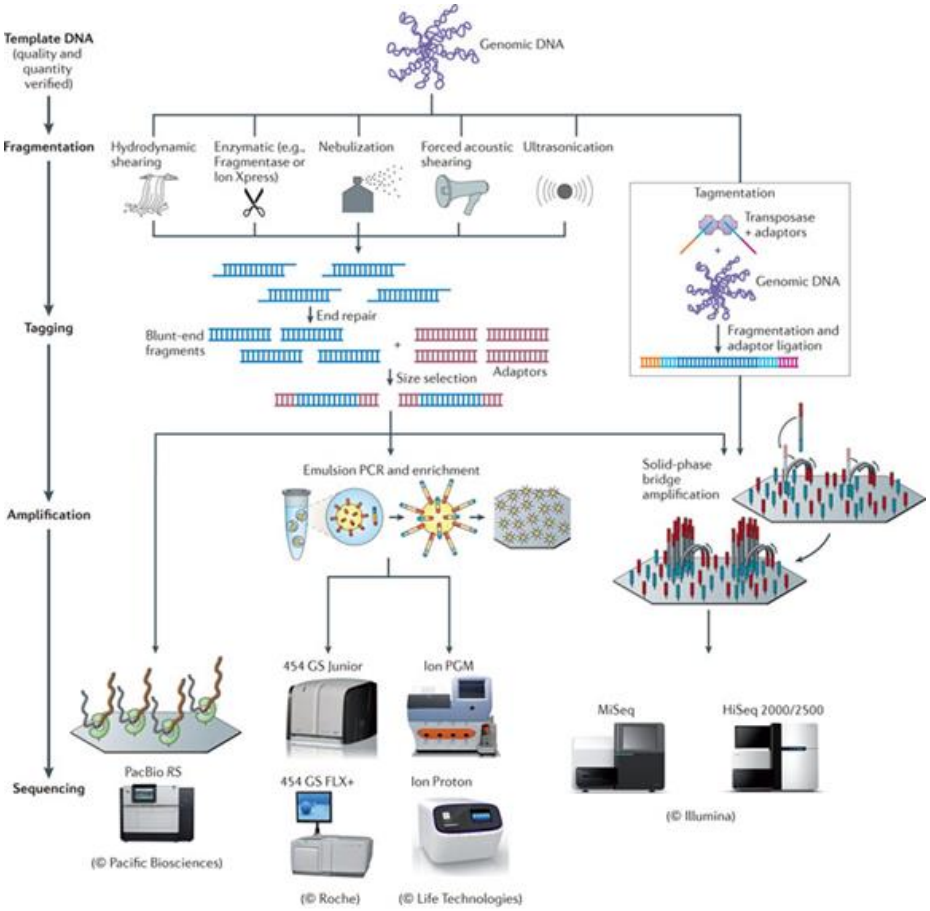
HELSE MIDT-NORGE

NTNU

NGS platforms

- Roche: Genome Sequencer FLX System (454), GS Jr.
- Illumina: NovaSeq, HiSeq, NextSeq, MiSeq, MiniSeq
- Life technologies: Ion Torrent (Personal Genome Machine)
- Pacific Bioscience (long read, singe molecule sequencer): PacBio RS II and Sequel System
- Oxford Nanopore (single molecule sequencer) – MinION/GridION

Platform	Sequence length	Throughput	Reads	Error type	Time
Pacific Bio RSII Pacific Bio Sequel	20 kb 8-12 kb	1 Gb 7 Gb	0.055 M 0.35 M	13% single pass, ≤1% circular consensus read, indels	4 h 6 h
Ion PGM Ion Proton	400 bp (SR) 200 bp (SR)	1 Gb 15 Gb	5.5 M 80 M	1% Indels	7 h 2.5 h
Illumina MiSeq *v3*	75 bp (PE) 300 bp (PE)	3.8 Gb 15 Gb	50 M (PE)	Substitutions 0.1%	21-56 h
Illumina NextSeq *Highoutput*	75 bp (SR) 75 bp (PE) 150 bp (PE)	30 Gb 60 Gb 120 Gb	400 M (SR) 800 M (PE)	Substitutions <0.1%	11 h 18 h 29 h
Illumina HiSeq 4000	50 bp (SR) 75 bp (PE) 150 bp (PE)	125 Gb 375 Gb 750 Gb	3.5 B (SR) 7 B (SR)	Substitutions 0.1%	1-3.5 d
Illumina Novaseq 5000 & 6000	50 bp (PE) 100 bp (PE) 150 bp (PE)	333 Gb 667 Gb 3000 Gb	3.2 B (S1) to 20 B (S4)	Substitutions ?%	2 d
Oxford MinION	200 kb	1.5 Gb	0.1 M	12% indels	48 h
10x Genomics	100 kb	HiSeq	HiSeq	HiSeq	



Clonal cluster sequencing

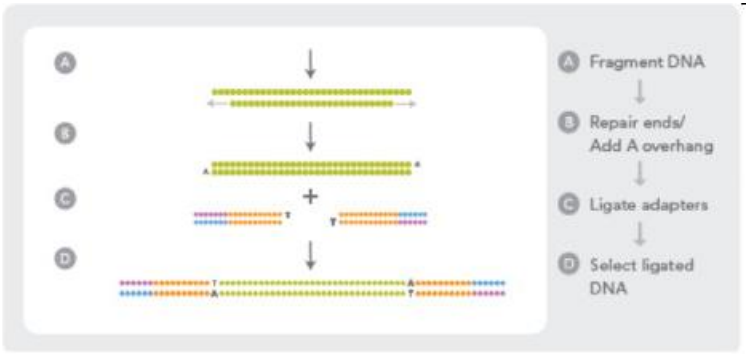
- **Illumina**
  - Sequencing by synthesis
- **Roche**
  - Pyrosequencing
- **Non optical, life technologies**
  - Ion Torrent/ Ion Proton
- **Single molecule sequencing**
  - Pacific Biosciences, Oxford Nanopores

Illumina workflow

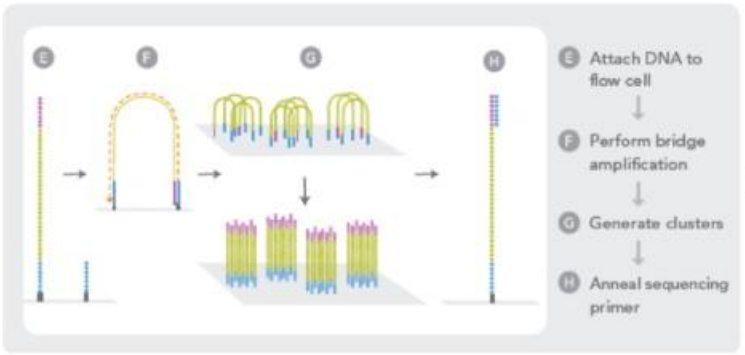
- Fragmentation and adapter ligation steps take place, before applying the library onto the solid surface of a **flow cell**.
- Attached DNA fragments form ‘bridge’ molecules which are subsequently amplified via an isothermal amplification process, leading to a cluster of identical fragments that are subsequently denatured for sequencing primer annealing.
- Amplified DNA fragments are subjected to **sequencing-by-synthesis** using 3′ blocked labelled nucleotides.

Sequencing by synthesis

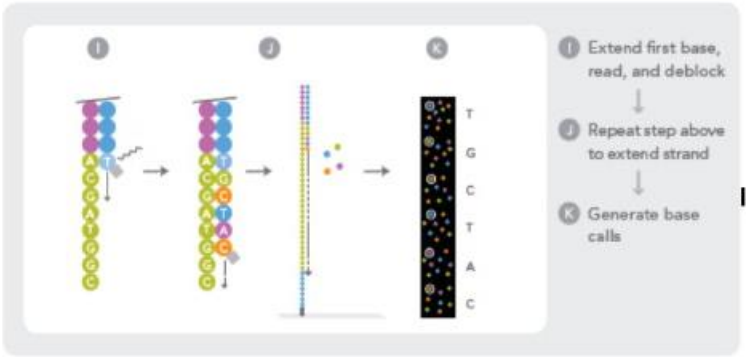
I.

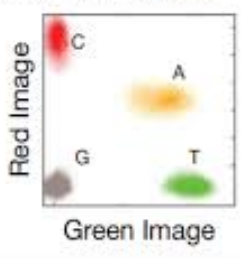


II.



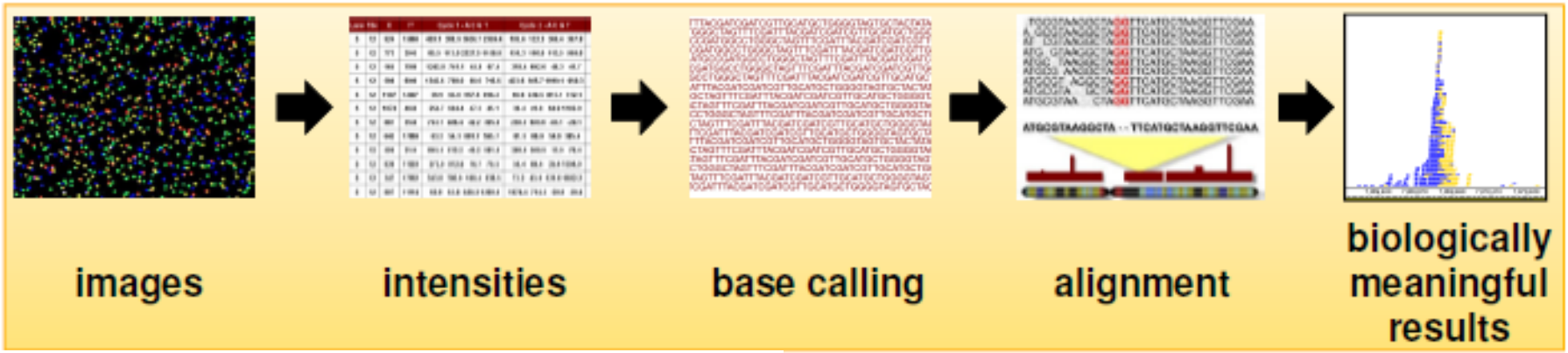
II.



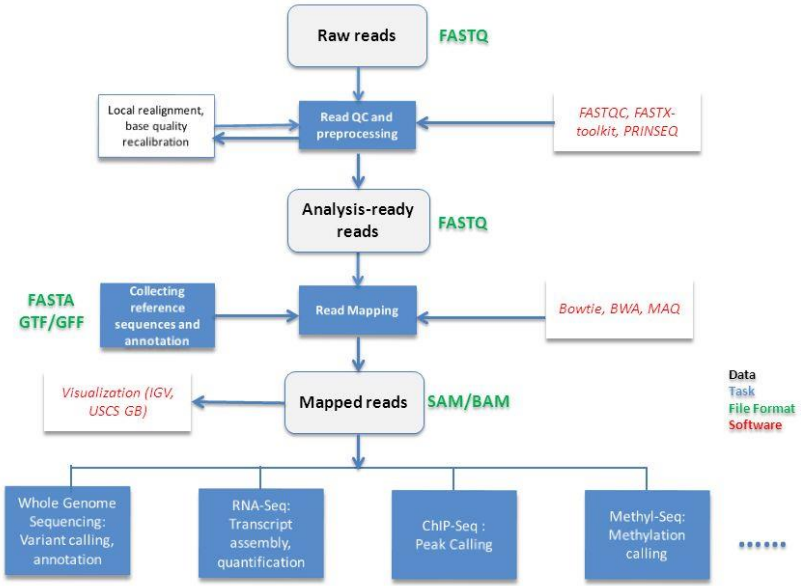


NTNU

Data Analysis



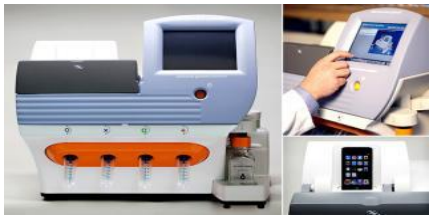
Data Analysis Pipeline



Important: Data security for human sequences

NTNU

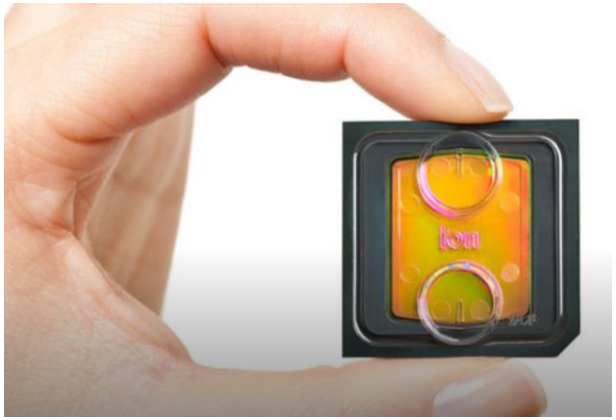
Ion Torrent

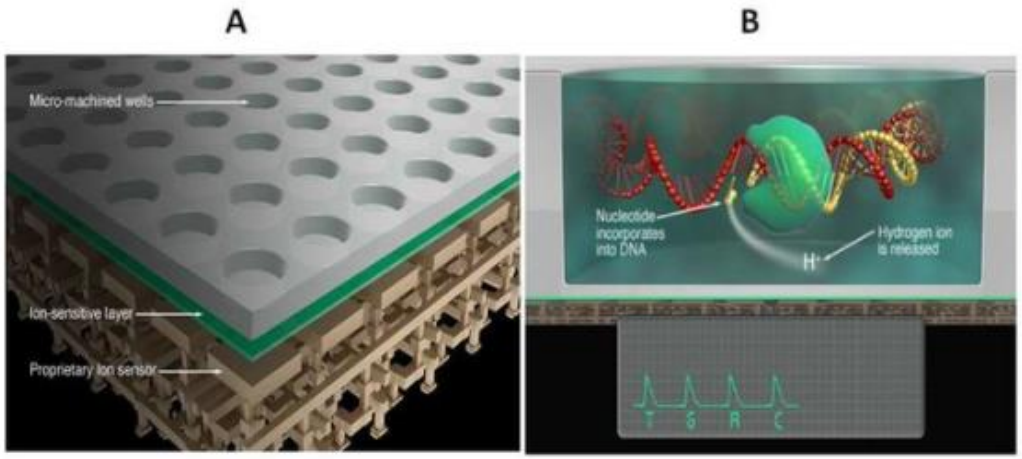


- Three types of semiconductor chips:
  - 314 – 20Mb, 316 - 200Mb and 318 – 1Gb
  - Read length depends on base composition 200-250bp
  - System is enabled for Paired End 2x100cycles
  - The fastest sequencing system on the market.
- Applications:
  - Resequencing applications which require fast turnaround of samples
  - Amplicons (PCR products)
  - Small and medium size genomes
  - Custom DNA capture applications

How it works:

H+ ion is released during base incorporation. Individual polymerases attached to beads are positioned in tiny wells that rest on a tiny pH meter.

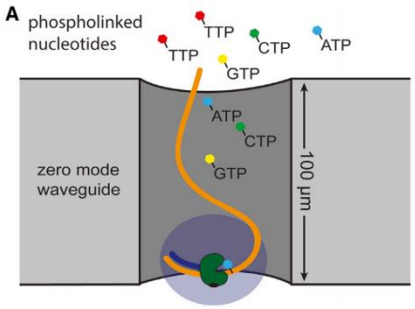


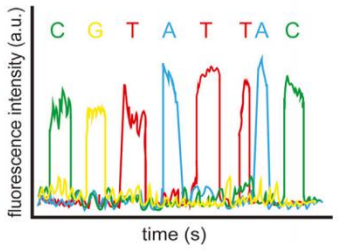


What's next?
True Single Molecule Sequencing

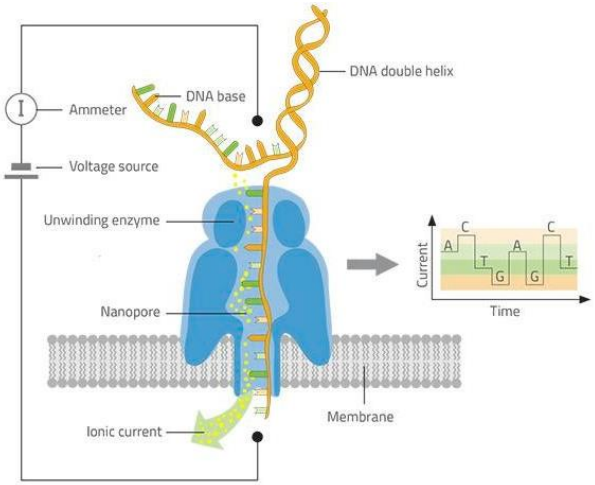
- Sequences directly single molecules of DNA
- No amplification
- Increased speed

PACIFIC BIOSCIENCES™





Oxford NANOPORE Technologies



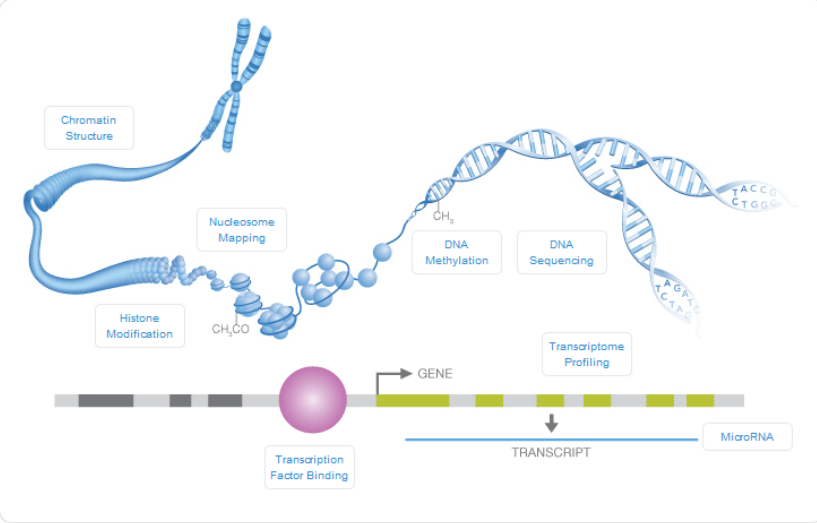
HELSE MIDT-NORGE

NTNU

Obtaining a genome sequence is a one step towards understanding biological processes

Questions that follow form the genome are:

- How is the genome organized?
- What is transcribed?
- Where do protein binds?
- What is methylated etc.?



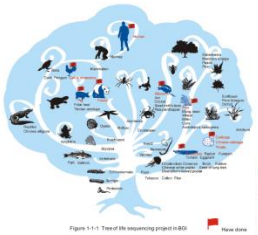
In other words; this can give us information about how the cell works.

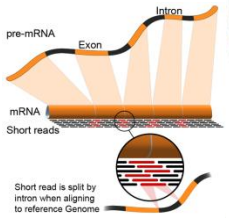
HELSE MIDT-NORGE

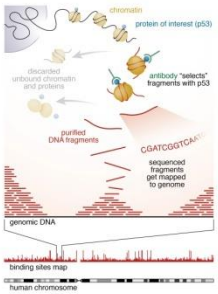
NTNU

HTS can be used for a variety of studies e.g.:

- de novo sequencing
- Whole genome or targeted resequencing
  - Complete exome sequencing and amplicon sequencing
- Whole genome transcriptome profiling including
  - Tag profiling
  - Quantification of transcript isoforms, fusion genes and small RNAs
- Gene regulation profiling by e.g.
  - ChIP-Seq to e.g.
    - Detect transcription binding sites across the genome (DNA-protein interactions)
    - Chromatin structure (histone modifications)
  - CpG methylation analysis
- Metagenomic sequencing of e.g.
  - Microflora genomes (human intestine, skin etc.)
  - Environmental samples (e.g. soil, water etc).





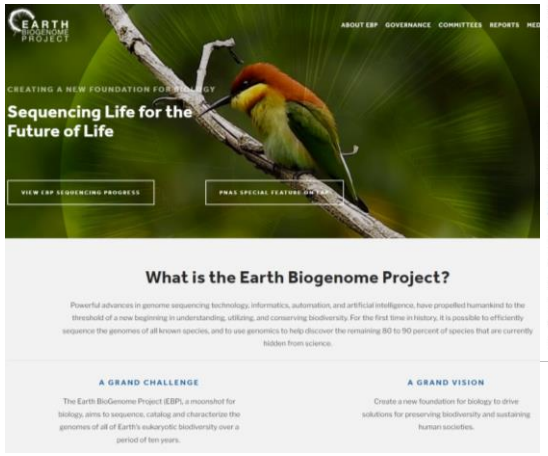


HELSE MIDT-NORGE

NTNU

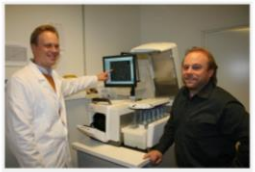
De Novo Sequencing

- *De novo* sequencing refers to sequencing a **novel genome** where there is **no reference sequence available** for alignment. NGS enables fast, accurate characterization of any species.
- **Sequence reads are assembled as contigs**, and the coverage quality of *de novo* sequence data depends on the size and continuity of the contigs (e.g., the number of gaps in the data).
- NGS enables faster, more accurate characterization of any species compared to traditional methods, such as Sanger sequencing. **Mate pair sequencing and long-read technology can complement shorter reads for accurate, complete characterization of any species.**



Norwegian Team Completes Cod Genome

OCTOBER 30, 2009



Lex Nederbragt and Kjetil S. Jakobsen in front of the machine that they use in their analysis. During one afternoon, the machine can read 500 million bases, which appears as dots on the computer screen. Photo: Elin Fugelsnes

Scientist from a consortium of Norwegian marine and aquaculture institutions has announced this week that they have generated the first ever draft sequence and assembly of the Atlantic cod genome.

The results, presented at the GenoFisk meeting in Oslo, Norway, include a high-quality genome assembly and preliminary annotation of the fish species, completed using only whole genome shotgun and paired-end data generated by the Genome Sequencer FLX System from 454 Life Sciences, a Roche Company.

By using the long-read GS FLX Titanium Series and the system's accompanying GS Assembler software, the consortium was able to produce a draft assembly at a fraction of the time and cost of traditional sequencing methods. The researchers hope to use this information and the fully annotated genome to identify genes and genetic variations underlying important traits for cod aquaculture.

nature

Explore content · About the journal · Publish with us

Open Access | Published: 18 April 2016

The Atlantic salmon genome provides insights into rediploidization

Sigbjørn Lien, Ben F. Koop, [...]William S. Davidson

Nature **533**, 200–205 (2016) | Cite this article

54k Accesses | 480 Citations | 227 Altmetric | Metrics

Abstract

The whole-genome duplication 80 million years ago of the common ancestor of salmonids (salmonid-specific fourth vertebrate whole-genome duplication, Ss4R) provides unique opportunities to learn about the evolutionary fate of a duplicated vertebrate genome in 70 extant lineages. Here we present a high-quality genome assembly for Atlantic salmon (*Salmo salar*), and show that large genomic reorganizations, coinciding with bursts of transposon-mediated repeat expansions, were crucial for the post-Ss4R rediploidization process. Comparisons of duplicate gene expression patterns across a wide range of tissues with orthologous genes from a pre-Ss4R outgroup unexpectedly demonstrate far more instances of neofunctionalization than subfunctionalization. Surprisingly, we find that genes that were retained as duplicates after the teleost-specific whole-genome duplication 320 million years ago were not more likely to be retained after the Ss4R, and that the duplicate retention was not influenced to a great extent by the nature of the predicted protein interactions of the gene products. Finally, we demonstrate that the Atlantic salmon assembly can serve as a reference sequence for the study of other salmonids for a range of purposes.

HELSE MIDT-NORGE

NTNU

Whole-Genome (re)Sequencing

- Whole-genome sequencing is the most comprehensive method for analyzing the genome.
- Rapidly dropping sequencing costs and the ability to obtain valuable information about the entire genetic code make whole-genome sequencing a powerful research tool.
- Analyzing the whole genome using HTS delivers a base-by-base view of all genomic alterations, including **single nucleotide variants (SNV), insertions and deletions, copy number changes, and structural variations**.

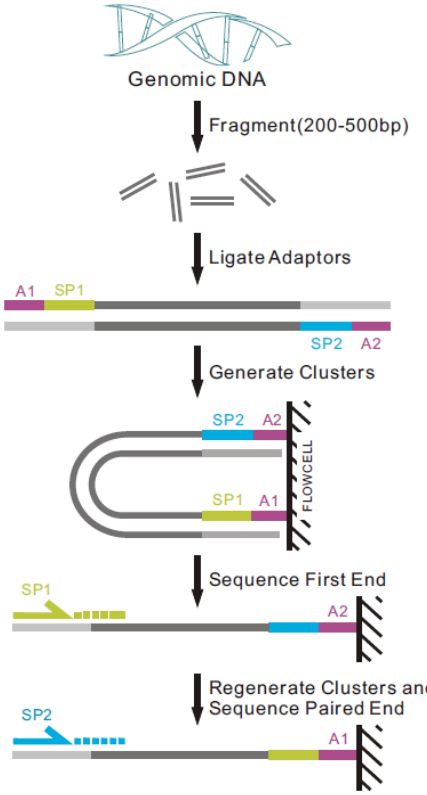
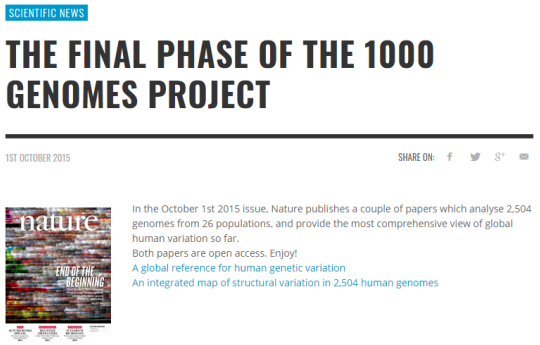


Figure 1-2-1 Pipeline of paired-end sequencing (www.illumina.com)



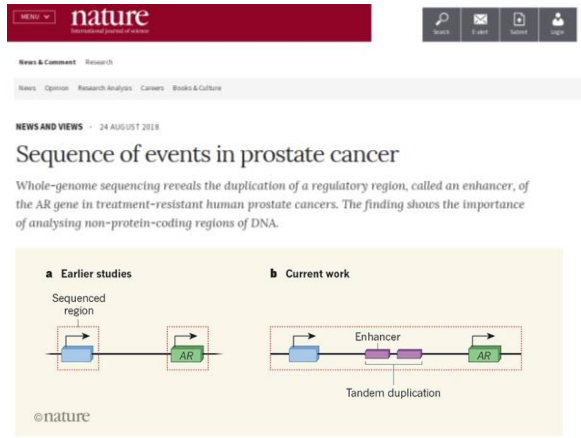
SCIENTIFIC NEWS

THE FINAL PHASE OF THE 1000 GENOMES PROJECT

1ST OCTOBER 2015

SHARE ON:

In the October 1st 2015 issue, Nature publishes a couple of papers which analyse 2,504 genomes from 26 populations, and provide the most comprehensive view of global human variation so far.
Both papers are open access. Enjoy!
A global reference for human genetic variation
An integrated map of structural variation in 2,504 human genomes



nature

NEWS AND VIEWS · 24 AUGUST 2018

Sequence of events in prostate cancer

Whole-genome sequencing reveals the duplication of a regulatory region, called an enhancer, of the AR gene in treatment-resistant human prostate cancers. The finding shows the importance of analysing non-protein-coding regions of DNA.

a Earlier studies
Sequenced region
AR
b Current work
Enhancer
AR
Tandem duplication
©nature

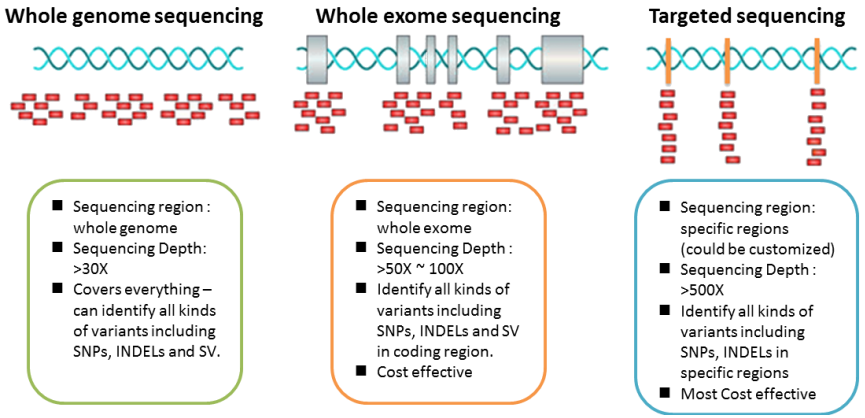
Figure 1 | Duplication of an enhancer region of the genome occurs frequently in prostate cancer. a, Many of the earlier DNA-sequencing studies of human prostate cancer have focused on the protein-coding regions of the genome, such as the gene shown in blue (the red dotted box indicates sequenced regions). This work identified alterations in the sequence of the *AR* gene, which encodes the androgen receptor (AR), as being a common driver of disease progression[7,8]. **b**, Takeda *et al.*[3], Viswanathan *et al.*[4] and Quigley *et al.*[5] demonstrate the utility of sequencing approaches that are not restricted to the protein-coding regions, and the advantages of sequencing tumour samples that have become resistant to therapy as a way of investigating why clinical treatment eventually fails. The three studies of late-stage prostate cancer report that the DNA sequence in a region upstream of *AR*, termed an enhancer, is commonly expanded, and this amplification is often in the form of a tandem duplication of the sequence. An enhancer amplification can drive expression of *AR*, which would enable tumours to evade the effects of clinical treatments that target the AR signalling pathway.

HELSE MIDT-NORGE

NTNU

Targeted (re)Sequencing

- With targeted resequencing, **a subset of genes or regions of the genome are "isolated" and sequenced**, allowing researchers to focus time, expenses, and analysis on specific areas of interest.
- Targeted resequencing enables researchers to focus on regions that are most likely to be involved in the phenotype under study, conserving resources and generating a smaller, more manageable data set.
- Such targeted analysis can include the **exome** (the protein-coding portion of the genome), specific genes of interest (custom content), targets within genes, or mitochondrial DNA.
- Targeted approaches can also deliver much **higher coverage** levels, allowing identification of variants that would be too rare and too expensive to identify with whole-genome sequencing.

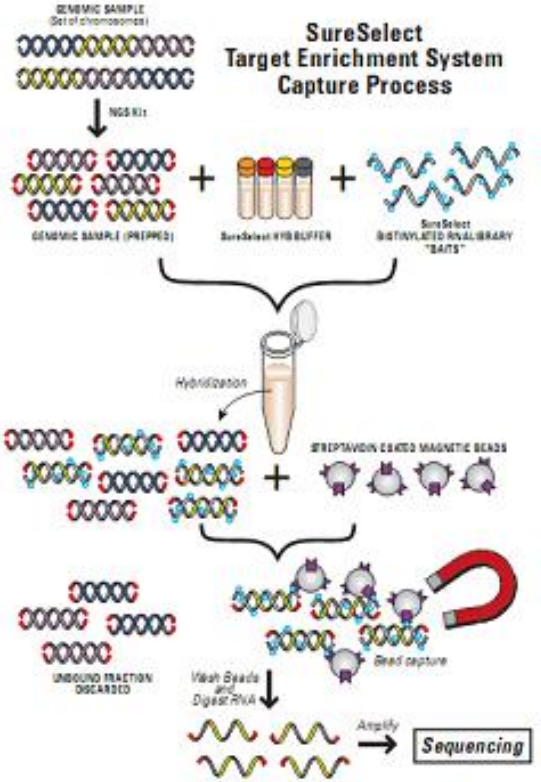


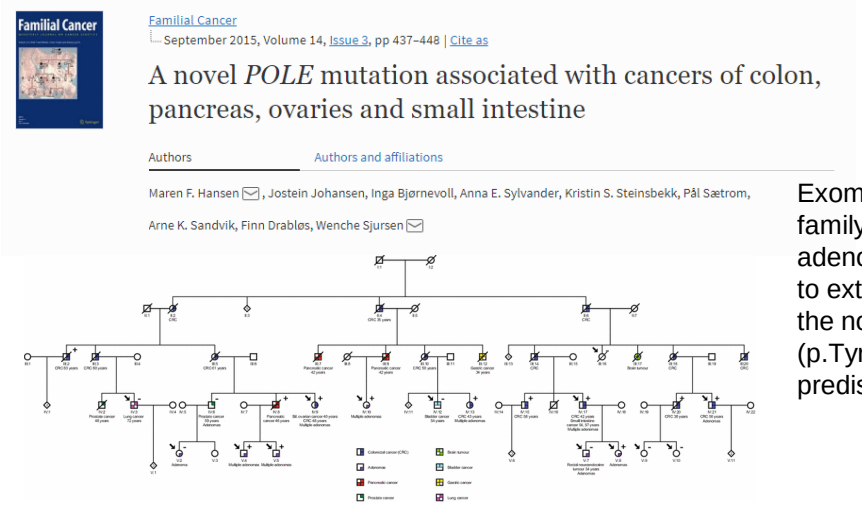
HELSE MIDT-NORGE

NTNU

Exome Sequencing

- Exome sequencing analyzes the protein-coding region of the genome, where many disease-causing variants are found. It is a cost-effective alternative to whole-genome sequencing.
- The exome (of the human genome) represents less than 2% of the genetic code, but contains ~85% of known disease-related variants, making whole-exome sequencing a cost-effective alternative to whole-genome sequencing.
- Exome sequencing can efficiently identify coding variants across a wide range of applications, including population genetics, genetic disease, and cancer studies.
- Exome sequencing detects variants in coding exons, with the capability to expand targeted content to include untranslated regions (UTRs) and microRNA for a more comprehensive view of gene regulation.

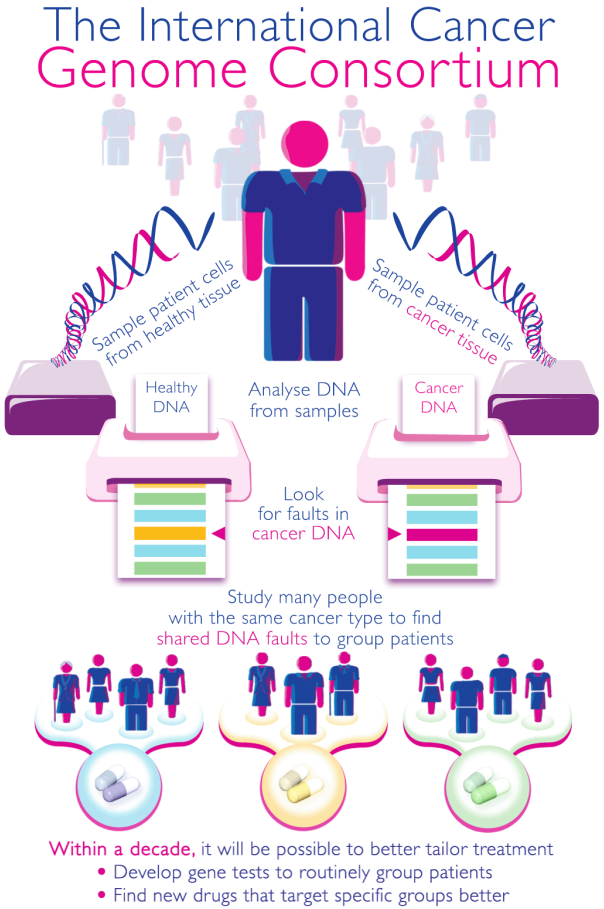




Exome sequencing of members of a family with high burden of colorectal adenomas and carcinomas, in addition to extra-colonic cancers, has identified the novel mutation c.1373A>T (p.Tyr458Phe) in *POLE* as a likely predisposing mutation.

The Complete Genomic Landscape of Cancer

- Cancer whole-genome sequencing (WGS) provides a base-by-base view of the unique mutations present in cancer tissue. It enables discovery of novel cancer-associated variants, including single nucleotide variants (SNVs), copy number changes, and structural variants. By comparing tumor and normal DNA, WGS can also provide a comprehensive view of changes to a specific tumor sample.
- Many cancer-associated variants have been discovered using cancer genome sequencing. Ongoing efforts to categorize and characterize mutations in cancer include The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC), and Catalogue of Somatic Mutations in Cancer (COSMIC).
- Through tumor-normal WGS, researchers can compare tumor mutations to a matched normal sample. Tumor-normal comparisons are crucial for identifying the somatic variants that act as driver mutations in cancer progression.



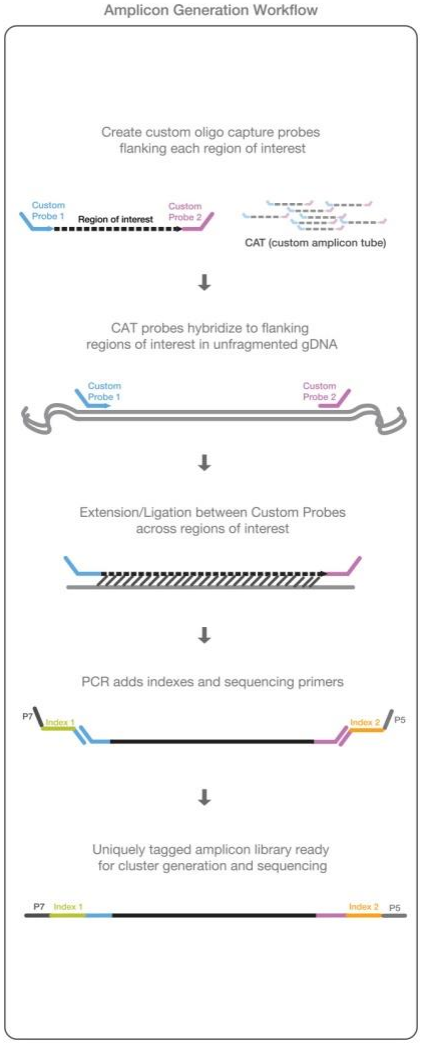
HELSE MIDT-NORGE

NTNU

Custom Targeted Gene Sequencing

- With custom designs, researchers can target regions of the genome relevant to their specific research interests. Custom targeted sequencing is ideal for examining genes in specific pathways, or for follow-up experiments from genome-wide association studies or whole-genome sequencing.

- Two methods for targeted sequencing: target enrichment and amplicon generation.
  - **Target enrichment** (for larger gene content, typically > 50 genes): Regions of interest are captured by hybridization to biotinylated probes and then isolated by magnetic pulldown. Target enrichment captures 20 kb–62 Mb regions, depending on the experimental design.
  - **Amplicon sequencing** (for Smaller gene content, typically < 50 genes): Regions of interest are amplified and purified using highly multiplexed oligo pools. This method allows researchers to sequence 26–1536 targets at a time, spanning 150–450 bp per target, depending on the library preparation kit used.



HELSE MIDT-NORGE

NTNU

Environmental DNA (eDNA) sequencing

- Environmental DNA (eDNA) sequencing is a rapidly emerging method for studying biodiversity and monitoring ecosystem changes. As organisms shed DNA into their environments, eDNA analysis can provide clues about the species present without disrupting the ecosystem.

- eDNA Metabarcoding

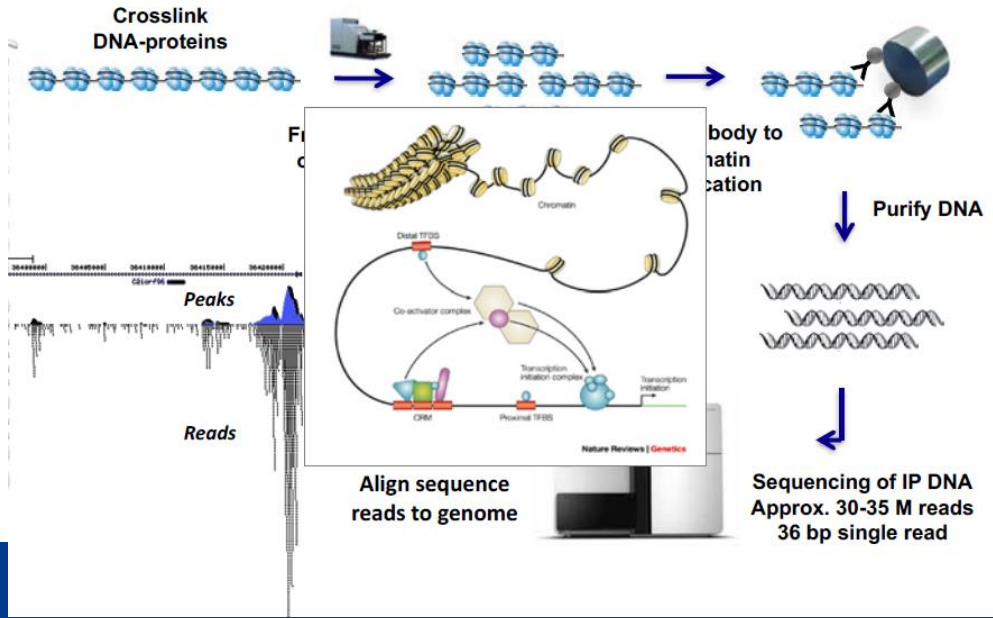
  Every organism has a unique DNA sequence, or barcode, associated with it. This DNA barcode is a highly variable region interspersed between conserved genomic regions. eDNA metabarcoding involves target-specific amplification and sequencing of these barcodes, often mitochondrial cytochrome oxidase 1 (CO1) or the 18S ribosomal subunit. These are useful approaches for distinguishing between higher-order eukaryotes.

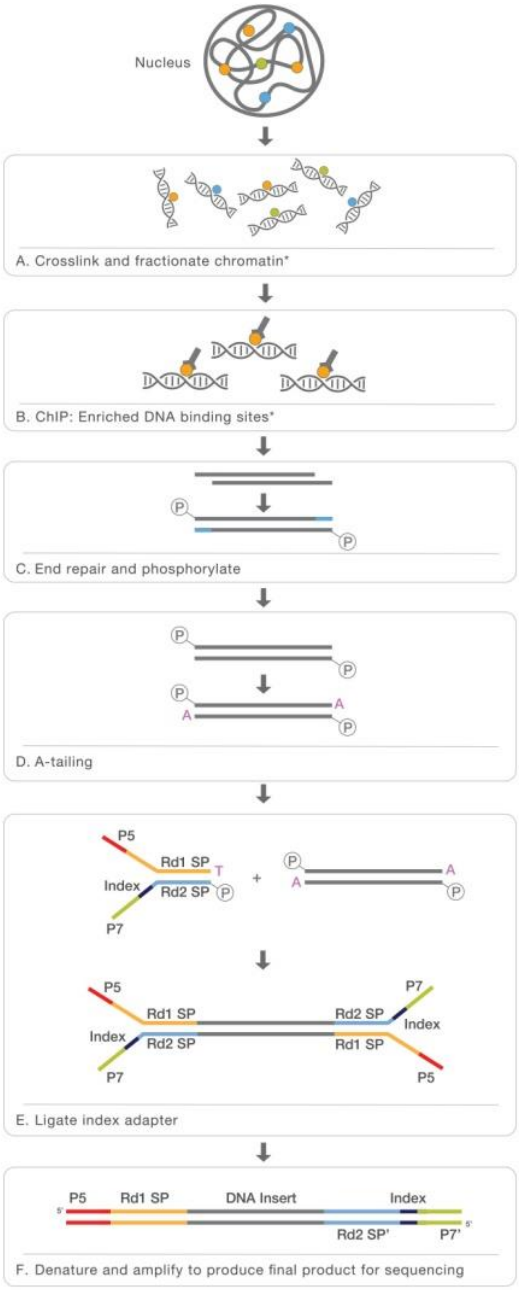
HELSE MIDT-NORGE

NTNU

ChIP Sequencing

- By combining **chromatin immunoprecipitation** (ChIP) assays and sequencing, ChIP sequencing (ChIP-Seq) is a powerful method to identify genome-wide DNA binding sites for transcription factors and other proteins.
- Following ChIP protocols, DNA-bound protein is immuno-precipitated using a specific antibody. The bound DNA is then coprecipitated, purified, and sequenced.
- The application of HTS to ChIP has revealed insights into gene regulation events that play a role in various diseases and biological pathways, such as development and cancer progression. ChIP-Seq enables thorough examination of the interactions between proteins and nucleic acids on a genome-wide scale.

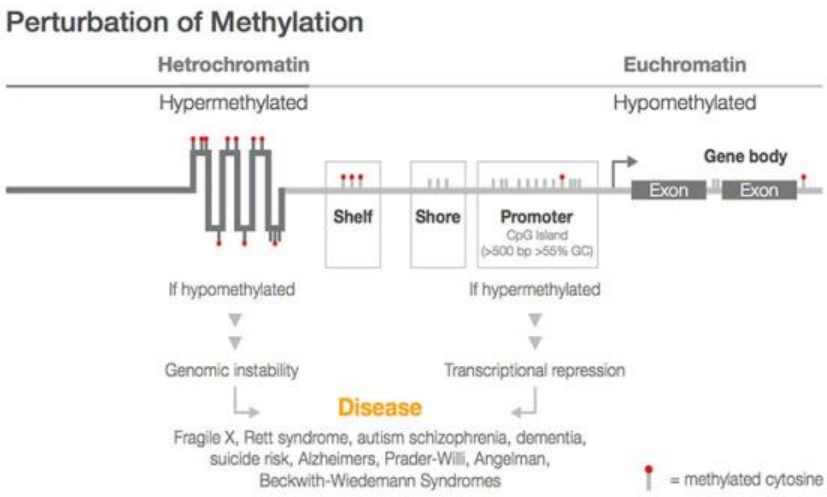


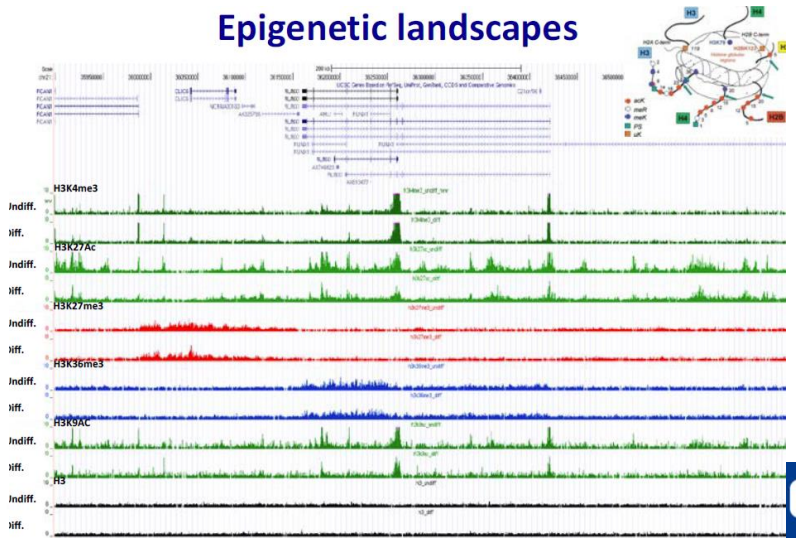


NTNU

Epigenetic (Methylation) Sequencing

- Cytosine methylation can significantly modify temporal and spatial gene expression and chromatin remodeling.
- Whole-genome bisulfite sequencing (WGBS) leverages the power of NGS and genome-wide analysis to provide a comprehensive view of methylation patterns at single-base resolution across the genome.
- Most methods rely on bisulfite conversion of DNA to detect unmethylated cytosines.
- **Bisulfite conversion changes unmethylated cytosines to uracil** during library preparation.
- Converted bases are identified (after PCR) as thymine in the sequencing data, and read counts are used to determine the % methylated cytosines.
- Bisulfite conversion sequencing can be done with targeted methods, such amplicon methyl-seq, or with whole-genome bisulfite sequencing (WGBS).

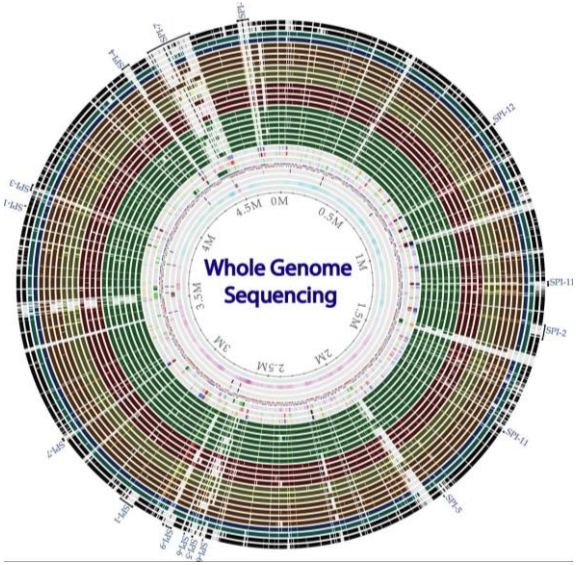




NTNU

Whole-Genome Sequencing (resequencing of small genomes)

- Small genome sequencing (≤ 5 Mb) involves sequencing the entire genome of a **bacterium, virus, or other microbe**, and then comparing the sequence to a known reference.
- **Without requiring bacterial culture**, researchers can sequence thousands of small organisms in parallel using NGS.
- Sequencing small microbial genomes can be useful for **food testing, in public health, infectious disease surveillance, molecular epidemiology studies, and environmental metagenomics.**

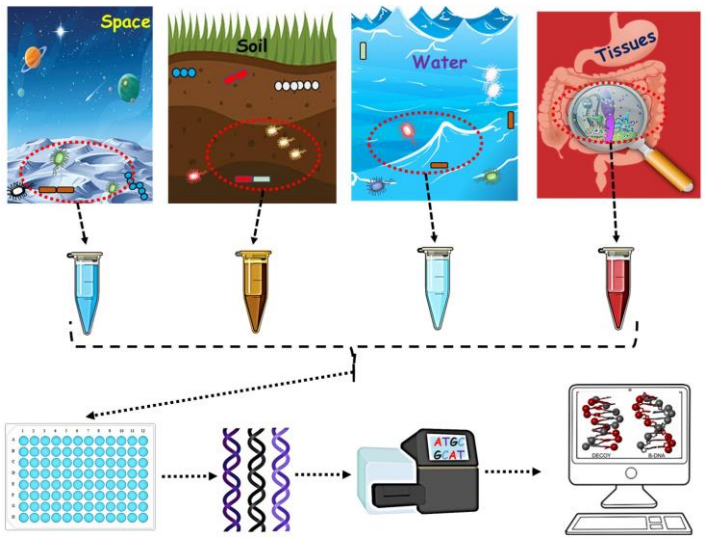


HELSE MIDT-NORGE

NTNU

Fast, Culture-Free Microbial Analysis

- Unlike traditional approaches, small genome sequencing studies using NGS do not rely on labor-intensive cloning steps.
- NGS also enables biologists to sequence hundreds of organisms simultaneously via multiplexing.
- NGS can identify low-frequency variants, genomic rearrangements, and other genetic changes that might be missed or are too costly to identify using other methods.
- For small genomes, DNA libraries can be prepared, sequenced, and analyzed in as little as 2 days.



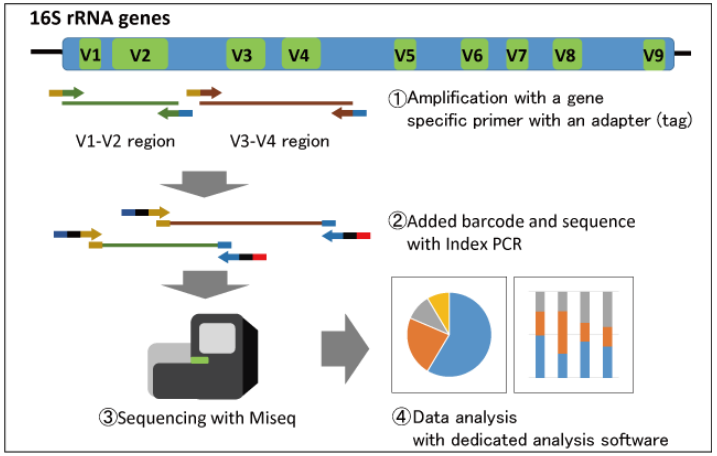
HELSE MIDT-NORGE

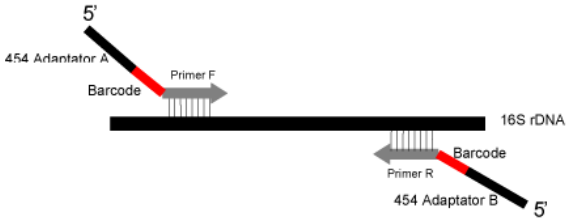
NTNU

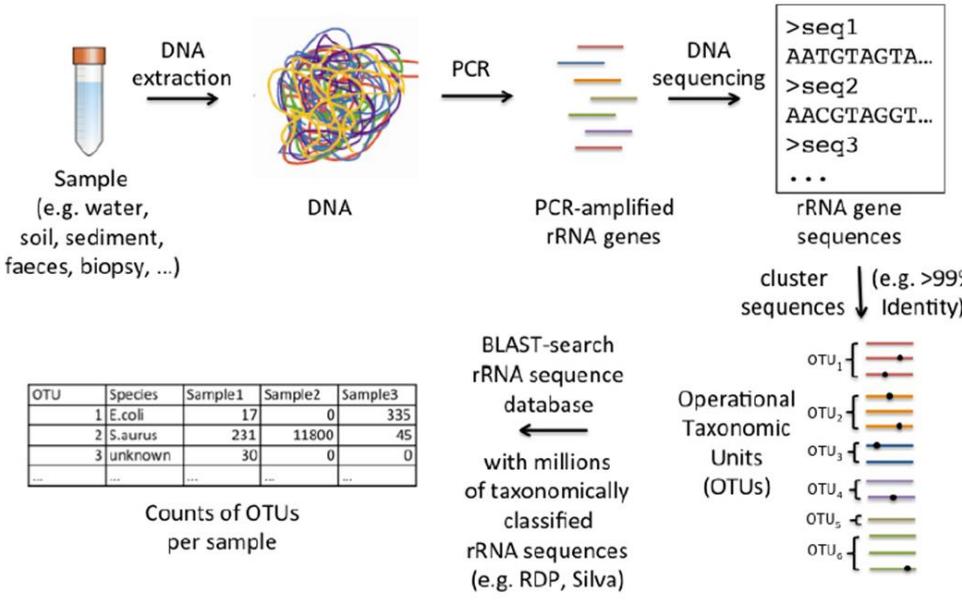
Metagenome analysis

16S rRNA Sequencing

- 16S ribosomal RNA (rRNA) sequencing is a common amplicon sequencing method used to identify and compare bacteria present within a given sample. NGS-based 16S rRNA sequencing can identify strains that might not be found using other methods.
- 16S rRNA gene sequencing is a well-established method for studying phylogeny and taxonomy of samples from complex microbiomes or environments that are difficult or impossible to study.
- Unlike capillary sequencing or PCR-based approaches, NGS is a culture-free method that enables analysis of the entire microbial community within a sample.







OTU	Species	Sample1	Sample2	Sample3
1	E.coli	17	0	335
2	S.aurus	231	11800	45
3	unknown	30	0	0
...	...	...	...	...

HELSE MIDT-NORGE

nature.com > scientific reports > articles > article

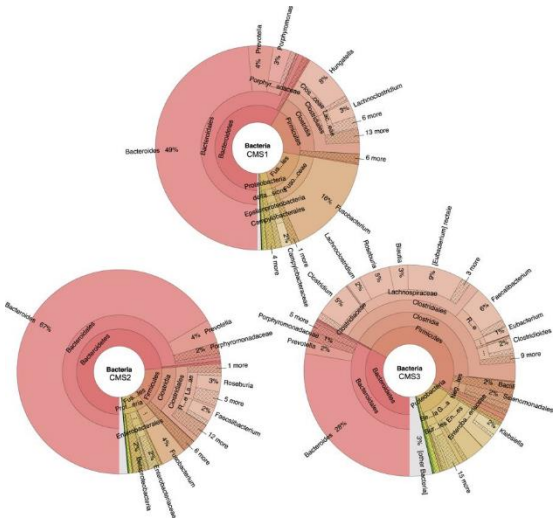
SCIENTIFIC REPORTS

Article | OPEN | Published: 14 September 2017

Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer

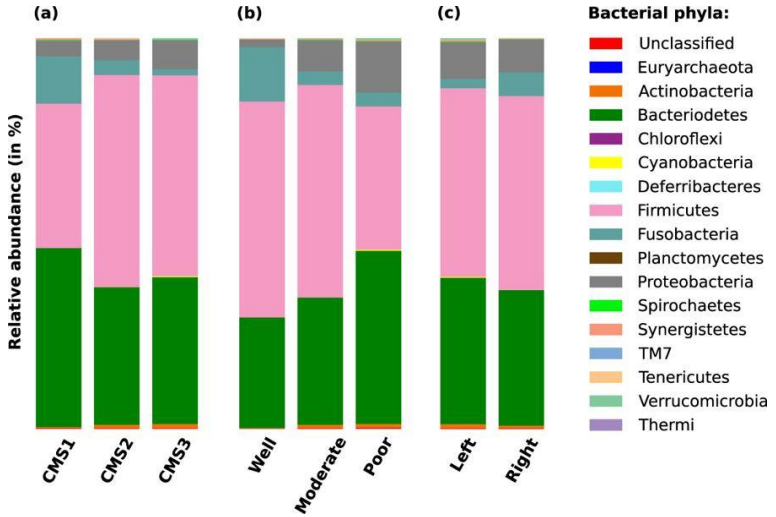
Rachel V. Purcell, Martina Visnovska, Patrick J. Biggs, Sebastian Schmeier & Frank A. Frizelle

Scientific Reports **7**, Article number: 11590 (2017) | Download Citation



Krona plots of for each CMS showing relative abundance of bacterial taxa at the genus level.

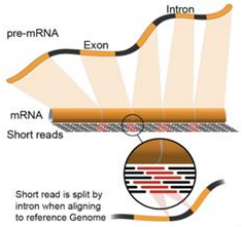
The contribution of bacterial species to CRC development is increasingly acknowledged, and here, Purcell et al sought to analyse CRC microbiomes and relate them to tumour consensus molecular subtypes (CMS), in order to better understand the relationship between bacterial species and the molecular mechanisms associated with CRC subtypes.
16S rRNA analysis showed enrichment of Fusobacteria and Bacteroidetes, and decreased levels of Firmicutes and Proteobacteria in CMS1. A more detailed analysis of bacterial taxa using non-human RNA-sequencing reads uncovered distinct bacterial communities associated with each molecular subtype.



Relative abundance of bacterial phyla in samples grouped by (**a**) consensus molecular subtype (CMS), (**b**) histological tumour differentiation, and (**c**) location of tumour.

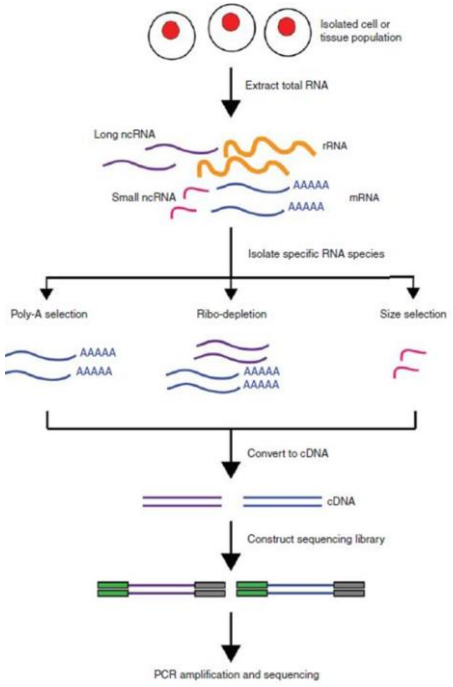
HELSE MIDT-NORGE

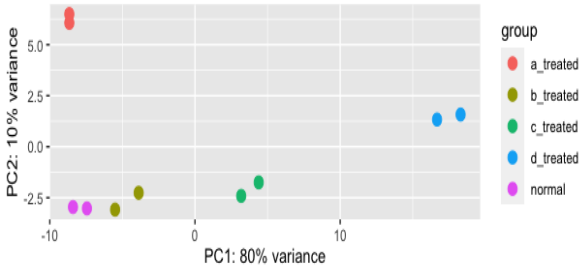
Detecting transcriptome changes with RNA sequencing



- RNA-Seq is revolutionizing the study of the transcriptome. A highly sensitive and accurate tool for measuring expression across the transcriptome, it is providing visibility to previously undetected changes occurring in disease states, in response to therapeutics, under different environmental conditions and across a broad range of other study designs.

- RNA-Seq allows us to detect both known and novel features in a single assay, **enabling the detection of transcript isoforms, gene fusions, single nucleotide variants, allele-specific gene expression** and other features without the limitation of prior knowledge.





HELSE MIDT-NORGE

NTNU

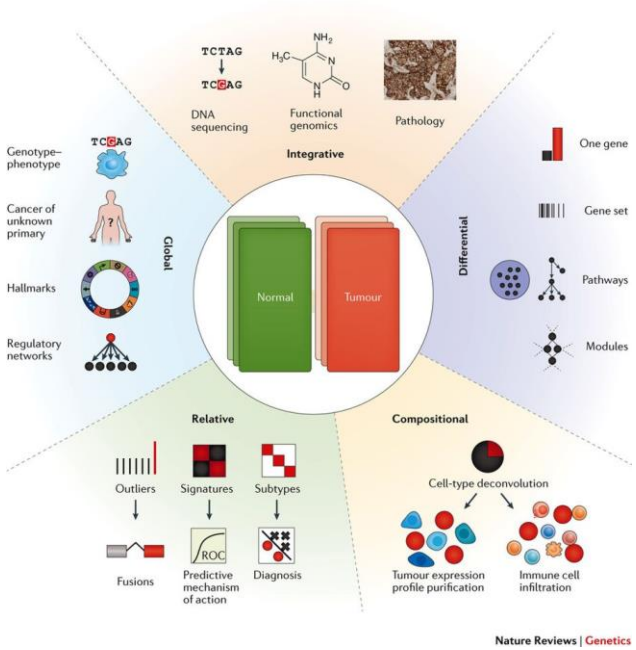
nature REVIEWS GENETICS

Review Article | Published: 27 December 2017

Cancer transcriptome profiling at the juncture of clinical translation

Marcin Cieślik & Arul M. Chinnaiyan

Nature Reviews Genetics **19**, 93–109 (2018) | Download Citation



Analyses of transcriptomic data fall into five broad categories:

Differential analyses focus on the differences between tumour and normal tissues at the gene, gene set, pathway or network level; they require at least two groups of paired or unpaired samples.

Relative analyses compare a single sample or a group of samples with the whole cohort and attempt to identify transcriptional outliers that are clinically useful signatures or subtypes.

Compositional analyses leverage the gene expression signatures of different cell types to assess (or control for) tumour cell purity, to deconvolute samples into constituent tumour and non-tumour cell types and to characterize immune infiltration.

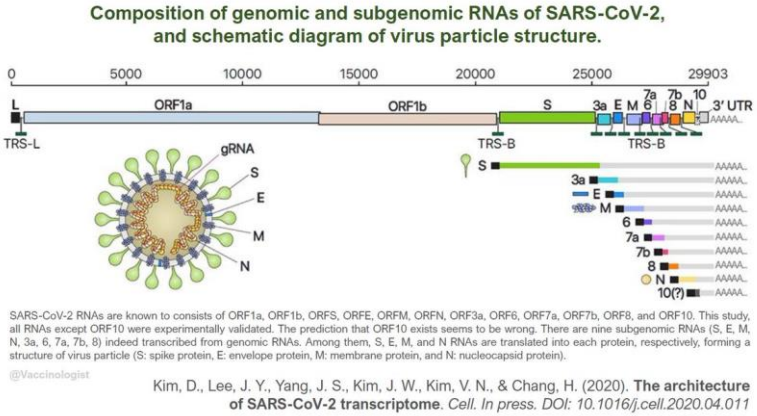
Global analyses compare a sample to a large reference compendium (often pan-tissue or pan-cancer) in order to characterize broad transcriptomic features, such as the accretion of cancer hallmarks, primary tissue type (if unknown) or genotype–phenotype relationships.

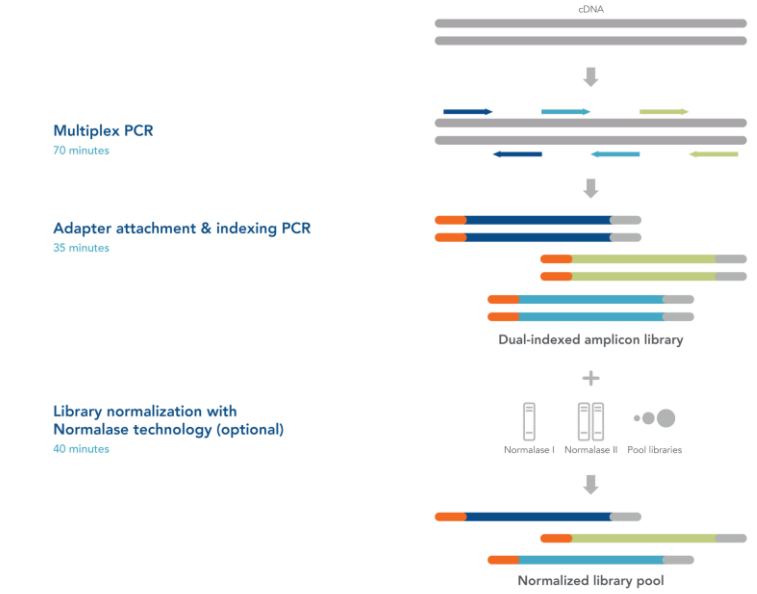
Integrative analyses attempt to supplement transcriptomic data with other data, such as DNA sequencing, functional genomics (for example, DNA CpG methylation) or clinical data (for example, pathology).

HELSE MIDT-NORGE

Targeted RNA Sequencing

- Genomic sequencing of SARS-CoV-2
  - Sequencing enabled the world to rapidly identify SARS-CoV-2 and develop diagnostic tests and other tools for outbreak management. Continued genome sequencing supports the monitoring of the disease's spread and evolution of the virus. Accelerated integration of genome sequencing into the practices of the global health community is required if we want to be better prepared for the future threats.



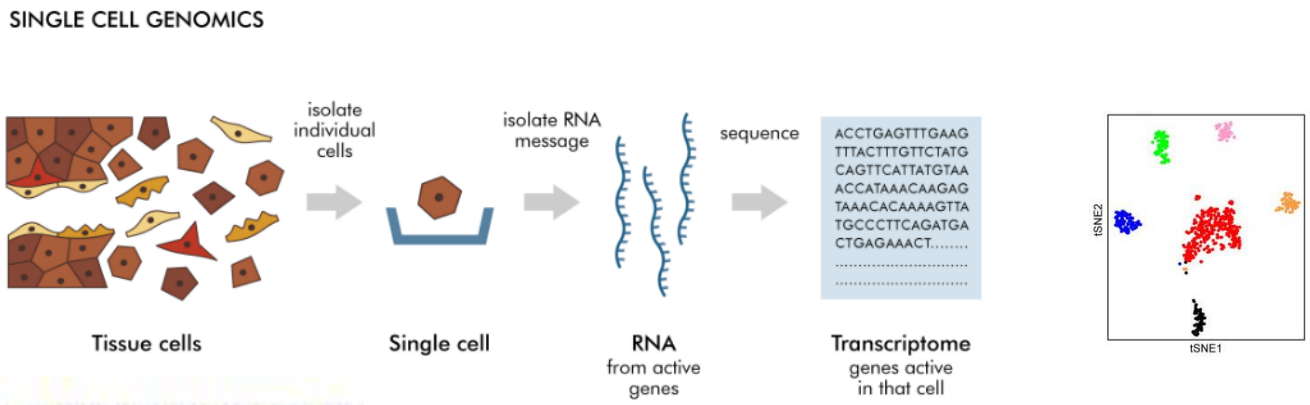


Single-Cell RNA-Seq

- Use RNA-Seq to examine the signals and behavior of a cell in the context of its surrounding environment. This method is advantageous for biologists studying cell function in time-dependent processes such as differentiation, proliferation, and tumorigenesis.

- RNA-Seq methods now enable gene expression analysis of very low-input samples and even single cells. **Single-cell sequencing is a method that examines the genomes or transcriptomes of individual cells, providing a high-resolution view of cell-to-cell variation.**

- With single-cell RNA-Seq, you can explore the distinct biology of different cells within an organ or a tumor, and understand subpopulation responses to environmental cues. These methods are advantageous for biologists studying cell function and heterogeneity in time-dependent processes such as differentiation, proliferation, and tumorigenesis.

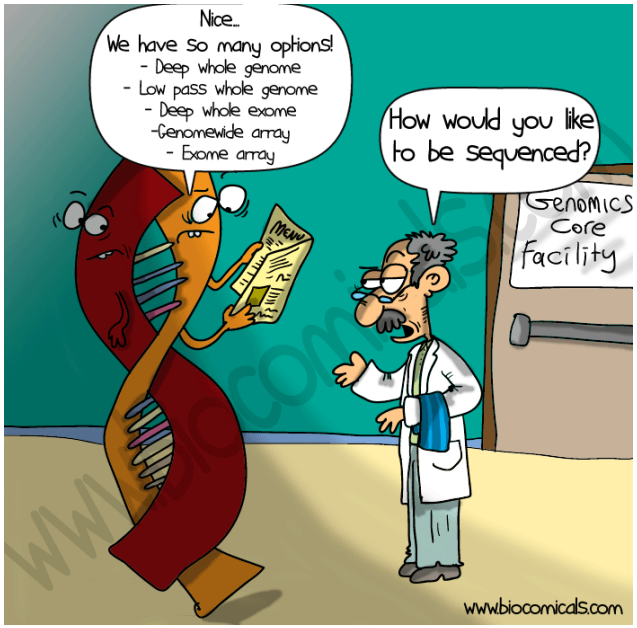


HELSE MIDT-NORGE

NTNU

Final Thoughts

- High throughput sequencing (HTS) technology is highly versatile, flexible and useful technology.
- HTS is becoming vastly faster and more affordable. 1000$ genome is now a reality.
- HTS will not only be an important research tool, it will also be common in the clinic.
- The border between research and clinic will be blurred.
- HTS will be an essential tool in other fields than medicine.
- Generating data is no longer the bottleneck, understanding it is.
- A close collaboration between project PI`s, lab personnel and bioinformaticans is important.



HELSE MIDT-NORGE

NTNU