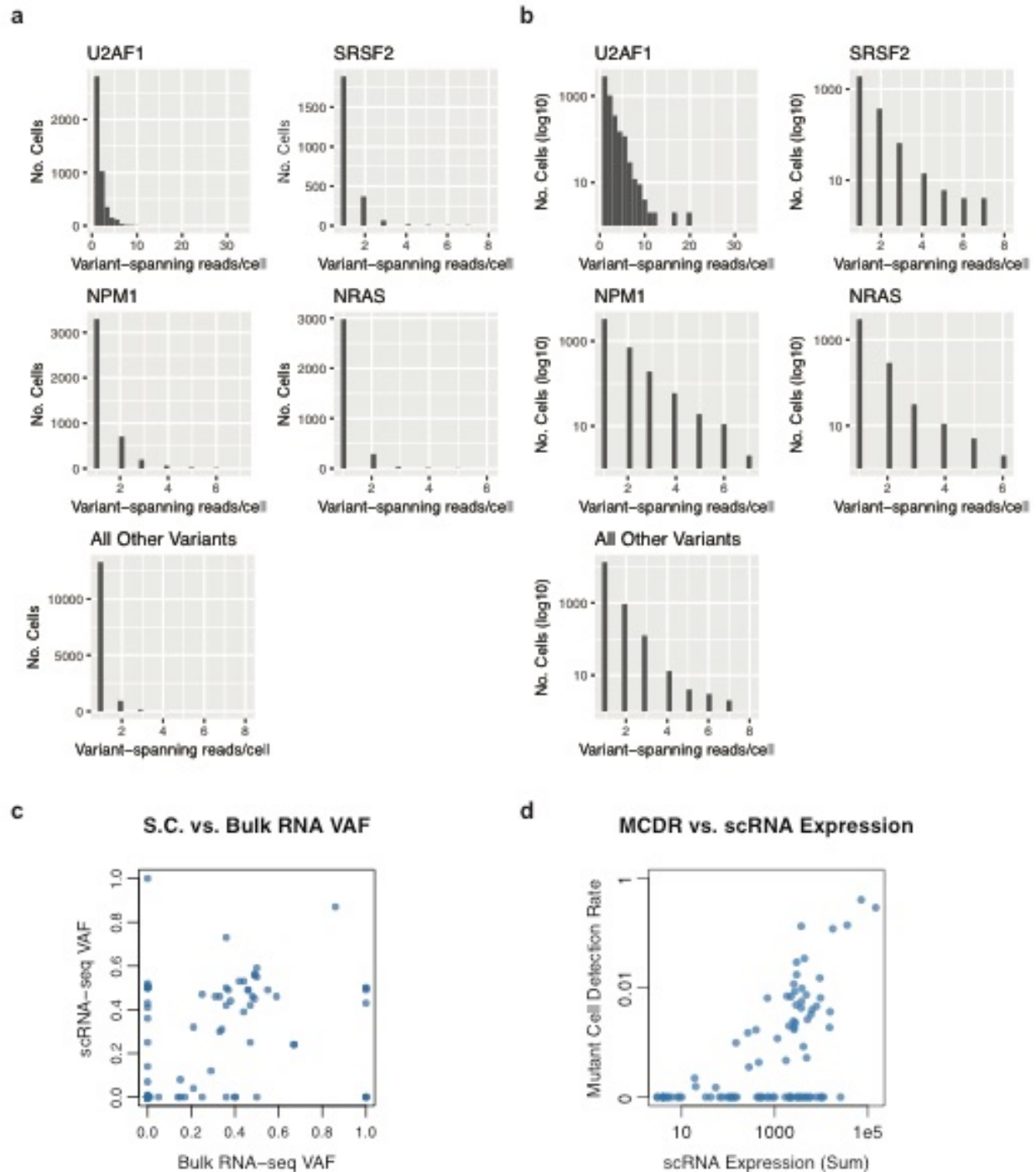


**Supplementary Information for:**

A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing

Petti et al.

## Supplementary Figures and Legends

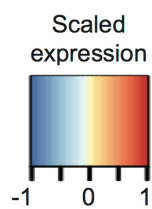
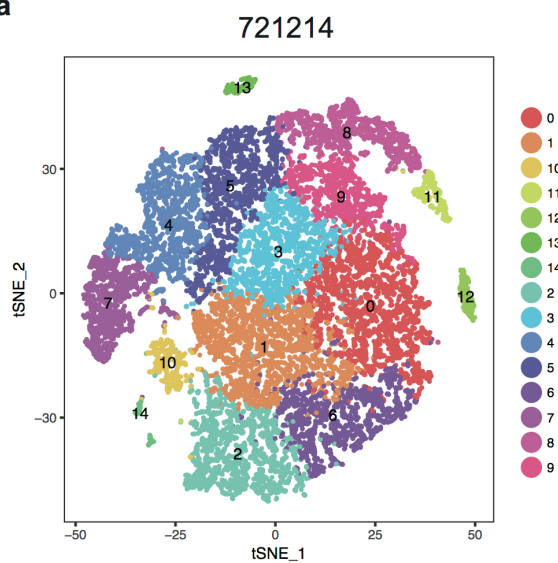


**Supplementary Figure 1. Additional performance metrics for single-cell variant detection.**

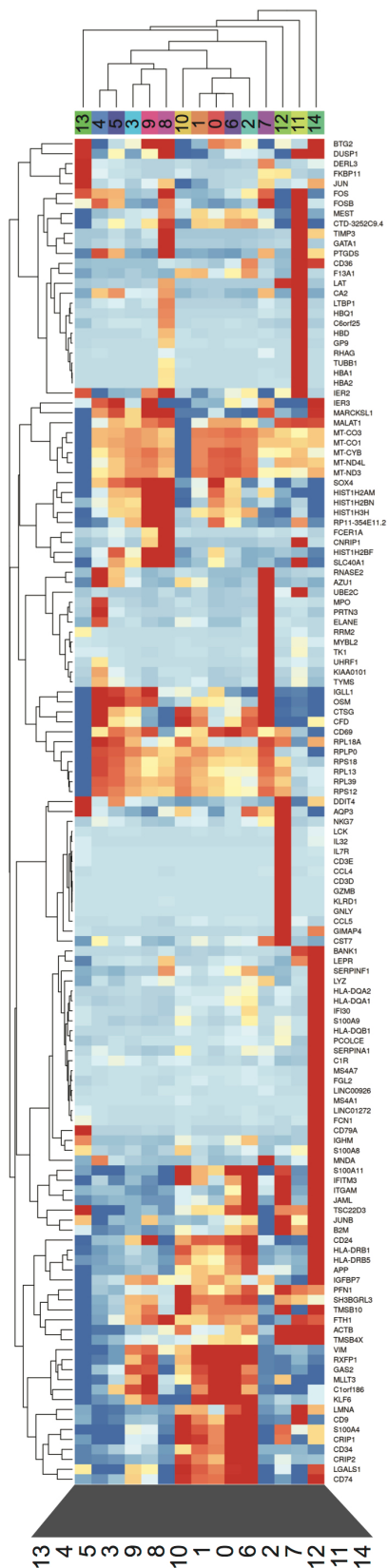
(a) Distribution of variant-spanning reads for mutations in the indicated gene(s). (b) Log-scale distribution of variant-spanning reads for mutations in the indicated gene(s). (c) Relationship

between single-cell and bulk RNA-seq VAF. (d) Mutant Cell Detection Rate as a function of gene expression level in the single-cell data.

**a**



**b**



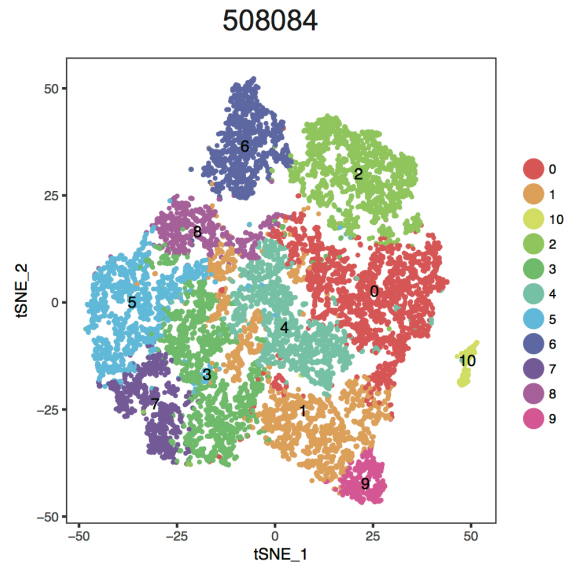
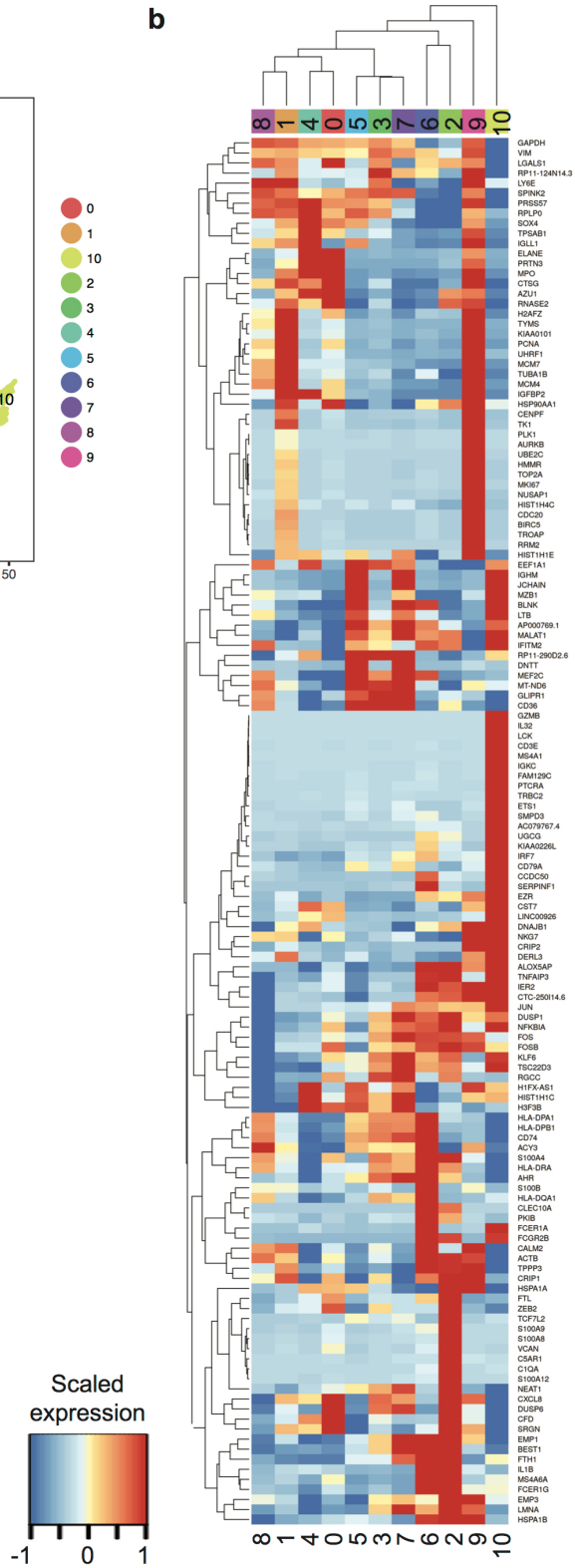
**Supplementary Figure 2. Clustering and overview of expression heterogeneity in 721214.**

(a) t-SNE plot of scRNA-seq data, cells colored according to graph-based cluster assignment; putative AML clusters circled. (b) Hierarchical clustering of the most heavily weighted genes in each principal component, averaged within graph-based clusters. Each column corresponds to a cluster in panel a.



**Supplementary Figure 3. Clustering and overview of expression heterogeneity in 548327.**

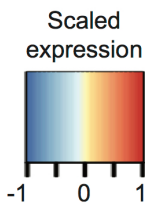
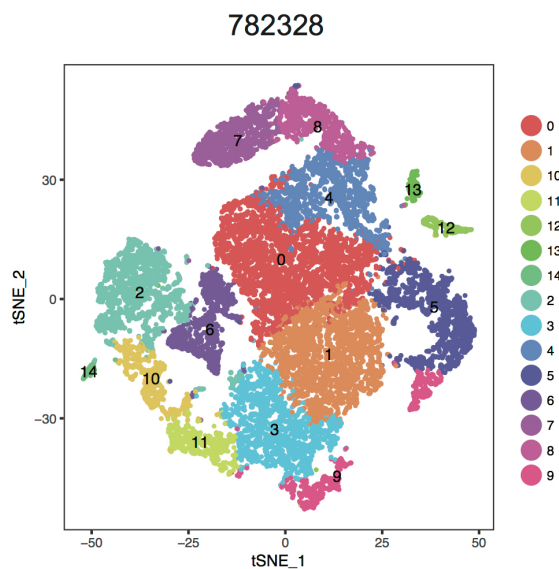
(a) t-SNE plot of scRNA-seq data, cells colored according to graph-based cluster assignment; putative AML clusters circled. (b) Hierarchical clustering of the most heavily weighted genes in each principal component, averaged within graph-based clusters.

**a****b**

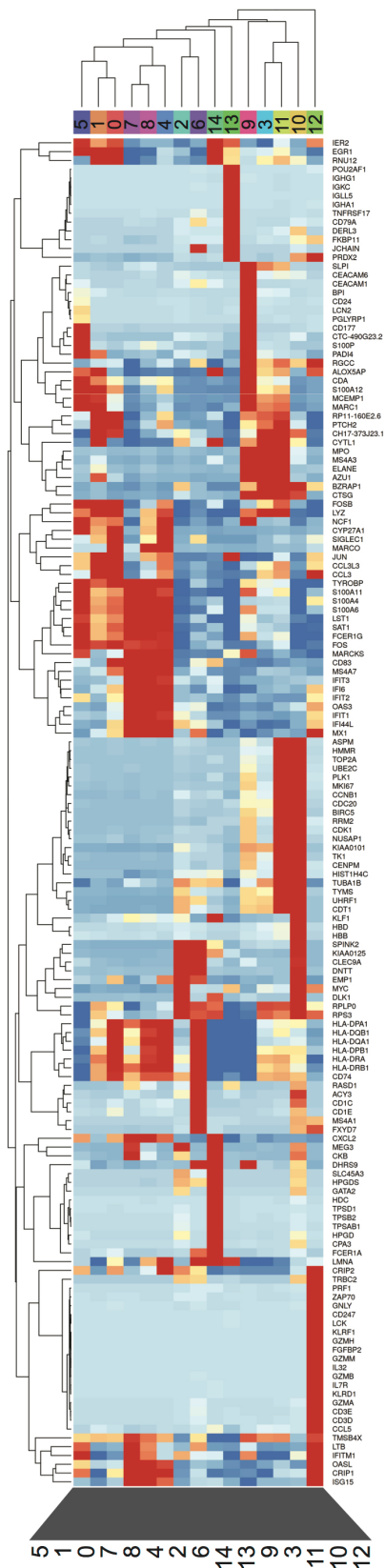
**Supplementary Figure 4. Clustering and overview of expression heterogeneity in 508084.**

(a) t-SNE plot of scRNA-seq data, cells colored according to graph-based cluster assignment; putative AML clusters circled. (b) Hierarchical clustering of the most heavily weighted genes in each principal component, averaged within graph-based clusters.

**a**

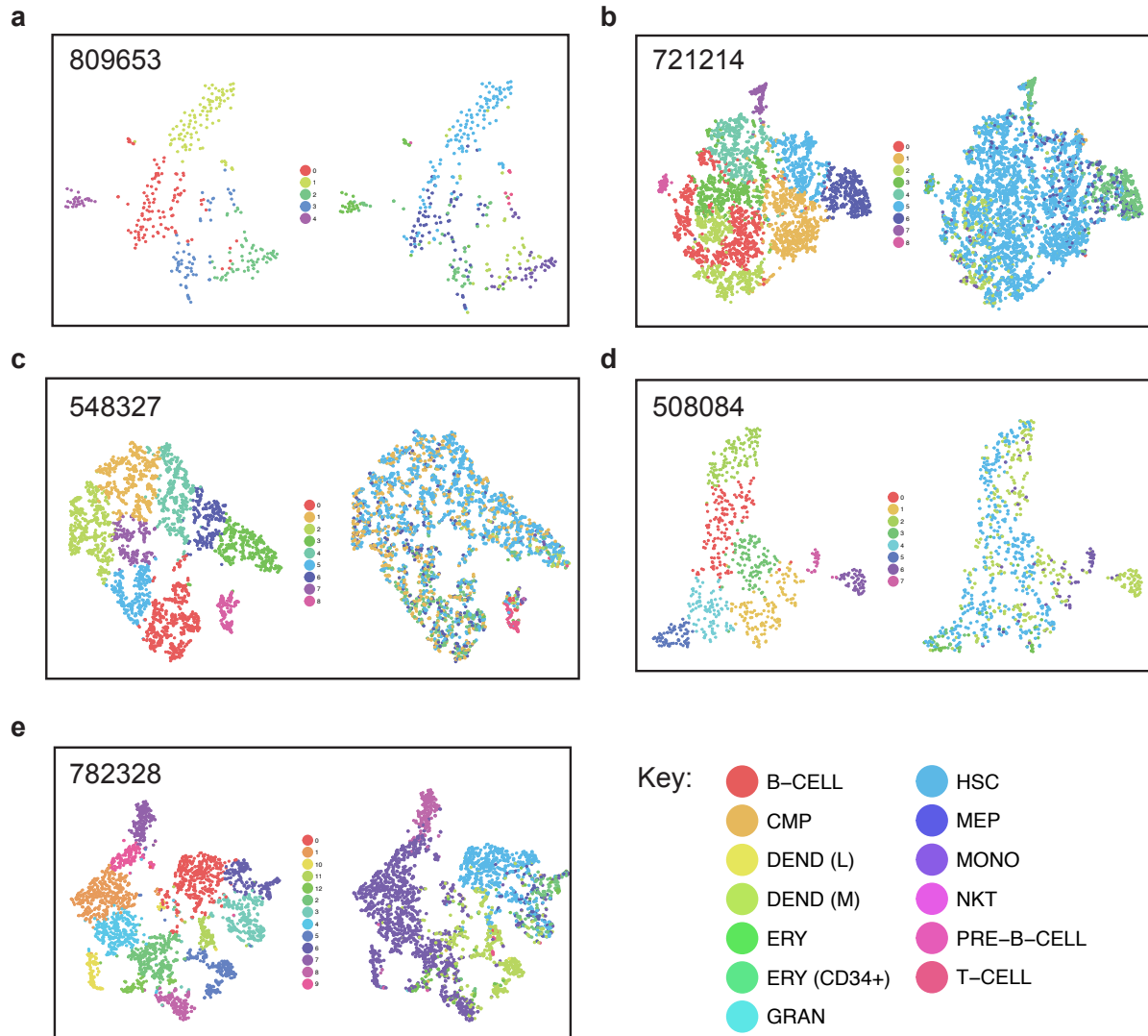


**b**

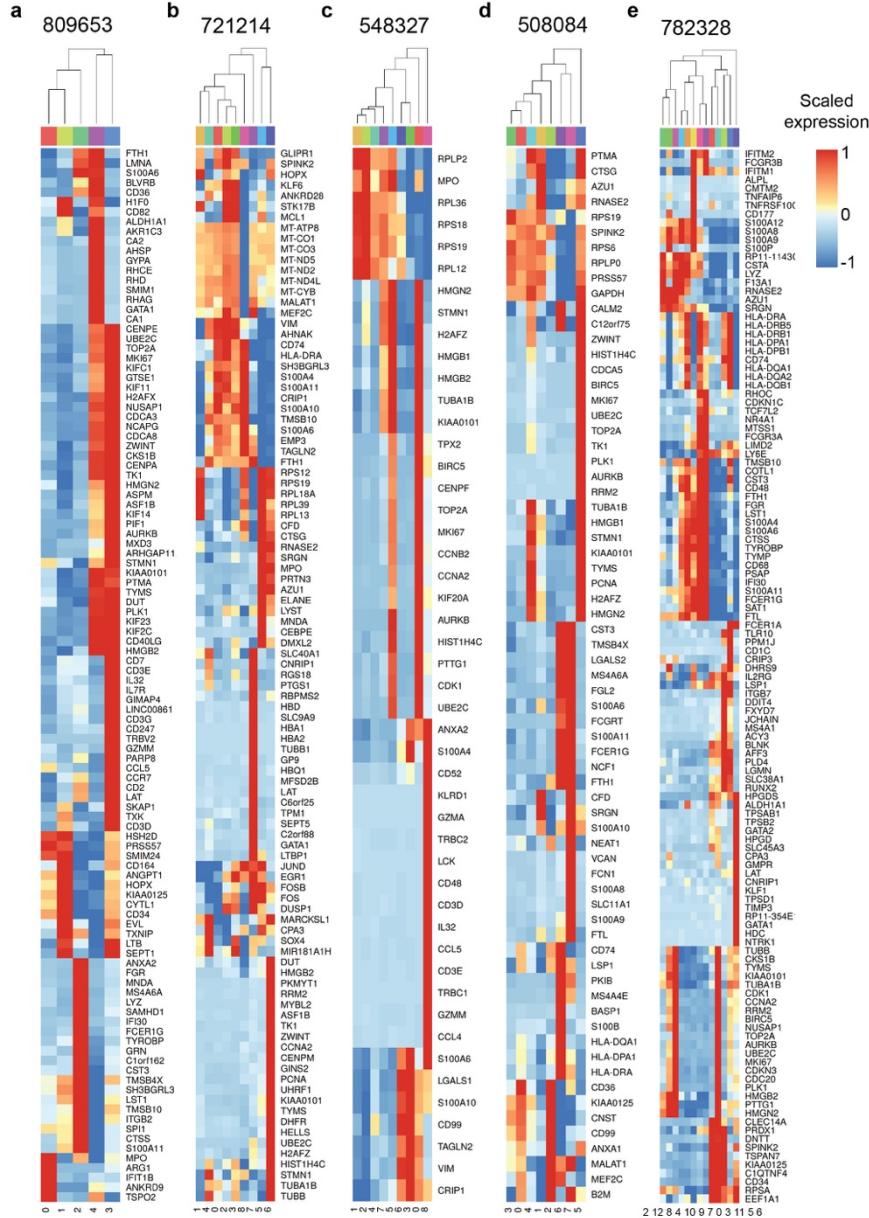


**Supplementary Figure 5. Clustering and overview of expression heterogeneity in 782328.**

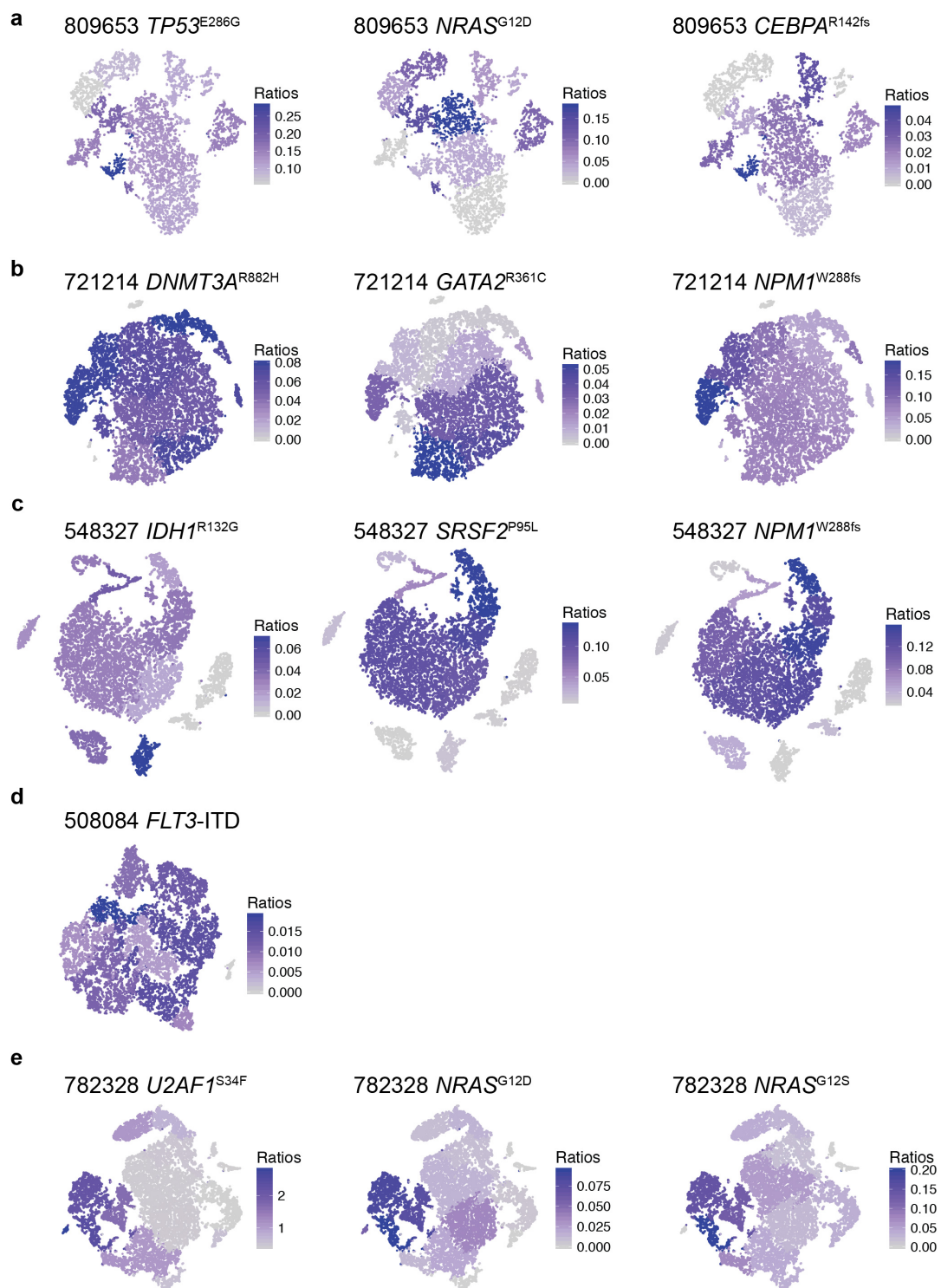
(a) t-SNE plot of scRNA-seq data, cells colored according to graph-based cluster assignment; putative AML clusters circled. (b) Hierarchical clustering of the most heavily weighted genes in each principal component, averaged within graph-based clusters.



**Supplementary Figure 6. Dimensionality reduction, clustering, and lineage inference of mutant cells.** (a) t-SNE projection of 809653 with cells colored according to graph-based cluster assignment (left) and inferred cell lineage (right); CMP = common myeloid progenitor, DEND (L) = lymphoid dendritic cell, DEND (M) = myeloid dendritic cell, ERY = erythrocyte, ERY (CD34+) = CD34+ erythrocyte, GRAN = granulocyte, HSC = hematopoietic stem cell, MEP = myeloid-erythroid progenitor, MONO = monocytic lineage, NKT = NK-T cell. (b) 721214. (c) 548327. (d) 508084. (e) 782328.



**Supplementary Figure 7. Cluster-averaged gene expression profiles of highly variable genes in mutant cells.** Each column corresponds to a graph-based cluster in Supp. Fig. 6. Genes were chosen based on their contributions to the top principal components (Methods). (a) 809653. (b) 721214. (c) 548327. (d) 508084. (e) 782328.



**Supplementary Figure 8. Clustered t-SNE plots colored according to expression-normalized mutation fraction in each cluster (selected genes).** “Ratio” is the ratio of mutation fraction to average expression of the mutated gene. (a) 809653, putative AML cells only. (b) 721214. (c) 548327. (d) 508084. (e) 782328.

**Supplementary Table 1. Representative functional enrichment of *GATA2*<sup>R361C</sup>-associated genes**

| GO ID      | GO Term                      | q-value  | Selected genes                                                       |
|------------|------------------------------|----------|----------------------------------------------------------------------|
| GO:0006955 | immune response              | 4.49E-10 | <i>CEBPB, HLA-DMB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2, CIITA</i> |
| GO:0012501 | programmed cell death        | 1.68E-06 | <i>HIF1A, HK1, CEBPB, KRAS, CAMK1D</i>                               |
| GO:0007159 | leukocyte cell-cell adhesion | 3.25E-06 | <i>CEBPB, HLA-DMB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2</i>        |
| GO:0005925 | focal adhesion               | 2.30E-11 | <i>AHNAK, FHL1, VIM, IRF2, GIT2, IQGAP1, PTPRC, KRAS</i>             |

**Supplementary Table 2. Representative enrichment of *VIM*-correlated, subclone-specific genes.**

| GO ID      | GO Term                             | q-value  | Selected genes                                                         |
|------------|-------------------------------------|----------|------------------------------------------------------------------------|
| GO:0006955 | immune response                     | 9.73E-06 | <i>WAS, CDC42, CEBPB</i>                                               |
| GO:0038094 | Fc-gamma receptor signaling pathway | 1.01E-04 | <i>SYK, FGR, RAC1, WAS, ARPC5, ARPC1B, ARPC2, CDC42, HCK</i>           |
| GO:0007010 | cytoskeleton organization           | 8.42E-05 | <i>VIM</i>                                                             |
| GO:0005925 | focal adhesion                      | 3.98E-09 | <i>ITGB1BP1, IQGAP1, PTPRC, AHNAK, VIM</i>                             |
| N/A        | WAS interactions                    | 6.14E-05 | <i>FGR, RAC1, WAS, GAS7, ARPC1B, ARPC2, RGS14, CDC42, HCK, PACSIN2</i> |

**Supplementary Table 3. Enrichment of genes that are more highly expressed in *RNF10* mutant-rich clusters.**

| GO ID      | GO Term                        | q-value  | Selected genes                                                                                             |
|------------|--------------------------------|----------|------------------------------------------------------------------------------------------------------------|
| GO:0032395 | MHC class II receptor activity | 6.25E-05 | <i>HLA-DPA1, HLA-DQA1, HLA-DQB1, HLA-DQB2</i>                                                              |
| GO:0098609 | cell-cell adhesion             | 9.63E-07 | <i>SOX4, FLT3, ICAM3, ITGA4, CD34, CD47, CD74</i>                                                          |
| GO:0070489 | T cell aggregation             | 1.08E-04 | <i>SOX4, FLT3, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DQB2, CD74, IL2RG, CDK6</i>                     |
| N/A        | VCAM1 interactions             | 1.38E-10 | <i>TUBB, STMN1, ITGA4, MSI2, HIST1H4C, RPLP0 (etc), RPS3 (etc), EEF1A1 (etc), EIF3L</i>                    |
| N/A        | MCM2 interactions              | 1.05E-05 | <i>SOX4, IMPDH2, TUBB, ITGA4, ADA, HIST1H4C, MACF1, CDK6, RPLP0 (etc), RPSA (etc), EIF3L, EEF1A1 (etc)</i> |

|     |                   |          |                                                                    |
|-----|-------------------|----------|--------------------------------------------------------------------|
| N/A | NPM1 interactions | 1.40E-05 | <i>IMPDH2, CPSF6, RPLP0 (etc), RPS3 (etc), EEF1A1 (etc), EIF3L</i> |
|-----|-------------------|----------|--------------------------------------------------------------------|

**Supplementary Table 4. Enrichment of genes that are more highly expressed in *CEBPA* mutant-rich clusters.**

| GO ID      | GO Term                            | q-value  | Selected genes                                                                                                                     |
|------------|------------------------------------|----------|------------------------------------------------------------------------------------------------------------------------------------|
| GO:0003735 | structural constituent of ribosome | 9.95E-43 | <i>RPL22, RPL23A, RPL26, etc.</i>                                                                                                  |
| GO:0042110 | T cell activation                  | 1.25E-03 | <i>HLA-DPA1, HLA-DPB1, HLA-DQB1, HLA-DRA, XBP1, CD74, TNFRSF4, FLT3, FOXP1, MYB, SOX4, KIT, RAC2, ZEB1, GSN, CDK6, RPL22, RPS6</i> |
| GO:0008134 | transcription factor binding       | 5.84E-03 | <i>MYC, GATA2, NPM1, TCF4, ZEB1, RUNX1, HLA-DQB1, XBP1, MEF2C, HOXA7, CD34, HDAC1</i>                                              |
| N/A        | MYC binding sites                  | 9.22E-03 | <i>FLT3, MYB, DNMT3A, NPM1, TCF4, SET, STMN1, HHEX, MAX, HOXA7, HOXA9, RPL22 (etc), RPS19 (etc), EIF3E (etc)</i>                   |