Lezione 5: Metodi di sequenziamento a singola molecola (Single Molecule Real Time Technology e Nanopore sequencing)

Single Molecule Real Time Technology (Pacific Bioscience)

https://www.youtube.com/watch?v=v8p4ph2MAvI
https://www.youtube.com/watch?v=NHCJ8PtYCFc

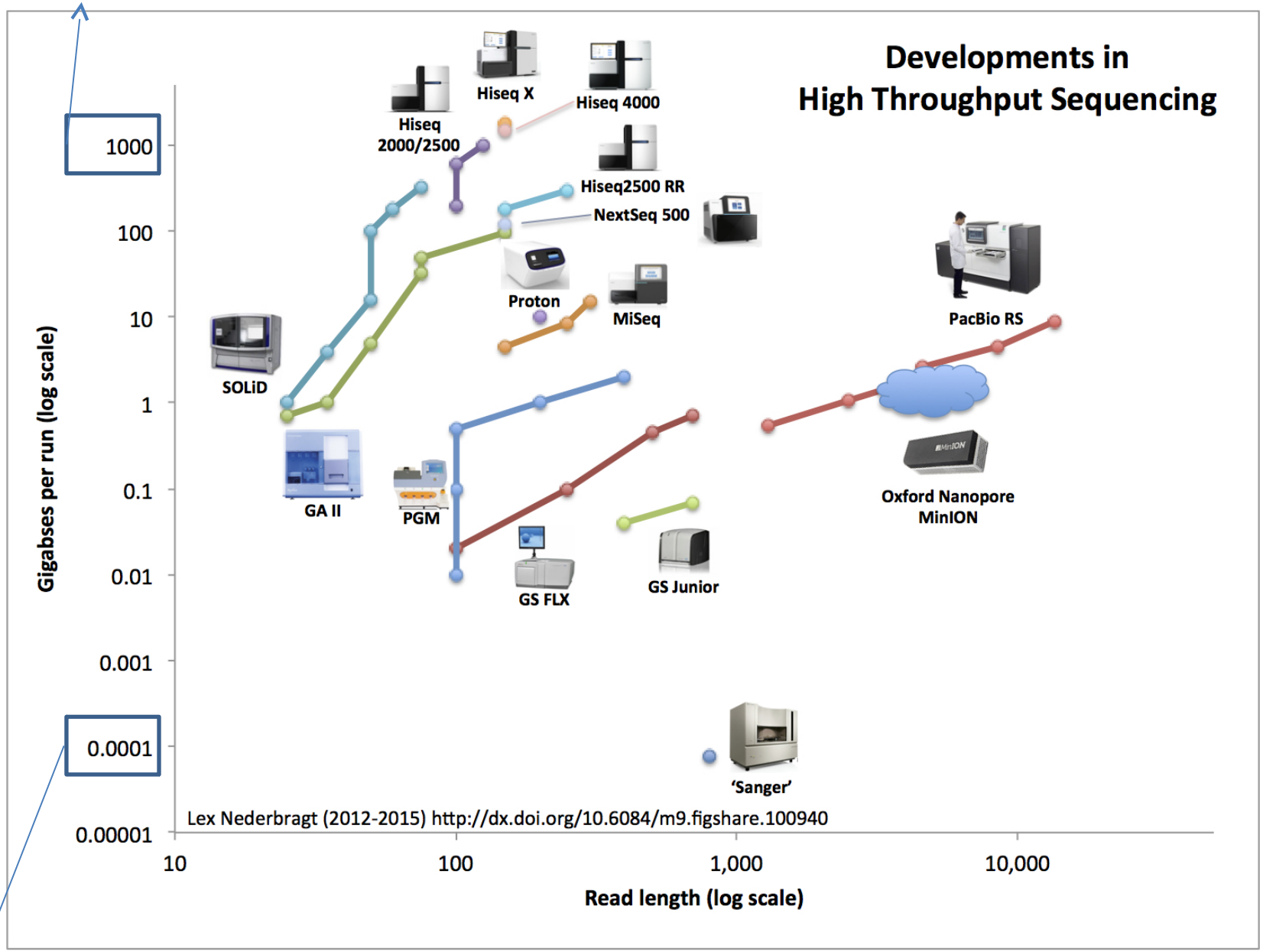
VS

Nanopore sequencing

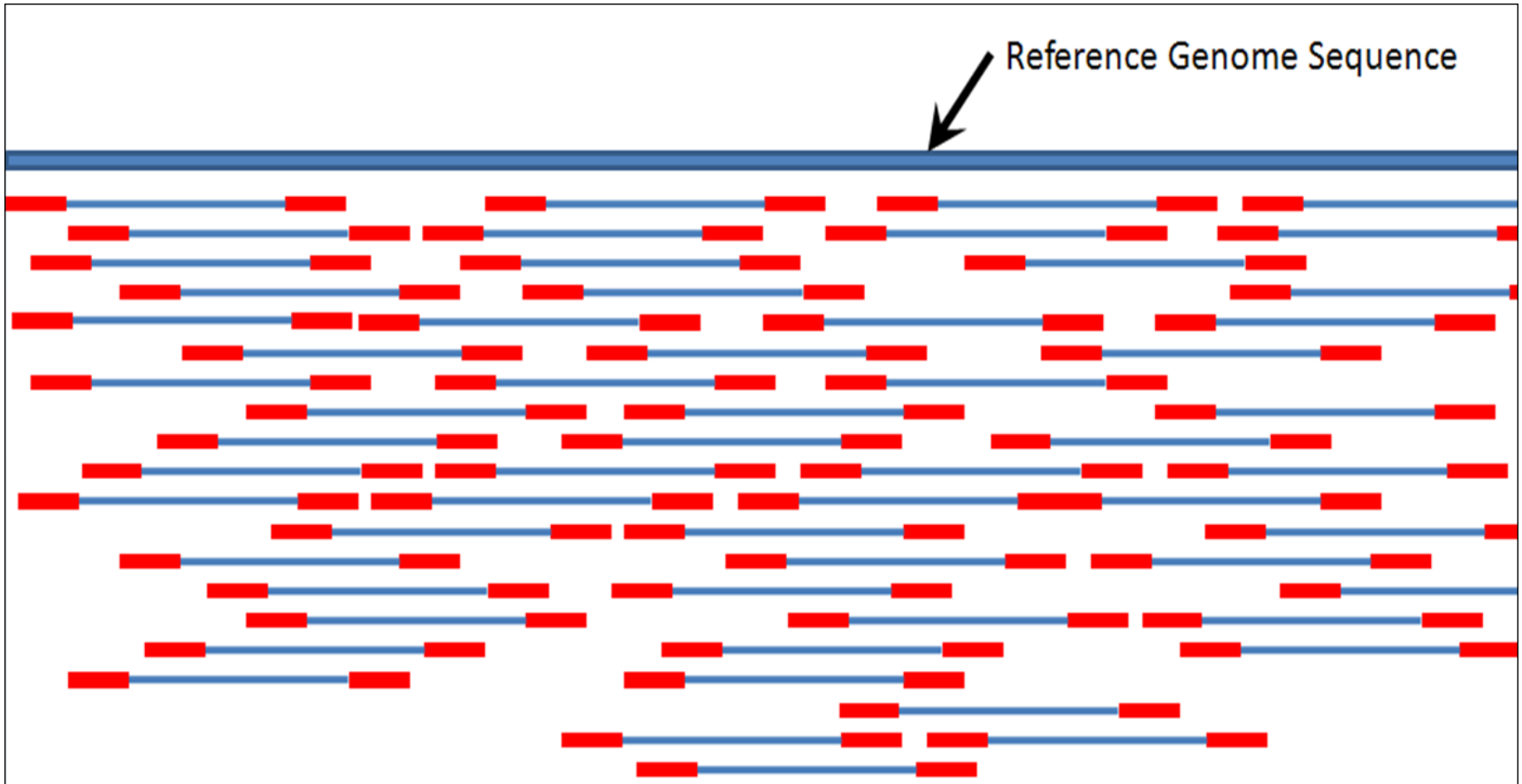
(Oxford Nanopore)

https://www.youtube.com/channel/UC5yMlYjHSgFfZ37LYq-dzig

Mille miliardi di basi



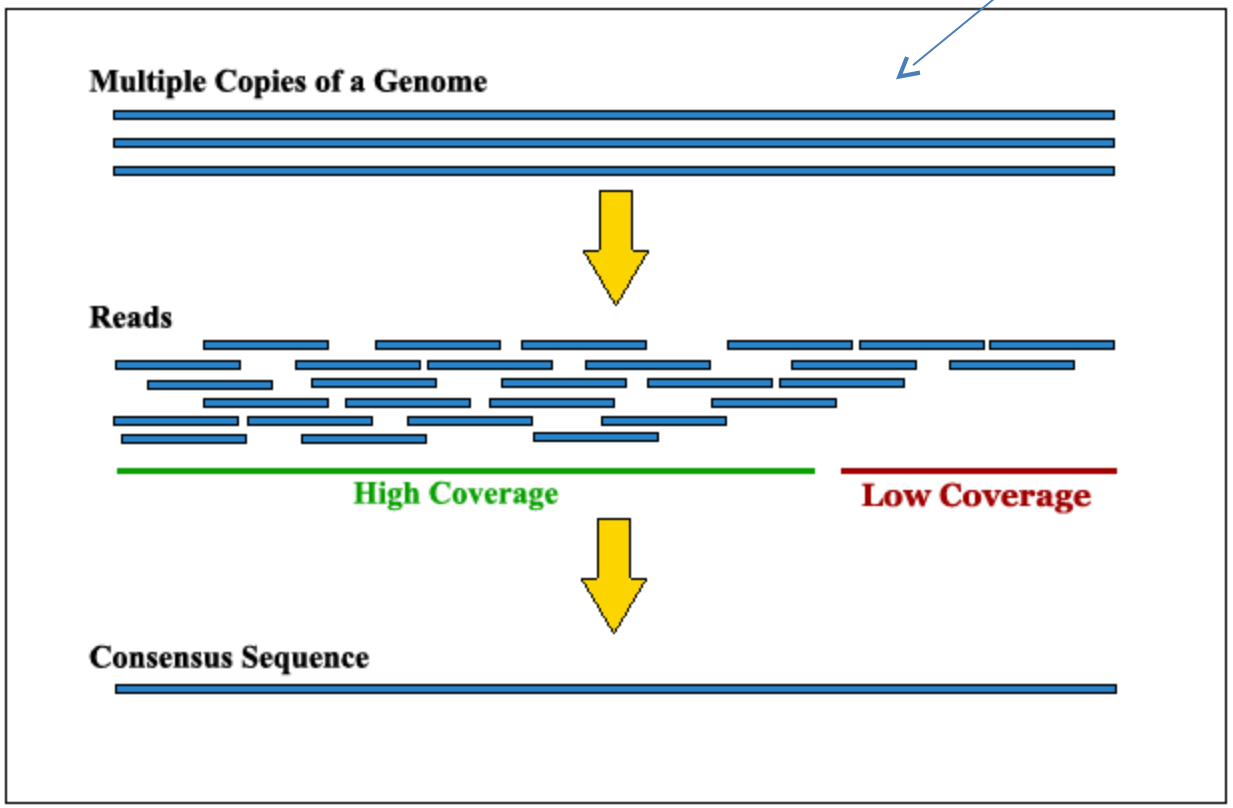
Centomila basi



Assemblaggio (assembly) di reads (i pezzettini blu) su un genoma di riferimento

E se non abbiamo il genoma di riferimento? Assemblaggio (assembly) di reads (i pezzettini blu) *de novo* per creare un consenso

Sequenza Sconosciuta!



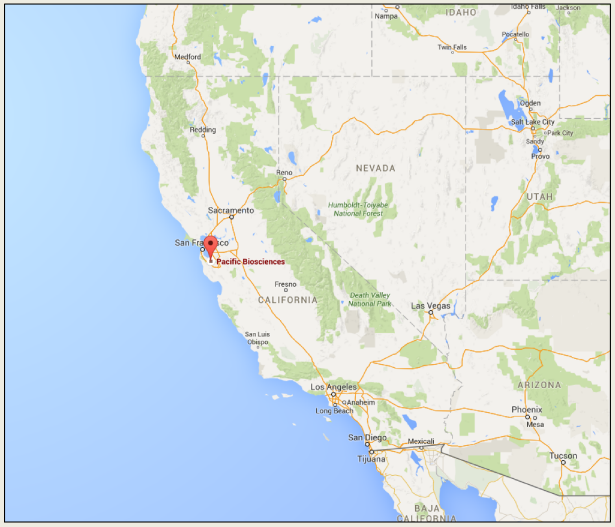
Long reads offrono un'eccellente opzione per **costruire uno scaffold** per assemblaggi ***de novo***.

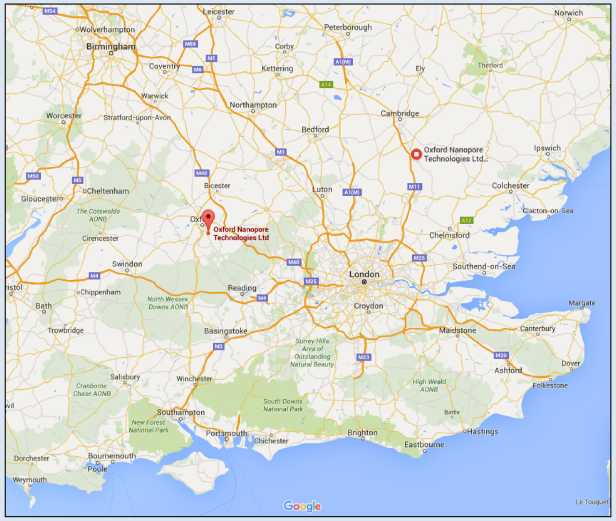
Long reads offrono la possibilità di conoscere la fase cromosomica (aplotipo).

Haplotype 1: AAG
Haplotype 2: GGA

PACIFIC BIOSCIENCES®

Oxford NANOPORE Technologies™





	PacBio RS	Oxford Nanopore
Single Molecule, Real Time sequencing	✔	✔
*Amplification free workflow	✔	✔
**Long reads	✔	✔
***Sequencing by synthesis (DNA polymerase)	✔	✗ No synthesis

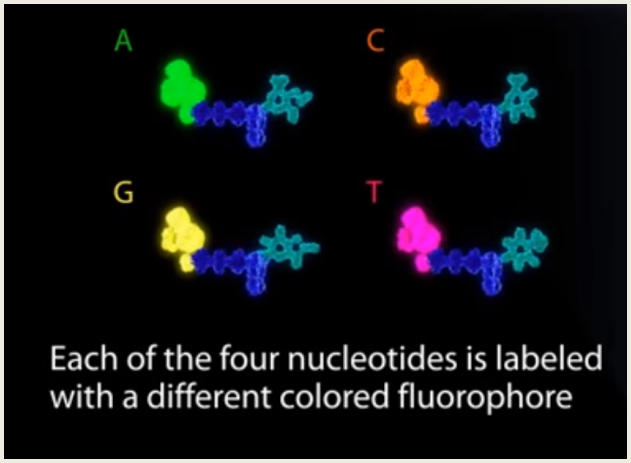
*Nelle tecnologie precedenti (Sanger, NGS) era necessario **moltiplicare i frammenti in modo da avere un segnale sufficientemente potente per poter essere leggibile**. Riuscire a leggere una **singola molecola di DNA**, monitorando in tempo reale l'aggiunta dei nucleotidi, significa poter usare piccole quantità di DNA risparmiando sui reagenti e sui tempi di preparazione.

****Long reads** offrono un'eccellente opzione di **costruire scaffolds** per assemblaggi *de novo* e per risolvere **regioni ripetitive**, ma in generale mancano di accuratezza.

*****PacBio utilizza DNA polimerasi e nucleotidi marcati**. **Oxford Nanopore**, al contrario, **legge una sequenza di DNA basandosi sui cambiamenti nel flusso di corrente che attraversa dei pori di pochi nanometri**.

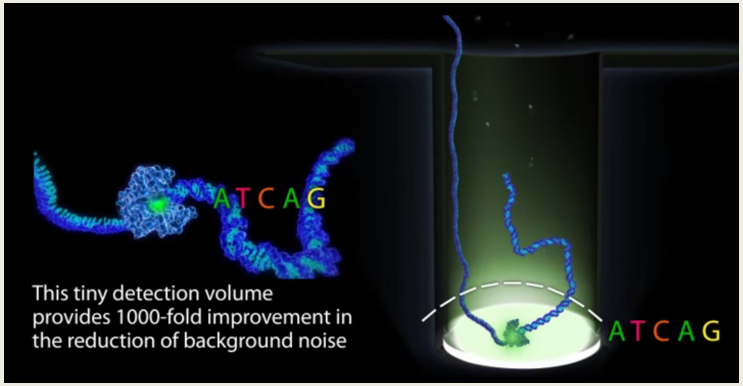
INNOVATIONS: PAC BIO RS

1° innovation: **Phospholinked nucleotides**.
Fluorecent label is attached at the terminal phosphate rather than on the base, like it is common in others NGS. The DNA Polymerase cleaves away the fluorecent label as part of the incorporation process, leaving a natural DNA strand.



This leads to advantages in the yield of the polymerase: higher speed, longer read length, higher fidelity.

2° innovation: **Nanophotonic visualization chamber ZMW** (Zero Mode Waveguide) whose diameter is 50 nm-wide > tiny detection volume > 1 nucleotide



Millisecond pulses of fluorescence can only be efficiently collected from the nucleotide as they are being added to the DNA.

FUNZIONAMENTO: PAC BIO RS

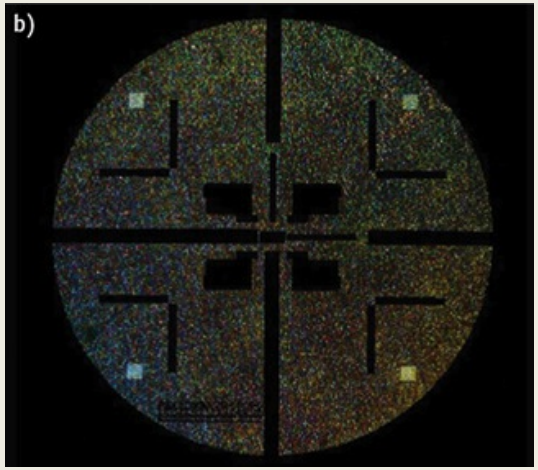
1. A DNA tamplate-polymerase complex is immobilized at the bottom of ZMW.

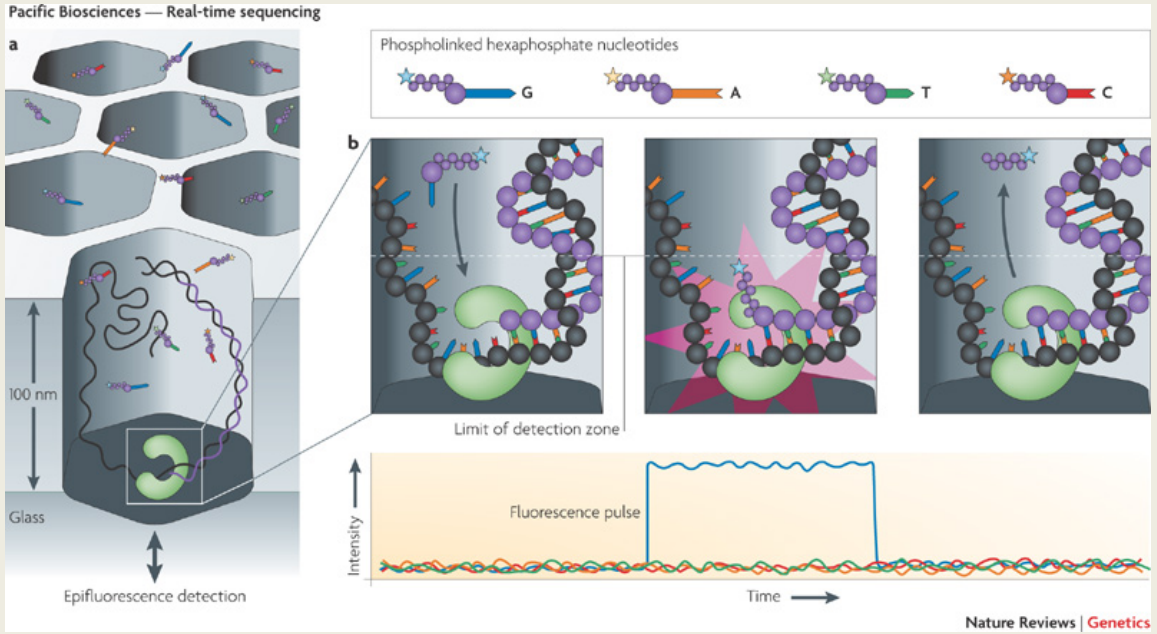
2. Phospholinked nucletides, labeled differently, are introduced into the ZMW chamber.

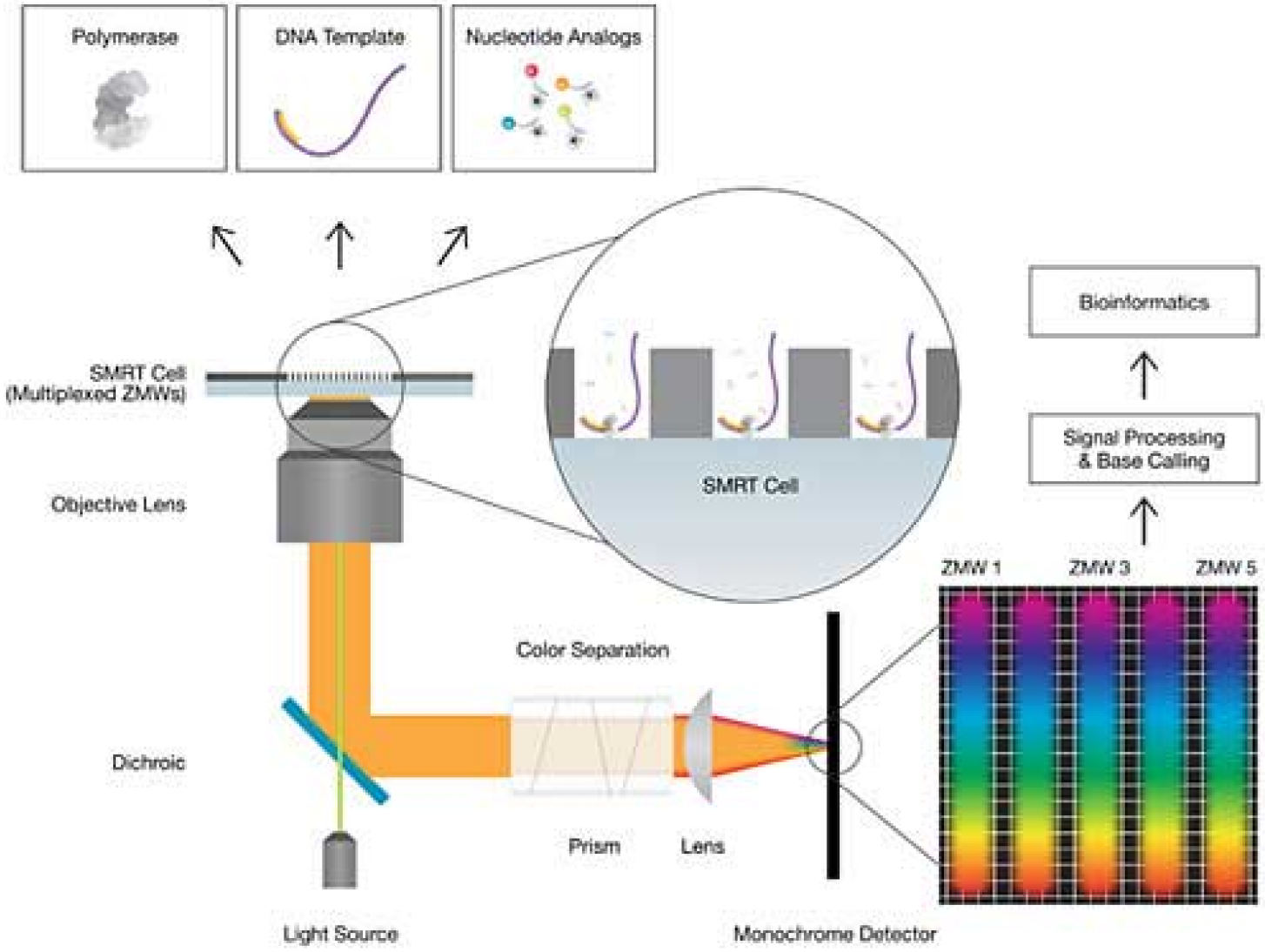
3. During incorporation, **the phosphate chain is cleaved**, **releasing the attached fluorophore which produces a light pulse**. The nucleotide that is incorporated into the new strand will be present for milliseconds while the polymerase is catalyzing the new phosphodiester bond. This time difference provides sufficient signal to enable accurate base-calling.

ZMW provides a tiny detection volume which enables observation of the individual molecules against the required background of labeled nucleotides, while mantaining high signal to noise.

4. This process occurs in parallel in every ZMWs which are monitored simultaneously, in real time.

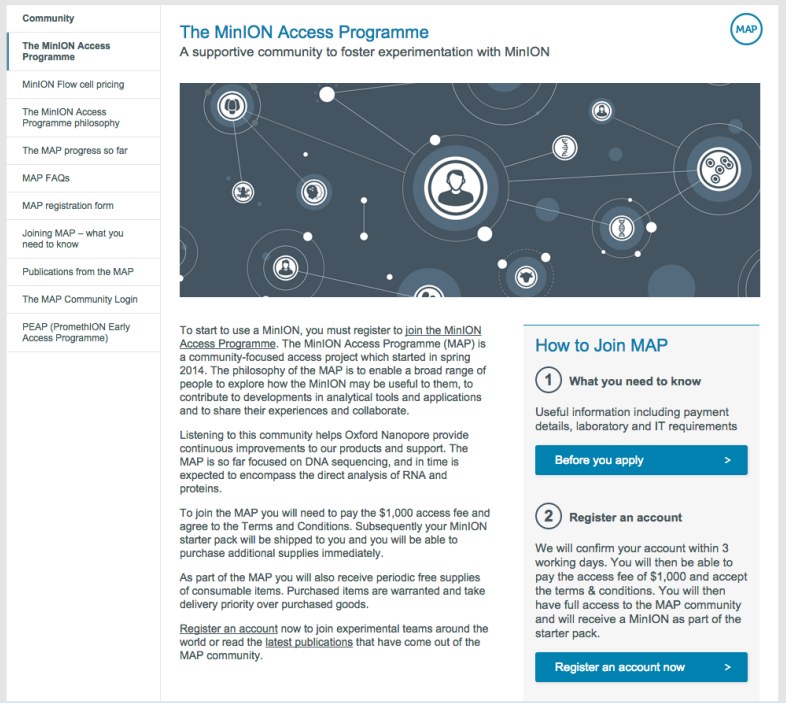





Polymerase
DNA Template
Nucleotide Analogs
Bioinformatics
Signal Processing & Base Calling
SMRT Cell (Multiplexed ZMWs)
Objective Lens
SMRT Cell
ZMW 1
ZMW 3
ZMW 5
Color Separation
Dichroic
Prism
Lens
Light Source
Monochrome Detector

OXFORD NANOPORE **MAP(MinION Access Program)**

In 2014 hundreds of MinIONs have been shipped to roughly 40 countries.



1. Registration and Deposit.

2. MinION Configuration (E. Coli).

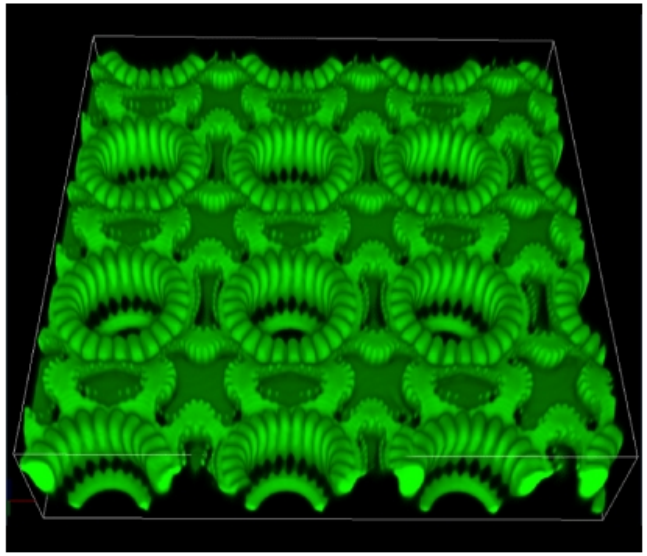
2. Burn-in experiments (fago Lambda).

3. Experimets using own samples.

Users of the Community are discovering what the instrument can do, rather than the company telling the costumer what to expect. This is why there are no specifications available on the company website.

INNOVATIONS: OXFORD NANOPORE

1° innovation: protein **Nanopore** (nanoscale diameter). As a Ionic flow is applied continuously any particle movement(DNA, RNA or even Protein) can disrupt the voltage across the channel which is detected by standard electrophysiological technique.
Readout is relied on the size difference between all dNMP.



Speed of reading: 30 bases per second per nanopore.

The **lack of polymerase** implies that nanopore sequencing is less sensitive to temperature.

2° innovation: the **small portable size** (90g, 10cm in length) is a way open to new applications for DNA sequencing.

NATURE | NEWS

Pint-sized DNA sequencer impresses first users

Portable device offers on-the-spot data to fight disease, catalogue species and more.

Erika Check Hayden

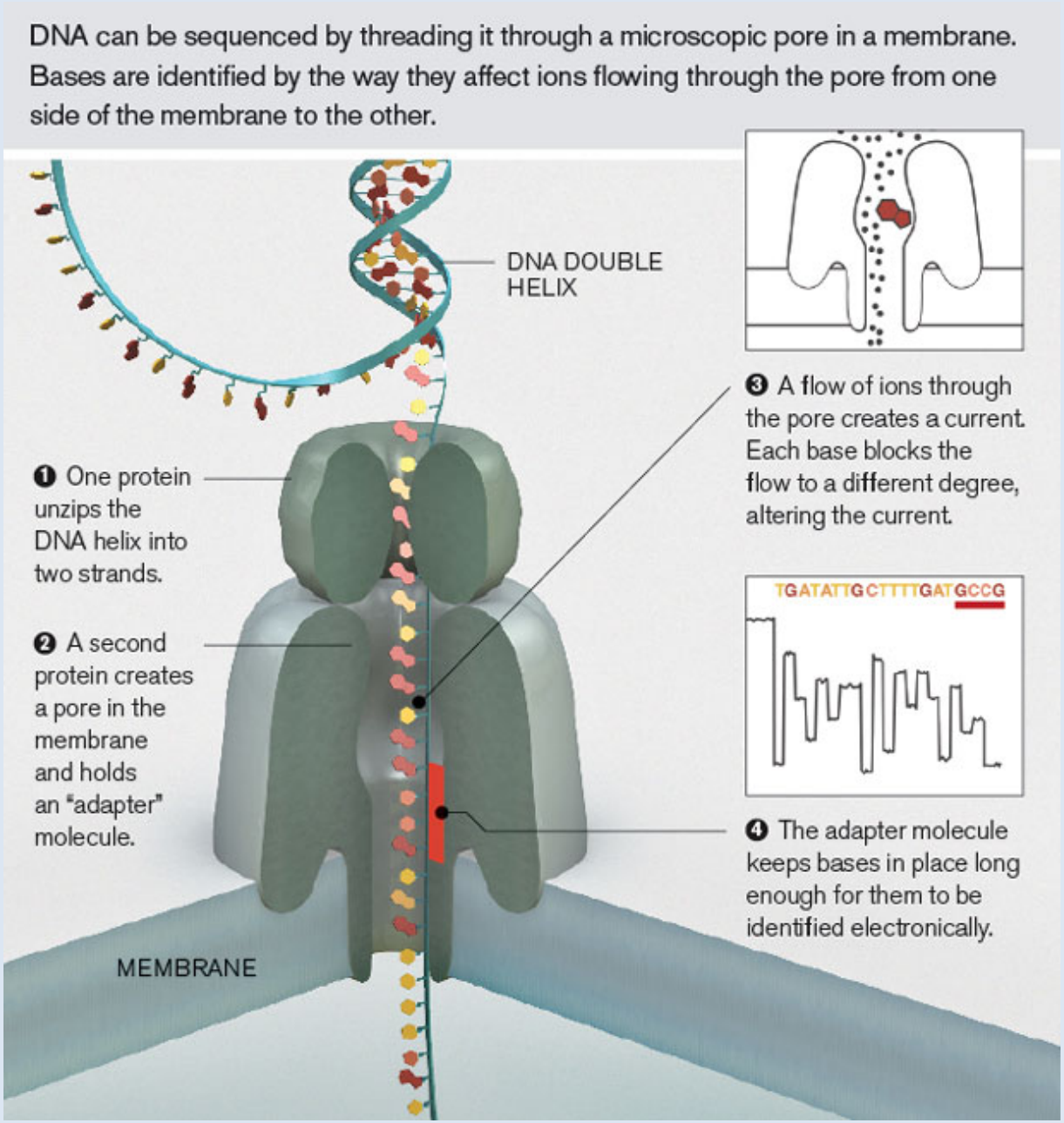
05 May 2015 | Corrected: 11 May 2015

PDF Rights & Permissions

MUSE/Science Museum of Trento

The MinION device can sequence small genomes, such as those of bacteria and viruses, displaying the

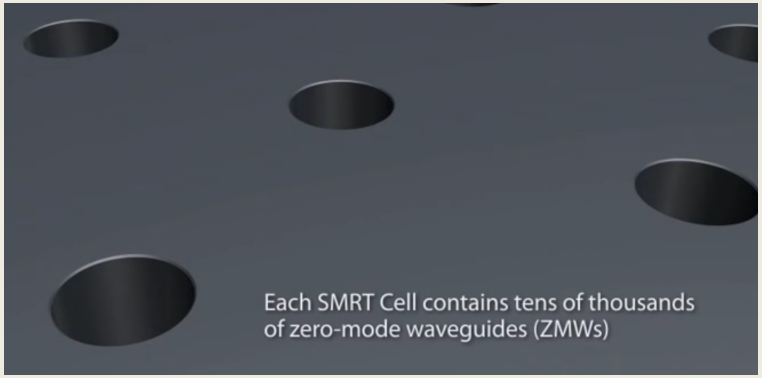
FUNZIONAMENTO: OXFORD



This process occurs in parallel in every pore, which are monitored simultaneously, in real time.

Single Molecule, Real-Time (SMRT) chip:

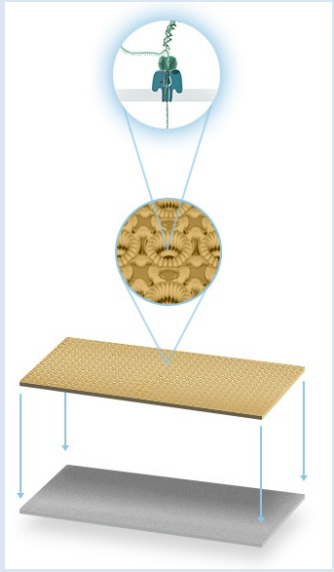
Up to 150,000 ZMWs (approximate diameter 50 nm) are patterned in an aluminum film (100 nm thickness) deposited on a clear silica substrate.



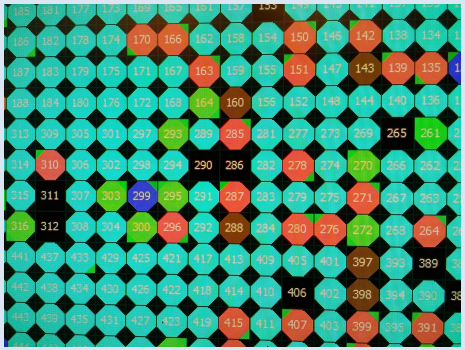
Flowcells:

The portable MinION device works with consumable Flowcells that contains the proprietary nanopore sensing apparatus required to perform experiments.

The MinION plugs into a PC or laptop using a USB3.0.

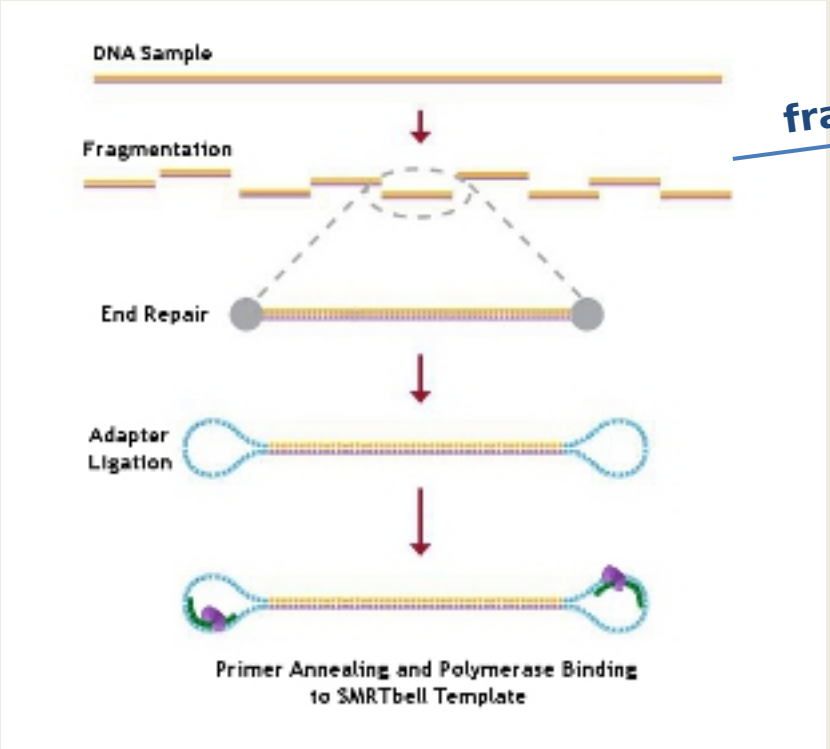


Flowcell contains 500 nanopores and a sensor Array of several hundred channels, to enable multiple nanopore experiments to be performed in parallel.

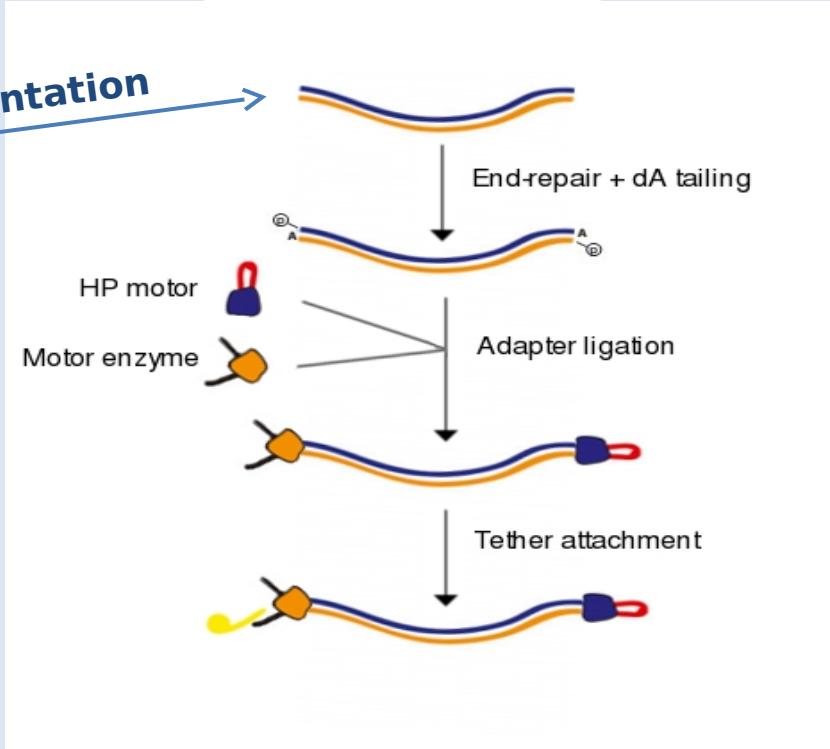


LIBRARIES

•Both requires fragmentation, end-repair and adapter ligation.



fragmentation



LIBRARIES

•Both requires fragmentation, end-repair and adapter ligation.

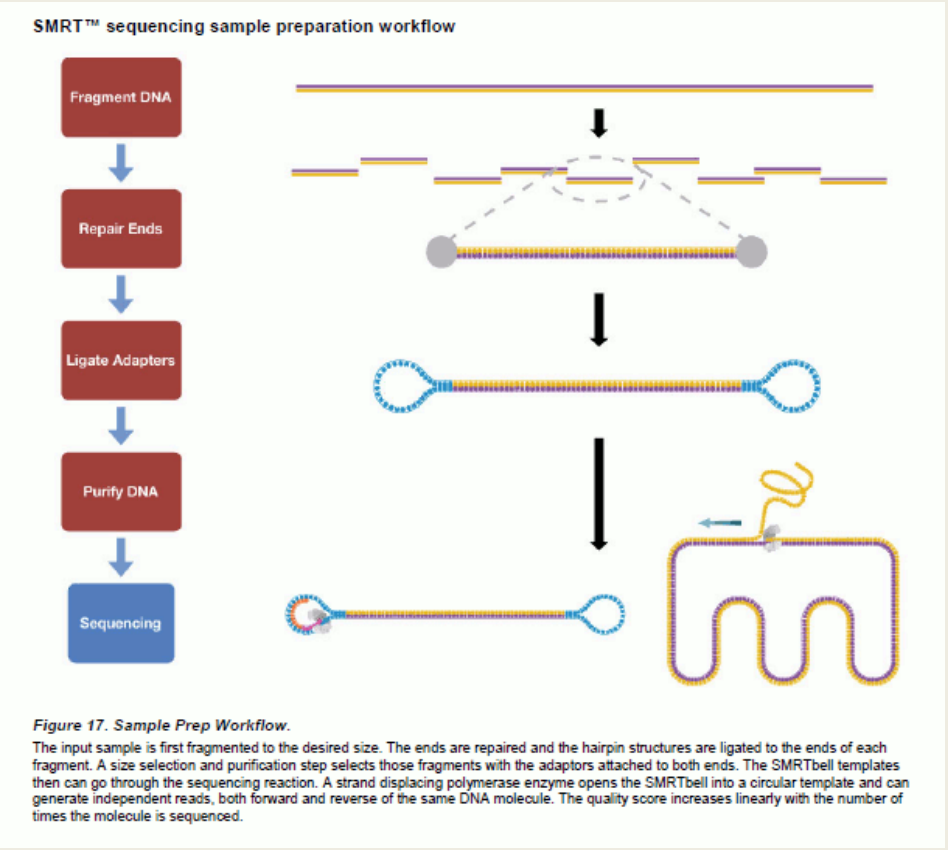
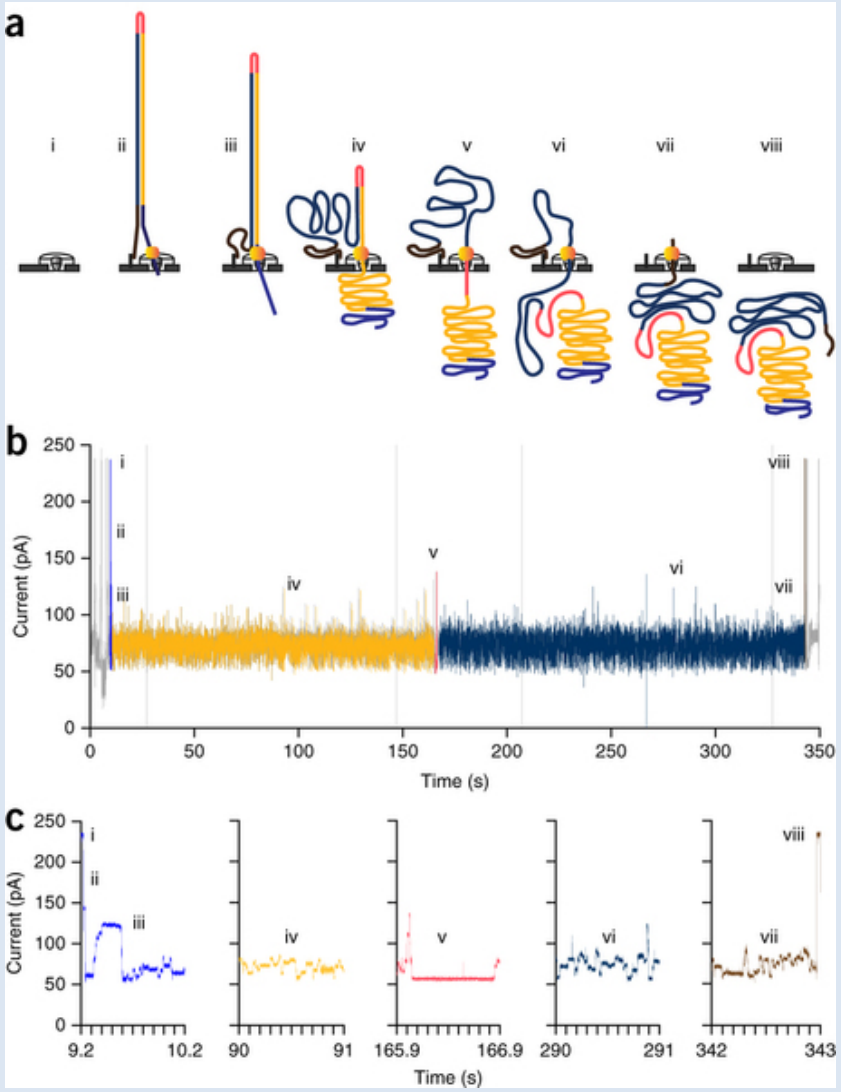


Figure 17. Sample Prep Workflow.
The input sample is first fragmented to the desired size. The ends are repaired and the hairpin structures are ligated to the ends of each fragment. A size selection and purification step selects those fragments with the adaptors attached to both ends. The SMRTbell templates then can go through the sequencing reaction. A strand displacing polymerase enzyme opens the SMRTbell into a circular template and can generate independent reads, both forward and reverse of the same DNA molecule. The quality score increases linearly with the number of times the molecule is sequenced.



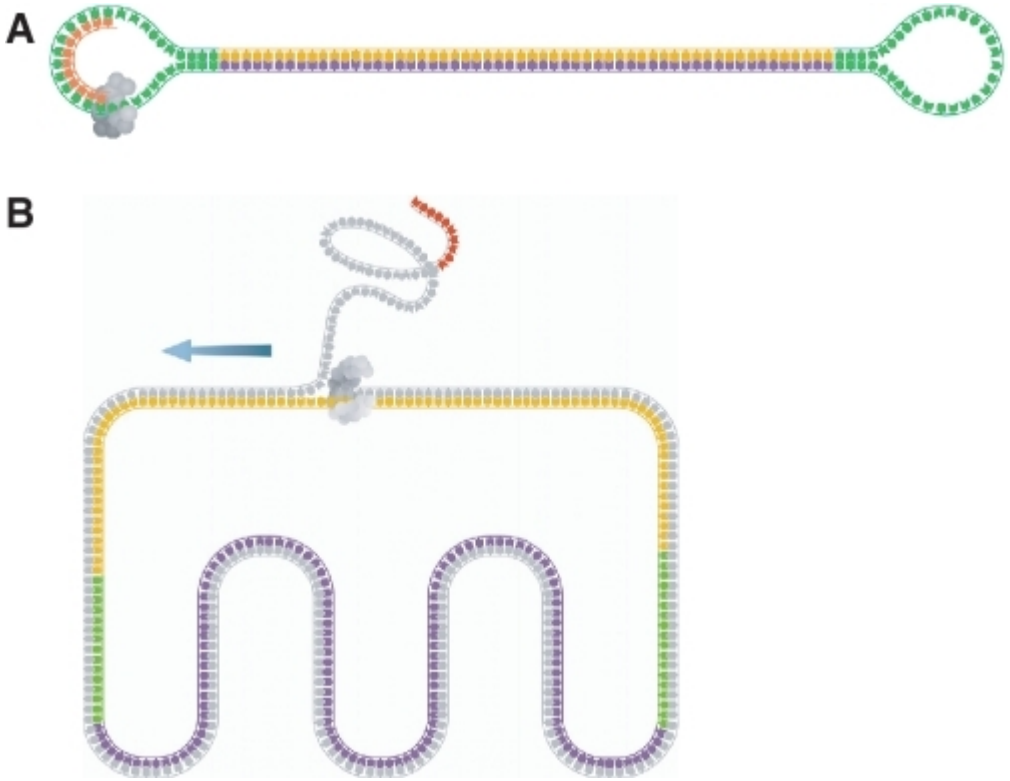


Figure 1: Schematic of a SMRTbell™ template. (A) A SMRTbell template consists of a double-stranded region (the insert) flanked by two hairpin loops. The hairpin loops present a single-stranded region to which a sequencing primer can bind (orange). (B) As a strand-displacing polymerase (gray) extends a primer from one of the hairpin loops, it uses one strand as the template strand and displaces the other. When the polymerase returns to the 5′-end of the primer, it begins strand displacement of the primer and continues to synthesize DNA (moving in the direction of the blue arrow). Therefore, the length of sequence obtained from these templates is not limited by the insert length. Furthermore, the resulting sequence is derived from both sense- and anti-sense strands.

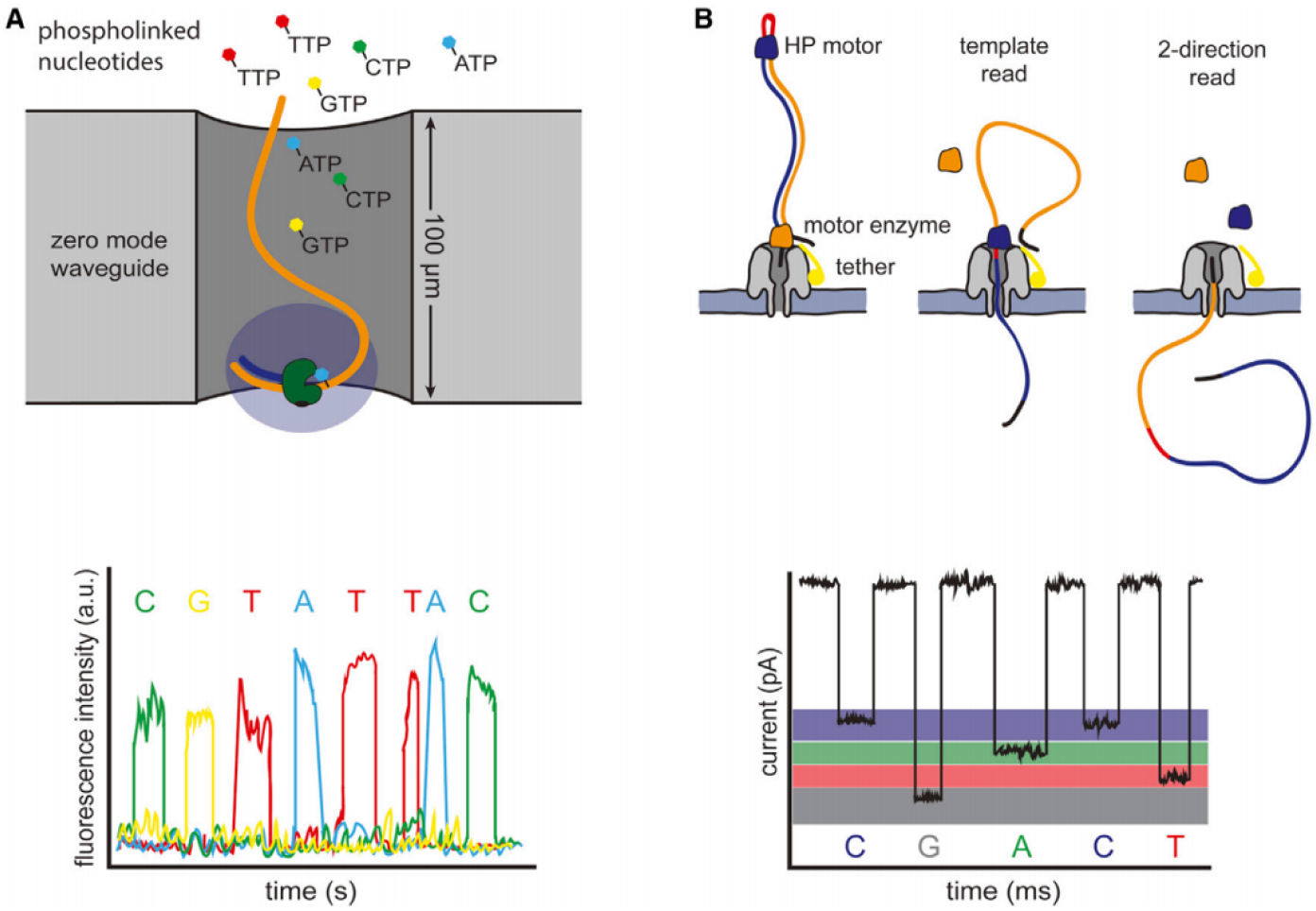


Figure 3. Single Molecule Sequencing Platforms

(A) Pacific Bioscience's SMRT sequencing. A single polymerase is positioned at the bottom of a ZMW. Phosphate-labeled versions of all four nucleotides are present, allowing continuous polymerization of a DNA template. Base incorporation increases the residence time of the nucleotide in the ZMW, resulting in a detectable fluorescent signal that is captured in a video.

(B) Oxford Nanopore's sequencing strategy. DNA templates are ligated with two adapters. The first adaptor is bound with a motor enzyme as well as a tether, whereas the second adaptor is a hairpin oligo that is bound by the HP motor protein. Changes in current that are induced as the nucleotides pass through the pore are used to discriminate bases. The library design allows sequencing of both strands of DNA from a single molecule (two-direction reads).

LIBRARIES

- Fast sample preparation for both coparing to other technologies.
- Both requires high standards of purity and fragment size.

	PacBio RS	Oxford Nanopore MinION
Typical DNA requirements	1ug (depends!)	1ug
Library preparation time	~8 hours	~5 hours

V.

The PacBio Sample Calculator should be used to estimate yield for all samples. The table below provides estimates of expected yields for various DNA libraries. Please note the assumptions used to generate the table.

Library Insert Size*	Recommended Quantity for submission	Min Concentration Required (Post-Shearing)	Est. Total Yield (Range)	
			MIN	MAX
250 bp	600 ng	250 ng	60 GB	125 GB
500 bp	600 ng	250 ng	10 GB	20 GB
1 kb	1.2 µg	500 ng	90 GB	180 GB
2 kb	1.2 µg	500 ng	45 GB	90 GB
5 kb	2.4 µg	1 µg	45 GB	91 GB
10 kb	2.4 µg	1 µg	20 GB	45 GB
10 kb (AMPure kit)	10 µg	5 µg	90 GB	182 GB
20 kb (AMPure kit)	15 µg	5 ug	45 GB	91 GB
20 kb (BluePippin™ kit)	15 µg	5 µg	9 GB	18 GB

* Amounts recommended for submission represent quantities needed for one SMRTbell library prep and includes extra quantity needed for any additional QC and conservative excess. Reported library yield is based on an assumption of a DNA loading concentration of 50 ng/µl and throughput of 200 MB per SMRT Cell using P4-C2 chemistry. For insert sizes ≥1 KB, a magnetic bead loading protocol is used in the SMRT Cell calculation. Two size-selection protocols for large-insert libraries are available using either AMPure or BluePippin strategies. Actual results may vary.

Part 1 - Template preparation

DNA QC

Before beginning library preparation, please ensure that your genomic DNA sample meets the following criteria:

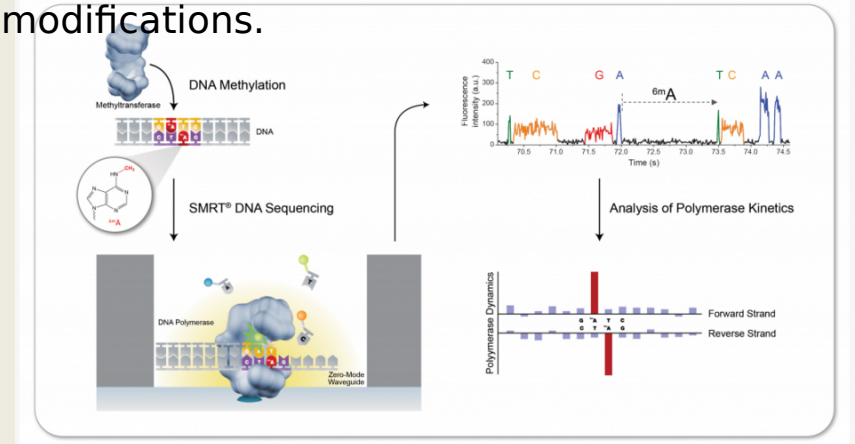
- Purity as measured using Nanodrop – OD 260/280 of 1.8 and OD 260/230 of 2.0–2.2
- Average fragment size, as measured by pulsed-field, or low percentage agarose gel analysis – >30 kb
- Input mass, as measured by Qubit – 1 μg

Further details on assessing the quality of the input sample can be found here. The Lambda DNA supplied meets these criteria.

OTHER FEATURES

•PacBio RS uses **circular libraries** which allows an error correction during sequencing. Highly processive polymerase used can synthesize complementary strands much longer than the insert length of the library. Multiple passes of the same strand can be aligned by the base-calling software to produce a circular consensus sequence, increasing accuracy.

•PacBio RS **uses Kinetic informations to detect DNA modifictions**(such as methylations). When the polymerase encounters a modified base in the template, pauses in order to adjust to the base. Using templates with known modifications, statistical models of incorporation can be built to identify previously uncharacterized DNA modifications.



DNA polymerization runs freely at ~3 bases/second. Alteration of this rate due to the incorporation of nucleotides across modified bases is detected and used to infer the presence of bases other than A, C, T or G. This information is automatically generated and processed during every run.

•To enable sequencing of both strands, a Oxford Nanopore library is constructed from **double-stranded DNA** (dsDNA). The presence of hairpin ties one strand(Template) of DNA to its complement, permitting the reading of both strands. When both Template and Complement strands are sequenced these are combined by the base-calling algorithm to produce **2D reads.**

"1D reads"	"2D reads"
:: Template 1D	:: Normal 2D
: only fwd stran	: mostly fwd, some rev
:: Complement 1D	:: Full 2D
: only rev strand	: most of fwd & rev
	: **these are high quality**

RELEASES AND CHEMISTRY UPGRADES

- Both have several and frequent adjustements.

2011:
- PacBio RS C1

2012:
- PacBio RS C2
- PacBio RS C2 XL

2013:
- PacBio RS II C2 XL

2014:
- PacBio RS II P5 C3
- PacBio RS II P6 C4

2014:
- MinION Mk0
- MinION MkI

- ....
- Flowcell R7
- Flowcell R7.3

- ....
- Sequencing Kit SQK-MAP004

2015:
- Sequencing Kit SQK-MAP005
- Sequencing Kit SQK-MAP006

FEATURES

	PacBio RS	Oxford Nanopore MinION
Read length	~10-20 Kb	~8-20 Kb
Sequence Yield per run	100 Mb-1Gb	50-500 Mb
Run Time	2 hours	6 to 48 hours
Millions of reads per run	0.03	0.1
Bases per read	3000	9000
Single read accuracy	~88%	~95%*

*based on what the Oxford Nanopore claims internally, although no members of the community reports those results, which are closer to ~75%.

ERROR RATES

	PacBio RS	Oxford Nanopore
Primary errors	Indel	Mismatches, Indel
Error Rate(%)	~13 %	~25 %
	Random error (no systematic bias)	

In PacBio Error model is almost completely random which means that with oversampling (multiple reads of the same molecule or of different molecules from the same genomic region) can yield very high quality consensus reads.

So far, the MinION error model isn't quite clear.

PRICES

	PacBio RS	Oxford Nanopore
Purchase Price	$695,000	$1,000 Access free (Access Program)
+ $85 Service contract (Annual maintenance agreements include on-site service)		
Cost/run	$100	$900
Cost/Gb	$1,111	$1,000

ADVANTAGES

PacBio RS	Oxford Nanopore
Single molecule real-time sequencing; Longest available read length; Ability to detect base modifications; Short instrument run time; Random error profile; Modest cost per sample; Many methods being developed; The longer reads also allow CCS (Circular Consensus Sequencing) to provide highly accurate long reads.	Single molecule real-time sequencing; Extremely long reads feasible (multiple kb) Detection of bases is fluorecent tag-free; No DNA polymerase; Small portable instrument; It is a USB device; Extremely low-cost instrument;

DISADVANTAGES

PacBio RS	Oxford Nanopore
High error rates; Low total number of reads per run; High cost per Mb; High capital cost; Many methods still in development; Weak company performance.	Early access program starting, but instruments not yet available; No data publicly available; Unknown official error-rate; Limited information available on sample prep or data yield during 6 hour flow cell life-time; High cost per Mb relative to other Nanopore sequencers in development.