Genome & Transcriptome studies using Next Generation Sequencing (NGS) Technologies : Past and Present

Fatma Guerfali, PhD
Institut Pasteur de Tunis
fatma.guerfali@gmail.com

www.pasteur.tn

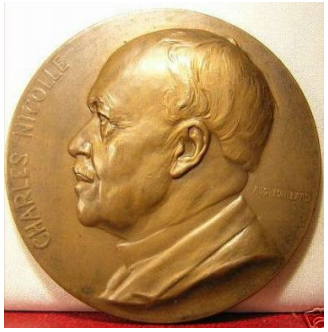
Institut Pasteur de Tunis

1893 – 3rd IP founded

Over a century at the service of Public Health

Vaccination centre

Major discoveries endowed IPT with an uncontested international renown



Charles Nicolle
(Nobel prize in Medicine, 1928)

Major public health impact: Eradication of trachoma, malaria and schistosomiasis

1956 (independence)
Preparation of vaccines
Epidemiology
Diagnostic centre
Education

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

- Next-generation sequencing (NGS), or high-throughput (HT) sequencing = catch-all term describing different modern sequencing technologies used by different platforms.

- Many variations

  - ‘DNA-Seq’ and other related ‘seq’ technologies allow to cover genome complexity : genomic DNA-Seq, Methyl-Seq, ChIP-Seq, exome sequencing…

INTRODUCTION

BASIC CONCEPTS OF NGS

1

NEED TO UNDERSTAND THE BASICS OF GENOME & TRANSCRIPTOME

2

NEED TO UNDERSTAND THE BASICS OF SEQUENCING TECHNOLOGIES & PLATFORMS

3

REFERENCE

Alignment/mapping

NEED TO UNDERSTAND THE BASICS OF SEQUENCING PROTOCOLS & FILE FORMATS

Assembly

OVERVIEW

PARTS 1 - 3

Part 1 Genomes and Transcriptomes : From central dogmas to ongoing discoveries

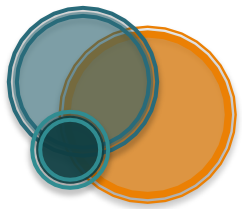
(Central dogma of molecular Biology, Human Genome Project, GENCODE, ENCODE…)

Part 2 Advances in Sequencing Technologies

(Examples of technologies developed for Closed/Open systems for gene expression analysis, Next-Generation Sequencing platforms and technologies …)

Part 3 Overview of NGS (DNA / RNA Seq) Protocols and related file formats

(Overview of protocols for Genomic and Transcriptomic analysis of standard samples using Next-Generation Sequencing and overview of the different formats generated at each step…)



OVERVIEW

PARTS 1 - 3

Part 1 Genomes and Transcriptomes : From central dogmas to ongoing discoveries

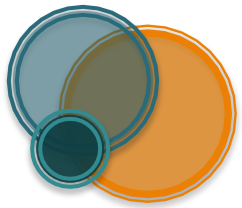
(Central dogma of molecular Biology, Human Genome Project, GENCODE, ENCODE…)

Part 2 Advances in Sequencing Technologies

(Examples of technologies developed for Closed/Open systems for gene expression analysis, Next-Generation Sequencing platforms and technologies …)

Part 3 Overview of NGS (DNA / RNA Seq) Protocols and related file formats

(Overview of protocols for Genomic and Transcriptomic analysis of standard samples using Next-Generation Sequencing and overview of the different formats generated at each step…)



GENOMES & TRANSCRIPTOMES

CENTRAL DOGMAS

- Proposed by Francis Crick 1958
- DNA holds the coded hereditary information in the nucleus
- This code is expressed at the ribosome during protein synthesis in the cytoplasm
- The protein produced by the genetic information is what is influenced by natural selection
- If a protein is modified it cannot influence the gene that codes for it. Therefore only one way flow of information was thought to occur:

DNA→RNA→Protein

GENOMES & TRANSCRIPTOMES

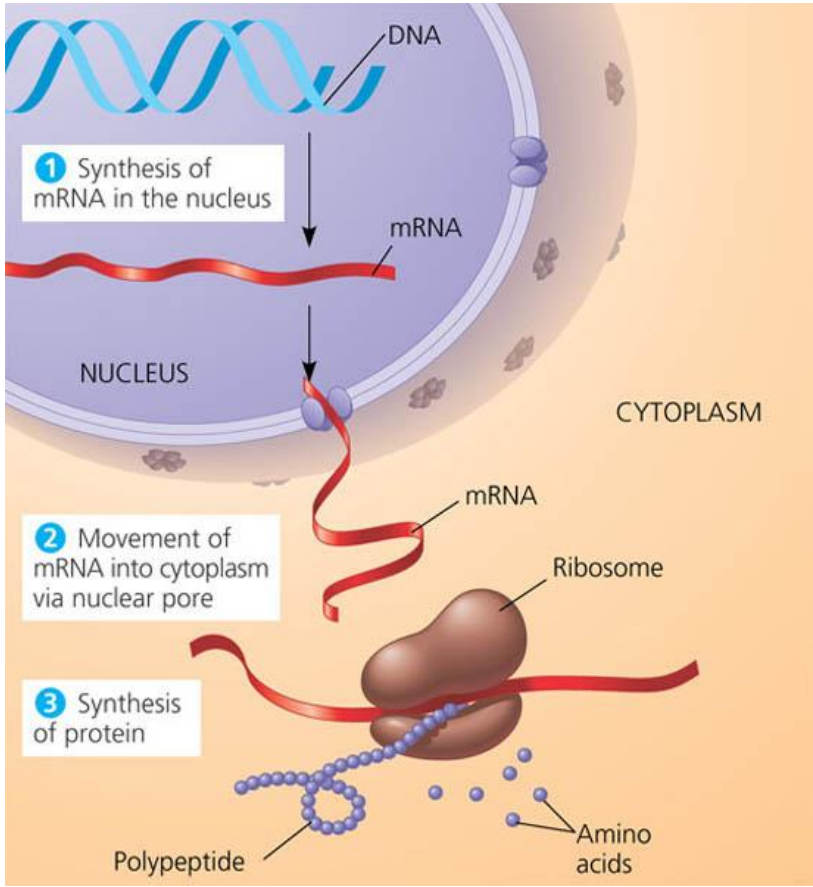
CENTRAL DOGMAS

What is a gene? (definition until less than 10 years ago)

Gene

Is defined as a DNA portion whose corresponding mRNA encodes a protein.

= RNA only considered to be a "bridge" in the transfer of biological information between DNA and proteins



(Costa et al., 2010)

GENOMES & TRANSCRIPTOMES

BASIC CONCEPTS OF GENE AND GENE EXPRESSION

What is a gene ?

Prokaryotes

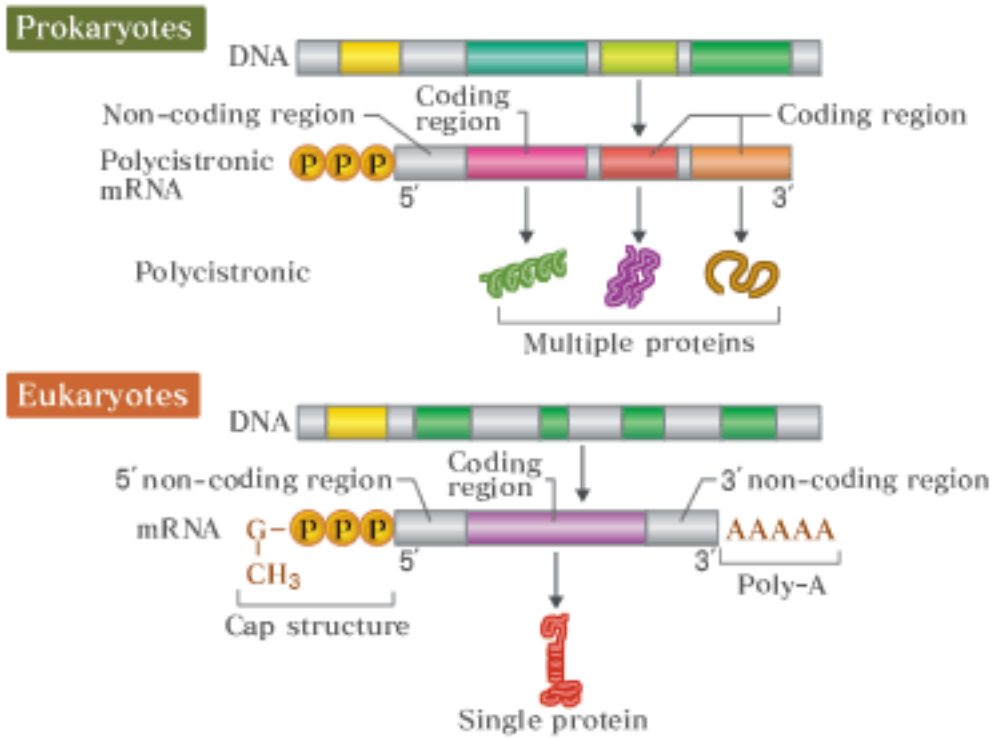
"cistron" = gene

"operon" = a gene unit controlled for gene expression by the same regulating region (e.g. Lactose, Histidine : if present in the medium, the prokarotic cell supress all genes related to its synthesis by the genes in operon).

Eukaryotes

In general: Eukaryotes do not have operons and hence no polycistronic RNA is generated

→ Exceptions !!!! (*Leishmania*)



GENOMES & TRANSCRIPTOMES

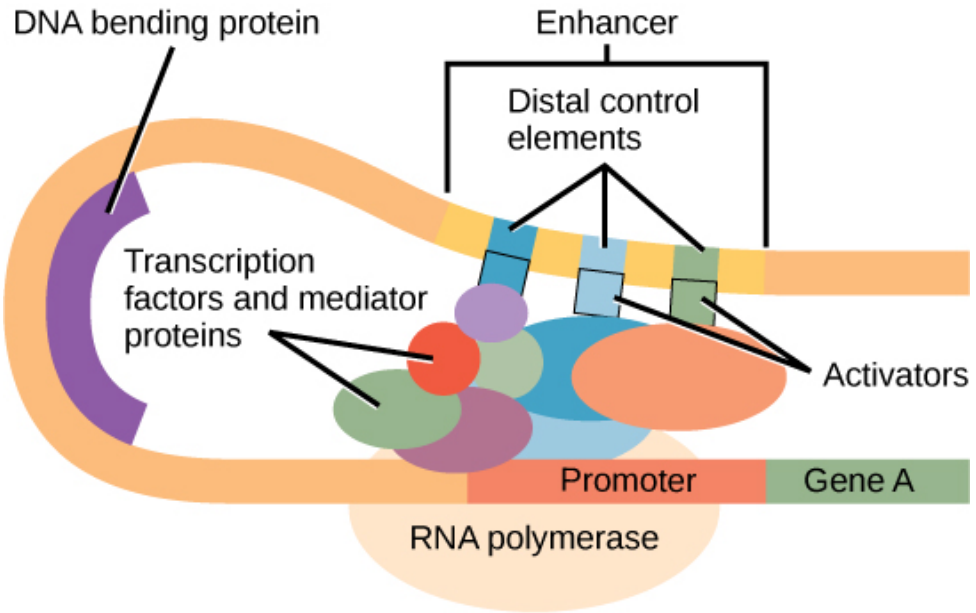
BASIC CONCEPTS OF GENE AND GENE EXPRESSION

What controls gene expression?

Transcription factors (TFs) = proteins that bind to the DNA and help to control gene expression. We call the sequences to which they bind *transcription factor binding sites (TFBSs)*, which are a type of *cis*-regulatory sequence.

There are two main type of *cis*-regulatory elements:

- promoters
- *cis*-regulatory modules (sometimes called "enhancers").

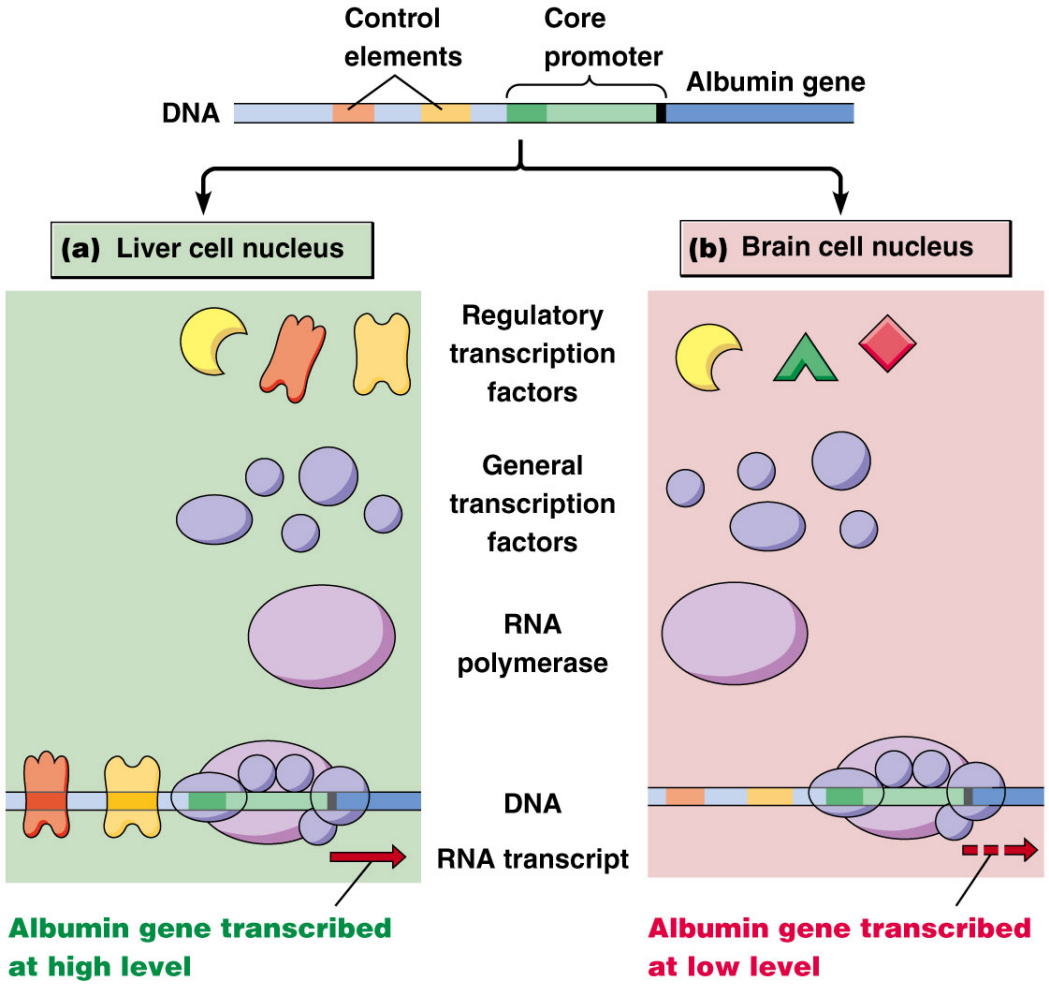


https://www.boundless.com/biology/textbooks/boundless-biology-textbook/

GENOMES & TRANSCRIPTOMES

BASIC CONCEPTS OF GENE AND GENE EXPRESSION

What controls gene expression?



It is the set of regulatory Transcription factors (TFs) that can be different in different tissues

GENOMES & TRANSCRIPTOMES

BASIC CONCEPTS OF GENE AND GENE EXPRESSION

What controls gene expression?

Prokaryotic Gene Regulation

- Regulation of the lac operon (dual control: repression and promotion)
- Regulation of the trp operon (a "riboswitch")

Eukaryotic Gene Regulation

- Genomic Control (DNA rearrangements…)
- Transcriptional Control (TFs, Enhancers…)
- Translational Control (isoforms, dose-dependant synthesis (iron…), siRNAs/ miRNAs…)
- Post-translational Control (active/inactive forms (kinases…), ubiquitin degradation…)

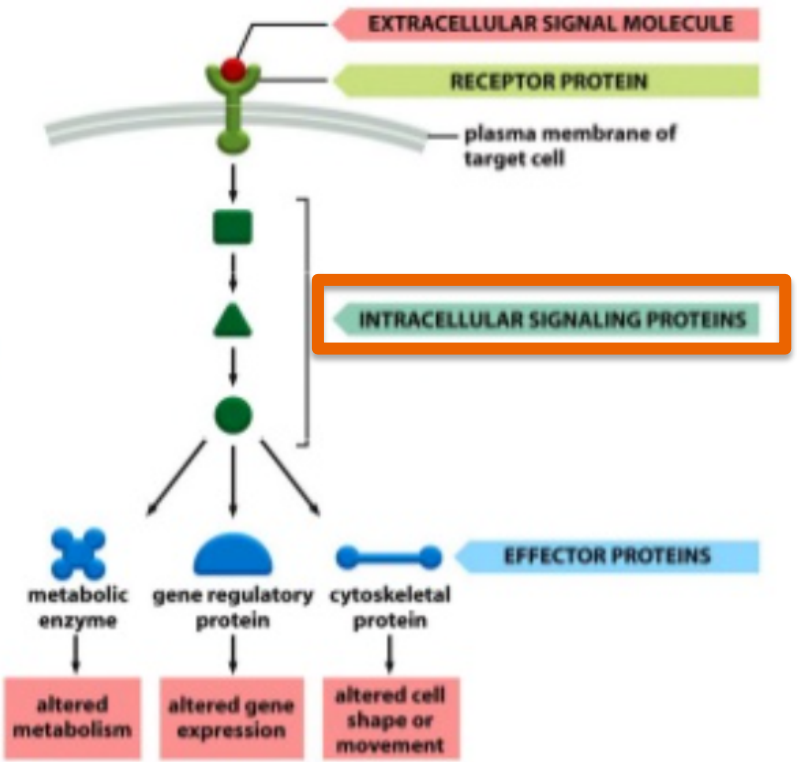
GENOMES & TRANSCRIPTOMES

BASIC CONCEPTS OF GENE AND GENE EXPRESSION

What controls gene expression?

Overview of Cell Signaling

- Sources of extracellular signal
  - Non-cellular environment
  - Cellular environment (cell-cell communication)
  - Hundreds of types of signals
- Cells signaling
  - Stimulus sensing; communication
  - Information processing; decision making
- ↓Receptors
  ↓Signal transducers
  ↓Effector proteins
- Signaling pathways regulate nearly all cellular functions.

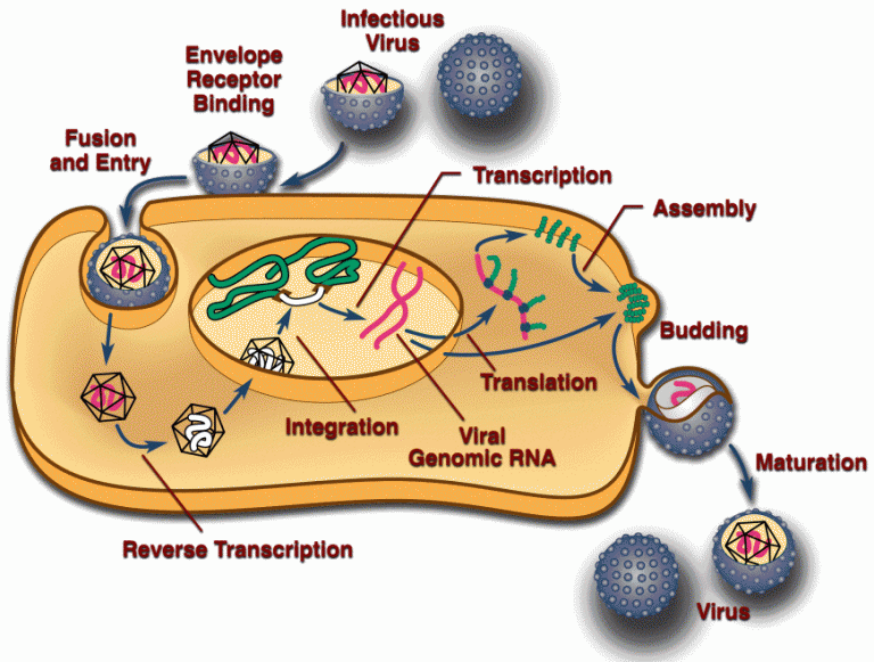


GENOMES & TRANSCRIPTOMES

CENTRAL DOGMAS

•1960s-70s: work on retroviruses changed the "central dogma"
-Temin (Rous sarcoma virus (RSV))
-Baltimore (Rauscher murine leukemia virus (R-MLV))

• **Retroviruses** (e.g. HIV) carry RNA as their genetic information
• When they invade their host cell they convert their RNA into a DNA copy using **reverse transcriptase**



https://www.extremetech.com/extreme/247618-ancient-dna-showcases-war-viruses-hominid-ancestors

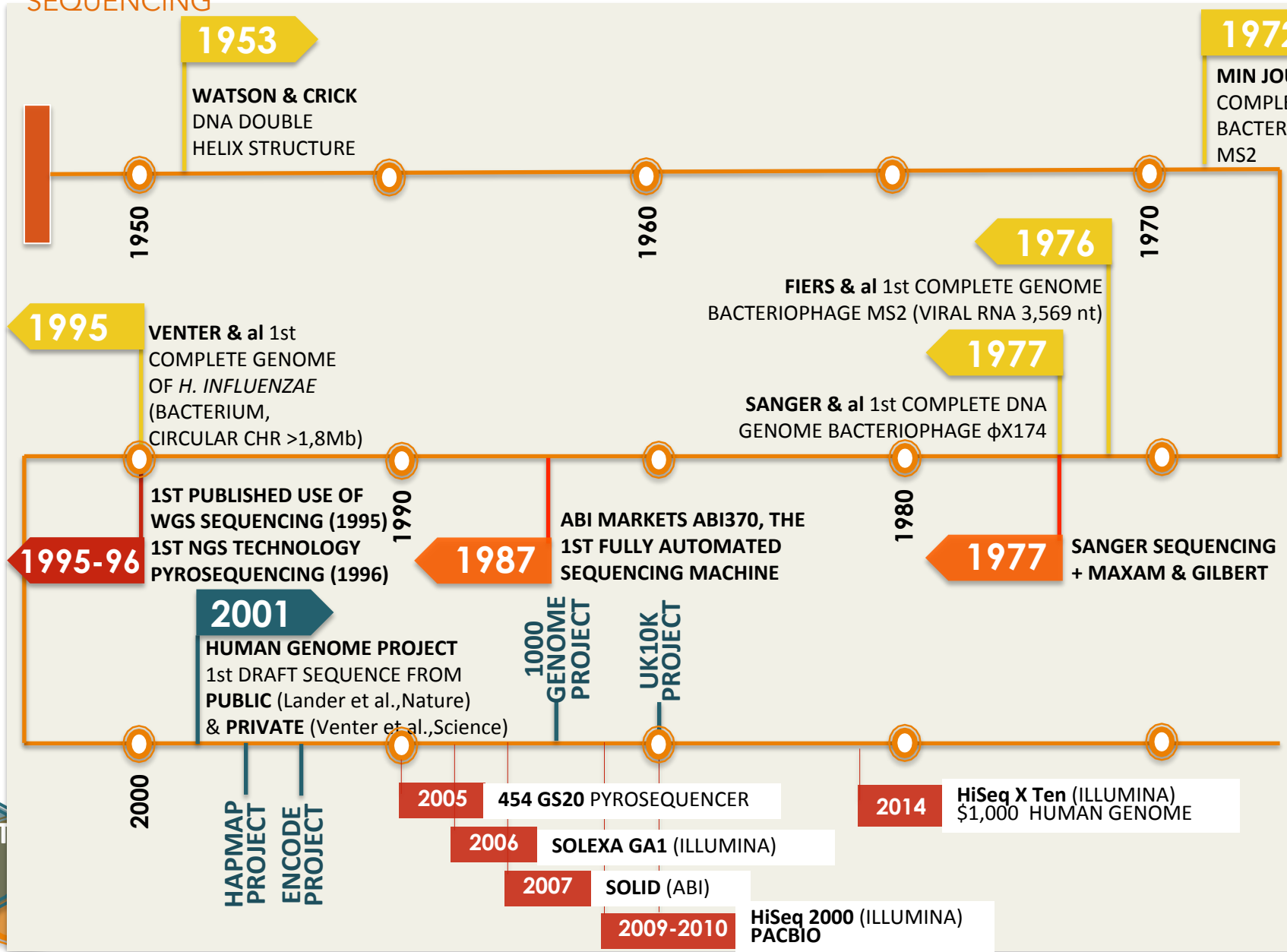
- Thus the central dogma has been modified:

$$DNA \leftrightarrow RNA \rightarrow Protein$$

- This has helped to explain an important paradox in the evolution of life.

GENOMES & TRANSCRIPTOMES

SEQUENCING



GENOMES & TRANSCRIPTOMES

HGP: THE METHODS

- The Human Genome Project (HGP)= a 13-years (1990-April 14, 2003) international effort to sequence the 3 billion "letters" of human DNA.

- $ 3 billion project, led by the U.S. DoE and the NIH.

- International Human Genome Sequencing Consortium (IHGSC)= group of publicly funded researchers

- 20 main sequencing centers in US, UK, Japan, France, China, Germany

http://www.genome.gov/sequencingcosts/
www.sanger.ac.uk

GENOMES & TRANSCRIPTOMES

HGP: THE METHODS

- 2 groups competing for sequencing:
  - Public
  - Private (Celera Genomics)

- Opposing philosophies :

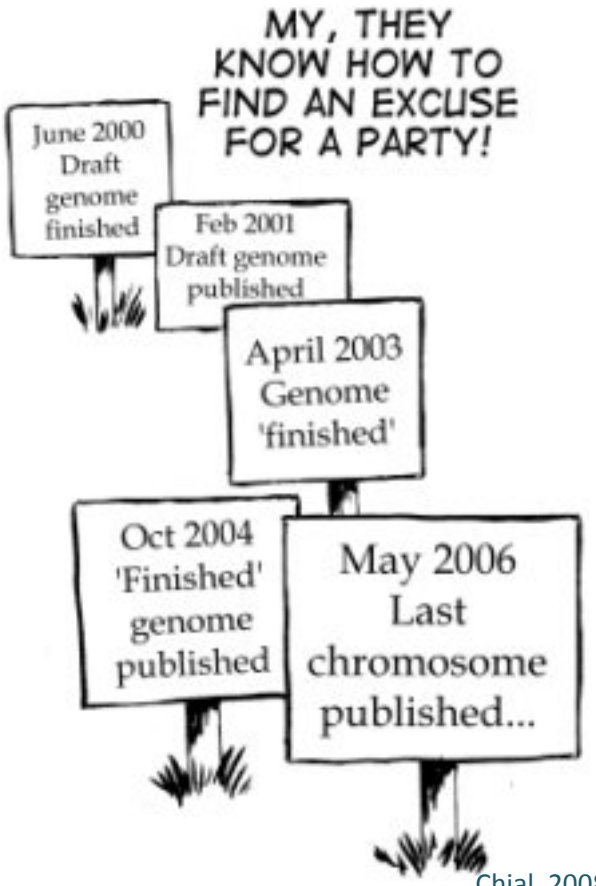
- **Public** : HGP Bermuda Agreement (1996)

→ all information from the project would be made freely available to all within 24h.

- **Private**

→ access restricted to paying customers !

- In February **2001, drafts of the human genome sequence were published simultaneously by both public-private** groups in separate articles (Lander et al (IHGSC)., Feb 2001 Nature) (Venter et al., Feb 2001 Science)



Chial, 2008

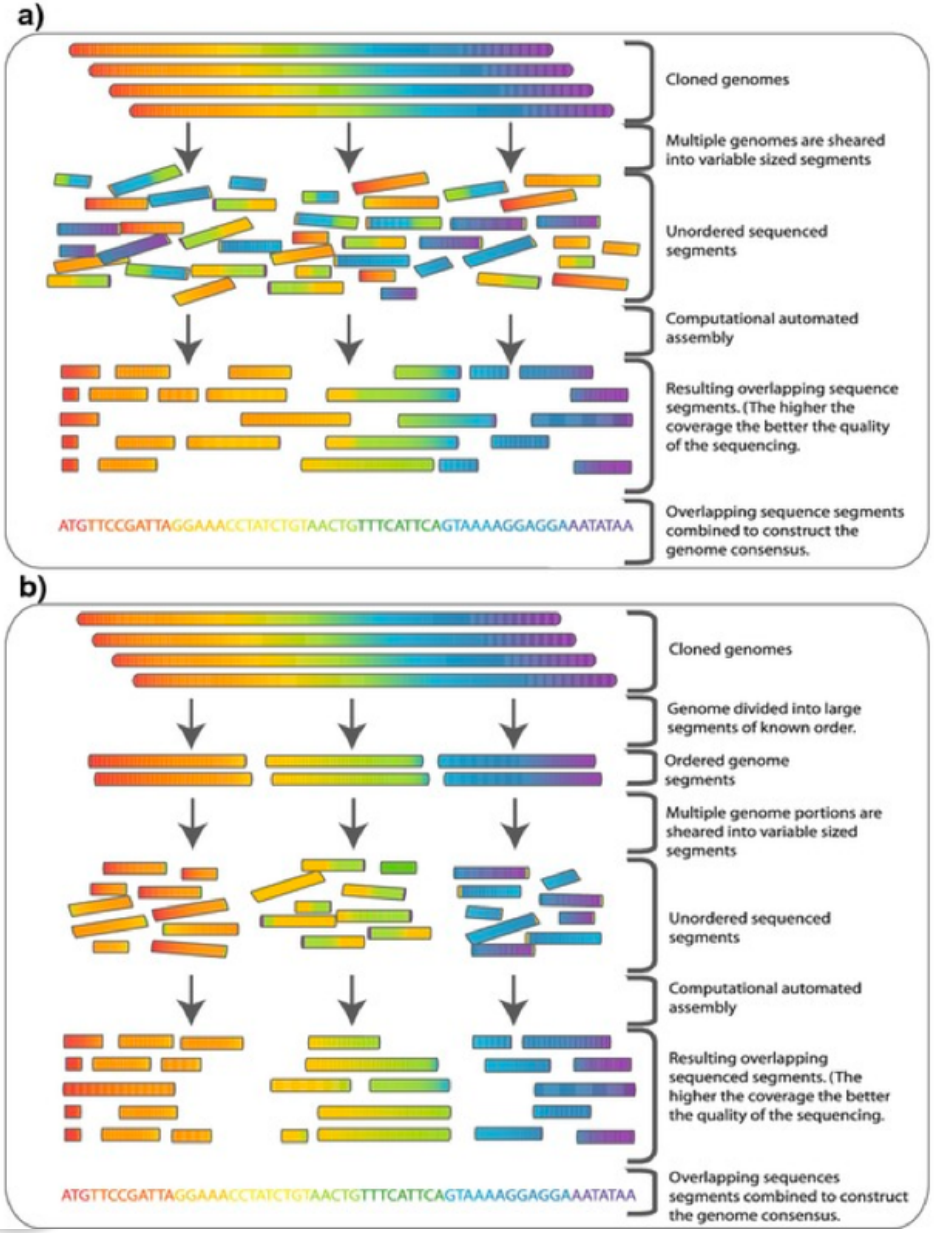
http://www.genome.gov/sequencingcosts/

http://www.yourgenome.org/

www.sanger.ac.uk

GENOMES & TRANSCRIPTOMES

HGP: THE METHODS



WHOLE GENOME SHOTGUN

(Celera genomics, Private)

→ Genome sheared randomly into small fragments (appropriately sized for sequencing)

→ Reassembly

HIERARCHICAL GENOME SHOTGUN (TOP-DOWN SEQUENCING)

→ low-resolution physical map of the genome is made prior to sequencing.

→ Genome fragmented into large segments.

→ Order of these segments is deduced

→ They are further sheared into fragments appropriately sized for sequencing.

→ Assembly relying on the physical map

→ **Finishing phase** : filling in gaps and resolving DNA sequences in ambiguous areas

PART 1

GENOMES & TRANSCRIPTOMES

HGP: THE METHODS

Variable degrees of completion of published genomes

- **Draft Sequencing**
  - high-throughput or shotgun phase (whole genome or clone-based approach)
  - Assembly using specific algorithms (whole-genome or single-clone assembly)

→ lower accuracy than finished sequence; some segments are missing or in the wrong order or orientation.

- **Finishing**
  - Accuracy in bases identification + Quality Check + few if any gaps.
  - Contiguous segments of sequence are ordered and linked to one another
  - No ambiguities or discrepancies about segments order and orientation

- **Complete Genome**

A Genome represented by a single contiguous sequence with no ambiguities

→ The sequences available are ***finished to*** *a certain high quality.*

(Mardis et al., 2002)
http://www.informatics.jax.org/

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

- The HGP project required that all human genome sequence information be freely and publicly available. The existing DNA sequences have been stored in databases available to anyone willing to exploit and analyze them.

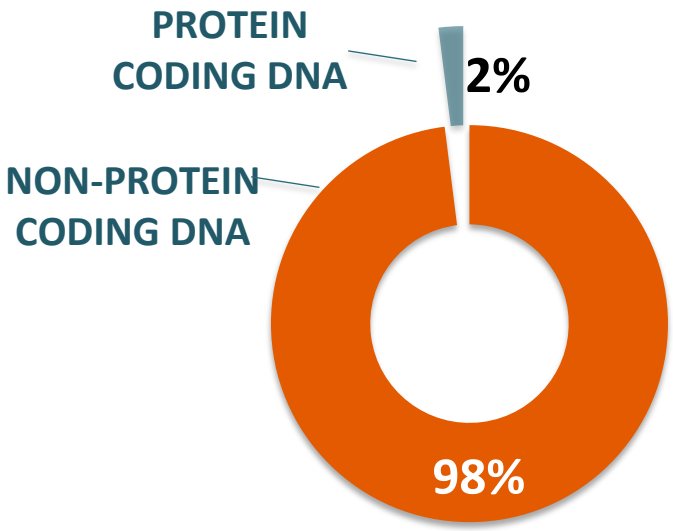
- Dedicated databases house various data for model organisms such as sequences of known and hypothetical genes and proteins (GenBank, NCBI). Other databases (Ensembl http://www.ensembl.org) present additional data and annotation as well as powerful tools for visualizing and searching it.

- Community efforts for non-model organisms like Eukaryotic Pathogens : EuPathDB (http://eupathdb.org/eupathdb/).

Computer programs have been developed to analyze and interpret the data.

GENOMES & TRANSCRIPTOMES

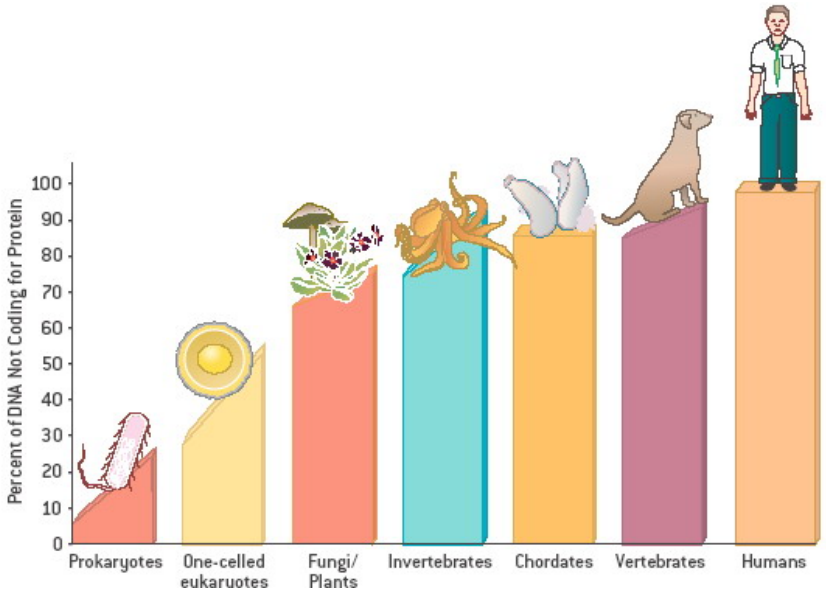
HGP: THE HERITAGE



≈ 20,000

The human genome contains only about 20,000 protein-coding genes : sequence alone is not enough to explain the whole complexity !

The proportion of non-coding DNA increases with organism complexity



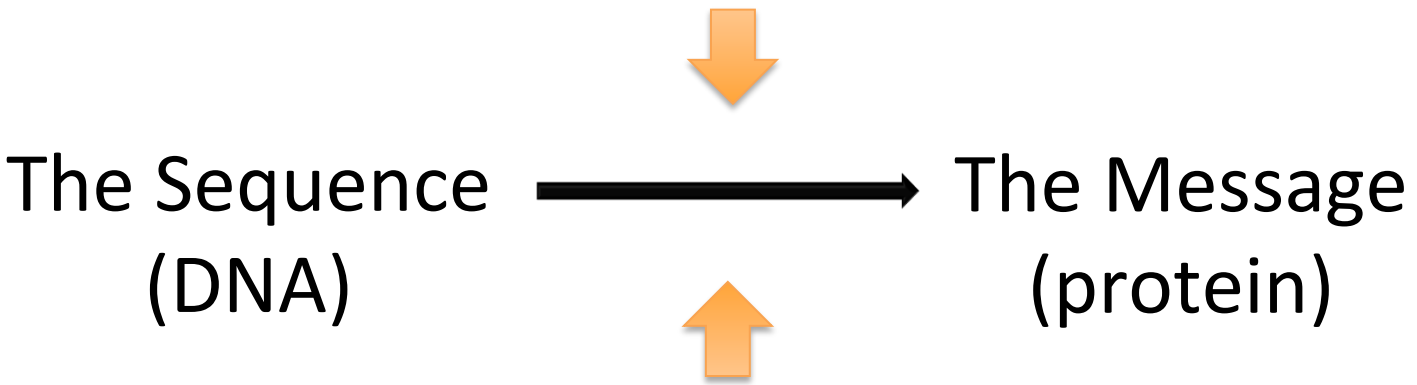
(Mattick, 2011)

http://www.yourgenome.org/stories/how-is-the-completed-human-genome-sequence-being-used

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

The Genome
The Transcriptome
XXXome…



The non-coding = Regulation

ncRNAs
Isoforms
Stability
Methylation
…

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

- *= It appears that the genetic programming of humans and other complex organisms has been misunderstood for the past 50 years, due to the assumption that most genetic information is transacted by proteins.*

- The majority of these sequences are dynamically transcribed, mainly into non-protein-coding RNAs, with tens if not hundreds of thousands that show specific expression patterns and subcellular locations, as well as many classes of small regulatory RNAs.

  Only infrastructural RNAs where known ! (rRNAs, tRNAs,…)

(Mattick, 2011)

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

- The emerging evidence indicates that these **RNAs control** the epigenetic states that underpin development, and that many are dysregulated in cancer and other complex diseases.
- Moreover it appears that animals, particularly primates, have **evolved plasticity in these RNA regulatory systems**, especially in the brain.
- Thus, it appears that what was dismissed as 'junk' because it was not understood holds the key to understanding human evolution, development, and cognition.

(Mattick, 2011)

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

"The Encyclopedia of DNA Elements (ENCODE) Project aims to provide a more biologically informative representation of the human genome by using high-throughput methods to identify and catalogue the functional elements encoded. "

ENCODE = Encyclopedia of DNA Elements

- ENCODE is a public research consortium launched by the US National Human Genome Research Institute (NHGRI) in September 2003.
- **The aim was to find the functional elements of the human genome.**
- All data that have been obtained in this project have been made available in public databases / repositories.

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

Non-Coding: the ENCODE Project

- Among the major key findings:

  - ≈ 20% of ncDNA in the human genome is functional (regulation of gene expression for coding genes).
  - ≈ 60% is transcribed without any known function yet.
  - The expression of each coding gene is under the control of several regulation sites that can be located near or far from the gene = unexpected complexity !

→ Much more complex transcriptional "landscape" than what was previously expected !!!

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

Coding: the GENCODE Project

- The human genome has been the focus of intensive manual annotation: GENCODE

  Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22, 1760–1774 (2012)

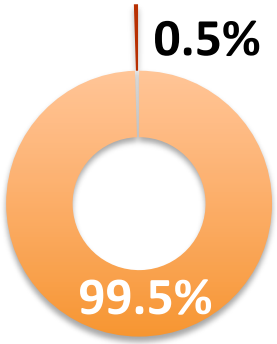
The GENCODE Consortium aims to identify all gene features in the human genome using a combination of:

- computational analysis
- manual annotation
- experimental validation.

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

HAPMAP



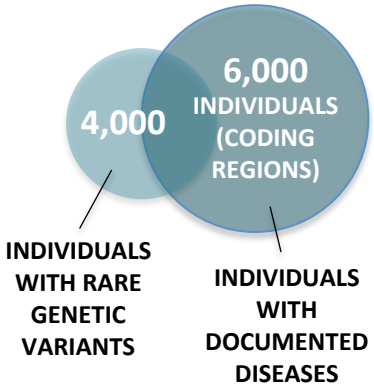
CATALOG OF COMMON GENETIC VARIATION AMONG HUMANS TO IDENTIFY DISEASE-RELATED GENES.

→ **THE DNA SEQUENCE OF ANY TWO PEOPLE IS 99.5% IDENTICAL.**

1,000 GENOMES

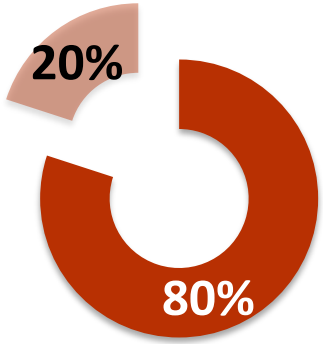
RESOURCE OF HUMAN GENETIC VARIATION

UK10K



LINK BETWEEN GENETIC VARIANTS AND RARE DISEASES

ENCODE



80% OF THE HUMAN GENOME SEQUENCE IS CLOSE TO REGIONS THAT CONTROL BIOLOGICAL FUNCTION.

→ **"JUNK DNA" IN THE HUMAN GENOME, IS ACTUALLY FUNCTIONAL.**

(Mattick, 2011)

http://www.yourgenome.org/stories/how-is-the-completed-human-genome-sequence-being-used

dbSNP: Database for Short Genetic Variations

Catalog of nucleotide changes for human and other model organisms

https://www.ncbi.nlm.nih.gov/snp/

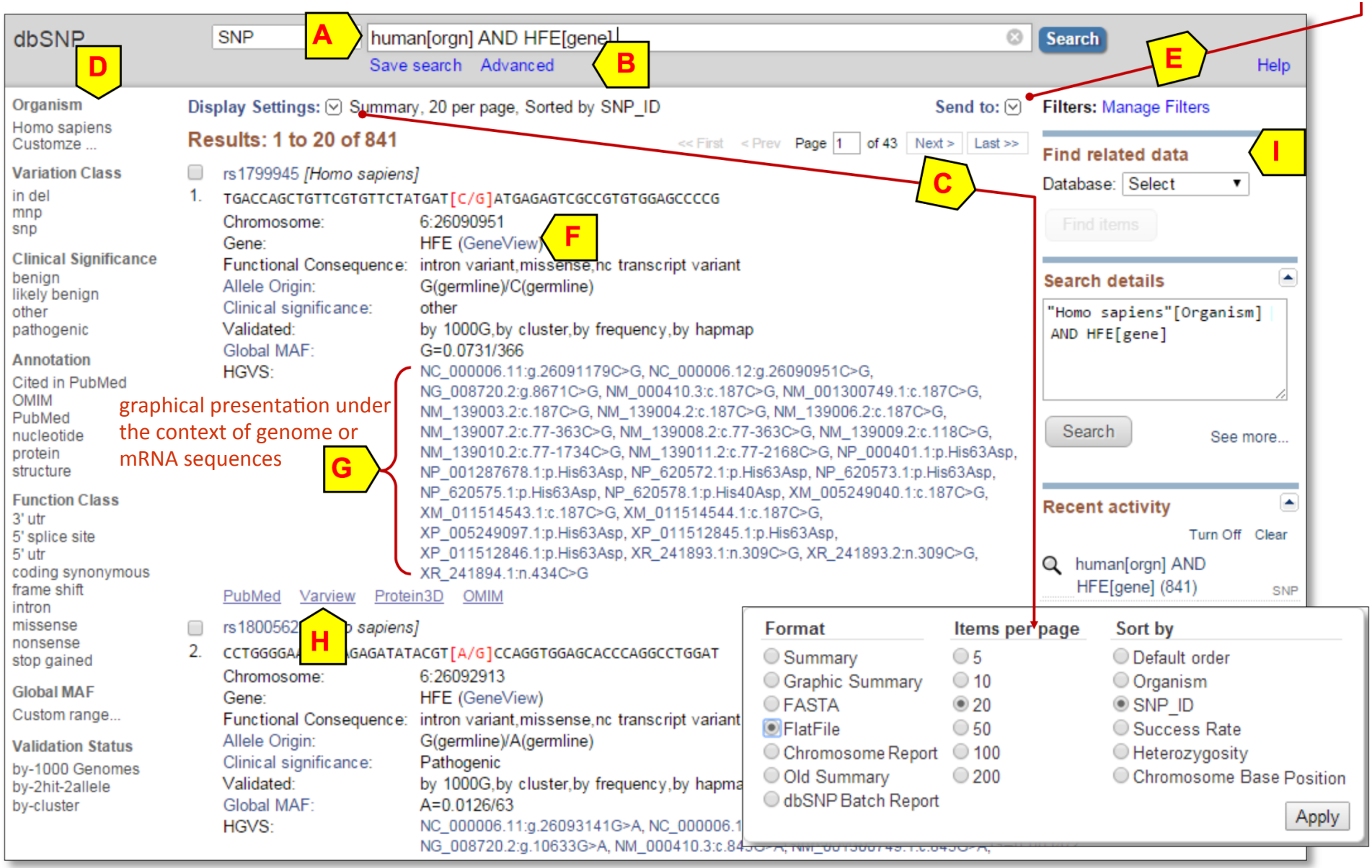
National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope and Access

The NCBI Short Genetic Variations database, commonly known as dbSNP, catalogs short variations in nucleotide sequences from a wide range of organisms. These variations include single nucleotide variations, short nucleotide insertions and deletions, short tandem repeats and microsatellites. Short Genetic Variations may be common, thus representing true polymorphisms, or they may be rare. Some rare human entries have additional information associated with them, including disease associations, genotype information and allele origin, as some variations are somatic rather than germline events.

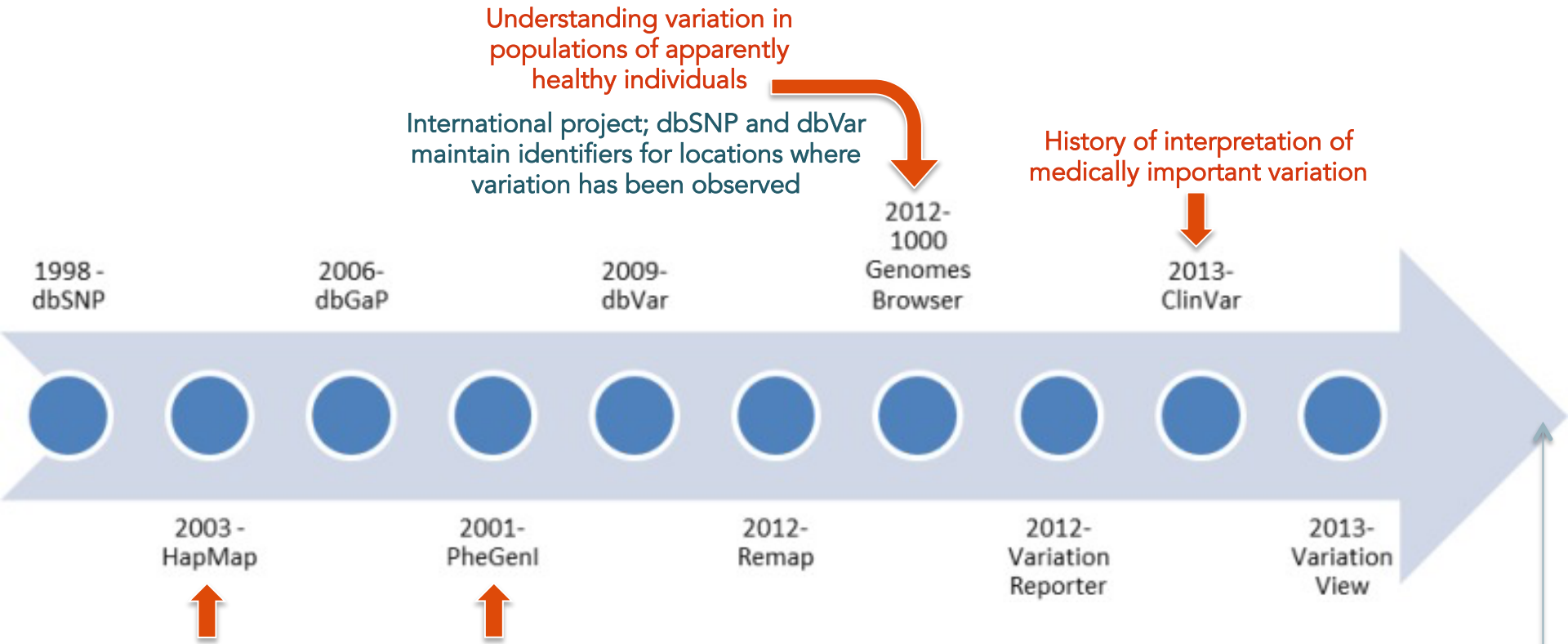
GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE



GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE



September 1, 2017 – dbSNP and dbVar stop accepting non-human variant data submissions
November 1, 2017 – dbSNP and dbVar interactive websites and related NCBI services stop presenting non-human variant data. (…)

Church et al., 2013 (https://www.ncbi.nlm.nih.gov/books/NBK169030/)

GENOMES & TRANSCRIPTOMES

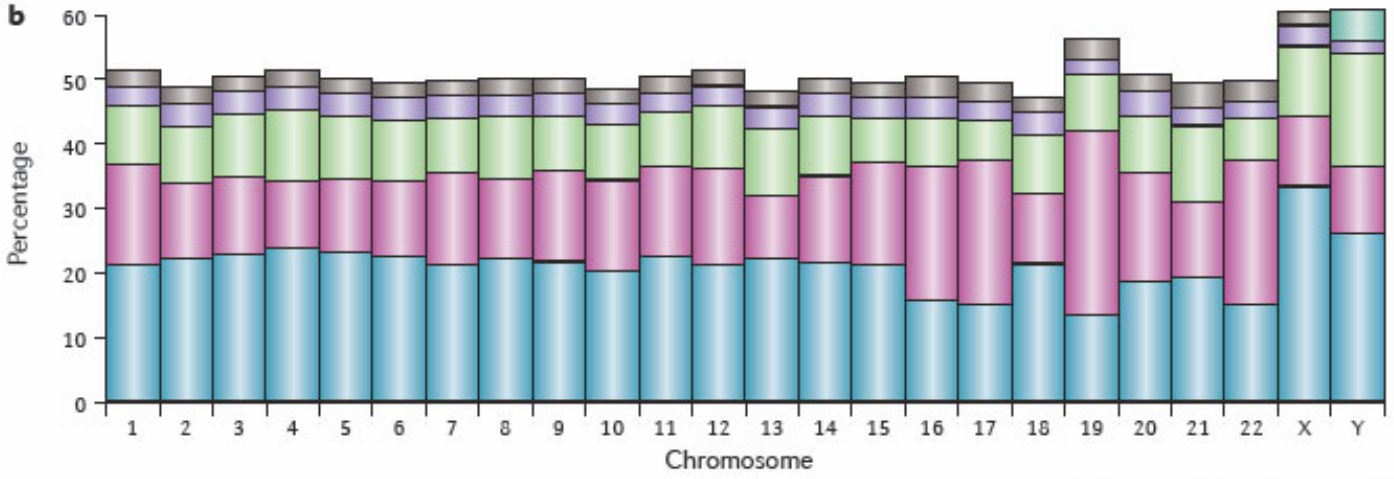
HGP: THE HERITAGE

- About 50% of the Human genome is composed of repeats

a

Repeat class	Repeat type	Number (hg19)	Cvg	Length (bp)
Minisatellite, microsatellite or satellite	Tandem	426,918	3%	2–100
SINE	Interspersed	1,797,575	15%	100–300
DNA transposon	Interspersed	463,776	3%	200–2,000
LTR retrotransposon	Interspersed	718,125	9%	200–5,000
LINE	Interspersed	1,506,845	21%	500–8,000
rDNA (16S, 18S, 5.8S and 28S)	Tandem	698	0.01%	2,000–43,000
Segmental duplications and other classes	Tandem or interspersed	2,270	0.20%	1,000–100,000

b



Nature Reviews | Genetics

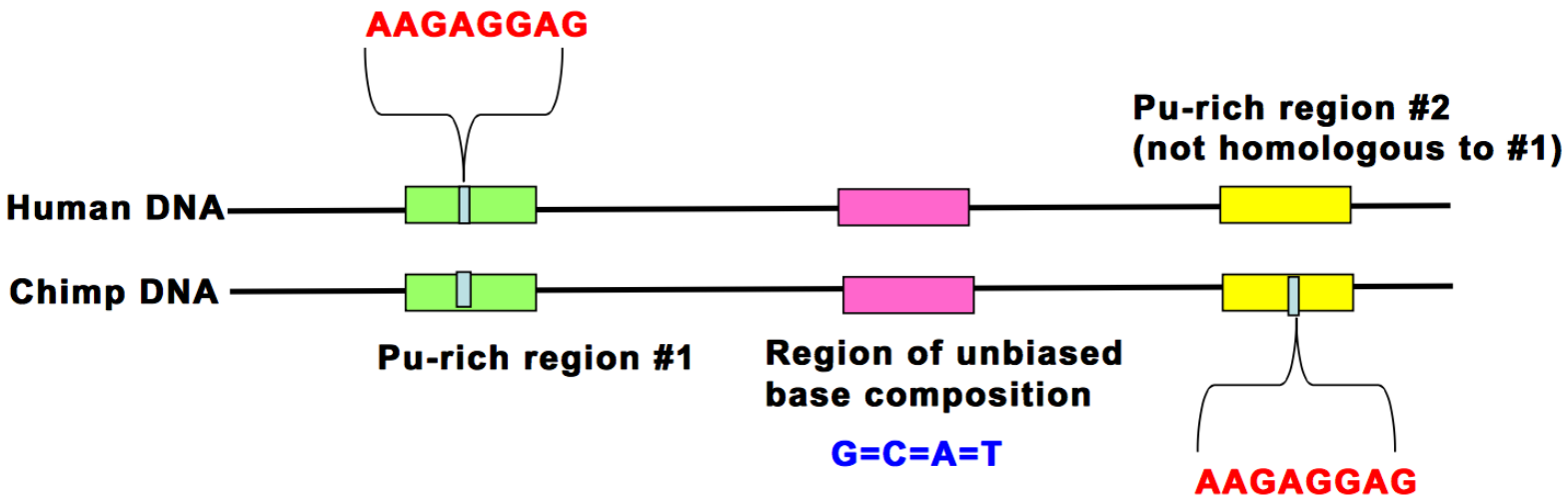
Treangen & Salzberg, 2012

GENOMES & TRANSCRIPTOMES

- Low complexity regions

Molecular structural region of biased composition—e.g.:

- homopolymeric runs
- short-period repeats (simple tandem repeats, polypurine, AT-rich…)
- subtle overrepresentation of one or a few residues
- …



When a sequence is of low complexity (and/or short length), a high % identity with another sequence does not reflect necessarily a shared evolutionary origin.

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

- General Overview
  - Only 2% of the Human genome is coding
  - Gaps in sequences remain
  - Annotated and 1000s of previously unannotated RNAs, including Families (tens to hundreds) of Non-coding RNA : lncRNAs, snmRNAs (microRNAs, snoRNAs, snRNAs), pseudogenes
  - ¾ of the human genome is capable of being transcribed
  - Pseudogenes can be transcribed
  - Over 100,000 splice junctions discovered that were not incorporated into transcript models (strand-specific RNA-seq).

"These observations, taken together, prompt a redefinition of the concept of a gene."

(Adapted from Djebali, S. et al. 2012)

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

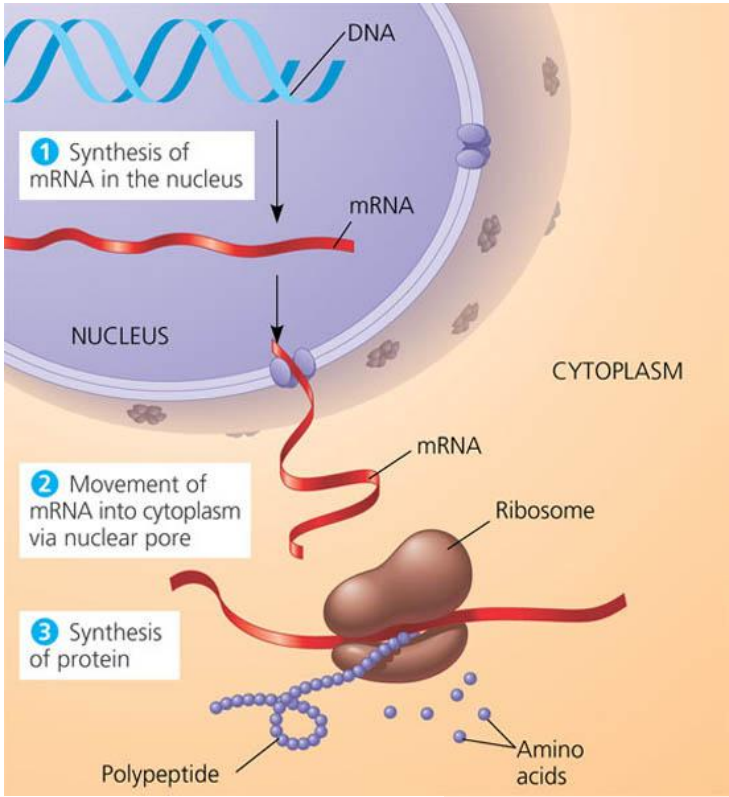
RNA is no longer considered to be a "bridge" in the transfer of biological information between DNA and proteins

Transcriptome consist of

- Coding RNA (mRNA (2–4%)
- Previously known non-coding RNAs
  - rRNA (80–90%)
  - tRNA (5–15%)

- Small fraction of intragenic (i.e., intronic) and intergenic non-coding RNA (1%, ncRNA) with undefined regulatory functions.

NB: Intragenic & Intergenic sequences
- Are enriched in repetitive elements
- have long been considered genetically inert



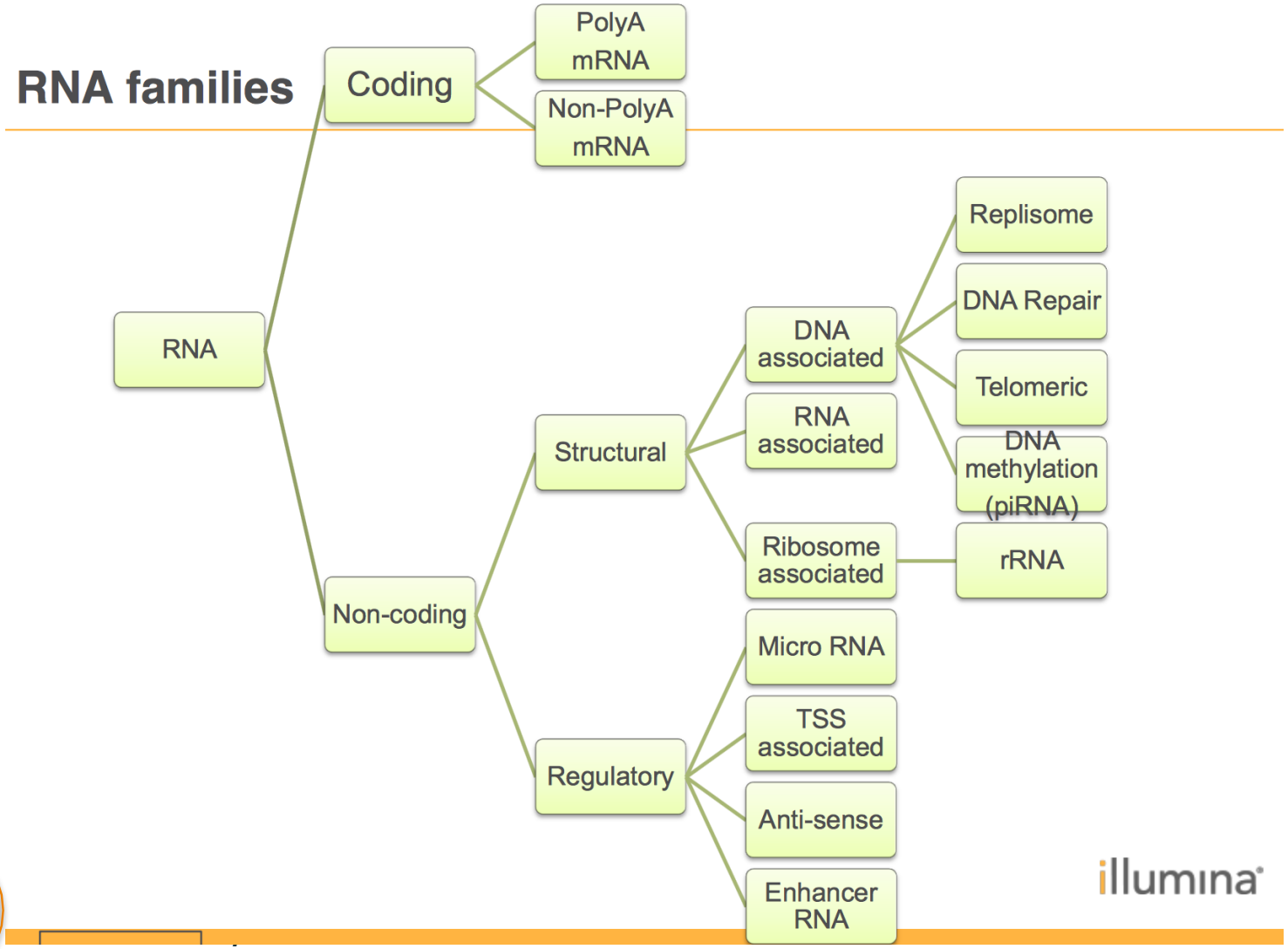
(Costa et al., 2010)

Transcriptome = The complete collection of transcribed elements of the genome (Affymetrix, 2004)

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

An overview of the eukaryotic transcriptome through examples of its products

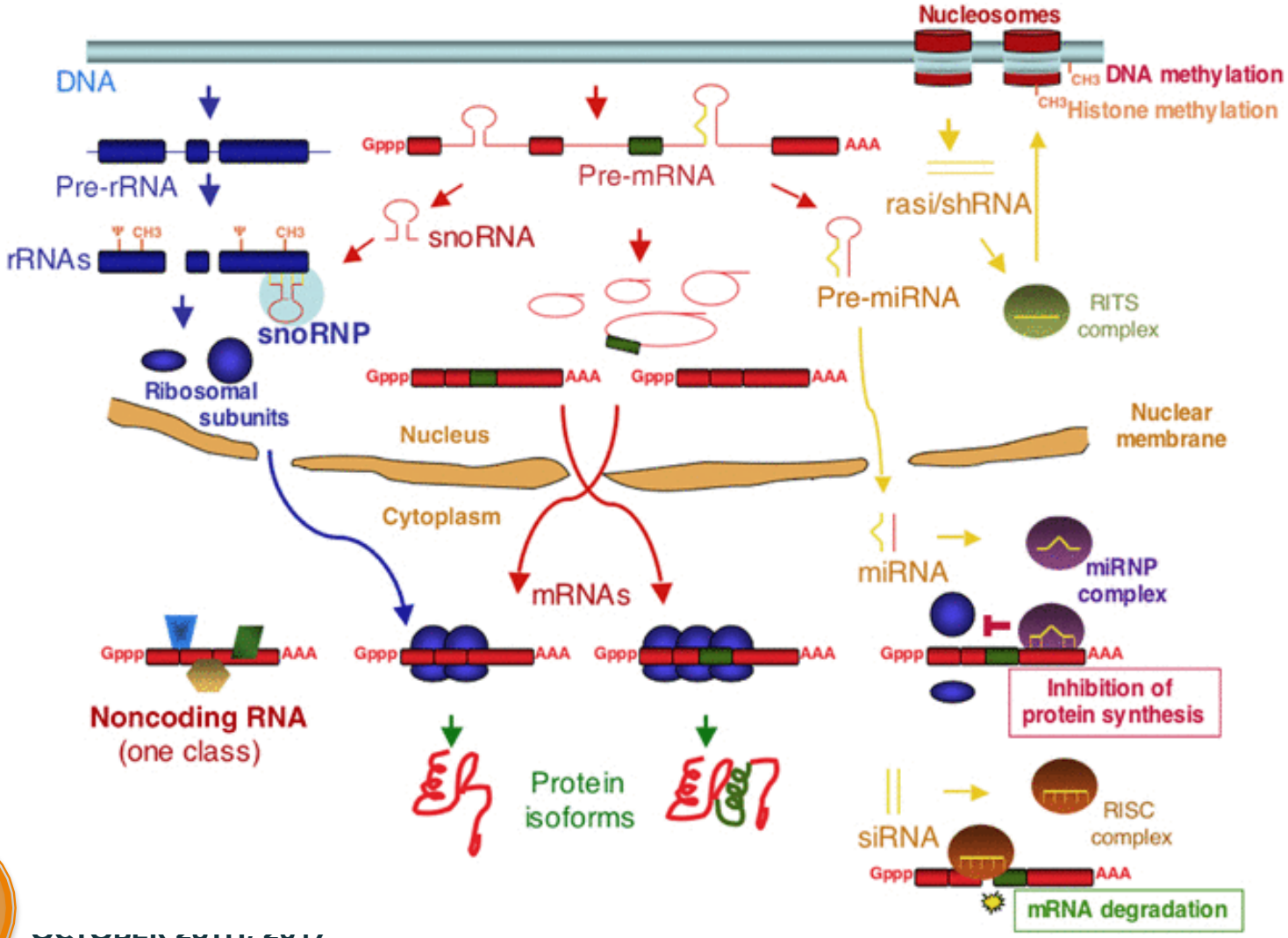


PART 1

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

An overview of the eukaryotic transcriptome through examples of its products



(Soares & Valcárcel, 2006)

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

An overview of the eukaryotic transcriptome through examples of its products

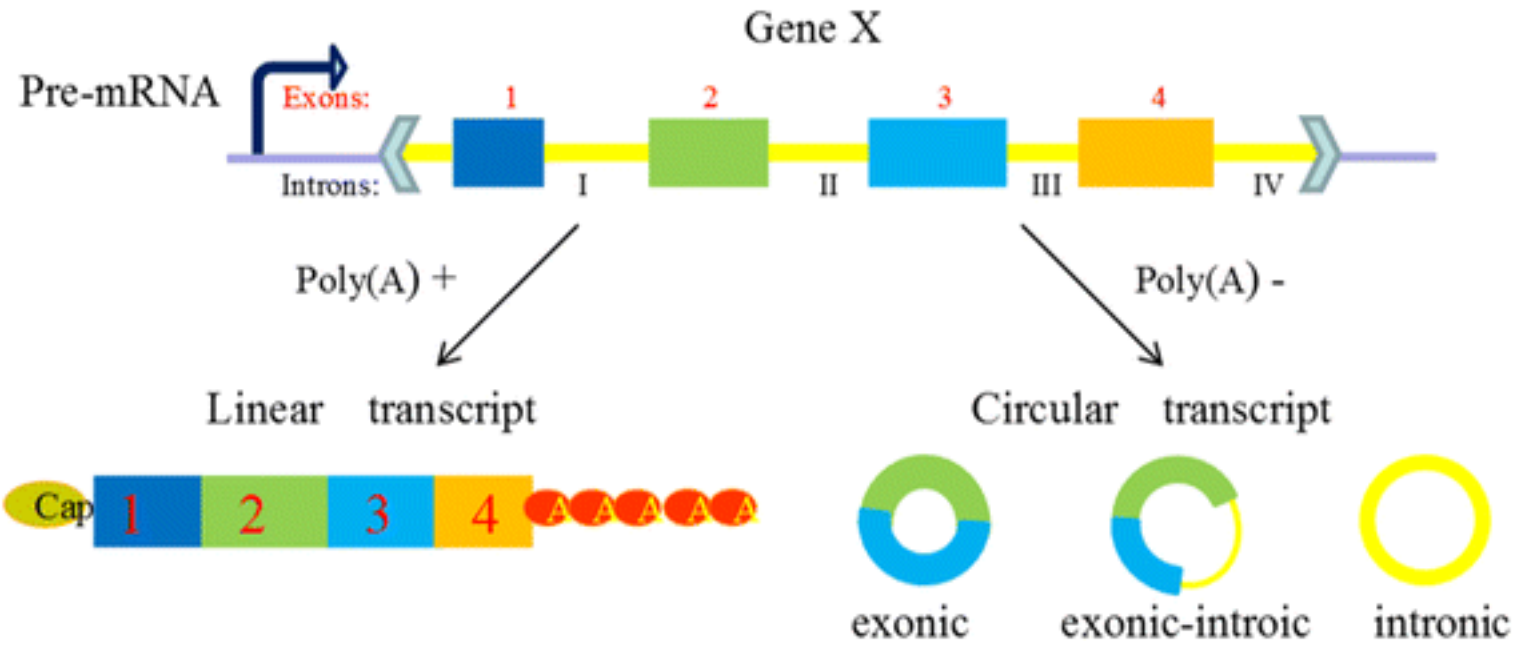
REVIEW | OPEN ACCESS

Circles reshaping the RNA world: from waste to treasure

Jing Liu, Tian Liu, Xiaman Wang and Aili He

Molecular Cancer 2017 **16**:58 | **DOI:** 10.1186/s12943-017-0630-y | © The Author(s). 2017

Received: 25 October 2016 | **Accepted:** 2 March 2017 | **Published:** 9 March 2017

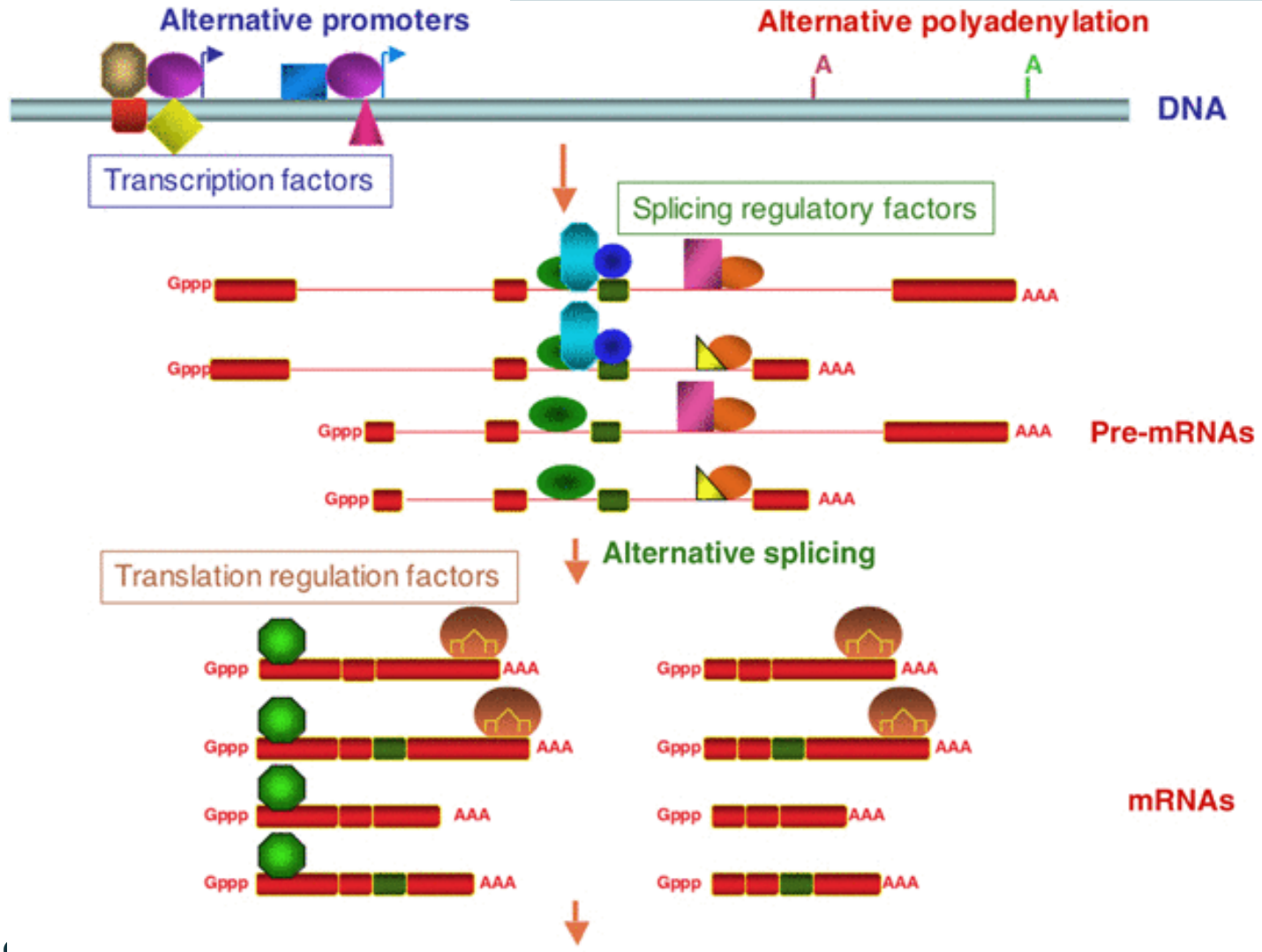


(Liu et al., 2017)

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

Transcript diversity and combinatorial regulation of gene expression



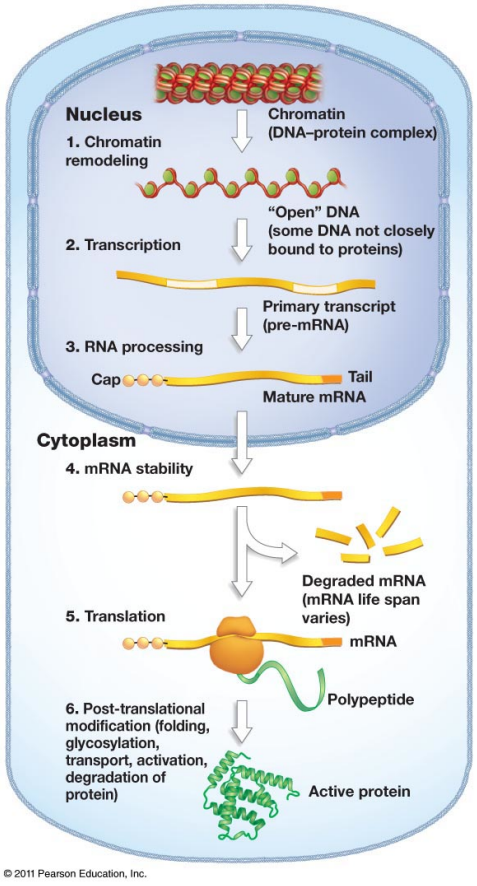
(Soares & Valcárcel, 2006)

PART 1

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

The regulation processes are complex, with several coding levels
Some methods used by eukaryotes for gene regulation:

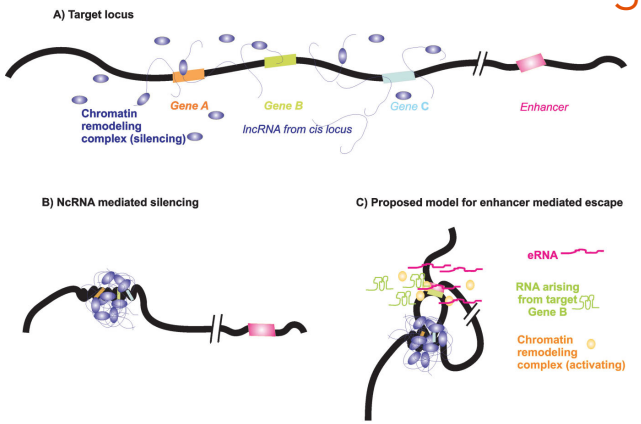


Acetylation
Ubiquitination
Methylation
Phosphorylation
Degradation
(...)

+

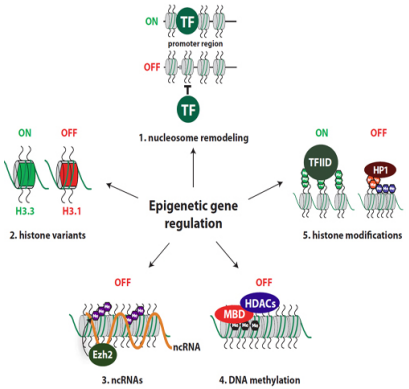
Response to stimuli
Transcription initiation

lncRNA mediated silencing



(Saxena & Carninci, 2011)

Epigenetic regulation



(Hahn, Dambacher & Schotta, 2010)

http://www.oryzon.com

GENOMES & TRANSCRIPTOMES

TAKE-HOME MESSAGE (OMICS : EXAMPLES)

With the increasing amount of data generated, people have been talking more and more about « omics »

Genomics	Studies the **genomes** of organisms.
Transcriptomics	Studies the **transcriptome** : it is the set of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA produced in one or a population of cells.
Proteomics	Studies the **proteome** : it is the entire set of proteins, including the modifications made to a particular set of proteins, produced by an organism or system. Large-scale study of proteins, particularly their structures and functions. Mass spectrometry techniques are used
Metabolomics	Scientific study of chemical processes involving metabolites. It is a "systematic study of the unique chemical fingerprints that specific cellular processes leave behind", the study of their small-molecule metabolite profiles
Pharmacogenomics	investigates the effect of the sum of variations within the genome on drugs
Proteogenomics	An emerging field of biological research at the intersection of proteomics and genomics. Proteomics data used for gene annotations.
Immunoproteomics	study of large sets of proteins (proteomics) involved in the immune response
Functional genomics	Gene and protein functions and interactions (uses **microarray** kind of techniques)

OVERVIEW

PARTS 1 - 3

Part 1 Genomes and Transcriptomes : From central dogmas to ongoing discoveries

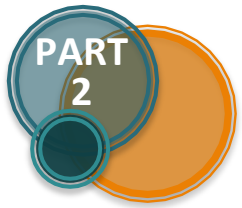
(Central dogma of molecular Biology, Human Genome Project, GENCODE, ENCODE…)

Part 2 Advances in Sequencing Technologies

(Examples of technologies developed for Closed/Open systems for gene expression analysis, Next-Generation Sequencing platforms and technologies …)

Part 3 Overview of NGS (DNA / RNA Seq) Protocols and related file formats

(Overview of protocols for Genomic and Transcriptomic analysis of standard samples using Next-Generation Sequencing and overview of the different formats generated at each step…)



GENOMES & TRANSCRIPTOMES

SEQUENCING

- In order to cover all these complex events, researchers seek to study entire genomes and transcriptomes.

  How is it possible nowadays to do so?

  → Sequencing using NGS

What is sequencing about ?

Determining the precise order of the smallest units of a query. Examples:

DNA	RNA	Protein	ChIP
			
determine the precise order of nucleotides	determine the precise order of nucleotides	determine the precise order of amino acids	determine genome-wide the DNA binding sites for **DNA-associated proteins** (TFs, …)

SEQ TECHNOLOGIES

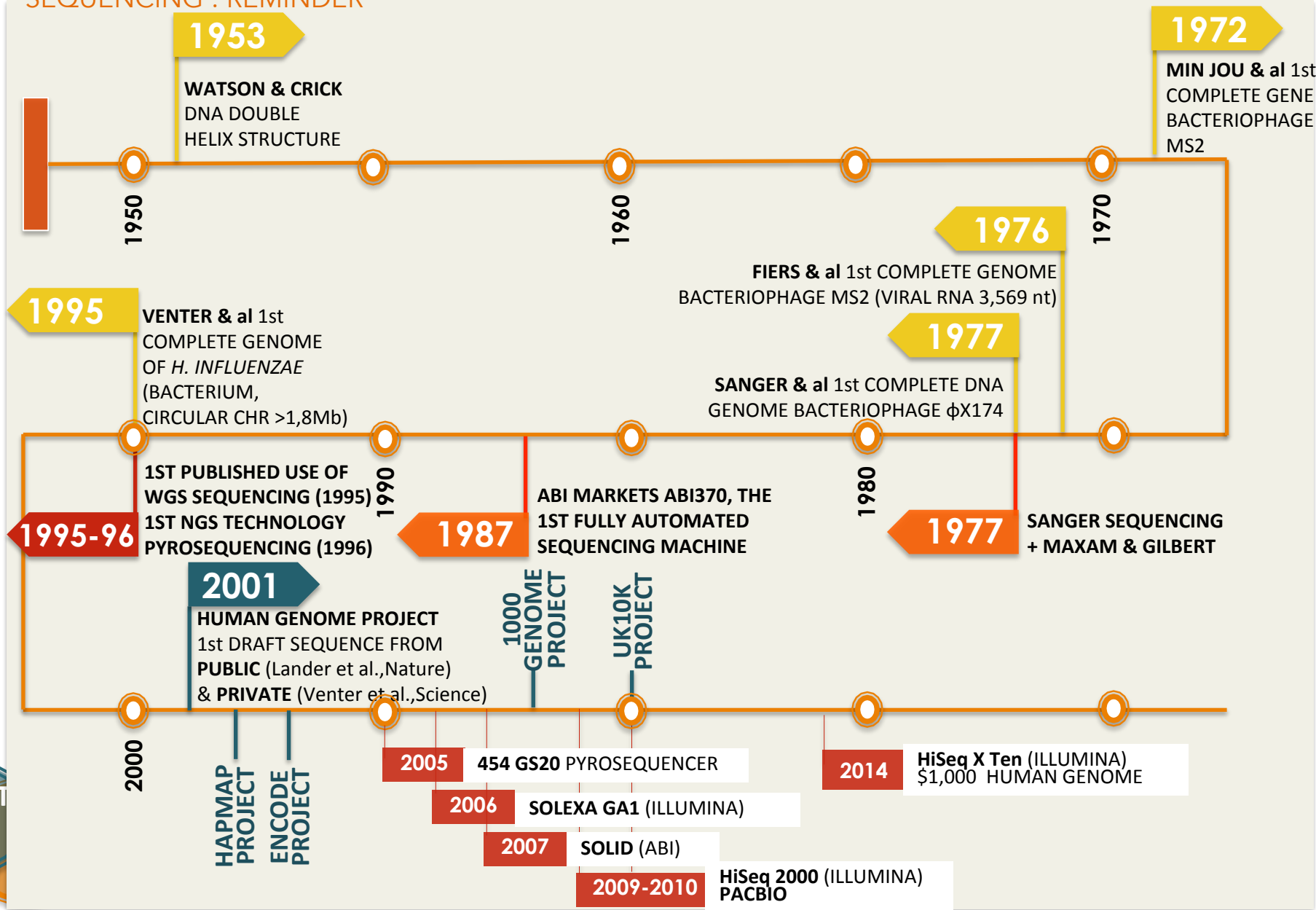
METHODS TO STUDY GENES AND GENE EXPRESSION

- Advances in Sequencing Technologies
- Next Generation Sequencing Platforms & Technologies

PART 2

GENOMES & TRANSCRIPTOMES

SEQUENCING : REMINDER



PART 2

SEQ TECHNOLOGIES

ADVANCES MADE IN SEQUENCING APPROACHES AND TECHNOLOGIES

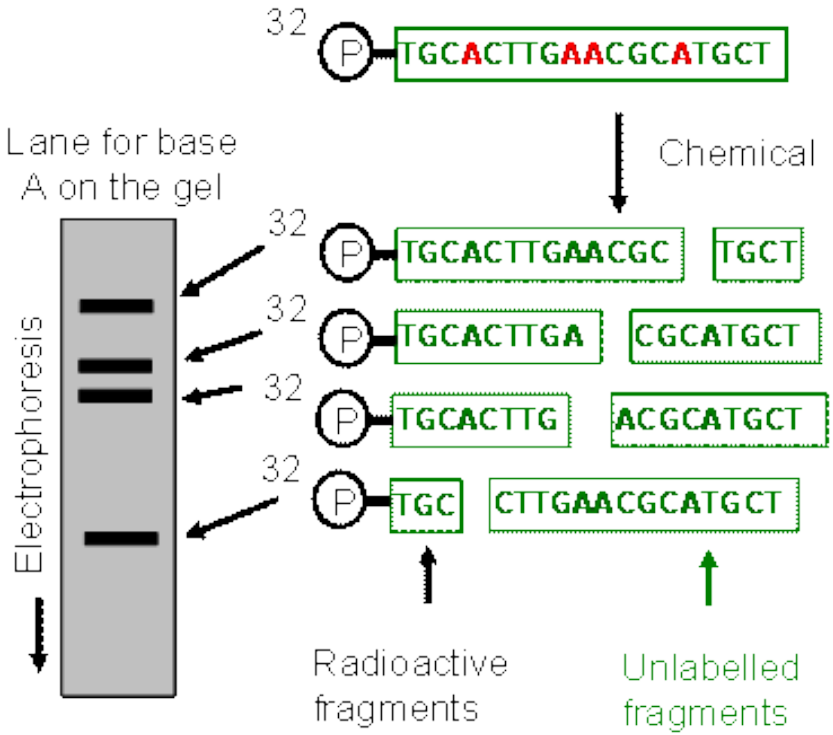
Maxam & Gilbert Sequencing (pre-NGS)

-The DNA fragment is labelled with ^{32}P at its 5' end.

- Then chemicals are used that break the DNA preferentially at each (or two) of the four nucleotide bases under conditions in which only one break per chain is made.

- After gel electrophoresis and autoradiography only the fragments possessing 5'-terminal ^{32}P-phosphate group show up on the gel.

- Time-consuming and requires handling of toxic chemicals.



wiki.biomine.skelleftea.se

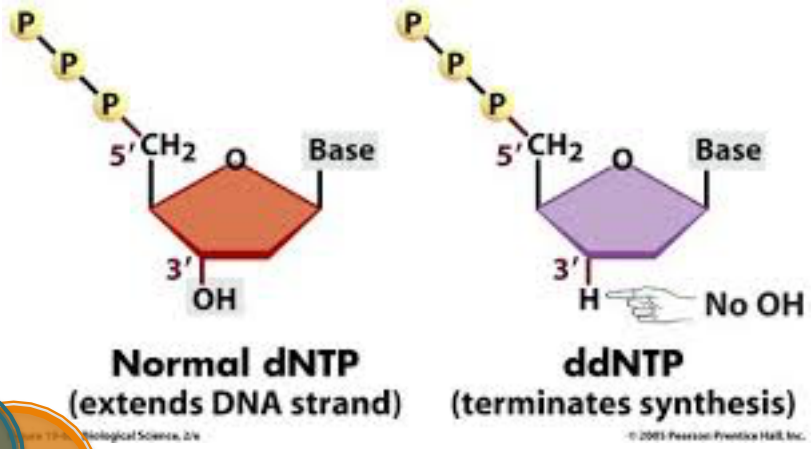
SEQ TECHNOLOGIES

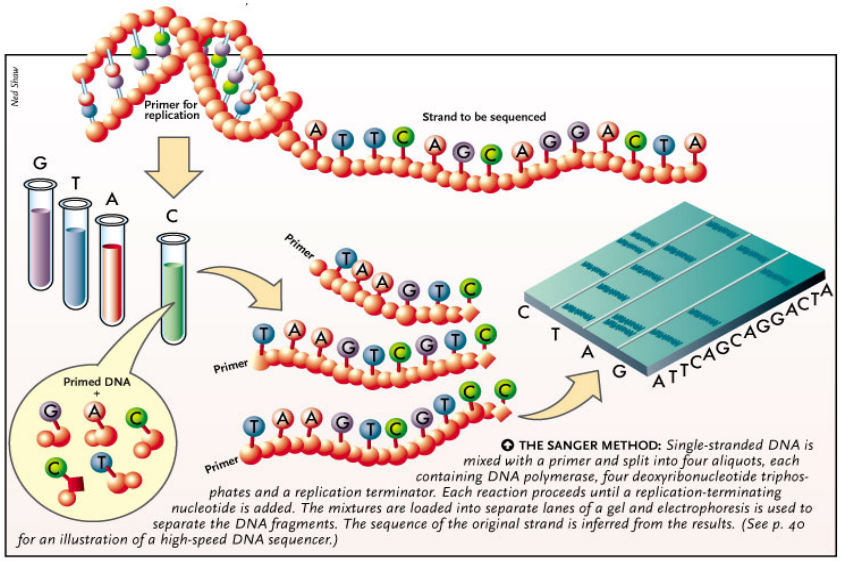
ADVANCES MADE IN SEQUENCING APPROACHES AND TECHNOLOGIES

Sanger Sequencing (pre-NGS)

- **Sanger sequencing** is the method developed by Frederick Sanger in 1977. This method involves copying single-stranded DNA with chemically altered bases called dideoxynucleotides (**ddNTPs**).

- ddNTPs when incorporated at the 3' end of the growing chain, terminate the chain selectively at A, C, G, or T. The terminated chains are then resolved by capillary electrophoresis.
- https://www.youtube.com/watch?v=FvHRio1yyhQ





SEQ TECHNOLOGIES

EXAMPLE OF TECHNIQUES DEVELOPED TO STUDY GENE EXPRESSION

- The principle of Sanger sequencing revolutionized the approaches developed to study genes and gene expression
- High accuracy
- Remained the most widely used sequencing method for about 40 years, until recently supplanted by "Next Generation Sequencing" (NGS) methods
- NGS are nowadays widely used, especially for large-scale automated genome sequencing
- However, Sanger sequencing remains the method of choice for small scale sequencing (few genes approach) and validation of microarray or NGS results, or for obtaining long reads (> 500 nucleotides)

https://en.wikipedia.org/wiki/Sanger_sequencing

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

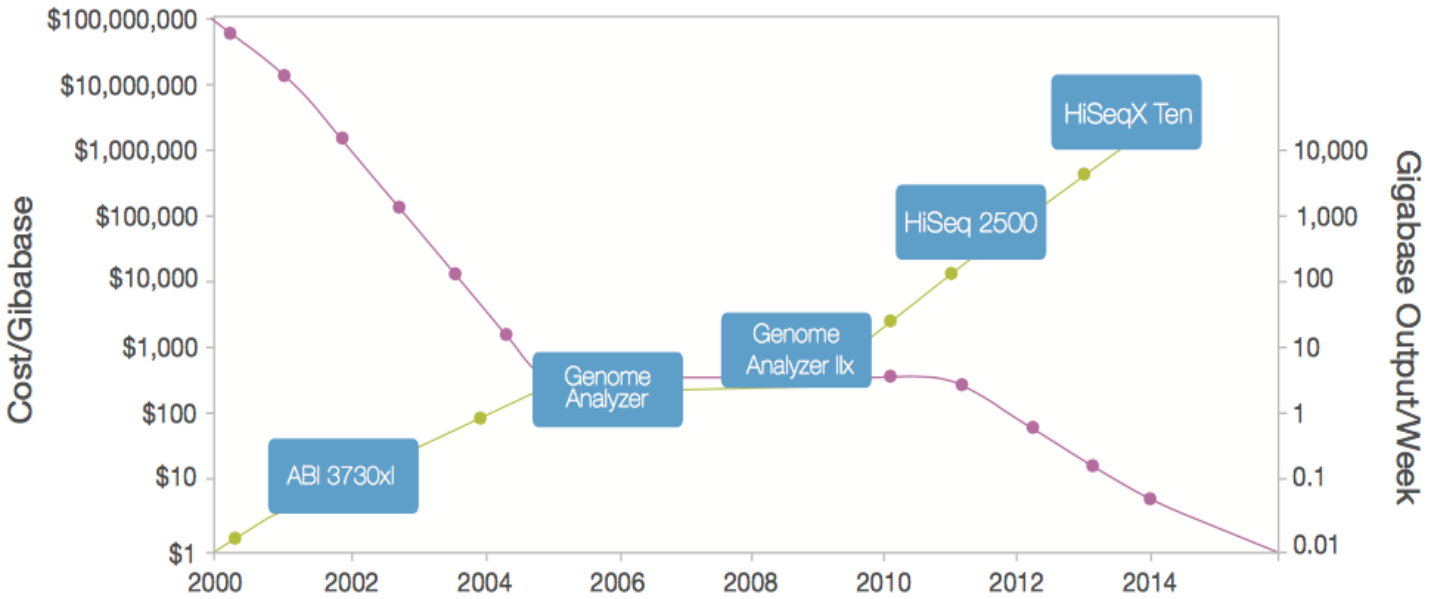
- 1977: Sanger Sequencing
- 1987: First automated sequencing instrument based on capillary electrophoresis (AB370, 1987 ; AB3730xl, 1998) → primary instrument for HGP (public and private).

→ "First generation instruments" : 84Kb / run

- 2005: Genome Analyzer

→ "Next generation instruments" start : 1Gb / run ---------- 2014: 1.8Tb

Sequencing Cost and Data Output Since 2000

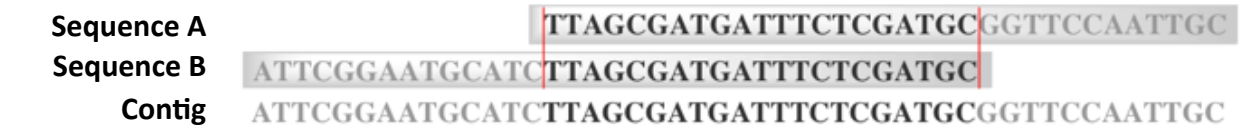


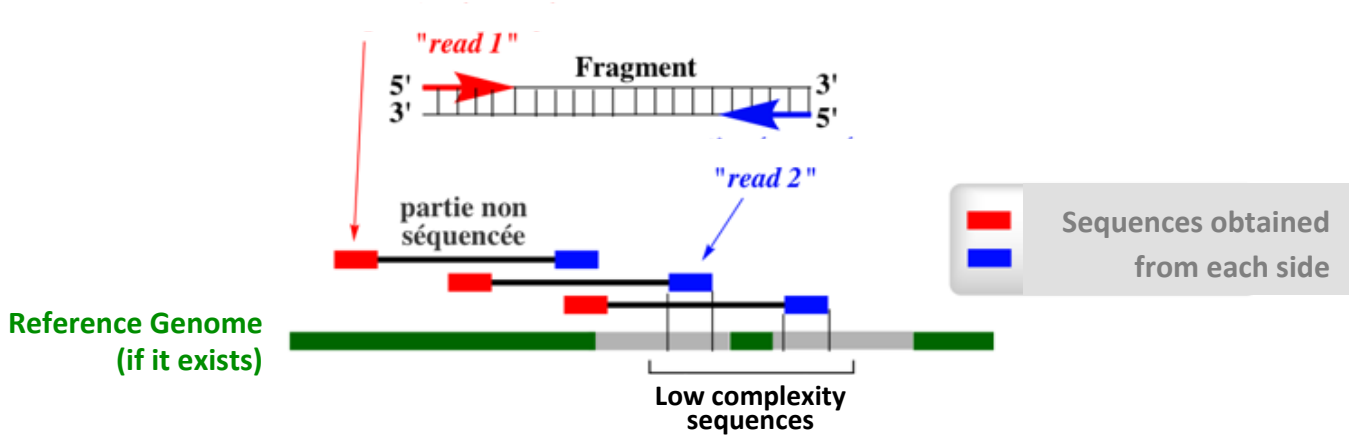
http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

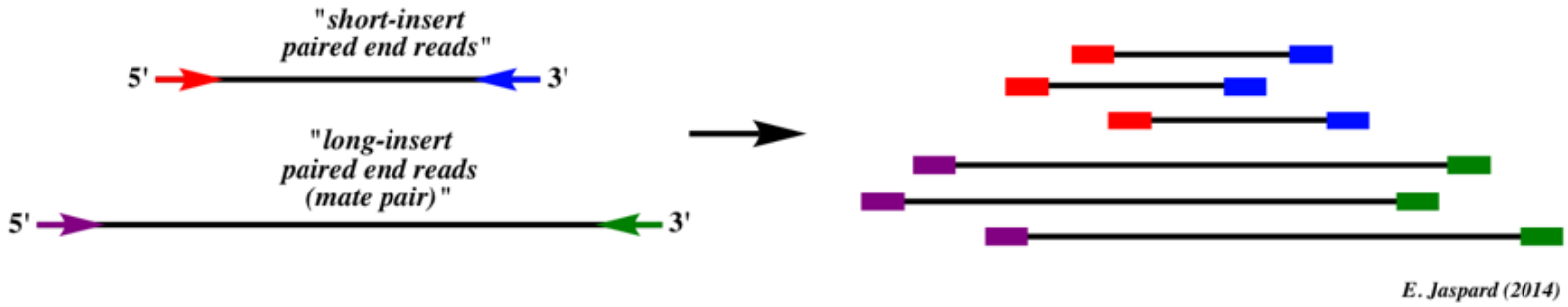
SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

- NGS







Next Generation of Sequencing Technologies

Technology	Company	Support	Chemistry
Massively Parallel Sequencing			
Solexa	Illumina	**Bridge PCR** on flowcell	**Seq-By-Synthesis**
454	Roche Applied Science	**emPCR** on beads	**Pyrosequencing**
SOLiD	AB / Life Technologies	emPCR on beads	Seq-By-Ligation
Ion Torrent	Life Technologies	emPCR on beads	Proton detection
Single Molecule Sequencing			
PacBio SMRT	Pacific Biosciences	Pol performance	Real-time-Seq
Nanopore	Oxford Nanopore Tech/McNally	Translocation	NA

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

• The principle of " Sequencing By Synthesis " (more commonly SBS) = tracking the addition of labeled nucleotides as the DNA chain is copied.

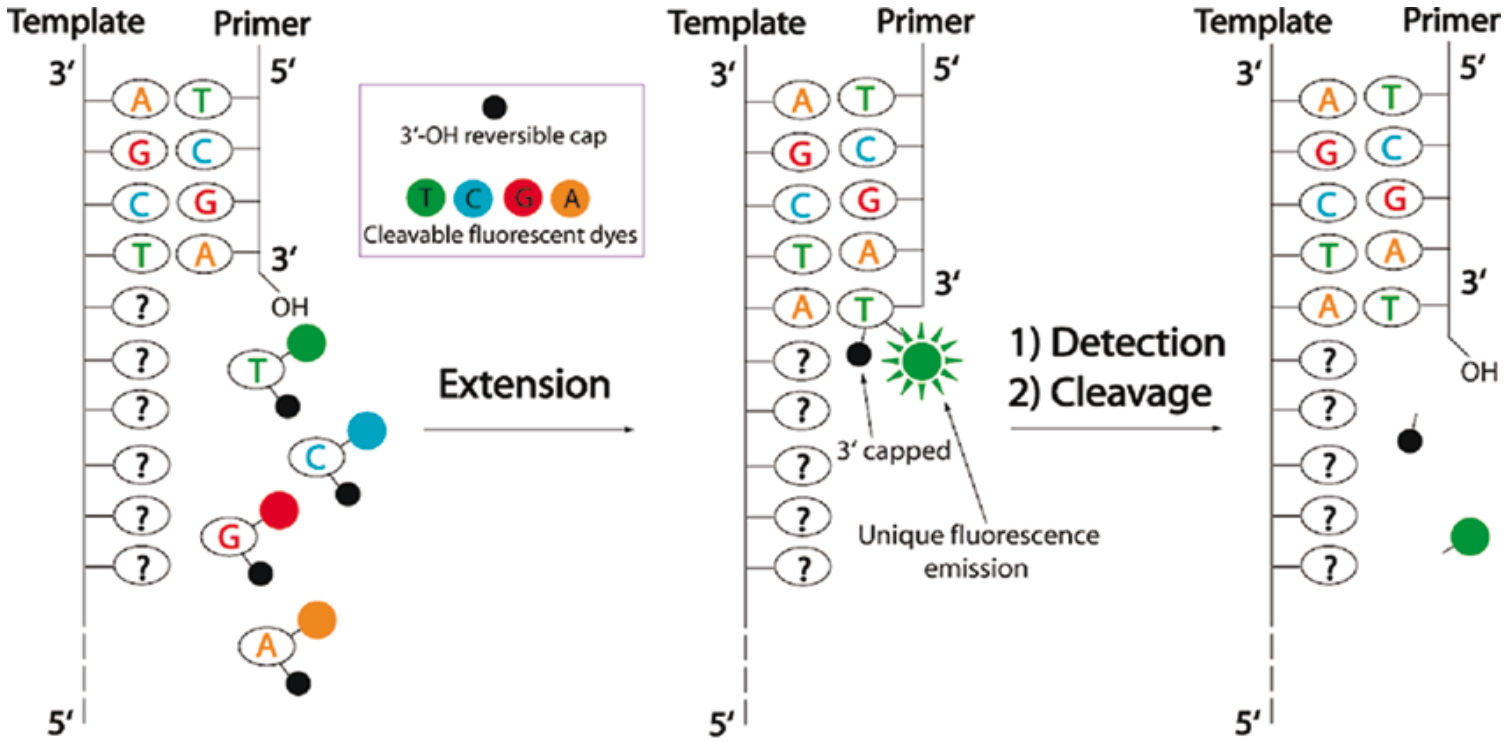
• This method involves to:
- Take a single-stranded DNA to sequence
- Synthesize its complementary strand enzymatically
- Detect light emitted sequentially

• The DNA sequence is determined by the light emitted during the incorporation of nucleotides, knowing that only one of the possible four bases A/T/C/G is incorporated.

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

- The DNA template is immobilized.
- Solutions of A, nucleotides C , G and T sequentially added and removed.
- Light is generated when a nucleotide complements the unpaired base.
- Chemiluminescent signal detected to determine the sequence.



SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

- Pyrosequencing : Basically, the process allows for sequencing a single stranded DNA by synthesizing the complementary strand over its entire length, a pair of bases at a time, and detection at each step of the base actually added.

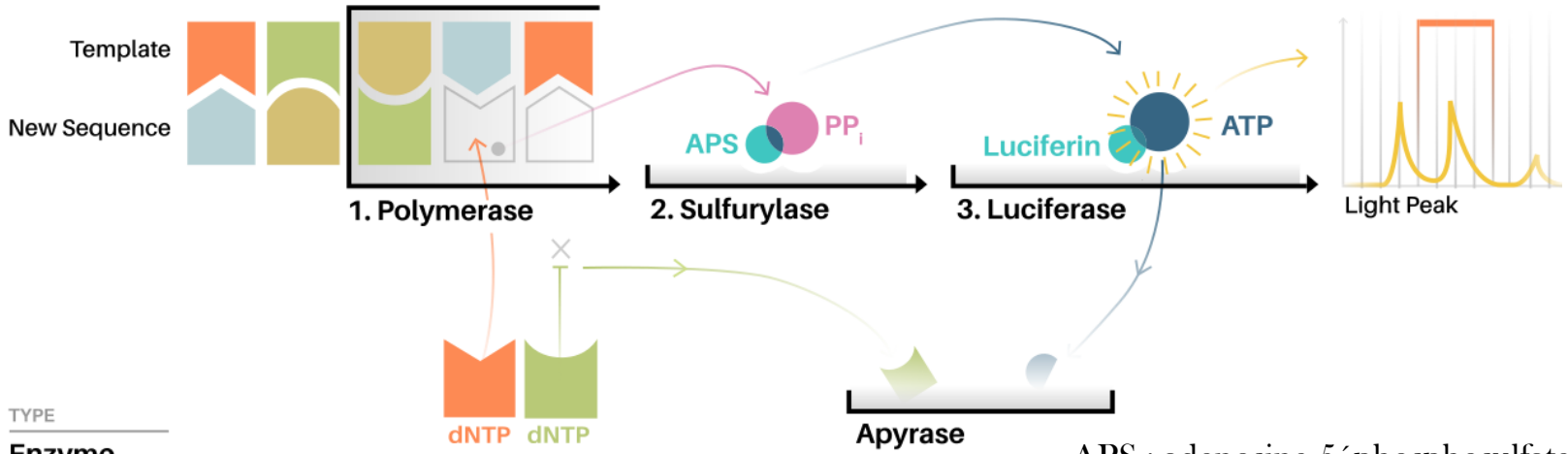
- Pyrosequencing is based on the detection of the activity of a DNA polymerase (the enzyme for DNA synthesis).

- The DNA sequence is determined by the light (luciferase activity) emitted during the incorporation of nucleotides, knowing that only one of the possible four bases A/T/C/G is incorporated.

SEQ TECHNOLOGIES

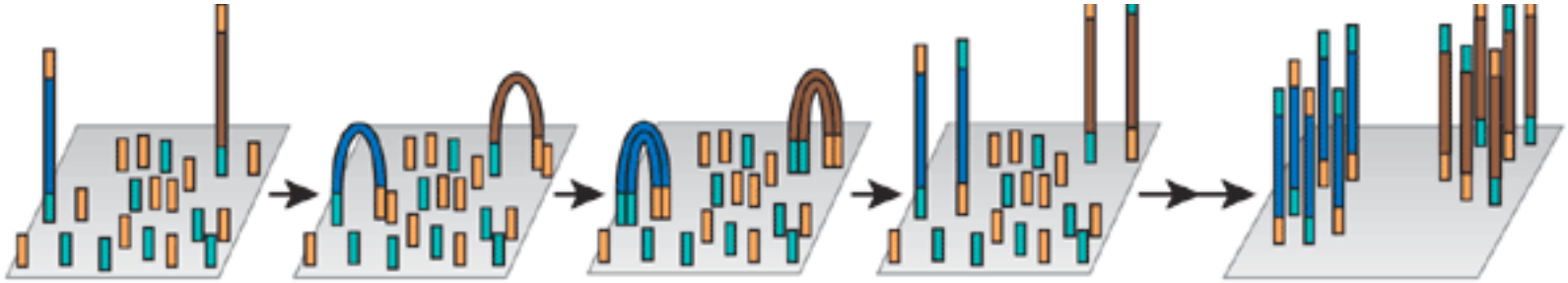
NEXT-GENERATION SEQUENCING

- Incorporation of the complementary dNTPs (added sequentially) by **DNApol** releases pyrophosphate (PPi).
- ATP sulfurylase converts PPi to ATP in the presence of APS.
- ATP = substrate for the luciferase-mediated conversion of luciferin to oxyluciferin This conversion generates visible light in amounts proportional to the amount of ATP detected by a camera and analyzed in a pyrogram.
- Unincorporated nucleotides and ATP are degraded by the apyrase, and the reaction can restart with another nucleotide



APS : adenosine 5´phosphosulfate
PPi : Pyrophosphate
DNApol: DNA Polymerase

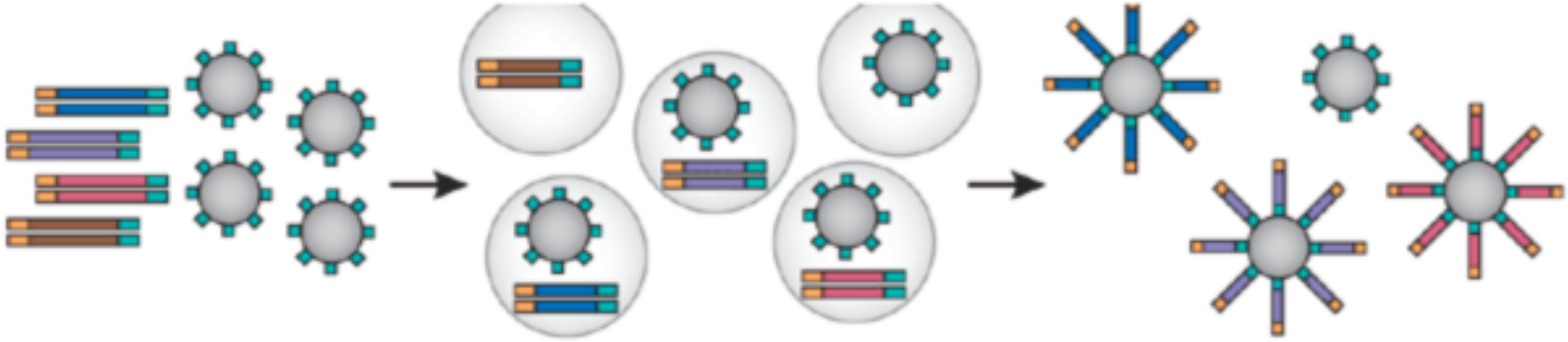
Bridge PCR



- The adaptor-flanked shotgun library is PCR amplified on a flow cell
- both primers coat the surface of a solid substrate
- Amplification products from any given member of the library remain locally fixed near the point of origin = cluster
- The PCR produces clonal clusters that contains copies of a single DNA.

(Shendure & Ji, 2008)

emPCR



- The adaptor-flanked shotgun library is PCR amplified in the context of a water-in-oil emulsion.
- PCR primer is 5'-attached on micron-scale beads.
- 1 bead-containing compartments = 0 or 1 template DNA.
- PCR amplicons are captured to the surface of the bead.
- 1 clonally amplified bead = PCR products corresponding to amplification of a single molecule from the library.

Next Generation of Sequencing Technologies

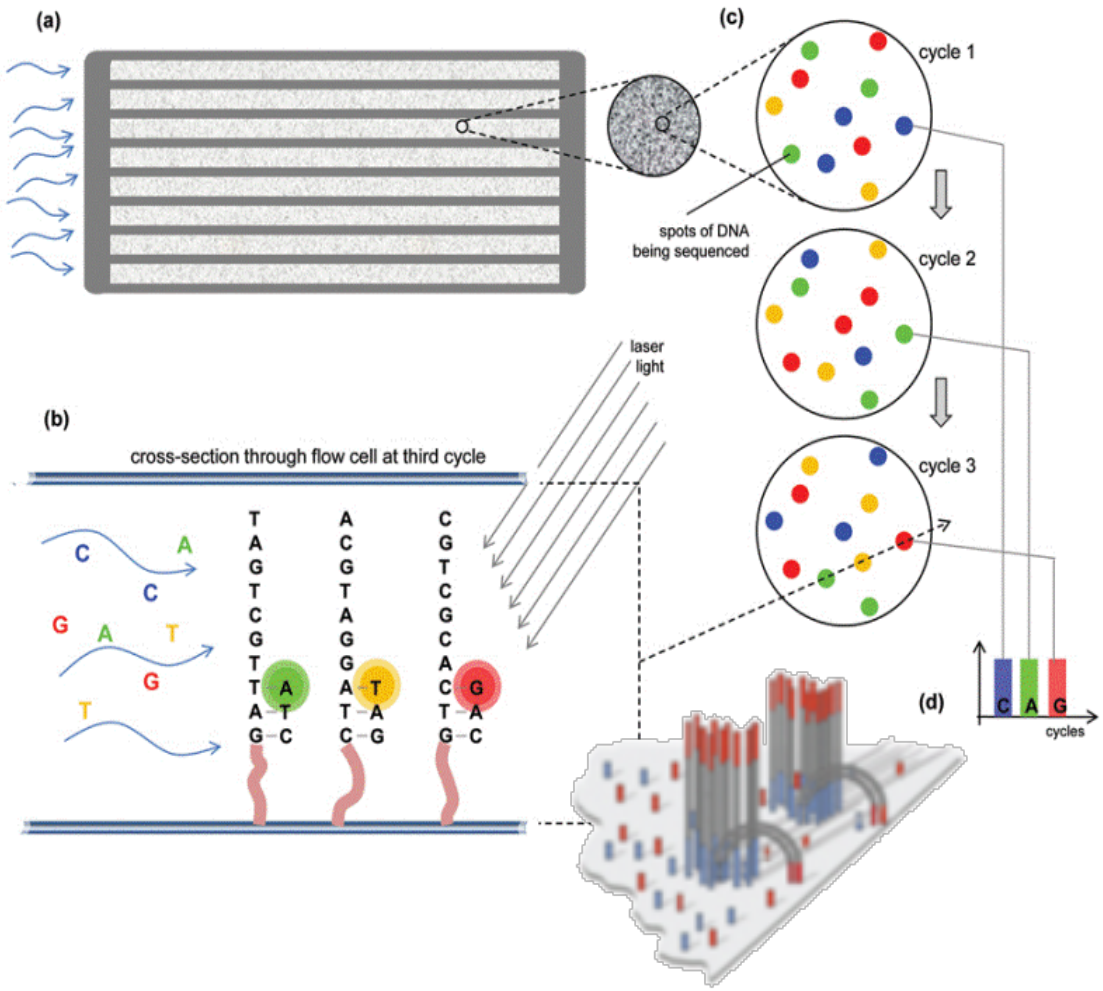
Technology	Company	Support	Chemistry
Massively Parallel Sequencing			
Solexa	**Illumina**	**Bridge PCR** on flowcell	**Seq-By-Synthesis**
454	Roche Applied Science	**emPCR** on beads	**Pyrosequencing**
SOLiD	AB / Life Technologies	emPCR on beads	Seq-By-Ligation
Ion Torrent	Life Technologies	emPCR on beads	Proton detection
Single Molecule Sequencing			
PacBio SMRT	Pacific Biosciences	Pol performance	Real-time-Seq
Nanopore	Oxford Nanopore Tech/McNally	Translocation	NA

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

Solexa (Illumina)

- The input sample must be cleaved into short sections.
- Fragments are ligated to adaptors and annealed to the slide using the adaptors.
- Fragments are separated into single strands to be sequenced.
- Nucleotides are modified so that each emits a different coloured light when excited by a laser.
- they have a terminator, so that only 1 base is added at a time.
- PCR, process repeated in cycles, images analyzed (SBS).



(modified from Gilchrist, 2010)

Solexa (Illumina)

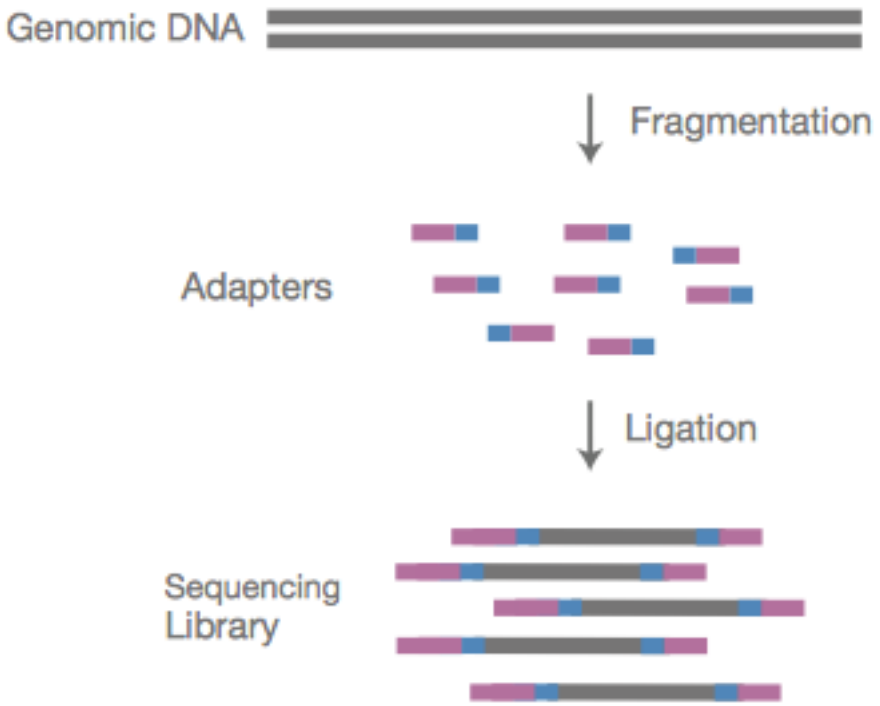
illumina

1. Library Preparation

- Random fragmentation of the DNA or cDNA sample
- 5’ and 3’ adapter ligation.

NB: Alternatively, “tagmentation” combines the fragmentation and ligation reactions into a single step that greatly increases the efficiency of the library preparation process.

- Adapter-ligated fragments are then PCR amplified and gel purified.

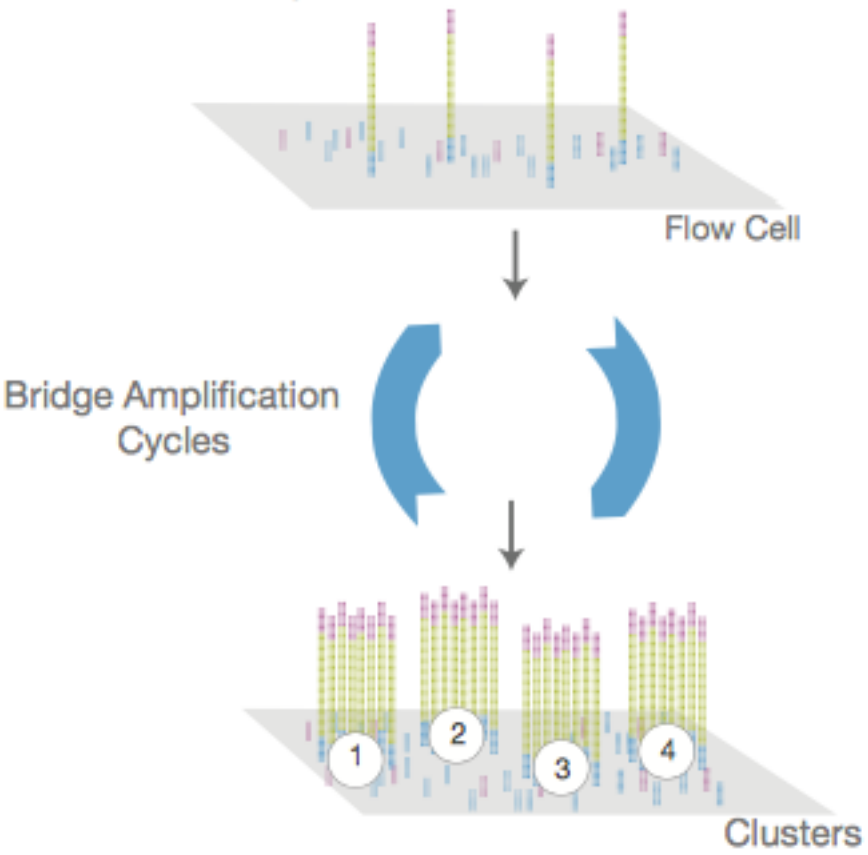


NGS library is prepared by fragmenting a gDNA sample and ligating specialized adapters to both fragment ends.

http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

Solexa (Illumina)

illumina



Library is loaded into a flow cell and the fragments hybridize to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.

2. Cluster Generation

- library is loaded into a flow cell
- fragments are captured on a surface-bound oligos complementary to the library adapters.
- Each fragment is then amplified into distinct, clonal clusters through bridge amplification.
- When cluster generation is complete, the templates are ready for sequencing.

Solexa (Illumina)

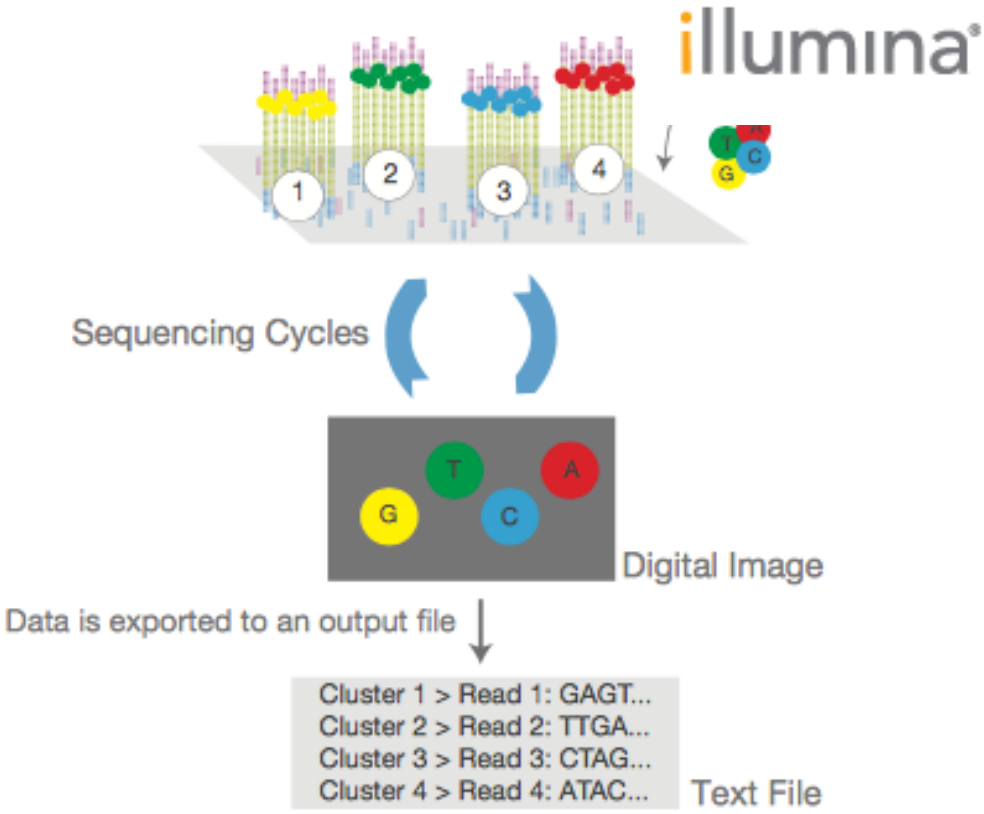
3. Sequencing

- Illumina SBS technology :

= reversible terminator–based method detecting single bases as they are incorporated into DNA template strands.

- All 4 reversible terminator-bound dNTPs are present during each sequencing cycle

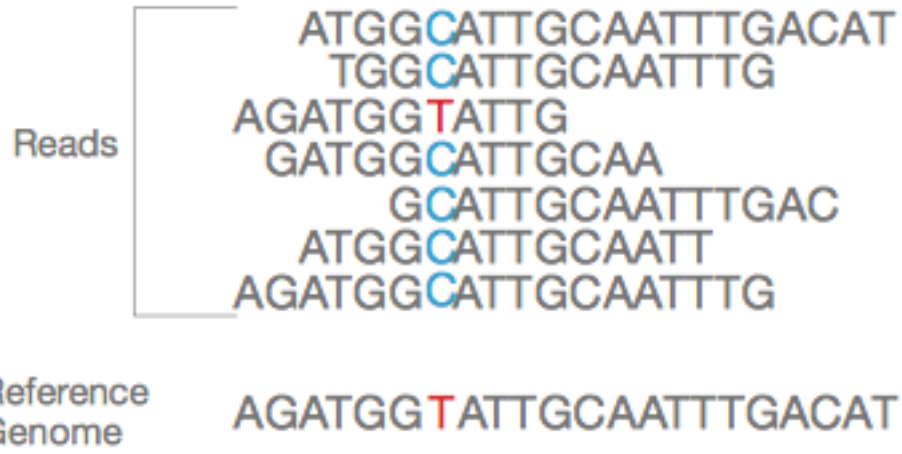
= natural competition minimizes incorporation bias and greatly reduces raw error rates.



Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated "n" times to create a read length of "n" bases.

http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

illumina



Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.

4. Alignment & Data Analysis

- The newly identified sequence reads are aligned to a reference genome.
- Following alignment, many variations of analysis are possible such as single nucleotide polymorphism (SNP) or insertion-deletion (indel) identification, read counting for RNA methods, phylogenetic or metagenomic analysis,…
- SBS sequencing: www.illumina.com/SBSvideo.

http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

Solexa (Illumina)

illumina

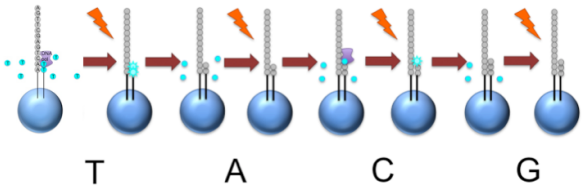
	 MiniSeq System	 MiSeq Series	 NextSeq Series	 HiSeq Series	 HiSeq X Series*
Key Methods	Amplicon, targeted RNA, small RNA, and targeted gene panel sequencing.	Small genome, amplicon, and targeted gene panel sequencing.	Everyday exome, transcriptome, and targeted resequencing.	Production-scale genome, exome, transcriptome sequencing, and more.	Population- and production-scale whole-genome sequencing.
Maximum Output	7.5 Gb	15 Gb	120 Gb	1500 Gb	1800 Gb
Maximum Reads per Run	25 million	25 million†	400 million	5 billion	6 billion
Maximum Read Length	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp
Run Time	4–24 hours	4–55 hours	12–30 hours	<1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	<3 days
Benchtop Sequencer	Yes	Yes	Yes	No	No
System Versions	• MiniSeq System for low-throughput targeted DNA and RNA sequencing	• MiSeq System for targeted and small genome sequencing • MiSeq FGx System for forensic genomics • MiSeqDx System for molecular diagnostics	• NextSeq 500 System for everyday genomics • NextSeq 550 System for both sequencing and cytogenomic arrays	• HiSeq 3000/HiSeq 4000 Systems for production-scale genomics • HiSeq 2500 Systems for large-scale genomics	• HiSeq X Five System for production-scale whole-genome sequencing • HiSeq X Ten System for population-scale whole-genome sequencing

PART 2

454 (Roche)

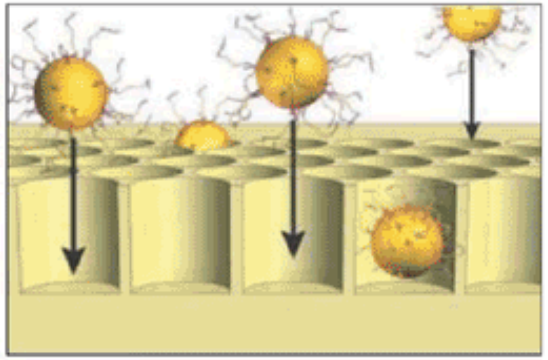
- As in Illumina, the DNA is fragmented.

- Adaptors added, end annealed to beads.
1 DNA fragment = 1 bead (theory).



- Fragments amplified by PCR using adaptor-specific primers.

- The sequence can then be determined computationally.

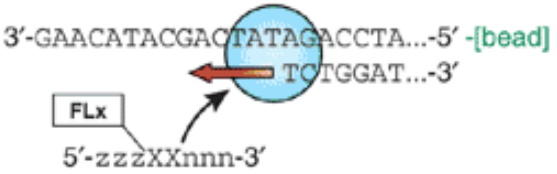


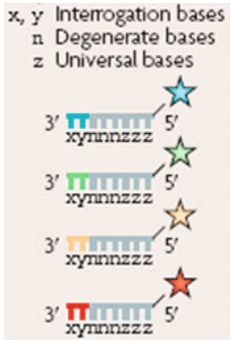
- Longer reads than Illumina, different lengths.

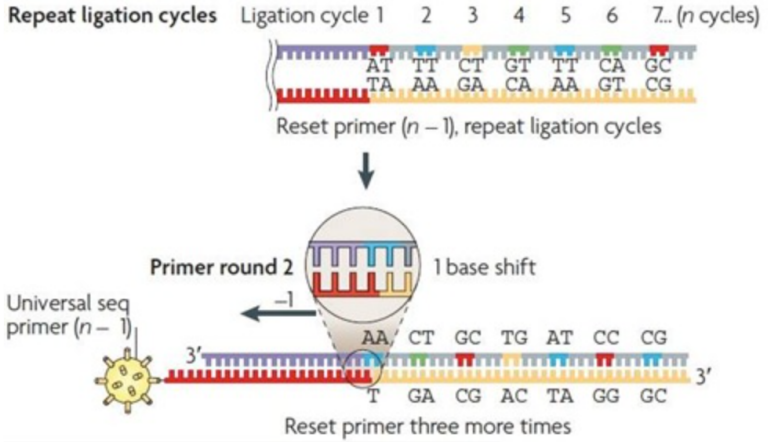
SOLID = Sequencing by Oligonucleotide Ligation and Detection

- Sequencing uses a ligase, rather than a polymerase → *Sequencing-By-Ligation*
- Each sequencing cycle introduces a **partially degenerated population of fluorescently labeled octamers**. The population is structured such that **the label correlates with the identity of the central 2 bp in the octamer**.
- **After ligation and imaging** in four channels, the **labeled portion of the octamer (that is, 'zzz') is cleaved** leaving a free end for another cycle of ligation.

→ Reading errors reduced compared to pyrosequencing.



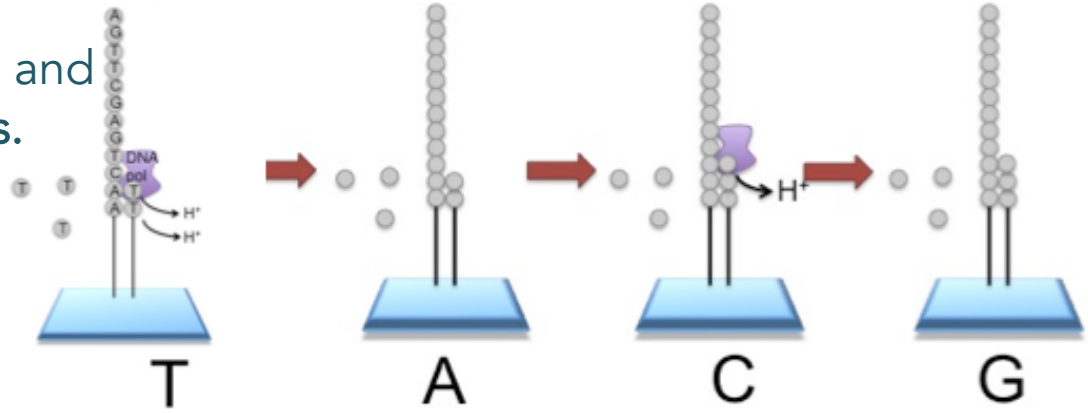




Adapted from (Shendure & Ji, 2008)
Adapted from (http://www.springer.com/978-1-4614-7725-9)
Adapted from (http://www.oezratty.net/)

ThermoFisher SCIENTIFIC

- As in other kinds of NGS, the input DNA is fragmented.
- Unlike other methods, Ion Torrent do not use optical signalling.
- **Adaptors** are added and **one molecule is placed onto a bead**.
- Amplification on the bead by **emulsion PCR**. Each bead is placed into 1 well of a slide.
- **The pH is detected is each of the wells, as each H+ ion released will decrease the pH**. The changes in pH allow us to determine if that base, and how many thereof, was added to the sequence read.
- The dNTPs are washed away, and the process is **repeated in cycles.**



Adapted from (http://www.ebi.ac.uk) and (http://www.springer.com/978-1-4614-7725-9)

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

Ion torrent

ThermoFisher SCIENTIFIC

Ion PGM System

Targeted DNA Sequencing

Targeted RNA Sequencing

Microbial Sequencing

Ion PGM System performance specifications

		Ion 314™ Chip v2 or Ion 314 Chip v2 BC	Ion 316™ Chip v2 or Ion 316 Chip v2 BC	Ion 318™ Chip v2 or Ion 318 Chip v2 BC
Output*	200 base 400 base†	30–50 Mb 60–100 Mb	300–600 Mb 600 Mb–1Gb	600 Mb–1Gb 1.2–2 Gb
Reads		400–550 thousand	2–3 million	4–5.5 million
Run time	200 base 400 base	2.3 hr 3.7 hr	3.0 hr 4.9 hr	4.4 hr 7.3 hr
Research areas		Cancer research, inherited disease research, microbial genomics, stem cell research, agriculture, epigenomics, metagenomics, forensic science, and ancient DNA genomics		
Key applications		Targeted DNA sequencing, copy number analysis, targeted RNA sequencing, small RNA sequencing, *de novo* microbial sequencing, bacterial typing research, viral typing research, ChIP sequencing, methylation analysis, SNP verification, and genotyping by sequencing		

ThermoFisher SCIENTIFIC

The Ion Proton™ System

Rapid, high-throughput benchtop sequencing

The Ion Proton™ System minimizes the high cost and complexity of high-throughput sequencing, bringing it to your research lab, on your budget, and on your schedule. This system enables high-quality exome and transcriptome sequencing in just a few days.

Ion Proton™ System performance specifications with Ion PI™ Chip			
Output	Up to 10 Gb		
Reads	60–80 million reads		
Read length	Up to 200-base reads		
Run time	2–4 hours		
Research areas	Cancer Genetic disorders	Agriculture Stem cells	Epigenetics
Key applications	Human-scale genome sequencing Exome sequencing	Transcriptome sequencing Copy number analysis	ChIP sequencing Targeted RNA sequencing

ThermoFisher SCIENTIFIC

Ion PGM™ Chip	Run time		Output	
	200 bp read	400 bp read	200 bp read	400 bp read
Ion 314™ Chip v2	2.3 hr	3.7 hr	30–50 Mb	60–100 Mb
Ion 316™ Chip v2	3.0 hr	4.9 hr	300–500 Mb	600 Mb–1 Gb
Ion 318™ Chip v2	4.4 hr	7.3 hr	600 Mb–1 Gb	1.2–2 Gb
Ion Proton™ Chip	Run time		Output	
	200 bp read		200 bp read	
Ion PI™ Chip	2–4 hr		Up to 10 Gb	

Next Generation of Sequencing Technologies

Technology	Company	Support	Chemistry
Massively Parallel Sequencing			
Solexa	**Illumina**	**Bridge PCR** on flowcell	**Seq-By-Synthesis**
454	Roche Applied Science	**emPCR** on beads	**Pyrosequencing**
SOLiD	AB / Life Technologies	emPCR on beads	Seq-By-Ligation
Ion Torrent	Life Technologies	emPCR on beads	Proton detection
Single Molecule Sequencing			
PacBio SMRT	Pacific Biosciences	Pol performance	Real-time-Seq
Nanopore	Oxford Nanopore Tech/McNally	Translocation	NA

Single Molecule Real Time (SMRT) DNA Seq technology

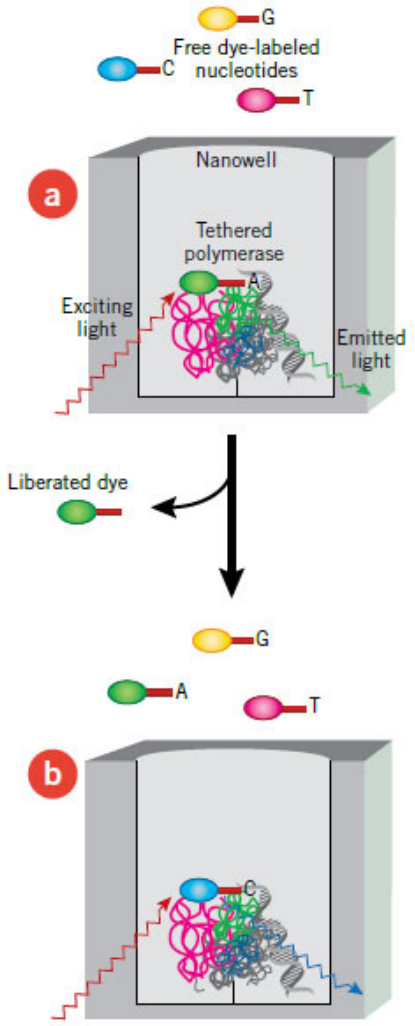
SMRT Sequencing enables real-time observation of DNA synthesis. SMRT Sequencing is built upon 2 key innovations:

- zero-mode waveguides (ZMWs)
- phospholinked nucleotides.

a → DNA polymerase molecule is tethered to the bottom of a nanowell → ZMW design ensures only one nucleotide-linked dye can be directly excited at a time.

b → Each incorporated phospholinked nucleotide will reside on the enzyme's active site for a few milliseconds, which is enough time for a fluorescent signal to be recorded.

Pacific Biosciences

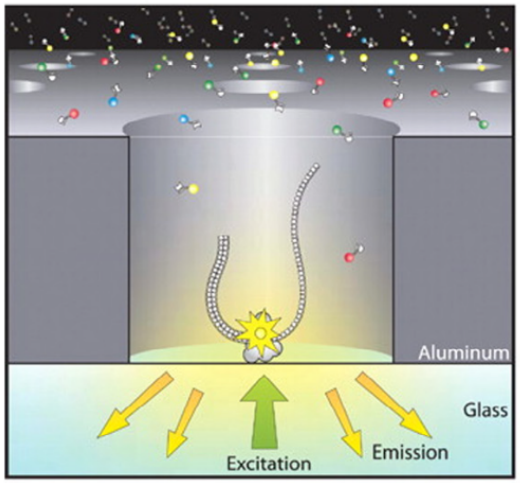


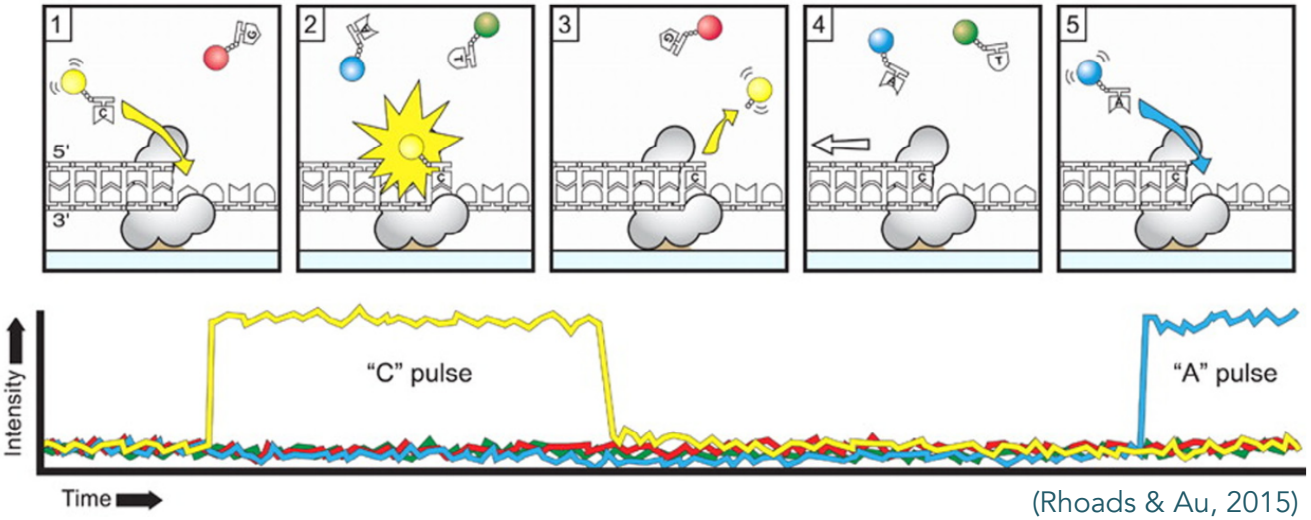
(Osherovich et al., 2010)
http://www.pacb.com/smrt-science/smrt-sequencing/

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

PacBio





(Rhoads & Au, 2015)

PCR	# PCR	% closed by Sanger	% closed by PacBio
<2.5 kb	246	64	73
>2.5 kb	113	0	88
hard stop	3	0	100

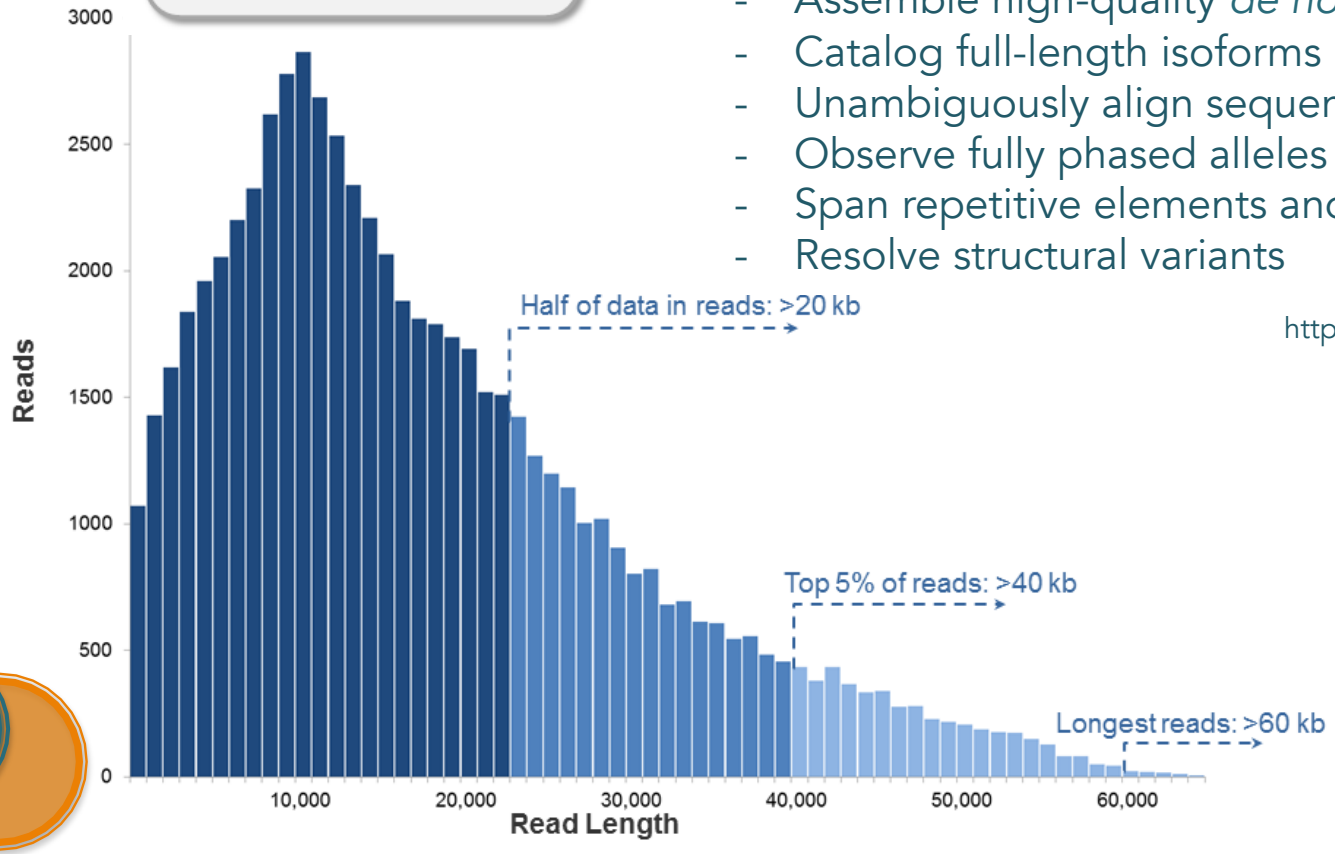
362 PCR products (each covering a different gap) were sequenced with both Sanger and PacBio technologies. While the majority of gaps less than 2.5 kb were closed with both Sanger (64%) and PacBio (73%) technologies, none of the gaps larger than 2.5 kb were closed with a single round of Sanger technology. Three hard stop gaps were all closed by PacBio.

(Zhang et al., 2012)

Single Molecule Real Time (SMRT) DNA Seq technology

Read Lengths > 20 kb

Data per SMRT Cell:
500 Mb – 1Gb



Read length:

- average > 10,000 bp,
- some reads > 60,000 bp

These long reads give you the ability to:

- Assemble high-quality *de novo* genomes
- Catalog full-length isoforms
- Unambiguously align sequences
- Observe fully phased alleles
- Span repetitive elements and complex regions
- Resolve structural variants

http://www.pacb.com/smrt-science/smrt-sequencing/read-lengths/

PART 2

Single Molecule Real Time (SMRT) DNA Seq technology

Sequel System: high-throughput, cost-effective access to SMRT Sequencing

The **Sequel System** idelivers ≥ 7X more reads with 1 million zero-mode waveguides (ZMWs) per SMRT Cell. The Sequel System is ideal for projects such as rapidly and cost-effectively generating **high-quality whole genome de novo assemblies**.

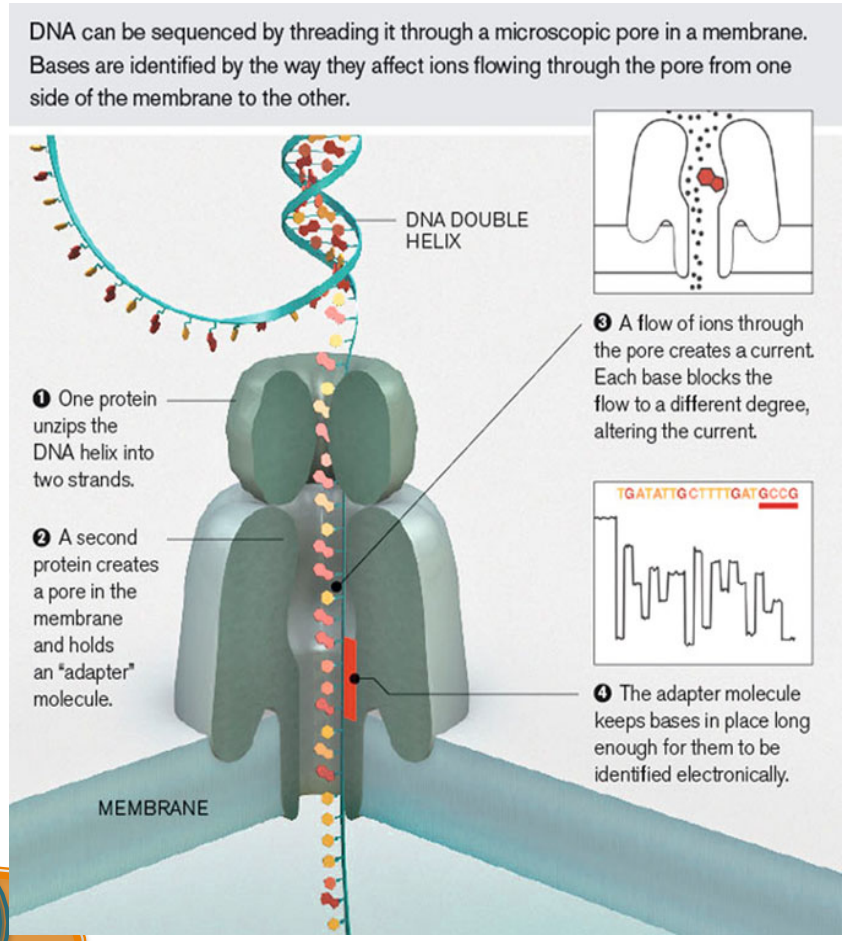
PacBio RS II: the original long-read sequencer

The **PacBio RS II** is ideal for **whole genome sequencing of small genomes, targeted sequencing, complex population analysis, RNA sequencing of targeted transcripts, and microbial epigenetics.**

http://www.pacb.com/

Nanopore

Single Molecule Real Time (SMRT) DNA Seq technology



Schematic representation of nanopore sequencing system.

- Upper protein → ssDNA.
- 2nd protein
  - forms a nanopore in a membrane.
  - contains an adaptor molecule that reduces the speed of passing DNA through the pore.
- Each base obstructs the flow to a different degree.

→ uses an electronic rather than an optical signal to identify DNA bases

(http://www.springer.com/978-1-4614-7725-9)
(picture from MIT's Technology Review)

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

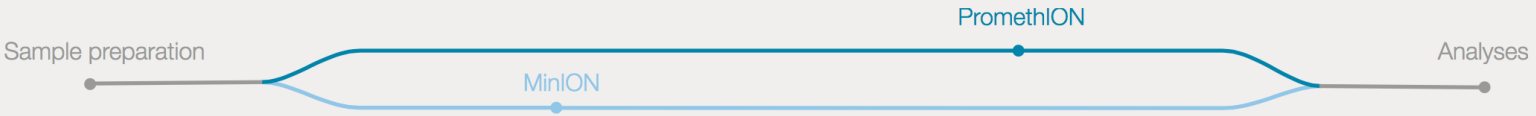
Nanopore

Single Molecule Real Time (SMRT) DNA

Weight
87g (103g with a flow cell)

Size
W 105, H 23, D 33mm

Simple sequencing workflows



Simple sample preparation (Coming soon: automated sample preparation from Voltrax)

Learn about Voltrax

Pocket-sized MinION for analysis anywhere

Learn about MinION

Desktop PromethION for high throughput, high sample number analysis

Learn about PromethION

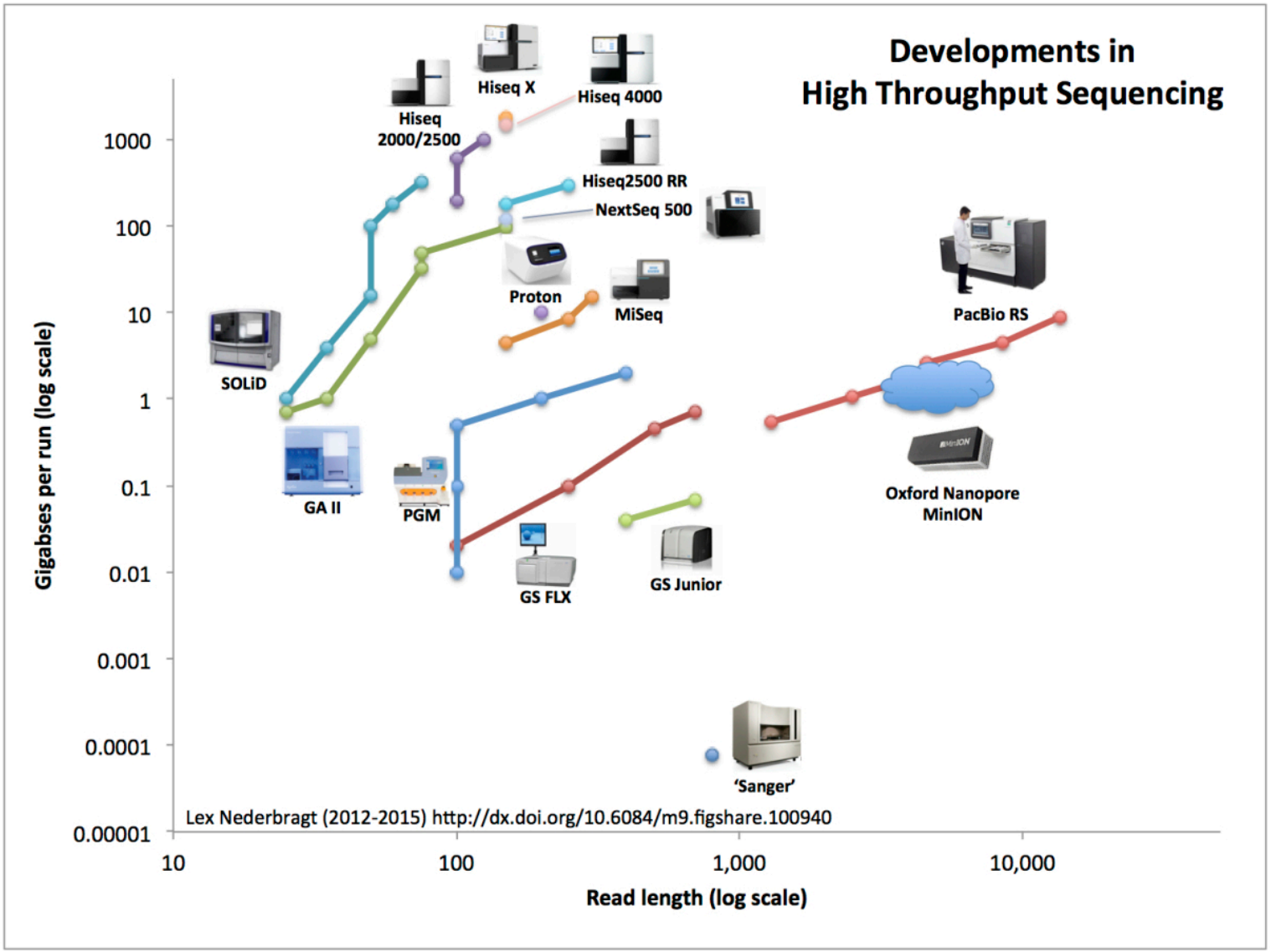
Real time, end-to-end analysis solutions from Metrichor

Learn about Metrichor

Read length : The longest read reported by a MinION user to date is more than 200Kb
Portability : The MinION can be used outside the traditional lab environment to perform field-based work

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING



SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

The four main advantages of NGS over classical Sanger sequencing

- **Speed**

NGS is quicker than Sanger sequencing in two ways.

- Chemical reaction may be combined with the signal detection, whereas in Sanger sequencing these are two separate processes.
- NGS is massively parallel.

- **Cost**

The human genome sequence cost 3 billion $.
Sequencing a human genome with Illumina allows to reach the $1,000 expected.

- **Sample size**

needs significantly less starting amount of DNA/RNA

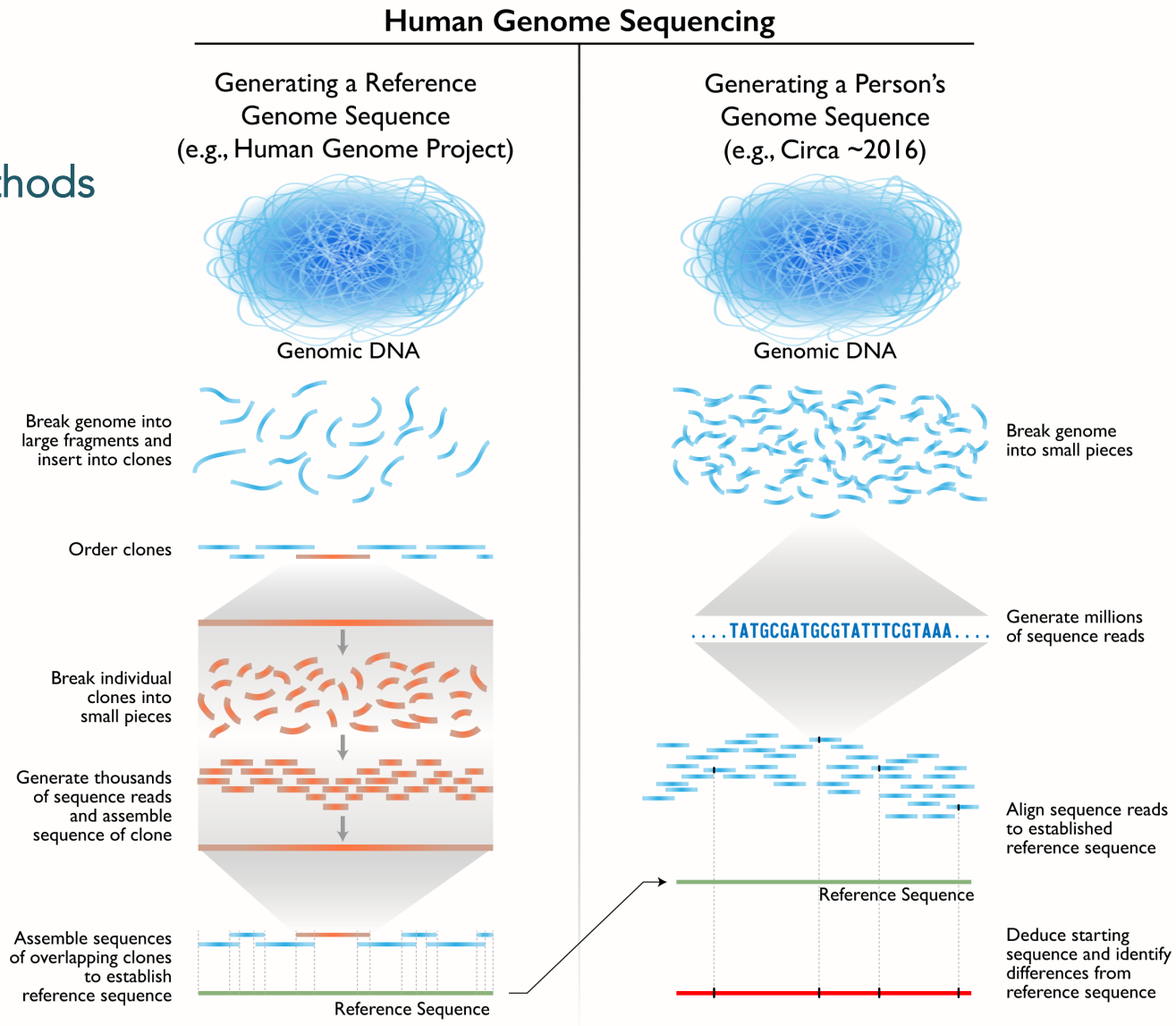
- **Accuracy**

More repeats than with Sanger sequencing → greater coverage, higher accuracy and sequence reliability (individual reads less accurate for NGS).

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

Comparison of human genome sequencing methods HGP vs.~ 2016

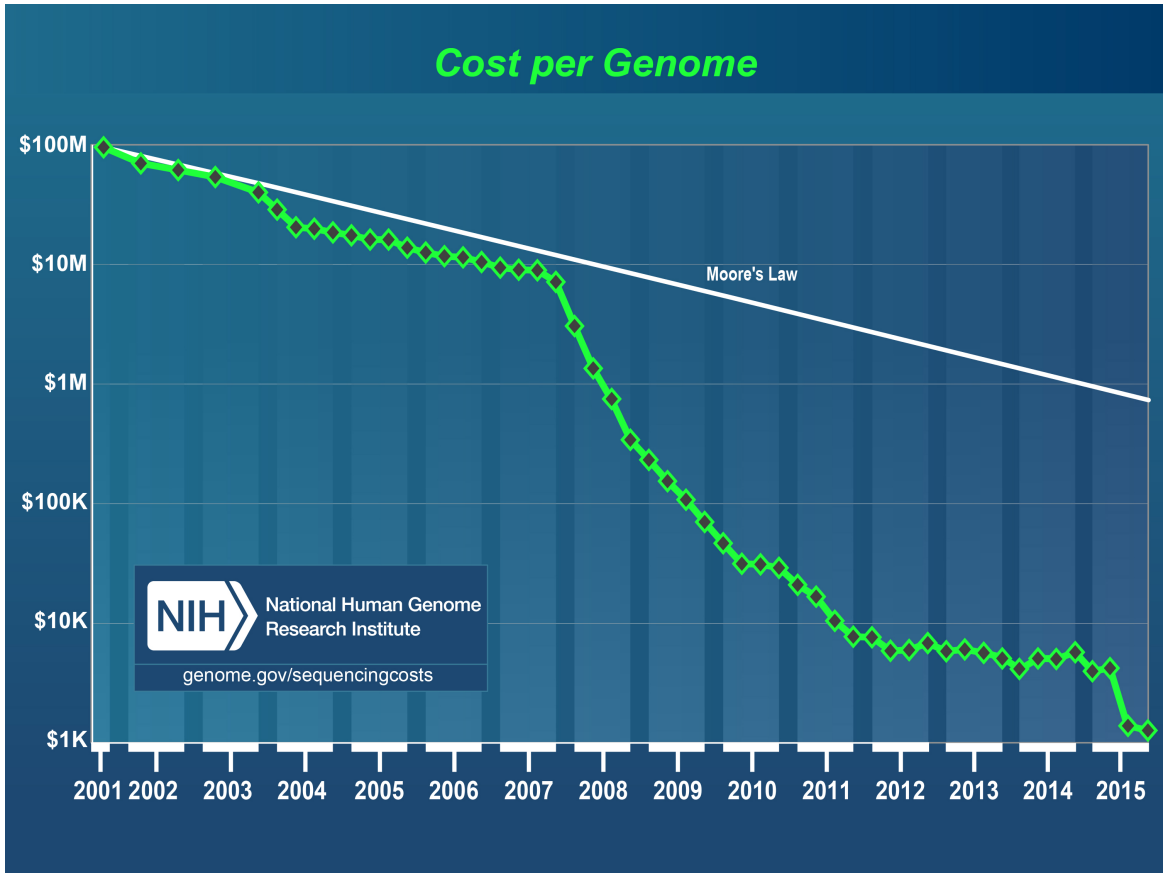


SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

Sequencing costs *vs.* technological advances

Data from the NHGRI (National Human Genome Research Institute)
→ Genome Sequencing Program (GSP)



PART 2

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

Sequencing costs *vs.* technological advances

HiSeq X Ten System

Parameter	Specification
Output per Run	Dual flow cell: 1.6-1.8 Tb
Single Reads Passing Filter	Dual flow cell: 5.3-6 billion
Supported Read Length	2 × 150 bp
Run Time	< 3 days
Quality	≥ 75% of bases above Q30 at 2 × 150 bp
Supported Library Preparation	TruSeq DNA PCR-Free Library Prep Kit TruSeq Nano DNA Library Prep Kit

Highlights

- **$1000 Genome Is a Reality**
  HiSeq X Ten System is the first and only platform to break the $1000 barrier human whole-genome sequencing
- **Population- and Production-Scale Whole-Genome Sequencing**
  HiSeq X Ten System delivers > 18,000 human genomes per year; HiSeq X Five System delivers > 9000 human genomes per year
- **Proven Performance**
  Take advantage of industry-leading data quality with the highly accurate Illumina sequencing by synthesis technology
- **Species Expansion**
  Now enabling cost-effective whole-genome sequencing of nonhuman species with unrivaled throughput

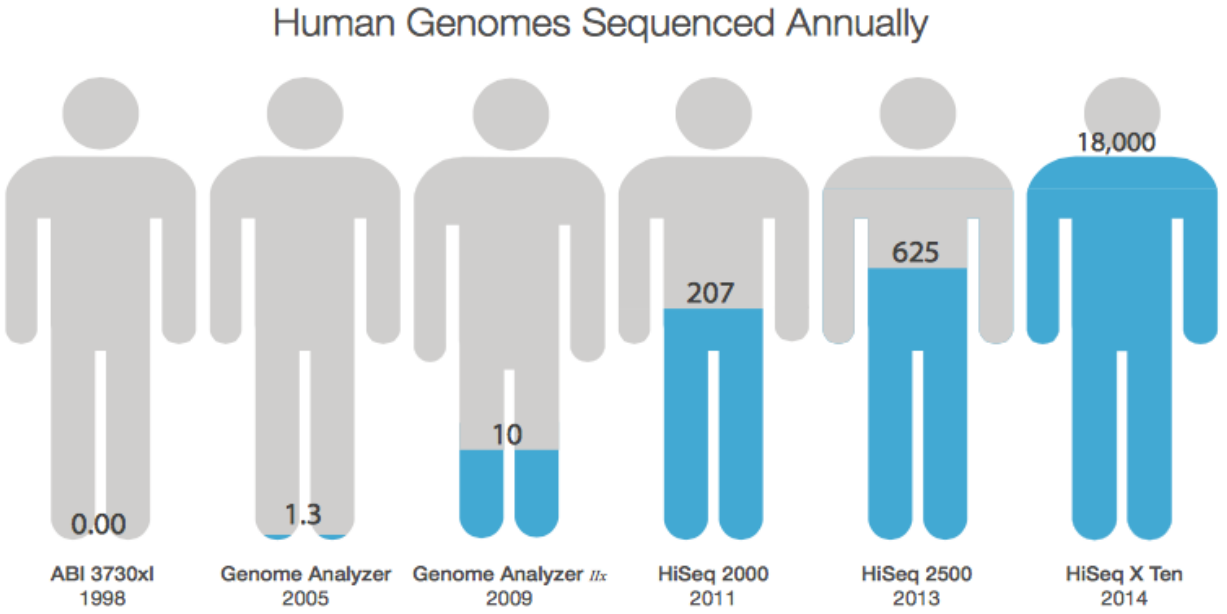
http://www.illumina.com/

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING

Sequencing costs *vs.* technological advances

1993-2001: First human genome: required 13-15 years. Cost ≈ 3 billion $.
2014: HiSeq X Ten can sequence over 45 human genomes in a single day for approximately $1000 each → ≈ 18,000 genomes / year.



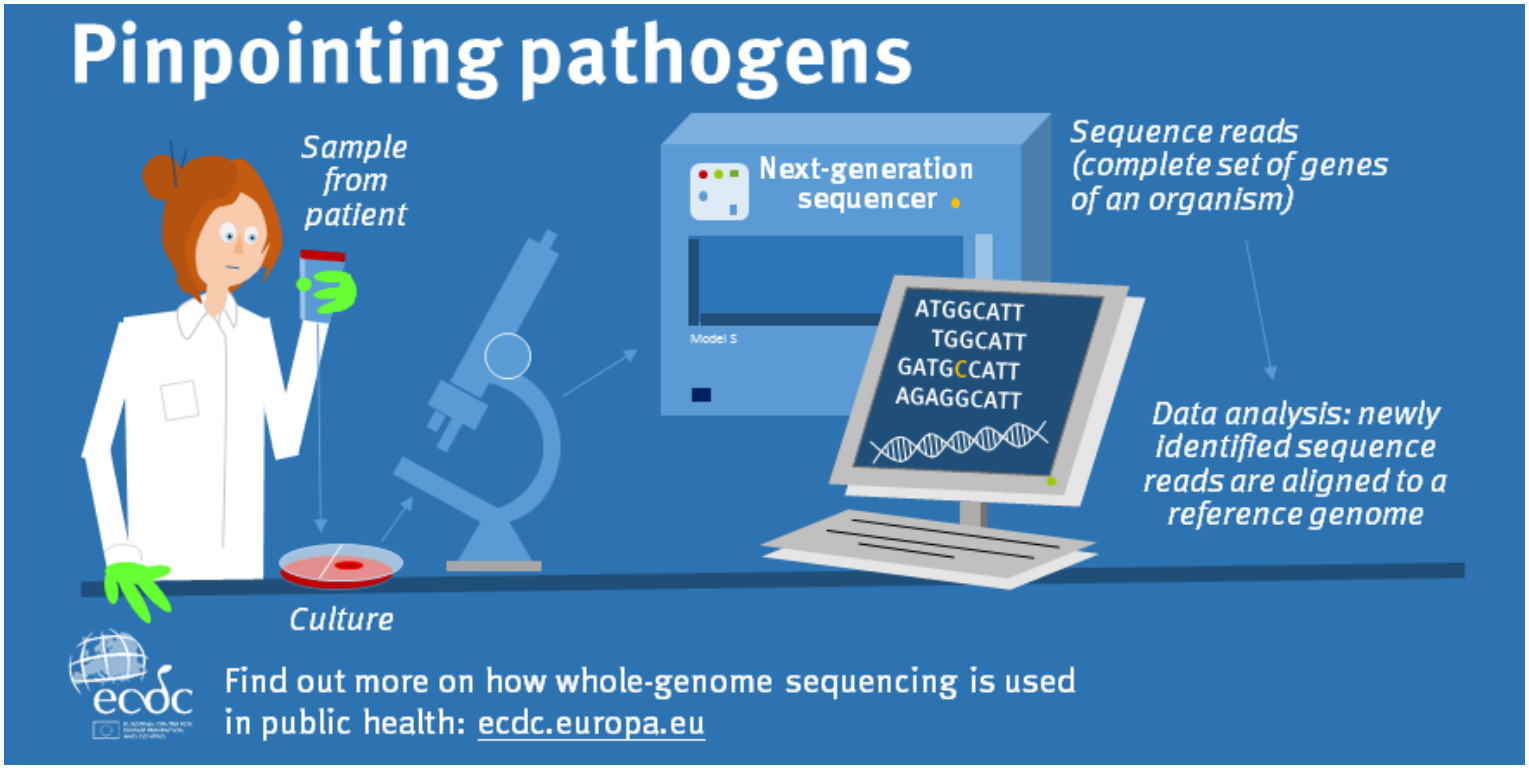
"Beyond the massive increase in data output, the introduction of NGS technology has transformed the way scientists think about genetic information."

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING : REAL_LIFE APPLICATIONS

NGS allows you to perform high-definition sequence-driven science
→ Epidemiological outbreaks…

ECDC (European Center for Disease Prevention and Control) : roadmap for integration of molecular typing and genomic typing into European-level surveillance and epidemic preparedness



SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING : REAL_LIFE APPLICATIONS

NGS allows you to perform high-definition sequence-driven science

→ Epidemiological outbreaks…

- *de novo* genome sequencing
- Genome assembly
- *In Vitro* Diagnostics
- Zoom in to deeply sequence target regions
- Simultaneously sequence enormous numbers of samples using multiplex sequencing with DNA barcode tags
- ….

(Fournier et al., 2013)

Table 1 | **Examples of infectious disease outbreaks that were investigated using next-generation sequencing**

Microorganism	Location	Year	Reference
Carbapenem-resistant *Klebsiella pneumoniae*	USA	2011	112
Clostridium difficile	Worldwide	2013	113
Escherichia coli O104:H4	Germany	2011	114,115
Legionella pneumophila serogroup 1	United Kingdom	2013	116
Methicillin-resistant *Staphylococcus aureus* (MRSA)	United Kingdom	2009	117
Mycobacterium tuberculosis	Canada	2006–2008	118
Vibrio cholerae O1 biovar El Tor	Haiti	2010–2011	119
Arenavirus	Australia	2008	120
Bas-Congo virus	Democratic Republic of the Congo	2009	121
Influenza A virus H1N1	Worldwide	2009	122

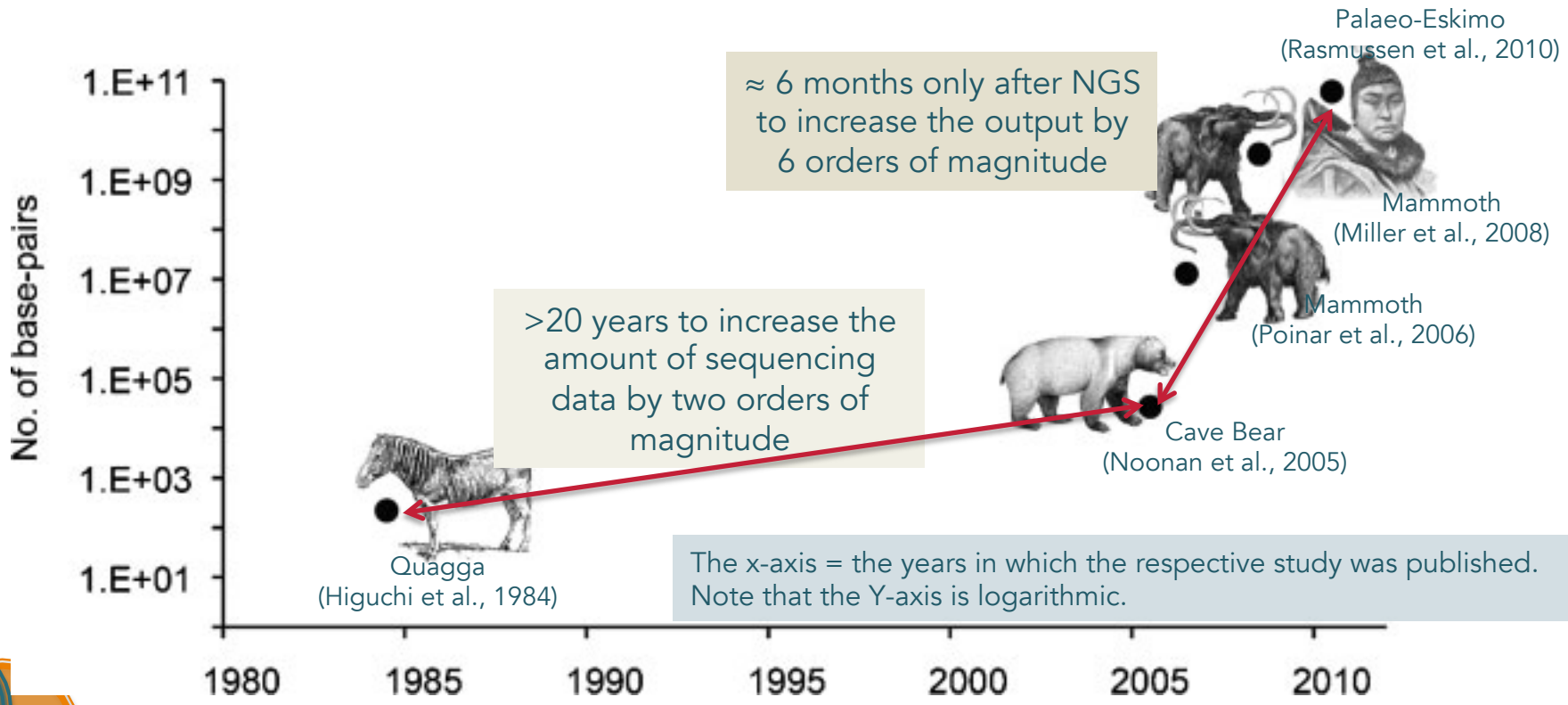
PART 2

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING : REAL_LIFE APPLICATIONS

NGS is used in many specific fields such as Paleogenomics

Ancient DNA sequencing (extinct organisms, limited sample)
→ increased output from ancient DNA after NGS



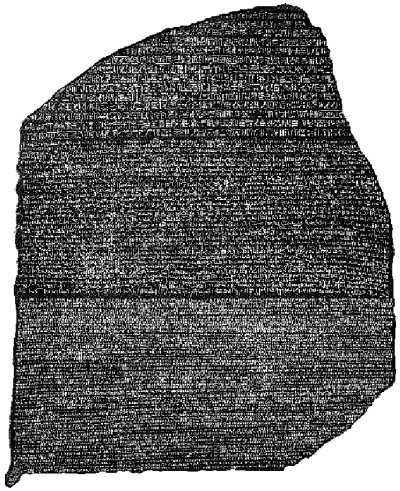
(Shapiro & Hofreiter, 2010)

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING : REAL_LIFE APPLICATIONS

NGS is used in many specific fields such as Comparative Genomics

Comparative genomics is based on homology and evolutionary dynamics between organisms



COMPARING CLOSELY RELATED SPECIES

COMPARING EVOLUTIONARILY DISTANT SPECIES

PART 2

GENOMES & TRANSCRIPTOMES

SEQUENCING

Moreover: thinking at a population scale is possible !

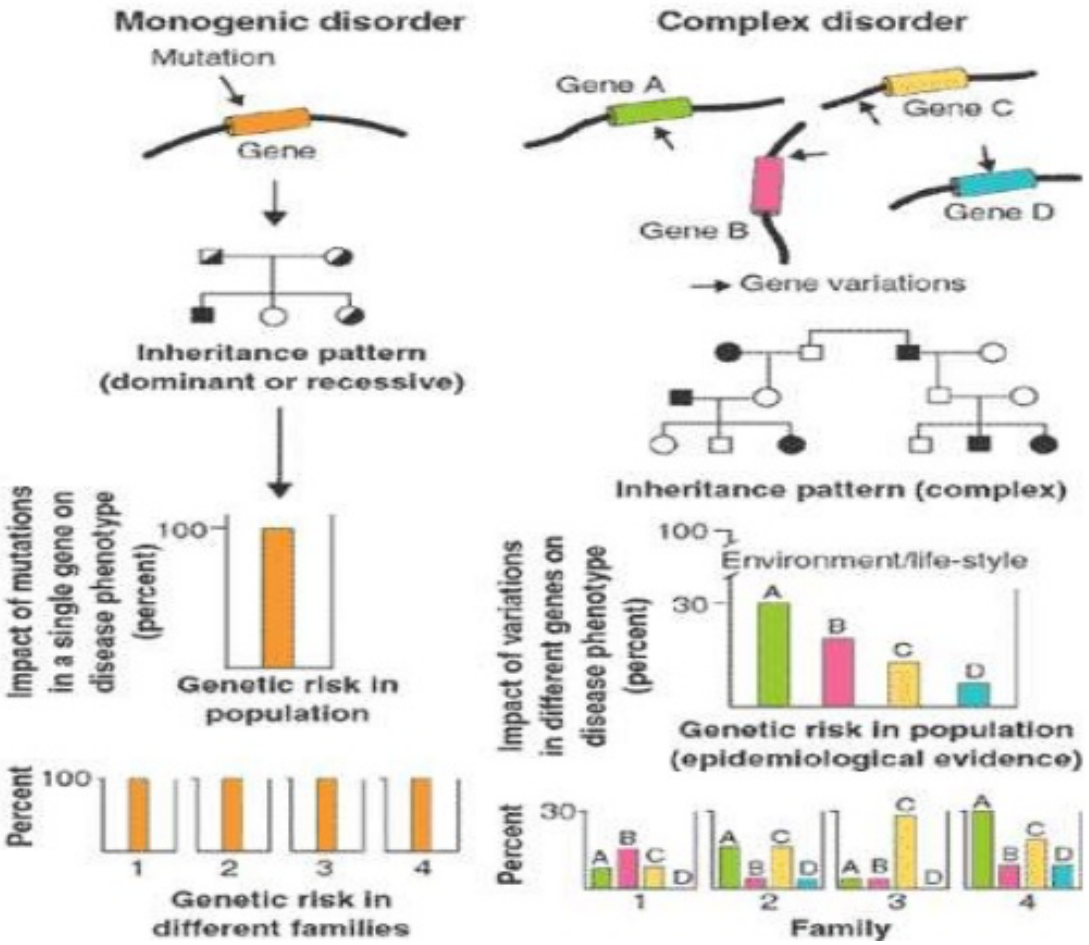
- NGS allows you to think at the individual and population scales.
- Explain what determines the complexity of living organisms, etc…

Humans as an example:

No two individuals are similar:

- What determines their different characteristics ?
- What determines their different susceptibility to diseases?

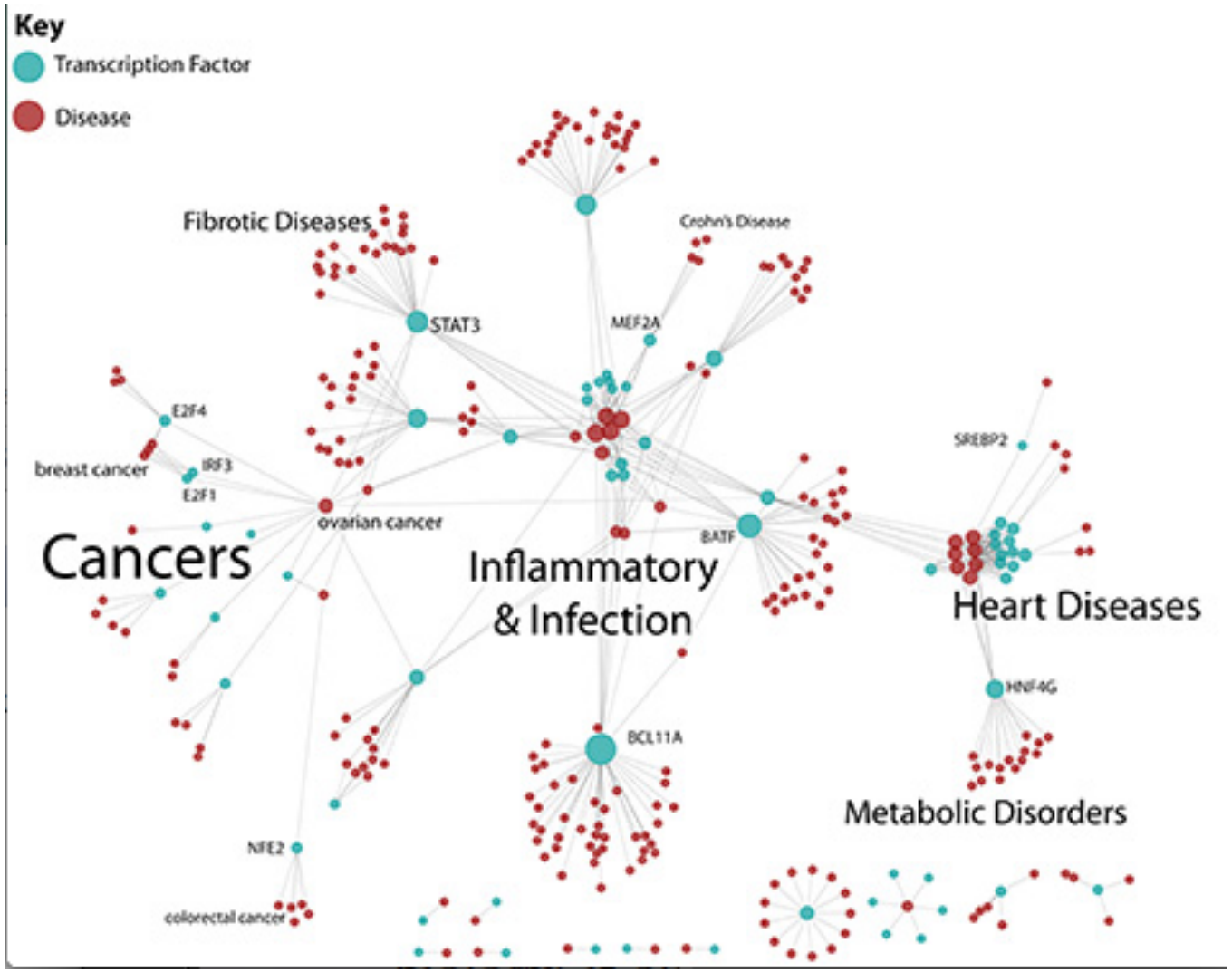
Monogenic vs. Complex Disorders



→ Link between thousands of genetic variants to complex diseases possible !

SEQ TECHNOLOGIES

ADVANCES MADE TO STUDY GENES AND GENE EXPRESSION

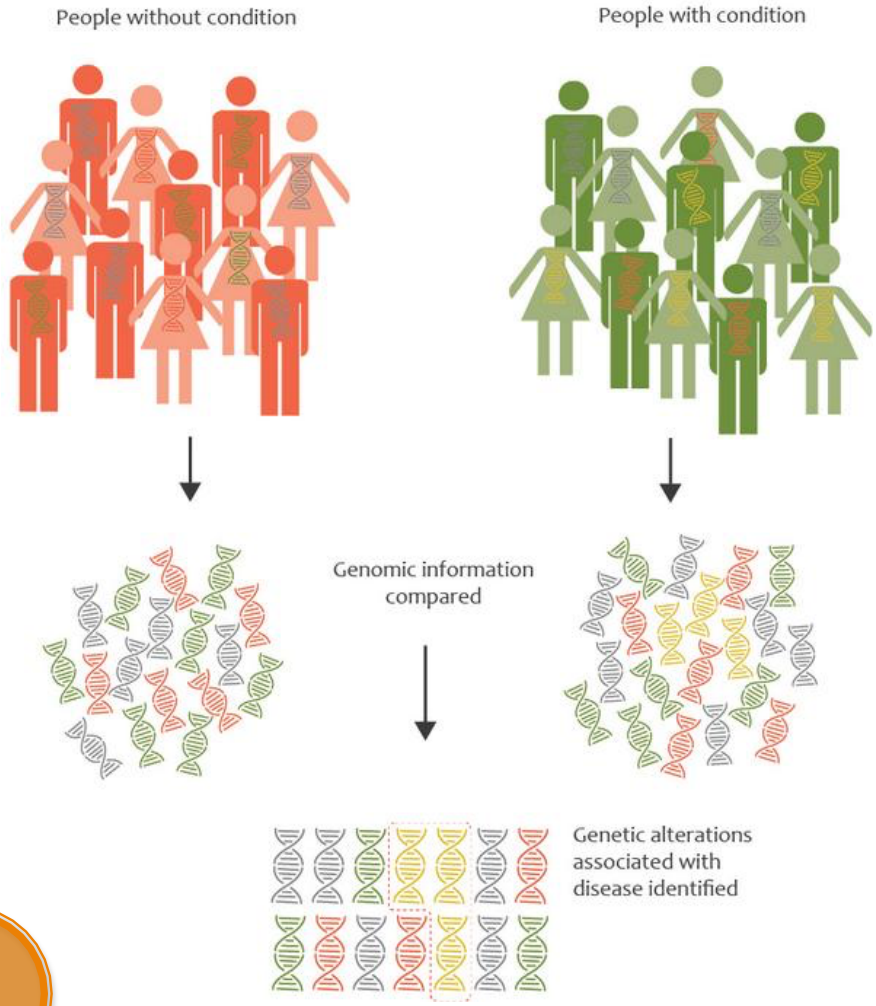


http://biomedicalcomputationreview.org/content/unlocking-genetics-complex-diseases-gwas-and-beyond

GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

How researchers compare genomic information to identify genetic alterations



Genomics

- ***Genomics*** is the study of genomes, including large chromosomal segments containing many genes.
- The *initial phase of genomics* aims to map and sequence an initial set of entire genomes.
- *Functional genomics* aims to deduce information about the function of DNA sequences.
  - Should continue long after the initial genome sequences have been completed.

https://twitter.com/genomicsedu/status/525219646476410880

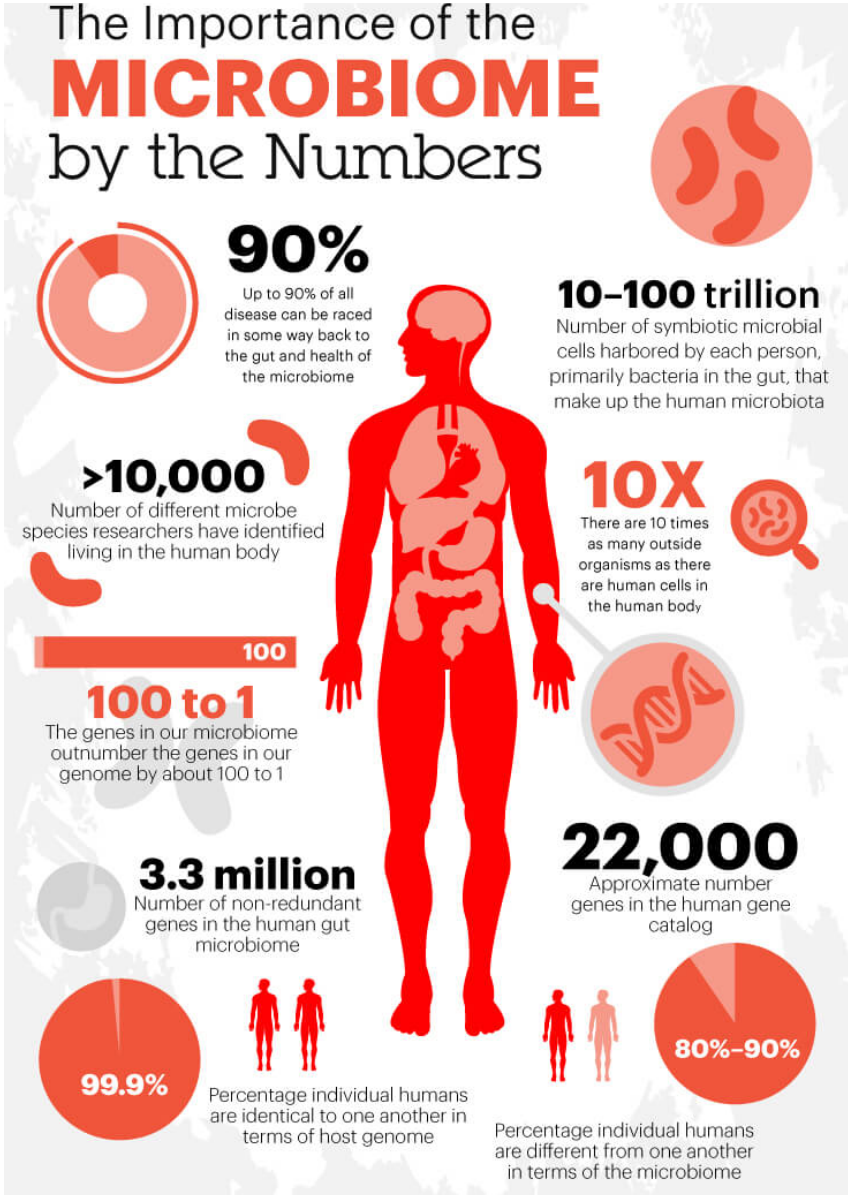
GENOMES & TRANSCRIPTOMES

HGP: THE HERITAGE

• "microbiome" :"micro" means small and "biome" means a habitat of living things.

- Popular among researchers : up to **90% of all diseases** can be traced back directly or indirectly to health of the microbiome.

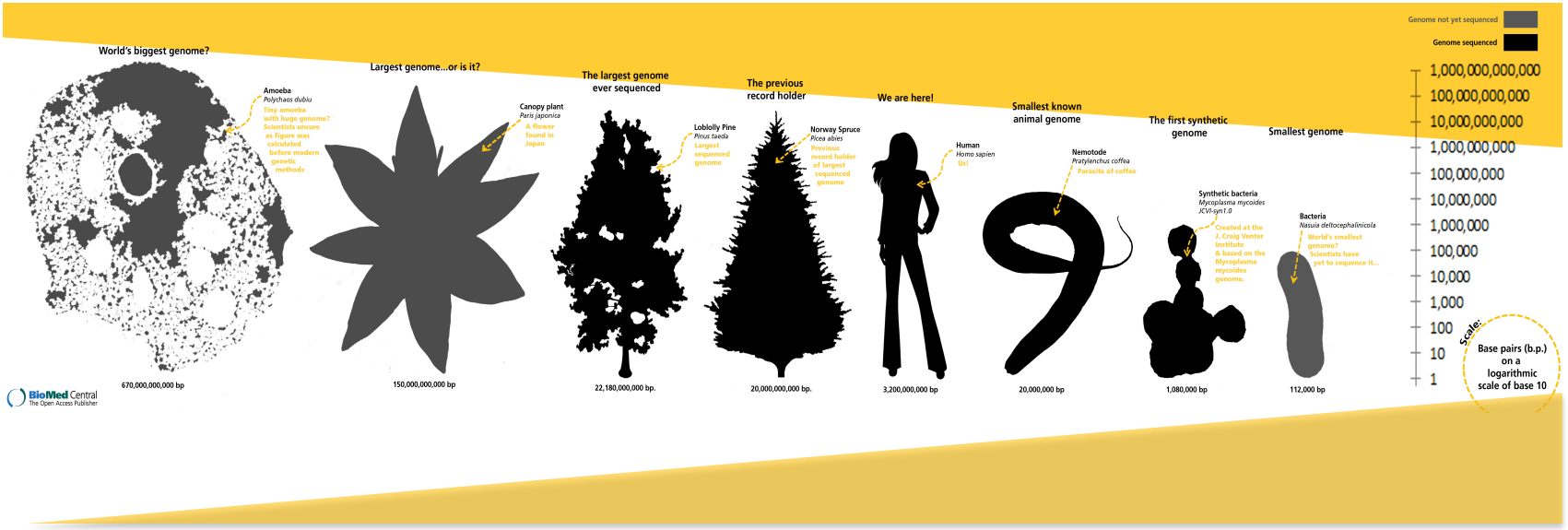
- Our microbiome is home to **trillions of microbes**, diverse organisms that help govern nearly every function of the human body in some way.



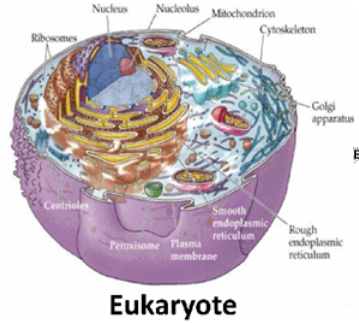
https://draxe.com/microbiome/

GENOMES & TRANSCRIPTOMES

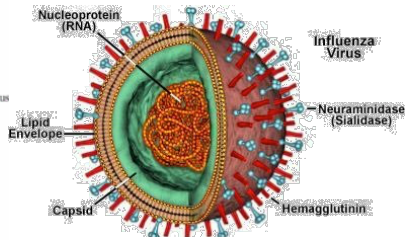
HGP: THE HERITAGE



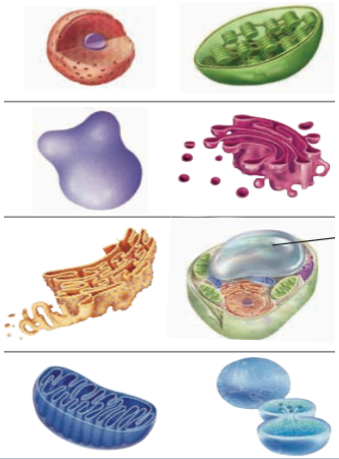
EUKARYOTES
3694



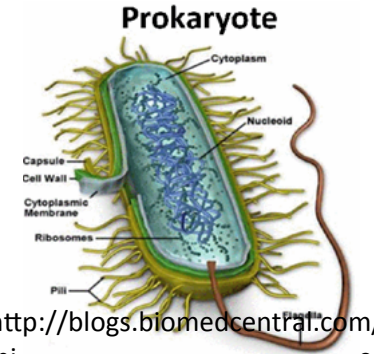
VIRUSES
5884



ORGANELLES
8634



PROKARYOTES
74156

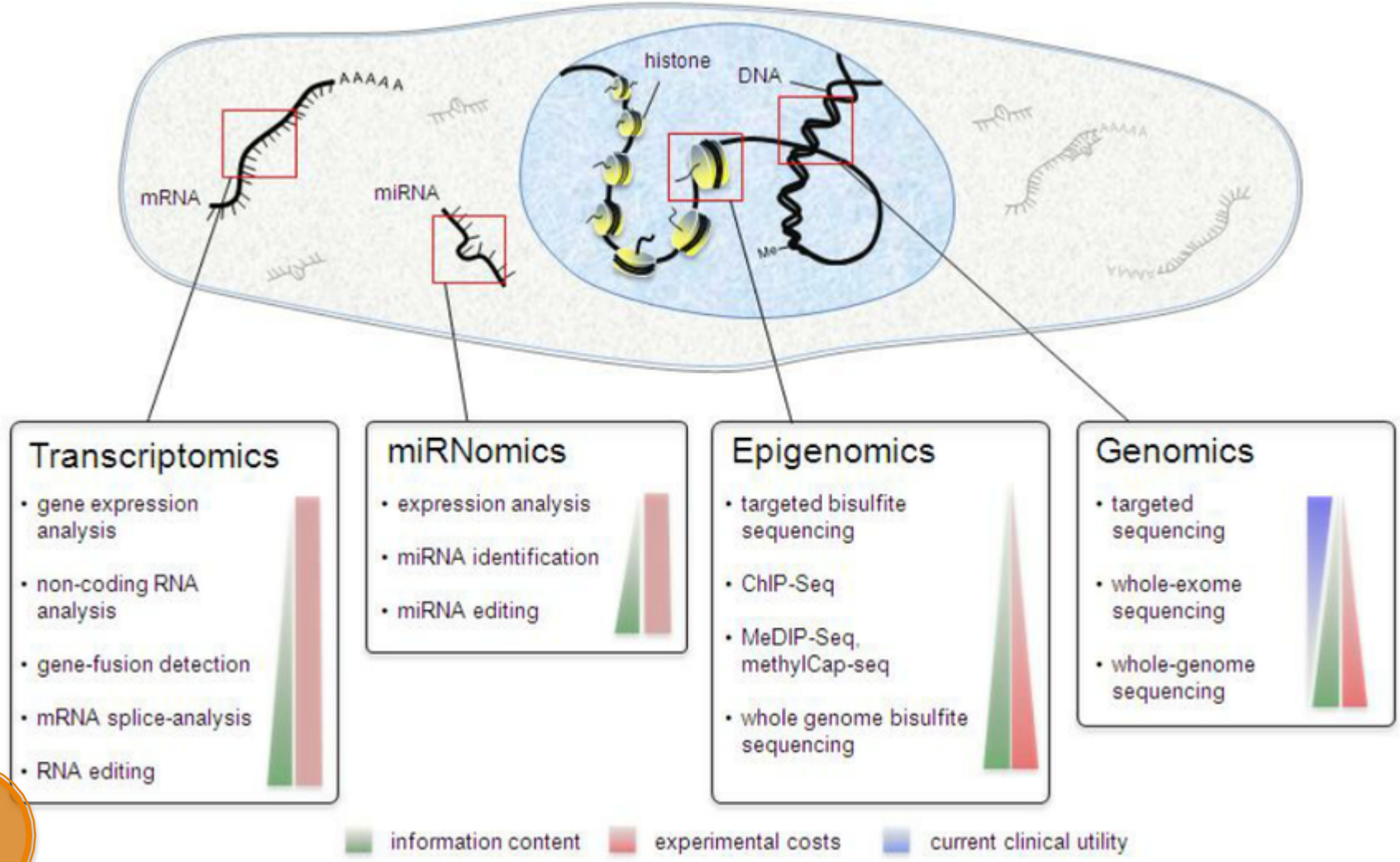


Perman, 2014: http://blogs.biomedcentral.com/
https://www.ncbi.nlm.nih.gov/genome/browse/
https://www.quora.com/

SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING : REAL_LIFE APPLICATIONS

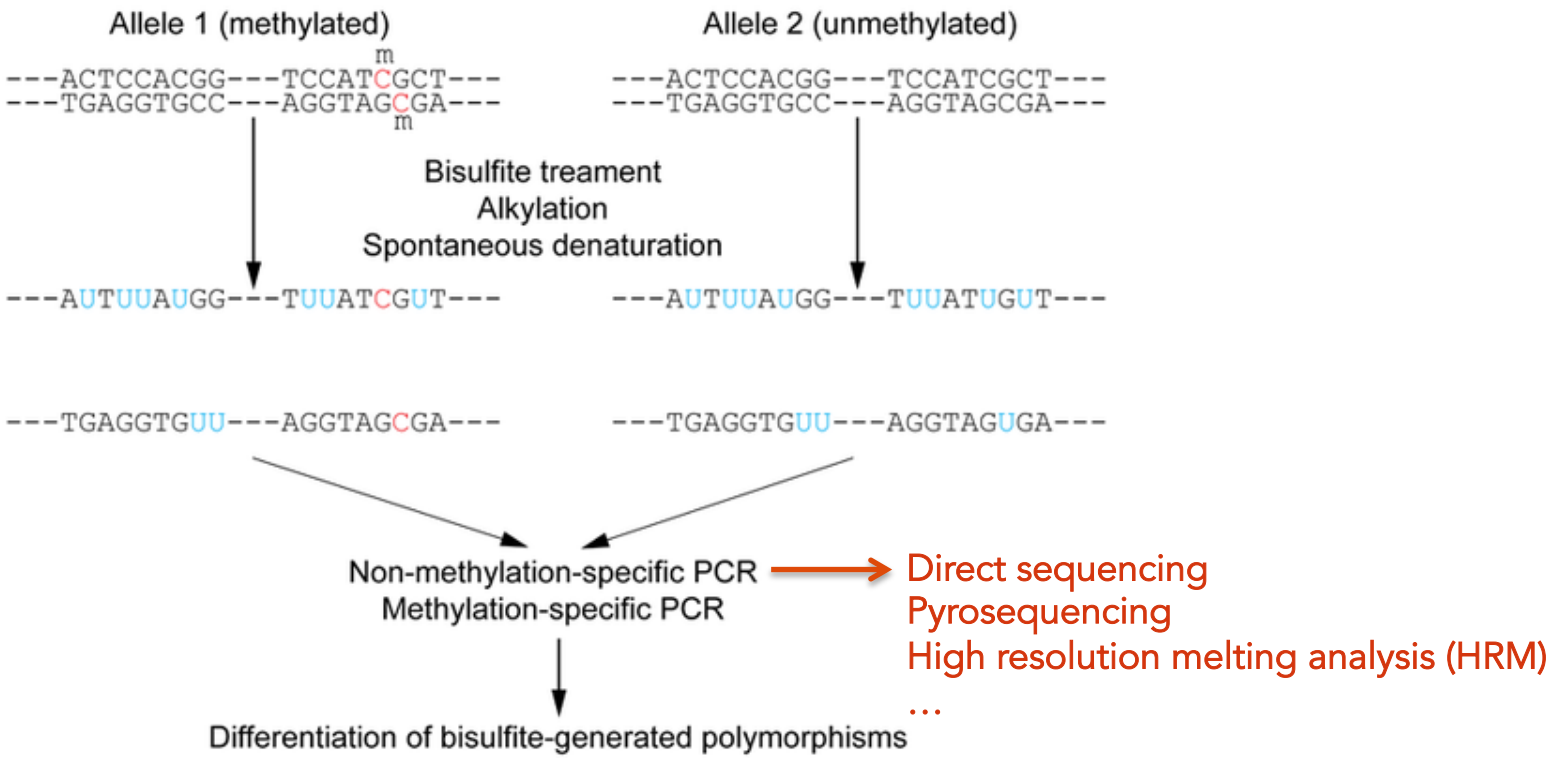
Sequencing today : a broad range of applications thanks to NGS



(Frese et al., 2013)

PART 2

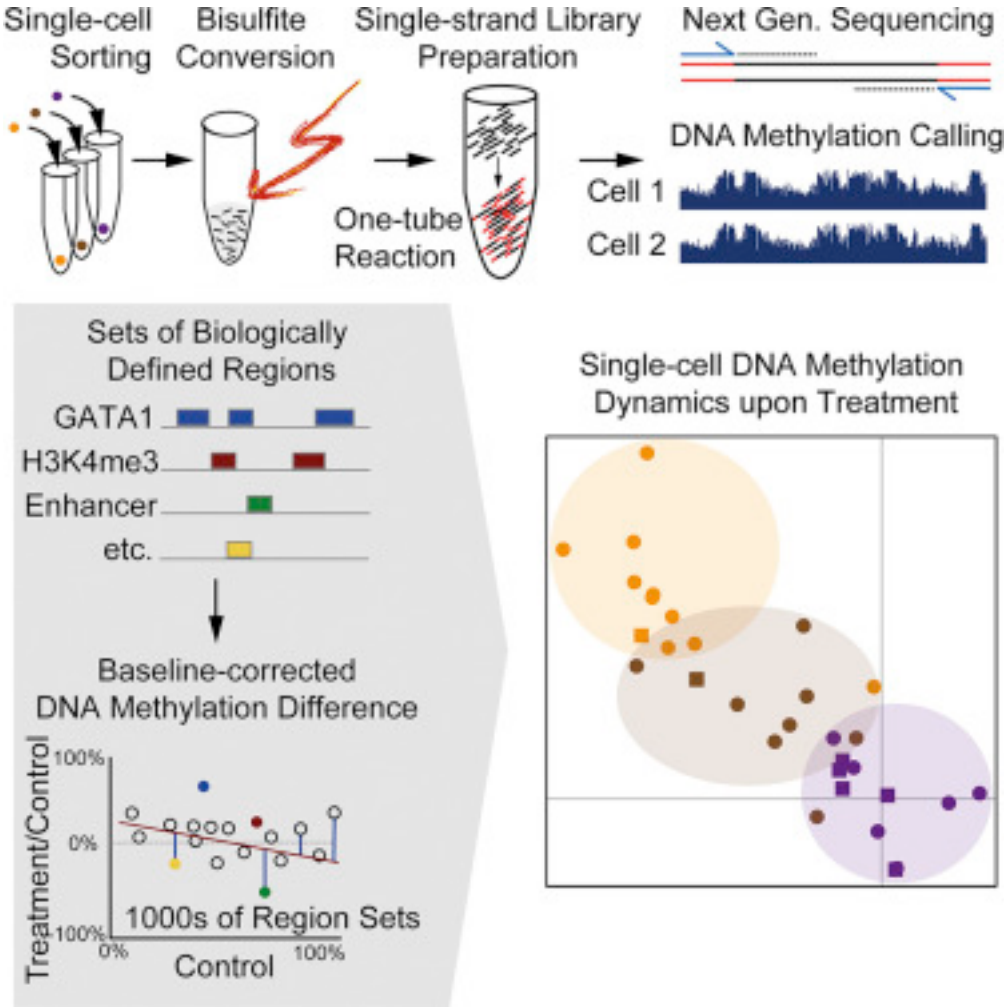
Sequencing today : a broad range of applications thanks to NGS



SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING : REAL_LIFE APPLICATIONS

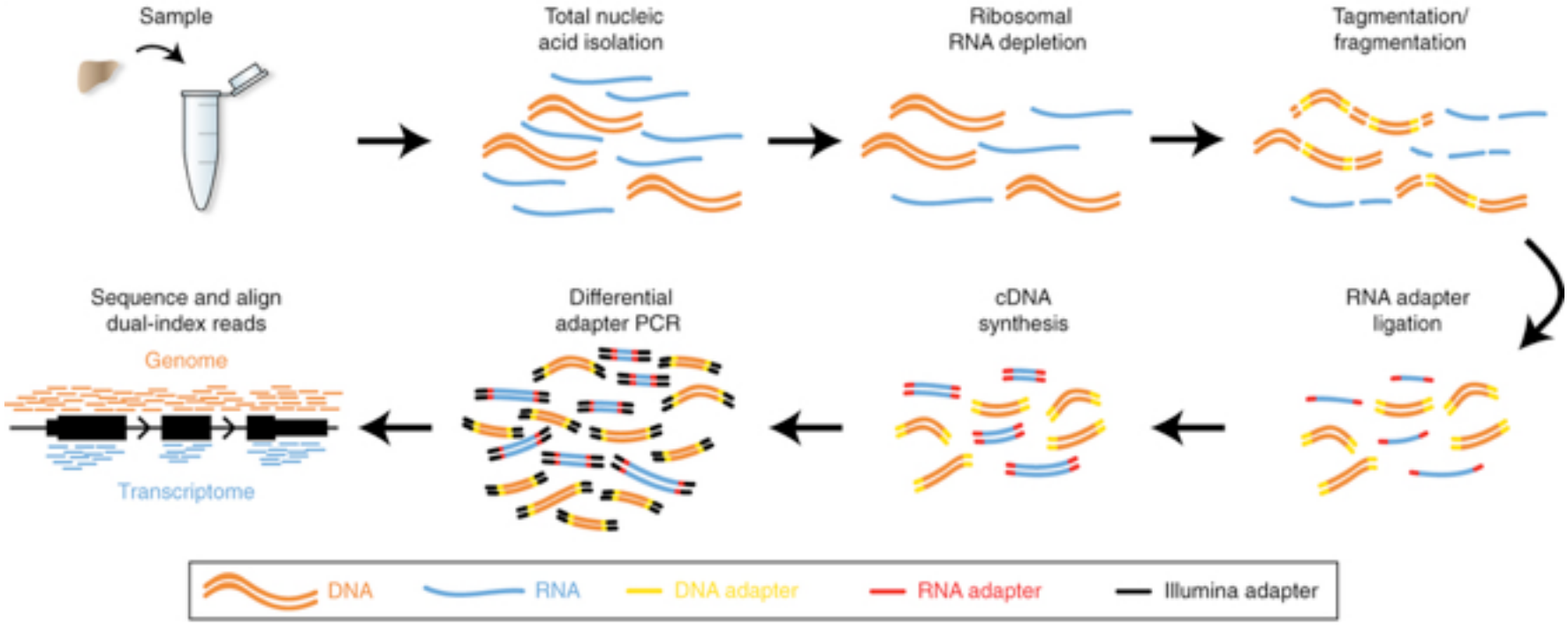
Sequencing today : a broad range of applications thanks to NGS



https://en.wikipedia.org/wiki/Single_cell_sequencing

- NGS: New improvements

Combined DNA-seq and RNA-seq (« Simul-seq »)



(Reuter et al., 2016)

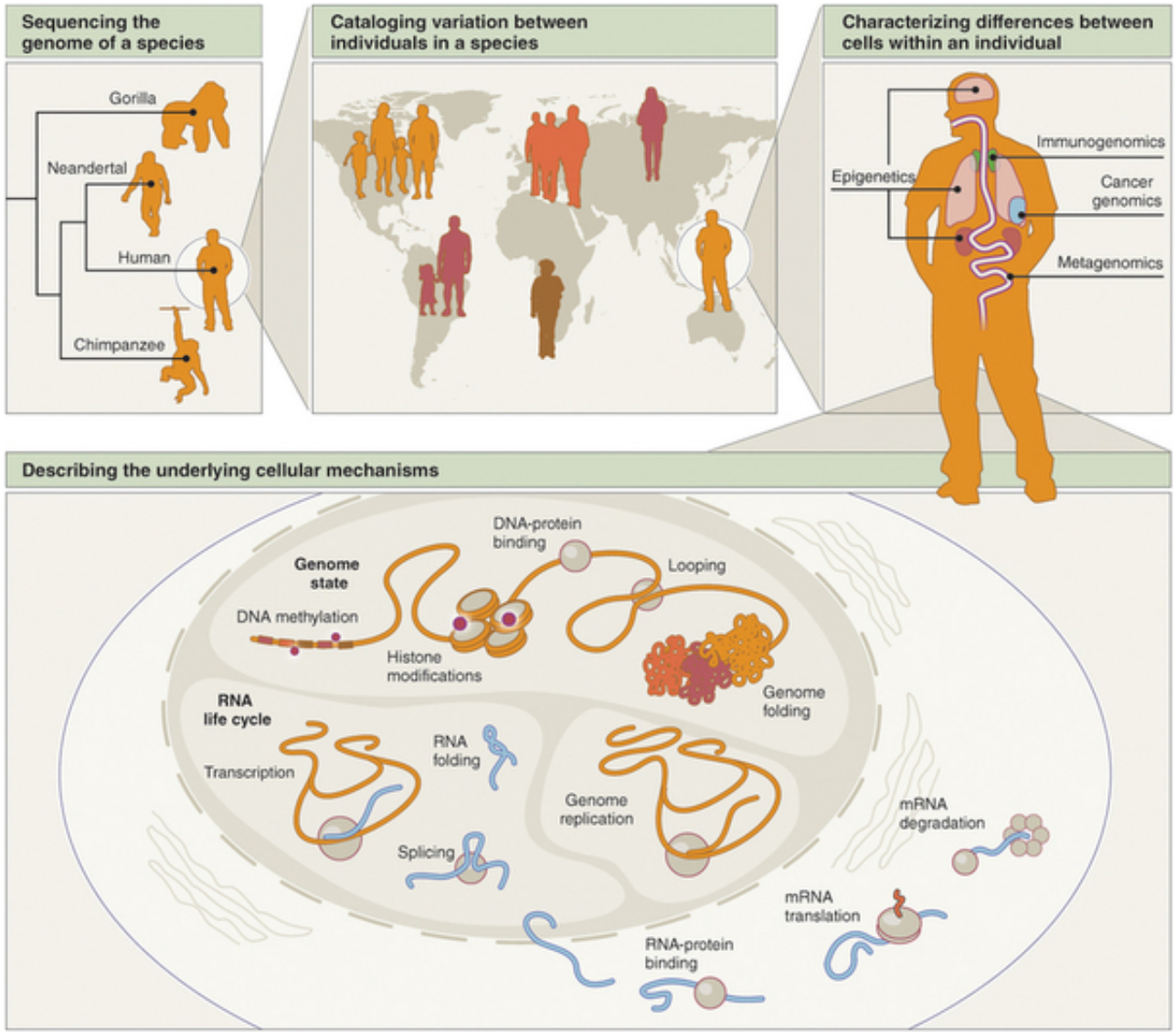
SEQ TECHNOLOGIES

NEXT-GENERATION SEQUENCING: REAL-LIFE APPLICATIONS

- NGS

Where NGS can help?

at all levels of information required !

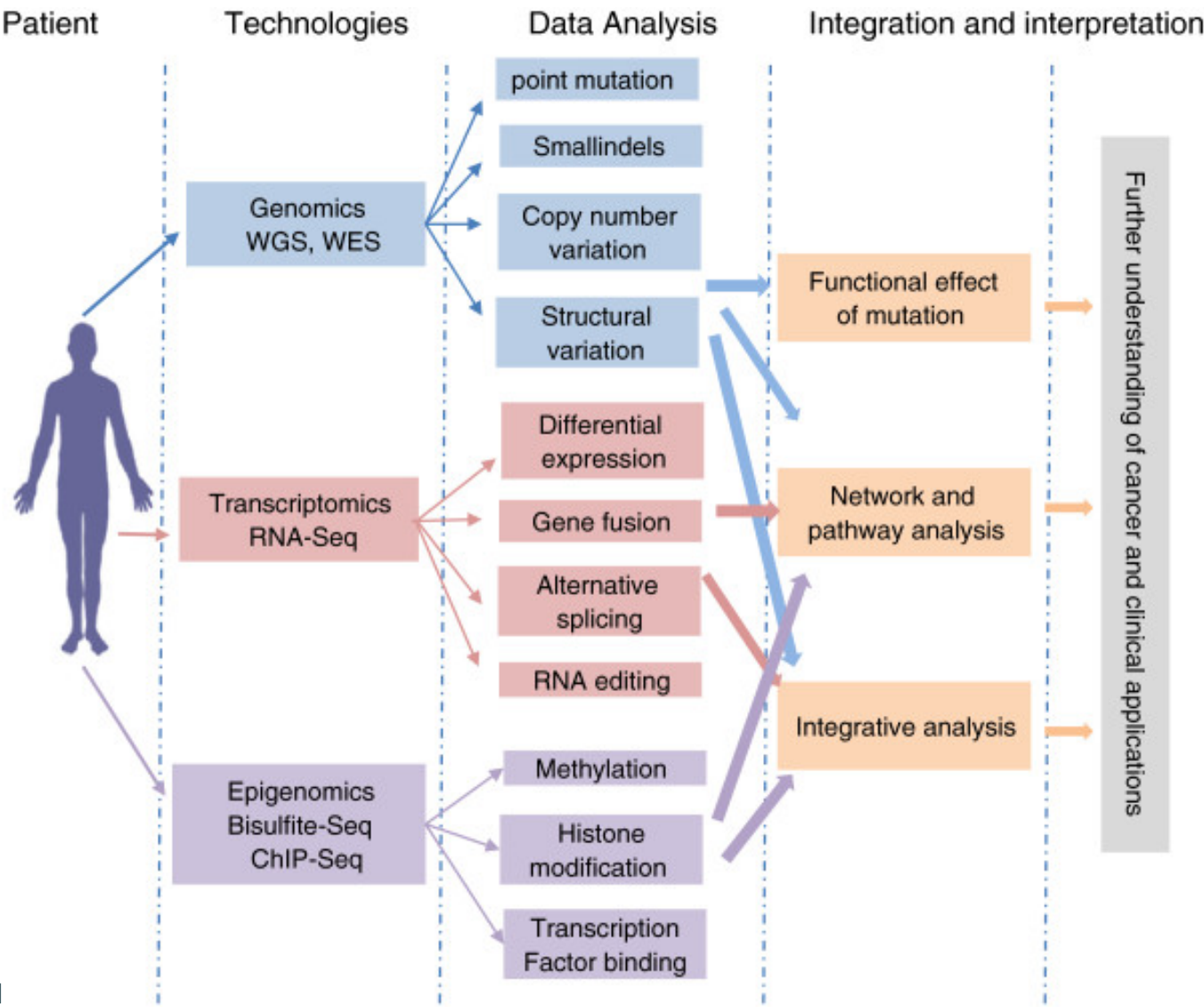


(Shendure & Aiden, 2012)

SEQ TECHNOLOGIES

TAKE-HOME MESSAGES : NEXT-GENERATION SEQUENCING

Sequencing today : a broad range of applications thanks to NGS



(Shyr and Liu, 2013)

SEQ TECHNOLOGIES

ADVANCES MADE TO STUDY GENES AND GENE EXPRESSION

High-throughput Approach

→ The results of HT experiments provide starting points of knowledge, and are not the end points !!!

Very useful to:

- Identify genetic variants
- Understand the effect of a particular treatment
- Understand the real interactions between genes expressed in a particular condition
- Drug design

…

But…

SEQ TECHNOLOGIES

TAKE-HOME MESSAGES

NGS or previous technologies?

There is no single method better than any other one, all depends on your question of interest !

One can envision a powerful symbiosis between previous technologies (microarrays…) and NGS technologies.

- Arrays may be best suited in classifying cohorts of samples, such as tumor tissues.

- Once samples of interest are defined, NGS could be used to provide comprehensive deep-sequence analysis of either genomic DNA to identify mutations, or RNA to report differences at the transcriptome level.

(Hurd and Nelson, 2009)