Implementation and Evaluation of Oxford Nanopore Technologies

Colloque EPGV

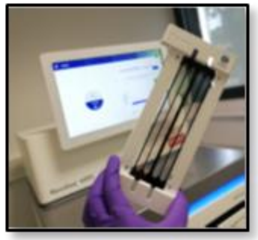
Octobre 2018

Camille Eché

http://get.genotoul.fr
get@genotoul.fr
@get_genotoul

GeT *Core Facility*

- **Genomics and Transcriptomics (GeT)** Platform of Genotoul hosted by INRA
- **NGS technologies :**

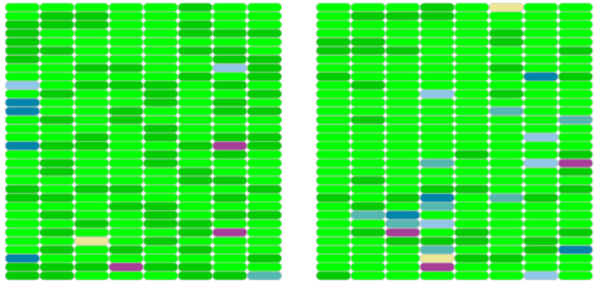


Short reads	Long reads	Synthetic long reads
Illumina : MiSeq HiSeq 3000 NovaSeq	ONT : MinION GridION PromethION	10x genomics : Chromium

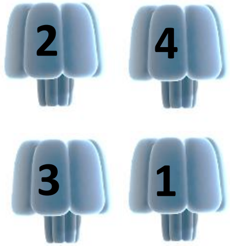
- **A strong partnership with a bioinformatics core facility**
- **A node of the National Distributed Infrastructure « France Génomique »**
- **Quality certifications ISO9001 & NFX 50 900, Propel**

Oxford Nanopore Technologies

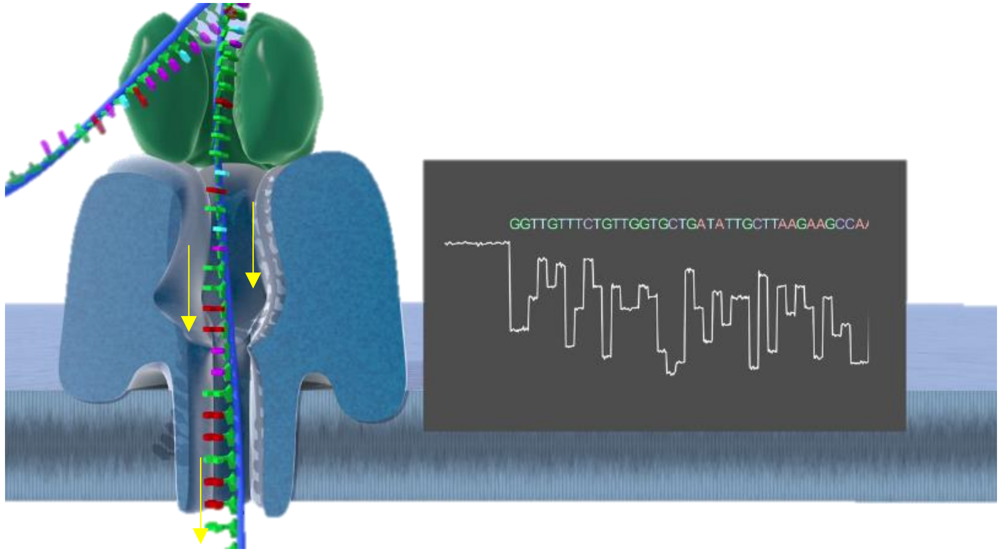
How it works



512 channels



4 pores per channel



- A **protein** is set in an electrically resistant polymer membrane
- An **ionic current** is passed through the nanopore
- The **event / base** creates a characteristic disruption in current
- Identification of G, A, T and C bases

Oxford Nanopore Technologies

*Validation in progress
Same quality and more quantity of data in comparison with the GridION*

ONT	MinION	GridION	PromethION
Number of FC	1	5	24 (up to 48)
Nanopore channels per FC	512	512	3000
Data per flow cell	~ 7 Gb	~ 7 Gb (35 Gb)	~ 25 to 100 Gb (1,2 Tb, up to 2,4 Tb)
Live basecalling	Minknow	Albacore	Albacore
Price per FC	790 $	790 $	2000 $
Price per Gb	112 $	112 $	40 $

Library preparation

- Have a **good DNA quality**

Nanodrop

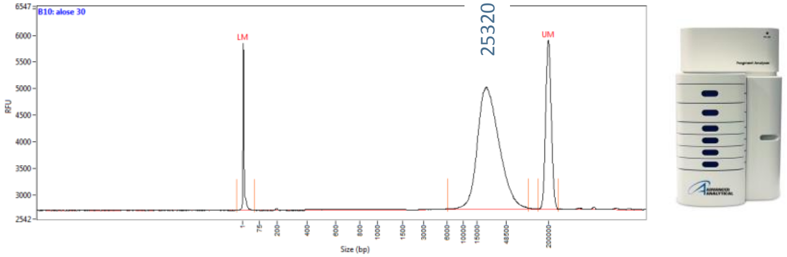
A260/280 = 1.8-2.0

A260/230 = 2.0-2.2

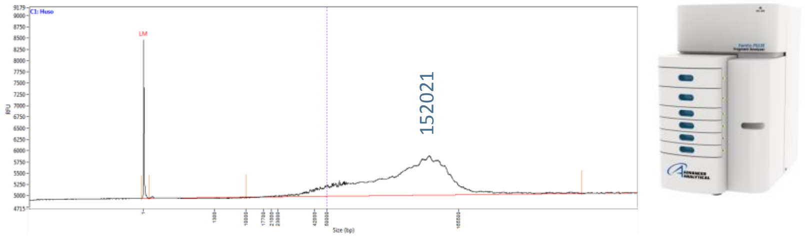
A bad ratio can warn of a library that blocks pores

- Start the library prep with the **right number of molecules**

Qubit



Fragment analyzer
DNA < 40 kb



FEMTO
DNA > 40 kb (up to 165 kb)

Importance of the quality controls

Library preparation

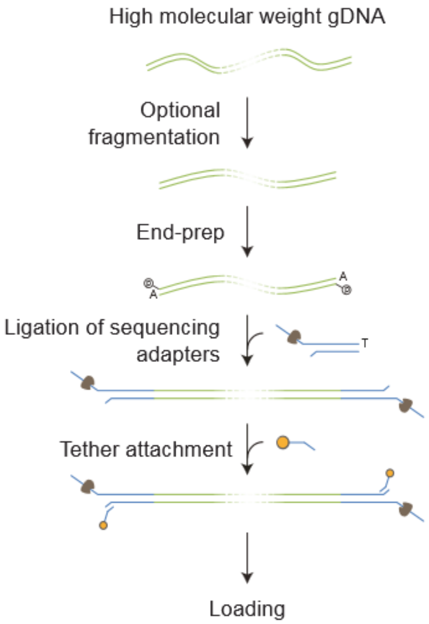
- Make a **suitable library prep**

Megaruptor shearing : 20kb
(improves the yield ; if necessary)

BluePippin : > 10kb
(Removes small fragments ; if necessary)

OPTIONAL

1D lib prep : LSK 109
(optimized for long fragments)



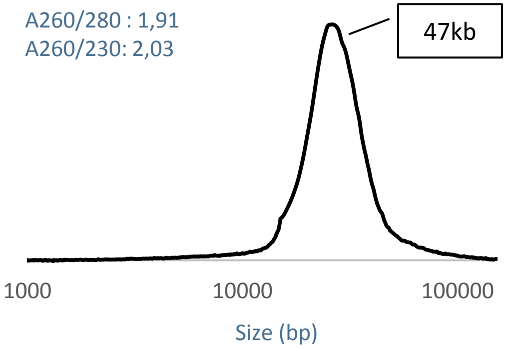
Less purification steps

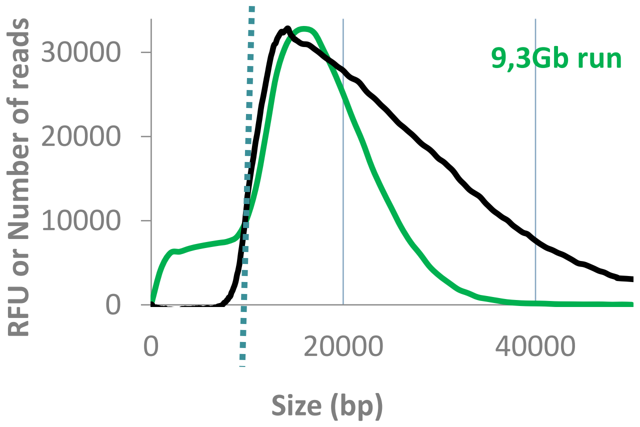
Shearing 20kb ➔ 1D lib prep : DNA repair, ligation adaptaters and motor ➔ Loading and sequencing

Quality impact

Non degraded DNA

A260/280 : 1,91
A260/230: 2,03



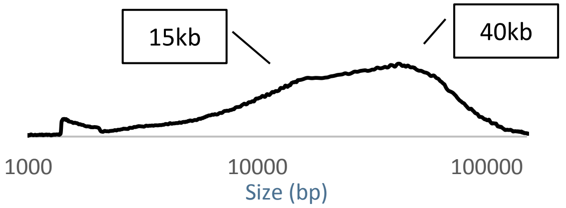


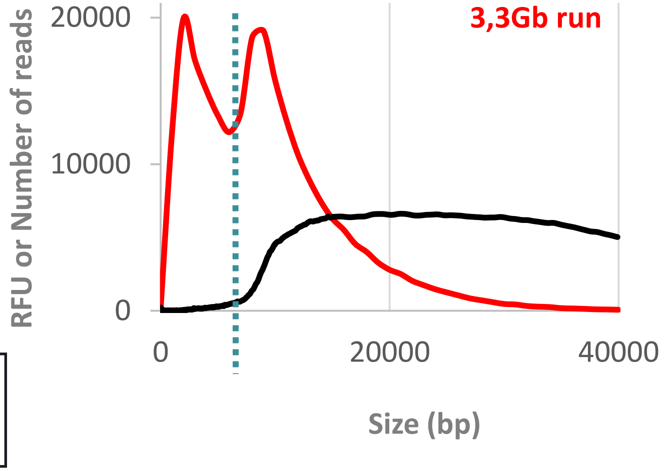
20kb shearing

>10kb selection

Degraded DNA

A260/280 : 1,94
A260/230: 1,97



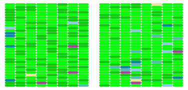


Library	Reads	Blue Pippin Cutoff
▬	🟩 🟥	⋯⋯

The read distribution depends on the DNA quality

Shearing impact

15kb – 20kb DNA : preferred size



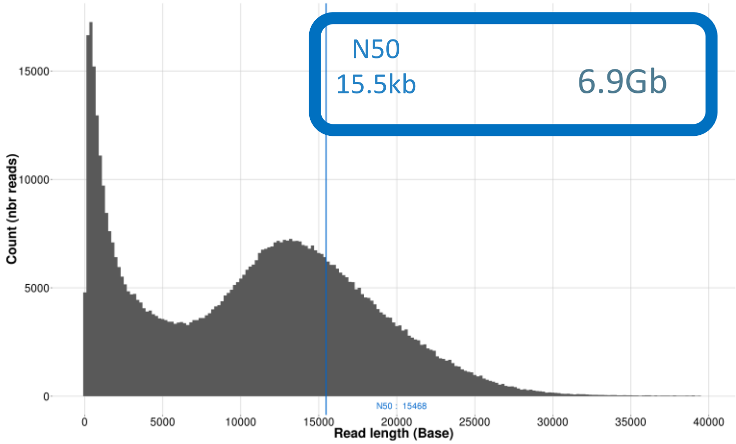
Similar number of active pores
Similar number of molecules loaded

20kb shearing

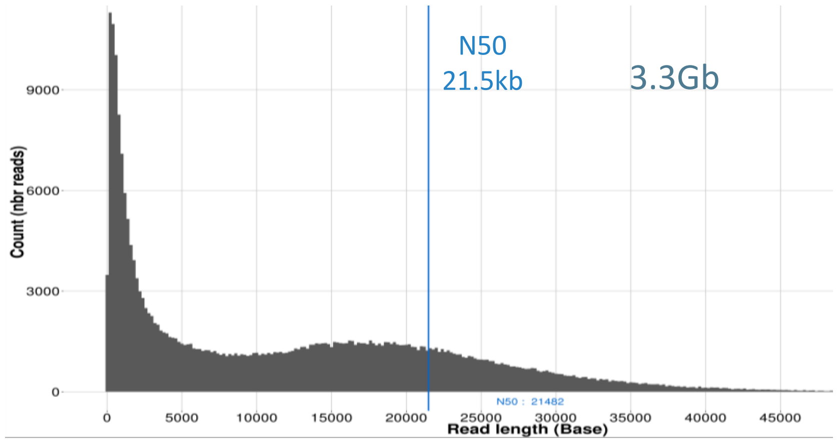
30kb shearing

40kb shearing

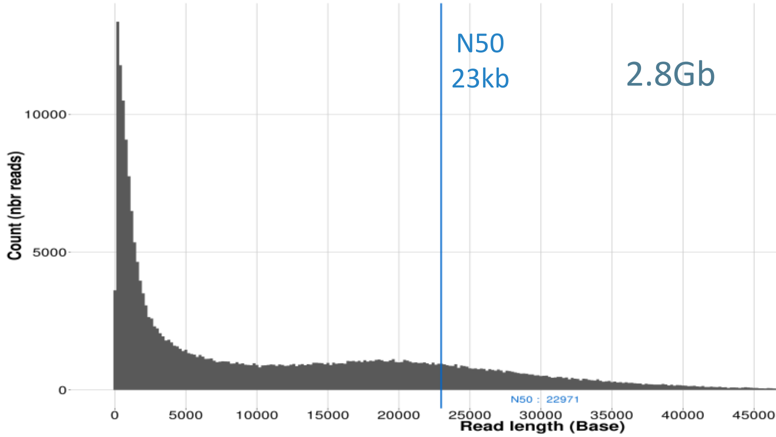
DNA 1



N50
15.5kb
6.9Gb

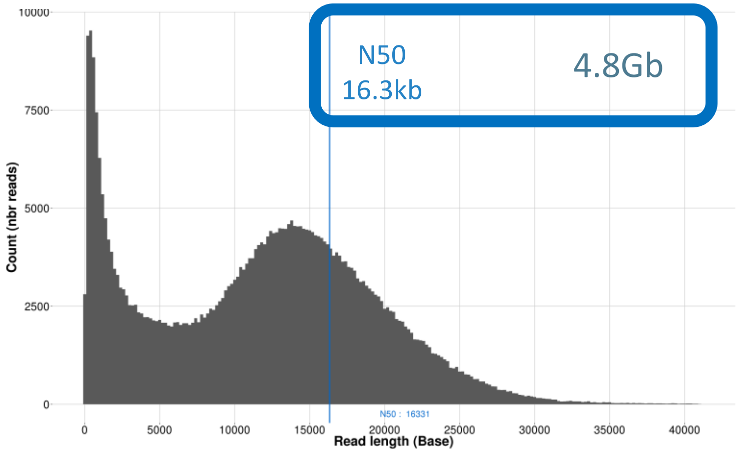


N50
21.5kb
3.3Gb

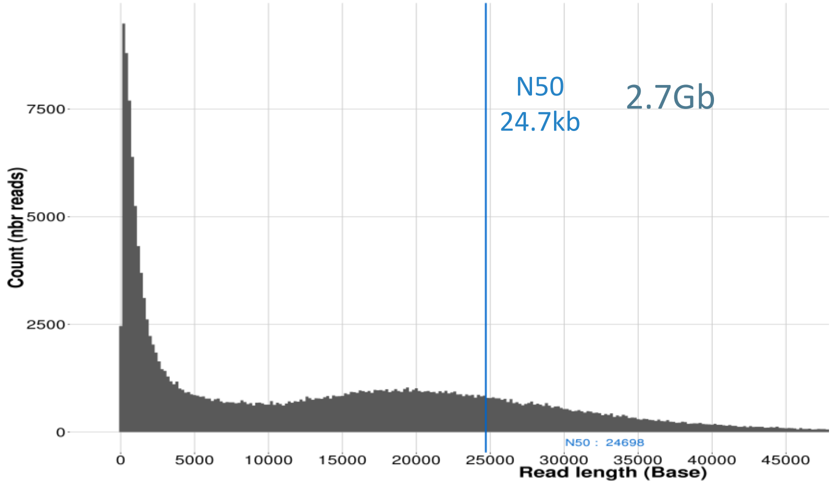


N50
23kb
2.8Gb

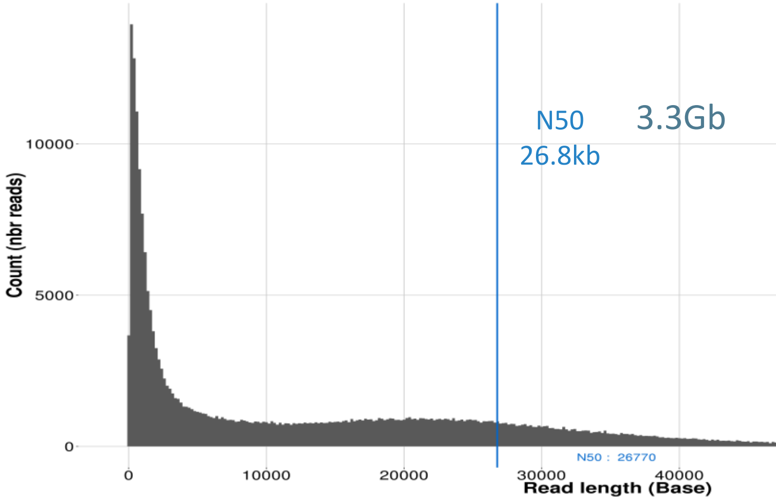
DNA 2



N50
16.3kb
4.8Gb



N50
24.7kb
2.7Gb



N50
26.8kb
3.3Gb

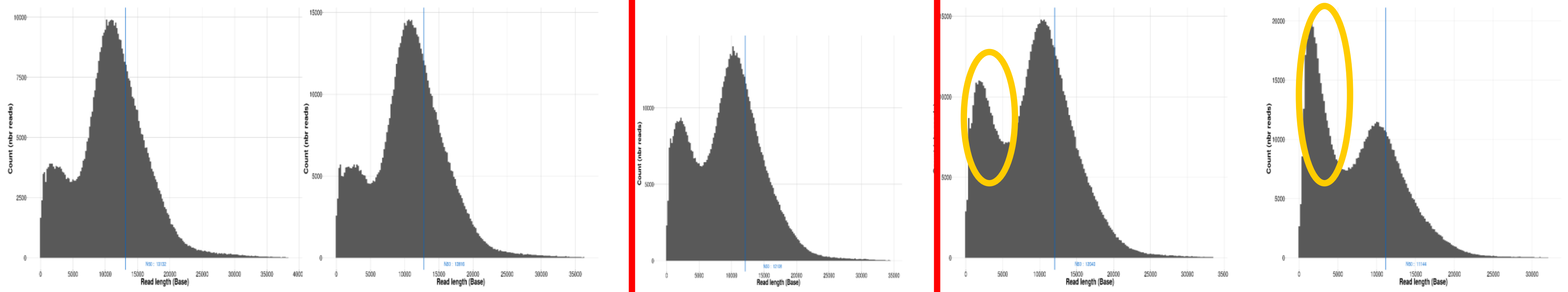
20kb DNA : preferred size

Concentration impact

ONT specification for R9.4.1 : 0,005 pM ➔ 0,05 pM

Flow cell loading test : Loading concentration range

Concentration pM	0,005	0,025	0,05	0,075	0,1
Gbases	6,1	8,5	7,6	8,3	6,3
N50 (kb)	13,15	12,8	12,13	12,06	11,16

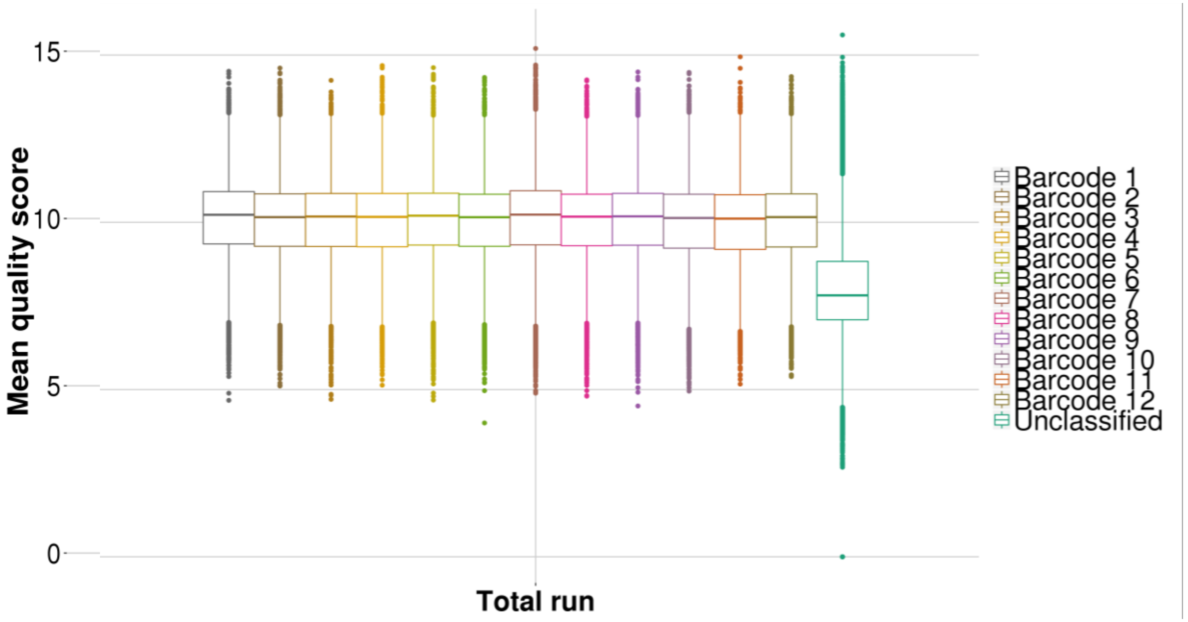


<u>0,05 pM DNA</u> : Due to the results : First option tested
Adaptation in progress

Increase loading ➔ Increase small sequenced reads
➔ Decreases the N50

Multiplex impact

Multisample run to lower the cost



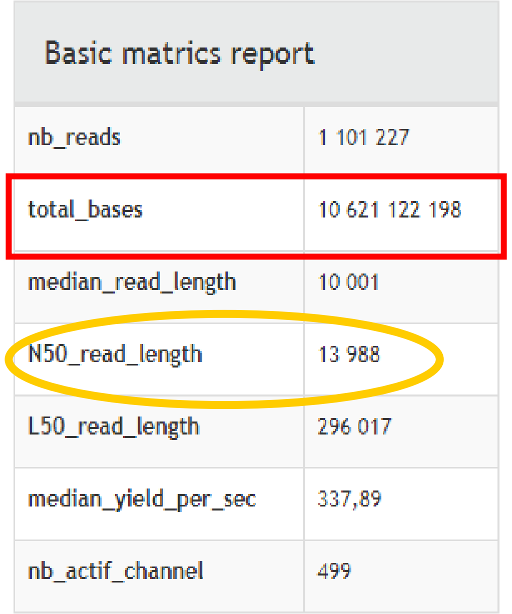
12plex – 20kb shearing

Similar quality between samples : equimolarity
More than 70% of reads assigned to a barcode
30% non assigned reads due to the technology

C : R9.4 FC / 13.4Gb run / Minknow 1.7.3 / Albacore 1.1.0

RevD impact : a new electronical system on the R9.4.1 flow cell

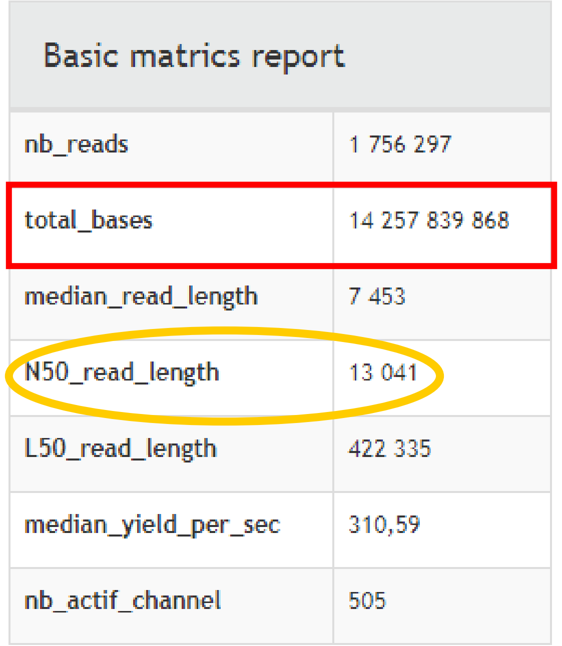
R9.4.1 flowcell



Basic matrics report	
nb_reads	1 101 227
total_bases	10 621 122 198
median_read_length	10 001
N50_read_length	13 988
L50_read_length	296 017
median_yield_per_sec	337,89
nb_actif_channel	499

Quality report	
median_read_quality	10,60
nb_read_Q>5	1 067 916 (96,98%)
nb_read_Q>10	698 914 (63,47%)
total_bases_Q>10	7 217 150 933 (67,95%)
nb_read_Q>15	0 (0,00%)

R9.4.1 _ RevD flowcell



Basic matrics report	
nb_reads	1 756 297
total_bases	14 257 839 868
median_read_length	7 453
N50_read_length	13 041
L50_read_length	422 335
median_yield_per_sec	310,59
nb_actif_channel	505

Quality report	
median_read_quality	10,30
nb_read_Q>5	1 688 028 (96,11%)
nb_read_Q>10	984 860 (56,08%)
total_bases_Q>10	8 944 105 463 (62,73%)
nb_read_Q>15	0 (0,00%)

➔ Increases the amount of data : 10 Gb to 14 Gb
➔ Same N50 and quality

A different ASIC (the electronics on the underside of the flow cell) there is less leakage current on these flow cells compared to the standard ones meaning that each group of pores can be used for longer

First test on our PromethION

- **Similar concentration loaded 0,05 pM**
- **Same library loaded**

	GridION	PromethION
Total bases (Gb)	7,2	65,5
N50 read lenght (kb)	15441	14623
Active pores	1509	7432

⚠

N50 calculated from all sequences :
- More date generated ➔ more small sequences too

Validation in progress :
Same quality and more quantity of data in comparison with the GridION

Remerciements

INRA SCIENCE & IMPACT

Genotoul GeT

Céline Roques, Céline Vandecasteele, Claire Kuchly,
Amalia Sayeh, Roxane Boyer, Maxime Manno,
Cécile Donnadieu , Gérald Salin,
Olivier Bouchez

LIPM Laboratoire Interactions Plantes Micro-organismes

Baptiste Mayjonade, Jérôme Gouzy, Fabrice Roux

Genotoul Bioinfo

Christophe Klopp

Physiologie Génomique Poissons

Yann Guiguen, Elodie Dupin De Beyssat

INP Toulouse ECOLE NATIONALE VETERINAIRE

Guillaume Croville, Jean-luc Guerin

IPS2 Institute of Plant Sciences Paris - Saclay

Abdelhafid Bendahmane

Assembly : Pacbio vs Nanopore

Completeness

Example of a bacterial genome assembly :
Same quantity of data / Same informatics resources

Same contig metrics :
1 contig, 5Mb

Assemblies assesment (BUSCOv2)

Completness based on Bacterial orthologues data base (148 genes)

Assembly sets	Complete genes	Fragmented genes	Missing genes
MinION 1D-CANU	53.4%	18.9%	27.7%
MinION 1D²-CANU	68.9%	14.9%	16.2%
MinION 1D or 1D²-CANU-PILON	95.3%	0	4.7%
RSII-HGAP3	95.3%	0	4.7%

Pilon : Illumina polishing