

## Description of Additional Supplementary Files

**File Name:** Supplementary Data 1

**Description:** Clone detection in vitro and in vivo from bulk GBC sequencing. Count statistics and classification of GBC sequences.

**File Name:** Supplementary Data 2

**Description:** Differential expression analysis at baseline (scRNA-Seq). DEA between each cluster and its complement at baseline in the same experiment (n = 2 each). P-values are computed with a generalised linear regression model (MAST method) and adjusted with Bonferroni correction.

**File Name:** Supplementary Data 3

**Description:** Functional annotation of differentially expressed genes at baseline. DEA between each cluster and its complement. REACTOME database is used for functional annotation. P-values are computed with a hypergeometric model and adjusted with Benjamini-Hochberg correction.

**File Name:** Supplementary Data 4

**Description:** Analysis of Pal et al.<sup>2</sup> dataset. Epithelial cells from four TNBC samples are reported.

**File Name:** Supplementary Data 5

**Description:** Cell annotation for scRNA-Seq of the MDA-MB-231 TGL sample.

**File Name:** Supplementary Data 6

**Description:** Differential expression analysis of MDA-MB-231 clusters at baseline (scRNA-Seq). DEA between each cluster and its complement (n = 1). P-values are computed with a generalised linear regression model (MAST method) and adjusted with Bonferroni correction.

**File Name:** Supplementary Data 7

**Description:** Cell annotation for scRNA-Seq and Multiome datasets (RNA-Seq library only).

**File Name:** Supplementary Data 8

**Description:** Differential expression analysis of SUM159PT derived tumours compared to parental cells. Expression level in TPM,  $\log_2(\text{FC})$  between TM4SF1-high and bulk. P-values are computed with a negative-binomial model (DESeq2 method) and adjusted with Benjamini-Hochberg correction.

**File Name:** Supplementary Data 9

**Description:** Differential expression analysis of TM4SF1-high cells compared to bulk population. Expression level in TPM,  $\log_2(\text{FC})$  between TM4SF1-high and bulk. P-values are computed with a negative binomial generalized linear models, with F-tests (edgeR-quasi method) and FDR are adjusted with Benjamini-Hochberg correction.

**File Name:** Supplementary Data 10

**Description:** Differential expression analysis at baseline (Multiome). DEA between each cluster and its complement at baseline in the same experiment (n = 2 each). P-values are

computed with a generalised linear regression model (MAST method) and adjusted with Bonferroni correction.

**File Name:** Supplementary Data 11

**Description:** CisTopic analysis and ATAC module detection. CisTopic is run for the two Multiome replicates separately. IDR is computed for each cross-replicate topic pair.

**File Name:** Supplementary Data 12

**Description:** Spearman's rho between module AUC (Module 1) and gene expression across cells, for the two replicates separately.

**File Name:** Supplementary Data 13

**Description:** Transcription factor enrichment by ChEA3. Meanrank score and overlapping genes using either Module 1 proximal genes (GREAT) or Module 1 correlated genes (Spearman's rho) as query.

**File Name:** Supplementary Data 14

**Description:** Copy number variants called on Whole Exome Sequencing data using CNVkit. Paclitaxel treated samples at day 15 (D15A-B-C) or day 90 are compared with parental cells. The table reports: "log2": weighted mean of log2 ratios of all the gene's bins; "depth": weighted mean of un-normalized read depths across all this gene's bins; "weight": sum of this gene's bins' weights; "cn": Integer copy number, merging adjacent with the same called value; "p\_ttest": z-test pvalues at each bin adjusted with Benjamini-Hochberg correction.

**File Name:** Supplementary Data 15

**Description:** Clone detection from single-cell RNA-Seq. Count statistics and classification of clones.

**File Name:** Supplementary Data 16

**Description:** Differential expression analysis after treatment. DEA between lineage 1 and lineage 2 after treatment in the same experiment (all time points). DEA between each cluster and its complement after treatment in exp2. P-values are computed with a generalised linear regression model (MAST method) and adjusted with Bonferroni correction.

**File Name:** Supplementary Data 17

**Description:** Functional annotation of differentially expressed genes after treatment. DEA between each cluster and its complement. REACTOME database is used for functional annotation. P-values are computed with a hypergeometric model and adjusted with Benjamini-Hochberg correction.

**File Name:** Supplementary Data 18

**Description:** RT-qPCR oligos, sgRNAs sequences, primers pair for GBC PCR on gDNA, primers pair for GBC PCR (DROP-seq), MULTI-Seq barcodes.