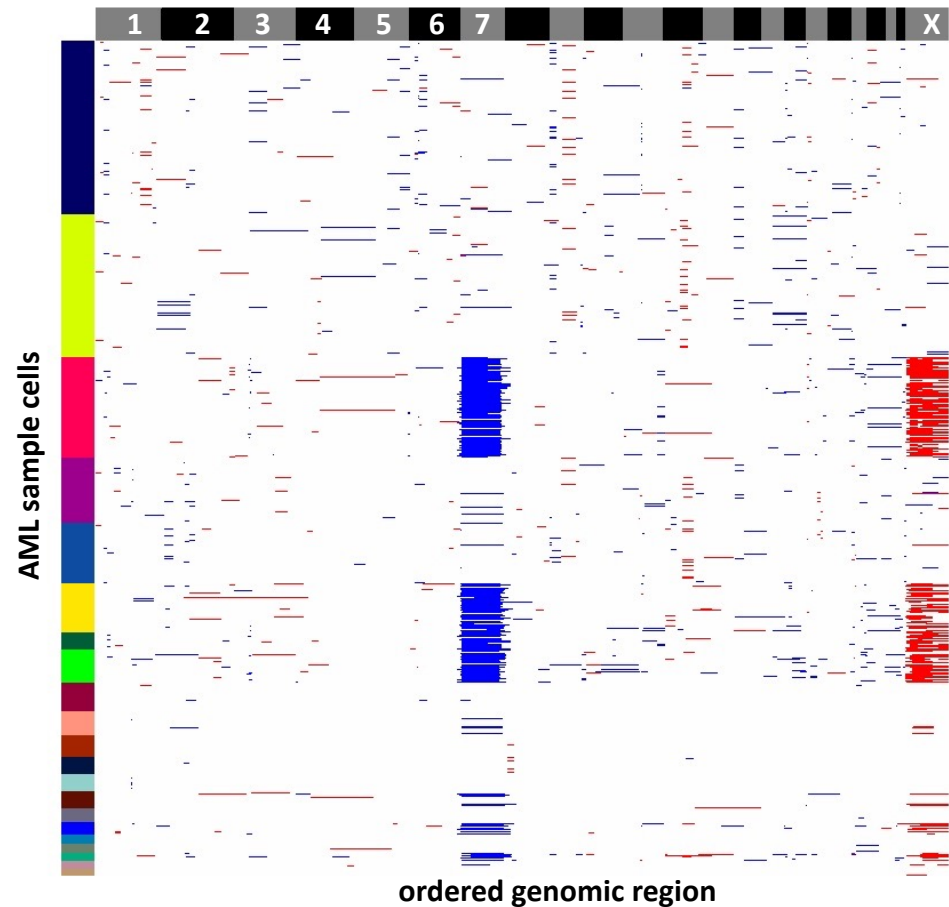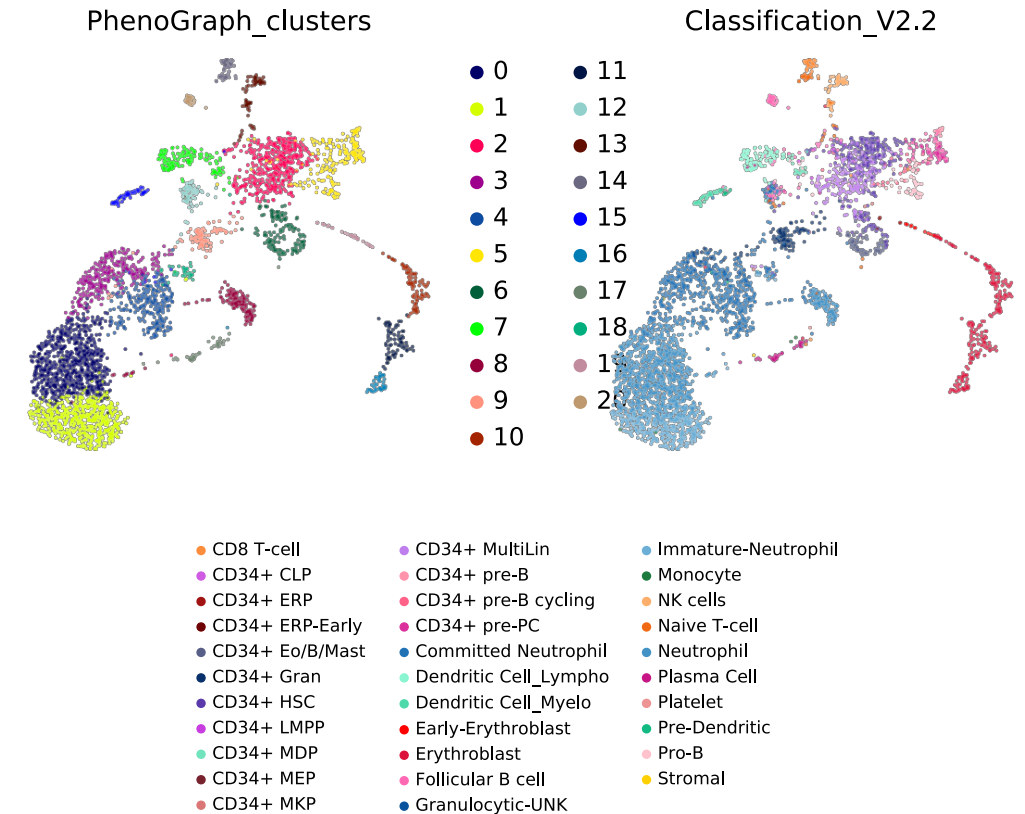We estimated copy number profiles from scRNA-seq data for AML samples, using a matched cell type approach with HCA as a reference to infer the copy number baseline. This approach allowed us to identify clusters with cytogenetic abnormalities. For each sample, we generated one clustered heatmap and two UMAP representations.

The clustered heatmap represents the scRNA-seq copy number profiles. Each row corresponds to a cell, color-coded and ordered according to its Phenograph cluster, and each column corresponds to an ordered genomic region. The black and gray boxes indicate the chromosomes (left: chromosome 1 -> right: chromosome X). The blue signal indicates a loss, while the red signal indicates a gain.

PhenoGraph_clusters

Classification_V2.2

0	11
1	12
2	13
3	14
4	15
5	16
6	17
7	18
8	19
9	20
10	

CD8 T-cell — CD34+ MultiLin — Immature-Neutrophil
CD34+ CLP — CD34+ pre-B — Monocyte
CD34+ ERP — CD34+ pre-B cycling — NK cells
CD34+ ERP-Early — CD34+ pre-PC — Naive T-cell
CD34+ Eo/B/Mast — Committed Neutrophil — Neutrophil
CD34+ Gran — Dendritic Cell_Lympho — Plasma Cell
CD34+ HSC — Dendritic Cell_Myelo — Platelet
CD34+ LMPP — Early-Erythroblast — Pre-Dendritic
CD34+ MDP — Erythroblast — Pro-B
CD34+ MEP — Follicular B cell — Stromal
CD34+ MKP — Granulocytic-UNK

The first UMAP (Left) corresponds to the sample's scRNA-seq data, colored by Phenograph cluster, while the second UMAP (Right) corresponds to the cell type annotation V2.2 predicted using our random forest classifier trained on HCA data.