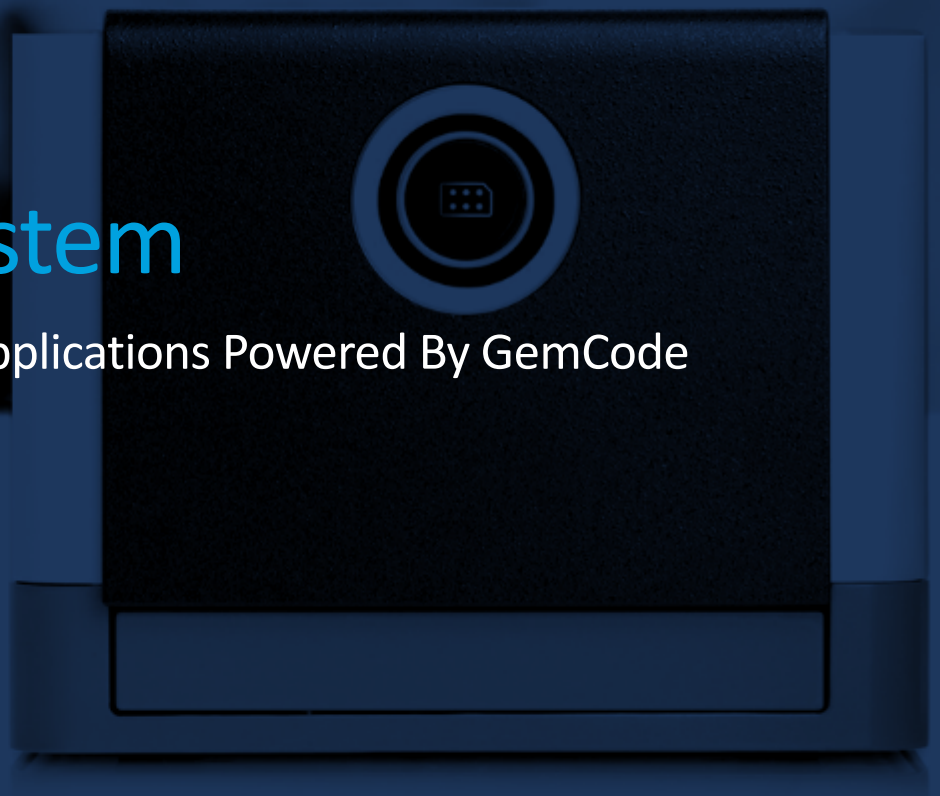
The Chromium™ System

Linked Read And Single Cell RNA-Seq Applications Powered By GemCode Technology



Chris Black, M.S.
Sales Executive
10x Genomics
chris.black@10xgenomics.com

Presentation Structure

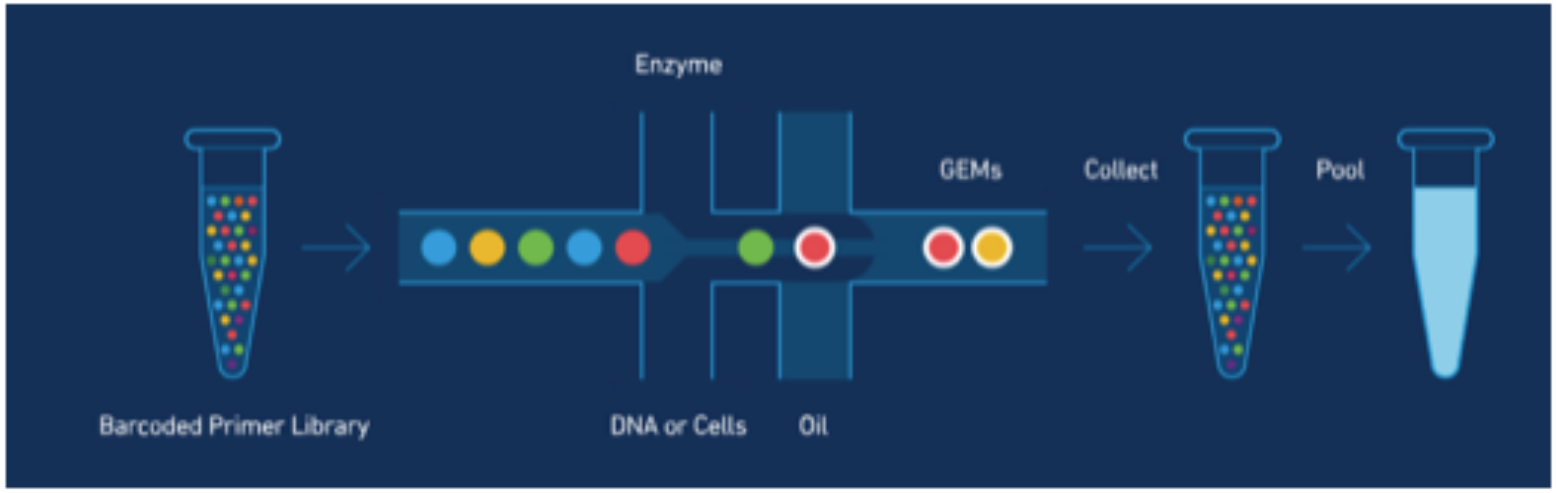
- GemCode™ Technology Overview
- Linked Read Applications with Chromium
  - Chromium Genome & Exome (Hybrid Capture)
  - *De Novo* Assembly Using Supernova
- Single Cell 3’ RNA-seq with Chromium
- Chromium Software Suite Overview

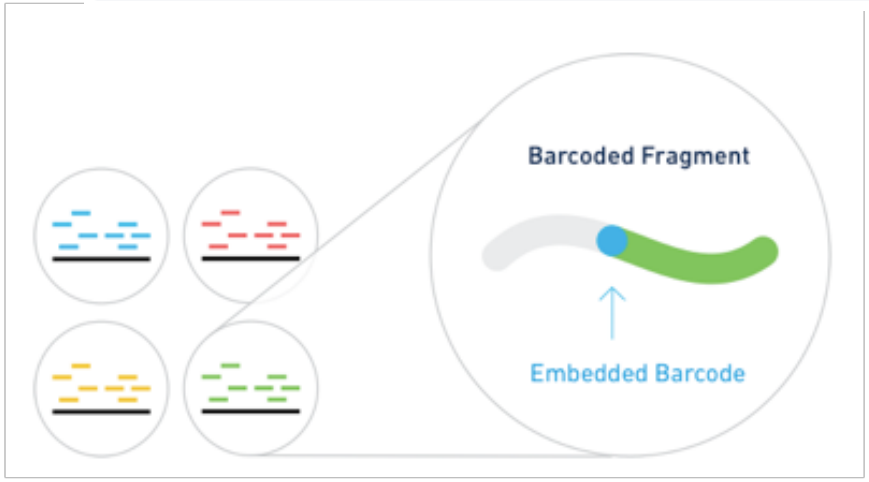
Confidential — Do not distribute

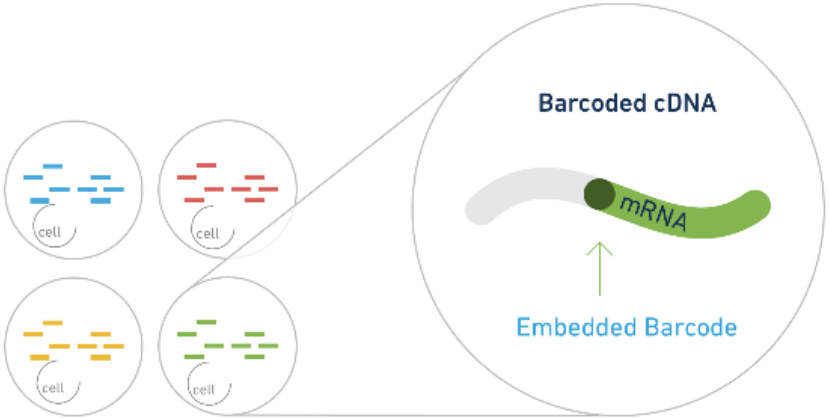
GemCode Technology Overview

GemCode Technology – Creating GEMs

- GEM – Gel Bead in EMulsion

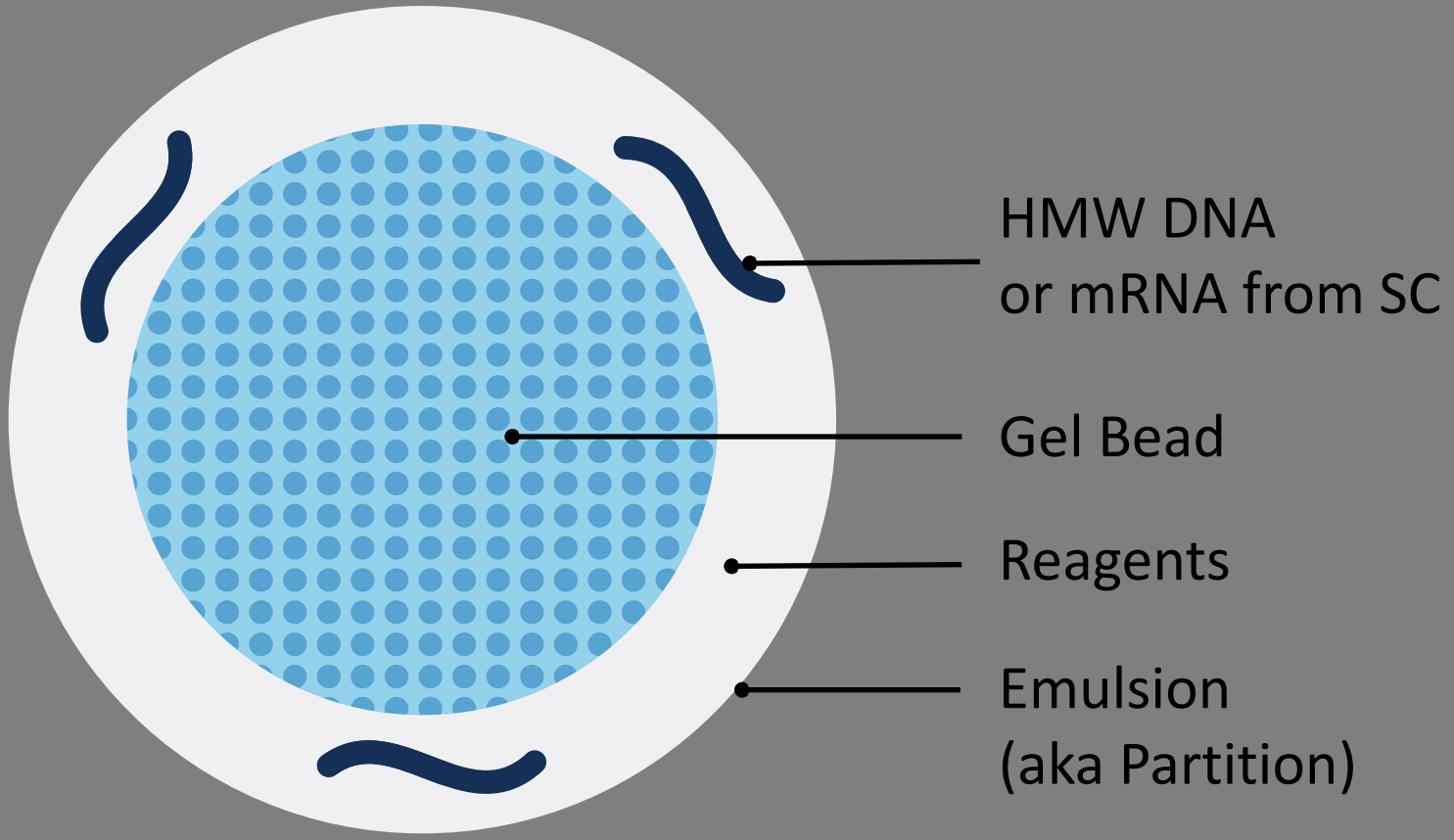






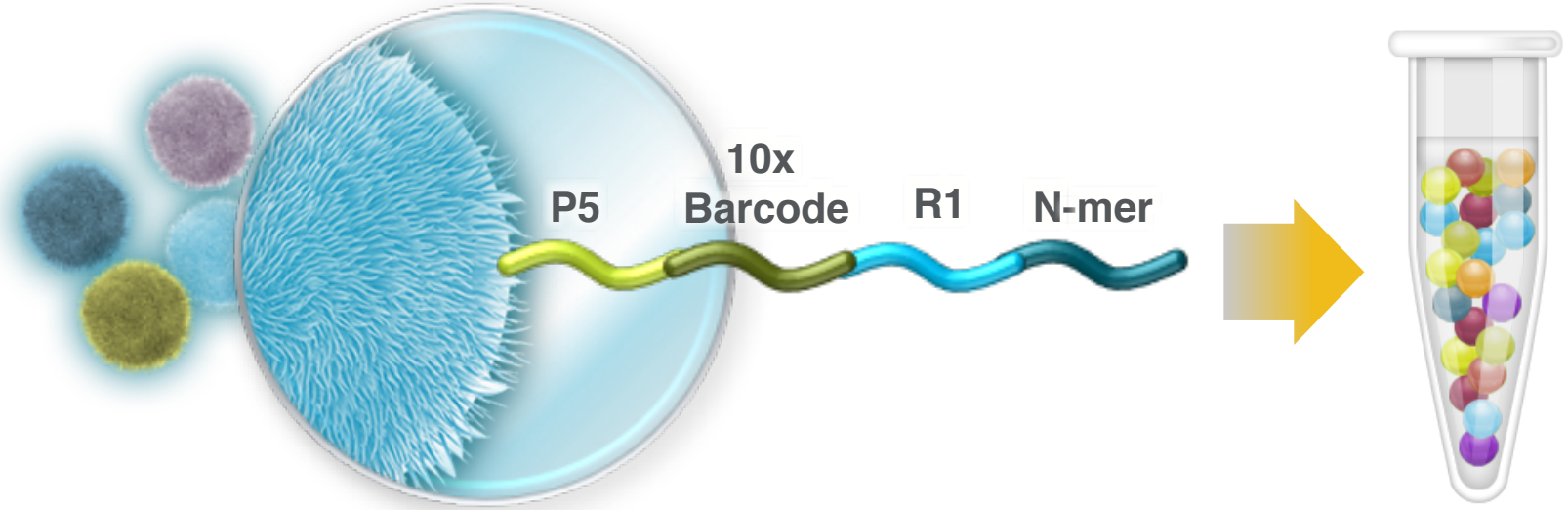
GEM Structure

Reagent Delivery, Sample Partitioning and Barcoding System



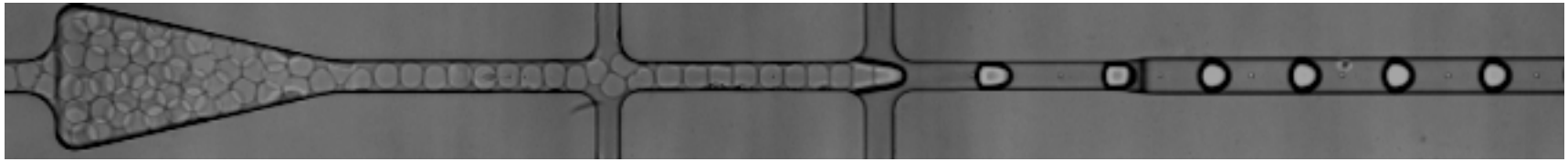
Gel Bead Oligo Delivery Substrate

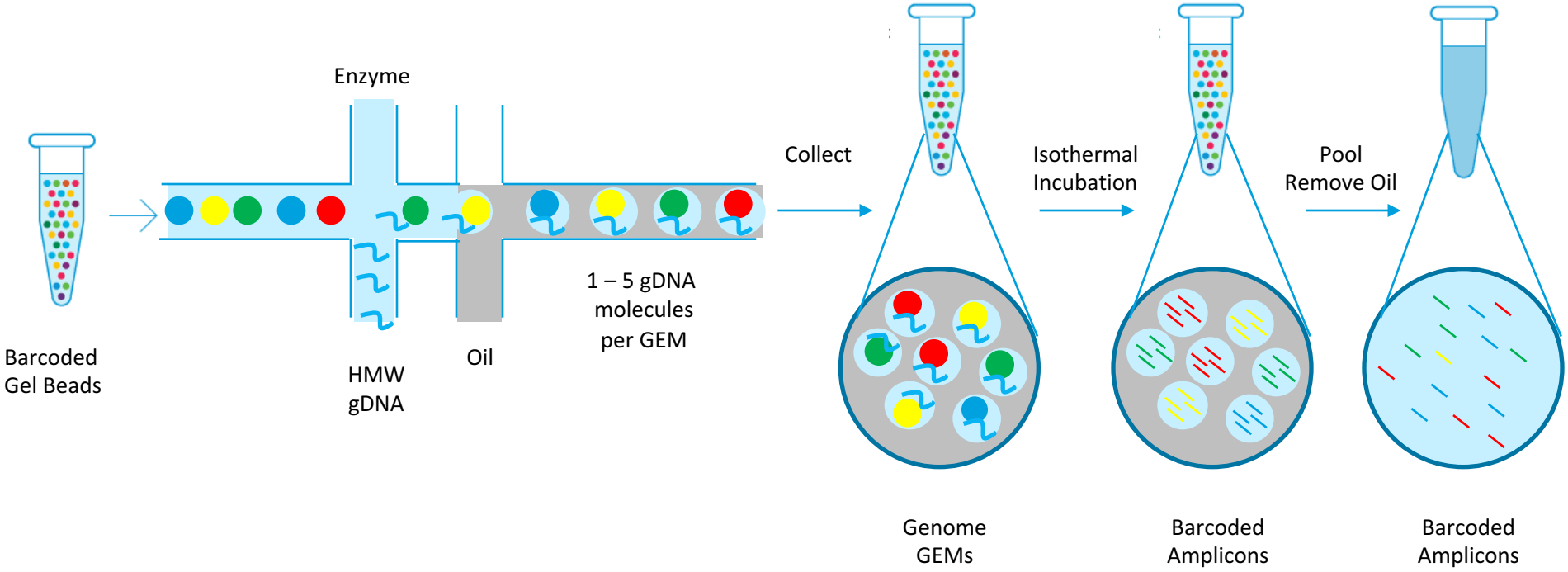
Gel bead scaffold → Functional oligo with barcode → High-diversity library



- Millions of copies of identical oiligos
- Defined barcode sequence
- Built-in sequencing adapter and primer content

Generating GEMs with the Chromium Controller





Solid phase reagent delivery → Fluid partitioning → Liquid phase biochemistry

Diversity Drives Performance

	STANDARD BARCODING	Chromium Single Cell 3' RNA-seq	Chromium Genome
Partitions	384	>1,000,000	>1,000,000
Barcode Pool	384	750,000	4,000,000
Input	100ng+ gDNA	2k – 12k cells	1ng HMW gDNA

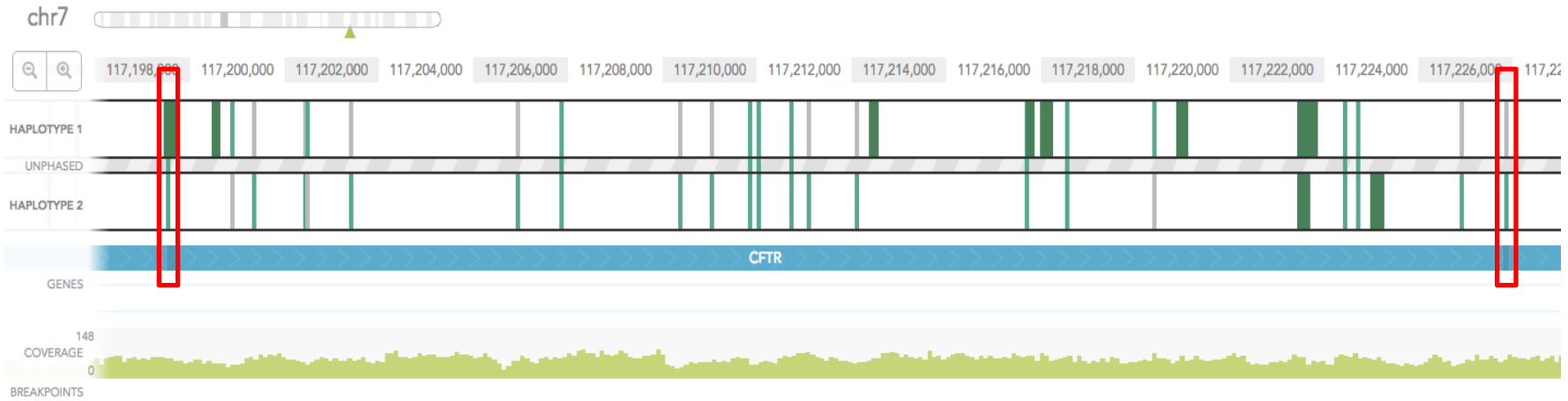
- Massive partitioning and barcode diversity enables
  - lower sequencing depth requirement (LR)
  - lower purity/quality samples (LR)
  - *de novo* Assembly (LR)
  - Minimal duplication rate even with 10k+ cells (SC)

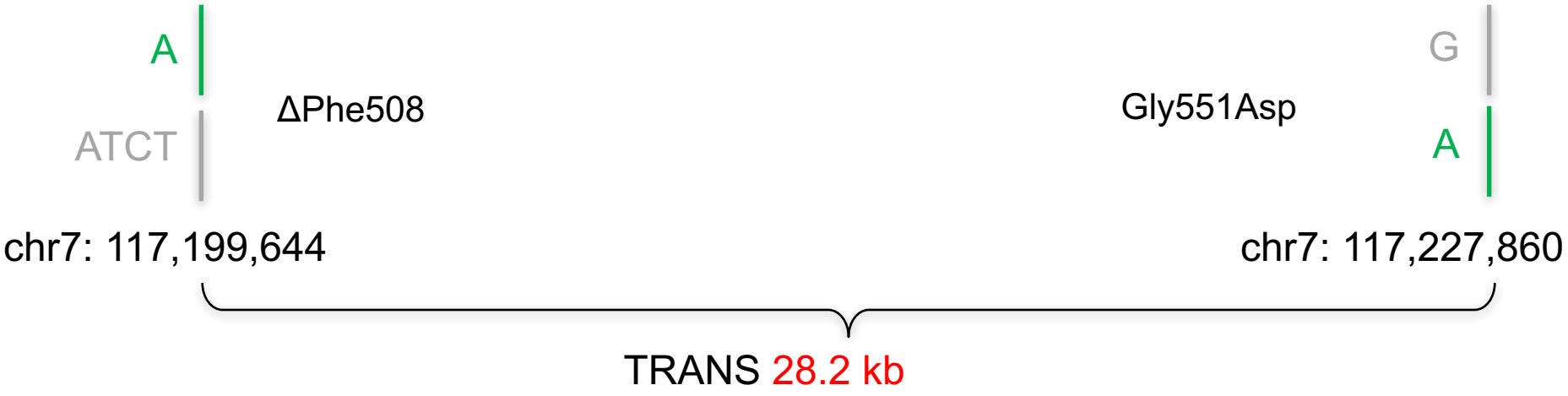
Linked Read Applications with Chromium:
Chromium Genome and Exome (Hybrid Capture)

Linked Reads & The Chromium Genome

- **Resolve the genome into multi-megabase phase blocks**
  - Phase the full spectrum of called variants
- **Detect structural variants**
  - Translocations, insertions, deletions, duplications, complex rearrangements
- **Recover variants in previously inaccessible parts of the genome**
  - Confidently map reads and call variants even in repetitive regions

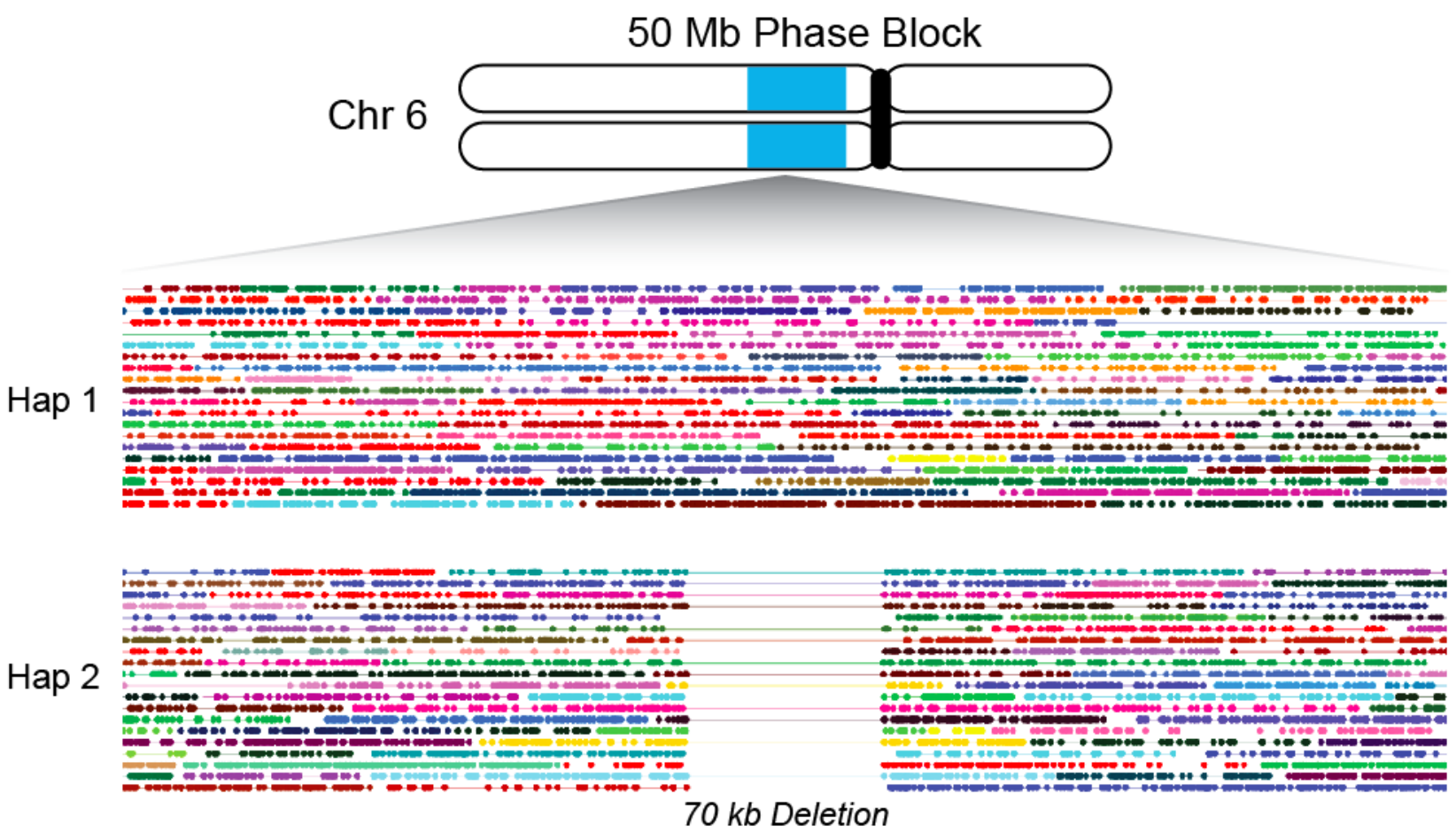
Chromium Genome: Phase Compound Hets





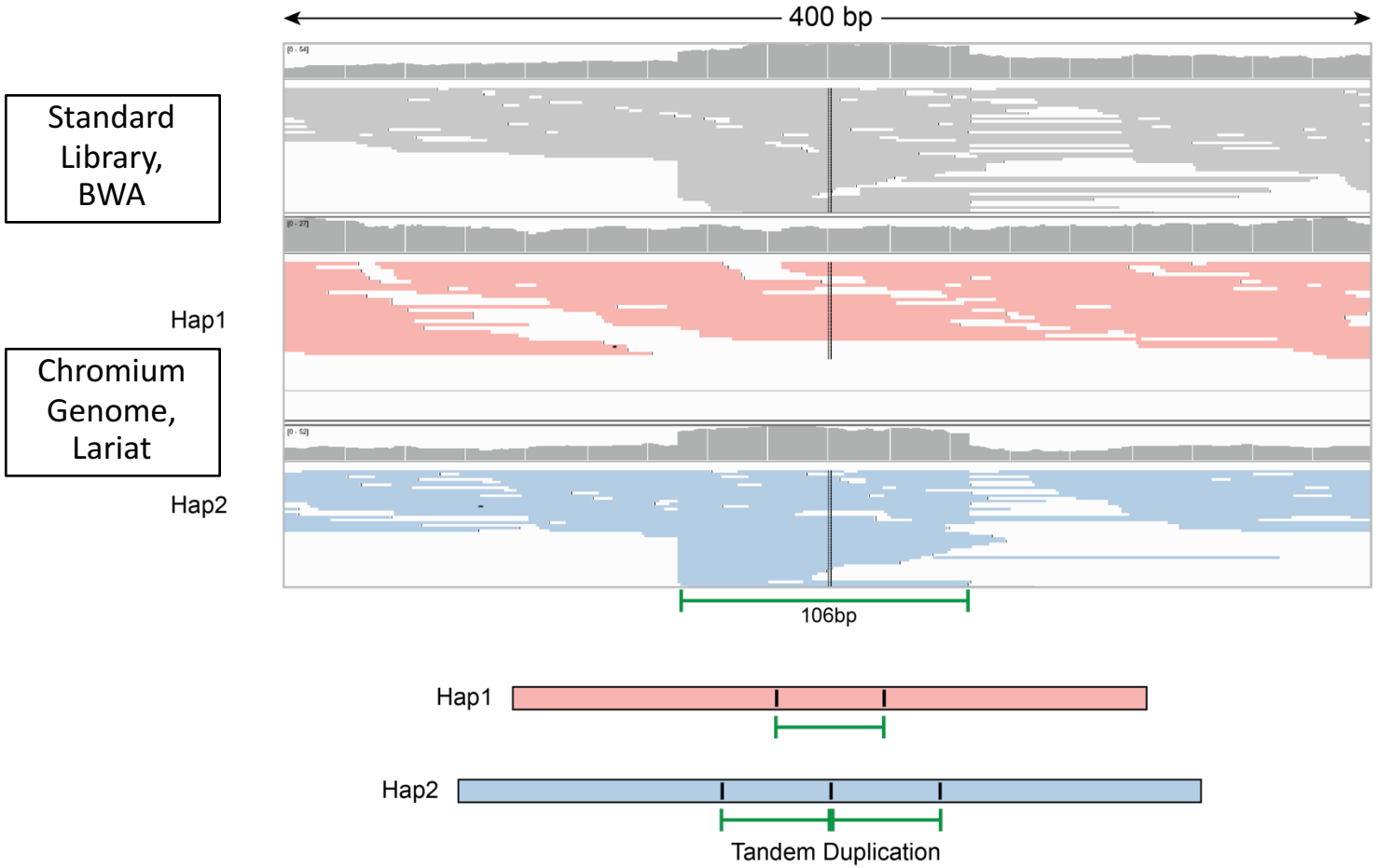
- NA11274 – Coriell sample from female affected with cystic fibrosis
- Compound heterozygous variants in cystic fibrosis transmembrane receptor
- Gene within 18Mb phase block

Chromium Genome: Phase Large Deletions



- NA12878 – Coriell sample from female
- 128Gb ILMN sequence data (2x150)
- N50 phase block – 8Mb
- 99% of all SNPs Phased

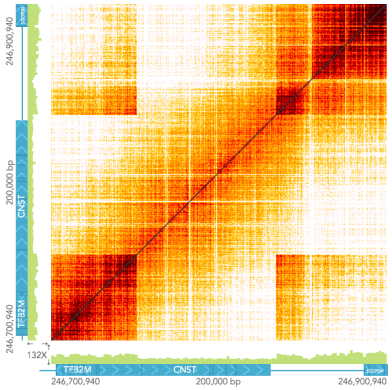
Chromium Genome: Phase Tandem Dupes



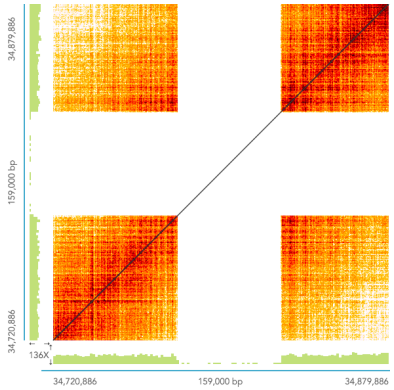
- NA12878 – Coriell sample from female
- 128Gb ILMN sequence data (2x150)
- N50 phase block – 8Mb
- 99% of all SNPs Phased

Chromium Genome: Detect SVs

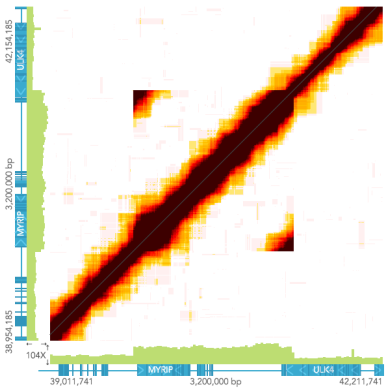
- Automated calling of structural variants < 50bp & > 30 kb
- Data supports detection of variants of all sizes



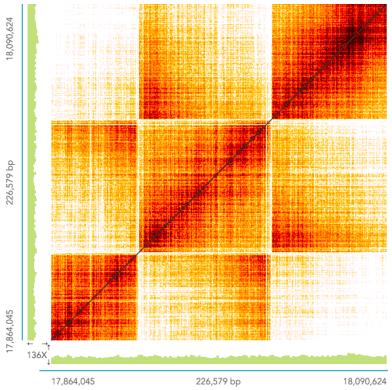
Heterozygous deletion
HCC1954T



Homozygous deletion
NA12878



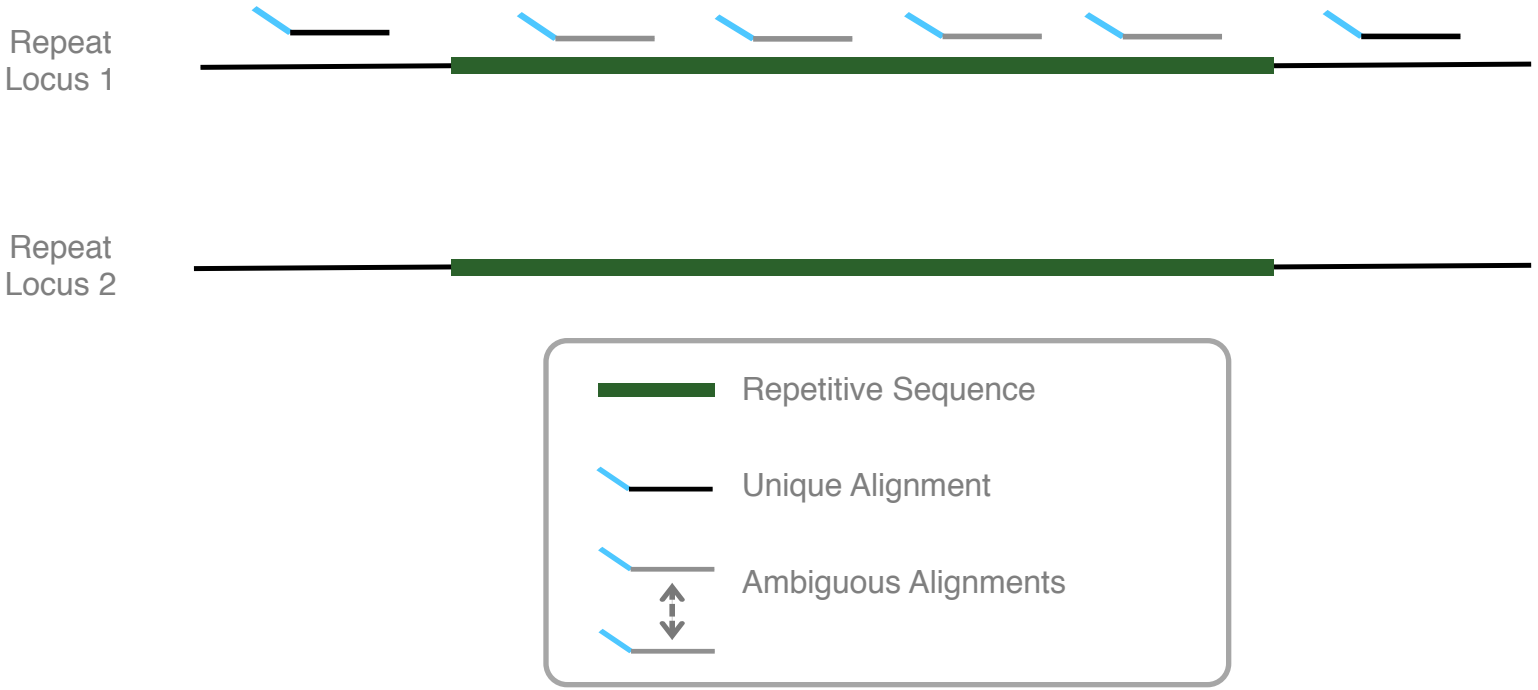
Tandem duplication
HCC1143T



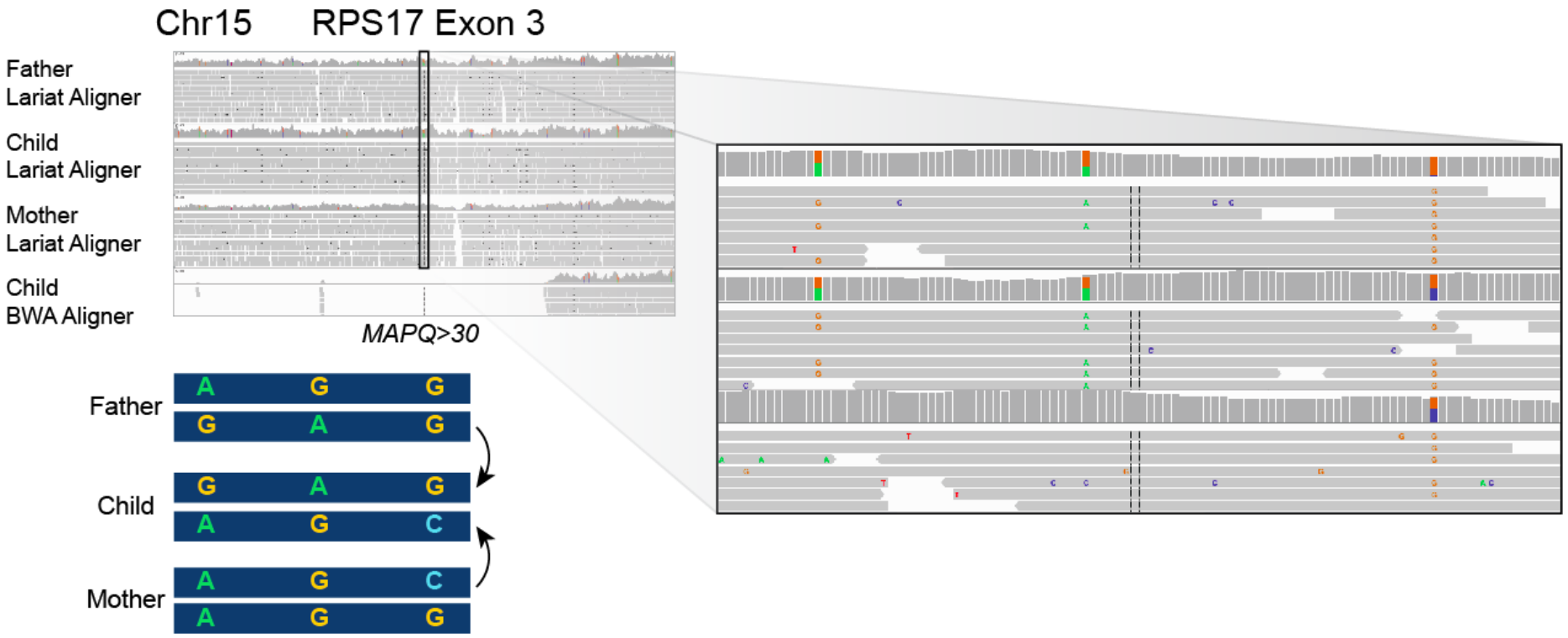
Homozygous inversion
NA12878

Chromium Genome: Resolve Mapping Issues

- BWA core generates multiple candidate alignments and fails to assign reads
- Lariat aligner uses linked read data from 10x barcodes to rescue unmapped reads

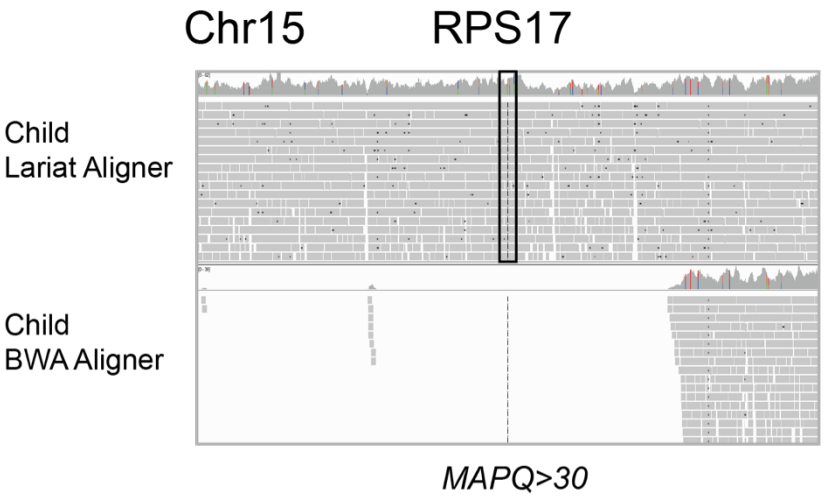


Chromium Genome: Rescue Repeat Regions

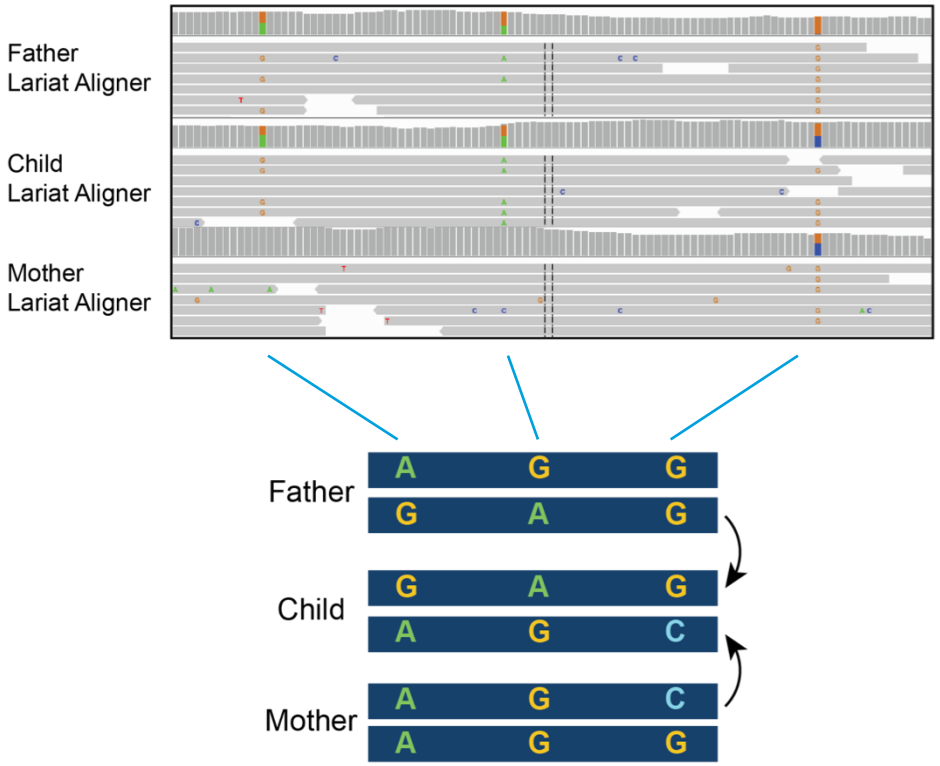


Chromium Genome: Eliminate Trio Analysis

Comparison of NA12878 aligned with Lariat and BWA



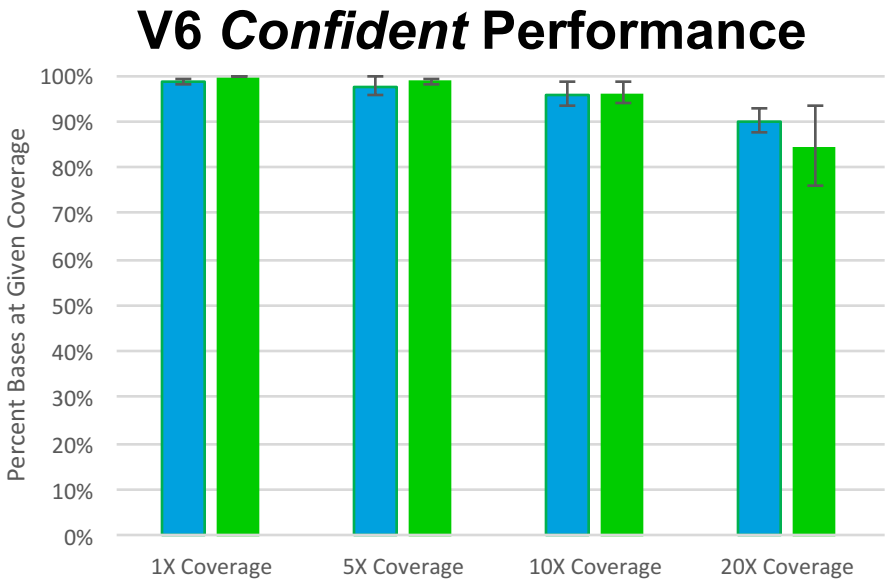
Trio validation of Lariat calls

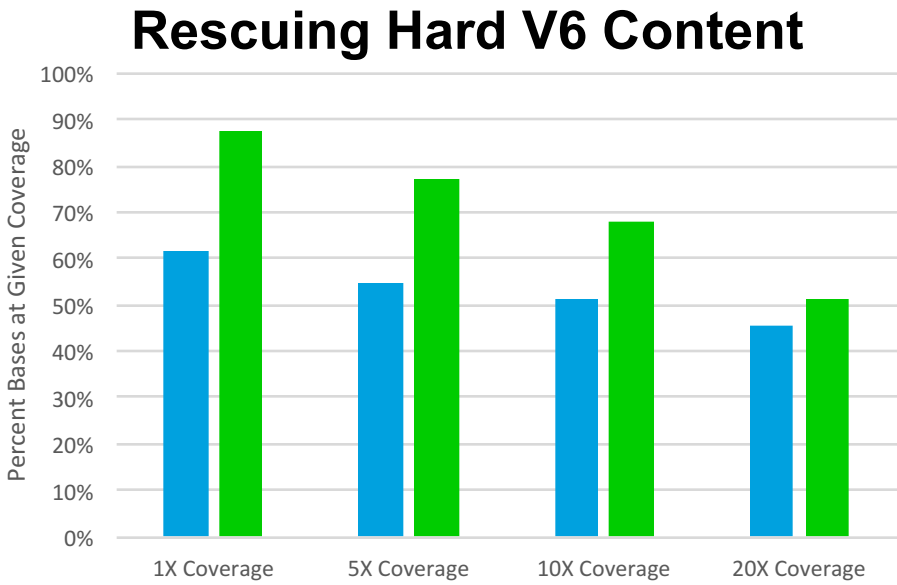


- NA12878 – Coriell sample from female child of NA12891 (P) and NA12892 (M)
- All three samples sequenced to 128Gb (ILMN 2x150)

Linked Reads for Exome (Hybrid Capture)

- Average of 5 different HapMap samples including NA12878
- V6 bait set; down-sampled to 6 Gb after removing barcode sequence; NIST Confident regions



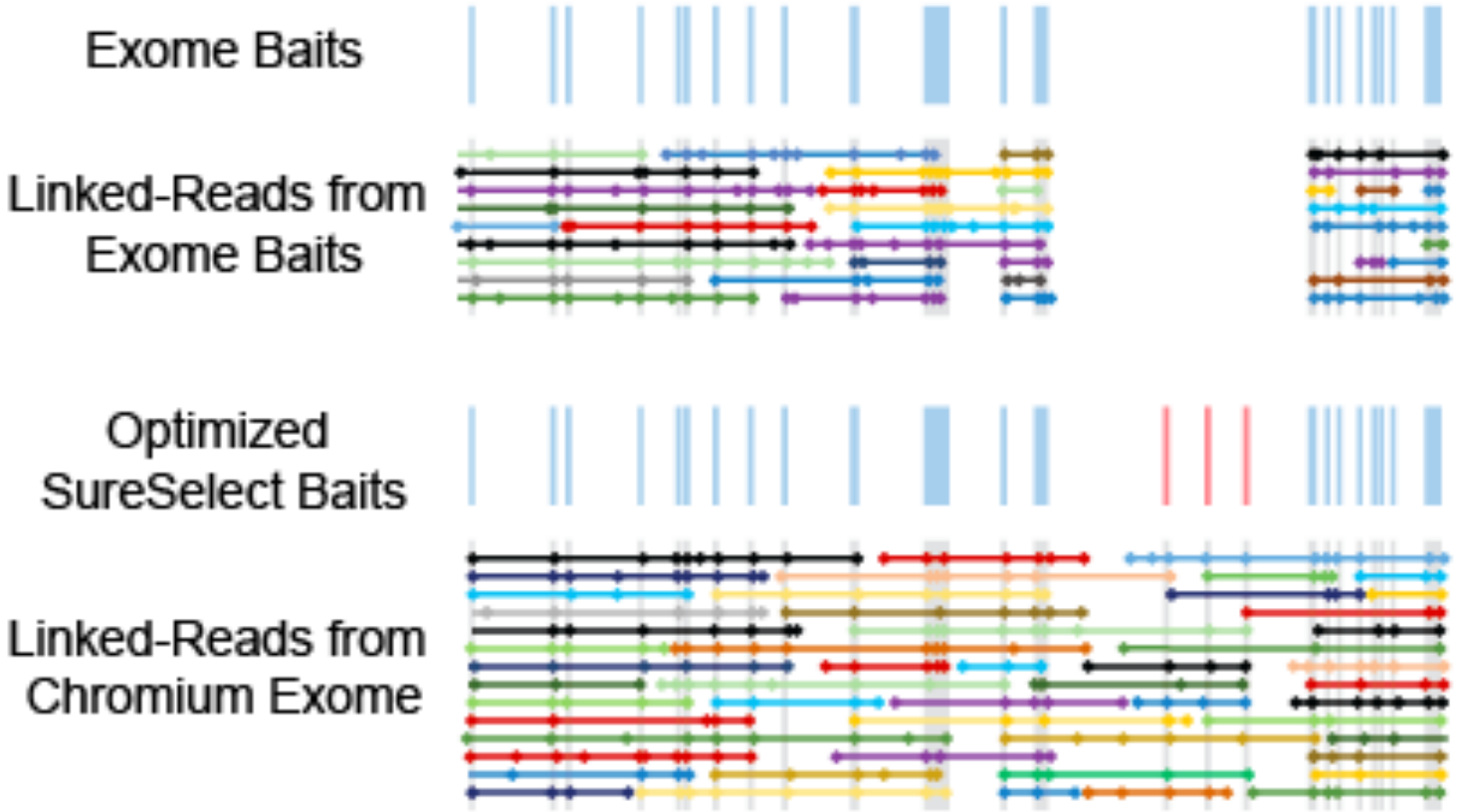


Agilent XT (200 ng) vs. Chromium Exome (1 ng)

Highly uniform and complex library from just 1 ng

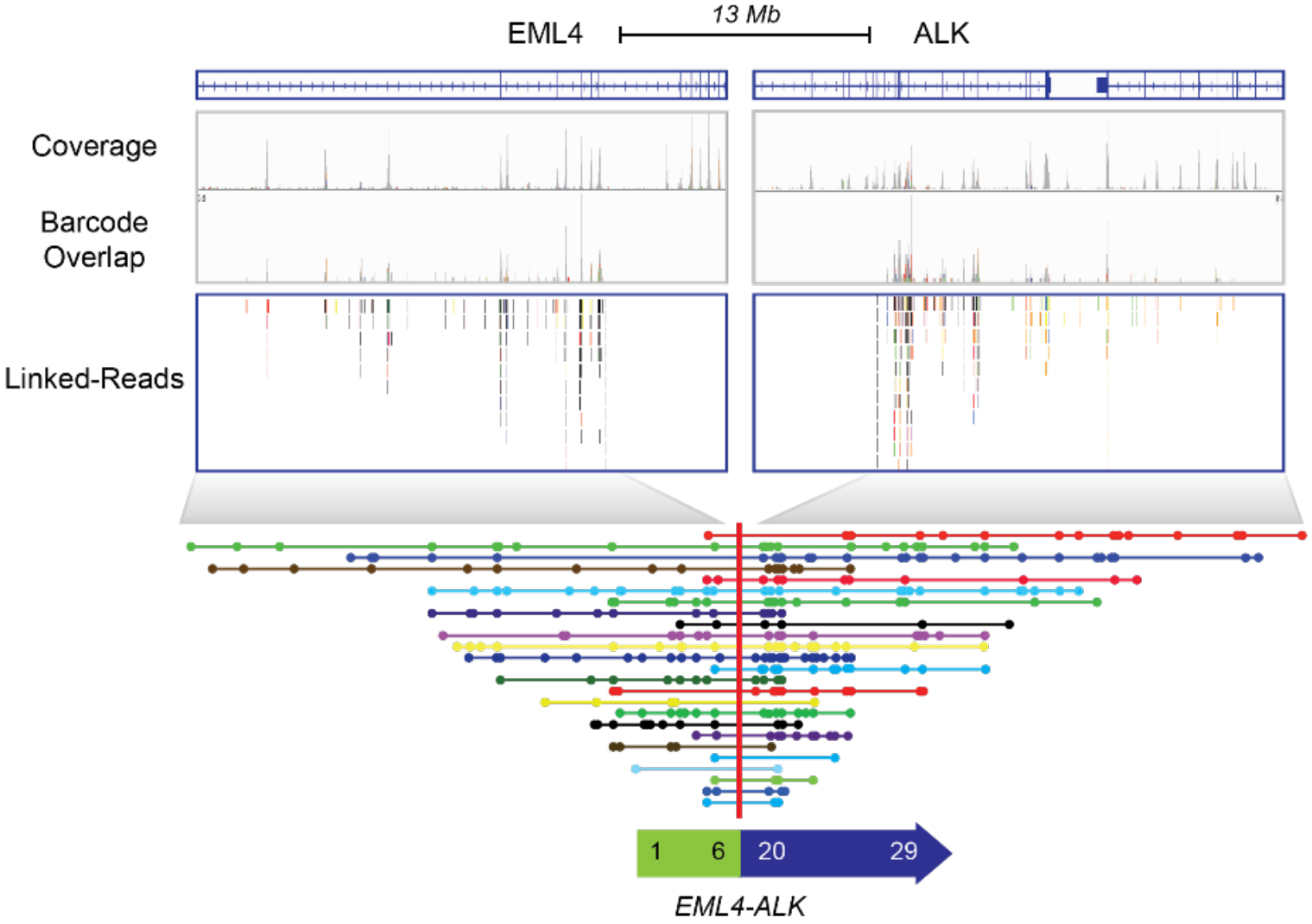
Optimized SureSelect Baits - Coming Soon

- Optimized SureSelect baits designed specifically for the Chromium system to improve gene phasing by closing gaps, and recovering hard-to-map loci in the genome.



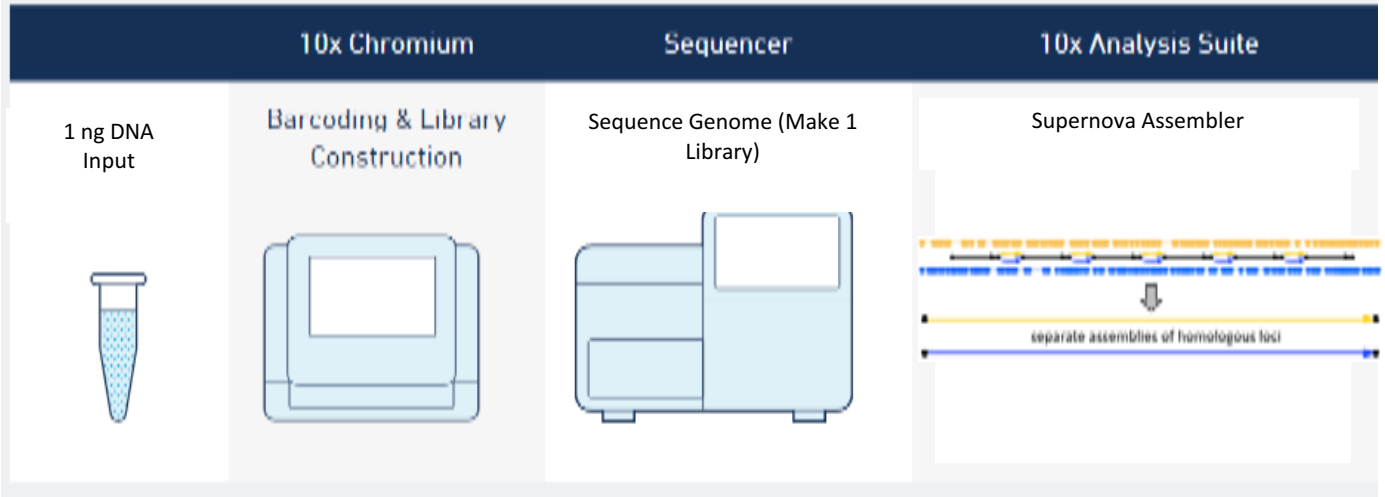
Identify Fusion Genes From Exome Data

- H2228: Non-small cell lung cancer cell line



Linked Read Applications with Chromium:
De Novo Assembly using Supernova

Supernova *De Novo* Assembly Workflow



DNA extraction (1hr-3 days)	Library prep (2 days)	Sequencers (< 3 days)	Supernova Human Assembly (2 days)
Assay requires 1ng of HMW DNA > 50 kb acceptable >100 kb preferred	Single linked read library per sample	• 2x150 reads • HiSeq recommended • >50X coverage recommended	• Run on single, standalone CentOS/RedHat or Ubuntu system • At least 24 cores • 512GB RAM • 2TB disk

Characteristics Of Supported Genomes

- Genome size: 1 - 3.2 GB total

- Ploidy
  - Haploid: not tested but likely to work well
  - Diploid: fully tested and supported
  - Polyploid: not tested, likely to require additional development

- Input should be from a single individual or clonal population

- Repeat content and G/C composition similar to human genome recommended.

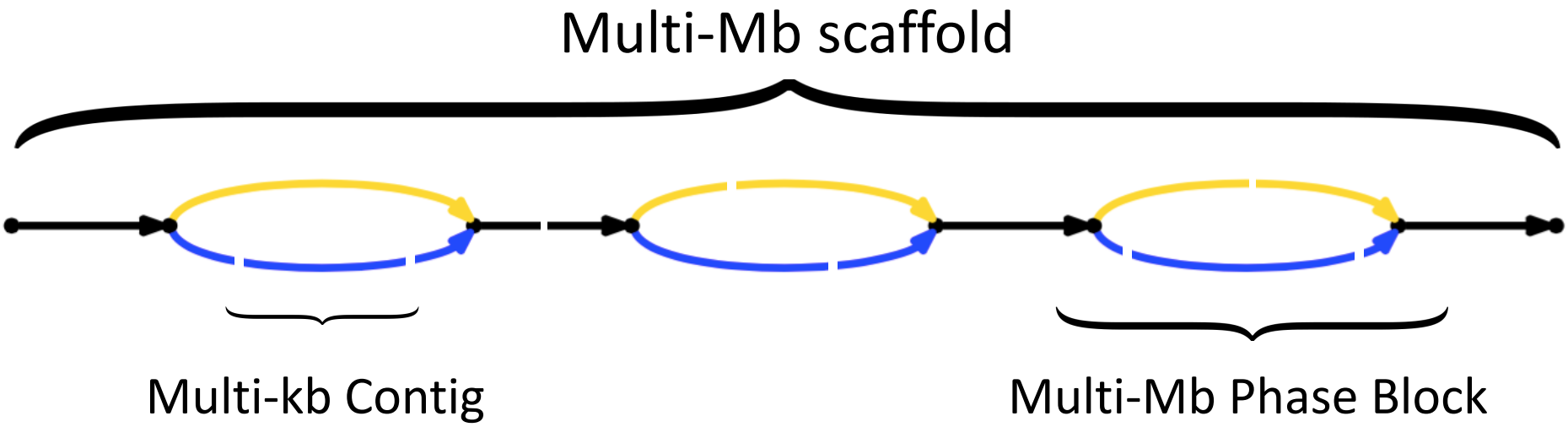
True Diploid Assembly From Linked Reads

Old Assembly Model: compress into a consensus

Supernova Assembly Model: represent both haplotypes

Church et al., 2011 PLoS Biology

Key Assembly Metrics/Terms



- **Contig**: an ungapped sequence
- **Scaffold**: a gapped sequence containing multiple contigs for which the order and orientation is asserted
- **Phase Block**: True diploid sequences; size is dependent on heterozygosity

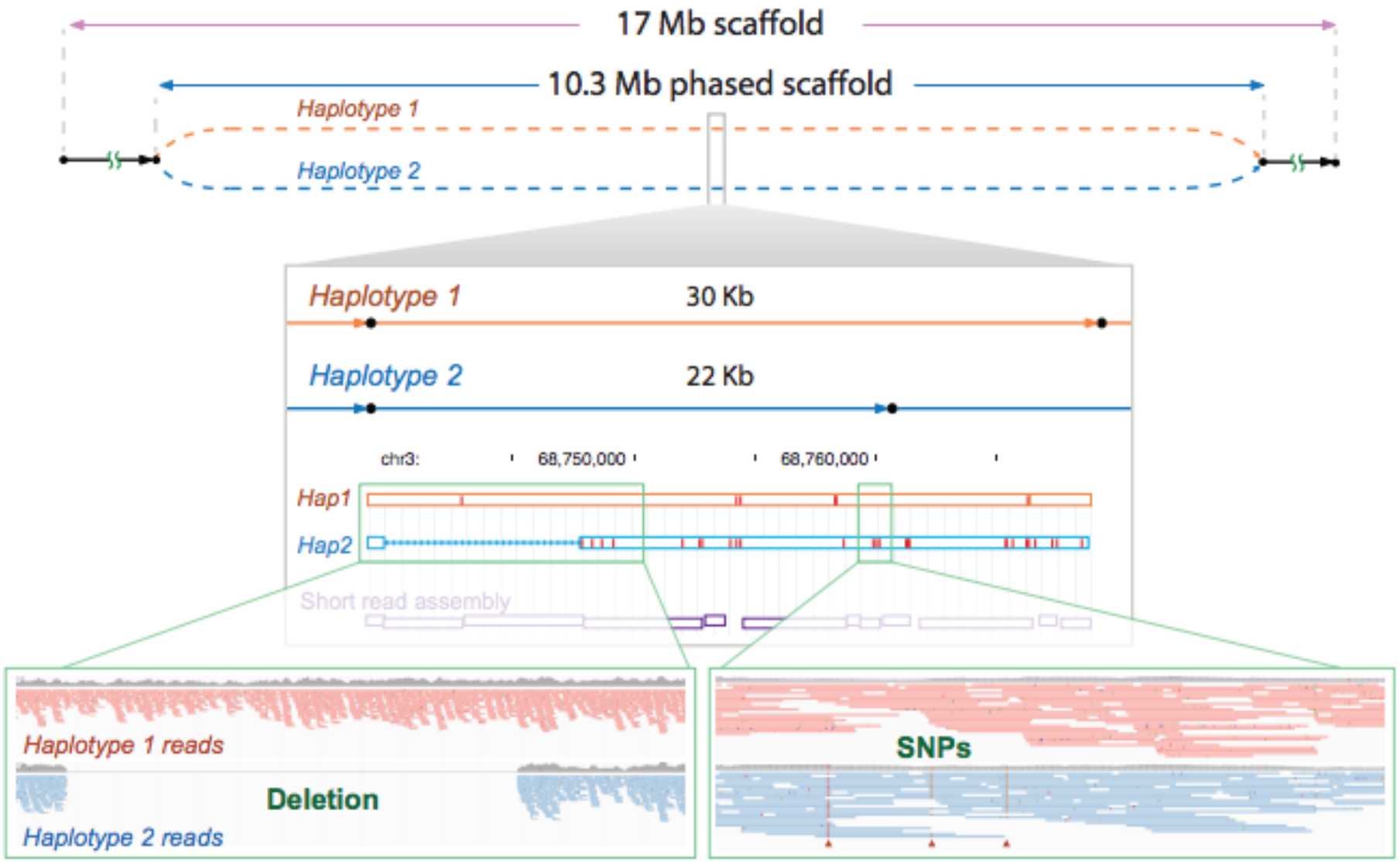
Supernova *De Novo* Assembly of NA12878

Sample Input and Sequencing	Supernova Compute Requirements	Supernova Assembly	
• Input DNA: 1.25ng • Molecule size: 80.3 kb • 2x150bp on Illumina HiSeq X Ten • 1200M reads (56x coverage)	• 28 cores (1 server) running for 48 hours • 1,344 total core-hours • 2TB of data generated	# of scaffolds > 10Kb	1651
		N50 contig size	103.67 kb
		N50 scaffold size	14.06 Mb
		Assembly size (scaffolds >= 10 kb)	2.74 Gb
		N50 phase block size	1.96 Mb

View and download at

http://software.10xgenomics.com/de-novo-assembly/overview/datasets

Phased *De Novo* Assembly



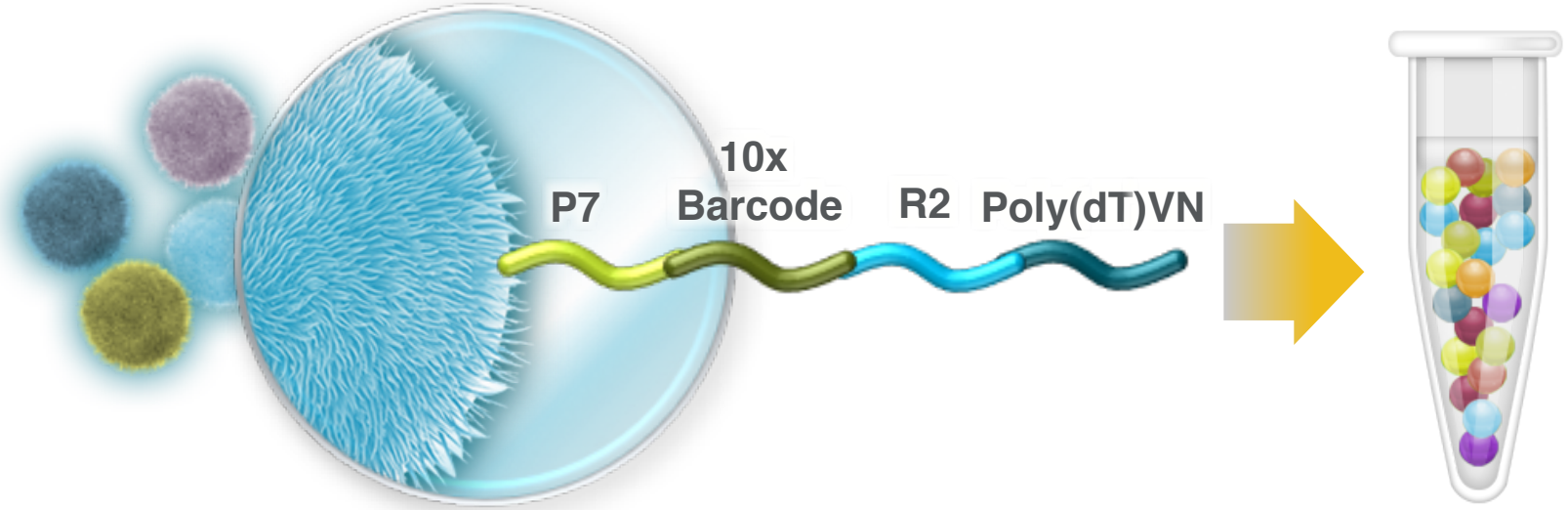
Supernova Assemblies From Diverse Genomes

	sample	Size (Gb)	DNA size (kb)	N50 contig (kb)	N50 scaffold (Mb)	HET SNP spacing (kb)	N50 phase block (Mb)
human	NA12878	3.2	95.5	85.0	12.8	1.7	2.8
human	NA24385	3.2	111.3	90.0	10.4	1.5	3.9
human	HGP	3.2	138.8	104.9	19.4	1.5	4.6
human	Yoruban	3.2	126.9	100.5	16.1	1.1	11.4
nonhuman	Komodo dragon	1.8	85.4	95.3	10.2	10.3	0.4
nonhuman	Spotted owl	1.5	72.2	118.3	10.1	7.1	0.2
nonhuman	Hummingbird	1.0	86.2	87.6	12.5	0.4	10.1
nonhuman	Monk seal	2.6	92.3	93.8	14.8	17.7	0.6
	Chili pepper	3.5	53.3	84.7	4.0	0.4	2.1

Single Cell 3’ RNA-seq with Chromium

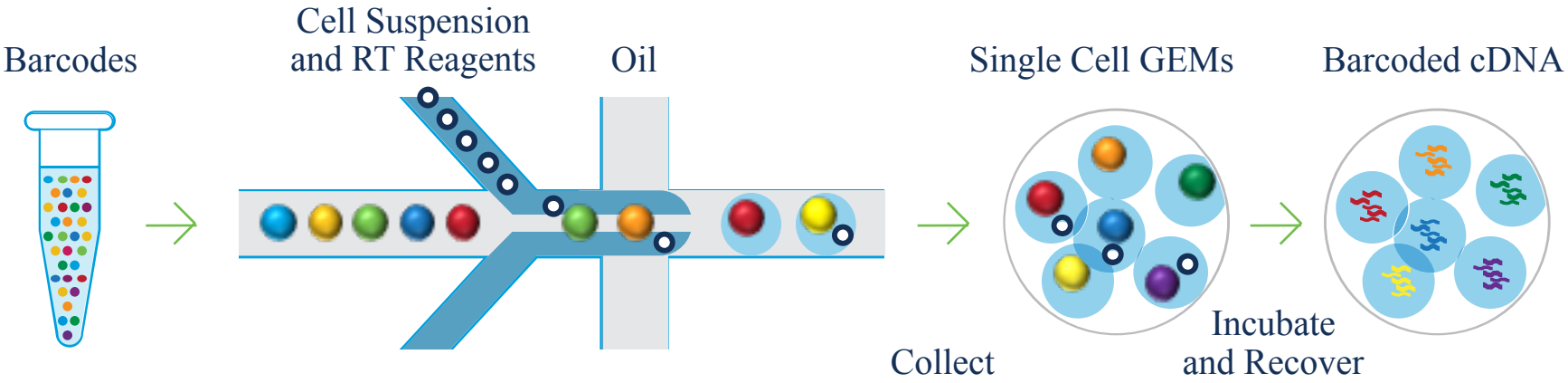
Gel Bead Oligo Delivery Substrate

Gel bead scaffold → Functional oligo with barcode → High-diversity library



- Millions of copies of identical oiligos
- Defined barcode sequence
- Built-in sequencing adapter and primer content

Partitioning and Barcoding of Single Cells



- High GEM fill ratio (~90% of droplets contain beads)
- Poisson loading of cells in GEMs
- Beads dissolve for efficient, liquid phase biochemistry
- Cell lysis starts immediately following encapsulation

Chromium Single Cell Sample Loading

Single-use microfluidics chip

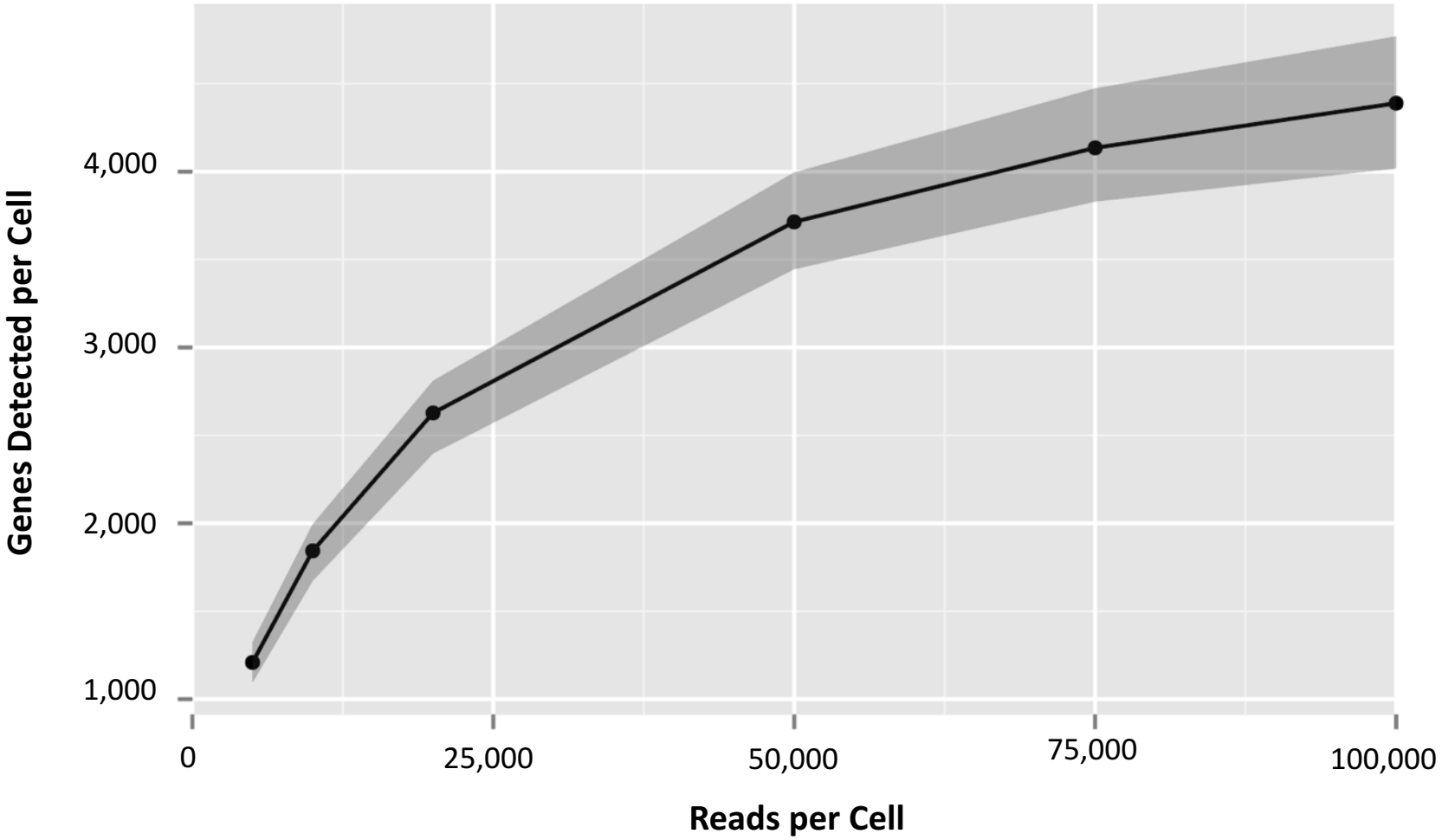
- Up to 8 channels processed in parallel
- 1,000 to 6,000 recovered cells per channel
- 10 minute run time per chip
- ~50 % cell processing efficiency

User controlled trade-off between cell numbers and doublet rate

Number of Recovered Cells	Expected Doublet Rate (%)*
1,200	~1.2
3,000	~2.9
6,000	~5.7

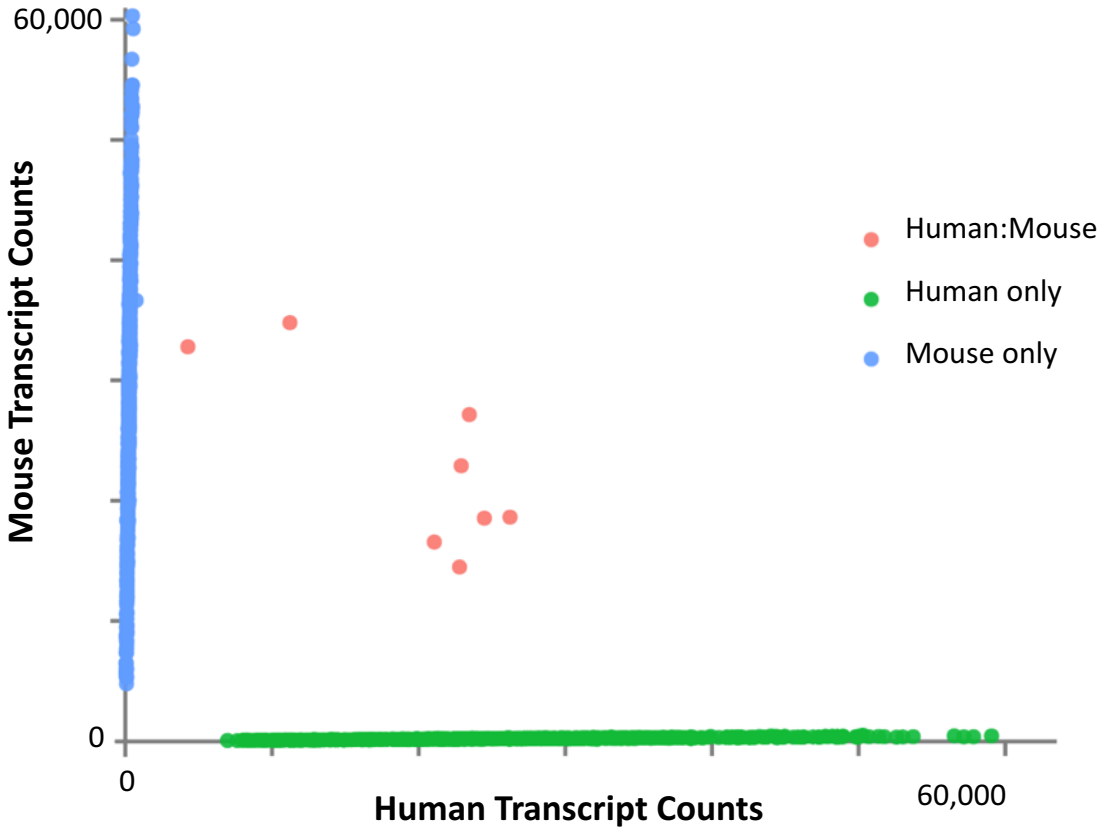
* Expected Doublet Rate assuming an ideal single cell suspension

Chromium Single Cell: Data Reproducibility



- Graph shows mean (black) and range (dark gray) over 16 independent experiments using HEK293T cells

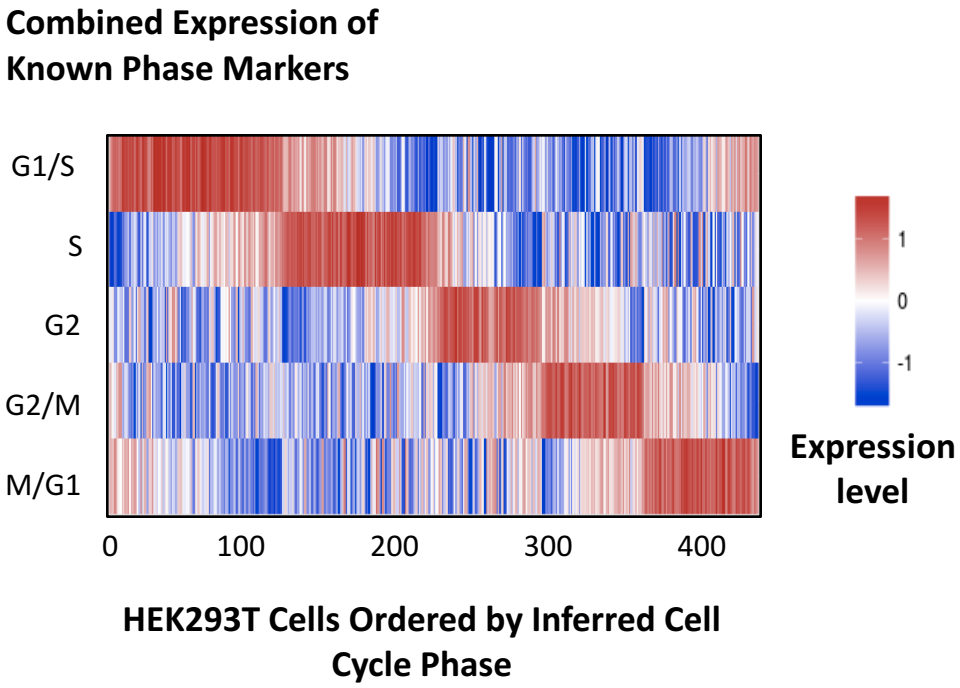
Chromium Single Cell: Doublet Rates



- 1:1 mixture of ~1,400 human (HEK293T) and mouse (NIH3T3) cells
- 99.4% of cell-occupied GEMs yielded reads mapping to only one species
- 1% inferred doublet rate*

- *includes unobserved human:human and mouse:mouse doublets
- Number of cells detected: ~1400 cells, Number of raw reads per cell: ~130k

Chromium Single Cell: Cell Cycle POP

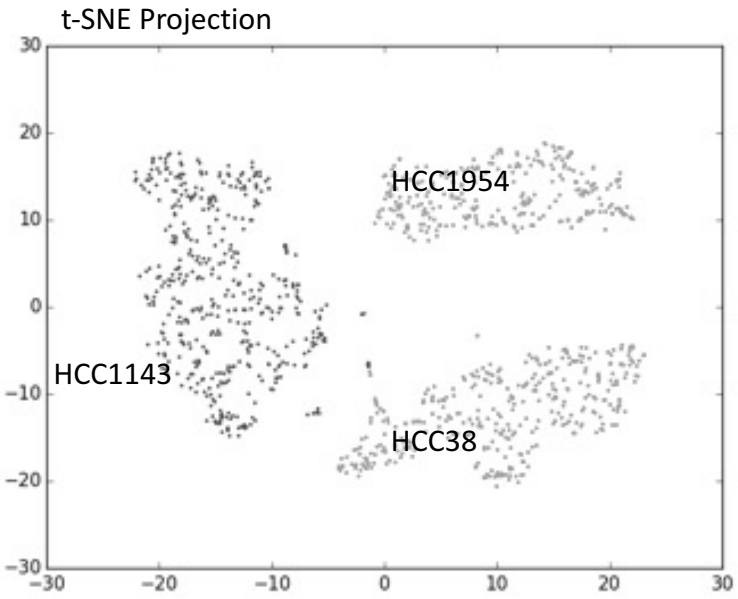


- Proliferating HEK293T cells were profiled and scored for expression of markers associated with each major cell cycle phase
- Cells from all phases were identified

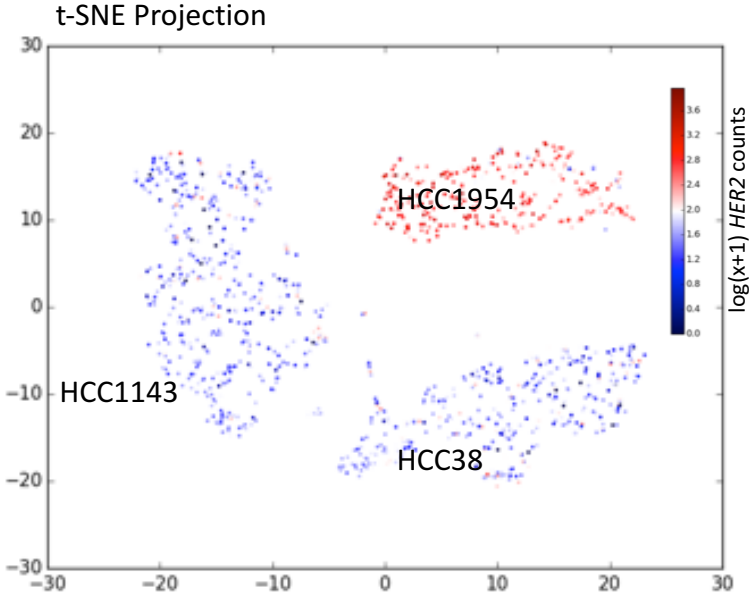
- Phase-specific genes derived from Whitfield et al., 2002
- Number of cells detected: ~400 cells, Number of raw reads per cell: ~40k

Data Clustering: No Cell Sorting Needed

Unbiased Automatic Clustering of Three Breast Cell Lines

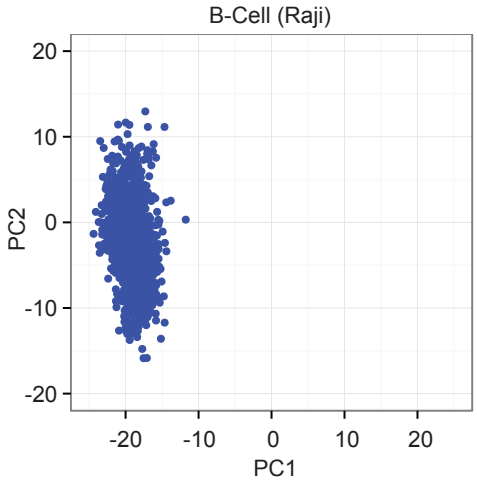


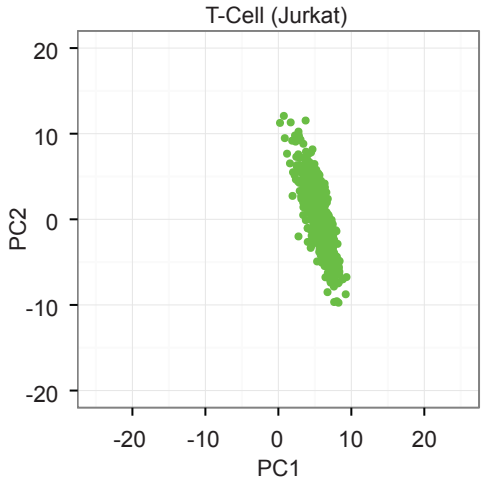
***HER2* Expression Matches Expected Cell Line Status**

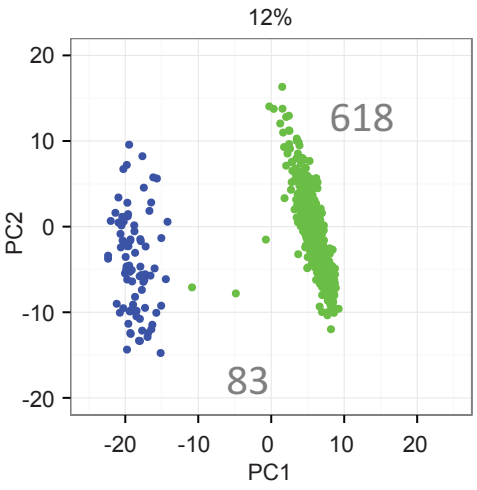


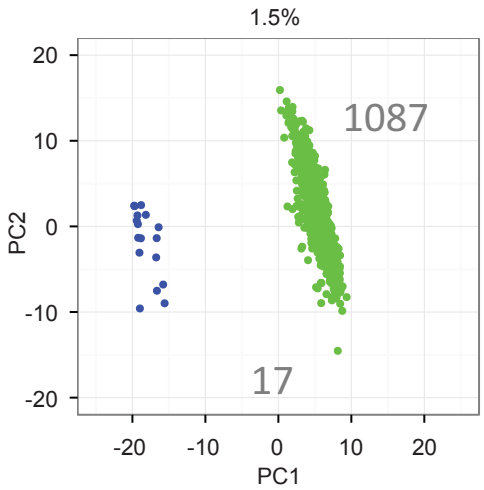
- Number of cells detected: ~1000 cells, Number of raw reads per cell: ~40k

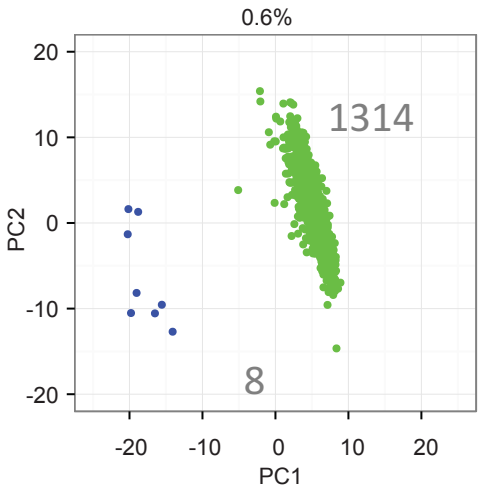
Identifying Rare Cell Types







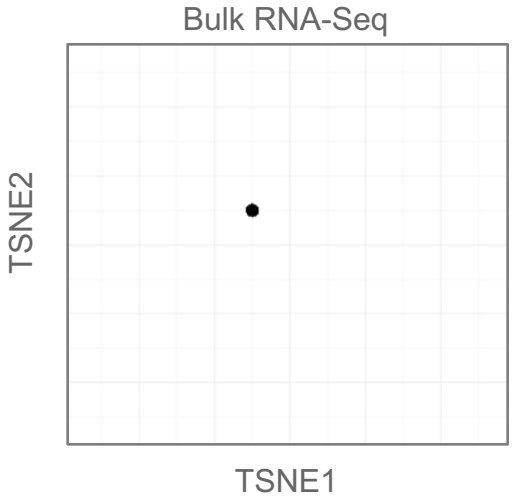




- Jurkat and Raji cells were combined at 9:1, 99:1 and 199:1 ratios and then profiled
- The minority Raji populations were identified in all three mixtures

- Number of cells detected: ~1000 cells, Number of raw reads per cell: ~60k

Increasing Cell Count Increases Resolution

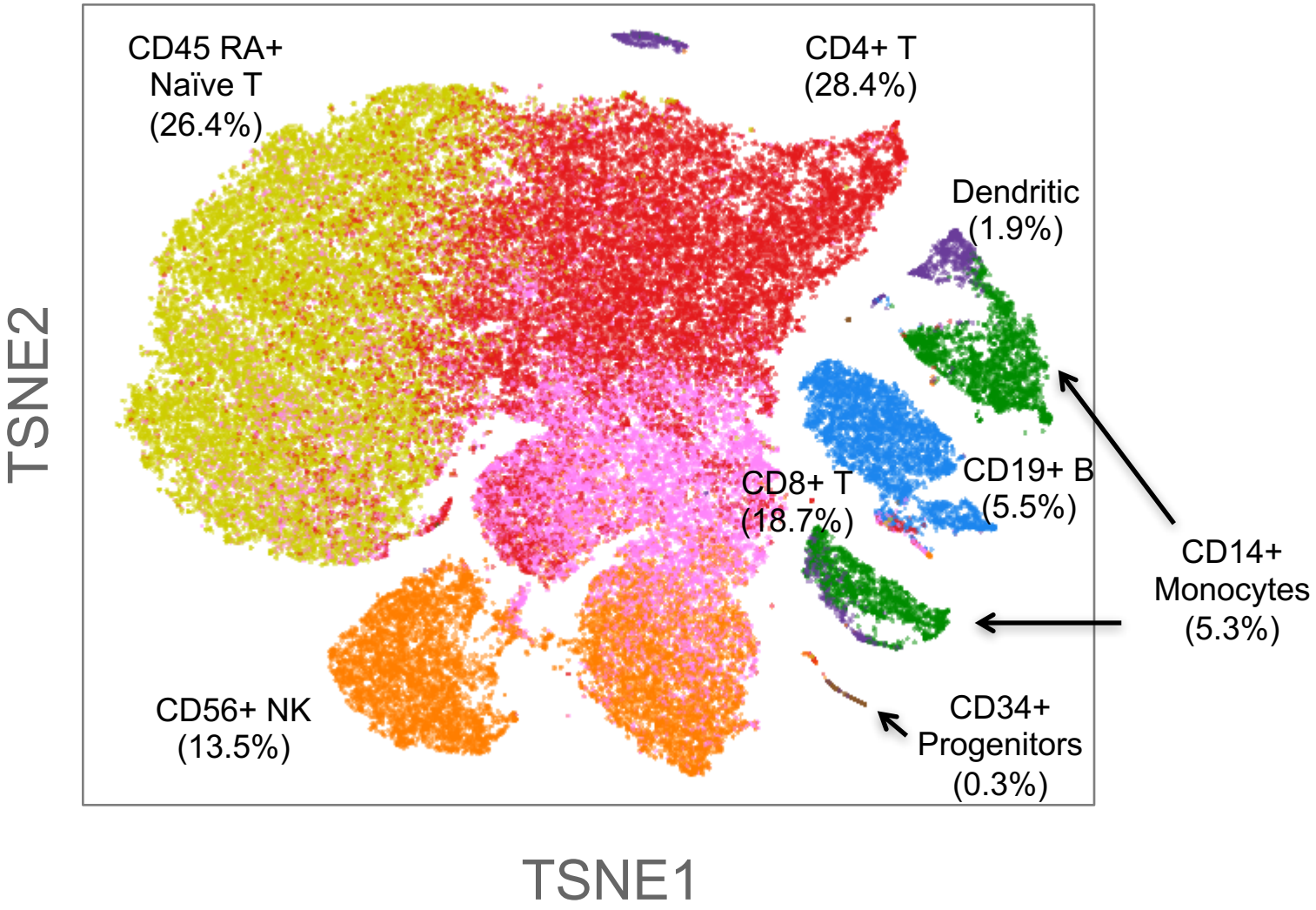


4,500 PBMCs

16,000 PBMCs

68,000 PBMCs

Major populations of PBMCs are detected



Chromium Single Cell: 68k PBMCs in One Run

- CD45RA+Naïve T Cells
- CD4+ T Cells
- CD8+ T Cells
- CD14+ Monocytes
- CD19+ B Cells
- CD34+ Myloid Progenitors
- CD56+ Natural Killer Cells

- Number of cells detected: ~68,000 cells, Number of raw reads per cell: ~21k

Chromium Software Suite Overview

Overview – Chromium Software Suite

Turn-key analysis and visualization software is included with all Chromium products:

Chromium Linked Read Data

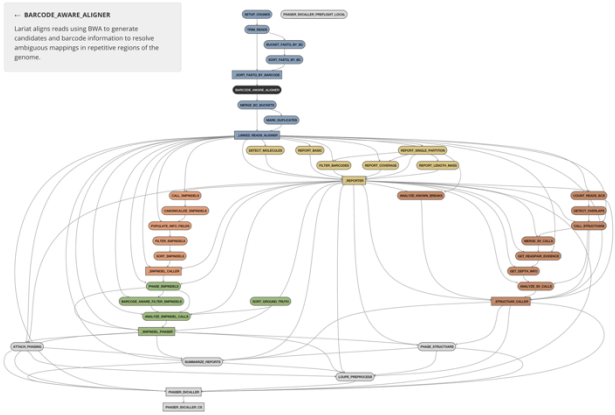
- Long Ranger Analysis Pipelines
- Loupe Genome Browser
- Supernova Assembler (Genome only)

Chromium Single Cell 3’ RNA-Seq

- Cell Ranger Analysis Pipelines
- Loupe Cell Browser (coming soon)

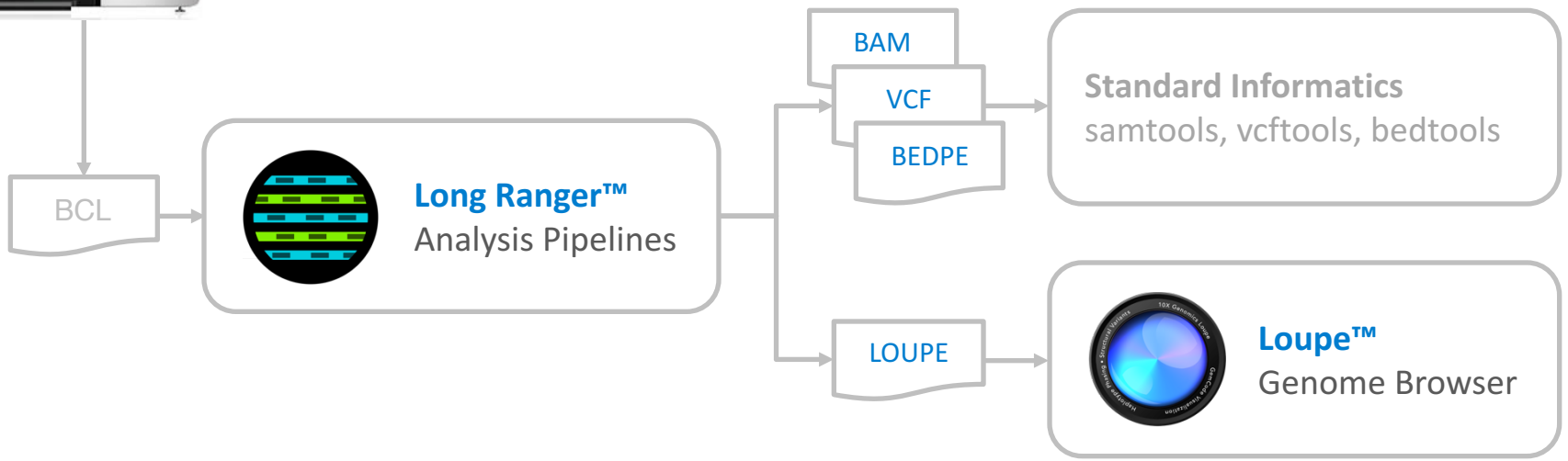
Chromium Software Platform

- **All software products built on a common platform**
  - Consistent look and feel and ease of use
- **Linux-based analysis pipelines**
  - Self-contained, simple to install
  - Run on workstations, and scale to clusters and cloud
  - Novel algorithms exploit Linked-Read data type
  - Multiple versions of pipelines can run side-by-side
- **Mac and Windows-based desktop visualization apps**



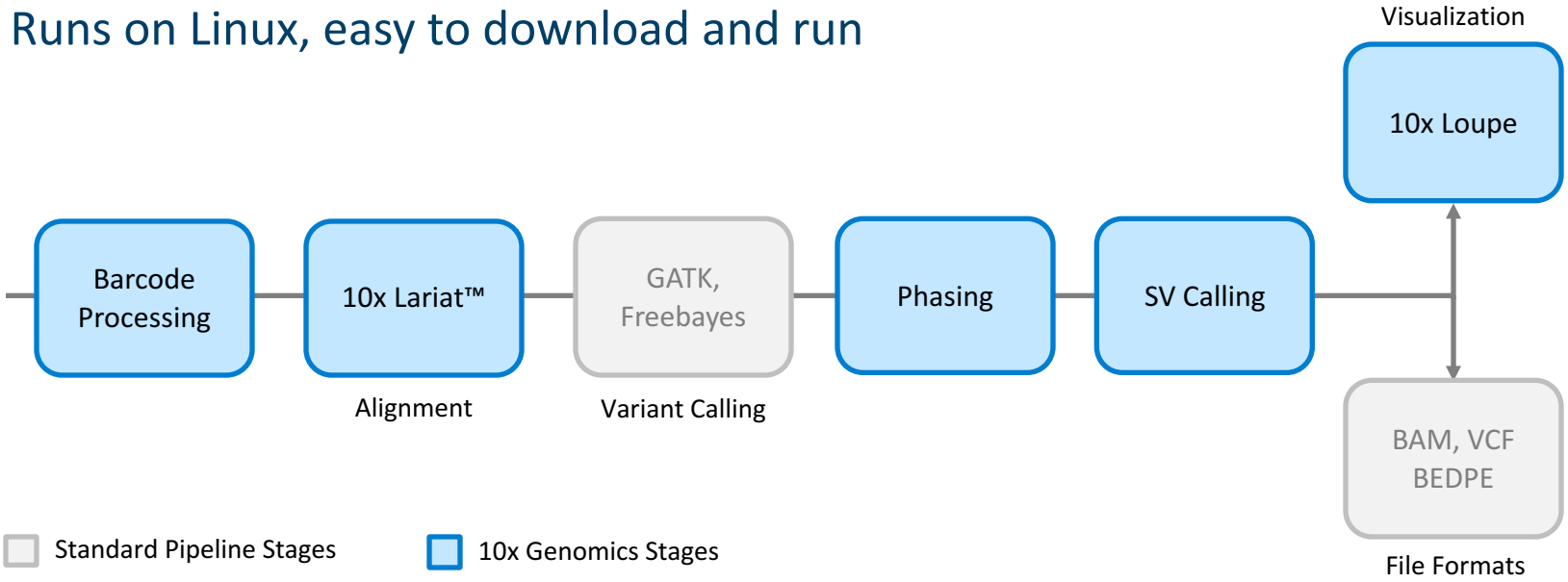
Linked Reads Informatics Workflow

- Complete software package for long range analysis
- Supports WGS and WES
- Outputs are standard formats plus Loupe visualization



Long Ranger Pipeline

- Complete, standalone pipeline optimized for human WGS and WES
- Exploits barcodes for alignment, variant calling, phasing, SVs
- Builds upon standard components and file formats
- Runs on Linux, easy to download and run



Long Ranger System Requirements

Local Mode

- Run on single, standalone Linux system
- CentOS/RedHat 5.2+ or Ubuntu 8.04+
- 16+ cores, 128GB RAM, and 2TB+ disk

Cluster Mode

- Run on SGE and LSF
- Each node must have 8+ cores and 8GB+ RAM/core
- Shared filesystem between nodes (e.g. NFS)

Runtime

- 30X (100Gb) Genome: 640 core-hrs (36 hrs on 20 cores)
- 10Gb Exome: 148 core-hrs

Loupe: Visualize Long Range Information

Loupe is a precision tool for inspecting GEMs

- Runs on output of Long Ranger pipeline
- Fully haplotype-enabled genome browser
- Visualization of breakpoints and structural variants
- Desktop application for Mac and Windows

Loupe: Summary View

Loupe · NA12878 · Whole Genome

History · New Window · Open File...

SUMMARY HAPLOTYPES STRUCTURAL VARIANTS LINKED-READS Search by gene name or locus

SNPs Phased
99.5%

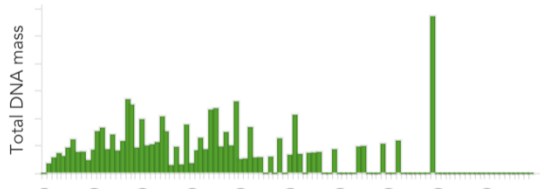
Longest Phase Block	GEMs Detected
39,989,694	1,267,827

GEM Performance

GEMs Detected	1,267,827
N50 Linked-Reads per Molecule (LPM)	71.0
Mean DNA per GEM	414,642 bp

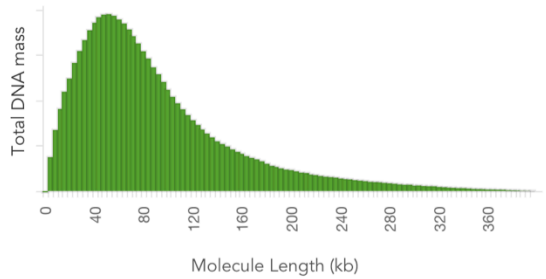
Phasing

SNPs Phased	99.5%
Longest Phase Block	39,989,694 bp
N50 Phase Block	14,428,091 bp



Input DNA

Molecule Length	μ 95,238 bp σ 73,250
DNA in Molecules >20kb	92.4%
DNA in Molecules >100kb	34.9%
Estimated DNA Loaded	0.191 ng



Structural Variants

Structural Variant Calls	22

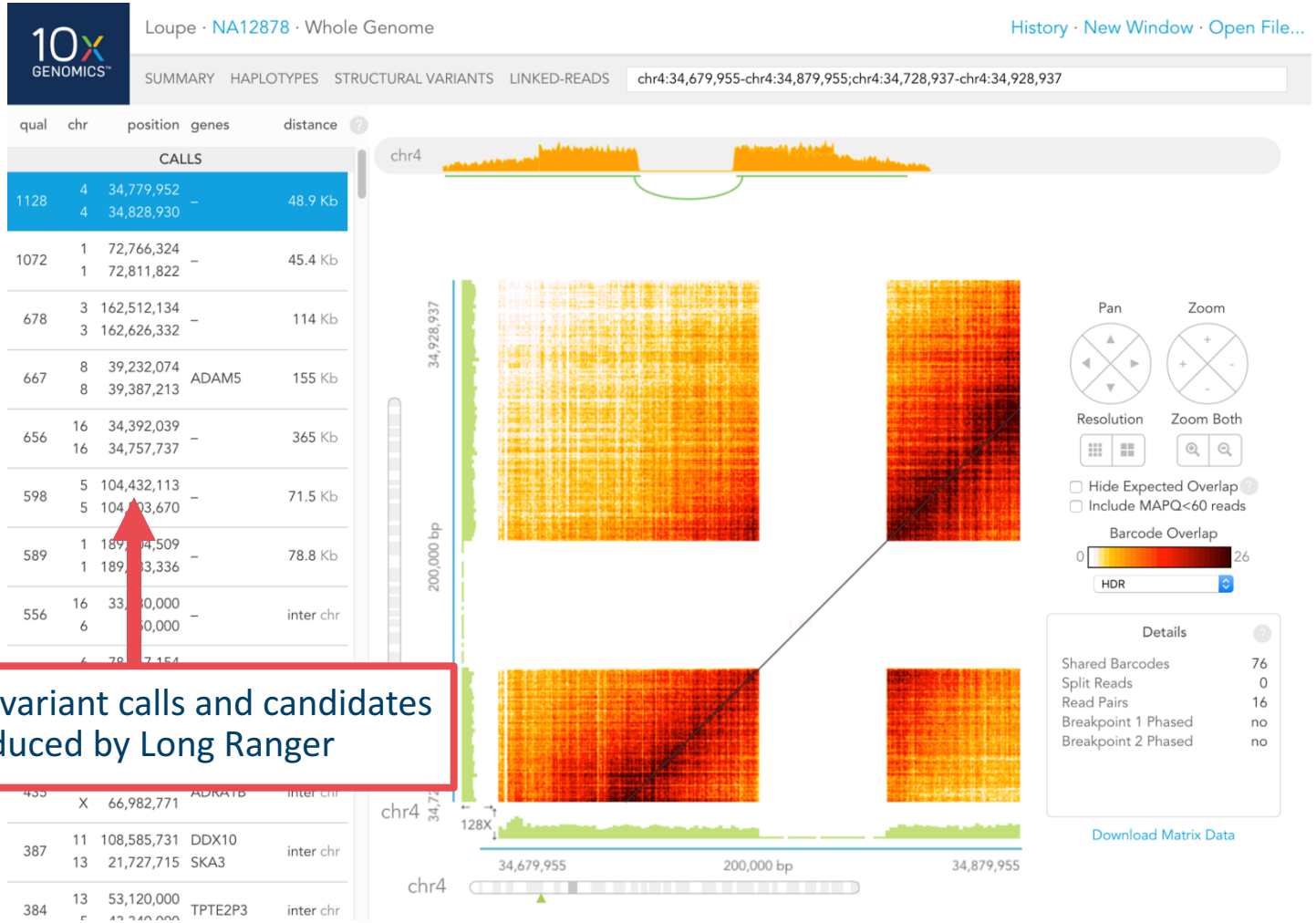
Sequencing

Number of Reads	1,024,463,130
Median Insert Size	382 bp
Mean Depth	40.0 X
Zero Coverage	8.24%

Loupe: Haplotype View

Fluidly search for and browse haplotype-resolved variants at multiple loci at once

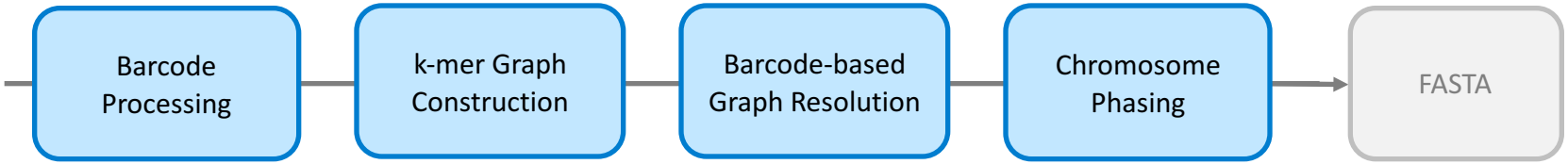
Loupe: Structural Variant View



Structural variant calls and candidates produced by Long Ranger

Supernova Informatics & Pipeline Overview

- Complete software package for *De Novo* assembly
- Demultiplexing to FASTA generation for contigs and scaffolds



File Formats

Supernova System Requirements

Local Mode

- Run on single, standalone Linux system
- CentOS/RedHat or Ubuntu
- 24+ cores, 512GB RAM, 2TB+ disk

Runtime

- 180Gb genome: 500 core-hours (36 hr on 28 cores)

Single Cell Informatics Workflow

- Complete software package for single cell analysis
- Outputs are standard formats plus Loupe visualization

BCL

Cell Ranger™
Analysis Pipelines

BAM
HDF5
MEX

Standard Informatics
samtools, Python, R

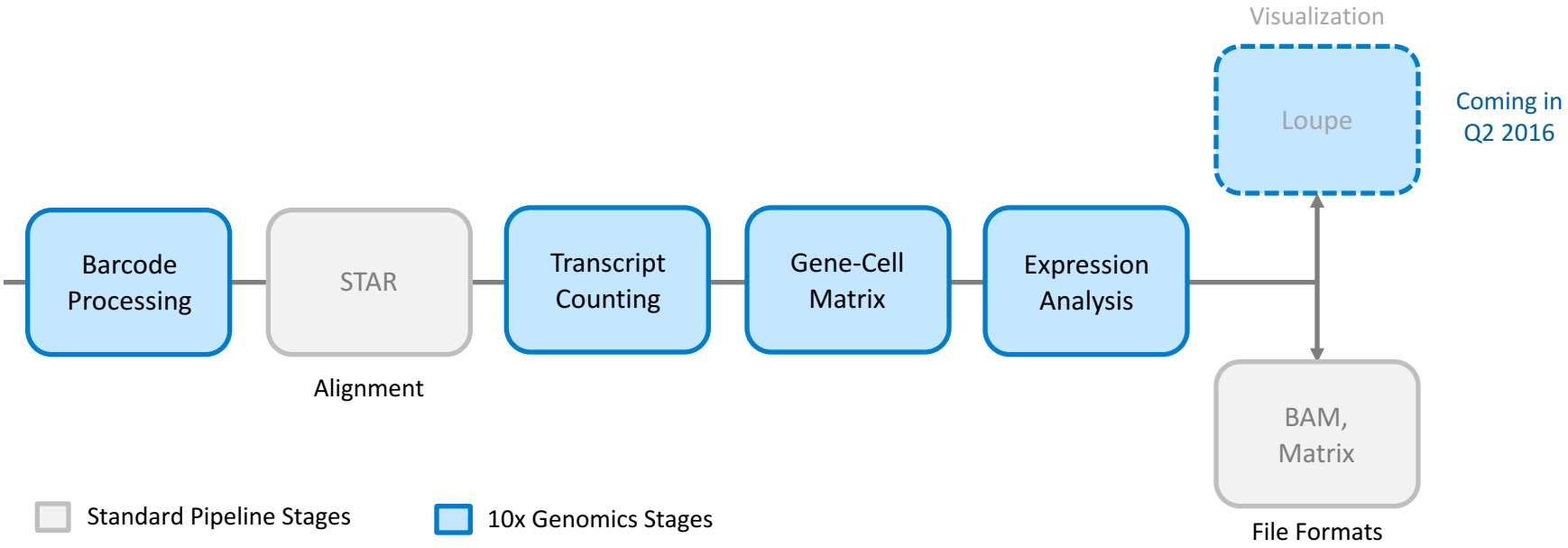
LOUPE

Loupe™
Cell Browser

Coming in Q2 2016

Cell Ranger Pipeline

- Complete, standalone single cell, gene expression pipeline
- Builds upon standard components and file formats
- Runs on Linux, easy to download and run



Cell Ranger System Requirements

Local Mode

- Run on single, standalone Linux system
- CentOS/RedHat 5.2+ or Ubuntu 8.04+
- 8+ cores, 64GB RAM

Cluster Mode

- Run on SGE and LSF
- Each node must have 8+ cores and 8GB+ RAM/core
- Shared filesystem between nodes (e.g. NFS)

Runtime

- 50 core-hours per 100M clusters
- 5000 cells, 40k reads/cell: 95 core-hours

Software Website

For more information, downloads,

file formats, and specifications:

http://software.10xgenomics.com

Thank you!

Questions?