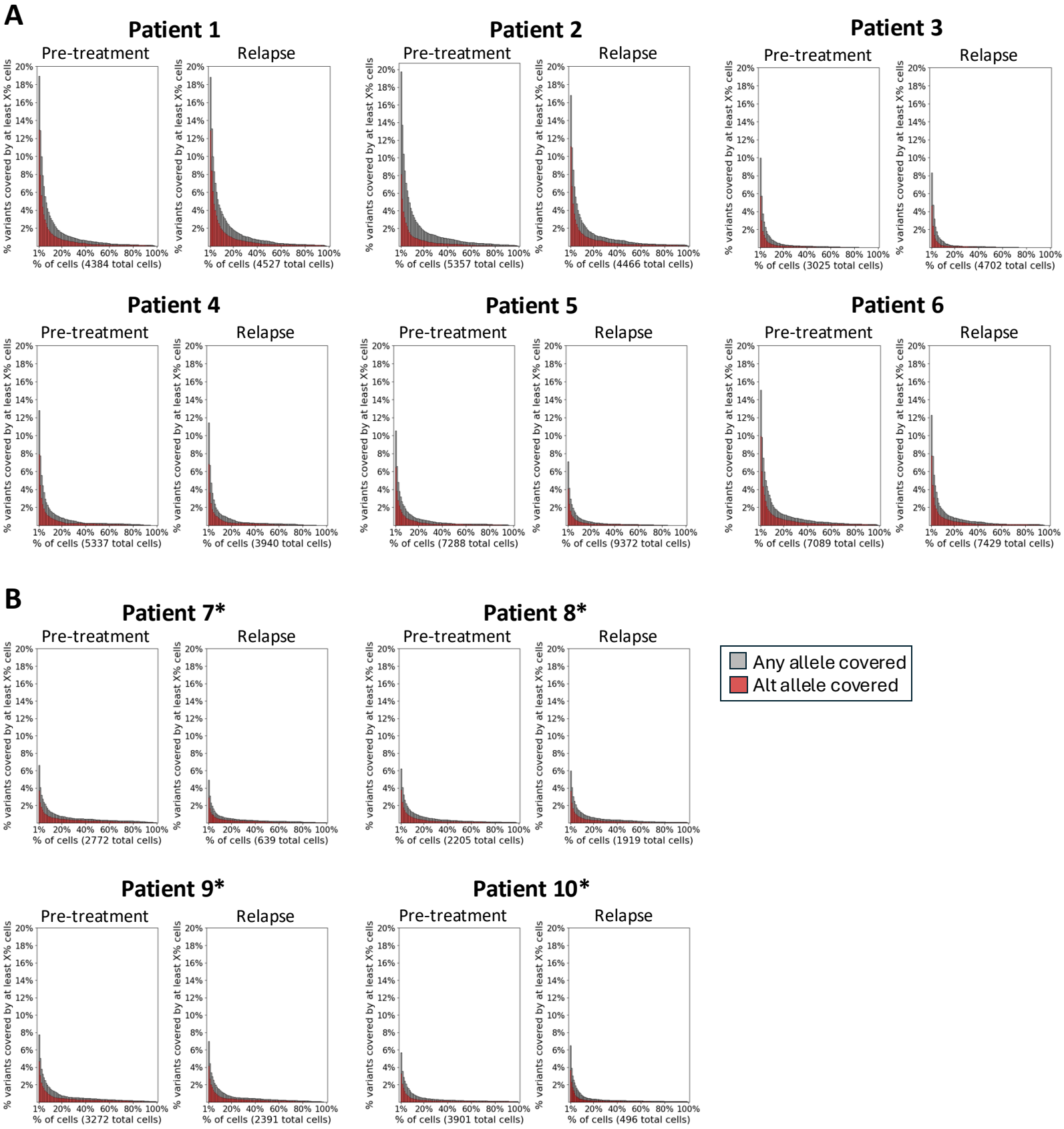
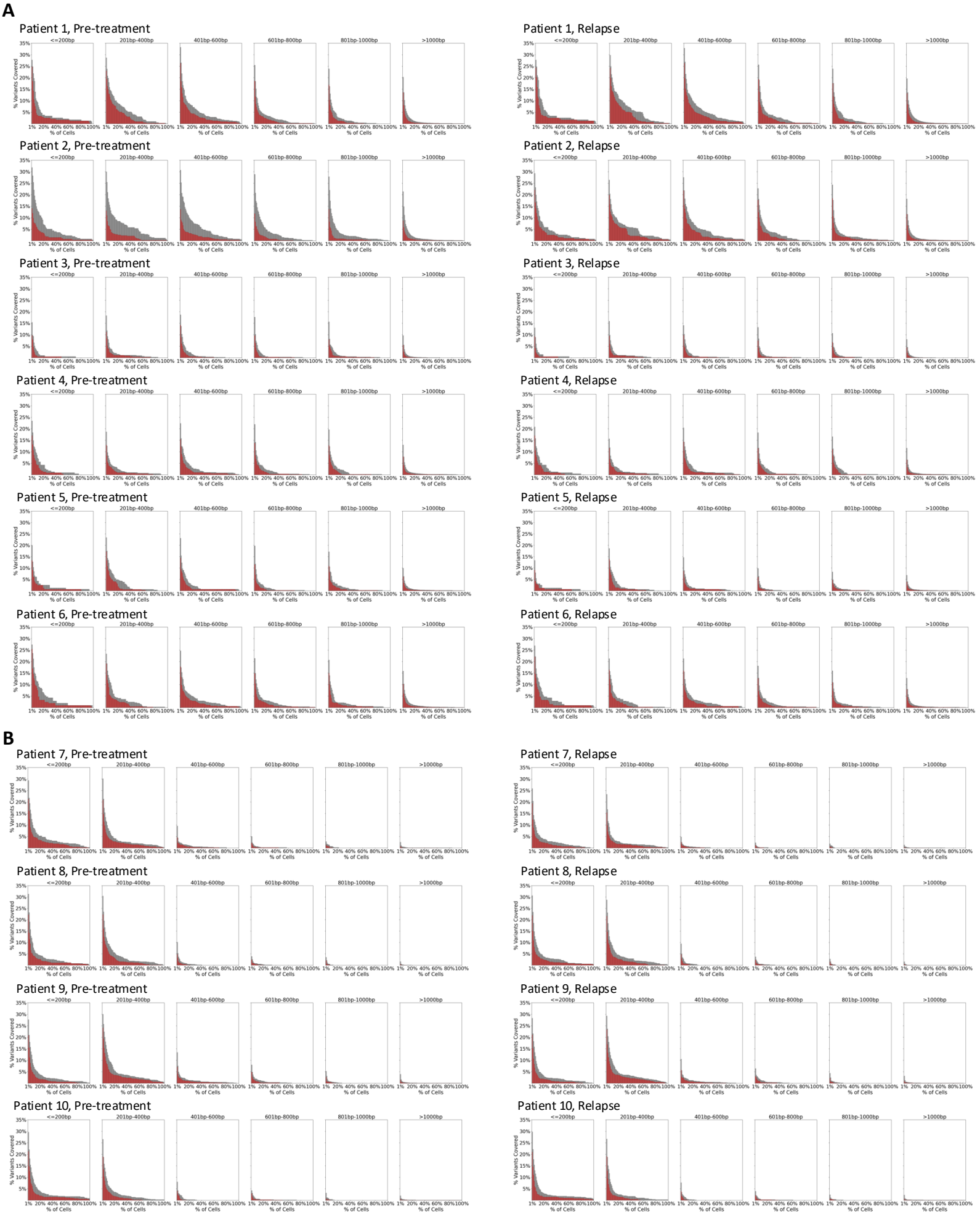
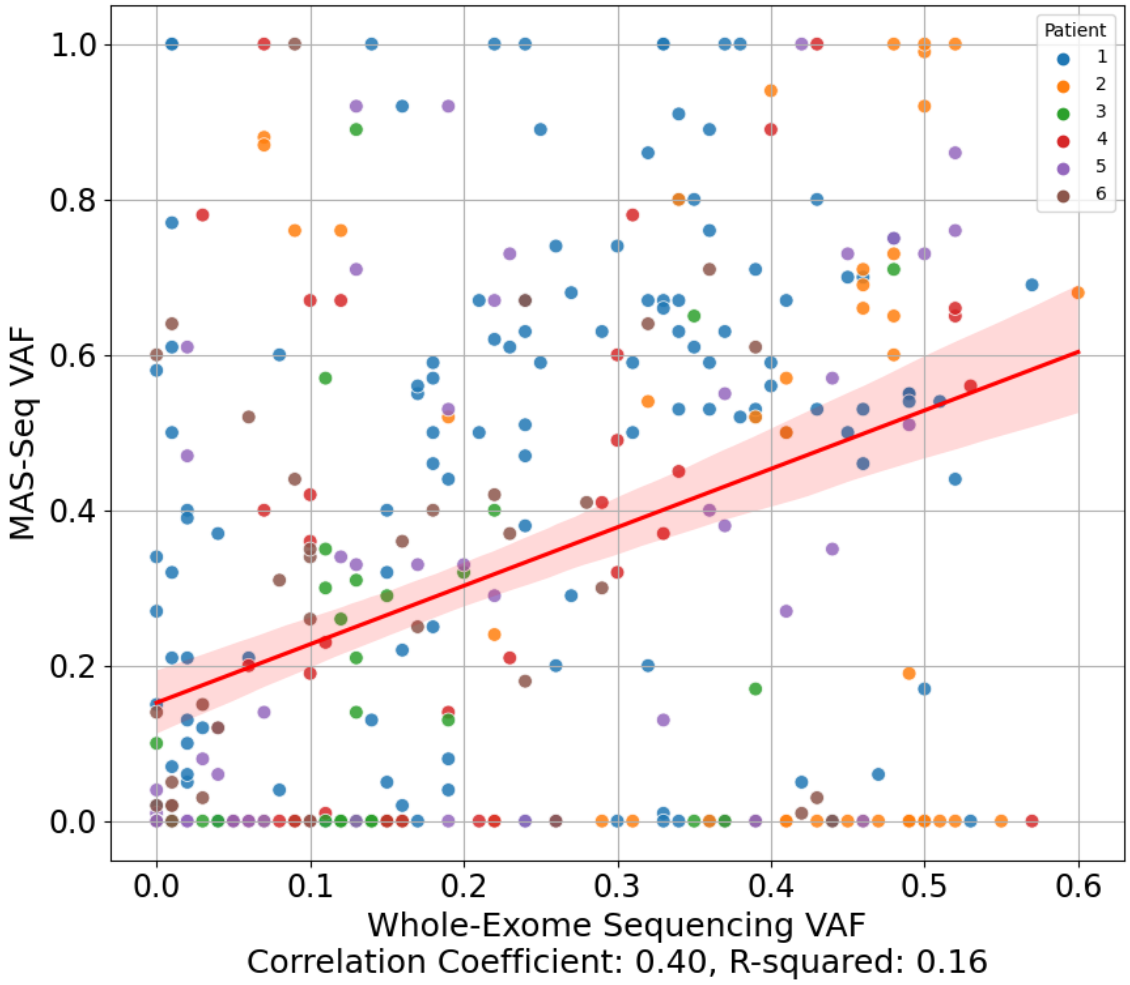


Supplemental Figure 1. The overall variant coverage for all pre-treatment and relapse samples included in this study. The percentage of cells covering each heterozygous germline variant in the patient's WES data is used to determine the percent of variants covered by at least X% of cells. A) Samples sequenced with MAS-Seq. Patients 1, 2, 5, and 6 were sequenced on the Revio machine, while Patients 3 and 4 were sequenced on the Sequel II machine. B) Samples sequenced using short-read sequencing technology. *Sequencing was done previously for these patients using 10X Genomics library prep, followed by short-read Illumina sequencing. These samples were not included in the remaining analyses of this study.

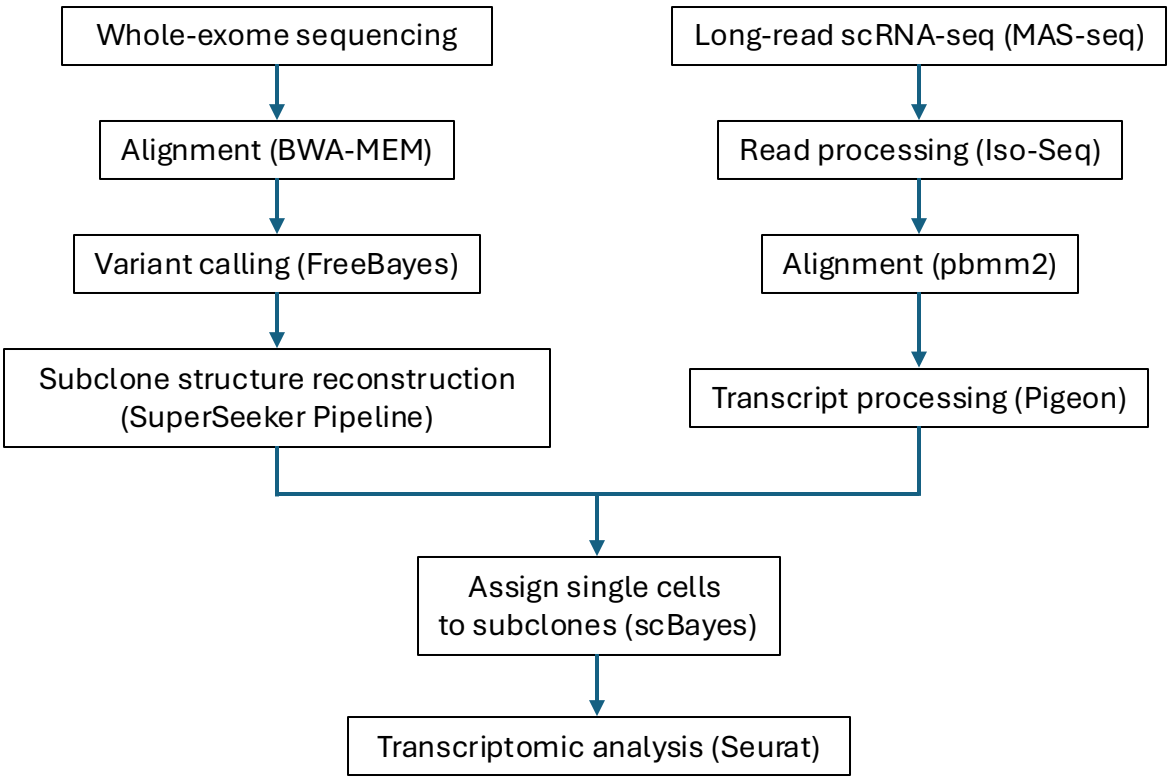


Supplemental Figure 2. Variant coverage for all samples, binned by distance from the priming site. The percentage of cells covering each heterozygous germline variant in the patient's WES data determines the percent of variants covered by at least X% of cells. A) MAS-Seq samples. B) Short-read sequenced samples.

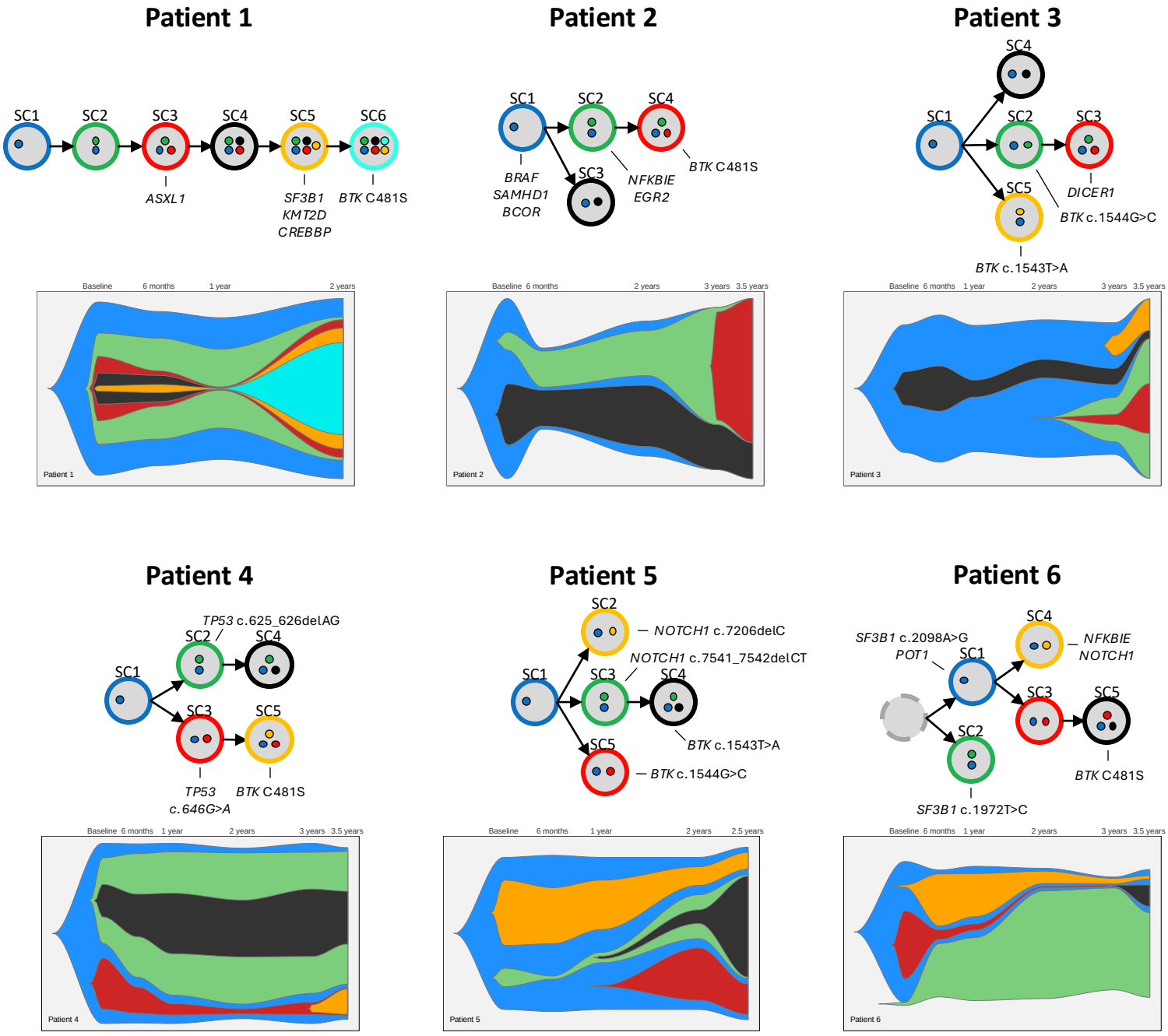
Any allele covered
Alt allele covered



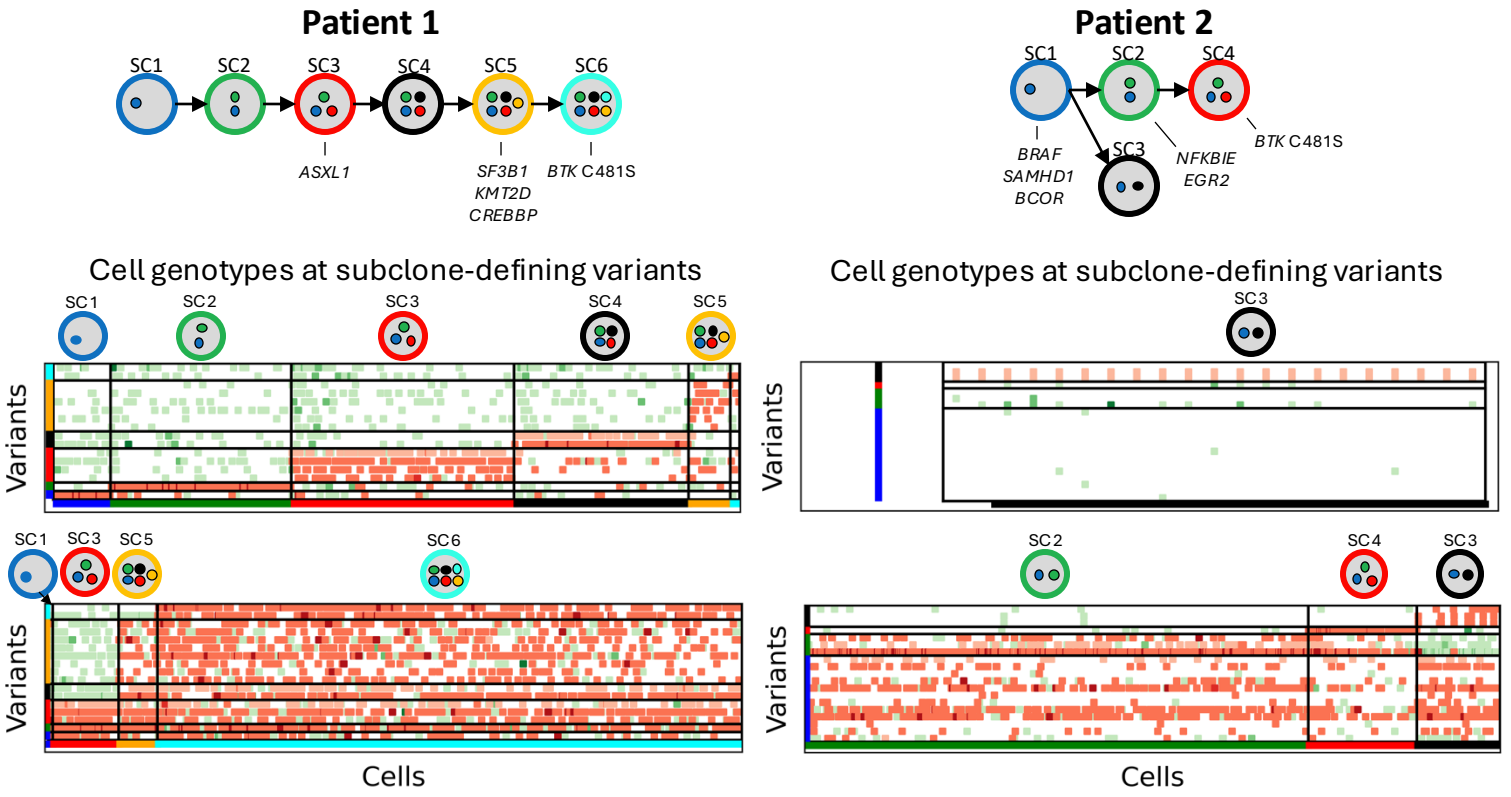
Supplemental Figure 3. Bivariate plot comparing the variant allele frequencies identified by whole-exome sequencing and MAS-Seq. Somatic mutations across all six patients are shown. Each dot represents a single variant, colored by the patient from which it belongs. Red line represents the regression line fitted to the data points.

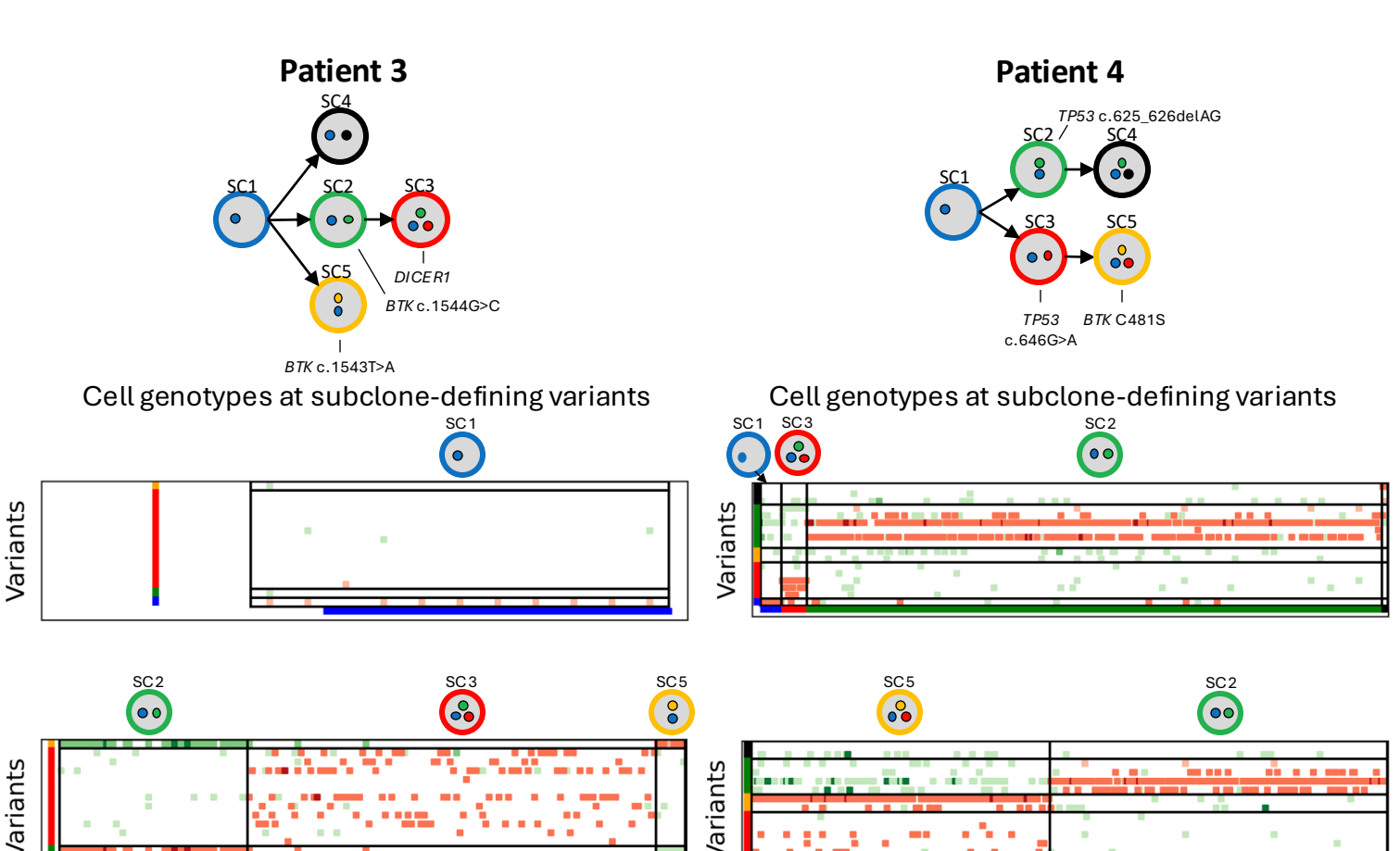


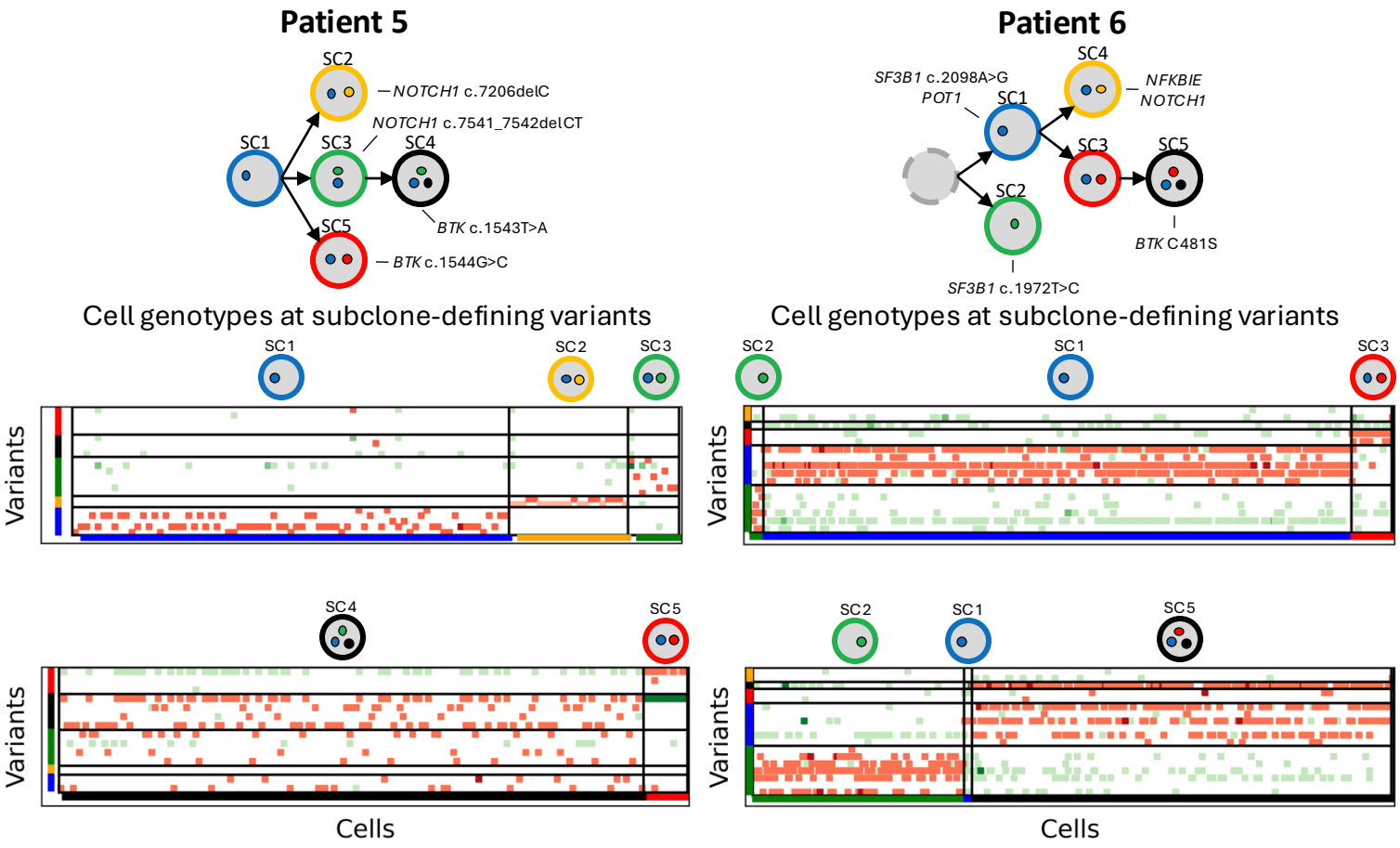
Supplemental Figure 4. Consort plot showing the analysis workflow. Whole-exome sequencing and long-read scRNA sequencing data are processed and combined to identify variants in subclones and analyze transcriptomic characteristics of subclones.



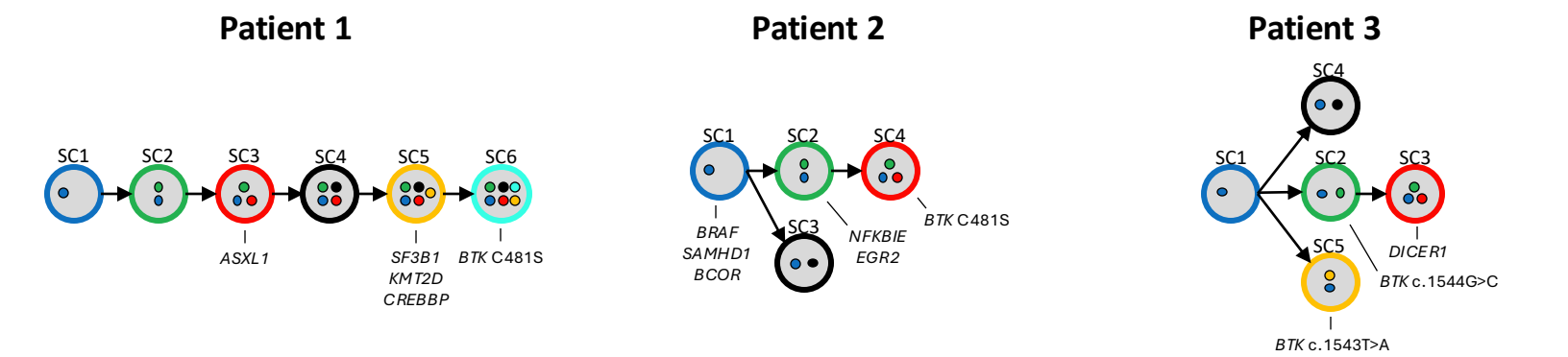
Supplemental Figure 5. Subclone structures for each patient are depicted using hierarchical trees. CLL-driver gene mutations are annotated on the subclone in which they arose. Subclone structures were identified initially in bulk DNA sequencing data and either confirmed or refined with the long-read scRNA-seq data. Fishplots depict the subclonal evolution in each patient over time, indicating sampling time points where WES data is available. Long-read scRNA-seq data was generated for the baseline/pre-treatment sample and the final sample from the patient.

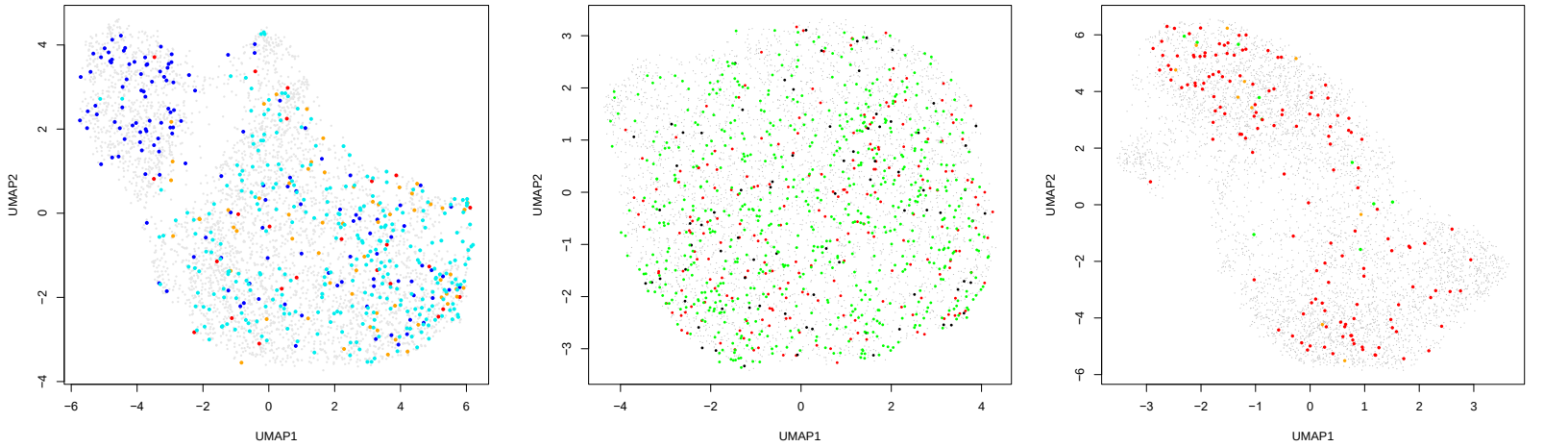

Patient 1
SC1
SC2
SC3
SC4
SC5
SC6
ASXL1
SF3B1
KMT2D
CREBBP
BTK C481S
Patient 2
SC1
SC2
SC4
SC3
BRAF
SAMHD1
BCOR
NFKBIE
EGR2
BTK C481S
Cell genotypes at subclone-defining variants
Variants
Cells

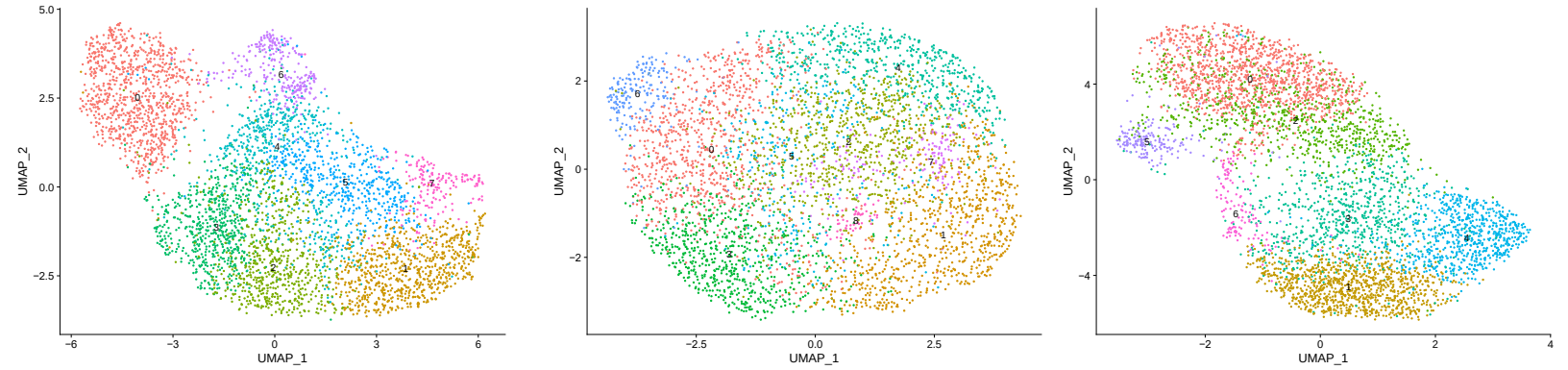

Patient 3
SC4
SC1
SC2
SC3
SC5
DICER1
BTK c.1544G>C
BTK c.1543T>A
Patient 4
TP53 c.625_626delAG
SC2
SC4
SC1
SC3
SC5
TP53
c.646G>A
BTK C481S
Cell genotypes at subclone-defining variants
Variants
Cells

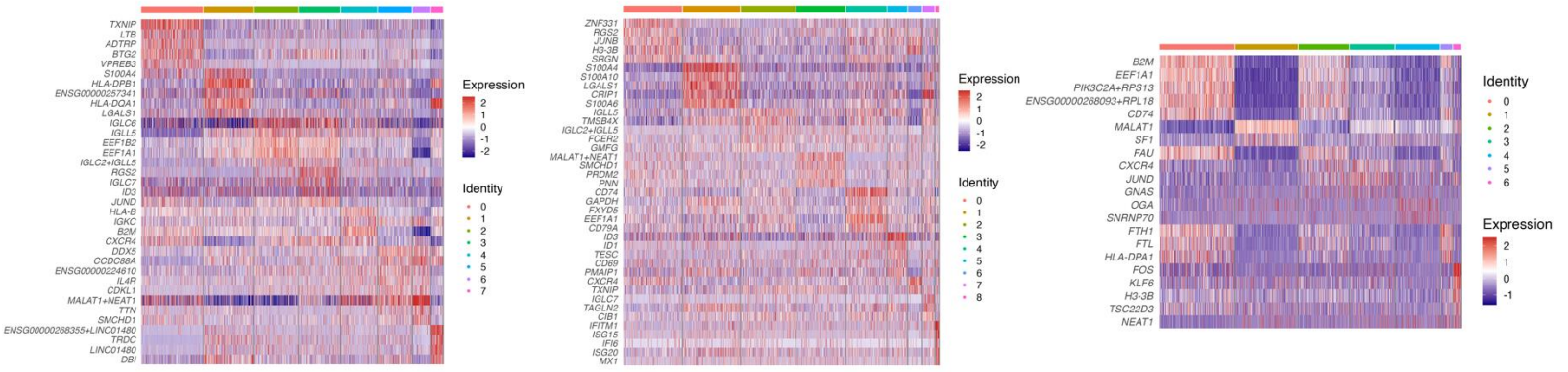


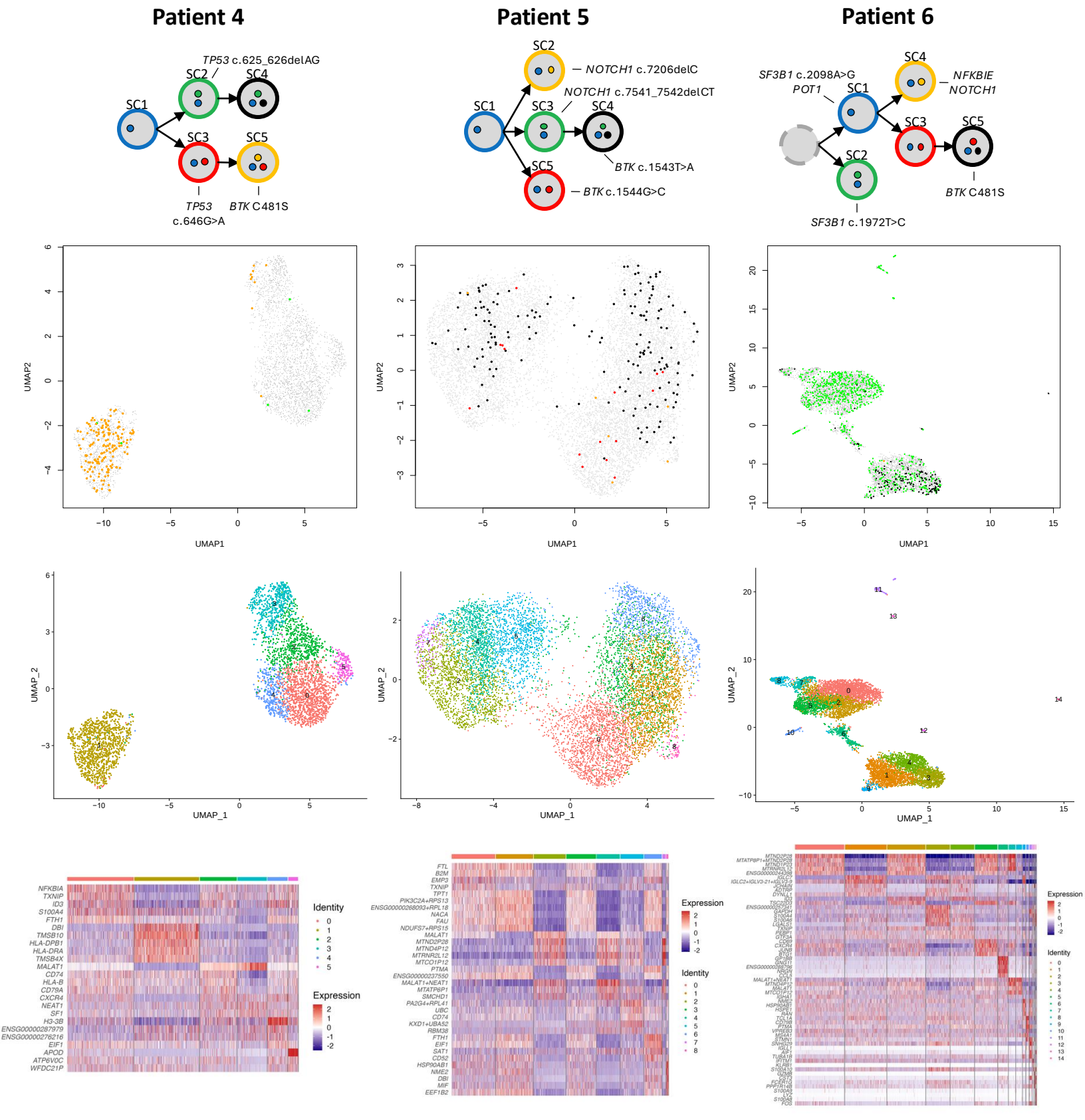
Supplemental Figure 6. Genotype Matrix Plots of the relapse sample for each patient is shown together with the patient's subclone structure to visualize the genotype at the time of relapse using the long-read scRNA-seq data. Pre-treatment plots are on top, followed by the relapse-sample plot on bottom. Each cell that was successfully assigned to a subclone is shown (X-axis), along with every variant that was present in the WES data (Y-axis).Green markers represent only reference alleles present in the scRNA-seq reads at the given variant location within the cell, while red markers indicate at least one scRNA-seq read in the cell contains the somatic variant allele. Darker marker coloring indicates an increased number of reads supporting that genotype. In the genotype matrix plot, variants and cells are grouped by their subclone assignment as denoted by the colored bars along the X and Y axis.


Patient 1
SC1
SC2
SC3
SC4
SC5
SC6
ASXL1
SF3B1
KMT2D
CREBBP
BTK C481S
Patient 2
SC1
SC2
SC4
SC3
BRAF
SAMHD1
BCOR
NFKBIE
EGR2
BTK C481S
Patient 3
SC4
SC1
SC2
SC3
SC5
DICER1
BTK c.1544G>C
BTK c.1543T>A

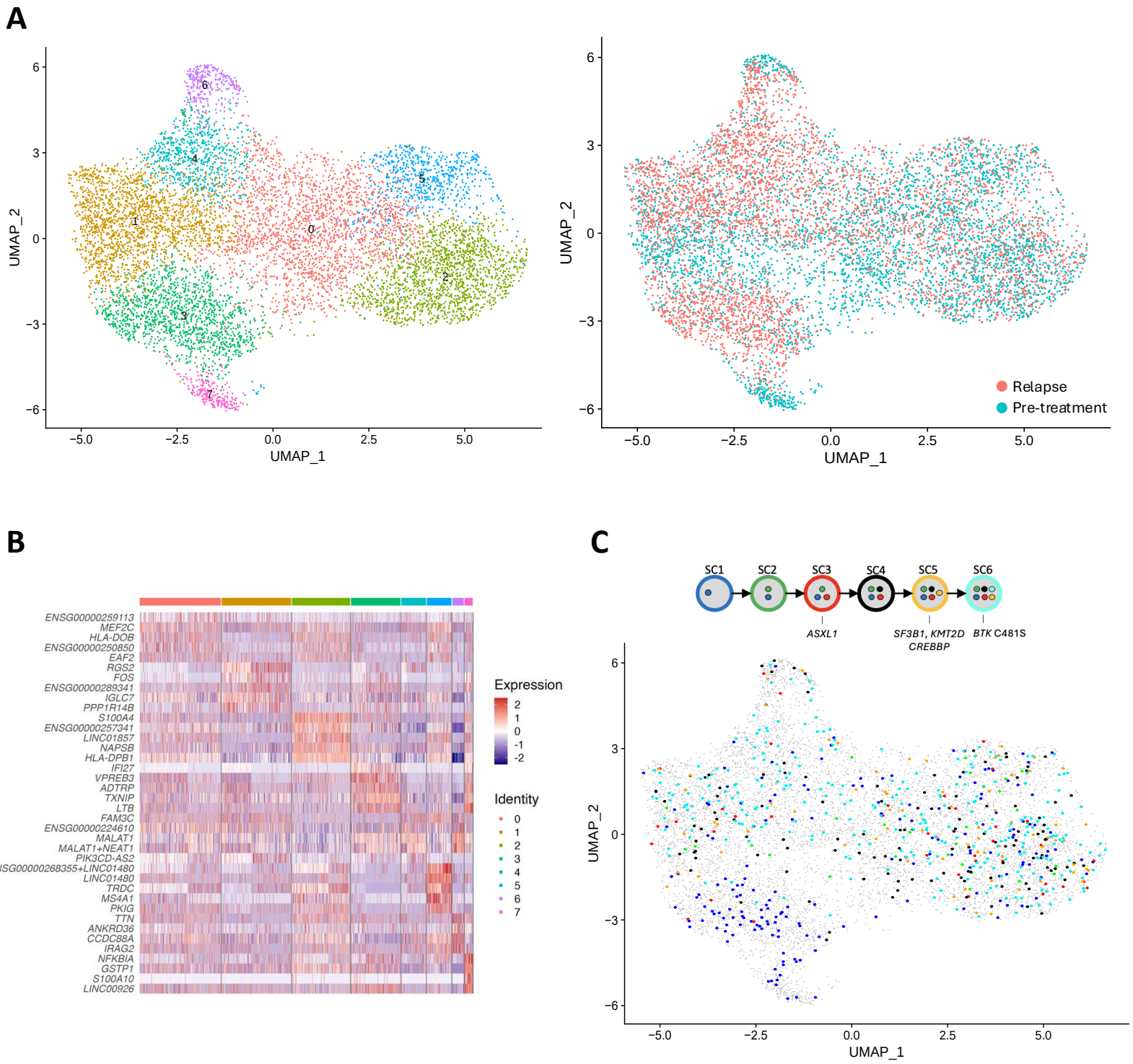

UMAP2
UMAP1


UMAP_2
UMAP_1

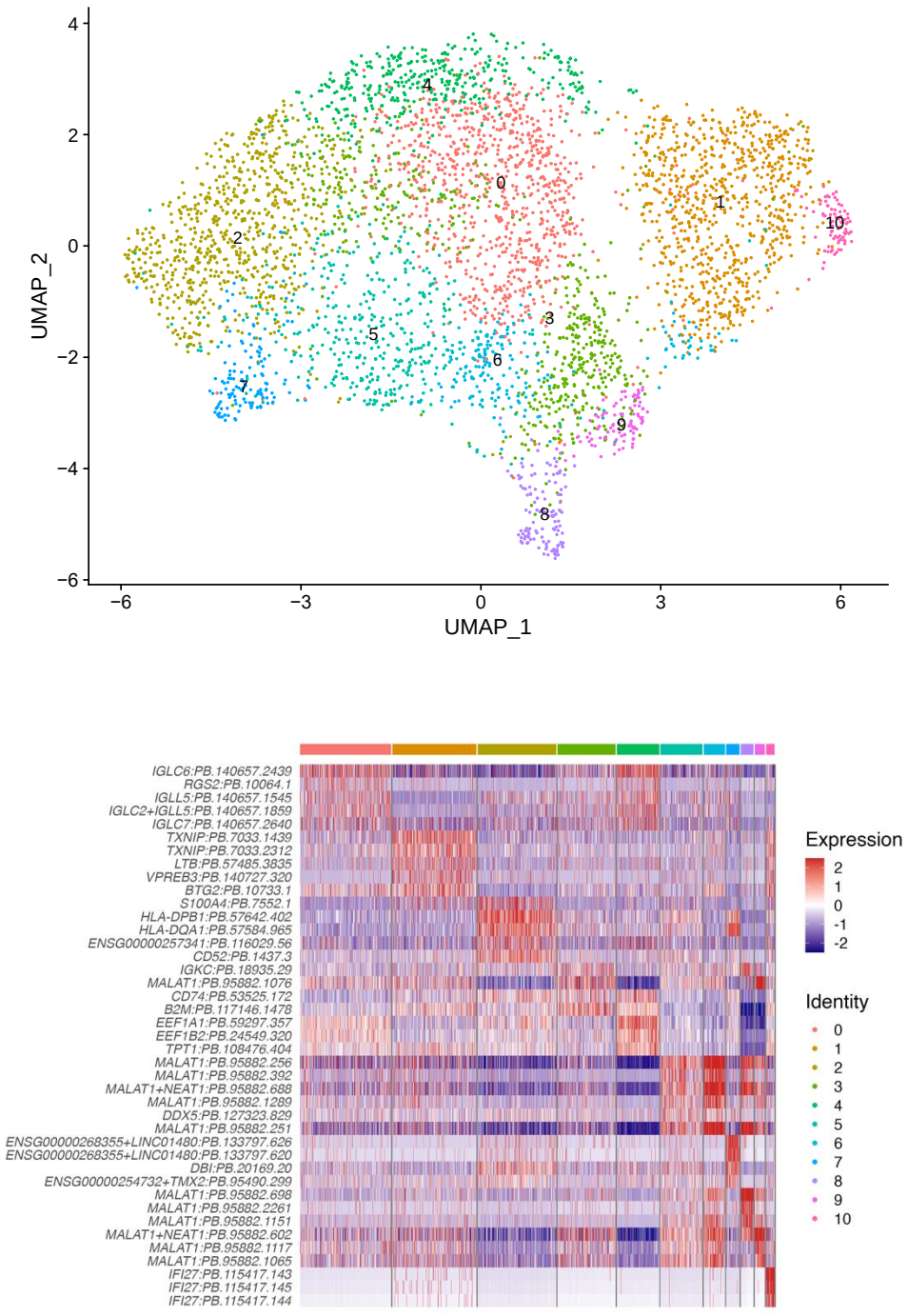

Expression
Identity



Supplemental Figure 7. Mapping subclone assignment to clustered cells enables the identification of phenotypically distinct subclones. For each patient, the subclone structure is shown, along with UMAPs of the relapse sample. Cells in the top UMAP are colored by subclone assignment, matching the subclone color scheme used in the subclone structure. The bottom UMAP is colored by the unsupervised cluster assignment. A heatmap is provided to indicate up-regulated genes in each unsupervised cluster.



Supplemental Figure 8. Comparison of the gene expression between the pre-treatment and relapse samples of Patient 1. A) Unsupervised clustering of the cells after merging the pre-treatment and relapse sample of Patient 1 (left UMAP). Overlap of all cells from the pre-treatment (blue) and relapse (red) samples is also shown (right UMAP). B) Heatmap showing the top up-regulated genes in each cluster identified with unsupervised clustering, as shown in panel A. C) Subclone assignment of cells from both samples showing the position of each subclonal population.



Supplemental Figure 9. Unsupervised clustering of cells by isoform expression in the post treatment sample of Patient 1. Heatmap shows the top up-regulated isoforms in each cluster.