Implementation and Evaluation of 10X Genomics Chromium technology

Claire Kuchly & Olivier Bouchez

28/11/2017

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

Chromium evaluation: pilot phase

- **Platform installed in november 2016**
- **Training: november 15th 2016**
- **3 pilot projects: Rabbit, Tomato & Fish**

Chromium 10XGENOMICS

- Library preparation for Illumina sequencing
- Long range genomics (>50 kb), haplotyping/genome phasing, structural variants detection, *de novo* sequencing
- Single cell analysis (500->10000)
- Exome sequencing

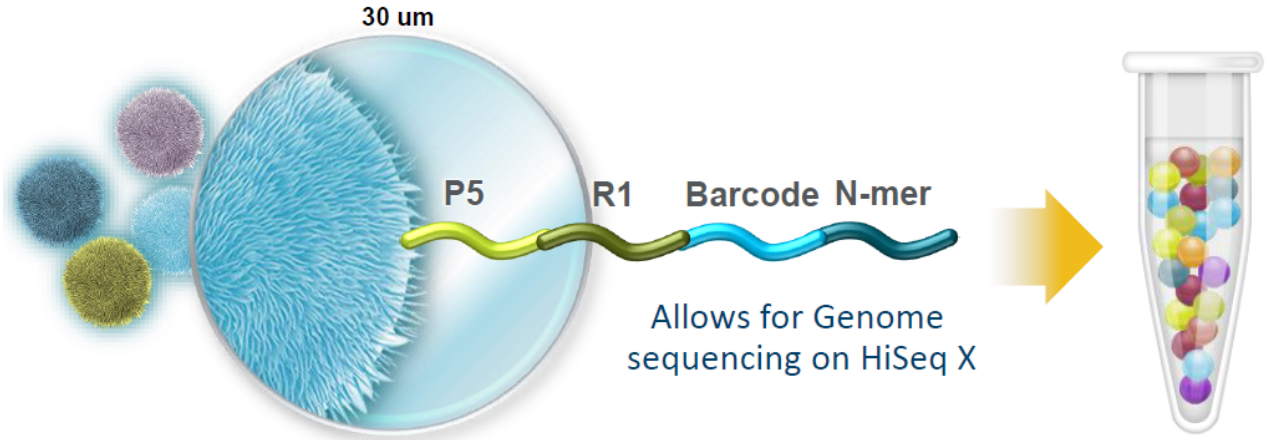
Chromium

HiSeq3000

NovaSeq

Chromium 10XGENOMICS

How does it work ?

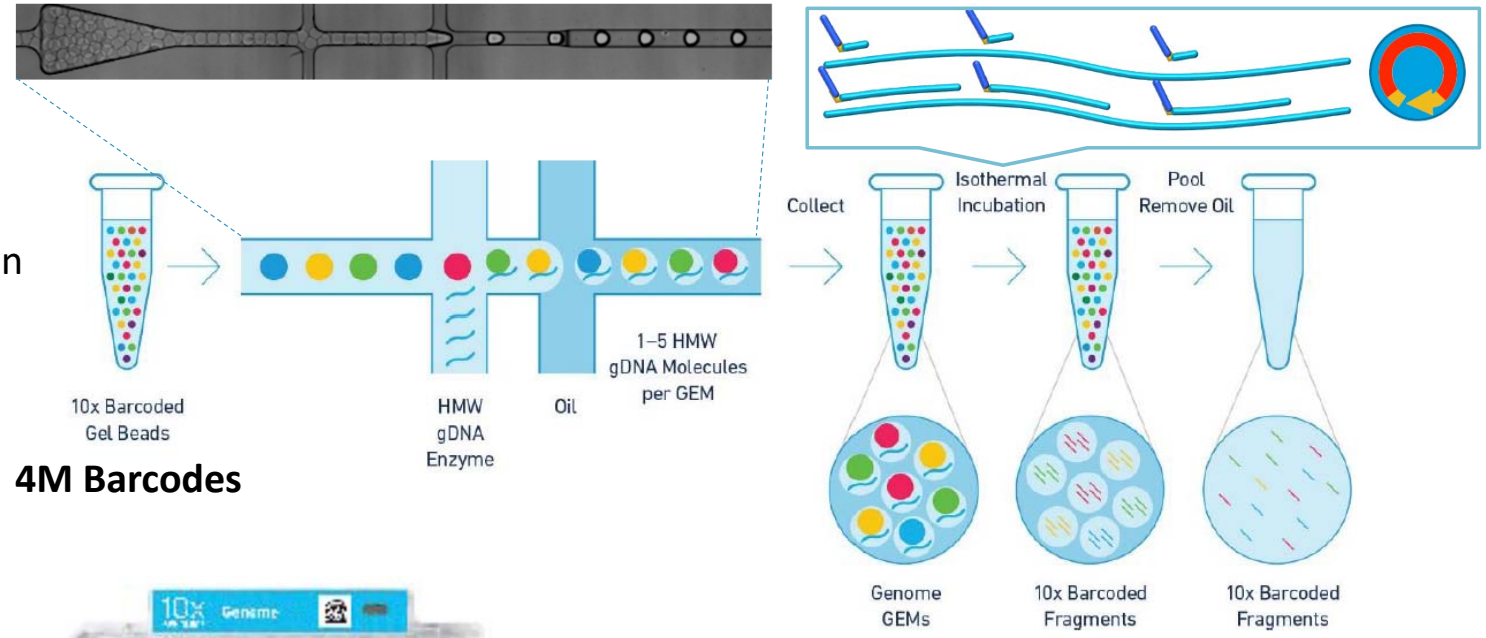


Chromium 10XGENOMICS

How does it work ?

Complete wetlab workflow: 2 days

GEM (Gel Bead in Emulsion)

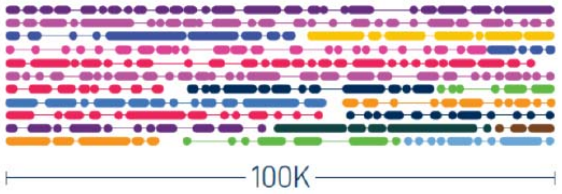


4M Barcodes

Input: ~1 ng HMW gDNA

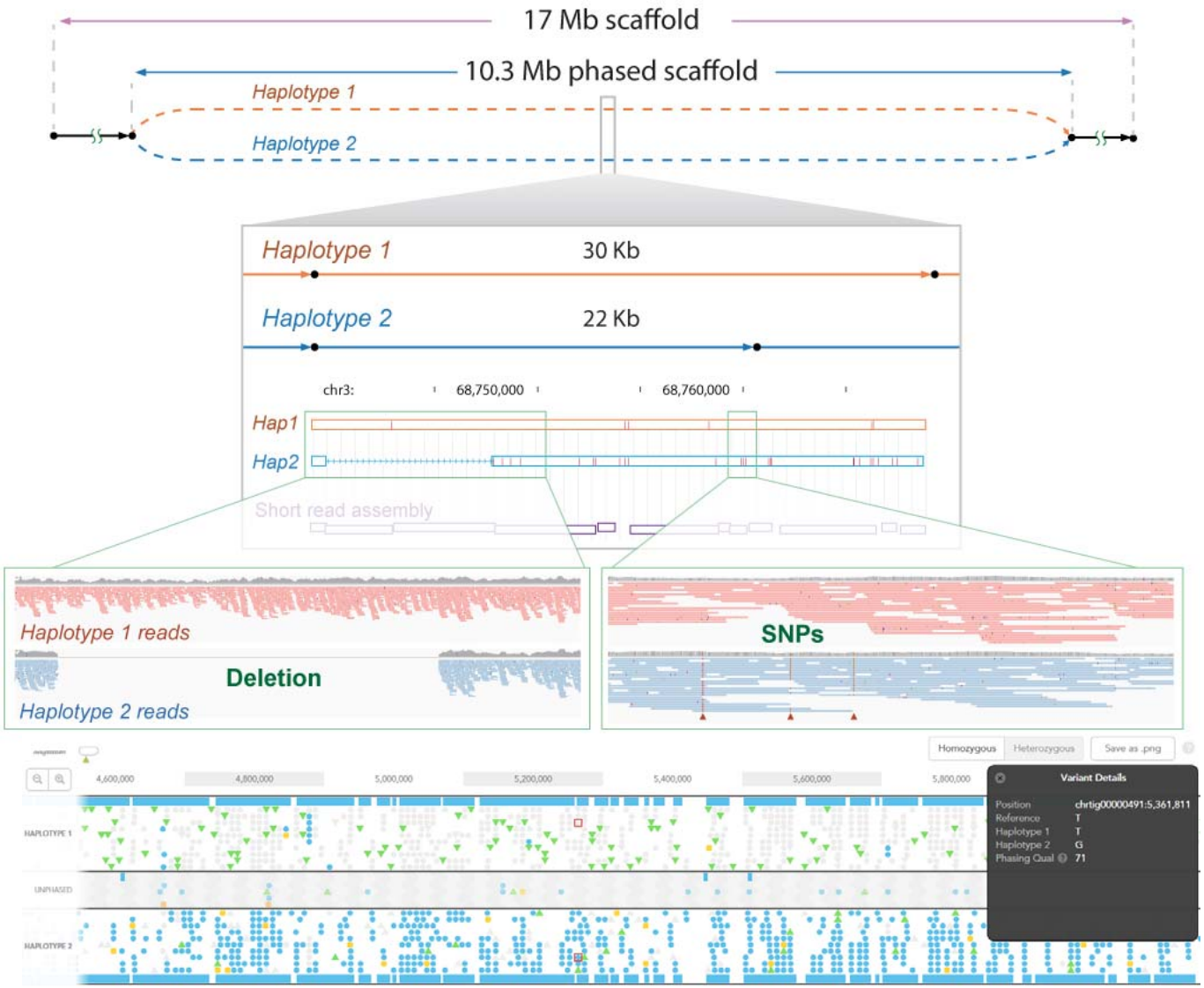
Linked-Reads

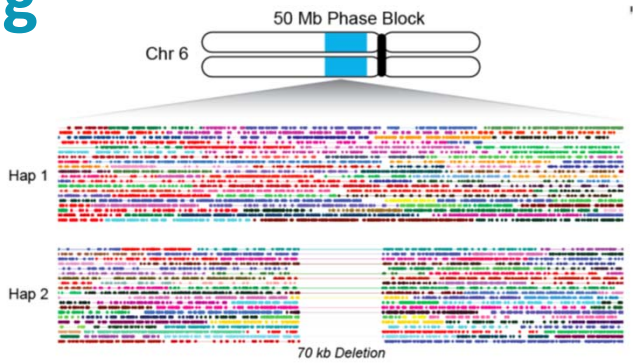
Sequencing on HiSeq3000 2x150pb



Genotoul GeT

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.

Genotoul GeT

Haplotyping using Chromium

Results via Loupe

- **Linked-reads and structural variant calling**

Single cell analysis

Estimated Number of Cells
552

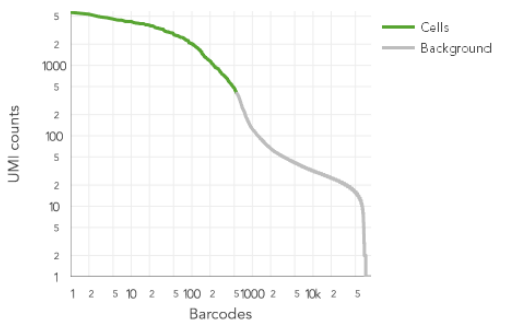
Mean Reads per Cell
650,811

Median Genes per Cell
585

Sequencing

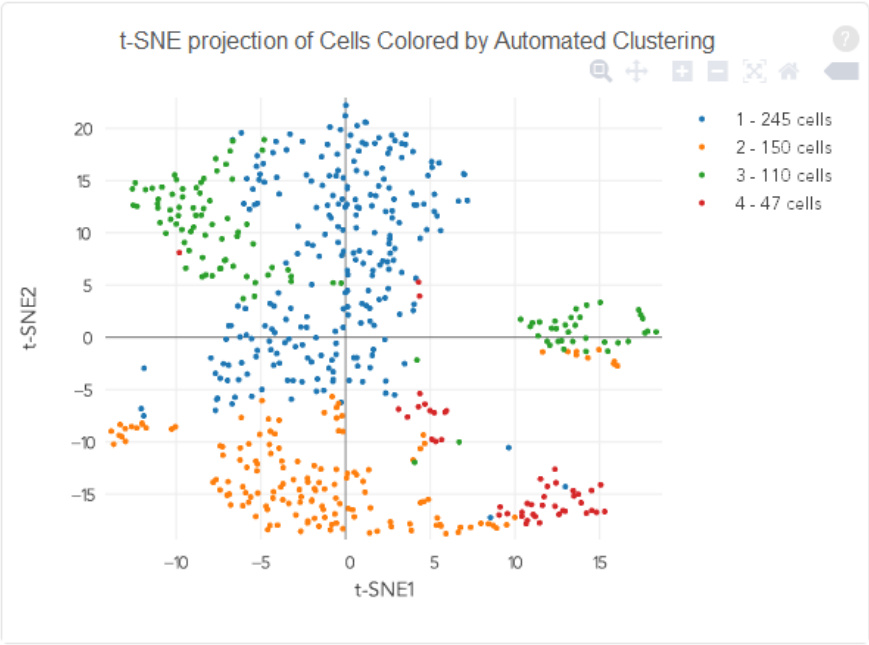
Number of Reads	359,248,007
Valid Barcodes	98.2%
Reads Mapped Confidently to Transcriptome	60.0%
Reads Mapped Confidently to Exonic Regions	63.6%
Reads Mapped Confidently to Intronic Regions	13.6%
Reads Mapped Confidently to Intergenic Regions	8.8%
Sequencing Saturation	98.9%
Q30 Bases in Barcode	98.7%
Q30 Bases in RNA Read	59.3%
Q30 Bases in Sample Index	89.5%
Q30 Bases in UMI	98.8%

Cells



Estimated Number of Cells	552
Fraction Reads in Cells	34.8%
Mean Reads per Cell	650,811
Median Genes per Cell	585
Total Genes Detected	14,168
Median UMI Counts per Cell	887

Sample



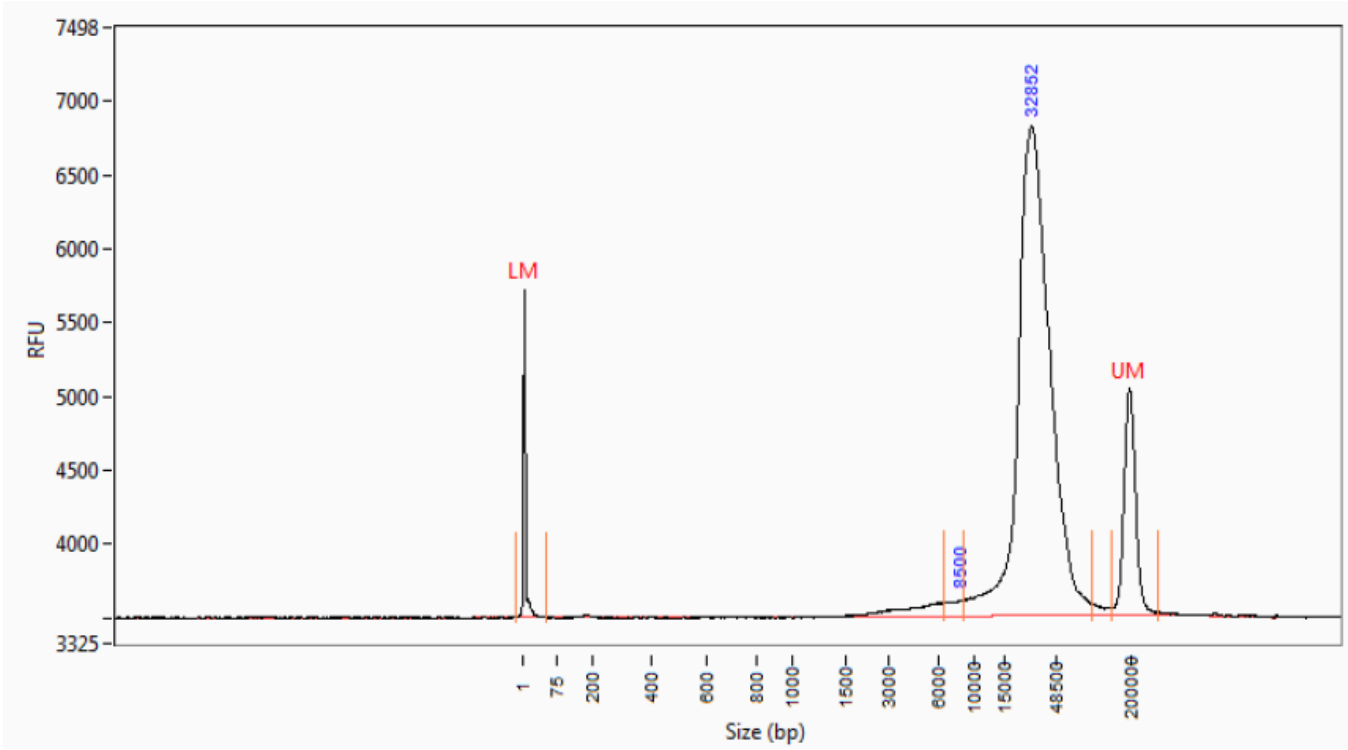
Chromium limits for genome sequencing

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

Haplotyping using Chromium

DNA prep, without BluePippin

- **Young rabbit = low amount of DNA**
- **Molecular weight assessed by Fragment Analyzer**



Haplotyping using Chromium

Results via Loupe software

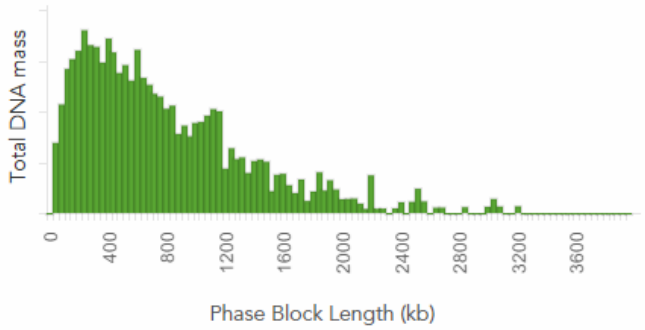
Without BluePippin sizing

GEM Performance

GEMs Detected	1,636,253
N50 Linked-Reads per Molecule (LPM)	13.0
Mean DNA per GEM	600,199 bp

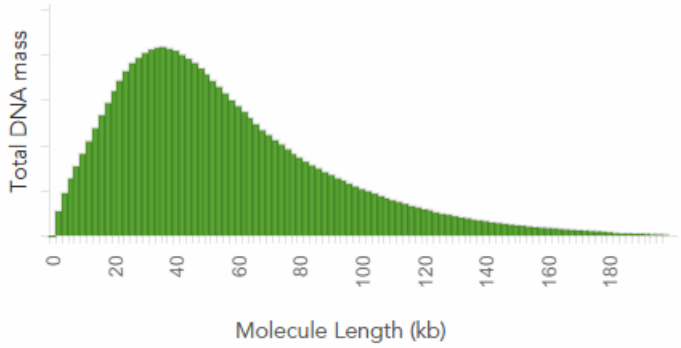
Phasing

SNPs Phased	97.1%
Longest Phase Block	3,241,488 bp
N50 Phase Block	632,948 bp
Phase Block View	



Input DNA

Molecule Length	μ 44,559 bp
DNA in Molecules >20kb	85.5%
DNA in Molecules >100kb	11.8%
Corrected Estimated of DNA Loaded	1.41 ng



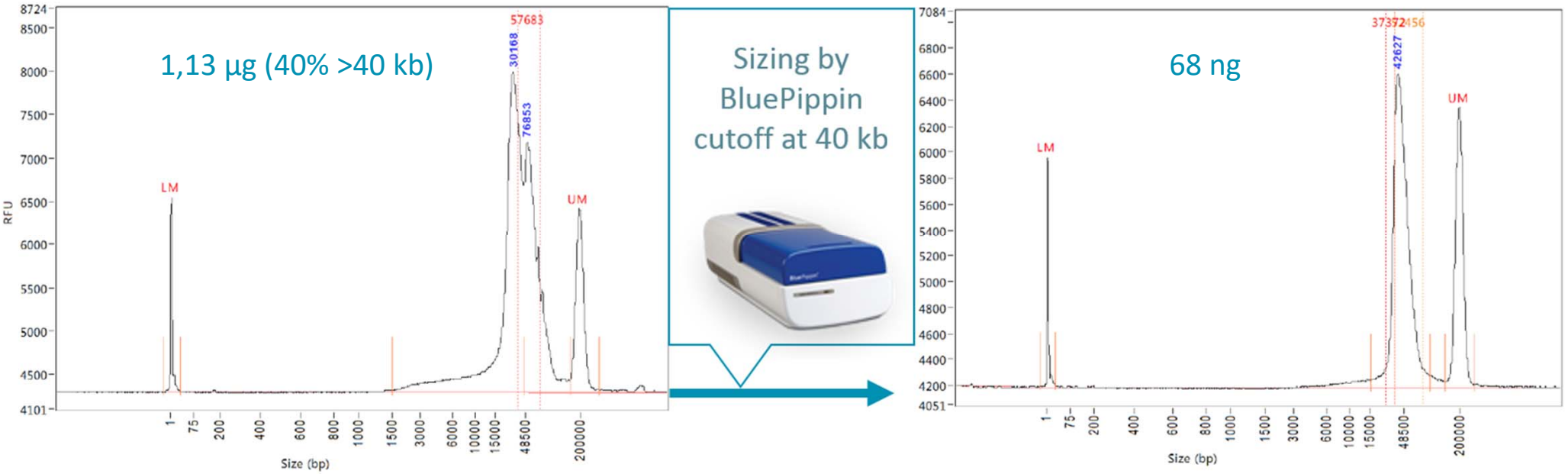
Structural Variants

Large Structural Variant Calls	117
Short Deletion Calls	47,585

Haplotyping using Chromium

DNA prep

- **Rabbit DNA from 2 parents & 2 babies – Genome ~ 2 Gb**
- **Extraction by MagAttract (Qiagen) from blood, molecular weight assessed by Fragment Analyzer**



Genotoul GeT

Haplotyping using Chromium

With BluePippin sizing : Results via Loupe software

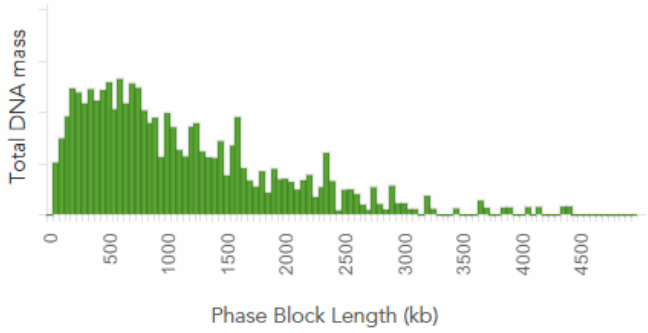
With BluePippin sizing

GEM Performance

GEMs Detected	1,593,499
N50 Linked-Reads per Molecule (LPM)	19.0
Mean DNA per GEM	576,954 bp

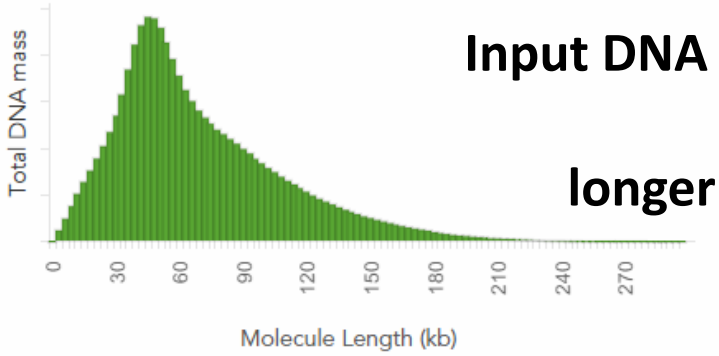
Phasing

SNPs Phased	97.2%
Longest Phase Block	4,417,547 bp
N50 Phase Block	875,494 bp
Phase Block View	



Input DNA

Molecule Length	μ 63,172 bp
DNA in Molecules >20kb	93.2%
DNA in Molecules >100kb	20.8%
Corrected Estimated of DNA Loaded	1.35 ng



Input DNA size improvment

=

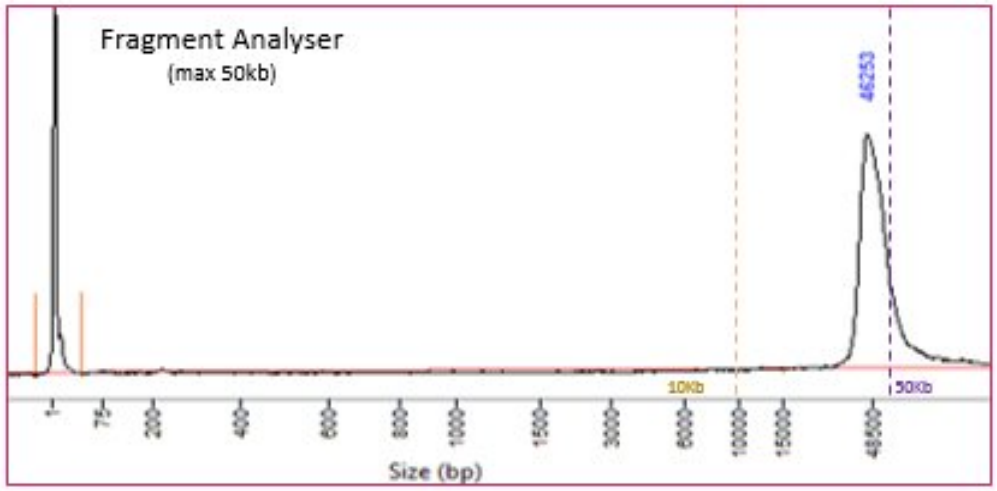
longer phase block

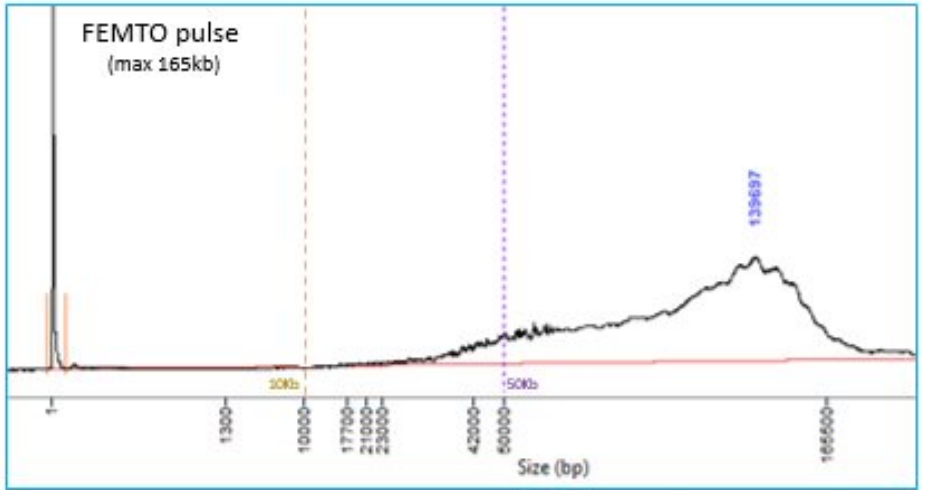
Structural Variants

Large Structural Variant Calls	162
Short Deletion Calls	49,846

Genotoul GeT

DNA QC: New Femto Pulse





Same DNA QC on Fragment Analyzer and Femto Pulse => better resolution

10X Applications

Our 10X Applications

Genome & Exome solution

Reference-Based
Workflow

Long Ranger
Analysis Pipelines

Loupe
Genome Browser

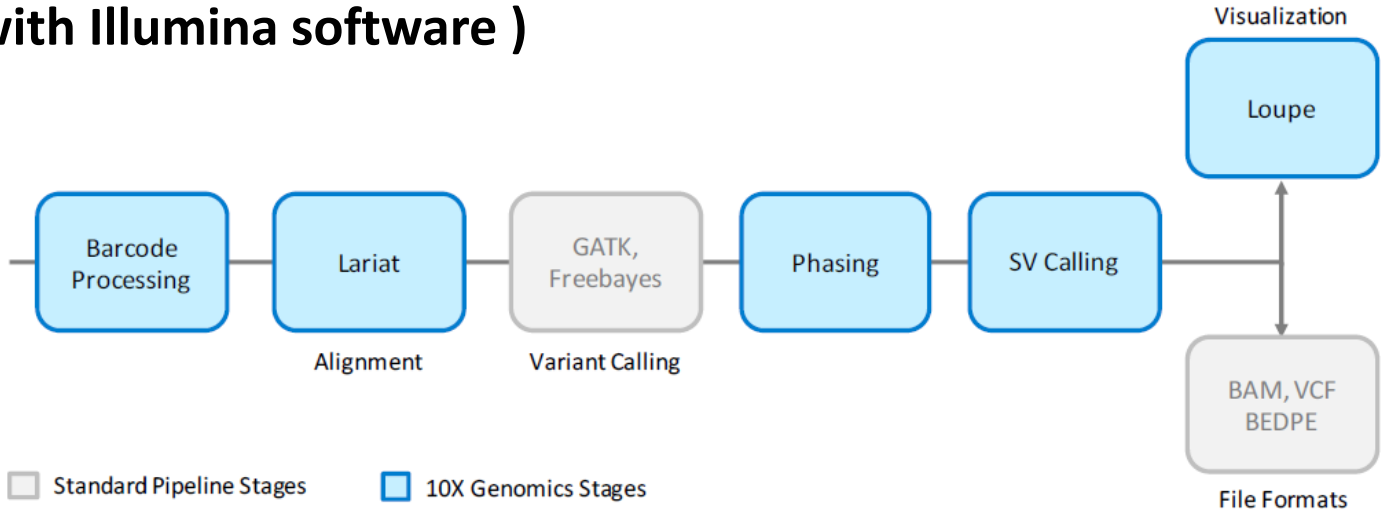
Linked-Reads analysis software

Genotoul GeT

LongRanger software

Analysis pipelines that perform :

- **Sample demultiplexing (with Illumina software)**
- **Barcode processing**
- **Alignment (Lariat)**
- **Quality control**
- **Variant calling**
- **Phasing**
- **Structural variant calling**



Software with few parameters for the analysis
Not easy to be run on our cluster

Sequencing Quality Control

Longest Phase Block	GEMs Detected
7,677,566	1,306,663

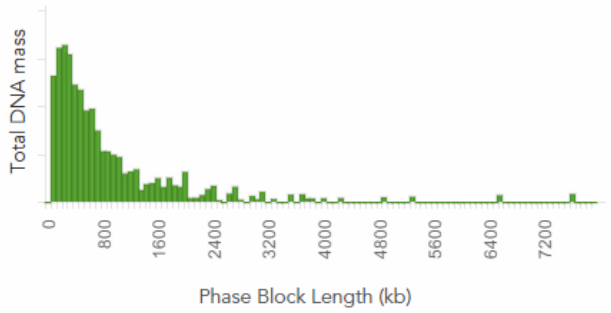
GEM Performance

GEMs Detected	1,306,663
N50 Linked-Reads per Molecule (LPM)	19.0
Mean DNA per GEM	644,628 bp

Phasing

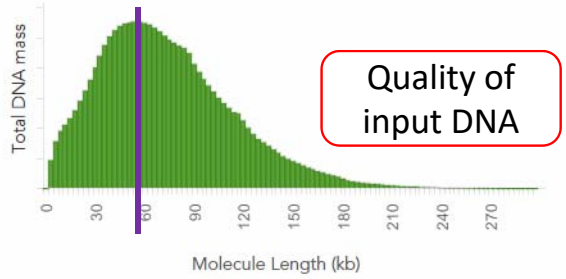
SNPs Phased	93.2%
Longest Phase Block	7,677,566 bp
N50 Phase Block	483,337 bp

Phase Block View



Input DNA

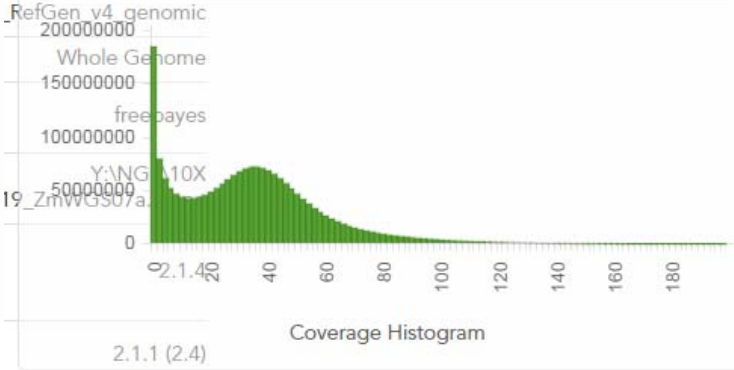
Molecule Length	μ 58,066 bp
DNA in Molecules >20kb	89.9%
DNA in Molecules >100kb	23.5%
Corrected Estimated of DNA Loaded	1.51 ng



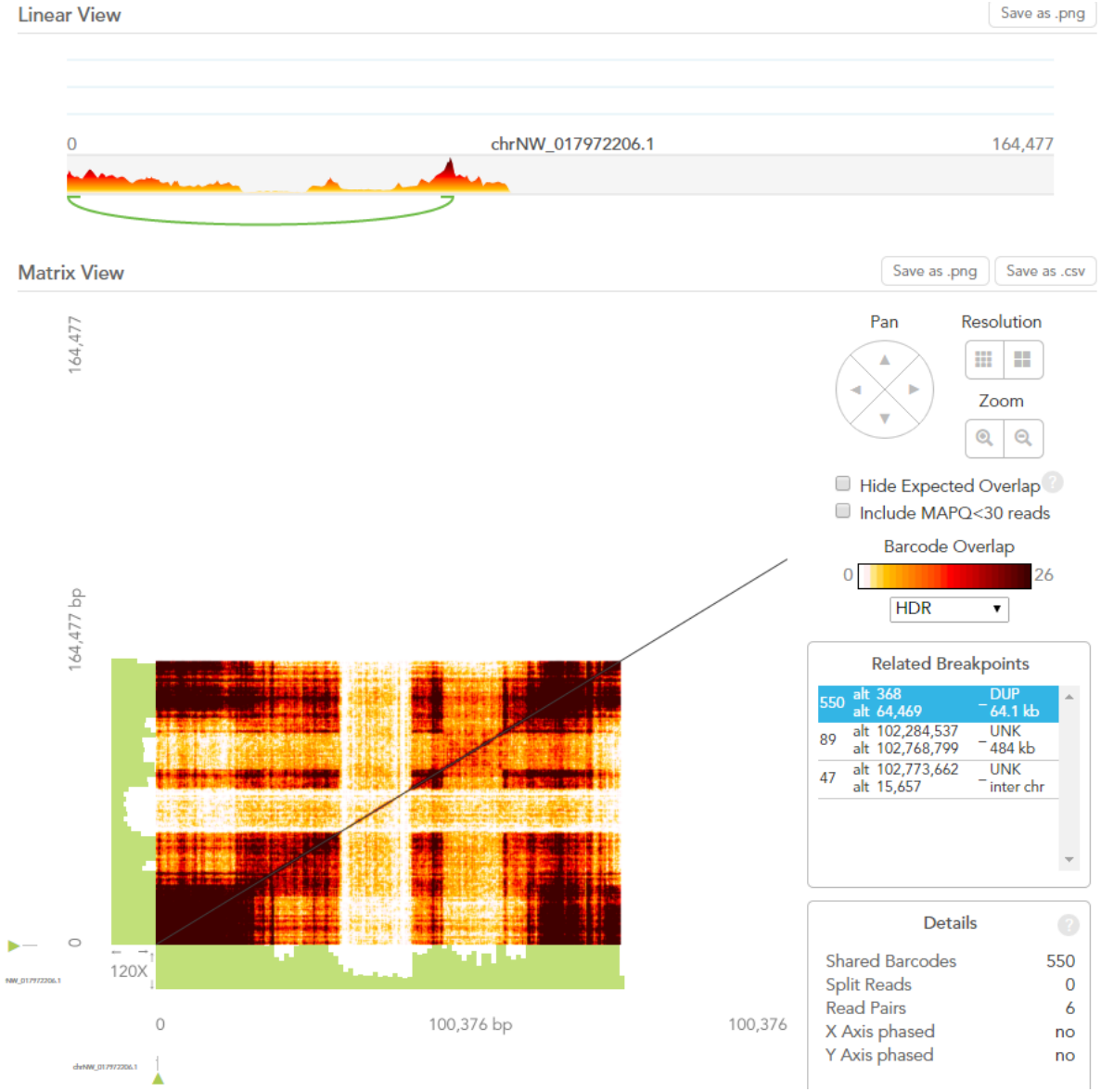
Quality of input DNA

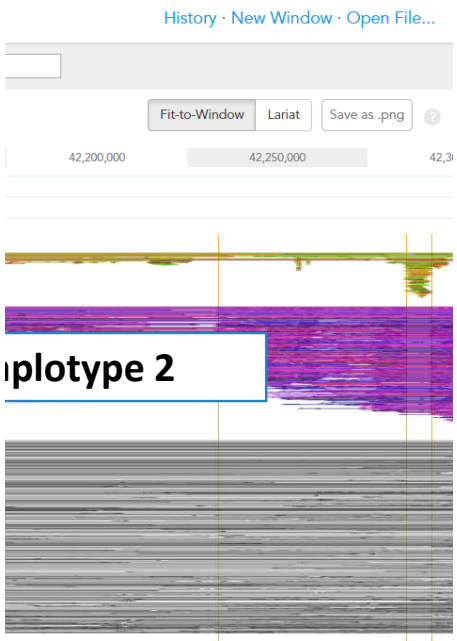
Sequencing

Number of Reads	633,906,276
Median Insert Size	377 bp
Mean Depth	38.4 X
Zero Coverage	6.15%
Mapped Reads	95.8%
PCR Duplication	4.95%
Q30 bases, Read 1	92.4%
Q30 bases, Read 2	72.7%



Differents views



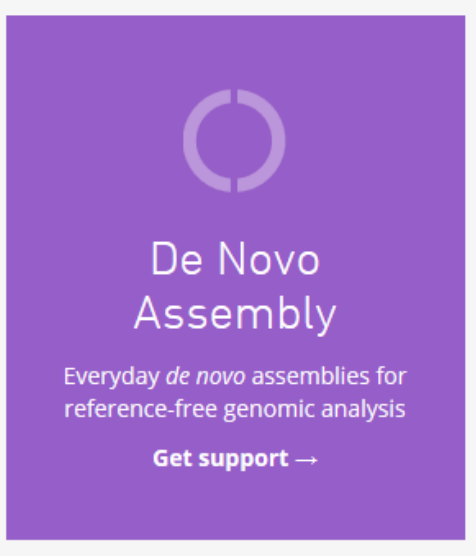


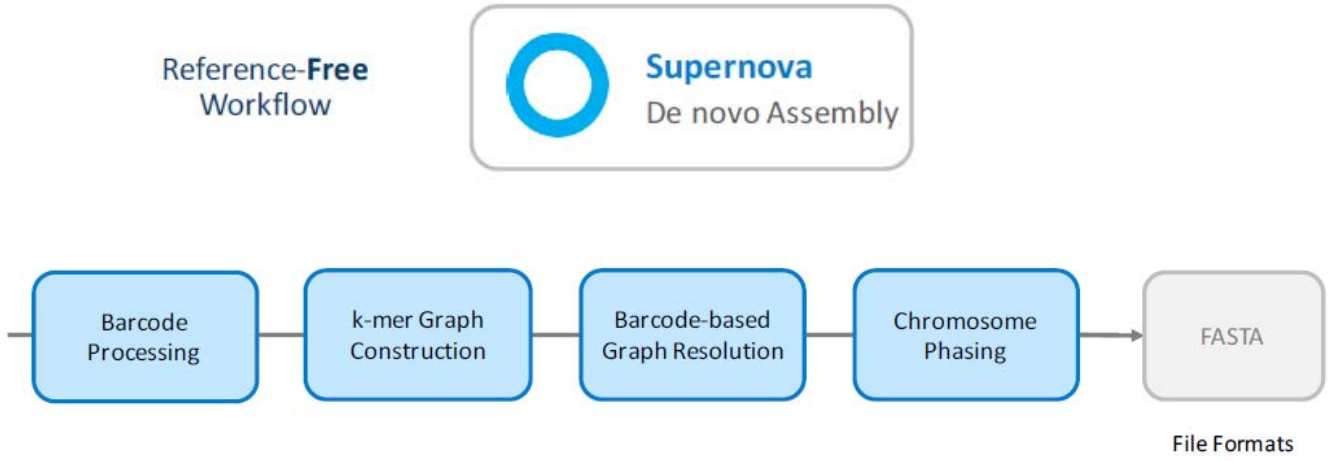
Phasing view – Indels, SNVs, SVs (50bp to up 30Kb)
Structural variant view – deletion, insertion

Our 10X Applications

Genotoul GeT

De Novo Assembly solution





Genomes supported for assembly

Genomes	Supported
Human germline genomes	☺
Mamalian genomes	☺
High repeat content	😐
Small genomes (100 Mb or greater)	😐
Genomes having ploidy > 2	☹
Microbes (<100 Mb)	☹
Human Non Germline (e.g Cancer)	☹
Large genomes (>3.2 Gb)	☹

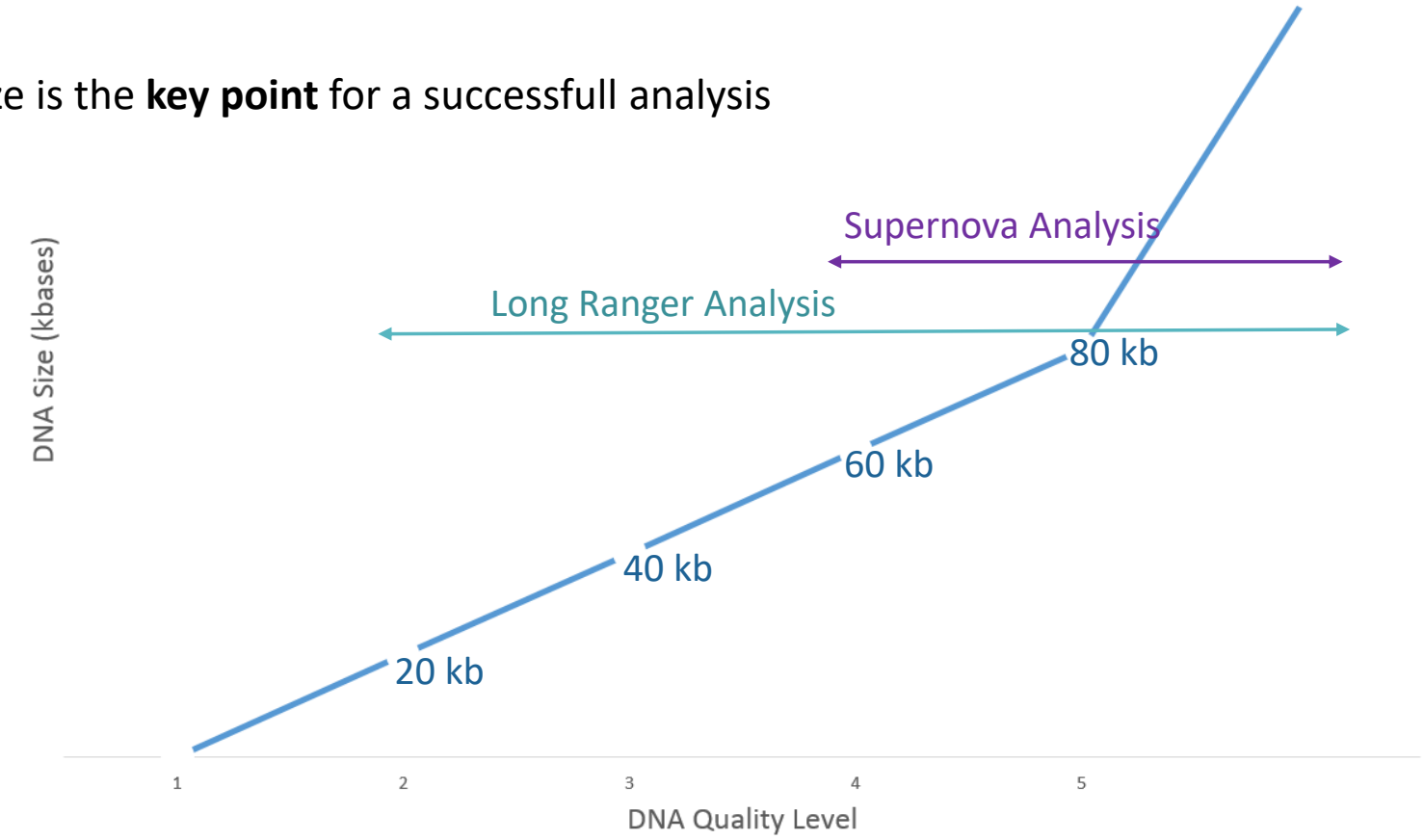
☺ Work well

😐 May work but may increase run time

☹ Unlikely to work well or risky

DNA Size recommendation

DNA size is the **key point** for a successfull analysis



Examples

Olivier Bouchez & Claire Kuchly

28/11/2017

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

Genotoul GeT

10X for Small Genomes ?

Genome size 400Mb

1 lane HiSeq3000
~200X cov

Supernova De novo Assembly

38X cov-> 17% fragments

Supernova De novo Assembly

56X cov-> 24% fragments

Supernova De novo Assembly

100X cov-> 50% fragments

All linked reads

Scaffolding (ARCS)

	Length	N50	L50	Contig number
38X Cov	388,8 Mb	303,528	262	34,913
56X Cov	389,3 Mb	2,104,491	41	24,198
56X Cov + linked reads	389,3 Mb	**3,646,511**	**27**	24,075
100X Cov	446,4 Mb	2,550,395	157	24,797
100X Cov + linked reads	446,4 Mb	**2,550,395**	**48**	24,694

- Good assembly for a small genome with 10X data only
- High number of contigs despite a great L50 and N50

PacBio RSII + 10X Chromium = ?

For assembly purposes, are PacBio RSII and 10X Chromium results similar or is there a benefit to use them together?

(~ 1Gb)

VS

Technology	Assembly analysis	Cov (X)	Tot bases (Mb)	# contigs	N50	L50	Completeness (BUSCO V2)
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	Canu	71	1 015	4 062	1,3 Mb	126	91 %

PacBio RSII + 10X Chromium = ?

For assembly purposes, are PacBio RSII and 10X Chromium results similar or is there a benefit to use them together?

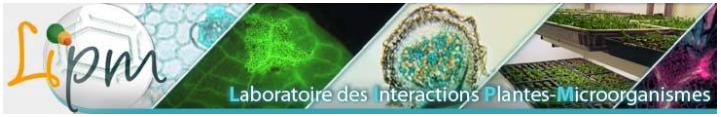
(~ 1Gb)

+

Technology	Assembly analysis	Cov (X)	Tot bases (Mb)	# contigs	N50	L50	Completeness (BUSCO V2)
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 %

With the information of the linked reads, we **can improve the primary assembly** and make an **scaffold assembly**

Thanks!



Jerôme Gouzy

Christophe Klopp

Laboratoire de Physiologie et génomique des poissons
Yann Guiguen

NGS team:
Olivier Bouchez
PacBio team:
Alain Roulet
Céline Roques
10X Genomics team:
Adeline Chaubet
Sophie Valière
Pauline Heuillard
Frédéric Martins
Bioinfo team:
Maxime Manno
Anaïs Poiraudeau
Claire Kuchly
Céline Vandecasteele

Stats assemblage 10X supernova