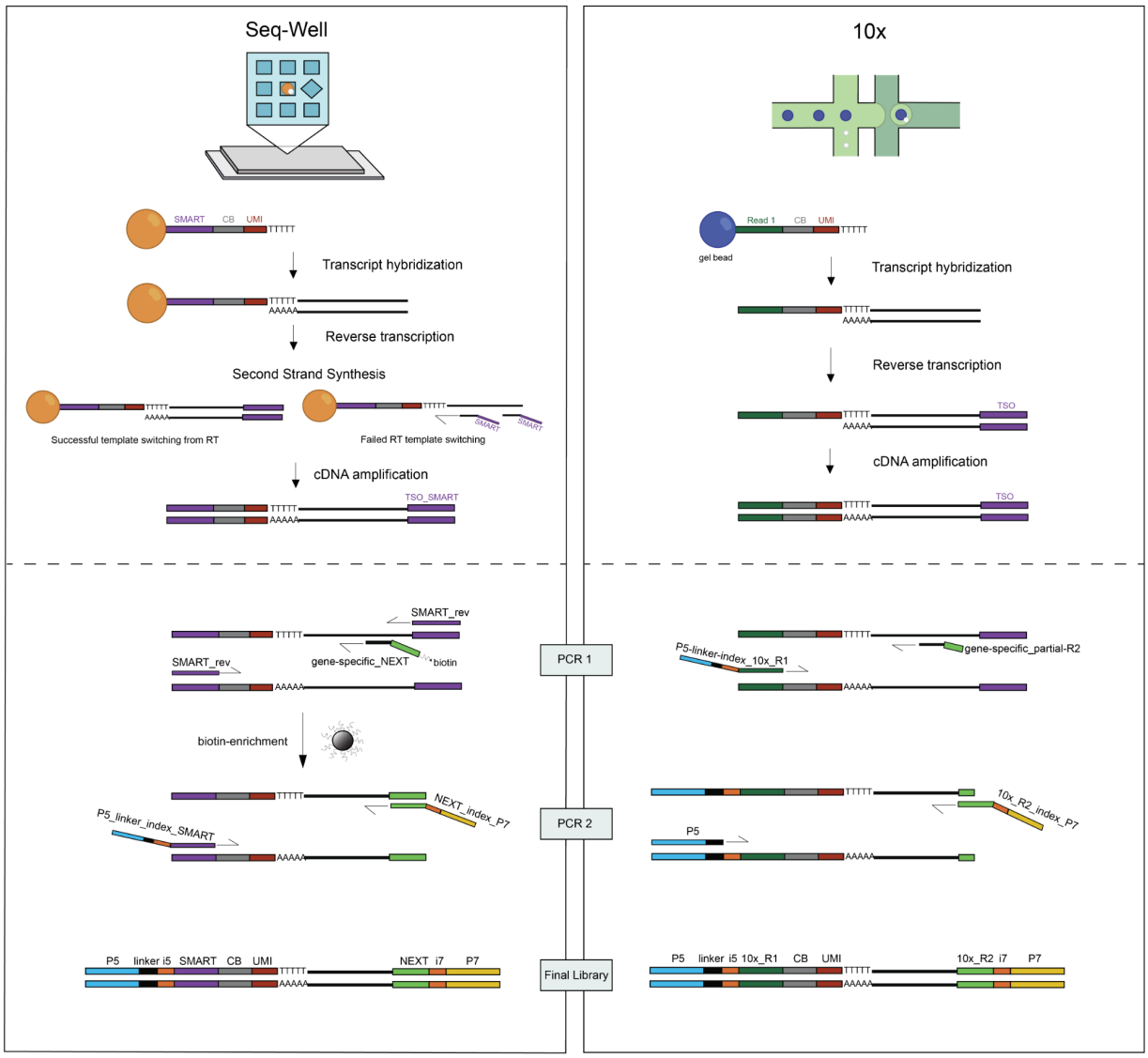
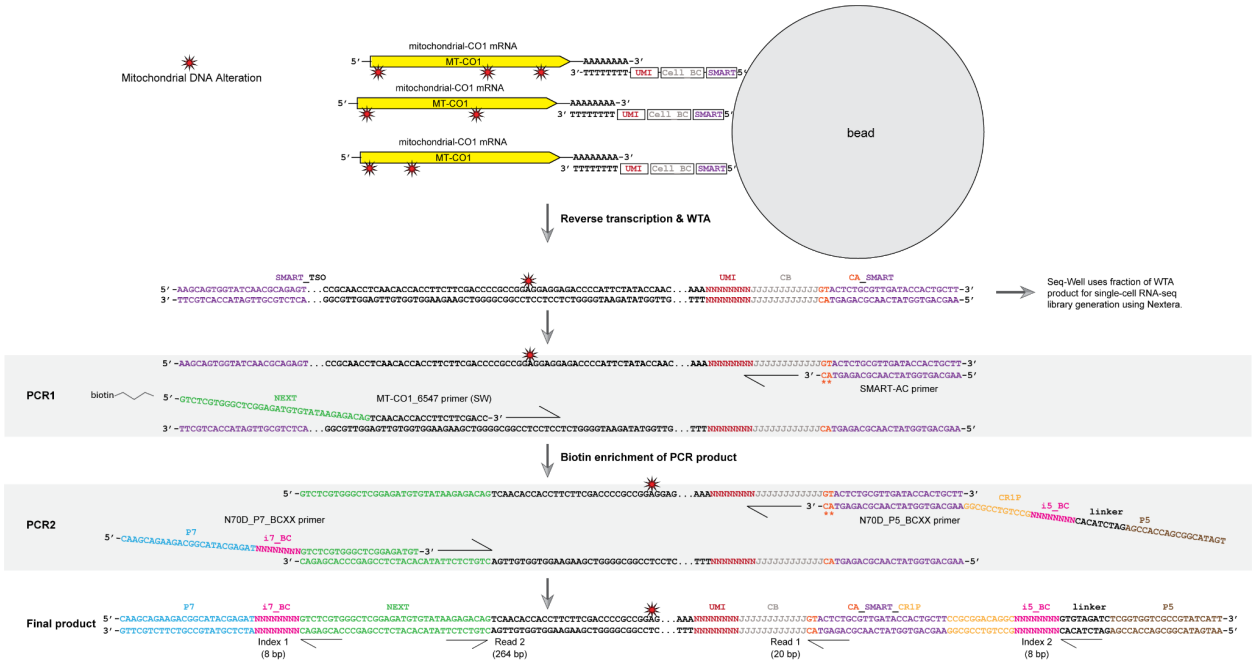
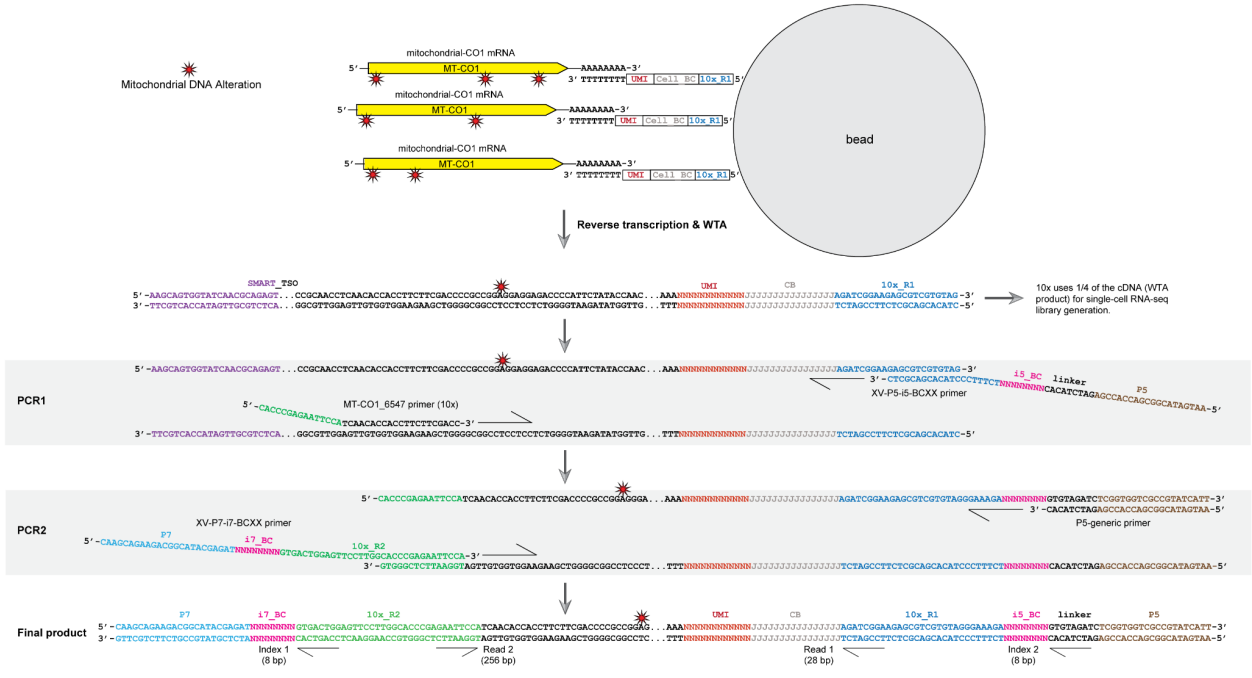


Supplementary Figure 1. Overview of MAESTER methods for Seq-Well or 10x scRNA-seq. Overview and comparison between the MAESTER method for Seq-Well S^3 or Drop-Seq scRNA-seq (left) and 10x procedures (right). Above the dotted line is the standard protocol to create the whole transcriptome amplified material (WTA) or cDNA library. Below the dotted line is the mitochondrial enrichment method from the WTA. Note: Seq-Well or Drop-Seq methods use SMART primer sequences on both ends of the cDNA transcripts, necessitating a biotin enrichment step. 10x uses different primers at each end, allowing for an asymmetric PCR without the need for biotin enrichment. CB: cell barcode, UMI: unique molecular identifier.

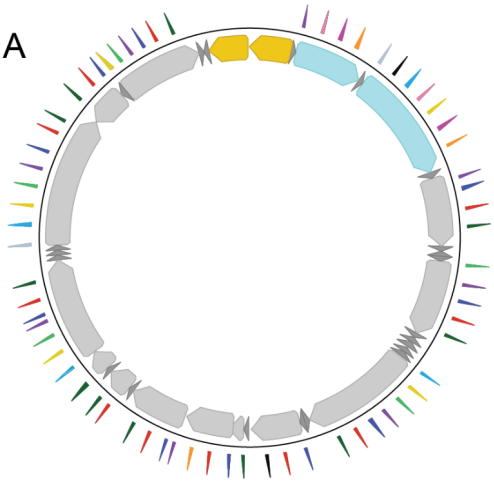


Supplementary Figure 2. Detailed schematic for MAESTER dial-out PCR protocol from Seq-Well or Drop-Seq WTA. Example for a single primer targeting one transcript (MT-CO1). In total, 65 targeted primers are used to enrich mitochondrial transcripts.



Supplementary Figure 3. Detailed schematic for MAESTER dial-out PCR protocol from 10x WTA (cDNA). Example for a single primer targeting one transcript (MT-CO1). In total, 65 targeted primers are used to enrich mitochondrial transcripts.

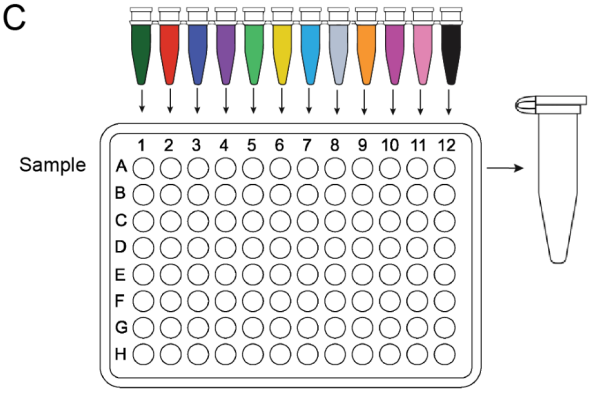
A



B

Transcript	Size	Number of Primers	Primer mix 1	2	3	4	5	6	7	8	9	10	11	12
RNR1	954	4				X					X	X	X	
RNR2	1159	7						X	X	X	X	X	X	X
ND1	956	4	X	X	X	X								
ND2	1042	5	X	X	X	X	X							
CO1	1542	7	X	X	X	X	X	X	X					
CO2	684	3		X	X									X
ATP8	207	1	X											
ATP6	681	3		X	X						X			
CO3	784	4	X	X	X	X								
ND3	346	2	X	X										
ND4L	297	1	X											
ND4	1378	7	X	X	X	X	X	X	X					
ND5	1812	8	X	X	X	X	X	X	X	X				
ND6	525	3	X	X	X									
CYB	1141	6	X	X	X	X	X	X						
Total			**11**	**11**	**10**	**8**	**5**	**5**	**4**	**2**	**3**	**2**	**2**	**2**

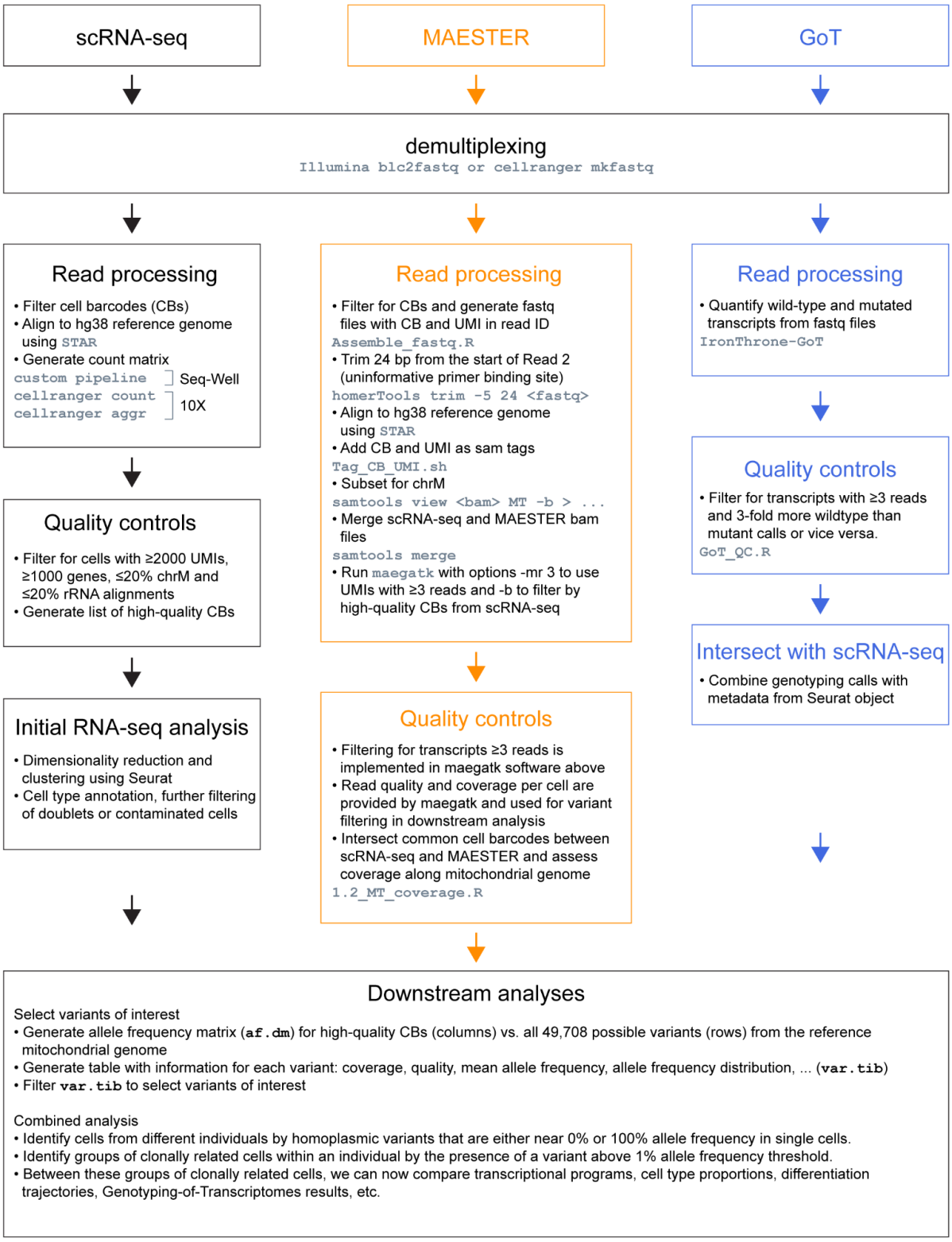
C



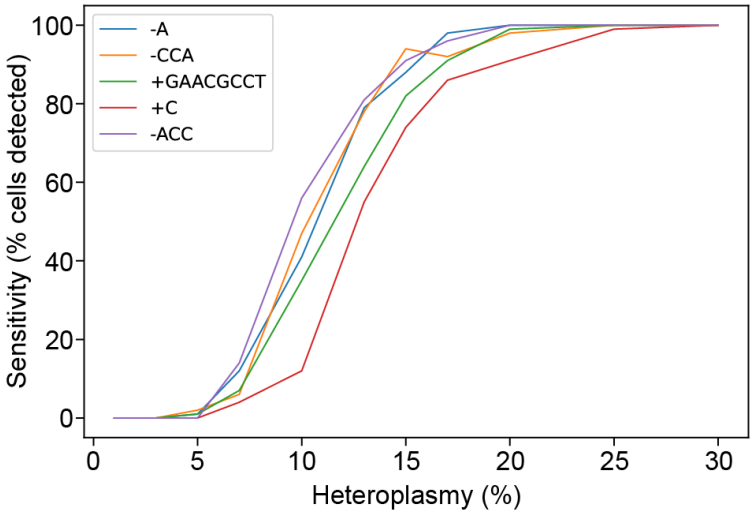
D

Primer mix	µl from reaction
Primer mix 1	20
Primer mix 2	25
Primer mix 3	25
Primer mix 4	25
Primer mix 5	25
Primer mix 6	25
Primer mix 7	25
Primer mix 8	25
Primer mix 9	5
Primer mix 10	5
Primer mix 11	5
Primer mix 12	10
Total Mix	**220**

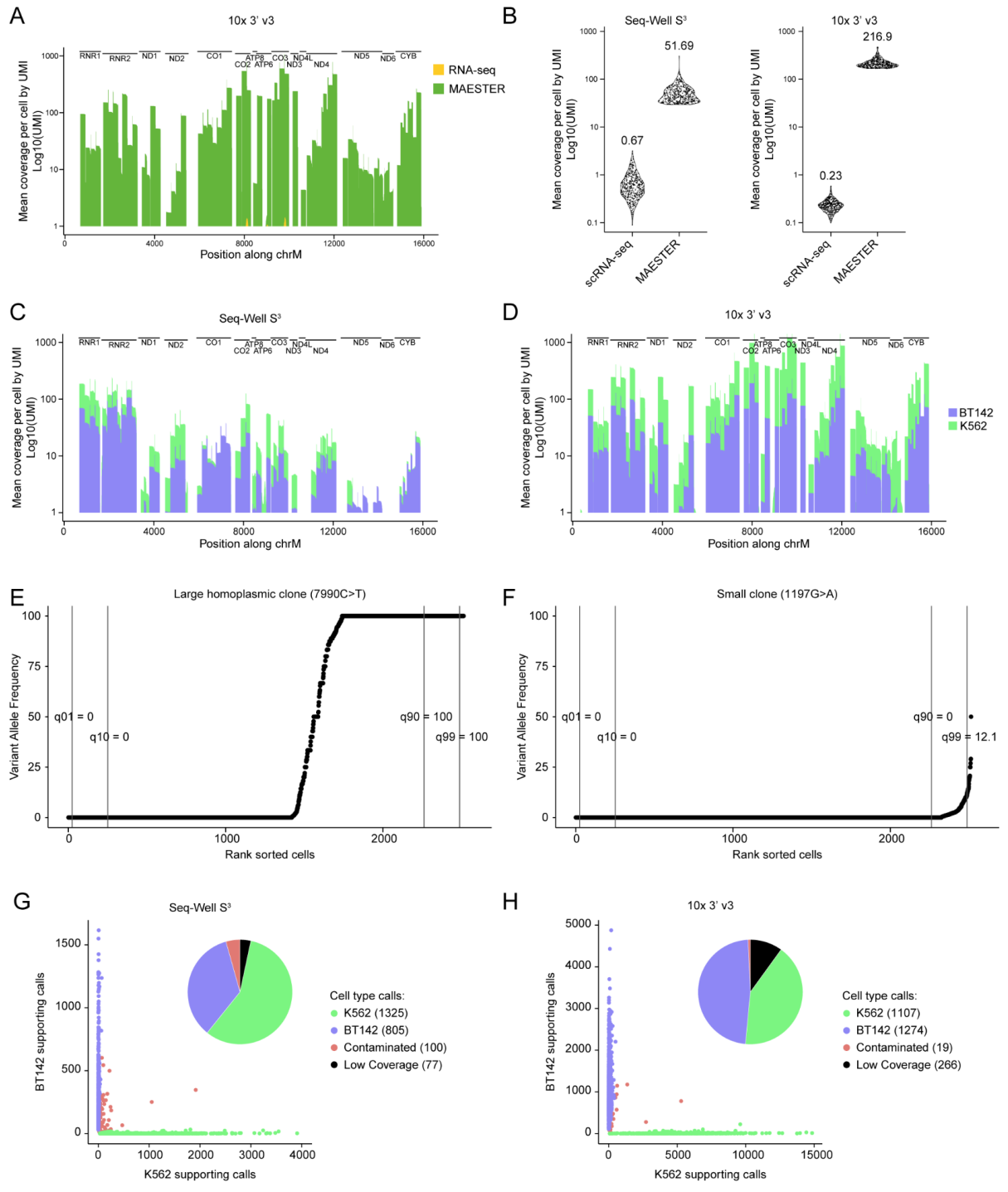
Supplementary Figure 4. Schematic and protocol for primer mixing.
A. Diagram depicts the circular mitochondrial genome. The colored triangles indicate where MAESTER primers bind. Colors indicate which primer mix they are in, as detailed in B and C. **B.** Table of 15 transcripts captured by scRNA-seq, with the length of the transcript and number of primers targeting that transcript. The table also indicates the primers for each transcript that are included in each of the 12 primer mixes for PCR1. This experimental design ensures that each transcript is targeted by no more than one primer per mix, that amplicons within the same PCR mix have similar sizes, and that highly expressed amplicons (mainly RNR1 and RNR2) do not overtake the reaction. **C.** Twelve PCR reactions with indicated primer mixes are completed for PCR1. Multiple samples can be processed in a 96 well plate. **D.** After PCR1, the 12 reactions are mixed together according to the volumes in the chart. This mixture is used as a template for PCR2 to generate sequencing libraries. For MAESTER from 10x genomics, since we used a 40 µl volume for PCR1, the indicated volumes were multiplied by 1.6.



Supplementary Figure 5. Overview of computational analyses.
Flow chart shows the major steps for the integrated analysis of scRNA-seq, MAESTER and Genotyping of Transcriptomes data from the same single cells. Indicated in grey are software tools (bcl2fastq, cellranger, STAR, homerTools, IronThrone-GoT, maegatk) and custom scripts (ending in .sh or .R), all of which are publicly available (see Data and Code availability).

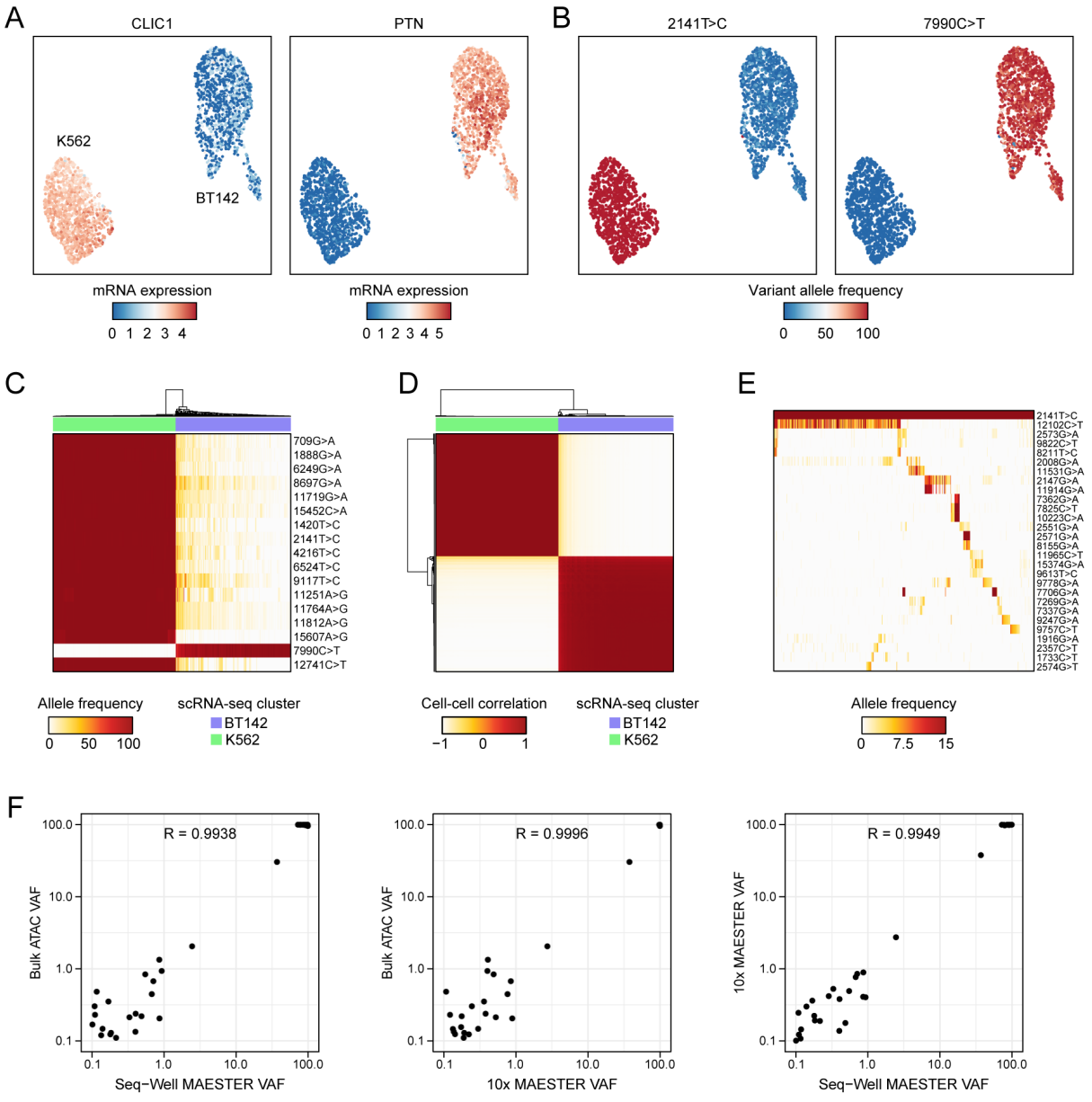


Supplementary Figure 6. Simulation of indel detection. Line graph shows the percent of cells in which different indels (indicated by color) are detected for different heteroplasmy levels (with 100-fold coverage per cell). This simulation shows that maegatk is able to pick up various indels with good sensitivity.



Supplementary Figure 7. Supporting data for Figure 1.
A. Barplot shows coverage of the mitochondrial genome with and without amplification from 10x libraries using MAESTER (see Figure 1C for Seq-Well). Mean coverage of 2,488 K562 and BT142 cells is shown. Each of the transcripts (UMIs) was sequenced ≥3 times. **B.** Sina plots show the mean coverage along all bases of the mitochondrial genome for the top 500 cells (symbols) analyzed by scRNA-seq alone or MAESTER as indicated. The mean is indicated on top. The difference in MAESTER coverage between Seq-Well and 10x is in part due to different sequencing depths: 200,395,346 and 683,202,846 reads, respectively. **C-D**. Barplots show mean coverage of the mitochondrial genome in the cell line mixing experiments from MAESTER

on Seq-Well (C) or 10x (D) cDNA. Colors indicate coverage in BT142 and K562 cells. Each of the transcripts (UMIs) was sequenced ≥3 times. **E-F.** Scatter plots depict an important strategy by which informative variants were selected. For each variant the allele frequency across all cells was sorted from low to high, followed by determining the VAF at different percentiles (e.g. the 1st, 10th, 90th and 99th percentile). All 6 (Seq-Well) and 17 (10x) homoplasmic variants that were used to distinguish between cell lines met the following parameters: mean coverage per cell >20, mean read quality >30, 10th VAF percentile of <10% and 90th VAF percentile of >90%. None of the other 49,691 possible variants met these parameters. **G-H.** Scatter plot shows the number of supporting calls (wild-type + mutant) for homoplasmic variants in K562 and BT142 cells. For Seq-Well (G), six variants shown in Figure 1F were used. For 10x (H), seventeen variants shown in Supplementary Figure 8C were used. Of the cells that were classified by mtDNA variants, 2,129/2,130 (100%) were concordant with mRNA clusters for Seq-Well and 2,380/2,381 (100%) were concordant with mRNA clusters for 10x.

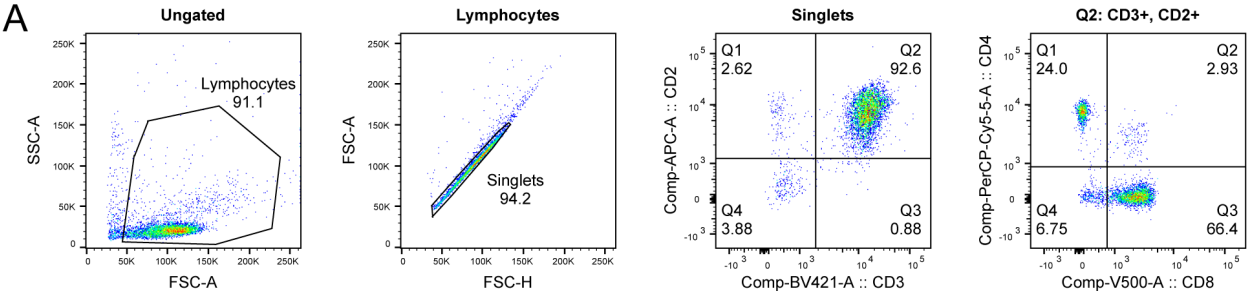


Supplementary Figure 8. Analysis of cell line mixing experiment with 10x scRNA-seq and bulk ATAC-sequencing.

A-B. UMAPs show detection of cell type-specific (A) gene expression from scRNA-seq and (B) homoplasmic mtDNA variants from MAESTER. Cluster identity is confirmed by mRNA expression of *CLIC1 (*a K562-specific gene) and *PTN (*a BT142-specific gene). Using MAESTER, we detected the mtDNA variant 2141T>C specifically in K562 cells and 7990C>T specifically in BT142 cells. **C.** Heatmap depicts separation of K562 and BT142 cells (columns) based on 17 mtDNA variants detected using MAESTER (rows). Cell type annotation from scRNA-seq is shown on top. **D.** Correlation matrix shows cell similarity based on the allele frequencies of 17 homoplasmic variants (rows and columns depict 2,083 cells). Unsupervised clustering identified two clusters that correlate with cell annotations from scRNA-seq. **E.** Heatmap shows VAF of 27 mtDNA variants detected by MAESTER (rows) for 788 K562 cells (columns, 71.2% of all 1107 K562 cells in this experiment) with informative subclonal variants. Homoplasmic K562 variant 2141T>C is shown for comparison. Heatmap is organized by clonal

structure (Methods). For B-E, only cells with >3-fold coverage of the indicated variants are shown. This figure shows analyses of MAESTER with 10x; see Figure 1D-H for similar analyses of MAESTER with Seq-Well. **F.** Scatter plots compare VAFs between MAESTER and bulk ATAC-seq, for all 57 informative K562 variants that were identified using Seq-Well or 10x MAESTER on the same passage of K562 cells (also shown in Figure 1D-H and Supplementary Figure 8A-H). For MAESTER, the VAF of K562-derived transcripts is shown. For bulk ATAC-seq, the VAF based on mitochondrial genome alignments is shown.

A



B

Bone Marrow Aspirate Smear	
BPDCN blasts (%)	None
Blasts	1%
Promyelocytes	1%
Myeloid	46%
Erythroid	37%
Lymphocytes	13%
Plasma cells	2%
Others	None
M:E ratio	1.2
Marrow dysplasia	No

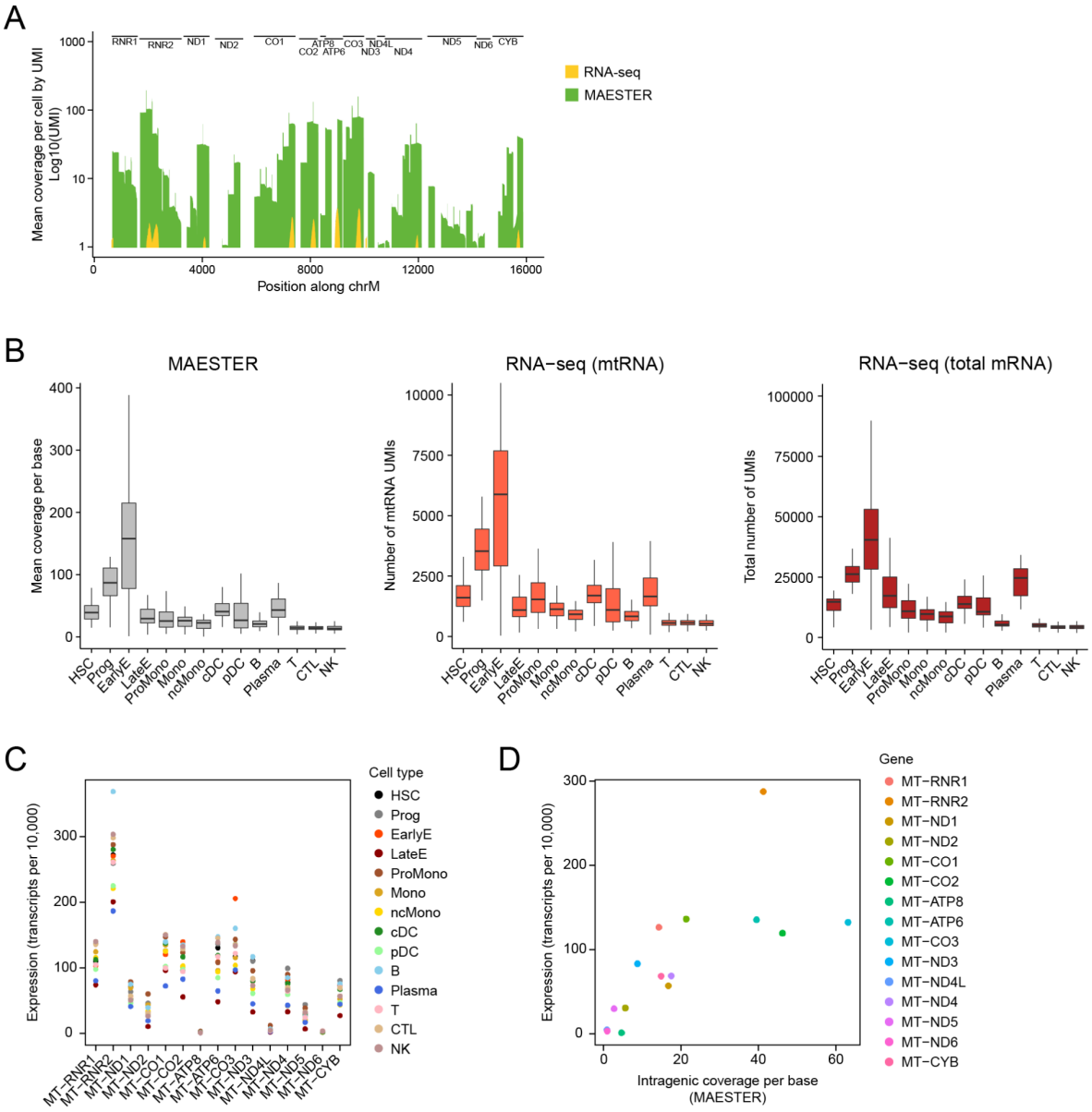
C

Peripheral Blood Counts	
WBC (4-10k/uL)	2.44
ANC (1.92-7.6k/uL)	0.85
AMC (0.16-1.1k/uL)	0.32
ALC (0.21-2.74k/uL)	1.27
Hgb (13.5-18g/dL)	15.2
MCV (80-95 fl)	86.9
Plt (150-450k/uL)	84

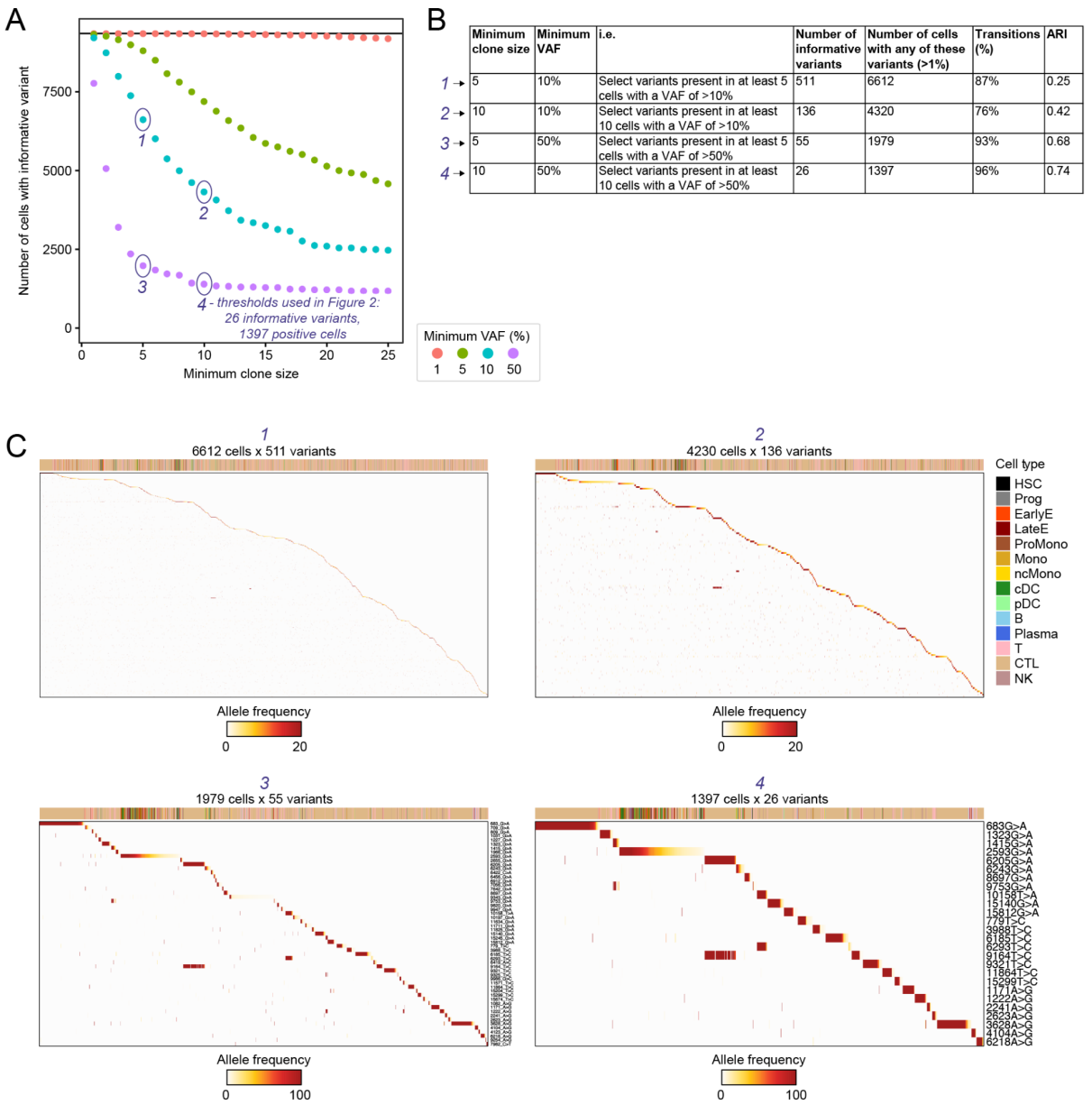
D

Gene	Mutation	VAF
TET2	S792*	29.5
	Q1034*	31.4
	R1216*	8.8
	H1380Y	6.9
ASXL1	G642fs*	44.7

Supplementary Figure 9. Patient characteristics of the clonal hematopoiesis sample. **A.** Flow cytometry plots of the same bone marrow aspirate used for 10x sequencing in Figure 2. Single lymphocytes were determined by forward and side scatter. CD2+, CD3+ cells represent T-cells (92.6%). CD2+, CD3- cells represent NK cells (2.6%). Of T-cells, 66% were CD8+. This is consistent with an enrichment of cytotoxic T-cells in the overall sample, in accordance with cell classification by scRNA-seq. **B**. Pathology summary of cells in bone marrow aspirate. **C**. Peripheral blood counts of the patient at the time of bone marrow aspirate. **D.** Mutations detected by targeted sequencing of the bone marrow aspirate.



Supplementary Figure 10. MAESTER coverage correlates with mtRNA content.
A. Barplot shows mean coverage of the mitochondrial genome in cells from the clonal hematopoiesis sample with and without amplification from 10x libraries using MAESTER. Each of the transcripts (UMIs) was sequenced ≥3 times. **B.** Left: boxplot shows the mean MAESTER coverage per base along the mitochondrial genome per cell (y-axis), split by cell types (on the x-axis). Middle: boxplot shows the count of mitochondrial RNAs (mtRNA) molecules per cell from scRNA-seq, split by cell types. Right: boxplot shows the total count of mRNA molecules per cell from scRNA-seq, split by cell types. The similarity between boxplots indicates that MAESTER coverage depends on gene expression / the availability of mRNA transcripts. Boxplots show the median, 25th and 75th percentiles, and smallest/largest values (geom_boxplot defaults). **C**. Plot of average gene expression by scRNA-seq per mitochondrial gene per cell type in the clonal hematopoiesis sample. **D**. Dotplot of the average scRNA-seq expression of each gene across all cells in the clonal hematopoiesis sample relative to MAESTER coverage of the same gene in those cells.

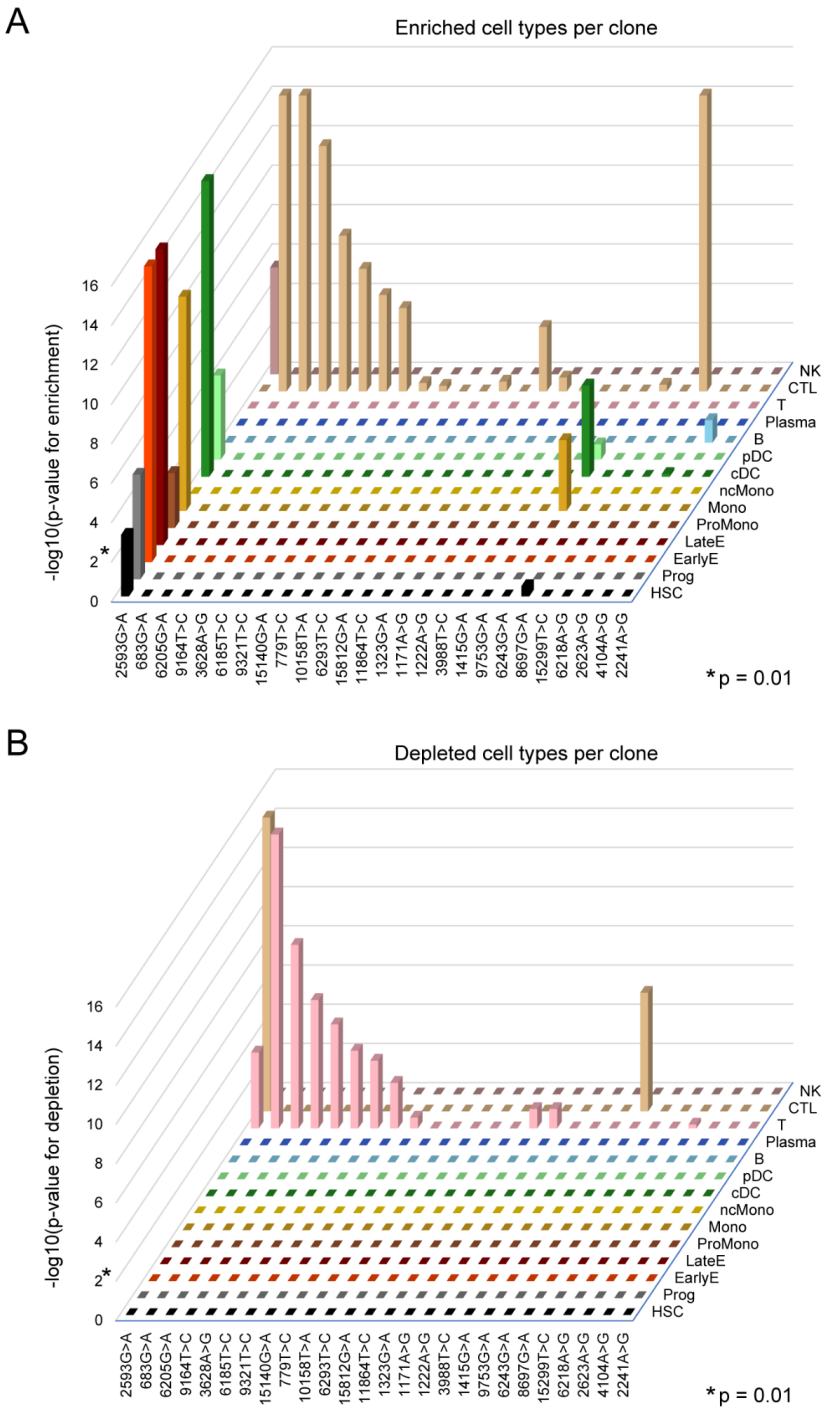


	Minimum clone size	Minimum VAF	i.e.	Number of informative variants	Number of cells with any of these variants (>1%)	Transitions (%)	ARI
1 →	5	10%	Select variants present in at least 5 cells with a VAF of >10%	511	6612	87%	0.25
2 →	10	10%	Select variants present in at least 10 cells with a VAF of >10%	136	4320	76%	0.42
3 →	5	50%	Select variants present in at least 5 cells with a VAF of >50%	55	1979	93%	0.68
4 →	10	50%	Select variants present in at least 10 cells with a VAF of >50%	26	1397	96%	0.74

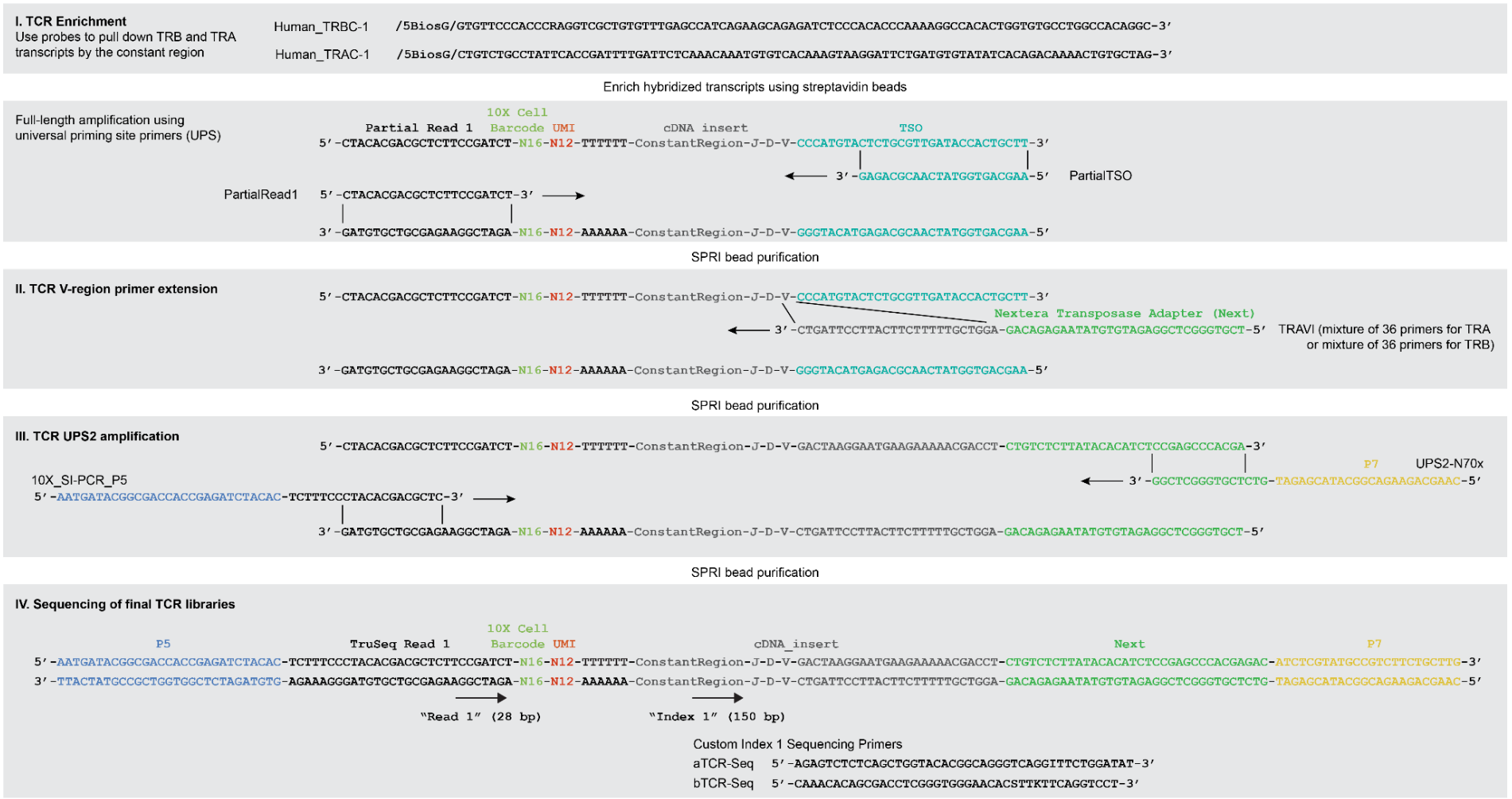
Supplementary Figure 11. Selection of informative variants in the clonal hematopoiesis sample. A. Scatter plot shows the number of cells with an informative variant (y-axis) using two selection criteria, minimum VAF (colors) and minimum clone size (x-axis). Horizontal line indicates the total number of cells in the sample (9,346). The number of cells that can be tracked increases by relaxing selection criteria to define informative variants. **B.** Table shows results when using four of the informative variant selection criteria highlighted in panel A. To select high-confidence and distinct clones for further analysis in this manuscript, we selected variants that were absent in 90% of the cells and had a VAF of >50% (minimum VAF) in at least 10 cells (minimum clone size). Most of these 26 variants were transitions as opposed to transversions, as expected. There were 1397 cells with a VAF of >1% for any of the 26 selected variants. Comparing clones that were called using these mtDNA variants to clones that were called based on TCR rearrangement yielded an Adjusted Rand Index (ARI) of 0.74, indicating good concordance (see also Figure 2F). **C.** Heatmaps show the VAF of variants (rows) in cells (columns) with at least 1% VAF for one of the selected variants. Four heatmaps correspond to variant selection criteria highlighted in panels A and B. The bottom right heatmap is similar to Figure 2B with some differences in the order of rows and columns.

Supplementary Figure 12. Mitochondrial variants inform clonally related cells in clonal hematopoiesis.

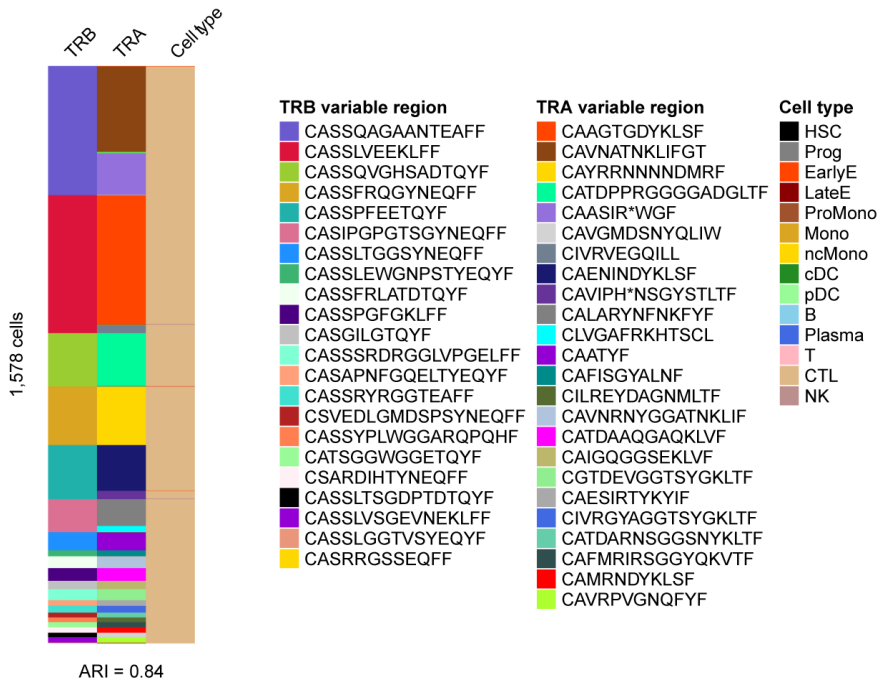
A. Chart shows the position of subclonal variants that were identified in the clonal hematopoiesis sample, along all 16,569 bases of the mitochondrial genome. At the bottom, the position of genes is indicated in blue. **B.** UMAPs show VAFs per cell for informative variants in Figure 2B. Cell coordinates are the same as in Figure 2A.



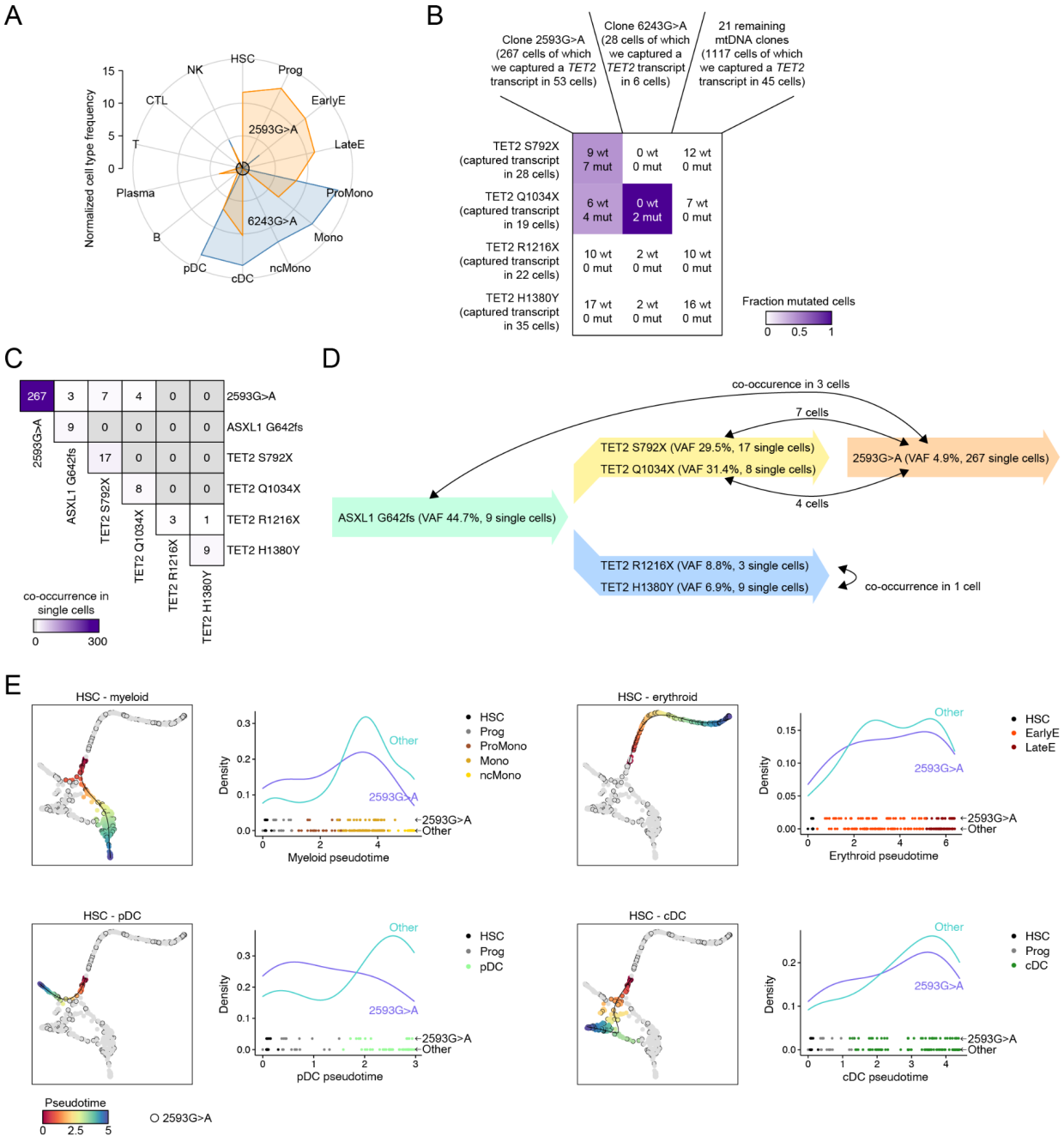
Supplementary Figure 13. Lineage bias in clones identified by mtDNA variants.
A-B. Enrichment (A) or depletion (B) of 14 cell types detected in each of the 26 clones compared to expected from the bulk sample, related to Figure 2D. *P*-values were calculated using the cumulative distribution function of the hypergeometric distribution separately for enrichment and depletion (one-sided). *P*-values were corrected for multiple testing by Bonferroni correction (n=364).



Supplementary Figure 14. Detailed schematic of the T-cell receptor sequencing (TREK-seq) protocol. These procedures are used to enrich variable regions from 10x 3' cDNA. See Supplemental Table 3 for primer sequences.



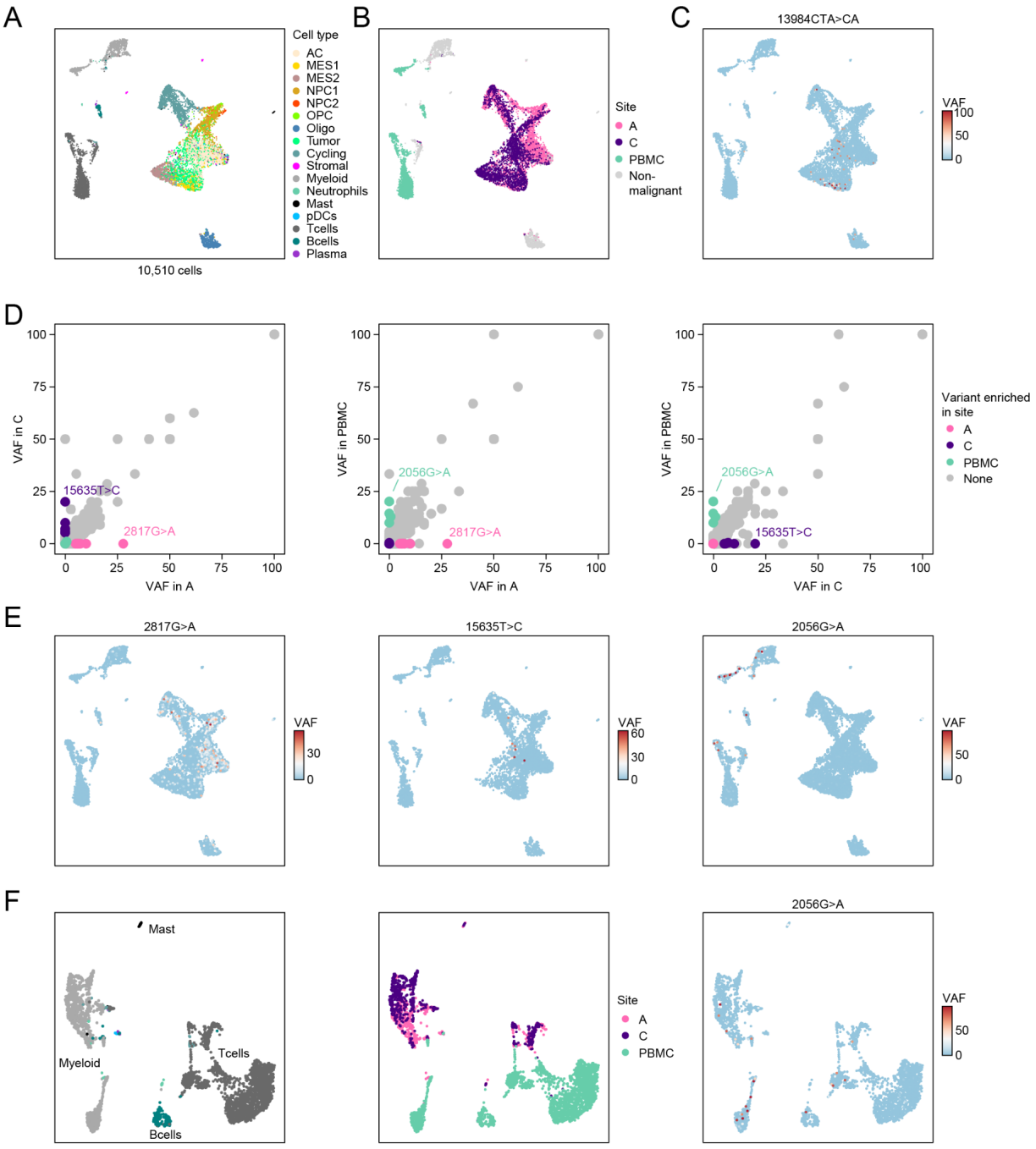
Supplementary Figure 15. Overlap between TRB and TRA variable regions. Plot shows cells (rows) in which both *TRB* and *TRA* variable region sequences were detected, sorted by *TRB* sequences. *TRB* and *TRA* sequences, along with cell type, are indicated by colors in legend.



Supplementary Figure 16. Cells marked by 2593G>A or 6243G>A exhibit lineage bias and *TET2* mutations.

A. Radar plot shows normalized cell type frequencies for clones marked by 2593G>A or 6243G>A. Frequency is normalized to the frequency of cell types for all cells, which is set to 1 and denoted by a black circle at the center. **B.** Chart shows the number of cells within mtDNA-defined subclones (detected by MAESTER) that express wild-type or mutated *TET2* transcripts (detected by GoT). The fraction of mutated cells in each box is designated by color scale. We recovered too few *ASXL1* transcripts to include in this analysis. **C.** Heatmap shows co-occurrence of the mtDNA mutation 2593G>A and *ASXL1* and *TET2* driver mutations in single cells. These co-occurrences, together with bulk VAFs, informed the phylogeny in the next panel. **D.** Visualization of the most likely order of nuclear and mitochondrial mutation

acquisitions. Clinical Rapid Heme Panel sequencing detected ASXL1 G642fs (VAF 44.7%), TET2 S792X (VAF 29.5%), TET2 Q1034X (VAF 31.4%), TET2 R1216X (VAF 8.8%) and TET2 H1380Y (VAF 6.9%). MAESTER detected 2593G>A at a VAF of 4.9%. Black arrows indicate all detected mutation co-occurrences from the single-cell genotyping protocols GoT and MAESTER. **E.** UMAPs depicts pseudotime of four differentiation trajectories. Cells in the 2593G>A clone are marked by a black circle. Line plots show relative cell density along pseudotime of differentiation in cells within clone 2593G>A and in cells not within that clone. Symbols at the bottom of the line graphs indicate the cells, colored according to the legend.



Supplementary Figure 17. Leveraging mitochondrial variants to resolve clonal relationships in glioblastoma cells.

A. Combined UMAP of three scRNA-seq samples from the same patient with a glioblastoma tumor analyzed using Seq-Well. Samples include two sections of tissue from different regions of the tumor and corresponding peripheral blood mononuclear cells (PBMCs) from the same patient taken at the time of surgery. Cell types were annotated using established gene signature scores. **B.** UMAP shows sample sources for malignant cells and PBMCs. Non-malignant cells from the tumor samples are also included in gray. **C.** UMAP is colored by the VAF of deletion 13984CTA>CA, which is nearly exclusively detected in malignant tumor cells. This is the only indel called by maegatk-indel with a Freebayes quality score >50, not containing 4 repeats of the same base, detected in at least 5 cells, and not present in the clonal hematopoiesis sample.

The Variant Effect Predictor (VEP) did not predict functional impact of 13984CTA>CA from Polyphen or SIFT, with no clinical significance or associated disease. **D.** Scatterplots show the 99th percentile VAF for each variant in cells from the indicated sample site. Colors indicate mtDNA variants that are specific to one tumor region or to PBMCs. Cumulatively we identified 19 informative variants marking 20.1% of cells. **E.** UMAPs are colored by the VAF of selected variants that were specific to a tumor region or to PBMCs. **F.** UMAPs of only immune cells from the 3 specimens are colored by cell type (left), sample source (middle), and the VAF of 2056G>A per cell (right). Of note, variant 2056G>A was found in myeloid cells within PBMCs and also myeloid cells in the tumor. This suggests that these tumor-associated myeloid cells were derived from circulating monocytes.