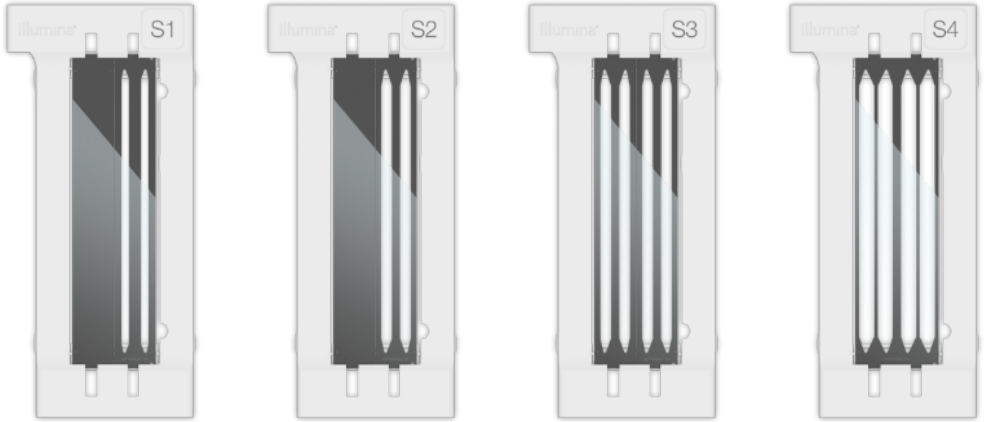
Sequencing Power for Every Scale

Systems for every application. For every lab.

illumina®

NovaSeq System Configurations

Max Output / Flow Cell: **0.5 Tb** **1 Tb** **2 Tb** **3 Tb**



NovaSeq 5000
$850K USD

NovaSeq 6000
$985K USD

NovaSeq 5000 Flow Cells

NovaSeq 6000 Flow Cells

NovaSeq 5000 **NovaSeq 6000**

NovaSeq Series

Any Genome. Any Method. Any Scale.

NovaSeq 5000		NovaSeq 6000
	PE 150 | Q30 ≥ 75%	
167 – 2000 Gb	OUTPUT	167 – 6000 Gb
1.6 – 6.6B	READ NUMBER	1.6 – 20B
Fast	RUN TIME	Fastest (40 Hr. for 2T Run)
2 Types	Flow Cells	4 Types

Output range shown based on currently unreleased flow cells

Scalable Throughput

Complete studies faster and more economically

Run times: <1 to ~2.5d based on system, FC and read length

Configure output to match your application and study size

Single flow cell output (1 or 2 can run simultaneously)

Flow Cell Type	NovaSeq 5000	NovaSeq 6000	Reads per Flow Cell	Output (Gb) per Flow Cell		
				100 cycles	200 cycles	300 cycles
S4*		✓	10B			3000
S3*		✓	6.6B			2000
S2	✓	✓	3.3 B	333	666	1000
S1*	✓	✓	1.6 B	167	333	500

*S1, S2 and S4 flow cells not currently released

illumina®

Highly Flexible

Configurable to support the broadest range of applications

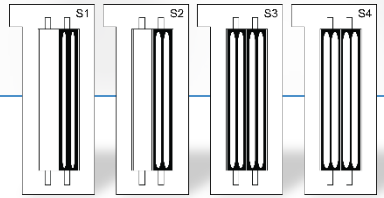
Open to all methods and species

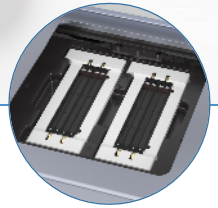
4* flow cells

Powered by 4 types of flow cells

1 | 2

Run 1 or 2 flow cells – mix and match different types





illumina®

*Based on flow cells not currently released

Streamlined Operation

Increase lab efficiency with a simplified workflow

Cartridge based reagents reduce hands on time and prevent misloading

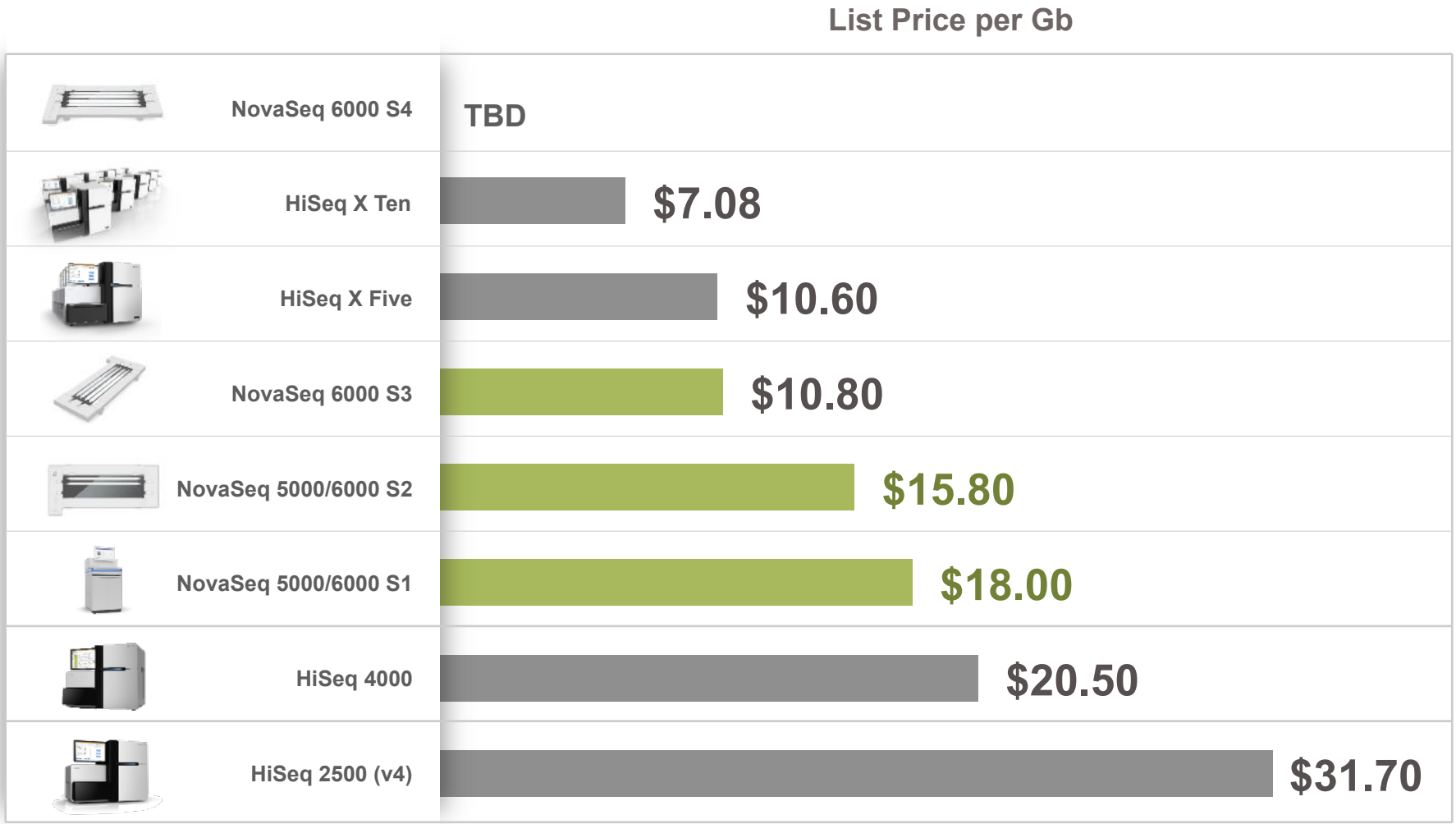
RFID encoded consumables provide traceability and ensure compatibility

Onboard cluster gen reduces hands on time and run variability

illumina®

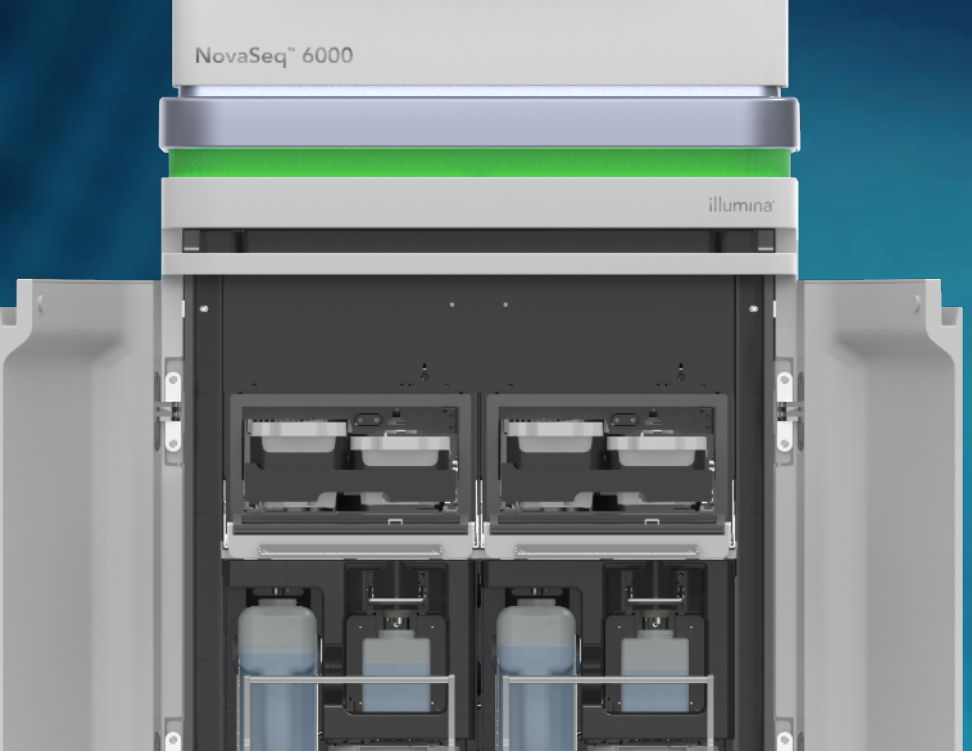
NovaSeq Series

Compelling price per data point enables highly-powered studies



HiSeq 2500 based on 250 cycle kit, all others based on 300 cycle kit

illumina®



Key Technology Enablers

Redesigned from the ground up

High Density Flow Cell

Higher density flow cell format (624nm pitch)

New surface chemistry

Increased signal:noise via smaller, brighter clusters

New superior imaging

4x faster scanning

Diffraction-limited performance

Data management

8x increase in primary analysis speed

Data footprint reduced by 25%

Reformulated chemistry

Reengineered nucleotide dyes

Optimization of 8 different sub-formulations

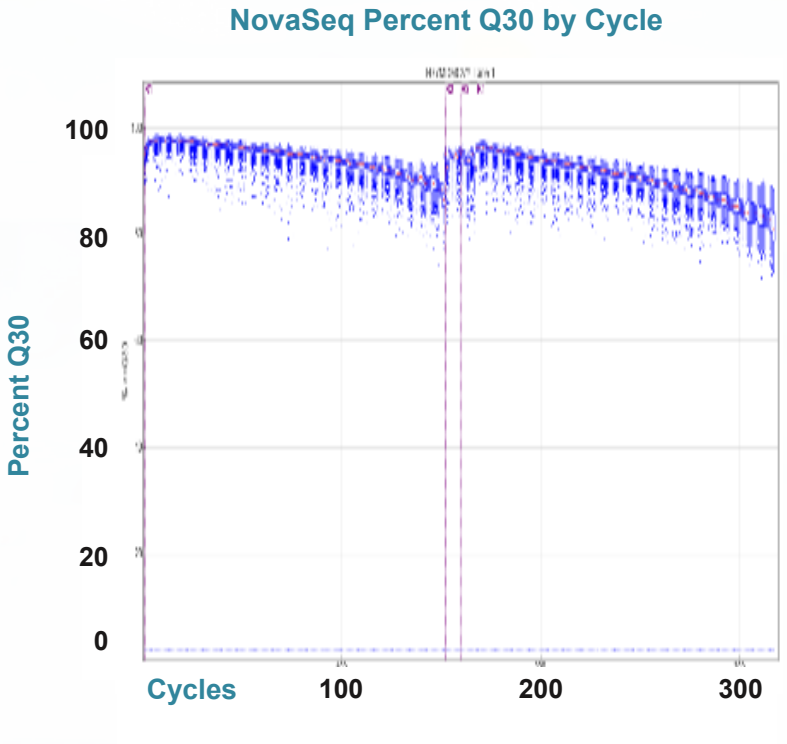
illumina®

NovaSeq Performance and Data Quality

Initial Data Quality

- **Data quality as good as HiSeq with initial release – significant opportunity for further improvements**
  - Major R&D focus on further optimizing 2-channel chemistry with patterned flow cells
- **High data quality enabled by superior optics and reformulated chemistry**
  - Diffraction-limited performance optics
  - New surface chemistry, dye-sets and enzymes

NovaSeq Performance



*NovaSeq Prototype Instrument running S2 flow cell
**NovaSeq Q30 based on calibrated, but non-final Q-table

Platform	Read Length	Output	Percent Q30**
NovaSeq*	2 x 151	>2000G	**92.3**
NovaSeq*	2 x 151	>2000G	**88.5**
HiSeq X	2 x 151	>2000G	**82.8**
HiSeq v4	2 x 126	~1000G	**85.7**

illumina®

Human Genome Performance on NovaSeq

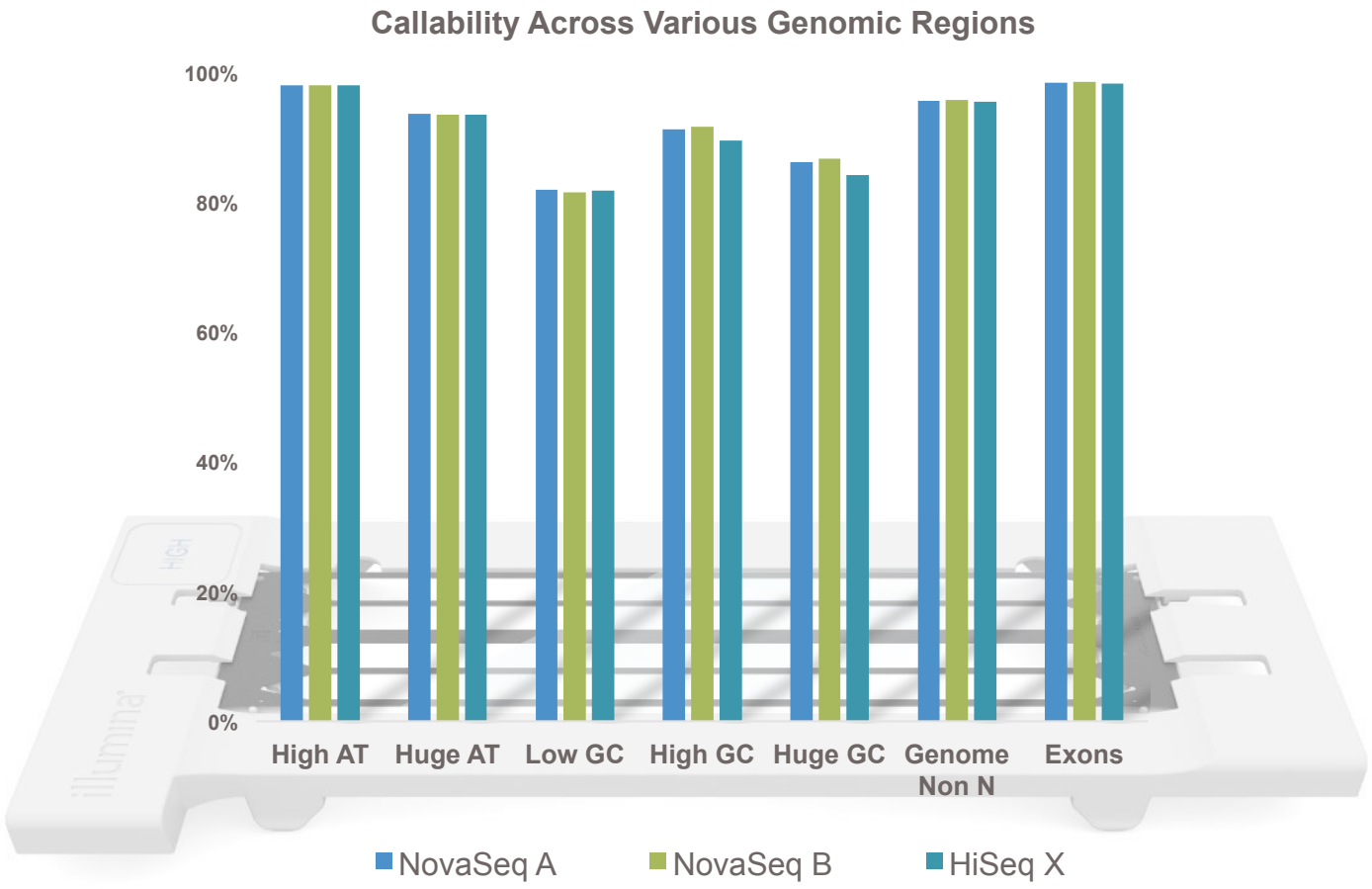
Genome build quality highly concordant with HiSeq

	NovaSeq (n4)	**HiSeq X** (n2)	**HiSeq v4** (n2)	**NextSeq** (n2)
Genome Coverage (x)	**30.6**	**30.5**	**29.8**	**30.1**
Autosome Coverage	**95%**	**95%**	**91%**	**94%**
Autosome Callability	**95%**	**95%**	**93%**	**93%**
Autosome Exon Callability	**98%**	**98%**	**91%**	**95%**
SNP Precision	**100%**	**100%**	**100%**	**100%**
SNP Recall	**97%**	**97%**	**96%**	**96%**
Indel Precision	**97%**	**98%**	**97%**	**96%**
Indel Recall	**95%**	**95%**	**88%**	**88%**

NovaSeq Prototype Instruments running S2 flow cell

Human Genome Performance on NovaSeq

Genome build quality meets or exceeds HiSeq



For Research Use Only. Not for use in diagnostic procedures.

NovaSeq Configurations and Key Methods

Configurable to support broad range of methods and accommodate rapid turn around

Flow Cell Type	S1	S2	S3	S4
Expression profiling	✓	✓		
Whole transcriptome analysis	✓	✓		
Exome	✓	✓	✓	
Low pass WGS	✓	✓	✓	
Liquid biopsy	✓	✓	✓	✓
WGS	✓	✓	✓	✓
T/N profiling	✓	✓	✓	✓

NovaSeq 5000 & 6000

NovaSeq 6000

illumina®

BaseSpace®

Workflow | Storage | Analysis | Sharing

>4,000
Instruments

>30,000
Users

>240,000
Runs

>60
Apps

Illumina Core Applications

Third Party Applications

BaseSpace Labs Apps

Illumina Core


16S Metagenomics


Amplicon DS


BWA & Isaac Enrichment

BWA & Isaac WGS


Cufflinks Assembly & DE


MethylSeq


RNA Express


Small RNA

TopHat Alignment


TruSeq Amplicon


TruSeq Long-Read Assembly


TruSeq Phasing Analysis


TruSeq Targeted RNA


Tumor Normal

VariantStudio

BaseSpace Labs


FASTQ Toolkit

FASTQC


Kraken Metagenomics


NextBio Annotates RNA-Seq


NextBio Transporter

PicardSpace

Prokka Genome Annotation

SRA Import

SRA Submission

SRST2

Variant Calling Assessment Tool


Velvet *de novo* Assembly

Third-Party

DNAStar


DeepChek-HBV, HCV, HIV

EDGC Annotator

Elastic Genome Browser


GeneTalk Variant Analyzer


GENIUS Metagenomics: Know Now

Genomatix Pathway System (GePS)


Genome Profiler


iPathwayGuide


LoFreq Rare Variant Caller


Melanoma Profiler


MetaPhlAn

miRNAS Analysis

MyFLq

Novoalign Generic DNA pipeline

OncoMD

PathSEQ Virome

PEDANT Sequence-Analyzer

Phy-Mer

Protein Expression Assembler

Protein Expression Extractor

Protein Expression Workflow

RNA-Seq Translator

SPAdes Genome Assembler

SWATHAtlas Ion Library Generator

The Broad's IGV

Variant Interpreter

Tiers and Introductory Pricing

	Basic	Professional	Enterprise
Available	Now!	February 10th, 2016	March 30th, 2016
Features	• Run set up, monitoring, data streaming, storage & sharing, push-button data analytics with 70+ Apps • Test Drive Sequencing Hub • Maximum 1TB storage (limit imposed from Feb 10th, 2016) • 250 iCredits for Apps (available March 30th, 2016) • Additional Storage and compute cannot be purchased	• All Basic features • Purchase additional storage and compute • Multiple user access to pooled resources • 8 hours of professional services support	• Private domain & single sign-on • Access Control • 24 hours of professional services support • Audit Trail & PHI Security/ Privacy compliance
Target Markets	• All customers	• Routine testing labs • Small research core labs	• Routine Testing Labs needing HIPAA • Large research core lab
Cost / Year	**Free**	**$4,995**	**1% of connected instrument value ($30K minimum)**

Prices in USD. Regional prices may vary

Cloud Storage and Compute Introductory Pricing

	Professional	Enterprise
Available	February 10th, 2016	March 30th, 2016
Features	• 1TB for free (one time) • Pre-paid yearly storage subscriptions • Available as 1TB, 5TB, 10TB, 50TB • Purchased via quote • No cancellations or refunds after 30 days	• Compute hours monitored via iCredits • 250 iCredits free (one time) • Pre-purchase • Purchased via quote • No cancellations or refund after 30 days
Cost	$360/TB/yr**	$1 / iCredit&

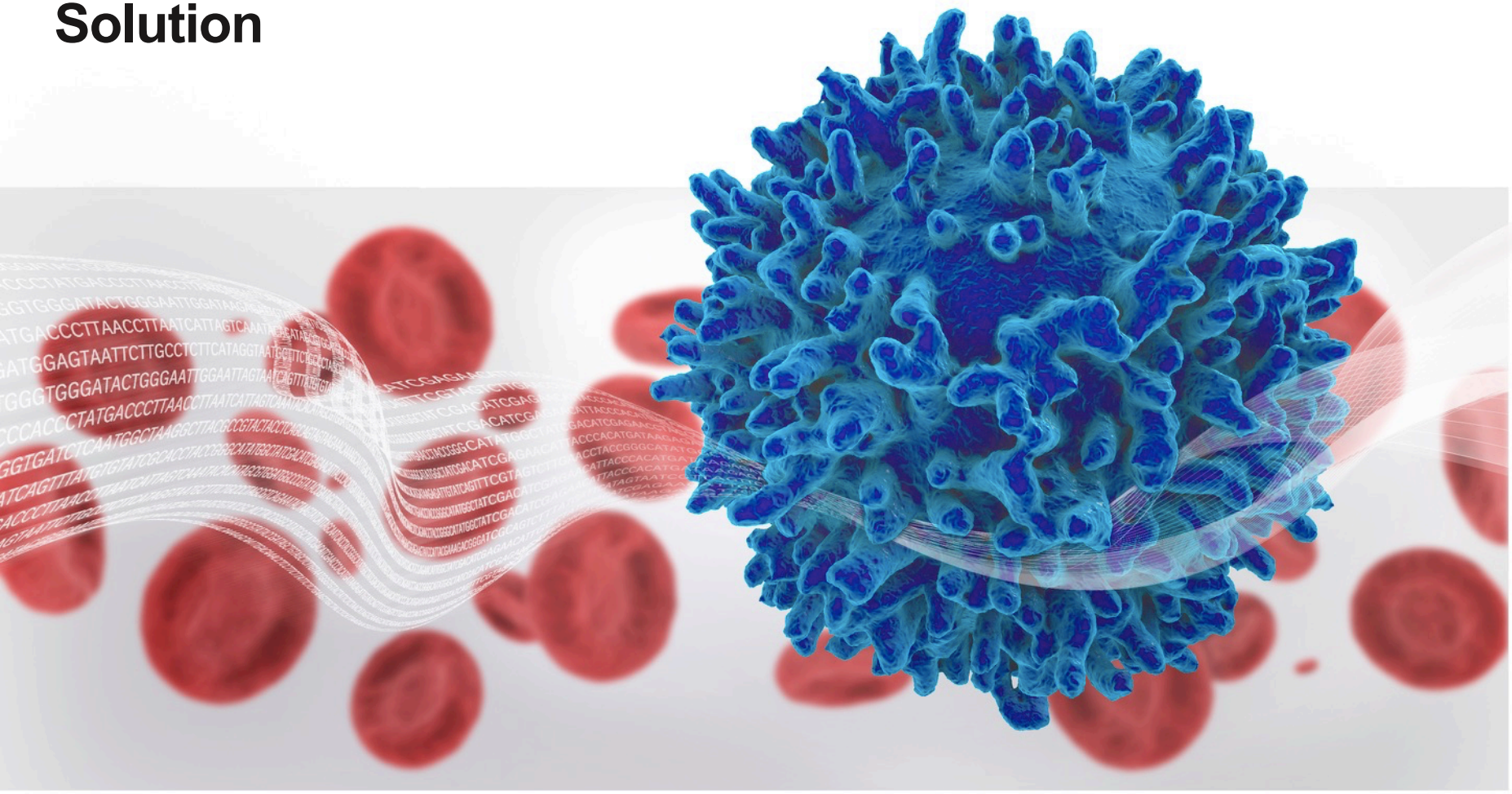
Prices in USD. Regional prices may vary

** Reflects AWS storage / compute prices

& Each compute hour will cost ~1-2 iCredits depending on application

illumina®

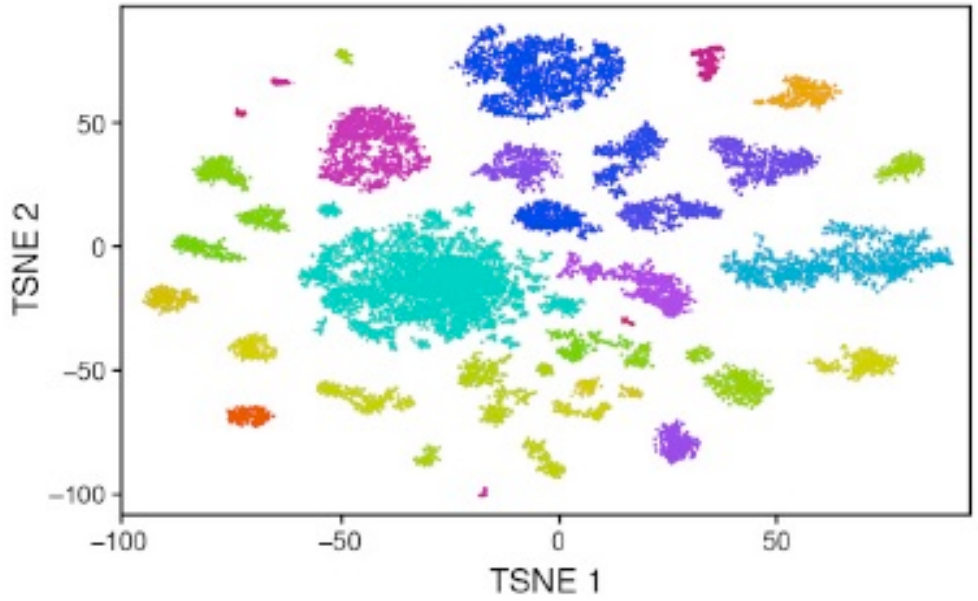
Illumina® | Bio-Rad® Single Cell Sequencing Solution



For Research Use Only. Not for use in diagnostic procedures.

illumina® | BIO-RAD

The importance of single cell sequencing



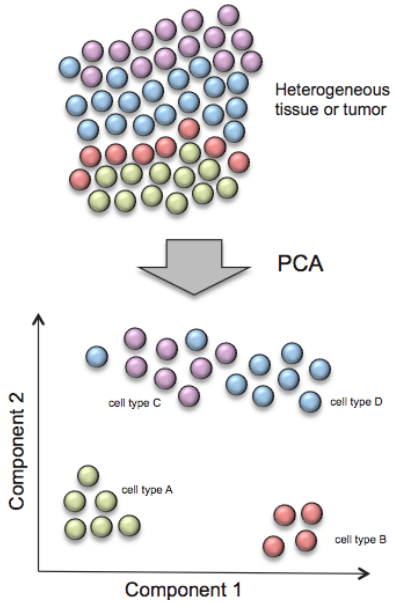
Macosko et al, Cell: May, 2-15

“The single cell ‘omics revolution is firmly underway. Nearly every expression study worth doing will be worth doing at single cell level…” Ewan Birney, EMBL

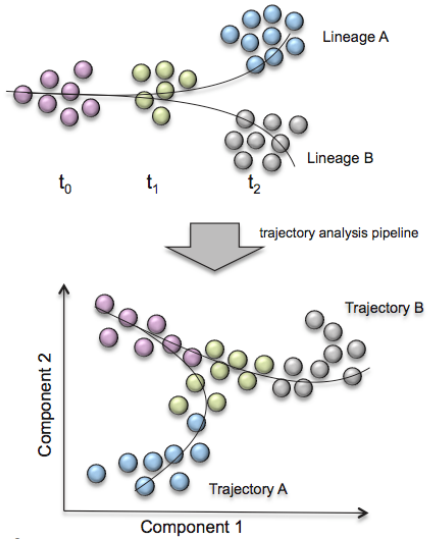
illumina® | BIO-RAD

Why Perform Single Cell Analysis ?

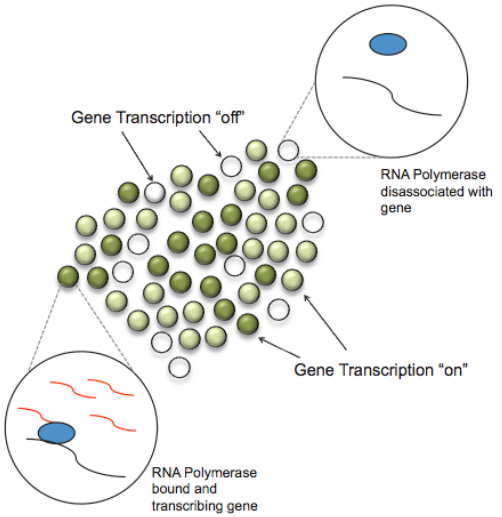
1. Assess cell-to-cell heterogeneity



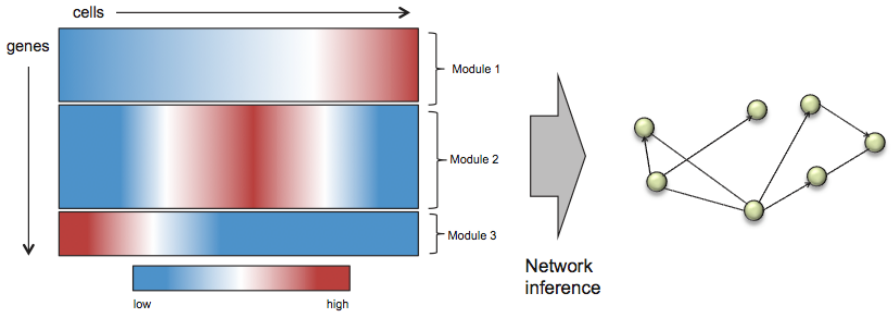
2. Map cell trajectories



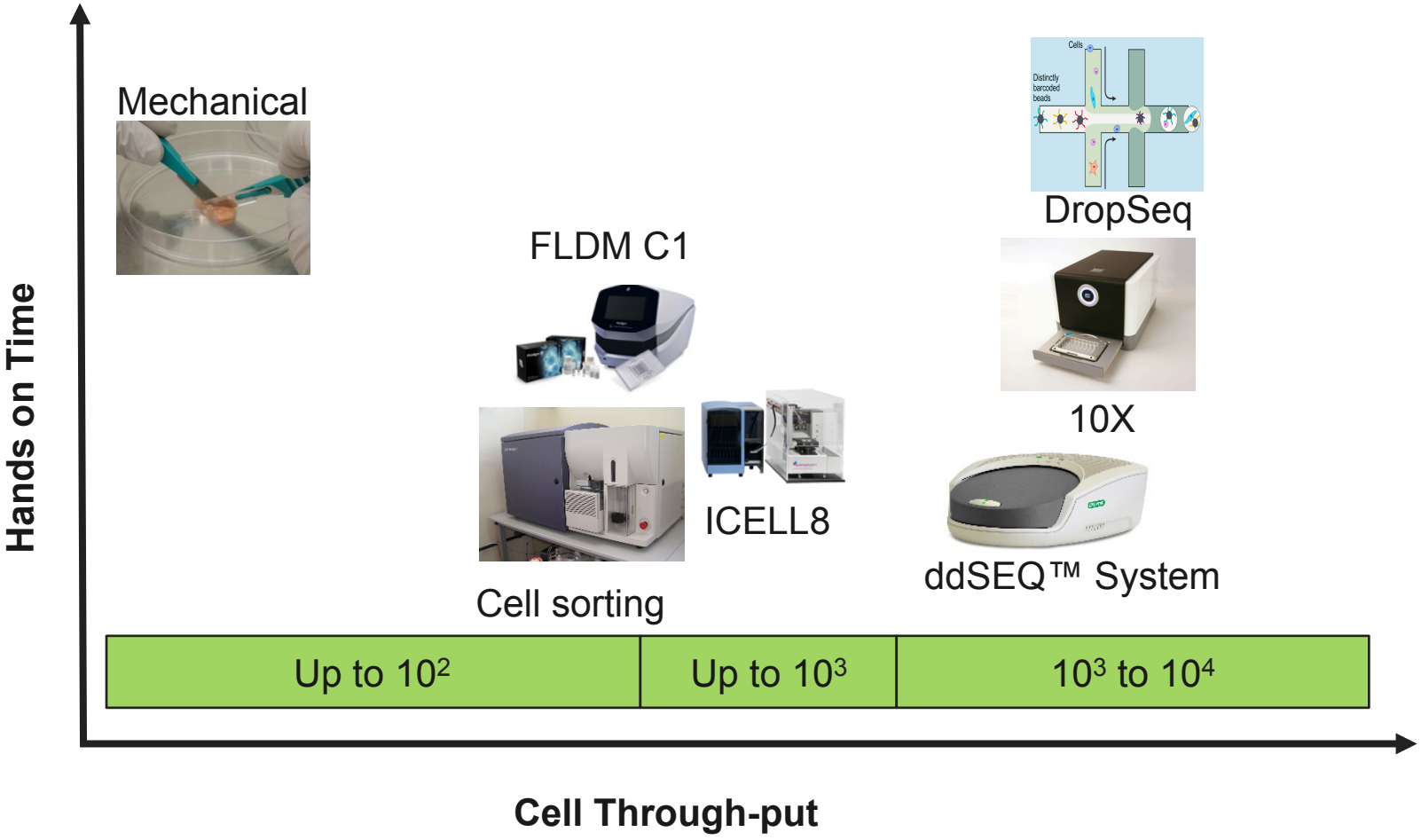
3. Dissect transcriptional mechanics



4. Infer gene regulatory networks

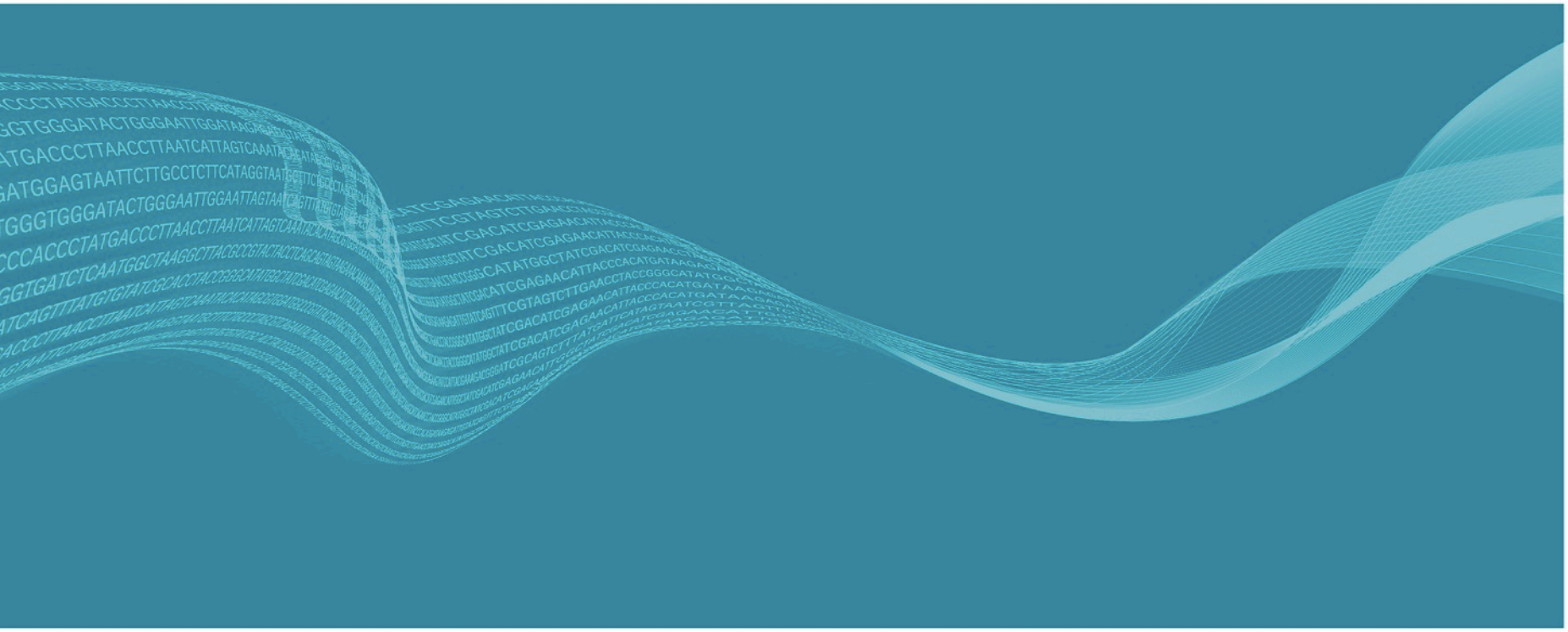


Single cell sample prep: Low vs High-Throughput



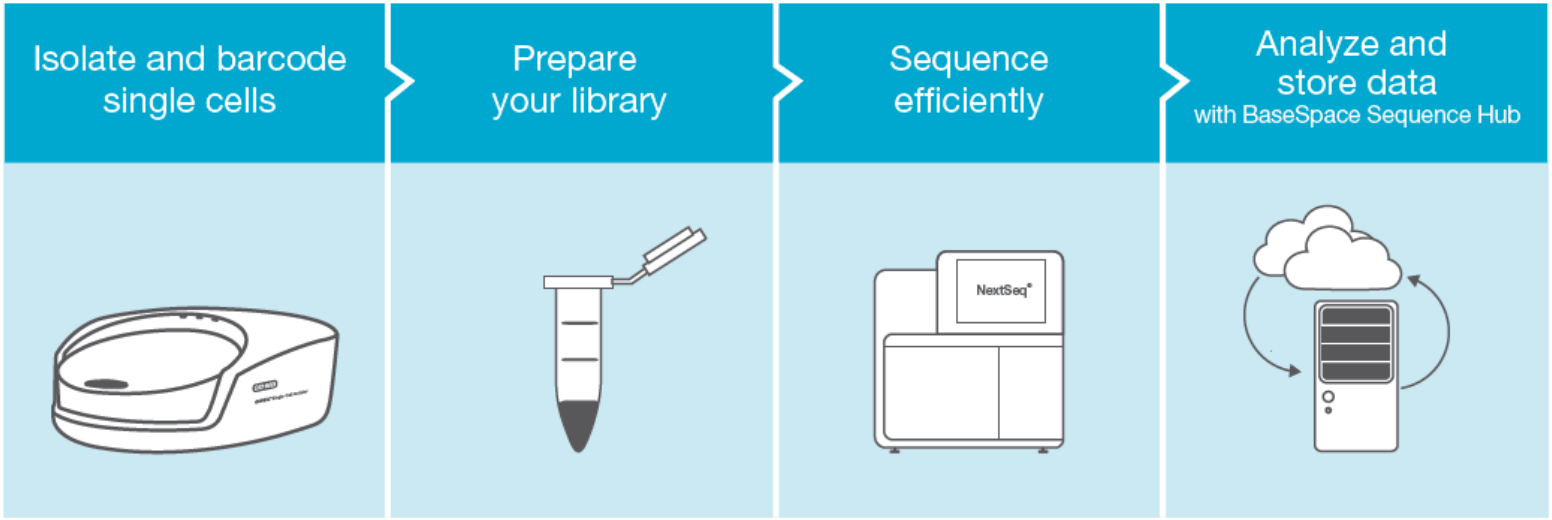
illumina® | BIO-RAD

Illumina® | Bio-Rad®
Single Cell Sequencing Solution



illumina® | BIO-RAD

The Illumina | Bio-Rad Single Cell Sequencing Solution



High-throughput single-cell sequencing workflow. Integrating the Bio-Rad industry-leading Droplet Digital partitioning technology with leading Illumina next-generation sequencing (NGS) technologies, Bio-Rad and Illumina will launch an isolation-to-analysis commercial solution that will enable high-throughput sequencing of thousands of individual cells quickly and cost effectively.

ddSEQ Single-Cell Isolator

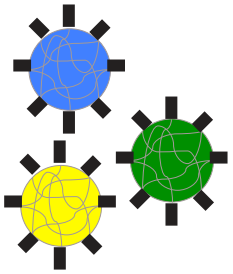
NextSeq & HiSeq

SureCell WTA 3' Library Prep Kit

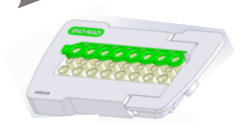
Single Cell RNA Seq v1.0.0

illumina® | BIO-RAD

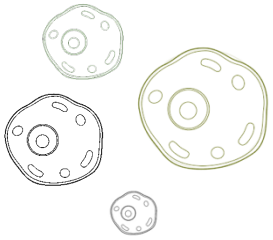
Encapsulate Thousands of Cells in < 5 min



Cell Barcoding Microspheres

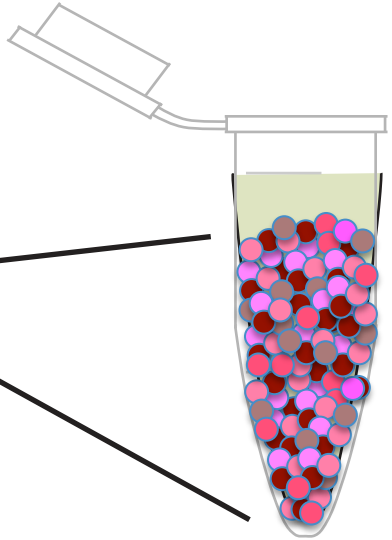


Disposable Cartridges



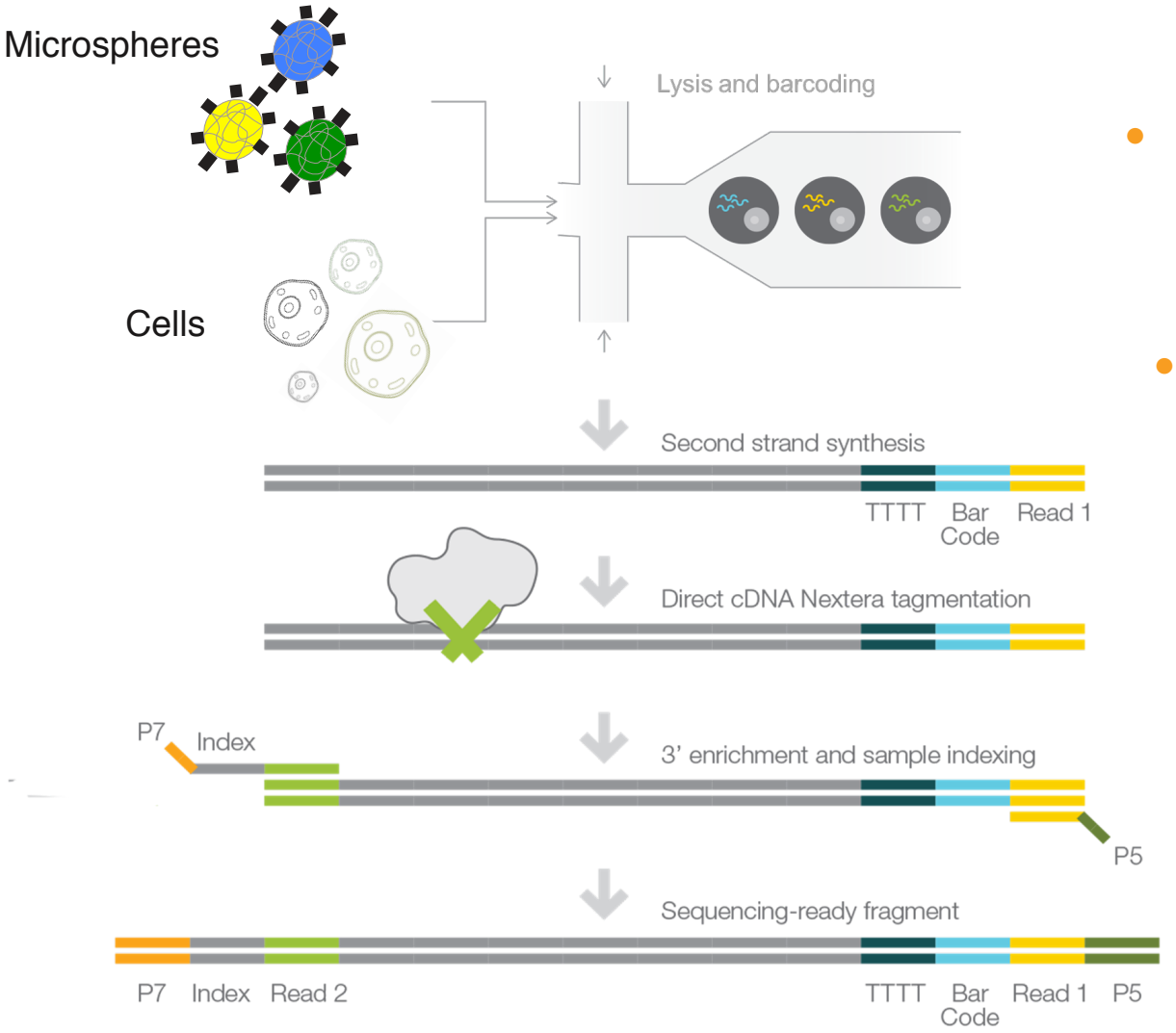
Cell Suspension

ddSEQ™ Single-Cell Isolator
with Droplet Digital™ Technology



Process Thousands of cells in minutes

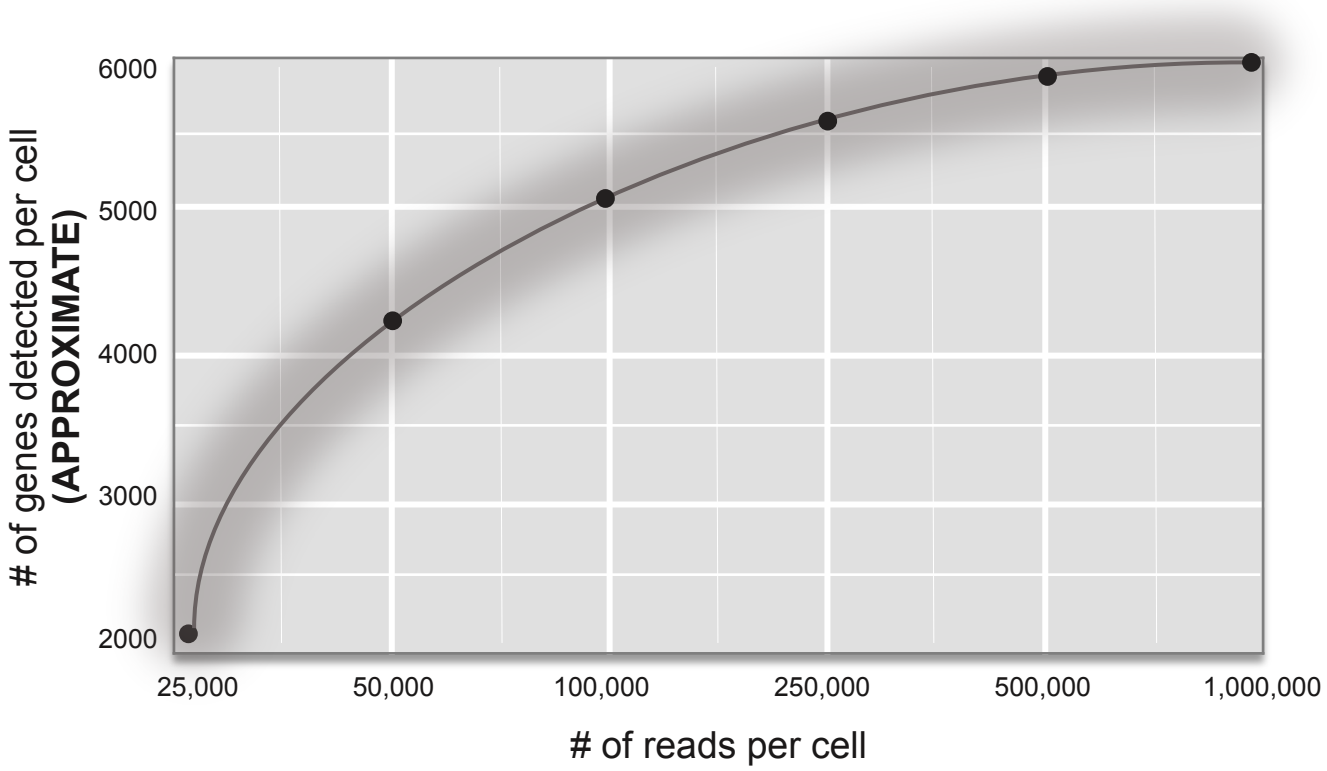
SureCell™ WTA 3’ Library Prep Kit for the ddSEQ™ System



- Single cell encapsulation with the ddSEQ™ Single-Cell Isolator
- Sensitive assay chemistry without pre-amplification
- Modified Nextera® Library prep

illumina® | BIO-RAD

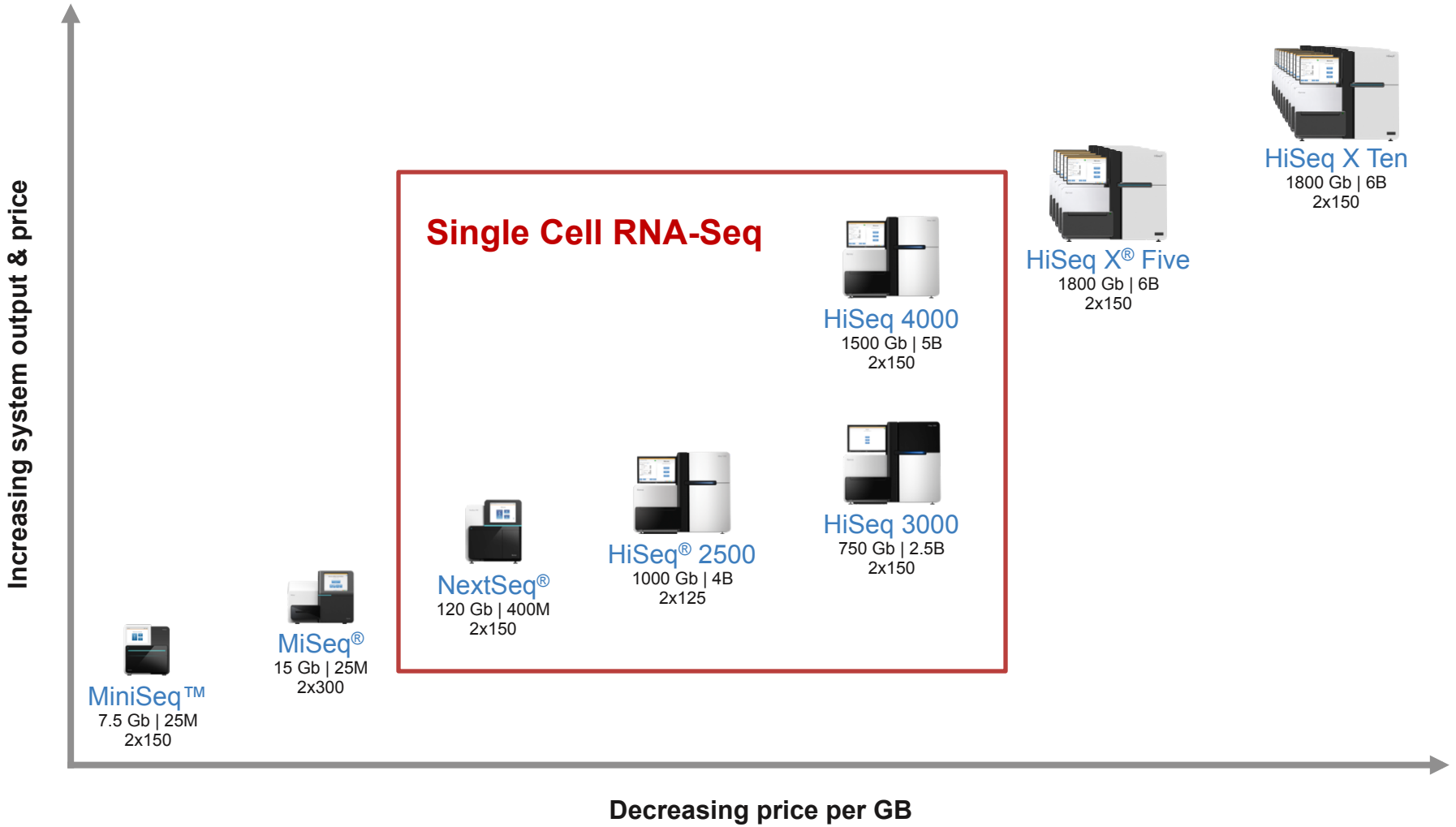
How much sequencing will I need?



- How many genes do I want to detect per cell?
- Am I looking for potentially rare cell types?
- Saturation of # detected genes requires more reads

Range depends on cell type and expression levels and also the biological question you're asking

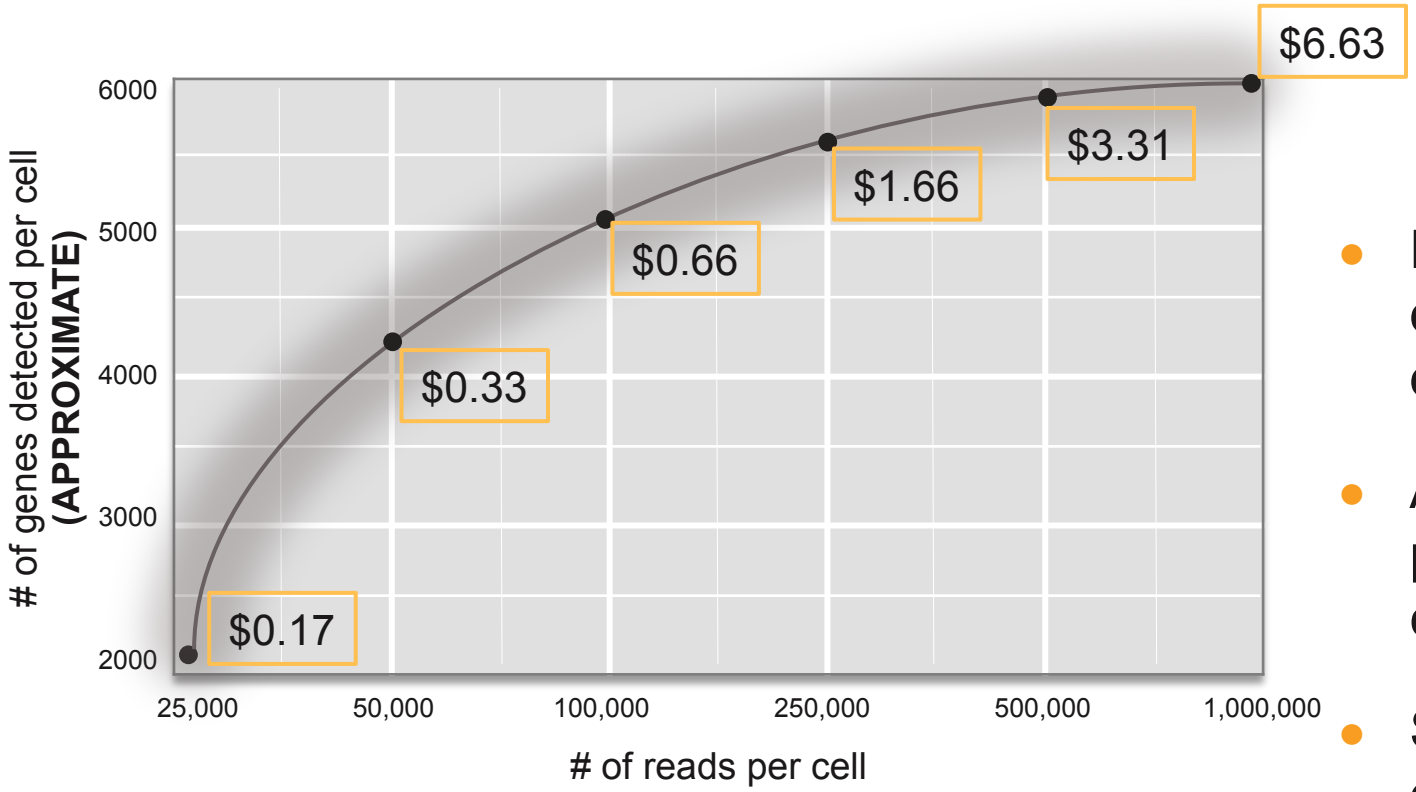
Sequencing Power for Every Scale



illumina® | BIO-RAD

Sequencing cost per cell

NextSeq 500

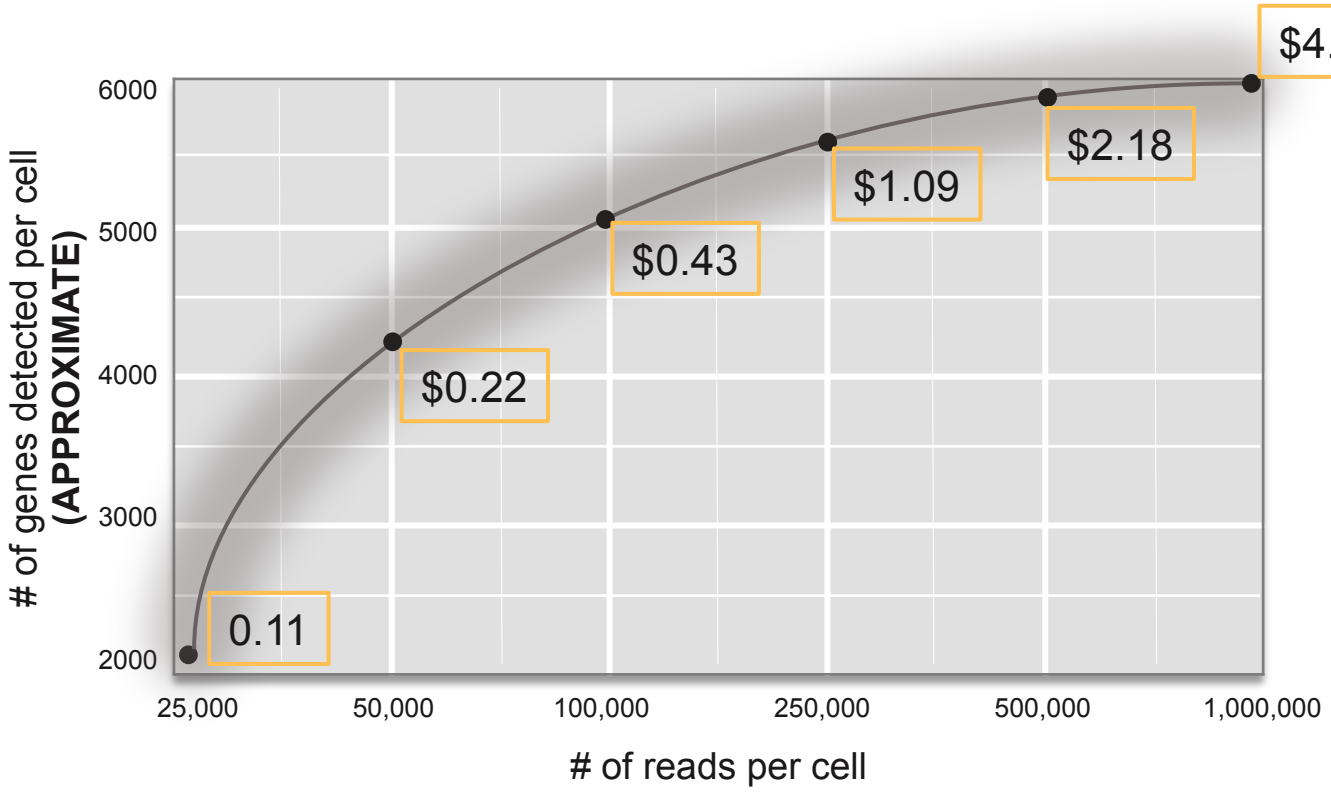


- How many genes do I want to detect per cell?
- Am I looking for potentially rare cell types?
- Saturation of # detected genes requires more reads

Range depends on cell type and expression levels and also the biological question you're asking

Sequencing cost per cell

HiSeq 3000



- How many genes do I want to detect per cell?
- Am I looking for potentially rare cell types?
- Saturation of # detected genes requires more reads

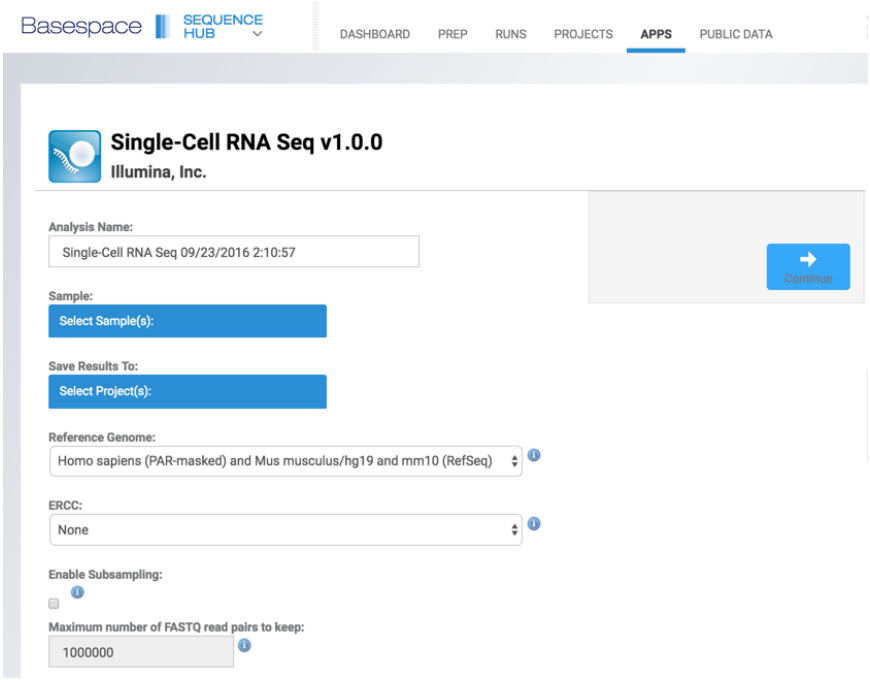
Range depends on cell type and expression levels and also the biological question you're asking

illumina | BIO-RAD

What about data analysis?

BaseSpace® Single-Cell RNA App

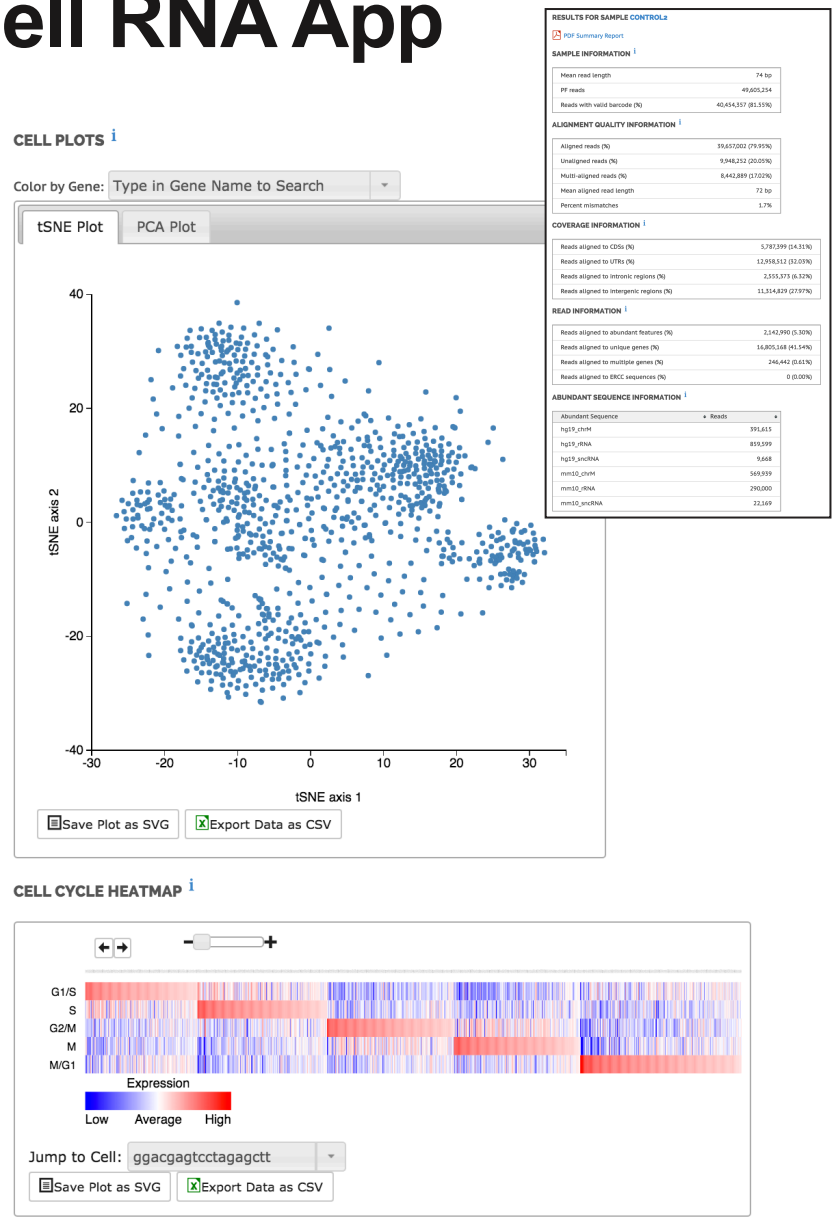
- **Simple analysis setup** for samples across multiple sequencing runs
  - Up to 96 samples per analysis
- **Easily choose analysis parameters**
  - Reference genome
  - ERCC spike-ins
  - Subsampling for QC
- **Rapid** alignment, cell and gene counting, and filtering



illumina® | BIO-RAD

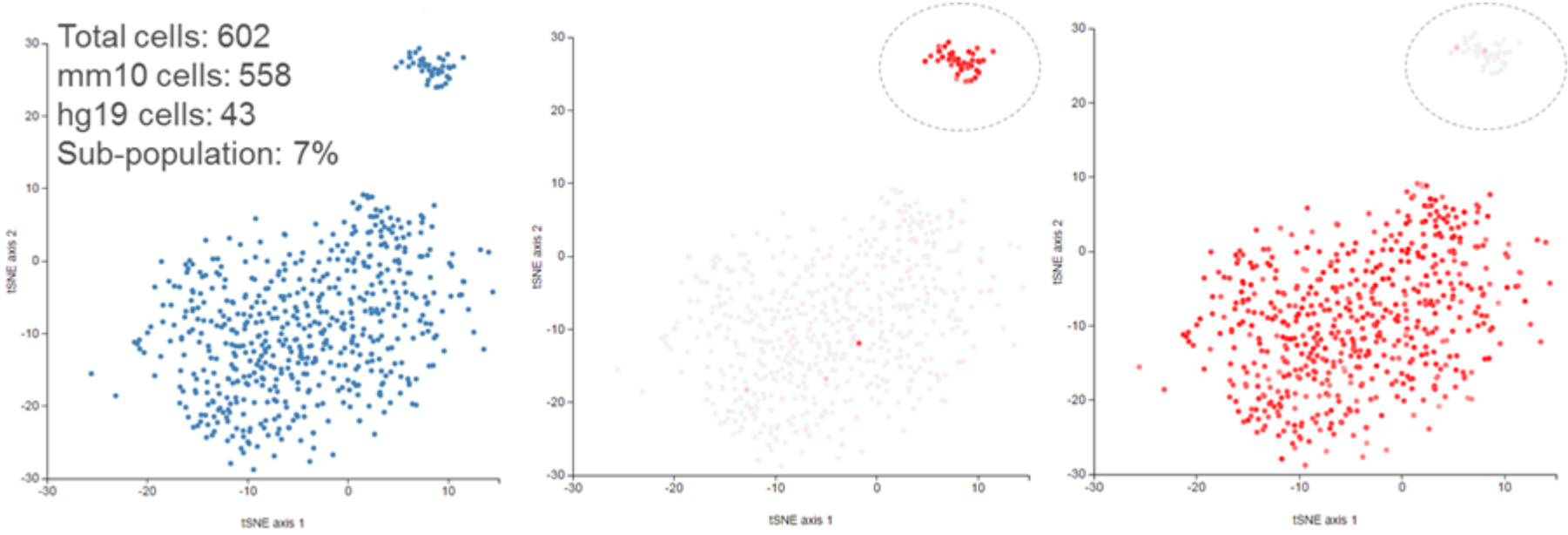
BaseSpace® Single-Cell RNA App

- **Per-sample reports** available in-browser and as PDF
  - Alignment quality
  - Coverage information
  - Abundant sequences
- **Perform global PCA and tSNE clustering**
- **Cell-cycle heatmap**
- **All output files available for download**, or as input into downstream applications
  - Includes cell-gene expression table



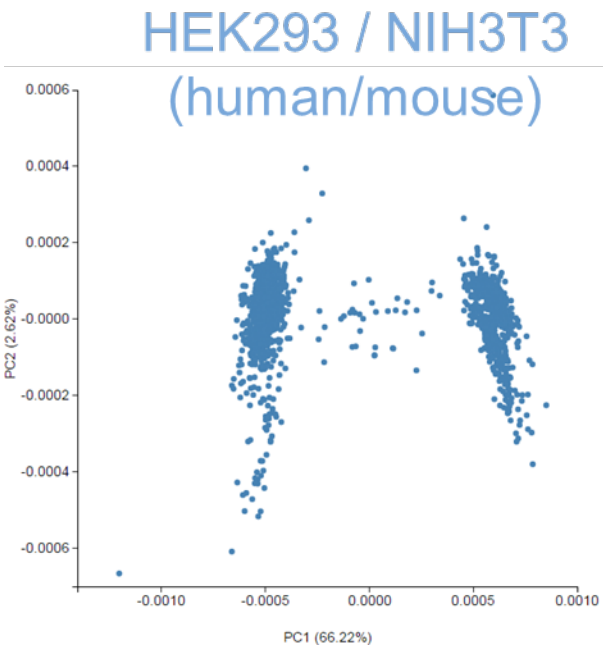
illumina® | BIO-RAD

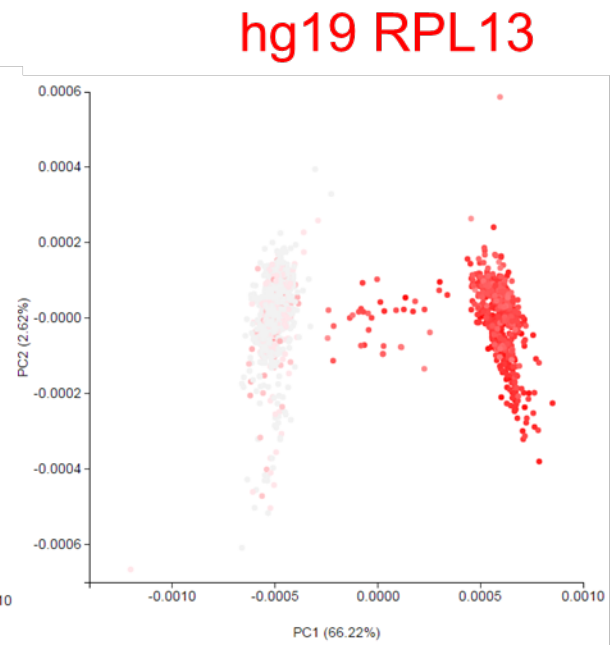
T-SNE analysis identifies a sub-population in a heterogeneous cell mixture mouse/human

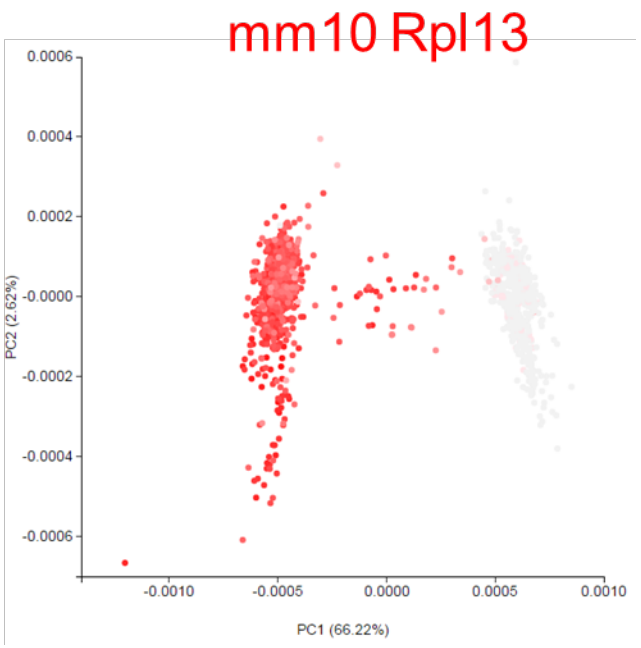


illumina® | BIO-RAD

PCA clustering of 1:1 mixture of 1,400 mouse and human cells detects distinct population

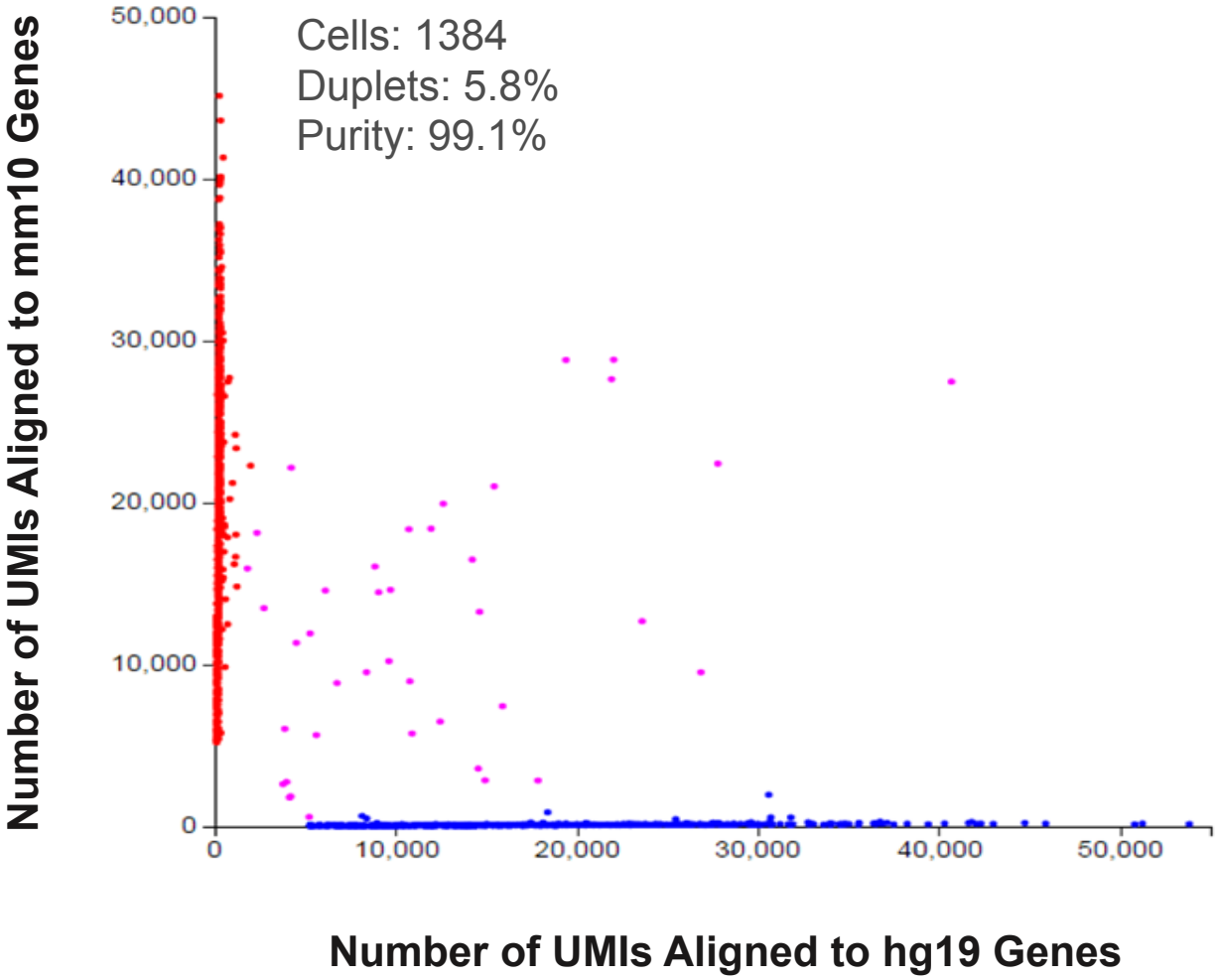






illumina® | BIO-RAD

Two-species cell mixture (HEK293/NIH3T3) demonstrates low crosstalk and high purity

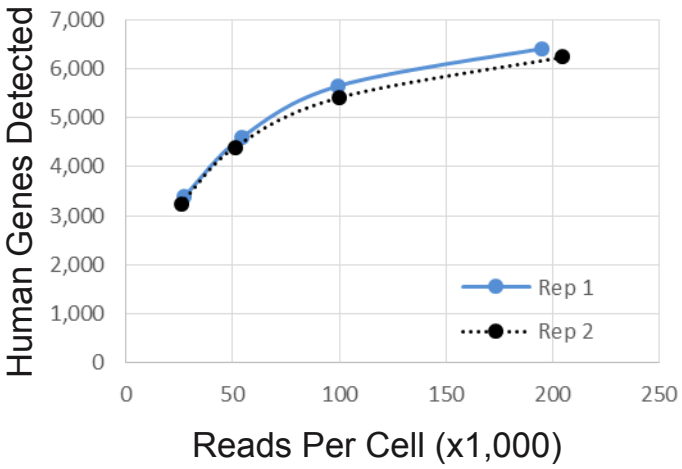


illumina® | BIO-RAD

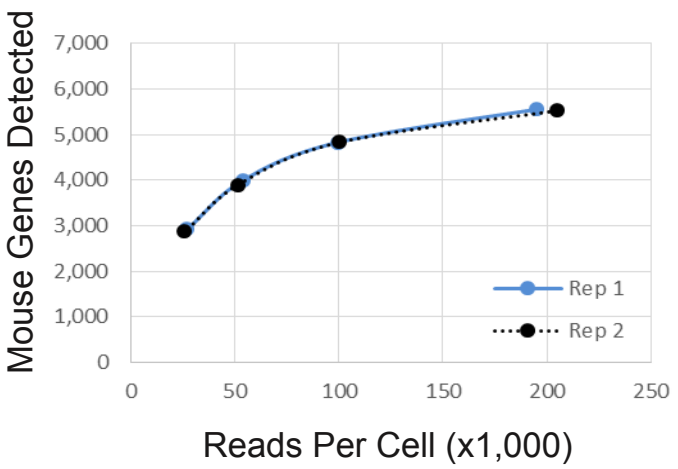
System performance on reference cells

Sensitivity & Reproducibility Across Cell Lines

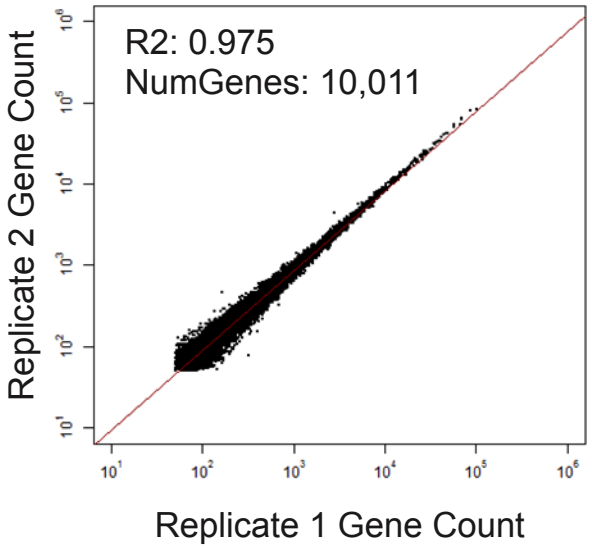
HEK293 Genes vs. Reads Per Cell



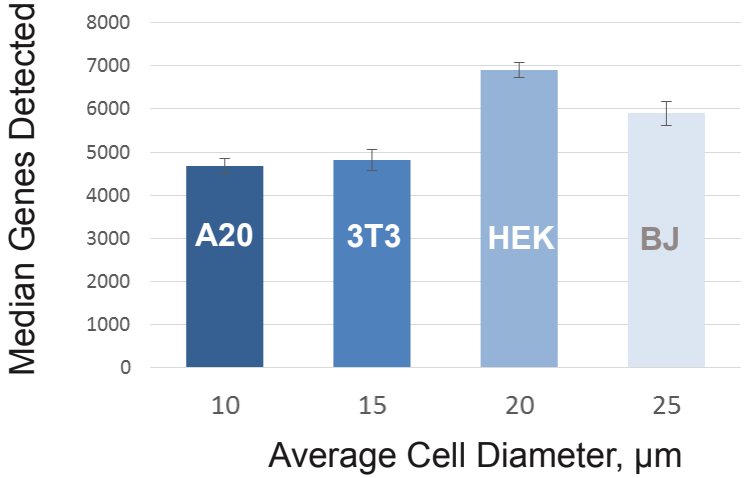
NIH3T3 Genes vs. Reads Per Cell



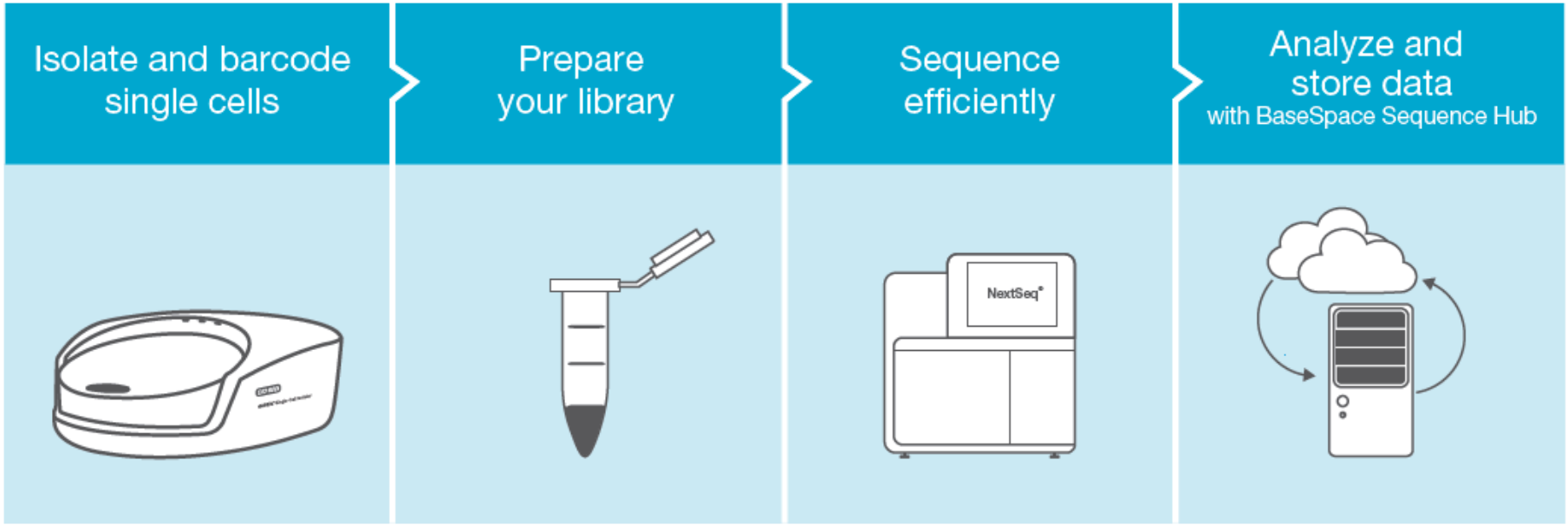
Gene Count Reproducibility



Genes Detected Across Different Cell Sizes



The Illumina | Bio-Rad Single Cell Sequencing Solution



1. **Isolation and barcoding of several thousand single cells in minutes**
2. **Affordable platform**
3. **Scalable to profile increasing number of cells in a cost-effective manner**
4. **Proven Illumina library prep with streamlined workflow**
5. **Single-Cell RNA-Seq BaseSpace® App available**

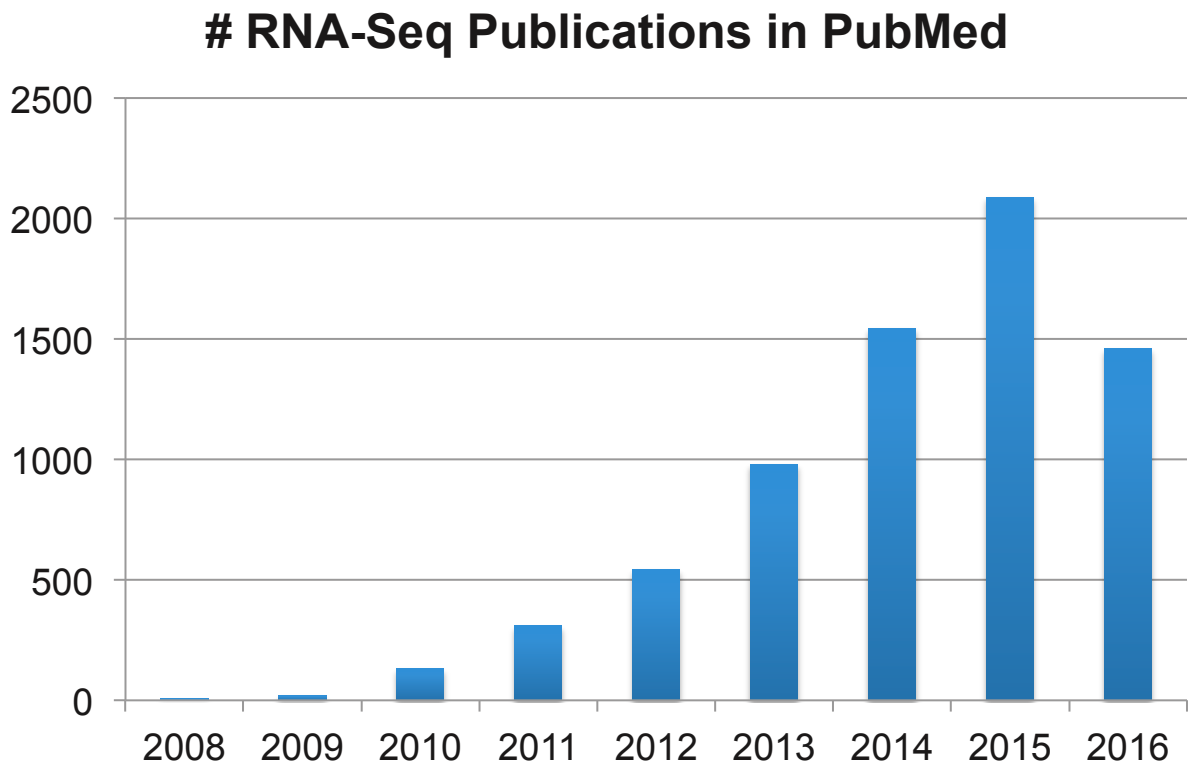
illumina® | BIO-RAD

The Next Step in Expression

Combining RNA-Seq and Methylation Analysis

illumina®

RNA-Seq is still growing in popularity



Source: NIH NCBI PubMed June 2016

illumina®

RNA-Seq is Impacting All Areas of Biological Research

Genomic Expression makes Red Herring "Top 100" for disprupting cancer care with RNA-Seq www.redherring.com June 2016

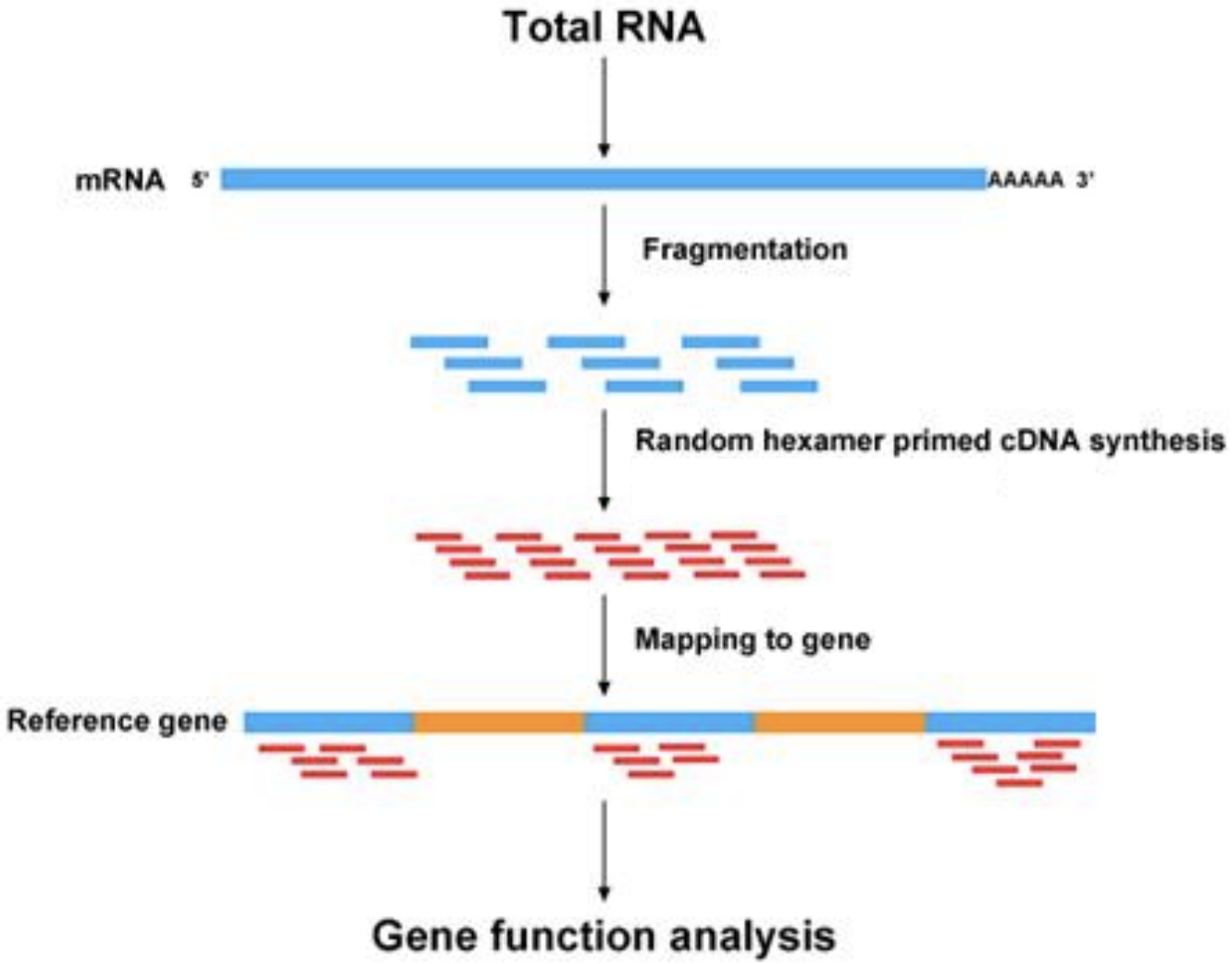
Using big data, scientists discover biomarkers that could help give cancer patients better survival estimates UCLA Newsroom June 2017

RNA-Seq identifies novel myocardial gene expression signatures of heart failure Yu et al., Science Direct February 2015

Single-cell RNA sequencing reveals human brain houses diverse populations of neurons RNASeqBlog June 2016

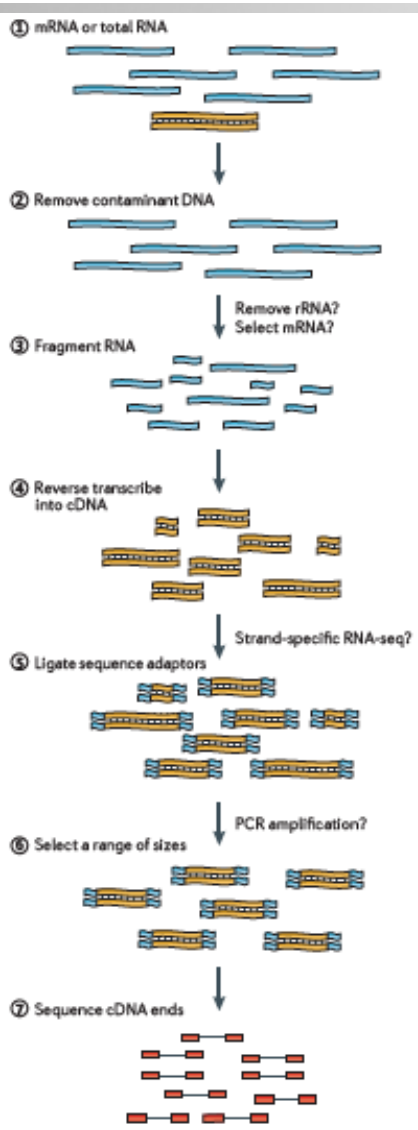
illumina®

How RNA-Seq Works



illumina®

Considerations for RNA-Seq Library Preparation

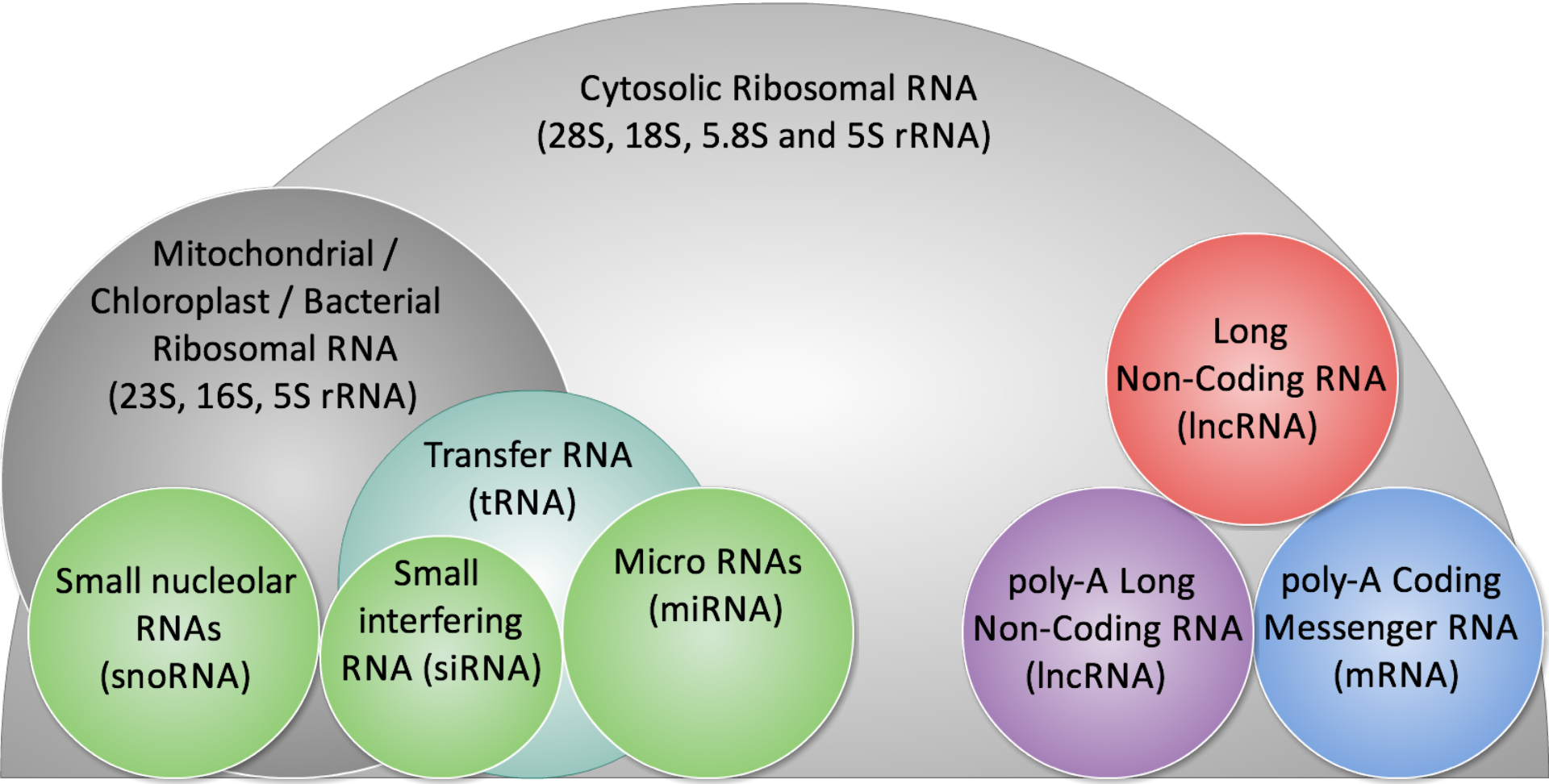


- What is the integrity of the input RNA?
- What is the source of the RNA? (FFPE?)
- Which RNA Type is of interest (mRNA or Total RNA)
- How much Total RNA is available per Sample?
- Which RNA-Seq application is planned (counting, discovery)?

From: Martin, J. A., and Z. Wang, 2011 Next-generation transcriptome assembly. Nat Rev Genet **12:** 671-682.

illumina®

Many different RNAs exist - Ribosomal RNA is most abundant RNA species – and least dynamic

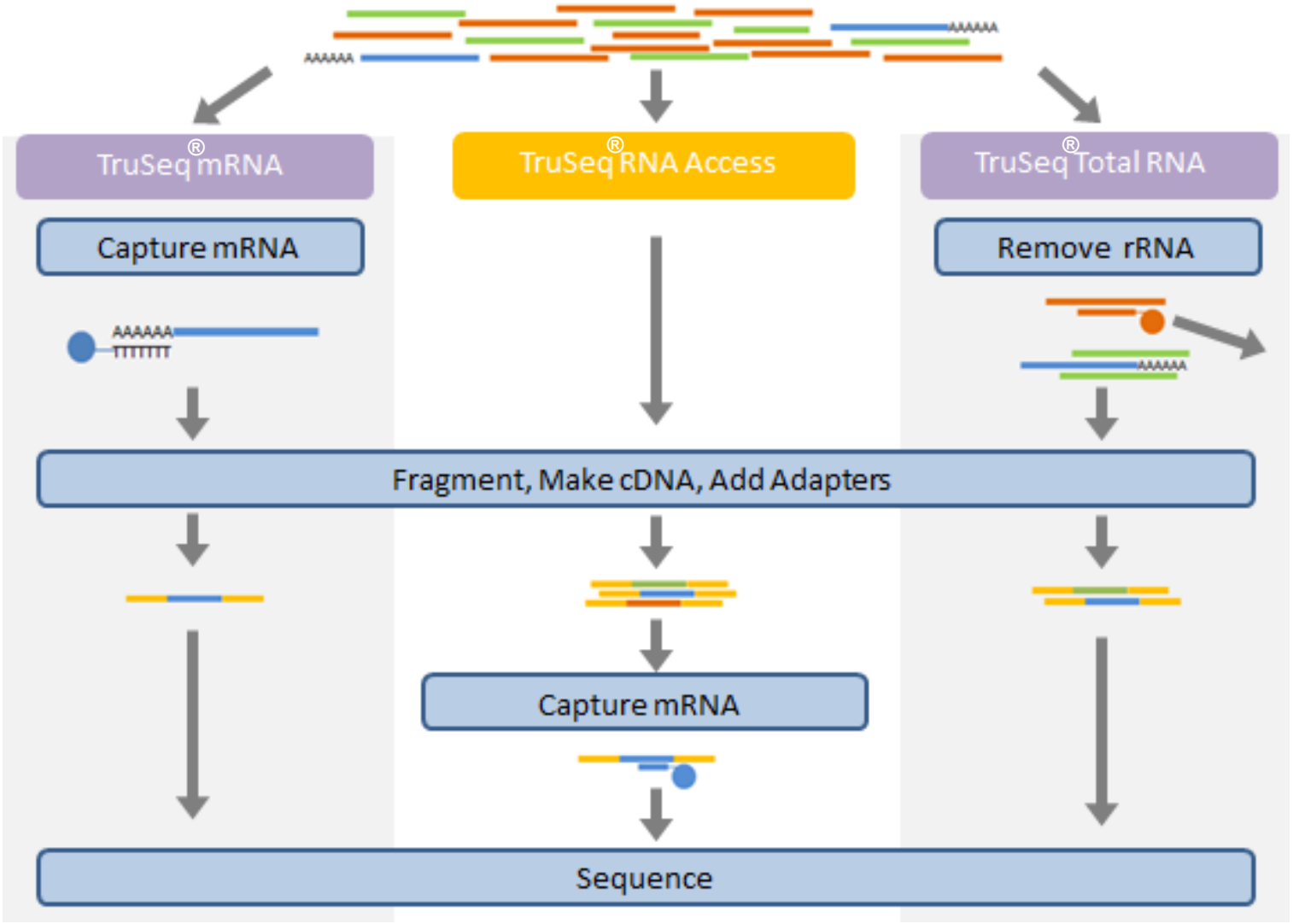


illumina®

Illumina's Suite of RNA Library Prep Solutions

Total RNA-Seq	mRNA-Seq/ GEx Profiling		Targeted Profiling	miRNA Analysis
				
TruSeq Stranded Total RNA	**TruSeq Stranded mRNA**	**TruSeq RNA Access**	**TruSeq Targeted RNA Expression**	**TruSeq small RNA**
• Coding + ncRNA • Transcript-level abundance • Splicing Analysis • Fusion Discovery • FFPE compatible	• Coding RNA • Transcript-level abundance • Splicing Analysis • Fusion Discovery	• Coding RNA • Transcript-level abundance • Splicing Analysis • Fusion Discovery • FFPE Compatible	• 10s-1,000s of targets • Coding + ncRNA • Transcript-level abundance • Fusion Validation • FFPE Compatible	• miRNA abundance • isomiR detection

RNA-Seq Experimental Design Depends on Your Target



illumina®

Four Easy Steps to RNA-Seq Results

1. Set up/run TopHat

2. QC of TopHat results

- Filter out challenging samples as needed

3. Set up/run Cufflinks

- Name/select control group
- Name/select comparison group

4. Visualize Group 1: Group 2

- GEX correlation

TopHat Alignment

Illumina, Inc.

App Session Name: TopHat Alignment 03/15/2015 10:50:29

Save Results To: Select Project(s):

Workshop ×

Samples: Select Sample(s):

Select All	Stranded
UHR-Access-2-5M	✓ ×
Brain-Access-2-5M	✓ ×
UHR-Access-1-5M	✓ ×
Brain-Access-1-5M	✓ ×

Reference Genome: Homo sapiens/hg19 (RefSeq)

Options

Call Fusions: ✓

Trim TruSeq Adapters: ☐

This app is free.

Continue

contact us

illumina

Analysis Info
Inputs
Output Files

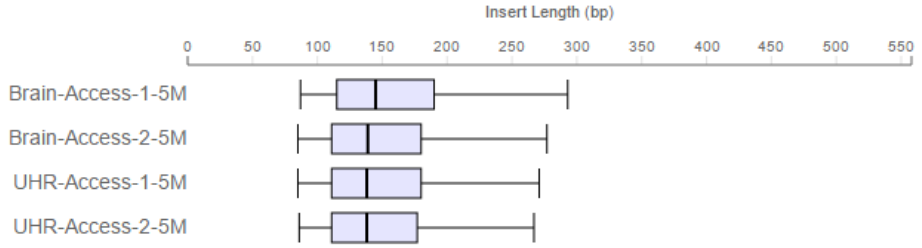
Analysis Reports
Summary
UHR-Access-1-5M
Brain-Access-1-5M
Brain-Access-2-5M
UHR-Access-2-5M

App Session TopHat Alignment & Call Fusions 4 Samples

Summary

	Reads	Number of Reads	% Total Aligned	% Abundant	% Unaligned	Median CV Coverage Uniformity	% Stranded
Brain-Access-1-5M	76/76	2,500,000	98.25%	3.75%	1.75%	1.03	98.21%
Brain-Access-2-5M	76/76	2,500,000	98.30%	3.06%	1.70%	1.03	98.45%
UHR-Access-1-5M	76/76	2,500,000	98.32%	3.29%	1.68%	0.93	99.44%
UHR-Access-2-5M	76/76	2,500,000	98.21%	3.64%	1.79%	0.94	99.41%

Insert Length Distribution



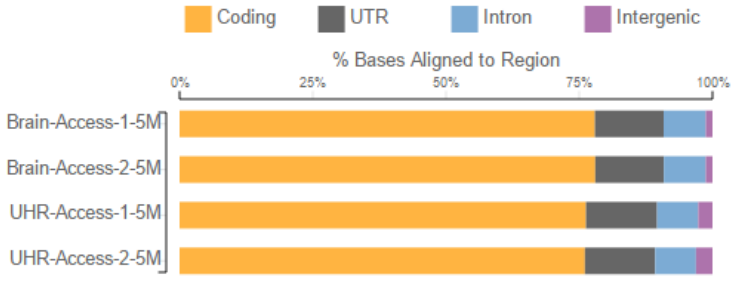
Save Plot as SVG

Alignment Distribution

Analysis Info
Inputs
Output Files

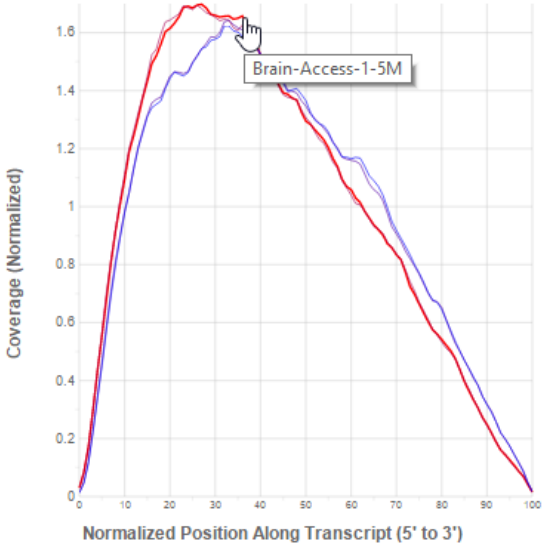
Analysis Reports
Summary
UHR-Access-1-5M
Brain-Access-1-5M
Brain-Access-2-5M
UHR-Access-2-5M

Alignment Distribution



Save Plot as SVG

Transcript Coverage



Save Plot as SVG

contact us

Analysis Info
Inputs
Output Files

Analysis Reports
Summary
UHR-Access-1-5M
Brain-Access-1-5M
Brain-Access-2-5M
UHR-Access-2-5M

Results for UHR-Access-1-5M

Primary Analysis Information

	Read 1	Read 2
Read Length	76	76
Number of Reads	2,500,000	2,500,000
Bases (GB)	0.19	0.19
Q30 Bases (GB)	0.18	0.18

Insert Information

Insert Length Mean	161.04
Insert Length S.D.	84.29
Duplicates (% Reads)	7.40%

Alignment Quality

	Read 1	Read 2
Total Aligned Reads (% Reads)	99.00%	97.63%
Abundant Reads (% Reads)	3.30%	3.28%
Unaligned Reads (% Reads)	1.00%	2.37%
Reads with spliced alignment (% Aligned Reads)	36.53%	39.93%
Reads aligned at multiple loci (% Aligned Reads)	8.26%	8.26%

Analysis Info
Inputs
Output Files

Analysis Reports
Summary
UHR-Access-1-5M
Brain-Access-1-5M
Brain-Access-2-5M
UHR-Access-2-5M

Coverage Uniformity

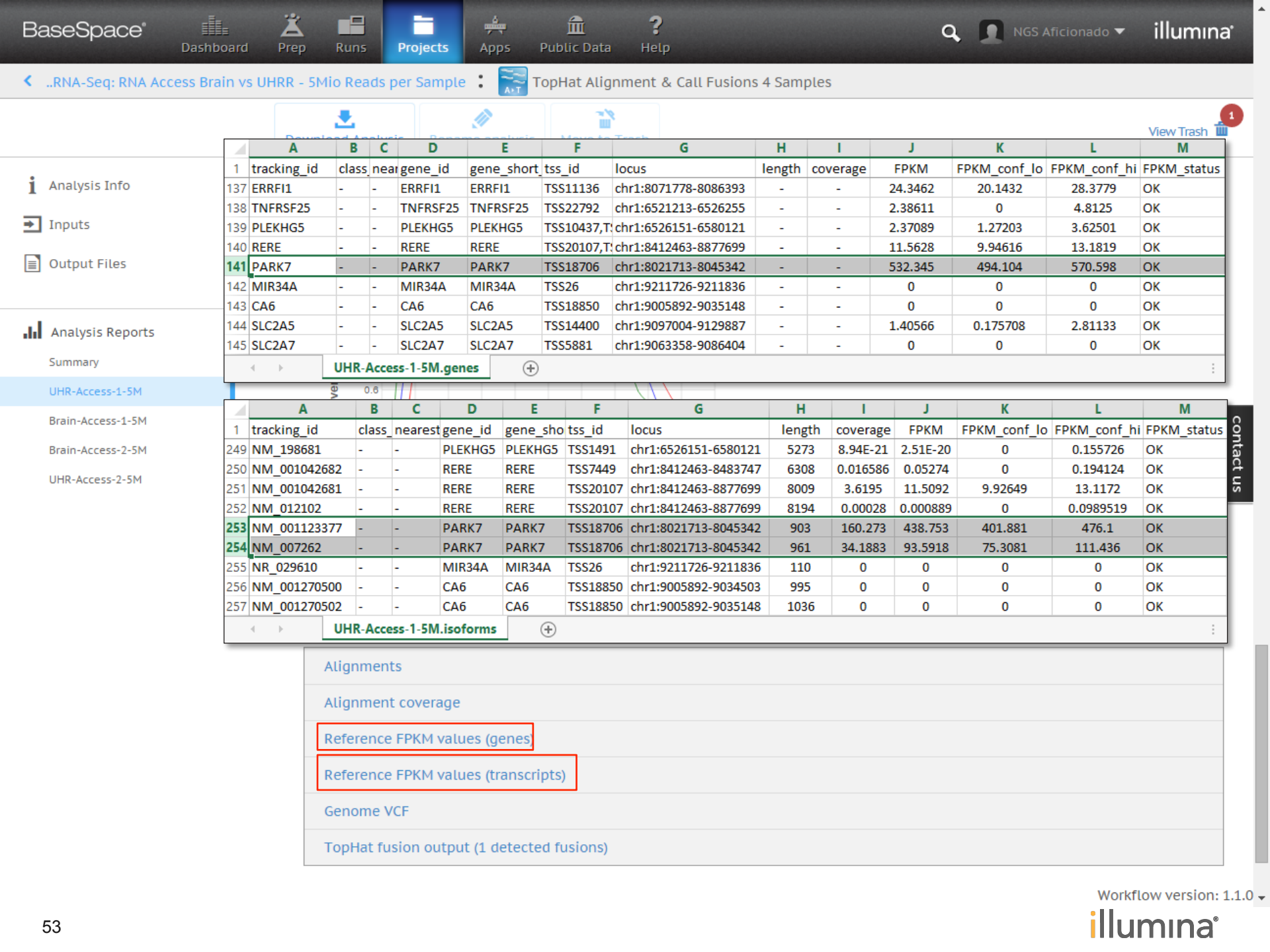
	Read 1	Read 2	Combined
Median CV	1.08	1.04	0.93
Median 3'	0.00	0.00	0.00
Median 5'	0.00	0.05	0.03
Reads aligned to correct strand	99.43%	99.44%	99.44%

Alignment Information

Region	Fold Coverage	% Bases
Coding	8.97x	76.27%
UTR	1.32x	13.34%
Intron	0.03x	7.73%
Intergenic	0.01x	2.67%

Variant Calls

Homozygous reference	759,661
Heterozygous	3,396
Homozygous variant	48
SNV	3,429
Indel	15
T_n/T_v	2.80

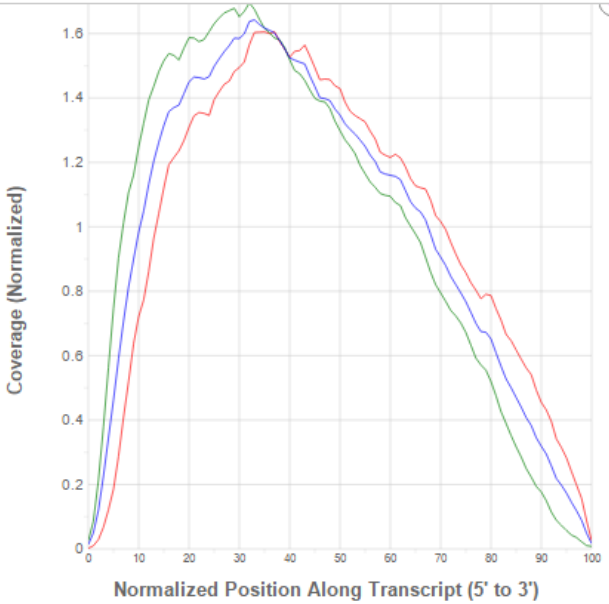


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	tracking_id	class_	near	gene_id	gene_short_	tss_id	locus	length	coverage	FPKM	FPKM_conf_lo	FPKM_conf_hi	FPKM_status
137	ERRFI1	-	-	ERRFI1	ERRFI1	TSS11136	chr1:8071778-8086393	-	-	24.3462	20.1432	28.3779	OK
138	TNFRSF25	-	-	TNFRSF25	TNFRSF25	TSS22792	chr1:6521213-6526255	-	-	2.38611	0	4.8125	OK
139	PLEKHG5	-	-	PLEKHG5	PLEKHG5	TSS10437,T	chr1:6526151-6580121	-	-	2.37089	1.27203	3.62501	OK
140	RERE	-	-	RERE	RERE	TSS20107,T	chr1:8412463-8877699	-	-	11.5628	9.94616	13.1819	OK
141	PARK7	-	-	PARK7	PARK7	TSS18706	chr1:8021713-8045342	-	-	532.345	494.104	570.598	OK
142	MIR34A	-	-	MIR34A	MIR34A	TSS26	chr1:9211726-9211836	-	-	0	0	0	OK
143	CA6	-	-	CA6	CA6	TSS18850	chr1:9005892-9035148	-	-	0	0	0	OK
144	SLC2A5	-	-	SLC2A5	SLC2A5	TSS14400	chr1:9097004-9129887	-	-	1.40566	0.175708	2.81133	OK
145	SLC2A7	-	-	SLC2A7	SLC2A7	TSS5881	chr1:9063358-9086404	-	-	0	0	0	OK

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	tracking_id	class_	nearest	gene_id	gene_sho	tss_id	locus	length	coverage	FPKM	FPKM_conf_lo	FPKM_conf_hi	FPKM_status
249	NM_198681	-	-	PLEKHG5	PLEKHG5	TSS1491	chr1:6526151-6580121	5273	8.94E-21	2.51E-20	0	0.155726	OK
250	NM_001042682	-	-	RERE	RERE	TSS7449	chr1:8412463-8483747	6308	0.016586	0.05274	0	0.194124	OK
251	NM_001042681	-	-	RERE	RERE	TSS20107	chr1:8412463-8877699	8009	3.6195	11.5092	9.92649	13.1172	OK
252	NM_012102	-	-	RERE	RERE	TSS20107	chr1:8412463-8877699	8194	0.00028	0.000889	0	0.0989519	OK
253	NM_001123377	-	-	PARK7	PARK7	TSS18706	chr1:8021713-8045342	903	160.273	438.753	401.881	476.1	OK
254	NM_007262	-	-	PARK7	PARK7	TSS18706	chr1:8021713-8045342	961	34.1883	93.5918	75.3081	111.436	OK
255	NR_029610	-	-	MIR34A	MIR34A	TSS26	chr1:9211726-9211836	110	0	0	0	0	OK
256	NM_001270500	-	-	CA6	CA6	TSS18850	chr1:9005892-9034503	995	0	0	0	0	OK
257	NM_001270502	-	-	CA6	CA6	TSS18850	chr1:9005892-9035148	1036	0	0	0	0	OK

Analysis Info
Inputs
Output Files

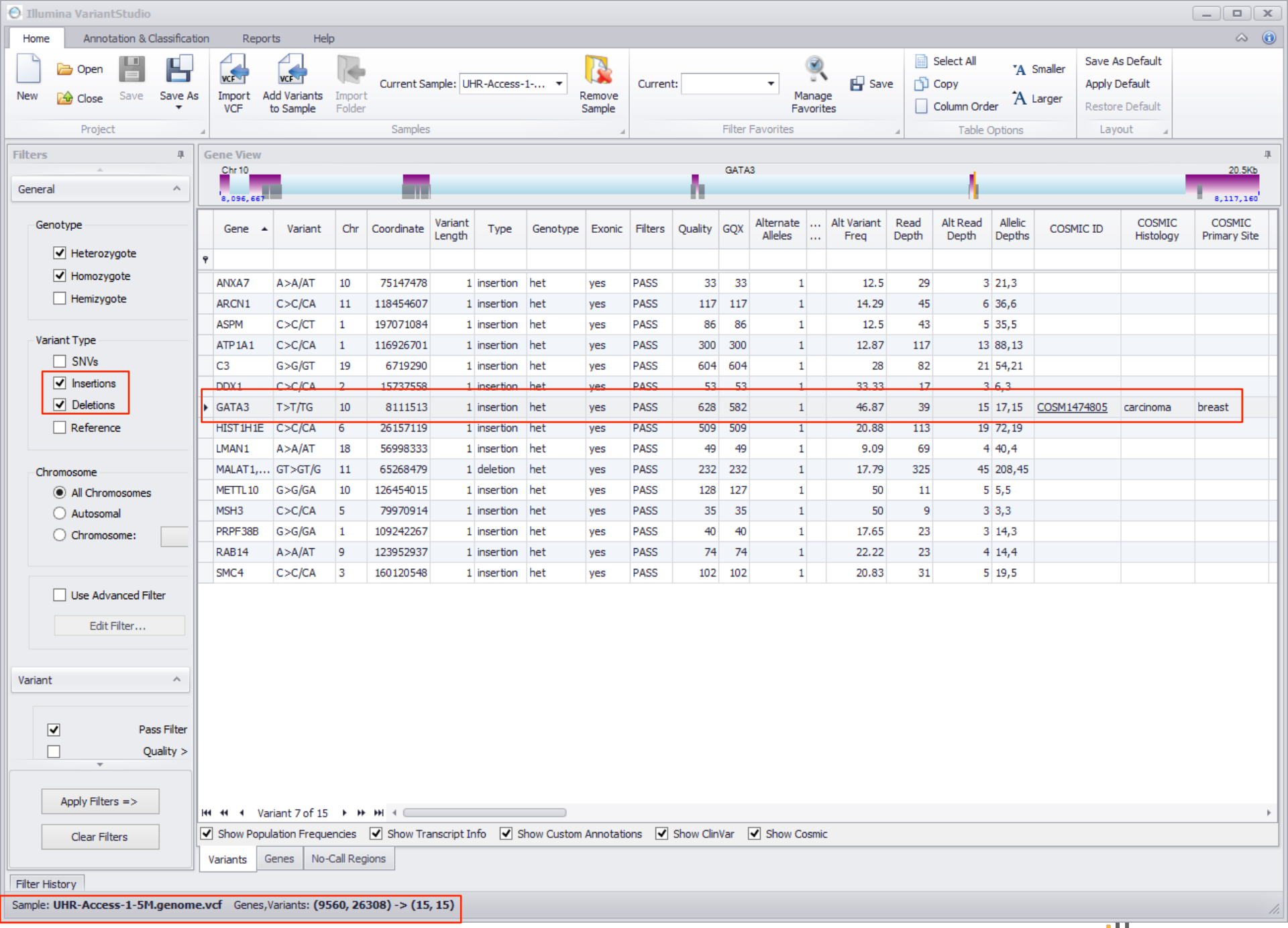
Analysis Reports
Summary
UHR-Access-1-5M
Brain-Access-1-5M
Brain-Access-2-5M
UHR-Access-2-5M



Save Plot as SVG | Export Data as CSV

Important Files for Download

- Alignments
- Alignment coverage
- Reference FPKM values (genes)
- Reference FPKM values (transcripts)
- Genome VCF
- TopHat fusion output (1 detected fusions)

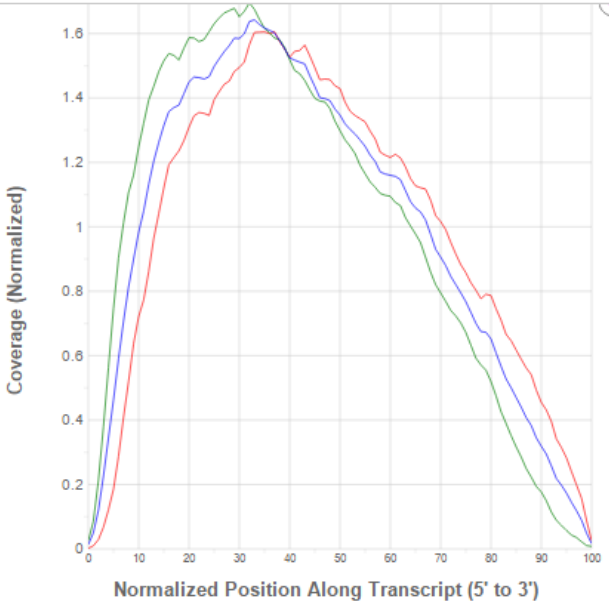

Illumina VariantStudio
Home
Annotation & Classification
Reports
Help
New
Open
Close
Save
Save As
Project
Import VCF
Add Variants to Sample
Import Folder
Current Sample: UHR-Access-1-...
Remove Sample
Samples
Current:
Manage Favorites
Save
Filter Favorites
Select All
Copy
Column Order
Smaller
Larger
Table Options
Save As Default
Apply Default
Restore Default
Layout
Filters
General
Genotype
Heterozygote
Homozygote
Hemizygote
Variant Type
SNVs
Insertions
Deletions
Reference
Chromosome
All Chromosomes
Autosomal
Chromosome:
Use Advanced Filter
Edit Filter...
Variant
Pass Filter
Quality >
Apply Filters =>
Clear Filters
Gene View
Chr 10
GATA3
20.5Kb
8,096,667
8,117,160
Gene
Variant
Chr
Coordinate
Variant Length
Type
Genotype
Exonic
Filters
Quality
GQX
Alternate Alleles
Alt Variant Freq
Read Depth
Alt Read Depth
Allelic Depths
COSMIC ID
COSMIC Histology
COSMIC Primary Site
ANXA7 A>A/AT 10 75147478 1 insertion het yes PASS 33 33 1 12.5 29 3 21,3
ARCN1 C>C/CA 11 118454607 1 insertion het yes PASS 117 117 1 14.29 45 6 36,6
ASPM C>C/CT 1 197071084 1 insertion het yes PASS 86 86 1 12.5 43 5 35,5
ATP1A1 C>C/CA 1 116926701 1 insertion het yes PASS 300 300 1 12.87 117 13 88,13
C3 G>G/GT 19 6719290 1 insertion het yes PASS 604 604 1 28 82 21 54,21
DDX1 C>C/CA 2 15737558 1 insertion het yes PASS 53 53 1 33.33 17 3 6,3
GATA3 T>T/TG 10 8111513 1 insertion het yes PASS 628 582 1 46.87 39 15 17,15 COSM1474805 carcinoma breast
HIST1H1E C>C/CA 6 26157119 1 insertion het yes PASS 509 509 1 20.88 113 19 72,19
LMAN1 A>A/AT 18 56998333 1 insertion het yes PASS 49 49 1 9.09 69 4 40,4
MALAT1,... GT>GT/G 11 65268479 1 deletion het yes PASS 232 232 1 17.79 325 45 208,45
METTL10 G>G/GA 10 126454015 1 insertion het yes PASS 128 127 1 50 11 5 5,5
MSH3 C>C/CA 5 79970914 1 insertion het yes PASS 35 35 1 50 9 3 3,3
PRPF38B G>G/GA 1 109242267 1 insertion het yes PASS 40 40 1 17.65 23 3 14,3
RAB14 A>A/AT 9 123952937 1 insertion het yes PASS 74 74 1 22.22 23 4 14,4
SMC4 C>C/CA 3 160120548 1 insertion het yes PASS 102 102 1 20.83 31 5 19,5
Variant 7 of 15
Show Population Frequencies
Show Transcript Info
Show Custom Annotations
Show ClinVar
Show Cosmic
Variants
Genes
No-Call Regions
Filter History
Sample: UHR-Access-1-5M.genome.vcf Genes,Variants: (9560, 26308) -> (15, 15)

illumina

Download Analysis | Rename analysis | Move to Trash | View Trash 1

- Analysis Info
- Inputs
- Output Files

Analysis Reports
- Summary
- UHR-Access-1-5M
- Brain-Access-1-5M
- Brain-Access-2-5M
- UHR-Access-2-5M



Save Plot as SVG | Export Data as CSV

Important Files for Download

- Alignments
- Alignment coverage
- Reference FPKM values (genes)
- Reference FPKM values (transcripts)
- Genome VCF
- TopHat fusion output (1 detected fusions)

contact us

Workflow version: 1.1.0

illumina

Candidate fusion list

The following tables show fusion candidates where fusions are grouped based on their genomic locations (table description).

1. chr9-chr22 rr

ABL1	chr9	133729450	BCR	chr22	23632599	6	2	8	392.06

table description

1. Gene on the "left" side of the fusion
2. Chromosome ID on the left
3. Coordinates on the left
4. Gene on the "right" side
5. Chromosome ID on the right
6. Coordinates on the right
7. Number of spanning reads
8. Number of spanning mate pairs
9. Number of spanning mate pairs where one end spans a fusion (reads spanning fusion with only a few bases are included)

If you follow the the 9th column, it shows coordinates "number1:number2" where one end is located at a distance of "number1" bases from the left genomic coordinate of a fusion and "number2" is similarly defined

ABL1	**chr9**	**133729450**	**BCR**	**chr22**	**23632599**	**6**	**2**	**8**	**<-------- <--------**

Left flanking sequence	Right flanking sequence
ACTCAGACCCTGAGGCTCAAAGTCAGATGCTACTGGCCGCTGAAGGGCTT	TTGAACTCTGCTTAAATCCAGTGGCTGAGTGGACGATGACATTCAGAAAC

illumina

Cufflinks Assembly & DE

Illumina, Inc.

App Session Name: Cufflinks Assembly & DE 03/15/2

Save Results To: Select Project(s):

Workshop ×

This app is free.

Continue

TopHat Alignments Selection Criteria

Reference Genome: Homo sapiens/hg19 (RefSeq)

Stranded: ☑

Options

Novel Transcript Assembly: ☑

Control Group

Group Label: UHR

TopHat Alignment App Result(s): Select App Result(s):

Adjust transcript assembly for samples without polyA selection: ☐

Comparison Group

Group Label: Brain

TopHat Alignment App Result(s): Select App Result(s):

contact us

Analysis Info
Inputs
Output Files
Analysis Reports
Cufflinks-Report

Overview

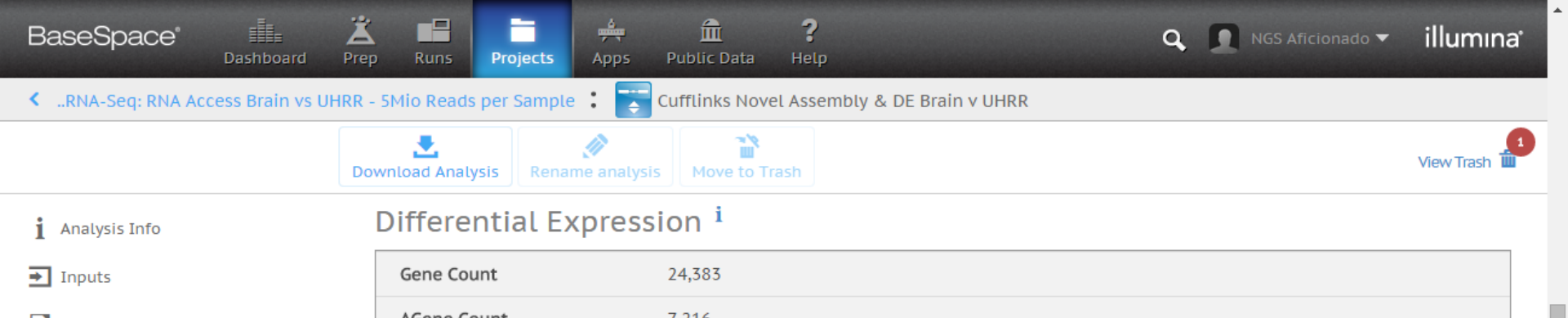
Control samples (UHRR)
- UHR-Access-1-5M
- UHR-Access-2-5M

Comparison samples (Brain)
- Brain-Access-1-5M
- Brain-Access-2-5M

FPKM tables: Genes / Transcripts

Assembly

	Control	Comparison	Merged
Gene Count	24,358	24,388	24,383
Transcript Count	49,045	50,236	52,862
Link to gene models	GTF result	GTF result	GTF result
Relation to reference transcripts			
Equal (=)	43,507	43,504	43,513
Potentially novel (j)	5,469	6,638	9,215
Unknown, intergenic (u)	54	70	99
Overlap with opposite-strand exon (x)	3	16	18
Other	12	8	17



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	test_id	gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
60	XLOC_000059	XLOC_000059	THAP3	chr1:6684924-6761966	UHRR	Brain	OK	5.544	7.417	0.420	0.571	0.45845	0.508114	no
61	XLOC_000060	XLOC_000060	CAMTA1	chr1:6845383-7829766	UHRR	Brain	OK	23.242	35.323	0.604	1.816	0.0282	0.042372	yes
62	XLOC_000061	XLOC_000061	VAMP3	chr1:7831328-7841492	UHRR	Brain	OK	9.196	8.234	-0.159	-0.445	0.5663	0.612817	no
63	XLOC_000062	XLOC_000062	PER3	chr1:7844762-7905237	UHRR	Brain	OK	7.813	28.922	1.888	7.467	5.00E-05	0.00013	yes
64	XLOC_000063	XLOC_000063	PARK7	chr1:8021713-8045342	UHRR	Brain	OK	568.514	517.327	-0.136	-0.668	0.3358	0.383859	no
65	XLOC_000064	XLOC_000064	SLC45A1	chr1:8384389-8404227	UHRR	Brain	OK	0.296	10.106	5.093	3.418	0.11235	0.147374	no
66	XLOC_000065	XLOC_000065	ENO1-AS1	chr1:8921058-8939943	UHRR	Brain	NOTEST	0.000	0.000	0.000	0.000	1	1	no
67	XLOC_000066	XLOC_000066	CA6	chr1:9005892-9035148	UHRR	Brain	NOTEST	0.000	0.000	0.000	0.000	1	1	no
68	XLOC_000067	XLOC_000067	H6PD	chr1:9294862-9331394	UHRR	Brain	OK	7.293	3.533	-1.046	-3.604	5.00E-05	0.00013	yes

UHRR_vs_Brain.gene_exp

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	test_id	gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
134	TCONS_00000133	XLOC_000062	PER3	chr1:7844762-7905237	UHRR	Brain	OK	5.213	7.826	0.586	0.711	0.3021	0.372398	no
135	TCONS_00000134	XLOC_000062	PER3	chr1:7844762-7905237	UHRR	Brain	OK	0.805	5.151	2.678	1.199	0.3896	0.455824	no
136	TCONS_00000135	XLOC_000062	PER3	chr1:7844762-7905237	UHRR	Brain	OK	0.933	9.629	3.367	2.073	0.17475	0.276365	no
137	TCONS_00000136	XLOC_000063	PARK7	chr1:8021713-8045342	UHRR	Brain	OK	456.931	288.738	-0.662	-2.505	0.00045	0.001868	yes
138	TCONS_00000137	XLOC_000063	PARK7	chr1:8021713-8045342	UHRR	Brain	OK	111.583	228.589	1.035	2.260	0.00635	0.019173	yes
139	TCONS_00000138	XLOC_000064	SLC45A1	chr1:8384389-8404227	UHRR	Brain	OK	0.296	10.106	5.093	3.418	0.11235	0.21789	no
140	TCONS_00000139	XLOC_000065	ENO1-AS1	chr1:8921058-8939943	UHRR	Brain	NOTEST	0.000	0.000	0.000	0.000	1	1	no
141	TCONS_00000140	XLOC_000066	CA6	chr1:9005892-9035148	UHRR	Brain	NOTEST	0.000	0.000	0.000	0.000	1	1	no
142	TCONS_00000141	XLOC_000066	CA6	chr1:9005892-9035148	UHRR	Brain	NOTEST	0.000	0.000	0.000	0.000	1	1	no

UHRR_vs_Brain.isoform_exp (1)

Differential Expression Gene Browser

- Analysis Info
- Inputs
- Output Files
- Analysis Reports
  - Cufflinks-Report

Filters

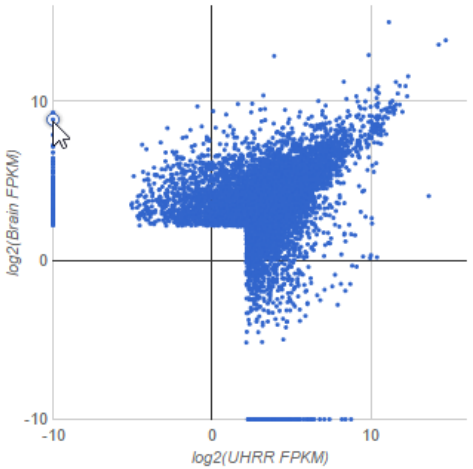
|log2(ratio)|

0.0 — 42.0

Significant: Choose a value...

Status: OK

Gene:



Save Plot as SVG

contact us

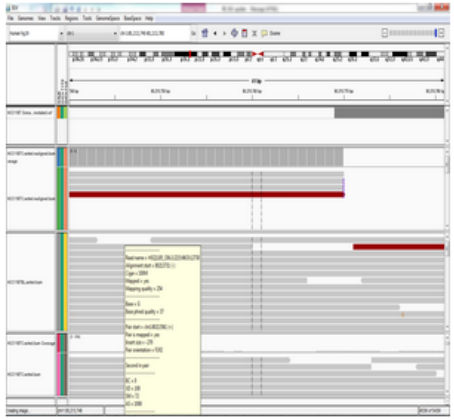
Test ID	Gene	Locus	Status	log 2 (UHRR FPKM)	log 2 (Brain FPKM)	log 2 (Ratio)	q Value	Significant
XLOC_003980	PHLDA2	chr11:2949502-2950650	OK	3.33	-10.00	-13.33	0.000	✓
XLOC_003981	NAP1L4	chr11:2965659-3013607	OK	5.96	5.71	-0.25	0.148	✗
XLOC_003983	CARS	chr11:3022151-3078681	OK	4.60	4.15	-0.46	0.030	✓
XLOC_003987	ZNF195	chr11:3379156-3400452	OK	3.26	2.50	-0.76	0.005	✓
XLOC_003991	NUP98	chr11:3696239-3847601	OK	5.48	4.50	-0.98	0.000	✓
XLOC_003992	RHOG	chr11:3848207-3862213	OK	3.11	2.17	-0.94	0.043	✓
XLOC_004009	HBB	chr11:5246695-5248301	OK	-10.00	8.85	18.85	0.000	✓
XLOC_004012	HBG1,HBG2	chr11:5269501-5276066	OK	9.96	0.15	-9.80	0.035	✓
XLOC_004013	HBE1	chr11:5289579-5291373	OK	8.76	-10.00	-18.76	0.000	✓
XLOC_004031	FAM160A2	chr11:6232563-6255941	OK	2.12	2.41	0.29	0.354	✗
XLOC_004033	APBB1	chr11:6416354-6440644	OK	2.24	7.09	4.85	0.000	✓

Integrative Genomics Viewer (IGV)

Broad Institute of MIT and Harvard

The Integrative Genomics Viewer (IGV) app is a powerful genome browser that displays next-generation sequencing data. It displays alignments and variants from multiple samples for performing complex variant analysis. The Broad Institute of MIT and Harvard developed IGV, and Illumina modified it to display BaseSpace data.

Your download should begin automatically. If the desktop app doesn't open automatically, you may need to manually open the ".jnlp" file that your browser downloaded.



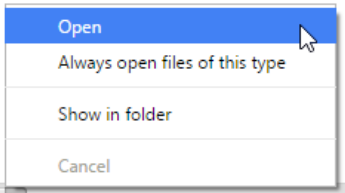
First time downloading?

1. Install or Upgrade Java

- Java version 7 or higher is required.

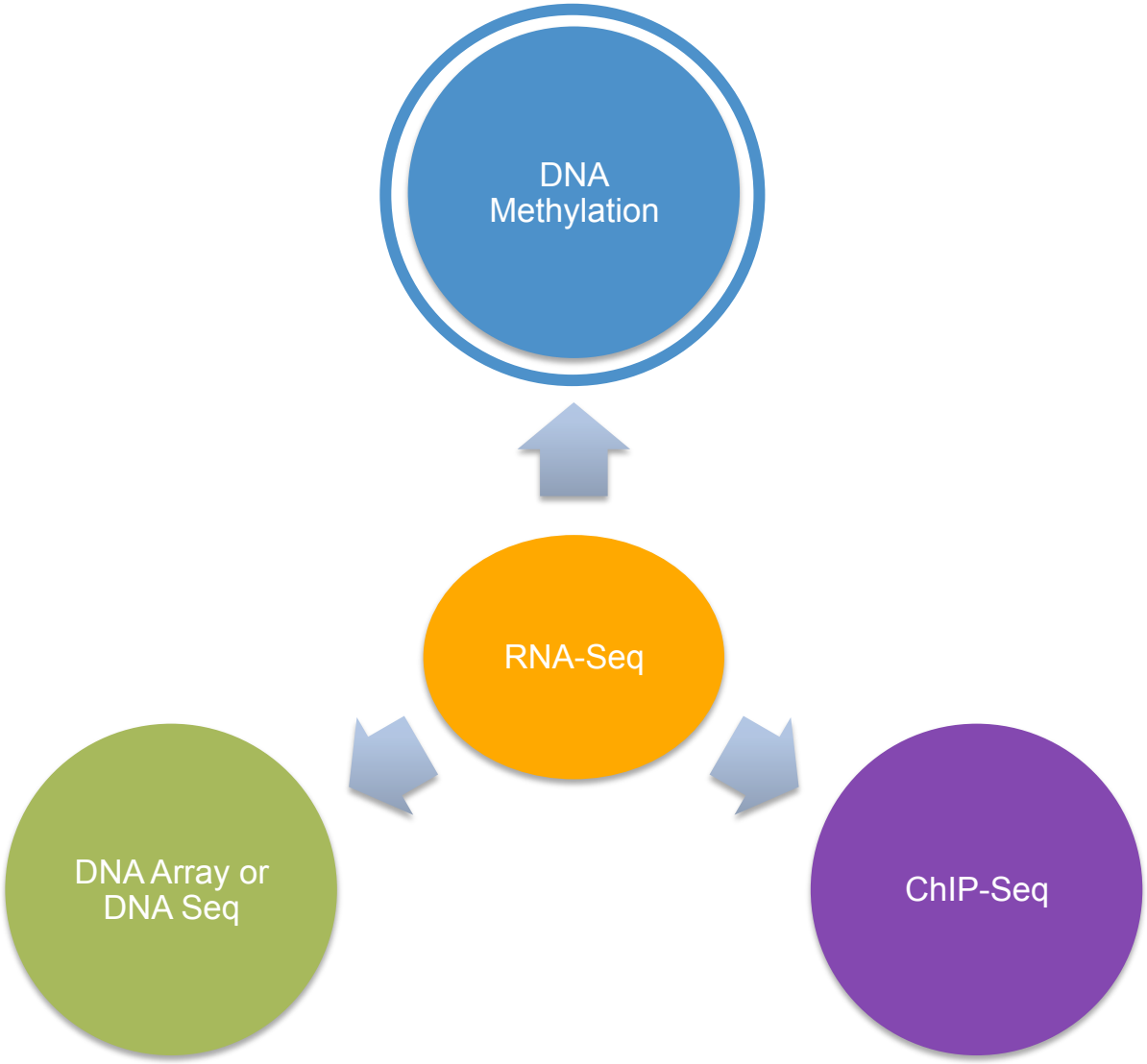
2. Launch IGV

- If the download doesn't start automatically, click the **Launch IGV** button above. This will download a jnlp file, which you may need to manually open.



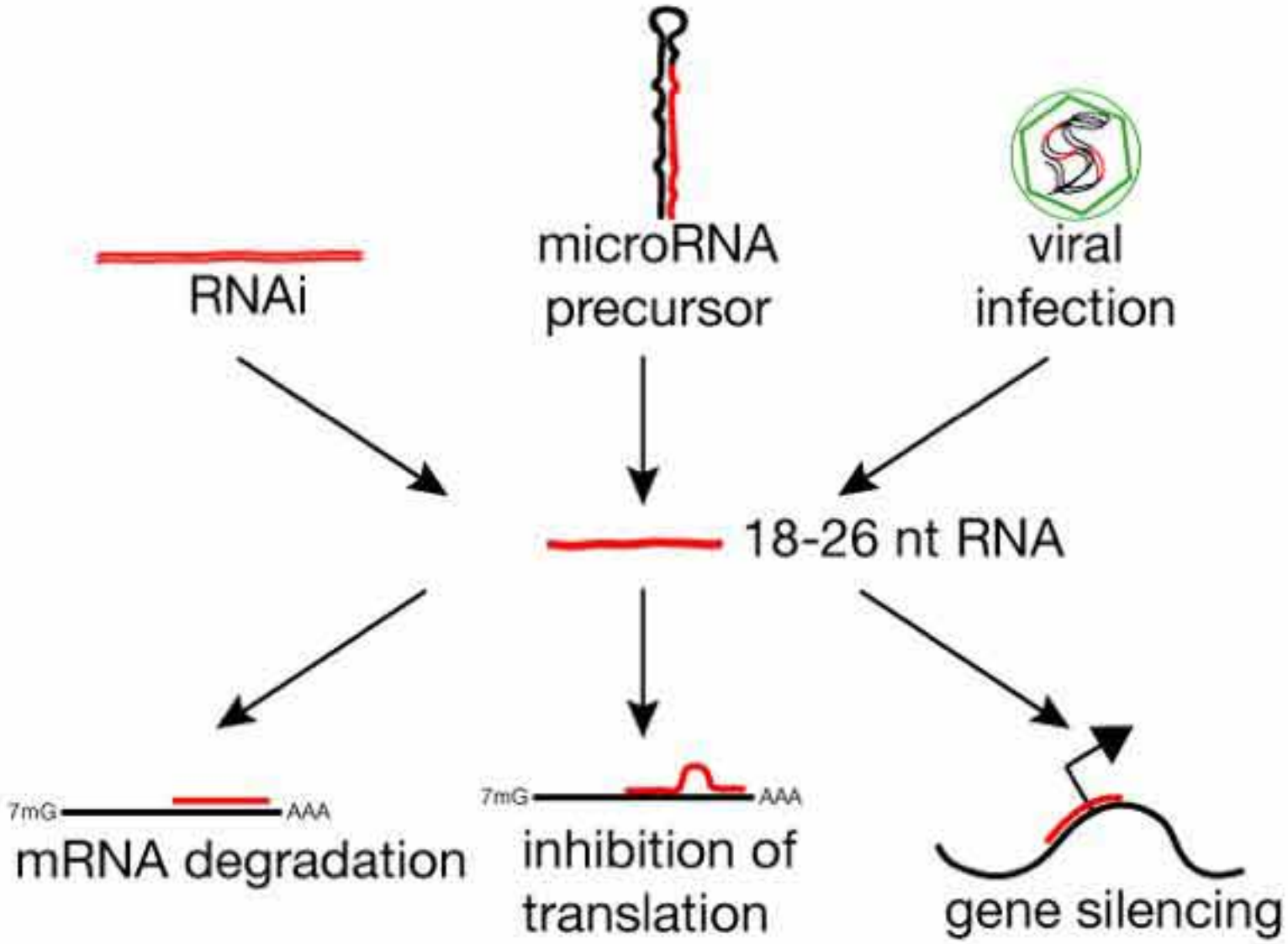
MultiOmics

Taking RNA-Seq to the Next Step

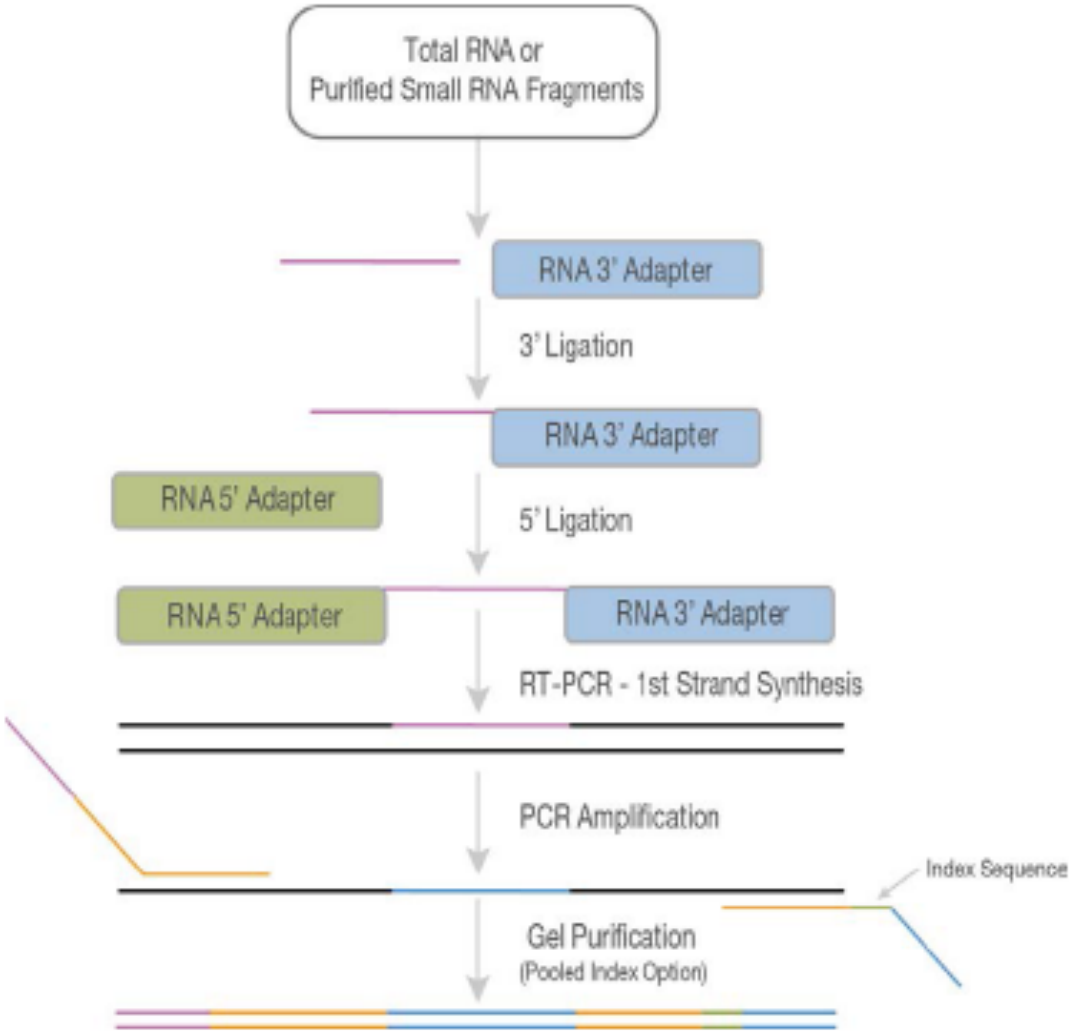


illumina®

Short RNAs are key players of gene regulation



TruSeq small RNA



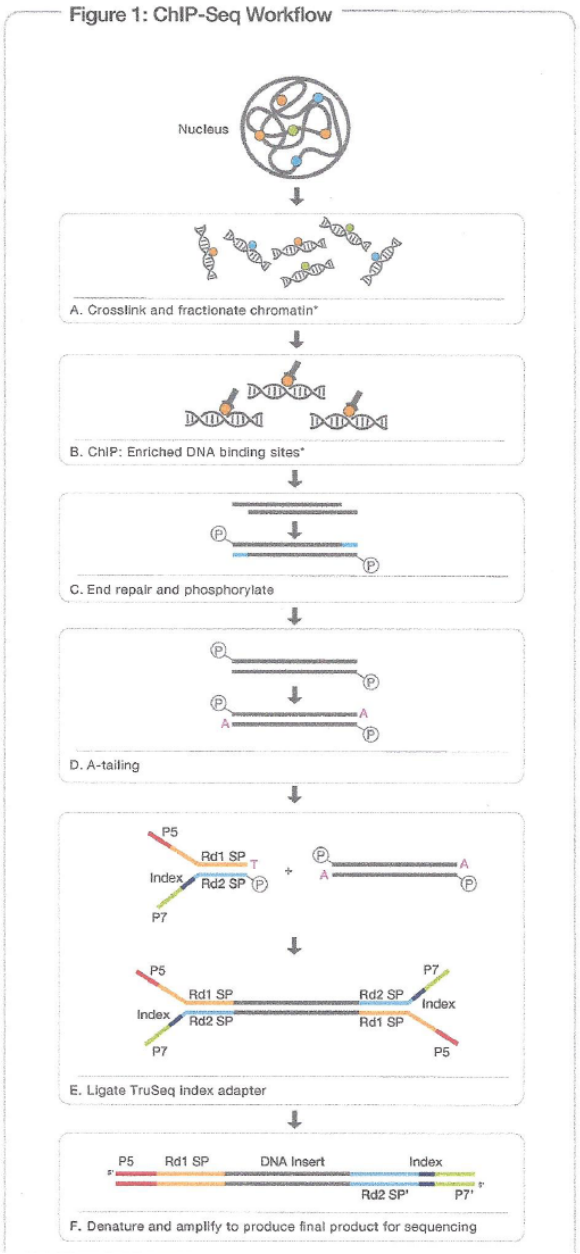
illumina®

TruSeq small RNA

App	Description	Details	Cons
 Small RNA Illumina, Inc.	Alignment to abundant, mature miRNA, other RNA and genomic. Outputs mature miRNAs, isomiRs and piRNAs. Optional novel precursor discovery and pairwise differential expression analysis.	• miRBase 21 supported for human, mouse & rat • 100 samples/200Gb limit • 25Gb/sample limit	• Adapter trimming not included • No target gene info available • Not compatible with Functional Impact or Pathway Analysis apps
 miRNAs Analysis B&Gu @ University of Torino	Differential miRNA expression analysis between 2 conditions	• Includes adapter trimming • miRBase 21 supported for human and mouse and optional download of latest miRBase DB • Choice of Outlier replacement from DESeq2	• Rat genome not supported • No miRNA precursor discovery • Not compatible with Functional Impact or Pathway Analysis apps

illumina®

TruSeq ChIP



illumina®

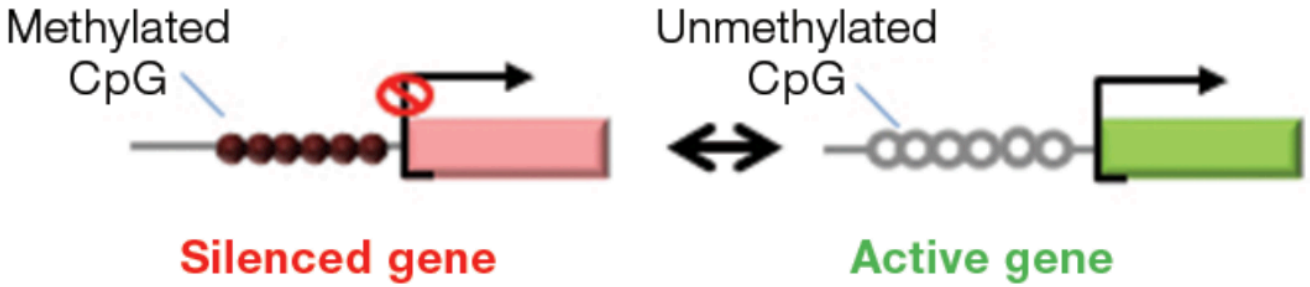
ChIP-Seq

ChIP-Seq for Protein-DNA interaction

App	Description	Details	Cons
 ChIPSeq BaseSpace Labs	Enriched region identification of Chromosome IP (**ChIP-Seq**) experiments and motive discovery.	• Alignment & Peak Calling: MACS • Peak annotation and Motive Enrichment: HOMER • Pairwise analysis recommended • ChIP vs Input and ChIP vs ChIP comparisons supported • Single ChIP analysis (no control) possible as well	• Identifies "narrow" peaks (i.e transcription factor binding sites), not "broad" peaks (i.e histone modifications). • hg19 only reference

How does DNA Methylation affect gene expression?

- Chemical modification to our DNA (typically cytosine) that compacts chromatin and effects gene expression.



illumina®

Why Is Methylation Important?

- **Methylation affects it all:** Cancer, development, Alzheimer's, aging, ADHD, obesity, diabetes, addiction, infection…

New Approaches for Breast Tumor Diagnostics: Epigenetic Profiles

Posted: 02/26/2015 9:45 am EST | Updated: 02/26/2015 9:59 am EST

Huffington Post April 2015

LiveScience January 2015

How Genes and Environment Conspire to Trigger Diabetes

Endurance training alters skeletal muscle 'at an epigenetic level'

WhatIsEpigenetics.com January 2015

Twins Data Reshaping Nature Versus Nurture Debate

NPR.org January 2012

Why Is Studying Methylation Important?

- **Methylation can be changed = actionable!**
  - Medicine can alter methylation
  - Exercise, your environment and actions can alter methylation
  - Methylation changes before DNA in tumors, giving us earlier warnings

Epigenetic Therapy of Cancer With 5-Aza-2′-Deoxycytidine (decitabine)

Momparler Seminars in oncology 2005

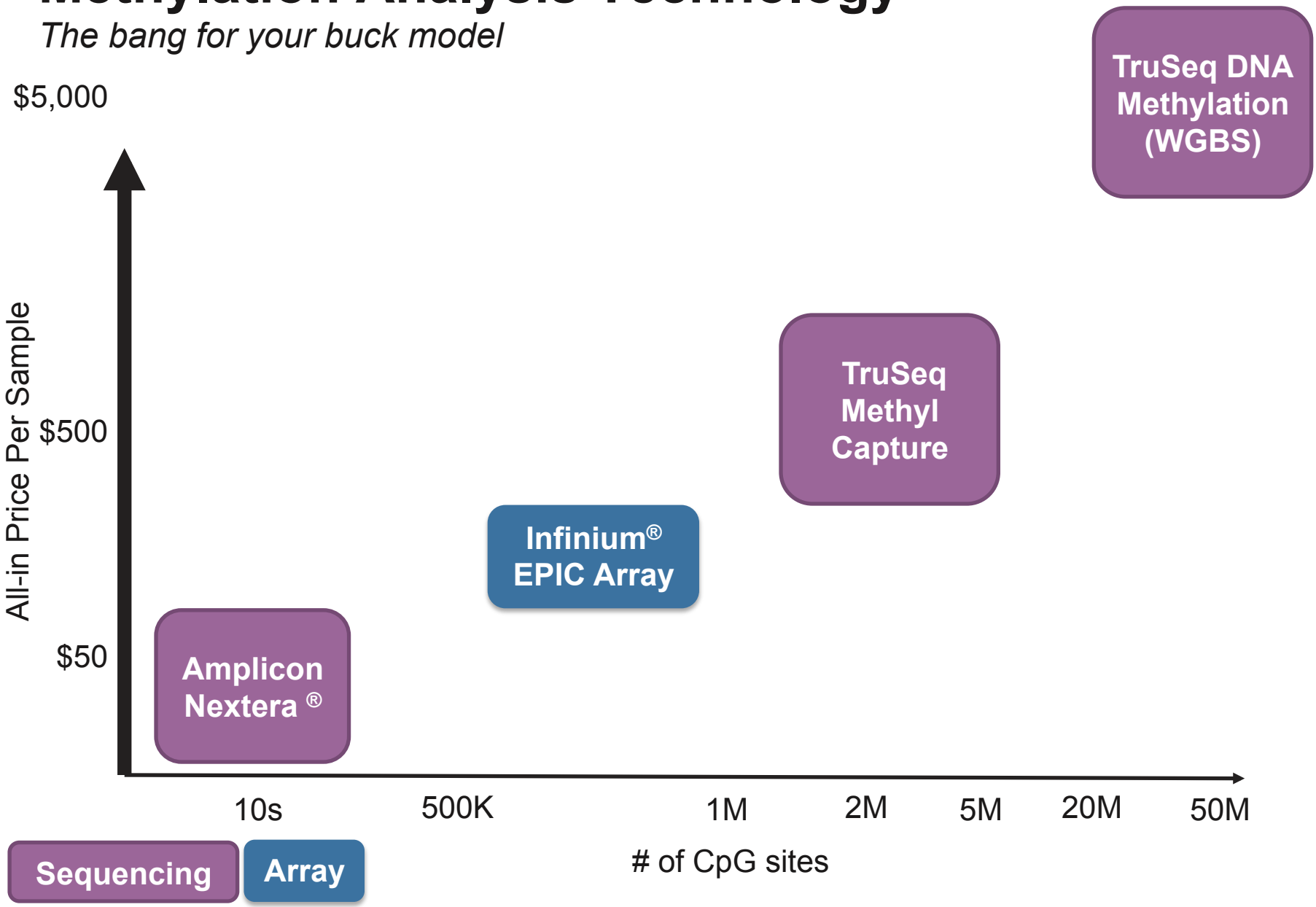
Effects of the Social Environment and Stress on Glucocorticoid Receptor Gene Methylation: A Systematic Review

Turecki et al., Biological psychiatry 2016

illumina®

Methylation Analysis Technology

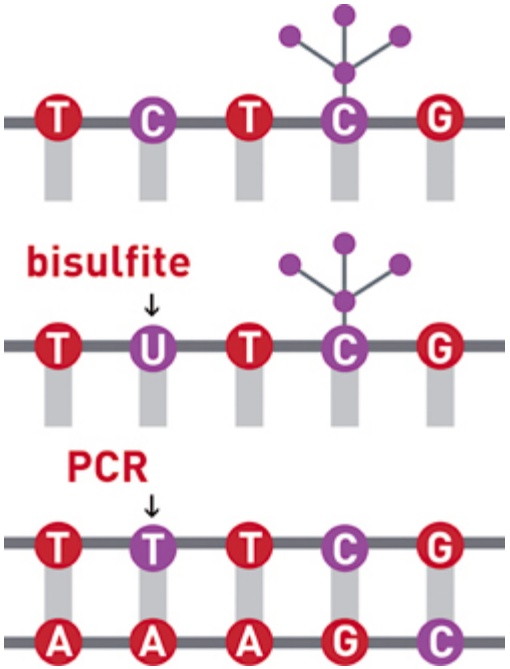
The bang for your buck model



illumina®

How does this technology work?

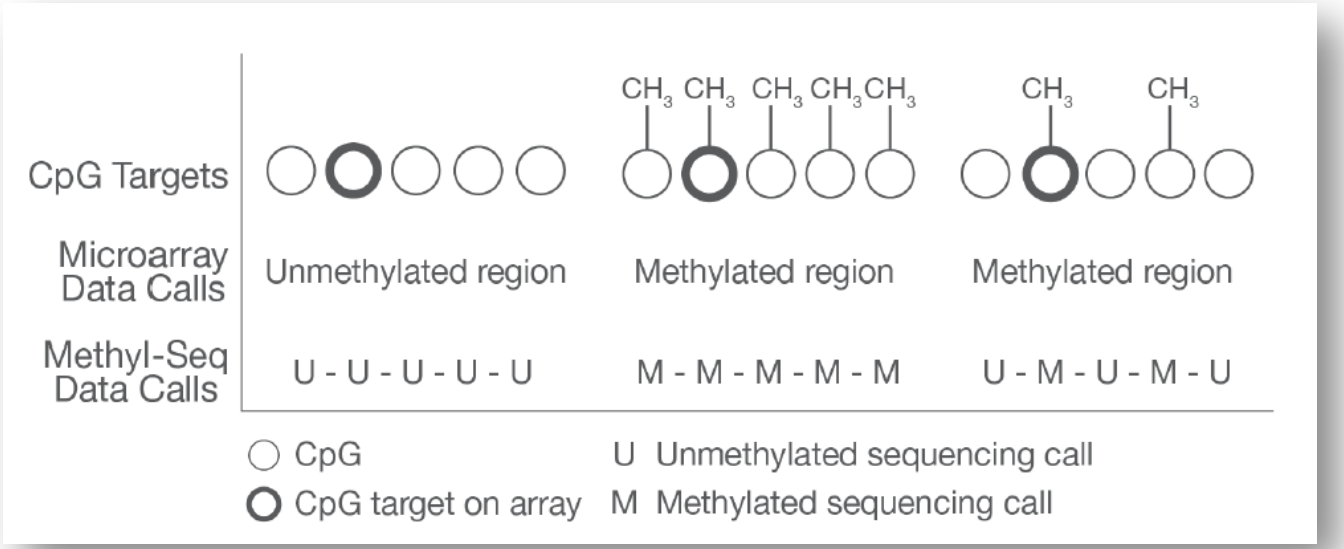
Methylation Analysis With Bisulfite Conversion



Methylation Arrays:
Identify the difference between a C and A at **~850,000 discreet sites of interest**

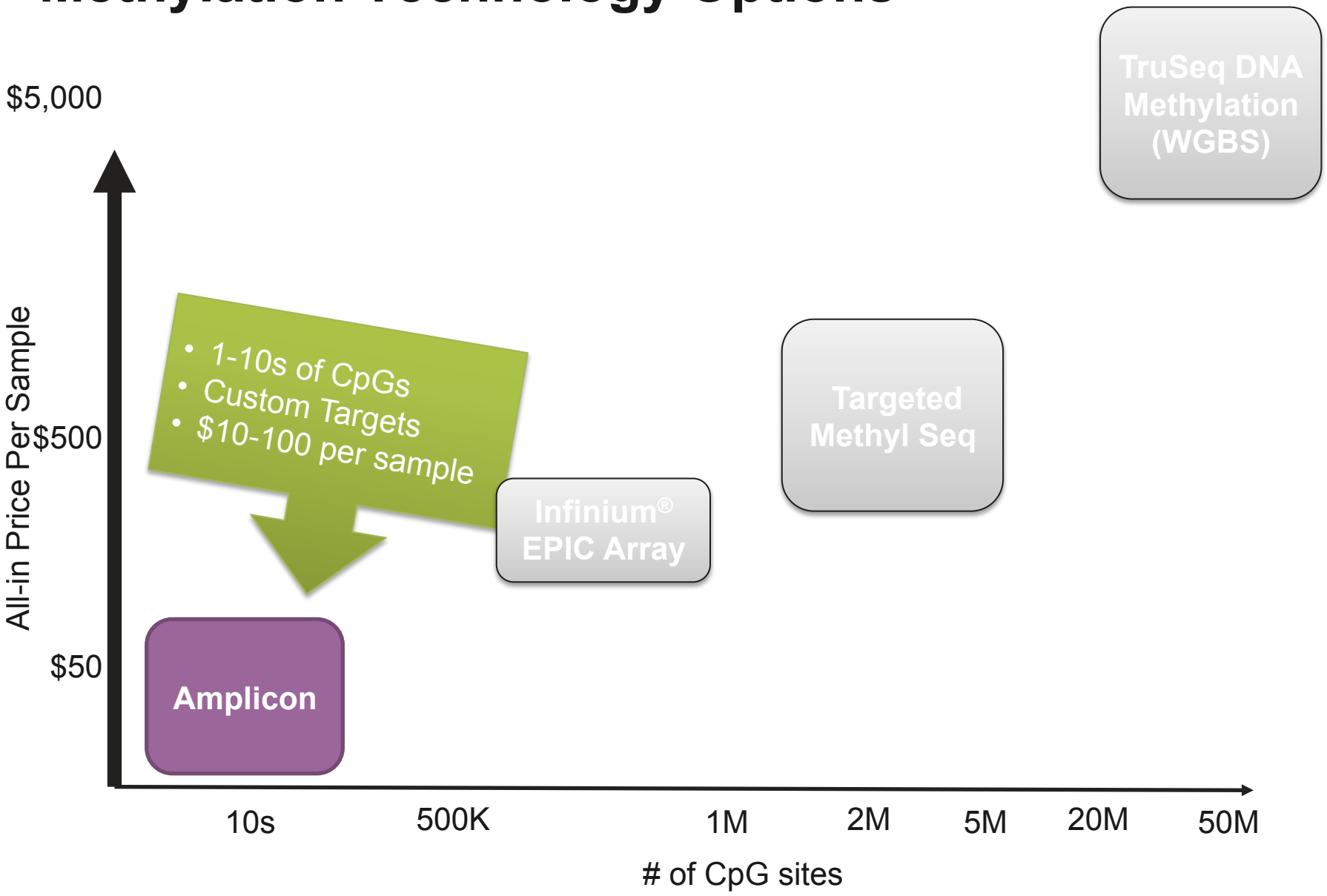
Methylation Sequencing:
Identify the difference between a C and A in **targeted regions up to 38M+ CpG sites in complete regions**

Methylation Arrays or Sequencing



- **Arrays are cost effective** for large scale screens
- **Sequencing provides deep information** across CpG rich regions and can call SNPs, indels within the region covered

Methylation Technology Options

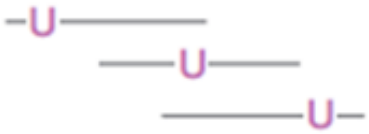


Amplicon Bisulfite Sequencing Workflow

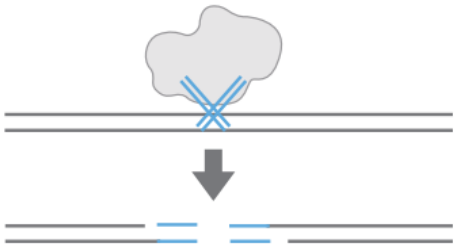
Prepare Libraries
Nextera XT

Sequence
MiniSeq™ or MiSeq®

Analyze
BaseSpace®

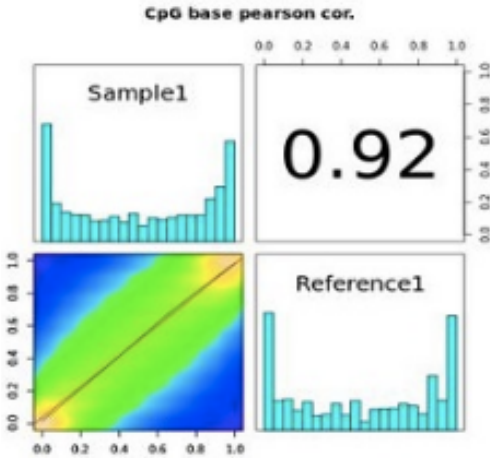


PCR Amplify converted DNA
(custom primers)



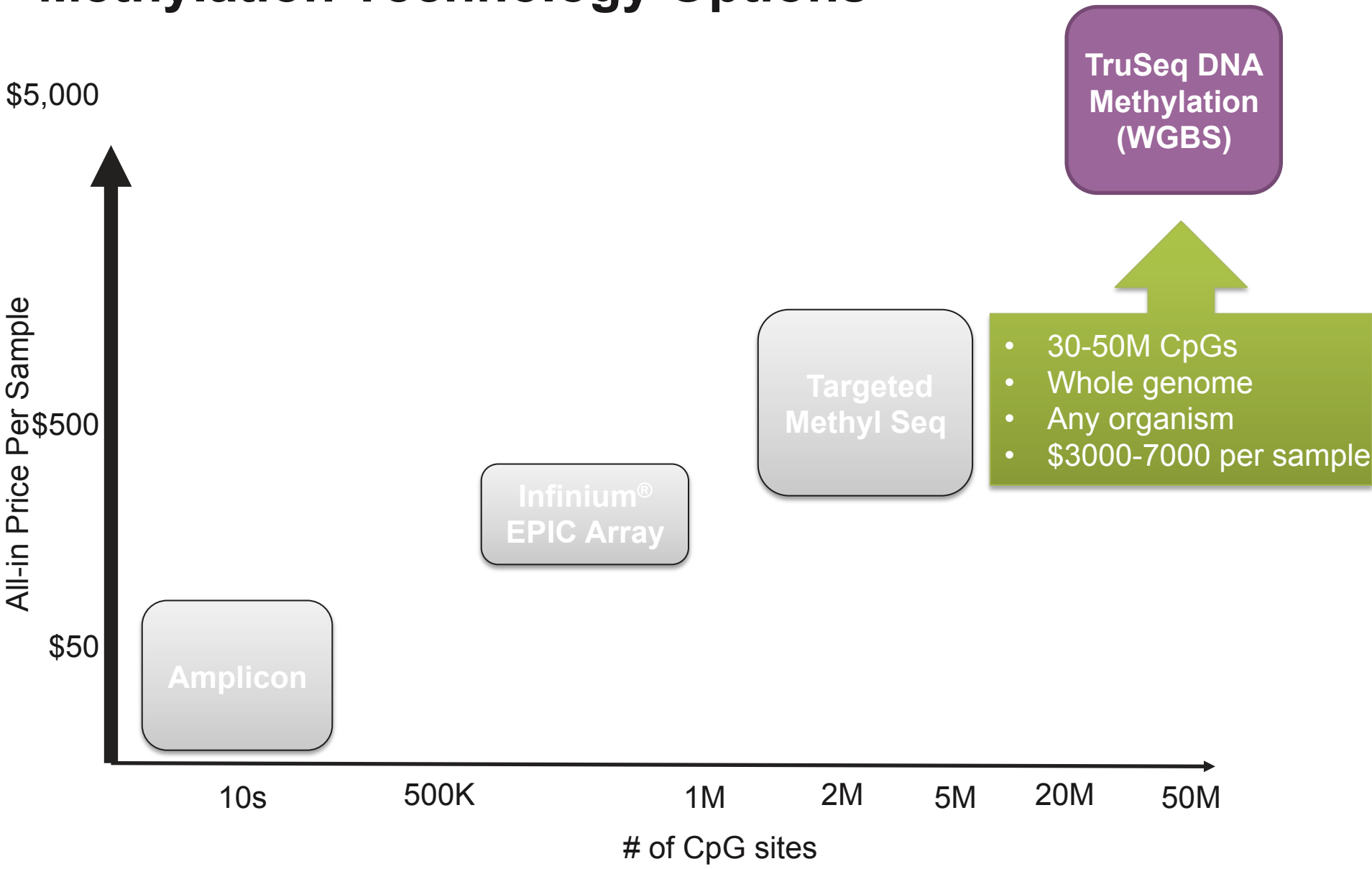
Nextera XT Library Prep

- 1-10s of CpGs
- Custom Targets
- $10-100 per sample



illumina®

Methylation Technology Options

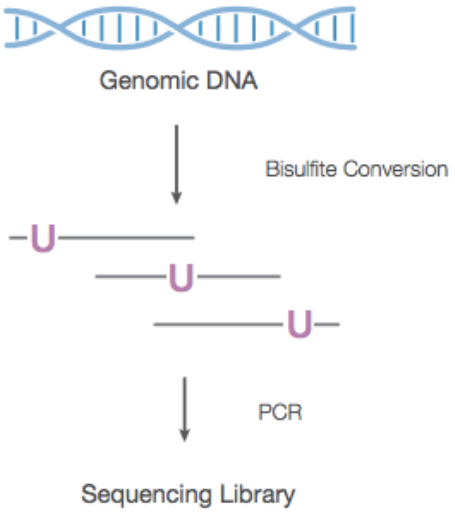


Whole Genome Bisulfite Sequencing Workflow

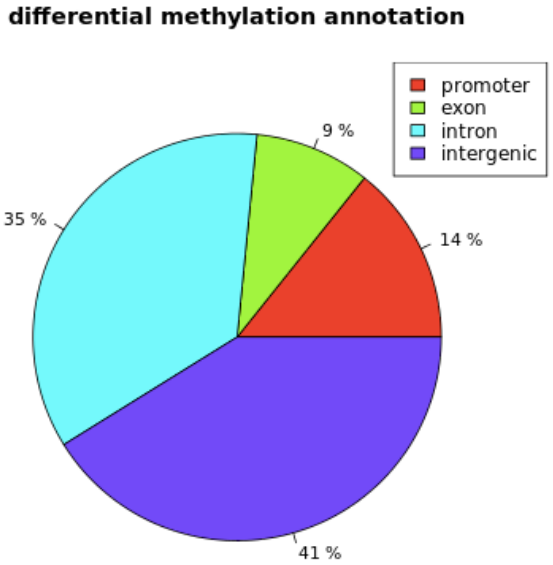
Prepare Libraries
TruSeq DNA Methylation

Sequence
NextSeq or HiSeq

Analyze
BaseSpace

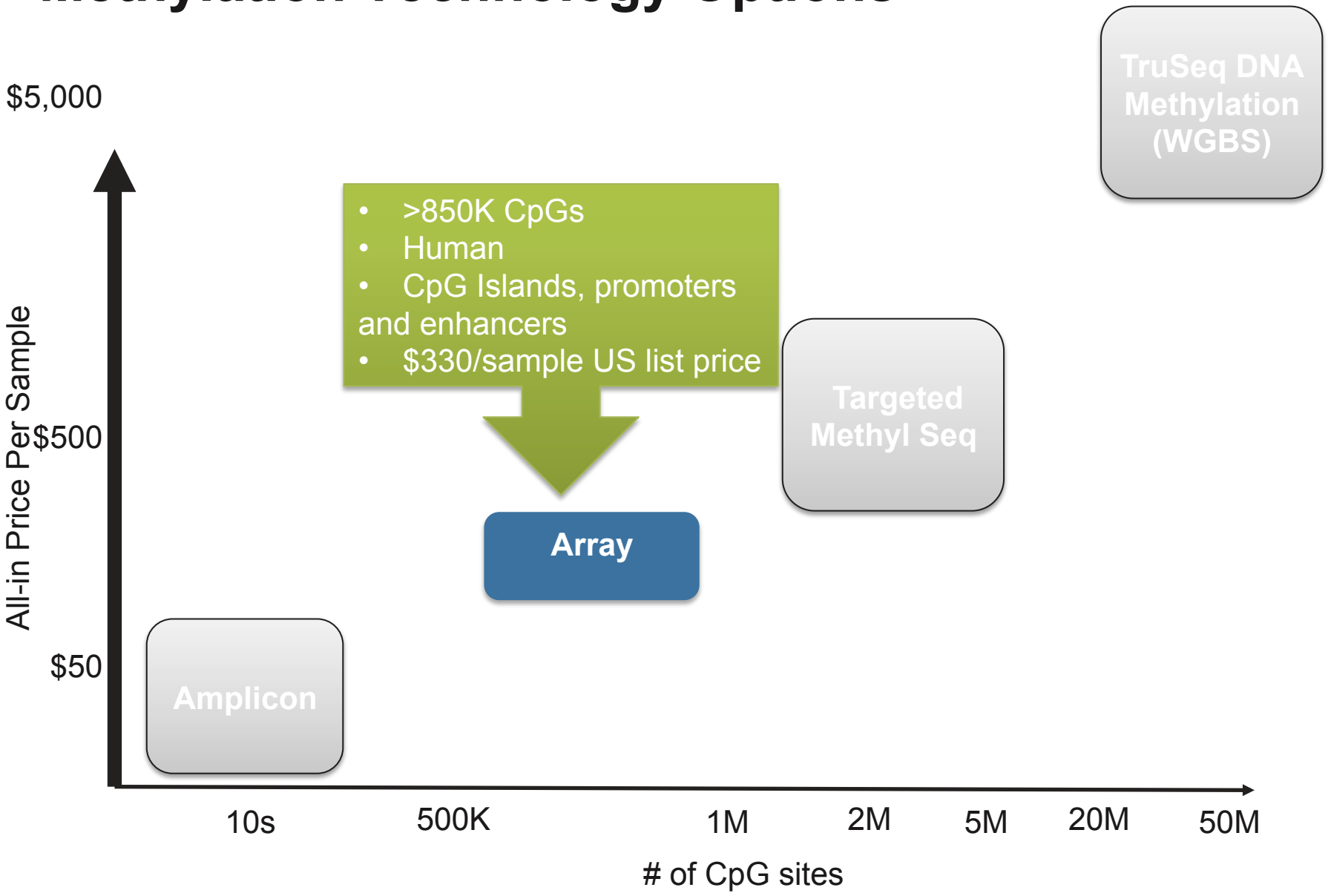


methylKit



illumina®

Methylation Technology Options



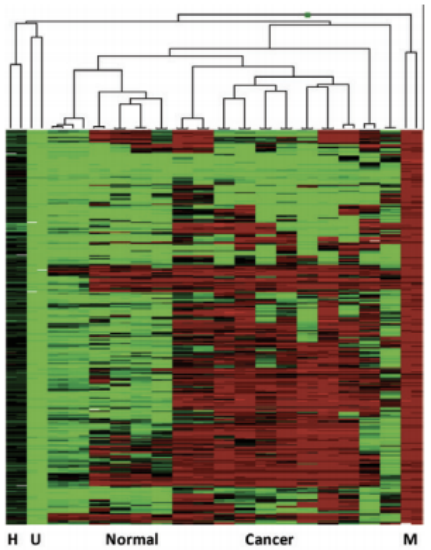
illumina®

Methylation Array Workflow

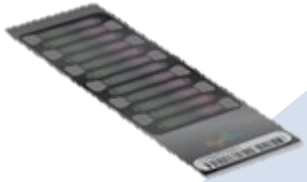
Prepare Samples
Infinium® MethylationEPIC

Scan
iScan® or HiScan®

Analyze
GenomeStudio and BioConductor Apps



Infinium® Methylation Arrays



2007: 27K Array

- First Infinium Methylation array

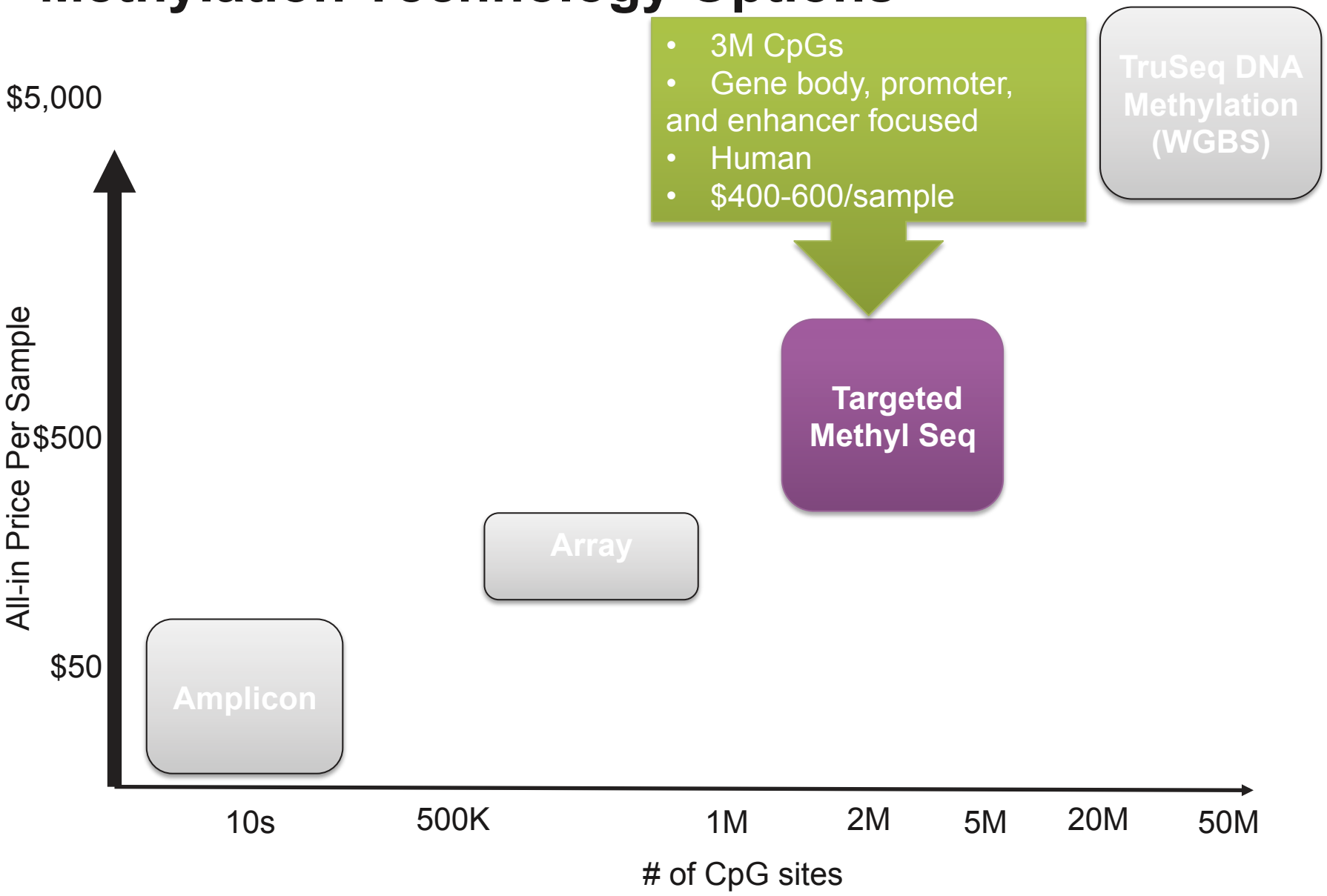
2010: 450K Array

- New chemistry
- 1000+ publications
- Focused on gene bodies and promoters
- EWAS studies

2015: EPIC Array

- Same trusted chemistry, refreshed content
- Enhancer regions (as identified by ENCODE, FANTOM5) added
- >90% backwards compatible with 450K content
- 98%+ R^2 for samples run on each array side by side

Methylation Technology Options

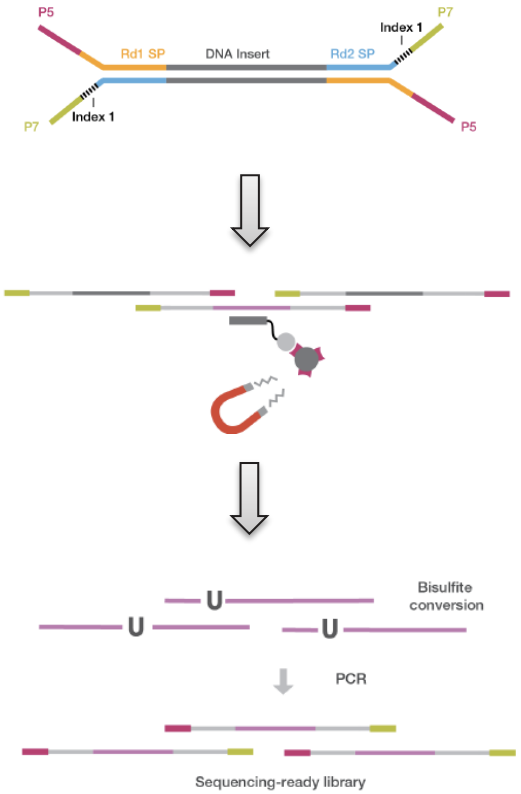


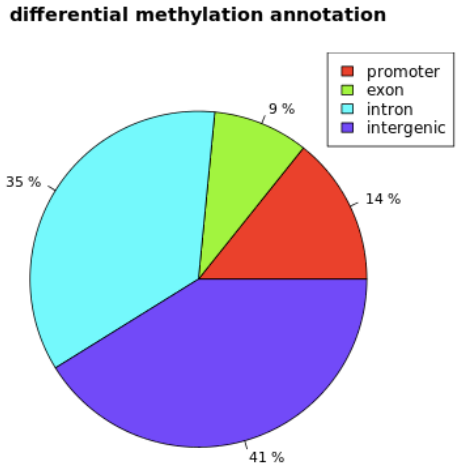
TruSeq Methyl Capture Workflow

Prepare Libraries
TruSeq Methyl Capture

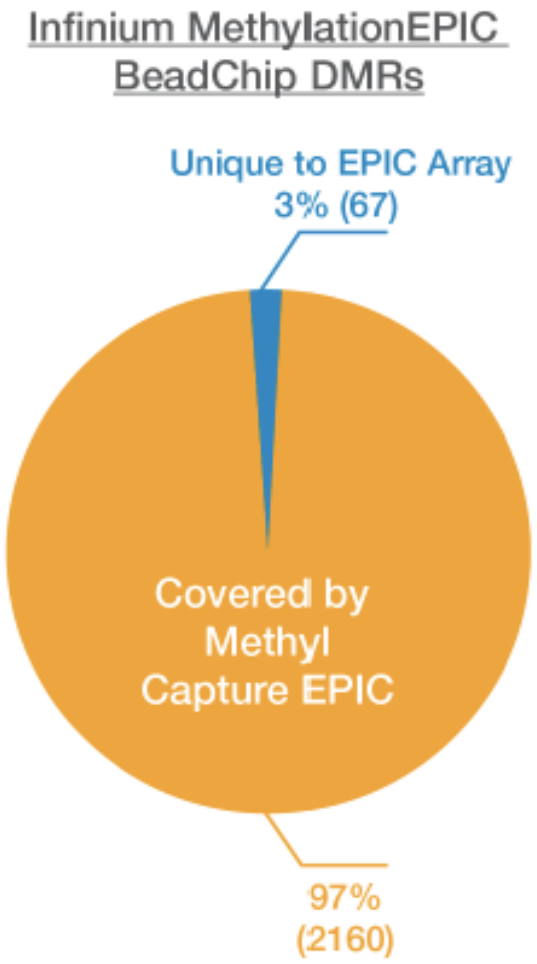
Sequence NextSeq® or HiSeq®

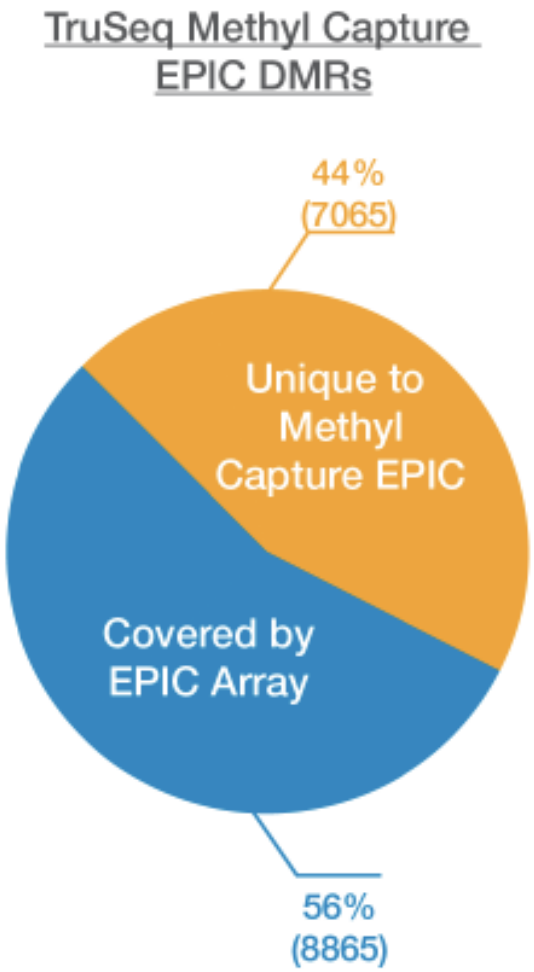
Analyze
BaseSpace





Sequencing Content Compared to Array



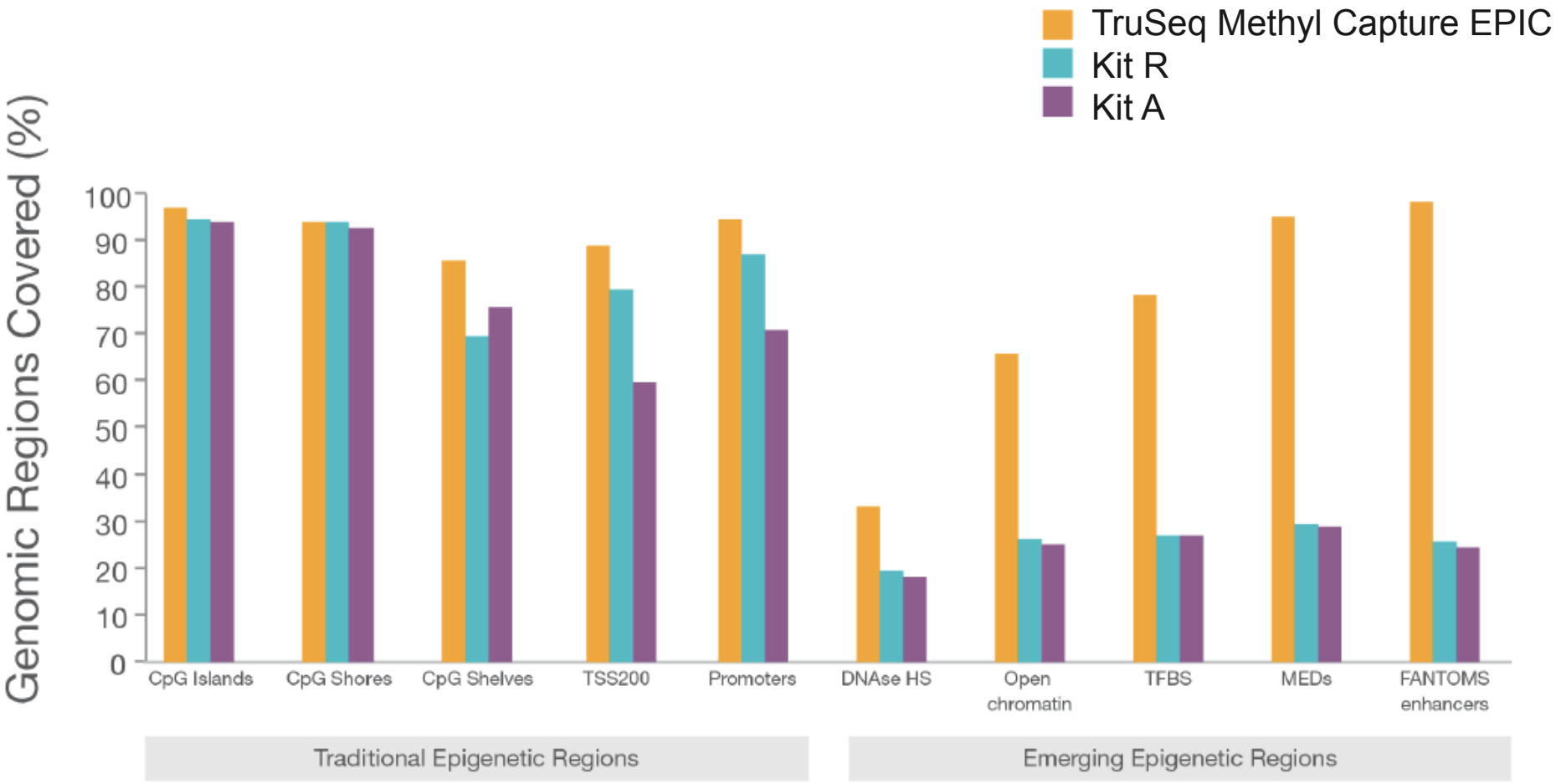


DMR
Differentially Methylated Region

TruSeq Methyl Capture Content

	Genome	TruSeq Methyl Capture, EPIC	Kit R	Kit A
Total Size	3000 Mb	107 Mb	84 Mb	81 Mb
CpG Sites	28.22M	3.34M	3.15M	2.80M
Differentially Methylated Regions from WGBS	1.9M	345K	240K	227K
Transcription Factor Binding Sites	229K	180K	62K	62K
FANTOM5 Enhancers	28.5K	28.2K	56.0K	53.7K

TruSeq Methyl Capture Content



Sequencing a TruSeq Methyl Capture Library

	Sequencing System	Samples / Run
NextSeq®	Mid-Output Flow Cell	2
	High-Output Flow Cell	8
HiSeq® 2500	Rapid Run Mode, Dual Flow Cell	10
	High-Output Mode, Dual Flow Cell	72
HiSeq® 3000	Single Flow Cell	45
HiSeq® 4000	Dual Flow Cell	90

Methylation Sequencing Analysis Workflow

MethylSeq

Illumina, Inc.

MethylKit

BaseSpace Labs

Align Data

- Map sequencing reads to a reference genome
- Define that reference for targeted sequencing (manifest)
- Bisulfite conversion specific aligner like BISMARK

Call CpG Methylation

- Based on bisulfite conversion rate
- Calculated as % methylated
- Also uses BISMARK

Compare Samples

- Differential methylation between samples with MethylKit
- Tumor/Normal, Treated/untreated, etc

illumina®

MethylSeq v1.0.1

Illumina, Inc.

Analysis Name: MethylSeq 07/29/2016 10:39:08

Save Results To: Select Project(s): MethylTest

Sample: Select Sample(s): NA18507-4M

Library prep kit (human only) is directional (2 strands) or non-directional (4 strands): directional / non-directional

▶ Advanced

illumina®

Analyses

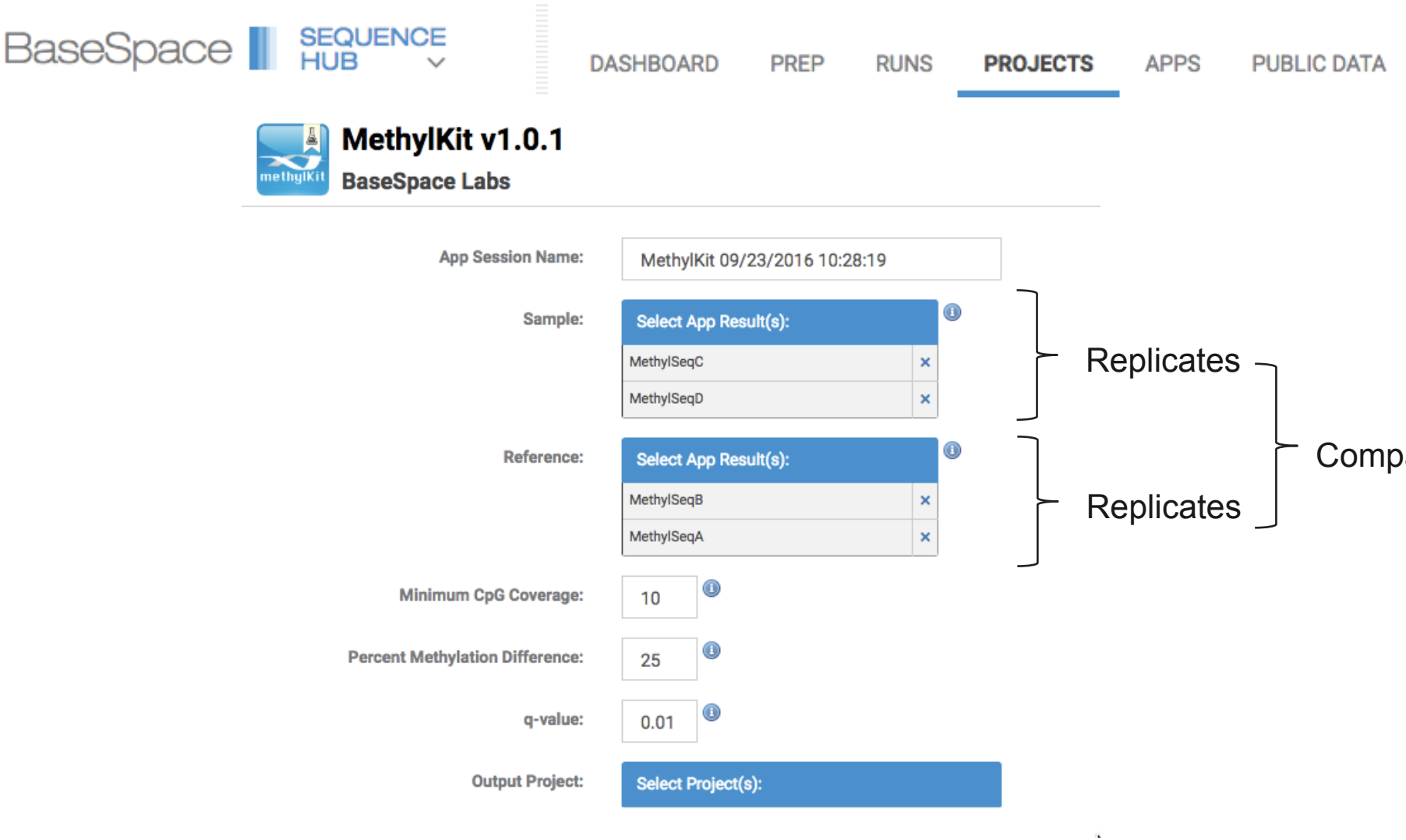
Showing 4 of 4

NAME	LAST MODIFIED	APPLICATION
MethylSeqv2 Beta 07/27/2016 1:52:15	Jul 28, 2016	MethylSeqv2 Beta
MethylKit 06/14/2016 9:17:25- WG	Jun 14, 2016	MethylKit
MethylKit 06/14/2016 10:00:13- Targeted	Jun 14, 2016	MethylKit
MethylSeqv2 06/13/2016 9:48:51	Jun 14, 2016	MethylSeqv2

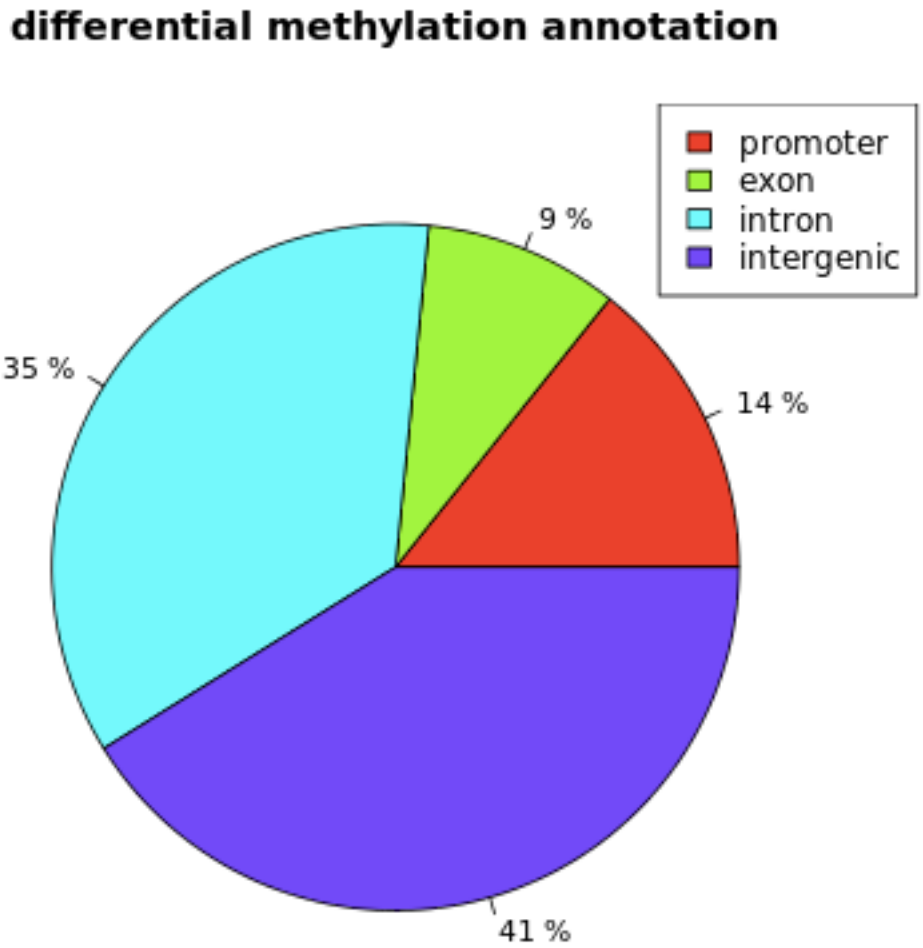
METHYLATION SUMMARY

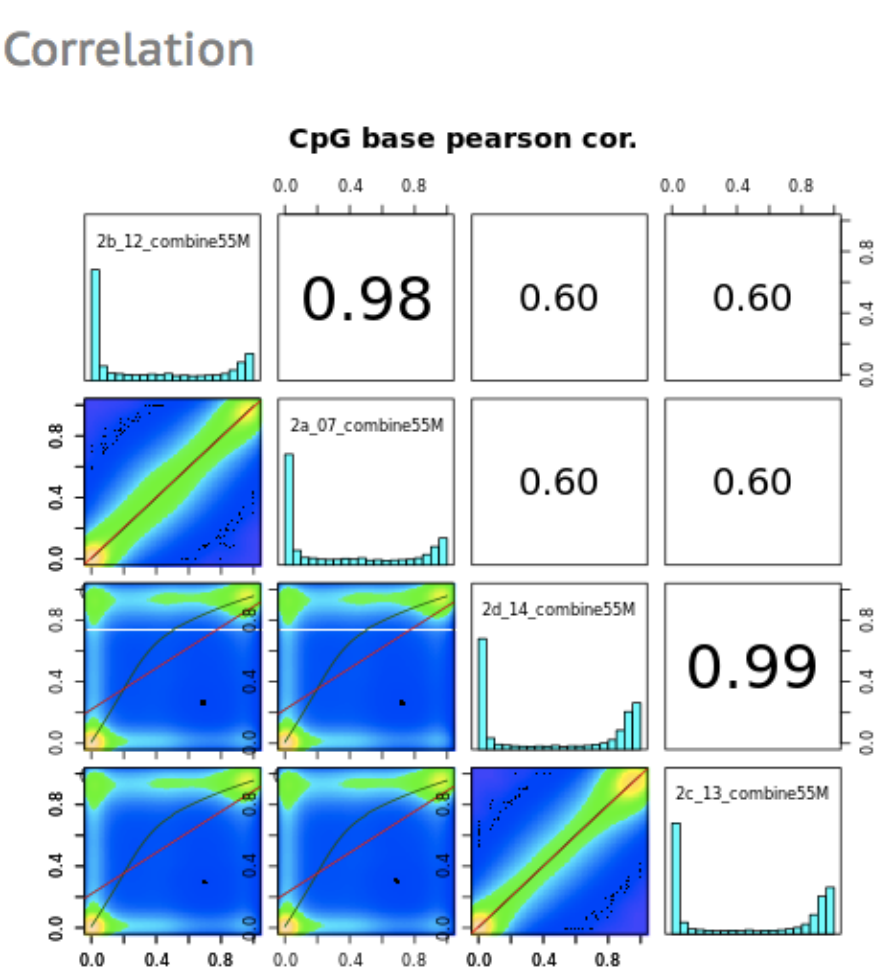
Category	C's in CpG	C's in CHG	C's in CHH
Methylated	89665270	1446786	3517887
Unmethylated	126210732	516721145	1455240390

Kicking off MethylKit analysis



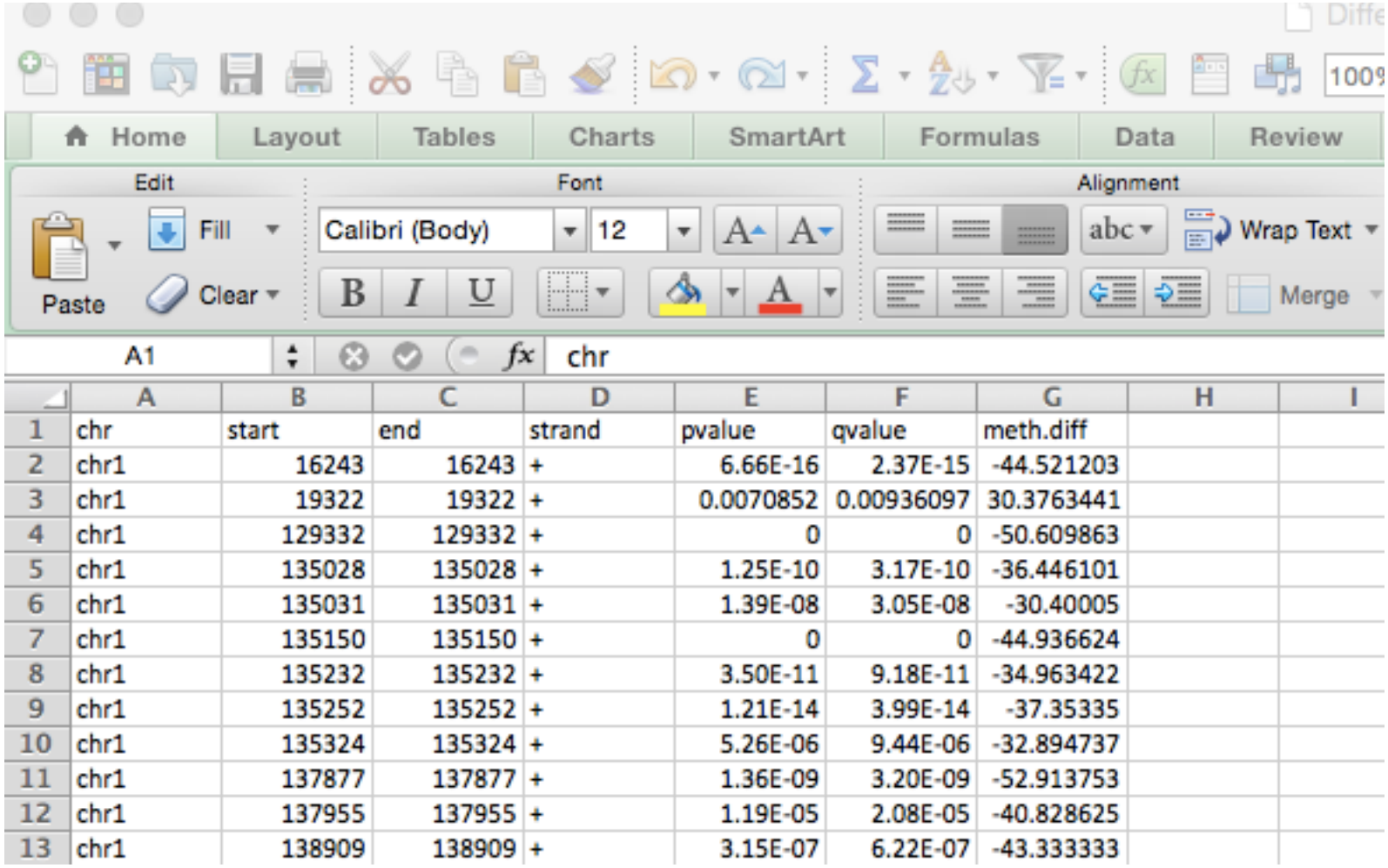
MethylKit comparing samples





illumina®

Comparing samples – Output Files



	A	B	C	D	E	F	G	H	I
1	chr	start	end	strand	pvalue	qvalue	meth.diff		
2	chr1	16243	16243	+	6.66E-16	2.37E-15	-44.521203		
3	chr1	19322	19322	+	0.0070852	0.00936097	30.3763441		
4	chr1	129332	129332	+	0	0	-50.609863		
5	chr1	135028	135028	+	1.25E-10	3.17E-10	-36.446101		
6	chr1	135031	135031	+	1.39E-08	3.05E-08	-30.40005		
7	chr1	135150	135150	+	0	0	-44.936624		
8	chr1	135232	135232	+	3.50E-11	9.18E-11	-34.963422		
9	chr1	135252	135252	+	1.21E-14	3.99E-14	-37.35335		
10	chr1	135324	135324	+	5.26E-06	9.44E-06	-32.894737		
11	chr1	137877	137877	+	1.36E-09	3.20E-09	-52.913753		
12	chr1	137955	137955	+	1.19E-05	2.08E-05	-40.828625		
13	chr1	138909	138909	+	3.15E-07	6.22E-07	-43.333333		

illumina®

Correlating Methylation with RNA-Seq Data

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

MethylMix

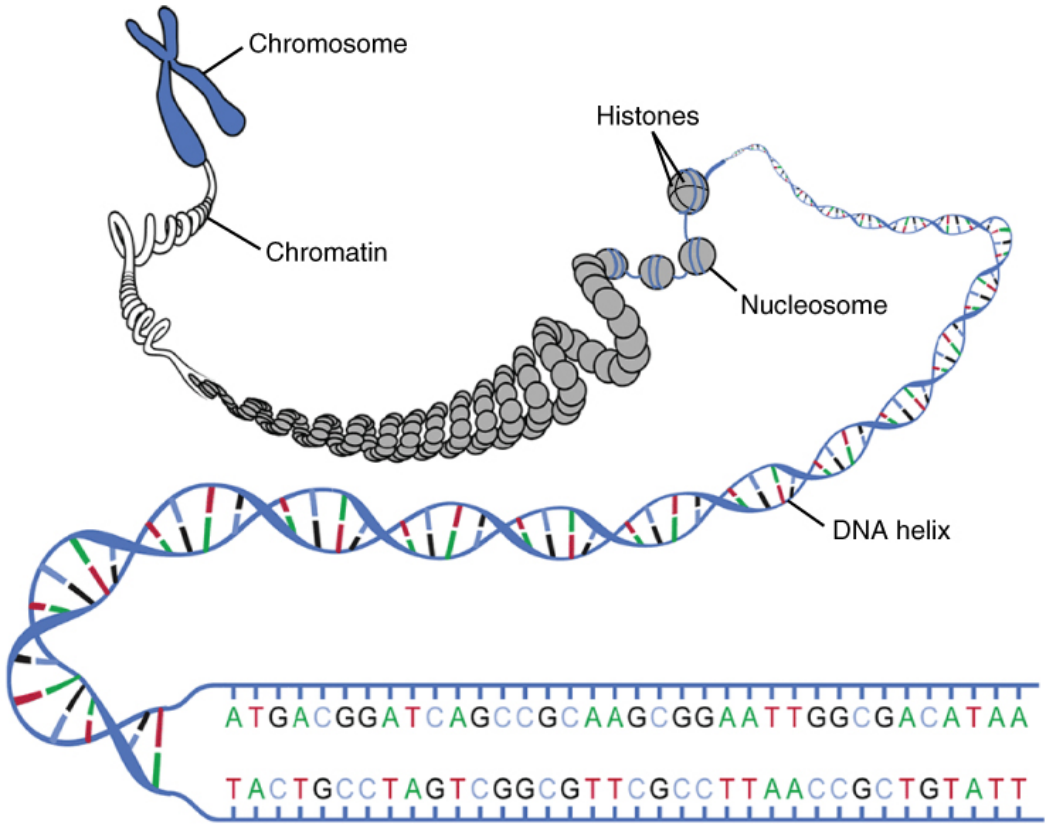
- Differential Methylation (array or sequencing) and matched RNA-Seq analysis
- Bioconductor Package @ bioconductor.org/packages/MethylMix/
- Demo data available from TCGA

COHCAP

- Differential Methylation (array or sequencing) and matched RNA-Seq analysis
- Bioconductor Package @ bioconductor.org/packages/COHCAP/
- Discussion groups online for Q&A
- Demo data included

illumina®

ATAC-seq



illumina®

What is ATAC-seq?

Assay for Transposase-Accessible Chromatin sequencing

October 2015

iCommunity NEWSLETTER

Surveying the Chromatin Landscape with Next-Generation Sequencing

Researchers develop novel sequencing methods with the MiSeq® and HiSeq® Systems to understand the epigenome and its impact on cancer and immune disease.

Introduction

Every cell in the human body has long strands of deoxyribonucleic acid (DNA) compactly folded inside its nucleus. That folding is made possible by chromatin, the complex of macromolecules that package each cell's DNA into that small, condensed volume—an architecture necessary to protect its structure and sequence. Understanding chromatin and this dynamic architecture are crucial to understanding how the genome works. Its tightly packed grooves and folds provide a unique physical landscape for gene transcription—one that has profound implications for our understanding of gene regulation, replication, and expression. Scientists are now finding new ways to delve into chromatin's many biochemical mysteries.

William Greenleaf, PhD, an assistant professor in Stanford University's renowned genetics department, is focused on understanding how the 2 meters of DNA in each cell nucleus are folded and stored. "About 95% of the genome is folded and sequestered away in the chromatin," Dr. Greenleaf said. "Only a small percentage is accessible to the transcription machinery. Deciphering how that all works is intriguing and important."

iCommunity spoke with Dr. Greenleaf about his team's development of 2 new next-generation sequencing (NGS) methods to better survey the enigmatic chromatin landscape: assay for transposase-accessible chromatin sequencing (ATAC-seq)[1] and single-cell ATAC-Seq (scATAC-seq).[2] He believes that these approaches might one day provide new insights into the development and treatment of cancer and autoimmune disease.

Q: What sparked your interest in applied physics?
William Greenleaf (WG): I was always interested in molecular biology—particularly DNA and the molecular machinery of the genome. But as an undergrad, I wanted to avoid chemistry, so I studied physics instead. I ended up getting my PhD in applied physics with a focus on single-molecule biophysics, because I was interested in understanding the mechanics by which individual molecules carry out tasks within the cell. During my postdoc, I was bitten by the high-throughput sequencing bug. We were thinking a lot about new ways to approach these different complex biological questions. A sequencer can make hundreds of millions or even billions of measurements across the genome and that's what is needed to understand the complexity of this biology.

Q: What does high-throughput sequencing provide over the other methods you used previously?
WG: As a grad student, I performed experiments on individual molecules. It's labor-intensive work—and you have to deal with a lot of handcrafted data. After a few years, I wanted to find a different way. I wanted to do the exact opposite—take an enormous number of measurements very quickly. So we've been working to repurpose the infrastructure associated with high-throughput sequencers to do massive scale biochemistry on nucleic acids.

Q: What inspired you to develop new tools to study chromatin?
WG: We have a great understanding of the structure of DNA—and a good understanding of a single nucleosome. However, that's where our high-resolution understanding of the nucleus ends. The question of how DNA is organized at the kilobase length scale remains a fundamental question to be answered. We don't know all that much about how the nucleosomes that bind to DNA tightly are shifting around, how the transcription factor binding sites might be competing for DNA, and how different transcription factors may cooperate to build enhancers. These things touch and interact mechanically to make things happen. We need to understand the logic of the physical regulatory landscape—the regulome, if you will—to see what makes a gene fire or not.

One of the significant questions is how a cell can mark and use these different elements to change their biological state. We know that all the different cells in a body have the same genome effectively, yet they do incredibly different things. I like to think of chromatin as a physical landscape that tells the cell which parts of the DNA to use and which parts to ignore. In a sense, it's a major organizational principle of biology.

Q: Has the data from the Encyclopedia of DNA Elements (ENCODE) Consortium and Epigenetics Roadmap provided a glimpse into the regulome?
WG: Recent work from the ENCODE consortium and the Epigenomics Roadmap have tried to illustrate how different elements in DNA are functional, and how they can be marked and used. That initial

Dr. William Greenleaf is an assistant professor in the Stanford University Genetics Department.

For Research Use Only. Not for use in diagnostic procedures.

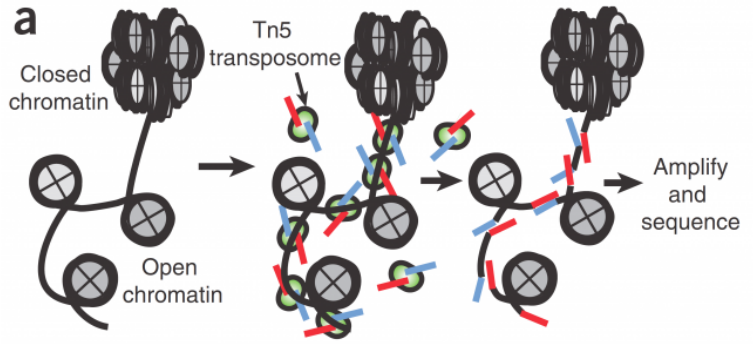
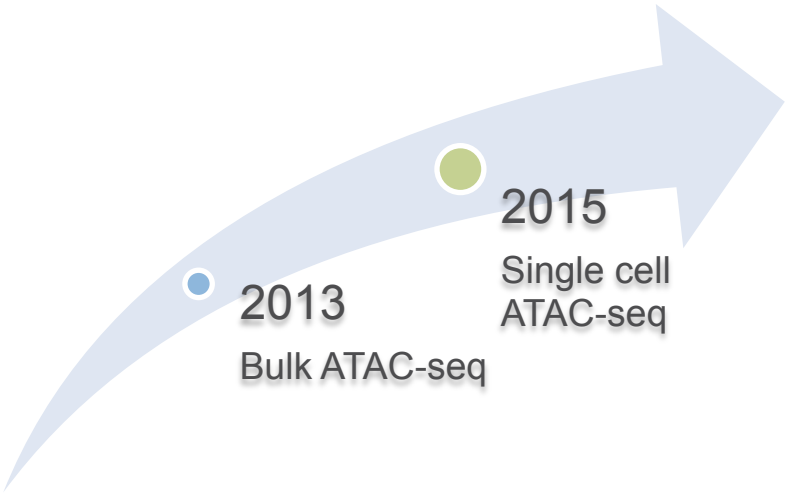


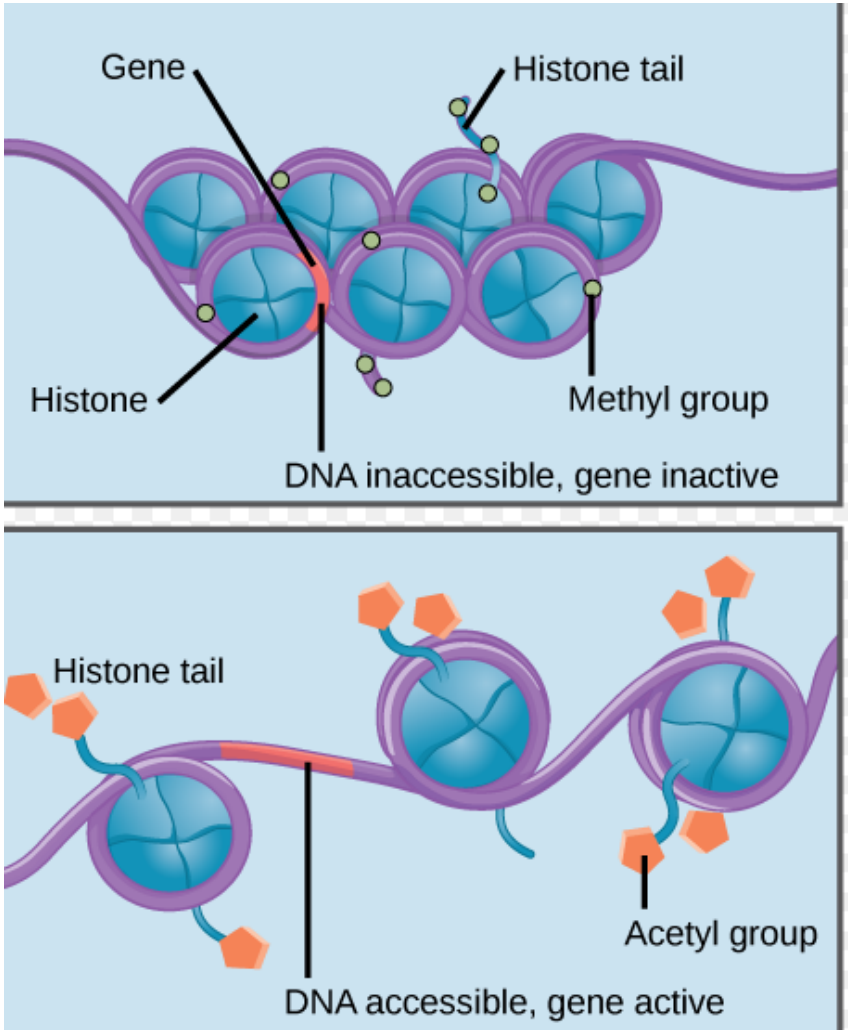
Figure 1. Scheme for ATAC-seq technic. Transposase enzyme (green), bearing sequencing adaptors (red and blue), is incorporated only in regions of open chromatin (between nucleosomes in grey). Allowing to amplify those open regions by PCR. | Credit: Buenrostro et al. 2013. Nat. Methods 10, 1213–8.



** Prof Greenleaf co-founded Epinomics company*

illumina®

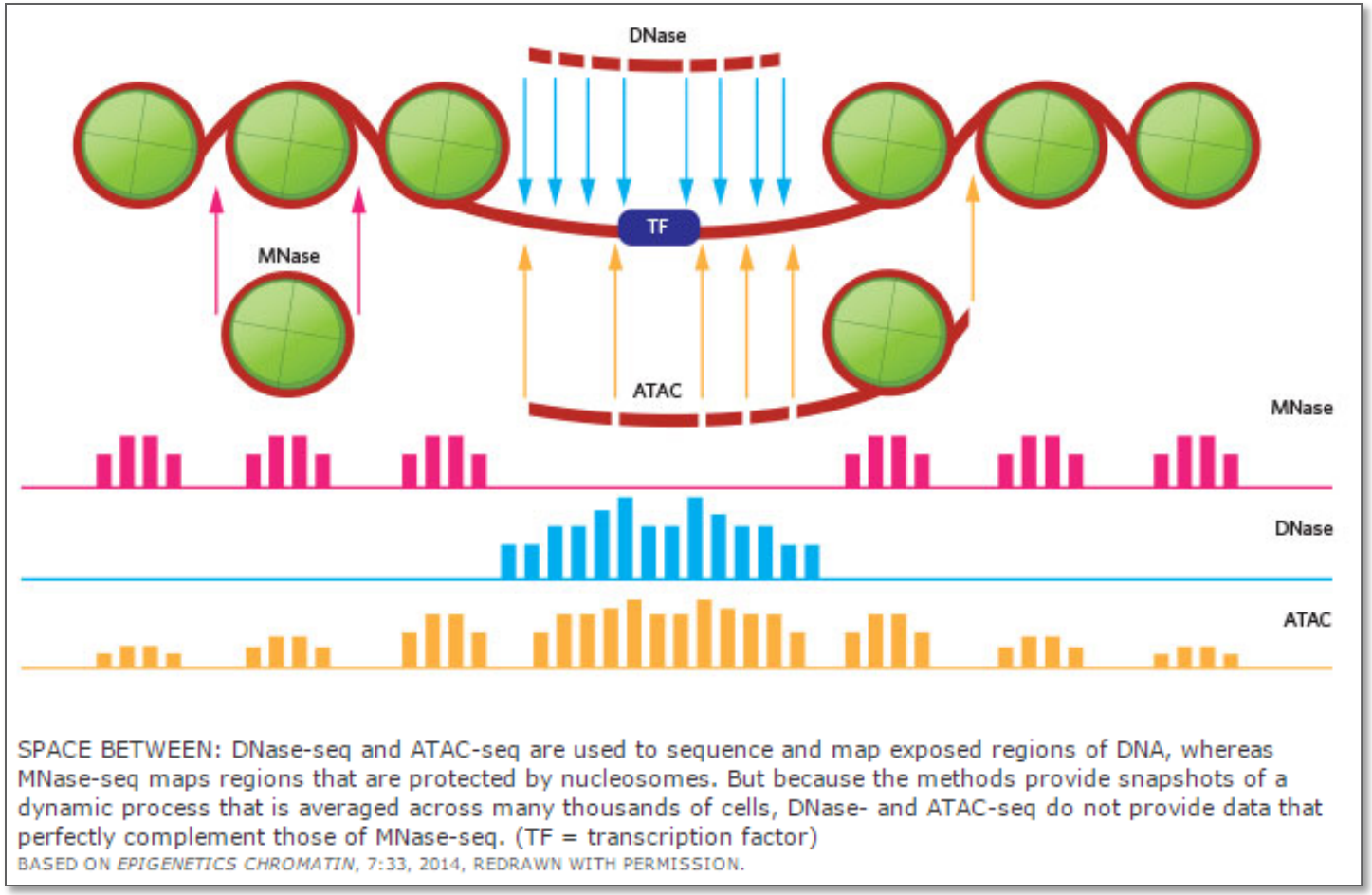
Why do researchers want to look at chromatin structures?



Methylation of DNA and histones causes nucleosomes to pack tightly together. Transcription factors cannot bind the DNA, and genes are not expressed.

Histone acetylation results in loose packing of nucleosomes. Transcription factors can bind the DNA and genes are expressed.

ATAC-seq combines data signal results of legacy methods MNase and DNase



SPACE BETWEEN: DNase-seq and ATAC-seq are used to sequence and map exposed regions of DNA, whereas MNase-seq maps regions that are protected by nucleosomes. But because the methods provide snapshots of a dynamic process that is averaged across many thousands of cells, DNase- and ATAC-seq do not provide data that perfectly complement those of MNase-seq. (TF = transcription factor)

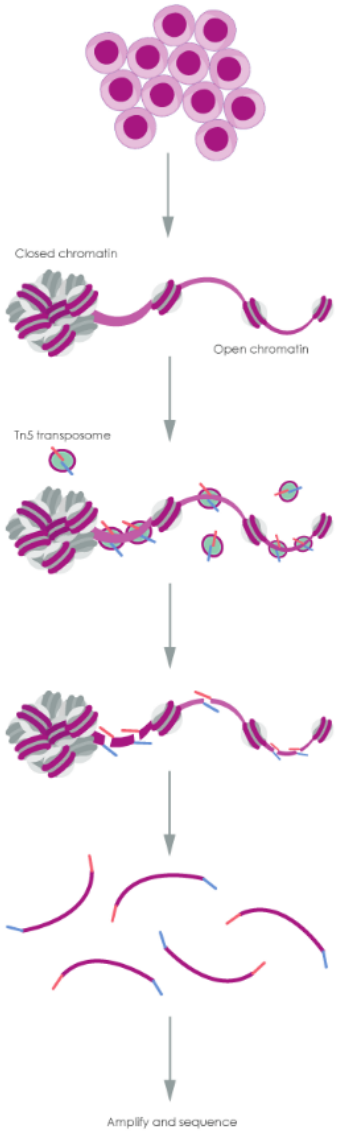
BASED ON *EPIGENETICS CHROMATIN*, 7:33, 2014, REDRAWN WITH PERMISSION.

http://www.the-scientist.com/?articles.view/articleNo/44772/title/Reveling-in-the-Revealed/

Tsompana and Buck: Chromatin accessibility: a window into the genome. Epigenetics & Chromatin 2014 7:33

ATAC-seq workflow

- Easy, 3 hour workflow
- Requires least amount of cells compared to other methods
- Number of reads for a region correlates with how open that chromatin is at a single nucleotide resolution.



Harvest cells. An accurate cell number is key to the success of the experiment.

Lyse cells to generate a crude nuclei preparation.

Tn5 tagmentation

Purify the fragmented and tagged DNA.

PCR amplify and purify the amplified DNA.

Sequence

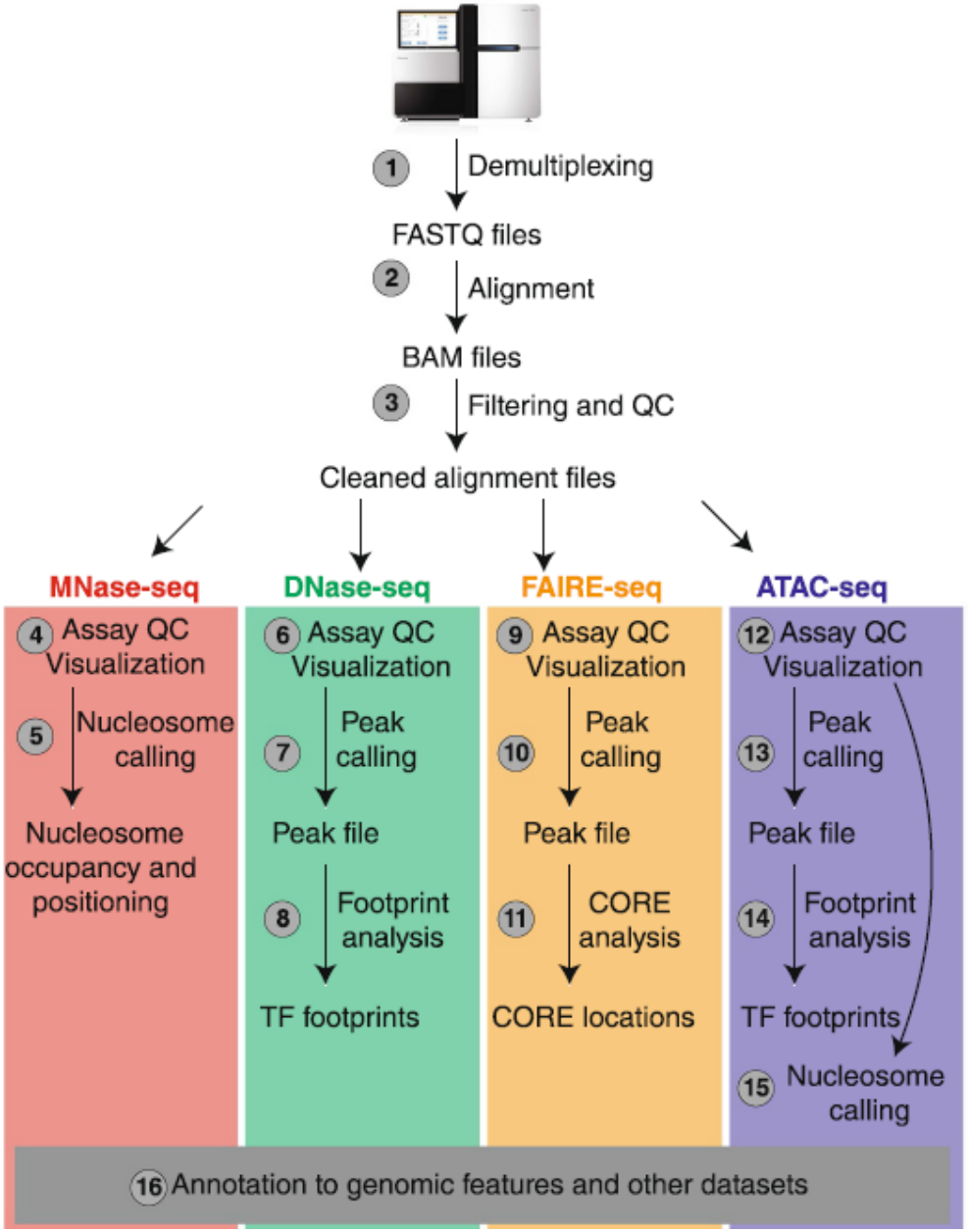
illumina®

How many ATAC-seq samples can you run on a HiSeq?

	HiSeq 3000	HiSeq 4000
		
Assuming 50M reads per ATAC-seq sample	Up to 50 samples	Up to 100 samples

ATAC-seq data analysis

- Analysis remains an evolving challenge for researchers.
- BaseSpace will launch an ATAC-seq analysis app coming soon



Tsompana and Buck: Chromatin accessibility: a window into the genome. Epigenetics & Chromatin 2014 7:33

illumina®

Explosion of methods…For All You Seq

http://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/ForAllYouSeqMethods.pdf

illumina®

Methods Selector

Methods for transcriptomic analysis:

Ribo-Seq/ART-Seq/GTI-Seq

Active mRNA Translation Sequencing (ART-seq), also called ribosome profiling (Ribo-Seq) or Global Translation Initiation Sequencing (GTI-Seq), isolates RNA that is being processed by the ribosome in order to monitor the translation process. In this method ribosome-bound RNA first undergoes digestion. The RNA is then extracted and the rRNA is depleted. Extracted RNA is reverse-transcribed to cDNA. Deep sequencing of the cDNA provides the sequences of RNAs bound by ribosomes during translation.

References:

ART-Seq/Ribo-Seq: Ingolia N. T., Ghaemmaghami S., Newman J. R. and Weissman J. S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324: 218-223

GTI-Seq: Wan J. and Qian S. B. (2014) TISdb: a database for alternative translation initiation in mammalian cells. Nucleic Acids Res 42: D845-850

Associated kits:

- ARTseq/TruSeq Ribosome Profiling kit

Find the right kit

http://www.illumina.com/science/sequencing-method-explorer.html

illumina®

Thank you!

Questions?

illumina®