Frederick National Laboratory for Cancer Research

CCR-Sequencing Facility

Jyoti Shetty

Jan 23, 2017

CCR-Sequencing Facility

Frederick National Laboratory for Cancer Research

The Mission of the Sequencing Facility is to utilize high-throughput sequencing technologies to enrich cancer research and ensure that the NCI community can remain at the leading-edge of next-generation sequencing technology.

Location

- Advanced Technology Center (ATC), Gaithersburg, MD (March, 2009)
- Advanced Technology Research Facility (ATRF), Frederick, MD (moved Oct, 2012)

Current Staff

- Director
- Team Leads
- R&D Scientist
- Research Associates
- QA/QC Specialists
- Bioinformaticians
- Other matrix support from throughout organization

CCR Sequencing Facility

Frederick National Laboratory for Cancer Research

Core facility dedicated to providing Next-Generation Sequencing (NGS) services to Center for Cancer Research investigators, operating one Illumina Hiseq 3000, two HiSeq 2500s, Three NextSeq 500, one MiSeq, one Pacbio RSII, 10X Chromium and BioNano Irys system.

DNA/RNA-Protein Interaction

Transcript Profiing and Discovery

Genome-wide Methylation study

Targeted Exome sequencing

Whole Genome Structural Variants

Single Cell Gene Expression Profiles

Illumina

De Novo Whole Genome Sequencing

Targeted Sequencing

Base Modifications

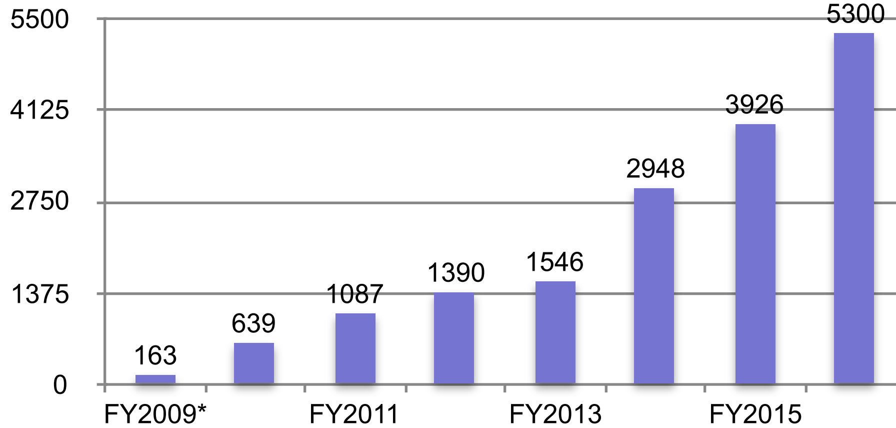
PacBio RSII

Transcript Isoform Detection

Services continue to be dedicated to and subsidized for CCR investigators.

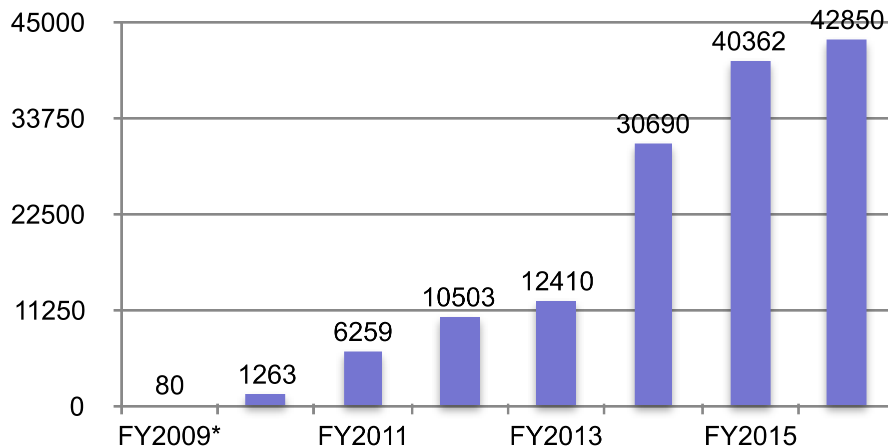
Provided service to more than 100 CCR Labs in FY2016

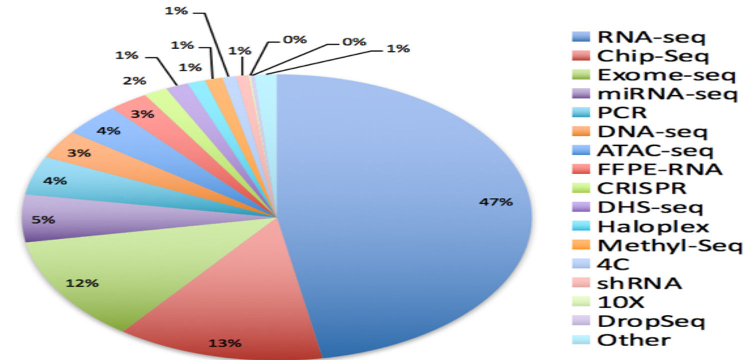
Frederick National Laboratory for Cancer Research



FY2016 Statistics:

- Serviced 109 CCR investigator labs
- Completed more than 5300 samples.
- Delivered over 43 trillion bases





CCR-Sequencing Facility

Sequencers:

HiSeq 3000:

- Read Length (75 bpx2, 150bpx2)
- Higher Yield (650-750 Gb/run),
- Runtime (3.5 days/run)

HiSeq 2500:

- Longer Read Length (125bpX2)
- Higher Yield (450-500 Gb/run),
- Runtime (6 days/run)

NextSeq 500:

- Desktop sequencer
- Fast turnaround (~11 hrs)

MiSeq :

- Low throughput desktop sequencer
- Fast turnaround (~5 hrs)

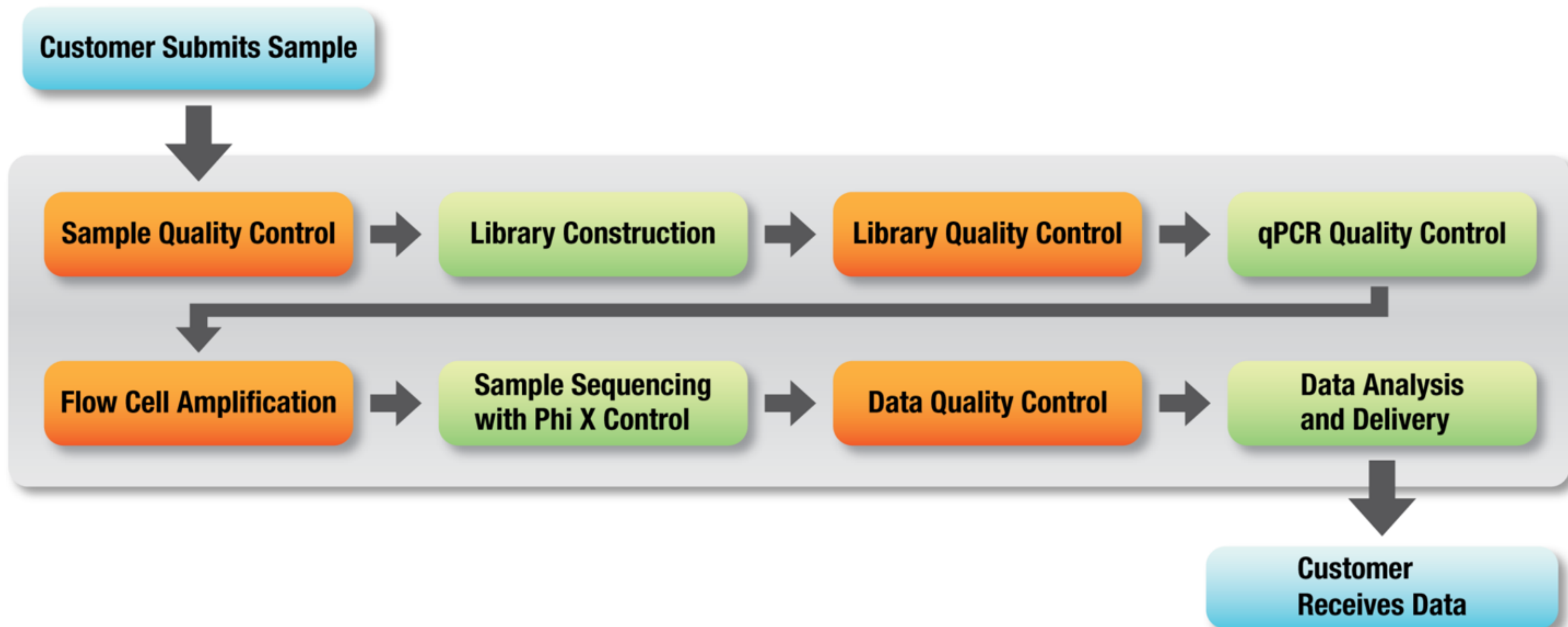
PacBio RSII:

- Long Read (avg.25 kb reads), 26 Gb/SMRT cell

SF Quality Control Process

Frederick National Laboratory for Cancer Research

At CCR–SF, producing high-quality data is our highest priority. To ensure this, we have quality control checks in place throughout the entire production pipeline.



Sample Requirements for Illumina Sequencing

Frederick National Laboratory for Cancer Research

Sample Quantity/Quality Requirements and Recommendations:

Type of Library	*Minimum DNA/RNA Requirement for Library Construction*	*Recommended DNA/RNA for Optimal Library Construction*	*Maximum Sample Volume Requirement for Library Construction*	*Additional requirements*
ChIP DNA Sequencing	5 ng	10 ng	30 µL	Bulk of the DNA fragments in the 100-300 bp range
ChIP-seq Ultra low input	**10 pg**	**1 ng**	**30 µL**	**Bulk of the DNA fragments in the 100-300 bp range**
gDNA Sequencing	100 ng	1 µg	30 µL	DNA should be as intact as possible with no contamination, OD260/280 1.8–2.0
mRNA Sequencing	100 ng	1 µg	50 µL	RIN should be at least 8.0, DNase treated
mRNA ultralow Clonetech	**100 pg**	**10 ng**	**10 µL**	**RIN should be at least 8.0, DNase treated**
microRNA Sequencing	100 ng	1 µg	6 µL	
Total RNA sequencing	100 ng	1 µg	10 µL	FFPE and degraded RNA can be used; however high quality RNA gives better data
Swift Methyl-Seq	50ng	50ng	30 µL	>2ng/ul

Sample Requirements for Illumina Sequencing

	Sample condition/ quality	Amount	Concentration
10XChromium single cell 3'mRNA-seq	washed, ready-to-load cells in PBS+0.04% BSA	~10,000 cells	1000 cells/ul
	Cells transferred in culture medium, to be washed at SF	~25,000 cells	Any
10XChromium Genome (SV analysis, whole genome or exome sequencing)	High molecular weight DNA (50-150kb)	20ng	>1ng/ul
Optical mapping using BioNano Irys	Cells	3-4 million cells	Any

Sample Requirements for PacBio Sequencing

Frederick National Laboratory for Cancer Research

Sequencing Application	Minimum DNA Quantity Requirement	Minimum DNA Quality Requirement
Circular Consensus (< 5kb fragment)	500 ng	OD260/280 1.8-2.0 OD260/230 1.7-2.2
Continuous Long Reads (5 - 10kb fragment)	1 ug	
Continuous Long Reads (> 10kb fragment)	5 ug	
Iso-seq (total RNA)	500 ng	RIN > 9
Low-Input	Please Inquire	

Criteria for sample submission

Frederick National Laboratory for Cancer Research

- Please QC your samples before you ship them to our facility to ensure they meet our minimum sample requirements.
- If you do not have access to a bioanalyzer you can contact :

  Liz Conner at connerl@dc37a.nci.nih.gov located at 37 Convent Dr. Rm 1044, Bethesda, MD, 20892.

- Ensure all samples are shipped in dry ice, in individual (1.5-2 ml) tubes labeled clearly.
- Please do not send samples in PCR strip tubes.

- The volumes of the samples in each tube should be in the 5 -30 ul range.

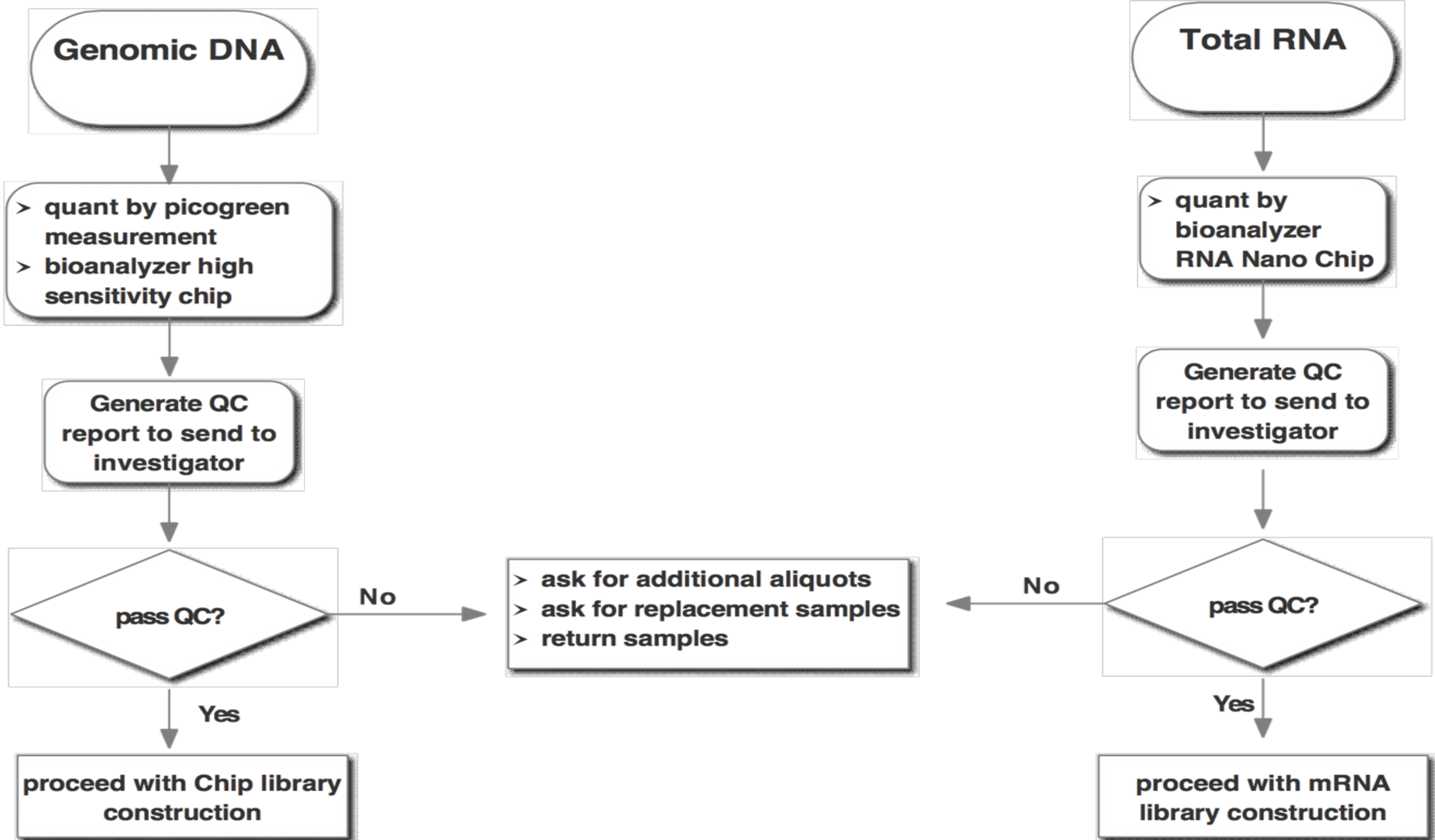
Criteria for sample submission

Frederick National Laboratory for Cancer Research

- **E-mail a copy of the sample manifest form and also include a copy of the form along with your sample shipment.**
- **Purified total RNA should be eluted in nuclease-free water. Please include a DNase step with the RNA isolation method for your samples.**
- **You can elute/resuspend your DNA samples in water. Low salt buffers such as TE with reduced EDTA (10 mM Tris, pH 8.0; 0.1 mM EDTA) are also fine.**

Sample QC for the SF Illumina Laboratory

Frederick National Laboratory for Cancer Research



General Recommendations to choose Sequencing platforms

Frederick National Laboratory for Cancer Research

Application	Sequencer
Chip Seq	NextSeq, Single Read 75 bp
Micro RNA	NextSeq, Single Read 75 bp
RNA Seq	HiSeq 2500 Paired End 125 cycles
WGS	HiSeq 3000 Paired End 150 cycles
Targeted Exome	HiSeq 2500 or 3000 Paired End
Single Cell	NextSeq Paired End 75 bp
Denovo WGS (small genome)	PacBio
Full Length Transcript	PacBio
Targeted Sequencing	PacBio

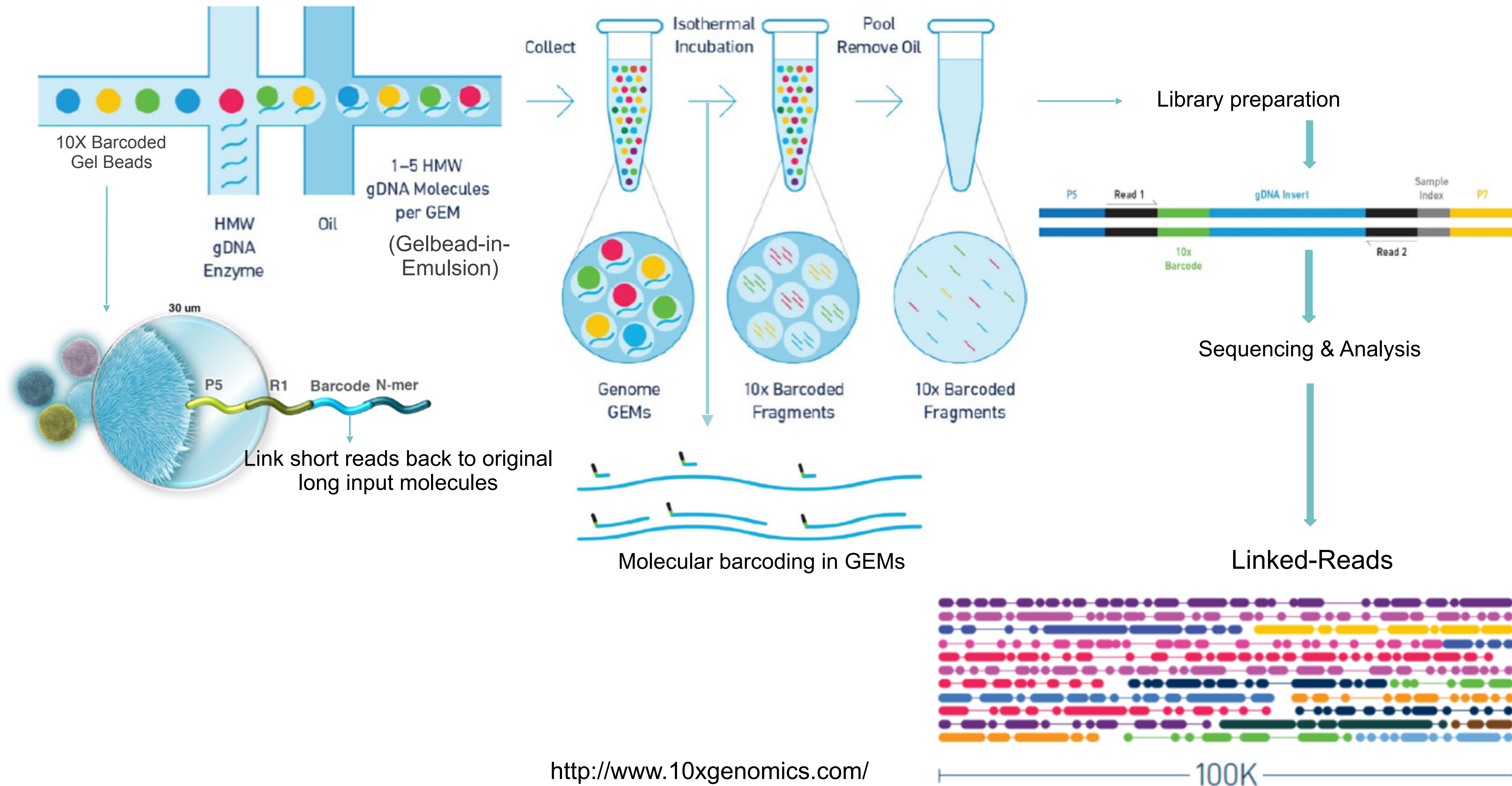
Recent additions

Frederick National Laboratory for Cancer Research

1. **Structural variation** in DNA
    1. 10XGenomics – WGS and target capture
    2. Genome mapping using BioNano Irys
2. **Single cell mRNA-seq**
    1. 10XChromium (3' mRNA-Seq)
        - Combines microfluidics with molecular barcoding to enable high-throughput single cell RNA sequencing
        - Helps uncover the variation masked by the average of heterogeneous subpopulations.
        - Transcriptional profiling of thousands of individual cells.
3. **Full length Iso-Seq**
    - Sequence full-length mRNA transcripts, no assembly required.
    - Complete information about alternatively spliced exons, transcriptional start sites, polyadenylation sites and strand orientation.
4. **DNA methylation analysis**: WGBS & targeted methylome sequencing
    - Increased mapping and improved detection of methylated cytosines with lower sample input requirement

10XGenomics – Overview

- Long range sequence information from short read sequencing by **partitioning and molecular barcoding**
- Delivers structural variants, haplotypes, and other valuable long range information for targeted, exome, and whole genome sequencing.
- >10Mb haplotype blocks with phased SNVs, indels, & SVs

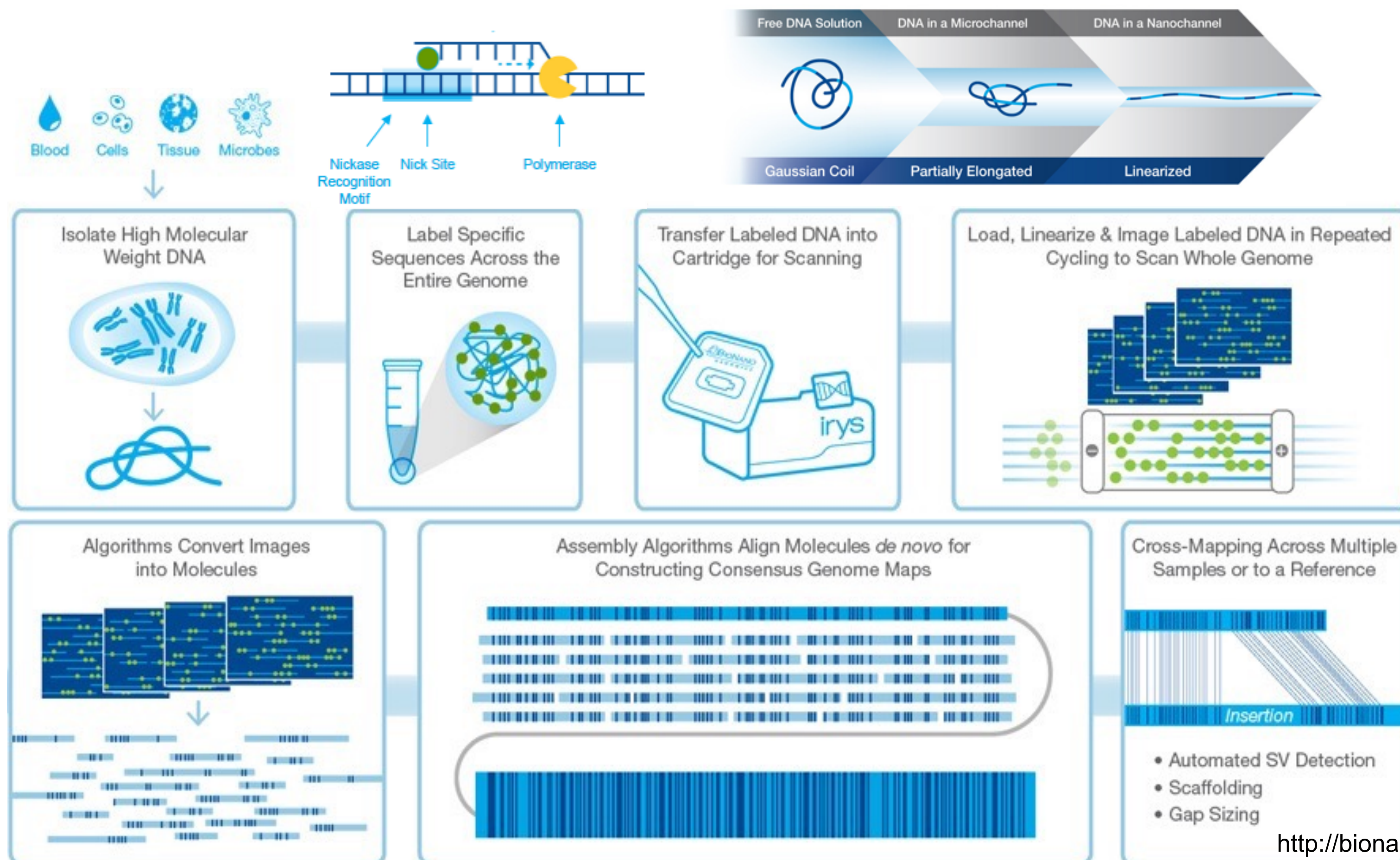


http://www.10xgenomics.com/

Next Generation Mapping with Irys System from BioNano Genomics

Frederick National Laboratory for Cancer Research

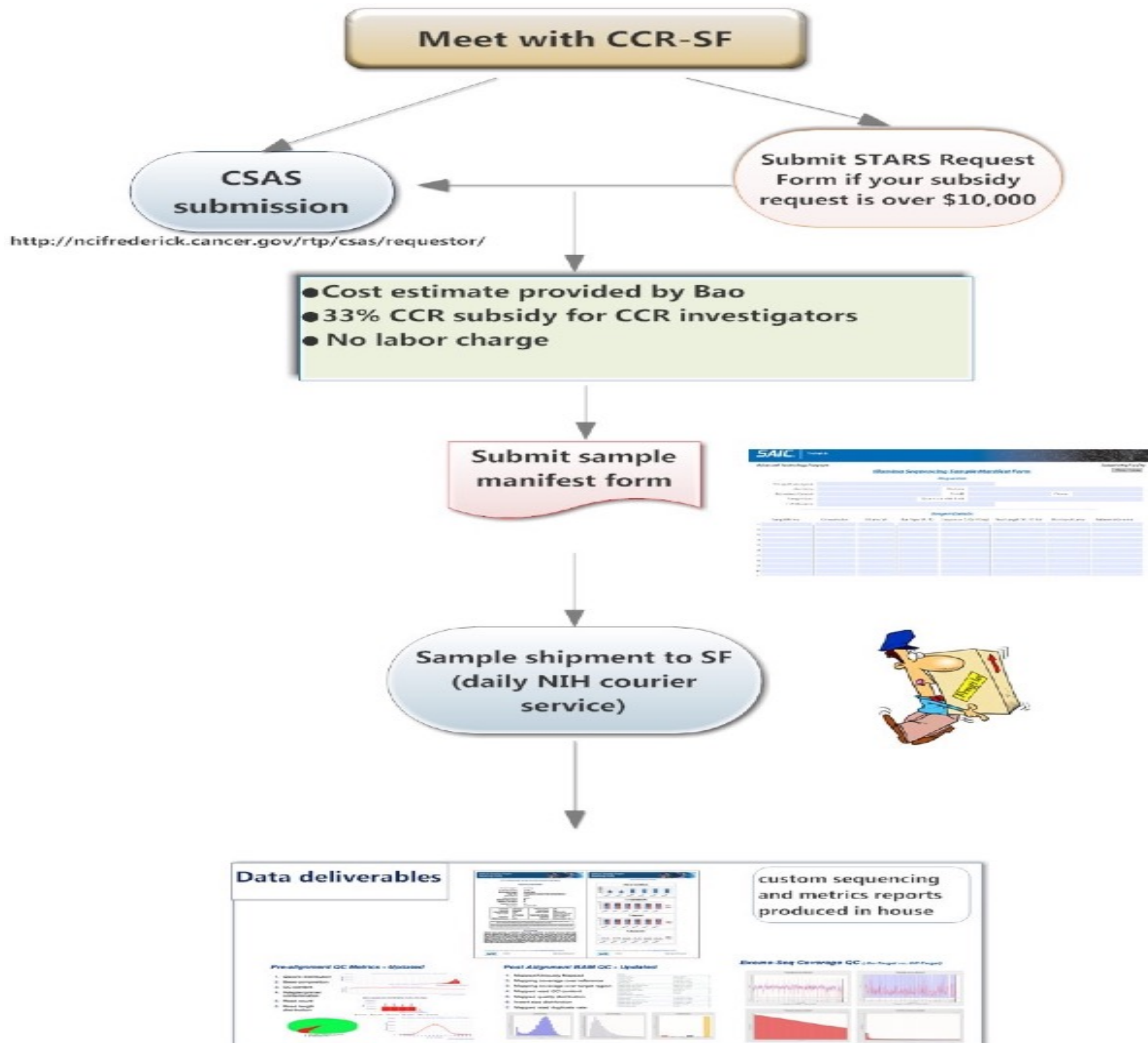
- Irys Next-Generation Mapping (NGM): Physical (optical) genome maps, provide long-range information to reveal true genome structure
  - Provide dense genome-wide anchor points for ordering and orienting sequencing contigs or scaffolds
  - Can be used alone to identify architecture and structural variation.



http://bionanogenomics.com/

Steps to Bring in Projects to CCR–SF

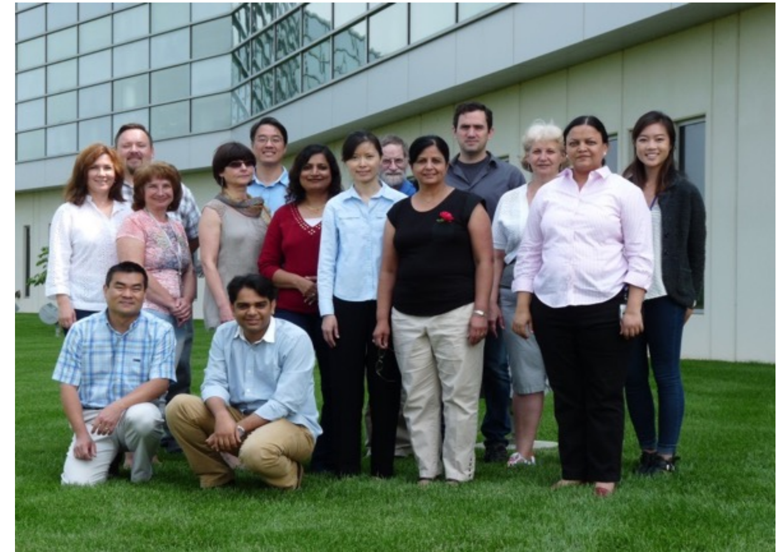
Frederick National Laboratory for Cancer Research



Talk to us….

Webpage:

https://ostr.cancer.gov/resources/fnl-cores/sequencing-facility



Phone:

301-360-3460

Email:

Lab director: tranb2@mail.nih.gov

Illumina lab manager: shettyju@mail.nih.gov

Pacbio scientist: raleyjc@mail.nih.gov

Bioinformatics manager: zhaoyong@mail.nih.gov

Frederick National Laboratory for Cancer Research

CCR Sequencing Facility Bioinformatics Support

Yongmei Zhao

Jan 23, 2017

Presentation Outline

Frederick National Laboratory for Cancer Research

- **Introduction**
  - – who we are and what service we provide
- **Production Data Analysis Support at CCR-SF**
- **R&D Project Data Analysis Support**
- **Data Delivery Method**
- **Summary**

Who We Are and What Service We Provide

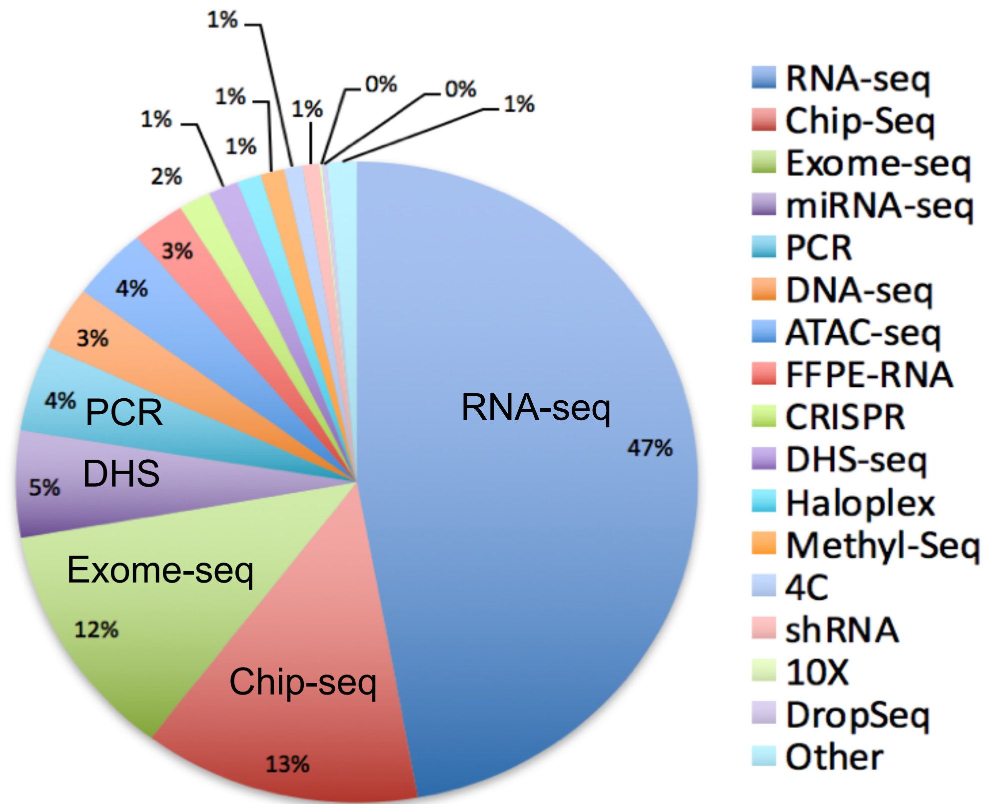
Frederick National Laboratory for Cancer Research

- **The CCR-SF bioinformatics group is a dedicated on-site informatics group that provides next generation sequencing and bioinformatics analysis support to CCR investigators**
- **The services we provide:**
  - Production sequencing data analysis support and quality assessment
  - R&D for new protocol and technology development
  - Data analysis workflow and pipeline development
  - LIMS and laboratory workflow tracking software support
  - NGS data management
  - Customer support for NGS data analysis related questions.

CCR-SF Sequencing and Data Analysis Support

FY2012 – FY2017

- **Completed more than 17,000 samples**
- **Delivered over 150,000 trillion bases**
- **Currently more than 20 library protocols are supported at SF**

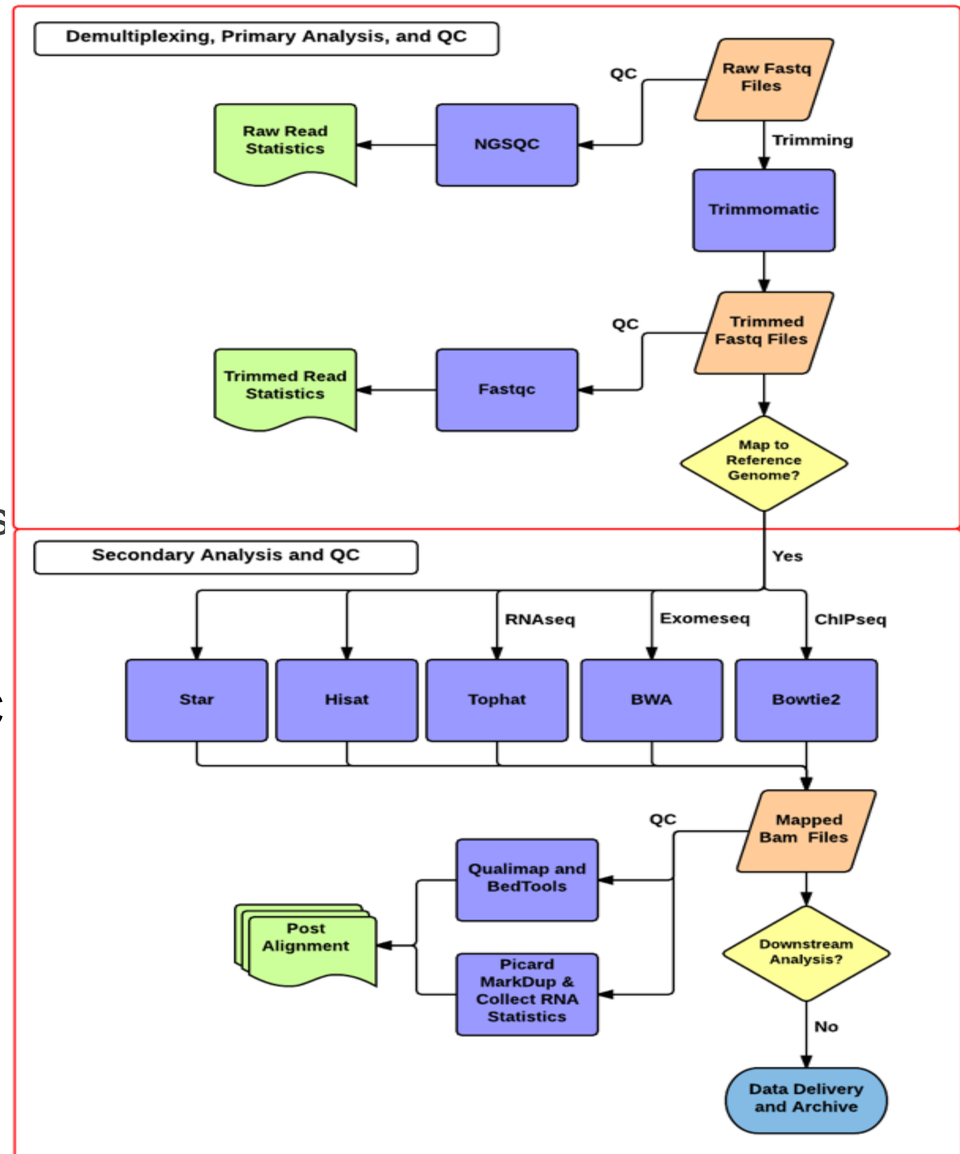


Production Data Analysis Workflow Overview

Steps for Data QC

- Run time processing and QC
- Demultiplex, pre-alignment QC
- Mapping and post-alignment QC
- Project specific secondary analysis and QC

- **Quality control and standard QC metrics development are the keys to ensure produce good quality data**



Primary Analysis and QC - help to determine systematic bias and library issue

Frederick National Laboratory for Cancer Research

➢ Pre-alignment QC metrics

- Sequencing Yield (sample, lane, undetermined barcodes)
- Per base quality score distribution
- Per sequence quality score
- Read length distribution
- Base composition
- GC content
- Duplications level
- Over-represented sequences
- Kmers
- Homopolymers

➢ Things to watch for:

- Pooling variation, missing barcodes?
- Lane batch effect?
- Sequencing quality issue?
- Unbalanced genome? High GC or AT?
- PCR bias or other artifacts?
- Adapter contaminant?

➢ Tools for QC, data cleaning:

- NGSQC toolkit, FASTQC, QC3, Trimmomatic, Cutadapt, Fastx-toolkit.

Post Alignment QC – measure library performance and biological variation

Post-alignment QC metrics

- Mapped/Uniquely Mapped
- Mapping coverage over reference
- Mapping coverage over target region(target capture)
- Mapped read GC content
- Mapped quality distribution
- Mismatch error rate
- Insert size distribution
- Duplicate rate
- RNA metrics for RNA-seq:

Things to watch for:

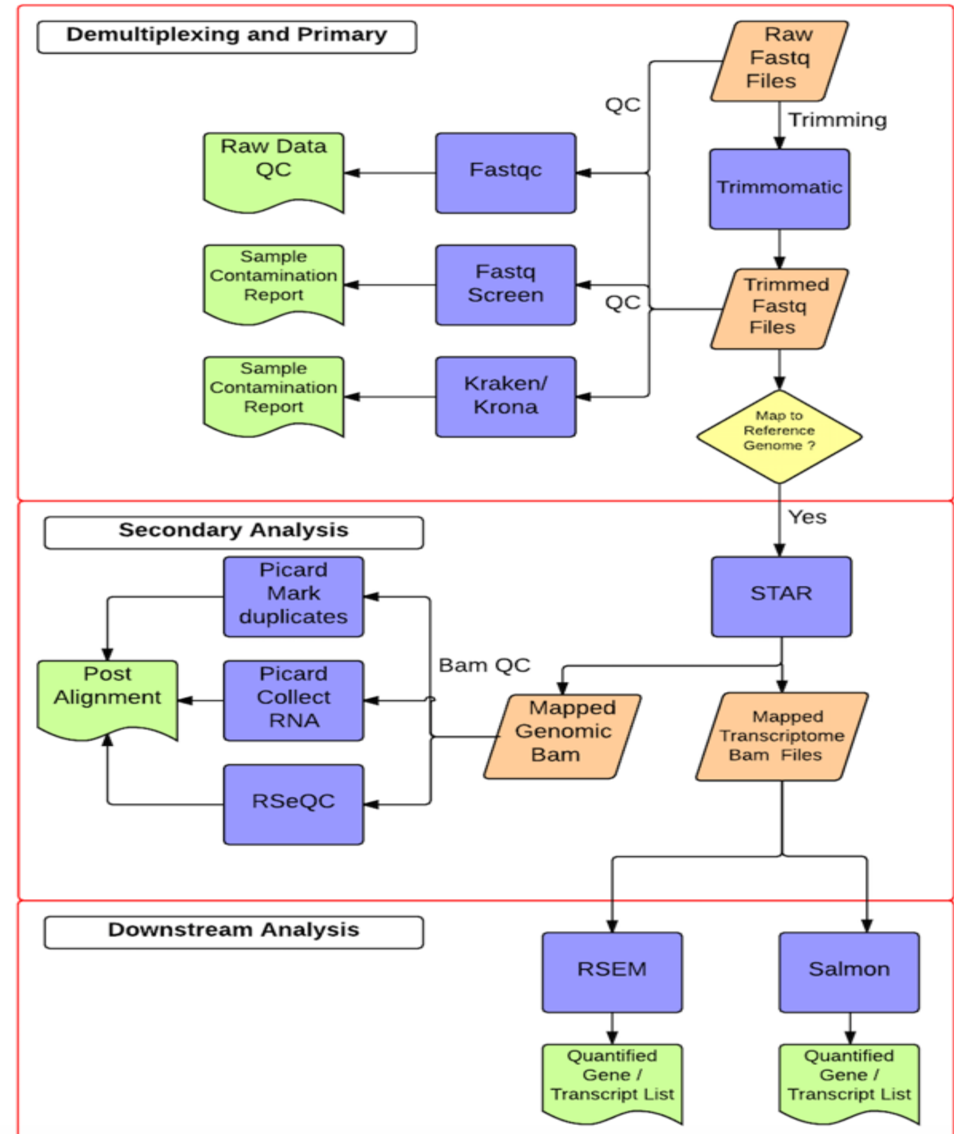
- Adapter and low quality trimming efficiency? Sample cross contamination?
- Target capture efficiency?
- Genome composition bias?
- Sequencing error? PCR bias?
- Unusual fragment sizes? PCR duplicates?
- Degraded RNA? strand bias? rRNA, pre-mRNA and other contaminant?

Tools for QC, batch effects and contamination screen:

- Qualimap, QC3, Bedtools, Picard, RSeQC, FastQ_Screen, BLAST

RNA-seq Data Analysis and Deliverable

- **FASTQ files**
- **Alignment BAM files**
- **QC statistics reports**
  - Mapped/Uniquely Mapped
  - Read count in intragenic, intergenic, exonic, intronic, mRNA, rRNA bases distribution,
  - 5'/3' bias, gene body coverage,
  - Strand specificity,
  - GC content
  - Pair-end inner distanct
  - Junction saturation
- **Gene/transcript quantification result files from Rsem and Salmon**



Production Data Report and Deliverables

Frederick National Laboratory for Cancer Research

- **Data deliverables for Rna-seq, exome-seq, chip-seq, WGS**
  - FASTQ/FASTA files
  - Alignment BAM files
  - QC Statistics Reports
- **Application specific:**
  - Gene quantification result for RNA-seq
  - miRNA count file for miRNA-seq
  - PCR product count file

Standard Report and Metrics Files

Frederick National Laboratory for Cancer Research

Sequencing Facility

CCR-Sequencing Facility Illumina Sequencing Report

Project Information

Principal Investigator:
PI Laboratory Contact:
Bioinformatics Contact:
Project Title:
CSAS Order ID: 17407
Samples Total in project: 25
Samples in This Report: 25
Completion of CSAS: yes
Report Date: April 29, 2015

Sequencing Details

Flowcell ID:	C6JH5ANXX	Sequence Control:	PhiX
Instrument:	HiSeq2500	Control Result:	Pass
		Library Protocol:	Clonetech Nextera RNA
Sequencing Type:	mRNA-seq		
Read Length:	126 (2x126cycles)	Sequencing Chemistry:	Illumina TruSeq V4.0
		Reference Genome:	Mm9
Multiplexed:	Yes (12 per lane)	Target Region File:	Ensembl_ncbim37 GTF

Run Comments

Twelve mRNA samples were sequenced on one HiSeq2500 lane using Illumina TruSeq v4 chemistry. Overall sample yields ranged from 23 to 44 million pass filtered reads. All the samples were of good quality, with percent Q30s between 87% and 90%. Sample reads were trimmed of adapters and low quality bases using Trimmomatic software and then aligned to the mouse (mm9) genome and ensemble ncbim37 transcripts using Tophat software. The sample total percent genome alignments ranged between 88% and 89%, with unique percent alignments all above 81%. RNA mapping statistics were calculated using Picard software. The percentage of mRNA sample bases was as expected, between 79% and 84%, with coding bases consisting of between 46% and 53% of these mRNA bases.

*Note: Residual samples will be retained up to **90 days** of the delivery of this report. To avoid shipping charges, please contact SFILLUMINALAB@mail.nih.gov to arrange pickup samples prior to this time.*

*Note: Sequencing data will be available to download for **two weeks** following delivery of this report. Please download the data files as soon as possible.*

For questions on any aspect of this report please contact SFILLUMINABIOINF@mail.nih.gov.

leidos
Leidos Biomedical Research, Inc.
http://atp.ncifcrf.gov/sf

Frederick National Laboratory for Cancer Research

Sequencing Facility

Analysis Workflow

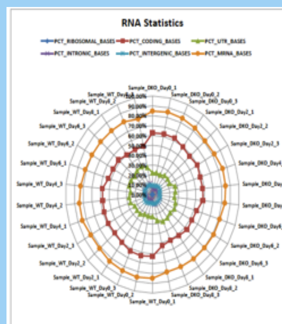
Base Calling and Pre Alignment Quality Check — RTA, bcl2fastq, NGS QC Toolkit
Trimming and Filtering — Trimmomatic
Alignment and Data Analysis — TOPHAT, AddOrReplaceReadGroups - Picard
Post Alignment Quality Check — CollectRnaSeqMetrics - Picard, RseQC, MarkDuplicates - Picard

Software and Parameters

Analysis Step	Software	Software Parameters / Notes
Basecalling	RTA 1.18.64	Illumina instrument run time analysis software
Demultiplexing	Bcl2fastq 1.8.4	Barcode demultiplexed allowing 1 mismatch
Filtering (Adaptor and quality)	Trimmomatic 0.30	PE -threads 16 -phred33 ILLUMINACLIP:TruSeq_and_nextera_adapters.fa:3:50:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:20 MAXINFO:50:0.8 MINLEN:25
Alignment	TopHat v2.0.8	tophat -G annonation.gtf -o ./ -r 10 --mate-std-dev 200 -p 16 --library-type fr-firststrand bowtie2_annoatation/prefix read1.fastq read2.fastq
RNAStatistics	Picard 1.84	CollectRnaSeqMetrics.jar REF_FLAT=annotation_refFlat.txt INPUT=sample.bam OUTPUT= RnaSeqMetrics.txt RIBOSOMAL_INTERVALS= ribosome_interval_list.txt STRAND_SPECIFICITY=NONE VALIDATION_STRINGENCY=LENIENT
Duplication Statistics	Picard 1.84	MarkDuplicates.jar INPUT=sample.bam OUTPUT=sample.MKDUP.bam METRICS_FILE=sample.bam.metric ASSUME_SORTED=true MAX_FILE_HANDLES_FOR_READ_ENDS_MAP=1000 VALIDATION_STRINGENCY=LENIENT
Insert Size Statistics	RseQC 2.3.5	inner_distance.py -i sample.bam -o ./ -r annotation.bed

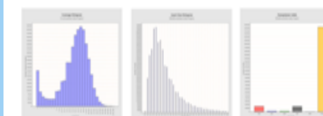
For questions on any aspect of this report please contact SFILLUMINABIOINF@mail.nih.gov.

leidos
Leidos Biomedical Research, Inc.
http://atp.ncifcrf.gov/sf

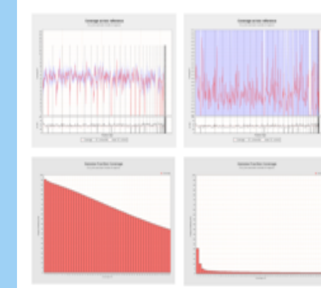


Post Alignment BAM QC - Updated

1. Mapped/Uniquely Mapped
2. Mapping coverage over reference
3. Mapping coverage over target region
4. Mapped read GC content
5. Mapped quality distribution
6. Insert size distribution
7. Mapped read duplicate rate



Exome-Seq Coverage QC (On-Target vs. Off-Target)



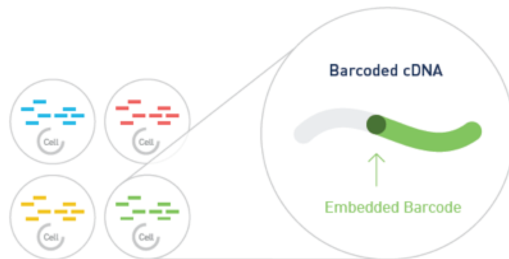
New Protocol Development and Data Analysis Support at SF

Frederick National Laboratory for Cancer Research

- Single Cell 3’ mRNA-seq analysis
- 10x Genomics for Large structural variation detection
- PacBio full length transcriptome sequencing analysis
- WGBS and targeted methylome sequencing analysis
- Genome optical mapping using BioNano Irys

Single Cell Digital Gene Expression Using 10x genomics chromium

Single Cell Digital Gene Expression



mRNA is transcribed by a reverse transcriptase that creates barcoded cDNA.

A barcode identifies transcripts originating from a single cell, which are then counted.

Sc V1		Sc V2	
Number of Cells	Expected Doublet Rate(%)	Number of Cells	Expected Doublet Rate(%)
NA	NA	500	~0.4
1200	~1.2	1000	~0.8
3000	~2.9	5000	~3.9
6000	~5.7	10,000	~7.6

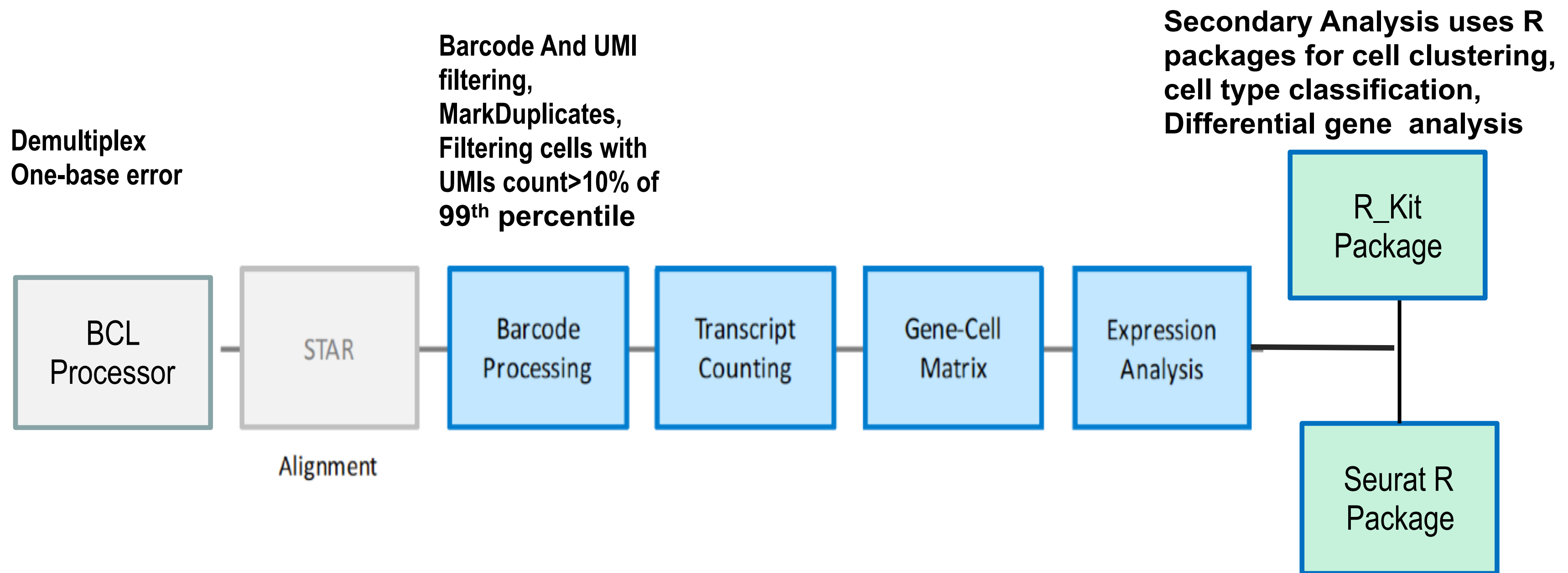
- Cell characterization and gene profiling
- Helps uncover the variation masked by the average of heterogeneous subpopulations.
- Profiling 1000-10,000s cells per experiment
- No cell size restrictions
- Up to 50%(Scv1) or 65%(Scv2) cell capture efficiency
- Multiplex rate is very low
- 50,000 -60,000 raw reads per cell is recommended min seq depth for for typical samples
- 30,000 raw reads per cell for RNA-poor cell types such as PBMCs
- Extra sequencing need given variability in cell counting, loading

10x Genomics: http://www.10xgenomics.com

Single Cell 3' mRNA-seq Analysis Pipelines

Frederick National Laboratory for Cancer Research

10X Genomics Single Cell Analysis Pipeline and Visualization Tools



- Filtering:
- Only use reads with valid cell barcodes (no more than 1 mismatch allowed) and and high Quality UMIs
- Only use confidently mapped reads aligning to transcriptome
- Select barcodes with total UMI count >10% of the 99th percentile of the expected recovered cells to product gene matrix

http://satijalab.org/seurat/pbmc-tutorial.html

10x Genomics: http://www.10xgenomics.com

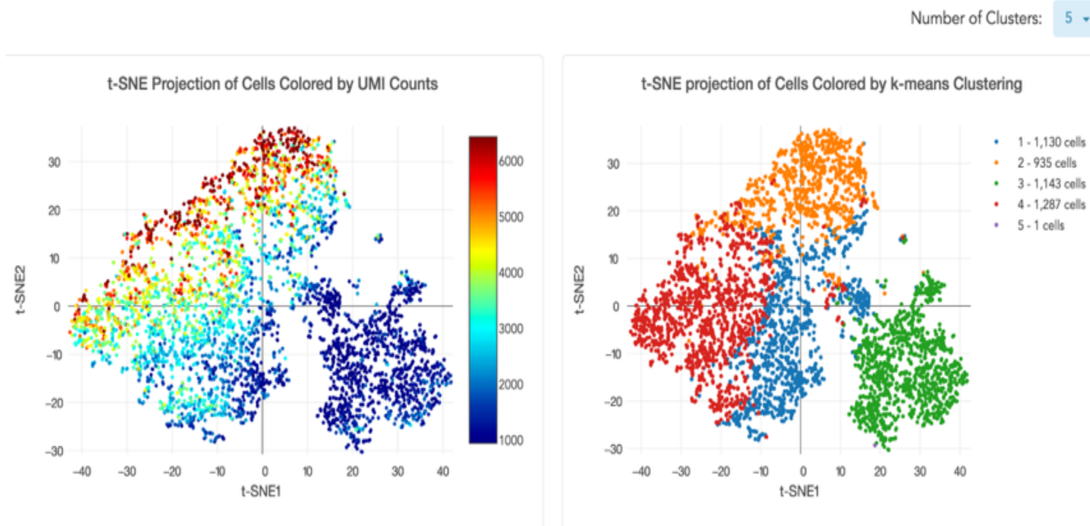
❖Sean Davis Web Site to see additional scRNA tools https://github.com/seandavi/awesome-single-cell/

10x scRNA-seq Data Delivery Files

Frederick National Laboratory for Cancer Research

- Raw fastq, genome and splice mapping BAM files and QC statistics
- Gene matrics files including gene, cell barcodes and UMI count
- PCA , t-SNE projection, and K-means clustering files for single cells data
- Differentially analysis result file and gene list for the single cell data

Examples of clusters generated by 10X Chromium: CD4+, Activated (Dr. Remy Bosselut's group)



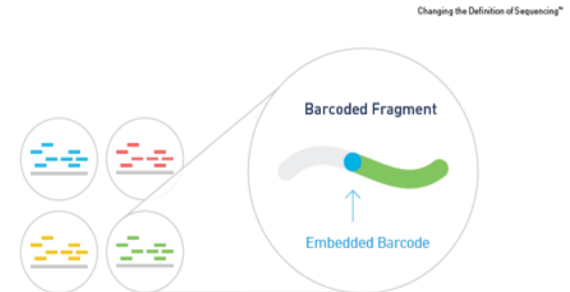
Top differentially expressed genes per cell cluster (UMI counts/cell)

Gene name	Cluster 1	Cluster 2	Cluster 3
Abc1	28.52	0.03	14.7
Xyz2	4.56	8.33	30.85
Fgh3	8.94	17.44	1.27

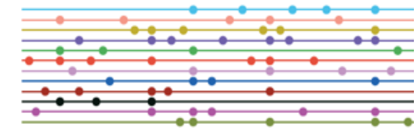
Large structural variation detection and phasing analysis using 10x genomics

Frederick National Laboratory for Cancer Research

- **Advantage of Linked-Reads:**
  - Identify intra/inter-chromosomal breakpoints and fusion genes
  - Distinguish haplotypes
  - Resolve ambiguous reference alignments
  - >10 Mb haplotype blocks with phased SNVs, indels, and structural variants.
- **Sequencing depth requirement:**
  - 30x -40x for WGS, 60x -70x for Exome-seq, 40x-56x for de nova assembly
- **Variant Detection Sensitivity**



Genome is replicated and barcoded via a low-level enzymatic replication.


Lines represent Linked-Reads. Dots represent reads. Color indicates barcode. Reads with same barcode originated from molecules encapsulated in the same partition.

Variant Class	Scale	Current Support
SNVs	1bp	Called and phased by Long Ranger
Indels	<50bp	Called and phased by Long Ranger
Small SVs (del, inv, ins)	50bp – 30Kb	Enabled by Linked-Reads, algorithms under development
Large-Scale SVs (del, inv, transloc)	>30Kb	Called and phased by Long Ranger

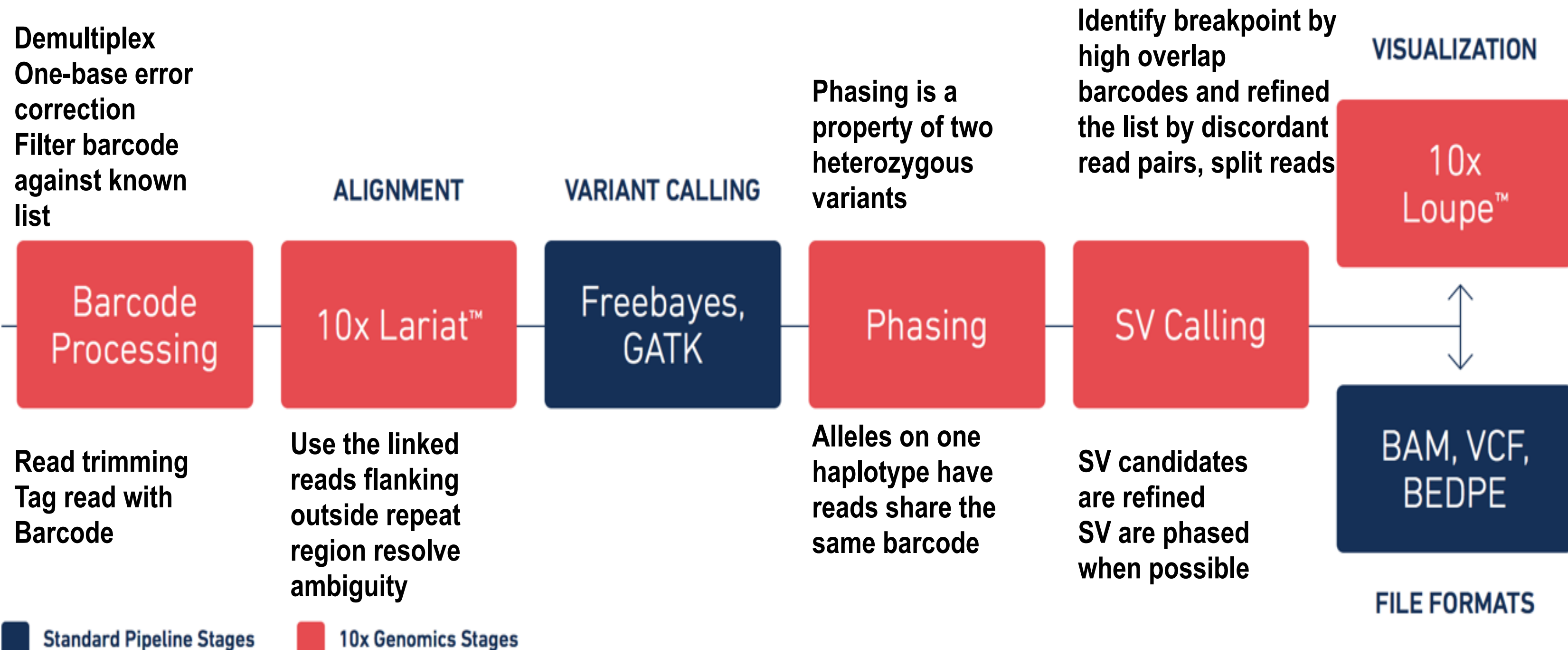
*Algorithms and workflow optimized for germline. Germline algorithms serving as initial prototypes for somatic calling.

10x Genomics: http://www.10xgenomics.com

Structural Detection Analysis Pipelines

Frederick National Laboratory for Cancer Research

10X Genomics Data Analysis Pipeline and Visualization Tools



10x Genomics: http://www.10xgenomics.com

10x Genomics Large Structural Variation Detection Data Deliverable

Frederick National Laboratory for Cancer Research

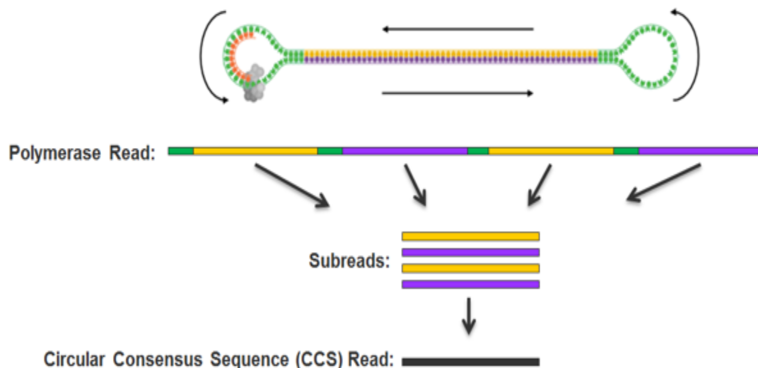
- **Raw FASTQ files and QC metrics**
  - Barcode exact match whitelisted 10x barcode, and post error correction match white list ratio
  - Barcode Q30 base ratio
  - Gem/droplet count estimate
  - Mean barcode Qscore
  - Total reads
- **Barcoded and phased BAM file**: position-sorted, has index code and phasing tag
- **Phased VCF file:** encode SNP and indel phasing in the GT (genotype) and PS (phase set) fields of VCF
- **Phased Structural Variants in VCF Format**
- **Phased Large-Scale Structural Variants in BEDPE Format**

PacBio Iso-seq Full Length Transcriptome Sequencing And Data Analysis

Frederick National Laboratory for Cancer Research

Advantages of using Iso-seq

- Extraordinarily long reads: full length transcript up to10kb. One read, one transcript, no assembly required
- Complete information about alternatively spliced exons, transcriptional start sites, polyadenylation sites and strand orientation
- High accuracy: sequencing individual molecule with >99% accuracy
- Least GC bias and no amplification bias
- Better and even gene coverage on both 5’ and 3’ end
- Cost saving with targeted Iso-seq for specific gene



Pacbio IsoSeq Pipeline

Polymerase Reads

Reads-of-Insert Protocol

Reads of Insert - Fasta

IsoSeq Classify

Incomplete Transcripts Fasta

Full-Length Transcripts Fasta

Secondary Analysis and Visualization - SF Workflow

Genomic Reference

GENCODE Annotation

Star Mapper

Mapped Exon Bam File

MatchAnnot

MatchAnnot Report

MatchAnnotation

Yes

Known Transcript and Count

No

GENCODE Annotation

SF Iso-Cluster program

Novel Transcripts BED files and BAM files

IGV Viewer

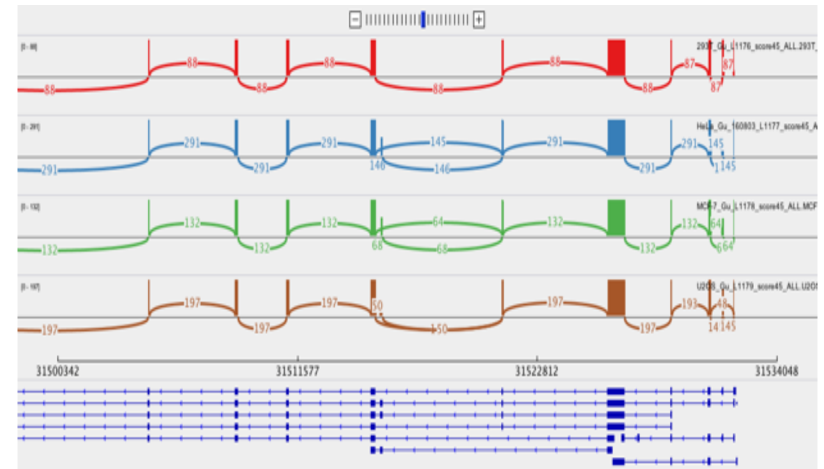
Iso-seq Data Delivery Files

Frederick National Laboratory for Cancer Research

- **Fasta file** of full length non-chimeric sequence files and non-full length sequence files
- **BAM file** of aligned sequences (STAR against genomic reference for full length non-chimeric reads)
- **BED files** for clustered potential novel and known iso-seq isoforms (targeted iso-seq only)
- **Split BAM and FASTA** files extracted for each isoform (targeted iso-seq only)
- **MatchAnnot annotation file** (optional)

Alternative Splicing

5' end



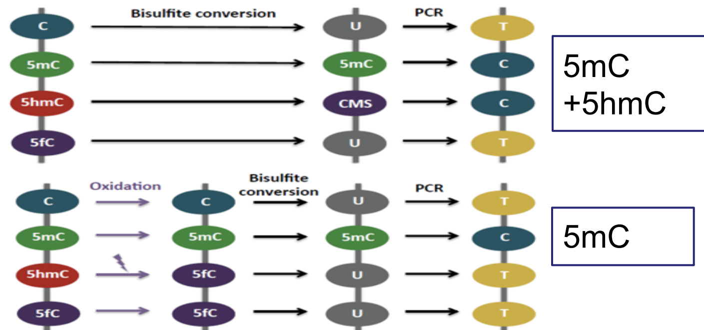
Full length View

Methylated DNA Sequencing Analysis

Frederick National Laboratory for Cancer Research

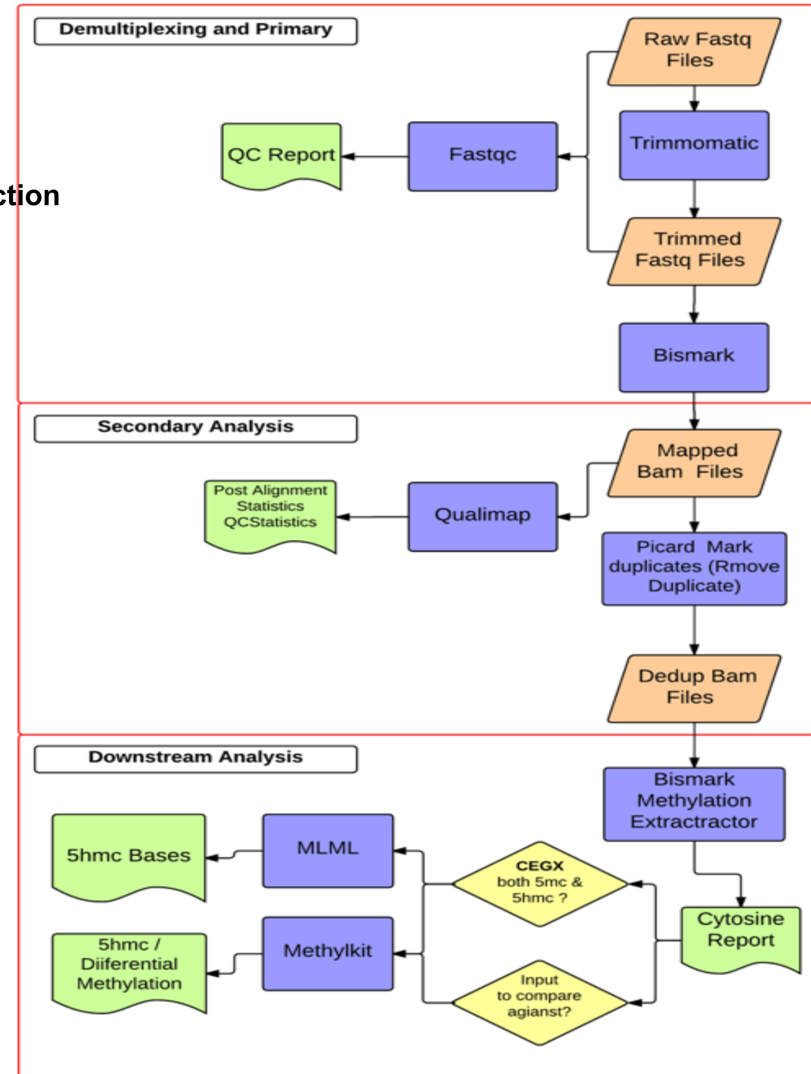
- DNA Methylation: regulation of cellular differentiation and development
- Bisulfite mediated conversion of cytosine to uracil enables detection of 5mC

CEGX TrueMethyl oxBS and BS enable true 5mC and 5hmC detection



Data delivery files:

- FASTQ and Bismark alignment BAM files
- Bsmark output raw file and summary
- Bsmark methylation extractor output files
- Differential methylation analysis to find true 5mc vs. 5hmc from CEGX libraries



Typical Data Delivery Sizes

Frederick National Laboratory for Cancer Research

Instrument and Run Type	Estimated Customer Data Per Lane
HiSeq3000 : 2x75 PE	65 - 140GB
HiSeq3000 : 2x150 PE	130 - 280 GB
Hiseq2500 -V4: 2x126 PE High throughput	125 - 200GB
NextSeq: 2X150 PE High throughput	200 - 260GB
NextSeq: 2x75 PE High throughput	100 - 130GB
NextSeq: 2x75 PE Mid throughput	80 - 120GB
NextSeq: 2 x75 PE Mid throughput	40 - 60GB
Miseq 1x36 SR	<1GB
Miseq 2x100 PE	4.5 - 9GB
Miseq 2x150 PE	6.5 - 14GB

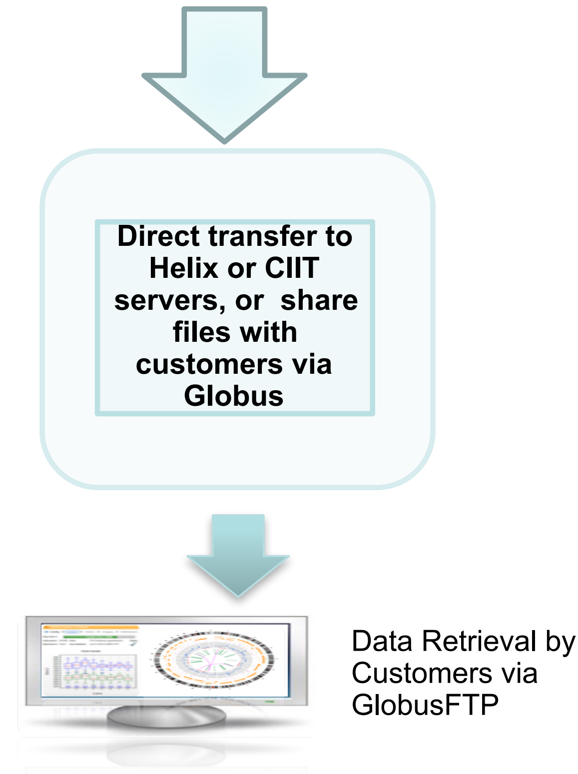
Standardized Data Delivery Method

Frederick National Laboratory for Cancer Research

- Steps for Using GlobusFTP

1. Register GlobusFTP account via:
   – *https://www.globus.org*
2. Send request to helix support to obtain approval to use helix GlobusFPT server
3. Follow helix tutorial to configure Globus endpoint:
   – *https://hpc.nih.gov/storage/globus.html*
4. Send Globus account name to CCR-SF IFX
5. Follow SF IFX provided link to download data files

Secure Data Delivery Via GlobusFTP or Other Transfer Protocol



Data Retrieval by Customers via GlobusFTP

Summary

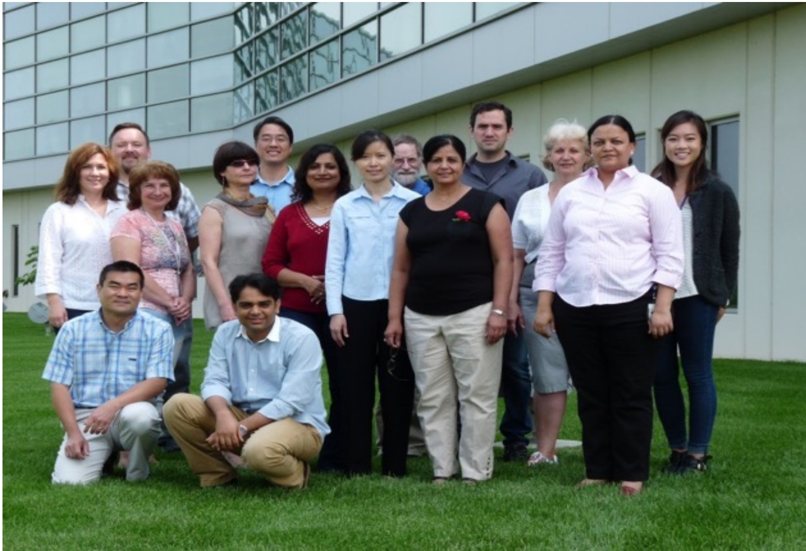
Frederick National Laboratory for Cancer Research

- **High quality of data comes from good experiment planning and good biological material. Ensure the sample input and quality to meet minimum protocol requirement.**
- **Consult your bioinformatics support earlier in the experiment planning stage to ensure good experimental design; discuss library protocol and sequencing strategy selection with SF/sequencing core to choose protocol that fits your project need**
- **The quality of the data have great impact on the success of the downstream analysis and biological discovery**
- **Choice between production protocols and new technologies dependent on project specific need, resources and timeline**

Acknowledgements

Frederick National Laboratory for Cancer Research

Sequencing Facility Teams



Leidos Biomedical

Dwight Nissley
Jack Collins
Bob Stephens
Xiaolin Wu
Daniel Soppet
Paul Griffith
Eric Stahlberg
Gregory Warth

Illumina

Dan Harmer
David Armistead

Pacific Biosciences

Yan Guo
Elizabeth Tseng

National Cancer Institute

David Goldstein
Mariam Malik
Paul Meltzer
Sean Davis
Javed Khan
Gordon Hager
Remy Bosselut
Isaac Brownell
Keisuke Nagao
Thomas Fountaine
Shuo Gu

10x Genomics

Ashley Wilson
Alex Wong

OSTR BTEP

Perter Fitzgerald
Anand Merchant

CCBR Bioinformatics

Maggie Cam
Parthav Jailwala
Jack Chen

Thank You!