



Implementation and Evaluation of 10X Genomics Chromium technology

Claire Kuchly, Olivier Bouchez

► To cite this version:

Claire Kuchly, Olivier Bouchez. Implementation and Evaluation of 10X Genomics Chromium technology. Retour d'expérience, GeT-PlaGe, Nov 2017, Castanet-Tolosan, France. <hal-04100440>

HAL Id: hal-04100440

<https://hal.science/hal-04100440v1>

Submitted on 17 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



Implementation and Evaluation of 10X Genomics Chromium technology

Claire Kuchly & Olivier Bouchez

28/11/2017



<http://get.genotoul.fr>
get@genotoul.fr
 [@get_genotoul](https://twitter.com/get_genotoul)



Chromium evaluation: pilot phase

- ⑤ Platform installed in november 2016
- ⑤ Training: november 15th 2016
- ⑤ 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



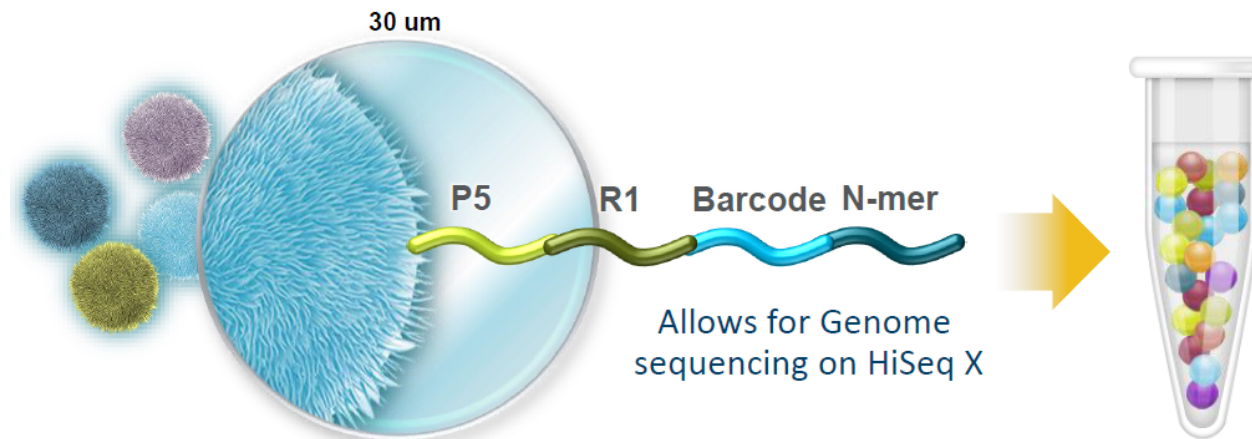
HiSeq3000



NovaSeq

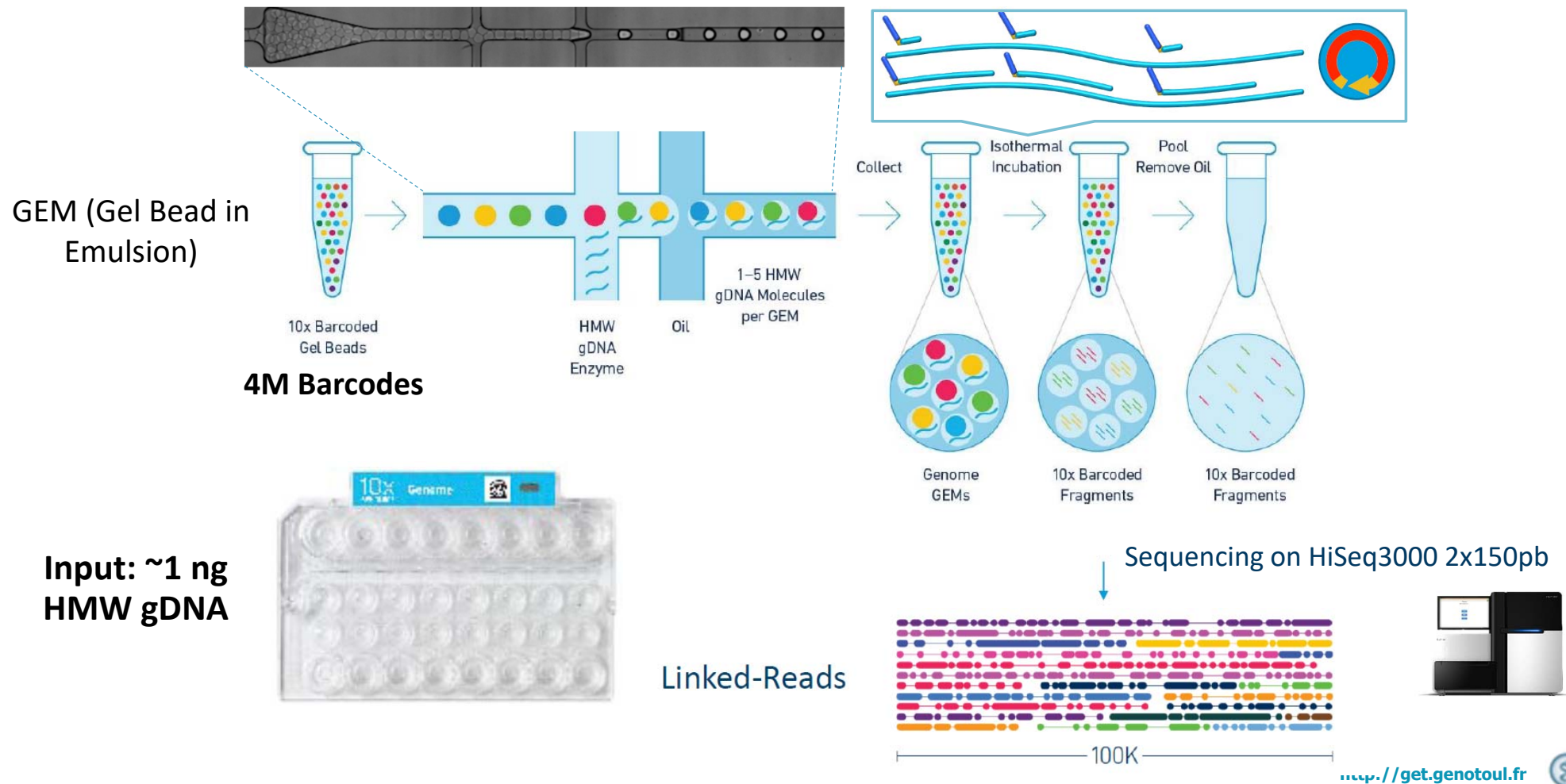
Chromium 10XGENOMICS

How does it work ?

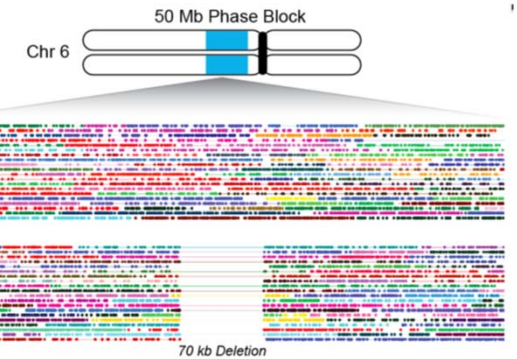
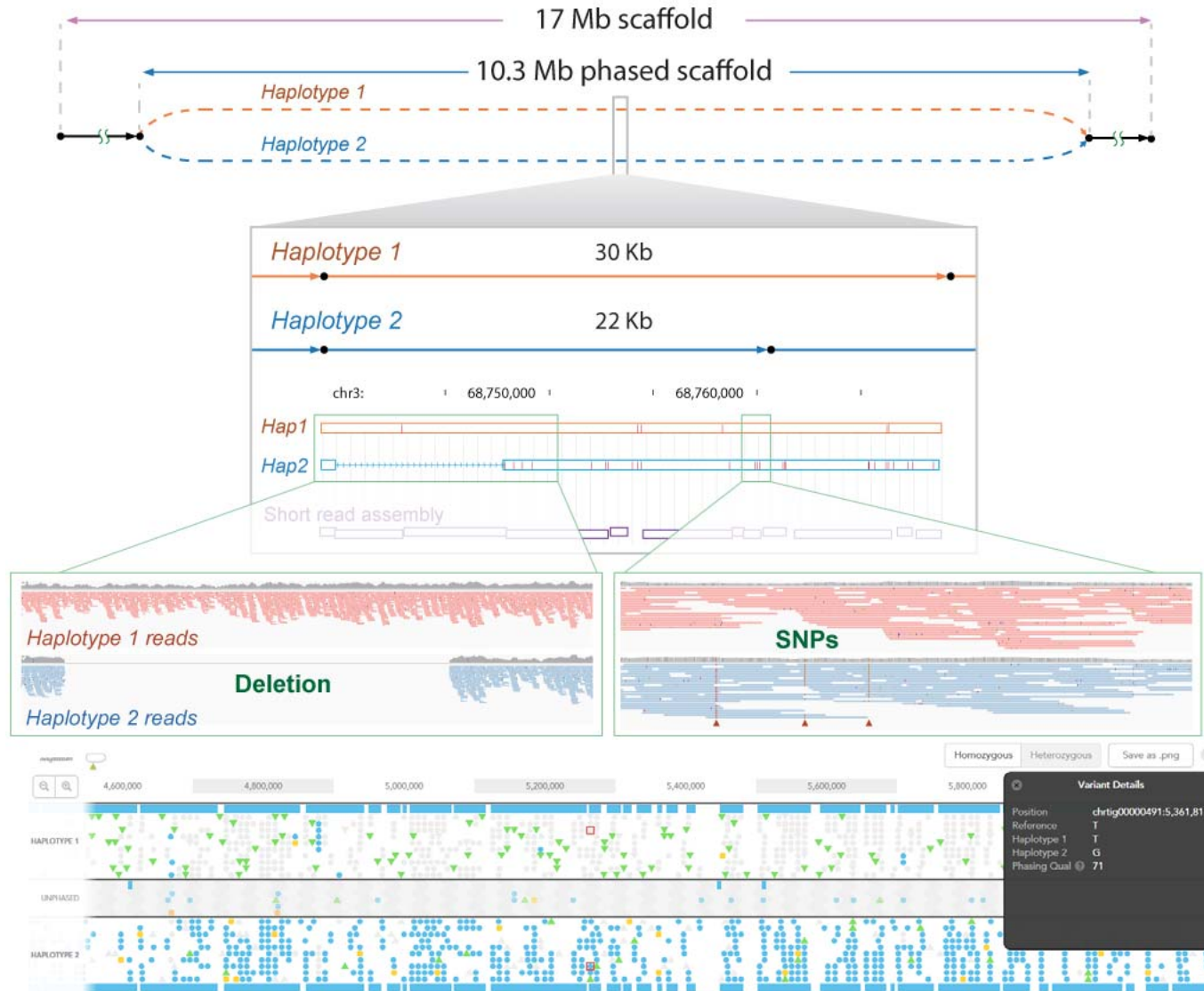


How does it work ?

Complete wetlab workflow: 2 days



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.

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

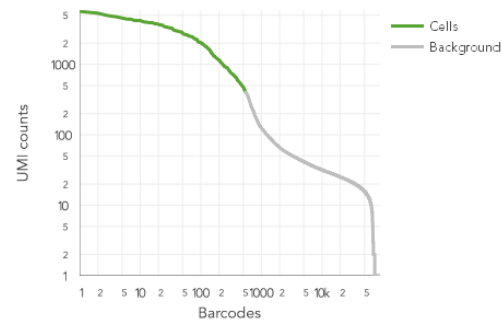
Median Genes per Cell

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%

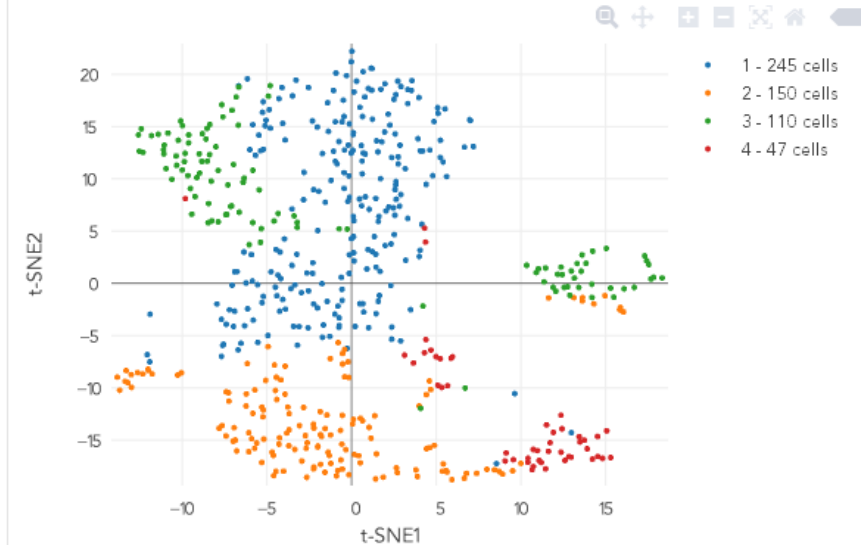
Cells



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

t-SNE projection of Cells Colored by Automated Clustering



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

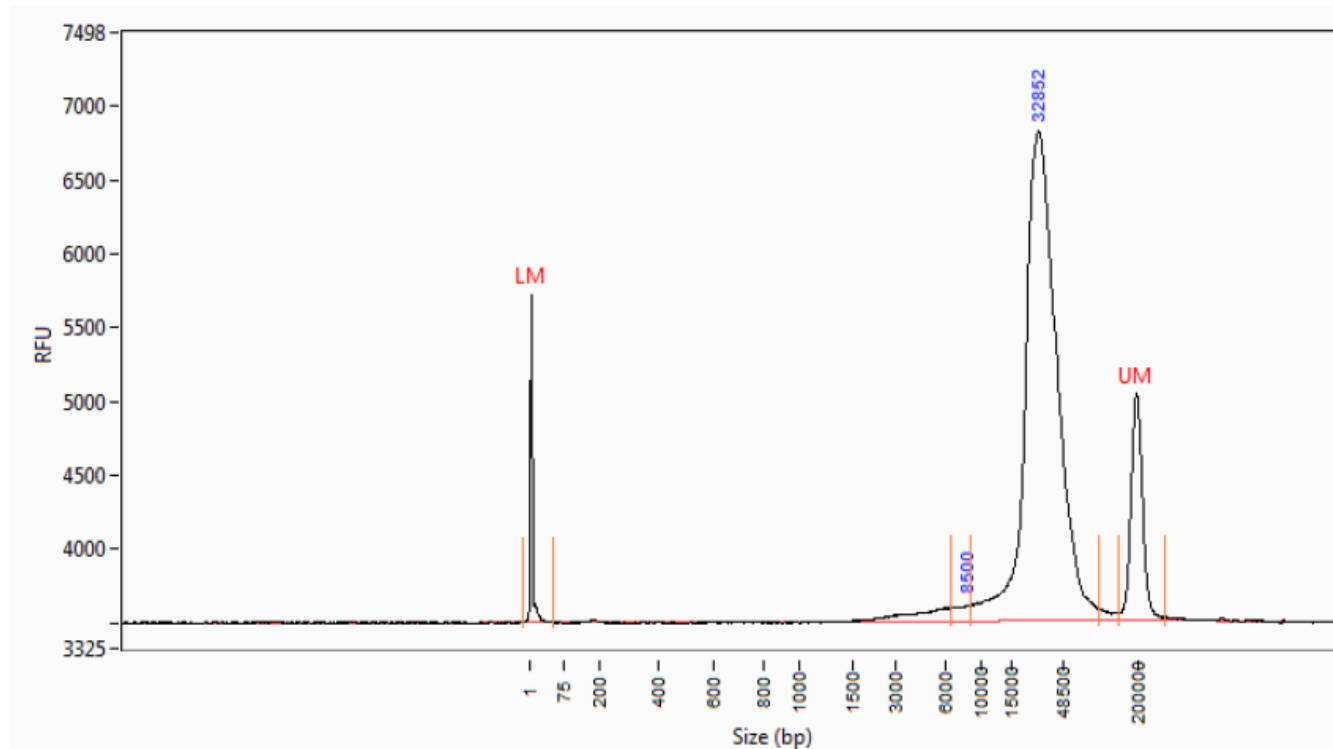


Haplotyping using Chromium

DNA prep, without BluePippin



- ② Young rabbit = low amount of DNA
- ② Molecular weight assessed by Fragment Analyzer



Haplotyping using Chromium

Results via Loupe software

Without BluePippin sizing



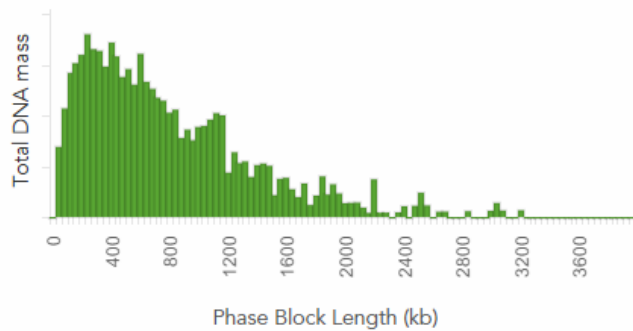
GEM Performance

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

Phasing

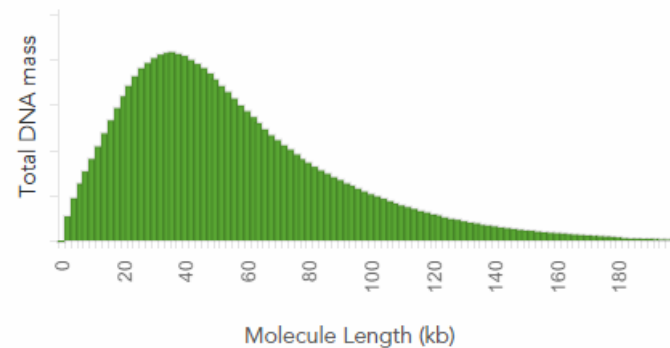
SNPs Phased	97.1%
Longest Phase Block	3,241,488 bp
N50 Phase Block	632,948 bp

[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



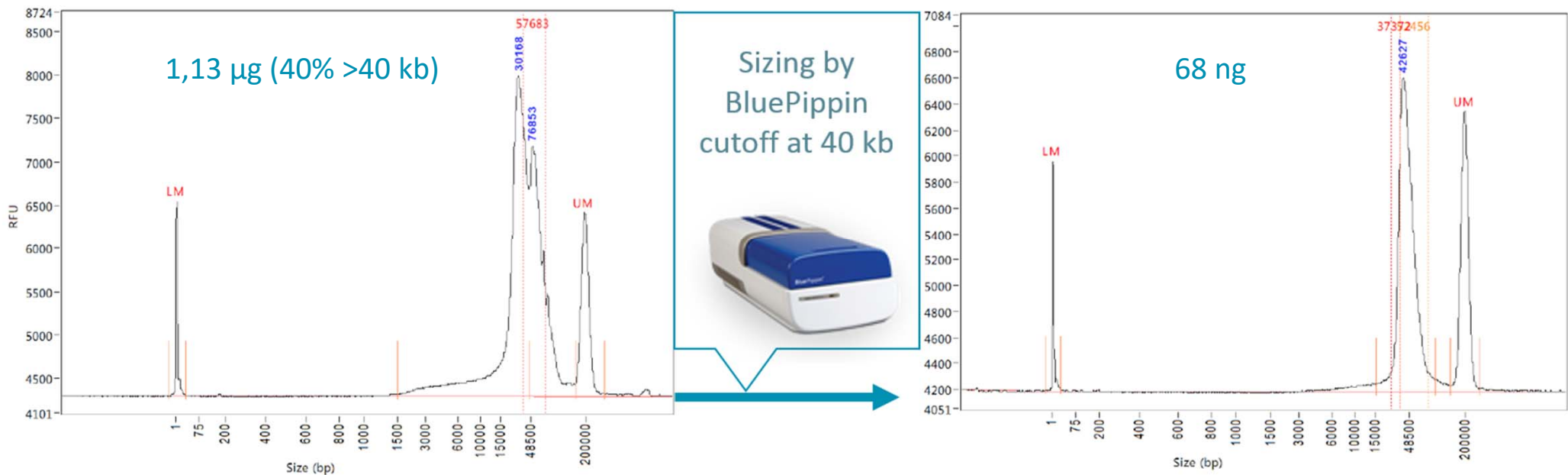
Structural Variants

Large Structural Variant Calls	117
Short Deletion Calls	47,585

Haplotyping using Chromium

DNA prep

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



Haplotyping using Chromium

With BluePippin sizing : Results via Loupe software

With BluePippin sizing

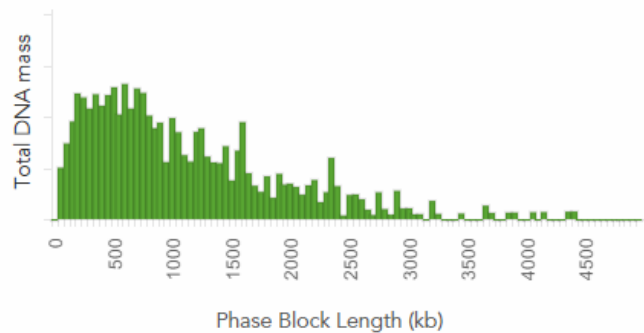
GEM Performance

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

Phasing

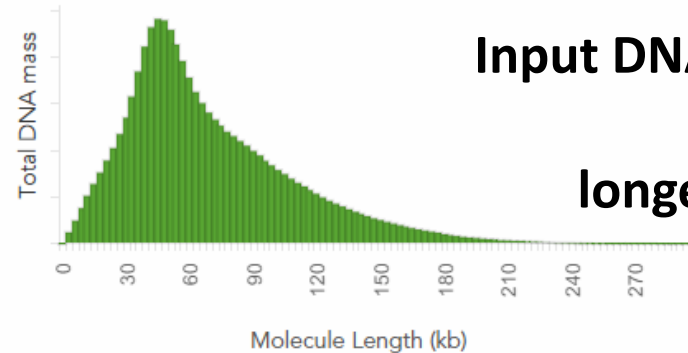
SNPs Phased	97.2%
Longest Phase Block	4,417,547 bp
N50 Phase Block	875,494 bp

[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



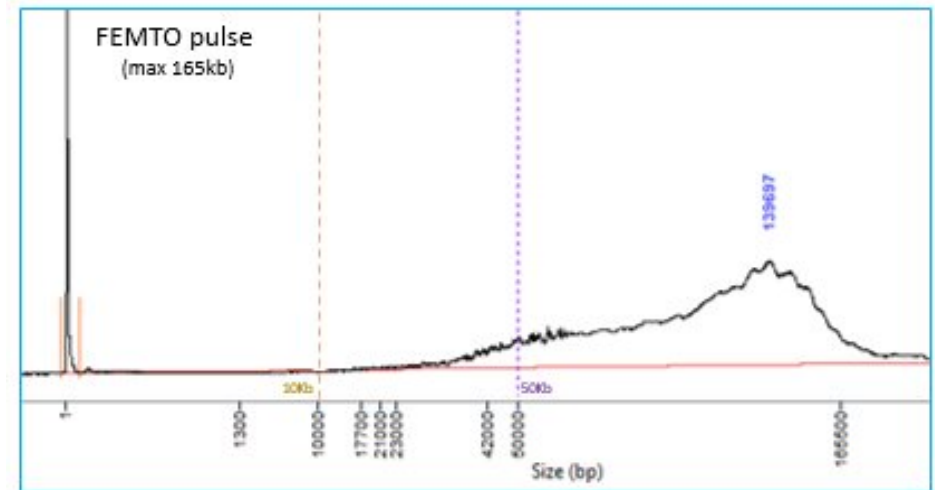
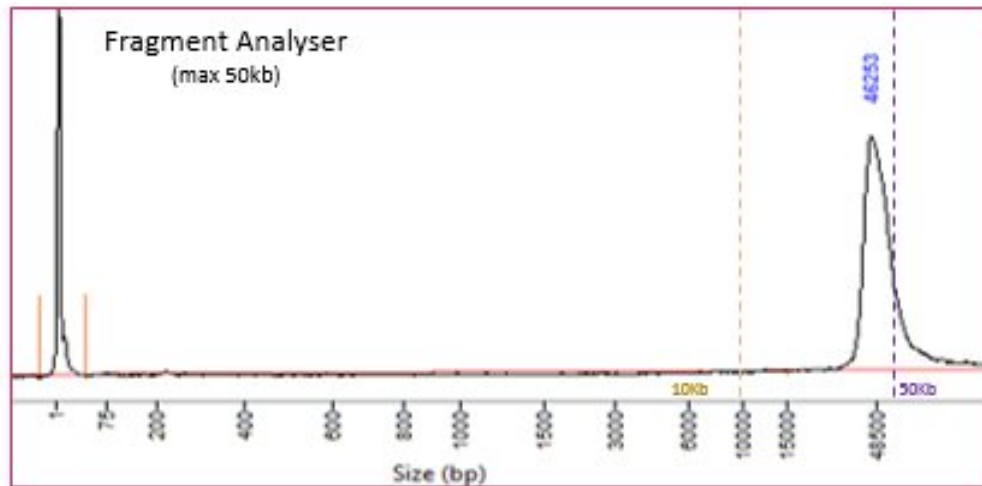
Input DNA size improvment
=
longer phase block

Structural Variants

Large Structural Variant Calls	162
Short Deletion Calls	49,846



DNA QC: New Femto Pulse



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



Genome & Exome

Long-range analysis and phasing of SNVs, indels, and structural variants

Get support →



Single Cell Gene Expression

Gene expression profiling at scale with single cell resolution

Get support →



Single Cell V(D)J

Paired V(D)J profiling at scale with single lymphocyte resolution

Get support →



De Novo Assembly

Everyday *de novo* assemblies for reference-free genomic analysis

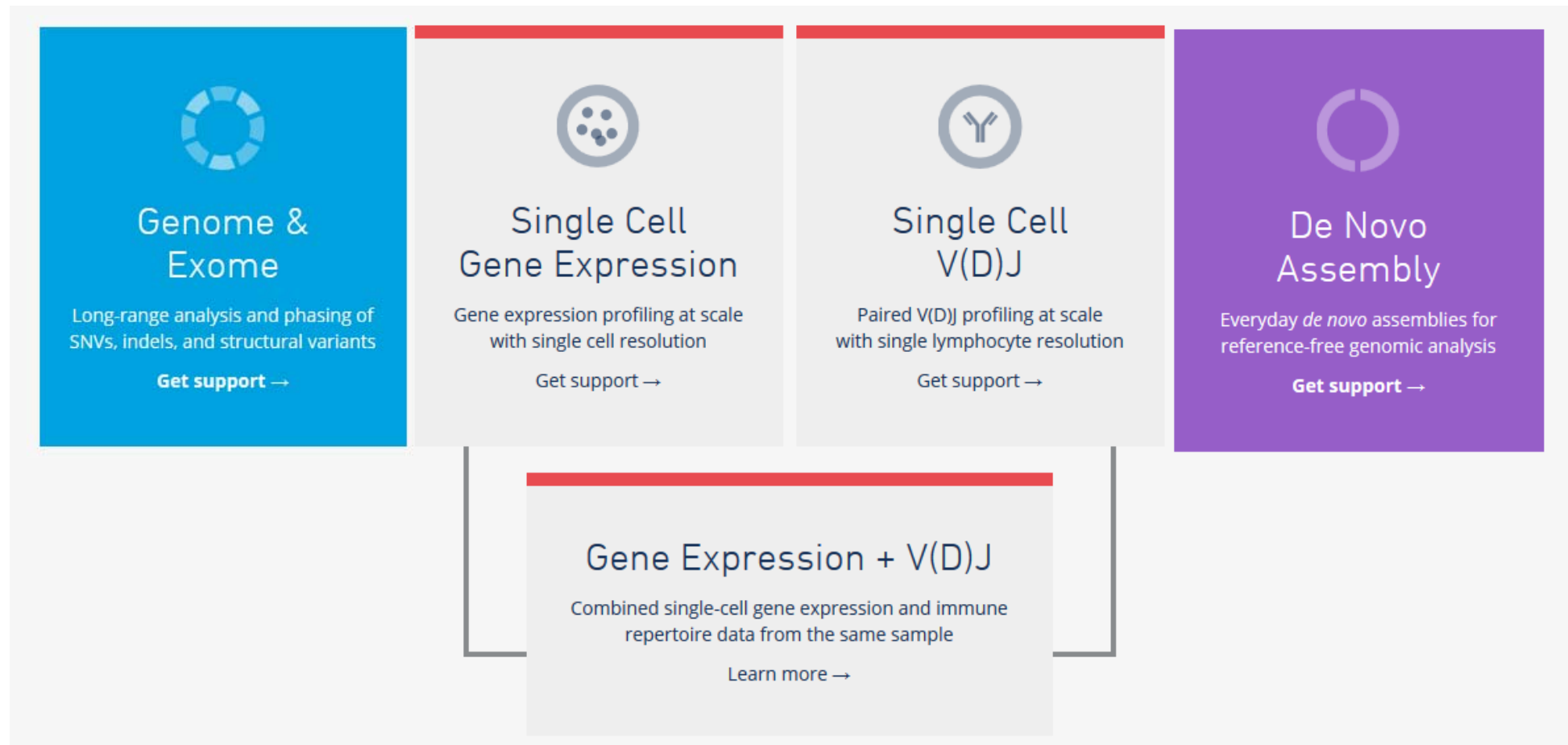
Get support →

Gene Expression + V(D)J

Combined single-cell gene expression and immune repertoire data from the same sample

Learn more →

Our 10X Applications



Genome & Exome solution



Reference-Based
Workflow

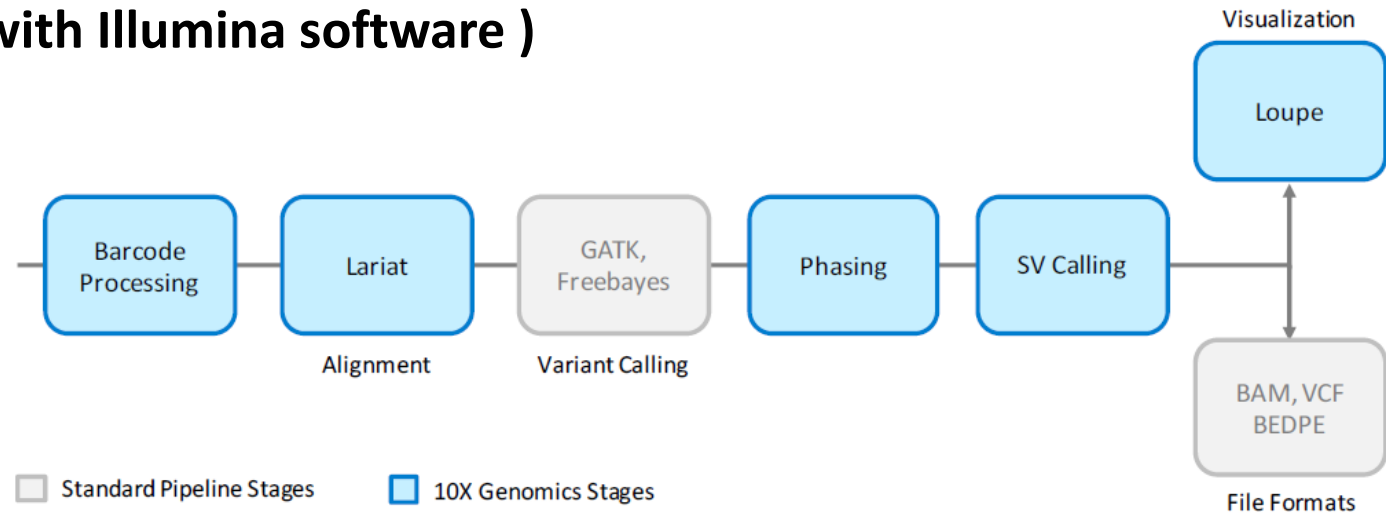


Linked-Reads analysis software

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

7,677,566

GEMs Detected

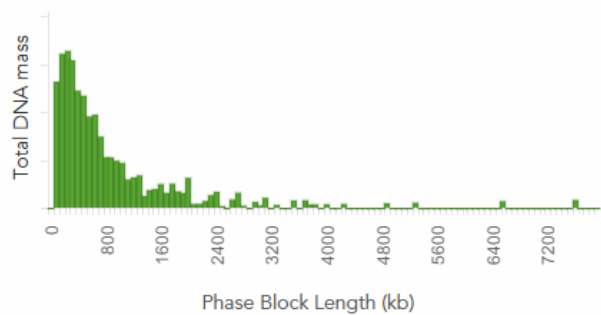
1,306,663

GEM Performance

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

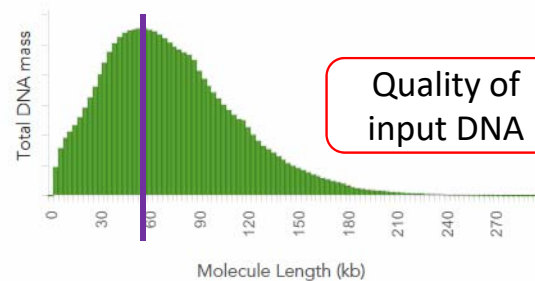
Phasing

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

[Phase Block View](#)

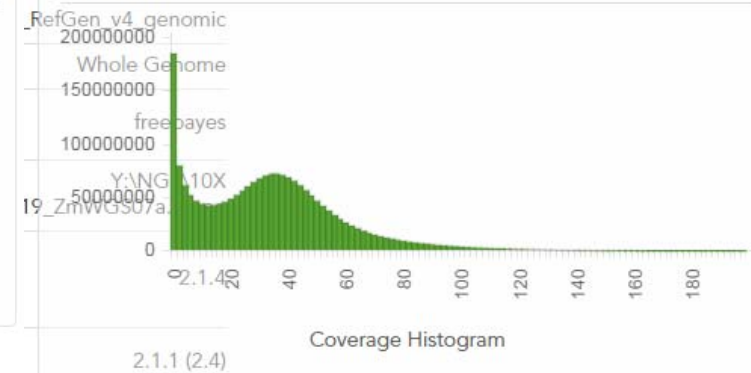
Input DNA

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



Sequencing

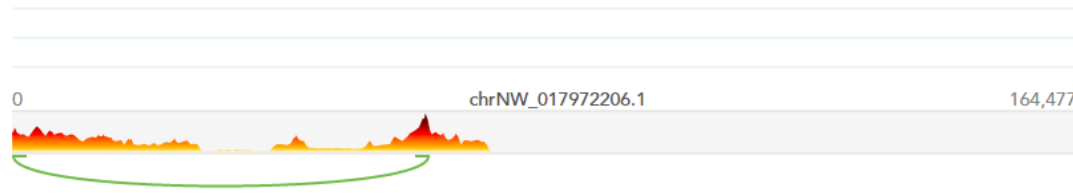
Number of Reads	633,906,276
Median Insert Size	377 bp
Mean Depth	38.4 X
Zero Coverage	6.15%
Mapped Reads	95.8%
PCR Duplication	4.95%
Q30 bases, Read 1	92.4%
Q30 bases, Read 2	72.7%





Linear View

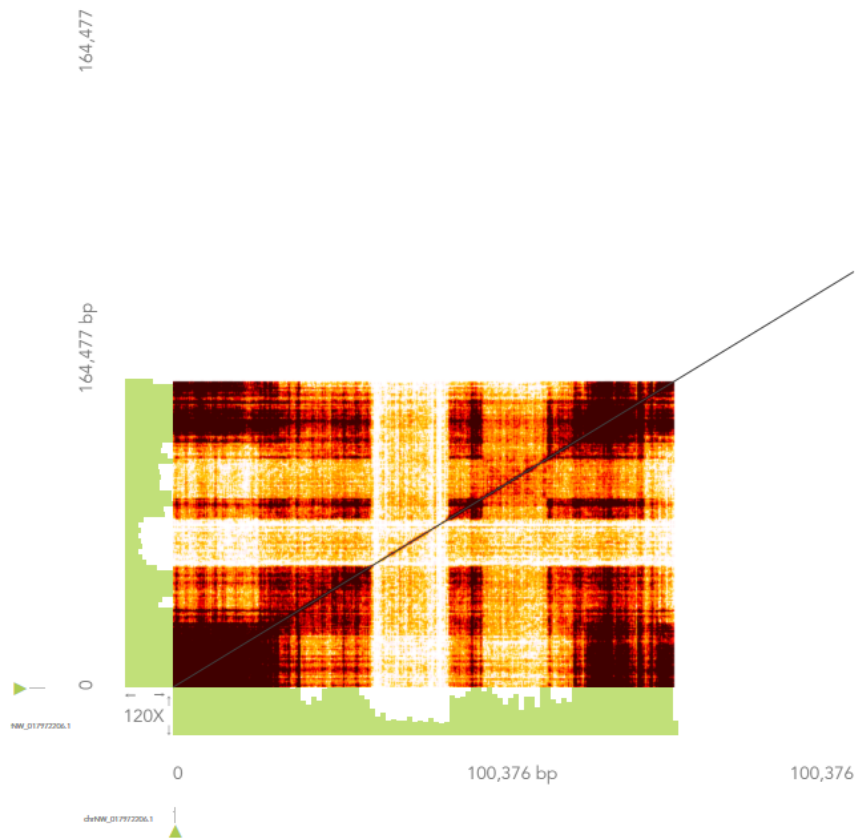
Save as .png



Matrix View

Save as .png

Save as .csv



Resolution



☐ Hide Expected Overlap ?

☐ Include MAPQ<30 reads

Barcode Overlap

0 26

HDR ▼

Related Breakpoints

550	alt 368	DUP
	alt 64,469	64.1 kb
89	alt 102,284,537	UNK
	alt 102,768,799	484 kb
47	alt 102,773,662	UNK
	alt 15,657	inter chr

Details ?

Shared Barcodes	550
Split Reads	0
Read Pairs	6
X Axis phased	no
Y Axis phased	no

Different views

History · New Window · Open File...

Fit-to-Window

Lariat

Save as .png

42,200,000

42,250,000

42,300,000



Phasing view – Indels, SNVs, SVs (50bp to up 30Kb)

Structural variant view – deletion, insertion

Our 10X Applications



Genome & Exome

Long-range analysis and phasing of SNVs, indels, and structural variants

Get support →



Single Cell Gene Expression

Gene expression profiling at scale with single cell resolution

Get support →



Single Cell V(D)J

Paired V(D)J profiling at scale with single lymphocyte resolution

Get support →



De Novo Assembly

Everyday *de novo* assemblies for reference-free genomic analysis

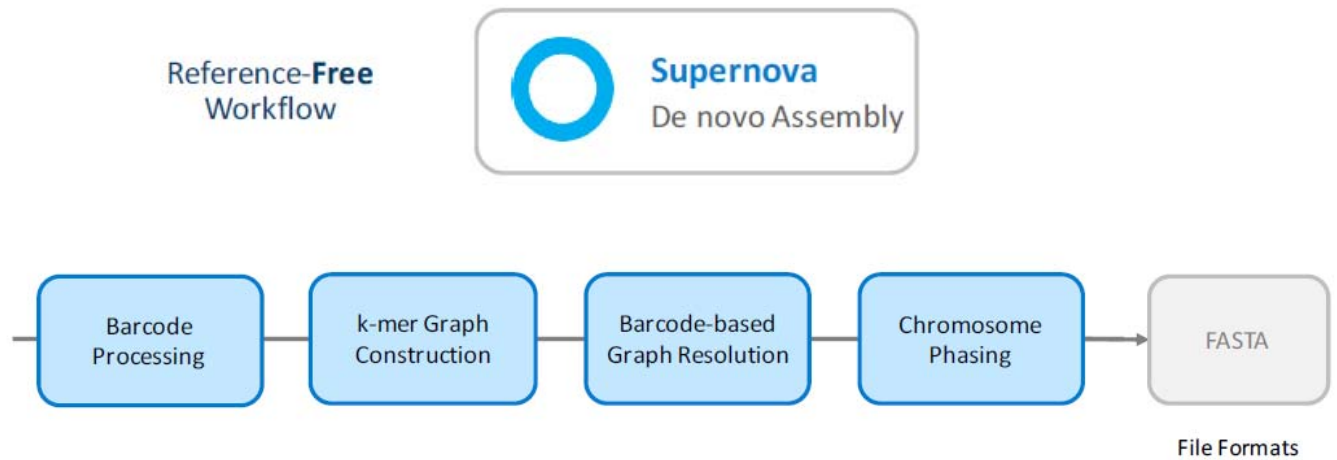
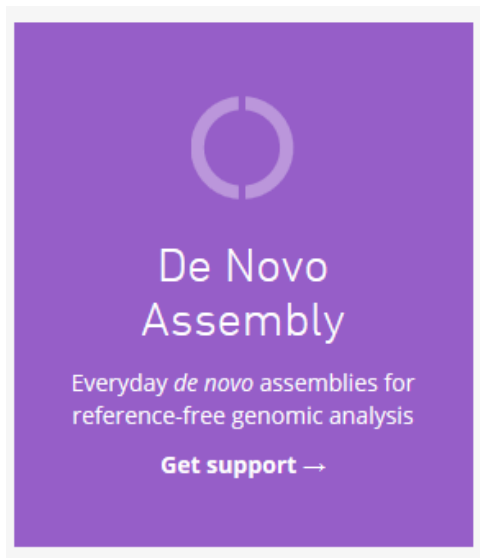
Get support →

Gene Expression + V(D)J













Combined single-cell gene expression and immune repertoire data from the same sample

Learn more →

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



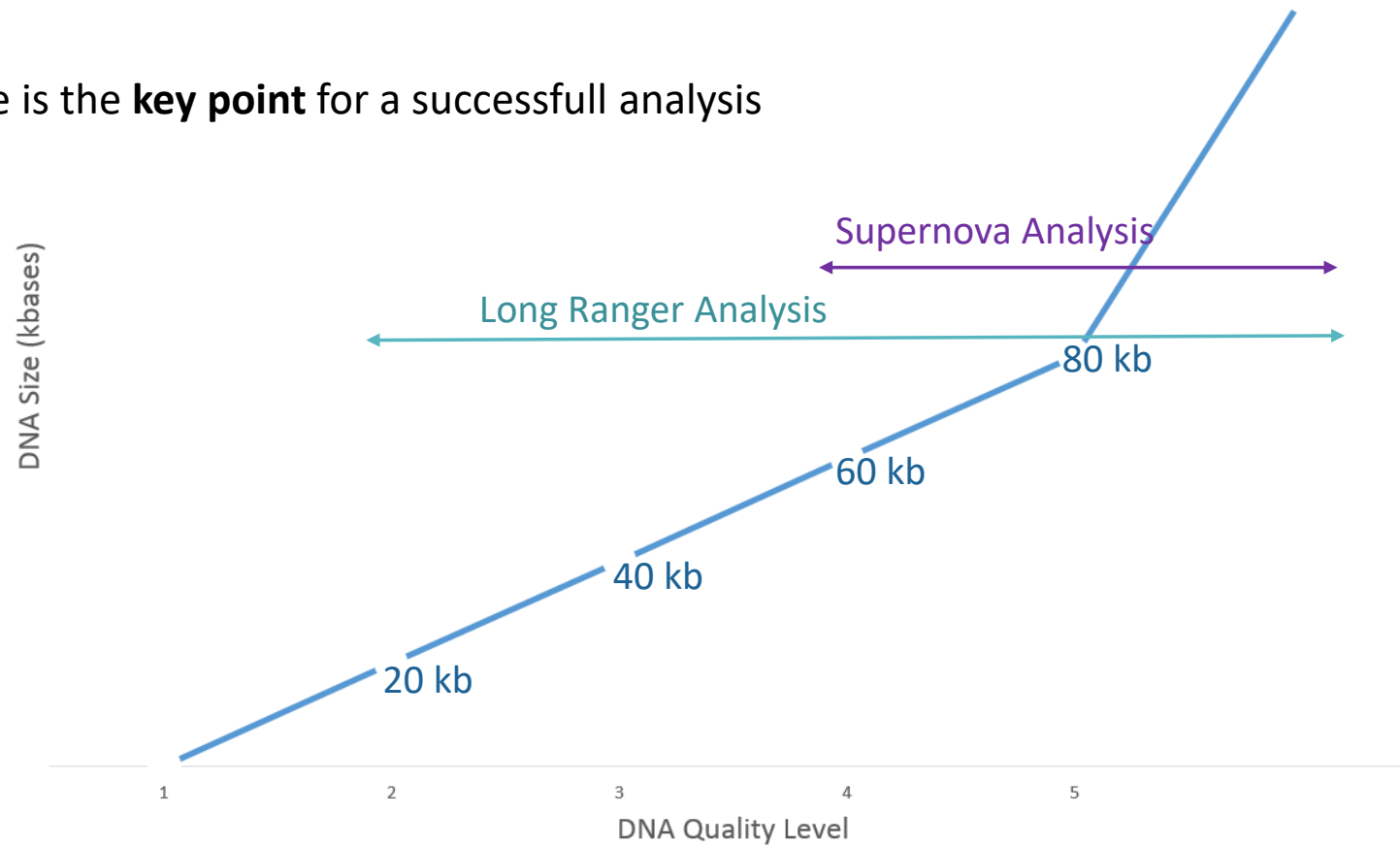
May work but may increase run time



Unlikely to work well or risky

DNA Size recommendation

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



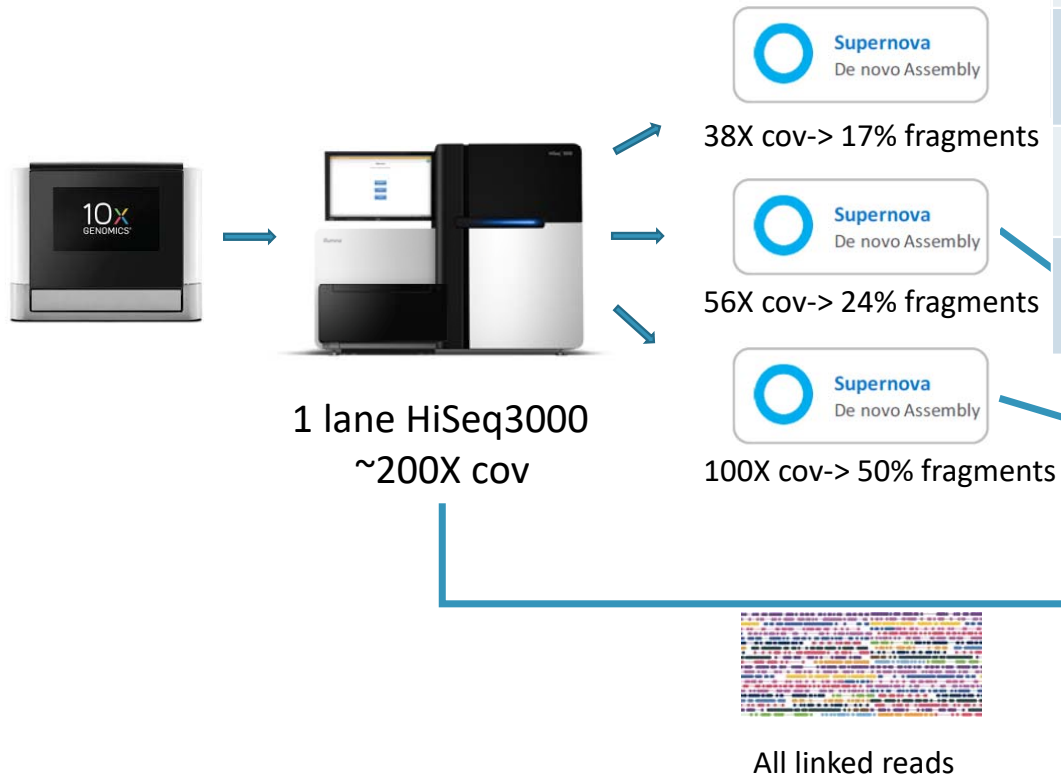
Examples

Olivier Bouchez & Claire Kuchly

28/11/2017

10X for Small Genomes ?

Genome size 400Mb



	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

Scaffolding (ARCS)

- 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?



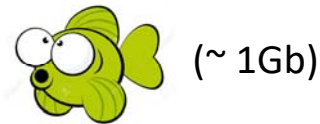
VS



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

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 Poiraudau

Claire Kuchly

Céline Vandecasteele



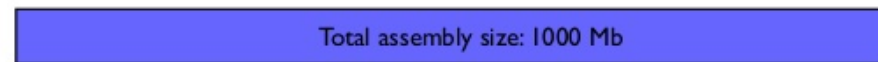
Christophe Klopp



Stats assemblage 10X supernova



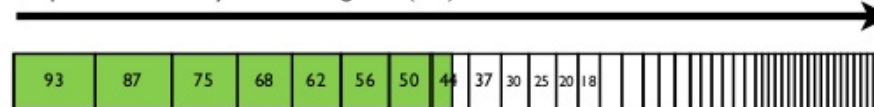
N50/L50



N50



Sequences order by descending size (Mb)



N50 = 7 sequences

L50 = 50 Mb