Interferon-Induced Bone Marrow Stromal Antigen 2 (BST2) Is A Functional Tumor-Initiating Cell Marker In Triple-Negative Breast Cancer

Eric P. Souto[1,2,3], Ping Gong[1], John D. Landua[1], R. Rajaram Srinivasan[1], Abhinaya Ganesan[1,2,3], Lacey E. Dobrolecki[1], Stephen C. Purdy[5], Heide L. Ford[5], Michael T. Lewis[1,2,3,4*]

[1]Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX, 77030, USA.

[2]Cancer and Cell Biology Graduate Program, Baylor College of Medicine, Houston, TX, 77030, USA.

[3]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, 77030, USA.

[4]Departments of Molecular and Cellular Biology and Radiology, Baylor College of Medicine, Houston, TX, 77030, USA.

[5]Department of Pharmacology, University of Colorado Anschutz Medical Campus (UC-AMC), Aurora, CO, 80045, USA; Pharmacology Graduate Program, UC-AMC, Aurora, CO, 80045, USA; University of Colorado Cancer Center, UC-AMC, Aurora, CO, 80045, USA.

[*]Corresponding Author

Michael T. Lewis

Baylor College of Medicine

One Baylor Plaza

BCM600; Room N1210

TEL: 713-798-1538

Email: mtlewis@bcm.edu

SUMMARY

Tumor-initiating cells (TIC) are a tumor cell subpopulation thought to be responsible for therapeutic resistance and metastasis. Using a Signal Transducer and Activator of Transcription (STAT) reporter, and a STAT-responsive lineage tracing system, we enriched for cells with enhanced mammosphere-forming potential in some, but not all, triple-negative breast cancer xenograft models (TNBC) indicating TIC-related and TIC-independent functions for STAT signaling. Single-cell RNA sequencing (scRNA-seq) of reporter-tagged xenografts identified a common interferon-associated transcriptional state, previously linked to inflammation and macrophage differentiation, in TIC. Similar transcriptional states exist in human breast cancer patient scRNA-seq datasets. Flow cytometric sorting using bone marrow stromal cell antigen 2 (BST2), a marker of this state, enriched for TIC, and BST2 knockdown reduced mammosphere-forming potential. These results suggest TIC may exploit the interferon response pathway to promote their activity in TNBC. Our results lay the groundwork to target interferon-associated pathways in TIC in a subset of TNBC.

Keywords: Cancer stem cells, lineage tracing, fluorescent reporter, single-cell RNA sequencing

INTRODUCTION

Breast cancer is a heterogeneous disease, which partly explains differences in prognosis, treatment response, and metastasis patterns between patients. Heterogeneity between patients (intertumoral heterogeneity) is reflected by clinical staging (Nagao et al., 2012), molecular classifications (Abd El-Rehim et al., 2005; Perou et al., 2000; Sotiriou et al., 2006), and expression of clinically-relevant biomarkers, such as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER2) (Hammond et al., 2010; Wolff et al., 2013). Tumors expressing ER(+/-PR) protein tend to respond well to hormonally targeted therapies, whereas HER2+ tumors respond well to agents targeting HER2. In contrast, triple-negative breast cancer (TNBC) is a breast cancer subtype defined by lack of ER, PR, and HER2 expression. As such, TNBC has limited treatment options, and is typically treated with combination chemotherapies.

In addition to intertumoral heterogeneity, cells within an individual tumor display intratumoral heterogeneity, and can be quite distinct from one another with respect to their genomic, transcriptomic, proteomic, epigenomic, and metabolomic profiles (Cancer Genome Atlas, 2012; Curtis et al., 2012; Gerlinger et al., 2012). Intratumoral heterogeneity is a major challenge for diagnosis, prognosis, and response prediction, particularly in TNBC, which itself represents at least six molecularly distinct subcategories of breast disease (Cancer Genome Atlas, 2012; Curtis et al., 2012; Lehmann et al., 2011).

Within tumors there exists a subpopulation of cells, typically relatively rare, termed cancer stem cells (CSC) or tumor-initiating cells (TIC). TIC are defined functionally by their ability to self-renew in sphere-formation assays, and to recapitulate clonally-derived cellular hierarchies of the tumor from which they were isolated upon transplantation (Reya et al., 2001). Therapeutic targeting of TIC is a significant challenge due to a lack of unique and phenotypically stable markers, and the need to conduct TIC assays to evaluate drug efficacy, which is rarely done.

In various tumors, enriched populations of TIC can be identified by biomarkers such as ALDH1 activity (Ginestier et al., 2007), the cell surface marker combination of $CD44^{+}/CD24^{low/-}$ (Al-Hajj et al., 2003), or fluorescent reporters for transcription factors that regulate TIC function, such as Wnt/β-catenin, Hedgehog, and Notch signaling, as well as POU5F1 (OCT4)/SOX2, and Signal Transducer and Activator of Transcription 3 (STAT3)-mediated signaling (Dierks et al., 2008; Harrison et al., 2010; Iliopoulos et al., 2011; Marotta et al., 2011; Wei et al., 2014; Zhang et al., 2010). STAT3 signaling is well known for its roles in both normal stem cells (Cartwright et al., 2005; Clarkson et al., 2006; Humphreys et al., 2002; Niwa et al., 1998) and malignant stem cells (Chung et al., 2013; Chung et al., 2014; Lin et al., 2011; Marotta et al., 2011; Philp et al., 1996) in several tissues, including the breast.

Some metastatic tumor cells express the TIC marker CD44 and have enhanced tumor-initiating capacity (Liu et al., 2010), suggesting TIC are responsible for metastasis. Following taxane- and anthracyclin-based chemotherapy, residual breast cancer cells in TNBC are enriched for TIC markers, including $CD44^{+}/CD24^{low/-}$ and aldehyde dehydrogenase (ALDH1) activity, and have enhanced mammosphere-forming efficiency, indicating that TIC are intrinsically resistant to some systemic therapies (Alamgeer et al., 2014; Korkaya et al., 2012; Li et al., 2008; Yu et al., 2007). Therefore, standard cancer therapies can eliminate many “bulk” tumor cells effectively, resulting in tumor shrinkage, but can also allow survival of TIC leading to eventual local and/or distant disease recurrence. Thus, selective targeting and eradicating TIC may be essential for effective treatment of breast cancer.

Here, we used an EGFP-expressing lentiviral signaling reporter responsive to STAT-mediated transcriptional activity (termed 4M67-EGFP, responsive to, at least, STAT1 and STAT3) and a STAT-responsive lineage tracing (LT) system to enrich cells with enhanced mammosphere-forming potential in some, but not all, xenograft models of TNBC. We performed single-cell RNA sequencing (scRNA-seq) on three xenograft models and identified a distinct interferon-

associated transcriptional state. Furthermore, we confirmed the presence of this transcriptional state in cells derived from human breast cancer patients using published scRNA-seq datasets. Using mammosphere-formation and limiting-dilution transplantation assays, we demonstrated that a marker of this cell state, bone marrow stromal cell antigen 2 (BST2), can be used to enrich for TIC, and BST2 expression was increased in residual tumors following chemotherapy. Finally, we demonstrated that BST2 knockdown reduces mammosphere-forming efficiency (MSFE). Taken together, our data suggest that TIC exploits the interferon-mediated inflammatory response pathway, a pathway best understood in the immune system in response to infection, to regulate other processes including self-renewal, and possibly also cell state decisions such as cellular quiescence/activation.

RESULTS

STAT reporter activity is heterogenous in a cohort of triple-negative breast cancer PDX models

We showed previously that STAT-mediated transcriptional activity using a lentiviral signaling reporter expressing EGFP downstream of four STAT response elements allows enrichment for TIC in two cell line xenograft (CDX) models of claudin-low breast cancer (SUM159 and MDA-MB-231) (Wei et al., 2014). While our focus was on STAT3 in these two models, STAT transcription factors bind similar DNA sequences (Ehret et al., 2001); therefore, this reporter may be responsive to other STAT family members, including STAT1.

STAT1 and STAT3, are expressed in many, but not all, human breast tumors (Wang et al., 2021; Widschwendter et al., 2002). At the mRNA level, 64.2% of breast tumors in the METABRIC cohort display comparatively elevated *STAT3* mRNA levels (Aleskandarany et al., 2016). At the protein level, pSTAT3 is observed in ~37.5% of TNBC patients, the highest of the three clinical subtypes (Wang et al., 2021). Although pSTAT1 has not been profiled as extensively in human breast tumors as pSTAT3, likely due to its known role as a tumor suppressor (Koromilas and Sexl, 2013), STAT1 activity has been detected using DNA binding assays and Western blot for pSTAT1 in a cohort of 73 breast cancer patients (Widschwendter et al., 2002).

To determine whether STAT signaling might also be associated with TIC function in a broader range of TNBC, and to determine whether our collection of PDX models represented a useful range of pSTAT3 expression, we screened a cohort of 39 TNBC PDX models for pSTAT using immunohistochemistry on tissue microarrays. Of the 39 TNBC PDX tested, 29 (74%) showed detectable pSTAT3 staining in the epithelium (data not shown), a value higher than, but arguably consistent with, results using clinical samples of TNBC given the known engraftment bias present in PDX collections toward immunologically “cold” tumors (Petrosyan et al., 2023).

This list of TNBC models was then cross-referenced with lists of PDX showing behaviors of interest with respect to the probable role of TIC, specifically local recurrence (or lack thereof) after apparent complete response to one or more chemotherapies, and/or metastasis to one or more organ sites. A subset of 12 PDX models of TNBC were then chosen for genetic tagging with the 4M67-EGFP lentiviral vector used previously (Wei et al., 2014), including potential negative control models. Three models, BCM-0046, BCM-7649, and WHIM12 showed no detectable EGFP expression (Supplementary Table 1), while the remaining eight PDX showed varying degrees of EGFP expression. Of these, three PDX models, BCM-3611, BCM-4013, and FCP699, were excluded from further consideration as useful models because the percentage of EGFP+ cells was too low (<<1%) to be experimentally tractable, thereby yielding five remaining candidate PDX models from which to choose for downstream analyses. At this time BCM-5097 was excluded due to slow growth kinetics upon transplantation leaving a final selection of four EGFP-expressing PDX.

The four EGFP-expressing candidates and two EGFP-negative PDX were then evaluated for pSTAT1 by IHC and EGFP expression by immunofluorescence microscopy in conjunction with the pSTAT3 IHC staining described above (Fig. 1A). Consistent with pSTAT1/3 positivity by IHC, PDX models displayed their anticipated STAT reporter activity via detection of EGFP (Fig. 1A), quantified by analytical flow cytometry (Fig. 1B). This revealed a continuum of EGFP reporter activity across models ranging from 7.88% (BCM-15006) to 31.7% (MC1) of tumor cells, ranges that should allow for discrimination of differential TIC function/gene expression in EGFP+ vs. EGFP- cells, should such differentials exist.

To examine the relationship between the chosen breast PDX models with respect to STAT1 and STAT3 mRNA expression and EGFP fluorescence, we generated heat maps based on Z-score normalized mRNA expression and hierarchical clustering (Fig. S1A–C), annotating the models chosen for analysis. This revealed models with no reporter activity (BCM-0046 and WHIM12)

clustered together and had predictably low STAT1 and STAT3 mRNA expression. In contrast, reporter-positive models BCM-4272, BCM-7821, and MC1 clustered closely and exhibited high STAT1 and STAT3 mRNA levels, whereas reporter-positive model BCM-15006 existed in a separate clade defined by median levels of STAT3 and comparatively lower levels of STAT1 mRNA expression.

To determine whether STAT reporter activity enriches for cells with high mammosphere-forming efficiency (MSFE) in reporter-positive TNBC PDX models (a surrogate for TIC function), we sorted $EGFP^+$, $EGFP^-$, and a mock sorted control from each reporter-positive model and plated them under low attachment conditions to test MSFE (Fig. 1C, S1D). STAT reporter positive cells in BCM-4272 had an 8.04% MSFE, representing a more than 100% increase relative to the $EGFP^-$ cell population (3.74% MSFE), and the MSFE of the $EGFP^-$ population was reduced by 59% relative to a mock sorted control (6.3% MSFE) (Fig. 1D). In contrast, STAT reporter activity did not enrich for cells with enhanced MSFE in BCM-15006, BCM-7821, or MC1 (Fig. 1D, S1E). Thus, while the 4M67 reporter can identify cells with enhanced MSFE in TNBC, STAT-mediated transcription must have MSFE-independent function(s) as well.

STAT lineage tracing reporter provides a long-term fluorescent label to STAT signaling cells

The 4M67-EGFP reporter provides a binary view of STAT activity, only capturing cells in a STAT on ($EGFP^+$) or STAT off ($EGFP^-$) state. Given that TIC marker expression, including STAT activity, is dynamically regulated (Liu et al., 2014; Meyer et al., 2009), the inability to lineage trace STAT signaling cells as they undergo cell state changes, or differentiate, over time is a major limitation of the 4M67-EGFP reporter. To augment the 4M67-EGFP reporter, we developed a TAM-activated, Cre recombinase-dependent, STAT signaling-specific, lentiviral LT system (Fig. S2A–B).

The LT system consists of two components: the first is a lentiviral vector that labels cells with active STAT signaling ($EGFP^+$), followed by a self-cleaving peptide and TAM-activated Cre recombinase (4M67-EGFP-P2A-$CreER^{T2}$) (Fig. S2A). The second component is a constitutively expressed dual-color switching Cre-dependent reporter vector (EFS-$_{loxP}$dsRed$_{loxP}$-mNeptune2) (Fig. S2B). Addition of TAM drives color switching from red (dsRed) to far-red (mNeptune2) via $CreER^{T2}$ recombination in STAT signaling cells in a specific temporal window where the TAM concentration is high enough to activate Cre recombinase, given its 6.8-hour half-life (Reid et al., 2014) (Fig. 2A). Use of the LT system results in four cell populations: long-lived STAT signaling cells ($EGFP^+$/mNeptune2$^+$), transient STAT signaling cells (or cells that did not undergo Cre recombination) ($EGFP^+$/dsRed$^+$), lineage-tagged cells ($EGFP^-$/mNeptune2$^+$), and bulk tumor cells ($EGFP^-$/dsRed$^+$) (Fig. 2B). These four LT cell populations can be visualized in situ by confocal microscopy or enriched using FACS. Ultimately, these four cell populations can be probed to understand the role of STAT signaling cells and their progeny, as well as identify candidate STAT TIC genetic vulnerabilities.

To demonstrate the sensitivity and specificity of the STAT LT reporter system, we used HepG2/C3A cells because they are responsive to IL-6 treatment in a STAT3-dependent manner (Horvath et al., 1995; Wei et al., 2014). Untreated STAT LT-transduced HepG2/C3A cells (HepG2/C3A STAT LT) had minimum EGFP and mNeptune2 expression and the addition of TAM (1 μM) induced mNeptune2 expression in a small percentage of cells. Addition of IL-6 (50 ng/mL) induced EGFP reporter activity, without mNeptune2 activation, but the addition of both IL-6 and TAM induced EGFP and mNeptune2 expression (Fig. S2C–D). These results demonstrate the LT reporter system is both stringent and sensitive to STAT signaling, and TAM addition allows mNeptune2 expression in STAT-responsive cells.

To demonstrate the ability of the LT system to provide a long-term fluorescent label on STAT signaling cells *in vivo*, we introduced the LT reporters into two TNBC cell lines where our

previous studies demonstrated STAT3 reporter activity enriches for TIC (SUM159 STAT LT and MDA-MB-231 STAT LT) (Wei et al., 2014). To ensure effective lineage-tagging of STAT signaling cells, we performed an in vivo TAM dose-response analysis (Fig. 2SE–F). Given that high levels of TAM can cause toxicity in mice (Guo et al., 2002), we aimed to find a dose that resulted in maximal loxP recombination without any adverse effects. At 4 mg, the recombination efficiency (the percentage of $EGFP^+/Nep^+$ cells out of all $EGFP^+$ cells) was 71%. The recombination efficiency and the total percentage of $mNeptune2^+$ cells decreased with each successive dose with the lowest dose (0.006 mg) being comparable to vehicle treatment. Given the high recombination frequency with no observable toxicity, we used 4 mg TAM treatments for the remainder of our studies.

SUM159 STAT LT or MDA-MB-231 STAT LT cell lines were engrafted into SCID/beige mice, and tumor-bearing mice were treated with 4 mg TAM when tumors reached 100 mm^3. Two days after TAM treatment, we visualized all four LT cell populations in situ using confocal microscopy and quantified their abundance in each tumor model using flow cytometry (Fig. 2C–F). In both cell line xenografts, TAM treatment increased the percentage of $EGFP^+/Nep^+$ cells (from 0.25% to 1.8% in SUM159 and from 0.02% to 0.07% in MDA-MB-231), which was accompanied by a decrease in $EGFP^+/dsRed^+$ cells (from 3.08% to 0.7% in SUM159 from 0.7% to 0.4% in MDA-MB-231). Likewise, TAM treatment increased the percentage of $mNeptune2^+$ cells (from 1.98% to 8.88% in SUM159 and from 0.07% to 3.15% in MDA-MB-231) with a corresponding decrease in the percentage of $dsRed^+$ cells (from 94.7% to 88.3% in SUM159 and from 96.7% to 93.3% in MDA-MB-231). These results indicate that TAM treatment leads to Cre-dependent mNeptune2 expression *in vivo*, demonstrating the utility of the STAT LT system to study STAT TIC biology.

Long-lived STAT signaling cells in SUM159 xenografts have enhanced mammosphere-forming efficiency

To confirm that the STAT LT system enriches cells with enhanced MSFE from SUM159 tumors in a manner similar to that of the 4M67-EGFP vector shown previously, we dissociated SUM159 xenografts eight days after TAM treatment, sorted the four fluorescent cell populations ($EGFP^+$/$mNeptune2^+$, $EGFP^+$/$dsRed^+$, $EGFP^-$/$mNeptune2^+$, and $EGFP^-$/$dsRed^+$), and performed a MSFE assay (Fig. 3A–B). Long-lived STAT signaling cells ($EGFP^+$/$mNeptune2^+$) had a significantly higher MSFE (7.2%) compared to all other cell populations. Furthermore, the MSFE of the mock (1.6%), $EGFP^+$/$dsRed^+$ (3.0%), $EGFP^-$/$mNeptune2^+$ (0.5%), and $EGFP^-$/$dsRed^+$ (2.2%) were not statistically different from each other. These results demonstrated that the long-lived STAT signaling cells have enhanced mammosphere-formation potential relative to all other STAT LT cell populations in SUM159 xenografts.

The cancer stem cell model postulates that cellular heterogeneity in tumors results, in part, from hierarchical organization of cell types observed, and that these hierarchies are fueled by rare, self-renewing tumor-initiating or “cancer stem” cells (Batlle and Clevers, 2017; Dalerba et al., 2007). To fuel this hierarchy, TIC undergo both symmetric (i.e., producing two TIC) and asymmetric divisions (producing a TIC and a daughter cell committed to differentiating) (Driessens et al., 2012). Considering the cancer stem cell model, we reasoned that long-lived STAT signaling cells give rise to progenitor cells ($EGFP^-$/$mNeptune2^+$) that become the non-TIC bulk tumor over time. To test this hypothesis, we performed a 16-day time course study. We harvested SUM159 tumors prior to TAM treatment, and again at days 2, 4, 8, and 16 after the TAM pulse (Fig. 3C–D). A total of 74% of the $EGFP^+$ cells recombined ($EGFP^+$/$dsRed^+$ converted to $EGFP^+$/$mNeptune2^+$ cells) two days after the TAM pulse, consistent with the value we demonstrated previously. Between days 2 and 16, the percentage of $EGFP^-$/Nep^+ cells increased from 18% to 35% of the tumor, indicating lineage-tagged cells proliferate and displace the $EGFP^-$/$dsRed^+$ cells, becoming bulk tumor cells over time. Furthermore, we noted the re-

emergence of $EGFP^+$/$dsRed^+$ cells on the 8th day, suggesting STAT signaling plasticity in SUM159 xenografts.

To extend the timeframe of this experiment, we serially transplanted tumor chunks from mice two days after the TAM pulse for a total of three transplant generations (TG) (Fig. 3E–F). The percentage of lineage-tagged cells increased from 18% to 61% from TG1 to TG3, respectively. Furthermore, the total percentage of STAT reporter positive cells remained unchanged over time and following serial passaging, indicating tight regulation of STAT signaling. Taken together, these results suggest long-lived STAT signaling cells sit atop the cellular hierarchy, giving rise to lineage-tagged progeny that generates the more differentiated bulk of the tumor in SUM159 xenografts.

Interferon response genes are differentially expressed in TIC

To determine whether STAT reporter-positive cells express a distinct set of genes that may regulate TIC biology, we performed scRNA-seq on all reporter populations from three TNBC xenograft models: two models where STAT reporter activity enriched for TIC (SUM159 and BCM-4272) and one model where STAT reporter activity did not enrich for TIC (BCM-15006) (Fig. 4A). Because STAT reporter positivity enriched for TIC in SUM159 and BCM-4272 xenografts, we overlayed sample information on Uniform Manifold Approximation and Projection (UMAP) plots to visualize whether reporter positive cells clustered together (Fig. 4B, Fig. S4.1A–C). Though EGFP+ cells can be found in all clusters, cluster 7 in SUM159 almost exclusively contains $EGFP^+$ cells (Fig. 4B, Fig S4.1A). Regarding BCM-4272, most $EGFP^+$ cells lie in clusters 0, 2, 4, 6, and 7, suggesting at least one of these clusters may represent a TIC transcriptional state (Fig. 4B, Fig. S4.1B). Because STAT reporter activity did not enrich TIC (Fig. 1C–D) in BCM-15006, we reasoned that a cluster that represented a TIC transcriptional state would have a roughly equal representation of $EGFP^+$ and $EGFP^-$ cells. Cluster 6 is the

only cluster that contains a roughly equal distribution of EGFP$^+$ and EGFP$^-$ cells, suggesting this cluster may represent a TIC transcriptional state (Fig. 4B, Fig S4.1C).

Because dynamic cellular processes, including TIC differentiation into bulk tumor cells, are captured through scRNA-seq, we used Monocle3 to construct lineage trajectories with the hypothesis that a TIC cluster must be predominantly EGFP$^+$ (or an equal distribution in the case of BCM-15006) and lie at one end of a trajectory (Fig. 4C). Based on this reasoning, cluster 7 from SUM159 (SUM159-C7) was the only cluster with an overrepresentation of EGFP$^+$ cells that laid at the end of the lineage trajectory. Cluster 3 was the latest in pseudotime, representing the most differentiated cell state, which had an overrepresentation of dsRed$^+$ and EGFP$^+$/dsRed$^+$ cells (Fig. 4B). Cluster 7 from BCM-4272 (BCM-4272-C7) was also defined by an overrepresentation of EGFP$^+$ cells that laid at the end of the lineage trajectory. Critically, constructing the lineage trajectory with BCM-4272-C7 as the root resulted in a predominantly EGFP$^-$ cluster as the latest in pseudotime, suggesting this transcriptional state is most differentiated, and may have the least tumorigenic potential. This is consistent with our previous data demonstrating that sorting EGFP$^-$ cells reduced MSFE (Fig. 1C). Because reporter-positive cells in BCM-15006 did not enrich nor deplete cells with enhanced MSFE (Fig. 1C), a TIC transcriptional state would be defined by a cluster with an equal distribution of EGFP$^+$ and EGFP$^-$ cells, and at the end of a lineage trajectory. Cluster 6 (BCM-15006-C6) was the only cluster that met both of these criteria. Interestingly, an EGFP$^+$ cluster was the latest in pseudotime, suggesting STAT activity may play a role in more differentiated cell states in some TNBC models. Taken together, these findings suggest SUM159-C7, BCM-4272-C7, and BCM-15006-C6 may represent a TIC transcriptional state.

To explore tumor cell heterogeneity in these three models further, we performed gene set enrichment analysis (GSEA) using the Hallmark gene set to gain insights into the biology of each cluster. Regarding SUM159 xenografts, we identified clusters defined by enrichment of

epithelial-mesenchymal transition (cluster 3 and cluster 8), oxidative phosphorylation (cluster 9), hypoxia (cluster 6), Myc targets V1 (cluster 0 and cluster 4), as well as interferon response and TNF-alpha signaling via NF-kB (cluster 5 and cluster 7) gene sets (Fig. S4.2). In BCM-4272 xenografts, we identified clusters defined by enrichment of epithelial-mesenchymal transition (cluster 0 and cluster 4), oxidative phosphorylation (cluster 1 and cluster 3), and mitotic spindle (cluster 5) gene sets (Fig. S4.3). For BCM-15006 xenografts, we identified clusters defined by enrichment of TNF-alpha signaling via NF-kB (clyster 0, cluster 3, and cluster 5), oxidative phosphorylation (cluster 1), and mTORC signaling (cluster 2) gene sets (Fig. S4.4). These findings reveal extensive cell state diversity within each xenograft model.

Because we reasoned SUM159-C7, 4272-C7, and 15006-C6 may represent a TIC transcriptional state, we examined our GSEA for these clusters to determine which gene sets may be relevant to TIC biology. All three clusters were defined by enrichment of interferon gamma response, interferon alpha response, and TNF-alpha signaling via NF-kB (Fig. 4D, Fig. S4.5). Furthermore, the gene set Myc targets V1 was downregulated in all three clusters. To evaluate enrichment across clusters within a single model further, we used Escape ssGSEA (Borcherding et al., 2021) to plot enrichment of interferon alpha response and interferon gamma response across each cluster in each xenograft model (Fig. 4E). SUM159-C7, 4272-C7, and 15006-C6 all displayed enrichment of interferon response relative to all other clusters. Although this enrichment was not as pronounced in SUM159-C7, likely because TNF-alpha signaling via NF-kB was the most enriched gene set in this cluster.

Given similar gene sets were enriched in SUM159-C7, 4272-C7, and 15006-C6, we hypothesized these clusters may share gene-level commonalities. To investigate the extent to which candidate TIC clusters share transcriptional features, we intersected the differentially expressed genes from each cluster (Fig. 4F). SUM159-C7 and 4272-C7 shared 48 features ($p=1.8e^{-24}$), SUM159-C7 and 15006-C6 shared 38 features ($p=5.7e^{-27}$), and 4272-C7 and

15006-C6 shared 79 features ($p=1.1e^{-110}$). In this process, we identified 25 common up-regulated genes across all three clusters. A list of genes from all intersections is shown in Supplementary Table 2 and a list of the 25 common up-regulated genes, and their known functions, are shown in Supplementary Table 3.

Next, we questioned whether this collection of genes exists in a gene regulatory network. To address the question, we used WebGestalt (Liao et al., 2019) and found that the majority of these 25 genes lie in the same gene regulatory network (Fig. 4G). Finally, we performed an over-representation analysis to determine whether these 25 genes are involved in the same biological processes (Fig. S4.6). We found that six of the 25 genes are involved in biological processes typically associated with viral response of the immune system, including interferon signaling.

To determine whether interferon alpha or interferon gamma can activate genes that define the interferon-associated TIC transcriptional state in the context of TNBC, we treated SUM159, BCM-4272, and BCM-15006 cells with interferon alpha or interferon gamma and compared the expression of six genes (PARP9, PARP14, STAT1, ADAR, OAS2, and BST2) from our 25 intersected TIC genes to untreated controls by qRT-PCR (Fig. S4.7). Treatment with interferon alpha or interferon gamma increased the expression of all six genes relative to the untreated controls in all three TNBC models. These findings suggest that interferon can induce genes that define the interferon-associated TIC transcriptional state, although functional studies are required to determine what is actually responsible for activating these genes in vivo. Taken together, these findings indicate interferon response genes are activated in SUM159, BCM-4272, and BCM-15006 TIC.

Interferon response genes are differentially expressed in a subset of epithelial cells in human breast cancer patient samples

To determine whether a similar interferon response transcriptional state is also reflected in TNBC patient samples, we analyzed scRNA-seq data from four TNBC patients retrieved from two previously published human breast cancer cell atlases (Pal et al., 2021; Wu et al., 2021). We restricted our analysis to the four TNBC patients with more than 1,000 non-cycling epithelial cells to increase our chances of identifying rare cell subpopulations, such as TIC (Al-Hajj et al., 2003; Ishizawa et al., 2010) (Fig. 5A). To determine whether clusters from the human breast cancer epithelial cells represent a similar transcriptional state to the TIC states we identified, we gave each cell a TIC score based on similarity to our intersected 25-gene signature (Fig. 5B). This revealed cluster 4 from patient CID44991, cluster 5 from patient B1-MH-4031, and cluster 9 from patient TN-MH-0135 had the highest score relative to all other clusters. However, there was no cluster present in patient B1-MH-0131 that scored higher than any other (Fig. S5A). These results suggest the presence of an interferon response transcriptional state in some TNBC patients.

To determine whether clusters in these patient samples identified by a high TIC score were defined by interferon response features, like our xenograft models, we performed GSEA using the Hallmark dataset (Fig. 5C). Similar to our findings in our three xenograft models, inflammatory genes sets, such as interferon alpha response, interferon gamma response, and TNF-alpha signaling via NF-kB, were enriched in all patient clusters defined by a high TIC score. Interferon response gene sets were most enriched in all clusters with a high TIC score (cluster 4 from CID44991, cluster 5 from B1-MH-4031, and cluster 9 from patient TN-MH-0135), followed by TNF-alpha signaling via NF-kB, similar to 4272-C7 and 15006-C6. In addition, we performed GSEA using the Reactome and GO Biological Processes datasets, which confirmed that gene sets associated with interferon and cytokine signaling are enriched (Fig. S5B).

To determine the extent to which clusters with high TIC scores share transcriptional features with the interferon response transcriptional state identified from our xenograft models (Fig. 4F),

we intersected the genes from each cluster with the 25 common genes from TIC clusters across all three xenograft models (Fig. 5D). All intersections were significant, with CID44991 cluster 4 sharing 13 genes ($p=1.1e^{-15}$), TN-MH-0135 cluster 9 sharing 17 genes ($p=7.7e^{-28}$), and B1-MH-4031 cluster 5 sharing 21 genes ($p=1.0e^{-37}$) with the 25 intersected TIC genes. A complete list of all intersections between patient clusters and xenograft clusters can be found in Supplementary Table 4. Taken together, our findings reveal an interferon response TIC transcriptional state in a subset of human TNBC.

BST2 is a TIC marker in a subset of TNBC models

To test whether tumor cells with activated interferon response genes represent TIC functionally, we needed to identify a biomarker that could be used to isolate these cells. To that end, we examined our 25 intersected genes and prioritized cell surface markers to facilitate enrichment using FACS. We identified BST2 as a cell surface protein among our list of 25 intersected genes (Mahauad-Fernandez and Okeoma, 2016). BST2 is an interferon-stimulated gene known for its tethering ability, which blocks the release of enveloped viruses from infected host cells (Neil et al., 2008; Van Damme et al., 2008). In breast cancer, BST2 is a predictor of overall survival and is responsible for invasiveness, proliferation, and apoptosis evasion (Mahauad-Fernandez et al., 2015; Mahauad-Fernandez et al., 2014; Mahauad-Fernandez et al., 2018; Sayeed et al., 2013).

To evaluate whether BST2 expression is enriched in mammospheres and tumors relative to 2D culture, we analyzed BST2 expression in SUM159 cells using analytical flow cytometry (Fig. 6A). The percentage of BST2 high cells increased from 17.5% in SUM159s grown *in vitro* to 40.7% in SUM159 mammospheres. Furthermore, the percentage of BST2 high cells from *in vivo*-grown SUM159s was 55.3%. These results indicate BST2 expression increases in both mammospheres and tumors, suggesting BST2 may be a TIC marker, and demonstrates that 3D mammospheres may represent a compromise between *in vivo* and *in vitro* with respect to TIC enrichment.

Next, we evaluated whether isolation of BST2 high cells could enrich cells with enhanced mammosphere forming potential in SUM159 and BCM-4272. To that end, we dissociated each tumor and FACS-enriched BST2 high (top 20%), BST2 low (bottom 20%), and mock sorted cells to test their mammosphere forming potential (Fig. 6B). In all models, BST2 high cells were enriched for MSFE relative to both BST2 low cells and the mock sorted control. For SUM159, there was no statistically significant difference between the mock control and BST2 low, whereas BST2 low cells had significantly less MSFE compared to the mock for BCM-4272. These results indicate that cells with high BST2 expression are enriched for mammosphere-forming potential.

To test whether BST2 can be used to enrich TIC from SUM159 xenografts, we compared the tumor-initiating potential of BST2 high (top 20%), BST2 low (bottom 20%), and mock-sorted cells using a limiting-dilution transplantation assay (Fig. 6C). Notably, the BST2 high cells were the only population capable of forming tumors at all doses, with a TIC frequency of 1 in 1,721 cells. The BST2 low cells had a 10-fold lower TIC frequency of 1 in 17,370 cells, whereas the mock-sorted group had a modestly lower TIC frequency of 1 in 2,501 cells. The modest TIC enrichment in the BST2 high compared to the mock may be because BST2 is expressed in most SUM159 xenograft cells, albeit to varying degrees (Fig. 6A). These results demonstrate cells with high BST2 expression have enhanced tumorigenic capacity compared to BST2 low cells in SUM159 xenografts. Taken together, our results suggest that a population of TIC are defined by the differential expression of interferon response genes, including BST2 that can be used to identify and enrich this tumorigenic population.

Because residual breast cancer cells in TNBC are enriched for TIC (Alamgeer et al., 2014; Korkaya et al., 2012; Li et al., 2008; Yu et al., 2007), we evaluated whether BST2 expression was increased following chemotherapy. We treated BCM-4272 with carboplatin or docetaxel (once per week for four weeks) and evaluated BST2 expression in the remaining tumors (6

docetaxel treated, 3 carboplatin treated) relative to a vehicle control by qRT-PCR (Fig. 6D). In BCM-4272 residual tumors, BST2 expression increased 7.2-fold following carboplatin treatment, and 6.9-fold following docetaxel treatment relative to the vehicle-treated control. These results demonstrate BST2 expression is increased in residual tumors following chemotherapy, suggesting an enrichment of BST2-expressing TIC in residual tumors or that BST2 expression is induced by chemotherapy. As a result, BST2 may contribute to therapeutic resistance and local recurrence.

BST2 knockdown reduces mammosphere-formation efficiency in SUM159 cells

To determine whether BST2 regulates TIC function, we generated two BST2 knockdown SUM159 cell lines (sh-BST2 #1 and sh-BST2 #2) and compared their MSFE relative to a non-targeting control (Fig. S7). BST2 knockdown reduced MSFE from 3% in the non-targeting control cells to 1.9% and 1.2% in sh-BST2 #1 and sh-BST2 #2, respectively (Fig. 7A–B). These results demonstrate BST2 expression functionally regulates MSFE in SUM159 cells, suggesting BST2 may be a regulator of TIC function. Taken together, our data demonstrate that BST2 is a functional TIC marker in a subset of TNBC models.

DISCUSSION

The cancer stem cell hypothesis postulates tumors maintain cellular hierarchical organization due to tumor-initiating, or “cancer stem”, cell self-renewal and asymmetric division. Numerous lines of evidence supporting this hypothesis have been generated primarily using enrichment of cell surface markers, enzymatic activity of ALDH, and lentiviral reporter vectors for signaling pathways associated with stem cell biology, such as STAT3 (Visvader and Lindeman, 2008, 2012; Wei et al., 2014). Because we previously reported that STAT reporter activity enriched TIC in two TNBC cell line xenograft models representing the claudin-low molecular subtype (Neve et al., 2006; Wei et al., 2014), we tested whether reporter activity could also isolate cells from TNBC PDX models with enhanced MSFE. By doing so, we identified BCM-4272 as an additional model where STAT reporter activity can enrich cells with enhanced MSFE. Despite this finding, the reporter did not enrich cells with enhanced MSFE in the three other TNBC PDX models tested, suggesting STAT-mediated transcription must have independent function(s). Indeed, STAT3 alone has numerous functions in breast cancer, including promoting immune evasion (Jones et al., 2016), metastasis (Kang et al., 2003), and chemoresistance (Wang et al., 2018), and the STAT reporter is likely identifying cells with different functions across models. Dissecting the different functions in response to STAT-mediated transcription across TNBC PDX models can be explored in future studies. To that end, we have transduced additional PDX models with the STAT reporter and identified additional models with reporter activity that can be used in future studies (Supplementary Table 4).

Because STAT activity is dynamically regulated, STAT reporters only offer a binary view of STAT-mediated transcription. To overcome this challenge, we developed a LT system to provide an inducible fluorescent label to permanently STAT signaling cells as they undergo cell state changes. We validated the ability of the STAT LT system to identify TIC in SUM159 xenografts, a model in which we previously showed that STAT reporter activity enriches TIC

(Wei et al., 2014). We demonstrated the utility of the LT system further by performing time course and serially transplantation experiments, both of which demonstrated the enrichment of lineage-tagged cells ($EGFP^-/Nep^+$) over time. These findings are consistent with the cancer stem cell hypothesis, that a small tumorigenic subpopulation proliferates and differentiates to maintain the cellular hierarchy within a tumor (Tan et al., 2006). The LT system is a powerful tool that can augment traditional biomarker isolation of TIC. Due to the flexible design of the promoter donor vector, the 4M67 promoter can be replaced by restriction cloning to generate LT systems for other signaling pathways, such as Wnt, SORE6, and Hedgehog. Additionally, there are many outstanding questions in the cancer stem cell field that the LT system can help address. It could be used to functionally address whether TIC from primary tumors function as metastasis-initiating cells (Batlle and Clevers, 2017; Massague and Ganesh, 2021) and whether enriched TIC populations following chemotherapy are the result of a therapeutic selective pressure or activated in response to the therapy (Batlle and Clevers, 2017; Kim et al., 2018).

In this study, we leveraged our newly developed STAT LT system and a STAT reporter to uncover gene expression patterns unique to TIC. To date, all efforts to uncover cancer stem cell gene signatures have relied on bulk sequencing enriched populations of TIC (Colacino et al., 2018; Creighton et al., 2009), or scRNA-seq using a limited panel of genes (Colacino et al., 2018). Others have noted the presence of “stem-like cells” in their single-cell sequencing data (Jiang et al., 2021). We combined scRNA-seq, which has the power to detect rare cell populations, with our reporter systems. Doing so allowed us to integrate our functional knowledge of whether reporter activity enriched populations of TIC with transcriptional data at a higher resolution than had been applied previously. Using three TNBC xenograft models, we demonstrated that interferon response genes are differentially expressed in TIC. By intersecting upregulated genes from candidate TIC transcriptional states across these three models, we

identified 25 common transcriptional features that can be prioritized to study their precise mechanistic role in TIC.

Interferons are pleiotropic cytokines known to exert both pro-tumorigenic and anti-tumorigenic effects (Gocher et al., 2022). Interferons have been shown to confer anti-tumorigenic effects by inducing immune-mediated tumor cell killing and directly inhibiting tumor cell growth through mediating the upregulation of p21 and p27 in tumor cells (Luth et al., 2011; Shankaran et al., 2001; Thyrell et al., 2002). In contrast, interferons have been shown to contribute to metastasis, epithelial-to-mesenchymal transition, and reduce antitumor immunity (Lane et al., 2018; Lo et al., 2019; Song et al., 2019). Furthermore, there is increasing evidence demonstrating the role of IFN signaling in cancer stem cells (Musella et al., 2022; Song et al., 2019). The activation of interferon response genes in tumor cells suggests that TIC may have adopted the interferon-mediated inflammatory response of the immune system to regulate other processes including self-renewal, and possibly also cell state decisions such as cellular quiescence/activation. Given multicellular organisms arise from a single self-renewing cell, it is also possible that the immune system may have co-opted a stem cell pathway to mediate inflammation in response to infection.

We grouped the genes from our 25-gene TIC signature into groups based on their known functions in the context of interferon-mediated inflammatory response (Supplementary Table 2). We identified genes associated with RNA binding (ISG20, OAS1, OAS2, OAS3, ADAR), cell viability (PLSCR1, IFI6, SHISA5, MT2A), antigen processing (HLA-C, PSME2, PSMB10), the endosomal pathway (IFITM1, IFITM3), cell-cell interaction (BST2, ISG15, GNB4), and transcriptional and/or signaling regulators (SP100, PML, STAT1, PARP9, PARP14, EIF2AK2, IFI44L). These processes may work in combination to contribute to the TIC phenotype or regulate the TIC transcriptional state. For example, PARP9 and PARP14 have been shown to have opposing roles in macrophage activation through ADP-ribosylation of STAT1. PARP14

silencing induces pro-inflammatory genes and STAT1 phosphorylation in IFNγ macrophages, whereas PARP9 silencing suppresses pro-inflammatory genes and STAT1 phosphorylation in IFNγ macrophages thereby cross-regulating macrophage functional states (Iwata et al., 2016). TIC may have co-opted a similar mechanism to regulate cell state decisions such as whether a TIC is in a quiescent or activated state.

The integration of STAT fluorescent reporters and scRNA-seq provided a unique opportunity to identify new candidate TIC markers. We identified BST2 as a candidate cell surface TIC marker, as it was differentially expressed in candidate TIC clusters across the three TNBC models we sequenced, and three out of four patients from publicly available TNBC patient data. BST2 is a transmembrane glycoprotein, whose expression is mediated by IFN, that is well-known for tethering newly-synthesized enveloped virions to the host cell plasma membrane (Van Damme et al., 2008; Venkatesh and Bieniasz, 2013). Furthermore, several studies have demonstrated the role of BST2 in breast cancer cell metastasis, proliferation, and survival (Jones et al., 2013; Mahauad-Fernandez et al., 2015; Mahauad-Fernandez et al., 2018; Sayeed et al., 2013). To our knowledge, BST2 has never been associated with TIC in any tumor type. Here, we demonstrated that FACS enrichment of cells that highly express BST2 can enrich for TIC, and that BST2 knockdown reduced MSFE, indicating BST2 also has a functional role in TIC, at least in SUM159. However, the mechanism by which BST2 regulates TIC function is unknown and can be explored in future studies.

In conclusion, our results provide evidence that interferon response genes are differentially expressed in TIC in a subset of TNBC models and that one of these genes, BST2, is a functional TIC marker in some TNBC models. This study lays the groundwork to mechanistically dissect the role of interferon response genes in TIC and provides candidate genetic targets that may represent vulnerabilities that can be exploited to eliminate TIC.

DECLARATION OF INTERESTS

M.T.L is a Founder of, and an uncompensated Manager in, StemMed Holdings L.L.C., an uncompensated Limited Partner in StemMed Ltd. M.T.L is also a Founder of, and equity holder in, Tvardi Therapeutics. L.E.D. is a compensated employee of StemMed Ltd. Selected BCM PDX models described herein are exclusively licensed to StemMed Ltd. resulting in tangible property royalties to M.T.L. and L.E.D.

ACKNOWLEDGMENTS

This work was supported, in part, by an F31 predoctoral fellowship (NCI CA278274) (to E.P.S), a P30 Cancer Center Support Grant CA125123 (To C. Kent Osborne), a U54 PDXNet PDX Development and Trials Center grant (NCI U54 CA224076) (to M.T.L.), an R01 grant (CA224867 to M.T.L. and H.L.F), a Core Facility Support Grant from the Cancer Research and Prevention Initiative of Texas RP170691 (to M.T.L.), and a generous gift from the Korell family for the study of triple-negative breast cancer. This project was also supported by the Cytometry and Cell Sorting Core at Baylor College of Medicine with funding from the CPRIT Core Facility Support Award (CPRIT-RP180672), the NIH (P30CA125123 and RR024574), the Single Cell Genomics Core at BCM with funding from NIH shared instrument grants (S10OD023469, S10OD025240) and P30EY002520, and the Genomic and RNA Profiling Core at Baylor College of Medicine with funding from NIH NCI (P30CA125123) and CPRIT (RP200504) grants. We would like to acknowledge Mu Wang for cloning the STAT lineage tracing vectors and Connor Hughes for some protein analysis.

MATERIALS & METHODS

STAT lineage tracing (LT) system construction

The STAT reporter (4M67-EGFP-P2A-CreER^{T2}) was generated by LR Gateway® Cloning of a promoter donor vector and a gene donor vector into a destination vector. To generate a flexible promoter donor vector, the EF-1α promoter sequence from the L4/R1 EF-1α promoter donor vector (ThermoFisher Scientific, cat. A11146) was replaced with a 194bp sequence containing multiple restriction digestion sites to facilitate simple restriction cloning of any desired promoter. The STAT-specific promoter donor vector was generated by inserting the STAT signaling-specific 4M67 promoter from the STAT regulatory reporter (4M67-EGFP) (Wei et al., 2014) into the flexible promoter donor vector to generate the L4-4M67-R1 promoter donor vector. To generate the gene donor vector, a CreER^{T2} sequence with attB1/B2 recombination sites and an EcoRI cutting site between attB1 and CreER^{T2} was amplified from pCAG-CreER^{T2} (Matsuda and Cepko, 2007) (a gift from Connie Cepko: Addgene plasmid #14797) by 2-step PCR, then incorporated into the pDonor233 vector (a kind gift from Dr. Kenneth L. Scott, Baylor College of Medicine, Houston, TX) with the P1/P2 compatible recombination sites by BP Gateway cloning. Next, GFP-P2A was amplified from pUltra (Lou et al., 2012) and inserted in front of CreER^{T2} using InFusion cloning to generate L1-EGFP-P2A-CreER^{T2}-L2 gene donor vector. Gateway® Cloning was used to incorporate the 4M67 promoter donor vector and EGFP-P2A-CreER^{T2} gene donor vector into the pLenti6.4/R4R2/V5-Dest destination vector (ThermoFisher Scientific, cat. A11146), generating 4M67-EGFP-P2A-CreER^{T2}.

To generate the dual color-switching lineage tracing reporter (EFS-loxP-dsRed-loxP-Nep), the loxP-dsRed-loxP-GFP sequence with attB1/B2 recombination sites was amplified from pMSCV-loxP-dsRed-loxP-eGFP-Puro-WPRE (Koo et al., 2011) (a gift from Hans Clevers: Addgene plasmid #32702) by 2-step PCR and then incorporated into the pDonor233 vector with the P1/P2 compatible recombination sites by BP Gateway cloning. Gateway® Cloning was used to

incorporate the EFS promoter donor vector and loxP-DsRed-loxP-EGFP gene donor vector into the pLenti6.4/R4R2/V5-Dest destination vector (ThermoFisher Scientific, cat. A11146). mNeptune2 was amplified from the mNeptune2-C1 vector (a kind gift from Dr. Michael Davidson; Addgene plasmid #54836) and used to replace the EGFP sequence by restriction cloning, generating EFS-$_{loxP}$dsRed$_{loxP}$-Nep.

Lentivirus production and transduction

Lineage tracing plasmids and envelope-expressing plasmids were prepared using a plasmid maxi kit (QIAGEN cat. 12162). To produce lentivirus, lineage tracing plasmids and envelope-expressing plasmids were transfected into HEK293T cells using jetPRIME® transfection reagent (Polyplus, cat. 114-15). Lentiviral supernatants were collected every 24 hours for three days, then concentrated. Concentrated viruses were titered using the p24-Elisa kit (Takara Bio, cat. 632200) following manufacturer's protocols. For lentiviral transduction, 5 mg/mL polybrene (Santa Cruz Biotechnology, cat. sc-134220) was added to cell lines prior to lentiviral spinfection at 2500 RPM, 25°C for 30 minutes (MOI of 5). Transduced cells were incubated at 37°C, 5% CO_2 for 24 hours, then lentivirus media was replaced with growth media. To generate LT cell lines (HepG2/C3A STAT LT, SUM159 STAT LT, and MDA-MB-231 STAT LT), 4M67-GFP-P2A-CreER^{T2} was transduced first, and transduced cells were selected using 10 mg/mL blasticidin (Thermo Fisher Scientific, cat. A1113903). Following blasticidin selection, cell lines were then transduced with EFS-$_{loxP}$dsRed$_{loxP}$-Nep. Transduced cells were selected by FACS-enrichment of dsRed^{+} cells. To generate BST2 knockdown cell lines, SUM159 cells were lentiviral transduced with either one of two shRNA vectors (Horizon Discovery, sh-BST2 #1: cat. RHS3979-201824601, sh-BST2 #2: cat. RHS3979-201829090) or one non-targeting control vector (Horizon Discovery, cat. RHS6848) as described above. Transduced cells were selected using 1 μg/mL puromycin (InvivoGen, cat. ant-pr-1).

Xenograft tumors were dissociated (isolation of single cells from human xenograft tumors). The resulting cells were transduced with the STAT regulatory reporter (4M67-GFP) (Wei et al., 2014) overnight in suspension in complete Mammary Epithelial Cell Growth Media ($MEGM^+$) (Lonza, cat. CC-3153, with supplements, cat. CC-4136) containing 20 ng/mL basic fibroblast growth factor (Invitrogen, cat. 13256-029), 10 ng/mL epidermal growth factor (Invitrogen, cat. PHG0311), 4 µg/mL Heparin (StemCell Technologies, cat. 07980), and 400 nL/mL B27 (Invitrogen, cat. 17504-044).

Cell culture and *in vitro* treatment

HepG2/C3A, MDA-MB-231, and HEK293T cells were cultured under adherent conditions with phenol red-free DMEM (Gibco, cat. 31053-028), 10% charcoal-stripped FBS (Gibco, cat. 12-676-029), Penicillin-Streptomycin (GenDEPOT, cat. CA005-010), 1 mM sodium pyruvate (Life Tech, cat. 11360070), and 2 mM L-Glutamine (Thermo Fisher Scientific, cat. 25030081). SUM159 cells were cultured under adherent conditions with phenol red-free Ham's F12 (Caisson labs, cat. HFL05-500ML) containing 5% charcoal-stripped FBS, 5 mg/mL insulin (Gemini Bioproducts, cat. 700-112P), 1 mg/mL hydrocortisone (Sigma-Aldrich, cat. H-4001), and Penicillin-Streptomycin (GenDEPOT, cat. CA005-010). Cells were incubated in a 5% CO_2 incubator at 37°C.

Human recombinant IL-6 (Sigma-Aldrich, cat. I1395) was prepared as a 50 mg/mL stock solution in PBS/1%BSA. Tamoxifen (TAM) (Sigma-Aldrich, cat. T5648) was prepared as a 5 mM solution in 100% ethanol. HepG2/C3A STAT LT cells were treated with either IL-6 (50 ng/mL), TAM (1 µM), or the combination and the cells were visualized by confocal microscopy (Leica, TCS-SP5) and quantified by flow cytometry (BD LSRII analyzer) after 72 hours.

Generation of cell line-derived and patient-derived xenografts

To prepare for cell injection, SUM159 STAT LT or MDA-MB-231 STAT LT cells were suspended in a 1:1 solution of their respective growth media and phenol red-free Matrigel® matrix (Corning,

cat. 356237). Cells were transplanted into the #4 epithelium-free “cleared” fat pad of 3–4-week-old SCID/Beige mice (Charles River Laboratories, Wilmington, MA, http://www.criver.com). When palpable tumors formed, mice were administered TAM by intraperitoneal injection to induce dsRed to mNeptune2 color switching in STAT signaling ($EGFP^+$) cells. TAM was prepared by dissolving in sunflower seed oil (Sigma-Aldrich, cat. S5007) and stirring overnight. All animal protocols were reviewed and approved by the Animal Protocol Review Committee at Baylor College of Medicine.

Isolation of single cells from human xenograft tumors

Xenograft tumors were harvested and digested using 5mL digestion media (phenol red-free DMEM/F12 [Corning cat. 16-405-CV], 100 mg/mL gentamicin [Worthington, cat. LS004176], 250 U/mL collagenase II [Worthington cat. LS004174] and 500 kunitz/mL DNase I [Stemcell, cat. 07900]) per 1 g tumor at 37°C in a gentleMACS dissociator (Miltenyi). The tumor digestion was passed through a 200 mM filter and centrifuged at 300 g for 1 minute. Red blood cells were lysed by adding 9 mL ddH_2O to the cell pellet for 20 seconds and lysis was stopped by the addition of 1 mL 10X PBS. To further dissociate cell clumps, the cell suspension was centrifuged at 300 g for 1 minute, then resuspended in TrypLE Express (Gibco, cat. 12605-010) for 5 minutes. To stop the digestion, 9 mL of PBS was added, and the cell suspension was passed through a 40 mm cell strainer.

Analytical flow cytometry and FACS

Cells from adherent cultures were trypsinized and pelleted by centrifuging at 300 g for 3 minutes. The cells were resuspended in $HBSS^+$ (HBSS [Cytiva Life Sciences cat. SH30588.01], 2% FBS). Single cells from digested xenograft tumors were labeled using a BUV395 anti-mouse cocktail (pooled and diluted 1:40: anti-mouse H2kD [BD Biosciences, cat. 742437], anti-mouse CD45 [BD Biosciences, cat. 567451], anti-mouse CD31 [BD Biosciences, cat. 740231], and anti-mouse Ter119 [BD Biosciences, cat. 563827]) to exclude mouse cells. A far-red BST2

antibody was used to label BST2-expressing cells (diluted 1:40, BioLegend, cat. 348416). Cell viability was determined by Sytox Blue (diluted 1:1000, V450, Invitrogen, cat. S34857) staining. Cells were applied to a BD LSRII analyzer for analytical flow cytometry or a BD AriaII for fluorescence-activated cell sorting (FACS). Quantification of each cell population was performed using FlowJoe (BD Biosciences) and statistical differences between cell populations were determined by one-way ANOVA followed by post-hoc Tukey's test, which was calculated using GraphPad Prism version 9 (GraphPad Software, San Diego, CA, http://www.graphpad.com).

Immunohistochemistry

Immunohistochemistry (IHC) was performed by the Lester and Sue Smith Breast Center Pathology Core. Tumors were fixed in 4% paraformaldehyde, paraffin-embedded, sectioned at 3–4 µm, and deparaffinized. The antigen retrieval buffer for pSTAT1 and pSTAT3 staining was 0.1 M Tris-HCL PH 9.0, respectively. Sections were then blocked in 3% hydrogen peroxide solution for 5 minutes and incubated by pSTAT1 Ser727 antibody (Cell Signaling Technology, cat. 8826, 1:200), or pSTAT3 Y705 antibody (Cell Signaling Technology, cat. 9145, 1:25), followed by Signalstain Boost IHC Detection Reagent (HRP, Rabbit) (Cell Signaling Technology, cat. 8114) for 30 minutes, visualized by 3, 3′ - Diaminobenzidine (DAB+) solution (DakoCytomation, Glostrup, Denmark), and counterstained by Harris hematoxylin.

Mammosphere forming efficiency assay

The protocol for mammosphere formation efficiency assays has been described previously (Dontu et al., 2003). Following tumor digestion and FACS enrichment of fluorescent cell populations, equal numbers of cells from each fraction were plated in complete Mammary Epithelial Cell Growth Media ($MEGM^{+}$) (Lonza, cat. CC-3153, with supplements, cat. CC-4136) containing 20 ng/mL basic fibroblast growth factor (Invitrogen, cat. 13256-029), 10 ng/mL epidermal growth factor (Invitrogen, cat. PHG0311), 4 µg/mL Heparin (StemCell Technologies, cat. 07980), and 400 nL/mL B27 (Invitrogen, cat. 17504-044) with 0.6% HPMC (Millipore Sigma,

cat. M7027) to prevent cell aggregation, in an ultra-low attachment plate. To prepare secondary mammospheres, primary mammospheres were collected and dissociated into single cells using TrypLE Express (Gibco, cat. 12605-010) for 5 minutes. To inactivate the TrypLE Express, 9 mL of PBS was added, and the cell suspension was passed through a 40 mm cell strainer. Then, an equal number of cells from each fraction were plated in $MEGM^+$ media with 0.6% HPMC. Secondary mammospheres were imaged on a Yokogawa CV8000, and CellPathfinder (Yokogawa) was used to quantify the number of mammospheres in each well. Spheres with an area greater than 300 mm^2 and a diameter greater than 50 mm were used to determine mammospheres. A BioTek Cytation 5 (Agilent, Santa Clara, CA) was used to take representative images. Statistical differences in mammosphere forming efficiency between cell populations were determined by one-way ANOVA followed by post-hoc Tukey's test, which was calculated using GraphPad Prism version 9 (GraphPad Software, San Diego, CA, http://www.graphpad.com).

Limiting-dilution transplantation assay

FACS-enriched cells were suspended in growth media at a concentration of 5,000 cells per 10ml and serially diluted using growth media. Five dilutions were prepared (5,000, 1,000, 200, 40 cells per 10ml), then phenol red-free Matrigel® matrix (Corning, cat. 356237) was added for a final 1:1 solution of growth media and Matrigel® prior to transplantation as described in "Generation of cell line-derived and patient-derived xenografts." The mice were monitored until the first tumor reached 10% body weight (~7-8 weeks) at which point the experiment was terminated and tumor outgrowth was recorded. Extreme limiting-dilution analysis was used to calculate the TIC frequency of each population (Hu and Smyth, 2009).

scRNA-seq data preprocessing and filtering

FASTQ files were mapped to the GRCh38 reference genome using the Cell Ranger Single Cell software v.6.1.2 (SUM159) or v7.0.1 (BCM-4272 and BCM-15006) (10X Genomics). Using

Seurat v4.1.3 in R v4.2.2, the count matrix of each sample was used to create a Seurat object. Cells expressing greater than 1000 and less than 9000 genes. Cells with greater than 10% mitochondrial transcripts or 50% ribosomal transcripts were excluded as they represented low quality or dying cells. Next, fluorescent samples from the same tumor were merged (ex. BCM-4272 EGFP+ and BCM-4272 EGFP- merged to BCM-4272) into a single Seurat object. To exclude the effects of cell cycle heterogeneity in the analysis, we used the CellCycleScoring function from Seurat. This function gives each cell a quantitative score for G1, G2/M, and S phases based on the expression of cell cycle phase genes (Kowalczyk et al., 2015). Cells with a predicted classification of G2M or S phase were regressed during data scaling.

scRNA-seq clustering and UMAP visualization

Seurat v4.1.3 was used for data normalization, dimensionality reduction, and clustering (Hao et al., 2021). The filtered gene matrices from each merged sample were normalized using the NormalizeData function. The normalized datasets were centered and scaled using ScaleData. Then, the most significant principal components were found using RunPCA, and ElbowPlot was used to determine the number of principal components to use for clustering each dataset. Different resolution parameters for unsupervised clustering were examined to find the optimal number of clusters based on clear segregation of clusters. For each dataset, the first 35 principal components were used for unsupervised clustering with a resolution = 0.1 for SUM159 (yielding nine clusters) and a resolution of 0.5 for BCM-4272 and BCM-15006 (yielding seven clusters and six clusters, respectively) using the FindNeighbors and FindClusters functions. To visualize the clustering for each dataset, the dimensionality was further reduced using the Seurat function RunUMAP. The same principal components used to determine the clustering were the same used to calculate the UMAP embedding. Each resulting cluster was further analyzed for potential doublets or low-quality cells using two methods: quality metrics (nCount_RNA, nFeature_RNA, and mitochondrial content) for each cluster were calculated and

any outlier values (clusters with greater, or less than, two standard deviations than the average of all clusters) were removed.

scRNA-seq lineage trajectory construction using Monocle3

To preserve the UMAP structure of each dataset for lineage trajectory construction, each Seurat object was converted to a cell data set using the as.cell_data_set function. Monocle3 was then used to construct lineage trajectories for each (Cao et al., 2019; Qiu et al., 2017; Trapnell et al., 2014). Next, the data were pre-processed using the preprocess_cds function using 100 principal components to capture the maximum variation within each dataset. Next, cluster_cells was used to assign clusters in each dataset using Louvain/Leiden community detection (Levine et al., 2015), and the learn_graph function was used to order cells in pseudotime by marking the root of the trajectory. Given that TIC sit atop the tumor cell hierarchy, proliferating and differentiating to give rise to bulk tumor cells (Cole et al., 2020; Liu et al., 2014), lineage trajectories were constructed with the hypothesis that a TIC cluster must be predominantly $EGFP^+$ (or an equal distribution for BCM-15006) and lie at the end of a trajectory. For SUM159, cluster 7 was selected as the root of the trajectory because it was the only cluster where $EGFP^+$ cells (which are enriched for TIC) were overrepresented. Cluster 6 was selected as the root of the trajectory for BCM-15006 because it was the only cluster with equal representation of $EGFP^+$ and $EGFP^-$ cells (in a model where EGFP reporter activity did not enrich or deplete TIC). For BCM-4272, cluster 7 was selected as the root of the trajectory because $EGFP^+$ cells (which are enriched for TIC) were overrepresented and the resulting trajectory ended in a cluster of $EGFP^-$ cells (which had less MSFE and therefore may represent differentiated bulk tumor cells). Constructed lineage trajectories were visualized using the plot_cells function from Monocle3.

scRNA-seq differential gene expression analysis and intersection of gene lists

To examine the different transcriptional states of tumor cells, FindAllMarkers was used to identify differentially expressed genes in each cluster using the “bimod” likelihood-ratio test

framework. Two additional parameters were used to include differentially expressed genes: genes that showed at least a 0.25 log2 fold expression difference, and genes that were expressed in at least 25% of the cells in a cluster. To compare the similarities between candidate TIC transcriptional states in different xenograft models, upregulated genes from each candidate TIC cluster were intersected using GeneOverlap (Shen, 2023).

scRNA-seq gene set enrichment analysis

To associate a biological phenotype with each cluster, gene set enrichment analysis was performed using EnrichR (Xie et al., 2021). A Wilcoxon test was used to determine significance ($p<0.05$) and genes with a log fold change threshold of 0.25 were included in the analysis. To view the enrichment of interferon gene sets across all clusters in a dataset simultaneously, Escape ssGSEA was used (Borcherding et al., 2021) to calculate gene set enrichment scores for each cluster.

Analysis of patient scRNA-seq data

Publicly available human breast cancer datasets were retrieved through GEO (GSE176078 and GSE161529) (Pal et al., 2021; Wu et al., 2021). Seurat v4.1.3 was used for cell cycle regression, data normalization, dimensionality reduction, and clustering, as described above. To identify clusters of epithelial cells, the expression of keratins (KRT8, KRT14, KRT18, and KRT19) was evaluated. Clusters with no keratin expression were filtered, and the epithelial dataset was re-normalized and clustered using Seurat v4.1.3. UCell was used to score each cluster based on similarity the 25 genes identified by intersecting SUM159-C7, 4272-C7, and 15006-C6 (Andreatta and Carmona, 2021). To identify differentially expressed genes between clusters, FindAllMarkers from Seurat v4.1.3 was used, then EnrichR was used to perform gene set enrichment analysis for each cluster.

Chemotherapy treatments

PDX-bearing mice (BCM-4272 and BCM-15006) were treated with four weekly cycles of docetaxel (20 mg/kg, IP) (low human equivalent dose), carboplatin (50 mg/kg, IP) (human equivalent dose AUC6), or left untreated for four weeks. Tumor volume was assessed twice weekly using calipers. Untreated tumors and residual tumors were harvested after four weekly cycles of chemotherapy treatment. Harvested tumors were cut into 10–20 mg chunks, which were snap-frozen in liquid nitrogen and stored at -80°C.

qRT-PCR

RNA extraction was performed using the RNeasy plus mini kit (Qiagen, cat. 74134) according to the manufacturer's protocol. Next, we performed cDNA synthesis using SuperScript™ III first-strand synthesis system (ThermoFisher Scientific, cat. 18080051) according to the manufacturer's protocol. Then, qPCR was performed using iTaq qPCR supermix (BioRad, cat. 1725130). Briefly, 10 µL 2X Taq, 9 µL cDNA, and 1 µL of a 20X TaqMan probe (ThermoFisher Scientific, cat. Hs00171632_m1 (BST2), Hs00241666_m1 (ADAR), Hs00942643_m1 (OAS2), Hs00967084_m1 (PARP9), Hs00393816_m1 (PARP14), Hs01013996_m1 (STAT1), Hs99999901_s1 (18S)) were added per reaction. The qPCR reaction was performed in a C1000-Touch-Thermocycler (BioRad) (Initial denaturation: 95°C 30 seconds, followed by 39 cycles of denaturation: 95°C 5 seconds and annealing/extension 60°C 30 seconds). Samples were normalized to the expression of a housekeeping gene, 18S. Statistical differences in BST2 expression following chemotherapy were determined for each PDX independently by one-way ANOVA followed by post-hoc Tukey's test. Statistical differences in the expression of interferon response genes were determined for each model independently by one-way ANOVA followed by post-hoc Tukey's test. All statistical differences were calculated using GraphPad Prism version 9 (GraphPad Software, San Diego, CA, http://www.graphpad.com).

REFERENCES

Abd El-Rehim, D.M., Ball, G., Pinder, S.E., Rakha, E., Paish, C., Robertson, J.F., Macmillan, D., Blamey, R.W., and Ellis, I.O. (2005). High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer *116*, 340-350.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A *100*, 3983-3988.

Alamgeer, M., Ganju, V., Kumar, B., Fox, J., Hart, S., White, M., Harris, M., Stuckey, J., Prodanovic, Z., Schneider-Kolsky, M.E.*, et al.* (2014). Changes in aldehyde dehydrogenase-1 expression during neoadjuvant chemotherapy predict outcome in locally advanced breast cancer. Breast Cancer Res *16*, R44.

Aleskandarany, M.A., Agarwal, D., Negm, O.H., Ball, G., Elmouna, A., Ashankyty, I., Nuglozeh, E., Fazaludeen, M.F., Diez-Rodriguez, M., Nolan, C.C.*, et al.* (2016). The prognostic significance of STAT3 in invasive breast cancer: analysis of protein and mRNA expressions in large cohorts. Breast Cancer Res Treat *156*, 9-20.

Andreatta, M., and Carmona, S.J. (2021). UCell: Robust and scalable single-cell gene signature scoring. Comput Struct Biotechnol J *19*, 3796-3798.

Batlle, E., and Clevers, H. (2017). Cancer stem cells revisited. Nat Med *23*, 1124-1134.

Borcherding, N., Vishwakarma, A., Voigt, A.P., Bellizzi, A., Kaplan, J., Nepple, K., Salem, A.K., Jenkins, R.W., Zakharia, Y., and Zhang, W. (2021). Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. Commun Biol *4*, 122.

Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61-70.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J.*, et al.* (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496-502.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885-896.

Chung, S.S., Aroh, C., and Vadgama, J.V. (2013). Constitutive activation of STAT3 signaling regulates hTERT and promotes stem cell-like traits in human breast cancer cells. PLoS One *8*, e83971.

Chung, S.S., Giehl, N., Wu, Y., and Vadgama, J.V. (2014). STAT3 activation in HER2-overexpressing breast cancer promotes epithelial-mesenchymal transition and cancer stem cell traits. Int J Oncol *44*, 403-411.

Clarkson, R.W., Boland, M.P., Kritikou, E.A., Lee, J.M., Freeman, T.C., Tiffen, P.G., and Watson, C.J. (2006). The genes induced by signal transducer and activators of transcription

(STAT)3 and STAT5 in mammary epithelial cells define the roles of these STATs in mammary development. Mol Endocrinol *20*, 675-685.

Colacino, J.A., Azizi, E., Brooks, M.D., Harouaka, R., Fouladdel, S., McDermott, S.P., Lee, M., Hill, D., Madden, J., Boerner, J.*, et al.* (2018). Heterogeneity of Human Breast Stem and Progenitor Cells as Revealed by Transcriptional Profiling. Stem Cell Reports *10*, 1596-1609.

Cole, A.J., Fayomi, A.P., Anyaeche, V.I., Bai, S., and Buckanovich, R.J. (2020). An evolving paradigm of cancer stem cell hierarchies: therapeutic implications. Theranostics *10*, 3083-3098.

Creighton, C.J., Li, X., Landis, M., Dixon, J.M., Neumeister, V.M., Sjolund, A., Rimm, D.L., Wong, H., Rodriguez, A., Herschkowitz, J.I.*, et al.* (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc Natl Acad Sci U S A *106*, 13820-13825.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y.*, et al.* (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346-352.

Dalerba, P., Cho, R.W., and Clarke, M.F. (2007). Cancer stem cells: models and concepts. Annu Rev Med *58*, 267-284.

Dierks, C., Beigi, R., Guo, G.R., Zirlik, K., Stegert, M.R., Manley, P., Trussell, C., Schmitt-Graeff, A., Landwerlin, K., Veelken, H.*, et al.* (2008). Expansion of Bcr-Abl-positive leukemic stem cells is dependent on Hedgehog pathway activation. Cancer Cell *14*, 238-249.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev *17*, 1253-1270.

Driessens, G., Beck, B., Caauwe, A., Simons, B.D., and Blanpain, C. (2012). Defining the mode of tumour growth by clonal analysis. Nature *488*, 527-530.

Ehret, G.B., Reichenbach, P., Schindler, U., Horvath, C.M., Fritz, S., Nabholz, M., and Bucher, P. (2001). DNA binding specificity of different STAT proteins. Comparison of in vitro specificity with natural target sites. J Biol Chem *276*, 6675-6688.

Gerlinger, M., Rowan, A.J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A.*, et al.* (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med *366*, 883-892.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S.*, et al.* (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555-567.

Gocher, A.M., Workman, C.J., and Vignali, D.A.A. (2022). Interferon-gamma: teammate or opponent in the tumour microenvironment? Nat Rev Immunol *22*, 158-172.

Guo, C., Yang, W., and Lobe, C.G. (2002). A Cre recombinase transgene with mosaic, widespread tamoxifen-inducible action. Genesis *32*, 8-18.

Hammond, M.E., Hayes, D.F., Wolff, A.C., Mangu, P.B., and Temin, S. (2010). American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Oncol Pract *6*, 195-197.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M.*, et al.* (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587 e3529.

Harrison, H., Farnie, G., Howell, S.J., Rock, R.E., Stylianou, S., Brennan, K.R., Bundred, N.J., and Clarke, R.B. (2010). Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. Cancer Res *70*, 709-718.

Horvath, C.M., Wen, Z., and Darnell, J.E., Jr. (1995). A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain. Genes Dev *9*, 984-994.

Hu, Y., and Smyth, G.K. (2009). ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J Immunol Methods *347*, 70-78.

Humphreys, R.C., Bierie, B., Zhao, L., Raz, R., Levy, D., and Hennighausen, L. (2002). Deletion of Stat3 blocks mammary gland involution and extends functional competence of the secretory epithelium in the absence of lactogenic stimuli. Endocrinology *143*, 3641-3650.

Iliopoulos, D., Hirsch, H.A., Wang, G., and Struhl, K. (2011). Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. Proc Natl Acad Sci U S A *108*, 1397-1402.

Ishizawa, K., Rasheed, Z.A., Karisch, R., Wang, Q., Kowalski, J., Susky, E., Pereira, K., Karamboulas, C., Moghal, N., Rajeshkumar, N.V.*, et al.* (2010). Tumor-initiating cells are rare in many human tumors. Cell Stem Cell *7*, 279-282.

Iwata, H., Goettsch, C., Sharma, A., Ricchiuto, P., Goh, W.W., Halu, A., Yamada, I., Yoshida, H., Hara, T., Wei, M.*, et al.* (2016). PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. Nat Commun *7*, 12849.

Jiang, G., Tu, J., Zhou, L., Dong, M., Fan, J., Chang, Z., Zhang, L., Bian, X., and Liu, S. (2021). Single-cell transcriptomics reveal the heterogeneity and dynamic of cancer stem-like cells during breast tumor progression. Cell Death Dis *12*, 979.

Jones, L.M., Broz, M.L., Ranger, J.J., Ozcelik, J., Ahn, R., Zuo, D., Ursini-Siegel, J., Hallett, M.T., Krummel, M., and Muller, W.J. (2016). STAT3 Establishes an Immunosuppressive Microenvironment during the Early Stages of Breast Carcinogenesis to Promote Tumor Growth and Metastasis. Cancer Res *76*, 1416-1428.

Jones, P.H., Mahauad-Fernandez, W.D., Madison, M.N., and Okeoma, C.M. (2013). BST-2/tetherin is overexpressed in mammary gland and tumor tissues in MMTV-induced mammary cancer. Virology *444*, 124-139.

Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massague, J. (2003). A multigenic program mediating breast cancer metastasis to bone. Cancer Cell *3*, 537-549.

Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., and Navin, N.E. (2018). Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell *173*, 879-893 e813.

Koo, B.K., Stange, D.E., Sato, T., Karthaus, W., Farin, H.F., Huch, M., van Es, J.H., and Clevers, H. (2011). Controlled gene expression in primary Lgr5 organoid cultures. Nat Methods *9*, 81-83.

Korkaya, H., Kim, G.I., Davis, A., Malik, F., Henry, N.L., Ithimakin, S., Quraishi, A.A., Tawakkol, N., D'Angelo, R., Paulson, A.K.*, et al.* (2012). Activation of an IL6 inflammatory loop mediates trastuzumab resistance in HER2+ breast cancer by expanding the cancer stem cell population. Mol Cell *47*, 570-584.

Koromilas, A.E., and Sexl, V. (2013). The tumor suppressor function of STAT1 in breast cancer. JAKSTAT *2*, e23353.

Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res *25*, 1860-1872.

Lane, R.S., Femel, J., Breazeale, A.P., Loo, C.P., Thibault, G., Kaempf, A., Mori, M., Tsujikawa, T., Chang, Y.H., and Lund, A.W. (2018). IFNgamma-activated dermal lymphatic vessels inhibit cytotoxic T cells in melanoma and inflamed skin. J Exp Med *215*, 3057-3074.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest *121*, 2750-2767.

Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir el, A.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R.*, et al.* (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell *162*, 184-197.

Li, X., Lewis, M.T., Huang, J., Gutierrez, C., Osborne, C.K., Wu, M.F., Hilsenbeck, S.G., Pavlick, A., Zhang, X., Chamness, G.C.*, et al.* (2008). Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst *100*, 672-679.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res *47*, W199-W205.

Lin, L., Liu, A., Peng, Z., Lin, H.J., Li, P.K., Li, C., and Lin, J. (2011). STAT3 is necessary for proliferation and survival in colon cancer-initiating cells. Cancer Res *71*, 7226-7237.

Liu, H., Patel, M.R., Prescher, J.A., Patsialou, A., Qian, D., Lin, J., Wen, S., Chang, Y.F., Bachmann, M.H., Shimono, Y.*, et al.* (2010). Cancer stem cells from human breast tumors are involved in spontaneous metastases in orthotopic mouse models. Proc Natl Acad Sci U S A *107*, 18115-18120.

Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., Martin-Trevino, R., Shang, L., McDermott, S.P., Landis, M.D.*, et al.* (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports *2*, 78-91.

Lo, U.G., Pong, R.C., Yang, D., Gandee, L., Hernandez, E., Dang, A., Lin, C.J., Santoyo, J., Ma, S., Sonavane, R.*, et al.* (2019). IFNgamma-Induced IFIT5 Promotes Epithelial-to-Mesenchymal Transition in Prostate Cancer via miRNA Processing. Cancer Res *79*, 1098-1112.

Lou, E., Fujisawa, S., Morozov, A., Barlas, A., Romin, Y., Dogan, Y., Gholami, S., Moreira, A.L., Manova-Todorova, K., and Moore, M.A. (2012). Tunneling nanotubes provide a unique conduit for intercellular transfer of cellular contents in human malignant pleural mesothelioma. PLoS One *7*, e33093.

Luth, S., Schrader, J., Zander, S., Carambia, A., Buchkremer, J., Huber, S., Reifenberg, K., Yamamura, K., Schirmacher, P., Lohse, A.W.*, et al.* (2011). Chronic inflammatory IFN-gamma signaling suppresses hepatocarcinogenesis in mice by sensitizing hepatocytes for apoptosis. Cancer Res *71*, 3763-3771.

Mahauad-Fernandez, W.D., Borcherding, N.C., Zhang, W., and Okeoma, C.M. (2015). Bone marrow stromal antigen 2 (BST-2) DNA is demethylated in breast tumors and breast cancer cells. PLoS One *10*, e0123931.

Mahauad-Fernandez, W.D., DeMali, K.A., Olivier, A.K., and Okeoma, C.M. (2014). Bone marrow stromal antigen 2 expressed in cancer cells promotes mammary tumor growth and metastasis. Breast Cancer Res *16*, 493.

Mahauad-Fernandez, W.D., Naushad, W., Panzner, T.D., Bashir, A., Lal, G., and Okeoma, C.M. (2018). BST-2 promotes survival in circulation and pulmonary metastatic seeding of breast cancer cells. Sci Rep *8*, 17608.

Mahauad-Fernandez, W.D., and Okeoma, C.M. (2016). The role of BST-2/Tetherin in host protection and disease manifestation. Immun Inflamm Dis *4*, 4-23.

Marotta, L.L., Almendro, V., Marusyk, A., Shipitsin, M., Schemme, J., Walker, S.R., Bloushtain-Qimron, N., Kim, J.J., Choudhury, S.A., Maruyama, R.*, et al.* (2011). The JAK2/STAT3 signaling pathway is required for growth of CD44(+)CD24(-) stem cell-like breast cancer cells in human tumors. J Clin Invest *121*, 2723-2735.

Massague, J., and Ganesh, K. (2021). Metastasis-Initiating Cells and Ecosystems. Cancer Discov *11*, 971-994.

Matsuda, T., and Cepko, C.L. (2007). Controlled expression of transgenes introduced by in vivo electroporation. Proc Natl Acad Sci U S A *104*, 1027-1032.

Meyer, M.J., Fleming, J.M., Ali, M.A., Pesesky, M.W., Ginsburg, E., and Vonderhaar, B.K. (2009). Dynamic regulation of CD24 and the invasive, CD44posCD24neg phenotype in breast cancer cell lines. Breast Cancer Res *11*, R82.

Musella, M., Guarracino, A., Manduca, N., Galassi, C., Ruggiero, E., Potenza, A., Maccafeo, E., Manic, G., Mattiello, L., Soliman Abdel Rehim, S.*, et al.* (2022). Type I IFNs promote cancer cell stemness by triggering the epigenetic regulator KDM1B. Nat Immunol *23*, 1379-1392.

Nagao, T., Kinoshita, T., Hojo, T., Tsuda, H., Tamura, K., and Fujiwara, Y. (2012). The differences in the histological types of breast cancer and the response to neoadjuvant

chemotherapy: the relationship between the outcome and the clinicopathological characteristics. Breast *21*, 289-295.

Neil, S.J., Zang, T., and Bieniasz, P.D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature *451*, 425-430.

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F.*, et al.* (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell *10*, 515-527.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev *12*, 2048-2060.

Pal, B., Chen, Y., Vaillant, F., Capaldo, B.D., Joyce, R., Song, X., Bryant, V.L., Penington, J.S., Di Stefano, L., Tubau Ribera, N.*, et al.* (2021). A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. EMBO J *40*, e107333.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A.*, et al.* (2000). Molecular portraits of human breast tumours. Nature *406*, 747-752.

Petrosyan, V., Dobrolecki, L.E., Thistlethwaite, L., Lewis, A.N., Sallas, C., Srinivasan, R.R., Lei, J.T., Kovacevic, V., Obradovic, P., Ellis, M.J.*, et al.* (2023). Identifying biomarkers of differential chemotherapy response in TNBC patient-derived xenografts with a CTD/WGCNA approach. iScience *26*, 105799.

Philp, J.A., Burdon, T.G., and Watson, C.J. (1996). Differential activation of STATs 3 and 5 during mammary gland development. FEBS Lett *396*, 77-80.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat Methods *14*, 979-982.

Reid, J.M., Goetz, M.P., Buhrow, S.A., Walden, C., Safgren, S.L., Kuffel, M.J., Reinicke, K.E., Suman, V., Haluska, P., Hou, X.*, et al.* (2014). Pharmacokinetics of endoxifen and tamoxifen in female mice: implications for comparative in vivo activity studies. Cancer Chemother Pharmacol *74*, 1271-1278.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105-111.

Sayeed, A., Luciani-Torres, G., Meng, Z., Bennington, J.L., Moore, D.H., and Dairkee, S.H. (2013). Aberrant regulation of the BST2 (Tetherin) promoter enhances cell proliferation and apoptosis evasion in high grade breast cancer cells. PLoS One *8*, e67191.

Shankaran, V., Ikeda, H., Bruce, A.T., White, J.M., Swanson, P.E., Old, L.J., and Schreiber, R.D. (2001). IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature *410*, 1107-1111.

Shen, L. (2023). GeneOverlap: Test and visualize gene overlaps. R Package *1*.

Song, M., Ping, Y., Zhang, K., Yang, L., Li, F., Zhang, C., Cheng, S., Yue, D., Maimela, N.R., Qu, J.*, et al.* (2019). Low-Dose IFNgamma Induces Tumor Cell Stemness in Tumor Microenvironment of Non-Small Cell Lung Cancer. Cancer Res *79*, 3737-3748.

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B.*, et al.* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst *98*, 262-272.

Tan, B.T., Park, C.Y., Ailles, L.E., and Weissman, I.L. (2006). The cancer stem cell hypothesis: a work in progress. Lab Invest *86*, 1203-1207.

Thyrell, L., Erickson, S., Zhivotovsky, B., Pokrovskaja, K., Sangfelt, O., Castro, J., Einhorn, S., and Grander, D. (2002). Mechanisms of Interferon-alpha induced apoptosis in malignant cells. Oncogene *21*, 1251-1262.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol *32*, 381-386.

Van Damme, N., Goff, D., Katsura, C., Jorgenson, R.L., Mitchell, R., Johnson, M.C., Stephens, E.B., and Guatelli, J. (2008). The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell Host Microbe *3*, 245-252.

Venkatesh, S., and Bieniasz, P.D. (2013). Mechanism of HIV-1 virion entrapment by tetherin. PLoS Pathog *9*, e1003483.

Visvader, J.E., and Lindeman, G.J. (2008). Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nat Rev Cancer *8*, 755-768.

Visvader, J.E., and Lindeman, G.J. (2012). Cancer stem cells: current status and evolving complexities. Cell Stem Cell *10*, 717-728.

Wang, T., Fahrmann, J.F., Lee, H., Li, Y.J., Tripathi, S.C., Yue, C., Zhang, C., Lifshitz, V., Song, J., Yuan, Y.*, et al.* (2018). JAK/STAT3-Regulated Fatty Acid beta-Oxidation Is Critical for Breast Cancer Stem Cell Self-Renewal and Chemoresistance. Cell Metab *27*, 136-150 e135.

Wang, Y., Wang, Q., Tang, C.H., Chen, H.D., Hu, G.N., Shao, J.K., Dong, X.F., Jin, L.L., and Wang, C.Q. (2021). p-STAT3 expression in breast cancer correlates negatively with tumor size and HER2 status. Medicine (Baltimore) *100*, e25124.

Wei, W., Tweardy, D.J., Zhang, M., Zhang, X., Landua, J., Petrovic, I., Bu, W., Roarty, K., Hilsenbeck, S.G., Rosen, J.M.*, et al.* (2014). STAT3 signaling is activated preferentially in tumor-initiating cells in claudin-low models of human breast cancer. Stem Cells *32*, 2571-2582.

Widschwendter, A., Tonko-Geymayer, S., Welte, T., Daxenbichler, G., Marth, C., and Doppler, W. (2002). Prognostic significance of signal transducer and activator of transcription 1 activation in breast cancer. Clin Cancer Res *8*, 3065-3074.

Wolff, A.C., Hammond, M.E., Hicks, D.G., Dowsett, M., McShane, L.M., Allison, K.H., Allred, D.C., Bartlett, J.M., Bilous, M., Fitzgibbons, P.*, et al.* (2013). Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical

Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol *31*, 3997-4013.

Wu, S.Z., Al-Eryani, G., Roden, D.L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J.R., Bartonicek, N.*, et al.* (2021). A single-cell and spatially resolved atlas of human breast cancers. Nat Genet *53*, 1334-1347.

Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M.*, et al.* (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90.

Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J.*, et al.* (2007). let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell *131*, 1109-1123.

Zhang, M., Atkinson, R.L., and Rosen, J.M. (2010). Selective targeting of radiation-resistant tumor-initiating cells. Proc Natl Acad Sci U S A *107*, 3522-3527.

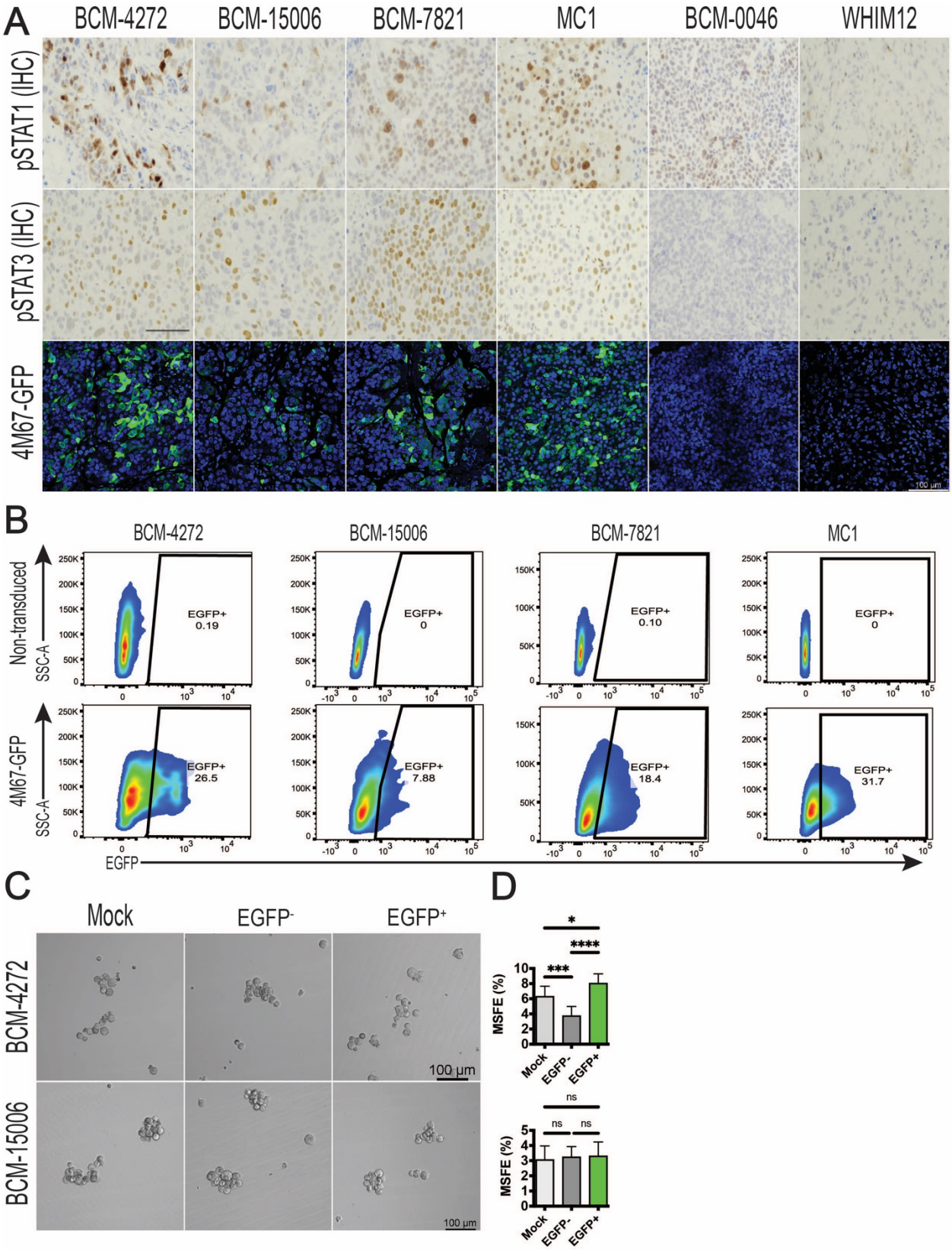


Figure 1: STAT reporter activity in breast PDX models. A) Representative immunohistochemistry (IHC) images showing levels of pSTAT1 and pSTAT3 in a panel of patient-

derived xenograft (PDX) models. 4M67-EGFP reporter activity was visualized by immunofluorescence. Scale bar = 50 μm. B) Flow cytometry plot illustrating the percentage of 4M67 reporter-positive cells in PDX models with demonstrated reporter activity. C) Representative images comparing the mammosphere forming efficiency between the $EGFP^+$, $EGFP^-$, and mock-sorted cells in two PDX models (BCM-4272: $n = 10$, BCM-15006: $n = 9$) and D) a bar graph of the mammosphere forming efficiency. Values are mean ± SD. One-way ANOVA and Tukey's post hoc test were used to determine statistical significance. *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$. MSFE: mammosphere forming efficiency.

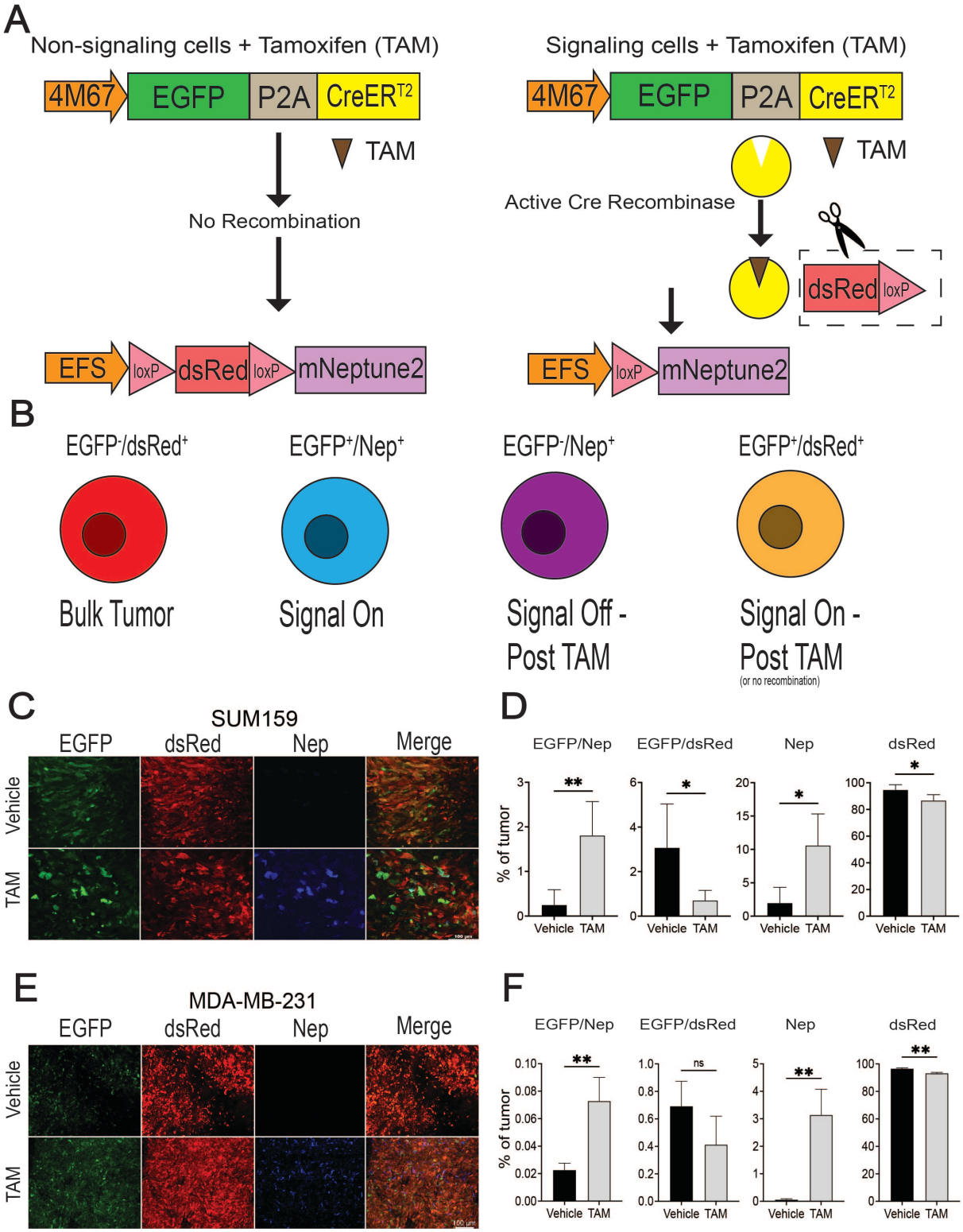


Figure 2: STAT lineage tracing reporter provides a long-term fluorescent label to STAT signaling cells. A) Schematic illustrating how the STAT lineage tracing system functions upon the addition of 4-OH-tamoxifen (TAM). Cre is not transcribed in non-signaling (EGFP$^-$) cells, thus no color-switching occurs in the presence of TAM. The Cre is transcribed in signaling (EGFP$^+$) cells; therefore, the addition of TAM drives color-switching from dsRed to mNeptune2. B) Graphic

illustrating the four different lineage tracing (LT) cell populations that can be identified by microscopy and flow cytometry. C) Confocal images of SUM159 STAT LT xenografts and D) flow cytometry quantification of the percent of each LT population two days after 4 mg TAM treatment (n = 4). E) Confocal images of MDA-MB-231 STAT LT xenografts and F) flow cytometry quantification of the percent of each LT population two days after 4 mg TAM treatment (n = 3). One-way ANOVA and Tukey's post hoc test were used to determine statistical significance. *$p < 0.05$, **$p < 0.01$, ns = not significant; values are mean ± SD. Nep: mNeptune2.

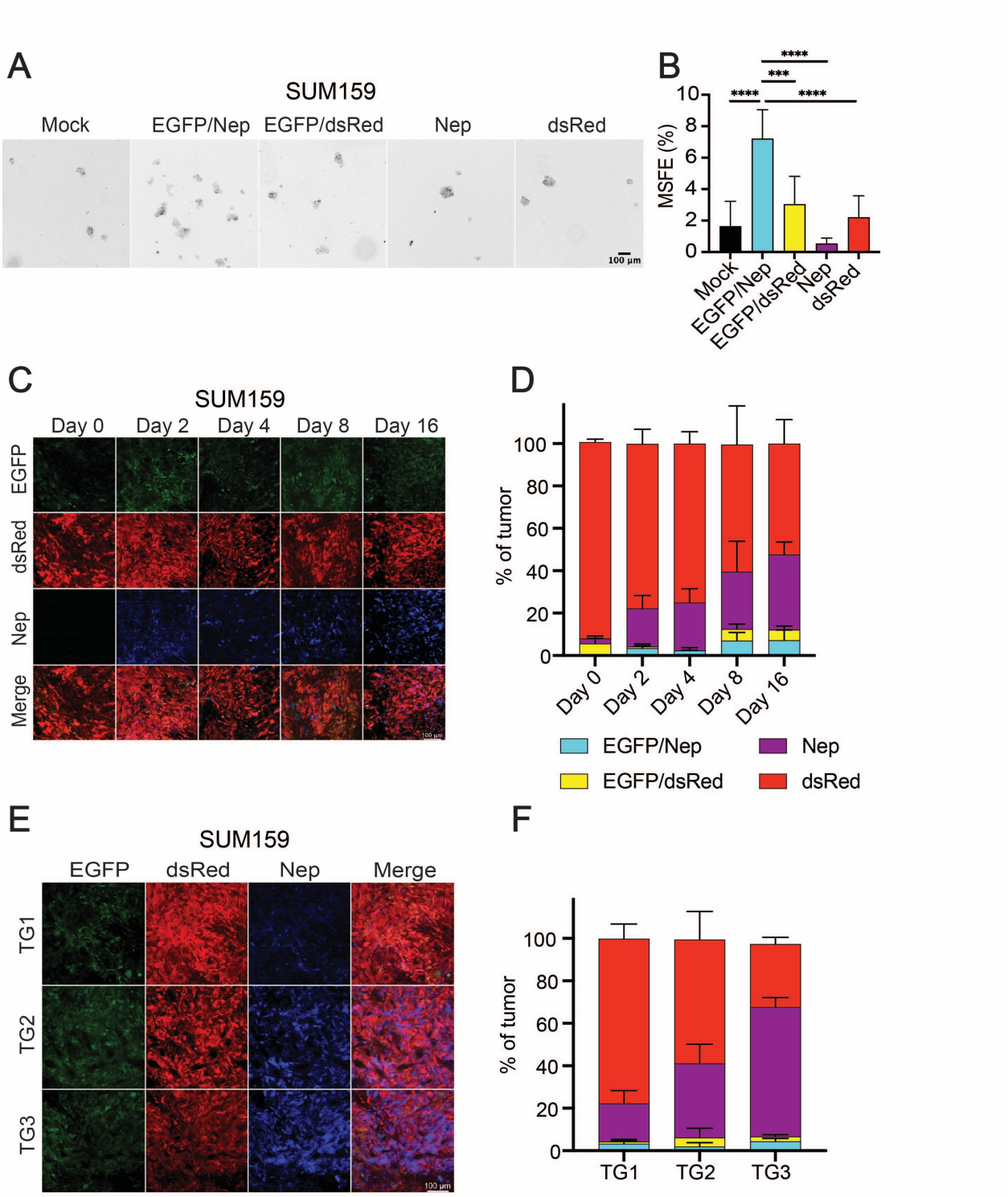


Figure 3: Long-lived STAT signaling cells are enriched for TIC in SUM159 xenografts. A) SUM159 STAT LT xenografts were dissociated, then each LT population was FACS enriched and grown as mammospheres. B) Mammosphere forming efficiency of each LT cell population from SUM159 xenografts (n = 6). One-way ANOVA and Tukey's post hoc test were used to determine statistical significance. C) Confocal images of a 16-day TAM time course in SUM159 xenografts. D) Flow cytometry quantification of the percentage of each LT cell population in SUM159

xenografts at five different time points. E) The time course was extended using a serial transplantation assay. Tumor chunks harvested two days post TAM (TG1), were transplanted into new hosts two additional times. F) Flow cytometry quantification of the percentage of each LT cell population in SUM159 xenografts at each transplant generation (TG). ***$p < 0.001$, ****$p < 0.0001$; values are mean ± SD. Nep: mNeptune2.

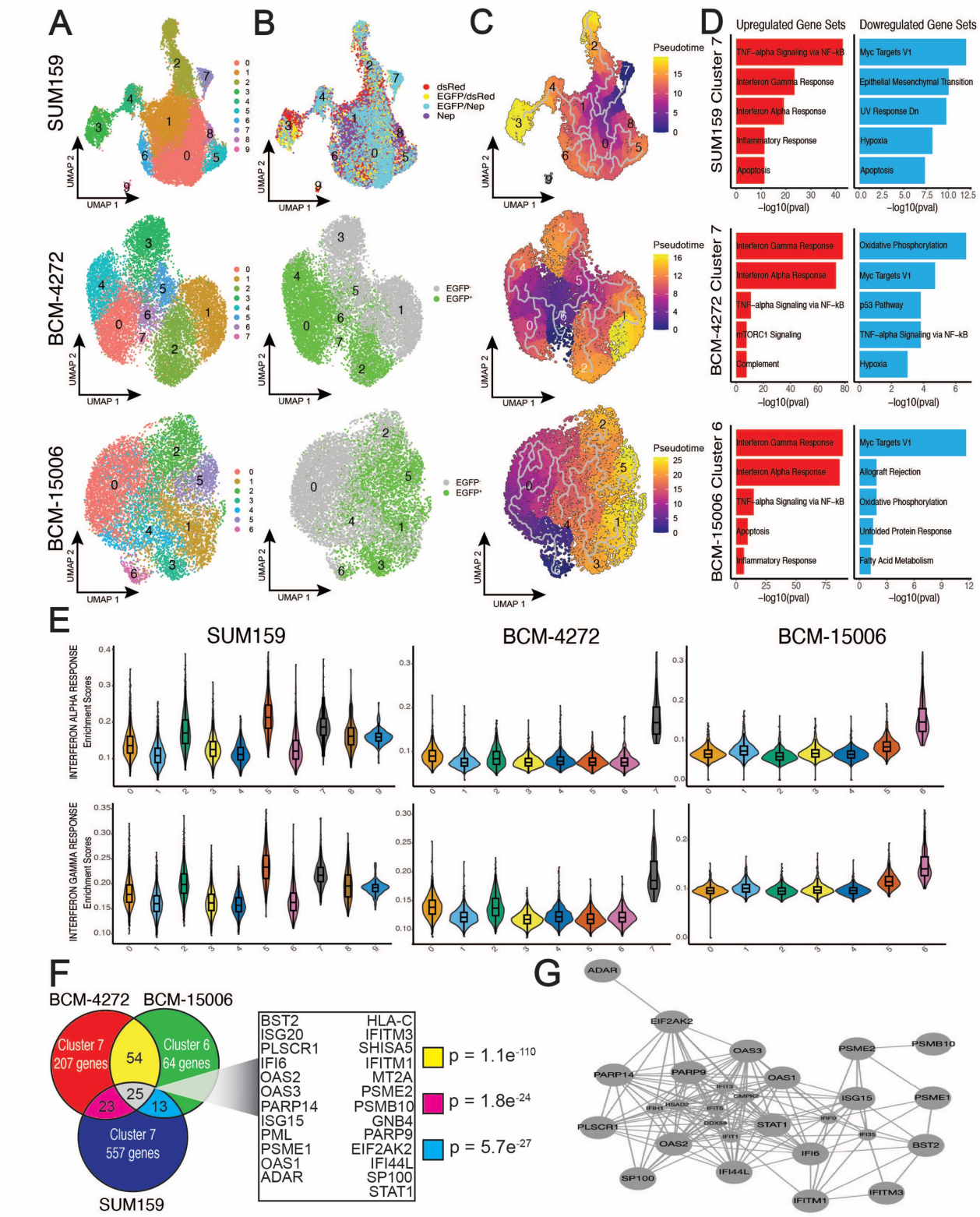


Figure 4: Interferon response genes are differentially expressed in TIC. SUM159 STAT3 LT, BCM-4272 4M67-EGFP, and BCM-15006 4M67-EGFP xenografts were harvested and dissociated, then each fluorescent population was enriched and sequenced using 10X genomics. A) Each xenograft was clustered and visualized using a Uniform Manifold Approximation and

Projection (UMAP). B) UMAP with cell identity overlayed on the plot. C) UMAP with Monocle3 lineage trajectory revealing which clusters lie at the end of the trajectories. D) Hallmark gene set enrichment analysis of candidate TIC transcriptional state from each xenograft. E) Hallmark gene set enrichment analysis (GSEA) comparing enrichment of interferon alpha and interferon gamma response gene sets across all clusters in each xenograft. F) Intersection of the upregulated genes in candidate TIC clusters with accompanying pairwise comparisons using Fisher's exact test. Intersection of candidate TIC clusters from each model reveals 25 shared genes. G) Illustration of a gene regulatory network demonstrating most of the 25 intersecting genes exist in the same gene regulatory network.

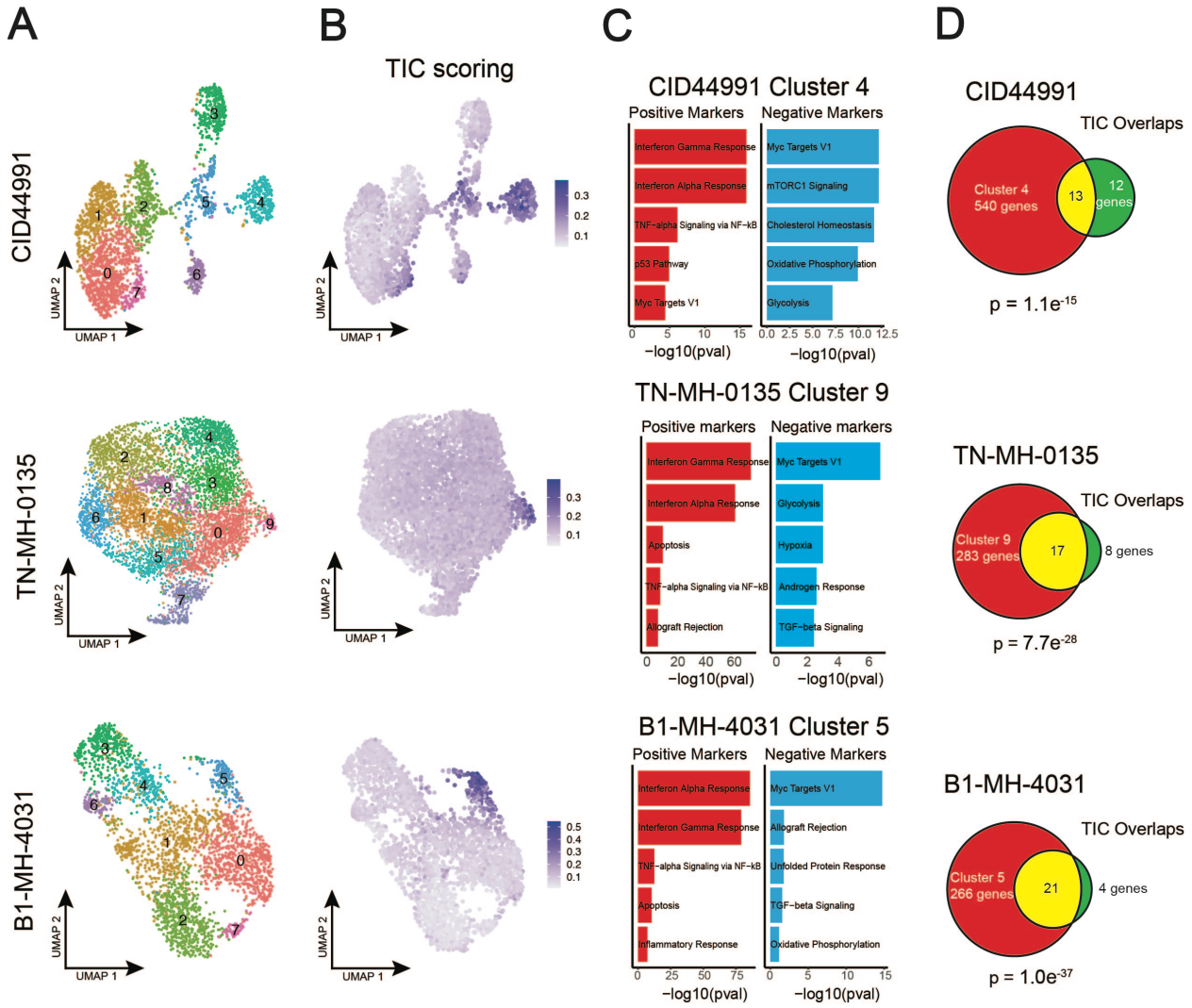
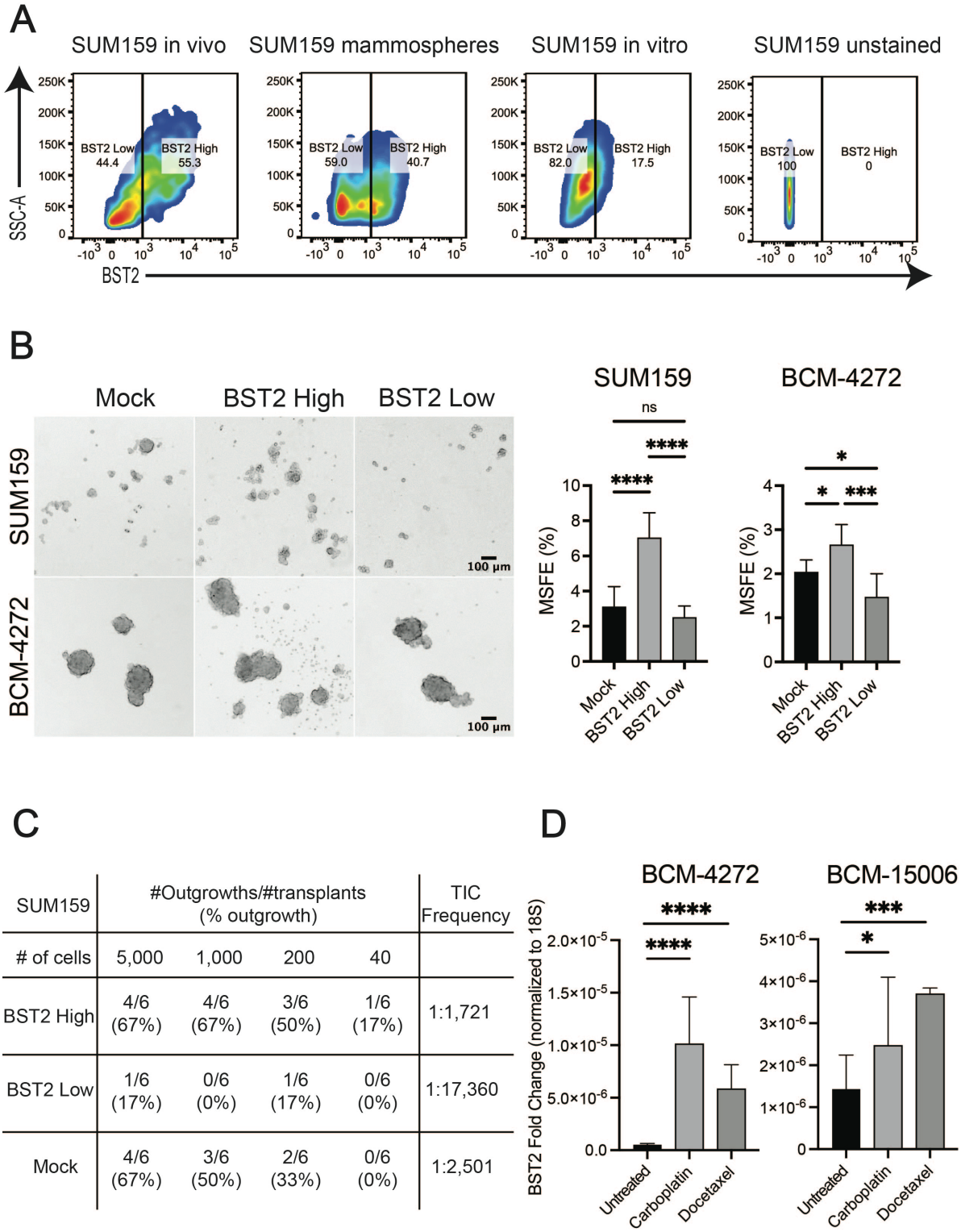


Figure 5: Interferon response genes are differentially expressed in a subset of epithelial cells in human breast cancer patient samples. A) UMAP visualization of three human breast cancer patient samples from publicly available single-cell RNA sequencing atlases (Pal et al., 2021; Wu et al., 2021). B) Feature plots illustrating TIC score based on each cell's similarity to the 25 genes shared between TIC transcriptional states across three xenograft models. C) GSEA for the cluster within each sample with a high TIC score relative to all other clusters. D) Intersections between clusters defined by differential expression of interferon response genes from each patient sample and the 25 intersecting genes common across three candidate TIC clusters.



SUM159	#Outgrowths/#transplants (% outgrowth)				TIC Frequency
# of cells	5,000	1,000	200	40	
BST2 High	4/6 (67%)	4/6 (67%)	3/6 (50%)	1/6 (17%)	1:1,721
BST2 Low	1/6 (17%)	0/6 (0%)	1/6 (17%)	0/6 (0%)	1:17,360
Mock	4/6 (67%)	3/6 (50%)	2/6 (33%)	0/6 (0%)	1:2,501

Figure 6: BST2 is a TIC marker in a subset of xenograft models. A) Flow cytometry plot illustrating the changes in BST2 expression in SUM159s when grown in vitro, as mammospheres, and in vivo. B) SUM159 and BCM-4272 xenografts were dissociated, then the BST2 high, BST2 low, and mock-sorted populations were FACS enriched and grown as mammospheres. Mammosphere forming efficiency of each cell population for each xenograft (n = 6). One-way

ANOVA and Tukey's post hoc test were used to determine statistical significance. Values are mean ± SD. C) Limiting dilution transplantation results of BST2 high, BST2 low, and mock-sorted populations from SUM159 xenografts. TIC frequency was calculated using extreme limiting dilution analysis (Hu and Smyth, 2009). D) qRT-PCR evaluation of BST2 RNA expression in residual BCM-4272 xenografts after carboplatin (50 mg/kg) or docetaxel (20 mg/kg) treatment relative to an untreated control (n = 6, carboplatin n = 3). One-way ANOVA and Tukey's post hoc test were used to determine statistical significance. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; values are mean ± SD.

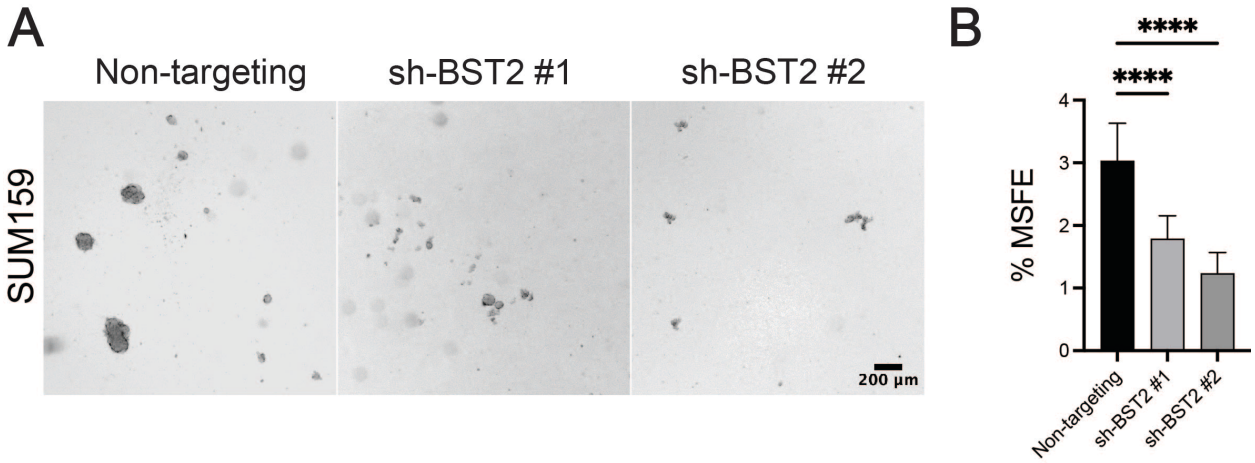


Figure 7: BST2 knockdown reduces mammosphere-formation efficiency in SUM159 cells. A) Representative images showing mammospheres formation in two SUM159 BST2 knockdown cells lines (sh-BST2 #1 and sh-BST2 #2) and a non-targeting control. B) Quantification of the mammosphere-forming efficiency of each cell line (n = 9). One-way ANOVA and Tukey's post hoc test were used to determine statistical significance. ****$p < 0.0001$; values are mean ± SD.