Single-Cell Sequencing in Cancer

Peter A. Sims, Columbia University

G4500: Cellular & Molecular Biology of Cancer

September 22, 2025

Learning Objectives

- Lecture will focus on technology for and applications of single-cell RNA-seq and DNA-seq
- Why are these two tools useful to cancer biologists and oncologists?
- How do they work and what biological questions can they answer?
- What are the key limitations of these technologies?

scDNA-seq

- Co-occurrence patterns for driver mutations / genotypic heterogeneity
- Clonal analysis / tumor cell phylogeny
- Identification of circulating, metastatic, or therapy-resistant clones

scRNA-seq

- Co-occurrence patterns for gene expression / phenotypic heterogeneity
- Differentiation trajectory analysis, identification of transitioning subpopulations
- Analysis of non-genetic mechanisms of metastis and therapy resistance

Bulk DNA-seq

- Can compute allele fractions, but can only infer co-occurrence.
- Rare clone discovery limited by sequencing depth.
- WGS can be used for both CNV and SNV detection.

scDNA-seq

- Direct observation of co-occurring mutations.
- Rare clone discovery limited by cell numbers.
- Difficult to call SNVs and CNVs accurately with the same technique.

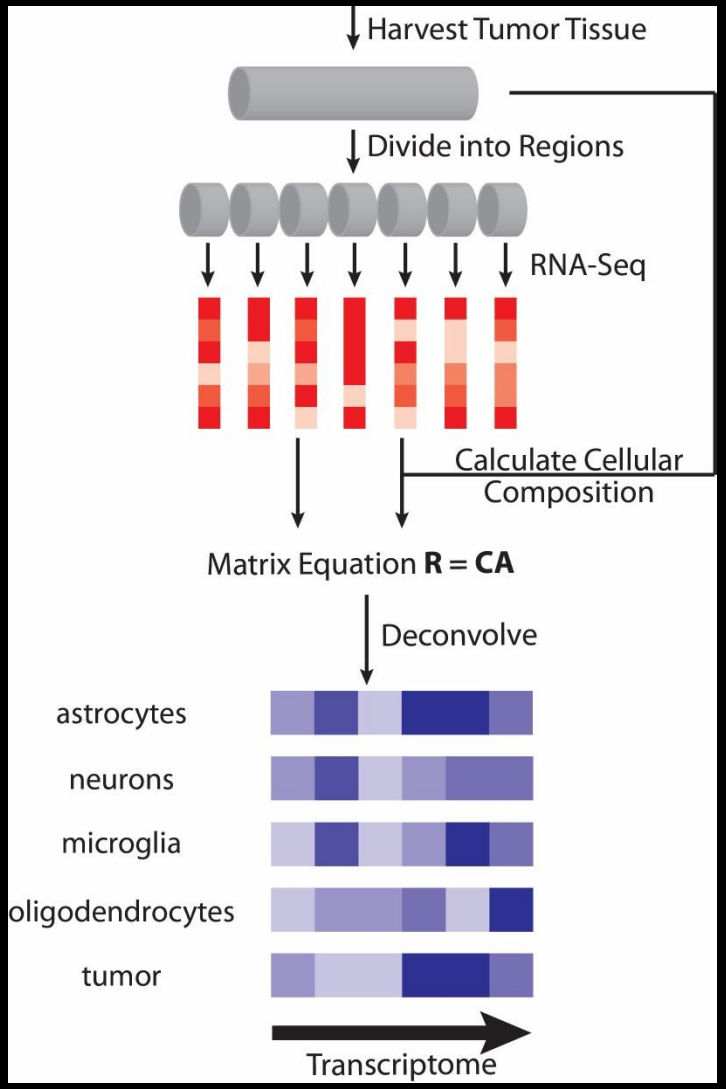
Bulk RNA-seq

- Essentially impossible to assess co-expression.
- Can make inferences about cellular composition by deconvolution.
- Works on homogenized tissue.
- Sensitivity limited mainly by sequencing depth – relatively easy to directly quantify lowly expressed genes.

scRNA-seq

- Allows direct (at best) or imputed (at worst) detection of co-expressed genes.
- Can directly measure cellular composition by unsupervised clustering.
- Requires dissociation to a single-cell suspension.
- Sensitivity limited mainly by molecular capture efficiency – must make inferences about lowly expressed genes.

Computational Deconvolution of Gene Expression



Total Expression Level of a Gene is the Sum of Contributions from Each Cell Type

$$G = C_1 g_1 + C_2 g_2 + C_3 g_3 + \ldots$$

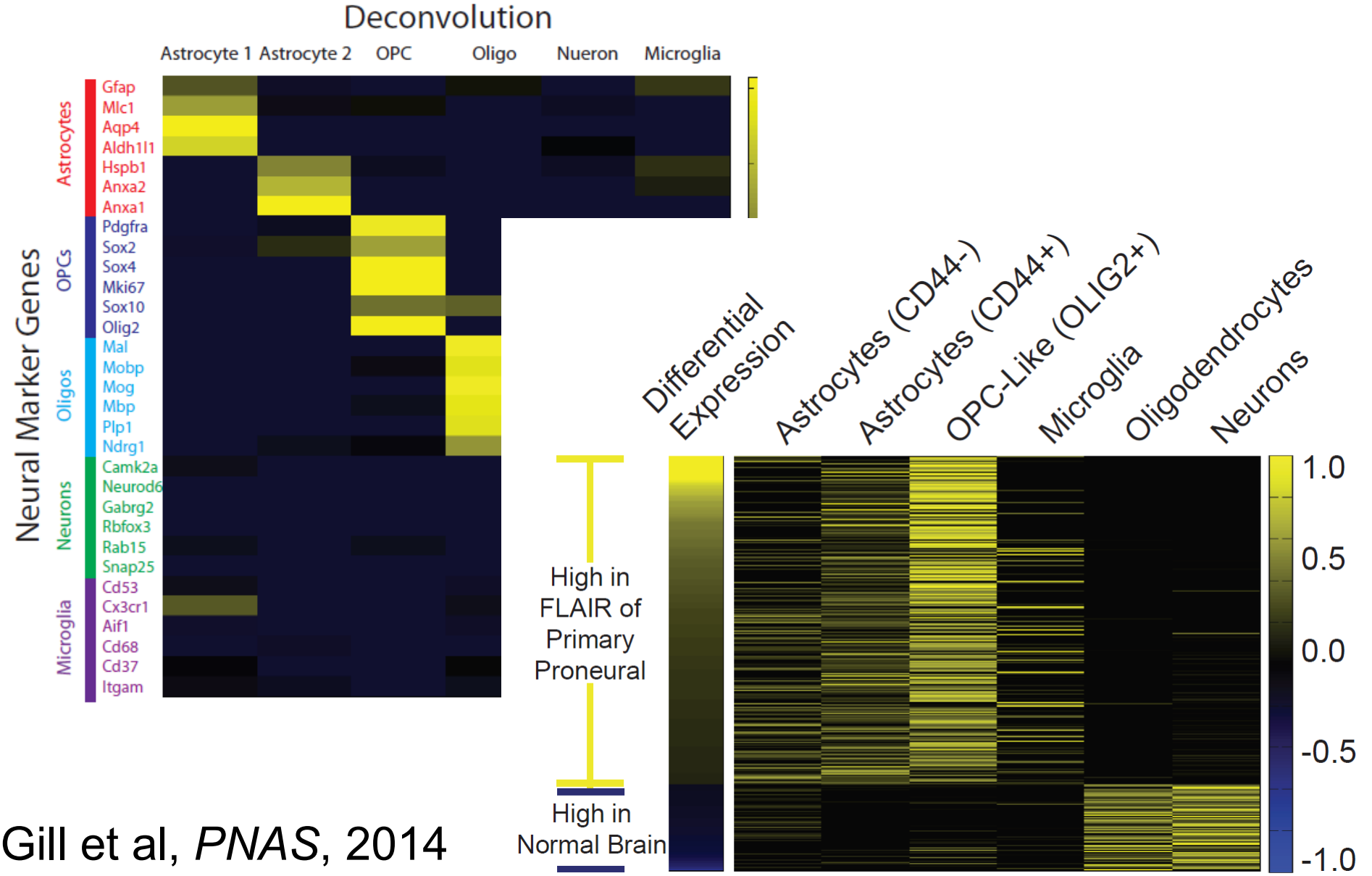
neuron astrocyte microglia

How can we solve this equation?

What if we could had this equation for many samples?

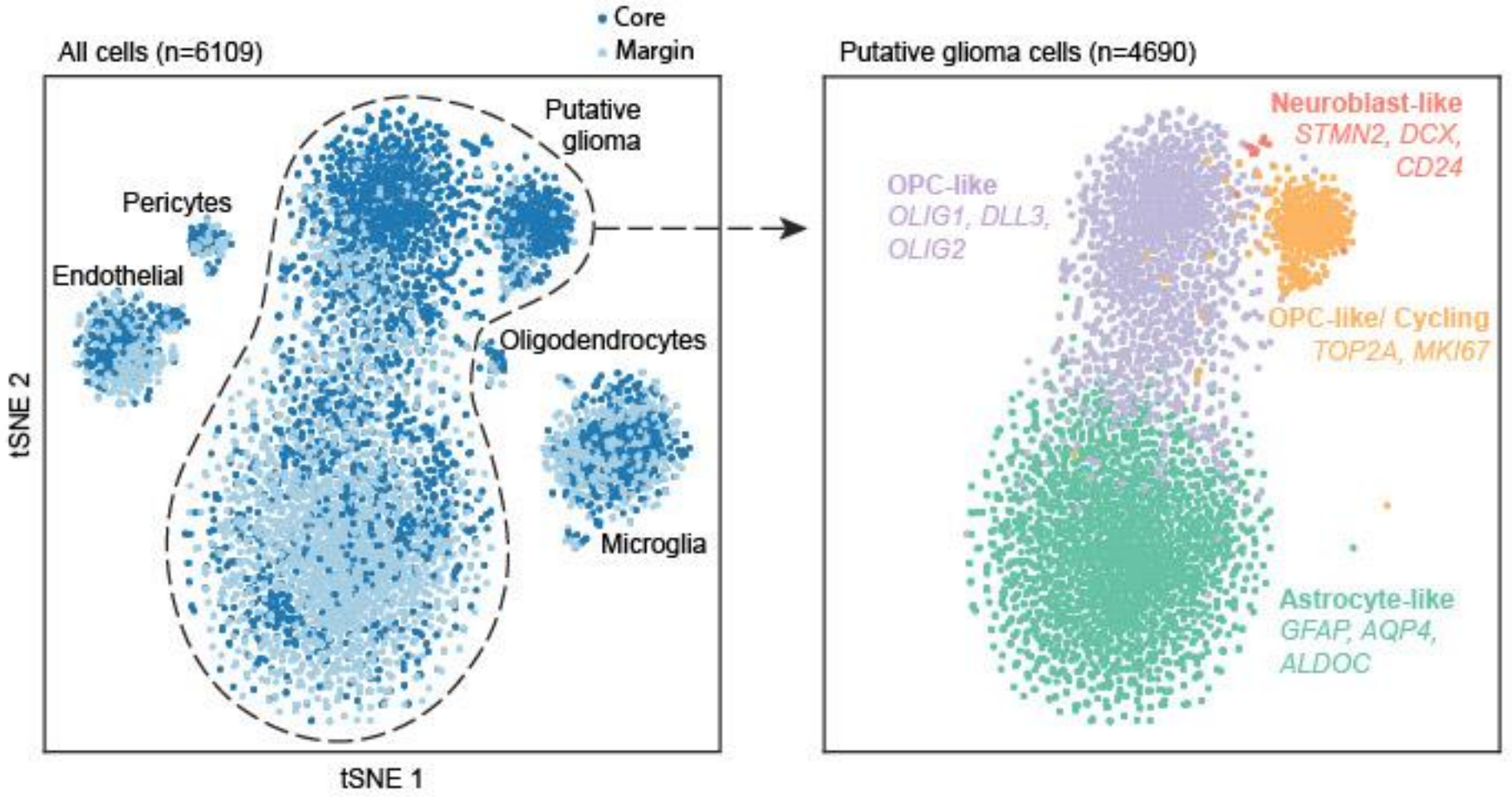
$$G_1 = C_{11} g_1 + C_{12} g_2 + C_{13} g_3 + \ldots$$
$$G_2 = C_{21} g_1 + C_{22} g_2 + C_{23} g_3 + \ldots$$
$$G_3 = C_{31} g_1 + C_{32} g_2 + C_{33} g_3 + \ldots$$
$$G_4 = C_{41} g_1 + C_{42} g_2 + C_{43} g_3 + \ldots$$

Deconvolution of Cell Type-Specific Gene Expression in Glioblastoma with Bulk RNA-seq



Gill et al, *PNAS*, 2014

Direct Observation of Cell Type-Specific Gene Expression in Glioblastoma with scRNA-seq

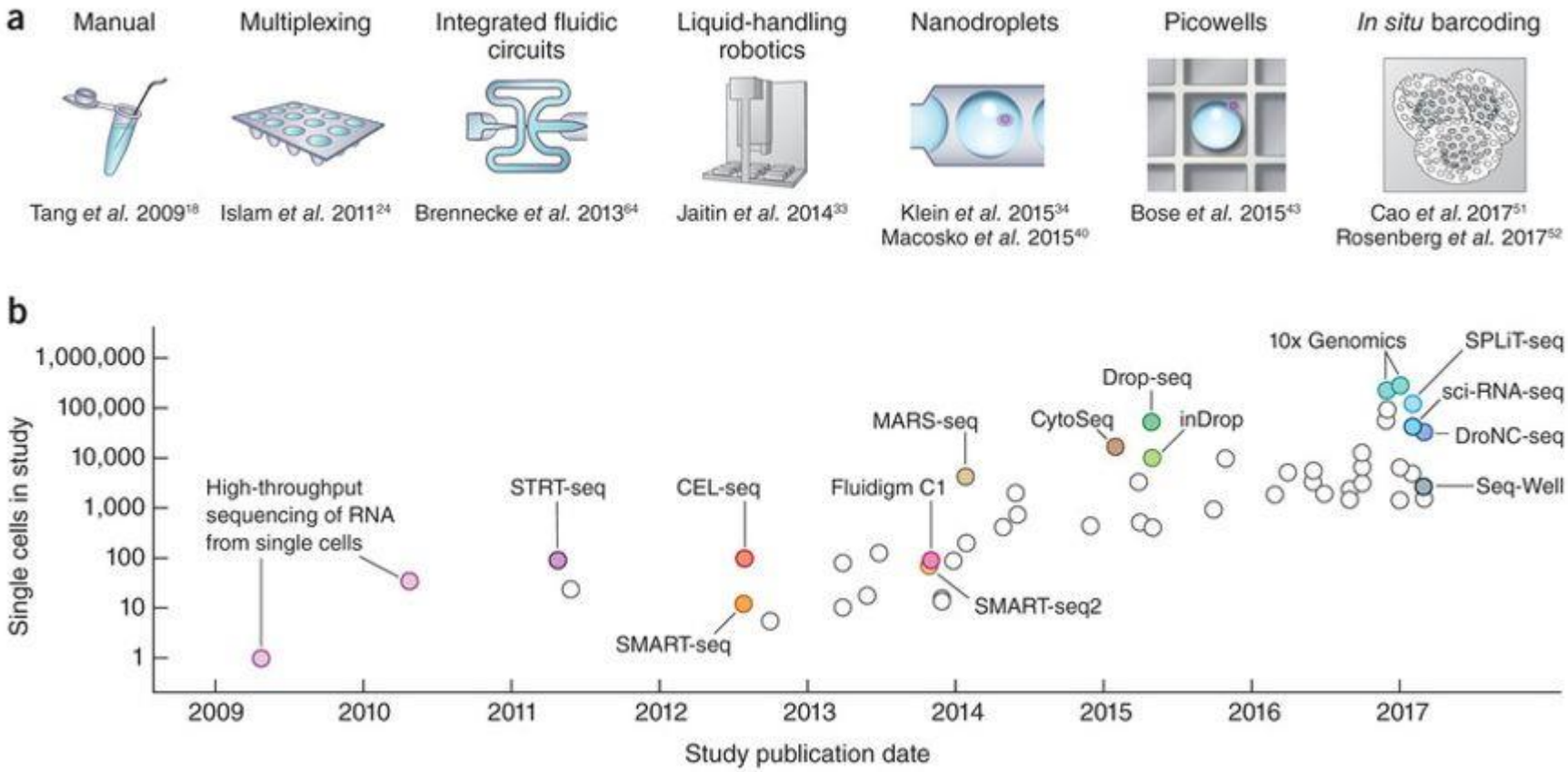


Levitin et al, *Molecular Systems Biology*, 2019

Challenges of scRNA-seq in Cancer

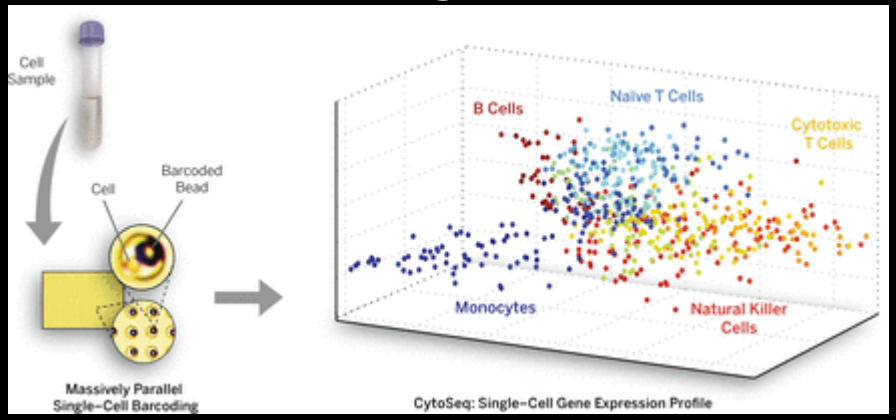
- Extreme compositional heterogeneity – multiple malignant lineages, asynchronous differentiation, immune microenvironment, stromal cells. Large cell numbers are required.
- Untransformed cell-of-origin also present – how to distinguish the malignant tumor cells from non-neoplastic cells in the microenvironment?
- Sample preparation – solid tumors must be dissociated without inducing massive expression changes and biasing cellular composition

Exponential Scaling of scRNA-seq over 10 Years

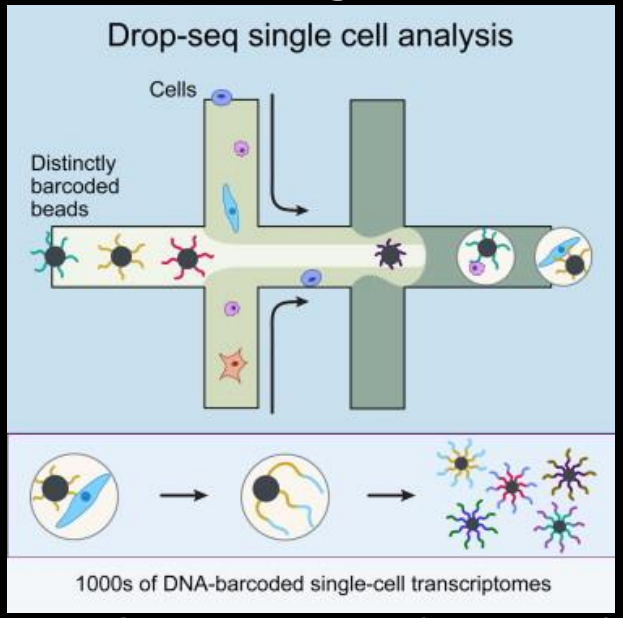


Svensson et al, *Nature Protocols*, 2018

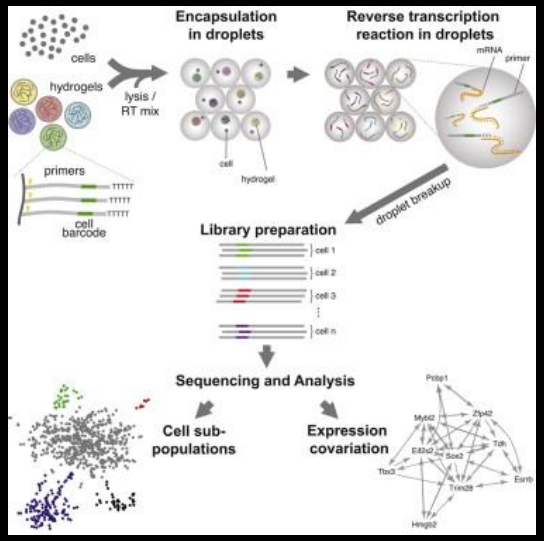
New Tools for Highly Multiplexed, Single-Cell Expression Profiling



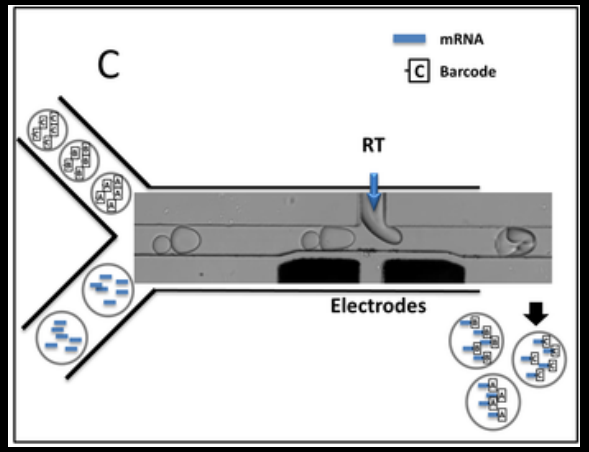
Targeted Expression Profiling in Open Microwells (CytoSeq)
Fan et al., *Science*, 2015.



RNA-Seq in Droplets (Drop-seq)
Macosko et al., *Cell*, 2015.



RNA-Seq in Droplets (inDrops)
Klein et al., *Cell*, 2015.

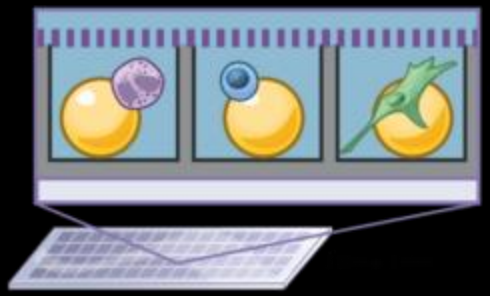


RNA-Seq in Droplets (Hi-SCL)
Rotem et al., *PLoS ONE*, 2015.

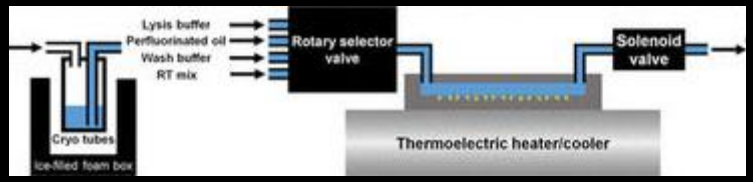
RNA-Seq in Sealable Microwells
Bose et al., *Genome Biology*, 2015.

Second-Generation Tools for Large-Scale scRNA-Seq

Microwells



Seq-Well
(Gierahn et al, *Nature Methods*, 2017)

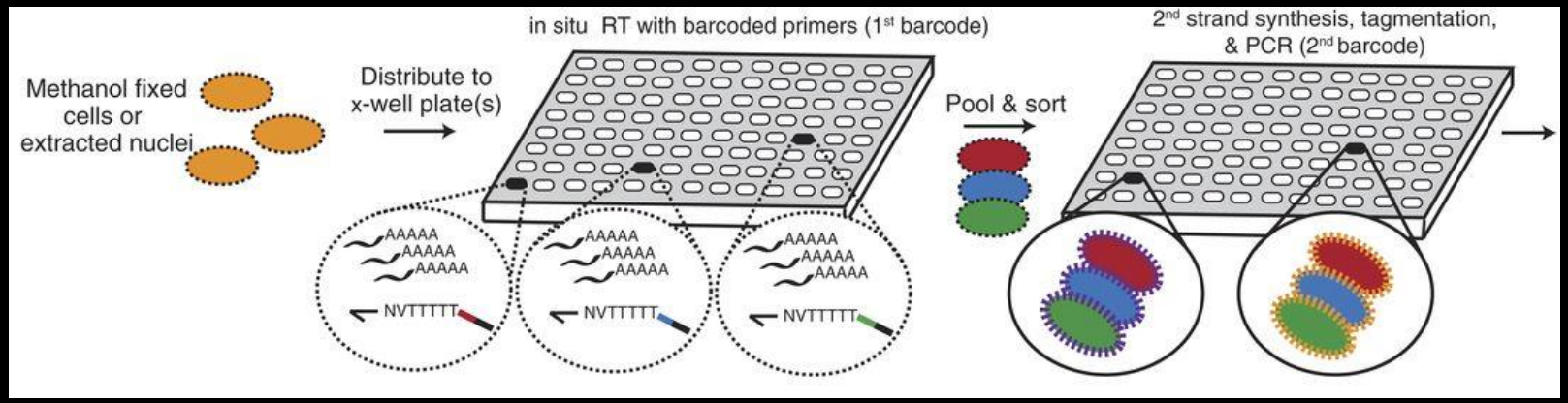


Automated Microwells
(Yuan and Sims, *Sci Rep*, 2016)

Droplets

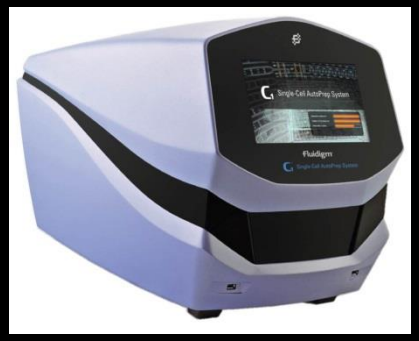
10x Genomics Chromium
(Zheng et al, *Nature Commun*, 2017)

Split-Pool Sorting

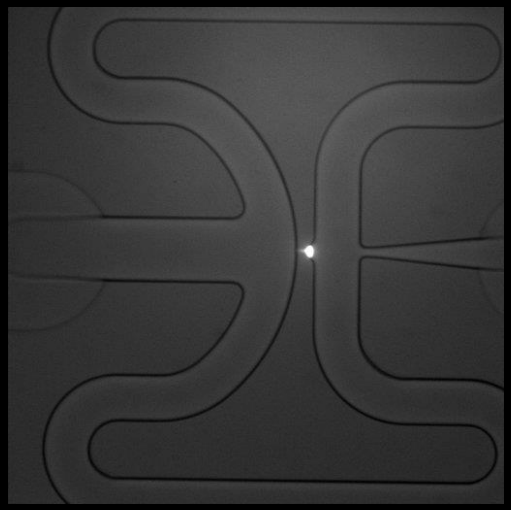


sci-RNA-Seq
(Cao et al, *Science*, 2017)

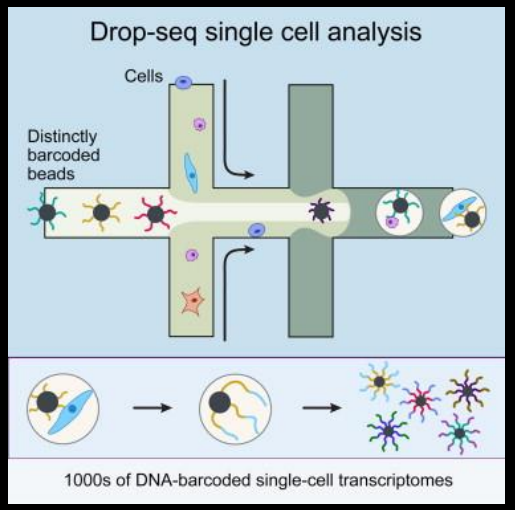
Old Way: Isolate Cells-of-Interest and Find Genes that Distinguish Them



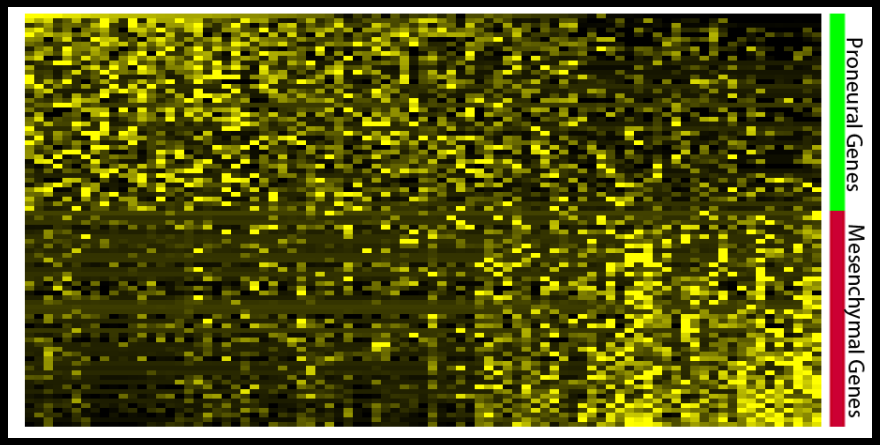
Fluidigm C1
96-cell Chip



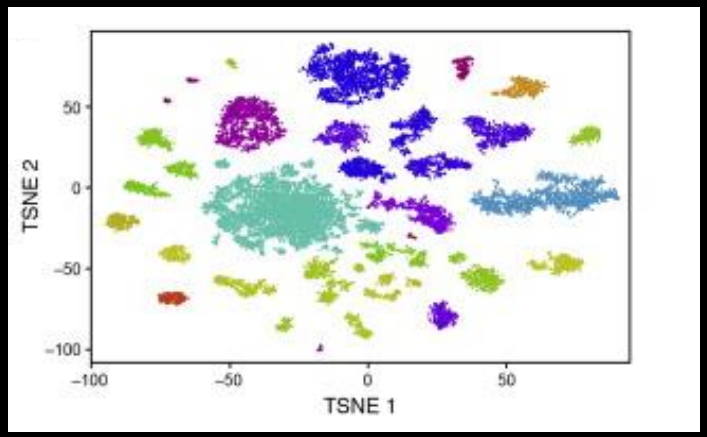
New Way: Get All of the Cells and Cluster



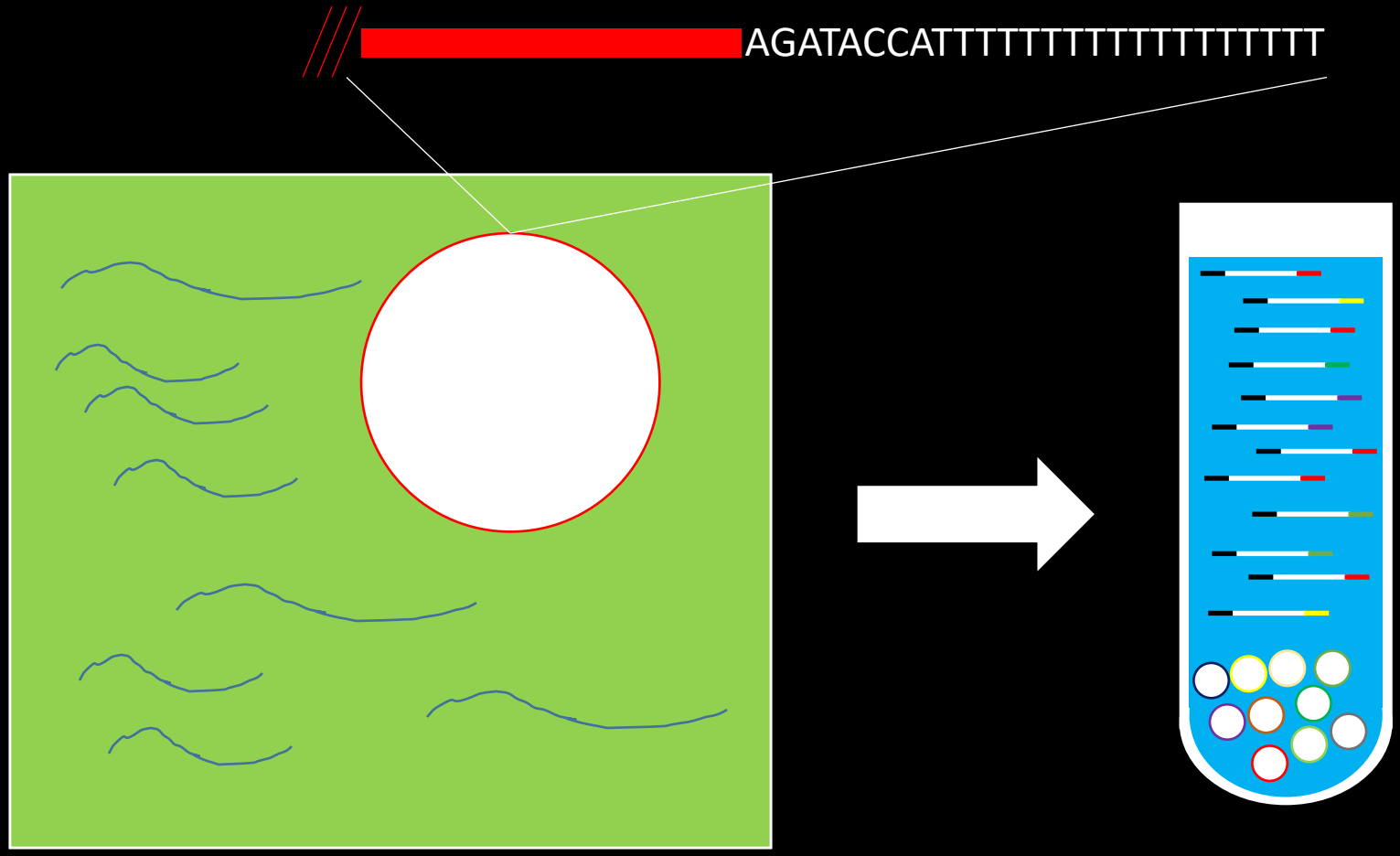
Transformed Cells from Proneural Tumor:



Cells from Dissociated Retina:

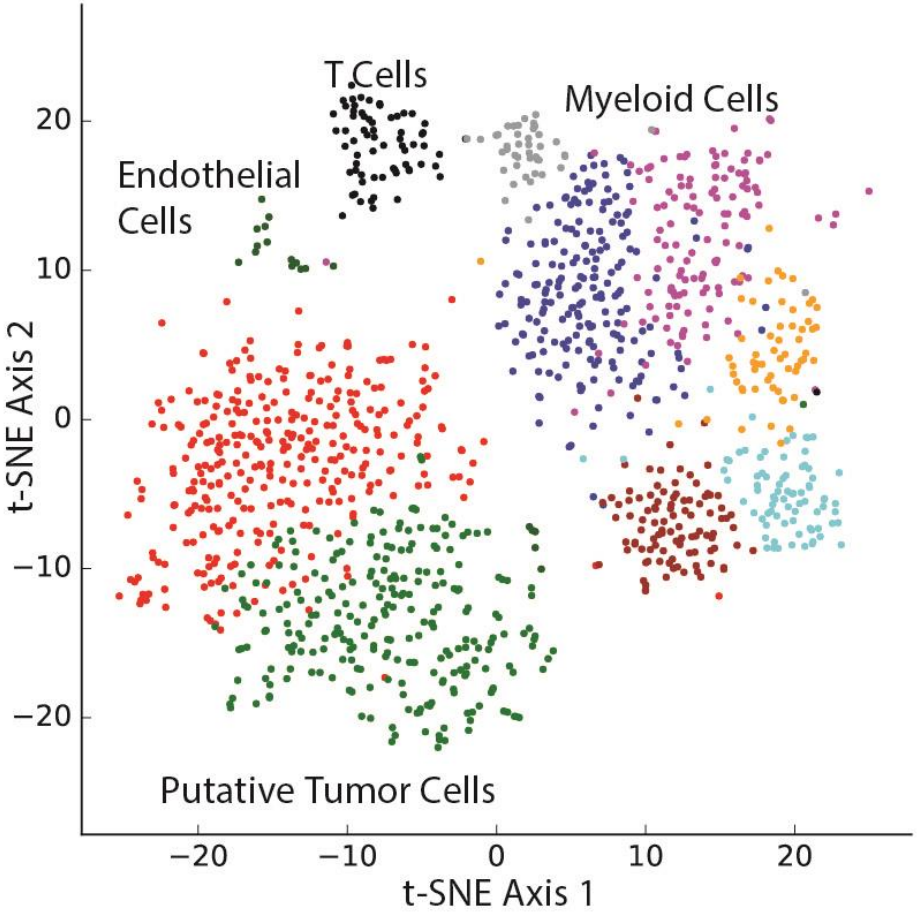


Microfluidic Pooled Barcode Approach to Single Cell RNA-Seq

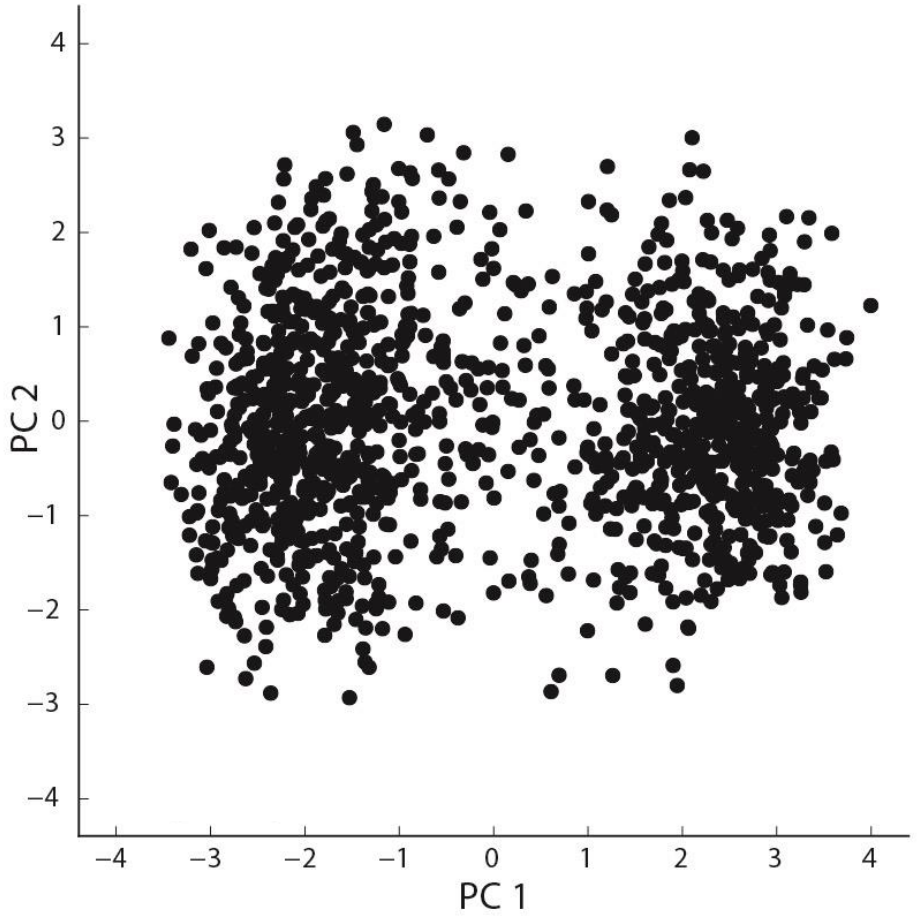


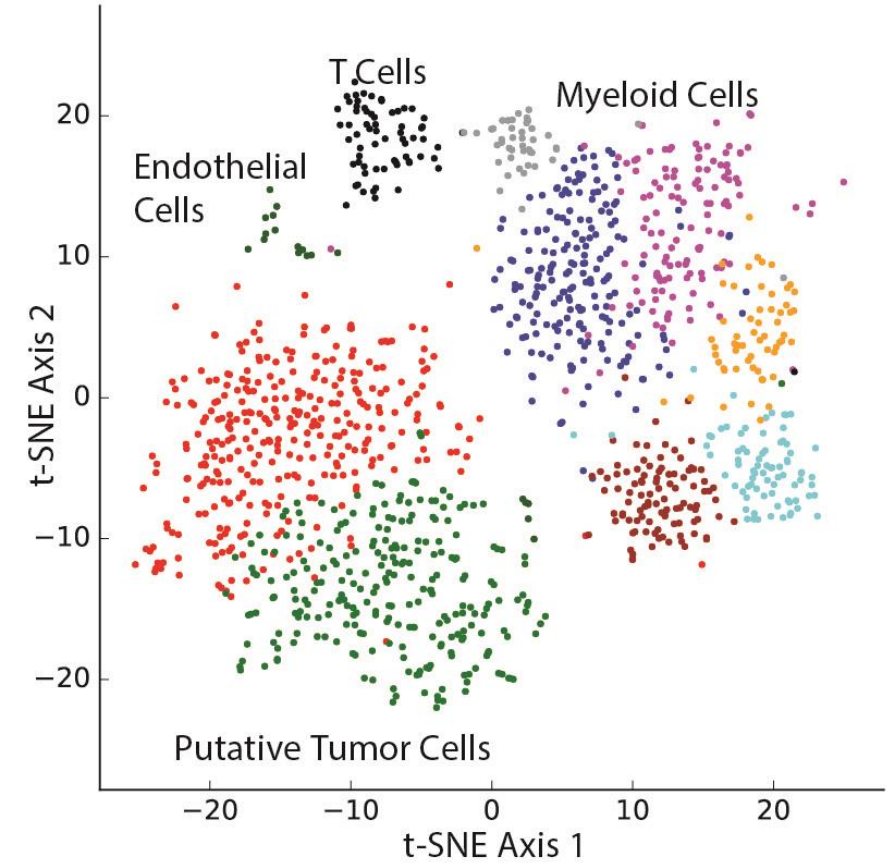
How do we identify the malignantly transformed tumor cells?

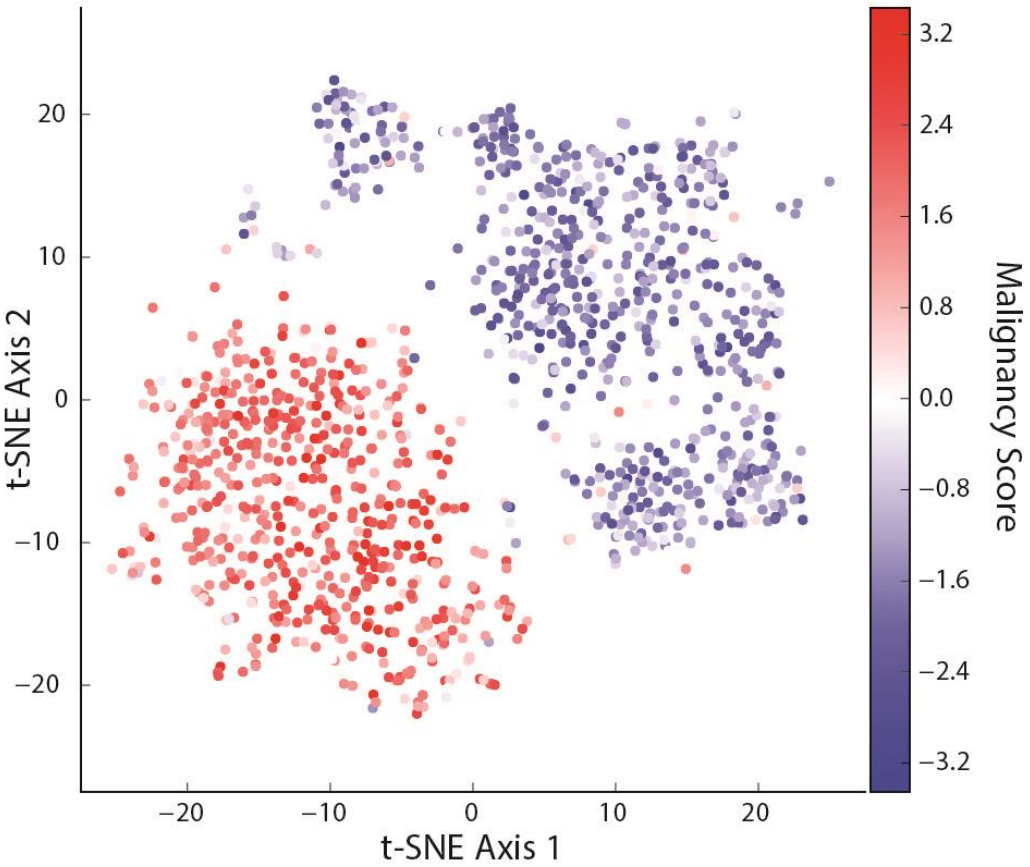
Cluster Cells based on Gene Expression

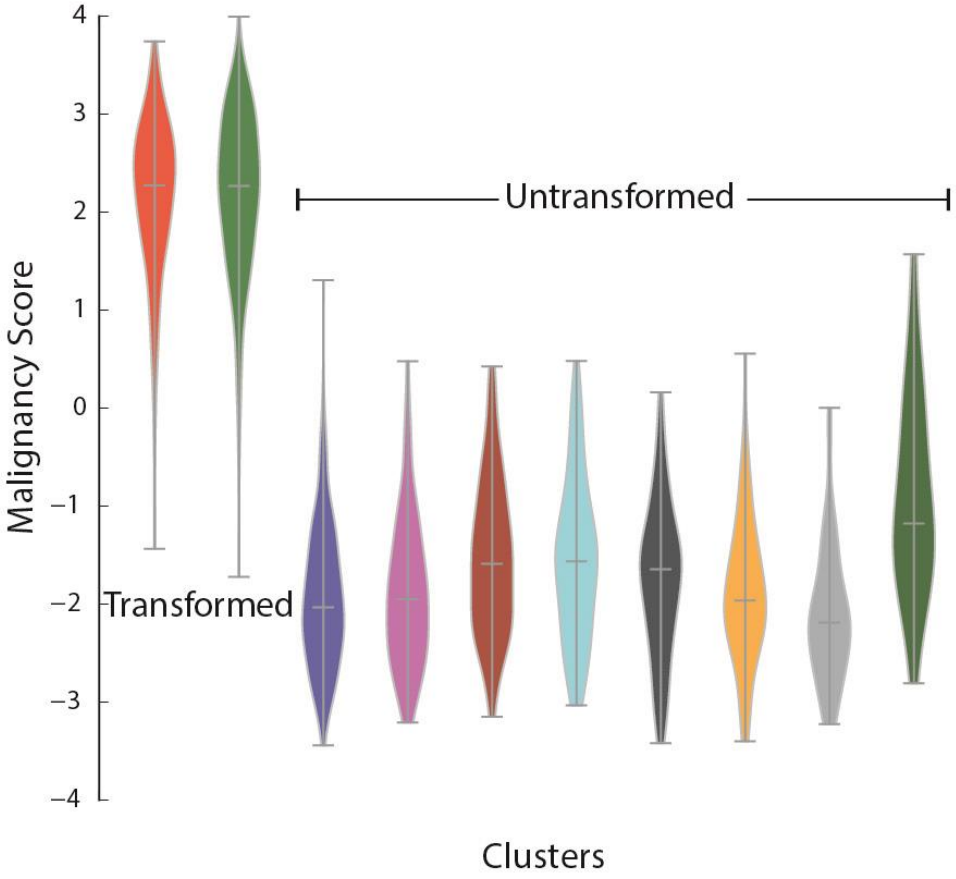


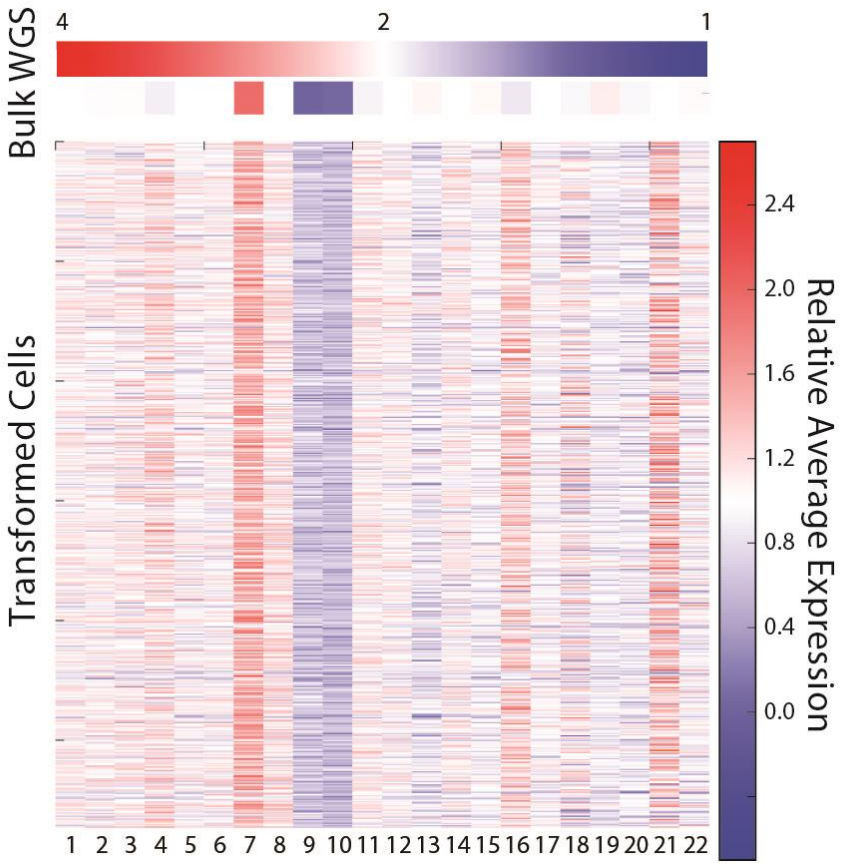
Cluster Cells based on Chromosomal Expression

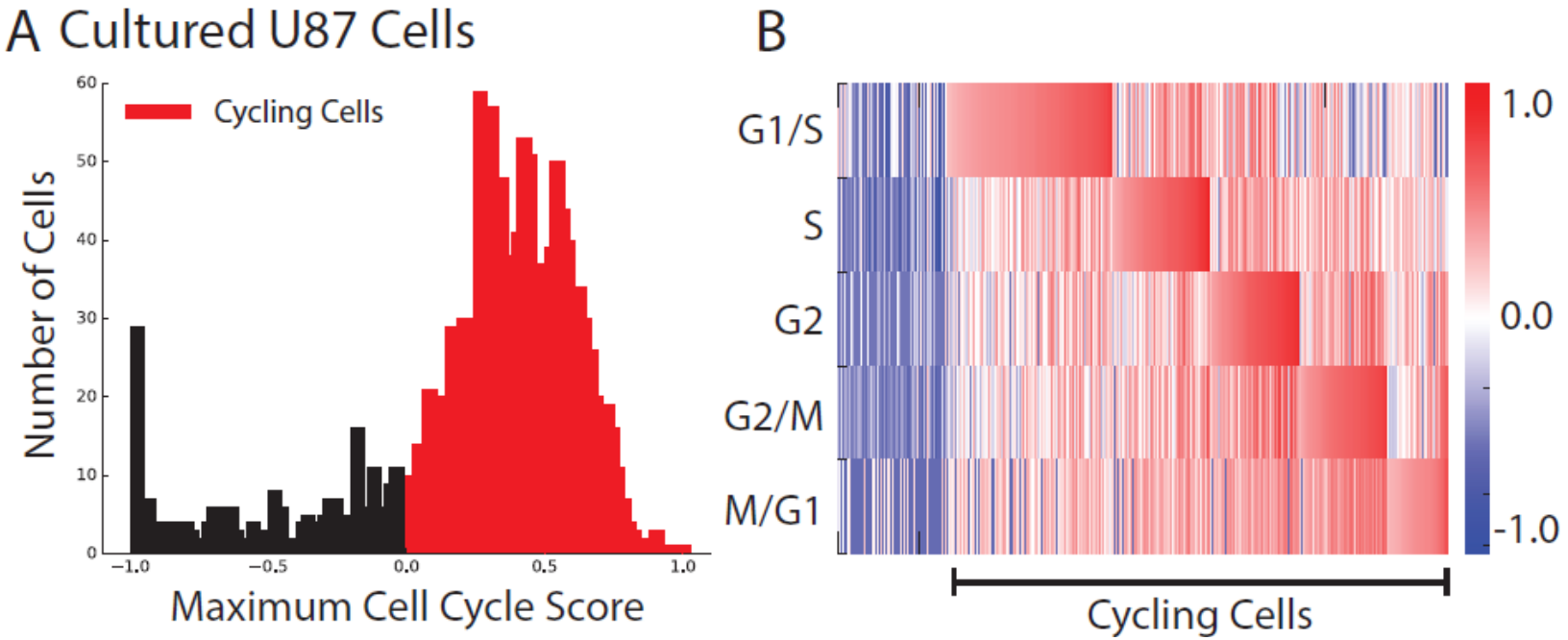



T Cells
Myeloid Cells
Endothelial Cells
Putative Tumor Cells
t-SNE Axis 1
t-SNE Axis 2
20
10
0
−10
−20
−20
−10
0
10
20

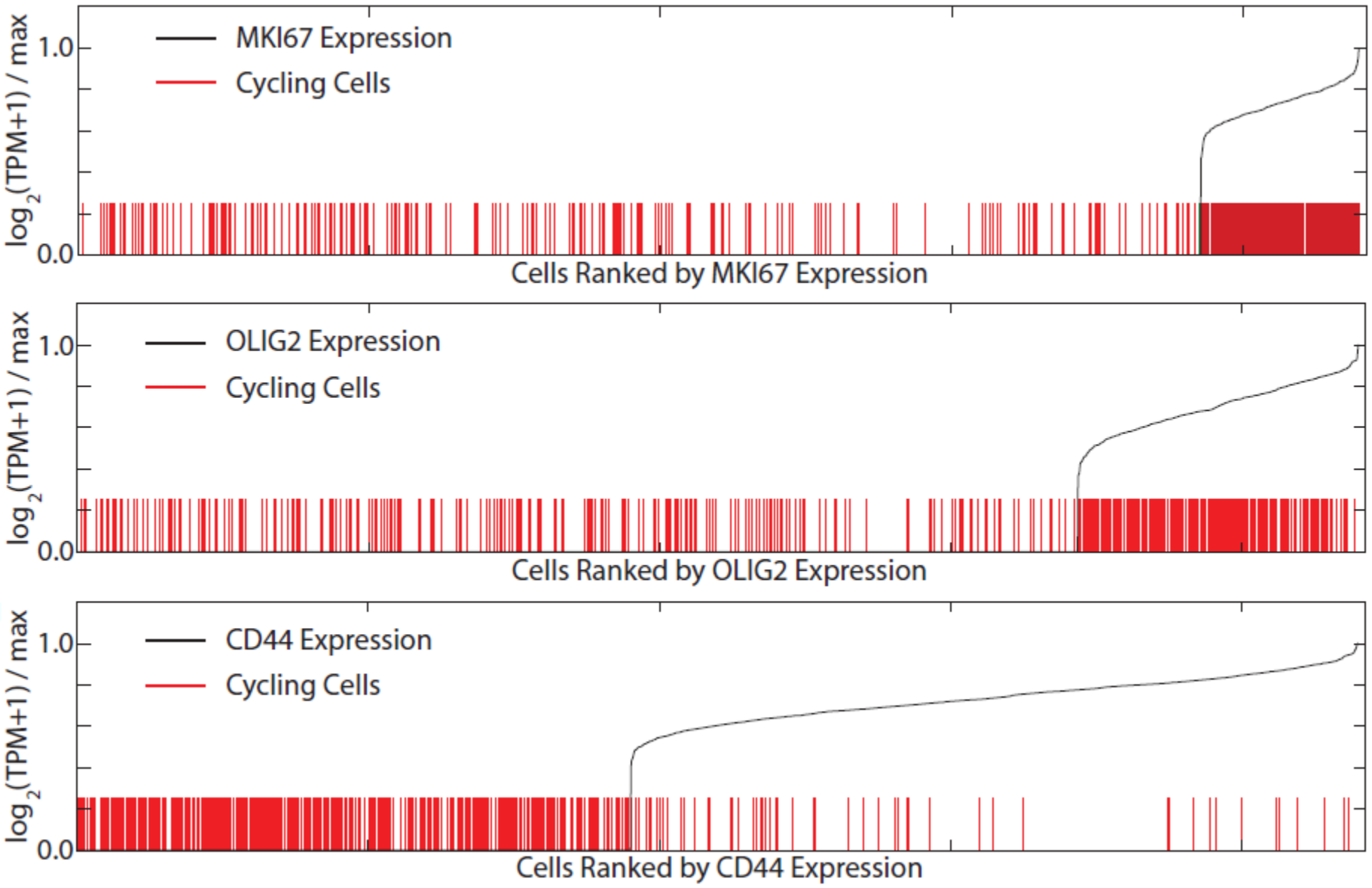

t-SNE Axis 1
t-SNE Axis 2
Malignancy Score
3.2
2.4
1.6
0.8
0.0
−0.8
−1.6
−2.4
−3.2
20
10
0
−10
−20
−20
−10
0
10
20


Malignancy Score
4
3
2
1
0
−1
−2
−3
−4
Transformed
Untransformed
Clusters


Bulk WGS
4
2
1
Transformed Cells
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Relative Average Expression
2.4
2.0
1.6
1.2
0.8
0.4
0.0


A Cultured U87 Cells
Cycling Cells
Number of Cells
60
50
40
30
20
10
0
−1.0
−0.5
0.0
0.5
1.0
Maximum Cell Cycle Score
B
G1/S
S
G2
G2/M
M/G1
1.0
0.0
-1.0
Cycling Cells

Simultaneous Analysis of “Cell State” vs. “Cell Type”



Challenges of scDNA-seq in Cancer

- Extreme compositional heterogeneity – rare clones can be just as important as highly prevalent ones. In many tumors, transformed cells aren't even in the majority and there is rarely a high-fidelity marker of the transformed cells.

- Simultaneously achieving highly uniform coverage and highly accurate sequence data remains challenging.

- Advances in pooled-barcode library construction do not have the same transformative impact on scDNA-seq as on scRNA-seq. At the end of the day, the genome is still large and so sequencing costs are limiting.

Multiple Displacement Amplification

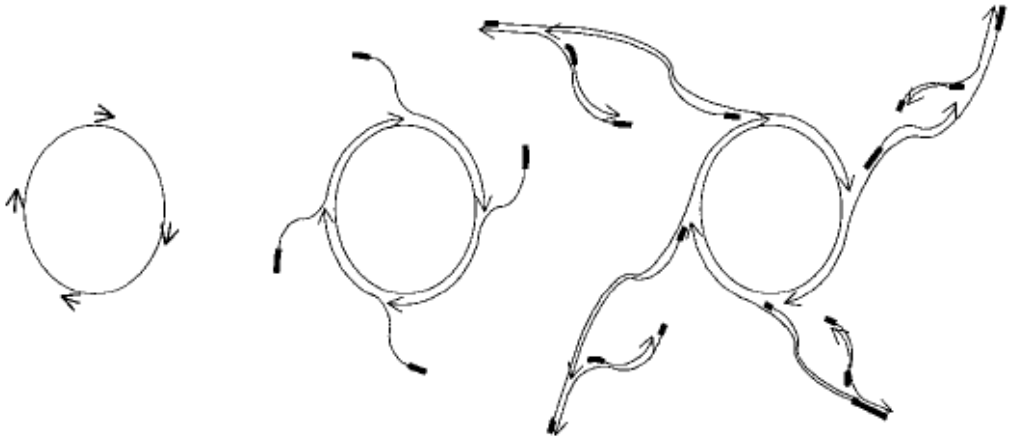
ADVANTAGES:

- Rapid, efficient isothermal amplification from single cells with no special equipment.
- Low error rate due to intrinsic proofreading activity of polymerase – good for calling SNVs
- Relatively high coverage (breadth)

Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification

Frank B. Dean,[1,3] John R. Nelson,[2,3] Theresa L. Giesler,[2] and Roger S. Lasken[1,4]

[1]Molecular Staging, Inc., New Haven, Connecticut 06511, USA; [2]Amersham Pharmacia Biotech, Piscataway, New Jersey 08855-1327, USA



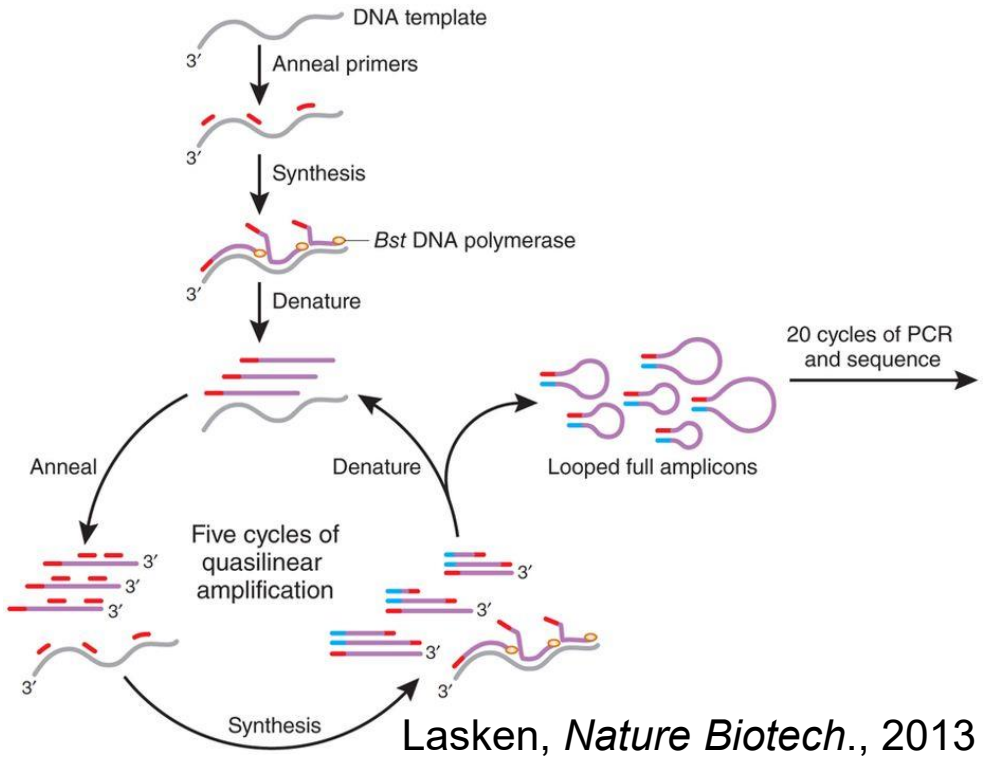
Disadvantages:

- Relatively low coverage uniformity compared to other methods, particularly for human-sized genomes.
- Hyperbranching process results in spurious chimera formation

Multiple Annealing and Looping-Based Amplification Cycles (MALBAC)

ADVANTAGES:

- Rapid and efficient amplification from single cells.
- Relatively high coverage (breadth), even for large genomes.
- Relatively uniform coverage (depth), even for large genomes.
- Looping step reduces chimera formation rate.



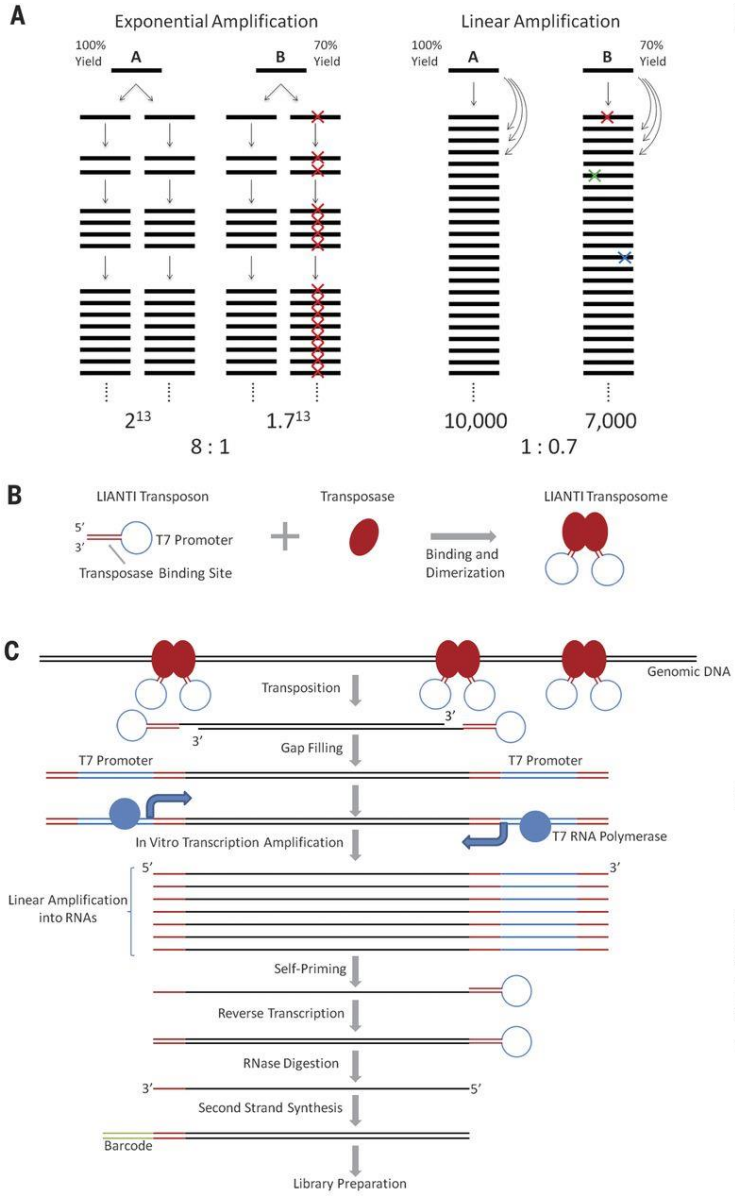
Lasken, *Nature Biotech*., 2013

Disadvantages:

- Uses *Bst* DNA polymerase to run primer extension reactions at higher temperature – enzyme has no proofreading, complicating accurate SNV calling.
- More complex reaction than MDA, requires thermocycling, multiple enzyme additions.

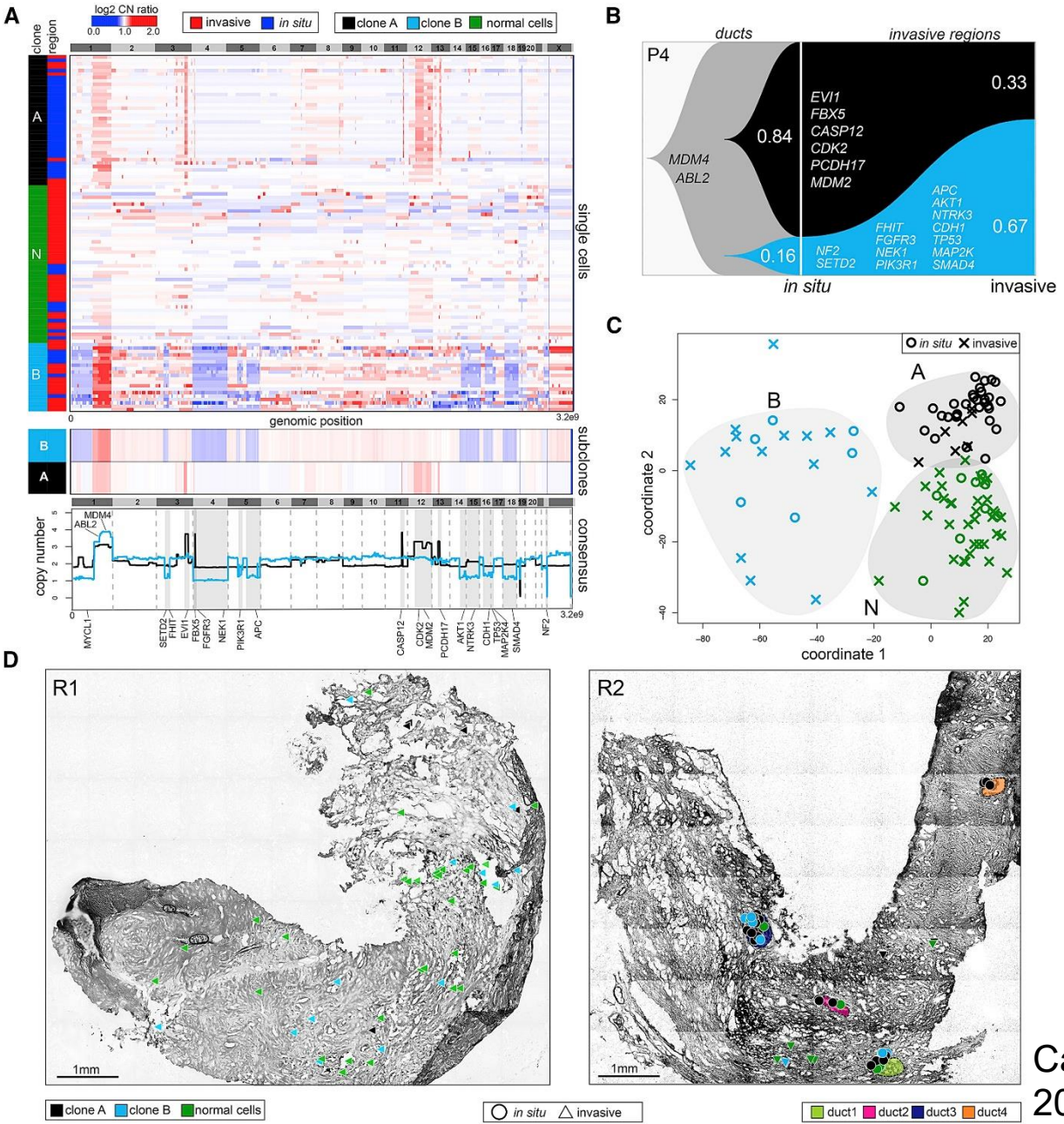
Zong et al, *Science*, 2012

Linear Amplification via Transposon Insertion (LIANTI)



Chen et al, *Science*, 2017

scDNA-seq Enables Simultaneous Spatial and Phylogenetic Profiling in Breast Cancer

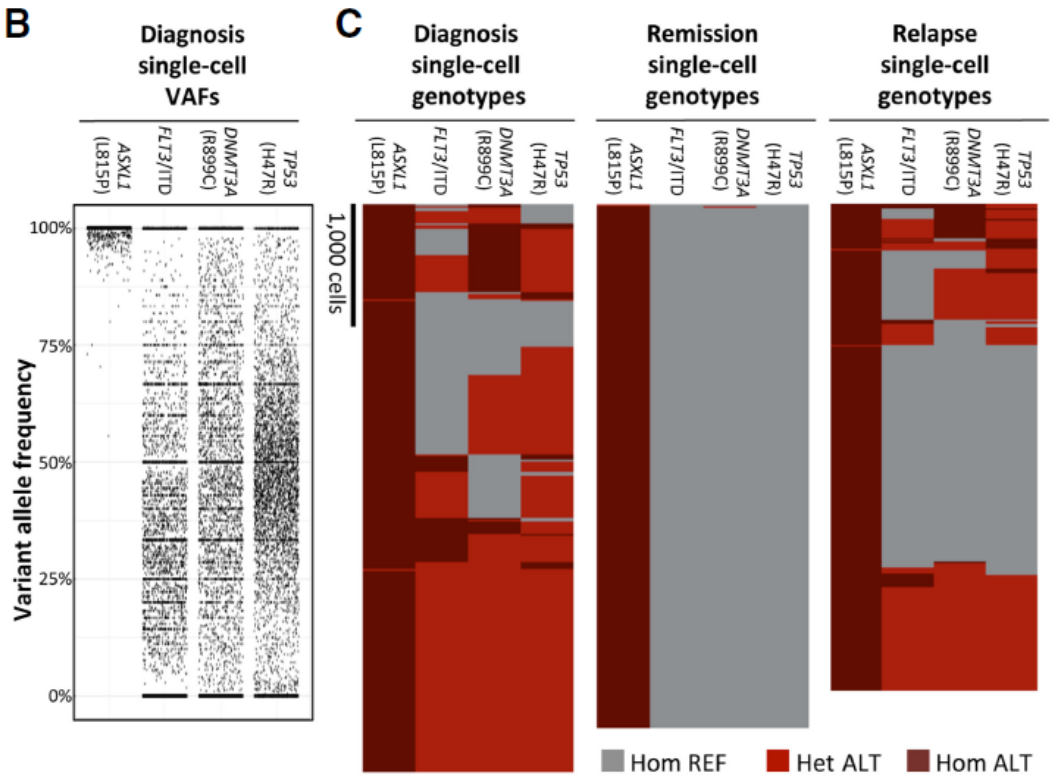


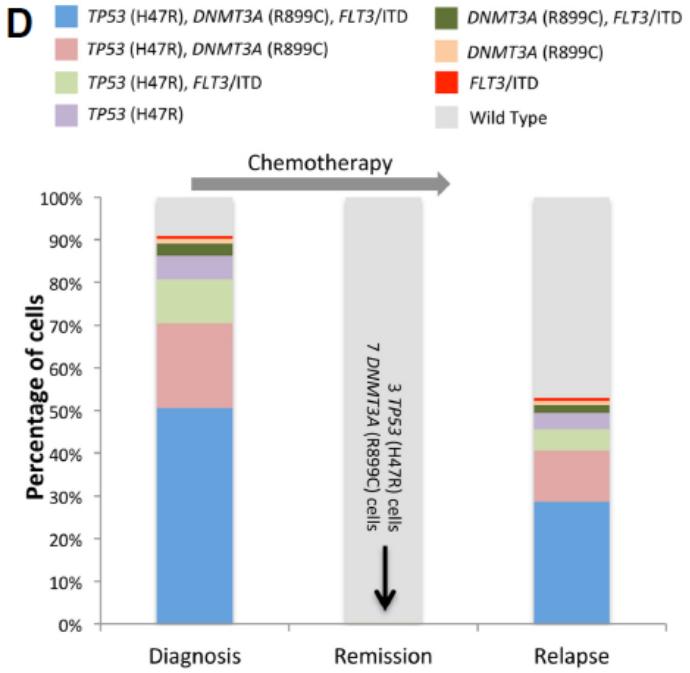
Casasent et al, *Cell*, 2018

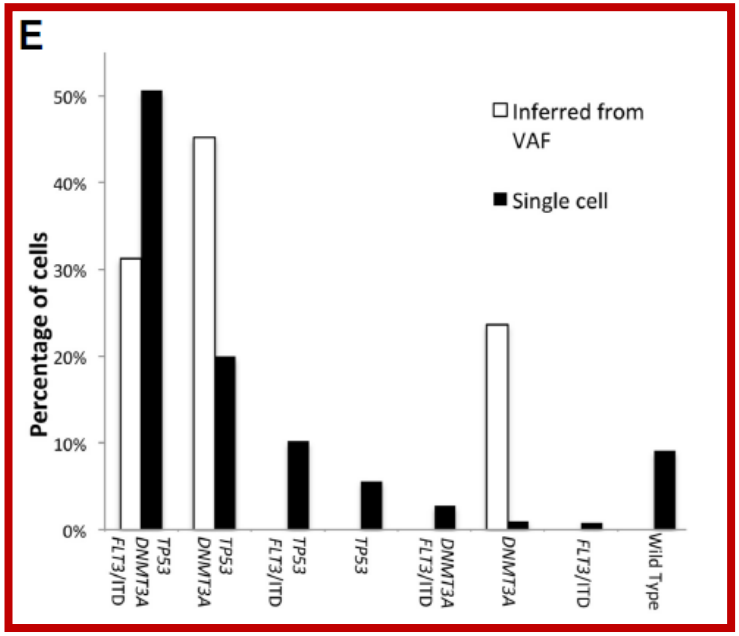
Single-cell Genomics of AML under Therapy

A

	Diagnosis	Remission	Relapse
Total read pairs	17,331,034	18,824,864	16,900,170
Reads mapping to cells	63%	85%	76%
Total cells found	7,364	5,605	5,498
Average reads per cell	1,578	2,974	2,188
Number of genotyped cells	4,748	4,384	4,236
Raji spike in detection rate	1.0%	1.3%	4.8%
Average allele dropout rate	2.1%	10.3%	8.7%







Pellegrino et al, *Genome Research,* 2018.