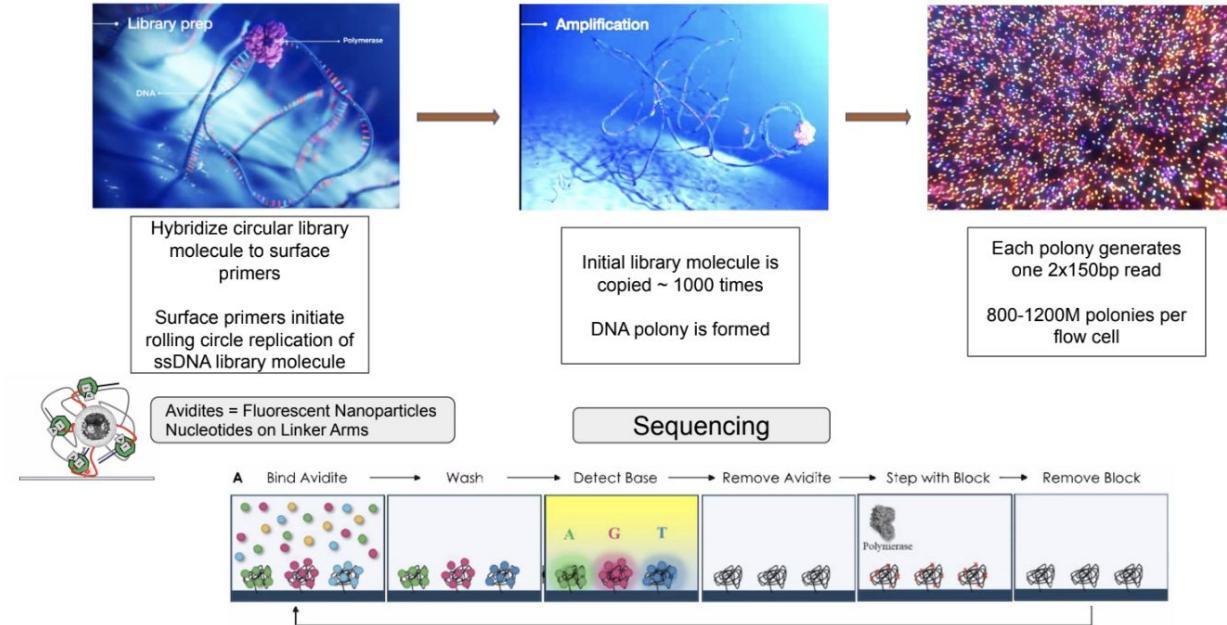


AVITI pilot results

Yiming Yang, Reshma Ramaiah, Gracie Gordon, Katie Geiger-Schuller and Bo Li

AVITI Technology

AVITI Benchtop Sequencer
from Element Biosciences



Pilot data overview

- Two runs
- Ch7: 3' v3.1, Gex and Hashing
- Ch10: 5' Gex, CITE, TCR and BCR

SampleID	IndexID	Pool	Percent Assigned (excluding PhiX)	Percent Assigned	Requested Assignment
Ch7_scRNA	SI-TT-B2	1	36.5%	65%	66%
Ch10_scRNA	SI-TT-A5		28.9%		
Ch7_CellPlex	SI-NN-A2	2	23.2%	24%	22%
Ch10_CITE	SI-TN-A1		0.8%		
Ch10_TCR	SI-TT-A1	3	5.4%	11%	12%
Ch10_BCR	SI-TT-A3		5.2%		

Total number of reads

- Downsampled fastq files are highlighted in red

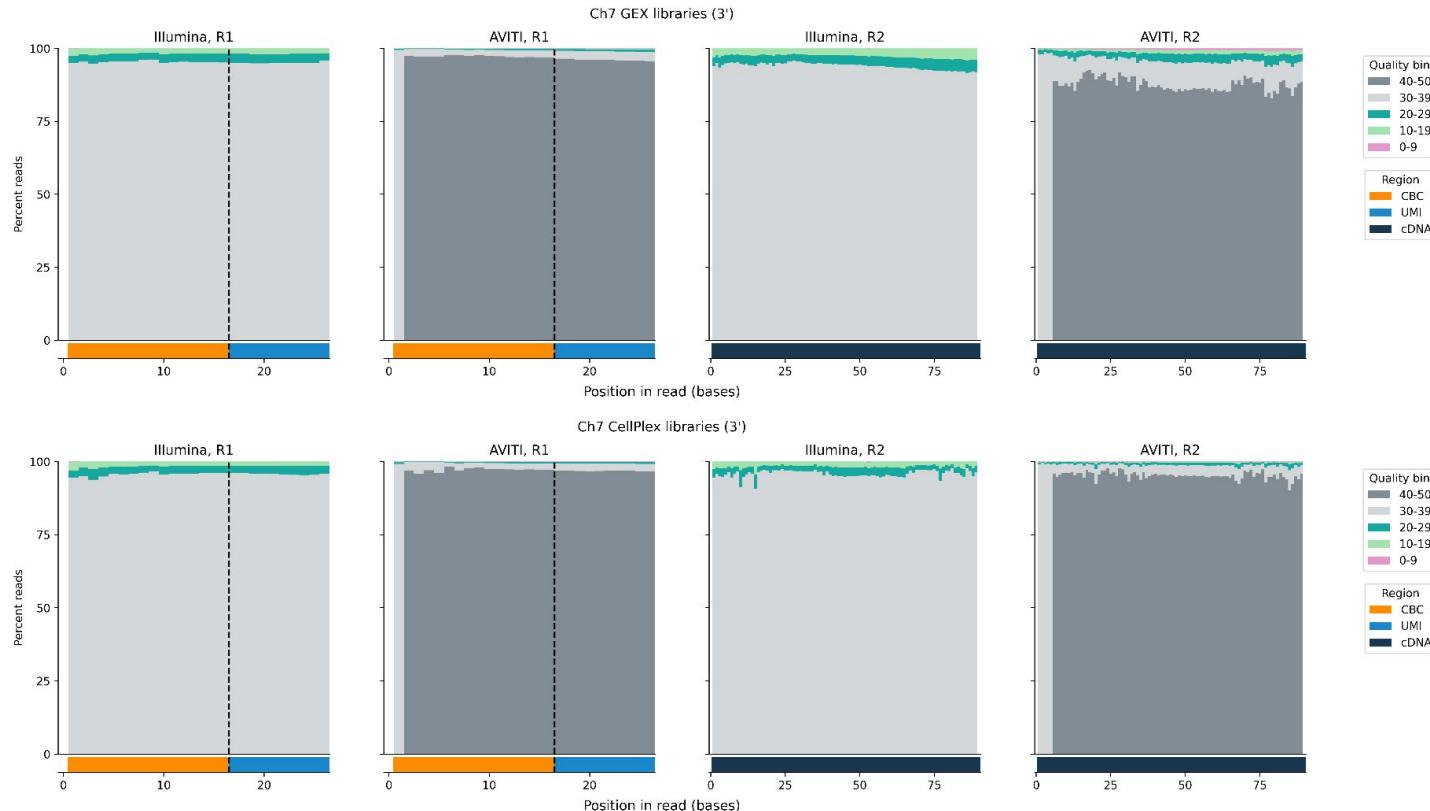
Library	ILLUMINA reads	AVITI reads
Ch7_scRNA	349,044,674	310,200,163
Ch7_CellPlex	118,710,802	197,573,702
Ch10_scRNA	290,756,618	245,394,995
Ch10_TCR	44,603,503	46,201,535
Ch10_BCR	47,792,057	44,548,596
Ch10_CITE	61,393,854	6,662,293

Sample Ch7

Experiment

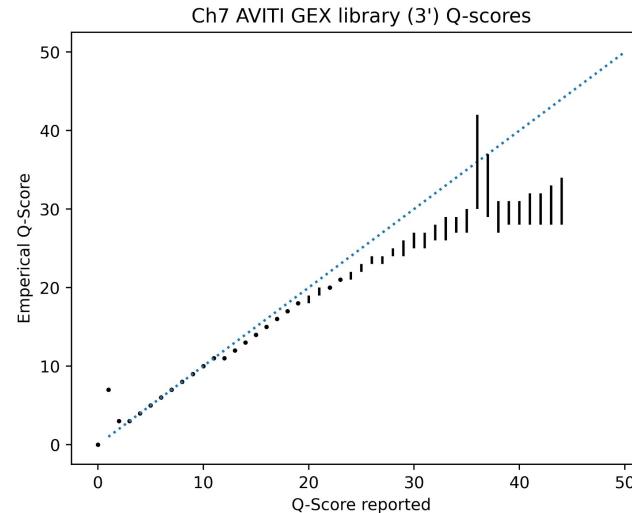
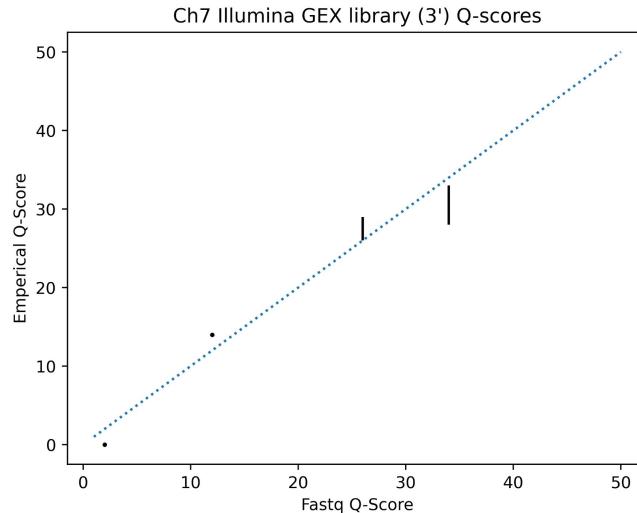
- Benchling lab notes
- 10x 3' v3.1 + CellPlex (40K cells loaded)
- 6 donors (from AllCells) mixed into 12 samples
 - Donor 1 (3052820): CMO301, CMO302
 - Donor 2 (3054238): CMO303, CMO304
 - Donor 3 (3053757): CMO305, CMO306
 - Donor 4 (3054136): CMO307, CMO308
 - Donor 5 (3054091): CMO309, CMO310
 - Donor 6 (3051144): CMO311, CMO312

AVITI has better instrument quality scores than Illumina



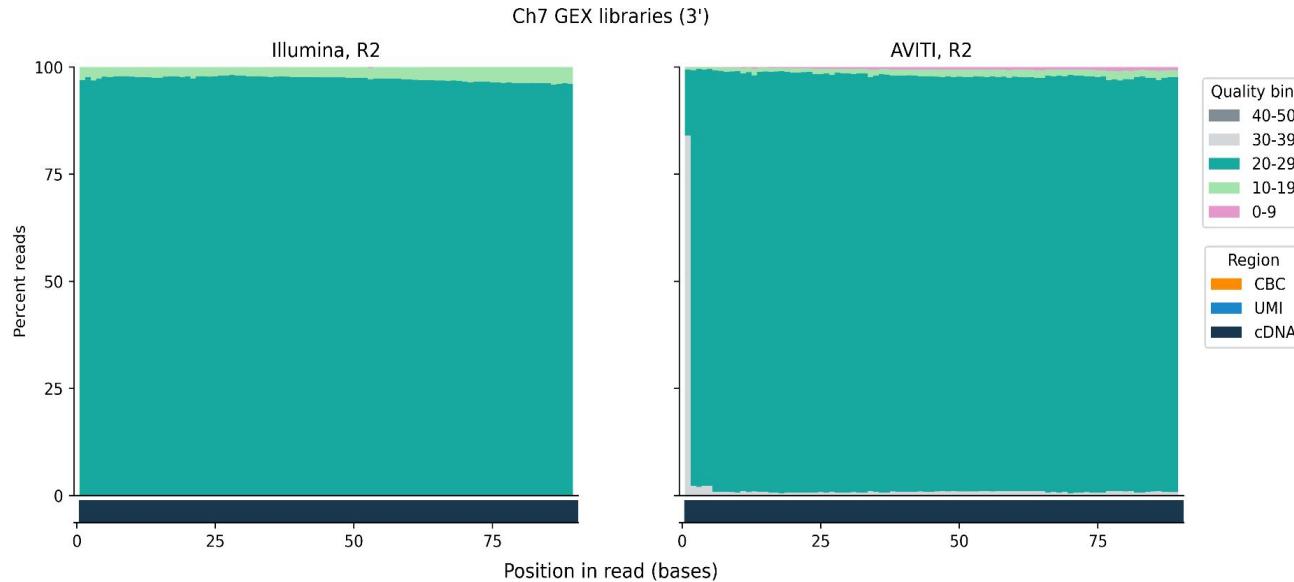
Estimating empirical quality scores from BAM files

- Align reads to the genome to generate the BAM file
- Calculate the error rate per read position based on the aligned BAM file (only consider mismatches) as lower bound
- Assume 1 SNP in 1000 bases and all SNPs are mismatches; previous error rate - 1e-3 as the upper bound



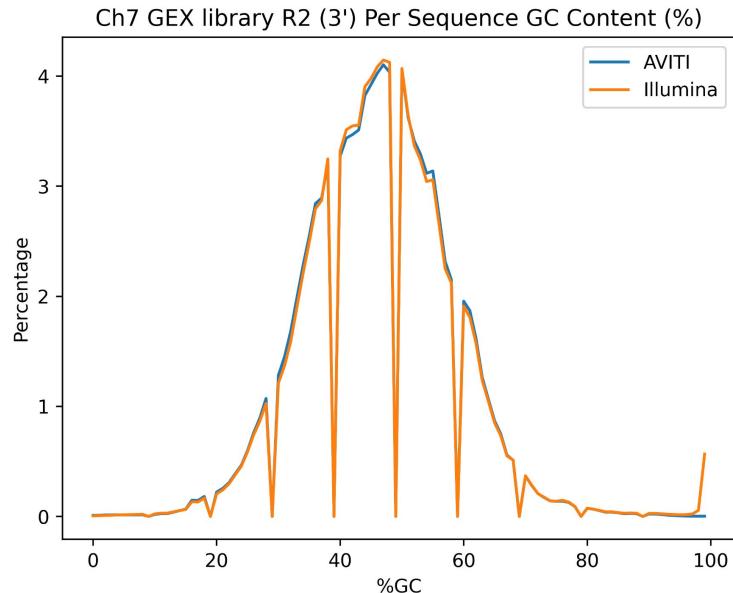
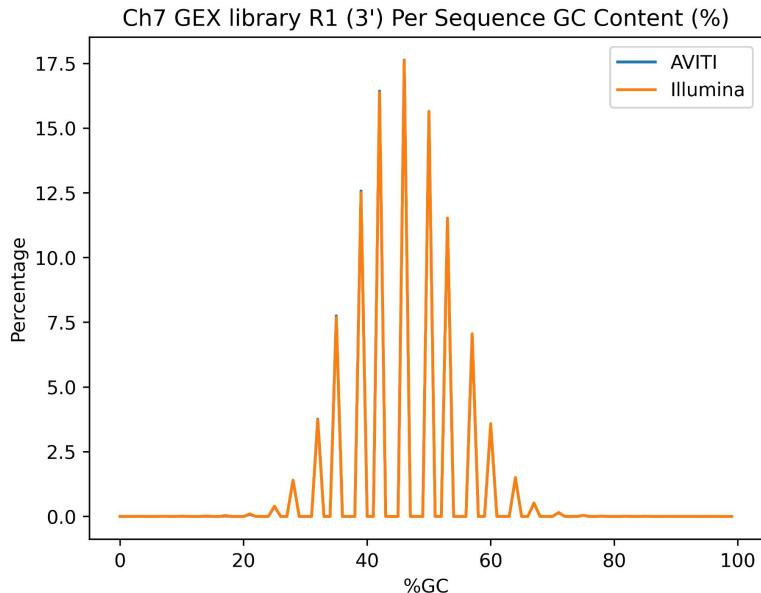
AVITI has better empirical quality scores than Illumina

- Empirical quality scores as means of the lower and upper bounds
- Replace instrument quality scores with empirical quality scores



AVITI and Illumina show similar GC content for the 3' assay

- For all reads, compute %GC as x axis
- Computer percentage of reads across different GC content as y axis



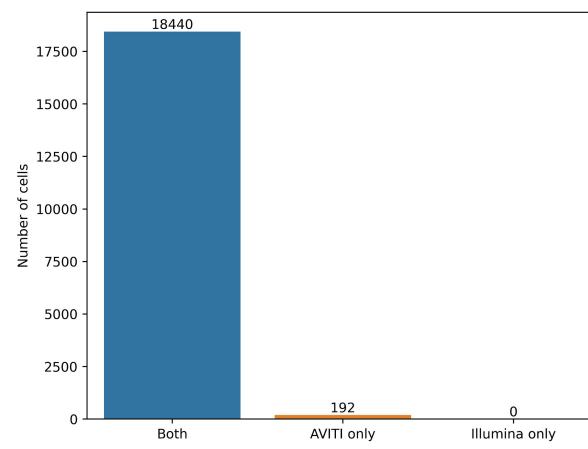
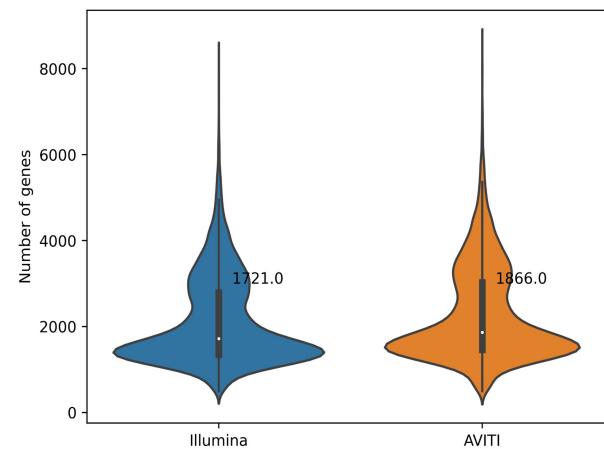
Align Ch7 GEX Reads to Genome

- Downsample Illumina reads to match the sequencing depth of the corresponding AVITI library
- Run both libraries with Cell Ranger v7.1.0; align to [GRCh38-2020-A](#) reference
- Alignment results are summarized below:

	ILLUMINA	AVITI
Estimated number of cells	18,450	18,512
Median genes per cell	1,721	1,875
Sequencing saturation	45.90%	39.60%
Reads Mapped Confidently to Genome	88.20%	89.10%
Reads Mapped Confidently to Transcriptome	70.90%	72.70%

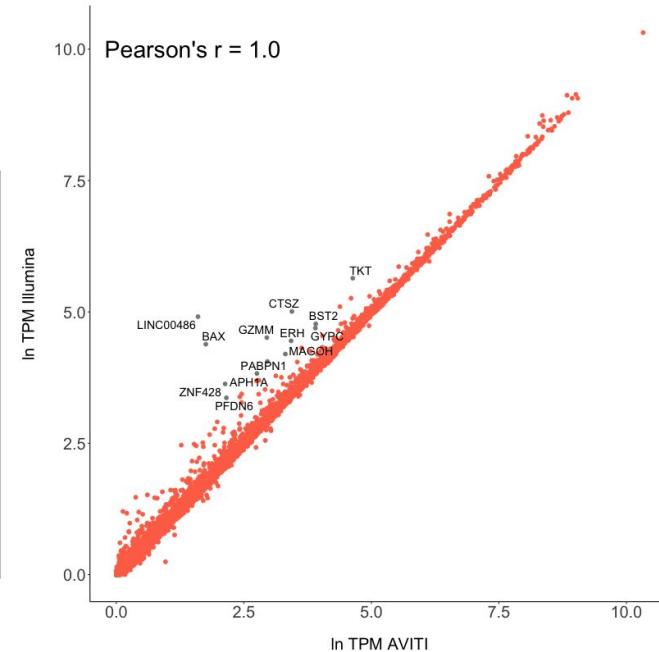
Align Ch7 GEX Reads to Genome

- Illumina data processing:
 - Downsample reads to match the sequencing depth of the corresponding AVITI library
- On both Illumina and AVITI libraries:
 - Run with Cell Ranger v7.1.0; align to [GRCh38-2020-A](#) reference



Distribution of number of genes is comparable

Number of cells identified is comparable



Gene expressions of Illumina and AVITI data are highly correlated

Demultiplexing based on CellPlex

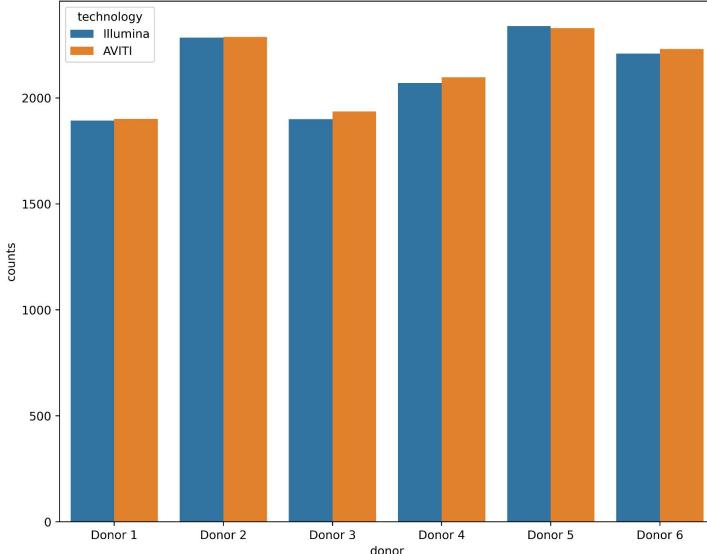
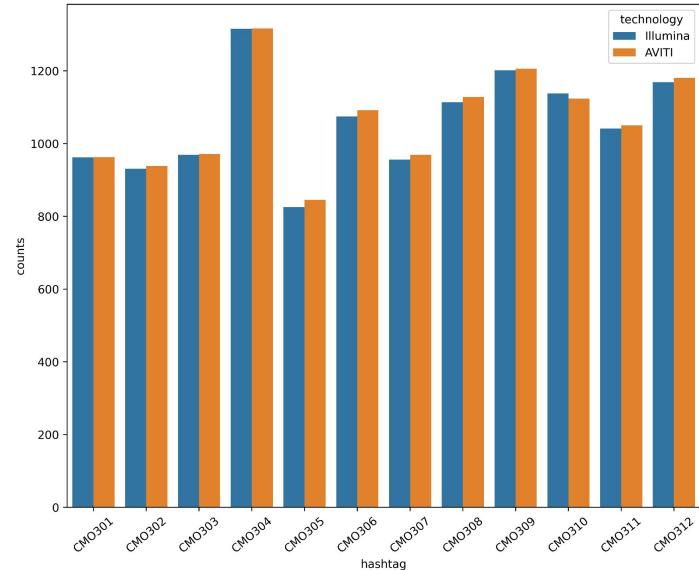
On CellPlex libraries:

- AVITI reads preprocessing:
 - Downsample to match the sequencing depth of the corresponding Illumina library
- Generate CMO count matrices:
 - Run [cumulus_feature_barcoding](#) to extract feature-count matrices

Demultiplexing:

- Run DemuxEM for each pair of GEX and CMO count matrices.

	Total singlets after demuxEM	Percentage of singlets (%)
Illumina	12698	68.86
AVITI	12785	68.62

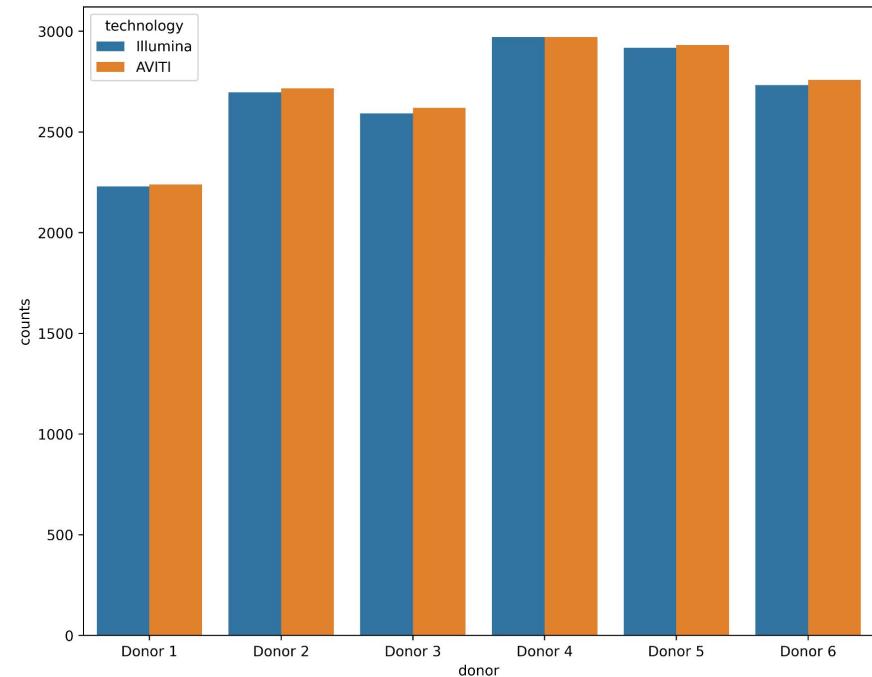


Demultiplexing based on genetics

Demultiplexing:

- Run souporcell for each pair of GEX count matrix and BAM tag file.

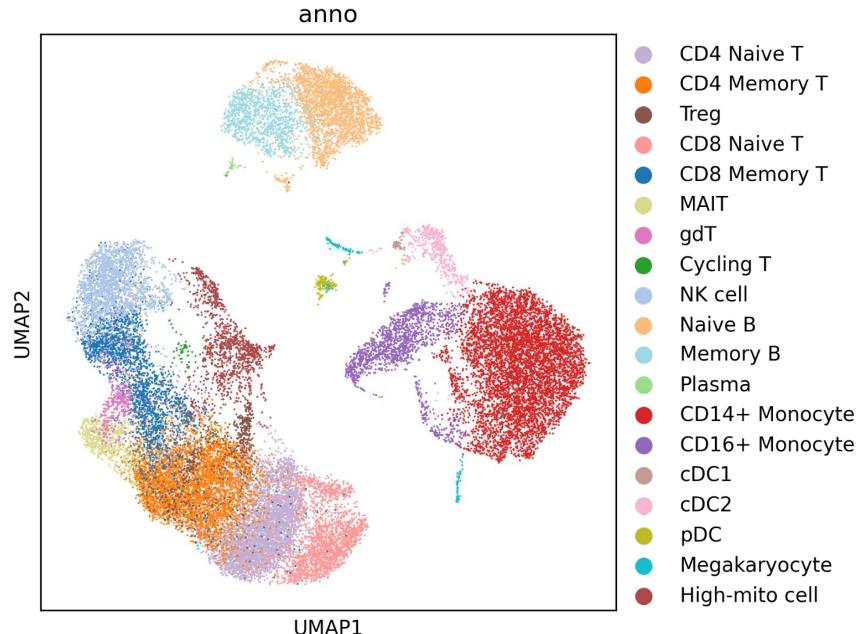
	Total singlets after Souporcell	Percentage of singlets (%)
Illumina	16141	87.53
AVITI	16235	87.14



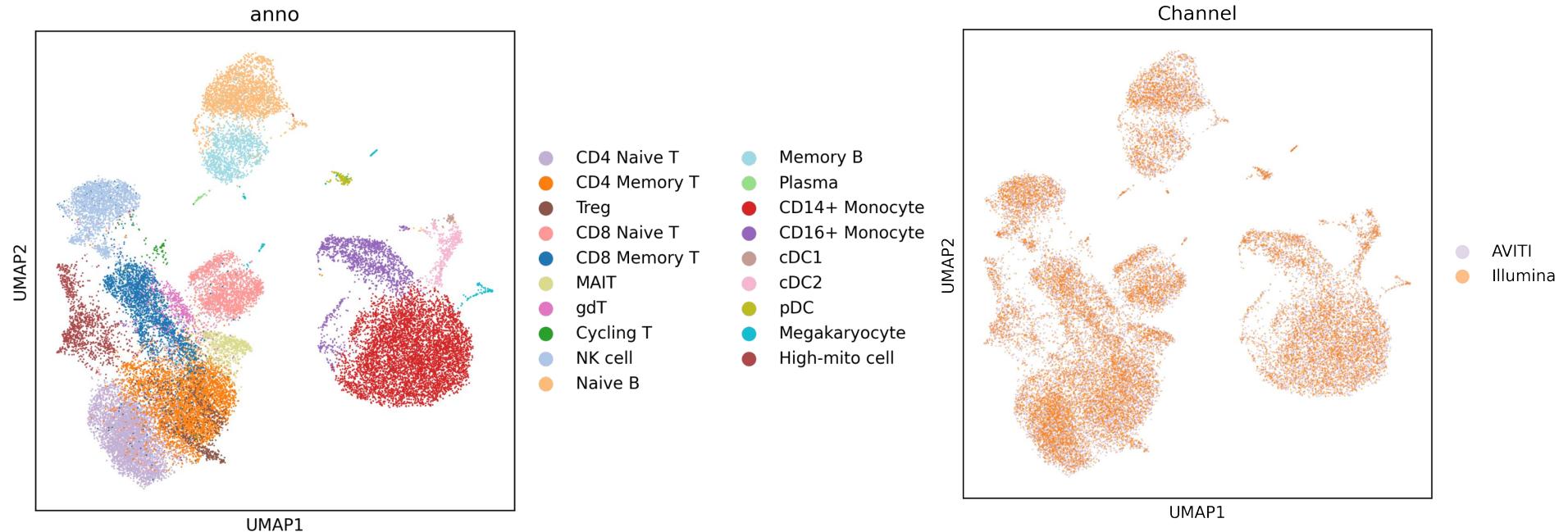
Cell type annotation procedure

Overview:

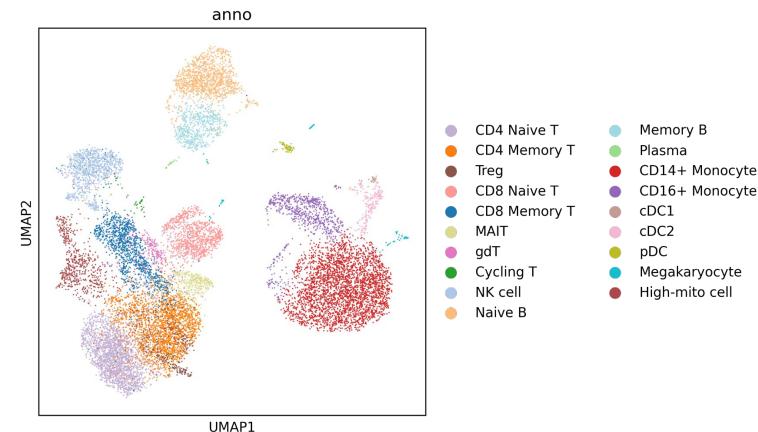
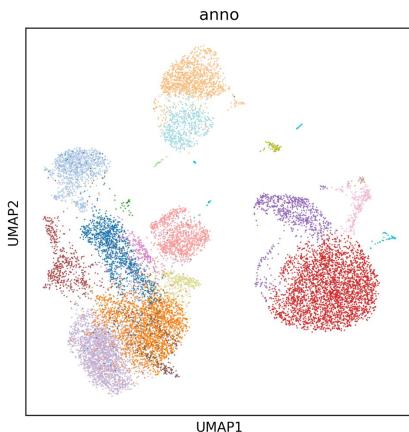
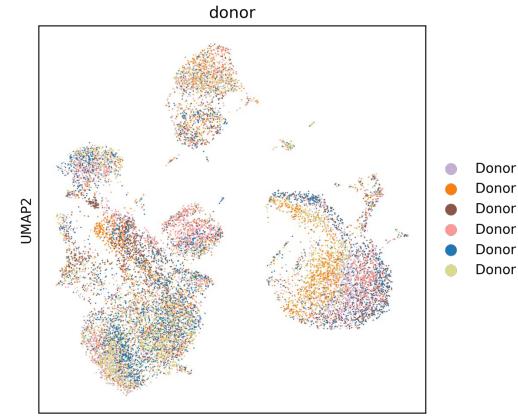
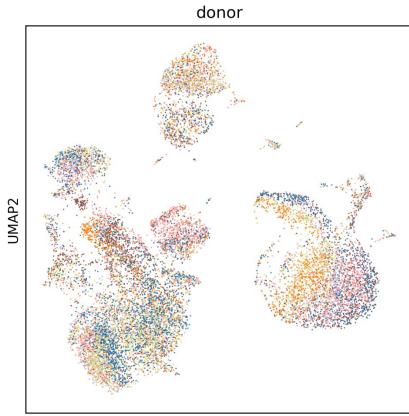
- Aggregate Illumina and AVITI count matrices, restricting to common cells to avoid tech bias
- QC: keep cells with # genes ≥ 500 and % mito $< 20\%$
- Use Scrublet-like method to remove doublets
- Select 2000 highly variable genes
- PCA with 36 PCs due to random matrix theory
- Data integration using Harmony algorithm
- Construct kNN graph with $k=100$
- Leiden clustering with resolution 1.3
- Subclustering:
 - B cell and Dendritic cell: subclustering to identify subtypes
 - T cell and NK cell: integrative NMF to identify subtypes
- Putative cell type annotation:
 - Using human immune markers
 - For T cells, using Azimuth markers



AVITI and Illumina data overlay perfectly without batch correction



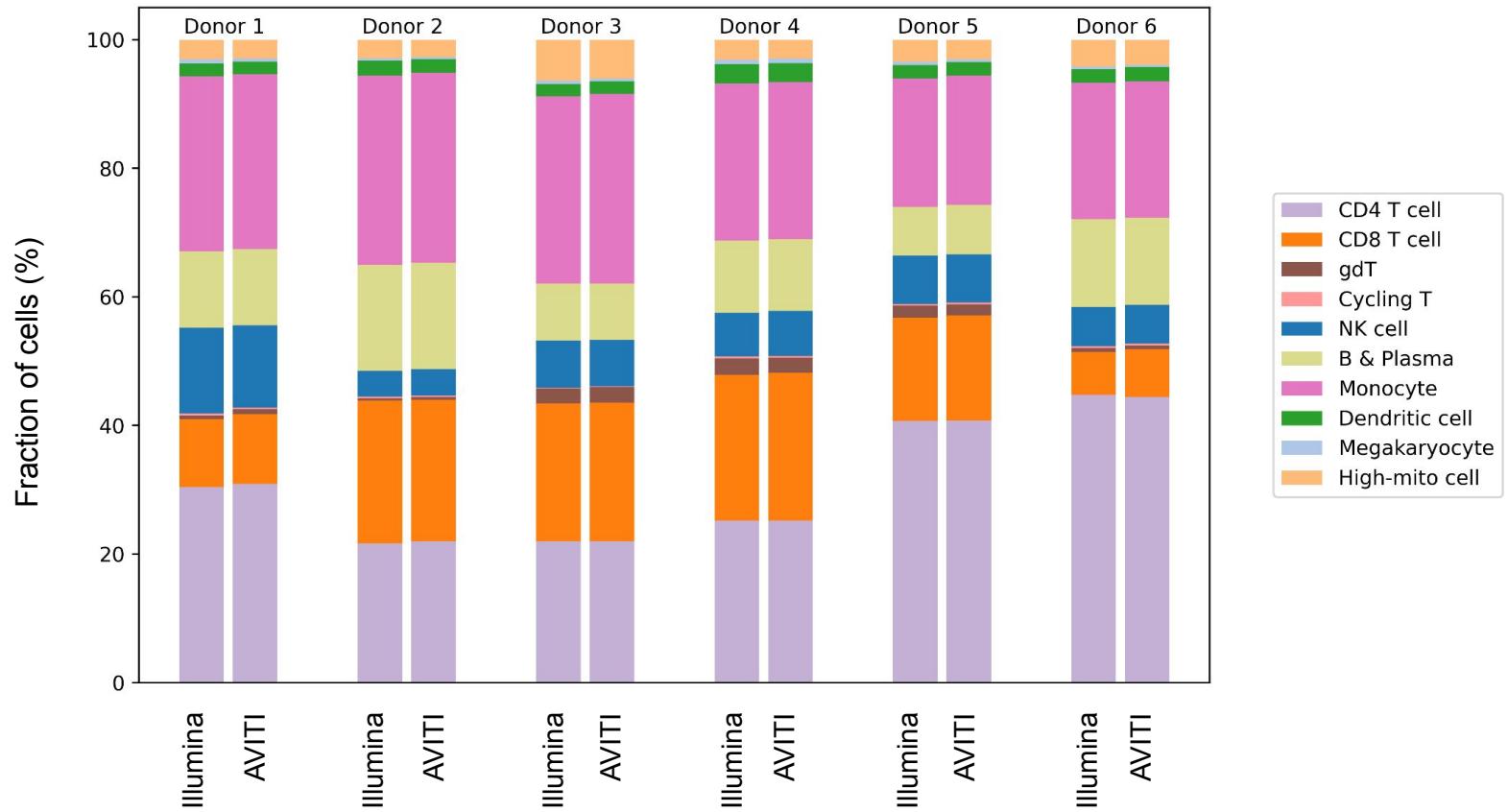
No Batch Correction: AVITI achieves comparable results



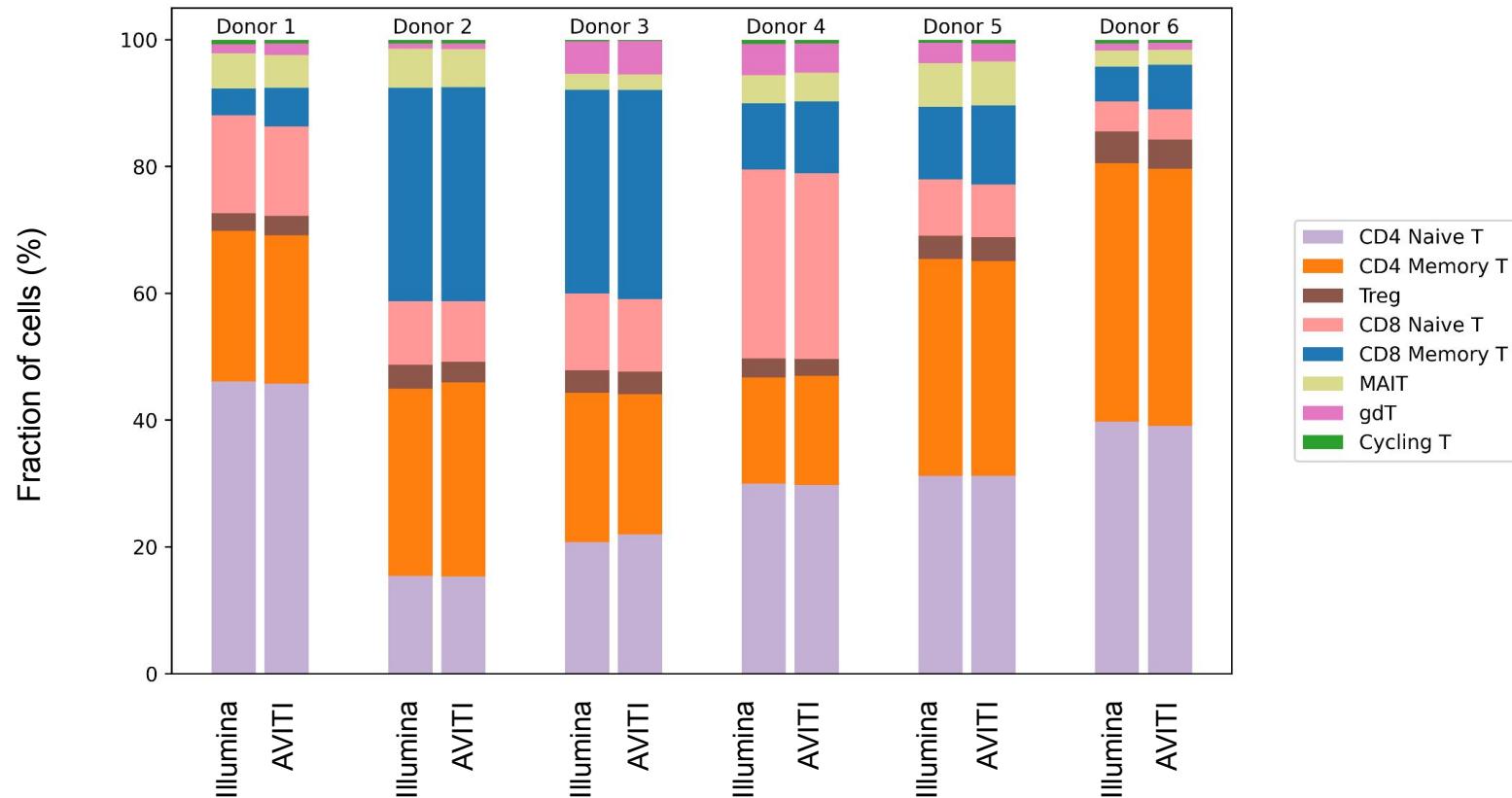
- Donor 1
- Donor 2
- Donor 3
- Donor 4
- Donor 5
- Donor 6

- CD4 Naive T
- CD4 Memory T
- Treg
- CD8 Naive T
- CD8 Memory T
- MAIT
- gdT
- Cycling T
- NK cell
- Naive B
- Memory B
- Plasma
- CD14+ Monocyte
- CD16+ Monocyte
- cDC1
- cDC2
- pDC
- Megakaryocyte
- High-mito cell

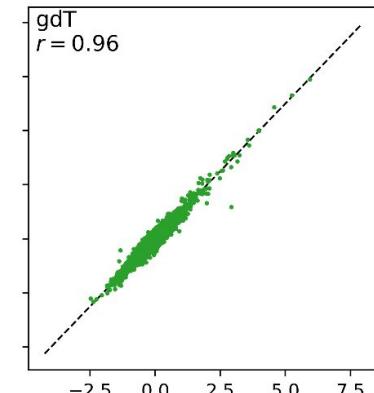
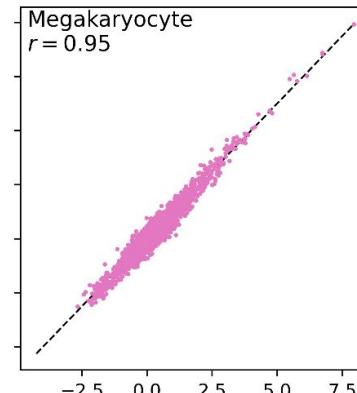
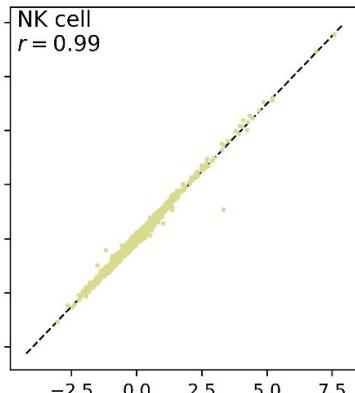
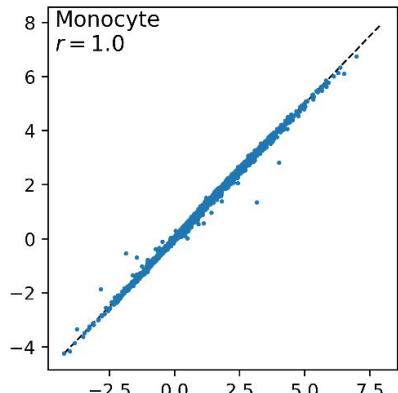
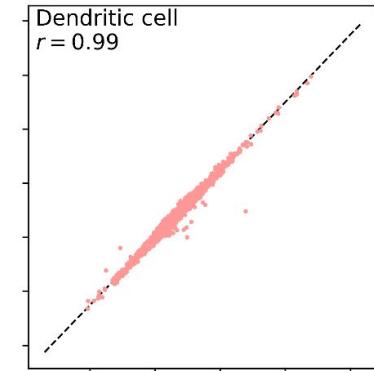
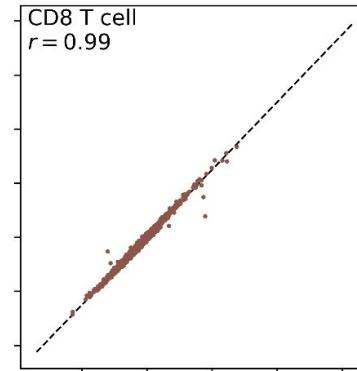
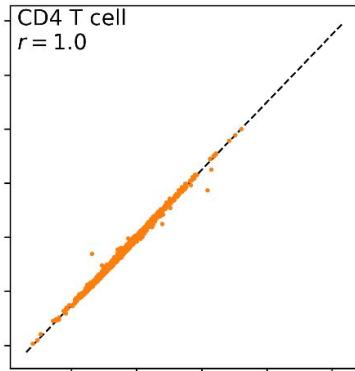
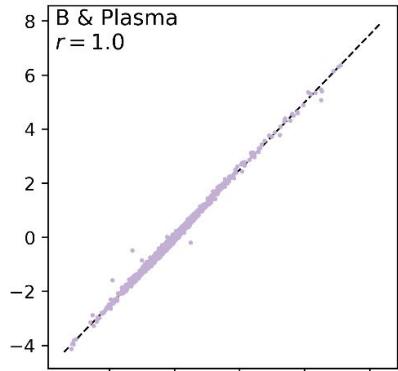
Cell type composition



T cell type composition



AVITI achieves consistent DE results across cell types



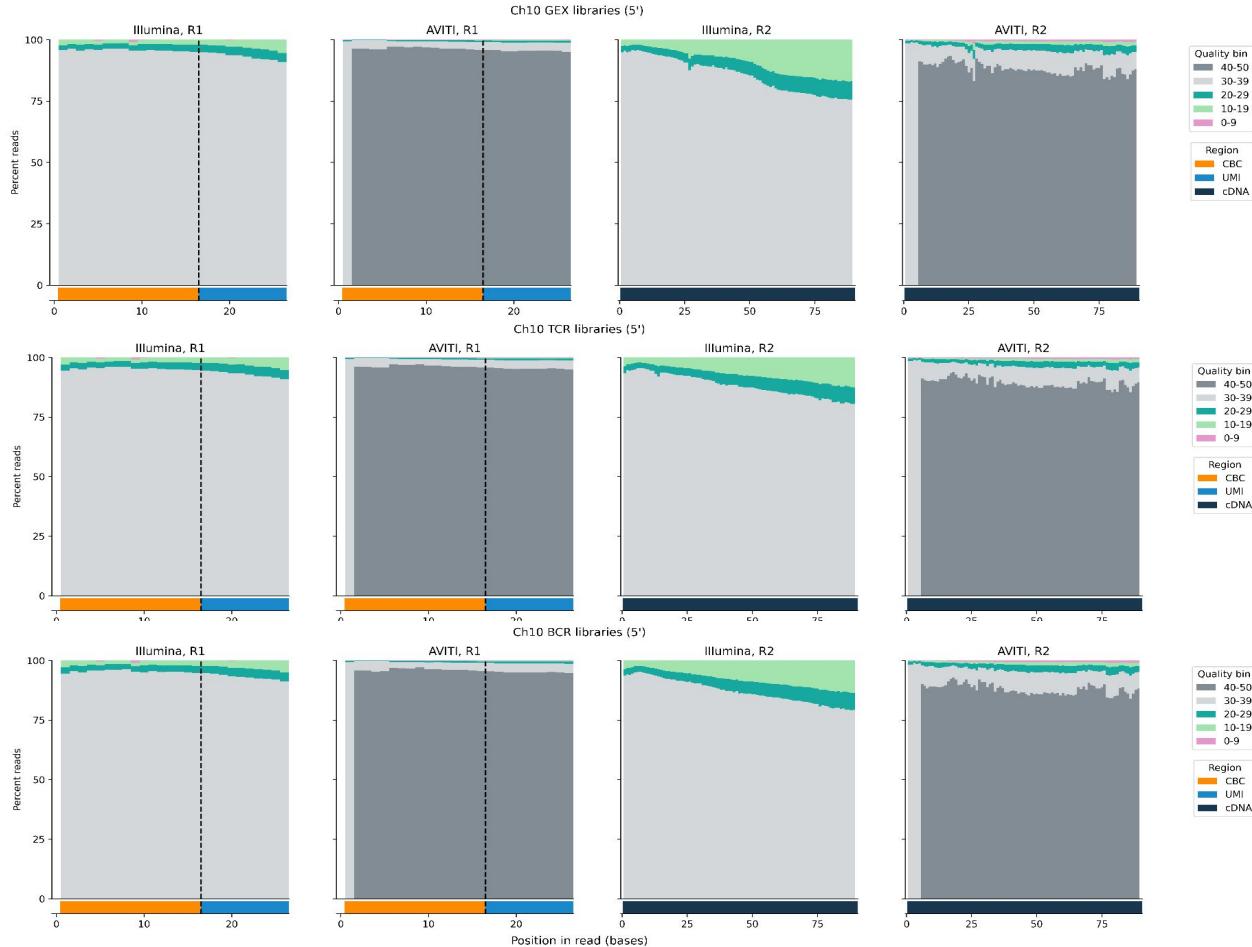
logFC of genes in Illumina (x-axis) vs AVITI (y-axis)

Sample Ch10

Experiment

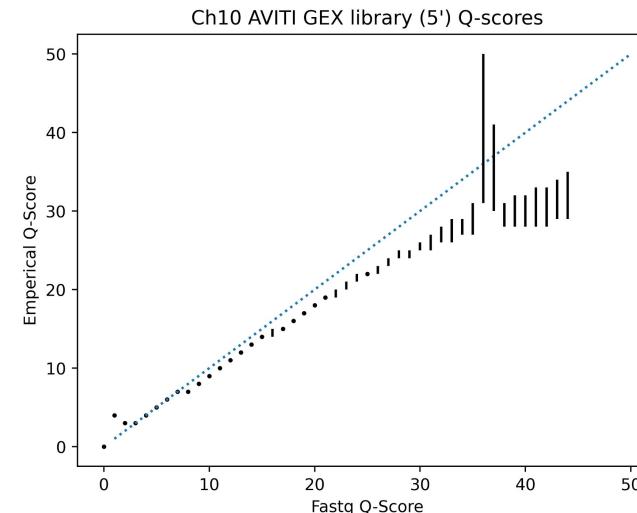
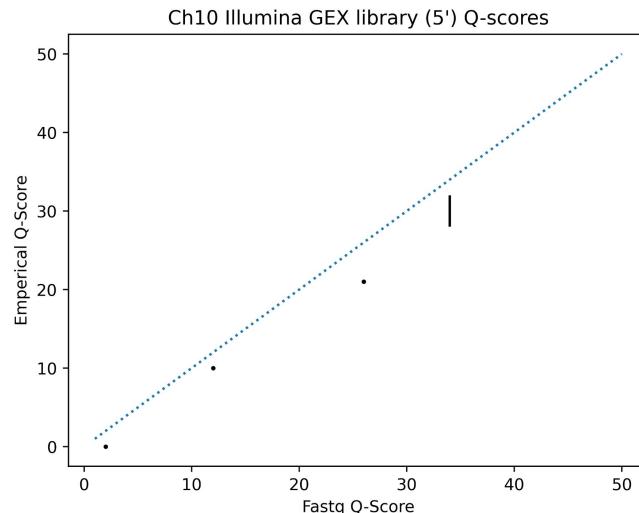
- [Benchling lab notes](#)
- 10x 5'+ CITE-Seq + TCR + BCR (?K cells loaded)
- Genetic multiplexing of 8 donors

AVITI has better instrument quality scores than Illumina



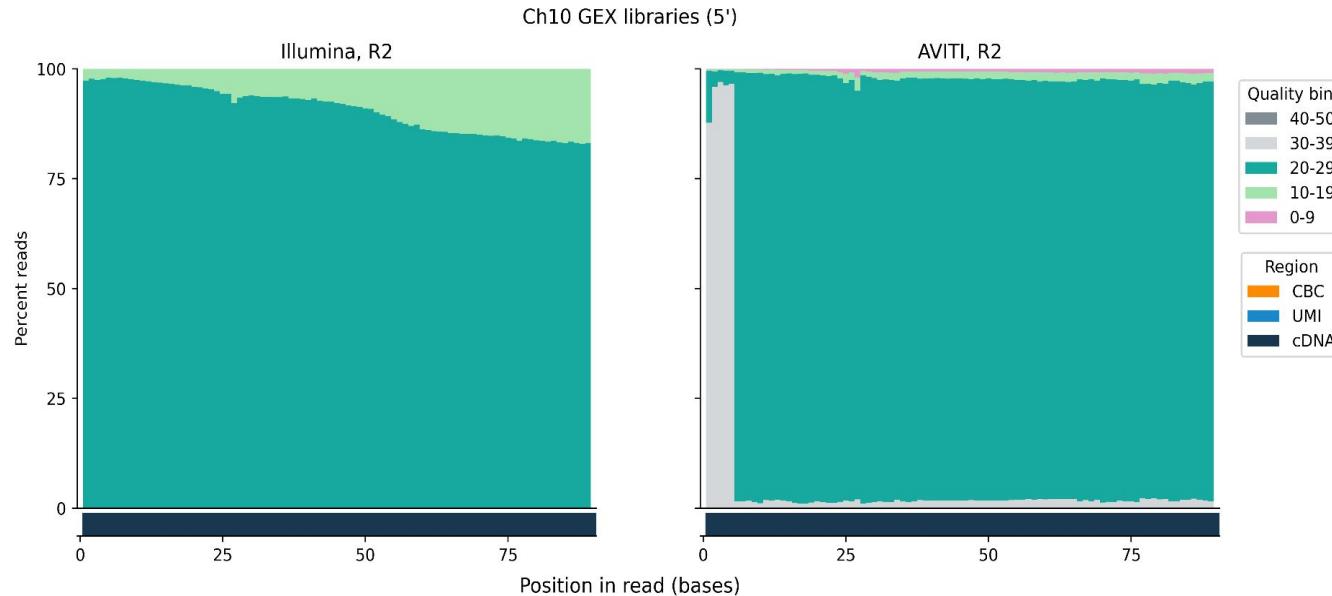
Estimating empirical quality scores from BAM files

- Align reads to the genome to generate the BAM file
- Calculate the error rate per read position based on the aligned BAM file (only consider mismatches) as lower bound
- Assume 1 SNP in 1000 bases and all SNPs are mismatches; previous error rate - 1e-3 as the upper bound



AVITI has better empirical quality scores than Illumina

- Empirical quality scores as means of the lower and upper bounds
- Replace instrument quality scores with empirical quality scores



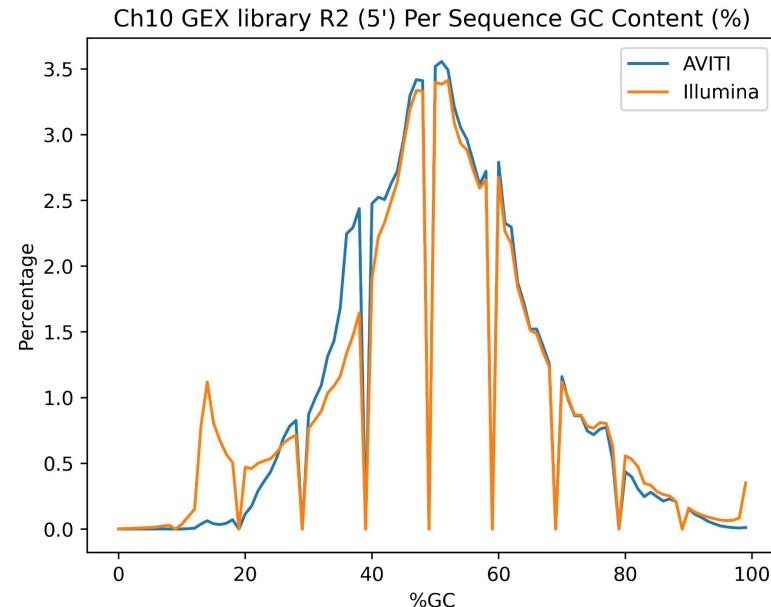
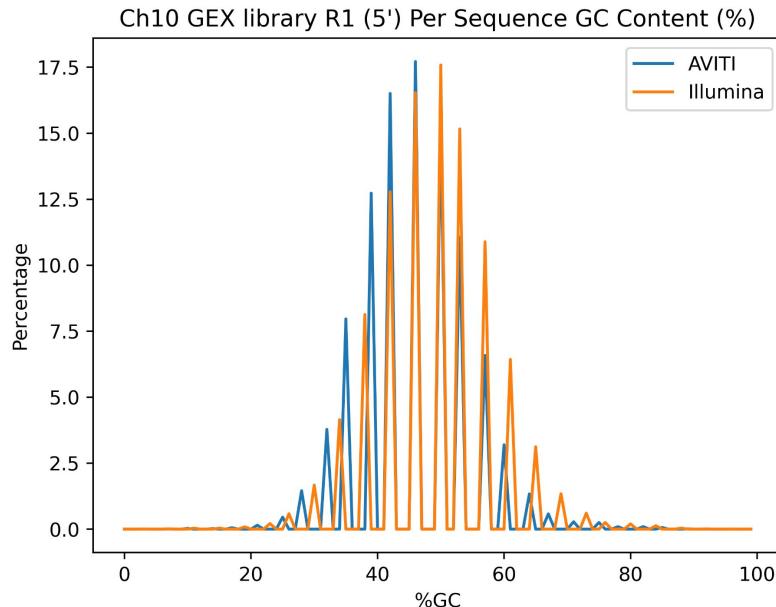
Genome alignment summary statistics

- For GEX data only
- Run Cell Ranger (run_count) and use web_summary.html

	ILLUMINA	AVITI
Estimated number of cells	28,221	28,507
Median genes per cell	782	857
Sequencing saturation	45.30%	41.40%
Reads Mapped Confidently to Genome	64.50%	68.10%
Reads Mapped Confidently to Transcriptome	53.50%	56.80%

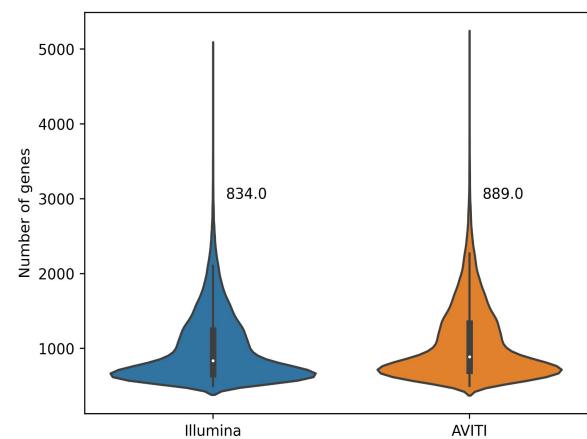
What's the pattern of GC bias?

- For GEX data only, calculate percentage of GCs
- Plot a histogram (number of reads vs. % GC)

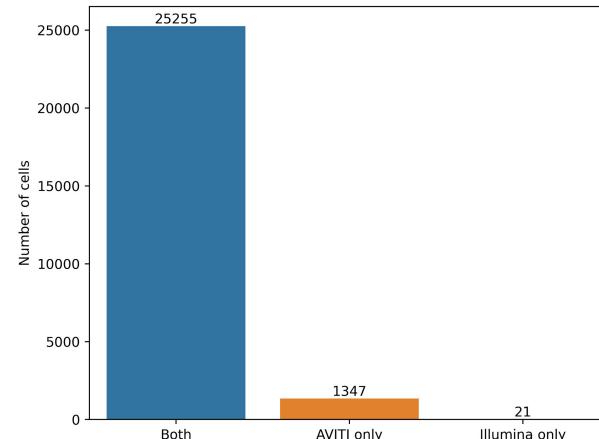


Align Ch10 GEX Reads to Genome

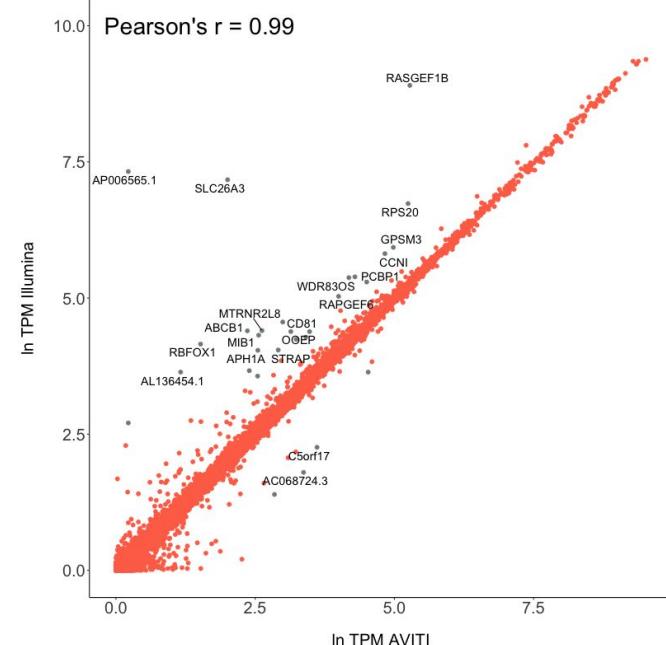
- Illumina data processing:
 - Downsample reads to match the sequencing depth of the corresponding AVITI library
- On both Illumina and Ultima libraries:
 - Run with Cell Ranger v7.1.0; align to [GRCh38-2020-A](#) reference



Distribution of number of genes is comparable



Number of cells identified is comparable



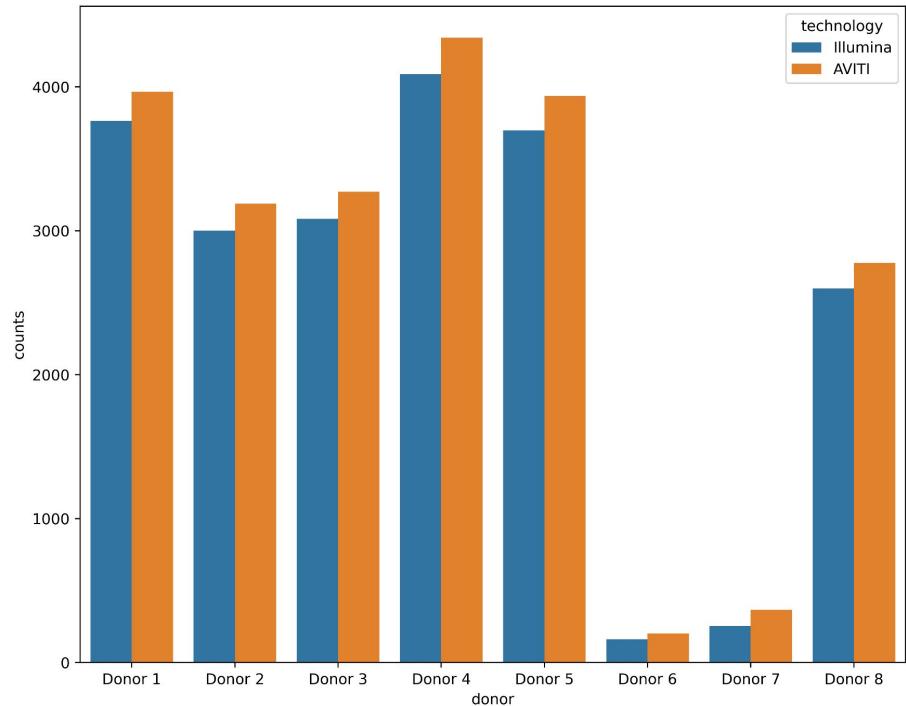
Gene expressions of Illumina and AVITI data are highly correlated

Demultiplexing based on genetics

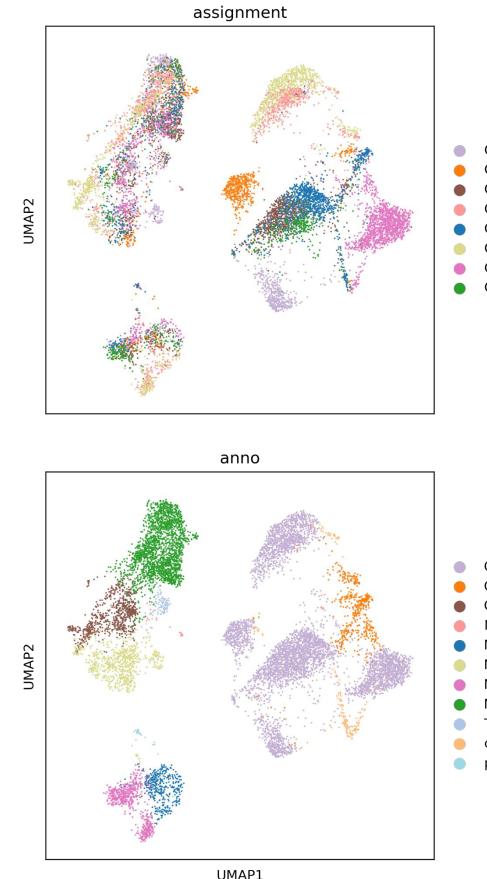
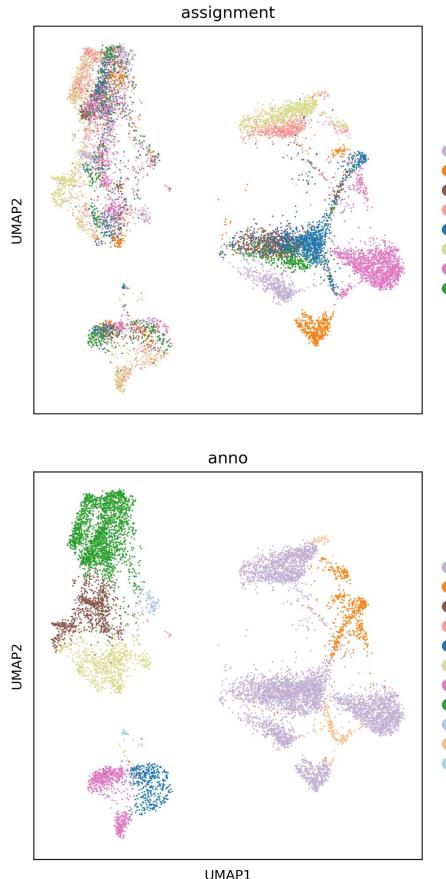
Demultiplexing:

- Run souporcell for each pair of GEX.

	Total singlets after Souporcell	Percentage of singlets (%)
Illumina	20,647	81.69
AVITI	22,050	82.89

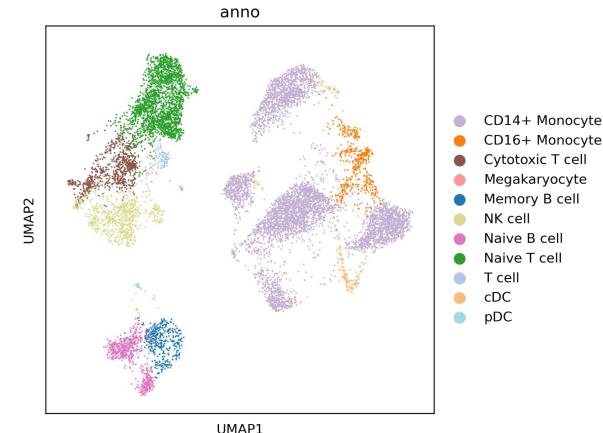
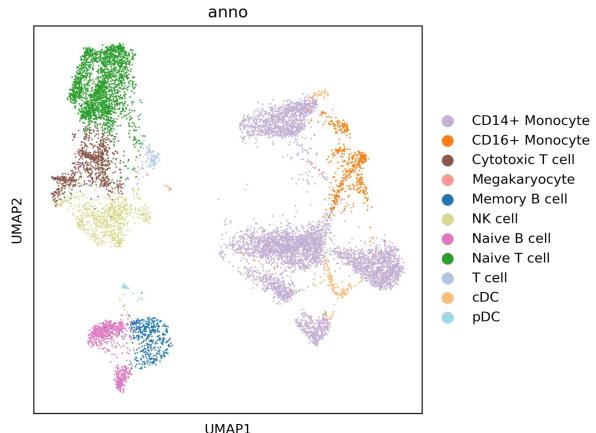


No Batch Correction: Ultima achieves comparable results

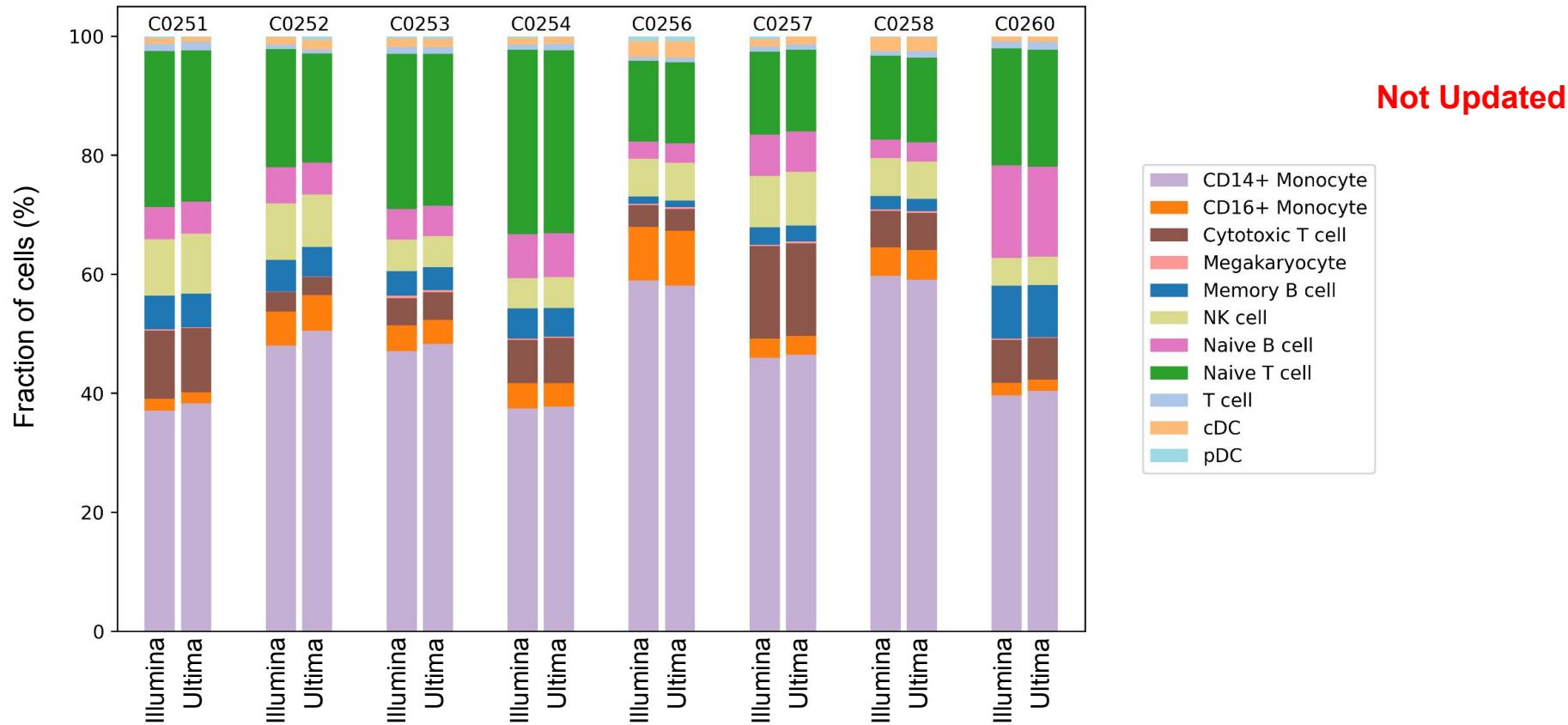


Illumina

Ultima

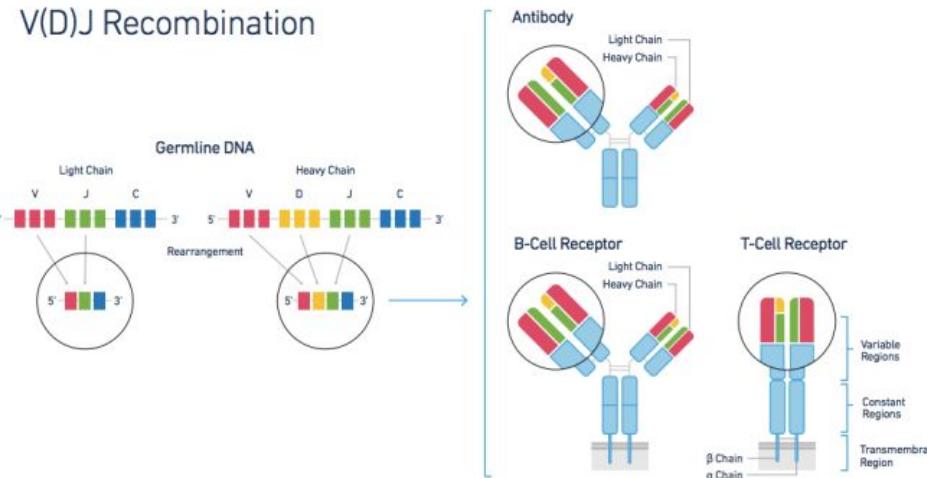


Cell type composition: Ultima achieves comparable results



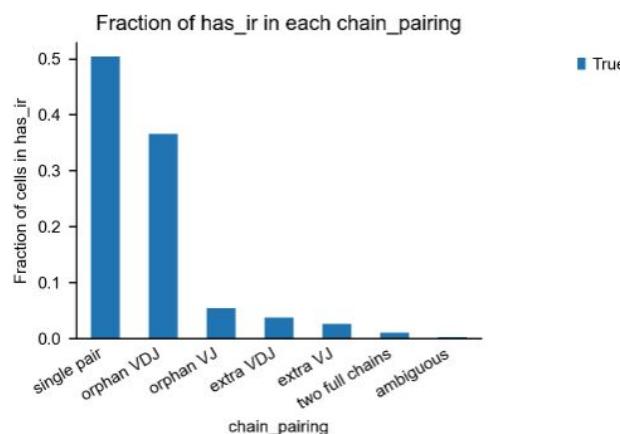
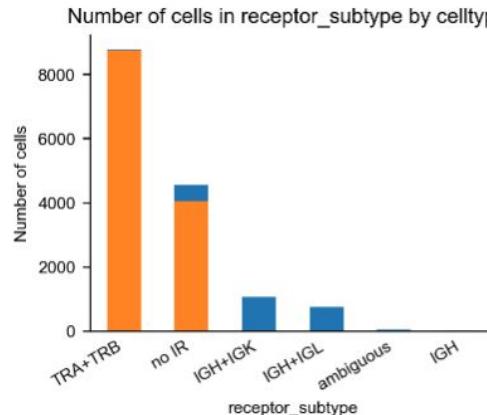
IR Comparison: Brief Introduction

- TCR and BCR Sequencing Data



- Questions
 - What are the similarities between called clonotypes between Illumina + AVITI?
 - Are the defined clonotype networks the same between Illumina + AVITI?
 - Are clonotype expansions detected equally with Illumina + AVITI?
 - Are VDJ annotations concordant between the two sequencing technologies?
 - Do we detect the same marker genes in the top clonotypes?

IR Comparison: Chain detection

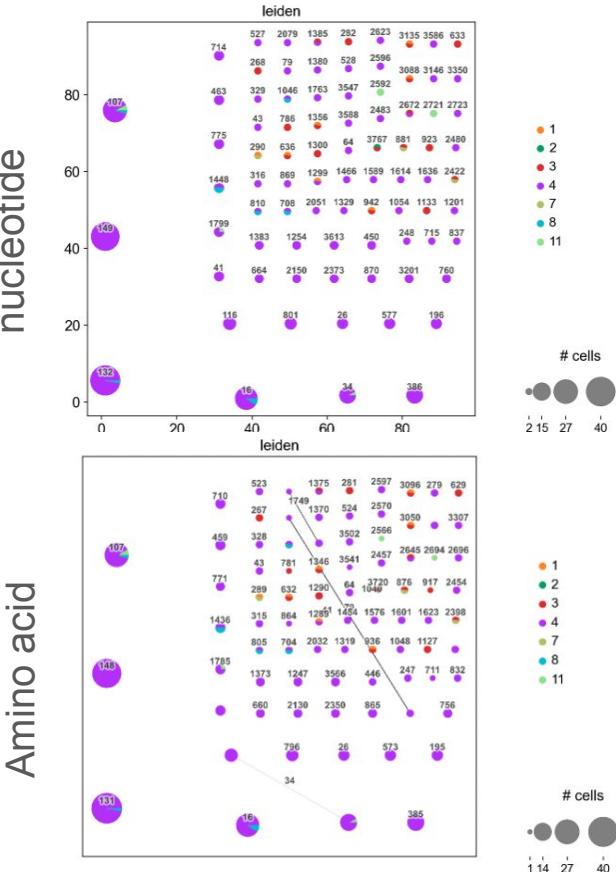


AVITI

Gracie Gordon

TCR Comparison: Clonotype Networks

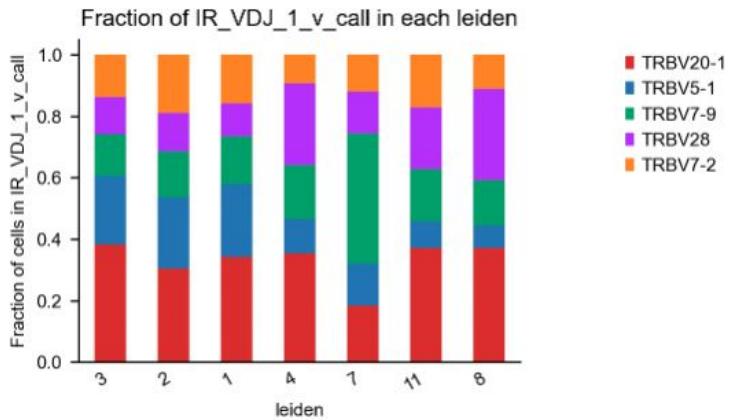
Illumina



Aviti

Gracie Gordon

TCR Comparison: V gene usage



AVITI

Illumina

Gracie Gordon

TCR Comparison: VDJ antigen annotation

Illumina

AVITI

Gracie Gordon

IR Comparison: T cells

Illumina

AVITI

Gracie Gordon