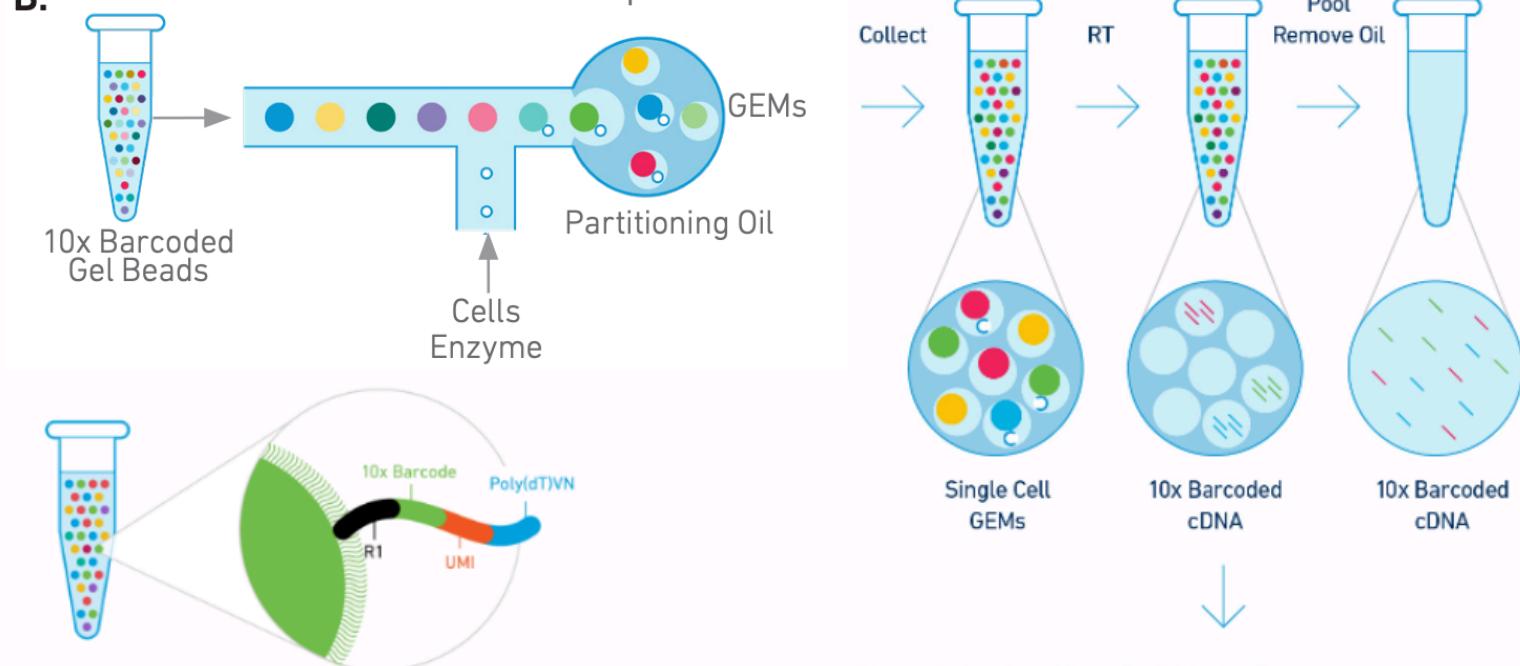
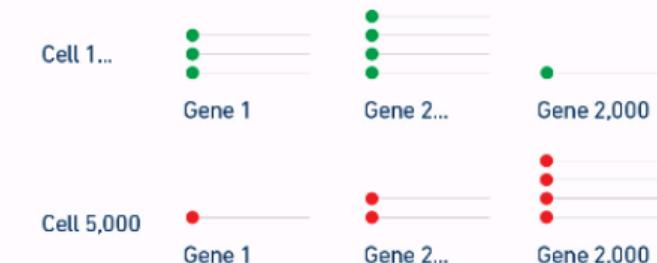


A.**B.** Chromium Next GEM Chip G

Transcriptional profiling of individual cells

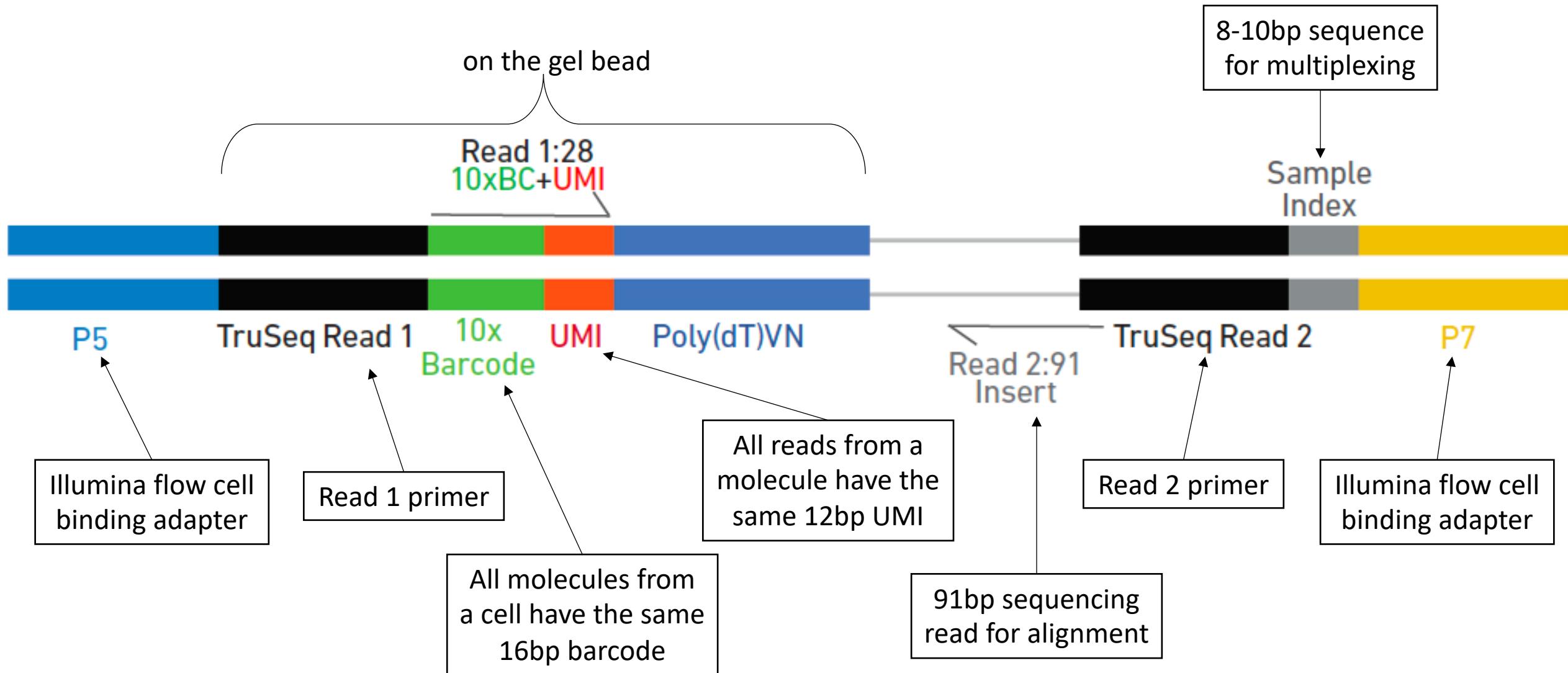


- Input: Single cells in suspension + 10x Gel Beads and Reagents
- Output: Digital gene expression profiles from every partitioned cell

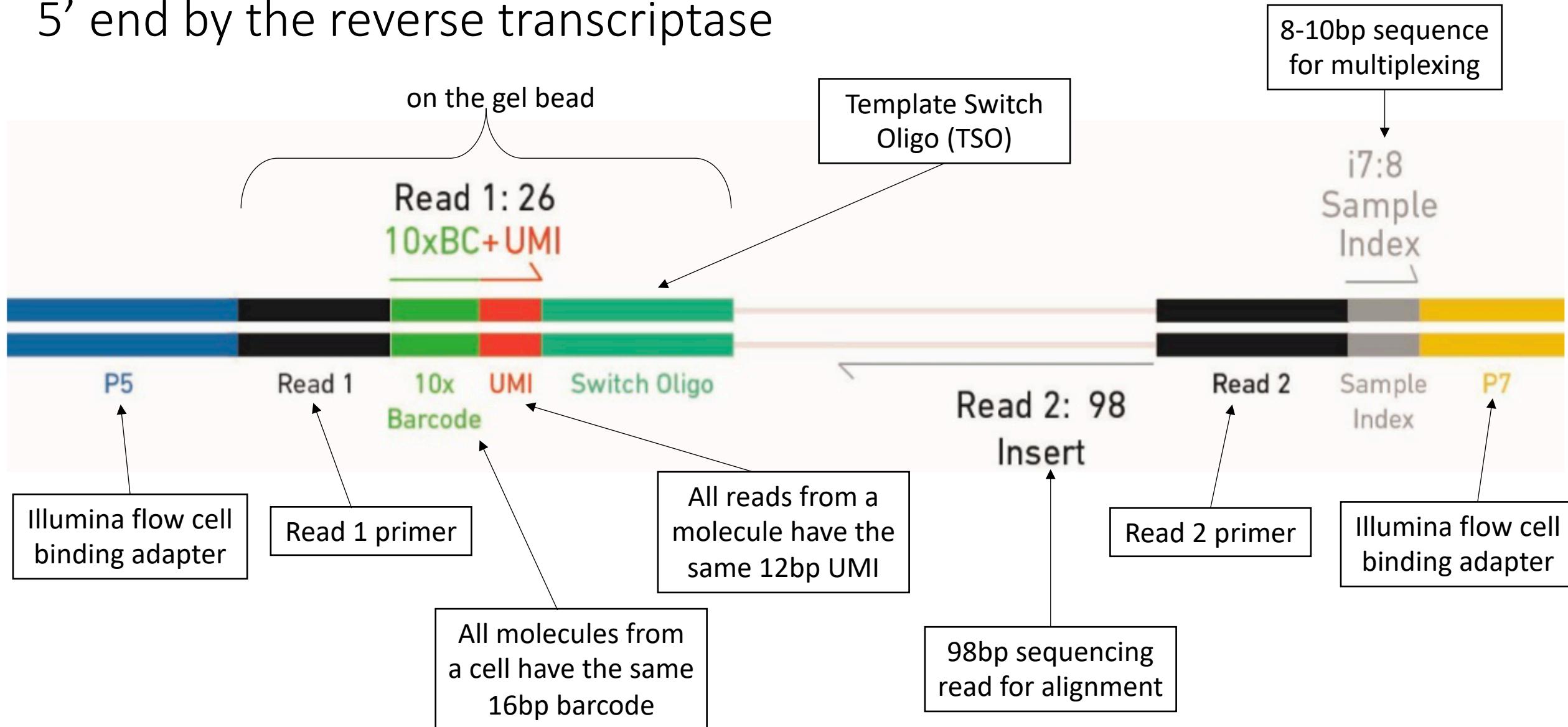
<https://www.10xgenomics.com/instrument/>

<https://ucdavis-bioinformatics-training.github.io/2017-June-RNA-Seq-Workshop/friday/scRNaseq.pdf>

3' GEX captures the 3' end of poly-A transcripts using a poly-T sequence on the gel bead

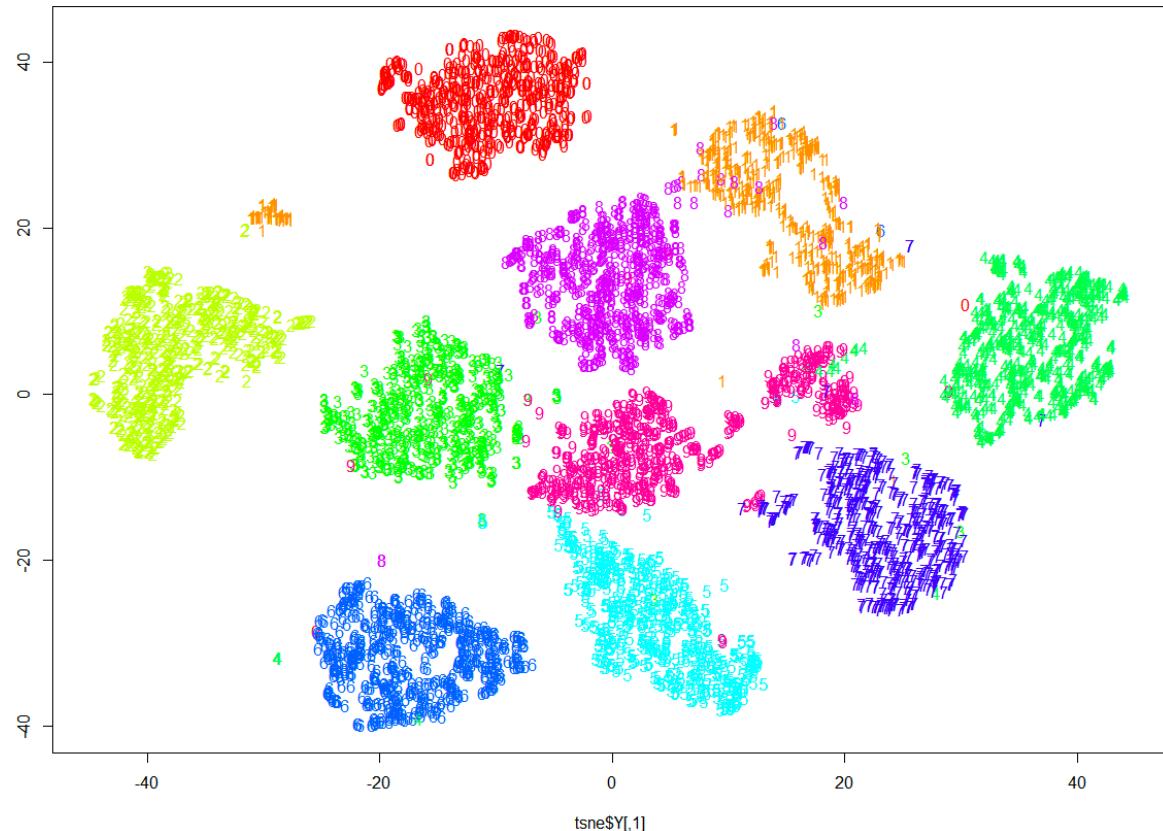


5' GEX captures the 5' end of poly-A transcripts using a Template Switch Oligo that hybridizes to a poly-C added at the 5' end by the reverse transcriptase



t-stochastic neighbor embedding (tSNE) is a useful but dangerous dimensionality reduction algorithm

- Numerical values on axes have no interpretable meaning
- Coloring may be from a clustering algorithm or from external metadata, pay attention!
- Cluster density is misleading
- Separation of clusters is misleading



<https://distill.pub/2016/misread-tsne/>

Check out this article + interactive visualizations!

ARTICLE

<https://doi.org/10.1038/s41467-019-11591-1>

OPEN

A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing

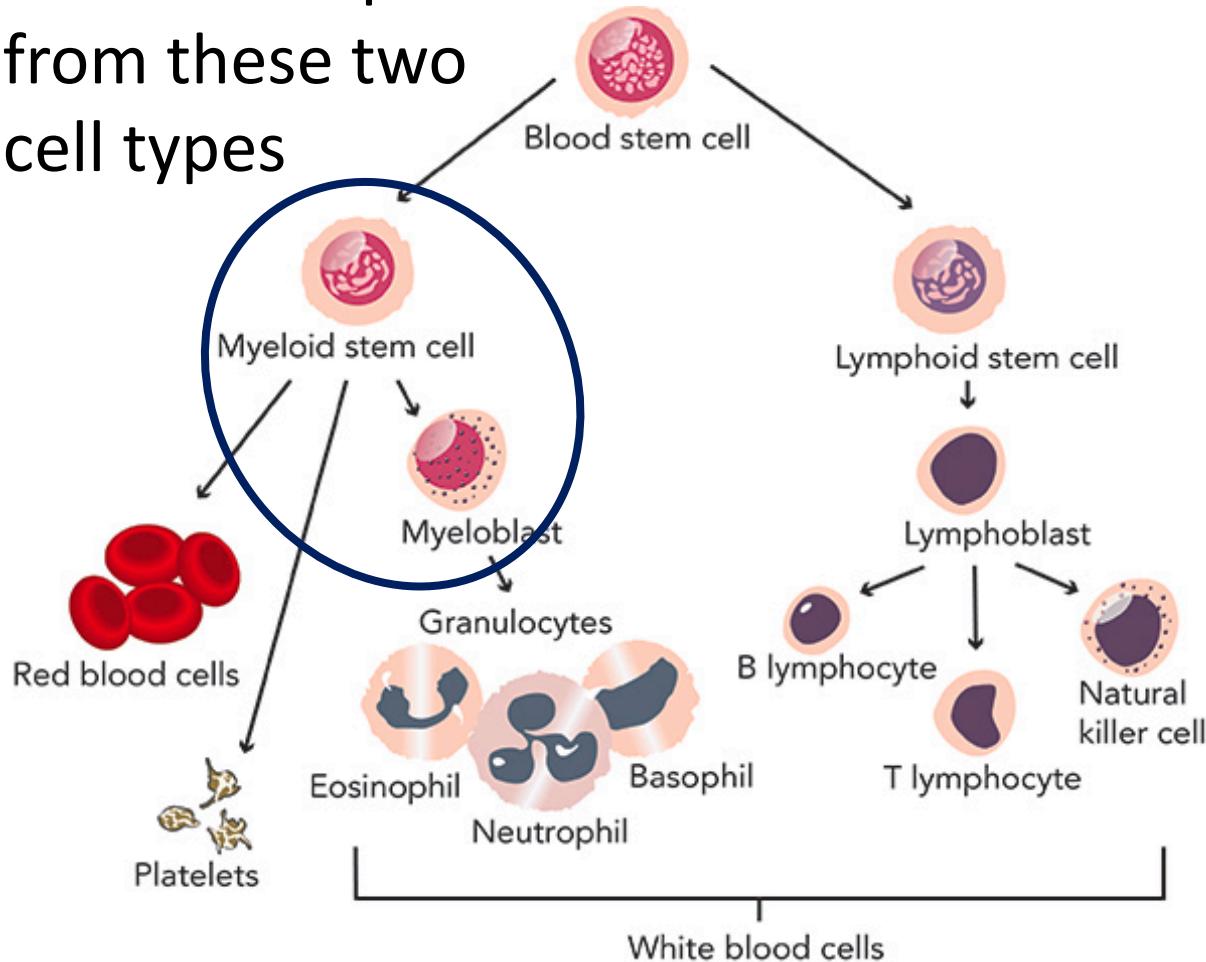
Allegra A. Petti^{1,2,7}, Stephen R. Williams^{3,7}, Christopher A. Miller^{1,2}, Ian T. Fiddes³, Sridhar N. Srivatsan¹, David Y. Chen⁴, Catrina C. Fronick², Robert S. Fulton², Deanna M. Church⁵ & Timothy J. Ley^{1,2,6}

Genomics Journal Club

09/18/2019

Charlotte Darby

AML develops from these two cell types



The American Cancer Society's estimates for leukemia in the United States for 2019 are:

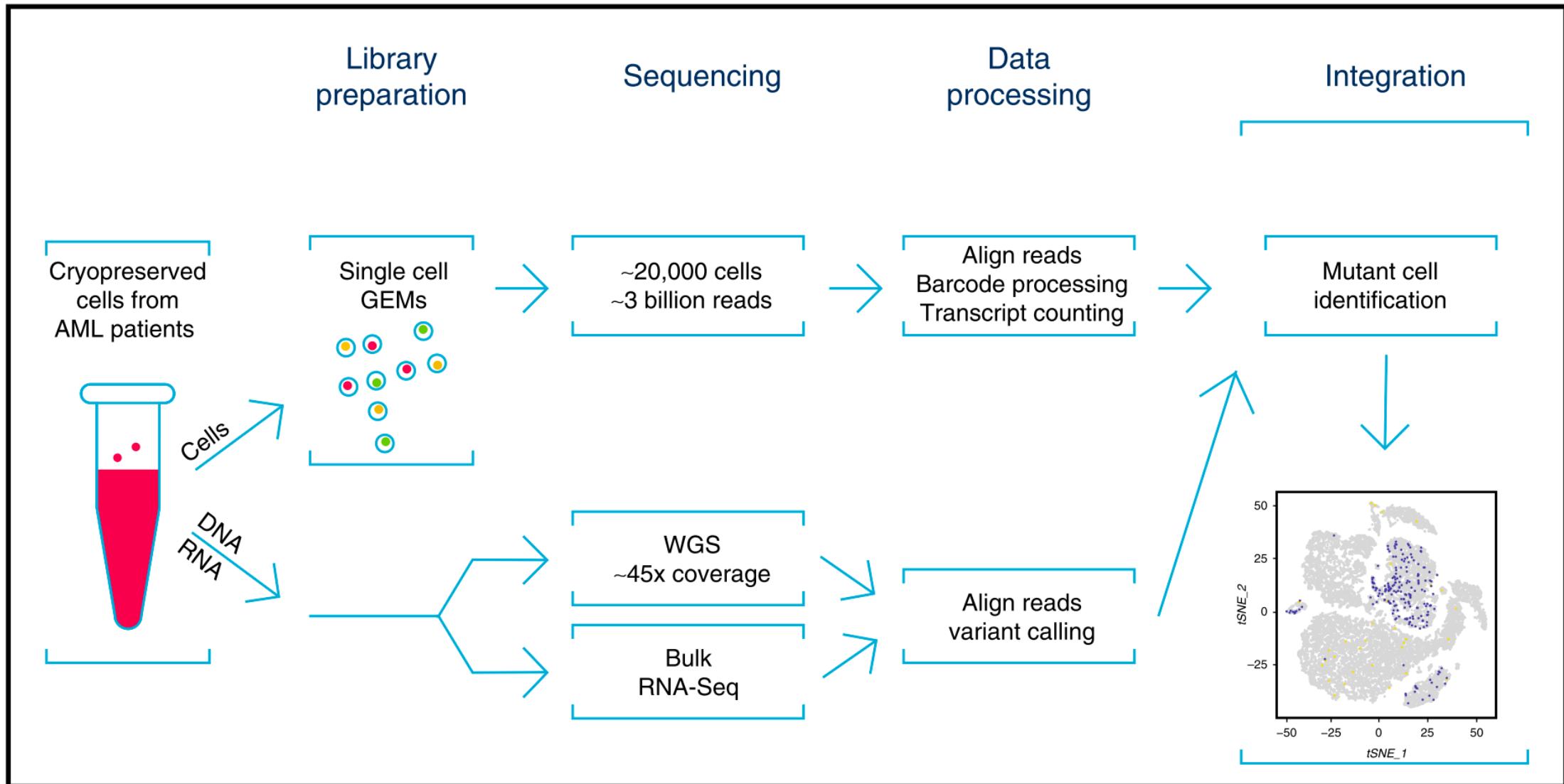
- About 61,780 new cases of leukemia (all kinds) and 22,840 deaths from leukemia (all kinds)
- About 21,450 new cases of acute myeloid leukemia (AML). Most will be in adults.
- About 10,920 deaths from AML. Almost all will be in adults.

AML is one of the most common types of leukemia in adults. Still, AML is fairly rare overall, accounting for only about 1% of all cancers.

AML is generally a disease of older people and is uncommon before the age of 45. The average age of people when they are first diagnosed with AML is about 68. But AML can occur in children as well.

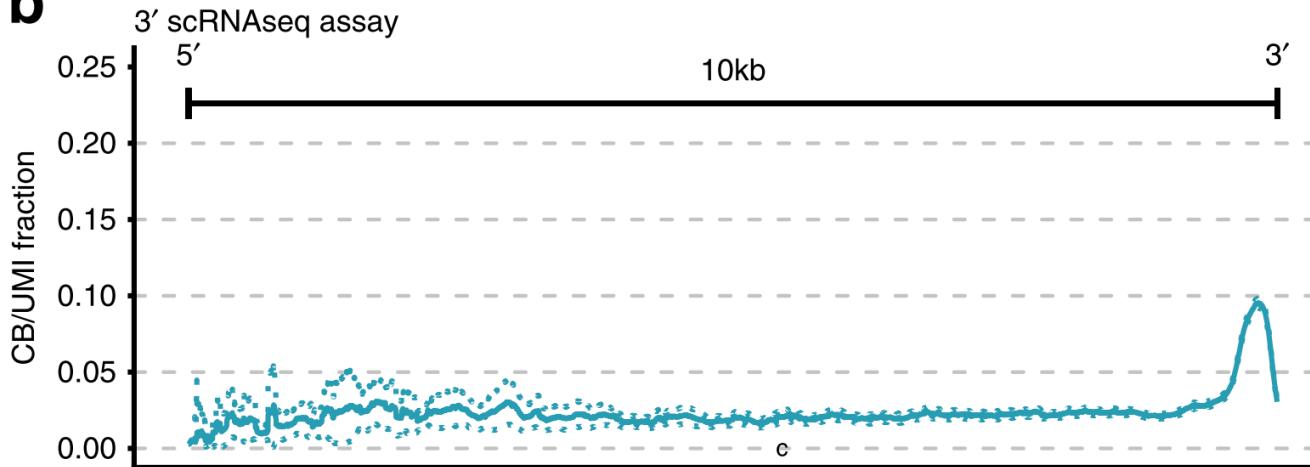
AML is slightly more common among men than women, but the average lifetime risk of getting AML in both sexes is about $\frac{1}{2}$ of 1%.

Experimental Workflow (Figure 1A)

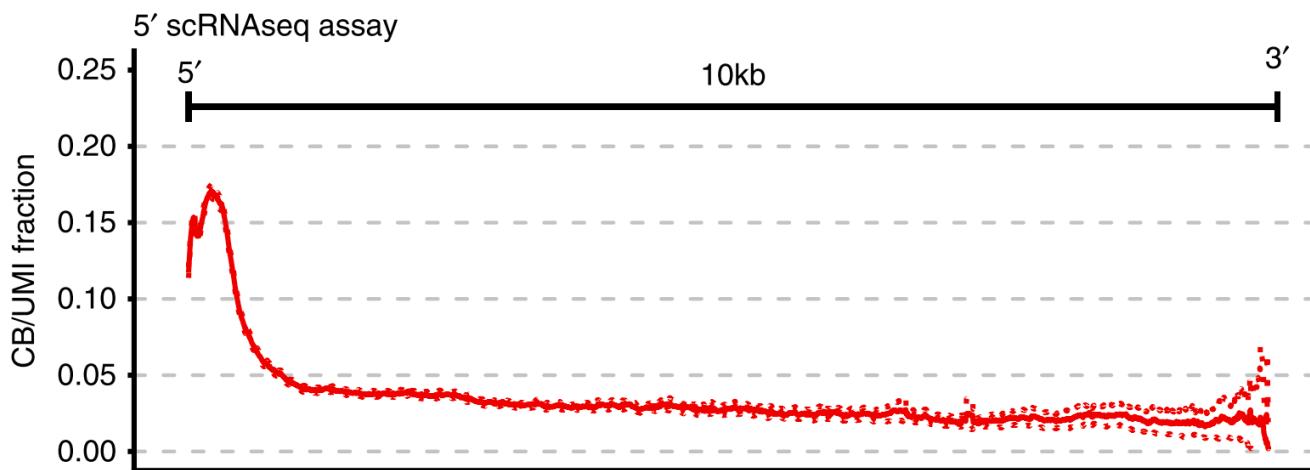


scRNA-seq coverage along a transcript (Figure 1B)

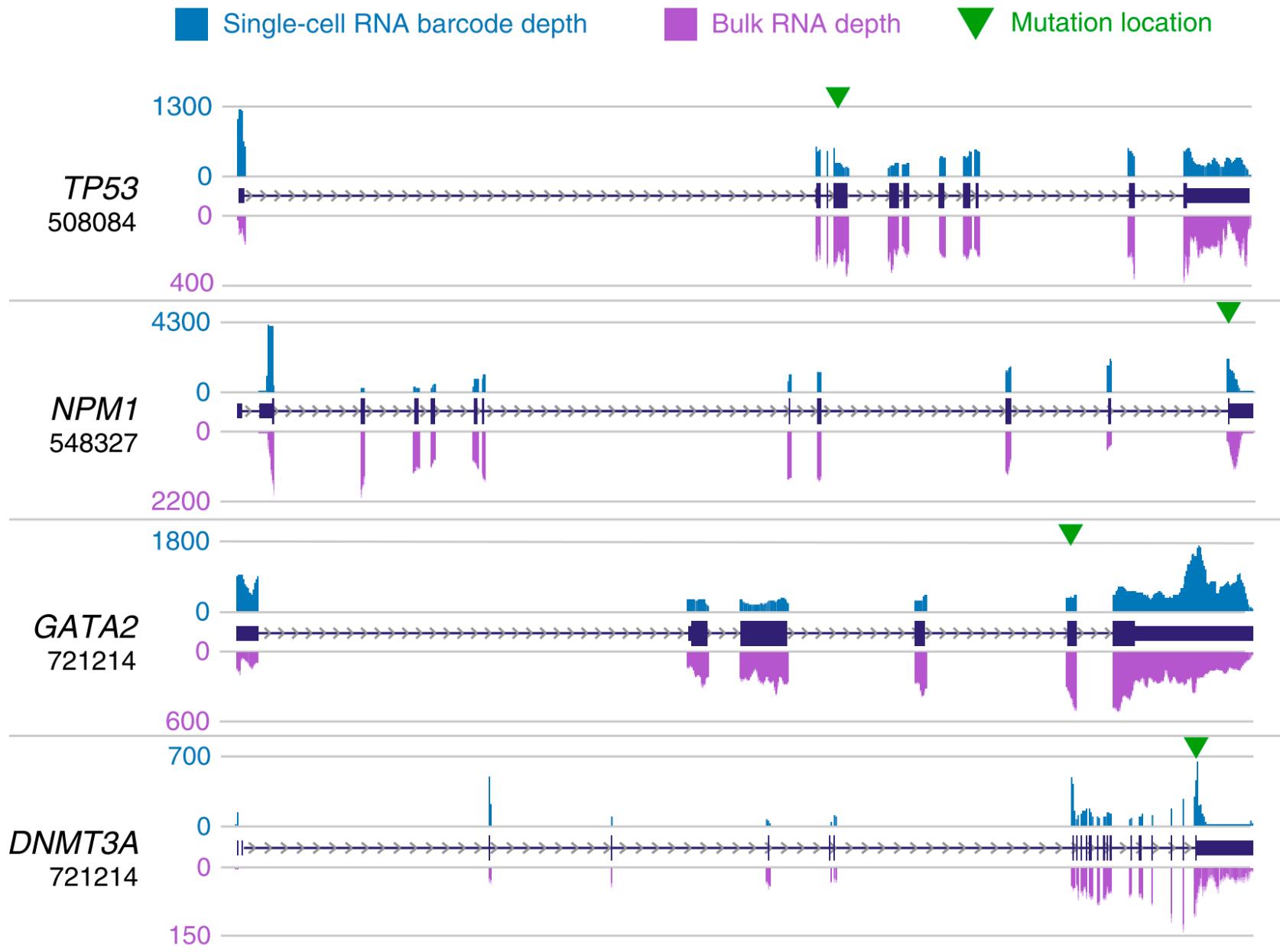
b



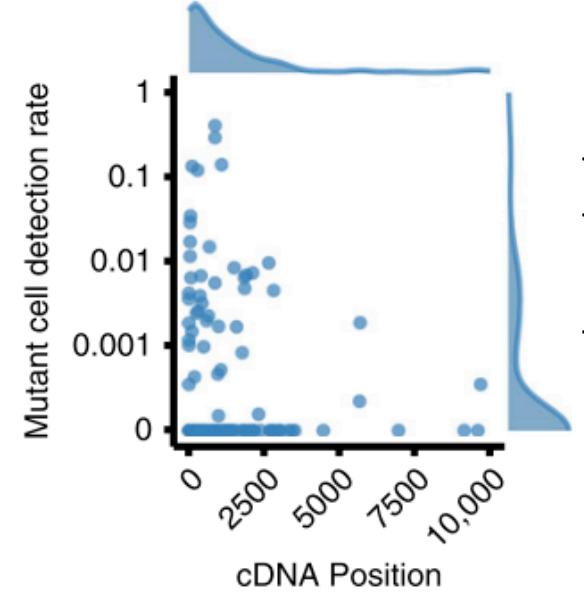
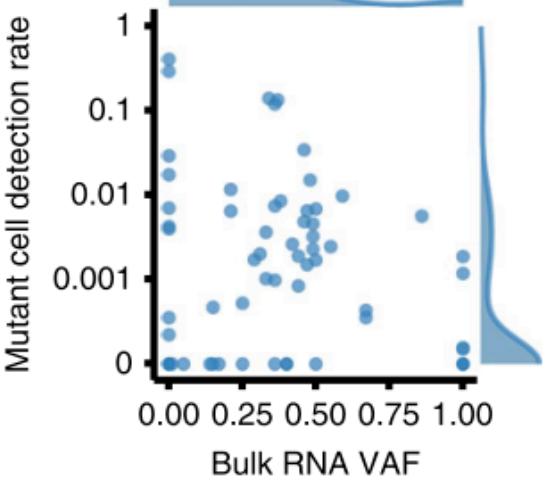
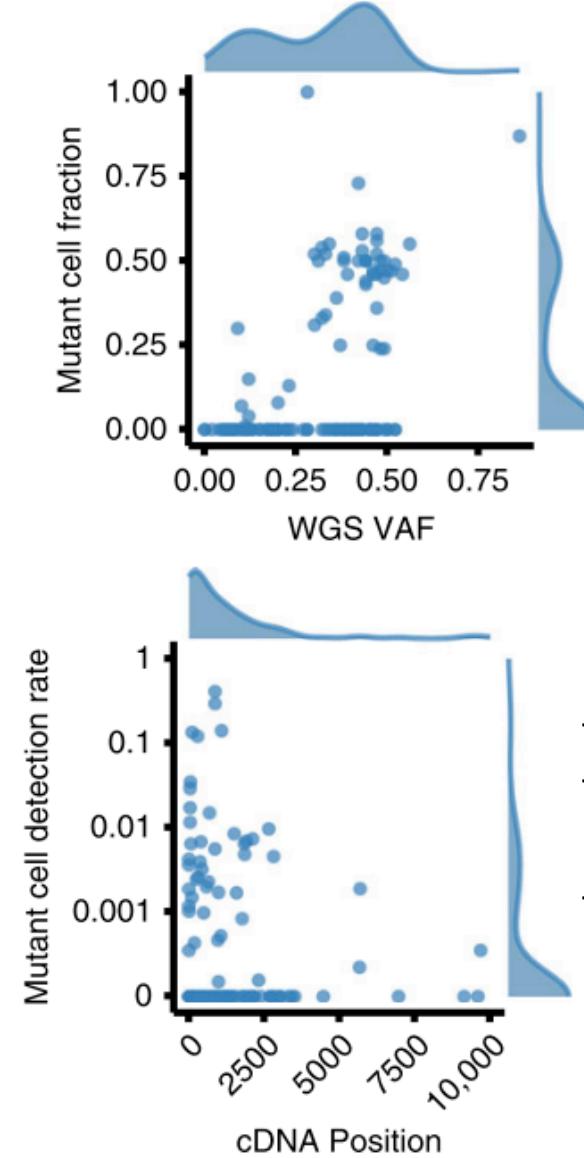
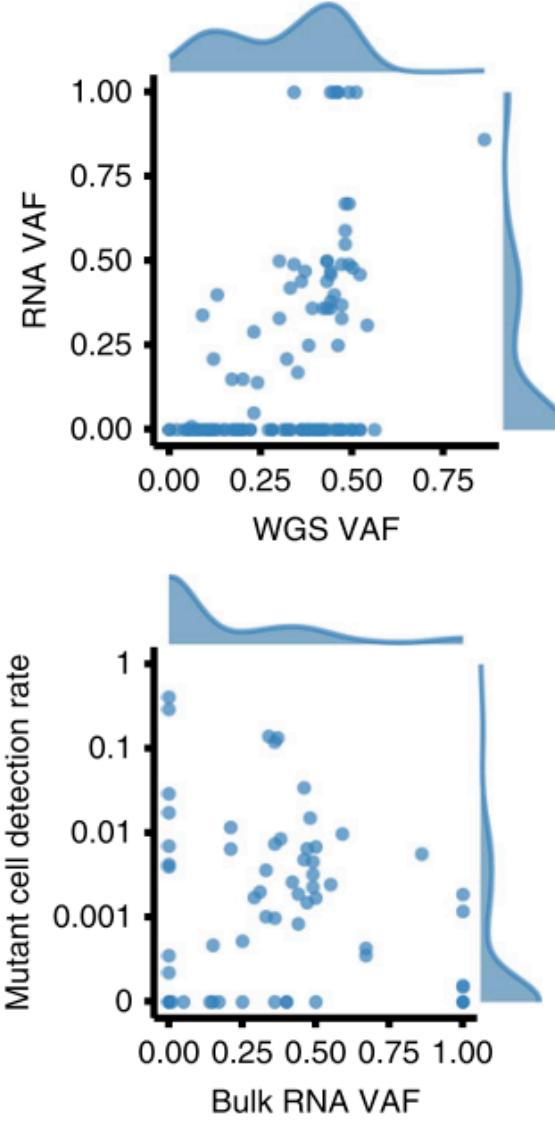
Fraction of unique transcripts (molecules) whose reads map to any given position up to 10 kbp away from the capture site in both the 5' and 3' kits.



Examples of variants found (Figure 1C)



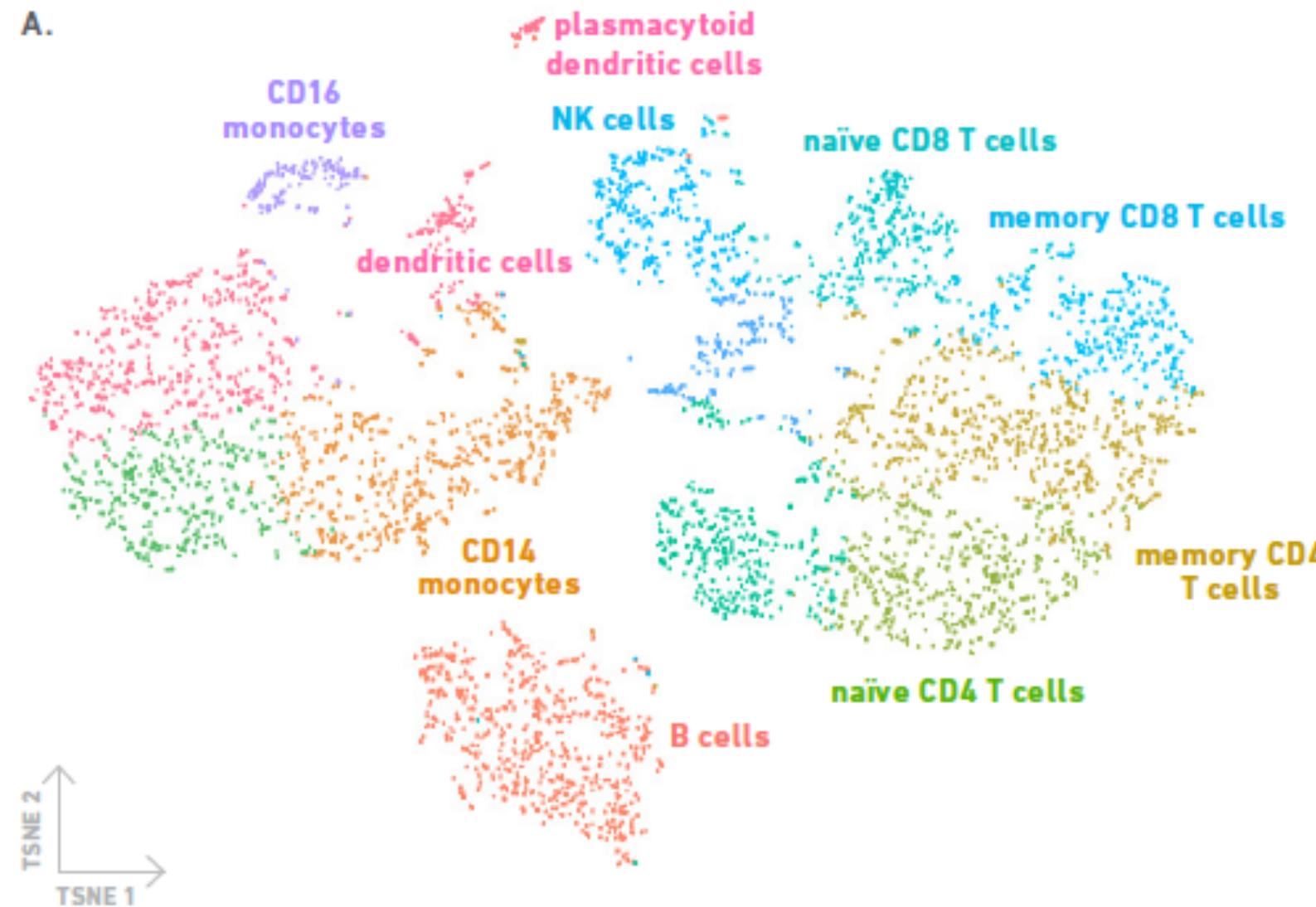
Detection limits (Figure 1D)



The two cell-based metrics were (1) Mutant Cell Fraction (MCF) and (2) Mutant Cell Detection Rate (MCDR)). The MCF for each variant was defined as M/T , where T is the number cells having coverage at the mutant site, and M is the number of cells having at least one mutant read at that site. The Mutant Cell Detection Rate (MCDR) was defined as the ratio of observed mutant cells to the number of expected mutant cells; for variants with coverage in bulk RNA-seq data, the number of expected mutant cells is twice the eWGS VAF.

- Variants at the 5' end of the transcript are detected more often.
- When the variant is detected in RNA-seq, the VAF in RNA-seq and WGS are correlated
- Frequently the variants from WGS are not detected in any cells

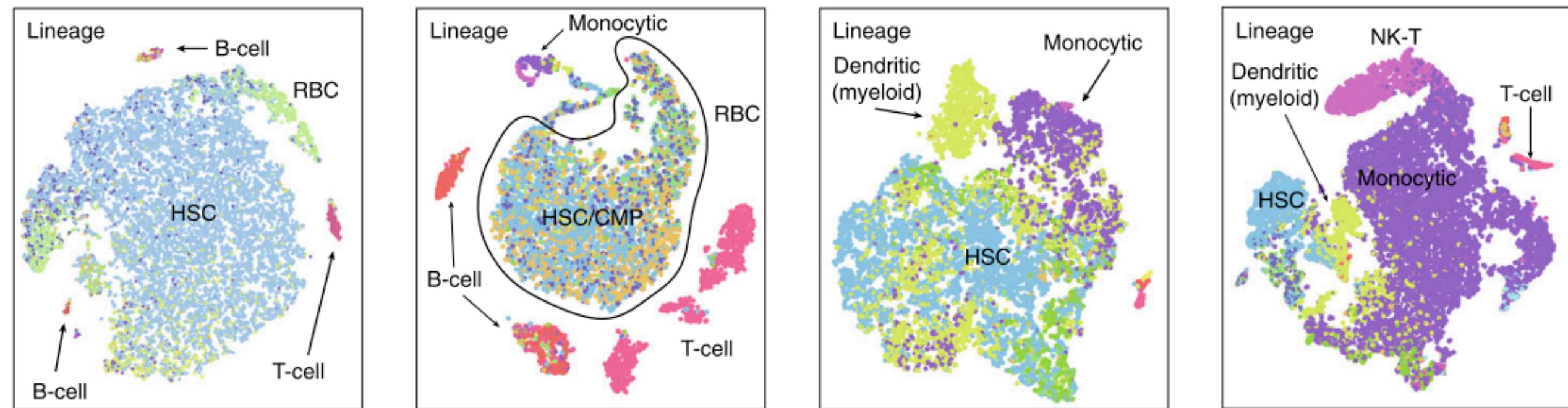
A.



Cell type identification of blood cells is usually performed using marker genes

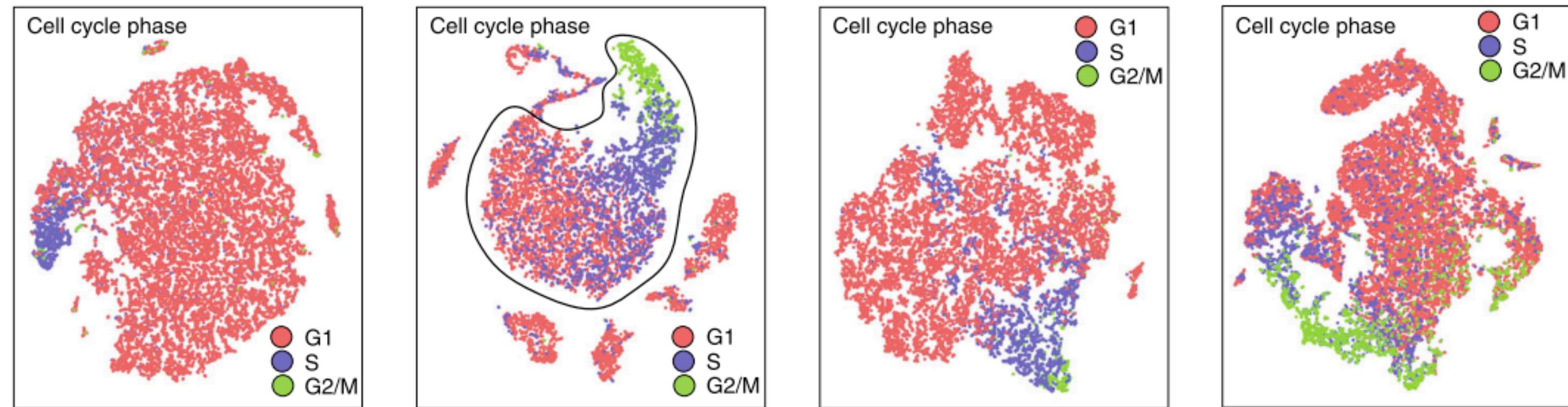
Figure from: 10x Genomics Applications Note “Improving Single Cell Characterization with Simultaneous Gene Expression and Cell Surface Protein Measurements at Scale”

Lineage identification (Figure 4)



"To investigate sample composition in a more unsupervised manner, we identified the nearest hematopoietic lineage of each cell by matching each cell's expression profile to the most similar lineage-specific expression profile in the DMAP database."

Cell cycle phase identification (Figure 4)



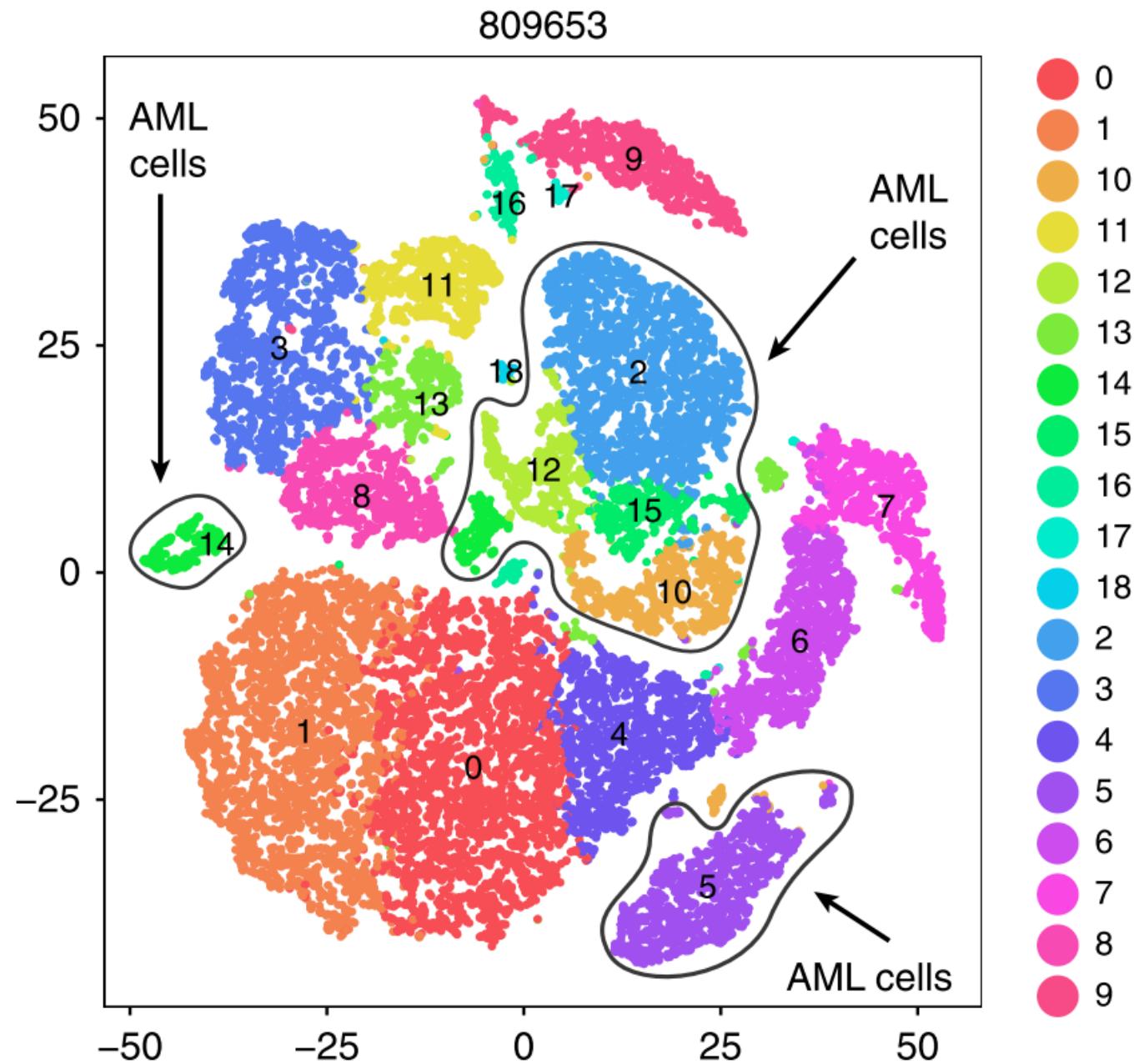
"Cell cycle phase was determined using methodology provided in Seurat, based on relative expression of phase-specific genes"

scRNA-seq is sparse, so each cell with a mutation has a very small chance of actually detecting it.

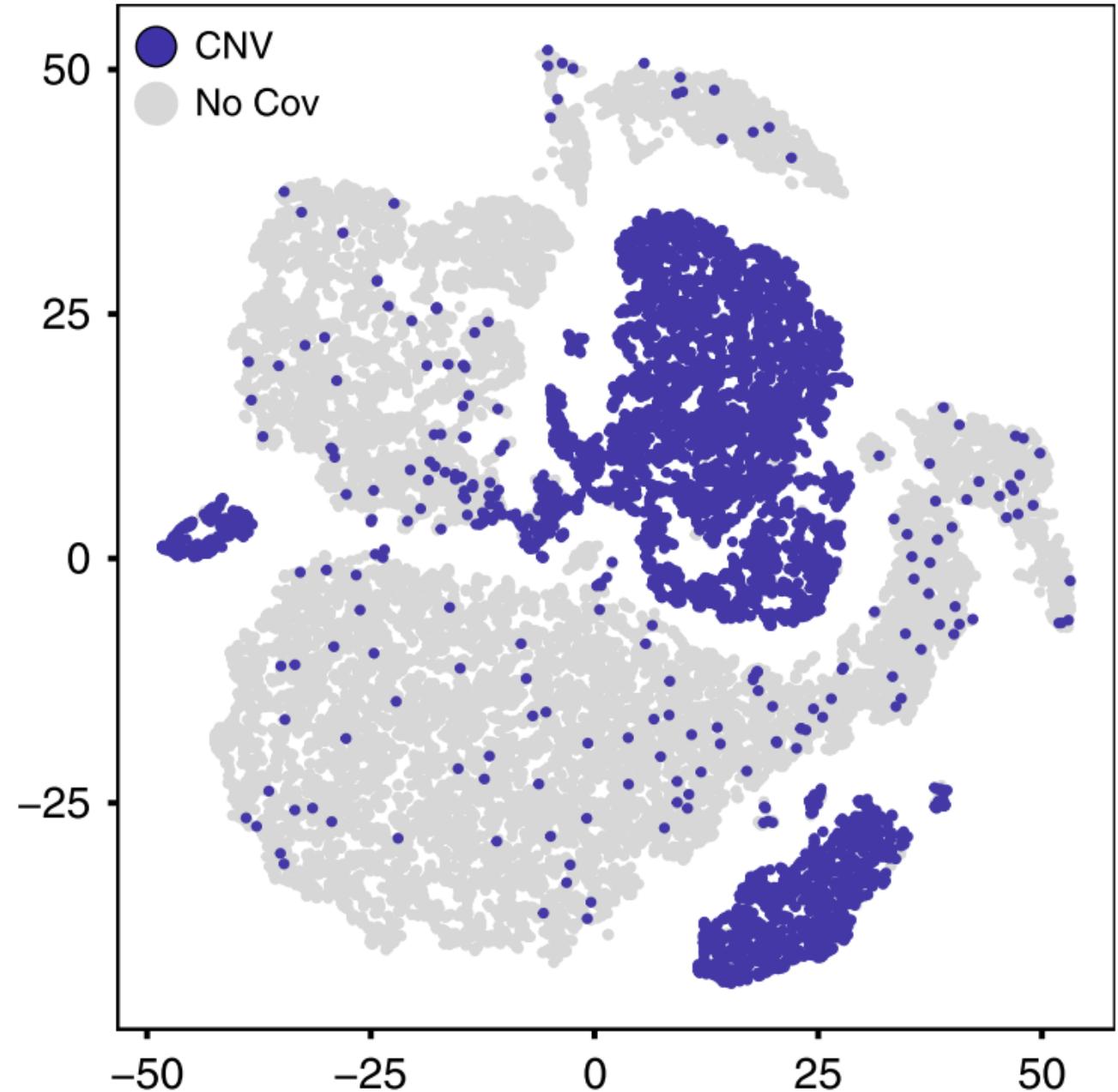
Median 1300 to 2400 genes detected per cell (depends on sample)

[...] most SNV locations were covered by a single read in most cells (although SNVs in several highly expressed, high-coverage genes (e.g. U2AF1, NPM1, SRSF2, and NRAS) were more likely to have multiple reads per cell) (Supplementary Fig. 1a). For a heterozygous mutation, therefore, there is a 50% chance that the observed transcript is mutant, and a 50% chance that it is wild-type, leading to the phenomenon known as allelic dropout. We therefore labeled a cell “mutant” if it contained at least one variant-containing read, and “unknown” if only wild-type reads or no reads were detected. We found an average of 49 mutant cells per variant (range: 1–3944), and 3732 mutant cells (22% of the total cells) per sample, but this varied widely among samples (range: 396–8200, or 1.8–52%), depending on the mutations present in each (Table 1). Most mutant cells contained one detected mutation, with one read mapping to the variant position.

t-SNE separates
cancer cells from
normal cells
(Figure 2A)



The AML cells have copy number variation (CNV) (Figure 2C)



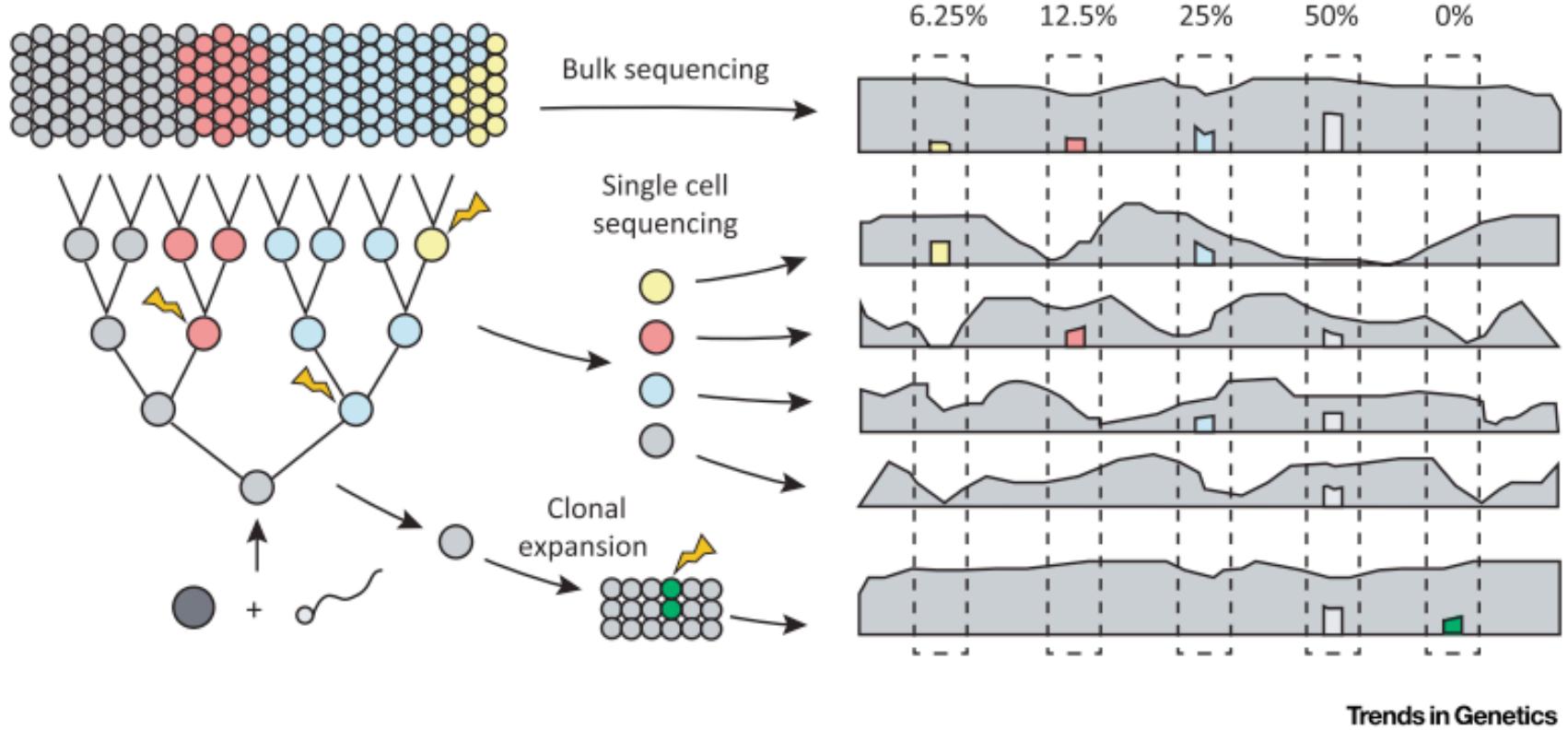
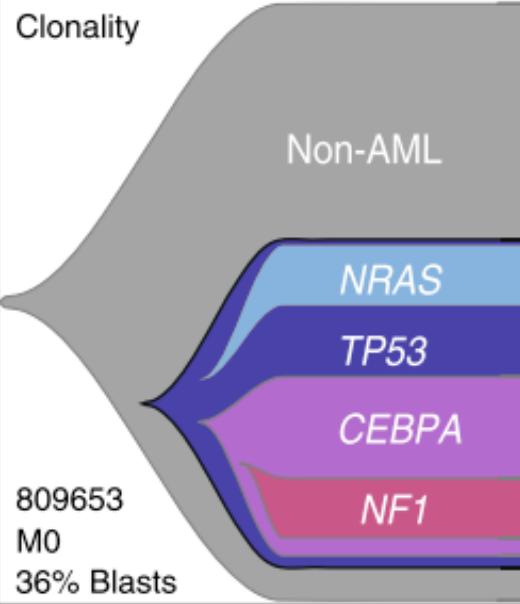


Figure 1. Somatic mutations arise during development and propagate to a subpopulation of cells (blue, 50% of cells; red, 25%; yellow, 12.5%). With bulk sequencing these somatic mutations are expected to be approximately half of the subpopulation frequency. Lower-frequency somatic mutations require higher sequencing depth to maintain detection sensitivity. With single-cell sequencing, somatic mutations can be detected as heterozygous variants that occur in a subset of cells. The ability to detect variants is dependent on uniformity of coverage and allelic balance in genome amplification, as well as on picking cells that contain variants. Clonal expansion followed by bulk sequencing does not suffer from the problems associated with single-cell sequencing, but artifactual mutations that occur early during expansion (green) can be difficult to distinguish from mutations in the original cell.



Single-cell mutation detection and interpretation in case 809653. (Figure 3)

d

