HHS Public Access

Author manuscript

Mol Cancer Res. Author manuscript; available in PMC 2022 April 01.

Published in final edited form as:

Mol Cancer Res. 2021 October ; 19(10): 1699–1711. doi:10.1158/1541-7786.MCR-21-0215.

HER2 isoforms uniquely program intratumor heterogeneity and predetermine breast cancer trajectories during the occult tumorigenic phase

Joshua D. Ginzel[#1], **Chaitanya R. Acharya**[#2], **Veronica Lubkov**[#1,2], **Hidetoshi Mori**[3], **Peter G. Boone**[1,2], **Lauren K. Rochelle**[1], **Wendy Roberts**[1], **Jeffrey I. Everitt**[4], **Zachary C. Hartman**[2], **Erika J. Crosby**[2], **Lawrence S. Barak**[1], **Marc G. Caron**[1], **Jane Q. Chen**[3], **Neil E. Hubbard**[3], **Robert D. Cardiff**[3,6], **Alexander D. Borowsky**[3,6], **H. Kim Lyerly**[2,5,6], **Joshua C. Snyder**[*,1,2,6]

[1]Department of Cell Biology, Duke University Medical Center, Durham, North Carolina, USA

[2]Department of Surgery, Division of Surgical Sciences, Duke University Medical Center

[3]Department of Pathology and Laboratory Medicine and The Center for Immunology and Infectious Disease, University of California-Davis

[4]Department of Pathology, Duke University Medical School, Durham, North Carolina, USA.

[5]Department of Immunology, Duke University School of Medicine

[6]These authors jointly supervised this work.

[#] These authors contributed equally to this work.

Abstract

Her2 positive breast cancers are among the most heterogeneous breast cancer subtypes. The early amplification of Her2 and its known oncogenic isoforms provide a plausible mechanism in which distinct programs of tumor heterogeneity could be traced to the initial oncogenic event. Here a Cancer rainbow mouse simultaneously expressing fluorescently barcoded wildtype (^{WT}HER2), exon-16 null (^{d16}HER2), and N-terminally truncated (^{p95}HER2) HER2 isoforms is used to trace tumorigenesis from initiation to invasion. Tumorigenesis was visualized using whole-gland fluorescent lineage tracing and single-cell molecular pathology. We demonstrate that within weeks of expression, morphologic aberrations were already present and unique to each HER2 isoform. Although ^{WT}HER2 cells were abundant throughout the mammary ducts, detectable lesions were exceptionally rare. In contrast, ^{d16}HER2 and ^{p95}HER2 induced rapid tumor development. ^{d16}HER2 incited homogenous and proliferative luminal-like lesions which

[*]Correspondence to Joshua C. Snyder, Ph.D., joshua.snyder@duke.edu, (919)684-3162, Duke University Medical Center, Box 2606, Durham, NC 27710.

Author Contributions

J.D.G., C.R.A, and V.L. performed experiments, collected data, analyzed data, and wrote the manuscript. H.M, P.G.B., and L.K.R. performed experiments and analyzed data. W.R. provided mouse husbandry expertise. J.I.E. performed histopathology and necropsy. E.J.C. and Z.C.H. provided conceptual and experimental advice for mammary cancer mouse modeling. L.S.B. and M.G.C. provided imaging and experimental design support. J.Q.C. and N.E.H. performed histopathology and specimen biobanking. R.D.C., A.D.B, H.K.L, and J.C.S. designed the experiments, supervised the work, and wrote the manuscript.

Declaration of Interests

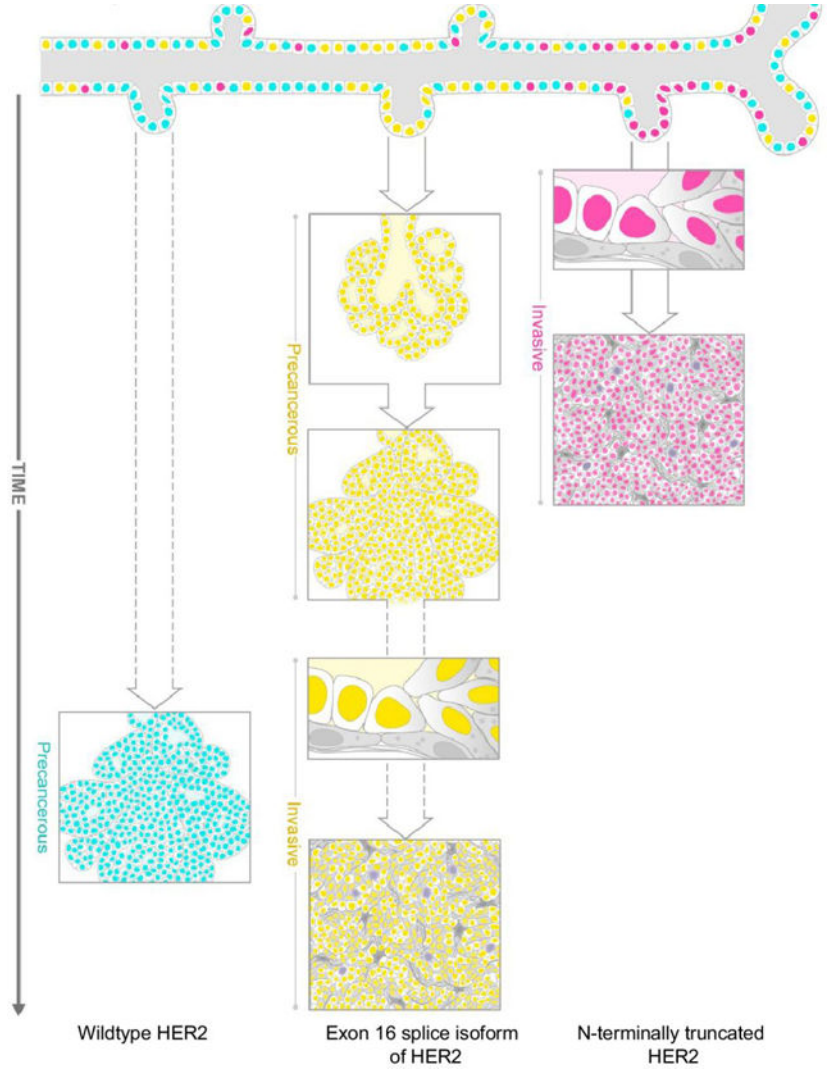
No declarations to declare.

No financial conflicts to declare.

infrequently progressed to invasive phenotypes whereas ^{p95}HER2 lesions were heterogenous and invasive at the smallest detectable stage. Distinct cancer trajectories were observed for ^{d16}HER2 and ^{p95}HER2 tumors as evidenced by oncogene dependent changes in epithelial specification and the tumor microenvironment. These data provide direct experimental evidence that intratumor heterogeneity programs begin very early and well in advance of screen or clinically detectable breast cancer.

Implications—Although all HER2 breast cancers are treated equally, we show a mechanism by which clinically undetected HER2 isoforms program heterogenous cancer phenotypes through biased epithelial specification and adaptations within the tumor microenvironment.

Graphical Abstract



Introduction

Amplification of the protooncogene receptor tyrosine-protein kinase erbB-2 (*HER2*) is a very early event found in approximately 20% of breast cancers and is prognostic of

poorest outcome(1–4). Treatment of *HER2* amplified breast cancer has been revolutionized by targeted HER2 therapy (5). However, most patients with metastatic HER2+ breast cancer will eventually progress and die (6–8). One explanation could be that diagnosis of HER2+ breast cancer relies upon simple measures of HER2 overexpression using immunohistochemistry or *in situ* hybridization assays. These assays fail to capture the true heterogeneity present within HER2+ tumors. For instance, molecular stratification shows that approximately half of HER2+ breast cancers are more similar to Luminal A, Luminal B, or basal-like subtypes and that this heterogeneity is predictive of treatment resistance(9).

Expression of alternative HER2 isoforms may provide a mechanistic basis for differences in heterogeneity and therapy resistance encountered in HER2+ breast cancers. Mouse models engineered to overexpress wild-type *HER2* (*^{WT}HER2*) develop tumors more slowly compared to mice expressing activating mutations(10). Even slow growing *^{WT}HER2* tumors occur only after secondary stochastic mutations to the extracellular domain(11). These oncogenic *HER2* isoforms, such as the *HER2* isoform lacking exon 16 (*^{d16}HER2*), incite mammary tumors in mice (12–16). Another variant, the carboxyl terminal fragment of *HER2* (*^{p95}HER2, aa611–1255*), has also been reported in a subset of breast cancers. Tumors expressing *^{p95}HER2* are resistant to trastuzumab due to the loss of the specific extracellular binding domain recognized by trastuzumab (17–20). These *HER2* isoforms can be expressed heterogeneously throughout the tumor epithelium and can also vary between patients and cancer types(21–23).

The heterogeneity of HER2 isoforms has never been fully reconstructed in an *in vivo* model. This has led to unresolved differences between HER2 models, even those expressing the same oncogene. For instance, Castagnoli and colleagues(15) described tumors from transgenic MMTV-^{d16}HER2 mice as a “solid pattern” that often resemble standard ^{WT}HER2 tumor phenotypes. In contrast, Turpin and colleagues(12) described their tetracycline inducible ^{d16}HER2 tumors as markedly different from HER2/neu phenotypes due to the appearance of metaplastic changes in the tumor and an associated stem-like transcriptional profile. One reason for this discrepancy could be that potential plasticity changes induced by each HER2 isoform have not been adequately visualized and quantitatively assessed throughout the entire tumorigenic process.

We recently developed a Cancer rainbow (Crainbow) mouse modelling platform for performing functional genomics and lineage tracing cell competition in a single mouse(24). Here we report on the implementation of a *HER2* Crainbow mouse (*HER2BOW*) to visualize the effects of HER2 isoform heterogeneity *in vivo*. *HER2BOW* mice enable controlled determination of progression timescales and tumor phenotypes for three *HER2* isoforms in a single, immune intact mouse. We use whole-gland FUnGI lineage tracing(25), molecular pathology, and scRNAseq in order to visualize the genetic origin and cellular trajectory during the prediagnostic phase of breast cancer. Our work ultimately illustrates how the expression of HER2 isoforms drive immediate, yet subtle differences in the mammary duct, that manifest as distinct programs of cancer heterogeneity.

Methods

Animal Studies

All animal experiments conformed to regulatory standards and were performed according to a Duke Institutional Animal Care and Use Committee-approved protocol. The Duke Transgenic and Knockout Mouse core targeted the mouse ROSA locus using a HER2 Crainbow targeting vector in accordance with previously published protocols(24,26). Crainbow-HER2 mice were crossed with MMTV-cre/Line D (MGI Cat# 3581599, RRID:MGI:3581599) from Jackson Laboratory and backcrossed to FVB/N background using Taconic speed congenics services. Animals used in the study were backcrossed for more than 6 generations using speed congenics. All experimental animals were female.

Microscopy and image analysis

Tissues were fixed with 4% neutral buffered formalin (NBF) for at least 48 hours and then transferred to ethanol prior to embedding. For confocal imaging of sectioned tumor tissues, fixed tissue was embedded in 4% agar and then sectioned in 100μm slices with a Vibratome. Sections were mounted onto glass slides and tiled images at 20× were acquired on a Zeiss LSM 880 microscope. To estimate the signal from Her2 reporters, the RGB image was manually segmented using custom-designed MATLAB (MATLAB,RRID:SCR_001622) software and the number of red (p95), green (d16), blue (WT) positive pixels were counted in the area of interest (lesion).

Histopathology

Samples were paraffin-embedded, sectioned at 4 μm, mounted on glass slides, and stained with Mayer's hematoxylin and eosin. Stained slides were scanned on an Aperio, and digital images were viewed using the ImageScope application (ImageScope, RRID:SCR_014311). Small foci (>1.25×1.25mm) were considered MIN and classified according to cell type, MIN border, inflammation and fibrosis. The inflammation was generally in the peripheral stroma but in more severe cases, infiltrated into the neoplasm. In a like manner the amount of fibrosis was scored 0–4+ depending on the relative amount of connective tissue surrounding and invading the neoplasm. Inflammation and fibrosis were scored on a sliding scale from none (0) to the most (4+). The borders of the lesions were either expansile or invasive. An expansile border did not invade, but compressed the surrounding tissues. Invasive invaded the surrounding tissues. Inflammation and fibrosis were scored on an arbitrary scale from none (0) to worse (4+). Immunohistochemistry (IHC) was performed as previously described(27). Antibodies used were anti- C-terminus of HER2 (Leica Biosystems Cat# NCL-CB11, RRID:AB_442046), anti-N-terminus HER2 (Cell Signaling Technology Cat# 8609, RRID:AB_10829036), anti-SMA (Abcam Cat# ab5694, RRID:AB_2223021), anti-Vimentin (Cell Signaling Technology Cat# 5741, RRID:AB_10695459), anti-CD3 (LSBio (LifeSpan) Cat# LS-C31900–100, RRID:AB_904742), anti-CD163 (Abcam Cat# ab182422, RRID:AB_2753196), CK8/18 (Fitzgerald Industries International Cat# 20R-CP004, RRID:AB_1284055), and CK14 (Thermo Fisher Scientific Cat# PA5–109330, RRID:AB_2854741).

Multiplex RNA FISH.—Advanced Cell Diagnostics RNAscope® Multiplex Fluorescent kit (323100) was used to detect up to 4 mRNA targets *in situ*. FFPE tissue sections were prepared and pretreated prior to probe hybridization and signal development following the Multiplex Fluorescent kit user manual. Probe labelling utilized Opal 7-color Kit dyes (PerkinElmer, Inc NEL801001KT). Tiled images were acquired for each channel on a Zeiss LSM 880 using a 40x objective. For RNA FISH quantification, Cell Profiler 3.0 was used to filter, threshold and watershed images in order to count transcript spots(28). The following probes were used for Multiplex RNA FISH: *mKO* (RNAscope Multiplex Probe - mKO-fp 540071, 488nm), *eYFP* (RNAscope Multiplex Probe - EYFP-C2 312131-C2, 550nm), *mTFP* (RNAscope Multiplex Probe - mTFP1-C3 500271-C3, 633nm), *Krt14* (RNAscope Multiplex Probe – Mm-Krt14-C2 422521-C2), *Krt8* (RNAscope Multiplex Probe – Mm-Krt8-C3 424521-C3), *Aldoc* (RNAscope Probe - Mm-Aldoc-C4 429531-C4), and *Il33* (RNAscope Probe - Mm-Il33 400591). Where necessary HER2BOW native fluorescent proteins could be imaged in FFPE tissue without RNA FISH, which we termed True Registry (TruRegistry). FFPE slides are washed twice for 5 minutes in 100% xylene and then washed twice for 2 minutes in 100% EtOH to remove paraffin wax. After EtOH has fully dried at room temperature, slides are DAPI stained and mounted. Native fluorescent proteins are imaged on a Zeiss LSM 880 confocal microscope.

Multiplex immunohistochemistry (mIHC)

Preparation of FFPE tissue sections was performed by following previously described procedure(29). The procedure for mIHC was followed by a manufacturer's protocol for Opal™ 7-color manual IHC kit (AKOYA BIOSCIENCE) with an additional procedure to reduce autofluorescence(29). Panel for mIHC contains CD3, CK8/18, CK14, N-terminus (extracellular domain: Ext) of HER2 and CD163. Slides stained for mIHC were scanned with multispectral imaging microscope (Vectra 3.0; Perkin Elmer), and unmixed multispectral images were performed with inForm ver. 2.4.0. Tissue segmentation, cell segmentation and phenotyping were performed as described previously(30). Tissue segmentation was performed with CK8/18 and CK14 positive cells to separate tumor and stroma. Cell segmentation was performed to separate nuclei, cytoplasm and membrane by using DAPI, keratins and HER2 staining respectively. To profile HER2 status, we used Ext-HER2 positiveness for phenotyping d16 (stained with Ext-HER2) and p95 (negative for Ext-HER2). The genotyping of each tumor colony was confirmed with ISH on a serial tissue section.

Whole Mount Mammary Gland Imaging.—Thoracic (third) glands were harvested from HER2BOW mice at 4, 8, and 16 weeks of age. For imaging of Crainbow barcodes and Keratin 14, clarification and staining of wholemount mammary glands were performed according to the previously described FUnGI protocol(25). Rabbit anti-Keratin-14 antibody (Lab Vision Cat# RB-9020-R7, RRID:AB_149789) was diluted 1:500 in wash buffer (PBS, 0.1% Triton, 0.02% SDS, 0.2% BSA, 50 μg/mL ascorbic acid and 0.05 ng/mL L-Glutathione). Tissues were incubated in primary antibody solution on an orbital shaker at 4°C for 2 days. Tissue was then washed and incubated with Goat anti-Rabbit IgG AlexaFluor 633 (Thermo Fisher Scientific Cat# A-21070, RRID:AB_2535731) was diluted 1:500 in wash buffer on an orbital shaker at 4°C overnight. After staining, tissues were

incubated in FUnGI solution at 4°C overnight. Imaging of clarified tissue was performed on a Zeiss LSM 880 confocal microscope at 10x and 40x. 3D rendering was performed using Imaris software (Imaris, RRID:SCR_007370).

Single Cell Capture.—Tumors were digested in collagenase (1mg/mL), Dnase (20 U/mL), and hyaluronidase (100 μg/mL) for 1.5 hours at 37°C. For a single cell solution, cell pellet from the digestion was resuspended in 1x trypsin with Dispase for up to 1 hour at 37°C. Cells were passed through a 40 μm filter and then resuspended in 0.1% BSA/1:250 anti-clumping agent with 10% Viahance viability dye. Cells were placed on a DynaMag™ Spin Magnet for 10 minutes before supernatant was removed and passed through a 20 μm filter. Single cells were resuspended in 0.1% BSA at 3000 cells/μL. ddSEQ Single-Cell Isolator (Biorad 12004336) was used to capture single cells according to manufacturer's protocol. Library preparation was performed using a compatible SureCell™ WTA 3' Library Prep Kit (Illumina, 220014279). Libraries were sequenced on an Illumina NextSeq 500 (Illumina, San Diego CA) in Hi-output mode at the Duke University Sequencing and Genomic Technologies Shared Resource. Raw sequence files were uploaded to BaseSpace Sequence Hub (BaseSpace, RRID:SCR_011881).

Mouse single-cell RNA sequencing (scRNAseq) data preprocessing.—BaseSpace (Illumina) was used to demultiplex cellular barcodes from 12 tumor samples, map reads to mouse transcriptome, down-sample reads as required to generate normalized aggregate data across samples, resulting in unique molecular identifier (UMIs) matrix of gene counts. Raw data is deposited at Gene Expression Omnibus (Gene Expression Omnibus (GEO), RRID:SCR_005012) under the accession id GSE152422. Thus, UMI gene count matrices, which was generated for all individual tumors in this study, were further processed in Seurat (version 3.1.3) (SEURAT, RRID:SCR_007322) (31). For each individual dataset, scores for S and G2/M cell cycle phases were computed based on previously defined gene sets using *CellCycleScoring* function. After log-normalizing individual tumor datasets, data were scaled after regressing out total number of UMI counts, percent mitochondrial gene expression and cell-cycle phase in order to account for the variability in gene expression. We selected top 5,000 highly variable genes from each individual tumor dataset immediately following variance-stabilizing transformation of the data. We then rank genes by the number of datasets they appear in (breaking ties by the median rank across datasets) and picked top 5,000 genes that can be used for data integration. We followed Seurat's reference-based integration workflow to find "anchorset" using *FindIntegrationAnchors* function, and integrate all the datasets using *IntegrateData* function. Scaled z-scores for each gene within the integrated data were calculated using *ScaleData* function, which were later used as input to principal component analysis (PCA). Ribosomal and mitochondrial genes/proteins were removed from the list of variable genes in order to prevent clustering based on ribosomal and mitochondrial content. The resulting integrated dataset had 4,960 genes and 8,486 cells. Clusters were identified using shared nearest neighbor (SNN)-based clustering on the basis of the first 20 principal components with k = 30 and resolution = 0.4. The same principal components were used to generate UMAP projections, which were generated with a minimum distance of 1 and 50 neighbors using cosine distance as the metric. A total of 3,843 epithelial cells were identified based on Epcam gene expression

and re-clustered as previously described. For UMAP visualization(32), we used the first 20 principal components, a minimum distance of 1 and 50 neighbors.

Annotation of cell clusters and cell states from single-cell trajectories

Individual clusters were annotated based on expression of known marker genes for epithelial cells (Epcam), endothelial cells (Cd34 and Vwf), immune cells (Ptprc/Cd45) and stromal fibroblasts (Vim). Different cell states of $HER2^{p95}$ and $HER2^{d16}$ epithelial cells obtained from single-cell trajectory mapping were annotated using a list of curated genesets obtained from literature search. Individual cells were then scored for enrichment of gene signatures using *AddModuleScore* function from the R package Seurat.

Trajectory mapping of $HER2^{p95}$ and $HER2^{d16}$ epithelial cells.—Monocle 2 (version 2.12.0) (Monocle2, RRID:SCR_016339)(33) was used for the trajectory analysis of the single cells. Trajectory mapping was performed separately for $HER2^{p95}$ and $HER2^{d16}$ cells. For each genotype, integrated and clustered data from Seurat were imported by Monocle 2. Following cluster-specific feature selection (dpFeature) at 1% FDR, dimensionality reduction (Reversed Graph Embedding)(34) and manifold learning, a tree-like structure was constructed that describes a projected path of gene expression for a given cell.

Statistical analyses.—All statistical analyses were performed in R (version 3.5.1) and GraphPad Prism (version 8.4.3) (GraphPad Prism, RRID:SCR_002798). Statistical significance for all the analyses was accepted at p-value ≤ 0.05 unless otherwise stated. All gene marker identifications (both cluster-specific and cellular trajectory state-specific) were performed using non-parametric Wilcoxon test. Gene expression was ordered along pseudotime (obtained from Monocle trajectory mapping) and the non-linear gene expression trend over pseudotime is fitted using generalized additive models.

Data Availability.—All the necessary R scripts from our analyses are made available as a Code Ocean compute capsule (https://codeocean.com/capsule/1839143).

Results

HER2 isoforms encode distinct developmental trajectories of cancer progression.

A *HER2* Crainbow (*HER2BOW*) ROSA-targeting vector was engineered using the Crainbow mouse modeling platform (24). The *HER2BOW* vector encodes Cre recombinase-inducible and fluorescently barcoded *HER2* variants – ^{WT}*HER2* co-expressed with nuclear *mTFP1* (cyan), ^{d16}*HER2* co-expressed with nuclear *EYFP* (yellow), and ^{p95}*HER2* co-expressed with nuclear *mKO* (magenta) (Fig. 1a–b, Supplementary Figure 1 and Supplementary File 1). Gene targeting in mouse embryonic stem cells established a viable *HER2BOW* mouse line. *HER2BOW* mice were crossed with mice expressing *MMTV-Cre* recombinase(35) to generate *HER2BOW+/MMTV-Cre+* mice ($HER2BOW^{MMTV\text{-}Cre}$). $HER2BOW^{MMTV\text{-}Cre}$ mice were subsequently used to induce *HER2* Crainbow lineage tracing in the mammary gland and model the intratumor and intertumor heterogeneity of HER2 isoforms.

The tumorigenic process was visualized by confocal imaging of FUnGI clarified(25) *HER2BOW*$^{MMTV\text{-}Cre}$ mammary glands (Fig. 1c). FUnGI clarified mammary glands from four-week old *HER2BOW*$^{MMTV\text{-}Cre}$ mice were imaged by confocal microscopy, which is approximately one week after expression of Cre recombination(35). Robust Crainbow lineage tracing was observed throughout the mammary gland (Fig. 1d). Terminal end bud (TEBs) of *HER2BOW*$^{MMTV\text{-}Cre}$ mice were highly recombined and were predominantly labelled for ^{WT}HER2:mTFP1 or ^{d16}HER2:EYFP lineages (Fig. 1e). Mixed-lineage cells of the luminal epithelium express the basal cell marker, Krt14, during mammary gland development and prior to their commitment to mature luminal cell types(36). Mixed-lineage cells were prevalent in the TEB and were also Crainbow lineage positive (Fig. 1e, arrows). Low magnification imaging of HER2BOW$^{MMTV\text{-}Cre}$ mammary ducts revealed normal appearing ^{WT}HER2:mTFP1 labelled ducts, excessive ^{d16}HER2:EYFP labeled side-buds, and a profound scarcity of ^{p95}HER2:mKO ducts (Fig. 1d). Modest positional effects in Crainbow transgenes result in non-stochastic bias toward the mTFP1 lineage (24). Thus, the profound increase in EYFP-traced side-buds demonstrates the increased fitness of the ^{d16}HER2:EYFP lineage. At higher magnification, ductal ^{d16}HER2:EYFP side-buds were circumscribed by Krt14 positive (Krt14^{pos}) basal cells that were HER2BOW lineage negative. Unlike immature TEBs, luminal cells in the more mature ducts were HER2BOW positive and Krt14 negative (Krt14^{neg}). (Fig. 1f, inset arrowheads).

Extensive ductal side-budding was even more pronounced in the ^{d16}HER2:EYFP labelled eight-week-old mammary glands (i.e., approximately 5 weeks post *Cre* expression). ^{WT}HER2:mTFP1 labelled ducts were smooth in appearance with very few side-buds. Infrequent ^{p95}HER2:mKO side buds were observed (Fig. 1g). Higher magnification imaging revealed significant differences in the luminal and basal compartments. Similar to 4 weeks, ^{d16}HER2:EYFP labelled side buds were circumscribed by HER2BOW negative, Krt14^{pos} basal cells (Fig. 1h, arrowheads, Movie 1). In contrast, the few ^{p95}HER2:mKO labelled side buds often contained ^{p95}HER2:mKO/Krt14 double positive cells located on the bud periphery (Fig. 1i, solid arrows, Movie 2).

At sixteen weeks of age, *HER2BOW*$^{MMTV\text{-}Cre}$ mice developed palpable tumors. Mammary glands were also FUnGI clarified and imaged to acquire the fluorescent barcode of each tumor. In addition to the Crainbow lineage tag, tumors were also categorized on the basis of size and morphology as either a carcinoma (CA, >1.5mm and invasive margins), microinvasive carcinoma (miCA, <1.5mm and invasive margins), or mammary intraepithelial neoplasm (MIN, <1.5mm and non-invasive margins). MINs are small, non-palpable early lesions in genetically engineered mice that are analogous to human hyperplastic *in situ* carcinomas and can be precursors of invasive carcinoma (27,37–39). ^{p95}HER2:mKO expressing carcinomas (CA^{p95}) and microinvasive carcinoma (miCA^{p95}) were readily observable (Fig. 2a). In contrast, several instances of ^{d16}HER2:EYFP expressing MIN (MIN^{d16}) were found, none of which appeared invasive. At higher magnification, Ca^{p95} contained multiple clusters of Krt14/^{p95}HER2:mKO double positive cells (Fig. 2b) as well as spindle-shaped Krt14/ ^{p95}HER2:mKO double positive cells invading the peripheral stroma (Movie 3). Despite a much smaller size, the miCA^{p95} lacked distinct borders, and also contained single Krt14/ ^{p95}HER2:mKO double positive cells invading the mammary fat pad – sometimes hundreds of microns from the tumor

periphery (Fig. 2c, Movie 4). In contrast, MIN^{d16} resembled bouquets of side-buds in an organized appearance without evidence of an invasive border. Compared to Ca^{p95}/miCa^{p95}, the ^{d16}HER2:EYFP/Krt14 double positive cells were infrequent in MIN^{d16} lesions (Fig. 2d, Movie 5). Together, these data suggest that oncogenic isoforms of HER2 program unique tumor phenotypes.

Cancer phenotypes are predicted by HER2 isoforms.

Histological analysis of *HER2BOW* precancers and invasive tumors was performed to determine if HER2 isoforms program distinct tumor phenotypes. *HER2BOW* tumorigenic potential was first determined for each *HER2* isoform. Because pregnancy accelerates tumor formation in mice(40), multiparous *HER2BOW*$^{MMTV\text{-}Cre}$ mice were used to screen and quantify the overall tumorigenic potential of each *HER2* isoform. Palpable tumors were isolated from *HER2BOW*$^{MMTV\text{-}Cre}$ mice between 17 and 21 weeks of age. Tumors were vibratome sectioned and imaged by confocal microscopy to retrieve the fluorescent protein barcode (XFP). One ^{WT}HER2:mTFP1, six ^{d16}HER2:EYFP, and seven ^{p95}HER2:mKO tumors were observed. In addition, eleven multiclonal tumors were imaged and were primarily ^{d16}HER2 and ^{p95}HER2 positive (Supplementary Figure 2a). Together, this demonstrates the differences in tumorigenic potential of ^{WT}HER2 and its isoforms. Differences in tumor morphology were also apparent by confocal imaging. As such, a strategy for co-registering pathology with fluorescent lineage tracing was developed. This strategy utilizes serial adjacent tissue sections for histopathology and the analysis of XFP lineage reporters through **M**ultiplex **a**nalysis of **X**FP mRNAs by **f**luorescence **i**n **s**itu **h**ybridization (MaXFISH) (Supplementary Figure 2b–d). Histopathology and co-registry with MaXFISH imaging illustrates that oncogenic variants of *HER2* elicit heterogeneous tumors that correlate with tumor genotype.

Nulliparous *HER2BOW*$^{MMTV\text{-}Cre}$ mice were used to analyze the progression of early lesions to invasive carcinomas. Genotype:phenotype analysis was performed on a total of 52 neoplasms from nulliparous *HER2BOW*$^{MMTV\text{-}}$Cre mammary glands (N = 10 mice). This included 24 MIN lesions (^{WT}HER2:mTFP1 = 1, ^{d16}HER2:EYFP = 11, and ^{p95}HER2:mKO = 12) and 28 tumors (^{WT}HER2:mTFP1 = 1, ^{d16}HER2:EYFP = 13, and ^{p95}HER2:mKO = 14).

MaXFISH staining and imaging was used to quantify lineage expression. Inflammation and fibrosis were scored on a sliding scale and margins defined as described in “Methods”. Genotype:phenotype clusters were constructed for *HER2BOW* neoplasms through unsupervised hierarchical clustering of quantified MaXFISH and H&E histopathology (inflammation, fibrosis, area). Six clusters were identified (Fig. 3a, and Supplementary Figure 2e). Two ^{WT}HER2:mTFP1 lineage labelled lesions clustered together. ^{WT}HER2:mTFP1 tumors were rare (N =2), despite the abundant mTFP1 labelled cells throughout the mammary ducts. Nonetheless, the protracted rates of ^{WT}HER2 tumors and the pathological assessment of the two ^{WT}HER2 tumors confirmed the previously documented tumor phenotypes in *MMTV-cNeu/HER2* mice(10,39) (Supplementary Figure 2e). Humane endpoints were reached in the ^{d16}HER2:EYFP/ ^{p95}HER2:mKO tumors before

adequate numbers of ^{WT}HER2:mTFP1 tumors could be observed, thereby preventing further analysis of ^{WT}HER2:mTFP1 tumors.

^{d16}HER2:EYFP labelled lesions were phenotypically similar to previously described ^{WT}HER2 transgenic mice and ^{d16}HER2 mice(10,39,41). This included expansile margins with little fibrosis and inflammation. Several lesions were beginning to show signs of progression, including inflammation and fibrosis. Accordingly, this cluster was named “d16-expansile progressors” (Fig. 3a, “d16 expansile progressors”). A few ^{d16}HER2:EYFP lesions fully progressed to an invasive, inflamed, and fibrotic phenotype and were named “d16-invasive” (Fig. 3a, “d16 invasive”). ^{p95}HER2:mKO labelled lesions were composed of larger and more dysplastic cells frequently forming gland-like spaces. Their margins were irregular, extending into the fat-pad and other connective tissue. ^{p95}HER2:mKO margins were also clearly invasive, fibrotic, and had significant inflammation. These lesions were generally indistinguishable except by area and together form the “p95-invasive” cluster (Fig. 3a, “p95-invasive”). A smaller cluster for the largest ^{p95}HER2:mKO labelled tumors, “p95:invasive-Large”, was also identified (Fig. 3a, “p95:invasive-Large”). Only a few lesions were present in the “p95 early-invasive” cluster (Fig. 3a, “p95 early-invasive”).

Representative images of MaXFISH and H&E co-registry further demonstrate the remarkable differences between HER2+ lesions. The majority of ^{d16}HER2:EYFP *in situ* lesions closely resemble MIN (Fig. 3b–c, “b1” and “c1”). However, four of the ^{d16}HER2:EYFP had become invasive with modest immune infiltration (Fig. 3b–c, “b2” and “c2”). These invasive ^{d16}HER2:EYFP lesions started to resemble the invasive lesions induced by ^{p95}HER2:mKO. The majority of the ^{p95}HER2:mKO lesions were inflamed, invasive, and had become fibrotic. Small (i.e., <1.5mm^2) ^{p95}HER2:mKO closely resemble microinvasive carcinoma (Fig. 3b, “b3”) whereas larger ^{p95}HER2:mKO lesions (>1.5mm^2) were consistent with carcinoma (Fig. 3c, “c3”). Together, the genotype:phenotype clusters and FUnGI whole-gland imaging are evidence for at least three *HER2*-dependent trajectories. ^{WT}HER2:mTFP1 expression incites indolent epithelial precursors. ^{d16}HER2:EYFP expression incites proliferative *in situ* lesions, often reflecting a homogeneous luminal-like phenotype and less immune and stromal composition. Some ^{d16}HER2:EYFP appear to eventually progresses to invasion. Strikingly, ^{p95}HER2:mKO expression incites carcinomas marked by the early appearance of epithelial, immune, and stromal heterogeneity.

Single-cell molecular pathology of HER2BOW tumors.

Single-cell RNA sequencing (scRNAseq) was used to objectively map the heterogeneity of ^{d16}HER2:EYFP and ^{p95}HER2:mKO tumors. HER2BOW tumors were dissected from the mammary gland. Total cells from twelve tumors were digested and used for Drop-Seq capture (Macosko et al., 2015). Sixteen cell types were identified and grouped as either epithelial, immune, endothelial, or stromal/fibroblast (Fig. 4a, Supplementary Figure 3a–b, Supplementary Table 1). Biopsies from the sequenced tumors were used to co-register scRNAseq results with MaXFISH/H&E imaging (Supplementary Figure 4). One sequenced ^{d16}HER2:EYFP tumor was of the expansile-progressor phenotype (d16-ep). This included the more homogeneous luminal-like appearance with an abundant epithelial compartment

but little immune and stromal infiltration (Fig. 4b, “b1” and Supplementary Figure 4). Two ^{d16}HER2:EYFP tumors were sequenced that had progressed to an invasive phenotype (d16-inv) (Fig. 4b, “b2” and Supplementary Figure 4). These tumors were more heterogenous in appearance and varied in their epithelial, immune, and stromal composition. Five ^{p95}HER2:mKO tumors were sequenced, all of which were invasive and characterized by the heterogeneous presence of larger, dysplastic cells and abundant inflammation (p95-inv.) (Fig. 4b, “b3” and Supplementary Figure 4). These tumors tended to have fewer epithelial cells with a corresponding increase in immune and fibroblast cells. Four tumors were unable to be MaXFISH coregistered, but were still included in the study as histopathological reference controls for expansile progressors (X-ep.) or phenotypically mixed tumors (X-Mix, expansile progressors and invasive phenotypes in the same biopsy) (Supplementary Figure 4). In general, scRNAseq/HER2BOW co-registry demonstrated that ^{p95}HER2:mKO tumors were characterized by more abundant epithelial, immune, and stromal heterogeneity (Fig. 4b–c). While the stromal compartment appeared to be increased in ^{p95}HER2:mKO tumors, significant variation was found in the ^{d16}HER2:EYFP samples due to the progressive invasive phenotype observed in one of the invasive ^{d16}HER2:EYFP tumors.

Next, the epithelial (Fig. 4d–f), immune (Fig. 4g–i), and fibroblast (Fig. 4j–l) compartments were systematically characterized. *Epcam+* epithelial cells (4,666 cells) were re-clustered into nine distinct clusters with known marker genes of the basal and luminal epithelial compartments (Fig. 4d, Supplementary Table 2). Cluster 2 was excluded from further studies due to the lack of statistically significant upregulated marker genes. Cluster 4 contained an epithelial cell expressing both luminal (*Krt18*) and basal (*Krt14*) keratins (Fig. 4e). Approximately 24% of ^{p95}HER2:mKO and 6% of ^{d16}HER2:EYFP cells were *Kr18/Krt14* double-positive (Figure 4e). Immunohistochemistry (IHC) was performed to confirm these findings (Fig. 4f). Krt18 immunoreactivity was similar between MaXFISH coregistered ^{d16}HER2:EYFP tumors and ^{p95}HER2:mKO tumors. Infrequent Krt14 immunoreactivity was found in ^{d16}HER2:EYFP tumors. In contrast, and consistent with the scRNAseq analysis, Krt14 immunoreactivity was prominent in ^{p95}HER2:mKO tumors.

CD45+ immune cells (1,803 cells) were re-clustered into seven clusters. An analysis of marker genes enabled cluster identification (Fig. 4g, Supplementary Figure 3c, Supplementary Table 3). M1 and M2 Macrophages, Neutrophils, T-regulatory lymphocytes (Tregs), Cd8 T lymphocytes, Monocytes, and plasmacytoid dendritic cells (pDC) were found in HER2BOW tumors (Fig. 4g,h). Confirming our pathological scoring (Fig. 3), the fractional contribution of each immune cell was increased in ^{p95}HER2:mKO tumors. IHC for CD163-macrophages and CD3 lymphocytes further validated these findings in MaXFISH coregistered ^{d16}HER2:EYFP and ^{p95}HER2:mKO tumors (Fig. 4i).

Fibroblasts (922 cells) were reclustered into 5 clusters of cancer associated fibroblasts (CAF). Gene signature analysis revealed CAFs included myofibroblast-like, inflammatory, antigen-presenting, matrix, and vascular sub-types (Fig. 4j, Supplementary Figure 3d, Supplementary Table 4). A trend toward an increase in CAFs in ^{p95}HER2:mKO tumors was observed, and was most evident in Myo-like CAFs (Fig. 4k). IHC for Vimentin (VIM) and Smooth-muscle actin (SMA) qualitatively confirmed these observations in MaXFISH

co-registered ^{d16}HER2:EYFP and ^{p95}HER2:mKO tumors (Fig. 4l). Together these data demonstrate that HER2 isoforms incite unique tumorigenesis programs each characterized by differences in heterogeneity in the epithelial compartment and adaptive changes within the associated tumor microenvironment.

Oncogenic HER2 isoforms establish unique epithelial lineage states within tumors

We were intrigued by the cluster of *Krt18/Krt14* double-positive epithelial cells. These cells resemble the Krt14+ cells found in the luminal compartment of side-buds at 4-weeks of age (Fig. 1). Single-plex IHC and serial staining (Fig. 4f) confirmed the presence of Krt14+ and Krt18+ cells in ^{p95}HER2:mKO tumors but was inadequate for quantifying co-expression of multiple lineage markers in the same cell. Therefore, luminal cells (Krt8/18^{pos}&Krt14^{neg}), basal cells (Krt8/18^{neg}&Krt14^{pos}), and double-positive cells (Krt8/18^{pos}&Krt14^{pos}) were quantified in *HER2BOW* tumors using multiplex IHC. MaXFISH coregistry was used to genotype each tumor and validate the expected ^{d16}HER2:EYFP expansile-progressor and ^{p95}HER2:mKO invasive phenotypes (Fig. 5a–b). Next, a multiplex IHC method was utilized for quantifying extracellular domain intact HER2 (N-terminal Antibody), total HER2 (C-terminal Antibody), basal lineages (Krt14 Antibody), and luminal lineages (Krt8/18 Antibody). ^{d16}HER2:EYFP tumors stained strongly for N-terminal HER2 whereas ^{p95}HER2:mKO tumors were expectedly N-terminal negative (Fig. 5c–d). Krt14 and Krt8/18 IHC demonstrated that ^{d16}HER2:EYFP lesions primarily contained luminal lineage positive cells (Krt8/18^{pos}&Krt14^{neg}) (Fig. 5e) with infrequent basal cells (Krt8/18^{neg}&Krt14^{pos}) (Fig. 5e, inset). Double-positive (Krt8/18^{pos}&Krt14^{pos}) cells were rarely observed. In contrast, despite a clear luminal compartment, abundant basal cells and double-positive cells were observed in ^{p95}HER2:mKO tumors (Fig. 5f, insets 1 and 2). The epithelial lineages in each tumor were then quantified (N = 12 ^{d16}HER2:EYFP and N = 9 ^{p95}HER2:mKO). ^{d16}HER2:EYFP lesions were 95% luminal, <1% basal, and approximately 1.5% double-positive. In contrast, ^{p95}HER2:mKO lesions were 71% luminal, 13% basal, and 14% double-positive (Fig. 5g).

A ‘TruRegistry’ method was also developed to precisely co-register HER2BOW lineage tracing and epithelial specification. Native fluorescence of mTFP1, EYFP, and mKO was preserved on FFPE slides (Supplementary Figure 5a). TruRegistry demonstrated exclusive transgene expression in the epithelial compartment but not the stromal or immune microenvironment in HER2BOW^{MMTV} mice (Supplementary Figure 5b–c). In addition, IHC using a c-terminal specific HER2 antibody also demonstrated similar HER2 protein levels in ^{d16}HER2 and ^{p95}HER tumors (Supplementary Figure 5d).

We returned to our scRNAseq analysis to further characterize differences in the epithelial compartment of HER2BOW tumors. Monocle trajectory analysis(33) of Epcam+ ^{d16}HER2:EYFP tumor cells revealed 5 terminal branches, in addition to eight transitional states that were consolidated into a single state (Fig. 6a, Supplementary Table 5a). Cellular identities were inferred in the terminal branches using gene set enrichment analysis of 30 cell type specific signatures from human and mouse sequencing studies of the mammary gland (Supplementary Figure 6). The largest proportion of cells were characteristic of luminal-like hormone receptor negative cells (HR-) which comprised

62% of the ^{d16}HER2:EYFP epithelium **(HR-,** Fig. 6a, **red highlight)**. Two differentiated luminal-like cell states **(L1.1 and L1.2,** Fig. 6a) and one alveolar-like state **(Av,** Fig. 6a) were apparent within the HR- population. In addition, two rare cell states were observed. One closely resembled tumor epithelial cells undergoing an epithelial to mesenchymal-like transition **(EMT,** Fig. 6a, **yellow highlight)** and the second had characteristics of a hormone-sensitive cell **(HS,** Fig. 6a, **green highlight)**. Each accounted for 4% of the ^{d16}HER2:EYFP epithelial cells and both were double positive for the expression of *Krt18/Krt14* gene expression. Four genes (*Il33, Aldoc, Krt18,* and *Krt14*) were found that could discriminate the HS, HR-, and EMT-like cells (Fig. 6b). Crainbow TruRegistry and mRNA FISH of ^{d16}HER2:EYFP tumor sections were used to evaluate each cell state *in situ* (Fig. 6c). HS cells (*Il33*$^{+}$*,Krt18*$^{+}$*, Krt14*$^{+}$) were infrequent and unexpectedly did not express Crainbow lineage tags. Abundant HR- cells (*Aldoc*$^{+}$*/Krt18*$^{+}$) were present and as expected were positive for the ^{d16}*HER2*:EYFP lineage tag. Likewise, the rare EMT-like cells (*Krt14*$^{+}$*/Krt18*$^{+}$) were ^{d16}*HER2*:EYFP positive (Fig. 6d). Low magnification images confirm the scarcity of *Il33*-expressing HS cells and *Krt14*-expressing EMT-like cells. ^{d16}HER2:EYFP tumors instead appear to be more homogenously comprised of *Aldoc*-expressing HR- epithelial cells (Supplementary Figure 7a).

A similar analysis was performed for ^{p95}HER2:mKO tumors. Gene set enrichment analysis identified three broad cell states within the tumor epithelium (Fig. 6e, Supplementary Table 5b). Similar to ^{d16}HER2:EYFP, this included HS-like cells (green highlight), EMT-like (yellow highlight), and HR- like cells (red highlight) (Fig. 6e). The HR- could be further classified into alveolar-like and a single differentiated luminal-like cell type. Overall, ^{p95}HER2:mKO were more heterogenous than ^{d16}HER2:EYFP due to the abundance of HS-like cells (21%) and EMT like cells (18%) in addition to a significant HR- population (43%). The same set of genes used for ^{d16}HER2:EYFP (*Il33, Aldoc, Krt18, and Krt14*) was used to discriminate HS, HR- and EMT-like cell states on ^{p95}HER2:mKO tumor sections (Fig. 6f–h). HS cells (*Il33*$^{+}$*,Krt18*$^{+}$*, Krt14*$^{+}$) were observed and, similar to ^{d16}HER2:EYFP HS cells, did not express Crainbow lineage tags. HR- cells (*Aldoc*$^{+}$*/Krt18*$^{+}$) and EMT-like cells (*Krt14*$^{+}$*/Krt18*$^{+}$) were present and as expected, were positive for the ^{p95}*HER2*:mKO lineage (Fig. 6h). In contrast to ^{d16}HER2:EYFP tumors, low magnification images illustrate the heterogeneity of *Il33*-expressing HS cells, *Krt14*-expressing EMT-like cells, and *Aldoc*-expressing HR- cells within ^{p95}HER2:mKO tumors (Supplementary Figure 7b). Together, these data demonstrate that expression of ^{d16}HER2 and ^{p95}HER2 can account for intrinsic differences in epithelial heterogeneity in HER2+ mammary cancers.

Discussion

Our results demonstrate that oncogenic HER2 isoforms can encode the intratumor heterogeneity observed in HER2+ breast cancers. The expression of ^{WT}*HER2* induces rare, slowly growing indolent tumors. In contrast, ^{d16}*HER2* expression induces proliferative *in situ* lesions and luminal-like tumors, that over time progress to invasive cancers. The expression of ^{p95}*HER2* initiates invasive cancers characterized by significant epithelial, immune, and stromal heterogeneity without a protracted non-invasive stage. In this manner, associated epithelial lineage specification programs and adaptations within the tumor

microenvironment are uniquely programmed by the oncogene and can also occur very early in the tumorigenic process.

The Crainbow mouse modeling platform(24) was instrumental in visualizing the earliest events underlying tumor heterogeneity. Common problems such as transgene variegation, differences in recombination specificity, and strain-dependent immune effects are avoided in Crainbow systems. This eliminates obvious concerns regarding the comparison of multiple lines of transgenic mice. Crainbow is also an autochthonous model and therefore able to overcome limitations associated with xenograft transplantations. By activating each HER2 isoform in the same mouse we are able to compete each isoform in a tractable and controlled manner. FUnGI whole-gland(25) fluorescent imaging together with classical molecular pathology demonstrated the power of state-of-the-art imaging modalities when paired with appropriate genetic tools. Using Crainbow, we were able to observe very early changes in epithelial heterogeneity. This included evidence of relatively normal ^{WT}HER2 ductal epithelial cells. In contrast, ^{d16}HER2 epithelium had features of early proliferative disease that included a remarkable ability to generate side-buds. Over time these appeared to grow into *in situ* disease with very little observable heterogeneity. Most remarkable was that even the smallest ^{p95}HER2 lesions were invasive and already heterogeneous. This shows how inciting oncogenes can program intratumor heterogeneity during the earliest and occult tumorigenic phase.

HER2BOW mice provided a reproducible tool for exploring intratumor heterogeneity. scRNAseq was useful for characterizing the epithelial, immune, and stromal compartments present within twelve tumors. In each case, H&E histopathology closely resembled the overall trend present in the scRNAseq data sets. That is, as tumors became more invasive, a decrease in total epithelial cells and a corresponding increase in immune and stromal infiltrate was found. While there were a fewer total number of epithelial cells in the invasive ^{p95}HER2 tumors compared to ^{d16}HER2 tumors, there was a surprising increase in epithelial cells co-expressing basal and luminal cytokeratin (i.e., *Krt14/Krt18*). Mutations downstream of *HER2*, like oncogenic *PI3KCA* mutations, reprogram luminal epithelial cells into tumor initiating multipotent progenitor cells that also co-express basal and luminal keratin (42,43). Again, Crainbow TruRegistry was invaluable because we were able to directly visualize these heterogeneous epithelial states *in situ* and in relation to Crainbow transformed cells. We found that two populations of mixed-lineage states exist. One is an EMT-like state. The second appears to be similar to a hormone sensitive cell that also expresses the multipotent progenitor marker gene *Il33*(42). Surprisingly, this cell was Crainbow lineage negative. Thus, the robust increase of HS-like cells in ^{p95}HER2 tumors suggests that non-transformed bystander cells may also contribute to intratumor heterogeneity.

The sporadic tumorigenic events observed in HER2BOW ducts suggest that not all luminal cells are capable of oncogenic reprogramming. One potential cell-of-origin is the rare adult mammary stem cell(44). Another possibility is that the Krt14/Krt18 double positive cells in HER2BOW tumors could originate from mixed-lineage developmental progenitors that fail to appropriately differentiate(36). *MMTV-Cre* dependent recombination of the *HER2BOW* transgene occurs approximately 3 weeks postnatally. Coincidentally, this occurs during a critical developmental milestone when mixed lineage cells are abundant. Thus

HER2-isoforms may differ in their intrinsic effects on lineage specification during postnatal mammary gland development.

For decades, oncologists have sought to explain how progressively earlier, and more aggressive surgical intervention has not continued to improve survival in breast cancer patients. Whether, or not, contemporary screening technologies can identify truly lethal cancers at a non-disseminated stage has led to significant debate(45). Some postulated invasive breast cancers were “biologically predetermined” to be lethal(46). Proponents of this theory argued that the formative early phase of tumor and microenvironment adaptation occurred decades before diagnosis and destined the overall trajectory of tumors to either lethal or indolent. While overly simplistic at the time, the basis for this theory, that cancer is determined by the balance struck between host and neoplasm, has become increasingly relevant in the wake of new molecular evidence(46). The model of adaptive oncogenesis by DeGregori (47,48) provides a similar explanation for the observed latency in tumor growth that occurs long after the initial genetic event. In this model tumor driver genes present years before diagnosis are held in check by a healthy epithelium. Later in life, when age or injury dependent declines in health are observed, previously quiescent genetic changes provide exceptional adaptive fitness for the rapid growth of tumorigenic cells. By extension, our work begins to show that isoforms of HER2 have their own intrinsic potential to differentially adapt, resulting in profound effects on the timescales and heterogeneous phenotypes of invasive breast cancer. More broadly, our work highlights the importance of deciphering the early genetic and cellular programs at work during the prolonged pre-diagnostic phase of breast cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by NCI K12 Training grant 512-CA100639-10, NCI 1K22CA212058, Sage Biosciences IACM Supplement 3U24CA209923-01S1, Duke Surgery Start-up funds to J.C.S. and the Duke Surgery Gardner Award to J.C.S. and H.K.L. This work was also supported by an NCI 1R01CA255372-01 to J.C.S., H.K.L, and Z.H.C. Z.H.C was supported by DOD W81XWH-20-1-0346 and NCI 5R01-CA238217. HKL is supported by Department of Defense Grant Number BC111085. J.D.G. was supported by an NICHD 5T32HD040372. C.A. was supported by T32-CA009111. L.S.B, H.K.L, and M.G.C were supported by NCI Grants R21CA173245 and 1R33CA191198. A.D.B and H.M were supported by a grant from NIH/NCI U01CA196406. H.K.L, Z.C.H, M.G.C, and J.C.S acknowledge membership in the Duke Cancer Institute. We thank the Duke University School of Medicine, and Duke Cancer Institute for the use of the Sequencing and Genomic Technologies Shared Resource, which provided next-generation sequencing services for scRNAseq studies. The authors also acknowledge the Duke Cancer Institute’s Transgenic and Knockout Mouse core, including Cheryl Bock, Gary Kucera, and Mei Lang Flowers, for the generation of HER2BOW mice.

References

1. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 1989;244(4905):707–12 doi 10.1126/science.2470152. [PubMed: 2470152]
2. Allred DC, Clark GM, Molina R, Tandon AK, Schnitt SJ, Gilchrist KW, et al. Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer. Hum Pathol 1992;23(9):974–9. [PubMed: 1355464]

3. Andrulis IL, Bull SB, Blackstein ME, Sutherland D, Mak C, Sidlofsky S, et al. neu/erbB-2 amplification identifies a poor-prognosis group of women with node-negative breast cancer. Toronto Breast Cancer Study Group. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 1998;16(4):1340–9 doi 10.1200/JCO.1998.16.4.1340. [PubMed: 9552035]
4. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature 2020;578(7793):122–8 doi 10.1038/s41586-019-1907-7. [PubMed: 32025013]
5. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, et al. Adjuvant trastuzumab in HER2-positive breast cancer. N Engl J Med 2011;365(14):1273–83 doi 10.1056/NEJMoa0910383. [PubMed: 21991949]
6. Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 2012;367(19):1783–91 doi 10.1056/NEJMoa1209124. [PubMed: 23020162]
7. Swain SM, Baselga J, Kim SB, Ro J, Semiglazov V, Campone M, et al. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. N Engl J Med 2015;372(8):724–34 doi 10.1056/NEJMoa1413513. [PubMed: 25693012]
8. Geyer CE, Forster J, Lindquist D, Chan S, Romieu CG, Pienkowski T, et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. N Engl J Med 2006;355(26):2733–43 doi 10.1056/NEJMoa064320. [PubMed: 17192538]
9. Prat A, Pascual T, Adamo B. Intrinsic molecular subtypes of HER2+ breast cancer. Oncotarget 2017;8(43):73362–3 doi 10.18632/oncotarget.20629. [PubMed: 29088709]
10. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. Proc Natl Acad Sci U S A 1992;89(22):10578–82. [PubMed: 1359541]
11. Finkle D, Quan ZR, Asghari V, Kloss J, Ghaboosi N, Mai E, et al. HER2-targeted therapy reduces incidence and progression of midlife mammary tumors in female murine mammary tumor virus huHER2-transgenic mice. Clin Cancer Res 2004;10(7):2499–511 doi 10.1158/1078-0432.ccr-03-0448. [PubMed: 15073130]
12. Turpin J, Ling C, Crosby EJ, Hartman ZC, Simond AM, Chodosh LA, et al. The ErbB2DeltaEx16 splice variant is a major oncogenic driver in breast cancer that promotes a pro-metastatic tumor microenvironment. Oncogene 2016;35(47):6053–64 doi 10.1038/onc.2016.129. [PubMed: 27157621]
13. Castagnoli L, Ghedini GC, Koschorke A, Triulzi T, Dugo M, Gasparini P, et al. Pathobiological implications of the d16HER2 splice variant for stemness and aggressiveness of HER2-positive breast cancer. Oncogene 2017;36(12):1721–32 doi 10.1038/onc.2016.338. [PubMed: 27641338]
14. Mitra D, Brumlik MJ, Okamgba SU, Zhu Y, Duplessis TT, Parvani JG, et al. An oncogenic isoform of HER2 associated with locally disseminated breast cancer and trastuzumab resistance. Mol Cancer Ther 2009;8(8):2152–62 doi 10.1158/1535-7163.MCT-09-0295. [PubMed: 19671734]
15. Castagnoli L, Iezzi M, Ghedini GC, Ciravolo V, Marzano G, Lamolinara A, et al. Activated d16HER2 homodimers and SRC kinase mediate optimal efficacy for trastuzumab. Cancer Res 2014;74(21):6248–59 doi 10.1158/0008-5472.CAN-14-0983. [PubMed: 25164009]
16. Marchini C, Gabrielli F, Iezzi M, Zenobi S, Montani M, Pietrella L, et al. The Human Splice Variant Δ16HER2 Induces Rapid Tumor Onset in a Reporter Transgenic Mouse. PLoS ONE 2011;6(4):e18727 doi 10.1371/journal.pone.0018727.
17. Molina MA, Saez R, Ramsey EE, Garcia-Barchino MJ, Rojo F, Evans AJ, et al. NH(2)-terminal truncated HER-2 protein but not full-length receptor is associated with nodal metastasis in human breast cancer. Clin Cancer Res 2002;8(2):347–53. [PubMed: 11839648]
18. Pedersen K, Angelini PD, Laos S, Bach-Faig A, Cunningham MP, Ferrer-Ramon C, et al. A naturally occurring HER2 carboxy-terminal fragment promotes mammary tumor growth and metastasis. Mol Cell Biol 2009;29(12):3319–31 doi 10.1128/MCB.01803-08. [PubMed: 19364815]
19. Xia W, Liu Z, Zong R, Liu L, Zhao S, Bacus SS, et al. Truncated ErbB2 expressed in tumor cell nuclei contributes to acquired therapeutic resistance to ErbB2 kinase inhibitors. Mol Cancer Ther 2011;10(8):1367–74 doi 10.1158/1535-7163.MCT-10-0991. [PubMed: 21673090]

20. Anido J, Scaltriti M, Bech Serra JJ, Santiago Josefat B, Todo FR, Baselga J, et al. Biosynthesis of tumorigenic HER2 C-terminal fragments by alternative initiation of translation. EMBO J 2006;25(13):3234–44 doi 10.1038/sj.emboj.7601191. [PubMed: 16794579]
21. Carvajal-Hausdorf DE, Patsenker J, Stanton KP, Villarroel-Espindola F, Esch A, Montgomery RR, et al. Multiplexed (18-Plex) Measurement of Signaling Targets and Cytotoxic T Cells in Trastuzumab-Treated Patients using Imaging Mass Cytometry. Clin Cancer Res 2019;25(10):3054–62 doi 10.1158/1078-0432.CCR-18-2599. [PubMed: 30796036]
22. Carvajal-Hausdorf DE, Schalper KA, Pusztai L, Psyrri A, Kalogeras KT, Kotoula V, et al. Measurement of Domain-Specific HER2 (ERBB2) Expression May Classify Benefit From Trastuzumab in Breast Cancer. J Natl Cancer Inst 2015;107(8) doi 10.1093/jnci/djv136.
23. Smith HW, Yang L, Ling C, Walsh A, Martinez VD, Boucher J, et al. An ErbB2 splice variant lacking exon 16 drives lung carcinoma. Proc Natl Acad Sci U S A 2020;117(33):20139–48 doi 10.1073/pnas.2007474117. [PubMed: 32727899]
24. Boone PG, Rochelle LK, Ginzel JD, Lubkov V, Roberts WL, Nicholls PJ, et al. A cancer rainbow mouse for visualizing the functional genomics of oncogenic clonal expansion. Nat Commun 2019;10(1):5490 doi 10.1038/s41467-019-13330-y. [PubMed: 31792216]
25. Rios AC, Capaldo BD, Vaillant F, Pal B, van Ineveld R, Dawson CA, et al. Intraclonal Plasticity in Mammary Tumors Revealed through Large-Scale Single-Cell Resolution 3D Imaging. Cancer cell 2019;35(4):618–32 e6 doi 10.1016/j.ccell.2019.02.010. [PubMed: 30930118]
26. Snyder JC, Rochelle LK, Ray C, Pack TF, Bock CB, Lubkov V, et al. Inhibiting clathrin-mediated endocytosis of the leucine-rich G protein-coupled Receptor-5 diminishes cell fitness. J Biol Chem 2017 doi 10.1074/jbc.M116.756635.
27. Mori H, Chen JQ, Cardiff RD, Penzvalto Z, Hubbard NE, Schuetter L, et al. Pathobiology of the 129:Stat1 (−/−) mouse model of human age-related ER-positive breast cancer with an immune infiltrate-excluded phenotype. Breast cancer research : BCR 2017;19(1):102 doi 10.1186/s13058-017-0892-8. [PubMed: 28865492]
28. McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: Next-generation image processing for biology. PLoS biology 2018;16(7):e2005970 doi 10.1371/journal.pbio.2005970.
29. Mori H, Cardiff RD. Methods of Immunohistochemistry and Immunofluorescence: Converting Invisible to Visible. Methods Mol Biol 2016;1458:1–12 doi 10.1007/978-1-4939-3801-8_1. [PubMed: 27581010]
30. Stack EC, Wang C, Roman KA, Hoyt CC. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. Methods 2014;70(1):46–58 doi 10.1016/j.ymeth.2014.08.016. [PubMed: 25242720]
31. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177(7):1888–902 e21 doi 10.1016/j.cell.2019.05.031. [PubMed: 31178118]
32. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 2018 doi 10.1038/nbt.4314.
33. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 2014;32(4):381–6 doi 10.1038/nbt.2859. [PubMed: 24658644]
34. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 2017;14(10):979–82 doi 10.1038/nmeth.4402. [PubMed: 28825705]
35. Wagner KU, McAllister K, Ward T, Davis B, Wiseman R, Hennighausen L. Spatial and temporal expression of the Cre gene under the control of the MMTV-LTR in different lines of transgenic mice. Transgenic Res 2001;10(6):545–53. [PubMed: 11817542]
36. Pal B, Chen Y, Vaillant F, Jamieson P, Gordon L, Rios AC, et al. Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nat Commun 2017;8(1):1627 doi 10.1038/s41467-017-01560-x. [PubMed: 29158510]

37. Cardiff RD, Moghanaki D, Jensen RA. Genetically engineered mouse models of mammary intraepithelial neoplasia. J Mammary Gland Biol Neoplasia 2000;5(4):421–37. [PubMed: 14973386]
38. Maglione JE, McGoldrick ET, Young LJ, Namba R, Gregg JP, Liu L, et al. Polyomavirus middle T-induced mammary intraepithelial neoplasia outgrowths: single origin, divergent evolution, and multiple outcomes. Mol Cancer Ther 2004;3(8):941–53. [PubMed: 15299077]
39. Cardiff RD, Anver MR, Gusterson BA, Hennighausen L, Jensen RA, Merino MJ, et al. The mammary pathology of genetically engineered mice: the consensus report and recommendations from the Annapolis meeting. Oncogene 2000;19(8):968–88. [PubMed: 10713680]
40. Landis MD, Seachrist DD, Abdul-Karim FW, Keri RA. Sustained trophism of the mammary gland is sufficient to accelerate and synchronize development of ErbB2/Neu-induced tumors. Oncogene 2006;25(23):3325–34 doi 10.1038/sj.onc.1209365. [PubMed: 16434967]
41. Castagnoli L, Iezzi M, Ghedini GC, Ciravolo V, Marzano G, Lamolinara A, et al. Activated d16HER2 Homodimers and SRC Kinase Mediate Optimal Efficacy for Trastuzumab. Cancer Research 2014;74(21):6248–59 doi 10.1158/0008-5472.can-14-0983. [PubMed: 25164009]
42. Van Keymeulen A, Lee MY, Ousset M, Brohee S, Rorive S, Giraddi RR, et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. Nature 2015;525(7567):119–23 doi 10.1038/nature14665. [PubMed: 26266985]
43. Koren S, Reavie L, Couto JP, De Silva D, Stadler MB, Roloff T, et al. PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. Nature 2015;525(7567):114–8 doi 10.1038/nature14669. [PubMed: 26266975]
44. Wang D, Cai C, Dong X, Yu QC, Zhang XO, Yang L, et al. Identification of multipotent mammary stem cells by protein C receptor expression. Nature 2015;517(7532):81–4 doi 10.1038/nature13851. [PubMed: 25327250]
45. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. N Engl J Med 2016;375(15):1438–47 doi 10.1056/NEJMoa1600249. [PubMed: 27732805]
46. Macdonald I. The natural history of mammary carcinoma. The American Journal of Surgery 1966;111(3):435–42 doi 10.1016/S0002-9610(66)80023-5. [PubMed: 5907009]
47. Laconi E, Marongiu F, DeGregori J. Cancer as a disease of old age: changing mutational and microenvironmental landscapes. British journal of cancer 2020;122(7):943–52 doi 10.1038/s41416-019-0721-1. [PubMed: 32042067]
48. Marusyk A, DeGregori J. Declining cellular fitness with age promotes cancer initiation by selecting for adaptive oncogenic mutations. Biochimica et biophysica acta 2008;1785(1):1–11 doi 10.1016/j.bbcan.2007.09.001. [PubMed: 17980163]

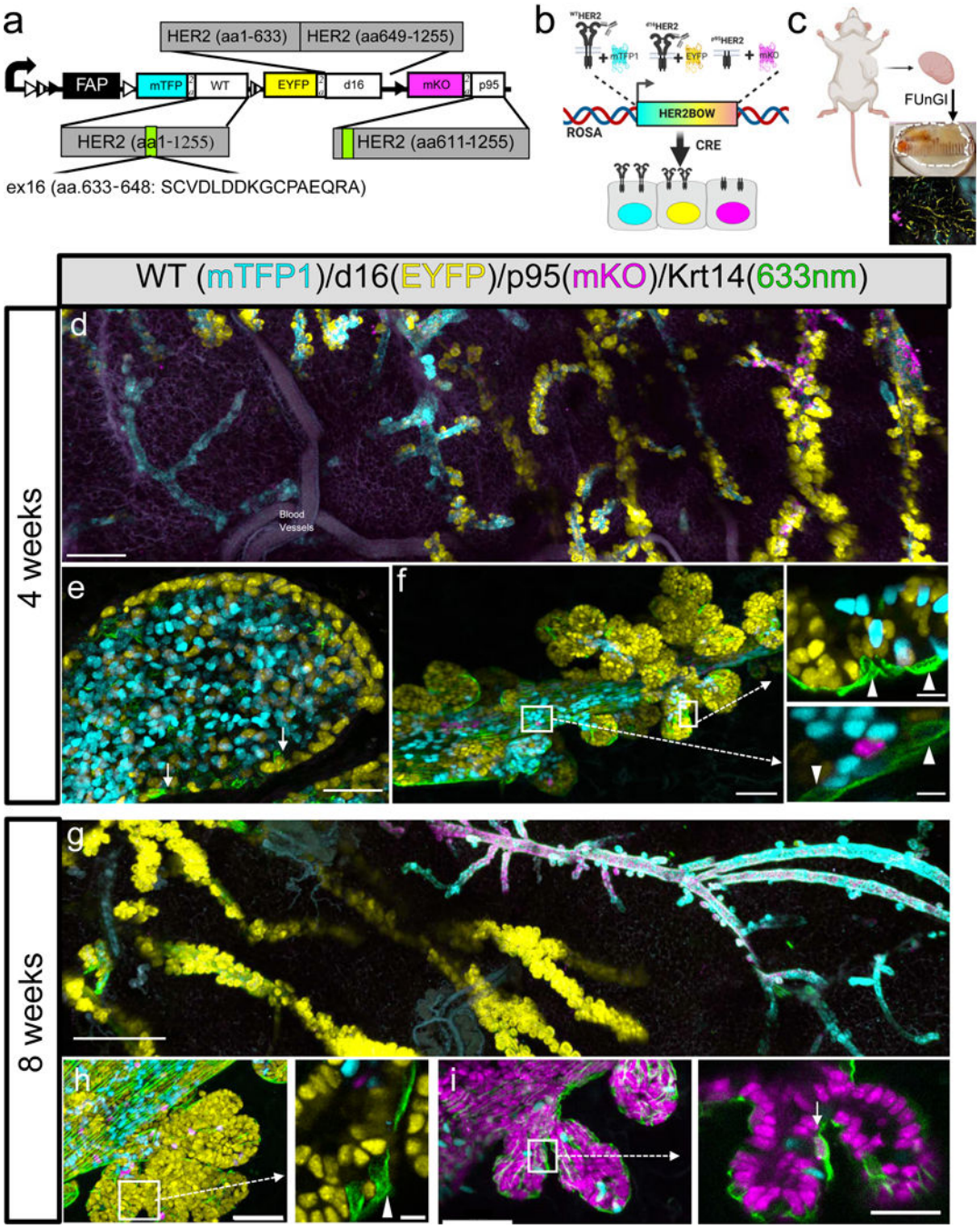


Figure 1. HER2 isoforms encode distinct developmental trajectories of cancer progression

(a) Fluorescent barcoding of human *HER2* oncogenic variants in a Crainbow mouse (*HER2BOW*). *mTFP1* (cyan, ex**λ**462*nm*/em**λ**492*nm*) is co-expressed with full length *HER2* (*^{WT}HER2:mTFP1*). *EYFP* (yellow, ex**λ**513*nm*/em**λ**527*nm*) is co-expressed with a *HER2* splice isoform lacking exon 16 (*^{d16}HER2:EYFP*). *mKO* (magenta, ex**λ**548*nm*/em**λ**559nm) is co-expressed with the N-terminally truncated fragment of HER2 (*^{p95}HER2:mKO, aa611–1255*). Green bar denotes exon 16 and orthogonal lox sites are denoted by triangles (LoxN: white triangle, Lox2272: hatched triangle, and LoxP: filled triangle). See plasmid

map in Supplementary Figure 1. **(b)** *Cre* activation of the *HER2BOW* construct in a cell results in expression of one of the three HER2 isoforms and nuclear fluorescent protein barcode. **(c)** Mammary glands from *HER2BOW+/MMTV-Cre+* mice were fixed, immunostained for Krt14, and FUnGI clarified. **(d)** Clarified mammary glands from four-week-old HER2BOW$^{MMTV\text{-}Cre}$ mice at 10x magnification. **(e-f)** 40X magnification confocal imaging and 3D reconstructions of **(e)** terminal end bud of a developing mammary gland with Krt14^{pos}/HER2BOW^{pos} cells (arrows) and **(f)** mammary duct with extensive ^{d16}HER2:EYFP side buds. Insets show Krt14^{pos}/HER2BOW^{neg} basal cells (arrow heads). **(g)** Clarified mammary glands from eight-week-old *HER2BOW*$^{MMTV\text{-}Cre}$ mice at 10x magnification. 40X magnification confocal imaging and 3D reconstructions of **(h)** ^{d16}HER2:EYFP and **(i)** ^{p95}HER2:mKO side buds. See also Movie 1 and Movie 2. Inset in **(h)** Krt14^{pos}/HER2BOW^{neg} basal cells (arrow heads). Inset in **(i)** Krt14/HER2BOW double positive cells (arrow). Scale bars: (d,g) 500μm, (e,f,h,i) 50μm, (e,f,h inset) 5μm, (i inset) 25μm

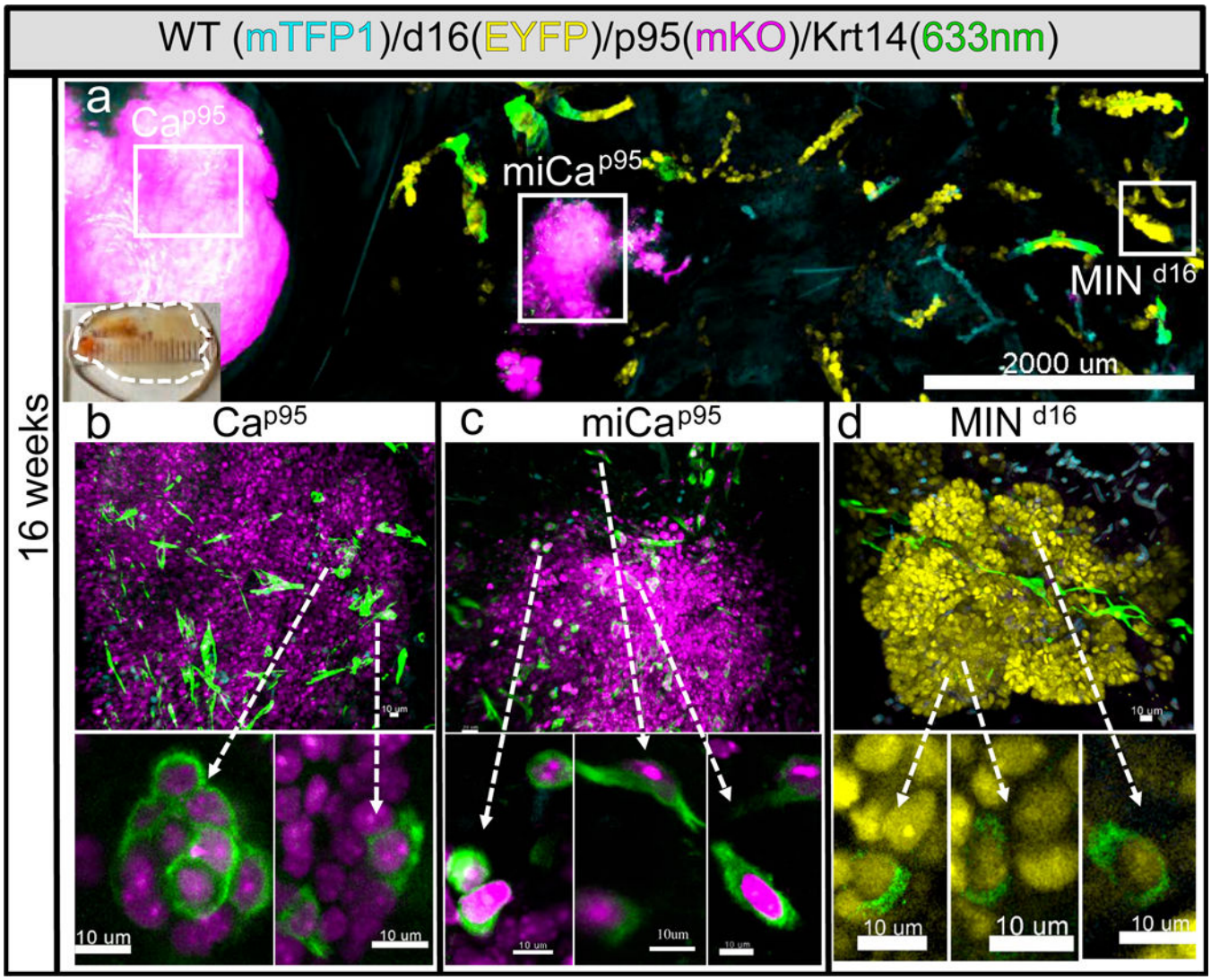


Figure 2. Three dimensionally reconstructing HER2+ tumors.

Mammary glands from 16-week-old HER2BOW mice were fixed, immunostained for Krt14 (green), and then clarified. **(a)** Whole mammary gland imaging (10x) reveals multiple advanced lesions expressing either ^{d16}HER2:EYFP or ^{p95}HER2:mKO. Inset shows clarified mammary gland (light micrograph). ROIs in "a" are shown in "b-d". High magnification (40X) confocal imaging and 3D reconstructions of **(b)** ^{p95}HER2:mKO invasive carcinoma (Ca^{p95}, See Movie 3), **(c)** ^{p95}HER2:mKO microinvasive carcinoma (miCa^{p95,} Movie 4), and **(d)** ^{d16}HER2:EYFP mammary intraepithelial neoplasm (MIN^{d16}, Movie 5). Insets depict optical slices of p95^{pos}/Krt14^{pos} cells in "b and c" and d16 ^{pos}/Krt14^{pos} cells in "d". Scale bars as indicated.

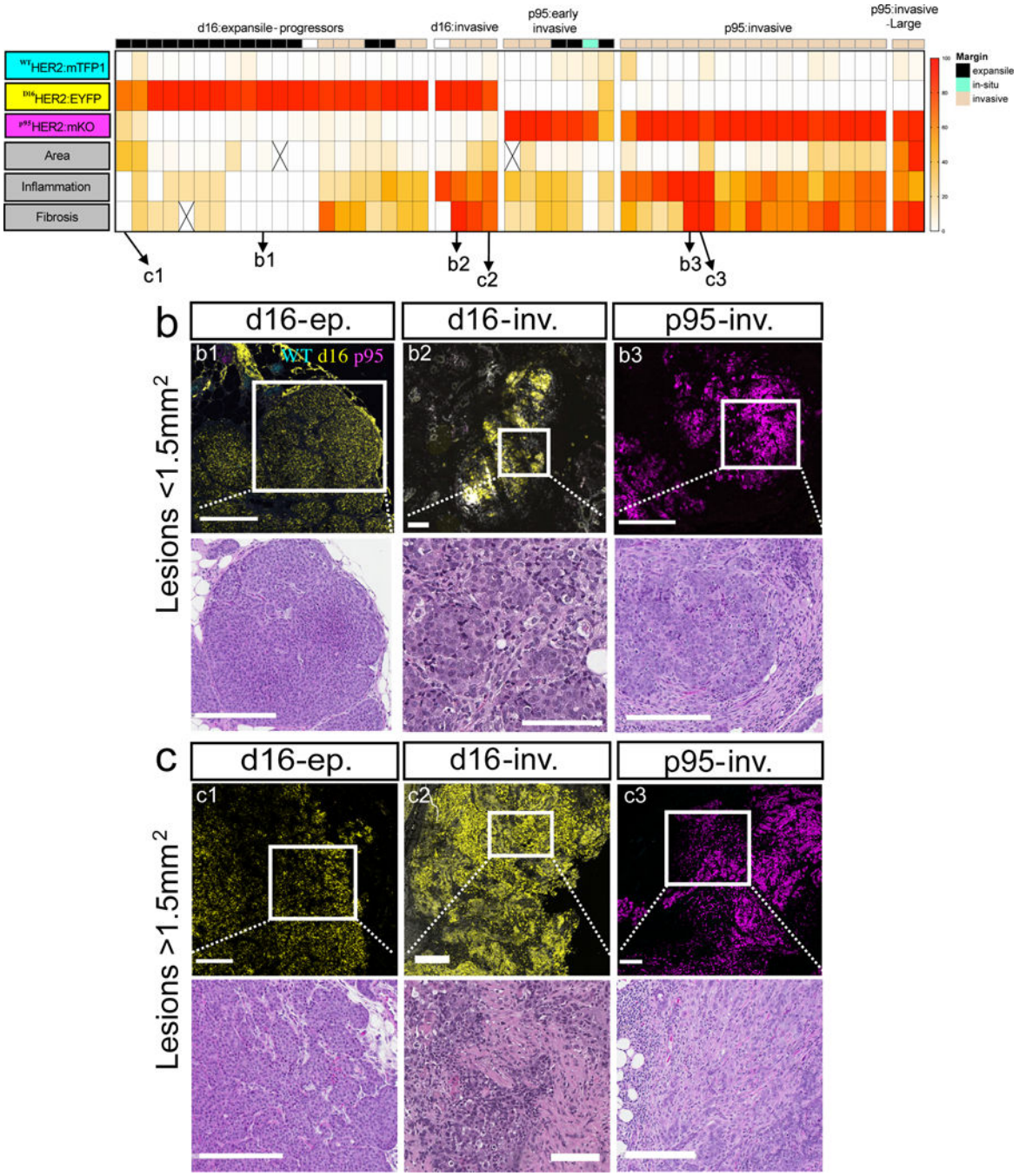


Figure 3. HER2 genotypes predict cancer phenotypes.

(a) Unsupervised hierarchical clustering of MaXFISH coregistry (mTFP1, EYFP, mKO), inflammation, fibrosis, and area of n=50 lesions. Clusters were ordered by genotype (left to right). Top row denotes expansile margin (black), in-situ margin (teal), invasive margin (tan), and indeterminant margins (white). Quantification of MaXFISH and H&E histopathology as described in text (X = undetermined). See also MaXFISH description in Supplementary Figure 2. Representative images of three genotype:phenotype clusters in **(b)** MaxFISH

coregistered lesions <1.5mm^2 (b1-b3) and in **(c)** MaxFISH coregistered lesions >1.5mm^2 (c1-c3). Scale bars: (b2, c2) 100μm, (b1, b3, c1, c3) 250μm.

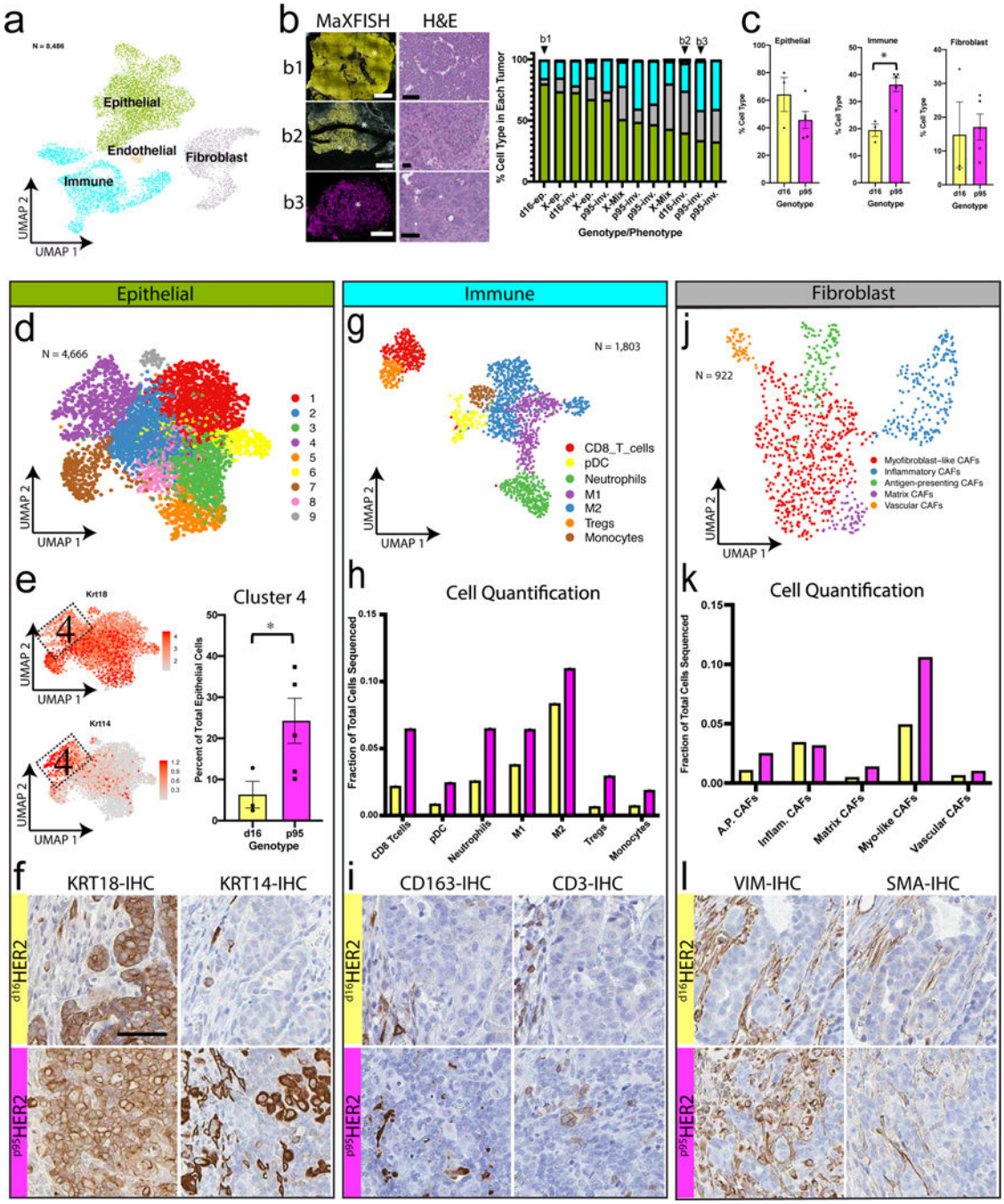


Figure 4. Single cell molecular pathology of HER2BOW tumors.

(a) Uniform manifold approximation and projection (UMAP) of 8,486 cells with clusters of inferred cell types from HER2BOW tumors (n=12 tumors). For cell type identities see also Supplementary Figure 3. **(b)** MaXFISH coregistry of a ^{d16}HER2:EYFP tumor with expansile progressor appearance (b1, d16-ep.), a ^{d16}HER2:EYFP tumor with invasive appearance (b2, d16-inv.), and an invasive ^{p95}HER2:mKO tumor (b3, p95-inv). For complete coregistry see Supplementary Figure 4. Quantification of the epithelial (green), fibroblast (gray), immune (cyan), and endothelial (orange) compartments in scRNAseq data

according to the genotype:phenotype of each tumor. Tumors where genotype coregistry was unavailable (X). **(c)** Comparison of the epithelial, immune, and fibroblast compartments in scRNAseq data between ^{d16}HER2:EYFP (n=3) and ^{p95}HER2:mKO (n=5). **(d)** UMAP of all 4,666 epithelial cells identified in "a" revealed 9 clusters. See also Supplementary Table 2. **(e)** Feature plots for *Krt18* (top) and *Krt14* (bottom) expression in epithelial clusters. Red indicates high expression. Quantification of the percent of total ^{p95}HER2:mKO or ^{d16}HER2:EYFP epithelial cells in cluster 4. **(f)** Immunohistochemistry for Krt18 and Krt14 in ^{d16}HER2:EYFP and ^{p95}HER2:mKO tumors. **(g)** UMAP of all 1803 immune cells identified in "**a**" revealed 7 clusters. See also Supplementary Figure 3 and Supplementary Table 3. **(h)** Quantification of the percent of total ^{p95}HER2:mKO or ^{d16}HER2:EYFP cells in each immune cluster. **(i)** Immunohistochemistry for CD163 and CD3 in ^{d16}HER2:EYFP and ^{p95}HER2:mKO tumors. **(j)** UMAP of all 902 cancer associated fibroblasts identified in "a" revealed 5 clusters. See also Supplementary Figure 3 and Supplementary Table 4. **(k)** Quantification of the percent of total ^{p95}HER2:mKO or ^{d16}HER2:EYFP cells in each fibroblast cluster. **(l)** Immunohistochemistry for VIM and SMA in ^{d16}HER2:EYFP and ^{p95}HER2:mKO tumors. (c,e) * = $p<0.05$. Data are represented as mean ± SEM.

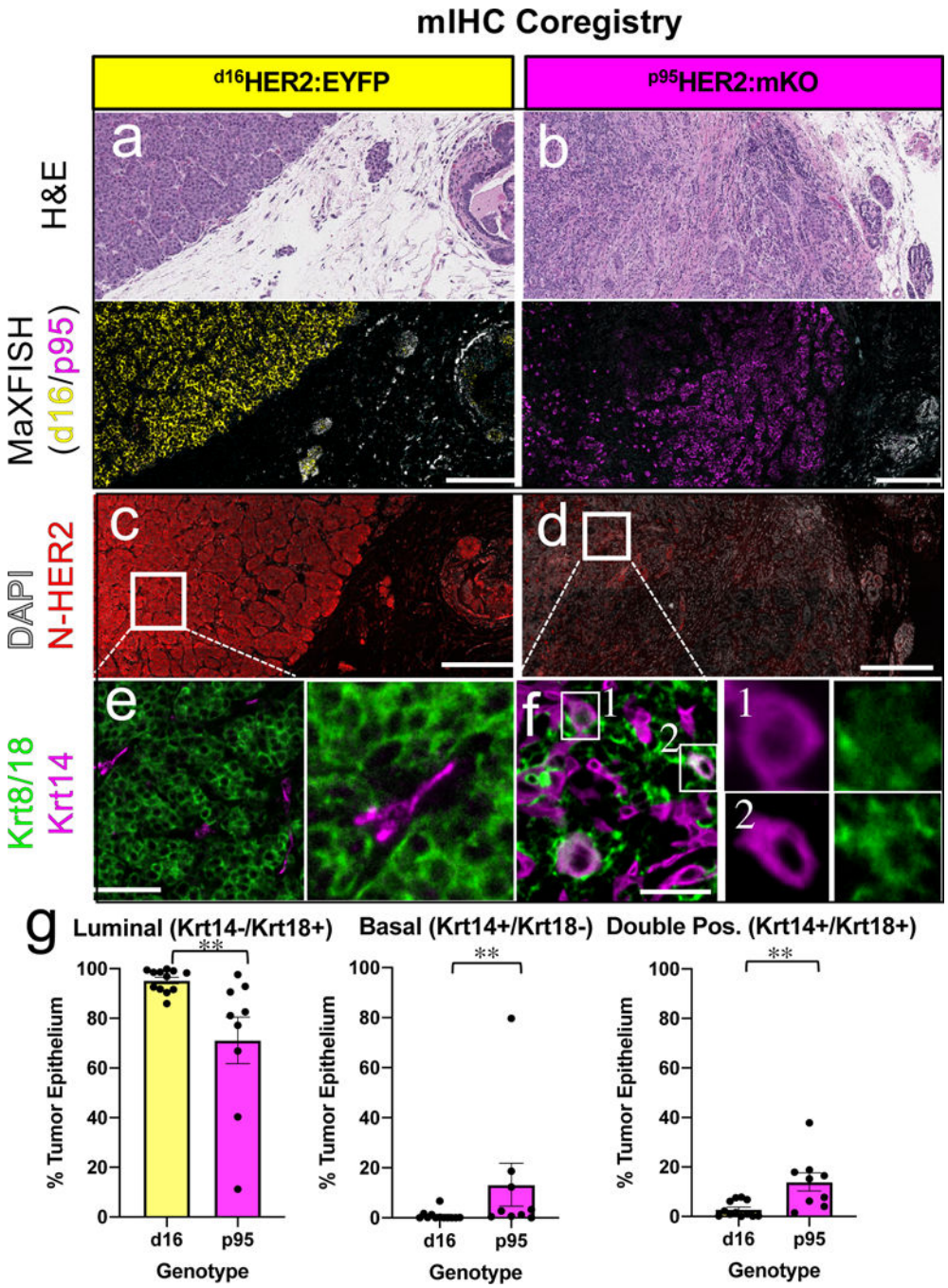


Figure 5. Oncogenic HER2 isoforms establish unique epithelial lineage states within tumors.

MaxFISH coregistry of **(a)** ^{d16}HER2:EYFP and **(b)** ^{p95}HER2:mKO tumors. Multiplex IHC (mIHC) for N-terminal HER2, C-terminal HER2, Krt14, and Krt8/18 was performed on a serial adjacent slide. Representative images of HER2 N-terminal domain intact (red) on **(c)** ^{d16}HER2:EYFP tumor, serial adjacent to "a" and **(d)** ^{p95}HER2:mKO tumor, serial adjacent to "b". Representative images of Krt8/18 (green) and Krt14 (magenta) in **(e)** ^{d16}HER2:EYFP tumor. Inset depicts luminal lineage cells (Krt8/18,green) and basal lineage cells (Krt14,magenta) **(f)** ^{p95}HER2:mKO tumor with mixed lineage Krt8/18pos/ Krt14pos

cells. Insets 1 and 2 show single color images of mixed lineage cells. **(g)** Quantification of each epithelial lineage as a percentage of total tumor epithelium. ^{d16}HER2:EYFP (n=12) and ^{p95}HER2:mKO (n=9). Mann-Whitney test; ** = $p < 0.005$. Scale bars: (b) 250μm, (c) 50μm

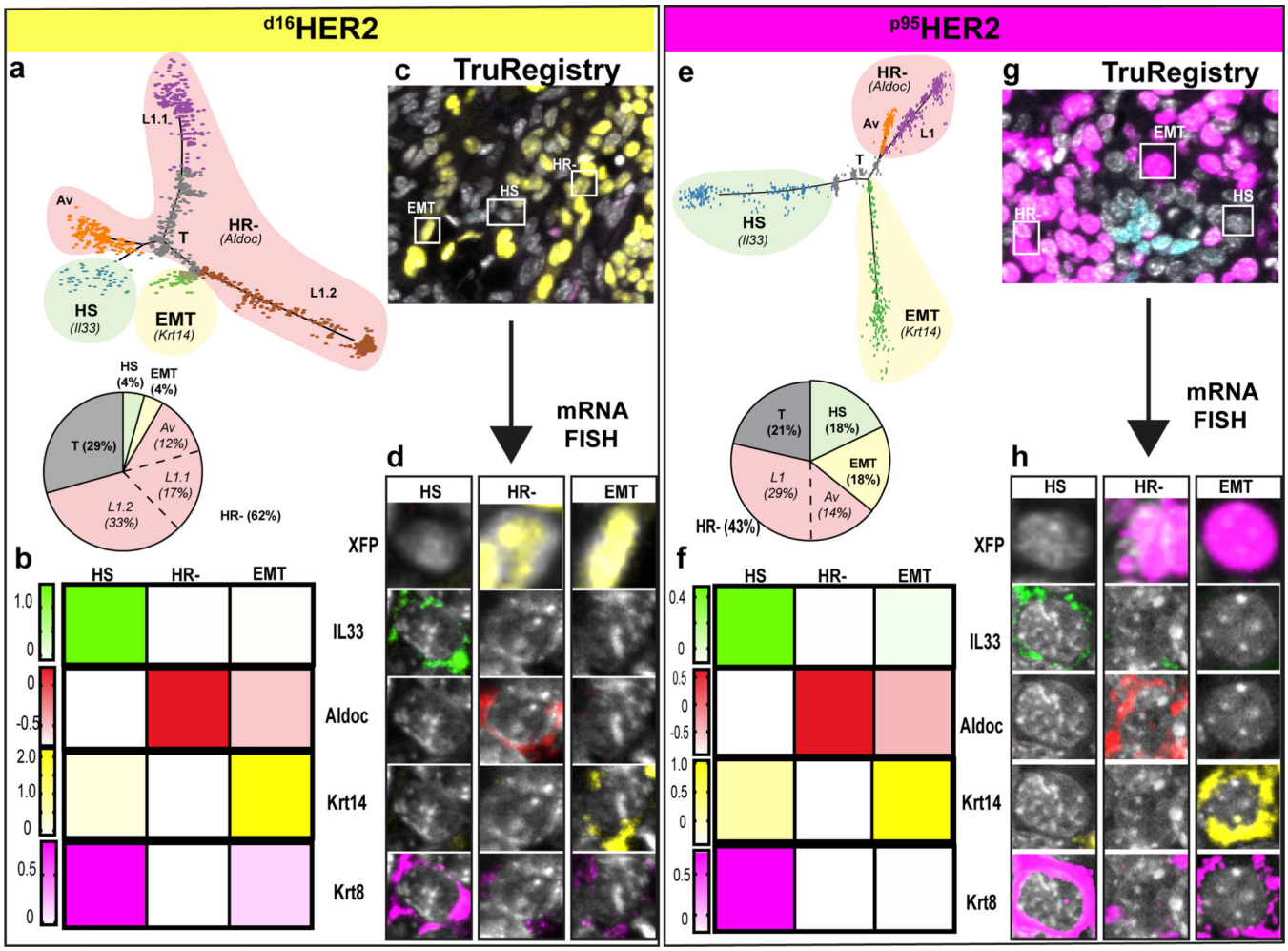


Figure 6.

HER2 isoforms specify unique epithelial cell trajectories. (a) Trajectory mapping of the ^{d16}HER2:EYFP epithelial compartment. Terminal branches were categorized into three cell states each with a prominent marker gene: Hormone sensitive (HS, green shade, *Il33*), Hormone receptor negative (HR-,red shade, *Aldoc*), and Epithelial to mesenchymal-like transition (EMT, yellow shade, *Krt14*). The HR- group could be further classified into Alveolar (Av, orange), Luminal 1.1 (L1.1, purple), and Luminal 1.2 (L1.2, brown). Multiple intermediate states were consolidated into a single transitional state (grey). (b) Heatmap of *Il33*, *Aldoc*, *Krt18*, and *Krt14* expression in the HS, HR-, and EMT branches. (c) Truregistry of Crainbow XFPs and (d) mRNA FISH of *Il33*, *Aldoc*, *Krt8*, and *Krt14* in a ^{d16}HER2:EYFP tumor. (e) Trajectory mapping of the ^{p95}HER2:mKO epithelial compartment. Terminal branches were categorized into three cell states each with a prominent marker gene: Hormone sensitive (HS, green shade, *Il33*), Hormone receptor negative (HR-,red shade, *Aldoc*), and Epithelial to mesenchymal-like transition (EMT, yellow shade, *Krt14*).The HR- group could be further classified into Alveolar (Av, orange) or Luminal (L1, purple). Multiple intermediate states were consolidated into a single transitional state (T, grey). (f) Heatmap of *Il33*, *Aldoc*, *Krt18*, and *Krt14* expression in the

HS, HR-, and EMT branches. (g) Truregistry of Crainbow XFPs and (h) mRNA FISH of *Il33*, *Aldoc*, *Krt18*, and *Krt14* in a ^{p95}HER2:mKO tumor.