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], 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.
*Corresponding author: spector@cshl.edu

Summary

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 better treatment options.

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 (Parker et al., 2009; Perou et al., 2000; Sørlie et al., 2003). 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 (Prat and Perou, 2011). 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 (Herschkowitz et al., 2006; Parker et al., 2009; Prat et al., 2010; Prat and Perou, 2011). 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 (Lehmann et al., 2015, 2011). TNBCs are typically presented as high-grade carcinomas (Howlader et al., 2014) and show a significantly poor 5-year prognosis compared to other subtypes (Curtis et al., 2012; Liedtke et al., 2008; Rueda et al., 2019).

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 (Fujii and Sato, 2021; Tuveson and Clevers, 2019)). 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 (Sato et al., 2011), pancreas (Boj et al., 2015), ovary (Hoffmann et al., 2020; Kopper et al., 2019), prostate (Beshiri et al., 2018) and breast (Sachs et al., 2018). 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 (Dekkers et al., 2020; Rosenbluth et al., 2020). While cancer cell lines have been a valuable resource for providing insights into cancer biology and drug development (Barretina et al., 2012; Garnett et al., 2012; Yu et al., 2016), 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 (Curtis et al., 2012; Koboldt et al., 2012a; Nik-Zainal et al., 2016; Shah et al., 2012) 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.

Results:

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

Breast cancer tissues along with paired normal breast tissues were acquired from female patients and developed into 3-D organoids. In addition to breast cancer tissues, we also collected normal reductive mammoplasty samples from 10 cancer-free patients to generate normal PDOs for downstream analyses. The majority of tumor samples were from patients with invasive ductal carcinomas (52/83, Fig. 1A), 11/83 were from invasive lobular carcinomas, 4/83 were from metastatic lymph nodes and a small percentage from other categories (Fig 1A, Table S1). The majority of the tumor samples (43/83) were luminal BC as defined by immunohistochemistry (ER/PR+/HER2-), 33/83 were from TNBCs (ER-PR-HER2-), 2/83 were from ER/PR+/HER2+ BCs, 2/83 were from ER-/PR-/HER2+ subtype, and 3/83 samples belonged to post-treatment residual tissue with no visible carcinoma (NA) (Fig 1B). Samples were collected from patients of various age groups (Fig 1C) and diverse ethnic and racial backgrounds (Fig 1D). An increased proportion of TNBC tumor samples was observed from black patients with African or west-Indian heritage (Fig 1D, Table S1) which recapitulates the higher incidence of TNBCs in the African-American community (Perou, 2011; Prakash et al., 2020). Using a DNA-seq panel of 143 cancer driver genes, we identified pathogenic single nucleotide variants (SNVs) in 45 tumor organoid lines (Fig 1E); *PIK3CA* was mutated in 33% of the tumor organoids, the majority of which were from patients with luminal breast cancer. 24% of the organoids (50% of TNBC PDOs 9/18) had a pathogenic *TP53* mutation and primarily represented the high grade TNBC cohort (Fig 1E). Two of the organoids are derived from

a patient with rare breast cancer showing adenoid cystic carcinoma (AdCC) like morphology, NH87T (primary tumor) and HCM-CSHL-0655-C50 (lymph node metastasis from the same patient) (Table S1). The AdCC organoids show a non-traditional TNBC mutation profile with pathogenic mutations in *APC*, *KDM6A* and *NOTCH1* (Fig 1E), which were previously observed in AdCCs of the breast (Fusco et al., 2016). Mutations in *KMT2C* and *GATA3* were also observed in a variety of luminal breast cancer organoids, *CDH1* mutations were present in 11% of PDOs all of which belonged to invasive lobular carcinomas, and *ARID1B* mutations were found in a few TNBC-derived organoids (Fig 1E). Overall, the mutation profiles of patient-derived BC organoids are in concordance with the mutational landscape of BCs (Curtis et al., 2012; Koboldt et al., 2012a). A relatively large subset of tissues (24/83) that resulted in cultured organoids did not have pathogenic SNVs (Fig 1F, Table S1), while 10/83 samples dropped off in culture at early passages (Fig 1F). The PDOs with pathogenic SNVs showed a range of growth properties *in vitro* with subsets exhibiting long-term continued expansion while others seem to have more limited cultures either in terms of maximum passages reached or limited expansion abilities (Table S1).

A comparison of the SNV profile between several pairs of patient tumor and PDOs showed high concordance of the pathogenic SNVs between the tumor and tumor derived organoids (Fig S1A), except in some cases where the SNVs are present at a much lower variant allele frequency (VAF) in the tumor, for instance, NH84TT-P0 showed <25% TP53 VAF compared to 100% VAF in NH84TT-p10 PDO (Fig S1A) suggesting normal contamination in the tumor tissues and a successful outgrowth of cancer cells in organoid cultures. Longitudinal analyses for some patient-derived organoids showed the stability

of primary driver mutations overtime in culture (Fig S1B). To determine whether the amount of tumor material was a limiting step for generation of organoids, we collected tumor scrapings (see methods) from a subset of tumor samples. A comparison of tumor tissue (labeled TT) and the scrapings (labeled Sc) from three different patients (Fig S1C) showed high concordance and indicates that, if necessary, small amounts of tumor material can be used to generate organoids. We also examined the pathogenic SNVs in TNBC tissue samples (p0) that did not result in a successful generation of organoids (establishment or long-term cultures) and 6/10 had TP53 mutations and some samples also had BRCA1/2 mutations (Fig S1D), the latter were largely absent in successful organoid cultures (Fig 1E).

Breast cancer organoids are enriched for copy number alterations

Having identified pathogenic driver mutations for the tumor-derived organoids, we performed copy number analysis using SMASH (short multiply aggregated sequence homologies) (Wang et al., 2016) for various PDOs. TP53 mutated TNBC PDOs make up the majority of the ER- subset and are highly genomically altered compared to the luminal ER+ organoids (Fig 2A, B, Table S2). Some organoids that were deemed to be derived from tumor based on pathogenic SNVs (eg. LNS12T, LNS18T, NH06T etc.) showed no prominent CNAs. Some TNBC PDOs that showed trace SNVs such as NH58T, NH72T and NH66T also showed minimal CNAs (Fig 1E and 2A). In keeping with large scale genomic datasets (Curtis et al., 2012; Koboldt et al., 2012a), a consistent gain of chromosome 1q (chr1q) and loss of chr16q was observed in the luminal organoids where CNAs were detected (Fig 2A). The TP53 mutated TNBC organoids were highly

genomically aberrant where multiple lines showed a gain of chromosomes (chrs) 1q, 8q, 19q and chrs 7, 20 and 21; loss of chrs 3p, 4, 5, 17p and Xp was also observed in multiple organoid lines (Fig 2A). These TNBC PDOs also showed copy number loss of chr4q and chr5q (Fig 2A) that have been previously reported to be over-represented in basal-like breast cancers (Weigman et al., 2012). Of note, the AdCC-like organoids, NH87T (primary tumor) and HCM-CSHL-0655-C50 (lymph node metastasis from the same patient), were considerably less genomically aberrant as compared to the other TNBC organoids and showed focal amplification of chr8q and deletion of chr6p (Fig 2A, C-D). Interestingly, the chrX deletion was only observed in the primary tumor (NH87T) sample (Fig 2A, C) and a focal deletion was observed in chr2 of the lymph met HCM-CSHL-0655-C50 (Fig 2C). The AdCC organoid CNA profiles are consistent with the low CNA profiles that are typically observed in AdCC-like breast cancers (Fusco et al., 2016).

For a subset of these samples, we compared the copy number profiles of the primary tumor with the paired tumor and normal organoids and observed more pronounced CNAs in the organoids compared to the tumors suggesting that the organoid culture enriches for tumor cells (Fig 2C-D). As observed in the SNV data (Fig S1) the NH84T tumor (p0 with <25% VAF of pathogenic TP53) was likely very heterogenous and contained a significant population of normal cells, however, the tumor cells successfully outgrew the normal and resulted in a highly pure tumor organoid culture over multiple passages (Fig 2C-D). For HCM-CSHL-0366-C50 and NH87T the early passage p1 or tumor tissue (p0) respectively, had relatively less normal contamination and maintained their copy number profiles over time in culture (Fig 2C-D). While copy number profiles

were enriched for in successful tumor PDOs, for some cultures we observed loss of copy number alterations in the organoids over-time, suggesting a normal outgrowth (Fig S1E).

Since tumor organoids are more enriched for tumor cells compared to the p0 tumor tissue, we profiled the putative cancer driver gene panel in individual organoid samples. TNBC PDOs typically had a higher copy number alteration frequency of these cancer driver genes (top bar-graph in Fig 2B). *TGFB2*, *MDM3*, *AKT3*, *DDR2* and *INSRR* showed the highest frequency of alteration (Fig 2B) and are all genes present on chr1q that is amplified in both luminal and TNBC PDOs (Fig 2A-B). Interestingly, we also found a higher frequency of deletion of *KDM6A*, *ARAF* and *RPS6KA3* tumor suppressors, located on chrXp, specifically in the TNBC-derived organoids (arrows in Fig 2B). Loss of *PPP2R2A* (chr8p), which was previously reported to be a tumor suppressor in breast cancer (Curtis et al., 2012), was also identified in 38% of the samples, the majority of which are of TNBC subtype (Fig 2B). Furthermore, there is an over-representation of loss of chr3p in the TNBC samples, that results in the deletion of potential tumor suppressors: *ACVR2B*, *BAP1*, *CTNNB1*, *MLH1*, *MYF88* and *PBRM1* (Fig 2B). Copy number loss of chr6q is also common in TNBC organoids and results in loss of *ARID1B*, *MAP3K4*, and *PARK2*, a master regulator of G1/S cyclins (Gong et al., 2014). *ESR1*, the gene encoding estrogen receptor (ER), present on chr6q is also frequently lost in these TNBC organoids (Fig 2A-B). We also identified gains of *CALR*, *JAK3*, *KEAP1*, *PIN1* and *SMARCA4* that are associated the amplification of chr19p and amplification of mismatch repair (MMR) genes *MSH2* and *MSH6* located on chr2p (Fig 2A-B) in subsets of TNBC PDOs. While MMR genes are commonly mutated in various cancers, their overexpression was recently associated with aggressive prostate cancers (Wilczak et al., 2017) and was shown to

promote genomic instability in yeast (Chakraborty et al., 2018) but its role is yet to be determined in TNBC disease progression.

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

TNBC organoids recapitulate the transcriptome of basal-like breast cancers

Next, we performed RNA-seq on various tumor and normal-derived organoids to profile their transcriptomes (Table S3). Hierarchical clustering of samples by euclidean distance divided them into six groups (Fig 3A). We used supervised clustering using 838 previously curated gene expression signatures (Fan et al., 2011; Garcia-Recio et al., 2020; Gatza et al., 2014) (Table S4) and found similar patterns of clustering among the various organoid lines (Fig 3B). Group 4 is largely comprised of the true normal organoids (labeled with prefix NM) derived from breast tissues obtained from individuals undergoing reductive mammoplasty (Normal in Fig 3A-B), paired normal PDOs derived from normal tissue adjacent or distal to the tumor and very few tumor-derived organoids; this cluster showed an enrichment of normal mammary stem cell (MaSC) signatures and low

proliferation related gene expression profiles (Fig 3B-C). Group 1 is comprised of a mixture of either paired normal organoids or tumor organoids that did not have a strong driver mutation (eg. *PIK3CA* or *TP53*) (Fig 3B, 1E). This subgroup had a signature profile that was distinct from the true normal group (Group 4) and is most similar to luminal cell signatures (Fig 3B). The samples in this group also have a higher proliferation signature compared to the true normal Group 4 (Fig 3B). Group 5 mostly contained PDOs derived from Luminal BC and showed luminal like gene-expression signatures (Fig 3A-C). Signature profiling also uncovered immune and stromal signatures that highly corresponded to samples in Group 3 (Fig 3A-B). Interestingly, two of the samples in this group NH63T and NH54T exhibited limited propagation in culture due to stromal outgrowth leading us to hypothesize that these samples had some fibroblast-like cells. We also found a strong interferon (IFN) signature in this group along with some additional PDOs from Group 6 (Fig 3B, S2B).

Groups 2 and 6 are comprised mostly of TNBC organoids and all organoid lines in these groups could be propagated to long-term cultures (>passage10, labeled LT-TNBCs) and showed continued expansion (Fig 3A-B). The only exception is the NH48N normal sample, which was confirmed to be mostly tumor by copy number analysis and identical to its counterpart NH48T (Fig S2A). The most prominent signatures of Groups 2 and 6 correspond to basal-like breast cancer gene-sets that are also defined by luminal progenitor (LP) like signatures (Fig 3B-C, S2B-C). LP-like gene expression has previously been shown to be associated with basal-like breast cancers (Lim et al., 2009). The organoids in this group also showed enrichment of proliferation signatures (Fig 3B-C) and the majority had a *TP53* mutation which was in conjunction with the *PIK3CA* mutations

for three samples (HCM-CSHL-0773-C50, HUB4T and HCM-CSHL-0155-C50) (Fig 3A). We also observed a MYC amplification signature in this subgroup (Fig 3D) that was accompanied by the copy number amplification of *cMYC* in many of these PDOs (Fig S2D). Additionally, gene sets associated with hypoxia, glycolysis, angiogenesis and fetal MaSC (fMaSC) metabolism signatures were also enriched in LT-TNBCs (Fig 3B-C, S2B). The *VEGF* 13-gene signature showed high correlation with basal-like breast cancers (Perou, 2011), has prognostic significance and is associated with poor outcome in breast and other cancers (Hu et al., 2009). Similarly, fMaSC metabolism signature is a refined 8-gene signature, which primarily comprises genes associated with glycolysis and fatty acid metabolism, and was previously shown to be associated with TNBCs and metastatic TNBC lesions (Giraddi et al., 2018). This suggests that the PDOs with these signatures are likely associated with tumors linked to poor outcomes.

Of note, none of the TNBCs propagated in culture showed expression profiles of the more mesenchymal-like claudin-low subgroup (Prat et al., 2010), possibly because the culture conditions are more favorable towards the propagation of epithelial cells. Overall, the gene expression signatures of the normal and tumor-derived organoids depict the expression profiles of luminal and basal-like breast cancers. The TNBC organoids that can be propagated to long-term cultures (LT-TNBCs), Groups 2 and 6, had very classic basal-like breast cancer signatures and were associated with proliferation, hypoxia and *c-MYC* amplification gene expression signatures (Fig 3B-D, S2C-D). Furthermore, about 40% of basal-breast cancers exhibit *c-MYC* amplification (Koboldt et al., 2012a) and the majority of our LT-TNBC organoids show an enrichment of *c-MYC*

(Fig S2D) signature suggesting that our organoid system results in long-term expansion of TNBCs that are *c-MYC* driven and have a LP-like basal breast cancer signature.

TNBC organoids exhibit heterogenous growth properties in culture

In order to understand the complexity of organoid cultures we profiled the growth properties of various TNBC and normal organoids. TNBC organoids appear as densely filled balls of cells which is in stark contrast to the normal organoids that have a predominant acinar structure with a central lumen and some organoids that were filled (NM04N in Fig 4A). The TNBC-organoids showed varying degrees of propagation in culture where some lines could be propagated for over 10 passages (LT-TNBCs) with continued expansion while many short-term culture lines dropped out of culture at various points (Fig 4B, S3A) and had limited material to assay. Low starting material, limited proliferation, normal and stromal outgrowths were among the primary reasons for short term culture growth (Fig S3A). We focused on the LT-TNBCs for all downstream analyses. IHC for the proliferation marker, Ki-67, in confluent PDO cultures showed large variability of expression amongst the various organoid lines and also within each culture (Fig 4A- lower panels, C). The LT-TNBC organoids were highly proliferative compared to normal and luminal organoids (Fig 4A, C). There was a lot of heterogeneity within organoid cultures with some organoids being mostly comprised of proliferating cells, for instance NH85TSc and NH95T, while other cultures were more mixed (Fig 4C). Normal-derived organoids typically had <10% of proliferating cells per organoid and showed luminal-basal organization with a hollow lumen, while TNBC-derived organoids tend to be highly proliferative and undifferentiated with little to no observed cellular organization (Fig

4A). When seeded as single cells, we observed a range of organoid formation amongst the various lines (Fig 4D) which did not always correlate with the proliferative index of the PDOs.

TNBC organoids can recapitulate tumor morphology *in-vivo*

Next, we examined the *in-vivo* tumor forming abilities of these patient-derived cultures. We transplanted organoids from eight different TNBC (LT-TNBCs) lines into the mammary-fat pads of NOD-SCID mice and assayed their tumor formation and metastatic potential over time (Fig 5A). We found striking differences in the tumor forming ability between the different organoid lines (Fig 5B) despite having highly aberrant genomic profiles (Fig 5C). TNBC organoids HCM-CSHL-0366-C50, NH85TSc and NH87T resulted in palpable masses around 30 days and steady tumor formation in 100% of injection sites (Fig 5B, S4A). NH85TSc-derived tumors grew rapidly and resulted in visibly necrotic masses that needed to be resected at ~100 days. HCM-CSHL-0366-C50 resulted in primary tumors that also developed fluid filled cysts on top of the tumors that drained on their own. NH87T and HCM-CSHL-0655-C50 are the AdCC-like primary tumor and lymph-node metastasis from the same patient. Interestingly, while NH87T formed tumors rapidly, HCM-CSHL-0655-C50 did not and the tumors remained small (Fig 5B, S4A), despite both of these lines having equivalent growth properties *in vitro* (Fig 4). This observation was consistent for the two independent transplant experiments done with different passage organoids (Fig S4A). NH95T resulted in small tumor masses at multiple injection sites that were slow to grow while NH84T and NH93T had small tumors observable only in fat-pad histology sections of some sites and NH64T did not result in

any primary tumors (Fig 5B, S4A-B) despite being highly genomically aberrant (Figs 5C, 1-2) and having high proliferation and organoid formation rates *in-vitro* (Fig 4C-D).

Remarkably, the xenograft tumors resulting from HCM-CSHL-0366-C50, NH85TSc, NH87T and NH95T organoids recapitulated the morphology of the patient tumor despite previously being in culture for up to 18 passages. The tumor and the organoid-xenograft from patient HCM-CSHL-0366-C50 were the most distinct and showed squamous differentiation with pleomorphic cells that were variable in size and had cells with either oval or pale nuclei with prominent nucleoli (Fig 5C). We also observed cells that had thin spinous connections with adjacent tumor cells which were also present in the organoids (black arrows Fig 5C). NH85TSc tumor and organoid-xenografts both showed highly undifferentiated morphology with large necrotic areas and stromal compartments infiltrating in between the tumors (Fig 5C). Lastly, NH87T tumor was a TNBC type with AdCC like features that are characterized by cribriform architectural patterns and pseudo-lumens (Fusco et al., 2016; Marchiò et al., 2010). While the organoid-derived xenograft had a lower stromal composition compared to the patient tumor it still recapitulated the key features of the original tumor, including the presence of cuboidal cells, cribriform architecture and pseudo-lumens (yellow arrows Fig 5C). While NH95T formed tumors at a much lower efficiency (Fig 5A, B), the tumors that did form had morphological similarities to the original patient tumor (Fig 5C). The patient tumor cells showed an organization pattern which was recapitulated in the organoid-derived xenografts and to some extent in the NH95T organoids (Fig 5C).

Of note, patients from whom the fast-growing organoids were derived, i.e., HCM-CSHL-0366-C50, NH85 and NH87, presented with poor outcomes and diagnosis (TMN

staging in Fig 5B). Patient HCM-CSHL-0366-C50 had rapid metastasis to the brain and succumbed from the disease, patient NH85 had local recurrence in the lymph node and patient NH87 presented with a positive lymph node at initial diagnosis (Table S1). Using IHC with a human mitochondrial antibody we only observed micro-metastasis and single-cell metastasis in our experiments (Fig S4A, C), however, altering the experimental conditions such as: prolonging the end-point, using a more immunocompromised NOD/SCID gamma (NSG) mouse model or tail-vein injections might result in more metastatic lesions and will be examined in future studies. Thus far, our data shows that TNBC PDOs recapitulate the tumor intrinsic properties of the original tumors at genomic, transcriptomic and morphological levels.

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

To fully assess the utility of PDOs as cancer models, we next asked what were the cell types represented within these organoids and how did they differ from the normal derived PDOs. In order to profile the cell-types present within the normal and TNBC organoids, we used a combination of flow cytometry and single cell RNA-seq (scRNA-seq). We assayed for mammary epithelial lineages using EPCAM and CD49f as luminal and basal cell markers respectively (Dekkers et al., 2020; Lim et al., 2009; Raouf et al., 2008; Rosenbluth et al., 2020). As previously shown (Rosenbluth et al., 2020), the normal derived organoids recapitulate the EPCAM+ luminal lineage, EPCAM+/CD49f+ luminal progenitor cells and EPCAM-/CD49f+ basal cell lineages (NM07NL in Fig 6A). The populations slightly fluctuated between the different normal lines and passages but overall, the different lineages were observed in all normal organoids (Fig 6A-C, S5A,

Table S6). TNBC organoids were less heterogeneous in the expression of these markers, and as observed by RNA-seq signatures (Fig 3B, S2C), showed an enrichment of luminal progenitor-like cells (Fig 6A-B) which had co-expression of EPCAM and CD49f but showed slightly different and patient-specific expression patterns of these markers (Fig 6A, S5A). TNBC organoids also showed a higher gene expression of EPCAM and CD49f (gene *ITGA6*) compared to the normal organoids (Fig S5B). Some TNBC PDOs, for example NH84T, had a more luminal cell flow-cytometry profile, despite clustering with the other LT-TNBCs at the transcriptome level. Interestingly, metastatic TNBC lines HCM-CSHL-0773-C50 and HUB4T also had an enrichment of more luminal-like cells (Fig 6B, S5D). NH85TSc which was derived from a patient that later relapsed and HCM-CSHL-0366-C50, which showed rapid progression in the patient, also had a more luminal cell profile with higher percentage of EPCAM only cells (Fig 6A-B) despite showing gene signatures of the basal-like BCs. The EPCAM/CD49f profile of the individual lines was stable overtime at different passages (Fig S5C). Additionally, while always observed in the normal-organoids, we rarely observed CD49f only cells in the TNBC-organoids (Fig 6A-B, S5A, D). We performed longitudinal analysis for one of the patient-derived organoid lines and observed an enrichment of the tumor LPs in culture over time in early passages (Fig 6C) which coincided with a more pronounced copy number profile of the TNBC organoids (Fig 6D) suggesting LP-like cells being the predominant cancer cell population in TNBCs and potentially the cell of origin of these cancers.

In order to better understand the cellular composition of these organoids, we performed scRNA-seq analysis on three normal and seven TNBC organoid lines. The study was done in three experimental batches (Fig S5E) and processed where each of

the samples was quality controlled and filtered to remove cells with high mitochondrial gene content and low gene identification (i.e. dead cells). The filtered matrix was SC-transformed using Seurat (Stuart et al., 2019) and the samples were integrated to account for the different batches (Fig 6E). Clustering of data identified 19 clusters (Fig 6E) with some clusters being highly representative of the normal organoids (clusters 0, 3, 5, 6, 9, 11 in Fig 6F) while other clusters were largely comprised of tumor cells. Tumor organoids were distinct from normal, with some overlap in clusters 0, 3 and 9 (Fig 6F-H). Expression profiles of *EPCAM* and *ITGA6* (gene encoding CD49f) (Fig 6H) recapitulated the data from flow cytometry (Fig 6A-B) where tumor organoids are predominantly of the luminal progenitor nature as measured by co-expression of *EPCAM* and *ITGA6* (CD49f) while the normal organoids recapitulate the three broad mammary epithelial lineages: mature luminal (EPCAM+), luminal progenitors (LPs, EPCAM+/CD49f+) and myoepithelial/basal-like cells (Epcam-low/CD49f+). Unlike the normal organoids, all tumor cells were *EPCAM*+ and a subset of those had low *ITGA6* (Fig 6H). In concordance with the flow cytometry data (Fig 6A), the *ITGA6*-low cells were largely present in HCM-CSHL-0366-C50 and NH85TSc. NH93T and the NH95T organoid line had similar profiles and occupied similar space as the normal LP cells (Fig 6H). Lines NH87T and HCM-CSHL-0655-C50 are paired primary and lymph node metastasis samples from the same patient and occupied very similar spaces with some overlap with normal LP cells (Fig 6H), while NH64T, NH85TSc and HCM-CSHL-0366-C50 are very distinct from the normal lines. Interestingly, there was a significant overlap between the clusters identified between NH85TSc and HCM-CSHL-0366-C50 samples (Fig 6H) in keeping with their transcriptome similarity (Fig 3A) and rapid tumor progression *in vivo* (Fig 5A).

This data builds on previous studies that have correlated an LP-like expression signature with basal-like breast cancers (Lim et al., 2009) and have shown that LP cells are the cell of origin of BRCA-mutated basal breast cancers (Molyneux et al., 2010). We suggest that LP-like cells are the possible cell of origin for a broader subset of TNBCs. Interestingly, a higher percent of EPCAM+ only cells seems to be associated with a greater degree of disease progression, however, this needs to be further investigated.

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

Since TNBC organoids had a large proportion of LP-like cells that seemed distinct from the normal LP cells (Figs 6, S5) we performed integrated single cell analysis by SC-transforming individual samples and performing an anchor-dependent integration for all individual samples. We identified thirteen clusters between the tumor and normal cells (Fig 7A) out of which clusters 6, 8, 10 and 9 represented cell cycle clusters (Fig S6A) with cluster 6 representing a population of G1-S phase cells, cluster 8 representing S-phase cells, cluster 9 S-G2M transition cells and cluster 10 marking G2M cells while the remaining clusters represented G1 cells (Fig S6A-C). The cell clusters identified represented biologically meaningful cell-types recently annotated by single cell sequencing of adult human breast epithelium (Nguyen et al., 2018). Cell types identified included: XBP1, AGR2 expressing mature luminal cells (Mat lum, cluster 1), APOE, KRT6A expressing basal stem-like cells (basal SC, cluster 4), TAGLN, TIMP1 expressing myoepithelial cells (Myo), SLPI expressing luminal progenitor cells (LP, cluster 7) and LTF expressing secretory cells (LP sec, cluster 2) (Fig 7B, S6C). To further validate the

lineage specification in organoids we overlaid the published mammary epithelial gene-sets (Lim et al., 2009) and computed a score for each of the three mammary epithelial cells. The normal cell clusters showed a much stronger cell type enrichment score while the tumor cells despite being predominantly EPCAM+/CD49f+ showed a more diffused enrichment of these signatures, suggesting some cell type heterogeneity within these tumor organoids (Fig 7C). In normal organoids, Clusters 2,7 and 11 had a higher enrichment for luminal progenitor (LP) cell score, while cluster 1 showed an enrichment for mature luminal (mature Lum) cells and cluster 3 was predominantly of the basal mammary stem cell (Basal SC) compartment (Fig 7C, S6C). The tumor organoids showed varied expression profiles of various luminal/basal markers (Fig S6D) and a diffused cluster specific enrichment of lineage scores (Fig 7C, S6E), suggesting that while they have lost proper cell type specification, the clusters identified using this integration approach have some similarity to the normal cell lineages.

With the aim to identify mechanisms that underlie the tumor luminal progenitor (LP) cells in TNBCs, we performed a differential expression analysis between tumor and normal cells of clusters 2, 7 and 11. We identified 1103 significantly differentially expressed genes (p_val_adj<0.05) between the tumor LPs vs normal LPs (Table S7). GSEA on the differentially expressed genes showed an enrichment of NOTCH signaling related genes in the tumor LPs versus an enrichment of BMP2 targets in the normal LPs (Fig 7D). Leading edge genes from these gene sets showed a consistent downregulation of NOTCH signaling related genes, *RPS27A*, *YBX1*, *JAG1*, *MDK*, *MYC* and *SEM1*, and an upregulation of BMP2 target genes, *LTF*, *MGP*, *KRT16*, *KRT7*, *PLAAT3* and *NTRK2* in normal LPs from all three patients (Fig 7E). MDK, JAG1, YBX1 are ligands of NOTCH1,

while *MYC* and *RSP27A* are downstream targets. While NOTCH activity was present in normal LP cells, it was more pronounced in the TNBC organoids (Fig S6F).

A further investigation into differentially regulated pathways showed upregulation of genes involved in ribosomal biogenesis and translation in the tumor LPs. We performed motif enrichment analysis on the differentially expressed LP genes and found a significant enrichment of MYC binding sites in the tumor LP expressed genes (Fig S6G). We used published MYC signatures to assess MYC activity in tumor and normal organoids and found a higher ubiquitous enrichment of MYC activation in tumor organoid cells, while in normal organoids it seemed to be higher in the proliferating cell clusters (Fig S6F). Similarly, while the NOTCH pathway was hyper-active in tumor LPs it was ubiquitously active across all tumor cells (Fig S6F). We tested whether NOTCH and MYC activation was required for organoid formation and seeded NH95T, NH85TSc and HCM-CSHL-0366-C50 as single cells in regular organoid growth medium in the presence of DAPT (NOTCH inhibitor) and MYCi975 (MYC inhibitor) (Han et al., 2019) and allowed for organoids to form for 12 days. We saw a significant reduction in organoid formation in the presence of both MYC and NOTCH inhibitors for NH85TSc and NH95T but not for HCM-CSHL-0366-C50 (Fig 7F). While HCM-CSHL-0366-C50 did not show reduction in the number of organoids formed, we did observe a significant difference in organoid size with DAPT and MYCi975 suggesting growth defects on inhibition of these pathways in all lines (Fig S6H).

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

formation of some, but not all, TNBC organoids and when perturbed result in proliferation defects in all lines examined.

TNBC organoids are comprised of heterogenous cancer cell populations

Since normal organoids have clearly defined lineages, the clustering of the integrated dataset (Fig 7) was largely driven by the normal organoids. To assess the heterogeneity that exists within TNBC organoid cultures we performed a similar integrated analysis on the TNBC organoids only and identified 12 different cell clusters (Fig 8A, S7A). Clusters 3,9,6,8 and 5 correspond to the different cell cycle phases (Fig 8A) and the remaining clusters are G1 cells (Fig 8A-B). We identified the markers that uniquely define each of the clusters (Fig 8C, Table S8) and performed GSEA on the marker genes to identify the phenotypes associated with each cell cluster (Fig 8D). Cluster 0 represented a mixed cell cluster with some enrichment of mature luminal-like cells. Cluster 1 was defined as Basal-like 1 as it showed a positive enrichment of basal breast cancer gene-set including specific markers *SAT1*, *GABRP*, *TM4SF1*, *TTYH1*, *KRT16* and *KRT6A* genes (Table S8). Cluster 2 is mesenchymal-like due to the selective expression of mesenchyme genes including *CCN2*, *TPM1* and *LAMB1*. Cluster 7, MGP-high cluster, is a relatively distinct cluster with a very specific high expression of *MGP* (Matrix Gla Protein) in all TNBC organoid lines (Fig 8C, S7E). MGP is normally expressed in smooth muscle cells perhaps suggesting that these cells might have contractile abilities. Cluster 12 had a very high expression of ribosomal and translation related genes (Fig 8C-D). Clusters 4,10 and 11 are the most distinct of the tumor cell clusters and are identified by the unique expression of certain genes. Cluster 10 represents an NFKB-active cluster

that was also found in normal organoids (Fig S6C), albeit with some differences in gene expression. As in normal organoids (Fig 7A, S6B, 8C). this cluster shows the expression of *CXCL1*, *CXCL3*, *NFKBIA* etc. Cluster 11 is also defined by a basal breast cancer signature (Fig 8C-D), is labeled Basal-like 2 and shows a selective expression of *KRT17*. Cluster 4 is a hypoxia cell cluster that shows selective high expression of *EGLN3*, *NDRG1*, *VEGFA* and likely represents cells at the center of these organoids (Fig 8C-D). This cluster is represented by a higher hypoxia score (Fig 8C-D), which interestingly correlates with basal stem cell (Basal SC) signatures and shows comparatively low activity of MYC and NOTCH pathways (Fig 8E and S7D-E). We also noted that the cluster specific genes for all clusters were fairly well conserved between the different PDO lines (Fig S7F). Overall, this data suggests that while the TNBC organoids retain patient intrinsic properties (Fig 1-5) there are common cell signatures that define the cell types within these organoids.

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

Discussion:

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

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

Patient-derived organoids are an exciting step towards personalized medicine with a promise to be used for real-time drug screens for guiding patient treatment (Tuveson and Clevers, 2019). When successfully established, breast cancer PDOs retain the genomic and transcriptomic features of breast cancers and their parent tissues. We found that, the TNBC organoids that showed long-term robust growth had activated MYC signaling, an LP-like gene expression signature and were overwhelmingly composed of LP-like cells. As these are among the most aggressive forms of breast cancers, organoid models appear to be an exciting avenue for their study. In addition to replicating the tumor specific features *in-vitro*, TNBC PDOs when transplanted into NOD/SCID mice generated tumors with remarkable morphological similarity to that of the original patient tumors, despite being in long-term organoid culture for up-to 18 passages. Long-term cultured organoids, thus, maintain the intrinsic ability to represent the tumor from which they were derived when placed in an *in vivo* environment. Interestingly, in our study not all PDOs readily generated primary tumors despite being highly proliferative and genomically aberrant organoids. The association of PDOs that generated tumors rapidly, with worse outcomes and diagnosis, suggests that there might be a biologically relevant explanation for this observed difference and must be investigated further along with long-term patient follow up information. Patient-derived xenograft (PDX) models also retain the histopathology of the original patient tumors (Bruna et al., 2016; DeRose et al., 2011) and a recent pre-print study showed successful derivation of breast cancer organoids from

PDX models (Guillen et al., 2021). Our models are complementary to this system and represent an opportunity to derive PDOs first, perform *in-vitro* assays and drug screens, followed by PDOX derivation for *in-vivo* validation studies.

While successfully established PDOs faithfully recapitulate the patient tumor properties, the efficiency of establishment of these cultures is currently less than ideal. The common challenges we faced during establishing organoid cultures included normal outgrowth in some tumor organoid cultures, stromal outgrowth, non-proliferating or dormant tumor cells, or limited growth potential *ex-vivo*. Since the medium composition for tumor organoid growth also support the culture of normal mammary organoids (Rosenbluth et al., 2020) the purity of the starting tumor tissue might govern the time to generation of tumor-only organoid cultures and early passages should be meticulously tested at a genomic level before any pharmacological studies are carried out with these models. Furthermore, some subtypes of breast cancers might be challenging to culture including the less proliferative Luminal A subtype or the more mesenchymal claudin-low group of TNBCs. Efforts to culture tumor-organoids in reduced complexity medium might alleviate some of these challenges and needs to be examined in future studies. However, the simultaneous growth of normal and TNBC organoids in the same culture conditions allows for comparison of similar cell types within these organoids.

We used an integrated scRNA-seq approach to compare normal LP cells with the tumor LP-like cells and identified hyperactivation of NOTCH and MYC signaling in the tumor compared to normal LPs. LP cells were previously shown to be the cell of origin of BRCA1 mutated basal-like BCs (Lim et al., 2009; Molyneux et al., 2010) and are speculated to be involved in tumorigenesis of all basal-like BCs due to similarity of gene

expression profiles (Lim et al., 2009). Our data provides strong support for this hypothesis by showing that TNBC PDOs are largely comprised of LP-like cells and suggests that hyperactivation of NOTCH and MYC signaling might be relevant in tumorigenesis from LP-cells (Fig 8F). Mouse studies have shown that Notch signaling is important in the maintenance of luminal lineage of the normal mammary gland (Bouras et al., 2008; Lilja et al., 2018; Rodilla et al., 2015; Yalcin-Ozuysal et al., 2010) and overactivation of Notch in luminal progenitors resulted in hyperplasia and acquisition of self-renewal properties (Bouras et al., 2008). Notch activation drives the luminal fate specificity in normal mammary gland and can reprogram committed basal mammary cells to an ER- luminal cell fate via multipotent embryonic cell states (Bland and Howard, 2018; Lilja et al., 2018), suggesting a role for Notch in fate specification and cellular plasticity. While our data provides strong support for hyperactivation of NOTCH signaling in LP-like cells in TNBCs it remains to be tested whether these cells arose from activation of NOTCH in normal LP cells or from reprograming of normal basal cells. Our data suggest that activation of MYC along with NOTCH might also have a role in this transformation. Given the complementarity of the NOTCH and MYC pathways, further studies are needed to gain better resolution of this mechanism. The PDO system, thus, can be exploited for identification of dysregulated pathways in cancers which can then be perturbed in the normal PDOs to understand the origins of cancers.

A detailed analysis of the cell type heterogeneity of TNBC organoids also revealed the presence of MYC/NOTCH-activity low and hypoxia high cells amongst various other cancer relevant cell types. Recent studies have shown that cancer cells can enter a MYC-low diapause-like state of dormancy upon chemotherapy treatment and result in therapy

escape (Dhimolea et al., 2021; Rehman et al., 2021). Furthermore, single cell analysis from TNBC tumors with residual disease after chemotherapy treatment showed an enrichment of hypoxia, angiogenesis, EMT and ECM degradation related genes in the persistent tumors (Kim et al., 2018). We hypothesize that in PDOs with these expression signatures, the cells that exhibit a MYC-low/hypoxia-high profile are the likely candidates for chemotherapy escape (Fig 8F). Additionally, fate mapping and other studies have shown that tumor cells exposed to hypoxia have a higher tendency to result in metastasis (Godet et al., 2019; Zhao et al., 2021). It remains to be tested to what extent these MYC-low/hypoxia-high cells contribute towards metastasis and/or chemotherapy resistance, and these patient-derived organoids will be an invaluable tool to answer these questions.

In summary, we developed a diverse biobank of BC organoids with a focus on TNBC-derived organoids. We have thoroughly characterized these models as valid systems that mimic the various aspects of patients' tumors, including genomic alternations, transcriptomic signatures, cell type specificity and morphological characteristics. Comparison of TNBC and normal-derived organoids provides important insights into the mechanisms regulating tumorigenesis that can then be validated by perturbation in normal PDOs (Dekkers et al., 2020). This comparison can also be used to identify novel tumor specific targets that may play an important role in tumor growth and progression (Chang et al., 2020). These next generation cancer models and the data derived from them offer vast utilities and can be used for drug-screens, co-culture experiments, metabolomics, and fate mapping studies to better understand the mechanisms driving cancers and for identifying more specific treatment options.

Acknowledgements

We would like to thank members of the Spector lab for critical discussions and advice throughout the course of this study. We thank Bodu Liu for critical review of this manuscript. We thank Dr. Dennis Plenker, Dr. Priya Sridevi and Hardik Patel from the laboratory of Dr. Dave Tuveson for their helpful insights and sharing their expertise on organoid cultures and QC. We also thank the various members of NCI-Human Cancer Model Initiative (HCMI) for general advice and inputs on organoid cultures. We acknowledge the CSHL Cancer Center Shared Resources (Animal, Flow Cytometry, Histology, Microscopy, Organoid, Next-Gen Sequencing, and Single Cell Biology) for services and technical expertise (NCI 2P3OCA45508). The sequencing analysis was performed using equipment purchased through NIH grant S10OD028632-01. We thank Dr. Zihua Wang for advice and a protocol for SMASH library preparation, Dr. Qing Gao for assistance with histological sample preparation and optimization, Jill Habel for assistance with mouse experiments. We are grateful for the support of Dr. Tawfiqui Bhuiya (Northwell Health) for tumor tissue sections and regret his untimely passing. We deeply appreciate the efforts of Northwell Health Biobanking team for sample acquisition and the patients and their families who donated tissues for research. This research was supported by CSHL/Northwell Health support (D.L.S.), NCI 5P01CA013106-Project 3 (D.L.S.), Leidos Biomedical HHSN26100008 (D.L.S), and the Manhasset Women's Coalition Against Breast Cancer (S.B.). This research was also supported by NCI Breast SPORE program P50-CA58223 (C.M.P), and U01CA238475-01 (C.M.P).

Author Contributions

Conceptualization, S.B., G.A. and D.L.S; Methodology, S.B, S.R., P.N., W.R.M. and G.A.; Formal Analysis, S.B., M.K., P.A., C.F., C.M.P., T.H., J.P. and J.E.W; Investigation, S.B., S.R., P.N., K.B. and J.P.; Resources, A.R. and K.K; Data Curation, S.B. and M.K.; Writing-Original Draft, S.B. and D.L.S; Writing- Review & Editing- S.B., C.M.P, J.P., G.A. and D.L.S; Visualization- S.B., M.K. and T.H.; Supervision, D.L.S; Funding Acquisition, S.B, D.L.S, and C.M.P.

Declaration of Interests

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.

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 (45 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 box in C.

Figure 3. TNBC organoids recapitulate the transcriptome of basal-like breast cancers. A.) Heatmap of sample-by-sample distance matrix showing Euclidean distances between the different PDOs. Samples were clustered using hierarchical clustering with Ward linkage. 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.)** Molecular signatures associated with the different organoid lines. The sample legends are as in 3A. **C & D.)** Box-plots showing the **C**.) breast cancer related and **D.**) 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= 16, TNBC met= 3. 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).

Figure 4. TNBC organoids show heterogenous growth properties in culture. A.) Light microscopy images of the various TNBC- and normal (NM04N) derived organoid lines, along with Ki67-IHC, scale bars=100μm **B.)** Distribution of maximum passage numbers tested for the various TNBC PDOs. Long-term cultures (long) are defined by

p>10 with continued expansion **C.)** Distribution of percent Ki67 positive cells per organoid for the various normal and tumor-derived organoids. The numbers on top of violin points indicate the number of organoids counted. The black line in the violin represents median. Data is plotted using GraphPad Prism. **D.)** Growth rate of organoid formation as measure by 3D Cell Titre Glo assay d6 post seeding, bars represent average with standard error and individual data points are measurements at different passages per PDO. Data is plotted using GraphPad Prism.

Figure 5. 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.)** 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.

Figure 6. 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 S6). Data is 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 A split by individual tumor and normal samples and showing normalized expression of *EPCAM* and *ITGA6* (gene encoding CD49f) expression patterns in individual cells.

Figure 7. Tumor LP 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 (Nguyen et al., 2018). **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 (Lim et al., 2009). 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 8. 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 for 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.

METHODS

Animals

Six-week-old female NOD scid mice (NOD.Cg-*Prkdc*^{scid}/J) were obtained from the Jackson laboratory (JAX stock #001303) and acclimated at the Cold Spring Harbor Laboratory Animal Shared Resource for a minimum of 1week. All animal experiments were performed in accordance with the Institutional Animal Care and Use Committee.

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 created, and data generated from them can be shared without exceptions with researchers in the scientific community.

METHOD DETAILS

Patient-derived organoid culture: Patient-derived organoids were established and propagated using a previously published protocol (Sachs et al., 2018). In summary, the tissues were manually cut into smaller pieces and treated with 2mg/ml collagenase IV in base media (ADF+++: Advanced DMEM-F12 (Invitrogen 12634-034) with 1x Glutamax (Invitrogen 12634-034), 10mM Hepes (Invitrogen 15630-056), 100U/ml Pen-Strep (Invitrogen 15140-122)) at 37°C for 45-90mins with gentle agitation to break the tissue

into small clusters of cells. The suspension was intermittently resuspended by pipetting multiple times to ensure proper digestion of the tissue. The cell suspension was centrifuged at 300g for 5mins and the pellet was treated with red blood cell lysis buffer (Cat # 11814389001, Sigma) for 5mins at room temperature if it appeared bloody. The suspension was washed 2x with ADF+++ and plated in a matrigel (lot test for concentration of 8-10mg/ml, Cat # 356231, Corning) dome on pre-warmed tissue culture plates. The dome was incubated at 37°C for 15mins and supplemented with completed medium: 10% R-Spondin1 conditioned medium, 5nM Neuregulin 1 (Peprotech 100-03), 5ng/ml FGF7 (Peprotech 100-19), 20ng/ml FGF10 (Peprotech 100-26), 5ng/ml EGF (Peprotech AF-100-15), 100ng/ml Noggin (Peprotech 120-10C), 500nM A83-01 (Tocris 2939), 5uM Y-27632 (Abmole Y-27632), 1.2uM SB202190 (Sigma S7067), 1x B27 supplement (Gibco 17504-44), 1.25mM N-Acetylcysteine (Sigma A9165), 5mM Nicotinamide (Sigma N0636), 50ug/ml Primocin (Invitrogen ant-pm-1) in ADF+++. The organoids were passaged every 15-30 days using TrypLE^{TM} (Thermo Fischer 12605028) to break down the organoids into smaller clusters of cells and re-plating them in matrigel domes as described above. For tumor scrapings, the tumor surface was shaved on multiple sides to collect material which was subsequently manually broken down, treated with red blood cell lysis buffer and seeded in matrigel followed by regular PDO culture.

Organoid models labelled with the prefix HCM-CSHL were acquired as part of the Human Cancer Model Initiative (HCMI) https://ocg.cancer.gov/programs/HCMI and a subset of those models are or will be available for access from ATCC. The data for these models can or will be accessed here: dbGaP accession number phs001486. Organoid nomenclature: prefixes LNS, NH, DS, HCM-CSHL are de-identified patient IDs and are

not distinct in any features other than protocols used for sample acquisition; prefix NM designates true normal samples collected from patients undergoing reductive mammoplasty (NM: Normal Mammoplasty); suffixes: T=tumor, N=normal, ND=normal distal, NAdj=normal adjacent, and TSc= tumor scraping. Organoid freeze thaws are indicated in parenthesis: (passage frozen down)passage after thaw eg. NH85TSc (p4)p4.

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

Targeted gene panel sequencing: We performed capture based targeted gene panel sequencing (Chen et al., 2015) for a panel of potential cancer driver genes. Briefly, we used a panel of 143 cancer genes with a total of ~4000 probes for capture. The captured DNA was paired-end sequenced with 150bp reads and a coverage of about 300-500x. Library preparation and sequencing of the targeted gene panel was performed by the CSHL Next Generation Sequencing Core Facility. We developed an analysis pipeline to prioritize identification of driver mutations. The sequencing reads are aligned to the hg19 reference genome using BWA (Li and Durbin, 2009), followed by conversion to BAM format and sorting with Samtools (Li et al., 2009), removal of PCR duplicates with Picard (https://broadinstitute.github.io/picard/), and filtering with Bamtools (Barnett et al., 2011) for mapping quality and proper read pairing. Coverage of the target regions is assessed

for breadth and depth using Picard HSmetrics to ensure adequate coverage for confident variant detection. Variants are then called using VarScan2 (Koboldt et al., 2012b) in somatic mode to stratify germline versus somatic variants, and are annotated with Annovar (Wang et al., 2010) to cover a broad range of variant assessment tools. We then select rare loss of function variants (nonsense, frameshift, splice site) with frequency less than 1% in the Gnomad, ExAC, EVS and 1000 Genomes databases. Missense and in-frame indel variants are selected if they are noted as pathogenic by ClinVar (Landrum et al., 2018), or if they are both rare and annotated as pathogenic by COSMIC (Tate et al., 2019), or if they are both rare and found to be present in the TCGA cohorts. Finally, missense variants are selected if they are annotated as potentially deleterious by the ensemble tools REVEL (Ioannidis et al., 2016) and MCAP (Jagadeesh et al., 2016). Variants that were deleterious by REVEL and MCAP but did not have population level data were discarded from the final oncoplot. Oncoplots are generated from these candidate variants using Maftools (Mayakonda et al., 2018).

RNA-seq: All RNA samples were quality controlled using a nanodrop followed by a 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 (Dobin et al., 2013) 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 (Love et al., 2014). 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 (Fan et al., Gatza et al., Garcia-Recio et al.) and visualized using Java TreeView v1.2.0. For individual signature comparisons signature scores in experimental groups were compared using Kruskal-Wallis test followed by pairwise comparisons using Wilcoxon rank-sum test. Family-wise error rate was adjusted using Bonferroni-Holm method.

SMASH copy number: SMASH was performed as per the published protocol (Wang et al., 2016) starting with 750ng of genomic DNA. SMASH was performed in batches of 10 samples and sequenced on a MiSeq PE150bp run. The SMASH analysis is based on identification of Maximal Unique Matches (MUMs) to the human genome in all read pairs (Wang et al., 2016). These MUMs were filtered to remove matches <20 bp, matches with <4 bp of excess unique sequence, and matches on read 2 that are within 1000 bases of the matches from read 1. Raw copy number profiles are then generated from the remaining 3–4 matches per read pair which are then added to empirically sized bins spanning the genome. Regions with identical copy are expected to yield similar bin counts using these empirical bins. We next perform GC correction by normalizing counts based on LOWESS smoothing of count vs. GC content data in each bin. Final copy number profiles are normalized so that the autosome has an average copy number of 2. Plots were generated with G-Graph MUMdex software (https://mumdex.com/) and IGV browser v2.9.2 (Robinson et al., 2011).

Organoid Histology: Organoid domes in complete medium were scraped from the tissue culture plate and collected in falcon tube precoated with BSA. The organoids were collected and washed 1x with PBS by spinning at 300g for 5mins. Organoid harvesting solution (Cat # 3700-100-01, Trevigen) was added to the organoids (3x the volume of matrigel) and incubated at 4°C on ice for 30minutes to ensure that matrigel was removed and the organoids were concentrated at the bottom. The organoids were washed 1x with ample PBS and fixed with fresh 4% PFA at room temperature for 10minutes. 1:1 (v/v) BSA was added to the tube and spun at 300g for 5mins to remove the PFA. The organoids

were washed 2X with ample amounts of PBS and embedded in 2% agarose in dH2O). The agarose organoid molds were then paraffin embedded and cut into 5um sections.

Organoid Hematoxylin & Eosin (H&E) staining and Immunohistochemistry (IHC): H&E and IHC staining were performed at the CSHL Histology Core Facility. PFA fixed organoids in agarose were processed in Thermo Excelsior ES processor and embedded with Thermo HistoStar embedding system following manufacturer's protocols. Paraffin blocks were cut into 5um sections and mounted onto positively charged slides (VWR superfrost plus micro slide).

For H&E staining, slides were stained in a Leica Multistainer (ST5020). Briefly, slides were deparaffinized and rehydrated and then stained in hematoxylin (Hematoxylin 560 MX, Leica) for 1 min, followed by destaining in Define MX-aq (Leica) for 30 sec, bluing in Blue Buffer 8 (Leica) for 1min and subsequently stained in eosin (EOSIN 515 LT, Leica) for 30sec. After dehydration, coverslips were placed onto glass slides using a robotic coversliper (Leica CV5030).

IHC slides were stained in Discovery Ultra automatic IHC stainer (Roche) following standard protocols. Briefly, slides were subjected to antigen retrieval (Benchmark Ultra CC1, Roche) at 96˚C for 1hr; primary antibodies were incubated at 37˚C for 1hr and Discovery multimer detection system (Discovery OmniMap HRP, Discovery DAB and Purple, Roche) was used to detect and amplify immuno-signals. Antibodies used: Ki67 (Spring Bioscience, #M3062, 1:500).

Organoid formation assay: Organoids were processed using TrypLE and 1500 single cells per well of a 96 well plate were seeded in 10% matrigel + complete growth medium.

Cell viability was assessed using Cell Titre Glo 3D luminescence assay (Promega G9683 CellTiter-Glo 3D Cell Viability Assay). Baseline cells were measured using Cell Titre Glo 3D assay at 24 hrs (d1) post seeding and growth was measured at 6 days after seeding (d6). Each organoid line was evaluated for multiple passages n=2 or n=3 per PDO.

For MYCi and DAPT experiments: Organoids were processed to single cells using TrypLE and seeded in 50ul matrigel domes as 10,000 cells per well of a 24 well-plate. Complete medium or medium supplemented with DMSO, DAPT (Selleckchem S2215 DAPT) and MYC-inhibitor (Selleckchem S8906 MYCi975) were added to the respective wells. Organoids were allowed to form for 12 days, images were acquired using microscope and organoids were manually counted for each condition. Statistical analysis was performed using one-way ANOVA test followed by pairwise comparisons using two-sample t-test. Family-wise error rate was adjusted using Bonferroni-Hold method.

Organoid proliferation index analysis: Paraffin embedded organoid sections were IHC stained for Ki67. Slides were scanned and viewed using Aperio ImageScope 12.3.3. Images were analyzed in FIJI (Schindelin et al., 2012). Briefly, the images were deconvolved using Colour Deconvolution for hematoxylin and DAB, converted to 8-bit binary images and analyzed using the BioVoxxel Toolbox plugin (https://www.biovoxxel.de/#/) to evalute %Ki67 positive cells per organoid. Multiple passages for each organoid line were evaluated, n=2 or n=3 per PDO.

In-vivo* transplant experiments*: TNBC organoids were harvested using the organoid harvesting solution (Trevigen Cat# 3700-100-01) and manually counted. Organoids were

resuspended in a 1:1 mixture of PBS:matrigel and 50K organoids were injected into the bilateral mammary fat pads by the fourth nipple of female NOD scid mice (NOD.Cg-*Prkdc*^{scid}/J). Mice were anesthetized with 1.5-2% isoflurane and weighed before the injections. The animals were regularly monitored for their weight, tumor size and any other discrepancies. Mice were sacrificed when the tumors reached any of the following end-points: 2cm tumors, ulceration, visible necrosis, blistering of tumors or deteriorating health of the mice. At end-point, dissections were performed and the tumors along with lungs, liver, lymph nodes and the femur were fixed in 4% PFA overnight at 4°C. If the tumors were not observed the mammary fat pads were collected instead. The transplant experiments were done with 2 independent passages of PDOs, with 4-6 injections per PDO per passage.

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

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

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

Single cell analysis was performed in three different batches (Fig S5E). Batch 3 was a multiplexed pool of 4 samples, which were demultipexed using a custom genotype-aware pipeline. At the time of 10X Genomics library preparation, ~20,000 cells from each of the four organoids were set aside to prepare low-input bulk RNA-seq libraries tagged with unique i7 barcodes. These bulk libraries share the same adapter structure as 10X Genomics libraries, and were spiked into the Illumina NextSeq500 flow cell at a 5% molar ratio to obtain ~5M reads per organoid. These barcoded bulk libraries were then used to create reference VCF files using cellSNP v0.3.2 by searching a list of 7.4M common human SNPs from the 1000 Genomes Project (http://ufpr.dl.sourceforge.net/project/cellsnp/SNPlist/genome1K.phase3.SNP_AF5e2.chr1toX.hg38.vcf.gz). Genotype profiles were filtered to include only positions with < 10% minor allele frequency and >20 UMI counts. In parallel, per-cell VCF files were generated from the multiplexed single cell library using the cellSNP with the same parameters. Cells from the single-cell pool were assigned to their respective donors using Vireo v0.4.2 (Huang et al., 2019).

Genotyping low-input bulk RNAseq library prep:

RT Primer Design:

CTACACGACGCTCTTCCGATCTSSSSSSSSNNNNNNNNNNVVVVVTTTTTTTT

TTTTTTTTTTTTTTTTTTTTTTVN

where:

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

GACAGTGC=HCM-CSHL-0366-C50

GAGTTAGC=NH85TSc

GATGAATC=NH95T

GCCAAGAC=NH93T

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

Template Switch Oligo: AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG

cDNA_amplification_Forward:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG

cDNA_amplification_Reverse: AAGCAGTGGTATCAACGCAGAGT

RT was performed using the SuperScript IV (Life Technologies #18091050) according to the manufacturer's instructions except for the addition of 1uM Template Switch Oligo during first strand synthesis. All custom oligos were synthesized by IDT. After cDNA amplification, molar concentration was estimated using the Agilent Bioanalyzer 2100, libraries were pooled at an equimolar ratio, and prepared for Illumina sequencing using the Nextera XT DNA Library Prep Kit (Illumina) according to the manufacturer's

instructions. Final fragmented libraries were again checked and quantified by Bioanalyzer prior to mixing at a 1:20 molar ratio with 10X Genomics libraries for sequencing. *Sequencing and Mapping:* The libraries were sequenced on Illumina NextSeq 500 High Output 75 cycle kits using the read format: 8bp (i7 index) x 28bp (Read 1) x 56bp (Read2). 10X Genomics libraries were mapped using Cell Ranger version 4.0.0 (10X Genomics) with default settings and a custom genome reference based on the comprehensive gene annotation set from Gencode Release 32 (GRCh38.p13) (http://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_32/gencode.v32.annotation.gtf.gz). For the multiplexed pool, sample identities for each cell were assigned using Vireo as described above, such that each sample could be subset from the pooled matrix during analysis.

Filtering, feature selection, clustering, and other secondary analyses were carried out in R using Seurat v4.0.3 (https://satijalab.org/seurat/) (Butler et al., 2018; Stuart et al., 2019). Gene set enrichment analyses were performed using GSEA v4.1.0 (Subramanian et al., 2005). Following gene sets were used for the various signature scores: adult human breast epithelium markers from (Nguyen et al., 2018), Mammary epithelial lineage scores from (Lim et al., 2009), NOTCH signaling: REACTOME_SIGNALING_BY_NOTCH, BMP2 targets: LEE_BMP2_TARGETS_UP, MYC signature from (Chandriani et al., 2009), Hypoxia signature: HALLMARK_HYPOXIA, Basal mammary stem cell signature from (Lim et al., 2009).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using R (version 4.1.0) on RStudio and GraphPad Prism software (v9.1.2, GraphPad Software, San Diego, California USA, www.graphpad.com). Specific tests are indicated in the figure legends along with the statistical significance.

Supplementary Materials

Supplemental Figure Legends

Figure S1. related to Figure 1,2 **A. & B.)** Variant allele frequencies comparisons on a patient-wise manner for the various SNVs identified in **A.)** tumor and/or PDOs **B.)** different passage PDOs. R represents Pearson correlation. **C.)** Oncoplot for pathogenic variants in tumor vs scraping derived organoids. **D.)** Oncoplot for tumor tissue (p0 organoids) for samples that did not result in successful organoid cultures. **E.**) Loss of copy number alterations in one PDO model overtime in culture.

Figure S2. related to Figure 3 **A.)** Copy number profiles of NH48N and NH48T **B.)** Gene expression signatures for selected modules. Pink box highlights the samples defined a LT-TNBCs **C.)** Gene expression signatures for luminal-progenitor gene sets and c-MYC amplification. Each dot represents a different PDO; Luminal N=12, Normal N= 7, TNBC N= 16, TNBC met= 3. **D.**) Copy number data for selected MYC genes

Figure S3. related to Figure 4 **A.)** Selected culture images for organoids that did not result in successful cultures. **B.)** Ki67 IHC and H&E images for PDOs derived from normal and tumor pairs of the same patient, scale bars= 100μm.

Figure S4. related to Figure 5 **A.)** Overview of the PDO xenotransplant experiments N1 and N2 **B.)** Visualization of micro-metastasis and single cell metastasis using IHC for human mitochondria staining in lung, liver and lymph nodes of the NOD/SCID mice transplanted with various PDOs. Pink arrows point to human mitochondria stained micro- or single cell-metastasis.

Figure S5. related to Figure 6 **A.)** Flow-cytometry contour plots for EPCAM and CD49f for selected tumor PDOs (in red contours) overlayed onto normal-derived organoids (in blue contours). **D.**) Expression of ITGA6 and EPCAM from RNA-seq data for the different PDOs. Dark pink dots: true normal PDOs, light pink dots: paired normal PDOs and black dots: tumors. Circled region represents long term propagating organoids. **E.)** Different passage flow-cytometry plots for NH95T, NH85TSc and HCM-CSHL-0366-C50 showing stability of the cell populations over time in culture. **F.**) Flow-cytometry plots for some more TNBC and lymph-met derived organoids.

Figure S6. related to Figure 7 **A. & B.)** Cell cycle related **A.)** UMAP-plot and **B.)** gene expression in the cycling clusters **C.)** Cluster specific markers associated with each G1 cluster identified **D.)** Expression patterns of the various luminal/basal markers in normal and TNBC PDOs. HC-0366 is PDO HCM-CSHL-0366-C50 and HC-0655 is PDO HCM-

CSHL-0655-C50 **E.)** Mammary lineage scores for each of the PDOs in specific clusters **F.)** Gene-set scores for each of the specified gene-set between the tumor and normal samples. **G.)** Motif enrichment for differentially expressed genes that were significantly upregulated in tumor LPs compared to normal LPs (clusters 2, 7 and 11). Score= -10*NES*padj.value **H.)** Representative culture images for three different PDOs treated with DAPT and MYCi

Figure S7. related to Figure 8 **A.)** Cluster distribution for each of the TNBC PDOs tested. **B.)** EPCAM and ITGA6 expression for all the TNBC samples **C.)** Feature plots showing mammary lineage scores for each of the PDOs in specific clusters **D.**) Boxplots showing scores for the different gene-sets tested **E**.) Feature plots showing the gene-set score distribution among the different TNBC PDOs. The first panel is the combined score distribution for all samples and provides the legend for the whole row. **F.**) Marker expression for one gene per cluster for the combined tumor organoids (first panel) and for individual TNBC organoids.

Supplementary Table details:

Table S1. Clinical and culture properties of the 84 patient tumor tissues used in this study. Highlighted green and yellow samples are paired samples from two different patients.

Table S2. Final copy number segment file from SMASH data for all samples

Table S3. Sample details for samples used in transcriptomic profiling

Table S4. Gene expression signatures analysis for the various PDOs

Table S5. Percent Ki67 per organoid analysis for various normal and tumor PDOs

Table S6. Summary of EPCAM/CD49f flow cytometry data for the various lines. N= multiple passages for the tumor PDOs, and multiple different patient-derived normal organoid lines for the normal (NM0s) PDOs

Table S7. Table of differential gene expression for comparing tumor LPs vs normal LPs

Table S8. Cluster specific markers for each of the clusters identified in TNBC PDOs

References

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., Marth, G.T., 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27, 1691–1692. https://doi.org/10.1093/bioinformatics/btr174

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M.F., Monahan, J.E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., Mapa, F.A., Thibault, J., Bric-Furlong, E., Raman, P., Shipway, A., Engels, I.H., Cheng, J., Yu, G.K., Yu, J., Aspesi, P., de Silva, M., Jagtap, K., Jones, M.D., Wang, L., Hatton, C., Palescandolo, E., Gupta, S., Mahan, S., Sougnez, C., Onofrio, R.C., Liefeld, T., MacConaill, L., Winckler, W., Reich, M., Li, N., Mesirov, J.P., Gabriel, S.B., Getz, G., Ardlie, K., Chan, V., Myer, V.E., Weber, B.L., Porter, J., Warmuth, M., Finan, P., Harris, J.L., Meyerson, M., Golub, T.R., Morrissey, M.P., Sellers, W.R., Schlegel, R., Garraway, L.A., 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607. https://doi.org/10.1038/nature11003

Beshiri, M.L., Tice, C.M., Tran, C., Nguyen, H.M., Sowalsky, A.G., Agarwal, S., Jansson, K.H., Yang, Q., McGowen, K.M., Yin, J., Alilin, A.N., Karzai, F.H., Dahut, W.L., Corey, E., Kelly, K., 2018. 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 24, 4332–4345. https://doi.org/10.1158/1078-0432.CCR-18-0409

Bland, P., Howard, B.A., 2018. Mammary lineage restriction in development. Nature Cell Biology 20, 637–639. https://doi.org/10.1038/s41556-018-0111-6

Boj, S.F., Hwang, C.-I., Baker, L.A., Chio, I.I.C., Engle, D.D., Corbo, V., Jager, M., Ponz-Sarvise, M., Tiriac, H., Spector, M.S., Gracanin, A., Oni, T., Yu, K.H., van Boxtel, R., Huch, M., Rivera, K.D., Wilson, J.P., Feigin, M.E., Öhlund, D., Handly-Santana, A., Ardito-Abraham, C.M., Ludwig, M., Elyada, E., Alagesan, B., Biffi, G., Yordanov, G.N., Delcuze, B., Creighton, B., Wright, K., Park, Y., Morsink, F.H.M., Molenaar, I.Q., Borel Rinkes, I.H., Cuppen, E., Hao, Y., Jin, Y., Nijman, I.J., Iacobuzio-Donahue, C., Leach, S.D., Pappin, D.J., Hammell, M., Klimstra, D.S., Basturk, O., Hruban, R.H., Offerhaus, G.J., Vries, R.G.J., Clevers, H.,

Tuveson, D.A., 2015. Organoid Models of Human and Mouse Ductal Pancreatic Cancer. Cell 160, 324–338. https://doi.org/10.1016/j.cell.2014.12.021

Bouras, T., Pal, B., Vaillant, F., Harburg, G., Asselin-Labat, M.-L., Oakes, S.R., Lindeman, G.J., Visvader, J.E., 2008. Notch Signaling Regulates Mammary Stem Cell Function and Luminal Cell-Fate Commitment. Cell Stem Cell 3, 429–441. https://doi.org/10.1016/j.stem.2008.08.001

Bruna, A., Rueda, O.M., Greenwood, W., Batra, A.S., Callari, M., Batra, R.N., Pogrebniak, K., Sandoval, J., Cassidy, J.W., Tufegdzic-Vidakovic, A., Sammut, S.-J., Jones, L., Provenzano, E., Baird, R., Eirew, P., Hadfield, J., Eldridge, M., McLaren-Douglas, A., Barthorpe, A., Lightfoot, H., O'Connor, M.J., Gray, J., Cortes, J., Baselga, J., Marangoni, E., Welm, A.L., Aparicio, S., Serra, V., Garnett, M.J., Caldas, C., 2016. A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. Cell 167, 260-274.e22. https://doi.org/10.1016/j.cell.2016.08.041

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411–420. https://doi.org/10.1038/nbt.4096

Chakraborty, U., Dinh, T.A., Alani, E., 2018. Genomic Instability Promoted by Overexpression of Mismatch Repair Factors in Yeast: A Model for Understanding Cancer Progression. Genetics 209, 439–456. https://doi.org/10.1534/genetics.118.300923

Chandriani, S., Frengen, E., Cowling, V.H., Pendergrass, S.A., Perou, C.M., Whitfield, M.L., Cole, M.D., 2009. A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. PloS one 4, e6693–e6693. https://doi.org/10.1371/journal.pone.0006693

Chang, K.-C., Diermeier, S.D., Yu, A.T., Brine, L.D., Russo, S., Bhatia, S., Alsudani, H., Kostroff, K., Bhuiya, T., Brogi, E., Pappin, D.J., Bennett, C.F., Rigo, F., Spector, D.L., 2020. MaTAR25 lncRNA regulates the Tensin1 gene to impact breast cancer progression. Nature Communications 11, 6438–6438. https://doi.org/10.1038/s41467-020-20207-y

Chen, R., Im, H., Snyder, M., 2015. Whole-Exome Enrichment with the Roche NimbleGen SeqCap EZ Exome Library SR Platform. Cold Spring Harb Protoc 2015, pdb.prot084855. https://doi.org/10.1101/pdb.prot084855

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J.D., Tavaré, S., Caldas, C., Aparicio, S., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346–352. https://doi.org/10.1038/nature10983

Dekkers, J.F., Whittle, J.R., Vaillant, F., Chen, H.-R., Dawson, C., Liu, K., Geurts, M.H., Herold, M.J., Clevers, H., Lindeman, G.J., Visvader, J.E., 2020. Modeling Breast Cancer Using CRISPR-Cas9-Mediated Engineering of Human Breast Organoids. Journal of the National Cancer Institute 112, 540–544. https://doi.org/10.1093/jnci/djz196

DeRose, Y.S., Wang, G., Lin, Y.-C., Bernard, P.S., Buys, S.S., Ebbert, M.T.W., Factor, R., Matsen, C., Milash, B.A., Nelson, E., Neumayer, L., Randall, R.L., Stijleman, I.J., Welm, B.E., Welm, A.L., 2011. Tumor grafts derived from women with breast cancer authentically reflect

tumor pathology, growth, metastasis and disease outcomes. Nature Medicine 17, 1514–1520. https://doi.org/10.1038/nm.2454

Dhimolea, E., de Matos Simoes, R., Kansara, D., Al'Khafaji, A., Bouyssou, J., Weng, X., Sharma, S., Raja, J., Awate, P., Shirasaki, R., Tang, H., Glassner, B.J., Liu, Z., Gao, D., Bryan, J., Bender, S., Roth, J., Scheffer, M., Jeselsohn, R., Gray, N.S., Georgakoudi, I., Vazquez, F., Tsherniak, A., Chen, Y., Welm, A., Duy, C., Melnick, A., Bartholdy, B., Brown, M., Culhane, A.C., Mitsiades, C.S., 2021. An Embryonic Diapause-like Adaptation with Suppressed Myc Activity Enables Tumor Treatment Persistence. Cancer Cell 39, 240-256.e11. https://doi.org/10.1016/j.ccell.2020.12.002

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Fan, C., Prat, A., Parker, J.S., Liu, Y., Carey, L.A., Troester, M.A., Perou, C.M., 2011. Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. BMC Medical Genomics 4, 3–3. https://doi.org/10.1186/1755-8794-4-3

Fujii, M., Sato, T., 2021. Somatic cell-derived organoids as prototypes of human epithelial tissues and diseases. Nature Materials 20, 156–169. https://doi.org/10.1038/s41563-020-0754-0

Fusco, N., Geyer, F.C., De Filippo, M.R., Martelotto, L.G., Ng, C.K.Y., Piscuoglio, S., Guerini-Rocco, E., Schultheis, A.M., Fuhrmann, L., Wang, L., Jungbluth, A.A., Burke, K.A., Lim, R.S., Vincent-Salomon, A., Bamba, M., Moritani, S., Badve, S.S., Ichihara, S., Ellis, I.O., Reis-Filho, J.S., Weigelt, B., 2016. 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 29, 1292–1305. https://doi.org/10.1038/modpathol.2016.134

Garcia-Recio, S., Thennavan, A., East, M.P., Parker, J.S., Cejalvo, J.M., Garay, J.P., Hollern, D.P., He, X., Mott, K.R., Galván, P., Fan, C., Selitsky, S.R., Coffey, A.R., Marron, D., Brasó-Maristany, F., Burgués, O., Albanell, J., Rojo, F., Lluch, A., de Dueñas, E.M., Rosen, J.M., Johnson, G.L., Carey, L.A., Prat, A., Perou, C.M., 2020. FGFR4 regulates tumor subtype differentiation in luminal breast cancer and metastatic disease. Journal of Clinical Investigation 130, 4871–4887. https://doi.org/10.1172/JCI130323

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., Liu, Q., Iorio, F., Surdez, D., Chen, L., Milano, R.J., Bignell, G.R., Tam, A.T., Davies, H., Stevenson, J.A., Barthorpe, S., Lutz, S.R., Kogera, F., Lawrence, K., McLaren-Douglas, A., Mitropoulos, X., Mironenko, T., Thi, H., Richardson, L., Zhou, W., Jewitt, F., Zhang, T., O'Brien, P., Boisvert, J.L., Price, S., Hur, W., Yang, W., Deng, X., Butler, A., Choi, H.G., Chang, J.W., Baselga, J., Stamenkovic, I., Engelman, J.A., Sharma, S.V., Delattre, O., Saez-Rodriguez, J., Gray, N.S., Settleman, J., Futreal, P.A., Haber, D.A., Stratton, M.R., Ramaswamy, S., McDermott, U., Benes, C.H., 2012. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 483, 570–575. https://doi.org/10.1038/nature11005

Gatza, M.L., Silva, G.O., Parker, J.S., Fan, C., Perou, C.M., 2014. An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. Nature Genetics 46, 1051–1059. https://doi.org/10.1038/ng.3073

Giraddi, R.R., Chung, C.-Y., Heinz, R.E., Balcioglu, O., Novotny, M., Trejo, C.L., Dravis, C., Hagos, B.M., Mehrabad, E.M., Rodewald, L.W., Hwang, J.Y., Fan, C., Lasken, R., Varley, K.E., Perou, C.M., Wahl, G.M., Spike, B.T., 2018. Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development. Cell reports 24, 1653-1666.e7. https://doi.org/10.1016/j.celrep.2018.07.025

Godet, I., Shin, Y.J., Ju, J.A., Ye, I.C., Wang, G., Gilkes, D.M., 2019. Fate-mapping post-hypoxic tumor cells reveals a ROS-resistant phenotype that promotes metastasis. Nature Communications 10, 4862–4862. https://doi.org/10.1038/s41467-019-12412-1

Gong, Y., Zack, T.I., Morris, L.G.T., Lin, K., Hukkelhoven, E., Raheja, R., Tan, I.-L., Turcan, S., Veeriah, S., Meng, S., Viale, A., Schumacher, S.E., Palmedo, P., Beroukhim, R., Chan, T.A., 2014. Pan-cancer genetic analysis identifies PARK2 as a master regulator of G1/S cyclins. Nature genetics 46, 588–594. https://doi.org/10.1038/ng.2981

Guillen, K.P., Fujita, M., Butterfield, A.J., Scherer, S.D., Bailey, M.H., Chu, Z., DeRose, Y.S., Zhao, L., Cortes-Sanchez, E., Yang, C.-H., Toner, J., Wang, G., Qiao, Y., Huang, X., Greenland, J.A., Vahrenkamp, J.M., Lum, D.H., Factor, R.E., Nelson, E.W., Matsen, C.B., Poretta, J.M., Rosenthal, R., Beck, A.C., Buys, S.S., Vaklavas, C., Ward, J.H., Jensen, R.L., Jones, K.B., Li, Z., Oesterreich, S., Dobrolecki, L.E., Pathi, S.S., Woo, X.Y., Berrett, K.C., Wadsworth, M.E., Chuang, J.H., Lewis, M.T., Marth, G.T., Gertz, J., Varley, K.E., Welm, B.E., Welm, A.L., 2021. A breast cancer patient-derived xenograft and organoid platform for drug discovery and precision oncology. bioRxiv 2021.02.28.433268. https://doi.org/10.1101/2021.02.28.433268

Han, H., Jain, A.D., Truica, M.I., Izquierdo-Ferrer, J., Anker, J.F., Lysy, B., Sagar, V., Luan, Y., Chalmers, Z.R., Unno, K., Mok, H., Vatapalli, R., Yoo, Y.A., Rodriguez, Y., Kandela, I., Parker, J.B., Chakravarti, D., Mishra, R.K., Schiltz, G.E., Abdulkadir, S.A., 2019. Small-Molecule MYC Inhibitors Suppress Tumor Growth and Enhance Immunotherapy. Cancer cell 36, 483-497.e15. https://doi.org/10.1016/j.ccell.2019.10.001

Herschkowitz, J., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L., Assefnia, S., Chandrasekharan, S., Backlund, M., Yin, Y., Khramtsov, A., Bastein, R., Quackenbush, J.F., Glazer, R., Brown, P., Green, J., Kopelovich, L., Furth, P., Palazzo, J., Olopade, O., Bernard, P., Churchill, G., van Dyke, T., Perou, C., 2006. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biology 8, R76–R76.

Hoffmann, K., Berger, H., Kulbe, H., Thillainadarasan, S., Mollenkopf, H.-J., Zemojtel, T., Taube, E., Darb-Esfahani, S., Mangler, M., Sehouli, J., Chekerov, R., Braicu, E.I., Meyer, T.F., Kessler, M., 2020. Stable expansion of high-grade serous ovarian cancer organoids requires a low-Wnt environment. The EMBO journal 39, e104013–e104013. https://doi.org/10.15252/embj.2019104013

Howlader, N., Altekruse, S.F., Li, C.I., Chen, V.W., Clarke, C.A., Ries, L.A.G., Cronin, K.A., 2014. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. Journal of the National Cancer Institute 106. https://doi.org/10.1093/jnci/dju055

Hu, Z., Fan, C., Livasy, C., He, X., Oh, D.S., Ewend, M.G., Carey, L.A., Subramanian, S., West, R., Ikpatt, F., Olopade, O.I., van de Rijn, M., Perou, C.M., 2009. A compact VEGF signature associated with distant metastases and poor outcomes. BMC Medicine 7, 9–9. https://doi.org/10.1186/1741-7015-7-9

Huang, Y., McCarthy, D.J., Stegle, O., 2019. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. Genome Biology 20, 273. https://doi.org/10.1186/s13059-019-1865-2

Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L.A., Teerlink, C.C., Stanford, J.L., Isaacs, W.B., Xu, J., Cooney, K.A., Lange, E.M., Schleutker, J., Carpten, J.D., Powell, I.J., Cussenot, O., Cancel-Tassin, G., Giles, G.G., MacInnis, R.J., Maier, C., Hsieh, C.-L., Wiklund, F., Catalona, W.J., Foulkes, W.D., Mandal, D., Eeles, R.A., Kote-Jarai, Z., Bustamante, C.D., Schaid, D.J., Hastie, T., Ostrander, E.A., Bailey-Wilson, J.E., Radivojac, P., Thibodeau, S.N., Whittemore, A.S., Sieh, W., 2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. The American Journal of Human Genetics 99, 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

Jagadeesh, K.A., Wenger, A.M., Berger, M.J., Guturu, H., Stenson, P.D., Cooper, D.N., Bernstein, J.A., Bejerano, G., 2016. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nature Genetics 48, 1581–1586. https://doi.org/10.1038/ng.3703

Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., Navin, N.E., 2018. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell 1–15. https://doi.org/10.1016/j.cell.2018.03.041

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., Wilson, R.K., Ally, A., Balasundaram, M., Butterfield, Y.S.N., Carlsen, R., Carter, C., Chu, Andy, Chuah, E., Chun, H.J.E., Coope, R.J.N., Dhalla, N., Guin, R., Hirst, C., Hirst, M., Holt, R.A., Lee, D., Li, H.I., Mayo, M., Moore, R.A., Mungall, A.J., Pleasance, E., Robertson, A.G., Schein, J.E., Shafiei, A., Sipahimalani, P., Slobodan, J.R., Stoll, D., Tam, A., Thiessen, N., Varhol, R.J., Wye, N., Zeng, T., Zhao, Y., Birol, I., Jones, S.J.M., Marra, M.A., Cherniack, A.D., Saksena, G., Onofrio, R.C., Pho, N.H., Carter, S.L., Schumacher, S.E., Tabak, B., Hernandez, B., Gentry, J., Nguyen, H., Crenshaw, A., Ardlie, K., Beroukhim, R., Winckler, W., Getz, G., Gabriel, S.B., Meyerson, M., Chin, L., Kucherlapati, R., Hoadley, K.A., Auman, J.T., Fan, C., Turman, Y.J., Shi, Y., Li, L., Topal, M.D., He, X., Chao, H.H., Prat, A., Silva, G.O., Iglesia, M.D., Zhao, W., Usary, J., Berg, J.S., Adams, M., Booker, J., Wu, J., Gulabani, A., Bodenheimer, T., Hoyle, A.P., Simons, J.V., Soloway, M.G., Mose, L.E., Jefferys, S.R., Balu, S., Parker, J.S., Hayes, D.N., Perou, C.M., Malik, S., Mahurkar, S., Shen, H., Weisenberger, D.J., Triche, T., Lai, P.H., Bootwalla, M.S., Maglinte, D.T., Berman, B.P., Van Den Berg, D.J., Baylin, S.B., Laird, P.W., Creighton, C.J., Donehower, L.A., Noble, M., Voet, D., Gehlenborg, N., Di Cara, D., Zhang, Juinhua, Zhang, H., Wu, C.J., Yingchun Liu, S., Lawrence, M.S., Zou, L., Sivachenko, A., Lin, P., Stojanov, P., Jing, R., Cho, J., Sinha, Raktim, Park, R.W., Nazaire, M.D., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P.J., Reynolds, S., Kreisberg, R.B., Bernard, B., Bressler, R., Erkkila, T., Lin, J., Thorsson, V., Zhang, W., Shmulevich, I., Ciriello, G., Weinhold, N., Schultz, N., Gao, J., Cerami, E.,

Gross, B., Jacobsen, A., Sinha, Rileen, Aksoy, B.A., Antipin, Y., Reva, B., Shen, R., Taylor, B.S., Ladanyi, M., Sander, C., Anur, P., Spellman, P.T., Lu, Y., Liu, W., Verhaak, R.R.G., Mills, G.B., Akbani, R., Zhang, N., Broom, B.M., Casasent, T.D., Wakefield, C., Unruh, A.K., Baggerly, K., Coombes, K., Weinstein, J.N., Haussler, D., Benz, C.C., Stuart, J.M., Benz, S.C., Zhu, J., Szeto, C.C., Scott, G.K., Yau, C., Paull, E.O., Carlin, D., Wong, C., Sokolov, A., Thusberg, J., Mooney, S., Ng, S., Goldstein, T.C., Ellrott, K., Grifford, M., Wilks, C., Ma, S., Craft, B., Yan, C., Hu, Y., Meerzaman, D., Gastier-Foster, J.M., Bowen, J., Ramirez, N.C., Black, A.D., Pyatt, R.E., White, P., Zmuda, E.J., Frick, J., Lichtenberg, T.M., Brookens, R., George, M.M., Gerken, M.A., Harper, H.A., Leraas, K.M., Wise, L.J., Tabler, T.R., McAllister, C., Barr, T., Hart-Kothari, M., Tarvin, K., Saller, C., Sandusky, G., Mitchell, C., Iacocca, M.V., Brown, J., Rabeno, B., Czerwinski, C., Petrelli, N., Dolzhansky, O., Abramov, M., Voronina, O., Potapova, O., Marks, J.R., Suchorska, W.M., Murawa, D., Kycler, W., Ibbs, M., Korski, K., Spychała, A., Murawa, P., Brzeziński, J.J., Perz, H., Łaźniak, R., Teresiak, M., Tatka, H., Leporowska, E., Bogusz-Czerniewicz, M., Malicki, J., Mackiewicz, A., Wiznerowicz, M., Van Le, X., Kohl, B., Viet Tien, N., Thorp, R., Van Bang, N., Sussman, H., Phu, B.D., Hajek, R., Hung, N.P., Phuong, T.V.T., Thang, H.Q., Khan, K.Z., Penny, R., Mallery, D., Curley, E., Shelton, C., Yena, P., Ingle, J.N., Couch, F.J., Lingle, W.L., King, T.A., Gonzalez-Angulo, A.M., Dyer, M.D., Liu, S., Meng, X., Patangan, M., Waldman, F., Stöppler, H., Rathmell, W.K., Thorne, L., Huang, M., Boice, L., Hill, A., Morrison, C., Gaudioso, C., Bshara, W., Daily, K., Egea, S.C., Pegram, M.D., Gomez-Fernandez, C., Dhir, R., Bhargava, R., Brufsky, A., Shriver, C.D., Hooke, J.A., Campbell, J.L., Mural, R.J., Hu, H., Somiari, S., Larson, C., Deyarmin, B., Kvecher, L., Kovatich, A.J., Ellis, M.J., Stricker, T., White, K., Olopade, O., Luo, C., Chen, Y., Bose, R., Chang, L.W., Beck, A.H., Pihl, T., Jensen, M., Sfeir, R., Kahn, A., Chu, Anna, Kothiyal, P., Wang, Z., Snyder, E., Pontius, J., Ayala, B., Backus, M., Walton, J., Baboud, J., Berton, D., Nicholls, M., Srinivasan, D., Raman, R., Girshik, S., Kigonya, P., Alonso, S., Sanbhadti, R., Barletta, S., Pot, D., Sheth, M., Demchok, J.A., Shaw, K.R.M., Yang, L., Eley, G., Ferguson, M.L., Tarnuzzer, R.W., Zhang, Jiashan, Dillon, L.A.L., Buetow, K., Fielding, P., Ozenberger, B.A., Guyer, M.S., Sofia, H.J., Palchik, J.D., 2012a. Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70. https://doi.org/10.1038/nature11412

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012b. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576. https://doi.org/10.1101/gr.129684.111

Kopper, O., de Witte, C.J., Lohmussaar, K., Valle-Inclan, J.E., Hami, N., Kester, L., Balgobind, A.V., Korving, J., Proost, N., Begthel, H., van Wijk, L.M., Revilla, S.A., Theeuwsen, R., van de Ven, M., van Roosmalen, M.J., Ponsioen, B., Ho, V.W.H., Neel, B.G., Bosse, T., Gaarenstroom, K.N., Vrieling, H., Vreeswijk, M.P.G., van Diest, P.J., Witteveen, P.O., Jonges, T., Bos, J.L., van Oudenaarden, A., Zweemer, R.P., Snippert, H.J.G., Kloosterman, W.P., Clevers, H., 2019. An organoid platform for ovarian cancer captures intra- and interpatient heterogeneity. Nature medicine 25, 838–849. https://doi.org/10.1038/s41591-019-0422-6

Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A.,

McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J.B., Kattman, B.L., Maglott, D.R., 2018. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research 46, D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., Pietenpol, J.A., 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. The Journal of clinical investigation 121, 2750–2767. https://doi.org/10.1172/JCI45014

Lehmann, B.D., Pietenpol, J.A., Tan, A.R., 2015. Triple-Negative Breast Cancer: Molecular Subtypes and New Targets for Therapy. American Society of Clinical Oncology Educational Book e31–e39. https://doi.org/10.14694/edbook_am.2015.35.e31

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liedtke, C., Mazouni, C., Hess, K.R., Andre, F., Tordai, A., Mejia, J.A., Symmans, W.F., Gonzalez-Angulo, A.M., Hennessy, B., Green, M., Cristofanilli, M., Hortobagyi, G.N., Pusztai, L., 2008. 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 26, 1275–1281. https://doi.org/10.1200/JCO.2007.14.4147

Lilja, A.M., Rodilla, V., Huyghe, M., Hannezo, E., Landragin, C., Renaud, O., Leroy, O., Rulands, S., Simons, B.D., Fre, S., 2018. 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 20, 677–687. https://doi.org/10.1038/s41556-018-0108-1

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.-L., Gyorki, D.E., Ward, T., Partanen, A., Feleppa, F., Huschtscha, L.I., Thorne, H.J., Fox, S.B., Yan, M., French, J.D., Brown, M.A., Smyth, G.K., Visvader, J.E., Lindeman, G.J., 2009. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nature Medicine 15, 907–913. https://doi.org/10.1038/nm.2000

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Marchiò, C., Weigelt, B., Reis-Filho, J.S., 2010. 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 63, 220–228. https://doi.org/10.1136/jcp.2009.073908

Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., Koeffler, H.P., 2018. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 28, 1747–1756. https://doi.org/10.1101/gr.239244.118

Molyneux, G., Geyer, F.C., Magnay, F.-A., McCarthy, A., Kendrick, H., Natrajan, R., MacKay, A., Grigoriadis, A., Tutt, A., Ashworth, A., Reis-Filho, J.S., Smalley, M.J., 2010. BRCA1 Basal-like Breast Cancers Originate from Luminal Epithelial Progenitors and Not from Basal Stem Cells. Cell Stem Cell 7, 403–417. https://doi.org/10.1016/j.stem.2010.07.010

Nguyen, Q.H., Pervolarakis, N., Blake, K., Ma, D., Davis, R.T., James, N., Phung, A.T., Willey, E., Kumar, R., Jabart, E., Driver, I., Rock, J., Goga, A., Khan, S.A., Lawson, D.A., Werb, Z., Kessenbrock, K., 2018. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nature Communications 9, 2028. https://doi.org/10.1038/s41467-018-04334-1

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., Van Loo, P., Ju, Y.S., Smid, M., Brinkman, A.B., Morganella, S., Aure, M.R., Lingjaerde, O.C., Langerod, A., Ringner, M., Ahn, S.-M., Boyault, S., Brock, J.E., Broeks, A., Butler, A., Desmedt, C., Dirix, L., Dronov, S., Fatima, A., Foekens, J.A., Gerstung, M., Hooijer, G.K.J., Jang, S.J., Jones, D.R., Kim, H.-Y., King, T.A., Krishnamurthy, S., Lee, H.J., Lee, J.-Y., Li, Y., McLaren, S., Menzies, A., Mustonen, V., O'Meara, S., Pauporte, I., Pivot, X., Purdie, C.A., Raine, K., Ramakrishnan, K., Rodriguez-Gonzalez, F.G., Romieu, G., Sieuwerts, A.M., Simpson, P.T., Shepherd, R., Stebbings, L., Stefansson, O.A., Teague, J., Tommasi, S., Treilleux, I., Van den Eynden, G.G., Vermeulen, P., Vincent-Salomon, A., Yates, L., Caldas, C., van't Veer, L., Tutt, A., Knappskog, S., Tan, B.K.T., Jonkers, J., Borg, A., Ueno, N.T., Sotiriou, C., Viari, A., Futreal, P.A., Campbell, P.J., Span, P.N., Van Laere, S., Lakhani, S.R., Eyfjord, J.E., Thompson, A.M., Birney, E., Stunnenberg, H.G., van de Vijver, M.J., Martens, J.W.M., Borresen-Dale, A.-L., Richardson, A.L., Kong, G., Thomas, G., Stratton, M.R., 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534, 47–54. https://doi.org/10.1038/nature17676

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J.F., Stijleman, I.J., Palazzo, J., Marron, J.S., Nobel, A.B., Mardis, E., Nielsen, T.O., Ellis, M.J., Perou, C.M., Bernard, P.S., 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 27, 1160–1167. https://doi.org/10.1200/JCO.2008.18.1370

Perou, C., Sørlie, T., Eisen, M., Rijn, M.V., Jeffrey, S., Rees, C., Pollack, J., Ross, D.T., Johnsen, H., Akslen, L., Fluge, Ø., Pergamenschikov, A., Williams, C., Zhu, S., Lønning, P., Børresen-Dale, A.-L., Brown, P., Botstein, D., 2000. Molecular portraits of human breast tumours. Nature 406, 747–752.

Perou, C.M., 2011. Molecular Stratification of Triple-Negative Breast Cancers. The Oncologist 16, 61–70. https://doi.org/10.1634/theoncologist.2011-S1-61

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

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., Perou, C.M., 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Research 12, R68. https://doi.org/10.1186/bcr2635

Prat, A., Perou, C.M., 2011. Deconstructing the molecular portraits of breast cancer. Molecular Oncology 5, 5–23. https://doi.org/10.1016/j.molonc.2010.11.003

Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., Iscove, N., Jones, S., McKinney, S., Emerman, J., Aparicio, S., Marra, M., Eaves, C., 2008. Transcriptome Analysis of the Normal Human Mammary Cell Commitment and Differentiation Process. Cell Stem Cell 3, 109–118. https://doi.org/10.1016/j.stem.2008.05.018

Rehman, S.K., Haynes, J., Collignon, E., Brown, K.R., Wang, Y., Nixon, A.M.L., Bruce, J.P., Wintersinger, J.A., Singh Mer, A., Lo, E.B.L., Leung, C., Lima-Fernandes, E., Pedley, N.M., Soares, F., McGibbon, S., He, H.H., Pollet, A., Pugh, T.J., Haibe-Kains, B., Morris, Q., Ramalho-Santos, M., Goyal, S., Moffat, J., O'Brien, C.A., 2021. Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. Cell 184, 226-242.e21. https://doi.org/10.1016/j.cell.2020.11.018

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat Biotechnol 29, 24–26. https://doi.org/10.1038/nbt.1754

Rodilla, V., Dasti, A., Huyghe, M., Lafkas, D., Laurent, C., Reyal, F., Fre, S., 2015. Luminal Progenitors Restrict Their Lineage Potential during Mammary Gland Development. PLOS Biology 13, e1002069. https://doi.org/10.1371/journal.pbio.1002069

Rosenbluth, J.M., Schackmann, R.C.J., Gray, G.K., Selfors, L.M., Li, C.M.-C., Boedicker, M., Kuiken, H.J., Richardson, A., Brock, J., Garber, J., Dillon, D., Sachs, N., Clevers, H., Brugge, J.S., 2020. Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. Nature Communications 11, 1711–1711. https://doi.org/10.1038/s41467-020-15548-7

Rueda, O.M., Sammut, S.-J., Seoane, J.A., Chin, S.-F., Caswell-Jin, J.L., Callari, M., Batra, R., Pereira, B., Bruna, A., Ali, H.R., Provenzano, E., Liu, B., Parisien, M., Gillett, C., McKinney, S., Green, A.R., Murphy, L., Purushotham, A., Ellis, I.O., Pharoah, P.D., Rueda, C., Aparicio, S., Caldas, C., Curtis, C., 2019. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. Nature 567, 399–404. https://doi.org/10.1038/s41586-019-1007-8

Sachs, N., de Ligt, J., Kopper, O., Gogola, E., Bounova, G., Weeber, F., Balgobind, A.V., Wind, K., Gracanin, A., Begthel, H., Korving, J., van Boxtel, R., Duarte, A.A., Lelieveld, D., van Hoeck, A., Ernst, R.F., Blokzijl, F., Nijman, I.J., Hoogstraat, M., van de Ven, M., Egan, D.A., Zinzalla, V., Moll, J., Boj, S.F., Voest, E.E., Wessels, L., van Diest, P.J., Rottenberg, S., Vries, R.G.J., Cuppen, E., Clevers, H., 2018. A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. Cell 172, 373-386.e10. https://doi.org/10.1016/j.cell.2017.11.010

Sato, T., Stange, D.E., Ferrante, M., Vries, R.G.J., van Es, J.H., van den Brink, S., van Houdt, W.J., Pronk, A., van Gorp, J., Siersema, P.D., Clevers, H., 2011. Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. Gastroenterology 141, 1762–1772. https://doi.org/10.1053/j.gastro.2011.07.050

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K.,

Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., Bashashati, A., Prentice, L.M., Khattra, J., Burleigh, A., Yap, D., Bernard, V., McPherson, A., Shumansky, K., Crisan, A., Giuliany, R., Heravi-Moussavi, A., Rosner, J., Lai, D., Birol, I., Varhol, R., Tam, A., Dhalla, N., Zeng, T., Ma, K., Chan, S.K., Griffith, M., Moradian, A., Cheng, S.W.G., Morin, G.B., Watson, P., Gelmon, K., Chia, S., Chin, S.F., Curtis, C., Rueda, O.M., Pharoah, P.D., Damaraju, S., MacKey, J., Hoon, K., Harkins, T., Tadigotla, V., Sigaroudinia, M., Gascard, P., Tlsty, T., Costello, J.F., Meyer, I.M., Eaves, C.J., Wasserman, W.W., Jones, S., Huntsman, D., Hirst, M., Caldas, C., Marra, M.A., Aparicio, S., 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486, 395–399. https://doi.org/10.1038/nature10933

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C., Lønning, P., Brown, P., Børresen-Dale, A.-L., Botstein, D., 2003. 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 100, 8418–8423.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S.C., Kok, C.Y., Noble, K., Ponting, L., Ramshaw, C.C., Rye, C.E., Speedy, H.E., Stefancsik, R., Thompson, S.L., Wang, S., Ward, S., Campbell, P.J., Forbes, S.A., 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Research 47, D941–D947. https://doi.org/10.1093/nar/gky1015

Tuveson, D., Clevers, H., 2019. Cancer modeling meets human organoid technology. Science 364, 952–955. https://doi.org/10.1126/science.aaw6985

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research 38, e164–e164. https://doi.org/10.1093/nar/gkq603

Wang, Z., Andrews, P., Kendall, J., Ma, B., Hakker, I., Rodgers, L., Ronemus, M., Wigler, M., Levy, D., 2016. SMASH, a fragmentation and sequencing method for genomic copy number analysis. Genome research 26, 844–851. https://doi.org/10.1101/gr.201491.115

Weigman, V.J., Chao, H.-H., Shabalin, A.A., He, X., Parker, J.S., Nordgard, S.H., Grushko, T., Huo, D., Nwachukwu, C., Nobel, A., Kristensen, V.N., Børresen-Dale, A.-L., Olopade, O.I., Perou, C.M., 2012. Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. Breast Cancer Res Treat 133, 865–880. https://doi.org/10.1007/s10549-011-1846-y

Wilczak, W., Rashed, S., Hube-Magg, C., Kluth, M., Simon, R., Büscheck, F., Clauditz, T.S., Grupp, K., Minner, S., Tsourlakis, M.C., Möller-Koop, C., Graefen, M., Adam, M., Haese, A., Wittmer, C., Sauter, G., Izbicki, J.R., Huland, H., Schlomm, T., Steurer, S., Krech, T., Lebok, P., 2017. 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 38, 19–27. https://doi.org/10.1093/carcin/bgw116

Yalcin-Ozuysal, Ö., Fiche, M., Guitierrez, M., Wagner, K.-U., Raffoul, W., Brisken, C., 2010. Antagonistic roles of Notch and p63 in controlling mammary epithelial cell fates. Cell Death & Differentiation 17, 1600–1612. https://doi.org/10.1038/cdd.2010.37

Yu, C., Mannan, A.M., Yvone, G.M., Ross, K.N., Zhang, Y.-L., Marton, M.A., Taylor, B.R., Crenshaw, A., Gould, J.Z., Tamayo, P., Weir, B.A., Tsherniak, A., Wong, B., Garraway, L.A., Shamji, A.F., Palmer, M.A., Foley, M.A., Winckler, W., Schreiber, S.L., Kung, A.L., Golub, T.R., 2016. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. Nature Biotechnology 34, 419–423. https://doi.org/10.1038/nbt.3460

Zhao, Y., Fu, X., Lopez, J.I., Rowan, A., Au, L., Fendler, A., Hazell, S., Xu, H., Horswell, S., Shepherd, S.T.C., Spain, L., Byrne, F., Stamp, G., O'Brien, T., Nicol, D., Augustine, M., Chandra, A., Rudman, S., Toncheva, A., Pickering, L., Sahai, E., Larkin, J., Bates, P.A., Swanton, C., Turajlic, S., Litchfield, K., 2021. Selection of metastasis competent subclones in the tumour interior. Nature Ecology & Evolution 1–13. https://doi.org/10.1038/s41559-021-01456-6

Figure
Figure 1

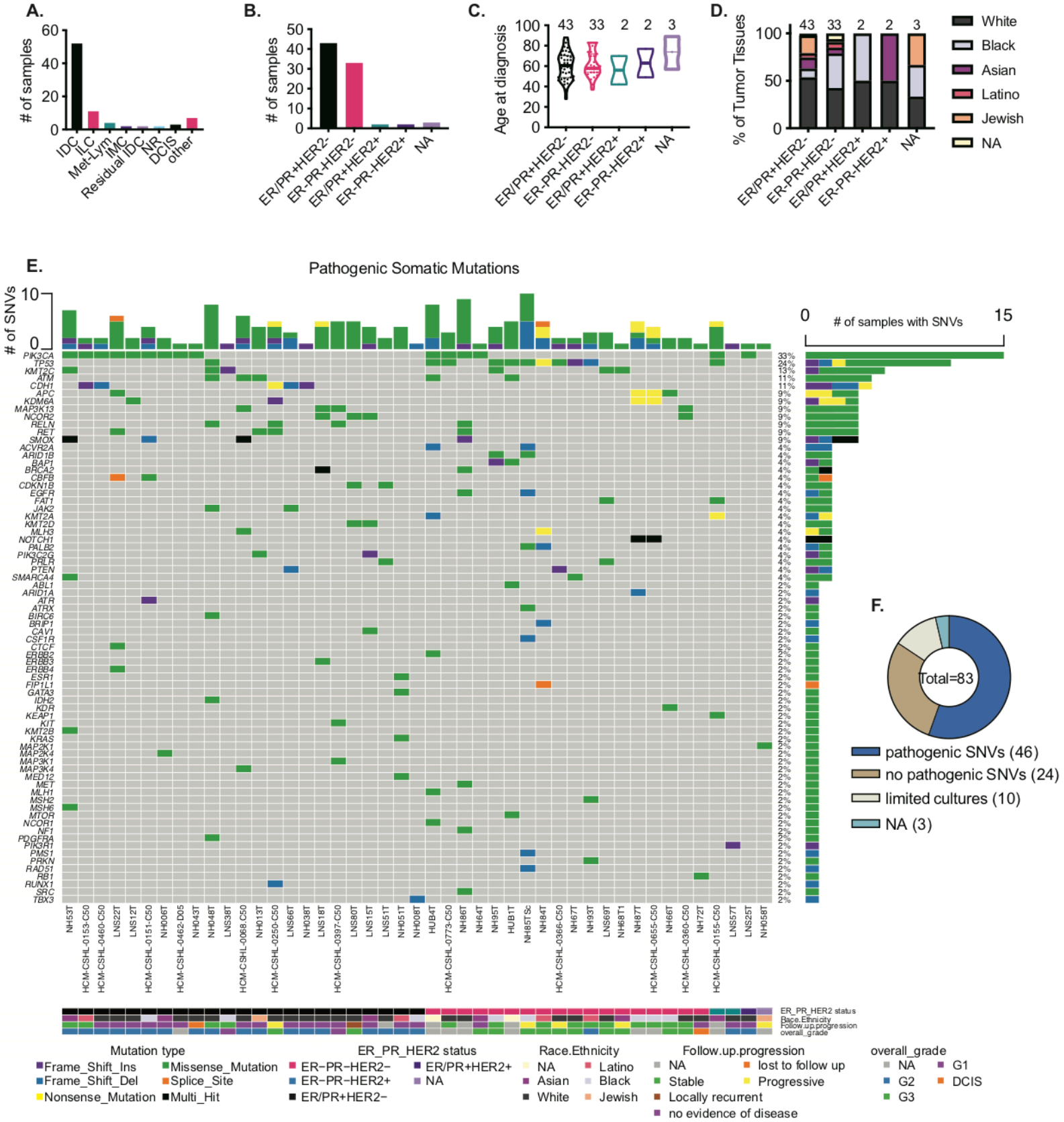


Figure 2

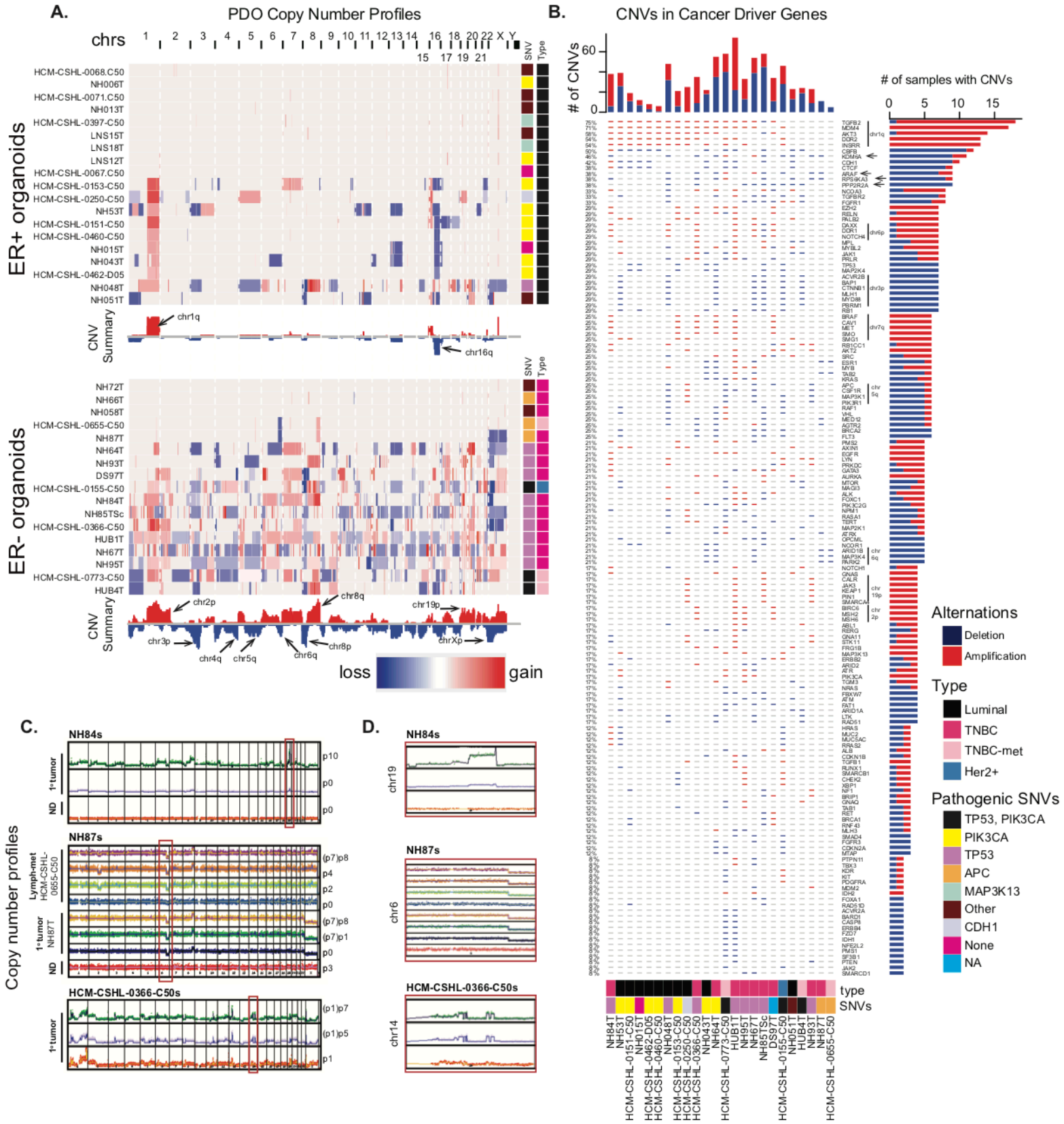

A.
PDO Copy Number Profiles
chrs
ER+ organoids
ER- organoids
CNV Summary
chr1q
chr16q
chr2p
chr8q
chr19p
chr3p
chr4q
chr5q
chr6q
chr8p
chrXp
loss
gain
B.
CNVs in Cancer Driver Genes
# of CNVs
# of samples with CNVs
Alternations
Deletion
Amplification
Type
Luminal
TNBC
TNBC-met
Her2+
Pathogenic SNVs
TP53, PIK3CA
PIK3CA
TP53
APC
MAP3K13
Other
CDH1
None
NA
type
SNVs
C.
NH84s
NH87s
HCM-CSHL-0366-C50s
Copy number profiles
D.
chr19
chr6
chr14

Figure 3

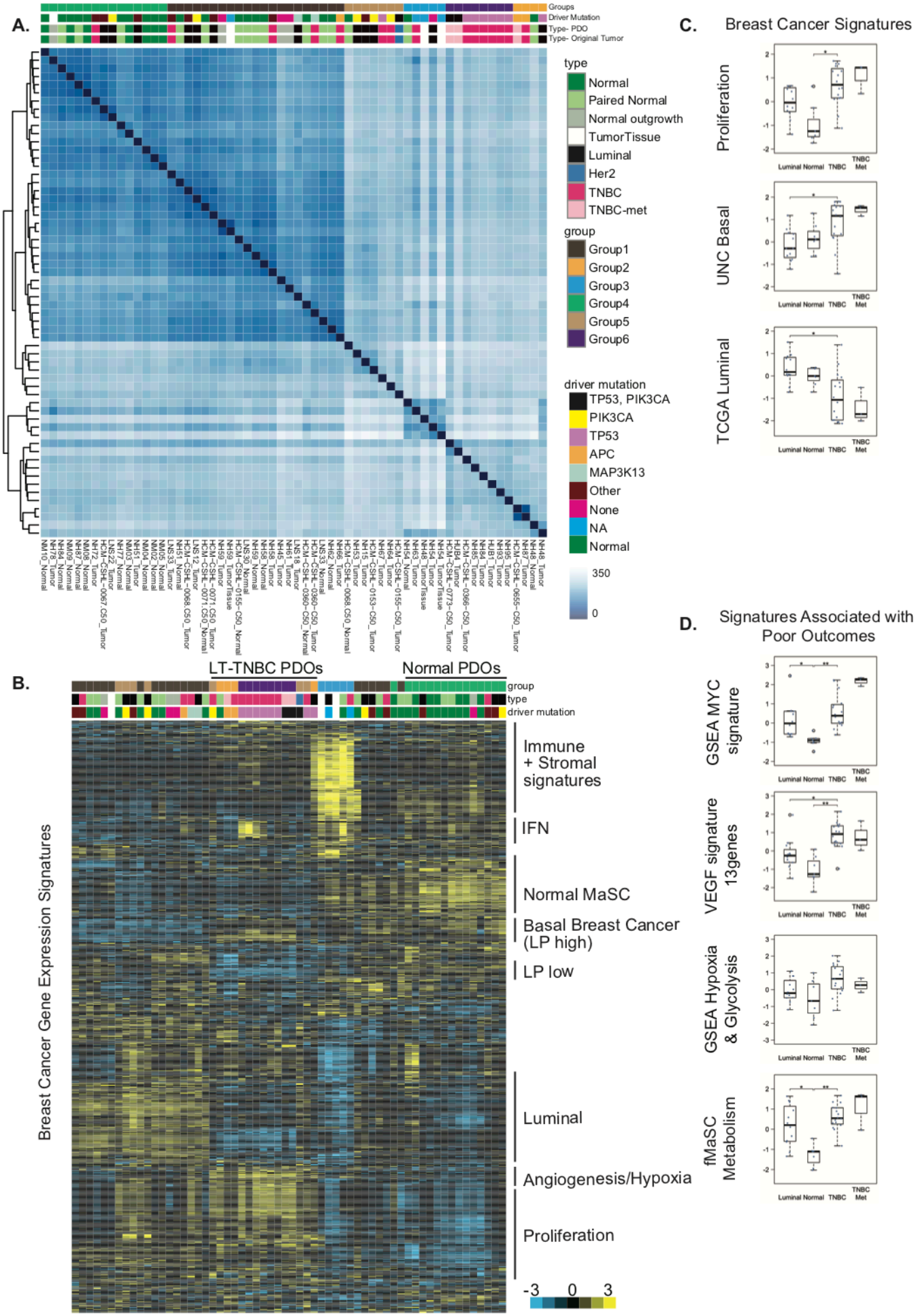

A.
Groups
Driver Mutation
Type- PDO
Type- Original Tumor
type
Normal
Paired Normal
Normal outgrowth
TumorTissue
Luminal
Her2
TNBC
TNBC-met
group
Group1
Group2
Group3
Group4
Group5
Group6
driver mutation
TP53, PIK3CA
PIK3CA
TP53
APC
MAP3K13
Other
None
NA
Normal
350
0
B.
LT-TNBC PDOs
Normal PDOs
group
type
driver mutation
Breast Cancer Gene Expression Signatures
Immune + Stromal signatures
IFN
Normal MaSC
Basal Breast Cancer (LP high)
LP low
Luminal
Angiogenesis/Hypoxia
Proliferation
-3
0
3
C.
Breast Cancer Signatures
Proliferation
UNC Basal
TCGA Luminal
Luminal Normal
TNBC
TNBC Met
D.
Signatures Associated with Poor Outcomes
GSEA MYC signature
VEGF signature 13genes
GSEA Hypoxia & Glycolysis
fMaSC Metabolism

Figure 4

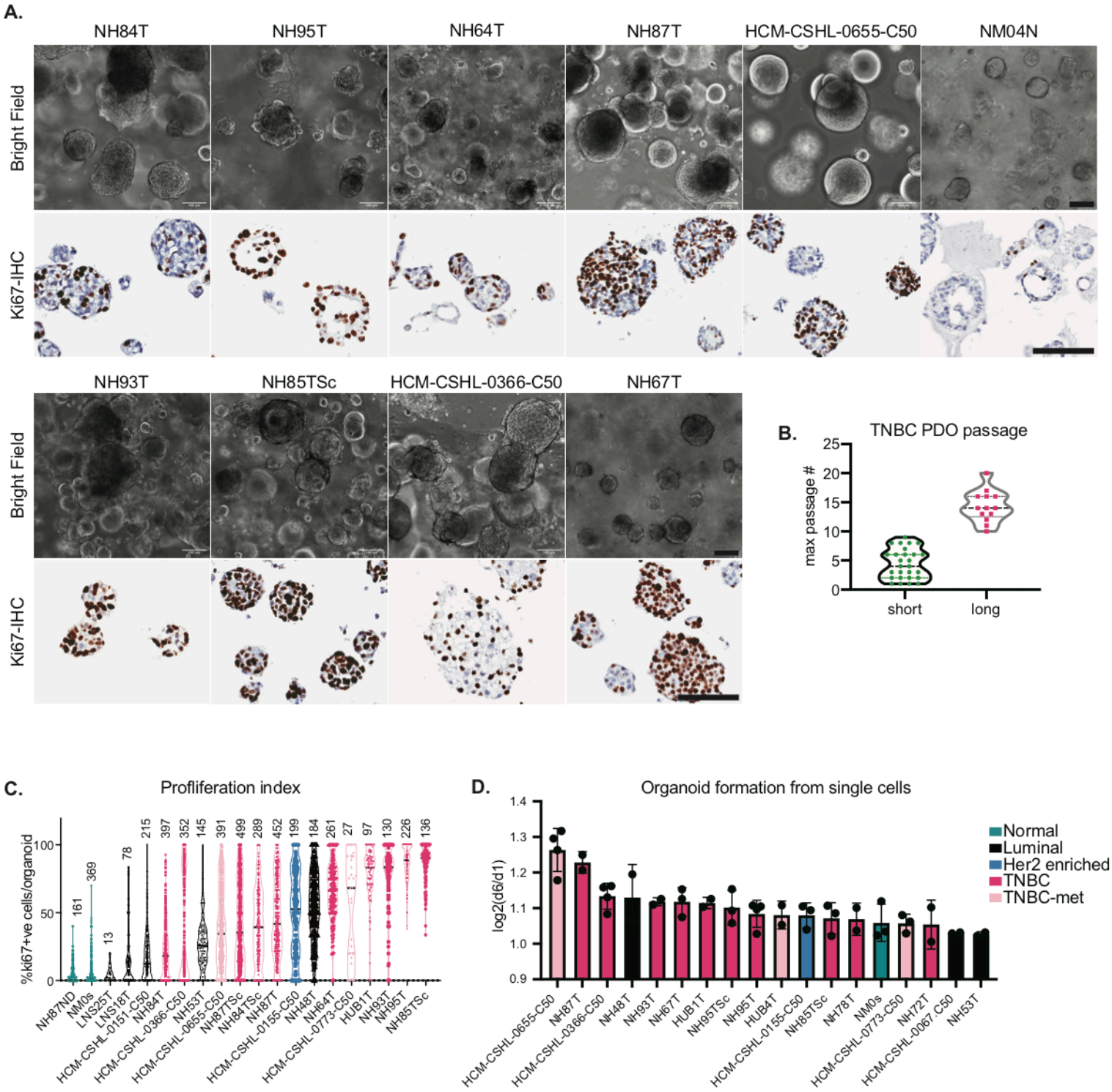

A.
NH84T
NH95T
NH64T
NH87T
HCM-CSHL-0655-C50
NM04N
Bright Field
Ki67-IHC
NH93T
NH85TSc
HCM-CSHL-0366-C50
NH67T
B.
TNBC PDO passage
max passage #
short
long
C.
Profliferation index
%ki67+ve cells/organoid
NH87ND
NM0s
LNS25T
LNS18T
HCM-CSHL-0151-C50
NH84T
HCM-CSHL-0366-C50
NH53T
HCM-CSHL-0655-C50
NH87TSc
NH84TSc
NH87T
HCM-CSHL-0155-C50
NH48T
NH64T
HCM-CSHL-0773-C50
HUB1T
NH93T
NH95T
NH85TSc
D.
Organoid formation from single cells
log2(d6/d1)
Normal
Luminal
Her2 enriched
TNBC
TNBC-met
HCM-CSHL-0655-C50
NH87T
HCM-CSHL-0366-C50
NH48T
NH93T
NH67T
HUB1T
NH95TSc
NH95T
HUB4T
HCM-CSHL-0155-C50
NH85TSc
NH78T
NM0s
HCM-CSHL-0773-C50
NH72T
HCM-CSHL-0067.C50
NH53T

Figure 5

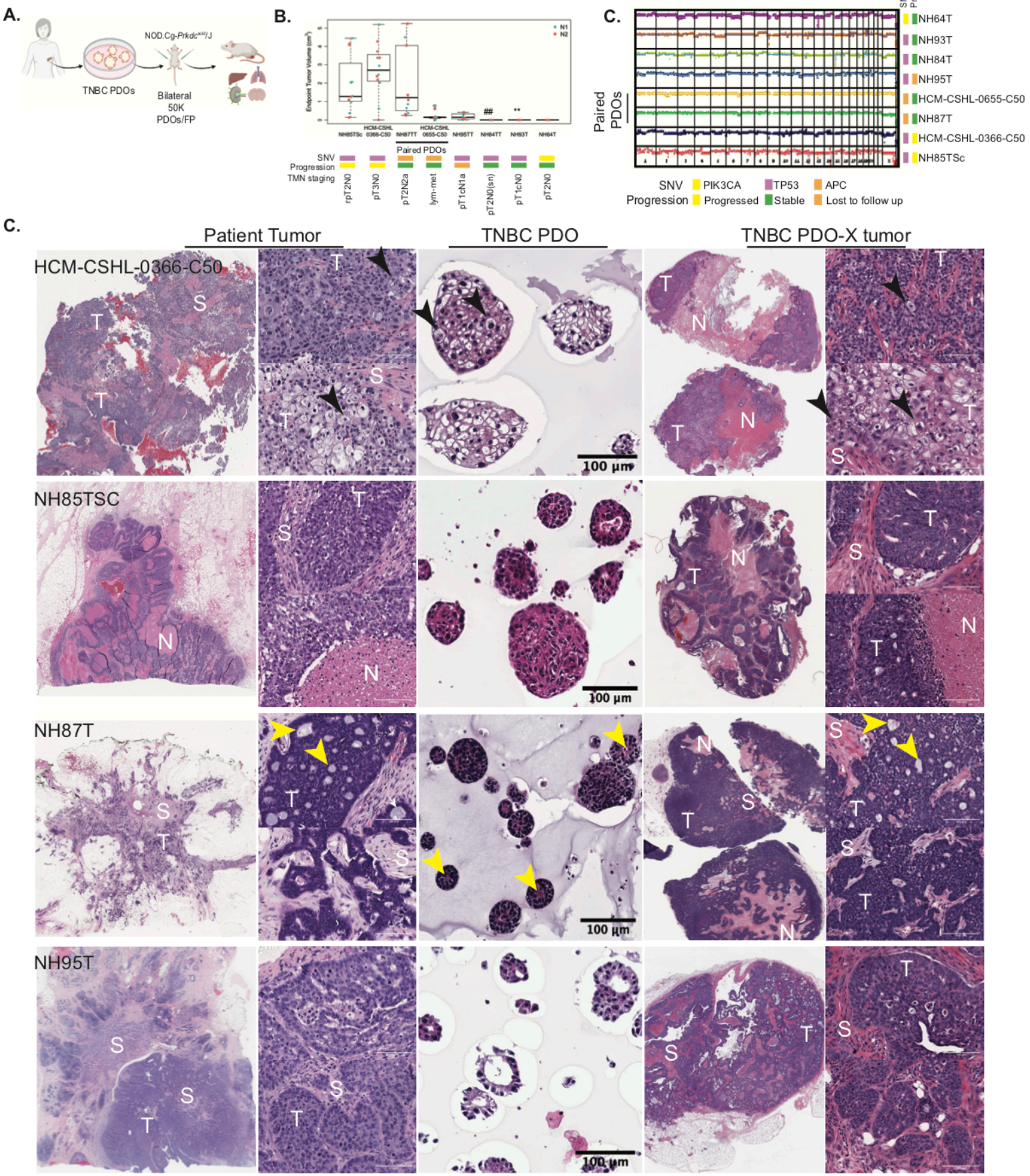

A.
NOD.Cg-Prkdc scid/J
TNBC PDOs
Bilateral
50K
PDOs/FP
B.
Endpoint Tumor Volume (cm³)
N1
N2
##
**
NH85TSc
HCM-CSHL 0366-C50
NH87TT
HCM-CSHL 0655-C50
NH95TT
NH84TT
NH93T
NH64T
Paired PDOs
SNV
Progression
TMN staging
rpT2N0
pT3N0
pT2N2a
lym-met
pT1cN1a
pT2N0(sn)
pT1cN0
pT2N0
C.
SNV
Progression
NH64T
NH93T
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
C.
Patient Tumor
TNBC PDO
TNBC PDO-X tumor
HCM-CSHL-0366-C50
NH85TSC
NH87T
NH95T
100 µm

Figure 6

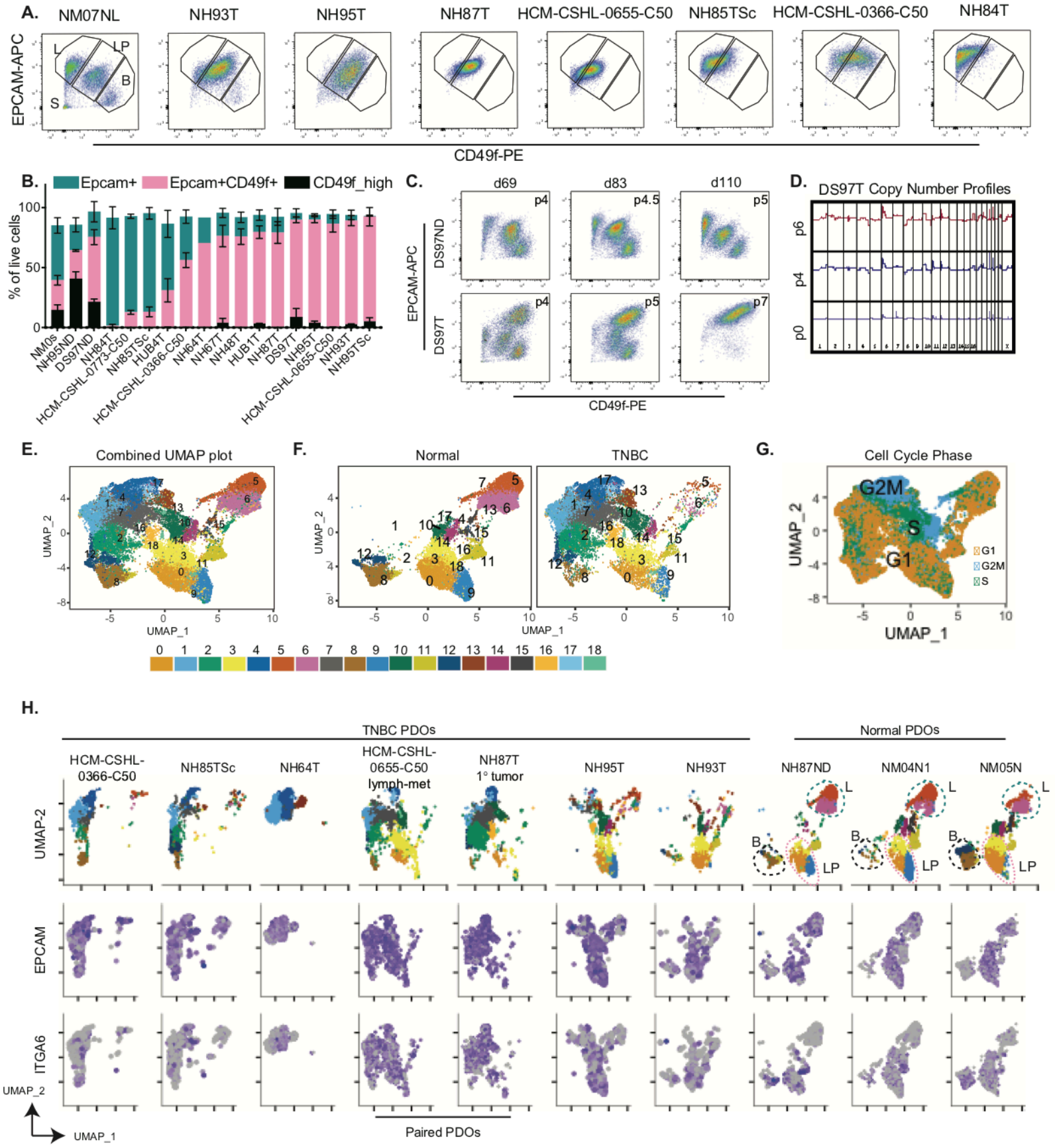

A.
NM07NL
NH93T
NH95T
NH87T
HCM-CSHL-0655-C50
NH85TSc
HCM-CSHL-0366-C50
NH84T
EPCAM-APC
L
LP
B
S
CD49f-PE
B.
Epcam+
Epcam+CD49f+
CD49f_high
% of live cells
NM0s
NH95ND
DS97ND
NH84T
HCM-CSHL-0773-C50
NH85TSc
HUB4T
HCM-CSHL-0366-C50
NH64T
NH67T
NH48T
HUB1T
NH87T
DS97T
NH95T
HCM-CSHL-0655-C50
NH93T
NH95TSc
C.
d69
d83
d110
DS97ND
DS97T
p4
p4.5
p5
p7
EPCAM-APC
CD49f-PE
D.
DS97T Copy Number Profiles
p6
p4
p0
E.
Combined UMAP plot
UMAP_1
UMAP_2
F.
Normal
TNBC
G.
Cell Cycle Phase
G2M
S
G1
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
UMAP_2
UMAP_1
Paired PDOs

Figure 7

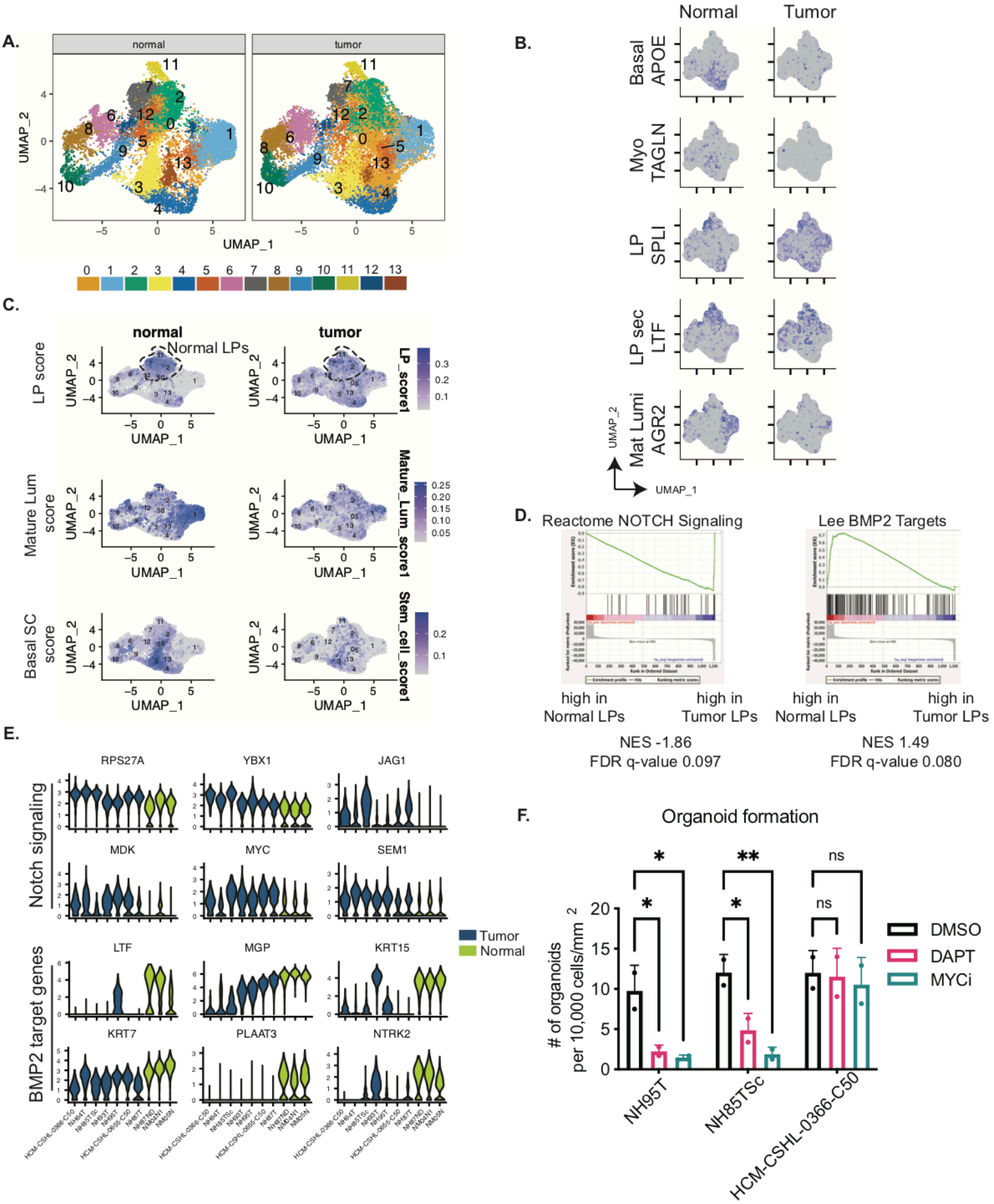

A.
normal
tumor
UMAP_2
UMAP_1
0 1 2 3 4 5 6 7 8 9 10 11 12 13
B.
Normal
Tumor
Basal APOE
Myo TAGLN
LP SPLI
LP sec LTF
Mat Lumi AGR2
UMAP_2
UMAP_1
C.
normal
tumor
Normal LPs
LP score
Mature Lum score
Basal SC score
LP_score1
Mature_Lum_score1
Stem_cell_score1
UMAP_1
UMAP_2
D.
Reactome NOTCH Signaling
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
Notch signaling
LTF
MGP
KRT15
KRT7
PLAAT3
NTRK2
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 8

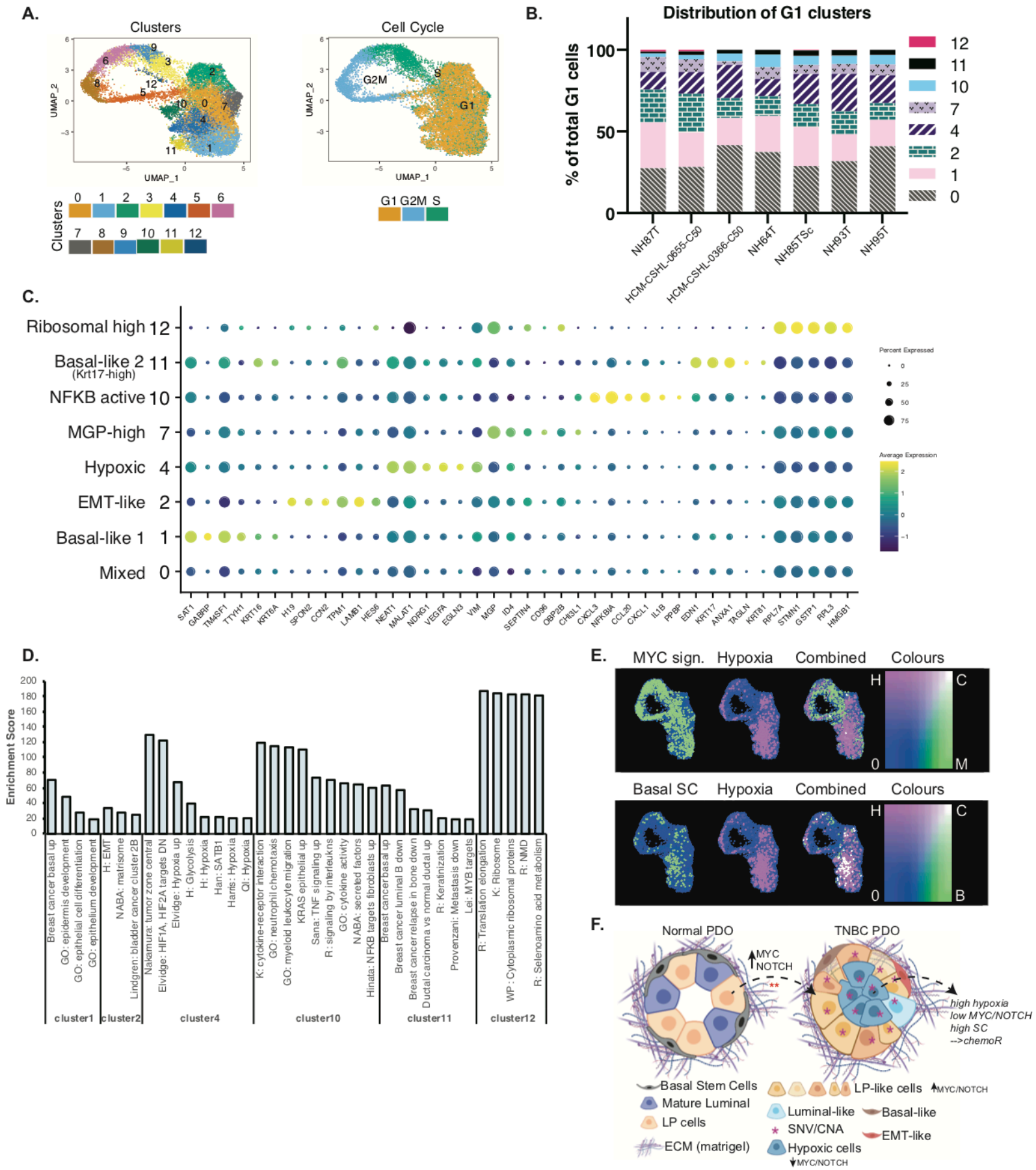

A.
Clusters
Cell Cycle
UMAP_1
UMAP_2
G2M
S
G1
Clusters
0 1 2 3 4 5 6
7 8 9 10 11 12
G1 G2M S
B.
Distribution of G1 clusters
% of total G1 cells
NH87T
HCM-CSHL-0655-C50
HCM-CSHL-0366-C50
NH64T
NH85TSc
NH93T
NH95T
C.
Ribosomal high 12
Basal-like 2 (Krt17-high) 11
NFKB active 10
MGP-high 7
Hypoxic 4
EMT-like 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

Fig S1

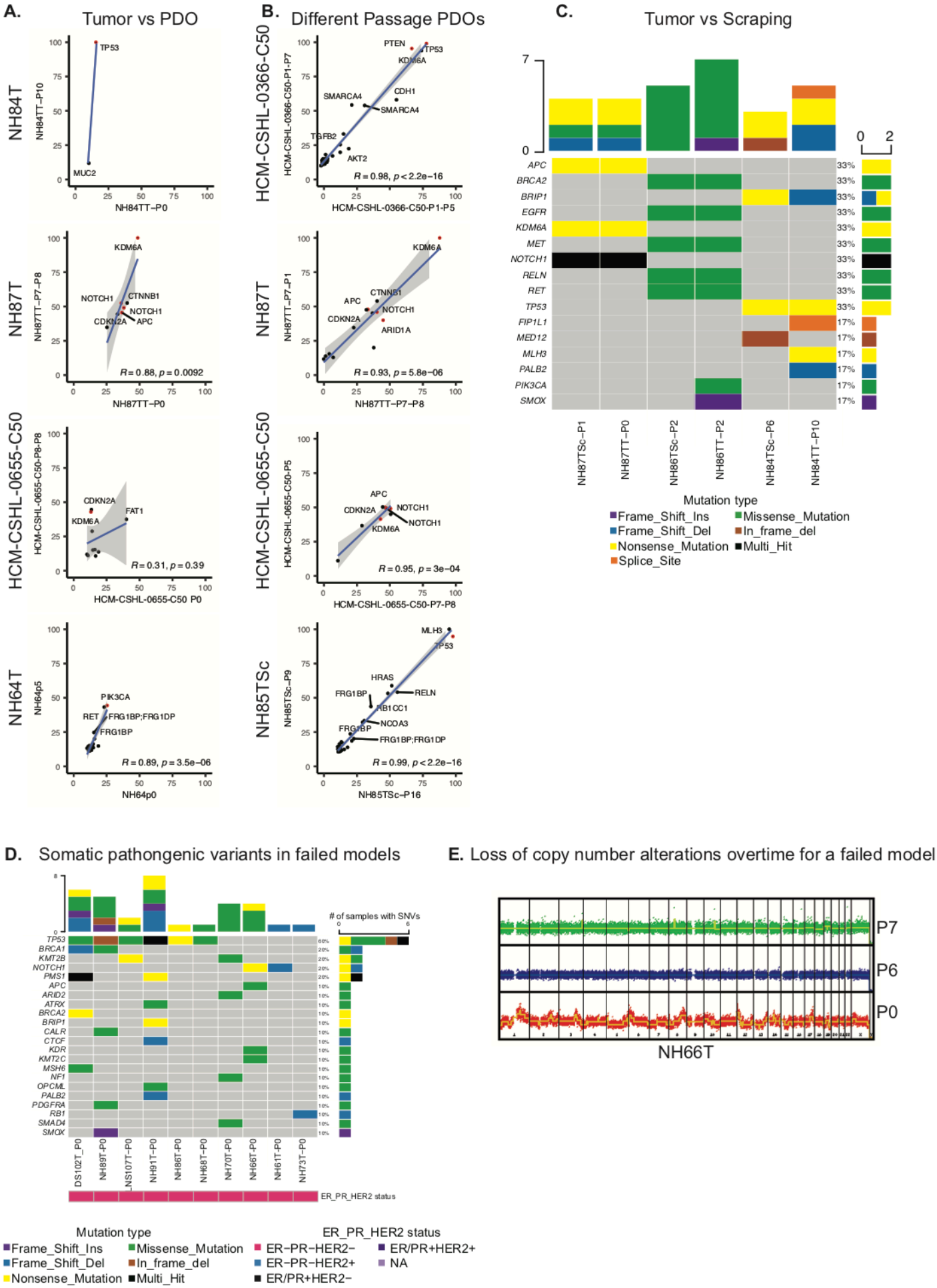

A.
Tumor vs PDO
NH84T
NH84TT−P10
NH84TT−P0
TP53
MUC2
NH87T
NH87TT−P7−P8
NH87TT−P0
KDM6A
NOTCH1
CTNNB1
CDKN2A
APC
R = 0.88, p = 0.0092
HCM-CSHL-0655-C50
HCM-CSHL-0655-C50-P8-P8
HCM-CSHL-0655-C50 P0
CDKN2A
KDM6A
FAT1
R = 0.31, p = 0.39
NH64T
NH64p5
NH64p0
PIK3CA
RET
FRG1BP;FRG1DP
FRG1BP
R = 0.89, p = 3.5e−06
B.
Different Passage PDOs
HCM-CSHL-0366-C50
HCM-CSHL-0366-C50-P1-P7
HCM-CSHL-0366-C50-P1-P5
PTEN
TP53
KDM6A
CDH1
SMARCA4
SMARCA4
TGFB2
AKT2
R = 0.98, p < 2.2e−16
NH87T
NH87TT−P7−P1
NH87TT−P7−P8
KDM6A
CTNNB1
APC
NOTCH1
CDKN2A
ARID1A
R = 0.93, p = 5.8e−06
HCM-CSHL-0655-C50
HCM-CSHL-0655-C50-P5
HCM-CSHL-0655-C50-P7-P8
APC
CDKN2A
NOTCH1
NOTCH1
KDM6A
R = 0.95, p = 3e−04
NH85TSc
NH85TSc−P9
NH85TSc−P16
MLH3
TP53
HRAS
FRG1BP
RELN
NB1CC1
NCOA3
FRG1BP
FRG1BP;FRG1DP
R = 0.99, p < 2.2e−16
C.
Tumor vs Scraping
APC
BRCA2
BRIP1
EGFR
KDM6A
MET
NOTCH1
RELN
RET
TP53
FIP1L1
MED12
MLH3
PALB2
PIK3CA
SMOX
33%
17%
NH87TSc−P1
NH87TT−P0
NH86TSc−P2
NH86TT−P2
NH84TSc−P6
NH84TT−P10
Mutation type
Frame_Shift_Ins
Missense_Mutation
Frame_Shift_Del
In_frame_del
Nonsense_Mutation
Multi_Hit
Splice_Site
D.
Somatic pathogenic variants in failed models
# of samples with SNVs
TP53
BRCA1
KMT2B
NOTCH1
PMS1
APC
ARID2
ATRX
BRCA2
BRIP1
CALR
CTCF
KDR
KMT2C
MSH6
NF1
OPCML
PALB2
PDGFRA
RB1
SMAD4
SMOX
DS102T_P0
NH89T-P0
NS107T-P0
NH91T-P0
NH86T-P0
NH68T-P0
NH70T-P0
NH66T-P0
NH61T-P0
NH73T-P0
ER_PR_HER2 status
Mutation type
Frame_Shift_Ins
Missense_Mutation
Frame_Shift_Del
In_frame_del
Nonsense_Mutation
Multi_Hit
ER_PR_HER2 status
ER−PR−HER2−
ER/PR+HER2+
ER−PR−HER2+
NA
ER/PR+HER2−
E.
Loss of copy number alterations overtime for a failed model
P7
P6
P0
NH66T

Fig S2

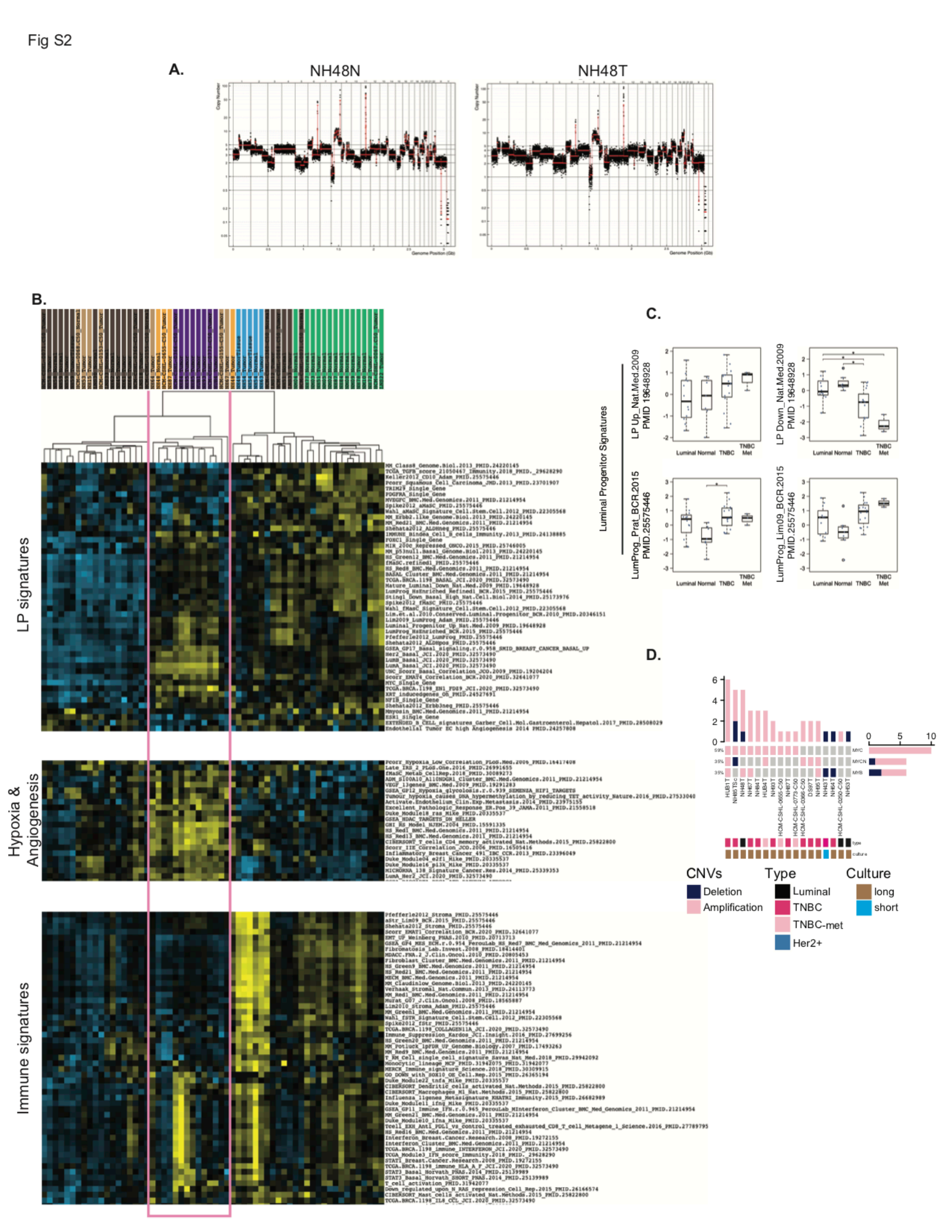

A.
NH48N
NH48T
Copy Number
Genome Position (Gb)
B.
LP signatures
Hypoxia & Angiogenesis
Immune signatures
C.
Luminal Progenitor Signatures
LP Up_Nat.Med.2009 PMID 19648928
LP Down_Nat.Med.2009 PMID 19648928
LumProg_Prat_BCR.2015 PMID.25575446
LumProg_Lim09_BCR.2015 PMID.25575446
Luminal Normal
TNBC
TNBC Met
D.
MYC
MYCN
MYB
type
culture
CNVs
Deletion
Amplification
Type
Luminal
TNBC
TNBC-met
Her2+
Culture
long
short

Fig S3

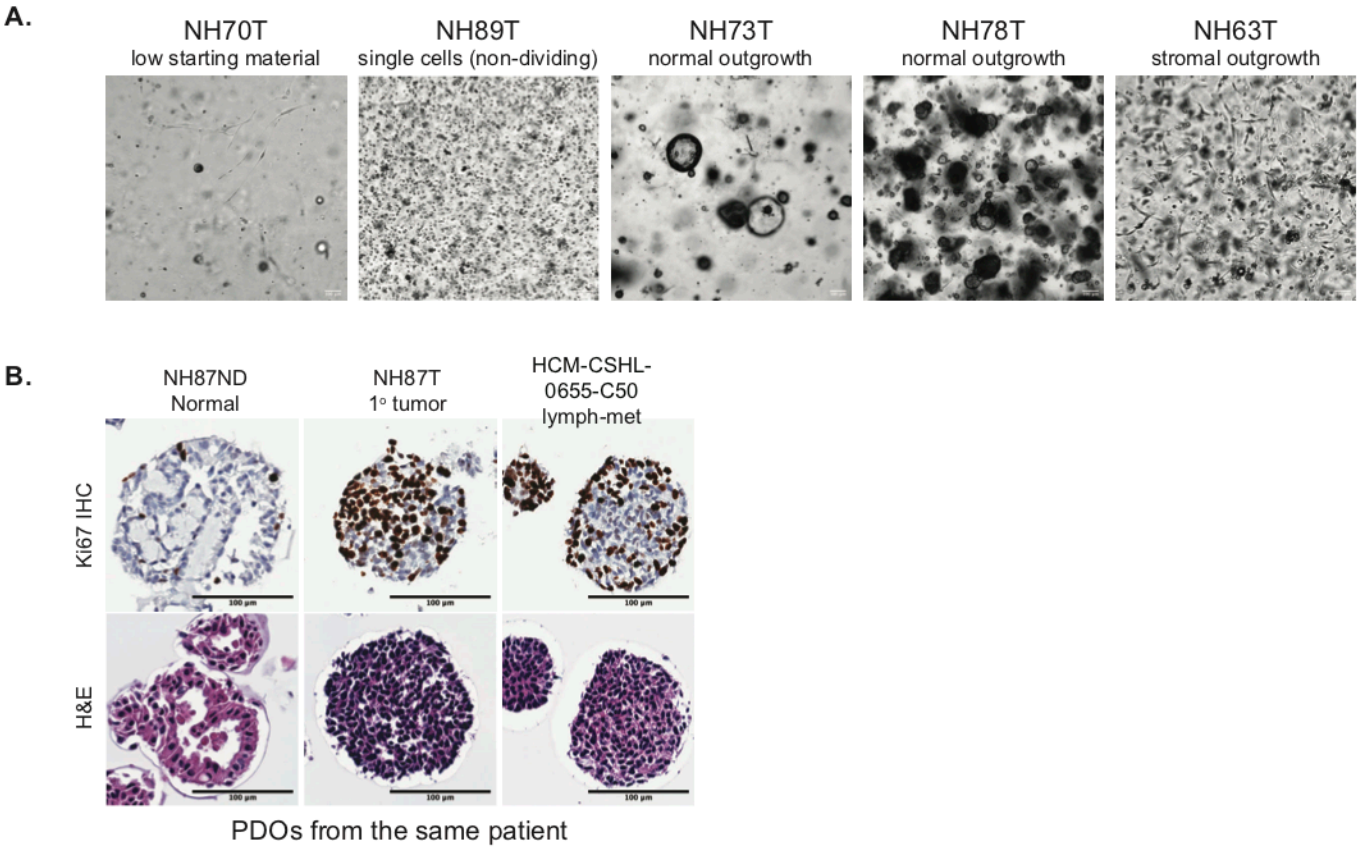

A.
NH70T
low starting material
NH89T
single cells (non-dividing)
NH73T
normal outgrowth
NH78T
normal outgrowth
NH63T
stromal outgrowth
B.
NH87ND
Normal
NH87T
1° tumor
HCM-CSHL-
0655-C50
lymph-met
Ki67 IHC
H&E
PDOs from the same patient

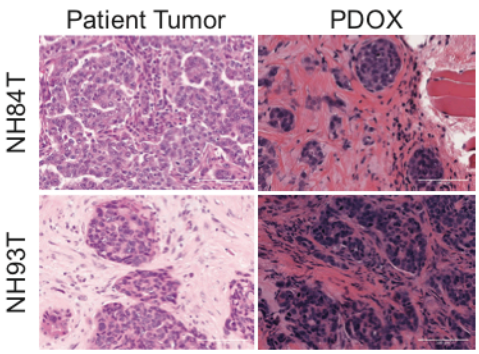
Fig S4

A.

	PDO transplant #1			PDO transplant #2			Metastasis (assayed by hMito IHC)			
	N1			N2						
	PDO passage	Experiment design	Primary tumors observed	PDO passage	Experiment design	Primary tumors observed	Lung mets	Lymph node mets	Liver	Brain
NH87T	(p6)p6	3 mice: 6 injections	6 of 6	(p6)p12	3 mice: 6 injections	6 of 6	micro-mets	single-cell	no	no
HCM-CSHL-0655-C50	(p7)p4	3 mice: 6 injections	4 of 6	(p4)p6	3 mice: 6 injections	6 of 6	no	single-cell	single-cell	no
HCM-CSHL-0366-C50	(p1)p5p4	3 mice: 6 injections	6 of 6	(p4)p7	3 mice: 6 injections	6 of 6	micro-mets	single-cell	single-cell	no
NH85TSc	(p5)p3p2	3 mice: 6 injections	6 of 6	(p5)p3p4	3 mice: 6 injections	6 of 6	micro-mets	single-cell	single-cell	no
NH95T	(p8)p2	3 mice: 6 injections	5 of 6	(p8)p2	3 mice: 6 injections	4 of 6	No	No	No	No
NH84T	(p6)p5	2 mice: 4 injections	0 of 4	(p6)p12	3 mice: 6 injections	1 of 6#	No	No	No	No
NH93T	(p5)p4	2 mice: 4 injections	2 of 4#	(p5)p10	3 mice: 6 injections	4 of 6#	No	No	No	No
NH64T	(p4)p2	2 mice: 4 injections	0 of 3	(p4)p2 *	3 mice: 6 injections	0 of 6	No	No	No	No

* different thaw; BM: Bone Marrow; hMito: human mitochondria
\# only observable by histology of fat-pad tissue; Bold samples: paired PDOs from same patient

B.



C.

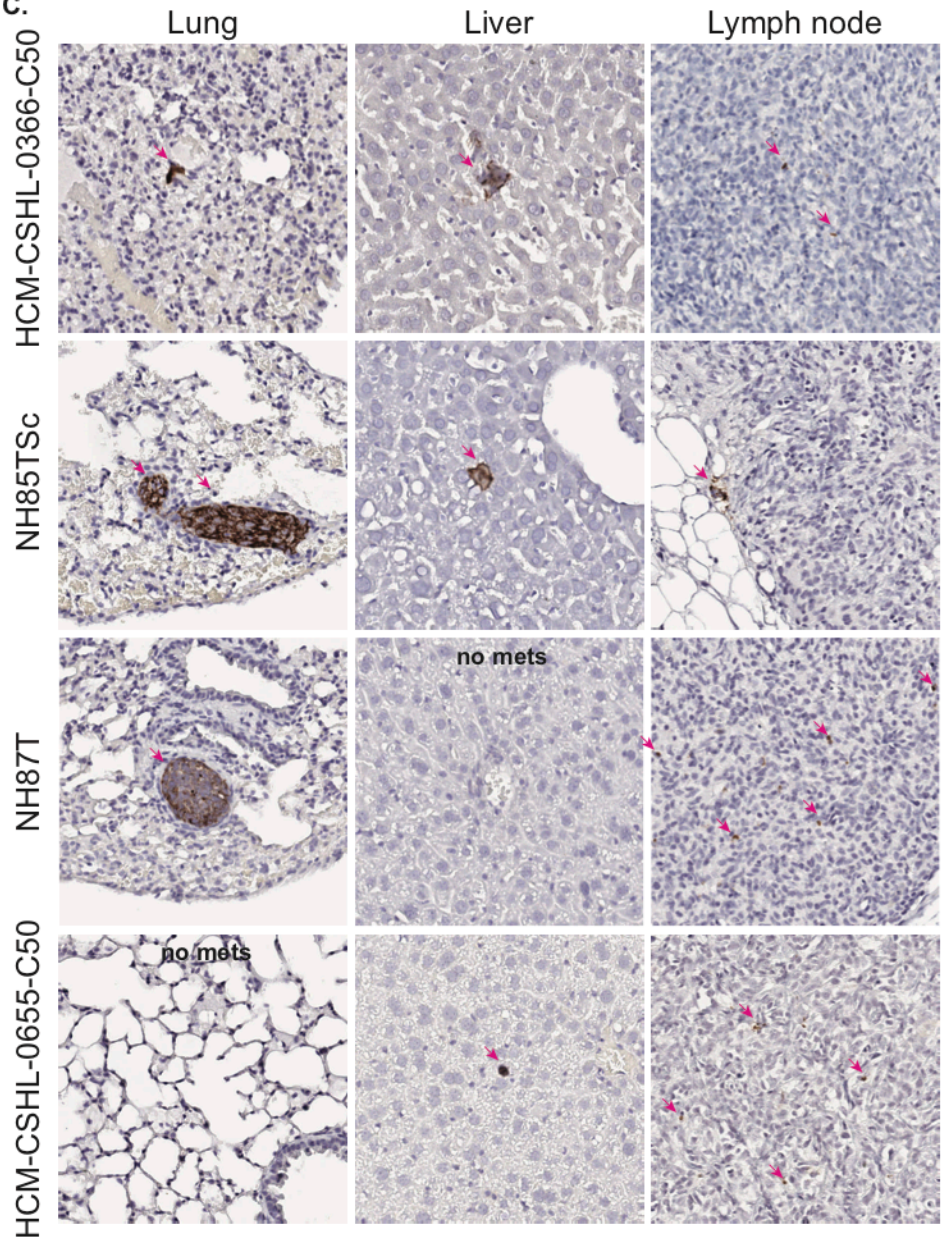
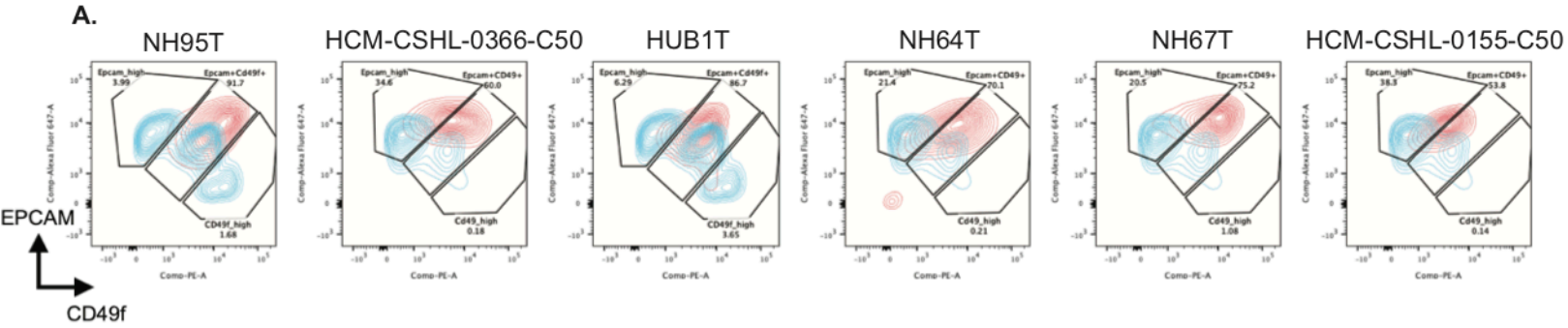
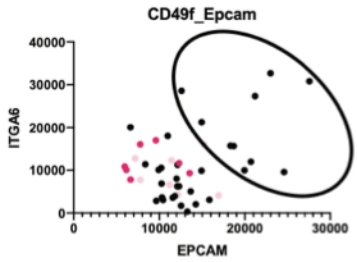
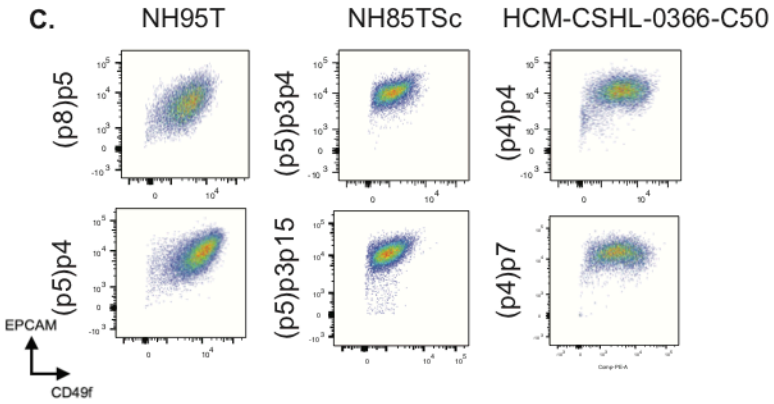
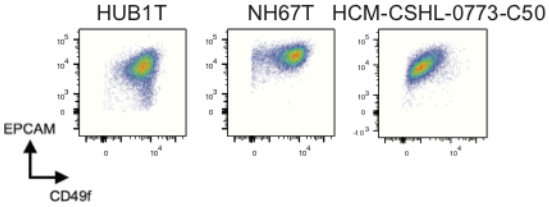


Fig S5









E.

sample_name	type	batch	patientID
NH64T	tumor	batch_1	NH64
NM04N	normal	batch_1	NM04
NM05N	normal	batch_1	NM05
NH87T	tumor	batch_2	NH87
NH87ND	normal	batch_2	NH87
HCM-CSHL-0655-C50	tumor	batch_2	NH87
NH85T	tumor	batch_3	NH85
NH93T	tumor	batch_3	NH93
NH95T	tumor	batch_3	NH95
HCM-CSHL-0366-C50	tumor	batch_3	HCM-CSHL-0366-C50

Fig S6

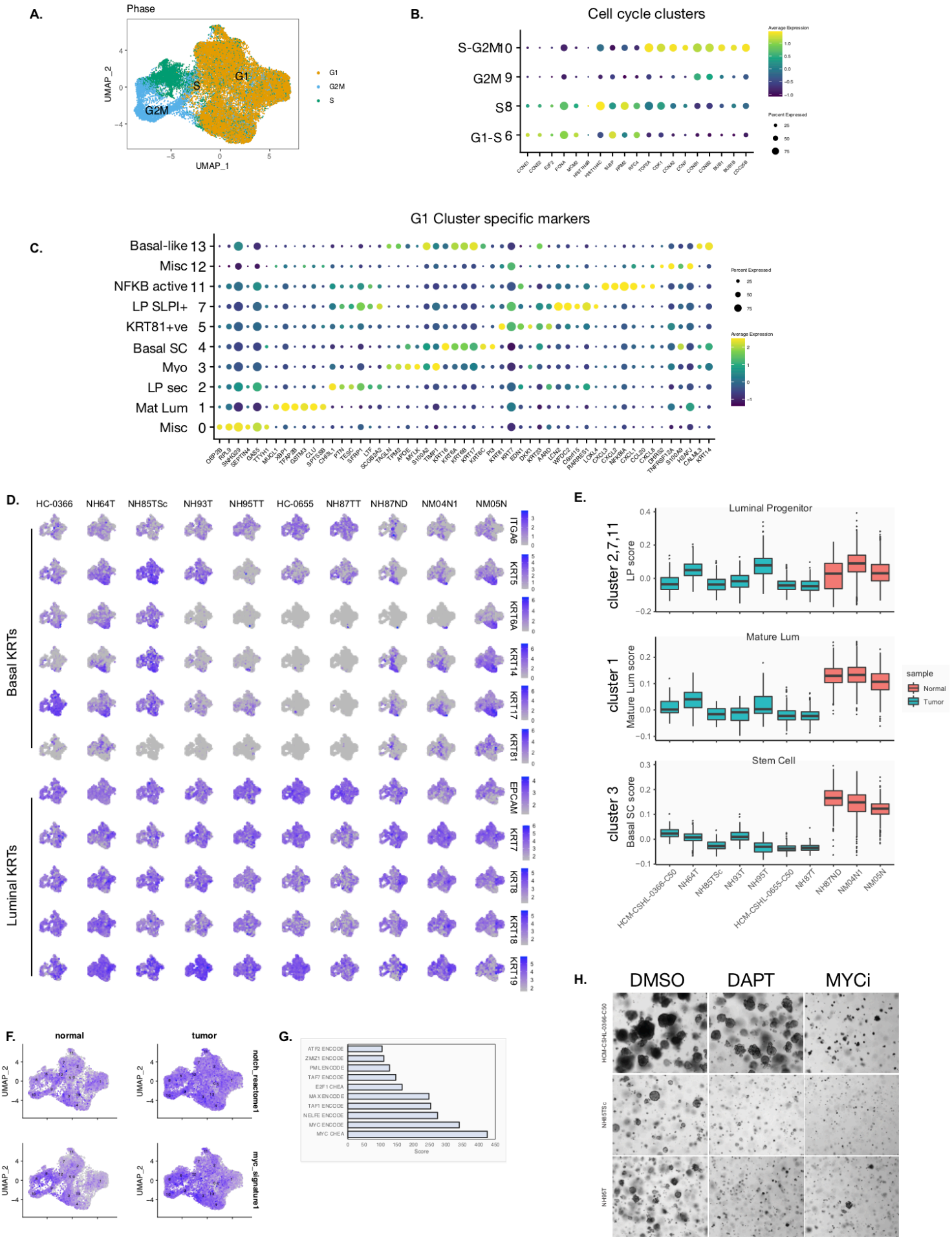
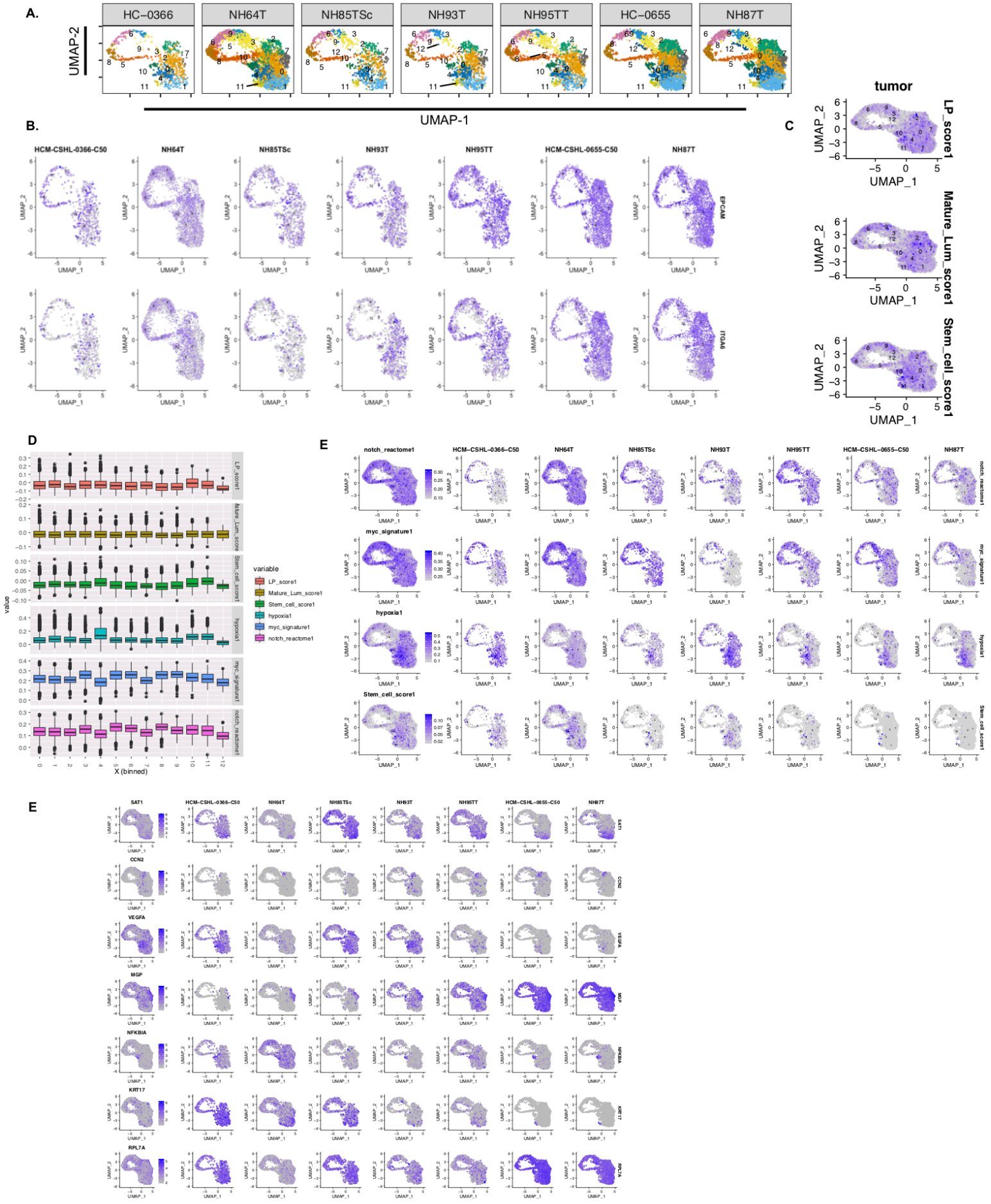

A.
Phase
UMAP_2
UMAP_1
G1
S
G2M
B.
Cell cycle clusters
S-G2M 10
G2M 9
S 8
G1-S 6
Average Expression
Percent Expressed
C.
G1 Cluster specific markers
Basal-like 13
Misc 12
NFKB active 11
LP SLPI+ 7
KRT81+ve 5
Basal SC 4
Myo 3
LP sec 2
Mat Lum 1
Misc 0
Percent Expressed
Average Expression
D.
HC-0366
NH64T
NH85TSc
NH93T
NH95TT
HC-0655
NH87TT
NH87ND
NM04N1
NM05N
Basal KRTs
Luminal KRTs
ITGA6
KRT5
KRT6A
KRT14
KRT17
KRT81
EPCAM
KRT7
KRT8
KRT18
KRT19
E.
Luminal Progenitor
cluster 2,7,11
LP score
Mature Lum
cluster 1
Mature Lum score
Stem Cell
cluster 3
Basal SC score
sample
Normal
Tumor
HCM-CSHL-0366-C50
NH64T
NH85TSc
NH93T
NH95T
HCM-CSHL-0655-C50
NH87T
NH87ND
NM04N1
NM05N
F.
normal
tumor
UMAP_2
notch_reactome1
myc_signature1
G.
ATF2 ENCODE
ZMIZ1 ENCODE
PML ENCODE
TAF7 ENCODE
E2F1 CHEA
MAX ENCODE
TAF1 ENCODE
NELFE ENCODE
MYC ENCODE
MYC CHEA
Score
H.
DMSO
DAPT
MYCi
HCM-CSHL-0366-C50
NH85TSc
NH95T

Fig S7


A.
HC−0366
NH64T
NH85TSc
NH93T
NH95TT
HC−0655
NH87T
UMAP-2
UMAP-1
B.
HCM-CSHL-0366-C50
NH64T
NH85TSc
NH93T
NH95TT
HCM-CSHL-0655-C50
NH87T
EPCAM
ITGA6
C
tumor
LP_score1
Mature_Lum_score1
Stem_cell_score1
UMAP_1
UMAP_2
D
LP_score1
Mature_Lum_score
Stem_cell_score1
hypoxia1
myc_signature1
notch_reactome1
value
X (binned)
variable
LP_score1
Mature_Lum_score1
Stem_cell_score1
hypoxia1
myc_signature1
notch_reactome1
E
notch_reactome1
myc_signature1
hypoxia1
Stem_cell_score1
HCM−CSHL−0366−C50
NH64T
NH85TSc
NH93T
NH95TT
HCM−CSHL−0655−C50
NH87T
E
SAT1
CCN2
VEGFA
MGP
NFKBIA
KRT17
RPL7A