The interferon/STAT1 signaling axis is a common feature of tumor-initiating cells in 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], Xingxin Pan[6], Michael Zeosky[6], Anna Chung[6], S. Stephen Yi[6], 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.

[6]Livestrong Cancer Institutes and Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA; Oden Institute for Computational Engineering and Sciences and Interdisciplinary Life Sciences Graduate Programs, The University of Texas at Austin, Austin, TX 78712, USA

[*]Corresponding Author
Michael T. Lewis
Baylor College of Medicine
One Baylor Plaza
BCM600; Room N1210

28 TEL: 713-798-1538
29 Email: mtlewis@bcm.edu

30 **Running Title:** Interferon/STAT1 signaling in tumor-initiating cells

31 **DECLARATION OF INTERESTS**

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

37

ABSTRACT

A tumor cell subpopulation of tumor-initiating cells (TIC), or “cancer stem cells”, are associated with therapeutic resistance, as well as both local and distant recurrence. Enriched populations of TIC are identified by markers including aldehyde dehydrogenase (ALDH1) activity, the cell surface marker combination $CD44^{+}/CD24^{-}$, or fluorescent reporters for signaling pathways that regulate TIC function. We showed previously that Signal Transducer and Activator of Transcription (STAT)-mediated transcription allows enrichment for TIC in claudin-low models of human triple-negative breast cancer using a STAT-responsive reporter. However, the molecular phenotypes of STAT TIC are not well understood, and there is no existing method to lineage-trace TIC as they undergo cell state changes. Using a new STAT-responsive lineage-tracing (LT) system in conjunction with our original reporter, we enriched for cells with enhanced mammosphere-forming potential in some, but not all, basal-like triple-negative breast cancer (TNBC) xenograft models (TNBC) indicating TIC-related and TIC-independent functions for STAT signaling. Single-cell RNA sequencing (scRNAseq) of reporter-tagged xenografts and clinical samples identified a common interferon (IFN)/STAT1-associated transcriptional state, previously linked to inflammation and macrophage differentiation, in TIC. Surprisingly, most of the genes we identified are not present in previously published TIC signatures derived using bulk RNA sequencing. Finally, we demonstrated that bone marrow stromal cell antigen 2 (BST2), is a cell surface marker of this state, and that it functionally regulates TIC frequency. These results suggest TIC may exploit the IFN/STAT1 signaling axis to promote their activity, and that targeting this pathway may help eliminate TIC.

61 **Significance**

62 TIC differentially express interferon response genes, which were not previously reported in bulk
63 RNA sequencing-derived TIC signatures, highlighting the importance of coupling single-cell
64 transcriptomics with enrichment to derive TIC signatures.

65 **INTRODUCTION**

66 Breast cancer is a heterogeneous disease that can be classified into several histological and
67 molecular subtypes (1,2). Triple-negative breast cancer (TNBC) is a histological subtype defined
68 clinically by lack of estrogen receptor (ER) and/or progesterone receptor (PR), and without
69 amplification or overexpression of the ERBB2 gene (HER2), and is an aggressive disease with
70 poor outcomes (3).

71 There exists considerable intertumoral and intratumoral heterogeneity within each histological
72 subtype. Intertumoral heterogeneity refers to variations across tumors, whereas intratumoral
73 heterogeneity refers to the presence of molecularly distinct tumor cell populations within the same
74 tumor. Both present major challenges clinically, particularly in TNBC, which itself represents at
75 least six molecularly distinct subtypes (4,5). Due to the relative lack of targeted therapeutics,
76 TNBC is typically treated with combination chemotherapies, with or without immuno-oncology
77 agents or radiation.

78 With respect to intratumoral heterogeneity, there exists a subpopulation of cells, typically rare,
79 termed tumor-initiating cells (TIC) (a.k.a. cancer stem cells (CSC)). TIC are defined functionally
80 by their ability to self-renew, and to recapitulate clonally-derived cellular hierarchies present in the
81 tumor from which they were isolated upon transplantation (6). Enriched TIC populations can be
82 identified by aldehyde dehydrogenase (ALDH1) activity (7), the cell surface marker combination
83 of $CD44^{+}/CD24^{low/-}$ (8), or fluorescent signaling reporters expressing pathway-specific
84 transcription factors that regulate TIC function, including Wnt/β-catenin, Hedgehog, and Notch
85 signaling, as well as POU5F1 (OCT4)/SOX2, and Signal Transducer and Activator of
86 Transcription 3 (STAT3)-mediated signaling (9-14). However, no single marker is present in all
87 breast cancers.

88 Following neoadjuvant (before surgery) taxane- or anthracyclin-based chemotherapy in TNBC,
89 residual breast cancer cells are enriched for TIC markers, including $CD44^{+}/CD24^{low/-}$ and aldehyde

90 dehydrogenase (ALDH1) activity, and isolated cells show enhanced mammosphere-forming
91 efficiency (MSFE), indicating that TIC are intrinsically resistant to some systemic therapies (15-
92 18). Thus, standard cancer therapies can eliminate many “bulk” tumor cells to shrink tumors, but
93 TIC survive preferentially, leading to local and/or distant recurrence. Indeed, some metastatic
94 tumor cells express the TIC marker CD44, and such cells have enhanced tumor-initiating capacity
95 (19). Thus, selective targeting of TIC, especially at metastatic sites, may be essential for effective
96 treatment. Unfortunately, TIC-specific assays to evaluate drug efficacy are rarely done in clinical
97 trials.

98 Here, we exploit a STAT signaling reporter, and a new STAT-responsive lineage-tracing (LT)
99 system, to enrich cells with enhanced mammosphere-forming potential in some, but not all, non-
100 claudin-low xenograft models of TNBC. Single-cell RNA sequencing (scRNAseq) on three
101 xenograft models identified a distinct interferon (IFN)/STAT1-associated transcriptional state in
102 TIC that is also present in most clinical samples evaluated, yet contains genes that have not been
103 included in at least five TIC signatures derived through bulk RNA sequencing. Finally, we
104 validated our signature by demonstrating that one of the genes, bone marrow stromal antigen 2
105 (BST2), can enrich for TIC, and regulates TIC frequency, but not self-renewal. Taken together,
106 our data suggest that TIC exploit the IFN-mediated response pathway to regulate stem cell
107 frequency independent of self-renewal. Thus, this pathway may represent an attractive candidate
108 therapeutic target in a subset of breast cancers.

109

110 **MATERIALS & METHODS**

111 **STAT lineage-tracing (LT) system construction**

112 The STAT reporter (4M67-EGFP-P2A-CreER^{T2}) was generated by LR Gateway® Cloning of a
113 promoter donor vector and a gene donor vector into a destination vector. To generate a flexible
114 promoter donor vector, the EF-1α promoter sequence from the L4/R1 EF-1α promoter donor
115 vector (ThermoFisher Scientific, cat. A11146) was replaced with a 194bp sequence containing
116 multiple restriction digestion sites to facilitate simple restriction cloning of any desired promoter.
117 The STAT-specific promoter donor vector was generated by inserting the STAT signaling-specific
118 4M67 promoter from the STAT regulatory reporter (4M67-EGFP) (11) into the flexible promoter
119 donor vector to generate the L4-4M67-R1 promoter donor vector. To generate the gene donor
120 vector, a CreER^{T2} sequence with attB1/B2 recombination sites and an EcoRI cutting site between
121 attB1 and CreER^{T2} was amplified from pCAG-CreER^{T2} (20) (a gift from Connie Cepko: Addgene
122 plasmid #14797) by 2-step PCR, then incorporated into the pDonor233 vector (a kind gift from Dr.
123 Kenneth L. Scott, Baylor College of Medicine, Houston, TX) with the P1/P2 compatible
124 recombination sites by BP Gateway cloning. Next, GFP-P2A was amplified from pUltra (21) and
125 inserted in front of CreER^{T2} using InFusion cloning to generate L1-EGFP-P2A-CreER^{T2}-L2 gene
126 donor vector. Gateway® Cloning was used to incorporate the 4M67 promoter donor vector and
127 EGFP-P2A-CreER^{T2} gene donor vector into the pLenti6.4/R4R2/V5-Dest destination vector
128 (ThermoFisher Scientific, cat. A11146), generating 4M67-EGFP-P2A-CreER^{T2}.

129 To generate the dual color-switching lineage-tracing reporter (EFS-loxP-dsRed-loxP-Nep), the
130 loxP-dsRed-loxP-GFP sequence with attB1/B2 recombination sites was amplified from pMSCV-
131 loxP-dsRed-loxP-eGFP-Puro-WPRE (22) (a gift from Dr. Hans Clevers: Addgene plasmid
132 #32702) by 2-step PCR and then incorporated into the pDonor233 vector with the P1/P2
133 compatible recombination sites by BP Gateway cloning. Gateway® Cloning was used to
134 incorporate the EFS promoter donor vector and loxP-DsRed-loxP-EGFP gene donor vector into

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

139 **Lentivirus production and transduction**

140 Lineage-tracing plasmids and envelope-expressing plasmids were prepared using a plasmid maxi
141 kit (QIAGEN cat. 12162). To produce lentivirus, lineage-tracing plasmids and envelope-
142 expressing plasmids were transfected into HEK293T cells using jetPRIME® transfection reagent
143 (Polyplus, cat. 114-15). Lentiviral supernatants were collected every 24 hours for three days, then
144 concentrated. Concentrated viruses were titered using the p24-Elisa kit (Takara Bio, cat. 632200)
145 following manufacturer's protocols. For lentiviral transduction, 5mg/mL polybrene (Santa Cruz
146 Biotechnology, cat. sc-134220) was added to cell lines prior to lentiviral spinfection at 2500 RPM,
147 25°C for 30 minutes (MOI of 5). Transduced cells were incubated at 37°C, 5% CO_2 for 24 hours,
148 then lentivirus media was replaced with growth media. To generate LT cell lines (HepG2/C3A
149 STAT LT, SUM159 STAT LT, and MDA-MB-231 STAT LT), 4M67-GFP-P2A-CreER^{T2} was
150 transduced first, and transduced cells were selected using 10mg/mL blasticidin (Thermo Fisher
151 Scientific, cat. A1113903). Following blasticidin selection, cell lines were then transduced with
152 EFS-$_{loxP}$dsRed$_{loxP}$-Nep. Transduced cells were selected by FACS-enrichment of dsRed^{+} cells. To
153 generate BST2 knockdown cell lines, SUM159 cells were lentiviral transduced with either one of
154 two shRNA vectors (Horizon Discovery, sh-BST2 #1: cat. RHS3979-201824601, sh-BST2 #2: cat.
155 RHS3979-201829090) or one non-targeting control vector (Horizon Discovery, cat. RHS6848) as
156 described above. Transduced cells were selected using 1μg/mL puromycin (InvivoGen, cat. ant-
157 pr-1).

158 Xenograft tumors were dissociated (isolation of single cells from human xenograft tumors). The
159 resulting cells were transduced with the STAT regulatory reporter (4M67-GFP) (11) overnight in

160 suspension in complete Mammary Epithelial Cell Growth Media (MEGM$^+$) (Lonza, cat. CC-3153,
161 with supplements, cat. CC-4136) containing 20ng/mL basic fibroblast growth factor (Invitrogen,
162 cat. 13256-029), 10ng/mL epidermal growth factor (Invitrogen, cat. PHG0311), 4µg/mL Heparin
163 (StemCell Technologies, cat. 07980), and 40 nL/mL B27 (Invitrogen, cat. 17504-044).

164 **Cell culture and *in vitro* treatment**

165 HepG2/C3A, MDA-MB-231, and HEK293T cells were purchased from ATCC and were cultured
166 under adherent conditions with phenol red-free DMEM (Gibco, cat. 31053-028), 10% charcoal-
167 stripped FBS (Gibco, cat. 12-676-029), Penicillin-Streptomycin (GenDEPOT, cat. CA005-010),
168 1mM sodium pyruvate (Life Tech, cat. 11360070), and 2mM L-Glutamine (Thermo Fisher
169 Scientific, cat. 25030081). SUM159 cells were purchased from BioIVT and were cultured under
170 adherent conditions with phenol red-free Ham's F12 (Caisson labs, cat. HFL05-500ML) containing
171 5% charcoal-stripped FBS, 5mg/mL insulin (Gemini Bioproducts, cat. 700-112P), 1mg/mL
172 hydrocortisone (Sigma-Aldrich, cat. H-4001), and Penicillin-Streptomycin (GenDEPOT, cat.
173 CA005-010). Cells were incubated in a 5% CO_2 incubator at 37°C. All cell lines were authenticated
174 using short tandem repeat profiling, which was performed by the Cytogenetics and Cell Line
175 Authentication Core at MD Anderson Cancer Center (Houston, TX). Mycoplasma detection was
176 routinely performed using Venor®GeM qOneStep Mycoplasma detection kit (Minerva Biolabs,
177 cat. 11-91100). All experiments were performed with five passages of STR profiling and
178 mycoplasma testing.

179 Human recombinant IL-6 (Sigma-Aldrich, cat. I1395) was prepared as a 50mg/mL stock solution
180 in PBS/1%BSA. Recombinant human IFN alpha (IFNα) (Bio-Techne, cat. NBP2-35891-100ug)
181 and recombinant human IFN gamma (IFNγ) (Bio-Techne, cat. 285-IF-100) were prepared as a
182 200mg/mL stock in PBS. Tamoxifen (TAM) (Sigma-Aldrich, cat. T5648) was prepared as a 5mM
183 solution in 100% ethanol. HepG2/C3A STAT LT cells were treated with either IL-6 (50ng/mL),

184 TAM (1μM), or the combination and the cells were visualized by confocal microscopy (Leica, TCS-
185 SP5) and quantified by flow cytometry (BD LSRII analyzer) after 72 hours.

186 **Generation of cell line-derived and patient-derived xenografts**

187 To prepare for cell injection, SUM159 STAT LT or MDA-MB-231 STAT LT cells were suspended
188 in a 1:1 solution of their respective growth media and phenol red-free Matrigel® matrix (Corning,
189 cat. 356237). Cells were transplanted into the #4 epithelium-free "cleared" fat pad of 3–4-week-
190 old SCID/Beige mice (Charles River Laboratories, Wilmington, MA, http://www.criver.com). When
191 palpable tumors formed, mice were administered TAM by intraperitoneal injection to induce
192 dsRed to mNeptune2 color switching in STAT signaling ($EGFP^+$) cells. TAM was prepared by
193 dissolving in sunflower seed oil (Sigma-Aldrich, cat. S5007) and stirring overnight. All animal
194 protocols were reviewed and approved by the Animal Protocol Review Committee at Baylor
195 College of Medicine.

196 **Isolation of single cells from human xenograft tumors**

197 Xenograft tumors were harvested and digested using 5mL digestion media (phenol red-free
198 DMEM/F12 [Corning cat. 16-405-CV], 100mg/mL gentamicin [Worthington, cat. LS004176],
199 250U/mL collagenase II [Worthington cat. LS004174] and 2μg/mL DNase I [Stemcell, cat. 07900])
200 per 1g tumor at 37°C in a gentleMACS dissociator (Miltenyi). The tumor digestion was passed
201 through a 200mM filter and centrifuged at 300g for 1 minute. Red blood cells were lysed by adding
202 9mL ddH_2O to the cell pellet for 20 seconds and lysis was stopped by the addition of 1mL 10X
203 PBS. To dissociate cell clumps further, the cell suspension was centrifuged at 300 g for 1 minute,
204 then resuspended in TrypLE Express (Gibco, cat. 12605-010) for 5 minutes. To stop the digestion,
205 9mL of PBS was added, and the cell suspension was passed through a 40mm cell strainer.

206 **Analytical flow cytometry and FACS**

207 Cells from adherent cultures were trypsinized and pelleted by centrifuging at 300g for 3 minutes.
208 The cells were resuspended in $HBSS^+$ (HBSS [Cytiva Life Sciences cat. SH30588.01], 2% FBS).

209 Single cells from digested xenograft tumors were labeled using a BUV395 anti-mouse cocktail
210 (pooled and diluted 1:40: anti-mouse H2kD [BD Biosciences, cat. 742437], anti-mouse CD45 [BD
211 Biosciences, cat. 567451], anti-mouse CD31 [BD Biosciences, cat. 740231], and anti-mouse
212 Ter119 [BD Biosciences, cat. 563827]) to exclude mouse cells. A far-red BST2 antibody was used
213 to label BST2-expressing cells (diluted 1:40, BioLegend, cat. 348416). CD44 was labeled with a
214 PE conjugated antibody (diluted 1:20, BD Biosciences, cat. 566803), and CD24 was labeled with
215 a FITC conjugated antibody (diluted 1:20, BD Biosciences, cat. 655154). Cell viability was
216 determined by Sytox Blue (diluted 1:1000, V450, Invitrogen, cat. S34857) staining. Cells were
217 applied to a BD LSRII analyzer for analytical flow cytometry or a BD AriaII for fluorescence-
218 activated cell sorting (FACS). Quantification of each cell population was performed using FlowJo
219 (BD Biosciences) and statistical differences between cell populations were determined by one-
220 way ANOVA followed by post-hoc Tukey's test, which was calculated using GraphPad Prism
221 version 9 (GraphPad Software, San Diego, CA, http://www.graphpad.com).

222 **ALDEFLUOR assay**

223 ALDH activity was assessed using the ALDEFLUOR kit (StemCell technologies, cat. 01700)
224 according to the manufacturer's protocol. Briefly, cells were incubated in ALDEFLUOR assay
225 buffer containing ALDH substrate for 30 minutes at 37°C. In each experiment, a sample of cells
226 was stained under identical conditions with diethylaminobenzaldehyde (DEAB), a specific ALDH
227 inhibitor, as a negative control. Any antibody incubation was performed immediately after
228 ALDEFLUOR incubation, as described in the "analytical flow cytometry and FACS" section.

229 **Propidium Iodide cell cycle assay**

230 Single cell suspensions were fixed in 70% ethanol overnight at -20°C. Then, cells were washed
231 twice with PBS and stained in 100μL propidium iodide solution (0.001% Triton X-100, 0.2mg/mL
232 DNase-free RNase A (Thermo Fisher, cat. 12091-021), 20μg/mL propidium iodide (Thermo

233 Fisher, cat. P3566)) for 15 minutes at 37°C. After incubations, the cells were immediately applied
234 to a BD LSRII analyzer for analytical flow cytometry.

235 **Immunohistochemistry and Immunofluorescence**

236 Immunohistochemistry (IHC) and immunofluorescence (IF) were performed by the Lester and
237 Sue Smith Breast Center Pathology Core. Tumors were fixed in 4% paraformaldehyde, paraffin-
238 embedded, sectioned at 3–4 μm, and deparaffinized. The antigen retrieval buffer for pSTAT1,
239 pSTAT3, and pSTAT5 staining was 0.1M Tris-HCL PH 9.0. Sections were then blocked in 3%
240 hydrogen peroxide solution for 5 minutes and incubated by pSTAT1 Ser727 antibody (Cell
241 Signaling Technology, cat. 8826, 1:200), pSTAT3 Y705 antibody (Cell Signaling Technology, cat.
242 9145, 1:25) or pSTAT5 antibody Tyr694 (Cell Signaling Technology, cat. 9359, 1:25), followed
243 by Signalstain Boost IHC Detection Reagent (HRP, Rabbit) (Cell Signaling Technology, cat.
244 8114) for 30 minutes, visualized by 3, 3′ - Diaminobenzidine (DAB+) solution (DakoCytomation,
245 Glostrup, Denmark), and counterstained by Harris hematoxylin. SUM159 tumors, known for
246 activation of pSTAT3 (11), were used as a positive control for pSTAT3, and a human tissue array,
247 which included the spleen, was used as a positive control for pSTAT1.

248 For IF, Five IF images were quantified for each PDX model using CellProfiler (23). 0.1M Tris-HCL
249 PH 9.0. was used as the antigen retrieval buffer, then sections were blocked in 3% hydrogen
250 peroxide solution for 5 minutes and incubated with pSTAT1 Ser727 antibody (Cell Signaling
251 Technology, cat. 8826, 1:200), pSTAT3 Y705 antibody (Cell Signaling Technology, cat.
252 9145, 1:25), or a GFP antibody (Fisher Scientific, cat. NC9777966) overnight at 4°C. The following
253 day, slides were incubated with secondary antibodies: anti-mouse (for pSTAT1 and pSTAT3,
254 Thermo Fisher, cat. A-11032) and/or anti-rabbit (for GFP, Thermo Fisher, cat. A-11008) for one
255 hour at room temperature. Slides were then counterstained with DAPI and imaged. For each PDX
256 model, five representative images were quantified using FIJI (24).

257 **Mammosphere-forming efficiency assay**

258 Following tumor digestion and FACS enrichment of fluorescent cell populations, equal numbers
259 of cells from each fraction were plated in complete Mammary Epithelial Cell Growth Media
260 ($MEGM^+$) (Lonza, cat. CC-3153, with supplements, cat. CC-4136) containing 20ng/mL basic
261 fibroblast growth factor (Invitrogen, cat. 13256-029), 10ng/mL epidermal growth factor (Invitrogen,
262 cat. PHG0311), 4μg/mL Heparin (StemCell Technologies, cat. 07980), and 400nL/mL B27
263 (Invitrogen, cat. 17504-044) with 0.6% HPMC (Millipore Sigma, cat. M7027) to prevent cell
264 aggregation, in an ultra-low attachment plate.

265 To prepare secondary mammospheres, primary mammospheres were collected and dissociated
266 into single cells using TrypLE Express (Gibco, cat. 12605-010) for 5 minutes. To inactivate the
267 TrypLE Express, 9mL of PBS was added, and the cell suspension was passed through a 40mm
268 cell strainer. Then, an equal number of cells from each fraction were plated in $MEGM^+$ media with
269 0.6% HPMC. Secondary mammospheres were imaged on a Yokogawa CV8000, and
270 CellPathfinder (Yokogawa) was used to quantify the number of mammospheres in each well.
271 Spheres with an area greater than $300mm^2$ and a diameter greater than 50mm were used to
272 determine mammospheres. A BioTek Cytation 5 (Agilent, Santa Clara, CA) was used to take
273 representative images. Statistical differences in mammosphere forming efficiency between cell
274 populations were determined by one-way ANOVA followed by post-hoc Tukey's test, which was
275 calculated using GraphPad Prism version 9 (GraphPad Software, San Diego, CA,
276 http://www.graphpad.com).

277 **Limiting-dilution transplantation assay**

278 FACS-enriched cells were suspended in growth media at a concentration of 5,000 cells per 10μl
279 (SUM159) or 50,000 cells per 10μl (BCM-4272) and serially diluted using growth media. Four
280 dilutions were prepared (SUM159: 5,000, 1,000, 200, 40 cells per 10μl, BCM-4272: 50,000, 5,000,
281 500, 50 cells per 10μl), then phenol red-free Matrigel® matrix (Corning, cat. 356237) was added
282 for a final 1:1 solution of growth media and Matrigel® before transplantation as described in

283 “Generation of cell line-derived and patient-derived xenografts.” The mice were monitored until
284 the first tumor reached 10% body weight (~7-8 weeks) at which point the experiment was
285 terminated and tumor outgrowth was recorded. Extreme limiting-dilution analysis was used to
286 calculate the TIC frequency of each population (25).

287 **scRNAseq library prep and sequencing**

288 SUM159, BCM-4272, and BCM-15006 were sorted based on fluorescent reporter activity
289 (SUM159: $EGFP^+$/$mNeptune2^+$, $EGFP^+$/$dsRed^+$, $EGFP^-$/$mNeptune2^+$, $EGFP^-$/$dsRed^+$; BCM-4272
290 and BCM-15006: $EGFP^+$ or $EGFP^-$), and single-cell gene expression libraries were prepared
291 according to Chromium Single Cell Gene Expression 3v3.1 kit (10x Genomics). In brief, single
292 cells, reverse transcription (RT) reagents, Gel Beads containing barcoded oligonucleotides, and
293 oil were loaded on a Chromium controller (10x Genomics) to generate single-cell GEMS (Gel
294 Beads-In-Emulsions) where full-length cDNA was synthesized and barcoded for each single cell.
295 Subsequently the GEMS are broken and cDNA from each single cell are pooled. Following
296 cleanup using Dynabeads MyOne Silane Beads, cDNA is amplified by PCR. The amplified
297 product is fragmented to optimal size before end-repair, A-tailing, and adaptor ligation. The final
298 library was generated by amplification. Finally, single-cell sequencing libraries were sequenced
299 using NovaSeq 6000 (read configuration: 28-10-10-90; 1% PhiX Spike-in; read depth
300 200M/sample).

301 For single nuclei sequencing, Tissue processing and nuclei RNA fixation from PDX samples were
302 performed using Chromium Next GEM Single Cell Fixed RNA Sample Preparation Kit (10x
303 Genomics). Fixation was then quenched and samples were stored at -80°C until sequencing.
304 Single nuclei RNAseq libraries were prepared according to Chromium Single Cell Gene
305 Expression 3v3.1 kit (10x Genomics), similar to the cDNA synthesis and barcoding process for
306 each single cell. Single-cell nuclei RNA libraries were sequenced using NovaSeq (read
307 configuration: Paired End 28-10-10-90; 1% PhiX Spike-in; read depth 100M/sample).

308 **scRNAseq data preprocessing and filtering**

309 FASTQ files were mapped to the GRCh38 reference genome using the Cell Ranger Single Cell
310 software v.6.1.2 (SUM159) or v7.0.1 (BCM-4272 and BCM-15006) (10X Genomics). Single nuclei
311 sequencing data were integrated with the Texas Advanced Computing Center (TACC)
312 supercomputer resources for data analysis and interpretation. In brief, gene expression counts
313 for individual nuclei were generated from raw FASTQ files using Cell Ranger (version 3). Reads
314 with low mapping quality (<10) or that map to multiple genomic locations were filtered out. Using
315 Seurat v4.1.3 in R v4.2.2, the count matrix of each sample was used to create a Seurat object.
316 Cells expressing greater than 1000 and less than 9000 genes. Cells with greater than 10%
317 mitochondrial transcripts were excluded as they represented low-quality or dying cells. Next,
318 fluorescent samples from the same tumor were merged into a single Seurat object (ex. BCM-
319 4272 EGFP+ and BCM-4272 EGFP- merged to BCM-4272). To exclude the effects of cell cycle
320 heterogeneity in the analysis, we used the CellCycleScoring function from Seurat. This function
321 gives each cell a quantitative score for G1, G2/M, and S phases based on the expression of cell
322 cycle phase genes (26). Subsequently, effects of cell cycle heterogeneity were regressed out of
323 the data during data scaling.

324 **scRNAseq clustering and UMAP visualization**

325 Seurat v4.1.3 was used for data normalization, dimensionality reduction, and clustering (27). The
326 filtered gene matrices from each merged sample were normalized using the NormalizeData
327 function. The normalized datasets were centered and scaled using ScaleData. Then, the most
328 significant principal components were found using RunPCA, and ElbowPlot was used to
329 determine the number of principal components to use for clustering each dataset. Different
330 resolution parameters for unsupervised clustering were examined to find the optimal number of
331 clusters based on clear segregation of clusters. For each dataset, the first 35 principal
332 components were used for unsupervised clustering with a resolution = 0.1 for SUM159 (yielding

333 nine clusters) and a resolution of 0.5 for BCM-4272 and BCM-15006 (yielding seven clusters and
334 six clusters, respectively) using the FindNeighbors and FindClusters functions. To visualize the
335 clustering for each dataset, the dimensionality was further reduced using the Seurat function
336 RunUMAP. The same principal components used to determine the clustering were the same used
337 to calculate the UMAP embedding. Each resulting cluster was further analyzed for potential
338 doublets or low-quality cells using two methods: quality metrics (nCount_RNA, nFeature_RNA,
339 and mitochondrial content) for each cluster were calculated and any outlier values (clusters with
340 greater, or less than, two standard deviations than the average of all clusters) were removed.
341 Three PDX models from an independent, single nuclei RNA seq, dataset were processed,
342 clustered, and visualized as described above. UCell was used to score each cell based on
343 similarity the 25 genes identified by intersecting SUM159-C7, 4272-C7, and 15006-C6 (28).

344 **scRNAseq differential gene expression analysis and intersection of gene lists**

345 To examine the different transcriptional states of tumor cells, FindAllMarkers was used to identify
346 differentially expressed genes in each cluster using the “bimod” likelihood-ratio test framework.
347 Two additional parameters were used to include differentially expressed genes: genes that
348 showed at least a 0.25 log2 fold expression difference, and genes that were expressed in at least
349 25% of the cells in a cluster. To compare the similarities between candidate TIC transcriptional
350 states in different xenograft models, upregulated genes from each candidate TIC cluster were
351 intersected using GeneOverlap (29). To examine the similarity between the TIC genes we
352 identified and previously published TIC signatures derived using bulk RNA sequencing (30-34),
353 we intersected the TIC genes common between any two xenograft models with previously
354 published TIC signatures using GeneOverlap (29). To examine the extent to which the 25
355 common TIC genes are unique to a given transcriptional state in each xenograft model, we
356 performed Marker Enrichment Modeling (MEM) using the 25 common TIC genes across models

357 as MEM labels. Similarity scores based on MEM labels were calculated and visualized with a
358 heatmap using the MEM_RMSD function (35).

359 **scRNAseq gene set enrichment analysis**

360 To associate a biological phenotype with each cluster, gene set enrichment analysis was
361 performed using ClusterProfiler (36). Differentially expressed genes from each cluster, as
362 determine using the FindAllMarkers function described above, with a log fold change threshold of
363 0.25 were included in the analysis. Gene sets with a false discovery rate (FDR) < 0.01 and a
364 Benjamini-Hochberg (BH)-adjusted p-value < 0.05 were considered significantly enriched, and
365 these cutoffs were established prior to analyzing the data.

366 **Analysis of patient scRNAseq data**

367 Publicly available human breast cancer datasets were retrieved through GEO (GSE161529,
368 GSE176078) (37,38). Seurat v4.1.3 was used for cell cycle regression, data normalization,
369 dimensionality reduction, and clustering, as described above. To identify clusters of epithelial
370 cells, the expression of keratins (KRT8, KRT14, KRT18, and KRT19) was evaluated. Clusters
371 with no keratin expression were filtered, and the epithelial dataset was re-normalized and
372 clustered using Seurat v4.1.3. UCell was used to score every cell in each patient samples with
373 more than 1,500 non-cycling epithelial cells based on similarity using the 25 genes identified by
374 intersecting SUM159-C7, 4272-C7, and 15006-C6 (28). Harmony was used to integrate common
375 cell states across patient samples (39). The patient samples were integrated by breast cancer
376 subtype, pre-neoplastic, or normal breast. UCell was used to score each cell in the integrated
377 datasets based on similarity using the 25 genes identified by intersecting SUM159-C7, 4272-C7,
378 and 15006-C6 (28). A two-sided Wilcoxon test was used to determine significance of TIC scores.
379 To identify differentially expressed genes between clusters, FindAllMarkers from Seurat v4.1.3
380 was used. To compare the similarities between clusters with high TIC scores across breast cancer

381 subtypes and our 25-gene signature, upregulated genes from each cluster were intersected using
382 GeneOverlap (29).

383 **Patient survival and chemotherapy response**

384 To determine whether there is any association between our 25-gene TIC signature and overall
385 survival in breast cancer patients, we used the RNAseq datasets from KM-plotter (40). ROC
386 plotter was used to determine whether our 25-gene TIC signature is associated with
387 chemotherapy response in breast cancer patients (41). Statistical significance was set at $p < 0.05$
388 before the analysis was performed.

389 **qRT-PCR**

390 RNA extraction was performed using the RNeasy plus mini kit (Qiagen, cat. 74134) according to
391 the manufacturer's protocol. Next, we performed cDNA synthesis using SuperScript™ III first-
392 strand synthesis system (ThermoFisher Scientific, cat. 18080051) according to the
393 manufacturer's protocol. Then, qPCR was performed using iTaq qPCR supermix (BioRad, cat.
394 1725130). Briefly, 10μL 2X Taq, 9μL cDNA, and 1μL of a 20X TaqMan probe (ThermoFisher
395 Scientific, cat. Hs00171632_m1 (BST2), Hs00241666_m1 (ADAR), Hs00942643_m1 (OAS2),
396 Hs00967084_m1 (PARP9), Hs00393816_m1 (PARP14), Hs01013996_m1 (STAT1),
397 Hs99999901_s1 (18S)) were added per reaction. The qPCR reaction was performed in a C1000-
398 Touch-Thermocycler (BioRad) (Initial denaturation: 95°C 30 seconds, followed by 39 cycles of
399 denaturation: 95°C 5 seconds and annealing/extension 60°C 30 seconds). Samples were
400 normalized to the expression of a housekeeping gene, 18S. Statistical differences in BST2
401 expression following chemotherapy were determined for each PDX independently by one-way
402 ANOVA followed by post-hoc Tukey's test. Statistical differences in the expression of IFN
403 response genes were determined for each model independently by one-way ANOVA followed by
404 post-hoc Tukey's test. All statistical differences were calculated using GraphPad Prism version 9
405 (GraphPad Software, San Diego, CA, http://www.graphpad.com).

406 **Jess™ Automated Western Blot**

407 BST2 expression levels in BST2 knockdown cells relative to the non-targeting control cells was
408 evaluated using ProteinSimple's Jess™ fully automated western blot analysis (Bio-Techne, cat.
409 004-650) according to the manufacturer's protocol. A total of 0.5mg/mL protein was loaded per
410 sample and a BST2 antibody (1:25, Novus Biologicals, cat. NBP2-27154) was used to detect
411 BST2 levels in each sample. The ProteinSimple's Jess™ protein normalization was used to
412 normalize variation in protein loading.

413

414 **RESULTS**

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

417 We showed previously that STAT-mediated transcriptional activity allows enrichment for TIC in
418 two cell line xenograft (CDX) models of claudin-low breast cancer (SUM159 and MDA-MB-231)
419 (11). While our focus was on STAT3 in these two models, STAT1 and STAT3 transcription factors
420 bind similar DNA sequences, and both can activate the reporter (42).

421 To determine whether STAT signaling might also be associated with TIC function in a broader
422 range of TNBC, we screened a cohort of 56 TNBC PDX models for pSTAT3 expression, of which
423 38 (68%) had detectable pSTAT3 staining in the epithelium (Supplementary Table 1). This value
424 is higher than, but arguably consistent with, results using clinical samples of TNBC (~37.5% (43))
425 given the known engraftment bias present in PDX collections toward immunologically "cold"
426 tumors (44).

427 The 38 pSTAT3 positive TNBC models were then cross-referenced with PDX showing behaviors
428 of interest with respect to putative roles of TIC, specifically local recurrence after apparent
429 complete response to one or more chemotherapies, and/or metastasis to one or more organ sites
430 (44,45). Based on these properties, a subset of 11 PDX models of TNBC were chosen for genetic
431 tagging with the 4M67-EGFP lentiviral vector (11). Nine PDX showed varying degrees of EGFP
432 expression, whereas candidate negative control models BCM-0046 and WHIM12, showed no
433 detectable EGFP expression (Supplementary Table 2). Of nine EGFP+ PDX, five models (BCM-
434 3611, BCM-4013, BCM-5097, HCI-028, and HCI-002) were excluded from further consideration
435 because the percentage of $EGFP^+$ cells was too low (<1.2%) to be experimentally tractable,
436 thereby yielding four remaining PDX models for use.

437 The chosen PDX were then evaluated for pSTAT1, pSTAT3, and pSTAT5 by IHC, as well as
438 EGFP expression by immunofluorescence (Fig. 1A, Fig. S1A). Consistent with pSTAT1/3

439 positivity by IHC, PDX models displayed their anticipated STAT reporter activity via detection of
440 EGFP (Fig. 1A). Reporter-positive PDX models did not exhibit pSTAT5 staining, indicating STAT5
441 signaling is not responsible for reporter activity. In $EGFP^+$ PDX, analytical flow cytometry (Fig.
442 1B) showed a continuum of EGFP reporter activity across models ranging from 7.88% (BCM-
443 15006) to 31.7% (MC1) of tumor cells, ranges that should allow for discrimination of differential
444 TIC function/gene expression in $EGFP^+$ vs. $EGFP^-$ cells, should such differences exist. Although
445 EGFP percentage is comparable to pSTAT1 and pSTAT3 percentage (Fig. 1A), co-IF staining
446 revealed little direct correlation between the reporter and pSTAT1/3 (Fig. S1B), likely due to the
447 26-hour half-life of GFP (46) relative to the comparatively rapid rate of STAT dephosphorylation
448 (47).

449 To determine whether STAT reporter activity enriches for cells with increased MSFE in reporter-
450 positive TNBC PDX models (a surrogate for TIC function), we sorted $EGFP^+$, $EGFP^-$, and a mock
451 sorted control from each reporter-positive model and plated them under low attachment conditions
452 (Fig. 1C). STAT reporter positive cells in BCM-4272 had an 8.04% MSFE, representing a more
453 than 100% increase relative to the $EGFP^-$ cell population (3.74% MSFE), and the MSFE of the
454 $EGFP^-$ population was reduced by 59% relative to a mock sorted control (6.3% MSFE) (Fig. 1C).
455 In contrast, STAT reporter activity did not enrich for cells with enhanced MSFE in BCM-15006,
456 BCM-7821, or MC1 (Fig. 1C). Thus, while the 4M67 reporter can identify cells with enhanced
457 MSFE in some TNBC, STAT-mediated transcription must have TIC-independent function(s).

458 **A STAT lineage-tracing reporter provides a long-term fluorescent label to STAT signaling**
459 **cells**

460 Given that TIC marker expression, including STAT transcriptional activity, is dynamically
461 regulated (48), the inability to lineage-trace STAT signaling cells as they undergo cell state
462 changes, or differentiate, is a major limitation of the 4M67-EGFP reporter.

463 To augment the 4M67-EGFP reporter, we developed a TAM-activated, Cre recombinase-
464 dependent, STAT signaling-specific, lentiviral LT system (Fig. S2A–B). This system consists of
465 two components: the first is a lentiviral vector that labels cells with active STAT signaling ($EGFP^+$),
466 followed by a self-cleaving P2A peptide and a TAM-activated Cre recombinase (4M67-EGFP-
467 P2A-$CreER^{T2}$) (Fig. S2A). The second component is a constitutively-expressed dual-color
468 switching Cre-dependent reporter vector (EFS-$_{loxP}$dsRed$_{loxP}$-mNeptune2) (Fig. S2B). Addition of
469 TAM drives color switching from red (dsRed) to far-red (mNeptune2) via $CreER^{T2}$ recombination
470 in STAT signaling cells within a specific temporal window where the TAM concentration remains
471 high enough to activate Cre recombinase (6.8-hour half-life) (49) (Fig. 2A). Use of the LT system
472 results in four cell populations: long-lived (label-retaining) STAT signaling cells
473 ($EGFP^+$/mNeptune2$^+$), transient STAT signaling cells post-TAM (or cells that did not undergo Cre
474 recombination) ($EGFP^+$/dsRed$^+$), lineage-tagged cells ($EGFP^-$/mNeptune2$^+$), and bulk tumor cells
475 ($EGFP^-$/dsRed$^+$) (Fig. 2B).

476 To demonstrate the sensitivity and specificity of the STAT LT reporter system, we used
477 HepG2/C3A cells (IL-6 responsive) in a STAT3-dependent manner (11,50). Untreated STAT LT-
478 transduced HepG2/C3A cells (HepG2/C3A STAT LT) had minimal EGFP and mNeptune2
479 expression. TAM (1μM) induced mNeptune2 expression in a small percentage of cells. Addition
480 of IL-6 (50ng/mL) induced the EGFP reporter activity, without mNeptune2 activation, but the
481 addition of both IL-6 and TAM induced EGFP and mNeptune2 expression (Fig. S2C–D). These
482 results demonstrate the LT reporter system is both stringent and sensitive to STAT signaling, and
483 TAM addition induced mNeptune2 expression in STAT-responsive cells.

484 To demonstrate the ability of the LT system to provide a long-term fluorescent label on STAT
485 signaling cells *in vivo*, we introduced the LT reporters into the two TNBC cell lines used previously
486 to demonstrate STAT reporter activity enriches for TIC (SUM159 STAT LT and MDA-MB-231
487 STAT LT) (11). To ensure effective lineage-tagging of STAT signaling cells, we performed an *in*

488 *vivo* TAM dose-response analysis (Fig. 2SE–F). Given that high levels of TAM can cause toxicity
489 in mice (51), we aimed to find a dose that resulted in maximal loxP recombination without any
490 adverse effects. At 4mg, the recombination efficiency (the percentage of $EGFP^+/Nep^+$ cells out of
491 all $EGFP^+$ cells) two days after TAM treatment was 71%. The recombination efficiency, and the
492 total percentage of $mNeptune2^+$ cells, decreased with each successive dose, with the lowest dose
493 (0.006mg) being comparable to vehicle. Given high recombination with no observable toxicity at
494 4mg TAM, we used this dose for the remainder of our studies.

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

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

509 To confirm that the STAT LT system enriches for cells with enhanced MSFE from SUM159 tumors
510 in a manner similar to that of the 4M67-EGFP vector shown previously (11), we dissociated
511 SUM159 xenografts eight days after TAM treatment, sorted the four fluorescent cell populations
512 ($EGFP^+/mNeptune2^+$, $EGFP^+/dsRed^+$, $EGFP^-/mNeptune2^+$, and $EGFP^-/dsRed^+$), and performed

513 MSFE assays (Fig. 3A–B). Long-lived, label-retaining STAT signaling cells ($EGFP^+$/$mNeptune2^+$)
514 had a significantly higher MSFE (7.2%) compared to all other cell populations, which were not
515 statistically different from each other.

516 The cancer stem cell hypothesis postulates that cellular heterogeneity in tumors results, in part,
517 from hierarchical organization of cell types present, and that these hierarchies are fueled by rare,
518 self-renewing TIC (52). We reasoned that long-lived STAT signaling cells give rise to progenitor
519 cells ($EGFP^-$/$mNeptune2^+$) that will differentiate into non-TIC bulk tumor cells over time.

520 To test this hypothesis, we performed a 16-day time course study. We harvested SUM159 tumors
521 prior to TAM treatment, and again at days 2, 4, 8, and 16 after the TAM pulse (Fig. 3C–D). A total
522 of 74% of the $EGFP^+$ cells recombined ($EGFP^+$/$dsRed^+$ converted to $EGFP^+$/$mNeptune2^+$ cells)
523 two days after the TAM pulse. Between days 0 and 16, the percentage of $EGFP^-$/Nep^+ cells
524 increased from 2% to 35% of the tumor ($p < 0.001$), indicating lineage-tagged cells proliferate and
525 displace the $EGFP^-$/$dsRed^+$ cells. Furthermore, we noted the re-emergence of $EGFP^+$/$dsRed^+$
526 cells on the 8th day, suggesting STAT signaling activation in cells not defined by STAT signaling
527 at the time of the TAM pulse.

528 To extend the timeframe of this experiment, we serially transplanted tumor fragments two days
529 after the TAM pulse for a total of three transplant generations (TG) (Fig. 3E–F). Each TG was
530 allowed to grow for 21 days and harvested. The percentage of lineage-tagged cells increased
531 from 18% to 35% from TG1 to TG2, then to 61% in TG3. Furthermore, the total percentage of
532 $EGFP^+$/$mNeptune2^+$ cells remained unchanged over time and following serial passaging,
533 indicating tight regulation of STAT signaling and TIC frequency.

534 **Interferon response genes are differentially expressed in TIC**

535 To determine whether STAT reporter-positive cells express a distinct set of genes that may
536 regulate TIC biology, we FACS-enriched all reporter populations from three TNBC xenograft
537 models: two models where STAT reporter activity enriched for TIC (SUM159 and BCM-4272),

538 and one model where STAT reporter activity did not enrich for TIC (BCM-15006), and performed
539 scRNAseq (Fig. 4A).

540 To determine whether EGFP reporter activity could be used to identify TIC transcriptional states
541 in each model, we overlayed sample information on Uniform Manifold Approximation and
542 Projection (UMAP) plots to visualize whether reporter positive cells clustered together (Fig. 4B,
543 Fig. S3B–D). Then, we looked at the percent representation of each cluster within each reporter
544 population to identify clusters that were overrepresented in EGFP positive samples. In SUM159,
545 though all clusters are represented in the EGFP$^+$ cells, positive cells are highly overrepresented
546 in cluster 7 (SUM159-C7) (9.8% of the EGFP$^+$/mNeptune2$^+$, 8.2% of the EGFP$^+$/dsRed$^+$ cells, vs.
547 0.1% of mNeptune2$^+$ cells, and 0.07% of dsRed$^+$ cells) (Fig. 4B, Fig S3B), suggesting it represents
548 a TIC transcriptional state. In BCM-4272, clusters 0 (35% of EGFP$^+$ cells, 0.5% of EGFP$^-$ cells),
549 2 (25% of EGFP$^+$ cells, 10% of EGFP$^-$ cells), 4 (22% of EGFP$^+$ cells, 0.7% of EGFP$^-$ cells), and 7
550 (0.5% of EGFP$^+$ cells, 0.1% of EGFP$^-$ cells) are highly represented in the EGFP$^+$ cells, suggesting
551 at least one of these clusters may represent a TIC transcriptional state (Fig. 4B, Fig. S3A).

552 To clarify which of these clusters most likely represents a TIC transcriptional state, we performed
553 differential gene expression analysis between these four clusters and found that cluster 7
554 differentially expressed genes associated with TIC, including *ABCD1*, *CXCL11*, *NOTCH1*, *RELA*,
555 and *TCF7*, suggesting cluster 7 (4272-C7) represents TIC (Fig. S3B).

556 Because STAT reporter activity did not enrich TIC (Fig. 1C–D) in BCM-15006, we reasoned that
557 a cluster representing a TIC transcriptional state would not be overrepresented in either EGFP$^+$
558 and EGFP$^-$ cells. Cluster 6 (15006-C6) is the only cluster that was equally represented in both
559 EGFP$^+$ and EGFP$^-$ cells (2.1% of EGFP$^+$, 1.9% of EGFP$^-$ cells), suggesting this cluster may
560 represent a TIC transcriptional state (Fig. 4B, Fig S3A).

561 To determine whether established TIC markers CD44$^+$/CD24$^{low/-}$ and ALDH1A1 could be used to
562 confirm the identification of TIC transcriptional states in our data, we generated feature plots

563 based on their expression (Fig. S3C). In all three xenografts, CD44 had no distinct expression
564 pattern and CD24 was highly expressed in all cells in BCM-4272 and BCM-15006, but not
565 expressed in many SUM159 cells. Furthermore, ALDH1A1 was not a distinct marker of any
566 clusters.

567 To gain insights into the biological states of each cluster, we used the differentially expressed
568 genes that define the cells in a given cluster relative to all other cells to perform GSEA
569 (Supplementary Tables 4–6). SUM159-C7, 4272-C7, and 15006-C6, which we hypothesized may
570 represent a TIC transcriptional state, were defined by enrichment of IFNγ response, IFNα
571 response, and TNF-alpha signaling via NF-kB (Fig. 4D, Supplementary Table 3). To investigate
572 the extent to which candidate TIC clusters share transcriptional features, we intersected the
573 differentially expressed genes from each cluster and performed Fisher's exact test for significance
574 (Fig. 4E). SUM159-C7 and 4272-C7 shared 48 features ($p=1.8e^{-24}$), SUM159-C7 and 15006-C6
575 shared 38 features ($p=5.7e^{-27}$), and 4272-C7 and 15006-C6 shared 79 features ($p=1.1e^{-110}$). In
576 this process, we identified only 25 common up-regulated genes across all three clusters, that may
577 represent a new TIC signature. A list of genes from all intersections is shown in Supplementary
578 Table 7.

579 To determine whether these 25 genes participate in shared biological processes, we performed
580 a functional protein-protein interaction analysis using the STRING database (53) (Fig. 4F). We
581 found all of our genes, except GNB4, exist within the same protein-protein interaction network
582 (PPI enrichment: $p<1.0e^{-16}$), which suggests the genes in our TIC signature may be part of a
583 biological process responsible their regulation. To uncover this biological process, we performed
584 an over-representation analysis using the 25-gene TIC signature, which confirmed these genes
585 are overrepresented in gene sets and pathways related to IFN signaling (Fig. 4F, Fig. S4).

586 To determine whether our 25-gene signature can effectively distinguish cell states within each
587 xenograft, we used marker enrichment modeling (35) to generate similarity scores based on the

588 expression of each gene in the 25-gene signature and visualized the results with a heatmap (Fig.
589 4G). In the PDX models, 4272-C7 and 15006-C6 clustered separately from every other cell state
590 in their respective xenografts, demonstrating the 25-gene signature is unique to these clusters. In
591 contrast, SUM159-C7 is a member of the same clade as cluster 5, indicating cluster 5 also is
592 defined by the expression of IFN genes. To distinguish these two clusters, we performed
593 differential gene expression analysis and found SUM159-C7 differentially expressed genes
594 associated with TIC, including *TP63*, *POU5F1*, *ABCA5*, *FOXO1*, and *TCF4*, supporting our
595 hypothesis that SUM159-C7 represents TIC (Fig. S3D).

596 To determine whether additional PDX models have a similar transcriptional state, we analyzed an
597 independent dataset of three PDX models sequenced using isolated nuclei rather than whole cells
598 and gave each cell a TIC score based on the expression of the genes in our 25-gene signature
599 (Fig. S5). All three PDX models (BCM-4013, BCM-15166, and WHIM30) had a single cluster with
600 a high TIC score relative to all other clusters. Thus, our TIC signature defines distinct cell states
601 in three additional PDXs, suggesting this cell state may be represented in a wider cohort of TNBC
602 PDXs.

603 To confirm that IFNs can induce EGFP in cells transduced with the STAT reporter, we treated
604 SUM159 STAT LT, BCM-4272 4M67-EGFP, and BCM-15006 4M67-EGFP cells with IFNα or IFNγ
605 and confirmed that IFNγ activated the reporter in all three models more robustly than did IFNα.
606 IFNα did not activate the reporter in SUM159 (Fig. S6A). To determine whether IFN treatment
607 activates genes that define the IFN-associated TIC transcriptional state, we input our 25 genes
608 into INTERFEROME (54), an online database of IFN-regulated genes, which revealed all 25
609 genes are regulated by IFN (Fig. S6B).

610 To demonstrate these genes can be regulated by IFN in our models, we compared the expression
611 of three upstream regulators of IFN signaling (PARP9, PARP14, STAT1) and three genes
612 activated in response to IFN signaling (ADAR, OAS2, and BST2) from our 25 intersected TIC

613 genes after IFN treatment to untreated controls by qRT-PCR (Fig. S6C). IFNα or IFNγ treatment
614 increased the expression of all six genes relative to the untreated controls in all three TNBC
615 models.

616 To evaluate whether any candidate TIC genes we identified have been previously reported, we
617 compared intersections from any two xenograft models to five previously published TIC signatures
618 that were derived using bulk RNA sequencing, including our own (30-34), and a previously
619 published panel of stem cell and epithelial-to-mesenchymal transition markers (55)
620 (Supplementary Table 8). None of the 25 genes we identified overlapped significantly with any of
621 the six previously published TIC signatures. The Pece et al. (33) signature contained one of our
622 TIC genes. LY6E, from the 4272-C7 and 15006-C6 intersection, whereas the Colacino et al. (55)
623 custom panel only contained CDH3 (SUM159-C7 and 4272-C7 intersection) and MCL1 and
624 ITGA6 (SUM159-C7 and 15006-C6 intersection). Furthermore, a comparison of these previously
625 published TIC signatures did not reveal any significant gene-level commonalities, suggesting that
626 single cell methods may be required for accurate identification of TIC.

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

629 To determine whether a similar IFN/STAT1 transcriptional state could be detected in human
630 clinical samples, we analyzed scRNAseq data retrieved from a published human breast cancer
631 cell atlas that contained TNBC, HER2^{+}, ER^{+}, pre-neoplastic (BRCA1$^{+/-}$), and normal breast
632 samples (37). To identify epithelial cells from breast cancer patients with a similar transcriptional
633 state to our IFN/STAT1 TIC, we gave each cell a TIC score based on the expression of the genes
634 in our 25-gene TIC signature (Fig. 5A–C). This identified cells in cluster 7 in TNBC patients
635 ($p=3.42e^{-113}$), cluster 6 in HER2^{+} patients ($p=3.16e^{-287}$), and cluster 3 in ER^{+} patients ($p=4.71e^{-96}$)
636 had the highest TIC scores relative to cells in all other clusters. To determine whether established
637 TIC markers CD44^{+}/CD24$^{low/-}$ and ALDH1A1 could be used to identify TIC transcriptional states

638 in patient data, we generated feature plots based on their expression (Fig. S7A). In all three breast
639 cancer subtypes, established TIC markers were not markers of any particular cluster, similar to
640 our observations in PDX models.

641 To determine whether a similar transcriptional state exists in pre-neoplastic or normal breast
642 tissue, we applied the same TIC scoring method (Fig. 5D–E). This revealed a similar IFN/STAT1
643 transcriptional state in pre-neoplastic tissue (cluster 2, $p=4.76e^{-115}$), but not in normal breast.
644 These results suggest the IFN/STAT1 transcriptional state is only present in breast cancer and
645 pre-malignant tissue.

646 To determine the extent to which clusters with high TIC scores share transcriptional features with
647 our 25-gene TIC signature, we intersected the genes from each cluster with a high TIC score from
648 each breast cancer subtype, then intersected that gene list with our 25-genes TIC signature (Fig.
649 5F). This revealed the IFN/STAT1 transcriptional state from breast cancer patients shared 9
650 genes with our 25-gene signature (Fisher's exact test: $p=9.6e^{-23}$).

651 To determine whether this similarity extends to pre-neoplastic tissue, we intersected the genes
652 from the high TIC score cluster from pre-neoplastic samples with the high TIC score intersected
653 gene list from all breast cancer patients, then intersected that list with our 25-gene signature (Fig.
654 5G). This revealed the IFN/STAT1 transcriptional state from both breast cancer patients and pre-
655 neoplastic tissue shared 4 genes with our 25-gene signature (Fisher's exact test: $p=2.7e^{-11}$).

656 To determine whether the IFN/STAT1 transcriptional state can be detected in individual breast
657 cancer patients, we analyzed each breast cancer patient of each subtype with more than 1,500
658 non-cycling epithelial cells to increase our chances of identifying rare cell subpopulations, such
659 as TIC (8) (Fig. S7B). We found that the IFN/STAT1 transcriptional state in most, but not all,
660 patients analyzed, which is consistent with TIC heterogeneity. To establish the robustness of the
661 TIC signature, we examined an additional cohort of TNBC patients from an additional human
662 breast cancer scRNAseq dataset (38). We gave each cell a TIC score based on the expression

663 of the genes in our 25-gene TIC signature (Fig. S8A), which revealed cluster 5 had a significantly
664 higher TIC score relative to all other clusters ($p=4.5e^{-113}$). Furthermore, the only patient with more
665 than 1,500 non-cycling epithelial cells also had a cluster defined by a higher TIC score (cluster 4,
666 $p=2.3e^{-03}$) (Fig. S8B). Taken together, our findings reveal that the IFN/STAT1 signaling axis we
667 observe in PDX models is present in human breast cancer patients.

668 **The IFN/STAT1 signaling axis is associated with worse overall survival and lack of**
669 **response to chemotherapy in breast cancer patients**

670 To evaluate the clinical significance of our TIC signature, we examine the association between
671 high expression of our 25-gene TIC signature and overall survival using RNAseq datasets from
672 KM-plotter (40). In all breast cancer patients, regardless of treatment status, the 25-gene TIC
673 signature was associated with a worse overall survival (HR=1.46, p=0.0067) (Fig. 6A). This
674 association with worse overall survival was more pronounced when stratifying for untreated
675 patients (HR=2.33, p=0.0022), but was eliminated when stratifying for patients treated with
676 chemotherapy (HR=0.77, p=0.23) (Fig. 6A), suggesting the TIC signature is only prognostic in
677 untreated patients.

678 Next, we stratified patients by PAM50 subtype and examined differences in overall survival in
679 untreated patients compared to chemotherapy-treated patients (Fig. 6B). In basal, luminal A, and
680 luminal B untreated patients, there was a trend of high TIC score with worse overall survival,
681 which reduced in luminal A chemotherapy-treated patients or reversed in basal and luminal B
682 chemotherapy-treated patients (Fig. 6B). High expression of TIC gene was associated with
683 favorable overall survival in HER2 untreated patients, but this was reduced in chemotherapy-
684 treated patients (Fig. 6B). Normal-like patients had a similar overall survival regardless of
685 chemotherapy treatment (Fig. 6B).

686 We and others showed previously that TIC demonstrate intrinsic resistance to systemic
687 chemotherapies. Thus, we evaluated whether expression of our 25-gene TIC signature is

688 associated with response to chemotherapy in breast cancer patients using ROC plotter (41).
689 Consistent with intrinsic chemotherapy resistance, high expression of our TIC signature was
690 associated with non-responders, and was predictive of lack of response to chemotherapy (AUC:
691 0.593, $p=9.7e^{-04}$) (Fig. 6C). Furthermore, this finding was consistent for all breast cancer subtypes,
692 except for luminal A (Fig. 6C).

693 **BST2 allows TIC enrichment in breast cancer xenograft models.**

694 To demonstrate conclusively that the TIC transcriptional state identifies a subpopulation enriched
695 for TIC function, we leveraged the observation that BST2 was a FACS-compatible cell surface
696 found in the intersection between IFN/STAT1 transcriptional states from all breast cancer
697 subtypes and pre-neoplastic tissue and our 25-gene signature. We performed analytical flow
698 cytometry to examine the relationship between BST2 protein levels and TIC marker
699 $CD44^{+}/CD24^{low/-}$ and ALDH activity. In SUM159 xenografts, $BST2^{+}$ cells are ~40% of the total
700 tumor but represent 96% of $CD44^{+}/CD24^{low/-}$ cells and 58% of $ALDH^{+}$ cells. In BCM-4272
701 xenografts, $BST2^{+}$ cells represented ~20% of the tumor, but 38% of $CD44^{+}/CD24^{low/-}$ cells and
702 92% of $ALDH^{+}$ cells. These results indicate BST2 expression correlates with established TIC
703 markers, suggesting BST2 may also allow enrichment for TIC.

704 To determine whether cells that highly express BST2 are enriched for mammosphere forming
705 potential and/or self-renewal activity we dissociated SUM159 xenografts and FACS-enriched
706 $BST2^{high}$ (top 20%), $BST2^{low/-}$ (bottom 20%), and mock sorted cells to test their primary and
707 secondary mammosphere forming frequencies (Fig. 7B). $BST2^{high}$ cells formed both primary and
708 secondary mammospheres at a greater frequency than either the $BST2^{low/-}$ cells or the mock
709 sorted control, with no statistically significant difference between the mock control and $BST2^{low/-}$.

710 To determine whether self-renewal capacity differed between the two populations, we compared
711 the rate of MSFE between the primary and secondary mammospheres and found the $BST2^{high}$,
712 $BST2^{low/-}$, and mock sorted control form mammospheres at the same rate over passaging (Fig.

713 7C) indicating no differences in self-renewal capacity. Similar results were observed for BCM-
714 4272 (Fig. 7D and 7E). These results indicate that cells with high BST2 expression have enhanced
715 MSFE, but not increased self-renewal capacity.

716 Finally, to test whether BST2 can be used to enrich TIC directly, we compared the tumor-initiating
717 potential of BST2^{high} (top 20%), BST2$^{low/-}$ (bottom 20%), and mock-sorted cells from SUM159
718 xenografts using a limiting-dilution transplantation assay (Fig. 7F). Notably, the BST2^{high} cells
719 were the only population capable of forming tumors at all doses, with a TIC frequency of 1 in
720 1,721 cells. The BST2$^{low/-}$ cells had a 10-fold lower TIC frequency of 1 in 17,370 cells, whereas
721 the mock-sorted group had a modestly lower TIC frequency of 1 in 2,501 cells, likely due to the
722 presence of TIC with lower BST2 expression present in the 60% of cells not assayed.

723 Qualitatively similar results were obtained using BCM-4272 xenografts using a limiting-dilution
724 transplantation assay (Fig. 7G). The BST2^{high} cells were the only population capable of forming
725 tumors at all doses, with a TIC frequency of 1 in 1,102 cells, a much higher TIC frequency than
726 the mock sorted control, which had a TIC frequency of 1 in 8,859 cells. In BST2$^{low/-}$ cells, tumor-
727 initiating potential was depleted relative to the mock sorted control, with a TIC frequency of 1 in
728 18,796 cells. Taken together, our findings indicate BST2 can enrich for IFN/STAT1 signaling TIC.

729 **BST2 regulates TIC frequency in SUM159**

730 To explore whether BST2 regulates TIC function, we generated two BST2 knockdown SUM159
731 cell lines (sh-BST2 #1 and sh-BST2 #2) (Fig. 8A) and compared their primary and secondary
732 MSFE relative to a non-targeting control (Fig. 8B). BST2 knockdown reduced primary MSFE from
733 2.5% in the non-targeting control cells to 1.5% and 1.3% in shBST2 #1 and shBST2 #2,
734 respectively, whereas BST2 knockdown reduced secondary MSFE from 3% in the non-targeting
735 control cells to 1.6% and 1.2% in shBST2 #1 and shBST2 #2, respectively. These results
736 demonstrate BST2 regulates mammosphere-forming frequency in SUM159 cells.

737 To determine whether BST2 regulates self-renewal, we compared the rate of MSFE between
738 primary and secondary mammospheres in BST2 knockdown cells relative to a non-targeting
739 control (Fig. 8C). We found both BST2 knockdown cells formed mammospheres at the same rate
740 over passaging compared to the non-targeting control, indicating that BST2 does not regulate
741 self-renewal. Taken together, our data suggest BST2 functionally regulates TIC frequency
742 independent of self-renewal.

743 Because IFN regulates the expression of genes in our TIC signature, including BST2 (Fig. S6B–
744 C), we grew SUM159 mammospheres in the presence of IFNα or IFNγ and compared primary
745 and secondary mammosphere-forming frequency to an untreated control (Fig. 8D). Both IFNα
746 and IFNγ increased primary mammosphere frequency to 6.2% and 7.5%, respectively, compared
747 to the untreated control (3.4%). Similarly, IFNα and IFNγ increased secondary mammosphere
748 frequency to 5% and 9.3%, respectively, compared to the untreated control (2.9%). However, only
749 IFNγ increased rate of MSFE over passaging (from 7.5% to 9.3%), suggesting IFNγ regulates
750 both TIC frequency and self-renewal in SUM159 (Fig. 8E).

751 To determine whether BST2 is responsible for the IFN-mediated increase in MSFE, we compared
752 the effect of IFN on MSFE in two BST2 knockdown SUM159 cell lines (shBST2 #1 and shBST2
753 #2) compared to a non-targeting control (Fig. 8F). In the non-targeting control, IFNα increased
754 MSFE to 4.4% and IFNγ increased MSFE to 5.1% relative to the untreated control (2.7%). In both
755 BST2 knockdown lines, neither IFNα nor IFNγ increased MSFE relative to the untreated control,
756 suggesting BST2 is a mediator of the IFN-mediated increase in MSFE.

757 ALDH activity is a TIC marker in SUM159 (56). To evaluate whether BST2 knockdown reduces
758 ALDH activity, we performed an ALDEFLUOR assay in BST2 knockdown SUM159 cell lines
759 relative to a non-targeting control. BST2 knockdown reduced ALDH activity from 7.0% in the non-
760 targeting control cells to 1.6% and 0.8% in shBST2 #1 and shBST2 #2, respectively (Fig. 8G).

761 To determine whether this increase in TIC is due to proliferation, we compared cell cycle changes
762 following BST2 knockdown. Of the non-targeting control cells, 33.7% were proliferative, whereas
763 only 25.4% in shBST2 #1, and 13.1% in shBST2 #2, were proliferative, indicating BST2
764 knockdown reduces proliferation (Fig. 8H). Taken together, our data suggest that BST2 regulates
765 TIC frequency in SUM159, perhaps through modulating proliferation.

766

767 **DISCUSSION**

768 Herein, we extend our previous finding that STAT reporter activity enriches for TIC in two cell line
769 xenograft models of claudin-low TNBC (11,57), and identified PDX BCM-4272 as an additional
770 non-claudin-low model where STAT reporter activity is associated with TIC functions. Despite this
771 finding, the reporter did not enrich cells with enhanced MSFE in the three other TNBC PDX
772 models. Data are consistent with previous findings that STAT-mediated transcription has
773 function(s) not necessarily associated with TIC. For example, STAT3 can promote immune
774 evasion, metastasis, and chemoresistance (58), whereas STAT1 is associated with both tumor-
775 promoting functions, such as immune suppression and metastasis, and tumor-suppressor
776 functions, such as promoting apoptosis and immunosurveillance (59).

777 To address shortcomings of our original STAT-EGFP reporter, we developed the LT system to
778 provide an inducible fluorescent label to tag STAT signaling TIC and other cells permanently as
779 they undergo cell state changes. In addition, there are key questions in the cancer stem cell field
780 that this system can address, including whether TIC from primary tumors function as metastasis-
781 initiating cells (52,60), and whether enriched TIC populations following chemotherapy are the
782 result of a therapeutic selective pressure, or are activated in response to the therapy (52,61).

783 We leveraged a synergy between scRNAseq, which has the power to detect rare cell populations,
784 with our reporter systems to identify more precisely the TIC-enriched subpopulation of cells in
785 multiple models of TNBC. The established TIC markers $CD44^{+}$/ $CD24^{low/-}$ and ALDH1A1 did not
786 identify any unique cell states in our scRNAseq data, suggesting the expression of established
787 TIC markers alone may not be sufficient to identify TIC in scRNAseq datasets. However, our
788 analyses revealed that IFN response genes are differentially expressed in TIC in all three of our
789 reporter-tagged TNBC, as well as three PDXs from an additional cohort, indicating the IFN/STAT1
790 signaling axis may be a common feature of TIC in a subset of TNBC. Critically, the IFN/STAT1
791 transcriptional state defines distinct cell states in all three breast cancer subtypes, as well as pre-

792 neoplastic tissue. The absence of the IFN/STAT1 state in normal breast suggests this
793 transcriptional state is a unique feature of some pre-neoplastic and neoplastic tissues, and may
794 represent a candidate cancer-specific therapeutic target.

795 From an evolutionary perspective, the activation of IFN response genes in tumor cells presents
796 multiple possibilities. On the one hand, TIC may have co-opted the IFN-mediated inflammatory
797 response of the immune system to regulate TIC processes related to their frequency, such as
798 differentiation, or cell state decisions (e.g. cellular quiescence/activation). Alternatively, and
799 arguably more evolutionarily likely given that multicellular organisms arose from a single self-
800 renewing cell, it is also possible that the immune system may have co-opted a stem cell pathway
801 to mediate inflammatory responses.

802 Using the common genes across TIC states in all three PDX models in scRNAseq data, we
803 derived a 25-gene TIC signature. Almost all genes we identified were not reported in previously
804 published TIC signatures, all of which were derived using bulk RNA sequencing (30-34,55).
805 Indeed, we found little agreement among the previously published TIC signatures themselves.
806 This suggests that it is essential to combine TIC enrichment with scRNAseq to derive high-
807 resolution TIC signatures.

808 There are limitations to consider when comparing bulk and single-cell samples. Bulk RNA
809 sequencing is frequently performed at much higher coverage and depth than typical scRNAseq,
810 which can allow detection of low expressing genes that may be missed by single-cell approaches.
811 On the other hand, scRNAseq can dissect transcriptome heterogeneity, allowing discovery of rare
812 cell types/states that would otherwise be obscured in bulk RNAseq.

813 With respect to translational potential, our 25-gene TIC signature was associated with worse
814 overall survival in untreated breast cancer patients, but not in patients treated with chemotherapy.
815 However, the signature was predictive of lack of response to chemotherapy, in line with previous
816 findings that TIC are intrinsically resistant to systemic therapies (15). These data are also

817 consistent with previous studies demonstrating enhanced expression of IFN genes following
818 cytotoxic chemotherapy (62). Indeed, a tumor cell state defined by IFN gene expression was
819 recently identified in TNBC scRNAseq, and was shown to be associated with tumors resistant to
820 standard-of-care chemotherapies (63). This raises the possibility that TIC exploit the IFN/STAT1
821 axis to promote their activity in cancer, including mediating chemoresistance (64).

822 We identified BST2 as a candidate cell surface TIC marker, as it was differentially expressed in
823 candidate TIC clusters across the three TNBC models we sequenced, and was a marker of
824 clusters with a high TIC score in each breast cancer subtype, as well as pre-neoplastic tissue.
825 BST2 is a transmembrane glycoprotein, whose expression is mediated by IFN, that is well-known
826 for tethering newly-synthesized enveloped virions to the host cell plasma membrane (65) and
827 activating NFκB to induce an antiviral state (66). To our knowledge, BST2 has never been
828 associated with TIC in any tumor type.

829 Beyond utility as a TIC marker, our results demonstrate that BST2 is a regulator of TIC frequency
830 in SUM159 and PDX BCM-4272, yet its knockdown in SUM159 did not affect self-renewal. This
831 indicates BST2 must regulate TIC frequency by a different mechanism, such as quorum sensing.
832 Quorum sensing, which is most commonly studied in bacteria, is a process of cell-to-cell
833 communication that allows cells to share information about cell density and adjust their gene
834 expression to regulate cell number (67). Organ-level quorum sensing has also been studied in
835 stem cells, such as hair stem cell populations (68). To the extent that we have studied it, tumors
836 appear to have specific TIC frequencies that are intrinsic properties of each tumor (69). If so, it
837 stands to reason quorum sensing is one mechanism that regulates TIC frequency. As a cell
838 surface protein, BST2 may be an integral component of the quorum sensing mechanism that
839 regulates TIC frequency.

840

841 **ACKNOWLEDGMENTS**

842 This work was supported, in part, by an F31 predoctoral fellowship (NCI CA278274) (to E.P.S), a
843 U54 PDXNet PDX Development and Trials Center grant (NCI U54 CA224076) (to M.T.L.), an R01
844 grant (CA224867 to M.T.L. and H.L.F), a P30 Cancer Center Support Grant CA125123 (To C.
845 Kent Osborne), a Core Facility Support Grant from the Cancer Research and Prevention Initiative
846 of Texas RP170691 (to M.T.L.), the authors also acknowledge the joint participation by Diana
847 Henry Helis Medical Research Foundation through its direct engagement in the continuous active
848 conduct of medical research in conjunction with Baylor College of Medicine, the Komen for the
849 Cure Promise grant, and a generous gift from the Korell family for the study of triple-negative
850 breast cancer. S.S.Y. was supported by NIH grant R35GM133658 and Susan G. Komen
851 Foundation grant CCR19609287. This project was also supported by the Cytometry and Cell
852 Sorting Core at Baylor College of Medicine with funding from the CPRIT Core Facility Support
853 Award (CPRIT-RP180672), the NIH (P30CA125123 and RR024574), the Single Cell Genomics
854 Core at BCM with funding from NIH shared instrument grants (S10OD023469, S10OD025240)
855 and P30EY002520, and the Genomic and RNA Profiling Core at Baylor College of Medicine with
856 funding from NIH NCI (P30CA125123) and CPRIT (RP200504) grants. We thank Mu Wang for
857 cloning the STAT lineage-tracing vectors as well as Dr. Koen Venken for helpful suggestions for
858 vector design.

859

860 **REFERENCES**

861 1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA*, et al.* Molecular
862 portraits of human breast tumours. Nature **2000**;406:747-52

863 2. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J*, et al.* Gene expression profiling in
864 breast cancer: understanding the molecular basis of histologic grade to improve
865 prognosis. J Natl Cancer Inst **2006**;98:262-72

866 3. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA*, et al.* Triple-negative
867 breast cancer: clinical features and patterns of recurrence. Clin Cancer Res
868 **2007**;13:4429-34

869 4. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y*, et al.*
870 Identification of human triple-negative breast cancer subtypes and preclinical models for
871 selection of targeted therapies. J Clin Invest **2011**;121:2750-67

872 5. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours.
873 Nature **2012**;490:61-70

874 6. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells.
875 Nature **2001**;414:105-11

876 7. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M*, et al.* ALDH1 is
877 a marker of normal and malignant human mammary stem cells and a predictor of poor
878 clinical outcome. Cell Stem Cell **2007**;1:555-67

879 8. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective
880 identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A **2003**;100:3983-
881 8

882 9. Iliopoulos D, Hirsch HA, Wang G, Struhl K. Inducible formation of breast cancer stem cells
883 and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. Proc Natl Acad
884 Sci U S A **2011**;108:1397-402

885 10. Marotta LL, Almendro V, Marusyk A, Shipitsin M, Schemme J, Walker SR*, et al.* The
886 JAK2/STAT3 signaling pathway is required for growth of CD44(+)CD24(-) stem cell-like
887 breast cancer cells in human tumors. J Clin Invest **2011**;121:2723-35

888 11. Wei W, Tweardy DJ, Zhang M, Zhang X, Landua J, Petrovic I*, et al.* STAT3 signaling is
889 activated preferentially in tumor-initiating cells in claudin-low models of human breast
890 cancer. Stem Cells **2014**;32:2571-82

891 12. Zhang M, Atkinson RL, Rosen JM. Selective targeting of radiation-resistant tumor-initiating
892 cells. Proc Natl Acad Sci U S A **2010**;107:3522-7

893 13. Dierks C, Beigi R, Guo GR, Zirlik K, Stegert MR, Manley P*, et al.* Expansion of Bcr-Abl-
894 positive leukemic stem cells is dependent on Hedgehog pathway activation. Cancer Cell
895 **2008**;14:238-49

896 14. Harrison H, Farnie G, Howell SJ, Rock RE, Stylianou S, Brennan KR*, et al.* Regulation of
897 breast cancer stem cell activity by signaling through the Notch4 receptor. Cancer Res
898 **2010**;70:709-18

899 15. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF*, et al.* Intrinsic resistance of
900 tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst **2008**;100:672-9

901 16. Korkaya H, Kim GI, Davis A, Malik F, Henry NL, Ithimakin S*, et al.* Activation of an IL6
902 inflammatory loop mediates trastuzumab resistance in HER2+ breast cancer by
903 expanding the cancer stem cell population. Mol Cell **2012**;47:570-84

904 17. Alamgeer M, Ganju V, Kumar B, Fox J, Hart S, White M*, et al.* Changes in aldehyde
905 dehydrogenase-1 expression during neoadjuvant chemotherapy predict outcome in locally
906 advanced breast cancer. Breast Cancer Res **2014**;16:R44

907 18. Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C*, et al.* let-7 regulates self renewal and
908 tumorigenicity of breast cancer cells. Cell **2007**;131:1109-23

909 19. Liu H, Patel MR, Prescher JA, Patsialou A, Qian D, Lin J*, et al.* Cancer stem cells from
910 human breast tumors are involved in spontaneous metastases in orthotopic mouse
911 models. Proc Natl Acad Sci U S A **2010**;107:18115-20

912 20. Matsuda T, Cepko CL. Controlled expression of transgenes introduced by in vivo
913 electroporation. Proc Natl Acad Sci U S A **2007**;104:1027-32

914 21. Lou E, Fujisawa S, Morozov A, Barlas A, Romin Y, Dogan Y*, et al.* Tunneling nanotubes
915 provide a unique conduit for intercellular transfer of cellular contents in human malignant
916 pleural mesothelioma. PLoS One **2012**;7:e33093

917 22. Koo BK, Stange DE, Sato T, Karthaus W, Farin HF, Huch M*, et al.* Controlled gene
918 expression in primary Lgr5 organoid cultures. Nat Methods **2011**;9:81-3

919 23. Stirling DR, Swain-Bowden MJ, Lucas AM, Carpenter AE, Cimini BA, Goodman A.
920 CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics
921 **2021**;22:433

922 24. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T*, et al.* Fiji: an
923 open-source platform for biological-image analysis. Nat Methods **2012**;9:676-82

924 25. Hu Y, Smyth GK. ELDA: extreme limiting dilution analysis for comparing depleted and
925 enriched populations in stem cell and other assays. J Immunol Methods **2009**;347:70-8

926 26. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ*, et al.* Single-cell RNA-seq
927 reveals changes in cell cycle and differentiation programs upon aging of hematopoietic
928 stem cells. Genome Res **2015**;25:1860-72

929 27. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A*, et al.* Integrated
930 analysis of multimodal single-cell data. Cell **2021**;184:3573-87 e29

931 28. Andreatta M, Carmona SJ. UCell: Robust and scalable single-cell gene signature scoring.
932 Comput Struct Biotechnol J **2021**;19:3796-8

933 29. Shen L. GeneOverlap: Test and visualize gene overlaps. R Package **2023**;1

934 30. Charafe-Jauffret E, Ginestier C, Iovino F, Wicinski J, Cervera N, Finetti P*, et al.* Breast
935 cancer cell lines contain functional cancer stem cells with metastatic capacity and a
936 distinct molecular signature. Cancer Res **2009**;69:1302-13

937 31. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A*, et al.* Residual breast
938 cancers after conventional therapy display mesenchymal as well as tumor-initiating
939 features. Proc Natl Acad Sci U S A **2009**;106:13820-5

940 32. Leth-Larsen R, Terp MG, Christensen AG, Elias D, Kuhlwein T, Jensen ON*, et al.*
941 Functional heterogeneity within the CD44 high human breast cancer stem cell-like
942 compartment reveals a gene signature predictive of distant metastasis. Mol Med
943 **2012**;18:1109-21

944 33. Pece S, Disalvatore D, Tosoni D, Vecchi M, Confalonieri S, Bertalot G*, et al.* Identification
945 and clinical validation of a multigene assay that interrogates the biology of cancer stem
946 cells and predicts metastasis in breast cancer: A retrospective consecutive study.
947 EBioMedicine **2019**;42:352-62

948 34. Pei J, Wang Y, Li Y. Identification of key genes controlling breast cancer stem cell
949 characteristics via stemness indices analysis. J Transl Med **2020**;18:74

950 35. Diggins KE, Greenplate AR, Leelatian N, Wogsland CE, Irish JM. Characterizing cell
951 subsets using marker enrichment modeling. Nat Methods **2017**;14:275-8

952 36. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z*, et al.* clusterProfiler 4.0: A universal enrichment
953 tool for interpreting omics data. Innovation (Camb) **2021**;2:100141

954 37. Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X*, et al.* A single-cell RNA
955 expression atlas of normal, preneoplastic and tumorigenic states in the human breast.
956 EMBO J **2021**;40:e107333

957 38. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A*, et al.* A single-cell
958 and spatially resolved atlas of human breast cancers. Nat Genet **2021**;53:1334-47

959 39. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K*, et al.* Fast, sensitive and
960 accurate integration of single-cell data with Harmony. Nat Methods **2019**;16:1289-96

961 40. Gyorffy B. Transcriptome-level discovery of survival-associated biomarkers and therapy
962 targets in non-small-cell lung cancer. Br J Pharmacol **2024**;181:362-74

963 41. Fekete JT, Gyorffy B. ROCplot.org: Validating predictive biomarkers of
964 chemotherapy/hormonal therapy/anti-HER2 therapy using transcriptomic data of 3,104
965 breast cancer patients. Int J Cancer **2019**;145:3140-51

966 42. Ehret GB, Reichenbach P, Schindler U, Horvath CM, Fritz S, Nabholz M*, et al.* DNA
967 binding specificity of different STAT proteins. Comparison of in vitro specificity with natural
968 target sites. J Biol Chem **2001**;276:6675-88

969 43. Wang Y, Wang Q, Tang CH, Chen HD, Hu GN, Shao JK*, et al.* p-STAT3 expression in
970 breast cancer correlates negatively with tumor size and HER2 status. Medicine (Baltimore)
971 **2021**;100:e25124

972 44. Petrosyan V, Dobrolecki LE, Thistlethwaite L, Lewis AN, Sallas C, Srinivasan RR*, et al.*
973 Identifying biomarkers of differential chemotherapy response in TNBC patient-derived
974 xenografts with a CTD/WGCNA approach. iScience **2023**;26:105799

975 45. Zhang X, Claerhout S, Prat A, Dobrolecki LE, Petrovic I, Lai Q*, et al.* A renewable tissue
976 resource of phenotypically stable, biologically and ethnically diverse, patient-derived
977 human breast cancer xenograft models. Cancer Res **2013**;73:4885-97

978 46. Corish P, Tyler-Smith C. Attenuation of green fluorescent protein half-life in mammalian
979 cells. Protein Eng **1999**;12:1035-40

980 47. Groner B. Determinants of the extent and duration of STAT3 signaling. JAKSTAT
981 **2012**;1:211-5

982 48. Meyer MJ, Fleming JM, Ali MA, Pesesky MW, Ginsburg E, Vonderhaar BK. Dynamic
983 regulation of CD24 and the invasive, CD44posCD24neg phenotype in breast cancer cell
984 lines. Breast Cancer Res **2009**;11:R82

985 49. Reid JM, Goetz MP, Buhrow SA, Walden C, Safgren SL, Kuffel MJ*, et al.*
986 Pharmacokinetics of endoxifen and tamoxifen in female mice: implications for comparative
987 in vivo activity studies. Cancer Chemother Pharmacol **2014**;74:1271-8

988 50. Horvath CM, Wen Z, Darnell JE, Jr. A STAT protein domain that determines DNA
989 sequence recognition suggests a novel DNA-binding domain. Genes Dev **1995**;9:984-94

990 51. Guo C, Yang W, Lobe CG. A Cre recombinase transgene with mosaic, widespread
991 tamoxifen-inducible action. Genesis **2002**;32:8-18

992 52. Batlle E, Clevers H. Cancer stem cells revisited. Nat Med **2017**;23:1124-34

993 53. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R*, et al.* The STRING
994 database in 2023: protein-protein association networks and functional enrichment
995 analyses for any sequenced genome of interest. Nucleic Acids Res **2023**;51:D638-D46

996 54. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H*, et al.* Interferome v2.0: an
997 updated database of annotated interferon-regulated genes. Nucleic Acids Res
998 **2013**;41:D1040-6

999 55. Colacino JA, Azizi E, Brooks MD, Harouaka R, Fouladdel S, McDermott SP*, et al.*
1000 Heterogeneity of Human Breast Stem and Progenitor Cells as Revealed by Transcriptional
1001 Profiling. Stem Cell Reports **2018**;10:1596-609

1002 56. Lin L, Hutzen B, Lee HF, Peng Z, Wang W, Zhao C*, et al.* Evaluation of STAT3 signaling
1003 in ALDH+ and ALDH+/CD44+/CD24- subpopulations of breast cancer cells. PLoS One
1004 **2013**;8:e82821

1005 57. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T*, et al.* A collection of breast
1006 cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell
1007 **2006**;10:515-27

1008 58. Huynh J, Chand A, Gough D, Ernst M. Therapeutically exploiting STAT3 activity in cancer
1009 - using tissue repair as a road map. Nat Rev Cancer **2019**;19:82-96

1010 59. Meissl K, Macho-Maschler S, Muller M, Strobl B. The good and the bad faces of STAT1
1011 in solid tumours. Cytokine **2017**;89:12-20

1012 60. Massague J, Ganesh K. Metastasis-Initiating Cells and Ecosystems. Cancer Discov
1013 **2021**;11:971-94

1014 61. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T*, et al.* Chemoresistance Evolution
1015 in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell
1016 **2018**;173:879-93 e13

1017 62. Sistigu A, Yamazaki T, Vacchelli E, Chaba K, Enot DP, Adam J*, et al.* Cancer cell-
1018 autonomous contribution of type I interferon signaling to the efficacy of chemotherapy. Nat
1019 Med **2014**;20:1301-9

1020 63. Jacobo Jacobo M, Donnella HJ, Sobti S, Kaushik S, Goga A, Bandyopadhyay S. An
1021 inflamed tumor cell subpopulation promotes chemotherapy resistance in triple negative
1022 breast cancer. Sci Rep **2024**;14:3694

1023 64. Zaidi MR, Merlino G. The two faces of interferon-gamma in cancer. Clin Cancer Res
1024 **2011**;17:6118-24

1025 65. Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC*, et al.* The
1026 interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the
1027 cell surface by the viral Vpu protein. Cell Host Microbe **2008**;3:245-52

1028 66. Tokarev A, Suarez M, Kwan W, Fitzpatrick K, Singh R, Guatelli J. Stimulation of NF-
1029 kappaB activity by the HIV restriction factor BST2. J Virol **2013**;87:2046-57

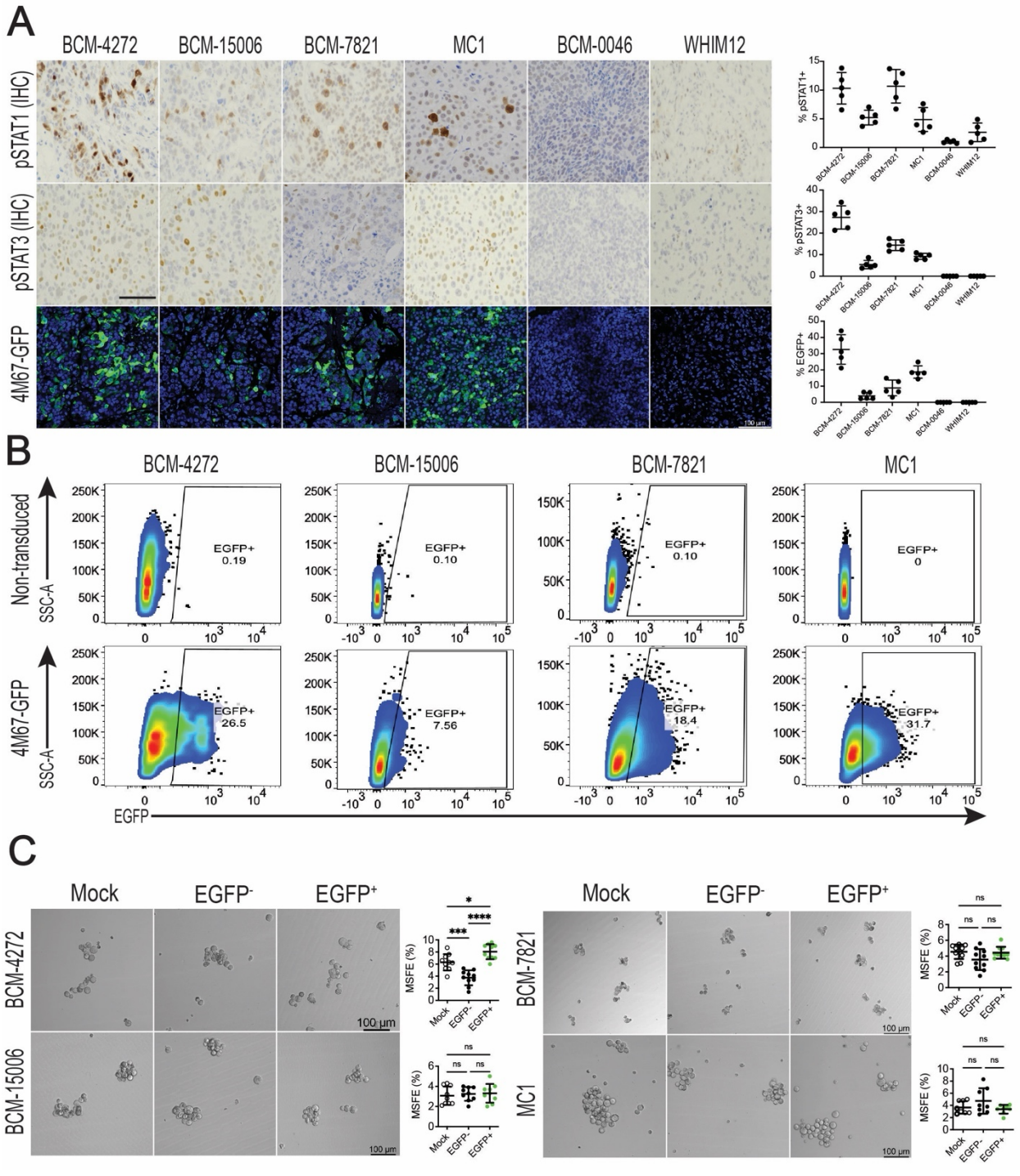
1030 67. Rutherford ST, Bassler BL. Bacterial quorum sensing: its role in virulence and possibilities
1031 for its control. Cold Spring Harb Perspect Med **2012**;2

1032 68. Chen CC, Wang L, Plikus MV, Jiang TX, Murray PJ, Ramos R*, et al.* Organ-level quorum
1033 sensing directs regeneration in hair stem cell populations. Cell **2015**;161:277-90

1034 69. Schatton T, Murphy GF, Frank NY, Yamaura K, Waaga-Gasser AM, Gasser M*, et al.*
1035 Identification of cells initiating human melanomas. Nature **2008**;451:345-9
1036

1037

1038 **FIGURES**



1039

1040 **Figure 1: STAT reporter activity in breast PDX models.** A) Representative IHC images

1041 showing levels of pSTAT1 and pSTAT3 in a panel of PDX models. 4M67-EGFP reporter activity

1042 was visualized by immunofluorescence (IF). Scale bar = 50 µm. IHC and IF were quantified using

1043 5 representative images per PDX model. B) Flow cytometry plot illustrating the percentage of
1044 4M67 reporter-positive cells in PDX models with demonstrated reporter activity. C)
1045 Representative images comparing the mammosphere forming efficiency between the $EGFP^+$,
1046 $EGFP^-$, and mock-sorted cells in four PDX models (BCM-4272: n = 10, BCM-15006: n = 9, BCM-
1047 7821: n = 10, MC1: n = 8) with a bar graph of the mammosphere forming efficiency. Values are
1048 mean ± SD. One-way ANOVA and Tukey's post hoc test were used to determine statistical
1049 significance. *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$. MSFE: mammosphere forming efficiency.
1050

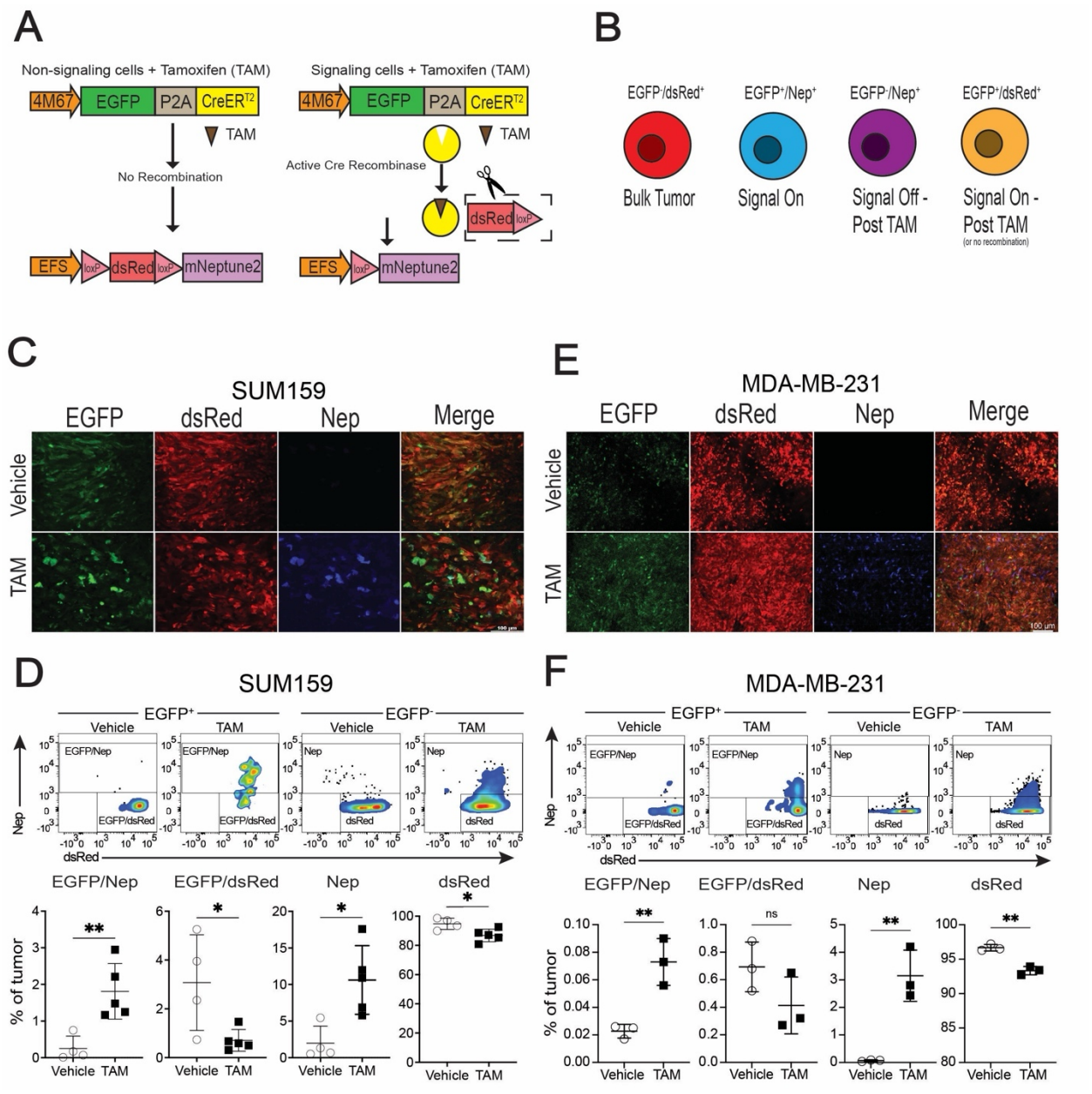


Figure 2: STAT lineage-tracing reporter provides a long-term fluorescent label to STAT signaling cells. A) Schematic illustrating how the STAT lineage-tracing system functions upon the addition of TAM. Cre is not transcribed in non-signaling (EGFP$^-$) cells, thus no color-switching occurs in the presence 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 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

1059 each LT population two days after 4 mg TAM treatment (Untreated: n = 4, TAM-treated: n = 5).
1060 E) Confocal images of MDA-MB-231 STAT LT xenografts and F) flow cytometry quantification of
1061 the percent of each LT population two days after 4 mg TAM treatment (n = 3). One-way ANOVA
1062 and Tukey's post hoc test were used to determine statistical significance. *$p < 0.05$, **$p < 0.01$,
1063 ns = not significant; values are mean ± SD. Nep: mNeptune2.
1064

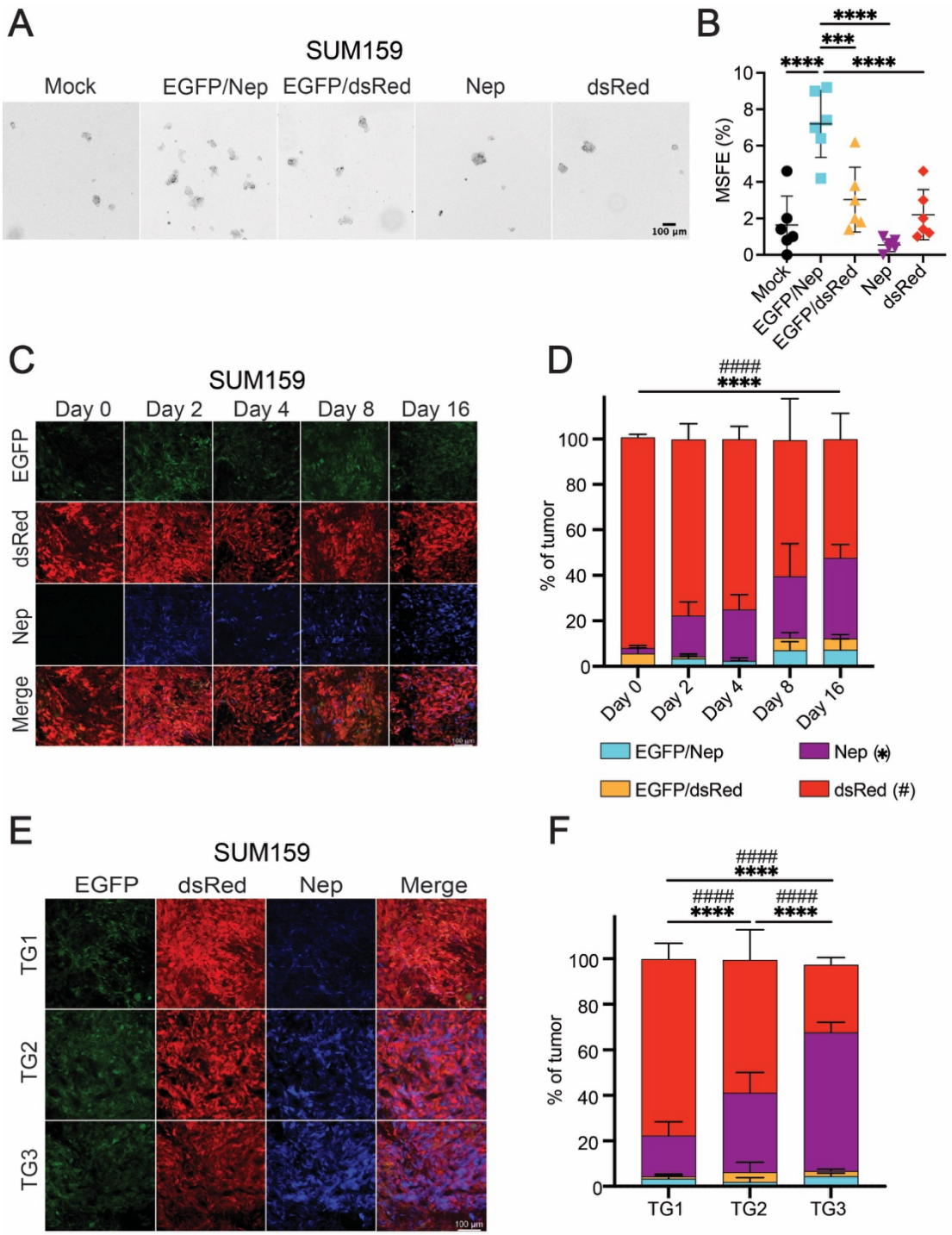
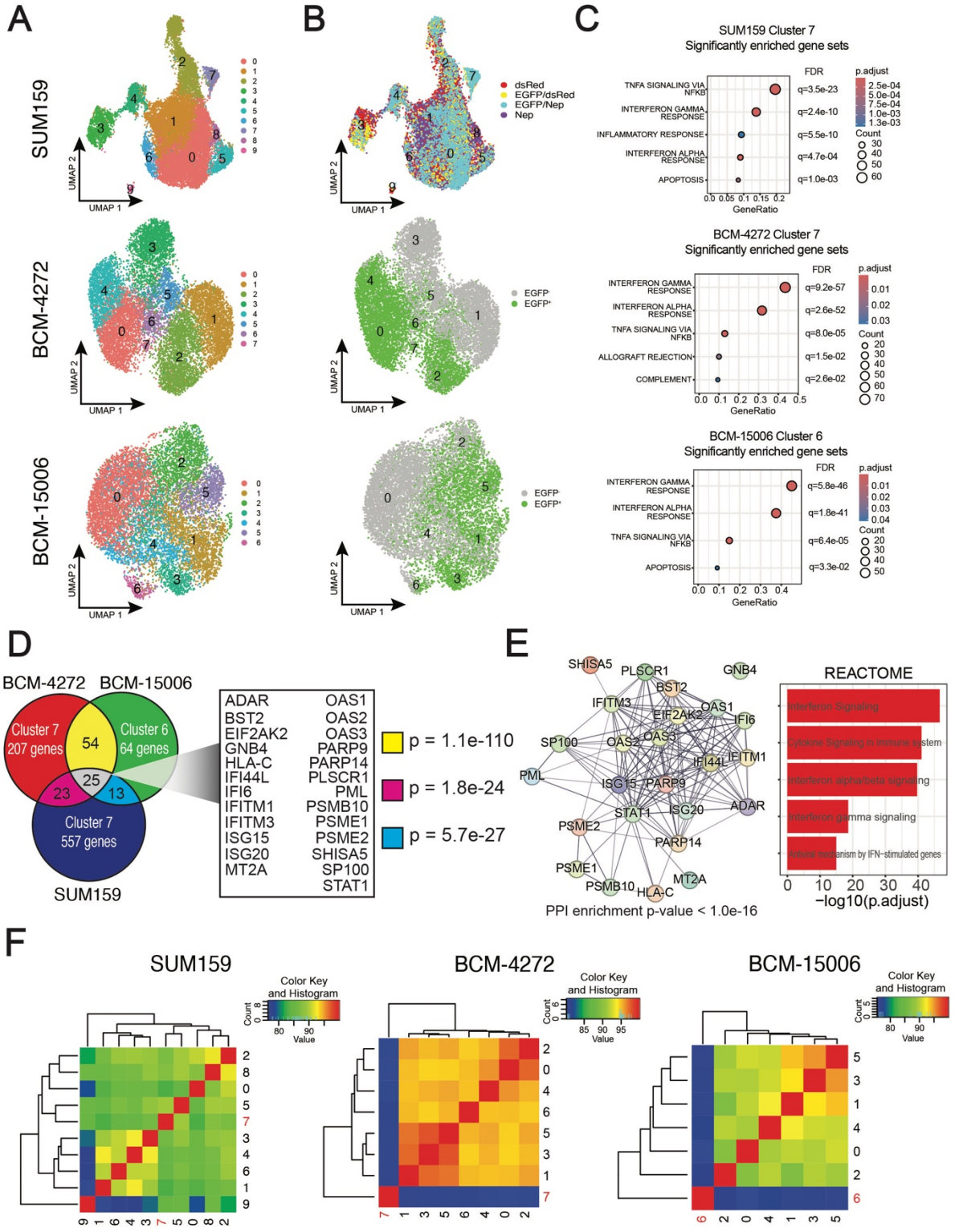


Figure 3: Long-lived STAT signaling cells are enriched for TIC in SUM159 xenografts. A) SUM159 STAT LT xenografts were dissociated, then each LT population was FACS enriched and grown as mammospheres. B) Mammosphere forming efficiency of each LT cell population from SUM159 xenografts (*n* = 6). One-way ANOVA and Tukey's post hoc test were used to determine

1070 statistical significance. C) Confocal images of a 16-day TAM time course in SUM159 xenografts
1071 (*n* = 6). D) Flow cytometry quantification of the percentage of each LT cell population in SUM159
1072 xenografts at five different time points. E) The time course was extended using a serial
1073 transplantation assay. Tumor fragments harvested two days post TAM (TG1), were transplanted
1074 into new hosts two additional times (*n* = 6). F) Flow cytometry quantification of the percentage of
1075 each LT cell population in SUM159 xenografts at each transplant generation (TG). ***$p < 0.001$,
1076 ****$p < 0.0001$, $^{\#\#\#\#}p < 0.0001$; values are mean ± SD. Nep: mNeptune2.
1077


A
SUM159
BCM-4272
BCM-15006
UMAP 1
UMAP 2
B
dsRed
EGFP/dsRed
EGFP/Nep
Nep
EGFP-
EGFP+
C
SUM159 Cluster 7
Significantly enriched gene sets
TNFA SIGNALING VIA NFKB
INTERFERON GAMMA RESPONSE
INFLAMMATORY RESPONSE
INTERFERON ALPHA RESPONSE
APOPTOSIS
q=3.5e-23
q=2.4e-10
q=5.5e-10
q=4.7e-04
q=1.0e-03
BCM-4272 Cluster 7
Significantly enriched gene sets
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
TNFA SIGNALING VIA NFKB
ALLOGRAFT REJECTION
COMPLEMENT
q=9.2e-57
q=2.6e-52
q=8.0e-05
q=1.5e-02
q=2.6e-02
BCM-15006 Cluster 6
Significantly enriched gene sets
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
TNFA SIGNALING VIA NFKB
APOPTOSIS
q=5.8e-46
q=1.8e-41
q=6.4e-05
q=3.3e-02
FDR
p.adjust
Count
GeneRatio
D
BCM-4272
BCM-15006
Cluster 7 207 genes
Cluster 6 64 genes
Cluster 7 557 genes
54
25
23
13
SUM159
ADAR
BST2
EIF2AK2
GNB4
HLA-C
IFI44L
IFI6
IFITM1
IFITM3
ISG15
ISG20
MT2A
OAS1
OAS2
OAS3
PARP9
PARP14
PLSCR1
PML
PSMB10
PSME1
PSME2
SHISA5
SP100
STAT1
p = 1.1e-110
p = 1.8e-24
p = 5.7e-27
E
REACTOME
Interferon Signaling
Cytokine Signaling in Immune system
Interferon alpha/beta signaling
Interferon gamma signaling
Antiviral mechanism by IFN-stimulated genes
−log10(p.adjust)
PPI enrichment p-value < 1.0e-16
F
SUM159
BCM-4272
BCM-15006
Color Key and Histogram
Count
Value

1078

1079 **Figure 4: Interferon response genes are differentially expressed in TIC.** SUM159 STAT3 LT,
1080 BCM-4272 4M67-EGFP, and BCM-15006 4M67-EGFP xenografts were harvested and
1081 dissociated, then each fluorescent population was enriched and sequenced using 10X genomics.
1082 A) Each xenograft was clustered and visualized using a UMAP plot. B) UMAP with cell identity
1083 overlayed on the plot. C) Hallmark gene set enrichment analysis of candidate TIC transcriptional
1084 state from each xenograft. D) Intersection of the upregulated genes in candidate TIC clusters with
1085 accompanying pairwise comparisons using Fisher's exact test. Intersection of candidate TIC
1086 clusters from each model reveals 25 shared genes. E) STRING database illustration of protein-
1087 protein interactions between the 25-gene TIC signature and overrepresentation analysis results
1088 using the 25-gene list. F) Marker enrichment modeling heatmap illustrating cluster-to-cluster
1089 similarity based on expression of the genes in the 25-gene TIC signature. Nep: mNeptune2, FDR:
1090 false discovery rate, p.adjust: Benjamini-Hochberg (BH)-adjusted p-values.
1091

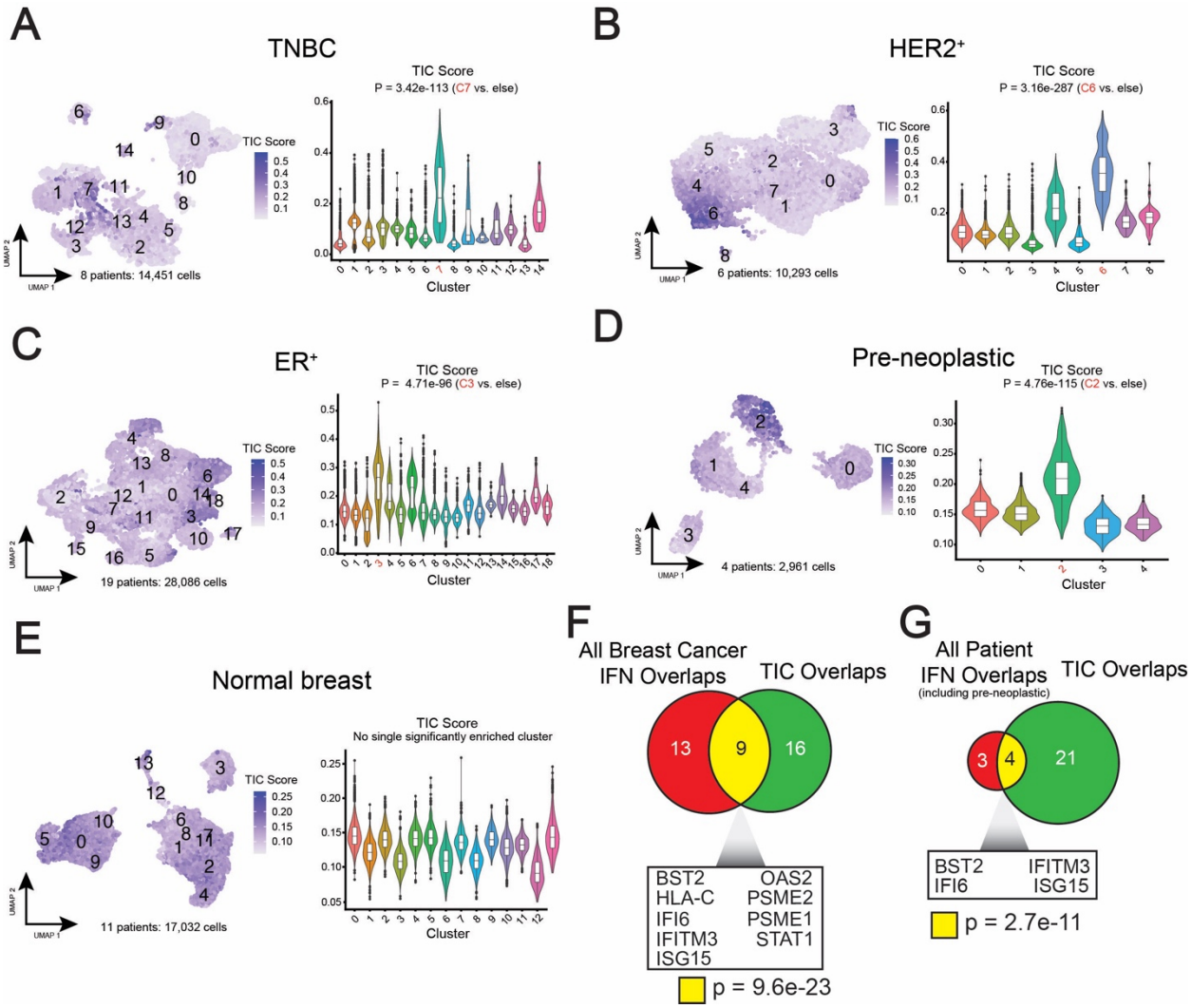


Figure 5: Interferon response genes are differentially expressed in a subset of epithelial cells in human breast cancer patient samples. A) UMAP visualization of eight human TNBC patient samples with feature plot and violin plot showing TIC score, derived from based on the expression of our 25-gene TIC signature. B) UMAP visualization of six human HER2^{+} patient samples with TIC score feature and violin plots. C) UMAP visualization of 19 human ER^{+} patient samples with TIC score feature and violin plots. D) UMAP visualization of four human pre-neoplastic (*BRCA1*$^{+/-}$ tissue) patient samples with TIC score feature and violin plots. E) UMAP visualization of 11 human normal breast samples with TIC score feature and violin plots. In all TIC score violin plots, P values were calculated using a two-sided Wilcoxon test. F) The intersection

1102 between the common differentially expressed genes from clusters that were defined by a high
1103 TIC score across all three breast cancer subtypes and our 25-gene TIC signature. G) The
1104 intersection between the common differentially expressed genes from clusters that were defined
1105 by a high TIC score across all three breast cancer subtypes and pre-neoplastic tissue and our 25-
1106 gene TIC signature. The significance of gene list comparisons was calculated using Fisher's exact
1107 test. IFN: interferon
1108

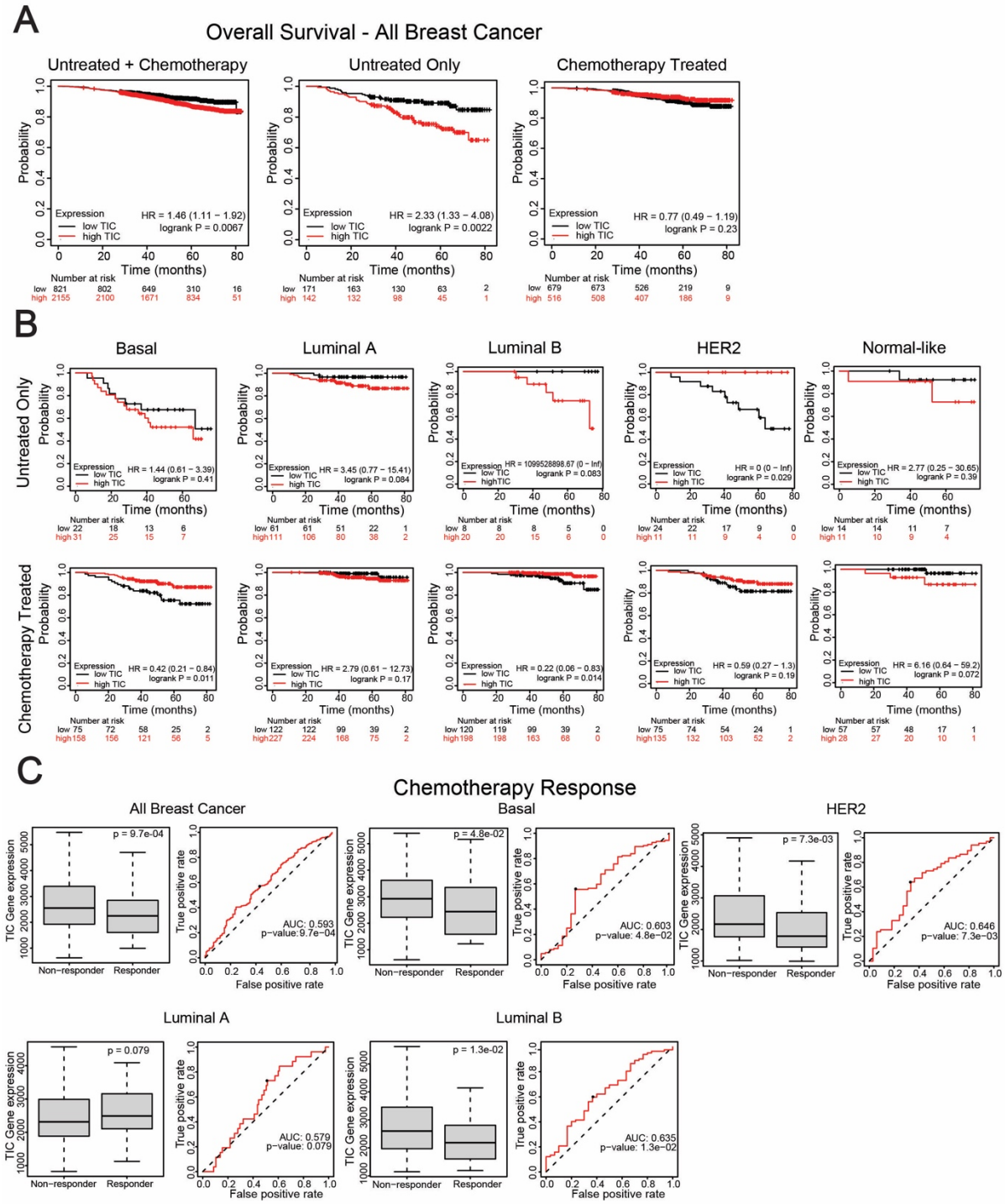
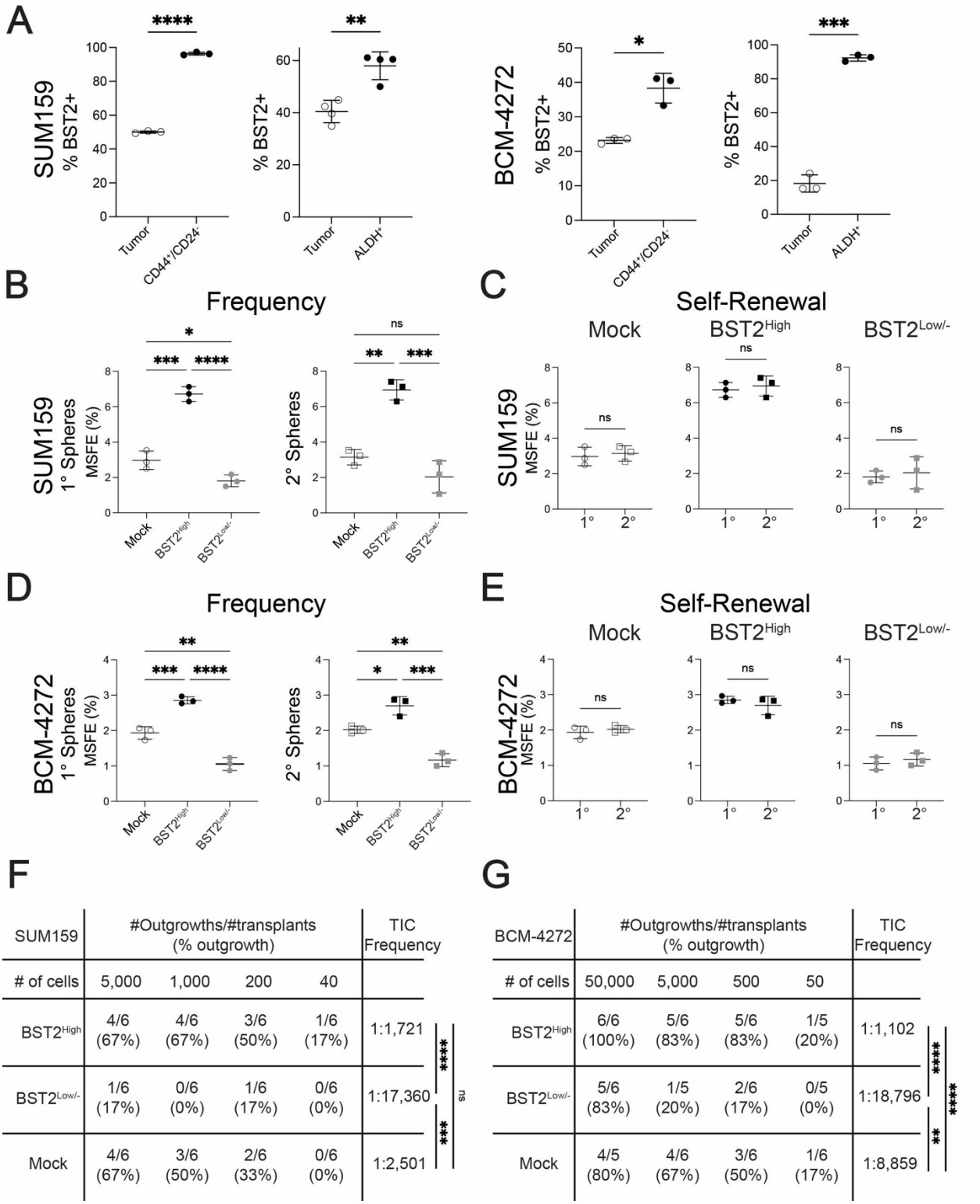


Figure 6: The IFN/STAT1 signaling axis is associated with worse overall survival and lack of response to chemotherapy in breast cancer patients. A) Kaplan-Meier plots showing overall survival between breast cancer patients with high or low expression of the 25-gene TIC signature including all patients (untreated + chemotherapy), only including untreated patients, and

1114 only including patients treated with chemotherapy. B) Kaplan-Meier plots showing overall survival
1115 between breast cancer patients with high or low expression of the 25-gene TIC signature
1116 comparing untreated patients and patients treated with chemotherapy by PAM50 subtype. C)
1117 ROC curve and box plots showing the association between the 25-gene TIC signature and
1118 chemotherapy response by breast cancer subtype.
1119

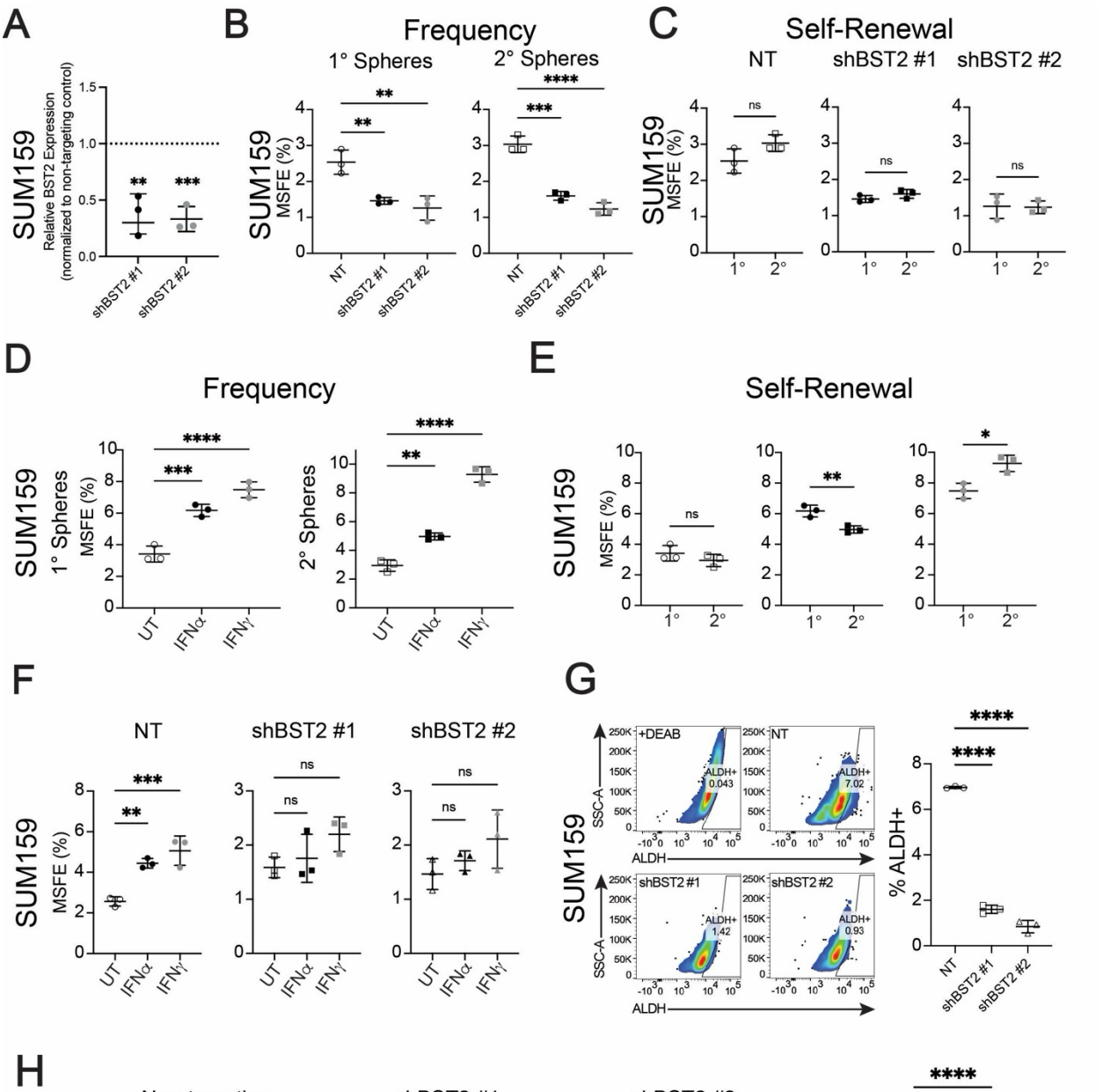


SUM159	#Outgrowths/#transplants (% outgrowth)				TIC Frequency
# of cells	5,000	1,000	200	40	
$BST2^{High}$	4/6 (67%)	4/6 (67%)	3/6 (50%)	1/6 (17%)	1:1,721
$BST2^{Low/-}$	1/6 (17%)	0/6 (0%)	1/6 (17%)	0/6 (0%)	1:17,360
Mock	4/6 (67%)	3/6 (50%)	2/6 (33%)	0/6 (0%)	1:2,501

BCM-4272	#Outgrowths/#transplants (% outgrowth)				TIC Frequency
# of cells	50,000	5,000	500	50	
$BST2^{High}$	6/6 (100%)	5/6 (83%)	5/6 (83%)	1/5 (20%)	1:1,102
$BST2^{Low/-}$	5/6 (83%)	1/5 (20%)	2/6 (17%)	0/5 (0%)	1:18,796
Mock	4/5 (80%)	4/6 (67%)	3/6 (50%)	1/6 (17%)	1:8,859

Figure 7: BST2 allows TIC enrichment in breast cancer xenograft models. A) Flow cytometry plot examining the correlation between BST2 and TIC markers $CD44^{+}/CD24^{-}$ and ALDEFLUOR

1123 activity in SUM159 and BCM-4272 xenografts. A two-tailed T-test was used to determine
1124 significance. B) SUM159 xenografts were dissociated, then the BST2 high, BST2 low, and mock-
1125 sorted populations were FACS enriched and grown as primary mammospheres (1°), then
1126 passaged and grown as secondary mammospheres (2°) (n = 3 biological replicates). One-way
1127 ANOVA and Tukey's post hoc test were used to determine statistical significance. C) Differences
1128 in MSFE between primary and secondary mammospheres for each flow-sorted fraction. A two-
1129 tailed T-test was used to determine significance. D) BCM-4272 xenografts were dissociated, then
1130 the BST2 high, BST2 low, and mock-sorted populations were FACS enriched and grown as
1131 primary mammospheres, then passaged and grown as secondary mammospheres. MSFE of
1132 each cell population for each xenograft (n = 3 biological replicates). One-way ANOVA and Tukey's
1133 post hoc test were used to determine statistical significance. E) Differences in MSFE between
1134 primary and secondary mammospheres for each flow-sorted fraction A two-tailed T-test was used
1135 to determine significance. F) Limiting dilution transplantation results of BST2 high, BST2 low, and
1136 mock-sorted populations from SUM159 xenografts. TIC frequency was calculated using extreme
1137 limiting dilution analysis and a Fisher's exact test was used to determine statistical significance
1138 (25). G) Limiting dilution transplantation results of BST2 high, BST2 low, and mock-sorted
1139 populations from BCM-4272 xenografts. TIC frequency was calculated using extreme limiting
1140 dilution analysis and a Fisher's exact test was used to determine statistical significance (25). *$p <$
1141 0.05, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; values are mean $\pm$ SD.
1142


A
SUM159
Relative BST2 Expression (normalized to non-targeting control)
shBST2 #1
shBST2 #2
B
Frequency
1° Spheres
2° Spheres
SUM159
MSFE (%)
NT
shBST2 #1
shBST2 #2
C
Self-Renewal
NT
shBST2 #1
shBST2 #2
ns
1°
2°
D
Frequency
SUM159
1° Spheres
2° Spheres
MSFE (%)
UT
IFNα
IFNγ
E
Self-Renewal
SUM159
MSFE (%)
F
NT
shBST2 #1
shBST2 #2
SUM159
MSFE (%)
G
+DEAB
NT
shBST2 #1
shBST2 #2
SUM159
SSC-A
ALDH
ALDH+ 0.043
ALDH+ 7.02
ALDH+ 1.42
ALDH+ 0.93
% ALDH+
H
Non-targeting
shBST2 #1
shBST2 #2
SUM159
Count
G1 64.9
S 12.9
G2 21.6
G1 75.4
S 10.2
G2 13.9
G1 83.8
S 5.88
G2 10.0
Propidium Iodide
% Proliferating (S + G2)

1145 **Figure 8: BST2 regulates TIC frequency in SUM159.** A) BST2 was knocked down in SUM159
1146 cells using two different shRNAs (shBST2 #1 and shBST2 #2), then BST2 knockdown was
1147 confirmed using Jess automated western blot. A two-tailed T-test was used to evaluate
1148 significance between each knockdown line and a non-targeting control (NT). B) SUM159 NT,
1149 shBST2 #1, and shBST2 #2 cells were grown as primary mammospheres (1°), then passaged
1150 and grown as secondary mammospheres (2°) (*n* = 3 biological replicates). One-way ANOVA and
1151 Tukey's post hoc test were used to determine statistical significance. C) Differences in MSFE
1152 between primary and secondary mammospheres for each cell line. A two-tailed T-test was used
1153 to determine significance. D) SUM159 cells were grown as primary mammospheres (1°) in the
1154 presence of IFNα (100 ng/μL), IFNγ (100 ng/μL), or left untreated as a control, then passaged and
1155 grown as secondary mammospheres (2°) in the presence of IFNα (100 ng/μL), IFNγ (100 ng/μL),
1156 or left untreated as a control (*n* = 3 biological replicates). One-way ANOVA and Tukey's post hoc
1157 test were used to determine statistical significance. E) Differences in MSFE between primary and
1158 secondary mammospheres for treatment. A two-tailed T-test was used to determine significance.
1159 F) SUM159 NT, shBST2 #1, and shBST2 #2 cells were grown as primary mammospheres (1°) in
1160 the presence of IFNα (100 ng/μL), IFNγ (100 ng/μL), or left untreated (UT) as a control (*n* = 3
1161 biological replicates). One-way ANOVA and Tukey's post hoc test were used to determine
1162 statistical significance. G) Representative flow cytometry plots illustrating ALDH activity in
1163 SUM159 NT, shBST2 #1, and shBST2 #2 cells and quantification of three biological replicates.
1164 H) Representative cell cycle flow cytometry plots using propidium iodide in SUM159 NT, shBST2
1165 #1, and shBST2 #2 cells and quantification of three biological replicates. *$p < 0.05$, **$p < 0.01$,
1166 ***$p < 0.001$, ****$p < 0.0001$; values are mean ± SD. 1°: primary mammospheres, 2°: secondary
1167 mammospheres, UT: untreated, IFN: interferon.
1168
1169

1170