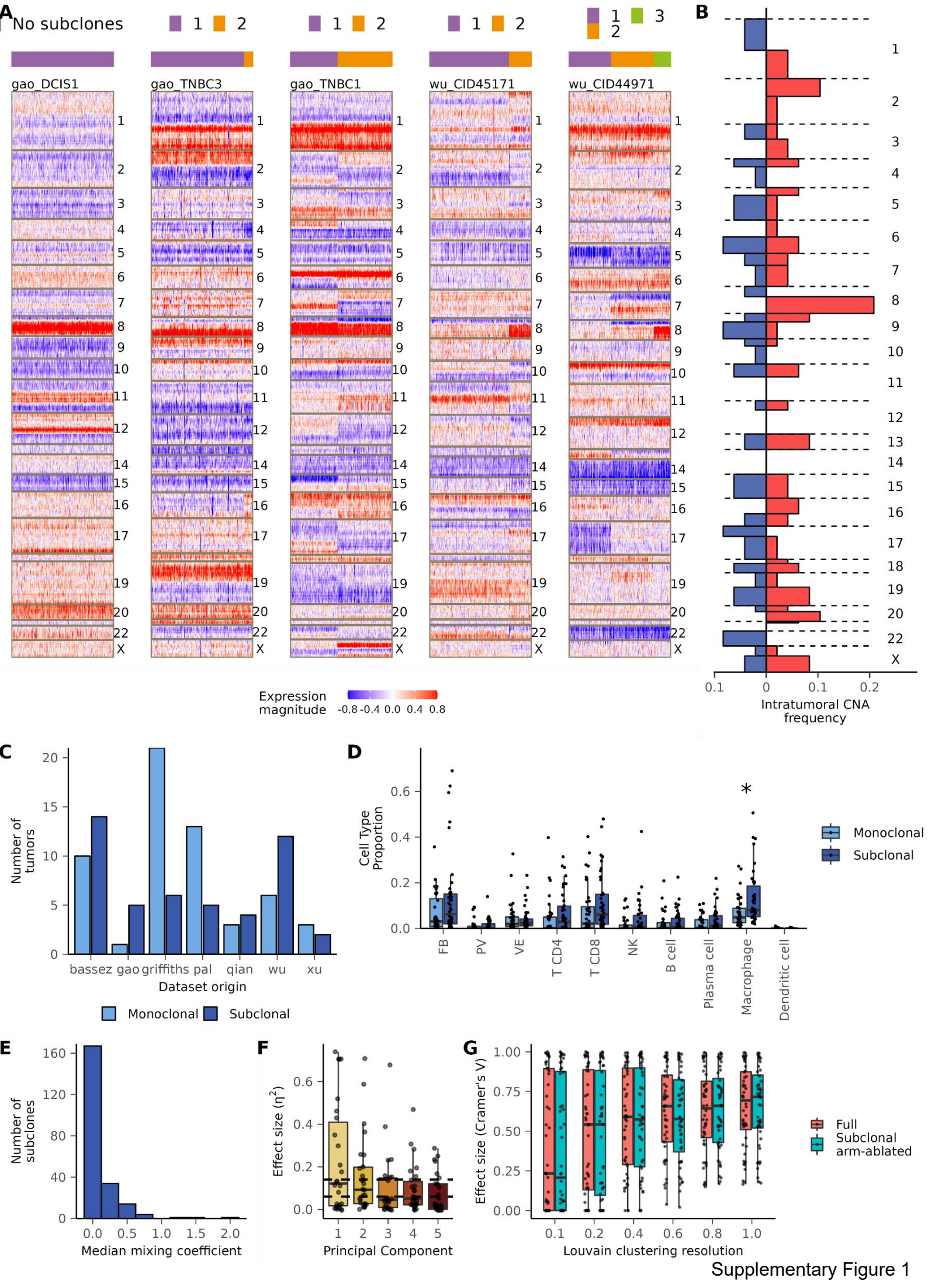
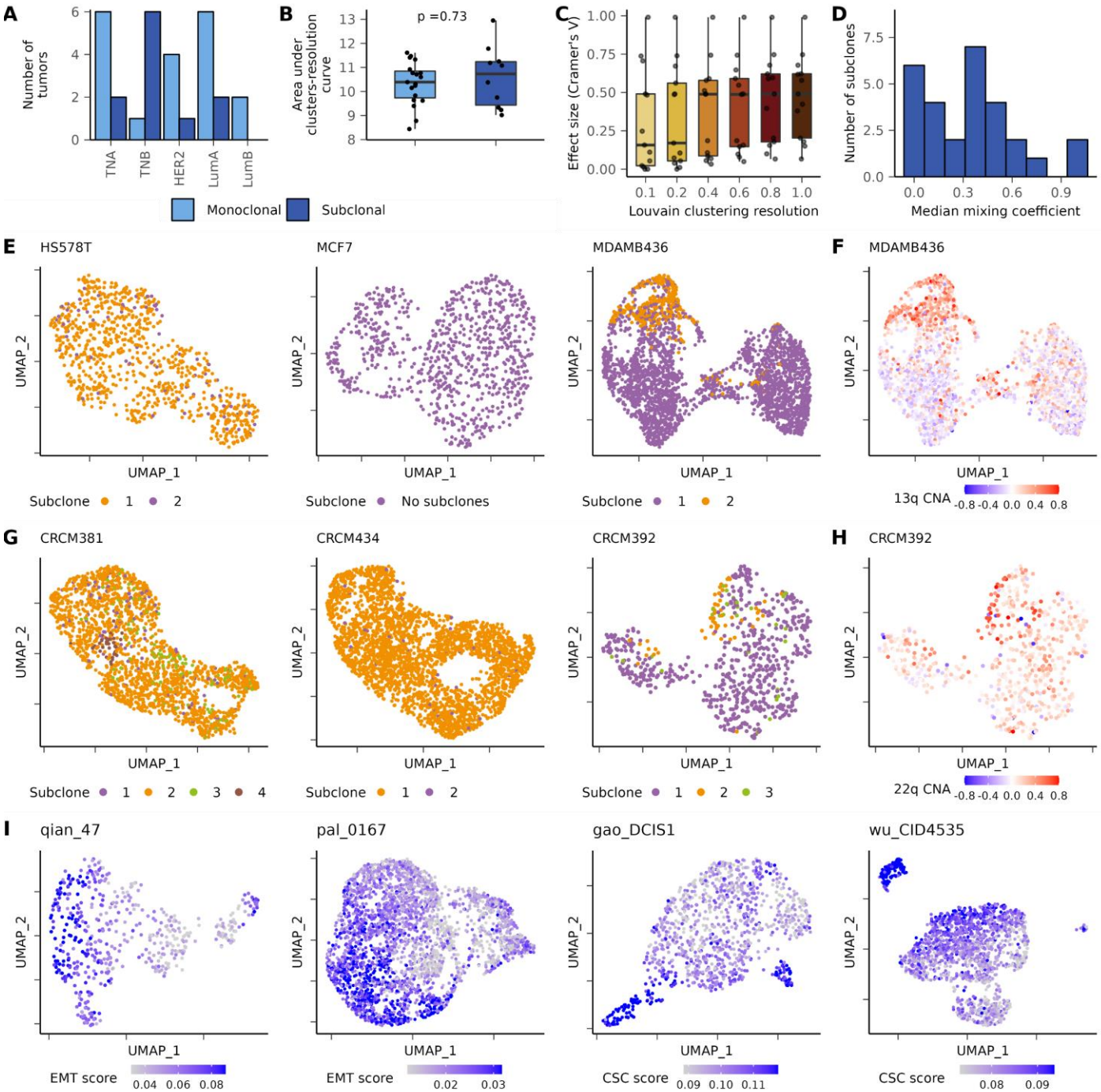
SUPPLEMENTARY FIGURES



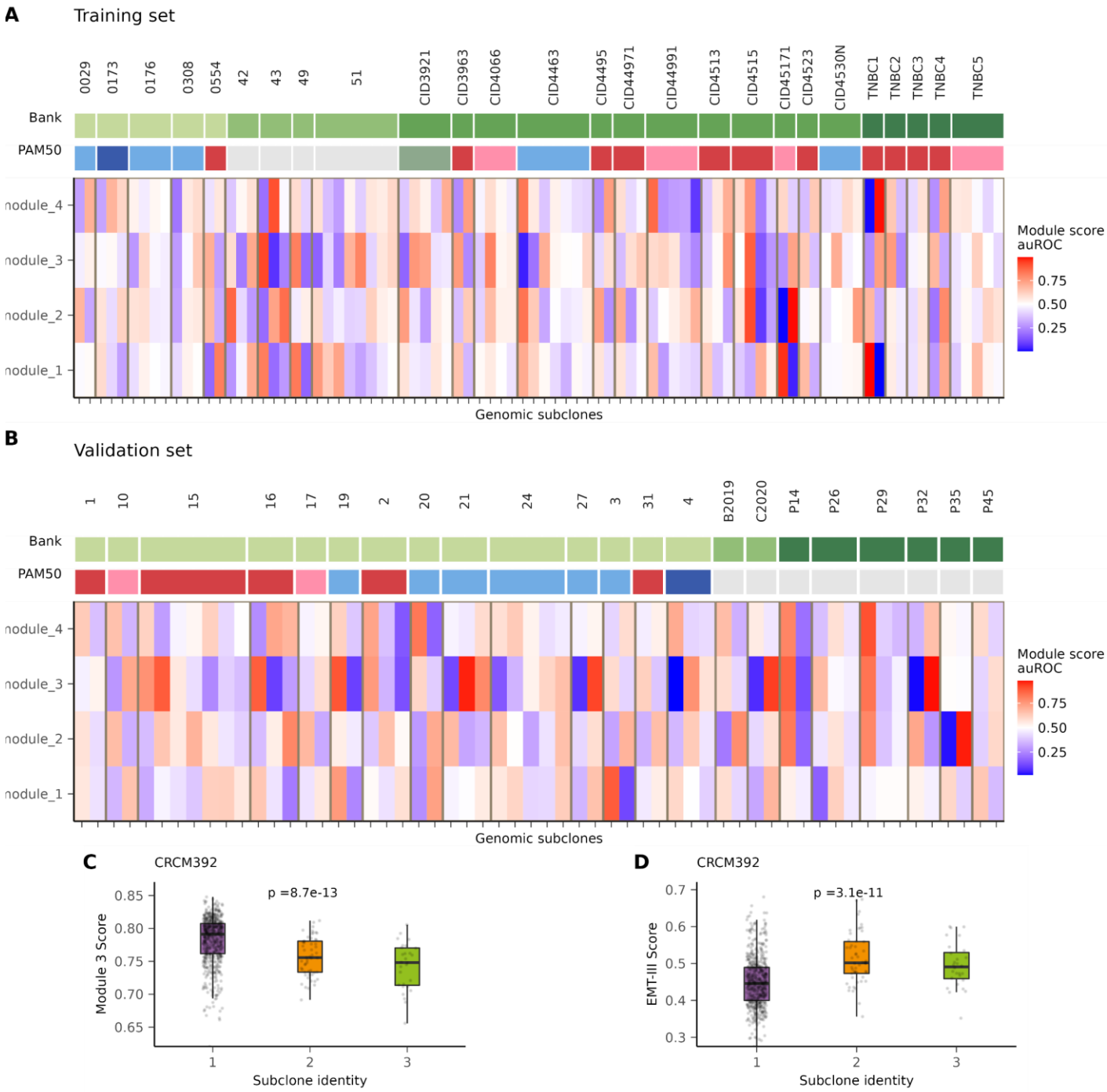
Supplementary Figure 1

Supplementary Figure 1. Examples of subclonal architectures in primary breast tumors and association with cell types from the microenvironment A. Inferred CNA profiles of DCIS1, TNBC3 and TNBC1 from Gao et al. 2021, CID45171 and CID44971 from Wu et al. 2021, with subclone annotation. **B.** CNA frequencies associated with subclonal architecture in primary breast tumors. **C.** Number of monoclonal and subclonal tumors according to dataset origin. **D.** Proportion of normal cell types from the microenvironment for monoclonal and subclonal tumors. **E.** Median mixing coefficient by subclone. **F.** Effect size of genomic subclonal identity on PC embeddings in the first 5 PCs for each tumor. η^2>6% is considered as a moderate effect. η^2>14% is considered as a large effect. **G.** Effect size of genomic subclone identity on transcriptomic clusters belonging as assessed by Cramer's V for Louvain clustering resolution from 0.1 to 1.0. Cramer's V is computed for full matrices and matrices without the genes located on the chromosome arms responsible for genomic differences within the tumor. In **D** and **F**, box plots depict first and third quartiles as lower and upper bounds, respectively. The whiskers represent 1.5× the IQR and the center depicts the median. Wilcoxon rank-sum test with Bonferroni correction, *P<0.05 ; **P<0.01 (**D**) | Kruskal-Wallis rank-sum test (**F**)



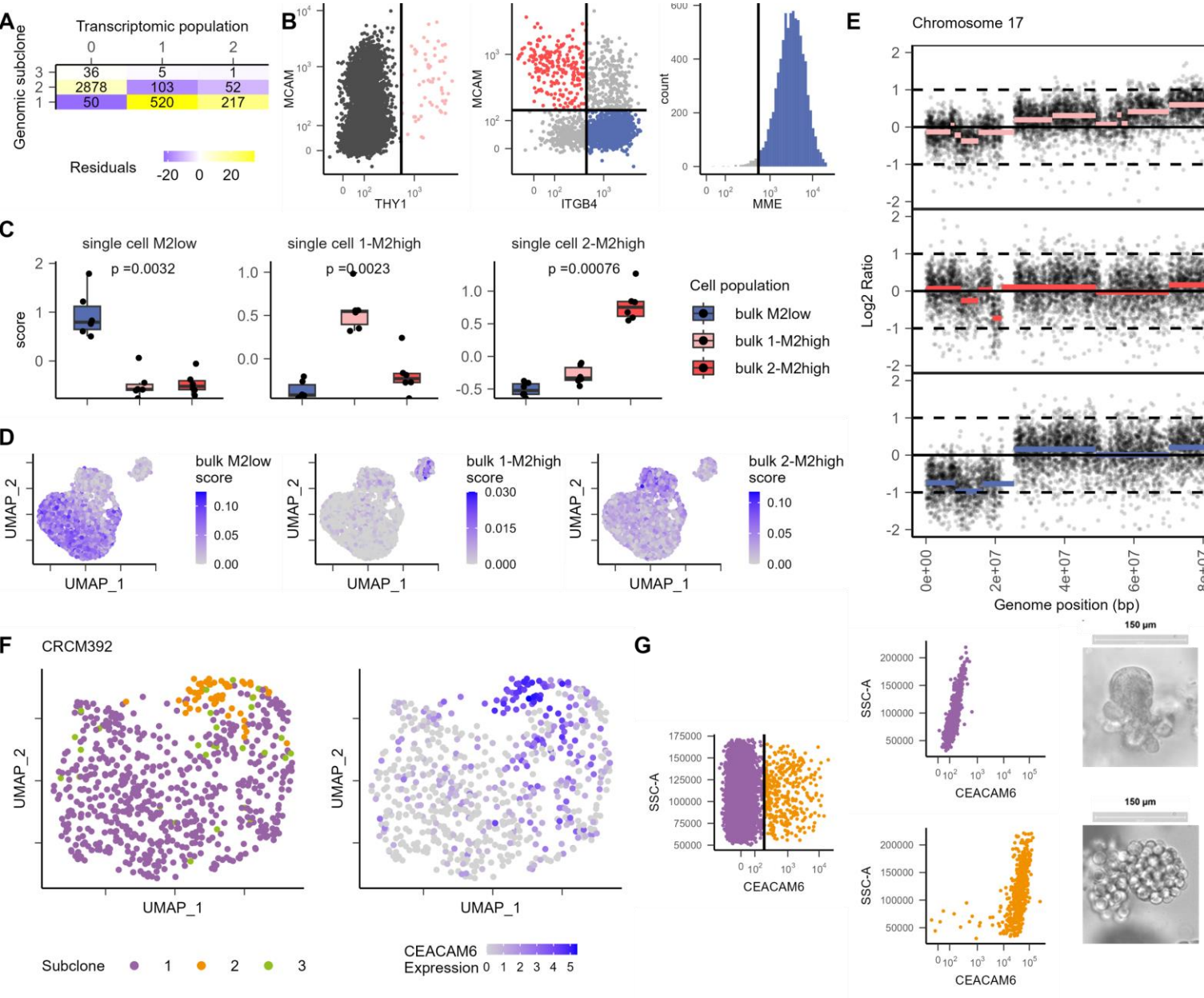
Supplementary Figure 2

Supplementary Figure 2. Intratumoral heterogeneity in cancer cell models and monoclonal tumors A. Number of monoclonal and subclonal breast cancer cell lines according to breast cancer subtype as previously annotated[51]. **B.** Area under clusters-resolution curve (AUC) for monoclonal and subclonal cell lines. **C.** Effect size of genomic subclone identity on transcriptomic clusters belonging for subclonal cell lines as assessed by Cramer's V for Louvain clustering resolution from 0.1 to 1.0. **D.** Median mixing coefficient by subclone for subclonal cell lines. **D.** UMAP plots of HS578T, MCF7 and MDAMB436 cell lines. Cells are colored according to subclone annotation. **F.** UMAP plot of MDAMB436 cell line. Cells are colored according to inferred 13q CNA level. **G.** UMAP plots of CRCM381, CRCM434 and CRCM392 PDXs. Cells are colored according to subclone annotation. **H.** UMAP plot of CRCM392 PDX. Cells are colored according to inferred 22q CNA level. **I.** Representative examples of phenotypic diversity in monoclonal tumors:UMAP plots of sample 47 from Qian et al. 2020, sample 0167 from Pal et al. 2020, DCIS1 from Gao et al. 2021 and CID4535 from Wu et al. 2021. Cells are colored according to the EMT score[55] and CSC score[52]. In **B** and **C**, box plots depict first and third quartiles as lower and upper bounds, respectively. The whiskers represent 1.5× the IQR and the center depicts the median. Wilcoxon rank-sum test (**B** and **C**).



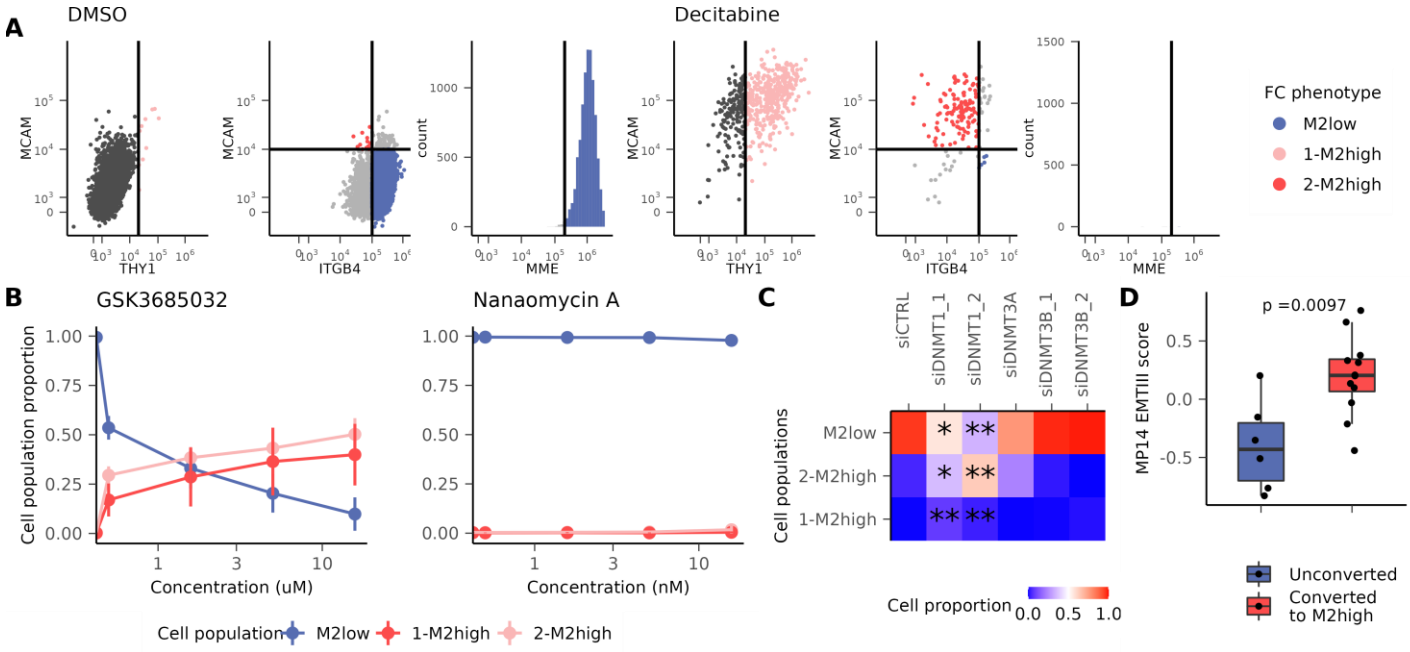
Supplementary Figure 3

Supplementary Figure 3. Regulatory modules are recurrently differentially activated in primary tumors and breast cancer models. A-B. Heatmap of auROC analysis of module activity between subclones for the training set (**A**, Pal et al., Qian et al., Wu et al. and Gao et al. datasets are represented from left to right) and the validation set (**B**, Bassez et al., Xu et al. and Griffiths et al. are represented from left to right). PAM50 Basal, HER2, Luminal A and B are colored in red, green, light blue and blue, respectively. **C.** Module 3 activity according to subclone identity for CRCM392. **D.** Number of significantly differentially activated MPs by PDX tumor. Each bar indicates the frequency of tumors exhibiting a specific MP count. **F.** EMT-III score (MP14 activity) according to subclone identity for CRCM392. In **C** and **E**, box plots depict first and third quartiles as lower and upper bounds, respectively. The whiskers represent 1.5× the IQR and the center depicts the median. Kruskal-Wallis rank-sum test (**C** and **E**).



Supplementary Figure 4

Supplementary 4. Identification of cell surface marker combination to FACS-sort and characterize cell subpopulations in SUM159 and CRCM392. A. The number of cells and relative enrichment (Pearsons' residual, calculated as $(observed\ number\ of\ cells - expected\ number\ of\ cells)/\sqrt{expected\ number\ of\ cells}$, color bar) from each transcriptomic cluster (columns) in each subclone (rows) in SUM159 cell line. **B.** Flow cytometry plots of unsorted SUM159 cells, stained with antibodies against THY1, MCAM, ITGB4 and MME. **C.** $M2^{low}$, 1-$M2^{high}$ and 2-$M2^{high}$ signature score of FACS-sorted populations bulk RNA-seq. The signatures represented were derived from SUM159 scRNA-seq analysis. **D.** UMAP plots of SUM159 scRNA-seq data with cell-cycle regression. Cells are colored according to $M2^{low}$, 1-$M2^{high}$ and 2-$M2^{high}$ signature scores. The signatures represented were derived from FACS-sorted populations from SUM159 using bulk RNA-seq. **E.** aCGH chromosome 17 analysis on FACS-sorted $M2^{low}$, 1-$M2^{high}$ and 2-$M2^{high}$ SUM159 populations. Data segmentation is colored according to population. **F.** UMAP plots of CRCM392 with cell-cycle regression. Cells are colored according to genomic subclone and gene expression level of CEACAM6. **G.** Flow cytometry plot of unsorted CRCM392 (left). Flow cytometry plots of FACS-sorted CRCM392 CEACAM6- and CEACAM6+ populations (middle). Cells are stained with antibodies against CEACAM6. Brightfield representative images of CRCM392 CEACAM6- and CEACAM6+ PDxOs after 14 days of culture (right). In **B**, box plots depict first and third quartiles as lower and upper bounds, respectively. The whiskers represent 1.5× the IQR and the center depicts the median. Kruskal-Wallis rank-sum test (**B**).



Supplementary Figure 5

Supplementary 5. DNMT1 is necessary to maintain M2^{low} cell state. A. Representative flow cytometry plots of M2^{low} cells untreated (left) or treated with Decitabine (right). Cells are stained with antibodies against THY1, MCAM, ITGB4 and MME. **B.** M2^{low}, 1-M2^{high} and 2-M2^{high} phenotype cell proportions assessed by flow cytometry, according to DNMT inhibitor concentration. Initially M2^{low} cells were treated with GSK3685032 (DNMT1 inhibitor) and Nanaomycin A (DNMT3B inhibitor) at multiple concentrations. 15.8 µM and 15.8 nM are maximum viable concentrations for GSK3685032 and Nanaomycin A, respectively. **C.** M2^{low}, 1-M2^{high} and 2-M2^{high} phenotype cell proportions assessed by flow cytometry, according to siRNA targeting DNMTs. Initially M2^{low} cells were lipofected with Negative Control, two siRNA targeting DNMT1, one siRNA targeting DNMT3A and two siRNA targeting DNMT3B. **D.** MP14 activity of FACS-sorted 5-azacytidine treated M2^{low} cells, unconverted and converted towards M2^{high} phenotype using bulk RNA-seq. In **C**, each cell population proportion is compared to its equivalent lipofected with siCTRL. In **D**, box plots depict first and third quartiles as lower and upper bounds, respectively. The whiskers represent 1.5× the IQR and the center depicts the median. Wilcoxon rank-sum test, *P<0.05 ; **P<0.01 (**C** and **D**).