



**Weill Cornell
Medicine**

Core Laboratories Center

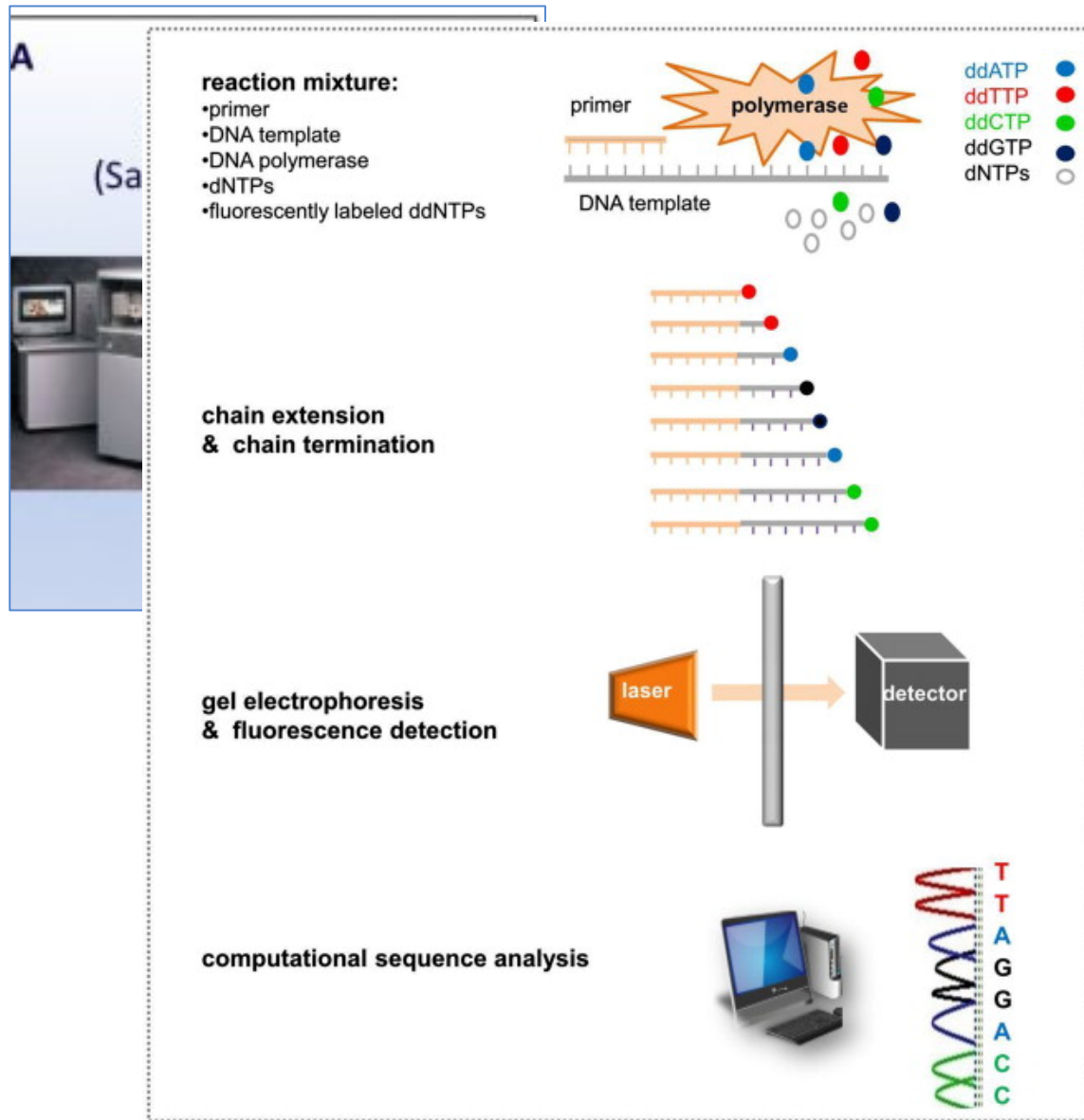
Epigenomics Core Facility:
“I have samples for sequencing”, what does it mean?



Alicia Alonso, PhD.
Director, Epigenomics Services
Genomics and Epigenomics Core Facility
Assistant Research Professor, Department of
Hematology and Oncology

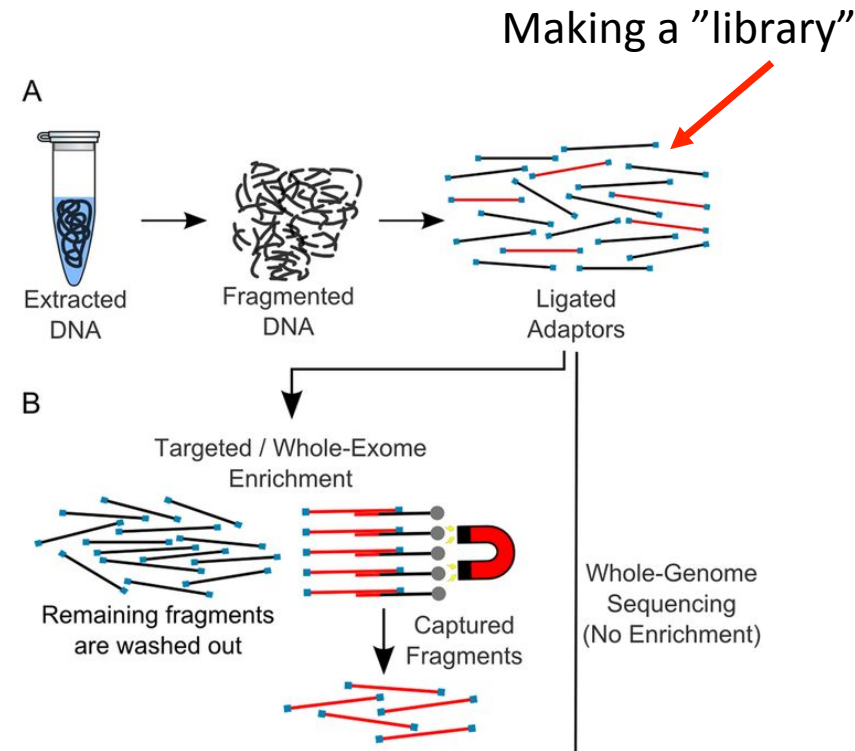
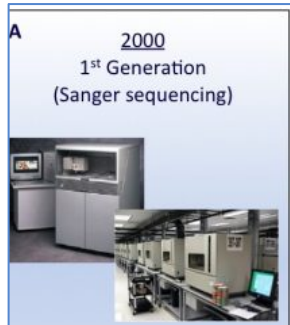
04022019-ICB students

Evolution of Sequencing Technologies



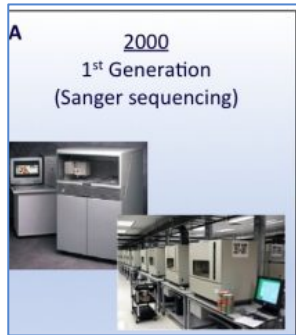
- Custom sequencing (targeted)
- Sequencing length: 300-1000bp
- Output: 384 independent samples at a time
- ABI Prism 375 (Applied Biosystems)

Evolution of Sequencing Technologies



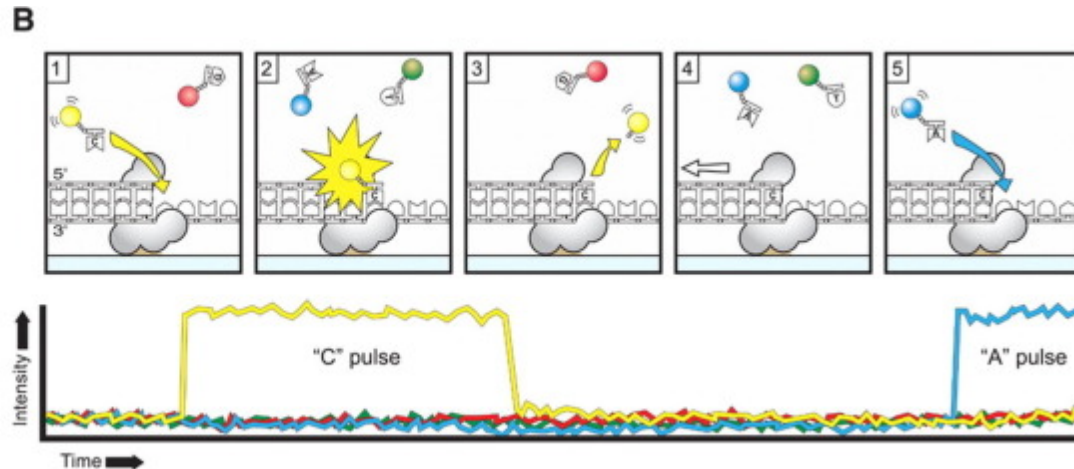
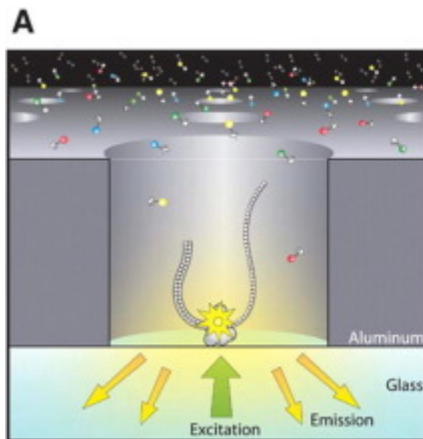
- Generation of many millions of short reads **to be sequenced** in parallel (150-600bp)
- Speed of sequencing (compared to 1st generation)
- Cost of sequencing (lower per base)
- Output is detected directly (ie no electrophoresis)
- Output varies from 15M – 400B “reads”
- Short reads are aligned to a **previously sequenced genome**
- Examples of 2nd generation competing technologies
 - Roche/454 (pyrosequencing – extinct- up 1000bp, 1Gb)
 - Ion torrent (detects the Hydrogen ion released, change of pH-Life Technologies, 2010) 200-600bp, 10Gb (2-8 hours)
 - ABI/SOLiD 35-75 bp 30Gb/run (each base is read 2x, high accuracy)
 - ***Illumina/Solexa sequencing: Sequencing by Synthesis SBS (50-250bp)***

Evolution of Sequencing Technologies

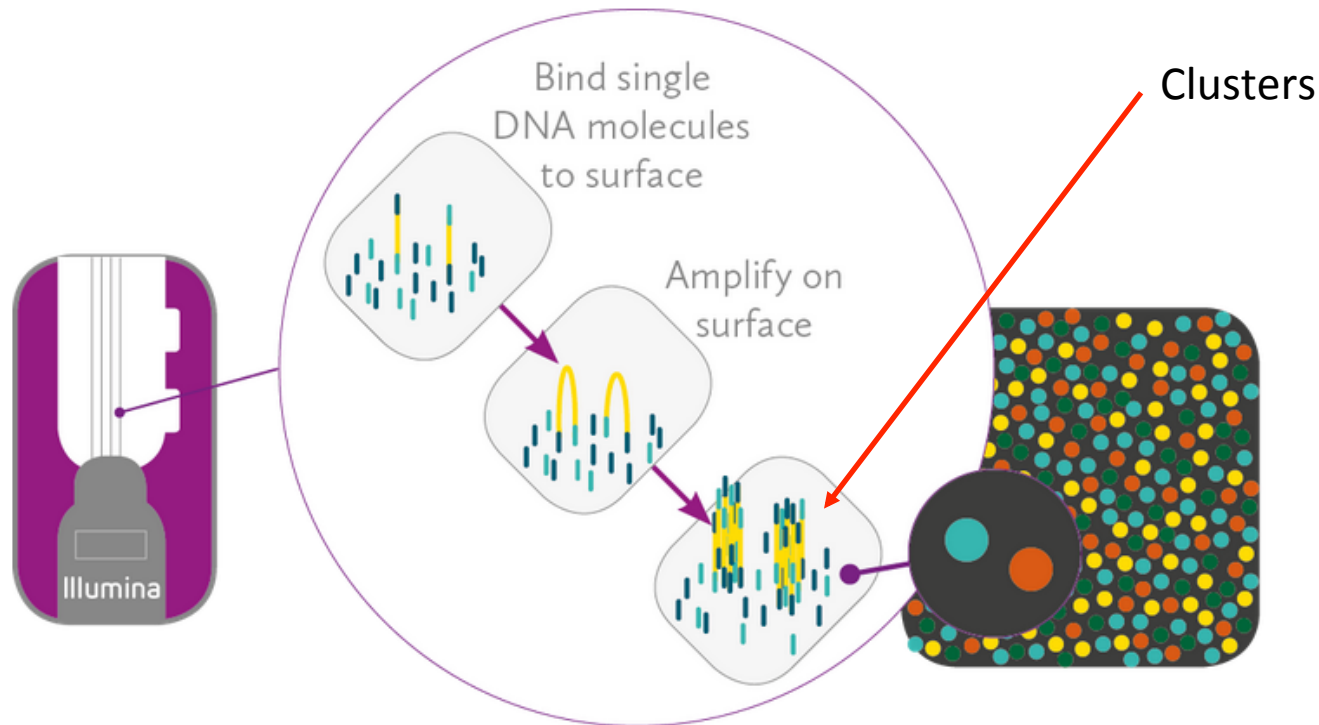


SMRT : single molecule real time sequencing

- Pacific Biosciences: Sequel
- Oxford Nanopore: MinION, PromethION
- Sequencing length: up to 100kb
- Allows "de novo" mapping
- Sequencing of modifications of the bases (methylation for example)



Illumina Technology for 2nd Generation Sequencing Rules the Field!



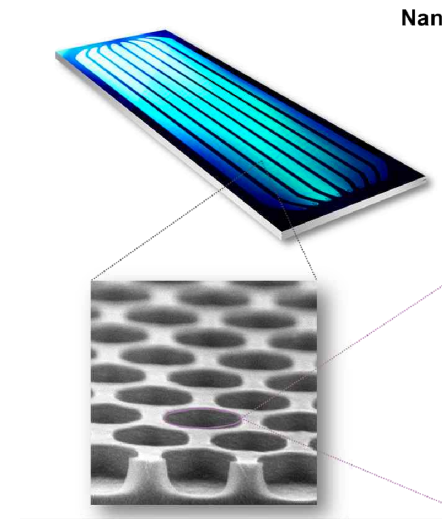
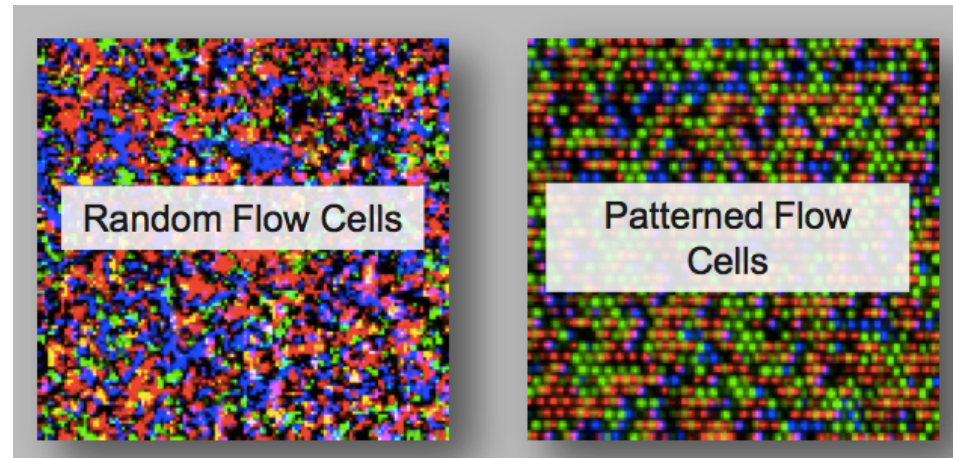
Processing: base calls are made directly from the signal intensity using Illumina's RTA software, using Casava 2.1.7 raw reads and quality scores are generated.

Clusters: varies with the sequencer used and the chemistry used
A HiSeq 2500 (our most common sequencing run) gives ~ 800k/mm²

reads: total number of clusters X read length
for SR50, our most common length ~280M reads, 7GB per lane, 8 lanes

Illumina Technology continues improving!

Efficient clustering



Illumina Technology continues improving!

Faster “sequencing”

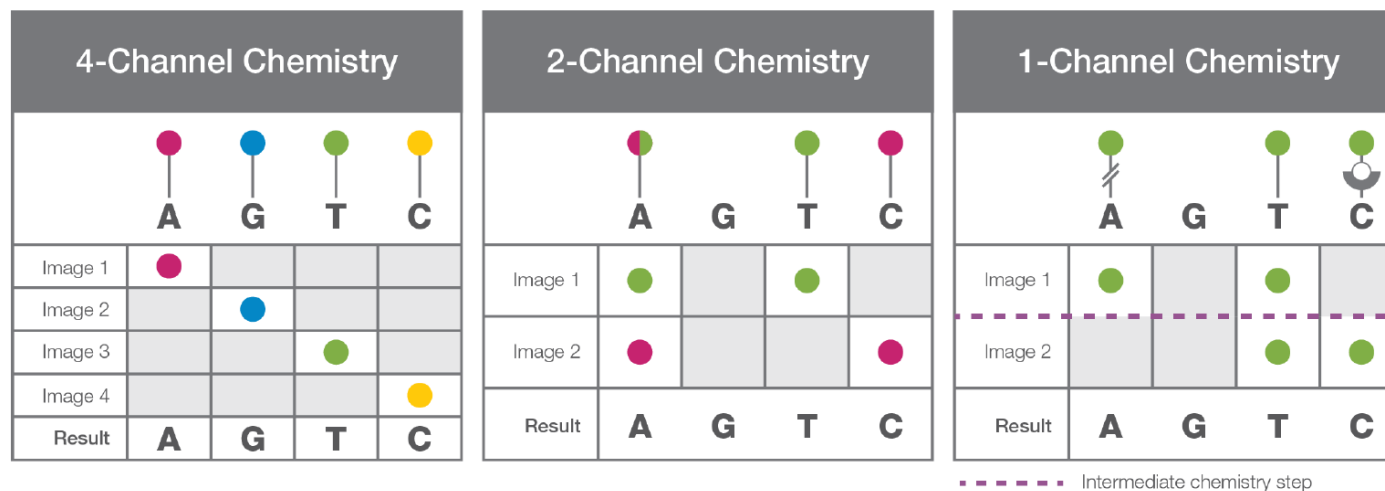
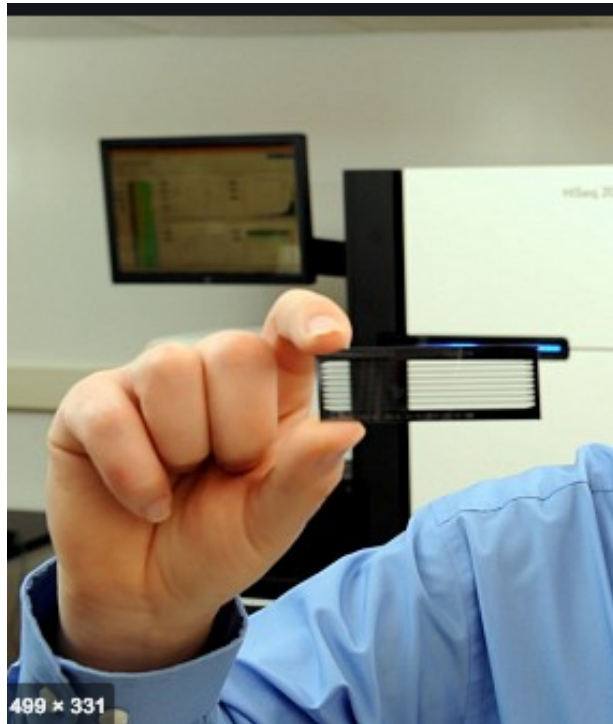


Figure 2: Four-, Two-, and One-Channel Chemistry—Four-channel chemistry uses a mixture of nucleotides labeled with four different fluorescent dyes. Two-channel chemistry uses two different fluorescent dyes, and one-channel chemistry uses only one dye. The images are processed by image analysis software to determine nucleotide identity.

Increasing throughput

Sequencing System	iSeq [™]	MiniSeq [™]	MiSeq [™]	NextSeq [™]	HiSeq [™]	HiSeq [™] X	NovaSeq [™]
					4000	Five/Ten	6000
Output per run	1.2 Gb	7.5 Gb	15 Gb	120 Gb	1.5 Tb	1.8 Tb	1 Tb - 6 Tb ¹
Instrument price	\$19.9K	\$49.5K	\$99K	\$275K	\$900K	\$6M ² /\$10M ²	\$985K
Clustering	patterned	random	random	random	random & patterned	random	patterned
Chemistry	1-color	4-color	4-color	2-color	4-color	4-color	2-color

Illumina Technology continues improving!



Illumina Technology continues improving, the need to MULTIPLEX!

- Coverage (or depth):
number of reads that are likely to be aligned at a given reference position

$$\text{Coverage} = \frac{\text{read length} \times \text{number of reads}}{\text{genome length (haploid)}}$$

- How many reads would one need to sequence ONE human genome at 30x coverage in a PE100 run?

$$\text{number of reads} = \frac{30 \times 3 \times 10^9}{200} \quad 450\text{M reads}$$

On a HiSeq 2500: two sequencing lanes ~\$5,000

On a HiSeq 4000: one sequencing lane ~\$2,500

On a Novaseq S4: 1/6 of a sequencing lane ~\$1,000

- Multiplexing: being able to physically pool libraries from different origins, sequence and then bioinformatically re-assign each library to its origin

For ChIPseq, 40M reads



Overview

The Epigenomics Core Facility of Weill Cornell Medicine provides an array of epigenomics and bioinformatics research resources and services that include:

- DNA methylation profiling [\[More\]](#)
- Protein-nucleic acid association [\[ChIP-Seq, ATAC-Seq\]](#)
- Single Cell RNA and Long Range DNA Sequencing [\[More\]](#)
- RNA-seq [\[More\]](#)
- Bioinformatics analysis [\[More\]](#)

Core resources and services include sample preparation services and data generation on the Illumina HiSeq 2500, MiSeq platforms and Sequenom MassArray platform.

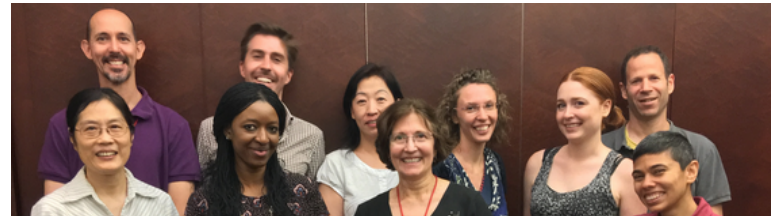
Getting Started

Request Services

Pubshare

FAQ

Pricelist



Wet Lab

Scientific Director
Lab Manager
Research Specialists

Research Technician

Alicia Alonso, PhD
Yushan Li
Marisa Mariani, PhD
Caroline Sheridan
Yuan Xin

Dry Lab

Bioinformatics Director:
Computational Biologist:
Software Developers:

Doron Betel, PhD
Piali Mukherjee
Thadeous Kacmarczyk, PhD
Simon Johnson

A project with the Epigenomics core includes

Consultation

Sample
submission

Library prep

Sequencing

Data release

Bioinformatics
support

Sample Submission Tracking

Epigenomics Core

Bioinformatics Support 99. QC RNA 98. QC DNA 14. RNAseq_SingleCell 13. RNAseq_Total_LowInput 12. RNAseq_L

Show All entries

Searching within 02. RRBS: Results in this tab are restricted by Status, Keywords (show details)
Please use the filter panels in the left-hand menu to drill down to requests of interest. You can also save filters to custom tabs! Click

Show Filters

date	for	service id
May 11 (May 07 2018)	Reinhard Stoger Stoger, Reinhard (UON) Lab	EC-RS-5126 L-MiniLIMS RRBSx8; SR50x3
Apr 25 (Apr 25 2018)	Ronan Chaligne Landau, Dan (WCMC) Lab	EC-RC-5080 L-MiniLIMS Plate 4 scRRBS_RNAseq, I
Apr 25 (Apr 25 2018)	Shannon Odell Toth, Miklos (WCM) Lab	EC-SO-5077 L-MiniLIMS RRBSx9; SR50x2
Apr 16 (Apr 16 2018)	Ronan Chaligne Landau, Dan (WCMC) Lab	EC-RC-5054 L-MiniLIMS Plate 3 scRRBS/RNAseqx1
Apr 16 (Apr 16 2018)	Ronan Chaligne Landau, Dan (WCMC) Lab	EC-RC-5051 L-MiniLIMS Plate 2 scRRBS/RNAseqx1
Apr 09 (Apr 09 2018)	Ronan Chaligne Landau, Dan (WCMC) Lab	EC-RC-5032 L-MiniLIMS Plate 1 scRRBS_RNAseqx1
Mar 23 (Mar 23 2018)	Shannon Odell Toth, Miklos (WCM) Lab	EC-SO-5001 L-MiniLIMS RRBSx6; SR50x1
Jan 11 (Jan 11 2018)	Alicia Alonso WCM Epigenomics Core - lab	EC-AA-4891 L-MiniLIMS RRBSx2 (RRoxBS) ; TEST

Sequencing_Request Summary

Sequencing Request

Sequencing Request 2659

Property	Value
Date Modified	2018-04-16
Date Request Submitted	2018-04-13
ILabs Service ID	EC-JMGS-5049 ILab
Submitter	Jesús María Gómez Saline
PI	Shido Koji
Request ID	mouse adult liver ECs
Organism	Mus musculus
Library Preparation Requested	Yes
Library Type	RNA-seq-Chromium-single
Run Type	Paired End
Read Length	26-8-98
Number of Samples	1
Multiplexing Requested	Yes
Number of Lanes	1
Alignment Requested	Yes
Reference Genome	NA
Sequencing Platform	Illumina HiSeq
CASAVA version	cellranger2.0
Illumina Flowcell	Illumina Flowcell 1244 SM
Status	completed
Assigned to	Yushan

LibraryTableAttributes

show / hide columns

Show All entries

Sample Number	Lib
1	mouse_adult_liver_ECs

RB_Plus_P53-1	1	human	harold_varmus	huc2003@med.cornell.edu	EC-HC-50!
RB_Plus_P53-2	1	human	harold_varmus	huc2003@med.cornell.edu	EC-HC-50!
CCND2WT1_E13_5_04_10_2018	pool1	rna-seq-chromium-single-cell-v2	margaret_ross	shs2039@med.cornell.edu	EC-SS-50!
CCND2WT1_E13_5_04_10_2018	pool1	rna-seq-chromium-single-cell-v2	margaret_ross	shs2039@med.cornell.edu	EC-SS-50!

Internal LIMS

Illumina Flowcell

Illumina Flowcell 1244

Property	Value
Date Created	2018-04-26
Date Modified	2018-04-26
Run ID	180427_D00796_0325_ACCZKCANKX SM
Flowcell ID	CCZKCANKX
Run Type	Paired End
Read Length (bp)	26-8-98
Cluster Generator	cBot1028
SBS kit	HiSeq SBS Kit v4
Sequencer	D00796
User	yu2008
Status	Released Withhold

Write Datasheet XLS Valid Flowcell Generate Flowcell Form

EC-JMGS-5049 IEM EC-JMGS-5049 SeqMon EC-SS-5047 IEM EC-SS-5047 SeqMon EC-PB-5025 IEM EC-PB-5025 SeqMon EC-HC-5058 IEM EC-HC-5058 SeqMon EC-SS-5079 IEM EC-SS-5079 SeqMon

Flowcell Samples

show/hide

Library name	Group/Pool	Library type	Species	PI	Submitter email	ILabs ID
Young_liver	1		mouse	vicente_andres	ignacio_benedicto@hotmail.com	EC-VA-495
Aged_liver	1		mouse	vicente_andres	ignacio_benedicto@hotmail.com	EC-VA-495
ami-hma-pre-1	2		human	chia-lin_wei	sheng.li@jax.org	EC-SL-495
ami-hma-post-1	2		human	chia-lin_wei	sheng.li@jax.org	EC-SL-495
ami-hma-pre-2	2		human	chia-lin_wei	sheng.li@jax.org	EC-SL-495
ami-hma-post-2	2		human	chia-lin_wei	sheng.li@jax.org	EC-SL-495
IP12424	1		human	giorgio_inghirami	giorgio.inghirami@unibo.it	EC-GI-499
B1	1		human	giorgio_inghirami	giorgio.inghirami@unibo.it	EC-GI-499
C1	1		human	giorgio_inghirami	giorgio.inghirami@unibo.it	EC-GI-499
HD	1		human	virginia_pascual	prb2011@med.cornell.edu	EC-PB-502
SLE	1		human	virginia_pascual	prb2011@med.cornell.edu	EC-PB-502
CCND2WT1_E13_5_04_18_2018	1		mouse	margaret_ross	shs2039@med.cornell.edu	EC-SS-504
CCND2WT1_E13_5_04_18_2018	1		mouse	margaret_ross	shs2039@med.cornell.edu	EC-SS-504
mouse_adult_liver_ECs			mouse	koji_shido	jmg2008@med.cornell.edu	EC-JMGS-504
RB_Plus_P53-1	1		human	harold_varmus	huc2003@med.cornell.edu	EC-HC-50!
RB_Plus_P53-2	1		human	harold_varmus	huc2003@med.cornell.edu	EC-HC-50!
CCND2WT1_E13_5_04_10_2018	pool1	rna-seq-chromium-single-cell-v2	margaret_ross	shs2039@med.cornell.edu	EC-SS-50!	
CCND2WT1_E13_5_04_10_2018	pool1	rna-seq-chromium-single-cell-v2	margaret_ross	shs2039@med.cornell.edu	EC-SS-50!	

 Epigenomics Core

Jobs

Update jobs

Show 25 entries Search:

JobId	Name	Dataset	State	Exit status	Failed	Output	Hostname	Submitted	Started	Completed
28856	errbs	vcl10612	complete	0	0	view	epicore06	2017-09-14 16:57:44	2017-09-14 23:50:05	2017-09-14 23:50:05
28855	errbs	vcl10611	complete	0	0	view	epicore07	2017-09-14 16:57:42	2017-09-14 16:57:50	2017-09-14 16:57:50
28854	errbs	vcl10608	complete	0	0	view	epicore07	2017-09-14 16:57:39	2017-09-14 16:57:50	2017-09-14 16:57:50
28853	errbs	vcl10600	complete	0	0	view	epicore07	2017-09-14 16:57:35	2017-09-14 16:57:50	2017-09-14 16:57:50
28852	errbs	vcl10600	complete	0	0	view	epicore07	2017-09-14 16:57:32	2017-09-14 16:57:35	2017-09-14 16:57:35

Doron Betel

Home

Welcome to PubShare

Please use the [Data Browser](#) to access data generated for you through the [Epigenomics Core Facility](#) and [Microbiome Core of Weill Cornell Medicine](#). Use our new **cart** feature to collect samples for a project or to share and download a collection of samples to a server with a one line command provided for you or a list of links as preferred.

Pubshare allows you to search, filter, sort and create custom groupings (carts) of samples and/or files to download or share. Should you need assistance using features available to you, please click on the  icon next to the feature or visit our [Help](#) section for assistance. Should you have additional questions, please contact us at epigenomicscore@med.cornell.edu.

Please visit the [Data Processing and File Type Descriptions](#) section of our [Frequently Asked Questions](#) for further information.

Please Note: We require that core clients acknowledge the Epigenomics Core of Weill Cornell Medicine in publications and presentations enabled by Epigenomics core resources.

Recent Data

Last login: 2018-05-11 14:00:59 UTC

[Load stats and summary data](#)

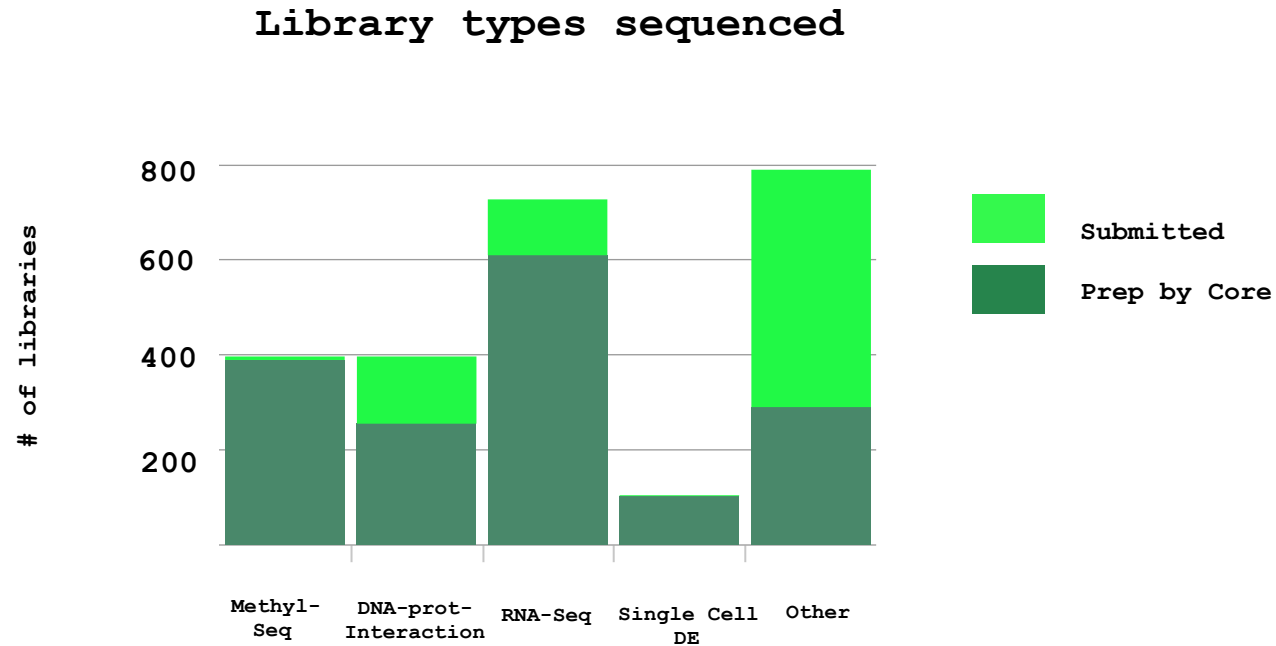


Data Summary

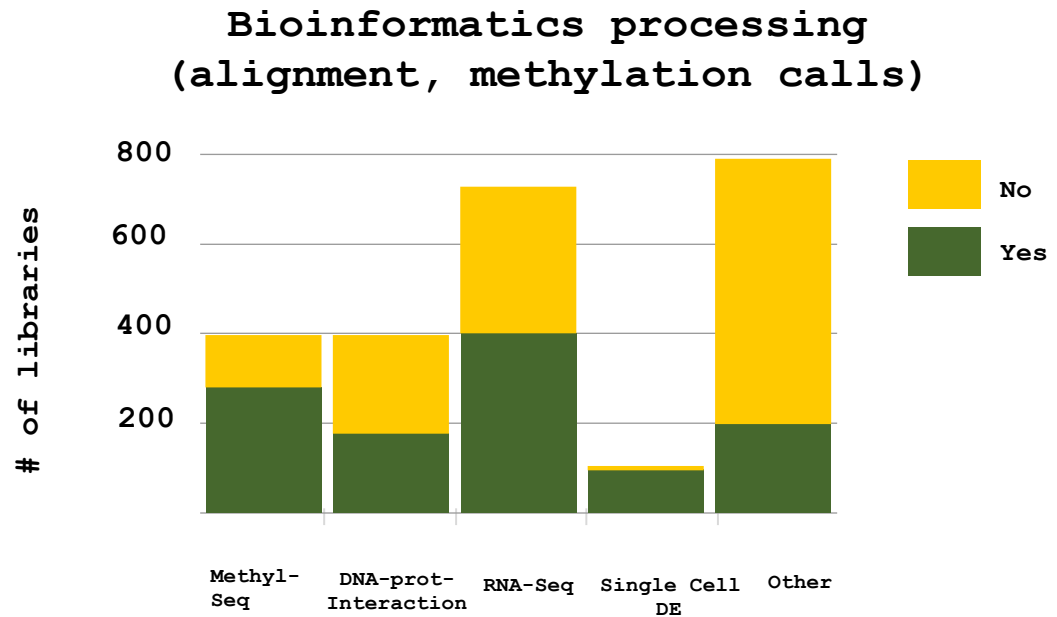
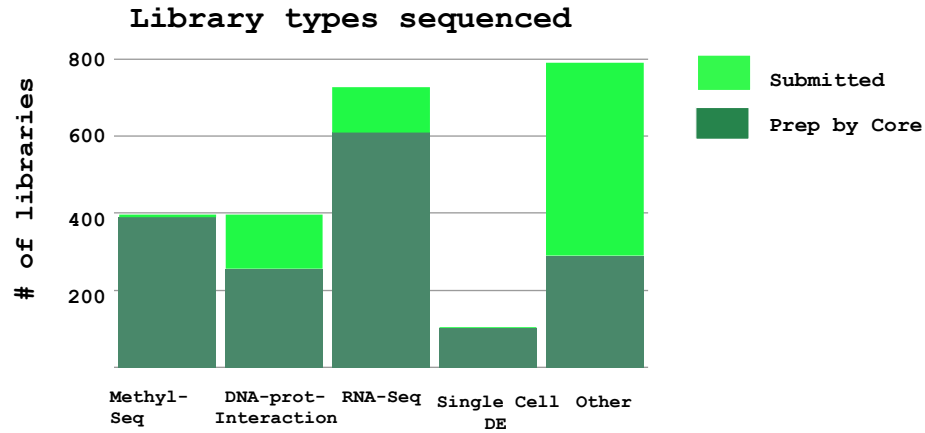
Summary table of all groups, projects by organism and library type, file counts and sizes

[Load stats and summary data](#)

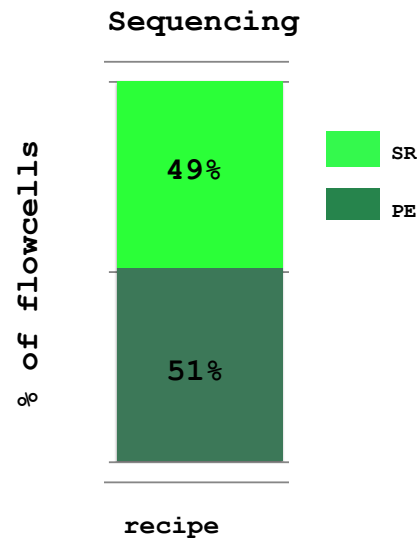
Epicore metrics FY18



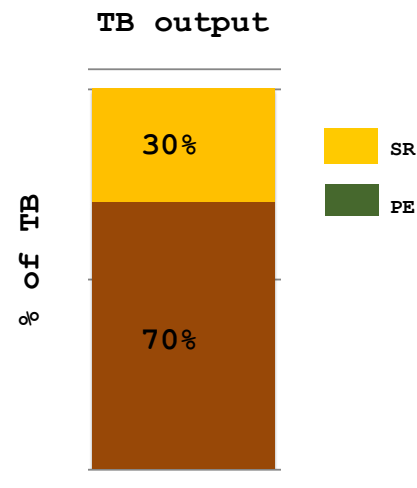
Epicore metrics FY18



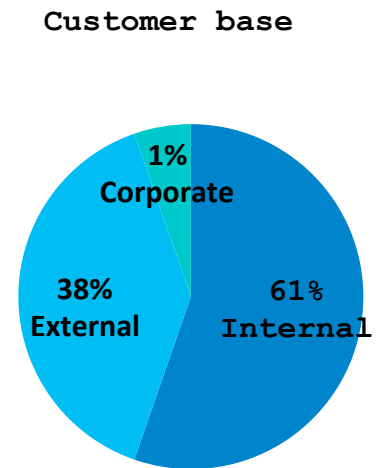
Epicore metrics FY18



90% HiSeq2500
9% HiSeq4000
1% MiSeq

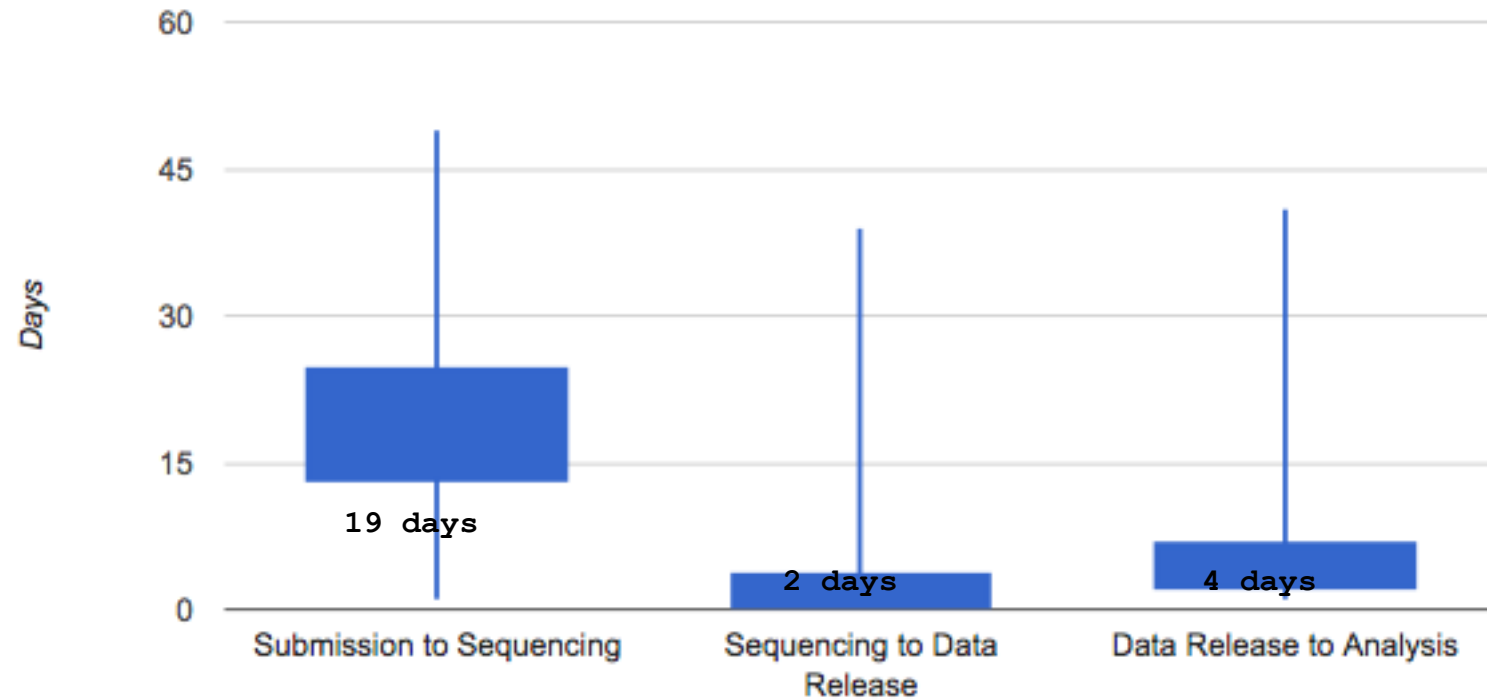


~32TB



~100 labs

Processing time FY18



Illumina NGS workflow

- **Good quality starting material:** DNA and RNA, in the past 2 years .. Add single cell!
 - DNA eukaryotic or prokaryotic, DNA methylation, from immunoprecipitations, from ATACseq, from Hi-C
 - Total RNA, immunoprecipitated RNA, RNA from FFPE (degraded)
 - Single cell suspensions
- **Library preparation:** (from DNA)
 - random fragmentation of DNA, addition of 5' and 3' adapters (which will allow binding to the Illumina flowcell) and then undergo PCR amplification if needed.

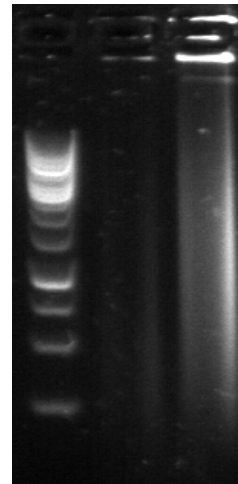
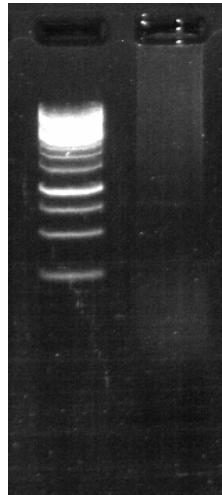
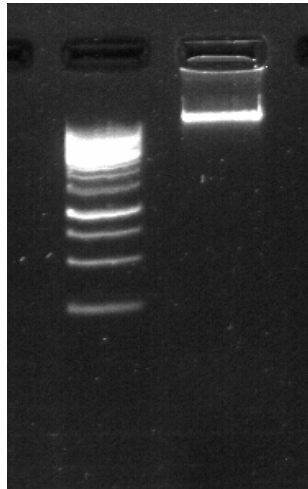
Sequencing indexing strategy must be thought out before the libraries are made

- **Cluster Generation:**
 - library is loaded into an Illumina flowcell, DNA fragments are captured on a lawn of surface bound oligos complementary to the library adapters. Fragments are amplified into clonal clusters through bridge amplification.
- **Sequencing:**
 - proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands.

Quality Control of DNA

1. Genomic DNA sample QC: ~40kb, no RNA contamination

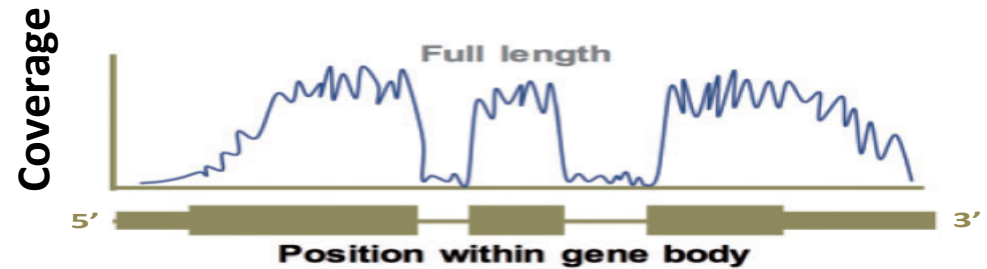
(whole genome sequencing, targeted sequencing, methylation sequencing)



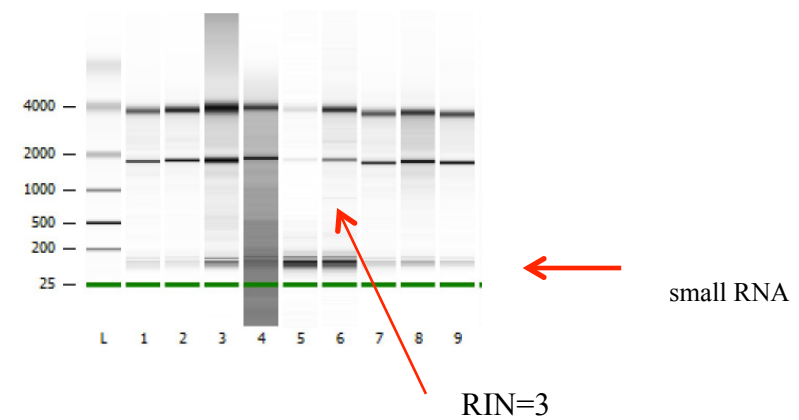
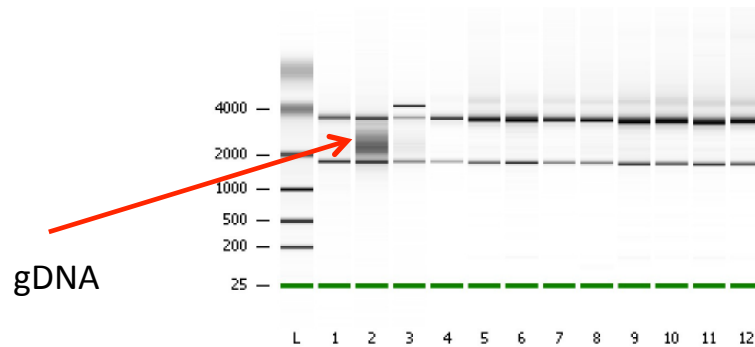
Quality Control for RNA

Transcriptome analysis

RIN number (RNA integrity number)

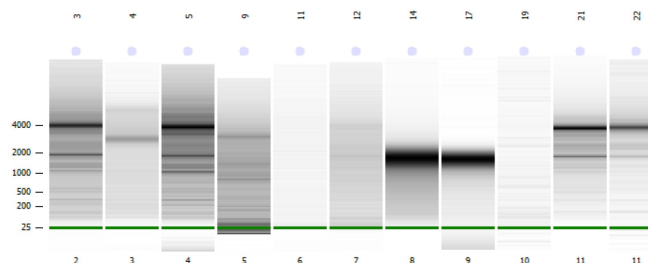


250ng RNA, RIN >- 8



Total stranded RNA prep

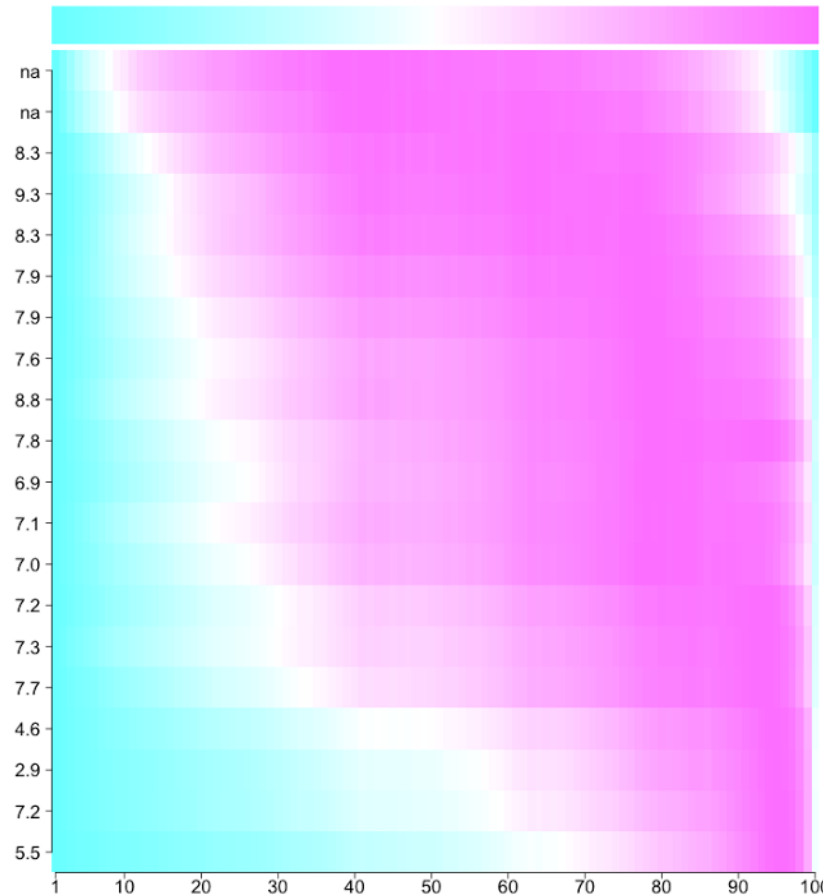
Requires 500ng RNA, RIN >- 2



Effect of RIN on gene body counts

Gene body coverage heat map

RIN



Gene body coverage (5' to 3')

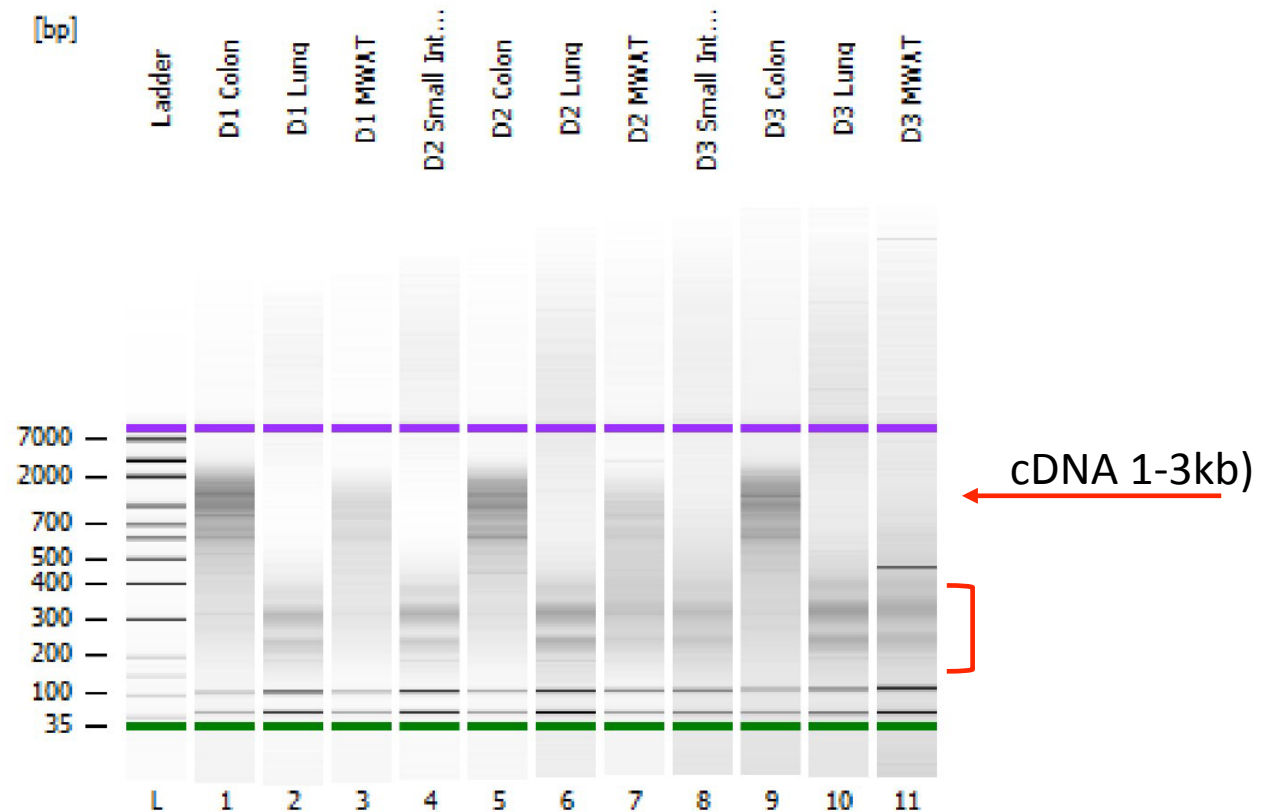
Piali Mukherjee

Quality Control for RNA (low input)

UltraLowInput RNAseq (SMART technology)

Requires 1-10ng RNA, RIN >- 8; or 10-500 cells

Quality control is done after cDNA preparation



Genome vs Epigenome

Genetic information- DNA (constant)



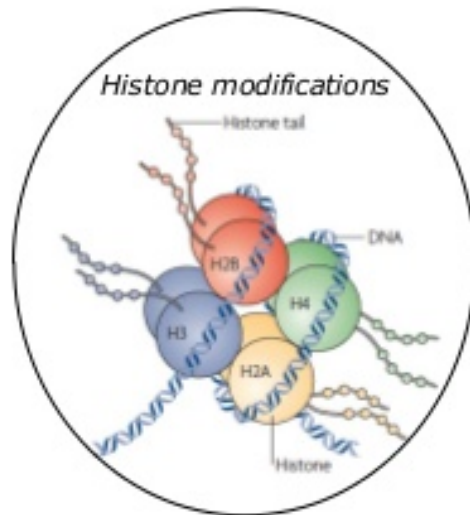
Epi-genetic
information

The Epigenetic code, which modifies gene expression, has **plasticity** and can be **reprogrammed**

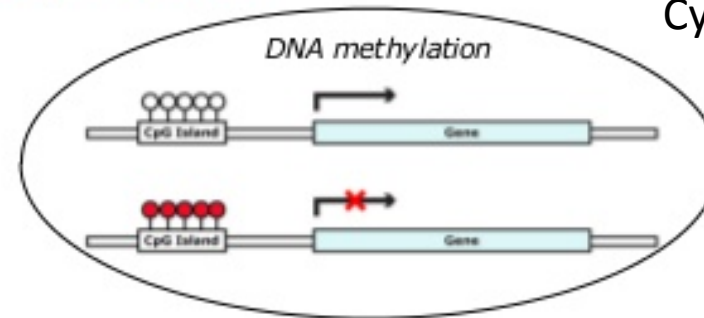
Epigenetic code affects gene expression

Many layers of epigenetic regulation

DNA-protein interactions
ChIPseq; ATACseq



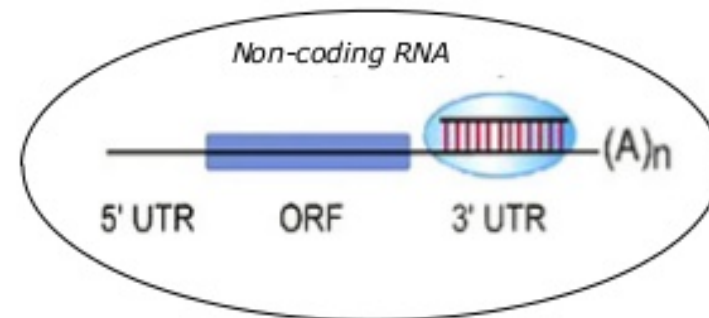
N.Tsankova,
Nat Rev Neurosc 2007



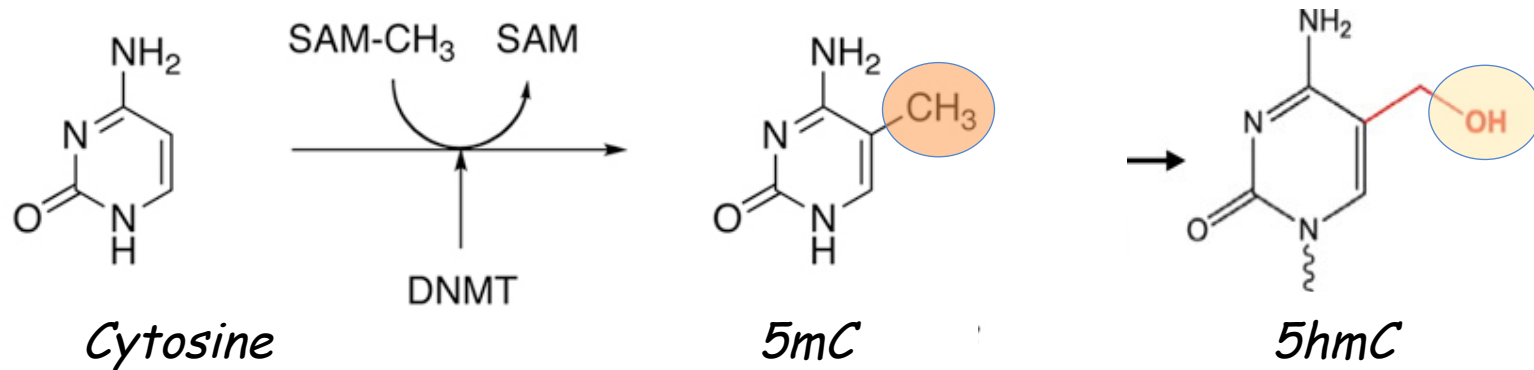
Methylation
Sequencing;
Cytosine on CpG

Gene Expression

RNAseq



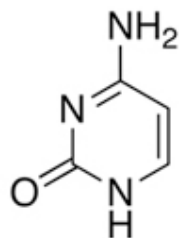
A modified C cannot be identified by sequencing by synthesis methods



Sequencing by synthesis depends on a DNA polymerase, C methylated C and all oxidative products will be amplified as a G

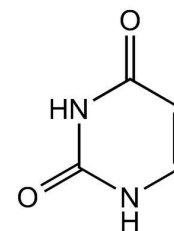
Methylation sequencing refers to a chemical step prior to library preparation (Bisulfite conversion - BS)

Step 1: conversion of Cytosine to Uracil



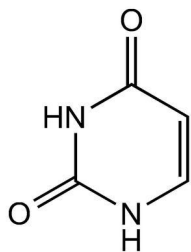
Cytosine

*Deamination
by bisulfite
treatment*



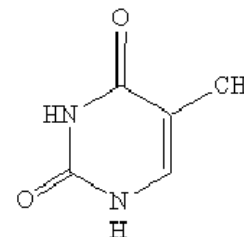
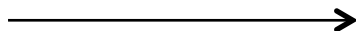
Uracil

Step 2: PCR amplification



Uracil

*PCR amplification results in
every C converted to a T*



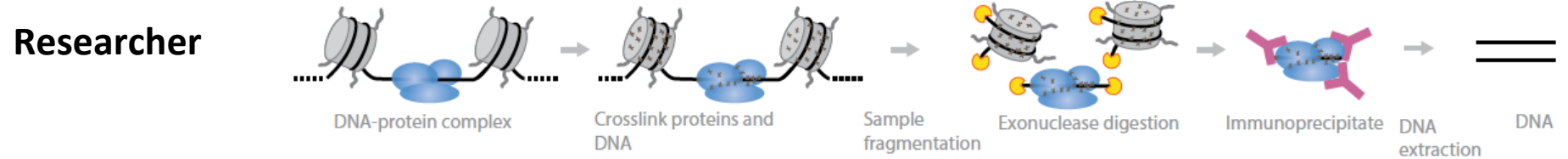
Thymine

*This chemical step preserves both 5mC and 5hmC, which are about 1% of the total Cs
It reduces the complexity of the genome (mostly) to a 3-letter code*

Quality Control of DNA

2. Genome wide mapping of DNA binding proteins

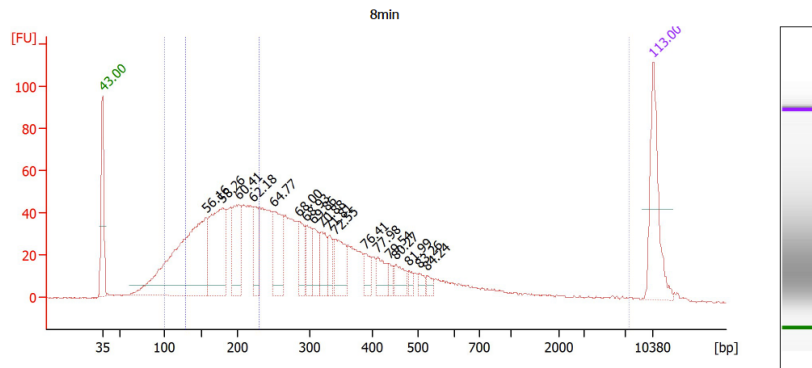
- Chromatin Immunoprecipitation.



Core:

Input requirements

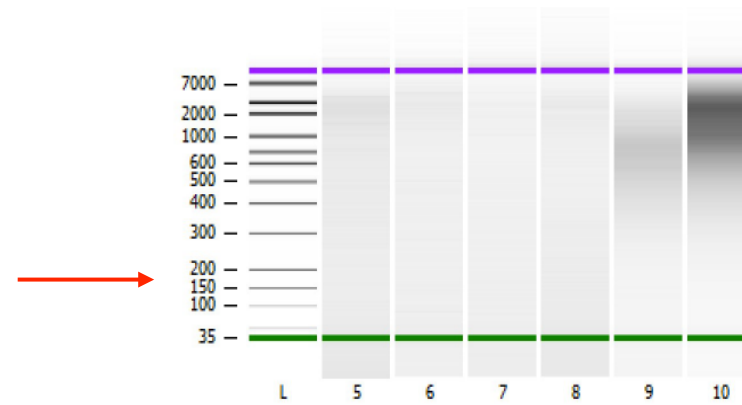
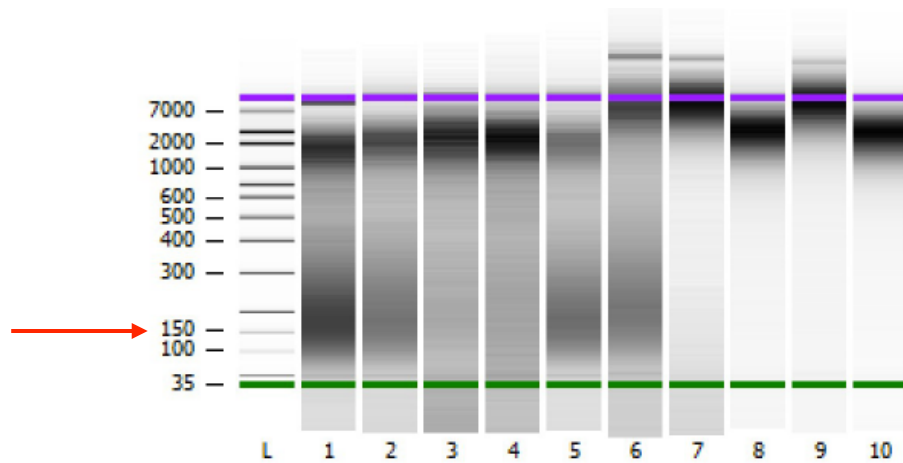
- a minimum of 11ng of sonicated ChIP DNA at $\leq 1\text{ng}/\mu\text{l}$, determined by Qubit
- 10% of total material must be in the size range of the DNA fraction required for library preparation (130-230bp)



Region table for sample 4 :				8min
From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]
100	7,500	1,795.4	95	390
130	230	675.3	36	182

Samples submitted most often look like this

Agilent Bioanalyser trace

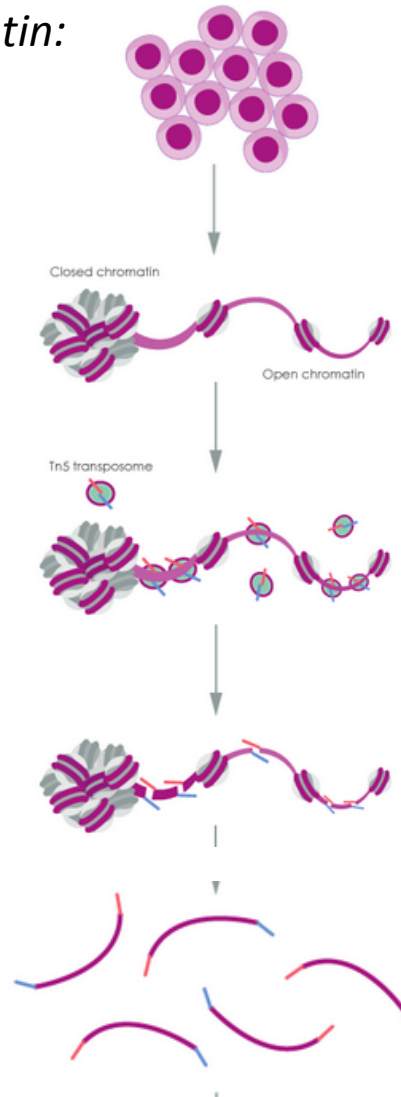


Quality control of cells

Functional state of chromatin:

Assay for transposase-accessible chromatin:
ATACseq

Technique was published by Greenleaf's lab in 2013



1. 90% live cells in suspension

2. Prepare nuclei (50,000 cells)

3. Add Tagmentase

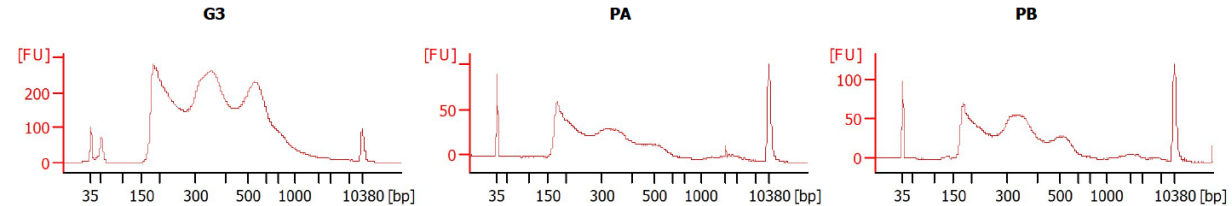
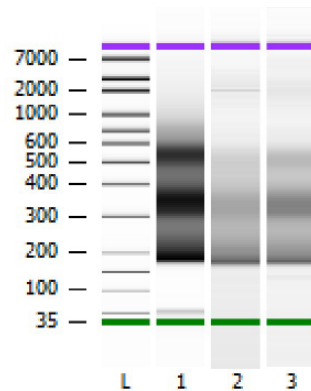
4. Short **barcoded** fragments
are released by tagmentation

5. PCR amplify and Sequence

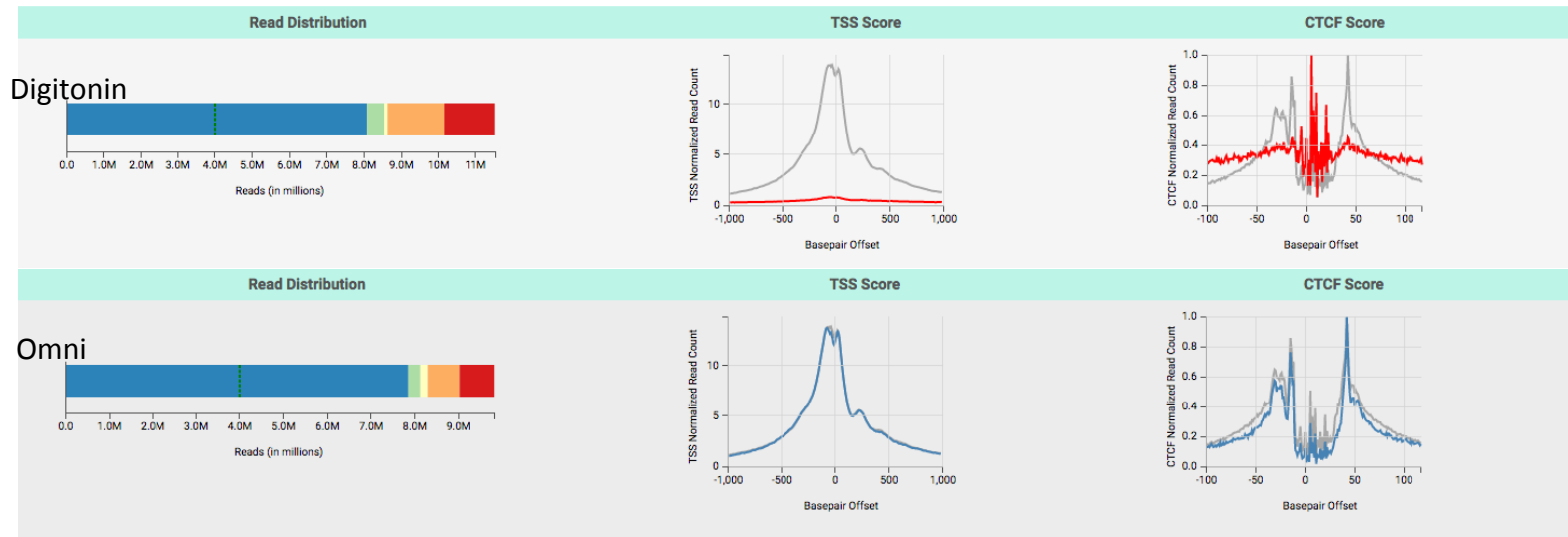
Quality control of ATACseq libraries is at the sequencing stage!

Bio

Bioanalyzer QC



QC using Epinomics website (www.epinomics.co/) Greenleaf initiative)



Illumina NGS workflow

- **Good quality nucleic acid:** DNA and RNA
 - DNA eukaryotic or prokaryotic, DNA methylation, from immunoprecipitations, from ATACseq, from Hi-C
 - Total RNA, immunoprecipitated RNA, RNA from FFPE (degraded)
- **Library preparation:** (from DNA)
 - random fragmentation of DNA, addition of 5' and 3' adapters (which will allow binding to the Illumina flowcell) and then undergo PCR amplification if needed.

Sequencing indexing strategy must be thought out before the libraries are made
- **Cluster Generation:**
 - library is loaded into an Illumina flowcell, DNA fragments are captured on a lawn of surface bound oligos complementary to the library adapters. Fragments are amplified into clonal clusters through bridge amplification.
- **Sequencing:**
 - proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands.

Library preparation

1. Shear substrate to ~200bp



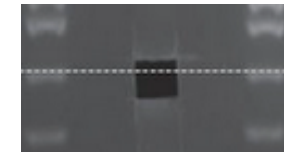
2. End repair, A-tailing



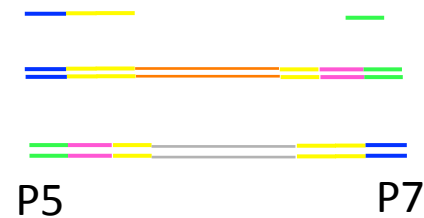
3. Adapter ligation



4. Size selection of DNA from 250-350bp
(original DNA ~130-230bp)

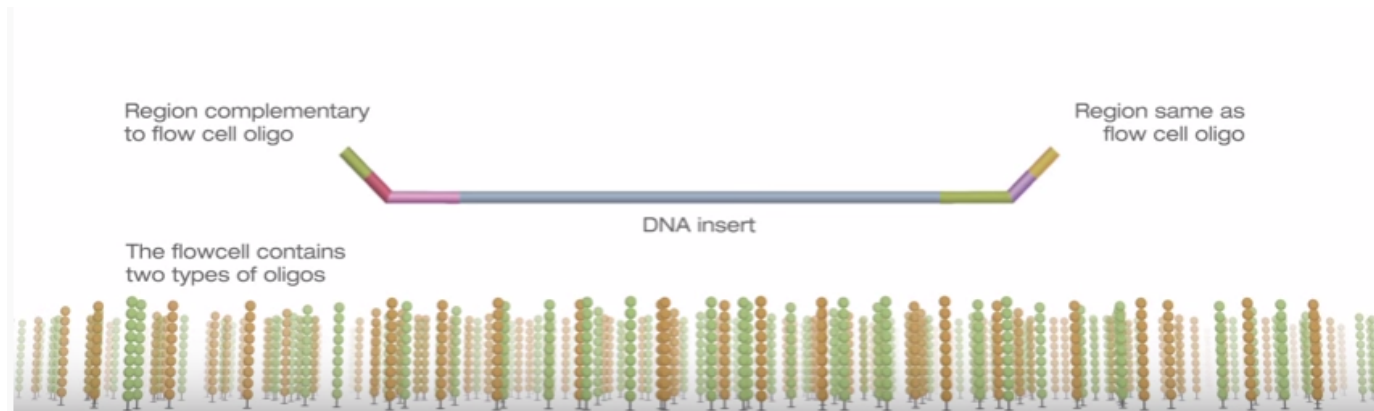


5. PCR amplification



Library preparation

- **Library preparation:**
 - fragmentation of DNA
 - addition of 5' and 3' adapters (will tether the DNA to the Illumina flowcell)
 - PCR amplification if needed.
 - 5' end has a UNIVERSAL primer



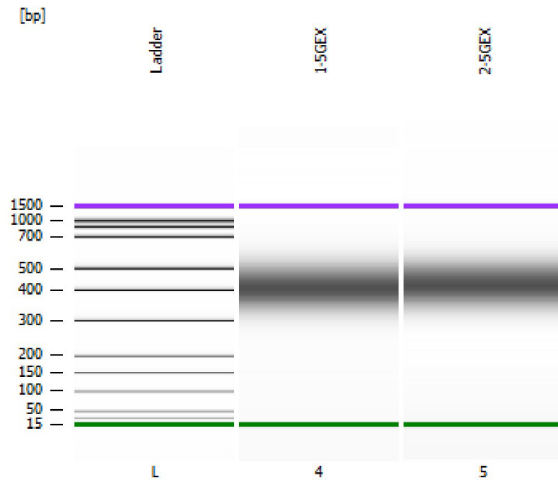
Flow cell



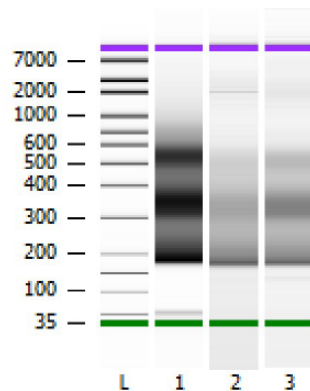
Quality Control of Library preparation

Before sequencing, libraries are analyzed for quantity and quality

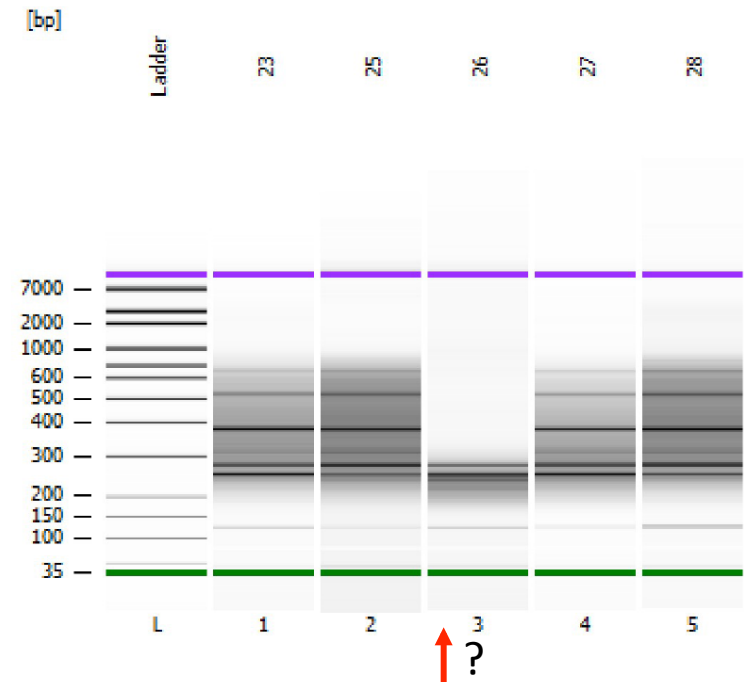
RNAseq libraries



ATACseq libraries



36 gDNA for methylation sequencing (RRBS)
, all passed DNA QC ...



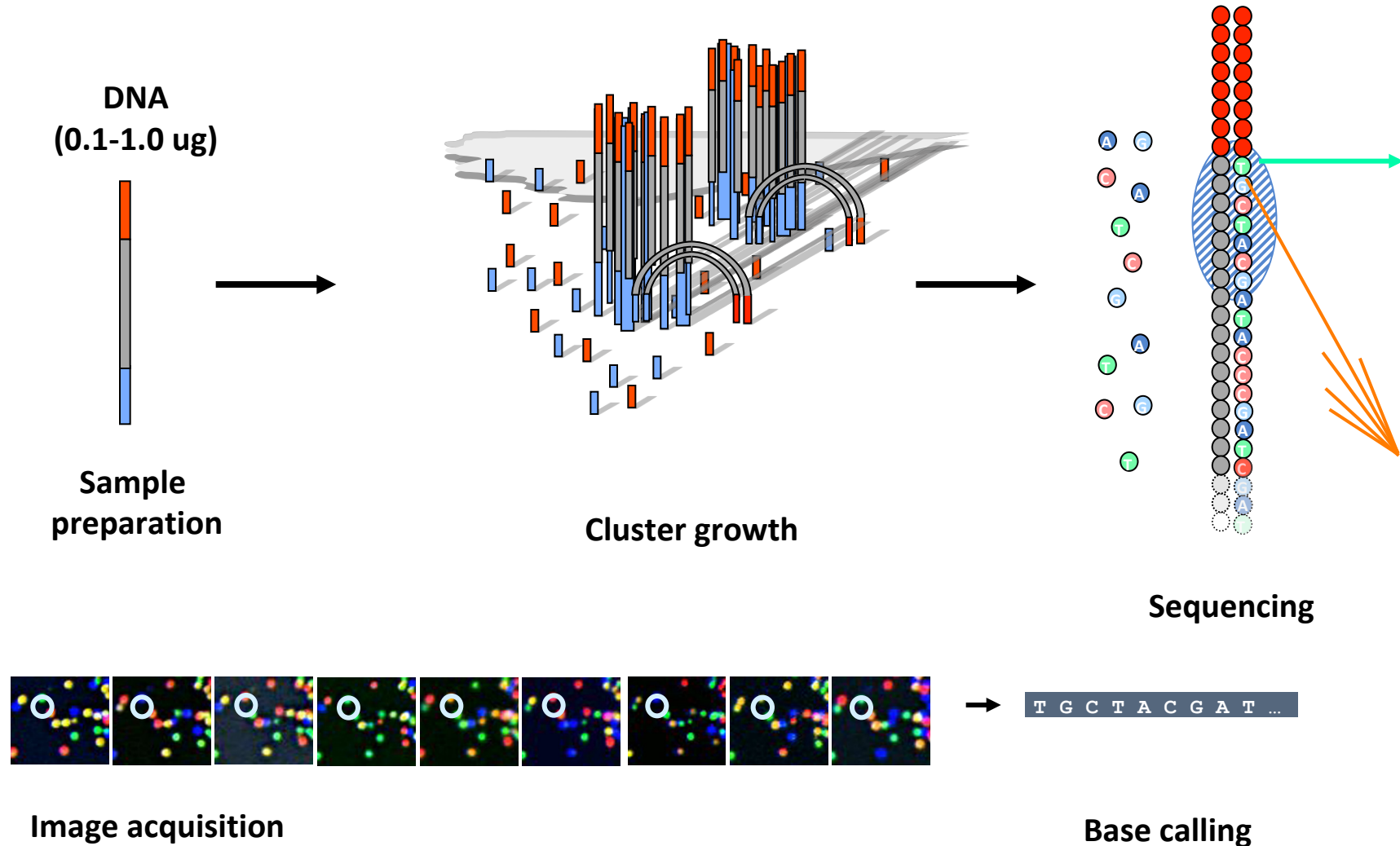
Illumina NGS workflow

- **Good quality nucleic acid:** DNA and RNA
 - DNA eukaryotic or prokaryotic, DNA methylation, from immunoprecipitations, from ATACseq, from Hi-C
 - Total RNA, immunoprecipitated RNA, RNA from FFPE (degraded)
- **Library preparation:** (from DNA)
 - random fragmentation of DNA, addition of 5' and 3' adapters (which will allow binding to the Illumina flowcell) and then undergo PCR amplification if needed.

Sequencing indexing strategy must be thought out before the libraries are made

- **Cluster Generation:**
 - library is loaded into an Illumina flowcell, DNA fragments are captured on a lawn of surface bound oligos complementary to the library adapters. Fragments are amplified into clonal clusters through bridge amplification.
- **Sequencing:**
 - proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands.

Illumina Clustering and Sequencing



Sequencing run quality control

This is a good run:

Run: 181114_D00796_ACCVYFANXX: Charts

SUMMARY

BIO SAMPLES

CHARTS

METRICS

INDEXING QC

SAMPLE SHEET

FILES

Launch App

File

Status

Share

Move to Trash

Flow Cell: CCVYFANXX

Extracted: 110

Called: 110

Scored: 110

Flow Cell Chart

Chart

Intensity

Surface

Both Surfaces

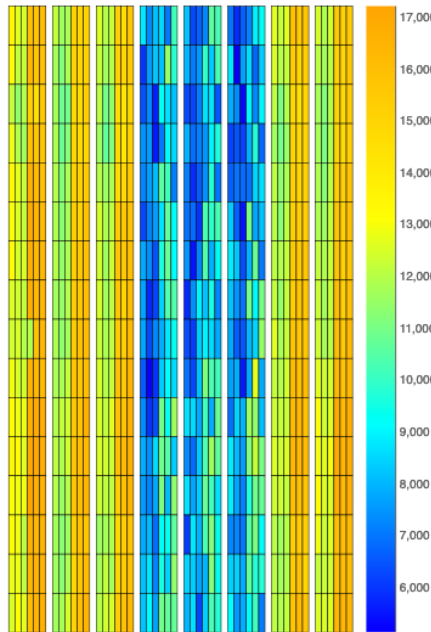
Cycle

Cycle 1

Channel

A

☐ Fix Scale



Data By Cycle

Chart

% Base

Surface

Both Surfaces

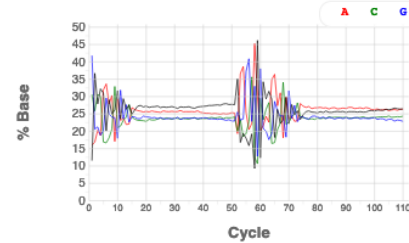
Lane

All Lanes

Base

All Bases

☐ Fix Scale



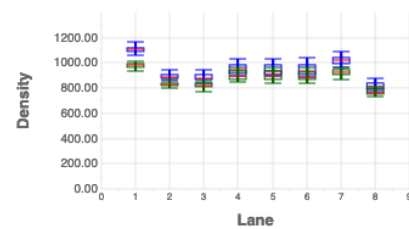
Data By Lane

Show

Density

Surface

Both Surfaces



QScore Distribution

Surface

Both Surfaces

Lane

All Lanes

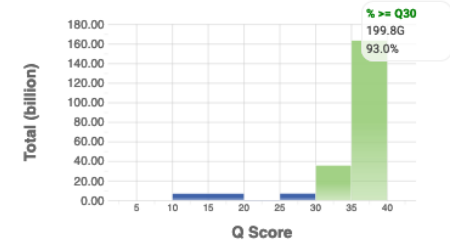
Read

All Reads

Cycle

All Cycles

☐ Fix Scale



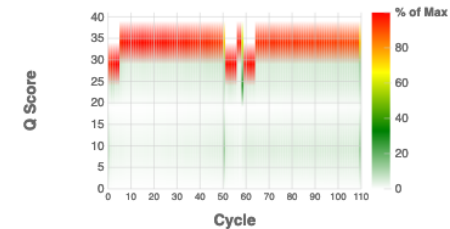
QScore Heatmap

Surface

Both Surfaces

Lane

All Lanes



Sequencing run quality control

Some lanes on this run need to be repeated

Run: 190325_K00237_BH3GJJBBXY: Charts

Launch App File Status Share Move to Trash

Flow Cell: H3GJJBBXY Extracted: 59 Called: 59 Scored: 59

Flow Cell Chart

Chart

Intensity

Surface

Both Surfaces

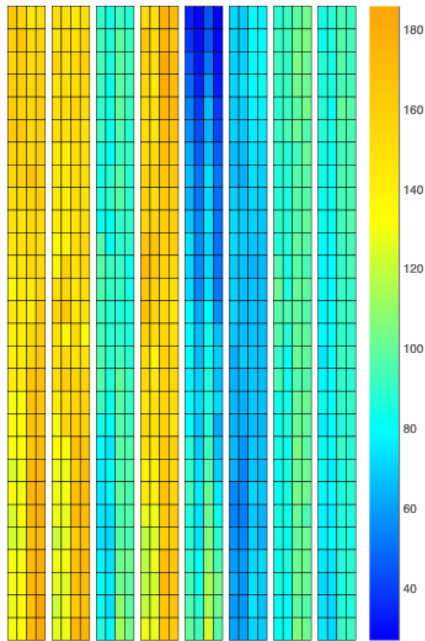
Cycle

Cycle 1

Channel

A

Fix Scale



Data By Cycle

Chart

% Base

Surface

Both Surfaces

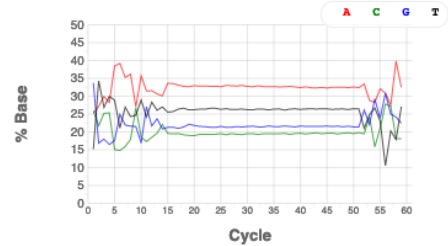
Lane

All Lanes

Base

All Bases

Fix Scale



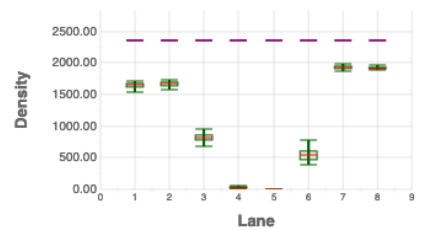
Data By Lane

Show

Density

Surface

Both Surfaces



QScore Distribution

Surface

Both Surfaces

Lane

All Lanes

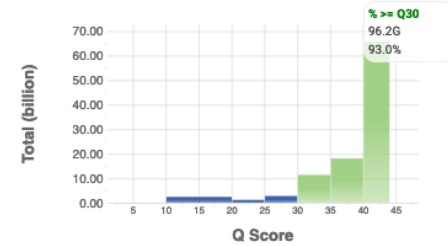
Read

All Reads

Cycle

All Cycles

Fix Scale



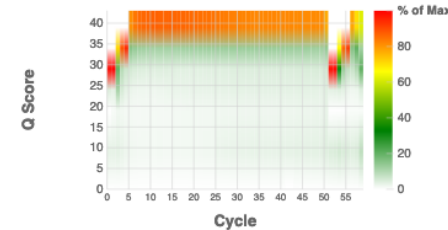
QScore Heatmap

Surface

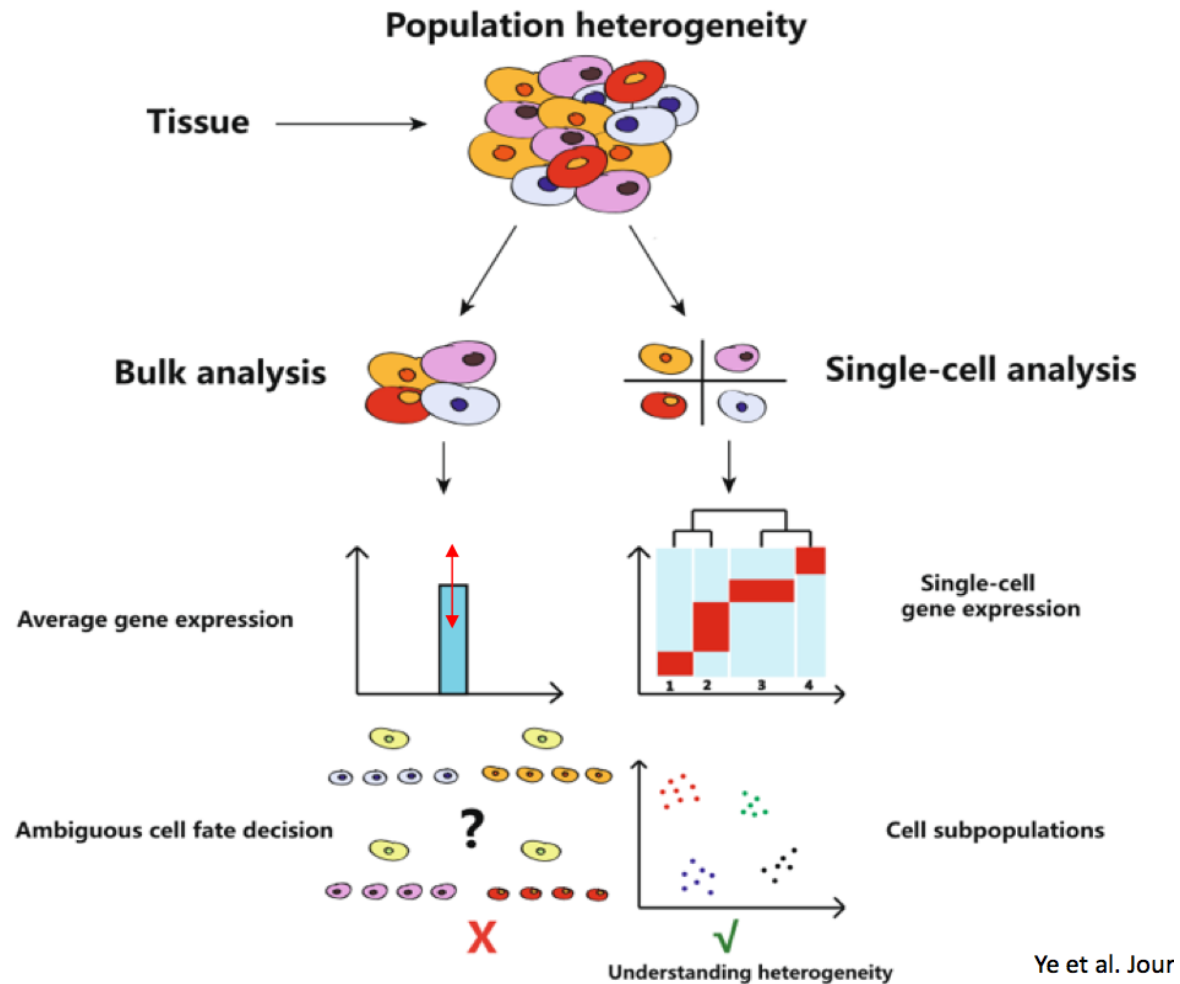
Both Surfaces

Lane

All Lanes

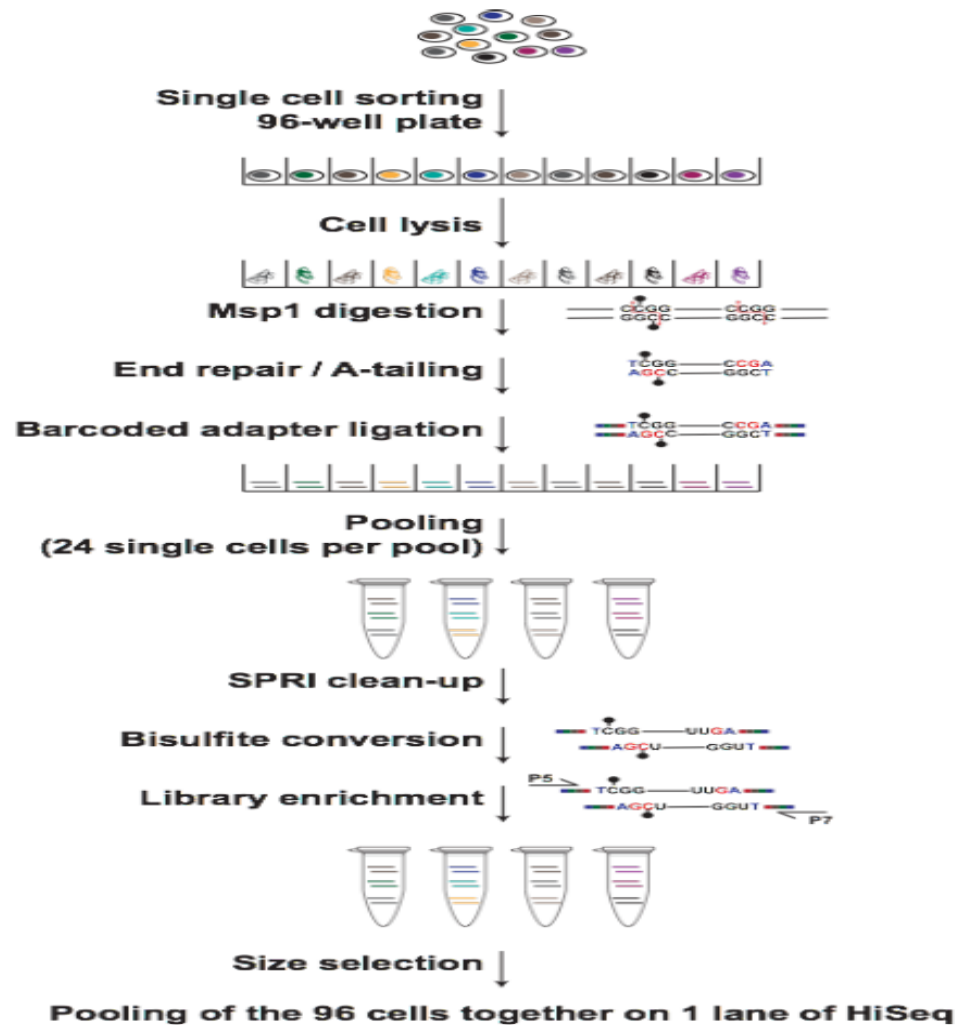


Bulk vs Single Cell Genomics



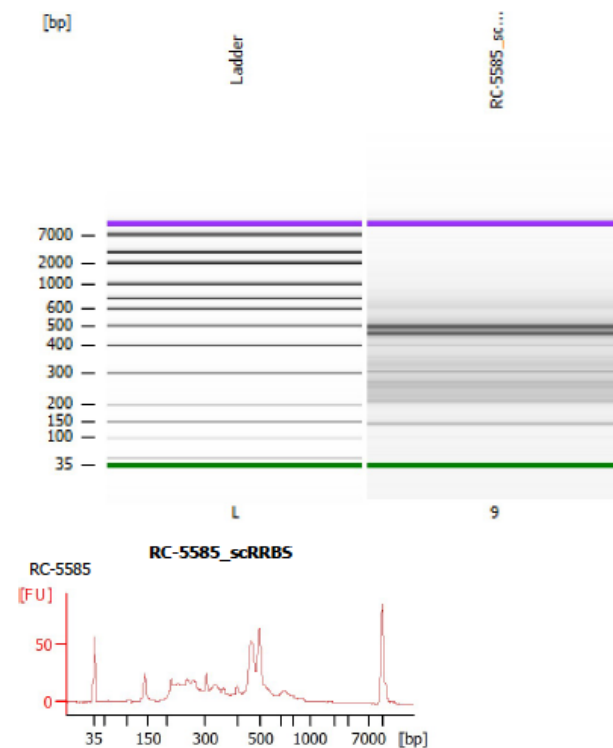
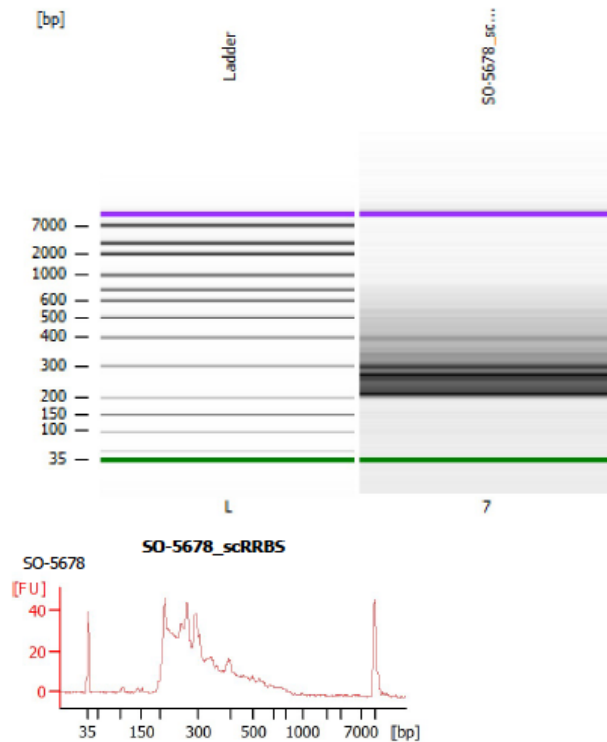
Single cells methylation sequencing (RRBS)

Manual, 96 cells at a time



Single Cell Quality Control of Library preparation

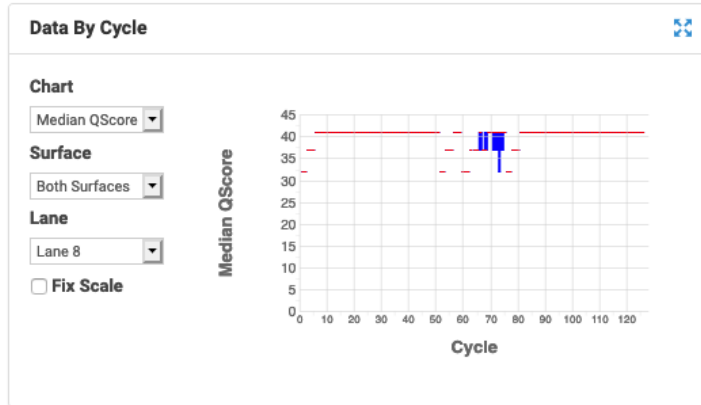
But, for single cell ...



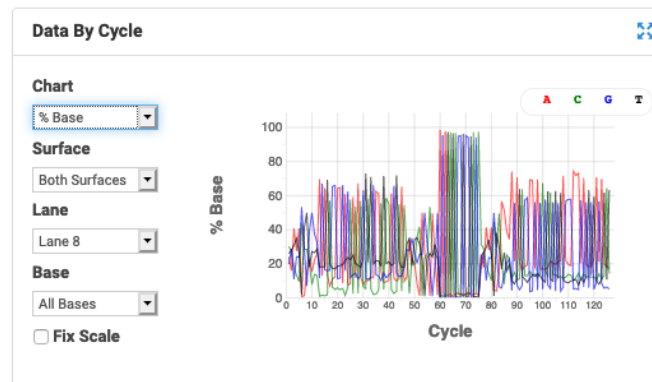
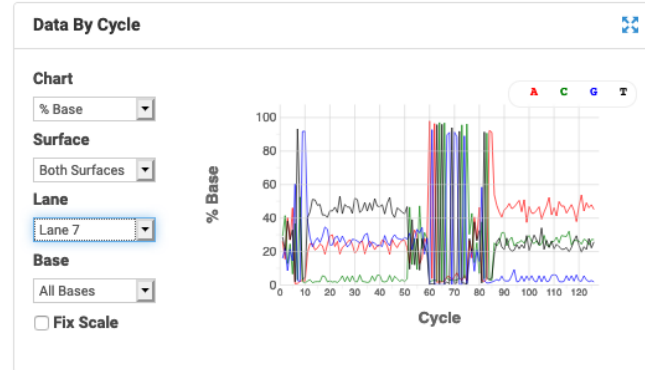
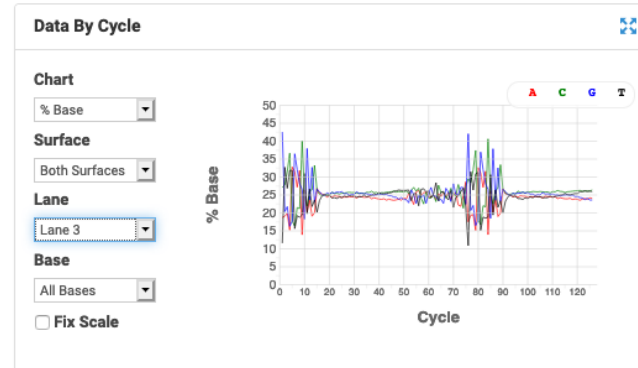
Single Cell Quality Control of Sequencing Run

But, for single cell ...

Sequencing QC



Base Composition

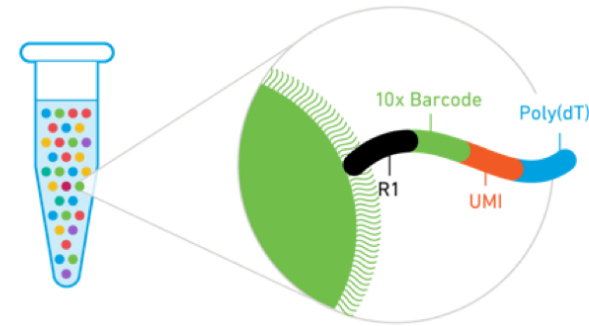


Single cell transcriptomics

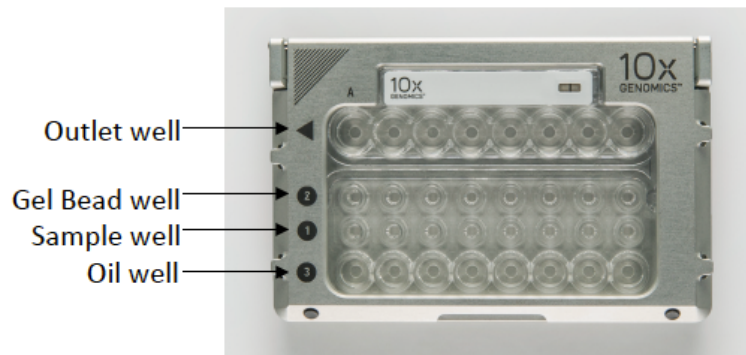
10 X Chromium Platform:
Single Cell 3' Digital Expression 500-10,000 cells

3 Components:

1. Gel beads with 750,000 unique barcodes
2. Cell suspension with RT reagents
3. Partitioning oil



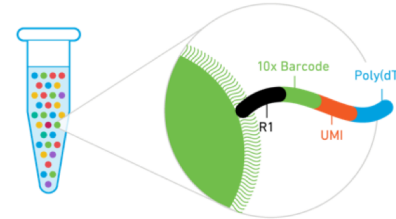
Single-use microfluidics chip



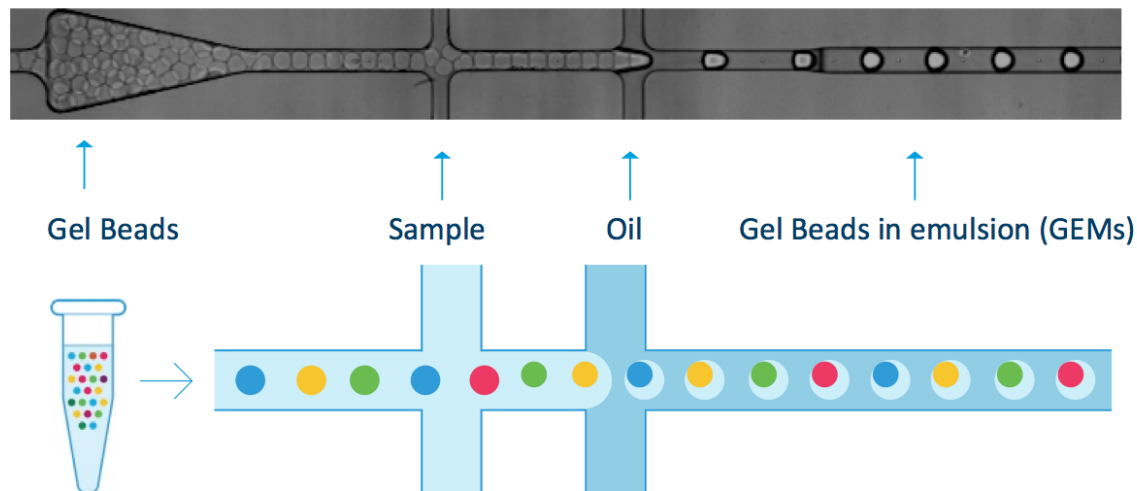
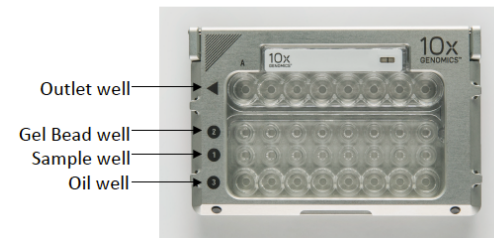
Single Cell transcriptomics

3 Components:

1. Gel beads with 750,000 unique barcodes
2. Cell suspension with RT reagents
3. Partitioning oil

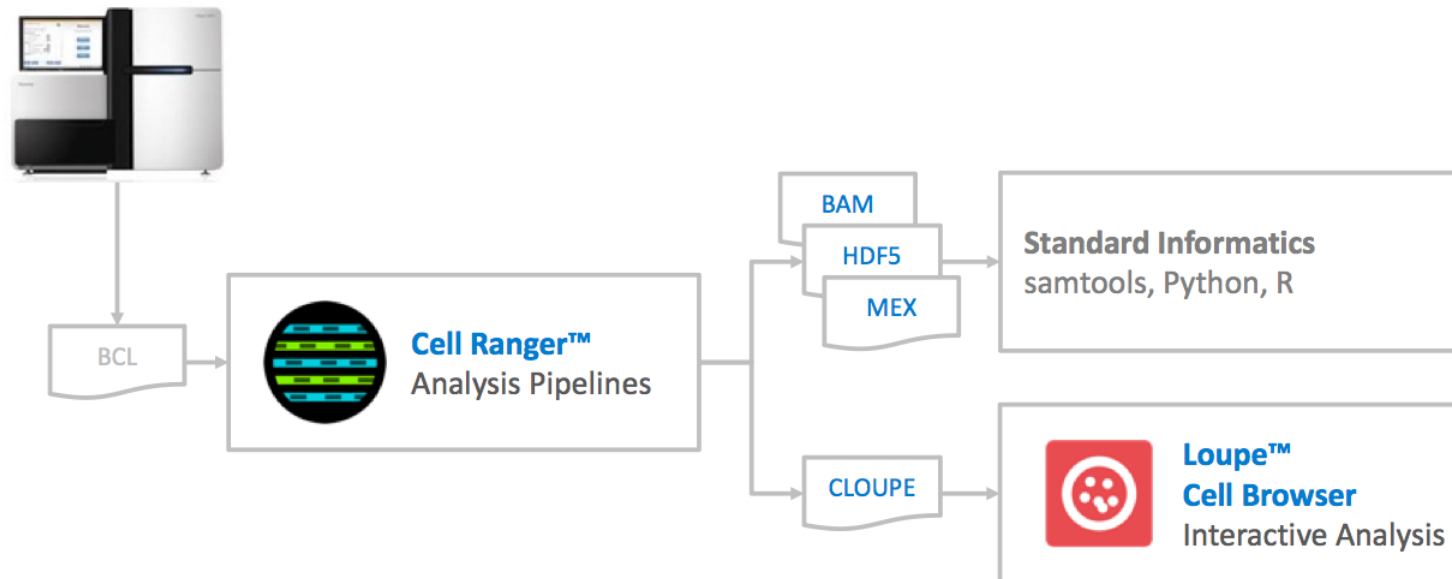


Single-use microfluidics chip



Chromium platform, Bioinformatics Support

- Sequence Chromium libraries
- Cell Ranger™ pipeline converts sequence data to single cell gene expression profiles
- Loupe™ Cell Browser enables interactive analysis



10 X Chromium Platform: Single Cell 3' Digital Expression

Cell Ranger analysis pipeline

Estimated Number of Cells

3,520

Mean Reads per Cell

41,267

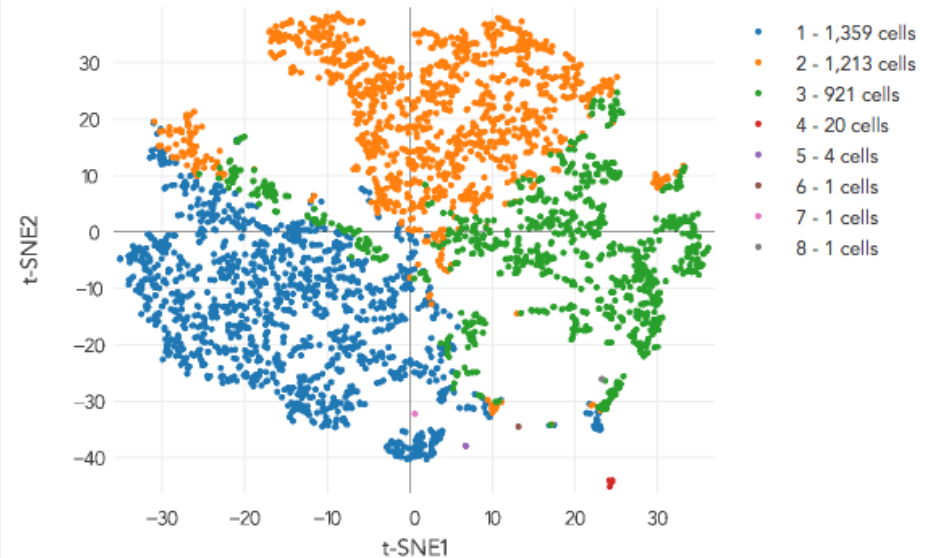
Median Genes per Cell

2,151

Sequencing

Number of Reads	145,260,501
Valid Barcodes	98.3%
Reads Mapped Confidently to Transcriptome	55.5%
Reads Mapped Confidently to Exonic Regions	58.6%
Reads Mapped Confidently to Intronic Regions	15.3%
Reads Mapped Confidently to Intergenic Regions	3.4%
Sequencing Saturation	61.4%
Q30 Bases in Barcode	95.9%
Q30 Bases in RNA Read	87.5%
Q30 Bases in Sample Index	95.6%
Q30 Bases in UMI	96.4%

t-SNE projection of Cells Colored by k-means Clustering



10 X Chromium Platform: Single Cell 3' Digital Expression

Sometimes, it fails ...

Estimated Number of Cells
3,574

Mean Reads per Cell
48,598

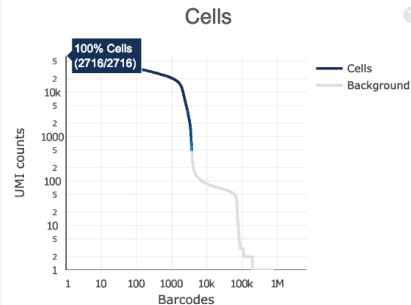
Median Genes per Cell
3,255

Sequencing

Number of Reads	173,689,296
Valid Barcodes	98.5%
Sequencing Saturation	46.1%
Q30 Bases in Barcode	98.9%
Q30 Bases in RNA Read	61.3%
Q30 Bases in Sample Index	95.5%
Q30 Bases in UMI	98.7%

Mapping

Reads Mapped to Genome	90.7%
Reads Mapped Confidently to Genome	88.0%
Reads Mapped Confidently to Intergenic Regions	4.8%
Reads Mapped Confidently to Intronic Regions	19.2%
Reads Mapped Confidently to Exonic Regions	64.1%
Reads Mapped Confidently to Transcriptome	61.4%
Reads Mapped Antisense to Gene	0.6%



Estimated Number of Cells	3,574
Fraction Reads in Cells	90.9%
Mean Reads per Cell	48,598
Median Genes per Cell	3,255
Total Genes Detected	21,018
Median UMI Counts per Cell	14,527

Sample

Name	1412A
Description	
Transcriptome	hg19
Chemistry	Single Cell 3' v3
Cell Ranger Version	3.0.0

Estimated Number of Cells

184

Mean Reads per Cell
959,423

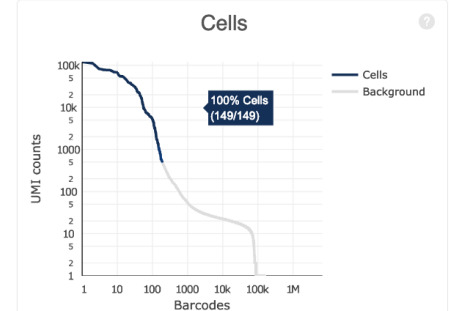
Median Genes per Cell
1,856

Sequencing

Number of Reads	176,533,869
Valid Barcodes	98.6%
Sequencing Saturation	95.8%
Q30 Bases in Barcode	98.5%
Q30 Bases in RNA Read	61.7%
Q30 Bases in Sample Index	95.4%
Q30 Bases in UMI	98.1%

Mapping

Reads Mapped to Genome	92.1%
Reads Mapped Confidently to Genome	88.4%
Reads Mapped Confidently to Intergenic Regions	5.3%
Reads Mapped Confidently to Intronic Regions	19.3%
Reads Mapped Confidently to Exonic Regions	63.7%
Reads Mapped Confidently to Transcriptome	61.1%
Reads Mapped Antisense to Gene	0.7%



Estimated Number of Cells	184
Fraction Reads in Cells	83.7%
Mean Reads per Cell	959,423
Median Genes per Cell	1,856
Total Genes Detected	16,925
Median UMI Counts per Cell	5,787

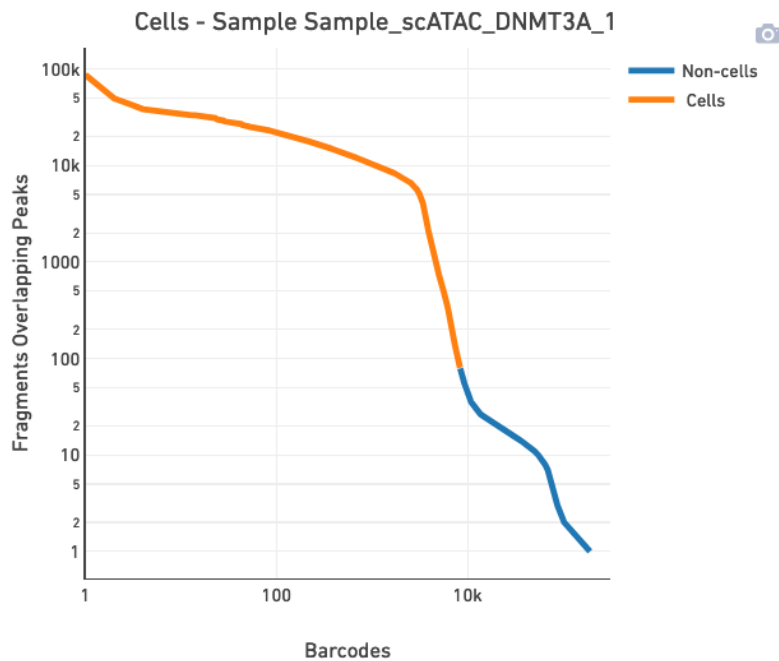
Sample

Name	1429A
Description	
Transcriptome	hg19
Chemistry	Single Cell 3' v3
Cell Ranger Version	3.0.0

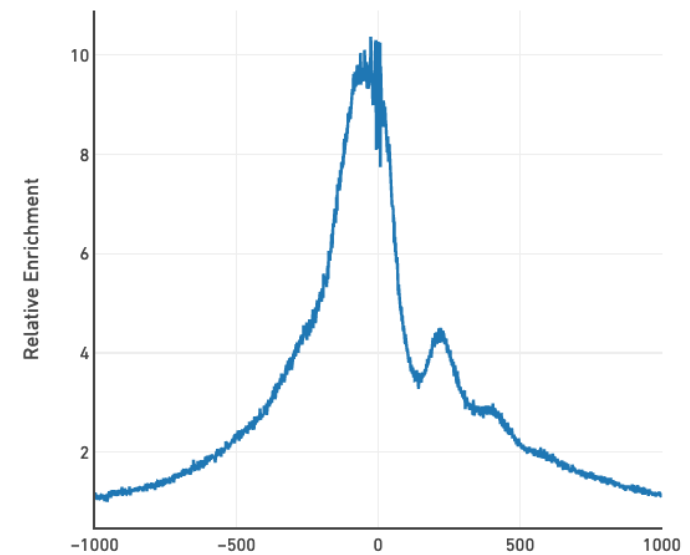
10 X Chromium Platform: Single Cell ATACseq

Cells ?

Estimated number of cells	8,352
Lower threshold on the number of fragments overlapping peaks per barcode to annotate barcode as cell	80.00
Median fragments per cell	2,380
Median fragments per non-cell barcode	1



Enrichment around TSS (normalized) - Sample Sample_scATAC_DNMT3A_1



Genomic analysis keep expanding, we keep expanding!

In development (collaboration with Landau's lab)

- Single cell methylation (RRBS)/RNAseq – single cell plate
- Single cell ATACseq/Rnaseq –single cell plate
- With 10x platform
 - Single cell expression with associated genotyping (paper re-submitted)

Testing commercial platforms

- Tapestry, Mission Bio –high throughput single cell DNA tumor variant phenotypes
- GeoMx DPS, Nanostring – spatial genomics –allows to detect both protein and RNA expression at a 10 μ M/100 μ M resolution

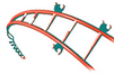
With Genomics Core

- Transition from HiSeq2500/4000 to NovaSeq

In spite of daily challenges, our core thrives!



Epigenomics Core



Home

Services

People

Resources

Overview

The **Epigenomics Core Facility** of [Weill Cornell Medicine](#) provides an array of epigenomics and bioinformatics research resources and services that include:

- [DNA methylation profiling](#) [\[More\]](#)
- [Protein-nucleic acid association](#) [\[More\]](#)
- [Single Cell Transcriptomics, Immune Profiling and Epigenomics](#) [\[More\]](#)
- [Long Range DNA Sequencing](#) [\[More\]](#)
- [Bioinformatics analysis](#) [\[More\]](#)

Core resources and services include sample preparation services and data generation on the Illumina HiSeq 2500, HiSeq 4000, NextSeq and MiSeq platforms.

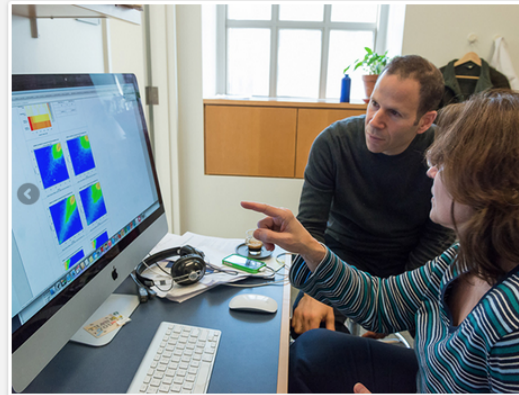
Getting Started

Request Services

Pubshare

FAQ

Pricelist



Recent Publications

We require that core clients acknowledge the Epigenomics Core of Weill Cornell Medical College in publications and presentations enabled by our resources.

Check out our latest paper on the [comprehensive evaluation of four DNA methylation measurement platforms](#):

Kacmarczyk T.J. et al. *Epigenetics Chromatin*. 2018. May 25;11(1):21.

[\[PubMed\]](#)

- Moriyama, S. et al. *Science*. 2018. 359(6379):1056-1061. [\[PubMed\]](#)
- Klose, C.S.N. et al. *Nature*. 2017. 549, 282-286. [\[PubMed\]](#)
- Wang, H. et al. *Nature Commun.* 2017. 8(1):767. [\[PubMed\]](#)
- McIntyre, ABR. et al. *Genome Biol.* 2017. 18(1):182. [\[PubMed\]](#)
- Vu, LP, Pickering BF., Cheng Y. et al. *Nat Med.* 2017. Sept 18 EPub. [\[PubMed\]](#)
- Zhang, T. et al. *Blood Cancer J.* 2017. 7(9):e606. [\[PubMed\]](#)
- Xu, R. et al. *J Bone Miner Res.* 2017. 32(9):1811-1815. [\[PubMed\]](#)
- Glass, J.L. et al. *Cancer Discov.* 2017. 7(8):868-883. [\[PubMed\]](#)
- Dhirgra, P. et al. *Genome Biol.* 2017. 18(1):141. [\[PubMed\]](#)
- Murata, K. et al. *Immunity*. 2017. 47(1):66-79.e5. [\[PubMed\]](#)
- Loupasakis, K. et al. *PLoS One*. 2017. 12(7):e0179762. [\[PubMed\]](#)
- Kan, L. et al. *Nat Commun.* 2017. 8:15737. [\[PubMed\]](#)
- Kurt, IC. et al. *Cel Death Dis.* 2017. 8(6):e2897 [\[PubMed\]](#)
- Siqueira, LG. et al. *Biol Reprod.* 2017. 96(4):743-757. [\[PubMed\]](#)
- Anelli, V. et al. *Elife*. 2017. pii: e20728. [\[PubMed\]](#)

News

Sep 2018: In order to serve our scientific community more efficiently, all RNASeq, DNaseSeq and sequencing will be performed by the Weill Cornell Medicine Genomics Resources Core Facility (GRCF). **If you require these services** - please submit your samples directly to the [GRCF](#). Please contact [Dr. Jenny Xiang](#) if you have any questions about these services. If you have [ongoing/legacy projects](#), questions or concerns, please contact us at epigenomicscore@med.cornell.edu. **If you require an Epigenomics Core library preparation service**, submit a request as before (see the Getting Started section in our [Services](#) page).

Aug 2018: The Epigenomics Core will be **discontinuing RNASeq, DNaseSeq and Sequencing Only services next month**. To continue availing of these services, please contact [Dr. Jenny Xiang](#) at the Genomics Resources Core Facility (GRCF). [\[GRCF Service Request Portal\]](#)

May 2018: Introducing **Microbiome Services** - officially launching June 2018! [\[Microbiome Core\]](#)

Tweets by @wcmepigenomics

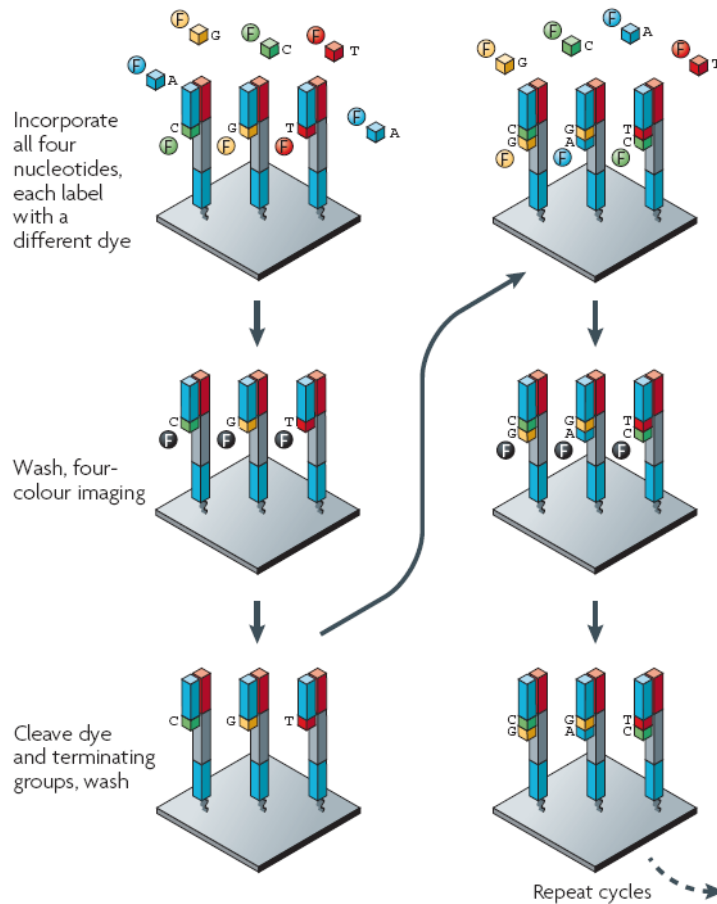
Epigenomics Core Retweeted

Doron Betel
@brthalal

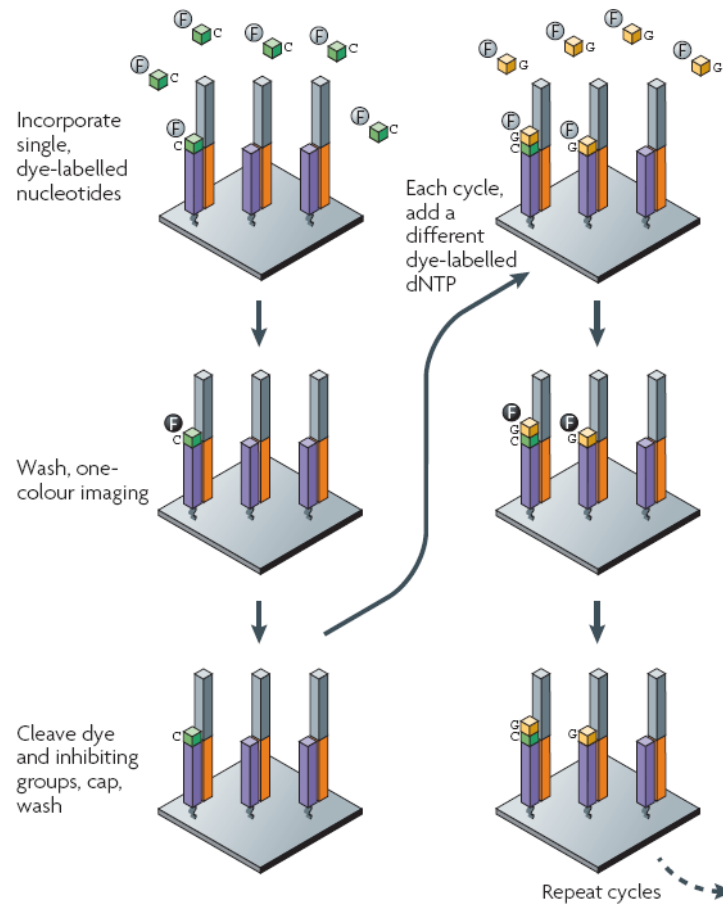
THE END

Sequencing by Synthesis (SBS)

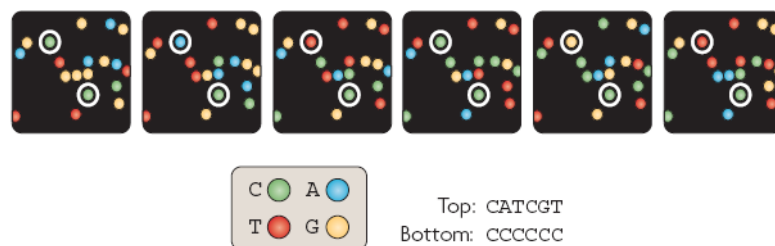
a Illumina/Solexa — Reversible terminators



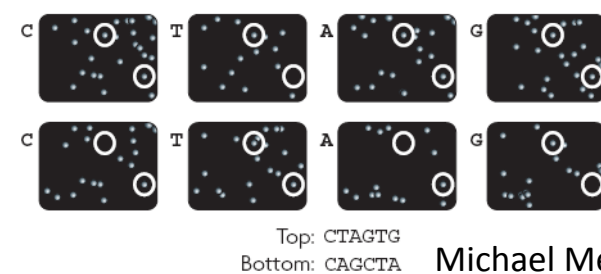
c Helicos BioSciences — Reversible terminators



b



d



Now three kinds of chemistry

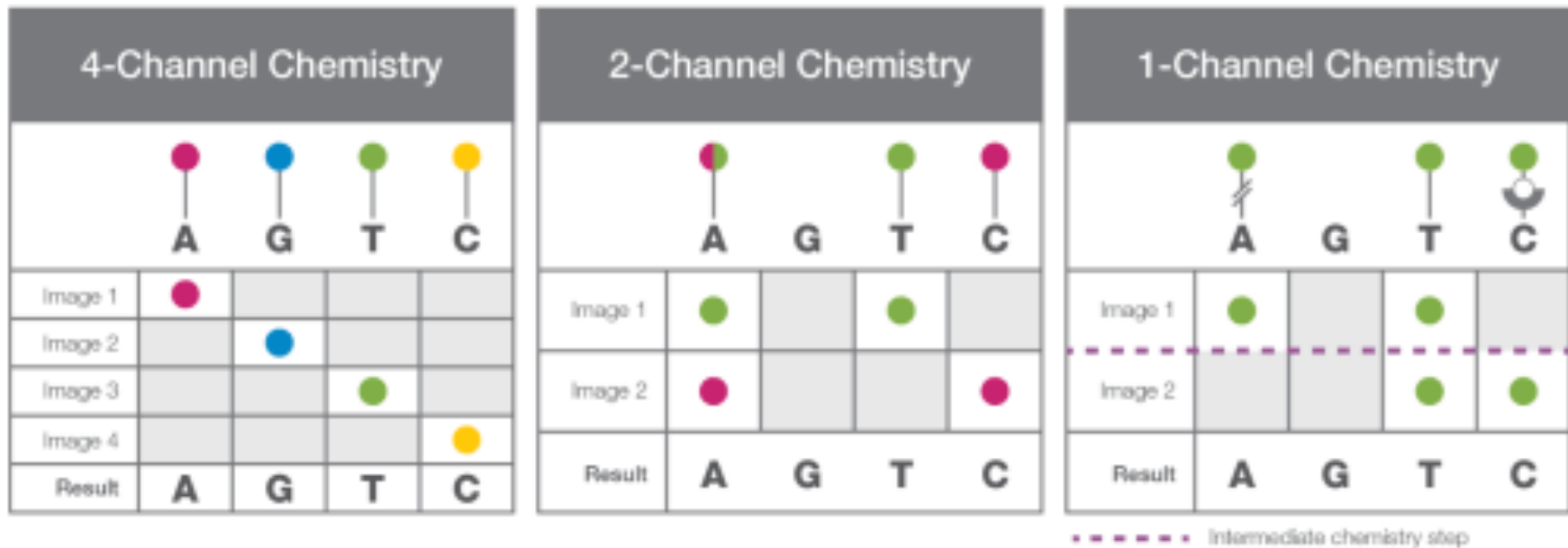
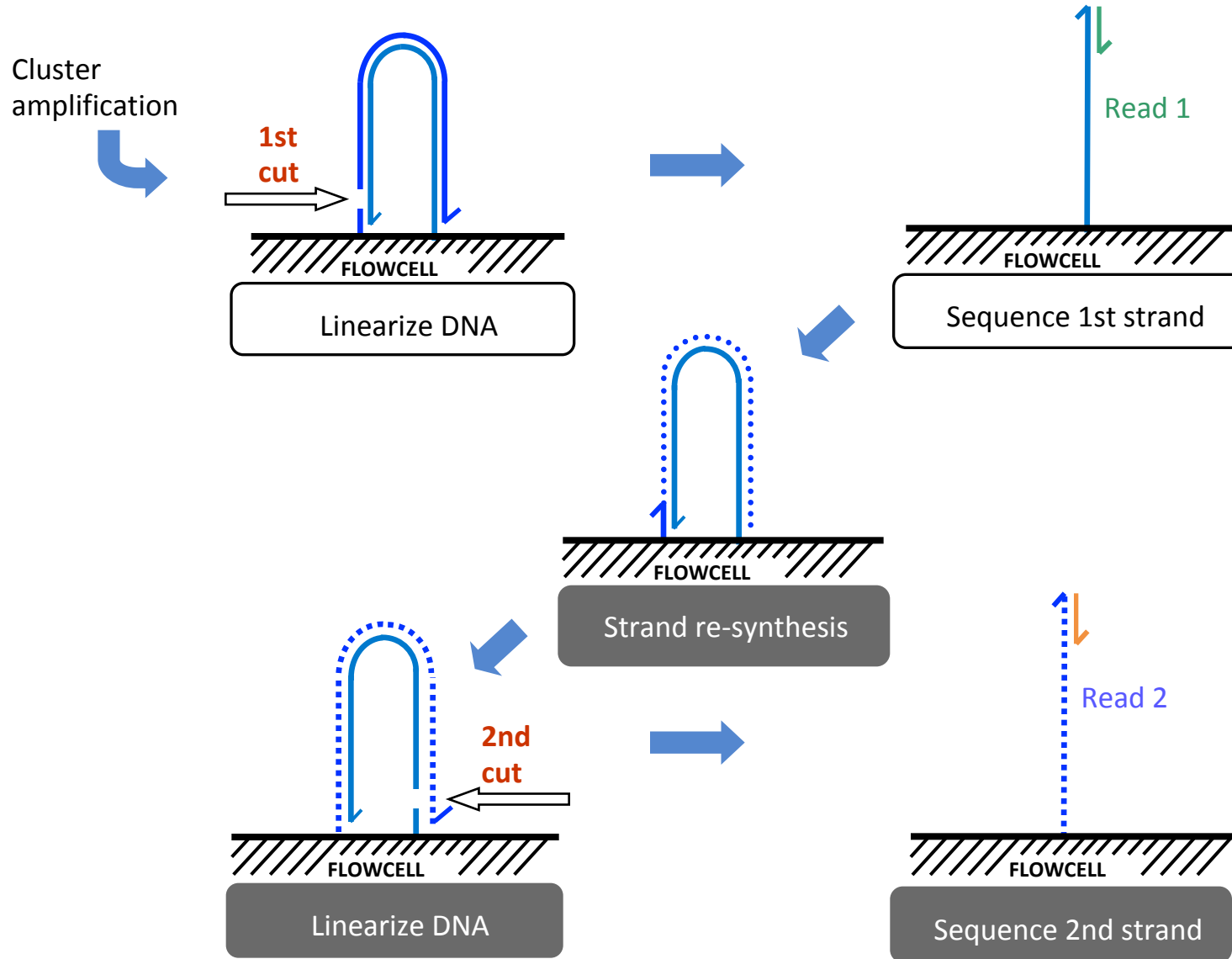
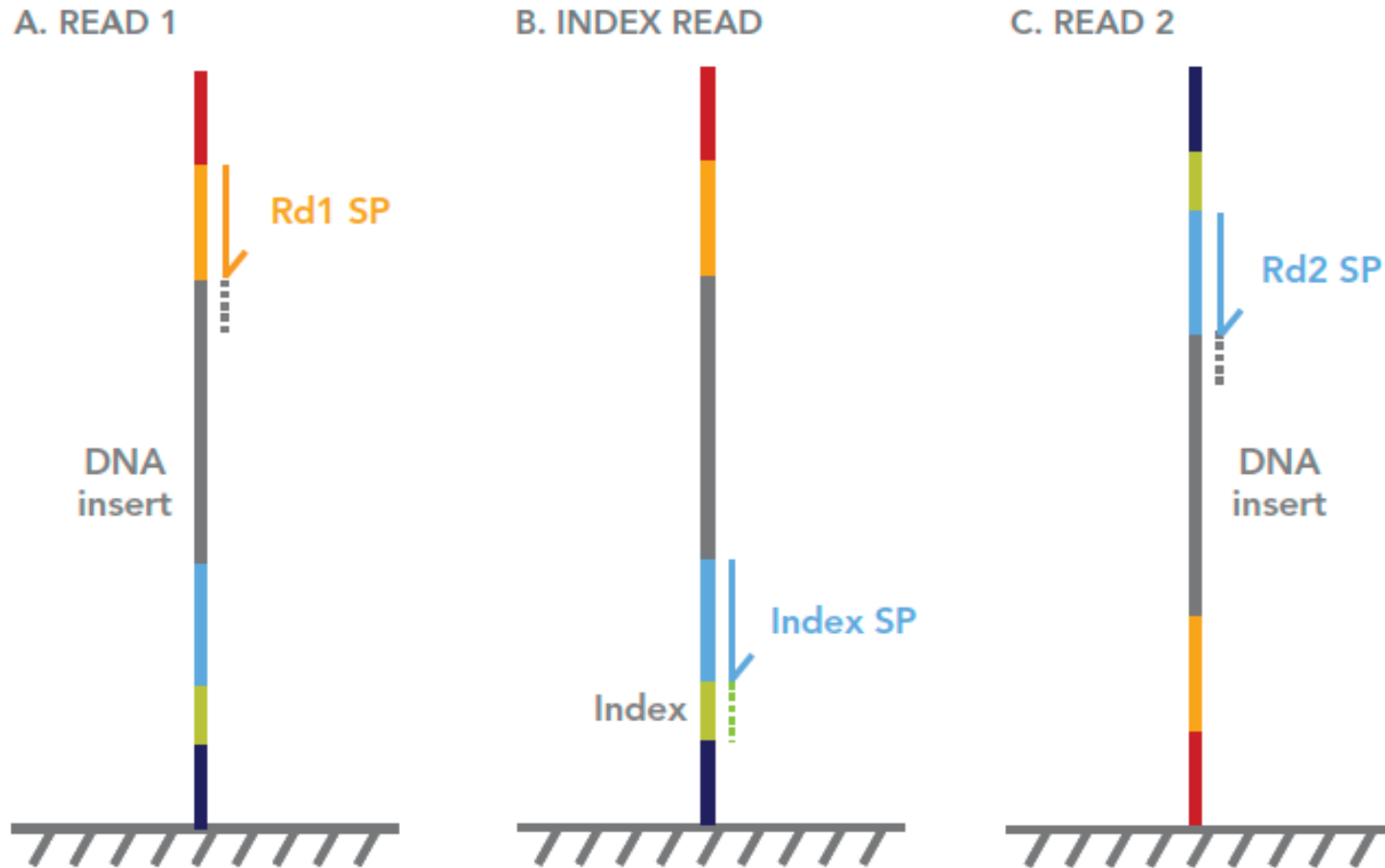


Figure 2: Four-, Two-, and One-Channel Chemistry—Four-channel chemistry uses a mixture of nucleotides labeled with four different fluorescent dyes. Two-channel chemistry uses two different fluorescent dyes, and one-channel chemistry uses only one dye. The images are processed by image analysis software to determine nucleotide identity.

Paired-End Sequencing allows for two looks at a sequence

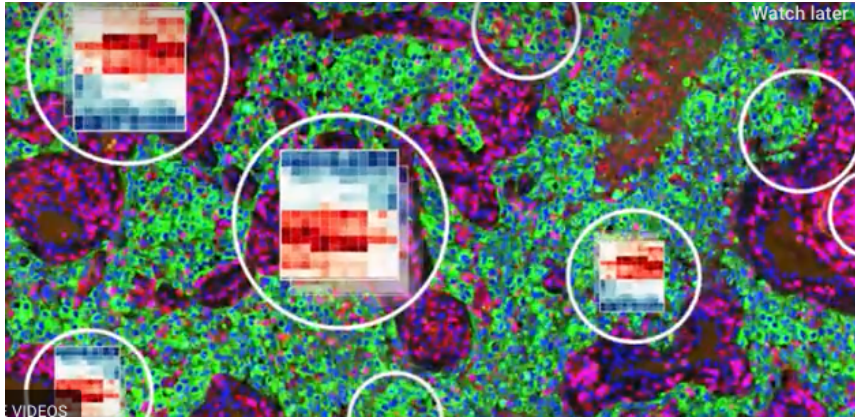


Indexed sequencing method is now standard for single and paired reads

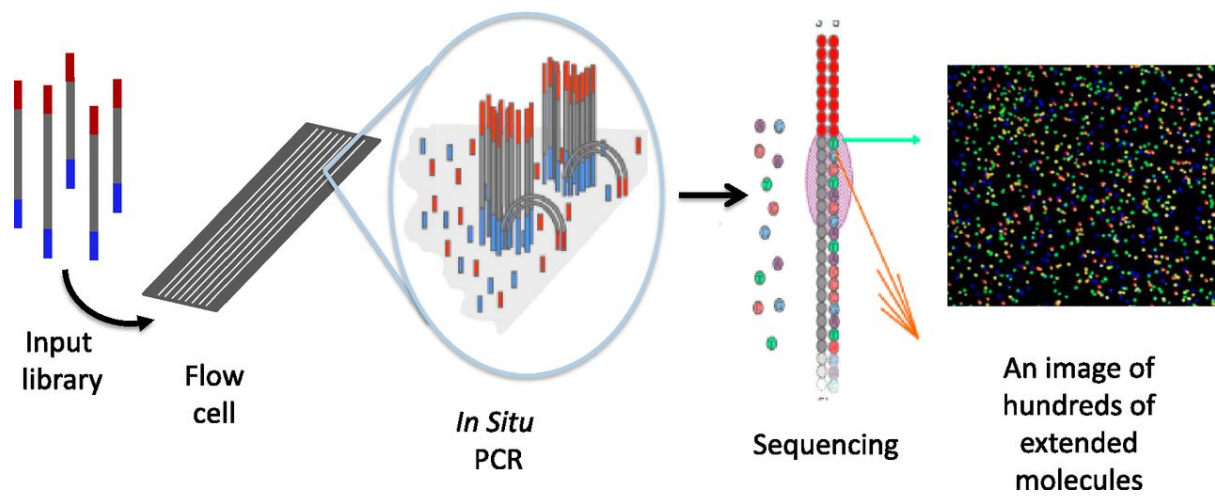
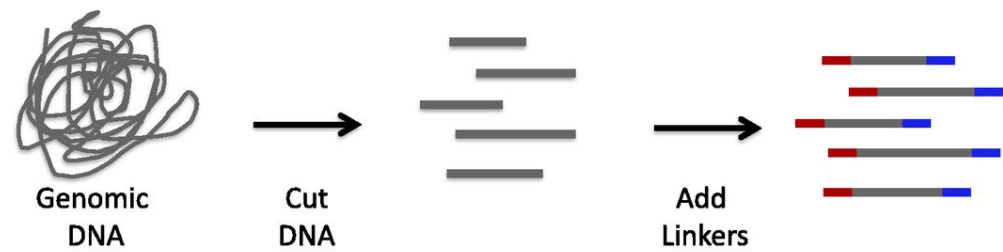


THE TOPIC: SPATIAL GENOMICS

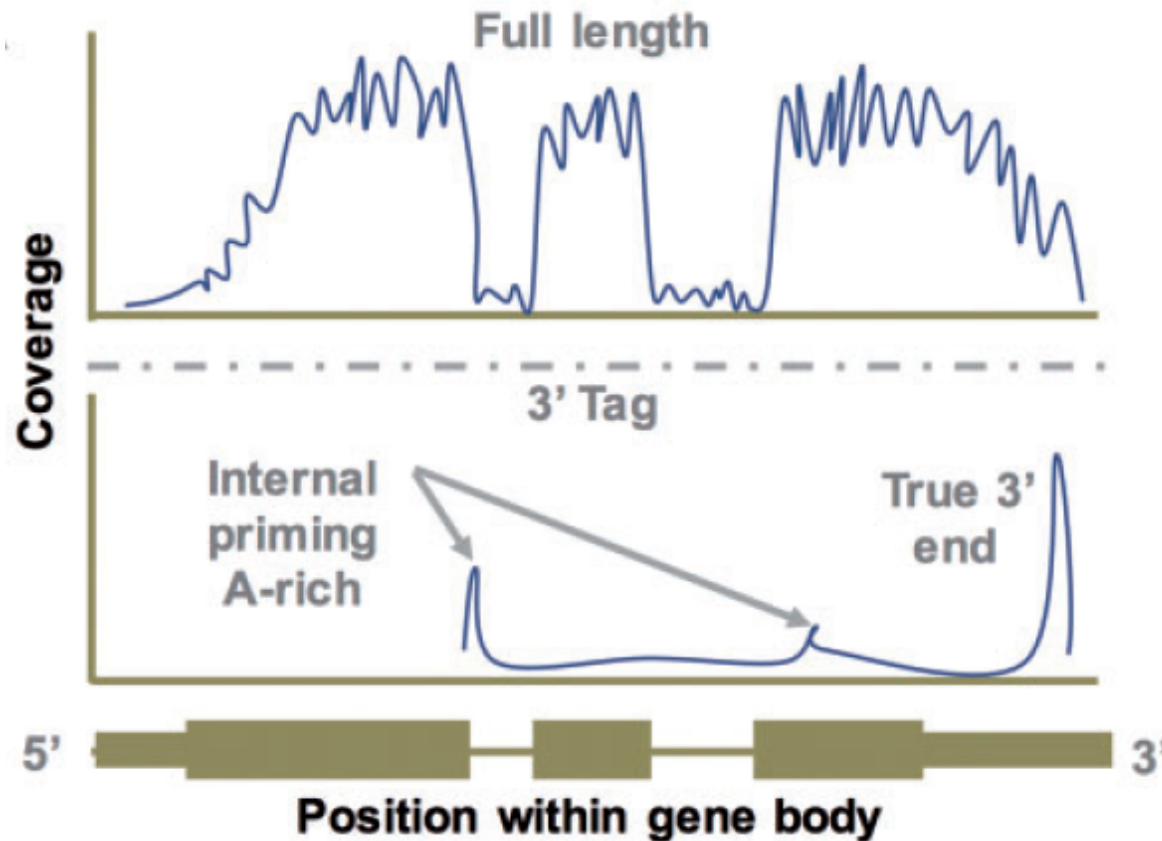
KEEPING SPATIAL INFORMATION – SPATIAL GENOMICS







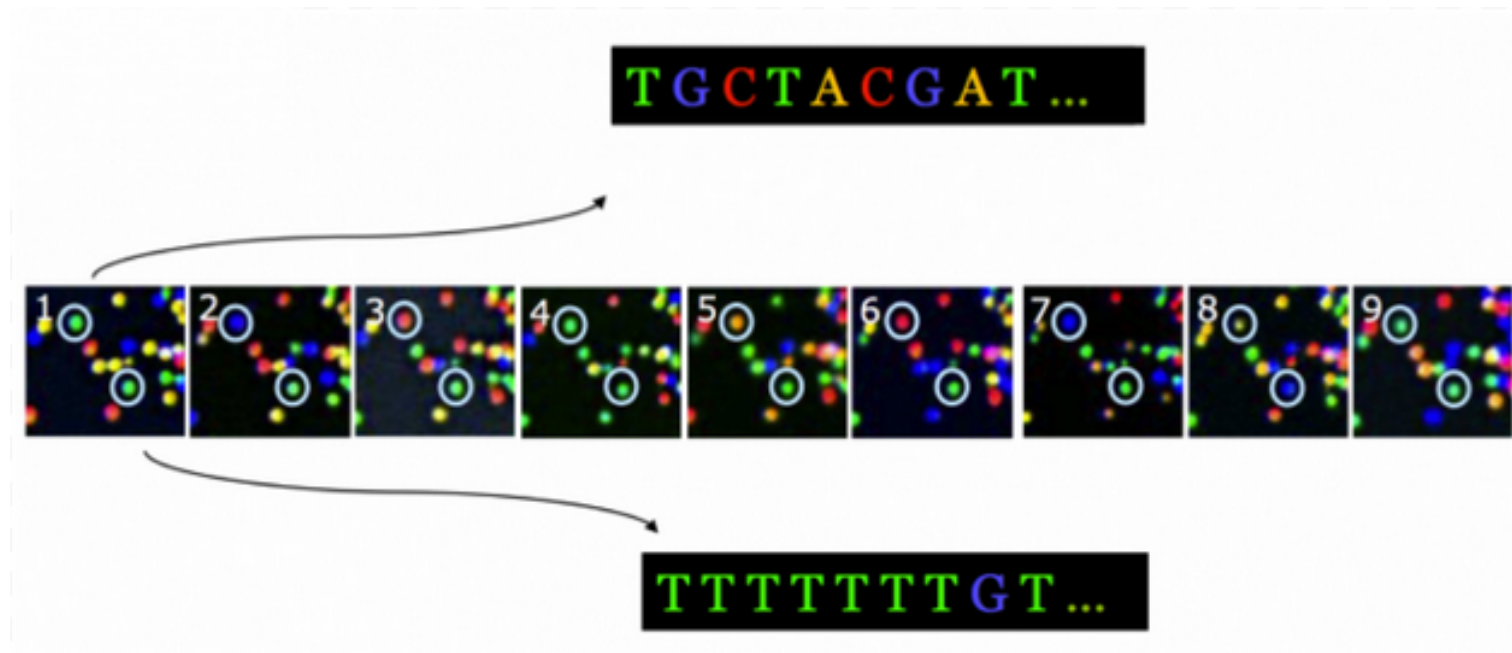
Single cell transcriptomics: full gene body coverage vs 3' gene body coverage



SMARTseq/SMARTseq v2
C1 Fluidigm
NO UMI

Cell-seq, MARS-seq,
Dropseq, InDrop, ddSeq,
Chromium

Genomic Diversity and Index Diversity Matter!!!



- Algorithm that calls the base on a cluster is set up using the 1st 25 bases of the run. These bases are expected to be color-balanced for A,G,C,T; if not the base call is incorrect
Index run also needs to be balanced

Sequencing Coverage

- Coverage (or depth):
number of reads that are likely to be aligned at a given reference position

$$\text{Coverage} = \frac{\text{read length} \times \text{number of reads}}{\text{genome length (haploid)}}$$

- Coverage for Methylation sequencing of the human genome
Using one lane of a HiSeq 2500 v4 chemistry

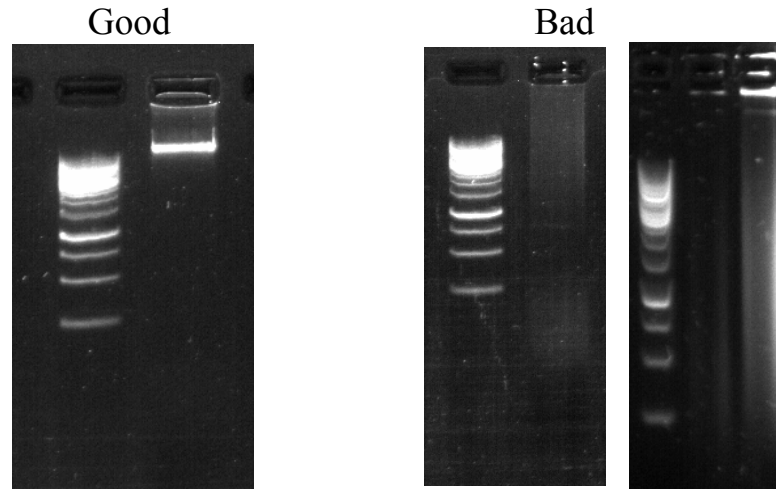
$$\frac{(100\text{bp}) \times (280 \times 10^6)}{3 \times 10^9} = 9.3\text{X coverage}$$

- Illumina's website has a coverage calculator app

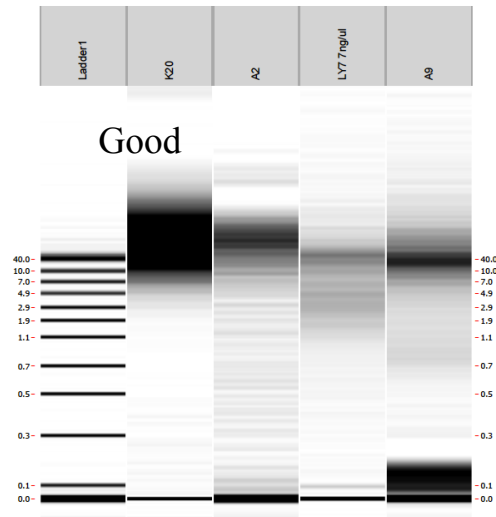
http://support.illumina.com/downloadssequencing_coverage_calculator.html

Genomic DNA sample QC: ~40kb, no RNA contamination

Using an agarose gel



Using the Perkin Elmer Labchip GX



Enhanced Reduced Representation of Bisulfate Conversion (ERRBS)

1. MspI digestion



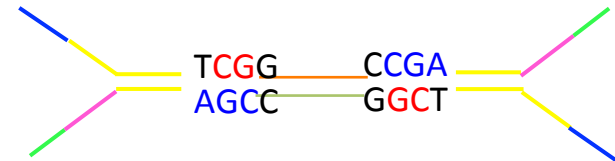
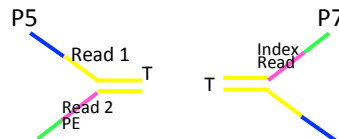
2. End Repair, generates a de novo C, unmethylated and must be removed for data analysis



3. A-tail (for complementing the Illumina Adapter)



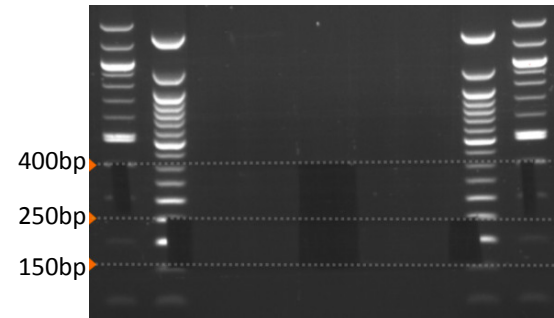
4. Adapter ligation (methylated version)



Enhanced Reduced Representation of Bisulfate Conversion (ERRBS)

5. Size selection (150-250bp and 250-400bp)
(DNA size 30-280bp)

Two fractions are carried over for the rest of the preparation



6. Bisulfite conversion, all C are modified to U
DNA is no longer complementary ie single stranded

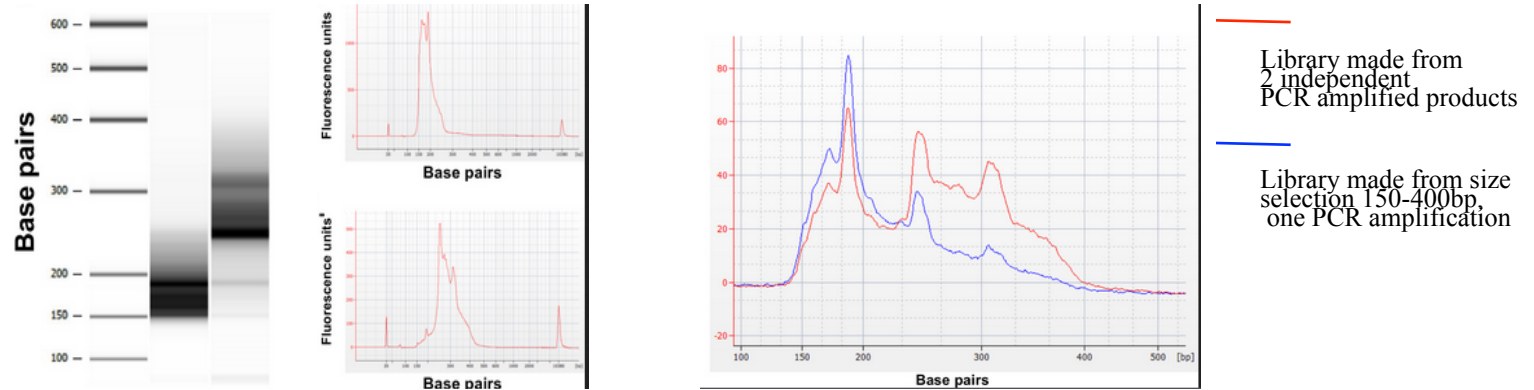


7. PCR amplification

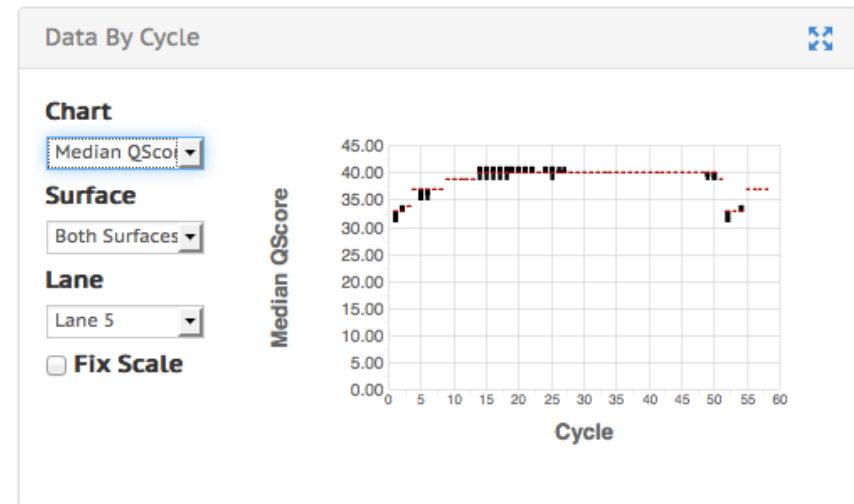
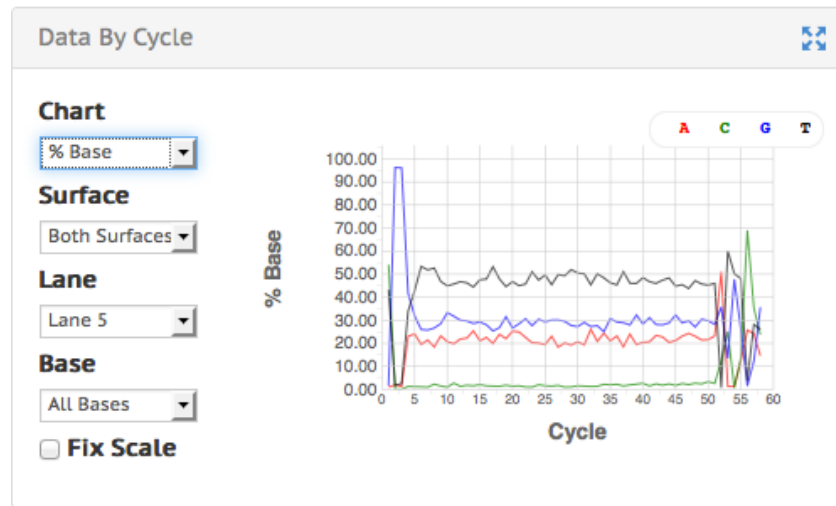


Enhanced Reduced Representation of Bisulfate Conversion (ERRBS)

Library QC



Sequencing QC



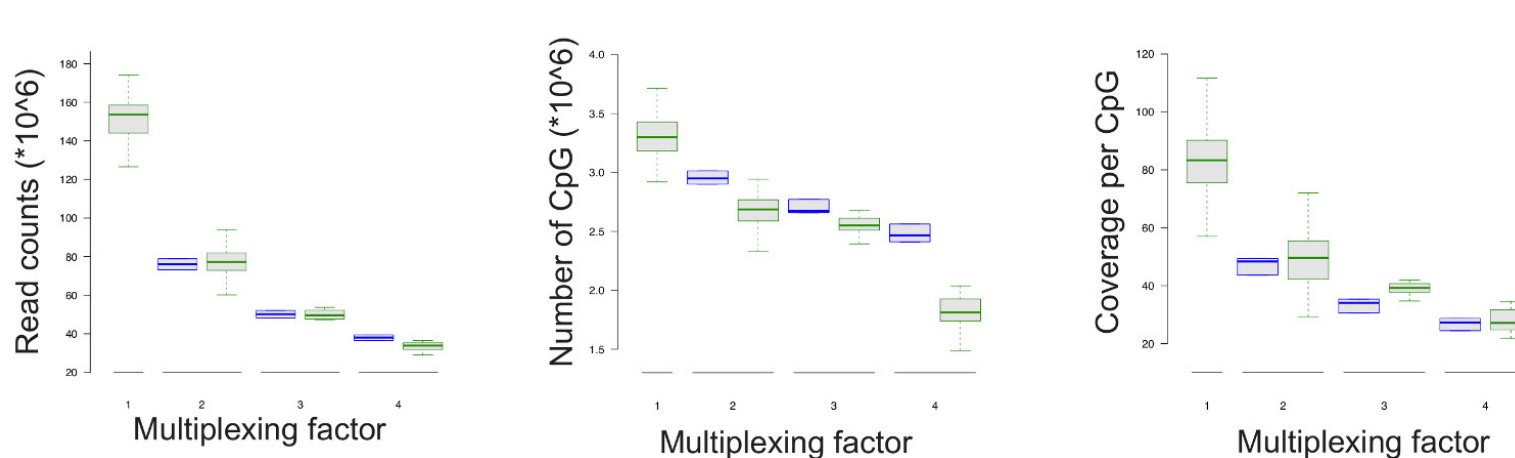
- Akalin AI, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtkova I, Milne TA, Huang Y, Biswas D, Hess JL, Allis CD, Roeder RG, Valk PJ, Löwenberg B, Delwel R, Fernandez HF, Paietta E, Tallman MS, Schroth GP, Mason CE, Melnick A, Figueroa ME. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. [PLoS Genet.](https://doi.org/10.1371/journal.pgen.1002781) 2012;8(6):e1002781. doi: 10.1371/journal.pgen.1002781. Epub 2012 Jun 21.
- Garrett-Bakelman, F.E., Sheridan, C.K., Kacmarczyk, T.J., Ishii, J., Betel, D., Alonso, A., Mason, C.E., Figueroa, M.E., Melnick, A.M. Enhanced Reduced Representation Bisulfite Sequencing for Assessment of DNA Methylation at Base Pair Resolution. *J. Vis. Exp.* (96), e52246, doi:10.3791/52246 (2015).

Enhanced Reduced Representation of Bisulfate Conversion(ERRBS)

Sequencing depth

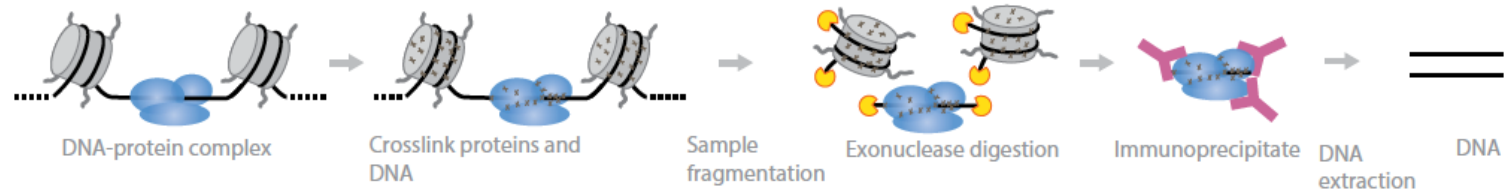
Sequencing recipe	Number of CpG covered (>10X)	Mean CpG coverage (>10X)
SR50	2,979,084	96.62
PE50	5,930,668	88.53
SR100	4,889,616	93.15
PE100	6,607,470	85.48

How does multiplexing affect ERRBS data?



ChIPseq

Chromatin Immunoprecipitation



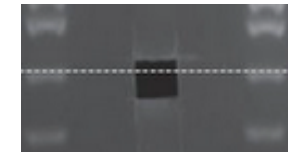
1. End repair, A-tailing



2. Adapter ligation



3. Size selection of DNA from 250-350bp
(original DNA ~130-230bp)



4. PCR amplification



ChIPseq

Sample QC

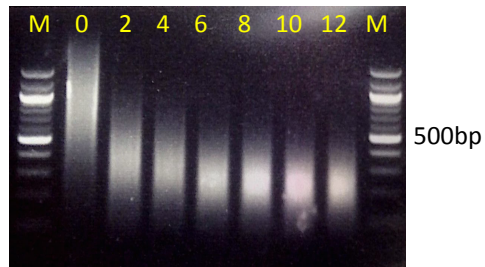
QC1: quantitation

a minimum of 11ng of ChIP DNA at $\leq 1\text{ng/ul}$, determined by Qubit

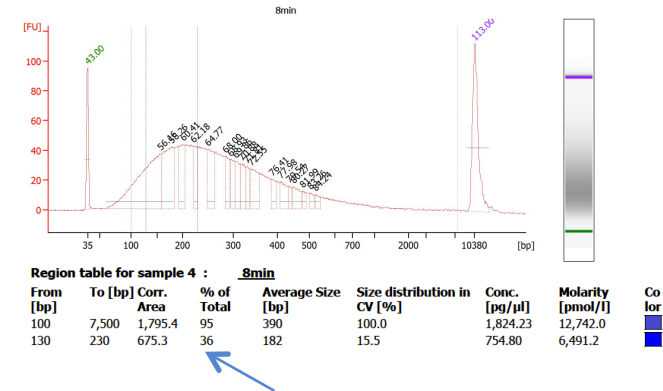
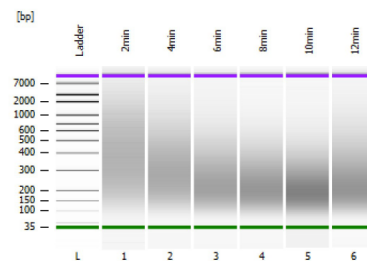
QC2: qualitative measurement

10% of total material must be in the size range of the DNA fraction required for library preparation (130-230bp)

1.5% agarose gel



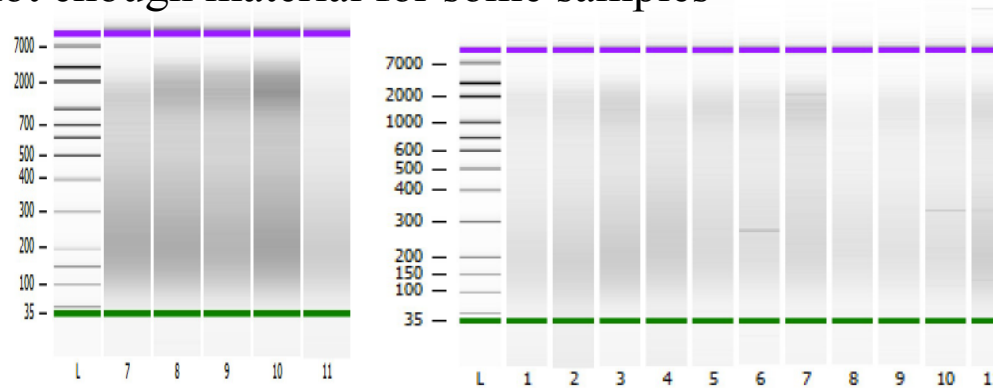
BioAnalyzer High Sensitivity



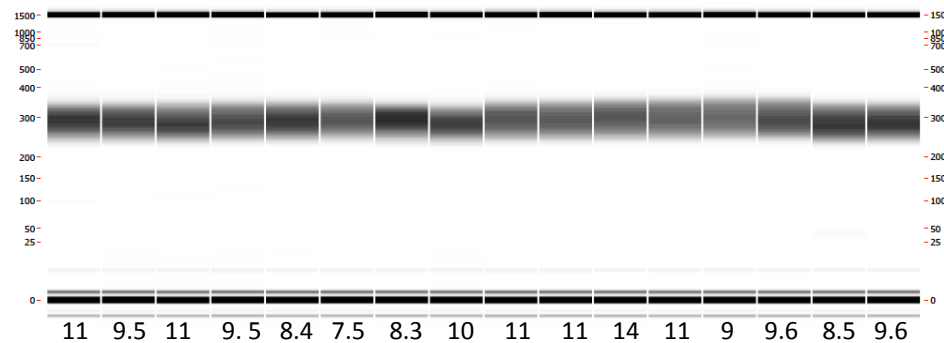
Typical ChIPseq project

QC2 BioA all samples >-10% within range

QC1 not enough material for some samples



Library QC



% duplication

Core: will send an email informing libraries are ready to be tested by qPCR

Fast QC report will give % duplication, which is one parameter to follow

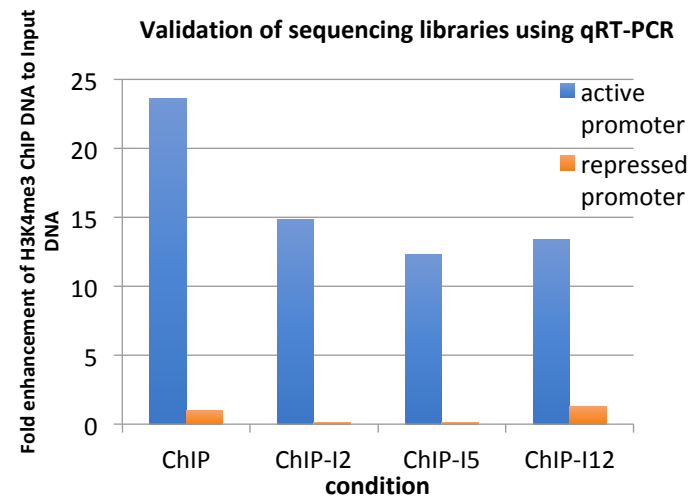
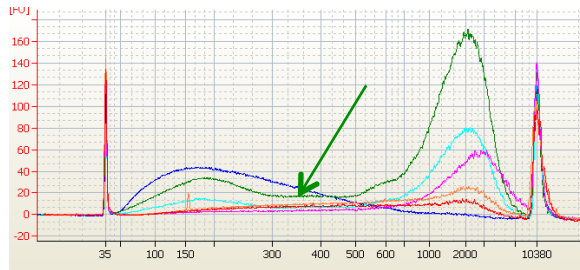
ChIPseq

Sequencing depth:

depends on the antibody, number of binding sites, cell type, quality of the IP

Multiplexing of ChIP-Seq Samples in an Optimized Experimental Condition Has Minimal Impact on Peak Detection Thadeous J. Kacmarczyk, Caitlin Bourque, Xihui Zhang, Yanwen Jiang, Yariv Houvras, Alicia Alonso, and Doron Betel [PLoS One](https://doi.org/10.1371/journal.pone.0129350). 2015 Jun 11;10(6):e0129350. doi: 10.1371/journal.pone.0129350. eCollection 2015.

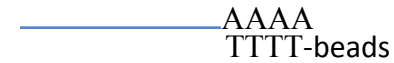
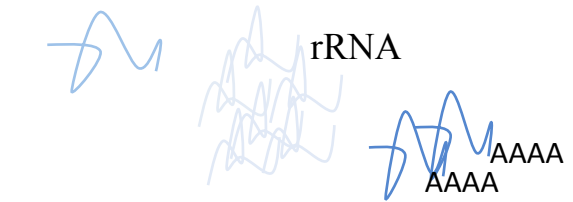
QC2



Multiple libraries were made and sequenced on SR50 at different depths, from 181M to 27M

TruSeq mRNA prep

1. rRNA depletion: oligo dT hybridization, and purification
2. mRNA fragmentation: chemical cleavage (~150bp)
3. Random priming to obtain cDNA strand and dscDNA
4. End repair, A-tailing
5. Adapter ligation
6. PCR amplification



TruSeq stranded mRNA prep

1. rRNA depletion: oligo dT hybridization, and purification

2. mRNA fragmentation: chemical cleavage (~150bp)

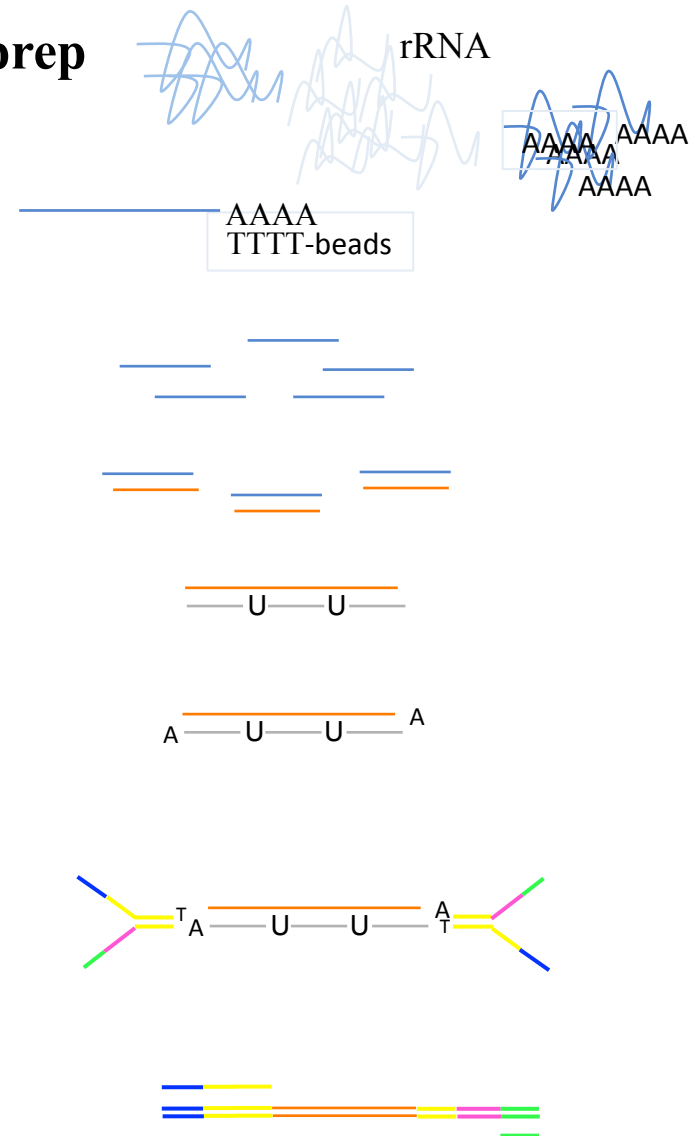
3. Random priming to obtain cDNA strand

4. dUTP incorporation on the dscDNA

5. A-tailing

6. Adapter ligation

**7. PCR amplification,
removal of dUTP containing strand**



TruSeq Total stranded RNA prep

1. **rRNA depletion: rRNA hybridization, and purification**

2. mRNA fragmentation: chemical cleavage (~150bp)

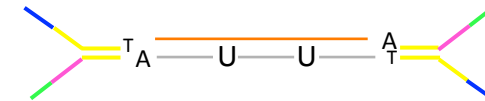
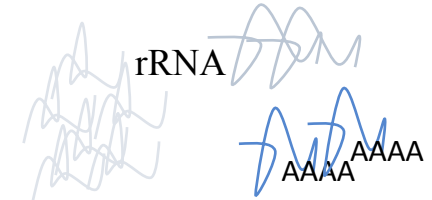
3. Random priming to obtain cDNA strand

4. **dUTP incorporation on the dscDNA**

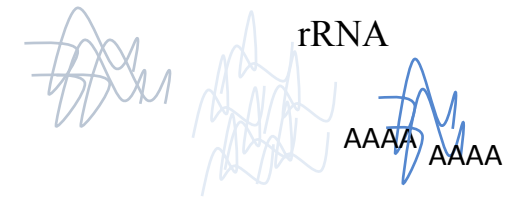
5. A-tailing

6. Adapter ligation

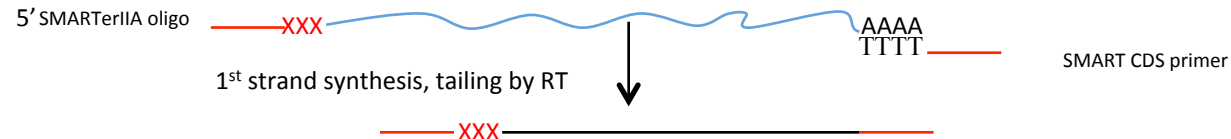
7. **PCR amplification, removal of dUTP containing strand**



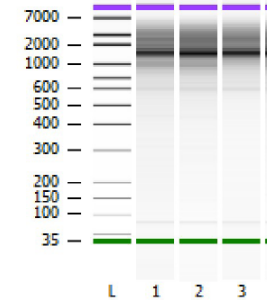
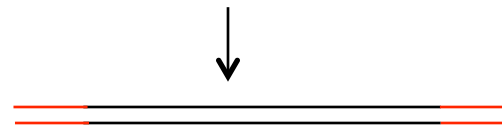
UltraLowInput RNaseq (SMART technology) (Switching Mechanism At 5' End of RNA Transcript)



1. SMARTer Reverse transcriptase



2. PCR amplification from SMARTer primers (Full length dscDNA)



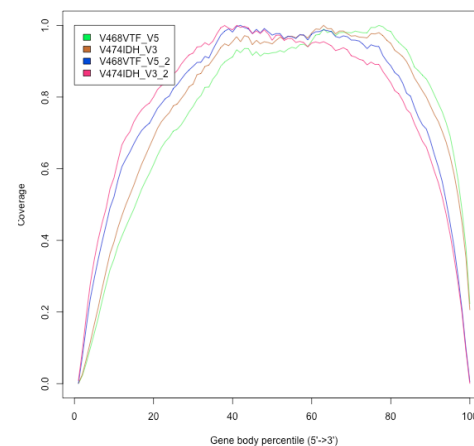
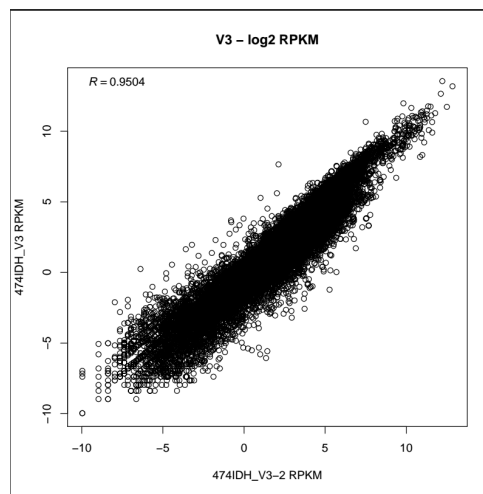
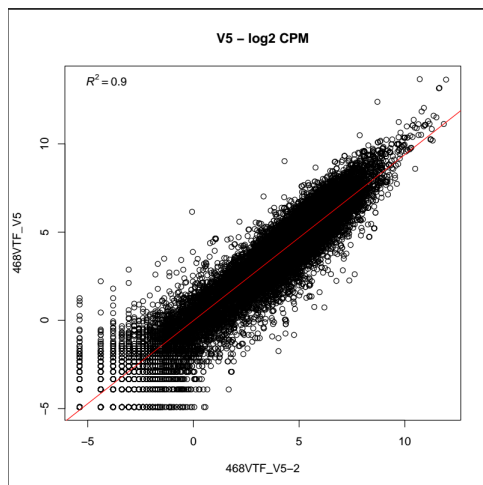
3. Sonication to ~200bp

4. End repair, A-tailing, ligation of TruSeq adapters

5. PCR amplification

Comparison between data generated by RNAseq and Ultra Low Input RNAseq

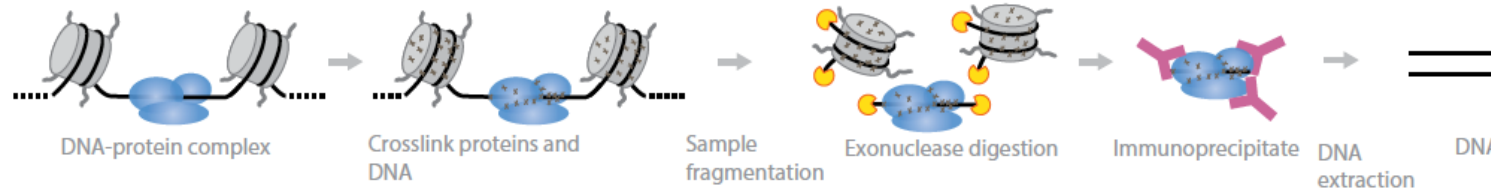
	468VTF_V5-2 (RNAseq)	468VTF-5 (ultralow)	474IDH_V3-2 (RNAseq)	474IDH_V3 (ultralow)
RIN	8.3	8.3	8.3	8.3
% duplication	48.22	45.95	47.99	44.53
% mapped reads	98.90	97.85	98.80	98.02
RefSeq IDs with > 10 counts	18602	17997	17868	17168
% rRNA in bam(RSeQC)	2.71	7.42	3.3	7.69



Xihui, Yushan, Piali

Quality Control of ChIP DNA

Chromatin Immunoprecipitation



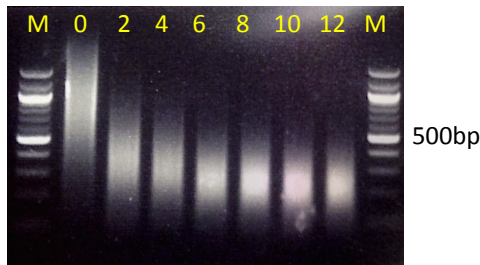
QC1: quantitation

a minimum of 11ng of ChIP DNA at $\leq 1\text{ng/ul}$, determined by Qubit

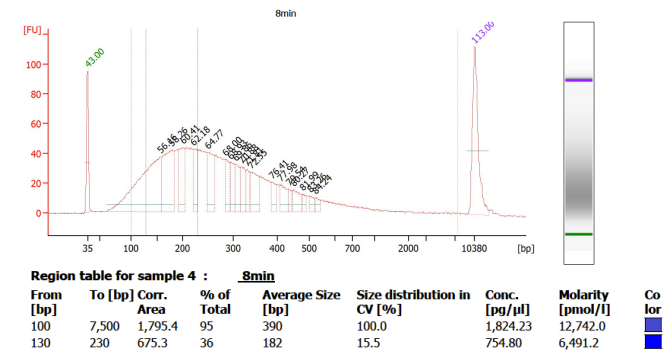
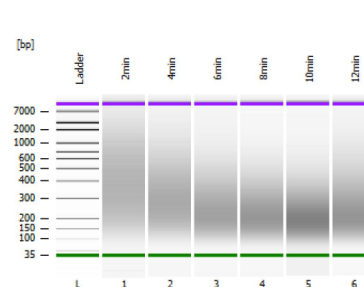
QC2: qualitative measurement

10% of total material must be in the size range of the DNA fraction required for library preparation (130-230bp)

1.5% agarose gel



BioAnalyzer High Sensitivity





Overview

The **Epigenomics Core Facility** of [Weill Cornell Medical College](#) provides an array of epigenomics and bioinformatics research resources and services that include:

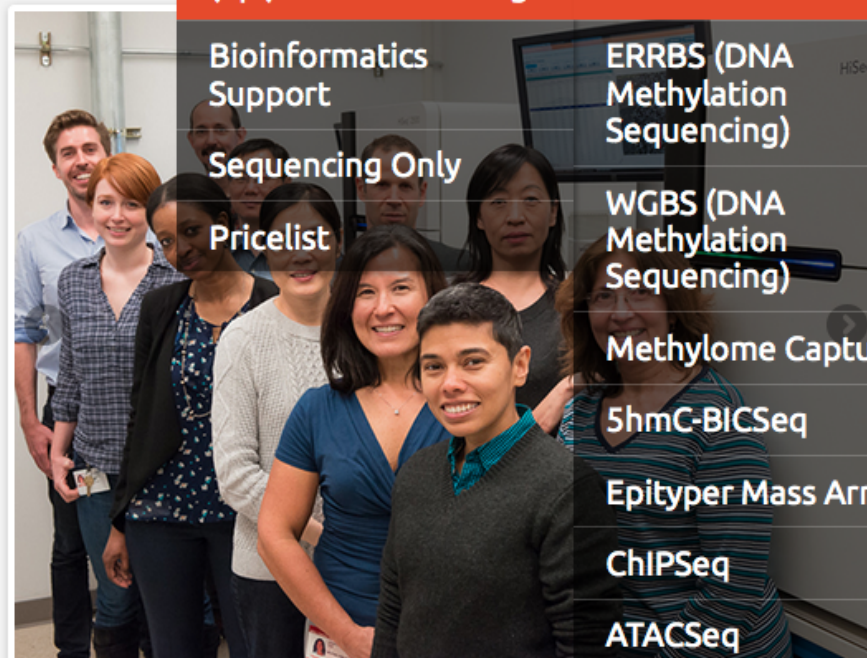
- **DNA methylation profiling** [\[More\]](#)
- **Protein-nucleic acid association** [\[ChIP-Seq\]](#)
- **Genome sequencing** [\[More\]](#)
- **Exome capture** [\[More\]](#)
- **RNA-seq** [\[More\]](#)
- **Bioinformatics analysis** [\[More\]](#)

Core resources and services include sample preparation services and data generation on the Illumina HiSeq 2500, MiSeq platforms and Sequenom MassArray platform.

[Getting Started](#)[iLab](#)[Pubshare](#)[FAQ](#)[\\$ Pricelist](#)

(Epi)Genetic Profiling

Overview

[Bioinformatics Support](#)[Sequencing Only](#)[Pricelist](#)[ERRBS \(DNA Methylation Sequencing\)](#)[WGBS \(DNA Methylation Sequencing\)](#)[Methylome Capture](#)[5hmC-BICSeq](#)[Epityper Mass Array](#)[ChIPSeq](#)[ATACSeq](#)[Exome Capture](#)[DNaseq](#)[Targeted Resequencing](#)[RNASeq](#)

Illumina SBS Technology

Reversible Terminator Chemistry Foundation

