Patient-derived triple negative breast cancer organoids provide robust model systems that recapitulate tumor intrinsic characteristics

Sonam Bhatia[1], Melissa Kramer[1#], Suzanne Russo[1#], Payal Naik[1#], Gayatri Arun[1$], Kyle Brophy[1], Peter Andrews[1], Cheng Fan[2], Charles M. Perou[2], Jonathan Preall[1], Taehoon Ha[1], Dennis Plenker[1$$], David A. Tuveson[1], Arvind Rishi[3], J Erby Wilkinson[4], W. Richard McCombie[1], Karen Kostroff[5] and David L. Spector[1*]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724
[2]University of North Carolina, Lineberger Comprehensive Cancer Center, Department of Genetics, Chapel Hill, N.C. 27599
[3]Northwell Health, Department of Pathology, Lake Success, NY 11042
[4]University of Michigan, Department of Pathology, Ann Arbor, MI 48109, USA
[5]Northwell Health, Department of Surgical Oncology, Lake Success, NY 11042

[#]Contributed equally
[$]Currently at: Envisagenics Inc.
[$$]Currently at: Loxo Oncology at Eli Lilly
*Corresponding author: spector@cshl.edu

Conflict of Interest

C.M.P is an equity stock holder and consultant of BioClassifier LLC; C.M.P is also listed as an inventor on patent applications for the Breast PAM50 Subtyping assay.

Abstract

Triple negative breast cancer (TNBC) is an aggressive form of breast cancer with poor patient outcomes, and an unmet clinical need for targeted therapies and better model systems. Here, we developed and comprehensively characterized a diverse biobank of normal and breast cancer patient-derived organoids (PDOs) with a focus on TNBCs. PDOs recapitulated patient tumor intrinsic properties and a subset of PDOs can be propagated for long-term culture (LT-TNBCs). Single cell profiling of PDOs identified cell types and gene candidates affiliated with different aspects of cancer progression. The LT-TNBC organoids exhibit signatures of aggressive MYC-driven basal-like breast cancers and are largely comprised of luminal progenitor (LP)-like cells. The TNBC LP-like cells are distinct from normal LPs and exhibit hyperactivation of NOTCH and MYC signaling. Overall, our study validates TNBC PDOs as robust models for understanding breast cancer biology and progression, paving the way for personalized medicine and tailored treatment options.

Statement of Significance

A comprehensive analysis of TNBC patient-derived organoids is presented by genomic, transcriptomic, and *in-vivo* analyses, providing insights into cellular heterogeneity and mechanisms of tumorigenesis at the single cell level.

Introduction

Breast cancer is characterized into several histopathological subtypes based on the expression of various receptors: estrogen (ER), progesterone (PR) and human epidermal growth factor receptor 2 (HER2/ERBB2). Molecular subtyping revealed multiple subgroups based on gene expression patterns that have been refined overtime and constitute luminal A, luminal B, HER2-enriched, basal-like, and normal-like breast cancers (1–3). Luminal A (ER+/PR+/Her2-) is the most prevalent and constitutes ~70% of all breast cancer cases, luminal B (ER+/PR+/HER2+) make up ~12% of cases, HER2-amplified (ER-/PR-/HER2+) represents ~5% of total cases and the basal or triple negative breast cancer (TNBC, ER-/PR-/HER2-) makes up ~12% of total breast cancer cases (4). The majority of luminal breast cancers exhibit low proliferation rates while the TNBCs are highly aggressive and proliferative in nature. Since TNBCs lack ER/PR and HER2 receptors, patients are exempt from any targeted endocrine or HER2 therapies resulting in non-specific cytotoxic chemotherapy as the standard of care. The majority, but not all, TNBCs fall under the basal-like subgroup of breast cancer by gene expression based PAM50 profiling and a small subset belong to the highly invasive claudin-low subgroup (1,4–6). Further molecular characterization of TNBCs revealed heterogeneity within this subset and at least 6 molecular subtypes have been identified that can be used to design more targeted therapies (7,8). TNBCs are typically presented as high-grade carcinomas (9) and show a significantly poor 5-year prognosis compared to other subtypes (10–12). Therefore, there is an unmet clinical need to better understand triple-negative disease progression and identify more precise druggable targets, and specific therapies for this cohort of breast cancer patients.

3-D organoid models have gained traction over recent years as the next generation of models for studying disease and development (reviewed in (13,14)). While cell lines, spheroids and mouse model-derived organoids have been the primary *in vitro* model systems for studying cancer biology; patient-derived organoid (PDO) models have now been developed for a variety of cancers including those originating in the colon (15), pancreas (16), ovary (17,18), prostate (19) and breast (20). Sachs et al. developed a methodology for growing human breast tumors *ex vivo* as organoid models that has now been expanded to the growth of organoids from normal mammary tissue as well (21,22). While cancer cell lines have been a valuable resource for providing insights into cancer biology and drug development (23–25), PDO models are not only patient-specific but also provide a three-dimensional context which is closer to that of the actual tumor microenvironment. PDOs, therefore, represent unique model systems to study disease progression and for the identification and validation of better treatment options. However, we do not yet fully understand the extent to which these models recapitulate the cellular heterogeneity and complexities of triple negative disease.

Here, we developed a diverse breast cancer PDO biobank and performed comprehensive genomic, transcriptomic and cellular characterization of organoids with an emphasis on TNBCs. Using genomic assays we show that our organoid models recapitulate pathogenic single nucleotide variants (SNVs) and copy number alterations (CNAs) of breast cancers as portrayed in large scale breast cancer genomic datasets (10,26–28) and reveal lesser studied cancer driver genes. Transcriptomically, our biobank recapitulates the various subtypes and signatures of breast cancers, with a subset of

organoids exhibiting signature profiles associated with poor patient outcomes. *In-vivo* transplants of these organoids highly recapitulated the patient-tumor morphology, providing strong evidence of retention of individual tumor intrinsic properties in long-term organoid cultures, even for models derived from rare BCs. In addition, we find that while normal PDOs retain the major cell types found within the mammary epithelium, the TNBC PDOs have lost this lineage specificity and are predominantly enriched for luminal progenitor (LP)-like cells. Single cell RNA-sequencing (scRNA-seq) of TNBC and normal PDOs identified differential signatures between the tumor and normal LP cells, providing insights into putative mechanisms of tumorigenesis. Lastly, we identified cells with various gene expression signatures in TNBC organoids that can be used to model and perturb various aspects of cancer biology, including tumorigenesis, hypoxia response, and EMT. Overall, our comprehensive characterization of TNBC organoids identified them as valid cancer models for studying cancer biology and for applications in precision medicine.

Materials and Methods:

Patient Material

Tumor resections from breast cancer patients along with the distal and adjacent normal tissue were obtained from Northwell Health in accordance with Institutional Review Board protocol IRB-03-012 and IRB 20-0150. Specific information for all samples is available in Table S1. The collection of genomic and phenotypic data was consistent with 45 CFR Part 46 (Protection of Human Subjects) and the NIH Genomic Data Sharing (GDS) Policy. Informed consent ensured that the de-identified materials collected, the models

70 created, and data generated from them can be shared without exceptions with
71 researchers in the scientific community.
72
73 **Patient-derived organoid culture:** Patient-derived organoids were established and
74 propagated using a previously published protocol (20). In summary, the tissues were
75 manually cut into smaller pieces and treated with 2mg/ml collagenase IV in base media
76 (ADF+++: Advanced DMEM-F12 (Invitrogen 12634-034) with 1x Glutamax (Invitrogen
77 12634-034), 10mM Hepes (Invitrogen 15630-056), 100U/ml Pen-Strep (Invitrogen 15140-
78 122)) at 37°C for 45-90mins with gentle agitation to break the tissue into small clusters of
79 cells. The suspension was intermittently resuspended by pipetting multiple times to
80 ensure proper digestion of the tissue. The cell suspension was centrifuged at 300g for
81 5mins and the pellet was treated with red blood cell lysis buffer (Cat # 11814389001,
82 Sigma) for 5mins at room temperature if it appeared bloody. The suspension was washed
83 2x with ADF+++ and plated in a matrigel (lot test for concentration of 8-10mg/ml, Cat #
84 356231, Corning) dome on pre-warmed tissue culture plates. The dome was incubated
85 at 37°C for 15mins and supplemented with completed medium: 10% R-Spondin1
86 conditioned medium, 5nM Neuregulin 1 (Peprotech 100-03), 5ng/ml FGF7 (Peprotech
87 100-19), 20ng/ml FGF10 (Peprotech 100-26), 5ng/ml EGF (Peprotech AF-100-15),
88 100ng/ml Noggin (Peprotech 120-10C), 500nM A83-01 (Tocris 2939), 5uM Y-27632
89 (Abmole Y-27632), 1.2uM SB202190 (Sigma S7067), 1x B27 supplement (Gibco 17504-
90 44), 1.25mM N-Acetylcysteine (Sigma A9165), 5mM Nicotinamide (Sigma N0636),
91 50ug/ml Primocin (Invitrogen ant-pm-1) in ADF+++. The organoids were passaged every
92 15-30 days using TrypLE^{TM} (Thermo Fischer 12605028) to break down the organoids into
93 smaller clusters of cells and re-plating them in matrigel domes as described above. For

94 tumor scrapings, the tumor surface was shaved on multiple sides to collect material which
95 was subsequently manually broken down, treated with red blood cell lysis buffer and
96 seeded in matrigel followed by regular PDO culture.

97 Organoid models labelled with the prefix HCM-CSHL were acquired as part of the
98 Human Cancer Model Initiative (HCMI) https://ocg.cancer.gov/programs/HCMI and a
99 subset of those models are or will be available for access from ATCC. The data for these
100 models can be accessed here: dbGaP accession number phs001486. Organoid
101 nomenclature: prefixes LNS, NH, DS, HCM-CSHL are de-identified patient IDs and are
102 not distinct in any features other than protocols used for sample acquisition; prefix NM
103 designates true normal samples collected from patients undergoing reductive
104 mammoplasty (NM: Normal Mammoplasty); suffixes: T=tumor, N=normal, ND=normal
105 distal, NAdj=normal adjacent, and TSc= tumor scraping. Organoid freeze thaws are
106 indicated in parenthesis: (passage frozen down)passage after thaw eg. NH85TSc (p4)p4.
107

108 **Organoid DNA and RNA extraction**: Organoid RNA was extracted using TRIzol®
109 (Thermo Fischer 15596018) RNA extraction protocol. DNA was extracted by removing
110 matrigel from organoids using ice cold PBS or TrypLE following by DNA extraction using
111 Qiagen DNeasy Blood and Tissue kit (Qiagen 69504) with elution in nuclease free water
112 (Thermo Fischer/Ambion 4387936).
113

114 **Targeted gene panel sequencing**: We performed capture based targeted gene panel
115 sequencing (65) for a panel of potential cancer driver genes. Briefly, we used a panel of
116 143 cancer genes with a total of ~4000 probes for capture. The captured DNA was paired-

117 end sequenced with 150bp reads and a coverage of about 300-500x. Library preparation
118 and sequencing of the targeted gene panel was performed by the CSHL Next Generation
119 Sequencing Core Facility. We developed an analysis pipeline to prioritize identification of
120 driver mutations. The sequencing reads are aligned to the hg19 reference genome using
121 BWA (66), followed by conversion to BAM format and sorting with Samtools (67), removal
122 of PCR duplicates with Picard (https://broadinstitute.github.io/picard/), and filtering with
123 Bamtools (68) for mapping quality and proper read pairing. Coverage of the target regions
124 is assessed for breadth and depth using Picard HSmetrics to ensure adequate coverage
125 for confident variant detection. Variants are then called using VarScan2 (69) in somatic
126 mode to stratify germline versus somatic variants, and are annotated with Annovar (70)
127 to cover a broad range of variant assessment tools. We then select rare loss of function
128 variants (nonsense, frameshift, splice site) with frequency less than 1% in the Gnomad,
129 ExAC, EVS and 1000 Genomes databases. Missense and in-frame indel variants are
130 selected if they are noted as pathogenic by ClinVar (71), or if they are both rare (<1% in
131 all genome databases) and annotated as pathogenic by COSMIC (72), or if they are both
132 rare and found to be present in the TCGA cohorts. Finally, missense variants are selected
133 if they are annotated as potentially deleterious by the ensemble tools REVEL (73) and
134 MCAP (74). Variants that were deleterious by REVEL and MCAP but did not have
135 population level data were discarded from the final oncoplot. Oncoplots are generated
136 from these candidate variants using Maftools (75).
137
138 **RNA-seq**: All RNA samples were quality controlled using a nanodrop followed by a
139 bioanalyzer (RNA nano-kit Cat # 5067-1511) and only samples with RIN >7 were used

for RNA sequencing. 750ng of RNA was used to prepared un-stranded RNA-seq libraries using Illumina TruSeq RNA Library prep kit v2 (RS-122-2001) and sequenced as 75bp paired-end reads.

RNA-seq analysis: The sequencing fastq files were quality checked using fastQC to make sure the reads were of consistent quality between different runs. The reads were aligned using STAR-aligner STAR-2.5.3a (76) using the following parameters:

```
--outFilterMismatchNmax 8 --alignIntronMax 1000000 --alignMatesGapMax
1000000 --outSAMstrandField intronMotif --outFilterIntronMotifs
RemoveNoncanonicalUnannotated
```

against the gencode v27 gtf reference file. Any PCR duplicates were marked in the aligned files using STAR with the following parameters:

```
--runMode inputAlignmentsFromBAM --bamRemoveDuplicatesType
UniqueIdenticalNotMulti --runThreadN $thread --inputBAMfile $input_bam
--outSAMtype BAM SortedByCoordinate
```

followed by removing duplicated reads using `samtools view -bF 0x400`. HTseq-count was used to count the reads per gene using the gencode v27 gtf file. The counts files were exported into R 4.1.0 and analyzed for differential expression using DeSeq2 1.32.0 (77). Concordance between technical replicates was ensured using PCA and sample distance matrix before summing them together for downstream analysis. Sample distance matrix was generated using euclidean distances between the samples and hierarchical clustering was performed using “ward.D2” linkage method followed by cutree with kmeans=6 with the R package “stats” (v4.1.0). For signature correlation, all the samples were used and an unsupervised clustering was performed for 838 previously curated gene expression signatures (37–39) and visualized using Java TreeView v1.2.0. For individual signature comparisons signature scores in experimental groups were

167 compared using Kruskal-Wallis test followed by pairwise comparisons using Wilcoxon
168 rank-sum test. Family-wise error rate was adjusted using Bonferroni-Holm method.
169
170 **SMASH copy number**: SMASH was performed as per the published protocol (32)
171 starting with 750ng of genomic DNA. SMASH was performed in batches of 10 samples
172 and sequenced on a MiSeq PE150bp run. The SMASH analysis is based on identification
173 of Maximal Unique Matches (MUMs) to the human genome in all read pairs (32). These
174 MUMs were filtered to remove matches <20 bp, matches with <4 bp of excess unique
175 sequence, and matches on read 2 that are within 1000 bases of the matches from read
176 1. Raw copy number profiles are then generated from the remaining 3–4 matches per
177 read pair which are then added to empirically sized bins spanning the genome. Regions
178 with identical copy are expected to yield similar bin counts using these empirical bins. We
179 next perform GC correction by normalizing counts based on LOWESS smoothing of count
180 vs. GC content data in each bin. Final copy number profiles are normalized so that the
181 autosome has an average copy number of 2. Plots were generated with G-Graph
182 MUMdex software (https://mumdex.com/) and IGV browser v2.9.2 (78).
183
184 **Organoid Histology:** Organoid domes in complete medium were scraped from the tissue
185 culture plate and collected in falcon tube precoated with BSA. The organoids were
186 collected and washed 1x with PBS by spinning at 300g for 5mins. Organoid harvesting
187 solution (Cat # 3700-100-01, Trevigen) was added to the organoids (3x the volume of
188 matrigel) and incubated at 4°C on ice for 30minutes to ensure that matrigel was removed
189 and the organoids were concentrated at the bottom. The organoids were washed 1x with

190 ample PBS and fixed with fresh 4% PFA at room temperature for 10minutes. 1:1 (v/v)
191 BSA was added to the tube and spun at 300g for 5mins to remove the PFA. The organoids
192 were washed 2X with ample amounts of PBS and embedded in 2% agarose in dH2O).
193 The agarose organoid molds were then paraffin embedded and cut into 5um sections.
194
195 **Organoid Hematoxylin & Eosin (H&E) staining and Immunohistochemistry (IHC)**:
196 H&E and IHC staining were performed at the CSHL Histology Core Facility. PFA fixed
197 organoids in agarose were processed in Thermo Excelsior ES processor and embedded
198 with Thermo HistoStar embedding system following manufacturer's protocols. Paraffin
199 blocks were cut into 5um sections and mounted onto positively charged slides (VWR
200 superfrost plus micro slide).
201 For H&E staining, slides were stained in a Leica Multistainer (ST5020). Briefly,
202 slides were deparaffinized and rehydrated and then stained in hematoxylin (Hematoxylin
203 560 MX, Leica) for 1 min, followed by destaining in Define MX-aq (Leica) for 30 sec, bluing
204 in Blue Buffer 8 (Leica) for 1min and subsequently stained in eosin (EOSIN 515 LT, Leica)
205 for 30sec. After dehydration, coverslips were placed onto glass slides using a robotic
206 coversliper (Leica CV5030).
207 IHC slides were stained in Discovery Ultra automatic IHC stainer (Roche) following
208 standard protocols. Briefly, slides were subjected to antigen retrieval (Benchmark Ultra
209 CC1, Roche) at 96˚C for 1hr; primary antibodies were incubated at 37˚C for 1hr and
210 Discovery multimer detection system (Discovery OmniMap HRP, Discovery DAB and
211 Purple, Roche) was used to detect and amplify immuno-signals. Antibodies used: Ki67
212 (Spring Bioscience, #M3062, 1:500).

213

214 **Organoid formation assay**: Organoids were processed using TrypLE and 1500 single
215 cells per well of a 96 well plate were seeded in 10% matrigel + complete growth medium.
216 Cell viability was assessed using Cell Titre Glo 3D luminescence assay (Promega G9683
217 CellTiter-Glo 3D Cell Viability Assay). Baseline cells were measured using Cell Titre Glo
218 3D assay at 24 hrs (d1) post seeding and growth was measured at 6 days after seeding
219 (d6). Each organoid line was evaluated for multiple passages n=2 or n=3 per PDO.

220 For MYCi and DAPT experiments: Organoids were processed to single cells using
221 TrypLE and seeded in 50ul matrigel domes as 10,000 cells per well of a 24 well-plate.
222 Complete medium or medium supplemented with DMSO, DAPT (Selleckchem S2215
223 DAPT) and MYC-inhibitor (Selleckchem S8906 MYCi975) were added to the respective
224 wells. For normal PDOs, organoids were dissociated with trypLE and live cells were
225 sorted using fluorescence activated cell sorting (FACS) based on staining of CD46f,
226 EPCAM and 7AAD. Organoids were allowed to form for 12 days, images were acquired
227 using microscope and organoids were manually counted for each condition. Statistical
228 analysis was performed using one-way ANOVA test followed by pairwise comparisons
229 using two-sample t-test. Family-wise error rate was adjusted using Bonferroni-Hold
230 method.

231

232 **Organoid proliferation index analysis**: Paraffin embedded organoid sections were IHC
233 stained for Ki67. Slides were scanned and viewed using Aperio ImageScope 12.3.3.
234 Images were analyzed in FIJI (79). Briefly, the images were deconvolved using Colour
235 Deconvolution for hematoxylin and DAB, converted to 8-bit binary images and analyzed

236 using the BioVoxxel Toolbox plugin (https://www.biovoxxel.de/#/) to evalute %Ki67
237 positive cells per organoid. Multiple passages for each organoid line were evaluated, n=2
238 or n=3 per PDO.
239 **Drug dose response assays**: Organoids were digested into single cells and seeded as
240 1500 cells/well of a 384 wp (USA-Scientific Cat # 5678-1976) as suspension cultures in
241 complete medium with 10% matrigel using a liquid handler. Organoids were incubated at
242 37C for 24hrs and drugs were added using a drug dispenser Beckman Echo 650. The
243 organoids were incubated for 5 days at 37C and assayed using the CellTiter-Glo® 3D
244 Cell Viability Assay (Promega Cat # G9681) and read for luminescence using the
245 EnVision 2105 plate reader. The data was analyzed using GraphPad Prism v9.0.0. The
246 screens were performed in 3 technical replicates with 2-4 experimental replicates per
247 organoid line. The raw reads were first normalized against untreated controls and a non-
248 linear model was fit for the mean, after removing any detected outliers, using the
249 log(inhibitor) vs normalized response with a variable slope.
250 **Animals**: Six-week-old female NOD scid mice (NOD.Cg-*Prkdc*^{scid}/J) were obtained from
251 the Jackson laboratory (JAX stock #001303) and acclimated at the Cold Spring Harbor
252 Laboratory Animal Shared Resource for a minimum of 1week. All animal experiments
253 were performed in accordance with the Institutional Animal Care and Use Committee.
254
255 ***In-vivo* transplant experiments*:*** TNBC organoids were harvested using the organoid
256 harvesting solution (Trevigen Cat# 3700-100-01) and manually counted. Organoids were
257 resuspended in a 1:1 mixture of PBS:matrigel and 50K organoids were injected into the
258 bilateral mammary fat pads by the fourth nipple of female NOD scid mice (NOD.Cg-

259 *Prkdc*^{scid}/J). Mice were anesthetized with 1.5-2% isoflurane and weighed before the
260 injections. The animals were regularly monitored for their weight, tumor size and any other
261 discrepancies. Mice were sacrificed when the tumors reached any of the following end-
262 points: 2cm tumors, ulceration, visible necrosis, blistering of tumors or deteriorating health
263 of the mice. At end-point, dissections were performed and the tumors along with lungs,
264 liver, lymph nodes and the femur were fixed in 4% PFA overnight at 4°C. If the tumors
265 were not observed the mammary fat pads were collected instead. The transplant
266 experiments were done with 2 independent passages of PDOs, with 4-6 injections per
267 PDO per passage.

268 Fixed tumors and tissues were processed for histology as above. Metastases and
269 micro-metastases were assessed using IHC with a human mitochondria antibody
270 (Millipore MAB1273 Anti-Mitochondria clone 113-1).

271

272 **Flow Cytometry**: Organoids were scraped in the culture medium and washed 1X with
273 PBS. TrypLE^{TM} was used to fully digest the organoids into single cells. The cells were
274 counted, diluted to 200,000 cells/100ul and stained in 100ul of ADF+++ using anti-Epcam
275 (1:50), anti-CD49f (1:50), 7AAD (1:50). The following antibodies were used: PE Mouse
276 IgG2a, κ Isotype Control (BD 555574), APC Mouse IgG2b κ Isotype Control RUO (BD
277 555745), Alexa Fluor® 647 anti-human CD326 (EpCAM) Antibody (Biolegend 324212),
278 PE anti-human/mouse CD49f Antibody (Biolegend 313612). The cells were read using a
279 BD Fortessa and analyzed using the FACS DIVA and FlowJo v10 software.

280

281 **Single cell RNA-seq**: Organoids were digested into single cells using TrypLE™,
282 resuspended in 0.04% BSA in PBS as 1 million cells/ml. ~12,000 cells were loaded into
283 one well of a 10x Chromium microfluidics chip. Single cell barcoding and libraries were
284 prepared using the 10x Chromium v3 chemistry (Cat # V3 reagents 1000075, V3 chips -
285 1000153 or NextGEM reagents- 1000121, chips 1000120). Libraries were quality
286 checked using a Bioanalyzer HS kit for cDNA yield and final library size and qubit to
287 quantify.

288 Single cell analysis was performed in three different batches (Table S7). Batch 3
289 was a multiplexed pool of 4 samples, which were demultipexed using a custom genotype-
290 aware pipeline. At the time of 10X Genomics library preparation, ~20,000 cells from each
291 of the four organoids were set aside to prepare low-input bulk RNA-seq libraries tagged
292 with unique i7 barcodes. These bulk libraries share the same adapter structure as 10X
293 Genomics libraries, and were spiked into the Illumina NextSeq500 flow cell at a 5% molar
294 ratio to obtain ~5M reads per organoid. These barcoded bulk libraries were then used to
295 create reference VCF files using cellSNP v0.3.2 by searching a list of 7.4M common
296 human SNPs from the 1000 Genomes Project
297 (http://ufpr.dl.sourceforge.net/project/cellsnp/SNPlist/genome1K.phase3.SNP_AF5e2.ch
298 r1toX.hg38.vcf.gz). Genotype profiles were filtered to include only positions with < 10%
299 minor allele frequency and >20 UMI counts. In parallel, per-cell VCF files were generated
300 from the multiplexed single cell library using the cellSNP with the same parameters. Cells
301 from the single-cell pool were assigned to their respective donors using Vireo v0.4.2 (80).
302 *Genotyping low-input bulk RNAseq library prep:*

303 RT Primer Design:

304 CTACACGACGCTCTTCCGATCTSSSSSSSSNNNNNNNNNNNNVVVVVTTTTTTTTT

305 TTTTTTTTTTTTTTTTTTTTTTTVN

306 where:

307 SSSSSSSS = 8bp sample barcode, with following multi-plexing key:

308 GACAGTGC=HCM-CSHL-0366-C50

309 GAGTTAGC=NH85TSc

310 GATGAATC=NH95T

311 GCCAAGAC=NH93T

312

313 NNNNNNNNNNNVVVVV = 15bp UMI with 5 non-T residues at 3' end

314 Template Switch Oligo: AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG

315 cDNA_amplification_Forward:

316 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG

317 cDNA_amplification_Reverse: AAGCAGTGGTATCAACGCAGAGT

318

319 RT was performed using the SuperScript IV (Life Technologies #18091050) according to

320 the manufacturer’s instructions except for the addition of 1uM Template Switch Oligo

321 during first strand synthesis. All custom oligos were synthesized by IDT. After cDNA

322 amplification, molar concentration was estimated using the Agilent Bioanalyzer 2100,

323 libraries were pooled at an equimolar ratio, and prepared for Illumina sequencing using

324 the Nextera XT DNA Library Prep Kit (Illumina) according to the manufacturer's

325 instructions. Final fragmented libraries were again checked and quantified by Bioanalyzer

326 prior to mixing at a 1:20 molar ratio with 10X Genomics libraries for sequencing.
327 *Sequencing and Mapping:* The libraries were sequenced on Illumina NextSeq 500 High
328 Output 75 cycle kits using the read format: 8bp (i7 index) x 28bp (Read 1) x 56bp
329 (Read2). 10X Genomics libraries were mapped using Cell Ranger version 4.0.0 (10X
330 Genomics) with default settings and a custom genome reference based on the
331 comprehensive gene annotation set from Gencode Release 32 (GRCh38.p13)
332 (http://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_32/gencode.v32.
333 annotation.gtf.gz). For the multiplexed pool, sample identities for each cell were assigned
334 using Vireo as described above, such that each sample could be subset from the pooled
335 matrix during analysis.
336 Filtering, feature selection, clustering, and other secondary analyses were carried
337 out in R using Seurat v4.0.3 (https://satijalab.org/seurat/) (47,81). Gene set enrichment
338 analyses were performed using GSEA v4.1.0 (82). Following gene sets were used for the
339 various signature scores: adult human breast epithelium markers from (49), Mammary
340 epithelial lineage scores from (40), NOTCH signaling:
341 REACTOME_SIGNALING_BY_NOTCH, BMP2 targets: LEE_BMP2_TARGETS_UP,
342 MYC signature from (83), Hypoxia signature: HALLMARK_HYPOXIA, Basal mammary
343 stem cell signature from (40).
344
345 **Quantification and statistical analysis**: Statistical analyses were performed using R
346 (version 4.1.0) on RStudio and GraphPad Prism software (v9.1.2, GraphPad Software,
347 San Diego, California USA, www.graphpad.com). Specific tests are indicated in the figure
348 legends along with the statistical significance.

349

350 **Materials, data and code availability**: Organoid lines generated under the HCMI project
351 (starting with HCM) will be available for purchase from the American Type Culture
352 Collection (ATCC). All raw data will be available to download from dbGaP
353 (phs002722.v1.p1 and phs001486). Copy number segment file is available in Table S2,
354 transcriptome signature scores are in Table S4. Metadata for the different PDOs is
355 available in Table S1. All original code will be deposited at https://github.com/bhatia-
356 sonam/manuscript-v1

357

358 **Results:**

359 **Establishment and somatic variant profiling of a diverse patient-derived breast**
360 **cancer organoid biobank**

361 Breast cancer tissues along with paired normal breast tissues were acquired from
362 female patients and developed into 3-D organoids. In addition to breast cancer tissues,
363 we also collected normal reductive mammoplasty samples from 10 cancer-free patients
364 to generate normal PDOs for downstream analyses. The majority of tumor samples were
365 from patients with invasive ductal carcinomas (56/87, Fig. 1A), 11/87 were from invasive
366 lobular carcinomas, 4/87 were from metastatic lymph nodes and a small percentage from
367 other categories (Fig 1A, Table S1). The majority of the tumor samples (43/87) were
368 luminal BC as defined by immunohistochemistry (ER/PR+/HER2-), 37/87 were from
369 TNBCs (ER-PR-HER2-), 2/87 were from ER/PR+/HER2+ BCs, 2/87 were from ER-/PR-
370 /HER2+ subtype, and 3/87 samples belonged to post-treatment residual tissue with no
371 visible carcinoma (NA) (Fig 1B). Samples were collected from patients of various age

372 groups (Fig 1C) and diverse ethnic and racial backgrounds (Fig 1D). An increased
373 proportion of TNBC tumor samples was observed from black patients with African or west-
374 Indian heritage (Fig 1D, Table S1) which recapitulates the higher incidence of TNBCs in
375 the African-American community (29,30). Using a DNA-seq panel of 143 cancer driver
376 genes, we identified pathogenic single nucleotide variants (SNVs) in 49 tumor organoid
377 lines (Fig 1E); *PIK3CA* was mutated in 33% of the tumor organoid lines, the majority of
378 which were from patients with luminal breast cancer. 24% of the organoid lines (45% of
379 TNBC PDOs 10/22) had a pathogenic *TP53* mutation and primarily represented the high
380 grade TNBC cohort (Fig 1E). Two of the organoid lines are derived from a patient with
381 rare breast cancer showing adenoid cystic carcinoma (AdCC) like morphology, NH87T
382 (primary tumor) and HCM-CSHL-0655-C50 (lymph node metastasis from the same
383 patient) (Table S1). The AdCC organoids show a non-traditional TNBC mutation profile
384 with pathogenic mutations in *APC*, *KDM6A* and *NOTCH1* (Fig 1E), which were previously
385 observed in AdCCs of the breast (31). Mutations in *KMT2C* and *GATA3* were also
386 observed in a variety of luminal breast cancer organoids, *CDH1* mutations were present
387 in 10% of PDOs all of which belonged to invasive lobular carcinomas, and *ARID1B*
388 mutations were found in a few TNBC-derived organoids (Fig 1E). Overall, the mutation
389 profiles of patient-derived BC organoids are in concordance with the mutational
390 landscape of BCs (10,26). A relatively large subset of tissues (24/87) that resulted in
391 cultured organoids did not have pathogenic SNVs (Fig 1F, Table S1), while 10/87
392 samples dropped off in culture at early passages (Fig 1F). The PDOs with pathogenic
393 SNVs showed a range of growth properties *in vitro* with subsets exhibiting long-term

394 continued expansion while others seem to have more limited cultures either in terms of
395 maximum passages reached or limited expansion abilities (Table S1).
396 A comparison of the SNV profile between several pairs of patient tumor and PDOs
397 showed high concordance of the pathogenic SNVs between the tumor and tumor derived
398 organoids (Fig S1A), except in some cases where the SNVs are present at a much lower
399 variant allele frequency (VAF) in the tumor, for instance, NH84TT-P0 showed <25% TP53
400 VAF compared to 100% VAF in NH84TT-p10 PDO (Fig S1A) suggesting normal
401 contamination in the tumor tissues and a successful outgrowth of cancer cells in organoid
402 cultures. Longitudinal analyses for some patient-derived organoids showed the stability
403 of primary driver mutations overtime in culture (Fig S1B). To determine whether the
404 amount of tumor material was a limiting step for generation of organoids, we collected
405 tumor scrapings (see methods) from a subset of tumor samples. A comparison of tumor
406 tissue (labeled TT) and the scrapings (labeled Sc) from three different patients (Fig S1C)
407 showed high concordance and indicates that, if necessary, small amounts of tumor
408 material can be used to generate organoids. We also examined the pathogenic SNVs in
409 TNBC tissue samples (p0) that did not result in a successful generation of organoids
410 (establishment or long-term cultures) and 6/10 had TP53 mutations and some samples
411 also had BRCA1/2 mutations (Fig S1D), the latter were largely absent in successful
412 organoid cultures (Fig 1E).
413
414 **Breast cancer organoids are enriched for copy number alterations**
415 Having identified pathogenic driver mutations for the tumor-derived organoids, we
416 performed copy number analysis using SMASH (short multiply aggregated sequence

417 homologies) (32) for various PDOs. TP53 mutated TNBC PDOs make up the majority of
418 the ER- subset and are highly genomically altered compared to the luminal ER+
419 organoids (Fig 2A, B, Table S2). Some organoids that were deemed to be derived from
420 tumor based on pathogenic SNVs (eg. LNS12T, LNS18T, NH06T etc.) showed no
421 prominent CNAs. Some TNBC PDOs that showed trace SNVs such as NH58T, NH72T
422 and NH66T also showed minimal CNAs (Fig 1E and 2A). In keeping with large scale
423 genomic datasets (10,26), a consistent gain of chromosome 1q (chr1q) and loss of chr16q
424 was observed in the luminal organoids where CNAs were detected (Fig 2A). The TP53
425 mutated TNBC organoids were highly genomically aberrant where multiple lines showed
426 a gain of chromosomes (chrs) 1q, 8q, 19q and chrs 7, 20 and 21; loss of chrs 3p, 4, 5,
427 17p and Xp was also observed in multiple organoid lines (Fig 2A). These TNBC PDOs
428 also showed copy number loss of chr4q and chr5q (Fig 2A) that have been previously
429 reported to be over-represented in basal-like breast cancers (33). Of note, the AdCC-like
430 organoids, NH87T (primary tumor) and HCM-CSHL-0655-C50 (lymph node metastasis
431 from the same patient), were considerably less genomically aberrant as compared to the
432 other TNBC organoids and showed focal amplification of chr8q and deletion of chr6p (Fig
433 2A, C-D). Interestingly, the chrX deletion was only observed in the primary tumor (NH87T)
434 sample (Fig 2A, C) and a focal deletion was observed in chr2 of the lymph met HCM-
435 CSHL-0655-C50 (Fig 2C). The AdCC organoid CNA profiles are consistent with the low
436 CNA profiles that are typically observed in AdCC-like breast cancers (31).

437 For a subset of these samples, we compared the copy number profiles of the
438 primary tumor with the paired tumor and normal organoids and observed more
439 pronounced CNAs in the organoids compared to the tumors suggesting that the organoid

440 culture enriches for tumor cells (Fig 2C-D). As observed in the SNV data (Fig S1) the
441 NH84T tumor (p0 with <25% VAF of pathogenic TP53) was likely very heterogenous and
442 contained a significant population of normal cells, however, the tumor cells successfully
443 outgrew the normal and resulted in a highly pure tumor organoid culture over multiple
444 passages (Fig 2C-D). For HCM-CSHL-0366-C50 and NH87T the early passage p1 or
445 tumor tissue (p0) respectively, had relatively less normal contamination and maintained
446 their copy number profiles over time in culture (Fig 2C-D). While copy number profiles
447 were enriched for in successful tumor PDOs, for some cultures we observed loss of copy
448 number alterations in the organoids over-time, suggesting a normal outgrowth (Fig S1E).
449 Since tumor organoids are more enriched for tumor cells compared to the p0 tumor
450 tissue, we profiled the putative cancer driver gene panel in individual organoid samples.
451 TNBC PDOs typically had a higher copy number alteration frequency of these cancer
452 driver genes (top bar-graph in Fig 2B). *TGFB2*, *MDM3*, *AKT3*, *DDR2* and *INSRR* showed
453 the highest frequency of alteration (Fig 2B) and are all genes present on chr1q that is
454 amplified in both luminal and TNBC PDOs (Fig 2A-B). Interestingly, we also found a
455 higher frequency of deletion of *KDM6A*, *ARAF* and *RPS6KA3* tumor suppressors, located
456 on chrXp, specifically in the TNBC-derived organoids (arrows in Fig 2B). Loss of
457 *PPP2R2A* (chr8p), which was previously reported to be a tumor suppressor in breast
458 cancer (10), was also identified in 38% of the samples, the majority of which are of TNBC
459 subtype (Fig 2B). Furthermore, there is an over-representation of loss of chr3p in the
460 TNBC samples, that results in the deletion of potential tumor suppressors: *ACVR2B*,
461 *BAP1*, *CTNNB1*, *MLH1*, *MYF88* and *PBRM1* (Fig 2B). Copy number loss of chr6q is also
462 common in TNBC organoids and results in loss of *ARID1B*, *MAP3K4*, and *PARK2*, a

463 master regulator of G1/S cyclins (34). *ESR1*, the gene encoding estrogen receptor (ER),
464 present on chr6q is also frequently lost in these TNBC organoids (Fig 2A-B). We also
465 identified gains of *CALR*, *JAK3*, *KEAP1*, *PIN1* and *SMARCA4* that are associated with
466 the amplification of chr19p and amplification of mismatch repair (MMR) genes *MSH2* and
467 *MSH6* located on chr2p (Fig 2A-B) in subsets of TNBC PDOs. While MMR genes are
468 commonly mutated in various cancers, their overexpression was recently associated with
469 aggressive prostate cancers (35) and was shown to promote genomic instability in yeast
470 (36) but their role is yet to be determined in TNBC disease progression.

471 Taken together, the SNV and copy number profiling of PDOs shows robust
472 retention of genomic features of various types of breast cancers, including luminal, TNBC,
473 and rare AdCC-like carcinomas. While there might be small alterations overtime in PDO
474 cultures, the key pathogenic mutations and overall copy number profiles are conserved
475 throughout organoid culture (Fig S1, 2C-D) supporting their utilization as valid cancer
476 models. Additionally, we find overrepresentation of some lesser studied copy number
477 variants in our data such as loss of tumor suppressors *RPS6KA3*, *PPP2R2A* and *PARK2*
478 and copy number gains of MMR genes *MSH2* and *MSH6* that might have important
479 unexplored consequences in breast cancer progression.

480

481 **A subset of TNBC organoids recapitulate signatures of aggressive basal-like breast**
482 **cancers**

483 Next, we performed RNA-seq on various tumor and normal-derived organoids to
484 profile their transcriptomes (Table S3). Hierarchical clustering of samples by euclidean
485 distance divided them into six groups (Fig S2A). We used supervised clustering with 838

486 previously curated gene expression signatures (37–39) (Table S4) and found similar
487 patterns of clustering among the various organoid lines (Fig 3A). Group 4 is largely
488 comprised of the true normal organoids (labeled with prefix NM in Fig S2A) derived from
489 breast tissues obtained from individuals undergoing reductive mammoplasty (Normal in
490 Fig 3A-B, S2A, C), paired normal PDOs derived from normal tissue adjacent or distal to
491 the tumor and very few tumor-derived organoids; this cluster showed an enrichment of
492 normal mammary stem cell (MaSC) signatures and low proliferation related expression
493 profiles (Fig 3A-B). Group 1 is comprised of a mixture of either paired normal organoids
494 or tumor organoids that did not have a strong driver mutation (eg. *PIK3CA* or *TP53*) (Fig
495 3A, S2A, 1E). This group had a signature profile that was distinct from the true normal
496 group (Group 4) and is most similar to luminal cell signatures (Fig 3A). The samples in
497 this group also have a higher proliferation signature compared to the true normal Group
498 4 (Fig 3A). Group 5 mostly contained PDOs derived from Luminal BC and showed luminal
499 like gene-expression signatures (Fig 3A-B). Signature profiling also uncovered immune
500 and stromal signatures that highly corresponded to samples in Group 3 (Fig 3A, S2C).
501 Interestingly, two of the samples in this group NH63T and NH54T exhibited limited
502 propagation in culture due to stromal outgrowth leading us to hypothesize that these
503 samples had some fibroblast-like cells. We also found a strong interferon (IFN) signature
504 in this group along with some additional PDOs from Group 6 (Fig 3A, S2C).

505 Groups 2 and 6 are comprised mostly of TNBC organoids and all organoid lines in
506 these groups could be propagated to long-term cultures (>passage10, labeled LT-
507 TNBCs) and showed continued expansion (Fig 3A, S2A). The only exception is the
508 NH48N normal sample, which was confirmed to be mostly tumor by copy number analysis

509 and identical to its counterpart NH48T (Fig S2B). The most prominent signatures of
510 Groups 2 and 6 correspond to basal-like breast cancer gene-sets that are also defined
511 by luminal progenitor (LP) like signatures (Fig 3A-B, S2C-D). LP-like gene expression has
512 previously been shown to be associated with basal-like breast cancers (40). The
513 organoids in this group also showed enrichment of proliferation signatures (Fig 3A-B) and
514 the majority had a *TP53* mutation which was in conjunction with the *PIK3CA* mutations
515 for three samples (HCM-CSHL-0773-C50, HUB4T and HCM-CSHL-0155-C50) (Fig 3A,
516 S2A). We also observed a MYC amplification signature in this subgroup (Fig 3B) that was
517 accompanied by the copy number amplification of *cMYC* in many of these PDOs (Fig
518 S2E). Additionally, gene sets associated with hypoxia, glycolysis, angiogenesis and fetal
519 MaSC (fMaSC) metabolism signatures were also enriched in LT-TNBCs (Fig 3A-B, S2C).
520 The *VEGF* 13-gene signature showed high correlation with basal-like breast cancers (29),
521 has prognostic significance and is associated with poor outcome in breast and other
522 cancers (41). Similarly, fMaSC metabolism signature is a refined 8-gene signature, which
523 primarily comprises genes associated with glycolysis and fatty acid metabolism, and was
524 previously shown to be associated with TNBCs and metastatic TNBC lesions (42). This
525 suggests that the PDOs with these signatures are likely associated with tumors linked to
526 poor outcomes.

527 In order to understand the complexity of organoid cultures we profiled the growth
528 properties of various TNBC and normal organoids. TNBC organoids appear as densely
529 filled balls of cells which is in stark contrast to the normal organoids that have a
530 predominant acinar structure with a central lumen and some organoids that were filled
531 (NM04N in Fig 3C). The TNBC-organoids showed varying degrees of propagation in

532 culture where some lines could be propagated for over 10 passages (LT-TNBCs) with
533 continued expansion while many short-term culture lines dropped out of culture at various
534 points (Fig 3D, S3A) and had limited material to assay. Low starting material, limited
535 proliferation, normal and stromal outgrowths were among the primary reasons for short
536 term culture growth (Fig S3A). We focused on the LT-TNBCs for all downstream
537 analyses. IHC for the proliferation marker, Ki-67, in confluent PDO cultures showed large
538 variability of expression amongst the various organoid lines and also within each culture
539 (Fig 3C- lower panels). In concordance with expression signatures (Fig 3B), LT-TNBC
540 organoids were highly proliferative compared to normal and luminal organoids (Fig 3C,
541 S3B-C). There was a lot of heterogeneity within organoid cultures with some organoids
542 being mostly comprised of proliferating cells, for instance NH85TSc and NH95T, while
543 other cultures were more mixed (Fig S3C). Normal-derived organoids typically had <10%
544 of proliferating cells per organoid and showed luminal-basal organization with a hollow
545 lumen, while TNBC-derived organoids tend to be highly proliferative and undifferentiated
546 with little to no observed cellular organization (Fig S3C). When seeded as single cells, we
547 observed a range of organoid formation amongst the various lines (Fig S3D) which did
548 not always correlate with the proliferative index of the PDOs, and might be due to the
549 requirement of cell-cell contact in some but not all PDOs that is apparent when seeding
550 at low density as single cells.

551 Of note, none of the TNBCs propagated in culture showed expression profiles of
552 the more mesenchymal-like claudin-low subgroup (6), possibly because the culture
553 conditions are more favorable towards the propagation of epithelial cells. Overall, the
554 gene expression signatures of the normal and tumor-derived organoids depict the

555 expression profiles of luminal and basal-like breast cancers. The TNBC organoids that
556 can be propagated to long-term cultures (LT-TNBCs), Groups 2 and 6, had very classic
557 basal-like breast cancer signatures and were associated with proliferation (Fig 3A-D),
558 hypoxia (Fig 3A-B) and *c-MYC* amplification gene expression signatures (Fig 3B, S2E).
559 Furthermore, about 40% of basal-breast cancers exhibit *c-MYC* amplification (26) and the
560 majority of our LT-TNBC organoids show an enrichment of *c-MYC* (Fig S2E) signature
561 suggesting that our organoid system results in long-term expansion of TNBCs that are *c-*
562 *MYC* driven and have a LP-like basal breast cancer signature.

563 As a proof-of-concept, we tested whether these PDOs were amenable to drug
564 response assays, and performed dose response curves using 6 different PDO lines with
565 three different drugs (Fig S3E). TNBC PDOs (Fig S3D) showed high sensitivity to the
566 cytotoxic chemotherapeutic agent paclitaxel compared to the luminal line NH53T and the
567 TNBC-met HCM-CSHL-0773-C50 (Fig S3E). Conversely, capivasertib (AZD5363), a pan-
568 AKT inhibitor used in PIK3C mutated cancers (43), showed greater sensitivity against the
569 PIK3CA mutated luminal NH53T PDO (Fig S3E). Similarly, afatinib a targeted inhibitor for
570 EGFR, showed the most activity in the EGFR amplified NH84T PDO line (Fig S3E).

571

572 **TNBC organoids can recapitulate tumor morphology *in-vivo***

573 Next, we examined the *in-vivo* tumor forming abilities of these patient-derived
574 cultures. We transplanted organoids from eight different TNBC (LT-TNBCs) lines into the
575 mammary-fat pads of NOD-SCID mice and assayed their tumor formation and metastatic
576 potential over time (Fig 4A). We found striking differences in the tumor forming ability
577 between the different organoid lines (Fig 4B) despite having highly aberrant genomic

578 profiles (Fig 2, 4C). TNBC organoids HCM-CSHL-0366-C50, NH85TSc and NH87T
579 resulted in palpable masses around 30 days and steady tumor formation in 100% of
580 injection sites (Fig 4B, Table S6). NH85TSc-derived tumors grew rapidly and resulted in
581 visibly necrotic masses that needed to be resected at ~100 days. HCM-CSHL-0366-C50
582 resulted in primary tumors that also developed fluid filled cysts on top of the tumors that
583 drained on their own. NH87T and HCM-CSHL-0655-C50 are the AdCC-like primary tumor
584 and lymph-node metastasis from the same patient. Interestingly, while NH87T formed
585 tumors rapidly, HCM-CSHL-0655-C50 did not and the tumors remained small (Fig 4B,
586 Table S6), despite both of these lines having equivalent growth properties *in vitro* (Fig 3,
587 S3). This observation was consistent for the two independent transplant experiments
588 done with different passage organoids (Table S6). NH95T resulted in small tumor masses
589 at multiple injection sites that were slow to grow while NH84T and NH93T had small
590 tumors observable only in fat-pad histology sections of some sites and NH64T did not
591 result in any primary tumors (Figs 4B, S4A, Table S6) despite being highly genomically
592 aberrant (Figs 4C, 1-2) and having high proliferation and organoid formation rates *in-vitro*
593 (Figs S3C-D).

594 Remarkably, the PDO-derived xenograft (PDO-X) tumors resulting from HCM-
595 CSHL-0366-C50, NH85TSc, NH87T and NH95T organoids recapitulated the morphology
596 of the patient tumor despite previously being in culture for up to 18 passages. The tumor
597 and the organoid-xenograft from patient HCM-CSHL-0366-C50 were the most distinct
598 and showed squamous differentiation with pleomorphic cells that were variable in size
599 and had cells with either oval or pale nuclei with prominent nucleoli (Fig 4D). We also
600 observed cells that had thin spinous connections with adjacent tumor cells which were

601 also present in the organoids (black arrows Fig 4D). NH85TSc tumor and organoid-
602 xenografts both showed highly undifferentiated morphology with large necrotic areas and
603 stromal compartments infiltrating in between the tumors (Fig 4D). Lastly, NH87T tumor
604 was a TNBC type with AdCC like features that are characterized by cribriform
605 architectural patterns and pseudo-lumens (31,44). While the organoid-derived xenograft
606 had a lower stromal composition compared to the patient tumor it still recapitulated the
607 key features of the original tumor, including the presence of cuboidal cells, cribriform
608 architecture and pseudo-lumens (yellow arrows Fig 4D). This particular subtype is also
609 characterized by the expression of CD117 and CK5/6 (45). We performed IHC for CD117
610 and CK5/6 on PDO-X tumor sections from NH87T and observed positive membrane
611 labeling for CD117 along with areas of high and low CK5/6 labeling as observed in the
612 clinical IHC labeling of the patient slide (Fig S4C).

613 While NH95T formed tumors at a much lower efficiency (Fig 4A, B), the tumors
614 that did form had morphological similarities to the original patient tumor (Fig 4D). The
615 patient tumor cells showed an organization pattern which was recapitulated in the
616 organoid-derived xenografts and to some extent in the NH95T organoids (Fig 4D). Of
617 note, patients from whom the fast-growing organoids were derived, i.e., HCM-CSHL-
618 0366-C50, NH85 and NH87, presented with poor diagnoses and outcomes (TMN staging
619 in Fig 4B). Patient HCM-CSHL-0366-C50 had rapid metastasis to the brain and
620 succumbed from the disease, patient NH85 had local recurrence in the lymph node and
621 patient NH87 presented with a positive lymph node at initial diagnosis (Table S1). Using
622 IHC with a human mitochondrial antibody we only observed micro-metastasis and single-
623 cell metastasis in our experiments (Fig 4D, S4B, Table S6), however, altering the

624 experimental conditions such as: prolonging the end-point, using a more
625 immunocompromised NOD/SCID gamma (NSG) mouse model or tail-vein injections
626 might result in more metastatic lesions and will be examined in future studies. Thus far,
627 our data shows that TNBC PDOs recapitulate the tumor intrinsic properties of the original
628 tumors at genomic, transcriptomic and morphological levels.
629

630 **TNBC derived organoids are enriched for luminal-progenitor-like cells**

631 To fully assess the utility of PDOs as cancer models, we next asked what were the
632 cell types represented within these organoids and how did they differ from the normal
633 derived PDOs. In order to profile the cell-types present within the normal and TNBC
634 organoids, we used a combination of flow cytometry and single cell RNA-seq (scRNA-
635 seq). We assayed for mammary epithelial lineages using EPCAM and CD49f as luminal
636 and basal cell markers respectively (21,22,40,46). As previously shown (22), the normal
637 derived organoids recapitulate the EPCAM+ luminal lineage, EPCAM+/CD49f+ luminal
638 progenitor cells and EPCAM-/CD49f+ basal cell lineages (NM07NL in Fig 5A). The
639 populations slightly fluctuated between the different normal lines and passages but
640 overall, the different lineages were observed in all normal organoids (Fig 5A-C, S5A,
641 Table S7). TNBC organoids were less heterogeneous in the expression of these markers,
642 and as observed by RNA-seq signatures (Fig 3B, S2D), showed an enrichment of luminal
643 progenitor-like cells (Fig 5A-B) which had co-expression of EPCAM and CD49f but
644 showed slightly different and patient-specific expression patterns of these markers (Fig
645 5A, S5A). TNBC organoids also showed a higher gene expression of EPCAM and CD49f
646 (gene *ITGA6*) compared to the normal organoids (Fig S5B). Some TNBC PDOs, for

647 example NH84T, had a more luminal cell flow-cytometry profile, despite clustering with
648 the other LT-TNBCs at the transcriptome level. Interestingly, metastatic TNBC lines HCM-
649 CSHL-0773-C50 and HUB4T also had an enrichment of more luminal-like cells (Fig 5B,
650 S5D). NH85TSc which was derived from a patient that later relapsed and HCM-CSHL-
651 0366-C50, which showed rapid progression in the patient, also had a more luminal cell
652 profile with higher percentage of EPCAM only cells (Fig 5A-B) despite showing gene
653 signatures of the basal-like BCs. The EPCAM/CD49f profile of the individual lines was
654 stable overtime at different passages (Fig S5C). Additionally, while always observed in
655 the normal-organoids, we rarely observed CD49f only cells in the TNBC-organoids (Fig
656 5A-B, S5A, D). We performed longitudinal analysis for one of the patient-derived organoid
657 lines and observed an enrichment of the tumor LPs in culture over time in early passages
658 (Fig 5C) which coincided with a more pronounced copy number profile of the TNBC
659 organoids (Fig 5D) suggesting LP-like cells being the predominant cancer cell population
660 in TNBCs and potentially the cell of origin of these cancers.

661 In order to better understand the cellular composition of these organoids, we
662 performed scRNA-seq analysis on three normal and seven TNBC organoid lines. The
663 study was done in three experimental batches (Table S7) and processed where each of
664 the samples was quality controlled and filtered to remove cells with high mitochondrial
665 gene content and low gene identification (i.e. dead cells). The filtered matrix was SC-
666 transformed using Seurat (47) and the samples were integrated to account for the
667 different batches (Fig 5E). Clustering of data identified 19 clusters (Fig 5E) with some
668 clusters being highly representative of the normal organoids (clusters 0, 3, 5, 6, 9, 11 in
669 Fig 5F) while other clusters were largely comprised of tumor cells. Tumor organoids were

670 distinct from normal, with some overlap in clusters 0, 3 and 9 (Fig 5F-H). Expression
671 profiles of *EPCAM* and *ITGA6* (gene encoding CD49f) (Fig 5H) recapitulated the data
672 from flow cytometry (Fig 5A-B) where tumor organoids are predominantly of the luminal
673 progenitor nature as measured by co-expression of *EPCAM* and *ITGA6* (CD49f) while the
674 normal organoids recapitulate the three broad mammary epithelial lineages: mature
675 luminal (EPCAM+), luminal progenitors (LPs, EPCAM+/CD49f+) and myoepithelial/basal-
676 like cells (Epcam-low/CD49f+). Unlike the normal organoids, all tumor cells were
677 *EPCAM*+ and a subset of those had low *ITGA6* (Fig 5H). In concordance with the flow
678 cytometry data (Fig 5A), the *ITGA6*-low cells were largely present in HCM-CSHL-0366-
679 C50 and NH85TSc. NH93T and the NH95T organoid line had similar profiles and
680 occupied similar space as the normal LP cells (Fig 5H). Lines NH87T and HCM-CSHL-
681 0655-C50 are paired primary and lymph node metastasis samples from the same patient
682 and occupied very similar spaces with some overlap with normal LP cells (Fig 5H), while
683 NH64T, NH85TSc and HCM-CSHL-0366-C50 are very distinct from the normal lines.
684 Interestingly, there was a significant overlap between the clusters identified between
685 NH85TSc and HCM-CSHL-0366-C50 samples (Fig 5H) in keeping with their
686 transcriptome similarity (Fig 3A, S2A) and rapid tumor progression *in vivo* (Fig 4A).
687 This data builds on previous studies that have correlated a LP-like expression
688 signature with basal-like breast cancers (40) and have shown that LP cells are the cell of
689 origin of BRCA-mutated basal breast cancers (48). We suggest that LP-like cells are the
690 possible cell of origin for a broader subset of TNBCs. Interestingly, a higher percent of
691 EPCAM+ only cells seems to be associated with a greater degree of disease progression,
692 however, this needs to be further investigated.

693

694 **Tumor LP-like cells exhibit altered expression and have an upregulation of NOTCH**
695 **and MYC downstream pathways**

696 Since TNBC organoids had a large proportion of LP-like cells that seemed distinct
697 from the normal LP cells (Figs 5, S5) we performed integrated single cell analysis by SC-
698 transforming individual samples and performing an anchor-dependent integration for all
699 individual samples. We identified thirteen clusters between the tumor and normal cells
700 (Fig 6A) out of which clusters 6, 8, 10 and 9 represented cell cycle clusters (Fig S6A) with
701 cluster 6 representing a population of G1-S phase cells, cluster 8 representing S-phase
702 cells, cluster 9 S-G2M transition cells and cluster 10 marking G2M cells while the
703 remaining clusters represented G1 cells (Fig S6A-C). The cell clusters identified
704 represented biologically meaningful cell-types recently annotated by single cell
705 sequencing of adult human breast epithelium (49). Cell types identified included: XBP1,
706 AGR2 expressing mature luminal cells (Mat lum, cluster 1), APOE, KRT6A expressing
707 basal stem-like cells (basal SC, cluster 4), TAGLN, TIMP1 expressing myoepithelial cells
708 (Myo, cluster 3), SLPI expressing luminal progenitor cells (LP, cluster 7) and LTF
709 expressing secretory cells (LP sec, cluster 2) (Fig 6B, S6C) (49). To further validate the
710 lineage specification in organoids we overlaid the published mammary epithelial gene-
711 sets (40) and computed a score for each of the three mammary epithelial cells. The
712 normal cell clusters showed a much stronger cell type enrichment score while the tumor
713 cells despite being predominantly EPCAM+/CD49f+ showed a more diffused enrichment
714 of these signatures, suggesting some cell type heterogeneity within these tumor
715 organoids (Fig 6C). In normal organoids, Clusters 2,7 and 11 had a higher enrichment for

716 luminal progenitor (LP) cell score, while cluster 1 showed an enrichment for mature
717 luminal (mature Lum) cells and cluster 3 was predominantly of the basal mammary stem
718 cell (Basal SC) compartment (Fig 6C, S6C). The tumor organoids showed varied
719 expression profiles of various luminal/basal markers (Fig S6D) and a diffused cluster
720 specific enrichment of lineage scores (Fig 6C, S6E), suggesting that while they have lost
721 proper cell type specification, the clusters identified using this integration approach have
722 some similarity to the normal cell lineages.

723 With the aim to identify mechanisms that underlie the tumor luminal progenitor-like
724 (LP-like) cells in TNBCs, we performed a differential expression analysis between tumor
725 and normal cells of clusters 2, 7 and 11. We identified 1103 significantly differentially
726 expressed genes (p_val_adj<0.05) between the tumor LPs vs normal LPs (Table S8).
727 GSEA on the differentially expressed genes showed an enrichment of NOTCH signaling
728 related genes in the tumor LPs versus an enrichment of BMP2 targets in the normal LPs
729 (Fig 6D). Leading edge genes from these gene sets showed a consistent downregulation
730 of NOTCH signaling related genes, *RPS27A*, *YBX1*, *JAG1*, *MDK*, *MYC* and *SEM1*, and
731 an upregulation of BMP2 target genes, *LTF*, *MGP*, *KRT16*, *KRT7*, *PLAAT3* and *NTRK2*
732 in normal LPs from all three patients (Fig 6E). MDK, JAG1, YBX1 are ligands of NOTCH1,
733 while *MYC* and *RSP27A* are downstream targets. While NOTCH activity was present in
734 normal LP cells, it was more pronounced in the TNBC organoids (Fig S6F).

735 A further investigation into differentially regulated pathways showed upregulation
736 of genes involved in ribosomal biogenesis and translation in the tumor LPs. We performed
737 motif enrichment analysis on the differentially expressed LP genes and found a significant
738 enrichment of MYC binding sites in the tumor LP expressed genes (Fig S6G). We used

739 published MYC signatures to assess MYC activity in tumor and normal organoids and
740 found a higher ubiquitous enrichment of MYC activation in tumor organoid cells, while in
741 normal organoids it seemed to be higher in the proliferating cell clusters (Fig S6F).
742 Similarly, while the NOTCH pathway was hyper-active in tumor LPs it was ubiquitously
743 active across all tumor cells (Fig S6F). We tested whether NOTCH and MYC activation
744 was required for organoid formation and seeded NH95T, NH85TSc and HCM-CSHL-
745 0366-C50 as single cells in regular organoid growth medium in the presence of DAPT
746 (NOTCH inhibitor) or MYCi975 (MYC inhibitor) (50) and allowed for organoids to form for
747 12 days. We saw a significant reduction in organoid formation in the presence of both
748 MYC and NOTCH inhibitors for NH85TSc and NH95T but not for HCM-CSHL-0366-C50
749 (Fig 6F). While HCM-CSHL-0366-C50 did not show reduction in the number of organoids
750 formed, we did observe a significant difference in organoid size with DAPT and MYCi975
751 suggesting growth defects on inhibition of these pathways in all lines (Fig S6H). We
752 further tested whether inhibition of NOTCH and MYC is necessary for the formation of
753 normal PDOs. Since, normal PDOs have multiple populations we flow-sorted the basal
754 stem-like CD49f+ve population and EPCAM+/CD49f+ LP cells and seeded them as single
755 cells in the presence of DAPT or MYCi975 (Fig S6I). While DAPT had no effect on
756 generation of normal PDOs from both LP and basal stem-like cells, MYC inhibition
757 resulted in no organoid formation.

758 Thus, our data shows that TNBC organoids have an underlying LP expression
759 signature which is driven by LP-like cells that exhibit altered expression from normal LPs
760 by hyperactivation of NOTCH and MYC signaling. These pathways are necessary for the

761 formation of some, but not all, TNBC organoids and when perturbed result in proliferation
762 defects in all tumor lines examined.
763
764 **TNBC organoids are comprised of heterogenous cancer cell populations**
765 Since normal organoids have clearly defined lineages, the clustering of the
766 integrated dataset (Fig 6) was largely driven by the normal organoids. To assess the
767 heterogeneity that exists within TNBC organoid cultures we performed a similar integrated
768 analysis on the TNBC organoids only and identified 13 different cell clusters (Fig 7A,
769 S7A). Clusters 3,9,6,8 and 5 correspond to the different cell cycle phases (Fig 7A) and
770 the remaining clusters are G1 cells (Fig 7A-B). We identified the markers that uniquely
771 define each of the clusters (Fig 7C, Table S9) and performed GSEA on the marker genes
772 to identify the phenotypes associated with each cell cluster (Fig 7D). Cluster 0
773 represented a mixed cell cluster with some enrichment of mature luminal-like cells.
774 Cluster 1 was defined as Basal-like 1 as it showed a positive enrichment of a basal breast
775 cancer gene-set including specific markers *SAT1*, *GABRP*, *TM4SF1*, *TTYH1*, *KRT16* and
776 *KRT6A* genes (Table S9). Cluster 2 is mesenchymal-like due to the selective expression
777 of mesenchyme genes including *CCN2*, *TPM1* and *LAMB1*. Cluster 7, MGP-high cluster,
778 is a relatively distinct cluster with a very specific high expression of *MGP* (Matrix Gla
779 Protein) in all TNBC organoid lines (Fig 7C, S7E). MGP is normally expressed in smooth
780 muscle cells perhaps suggesting that these cells might have contractile abilities. Cluster
781 12 had a very high expression of ribosomal and translation related genes (Fig 7C-D).
782 Clusters 4,10 and 11 are the most distinct of the tumor cell clusters and are identified by
783 the unique expression of certain genes. Cluster 10 represents an NFKB-active cluster

784 that was also found in normal organoids (Fig S6C), albeit with some differences in gene
785 expression. As in normal organoids (Fig 6A, S6B, 8C). this cluster shows the expression
786 of *CXCL1*, *CXCL3*, *NFKBIA* etc. Cluster 11 is also defined by a basal breast cancer
787 signature (Fig 7C-D), is labeled Basal-like 2 and shows a selective expression of *KRT17*.
788 Cluster 4 is a hypoxia cell cluster that shows selective high expression of *EGLN3*,
789 *NDRG1*, *VEGFA* and likely represents cells at the center of these organoids (Fig 7C-D).
790 This cluster is represented by a higher hypoxia score (Fig 7C-D), which interestingly
791 correlates with basal stem cell (Basal SC) signatures and shows comparatively low
792 activity of MYC and NOTCH pathways (Fig 7E and S7D-E). We also noted that the cluster
793 specific genes for all clusters were fairly well conserved between the different PDO lines
794 (Fig S7F). Overall, this data suggests that while the TNBC organoids retain patient
795 intrinsic properties (Fig 1-4) there are common cell signatures that define the cell types
796 within these organoids.

797 The distinct gene expression patterns within these cell clusters also suggests that
798 the TNBC organoids are comprised of multiple cell types that are known to be involved in
799 tumorigenesis, tumor progression and metastasis, and can therefore be used as models
800 to study the various aspects of breast cancer progression.

801

802 **Discussion**:

803 We developed a patient-derived biobank of normal and breast cancer organoids
804 from a diverse group of patients with a focus on the highly aggressive TNBC subtype.
805 Our biobank is heterogenous in terms of ethnic/racial backgrounds, patient age and
806 breast cancer subtypes. A comprehensive genomic, transcriptomic, and cellular

807 characterization of these PDO models demonstrate their faithful recapitulation of the
808 patient tumor intrinsic properties and hence validates them as cancer models to study
809 various aspects of breast cancer progression and treatment.

810 Patient-derived organoids are an exciting step towards personalized medicine with
811 a promise to be used for real-time drug screens for guiding patient treatment (14). When
812 successfully established, breast cancer PDOs retain the genomic and transcriptomic
813 features of breast cancers and their parent tissues. We found that, the TNBC organoids
814 that showed long-term robust growth had activated MYC signaling, an LP-like gene
815 expression signature and were overwhelmingly composed of LP-like cells. As these are
816 among the most aggressive forms of breast cancer, organoid models appear to be an
817 exciting avenue for their study. In addition to replicating the tumor specific features *in-*
818 *vitro*, TNBC PDOs when transplanted into NOD/SCID mice generated tumors with
819 remarkable morphological similarity to that of the original patient tumors, despite being in
820 long-term organoid culture for up-to 18 passages. Long-term cultured organoids, thus,
821 maintain the intrinsic ability to represent the tumor from which they were derived when
822 placed in an *in vivo* environment. Interestingly, in our study not all PDOs readily generated
823 primary tumors *in vivo* despite being highly proliferative and genomically aberrant
824 organoids. The association of PDOs that generated tumors rapidly, with worse outcomes
825 and diagnosis, suggests that there might be a biologically relevant explanation for this
826 observed difference and must be investigated further along with long-term patient follow
827 up information. Patient-derived xenograft (PDX) models also retain the histopathology of
828 the original patient tumors (51,52) and a recent pre-print study showed successful
829 derivation of breast cancer organoids from PDX models (53). Our models are

830 complementary to this system and represent an opportunity to derive PDOs first, perform
831 *in-vitro* assays and drug screens, followed by PDOX derivation for *in-vivo* validation
832 studies.

833 While successfully established PDOs faithfully recapitulate the patient tumor
834 properties, the efficiency of establishment of these cultures is currently less than ideal.
835 The common challenges we faced during establishing organoid cultures included normal
836 outgrowth in some tumor organoid cultures, stromal outgrowth, non-proliferating or
837 dormant tumor cells, or limited growth potential *ex-vivo*. Since the medium composition
838 for tumor organoid growth also supports the culture of normal mammary organoids (22)
839 the purity of the starting tumor tissue might govern the time to generation of tumor-only
840 organoid cultures and early passages should be meticulously tested at a genomic level
841 before any pharmacological studies are carried out with these models. Furthermore,
842 some subtypes of breast cancer might be challenging to culture including the less
843 proliferative Luminal A subtype or the more mesenchymal claudin-low group of TNBCs.
844 Efforts to culture tumor-organoids in reduced complexity medium (53), might alleviate
845 some of these challenges and needs to be examined in future studies. The PDO culture
846 conditions are comprised of complex growth factors including, EGF, FGF10, Neuregulin,
847 R-spondin etc. (20), and it remains to be tested whether the removal or alteration of these
848 components would be beneficial in selecting out cancer cells over normal epithelial cells,
849 resulting in higher efficiency in the establishment of TNBC PDOs. However, the
850 simultaneous growth of normal and TNBC organoids in the same culture conditions allows
851 for comparison of similar cell types within these organoids.

852 We used an integrated scRNA-seq approach to compare normal LP cells with the
853 tumor LP-like cells and identified hyperactivation of NOTCH and MYC signaling in the
854 tumor compared to normal LPs. LP cells were previously shown to be the cell of origin of
855 BRCA1 mutated basal-like BCs (40,48) and are speculated to be involved in
856 tumorigenesis of all basal-like BCs due to similarity of gene expression profiles (40). Our
857 data provides strong support for this hypothesis by showing that TNBC PDOs are largely
858 comprised of LP-like cells and suggests that hyperactivation of NOTCH and MYC
859 signaling might be relevant in tumorigenesis from LP-cells (Fig 7F). Mouse studies have
860 shown that Notch signaling is important in the maintenance of luminal lineage of the
861 normal mammary gland (54–57) and overactivation of Notch in luminal progenitors
862 resulted in hyperplasia and acquisition of self-renewal properties (54). Notch activation
863 drives the luminal fate specificity in normal mammary gland and can reprogram committed
864 basal mammary cells to an ER- luminal cell fate via multipotent embryonic cell states
865 (55,58), suggesting a role for Notch in fate specification and cellular plasticity. While our
866 data provides strong support for hyperactivation of NOTCH signaling in LP-like cells in
867 TNBCs it remains to be tested whether these cells arose from activation of NOTCH in
868 normal LP cells or from reprograming of normal basal cells. Our data suggest that
869 activation of MYC along with NOTCH might also have a role in this transformation. Given
870 the complementarity of the NOTCH and MYC pathways, further studies are needed to
871 gain better resolution of this mechanism. The PDO system, thus, can be exploited for
872 identification of dysregulated pathways in cancers which can then be perturbed in the
873 normal PDOs to understand aspects of the origins of cancers.

874 A detailed analysis of the cell type heterogeneity of TNBC organoids also revealed
875 the presence of MYC/NOTCH-activity low and hypoxia high cells amongst various other
876 cancer relevant cell types. Recent studies have shown that cancer cells can enter a MYC-
877 low diapause-like state of dormancy upon chemotherapy treatment and result in therapy
878 escape (59,60). Furthermore, single cell analysis from TNBC tumors with residual disease
879 after chemotherapy treatment showed an enrichment of hypoxia, angiogenesis, EMT and
880 ECM degradation related genes in the persistent tumors (61). We hypothesize that in
881 PDOs with these expression signatures, the cells that exhibit a MYC-low/hypoxia-high
882 profile are the likely candidates for chemotherapy escape (Fig 7F). Additionally, fate
883 mapping and other studies have shown that tumor cells exposed to hypoxia have a higher
884 tendency to result in metastasis (62,63). It remains to be tested to what extent these MYC-
885 low/hypoxia-high cells contribute towards metastasis and/or chemotherapy resistance,
886 and these patient-derived organoids will be an invaluable tool to answer these questions.
887 In summary, we developed a diverse biobank of BC organoids with a focus on
888 TNBC-derived organoids. We have thoroughly characterized these models as valid
889 systems that mimic the various aspects of patients' tumors, including genomic
890 alternations, transcriptomic signatures, cell type specificity and morphological
891 characteristics. Comparison of TNBC and normal-derived organoids provides important
892 insights into the mechanisms regulating tumorigenesis that can then be validated by
893 perturbation in normal PDOs (21). This comparison can also be used to identify novel
894 tumor specific targets that may play an important role in tumor growth and progression
895 (64). These next generation cancer models and the data derived from them offer vast
896 utilities and can be used for drug-screens, co-culture experiments, metabolomics, and

897 fate mapping studies to better understand the mechanisms driving cancers and for
898 identifying more specific treatment options.
899

900 **Acknowledgements**

901 We would like to thank members of the Spector lab for critical discussions and
902 advice throughout the course of this study. We thank Bodu Liu for critical review of this
903 manuscript. We also thank the various members of NCI-Human Cancer Model Initiative
904 (HCMI) for general advice and inputs on organoid cultures. We acknowledge the CSHL
905 Cancer Center Shared Resources (Animal, Flow Cytometry, Histology, Microscopy,
906 Organoid, Next-Gen Sequencing, and Single Cell Biology) for services and technical
907 expertise (NCI 2P3OCA45508). The sequencing analysis was performed using
908 equipment purchased through NIH grant S10OD028632-01. We thank Dr. Zihua Wang
909 for advice and a protocol for SMASH library preparation, Dr. Qing Gao for assistance with
910 histological sample preparation and optimization, Jill Habel for assistance with mouse
911 experiments. We are grateful for the support of Dr. Tawfiqui Bhuiya (Northwell Health) for
912 tumor tissue sections and regret his untimely passing. We deeply appreciate the efforts
913 of the Northwell Health Biobanking team for sample acquisition and the patients and their
914 families who donated tissues for research. This research was supported by
915 CSHL/Northwell Health support (D.L.S.), NCI 5P01CA013106-Project 3 (D.L.S.), Leidos
916 Biomedical HHSN26100008 (D.A.T, D.L.S), and the Manhasset Women's Coalition
917 Against Breast Cancer (S.B.). This research was also supported by NCI Breast SPORE
918 program P50-CA58223 (C.M.P), and U01CA238475-01 (C.M.P).

References

1. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009 Mar;27(8):1160–7.

2. Perou C, Sørlie T, Eisen M, Rijn MV, Jeffrey S, Rees C, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747–52.

3. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences of the United States of America. 2003;100:8418–23.

4. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Molecular Oncology. 2011 Feb;5(1):5–23.

5. Herschkowitz J, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biology. 2006;8:R76–R76.

6. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Research. 2010 Sep 2;12(5):R68.

7. Lehmann BD, Pietenpol JA, Tan AR. Triple-Negative Breast Cancer: Molecular Subtypes and New Targets for Therapy. American Society of Clinical Oncology Educational Book. 2015;(35):e31–9.

8. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. The Journal of clinical investigation. 2011 Jul;121(7):2750–67.

9. Howlader N, Altekruse SF, Li CI, Chen VW, Clarke CA, Ries LAG, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. Journal of the National Cancer Institute. 2014 Apr;106(5).

10. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012 Jun;486(7403):346–52.

11. Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008 Mar;26(8):1275–81.

12. Rueda OM, Sammut S-J, Seoane JA, Chin S-F, Caswell-Jin JL, Callari M, et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. Nature. 2019 Mar;567(7748):399–404.

13. Fujii M, Sato T. Somatic cell-derived organoids as prototypes of human epithelial tissues and diseases. Nature Materials. 2021 Feb;20(2):156–69.

14. Tuveson D, Clevers H. Cancer modeling meets human organoid technology. Science. 2019 Jun 7;364(6444):952–5.

15. Sato T, Stange DE, Ferrante M, Vries RGJ, van Es JH, van den Brink S, et al. Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. Gastroenterology. 2011 Nov 1;141(5):1762–72.

16. Boj SF, Hwang C-I, Baker LA, Chio IIC, Engle DD, Corbo V, et al. Organoid Models of Human and Mouse Ductal Pancreatic Cancer. Cell. 2015 Jan 15;160(1):324–38.

17. Hoffmann K, Berger H, Kulbe H, Thillainadarasan S, Mollenkopf H-J, Zemojtel T, et al. Stable expansion of high-grade serous ovarian cancer organoids requires a low-Wnt environment. The EMBO journal. 2020 Mar;39(6):e104013–e104013.

18. Kopper O, de Witte CJ, Lohmussaar K, Valle-Inclan JE, Hami N, Kester L, et al. An organoid platform for ovarian cancer captures intra- and interpatient heterogeneity. Nature medicine. 2019 May;25(5):838–49.

19. Beshiri ML, Tice CM, Tran C, Nguyen HM, Sowalsky AG, Agarwal S, et al. A PDX/Organoid Biobank of Advanced Prostate Cancers Captures Genomic and Phenotypic Heterogeneity for Disease Modeling and Therapeutic Screening. Clinical cancer research : an official journal of the American Association for Cancer Research. 2018 Sep;24(17):4332–45.

20. Sachs N, de Ligt J, Kopper O, Gogola E, Bounova G, Weeber F, et al. A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. Cell. 2018 Jan;172(1–2):373-386.e10.

21. Dekkers JF, Whittle JR, Vaillant F, Chen H-R, Dawson C, Liu K, et al. Modeling Breast Cancer Using CRISPR-Cas9-Mediated Engineering of Human Breast Organoids. Journal of the National Cancer Institute. 2020 May;112(5):540–4.

22. Rosenbluth JM, Schackmann RCJ, Gray GK, Selfors LM, Li CM-C, Boedicker M, et al. Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. Nature Communications. 2020 Dec;11(1):1711–1711.

23. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012 Mar;483(7391):603–7.

24. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012 Mar;483(7391):570–5.

25. Yu C, Mannan AM, Yvone GM, Ross KN, Zhang Y-L, Marton MA, et al. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. Nature Biotechnology. 2016 Apr;34(4):419–23.

26. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61–70.

27. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016 Jun;534(7605):47–54.

28. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012;486(7403):395–9.

29. Perou CM. Molecular Stratification of Triple-Negative Breast Cancers. The Oncologist. 2011 Jan;16(S1):61–70.

30. Prakash O, Hossain F, Danos D, Lassak A, Scribner R, Miele L. Racial Disparities in Triple Negative Breast Cancer: A Review of the Role of Biologic and Non-biologic Factors. Frontiers in Public Health. 2020;8:762–762.

31. Fusco N, Geyer FC, De Filippo MR, Martelotto LG, Ng CKY, Piscuoglio S, et al. Genetic events in the progression of adenoid cystic carcinoma of the breast to high-grade triple-negative breast cancer. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2016 Nov;29(11):1292–305.

32. Wang Z, Andrews P, Kendall J, Ma B, Hakker I, Rodgers L, et al. SMASH, a fragmentation and sequencing method for genomic copy number analysis. Genome research. 2016 Jun;26(6):844–51.

33. Weigman VJ, Chao H-H, Shabalin AA, He X, Parker JS, Nordgard SH, et al. Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. Breast Cancer Res Treat. 2012 Jun;133(3):865–80.

34. Gong Y, Zack TI, Morris LGT, Lin K, Hukkelhoven E, Raheja R, et al. Pan-cancer genetic analysis identifies PARK2 as a master regulator of G1/S cyclins. Nature genetics. 2014 Jun;46(6):588–94.

35. Wilczak W, Rashed S, Hube-Magg C, Kluth M, Simon R, Büscheck F, et al. Up-regulation of mismatch repair genes MSH6, PMS2 and MLH1 parallels development of genetic instability

and is linked to tumor aggressiveness and early PSA recurrence in prostate cancer. Carcinogenesis. 2017 Jan 1;38(1):19–27.

36. Chakraborty U, Dinh TA, Alani E. Genomic Instability Promoted by Overexpression of Mismatch Repair Factors in Yeast: A Model for Understanding Cancer Progression. Genetics. 2018 Jun 1;209(2):439–56.

37. Fan C, Prat A, Parker JS, Liu Y, Carey LA, Troester MA, et al. Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. BMC Medical Genomics. 2011 Dec;4(1):3–3.

38. Garcia-Recio S, Thennavan A, East MP, Parker JS, Cejalvo JM, Garay JP, et al. FGFR4 regulates tumor subtype differentiation in luminal breast cancer and metastatic disease. Journal of Clinical Investigation. 2020 Aug;130(9):4871–87.

39. Gatza ML, Silva GO, Parker JS, Fan C, Perou CM. An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. Nature Genetics. 2014 Oct;46(10):1051–9.

40. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nature Medicine. 2009 Aug;15(8):907–13.

41. Hu Z, Fan C, Livasy C, He X, Oh DS, Ewend MG, et al. A compact VEGF signature associated with distant metastases and poor outcomes. BMC Medicine. 2009 Mar;7(1):9–9.

42. Giraddi RR, Chung C-Y, Heinz RE, Balcioglu O, Novotny M, Trejo CL, et al. Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development. Cell reports. 2018 Aug;24(6):1653-1666.e7.

43. Davies BR, Greenwood H, Dudley P, Crafter C, Yu D-H, Zhang J, et al. Preclinical Pharmacology of AZD5363, an Inhibitor of AKT: Pharmacodynamics, Antitumor Activity, and Correlation of Monotherapy Activity with Genetic Background. Mol Cancer Ther. 2012 Apr 1;11(4):873–87.

44. Marchiò C, Weigelt B, Reis-Filho JS. Adenoid cystic carcinomas of the breast and salivary glands (or “The strange case of Dr Jekyll and Mr Hyde” of exocrine gland carcinomas). Journal of clinical pathology. 2010 Mar;63(3):220–8.

45. Weigelt B, Horlings H, Kreike B, Hayes M, Hauptmann M, Wessels L, et al. Refinement of breast cancer classification by molecular characterization of histological special types. The Journal of Pathology. 2008;216(2):141–50.

46. Raouf A, Zhao Y, To K, Stingl J, Delaney A, Barbara M, et al. Transcriptome Analysis of the Normal Human Mammary Cell Commitment and Differentiation Process. Cell Stem Cell. 2008 Jul 3;3(1):109–18.

47. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-1902.e21.

48. Molyneux G, Geyer FC, Magnay F-A, McCarthy A, Kendrick H, Natrajan R, et al. BRCA1 Basal-like Breast Cancers Originate from Luminal Epithelial Progenitors and Not from Basal Stem Cells. Cell Stem Cell. 2010 Sep 3;7(3):403–17.

49. Nguyen QH, Pervolarakis N, Blake K, Ma D, Davis RT, James N, et al. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nature Communications. 2018 May 23;9(1):2028.

50. Han H, Jain AD, Truica MI, Izquierdo-Ferrer J, Anker JF, Lysy B, et al. Small-Molecule MYC Inhibitors Suppress Tumor Growth and Enhance Immunotherapy. Cancer cell. 2019 Nov;36(5):483-497.e15.

51. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, et al. A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. Cell. 2016 Sep;167(1):260-274.e22.

52. DeRose YS, Wang G, Lin Y-C, Bernard PS, Buys SS, Ebbert MTW, et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. Nature Medicine. 2011 Nov;17(11):1514–20.

53. Guillen KP, Fujita M, Butterfield AJ, Scherer SD, Bailey MH, Chu Z, et al. A breast cancer patient-derived xenograft and organoid platform for drug discovery and precision oncology. bioRxiv. 2021 Mar 11;2021.02.28.433268.

54. Bouras T, Pal B, Vaillant F, Harburg G, Asselin-Labat M-L, Oakes SR, et al. Notch Signaling Regulates Mammary Stem Cell Function and Luminal Cell-Fate Commitment. Cell Stem Cell. 2008 Oct;3(4):429–41.

55. Lilja AM, Rodilla V, Huyghe M, Hannezo E, Landragin C, Renaud O, et al. Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. Nature Cell Biology. 2018 Jun;20(6):677–87.

56. Rodilla V, Dasti A, Huyghe M, Lafkas D, Laurent C, Reyal F, et al. Luminal Progenitors Restrict Their Lineage Potential during Mammary Gland Development. PLOS Biology. 2015 Feb 17;13(2):e1002069.

57. Yalcin-Ozuysal Ö, Fiche M, Guitierrez M, Wagner K-U, Raffoul W, Brisken C. Antagonistic roles of Notch and p63 in controlling mammary epithelial cell fates. Cell Death & Differentiation. 2010 Oct;17(10):1600–12.

58. Bland P, Howard BA. Mammary lineage restriction in development. Nature Cell Biology. 2018 Jun;20(6):637–9.

59. Dhimolea E, de Matos Simoes R, Kansara D, Al'Khafaji A, Bouyssou J, Weng X, et al. An Embryonic Diapause-like Adaptation with Suppressed Myc Activity Enables Tumor Treatment Persistence. Cancer Cell. 2021 Feb;39(2):240-256.e11.

60. Rehman SK, Haynes J, Collignon E, Brown KR, Wang Y, Nixon AML, et al. Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. Cell. 2021 Jan;184(1):226-242.e21.

61. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell. 2018;1–15.

62. Godet I, Shin YJ, Ju JA, Ye IC, Wang G, Gilkes DM. Fate-mapping post-hypoxic tumor cells reveals a ROS-resistant phenotype that promotes metastasis. Nature Communications. 2019;10(1):4862–4862.

63. Zhao Y, Fu X, Lopez JI, Rowan A, Au L, Fendler A, et al. Selection of metastasis competent subclones in the tumour interior. Nature Ecology & Evolution. 2021 May 17;1–13.

64. Chang K-C, Diermeier SD, Yu AT, Brine LD, Russo S, Bhatia S, et al. MaTAR25 lncRNA regulates the Tensin1 gene to impact breast cancer progression. Nature Communications. 2020 Dec;11(1):6438–6438.

65. Chen R, Im H, Snyder M. Whole-Exome Enrichment with the Roche NimbleGen SeqCap EZ Exome Library SR Platform. Cold Spring Harb Protoc. 2015 Jul 1;2015(7):pdb.prot084855.

66. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

68. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011 Jun 15;27(12):1691–2.

69. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012 Mar 1;22(3):568–76.

70. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010 Sep 1;38(16):e164–e164.

71. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research. 2018 Jan 4;46(D1):D1062–7.

72. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Research. 2019 Jan 8;47(D1):D941–7.

73. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. The American Journal of Human Genetics. 2016 Oct 6;99(4):877–85.

74. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nature Genetics. 2016 Dec;48(12):1581–6.

75. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018 Nov;28(11):1747–56.

76. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013 Jan;29(1):15–21.

77. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014 Dec 5;15(12):550.

78. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.

79. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012 Jul;9(7):676–82.

80. Huang Y, McCarthy DJ, Stegle O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. Genome Biology. 2019 Dec 13;20(1):273.

81. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 May;36(5):411–20.

82. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS. 2005 Oct 25;102(43):15545–50.

83. Chandriani S, Frengen E, Cowling VH, Pendergrass SA, Perou CM, Whitfield ML, et al. A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. PloS one. 2009 Aug;4(8):e6693–e6693.

Figure Legends:

Figure 1. Establishment and somatic variant profiling of the breast cancer organoid biobank. A.) Summary of cancer type of the various tumor tissues that were used to generate organoids. IDC: Invasive ductal carcinoma, ILC: Invasive lobular carcinoma, Met-Lym: lymph node metastasis, IMC: Invasive mucinous carcinoma, NR: no residual tumor seen, DCIS: Ductal carcinoma in-situ, other: see Table S1 **B.)** Histopathological subtypes of the tumor tissues, ER/PR: Estrogen receptor (ER) and/or Progesterone receptor (PR), NA: not assessed **C.)** Age at diagnosis of the various subgroups of patient tumor tissues **D.)** Subtype specific, self-identified racial and ethnic breakdown of the patients represented in this biobank **E.)** Pathogenic single nucleotide variants (SNVs) identified in putative cancer driver genes in patient-derived organoids. **F.)** Proportion of organoids with pathogenic SNVs identified. Pathogenic SNVs: SNVs called pathogenic by ClinVar, COSMIC or REVEL and MCAP scores from targeted gene-panel sequence (49 samples) or whole exome sequencing (1 sample) (see Table S1), limited cultures: cultures where organoids were established at early passages (p0-p1) but could not be propagated, NA: not assessed

Figure 2. Copy number alterations (CNAs) enriched in the organoid models. A.) Copy number profiles, from IGV, of the various ER+ and ER- tumor derived organoids, along with the summary of overall copy number alterations across all samples. Side panel shows the pathogenic SNVs identified in that sample from Fig 1E. **B.)** Copy number amplifications or deletions identified in putative cancer driver genes (from Fig 1E). **C.)**

Copy number across different passages of three different sets of patient-derived organoids. **D.)** Magnified view of the chromosome regions in the red boxes in C.

Figure 3. A subset of TNBC organoids recapitulate signatures of aggressive basal-like breast cancers. A.) Molecular signatures associated with the different organoid lines. The sample legends are type: Normal= reductive mammoplasty derived normal organoids, Paired Normal= Adjacent or Distal to the tumor paired normal, Normal outgrowth= no pathogenic mutations were found, Luminal= ER/PR+ organoids; driver mutation: Other= trace mutations (see Fig 1E), None= no pathogenic mutations were found, NA= not assessed **B.)** Box-plots showing the breast cancer related and TNBC-specific scores for various gene signatures associated with poor outcomes. Each dot represents a different PDO; Luminal N=12, Normal N= 7, TNBC N= 19, TNBC met= 4. Differences in experimental groups were compared using Kruskal-Wallis test followed by pairwise comparisons using Wilcoxon rank-sum test. Bonferroni-Holm method was used to adjust the family-wise error (** adjusted p-value < 0.005, * adjusted p-value < 0.05) **C.)** Light microscopy images of the various TNBC- and normal (NM04N) derived organoid lines, along with Ki67-IHC, scale bars=100μm **D.)** Distribution of maximum passage numbers tested for the various TNBC PDOs. Long-term cultures (long) are defined by p>10 with continued expansion

Figure 4. TNBC Organoids can recapitulate tumor morphology and progression *in-vivo.* A.) Overview of PDO xenotransplant experiment using TNBC PDOs. **B.)** Box plots showing the end point tumor volume for the various organoid lines transplanted into the

fat-pads of NOD-SCID mice. Each dot represents tumor volume from one injection, N1= experiment 1, N2= experiment 2. ## For NH84T microscopic primary tumors observed in histology sections from 1/10 sites. ** For NH93T microscopic primary tumors observed in histology sections from 6/10 sites. TMN staging: pathologic TMN staging from patient pathology report (Table S1) **C.)** CNV profiles of the PDOs lines selected for *in vivo* transplant experiments **D.)** H&E images of the paired patient tumor tissue, TNBC patient-derived organoids (PDO) and xenografts generated form patient-derived organoids (PDO-X). S= Stroma, T= tumor, N=necrosis. Black arrows point to the cells with spinous connections with adjacent tumor cells. Yellow arrows point to the pseudo-lumen observed within AdCC-like breast cancers. Last column shows human mitochondria IHC in the lung and the liver from a representative mouse injected with the respective PDO (scale bar=100μm).

Figure 5. TNBC organoids are enriched in luminal progenitor-like cells. A.) Representative flow-cytometry plots for normal derived organoid (NM07NL) and various TNBC organoids stained for CD49f-PE on the x-axis and EPCAM-AF647 on the y-axis. The gates are subsets of live single cells within each organoid line and represent various cell types of the mammary epithelium. L=EPCAM-high mature luminal cells, LP= EPCAM+CD49f+ luminal progenitors, B= CD49f+ basal cells, S= stromal compartment. **B.)** Quantitation of the L, LP and B gates in panel B for multiple TNBC and normal organoid lines over multiple passages (see Table S7). Data-points are plotted as mean ±SEM using GraphPad Prism. NM0s: comprises multiple normal mammoplasty derived organoids from different patients **C.)** Flux of the epithelial cells during the early passages

of organoid derivation for normal distal (DS97ND) and TNBC tumor (DS97T) samples from the same patient **D.)** Copy number plots of the TNBC organoids DS97T over multiple passages **E.)** UMAP plot of batch corrected scRNA-seq data from 3 normal and 7 TNBC organoids. Numbers on the plot represent cluster IDs. **F. & G.)** UMAP plot in A but **F.)** separated by the normal and tumor samples and **G.**) colored by cell cycle. **H.)** UMAP plot in E split by individual tumor and normal samples and showing normalized expression of *EPCAM* and *ITGA6* (gene encoding CD49f) expression patterns in individual cells.

Figure 6. Tumor LP-like cells exhibit altered gene expression and have an upregulation of NOTCH and MYC downstream pathways. A.) UMAP plots for Integrated scRNA-seq data for all samples. The numbers indicate cluster IDs. **B.)** Marker expression of various cell type specific genes in the adult human breast epithelium (49). **C.)** Plots showing combined scores for the three mammary epithelial lineages: LP score: Luminal Progenitor score, Mature Lum score: Mature luminal score, MaSC score: Mammary stem cell score (40). Dashed region indicates LP clusters 2,7 and 11 that were used to perform differential expression analysis between normal and tumor LPs. **D.)** GSEA plots showing enrichment of the differentially expressed genes between normal and tumor LPs. **E.)** Violin plots showing combined expression in clusters 2,7 and 11 of the leading-edge NOTCH signaling genes and BMP2 target genes as identified in E. **F.)** Organoid formation from single cells. Significance was assessed by two-way ANOVA ns= not significant, ** pvalue<0.005, * pvalue<0.05

Figure 7. TNBC Organoids are comprised of heterogenous cell populations. A.) UMAP plot of TNBC only integrated scRNA-seq data showing clusters identified and cell cycle phases. **B.)** Distribution of cells in each of the G1 clusters identified per organoid line. **C.)** Dot-plot showing the marker genes for each of the G1 clusters and the associated phenotypic identity of that cell cluster. **D.)** Enrichment scores from GSEA of each of the G1 clusters that showed strong enrichment of some specific pathways or phenotypes. Enrichment score is represented by -10*NES*padj.value. **E**.) Combined gene set scores for the various phenotypes. Top panel: green= MYC signature, pink= Hypoxia signature. Bottom panel: green= Basal mammary stem cell (SC) signature, pink= Hypoxia signature. White represents positive correlation of the two signatures. **F.)** Schematic (created using BioRender.com) showing the cellular composition and heterogeneity observed in normal vs TNBC PDOs when cultured in matrigel. TNBC PDOs retain the tumor SNV/CNA profiles, are largely comprised of LP-like cells that might have originated from normal LP cells by the hyperactivation of NOTCH/MYC pathways. TNBC PDOs also have cells with signatures of hypoxia which is anti-correlated with NOTCH/MYC and positively associated with basal mammary stem cell signatures.

Figure 1

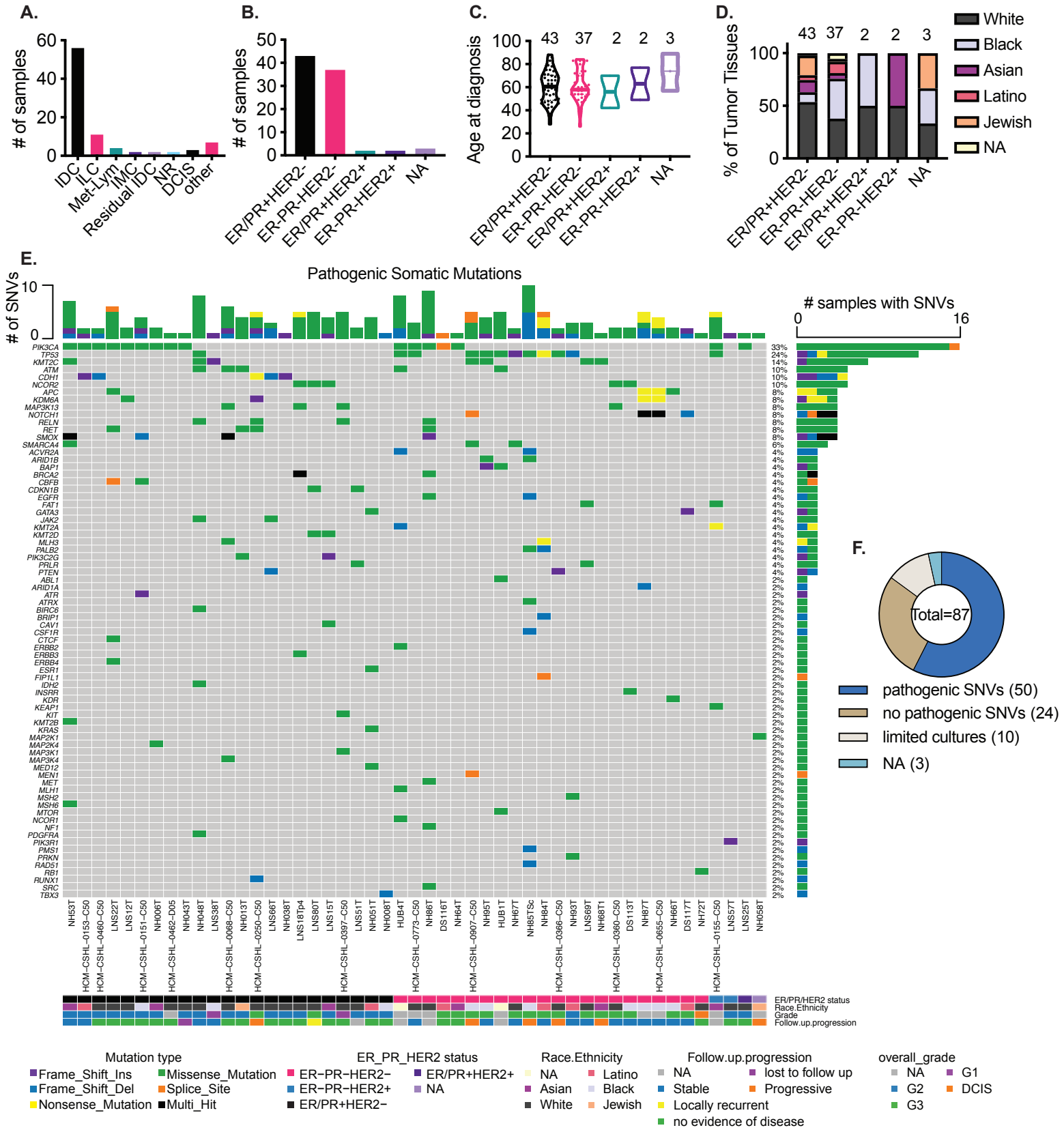

A.
# of samples
IDC
ILC
Met-Lym
IMC
Residual IDC
NR
DCIS
other
B.
# of samples
ER/PR+HER2-
ER-PR-HER2-
ER/PR+HER2+
ER-PR-HER2+
NA
C.
Age at diagnosis
43 37 2 2 3
D.
% of Tumor Tissues
White
Black
Asian
Latino
Jewish
NA
E.
Pathogenic Somatic Mutations
# of SNVs
samples with SNVs
ER/PR/HER2 status
Race.Ethnicity
Grade
Follow.up.progression
F.
Total=87
pathogenic SNVs (50)
no pathogenic SNVs (24)
limited cultures (10)
NA (3)
Mutation type
Frame_Shift_Ins
Frame_Shift_Del
Nonsense_Mutation
Missense_Mutation
Splice_Site
Multi_Hit
ER_PR_HER2 status
ER-PR-HER2-
ER-PR-HER2+
ER/PR+HER2-
ER/PR+HER2+
NA
Race.Ethnicity
NA
Asian
White
Latino
Black
Jewish
Follow.up.progression
NA
Stable
Locally recurrent
no evidence of disease
lost to follow up
Progressive
overall_grade
NA
G1
G2
G3
DCIS

Figure 2

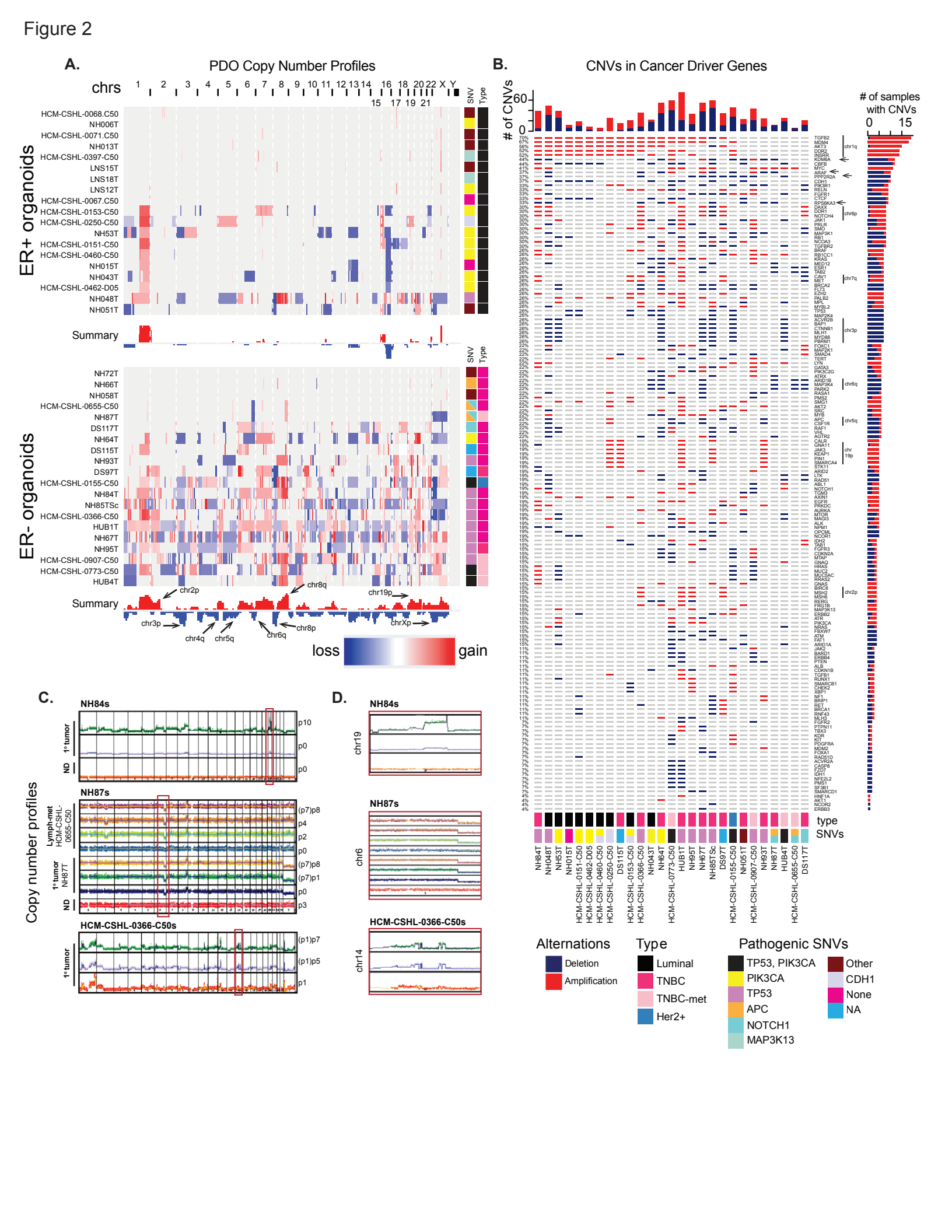

A.
PDO Copy Number Profiles
chrs
SNV
Type
ER+ organoids
ER- organoids
Summary
chr2p
chr8q
chr19p
chr3p
chr4q
chr5q
chr6q
chr8p
chrXp
loss
gain
B.
CNVs in Cancer Driver Genes
# of CNVs
# of samples with CNVs
type
SNVs
Alternations
Deletion
Amplification
Type
Luminal
TNBC
TNBC-met
Her2+
Pathogenic SNVs
TP53, PIK3CA
PIK3CA
TP53
APC
NOTCH1
MAP3K13
Other
CDH1
None
NA
C.
NH84s
NH87s
HCM-CSHL-0366-C50s
Copy number profiles
1°tumor
ND
Lymph-met HCM-CSHL-0655-C50
1°tumor NH87T
D.
chr19
chr6
chr14

Figure 3

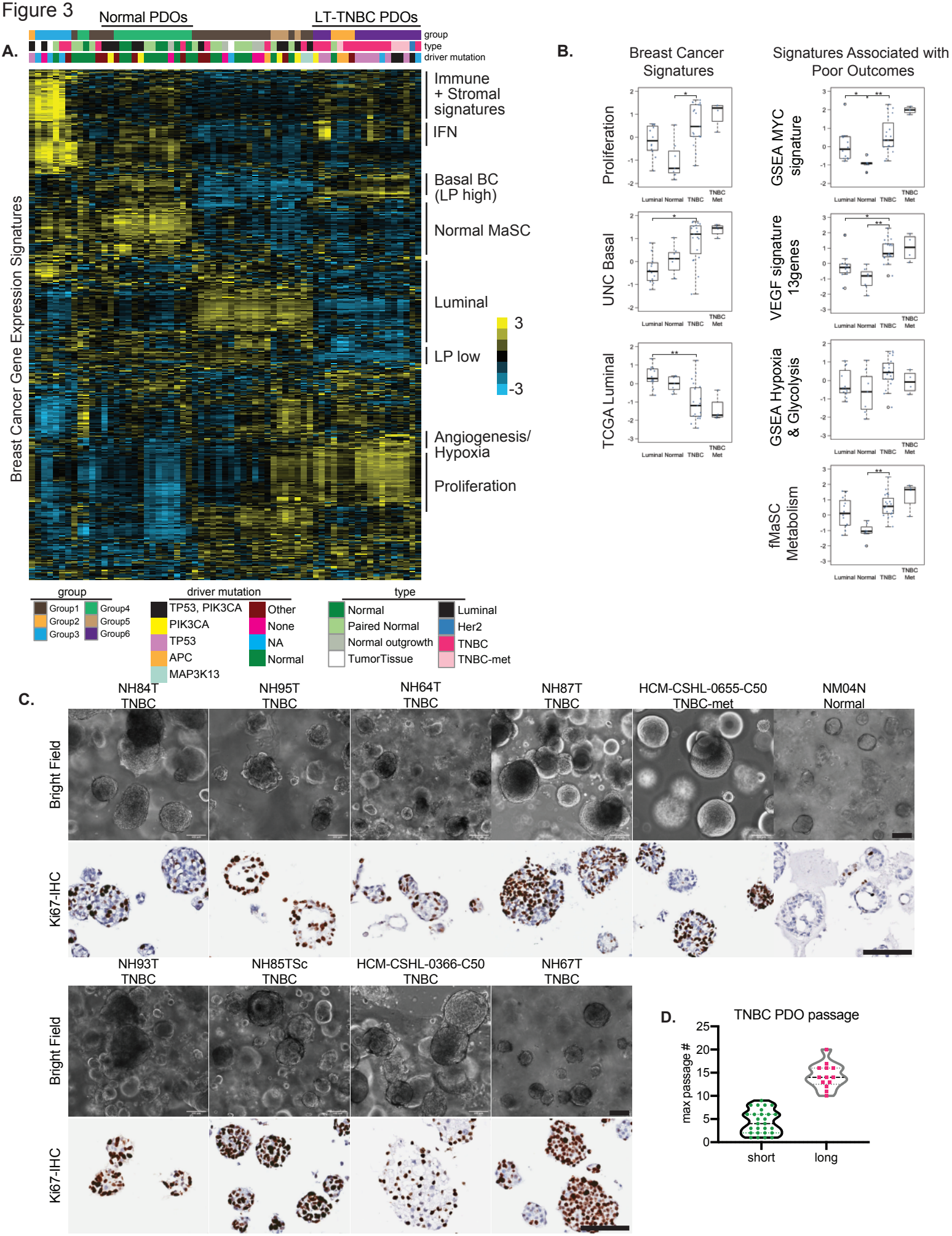

A.
Normal PDOs
LT-TNBC PDOs
group
type
driver mutation
Breast Cancer Gene Expression Signatures
Immune + Stromal signatures
IFN
Basal BC (LP high)
Normal MaSC
Luminal
LP low
Angiogenesis/ Hypoxia
Proliferation
3
-3
group
Group1
Group2
Group3
Group4
Group5
Group6
driver mutation
TP53, PIK3CA
PIK3CA
TP53
APC
MAP3K13
Other
None
NA
Normal
type
Normal
Paired Normal
Normal outgrowth
TumorTissue
Luminal
Her2
TNBC
TNBC-met
B.
Breast Cancer Signatures
Signatures Associated with Poor Outcomes
Proliferation
UNC Basal
TCGA Luminal
GSEA MYC signature
VEGF signature 13genes
GSEA Hypoxia & Glycolysis
fMaSC Metabolism
Luminal Normal TNBC TNBC Met
C.
NH84T TNBC
NH95T TNBC
NH64T TNBC
NH87T TNBC
HCM-CSHL-0655-C50 TNBC-met
NM04N Normal
Bright Field
Ki67-IHC
NH93T TNBC
NH85TSc TNBC
HCM-CSHL-0366-C50 TNBC
NH67T TNBC
Bright Field
Ki67-IHC
D.
TNBC PDO passage
max passage #
25
20
15
10
5
0
short
long

Figure 4

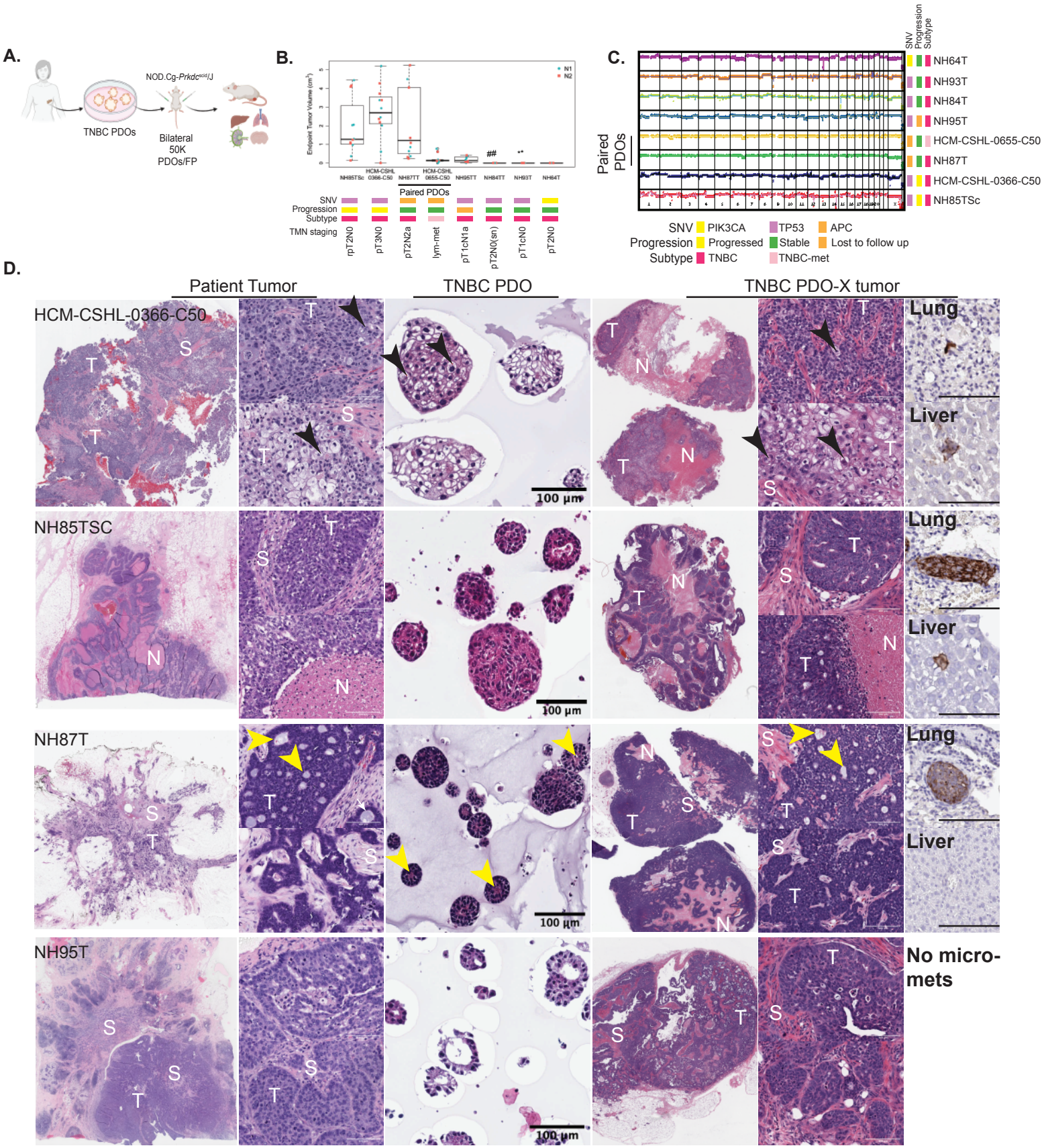

A.
TNBC PDOs
NOD.Cg-Prkdc scid/J
Bilateral
50K
PDOs/FP
B.
Endpoint Tumor Volume (cm³)
N1
N2
##
**
NH85TSc
HCM-CSHL 0366-C50
NH87TT
HCM-CSHL 0655-C50
NH95TT
NH84TT
NH93T
NH64T
Paired PDOs
SNV
Progression
Subtype
TMN staging
pT2N0
pT3N0
pT2N2a
lym-met
pT1cN1a
pT2N0(sn)
pT1cN0
pT2N0
C.
SNV
Progression
Subtype
NH64T
NH93T
NH84T
NH95T
HCM-CSHL-0655-C50
NH87T
HCM-CSHL-0366-C50
NH85TSc
Paired PDOs
SNV PIK3CA TP53 APC
Progression Progressed Stable Lost to follow up
Subtype TNBC TNBC-met
D.
Patient Tumor
TNBC PDO
TNBC PDO-X tumor
HCM-CSHL-0366-C50
NH85TSC
NH87T
NH95T
Lung
Liver
No micro-mets
100 µm

Figure 5

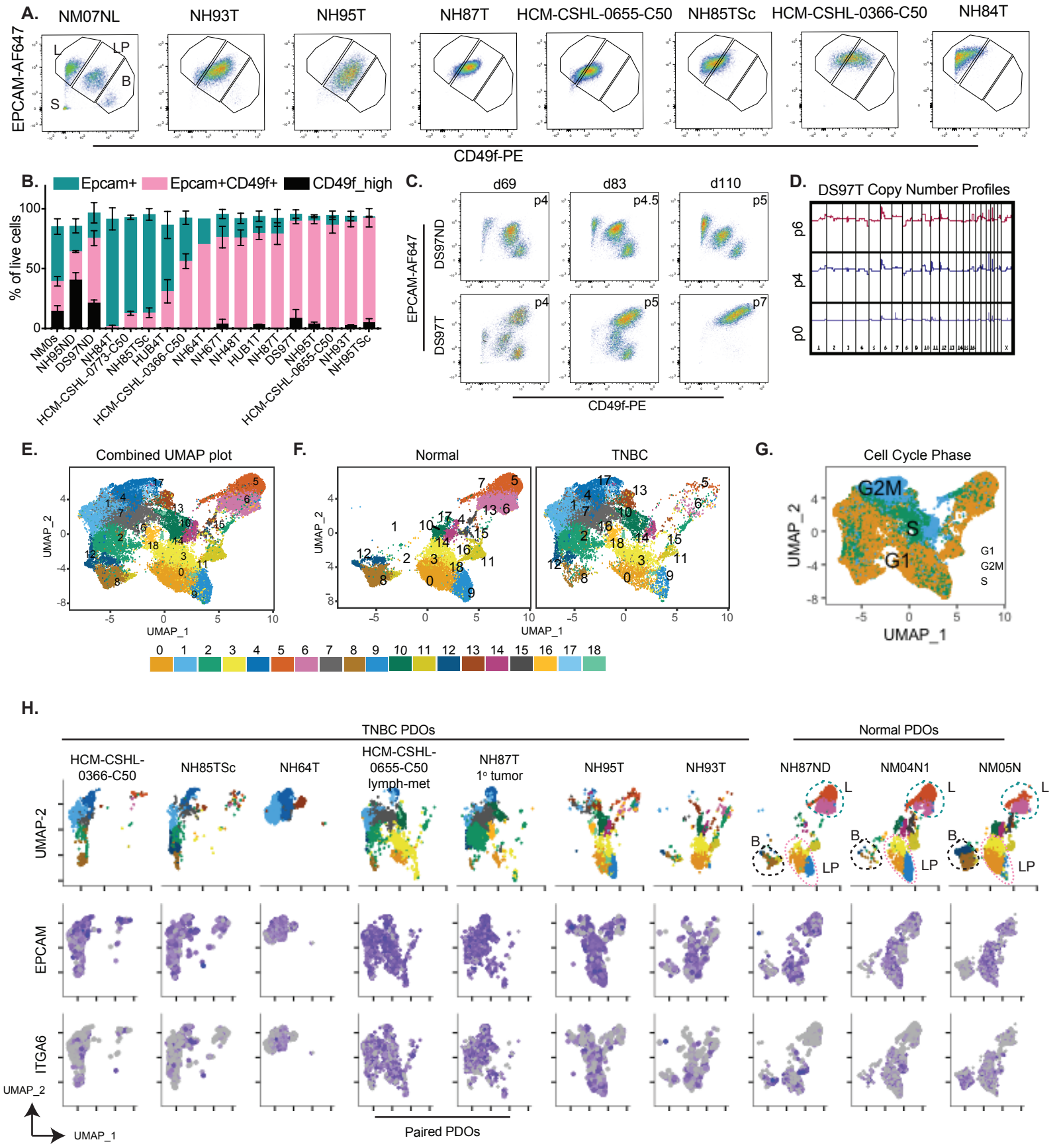

A.
NM07NL
NH93T
NH95T
NH87T
HCM-CSHL-0655-C50
NH85TSc
HCM-CSHL-0366-C50
NH84T
EPCAM-AF647
L
LP
B
S
CD49f-PE
B.
Epcam+
Epcam+CD49f+
CD49f_high
% of live cells
100
50
0
NM0s
NH95ND
DS97ND
NH84T
HCM-CSHL-0773-C50
NH85TSc
HUB4T
HCM-CSHL-0366-C50
NH64T
NH67T
NH48T
HUB1T
NH87T
DS97T
NH95T
HCM-CSHL-0655-C50
NH93T
NH95TSc
C.
d69
d83
d110
p4
p4.5
p5
DS97ND
DS97T
p4
p5
p7
EPCAM-AF647
CD49f-PE
D.
DS97T Copy Number Profiles
p6
p4
p0
E.
Combined UMAP plot
UMAP_2
UMAP_1
F.
Normal
TNBC
UMAP_2
UMAP_1
G.
Cell Cycle Phase
G2M
S
G1
G1
G2M
S
UMAP_2
UMAP_1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
H.
TNBC PDOs
Normal PDOs
HCM-CSHL-0366-C50
NH85TSc
NH64T
HCM-CSHL-0655-C50 lymph-met
NH87T 1° tumor
NH95T
NH93T
NH87ND
NM04N1
NM05N
UMAP-2
EPCAM
ITGA6
L
B
LP
UMAP_2
UMAP_1
Paired PDOs

Figure 6

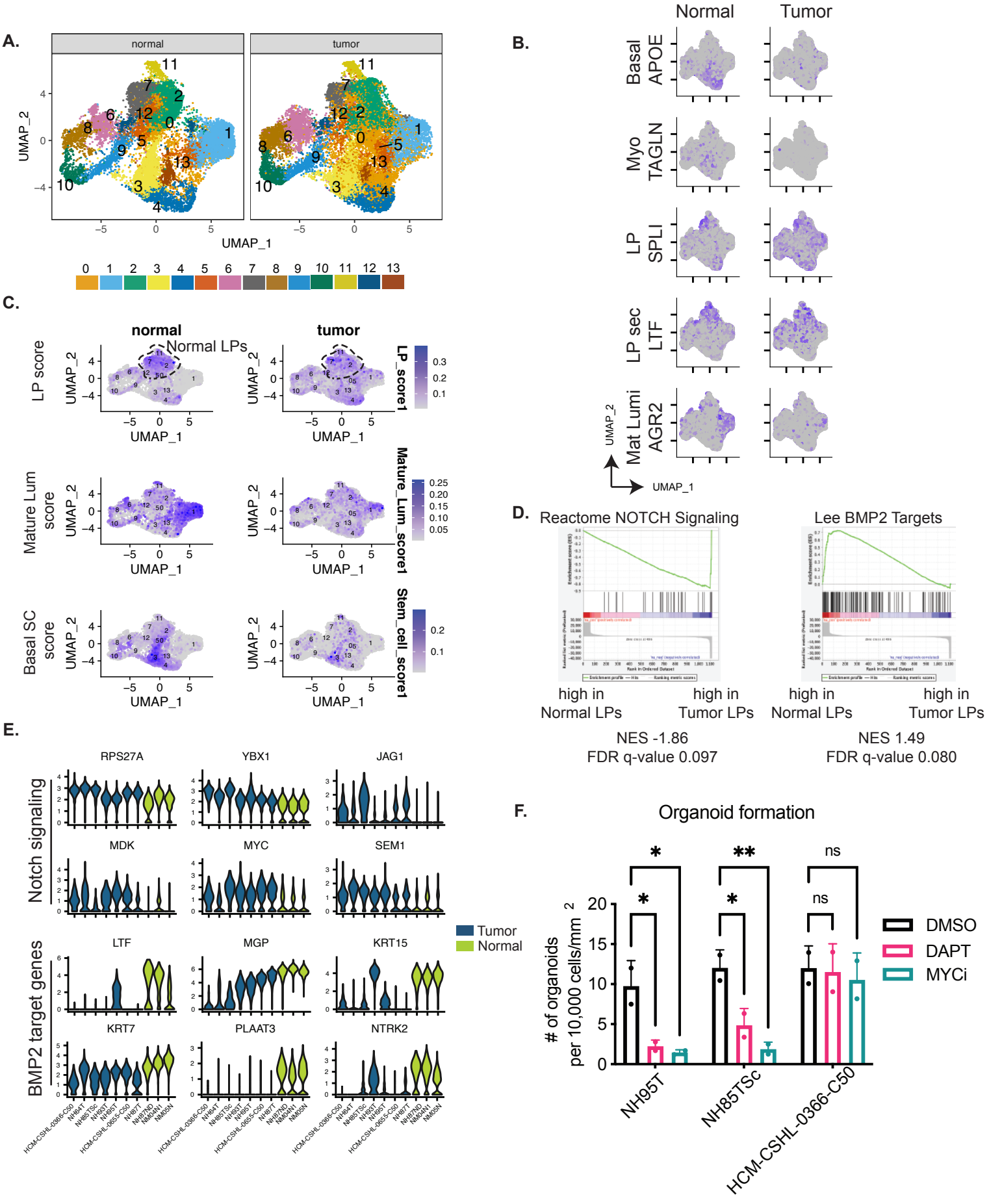
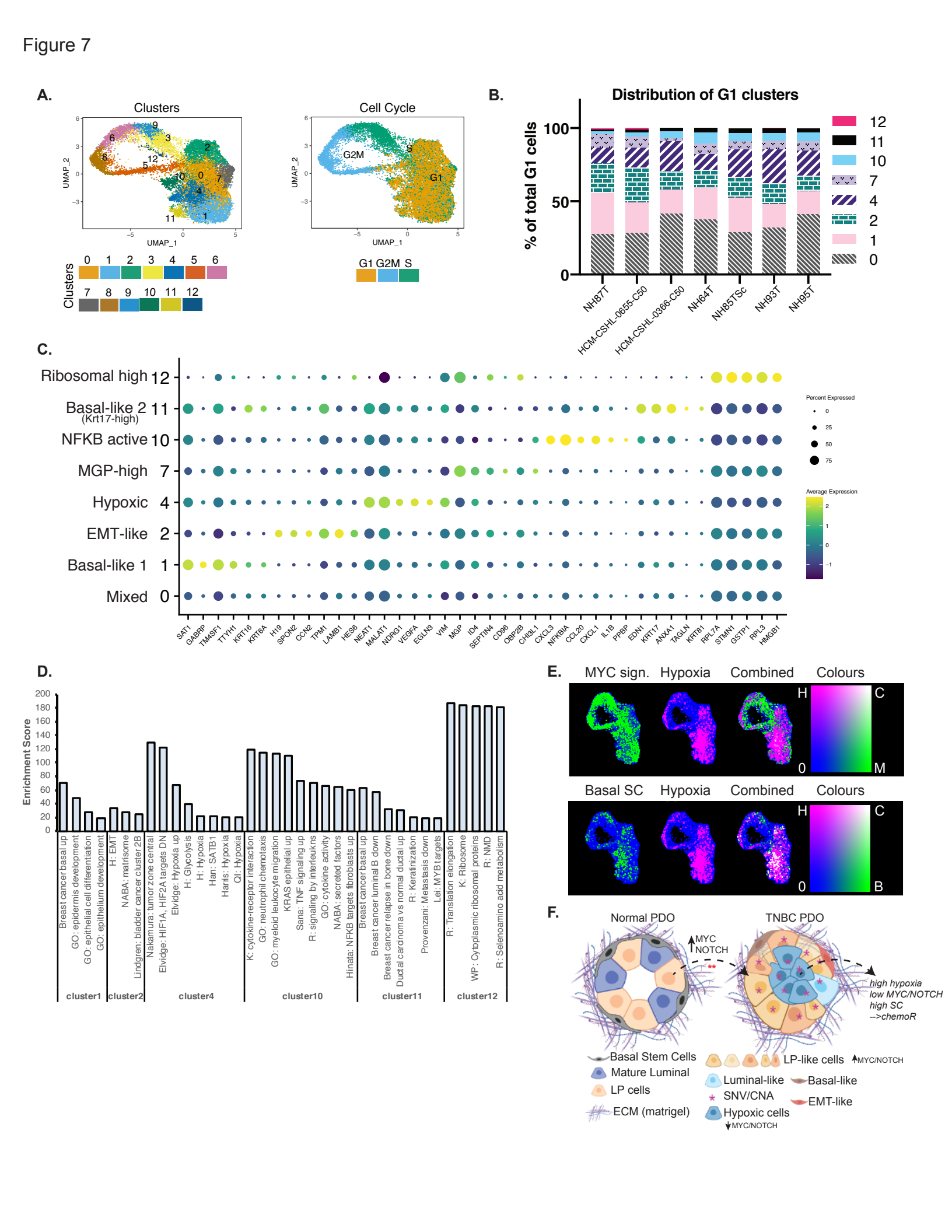

A.
normal
tumor
UMAP_2
UMAP_1
0 1 2 3 4 5 6 7 8 9 10 11 12 13
B.
Normal
Tumor
Basal APOE
Myo TAGLN
LP SPLI
LP sec LTF
Mat Lumi AGR2
UMAP_2
UMAP_1
C.
normal
tumor
Normal LPs
LP score
Mature Lum score
Basal SC score
LP_score1
Mature_Lum_score1
Stem_cell_score1
UMAP_2
UMAP_1
D.
Reactome NOTCH Signaling
Lee BMP2 Targets
high in Normal LPs
high in Tumor LPs
NES -1.86
FDR q-value 0.097
NES 1.49
FDR q-value 0.080
E.
RPS27A
YBX1
JAG1
MDK
MYC
SEM1
LTF
MGP
KRT15
KRT7
PLAAT3
NTRK2
Notch signaling
BMP2 target genes
Tumor
Normal
F.
Organoid formation
# of organoids per 10,000 cells/mm 2
DMSO
DAPT
MYCi
NH95T
NH85TSc
HCM-CSHL-0366-C50
ns

Figure 7


A.
Clusters
Cell Cycle
UMAP_1
UMAP_2
G2M
S
G1
Clusters
0 1 2 3 4 5 6
7 8 9 10 11 12
G1 G2M S
B.
Distribution of G1 clusters
% of total G1 cells
NH87T
HCM-CSHL-0655-C50
HCM-CSHL-0366-C50
NH64T
NH85TSc
NH93T
NH95T
C.
Ribosomal high 12
Basal-like 2 (Krt17-high) 11
NFKB active 10
MGP-high 7
Hypoxic 4
EMT-like 2
Basal-like 1 1
Mixed 0
Percent Expressed
Average Expression
D.
Enrichment Score
cluster1
cluster2
cluster4
cluster10
cluster11
cluster12
E.
MYC sign.
Hypoxia
Combined
Colours
Basal SC
F.
Normal PDO
TNBC PDO
MYC
NOTCH
high hypoxia
low MYC/NOTCH
high SC
-->chemoR
Basal Stem Cells
Mature Luminal
LP cells
ECM (matrigel)
LP-like cells
Luminal-like
SNV/CNA
Hypoxic cells
Basal-like
EMT-like