Comparison of Illumina and Ultima Sequencing Platforms for 10x Genomics Single-Cell RNA Sequencing

INTRODUCTION

Widespread adoption of new sequencing technologies is often hampered in part due to high costs and uncertainty in data quality compared to gold standard assays. 10x Genomics has popularized droplet-based single cell RNA sequencing (scRNA-sequencing), whereby single cells are passed through a microfluidics chip, captured by gel beads, and separated by oil emulsion [1].

However, large scRNA-sequencing studies have been limited due to the high costs associated with cell capture and next generation sequencing. Recently, Ultima Genomics revealed the UG 100, which utilizes "flow-based" sequencing combined with data capture cycles where only one nucleotide at a time is "flowed" through its own unique channel [2]. This flow system reduces errors by preventing base-calling errors as only one nucleotide is present at a time during the sequencing run. Next, high-speed cameras acquire data as sequencing is performed on a flat, rotating silicon wafer [2]. Using a moving wafer allows even reagent dispensing across the entire surface area, which yields higher quality and larger datasets per run. This directly relates to cost savings: the UG 100 10B read wafer provides end users with cost savings over traditional next-generation sequencing by synthesis (SBS) based flow cell chemistry, with significant cost savings per sequencing run.

Here, we compare two different projects captured using either the 10x Chromium Next GEM 3' or 5' technology for their data quality and consistency between the Illumina NovaSeq X Plus and the Ultima UG 100 and report comparable yields and data quality between the two platforms.

MATERIALS AND METHODS

Sequencing

We previously prepared and sequenced two libraries on the Illumina NovaSeq X Plus, the first a 10x 3' Chromium Next GEM v3.1 library of 18 human brain samples and the second a 10x 5' Chromium Next GEM HT v2 library of 9 human PBMCs. The 3' libraries targeted 10k cells/sample and 40k reads/cell and the 5' libraries targeted 5k cells/sample, 20k reads/cell (gene expression, GEX), and 5k reads/cell (BCR-seq; Illumina only). Ultima library conversion success was verified by a read length increase of ~35 bp by BioAnalyzer HS and TapeStation D1000. Both Ultima-converted GEX libraries were sequenced at 600 pM. 5' VDJ-sequencing (BCR and TCR) is currently not supported by the Ultima UG 100 and was omitted from this comparison.

Data Format

10x Genomics' cellranger pipelines require paired end FASTQ files for analysis, but raw data output from the Ultima UG 100 is a single stranded CRAM file. CRAM conversion to a forward strand FASTQ file is performed automatically on the

sequencer (APL 5.3.2.14). A paired FASTQ file is created by generating the reverse complement sequence from the forward strand.

Data Analysis

Converted FASTQ files were passed onto cellranger count v8.0.1 and the filtered feature matrix was loaded into R v4.4.1 for downstream data QC and analysis using the packages tidyverse v2.0.0, ggpubr v0.6.0, and Seurat v5.1.0. The 3′ datasets were filtered by 500 < features < 3,700 and % mitochondrial reads < 5%. 25 dimensions were used to find neighbors and a resolution of 0.6 was used for clustering.

The 5′ datasets were filtered by 300 < features < 2,000 and % mitochondrial reads < 5%. 25 dimensions were used to find neighbors and a resolution of 0.6 was used for clustering. Batch effect correction was performed using the harmony package v1.2.3. Statistics were calculated using either Welch's T-test, Wilcoxon rank sum test, or Spearman's correlation coefficient with p < 0.05 statistically significant.

RESULTS

Ultima library outputs are comparable to Illumina after Cell Ranger analysis

To determine if Ultima UG 100 gave equivalent sequencing depth compared to the Illumina NovaSeq X Plus, we converted 10x 3′ and 5′ Chromium Illumina libraries to Ultima format and sequenced both. Mean total reads were comparable between libraries for the 3′ samples (Table 1, Figure 1), but in contrast, the 5′ Ultima library generated 2.84x the number of reads per sample over Illumina. This increase in read number was concomitant with increased sequencing saturation, total genes per sample, mean reads per cell, and median genes per cell, but not total cells detected (Table 1). The disparity in 5′ reads

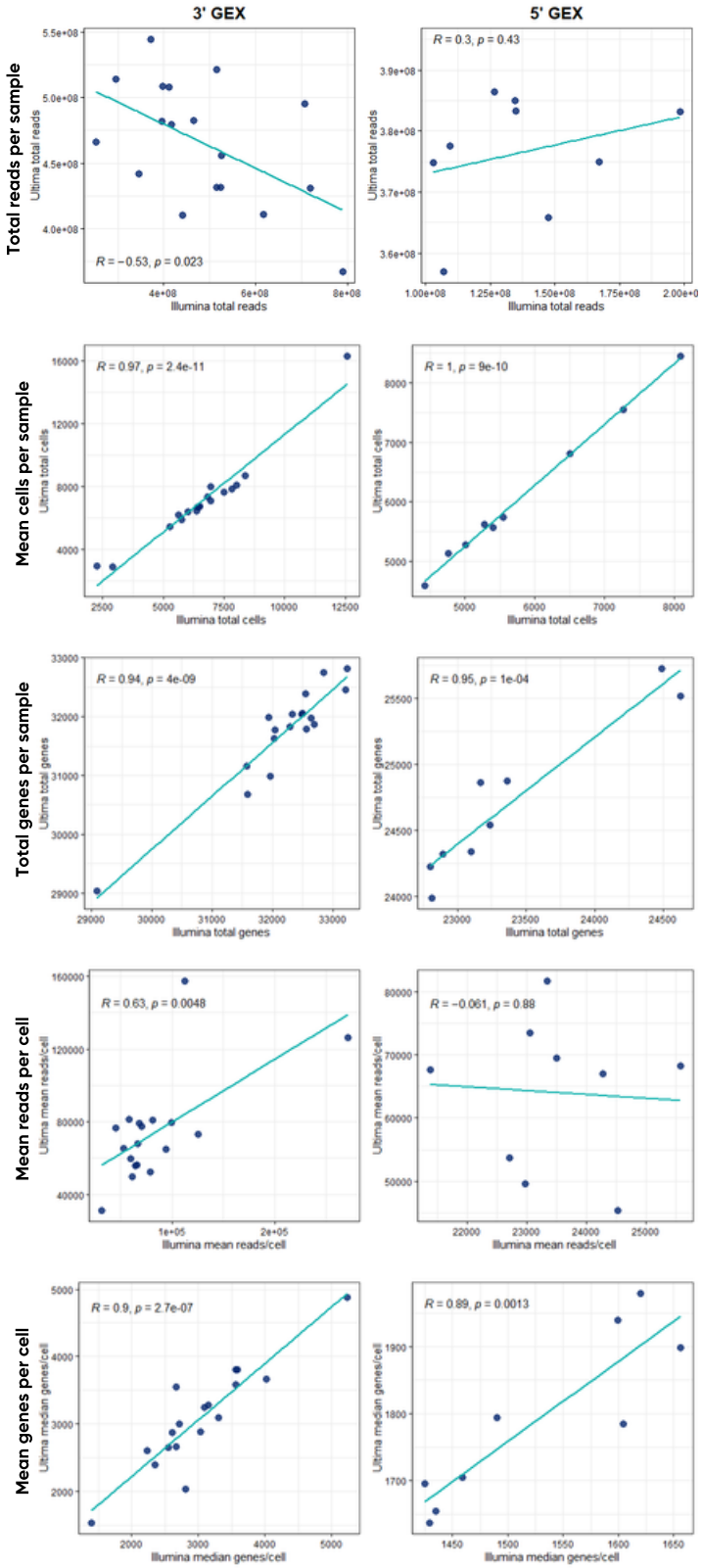


Figure 1: QC metrics are generally correlative between Illumina NovaSeq X Plus and Ultima UG 100. 10x Genomics cellranger QC metrics following count analysis for 3′ GEX (left) and 5′ GEX (right). Individual dots represent a single sample and are provided as an average across all cells for the bottom two rows. R values are Spearman's correlation coefficients.

was intentional, set by a higher target reads per sample, and not reflective of sequencing quality differences between the UG 100 and NovaSeq X Plus.

Table 1: cellranger count output summary.

	10x 3' GEX		10x 5' GEX	
	Illumina	Ultima	Illumina	Ultima
Mean total reads per sample	483,459,385	465,665,371	136,925,019	374,397,230
Estimated mean number of cells	6,581	7,078	5,997	6,381
Mean total genes	32,198	31,375	23,311	24,671
Fraction of reads in cells	85.58%	85.35%	88.79%	88.82%
Mean reads per cell	83,802	74,210	22,950	61,546
Median UMI counts per cell	8,543	9,139	3,313	4,309
Median genes per cell	3,032	3,085	1,464	1,714
Valid barcodes	94.89%	100%	88.43%	100%
Valid UMIs	99.87%	99.92%	99.13%	100%
Sequencing saturation	61.93%	51.37%	67.98%	85.74%
Reads mapped to genome	96.04%	97.04%	91.62%	97.66%
Reads mapped confidently to genome	88.82%	89.02%	74.35%	80.75%
Reads mapped confidently to intergenic regions	6.12%	6.91%	4.88%	4.45%
Reads mapped confidently to intronic regions	62.6%	62.38%	8.65%	9.24%
Reads mapped confidently to exonic regions	20.09%	19.75%	60.83%	67.07%
Reads mapped confidently to transcriptome	67.62%	69.55%	62.03%	70.49%
Reads mapped antisense to gene	14.1%	11.52%	6.92%	5.18%

Q scores do not predict Ultima sequencing quality

Sequencing-by-synthesis platforms such as Illumina report a quality (Q) score, which is a prediction of the probability of an error in base calling. The most commonly reported metric is the Q30 score, where 1 base in 1,000 is predicted to be incorrect [3]. In contrast, Ultima's flow-based sequencing effectively eliminates the possibility a base could be incorporated incorrectly, as only one nucleotide is flowed through the sequencing reaction at a time [2]. As Q scores are a gold-standard QC metric for next generation sequencing, we asked if these values had any predictive value for Ultima sequencing, despite the chemistry differences.

Analysis of the CRAM to FASTQ converted files revealed a decrease in percent Q20 scores and a sharp decrease for percent Q30 in both 3' and 5' GEX assays for Ultima relative to Illumina samples (Table 2, Figure 2). This was reflected in the cellranger count outputs, as the Q30 bases in RNA read Ultima runs showed a large decrease compared to the Illumina while Q30 bases in barcode was unaffected and Q30 bases in UMI decreased in the 3' GEX only (Table 2).

Table 2: Single cell sequencing Q score statistics.

	10x 3' GEX		**10x 5' GEX**	
	Illumina	Ultima	Illumina	Ultima
Reads percent Q20	93.25%	92.35%	96.79%	92.71%
Reads percent Q30	95.9%	80.15%	92.68%	79.16%
Q30 bases in barcode	95.8%	91.69%	94.35%	92.88%
Q30 bases in RNA read	95.82%	80.18%	92.04%	79.21%
Q30 bases in UMI	96.81%	76.91%	95.44%	98.81%

To determine if Q scores had any predictive value for cellranger outputs, we determined the Spearman correlation coefficient between the percent Q20 and Q30 with each cellranger count metric. Illumina library Q scores had no predictive power for Mean Reads per Cell, Median UMI Counts, or Median Genes per Cell, but these metrics were positively and negatively correlated for Ultima 5' and 3' libraries, respectively. Overall, these data indicate that although Q scores are lower in Ultima libraries, they do not have any consequences on data quality and low values can be safely ignored.

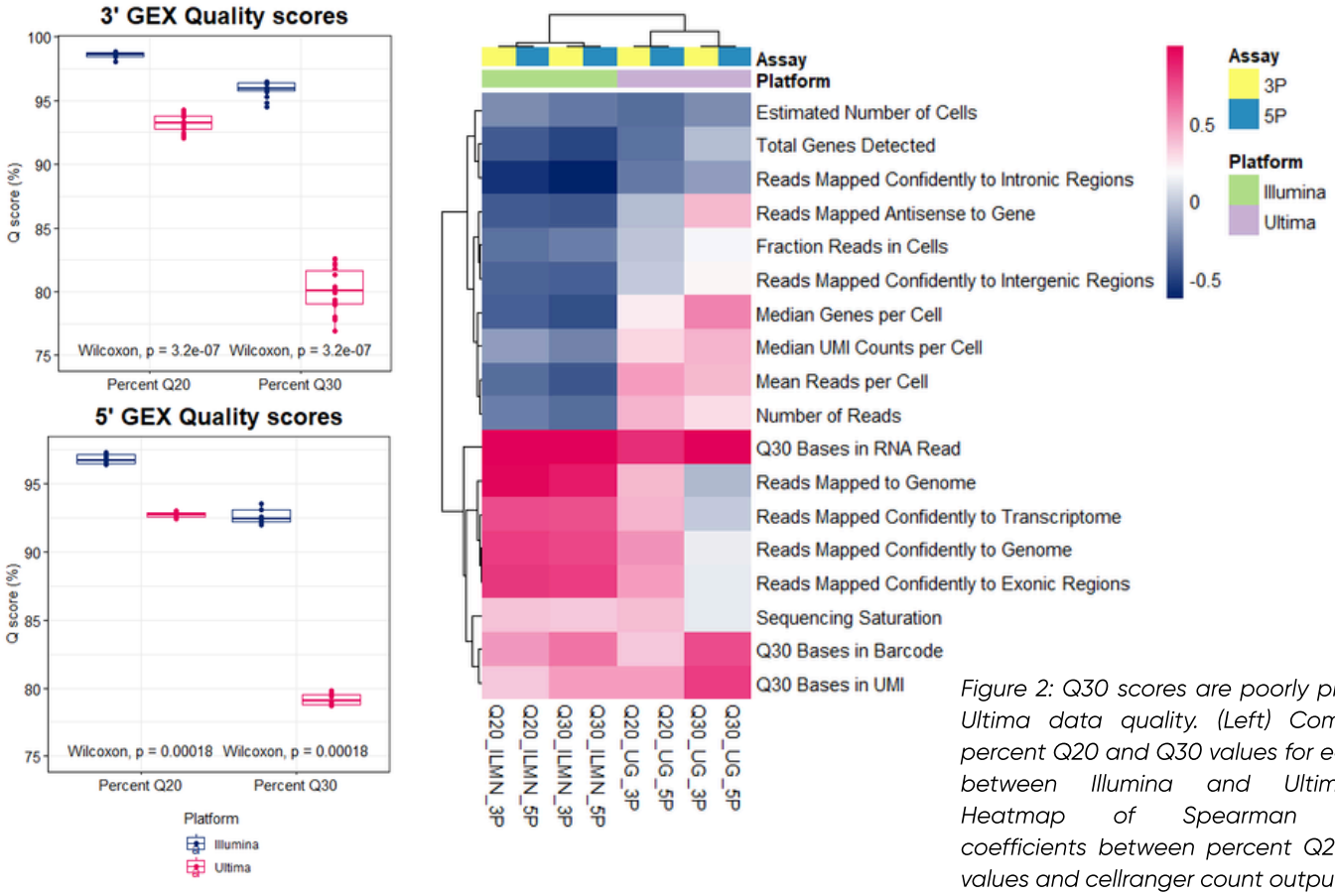


Figure 2: Q30 scores are poorly predictive for Ultima data quality. (Left) Comparison of percent Q20 and Q30 values for each sample between Illumina and Ultima. (Right) Heatmap of Spearman correlation coefficients between percent Q20 and Q30 values and cellranger count output metrics.

10x 3′ GEX cell clustering is consistent across sequencing platforms

After demonstrating the QC statistics were consistent between platforms, we then asked if gene expression data was also consistent. The 3′ dataset consisted of 36 samples from 2 libraries, with a total size of 33,682 genes and 144,392 single cells. The number of genes, UMIs, and percentage of mitochondrial and ribosomal reads were highly consistent between sequencing platforms (Table 3).

Table 3: 10x 3′ dataset per cell summaries.

Metric	Platform	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
UMIs	Illumina	499	3,170	7,126	14,625	17,916	271,322
	Ultima	499	3,122	7,857	16,298	19,583	322,827
Genes	Illumina	111	1,754	2,865	3,547	4,937	13,184
	Ultima	135	1,699	2,987	3,597	4,973	13,616
% Mitochondrial Reads	Illumina	0	0.03	0.088	0.307	0.247	56.81
	Ultima	0	0.033	0.094	0.351	0.271	63.67
% Ribosomal Reads	Illumina	0	0.259	0.366	0.469	0.553	14.341
	Ultima	0	0.273	0.387	0.51	0.593	14.437

We then performed normalization and cell clustering without batch correction and proceeded to visualize cellular transcriptomic relationships via UMAP, which identified a total of 37 clusters within the 3′ dataset (Figure 3). While certain individual clusters were specific for given samples (e.g. Clusters 3 and 4 both in S8), we did not find any significant bias or deviation in cell cluster composition by sequencing platform (Figure 4), with no more than 10% difference in contribution by platform. This finding meant that sequencing on the Ultima UG 100 did not create any cellular artifacts relative to the Illumina NovaSeq X Plus for 10x Genomics 3′ GEX.

Figure 3: Illumina and Ultima UMAP plots for 3′ GEX. UMAP projections of cell clusters grouped by (A) sequencing platform, navy Illumina and pink Ultima, (B) individual cluster, (C) pooled (Illumina + Ultima) sample IDs, and (D) cell types.

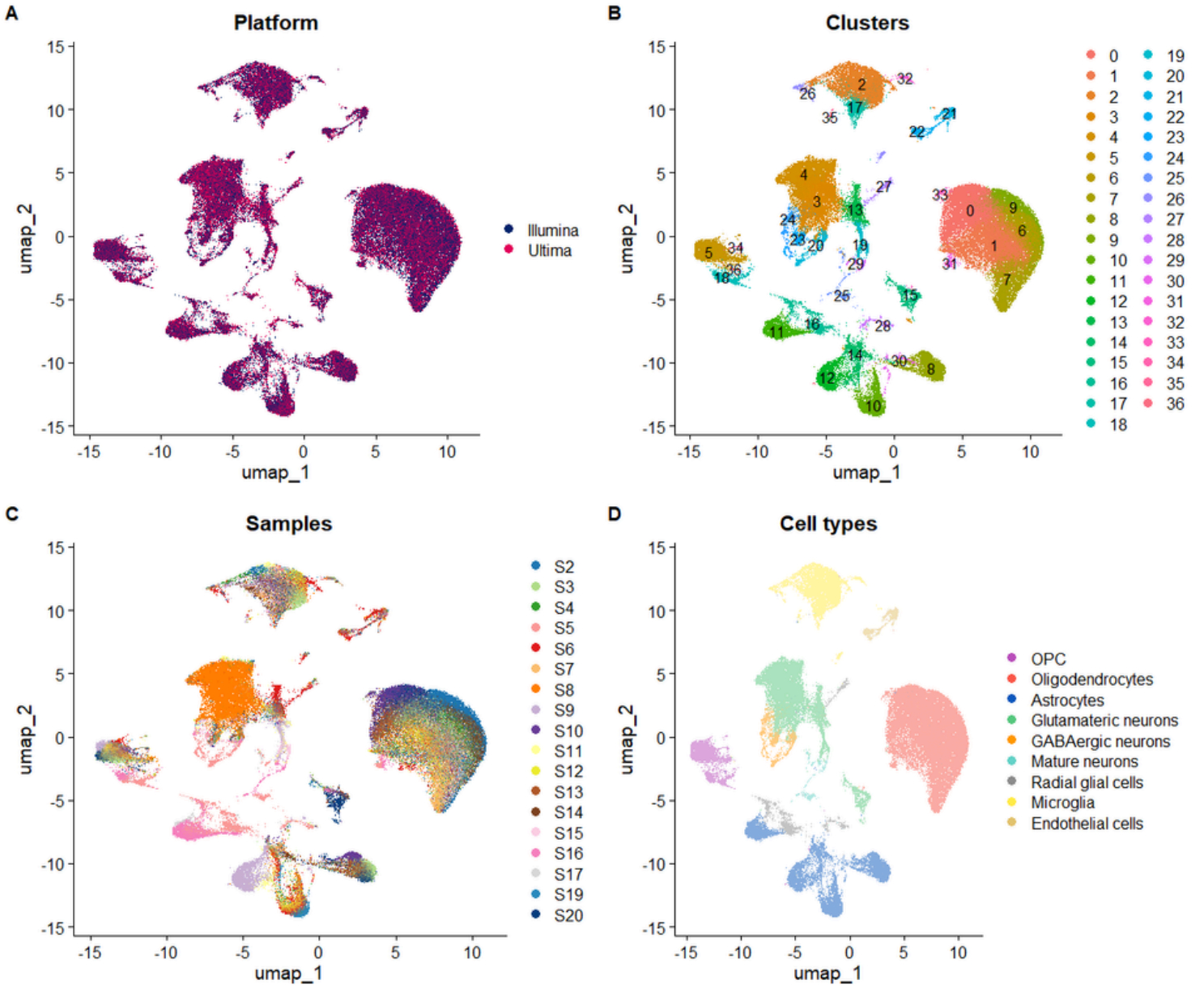
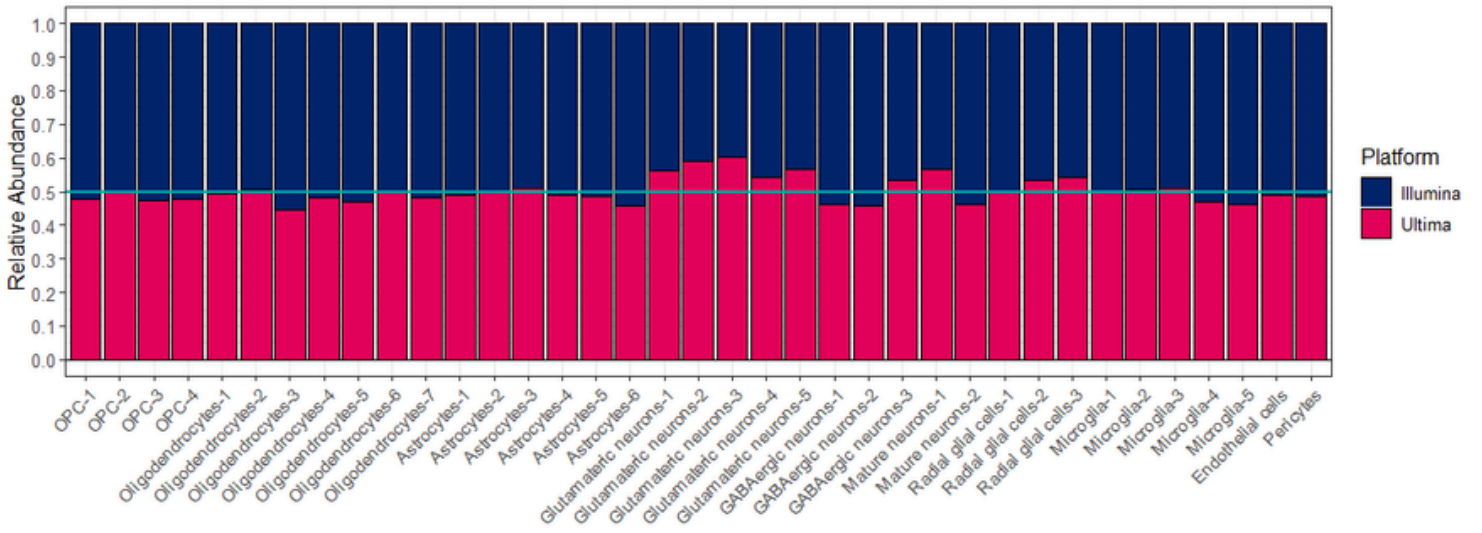


Figure 4: Platform does not bias cell cluster composition in 3′ GEX. Relative contribution by sequencing platform to each cell cluster. Teal line indicates a theoretical 50/50 split in cell counts, blue bars Illumina and pink bars Ultima sequenced cells.



10x 5′ GEX cell clustering is largely consistent across sequencing platforms after batch correction

The 5′ GEX dataset consisted of 18 samples from 2 libraries, with a total size of 24,603 genes and 52,407 single cells (Table 4). Cell clustering by UMAP revealed clear differences within specific clusters by sequencing platform (Figure 5A).

Table 4: 10x 5′ dataset per cell summaries.

Metric	Platform	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
UMIs	Illumina	499	2,386	3,497	3,968	4,825	56,914
	Ultima	499	3,141	4,533	5,198	6,376	73,852
Genes	Illumina	19	1,184	1,520	1,595	1,895	6,997
	Ultima	2	1,394	1,776	1,877	2,269	7,627
% Mitochondrial Reads	Illumina	0	2.78	3.62	5.18	4.96	99.1
	Ultima	0	2.84	3.67	5.53	5.02	98.69
% Ribosomal Reads	Illumina	0	13.74	20.06	20.26	27.5	58.8
	Ultima	0	11.15	18.28	18.87	26.92	57.16

Figure 5: Illumina and Ultima UMAP plots for 5' GEX. UMAP projections of cell clusters grouped by (A) sequencing platform pre-harmony batch correction, navy Illumina and pink Ultima, (B) sequencing platform post-hamony batch correction, (C) sample IDs, and (D) number of clusters.

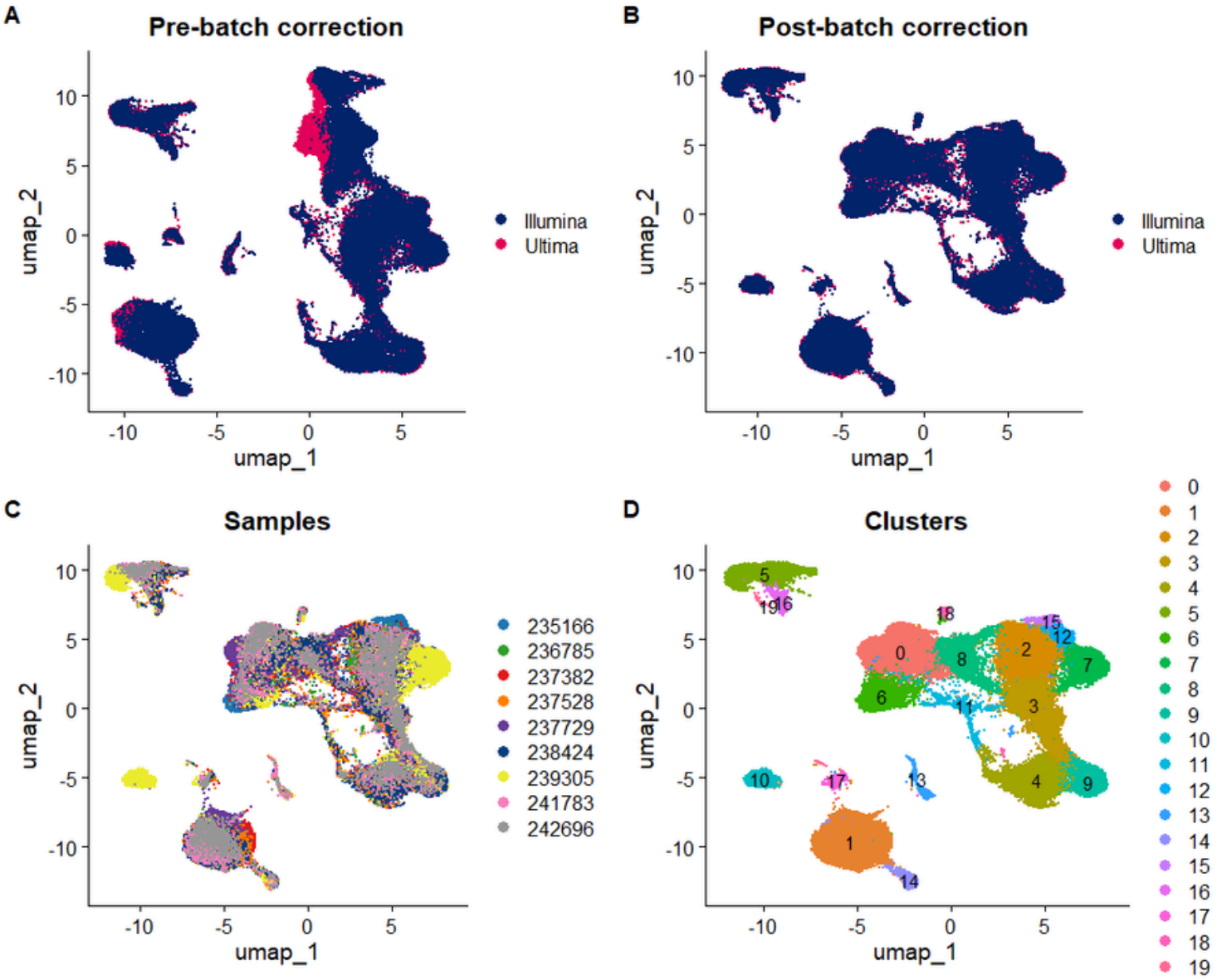
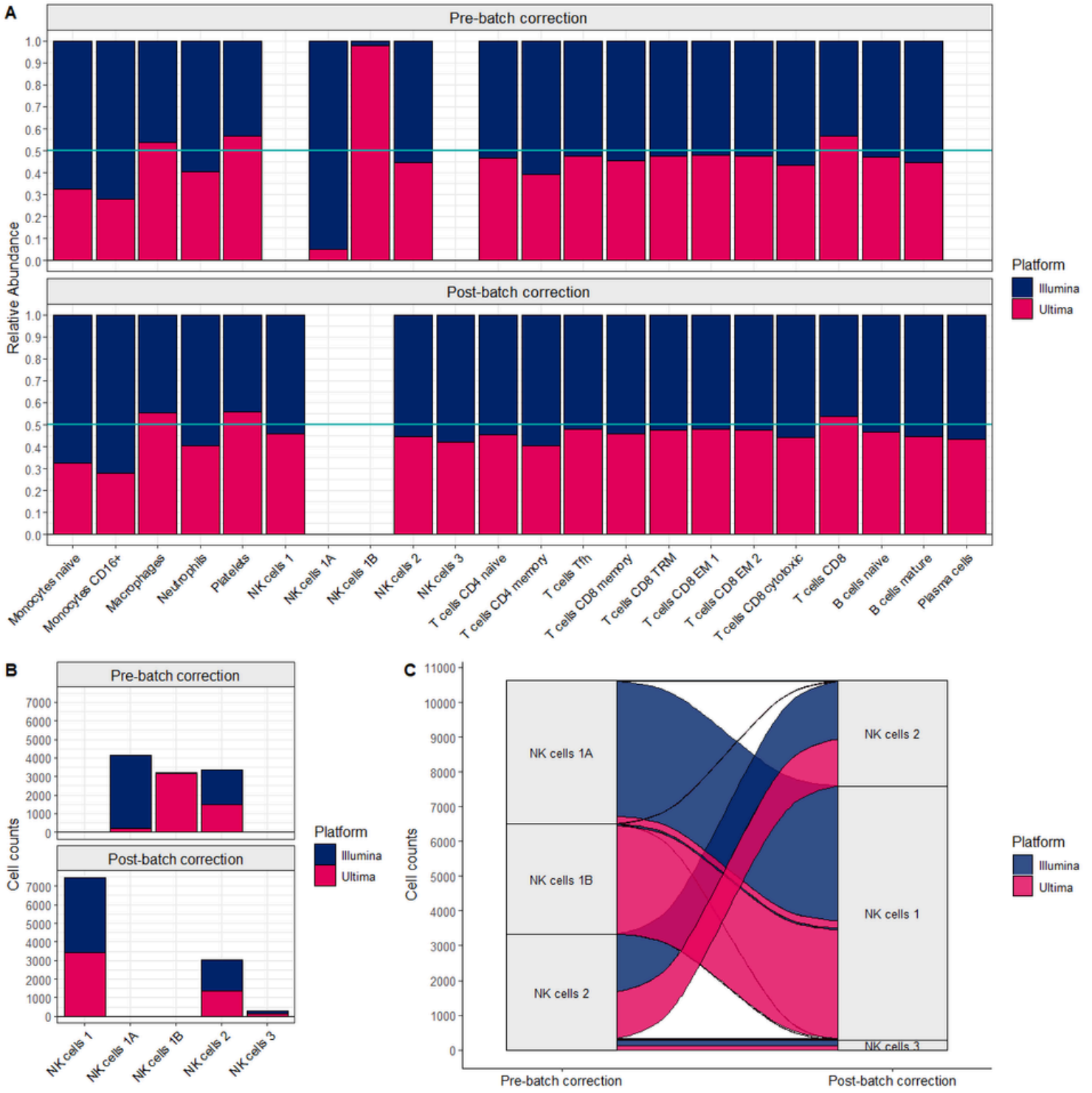


Figure 6: Platform bias in 5' GEX NK cell populations is removed by batch correction. (A) Relative contribution by sequencing platform to each cell cluster. Cell clusters shown pre- (top) or post-batch effect correction (bottom). Teal line indicates a theoretical 50/50 split in cell counts, blue bars Illumina and pink bars Ultima sequenced cells. Empty spaces represent a cell cluster not identified either pre- or post-batch correction. (B) Cell counts per NK cell cluster, separated pre- (top) or post-batch correction (bottom). (C) NK cells were matched by cell barcode ID, with clusters pre- (left) or post-batch correction (right). Each line indicates an individual cell.



We performed batch correction on the datasets using the harmony package and re-clustered cells, which abrogated the platform-specific clustering (Figures 5B-5D). This was further verified by comparing the relative contribution of each cluster, which aside from two issues noted below, averaged 3.1% deviation between sequencing platforms (Figure 6A).

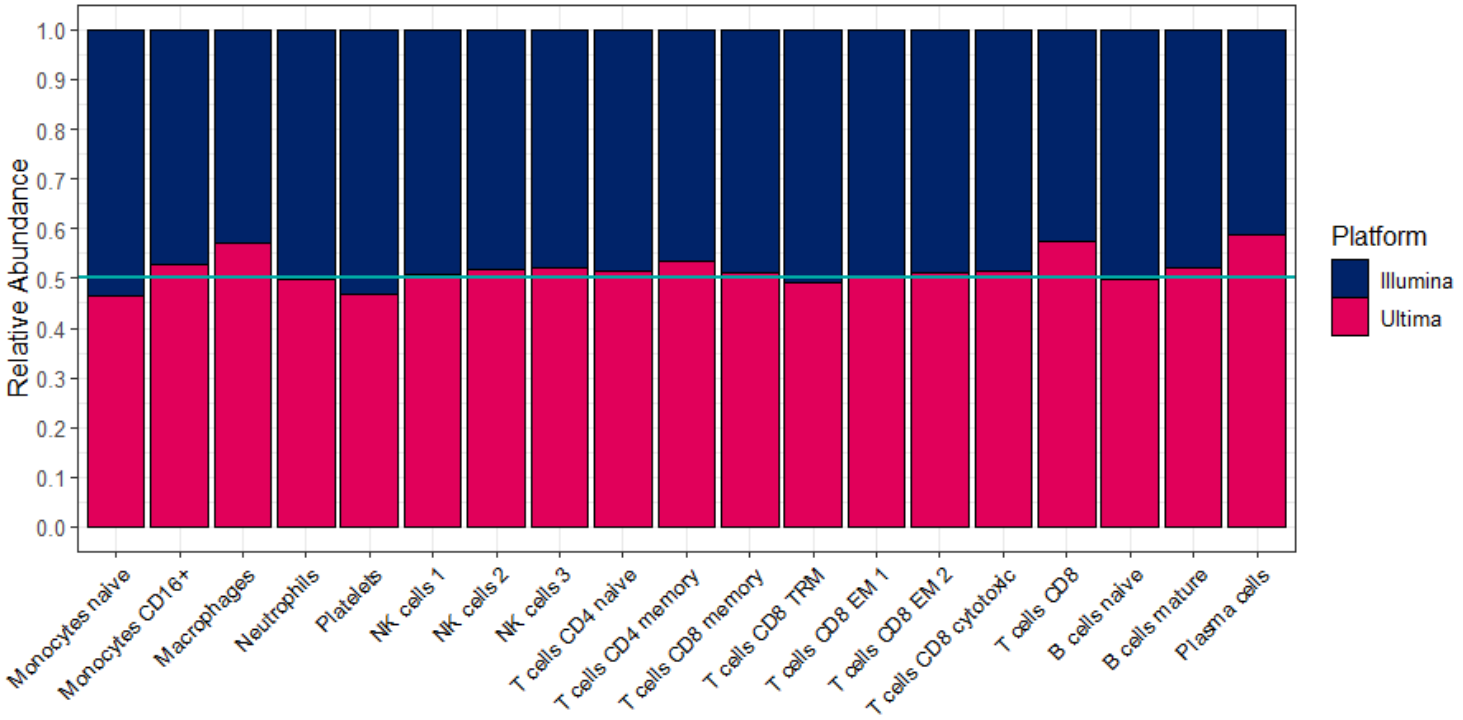
Pre-batch correction identified two major biases in this dataset: (1) approximately 70% of monocytes came from Illumina samples and (2) NK cells were split into three clusters, one with similar contribution from both platforms and two with 96% of cells sequenced on either platform (Figures 6A, 6B). Following batch correction, the Illumina bias was still observed in monocytic populations, while the two aberrant NK clusters merged together. We confirmed this by comparing the cell barcodes from the NK cell clusters pre- and post-batch correction (Figure 6C), demonstrating the cell clusters were integrated.

We then compared the monocytes pre- and post-filtering, which revealed a majority of cells filtered out were sequenced on the UG 100 (Table 5). Total UMIs and genes per cell were significantly enriched in Ultima-sequenced cells (Table 1), and as both 5′ datasets were filtered by the same thresholds (% mitochondrial reads < 5% and 300 < nFeature_RNA < 2000), this skewed the cells towards Illumina enrichment. We matched the percentiles for 300 and 2,000 features from Illumina (3.76% and 78.78%, respectively) to Ultima and repeated cell filtering at those thresholds (% mitochondrial reads < 5% and 335 < nFeature_RNA < 2411). This abrogated the skew previously observed in monocytes and also neutrophil clusters (Figure 7), with a 50.4/49.6% Illumina/Ultima split.

Table 5: 10x 5′ monocyte cell counts pre- and post-filtering.

	Platform	**Filtered out**	**Retained**		
			Monocytes naive	**Monocytes CD16+**	**Other**
Monocytes naive 1	Illumina	5,075	4,847	44	3
	Ultima	7,981	2,348	15	6
Monocytes naive 2	Illumina	623	45	0	0
	Ultima	654	20	0	1
Monocytes CD16+ 1	Illumina	2,421	17	362	0
	Ultima	2,672	5	147	0
Monocytes CD16+ 2	Illumina	354	0	44	0
	Ultima	381	0	12	0
Other	Illumina	N/A	19	7	N/A
	Ultima	N/A	23	2	N/A

Figure 7: Platform bias in 5' GEX monocytes is removed by altering cell filtration. Relative contribution by sequencing platform to each cell cluster. Both platforms were filtered for % mitochondrial reads < 5, while features per cell was variable: 300 < Illumina < 2,000 and 335 < Ultima < 2,411. Teal line indicates a theoretical 50/50 split in cell counts, blue bars Illumina and pink bars Ultima sequenced cells.



These revised analyses proved that the sequencer-specific biases initially observed were due to batch effects and filter settings. Overall, we find no difference in yield or data quality between scRNA-sequencing libraries on the Illumina NovaSeq X Plus and the Ultima UG 100.

CONCLUSIONS

Cost-effective solutions are critical for emerging next-generation sequencing technologies. Here, we compared the use of the "flow-based" sequencing of the Ultima UG 100 against the gold standard Illumina NovaSeq X Plus for efficacy using the 10x Genomics Chromium Next GEM 5' and 3' scRNA-sequencing kits as test subjects. We found that 10x Genomics libraries, originally prepared for Illumina sequencing, could be adapted for the UG 100 and returned a comparable number of raw reads compared to the Illumina. Traditionally used metrics for sequencing quality such as Q30 scores had no predictive value for Ultima outputs.

Although the 3' dataset showed minimal variability or bias, the 5' dataset required both batch correction and altered filtering settings to correct bias in NK cells and monocytes, respectively. Collectively, these data show that the Ultima UG 100 performs comparably to the Illumina NovaSeq X Plus for scRNA-sequencing, but at much lower cost.

LITERATURE CITED

1. Ortolano N. 2024. The neXt generation of single cell RNA-seq: an introduction to GEM-X technology. https://www.10xgenomics.com/blog/the-next-generation-of-single-cell-rna-seq-an-introduction-to-gem-x-technology. Accessed August 11, 2025.
2. Almogy G, M Pratt, F Oberstrass, L Lee, D Mazur, N Beckett, O Barad, I Soifer, E Perelman, Y Etzioni, M Sosa, A Jung, T Clark, E Trepagnier, G Lithwick-Yanai, S Pollock, G Hornung, M Levy, M Coole, T Howd, M Shand, Y Farjoun, J Emery, G Hall, S Lee, T Sato, R Magner, S Low, A Bernier, B Gandi, J Stohlman, N Lennon, S Gabriel, D Lipson. 2022. Cost-efficient whole genome-sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics platform. bioRxiv 2022.05.29.493900.
3. Ewing B, P Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research 8:186–194.