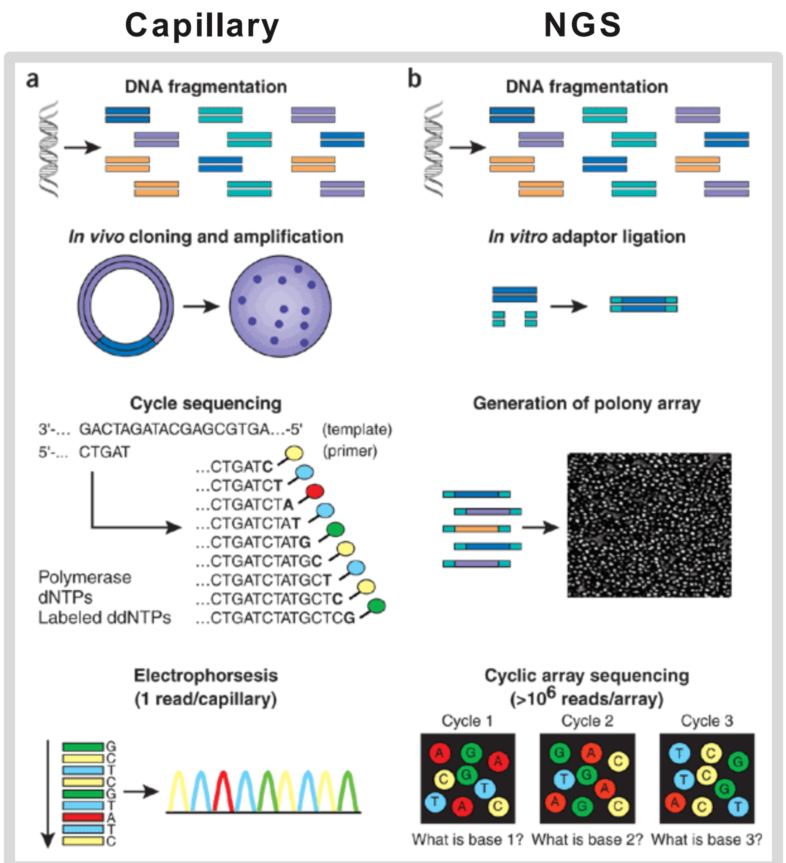
Intro to NGS & Applications

Matt Angel
Sr. Sequencing Specialist | 06/12/2018
mangel@illumina.com

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

illumina

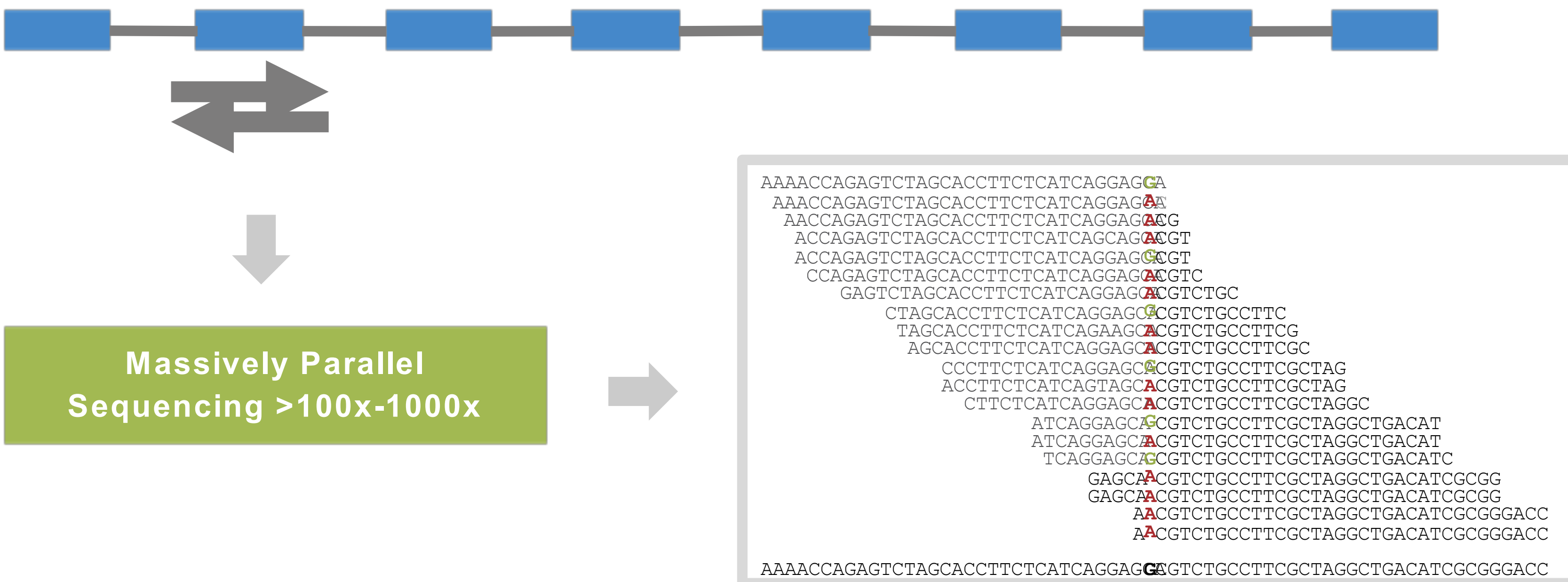
Next-Generation Sequencing (NGS)



2016 capacity:
10^6 – 10^9 reads/array

Shendure, J. et al. Next-generation DNA sequencing. *Nature Biotechnology* 26 1135 – 1145 (2008)
http://www.illumina.com/systems/sequencing.html

Next-Generation Sequencing



NGS provides "digital" data, each read is analysed independently and is quantitative

Overlapping reads are aligned together, resulting in quantitative and high confidence variant calling

illumina

Sanger Versus NGS (Analogue Versus Digital)

Sensitivity / Limit of Detection (LOD)

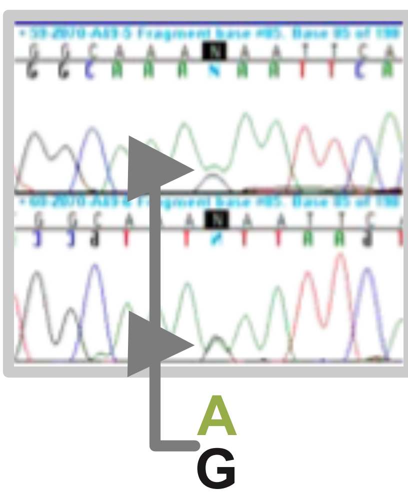
Sanger Sequencing

Limit of Detection ~20%

Molecular Pathology Checklist

“Detection of **20% variant allele frequency**, which is typically equivalent to a 40% proportion of mutation positive cells, is commonly measured as the **LOD for Sanger Sequencing**.”

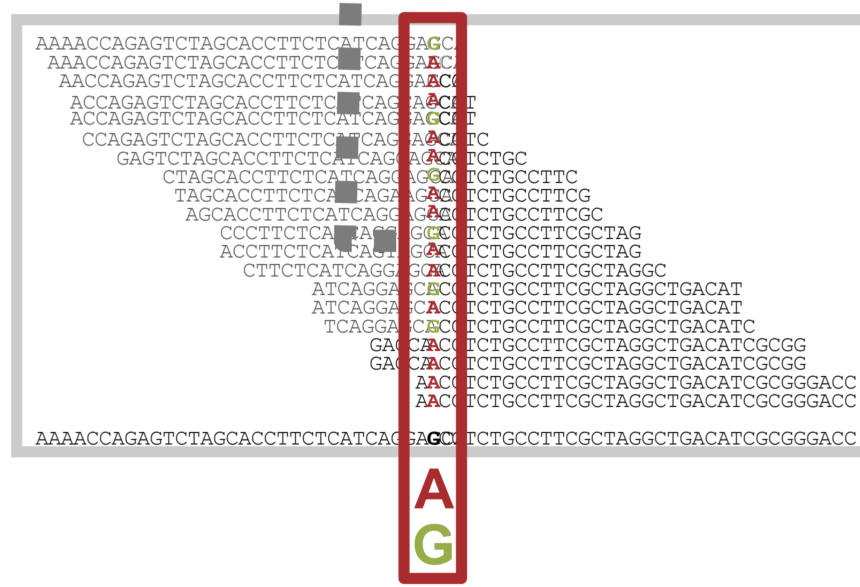
- College of American Pathologists (CAP)



NGS

Limit of Detection ~ 3-5%

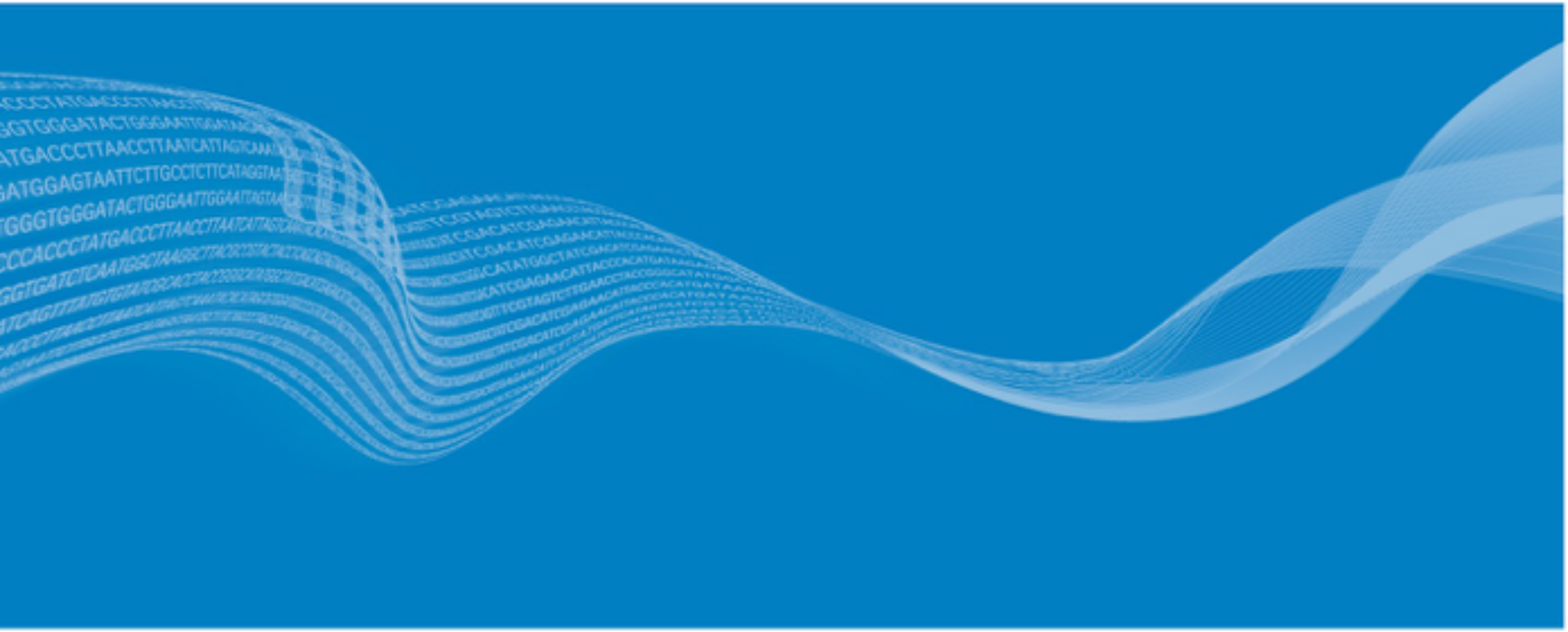
Coverage	30X	100X	1,000X	10,000X
A	28 – 29	95 – 97	950 – 970	9500 – 9700
G	1 – 2	3 – 5	30 – 50	300 – 500
Limit of Detection	~3 – 6%	3 – 5%	3 – 5%	3 – 5%



Kloss-Brandstatter A., Weissensteiner H., Erhart G., Schafer G., Forer L., et al. (2015) Validation of Next-Generation Sequencing of Entire Mitochondrial Genomes and the Diversity of Mitochondrial DNA Mutations in Oral Squamous Cell Carcinoma. PLoS One 10: e0135643

illumina

Why do labs use NGS?



illumina

Traditional Genomic Technologies

qPCR
First instrument released in 1996

Low complexity assays	Single assay reactions
Lack of discovery power	Cumbersome workflow

Sanger/CE Sequencing
First instrument released in 1987

Low complexity assays	Low sensitivity and detection power
Lack of discovery power	Missed calls

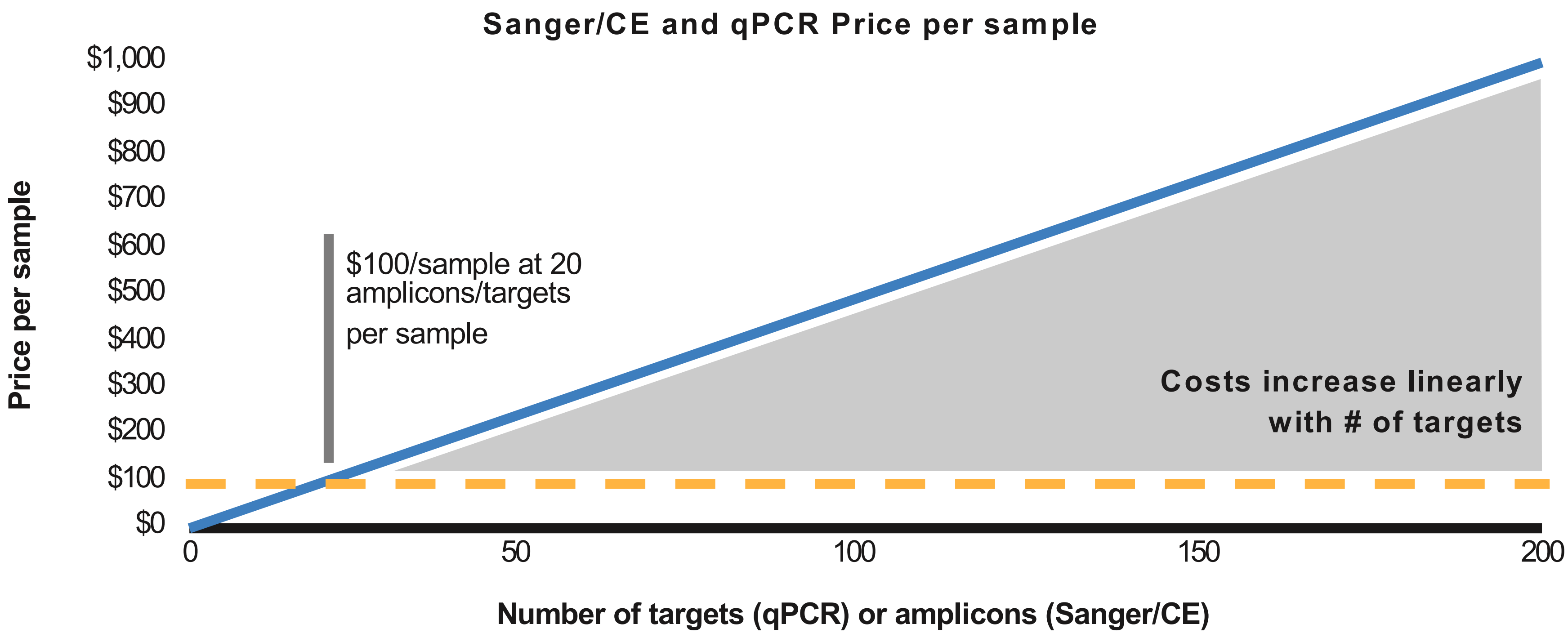
The Challenge

How do I cost-effectively expand my research?

illumina

Cost of Traditional Genomics Technologies

Gene Expression on qPCR & Targeted Resequencing Sanger/CE



Time, cost per sample, and sample input quantity all increase linearly with number of targets:

limiting discovery power

Assuming an average cost of $5/reaction/sample based on customer conversations (qPCR and Sanger)

illumina

Ability to Scale

Inherently limited in qPCR & Sanger/CE

Current Capabilities

Surveying a limited number of targets on Sanger or qPCR

What are you missing out on…

Ability to efficiently run mid-complexity panels (>20)

Current Options	Outsource	Purchase higher capacity systems	Run multiple plates on your current qPCR or Sanger system
Cost of Solution	Long wait times for custom projects Lose control of your samples and experiment	High throughput systems require a large investment–$100k+	Substantial time investment for optimization and sample processing

A Solution To Overcome These Challenges Would Require…

Ability to **simultaneously survey 100s of targets**

Maintain a manageable price per sample **as you scale**

Low sample input to **conserve precious sample**

Maintain sensitivity and **data resolution**

Enhanced discovery of genetic variation

NGS provides a broader understanding of genetics

- **Qualitative information**
  - Mutations or changes from a standard reference. SNPs, insertions, deletions, duplications, inversions.
    - Cancer, heritable disorders
  - Pairwise differences. What is it about their genetic makeup makes sample A and sample B different?
    - Disease resistance, genetic rick factors, morphological differences
  - Validation. Did my breeding, genetic modification, or construct come out as planned?
    - Genetic engineering, agriculture, synthetic biology, cloning

Understanding the genetic drivers of disease

NGS is a highly sensitive method to quantify genetic effects

- **Quantitative information**
  - Copy number variation
    - Reproductive health, genetic engineering
  - Gene expression
    - Host/pathogen interactions, drug response
  - Gene regulation
    - Small RNA sequencing, antisense expression
  - Protein/DNA interactions
    - DNA binding sites & regulatory pathways
  - Epigenetics
    - DNA methylation
  - Metagenomics
    - Microbial community profiling


Resequencing


RNA-Seq


ChIP-Seq


16S Metagenomics

Next-Generation Sequencing Workflow

Common Workflow

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

Library Preparation

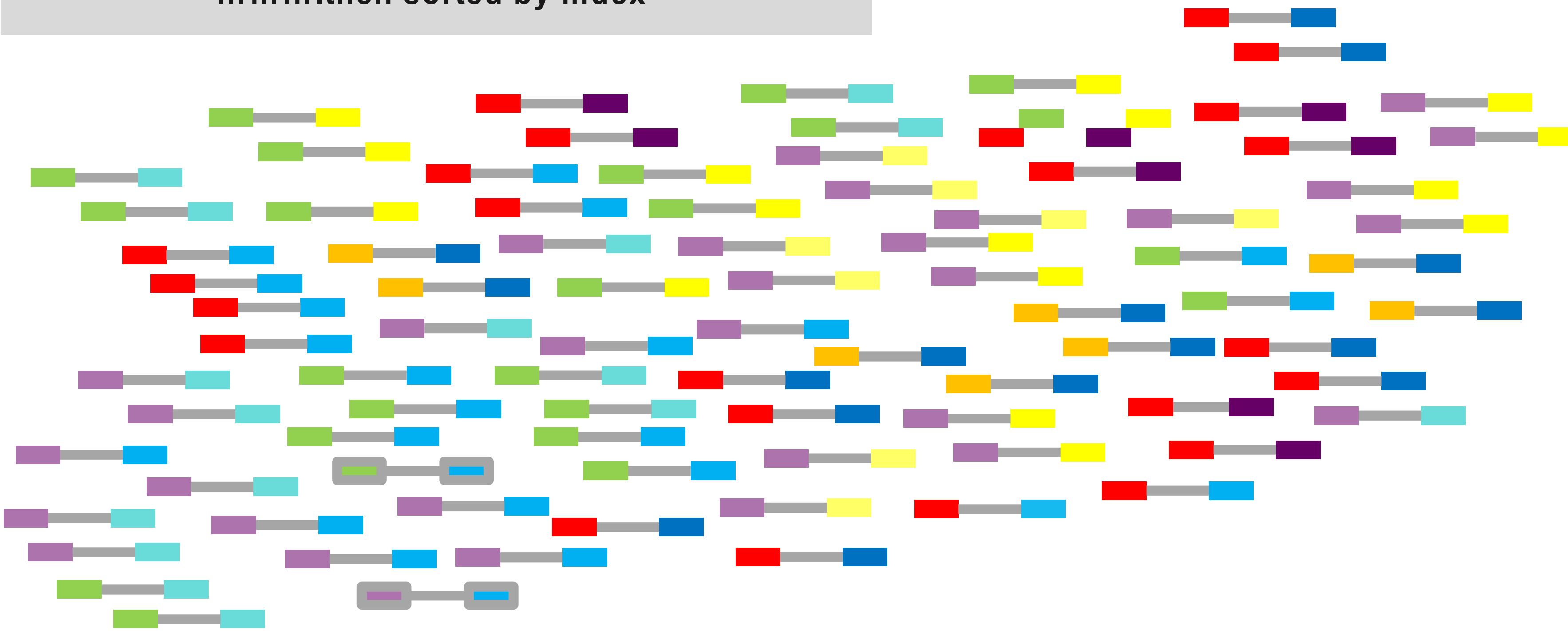
No matter the application, all libraries end up looking similar

Dual Index Library Shown

The aim of library preparation is to obtain nucleic acid fragments with adapters attached on both ends

How and Why Samples Can Be Mixed Together

………then sorted by index



Sample 6 | Sample 7 | Sample 8 | Sample 9 | Sample 10

Sample 1 | Sample 2 | Sample 3 | Sample 4 | Sample 5

Sequencing

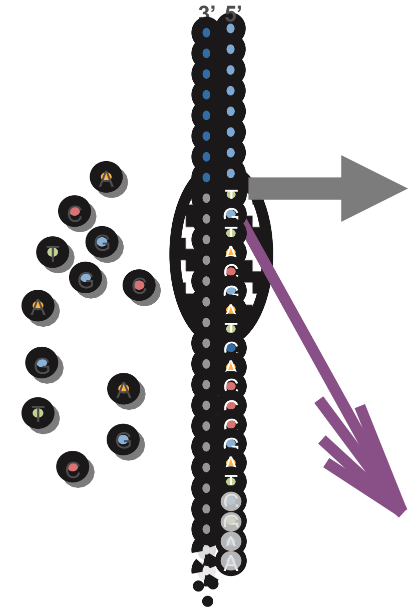
Sequencing-by-Synthesis (SBS)

DNA
(0.01-1.0 µg)

Library Preparation

Single molecule array

Cluster Growth



Sequencing

1 2 3 4 5 6 7 8 9

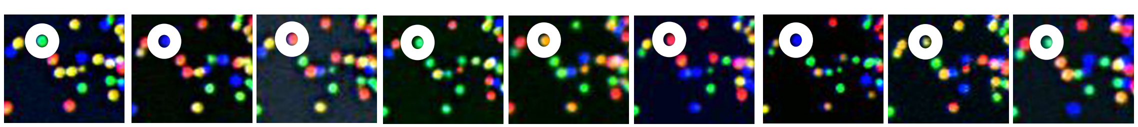


Image Acquisition

T G T A C G A T ...

Base Calling

Data Analysis

Sequencing reads aligned to a reference genome

CGATTAGTAC
ACTCGATTAG
ACGGGCTCGA
ACGACGGACT

Reference Sequence CATACGACGGACTCGATTAGTACTCGTA

Variant/genotype calling

Annotation & filtering

Functional effect

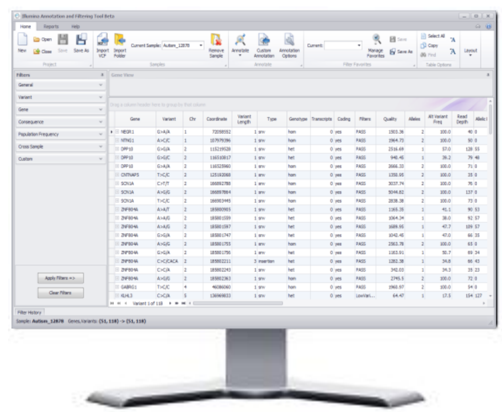
20bp region
6bp @ 1X
8bp @ 2X
6bp @ 3X
Average = 2X

Analysis and Interpretation of Sequencing Data

Data in, biological knowledge out

Public Databases & Private Data Centres

Sequencing data

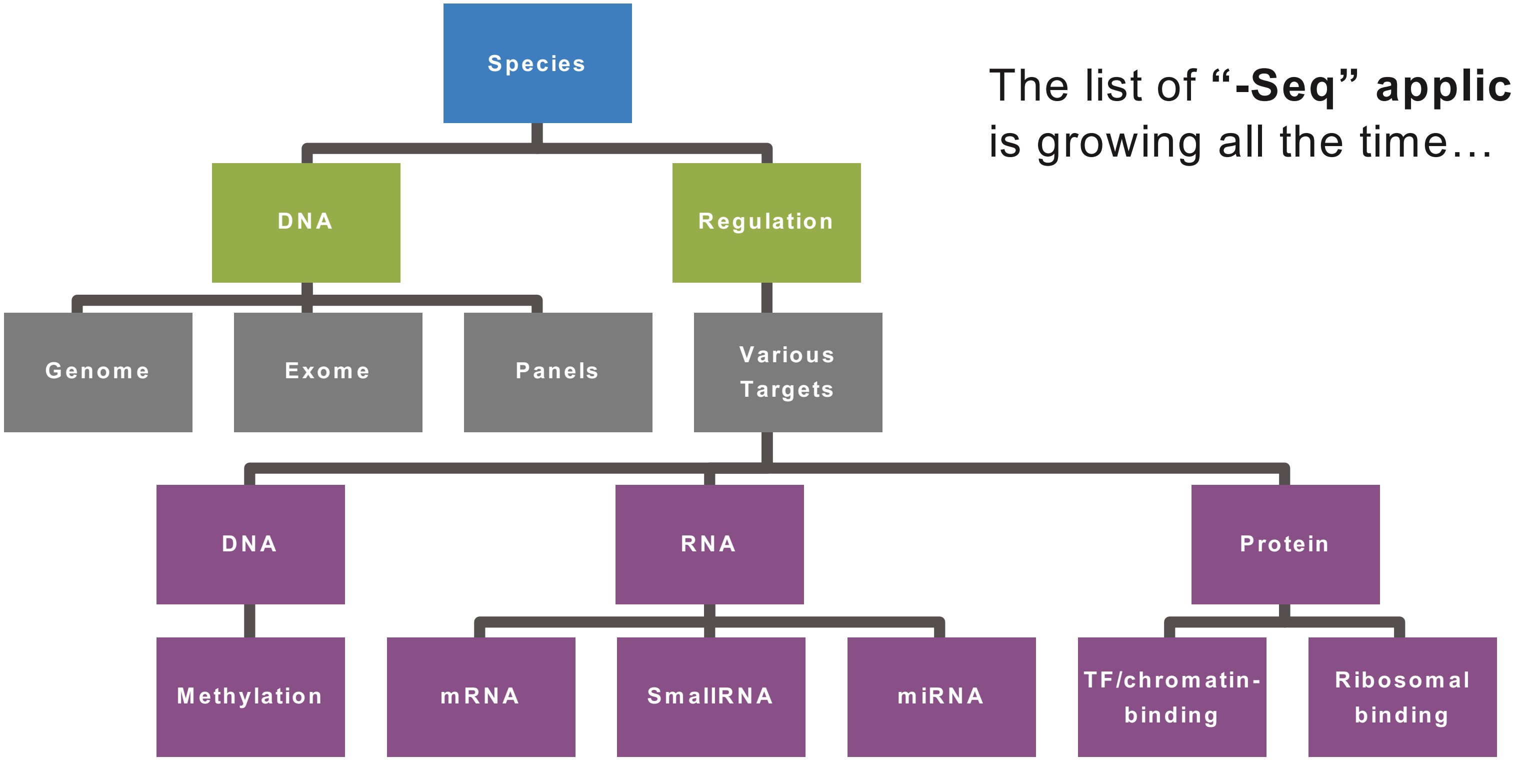


Data Analysis

Export annotated/filtered variants

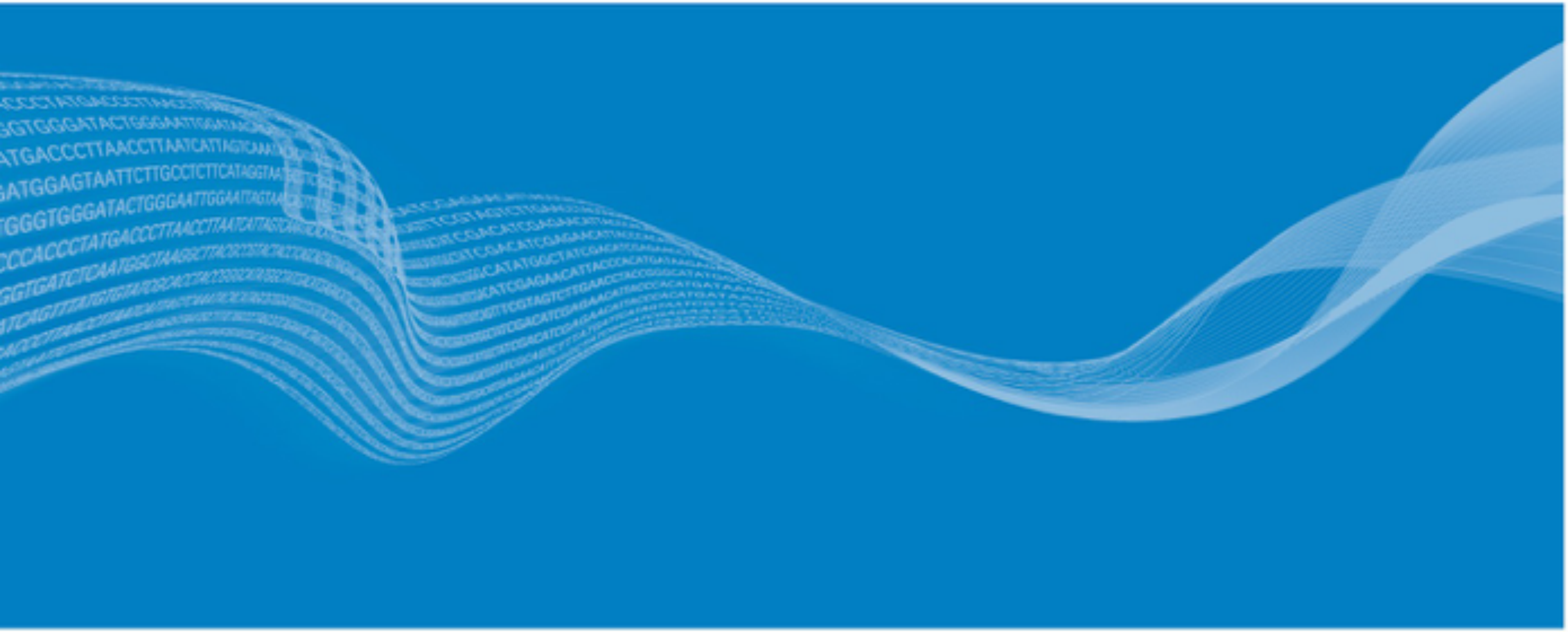
For Research Use Only. Not for use in diagnostic procedures.

Sequencing Applications



The list of **"-Seq" applications** is growing all the time…

Next Generation Sequencing Applications



RNA-Seq

RNA-Seq is Impacting Cancer

Genomic Expression makes Red Herring "Top 100" for disprupting cancer care with RNA-Seq

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

> *"In cancer, sometimes a single gene produces two isoforms, one of which promotes metastasis and one of which represses metastasis," he said, adding that understanding the differences between the two is extremely important in combatting cancer.*
>
> *"We have just scratched the surface," Xing said. "We will apply the method to much larger data sets, and we expect to learn a lot more."*

Yi Xing, UCLA, RNASeqBlog.com

illumina

RNA-Seq is Impacting Genetic Disease Research

RNA-Seq identifies novel myocardial gene expression signatures of heart failure

Genome-wide RNA-Seq of Human Motor Neurons Implicates Selective ER Stress Activation in Spinal Muscular Atrophy

RNA-Seq reveals genetic link between two chronic lung diseases

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

illumina

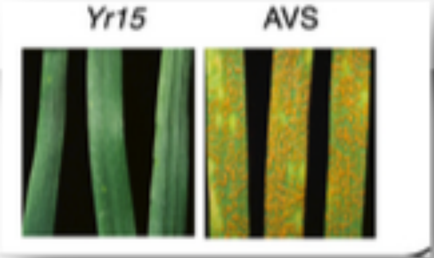
RNA-Seq is Impacting Our Environment

Researchers use RNA-Seq to assess how a tiny dinoflagellate could have a big impact on the pacific food chain

University of Hawai’i Manoa

Gene expression analysis of bud and leaf color in tea.

RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat



illumina

Consortium RNA-seq studies

Multi-platfo
using RNA-
sequencin

Sheng Li, Scott W T
Farmerie, Agnes Vi
Boland, Belynda Hi
Raghavachari, Jorg
Roberson, Jeffrey F

Affiliations | Contri

Nature Biotechnolog
Received 14 May 20

A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium

SEQC/MAQC-III Consortium

Affiliations | Contributions | Corresponding authors

Nature Biotechnology **32**, 903–914 (2014) | doi:10.1038/nbt.2957
Received 13 June 2013 | Accepted 11 May 2014 | Published online 24 August 2014 | Corrected online 09 September 2014

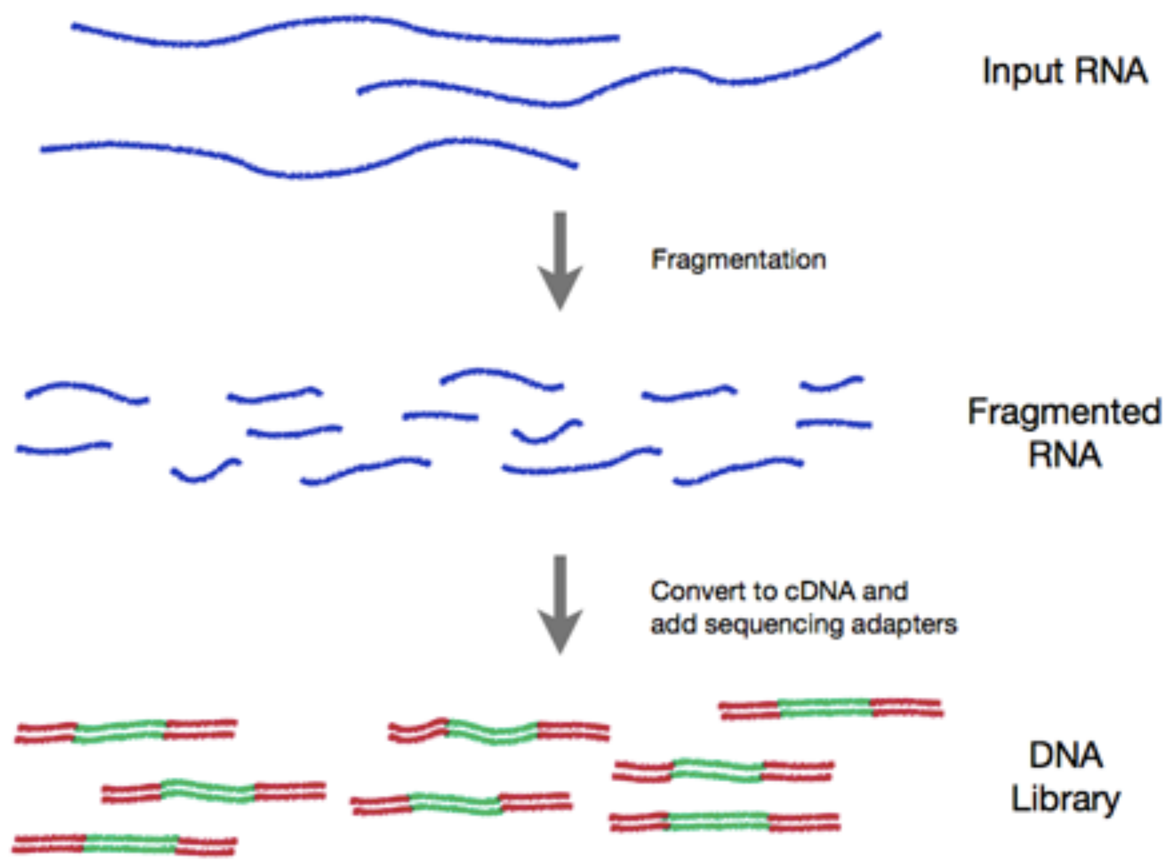
Key Take away from Studies:

1. RNA-Seq is now about as inexpensive as microarrays, and comes with the advantages of higher reproducibility and a broader dynamic range.
2. All of the sequencing platforms are producing high-quality, consistent data
3. Library prep methods matter more than you might think.

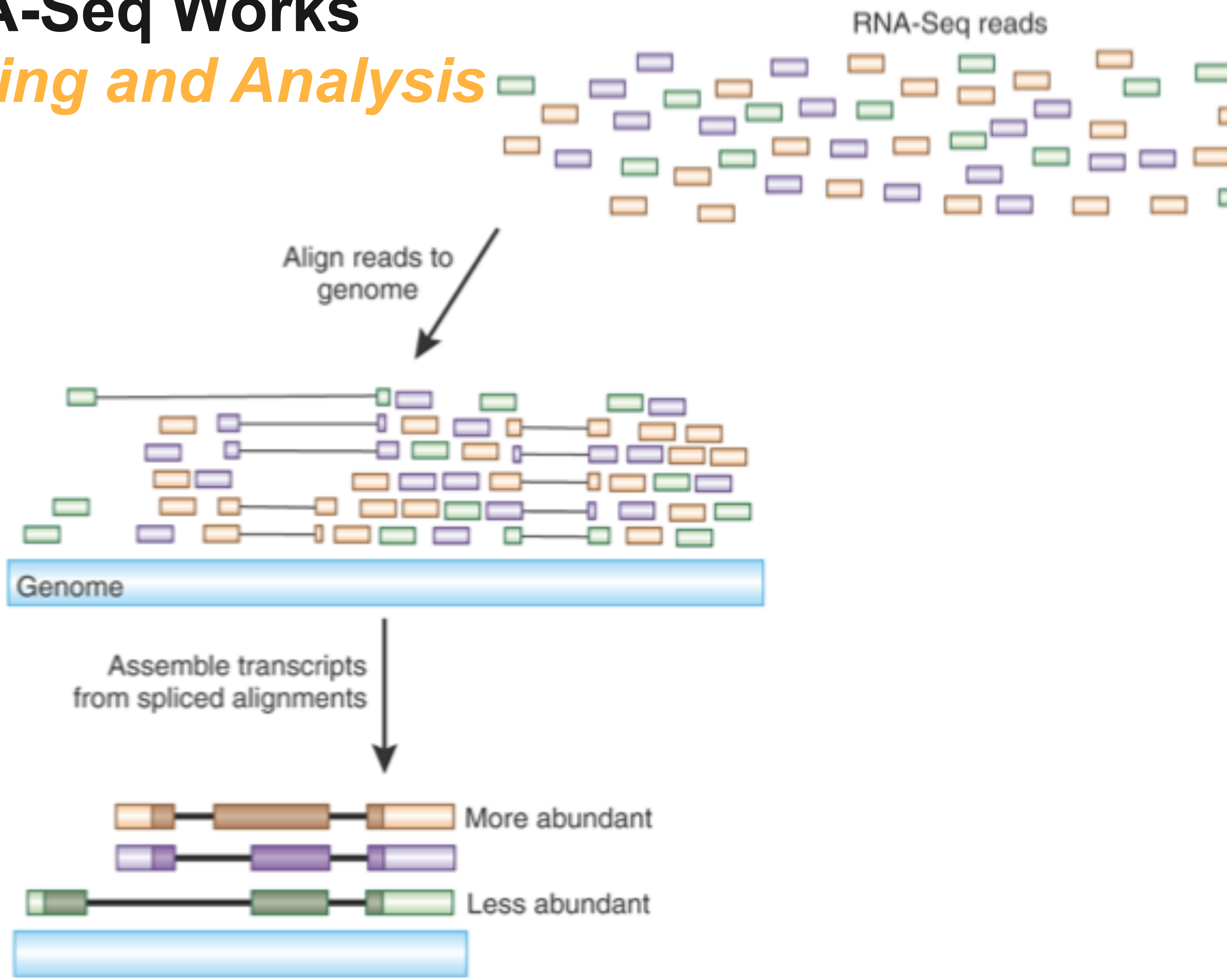
http://cofactorgenomics.com/lates-rna-seq-need-know/

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

How RNA-Seq Works
Library Prep



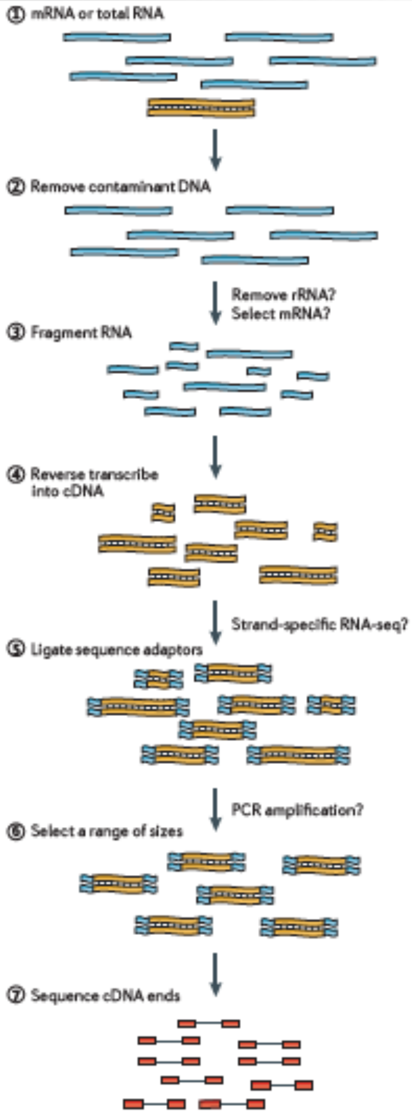
How RNA-Seq Works
Sequencing and Analysis



RNA-Seq | Key Benefits

1. **Limitless:** Most organisms, even unsequenced
2. **Precise:** Can reveal location of transcriptome boundaries (single-base resolution) and SNPs, fusions, splicing, etc
3. **Flexible:** Large dynamic range (5-6 orders of magnitude)

Considerations for RNA-Seq Library Preparation



- **What is the integrity of the RNA?**
- **How much Total RNA is available?**
- **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.

RNA Input Quality and Quantity

- **Input quality should be assessed with a BioAnalyzer or Fragemnt Analyzer**
  - Some kits do not work on degraded samples
  - Some kits optimized specifically for FFPE

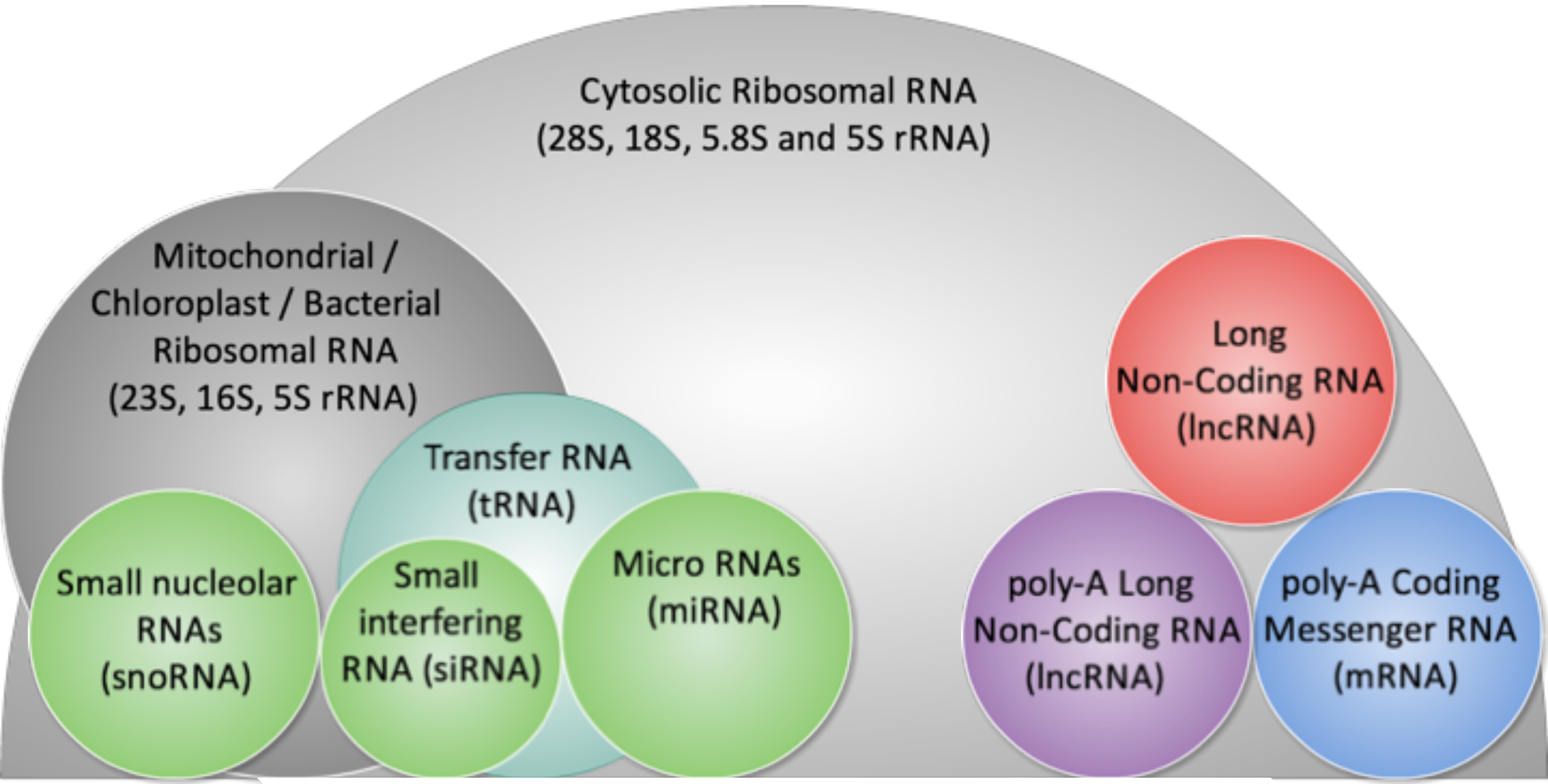
- **Input quantity should be assessed with a fluorometric method**
  - mRNA and Total RNA preps require 100ng
  - RNA Access kits can go as low as 10ng

Input quality and quantity will guide you to the correct RNA-Seq Library Prep Solution

RNA Isolation → Library Prep → Sequencing → Data Analysis

Many different RNAs exist

Ribosomal RNA is most abundant RNA species



RNA Isolation | Library Prep | Sequencing | Data Analysis

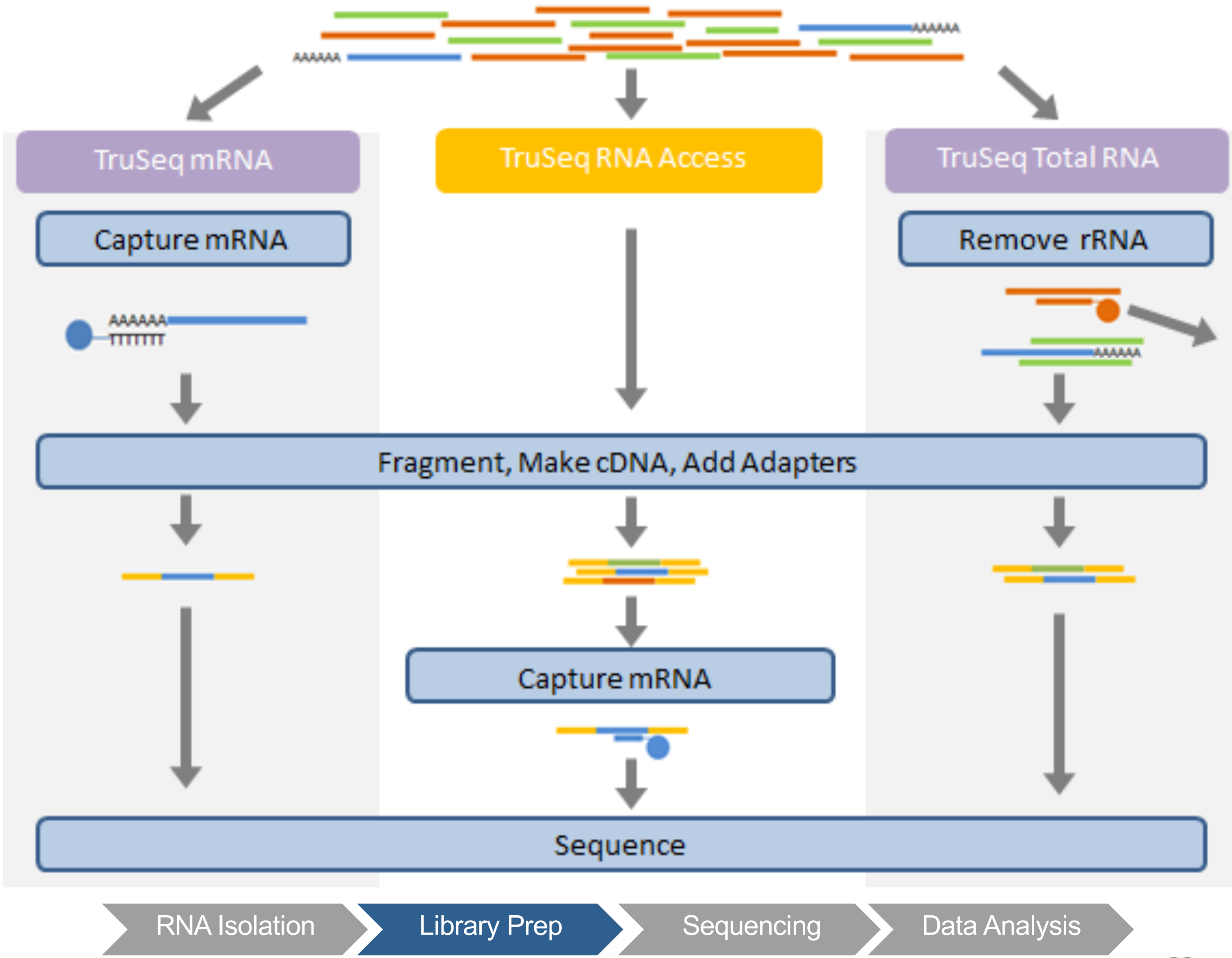
illumina

Illumina's Suite of RNA Library Prep Solutions

Whole Transcriptome	mRNA-Seq/ GEx Profiling	
TruSeq Stranded Total RNA	**TruSeq Stranded mRNA**	**TruSeq Stranded RNA Access**
• Coding + ncRNA • FFPE compatible • Many species	• Coding RNA • High quality RNA • Mammals	• RNA Exome • FFPE /low input • Humans

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

RNA Library Prep Chemistries



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

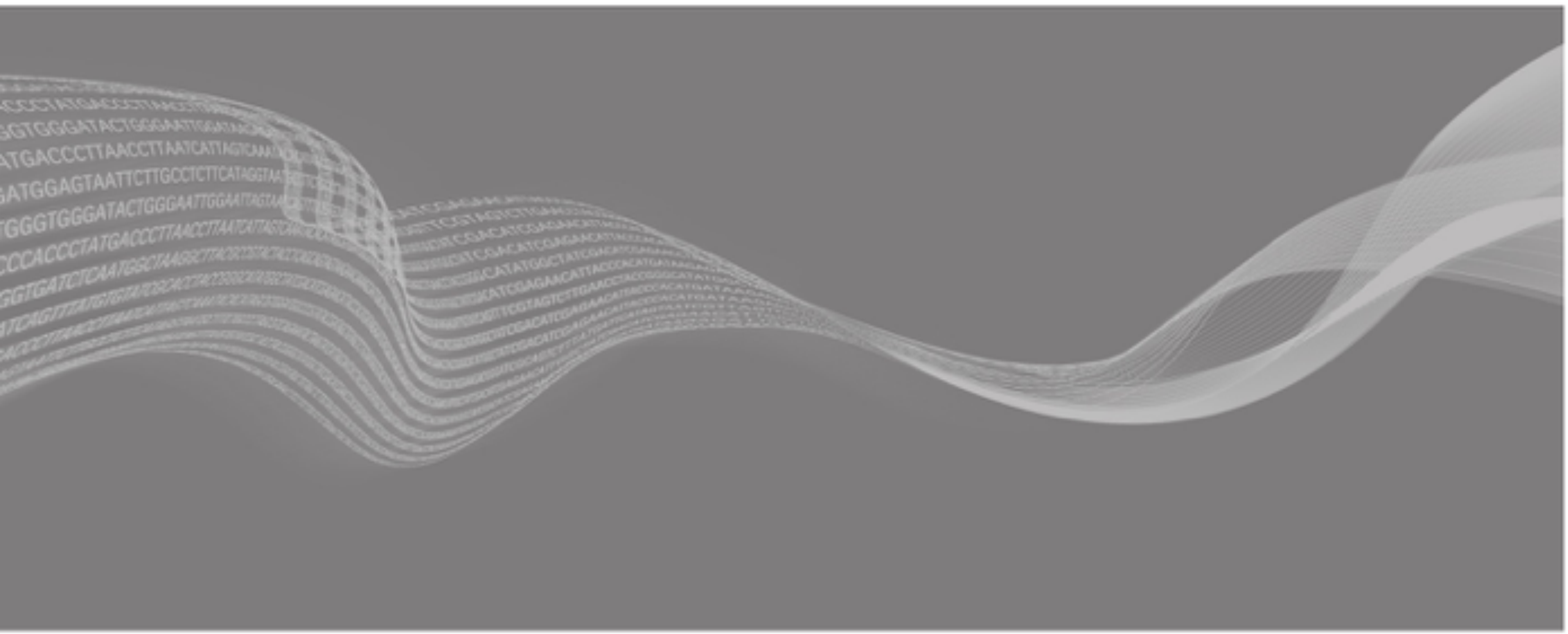
Sequencing costs vary by method/project

Application	Description	Read Length	Number of Reads
TruSeq Stranded mRNA	• Transcript level abundance • Discover novel features • Variant detection	2x75 bp	≥25M
RNA Access Deep Coverage	• Coding Region Interrogation • High Splice Junction Sensitivity • Discover rare mRNA features • Variant detection	2x75 bp	≥25M
TruSeq Stranded Total RNA	• Coding + ncRNA Interrogation • Transcript level abundance • Discover novel features • Variant detection	2x75 bp Or 2x100 bp	≥50M

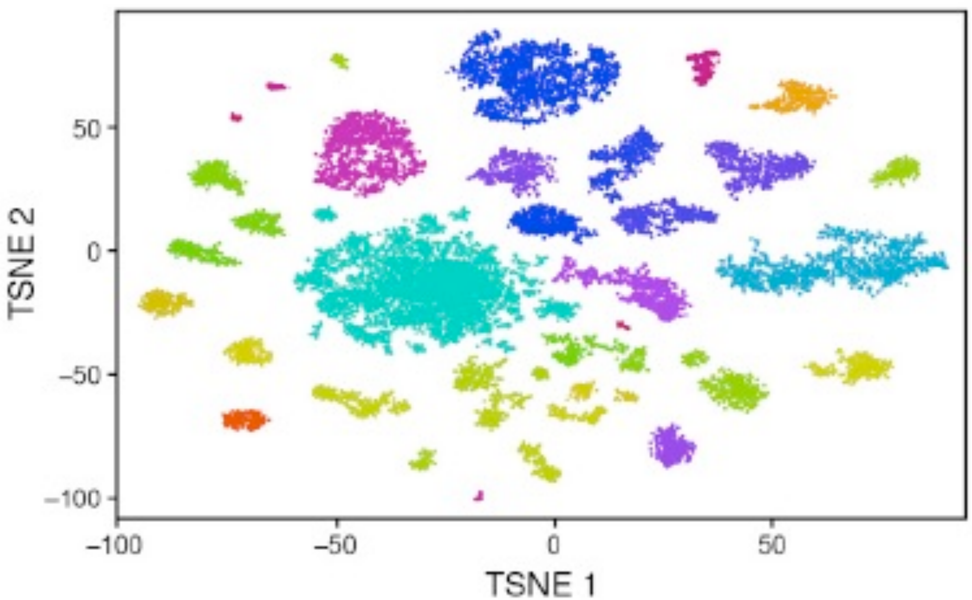
RNA Isolation → Library Prep → Sequencing → Data Analysis

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

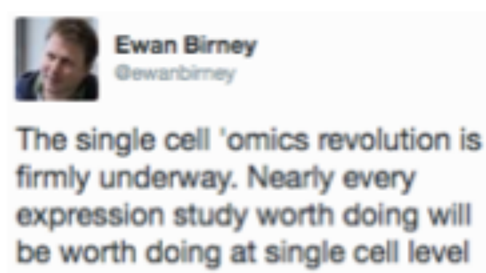
Single-Cell Sequencing



The importance of single cell sequencing

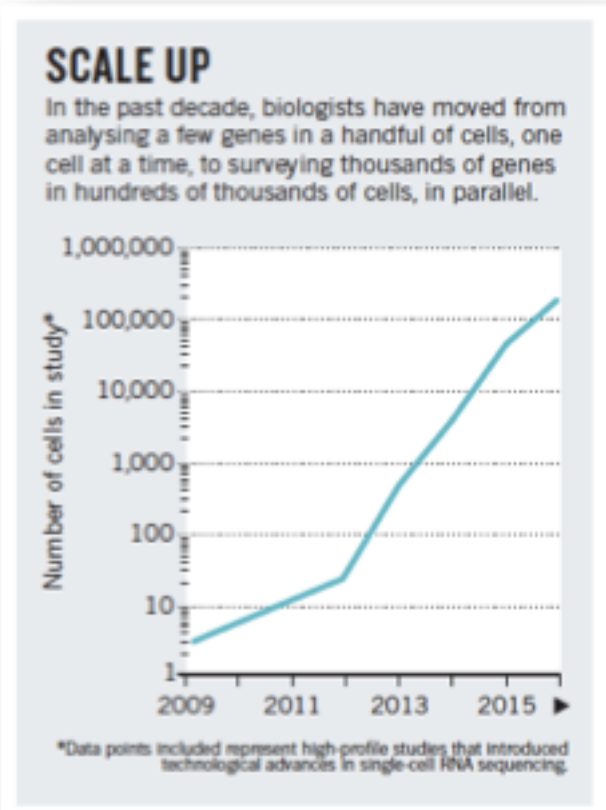


Macosko et al, Cell: May, 2-15



illumina

Single-Cell: Atlas or Characterize



Amir Giladi and Ido Amit

28 | NATURE | VOL 547 | 6 JULY 2017

Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq

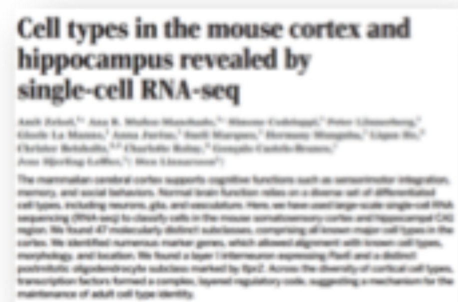
198 lung epithelial cells

Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types

4000 immune cells

Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity

722 immune cells



Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq

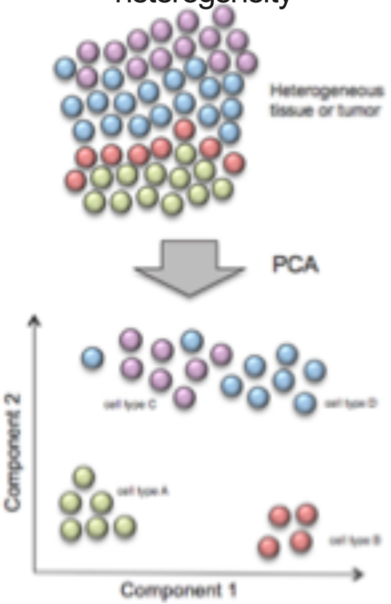
3000 neuronal cells

Massively parallel single-nucleus RNA-seq with DroNc-seq

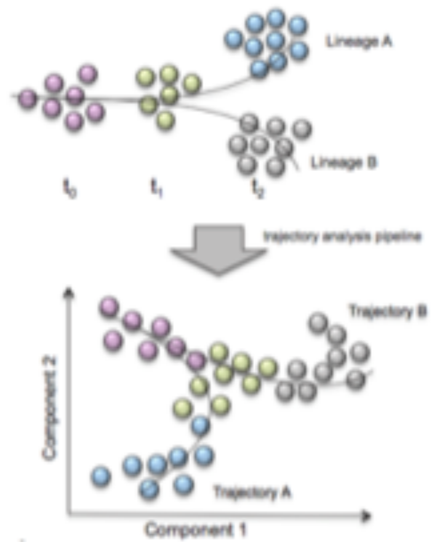
200,000 cells

Why single cells?

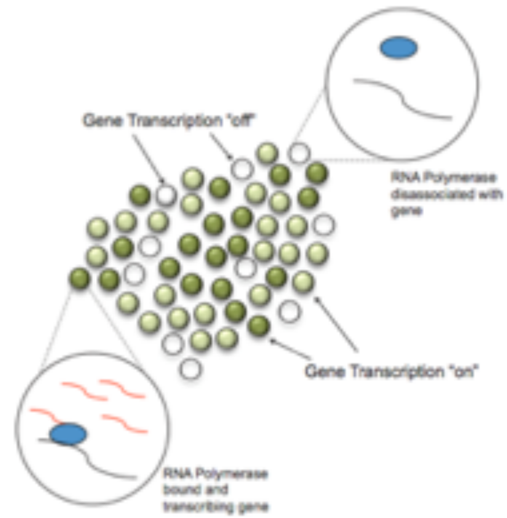
1. Assess cell-to-cell heterogeneity



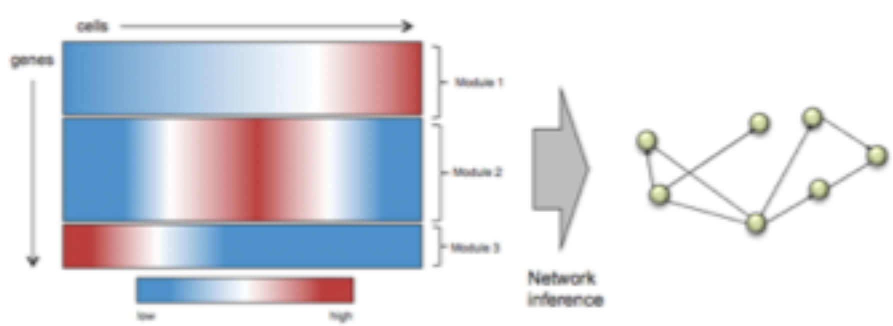
2. Map cell trajectories



3. Dissect transcriptional mechanics



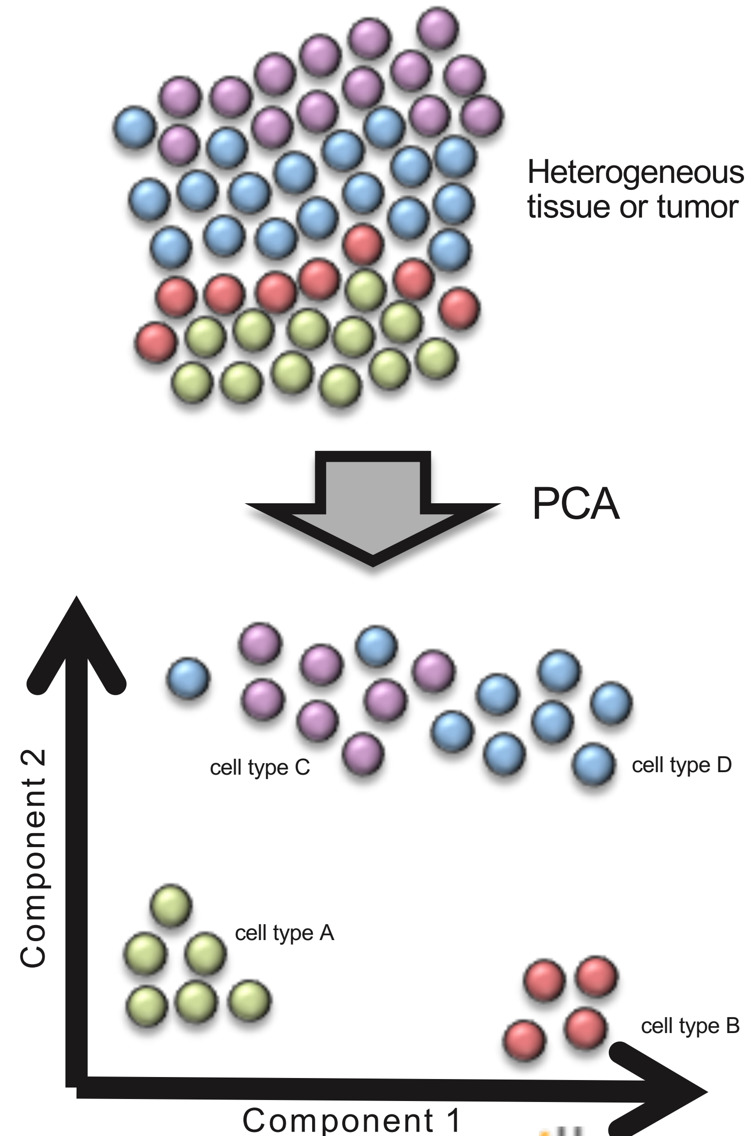
4. Infer gene regulatory networks



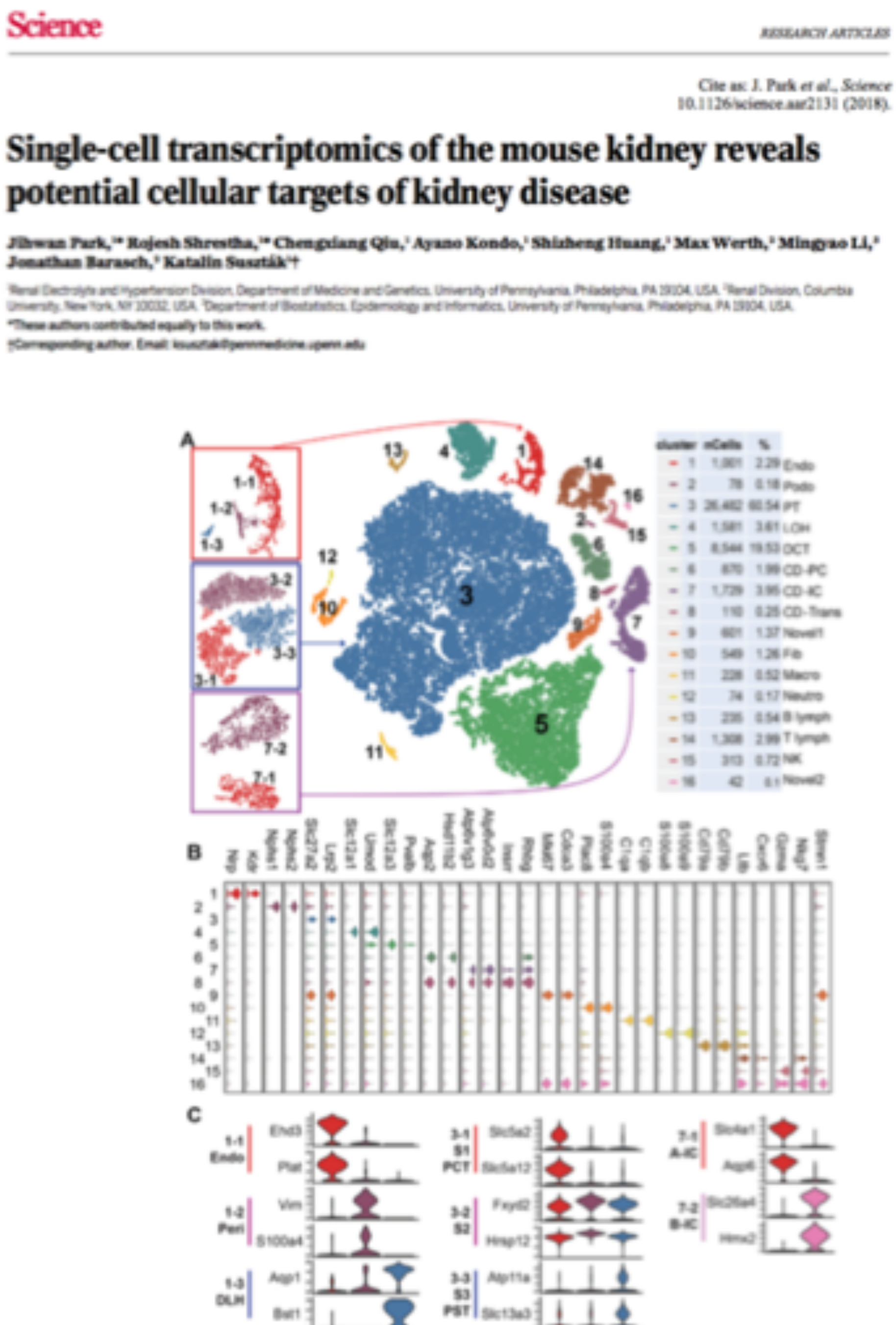
Adapted from Liu and Trapnell (2016)
10.12688/f1000research.7223.1

Assessing Cell-to-Cell Heterogeneity

- Understand composition of complex cell mixtures
- Discover rare cell types
- Determine ratios of cell types within a complex tissue or tumor
- Determine specific cell types driving a disease pathology

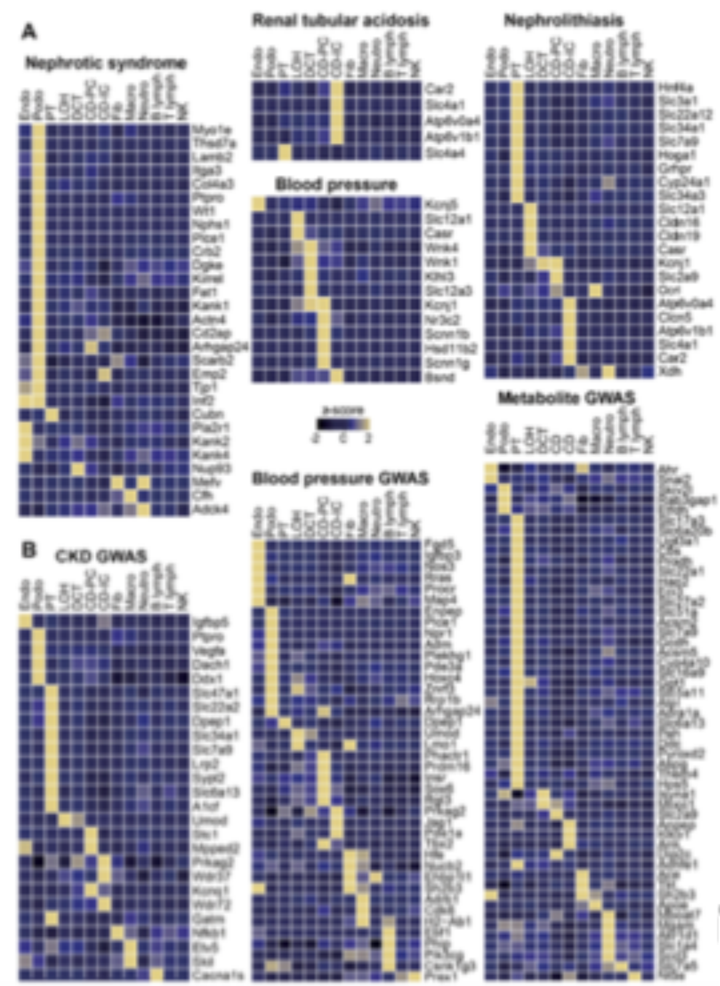


Trapnell, *et. al. F1000 Research,* 2016



Classified >57K cells by scRNA-Seq

Compared genetics of human disease to transcriptional readout of single cells - in several cases, the genes indicative of a particular disease are only expressed in a single cell type



Hot Single-Cell Applications

Single-Cell Pooled CRISPR Screens

Single-Nuclei sequencing

Single-Cell T Cell or B Cell Receptor Sequencing

Single-Cell Epitope Detection

Single-Cell Multiplexing, Multiplet Detection, and Batch Effect

Single-Cell Preservation methods and considerations

Single-Cell ATAC-seq

SNV detection in Single Cell DNA

Single-Cell Pooled CRISPR Screens

Cell

Volume 167, Issue 7, 15 December 2016, Pages 1883–1896.e15

Resource

Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq

Diego Adhemar Jaitin[1,5], Assaf Weiner[1,2,5], Ido Yofe[1,5], David Lara-Astiaso[1], Hadas Keren-Shaul[1], Eyal David[1], Tomer Meir Salame[3], Amos Tanay[4], Alexander van Oudenaarden[2], Ido Amit[1,6]

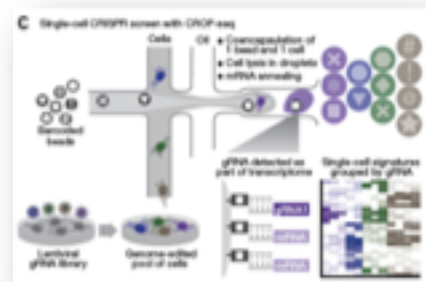
CRISP-seq

Pooled CRISPR screening with single-cell transcriptome readout

Paul Datlinger, André F Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar & Christoph Bock

Affiliations | Contributions | Corresponding author

Nature Methods 14, 297–301 (2017) | doi:10.1038/nmeth.4177
Received 11 October 2016 | Accepted 10 January 2017 | Published online 18 January 2017



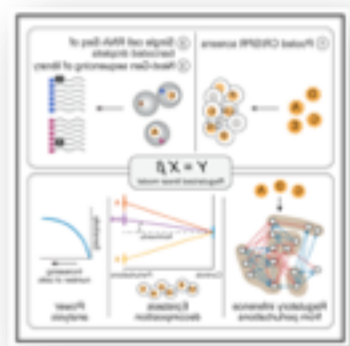
CROP-seq

Resource

Cell

Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens

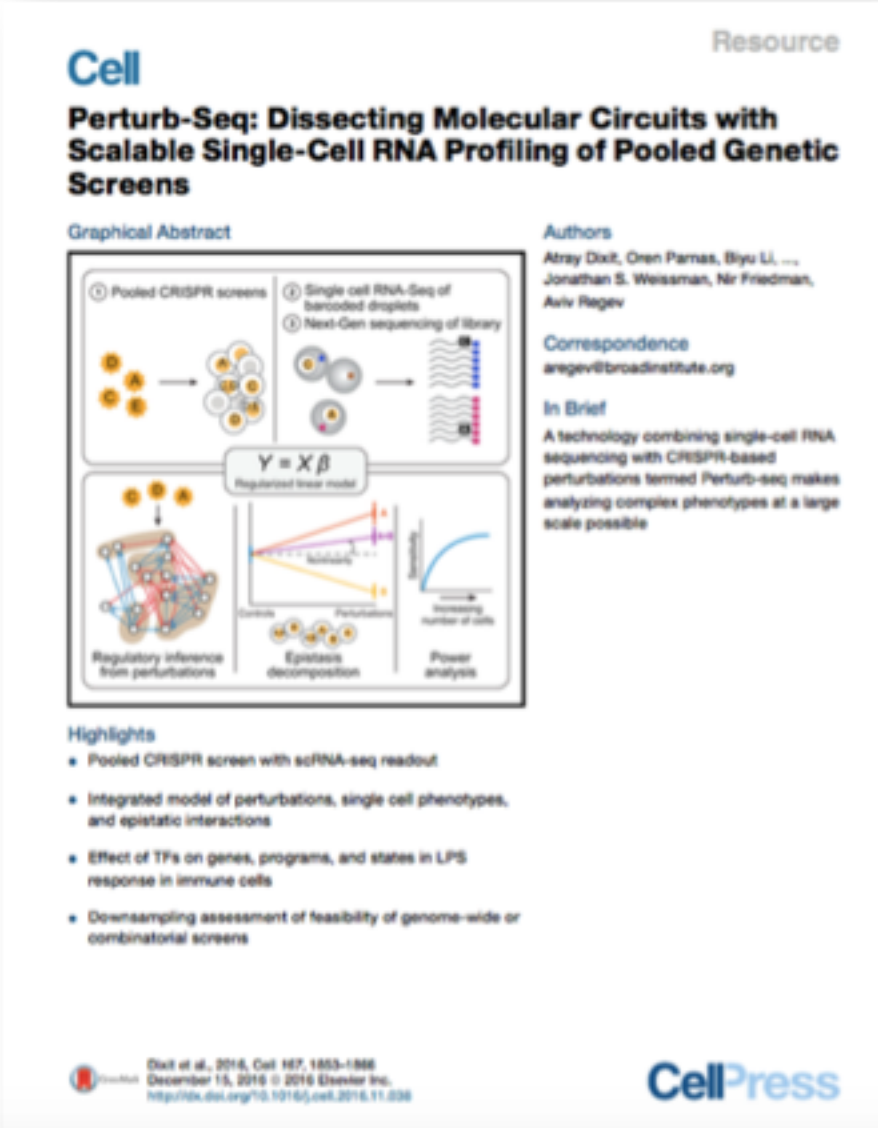
Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P. Fulco, Livnat Jerby-Arnon, Nemanja D. Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, Britt Adamson, Thomas M. Norman, Eric S. Lander, Jonathan S. Weissman, Nir Friedman, and Aviv Regev



Perturb-seq

illumina

Combining CRISPR and Single Cells



Resource

Cell

Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens

Graphical Abstract

Authors

Atray Dixit, Oren Parnas, Biyu Li, ..., Jonathan S. Weissman, Nir Friedman, Aviv Regev

Correspondence

aregev@broadinstitute.org

In Brief

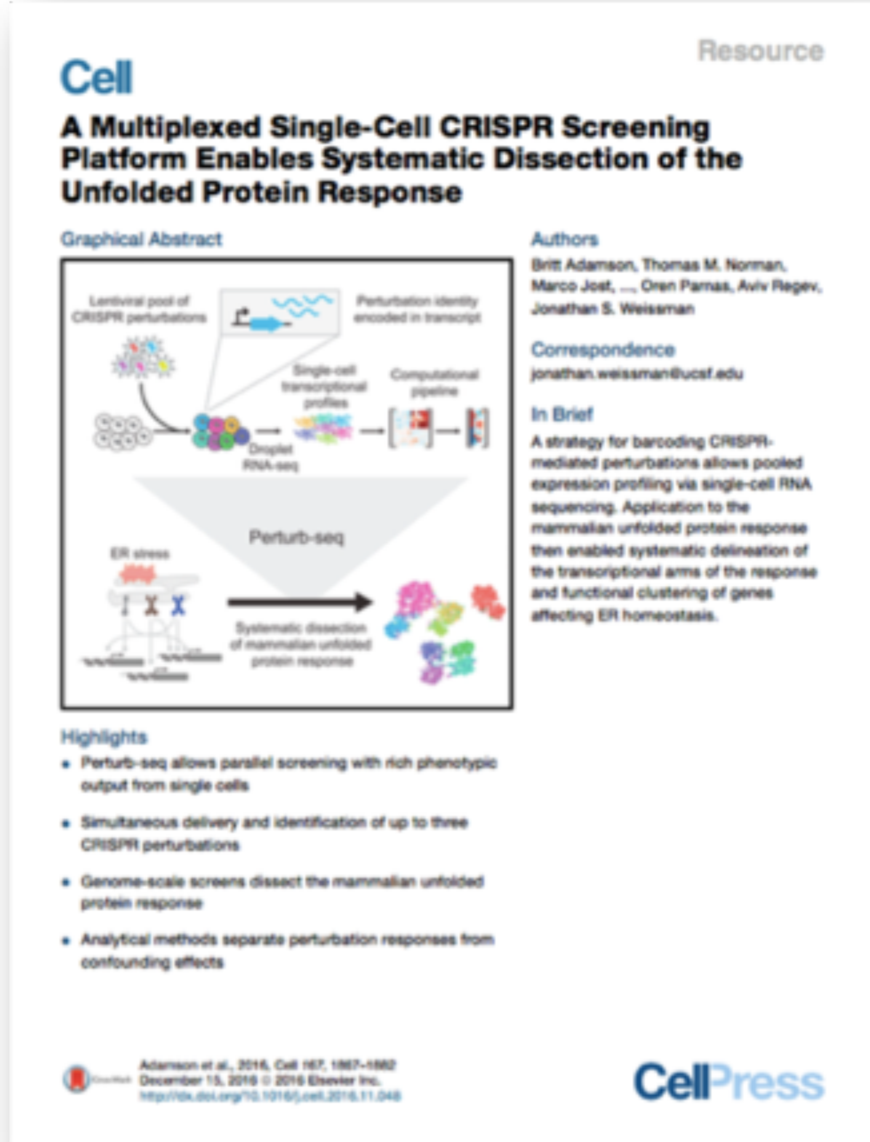
A technology combining single-cell RNA sequencing with CRISPR-based perturbations termed Perturb-seq makes analyzing complex phenotypes at a large scale possible

Highlights

- Pooled CRISPR screen with scRNA-seq readout
- Integrated model of perturbations, single cell phenotypes, and epistatic interactions
- Effect of TFs on genes, programs, and states in LPS response in immune cells
- Downsampling assessment of feasibility of genome-wide or combinatorial screens

Dixit et al., 2016, Cell 167, 1853–1866
December 15, 2016 © 2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2016.11.038

CellPress



Resource

Cell

A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response

Graphical Abstract

Authors

Britt Adamson, Thomas M. Norman, Marco Jost, ..., Oren Parnas, Aviv Regev, Jonathan S. Weissman

Correspondence

jonathan.weissman@ucsf.edu

In Brief

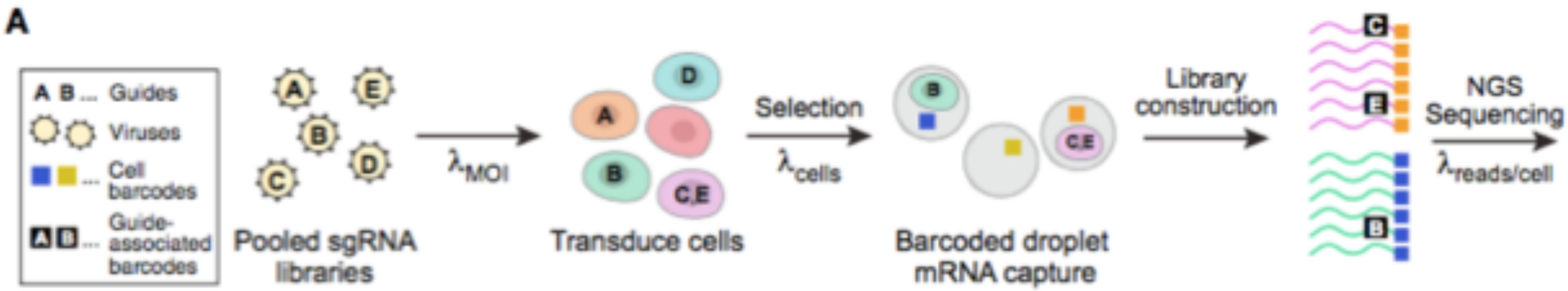
A strategy for barcoding CRISPR-mediated perturbations allows pooled expression profiling via single-cell RNA sequencing. Application to the mammalian unfolded protein response then enabled systematic delineation of the transcriptional arms of the response and functional clustering of genes affecting ER homeostasis.

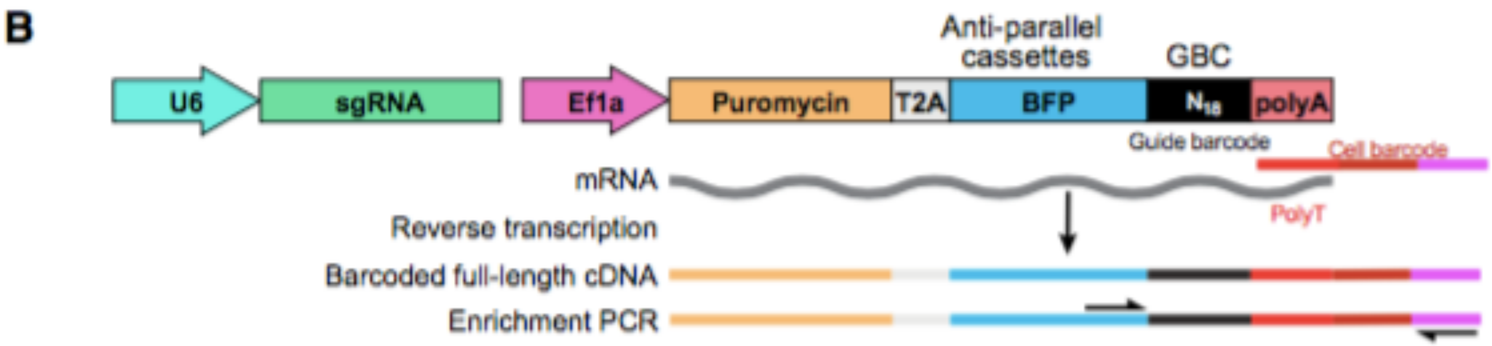
Highlights

- Perturb-seq allows parallel screening with rich phenotypic output from single cells
- Simultaneous delivery and identification of up to three CRISPR perturbations
- Genome-scale screens dissect the mammalian unfolded protein response
- Analytical methods separate perturbation responses from confounding effects

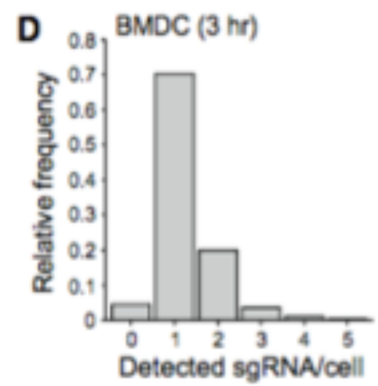
Adamson et al., 2016, Cell 167, 1867–1882
December 15, 2016 © 2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2016.11.048

CellPress





Cell type	sgRNA pool	Total cells	Time points
Mouse BMDC	Transcription factors (67 guides)	70,000	0 and 3 hr post-LPS
Human K562	Transcription factors (46 guides)	104,000	7 and 13 days
Human K562	Cell cycle regulators (36 guides)	26,000	7 days



http://dx.doi.org/10.1016/j.cell.2016.11.038

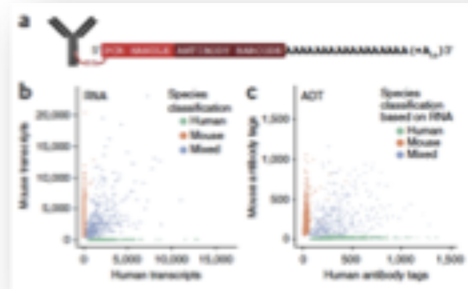
Single-Cell Epitope Detection

Simultaneous epitope and transcriptome measurement in single cells

Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija & Peter Smibert

Affiliations | Contributions | Corresponding author

Nature Methods **14**, 865–868 (2017) | doi:10.1038/nmeth.4380
Received 02 March 2017 | Accepted 07 July 2017 | Published online 31 July 2017

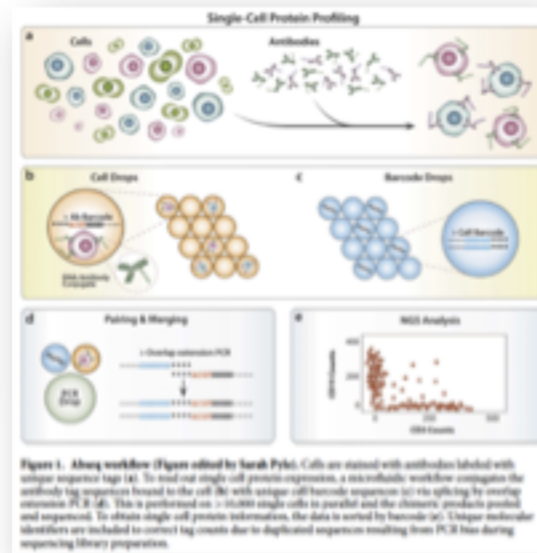


CITE-seq

Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding

Payam Shahi, Samuel C. Kim, John R. Haliburton, Zev J. Gartner & Adam R. Abate

Scientific Reports 7, Article number: 44447 (2017)
doi:10.1038/srep44447
Received: 02 November 2016
Accepted: 08 February 2017
Published online: 14 March 2017



Abseq

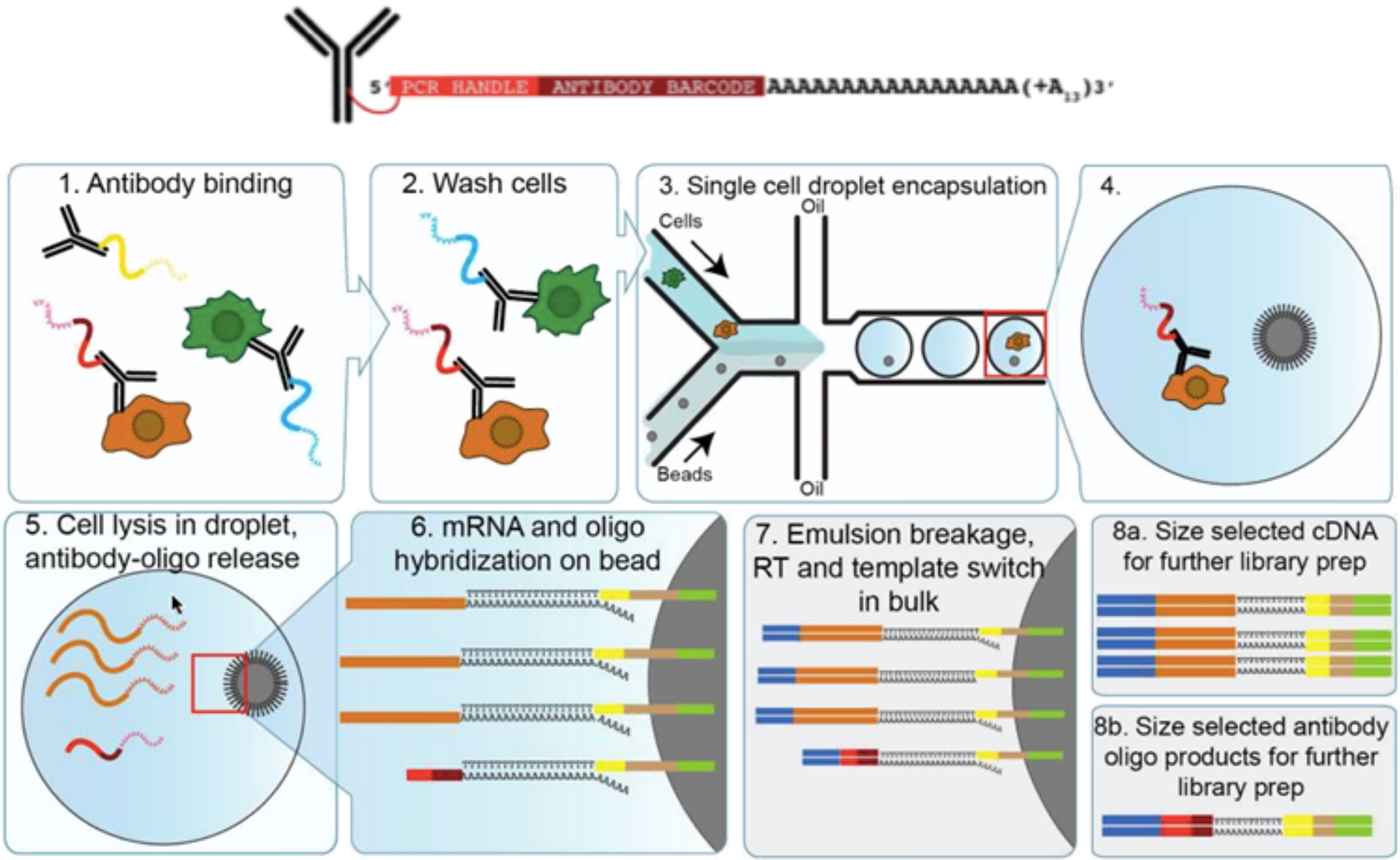
Multiplexed quantification of proteins and transcripts in single cells

Vanessa M Peterson[1,5], Kelvin Xi Zhang[2,5], Namit Kumar[1], Jerelyn Wong[3], Lixia Li[1], Douglas C Wilson[3], Renee Moore[4], Terrill K McClanahan[3], Svetlana Sadekova[3] & Joel A Klappenbach[1]

REAP-seq

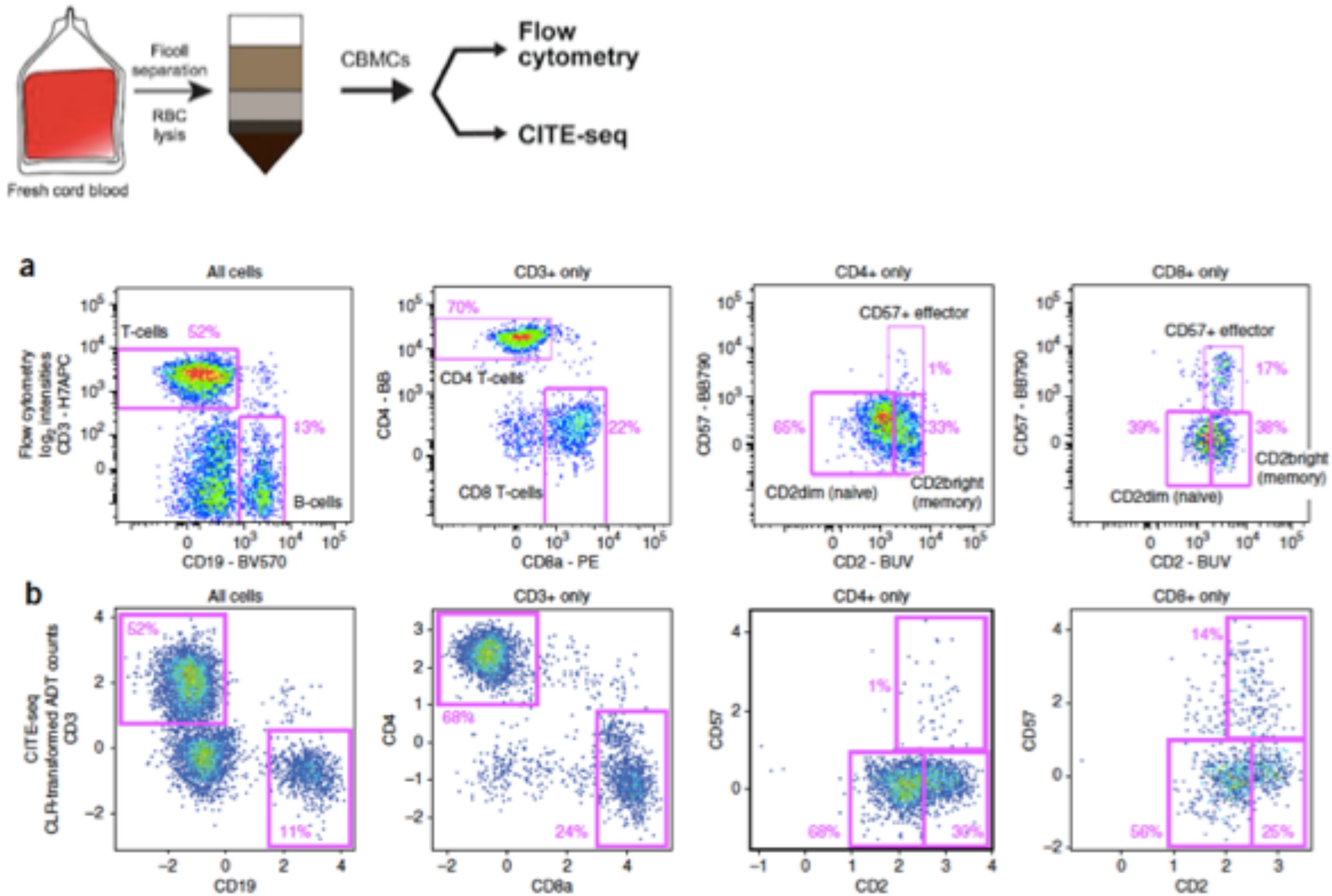
CITE-seq workflow

Cellular Indexing of Transcriptomes and Epitopes by sequencing



Stoeckius NATURE METHODS | VOL.14 NO.9 | SEPTEMBER 2017 | 865

Concordance between Flow and CITE-seq

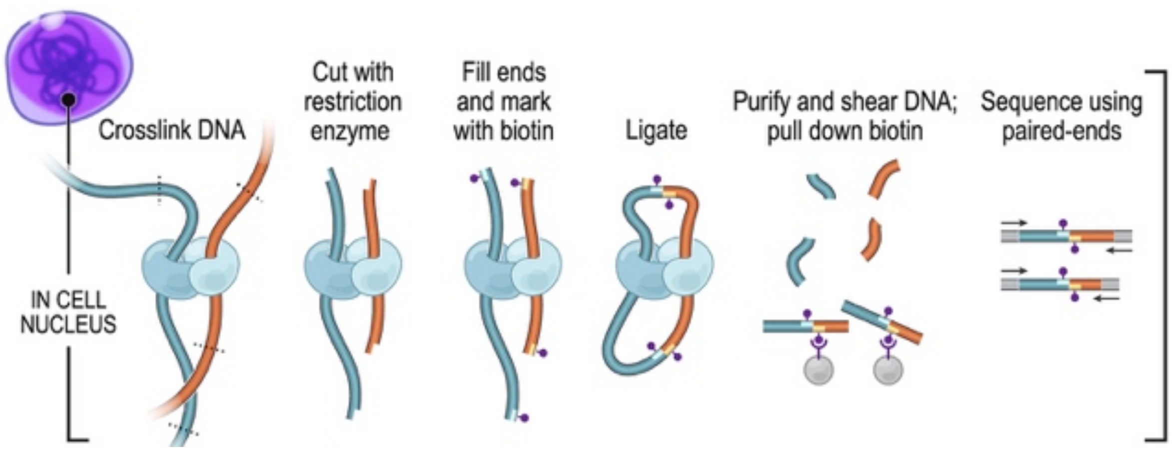


Stoeckius NATURE METHODS | VOL.14 NO.9 | SEPTEMBER 2017 | 865

Hi-C Sequencing

Hi-C Sequencing

- **Chromatin conformation capture sequencing**
- **Used to analyze chromatin interactions**
  - DNA/protein complexes are crosslinked
  - sample is fragmented and DNA ligated and digested
  - DNA fragments are PCR-amplified and sequenced



Hi-C Sequencing in Metagenomics

Species-Level Deconvolution of Metagenome Assemblies with Hi-C–Based Contact Probability Maps

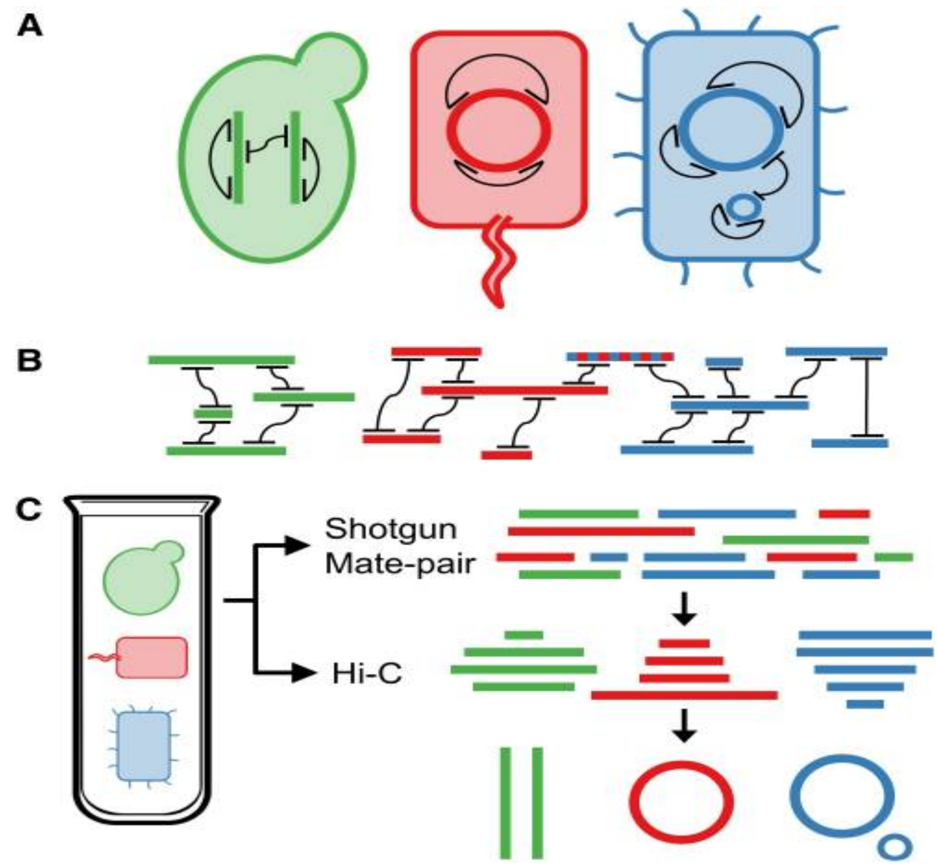
Joshua N. Burton,[1] Ivan Liachko,[1] Maitreya J. Dunham,[2] and Jay Shendure[2]

- Demonstrated that Hi-C provides a signal of contiguity that is intracellular and contains intra and inter chromosomal information
- Signal from Hi-C can be used to reconstruct individual genomes of microbial species present within a metagenomic sample
- Results of the clustering of fungal, bacterial, and archaelal species were 99% concordant with published reference genomes
- Hi-C can also be used to create scaffolded genome assemblies of individual species present in the community

Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-Level Deconvolution of Metagenome Assemblies with Hi-C–Based Contact Probability Maps. G3: Genes|Genomes|Genetics 4, 1339–1346 (2014).

Hi-C Sequencing in Metagenomics

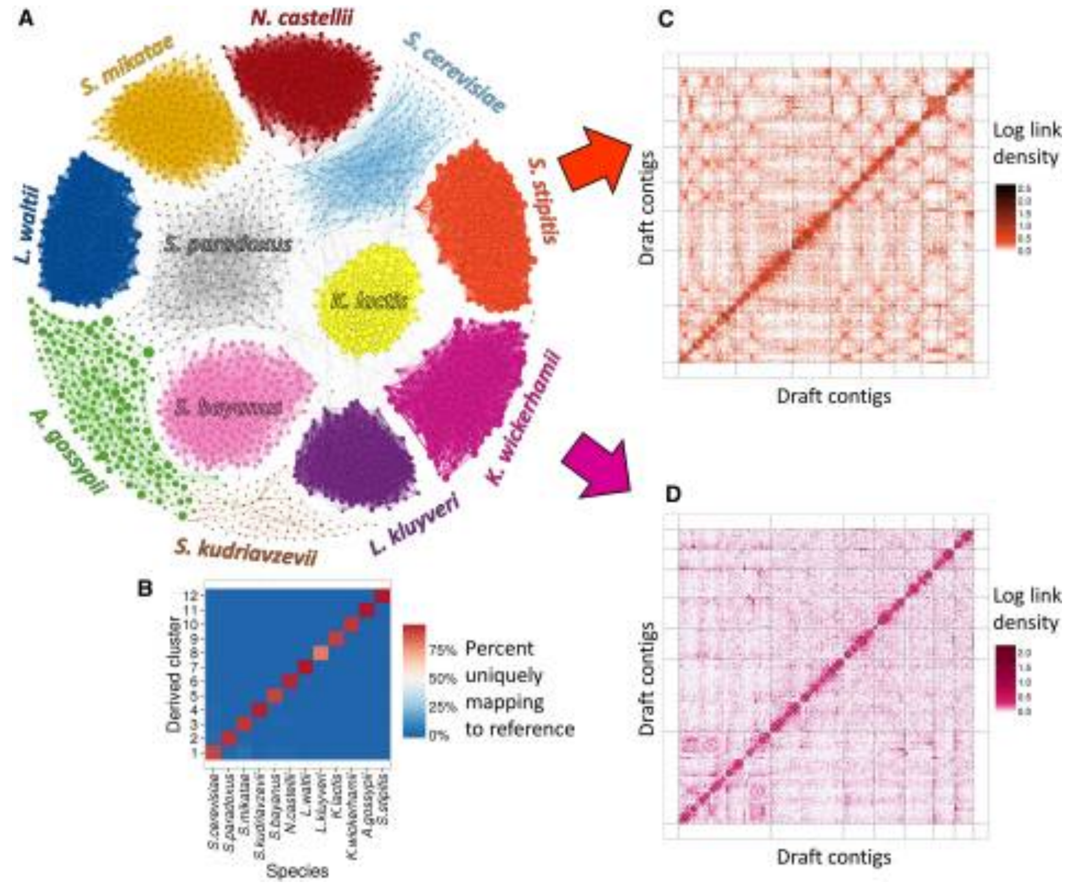
- Library prep done with Illumina Nextera and Illumina Nextera Mate Pair
- 2 x100 bp read length
- Sequenced on HiSeq
- 81-85M Reads per sample (Yeast Mixture sample) for Hi-C library
- 92M Reads from shotgun library
- Shotgun and mate-pair libraries used to generate a draft *de novo* metagenome assembly



Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-Level Deconvolution of Metagenome Assemblies with Hi-C–Based Contact Probability Maps. G3: Genes|Genomes|Genetics 4, 1339–1346 (2014).

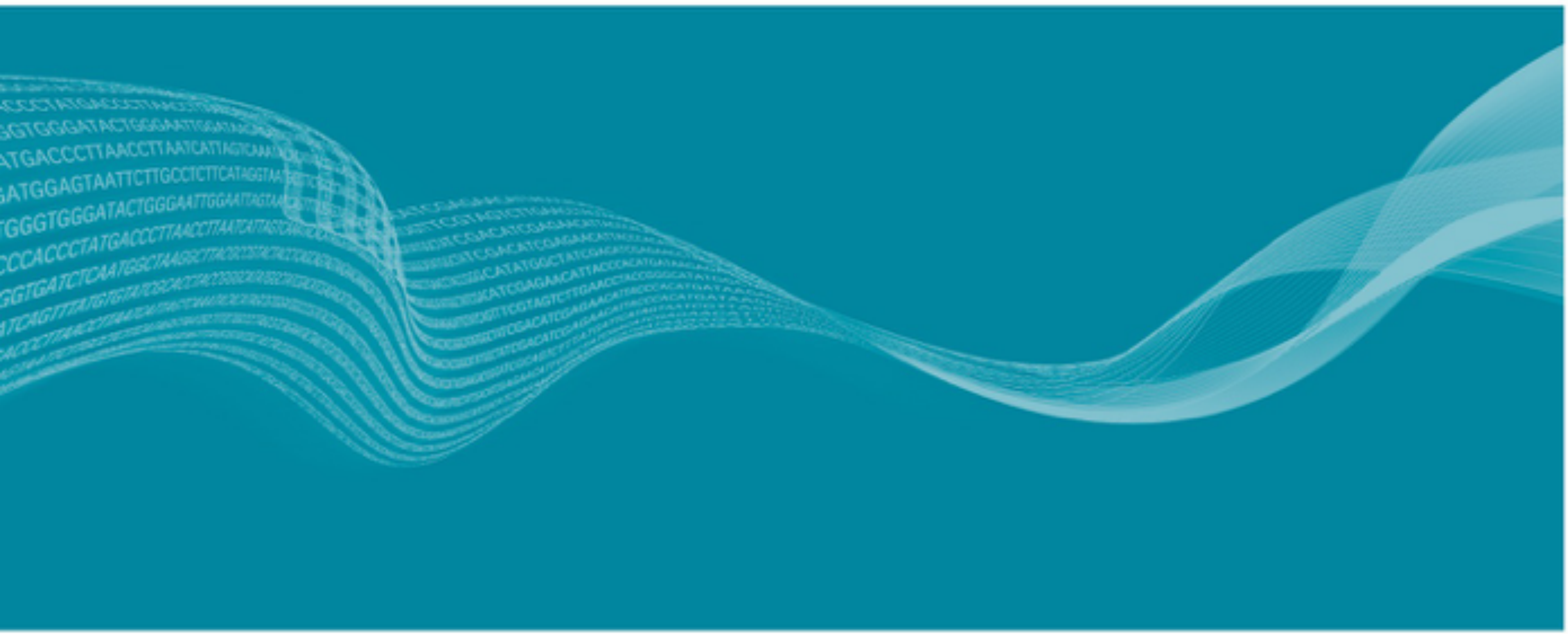
Hi-C Sequencing in Metagenomics

- Contact probability maps from Hi-C enable deconvolution of shotgun metagenomic assemblies
- Hi-C enables two different signals
  - Intracellularity of each pair which enables species level deconvolution
  - Correlation of Hi-C linkage with chromosomal distance, which enbales scaffolding of *de novo* assemblies



Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-Level Deconvolution of Metagenome Assemblies with Hi-C–Based Contact Probability Maps. G3: Genes|Genomes|Genetics 4, 1339–1346 (2014).

Epigenetics



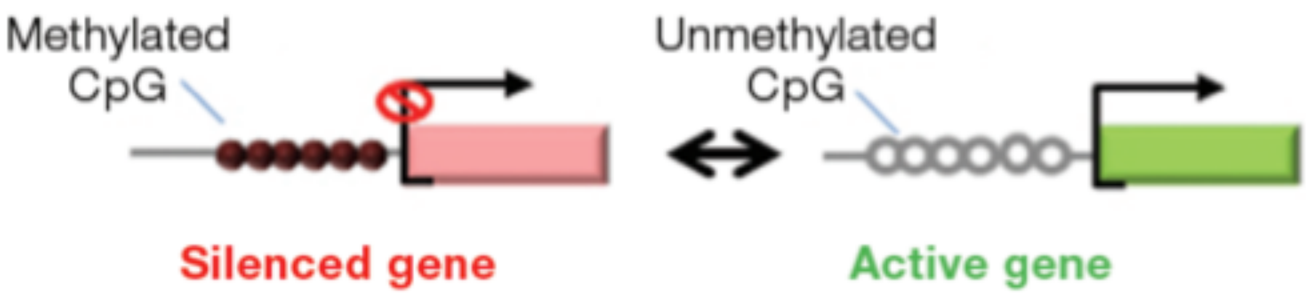
For Research Use Only. Not for use in diagnostic procedures.

Epigenetics

The study of changes in organisms caused by modification of gene expression rather than alteration of the genetic code itself.

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)

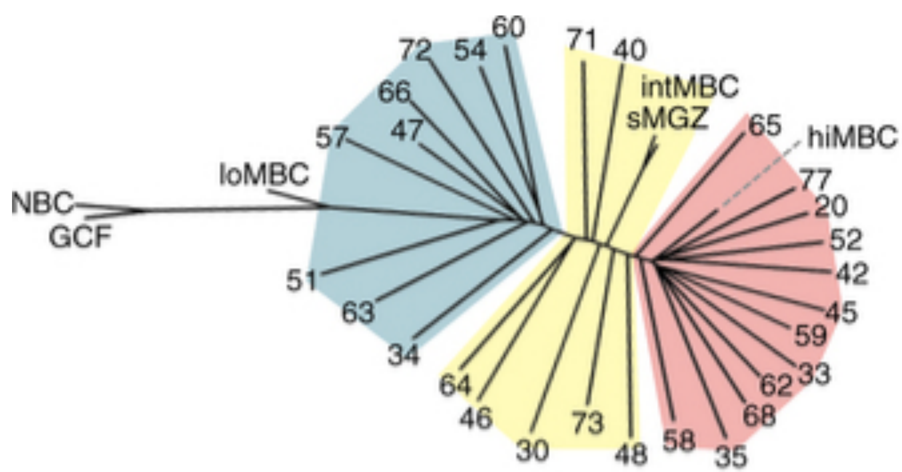
Momparler Seminars in oncology 2005

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

Turecki et al., Biological psychiatry 2016

DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia

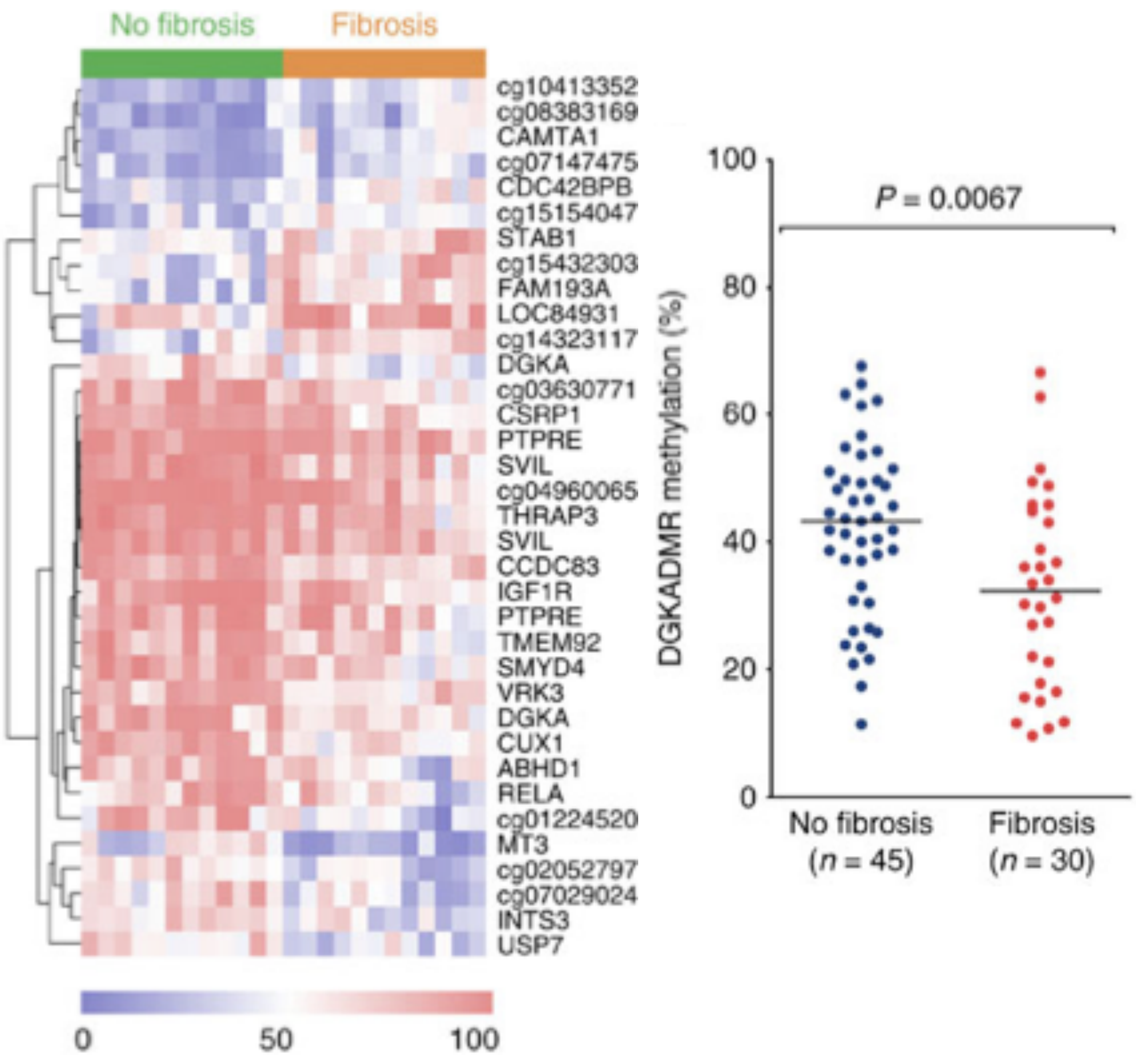
- Combined analysis of **methylation array, deep methylation sequencing, and RNA-Seq**
- Identified methylation dysregulation during CLL (chronic lymphocytic leukemia).
- Conducted RNA-Seq on methylation-mapped cell phylogenies
- Mapped the dysregulated pathway based on expression changes



Nature Genetics 48, 253–264 (2016)

DNA Methylation affects Breast Cancer Treatment

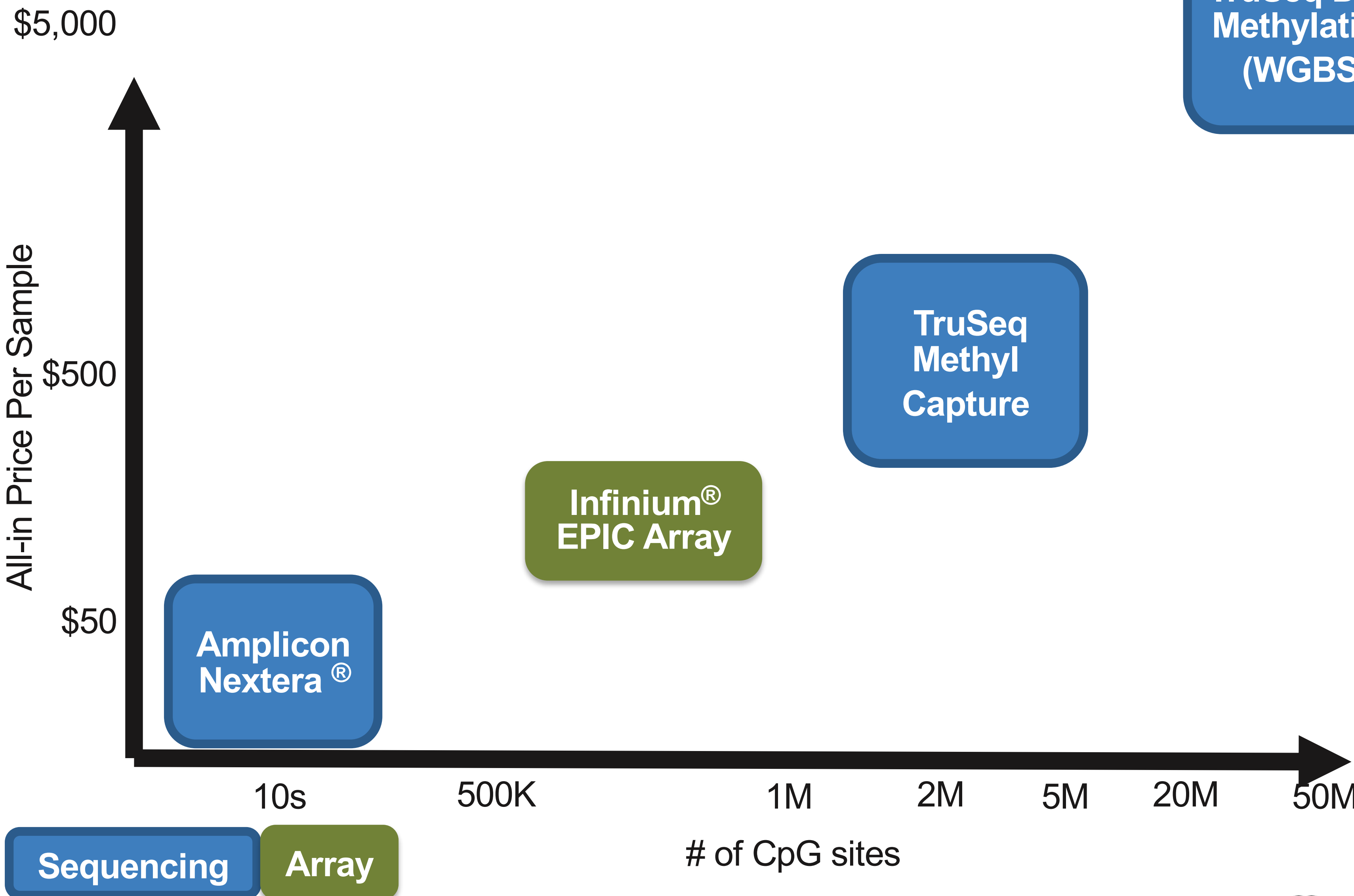
- Radiation of breast cancer can result in tissue fibrosis, limiting treatment
- Researchers used a large scale **methylation array screen** correlated with **RNA-Seq data** and **ChIP-Seq data**
- Identified *DGKA*, an enhancer involved in profibrotic transcription factor activation, as a key regulator of fibrosis and potential therapeutic target



Weigel, Christoph, et al. "Epigenetic regulation of diacylglycerol kinase alpha promotes radiation-induced fibrosis." Nature communications 7 (2016).

Methylation Analysis Technology

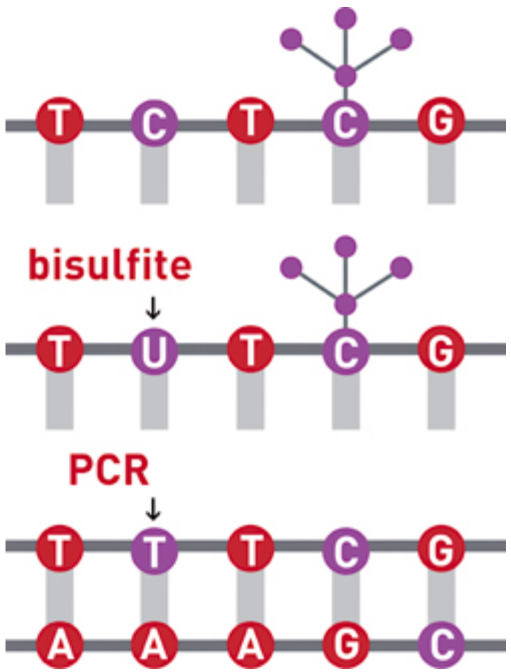
The bang for your buck model



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

How does this technology work?

Methylation Analysis With Bisulfite Conversion

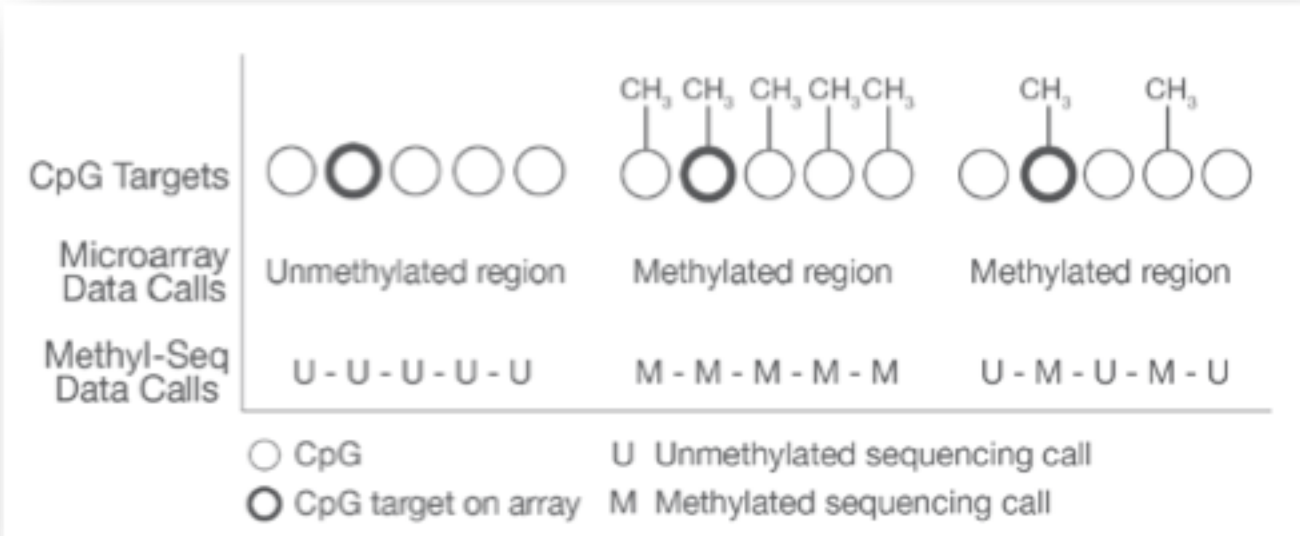


Methylation Arrays:
Identify the difference between a C and T at **~850,000 discreet sites of interest**

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

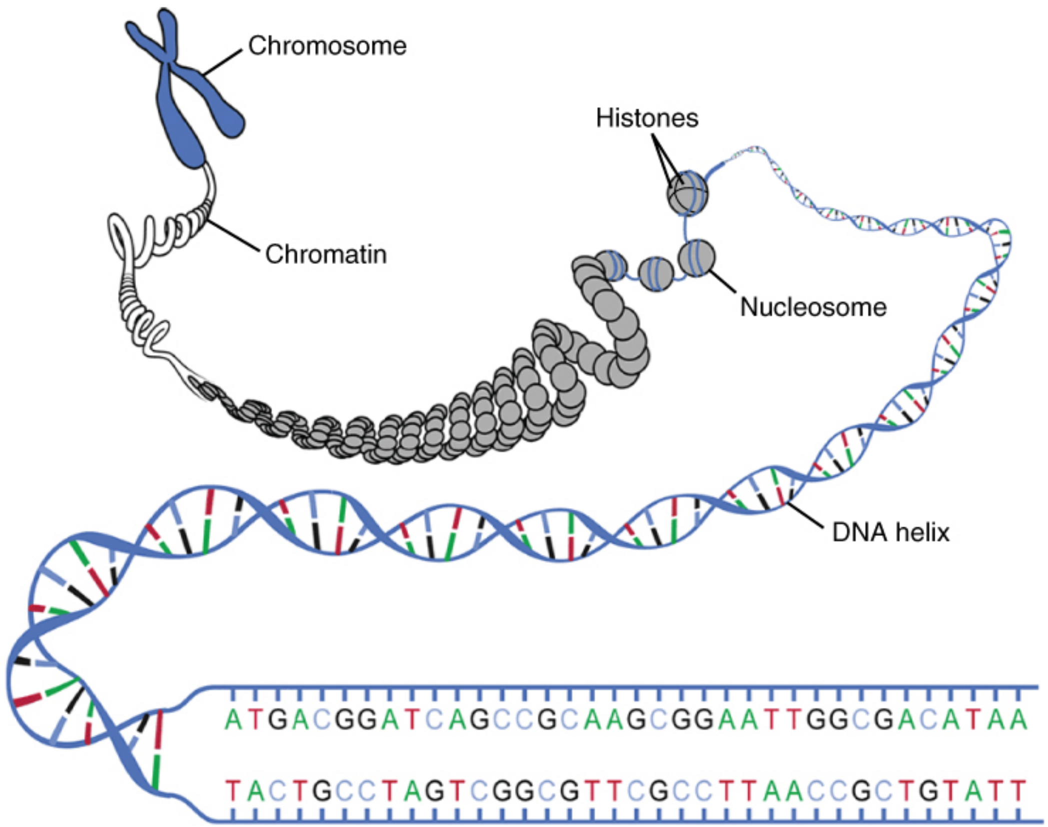
illumina

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

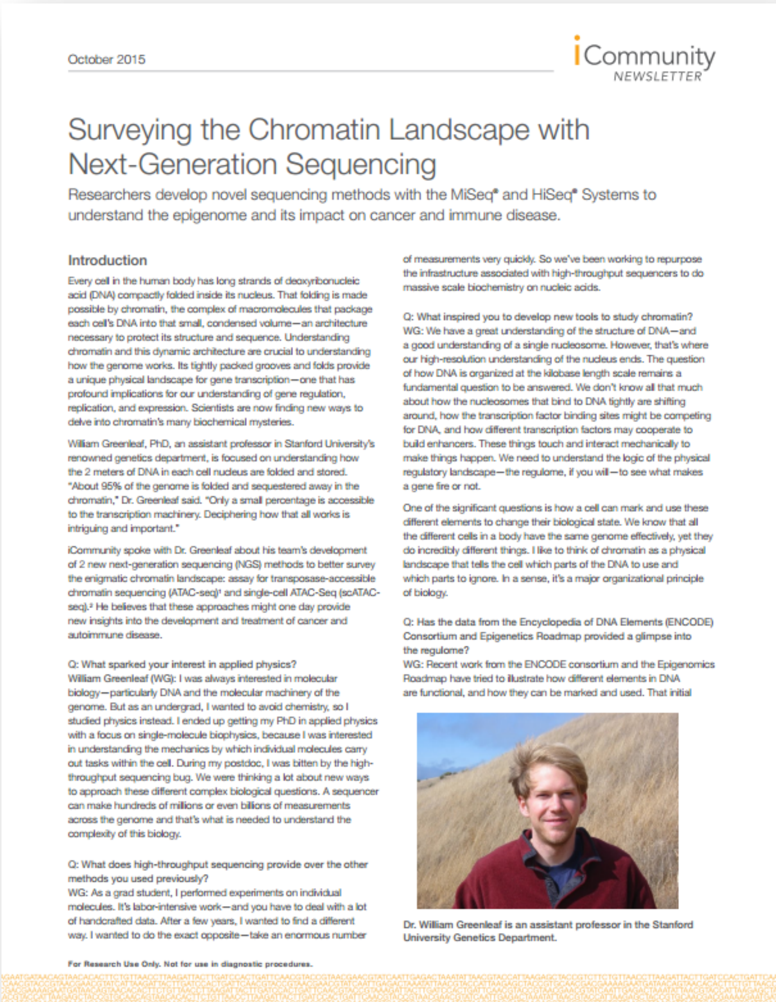
ATAC-seq



Chromosome
Histones
Chromatin
Nucleosome
DNA helix
ATGACGGATCAGCCGCAAGCGGAATTGGCGACATAA
TACTGCCTAGTCGGCGTTCGCCTTAACCGCTGTATT

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.

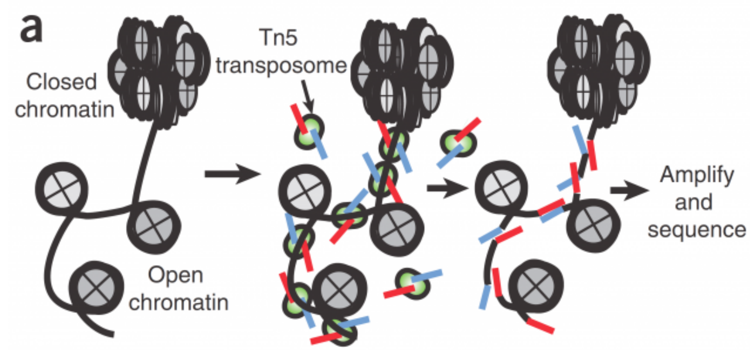
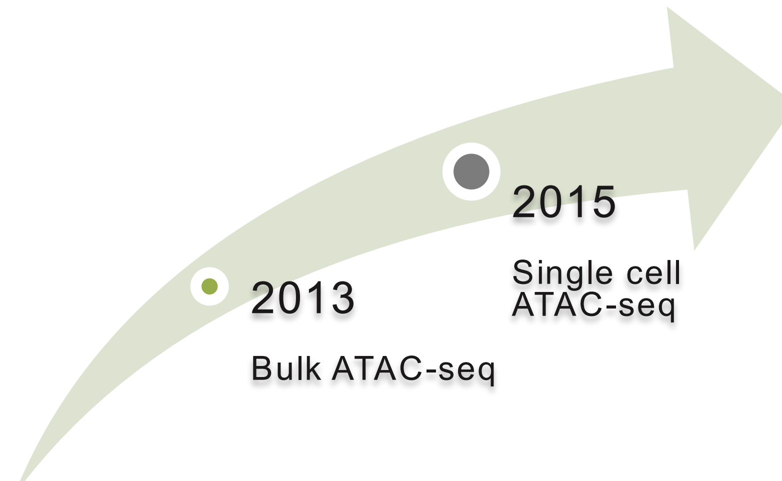


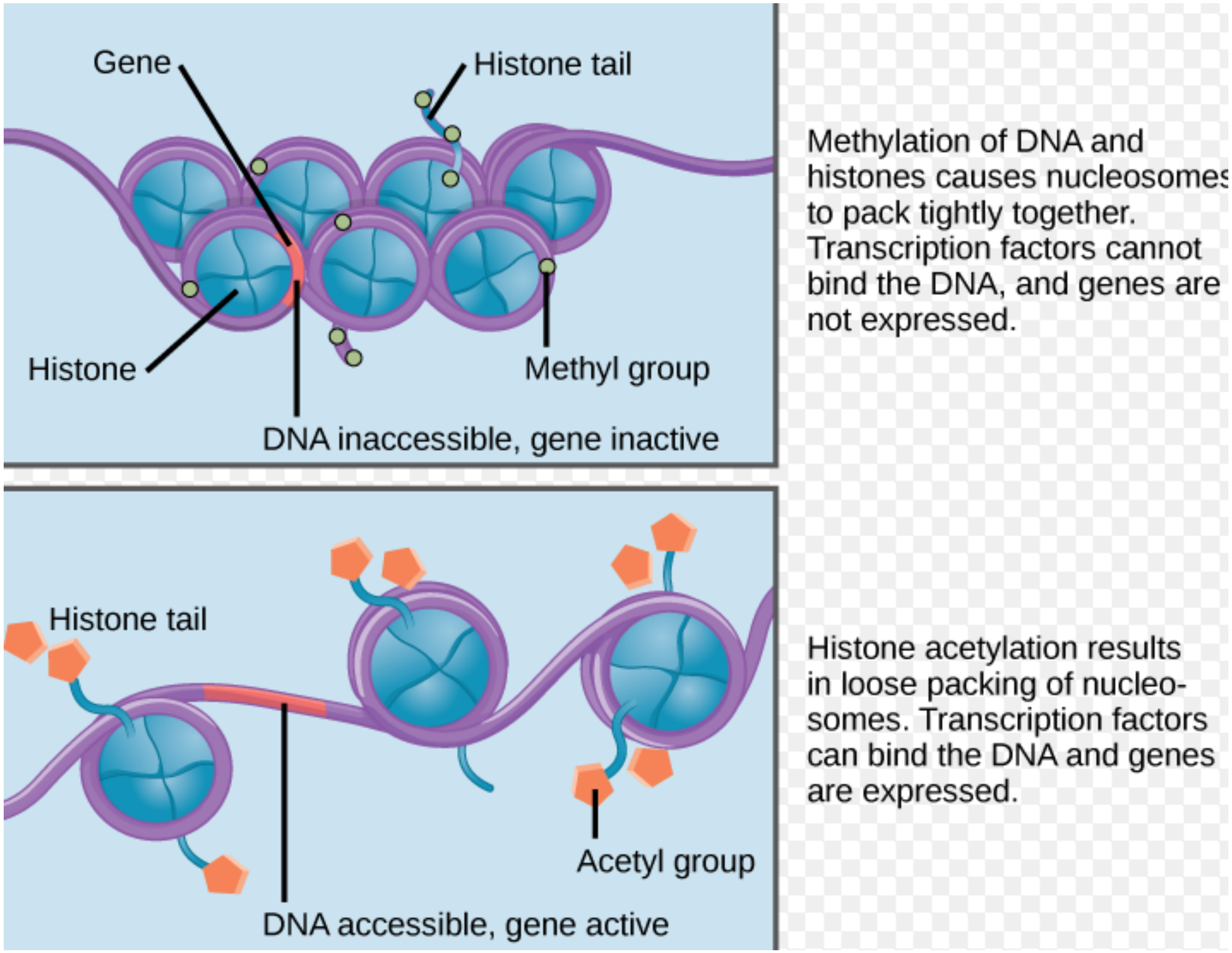
Figure 1. Scheme for ATAC-seq technic. Transposase enzyme (green), bearing sequencing adaptors (red and blue), is incorporated only in regions of open chromatin (between nucleosomes in grey). Allowing to amplify those open regions by PCR. | Credit: Buenrostro et al. 2013. Nat. Methods 10, 1213–8.



** Prof Greenleaf co-founded Epinomics company*

illumina®

Why do researchers want to look at chromatin structures?

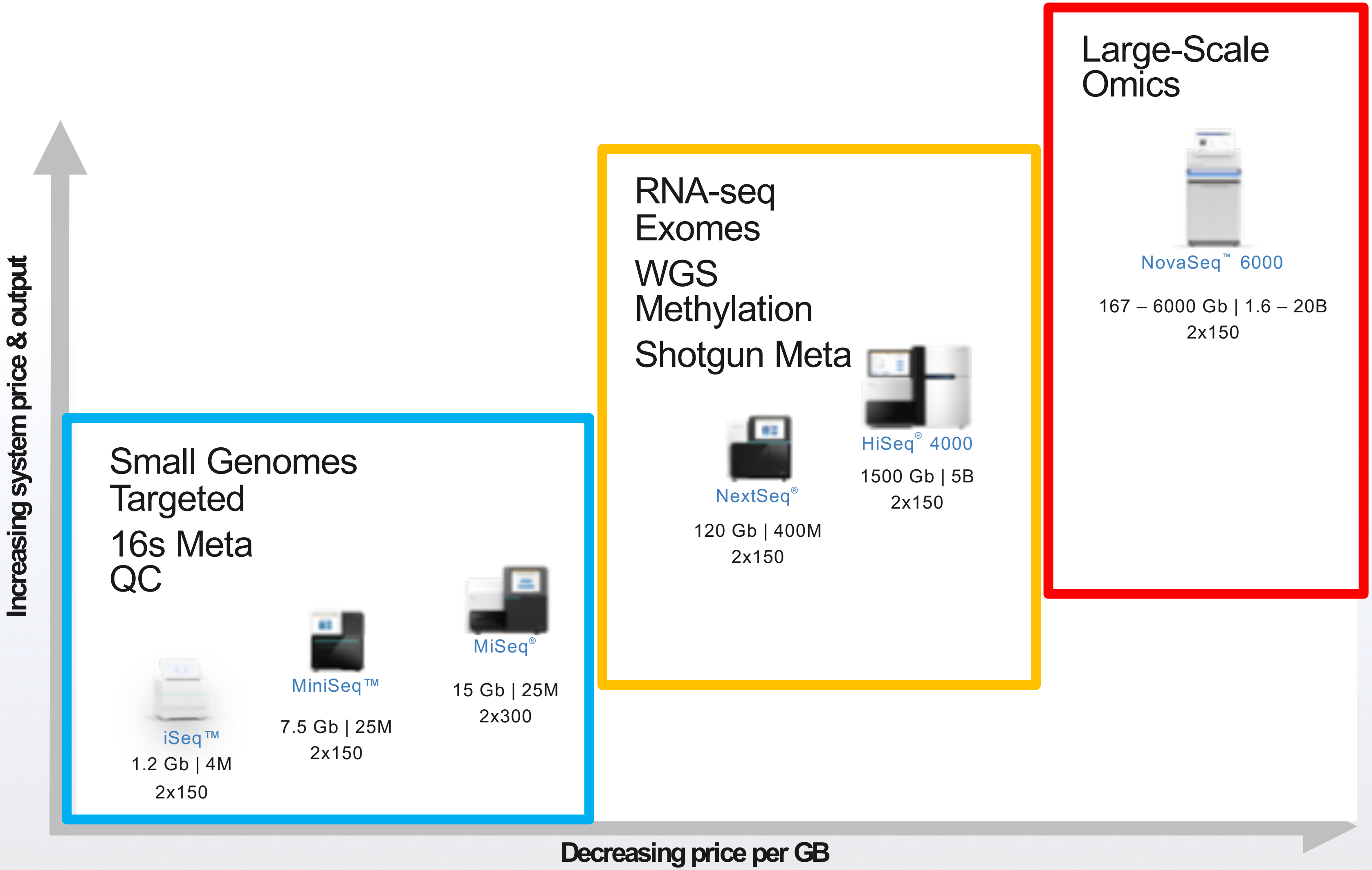


Application examples using ATAC-seq

Application	Nucleosome mapping	Transcription factor occupancy analysis in specific cell types	Identify novel enhancers during development	Explore various pathological conditions
Example and publication	Identify changes in nucleosome position btwn experimental conditions and correlation with sequence context. **Schep et al. (2015)**	Find lineage-specific factors during hematopoesis. **Lara-Astiaso et al. (2014)**	Explore evolution of neural crest cis-regulatory element by comparing human and chimp development. **Prescott et al. (2015)**	Identify ectopically-active regions during Ras-dependent oncogenesis. **Davie et al. (2015)**
Market segment	Cell biology	Immunogenetics	Cell biology	Oncology

illumina

Sequencing Power for Virtually Every Scale



Thank You!

mangel@illumina.com