dcvmn
Developing Countries Vaccine Manufacturers Network

Webinar 13 October 2021 9:00-10:00 CET

Next generation sequencing

Core technology in vaccinomics

Domenico Genovese
Centro Nazionale Controllo e Valutazione dei Farmaci
Istituto Superiore di Sanità

❑ Outline (1)

- Introduction
- NGS: second generation sequencing
- NGS: third generation sequencing
- NGS: fourth generation sequencing
- Data Analysis: A new syntax

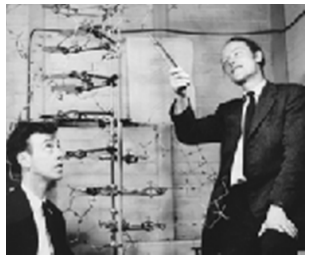
APPLICATIONS OF NGS

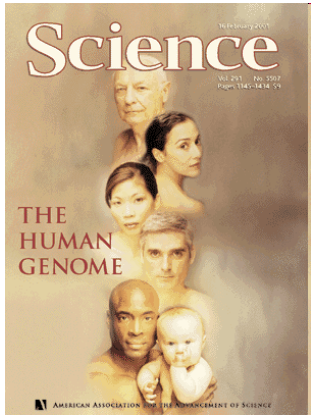
NGS allows labs to:

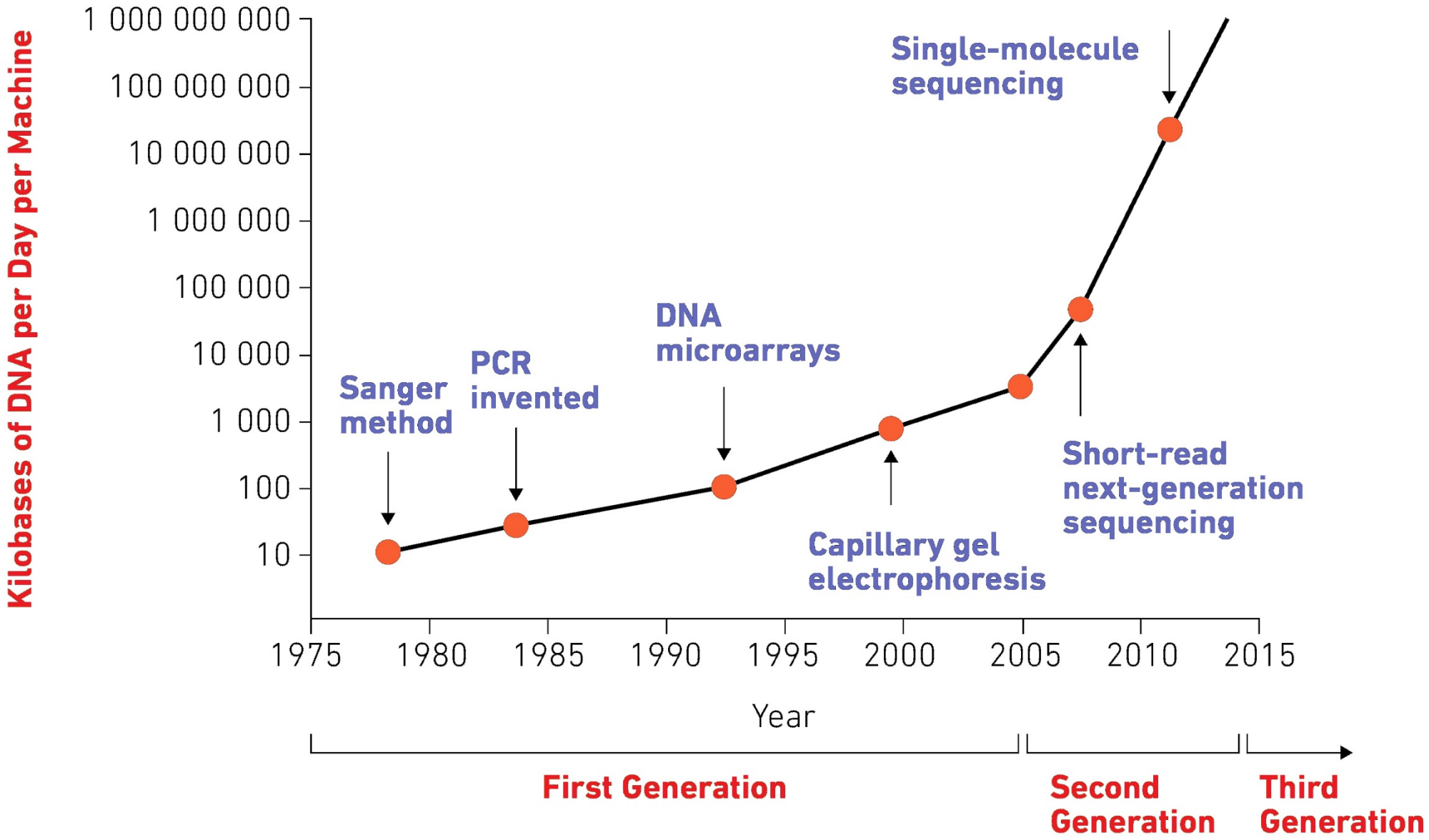
- Mutation discovery
- Transcriptome Analysis – RNA-Seq to discover novel RNA variants and splice sites, or quantify mRNAs for gene expression analysis
- Sequencing clinical isolates in strain-to-reference mechanisms.
- Enabling Metagenomics
- Defining DNA-Protein interactions – ChIP-Seq
- Discovering non-coding RNAs
- Molecular diagnostics for Oncology & Inherited Disease study.
- Gene Regulation Analysis
- Whole Genome Sequencing
- Exploring Chromatin Packaging
- Deeply sequence target regions
- Analyze epigenetic factors such as genome-wide DNA methylation and DNA-protein interactions
- Study the human microbiome

A quick history of sequencing

1869 – Discovery of DNA
1909 – Chemical characterisation
1953 – Structure of DNA solved
1977 – First genome (ΦX174) Sequencing by synthesis (Sanger) - Sequencing by degradation (Maxam- Gilbert)
1986 – First automated sequencing machine 1990 – Human Genome Project started
1992 – First "sequencing factory" at TIGR
1995 – First bacterial genome – *H. influenzae* (1.8 Mb) 1998 – First animal genome – *C. elegans* (97 Mb)
2003 – Completion of Human Genome Project (3 Gb)
– 13 years, $2.7 bn
2005 – First "next-generation" sequencing instrument
2013 – 10,000 genome sequences in NCBI database

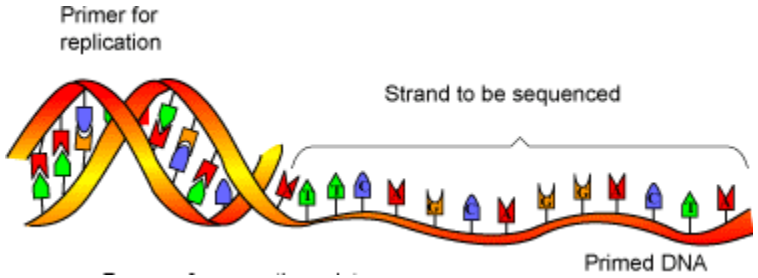


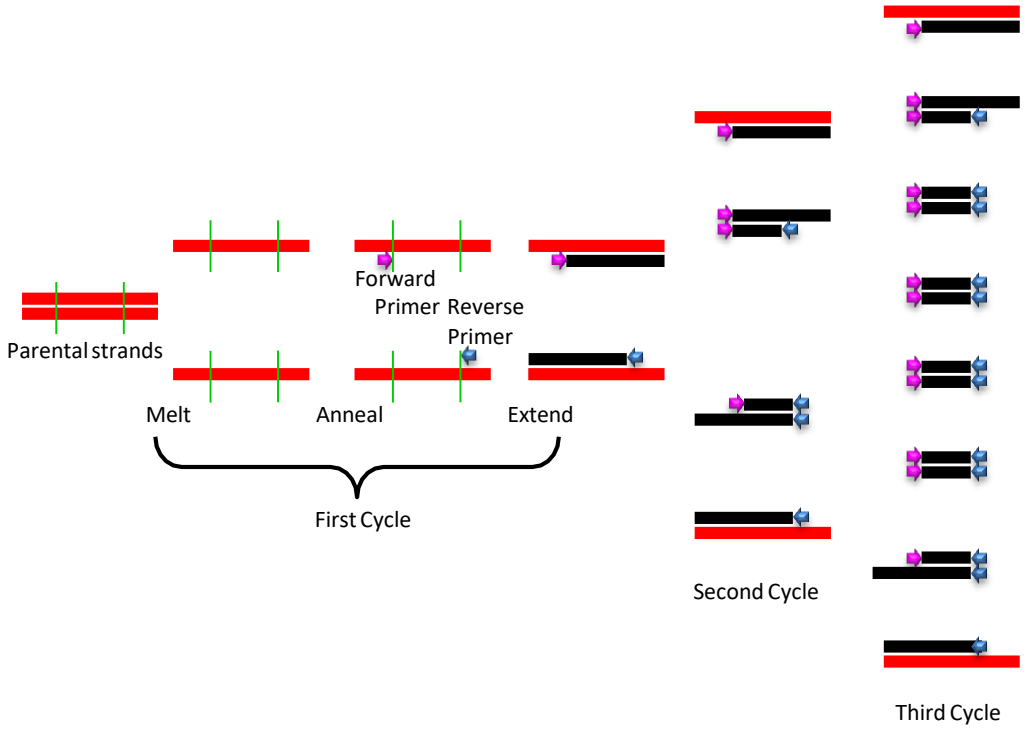




Sanger Method

- Sequence of interest is targeted via designed primers
- Massive amplification of target (e.g. by ca. 35 rounds of PCR amplification)



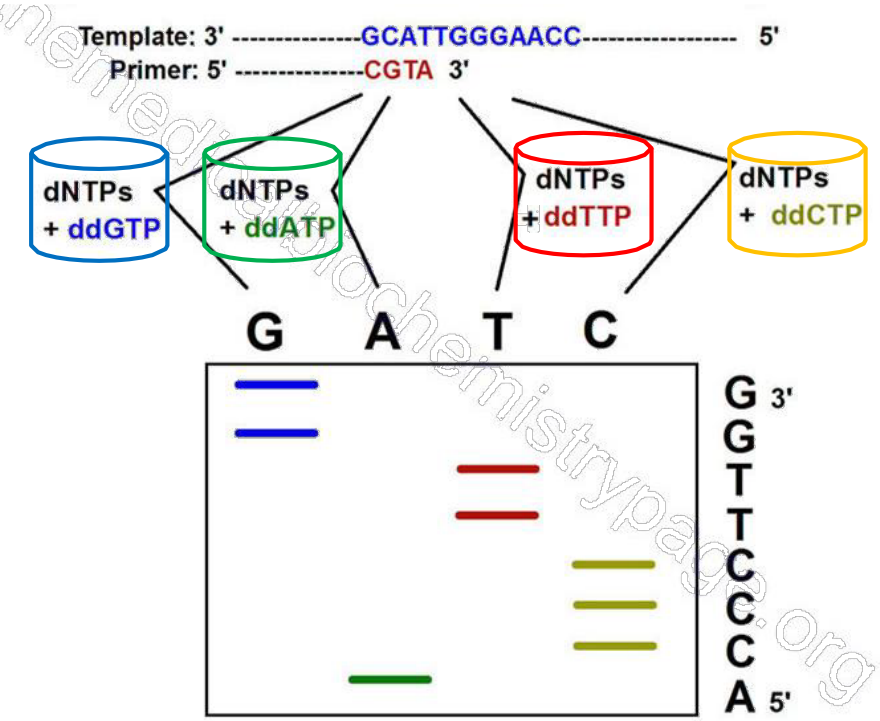


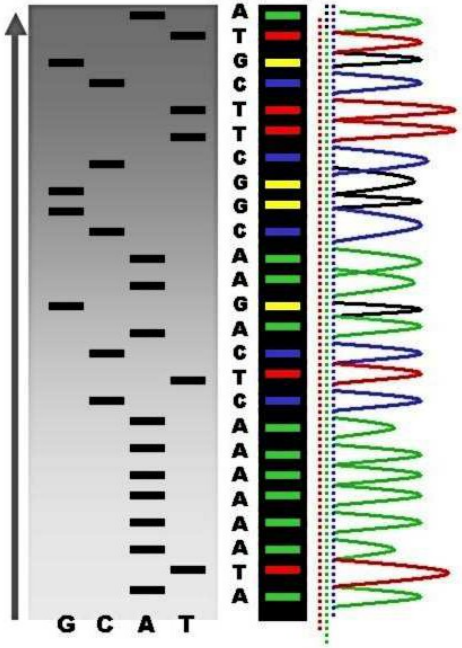
Sanger sequencing: dye-terminator sequencing

- Elongation of DNA sequences by polymerase
- Enzyme stops at a random position per copy (by **d**dNTP)
- Terminated copies are separated within a gel (smaller ones run further)
- Sequence can be read directly

→ Extremely accurate: „gold standard“ (error rate ~1:100,000)

→ Slow: poorly parallelizable (60x max.)

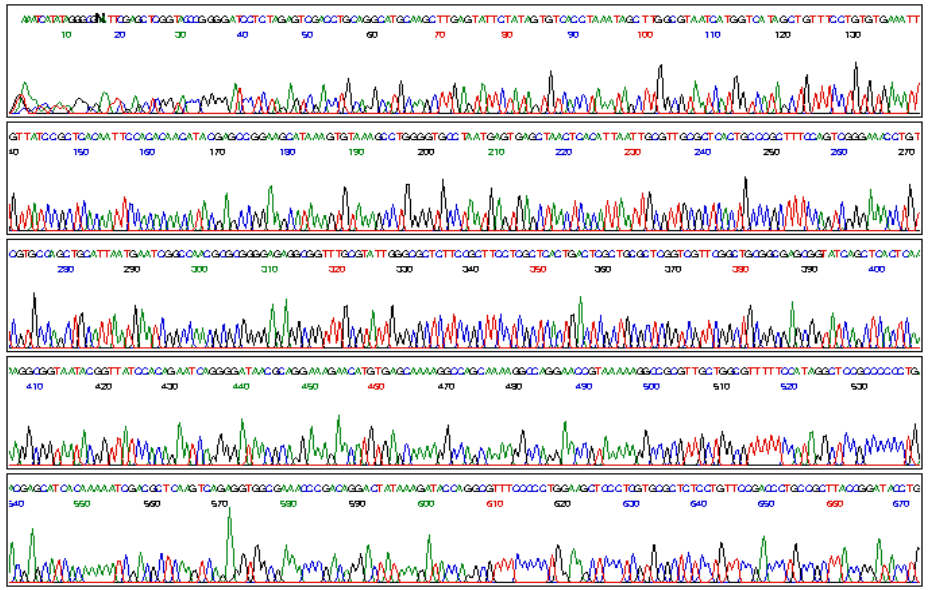




Automated DNA Sequencers

ABI 377 Plate Electrophoresis

ABI 3730 xl Capillary Electrophoresis



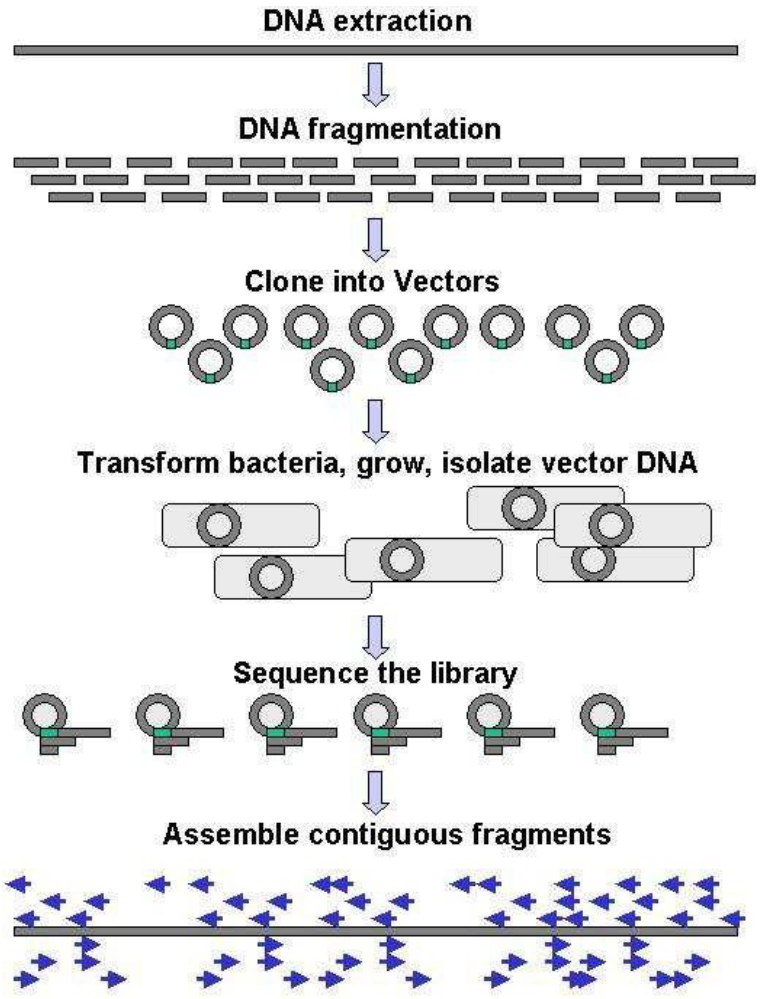
Human Genome Project

Launched in 1989 –expected to take 15 years

Competing Celera project launched in 1998

- 1° Draft released in 2000
  - "Complete" genome released in 2003
  - Sequence of last chromosome published in 2006
- Cost: ~$3 billion
  - Celera ~$300 million

Shotgun library preparation



NGS: second generation sequencing

Second-generation sequencing methods are well-established and share many features in common.

They can be subdivided according to their underlying detection chemistries including sequencing by ligation (incorporating nanoball) and sequencing by synthesis (SBS), which further divides into proton detection, pyrosequencing and reversible terminator.

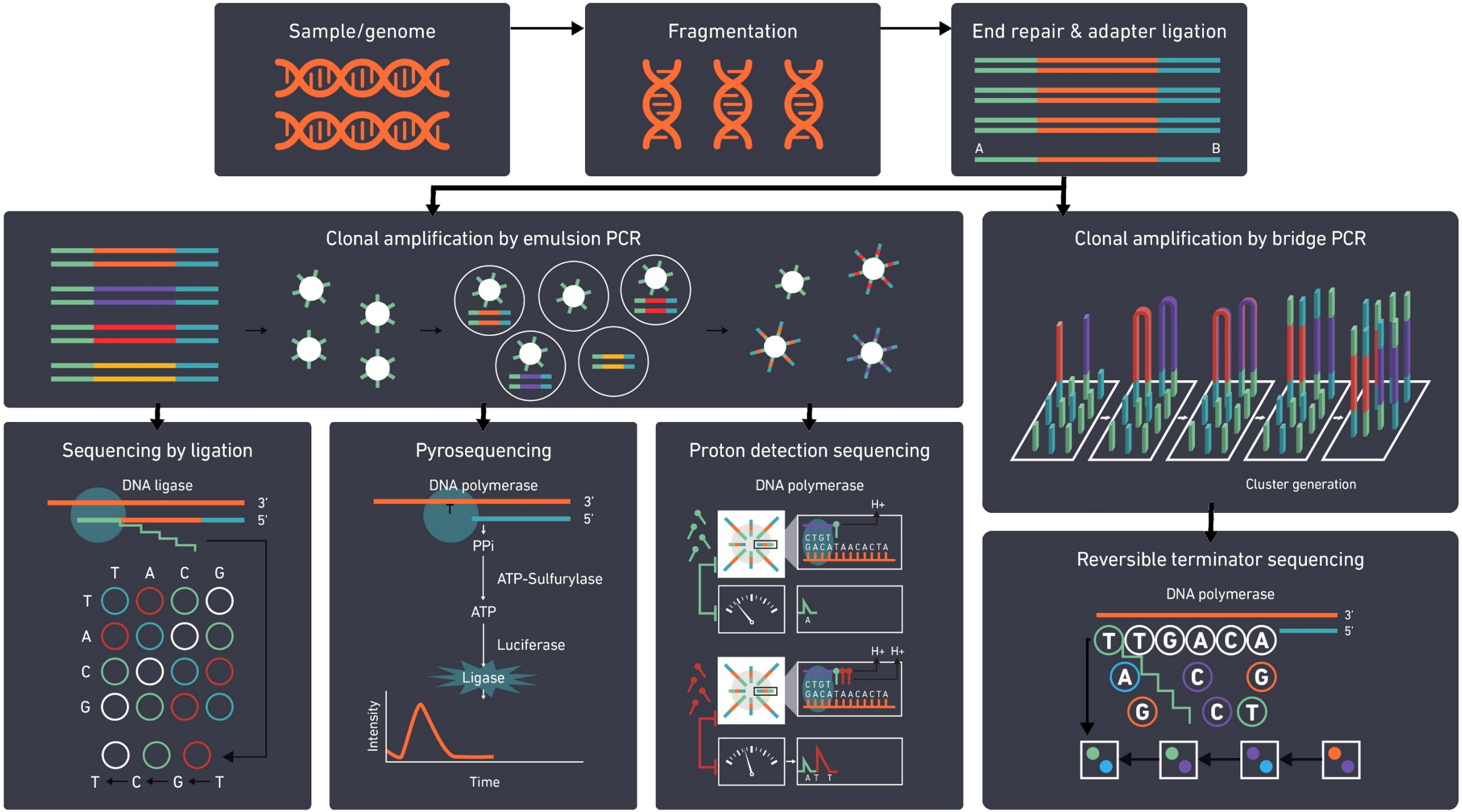
NGS

ADVANTAGES

- No *invivo* cloning, Transformation, Colony picking
- High degree of Parallelism then Capillary Sequencing
- Low Reagent Cost
- Reduced Sample Size
- Less Tme

DISADVANTAGES

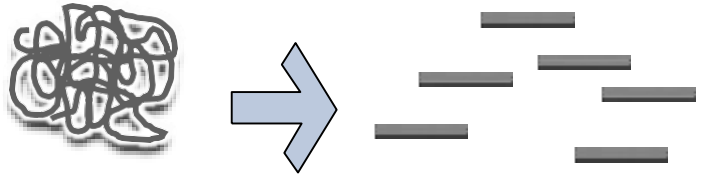
- Poor interpretation of homopolimers
- Incorporation of incorrect dNTPs by polymerases.
- Necessity for a deeper sequencing
- Need for PCR amplification prior to sequencing.


Sample/genome
Fragmentation
End repair & adapter ligation
A
B
Clonal amplification by emulsion PCR
Clonal amplification by bridge PCR
Cluster generation
Sequencing by ligation
DNA ligase
3'
5'
T A C G
T
A
C
G
T ← C ← G ← T
Pyrosequencing
DNA polymerase
T
3'
5'
PPi
ATP-Sulfurylase
ATP
Luciferase
Ligase
Intensity
Time
Proton detection sequencing
DNA polymerase
H+
CTGT
GACATAACACTA
A
H+ H+
CTGT
GACATAACACTA
A T T
Reversible terminator sequencing
DNA polymerase
3'
5'
T T G A C A
A C G
G C T

NGS WORKFLOW

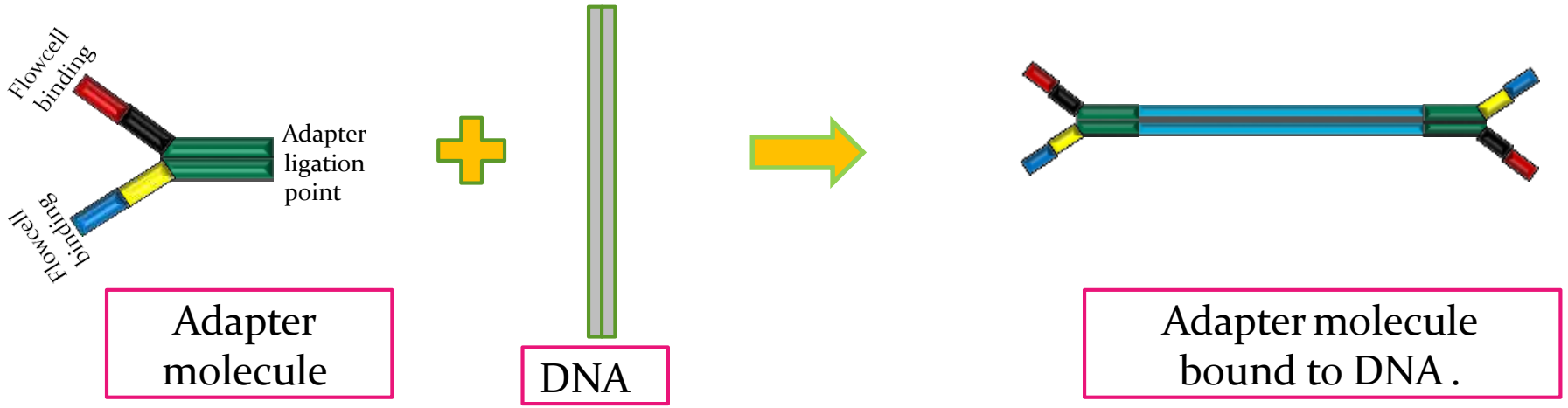
1. Sample extraction

2. Create DNA fragments

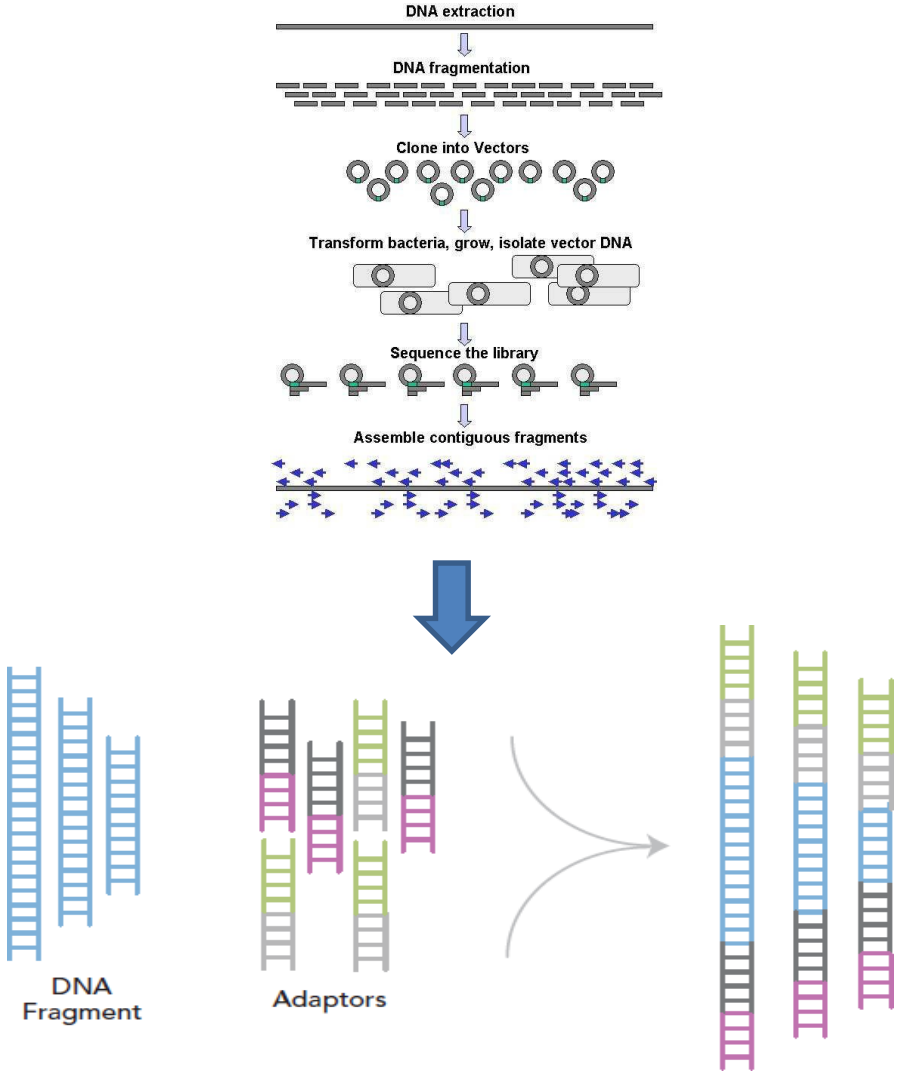


3. End repair & adapter ligation.

Add platform-specific adapter sequences to library. Adapter molecules bind every fragment to a flowcell or bead; add barcodes for multiplexing.

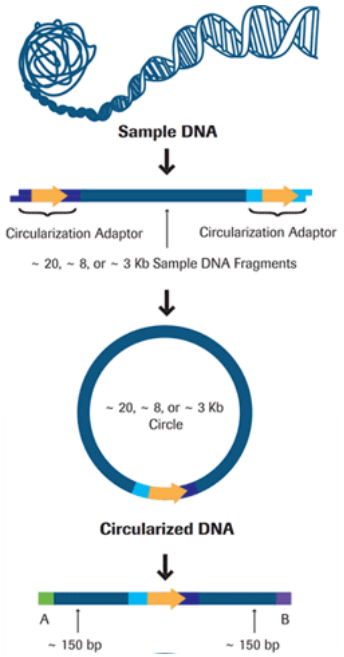


Next-gen sequencing: shotgun library preparation



Shotgun libraries

- Whole genome sequencing
  - Input: 100-1,000 ng of DNA
  - shear DNA (<1,000 bp)
  - *End repair*
  - *A-tailing*
  - *Ligation of sequencing adapters*

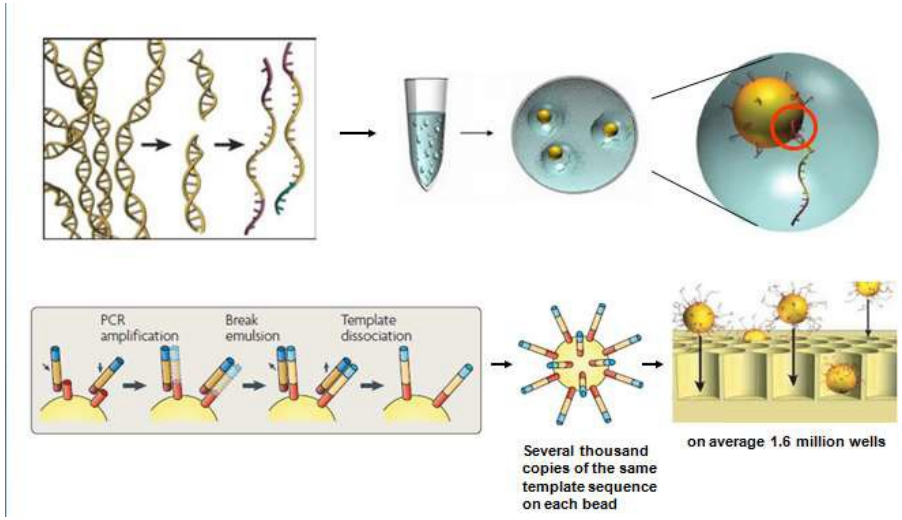


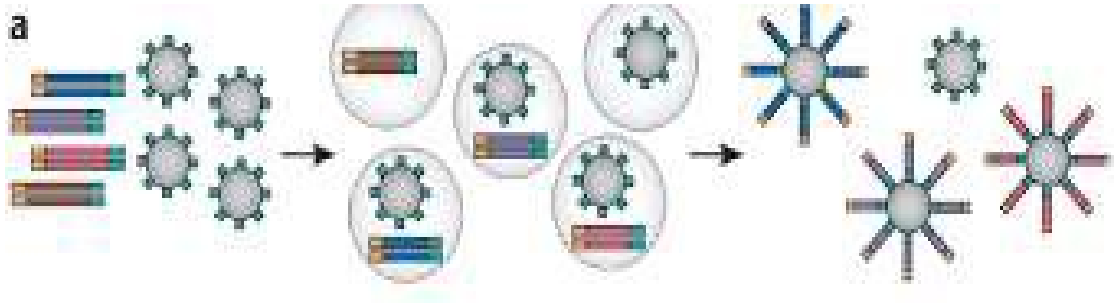
Mate pair libraries

- scafolding and structural variation
  - Input: 5-20 ug of DNA
  - Shear DNA to 3kb, 8kb and 20Kb fragments
  - Ligation of biotinylated circularization adapters
  - Shear circularized DNA
  - Isolate biotinylated mate pair junction
  - Ligate sequencing adapters

NGS WORKFLOW

4a. Clonal amplification by emulsion PCR



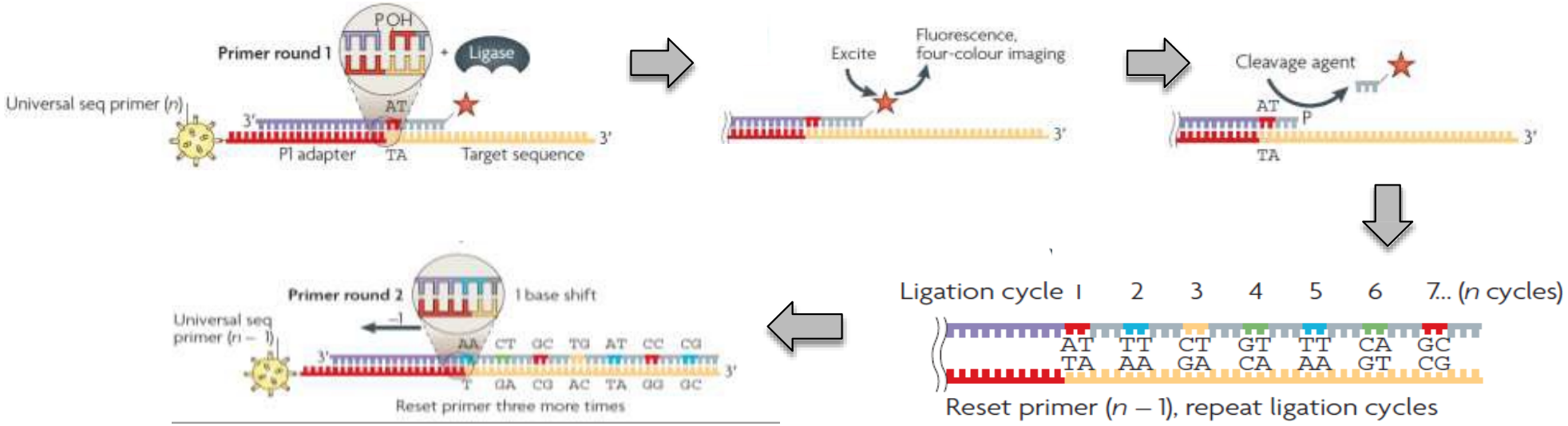


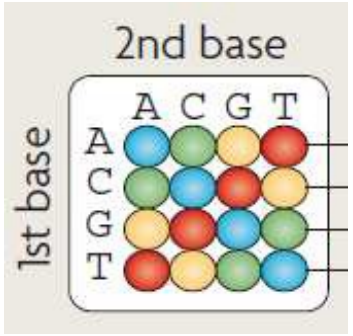
- Fragments with adaptors (the library) are PCR amplified within a water drop in oil.
- One PCR primer is attached to the surface of a **bead.**
- DNA molecules are synthesized on the beads in the water droplet. Each bead bears clonal DNA originated from a single DNA fragment
- Beads (with attached DNA) are then deposited into the wells of sequencing chips – one well, one bead.

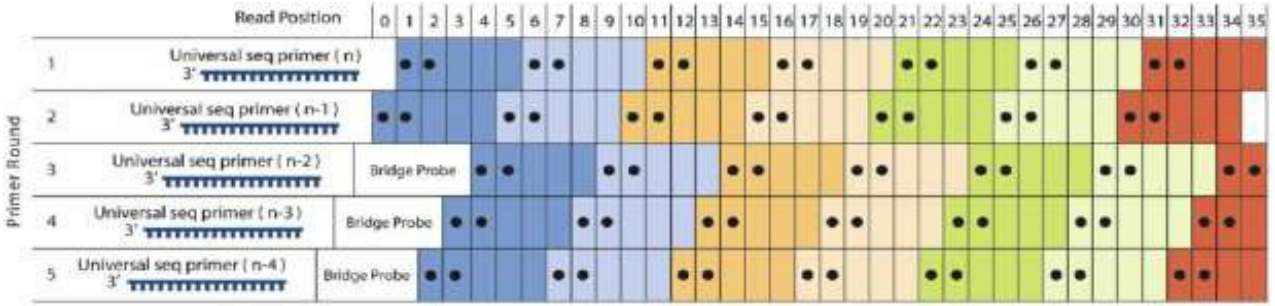
NGS WORKFLOW

5a. Sequencing by ligation

Sequencing by Ligation (SBL) uses the enzyme DNA ligase to identify the nucleotide present at a given position in a DNA sequence.



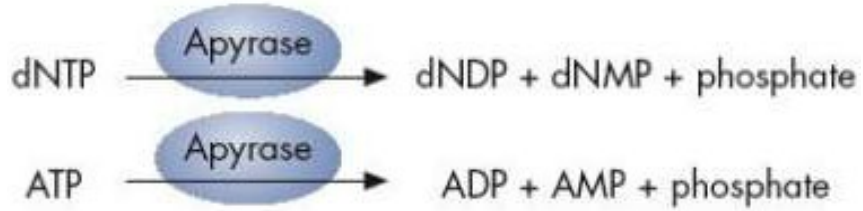


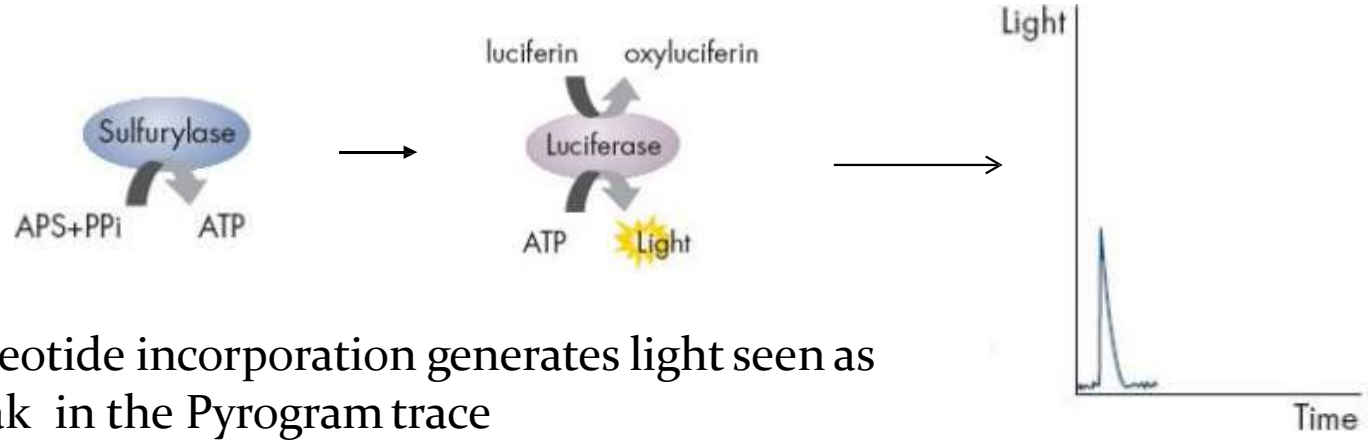


NGS WORKFLOW

5a. Pyrosequencing

Pyrosequencing: non-electrophoretic, bioluminescence method that measures the release of inorganic pyrophosphate by proportionally converting it into visible light using a series of enzymatic reaction

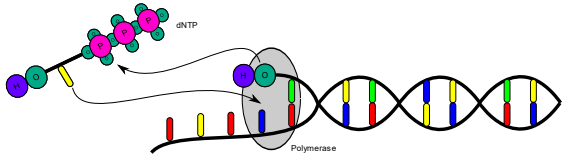




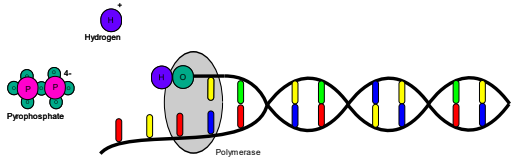
Nucleotide incorporation generates light seen as a peak in the Pyrogram trace

NGS WORKFLOW

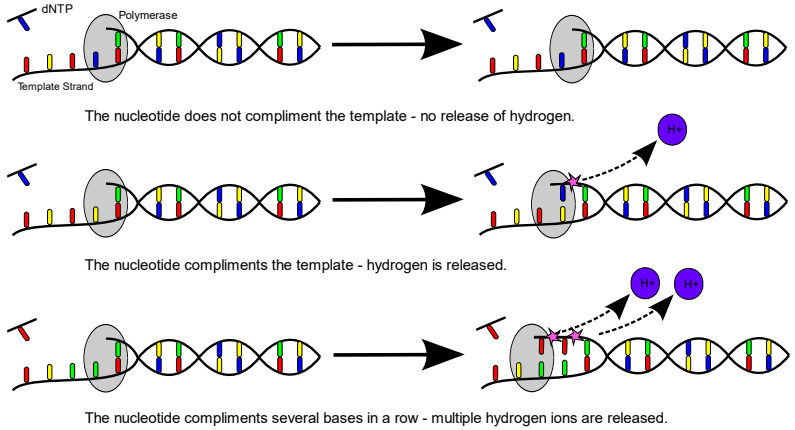
5a. Proton detection sequencing

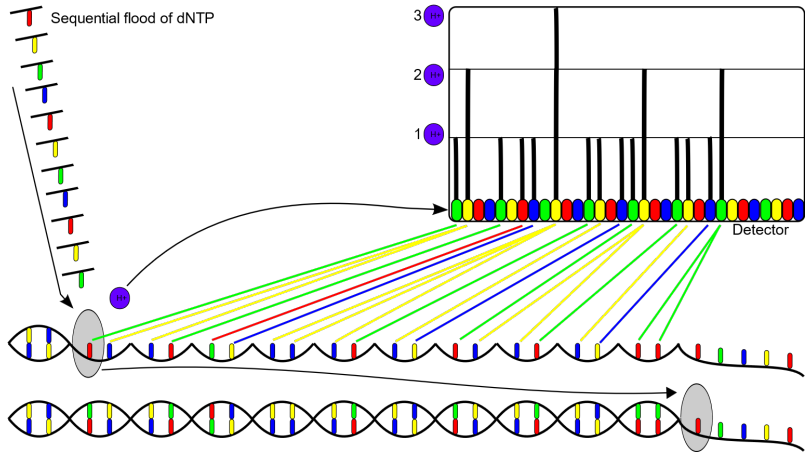


Polymerase integrates a nucleotide.



Hydrogen and pyrophosphate are released.





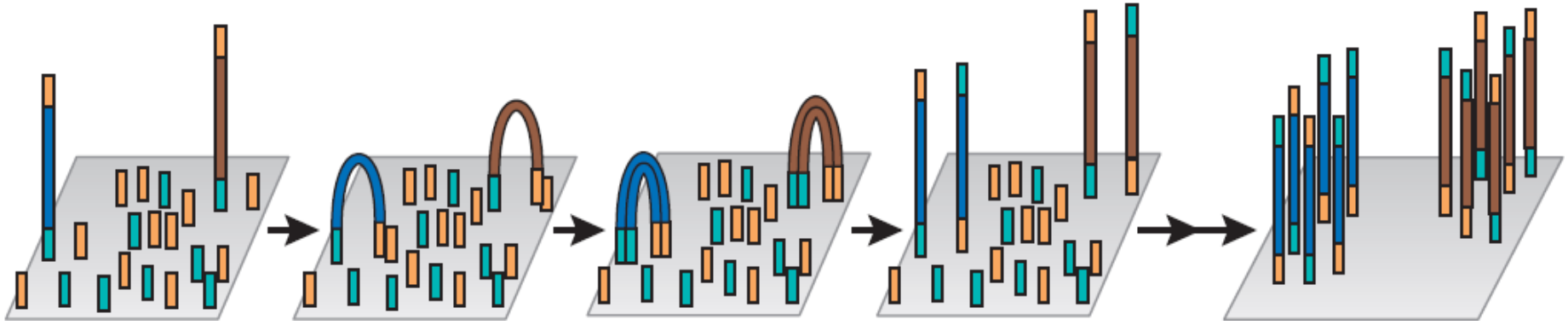
The incorporation of deoxyribonucleotide Triphosphate into a growing DNA strand causes the release of hydrogen and pyrophosphate.

The release of hydrogen ions indicate if zero, one or more nucleotides were incorporated.

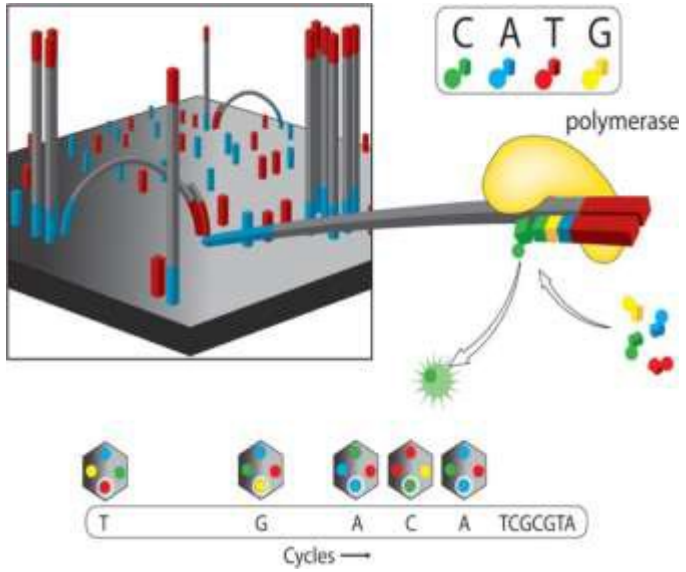
Released hydrogens ions are detected by an ion sensor. Multiple incorporations lead to a corresponding number of released hydrogens and intensity of signal.

NGS WORKFLOW

4b. Clonal amplification by bridge PCR



- DNA fragments are flanked with adaptors (**library**)
- A solid surface is coated with primers complementary to the two adaptor sequences
- Isothermal amplification, with one end of each “bridge” tethered to the surface
- **Clusters** of DNA molecules are generated on the chip. Each cluster is originated from a single DNA fragment, and is thus a clonal population.



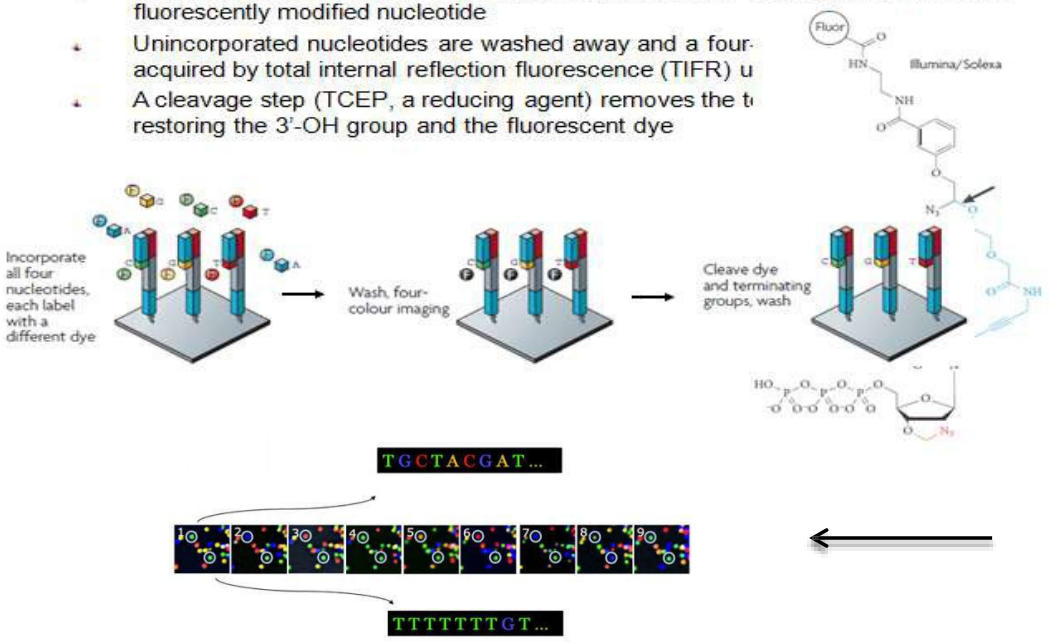
NGS WORKFLOW

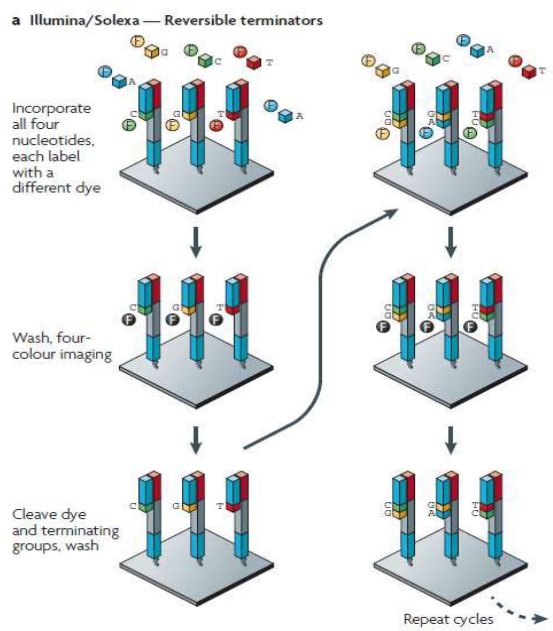
5b. Reversible terminator sequencing

By far the most popular SBS method is reversible terminator sequencing which utilizes ‘’bridge-amplification’’. During the synthesis reactions, the fragments bind to oligonucleotides on the flow cell, creating a bridge from one side of the sequence (P5 oligo on flow cell) to the other (P7), which is then amplified. The added fluorescently-labeled nucleotides are detected using direct imaging.

Sequencing by Cyclic Reversible Termination (CRT): CRT uses reversible terminators in a cyclic method that comprises nucleotide incorporation, fluorescence imaging

- a DNA polymerase, bound to the primed template, adds or incorporates just one fluorescently modified nucleotide
- Unincorporated nucleotides are washed away and a four- acquired by total internal reflection fluorescence (TIFR) u
- A cleavage step (TCEP, a reducing agent) removes the t restoring the 3’-OH group and the fluorescent dye





Third generation sequencing

The introduction of Third Generation Sequencing (TGS) circumvents the need for PCR, sequencing single molecules without prior amplification steps.

Sequence information is obtained with the use of DNA polymerase by monitoring the incorporation of fluorescently labeled nucleotides to DNA strands with single base resolution.

Genomics Proteomics Bioinformatics 13 (2015) 278–289

HOSTED BY

ELSEVIER

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

REVIEW

PacBio Sequencing and Its Applications

Anthony Rhoads [1,a], Kin Fai Au [1,2,*,b]

[1] *Department of Biostatistics, University of Iowa, Iowa City, IA 52242, USA*
[2] *Department of Internal Medicine, University of Iowa, Iowa City, IA 52242, USA*

Received 28 May 2015; revised 6 August 2015; accepted 11 August 2015
Available online 2 November 2015

Handled by Xuegong Zhang

Library construction

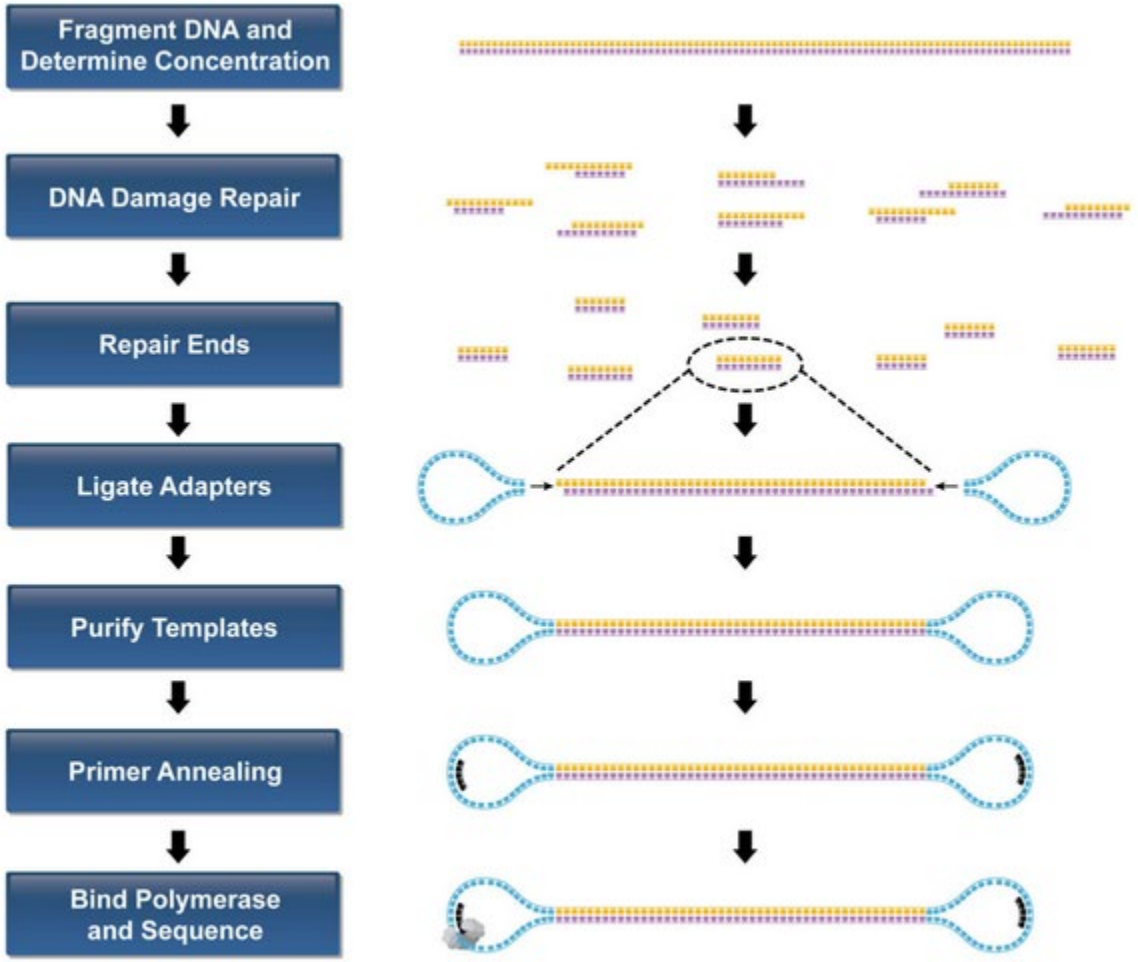
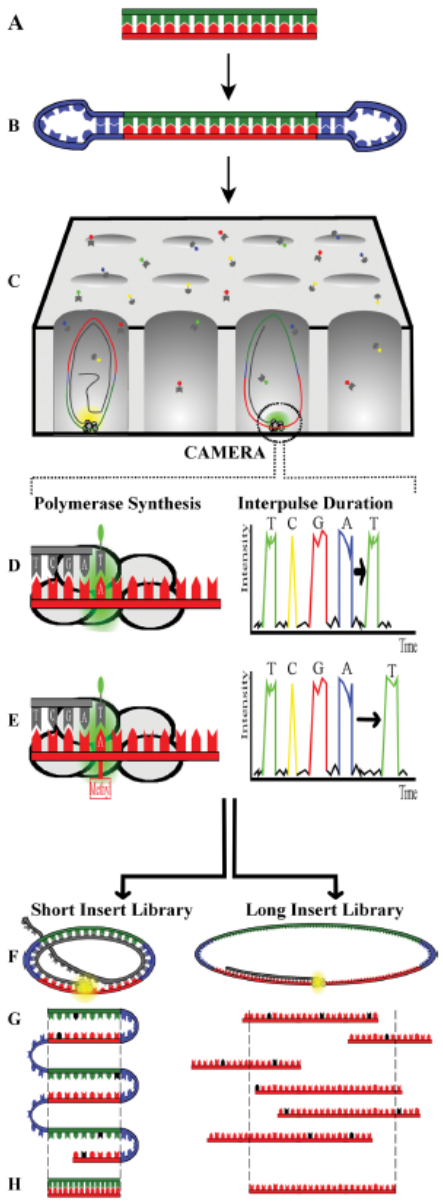


Figure 1. Template Preparation Workflow for PacBio RS II system.



Single-molecule real-time (SMRT) sequencing

Overview of SMRT Sequencing Technology. Sequencing starts with preparing a library from double stranded DNA (**A**) to which hairpin adapters are ligated (**B**). This library is thereafter loaded onto a SMRT Cell made up of nanoscale observation chambers (ZeroModeWaveguides (ZMWs)). The DNA molecules in the library will be pulled to the bottom of the ZMW where the polymerase will incorporate fluorescently labelled nucleotides (**C**). Note that not all ZMWs will contain a DNA molecule because the library is loaded by diffusion. The fluorescence emitted by the nucleotides is recorded by a camera in real-time. Hence, not only the fluorescence color can be registered, but also the time between nucleotide incorporation which is called the interpulse duration (IPD) (**D**, right panel). When a sequencing polymerase encounters nucleotides on the DNA strand containing an (epigenetic) modification, like for example a 6-methyl adenosine modification (E, left panel), then the IPD will be delayed (**E**, right panel) compared to non-methylated DNA (D, right panel). Due to the circular structure of the library, a short insert will be covered multiple times by the continuous long read (CLR). Each pass of the original DNA molecule is termed a subread, which can be combined into one highly accurate consensus sequence termed a circular consensus sequence (CCS) or reads-ofinsert (ROI) (**F**–**H**, left panel). Though SMRT sequencing always uses a circular template, long insert libraries typically only have a single pass and hence generate a linear sequence with single pass error rates (black nucleotides) (FG, right panel). Afterwards, overlapping single passes can be combined into one consensus sequence of high quality (H, right panel). Overall, CCS reads have the advantage of being very accurate while single passes stand out for their long read lengths (>20 kb).

Fourth generation sequencing

Genomics Proteomics Bioinformatics 13 (2015) 4–16

HOSTED BY

ELSEVIER

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

REVIEW

Nanopore-based Fourth-generation DNA Sequencing Technology

Yanxiao Feng [1,4,a], Yuechuan Zhang [1,2,b], Cuifeng Ying [1,3,c], Deqiang Wang [1,4,*,d], Chunlei Du [1,4,e]

Journal of Human Genetics (2020) 65:1
https://doi.org/10.1038/s10038-019-0683-4

EDITORIAL

Advent of a new sequencing era: long-read and on-site sequencing

Yutaka Suzuki[1]

Received: 30 September 2019 / Accepted: 2 October 2019 / Published online: 6 January 2020

ISTITUTO SUPERIORE DI SANITÀ
CENTRO NAZIONALE
CONTROLLO
E VALUTAZIONE DEI FARMACI

Nanopore technology requires no amplification and uses the concept of single molecule sequencing but with the integration of tiny biopores of nanoscale diameter (nanopores) through which the single molecule passes and is identified.

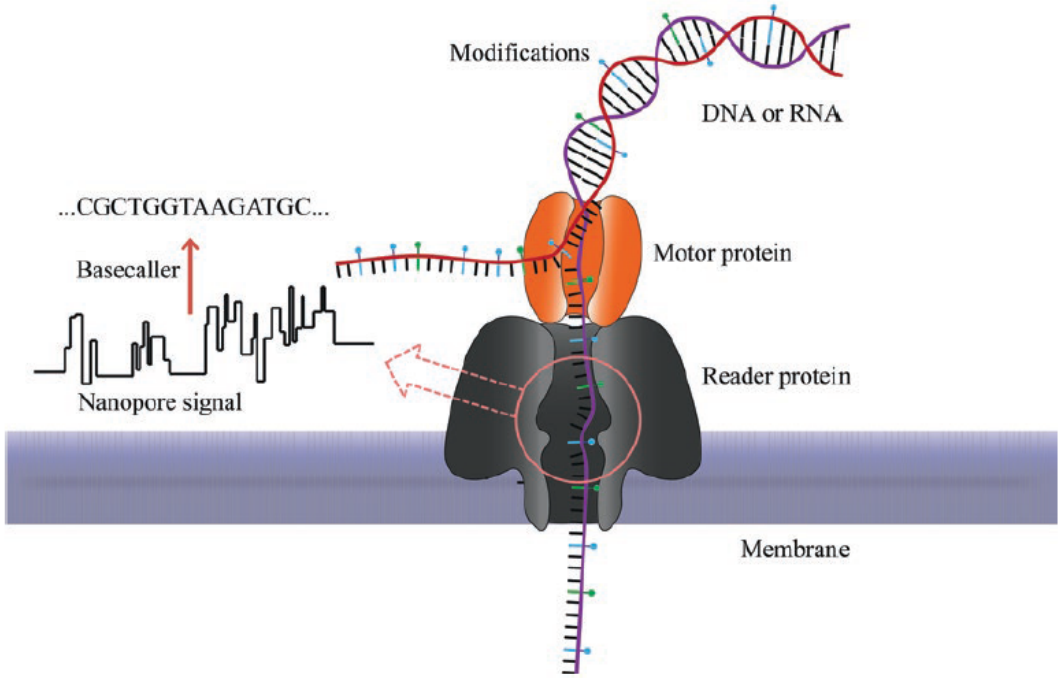
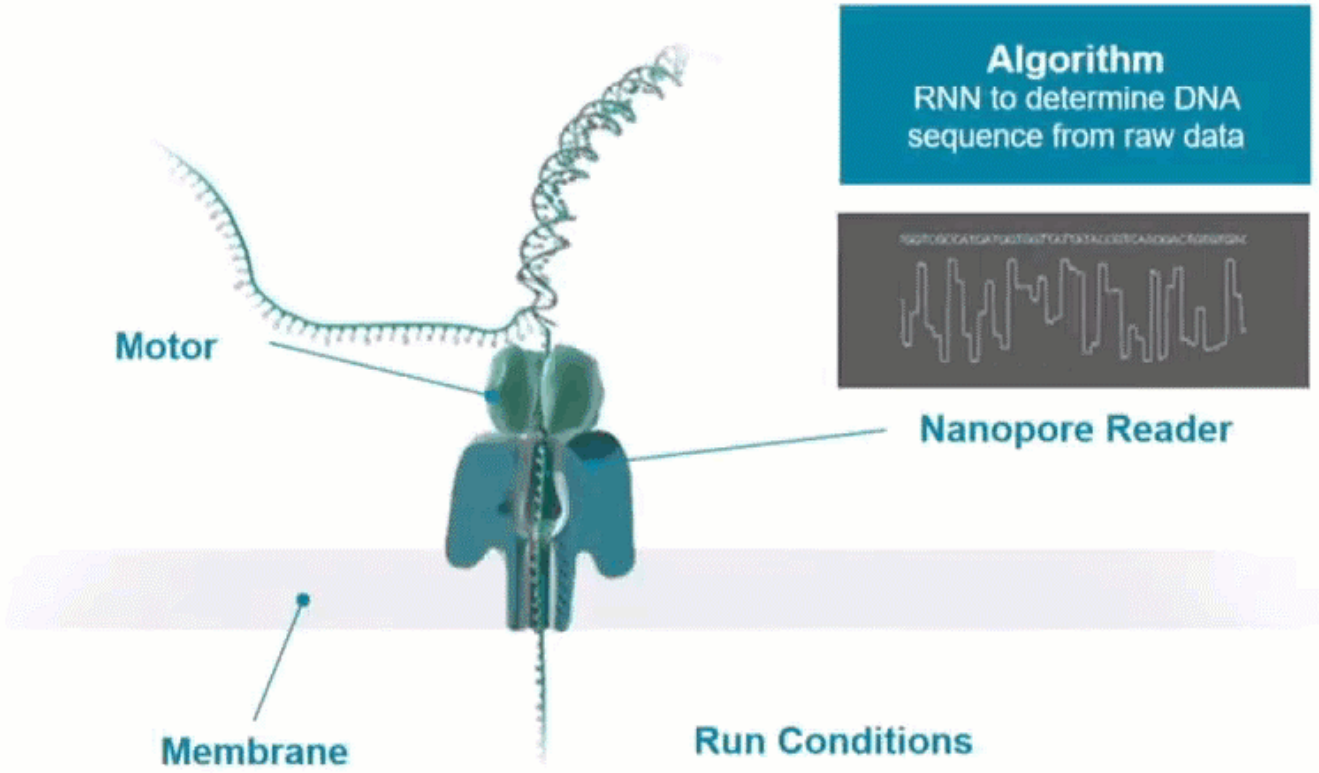


Fig. 1 Scheme for nanopore sequencing. The motor protein guides the DNA or RNA strand through the nanoscale pore provided by the reader protein. Passage of the nucleic acid molecules through a nanopore causes fluctuations of the current across the membrane. The basecaller converts the nanopore signal into the corresponding nucleic acid sequence

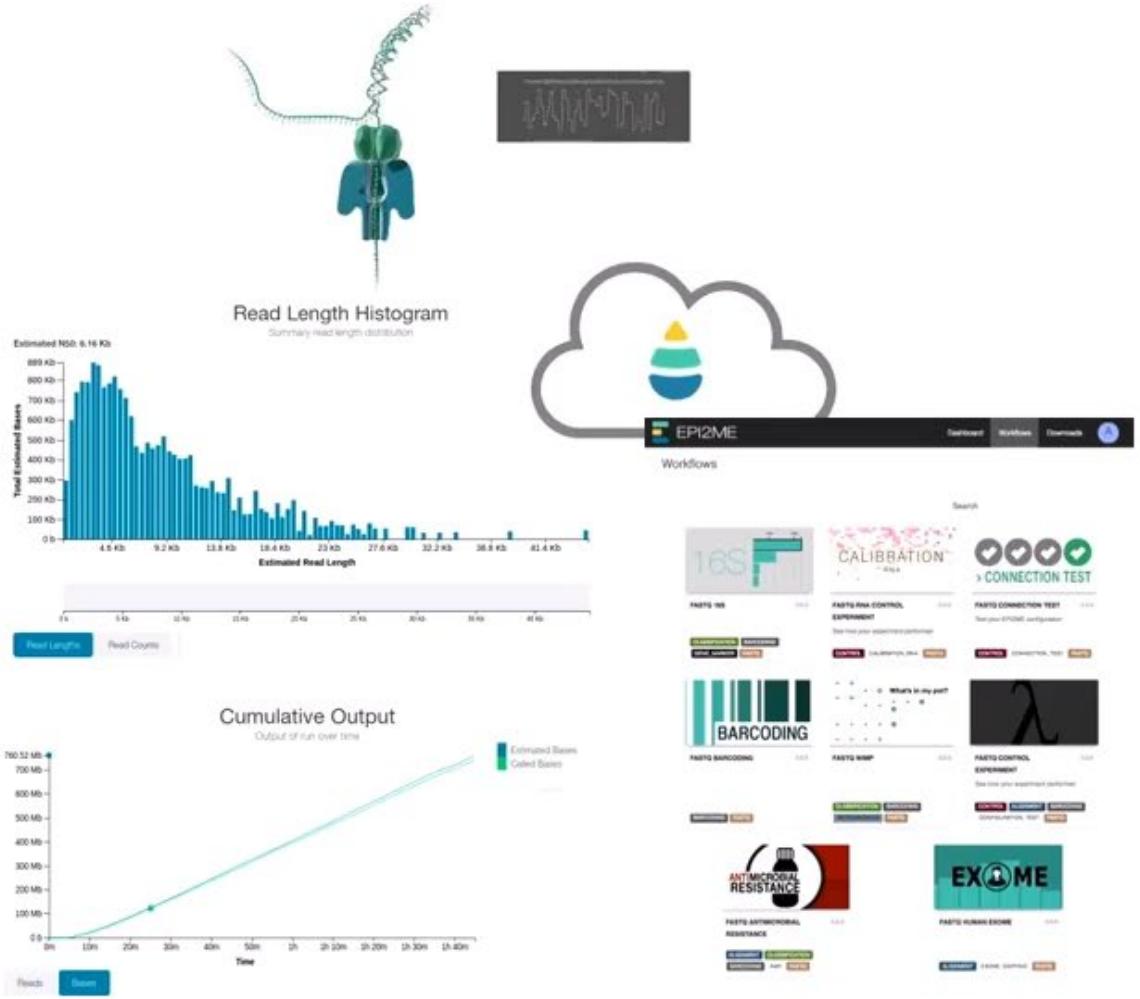
Nanopore sequencing



Oxford NANOPORE Technologies

ISTITUTO SUPERIORE DI SANITÀ
CENTRO NAZIONALE
CONTROLLO
E VALUTAZIONE DEI FARMACI

A new paradigm in sequencing



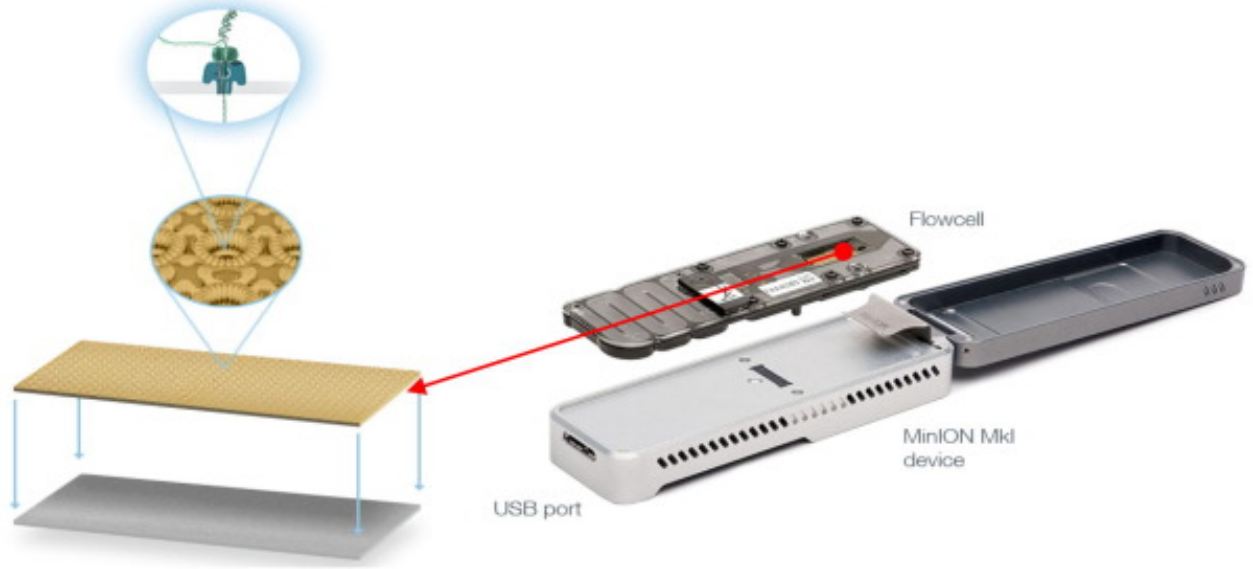
Real-time monitoring through MinKNOW

- No fixed run time
- Real-time basecalling
- Real-time data assessment
- On-demand sequencing
- Pause a sequencing run
- Efficient data acquisition

Oxford NANOPORE Technologies

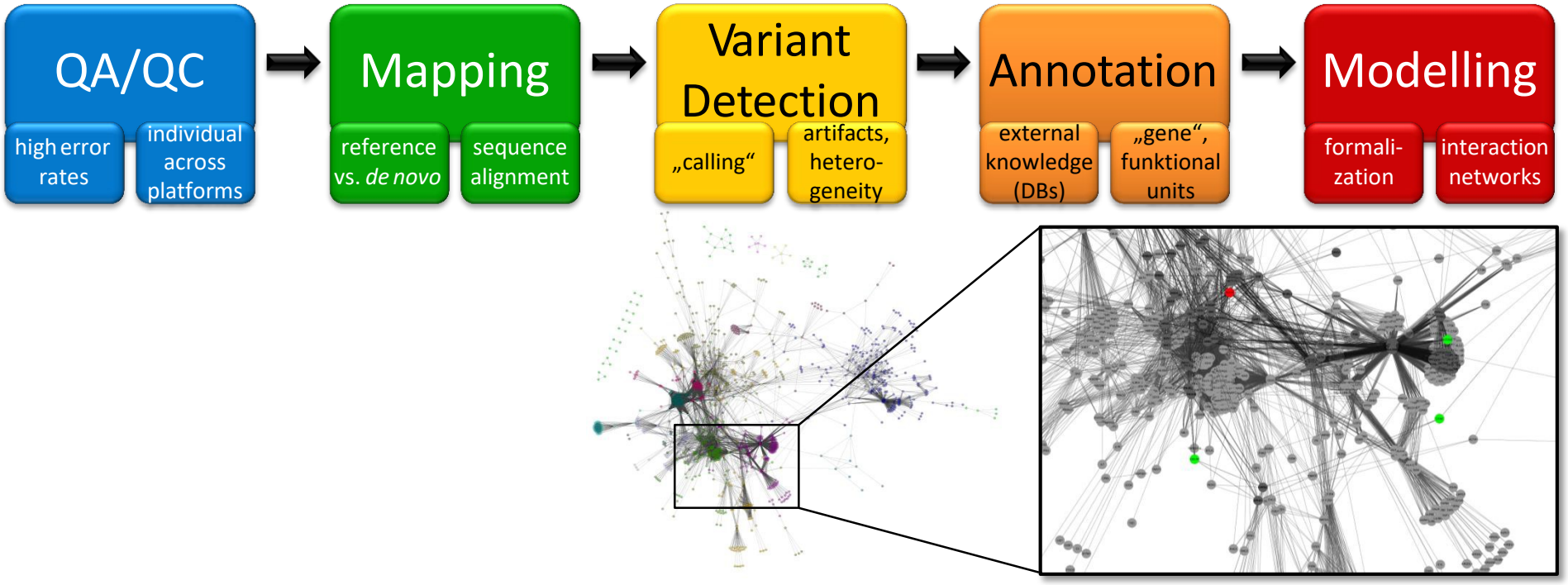
ISTITUTO SUPERIORE DI SANITÀ
CENTRO NAZIONALE CONTROLLO E VALUTAZIONE DEI FARMACI


PrometheION48
GridION Mk1


Flowcell
MinION Mkl
device
USB port


MacBook Pro

- NGS is just a technology generating data
- Scientists need assays in order to get from questions to answers
- Great variety of problems, scientific fields, target molecules, biological mechanism etc. determine the assay and the data analysis
- General scheme:



.fastq files

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

Phred quality score

$$Q = -10 \log_{10} P$$

```
@X1L6C:01561:00672
AAATATCACCAAATAAAAAACGCCTTAGTAAGTATTTTTCAGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTG
GATTAAAAAAAGAGTGTCTGATAGCAGCTTCTGAACTGGTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCAC
TAAATACTTTAACCAATATAGGCATAGCGCACAGACAGATAAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCA
CCACCATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAAGCCCGCA
CCTGACAGTGCGGGCTTTTTTCGACCAAAGGTAACGAGGTAACAACCATGCG
+
CC:9::FBC<CD7:88888(:>><C<CCCC<CCBBAAB/A@A8888,;<@;AABBB=?;B98992:B<
CGBBCGDCC??BCC;BB<ADEEED*CCCAAACCCBCABBDDBB>B??A;999;@8=>199A7>9::CBCH:B:>>>)999)
77037;<7==5=@@BBCC:C@BBB9B<E<D9>?><<6ADCBCBAABB@@@DDCCBA@@==+=.//?B<??AEB::6;DCD>
C:;;;-:9:BC<BBCCC9??<AA;AG<CB>GD@B;;;A<AE;AA<B??@9@C<BB<???BB;BBBAAAA:::BAB099/9>
@========(<<?)99997>>CCEBA>>=>2373333&3:99-33(3--717---43606704/47761
@X1L6C:01104:03031
AGAAGCTGCTATCAGACACTCTTTTTTTAATCCACACAGAGACATATTGCCCGTTGCAGTCAGAATGAAAAGCTGAAAAATA
CTTACTAAGGCGTTTTTTATTTGGTGATATTTTTTTCAATATCATGCAGCAAACGGTGCAACATTGCCGTGTCTCGTTGCTC
TAAAAGCCCCAGGCG
+
@AC=BCCC???B?@@CBB@???>>>>>*?8??>DAABEBCBABCAAA:@@>+9:8>;<;//.
98283988*44449;;9/88:?29:>>5;78333333&399298:6/./DCDDCC';>:ACBDAABB??9::+9<
1444@:?77-3<03368:8755888;:9833)3777'--'--
@X1L6C:03659:02717
GCTTCTGAACTGGTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAATATAGGCATA
GCGCACAGACAGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTA
CCACAGGTAACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAGACCCGCCACTGACCAGTGCG
+
???9?BB@<CAA;A8@?:?@@5::BCCCEC;C=CCC8CEJ8DE;AACF>CC?DDCCCBB:B@???9?;B=B=CAA@?;?BCG
CCCCCCBABBBBCCDDAA2:4;@???CAB@AAA9@@AB?C:;;C;CDCCC>ECCAA<AC<CB>DC<AB=CD=C9::A4::>
CC;@@@A?CI@DDAFKDDD:A@CBCDC::::99199+8;4746@CA?)<444/3:4934333-3888//
@X1L6C:02011:02071
TTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAATATAGGCATAGCGCACAGACAGATAAAAATTACAGAGTACA
CAACATCCATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGTGACGCGTACAG
GAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTCGACCAAAGTAACG
+
=0>>>19;;,;;7=CCDADC;?::::::,5;;==4>273:<@BBCF=CDH;@;MMFEED@?>>>:::::*5/55<
;::@:;:BC=BCBB<B@@@D<@@B:;3:::9@<BB=BD=AC;@B:???::CAC=CD;;;=BBAB>CC;AA;BAAAA9AD@>>
>>?955>4?94999855555&4<>2:;661499888...88/56666666$;6/.5:8(..+'++
@X1L6C:01333:03005
GCAATGCCAGGCAGGGCATGTACAGCTACGTACGTCTGAGCATCGATCGATGTACAGCTACGTACGTCTGAGCATCGATCGA
TGTACAGCTACGTACGTCTGAGCATCGATCGATGTACAGCTACG
+
555/55/(//(////(/8:9:<=>?<?@:98A??676<:;;@:5555555554444;=4443333;383338<68>>
68=333111831111111111933644588?==<76992---2+++0/
```

Coverage

Reads mapped on a reference genome

Ref seq

COVERAGE

Mapped reads

NGS and control of biologicals

NGS allows the sequencing of billions of molecules at the same time, a capability called highly parallel sequencing.

Drug manufacturing can also benefit from advances in sequencing.

NGS makes it possible to carry out cell line identification.

One NGS application is environmental monitoring of microbial contaminants in biopharma manufacturing

NGS and control of biologicals

NGS can support biopharmaceutical drug development for the design of clinical trials.

Control of adenovirus-associated vectors or other viral vectors

The pandemic has increased the biopharmaceutical industry’s use of NGS.

❑ Outline (2)

- Viral safety
- Applications of NGS in biological products
- Validation requirements for NGS
- Conclusion and Future Perspective

- Safety of vaccines and other biological products is critical.

- Safety is particularly critical for:
  - live vaccines
  - gene therapy viral vectors
  - cell therapy medicinal products

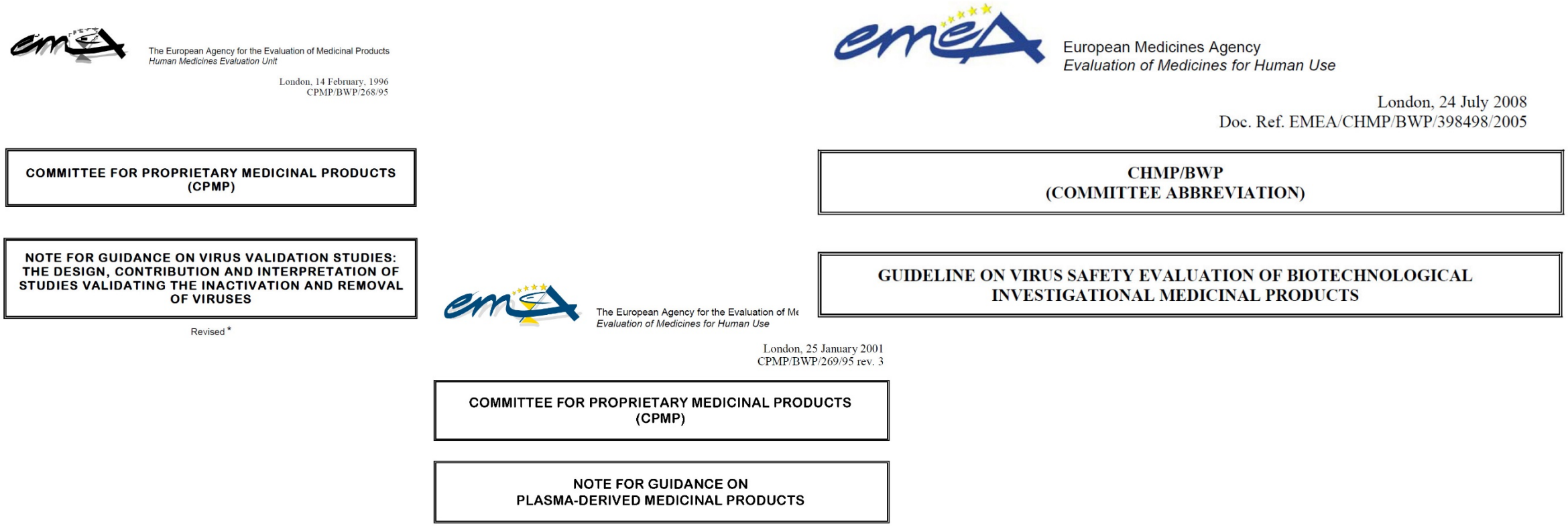
ICH HARMONISED TRIPARTITE GUIDELINE

VIRAL SAFETY EVALUATION OF BIOTECHNOLOGY PRODUCTS DERIVED FROM CELL LINES OF HUMAN OR ANIMAL ORIGIN

Q5A(R1)

- ICH Q5A, specifically requires that a manufacturer of biological products for human use demonstrate the capability of the manufacturing process to remove or inactivate known contaminants.

- Various EMA guidelines provide recommendations for validation of viral inactivation biopharmaceutical products.



- These recommendations also set specific values for virus clearance levels that had to be attained.

The safety assurance is achieved through the application of a robust and effective virus testing program, which adopts a three-tiered approach:

1. Selecting and testing of cell lines and other raw materials

2. Assessing capacity for viral clearance and inactivation by a manufacturing process

3. In-process and/or final product virus testing.

❑ Detection of viral contaminants in biopharmaceutical products

Non-specific test methods

Cell cultures

Infectivity Assay
(Retroviruses)

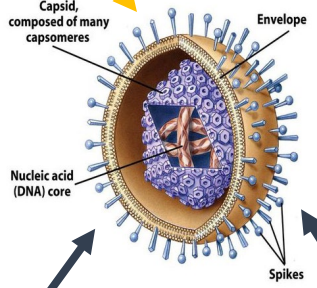
Tests in animals

Live Infectious virus

9CFR for Bovine & Porcine viruses

Test for reverse transcriptase

Transmission Electron Microscopy (TEM)



Viral components

PCRs

in vivo Antibody production tests (MAP, RAP, HAP)

Virus-specific test methods

in-vitro methods

in-vivo methods

Animal testing

one of the key differences between these testing methods is timing

❑ Limitation of conventional assays

- Cell lines may not be permissive for the virus
- Virus replication is not visible (no CPE)
- Cytotoxicity – Neutralization - Interference
- Virus not detected by PCR primers

- Regulatory expectations on viral safety of biopharmaceutical products have evolved over the past decade.
- Today, the concerns are much broader, encompassing unknown and uncharacterized agents.
- Increasingly stringent conditions are intended to decrease the risk of transmitting viruses.
- Next generation sequencing (NGS) is a sensitive and un-biased detection method for adventitious agents.

Evolution of European Pharmacopoeia

- Ph. Eur. Chapter 5.2.14: "Substitution of *in vivo* method(s) by *in vitro* method(s) for the quality control of vaccines", implemented 1/2018, version 9.3

5.2.14. SUBSTITUTION OF *IN VIVO* METHOD(S) BY *IN VITRO* METHOD(S) FOR THE QUALITY CONTROL OF VACCINES

PURPOSE

The purpose of this general chapter is to provide guidance to facilitate the implementation of *in vitro* methods as substitutes for existing *in vivo* methods, in cases where a typical one-to-one assay comparison is not appropriate for reasons unrelated to the suitability of one or more *in vitro* methods. This general chapter will not discuss the details of assay validation as such, since those principles are described elsewhere.

The general chapter applies primarily to vaccines for human or veterinary use, however the principles described may also apply to other biologicals such as sera.

Detection of viral extraneous agents by novel molecular methods

Detection of viral extraneous agents in cell banks, seed lots and cell culture harvests is currently conducted using a panel of *in vivo* and *in vitro* methods at different stages of the manufacturing process. Novel, sensitive molecular techniques with broad detection capabilities are available, including deep sequencing or high-throughput sequencing methods, degenerate polymerase chain reaction (PCR) for whole virus families or random-priming methods (associated or not with sequencing), hybridisation to oligonucleotide arrays and mass spectrometry. The use of these new molecular methods has highlighted gaps in the existing testing strategy by identifying previously undetected viral contaminants in final product, the cell banks from which it was produced and intermediate manufacturing stages. These new molecular methods (e.g.

Evolution of European Pharmacopoeia

- Ph. Eur. Chapter 5.2.5: “Cell Substrate for the production of vaccines for human use”, implemented 1/2018, version 9.3

5.2.3. CELL SUBSTRATES FOR THE PRODUCTION OF VACCINES FOR HUMAN USE

This general chapter deals with diploid cell lines and continuous cell lines used as cell substrates for the production of vaccines for human use; additional issues specifically related to vaccines prepared by recombinant DNA technology are covered by the monograph *Products of recombinant DNA technology (0784)*. The testing to be carried out at the various stages (cell seed, master cell bank (MCB), working cell bank (WCB), end of productions cells (EOPC) or extended cell bank (ECB) corresponding to cells at or beyond the maximum population doubling level used for production) is indicated in Table 5.2.3.-1. General provisions for the use of cell lines and test methods are given below. Where primary cells or cells that have undergone a few passages without constitution of a cell bank are used for vaccine production, requirements are given in the individual monograph for the vaccine concerned.

Tests for specific viruses. The list of specific viruses to be tested is defined based on a viral contamination risk assessment in accordance with the principles detailed in general chapter *5.1.7. Viral Safety*, and takes into account (but is not limited to) the origin of the cells and the potential sources of viral contamination (e.g. raw material of animal or plant origin). NAT tests (*2.6.21*) are carried out with or without prior amplification in cells. For cell lines of rodent origin, NAT (*2.6.21*) or antibody production tests in mice, rats or hamsters are used to detect species-specific viruses.

Tests for viruses using broad molecular methods. In agreement with the competent authority, broad molecular methods (e.g. High Throughput Sequencing) may be used either as an alternative to *in vivo* tests and specific NAT or as a supplement or alternative to *in vitro* culture tests based on the risk assessment.

For both NAT (*2.6.21*) and broad molecular methods, the stage at which testing is to be conducted (e.g. MCB, WCB, EOPC/ECB) is also based on the risk assessment and depends on the steps where viral contaminants may be introduced. In case of positive results with either broad molecular methods or NAT tests, a follow-up investigation must be conducted to determine whether detected nucleic acids are due to the presence of infectious extraneous agents and/or are known to constitute a risk to human health.

Evolution of European Pharmacopoeia

- Ph. Eur. Chapter 2.6.16: "Tests for extraneous agents in viral vaccines for human use", version 10.2

2.6.16. TESTS FOR EXTRANEOUS AGENTS IN VIRAL VACCINES FOR HUMAN USE

INTRODUCTION

A strategy for testing extraneous agents in viral vaccines must be developed based on a risk assessment following the principles of viral contamination risk detailed in general chapter *5.1.7. Viral safety*. This strategy includes a full package of suitable tests that are able to detect different families of extraneous agents that may infect the source of virus strains including cell substrates and raw material of animal or plant origin. It also takes into account the capacity of the manufacturing process to remove or inactivate viruses. The

New, sensitive molecular methods with broad detection capabilities are available. These new approaches include high-throughput sequencing (HTS) methods, nucleic acid amplification techniques (NAT) (e.g. polymerase chain reaction (PCR), reverse transcriptase PCR (RT-PCR), product-enhanced reverse transcriptase (PERT) assays) for whole virus families or random-priming methods (associated or not with sequencing), hybridisation to oligonucleotide arrays, and mass spectrometry with broad-spectrum PCR. These methods may be used either as an alternative to *in vivo* tests and specific NAT or as a supplement/alternative to *in vitro* culture tests based on the risk assessment and with the agreement of the competent authority.

WHO Focus on NGS

WHO - TRS 978, ECBS 2010:

"New, sensitive, molecular methods, with broad detection capabilities are being developed... The new generation of massively parallel (deep) sequencing (MPS) methods may have particular utility. They can be applied to detect virions after nuclease treatment to remove cellular DNA and unencapsidated genomes. Used in this mode, MPS has been used to discover new viruses in serum and other tissues and has revealed the contamination of human vaccines by porcine circovirus.“

"MPS can also be employed to screen cell substrates for both latent and lytic viruses by sequencing the transcriptome. In this mode, enormous quantities of data are generated, and robust bioinformatic methods are required to detect viral sequences by either positive selection against viral databases or negative selection to remove cellular sequences.”

WHO - TRS 878, Annex 1

"It is probable that application of methods of this type will be expected or required by regulatory agencies in future."

WHO TRS 993 Annex 2. Scientific principles for regulatory risk evaluation on finding an adventitious agent in a marketed vaccine

"WHO defined ***Next-generation sequencing (NGS) as "****high-throughput sequencing technology that processes sequences in parallel, producing thousands or millions of sequences at once from a sample… Significant bioinformatics using curated (trusted) databases are needed to analyze the considerable amount of data generated in each sequencing run."*

"New methods and technologies, such as NGS or microarrays, are powerful tools for the detection and identification of sequences from viruses and other adventitious agents without prior knowledge of the nature of the agent. In the future such new technologies may uncover the presence of other, as yet unrecognized, adventitious agents."

❑ Application and Usefulness of NGS -1

- Detection method for adventitious agents.

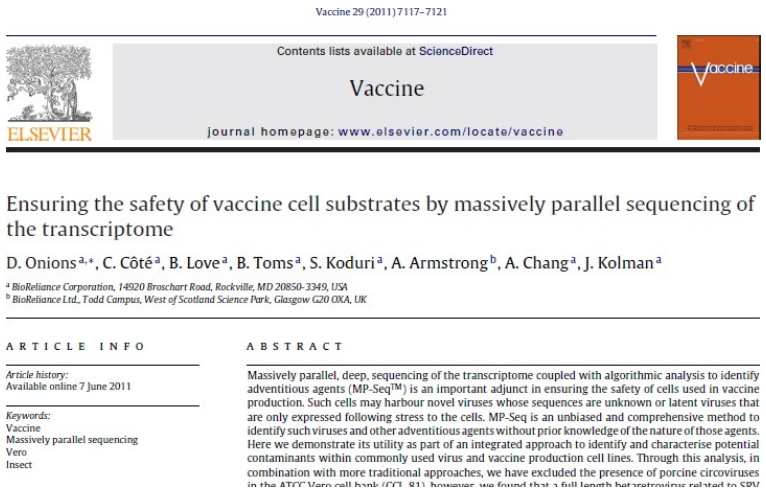
Removal, supplementation, replacement, substitution of in vivo adventitious agent tests

Substitution of in vitro nucleic acid based tests

Substitution of animal testing method

❑ Application and Usefulness of NGS -2

➢ Characterization, screening studies



Vaccine 29 (2011) 7117–7121

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Ensuring the safety of vaccine cell substrates by massively parallel sequencing of the transcriptome

D. Onions[a,*], C. Côté[a], B. Love[a], B. Toms[a], S. Koduri[a], A. Armstrong[b], A. Chang[a], J. Kolman[a]

[a] BioReliance Corporation, 14920 Broschart Road, Rockville, MD 20850-3349, USA
[b] BioReliance Ltd., Todd Campus, West of Scotland Science Park, Glasgow G20 0XA, UK

ARTICLE INFO

Article history:
Available online 7 June 2011

Keywords:
Vaccine
Massively parallel sequencing
Vero
Insect

ABSTRACT

Massively parallel, deep, sequencing of the transcriptome coupled with algorithmic analysis to identify adventitious agents (MP-Seq™) is an important adjunct in ensuring the safety of cells used in vaccine production. Such cells may harbour novel viruses whose sequences are unknown or latent viruses that are only expressed following stress to the cells. MP-Seq is an unbiased and comprehensive method to identify such viruses and other adventitious agents without prior knowledge of the nature of those agents. Here we demonstrate its utility as part of an integrated approach to identify and characterise potential contaminants within commonly used virus and vaccine production cell lines. Through this analysis, in combination with more traditional approaches, we have excluded the presence of porcine circoviruses in the ATCC Vero cell bank (CCL-81), however, we found that a full length betaretrovirus related to SRV

Pre Master Cell Bank / Pre Seed

Raw Materials

Biologicals 42 (2014) 218–219

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Short paper

Analysis by high throughput sequencing of Specific Pathogen Free eggs

Léa Gagnieur[a], Justine Cheval[b], Marielle Cochet[c], Emmanuel Breard[c], Marlène Gratigny[b], Charles Hébert[b], Erika Muth[b], Cyril Viarouge[c], Marine Dumarest[a], Muriel Coulpier[c], Marc Eloit[a,b,c,*]

[a] Institut Pasteur, Laboratory of Pathogen Discovery, Department of Virology, 28 rue du Docteur Roux, F-75724 Paris, France
[b] PathoQuest, Bâtiment François Jacob, 25 rue du Dr Roux, 75015, Paris, France
[c] Ecole Nationale Vétérinaire d'Alfort, UMR 1161 Virologie ENVA, INRA, ANSES, 7 Avenue Général de Gaulle, F-94704 Maisons Alfort, France

Biologicals 42 (2014) 145–152

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Unbiased analysis by high throughput sequencing of the viral diversity in fetal bovine serum and trypsin used in cell culture

Léa Gagnieur[a], Justine Cheval[b], Marlène Gratigny[b], Charles Hébert[b], Erika Muth[b], Marine Dumarest[a], Marc Eloit[a,b,c,*]

[a] Institut Pasteur, Laboratory of Pathogen Discovery, Department of Virology, 28 rue du Docteur Roux, F-75724 Paris, France
[b] PathoQuest, Bâtiment François Jacob, 25 rue du Dr Roux, 75015 Paris, France
[c] Ecole Nationale Vétérinaire d'Alfort, UMR 1161 Virologie ENVA, INRA, ANSES, 7 avenue Général de Gaulle, F-94704 Maisons Alfort, France

Biologicals 61 (2019) 1–7

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Characterization of the viral genomes present in commercial batches of horse serum obtained by high-throughput sequencing

W.P. Paim[a], M.N. Weber[a], S.P. Cibulski[b], M.S. da Silva[a], D.E. Puhl[a], R.F. Budaszewski[a], A.P.M. Varela[c], F.Q. Mayer[d], C.W. Canal[a,*]

[a] Laboratório de Virologia, Faculdade de Veterinária, Universidade Federal Do Rio Grande Do Sul (UFRGS), Porto Alegre, Rio Grande do Sul (RS), Brazil
[b] Laboratório de Biotecnologia Cellular e Molecular, Centro de Biotecnologia-CBiotec, Universidade Federal da Paraíba (UFPB), Cidade Universitária, João Pessoa, PB, Brazil
[c] Equipe de Virologia, Instituto de Ciências Básicas da Saúde, UFRGS, Porto Alegre, RS, Brazil
[d] Laboratório de Biologia Molecular, Instituto de Pesquisas Veterinárias Desidério Finamor (IPVDF), Fundação Estadual de Pesquisa Agropecuária (FEPAGRO), Eldorado Do Sul, RS, Brazil

❑ Application and Usefulness of NGS -3

- Investigational tool

For example: to clarify if an identified contaminant is replicative

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Use of a new RNA next generation sequencing approach for the specific detection of virus infection in cells

Audrey Brussel[a], Kerstin Brack[b], Erika Muth[c], Rudolf Zirwes[c], Justine Cheval[c], Charles Hebert[c], Jean-Marie Charpin[c], Alice Marinaci[b], Benoit Flan[a], Horst Ruppach[b], Pascale Beurdeley[c], Marc Eloit[c,d,e,*]

[a] LFB, Courtaboeuf, France
[b] Charles River Laboratories Germany GmbH, Erkrath, Germany
[c] PathoQuest, Paris, France
[d] National Veterinary School of Alfort, Paris-Est University, Maisons-Alfort, France
[e] Pathogen Discovery Laboratory, Biology of Infection Unit, Institut Pasteur, Paris, France

AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

RESEARCH ARTICLE
Applied and Environmental Science

Adventitious Virus Detection in Cells by High-Throughput Sequencing of Newly Synthesized RNAs: Unambiguous Differentiation of Cell Infection from Carryover of Viral Nucleic Acids

Justine Cheval,[a] Erika Muth,[a] Gaëlle Gonzalez,[b] Muriel Coulpier,[b] Pascale Beurdeley,[a] Stéphane Cruveiller,[a] Marc Eloit[a,c,d]

❑ Application and Usefulness of NGS -4

Biologicals 59 (2019) 29–36

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Use of a new RNA next generation sequencing approach for the specific detection of virus infection in cells

Audrey Brussel[a], Kerstin Brack[b], Erika Muth[c], Rudolf Zirwes[c], Justine Cheval[c], Charles Hebert[c], Jean-Marie Charpin[c], Alice Marinaci[b], Benoit Flan[a], Horst Ruppach[b], Pascale Beurdeley[c], Marc Eloit[c,d,e,*]

[a] LFB, Courtaboeuf, France
[b] Charles River Laboratories Germany GmbH, Erkrath, Germany
[c] PathoQuest, Paris, France
[d] National Veterinary School of Alfort, Paris-Est University, Maisons-Alfort, France
[e] Pathogen Discovery Laboratory, Biology of Infection Unit, Institut Pasteur, Paris, France

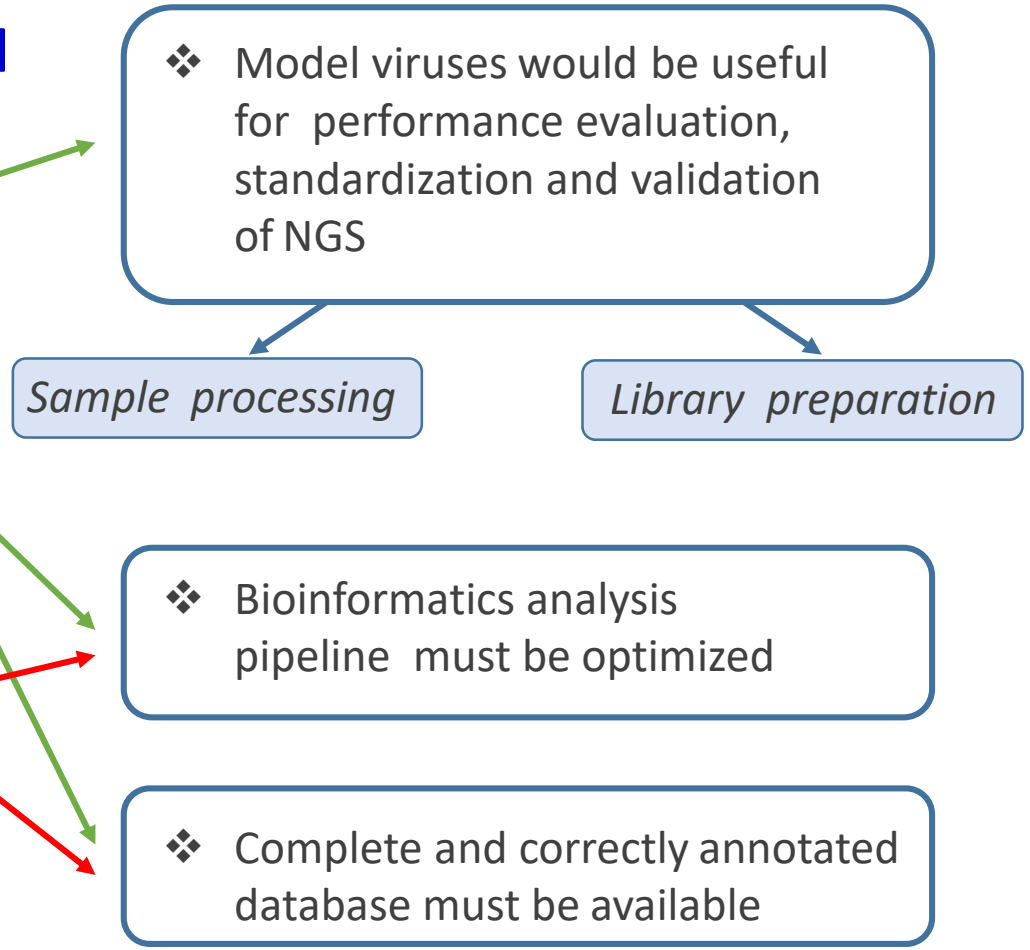
❑ Other potential applications of NGS

- NGS could also be used at different stages e.g. product development, manufacturing or finished product:
  - Identification and characterization of vaccine strains
  - Evaluation of genetic stability of vaccine strains after successive passages
  - Reversion to virulence of the attenuated vaccine strains

❑ Challenge to Use NGS to Detect Adventitious Agents

Major challenge → Validation of NGS Method

- Diversity of viral targets and biological matrices (e.g. cell banks, viral seeds, raw materials)
- Complexity of the NGS technologies and associated bioinformatics



❑ Due to the need to validate each step of NGS method, a coordinate work among specialists is important

In 2014 the Advanced Virus Detection Technologies Interest Group (AVDTIG), gathering together Regulatory and Government agencies, Industry, Service providers, Technology developers, and Academics from all over the world, has been formed.

Efforts of the Advanced Virus Detection Technologies Interest Group (AVDTIG) for NGS applications in Biologics

2nd Conference on Next Generation Sequencing for Adventitious Virus Detection in Human and Veterinary Biologics - An IABS-EU Meeting

November 13-14, 2019 - Het Pand, University of Ghent, Belgium

Jean-Pol Cassart and AVDTIG

❑ Validation and Standardization -1

➢ Preliminary consideration

- ❖ NGS is not a quantitative analysis
- ❖ Sample flow similar to PCR assays

➢ Sample and library preparation

- ❖ Extractions and recovery of viral nucleic acids controls (accuracy of the method)
- ❖ Extractions and recovery

❑ Validation and Standardization -2

- Appropriate model viruses for spiking studies (needs for a standard).

  - Efficiency of the different steps of the methodology
  - Evaluation of total NGS workflow in different biological matrices
  - Compare NGS with current assays for virus detection (PCR, *in vivo*, *in vitro*)
  - Generation of well-characterized datasets for evaluating bioinformatics pipelines
  - Sensitivity studies

Reagent available from NIBSC catalogue
www.nibsc.org/products ref: 11/242-001

Vaccine 34 (2016) 2035–2043

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Development of a candidate reference material for adventitious virus detection in vaccine and biologicals manufacturing by deep sequencing

Edward T. Mee[a,*], Mark D. Preston[b], CS533 Study Participants[1], Philip D. Minor[a], Silke Schepelmann[a]

[a] Division of Virology, National Institute for Biological Standards and Control, Medicines and Healthcare Products Regulatory Agency, South Mimms, Hertfordshire EN6 3QG, United Kingdom
[b] Division of Technology, Development and Infrastructure, National Institute for Biological Standards and Control, Medicines and Healthcare Products Regulatory Agency, South Mimms, Hertfordshire EN6 3QG, United Kingdom

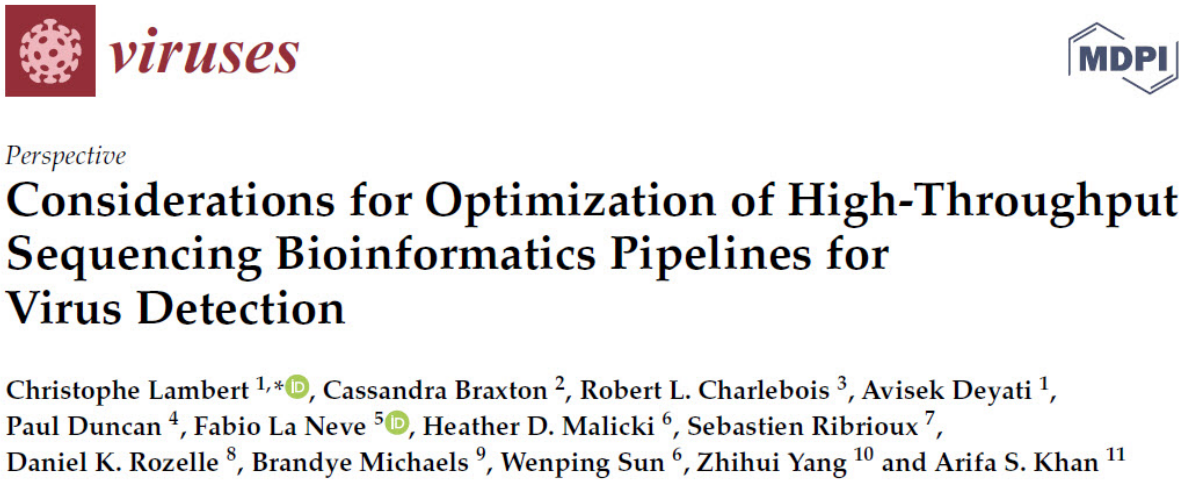
❑ Validation and Standardization -3

- Specificity
  - Demonstrated by a negative control extracted and sequenced in parallel
  - Breadth of detection confirmation

❑ Validation and Standardization -4

➢ Bioinformatics - Pipeline optimization

- ❖ Criteria for acceptable quality of reads
- ❖ Parameters for short read assembly
- ❖ hybrid assembly to correct high error-rate in long-read sequencing
- ❖ Strategies to identify novel viruses with minimal similarity to known sequences



❑ Validation and Standardization -5

- Development of a complete and correctly annotated, publicly available, Reference Virus Database

Database available at:

https://rvdb.dbi.udel.edu/

AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

RESEARCH ARTICLE
Applied and Environmental Science

Check for updates

A Reference Viral Database (RVDB) To Enhance Bioinformatics Analysis of High-Throughput Sequencing for Novel Virus Detection

Norman Goodacre,[a] Aisha Aljanahi,[a]* Subhiksha Nandakumar,[a] Mike Mikailov,[b] Arifa S. Khan[a]

❑ NGS positive sample: follow-up strategy

- Confirmation of a “true” hit
  - Can the results be confirmed by PCR or another assay?
  - Is a complete viral genome present?
- Determination of biological relevance and significance of a positive signal
  - Are particles present?
  - Are the particles infectious?
  - Is there a replication-competent virus?
  - Can the nucleic acid/particles be quantified?

❑ Summary of the steps which need validation to utilize NGS for Biological and Biotechnological Products -1

- **Sample preparation and processing**
  - Extraction efficiency of different virus structure (with/out envelope)
  - cDNA synthesis of different virus genome (Single/double strand; DNA/RNA)
  - Library preparation
  - Enrichment steps for viral nucleic acid Controls (reagents, method)
- **Sequencing platform**
  - Selection of sequencer to provide sufficient reads to detect a low level virus
  - Consider error rate of sequencing technology: short reads vs long reads

❑ Summary of the steps which need validation to utilize NGS for Biological and Biotechnological Products -2

- **Bioinformatics**
  - Strategies for detection of known and novel viruses (nucleotide vs amino acids, programs/tools, reads vs contigs, criteria and parameters for runs)
  - Databases
  - Unmapped reads?
  - Re-analysis?

❑ Conclusions

- Evaluation of NGS platforms for virus detection
- Standardization of the methods, including availability of virus references representing different virus families
- Developing bioinformatics tools and strategies for accurate virus detection and data interpretation

❑ Future perspective -1

- Further dialogue between researchers, developers, companies and regulators to understand **current hurdles** to approve the implementation of NGS.

- Improvement of experimental projects for an accurate standardization of all the steps involved in NGS and biologicals control.

- Collaboration between researchers, companies and regulators for the development of specific guidance on requirements for regulatory acceptance of NGS.

❑ Future perspective -2

- Coordination between regulatory bodies to **harmonize requirements**

- Organize collaborative studies to address technology complexity on common grounds

❑How to Balance Regulation and Innovation

- Need of specific guideline
- A strict Validation is requested
- Validation must cover all the steps
- all Validation steps must have a unique rationale
- Would facility acceptance of NGS by the regulatory organism

References

- http://www.eisenlab.org/
- http://www.summerschool.at/
- http://themedicalbiochemistrypage.org/
- https://www.technologynetworks.com/genomics/articles/an-overview-of-next-generation-sequencing-346532
- https://www.slideshare.net/arundhatimehta50/next-generation-sequencing-76687162
- https://nanoporetech.com/applications/dna-nanopore-sequencing

❑ Acknowledgements

- Christina Von Hunolstein
- Stefania Catone
- Giulio Pisani
- Patrizia Caprari

Address for enquiries: domenico.genovese@iss.it

ISTITUTO SUPERIORE DI SANITÀ
CENTRO NAZIONALE
CONTROLLO
E VALUTAZIONE DEI FARMACI