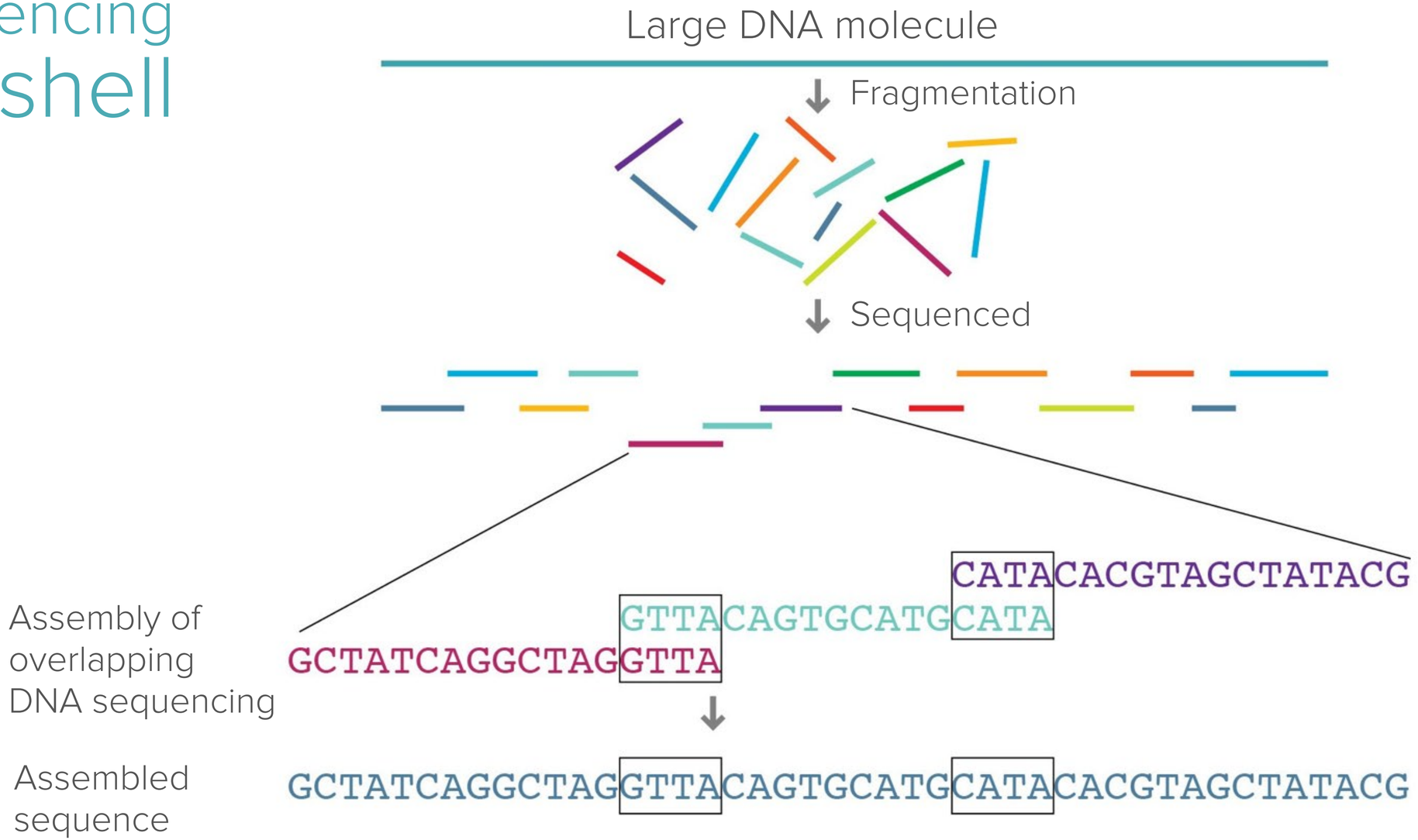
What is Genome Sequencing?

Whole genome sequencing (also known as WGS) – is a laboratory process that determines the complete DNA sequence of an organism's genome at a single time.

In other words:

Genome sequencing is a biotechnology that allows to figure out the order of DNA nucleotides, or bases, in a genome—the order of As, Cs, Gs, and Ts that make up an organism's DNA. The human genome is made up of over 3 billion of these genetic “letters”.

DNA sequencing in a nutshell



Courtesy: National Human Genome Research Institute

Sequencing Technologies

Current and emerging (3rd generation) platforms

First generation Sanger's Method of DNA sequencing

- The dideoxy or chain termination method - developed by Fred Sanger in 1977
- Additional links: http://www.youtube.com/watch?v=oYpIlbl0qF8&feature=related
  http://www.youtube.com/watch?v=UT9wqaVCH5s&NR=1

Second generation of sequencing platforms

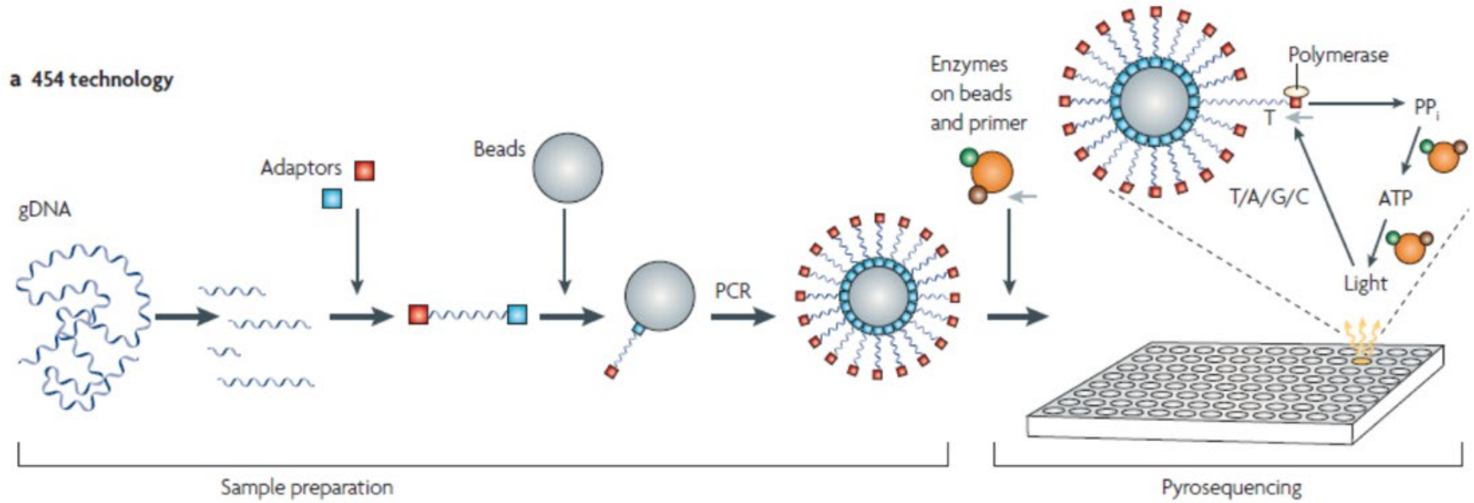
- Also known as Next Generation Sequencing or NGS
- Was developed around 2005 by the 454 Pirosequencing method,
  as well as by SOLiD (ABI) and by Solexa (Illumina) technologies

Next (Second) Generation DNA Sequencing

The 454 method

Overview

454 machines are able to read one Gigabase of DNA sequence in a just couple of days



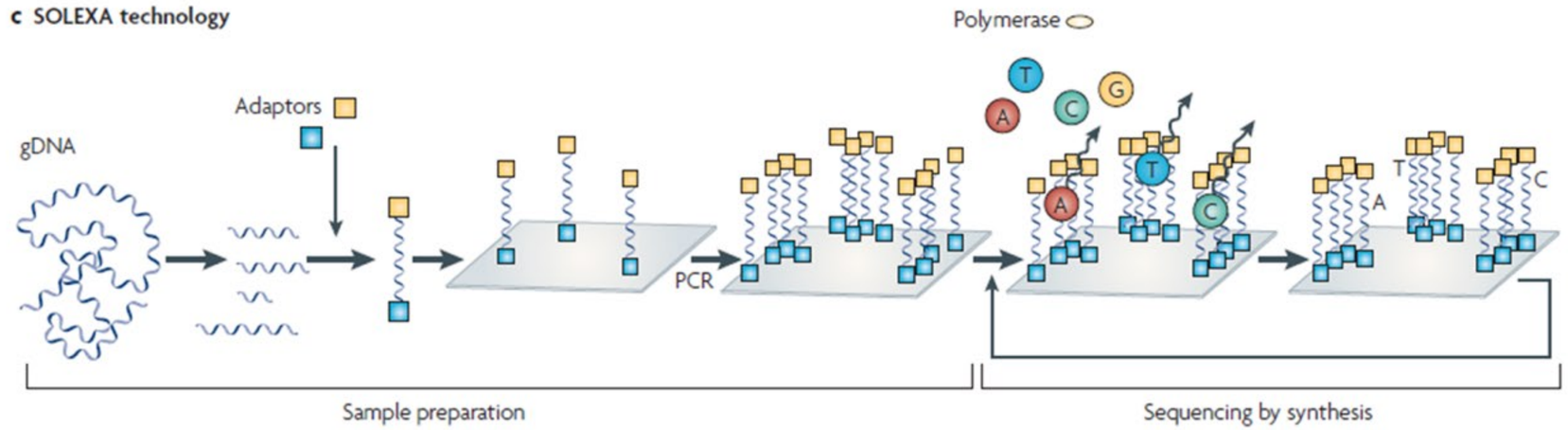
- First 2nd generation technology - 2005
- Can do long reads (400bp)
- Suitable for de novo genome sequencing of moderately sized genomes

Next (Second) Generation DNA Sequencing

The Illumina method

Overview

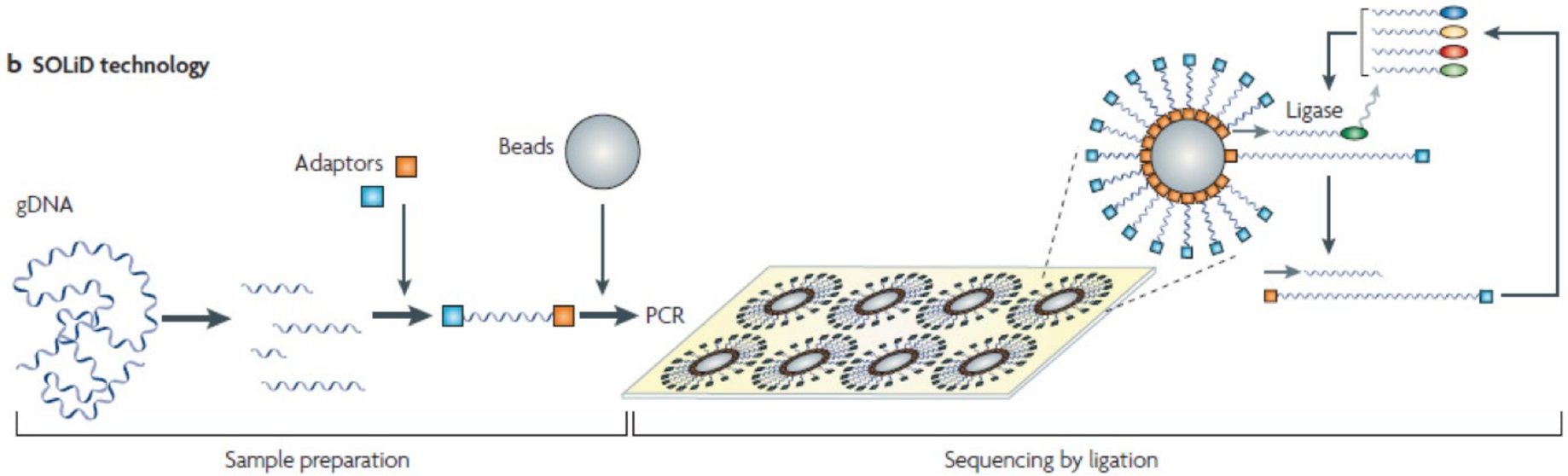
Illumina takes just half a day to read one gigabase and can run more samples simultaneously and significantly cheaper than the 454



- 20-200 gigabases per run (depending on instrument)

Next (Second) Generation DNA Sequencing

SOLiD (ABI)

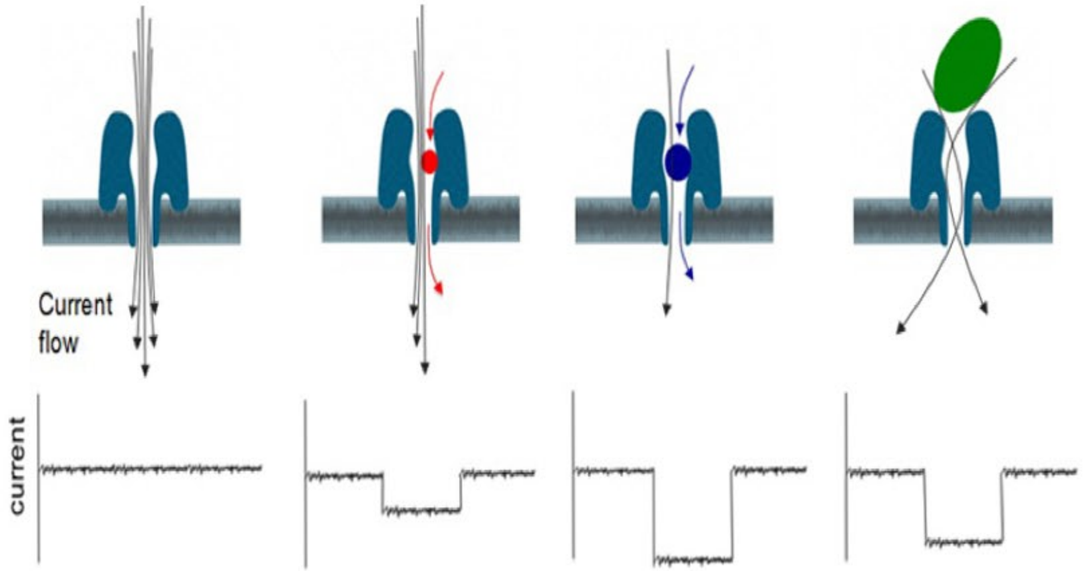


- 60 gigabases per run
- “Emulsion PCR”
- Sequencing by ligation
- Colorspace output

What's the latest NGS technology?

New 3rd Generation Technology:

Nanopore Sensing Oxford Nanopore Technologies



http://www.nanoporetech.com/news/movies#movie-24-nanopore-dna-sequencing

Sequencing Applications

Experimental Class

Whole Genome
(de novo, Resequencing)

Targeted Resequencing

Transcriptomics
(Expression, RNAseq)

Methytation

ChipSEQ

Metagenomics

.....(many more)

Analysis Type

de novo Assembly

Large Scale Variants
(CNV, Rearrangements)

Small Scale Variants
(SNPs, Indels)

Quantitative Profiling
(Counting)

Splice Variants

New... (e.g. WTSS)

.....(many more)

Bioinformatics Methods and Tools for NGS Data Analysis

1 • Customized Pipelines for alignment, assembly and variance search based on open source community supported software packages such as Tuxedo Suite; RSEM ; EdgeR and many others

2 • Galaxy Library for NGS analysis

3 • Commercial Tools from Major Vendors as part of NGS services

Bioinformatics Methods and Tools for NGS Data Analysis

4 • Public Data Repositories/Open Source Tools

1 • Sequencing Read Archive (SRA) http://www.ncbi.nlm.nih.gov/sra
makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms

2 • Regulome Explorer http://explorer.cancerregulome.org/

3 • Cancer Genomics Browser https://genome-cancer.ucsc.edu/

4 • International Cancer Genome Consortium http://dcc.icgc.org/

5 • Genomics Data Commons https://gdc.nci.nih.gov/

Bioinformatics Methods and Tools for NGS Data Analysis

Examples of Analysis workflows

- Whole Transcriptome Analysis using the SOLiD™ Sequencing System
  http://appliedbiosystems.cnpg.com/Video/flatFiles/1116/index.aspx
- View RNA-Seq data in Integrated Genome Browser
  http://www.youtube.com/watch?v=39yPoBEf9oI&feature=related

To Summarize: What is Genome Sequencing?

Genome sequencing - a biotechnology that allows to figure out the order of DNA nucleotides, or bases, in a genome - the order of As, Cs, Gs, and Ts that make up an organism's DNA. The human genome is made up of over 3 billion of these genetic letters.

1. Mutations could cause a disease
2. Gene mutations were studied one-by-one
3. Human Genome Project
   - completed in 2003
   - provided Important Reference to study mutations
4. Clear need for comprehensive profiling of mutations
5. In clinical practice - Sequencing applications for Personalized Medicine