GeT the latest technology news

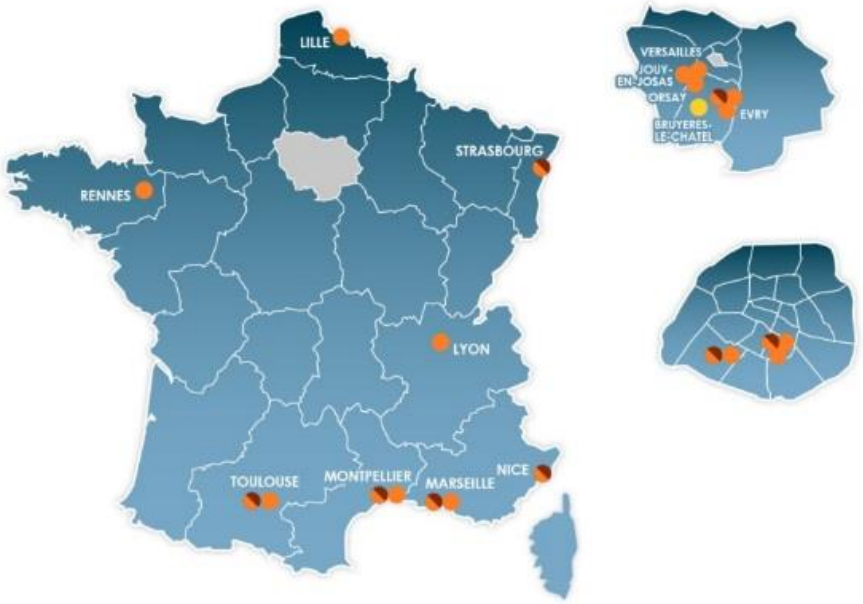
Céline Vandecasteele

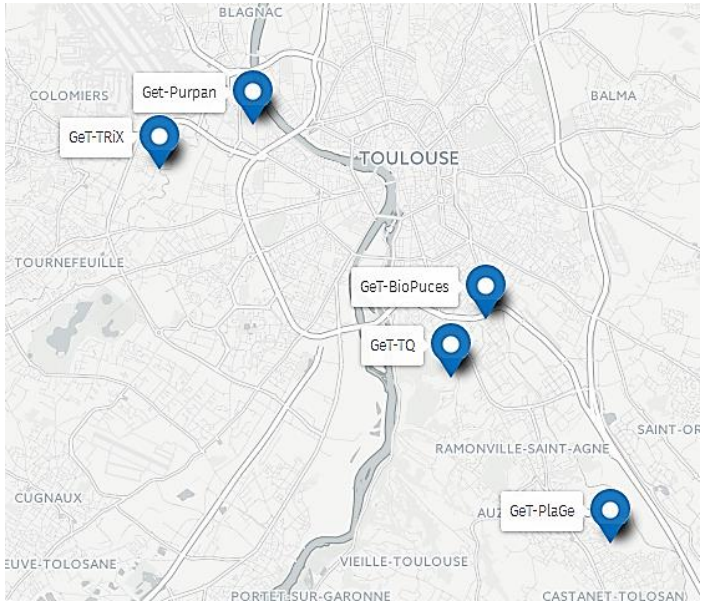
celine.vandecasteele@inra.fr

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

Who are we?

- Genomics and Transcriptomics (GeT) core facility of Genotoul located on 5 sites
- Regional node of National Infrastructure « France Génomique » PIA program
- IBISA Label and INRA strategic core-facility
- Quality certifications ISO9001 & NFX 50 900, Propel (Illumina certified)

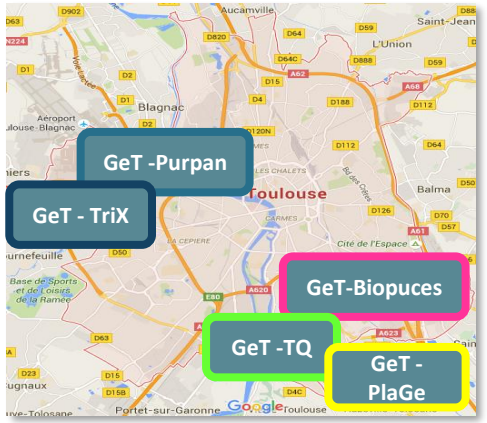




Team, Expertise and missions

35 people on 5 sites

- To provide innovating technologies for genome analysis to the scientific community
  - Sequencing / Genotyping
  - Gene expression
  - Epigenetics
- To Develop new protocols, new methodologies, acquire expertise and train in those technologies



A strong partnership with Genotoul Bioinformatics core-facility

Tools to improve the activity

- Sample and library quality controls



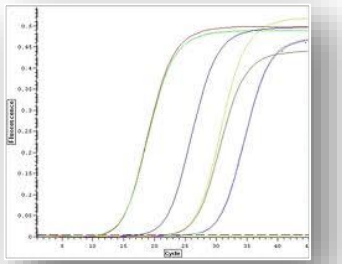
- Pipetting platforms for sample preparation
  - Partnership with Tecan (4 Evo), Agilent Bravo
  - Access array (fluidigm)



- Integration in NG6: Main quality control workflows for NGS data
- Upcoming soon: A new LIMS for NGS samples, sequencing and analysis tracking

Tools to analyse gene expression and genotype

Quantitative PCR

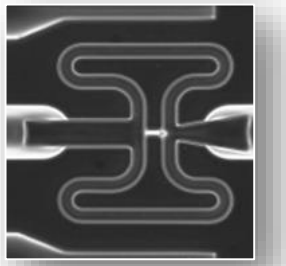


Viia7, QuantStudio, ABI7900HT, …

Q-PCR Microfluidic

BioMark (Fluidigm)

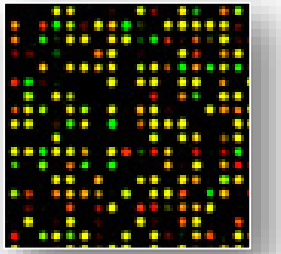
Single Cells



C1 (Fluidigm)

Chromium (10X Genomics)

Microarray



Affymetrix (Agilent)

ddPCR

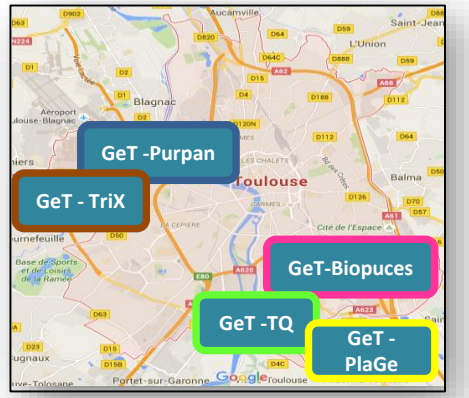
QX200 (Biorad)

A complete portfolio of sequencers at GeT

ABI
800 pb

Ion S5
200 pb
13 Gb

PGM
400 pb
1Gb



MiSeq
2x 300 pb
15 Gb

HiSeq3000
2x150 pb
700 Gb

2018

NovaSeq6000
2x150 pb
3 TB (60h) ?

PacBio RSII
~15 000 pb
1 Gb (6h)

Chromium
10X Genomics
50 000 – 100 000 pb

MinION
~15 000 pb
7 Gb (48h)

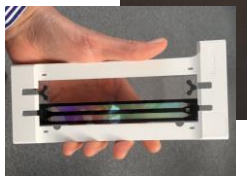
GridION
~15 000 pb
35 Gb (48h)

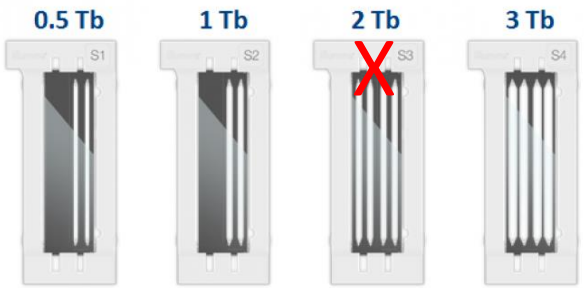
2018

PromethION
25 000 pb?
20Tb (48h)?

Genotoul GeT

NovaSeq 6000

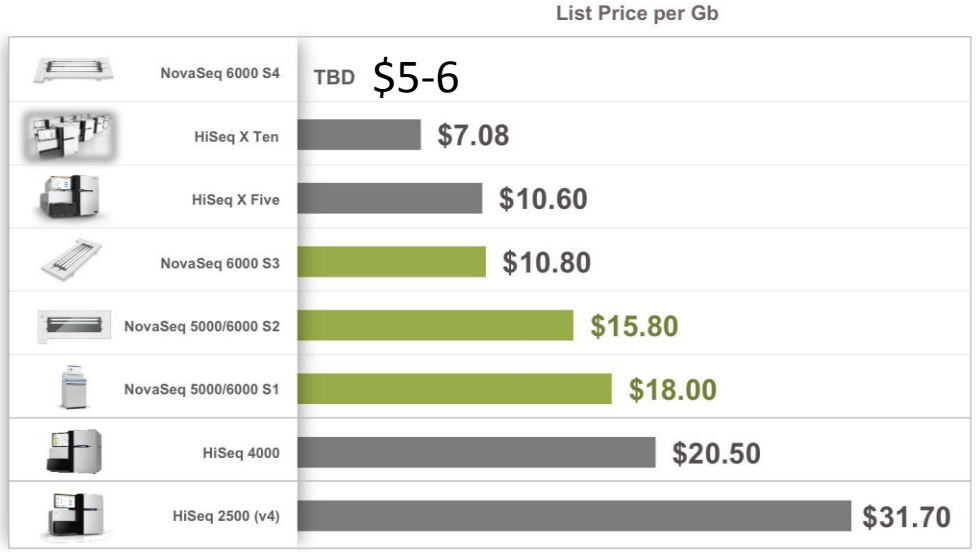




NovaSeq Series

Compelling price per data point enables highly-powered studies

illumina®



9 For Research Use Only. Not for use in diagnostic procedures. illumina

Run times: <1 to ~2.5d based on system, FC and read length

Configure output to match your application and study size

Single flow cell output (1 or 2 can run simultaneously)

Flow Cell Type	NovaSeq 5000	NovaSeq 6000	Reads per Flow Cell	Output (Gb) per Flow Cell: 100 cycles	200 cycles	300 cycles	Output/Run
S4*		✓	10B			3000	> 6 Tb
S3*		✓	6.6B			2000	
S2	✓	✓	3.3 B	333	666	1000	> 2 Tb
S1*	✓	✓	1.6 B	167	333	500	> 1 Tb

Focus on long read sequencing

Pacific Biosciences

PacBio RSII *vs* Sequel

- **PacBio RSII will be turned off at the end of the year**

Average length comparable to RSII
5-10 Gb /SMRT cell
10 hours runs

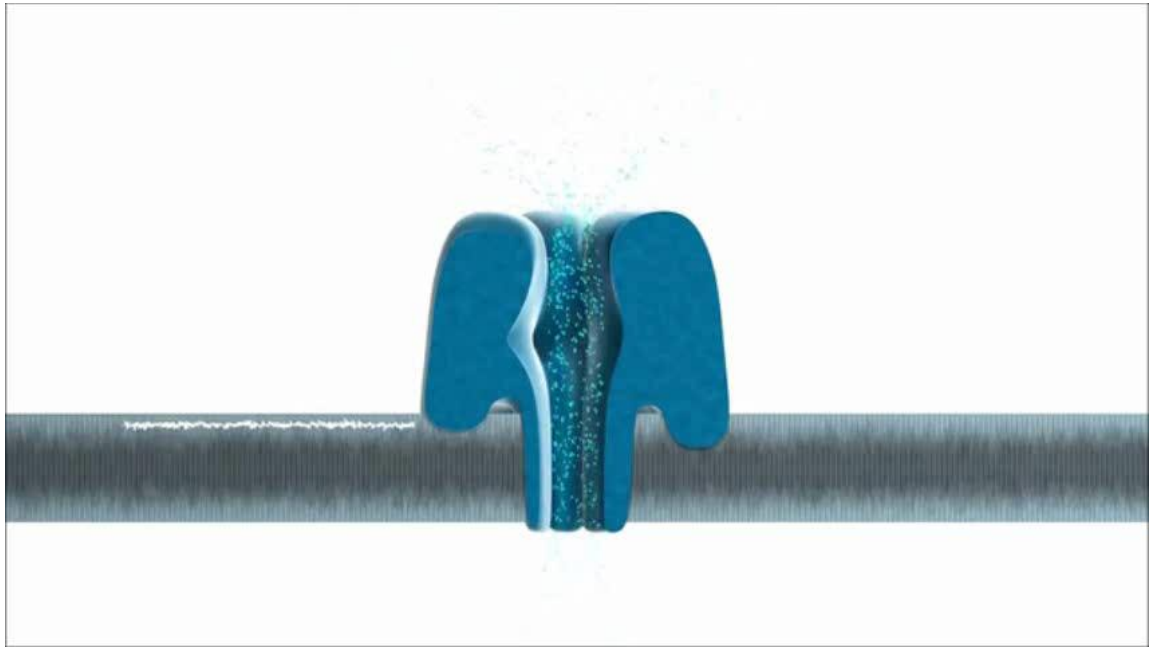
GENTYANE

Possibility to switch projets from GeT-PlaGe's RSII to Gentyane's Sequel (INRA Clermont)

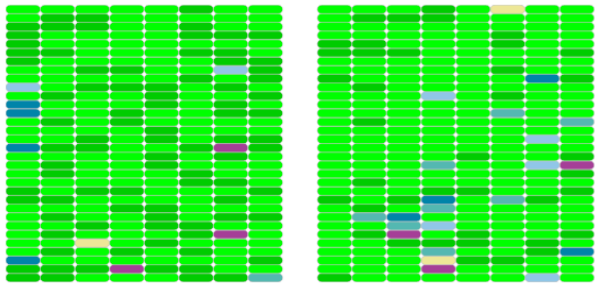
We are currently evaluating the Sequel to transfer some collaborative projects
(*de novo* sequencing, methylation analysis…)

Oxford Nanopore MinION & GridION

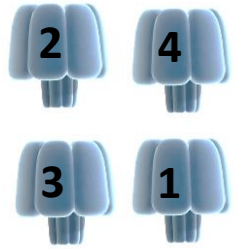
Oxford Nanopore Technology



- 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 > Base calling



512 channels



4 pores per channel

Run time = 48h

Oxford Nanopore Technology

A scalable technology

Only open for collaboration, not service

1 FC

MinION

~ 7Gb / 48h

5 FC

GridION

~ 35Gb / 48h

48 FC

PromethION

~2Tb / 48h ?

2016 → 2017 → 2018

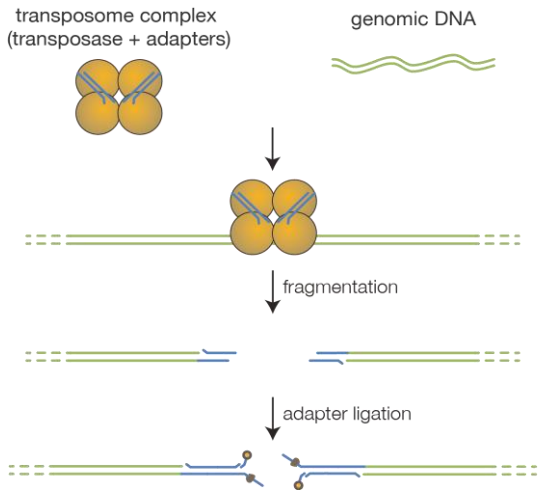
Oxford Nanopore Technology : Library preparation

Input : ~10 µg HMW DNA

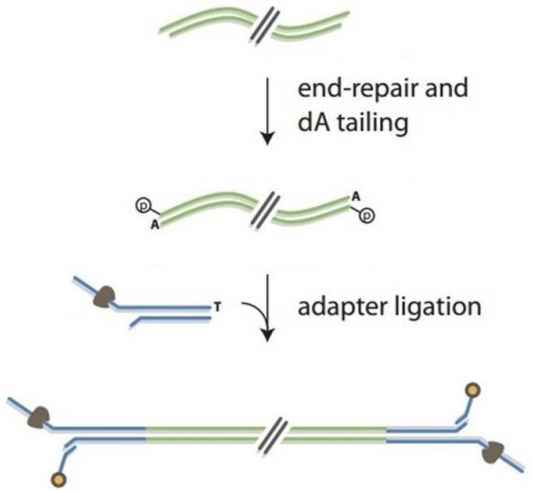
Shearing

Size selection

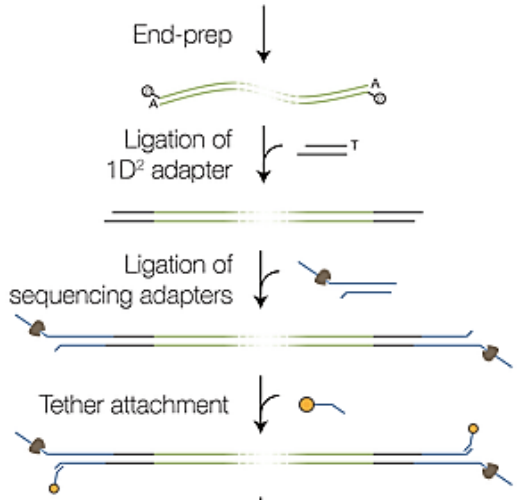
Rapid kit



1D kit



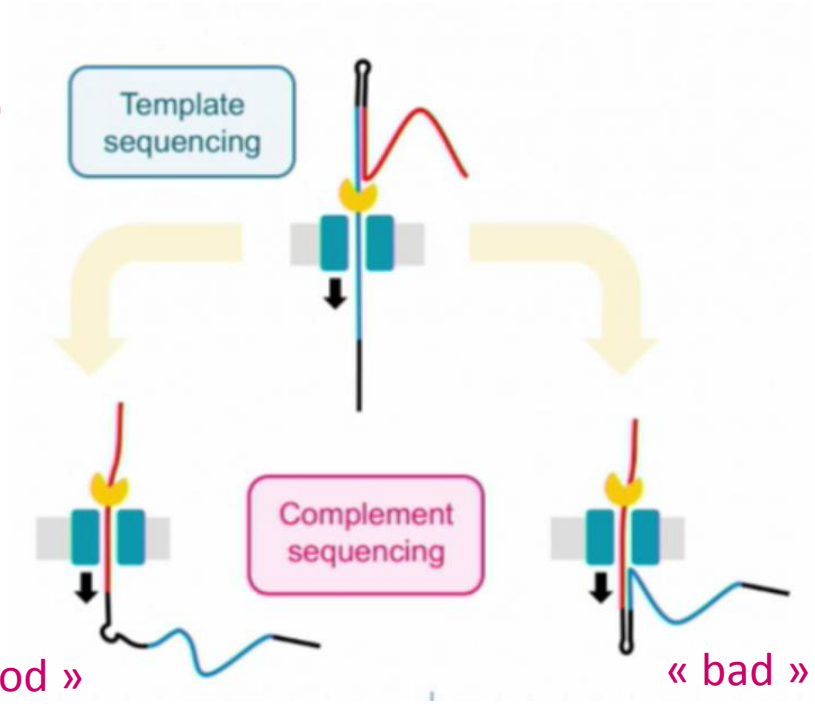
1D² kit

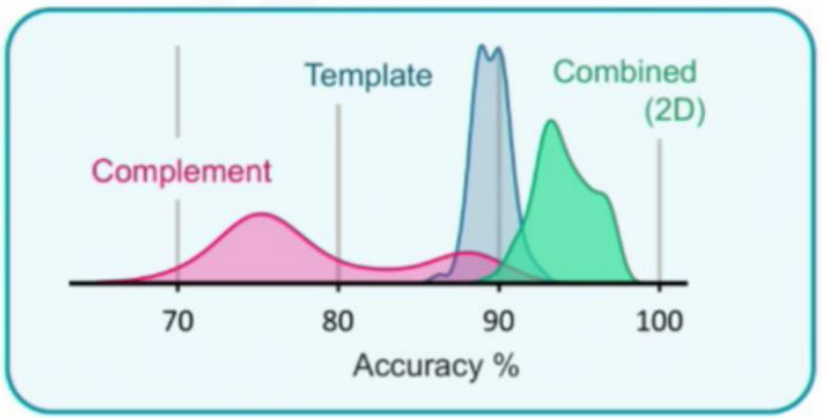


+	Fast lib prep : 10 minutes Very long reads	1-2 days High yield (2-15 Gb / FC)	1-2 days High yield (2-15 Gb / FC) Higher accuracy
–	Low yield (200 – 500 Mb)	large amount of DNA required	Only 30-40% of sequences in 1D² large amount of DNA required

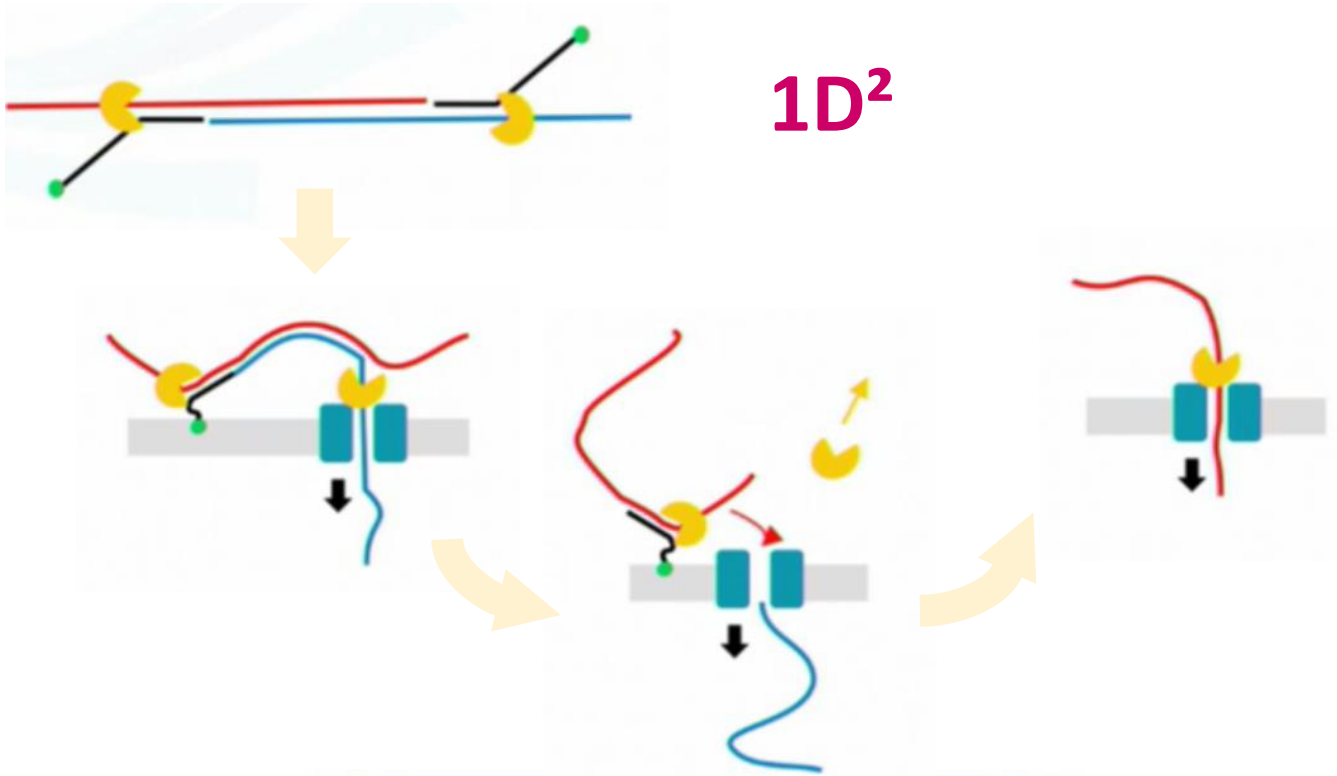
Oxford Nanopore Technology : 2D versus 1D²

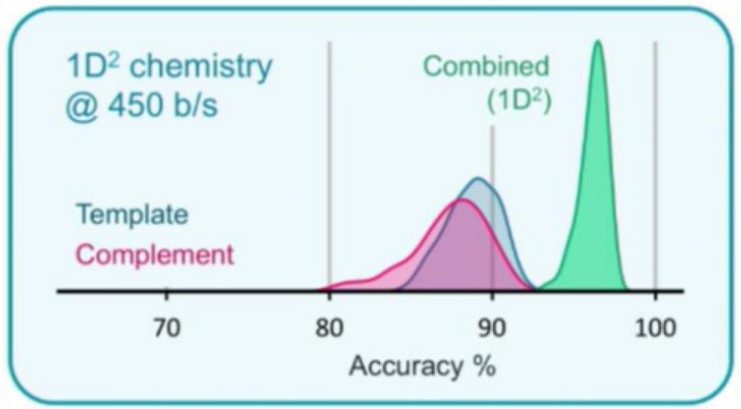
2D





1D²



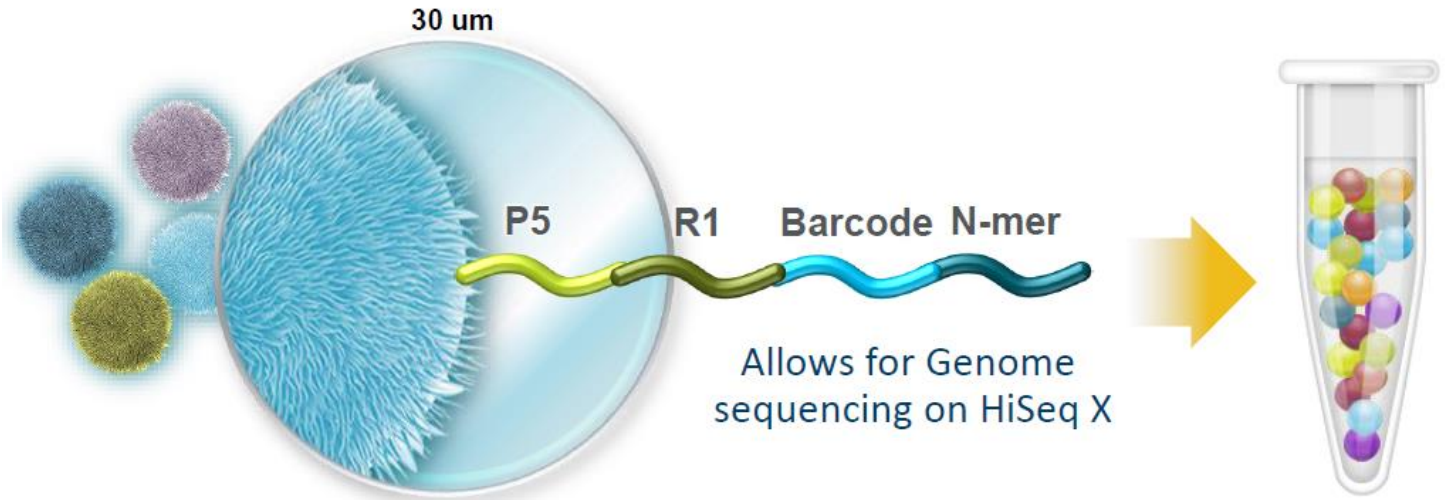


10XGENOMICS Chromium ®

Chromium 10XGENOMICS

Applications

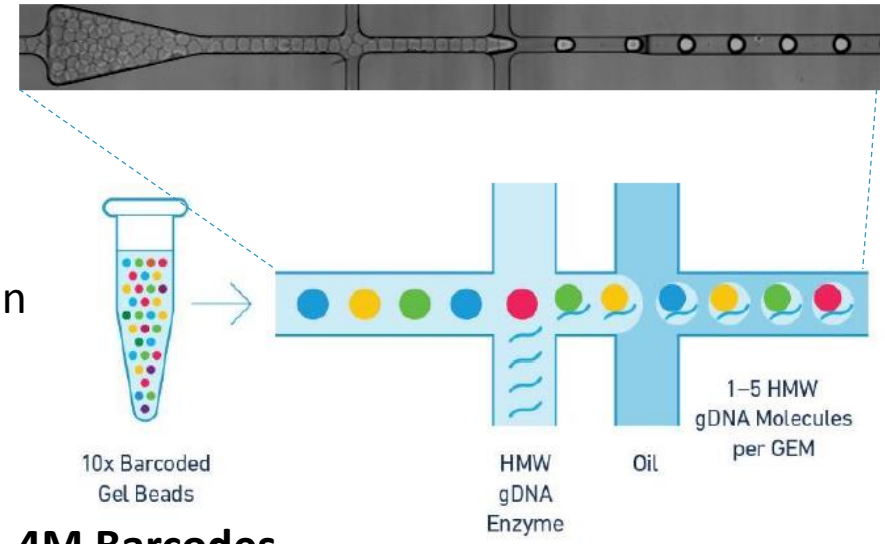
- **Library preparation for Illumina sequencing**
- **Long range genomics (>50 kb), haplotyping/genome phasing, structural variants detection, *de novo* sequencing**
- **Single cell analysis (Profiling 1,000s to 10,000s of cells per experiment increases sensitivity and accuracy for the detection of rare cell types)**
- **Exome sequencing**



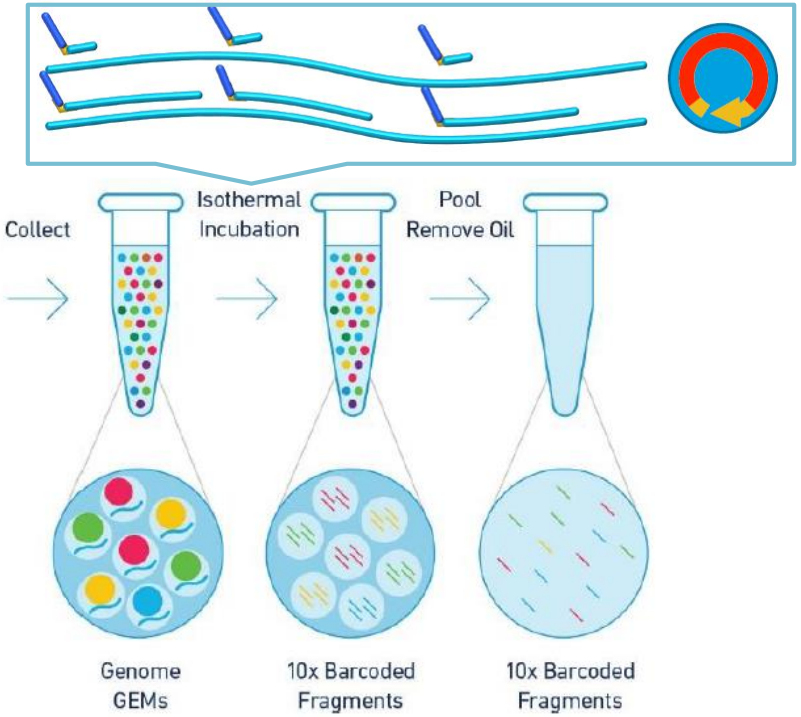
Chromium 10XGENOMICS

How does it work ?

GEM (Gel Bead in Emulsion)



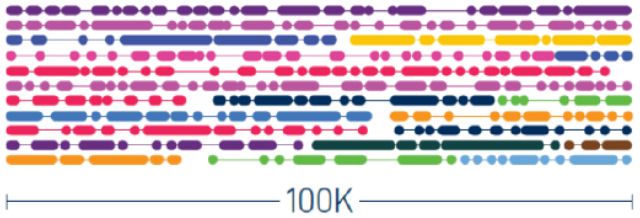
4M Barcodes



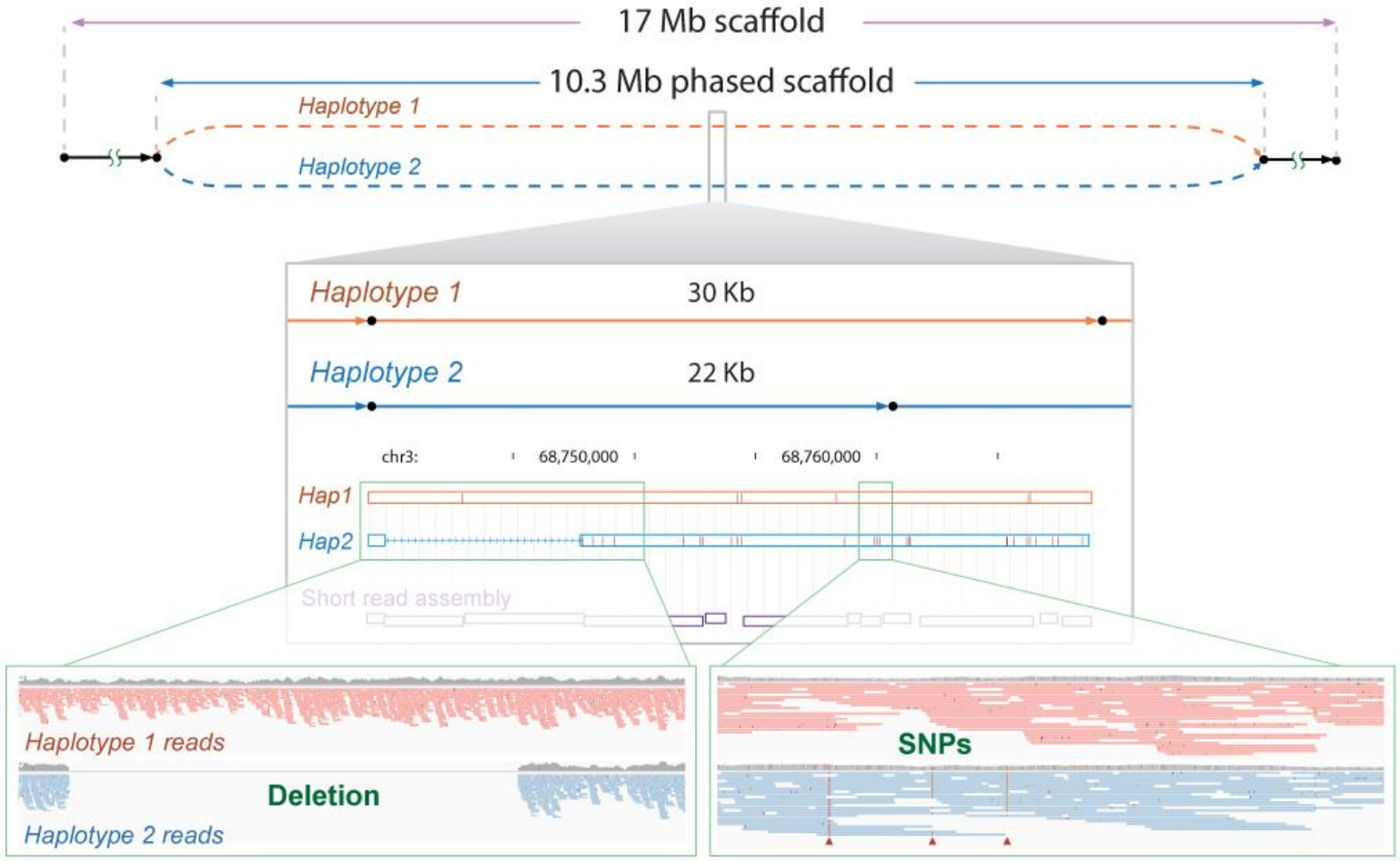
Input : ~1 ng HMW gDNA

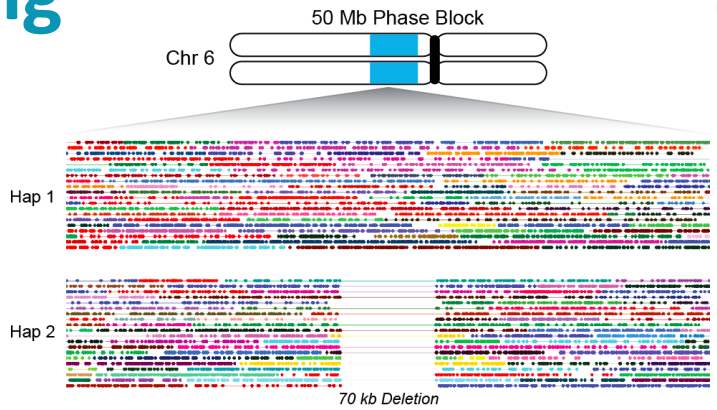
Linked-Reads

Sequencing on HiSeq3000 2x150pb



Advantage of linked-reads : genome phasing



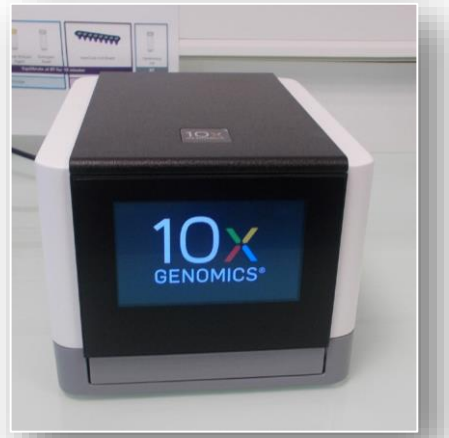


Resolve the Genome Into Multi-Megabase Phase Blocks

Phase the full spectrum of variants (SNVs, indels, and large-scale structural rearrangements) into ultra long multi-megabase phase blocks, enabling a full understanding of diploid genome sequence without the need for a reference.

Chromium limits

- **Development on human genome (3 Gb), useful for other genomes?**
- **Genome size**
  - 500 Mb minimum
- **Improvements to develop for smaller genomes**
- **DNA size**
  - 50 kb minimum, 100 kb for *de novo* assemblies



Genotoul GeT

To summarize and help to select a Long Read Technology

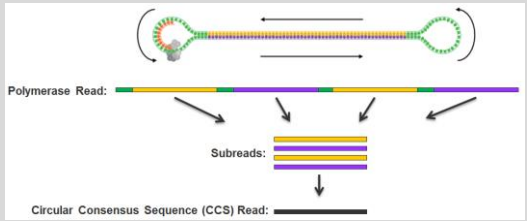
For Whole genome sequencing

SPECIFICATIONS

	PACIFIC BIOSCIENCES RSII	PACIFIC BIOSCIENCES Sequel	Oxford NANOPORE Technologies	10X GENOMICS / illumina
HMW gDNA quantity	~10µg 1lib/~10SM	~10µg 1lib/~30SM	~10µg 1lib/1FC	~1 ng 1 lib/~10 lanes (only for Genome size > 500 Mb)
Multiplexing (plex recommended)	4plex	6plex	12plex (1D) Very similar quality / data quantity between samples	-
Yield	0,5-2 Gb/SM	5-10 Gb/SM	5-10 Gb/FC (only 30-40% 1D²)	90 Gb/lane
# Reads (average)	~70 000	~500 000	450 000	600 M/lane
Read length (average)	~15 kb	~13 kb	~15 kb	50-100 kb linked-reads
Lib prep time	4 days		2 days (1D,1D²)	2-4 days
Run prep time	1 day		30 min	2,5 days
Run time	6h/SM	10h/SM	48h	3,5 days
Primary Errors	Indel (random)		deletions (no random in homopolymer)	substitution
Single-pass Error Rate (%)	~13 %		~13 % (1D), ~9% (1D²)	~0,1 (Illumina)
Final Error Rate (%)	≤1		≤1 (corrected by Illumina data)	~0,1 (Illumina)
Price (average)	+++	++	+ (+)	-

To summarize and help to select a Long Read Technology

For other applications

	PACIFIC BIOSCIENCES RSII / Sequel	Oxford NANOPORE Technologies	10X GENOMICS / illumina
EPIGENETICS	Direct methylation detection of 6mA and 4mC, not 5mC (but new algorithm for CpG sites) User-friendly analyses, in particular detects the sequence context of sites	Tools : Nanopolish and Nanoraw *In progress*	
METAGENOMICS	Full length 16S with ccs algorithm (PacBio tools : user-friendly analyses) + : multiple passes around a circular template 	Difficult with error rate ~9% (1D²)	
TRANSCRIPTOMICS	Reference transcriptome Isoform reconstruction	Reference transcriptome Direct RNA sequencing	Differential gene expression on a cell-by-cell level

Hybrid strategies to obtain genomes

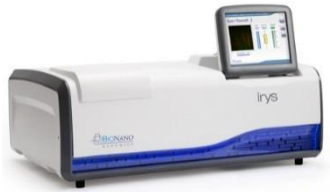
PacBio + 10X Genomics improve genome assembly

	Technology	Assembly analysis	Cov (X)	Tot bases (Mb)	# contigs	N50	L50	Completeness (BUSCO V2)
Fish genome (~1Gb)	10X	Supernova	78	818	45 319	1,1 Mb	157	82,4 %
	PacBio	(Pre-correction Canu) Smartdenovo	71	808	701	4,1 Mb	55	88,1 %
	PacBio + 10X	ARCS (PacBio Smartdenovo assembly + 10X)		808	534	5,4 Mb	44	88,1 %
	PacBio	Canu	71	1 015	4 062	1,3 Mb	126	91 %
	PacBio + 10X	ARCS (PacBio Canu assembly + 10X)		1015	3 583	1,6 Mb	102	91 %
Tomato genome (~800 Mb)	10X	Supernova	87	795	21 619	2,2 Mb	105	90,5 %
	PacBio	(Pre-correction Canu) Smartdenovo	81	768	857	2,1 Mb	112	92 %
	PacBio + 10X	ARCS (PacBio Smartdenovo assembly + 10X)		768	416	4,1 Mb	58	92 %
	PacBio	Canu	81	792	508	4,9 Mb	47	94 %
	PacBio + 10X	ARCS (PacBio Canu assembly + 10X)		792	284	13,6 Mb	19	94,2 %

Thanks to **Anaïs Poiraudeau**, Claire Kuchly, Christophe Klopp, Yann Guiguen lab and Mohammed Zouine lab

+ Bionano for a better Tomato Genome assembly

Chromium (10 X Genomics)



Irys (Bionano)

Technology	N50
PacBio (RSII 70 x)	3.2 Mb
+ Bionano (2 enzymes)	32 Mb
+ Chromium + Illumina (100x)	45 Mb

Pilot projects on Tomato for de novo assembly

Technology	N50
Chromium + Illumina (100x)	1.8 Mb
+ Bionano (2 enzymes)	17 Mb

GBF Lab : Mohammed Zouine, Pierre Frasse, Mondher Bouzayen
CNRGV lab (Bionano) : Sandrine Arribat, William Marrande, Hélène Bergès

For further informations

November 28th
Live transmission

https://seminaire.inra.fr/long-reads-dream-or-reality/Program

Long reads : Dream or Reality

Presentation of expertise and results obtained by research teams working on the latest technologies available on GeT

2017, November 28th • INRA Get-PlaGe, 24 Chemin de Borde Rouge 31326 Castanet-Tolosan

- 09h00 - 09h30 : Coffee
- 09h30 - 09h45 : **GeT Strategy**, Denis Milan - Genome & Transciptime (GeT) core facility
- 09h45 - 10h15 : **Implementation and Evaluation of Oxford Nanopore MinION and GridION sequencing**, Catherine Zanchetta & Maxime Manno – GeT-PlaGe, INRA, Genotoul, US1426
- 10h15 - 10h45 : **Minion Sequencing Provides New Insight On The Evolutionary History Of Seabird Mitochondrial Genomes**, Lucas Torres - Littoral Environnement et Sociétés (LIENSs) UMR 7266
- 10h45 - 11h15 : **Direct whole genome sequencing of avian poxvirus using Nanopore MinION**, Guillaume Croville - Université de Toulouse, UMR 1225, INRA/ENVT
- 11h15 - 11h45 : **INVITED SPEAKER - De novo assembly of teleost fishes using PacBio sequencing data: What is gained?**, Ole Kristian Tørresen - Centre for Ecological and Evolutionary Synthesis, OSLO
- 11h45 - 13h15 : Lunch
- 13h15 - 13h45 : **Implementation and Evaluation of Chromium technology**, Olivier Bouchez & Claire Kuchly - GeT-PlaGe, INRA, Genotoul, US1426
- 13h45 - 14h15 : **Phasing Haplotypes in Rabbit using Long Reads Technology**, Julie DEMARS - Génétique Physiologie et Systèmes d'Elevage (GenPhySE), INRA, UMR 1388
- 14h15 - 14h45 : **High-quality de novo genome assembly of the tomato genome using the latest longs reads sequencing and optical mapping technologies**, Mohamed Zouine - Laboratoire Génomique et Biotechnologie du Fruit GBF, UMR990, INRA/INP-ENSAT
- 14h45 - 15h15 : **Two examples of hard to assemble genomes, even with 3rd generation sequences**, Christophe Klopp - Unité de Mathématiques et Informatique Appliquées de Toulouse (MIAT) INRA
- 15h15 - 15h45 : Coffee
- 15h45 - 16H15 : **Not SMRT yet smart: 3D genomics with Hi-C sequencing**, Sylvain Foissac , Génétique Physiologie et Systèmes d'Elevage (GenPhySE), INRA, UMR 1388
- 16h15 - 16h45 : **Comparison of methylome profiles between closely related clones of the bacterial plant pathogen *Ralstonia solanacearum***, Alice Guidot - Laboratoire des Interactions Plantes Micro-organismes (LIPM) UMR CNRS-INRA 2594/441
- 16h45 - 17h15 : **Diversity of HEV genotype 3 based on full-length sequences**, Florence Nicot - Pôle Biologie IFB Hôpital Purpan

New Website !

http://get.genotoul.fr

@GeT_Genotoul

get@genotoul.fr

Thanks to Get team and all partners !!!