Supplementary Figures

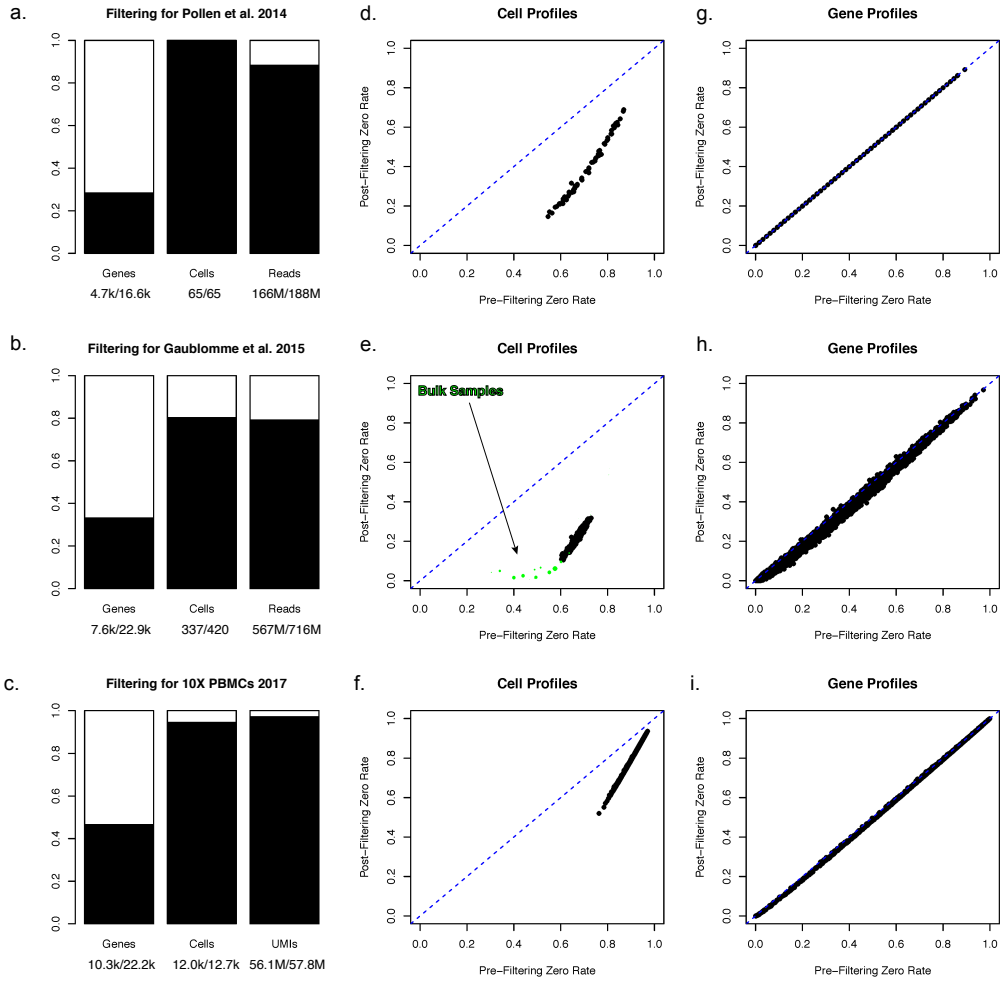

Figure S1 Related to Figure 1 and "Regression-based normalization" Section of STAR Methods. Data filtering for three scRNA-seq datasets (Pollen et al., 2014; Gaublomme et al., 2015; Zheng et al., 2017). **(a-c)** Boxplot representing (in black) proportions of genes, cells and counts preserved after sample and gene filtering. Note in (a) that no samples were filtered in Pollen et al. 2014. **(d-f)** Per-cell zero rates before and after data filtering. Datasets are presented in the same order as (a-c). In the case of Gaublomme et al. 2015, bulk RNA-seq samples from the same study were plotted in green along single-cell samples in (e). Point size corresponds to two-sided p-value of sample read count under log-normal model fit to single-cell samples passing filter. This is meant to highlight bulk samples with similar read coverage: other samples had very poor coverage and correspondingly high zero rates. **(g-i)** Per-gene zero rates before and after data filtering. Datasets are presented in the same order as (a-c).

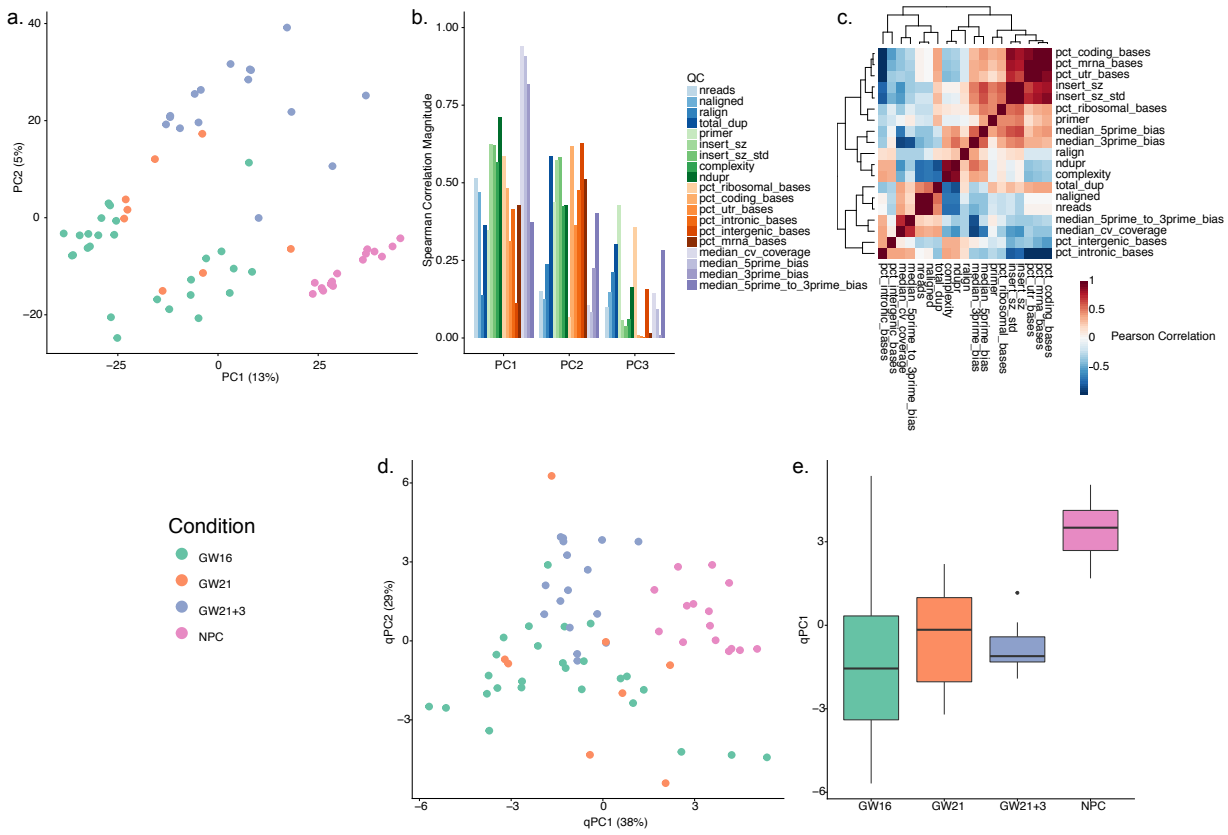

Figure S2 Related to Figure 1. Exploratory data analysis of human cortex cells (Pollen et al., 2014). **(a)** PCA of the log-transformed, TC-normalized read count data using all genes passing quality filtering. Cells are color-coded by biological condition. Cells cluster partially by biological condition, with significant intra-condition heterogeneity. The design of this study is fully confounded (one batch per biological condition): batch adjustment is not advisable in this case, as it would remove the biological effects of interest. **(b)** Absolute Spearman correlation coefficient between the first three PCs of the expression data (as computed in (a)) and a set of QC measures (Table S1). **(c)** Heatmap of pairwise Pearson correlation coefficients between QC measures. **(d)** PCA of the QC measures for all cells in (a). Single-cell QC profiles cluster by biological condition, suggestive of technical confounding. **(e)** Boxplot of the first qPC, stratified by biological condition. QC measures differ significantly between NPCs and other biological conditions / batches.

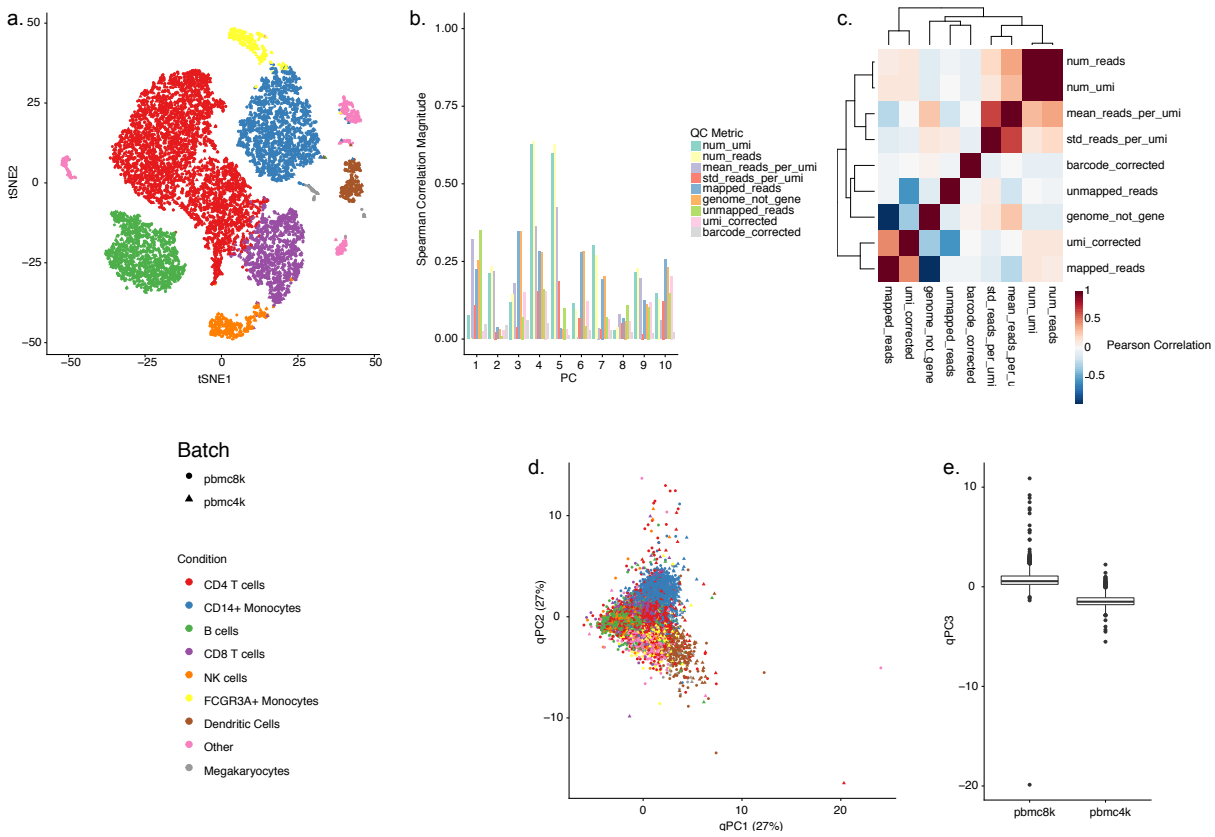

Figure S3 Related to Figure 1. Exploratory data analysis of human peripheral blood mononuclear cells (PBMCs) (Zheng et al., 2017). **(a)** tSNE of the first 10 PCs of the log-transformed, TC-normalized UMI count data for all genes and cells passing quality filtering. Cells are color-coded by a *Seurat*-based manual annotation of major PBMC subtypes; shape represents the 10x batch. Samples from both batches ("pmbc4k" and a larger "pbmc8k") originated from the same "healthy" human donor. Cells clearly cluster by data-derived biological condition, one consequence of being clustered jointly in *Seurat*. **(b)** Absolute Spearman correlation coefficient between the first ten PCs of the expression data (as computed in (a)) and a set of QC measures (Table S2). **(c)** Heatmap of pairwise Pearson correlation coefficients between QC measures. **(d)** PCA of the QC measures for all cells in (a). Single-cell QC profiles partially cluster by data-derived biology (especially CD14+ monocytes), with no clear clustering by batch. **(e)** boxplot of the third qPC, stratified by batch. The third qPC is the qPC with the highest correlation with batch.

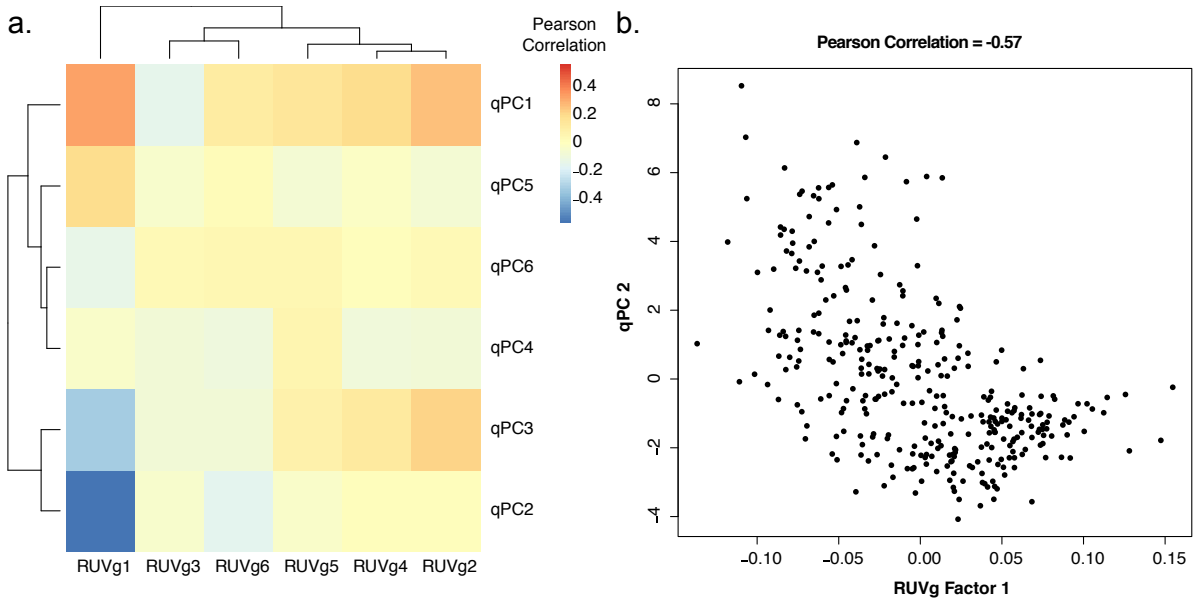

Figure S4 Related to Figure 2. Factors of unwanted variation in the Gaublomme et al. dataset (Gaublomme et al., 2015). **(a)** Heatmap of Pearson correlation coefficients between RUVg-derived factors of unwanted variation (Risso et al., 2014) and qPCs. Row and column clustering is generated from the R *hclust* function with default parameters. **(b)** Scatter plot of one anti-correlated pair of RUVg factor and qPC, selected based on their high correlation magnitude displayed in (a).

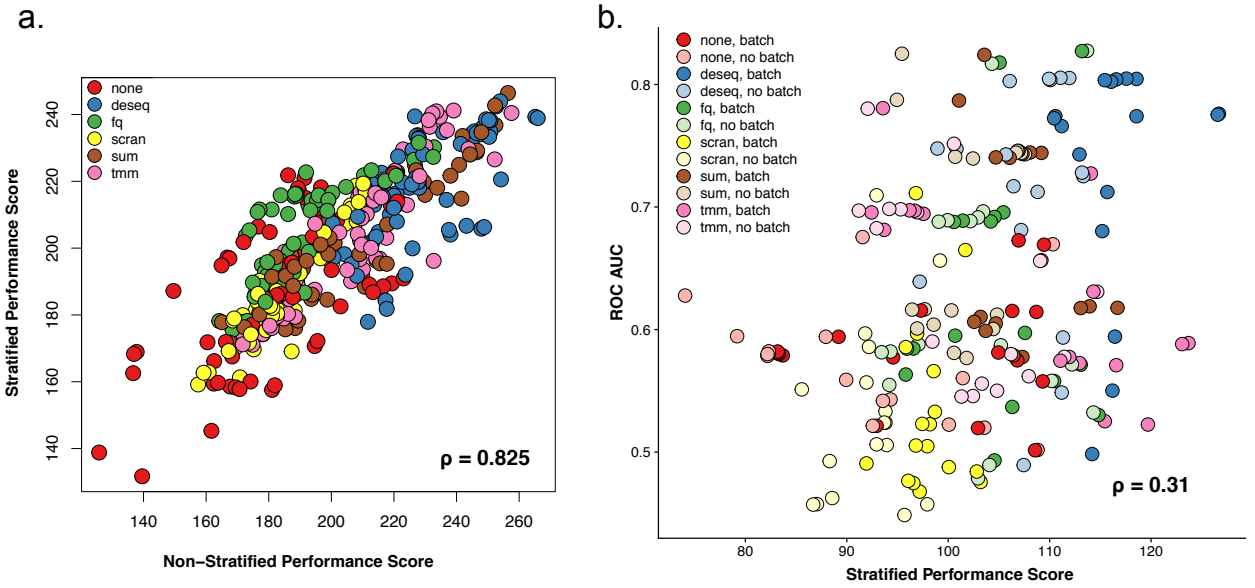

Figure S5 Related to Figure 2 and Figure 4. Stratified performance evaluation for human peripheral blood mononuclear cells (PBMCs) (Zheng et al., 2017). **(a)** Scatter plot representing *scone* performance scores for normalization procedures applied to the 10x PBMC data set. The x-axis measures the aggregation over non-stratified (default) performance scores. The y-axis measures the aggregation over scores computed via stratification by both batch and bio condition. Procedures are colored according to scaling method. Pearson correlation is denoted in the bottom-right corner of the plot. **(b)** ROC AUC vs. stratified *scone* performance score, as in 4c. Normalization procedures in the top-right corner are deemed best both by *scone* and by independent differential expression (DE) validation. Procedures are colored according to scaling method and batch adjustment.

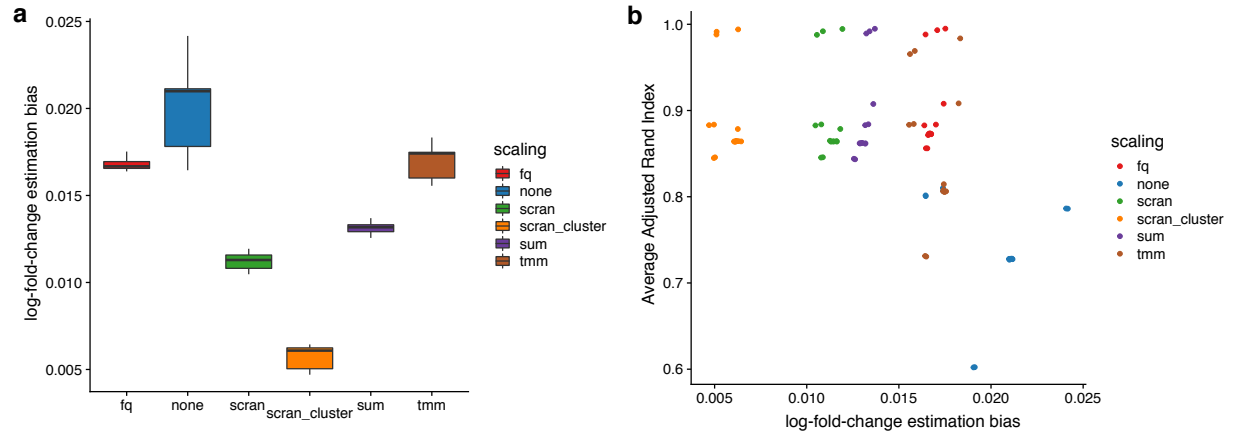

Figure S6 Related to Figure 5. Fold-change estimates of the simulated datasets. **(a)** Boxplot of the average log-fold-change estimation bias across 10 *splatter* simulations (see Methods). All methods led to small bias, with the scran method A. Lun et al., 2016, especially pooling samples after clustering, performing best. **(b)** Scatterplot of average ARI versus the average log-fold-change estimation bias. The absence of strong correlation between these two measures suggests that the bias in log-fold-change estimation is not enough to predict the impact of normalization methods on subsequent analyses.

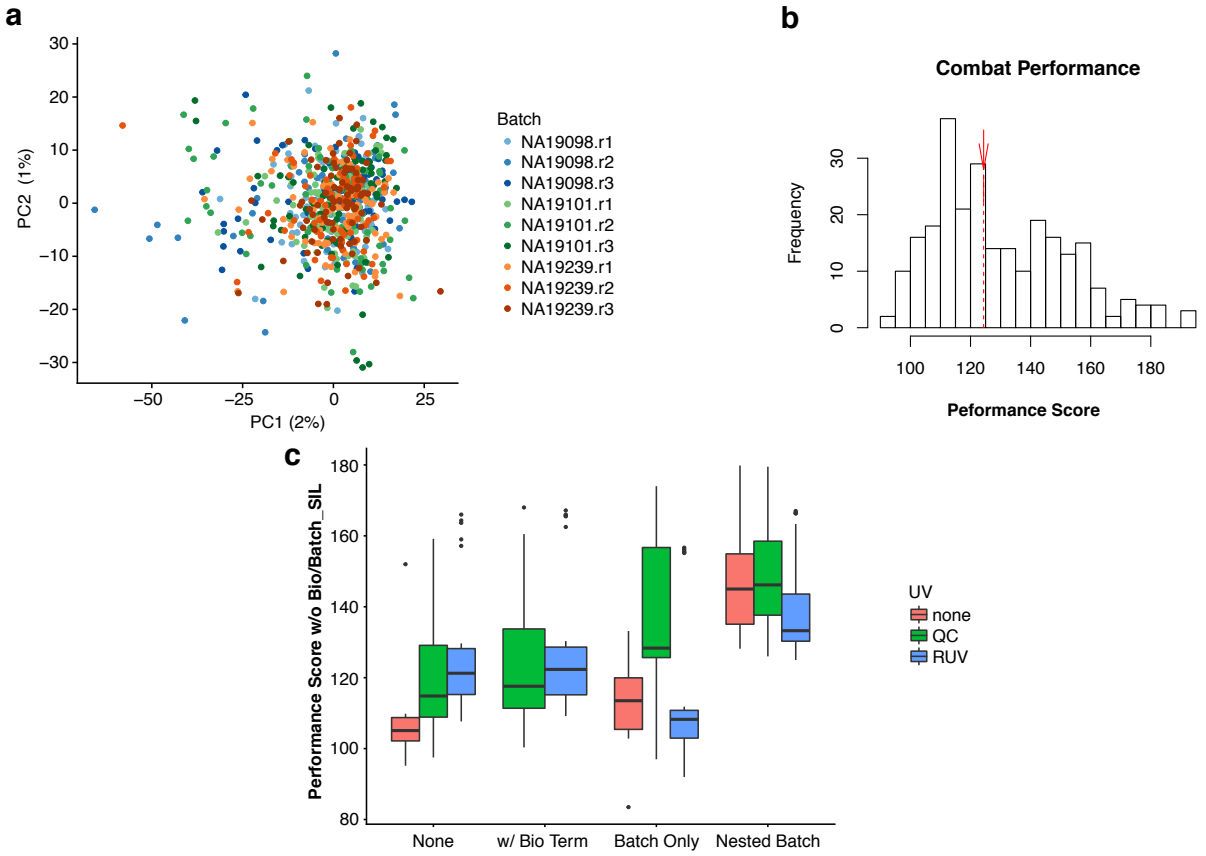

Figure S7 Related to Figure 6. Batch adjustment for the Tung et al. dataset (Tung et al., 2017). **(a)** PCA of ComBat (Leek et al., 2012) normalized data. Donor-specific effects are removed. **(b)** Histogram of *scone* performance scores recomputed to include ComBat (red arrow). **(c)** Boxplot of *scone* performance scores for various normalization procedures, excluding BIO_SIL and BATCH_SIL from the performance score calculation (see Methods).

Supplementary Tables

Table S1 Related to Figure 1. Sample-level quality control (QC) measures (non-10x Genomics).

Name	Description	Source
NREADS	Total number of sequenced reads	Picard
NALIGNED	Total number of aligned reads	Picard
RALIGN	Percentage of mapped reads	Picard
TOTAL_DUP	Number of duplicate reads	FastQC
PRIMER	Percentage of primer sequence reads	FastQC
INSERT_SZ	Average insert size	Picard
INSERT_SZ_STD	Insert size variance	Picard
COMPLEXITY	Sequence Complexity	Picard
NDUPR	Percentage of unique reads	Picard
PCT_RIBOSOMAL_BASES	Percentage of ribosomal bases	Picard
PCT_CODING_BASES	Percentage of coding bases	Picard
PCT_UTR_BASES	Percentage of UTR bases	Picard
PCT_INTRONIC_BASES	Percentage of intronic bases	Picard
PCT_INTERGENIC_BASES	Percentage of intergenic bases	Picard
PCT_MRNA_BASES	Percentage of mRNA bases	Picard
MEDIAN_CV_COVERAGE	Median coefficient of variation of coverage	Picard
MEDIAN_5PRIME_BIAS	Mean 5' coverage bias	Picard
MEDIAN_3PRIME_BIAS	Mean 3' coverage bias	Picard
MEDIAN_5PRIME_TO_3PRIME_BIAS	Mean 5' to 3' coverage bias	Picard

Table S2 Related to Figure 1. Sample-level quality control (QC) measures (10x Genomics).

Name	Description	Source
num_umi	Number of unique UMI sequences	Cell Ranger
num_reads	Total number of reads (regardless of mapping)	Cell Ranger
mean_reads_per_umi	The average number of reads supporting each UMI	Cell Ranger
std_reads_per_umi	Standard deviation of the number of reads supporting each UMI	Cell Ranger
mapped_reads	Proportion of reads which confidently mapped to a gene	Cell Ranger
genome_not_gene	Proportion of reads mapping to the genome, but not to a gene	Cell Ranger
unmapped_reads	Proportion of reads which did not align	Cell Ranger
umi_corrected	Proportion of reads whose UMI sequence was corrected by Cell Ranger	Cell Ranger
barcode_corrected	Proportion of reads whose barcode sequence was corrected by Cell Ranger	Cell Ranger