Supplemental information

Emergence of hybrid states of stem-like cancer cells correlates with poor prognosis in oral cancer

Kavya Vipparthi, Kishore Hari, Priyanka Chakraborty, Subhashis Ghosh, Ankit Kumar Patel, Arnab Ghosh, Nidhan Kumar Biswas, Rajeev Sharan, Pattatheyil Arun, Mohit Kumar Jolly, and Sandeep Singh
Supplementary Information:

Figure S1: Presence of ALD\(\text{Hi}\) cells in both CD24\(\text{Lo}\) and CD24\(\text{Hi}\) phenotypes in in-vitro cell lines, Related to Figure 1: Representative FACS dot plots of A) SCC070 B) SCC029B C) GBC035 D) SCC032 cell lines CD24/CD44 staining (Left bottom) with respect to Isotype control (Left top). ALDEFLUOR (ALDH) phenotype of CD24\(\text{Hi}/\text{CD44}\) subpopulation (Right top) and CD24\(\text{Lo}/\text{CD44}\) subpopulation (Right bottom) in presence or absence of DEAB.
Figure S2: Strategy for combined identification of ALDH cells and CD24\textsubscript{Low} and CD24\textsubscript{High} phenotypes in primary oral patient tissues, Related to Figure 1: A) Gating strategy of ALDEFLUOR assay combined with CD24/CD44 staining in a freshly resected and dissociated patient tumor. B) Dot plot showing all events. C, D) Dot plots showing single cell selection upon doublets elimination from analysis. E) Selection of propidium iodide negative live cells. F) Selection of hematopoietic lineage negative cells using Lineage antibody cocktail. G) All the hematopoietic lineage negative cells segregated into ALDH\textsuperscript{Hi} and ALDH\textsuperscript{Lo} cells. H) CD24/CD44 phenotype of all the lineage negative cells. I) CD24/CD44 phenotype of ALDH\textsuperscript{Hi} cells only.
Figure S3: ALDH$^{\text{High}}$ cells, although maximally enriched with CD24$^{\text{Low}}$ phenotypes, also existed within CD24$^{\text{High}}$ phenotypes in primary oral patient tissues, Related to Figure 1: Representative dot plots of A. KV017, B. AP039, C. AP037, D. AP034 patient samples (row wise). i) With DEAB Control sample for setting background fluorescence for ALDH Hi cells. ii) Without DEAB Test sample showing ALDH$^{\text{Hi}}$ (yellow) and ALDH$^{\text{Lo}}$ (pink) cells. iii) Control sample with isotype antibodies. iv) Test sample with CD24, CD44 antibodies. v) CD44/CD24 phenotype of ALDH$^{\text{Hi}}$ cells only.
Figure S4: The four subpopulations exhibit distinct repopulation patterns, Related to Figure 1:

A) (i) Representative FACS dot plots of the GBC02 parent culture showing its CD24 and ALDH phenotype with existence of four populations and their post-sort data. Re-analysis for CD24 and ALDH showing repopulation of the (ii) Red (iii) Orange (iv) Green and (v) Blue phenotypes on Day-10 of sorting.

B) Representative FACS dot plots of the SCC070 cell line’s Red, Orange, Green and Blue subpopulations repopulation pattern on Day-10 of sorting.

C) (i) Representative FACS dot plots of the SCC029b cell line’s Red, Orange, Green and Blue subpopulations repopulation pattern on Day-10 of sorting. (ii) Repopulation frequencies of each subpopulation on Day-0 and Day-10 of sorting. Error bars represent mean ± SEM from four biological repeats.

D) (i) Representative FACS dot plots of the SCC032 cell line’s Red and Orange subpopulations repopulation pattern on Day-10 of sorting. (ii) Repopulation frequencies of each subpopulation on Day-0 and Day-10 of sorting. Error bars represent mean ± SEM from two biological repeats.
**Figure S5**: CD24\textsuperscript{High} and CD24\textsuperscript{Low} phenotypes show distinct repopulation abilities, Related to Figure 1: 

**A) (i)** Sorting of CD24\textsuperscript{+} and CD24\textsuperscript{-} cells from SCC070 cell line. Established subline of CD24\textsuperscript{+} cells did not repopulate CD24\textsuperscript{-} cells whereas CD24\textsuperscript{-} cells repopulated both the subtypes. 

**(ii)** ALDH High and Low repopulation of SCC070 CD24\textsuperscript{+} subline’s sorted cells. 

**B)** ALDH High and Low repopulation of GBC035 sorted cells.
Figure S6: Clonal spheroid-cultures of the four subpopulations retained the distinct repopulation abilities, Related to Figure 2: A) Representative pictures of single spheroids and colonies generated from them in (i) Red (ii) Orange (iii) Green and (iv) Blue subpopulations of GBC02 cell line. B) Representative FACS dot plots of repopulation of single spheroids of Red, Orange, Green and Blue subpopulations, after 30 days of sorting.
Figure S7: Quantitative modeling of cell transitions showed unidirectional and bi-directional transitions on CD24 and ALDH axes respectively, Related to Figure 3: The phenotypic transition trajectories of Red, Orange, Green and Blue subpopulations of A) SCC070 B) SCC029B C) SCC032 and D) GBC035 cell lines. E) Demonstration of randomization of transition matrix for GBC02. F) Clustered scatter plot of distance of randomized transition matrices (x-axis) and the corresponding steady states (y-axis) from their wild-type (WT) counterparts, for GBC02. G) Cluster-wise heatmap for GBC02 of mean steady state values corresponding to the randomized transition matrices (TMs). The cell types are labelled as follows: C1 – CD24^{low}/ALDH^{high}, C2 – CD24^{low}/ALDH^{low}, C3 – CD24^{high}/ALDH^{high}, C4 – CD24^{low}/ALDH^{low}. The highlighted rows indicate the cluster of TMs with similar steady state as WT. H) Same as G, but for Standard deviation instead of mean. I) Same as G, but for the elements of TM. The x-axis labels are to be read as follows: C_{ij} = Transition rate of CellType C_{i} to CellType C_{j}, i,j = \{1,2,3,4\}. J) Same as I, but for standard deviation instead of mean. Low SD regions are highlighted in red. K) Cluster-wise heatmaps for SCC070 cell line showing the mean and L) standard deviation of the steady state fractions of the CellTypes (x-axis). Labelling of the cell types is the same as in GBC02 (G-H). M, N) Same as (K-L) but for transition matrices of SCC070.
Figure S8: Cisplatin enriches ALDH\textsuperscript{High} subpopulations in both CD24\textsuperscript{High} and CD24\textsuperscript{Low} phenotypes while following the treatment-naive repopulation patterns in SCC-070 cell line, Related to Figure 3: A) Representative dot plots of repopulation dynamics of SCC070 four subpopulations; (i) Red (ii) Green (iii) Orange and (iv) Blue in Untreated and Cisplatin (2µM) treated conditions. B) Repopulation frequencies of the four subpopulations (i) Red (ii) Green (iii) Orange and (iv) Blue in Untreated vs. Cisplatin (2µM) treated conditions from two biological repeats.
Figure S9: Cisplatin enriches ALDH^{High} subpopulations in both CD24^{High} and CD24^{Low} phenotypes while following the treatment-naive repopulation patterns in SCC-029 cell line, Related to Figure 3: A) Representative dot plots of repopulation dynamics of SCC029 four subpopulations; (i) Red (ii) Green (iii) Orange and (iv) Blue in Untreated and Cisplatin (2µM) treated conditions. B) Repopulation frequencies of the four subpopulations (i) Red (ii) Green (iii) Orange and (iv) Blue in Untreated vs. Cisplatin (2µM) treated conditions from two biological repeats.
Figure S10: Quantitative modeling of cell transitions under low dose Cisplatin treatment, Related to Figure 3: The phenotypic transition trajectories of the four subpopulations in Untreated vs. Cisplatin (2uM) treated conditions in A) SCC070 and B) SCC029B cell lines.
Figure S11: Differentially expressed genes from pairwise comparisons of the four subpopulations RNA sequencing, Related to Figure 4: Volcano plots of A) Red vs. Orange, B) Green vs. Blue, C) Red vs. Green and D) Orange vs. Green subpopulations pair-wise comparisons. Significantly differentially expressed genes are represented in pink color. A horizontal line is used to mark the adjusted P-value threshold > 0.05. The arrows point to ALDH1A3 gene down regulated in ALDHLo subpopulations compared to their respective ALDHHi subpopulations in figures A and B. Five up regulated (right) and down regulated (left) genes are shown in each plot.
Figure S12: Mosaic transcriptional signatures revealed hybrid states of stemness and differentiation among the four subpopulations, Related to Figure 4: A) Gene set enrichment analysis (GSEA) in (i) Orange (ii) Blue and (iii) Green subpopulations compared to Red subpopulation from GBC02 monolayer cultures showing depletion of GPCR ligand binding and receptor activity. B) Venn diagrams showing overlap of basal OEPC and basal differentiating gene sets (Jones et al, 2019) with upregulated genes from (i) Red vs. Orange comparisons in Red & Orange sorted subpopulations, and (ii) Red vs. Blue comparisons in Red & Blue sorted subpopulations, from GBC02 monolayer cultures. C) Heatmaps of mean normalized expression values from GBC02 monolayer cultures for (i) Basal OEPC TFs and (ii) Basal differentiating TFs adapted from Jones et al, 2019. D) ssGSEA scores of adult tissue stem cell (ATSC) gene signatures in the four subpopulations of GBC02 2D monolayers and 3D spheroids.
Figure S13: Survival analysis based on ssGSEA scores of gene signatures unique to the Red subpopulation, Related to Figure 5: A) Kaplan Meier curves of HNSC patients segregated into High expression (Red) and Low expression (Blue) groups based on top 60-100 uniquely up-regulated genes of Red subpopulation. B) Kaplan Meier curves of HNSC patients segregated into High expression (Red) and Low expression (Blue) groups based on top 50-100 uniquely down-regulated genes of Red subpopulation.
Figure S14: Differential expression of CD24 in response to Cisplatin in phenotypically heterogeneous GBC02 and SCC-032 cell lines, Related to Figure 5: 

A) Representative dot plots of CD24 (y-axis) and ALDH (x-axis) phenotyping of GBC02 parent cell line in Untreated vs. Cisplatin (2uM) treated conditions. 

B) Representative dot plots of CD24 (y-axis) and ALDH (x-axis) phenotyping of SCC032 parent cell line in Untreated vs. Cisplatin (2uM) treated conditions.
| Gene   | Forward (5’-3’)                      | Reverse (5’-3’)                      |
|--------|-------------------------------------|--------------------------------------|
| ALDH1A1| GATGCCGACTTGGACGAATGC               | TCTTAGCCCGCTCAACACTC                |
| ALDH1A3| CTGGATGCCCTGAGTCTGG                | TGGCTTCCCTGTATCCATCG                |
| SOX2   | AGTACAGGAGTTGTCAAGGC              | AGTCTCTAGTTAAAAGAAGGCA              |
| SOX9   | CCTGCCGTTCCTCCACCGAC            | GCTCTGGAGACTTCTGAAGAGAGC            |
| OCT4   | GACAGGGGGAGGGAGAGCTAGG          | CTTCCCTCAACCAGTGGCCCCAAAC          |
| NANO   | CAGCCCAGATCTCTCCACAGTCCC        | CGGAAGATTCCAGTGCGGTTTACC           |
| C-MYC  | GGACCGCGTCTTCCTGAAAGG              | TAACGTTGAGGGGCAATCGTC               |
| KLF4   | GGACACACGGGATGATGCTC            | CCGCTAATCAAGAGTGTTGG               |
| CK14   | CCCAGTTCTCTCTGGAATCG            | GCAGGAGGGAGGATCTTCCA              |
| CK5    | ATCCACCCTCTGGGAACAACA               | AGGCACAGTGGTGGTTGGAG              |
| CK19   | GAGCATGAAAGCTCGGCTTGG          | CTGGCTCAATACCCGCTGA              |
| CK13   | ATGGCGTGCCAAGAAGCACGAGAGG          | GGGAAACCAATCATTCTTGCGG            |
| CK1    | GATTGCCACCTACAGGACCC            | ACAGACACACTCGGTTTCC             |
| CK10   | AGAAGGCTCGTACTGTGTCG             | TTCTGGCACCACGGTTTACC               |
| ABCG2  | CACGAGAAACACCAATGCTG          | ACAGCTCTGATGAAATGCTTTC            |
| GAPDH  | GGTGCTCTCTGACTTCAACA            | GTTGCTGTAGCCAAATGCTTGT            |
| βActin | CATGGATGATATCGCCGC            | CAGATGGAGGGAAGACG              |